Defining cell fate specification of mouse Mammary Stem Cells in 4D

Claudia Carabana Garcia

To cite this version:

Claudia Carabana Garcia. Defining cell fate specification of mouse Mammary Stem Cells in 4D. Cancer. Université Paris sciences et lettres, 2022. English. NNT : 2022UPSLS055 . tel-04223526

HAL Id: tel-04223526
https://theses.hal.science/tel-04223526v1

Submitted on 30 Sep 2023

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Institut Curie
Équipe: La voie de signalisation Notch dans les cellules souches et les tumeurs

Defining cell fate specification of mouse Mammary Stem Cells in 4D

Spécification du destin cellulaire des cellules souches mammaires en 4D

Soutenue par
Claudia CARABANA GARCIA
Le 27 Septembre 2022

Ecole doctorale n° 515
Complexité du Vivant

Spécialité
Biologie Cellulaire

Composition du jury :

Philippe, CHAVRIER DR, Institut Curie, Paris	*Président*
Walid, KHALED Professeur, MRC Cambridge Stem Cell Institute	*Rapporteur*
Felicity Mae, DAVIS Professeure, Aarhus University, Denmark	*Rapporteur*
Lucie, PEDUTO DR, Institut Pasteur, Paris	*Examinatrice*
Silvia, FRE DR, Institut Curie, Paris	*Directrice de thèse*

institut Curie

Table of contents

ABBREVIATIONS **4**

RÉSUMÉ **7**

ABSTRACT **9**

INTRODUCTION **11**

1. COORDINATION BETWEEN BRANCHING MORPHOGENESIS AND CELL FATE SPECIFICATION **12**

2. CELLULAR HETEROGENEITY IN THE MAMMARY GLAND **13**
- 2.1. MAMMARY EPITHELIAL CELLS (MECs) 14
- 2.2. STROMAL CELLS 14
- 2.3. CROSSTALK BETWEEN MG EPITHELIUM AND STROMA 15
- 2.4. MAMMARY STEM CELLS 16

3. MaSC HIERARCHY **17**
- 3.1. TECHNIQUES TO STUDY MAMMARY EPITHELIAL CELL HIERARCHY 17
  - 3.1.1. Transplantation assays to assess the differentiation potential of MaSCs 18
  - 3.1.2. Fluorescence activated cell sorting to isolate mammary cell sub-populations with a combination of specific cell surface markers 19
  - 3.1.3. Lineage tracing to study the hierarchy and potency of MaSCs 20
  - 3.1.4. Single cell technologies to distinguish cell states during embryonic MG development 24
- 3.2. CELLULAR PLASTICITY OF MAMMARY CELLS 27

4. MAMMARY GLAND MORPHOGENESIS **27**
- 4.1. EMBRYONIC BRANCHING MORPHOGENESIS 28
  - 4.1.1. Molecular mechanisms driving embryonic mammary branching morphogenesis 29
    - *4.1.1.1. Specification of the mammary line at E10.5* 29
    - *4.1.1.2. Development of mammary placodes at E11.5* 30
    - *4.1.1.3. Formation of mature mammary bud at E13.5* 31
    - *4.1.1.4. Mammary mesenchyme differentiation at E15.5* 31
    - *4.1.1.5. Initiation of branching morphogenesis at E16.5* 32
  - 4.1.2. Cellular mechanisms driving embryonic mammary branching morphogenesis 32
- 4.2. POST-NATAL BRANCHING MORPHOGENESIS 34
  - 4.2.1. Hormonal signalling initiate post-natal MG development 34
  - 4.2.2. Physical signals are required to pattern ductal elongation 35
  - 4.2.3. Stromal signals mediating tip branching and termination 36

AIM OF THE STUDY **39**

MATERIALS AND METHODS 41

1. Mouse models.......... 42
2. Ethics Statement 42
3. Embryonic mammary gland dissection and *ex vivo* culture.......... 42
4. Mammary *ex vivo* culture whole-mount immunostaining 44
5. Immunofluorescence on 2D sections 44
6. Single molecule RNA fluorescence in situ hybridization (smRNA-FISH) 44
7. Microscopy and image acquisition 45
8. Single cell dissociation of embryonic mammary gland.......... 46
9. Cell labelling, flow cytometry and sorting 46
10. Mammary gland whole-mount carmine staining.......... 47
11. Image analysis and quantification.......... 47
12. scRNA-seq data analysis 48
13. Statistics and Reproducibility 51

RESULTS 53

CHAPTER 1: Cell fate specification underlies positional cues during branching morphogenesis of the embryonic mammary epithelium 54

1.1. Aim of the study.......... 54
1.2. Results 54

CHAPTER 2: Live-imaging of embryonic mammary explants to analyse cellular dynamic behaviour driving mammary gland morphogenesis 71

2.1. Aim of the study.......... 71
2.2. Results 71

CHAPTER 3: Canonical WNT/β-catenin signalling regulates embryonic branching morphogenesis....... 80

3.1. Aim of the study.......... 80
3.2. Results 80

DISCUSSION 89

1. Lineage specification from MaSCs to unipotent progenitors is a progressive process during embryonic development.......... 90
2. Spatial segregation of basal and luminal-like cells during embryonic MG development.......... 93
3. Live-imaging to study branching morphogenesis during embryonic development and multipotency restriction.......... 94
4. Embryonic mammary mesenchymal cells surrounding the epithelial bud represent two spatially distinct stromal cell populations.......... 98
5. Mesenchyme-epithelium communication is involved in mammary branching morphogenesis.......... 100
6. Wnt/β-catenin signalling dictates cell identity and controls branching morphogenesis in the embryonic mammary epithelium.......... 102
7. Link between multipotency of embryonic stem cells and cell plasticity in cancer 103

FUTURE PERSPECTIVES **105**

1. What is the exact timing of the switch from multipotent MaSCs to unipotent progenitors? 106

2. What are the molecular mechanisms regulating cell fate specification? 107

3. Does the subrabasal layer of the skin participate in the formation of the inner (luminal) compartment of the embryonic MG? 108

4. Does cell position in the bud instruct cell fate, or do cells move to a specific tissue region after lineage commitment? 109

5. Does orientation of cell division influence cell fate and positioning in the growing tissue? 109

6. Does cell proliferation drive branching morphogenesis in the embryonic MG? 110

7. Do the distinct mesenchymal cell clusters identified at birth represent spatially restricted stromal cell populations? 111

8. Does constitutive Wnt/β-catenin signalling induce squamous transdifferentiation of embryonic mammary epithelial cells, as found in adult development? 112

9. Can intrinsic signals determine cell fate? 112

ACKNOWLEDGEMENTS **115**

REFERENCES **119**

ANNEX 1 .... **131**

ANNEX 2 .... **163**

Abbreviations

3D	Three-dimensional
α-SMA	alpha- smooth muscle actin
AR	Androgen receptor
ATAC	Assay for transposase-accessible chromatin
AREG	Amphiregulin
BARWs	Branching and annihilating random walks
BCs	Basal cells
BM	Basement membrane
BMP4	Bone morphogenetic protein 4
BSA	Bovine Serum Albumin
CAFs	Cancer-associated fibroblasts
DKK1	Dickkopf-1
DLL1	Notch ligand Delta-like1
DOX	Doxycycline
DTA	Diptheria toxin A
E	Embryonic day
ECM	Extracellular matrix
Eda	Ectodysplasin
EdU	5-ethynyl-2′-deoxyuridine
EGF	Epidermal growth factor
EMT	Epithelial-mesenchymal transition
ERα	Estrogen receptor-alpha
EpCAM	Epithelial cell adhesion marker
FACS	Fluorescence-activated cell sorting
FAP	Fibroblast activation protein
FGF	Fibroblast growth factor
FSP1	Fibroblast specific protein 1
GaAsP	Gallium Arsenide Phosphide
GAM	Generalized additive model
GH	Growth hormone
GLI3	GLI family Zinc Finger 3
HDST	High-definition spatial transcriptomics
HGF	Hepatocyte growth factor

Hsps	Heat shock proteins
IGF1	Insulin-like growth factor 1
IGFR1	Insulin-like growth factor 1 receptor
IVM	Intravital microscopy
K	Keratin
LCs	Luminal cells
Lin^-	Lineage negative
LP	Luminal progenitors
LTBP1	Latent TGFβ-binding protein1
MaSCs	Mammary stem cells
ML	Mature luminal
MECs	Mammary epithelial cells
MEPs	Multipotent embryonic progenitors
MERFISH	Multiplexed error-robust fluorescence in situ hybridization
MG	Mammary gland
MIP	Maximum intensity Z-projection
MMPs	Matrix metalloproteinases
MMTV	Mouse mammary tumour virus
MST	Minimum spanning tree
N1	Notch1
N1IC	Notch1 intracellular
NRG3	Neuregulin 3
OCT	Optimum cutting temperature
P0	Post-natal day 0
PBS	Phosphate buffered saline
PCA	Principal Component Analysis
PDGFRα	Platelet-derived growth factor receptor α
PL	Prolactin
PTHrP	Parathyroid hormone-related protein
PTHR1	Parathyroid hormone 1 receptor
PR	Progesterone receptor
ProCR	Protein C receptor
PyMT	Polyoma middle T
R26	Rosa26
RT	Room temperature
RTK	Receptor tyrosine kinase

SCENIC	Single-Cell regulatory Network Inference and Clustering
SHH	Sonic hedgehog
scRNA-seq	Single cell RNA-sequencing
scATAC-seq	Single cell ATAC-sequencing
SCs	Stem cells
snATAC-seq	Single nucleus ATAC-sequencing
SNR	Signal-to-noise ratio
TAM	Tamoxifen
TBX3	T-box transcription factor 3
TEB	Terminal end bud
TGFBR	Transforming growth factor beta receptor
TNF	Tumour necrosis factor
TA	Transiently amplifying
UMAP	Uniform Manifold Approximation and Projection
UMI	Unique molecular identifiers
VEGFA	Vascular endothelial growth factor A
WAP	Whey acidic protein
WT	*Wild-type*

Résumé

Au cours du développement, une coordination entre la spécification du destin cellulaire et la morphogenèse tissulaire est nécessaire à la formation des organes. Par conséquent, la façon dont différents types de cellules se différencient dans le temps et l'espace reste une question majeure en biologie du développement.

La glande mammaire est constituée d'un épithélium ramifié composé d'une couche externe de cellules basales (BCs) et d'un compartiment interne de cellules luminales (LCs). Dans ce tissu, l'homéostasie adulte est exclusivement maintenue par des progéniteurs unipotents, alors que des cellules souches mammaires multipotentes (MaSC) ne se trouvent que dans la glande embryonnaire, ce qui en fait un paradigme tissulaire idéal pour étudier leur comportement dynamique et leur contribution à la morphogenèse tissulaire.

HYPOTHÈSE DE TRAVAIL

Nous avons récemment montré que les cellules souches mammaires deviennent unipotentes lors de la tubulogenèse de la glande, suggérant un lien étroit entre spécification cellulaire et morphogenèse. Notre hypothèse implique que la perte de multipotence soit liée aux remaniements cellulaires qui guident la morphogenèse. Cependant, le moment exact et les mécanismes responsables du passage de la multipotence à l'unipotence sont encore inconnus.

OBJECTIF ET MÉTHODOLOGIE

L'objectif principal de ce projet était d'intégrer les circuits transcriptionnels définissant le potentiel de différenciation des cellules souches à l'analyse de la morphogenèse cellulaire et tissulaire en temps réel, afin de déchiffrer les mécanismes qui sous-tendent l'acquisition de l'identité cellulaire et d'étudier la coordination entre spécification des cellules souches et morphogenèse. Nous avons abordé cet objectif ambitieux en combinant deux approches : 1) des analyses transcriptomiques en cellule unique intégrées dans l'espace, au cours du développement mammaire, et 2) une approche de suivi du lignage en temps réel dans des cultures mammaires embryonnaires *ex vivo*, afin d'étudier les comportements et réarrangements cellulaires au cours des premières phases de la croissance mammaire.

RÉSULTATS

Nous avons constaté que la restriction du potentiel des cellules souches est progressive pendant le développement embryonnaire. Par transcriptomique sur cellules uniques, nous avons pu distinguer trois types cellulaires distincts très tôt (E15). L'intégration dans l'espace de ces données transcriptomiques a montré que des cellules de type BC et LC sont positionnées différemment dans le bourgeon mammaire très précocement. Cette analyse a donc révélé de nouveaux marqueurs moléculaires des LC et BC embryonnaires, qui ne peuvent pas être distinguées avec les marqueurs mammaires adultes connus.

De plus, nous avons défini les signatures transcriptionnelles qui distinguent deux populations de cellules stromales mammaires embryonnaires délimitées dans l'espace. Nous avons aussi montré qu'une signalisation paracrine du mésenchyme aux cellules épithéliales, via FGF10-FGFR2, influence la morphogenèse de la glande mammaire.

Nous avons ensuite développé une approche d'analyse d'images semi-automatique aux niveaux cellulaire et tissulaire dans nos explants mammaires analysés par microscopie time-lapse. Nous montrons que les étapes initiales de la morphogenèse sont caractérisées par des réarrangements cellulaires très dynamiques. De plus, l'activation épithéliale de la beta-caténine empêche la formation de branches, indiquant que Wnt est un régulateur essentiel de la morphogenèse mammaire.

Ces travaux mettent en lumière les circuits transcriptionnels qui régissent le branchement mammaire et qui lient la différenciation des cellules souches à leur dynamique cellulaire pendant la morphogenèse. Les mécanismes ainsi dévoilés nous fournissent des biomarqueurs potentiels du cancer du sein, qui résulte souvent de la réactivation de programmes de multipotence embryonnaires.

Mots clés : Cellules souches mammaires, Traçage du lignage, Développement mammaire, Imagerie en temps réel, Séquençage de cellules uniques.

Abstract

Coordination of cell fate specification and branching morphogenesis is necessary to generate an organ with its specialized final structure and function. Accordingly, how different cell types are specified in a tightly regulated manner in time and space, in order to drive the morphogenesis of a complex tissue, remains a major question in the field of developmental biology.

The mammary gland (MG) consists in a branched bi-layered epithelium composed of an outer layer of basal cells (BCs) and an inner compartment of polarised luminal cells (LCs). In this tissue, adult homeostasis is exclusively maintained by lineage-restricted unipotent progenitors, whereas multipotent mammary stem cells (MaSCs) are only found in the embryonic gland, making it an ideal tissue paradigm to study stem cell dynamics and lineage specification, as well as their contribution to tissue morphogenesis.

WORKING HYPOTHESIS

Our recent results showed that multipotent MaSCs become lineage-restricted around embryonic day E15.5, coinciding with the first morphogenetic events that establish the mammary ductal network. We thus hypothesized that loss of multipotency in the mammary gland was linked to cell rearrangements, leading to the branching of embryonic mammary buds. However, the exact timing and the mechanisms responsible for the switch from multipotency to unipotency during embryonic MG are still unknown.

AIM AND METHODOLOGY

The overarching aim of this project was to characterise the stem cell dynamics underlying MaSCs differentiation during MG development, and to define the transcriptional signals underpinning this process. We have approached this ambitious objective combining two approaches: 1) single-cell RNA sequencing analysis at different embryonic times, to discover which signals determine cell identity during mammary development, and 2) a live lineage tracing approach in *ex vivo* embryonic mammary cultures to study dynamic cell behaviours and rearrangements during the earliest phases of mammary growth.

RESULTS

We found that lineage restriction is a progressive developmental process. By single cell transcriptomics, we identified a single population of mammary epithelial cells at E13.5, but we could distinguish three transcriptionally distinct cell subsets at E15.5, which included luminal-like, basal-like and hybrid cells co-expressing luminal and basal genes. Spatial transcriptomic analysis revealed that the basal-like and luminal-like clusters were indeed already spatially restricted in the embryonic mammary bud, being positioned either in close proximity to the basement membrane or in the inner bud region, respectively. Importantly, this analysis revealed novel molecular markers of committed LCs and BCs, that cannot be distinguished with known adult MG markers.

Additionally, we report the transcriptional signatures distinguishing two spatially restricted embryonic mammary mesenchymal cell populations, representing sub-epithelial and dermal mesenchyme. Long-term live-imaging revealed that paracrine signalling from embryonic mesenchyme to epithelial cells, via FGF10-FGFR2, influences epithelial branching.

We then developed a deep learning-based pipeline to semi-automatically track individual cells and tissue branches in embryonic mammary explants analysed by time-lapse microscopy. We show that the initial steps of morphogenesis are characterized by highly dynamic cell rearrangements in the growing branch tips. However, forced activation of the Wnt/β-catenin pathway in the embryonic mammary epithelium precluded branching *in vivo* and *ex vivo*, indicating that epithelial Wnt signalling is an essential regulator of mammary branching morphogenesis.

This work sheds light on the timing and mechanisms governing mammary cell fate decisions, providing potential biomarkers of breast cancer, which often arises from reactivation of embryonic multipotency programs.

Key words: Mammary stem cells, Lineage tracing, Mammary gland development, Time-lapse imaging, Single cell RNA sequencing.

Introduction

1. COORDINATION BETWEEN BRANCHING MORPHOGENESIS AND CELL FATE SPECIFICATION

Branching morphogenesis is a fundamental developmental program that generate arborized tree-like networks with the ultimate objective of maximising the surface of exchange between the epithelium and its lumen. Experimentalists and theoreticians have tried to understand the rules for generating these complex structures for centuries (Hannezo & Simons, 2019). Each organ has evolved to have its unique branching behaviour. It can occur in a stereotyped manner, as seen in the lung (Metzger et al., 2008) or kidney (Short et al., 2014), where an identical order of branching events happens; or in a stochastic manner, as in the mammary gland (MG), where each MG presents slight differences in the arborized structure (Hannezo et al., 2017; Myllymäki & Mikkola, 2019).

Branching morphogenesis is intimately linked with cell fate acquisition to generate an organ with its specialized final architecture and function. Several studies have started to uncover the crosstalk between the transcriptional changes and the physical mechanisms that are used to generate different branched tissues during embryonic development (Chan et al., 2017). For instance, smooth muscle differentiation plays an important role in shaping emerging domain branches in the mouse lung (Goodwin et al., 2019). FGF10 expressed in the distal mesenchyme drives sonic hedgehog (SHH) expression in the epithelium through FGFR2b. SHH, released by the epithelium, promotes differentiation of smooth muscle, whereas FGF9 inhibits smooth muscle differentiation (Goodwin & Nelson, 2020; McCulley et al., 2015; Morrisey & Hogan, 2010).

In the salivary gland, branching morphogenesis is initiated when mesenchymal FGF10 induces specification of distally located progenitors expressing SOX9 (Chatzeli et al., 2017; Myllymäki & Mikkola, 2019).

In the mammary gland, our lab has recently demonstrated that cell fate specification occurs progressively during embryonic development, coinciding with the first morphogenetic events (Lilja et al., 2018). Nonetheless, **the identity and origin of the signals involved in cell fate specification are still unknown, as is the orchestration of fate acquisition with branching morphogenesis to generate a functional mammary gland**.

In this Introduction to my PhD thesis, I will highlight recent advances in our understanding of how cell fate specification occurs and how branching morphogenesis is regulated, in particular in the mammary gland, the tissue I studied during my PhD, with the ultimate aim of understanding how these two processes might be coordinated.

2. CELLULAR HETEROGENEITY IN THE MAMMARY GLAND

The mammary gland is the defining biological feature of all mammals. It is a secretory organ composed of an elaborate network of branching ducts embedded within an adipocyte-rich stroma. Each mammary duct is formed by a bilayered epithelium, consisting of an inner layer of luminal cells (LCs) facing the ductal lumen, and an outer layer of contractile myoepithelial basal cells (BCs) that interface directly with the basement membrane (BM) and the stromal cells (Figure 1).

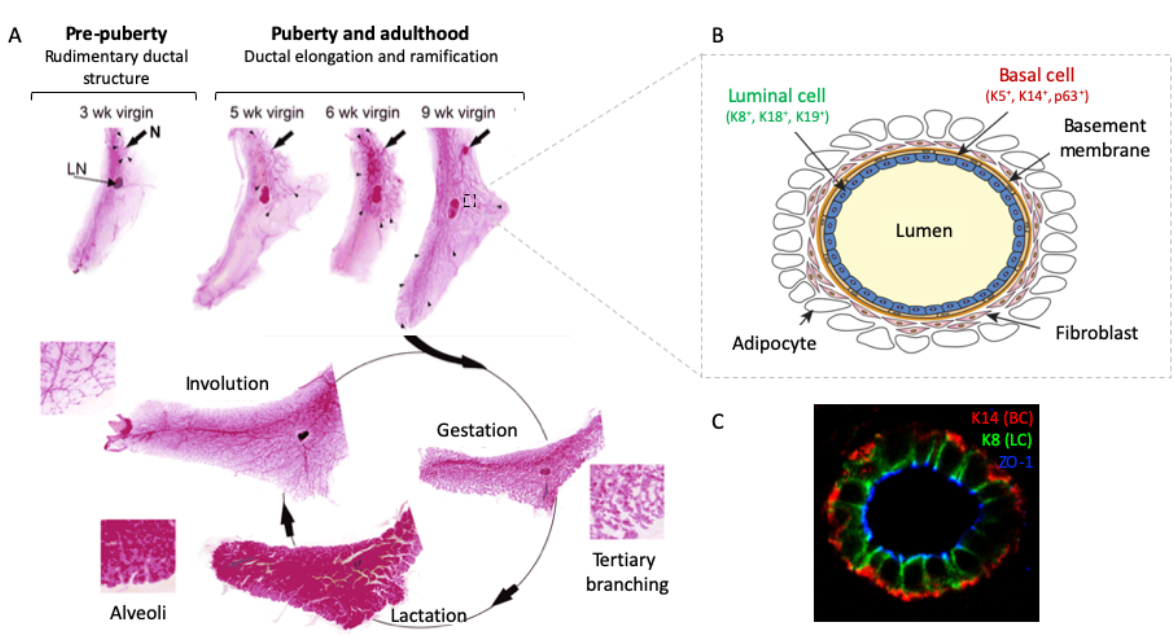


Figure 1 | Mammary gland postnatal development and cellular composition.
(A) Carmine-red staining to visualize the epithelial tree of the MG. Bold arrows indicate the position of the nipple (N), thin arrow indicates the lymph node (LN) and black arrowheads indicate ductal ends. During pre-pubertal stages MG remains relatively quiescent until puberty, when terminal end buds (TEBs) will drive ductal elongation under hormonal regulation. The ductal ends reach the edge of the fat pad before 9 weeks of age. Full mammary gland maturation is achieved during lactation with the appearance of alveolar units. In the absence of milk production, MG undergoes involution to a virgin-like state. Adapted from (Veltmaat, 2012). (B) Schematic representation of a MG duct and the different cell types composing the mammary fat pad. Adapted from (Fu et al., 2020). (C) Immunostaining of a MG duct illustrating K14$^+$ basal cells (in red), K8$^+$ luminal cells (in green) and ZO-1$^+$, a tight junction protein expressed by polarized cells facing the lumen (in blue).

Mammary epithelial cells (MECs) cooperate to achieve their function of nursing the offspring by producing and secreting milk. During gestation and lactation, the luminal cells undergo lactogenic differentiation and secrete milk proteins such as β-casein, whey acidic protein (WAP) and β-lactoglobulin. The myoepithelial cells contract – in response to Ca^{2+} signalling – in order to ease the flow of milk produced from secretory LCs in the alveolar units and allow its passage through the mammary ductal network (Stevenson et al., 2020).

Despite this apparently simple epithelial organization and minimal cellular hierarchy, the presence and potency of stem cells in this tissue is still under debate, as discussed below.

2.1. MAMMARY EPITHELIAL CELLS (MECs)

As mentioned above, the mammary epithelium is composed of two cell types. BCs are characterized by the expression of specific structural proteins like keratin (K) 5 and K14 and the transcription factor P63. They also express alpha-smooth muscle actin (α-SMA), mediating their contractile function to guide milk secretion throughout the ductal tree and nipple during lactation (Inman et al., 2015).

LCs are characterized by the expression of specific structural proteins like K8, K18 and K19. The luminal lineage can be further divided into two independent and self-sustained cell types, depending on the expression of estrogen receptor-α (ERα) and progesterone receptor (PR). LCs marked by the expression of ERα and PR are designated as hormone-sensing cells, since they perceive the presence of circulating hormones and induce proliferation of neighbouring cells, referred to as hormone-responsive cells, that lack expression of ERα and PR (Beleut et al., 2010; Mallepell et al., 2006). Hormone-sensing cells are also called Luminal Mature cells (LM) whereas hormone-responsive cells are considered Luminal Progenitors (LP).

2.2. STROMAL CELLS

The mammary stromal compartment is composed of different cell types, including fibroblasts, adipocytes that constitute the fat pad, vascular endothelial cells ensuring the vascularization of the tissue, immune cells and nerves, that work together to maintain a functional organ (Figure 1). The stromal compartment is the source of many instructive signals involved in the regulation of mammary epithelial development and function (Muschler & Streuli, 2010).

Interestingly, there are two types of mesenchymal cells in the embryonic mammary gland: the primary or fibrogenic mammary mesenchyme, composed of fibroblast-like cells that surround the mammary epithelium and the fat pad precursor mesenchyme, composed of pre-adipocytes, which are found posterior to the mammary epithelial rudiment (Sakakura et al., 1982).

Historically, the embryonic mammary mesenchyme has been overlooked and little is known about mammary fibroblast lineages. The presumptive dual origin of the embryonic mammary mesenchyme may lead one to think that there are two lineages of fibroblasts of distinct embryonic origin in the adult mammary gland (Sumbal et al., 2021). However, lineage tracing studies have not yet been performed and are needed to answer this question. Fibroblast lineage tracing has been limited due to the lack of unique markers that are not expressed in other cell types. For instance, vimentin or α-SMA are also expressed by myoepithelial basal cells. Some fibroblast-specific markers include Fibroblast specific protein 1 (FSP1) and Platelet-derived growth factor receptor α (PDGFRα). Other proposed markers are only expressed in specific subsets of fibroblasts, such as Fibroblast Activation Protein (FAP), which is expressed during bone and fat homeostasis (Sahai et al., 2020).

Based on their activity during tissue formation and homeostasis, fibroblasts can be divided in three subpopulations: normal or resting fibroblasts, activated fibroblasts or myofibroblasts, and cancer-associated fibroblasts (CAFs). Moreover, in the MG they can be located surrounding the epithelium, separating the ducts from the adipose stroma, referred as periductal fibroblasts, or dispersed between adipocytes, called interstitial fibroblasts, usually found in close proximity to perivascular fibroblasts (Sumbal et al., 2021).

In adult animal homeostasis, fibroblasts are the main producers of connective tissue, as they synthetize several extracellular matrix (ECM) components, such as collagens, proteoglycans and fibronectin. Fibroblasts also synthesize enzymes that remodel the microenvironment, such as matrix metalloproteases that help releasing growth factors and cytokines embedded within the ECM (Inman et al., 2015). Interestingly, fibroblasts also play crucial roles by communicating with other cell types during both tissue homeostasis and upon tissue damage. They can promote angiogenesis via the production of Vascular endothelial growth factor A (VEGFA), or coordinate immune responses via the production of chemokines and cytokines (Sahai et al., 2020).

Regarding the appearance of adipocytes in the mammary fat pad, the first lipid-filled adipocytes are observed two days after birth (Sakakura et al., 1982). In the adult and non-lactating gland, adipocytes with high lipid content comprise a large proportion of the stromal fat pad. During pregnancy and lactation, adipocytes reduce their lipid content, suggesting that this fat reservoir is necessary for the metabolically-demanding process of milk production (Inman et al., 2015).

Collectively, the mammary stroma can dynamically change in cellular composition during different developmental stages. Indeed, there is emerging data that cell proportions and gene expression of the different mammary stromal cell types change with age. For instance, in aged fibroblasts the expression of ECM-related genes is decreased in comparison with fibroblasts in young tissues, suggesting an impaired ability to maintain the stromal matrix (C. M. C. Li et al., 2020).

2.3. CROSSTALK BETWEEN MG EPITHELIUM AND STROMA

Crosstalk between the mammary epithelium and stroma is tightly regulated for the proper patterning and function of the MG (Wiseman & Werb, 2002).

The earliest evidence of the influence of the mesenchyme on mammary epithelial development came from tissue recombination experiments. Firstly, trypsin-isolated mammary epithelial rudiments did not develop in the absence of mesenchyme. When these isolated mammary epithelial rudiments were recombined with mammary mesenchyme *ex vivo*, they could develop successfully. Interestingly, when E12 and E16 mammary gland epithelium were recombined with salivary gland mesenchyme, the pattern of the resulting glandular structure was morphologically similar to that of the salivary gland (Kratochwil, 1969).

Further proof that the mammary mesenchyme provides key instructive signals that can induce morphological and functional mammary differentiation from non-mammary epithelium was obtained by Cunha et al. (Cunha et al., 1995). They performed tissue recombination experiments using E13 rat skin epithelium with E13 mouse mammary mesenchyme and were grafted under the renal capsule of female mouse hosts. In these tissue recombinants, mammary epithelial ductal structures derived from rat cells extended into the host mesenchyme. Moreover, when these female hosts were induced to undergo lactogenesis, rat epithelial cells in the tissue recombinants differentiated into milk-producing lobuloalveolar structures able to produce milk proteins (Cunha et al., 1995).

Reciprocally, it has also been shown that the mammary epithelium influences the differentiation of mesenchymal cells, by inducing its condensation to form several layers of fibroblast-rich cells by E14.5 (Wysolmerski et al., 1998).

Taken together, **the reciprocal inductive interactions between epithelium and mesenchyme are necessary to guide mammary cell fate specification** (Wansbury et al., 2011). Several molecular mechanisms have been postulated to mediate this crosstalk during the different stages of MG development. Some mechanistic insights into the molecular regulatory signals involved in these interactions are presented in the next section of this manuscript (*Introduction Chapter 4*).

2.4. MAMMARY STEM CELLS

Tissue-specific stem cells (SCs) are defined by their ability to self-renew and differentiate into the specialised cell types of their tissue of origin. SCs also contribute to tissue morphogenesis during development (van Keymeulen & Blanpain, 2012). SCs specify to the different cell types that compose the different tissues, generating the cellular diversity required by adult tissues to perform their distinct functions. After this stage, tissues stop expanding and enter a steady-state condition, characterised by the balance between cell proliferation and cell loss. The process of constant cell replacement is referred to as tissue homeostasis and is tightly controlled, as its deregulation can lead to tissue atrophy or cancer formation (Blanpain et al., 2007). Some epithelia have a very high cellular turnover in order to replace the cells that are continuously lost, such as the intestine or the skin (Barker et al., 2010), while other tissues renew at a slower pace in physiological conditions, such as the airway tracts that can take as long as 6 months to be replaced (Rock & Hogan, 2011). In addition, tissue-specific SCs are also activated upon injury in order to replace damaged cells (Blanpain et al., 2007).

Adult SCs give rise to proliferating progeny, referred to transiently amplifying (TA) cells or progenitor cells. These progenitor cells will actively divide for a restricted period of time, expanding the cellular pool that will then differentiate into a particular cell lineage (Blanpain et al., 2007).

With regard to the mammary gland, this tissue undergoes a massive expansion at puberty and is heavily remodelled during the cycle of pregnancy, lactation and involution. Both mammary morphogenesis and repeated tissue remodelling at each reproductive cycle implicate the existence of mammary stem cells (MaSCs). However, due to the lack of defined markers, the identity, differentiation potential and location of these cells within the mammary epithelium remains elusive.

3. MaSC HIERARCHY

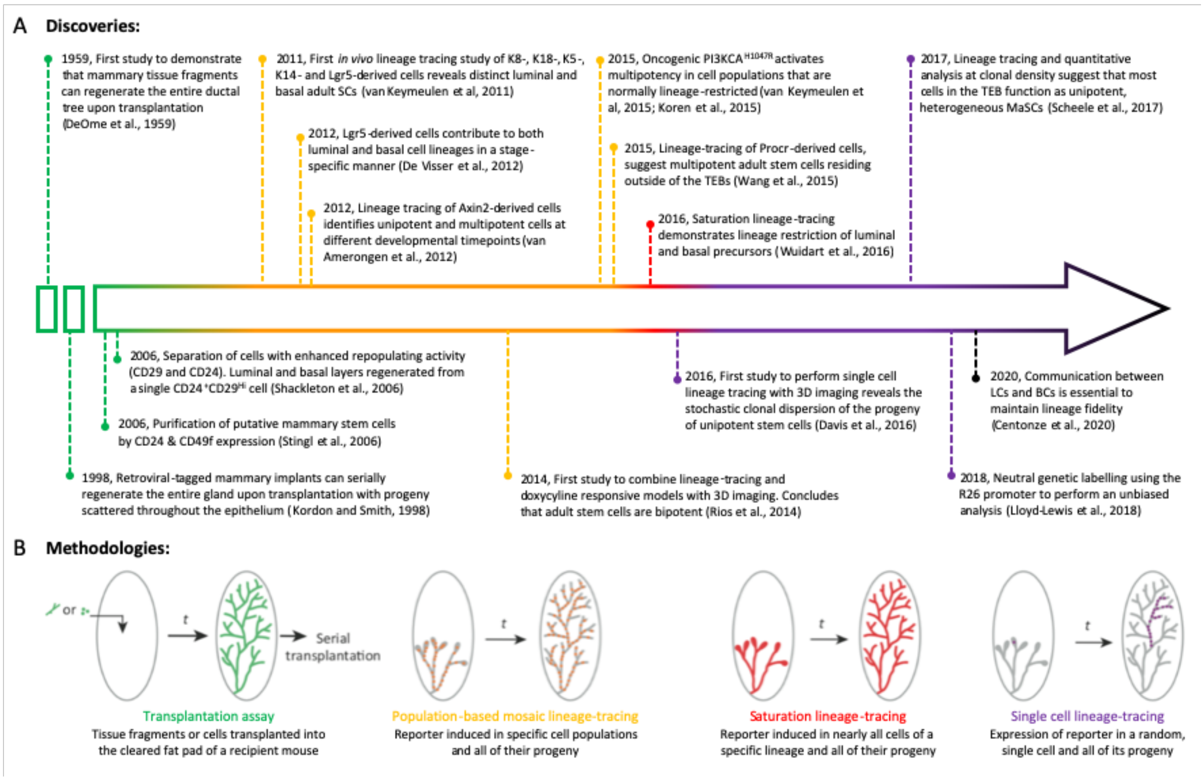
Clonal analyses based on lineage tracing approaches have demonstrated the existence of multipotent adult SCs able to differentiate into all different cell types of their tissue of origin throughout adulthood in some tissues, for example in the intestine (Barker et al., 2010). However, recent studies have uncovered that exclusively unipotent progenitors can sustain the development of many tissues, such as in the interfollicular epidemis of the skin (Clayton et al., 2007), sweat glands (C. P. Lu et al., 2012), sebaceous glands (Horsley et al., 2006), or prostate (Choi et al., 2012; Tika et al., 2019). In the mammary gland, the existence of adult multipotent MaSCs has generated considerable debate over recent years.

Furthermore, it is important to define the different subsets of stem and progenitor cells, because there is still no consensus in the terminology used. In the present manuscript, we call MaSC a SC that is multipotent (or at least bipotent), whereas progenitor cells are unipotent lineage-restricted cells; both these cells possess long-term self-renewal capacity and replicative potential (Watson & Khaled, 2020).

Understanding the hierarchical relationships between stem cells and lineage-committed progenitors that guide MG morphogenesis is also important to define the cellular origin of specific breast cancer types.

3.1. TECHNIQUES TO STUDY MAMMARY EPITHELIAL CELL HIERARCHY

The study of MaSCs hierarchy has lasted for more than 60 years. Three key techniques have enabled mammary gland biologists to investigate lineage commitment and stem cell characteristics (Figure 2): the isolation of each mammary epithelial cell type based on cell-surface markers for transplantation studies, the generation of lineage-specific mouse models for lineage tracing studies, and the development of transcriptomic and epigenetic analyses (Twigger & Khaled, 2021).



1959, First study to demonstrate that mammary tissue fragments can regenerate the entire ductal tree upon transplantation (DeOme et al., 1959)

2011, First *in vivo* lineage tracing study of K8-, K18-, K5-, K14- and Lgr5-derived cells reveals distinct luminal and basal adult SCs (van Keymeulen et al, 2011)

2012, Lgr5-derived cells contribute to both luminal and basal cell lineages in a stage-specific manner (De Visser et al., 2012)

2012, Lineage tracing of Axin2-derived cells identifies unipotent and multipotent cells at different developmental timepoints (van Amerongen et al., 2012)

2015, Oncogenic PI3KCA^{H1047R} activates multipotency in cell populations that are normally lineage-restricted (van Keymeulen et al, 2015; Koren et al., 2015)

2015, Lineage-tracing of Procr-derived cells, suggest multipotent adult stem cells residing outside of the TEBs (Wang et al., 2015)

2016, Saturation lineage-tracing demonstrates lineage restriction of luminal and basal precursors (Wuidart et al., 2016)

2017, Lineage tracing and quantitative analysis at clonal density suggest that most cells in the TEB function as unipotent, heterogeneous MaSCs (Scheele et al., 2017)

2006, Separation of cells with enhanced repopulating activity (CD29 and CD24). Luminal and basal layers regenerated from a single CD24^{+}CD29^{Hi} cell (Shackleton et al., 2006)

2006, Purification of putative mammary stem cells by CD24 & CD49f expression (Stingl et al., 2006)

1998, Retroviral-tagged mammary implants can serially regenerate the entire gland upon transplantation with progeny scattered throughout the epithelium (Kordon and Smith, 1998)

2014, First study to combine lineage-tracing and doxycyline responsive models with 3D imaging. Concludes that adult stem cells are bipotent (Rios et al., 2014)

2016, First study to perform single cell lineage tracing with 3D imaging reveals the stochastic clonal dispersion of the progeny of unipotent stem cells (Davis et al., 2016)

2020, Communication between LCs and BCs is essential to maintain lineage fidelity (Centonze et al., 2020)

2018, Neutral genetic labelling using the R26 promoter to perform an unbiased analysis (Lloyd-Lewis et al., 2018)

Figure 2 | Summary of the key discoveries in the study of MaSC hierarchy during adult MG development. (A) Timeline with the key discoveries presented in this work. (B) *In vivo* methodology used in these studies: transplantation assays (in green), population-based mosaic lineage-tracing (in orange), lineage-tracing at saturation (in red) and single cell lineage-tracing (in purple) (Adapted from (Lloyd-Lewis et al., 2017)).

In addition, emerging techniques to visualize individual cells in real-time either during post-natal mammary gland development by the surgical insertion of an imaging window for intravital microscopy (IVM), or during embryonic development by three-dimensional (3D) *ex vivo* cultures of embryonic mammary buds, together with barcode-based lineage tracing approaches, have the potential to finally unravel the mammary epithelial cell hierarchy (Watson, 2021).

Here, I summarize the main findings obtained up to now with each of these methodologies.

3.1.1. Transplantation assays to assess the differentiation potential of MaSCs

Exploration of the remarkable regenerative capacity of the MG has largely relied on transplantation experiments (Figure 2). In 1959, a seminal study by DeOme et al. postulated for the first time the existence of MaSCs. They demonstrated that an entire mammary ductal epithelium can be regenerated by transplantation of any epithelial fragment into the cleared fat pad of a recipient mouse, suggesting that mammary epithelium encompasses self-renewing stem cells (DeOme et al., 1959). This technique, called the mammary reconstitution assay, consists of surgically removing the region of the fat pad containing the endogenous mammary epithelium present in 3-week old pre-pubertal mice, followed by the injection of mammary tissue fragments or single cell suspensions from a donor animal.

Subsequent studies performing grafting of mammary tissue fragments from mice infected by the mouse mammary tumour virus (MMTV) into the cleared fat pad of a non-infected mice, confirmed that mammary cells with repopulating capacities were distributed throughout the adult ductal epithelium rather than localized in specific regions of the mammary tree (Kordon & Smith, 1998).

Collectively, these results supported the existence of multipotent adult MaSCs, revealed when such cells were removed them from their physiological environment.

3.1.2. Fluorescence activated cell sorting to isolate mammary cell sub-populations with a combination of specific cell surface markers

After the results obtained with transplantation assays, the main focus of researchers shifted to the identification and isolation of discrete MECs by fluorescence activated cell sorting (FACS) and their subsequent transplantation in a cleared mammary fat pad, to assess their differentiation potential.

The FACS technique is used to conduct multiparameter profiling and sorting of live cells based on fluorescent labelling of membrane proteins that bind to fluorophore-conjugated antibodies. Through a combination of cell surface markers, specific cell populations can be selected and recovered from a tissue (Twigger & Khaled, 2021).

The cell surface markers commonly used to isolate MECs are listed below. First, epithelial cells are negatively selected using markers against different stromal cell types (CD45 and Ter119 for hematopoietic and immune cells and CD31 for endothelial cells), known as Lineage negative (Lin$^-$). Within the MECs gate, BCs and LCs can be distinguished based on their differential expression levels of CD24 (heat stable antigen) or EpCAM (Epithelial cell adhesion marker) and CD29 (integrin β1) or CD49f (integrin α6). BCs correspond to the CD24$^+$CD29^{Hi} (Shackleton et al., 2006) or EpCAM$^-$CD49f$^+$ (Stingl et al., 2006) populations, whereas LCs correspond to CD24$^+$CD29^{Lo} cells, which can be further subdivided in mature luminal (ER$^+$/PR$^+$) cells (EpCAM$^+$/CD49f$^-$) and luminal progenitor (ER$^-$/PR$^-$) cells (EpCAM$^+$/CD49f$^+$).

Remarkably, Shackleton et al. initially observed that only Lin$^-$CD24$^+$CD29^{Hi} (BCs) single cells could reconstitute a completely functional mammary gland upon transplantation, supporting the hypothesis that the adult mouse mammary gland contains multipotent stem cells within the basal compartment. They demonstrated the self-renewing capacity of Lin$^-$CD24$^+$CD29^{Hi} BCs by serial transplantations, and their multi-lineage differentiation capacity by staining for luminal and myoepithelial markers in histological sections (Shackleton et al., 2006).

Interestingly, when BCs are transplanted together with LCs at non-limiting dilutions in a cleared fat pad, each epithelial cell lineage participate to the reconstitution of a functional mammary gland by generating exclusively its own lineage, demonstrating a strictly unipotent behaviour (van Keymeulen et al., 2011). Later studies showed that isolated luminal progenitors (LP) were also able to demonstrate multipotent capacity in transplantation assays (Regan et al., 2012; Rodilla et al., 2015), indicating that this technique reveals the multipotent potential and extent of plasticity of a given cell – defined as its ability to reprogram and reactivate multipotency programs – and not its physiological lineage potential in homeostatic conditions.

Furthermore, a recent study by Centonze et al. set out to investigate whether an active mechanism restricts MECs multipotency under normal homeostatic conditions. To this aim, they genetically ablated LCs both *in vivo* and in organoids and monitored to reactivation of multipotency of BCs by lineage tracing as a direct consequence of sensing the loss of ablated LCs. The targeted ablation of LCs by a diptheria toxin A (DTA)-based approach illustrated the impressive cellular plasticity of mammary epithelial cells, causing reactivation of multipotency in otherwise unipotent BCs. Mechanistically, Notch, Wnt and the Epidermal growth factor receptor (EGFR) signalling pathways were all found to be important for the observed reactivation of multipotency upon LC ablation. These elegant experiments brought the authors to propose that LCs restrict BCs multipotency under physiological conditions and one proposed mechanism involves the secretion of the cytokine tumour necrosis factor (TNF). These recent studies therefore corroborated the hypothesis of an active communication between BCs and LCs, that is necessary to maintain lineage fidelity and restrict multipotency of MaSCs (Centonze et al., 2020).

In conclusion, the work of several labs during the past 15 years has now unequivocally established that the adult MG is maintained exclusively by unipotent basal and luminal progenitors and that transplantation experiments assess the capacity of a cell to be reprogrammed to a regenerative multipotent state (i.e. its plasticity) and not its potency in physiological and homeostatic conditions.

3.1.3. Lineage tracing to study the hierarchy and potency of MaSCs

Following the results obtained by transplantation studies, lineage tracing approaches using lineage-specific promoters became the gold standard to assess SC potency and fate within their natural environment *in vivo.* In the MG, lineage tracing is been used by a growing number of laboratories to study stem cell differentiation and renewal properties in physiological conditions (van Keymeulen & Blanpain, 2012). This strategy has been instrumental to demonstrate that the MG is maintained exclusively by unipotent cell populations (van Keymeulen et al., 2011; Wuidart et al., 2016).

Genetic lineage tracing techniques allow the genetic tagging of individual cells and their progeny *in vivo*, thus avoiding most of the drawbacks associated with transplantation experiments (van

Keymeulen & Blanpain, 2012). Numerous transgenic mouse models have been used to label specific cell populations in a spatially and temporally controlled manner.

Lineage tracing experiments generally require the expression of two transgenes: the first one expresses a temporally-inducible form of the Cre recombinase under the control of a lineage-specific promoter. To achieve temporal control of the system, the Cre recombinase can be activated upon tamoxifen (TAM) or doxycycline (DOX) administration, depending on the system used.

The second essential element is a reporter gene, allowing tracking of targeted cells. A reporter gene can encode for the beta-galactosidase enzyme, or for a fluorescent protein. The transgenic construct harbours a loxP-STOP-loxP sequence. The STOP cassette that precedes the reporter gene prevents its expression in basal conditions. However, after Cre activation, the loxP-STOP-loxP sequence is excised, allowing targeted cells to be genetically and irreversibly labelled, as well as their progeny (or lineage), facilitating their visualization and tracking over time *in vivo* and *in situ* (Kretzschmar & Watt, 2012; Rodilla & Fre, 2022; van Keymeulen & Blanpain, 2012).

The first study using lineage tracing in the MG to assess the fate of MaSCs was conducted using transgenic mice expressing inducible Cre^{ERT2} either in the luminal lineage (using the K8 and K18 promoters) or in the basal lineage (using the K5 and K14 promoters). This approach allowed the authors to **demonstrate for the first time that tissue homeostasis is exclusively maintained by unipotent progenitors in the post-natal mammary gland** (van Keymeulen et al., 2011).

Subsequently, several groups have tackled this question by genetically marking specific cells within the basal or luminal compartment using different promoters (Figure 2): BCs were targeted using the promoters of K5 (Rios et al., 2014; van Keymeulen et al., 2011), K14 (Rios et al., 2014; Tao et al., 2014; van Keymeulen et al., 2011; Wuidart et al., 2016), Lgr5 (de Visser et al., 2012; Fu et al., 2017; Rios et al., 2014; van Keymeulen et al., 2011; Wuidart et al., 2016), Axin2 (van Amerongen et al., 2012), α-SMA (*Acta2*) (Prater et al., 2014), P63 (*Trp63*) (Sreekumar et al., 2017), Protein C Receptor (*Procr*) (D. Wang et al., 2015), and Lgr6 (Blaas et al., 2016; Wuidart et al., 2016); whereas LCs were investigated using K8 (Tao et al., 2014; van Keymeulen et al., 2011; Wuidart et al., 2016), K18 (van Keymeulen et al., 2011), K19 (Wuidart et al., 2016), Notch1 (Rodilla et al., 2015), Notch2 (Šale et al., 2013), Notch3 (Lafkas et al., 2013), Elf5 (Rios et al., 2014), ERα (*Esr1*) (van Keymeulen et al., 2017), Sox9 (C. Wang et al., 2017; Wuidart et al., 2016), Prominin 1 (*Prom1*) (C. Wang et al., 2017), and WAP (Chang et al., 2014).

Two of the above mentioned lineage tracing studies added a further layer of complexity, as they found that the expression of specific promoters can dynamically change depending on the developmental stage (de Visser et al., 2012; van Amerongen et al., 2012). De Visser et al. showed that immediately after birth *Lgr5* marks luminal lineage-committed cells, whereas 12 days after birth, Lgr5^{+} MECs and their progeny are committed to the basal compartment (de Visser et al., 2012).

In addition, Van Amerongen et al., demonstrated that Axin2$^+$ cells can contribute to either the luminal or basal compartment depending on the developmental stage at which tracing is initiated (van Amerongen et al., 2012). In brief, during embryonic development, Axin2$^+$-expressing cells are restricted to the luminal lineage. However, when tracing is induced at pre-puberty (post-natal days 14 and 16) and adult stage, Axin2$^+$-expressing cells contribute exclusively to the basal lineage, suggesting that a switch in Wnt/β-catenin signalling activity takes place around birth (van Amerongen et al., 2012). Altogether, although Lrg5$^+$ and Axin2$^+$ cells switch from the luminal to the basal compartment around birth, no multipotent MaSCs were identified.

Another study, performed by Lloyd-Lewis et al., employed a **neutral genetic labelling strategy** using the Rosa26 (R26)-Cre^{ERT2}/Confetti mice, in order to label single cells in the mammary gland independently of specific promoters, as the R26 promoter is ubiquitously expressed (Figure 2). This unbiased analysis further supported a model of lineage restriction and exclusive unipotency of MaSCs in the post-natal mammary gland (Lloyd-Lewis et al., 2018).

However, a **contradictory finding** was obtained in two studies using K14-Cre^{ER} and K5-rtTA transgenic and ProCR-Cre^{ER} knock-in mice (Rios et al., 2014; D. Wang et al., 2015). Rios et al. suggested the **existence of bipotent MaSCs in the adult mammary gland** giving rise to both BCs and LCs by using a stochastic multicolour Confetti model combined with a high-resolution 3D imaging strategy (Rios et al., 2014).

One default to the studies discussed is that they have performed **population-based mosaic lineage-tracing**, in which 5%-50% of the cells of a given lineage are labelled, depending on the dose of TAM or the efficiency of recombination (Wuidart et al., 2016). Moreover, general assumptions on the exclusive lineage specificity of a given Cre, or lack of thorough characterisation of short time points to confirm the initially-labelled cells, can explain these conflicting results. In order to overcome these limitations, an important study by Wuidart et al. performed **inducible lineage tracing at saturation**, to label all cells of a given population (Figure 2). A DOX-inducible (Tet-On) mouse model was used to achieve the highest level of recombination possible, avoiding the caveat of TAM toxicity when inducing at high doses. Using such lineage tracing at saturation, it is possible to mark also a rare population of multipotent MaSCs, which could escape the labelling when performing mosaic lineage tracing experiments, allowing to study the differentiation potential of a specific stem cell population and the potential flux of cells that might occur between different lineages during post-natal development.

When labelling all BCs of the MG (DOX was administered to K14-rtTA/TetO-Cre/Rosa-YFP mice during puberty in order to achieve complete recombination of BCs by the end of pubertal development), no LCs were labelled, confirming that indeed BCs are a self-sustained unipotent lineage during puberty (Wuidart et al., 2016), as previously suggested by chimeric lineage tracing (van Keymeulen et al., 2011). Reciprocally, assessment of the fate of all LCs cells with K8-rtTA/TetO-Cre mice revealed that LCs are not replaced by unlabelled multipotent BCs over time, demonstrating that the luminal lineage is also maintained by self-sustained unipotent progenitors with long-term renewal potential (Wuidart et al., 2016). Collectively, this study strongly suggests that rare multipotent stem cells do not exist in the post-natal mouse mammary gland.

Two additional studies performed **stochastic lineage tracing approach** avoiding the use of promoters for specific cell populations (Figure 2). Watson and colleagues used the R26$^{[CA]30;EYFP}$ "slippage" mouse model, in which genetic labelling is unbiased, and extremely rare, allowing to backtrack the clonal origin of single labelled cells with confidence. Such an approach demonstrated that clones arising from a single labelled cell can be distributed throughout the length of the ductal epithelium or alveoli, but they gave rise exclusively to either LCs or BCs, and never to both lineages (Davis et al., 2016). Similar conclusions on the absence of multipotent BCs were obtained in another unbiased study that used the R26-Cre^{ERT2}/Confetti model for stochastic labelling of any cell (Scheele et al., 2017). Here, the authors conclude that unipotent BCs or LCs are localized within TEBs, with cells residing at the border (but not in the front of the tip) depositing their progeny in the subtending duct as the TEB advances to drive the growth of mammary ducts (Scheele et al., 2017). These two latter studies have provided additional robust evidence that unipotent progenitors drive both ductal morphogenesis at puberty (Davis et al., 2016; Scheele et al., 2017) and alveolar formation during gestation (Davis et al., 2016).

The aforementioned studies also explored the heterogeneity within the luminal compartment. Indeed, while the *Notch3* and *Notch2* receptors label both ERα/PR-positive and ERα/PR-negative LCs (Lafkas et al., 2013; Šale et al., 2013), *Notch1* (Rodilla et al., 2015) and *Sox9* (C. Wang et al., 2017) promoters are very specific for ERα/PR-negative LP, and the *Esr1* (van Keymeulen et al., 2017) and *Prom1* (C. Wang et al., 2017) promoters mark exclusively ERα/PR-positive LCs.

Lineage tracing using Notch1 (N1)-Cre^{ERT2} (Rodilla et al., 2015) and Sox9-Cre^{ERT2} (C. Wang et al., 2017) mice showed that these cells are self-sustained during post-natal development, as they contribute specifically to the development and maintenance solely of ERα/PR-negative LCs. Consistent with this notion, when tracing ERα/PR-positive cells at saturation using ER-rtTA or Prom1-Cre^{ERT2} mice, Van Keymeulen et al. and Wang et al. showed that all ERα/PR-positive are exclusively maintained by ERα/PR-positive lineage-restricted progenitors and are not replaced over time by other cell types, neither ERα/PR-negative LPs nor BCs (van Keymeulen et al., 2017; C. Wang et al., 2017).

Collectively, all of the above-mentioned studies provide unequivocal evidence that during post-natal development and under homeostatic conditions, BCs and LCs are self-maintained by unipotent progenitors and that, within the luminal population, ERα/PR-positive and ERα/PR-negative cell populations are also generated by separate pools of luminal progenitors rather than by a common LC.

Therefore, lineage tracing methods have now overwhelmingly demonstrated the existence of three independent cell lineages with self-renewal capacity in the post-natal mammary gland under homeostatic conditions (Figure 3).

In conclusion, MaSCs should not be defined by their cell surface marker expression nor by their ability to repopulate a cleared fat pad, but by their physiological differentiation potential *in vivo* and *in situ* in the mammary gland during normal tissue homeostasis (Watson & Khaled, 2020).

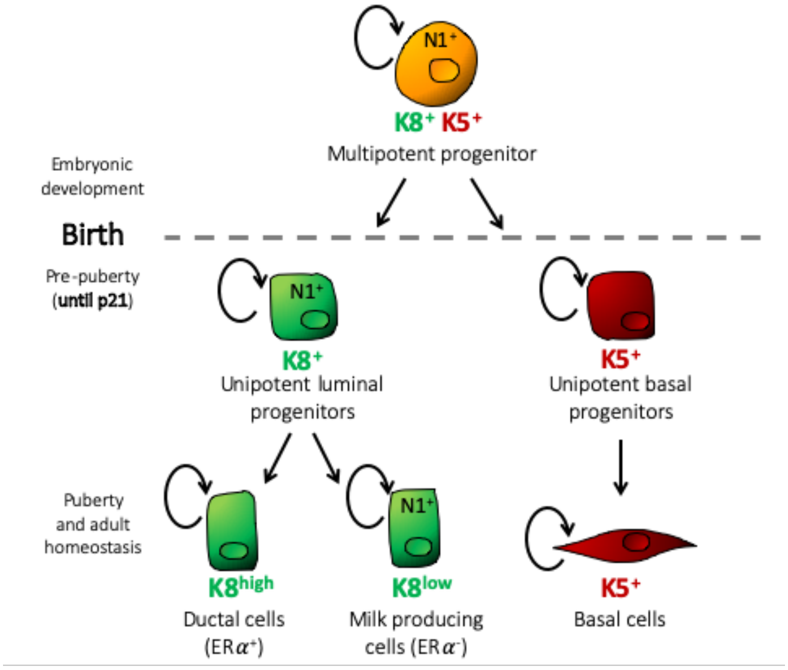


Figure 3 | The mammary epithelial cell hierarchy based on lineage tracing studies.
Multipotent MaSCs are found exclusively during embryonic development. In the post-natal MG, distinct unipotent progenitors are responsible to generate each cell type in order to maintain tissue growth and homeostasis.

3.1.4. Single cell technologies to distinguish cell states during embryonic MG development

Much progress has been recently made to disentangle the heterogeneity of mammary cells in the adult MG. However, previous population-based studies lacked the resolution required to address the important question of whether embryonic MaSCs constitute a transcriptionally homogeneous population with multipotent potential, or if they are composed of distinct stem cell subpopulations that are already committed towards the luminal or basal lineage (Rodilla & Fre, 2018).

Some studies suggested that embryonic MaSCs express a hybrid signature comprising markers of both luminal and basal cell types (Spike et al., 2012; Wansbury et al., 2011). This makes it difficult to find specific markers for embryonic cell types and for this reason **the timing and mechanisms responsible for the switch from multipotency to unipotency during embryonic MG development are still unknown**.

Single cell resolution data were needed to provide a comprehensive understanding of MEC heterogeneity during embryonic lineage specification. Thanks to recent advances in high-resolution single cell technologies, single cell RNA-sequencing (scRNA-seq) and single cell assay for transposase-accessible chromatin sequencing (scATAC-seq), we are now able to reveal the transcriptomes or chromatin accessibility profiles of single cells in order to investigate cellular heterogeneity and predict cell states in an unbiased fashion (Twigger & Khaled, 2021).

Two recent studies performed scRNA-seq analysis of embryonic MECs at distinct stages of development and proposed a model whereby the earliest stages of mammogenesis are driven by multipotent MaSCs, presenting a hybrid gene expression signature composed of both basal and luminal markers. Wuidart et al. identified a rather homogeneous population of embryonic MECs at E14 that expresses a hybrid transcriptional signature when compared to adult unipotent basal and luminal cells. They called this cell cluster MEPs, for Multipotent Embryonic Progenitors, and suggested that MEPs are not yet committed to a defined lineage (Wuidart et al., 2018). However, one limitation of this study is the small number of cells analysed (69 cells at E14, 51 adult BCs and 73 adult LCs), making the comparisons between different developmental time points limited.

The same year, an independent scRNA-seq analysis of embryonic mammary glands performed at E16 and E18 also showed that MECs co-express luminal and basal genes, and pseudotime analysis suggested that E18 cells give rise to distinct basal and luminal populations shortly after birth (Giraddi et al., 2018). This study was consistent with the results obtained by Bach et al., who investigated the mammary epithelial hierarchy by sequencing single cells from nulliparous and pregnant females and showed no transition states between the luminal and basal clusters (Bach et al., 2017). Thus, these data support the lack of a common progenitor for LCs and BCs.

In contrast, another scRNA-seq study proposed an alternative model whereby only MECs with basal characteristics are found in the early post-natal gland (14 days after birth), which generate the luminal lineages near the onset of puberty (Pal et al., 2017). They suggested that rare basal-like cells, that express CD45 before puberty, expand during puberty and reside within the luminal cell compartment (Pal et al., 2017). However, as discussed above, this model is inconsistent with mammary embryonic lineage tracing studies demonstrating unipotency after birth (Lilja et al., 2018; Lloyd-Lewis et al., 2018; Wuidart et al., 2018).

The switch from multipotency to unipotency was investigated in our lab using a combination of Notch1-dependent lineage tracing at each embryonic stage using the multicolour Confetti reporter mouse and mathematical modelling. In this work, **a progressive restriction in lineage potential was observed starting as early as E12.5, with multipotency becoming statistically undetectable after E15.5** (Figure 4). Strikingly, lineage potential restriction around E15.5 coincides with remarkable epithelial remodelling associated with the initiation of embryonic branching morphogenesis, thus providing the logical basis for our hypothesis that these two processes might be linked. It is important to stress that even unipotent embryonic progenitors co-expressed all known basal and luminal markers, making it impossible to distinguish multipotent and unipotent cells co-existing within the embryonic mammary bud (Lilja et al., 2018).

Bulk ATAC-seq analysis revealed that at E18.5 MaSCs presented open features at distal enhancer and proximal promoter regions for both luminal and basal genes (Dravis et al., 2018). Interestingly, subsequent single nucleus ATAC-seq (snATAC-seq) analysis showed that E18.5 MECs exhibit either a basal-like (*Krt5, Acta2*) or luminal progenitor-like (*Krt8, Krt18, Kit*) chromatin accessibility profile, potentially priming these cells to a lineage-restricted state immediately after birth (Chung et al., 2019). Of note, while at the transcription level E18.5 cells were indistinguishable (Giraddi et al., 2018), by snATAC-seq the authors obtained a better resolution of cell states that allowed them to conclude that at E18.5 the MECs can be divided in three separated subpopulations: basal progenitors, luminal progenitors and mature luminal cells (Chung et al., 2019).

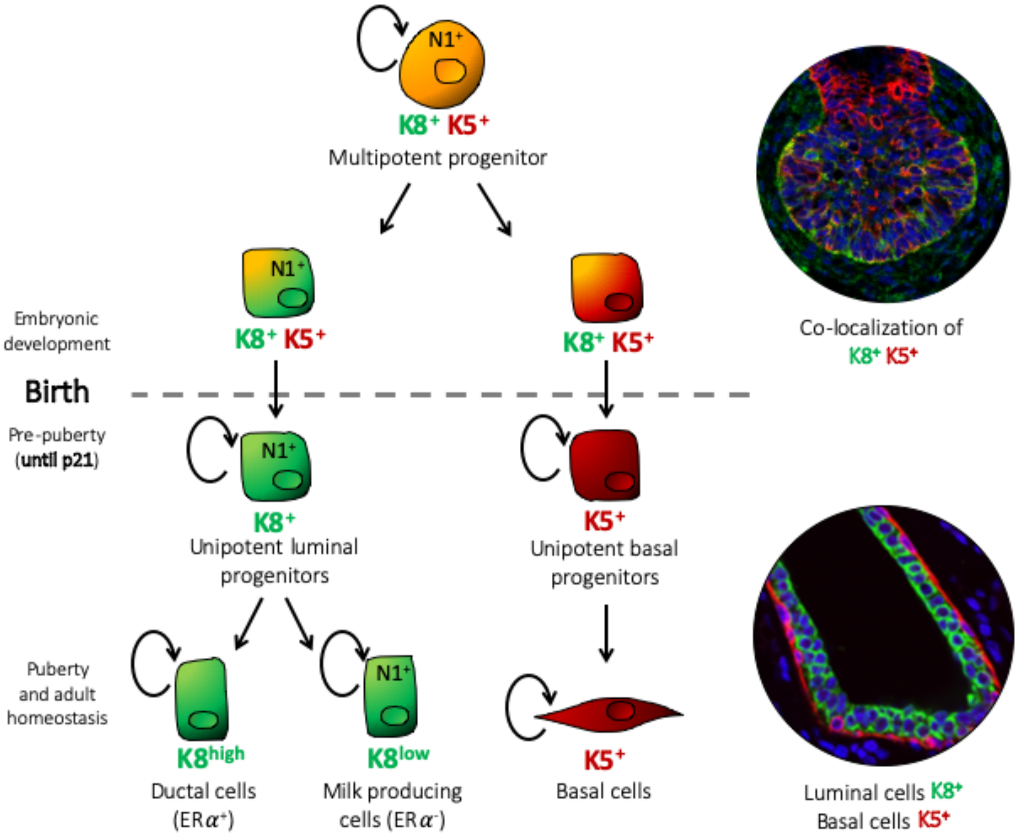


Figure 4 | Model of mammary epithelial cell hierarchy.
Multipotent MaSCs are present in the early mammary epithelial bud. Lineage restriction starts to occur as early as E12.5 and multipotency becomes statistically undetectable after E15.5. Therefore, pre-natal MG growth and branching is supported by unipotent luminal and basal progenitors. However, unipotent embryonic progenitors in the embryonic MG still co-expressed the commonly used basal (K5, in red) and luminal markers (K8, in green), making it impossible to distinguish multipotent and unipotent cells co-existing within the same mammary bud.

3.2. CELLULAR PLASTICITY OF MAMMARY CELLS

As extensively elaborated above, it is now well-established that adult mammary homeostasis is exclusively maintained by lineage-restricted unipotent progenitors, whereas multipotent MaSCs are only found in the embryonic gland. However, **lineage-committed adult mammary cells have been shown to retain a high degree of plasticity**, as they are able to revert to a multipotent stem cell state under different stress conditions, such as *de novo* tissue morphogenesis in transplantation experiments, tissue regeneration upon wounding or in tumorigenesis (Blanpain & Fuchs, 2014).

An obligatory molecular switch for the binary cell fate choice between basal and luminal differentiation is controlled by the Notch pathway. Indeed, our lab has discovered the essential role for Notch1 signalling in defining luminal cell fate, specifically towards the ERα/PR-negative luminal progenitor lineage, by examining the consequences of forced Notch1 signal activation (using Notch1 gain-of-function transgenic mice) (Murtaugh et al., 2003) in either embryonic or ectopically in basally committed adult mammary cells (Lilja et al., 2018).

Consistent with the reported role of Notch signalling in suppressing the expression of the transcription factor *Trp63* (Yalcin-Ozuysal et al., 2010), *Trp63* overexpression in adult K8-expressing LCs was shown to be sufficient to reprogram these cells into BCs (Wuidart et al., 2018). In conclusion, these two complementary studies genetically demonstrated that the Notch1-P63 axis controls the binary cell fate decision between basal and luminal differentiation. However, the mechanisms underlying lineage segregation in the embryonic mammary bud remain unknown.

Several important questions remain unanswered. For example, are multipotent MaSCs and unipotent progenitors localized in specific niches within the embryonic mammary bud? and how do they communicate with their microenvironment? What is the influence of environmental cues on cell plasticity? And how do cell plasticity and reactivation of multipotency programs underlie cancer?

4. MAMMARY GLAND MORPHOGENESIS

The hypothesis that guided my PhD work involved a connection or coordination between cell fate acquisition and morphogenetic events driving MG development (Figure 5). Thus, I will introduce here the distinct stages of mouse MG morphogenesis. The first one begins *in utero* during embryonic development and creates the rudimentary mammary ductal tree that is present at birth. The second morphogenetic event is triggered by pubertal production of hormones. This step is characterized by the formation of TEBs, club-shaped structures at the tips of the elongating ducts. The third morphogenetic step occurs at pregnancy, with the formation of tertiary branches that terminate in alveolar structures, in response to rising levels of progesterone and prolactin, accompanied by to the formation of secretory alveolar structures to produce milk. Finally, after lactation, a process called involution ensure the elimination of milk-producing cells and the return to a "virgin-like" state (Watson & Khaled, 2008).

4.1. EMBRYONIC BRANCHING MORPHOGENESIS

Mouse mammary gland development begins at embryonic day E10.5 with the formation of bilateral milk lines, followed by the appearance of five pairs of epithelial placodes (three thoracic and two inguinal). Each pair of placodes develop symmetrically at each side of the embryo, although asynchronously, with the appearance of pair number 3 first, then number 4, followed by the simultaneous formation of 1 and 5 and finally of placode 2. By E13, these placodes invaginate into the underlying mesenchyme to give rise to mammary buds. The mammary bud is surrounded by three to five layers of condensed mesenchyme, termed the primary mammary mesenchyme (Durnberger & Kratochwii, 1980). A condensed mesenchymal tissue becomes visible at E14.5, appearing separately, posterior to the mammary rudiment, referred to as fat pad precursor. Around E16.5 the first sprouting event takes place, when the epithelium, surrounded by the primary mammary mesenchyme, starts invading the fat pad precursor. At this stage, the fat pad precursor is composed of densely packed cells that start showing the first signs of lipid accumulation, known as pre-adipocytes (Sakakura et al., 1982). Paracrine interactions between MECs and pre-adipocytes in the fat pad precursor lead to the initiation of branching morphogenesis that gives rise to a small rudimentary ductal tree just prior to birth (Cowin & Wysolmerski, 2010; Spina & Cowin, 2021; Watson & Khaled, 2020) (Figure 5). Of note, this first round of branching morphogenesis occurs independently of hormonal inputs (Brisken & O'Malley, 2010). The mammary tree present at birth is composed of around 10-15 branches, which grow slowly until puberty, when the second round of branching morphogenesis and ductal elongation takes place. During late embryogenesis, two other important morphological processes must occur: the formation of a ductal lumen and the generation of the nipple sheath (Spina & Cowin, 2021).

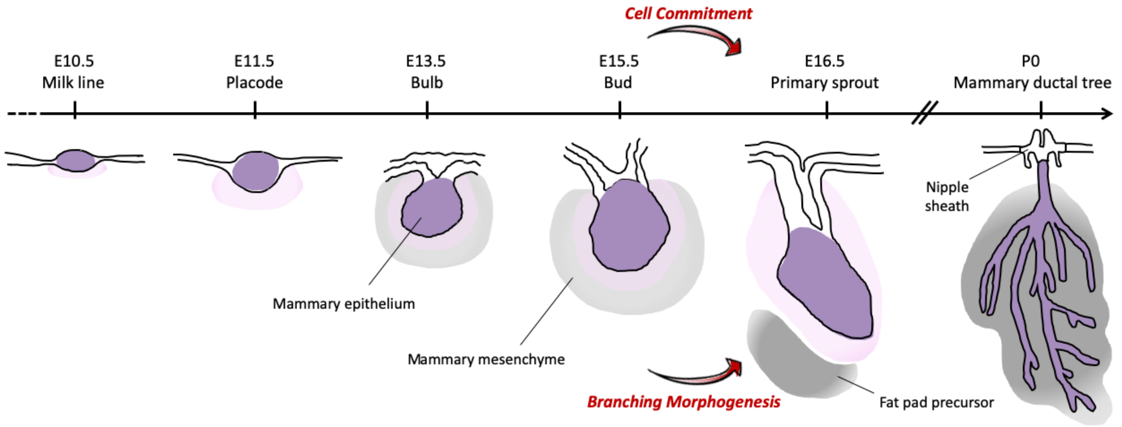


Figure 5 | Schematic representation of the stages of murine embryonic MG development.

Embryonic MG development starts with the formation of the milk line at 10.5 that subsequently resolved into an epithelial thickening known as placode by E11.5. At E13.5, placode invaginate to form the mammary buds. By

E16.5 the primary mammary sprout forms and extend toward the underlying fat pad precursor. A rudimentary mammary tree is present at birth. We hypothesized that there is a coordination between cell fate acquisition and morphogenetic events driving MG development (red arrows).

The regulatory processes driving mouse mammary branching morphogenesis have been identified to some extent. Studies using genetically modified mouse models and *ex vivo* explant cultures have shed some light onto the essential signalling pathways regulating the early stages of morphogenesis. However, the cellular mechanisms driving mammary branching morphogenesis are still poorly understood.

The next section will focus on the molecular signals underlying placode and bud formation, and mammary mesenchymal induction.

4.1.1. Molecular mechanisms driving embryonic mammary branching morphogenesis

4.1.1.1. Specification of the mammary line at E10.5

TBX3, WNT10B and FGF10 and its receptor FGFR2B are essential for the formation of most placodes. It has been proposed that the regulation of the dorso-ventral positioning of mammary buds is due to a mutual antagonism between ventral expression of the bone morphogenetic protein 4 (BMP4) and dorsal expression of the transcriptional repressor T-box transcription factor 3 (TBX3) (Robinson, 2007) (Figure 6). Indeed, the mammary line can be identified at E10.5 by a band of mesenchymal *Tbx3* mRNA expression (Jerome-Majewska et al., 2005). This is followed by the *Wnt10b* mRNA expression at E11.5 where the axillary and inguinal glands will develop (Chu et al., 2004; Veltmaat et al., 2004). TBX3 is essential for MG development since the absence of TBX3 fails to accumulate WNT10B, which in turn amplifies canonical Wnt signalling by upregulating *Lef1*, leading to a failure of mammary placode development (Jerome-Majewska et al., 2005).

In addition, the fibroblast growth factor (FGF) pathway is involved in defining the mammary line. Knock-out embryos for *Fgf10* and its receptor *Fgfr2b* fail to develop all buds, except for bud number 4 (Mailleux et al., 2002). FGF10 is initially expressed in thoracic somites when the mammary line is forming and subsequently in the mammary mesenchyme, and FGFR2b is expressed in the ectoderm (Veltmaat et al., 2006) (Figure 6). Of note, each pair of mammary placodes receive specific cues from the mesenchyme, as evidenced by the absence of defined placodes in various mouse genetic mutants.

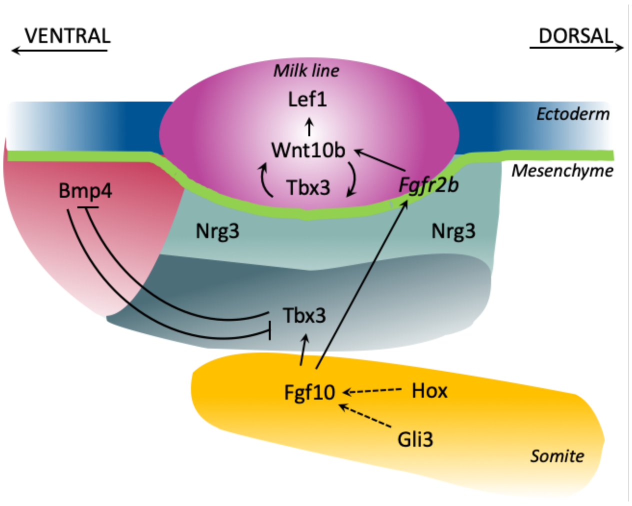


Figure 6 | Schematic representation of the mammary line specification.
The mammary line is specified by antagonism between dorsal TBX3 and ventral BMP4. FGF10 secreted from cells at the tip of the somite activates their receptor, FGFR2B. *Tbx3*, *Wnt10b*, and *Lef1* become upregulated along the mammary line (Adapted from (Spina & Cowin, 2021)).

4.1.1.2. Development of mammary placodes at E11.5

Wnt signalling participates in multiple developmental processes (Logan and Nusse, 2004) and is essential for embryonic MG development. This critical role for Wnt in placode formation and maintenance has been demonstrated by the finding that overexpression of *Dickkopf-1* (Dkk1), a secreted Wnt inhibitor, abolishes the formation of all placodes (Chu et al., 2004). Consistently, embryos deficient for *Lef1*, a downstream transcriptional mediator of canonical Wnt signalling, lack two pairs of placodes (number 2 and 3), with the remaining ones that do form eventually degenerating (van Genderen et al., 1994).

Also, Neuregulin (NRG3), a ligand of the receptor tyrosine-protein-kinase ErbB4 (a member of the EGF family), critically regulates the size and positions of buds number 3 and 4, with *Nrg3* mutants (*Nrg3^{ska}* or scaramanga) completely lacking placode number 3 (Howard et al., 2005). However, in *Nrg3* mutants, there is no loss of FGF10, indicating that NRG3 acts downstream of this factor. By contrast, experiments with *Nrg3* overexpression, or explant cultures exposed to NRG3-soaked beads, showed that ectopic NRG3 induces the formation of extra placodes around number 3 and 4 (Panchal et al., 2007). Taken together, these studies suggest that NGR3 plays a role in the formation of the mammary placodes by signalling to its receptor ErbB4 in the epithelium and modulating cell adhesion, promoting transduction of FGF10 signals from the somites underlying the milk lines (Macias & Hinck, 2012).

In addition, the ectodysplasin (Eda) signalling pathway regulates placodal cell fate along the entire mammary line. Eda is a TNF ligand and signals via its receptor, Edar, that delimits the radius of the placodes and defines the space between them. Indeed, overexpression of *Eda* results in the formation of extra placodes, particularly between placodes number 3 and 4 (Lindfors et al., 2013). In conclusion, BMP and Eda pathways define placodal size by stimulating expression of Wnt inhibitors, such as DKK1 (Cowin & Wysolmerski, 2010).

The earliest specified mesenchymal marker is latent TGFβ-binding protein1 (LTBP1). It is expressed in a halo surrounding each mammary rudiment to guide the movement of ectodermal cells into placodes numbers 1, 2 and 5 (Chandramouli et al., 2013).

4.1.1.3. Formation of mature mammary bud at E13.5

Mammary bud development displays differential requirements for specific signalling pathways, including members of the Insulin-like growth factor 1 (IGF1) and Wnt signalling pathways, as well as the combination of two homeodomain transcription factors, *Msx1* and *Msx2* (Cowin & Wysolmerski, 2010).

The IGF1/p190-B pathway is required for the formation of mature buds, as disruption of the IGF receptor (IGF1R) resulted in small epithelial buds (Heckman et al., 2007). Wnt signalling also plays a role in the transition from placodes to buds. *Pygopus 2* deletion (Pygo2, a Wnt modifier), as well as *Lrp5* or *Lrp6* (two Wnt coreceptors) genes, lead to smaller buds (Gu et al., 2009; Lindvall et al., 2006, 2009). In addition, as mentioned above, embryos deficient for *Lef1* only form placodes number 1, 4 and 5, which never transition to buds as epithelial cells fail to invaginate into the mammary mesenchyme (van Genderen et al., 1994).

Also, the embryonic MG is devoid of Hedgehog (HH) signalling activity (Hatsell & Cowin, 2006), whereas this is a major driver of cellular proliferation and organ growth in hair follicle placodes (Gritli-Linde et al., 2007). The HH pathway has been implicated in negatively regulating mammary bud invagination. In mice mutant for GLI family Zinc Finger 3 (*Gli3*, a repressor of the HH pathway), mammary buds often evaginate rather than invaginate (Lee et al., 2011). Thus, *Gli3* repression is required for acquiring mammary fate, by blocking hair follicle differentiation.

4.1.1.4. Mammary mesenchyme differentiation at E15.5

As previously highlighted, the interactions between mammary mesenchyme and epithelium are crucial for MG development and function. Parathyroid hormone-related protein (PTHrP) is considered to be the master regulator of mammary mesenchymal differentiation. PTHrP is secreted by MECs and activates the parathyroid hormone 1 receptor (PTHR1), expressed by the surrounding mesenchyme, to modulate Wnt and BMP signalling (Wysolmerski et al., 1998). PTHrP upregulates Bmp receptor-1A (*Bmpr1A*) expression in the mammary mesenchyme, and BMP4 signals through BMPR1A to induce expression of *Msx2* that ultimately inhibits hair follicle formation at the developing nipple sheath (Macias & Hinck, 2012).

Deletion of either gene results in the interruption of mammary development, demonstrating the importance of communication between the mammary mesenchyme and epithelium in the formation of the mammary bud and transition to a ductal system. In the absence of PTHrP, the undifferentiated mammary mesenchyme fails to send differentiative signals to the epithelium and, as a result, female buds lose their mammary identity and increase expression of epidermal specific markers (Wysolmerski et al., 1998).

4.1.1.5. Initiation of branching morphogenesis at E16.5

Once the mammary bud has been specified, the subsequent action of the Wnt, Eda, FGF and TGF-β signalling pathways is important to promote branching morphogenesis and lumen formation (Spina & Cowin, 2021).

Lumen formation in the embryonic mammary gland is thought to involve apoptosis, autophagy and changes in cell remodelling. First, intercellular spaces develop within the elongating ducts. Then, these spaces increase in size and number until a distinct lumen is visible at E18.5 (Macias & Hinck, 2012). However, the mechanisms underpinning lumen formation are still unknown. Of note, the early mesenchymal marker Ltbp1 is also expressed in cells facing a microlumen in the embryonic mammary tree at E17.5, and not in the multi-layered ductal tips (Chandramouli et al., 2013). This expression pattern suggests that the TGF-β pathway is involved in generating polarity and cavitation (Spina & Cowin, 2021).

4.1.2. Cellular mechanisms driving embryonic mammary branching morphogenesis

As detailed above, some signalling pathways regulating embryonic mammary branching morphogenesis have been studied, however the required cellular events remain elusive. A variety of cellular processes have been proposed to contribute to branching morphogenesis in different organs, including differential growth, collective cell migration, epithelial folding and matrix-driven tissue shape changes (Varner & Nelson, 2014).

During embryogenesis, the ectoderm gives rise to ectodermal appendages including the hair follicle, teeth, feathers, sweat glands and mammary glands. Multiple features appear to be conserved between these appendages during embryonic development (Biggs & Mikkola, 2014). Recent advances in live imaging techniques allows now researchers to investigate the dynamic changes occurring in 3D during embryonic tissue development.

Cellular mechanisms driving branching morphogenesis have been most extensively studied in the salivary gland thanks to *ex vivo* embryonic salivary gland cultures in combination with advanced time-lapse microscopy. It has been recently shown that **budding morphogenesis in the salivary gland is driven by the ECM**, specifically through a combination of strong cell-matrix adhesions and weak cell-cell adhesions (S. Wang et al., 2021).

In embryonic **molar tooth** development, it has been proposed that the transition from the placode to the bud stage is achieved by a combination of cell proliferation (Ahtiainen et al., 2016) and **horizontal contractile "canopy" and intercalation of superficially located cells that bring together both sides of the invagination**, thereby thickening the epithelium (Panousopoulou & Green, 2016). A recent study has introduced a new principle of epithelial invagination happening in teeth after the previous step of contractile canopy, designated as "**vertical telescoping**". They showed that initial invagination occurs through coordinated vertical cell-on-cell migration, with cells in the periphery of the placode migrating upwards with apical protrusions and central neighbours moving downwards (J. Li et al., 2020). This vertical telescoping, but not the canopy contraction, also participates in the invagination of the salivary gland (J. Li et al., 2020) and hair follicle (Morita et al., 2021). Interestingly, the mammary placodes have a similar shape as tooth placodes (Figure 7), raising the question of a conservation of this mechanism in the MG.

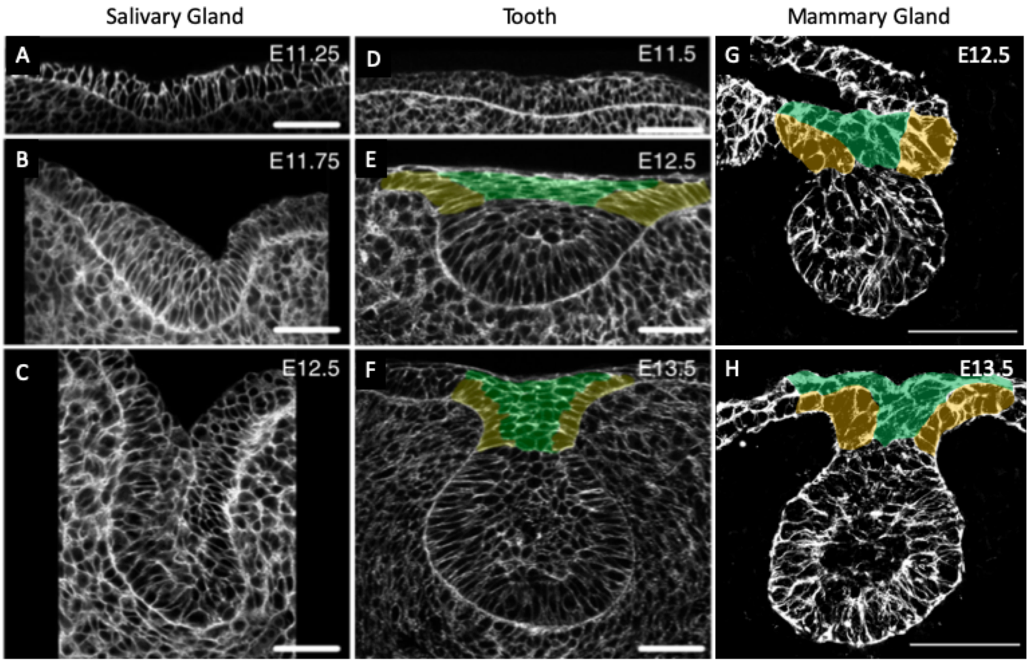


Figure 7 | Invagination of the salivary placode (A-C), tooth placode (D-F) and mammary placode (G-H). False-colouring indicates intercalating suprabasal cells (in green) and basal cells, that transmit the contractile force to the basal lamina (in orange). The mammary placode presents a structure similar to what we observe in the tooth placode, as highlighted with the same colours. Phalloidin staining (for salivary and tooth placodes) or K5 staining (for mammary placode) delineates cell membrane. Scale bars in all panels: 50 µm. Figures A-F were adapted from (J. Li et al., 2020).

Two studies indicated low levels of cell proliferation in mammary primordia (Balinsky, 1950; Lee et al., 2011). Furthermore, a recent study by Trela et al. implied that cell proliferation plays a minor role in placode formation (Trela et al., 2021). This study revealed that **the growth of mammary primordia up to placode formation is mainly driven by cell migration of adjacent epidermal cells.** This morphogenetic event is triggered by a ring-like structure of contractile keratinocytes encircling the invaginating mammary bud (Trela et al., 2021). Interestingly, both hair follicles (Ahtiainen et al., 2014) and tooth placodes (Ahtiainen et al., 2016) also form by cell influx, suggesting that cell migration could be a common driver of ectodermal appendages morphogenesis.

Notwithstanding these pioneering studies, several interesting questions remain to be tackled to better understand the initial development of a mammary gland, like: how are molecular signals translated into coordinated changes in cellular behaviour to ultimately drive tissue morphogenesis? How can branching morphogenesis instruct cell fate? What comes first: epithelial cell rearrangements or cell differentiation?

4.2. POST-NATAL BRANCHING MORPHOGENESIS

MG development is unique because, although it starts during embryogenesis, the tissue will reach its final shape only after pubertal growth, unlike most other organs that acquire their definitive morphology and function during embryogenesis.

After birth, mammary growth is conspicuously slowed down until puberty. During the pubertal stage (from 4 to 8 weeks of age), TEBs display high proliferation rates, resulting in extensive duct elongation and side branching, until the epithelium fills the entire fat pad by 9 weeks of age (Figure 1). TEBs are composed of an outer layer of cap cells –covered by a continuous BM– that generate BCs and an inner multi-layer of body cells giving rise to LCs (Inman et al., 2015) (Figure 8). Around 9 weeks of age the fat pad is filled, growth ceases and TEBs disappear. Lateral side branching in the mammary epithelium occurs during each oestrous cycle in preparation for prospective pregnancies. During pregnancy, the mammary epithelium undergoes extensive proliferation and promotes the differentiation of alveolar structures, dedicated to milk production and secretion during lactation (Watson & Khaled, 2008).

4.2.1. Hormonal signalling initiate post-natal MG development

The molecular programs that initiate post-natal MG expansion are primarily downstream of hormone signalling. These circulating hormones are released from the ovary, including ER and PR, and from the pituitary gland, including growth hormone (GH) and prolactin (PL) (Figure 8).

GH activates the release of IGF1 from stromal cells, which then signals to epithelial cells via IGF1 receptor (IGF1R) to promote branching. The expression of amphiregulin (AREG), a member of the EGF family, is induced in MECs by ER at puberty. AREG signals to its receptor, EGFR, on stromal fibroblasts. Activation of EGFR induces the expression of FGFs that, in turn, stimulate cell proliferation through FGFR2 expressed by MECs. Thus, FGF10 signalling through FGFR2 is also important for mammary gland development, just like in the salivary gland and lung (Goodwin & Nelson, 2020; Macias & Hinck, 2012) (Figure 8).

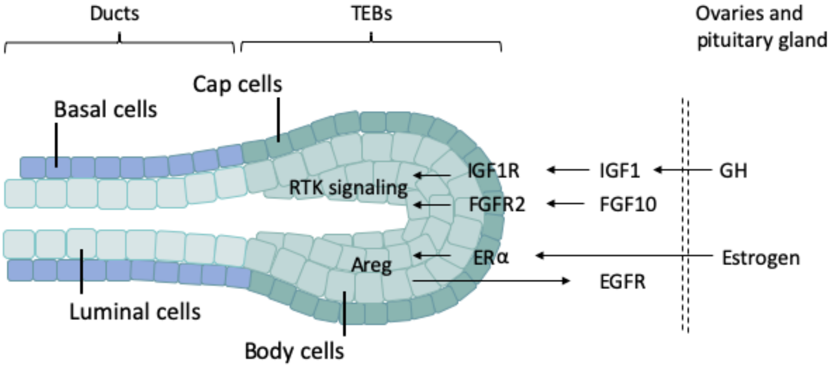


Figure 8 | Hormone signalling is essential for branching morphogenesis in the pubertal MG.
TEBs consist of cap and body cells, giving rise to basal and luminal cells, respectively. Stromal cells signals to epithelial cells via IGF1R and FGFR2 to promote growth and branching. Receptor tyrosine kinase (RTK) activity is essential for TEB growth and duct elongation. Signalling from epithelial cells of the MG to stromal cells via EGFR is also important for MG development during puberty (Adapted from (Goodwin & Nelson, 2020)).

ERα is required for both prepubertal development and at late pregnancy (Watson & Khaled, 2008). Terminal end buds (TEBs) are absent in ERα knock-out mammary glands (Mallepell et al., 2006). In addition, loss of ductal side-branching and lobuloalveolar structures have been shown upon the specific deletion of ERα in the mammary epithelium at late pregnancy and lactation, through the WAP promoter (Feng et al., 2007).

4.2.2. Physical signals are required to pattern ductal elongation

During puberty, the branched ductal network is created via 3 processes: (1) TEBs splitting, known as tip bifurcation or clefting, (2) TEBs extension and (3) side-branching of an existing duct. However, the physical signals that generate a branched network are still largely unknown.

One approach to tackle this question is represented by predictions developed via mathematical modelling (Myllymäki & Mikkola, 2019). For example, Hannezo and colleagues have proposed a theoretical model, termed branching and annihilating random walks (BARWs), capable of predicting the spatiotemporal dynamics of branching morphogenesis in the mouse mammary gland, as well as in the mouse kidney and human prostate. They proposed that growing ductal tips randomly explore the environment and compete for space, becoming inactive when they are in proximity with another duct. They concluded that branch elongation and orientation occur in a stochastic self-organizing manner, rather than through a pre-defined sequence of genetic events (Hannezo et al., 2017). Nonetheless, this study did not consider the dynamics of epithelial orientation observed at each developmental stage *in vivo* (Nerger et al., 2021). BARWs simulations were thus repeated using 2 new parameters: (a) from 4 to 10-week-old MGs the epithelial orientation is biased along the long axis of the fat pad, but in 10-week-old MGs there is a heterogeneous epithelial orientation due to the formation of lateral branches; and (b) the average bifurcation angle *in vivo* is of ~75°. The new modelisation finally proposed that local accumulation of collagen-rich ECM, which constrains the **angle of bifurcation of TEBs**, is sufficient to generate the global pattern of orientation observed in the developing mammary epithelium (Nerger et al., 2021).

In conclusion, although mammary organoids models have shed some light into the mechanisms of ductal elongation intrinsic to the epithelium, it is now known that the native stromal microenvironment plays an important role *in vivo*, not only as a source of growth factors but also through its mechanical properties (Myllymäki & Mikkola, 2019).

4.2.3. Stromal signals mediating tip branching and termination

The lack of stereotypical morphology of the mammary tree suggests that **TEB extension** is regulated by non-patterned secreted signals within the surrounding microenvironment. Local expression patterns of matrix metalloproteinases (MMPs), TGF-β, immune cells, collagen type I, sulphated glycosaminoglycans, hyaluronate, tenascin C, among others, have been reported to participate as guidance cues to TEB extension (Nerger et al., 2021).

Epithelial branching rate can be significantly stimulated by the addition of beads soaked with FGF10, a positive regulator of branching morphogenesis (Zhang et al., 2014), embedded into the mammary fat pad of 4-week-old pre-pubertal mice (Hannezo et al., 2017). On the other hand, TGF-β, an inhibitor of mammary ductal growth (Silberstein & Daniel, 1987), regulates termination of active TEBs in a local manner (Hannezo et al., 2017), suggesting that TGF-β may regulate ECM accumulation around bifurcating TEBs.

Additionally, it is now established that epithelial Wnt signalling is a regulator of TEB number and size (Myllymäki & Mikkola, 2019). Canonical Wnt ligands promote survival of cap cells in TEBs, by preventing the nuclear accumulation of *FoxO* transcription factors (Sreekumar et al., 2017). Macrophages envelope TEBs during ductal morphogenesis (Stewart et al., 2019) and were shown to be the source of Wnt ligands. A positive feedback loop between cap cells and macrophages was identified. Cap cells produce the Notch ligand Delta-like1 (DLL1), which activates the Notch2 and 3 receptors in stromal macrophages, leading to increased expression of Wnt ligands (Chakrabarti et al., 2018).

On the other hand, lateral branching occurs in the context of an epithelial bilayer, therefore the morphogenetic programs are likely different than in tip bifurcation, occurring within a compact cell mass forming the TEB (Myllymäki & Mikkola, 2019). Side branching occurs during the oestrous cycle and at pregnancy in response to PR signalling, whose effects are mediated by Wnt4, RANKL, a TNF family member (Joshi et al., 2015) and Id2 (Seong et al., 2018).

In conclusion, the development of powerful techniques, such as *ex vivo* cultures to dynamically study embryonic mammary gland morphogenesis, intravital imaging to examine live cell behaviour at single cell resolution, optical tissue clearing protocols for 3D imaging analysis, computational models and single-cell omics, will undoubtedly shed new lights to our understanding of branching morphogenesis, cell fate specification and their coordination in different developing organs.

Aim of the study

The embryonic mammary gland (MG) represents an ideal system for studying stem cell potential restriction and its coordination with tissue branching morphogenesis.

While the adult MG has been extensively investigated due to its accessibility and complete development happening post-natally, few studies have addressed the lineage potential of embryonic mammary cells. This is at least in part due to several technical shortcomings, including the difficulty in identifying multipotent embryonic Mammary Stem Cells (MaSCs) that lack specific markers, as well as the challenging procedures required for dissecting and establishing in culture embryonic mammary buds.

During my PhD work I have pursued three main objectives.

In *Aim 1*, I focused on understanding the developmental timing and reveal the molecular signatures dictating mammary cell differentiation. Thanks to extensive single cell transcriptomic analyses, we provide a comprehensive atlas of mammary cell differentiation during embryonic development. By sequencing stromal cells along with mammary epithelial cells (MECs), I also investigated the paracrine interactions between the embryonic mammary mesenchyme and MECs.

In *Aim 2*, I have established a time-lapse microscopy methodology and image analysis pipeline to visualize and quantify cellular dynamics and remodelling during embryonic MG growth. Dynamic cellular processes involved in the initial sprouting events leading to the rudimentary mammary tree remained unknown. This technique allowed me to study the behaviour of individual cells during embryonic MG branching morphogenesis at both cell and tissue scales.

Finally, the goal of *Aim 3* was to decipher how cell fate acquisition is coordinated with progressive tissue morphogenesis during mammary development. To tackle this question, I combined single cell sequencing and individual cell tracking analysis in embryonic *ex vivo* cultures using a β-catenin gain-of-function mutant mouse where MG development is arrested.

Overall, this PhD thesis aims to unravel how different cell types are specified in time and space in a tightly regulated manner in order to generate a branched tissue.

Materials and Methods

1. Mouse models

N1-Cre^{ERT2} (Fre et al., 2011) or K5-Cre^{ERT2} (Indra et al., 1999) mice were crossed to the double fluorescent reporter R26$^{mT/mG}$ (Muzumdar et al., 2007) and inducible mutant β-catenin (*Catnb*$^{+/lox(ex3)}$) (Harada et al., 1999) transgenic lines. All lines used were of mixed genetic background. Mice carrying a single copy of the Cre recombinase gene were used for analysis. Reporter expression and β-catenin stabilisation were induced in N1-Cre^{ERT2}/R26^{mTmG}/*Catnb*$^{+/lox(ex3)}$ and K5-Cre^{ERT2}/R26^{mTmG}/*Catnb*$^{+/lox(ex3)}$ females at day E12.5 of gestation by a single intraperitoneal injection of tamoxifen free base (MP Biomedicals, 156738) prepared in sunflower oil containing 10% ethanol (0.1 mg per g of mouse body weight) (Rodilla et al., 2015). For time-lapse experiments TAM doses were reduced 10-fold (0.01 mg per g of mouse body weight) to allow cell tracking during branching morphogenesis (Lloyd-Lewis et al., 2022). We exclusively analysed female mice. Mice were analysed at embryonic stages E13.5, E14.5 and E15.5, and during postnatal development at P0, as indicated in the figure legends. Plug detection at mid-day was considered 0.5 days-post-coitus (E0.5). Mice were genotyped by PCR analysis on genomic DNA extracted from tail tip. To distinguish the sex of the embryos, the following primer sequences and cycling conditions were used:

5’ - TGGATGGTGTGGCCAATG - 3’ , 3’ - CACCTG CACGTTGCCCTT - 5’.

94°C for 2 min; then 35 cycles of 94°C for 30 s, 58°C for 30 s, 72°C for 30 s; finally 72°C for 5 min.

2. Ethics Statement

All studies and procedures involving animals were in accordance with the recommendations of the European Community (2010/63/UE) for the Protection of Vertebrate Animals used for Experimental and other Scientific Purposes. Approval was provided by the ethics committee of the French Ministry of Research (reference APAFIS #34364-202112151422480). We comply with internationally established principles of replacement, reduction, and refinement in accordance with the Guide for the Care and Use of Laboratory Animals (NRC 2011). Husbandry, supply of animals, as well as maintenance and care in the Animal Facility of Institut Curie (facility license #C75–05–18) before and during experiments fully satisfied the animal’s needs and welfare. All mice were housed and bred in a specific-pathogen-free (SPF) barrier facility with a 12:12 hr light-dark cycle and food and water available *ad libitum*. Mice were sacrificed by cervical dislocation, whereas mice embryos were decapitated.

3. Embryonic mammary gland dissection and *ex vivo* culture

Mammary embryonic buds were dissected following the protocol developed by the laboratory of M. Mikkola (Voutilainen et al., 2013). Briefly, embryos were harvested from the uterus of a pregnant dam at day E13.5 of pregnancy. Under a dissecting microscope, an incision along the dorsal-lateral line from the hind limb to the forelimb in the right flank of the embryo was done using spring scissors. The flank of the embryo from the incision along the dorsal-lateral line to the midline was detached and the same steps were repeated for the left flank of the embryo, but this time cutting along the dorsal-lateral line from the forelimb to the hind limb. Tissues were collected in a 24-well plate with phosphate buffered saline (PBS) until all embryos were dissected.

Next, proteolytic digestion of dissected embryonic flanks was performed as previously described (Lan & Mikkola, 2020). Tissues were incubated with freshly prepared 1.25 U/ml Dispase II solution (Roche, 04942078001) at 4°C for 15 minutes. Then, with Pancreatin-Trypsin solution at room temperature (RT) for 4-5 minutes. To prepare Pancreatin-Trypsin working solution: first 0.225 g of Trypsin (Sigma-Aldrich, 85450C) were dissolved into 9 mL of Thyrode's solution [8 g/L NaCl (Sigma-Aldrich, S5886) + 0.2 g/L KCl (Sigma-Aldrich, P5405) + 0.05 g/L $NaH_2PO_4 \bullet H_2O$ (Sigma-Aldrich, S3522) + 1 g/L D-(+)-Glucose (Sigma-Aldrich, G7021) + 1 g/L $NaHCO_3$ (Sigma-Aldrich, S5761) dissolved in 1 L of distilled water and filter sterilised]. Then, 1 mL of 10X Pancreatin stock solution [0.85 g NaCl (Sigma-Aldrich, S5886) and 2.5 g Pancreatin (Sigma-Aldrich, P3292) dissolved into 100 mL of distilled water on a magnetic stirrer on ice for 4 hr and filter sterilised] and 20 µL of Penicillin-Streptomycin (10,000 U/ml in stock) (Sigma-Aldrich, P4333) were added. Finally, pH was adjusted to 7.4 with NaOH and the solution was filter sterilised (see in (Lan & Mikkola, 2020)).

When skin epithelium started to detach from the edges of the mammary mesenchyme, the Pancreatin-Trypsin solution was replaced with DMEM/F-12 (Gibco-Thermo Fisher Scientific, 21331020) embryonic culture medium to inactivate the enzyme activity. After incubating the tissue for 20-30 minutes in ice, the skin epidermis was removed away from the mesenchyme containing the embryonic mammary buds by using two needles.

Mammary embryonic buds were established in *ex vivo* culture as previously detailed in (Carabaña & Lloyd-Lewis, 2022). Collected embryonic mammary tissue was placed on a cell culture insert floating on embryonic culture medium into a 35 mm cover glass-bottomed tissue culture dish (Fluorodish, 81158). Embryonic culture medium is DMEM/F-12 (Gibco-Thermo Fisher Scientific, 21331020) supplemented with 2 mM GlutaMAX^{TM} (Gibco-Thermo Fisher Scientific, 35050-038), 10% fetal bovine serum (FBS) (v/v), 20 U/ml Penicillin-Streptomycin (Gibco-Thermo Fisher Scientific, 15140122) and 75 µg/mL Ascorbic acid (Sigma, A4544). Mammary cultures were maintained in a tissue culture incubator at 37°C and 5% CO_2 atmosphere. The culture media was replaced with fresh media every second day for the duration of the experiment. For growth factors assay, 1 nM FGF10 (Bio-techne, 6224-FG) was added in the embryonic culture medium at day 4 in *ex vivo* culture.

4. Mammary *ex vivo* culture whole-mount immunostaining

Ex vivo culture whole-mount immunostaining was performed as previously described (Carabaña & Lloyd-Lewis, 2022). Explants were transferred to a 24 well plate, washed in PBS and fixed with 4% PFA for 2 hr at RT. After a blocking step in PBS containing 5% FBS, 1% Bovine Serum Albumin (BSA) and 1% Triton x-100 (Euromedex, 2000-C) for 2 hr, *ex vivo* cultures were incubated with primary antibodies diluted in blocking buffer overnight at 4°C. Then, with secondary Alexa-fluor conjugated antibodies and DAPI (10µM) diluted in PBS for 5 hr at RT. *Ex vivo* cultures were mounted in a slide using Aqua-Polymount (Polysciences, 18606). The following primary antibodies were used: rabbit anti-SMA (1:300, Abcam, ab5694), rat anti-K8 (1:300, Developmental Studies Hybridoma Bank, clone TROMA-I), mouse anti-P63 (1:300, Abcam, ab735), rabbit anti-K5 (1:300, Covance, PRB-160P-100), rat anti-ZO-1 (1:100, Millipore, MABT11), rabbit anti-K14 (1:300, Abcam, ab181595).

EdU incorporation was visualized using Click-It chemistry (Invitrogen) by incubating *ex vivo* cultures for 2 hr with EdU solution (10 µM). EdU was then detected with freshly made Click-iT EdU Alexa Fluor 647 Imaging Kit (Invitrogen-Thermo Fisher Scientific, C10640), according to the manufacturer's protocol. Nuclei were stained with Hoechst33342 (10 µg/mL) for 30 minutes at RT.

5. Immunofluorescence on 2D sections

Embryos were harvested and fixed in 4% PFA overnight at 4°C, followed by another overnight incubation at 4°C in 30% sucrose. Then, embryos were embedded in optimum cutting temperature (OCT) compound and 7 µm-thick cryosections were cut using a cryostat (Leica CM1950). After a blocking step in PBS containing 5% FBS, 2% BSA and 0.2% Triton x-100 for 2 hr, sections were incubated with primary antibodies diluted in blocking buffer overnight at 4°C in a humidified chamber, then with secondary Alexa-fluor conjugated antibodies and DAPI (10µM) diluted in PBS for 2 hr at RT. Finally, sections were mounted in a slide using Aqua-Polymount (Polysciences, 18606). The following primary antibodies were used: rat anti-K8 (1:300, Developmental Studies Hybridoma Bank, clone TROMA-I), mouse anti-P63 (1:300, Abcam, ab735), mouse anti-ERalpha (1:20, Agilent-Dako, M7047), rabbit anti-K5 (1:300, Covance, PRB-160P-100), rabbit anti-PLAG1 (1:100) (Spengler et al., 1997).

6. Single molecule RNA fluorescence in situ hybridization (smRNA-FISH)

smRNA-FISH was performed using RNAscope Multiplex Fluorescent Reagent Kit v2 user manual (Advanced Cell Diagnostics, document number 323100-USM). The protocol was followed according to the manufacturer's recommendations. In brief, tissue sections were pre-treated with the target retrieval reagent (ACD, 322000) for 5 minutes and digested with Protease III (ACD, 322381) at 40°C during 15 minutes, before hybridization with the target oligonucleotide probes. Probe hybridization, amplification and binding of dye-labelled probes were performed sequentially. For subsequent immunostaining, sections were incubated in blocking buffer (PBS containing 5% FBS and 2% BSA) for 1 hr. For smRNA-FISH in the *ex vivo* cultures, blocking buffer also included 0,3 % Triton x-100 (Euromedex, 2000-C) to allow permeabilization. Incubation with primary antibodies diluted in blocking buffer was performed overnight at 4°C in a humidified chamber, then secondary antibodies and DAPI diluted in PBS were added for 2 hr at RT. The experiment was performed at least on three different embryos for each probe. Slides were mounted in ProLong Diamond Anti-fade Mountant (Invitrogen-Thermo Fisher Scientific, P36930) for imaging. The following RNAscope probes were used: Mm-Anxa1-C2 (ACD, 509291), Mm-Lgals3-C2 (ACD, 461471), Mm-Plet1-C1 (ACD, 557941), Mm-Ly6d-C1 (ACD, 532071), Mm-Cxcl14-C3 (ACD, 459741), Mm-Ndnf-C2 (ACD, 447471), Mm-Pthlh-C3 (ACD, 456521), Mm-Cd74-C1 (ACD, 437501), 3-plex Positive Control Probe-Mm (ACD, PN 320881) and 3-plex Negative Control Probe (ACD, PN 320891).

7. Microscopy and image acquisition

3D imaging: Images were acquired using a LSM780 or LSM880 inverted laser scanning confocal microscope (Carl Zeiss) equipped with 25x/0,8 OIL LD LCI PL APO or 40x/1,3 OIL DICII PL APO. For standard 4-colour imaging, laser power and gain were adjusted manually to give optimal fluorescence for each fluorophore with minimal photobleaching. Images were captured using the ZEN Imaging Software and processed in Fiji (ImageJ v1.53).

smRNA-FISH: Images were acquired using a LSM880 with an Airyscan system. The Airyscan system has 32-channel GaAsP (Gallium Arsenide Phosphide) detectors, which allows to obtain images with enhanced spatial resolution and improved signal-to-noise ratio (SNR) than in traditional LSM systems (Huff, 2015). A 63x/1,4 OIL DICII PL APO objective was used. Images processed in Fiji (ImageJ v1.53).

Live-imaging: Images were acquired using an LSM780 or LSM880 inverted laser scanning confocal microscope (Carl Zeiss) equipped with 10x/0,3 DICI EC PL NEOFLUAR, for imaging at the tissue scale; or with a 40x/1,0 W DIC PL APO, for imaging at the cellular resolution. Explants were cultured in a humid imaging chamber at 37°C with 5% CO_2 during the course of imaging. To analyse branching morphogenesis in embryonic mammary buds, images were acquired at 8 mm Z intervals over approximately 80 mm thickness and 60 minutes intervals for 12-48 hr. To track single cells, images were acquired at 2 mm Z intervals over approximately 80 mm thickness and 10 minutes intervals for 12-48 hr.

8. Single cell dissociation of embryonic mammary gland

The isolated embryonic mammary rudiments include both the mammary epithelium and the surrounding mesenchyme. 60-90 mammary rudiments were dissected for each experiment from 7-12 female embryos derived from 2-4 timed pregnant females. The scRNA-seq of each developmental time was performed in a separate dissection session to maximize the number of mammary buds analysed/timepoint.

Embryonic mammary buds together with the surrounding mesenchyme were dissected as detailed above (see *Embryonic mammary gland dissection and ex vivo culture* section). Single cell dissociation of the mammary rudiments was performed as previously described with modifications (Wuidart et al., 2018).

For mammary rudiments at E13.5, E14.5 and E15.5, single cell dissociation was performed through enzymatic digestion with 300 U/ml collagenase A (Roche, 10103586001) and 300 U/ml hyaluronidase (Sigma, H3884) for 90 minutes at 37°C under shaking. Mammary rudiments from each female embryo were dissociated in a separated 2 mL protein LoBind tube (Eppendorf, 022431102). Cells were further treated with 0.1 mg/ml DNase I (Sigma, D4527) for 3 minutes. 10% FBS diluted in PBS was added to quench the DNase I. Cells were pelleted by centrifugation at 320 g for 10 minutes.

For mammary glands at birth, the enzymatic digestion for single cell dissociation was optimized as followed. 600 U/ml collagenase A (Roche, 10103586001) and 150 U/ml hyaluronidase (Sigma, H3884) for 90 minutes at 37°C under shaking were used for enzymatic digestion. Cells were further treated with 0.1 mg/ml DNase I (Sigma, D4527) for 3 minutes and an additional incubation in 0.63% NH_4Cl for 1 minute allowed lysis of red blood cells. Cells were pelleted by centrifugation at 320 g for 10 minutes.

For all developmental times, after careful removal of the supernatant, cells were incubated in fluorescently labelled primary antibodies.

9. Cell labelling, flow cytometry and sorting

Single cell suspensions were incubated for 15 minutes on ice with fluorescently labelled primary antibodies diluted in HBSS with 2% FBS. Cells were washed from unbound antibodies with 2% FBS in HBSS and the cell suspension was passed through a 40 µm cell strainer filter to eliminate cell clumps.

Cell viability was determined with DAPI and doublets were systematically excluded during analysis. $CD45^+$, $CD31^+$, $Ter119^+$ (Lin^+) non-epithelial cells were excluded. FACS analysis was performed using an ARIA flow cytometer (BD).

The following primary antibodies were used at a 1:100 dilution: APC anti-mouse CD31 (Biolegend, 102510), APC anti-mouse Ter119 (Biolegend, 116212), APC anti-mouse CD45 (Biolegend, 103112), APC/Cy7 anti-mouse CD49f (Biolegend, 313628), and PE anti-mouse EpCAM (Biolegend, 118206), used for WT samples), or PE/Cy7 anti-mouse EpCAM (Biolegend, 118216, used for N1Cre/β-cat/Tom sample). The isotype controls were the following: PE rat IgM (Biolegend, 400808), PE/Cy7 rat IgG2a (Biolegend, 400522), APC/Cy7 rat IgG2a (Biolegend, 400524) and APC rat IgG2b (Biolegend, 400612). The results were analysed using FlowJo software (V10.0.7).

10. Mammary gland whole-mount carmine staining

Carmine alum solution was prepared by dissolving carmine (2g/L) (Sigma-Aldrich, 1390-65-4) and aluminium potassium sulfate (5g/L) (Sigma-Aldrich, A-7167) in distilled water and boiled for 20 minutes. Both flanks of newborns were collected, spread on glass slide and fixed 75% ethanol and 25% acetic acid glacial for either 4 hr at RT or overnight at 4°C. Then, tissues were rehydrated (70% ethanol for 30 minutes and distilled water for 5 minutes) and stained with Carmine alum overnight at RT. Stained embryonic flanks were dehydrated in increasing concentrations of ethanol (70%, 96%, 100% for 30 minutes each) and cleared in xylene (Voutilainen et al., 2013). Images were taken with a fluorescence stereo microscope (Leica M205 FCA). The acquired images were processed in Fiji (ImageJ v1.53).

11. Image analysis and quantification

For time-lapse live imaging analysis, first time-lapse reconstructions were generated using the Bio-Formats plugin (Linkert et al., 2010) in Fiji (ImageJ v1.53). Then, automated segmentation of mammary buds was performed using a homemade segmentation model based on U-Net (Ronneberger et al., 2015). Segmented masks and raw image were input in the ImageJ plugin, BTrack, for tracking the growing branch tips. BTrack allows the users to remove or create new end points to manually correct the obtained tracks. We obtained the average growth rate for each branch. Statistical analyses were performed in Prism (v9.2, GraphPad).

To determine bud surface area in the presence of FGF10 in the medium, segmented masks were obtained from each timepoint using the U-Net model previously described. Generated masks were manually checked and corrected against raw data for consistency prior to extracting area measurements. Surface area was measure for each timepoint and statistical analyses were performed in Prism (v9.2, GraphPad).

To analyse N1Cre/β-cat/Tom increase in area, buds were outlined manually for each time using the freehand selection tool in Fiji (ImageJ v1.53). The increase in area was calculated as the area at day 4 or day 6 in culture compared to the area measured at day 1. Statistical analyses were performed in Prism (v9.2, GraphPad).

For smRNA-FISH dot counting, Find Maxima tool in Fiji (ImageJ v1.53) was used to find the highest peak values in the images by using a threshold previously specified. Then, a custom-written ImageJ macro was used to create 3 parallel regions of interest (ROIs) with a ring-shaped surface. Finally, the number of dots in each ROI was calculated for each smRNA-FISH probe. The percentage of dots in each ring was calculated as the ratio of number of dots in a specific ROI to the total number of dots in the 3 ROIs (outer, middle and internal ring). Statistical analyses were performed in Prism (v9.2, GraphPad).

3D segmentation of single cells was performed using a customized semi-automated image analysis pipeline (Kapoor & Carabaña, 2021). Segmented masks were imported into the TrackMate Fiji plugin (Ershov et al., 2022), for semi-automatic 3D cell tracking. To calculate the distance of a cell to the BM, TrackMate computed the minimum distance from the centre of the cell (considering X and Y coordinates) to the surface rendering of the tissue. Tracks are color-coded based on the distance to the ROI.

Track velocities and lengths were measured using the publicly available *Chemotaxis and Migration* tool (Ibidi, free download from https://ibidi.com/chemotaxis-analysis/171-chemotaxis-and-migration-tool.html) (Zantl & Horn, 2011).

For EdU quantification 3 independent explants in each condition were analysed. For each explant, independent regions of interest were randomly selected in discrete Z-slides. The mammary epithelium was outlined manually in Fiji using the tdTomato or luminal lineage marker staining as a guide (ImageJ v1.53). Hoechst images were processed with a median filter (1-2px). StarDist (Schmidt et al., 2018; Weigert et al., 2020) was used to segment and quantify number of Nuclei and EdU$^+$ nuclei within the outlined mammary epithelial tree region in Fiji (ImageJ v1.53). EdU$^+$ nuclei were expressed as a percentage over total number of nuclei. Statistical analysis was performed in Prism (v9.2, GraphPad).

12. scRNA-seq data analysis

scRNA-seq experiments

Single cell capture and library construction were performed using the 10x Genomics Chromium Single Cell 3’ v3.1 kit following the manufacturer’s instructions, for samples of different developmental stages in WT or N1Cre/β-cat/Tom animals. The libraries were sequenced with an Illumina NovaSeq 6000 sequencer by the *Next Generation Sequencing* platform of Institut Curie.

Data pre-processing and quality control

The 10x Genomics Cell Ranger Single-Cell Software Suite was used for demultiplexing, read alignment and unique molecular identifier (UMI) quantification (http://software.10xgenomics.com/single-cell/overview/welcome). The pre-built mm10 reference genome obtained from the 10X Genomics website was used to align the reads. Then, the count matrices were individually loaded for each sample in R and analysed using the Seurat package v4.0.5 (Hao et al., 2021).

Genes expressed in less than 3 cells and cells with UMI count < 5000 and mitochondrial UMI count > 6% were removed. This resulted in the following total number of high-quality cells: 228 at E13.5, 59 at E14.5, 740 at E15.5, 409 at P0 in WT mice and 238 cells at P0 in N1Cre/β-cat/Tom mutant embryos.

Cell annotation, differential gene expression analysis and data visualization

Normalization: Objects were normalized separately using the SCTransform method, implemented in the "SCTransform" function from Seurat. Briefly, this method regresses out the sequencing depth variation between cells using a negative binomial regression model with regularized parameters (Hafemeister & Satija, 2019).

scRNA-seq data dimension reduction and clustering: Principal Component Analysis (PCA) was performed on the top 2000 highly variable genes of the SCT assay from the "SCTransform" step. The top 15 principal components (PCs) were further selected (based on inspection of PC elbow plot) to perform graph-based clustering and cell cluster detection. All the Uniform Manifold Approximation and Projection (UMAP) plots (McInnes et al., 2018) were computed using the "RunUMAP" Seurat function with default Seurat parameters.

Cell cluster identification: Cell clustering was performed using a two-step wise approach, using the "FindNeighbours" and "FindClusters" functions, respectively. "FindClusters" function was used setting the resolution parameter to 0.8.

Differential expression analysis: Cell-type marker genes for each cluster were identified using the function "FindAllMarkers" function in Seurat, with detected in minimum cell fraction > 10% and log-fold change > 0.1. Then, cell clusters were manually annotated based on cell type specific markers that are known to be enriched in each cell population (Table 1).

Cell population	Cell type specific markers
Proliferative cells	'Pclaf', 'Ncapg2', 'Smc2', 'Tyms', 'Tuba1b', 'Hmgb2', 'Top2a', 'Tacc3', 'Cenph', 'Cdk1', 'Tubb5', 'Diaph3', 'Cenpf'
Basal layer of the skin	'Col17a1', 'Bcl11b', 'Fam83g', 'Krt15'
Suprabasal layer of the skin	'Krt1', 'Krt10'
Mammary mesenchyme	'Tnc', 'Cav1', 'Esr1', 'Nrp2', 'Dlk1', 'Tcf4', 'Dcn', 'Sparc', 'Vim', 'Pdgfra', 'Col1a1'
Mammary epithelial cells	'Krt5', 'Krt8', 'Epcam', 'Krt14', 'Trp63', 'Cdh1'

Table 1 | Cell type specific markers used to annotate cell clusters.

Signature construction: a single-cell ID score for "basal-like" and "luminal-like" cells was calculated based on previously published transcriptomic analyses of adult MECs (Kendrick et al., 2008). The scores were computed using the Seurat function "AddModuleScore".

Detection of GFP-positive cells: The sequence was blasted in raw FASTQ files to the GFP sequence, and the reads aligning to the GFP were selected. Then, a filtering was applied to the alignment by 1) Matching length $\geqq$ 80, and 2) No mismatching basepairs. Finally, we extracted the 10X cell barcode and UMI sequences, which enabled us to count the UMI count per matched sequence per cell. The cells with GFP UMI count > 5 were marked as GFP-positive cells.

3D trajectory and pseudotime analysis

For this analysis, only the epithelial cell clusters from E13.5, E14.5, E15.5, and P0 were considered. The pre-processing steps previously described were re-applied (normalization, PCA, and basal and luminal score). Epithelial cells were then mapped in a 3D space including the luminal score and basal score on the x-axis and the PC related to developmental time on the y-axis. For each cell cluster, the coordinates of the centre in the 3D space with the median for each dimension were calculated and called "pseudo-bulks". A minimum spanning tree (MST) was generated to connect all the pseudo-bulks. Based on the MST, the basal and luminal trajectories were inferred.

To get the pseudotime of each cell along the basal or luminal trajectories, each cell was projected in the 3D space to the basal and luminal trajectories separately. Then, the pseudotime for each cell was defined as their distance from the initial point of the trajectory.

The luminal and basal gene expression heatmap was generated on the pseudotime with the "pheatmap" package. Briefly, the genes with the top 10% variation across cells within a lineage were selected. The gene expression values were smoothed versus the pseudotime using the generalized additive model (GAM). The hierarchical gene clusters were generated with Euclidean distance and Complete clustering algorithm.

Cell-cell interaction analysis

The cell-cell interaction analysis was done using the CellPhoneDB version 3.0.0 (Vento-Tormo et al., 2018) with a p-value threshold of 0.01. The CellPhoneDB database is publicly available at https://www.cellphonedb.org/. It is a curated database of ligand-receptor interactions that allow to predict cell-cell interactions in transcriptomic data. CellPhoneDB was used on our scRNA-seq E15.5 dataset between both mesenchymal clusters (sub-epithelial cluster and dermal mesenchyme cluster) against the basal-like epithelial cell cluster.

13. Statistics and Reproducibility

We have used at least n=3 animals for each experiment, and experiments with at least n=3 replicates were used to calculate the statistical value of each analysis. Statistical tests and further graphs were prepared in Prism (v9, GraphPad). All graphs show mean ± SEM. Differences between groups were assessed with two-tailed unpaired T-test with Welch's correction. Statistical analyses between the localization of two RNA probes were assessed with two-way ANOVA test. The significance threshold was $p < 0.05$. * indicates $p<0.05$, ** indicates $p<0.01$, *** indicates $p<0.001$, and **** indicates $p<0.0001$.

Results

CHAPTER 1

Cell fate specification underlies positional cues during branching morphogenesis of the embryonic mammary epithelium

1.1. Aim of the study

The precise timing and mechanisms of lineage specification during embryonic mammary gland morphogenesis remains unclear, hampered by the lack of specific markers for early fate specification. Here we profiled mammary epithelial cells at different embryonic developmental times using scRNA-seq to determine the developmental timing and molecular signatures dictating mammary cell differentiation. Furthermore, many previous studies have focused on the mammary epithelium, disregarding mesenchymal cells surrounding the developing mammary gland. Thus, the heterogeneity and molecular signatures of the mesenchymal cellular compartment are currently unknown. In this study, I investigated the heterogeneity of mesenchymal cells during embryonic mammary development, as well as the paracrine interactions mediating communication between the mammary mesenchyme and MECs.

The results of this *Chapter* are about to be submitted for publication.

1.2. Results

Lineage restriction is a progressive developmental process

How changes in mammary tissue architecture during morphogenesis translate into differential gene expression patterns that drive the lineage specification of individual cells during development remains unknown in many tissue contexts. To address this in the MG, we performed scRNA-seq analysis of mouse embryonic mammary glands at four developmental times spanning mammary bud invagination (E13.5), initial sprouting events at the presumptive onset of lineage segregation (E14.5 and E15.5) (Lilja et al., 2018) and post-natal branching morphogenesis (at birth or Post-natal day 0, P0) (Figure 9A). At each timepoint, we micro-dissected mammary buds from female mouse embryos (pooling tissues from 7- 12 embryos isolated from different pregnant dams) and isolated mammary epithelial (EpCAM$^+$) and stromal (EpCAM$^-$) cells by FACS for scRNA-seq using the 10x Chromium platform. Basal and luminal subpopulations are indistinguishable in embryonic mammary glands using standard EpCAM and CD49f gating strategies (Figure S1A).

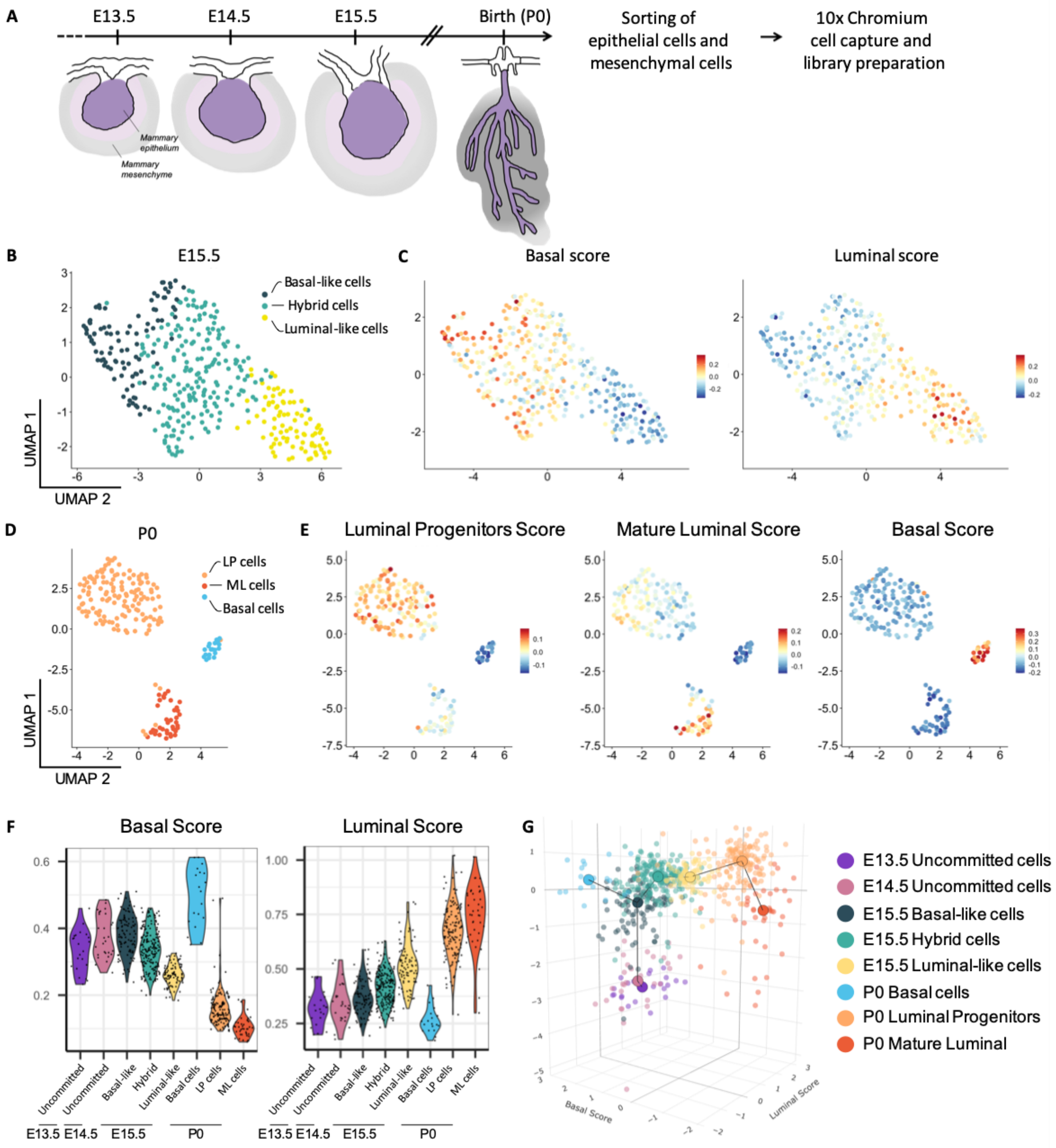
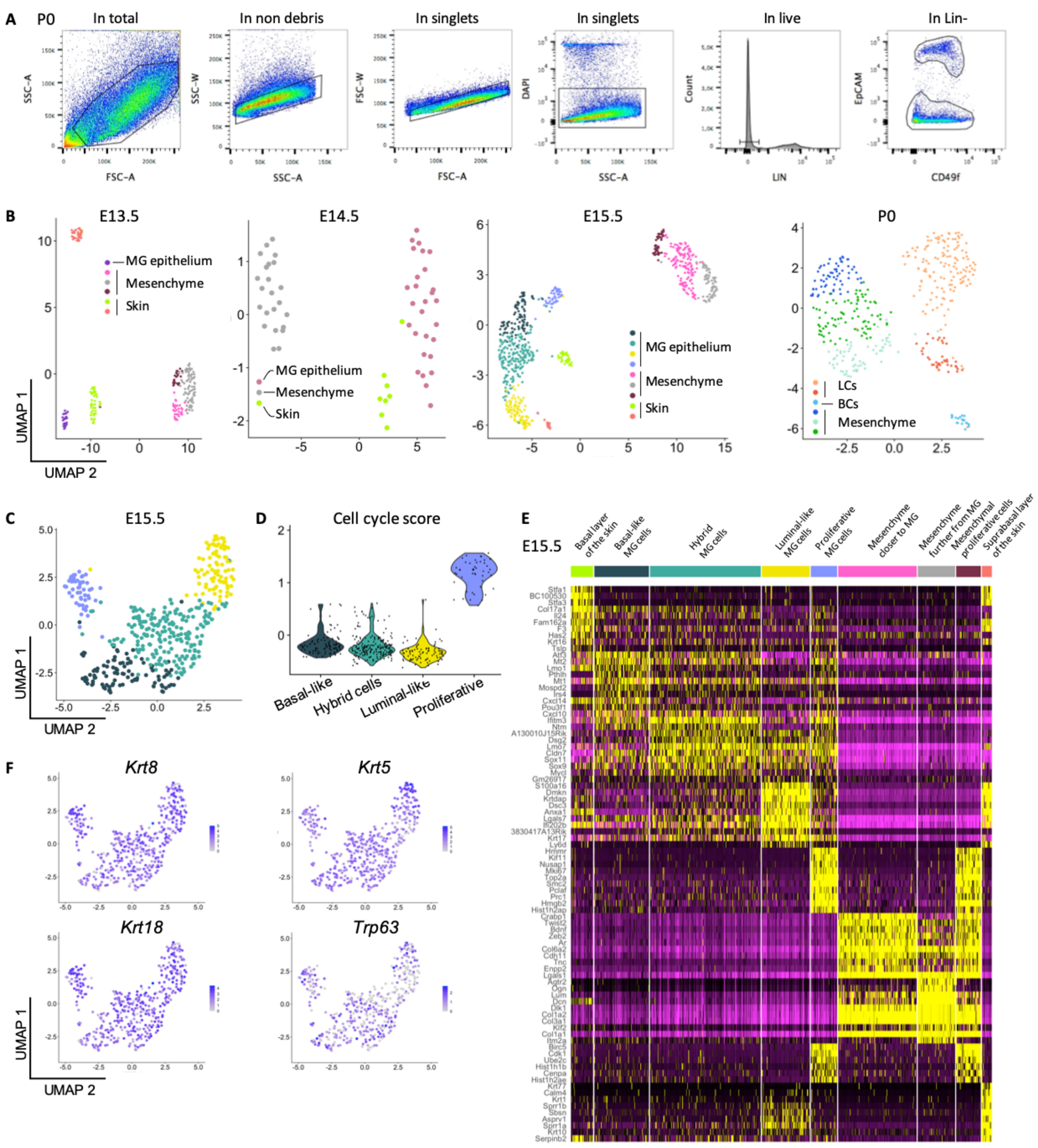


Figure 9 | Developmental atlas of the transcriptional signatures and 3D trajectory analysis of luminal and basal differentiation of single mammary epithelial cells from E13.5 until birth.

(A) Scheme showing the isolation and sequencing strategy of mammary embryonic cells at four developmental stages spanning embryonic MG development. (B) UMAP plot of embryonic MECs isolated at E15.5 after subset analysis of non-proliferative MG epithelial cells. Cells are colour-coded by cluster. (C) UMAP plots from (B) colour-coded according to the expression of the single-cell ID scores in MECs: basal score (left) and luminal score (right). (D) UMAP plot of MECs isolated at P0 after subset analysis of MG epithelial cells. (E) UMAP plots from (D) colour-coded according to the expression of luminal progenitors (LP), mature luminal (ML) and basal cell (BC) scores. (F) Violin plots showing the expression levels of the basal and luminal scores in each cluster. (G) 3D trajectory of MECs from E13.5 at the origin of the mammary cellular hierarchy to P0 MECs positioned at the end of two divergent differentiation routes.

Using the Seurat R package (Stuart et al., 2019), unsupervised clustering of single cell expression data revealed distinct cell clusters at E13.5, E14.5, E15.5 and P0, respectively (Figure S1B), which were manually annotated by matching enriched gene sets with known markers of mammary epithelium, mesenchyme and skin cells (See *Materials and Methods* section). With the objective of mapping MECs undergoing lineage commitment early in embryogenesis, we removed contaminating skin cells (Figure S1B) and performed a sub-clustering analysis of epithelial populations at each developmental timepoint. A cluster composed of proliferative epithelial cells was identified at E15.5, based on a list of cell cycle related genes (Figure S1C, 1SD, see *Materials and Methods* section). This cluster was excluded from further analysis to avoid a strong effect of the proliferative genes on the mammary differentiation trajectories. This analysis identified a single population of MECs at early developmental times, E13.5 and at E14.5, whereas 3 transcriptionally separate cell subsets were distinguishable at E15.5 and at P0 (Figure 9B, 9D, S1B). The identification of 3 MECs clusters at E15.5 was surprising, as previous studies observed a single population around this developmental stage (Giraddi et al., 2018). To investigate this further, we calculated a single-cell ID score for "basal-like" and "luminal-like" cells based on previously published transcriptomic analyses of adult MECs (Kendrick et al., 2008). A higher single-cell ID score reflects increasing similarity to the reference cell type: adult basal or luminal cells. Interestingly, this analysis revealed that E15.5 MECs can already be resolved into 3 distinct groups: luminal-like cells, basal-like cells and a hybrid cell population co-expressing luminal and basal genes (Figure 9C, S1E). As expected, lineage markers commonly used to distinguish LCs (*Krt8, Krt18*) from BCs (*Krt5, Trp63*) in the postnatal mammary gland were co-expressed in all 3 MECs clusters at E15.5 (Figure S1F). Importantly, alongside established markers for adult BCs (*Lmo1, Pthlh, Cxcl14*) and LCs (*Anxa1, Ly6d*) (Kendrick et al., 2008), this analysis also identified genes that have not been previously ascribed to distinct mammary BC or LC populations.

By applying a computed ID score for each epithelial adult cell type (Kendrick et al., 2008) to the 3 transcriptionally distinct cell populations observed at P0 (Figure 9D), BCs (*Acta2*$^+$*, Myh11*$^+$), luminal progenitors (LP) (*Notch1*$^+$, *Aldh1a*3$^+$, *Lypd3*$^+$) and mature luminal (ML) cells (*Prlr*$^+$, *Cited1*$^+$, *Esr1*$^+$) could be clearly distinguished (Figure 9E). This corroborates our previous findings (Lilja et al., 2018) indicating that MECs are already committed to specific lineages at birth. Moreover, these results are consistent with previous snATAC-seq analyses of the embryonic mammary gland, which also identified 3 separate clusters at E18.5 (Chung et al., 2019). Collectively, our data supports a model whereby mammary epithelial cell lineages are progressively being specified throughout development and are well segregated at birth.



Supplementary Figure 1. Related to Figure 9 | Lineage committed cells in early MG development. (A) Representative FACS dot plots of the gating strategy used to sort P0 epithelial and mesenchymal cells. (B) UMAP plots of embryonic MECs and surrounding mesenchymal cells isolated by scRNA-seq at E13.5, E14.5, E15.5 and P0. Cells are colour-coded by cluster. (C) UMAP plot of embryonic MECs isolated at E15.5 after subset analysis of all MECs (including proliferative cells shown in light blue). (D) Violin plot representation of the cell cycle score in each mammary epithelial cluster at E15.5. (E) Heatmap showing the expression of genes specific for each cell cluster at E15.5. Each column is colour-coded according to the cell cluster from (B). (F) UMAP plots from (C) showing the expression of specific luminal (*Krt8* and *Krt18*) and basal (*Krt5* and *Trp63*) genes commonly used to distinguish adult LCs and BCs but unable to discriminate distinct cell clusters at E15.5.

We next ordered the cells along pseudotemporal trajectories to infer the differentiation path of embryonic MECs towards a luminal or basal fate. To achieve this, we attributed the second principal component in PCA to developmental stage (y-axis) and plotted this against the basal and luminal scores (computed above (Kendrick et al., 2008); x-axis) (Figure 9F- G, see *Materials and Methods* section). The resulting plot indicates, as predicted, that E13.5 mammary cells lie at the origin of the mammary cellular hierarchy, with E15.5 cell populations occupying intermediate positions and P0 MECs positioned at the end of two divergent trajectories, representing the binary cell fate choice between basal or luminal differentiation. Remarkably, we notice that basal-like cells at E15.5 can either move towards the P0 basal cluster, or to a hybrid cell state that will give rise to LCs (Figure 9G), suggesting that they might lie at the origin of both lineages at this stage.

Together, our temporal scRNA-seq atlas reveals the molecular changes associated with progressive lineage restriction and identifies subsets of MECs that appear to be already biased towards basal or luminal cell fates at embryonic day E15.5. Thus, both committed (i.e. conceivably unipotent) and undifferentiated (possibly multipotent) cells (Lilja et al., 2018) likely exist at this important developmental stage in mammogenesis, which coincides with the first morphogenetic events of mammary epithelial branching and duct elongation.

Luminal and basal progenitors are already spatially segregated at E15.5

We next sought to identify differentially expressed genes for each mammary epithelial cluster by examining their dynamic expression profile towards luminal (Figure 10A) or basal (Figure 10B) differentiation trajectories. While our compiled scRNA-seq atlas emphasised the vast cellular heterogeneity of the embryonic mammary epithelium, this extended analysis identified different patterns of expression along the process of basal (Figure S2A) or luminal (Figure S2B) differentiation throughout embryonic development (from E13 to P0).

On the basal trajectory we found 5 distinct patterns of expression. Patterns 3 and 4 contained genes with sustained increased expression in early embryonic developmental times, at E13.5 and E14.5. Known key regulators of mammary bud epithelial cells are highly expressed only during early embryonic development, including *Ndnf, Pthlh, Msx1, Tbx3, Sostdc1,* whose expression is lost before birth. Moreover, multiple Wnt signalling genes, such as *Wnt3, Wnt6* and *Fzd10*, were enriched at this stage.

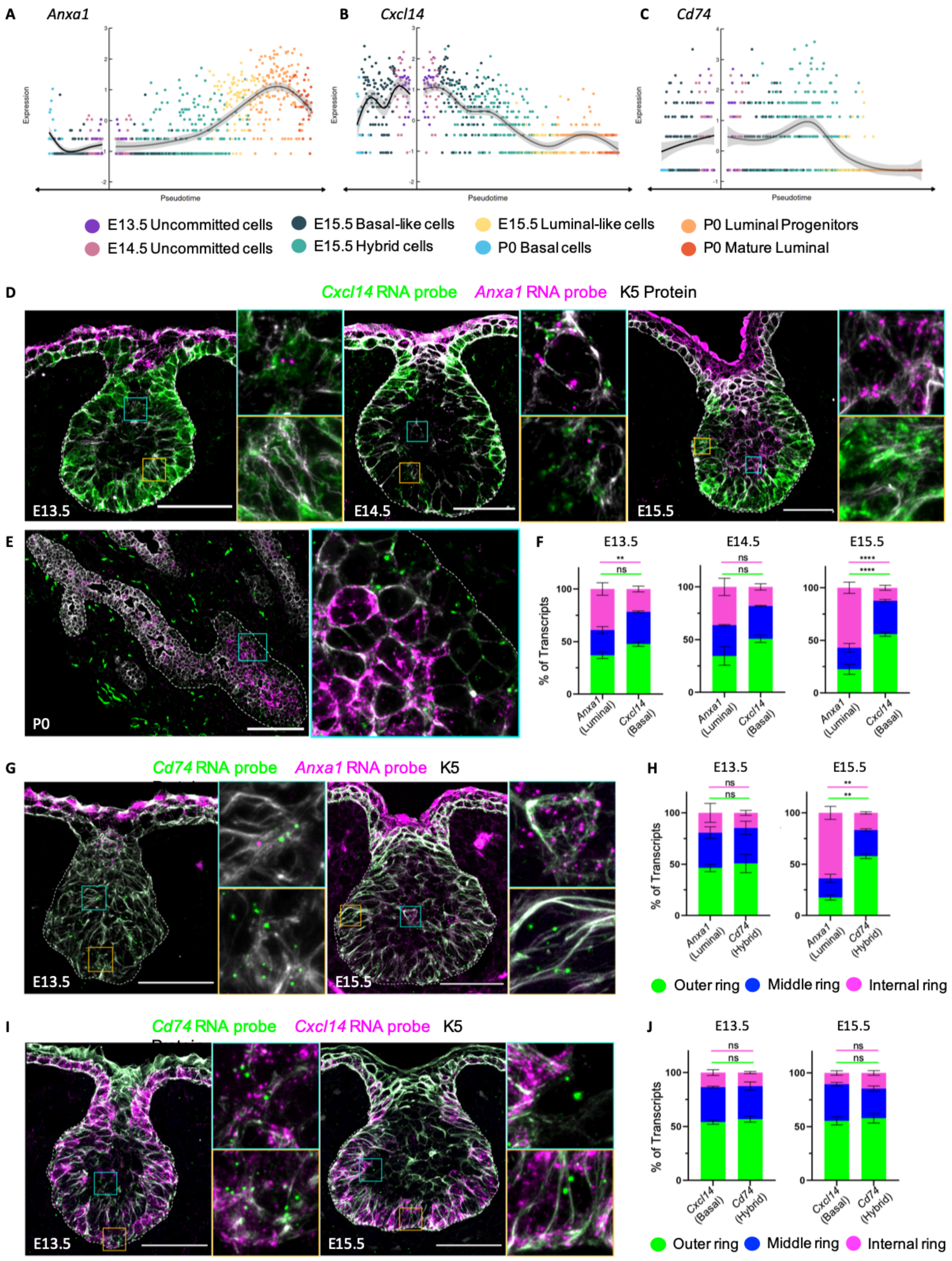


Figure 10 | Luminal and basal progenitors are already physically separated at E15.5.

(A, B and C) Examples of genes with pseudotime-dependent expression towards luminal differentiation (*Anxa1*, A), basal differentiation (*Cxcl14*, B) or with a higher expression in the hybrid cluster at E15.5 (*Cd74*, C). Cells are colour-coded by cluster. (D and E) Representative sections of embryonic mammary buds at E13.5, E14.5 and E15.5 (D) and P0 (E) showing the expression of *Cxcl14* (in green) and *Anxa1* (in magenta) detected by RNAscope and immunostained with K5 (in white). Dotted lines delineate the BM. Scale bars: 50 µm (D), 100 µm (E). (F) Quantification of the proportion of *Anxa1* and *Cxcl14* transcripts in each ring at each developmental stage. (G) Representative sections of embryonic mammary buds at E13.5 and E15.5, showing the expression of *Cd74* (in green) and *Anxa1* (in magenta) detected by RNAscope and immunostained with K5 (in white). Dotted lines delineate the BM. Scale bar: 50 µm. (H) Quantification of the proportion of *Cd74* and *Anxa1* transcripts in each ring at each developmental stage. (I) Representative sections of embryonic mammary buds at E13.5 and E15.5 showing the expression of *Cd74* (in green) and *Cxcl14* (in magenta) detected by RNAscope and immunostained with K5 (in white). Dotted lines delineate the BM. Scale bar: 50 µm. (J) Quantification of the proportion of *Cd74* and *Cxcl14* transcripts in each ring at each developmental stage. Statistical significance in (F), (H) and (J) was assessed with two-way ANOVA test between the two probes. The statistical analysis was performed between the outside ring (green line) and the inside ring (magenta line). ns: non-significant, ** indicates $p<0.01$ and **** indicates $p<0.0001$.

Another set of genes shows a transient phase of upregulation during the E15.5 development stage (Pattern 5) and is enriched for genes involved in migration (*Ptp4a1, Fam60a, Ralbp1*). Genes involved in basal myoepithelial differentiation in the mammary gland were found to be upregulated towards the P0 basal cluster (Pattern 1), these included myosin-related proteins (*Myl6, Myl9, Myh11, Mylk*) and genes involved in ECM organization (*Lama4, Adamts4, Itga1, Col9a1, Col4a1, Col11a1, Col16a1*). In addition, towards the P0 basal cluster we also found increased levels of genes regulating cell proliferation (*Top1, Cdkn1a, Runx1, Fosl1*), cytoskeletal organization (*Tuba1c, Tubb6*) and angiogenesis (*Tnfrsf12a, Serpine1, Tgfa, Hbegf*) in Pattern 2, suggesting that epithelial growth is highly regulated at this developmental stage.

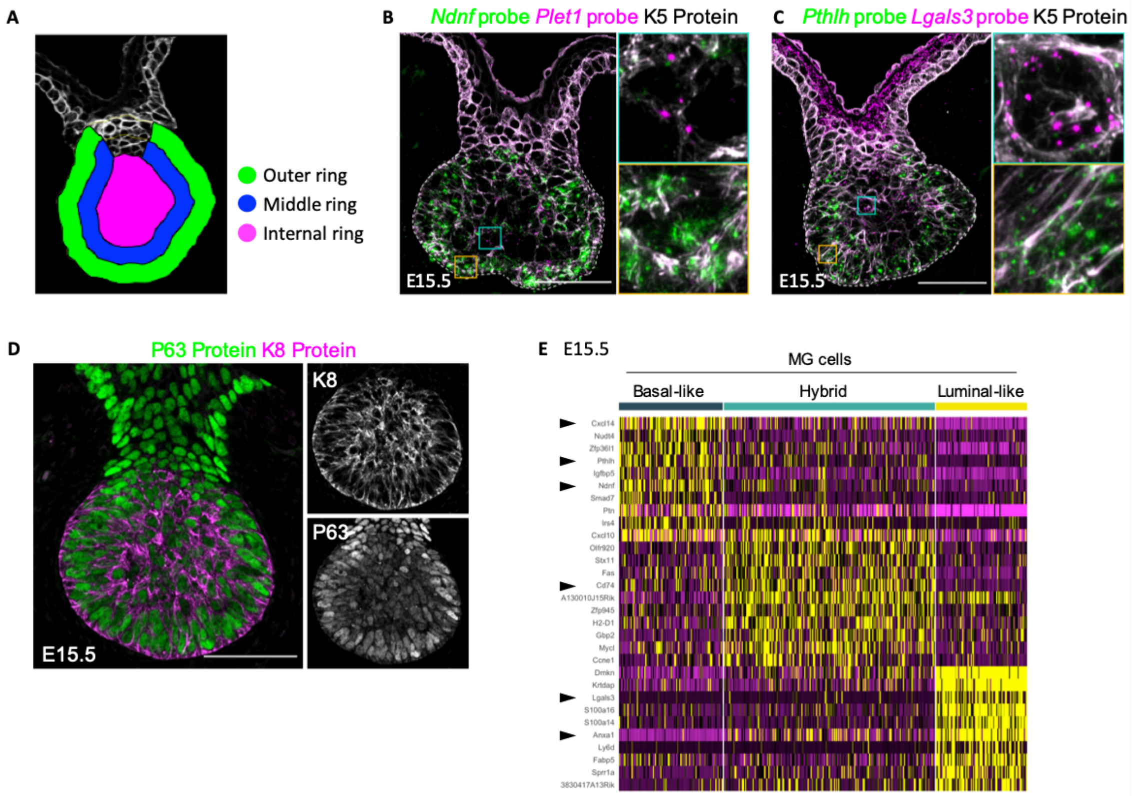
On the other hand, we found 7 distinct expression patterns along the luminal differentiation trajectory. As expected, the pattern presenting increasing expression contains genes with known luminal characteristics, such as *Krt8, Krt18* and *Krt19* (Pattern 3). A second group of genes that is switched on during late stages of differentiation is enriched for ML cells markers, such as *Cited1* and *Prlr* (Pattern 1). The set of genes that is activated at the beginning of the differentiation process and repressed along the luminal trajectory comprises typical basal markers, such as *Krt5* and *Krt14* (Pattern 7). *Sox11* also presents this dynamic pattern of expression – gradually decreasing its expression along the differentiation process. *Sox11* is expressed in MECs only during the early stages of MG embryonic development – when MG epithelial cells are largely quiescent – and it is no longer detected by E16.5, consistent with our results. Of interest, *Sox11* has been recently involved in cell fate regulation during MG embryonic development (Tsang et al., 2021).

Supplementary Figure 2. Related to Figure 10 | Pseudotime ordering identifies genes associated with early luminal and basal differentiation.

(A and B) Heatmaps illustrating genes exhibiting a differential pattern of expression along the pseudotime (from E13.5 to P0) towards the basal lineage (A) or the luminal lineage (B). Genes (rows) are clustered based on the dendrogram in the left and colour-coded by their expression levels (from blue to red). The gene expression levels were smoothed using the GAM and scaled by row. Genes of interest are indicated in the right. Each set of genes with a specific pattern is colour-coded in the left: 5 distinct patterns in the basal lineage (A) and 7 unique patterns in the luminal lineage (B).

Genes involved in epithelial stratification, such as *Lgals7, Dsc3* and *Krtdap*, are switched on only during the luminal-like stage at E15.5 (Pattern 6). Pattern 2 comprises genes encoding for several Heat shock proteins (Hsps). There is growing evidence that Hsps may impact neurodevelopment through specific pathways regulating cell differentiation, cell migration or angiogenesis (Miller & Fort, 2018).

This suggests a potential spatial segregation of cells acquiring luminal or basal characteristics during embryonic tissue morphogenesis. To further investigate this possibility, we first identified genes that exhibited a lineage-specific expression pattern along the differentiation trajectories (Figure 10A-C). These included *Cxcl14, Ndnf* and *Pthlh* (Figure S2A) and *Anxa1, Plet1* and *Lgals3* (Figure S2B) for basal and luminal lineage specification respectively. Using single molecule RNA-fluorescence *in situ* hybridization (smRNA-FISH), we subsequently examined the spatiotemporal expression pattern of selected genes at distinct stages of mammary embryonic development. Probes for the luminal specific membrane-associated protein Annexin A1 (*Anxa1*) (Fankhaenel et al., 2021) and the basal-specific secreted chemokine *Cxcl14* (Sjöberg et al., 2016) revealed that at early embryonic stages (E13.5), *Cxcl14* is expressed in all MECs, and *Anxa1* is lowly expressed in rare cells homogeneously distributed within the mammary bud (Figure 10D). However, at the critical developmental time of E15.5, the transcripts for these two genes show divergent spatial distribution patterns, with *Anxa1* expression being mainly confined to cells in the inner bud region and *Cxcl14* transcripts restricted to the external cell layers in contact or close proximity with the BM (Figure 10D). By P0, *Anxa1* and *Cxcl14* showed clear luminal and basal restricted expression respectively (Figure 10E). To quantify the spatial segregation of gene expression, we divided the mammary bud into three concentric "rings" (outer, middle and internal regions) (Figure S3A) and counted the number of RNA molecules (represented by each dot) within each ring for both markers. This unbiased approach confirmed the uniform expression pattern of *Anxa1* and *Cxcl14* transcripts in all 3 regions of the mammary buds at E13.5 (Figure 10F). By E15.5, however, *Anxa1* transcripts were prominently restricted to the middle and inner ring, while *Cxcl14* transcripts appeared preferentially localized to the middle and outer ring of the mammary bud (Figure 10F). This was particularly intriguing as all MECs still express K5 (in white in Figure 10D, S3B, S3C) and all other known markers of adult LCs and BCs at this developmental stage (Figure S3D). Thus, *Anxa1* and *Cxcl14* represent novel markers of MECs committed to luminal and basal lineages, respectively, as early as E15.5 during mammary development. Analogous smRNA-FISH analysis of E15.5 mammary buds with additional probes suggested that *Ndnf* and *Pthlh* are also expressed in embryonic basal committed MECs, while *Plet1* and *Lgals3* expression likely mark cells biased towards the luminal lineage (Figure S3B- C), further corroborating our temporal scRNA-seq analysis (Figure S2A- B).



Supplementary Figure 3. Related to Figure 10 | Identification of novel genes that distinguish lineage-biased embryonic mammary cells.
(A) Cartoon illustrating the outer ring (in green), middle ring (in blue) and internal ring (in magenta) used in our unbiased quantitative analysis. (B and C) Representative sections of embryonic mammary buds at E15.5 showing the expression of *Ndnf* (basal gene, in green) and *Plet1* (luminal gene, in magenta) (B) or *Pthlh* (basal gene, in green) and *Lgals3* (luminal gene, in magenta) (C), detected by RNAscope and immunostained with K5 (in white). (D) Single optical section showing the expression of the luminal epithelial marker K8 (in magenta), and the basal epithelial marker P63 (in green) in an embryonic mammary bud at E15.5. K8 and P63 are co-expressed by all MECs at E15.5. (E) Heatmap illustrating the expression of genes specific for each MEC cluster at E15.5. Each column is colour-coded according to the cell cluster from Figure 9B. Black arrowheads indicate genes previously used for RNAscope experiments.

In light of our findings that a proportion of MECs are already lineage committed at E15.5, we next sought to examine the spatial localisation of cells possessing a hybrid basal-luminal expression signature within the developing mammary bud. To this aim, we searched for genes associated with the hybrid cell cluster. A promising candidate marker gene for this cluster was the HLA class II cell surface receptor *Cd74* (Figure 10C, S3E), previously proposed as a mammary stem cell marker (dos Santos et al., 2013). smRNA-FISH analysis revealed that, while *Cd74* expression overlapped with both *Anxa1* and *Cxcl14* in early mammary embryonic development (E13.5), the vast majority of *Cd74* transcripts were located in the middle and outer regions of the mammary bud at E15.5, coinciding with *Cxcl14* expression (Figure 10G- J). Thus, the hybrid cells identified by transcriptomic analysis at E15.5 appear to be primarily localized in proximity with the BM, where basal-committed cells are also found within growing mammary buds.

Collectively, our spatial transcriptomic data reveal that the embryonic basal-like and luminal-like mammary cell clusters identified by scRNA-seq are already located in defined and mutually exclusive positions within the mammary bud at E15.5, at the onset of branching morphogenesis. Spatial segregation of mammary embryonic progenitors may conceivably underlie their state of differentiation and lineage commitment at this critical stage of embryonic mammary development.

Identification of two spatially distinct mesenchymal cell populations in the embryonic mammary stroma

Mammary epithelial buds at E13.5 are surrounded by a specialised mammary mesenchyme, subsequently undergoing sprouting to invade the underlying fat pat precursor at around E15.5 to initiate the first stages of branching morphogenesis. Paracrine signalling between mammary epithelial and surrounding mesenchymal cells is indispensable for this process (Spina & Cowin, 2021; Wansbury et al., 2011). To gain further insights into mammary mesenchymal patterning during embryonic development, we focused our analysis to the scRNA-seq data of mesenchymal cells at E13.5, E15.5 and at P0. Clustering of non-epithelial cells identified three mammary mesenchymal cell subsets at each stage (Figure 11A). By computing a cell cycle score based on a list of cell cycle-related genes (see *Materials and Methods* section), we identified proliferative cell clusters exclusively at early developmental timepoints, E13.5 and E15.5, indicating that proliferative cell populations are mostly absent at birth (Figure S4A).

We next examined the transcriptional signatures of the two non-proliferative mesenchymal clusters at E15.5 to identify markers defining each subset (Figure S4B). Candidate genes included *Esr1* (coding for the ERα) and *Plagl1* (coding for the zinc finger protein PLAGL1), which were highly expressed in opposing mesenchymal clusters (Figure 11B- C, S4B). Immunostaining for ERα showed clear expression in mesenchymal cells directly surrounding the mammary bud (Figure 11D), as previously reported (Wansbury et al., 2011). Immunofluorescence analysis for PLAGL1, on the other hand, revealed that PLAGL1^{+} mesenchymal cells are located further away from the mammary epithelium (Figure 11E). These results suggest that the two transcriptionally distinct mesenchymal populations are also differentially localized within the embryonic mammary stroma, and can be classified based on their proximity to the mammary epithelial bud: we thus termed the cells closest to the epithelium as sub-epithelial mesenchyme and the ones further away as dermal mesenchyme.

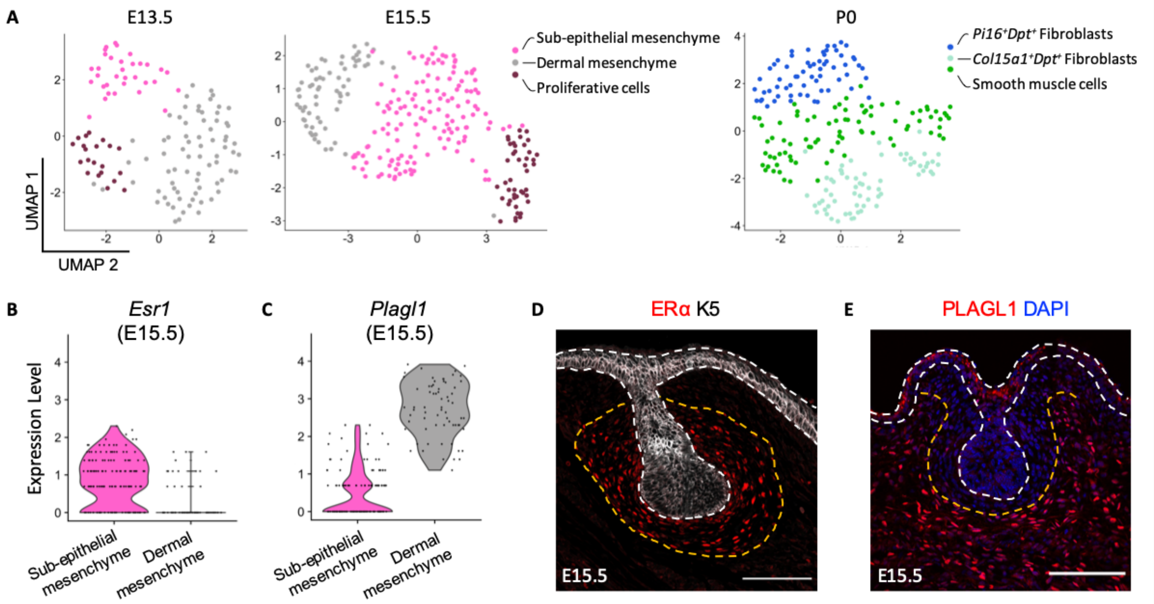
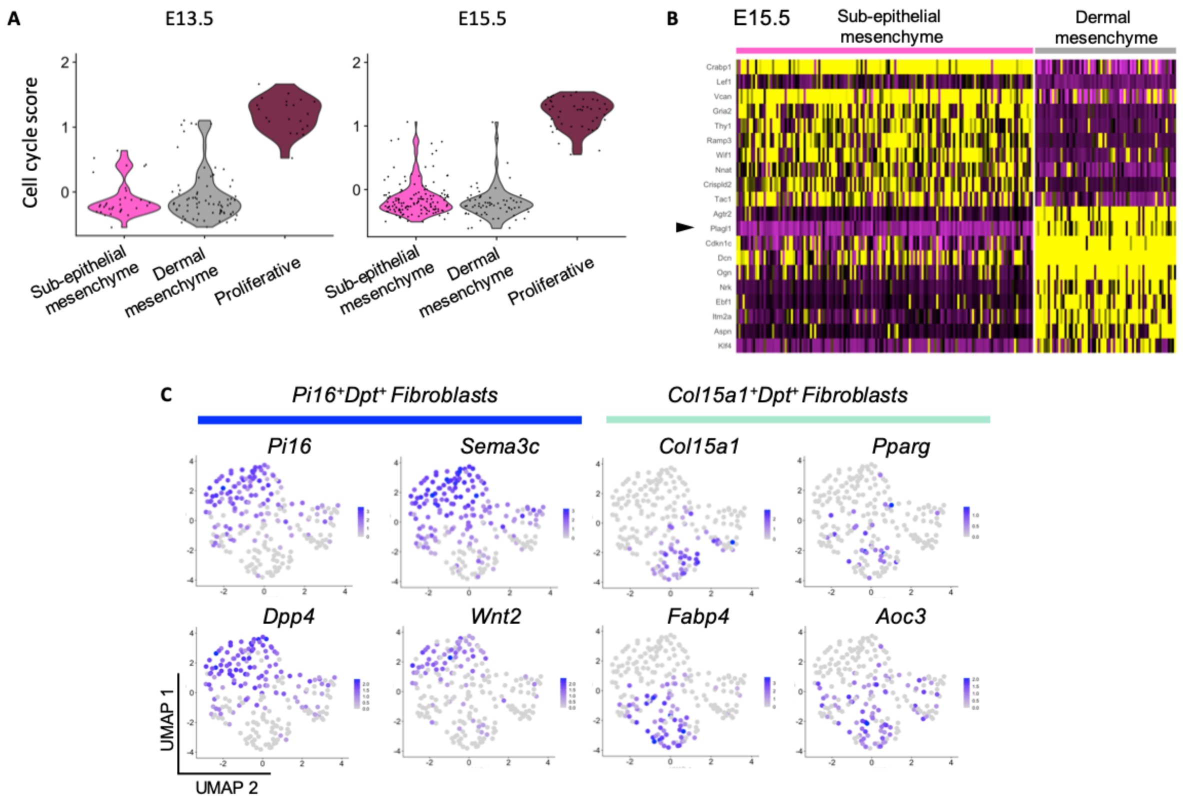


Figure 11 | The embryonic mammary mesenchyme contains two spatially distinct cell populations. (A) UMAP plots of embryonic mammary mesenchymal cells isolated at E13.5, E15.5 and P0 after subset analysis. Cells are colour-coded by cluster. (B and C) Violin plots representing the expression levels of *Esr1* (B) and *Plagl1* (C) in sub-epithelial and dermal mesenchyme respectively, at E15.5. (D and E) Representative sections of embryonic mammary buds at E15.5 immunostained for ERα (in red) and K5 (in white) (D) or PLAGL1 (in red) and DAPI (in blue) (E). Dotted lines delineate the BM (in white) and the two mesenchymal compartments (in orange). Scale bars: 100 μm.

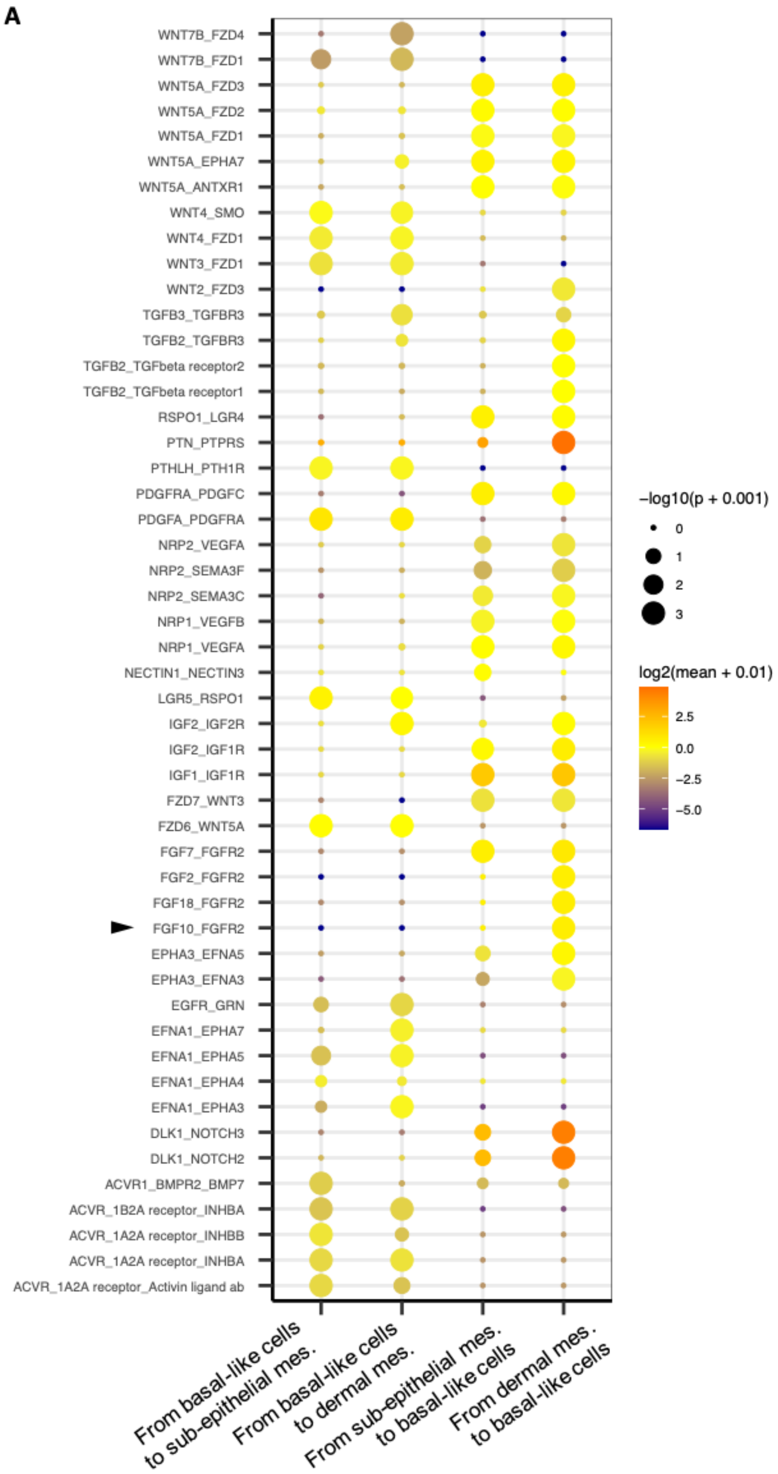
The heterogeneity of mesenchymal cells and the complexity of the mammary stroma increases at birth, where two clusters of *Dpt*$^{+}$ fibroblasts can be distinguished, a *Col15a1*$^{+}$ and a *Pi16*$^{+}$ cluster. Interestingly, the *Col15a1*$^{+}$*Dpt*$^{+}$ population also expresses *Fabp4*, *Pparg* and *Aoc3,* surface markers of pre-adipocytes, whereas the *Pi16*$^{+}$*Dpt*$^{+}$ population expresses *Dpp4*, *Sema3c* and *Wnt2*, which have been described to be upregulated in subcutaneous mesenchymal progenitors (Merrick et al., 2019) (Figure 11A, S4C). Structural and matricellular proteins of the ECM (*Col4a1, Col4a2, Col18a1, Mmp19, Sdc1, Sparcl1*) are also highly expressed in the *Col15a1*$^{+}$*Dpt*$^{+}$ population. Finally, the third mesenchymal population identified at P0 displays elevated expression of *Eln, Mfap4, Mgp*, genes typically expressed by smooth muscle cells.



Supplementary Figure 4. Related to Figure 11 | The heterogeneity of mesenchymal cells increases at birth. (A) Violin plots representing the cell cycle score in each mammary mesenchymal cluster at E13.5 and E15.5. (B) Heatmap illustrating the expression of genes specific for each mesenchymal cluster at E15.5. Each column is colour-coded according to the cell cluster from Figure 11A. The black arrowhead indicates *Plagl1*, previously used to label the dermal mesenchyme. (C) UMAP plots from Figure 11A illustrating the expression of cluster-specific genes in mesenchymal cells at P0.

FGF10 produced by the dermal mesenchyme is an important regulator of embryonic mammary morphogenesis

Given the importance of the mammary stroma on branching morphogenesis, in addition to our findings on the spatial patterning of mesenchymal cells at E15.5, we next sought to computationally predict specific paracrine interactions between the identified mesenchymal cell subsets and MECs using CellPhoneDB, a bioinformatic tool designed to predict highly significant ligand-receptor interactions between two cell types from scRNA-seq data (Vento-Tormo et al., 2018). We focused on ligand-receptor interaction pairs between the sub-epithelial or dermal mesenchyme and the basal-like cluster of MECs at E15.5, which we established to be in direct contact or in close proximity to the BM (Figure 10D- E). This approach highlighted several developmental signalling pathway components (including FGF, Wnt and Notch receptors and ligands) as putative mediators of the cross-talk between basal-like cells and the sub-epithelial or dermal mesenchyme at E15.5 (Figure S5A). Of particular interest, specific



Supplementary Figure 5. Related to Figure 12 | Ligand-receptor interaction pairs identified between the two mesenchymal populations and the basal-like cluster at E15.5.

(A) CellPhoneDB analysis with the predicted ligand-receptor interactions between the two mesenchymal populations, sub-epithelial or dermal mesenchyme, and basal-like cells at E15.5 and vice versa (p-value < 0.01). The arrowhead highlights the ligand-receptor interaction between FGF10 and FGFR2 that was functionally investigated in mammary embryonic *ex vivo* cultures.

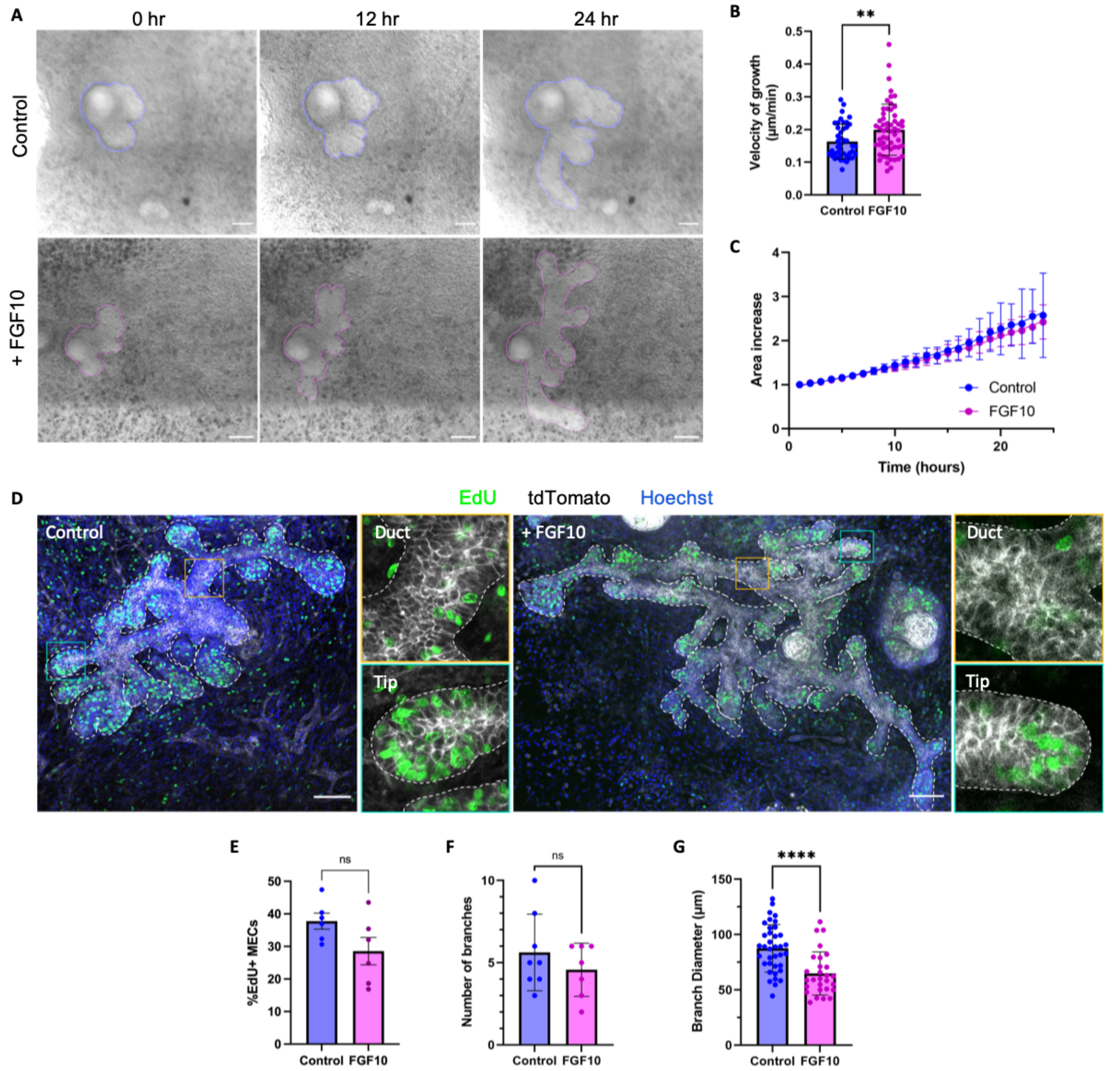


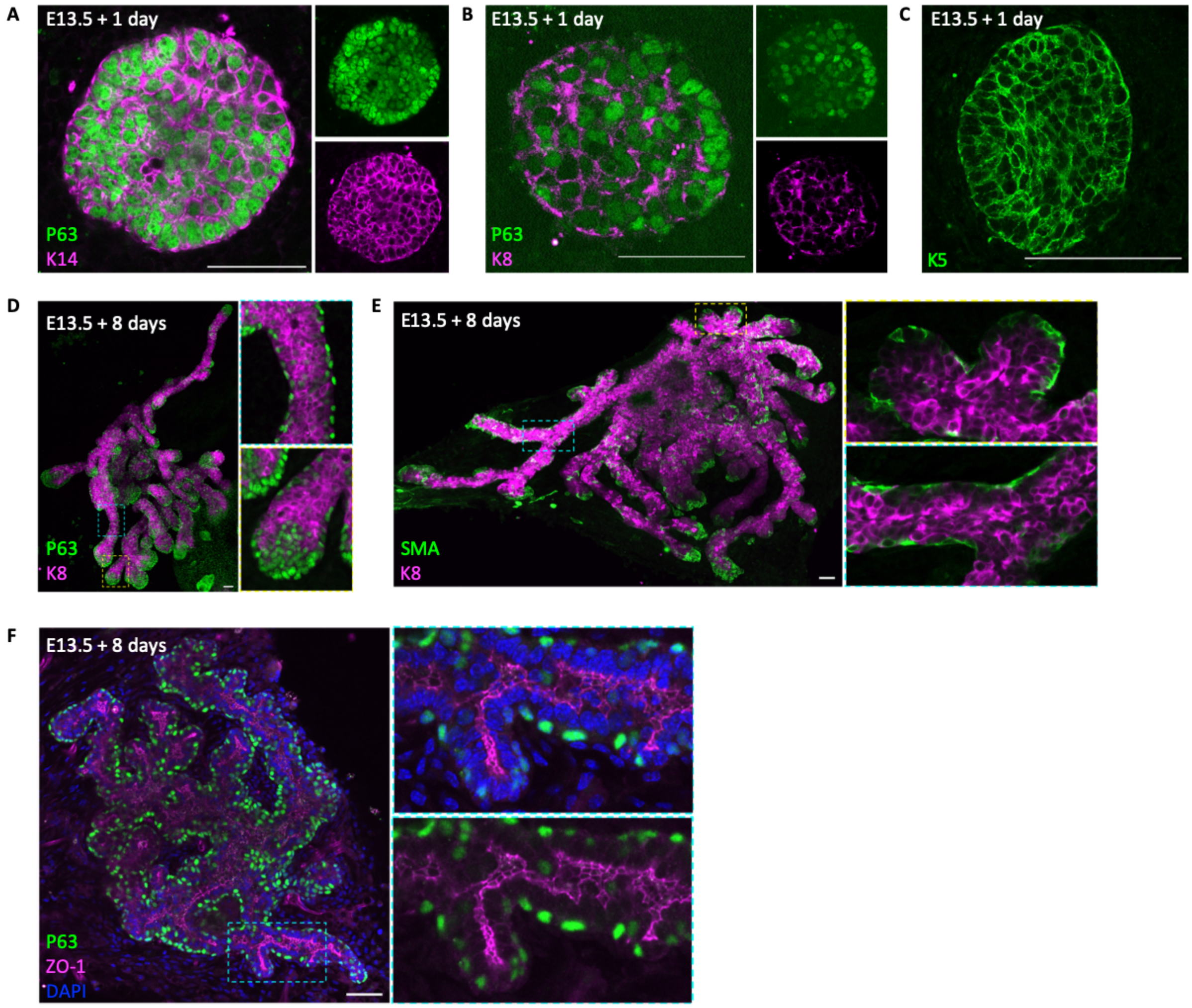
Figure 12 | FGF10 accelerates embryonic mammary branching without affecting cell proliferation.

(A) Time-lapse images of a mammary explant grown in control medium (top) or in the presence of FGF10 (bottom) for 24 hr. T= 0h refers to 4 days in culture. Scale bars: 100 µm. The rendered surface of the mammary epithelium is outlined in blue (in the control bud) and in magenta (in the FGF10 condition). (B) Quantification of the velocity of branch growth in control conditions (n= 43) and in the presence of FGF10 in the medium (n= 56). (C) Fold change increase in area in control and FGF10 conditions. In both cases, the area is doubled within 16 hr in culture. (D) Representative whole-mount immunostaining of an embryonic mammary gland cultured in control and FGF10 conditions showing Edu$^+$ cells (in green), membrane tdTomato (in white) and DAPI (in blue). Mammary buds were dissected at day E13.5 and cultured *ex vivo* for 7 days. Orange outlined insets show a duct region and blue outlined insets show a tip region. (E, F and G) Quantification of Edu$^+$ cells (E), number of branches (F) and branch diameter (G) in control and FGF10 conditions. Statistical significance was assessed with two-tailed unpaired T-test with Welch's correction. ns: non-significant, ** indicates p<0.01 and **** indicates p<0.0001.

interactions between the FGFR2 and its soluble ligand FGF10, as well as between the Transforming growth factor beta receptors TGFBR1 and TGFBR2 and their ligand TGFB2 were highly significant between basal-like MECs and the more distant dermal mesenchymal cells (Figure S5A). To functionally assess the validity of this computational prediction, we sought to investigate the impact of exogenous FGF10 on embryonic branching morphogenesis by live cell imaging of mammary buds established in *ex vivo* cultures (Carabaña & Lloyd-Lewis, 2022; Voutilainen et al., 2013). Explant cultures provide a highly tractable system for modelling embryonic mammary cell behaviour and branching morphogenesis (Carabaña & Lloyd-Lewis, 2022; Voutilainen et al., 2013). Embryonic mammary buds along with their surrounding mesenchyme were dissected at E13.5 and cultured *ex vivo* on a filter allowing to form an air-liquid interface (see *Materials and Methods* section for a detailed protocol). MECs within mammary embryonic buds expressed both basal and luminal markers (K5, K14 and P63 for basal cells and K8 for luminal cells) after 24 hours in culture (Figure S6A- C), consistent with *in vivo* observations (Wansbury et al., 2011). During 8 days in culture, embryonic mammary buds undergo sprouting and branching, recapitulating the morphogenetic events occurring *in vivo* (Figure S6D- E). Immunostaining of the resulting 8-day-old ductal tree (approximately corresponding to P0/P1 *in vivo*) revealed that MECs in the outer layer express basal markers such as P63 (Figure S6D, S6F) and α-SMA (Figure S6E), while inner layer cells express the luminal marker K8 (Figure S6D- E). In addition, polarity acquisition appeared normal, as revealed by apical ZO-1 staining in the inner layer of luminal cells (Figure S6F). Thus, key aspects of embryonic mammary morphogenesis and epithelial lineage segregation can be reconstituted *ex vivo* in explant cultures.

Taking advantage of this powerful system, we next investigated the impact of FGF signalling by undertaking live-imaging of embryonic mammary explants cultured in the presence of FGF10 (Figure 12A). To measure the velocity of branch growth, we traced the endpoint of each branch at sequential timepoints acquired every 60 min for 24 hours and measured the distance travelled over time in control and FGF10 treated conditions. These experiments indicated that mammary branches grow faster when cultured in the presence of FGF10 (Figure 12B).

Stromal-produced FGF10 may accelerate branching morphogenesis by increasing either epithelial cell proliferation or motility. To discriminate between these two possibilities, we measured the planar surface area of mammary buds over time and found that tissue growth was not significantly affected by FGF10, since the explant area increased 2-fold within 16 hours of culture in both control and FGF10 conditions (Figure 12C). Furthermore, although FGF10 can be a potent mitogen in certain tissues, it did not appear to promote epithelial cell proliferation during branch elongation, as assessed by 5-ethynyl-2′-deoxyuridine (EdU) incorporation experiments (Figure 12D- E). Also, the number of branches in embryonic explant cultures supplemented with FGF10 was equivalent to control cultures (Figure 12F). However, the branch diameter at the base of branches was reduced in the presence of FGF10 (Figure 12G), suggesting that while MEC numbers are equivalent, cells may move faster along extending ducts, which consequently become thinner in the presence of FGF10. Our data therefore shows that, similar to what was found during pubertal branching morphogenesis (Hannezo et al., 2017), FGF signalling promotes branching of the embryonic mammary ductal tree also during the initial stages of embryonic development, conceivably by promoting epithelial cell motility.



Supplementary Figure 6. Related to Figure 12 | Mammary bud *ex vivo* cultures recapitulate embryonic mammary morphogenesis and epithelial lineage segregation.

(A, B and C) Representative images of mammary embryonic buds dissected at day E13.5 and cultured *ex vivo* for 1 day, immunostained for the following lineage markers: P63 (in green) and K14 (in magenta) (A), P63 (in green) and K8 (in magenta) (B), and K5 (in green) (C). (D, E and F) Representative images of mammary embryonic buds dissected at day E13.5 and cultured *ex vivo* for 8 days, immunostained for the following lineage and polarity markers: P63 (in green) and K8 (in magenta) (D), α-SMA (in green) and K8 (in magenta) (E), and P63 (in green) and ZO-1 (in magenta) (F). Scale bars: 50 µm (at day 1 – A, B and C), 100 µm (at day 8 – D, E and F).

CHAPTER 2

Live-imaging of embryonic mammary explants to analyse cellular dynamic behaviour driving mammary gland morphogenesis

2.1. Aim of the study

In the second Chapter, I studied the cellular mechanisms associated with MG branching and tubulogenesis during embryonic development.

Our recent results showed that multipotent MaSCs become lineage-restricted around embryonic day E15.5, coinciding with the first morphogenetic events that establish the mammary ductal network (Lilja et al., 2018). I therefore hypothesize that the loss of multipotency in the MG is linked to the cellular movements and arrangements occurring during tissue morphogenesis that lead to the branching of embryonic mammary buds. To investigate this hypothesis, my goal was to record branching morphogenesis during this timeframe of lineage specification by time-lapse microscopy in 3D explant cultures.

2.2. Results

Ex vivo cultures of embryonic mammary buds are a powerful model to study dynamic cell behaviour during branching morphogenesis and lineage commitment

To perform high-resolution time-lapse analysis of cell behaviour during the initial morphogenetic events leading to the formation of the rudimentary embryonic MG, we have established a pipeline for live confocal microscopy in embryonic mammary *ex vivo* cultures, that enabled me to track individual cells in a physiological environment for extended time periods. Firstly, as briefly mentioned in the previous chapter, I have optimized a previously developed embryonic mammary bud dissection protocol to facilitate the isolation of buds and their surrounding mesenchyme at E13.5, before the onset of the first mammary morphogenetic events (Carabaña & Lloyd-Lewis, 2022; Voutilainen et al., 2013). We also demonstrated that physiological tissue development and lineage markers restriction can be recapitulated in this system, validating the suitability of *ex vivo* cultures as a relevant model to study embryonic mammary branching morphogenesis (Figure S6).

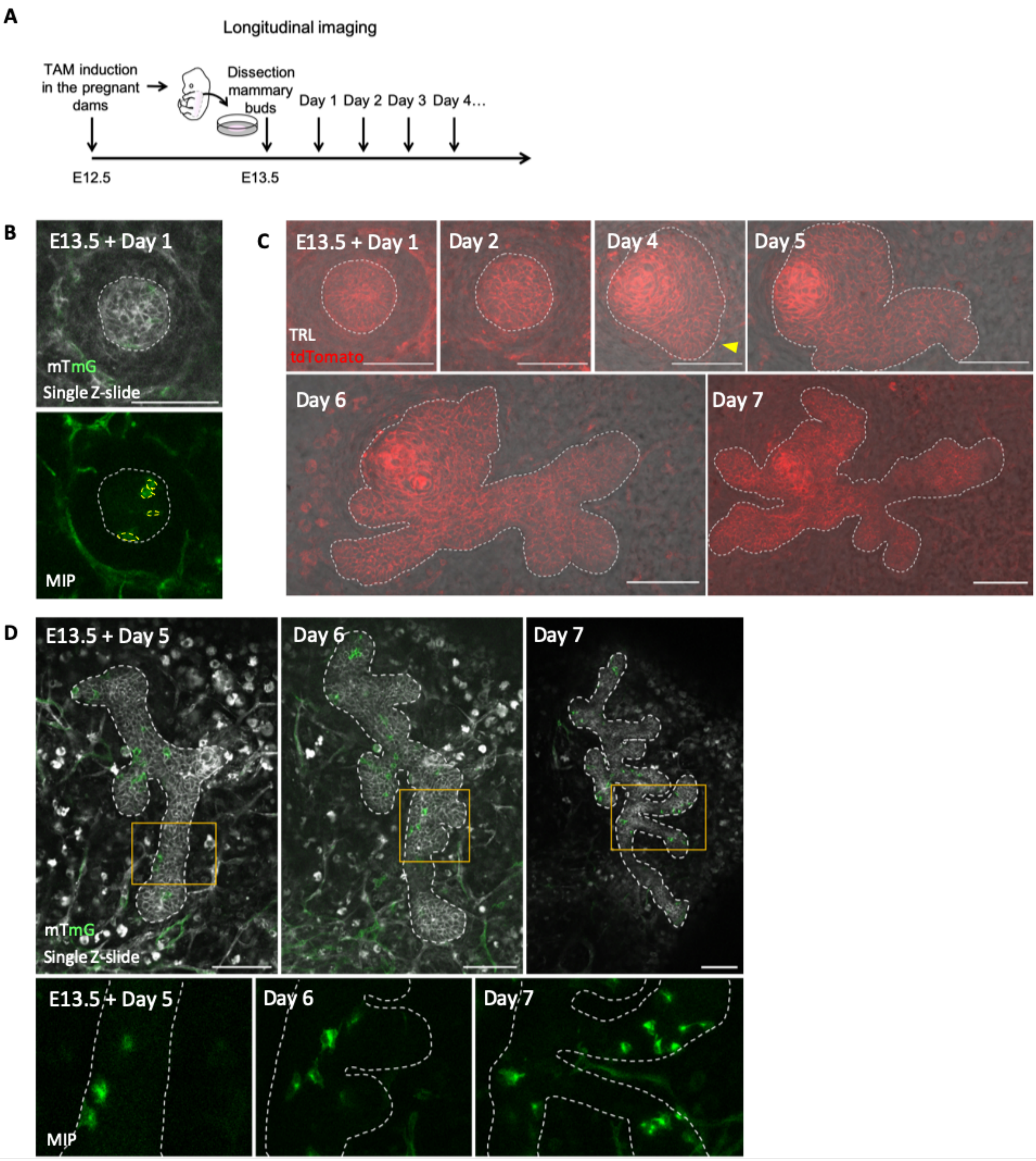


Figure 13 | Embryonic cultures allow to study cell dynamics during branching morphogenesis.
(A) Schematic representation of the experimental protocol for longitudinal imaging. N1Cre/Tom pregnant dams were administered TAM at day 12.5 of pregnancy. 24 hours later, mammary buds were dissected together with the surrounding mesenchyme. Mammary buds were imaged every day. (B) Optical section of a mammary bud after the first day in culture showing non-recombined tdTomato$^+$ cells (in white) and GFP$^+$ cells (in green) (top). Maximum intensity Z-projection (MIP) of GFP$^+$ cells (bottom). GFP labelling of individual Notch1-expressing cells in embryonic MECs and mesenchymal cells. Scale bar: 50 µm. (C) Representative images showing the growth of a N1Cre/Tom mammary bud dissected at E13.5 and imaged for 7 days. The yellow arrowhead indicates the first spouting event. Scale bar: 100 µm. (D) Longitudinal images showing the expansion of GFP$^+$ cells in a N1Cre/Tom mammary bud over time. Mammary buds were dissected at day E13.5 and cultured *ex vivo* for 5 days. Non-recombined tdTomato$^+$ cells (in white) and GFP$^+$ cells (in green). Insets show MIP of GFP$^+$ fluorescence. Scale bar: 100 µm. Dotted lines delineate the BM (A, B and C).

To investigate lineage specification during embryonic MG development, we used a lineage tracing approach to follow individual fluorescently-labelled cells within the same mammary bud over time. To achieve this, mammary buds were dissected from a N1-Cre^{ERT2}/R26^{mTmG} reporter mouse (Muzumdar et al., 2007) (henceforth referred to as N1Cre/Tom) allowing mosaic labelling of specific MECs and their progeny. In this model, all cells are labelled with a membrane-bound Tomato fluorescent protein. Tamoxifen administration in pregnant dams 24 hours before embryo dissection induces Cre-mediated recombination of the reporter allele in sporadic Notch1-expressing mammary cells, resulting in heritable membrane-bound EGFP fluorescent protein expression, allowing the progeny of individual labelled cells, as well as their fate outcomes, to be tracked over time. We have previously reported that the N1-Cre^{ERT2} line targets all MECs during embryogenesis and that these cells can give rise to all mammary lineages after birth (Lilja et al., 2018). By contrast, in the post-natal gland, Notch1 expression is restricted to unipotent hormone receptor-negative luminal progenitor cells (Rodilla et al., 2015).

To trace individual MECs during embryonic branching morphogenesis, E13.5 mammary buds were dissected and cultured as explant cultures 24 hours after *in vivo* TAM induction in pregnant N1Cre/Tom dams. Explant cultures were subsequently subjected to longitudinal confocal fluorescence imaging for up to 2 weeks (Figure 13A). After one day in culture, GFP-labelled MECs were homogeneously distributed in both the central region and the periphery of the bud (Figure 13B). This suggests that Notch1-expressing MECs do not have a specific localization within the mammary buds at early embryonic time points (Figure 13B). Typically, the first sprout appears between 3 to 4 days in *ex vivo* cultures, which corresponds to embryonic developmental days E16.5- E17.5 (Figure 13C). Strikingly, GFP-labelled cells were not clustered together in discrete locations of the branching epithelium, but were instead dispersed across several regions, suggesting that embryonic MECs undergo extensive rearrangements during branching morphogenesis (Figure 13D). Indeed, mammary epithelial cells appear to frequently exchange neighbours over time, as seen by the dispersion of labelled clonal progeny within the same branch (Figure 13D). Thus, this approach allows us to measure the dynamics of single mammary cells during embryonic MG development at high spatiotemporal resolution.

Dynamic behaviour of Notch1-expressing mammary cells by time-lapse imaging

Cell proliferation and hypertrophy (increase in individual cell size) play a minor role in the first epithelial invagination process that forms the mammary placode. Instead, migration-driven cell influx from surrounding epidermal cells drives the growth of the mammary rudiment (Trela et al., 2021). During the second morphological wave, the first sprout forms, triggering the onset of branching morphogenesis. Interestingly, EdU incorporation experiments revealed that cellular proliferation was more prevalent at the tips of the growing branches compared to ductal structures (Figure 12D), similar

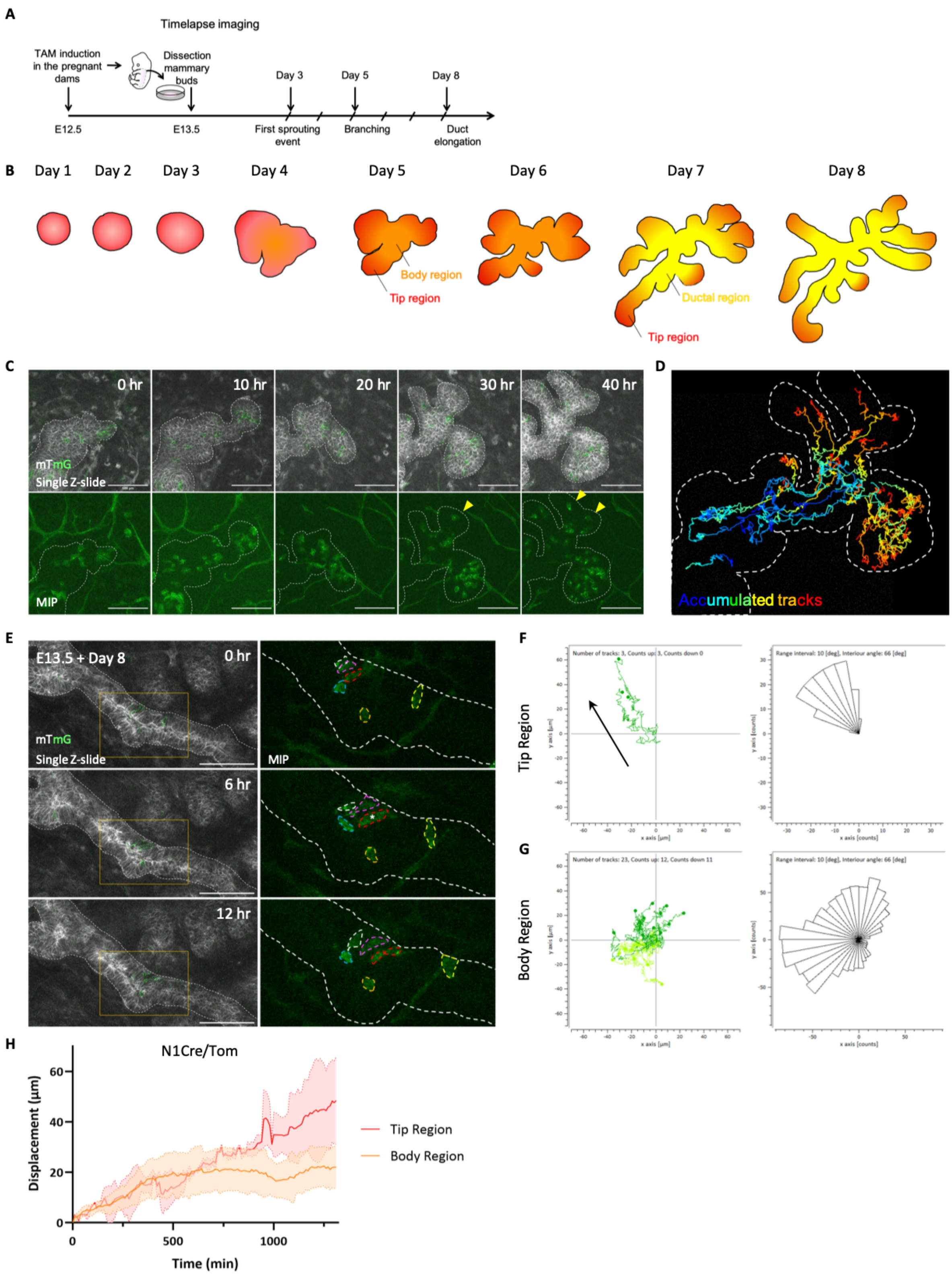


Figure 14 | Directional cell migration drives embryonic branching morphogenesis.

(A) Schematic representation of the experimental protocol for time-lapse imaging. N1Cre/Tom pregnant dams were administered TAM at day 12.5 of pregnancy. 24 hours later, mammary buds were dissected together with the surrounding mesenchyme. Branching morphogenesis and GFP^{+} cells dynamics was monitored for up to 24 hr by time-lapse confocal imaging. (B) Schematic representation of the growth of a mammary bud over time. The tip region is colour-coded in red, the body region in orange and the ductal region in yellow. (C) Time-lapse visualization of a N1Cre/Tom explant dissected at E13.5 and cultured for 4 days. Single optical section showing non-recombined tdTomato$^+$ cells (in white) and GFP$^+$ cells (in green) (top). Maximum intensity Z-projection (MIP) of GFP$^+$ cells (bottom). T= 0 hr refers to 4 days of culture. The yellow arrowheads indicate a dividing cell at 30 hr and its tracked daughter cells at 40 hr. Scale bar: 100 µm. (D) Trackmate figure illustrating all cellular tracks from time-lapse in (C) colour-coded according to time. Cold colours represent the beginning of the time-lapse, whereas warm colours are closer to the end of the time-lapse. (E) Time-lapse visualization of a N1Cre/Tom explant showing non-recombined tdTomato$^+$ cells (in white) and GFP$^+$ cells (in green) (left). Insets represent MIP of GFP$^+$ cells (right). T= 0 hr refers to 8 days after culture. Coloured dotted lines delineate the GFP$^+$ MECs. One dividing cell is indicated with an asterisk. Scale bar: 100 µm. (F and G) Graphic representations of the tracks of individual cells in a representative time-lapse of a N1Cre/Tom explant (left) and polar plots representing the distribution of trajectories angles of the same cells (right) in the tip region (F) and body region (G). T= 0 hr refers to 5 days after culture. Arrow indicates the branch growth direction. Angles show a preferentially orientated direction in cells in the branching tips (F), whereas random distribution for cells located in the body region (G). The lack of tracks in all directions (in G) is due to the presence of the tissue border preventing cell movements towards the right of the x-axis. (H) Net displacement of cells located in the tip (in red) and body region (in orange) in a representative time-lapse of a N1Cre/Tom explant. Net displacement was larger in cells located at the tips (0.037 µm/min vs 0.017 µm/min). Dotted lines demarcate the BM (in C, D and E).

to what is observed during pubertal branching morphogenesis (Scheele et al., 2017). This suggests that cellular proliferation within branching tips drive ductal elongation and invasion into the underlying secondary mammary mesenchyme, the precursor of the adult fat pad. To investigate this further, individual MECs in the branching tips were followed in four dimensions (x, y, z and t) to characterize cellular dynamics during the first morphological events spanning day 3 to day 6 in culture, as well as in already formed ducts after 7 days in culture (Figure 14A- B).

To thoroughly characterize cell movement patterns during branching morphogenesis, we first developed a segmentation and a tracking tool to analyse the dynamic behaviour of individual genetically labelled cells in embryonic mammary bud cultures (Figure 15A). As indicated by sequential daily imaging (Figure 13D), this cell tracking approach revealed that cells within the growing branches are extremely motile and continuously exchange position (Figure 14C- D), whereas cells present in the elongated ducts are remarkably static (Figure 14E). Interestingly, we observed that two daughter cells generated from a single division event could participate in the elongation of two different branches after branch bifurcation (Figure 14C). This means that a clone labelled with the same colour could be found scattered in different branches after extensive cell rearrangements occurring during embryonic branching morphogenesis, leading to misinterpretation of lineage tracing results. This finding, therefore, demonstrates the importance in understanding the clone size for clonal analysis experiments. In contrast, ductal cells mostly retained contact with the same neighbours over the imaging timeframes

(Figure 14E). For this reason all live imaging experiments were obtained from explants during the first few days of culture, as opposed to older explants grown for over 7 days, when elongating ducts have already formed. Even if I have not yet exploited and quantified these data, it is interesting to note that using this approach, I could also record the dynamic behaviour of Notch1-labelled mesenchymal cells surrounding the branching tips (note the green elongated cells in Figure 14C), probably pericytes, endothelial cells and fibroblasts, which could be involved in stroma remodelling to promote branching morphogenesis.

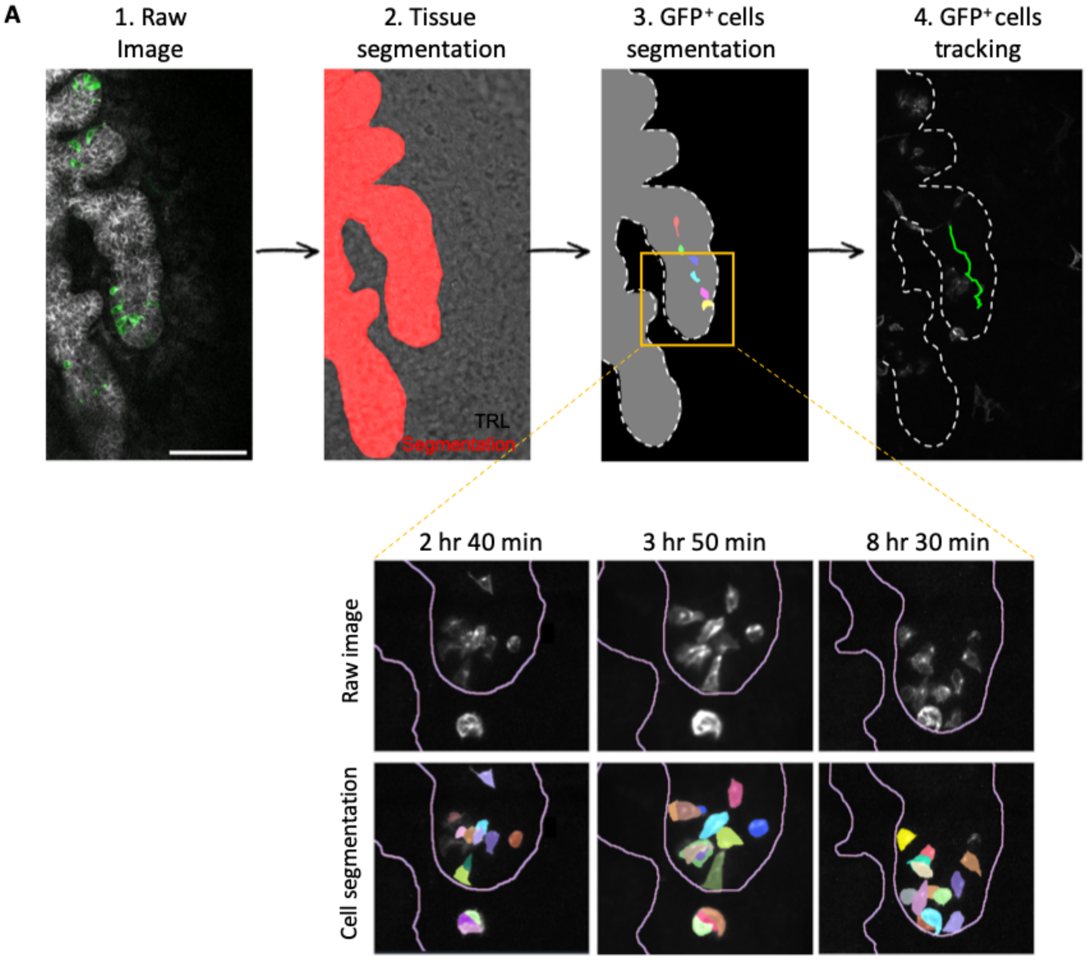


Figure 15 | Customized image analysis pipeline enables cell segmentation and cell tracking.

(A) (1) 3D visualization of a N1Cre/Tom *ex vivo* culture showing non-recombined tdTomato^{+} cells in white and GFP^{+} tracked cells in green. Scale bar: 100 µm. (2) Segmentation of a branched *ex vivo* culture (in red) overlaid on the raw image (Transmitted light). (3) 3D segmentation of the same GFP^{+} cell over time (each colour represents one timepoint). Segmentation of mammary epithelium is shown in grey. Inset: raw image (top inset) and 3D segmentation of all cells in three different timepoints along the time-lapse (bottom inset). Tissue outlined in purple. (4) Trajectory example (in green) for a GFP^{+} cell previously segmented in a growing embryonic mammary branch.

Next, I further characterized mammary epithelial cell movements by quantifying the directionality and net displacement of individual cells. While cell movements at the branch tips appeared to be preferentially orientated in the direction of the overall branch growth (Figure 14F), the movement of cells located in the body or ductal region appeared to be random (Figure 14G). Moreover, tip-located cells moved in a straight forward motion along the branch growth, compared to body-located cells. Therefore, the net displacement was also larger in the cells located in the tips (0.037 µm/min and 0.017 µm/min, respectively) (Figure 14H).

Thus, combining *ex vivo* mammary bud cultures and high-resolution time-lapse imaging allowed embryonic mammary cell dynamics to be assessed over different time scales ranging from hours to days. These findings suggest that embryonic mammary branching morphogenesis is driven by the collective action of cell migration and cell proliferation.

Cell fate acquisition during embryonic mammary development

As explained in *Results Chapter 1,* a major obstacle to study the origin of mammary epithelial lineages is the lack of specific markers that enable the identification and tracking of defined cell types. By evaluating cell displacement during branching morphogenesis, we sought to assess whether mammary embryonic buds are composed of two spatially distinct mammary cell populations i.e. internal vs surface (BM contacting) cell subsets, which may represent luminal and basal cells respectively, or if instead cells exchange freely between the outer and inner compartments.

We first colour-coded the 3D cell trajectories according to their distance to the basement membrane (Figure 16A). Interestingly, we found that MECs in the branching tips exhibit a wide range of migratory behaviours (Figure 16A- B). While some epithelial cells were observed to dynamically move across the external or internal compartments, a fraction of MECs (n= 6, corresponding to the 30% of the tracked cells) were found to crossover and remain into the new compartment. Additionally, we observed two modes of cell division: either (1) both daughter cells stay in the same cellular compartment, or (2) each daughter cell goes to a different compartment (Figure 16A- B). To quantify the different cellular behaviours observed during the lineage differentiation process, the distance of each tracked cell to the BM was plotted, which confirmed that MECs are able to stay within their original compartment, or cross into the other compartment (Figure 16C). Of note, a higher fraction of cells moves from the inner luminal to the outer basal compartment, whereas cells rarely move on the other direction, from a basal to a luminal localization.

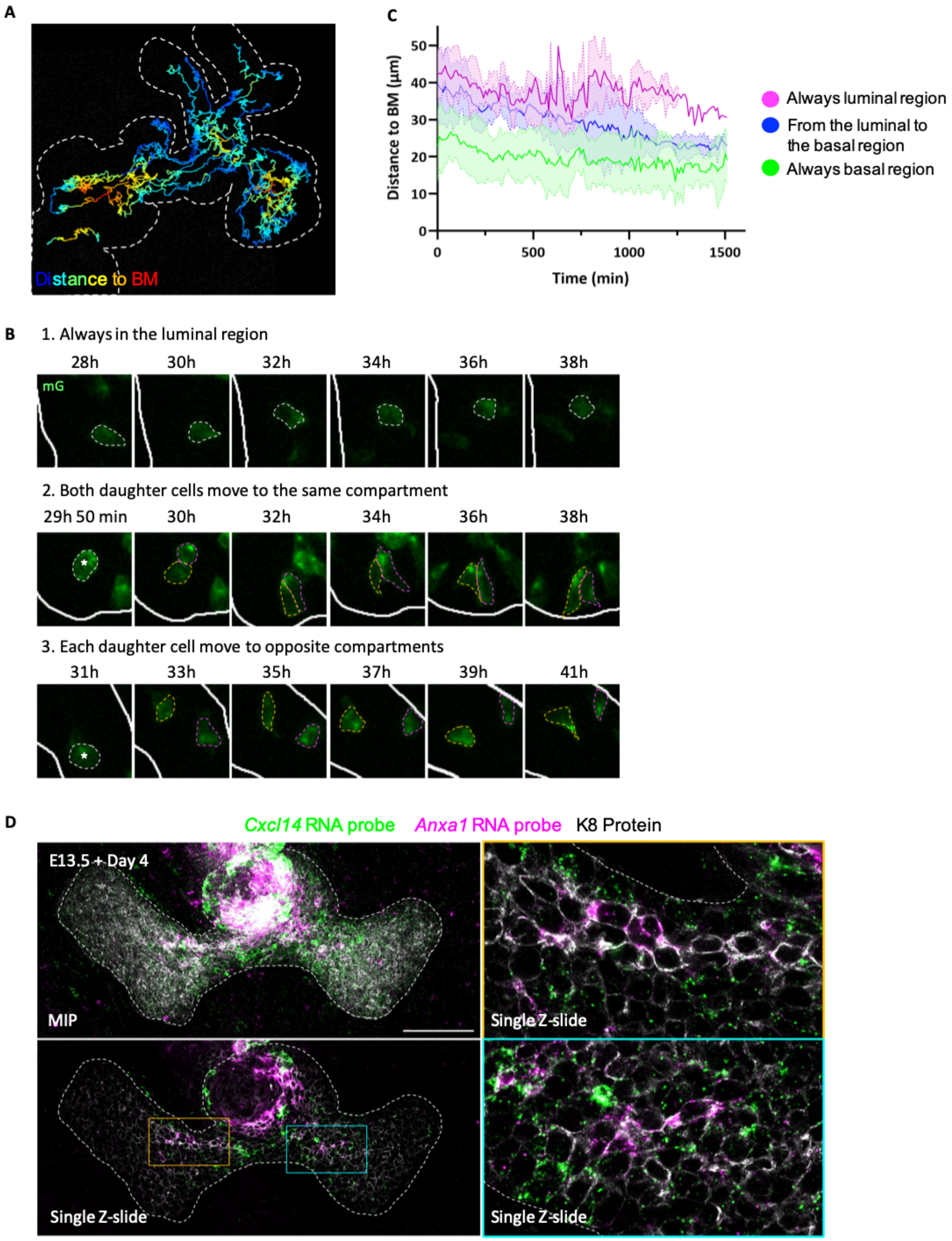


Figure 16 | Embryonic MECs exhibit different migratory behaviours.

(A) Trackmate figure illustrating all cellular tracks from time-lapse in (Figure 14C) colour-coded according to the distance to the BM. Cold colours represent tracks closer to the BM, whereas warm colours represent tracks further apart from the BM. (B) Snapshots from the time-lapse in (Figure 14C) showing either a cell that is always in the luminal compartment (1), two daughter cells that remain in the same compartment after division (2) or two daughter cells that move to opposite compartments after division (3). Dividing cells are indicated with an asterisk.

Recombined cells (GFP$^+$) are depicted in green. A white line demarcates the branching epithelium and colour-coded lines delineate MECs. T= 0 hr refers to 4 days after culture. (C) Plot illustrating the proportion of cells following the three modes of migratory behaviours observed: cells always in the luminal compartment (in magenta) (n= 4), cells that move from the luminal to the basal compartment (in blue) (n= 6) and cells always in the basal compartment (in green) (n= 10). (D) Maximum intensity Z-projection (MIP) (top) and a single optical slide (bottom) of an embryonic mammary bud dissected at day E13.5 and cultured *ex vivo* for 4 days. RNAscope for *Cxcl14* (in green) and *Anxa1* (in magenta) and immunostained with K8 (in white). Orange and blue outlined insets represent two different regions of the same explant in a single optical section (left). Scale bar: 100 µm. Dotted lines demarcate the BM (in A and D).

To determine whether cell fate determination relies on cell position within the tissue, we performed smRNA-FISH using the previously validated *Anxa1* and *Cxcl14* probes (Figure 10D- F), which represent novel markers of MECs committed to the luminal and basal lineage, respectively. We observed that *Cxcl14* is expressed in the majority of MECs. Interestingly, only *Cxcl14* transcripts were found in the most external cell layer of the explant. By contrast, *Anxa1* expression was exclusively detected in cells positioned in the innermost region of the mammary bud, with a preferential localization closest to the nipple area. In addition, *Anxa1* and *Cxcl14* were never found co-expressed in the same cells, and higher levels of K8 expression correlated with *Anxa1* expression (Figure 16D). This suggests that although the luminal and basal compartments are still not fully segregated at this developmental stage (mammary buds dissected at day E13.5 and cultured *ex vivo* for 4 days), lineage commitment has already started but MECs continue to dynamically exchange compartments until decisive cell fate specification.

Altogether, by using this lineage tracing approach, we are able to track and record the dynamics of a large number of cells across several divisions to understand how cell fate acquisition is coordinated with branching morphogenesis during MG development. However, many questions remain unanswered. Does a cell first specify to a specific lineage and only then it moves to that compartment? Or, on the contrary, does the position of a cell dictates its eventual fate? Additional experiments, including *in vivo* cell tracking using *Cxcl14*- or *Anxa1*-reporter mice models that would allow dynamic lineage progression to be captured as the tissue develops, would help to better understand the origin of mammary epithelial lineages during embryonic MG development.

CHAPTER 3

Canonical WNT/β-catenin signalling regulates embryonic branching morphogenesis

3.1. Aim of the study

In addition to the analysis of the progressive process of lineage restriction during embryonic MG development and the cellular dynamics during tissue morphogenesis, I have also started to address more mechanistic questions related to the coordination between tissue shape formation and cell fate specification, using a gain-of-function β-catenin mutant mice. Wnt/β-catenin signalling is required for the onset of MG morphogenesis (Chu et al., 2004) and plays critical roles in the determination of mammary basal cell fate (Yu et al., 2016). In addition, recent lineage-tracing studies undertaken in the lab revealed that mutant β-catenin stabilization during pubertal development and adult homeostasis leads to epidermal transdifferentiation of mammary epithelial cells when targeted to either basal or luminal cells (Lloyd-Lewis et al., 2022). Inspired by these results, I hypothesised that the comparative analysis of *wild-type* (WT) and β-catenin mutant embryonic mammary cells could inform us on how cells coordinate their dynamics to achieve correct cell fate specification during branching morphogenesis.

3.2. Results

β-catenin activation in embryonic epithelial cells impairs branching morphogenesis *in vivo*

To investigate the role of Wnt/β-catenin signalling in mammary cell differentiation and MG formation, I performed single cell transcriptomic data from β-catenin gain-of-function mutant mice (Harada et al., 1999) crossed with N1-Cre^{ERT2}/R26^{mTmG} mice (Rodilla et al., 2015) (henceforth referred to as N1Cre/β-cat/Tom) (Lloyd-Lewis et al., 2022) (Figure 17A). In these compound mice, the N1-Cre^{ERT2} line (Fre et al., 2011) is crossed to the double fluorescent reporter model R26^{mTmG} (Muzumdar et al., 2007), allowing the visualization of all cells with membrane-bound tdTomato fluorescence and of β-catenin mutant cells by membrane-bound EGFP upon tamoxifen-driven Cre activation.

I induced active β-catenin expression in E12.5 embryos and isolated mammary epithelial and mesenchymal cells by FACS 7 days later, at P0 (Figure 17B- C). Unsupervised clustering of the sequenced dataset for these mutant mice revealed 4 distinct cell clusters: 1 epithelial population and 3

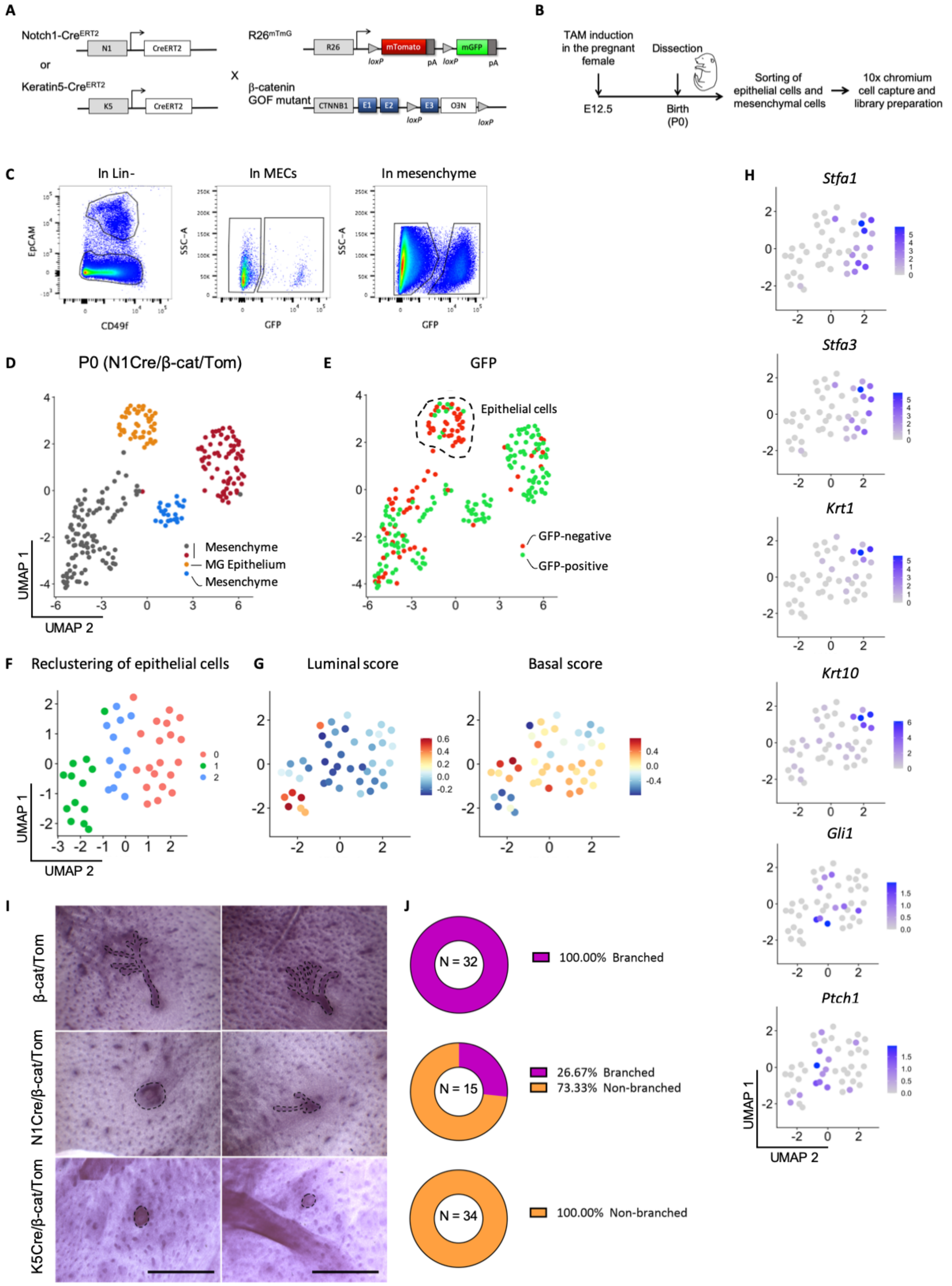


Figure 17 | β-catenin activation in embryonic epithelial cells impairs branching morphogenesis *in vivo*.

(A) Schematic representation of the N1Cre/Tom and N1Cre/β-cat/Tom mouse models used. All cells are labelled with a red membrane tomato fluorescence. TAM administration induces membrane GFP expression and β-catenin accumulation in the N1Cre/β-cat/Tom model. (B) Isolation and sequencing strategy of mammary embryonic cells at P0 in the N1Cre/β-cat/Tom model. Pregnant dams were administered TAM at day 12.5 of pregnancy. 7 days later (at birth), embryonic mammary buds were dissected together with the surrounding mesenchyme and processed for 10X scRNA-seq. (C) Representative FACS dot plots of the gating strategy used to sort GFP$^+$ epithelial and mesenchymal cells at P0 in N1Cre/β-cat/Tom mice 7 days after Cre induction. (D and E) UMAP plots of N1Cre/β-cat/Tom MECs and the surrounding mesenchymal cells isolated by scRNA-seq at P0. Cells are colour-coded by cluster (D) or by GFP expression (E). GFP$^+$ cells are illustrated in green and GFP$^-$ cells in red. (F) UMAP plots of epithelial cells isolated at P0 after re-clustering analysis. Cells are colour-coded by cluster. (G) UMAP plots from (F) colour-coded according to the expression of the single-cell ID scores in MECs: adult luminal score (left) or adult basal score (right). (H) UMAP plots from (F) colour-coded to show the expression of cluster-specific genes: for skin cells (*Stfa1* and *Stfa3*), suprabasal layer of the skin (*Krt1* and *Krt10*) and hair follicles (*Gli1* and *Ptch1*). (I) Representative images of whole-mount carmine red staining in β-cat/Tom control, N1Cre/β-cat/Tom and K5Cre/β-cat/Tom mice. Scale bar: 1 mm. (J) Percentage of branched (in magenta) and non-branched (in orange) mammary gland trees in β-cat/Tom control (n = 32), N1Cre/β-cat/Tom model (n = 15) and K5Cre/β-cat/Tom model (n = 34).

mesenchymal populations (Figure 17D). To distinguish the transcriptomes of β-catenin-mutant cells, we first identified them using their GFP expression (Figure 17E). Next, to determine the composition of the mammary epithelial cluster, we performed a re-clustering and applied the basal and luminal scores that we had previously generated to analyse the WT dataset (Figure 9C, 17F- G).

Cluster 1 (C1 in green) is composed of MECs, both LCs and BCs, as shown by the basal and luminal scores (Figure 17F- G). Regarding cluster 0 (C0 in pink), we believe that it represents a skin population (*Stfa1*$^+$, *Stfa3*$^+$) (Figure 17H). The basal layer of the skin is characterized by a higher basal score (Figure 17F- H), since many genes are shared between this layer of the skin and BCs of the MG. By contrast, the skin population with a lower basal score represents cells from the suprabasal layer of the skin, which also expresses *Krt1* and *Krt10* (Figure 17G- H). Given that in the WT dataset we did not find contamination of skin cells and that we isolated only the mammary fad pad (excluding the embryonic skin), we did not expect to find skin cells in our dataset. The analysed skin cells could be contaminants derived from the nipple area, but another possibility could be that Wnt activation in the embryonic mammary buds induces epidermal transdifferentiation, as we have seen happening in the pubertal and adult MG (Lloyd-Lewis et al., 2022) (see *Discussion* section, *page 102*).

Finally, cluster 2 (C2 in blue) is composed of hair follicle cells expressing several SHH signalling components, such as *Gli1, Ptch1* (Figure 17H). SHH is required to promote initial follicle growth during hair follicles development (Sennett & Rendl, 2012).

To analyse the *in vivo* effect of β-catenin stabilization on embryonic mammary development, we induced mutant expression at E12.5 and performed whole-mount carmine red staining to analyse the rudimentary mammary tree in female mice present at birth. In comparison to control littermates, β-catenin stabilization severely or entirely blocked branching in mammary rudiments, and reduced growth of the few branches that were able to emerge, likely as a result of mosaic mutant β-catenin expression (Figure 17I- J). Moreover, a similar phenotype was observed in all mammary glands, suggesting that the effect is not specific for each pair. Given that Notch1 is also expressed in cells residing in the mammary mesenchyme (Lilja et al., 2018), to assess whether the observed phenotype in the N1Cre/β-cat/Tom model was epithelial-specific or caused by stromal mutant cells, we performed the same experiments in K5Cre/β-cat/Tom mice where mutant β-catenin was exclusively targeted to the embryonic MECs. Since mammary branching was also severely impaired in the K5Cre/β-cat/Tom model (Figure 17I- J), we concluded that β-catenin stabilization in MECs precludes branching morphogenesis.

β-catenin activation in embryonic epithelial cells impairs branching morphogenesis *ex vivo*

To further characterize the failure in branching morphogenesis induced by mutant β-catenin stabilization, we established mammary buds from E13.5 N1Cre/β-cat/Tom embryos in explant cultures 24 hr after mutant β-catenin induction. Mutant β-catenin stabilization severely impaired branching initiation in *ex vivo* cultures (Figure 18A- B) confirming our *in vivo* observations. After 4 days in culture, all β-cat/Tom control mammary buds presented the first sprouting tip, whereas N1Cre/β-cat/Tom mammary buds failed to initiate branching. Consequently, the planar surface area of the few mutant explants that managed to form some branches after 6 days in culture was significantly decreased compared to β-cat/Tom control buds (Figure 18A, 18C).

Collectively, these findings demonstrate that aberrant Wnt/β-catenin signal activation in the embryonic mammary epithelium precludes normal branching. However, the molecular mechanisms underpinning this phenotype remain to be elucidated.

Analysis of the dynamic behaviour of β-catenin activated cells

Our ability to visualize the cellular dynamics of individual mammary cells by live imaging provided a tool to study the impact of sustained Wnt/β-catenin signalling on cellular movements and behaviour during embryonic morphogenesis. As described in the previous *Chapter*, a high number of mutant β-catenin-expressing embryonic MECs (obtained using TAM at the concentration of 0.1 mg per g of mouse body weight) blocked branching morphogenesis (Figure 17I- J). To explore the behaviour of mutant cells at clonal density, I have induced β-catenin stabilization in fewer mammary cells by reducing the TAM dose by 10-fold (0.01 mg per g of mouse body weight). Mammary buds from embryos exposed to this low TAM dose were then dissected, grown as *ex vivo* cultures and imaged

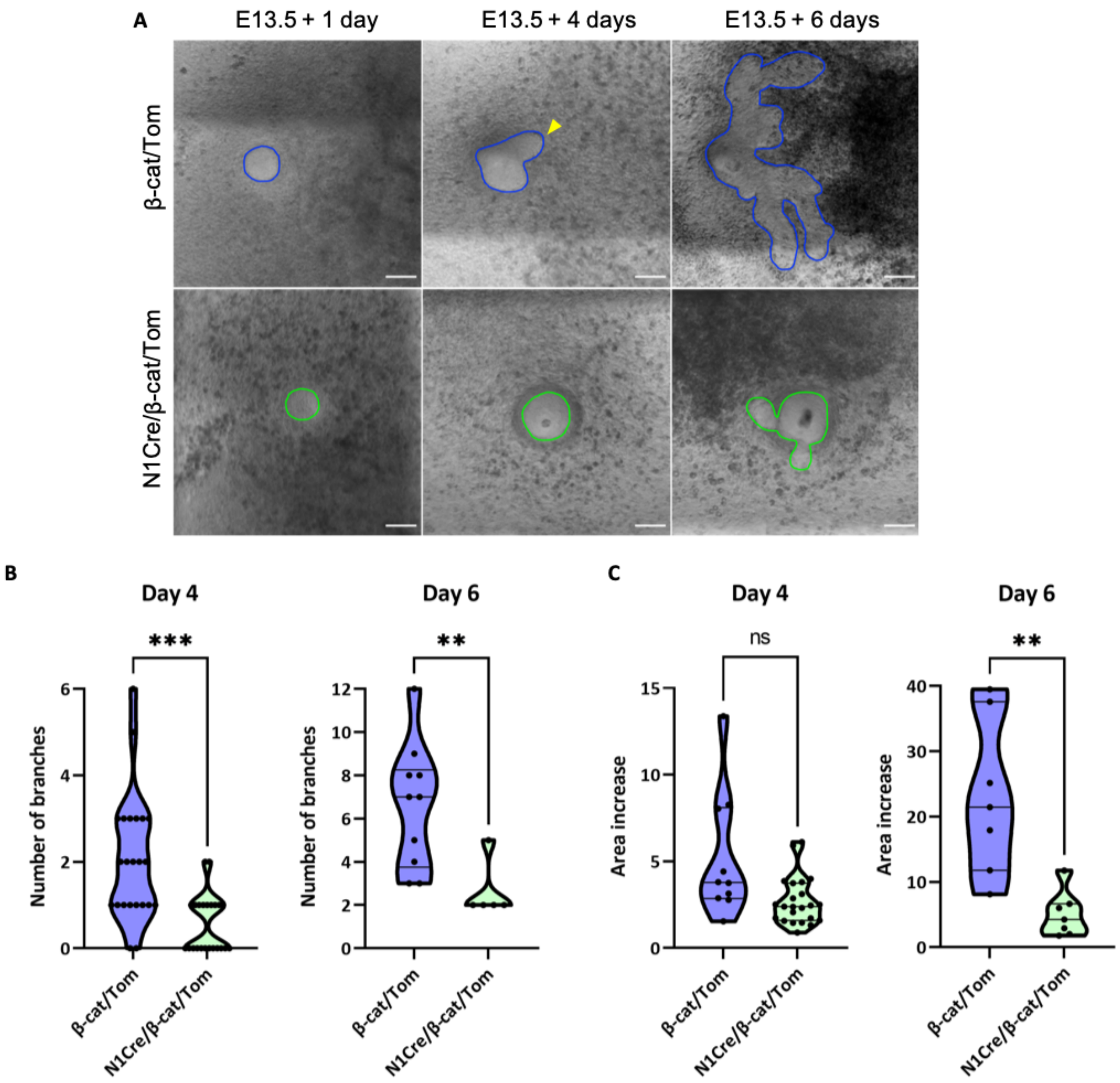


Figure 18 | β-catenin activation prevents branching morphogenesis initiation in cultured embryonic explants.

(A) Representative images of E13.5 mammary buds from control (β-cat/Tom) and N1Cre/β-cat/Tom mice at the indicated times in culture. Scale bar: 100 µm. The rendered surface of the mammary epithelium is outlined in blue (in β-cat/Tom control) or in green (in N1Cre/β-cat/Tom explants). The yellow arrowhead indicates the first sprouting event. (B) Quantification of the number of branches in control (β-cat/Tom) and N1Cre/β-cat/Tom explants at day 4 (left) and day 6 (right). (C) Quantification of the area increase in control and N1Cre/β-cat/Tom explants at day 4 (left) and day 6 (right) compared to the area measured at day 1 (fold-change). Statistical significance was assessed with two-tailed unpaired T-test with Welch's correction. ns: non-significant, ** indicates $p<0.01$ and *** indicates $p<0.001$.

daily for up to 2 weeks. Under these conditions, we observed that overall mammary branching morphogenesis in N1Cre/β-cat/Tom appeared to proceed normally when compared to N1Cre/Tom *ex vivo* cultures (Figure 13C, 19A), allowing us to analyse the behaviour of sporadic mutant cells in the context of an otherwise primarily WT tissue.

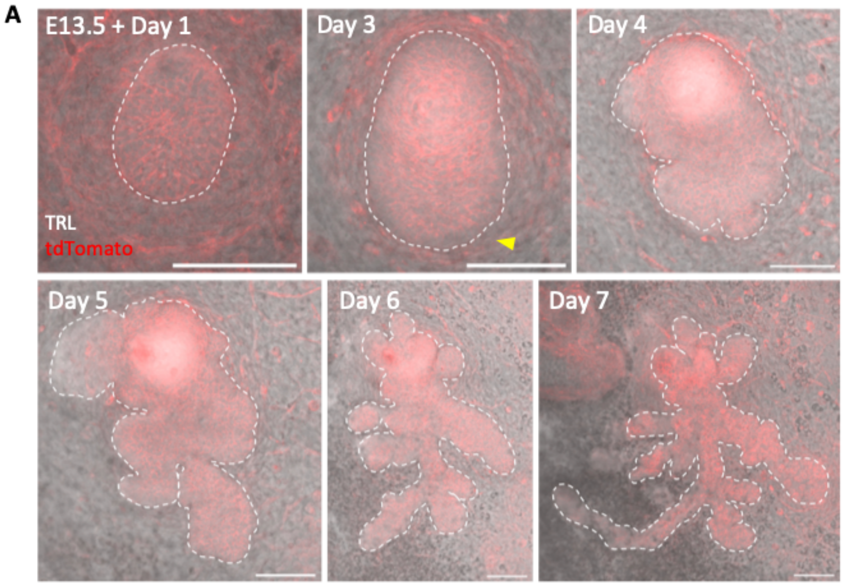


Figure 19 | Mammary branching morphogenesis in N1Cre/β-cat/Tom *ex vivo* cultures is not impaired when reducing the TAM dose.

(A) Representative images showing the growth of a N1Cre/β-cat/Tom mammary bud dissected at E13.5 and imaged over 7 days. Pregnant females were induced with a low-dose TAM (diluted 1:10). The yellow arrowhead indicates the first sprouting event. Dotted lines delineate the BM. Scale bar: 100 µm.

I thus addressed whether β-catenin stabilization could affect track length, track velocity and net displacement of individual mutant mammary cells (Figure 20A- B). I found that β-catenin stabilization does not significantly affect cell movement, since cells located in the tip region were oriented towards the direction of the elongating branch as observed in N1Cre/Tom control buds (Figure 14F, 20C). Also consistent with WT cells, non-oriented displacements were observed in the body region, where mammary cells continually move back and forth (Figure 14G, 20D). This suggests that, similarly to WT cells (Figure 14F), mutant β-catenin cells residing in the tips of mammary branches are still able to contribute to ductal elongation and fat pad invasion (Figure 20B- C). Additionally, cell displacement was significantly higher in the tips compared to the cells located in the body region of the mammary bud (0.046 µm/min and 0.019 µm/min, respectively) (Figure 20E). This could be correlated to the observed increased proliferation of tip cells compared to the ones located in the subtending duct (Figure 12D). In line with the cellular behaviours observed in N1Cre/Tom buds, N1Cre/β-cat/Tom cells also exhibited a dynamic cell displacement from the inner to the outer compartment, when we quantified the distance of each tracked cell to the BM (Figure 16C, 20F).

We then focused on cellular dynamics after 5 days in culture and compared cell movements in control and β-catenin mutant cells. Quantitative analysis showed that the mean track velocity was similar in control and mutant cells (0.183 µm/min and 0.198 µm/min, respectively) (Figure 20G). Also, the track length of mutant cells did not significantly differ from that of control cells (175.4 µm and 155.3 µm, respectively) (Figure 20H).

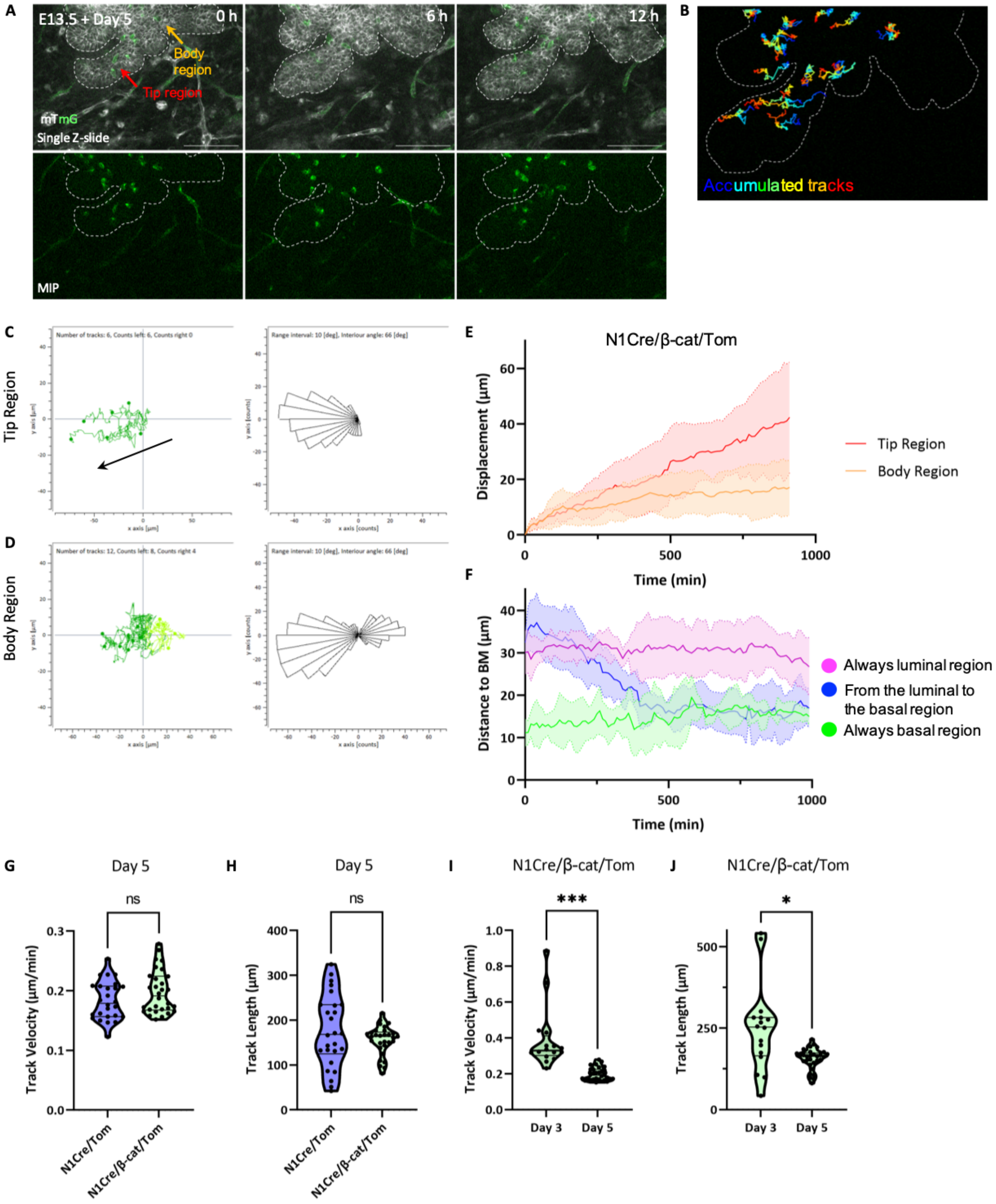


Figure 20 | Individual β-catenin mutant cells present a normal cell behaviour in *ex vivo* cultures at low induction levels.

(A) Time-lapse visualization of a N1Cre/β-cat/Tom explant dissected at E13.5 and cultured for 5 days. Pregnant females were induced with a low-dose TAM. Single optical section showing the expression of non-recombined tdTomato^{+} cells (in white) and GFP^{+} cells (in green). Tip and body regions are indicated. Maximum intensity Z-projection (MIP) of GFP^+ cells (in the bottom). Dotted lines delineate the BM. Scale bar: 100 µm. T= 0 hr refers to 5 days in culture. (B) Trackmate figure illustrating all cellular tracks from time-lapse in (A) colour-coded according to time. Cold colours represent the beginning of the time-lapse, whereas warm colours are closer to the end of the time-lapse. Dotted lines delineate the BM. (C and D) Plots representing tracks of individual cells in a representative time-lapse of a N1Cre/β-cat/Tom explant induced with low-dose TAM (left) and polar plots representing the distribution of trajectory angles of the same cells in the tip (C) and body region (D). T= 0 hr refers to 5 days in culture. Arrow indicates the branch growth direction. Angles show a preferentially orientated direction in cells in the branching tips (C) and random distribution for cells located in the body region (D). The lack of tracks in all directions in D is due to the presence of the tissue border, preventing 360 degrees movements. (E) Net displacement of cells located in the tip (in red) and body region (in orange) in a representative time-lapse of a N1Cre/β-cat/Tom mouse. Net displacement was larger in cells located at the tips (0.046 µm/min and 0.019 µm/min). (F) Plot illustrating the three modes of migratory behaviour observed in a N1Cre/β-cat/Tom explants: cells always in the luminal compartment (in magenta) (n= 8), cells that move from the luminal to the basal compartment (in blue) (n= 4) and cells always in the basal compartment (in green) (n= 4). (G and H) Quantification of cell movement parameters, track velocity (G) and track length (H), in N1Cre/Tom and N1Cre/β-cat/Tom explants (induced with low-dose TAM) cultured for 5 days. (I and J) Quantification of cell movement parameters, track velocity (I) and track length (J), in N1Cre/β-cat/Tom model after 3 or 5 days in culture. Statistical significance was assessed with two-tailed unpaired T-test with Welch's correction. ns: non-significant, * indicates $p<0.05$, and *** indicates $p<0.001$.

Then, we also compared mutant cell dynamics after 3 and 5 days in culture. During the first sprouting event (corresponding to day 3), the migration speed is highly variable but overall higher than at day 5, as indicated by their track velocity (0.388 µm/min at day 3 and 0.198 µm/min at day 5) (Figure 20I). This suggests that cell motility differs during the time-course of branching elongation. Since mutant cells at day 3 moved faster, they also travelled longer distances in comparison with mutant cells at day 5 (248.5 µm and 155.3 µm, respectively) (Figure 20J). This suggests that active cell migration may be an important mechanism that contributes to ductal elongation during the initial stages of branching morphogenesis.

In summary, low level induction of mutant Wnt/β-catenin signalling activity did not affect mammary cell movements or behaviour (velocity, length, or displacement) in comparison to control conditions. It is noteworthy, however, that these are still preliminary results and that further experiments at different developmental times, as well as in conditions with increased numbers of mutant β-catenin cells, are required to better understand how aberrant Wnt activation impairs branching morphogenesis, and whether lineage segregation is also impaired after Wnt stabilization, as we have observed during pubertal branching morphogenesis (Lloyd-Lewis et al., 2022).

Discussion

In this section, I will discuss the new embryonic mammary cellular hierarchy proposed in light of my results, the requirement for 3D *ex vivo* cultures of embryonic mammary buds surrounded by a specialised mammary mesenchyme to study branching morphogenesis and cell fate acquisition, as well as the role of Wnt/β-catenin signalling in embryonic mammary development.

1. Lineage specification from MaSCs to unipotent progenitors is a progressive process during embryonic development

In this work, we have delineated the progressive lineage specification of mammary epithelial cells during embryonic mammary gland development. We have shown, for the first time, that embryonic mammary epithelial cells at E15.5 can already be distinguished as three transcriptionally different populations: "basal-like" cells, "luminal-like" cells and a mixed cell state between these two cell types, which we called hybrid cells.

Using our scRNA-sequencing data, differentiation trajectories towards a luminal or basal fate were computationally reconstructed. Our analysis reveals the changing transcriptional landscapes during the progression towards cell fate commitment from E13.5 to P0 developmental stages, reaching lineage segregation at birth with epithelial cells with luminal or basal characteristics found at the opposite ends of the differentiation trajectory. Based on these results, we propose a dynamic hierarchical model of mammary epithelial cells fate specification during embryonic development (Figure 21A- B).

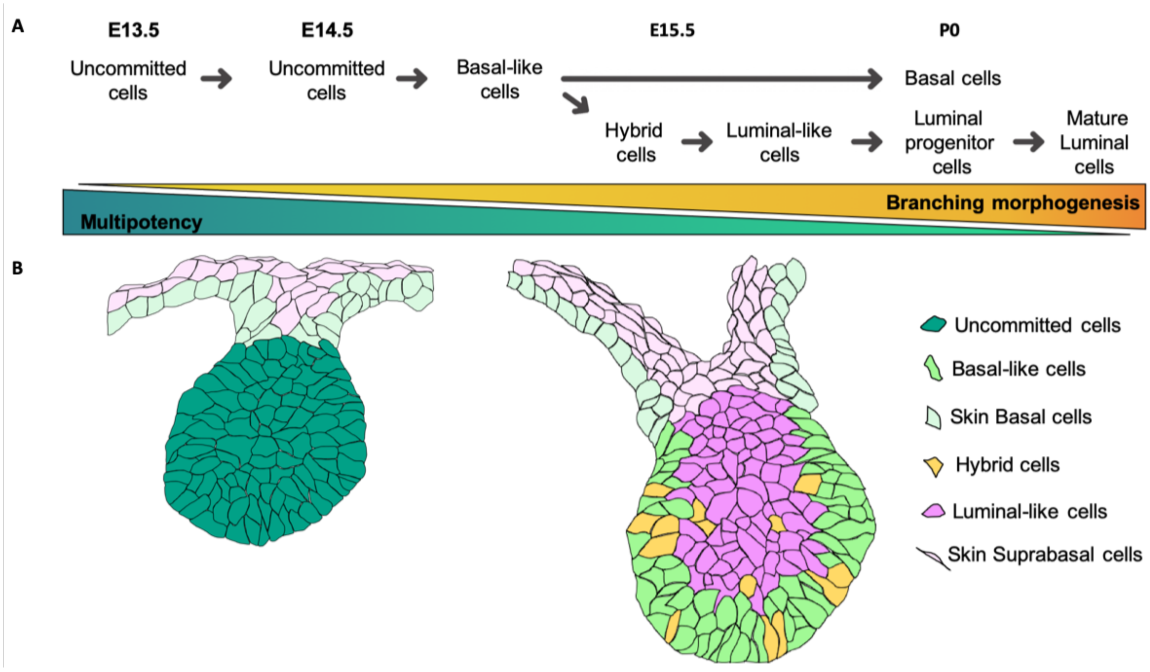


Figure 21 | Proposed model for lineage segregation of embryonic mammary epithelial cells during development.

(A) Proposed model of luminal and basal differentiation trajectories from E13.5 to P0. (B) Cartoon depicting the spatial localisation of the different cell types distinguishable in the embryonic mammary bud at E13.5 and E15.5.

Mammary epithelial cells at E13.5 are undifferentiated and have yet to engage in the process of lineage specification. As development and tissue morphogenesis progress, these supposedly multipotent embryonic MaSCs at E15.5 will give rise to what we called basal-like cells, based on their expression of several genes that define basal mammary cells postnatally. These basal-like cells will then either differentiate into basal unipotent progenitors by P0, giving rise to unipotent basal cells postnatally, or they will transition towards a transcriptionally hybrid state. Hybrid cells, whose lineage potential remain unclear at this stage, will gradually lose basal markers and acquire luminal gene expression, giving rise to unipotent luminal cells at birth (Figure 21A- B).

Previous studies concluded that basal-like embryonic mammary cells become lineage segregated in the early postnatal period (prior to 2 weeks) (Pal et al., 2021). In this study, they used ternary plots to determine the lineage identity of each cluster. In such graphic representations, basal, LP and ML epithelial subtypes are positioned each in a different corner. The closer one cluster is to a corner, the higher transcriptional similarity it exhibits to the specific cell type associated with that corner. No hybrid-linage signatures were found in this study at E18.5 (Figure 22A- B). The authors thus concluded that mammary epithelial cells at E18.5 are molecularly closer to the basal lineage and only in post-natal glands at day 5 a diffuse luminal population could be distinguished. Only later on in development, at the pre-puberty stage (2 weeks), a basal, LP and ML could be observed, although the LP and ML clusters only partially overlapped with the adult lineages at 10 weeks of age (Figure 22A- B).

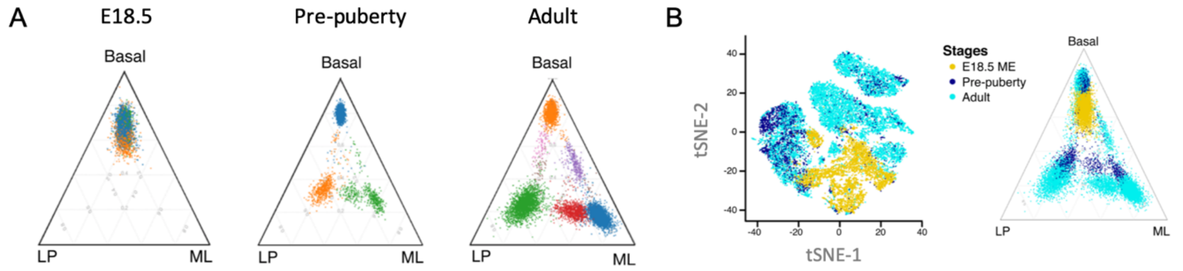


Figure 22 | Transcriptomic heterogeneity within the developing mammary epithelium previously proposed by Pal et al.

(A) Ternary plot of MECs from E18.5 embryos, pre-pubertal (2 weeks) and adult mice (10 weeks). Cells are positioned according to the proportion of basal, LP, or ML signature genes expressed in each cell. Cells are coloured according to cell population. (B) t-SNE and ternary plot of all cells after integration of all developmental times previously analysed. Cells are coloured according to developmental stage. Adapted from (Pal et al., 2021).

Interestingly, when we applied the analysis method employed by Pal et al. to our E15.5 dataset, we could observe that all epithelial cells at E15.5 also exhibited a higher homology with the basal cells (Figure 23A- B). This contrasts with our results when applying the basal and luminal single-cell ID score computed in *Results Chapter 1*, which showed that E15.5 MECs can clearly be resolved into 3 distinct groups: luminal-like cells, basal-like cells and a hybrid cell population co-expressing both luminal and basal genes (see Figure 9B from *Results Chapter 1*). These differences in the interpretation of our results implicate that ternary plots cannot be used to distinguish distinct embryonic MEC clusters and to draw final conclusions about lineage segregation.

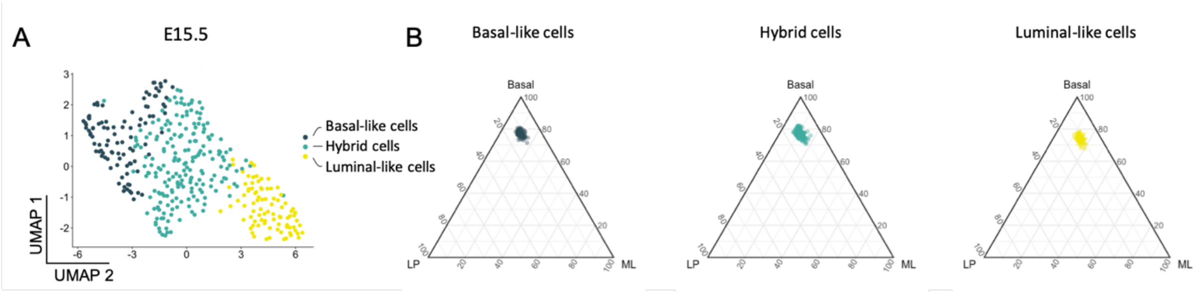


Figure 23 | Ternary plot analysis cannot be used to study MECs transcriptional heterogeneity during embryonic development.

(A) In our E15.5 scRNA-seq dataset, MECs can be resolved into 3 distinct groups: luminal-like cells, basal-like cells and a hybrid cell population, as shown in Figure 9B from *Results Chapter 1*. (B) Ternary plot analysis of cells from E15.5 mice. Cells are positioned according to the proportion of basal, LP, or ML signature genes expressed in each cell (derived from (Pal et al., 2021)). Cells are coloured according to cell population. Using the analysis method employed by Pal et al. all MECs at E15.5 exhibit higher similarity with basal adult cells.

On the other hand, two alternative scRNA-seq studies proposed that bipotent mammary stem cells, sharing luminal and basal characteristics exist throughout embryogenesis and that these cells will only generate basal, LP and ML cell types postnatally (Giraddi et al., 2018; Wuidart et al., 2018). Technical limitations, such as low numbers of reads per cell or fewer sequenced cells, may have contributed to the inability of these studies to identify cells with unique basal or luminal characteristics during embryonic development, contrary to our analysis. To better understand the potential of a cell population to differentiate into the specific mammary cell types, a more recent study performed snATAC-seq in the MG at E18.5 and adult stages (Chung et al., 2019). This analysis revealed that MG cells at E18.5, although still possessing fetal-specific features, start acquiring basal-, LP-, and ML-like characteristics, which represent 32%, 62%, and 4% of the total fetal population, respectively (Chung et al., 2019).

With regards to luminal cell specification, our computational analysis of the embryonic differentiation trajectories showed that luminal-like cells at E15.5 preferentially transition to LPs by P0, and that LPs will eventually give rise to the ML population, comprising cells expressing the hormone receptors ERα and PR. The finding of a common luminal progenitor is in agreement with Giraddi et al. (Giraddi et al., 2018). However, we have not yet elucidated how the cell fate choice from LP to ML is regulated at the transcriptional level. Further studies are required to better resolve the role of luminal progenitors in the differentiation of the two types of luminal cells, ER$^+$ and ER$^-$, during MG postnatal development. Afterwards, in adult homeostasis, several lineage tracing studies have clearly shown that these two luminal cell lineages are maintained by their own pool of unipotent progenitors (Chang et al., 2014; Rodilla et al., 2015; van Keymeulen et al., 2017; C. Wang et al., 2017).

Finally, our dataset provides a useful resource to identify specific genes that might be causally implicated in the differentiation process of multipotent MaSCs to basal and luminal cells during embryonic development. Nevertheless, further studies are necessary to functionally test the role of specific transcription factors in mammary cell fate specification.

2. Spatial segregation of basal and luminal-like cells during embryonic MG development

We then asked whether the three populations identified by transcriptomic analysis are localized in specific compartments. Importantly, early segregation of mammary cell types cannot be observed in the embryonic epithelium using the established differentiation markers that distinguish LCs and BCs in the adult MG. Indeed, to date, E17.5 is the earliest timepoint where segregation of the expression of basal and luminal markers has been observed, with the outermost layer of the embryonic MG exhibiting higher P63 levels compared to internal cells (Wuidart et al., 2018).

Fluorescence *in situ* hybridization using probes against several genes that we have found to display a differential expression pattern towards the basal or luminal lineage showed that basal-like and luminal-like cells indeed reside in spatially-distinct layers within the E15.5 mammary bud. Quantification of the number of RNA molecules as a function of distance from the BM showed that *Anxa1* and *Cxcl14* markers are present in opposing epithelial regions. While the *Anxa1* (luminal) probe predominantly labelled cells in the region furthest away from the BM, the *Cxcl14* (basal) probe preferentially labelled cells in the region adjacent to the BM. We observed similar patterns with the alternative markers *Ndnf* and *Pthlh* for basal-like cells, and *Plet1*, *Ly6d* and *Lgals3* for luminal-like cells. Interestingly, some of these markers are also expressed in the skin, as seen by both RNAscope (see Figure 10D from *Results Chapter 1*) and scRNA-seq (data not shown). *Cxcl14* is expressed specifically in the basal layer of the skin, whereas *Anxa1*, *Ly6d* and *Lgals3* are expressed in the suprabasal layer. Of note, *Ndnf, Pthlh* and *Plet1* are only expressed in the mammary epithelium (and not in the neck region connecting the mammary bud with the skin epidermis).

During embryogenesis, the surface ectoderm stratifies to become the epidermis. A local epithelial thickening, known as a placode, is the first morphological indicator of a forming ectodermal appendage (Biggs & Mikkola, 2014). At E13.5, the mature bud comprises a sphere of concentrically oriented epithelial cells connected to the skin surface by a stalk (Spina & Cowin, 2021). A migration-driven cell influx invagination process was recently described as mediating early mammary gland development from E12.5 to E13.5, driving the formation of the mammary bud and the neck region (Trela et al., 2021). This leads to the question whether the suprabasal layer of the skin participates in the MG development.

Further experiments to solve this question are proposed in the *Future perspectives* section below (*page 108*).

At early developmental times, basal genes are in general expressed at much more high levels in comparison to luminal genes. Conversely, at E15.5, luminal genes begin to exhibit higher expression levels. This finding correlates with the appearance of a luminal-like cell cluster revealed by scRNA-seq at E15.5. Interestingly, our smRNA-FISH data suggest that lumen formation does not coincide with the emergence of a luminal committed cell type, nor is the presence of a lumen necessary for luminal cell specification.

The developmental transcriptomic atlas that we have compiled by single cell analysis thus enables the spatial mapping of potentially multipotent and unipotent mammary progenitors. This dataset also allows the selection of lineage-specific genes that could be functionally important for dictating cell fate choices in this tissue. However, still images and snapshots of fixed samples taken at defined timepoints do not allow us to conclude whether a given **spatial position in the embryonic bud governs cell fate, or whether cells are first specified to a particular fate before moving to that cellular compartment.** In other words, do basal-like embryonic cells start to downregulate basal genes and acquire luminal features and this is what forces them to move towards the internal bud region? Consistent with this hypothesis, I found that the vast majority of cells expressing *Cd74*, a marker for the E15.5 hybrid cells, were in either the middle or the outer region of the mammary bud at E15.5, coinciding with expression of the basal-like *Cxcl14* marker gene. Nevertheless, additional smRNA-FISH experiments using probes for genes specifically expressed in the hybrid population at E15.5 are necessary to provide an assertive answer to this question.

To further study the "what comes first: the chicken or the egg, i.e. cell differentiation or position?" question, we established an *ex vivo* culture system to investigate whether mammary epithelial cells transition between luminal and basal compartments during embryonic development before acquiring their final cell fate. To enable fate-mapping and facilitate single cell tracking, we used Notch1-Cre-driven expression of the R26-mTmG fluorescent reporter to label sporadic Notch1-expressing cells that are scattered throughout the embryonic bud at E13.5, but will eventually become restricted to LP post-natally (Lilja et al., 2018).

3. Live-imaging to study branching morphogenesis during embryonic development and multipotency restriction

Several *ex vivo* culture methods have been developed to study adult mammary branching morphogenesis, including mammary epithelial organoids cultured in a 3D matrix (Huebner & Ewald, 2014). These models, however, are lacking the physiological stromal compartment, known to be essential for normal development *in vivo*. *Ex vivo* cultures of embryonic MG represent a highly tractable experimental system for disentangling the molecular mechanisms underpinning the initial stages of tissue development and the onset of branching events (Voutilainen et al., 2013). Yet, high-resolution analysis of cellular behaviours during the first morphogenetic events in the embryonic MG have, to date, been lacking. For this reason, we have put a substantial effort to optimize branching efficiency of *ex vivo* cultures as well as live-imaging of embryonic mammary buds, to address how cell fate specification is coordinated with branching morphogenesis. This protocol has been published in *Methods in Molecular Biology* (Carabaña & Lloyd-Lewis, 2022) (*Annex 1*).

The live-imaging experiments performed in this work revealed a surprising high level of cellular motility within the embryonic MG during branching morphogenesis and ductal elongation. Interestingly, we found that embryonic MECs lose the ability to change compartments when development proceeds. This might indicate that upon committing to a particular lineage, MECs are no longer able to move between inner and outer regions of the embryonic bud. Our data also revealed that a higher frequency of cells move from the inner to the outer compartment, whereas cells rarely move in the opposite direction.

Based on our analysis and their transcriptional similarity, we can assume that the basal-like cluster identified at E15.5 is composed of both committed basal cells, already located in the basal compartment, and of hybrid cells, which would be the *Cxcl14*-expressing cells that we found positioned in the internal bud region. Given that we found no cells expressing *Anxa1* in the outer bud compartment and that *Anxa1*-expressing cells present a higher expression of K8 at the protein level (see Figure 16D from *Results Chapter 2*), we believe that these cells represent luminal-like cells, which do not change compartment during mammary morphogenesis. Testing this hypothesis will require the generation of reporter mice for these selected genes, which will allow the tracking in real time of cells that differentiate towards the luminal or basal fate (described in detail in the *Future perspectives* section - *page 109*).

Further studies will be necessary to fully understand the mechanisms involved in lineage segregation during embryonic MG morphogenesis. What trigger the sprouting of new branches? Is branching necessary for cell differentiation? Which aspects of branching and the consequent cell rearrangements instruct cell fate? Does changing the form of the tissue affect patterns of cell fate acquisition?

Insight into mammary branch formation

Branching morphogenesis occurs through several different physical mechanisms that are not mutually exclusive. The exact mechanisms that drive morphogenesis in the embryonic MG remain to be elucidated (Varner & Nelson, 2014).

In this study, in agreement with the literature, we have shown the important role of epithelium-mesenchyme communication in embryonic mammary branching morphogenesis (Figure 24A). However, our findings do not exclude the involvement of other mechanisms in mammary branching, such as oriented cell division (Figure 24B) or differential proliferation in branch initiation (Figure 24C), which should be studied in future work (as further proposed in the *Future perspectives* section - *page 109*). Interestingly, when adding FGF10 to the culture medium, we found that mammary branches grew faster, but cell proliferation was not increased (see Figure 12D- E from *Results Chapter 1*). This suggests that MECs can move faster along the extending ducts when receiving signals from the mesenchyme. In addition, we observed by time-lapse imaging that MECs move faster when the first sprouting event occurs (at day 3 in culture) than at day 5 during duct elongation (see Figure 20I from *Results Chapter 3*), suggesting that active cell migration (Figure 24D) might be responsible for the directional elongation of branches during initial stages of branching morphogenesis. However, this observation was made only in N1Cre/β-cat/Tom *ex vivo* cultures, so additional experiments should be performed in WT conditions to ensure that this phenotype is not linked to the presence of mutant cells. In line with this hypothesis, it has been shown that both hair follicles (Ahtiainen et al., 2014) and teeth (Ahtiainen et al., 2016) are formed through directional cell migration.

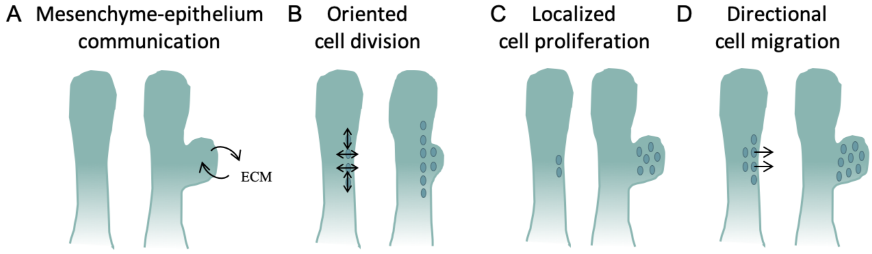


Figure 24 | Different proposed mechanisms of branching morphogenesis.
(A) Mesenchyme-epithelium communication. (B) Oriented cell divisions. (C) Localized increase in cell proliferation within the parental branch. (D) Directional cell migration driving ductal elongation or the formation of a new epithelial branch via an invasive migratory process.

In addition, we found that cell displacement was significantly higher in the tips compared to the cells located in the body region of the mammary bud (see Figure 14H from *Results Chapter 2*). This might be linked to the fact that cell proliferation is higher in the tip regions, as assessed by EdU incorporation experiments. In the embryonic salivary gland (S. Wang et al., 2021) and kidney (Packard et al., 2013), it has been shown that surface cells delaminate to a subsurface level to divide into two daughter cells that will then return back to the surface. This cellular behaviour plays a role in epithelial cell rearrangements that contribute to branching morphogenesis. The time-lapse imaging acquired in this study will be exploited to analyse the localisation of cell division events in embryonic mammary buds. Moreover, by time-lapse imaging of kidney *ex vivo* cultures (Packard et al., 2013), it has been observed that when a cell translocates to the inner compartment, it retains contact with the basal surface via a thin protrusion. These cellular protrusions that contact the BM have been observed in both the embryonic (data not shown) and adult MG (Lafkas et al., 2013; Lloyd-Lewis et al., 2022), however it is unclear whether these protrusions only form during cell division.

Overall, several of the above-mentioned cellular mechanisms may act synergistically to guide embryonic mammary branching. Based on our preliminary results, directional cell migration, but not proliferation, might be an important mechanism that contributes to the formation of new branches and ductal elongation during the initial stages of branching morphogenesis in the embryonic MG.

Image analysis challenges

I have worked together with an Image Analyst, Dr. Varun Kapoor, to develop a semi-automated image analysis pipeline for the analysis of the live-imaging datasets generated in this study. This pipeline has been published in *Proceedings of the Python in Science Conferences* (SciPy) (Kapoor & Carabaña, 2021) (*Annex 2*). Image analysis is performed in two main steps: first, the fluorescently-labelled cells being traced are detected as individual objects at each timepoint. Then, individual objects are linked to their previous positions in earlier timepoints to obtain their migration tracks over time.

Several state-of-the-art segmentation algorithms have been recently developed to automatically detect and track cells. However, these algorithms are optimized to detect round-shaped objects (i.e. cell nuclei). Therefore, the mTmG model that I have used posed a challenge for image analysis, as we needed to track cells with fluorescently-labelled membranes that are highly irregular in shape and size and have a low SNR. In addition, the frequent close contact between different GFP-labelled cells, inherent to the difficulty of labelling cells at clonal density *in vivo*, hindered the cell segmentation step.

Different methods have been developed for the segmentation task. *Stardist* directly predicts a shape representation as star-convex polygons for cell nuclei in 2D and 3D (Schmidt et al., 2018; Weigert et al., 2020). However, cell membrane segmentation is especially challenging as opposed to nuclei segmentation due to fewer boundary pixels and the need to separate touching cells. In contrast, U-Net classifies pixels as background or pixels as belonging to the cell (Ronneberger et al., 2015). Here, we developed a segmentation algorithm called *Vollseg*, in which we take advantage of the strengths of both *Stardist* and U-Net methods. Briefly, we use *Stardist* in 3D to obtain the cell centroids to distinguish two touching cells, combined with a 3D U-Net model to obtain a better approximation of the irregularly shaped cells. After cell segmentation was performed, the generated masks were imported into TrackMate, a Fiji plugin designed for tracking analysis (Ershov et al., 2022). TrackMate allowed us to obtain migration tracks for each cell, as well as their distances from the basement membrane.

A nuclear marker could be used to facilitate the tracking of epithelial cells, nevertheless cell membrane labelling enables us to obtain information related to the distance to the BM. For instance, the above-mentioned cellular protrusions that are in contact with the BM could not have been observed without a membrane signal.

In conclusion, high resolution imaging of mammary embryonic explant cultures, combined with our newly developed cell segmentation and tracking methods, has provided us with foundational insights into embryonic mammary cell proliferation, migration, rearrangements and differentiation. Further studies are needed however to build a detailed map of the dynamic spatiotemporal behaviour of mammary epithelial cells in developing embryonic mammary glands.

4. Embryonic mammary mesenchymal cells surrounding the epithelial bud represent two spatially distinct stromal cell populations

Previous studies using scRNA-seq at different stages of embryonic mammary development have not analysed mesenchymal cells. To our knowledge, our results present the first scRNA-seq dataset of embryonic mammary mesenchymal cells surrounding the developing buds. Interestingly, we identified transcriptional signatures that distinguish two spatially restricted mesenchymal populations in mammary embryonic glands at E13.5 and E15.5.

We hypothesized that genes shared between E13.5 and E15.5 stages may represent those with important roles in promoting mammary gland differentiation (Figure 25A- B). Comparison of the sub-epithelial mesenchyme at both developmental timepoints revealed several overlapping Wnt activators and targets of Wnt signalling, including *Axin2, Lef1, Nkd1, Tcf4* and *Tcf7*. Indeed, Wnt related genes expression has also been observed in the embryonic pulmonary mesenchyme, where Wnt signalling regulates the spatially-distinct compartments of the lung mesenchyme (Goodwin et al., 2022). Also, the Androgen receptor (Ar) is expressed by the sub-epithelial mesenchyme at both developmental times, since it is essential to cause degeneration of the buds in male embryos (Spina & Cowin, 2021).

When comparing mesenchymal cell populations residing in the dermal mesenchyme at these two developmental times, we found 18 common genes, such as *Gpc3*, involved in promoting mammary cell fate, as well as *Igf2* or *Ptn*. In addition, we observed that genes involved in invasive cell behaviour (*Cxcl12*) or in axon guidance (*Nrp2, Sema3a, Epha3, Epha5*) are highly expressed in the dermal mesenchyme. This suggests that the mesenchymal cells more distant from the epithelium might play important roles in sending the needed signals to promote branching morphogenesis.

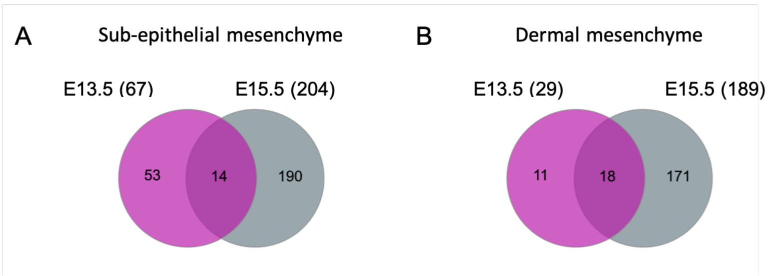


Figure 25 | Transcriptome analysis reveals similarities and differences within mammary mesenchyme at two different developmental times.
(A) Venn diagram showing the number of shared and unique genes expressed in the sub-epithelial mammary mesenchyme between E13.5 and E15.5 and (B) dermal mesenchyme between E13.5 and E15.5.

Similar to our results, two mesenchymal cell types can be distinguished also in the embryonic lung: sub-epithelial (closer to the epithelium) and sub-mesothelial (adjacent to the mesothelium). Interestingly, mesenchymal cells in the lung can move between these two compartments, as shown by time-lapse microscopy in embryonic explants (Goodwin et al., 2022). In the time-lapse datasets analysed in this study, we also recorded the dynamic behaviour of the Notch1-labelled mesenchymal cells surrounding the epithelium. This live-imaging data could be used to investigate whether mesenchymal cells dynamically move between the sub-epithelial and dermal compartments in the embryonic MG. Nevertheless, a gene expressed specifically in the mesenchyme should be used as reporter for this experiment. I propose to perform time-lapse analysis of *ex vivo* cultures using a Col1a2-Cre^{ERT2}/R26^{mTmG} lineage tracing approach, in which fibroblasts will be targeted.

At birth, three distinct mesenchymal populations could be resolved. Two fibroblasts subtypes were apparent: *Pi16*$^{+}$*Dpt*$^{+}$ and *Col15a1*$^{+}$*Dpt*$^{+}$ clusters, as previously identified across 17 other tissues (Buechler et al., 2021). *Col15a1*$^{+}$ fibroblasts (universally identified in all tissues analysed), have the capacity to secrete BM proteins, while the *Pi16*$^{+}$ subtype potentially serves as a progenitor for more specialized fibroblasts (Buechler et al., 2021). The third cluster identified by our scRNA-seq analysis is composed of smooth muscle related cells.

Two functionally and spatially distinct mesenchyme regions in the embryonic mammary gland at E18.5 were described in the 1980s. The primary mammary mesenchyme surrounding the mammary epithelium was described to be endowed with fibroblastic features, while the other type, known as the fat pad precursor mesenchyme, was related to adipogenesis (Sakakura et al., 1982). This seminal paper raised the question of the evolution of these mesenchymal populations.

In our P0 dataset, the *Col15a1*$^{+}$ cluster expresses genes that are typical surface markers of pre-adipocytes, such as *Fabp4, Aoc3* and *Pparg* (see Figure S4C from *Results Chapter 1*). This suggests that there could be two independent lineages, fibroblasts and pre-adipocytes by birth, supporting previous studies by Sakakura et al. (Sakakura et al., 1982). However, PDGFRα is expressed by all mesenchymal cells at P0 (data not shown). PDGFRα is a commonly used marker for fibroblasts and undifferentiated mesenchymal cells, that is not expressed by differentiated adipocytes (Sun et al., 2017). This suggests that cells acquire mature adipocyte differentiation and function only after birth. Lineage tracing experiments using stromal Cre drivers are required to understand how mesenchymal cells adopt a fibroblast or an adipocyte fate. Previous studies demonstrated that PDGFRα signalling regulates the balance between adipocyte and fibroblast differentiation trajectories (Sun et al., 2017). In addition, the ER-rtTA/TetO-Cre/Rosa-YFP mouse model used by Van Keymeulen et al. (van Keymeulen et al., 2017), could be used to label the sub-epithelial mesenchymal cells at E13.5 in order to assess the fate of the mesenchymal cells that express ERα in the embryonic gland.

Similar to the mammary mesenchyme at early embryonic stages, functional heterogeneity at birth may underlie spatial heterogeneity. Fibroblasts involved in morphogenesis may reside in close proximity to the embryonic mammary ducts, whereas adipogenic mesenchymal cells may be located further away. Additional smRNA-FISH using several candidate genes for each population will be performed to address this question.

5. Mesenchyme-epithelium communication is involved in mammary branching morphogenesis

The spatial localisation of the two mammary mesenchymal populations at early stages of mammogenesis raised questions about how and which mesenchymal compartment influences embryonic mammary branching.

Epithelial PTHrP to mesenchymal PTHR1 signalling is considered as the master epithelial inducer of mesenchymal differentiation (Spina & Cowin, 2021). The fact that *Pthlh* is restricted to the region closer to the BM at E15.5 (see Figure S3C from *Results Chapter 1*), prompted us to study the communication between the mesenchymal compartments and basal-like cells. Through this analysis, we provide valuable insights into the complex signalling crosstalk between epithelial and mesenchymal cells within the embryonic MG.

Using the CellphoneDB tool, we performed a cell-cell interaction analysis between basal-like cells in close proximity to the BM and the two spatially distinct mesenchymal compartments located close to, or further away from, the mammary embryonic bud. Our results highlighted several signalling receptor-ligand pairs of interest, including FGF pathway regulators such as FGF10.

The FGF pathway plays an important role in controlling placode formation, mediated by reciprocal signalling between the mammary epithelium and the surrounding mesenchyme. Indeed, knock-out mice for the FGF receptor *Fgfr2b* or its ligand *Fgf10* fail to develop mammary placodes, suggesting that the FGF10-FGFR2B signalling is required to initiate embryonic mammary gland development (Mailleux et al., 2002). However, this phenotype precluded studies into the role of the FGF10-FGFR2B signalling axis on mammary embryonic development. Interestingly, it has been shown that the FGF10-FGFR2B interaction is also important in embryonic lung branching morphogenesis (Kouros-Mehr & Werb, 2006). Of particular relevance, FGF signalling is known to be an important driver of postnatal mammary branching morphogenesis (Hannezo et al., 2017). We therefore asked whether FGF10 played a similar role during embryonic mammary development. Using mammary embryonic explants, we demonstrated that FGF10 is indeed important for mammary branching as it significantly increases the branching velocity (see Figure 12B in *Results Chapter 1*). Furthermore, it has been shown that FGFR2 contributes to postnatal mammary branching by stimulating epithelial cell proliferation in TEBs (P. Lu et al., 2008). *Fgfr2*-null epithelial cells are rapidly outcompeted in the TEBs by their WT neighbour counterparts during branching morphogenesis (P. Lu et al., 2008). Therefore, we assessed whether there is an increase of proliferative cells in the branching tips when adding FGF10 to the culture medium. However, quantification of proliferative cells by Edu incorporation experiments showed that proliferation does not substantially change in the presence of FGF10 in the medium of *ex vivo* cultures. This suggests that embryonic branching morphogenesis might not be mediated by localised increases in cell proliferation, but may instead be driven by the directional migration of cells, as previously discussed. High resolution imaging and tracking of individual cells within mammary embryonic *ex vivo* cultures exposed to exogenous FGF10 will be required to definitely address this question.

In conclusion, we show that the availability of this growth factor in the microenvironment perturbs branching dynamics in the embryo, indicating that the distribution of FGF10 plays a role in the embryonic branching process. Additionally, other growth factors, such as TGFβ and IGF, were identified in the CellPhoneDB analysis and they should be further explored. It has been shown that a local increase of TGFβ led to the premature termination of TEBs without the presence of a neighbouring duct in the pubertal MG (Hannezo et al., 2017). However, the role of TGFβ in embryonic mammary gland development has yet to be investigated. Agarose beads soaked with different growth factors can be used in embryonic explant cultures, in order to study their effect on branching directionality.

Although these experiments provide tantalising insights into the molecular modulators involved in branching, the molecular signals that regulate the initiation of mammary branching in the embryo remain unknown. What induces mesenchymal compartments specification? How are these compartments formed, both spatially and temporally, during branching of the epithelial mammary tree? What is the origin of the mechanical forces involved in branching morphogenesis?

6. Wnt/β-catenin signalling dictates cell identity and controls branching morphogenesis in the embryonic mammary epithelium

Multiple studies have shown that Wnt/β-catenin signalling is essential not only for the onset of MG morphogenesis but remains critical at all subsequent stages of mammary embryonic development (see *Introduction* section *4.1.1.*) (Spina & Cowin, 2021). In addition, two independent labs showed a switch in Wnt/β-catenin signalling in the mammary epithelium after birth, becoming active in unipotent adult BCs (de Visser et al., 2012; van Amerongen et al., 2012).

Most previous studies investigating the role of Wnt signalling on mammary embryonic development have used models whereby Wnt/β-catenin signalling is abrogated. Instead, in this work, I used an inducible β-catenin mutant mouse model that results in constitutive Wnt pathway activation. In these mice, Cre-mediated excision of loxP-flanked exon 3 of the endogenous *Catnb* gene, which contains all GSK3β phosphorylation target serine/threonine residues, leads to stabilization of β-catenin in the cytoplasm and subsequent translocation to the nucleus to activate Wnt target genes (Harada et al., 1999).

We selected this mutant model in order to define the molecular signatures dictating mammary cell differentiation during embryonic mouse development, by analysing how the differentiation potential may be skewed in response to constitutive Wnt pathway activation. However, we found that β-catenin stabilization blocks branching morphogenesis both *in vivo* and *ex vivo*, which precluded extensive analysis of its impact on mammary cell fate outcomes.

Unexpectedly, our scRNA-sequencing data revealed the presence of skin cells in the N1Cre/β-cat/Tom dataset, which were not observed in the WT dataset at the same developmental stage. It is important to note that, in contrast to the bud stage, at P0 the mammary fad pad can be easily micro-dissected without skin contamination (like in the adult MG).

Recent lineage-biased Wnt gain-of-function studies performed in the lab during pubertal development revealed that mutant β-catenin stabilization drives epidermal transdifferentiation of mammary epithelial cells (Lloyd-Lewis et al., 2022). Therefore, we hypothesised that a similar mechanism may occur when inducing constitutive Wnt/β-catenin signalling in the embryonic MG. To test this hypothesis, we plan to investigate whether hair follicle markers are expressed in the few mutant embryonic buds that were able to give rise to rare branching structures (either on the *in situ* gland, or in *ex vivo* cultures).

An interesting study from the laboratory of V. Greco analysed the effect of Wnt/β-catenin signalling on the behaviour of hair follicle stem cells. They show that β-catenin mutant cells activate Wnt signalling within the neighbouring WT cells via the secretion of Wnt ligands which induce new hair growths (Deschene et al., 2014). In addition, stimulation of both Wnt/β-catenin and ectodysplasin/NF-kB signalling increased cell motility and the number of cells committed to hair placodal fate (Ahtiainen et al., 2014). Although speculative, this might imply that Wnt signalling activation increases the number of cells committing to a hair placodal fate. Of note, hair follicle fate and morphogenetic events are controlled by the same molecular pathways in hair placode formation, suggesting a coordination of these two processes (Ahtiainen et al., 2014).

To further unravel the dynamic interactions of mutant cells with their WT neighbours during mammary branching, we performed longitudinal imaging in *ex vivo* cultures in which β-catenin stabilization was induced in few mammary cells (by reducing the TAM dose by 10-fold). We found that low level induction of mutant Wnt signalling activity, however, had no effect on mammary cell movements and behaviour in comparison to control conditions. We hypothesise that this is due to the mosaic induction of the mutation, with too few mutant β-catenin cells to perturb mammary embryonic development. Future experiments focused on deciphering whether Wnt/β-catenin signalling in the embryonic MG promotes a hair follicle-like fate are detailed in the *Future perspectives* section (*page 112*).

7. Link between multipotency of embryonic stem cells and cell plasticity in cancer

“Unlocking phenotypic plasticity” is now considered a new hallmark of cancer (Hanahan, 2022). There is increase evidence that dedifferentiating to an embryonic-like cell state is a critical event in the development of many cancer types (Hanahan, 2022). Recently, comparative transcriptome profiling revealed that invasive pancreatic neuroendocrine tumours acquire a progenitor-like molecular phenotype and are more similar to embryonic islet cell precursors (Saghafinia et al., 2021). In basal cell carcinomas of the skin, tumour-initiating cells are reprogrammed into an embryonic hair follicle progenitor-like fate as early as one week after oncogenic activation (Youssef et al., 2012). Another study using a zebrafish model of melanoma shows that reactivation of neural crest progenitor genes is a key event in melanoma initiation (Kaufman et al., 2016). In the MG, it has also been suggested that certain cancers may arise from reactivation of embryonic mammary signatures in postnatal differentiated cells (Spike et al., 2012; Zvelebil et al., 2013).

Two important studies published back-to-back in 2015 demonstrated that targeting the constitutively active PIK3CA mutation, one of the most common mutations found in breast cancer patients, in either basal or luminal unipotent mammary progenitors promotes the reacquisition of a multipotent embryonic stem cell-like state (Koren et al., 2015; van Keymeulen et al., 2015). Interestingly, targeting unipotent LCs gives rise to aggressive basal-like tumours, whereas targeting BCs gives rise to less aggressive luminal ERα$^+$ tumours. Additionally, it has been demonstrated that LCs are capable of producing BCs after activation of either polyoma middle T (PyMT) antigen or ErbB2 signalling (Hein et al., 2016). Collectively, these studies demonstrated that lineage-committed adult mammary cells retain a high degree of plasticity throughout adulthood and that they can be reprogrammed into stem-like cells under stress conditions. Therefore, the identification of genes and transcriptional regulators involved in maintaining an embryonic multipotent stage during early embryonic development may improve our understanding of how these genes are altered in cancer.

A recent study also revealed that the transcription factor *Sox10* is associated with an embryonic mammary stem cell state (Dravis et al., 2018). They observed that binding motifs for *Sox4, Sox9* and *Nf1* are enriched in chromatin that is accessible in stem/progenitor cells, but inaccessible in differentiated cells. In addition, they found that *Sox10* is expressed in 3 different breast cancer models and that tumour cells with high levels of *Sox10* have a stem/progenitor-like identity. Finally, they suggested that *Atf3, Nf1* and *Elf5* motifs may cooperate with *Sox10* in regulating mammary cell states (Dravis et al., 2018). Interestingly, in our E15.5 scRNA-seq dataset, *Atf3* presents a higher expression level in the basal-like cluster and it decreases along our differentiation pseudotime (Figure 26), while *Nf1* is expressed by all MECs at E15.5 and *Elf5* is very lowly expressed.

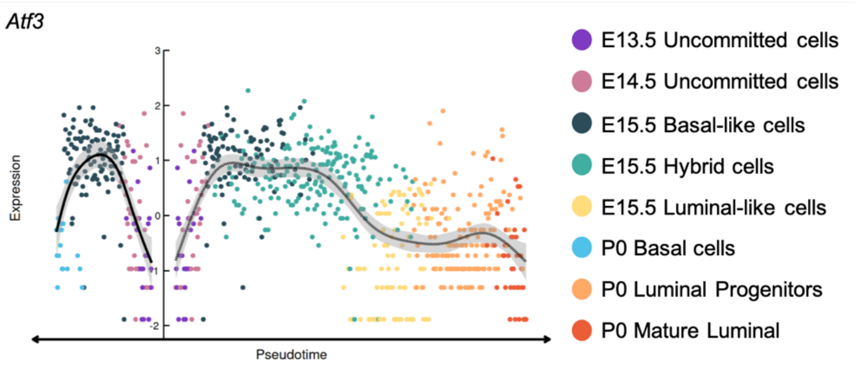


Figure 26 | *Atf3* pattern of expression along the mammary differentiation trajectory that we obtained. *Atf3* expression in our pseudotime analysis towards luminal differentiation (to the right) or basal differentiation (to the left). *Atf3* expression decreases along embryonic developmeny. Cells are coloured according to identified cell clusters.

To conclude, future studies should focus not only on the tumour cells of origin, but also on determining the contribution of deregulated genetic programs to cancer initiation and progression (Rodilla & Fre, 2018). Further studies are required to decipher the signals defining cell identity acquisition during embryonic development, to better understand the mechanisms that enable reactivation of multipotency programs in unipotent lineage-committed adult cells. Identifying specific targets that may be functionally critical in the early stages of tumour formation may translate into greater knowledge of breast cancer and help to define new therapy options for patients.

Future Perspectives

The findings obtained in this work have paved the road and provided both the methodology and the initial datasets for studying the coordination between cell fate and morphogenesis. However, several outstanding questions and challenges remain to be answered. How can we disentangle these two distinct processes that act within the same developmental window? How can we impair one to test the role of the other, and vice versa? (Chan et al., 2017).

Here, I will focus the new perspectives that these unanswered questions open.

1. What is the exact timing of the switch from multipotent MaSCs to unipotent progenitors?

The precise timing of lineage specification during mammary gland development remains unclear, hampered by the lack of specific markers for each early committed cell lineage. My work has identified novel markers of MECs committed to luminal and basal lineages as early as E15.5 during mammary development, providing new avenues for further lineage tracing analysis.

To date, lineage tracing studies have typically relied on widely expressed gene promoters, such as K14, K5 and K8. Embryonic mammary cells, however, co-express many of these markers that are commonly used for lineage tracing mammary cells in the postnatal gland. Moreover, the exact timing of the switch from multipotency to unipotency may vary depending on the promoter used for tracing. For instance, at birth, K5 is expressed only by the outermost layer of cells. Consistently, when inducing K5-Cre^{ER}/Rosa-YFP expression at P1, K5-expressing cells are already restricted to the basal cell fate. In contrast, K14 is still expressed by inner and outer cells at P1 and it only becomes restricted to BCs after P10. For this reason, doxycycline administration to K14-rtTA/TetO-Cre/Rosa-YFP mice at P1 marked both basal cells and also a small proportion of luminal cells (van Keymeulen et al., 2011). Therefore, early restricted gene promoters are necessary to distinguish multipotent embryonic cells from the ones that are already committed to a luminal or basal fate and that co-exist with multipotent MaSCs within the embryonic bud (Lilja et al., 2018), as well as to define the precise timing of their specification. To this end, in future work I would seek to perform **lineage tracing experiments with two genes targeting specifically luminal-like or basal-like cells** found in our E15.5 dataset, to assess their differentiation potential and contribution to each epithelial compartment.

In addition, a lineage tracing technique using a novel mouse line allowing barcoding of mammary cells *in vivo* (Perié et al., 2014) will shed some light in defining quantitatively the timing and proportion of multipotent and unipotent MECs. These experiments are currently being performed by Candice Merle, a post-doc in the lab. This mouse line, crossed with a luminal (N1-Cre^{ERT2}) or basal (K5-Cre^{ERT2}) Cre line, allows to induce barcode writing in MECs at different embryonic developmental times. By sequencing the barcode-bearing clonal progeny (sorted by FACS as BCs and LCs) of each cell at the onset of puberty, the goal here is to determine the ratio of multipotent and committed cells throughout mammary embryonic development.

2. What are the molecular mechanisms regulating cell fate specification?

The specific transcriptional regulators that control the bifurcation of embryonic cells towards a luminal or basal differentiation process during MG development remain unknown. To address this, I intend to perform a Single-Cell regulatory Network Inference and Clustering (SCENIC) analysis, a bioinformatic method that enables the identification of gene regulatory regions by inferring co-expression between transcription factors and their target genes (Aibar et al., 2017). By combining SCENIC analysis with the pseudotemporal ordering of cells we have already obtained, we could **identify putative transcription factors with branch-specific activation patterns**.

For instance, P63 acts as a major regulator promoting the switch from multipotency to basal unipotency during MG development, since overexpression of the transcription factor *Trp63* in committed luminal cells was sufficient to reprogram these cells into BCs (Wuidart et al., 2018). Moreover, Notch1 was found to have an opposing role, pushing MECs to acquire luminal identity, specifically ERα/PR$^-$ LCs (Lilja et al., 2018).

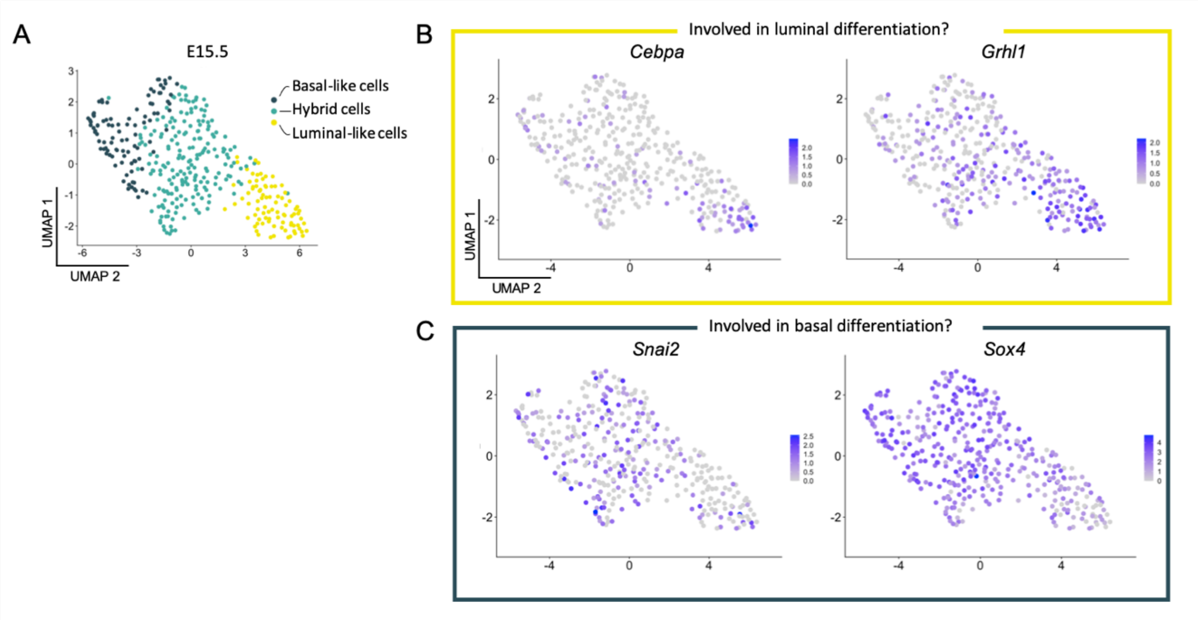


Figure 27 | Candidate transcription factors involved in luminal or basal differentiation.
(A) Schematic UMAP representation of the 3 MEC clusters identified at E15.5 (from Figure 9B in *Results Chapter 1*). (B) UMAP illustration of *Cebpa* and *Grhl1* expression levels at E15.5, mostly concentrated within the luminal-like cluster. (C) UMAP illustration of *Snai2* and *Sox4* expression levels at E15.5, spread among the basal-like and hybrid clusters.

One interesting study revealed that the E18.5 MaSC population is already composed of cells with basal and luminal-oriented chromatin features (Chung et al., 2019). The authors identified 148 transcription factors that are important to predict whether a cell belongs to a fetal state, or to a basal, LP or ML adult population. I focused on the transcription factors annotated to be enriched in fetal clusters and assessed their pattern of expression in our E15.5 dataset in order to obtain a list of candidate genes which could be involved in luminal and basal differentiation trajectories during embryonic development. Within this list, *Cebpb,* known to regulate luminal cell fate in the MG (LaMarca et al., 2010), *Cebpa* and *Grhl1,* transcription factors involved in regulation of epithelial-mesenchymal transition (EMT) (Lourenço et al., 2020; Sengez et al., 2019), are promising candidates involved in luminal differentiation (Figure 27A- B). This analysis also suggested that *Snai2*, a transcriptional repressor that may function as a repressor of luminal characteristics (Phillips et al., 2014), *Sox4*, one member of the Sox family that may contribute to cellular differentiation, or *Egr1* may have roles in basal differentiation (Figure 27A- C), and would represent promising candidates for future investigations.

In addition, recombinant adeno-associated virus (rAAV)-mediated gene transfer in embryonic mammary explants has been recently optimised (Lan & Mikkola, 2020) and could be performed to functionally validate these candidate transcription factors for the different lineages. Briefly, embryonic mammary buds could be dissected from a R26-Cre/Cas9-EGFP knock-in mouse (Platt et al., 2014), and infected with a rAAV expressing single-guide RNA (sgRNA) targeting one of these transcription factors, to allow us to study its role in cell fate specification during mammary branching morphogenesis.

3. Does the subrabasal layer of the skin participate in the formation of the inner (luminal) compartment of the embryonic MG?

Strikingly, there is a continuum from the suprabasal layer of the skin to the middle region of the mammary bud, as seen in Figure 10D from *Results Chapter 1*. To investigate whether *Anxa1*-expressing cells located in the neck that connects the mammary bud with the skin will only form the nipple area throughout development, or whether suprabasal skin cells migrate within the bud during the invagination process and participate in the formation of the luminal compartment, I propose to **perform clonal analysis at saturation** (Wuidart et al., 2016). To this end, DOX-inducible K8-rtTA/TetO-Cre/Rosa-tdTomato mice would be used, as K8 is one of the few markers known to be exclusively expressed in the mammary bud but not in the skin epidermis (see Figure S3D from *Results Chapter 1*). This model should allow us to label all MECs in the embryo. We would administer DOX in pregnant females starting from day E10.5, when mammary development starts, until E17.5, when K8 expression becomes progressively restricted to the inner region of the bud (Wuidart et al., 2018).

First, we would confirm that all MECs are labelled following DOX treatment. If the luminal lineage was composed of cells from the suprabasal skin, the percentage of labelled cells in the pubertal MG would be expected to decrease with time. On the other hand, if all LCs in the pubertal MG are labelled, it means that skin cells do not contribute to mammary lineages.

On the other hand, we could also perform lineage tracing using a K10 promoter to label specifically skin cells from the suprabasal layer, in order to assess their contribution to the luminal compartment of the adult MG.

4. Does cell position in the bud instruct cell fate, or do cells move to a specific tissue region after lineage commitment?

The transcriptomic results obtained in this study provide us with new avenues to investigate the coordination between cell fate and branching morphogenesis, a key question that will be addressed in the near future. To this aim, another PhD student in the lab, Robin Journot, is **generating fluorescent reporter mice for *Anxa1* and *Cxcl14***, which will facilitate the tracking of cells that differentiate towards the luminal or basal fate, respectively. We selected the mAmetrine and miRFP670 fluorophores as reporters for *Anxa1* and *Cxcl14*, respectively, in order to be able to cross these novel reporter mice with other transgenic models expressing tdTomato or mGFP fluorescence.

Is fate specification dictated at architecturally restricted sites within the primordial mammary gland? How do individual cells sense global changes in the tissue and translate these signals into differential gene expression? *Ex vivo* mammary cultures using Anxa1-mAmetrine and Cxcl14-miRFP670 reporter mice will allow to assess whether cell position, such as attachment to the basement membrane, or cell dynamic behaviour, such as cell division, cell migration or cell rearrangements during branching morphogenesis, contribute to cell fate specification.

Are the same molecular mechanisms involved in cell fate specification conserved in different organs? Our approaches could be extended to compare tissue morphology and lineage differentiation in other organs. These reporter mice could also be used to study cell fate specification in other tissues, such as the prostate or the submandibular salivary gland, since they also express *Anxa1* and *Cxcl14*. Like the MG, these glandular epithelia are also derived from multipotent progenitors that become lineage restricted during development (Rocchi et al., 2021; Tika et al., 2019).

5. Does orientation of cell division influence cell fate and positioning in the growing tissue?

It has been demonstrated that oriented cell division in the airway epithelium is responsible for the directional elongation of these tubes (Tang et al., 2011). By contrast, recent data show that the fate of epithelial cells in the whisker placode is determined by the cell position within the placode, regardless of the orientation of cell division (Morita et al., 2021). Whether cell division orientation influences branching in the embryonic MG is unknown.

Time-lapse imaging of embryonic mammary explant cultures revealed that dividing cells possess two different behaviours: both daughter cells stay in the same cellular compartment, or each daughter cell goes to opposite cellular compartments. To assess the relationship between cell division orientation and cell fate acquisition in the early development of embryonic MG, we aim to measure the orientation of cell division against the BM in our live imaging data. To this end, we will collaborate with the laboratory of Dr. Yohanns Bellaïche (Institut Curie) that have trained a deep learning model able to detect the sequential cellular events associated with cell division (Figure 28). Using this systematic and high-throughput approach, we will **categorize dividing cells** within the interior and external regions of the bud into groups **based on the direction of their division axis in relation to the BM (i.e. parallel or perpendicular to the BM)**.

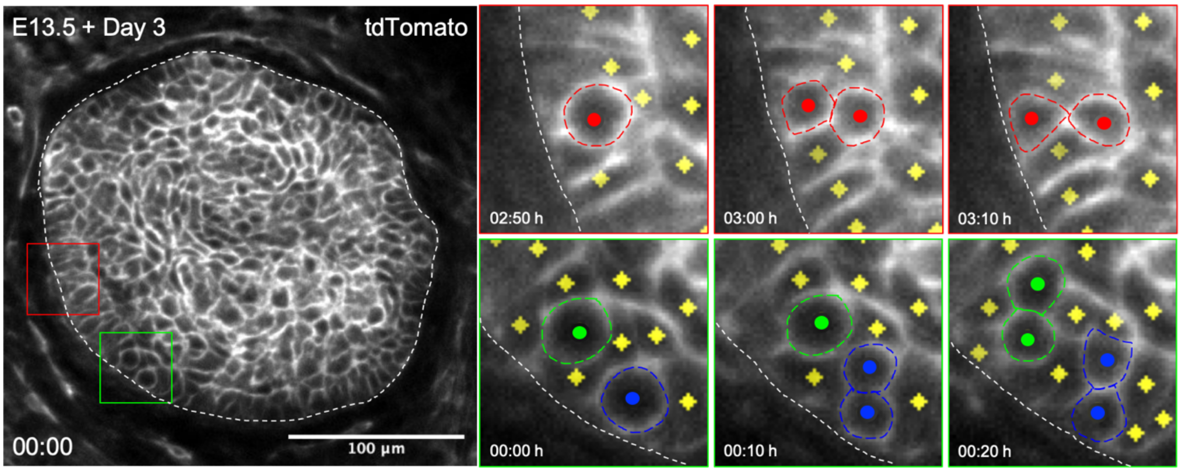


Figure 28 | Cell division orientation analysis in a mammary explant after 3 days in culture.
Three cell division events perpendicular to the BM are shown in red, green and blue. Each cell recognized by the algorithm is labelled with a yellow dot. Dotted lines delineate the BM.

6. Does cell proliferation drive branching morphogenesis in the embryonic MG?

Cell proliferation minimally contributes to hair placode formation (Ahtiainen et al., 2014). In addition, mammary rudiment formation at E13.5 is not driven by cell proliferation (Trela et al., 2021). Interestingly, the scRNA-seq dataset obtained at E15.5 revealed a population of proliferating epithelial cells, which was composed of both basal-like and luminal-like cells (Figure 29A- B). To assess whether cell proliferation is the driving force behind branch initiation in the embryonic ductal tree, rather than a consequence of branch formation, **live-imaging in *ex vivo* mammary bud cultures expressing the fluorescent ubiquitination-based cell cycle indicator (Fucci) transgenes** could be performed (Sakaue-Sawano et al., 2008). In Fucci mice, nuclei in the G1 phase of the cell cycle are marked in red, while nuclei labelled in green are in the S/G2/M phase of the cell cycle. This would allow us to define the location of highly proliferative cells during branching morphogenesis.

Adding **Mitomycin C, which inhibits DNA synthesis, to *ex vivo* cultures** could be used to assess whether cell proliferation contributes to the formation of new embryonic mammary branches. If embryonic mammary buds failed to grow and branch under these conditions, it would suggest that cell proliferation is a mandatory cellular event for mammary branching. In addition, we observed that in both N1Cre/Tom control and N1Cre/β-cat/Tom mutant buds, cell displacement was higher at the tips than in the body regions. By adding Mitomycin C in the *ex vivo* culture medium, we could test whether cells maintain this behaviour regardless of cell proliferation.

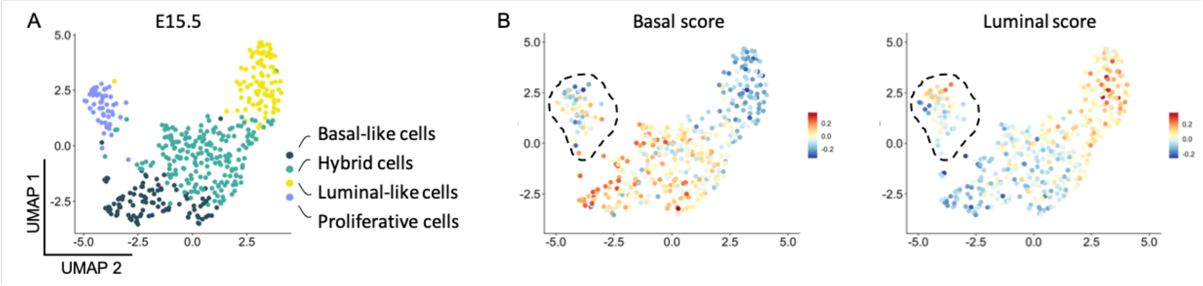


Figure 29 | Transcriptional heterogeneity of the proliferative cells at E15.5.

(A) UMAP representation of the E15.5 MECs: luminal-like cells, basal-like cells, a hybrid cell population and a cluster composed of proliferative cells (from Figure S1C in *Results Chapter 1*). (B) UMAP plots from (A) colour-coded according to the expression of the single-cell ID scores in MECs: basal score (left) and luminal score (right). The proliferative cell cluster is delineated in black.

7. Do the distinct mesenchymal cell clusters identified at birth represent spatially restricted stromal cell populations?

Our scRNA-seq data provided us with insights that permitted the spatial reconstruction of cells expressing specific genes within the mammary embryonic bud, allowing the positioning of cells acquiring luminal-like and basal-like characteristics at different developmental timepoints in the mouse embryo by smRNA-FISH. The RNAscope technology allows RNA expression measurements at the single cell level with high sensitivity and specificity within an intact tissue (F. Wang et al., 2012). However only 2-3 mRNAs can be measured simultaneously using this technique. Several methods using multiplexed smRNA-FISH, such as Seq-FISH (Sequential Barcoding FISH) (Eng et al., 2019) or MERFISH (Multiplexed error-robust fluorescence in situ hybridization) (Chen et al., 2015), have since been developed to increase the number of detectable mRNA species in each cell.

The ability to perform ***in situ* RNA sequencing directly on tissue sections** has garnered much interest. However, these recent techniques do not yet have the resolution for sequencing individual cells. For instance, a recent spatial transcriptomics method used a customized slide with microwells of 100 µm of diameter and a center-to-center distance of 200 µm to capture mRNAs (Ståhl et al., 2016). Each well was coated with an oligonucleotide containing unique positional barcodes and a poly T tail to capture mRNAs with poly A. Then, cDNA synthesis is carried out on-slide and finally, NGS-based RNA-seq is used to obtain the transcriptome (Liao et al., 2021). More recently, high-definition spatial transcriptomics (HDST) has been developed to produce a high-definition bead array with 2 µm resolution (Vickovic et al., 2019). However, HDST harnesses few UMIs per barcode location (an average of 7 UMIs), limiting its ability to measure high abundance genes. Thus, new technologies enabling high spatial resolution accompanied by high-throughput scRNA-seq represent the next challenge in this rapidly developing field of spatial transcriptomics (Liao et al., 2021).

Future experiments will undoubtedly exploit these emerging techniques to simultaneously gather the molecular characteristics and the positions of single cells within the embryonic mammary epithelium and the surrounding mesenchyme.

8. Does constitutive Wnt/β-catenin signalling induce squamous transdifferentiation of embryonic mammary epithelial cells, as found in adult development?

To investigate whether constitutive Wnt/β-catenin activation promotes hair follicle-like fate in embryonic mammary cells, we propose to **analyse the expression of epidermal and hair follicle markers in the few branches that manage to emerge in the N1Cre/β-cat/Tom model**.

Moreover, we observed that after 7 days of β-catenin stabilization, mammary epithelial cells do not express myoepithelial markers, such as *Acta2* (α-SMA). Of interest, the lesions obtained when inducing mutant β-catenin during puberty also lacked α-SMA expression. Thus, we could perform **immunostaining to compare the levels of α-SMA expression in the WT and mutant β-catenin epithelium**. Additionally, **RNAscope using an *Axin2* probe** could be performed to investigate whether and in which cells there is accumulation of *Axin2* transcripts after only a few days of Cre recombination, before birth, when branching morphogenesis is already blocked. These set of experiments in the embryonic mammary gland would determine whether there are similarities with the phenotype observed in the pubertal mammary gland in response to β-catenin activation.

Finally, we propose to perform time-lapse imaging using the N1Cre/β-cat/Tom model, but this time using the same dose as in the scRNA-seq experiment (same concentration as used in the *ex vivo* cultures in Figure 18A from *Results Chapter 3*). This will enable us to study the impact of β-catenin stabilization on the dynamic behaviour of embryonic MECs.

9. Can intrinsic signals determine cell fate?

We have previously reported that constitutive Notch1 activation in multipotent embryonic MaSCs imposes a unipotent luminal fate (Lilja et al., 2018). Since Notch1 can skew cell fate choices during MG development, we hypothesized that the comparative analysis of differentially expressed genes between WT and gain-of-function Notch1 (N1IC) mutant cells would reveal the transcriptional signals and molecular pathways driving luminal lineage commitment and loss of multipotency during embryonic mammary development.

We have thus performed a scRNA-seq experiment from an inducible gain-of-function Notch1 transgenic model crossed to a Notch1-Cre^{ERT2} line (henceforth N1Cre/N1IC) (Figure 30A). We induced the mosaic expression of the active form of the Notch1 receptor at E12.5 and sorted both mammary epithelial and mesenchymal populations 72 hours after Notch activation (at E15.5). Here, I present the preliminary data I have obtained from a first experiment.

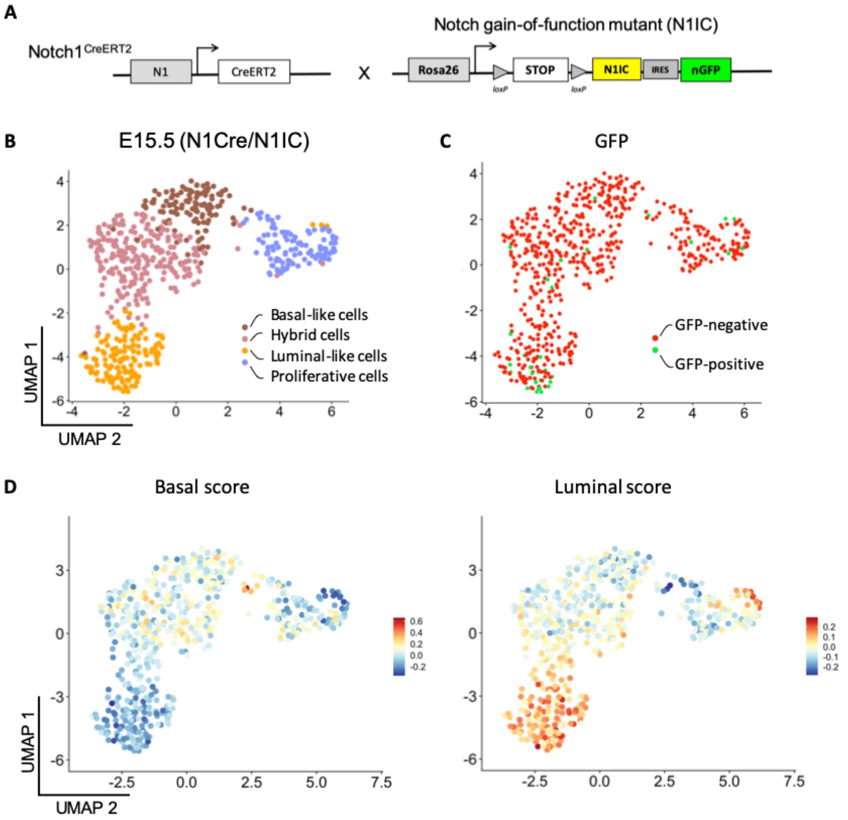


Figure 30 | Notch1 activation in embryonic mammary cells increases the proportion of epithelial cells undergoing luminal differentiation.

(A) Schematic representation of the N1Cre/N1IC mouse model. TAM administration induces Notch1 activation. The Notch-activated cells are labelled by GFP. (B and C) UMAP plots after subset analysis of N1Cre/N1IC MECs isolated by scRNA-seq at E15.5. Cells are colour-coded by cell type (in B) or by GFP expression (in C). (D) UMAP plots from (B), colour-coded according to the expression of the single-cell ID scores in MECs: basal score (left) and luminal score (right).

Unsupervised clustering of the sequenced dataset for these mutant mice revealed 9 distinct cell clusters: 4 MEC populations, 2 skin clusters and 3 mesenchymal populations (data not shown), just like in the WT dataset at E15.5 (see Figure S1B from *Results Chapter 1*). Then, we performed a sub-clustering analysis of MECs (Figure 30B). A cluster of proliferative epithelial cells was identified based on a list of cell cycle related genes. To understand the differences in the transcriptome due to Notch activation, we first detected the GFP$^+$ cells in the N1Cre/N1IC mutant sample (Figure 30C). Few mutant cells were obtained, likely due to an induction of only 72 hr. We then applied the basal and luminal scores that we had previously generated to analyse the WT dataset (see Figure 9C from *Results Chapter 1*) and observed that MECs in the N1Cre/N1IC dataset can also be divided into three populations: basal-like cells, luminal-like cells and hybrid cells. Interestingly, the GFP$^+$ cells are mainly located in the luminal-like cluster (Figure 30D), confirming that Notch1 activation drives luminal fate choices.

To establish the early signals involved in luminal differentiation, an additional scRNA-seq experiment should be performed, where we would induce Notch1 activation at E12.5 and analyse the glands 7 days after (at P0). We could then order the cells along a pseudotemporal trajectory and perform a comparative analysis of the genes presenting a distinguishing pattern of expression along luminal differentiation in the WT and N1Cre/N1IC datasets.

Finally, performing *ex vivo* cultures using this mutant model could yield important information on how cell fate and morphogenesis are coordinated, since it would enable us to study how branching morphogenesis occurs when luminal differentiation is enforced.

In conclusion, during my PhD, I have deciphered the heterogeneity of the mammary epithelium and mesenchyme throughout embryonic development and defined the transcriptional programs orchestrating the lineage restriction of multipotent MaSCs to unipotent progenitors. To achieve this, I coupled single cell transcriptional mapping across developmental timescales with *ex vivo* live imaging of mammary embryonic cell dynamics during branching morphogenesis. Importantly, this integrative approach prospectively identified new markers for specific mammary epithelial cells at early embryonic stages. Moreover, this work provides important insights into the poorly explored resident mammary embryonic mesenchymal cell populations that direct epithelial branching morphogenesis. To further understand the intimate connection between lineage acquisition and morphogenesis, we also used an inducible β-catenin mutant mouse model that results in constitutive Wnt/β-catenin pathway activation, which blocked embryonic mammary branching.

Acknowledgements

I would like to thank everyone that has made this thesis possible.

This thesis was carried out at Institut Curie in the "Genetics and Developmental Biology" Department with the financial support from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 666003 and the Medical Research Foundation FRM Project 12917.

I would like to first acknowledge the jury members of my thesis, Felicity Davis, Walid Khaled, Lucie Peduto and Philippe Chavrier for kindly accepting to participate in evaluating this scientific work and reviewing this manuscript.

Thank you, Silvia, for giving me the opportunity of joining the lab and start this PhD adventure in Paris. You have found the balance between guidance and freedom, allowing me to grow as a scientist. Also, thank you for encouraging me to go to several meetings and courses abroad during these years. I have had the great opportunity to go to Weggis, Los Angeles, Heidelberg and Roscoff.

I would like to thank my Thesis Committee members. During these years I have received support from Lucie Peduto, Marie Breau and Olivier Renaud. I am grateful for all your feedback and ideas to progress with my research. Also, thanks to Yohanns Belaïche and Jean-Léon Maître for the fruitful discussions we have had during these years.

I have had the opportunity to work close with 3 amazing women. Bethan, you were the first person I met when I arrived to the lab and since then you have been a reference for me. I am truly thankful for your time, guidance and support. I also want to thank you to Meghan. I have been the luckiest person to have the opportunity to meet you. You are a fantastic scientist, a wonderful person and a good friend. Y por último, pero no menos importante, gracias Marisa. Has sido una persona vital para mí. Gracias por tu infinita ayuda, por estar siempre pendiente del progreso de mis experimentos, por todas las ideas que me has dado y, sobre todo, por conocerme tan bien y saber cuándo necesitaba un bizcocho, unas aceitunas o un poco de jamón para desconectar.

I would like to thank all my colleagues in the lab: Candice, Jakub, Robin, Mathilde and Calvin; and the past members: Fatima, Fairouz, Francesca, Ulysse and Guillaume. Thank you all for both the scientific and personal discussions, and for helping me with all my French-related questions. Merci beaucoup!

Special thanks to Wenjie, for all your help and patience with the bioinformatic analysis, I have learnt a lot from you.

Thanks to Olivier and Olivier for all your help with all the microscopes that I have tried to be able to image the explants. Thanks to Anne Sophie for your help in the image analysis. *Iris* says that I am the first user in both the LSM780 and LSM880! And this time in the microscope would not have been useful if I had not started working with Varun. We have done a good team working together to develop the image analysis pipeline. I didn't know anything about deep learning when I started and you have taught me everything. Thanks a lot, Varun.

I would also like to say thank you to Ana Rita, one of the most wonderful people I have come across along the way. To my amazing neighbours Francesca, Mathilde and Edgar, who made my days in the lab more fun. And to all the amazing people that I have met from all over the world: Deep, Ram, Javiera, Jaime, Sam, Özgë, Tommaso, Pallavi and Marci.

When I arrived to the lab, no one was working with the embryos. I have spent countless hours in the microscopy room and use all my energy in the binocular dissecting hundreds of buds. I have the feeling that I could start now my second PhD to try to answer all the new questions that these results have opened. I want to use one last reference in this work: "How To Choose a Good Scientific Problem" by Uri Alon, 2009 (and his TED talk). I think all students should know about this concept of "the cloud" of the unknown and all the process between a question and the final results.

Now, with your permission, I am going to switch to Spanish…

Siempre me acuerdo de ti, Ilenia. Tú fuiste quien me enseñó cómo es el trabajo de laboratorio, y lo más importante, quien me ayudó a darme cuenta de que esto es lo que yo quería hacer. Si ahora soy investigadora es gracias a ti. También quiero dar las gracias a Tamara, junto a quien trabajé codo con codo. Vosotras me enseñasteis todo lo que sé sobre glándula mamaria.

Miguel, gracias por confiar en mí, por enseñarme tantas técnicas nuevas y por hacer del laboratorio un lugar cómodo, divertido y apasionante.

No puedo olvidarme de mis amigos de Portuñol, Xavi y Gonçalo. I could not have done this without you, *chicos*. Xavi, gracias por estar siempre dispuesto a ayudarme. De ti valoro tu capacidad para encarar los problemas, tu disponibilidad para ayudar a todo el mundo sin recibir nada a cambio, tu don para hacer croquetas y tu carácter cariñoso, sin el que no habría podido sobrevivir durante mis crisis arácnidas, entre otras. ¿Quién dijo que un madrileño y un catalán no pueden llevarse bien? Muito obrigada, Gonçalo. We met the first day that you arrived to Paris and since then we have never been separated. Thank you so much for taking care of me, for listening to me and for always being there. Thanks for feeding me and *Poisson*. We have been *compañeros de* Bacalhau à Brás, microscope and quarantine. You both are the most important part that I take with me from Paris.

Gracias a Silvia, Sandra y Darine; habéis hecho de París un lugar mejor.

A mis amigos Paloma, Leti, Sara, Lucía, Carlos, Dani y Germán (el orden de los factores no altera el producto), gracias por estar tan cerca a pesar de estar tan lejos. Siempre habéis sacado tiempo para verme, haciéndome sentir una Kawi más desde París. Gracias, chicas, por venir a visitarme, hemos demostrado que una habitación de 20 metros es más que suficiente para que convivan cuatro personas.

Quiero dar las gracias a mi padre, mi madre y mi hermano, por recibirme siempre en casa con los brazos abiertos, por las vídeollamadas cada semana, por los tuppers de paella sin los que no habría podido sobrevivir, por estar pendiente de mí y por escucharme. Esto no habría sido posible sin vosotros. Papá, ya he llegado al final de esta carrera de obstáculos. Ahora voy a por la siguiente.

También quiero acordarme de la Claudia de 2018, esa chica que no sabía si dar el paso y lanzarse a vivir una nueva vida en París. Gracias a tu valentía y determinación ahora soy una persona mejor y más completa.

Estos 4 años se los dedico, sin duda alguna, a Víctor. Gracias por haber sido siempre mi mayor apoyo. Hoy, 24 de julio de 2022, lo hemos conseguido. ¿Estás preparado para nuestra siguiente aventura?

Y, por último, gracias a todos los que van a leer esta tesis. Espero que disfrutéis tanto su lectura como yo he disfrutado su escritura - I hope that you enjoy reading this work as much as I have enjoyed writing it.

Claudia

References

Ahtiainen, L., Lefebvre, S., Lindfors, P., Renvoisé, E., Shirokova, V., Vartiainen, M., Thesleff, I., & Mikkola, M. (2014). Directional Cell Migration, but Not Proliferation, Drives Hair Placode Morphogenesis. *Dev Cell*, *28*(5), 588–602. https://doi.org/10.1016/j.devcel.2014.02.003

Ahtiainen, L., Uski, I., Thesleff, I., & Mikkola, M. L. (2016). Early epithelial signaling center governs tooth budding morphogenesis. *Journal of Cell Biology*, *214*(6). https://doi.org/10.1083/jcb.201512074

Aibar, S., González-Blas, C. B., Moerman, T., Huynh-Thu, V. A., Imrichova, H., Hulselmans, G., Rambow, F., Marine, J. C., Geurts, P., Aerts, J., van den Oord, J., Atak, Z. K., Wouters, J., & Aerts, S. (2017). SCENIC: Single-cell regulatory network inference and clustering. *Nature Methods*, *14*(11), 1083–1086. https://doi.org/10.1038/nmeth.4463

Bach, K., Pensa, S., Grzelak, M., Hadfield, J., Adams, D. J., Marioni, J. C., & Khaled, W. T. (2017). Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-02001-5

Balinsky, B. I. (1950). On the prenatal growth of the mammary gland rudiment in the mouse. *J Anat*, *84*(3), 227–235.

Barker, N., Bartfeld, S., & Clevers, H. (2010). Tissue-resident adult stem cell populations of rapidly self-renewing organs. *Cell Stem Cell*, *7*(6), 656–670. https://doi.org/10.1016/j.stem.2010.11.016

Beleut, M., Rajaram, R. D., Caikovski, M., Ayyanan, A., Germano, D., Choi, Y., Schneider, P., & Brisken, C. (2010). Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(7), 2989–2994. https://doi.org/10.1073/pnas.0915148107

Biggs, L. C., & Mikkola, M. L. (2014). Early inductive events in ectodermal appendage morphogenesis. *Seminars in Cell and Developmental Biology*, *25–26*, 11–21. https://doi.org/10.1016/j.semcdb.2014.01.007

Blaas, L., Pucci, F., Messal, H. A., Andersson, A. B., Ruiz, E. J., Gerling, M., Douagi, I., Spencer-Dene, B., Musch, A., Mitter, R., Bhaw, L., Stone, R., Bornhorst, D., Sesay, A. K., Jonkers, J., Stamp, G., Malanchi, I., Toftgard, R., & Behrens, A. (2016). Lgr6 labels a rare population of mammary gland progenitor cells that are able to originate luminal mammary tumours. *Nature Cell Biology*, *18*(12), 1346–1356. https://doi.org/10.1038/ncb3434

Blanpain, C., & Fuchs, E. (2014). Plasticity of epithelial stem cells in tissue regeneration. *Science*, *344*(6189). https://doi.org/10.1126/science.1242281

Blanpain, C., Horsley, V., & Fuchs, E. (2007). Epithelial Stem Cells: Turning over New Leaves. *Cell*, *128*(3), 445–458. https://doi.org/10.1016/j.cell.2007.01.014

Brisken, C., & O'Malley, B. (2010). Hormone action in the mammary gland. *Cold Spring Harbor Perspectives in Biology*, *2*(12). https://doi.org/10.1101/cshperspect.a003178

Buechler, M. B., Pradhan, R. N., Krishnamurty, A. T., Cox, C., Calviello, A. K., Wang, A. W., Yang, Y. A., Tam, L., Caothien, R., Roose-Girma, M., Modrusan, Z., Arron, J. R., Bourgon, R., Müller, S., & Turley, S. J. (2021). Cross-tissue organization of the fibroblast lineage. *Nature*, *593*(7860), 575–579. https://doi.org/10.1038/s41586-021-03549-5

Carabaña, C., & Lloyd-Lewis, B. (2022). Multidimensional Fluorescence Imaging of Embryonic and Postnatal Mammary Gland Development. In Vivanco MM (Ed.), *Methods Mol Biol* (Second, Vol. 2471, pp. 19–48). https://doi.org/10.1007/978-1-0716-2193-6_2

Centonze, A., Lin, S., Tika, E., Sifrim, A., Fioramonti, M., Malfait, M., Song, Y., Wuidart, A., van Herck, J., Dannau, A., Bouvencourt, G., Dubois, C., Dedoncker, N., Sahay, A., de Maertelaer, V., Siebel, C. W., van Keymeulen, A., Voet, T., & Blanpain, C. (2020). Heterotypic cell–cell communication regulates glandular stem cell multipotency. *Nature*, *584*(7822), 608–613. https://doi.org/10.1038/s41586-020-2632-y

Chakrabarti, R., Celià-Terrassa, T., Kumar, S., Hang, X., Wei, Y., Choudhury, A., Hwang, J., Peng, J., Nixon, B., Grady, J., DeCoste, C., Gao, J., van Es, J., Li, M., Aifantis, I., Clevers, H., &

Kang, Y. (2018). Notch ligand Dll1 mediates cross-talk between mammary stem cells and the macrophageal niche. *Science*, *360*(6396). https://doi.org/10.1126/science.aan4153

Chan, C. J., Heisenberg, C. P., & Hiiragi, T. (2017). Coordination of Morphogenesis and Cell-Fate Specification in Development. *Current Biology*, *27*(18), R1024–R1035. https://doi.org/10.1016/j.cub.2017.07.010

Chandramouli, A., Simundza, J., Pinderhughes, A., Hiremath, M., Droguett, G., Frendewey, D., & Cowin, P. (2013). Ltbp1L is focally induced in embryonic mammary mesenchyme, demarcates the ductal luminal lineage and is upregulated during involution. *Breast Cancer Res*, *15*(6). https://doi.org/10.1186/bcr3578

Chang, T., Kunasegaran, K., Tarulli, G., de Silva, D., Voorhoeve, P., & Pietersen, A. (2014). New insights into lineage restriction of mammary gland epithelium using parity-identified mammary epithelial cells. *Breast Cancer Research*, *16*(1). https://doi.org/10.1186/bcr3593

Chatzeli, L., Gaete, M., & Tucker, A. S. (2017). Fgf10 and Sox9 are essential for the establishment of distal progenitor cells during mouse salivary gland development. *Development (Cambridge)*, *144*(12), 2294–2305. https://doi.org/10.1242/dev.146019

Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. *Science*, *348*(6233). https://doi.org/10.1126/science.aaa6090

Choi, N., Zhang, B., Zhang, L., Ittmann, M., & Xin, L. (2012). Adult Murine Prostate Basal and Luminal Cells Are Self-Sustained Lineages that Can Both Serve as Targets for Prostate Cancer Initiation. *Cancer Cell*, *21*(2), 253–265. https://doi.org/10.1016/j.ccr.2012.01.005

Chu, E. Y., Hens, J., Andl, T., Kairo, A., Yamaguchi, T. P., Brisken, C., Glick, A., Wysolmerski, J. J., & Millar, S. E. (2004). Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis. *Development*, *131*(19), 4819–4829. https://doi.org/10.1242/dev.01347

Chung, C.-Y., Ma, Z., Dravis, C., Preissl, S., Poirion, O., Luna, G., Hou, X., Giraddi, R. R., Ren, B., & Wahl, G. M. (2019). Single-Cell Chromatin Analysis of Mammary Gland Development Reveals Cell-State Transcriptional Regulators and Lineage Relationships. *Cell Rep*, *29*(2), 495–510. https://doi.org/10.1016/j.celrep.2019.08.089

Clayton, E., Doupé, D. P., Klein, A. M., Winton, D. J., Simons, B. D., & Jones, P. H. (2007). A single type of progenitor cell maintains normal epidermis. *Nature*, *446*(7132), 185–189. https://doi.org/10.1038/nature05574

Cowin, P., & Wysolmerski, J. (2010). Molecular mechanisms guiding embryonic mammary gland development. *Cold Spring Harbor Perspectives in Biology*, *2*(6). https://doi.org/10.1101/cshperspect.a003251

Cunha, G., Young, P., Christov, K., Guzman, R., Nandi, S., Talamantes, F., & Thordarson, G. (1995). Mammary phenotypic expression induced in epidermal cells by embryonic mammary mesenchyme. *Acta Anat (Basel)*, *152*(3), 195–204. https://doi.org/10.1159/000147698

Davis, F. M., Lloyd-Lewis, B., Harris, O. B., Kozar, S., Winton, D. J., Muresan, L., & Watson, C. J. (2016). Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. *Nature Communications*, *7*(13053). https://doi.org/10.1038/ncomms13053

de Visser, K., Ciampricotti, M., Michalak, E., Tan, D., Speksnijder, E., Hau, C., Clevers, H., Barker, N., & Jonkers, J. (2012). Developmental stage-specific contribution of LGR5(+) cells to basal and luminal epithelial lineages in the postnatal mammary gland. *J Pathol*, *228*(3), 300–309. https://doi.org/10.1002/path.4096

DeOme, K., Faulkin, L., Bern, H., & Blair, P. (1959). Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res.*, *19*(5), 515–520.

Deschene, E., Myung, P., Rompolas, P., Zito, G., Sun, T., Taketo, M., Saotome, I., & Greco, V. (2014). β-Catenin activation regulates tissue growth non-cell autonomously in the hair stem cell niche. *Science*, *343*(6177), 1353–1356. https://doi.org/10.1126/science.1248373

dos Santos, C. O., Rebbeck, C., Rozhkova, E., Valentine, A., Samuels, A., Kadiri, L. R., Osten, P., Harris, E. Y., Uren, P. J., Smith, A. D., & Hannon, G. J. (2013). Molecular hierarchy of mammary differentiation yields refined markers of mammary stem cells. *Proceedings of the*

National Academy of Sciences of the United States of America, *110*(18), 7123–7130. https://doi.org/10.1073/pnas.1303919110

Dravis, C., Chung, C. Y., Lytle, N. K., Herrera-Valdez, J., Luna, G., Trejo, C. L., Reya, T., & Wahl, G. M. (2018). Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. *Cancer Cell*, *34*(3), 466-482.e6. https://doi.org/10.1016/j.ccell.2018.08.001

Durnberger, H., & Kratochwii, K. (1980). Specificity of Tissue Interaction and Origin of Mesenchymal Cells in the Androgen Response of the Embryonic Mammary Gland. *Cell*, *19*(2), 465–471. https://doi.org/10.1016/0092-8674(80)90521-8

Eng, C. H. L., Lawson, M., Zhu, Q., Dries, R., Koulena, N., Takei, Y., Yun, J., Cronin, C., Karp, C., Yuan, G. C., & Cai, L. (2019). Transcriptome-scale super-resolved imaging in tissues by RNA seqFISH+. *Nature*, *568*(7751), 235–239. https://doi.org/10.1038/s41586-019-1049-y

Ershov, D., Phan, M. S., Pylvänäinen, J. W., Rigaud, S. U., le Blanc, L., Charles-Orszag, A., Conway, J. R. W., Laine, R. F., Roy, N. H., Bonazzi, D., Duménil, G., Jacquemet, G., & Tinevez, J. Y. (2022). TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. *Nature Methods*, *19*(7), 829–832. https://doi.org/10.1038/s41592-022-01507-1

Fankhaenel, M., Sadat Golestan Hashemi, F., Mosa Hosawi, M., Mourao, L., Skipp, P., Morin, X., LGJ Scheele, C., & Elias, S. (2021). Annexin A1 is a polarity cue that directs planar mitotic spindle orientation during mammalian epithelial morphogenesis. *BioRxiv*, *454117*. https://doi.org/10.1101/2021.07.28.454117

Feng, Y., Manka, D., Wagner, K.-U., & Khan, S. A. (2007). Estrogen receptor-expression in the mammary epithelium is required for ductal and alveolar morphogenesis in mice. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(37), 14718–14723. https://doi.org/10.1073/pnas.0706933104

Fre, S., Hannezo, E., Sale, S., Huyghe, M., Lafkas, D., Kissel, H., Louvi, A., Greve, J., Louvard, D., & Artavanis-Tsakonas, S. (2011). Notch lineages and activity in intestinal stem cells determined by a new set of knock-in mice. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0025785

Fu, N. Y., Nolan, E., Lindeman, G. J., & Visvader, J. E. (2020). Stem cells and the differentiation hierarchy in mammary gland development. *Physiological Reviews*, *100*(2), 489–523. https://doi.org/10.1152/physrev.00040.2018

Fu, N. Y., Rios, A. C., Pal, B., Law, C. W., Jamieson, P., Liu, R., Vaillant, F., Jackling, F., Liu, K. H., Smyth, G. K., Lindeman, G. J., Ritchie, M. E., & Visvader, J. E. (2017). Identification of quiescent and spatially restricted mammary stem cells that are hormone responsive. *Nature Cell Biology*, *19*(3), 164–176. https://doi.org/10.1038/ncb3471

Giraddi, R. R., Chung, C.-Y., Heinz, R. E., Balcioglu, O., Novotny, M., Trejo, C. L., Dravis, C., Hagos, B. M., Mehrabad, E. M., Rodewald, L. W., Hwang, J. Y., Fan, C., Lasken, R., Varley, K. E., Perou, C. M., Wahl, G. M., & Spike, B. T. (2018). Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in Early Mammary Gland Development. *Cell Rep*, *24*(6), 1653–1666. https://doi.org/10.1016/j.celrep.2018.07.025

Goodwin, K., Jaslove, J. M., Tao, H., Zhu, M., Hopyan, S., & Nelson, C. M. (2022). Patterning the embryonic pulmonary mesenchyme. *IScience*, *25*(3). https://doi.org/10.1016/j.isci.2022.103838

Goodwin, K., Mao, S., Guyomar, T., Miller, E., Radisky, D. C., Košmrlj, A., & Nelson, C. M. (2019). Smooth muscle differentiation shapes domain branches during mouse lung development. *Development (Cambridge)*, *146*(22). https://doi.org/10.1242/dev.181172

Goodwin, K., & Nelson, C. M. (2020). Branching morphogenesis. *Development*, *147*(10). https://doi.org/10.1242/dev.184499

Gritli-Linde, A., Hallberg, K., Harfe, B. D., Reyahi, A., Kannius-Janson, M., Nilsson, J., Cobourne, M. T., Sharpe, P. T., McMahon, A. P., & Linde, A. (2007). Abnormal Hair Development and Apparent Follicular Transformation to Mammary Gland in the Absence of Hedgehog Signaling. *Developmental Cell*, *12*(1), 99–112. https://doi.org/10.1016/j.devcel.2006.12.006

Gu, B., Sun, P., Yuan, Y., Moraes, R. C., Li, A., Teng, A., Agrawal, A., Rhéaume, C., Bilanchone, V., Veltmaat, J. M., Takemaru, K. I., Millar, S., Lee, E. Y. H. P., Lewis, M. T., Li, B., & Dai, X. (2009). Pygo2 expands mammary progenitor cells by facilitating histone H3 K4 methylation. *Journal of Cell Biology*, *185*(5), 811–826. https://doi.org/10.1083/jcb.200810133

Hafemeister, C., & Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1874-1

Hanahan, D. (2022). Hallmarks of Cancer: New Dimensions. *Cancer Discovery*, *12*(1), 31–46. https://doi.org/10.1158/2159-8290.CD-21-1059

Hannezo, E., Scheele, C. L. G. J., Moad, M., Drogo, N., Heer, R., Sampogna, Rosemary. V., van Rheenen, J., & Simons, B. D. (2017). A Unifying Theory of Branching Morphogenesis. *Cell*, *171*(1), 242-255.e27. https://doi.org/10.1016/j.cell.2017.08.026

Hannezo, E., & Simons, B. D. (2019). Multiscale dynamics of branching morphogenesis. *Curr Opin Cell Biol*, *60*, 99–105. https://doi.org/10.1016/j.ceb.2019.04.008

Hao, Y., Hao, S., Andersen-Nissen, E., Mauck, W. M., Zheng, S., Butler, A., Lee, M. J., Wilk, A. J., Darby, C., Zager, M., Hoffman, P., Stoeckius, M., Papalexi, E., Mimitou, E. P., Jain, J., Srivastava, A., Stuart, T., Fleming, L. M., Yeung, B., … Satija, R. (2021). Integrated analysis of multimodal single-cell data. *Cell*, *184*(13), 3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048

Harada, N., Tamai, Y., Ishikawa, T.-O., Sauer, B., Takaku, K., Oshima, M., & Taketo, M. M. (1999). Intestinal polyposis in mice with a dominant stable mutation of the β-catenin gene. *The EMBO Journal*, *18*(21), 5931–5942. https://doi.org/10.1093/emboj/18.21.5931

Hatsell, S. J., & Cowin, P. (2006). Gli3-mediated repression of Hedgehog targets is required for normal mammary development. *Development*, *133*(18), 3661–3670. https://doi.org/10.1242/dev.02542

Heckman, B. M., Chakravarty, G., Vargo-Gogola, T., Gonzales-Rimbau, M., Hadsell, D. L., Lee, Adrian. v., Settleman, J., & Rosen, J. M. (2007). Crosstalk between the p190-B RhoGAP and IGF signaling pathways is required for embryonic mammary bud development. *Developmental Biology*, *309*(1), 137–149. https://doi.org/10.1016/j.ydbio.2007.07.002

Hein, S. M., Haricharan, S., Johnston, A. N., Toneff, M. J., Reddy, J. P., Dong, J., Bu, W., & Li, Y. (2016). Luminal epithelial cells within the mammary gland can produce basal cells upon oncogenic stress. *Oncogene*, *35*(11), 1461–1467. https://doi.org/10.1038/onc.2015.206

Horsley, V., O'Carroll, D., Tooze, R., Ohinata, Y., Saitou, M., Obukhanych, T., Nussenzweig, M., Tarakhovsky, A., & Fuchs, E. (2006). Blimp1 Defines a Progenitor Population that Governs Cellular Input to the Sebaceous Gland. *Cell*, *126*(3), 597–609. https://doi.org/10.1016/j.cell.2006.06.048

Howard, B., Panchal, H., McCarthy, A., & Ashworth, A. (2005). Identification of the scaramanga gene implicates Neuregulin3 in mammary gland specification. *Genes and Development*, *19*(17), 2078–2090. https://doi.org/10.1101/gad.338505

Huebner, R. J., & Ewald, A. J. (2014). Cellular foundations of mammary tubulogenesis. *Seminars in Cell and Developmental Biology*, *31*, 124–131. https://doi.org/10.1016/j.semcdb.2014.04.019

Huff, J. (2015). The Airyscan detector from ZEISS: confocal imaging with improved signal-to-noise ratio and super-resolution. *Nature Methods*, *12*(12), i–ii. https://doi.org/10.1038/nmeth.f.388

Indra, A. K., Warot, X., Brocard, J., Bornert, J.-M., Xiao, J.-H., Chambon, P., & Metzger, D. (1999). Temporally-controlled site-specific mutagenesis in the basal layer of the epidermis: comparison of the recombinase activity of the tamoxifen-inducible Cre-ER T and Cre-ER T2 recombinases. *Nucleic Acids Research*, *27*(22), 4324–4327. https://doi.org/10.1093/nar/27.22.4324

Inman, J. L., Robertson, C., Mott, J. D., & Bissell, M. J. (2015). Mammary gland development: Cell fate specification, stem cells and the microenvironment. *Development (Cambridge)*, *142*(6), 1028–1042. https://doi.org/10.1242/dev.087643

Jerome-Majewska, L. A., Jenkins, G. P., Ernstoff, E., Zindy, F., Sherr, C. J., & Papaioannou, V. E. (2005). Tbx3, the ulnar-mammary syndrome gene, and Tbx2 interact in mammary gland development through a p19Arf/p53-independent pathway. *Developmental Dynamics*, *234*(4), 922–933. https://doi.org/10.1002/dvdy.20575

Joshi, P. A., Waterhouse, P. D., Kannan, N., Narala, S., Fang, H., di Grappa, M. A., Jackson, H. W., Penninger, J. M., Eaves, C., & Khokha, R. (2015). RANK Signaling Amplifies WNT-Responsive Mammary Progenitors through R-SPONDIN1. *Stem Cell Reports*, *5*(1), 31–44. https://doi.org/10.1016/j.stemcr.2015.05.012

Kapoor, V., & Carabaña, C. (2021). Cell Tracking in 3D using deep learning segmentations. *Proceedings of the 20th Python in Science Conference*. https://doi.org/10.25080/majora-1b6fd038-02b

Kaufman, C. K., Mosimann, C., Fan, Z. P., Yang, S., Thomas, A. J., Ablain, J., Tan, J. L., Fogley, R. D., van Rooijen, E., Hagedorn, E. J., Ciarlo, C., White, R. M., Matos, D. A., Puller, A. C., Santoriello, C., Liao, E. C., Young, R. A., & Zon, L. I. (2016). A zebrafish melanoma model reveals emergence of neural crest identity during melanoma initiation. *Science*, *351*(6272). https://doi.org/10.1126/science.aad2197

Kendrick, H., Regan, J. L., Magnay, F. A., Grigoriadis, A., Mitsopoulos, C., Zvelebil, M., & Smalley, M. J. (2008). Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. *BMC Genomics*, *9*, 591. https://doi.org/10.1186/1471-2164-9-591

Kordon, E., & Smith, G. (1998). An entire functional mammary gland may comprise the progeny from a single cell. *Development*, *125*(10), 1921–1930. https://doi.org/10.1242/dev.125.10.1921

Koren, S., Reavie, L., Couto, J. P., de Silva, D., Stadler, M. B., Roloff, T., Britschgi, A., Eichlisberger, T., Kohler, H., Aina, O., Cardiff, R. D., & Bentires-Alj, M. (2015). PIK3CAH1047R induces multipotency and multi-lineage mammary tumours. *Nature*, *525*(7567), 114–118. https://doi.org/10.1038/nature14669

Kouros-Mehr, H., & Werb, Z. (2006). Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis. *Dev Dyn*, *235*(12), 3404–3412. https://doi.org/10.1002/dvdy.20978

Kratochwil, K. (1969). Organ Specificity in Mesenchymal Induction Demonstrated in the Embryonic Development of the Mammary Gland of the Mouse. *Developmental Biology*, *20*, 46–71. https://doi.org/10.1016/0012-1606(69)90004-9

Kretzschmar, K., & Watt, F. M. (2012). Lineage tracing. In *Cell* (Vol. 148, Issues 1–2, pp. 33–45). Elsevier B.V. https://doi.org/10.1016/j.cell.2012.01.002

Lafkas, D., Rodilla, V., Huyghe, M., Mourao, L., Kiaris, H., & Fre, S. (2013). Notch3 marks clonogenic mammary luminal progenitor cells in vivo. *Journal of Cell Biology*, *203*(1), 47–56. https://doi.org/10.1083/jcb.201307046

LaMarca, H. L., Visbal, A. P., Creighton, C. J., Liu, H., Zhang, Y., Behbod, F., & Rosen, J. M. (2010). CCAAT/enhancer binding protein beta regulates stem cell activity and specifies luminal cell fate in the mammary gland. *Stem Cells*, *28*(3), 535–544. https://doi.org/10.1002/stem.297

Lan, Q., & Mikkola, M. L. (2020). Protocol: Adeno-Associated Virus-Mediated Gene Transfer in Ex Vivo Cultured Embryonic Mammary Gland. *Journal of Mammary Gland Biology and Neoplasia*, *25*(4), 409–416. https://doi.org/10.1007/s10911-020-09461-4

Lee, M. Y., Racine, V., Jagadpramana, P., Sun, L., Yu, W., Du, T., Spencer-Dene, B., Rubin, N., Le, L., Ndiaye, D., Bellusci, S., Kratochwil, K., & Veltmaat, J. M. (2011). Ectodermal influx and cell hypertrophy provide early growth for all murine mammary rudiments, and are differentially regulated among them by Gli3. *PLoS ONE*, *6*(10). https://doi.org/10.1371/journal.pone.0026242

Li, C. M. C., Shapiro, H., Tsiobikas, C., Selfors, L. M., Chen, H., Rosenbluth, J., Moore, K., Gupta, K. P., Gray, G. K., Oren, Y., Steinbaugh, M. J., Guerriero, J. L., Pinello, L., Regev, A., & Brugge, J. S. (2020). Aging-Associated Alterations in Mammary Epithelia and Stroma Revealed by Single-Cell RNA Sequencing. *Cell Reports*, *33*(13). https://doi.org/10.1016/j.celrep.2020.108566

Li, J., Economou, A., Vacca, B., & Green, J. (2020). Epithelial invagination by a vertical telescoping cell movement in mammalian salivary glands and teeth. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-16247-z

Liao, J., Lu, X., Shao, X., Zhu, L., & Fan, X. (2021). Uncovering an Organ's Molecular Architecture at Single-Cell Resolution by Spatially Resolved Transcriptomics. *Trends in Biotechnology*, *39*(1), 43–58. https://doi.org/10.1016/j.tibtech.2020.05.006

Lilja, A. M., Rodilla, V., Huyghe, M., Hannezo, E., Landragin, C., Renaud, O., Leroy, O., Rulands, S., Simons, B. D., & Fre, S. (2018). Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells that retain long-term plasticity in the embryonic mammary gland. *Nature Cell Biology*, *20*(6), 677–687. https://doi.org/10.1038/s41556-018-0108-1

Lindfors, P. H., Voutilainen, M., & Mikkola, M. L. (2013). Ectodysplasin/NF-κB signaling in embryonic mammary gland development. *Journal of Mammary Gland Biology and Neoplasia*, *18*(2), 165–169. https://doi.org/10.1007/s10911-013-9277-5

Lindvall, C., Evans, N. C., Zylstra, C. R., Li, Y., Alexander, C. M., & Williams, B. O. (2006). The Wnt signaling receptor Lrp5 is required for mammary ductal stem cell activity and Wnt1-induced tumorigenesis. *Journal of Biological Chemistry*, *281*(46), 35081–35087. https://doi.org/10.1074/jbc.M607571200

Lindvall, C., Zylstra, C. R., Evans, N., West, R. A., Dykema, K., Furge, K. A., & Williams, B. O. (2009). The Wnt co-receptor Lrp6 is required for normal mouse mammary gland development. *PLoS ONE*, *4*(6). https://doi.org/10.1371/journal.pone.0005813

Linkert, M., Rueden, C. T., Allan, C., Burel, J. M., Moore, W., Patterson, A., Loranger, B., Moore, J., Neves, C., MacDonald, D., Tarkowska, A., Sticco, C., Hill, E., Rossner, M., Eliceiri, K. W., & Swedlow, J. R. (2010). Metadata matters: Access to image data in the real world. *Journal of Cell Biology*, *189*(5), 777–782. https://doi.org/10.1083/jcb.201004104

Lloyd-Lewis, B., Davis, F. M., Harris, O. B., Hitchcock, J. R., & Watson, C. J. (2018). Neutral lineage tracing of proliferative embryonic and adultmammary stem/progenitor cells. *Development (Cambridge)*, *145*(14). https://doi.org/10.1242/dev.164079

Lloyd-Lewis, B., Gobbo, F., Perkins, M., Jacquemin, G., Huyghe, M., Faraldo, M. M., & Fre, S. (2022). In vivo imaging of mammary epithelial cell dynamics in response to lineage-biased Wnt/β-catenin activation. *Cell Reports*, *38*(10). https://doi.org/10.1016/j.celrep.2022.110461

Lloyd-Lewis, B., Harris, O. B., Watson, C. J., & Davis, F. M. (2017). Mammary Stem Cells: Premise, Properties, and Perspectives. *Trends in Cell Biology*, *27*(8), 556–567. https://doi.org/10.1016/j.tcb.2017.04.001

Lourenço, A. R., Roukens, M. G., Seinstra, D., Frederiks, C. L., Pals, C. E., Vervoort, S. J., Margarido, A. S., van Rheenen, J., & Coffer, P. J. (2020). C/EBPα is crucial determinant of epithelial maintenance by preventing epithelial-to-mesenchymal transition. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-14556-x

Lu, C. P., Polak, L., Rocha, A. S., Pasolli, H. A., Chen, S. C., Sharma, N., Blanpain, C., & Fuchs, E. (2012). Identification of stem cell populations in sweat glands and ducts reveals roles in homeostasis and wound repair. *Cell*, *150*(1), 136–150. https://doi.org/10.1016/j.cell.2012.04.045

Lu, P., Ewald, A. J., Martin, G. R., & Werb, Z. (2008). Genetic mosaic analysis reveals FGF receptor 2 function in terminal end buds during mammary gland branching morphogenesis. *Developmental Biology*, *321*(1), 77–87. https://doi.org/10.1016/j.ydbio.2008.06.005

Macias, H., & Hinck, L. (2012). Mammary gland development. *Wiley Interdisciplinary Reviews: Developmental Biology*, *1*(4), 533–557. https://doi.org/10.1002/wdev.35

Mailleux, A., Spencer-Dene, B., Dillon, C., Ndiaye, D., Savona-Baron, C., Itoh, N., Kato, S., Dickson, C., Thiery, J., & Bellusci, S. (2002). Role of FGF10/FGFR2b signaling during mammary gland development in the mouse embryo. *Development*, *129*(1), 53–60. https://doi.org/10.1242/dev.129.1.53

Mallepell, S., Krust, A. E., Chambon, P., & Brisken, C. (2006). Paracrine signaling through the epithelial estrogen receptor is required for proliferation and morphogenesis in the mammary gland. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(7), 2196–2201. https://doi.org/10.1073/pnas.0510974103

McCulley, D., Wienhold, M., & Sun, X. (2015). The pulmonary mesenchyme directs lung development. *Current Opinion in Genetics and Development*, *32*, 98–105. https://doi.org/10.1016/j.gde.2015.01.011

McInnes, L., Healy, J., & Melville, J. (2018). *UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction*. http://arxiv.org/abs/1802.03426

Merrick, D., Sakers, A., Irgebay, Z., Okada, C., Calvert, C., Morley, M. P., Percec, I., & Seale, P. (2019). Identification of a mesenchymal progenitor cell hierarchy in adipose tissue. *Science*, *364*(6438). https://doi.org/10.1126/science.aav2501

Metzger, R. J., Klein, O. D., Martin, G. R., & Krasnow, M. A. (2008). The branching programme of mouse lung development. *Nature*, *453*(7196), 745–750. https://doi.org/10.1038/nature07005

Miller, D. J., & Fort, P. E. (2018). Heat shock proteins regulatory role in neurodevelopment. *Frontiers in Neuroscience*, *12*(821). https://doi.org/10.3389/fnins.2018.00821

Morita, R., Sanzen, N., Sasaki, H., Hayashi, T., Umeda, M., Yoshimura, M., Yamamoto, T., Shibata, T., Abe, T., Kiyonari, H., Furuta, Y., Nikaido, I., & Fujiwara, H. (2021). Tracing the origin of hair follicle stem cells. *Nature*, *594*(7864), 547–552. https://doi.org/10.1038/s41586-021-03638-5

Morrisey, E. E., & Hogan, B. L. M. (2010). Preparing for the First Breath: Genetic and Cellular Mechanisms in Lung Development. *Developmental Cell*, *18*(1), 8–23. https://doi.org/10.1016/j.devcel.2009.12.010

Murtaugh, L. C., Stanger, B. Z., Kwan, K. M., & Melton, D. A. (2003). Notch signaling controls multiple steps of pancreatic differentiation. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(25), 14920–14925. https://doi.org/10.1073/pnas.2436557100

Muschler, J., & Streuli, C. H. (2010). Cell-matrix interactions in mammary gland development and breast cancer. *Cold Spring Harbor Perspectives in Biology*, *2*(10). https://doi.org/10.1101/cshperspect.a003202

Muzumdar, M. D., Tasic, B., Miyamichi, K., Li, N., & Luo, L. (2007). A global double-fluorescent cre reporter mouse. *Genesis*, *45*(9), 593–605. https://doi.org/10.1002/dvg.20335

Myllymäki, S. M., & Mikkola, M. L. (2019). Inductive signals in branching morphogenesis – lessons from mammary and salivary glands. *Current Opinion in Cell Biology*, *61*, 72–78. https://doi.org/10.1016/j.ceb.2019.07.001

Nerger, B. A., Jaslove, J. M., Elashal, H. E., Mao, S., Košmrlj, A., Link, A. J., & Nelson, C. M. (2021). Local accumulation of extracellular matrix regulates global morphogenetic patterning in the developing mammary gland. *Current Biology*, *31*(9), 1903-1917.e6. https://doi.org/10.1016/j.cub.2021.02.015

Packard, A., Georgas, K., Michos, O., Riccio, P., Cebrian, C., Combes, A. N., Ju, A., Ferrer-Vaquer, A., Hadjantonakis, A. K., Zong, H., Little, M. H., & Costantini, F. (2013). Luminal Mitosis Drives Epithelial Cell Dispersal within the Branching Ureteric Bud. *Developmental Cell*, *27*(3), 319–330. https://doi.org/10.1016/j.devcel.2013.09.001

Pal, B., Chen, Y., Milevskiy, M. J. G., Vaillant, F., Prokopuk, L., Dawson, C. A., Capaldo, B. D., Song, X., Jackling, F., Timpson, P., Lindeman, G. J., Smyth, G. K., & Visvader, J. E. (2021). Single cell transcriptome atlas of mouse mammary epithelial cells across development. *Breast Cancer Research*, *23*(1). https://doi.org/10.1186/s13058-021-01445-4

Pal, B., Chen, Y., Vaillant, F., Jamieson, P., Gordon, L., Rios, A. C., Wilcox, S., Fu, N., Liu, K. H., Jackling, F. C., Davis, M. J., Lindeman, G. J., Smyth, G. K., & Visvader, J. E. (2017). Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nature Communications*, *8*(1). https://doi.org/10.1038/s41467-017-01560-x

Panchal, H., Wansbury, O., Parry, S., Ashworth, A., & Howard, B. (2007). Neuregulin3 alters cell fate in the epidermis and mammary gland. *BMC Developmental Biology*, *7*, 105. https://doi.org/10.1186/1471-213X-7-105

Panousopoulou, E., & Green, J. B. A. (2016). Invagination of Ectodermal Placodes Is Driven by Cell Intercalation-Mediated Contraction of the Suprabasal Tissue Canopy. *PLoS Biology*, *14*(3). https://doi.org/10.1371/journal.pbio.1002405

Perié, L., Hodgkin, P. D., Naik, S. H., Schumacher, T. N., deBoer, R. J., & Duffy, K. R. (2014). Determining Lineage Pathways from Cellular Barcoding Experiments. *Cell Reports*, *6*(4), 617–624. https://doi.org/10.1016/j.celrep.2014.01.016

Phillips, S., Prat, A., Sedic, M., Proia, T., Wronski, A., Mazumdar, S., Skibinski, A., Shirley, S. H., Perou, C. M., Gill, G., Gupta, P. B., & Kuperwasser, C. (2014). Cell-state transitions regulated by SLUG are critical for tissue regeneration and tumor initiation. *Stem Cell Reports*, *2*(5), 633–647. https://doi.org/10.1016/j.stemcr.2014.03.008

Platt, R. J., Chen, S., Zhou, Y., Yim, M. J., Swiech, L., Kempton, H. R., Dahlman, J. E., Parnas, O., Eisenhaure, T. M., Jovanovic, M., Graham, D. B., Jhunjhunwala, S., Heidenreich, M., Xavier, R. J., Langer, R., Anderson, D. G., Hacohen, N., Regev, A., Feng, G., … Zhang, F. (2014). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. *Cell*, *159*(2), 440–455. https://doi.org/10.1016/j.cell.2014.09.014

Prater, M. D., Petit, V., Alasdair Russell, I., Giraddi, R. R., Shehata, M., Menon, S., Schulte, R., Kalajzic, I., Rath, N., Olson, M. F., Metzger, D., Faraldo, M. M., Deugnier, M. A., Glukhova,

M. A., & Stingl, J. (2014). Mammary stem cells have myoepithelial cell properties. *Nature Cell Biology*, *16*(10), 942–950. https://doi.org/10.1038/ncb3025

Regan, J. L., Kendrick, H., Magnay, F. A., Vafaizadeh, V., Groner, B., & Smalley, M. J. (2012). C-Kit is required for growth and survival of the cells of origin of Brca1-mutation-associated breast cancer. *Oncogene*, *31*(7), 869–883. https://doi.org/10.1038/onc.2011.289

Rios, A. C., Fu, N. Y., Lindeman, G. J., & Visvader, J. E. (2014). In situ identification of bipotent stem cells in the mammary gland. *Nature*, *506*(7488), 322–327. https://doi.org/10.1038/nature12948

Robinson, G. W. (2007). Cooperation of signalling pathways in embryonic mammary gland development. *Nature Reviews Genetics*, *8*(12), 963–972. https://doi.org/10.1038/nrg2227

Rocchi, C., Barazzuol, L., & Coppes, R. P. (2021). The evolving definition of salivary gland stem cells. *Npj Regenerative Medicine*, *6*(1). https://doi.org/10.1038/s41536-020-00115-x

Rock, J. R., & Hogan, B. L. M. (2011). Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annual Review of Cell and Developmental Biology*, *27*, 493–512. https://doi.org/10.1146/annurev-cellbio-100109-104040

Rodilla, V., Dasti, A., Huyghe, M., Lafkas, D., Laurent, C., Reyal, F., & Fre, S. (2015). Luminal Progenitors Restrict Their Lineage Potential during Mammary Gland Development. *PLoS Biol*, *13*(2). https://doi.org/10.1371/journal.pbio.1002069

Rodilla, V., & Fre, S. (2018). Cellular Plasticity of Mammary Epithelial Cells Underlies Heterogeneity of Breast Cancer. *Biomedicines*, *6*(4). https://doi.org/10.3390/biomedicines6040103

Rodilla, V., & Fre, S. (2022). Tracing Methods to Study Mammary Epithelial Hierarchies In Vivo. In Vivanco MM (Ed.), *Methods Mol Biol* (Second, Vol. 2471, pp. 141–157). https://doi.org/10.1007/978-1-0716-2193-6_7

Ronneberger, O., Fischer, P., & Brox, T. (2015). *U-Net: Convolutional Networks for Biomedical Image Segmentation*. http://arxiv.org/abs/1505.04597

Saghafinia, S., Homicsko, K., di Domenico, A., Wullschleger, S., Perren, A., Marinoni, I., Ciriello, G., Michael, I. P., & Hanahan, D. (2021). Cancer cells retrace a stepwise differentiation program during malignant progression. *Cancer Discovery*, *11*(10), 2638–2657. https://doi.org/10.1158/2159-8290.CD-20-1637

Sahai, E., Astsaturov, I., Cukierman, E., DeNardo, D. G., Egeblad, M., Evans, R. M., Fearon, D., Greten, F. R., Hingorani, S. R., Hunter, T., Hynes, R. O., Jain, R. K., Janowitz, T., Jorgensen, C., Kimmelman, A. C., Kolonin, M. G., Maki, R. G., Powers, R. S., Puré, E., … Werb, Z. (2020). A framework for advancing our understanding of cancer-associated fibroblasts. *Nature Reviews Cancer*, *20*(3), 174–186. https://doi.org/10.1038/s41568-019-0238-1

Sakakura, T., Sakagami, Y., & Nishizuka, Y. (1982). Dual Origin of Mesenchymal Tissues Participating in Mouse Mammary Gland Embryogenesis. *Developmental Biology*, *91*, 202–207. https://doi.org/10.1016/0012-1606(82)90024-0

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., Kashiwagi, S., Fukami, K., Miyata, T., Miyoshi, H., Imamura, T., Ogawa, M., Masai, H., & Miyawaki, A. (2008). Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell*, *132*(3), 487–498. https://doi.org/10.1016/j.cell.2007.12.033

Šale, S., Lafkas, D., & Artavanis-Tsakonas, S. (2013). Notch2 genetic fate mapping reveals two previously unrecognized mammary epithelial lineages. *Nature Cell Biology*, *15*(5), 451–460. https://doi.org/10.1038/ncb2725

Scheele, C. L. G. J., Hannezo, E., Muraro, M. J., Zomer, A., Langedijk, N. S. M., van Oudenaarden, A., Simons, B. D., & van Rheenen, J. (2017). Identity and dynamics of mammary stem cells during branching morphogenesis. *Nature*, *542*(7641), 313–317. https://doi.org/10.1038/nature21046

Schmidt, U., Weigert, M., Broaddus, C., & Myers, G. (2018). Cell Detection with Star-convex Polygons. *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. https://doi.org/10.1007/978-3-030-00934-2_30

Sengez, B., Aygün, I., Shehwana, H., Toyran, N., Tercan Avci, S., Konu, O., Stemmler, M., & Alotaibi, H. (2019). The Transcription Factor Elf3 Is Essential for a Successful Mesenchymal to Epithelial Transition. *Cells*, *8*(8), 858. https://doi.org/10.3390/cells8080858

Sennett, R., & Rendl, M. (2012). Mesenchymal-epithelial interactions during hair follicle morphogenesis and cycling. *Seminars in Cell and Developmental Biology*, *23*(8), 917–927. https://doi.org/10.1016/j.semcdb.2012.08.011

Seong, J., Kim, N. S., Kim, J. A., Lee, W., Seo, J. Y., Yum, M. K., Kim, J. H., Park, I., Kang, J. S., Bae, S. H., Yun, C. H., & Kong, Y. Y. (2018). Side branching and luminal lineage commitment by ID2 in developing mammary glands. *Development (Cambridge)*, *145*(14). https://doi.org/10.1242/dev.165258

Shackleton, M., Vaillant, F., Simpson, K. J., Stingl, J., Smyth, G. K., Asselin-Labat, M. L., Wu, L., Lindeman, G. J., & Visvader, J. E. (2006). Generation of a functional mammary gland from a single stem cell. *Nature*, *439*(7072), 84–88. https://doi.org/10.1038/nature04372

Short, K. M., Combes, A. N., Lefevre, J., Ju, A. L., Georgas, K. M., Lamberton, T., Cairncross, O., Rumballe, B. A., McMahon, A. P., Hamilton, N. A., Smyth, I. M., & Little, M. H. (2014). Global quantification of tissue dynamics in the developing mouse kidney. *Developmental Cell*, *29*(2), 188–202. https://doi.org/10.1016/j.devcel.2014.02.017

Silberstein, G. B., & Daniel, C. W. (1987). Reversible inhibition of mammary gland growth by transforming growth factor-β. *Science*, *237*(4812), 291–293. https://doi.org/10.1126/science.3474783

Sjöberg, E., Augsten, M., Bergh, J., Jirström, K., & Östman, A. (2016). Expression of the chemokine CXCL14 in the tumour stroma is an independent marker of survival in breast cancer. *British Journal of Cancer*, *114*(10), 1117–1124. https://doi.org/10.1038/bjc.2016.104

Spengler, D., Villalba, M., Hoffmann, A., Pantaloni, C., Houssami, S., Bockaert, J., & Journot, L. (1997). Regulation of apoptosis and cell cycle arrest by Zac1, a novel zinc finger protein expressed in the pituitary gland and the brain. *The EMBO Journal*, *16*(10), 2814–2825. https://doi.org/10.1093/emboj/16.10.2814

Spike, B. T., Engle, D. D., Lin, J. C., Cheung, S. K., La, J., & Wahl, G. M. (2012). A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell*, *10*(2), 183–197. https://doi.org/10.1016/j.stem.2011.12.018

Spina, E., & Cowin, P. (2021). Embryonic mammary gland development. *Seminars in Cell and Developmental Biology*, *114*, 83–92. https://doi.org/10.1016/j.semcdb.2020.12.012

Sreekumar, A., Toneff, M. J., Toh, E., Roarty, K., Creighton, C. J., Belka, G. K., Lee, D. K., Xu, J., Chodosh, L. A., Richards, J. A. S., & Rosen, J. M. (2017). WNT-Mediated Regulation of FOXO1 Constitutes a Critical Axis Maintaining Pubertal Mammary Stem Cell Homeostasis. *Developmental Cell*, *43*(4), 436-448.e6. https://doi.org/10.1016/j.devcel.2017.10.007

Ståhl, P., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J., Magnusson, J., Giacomello, S., Asp, M., Westholm, J., Huss, M., Mollbrink, A., Linnarsson, S., Codeluppi, S., Borg, Å., Pontén, F., Costea, P., Sahlén, P., Mulder, J., Bergmann, O., … Frisén, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, *353*(6294), 78–82. https://doi.org/10.1126/science.aaf2403

Stevenson, A. J., Vanwalleghem, G., Stewart, T. A., Condon, N. D., Lloyd-Lewis, B., Marino, N., Putney, J. W., Scott, E. K., Ewing, A. D., & Davis, F. M. (2020). Multiscale imaging of basal cell dynamics in the functionally mature mammary gland. *Proceedings of the National Academy of Sciences of the United States of America*, *117*(43), 26822–26832. https://doi.org/10.1073/pnas.2016905117/-/DCSupplemental

Stewart, T. A., Hughes, K., Hume, D. A., & Davis, F. M. (2019). Developmental Stage-Specific Distribution of Macrophages in Mouse Mammary Gland. *Frontiers in Cell and Developmental Biology*, *7*, 250. https://doi.org/10.3389/fcell.2019.00250

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H. I., & Eaves, C. J. (2006). Purification and unique properties of mammary epithelial stem cells. *Nature*, *439*(7079), 993–997. https://doi.org/10.1038/nature04496

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W. M., Hao, Y., Stoeckius, M., Smibert, P., & Satija, R. (2019). Comprehensive Integration of Single-Cell Data. *Cell*, *177*(7), 1888-1902.e21. https://doi.org/10.1016/j.cell.2019.05.031

Sumbal, J., Belisova, D., & Koledova, Z. (2021). Fibroblasts: The grey eminence of mammary gland development. *Seminars in Cell and Developmental Biology*, *114*, 134–142. https://doi.org/10.1016/j.semcdb.2020.10.012

Sun, C., Berry, W. L., & Olson, L. E. (2017). PDGFRα controls the balance of stromal and adipogenic cells during adipose tissue organogenesis. *Development (Cambridge)*, *144*(1), 83–94. https://doi.org/10.1242/dev.135962

Tang, N., Marshall, W. F., Mcmahon, M., Metzger, R. J., & Martin, G. R. (2011). Control of Mitotic Spindle Angle by the RAS-Regulated ERK1/2 Pathway Determines Lung Tube Shape. *Science*, *333*(6040), 342–345. https://doi.org/10.1126/science.1204831

Tao, L., van Bragt, M. P. A., Laudadio, E., & Li, Z. (2014). Lineage tracing of mammary epithelial cells using cell-type-specific cre-expressing adenoviruses. *Stem Cell Reports*, *2*(6), 770–779. https://doi.org/10.1016/j.stemcr.2014.04.004

Tika, E., Ousset, M., Dannau, A., & Blanpain, C. (2019). Spatiotemporal regulation of multipotency during prostate development. *Development (Cambridge)*, *146*(20). https://doi.org/10.1242/dev.180224

Trela, E., Lan, Q., Myllymäki, S. M., Villeneuve, C., Lindström, R., Kumar, V., Wickström, S., & Mikkola, M. L. (2021). Cell influx and contractile actomyosin force drive mammary bud growth and invagination. *Journal of Cell Biology*, *220*(8). https://doi.org/10.1083/jcb.202008062

Tsang, S. M., Kim, H., Oliemuller, E., Newman, R., Boateng, N. A., Guppy, N., & Howard, B. A. (2021). Sox11 regulates mammary tumour-initiating and metastatic capacity in Brca1-deficient mouse mammary tumour cells. *DMM Disease Models and Mechanisms*, *14*(5). https://doi.org/10.1242/DMM.046037

Twigger, A. J., & Khaled, W. T. (2021). Mammary gland development from a single cell 'omics view. *Seminars in Cell and Developmental Biology*, *114*, 171–185. https://doi.org/10.1016/j.semcdb.2021.03.013

van Amerongen, R., Bowman, A. N., & Nusse, R. (2012). Developmental stage and time dictate the fate of Wnt/β-catenin- responsive stem cells in the mammary gland. *Cell Stem Cell*, *11*(3), 387–400. https://doi.org/10.1016/j.stem.2012.05.023

van Genderen, C., Okamura, R. M., Farinas, I., Quo, R.-G., Parslow, T. G., Bruhn, L., & Grosschedl, R. (1994). Development of several oreans that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. *Genes Dev.*, *8*(22), 2691–2703. https://doi.org/10.1101/gad.8.22.2691

van Keymeulen, A., & Blanpain, C. (2012). Tracing epithelial stem cells during development, homeostasis, and repair. *Journal of Cell Biology*, *197*(5), 575–584. https://doi.org/10.1083/jcb.201201041

van Keymeulen, A., Fioramonti, M., Centonze, A., Bouvencourt, G., Achouri, Y., & Blanpain, C. (2017). Lineage-Restricted Mammary Stem Cells Sustain the Development, Homeostasis, and Regeneration of the Estrogen Receptor Positive Lineage. *Cell Rep*, *20*(7), 1525–1532. https://doi.org/10.1016/j.celrep.2017.07.066

van Keymeulen, A., Lee, M. Y., Ousset, M., Brohée, S., Rorive, S., Giraddi, R. R., Wuidart, A., Bouvencourt, G., Dubois, C., Salmon, I., Sotiriou, C., Phillips, W. A., & Blanpain, C. (2015). Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. *Nature*, *525*(7567), 119–123. https://doi.org/10.1038/nature14665

van Keymeulen, A., Rocha, A. S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., Sharma, N., Dekoninck, S., & Blanpain, C. (2011). Distinct stem cells contribute to mammary gland development and maintenance. *Nature*, *479*(7372), 189–193. https://doi.org/10.1038/nature10573

Varner, V., & Nelson, C. (2014). Cellular and physical mechanisms of branching morphogenesis. *Development*, *141*(14), 2750–2759. https://doi.org/10.1242/dev.104794

Veltmaat, J. M. (2012). *Laboratory of mammary gland development*. http://www.veltmaatlab.net/research.html

Veltmaat, J. M., Relaix, F., Le, L. T., Kratochwil, K., Sala, F. G., van Veelen, W., Rice, R., Spencer-Dene, B., Mailleux, A. A., Rice, D. P., Thiery, J. P., & Bellusci, S. (2006). Gli3-mediated somitic Fgf10 expression gradients are required for the induction and patterning of mammary epithelium along the embryonic axes. *Development*, *133*(12), 2325–2335. https://doi.org/10.1242/dev.02394

Veltmaat, J. M., van Veelen, W., Thiery, J. P., & Bellusci, S. (2004). Identification of the Mammary Line in Mouse by Wnt10b Expression. *Developmental Dynamics*, *229*(2), 349–356. https://doi.org/10.1002/dvdy.10441

Vento-Tormo, R., Efremova, M., Botting, R. A., Turco, M. Y., Vento-Tormo, M., Meyer, K. B., Park, J. E., Stephenson, E., Polański, K., Goncalves, A., Gardner, L., Holmqvist, S., Henriksson, J., Zou, A., Sharkey, A. M., Millar, B., Innes, B., Wood, L., Wilbrey-Clark, A., … Teichmann, S. A. (2018). Single-cell reconstruction of the early maternal–fetal interface in humans. *Nature*, *563*(7731), 347–353. https://doi.org/10.1038/s41586-018-0698-6

Vickovic, S., Eraslan, G., Salmén, F., Klughammer, J., Stenbeck, L., Schapiro, D., Äijö, T., Bonneau, R., Bergenstråhle, L., Navarro, J., Gould, J., Griffin, G., Borg, Å., Ronaghi, M., Frisén, J., Lundeberg, J., Regev, A., & Ståhl, P. (2019). High-definition spatial transcriptomics for in situ tissue profiling. *Nature Methods*, *16*(10), 987–990. https://doi.org/10.1038/s41592-019-0548-y

Voutilainen, M., Lindfors, P. H., & Mikkola, M. L. (2013). Protocol: Ex vivo culture of mouse embryonic mammary buds. *Journal of Mammary Gland Biology and Neoplasia*, *18*(2), 239–245. https://doi.org/10.1007/s10911-013-9288-2

Wang, C., Christin, JR., Oktay, MH., & Guo, W. (2017). Lineage-Biased Stem Cells Maintain Estrogen-Receptor-Positive and -Negative Mouse Mammary Luminal Lineages. *Cell Reports*, *18*(12), 2825–2835. https://doi.org/10.1016/j.celrep.2017.02.071

Wang, D., Cai, C., Dong, X., Yu, Q. C., Zhang, X. O., Yang, L., & Zeng, Y. A. (2015). Identification of multipotent mammary stemcells by protein C receptor expression. *Nature*, *517*(7532), 81–84. https://doi.org/10.1038/nature13851

Wang, F., Flanagan, J., Su, N., Wang, L., Bui, S., Nielson, A., Wu, X., Vo, H., Ma, X., & Luo, Y. (2012). RNAscope: A novel in situ RNA analysis platform for formalin-fixed, paraffin-embedded tissues. *Journal of Molecular Diagnostics*, *14*(1), 22–29. https://doi.org/10.1016/j.jmoldx.2011.08.002

Wang, S., Matsumoto, K., Lish, S., Cartagena-Rivera, A., & Yamada, K. (2021). Budding epithelial morphogenesis driven by cell-matrix versus cell-cell adhesion. *Cell*, *184*(14), 3702-3716.e30. https://doi.org/10.1016/j.cell.2021.05.015

Wansbury, O., Mackay, A., Kogata, N., Mitsopoulos, C., Kendrick, H., Davidson, K., Ruhrberg, C., Reis-Filho, J. S., Smalley, M. J., Zvelebil, M., & Howard, B. A. (2011). Transcriptome analysis of embryonic mammary cells reveals insights into mammary lineage establishment. *Breast Cancer Research*, *13*(4). https://doi.org/10.1186/bcr2928

Watson, C. J. (2021). How should we define mammary stem cells? *Trends in Cell Biology*, *31*(8), 621–627. https://doi.org/10.1016/j.tcb.2021.03.012

Watson, C. J., & Khaled, W. T. (2008). Mammary development in the embryo and adult: A journey of morphogenesis and commitment. *Development*, *135*(6), 995–1003. https://doi.org/10.1242/dev.005439

Watson, C. J., & Khaled, W. T. (2020). Mammary development in the embryo and adult: New insights into the journey of morphogenesis and commitment. *Development*, *147*(22). https://doi.org/10.1242/dev.169862

Weigert, M., Schmidt, U., Haase, R., Sugawara, K., & Myers, G. (2020). Star-convex Polyhedra for 3D Object Detection and Segmentation in Microscopy. *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 3666–3673. https://doi.org/10.48550/arXiv.1908.03636

Wiseman, B. S., & Werb, Z. (2002). Stromal Effects on Mammary Gland Development and Breast Cancer. *Science*, *296*(5570), 1046–1049. https://doi.org/10.1126/science.1067431

Wuidart, A., Ousset, M., Rulands, S., Simons, B. D., van Keymeulen, A., & Blanpain, C. (2016). Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. *Genes Dev*, *30*(11), 1261–1277. https://doi.org/10.1101/gad.280057.116

Wuidart, A., Sifrim, A., Fioramonti, M., Matsumura, S., Brisebarre, A., Brown, D., Centonze, A., Dannau, A., Dubois, C., van Keymeulen, A., Voet, T., & Blanpain, C. (2018). Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nature Cell Biology*, *20*(6), 666–676. https://doi.org/10.1038/s41556-018-0095-2

Wysolmerski, J., Philbrick, W., Dunbar, M., Lanske, B., Kronenberg, H., & Broadus, A. (1998). Rescue of the parathyroid hormone-related protein knockout mouse demonstrates that parathyroid hormone-related protein is essential for mammary gland development. *Development*, *125*(7), 1285–1294. https://doi.org/10.1242/dev.125.7.1285

Yalcin-Ozuysal, Ö., Fiche, M., Guitierrez, M., Wagner, K. U., Raffoul, W., & Brisken, C. (2010). Antagonistic roles of Notch and p63 in controlling mammary epithelial cell fates. *Cell Death and Differentiation*, *17*(10), 1600–1612. https://doi.org/10.1038/cdd.2010.37

Youssef, K. K., Lapouge, G., Bouvrée, K., Rorive, S., Brohée, S., Appelstein, O., Larsimont, J. C., Sukumaran, V., van de Sande, B., Pucci, D., Dekoninck, S., Berthe, J. V., Aerts, S., Salmon, I., del Marmol, V., & Blanpain, C. (2012). Adult interfollicular tumour-initiating cells are reprogrammed into an embryonic hair follicle progenitor-like fate during basal cell carcinoma initiation. *Nature Cell Biology*, *14*(12), 1282–1294. https://doi.org/10.1038/ncb2628

Yu, Q. C., Verheyen, E. M., & Zeng, Y. A. (2016). Mammary development and breast cancer: A Wnt perspective. *Cancers*, *8*(7), 65. https://doi.org/10.3390/cancers8070065

Zantl, R., & Horn, E. (2011). Chemotaxis of slow migrating mammalian cells analysed by video microscopy. In *Methods Mol Biol* (Vol. 769, pp. 191–203). https://doi.org/10.1007/978-1-61779-207-6_13

Zhang, X., Martinez, D., Koledova, Z., Qiao, G., Streuli, C. H., & Lu, P. (2014). FGF ligands of the postnatal mammary stroma regulate distinct aspects of epithelial morphogenesis. *Development (Cambridge)*, *141*(17), 3352–3362. https://doi.org/10.1242/dev.106732

Zvelebil, M., Oliemuller, E., Gao, Q., Wansbury, O., MacKay, A., Kendrick, H., Smalley, M. J., Reis-Filho, J. S., & Howard, B. A. (2013). Embryonic mammary signature subsets are activated in Brca1-/-and basal-like breast cancers. *Breast Cancer Research*, *15*(2). https://doi.org/10.1186/bcr3403

Annex 1

Chapter 2

Multidimensional Fluorescence Imaging of Embryonic and Postnatal Mammary Gland Development

Claudia Carabaña and Bethan Lloyd-Lewis

Abstract

Multidimensional fluorescence imaging represents a powerful approach for studying the dynamic cellular processes underpinning the development, function, and maintenance of the mammary gland. Here, we describe key multidimensional imaging strategies that enable visualization of mammary branching morphogenesis and epithelial cell fate dynamics during postnatal and embryonic mammary gland development. These include 4-dimensional intravital microscopy and ex vivo imaging of embryonic mammary cultures, in addition to methods that facilitate 3-dimensional imaging of the ductal epithelium at single-cell resolution within its native stroma. Collectively, these approaches provide a window into mammary developmental dynamics, and the perturbations underlying tissue dysfunction and disease.

Key words Intravital microscopy, Multidimensional imaging, 3D imaging, 4D imaging, Mammary gland morphogenesis, Mammary embryonic development, Ex vivo culture, Optical tissue clearing, In vivo imaging

1 Introduction

Biological imaging is a fundamental and universal tool in the life sciences. For over 80 years, researchers have harnessed microscopy to reveal the inner workings of the mammary gland, a secretory organ essential for the survival of over 5000 mammalian species [1]. Light and electron microscopy of fixed tissue sections and cells are mainstay tools in a mammary gland biologist's armory. While immensely useful, two-dimensional (2D) static analyses are limited in their ability to provide detailed topological and/or dynamic functional information. Thus, to fully comprehend the intricate organization of the branched mammary epithelium, including the dazzling architecture of the lactating mammary gland, three-dimensional (3D; *x*-, *y*-, *z*-) imaging is essential. Moreover, interrogating the inherently dynamic cellular processes underpinning this complex tissue demands the ability to visualize mammary

Maria dM. Vivanco (ed.), *Mammary Stem Cells: Methods and Protocols*,
Methods in Molecular Biology, vol. 2471, https://doi.org/10.1007/978-1-0716-2193-6_2,

epithelial and stromal cells by live, four-dimensional (4D; *x*-,*y*-, *z*-, *t*-) imaging [2].

Due to its widespread accessibility and flexible multicolor acquisition capabilities, confocal laser scanning microscopy (CLSM) is a commonly used optical sectioning technique for volumetric fluorescence imaging [3]. However, conventional confocal modalities rely on single-photon excitation wavelengths in the visible range for imaging, which suffer from tissue light scattering and absorption. This limits confocal microscopy to depths of ~100 μm in most tissues [3, 4]. As such, microscopes equipped with pulsed, infrared multiphoton lasers are favored for deep tissue imaging. Alongside decreased photo-toxicity, multiphoton infrared excitation is less prone to light scattering and absorption, enabling deeper penetration and tissue imaging depths of up to 1 mm [5, 6]. These features make multiphoton systems particularly useful for in vivo fluorescence imaging by intravital microscopy (IVM).

In this chapter, we describe strategies for high-resolution, multidimensional imaging of live and fixed mammary tissues using CLSM and multiphoton microscopy. These include step-by-step protocols for 4D imaging of the adult and embryonic mammary gland using IVM and ex vivo imaging of embryonic cultures, respectively. We also describe approaches that facilitate high-resolution volumetric imaging of the mammary epithelium within its native stroma. These methodologies are particularly useful in contexts that demand the ability to visualize large regions of the mammary epithelium at high spatiotemporal resolutions, such as in genetic-fate mapping studies focused on delineating the differentiation potential and capacity of distinct populations of mammary progenitor cells [2, 7, 8]. Importantly, the approaches outlined enable in situ visualization of the dynamic interplay between mammary epithelial cells and their stromal microenvironment. While not discussed herein, many of the described protocols are also compatible with mammary tumor imaging, providing important insights into breast cancer cell biology [9–12].

IVM is a powerful technique that facilitates high-resolution, real-time fluorescence imaging of cells deep inside live animals [2, 13]. For short-term, acute IVM under non-recovery anesthesia, mammary gland tissues and tumors can be made accessible to the microscope's objective via a surgical "skin-flap" incision [14, 15]. Longitudinal IVM studies, however, require the surgical implantation of an imaging window to protect and provide optical access to tissues during repeated imaging sessions [9, 16–20]. Here, we describe the surgical procedure for implanting conventional metal mammary imaging windows (MIW), based on protocols developed for similarly designed models [17, 20–22]. These imaging windows consist of a glass cover-slipped titanium frame that can be fixed above tissues using a purse-string suture. Alongside, we suggest methods for retracing regions of interest in consecutive imaging sessions, facilitating in situ visualization of specific mammary epithelial cells and tissue structures for extended periods of time. For studies focused on the embryonic mammary gland, we describe methods for high-resolution, time-resolved CLSM and multiphoton imaging of embryonic mammary buds established in ex vivo culture [23–25]. This represents a highly tractable system for interrogating epithelial cell fate dynamics and behaviors during embryonic mammogenesis, a fundamental phase in mammary gland development that remains inaccessible to IVM.

Finally, we outline strategies for high-resolution, wholemount immunostaining and 3D visualization of fixed mammary tissues, aimed at circumventing issues associated with antibody penetration and depth of imaging in this optically opaque organ. This includes an enzymatic digestion-based procedure that facilitates improved deep tissue immunostaining and whole-gland imaging of slide-mounted tissues [26–29]. This approach, however, risks proteolytic-mediated damage to mammary epithelial and stromal cells, limiting its utility in some contexts. Consequently, we also provide two optical tissue clearing methods that enable high-resolution, deep tissue mammary gland imaging in the absence of enzymatic digestion or mechanical dissection, adapted from "see deep brain" (SeeDB) [30] and "clear unobstructed brain imaging cocktails" (CUBIC) [31] protocols. By mitigating tissue light scattering caused by cellular and extracellular components with different refractive indices (RIs), these methods allow for the visualization of expansive regions of the mammary epithelial tree at single-cell resolution within its native stroma [11, 32–37]. These fixed tissue methods are particularly valuable when applied downstream of in vivo/ex vivo live-imaging, allowing mammary cells visualized in 4D to be further characterized by immunostaining for biomarkers of interest. The refinement of the imaging approaches described herein—and their application to novel experimental models and methods for visualizing distinct cells and cellular processes—will continue to provide important insights into mammary gland and breast cancer biology.

2 Materials

2.1 Imaging Window (MIW) Preparation

1. Custom-made titanium MIW (*see* **Note 1**).
2. Glass coverslips (12 mm).
3. 100% Ethanol—prepare 70% (v/v) for use.
4. 100% acetone.
5. Cyanoacrylate-based glass glue.
6. Cotton swabs.
7. Sterile 0.9% saline.

8. *Optional step*: Sterile 1 ng/mL Poly(L-amino acid)-polyethylene glycol/poly(L-lysine)-polyethylene glycol (PLL-g-PEG) solution prepared in 10 mM HEPES buffer, pH 7.4.

2.2 Imaging Window Implantation

1. Extra fine Graefe forceps.
2. Straight Bonn surgical scissors.
3. *Optional*: Needle holder.
4. Autoclave pouches.
5. Bead sterilizer.
6. Sterile surgical gloves, gown, hair net, and face mask.
7. Sterile surgical drapes.
8. Sterile cotton gauze swabs.
9. Isoflurane inhalation anesthetic.
10. Isoflurane vaporizer with oxygen supply.
11. Anesthetic scavenger unit.
12. Anesthesia induction box/cage.
13. Heat pad/heated operating stage.
14. Analgesics, e.g., 0.3 mg/mL buprenorphine hydrochloride.
15. Physiological saline solution.
16. 1 mL syringe and 25-guage needle.
17. Ophthalmic ointment.
18. Tape.
19. Pet hair clipper.
20. *Optional*: depilatory cream.
21. Antiseptic surgical scrub—e.g., Betadine.
22. Sterile non-absorbable polypropylene surgical suture (4-0) (*see* **Note 2**).

2.3 Intravital Microscopy (IVM)

1. Multiphoton confocal microscope surrounded by a heated dark box (e.g., Nikon A1R MP, equipped with a Spectra-Physics Insight Deepsee laser, conventional and resonant scanners, and GaAsP non-descanned detectors).
2. Long-working distance objectives (e.g., Nikon, 25×/NA 1.1 2.0 mm WD water immersion objective).
3. Custom-made imaging window stabilizers, e.g., microscope stage insert or imaging box.
4. Physiological saline solution.
5. Syringe and butterfly-wing needle (for long-term (>3 h) imaging).

6. Mouse vital sign monitor (e.g., MouseOx Plus, Starr Life Sciences).
7. *Optional*: Injectable fluorescent probes to label mammary cells or structures of interest.
8. Image analysis software and processing tools: e.g., ImageJ (https://imagej.nih.gov/ij/), Imaris (http://www.bitplane.com/).

2.4 Embryonic Mammary Gland Dissection

1. Timed-pregnant mice, e.g., fluorescent reporter mouse models (*Lgr5-GFP* [38] or *Rosa26-mTmG* reporter strains [39]).
2. 1 mL syringe and 25-gauge needles.
3. Dissection tools: micro-dissecting scissors, spring scissors, Dumont #5 forceps, and curved micro-dissecting forceps.
4. 100 mm diameter Petri dish.
5. 24-well plates.
6. Stereoscopic dissecting microscope with transmitted illumination.
7. Sterile phosphate-buffered saline solution (PBS).
8. Embryonic culture medium: DMEM/F-12, 2 mM GlutaMAX™, 10% Fetal Bovine Serum (FBS) (v/v), and 20 U/mL Penicillin-Streptomycin (PS).
9. *Optional (for Cre-inducible reporter models)*: Tamoxifen-free base (MP Biomedicals), dissolved in sunflower oil.

2.5 Ex Vivo Embryonic Mammary Gland Culture and 4D Live-Cell Imaging

1. Tissue Culture Dish with Cover Glass Bottom (35 mm).
2. Cell culture inserts (0.4 μm, 30 mm diameter; Millicell).
3. 10× Pancreatin solution pH 7.0: 2.5 g of pancreatin from porcine pancreas and 0.85 g NaCl dissolved in 100 mL of cold MilliQ water and filter-sterilized (for preparation guidelines, *see* **Note 3**).
4. Porcine Trypsin, 1:250 (*see* **Note 4**).
5. Thyrode's solution (pH 7.4): 8 g/L NaCl, 0.2 g/L KCl, 0.05 g/L $NaH_2PO_4 + H_2O$, 1 g/L glucose, 1 g/L $NaHCO_3$ dissolved in 1 L of distilled water and filter-sterilized.
6. Pancreatin-trypsin working solution pH 7.4: 0.225 g porcine trypsin made up in a final volume of 10 mL Thyrode's solution containing 1 mL 10× pancreatin stock solution and 20 μL of PS. Filter-sterilize before use (for preparation guidelines, *see* **Note 4**).
7. Ascorbic acid: 10 mg/mL in distilled water and filter-sterilized.
8. Horizontal laminar flow hood.
9. Inverted CLSM microscope (e.g., Zeiss LSM780/880) or two-photon confocal microscope (e.g., Leica SP8 microscope

with a Chameleon Vision II laser), equipped with a heated environmental chamber.

10. Long-working distance objective (e.g., Zeiss 40×/1.0 NA water immersion objective, Leica 25×/0.95 NA water immersion objective).
11. Image analysis software and processing tools: e.g., ImageJ (https://imagej.nih.gov/ij/), Imaris (http://www.bitplane.com/).

2.6 Enzymatic Digestion and Wholemount Immunostaining of Mammary Tissues

1. Enzyme digestion (ED) mix: HBSS containing 300 U/mL collagenase Catologue numbers and suppliers for digestive enzymes are important information for researchers to replicate protocols as there are many types available (Cat. No. C0130, Sigma) and 300 μg/mL hyaluronidase Same comment as per collagenase - cat numbers/supplier details are helpful for protocols requiring digestive enzymes (Cat. No. 4272, Sigma).
2. 4% Paraformaldehyde in PBS.
3. 0.5 M NH_4Cl.
4. ED Blocking buffer: 1% (w/v) bovine serum albumin (BSA), 5% (v/v) horse serum, and 0.8% (w/v) Triton X-100 in PBS.
5. PBST: PBS + 0.2% (w/v) Tween20.
6. 4′,6-diamidino-2-phenylindole (DAPI) dilactate.
7. Primary and fluorescent secondary antibodies (as required).
8. Microscope slides and coverslips (No. 1.5).
9. Aqua Poly/Mount.

2.7 CUBIC-Based Optical Tissue Clearing

1. 10% Neutral Buffered Formalin (NBF).
2. Card or glass microscope slides.
3. CUBIC Reagent 1A: 10% (w/w) urea, 5% (w/w) *N*,*N*,*N*′,*N*′-tetrakis(2-hydroxypropyl)ethylenediamine, 10% (w/w) Triton X-100 and NaCl (25 mM) in distilled water (*see* **Notes 5** and **6**).
4. CUBIC Reagent 2: 44% (w/w) sucrose, 22% (w/w) urea, 9% (w/w) 2,2′,2″-nitrilotriethanol, and 0.1% (w/w) Triton X-100 in distilled water (*see* **Note 5**).
5. CUBIC Blocking buffer: 10% (v/v) goat serum, 0.5% (w/v) Triton X-100 in PBS.
6. 4′,6-diamidino-2-phenylindole (DAPI) dilactate.
7. *Optional*: Primary and fluorescent secondary antibodies as necessary.
8. Glass-bottomed imaging dishes.

2.8 SeeDB-Based Optical Tissue Clearing

1. 10% Neutral Buffered formalin (NBF).
2. Card or glass microscope slides.
3. α-thioglycerol.
4. Serial fructose solutions made up in distilled water—20%, 40%, 60%, 80%, 100% (w/v), SeeDB (115% (w/v)) (*see* **Note 7**).
5. SeeDB Blocking buffer: 10% (w/v) bovine serum albumin (BSA), 1% (w/v) Triton X-100 in PBS.
6. 4′,6-diamidino-2-phenylindole (DAPI) dilactate.
7. *Optional*: Primary and fluorescent secondary antibodies as necessary.
8. Glass-bottomed imaging dishes.

3 Methods

3.1 Mammary Imaging Window Implantation for 4D-IVM

All animal studies must be approved by institutional ethical committees and national authorities as required. Perform surgeries according to the LASA Guiding Principles of Preparing and Undertaking Aseptic Surgery to minimize infection risk (*see* **Note 8**). Due to its fast induction and recovery, alongside the ability to precisely control anesthesia depth and length, inhalation anesthetics should be used for window implantation and intravital imaging [14, 17]. While acute intravital imaging of the mammary gland under non-recovery anesthesia (i.e., where the animal is euthanized immediately at the end of imaging) can be performed via a skin-flap incision [14, 40, 41], this approach poses challenges for maintaining optimal mouse hydration and physiology. Thus, implanting a MIW is also beneficial when performing non-recovery imaging studies spanning several hours. When applied to fluorescent reporter mouse models, IVM is a powerful approach for tracing the fate, migration, proliferation, geometry, and re-arrangements of individual cells in real time during mammary gland postnatal development and tissue homeostasis.

3.1.1 Mammary Imaging Window (MIW) Preparation

1. Apply a thin layer of cyanoacrylate-based glass glue on the etched inset of the MIW that will contact the coverslip.
2. Using forceps, place the coverslip on the glue-covered inset and apply gentle pressure using a cotton swab for 1 min. If necessary, apply glue at the interface of the coverslip and the window.
3. Allow the glue to dry completely by placing it on its side in a flow cabinet for a minimum of 2 h.
4. If required, use a cotton swab soaked in 100% acetone to remove excessive and condensed glue from the coverslip. Use a cotton swab soaked in 70% (v/v) ethanol to remove the acetone.

5. Place the window glass-side down on a dry tissue and fill with physiological saline to assess for leaks. If the window is watertight, the tissue will remain dry (*see* **Note 9**).
6. Sterilize the MIW by steam sterilization, or by placing it in 70% (v/v) ethanol for a minimum of 30 min. Do not autoclave cover-slipped windows as it will degrade the glue.
7. *Optional*: In a sterile flow cabinet, coat the interior side of the coverslip with PLL-g-PEG solution. Incubate for 1 h at room temperature. Subsequently, wash the window in sterile PBS (*see* **Note 10**).

3.1.2 Surgery Preparation

1. Sterilize surgical tools by autoclaving or dry sterilization (*see* **Note 11**). Thoroughly disinfect the entire operating station (including heat pads, anesthesia nose cones, and adjustment controls) using 70% ethanol.
2. Anaesthetize the mouse in an induction chamber using 3–4% (v/v) gaseous isoflurane (*see* **Note 12**).
3. Once anaesthetized, transfer the mouse onto a heat pad situated away from the sterile operating area, and continue to supply anesthesia via a nose cone. Reduce the isoflurane concentration to 1.5–2% (v/v) (*see* **Note 13**).
4. Apply ophthalmic ointment to both eyes to prevent corneal drying.
5. Administer Buprenorphine at a dose of 0.1 mg/kg mouse body weight by sub-cutaneous injection (*see* **Note 14**).
6. Shave the skin over the fourth mammary gland using electronic pet clippers. As stray hairs under the window pose an infection risk and can obstruct imaging, it is important to remove as much residual hairs as possible from the surgical field. Brief exposures to depilatory cream can be used if necessary.
7. Disinfect the shaved skin using Betadine solution and a cotton gauze. Begin at the center of the surgical site and move out towards the periphery using an ever-widening circular motion. Repeat 3 times using a clean gauze each time.
8. Gently transfer the mouse to the disinfected heated operating stage. Continue to supply the mouse with 1.5–2% (v/v) isoflurane via a nose cone to maintain anesthesia.
9. Loosely immobilize the hind legs using tape. Cover the mouse with a sterile surgical drape equipped with an opening to allow access to the prepared surgical field.

3.1.3 Surgical Implantation of the MIW

1. Verify that the mouse is sufficiently anaesthetized by pinching a hind paw through the overlying sterile drape. If the mouse is unresponsive, proceed to the next step. Otherwise, adjust the isoflurane dose and wait until reflex behaviors are absent.

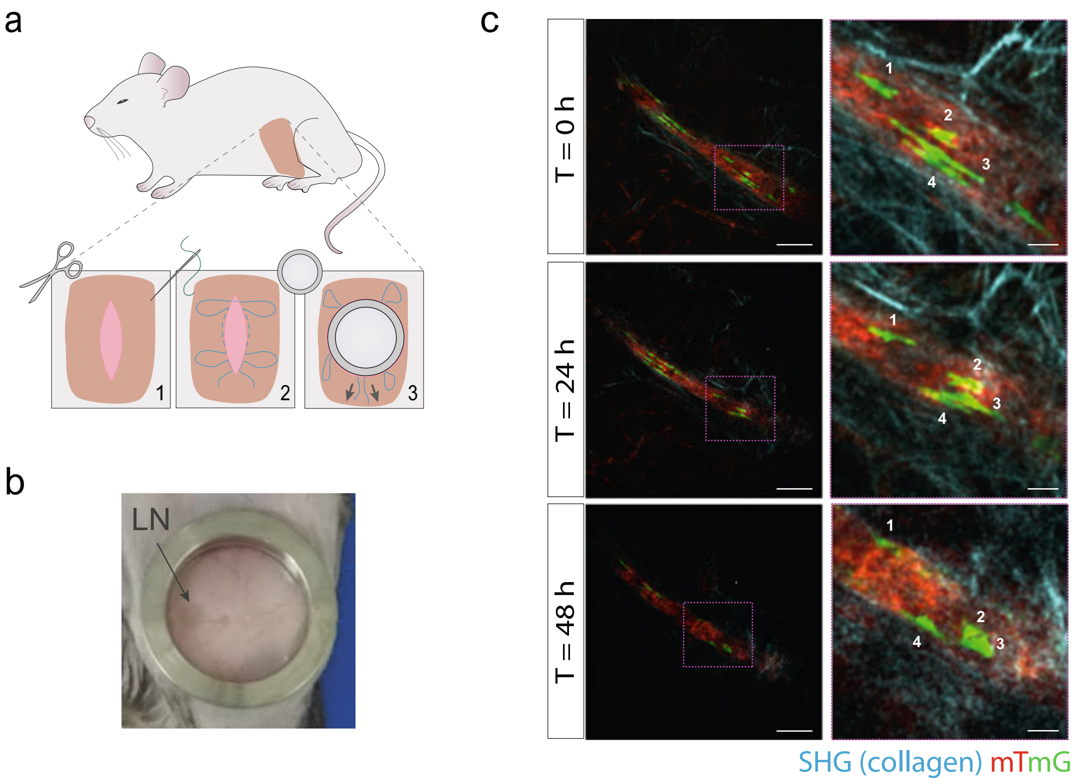


Fig. 1 Intravital imaging through the mammary imaging window (MIW). (**a**) Schematic representation of the surgical procedure. (1) Make an incision above the fourth mammary gland. (2) Place a purse-string suture around the incision. Dotted line denotes internal suture position. (3) Insert the MIW above the mammary gland and place the sutured skin within the groove. Secure the skin around the MIW by tightening the purse-string suture (protocol based on [17, 19]). (**b**) Photograph of the MIW implanted over the fourth abdominal mammary gland. *LN;* lymph node. (**c**) Longitudinal imaging of mammary epithelial cell dynamics by multiphoton IVM. Images show a mammary epithelial duct in a 5-week-old *K5-Cre*^{ERT2}*;Rosa26-mTmG* female mouse. In this model, all cells are labeled with a membrane-bound Tomato fluorescent protein (mT, red). Tamoxifen administration induces Cre-mediated recombination of the reporter allele in sporadic Keratin (K)5-expressing mammary myoepithelial cells, resulting in membrane-bound EGFP fluorescent protein expression (mG, green). Images show the migratory behavior of four EGFP^{+ve} mammary myoepithelial cells over time. Second harmonic generation (SHG) microscopy reveals the collagen organization surrounding the ductal structure. Scale bars: 100 μm (left panels), 25 μm (right panels)

2. Use sterile forceps to gently lift the skin away from the abdominal wall and make a ~12 mm long incision in the flank area above the fourth abdominal mammary gland using sterile scissors (Fig. 1a, panel 1, *see* **Notes 15** and **16**).
3. Use sterile forceps to blunt dissect the skin away from the underlying mammary gland, taking care not to damage the tissue.

4. Place a circular purse-string suture around the incision, approximately 4 mm from skin edges. Keep the external sections of the suture loose at this stage, leaving 4 butterfly-wing shaped outer loops (Fig. 1a, panel 2).
5. Using forceps, carefully insert the window on the exposed mammary gland and gently place the sutured skin into the window groove (Fig. 1a, panel 3).
6. Carefully pull the loops of the purse-string to tighten the skin in the window groove, securing it in place (Fig. 1a, panel 3). Place a double knot to fix the purse-string suture, making sure to hide the knot underneath the upper ring of the window to prevent the mice from biting it open (Fig. 1b).
7. Proceed immediately to intravital imaging, or allow the mouse to recover from the anesthesia in a cage placed on a heating pad at 37 °C. Once recovered, the mouse may be group-housed with other window-bearing mice.
8. Post-operative care: Closely monitor mice for signs of pain and discomfort after surgery. If required, provide post-operative analgesics in accordance with local veterinary guidance (*see* **Note 17**). Inspect the surrounding skin and the tissue underneath the MIW daily for signs of inflammation and necrosis.

3.2 Longitudinal 4D-IVM by Multiphoton Microscopy

1. If not proceeding directly from MIW surgery, anaesthetize the mouse in an induction chamber using 3–4% (v/v) gaseous isoflurane.
2. *Optional*: Administer fluorescent probes by tail vein injection to label mammary cells and structures as necessary (*see* **Note 18**).
3. To maintain mouse hydration during short-term experiments (<3 h), administer a maximum of 500 μL saline by sub-cutaneous injection prior to imaging. In experiments exceeding 3 h, saline should be provided periodically (~50–100 μL/h) using an indwelling intraperitoneal line [14, 17] (*see* **Note 19**).
4. Transfer the mouse onto the stage insert of a multiphoton microscope for imaging, continuing to supply 1.5–2% (v/v) isoflurane via a nose cone to maintain anesthesia.
5. Position and immobilize the MIW for imaging. On an inverted setup, a microscope insert or box customized with a hole that precisely fits the window can be used for stabilization [19]. To image in the upright configuration, the window can be fixed using custom-made holders or microstage devices [40] (*see* **Note 20**). Maintain the mouse's body temperature at 37 °C during imaging, ideally using a dark heated chamber that surrounds the microscope stage.

6. Once the mouse is secure and stabilized, reduce the isoflurane concentration to between 0.8% and 1.2% (*see* **Note 21**).
7. Closely monitor the mouse's vital parameters during imaging:
   (a) Regularly check the breathing rate and adjust the isoflurane supply accordingly if irregular or abnormal breathing is observed.

   (b) Measure the mouse's temperature during imaging using a rectal probe. Adjust the temperature of the heated imaging chamber as necessary.

   (c) Monitor capillary blood flow to ensure optimal conditions for imaging.

   (d) Continuous assessment of vital parameters can be performed using a non-invasive pulse oximetry monitoring system (e.g., MouseOx system) that measures the mouse's temperature, arterial oxygen saturation, breathing distension, and heart and respiratory rates during imaging.
8. When anesthesia levels are tightly controlled and animal vitals well-maintained, 4D-IVM can be performed continuously for up to ~40 h under non-recovery anesthesia [41]. Alternatively, longitudinal IVM can be performed whereby mice undergo repeated, shorter imaging sessions over extended periods of time (*see* **Note 22**).
9. For longitudinal imaging of specific cells and tissue structures, a number of approaches can be used to help retrace regions of interest in consecutive imaging sessions, including:
   (a) Using a motorized stage: In configurations where the relative position of the imaging window is fixed in repeated imaging sessions, the coordinates of individual regions can be stored and used to relocate imaging areas.

   (b) Morphological landmarks: The branched mammary epithelial network is structurally heterogeneous, providing unique patterns of ductal structures and adjacent blood vessels that can be readily recognized in repeat imaging sessions though the eyepiece or by manual scanning (*see* **Note 23**). Intravenous injection of fluorescent agents to label vasculature (e.g., fluorescently labeled dextrans), in addition to second harmonic generation (SHG) imaging of tissue collagen organization, can also assist with retracing regions of interest in consecutive imaging sessions (Fig. 1c).

   (c) Fluorescent reporter mouse models: Heterogeneous labeling of mammary epithelial cells gives rise to unique and identifiable color patterns for serial imaging, e.g., using fluorescent reporter mouse models such as *R26R*-

Confetti [9, 42] or *Rosa26-mTmG* [39] mouse strains (Fig. 1c).

10. *Optional*: To aid the re-identification of intravitally imaged mammary tissue regions in downstream 3D or 2D histological analyses, at the end of the final imaging session use focused high laser power to generate distinct autofluorescent "photo-tattoos" in nearby tissue areas, taking care not to damage cells and tissue structures of interest [43].
11. At experimental endpoints, euthanize the mouse according to institutional/national guidelines.
12. Taking note of the orientation of the tissue relative to the coverslip, harvest and fix the intravitally imaged mammary gland for downstream histological analysis (e.g., by 3D whole-mount immunostaining as described in Subheading 3.5). Harvest the contra-lateral, non-window mammary gland as a control.
13. To reuse the MIW, clean the titanium ring using soap and hot water. Incubate overnight in 100% acetone to remove the glue and release the used coverslip. Prepare the MIW for future implantation following the instructions detailed in Subheading 3.1.1.
14. Process and analyze intravital images using ImageJ (https://imagej.nih.gov/ij/) and/or commercial software such as Imaris (http://www.bitplane.com/), depending on availability.

3.3 Establishing Mammary Embryonic Buds in Culture for 4D Ex Vivo Imaging

The mammary epithelium is first specified as placodes at approximately embryonic (E) Day 11, which invaginate into the underlying mesenchyme to form mammary buds by E12. After E15.5, buds undergo sprouting and branching morphogenesis to give rise to a rudimentary epithelial tree by E18.5 [44]. To visualize this process in real time, mammary embryonic buds must be established in ex vivo culture [23–25]. In this section, we describe methods for high-resolution longitudinal or time-lapse fluorescence imaging of ex vivo cultured embryonic mammary buds by CLSM and multiphoton microscopy. Similarly to IVM, when applied to fluorescent reporter mouse models this approach allows for the dynamic behavior and fate of individual embryonic mammary cells to be traced in real time during embryonic mammary morphogenesis.

3.3.1 Dissection of the Embryonic Mammary Gland

1. Set up timed mouse matings to obtain pregnant females bearing embryos of the desired genotype and age. Embryo stage is determined by the detection of a vaginal plug the following day (detection at mid-day defined as 0.5 days-post-coitus, i.e., E0.5).

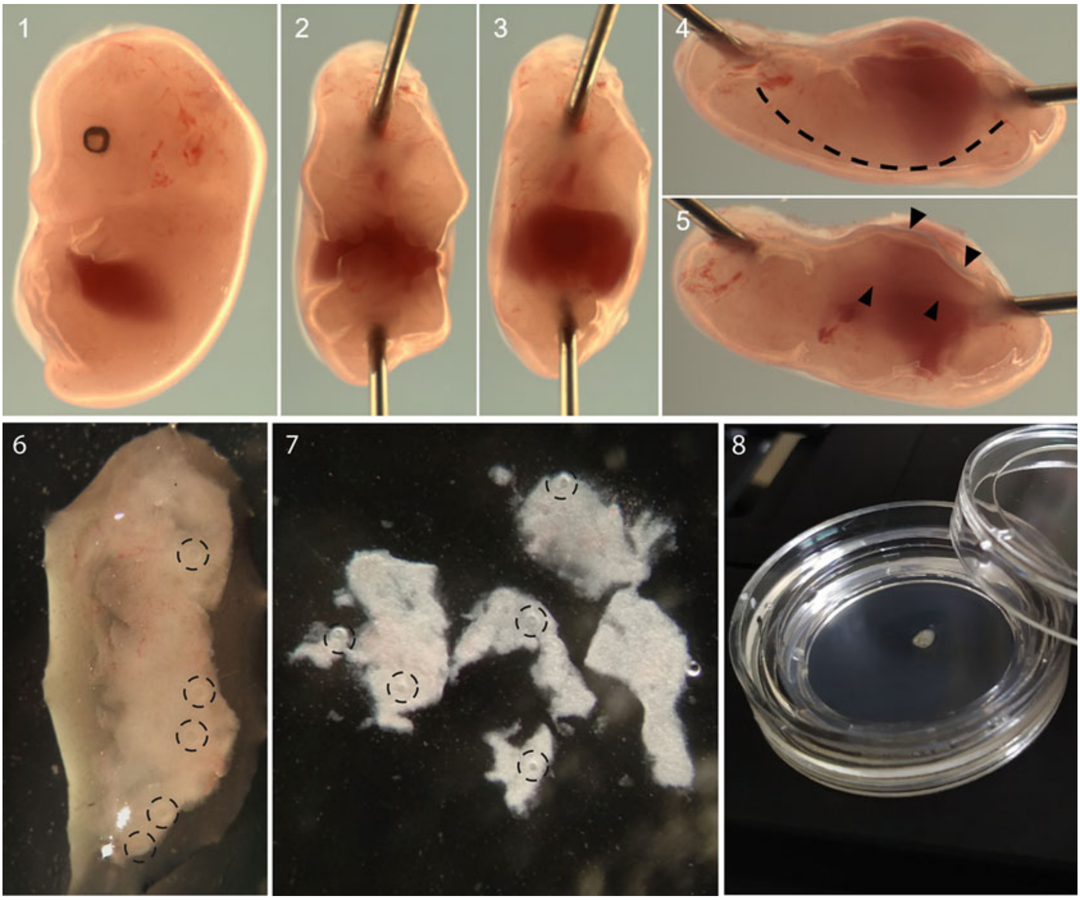


Fig. 2 Embryonic mammary gland dissection protocol. (1) Collect all embryos from the uterus of a pregnant female mouse. (2) Attach the decapitated embryo to a dish filled with set silicon. (3) Remove the limbs. (4) Cut along the dorsal-lateral line of the embryo (dashed line). (5) Detach the flank (see arrowheads) to the midline. (6) Collect both flanks of the embryo containing the embryonic mammary buds (marked with dashed circles). (7) Remove skin epithelium after enzymatic digestion with pancreatic-trypsin working solution. Dashed circles denote the location of embryonic mammary buds within the mesenchyme. (8) Culture the isolated tissues on an air-liquid interface. (Protocol based on [25])

2. *Optional*: If using Cre-inducible fluorescent reporter mice, administer a low dose of tamoxifen to pregnant females 24 h before tissue dissection to induce fluorescent labeling in embryos. While a tamoxifen dose of 0.1 mg/g of mouse body weight is commonly used, precise doses will vary depending on specific study requirements [28].
3. At the desired embryonic stage, sacrifice the pregnant female and harvest the embryos from the uterus in a 100 mm diameter Petri dish filled with cold PBS.
4. Remove the yolk sac and separate each embryo from its placenta, taking care not to damage the tissue (Fig. 2, panel 1).

5. Sacrifice each embryo by decapitation and place in separate wells of a 24-well plate filled with cold PBS. Keep tissues to confirm genotypes and sex by PCR (*see* **Note 24**).
6. Place one embryo in a 35 mm culture dish filled with set silicon. Secure the embryo in place by pinning the neck and tail joint using needles or dissection pins (Fig. 2, panel 2).
7. Remove the limbs. This makes the first and fifth mammary buds accessible for dissection (Fig. 2, panel 3).
8. Perform a small incision above the tail joint of the embryo. Using spring scissors, cut along the dorsal-lateral line from the hind limb to the forelimb in the right flank of the embryo (Fig. 2, panel 4).
9. Detach the flank of the embryo from the incision along the dorsal-lateral line to the midline (Fig. 2, panel 5). Hold the tissue using Dumont #5 forceps, and use the spring scissors to trim the right flank (*see* **Note 25**).
10. Repeat **steps 8** and **9** with the left flank of the embryo, but this time cutting along the dorsal-lateral line from the forelimb to the hind limb.
11. Transfer both flanks to a new 24-well plate with PBS (Fig. 2, panel 6).

3.3.2 Separating the Embryonic Skin Epithelium and Mesenchyme

This procedure entails proteolytic digestion of dissected embryonic flanks, based on a protocol developed by the laboratory of M. Mikkola [25].

1. Replace the PBS with pancreatic-trypsin working solution and incubate for 4–5 min. Optimal incubation times are heavily dependent on embryo stage, in addition to the particular enzyme batches used [25]. Thus, closely monitor the tissue under the stereomicroscope during enzyme treatment.
2. When the skin epithelium starts to detach from the edges of the mammary mesenchyme, inactivate the enzymes by replacing the pancreatic-trypsin solution with DMEM/F-12 embryonic culture medium (*see* **Note 26**).
3. Incubate the mammary tissue on ice for 30–45 min.
4. Place the mammary tissue in a 35 mm culture dish. Using two needles, gently peel the skin epidermis away from the mesenchyme containing the embryonic mammary buds. The mesenchyme is required for embryonic mammary epithelial development [23] and should be maintained as intact as possible around mammary buds (Fig. 2, panel 7).
5. Once isolated, use a plastic pipette to transfer the intact mesenchyme containing the embryonic mammary buds to fresh

DMEM/F-12 embryonic culture medium in a 24-well plate. Repeat this process for all harvested tissues.

3.3.3 Establishing Mammary Buds in Ex Vivo Culture

1. Per embryo, prepare a 35 mm cover glass-bottomed tissue culture dish containing 1 mL of DMEM/F-12 embryonic culture medium freshly supplemented with 75 μg/mL ascorbic acid.
2. Place the embryonic mammary tissue on a cell culture insert (containing 0.4 μm pores) using a 1000 mL pipette.
3. Using curved micro-dissecting forceps, carefully and slowly place the cell culture insert into the prepared glass-bottomed tissue culture dish to avoid bubble formation. This ensures mammary embryonic buds are cultured on an air-liquid interface, whereby the tissue remains exposed to air while maintaining contact with the embryonic cell culture medium through the pores of the cell culture insert (Fig. 2, panel 8).
4. Maintain mammary cultures in a tissue culture incubator at 37 °C and 5% CO_2 atmosphere.
5. Exchange the culture media with fresh media every second day for the duration of the experiment. Cultures can be maintained ex vivo for up to 2 weeks.

3.4 4D Time-Lapse and Longitudinal Imaging of Ex Vivo Embryonic Mammary Cultures

Once established in ex vivo culture, embryonic mammary branching development can be recorded daily by 3D fluorescence imaging for up to 2 weeks. Use an inverted CLSM or multiphoton microscope (e.g., Zeiss LSM780/880 or Leica SP8) equipped with long-working distance objectives to acquire high-resolution 3D image stacks at each time-point (Fig. 3a). Alternatively, time-lapse imaging can be performed for up to 48 h (Fig. 3b). Below, we provide an overview of the experimental conditions that facilitate high-resolution, ex vivo imaging of fluorescently labeled embryonic mammary epithelial cells over time.

1. Isolate embryonic mammary buds from a fluorescent reporter mouse model and establish in ex vivo culture as described in Subheading 3.3. Figure 3a, b show a cultured embryonic mammary bud established from a *Lgr5-GFP* [38] or *Rosa26-mTmG* [39] reporter mouse embryo, respectively. In the absence of Cre-mediated recombination in the *Rosa26-mTmG* mammary bud (Fig. 3b), all embryonic mammary epithelial and stromal cells are labeled with a membrane-bound Tomato fluorescent protein.
2. Due to superior depth of imaging and low photo-toxicity, a multiphoton confocal microscope is preferred for time-lapse imaging of ex vivo embryonic mammary cultures, e.g., an inverted Leica SP8 microscope equipped with a femtosecond Chameleon Vision II multiphoton laser (680–1350 nm;

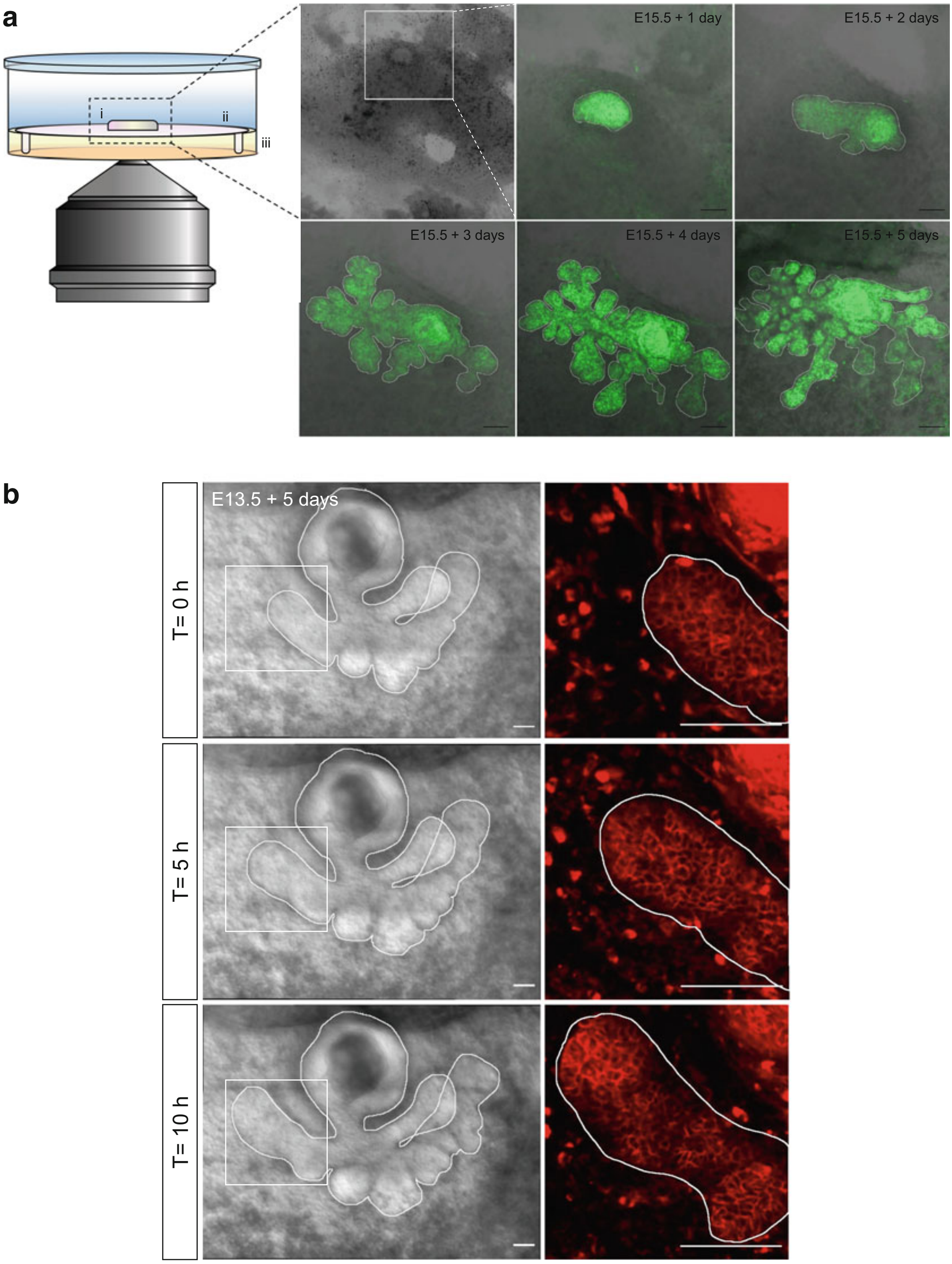


Fig. 3 4D ex vivo imaging of mammary embryonic bud cultures (**a**) Longitudinal fluorescence imaging of embryonic mammary branching morphogenesis. Schematic representation of the microscope configuration for live-cell imaging of embryonic mammary bud cultures. The tissue (i) is deposited on a cell culture insert (ii) placed in contact with the culture medium (iii) in a glass bottom dish. Use an inverted confocal or multiphoton microscope equipped with a long-working distance objective for imaging. Images show the growth of an E15.5 embryonic mammary bud isolated from a *Lgr5-GFP* [38] mouse embryo over 5 days.

Coherent, Inc.) and long-working distance objectives (e.g., Leica 25×/0.95 NA water immersion objective).

3. The microscope should be equipped with an incubation chamber that maintains tissues at 37 °C, 5% CO_2 atmosphere and 95% humidity (*see* **Note 27**). To avoid sample drift in the *Z*-axis, allow environmental conditions of the microscope chamber to stabilize for 30–60 min before commencing image acquisition.
4. Acquire image sequences as *Z*-stacks of the desired volume. Image acquisition time may vary between 5 and 60 min, depending on tissue thickness, required z resolution and area imaged, as well as the number of fluorophores to detect.
5. To image the whole embryonic mammary epithelium, overview tile scans can be acquired. Ensure at least a 10% overlap between tiled images, which can be stitched into larger mosaics using the microscope's acquisition software, or ImageJ Plugins (e.g., the Grid Collection/Stitching plugin) (https://imagej.nih.gov/ij/).
6. Time-lapse processing and analysis can be performed using ImageJ and/or commercial software such as Imaris (http://www.bitplane.com/), depending on availability.

3.5 3D Fluorescence Imaging of Fixed Mammary Gland Tissues

In this section, we provide three protocols for wholemount immunostaining and 3D visualization of fixed mammary gland tissues, encompassing proteolytic digestion and two optical tissue clearing methods based on published CUBIC [31] and SeeDB [30] protocols. These can be applied to freshly harvested tissues, or after completing 4D in vivo/ex vivo imaging to further characterize live-imaged tissue regions by immunostaining. While proteolytic digestion methods are particularly useful for whole-gland imaging of immunostained mammary tissues [26–29], CUBIC and SeeDB-basedoptical tissue clearing methods facilitate visualization of the mammary epithelial tree at single-cell resolutions within its intact stroma [11, 32–36] (*see* **Note 28**). For an overview of the optical

Fig. 3 (continued) Maximum intensity projections of image sequences are displayed, and the rendered surface of the mammary epithelium is outlined in white. Scale bars: 100 μm. (**b**) Time-lapse imaging of embryonic mammary cultures. Close-up of one embryonic mammary bud dissected from a *Rosa26-mTmG* [39] mouse embryo at Day E13.5, and cultured ex vivo for 5 days prior to time-lapse imaging. Images show the growth of an epithelial branch over 10 h (white box). In this reporter mouse strain, all cells are labeled with a membrane-bound Tomato fluorescent protein (red). Inset images show a thin optical slice. Individual mammary epithelial cells can be clearly visualized in ex vivo cultures by high-resolution multiphoton microscopy. Branching morphogenesis can be monitored for up to 48 h by time-lapse confocal or multiphoton imaging. The rendered surface of the mammary epithelium is outlined in white. Scale bars: 100 μm

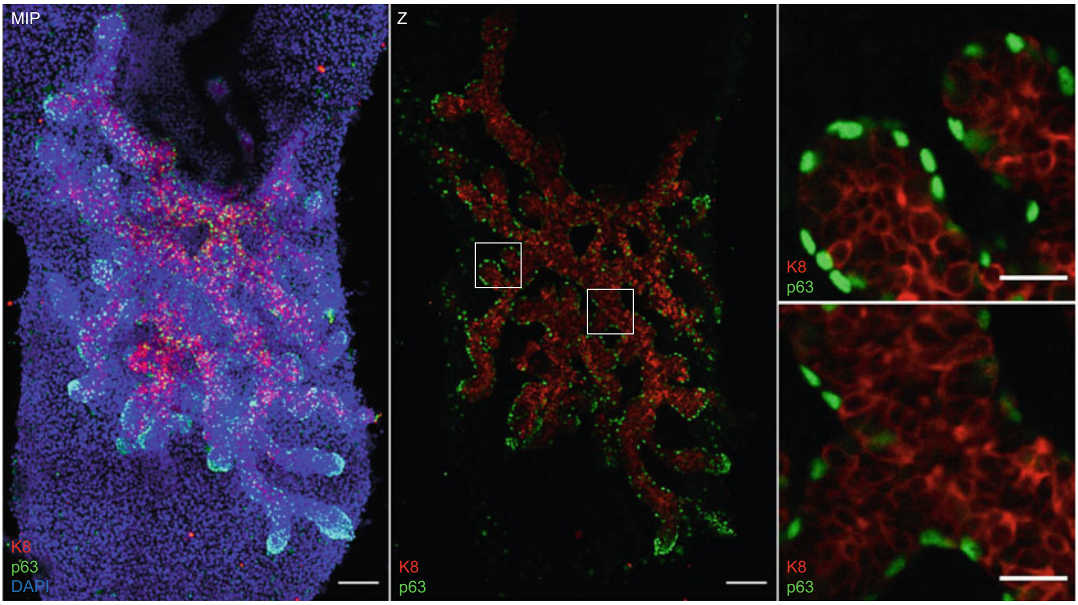


Fig. 4 Wholemount immunostaining of embryonic mammary cultures. Immunofluorescence staining of an embryonic mammary gland dissected at Day E13.5 and grown in ex vivo culture for 8 days. Left panel: Maximum intensity projection (MIP) showing the expression of the luminal epithelial marker protein Keratin 8 (K8, red), and the basal epithelial marker protein p63 (green) in embryonic mammary cells. Nuclei are stained with DAPI (blue). Center panel: A single optical section (*z*). White boxes mark regions displayed in the right panels. Scale bar: 100 μm. Right panel: Close-up images of K8 (red) and p63 (green) expressing cells in a tip (upper panel) and branch (lower panel) region of the embryonic mammary epithelium. Scale bar: 25 μm

tissue clearing methods previously applied to the mammary gland, in addition to their advantages and disadvantages, *see* [2, 11].

3.5.1 Proteolytic Digestion-Based Immunostaining of Mammary Gland Tissues

1. Euthanize mice according to institutional/national guidelines.
2. Excise mammary glands and immerse in Enzyme Digestion solution for 30–60 min (depending on tissue size) at 37 °C with gentle agitation. Due to its small size, high quality wholemount immunostaining can be performed in embryonic mammary gland tissues and explants without the need for enzymatic digestion [45] (Fig. 4).
3. Wash tissues 3 times in HBSS (5 min each time).
4. Fix tissues for 2 h at room temperature in 4% PFA.
5. Wash samples twice in 0.5 M NH_4Cl (10 min each time).
6. Wash tissues 3 times in PBS (10 min each time).
7. Incubate tissues in ED blocking buffer for 3 h at room temperature, or overnight at 4 °C with gentle agitation.
8. Incubate samples in primary antibodies diluted in blocking buffer overnight at room temperature.
9. Wash samples 3 times in PBST (10 min each time).

10. Incubate samples with secondary antibodies diluted in blocking buffer for 5 h at room temperature, or overnight at 4 °C with gentle agitation.
11. Wash samples 3 times in PBST (10 min each time).
12. Incubate tissues with DAPI (10 μM) for 30 min to 1 h at room temperature with gentle agitation to stain nuclei.
13. Mount tissues on microscope slides using Aqua Poly/Mount, taking care to avoid bubbles.
14. Acquire images using CLSM (e.g., Zeiss LSM780/880 or Leica SP8) with a long-working distance objective to facilitate deep tissue imaging (e.g., Zeiss 25×/0.8 oil immersion objective or Leica 25×/0.95 NA water immersion objective).
15. Imaging considerations: Adjust laser power and gain manually to give optimal fluorescence intensity for each fluorophore with minimal photobleaching. Acquire *Z*-stacks using step-sizes and line averaging appropriate for the desired resolution. If performing tile scans, at least a 10% overlap is recommended for optimal stitching of tiled images into larger mosaics (Fig. 4). Stitching can be performed using the microscope's acquisition software, or using ImageJ Plugins (e.g., the Grid Collection/Stitching plugin) (https://imagej.nih.gov/ij/).
16. Process and analyze image stacks using ImageJ (https://imagej.nih.gov/ij/) and/or commercial software such as Imaris (http://www.bitplane.com/), depending on availability.

3.5.2 Modified CUBIC Tissue Clearing and Immunostaining of Mammary Gland Tissues

1. Euthanize mice according to institutional/national guidelines.
2. Excise mammary glands and spread immediately on card (Tetra Pak) or glass microscope slides (*see* **Note 29**).
3. Fix tissues by immersing in 10% NBF for 6–9 h (according to tissue size/thickness) at room temperature (*see* **Note 30**).
4. Wash tissues briefly in PBS with gentle agitation to remove residual NBF. Cut tissues into large (~15 × 15 × 2 mm) pieces if necessary (*see* **Note 31**).
5. Immerse tissues in CUBIC Reagent 1A at 37 °C for 2–3 days, depending on the size of the tissue, exchanging the solution with fresh R1A each day. For timeline, *see* Fig. 5a.
6. Wash samples 3 times in PBS (10 min each time) to remove excess R1A solution. If tissues are to be imaged for genetically encoded fluorescent proteins and do not require immunostaining, proceed to **step 12**.
7. *Optional immunostaining*: Immerse tissues in CUBIC blocking buffer and incubate overnight at 4 °C with gentle agitation.

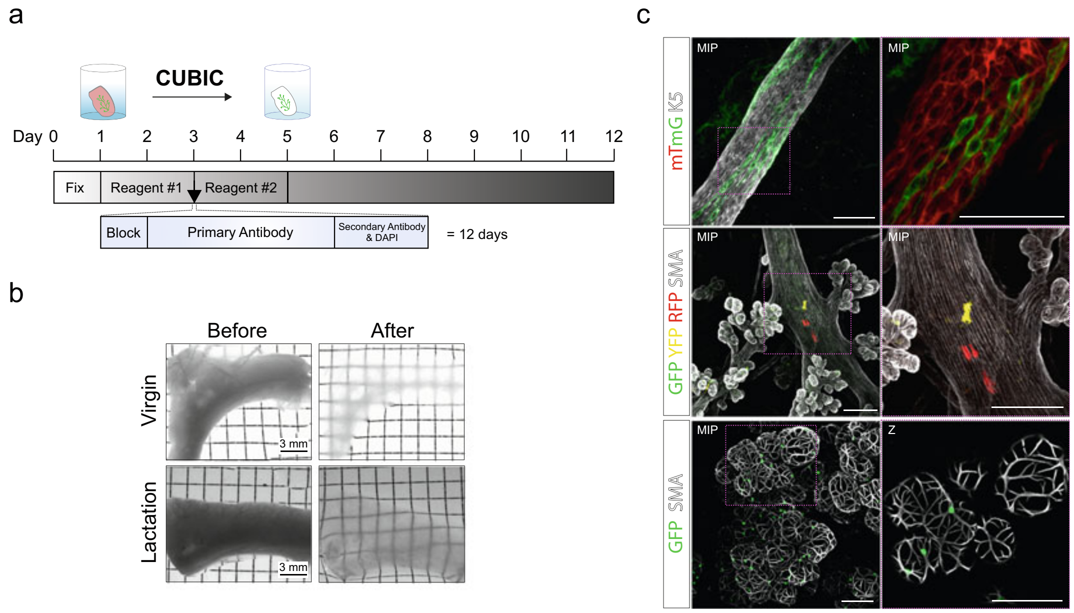


Fig. 5 Clear unobstructed brain imaging cocktails (CUBIC) optical clearing and 3D imaging of mouse mammary tissues. (**a**) CUBIC optical tissue clearing and immunostaining protocol and timeline. Black arrow shows the stage at which (optional) immunostaining may be performed. The experimental timeline may be adapted depending on the desired degree of transparency, and the size and nature of the tissue. (**b**) Transmission images of whole abdominal (fourth) virgin and lactating mammary glands, before and after CUBIC clearing (adapted from [11] with permission from Springer under http:/creativecommons.org/licenses/by/4.0/). (**c**) Example three-dimensional confocal images of cleared mammary tissues immunostained with basal mammary epithelial cell markers (K5 or smooth muscle actin (SMA)) showing compatibility of CUBIC clearing with high-resolution imaging of genetically encoded fluorescent proteins. Top panels show a close-up image of a mammary duct in a virgin *SMA-Cre*^{ERT2}*; Rosa26-mTmG* reporter mouse. Membrane labeling in recombined (mG, green) mammary basal cells can be observed at high resolutions by CUBIC clearing. Non-recombined cells express membrane-bound Tomato fluorescent protein (mT, red). Middle and bottom panels show 3D images of cleared mammary tissues from involuting and lactating *R26R-Confetti* fluorescent reporter mice, respectively. Reporter expression (nuclear GFP, cytosolic YFP, and cytosolic RFP) is induced at low, sporadic levels. Scale bars: 50 μm (top two panels), 100 μm (middle and bottom panels). MIP, maximum intensity projection; *z*, single optical section

8. Incubate tissues with primary antibodies diluted in CUBIC blocking buffer at 4 °C for 4 days with gentle agitation. Incubation times may be optimized for specific antibodies.
9. Wash tissues 3 times in PBS (1 h each time) with gentle agitation at room temperature.
10. Incubate samples with fluorescent-dye conjugated secondary antibodies made up in PBS for 2 days at 4 °C with gentle agitation (*see* **Note 32**).
11. Wash tissues 3 times in PBS (1 h each time) with gentle agitation at room temperature.

12. Incubate with DAPI (10 μM) for 2–3 h at room temperature with gentle agitation.
13. Transfer samples to CUBIC Reagent 2 and incubate at 37 °C for at least 24 h for refractive index matching before imaging (Fig. 5b).
14. Image samples in CUBIC Reagent 2 by CLSM or multiphoton microscopy within 1 week (e.g., Zeiss LSM780/880 or Leica SP8 equipped with long-working distance objectives to facilitate deep tissue imaging). Samples may be imaged on inverted microscopes using glass-bottomed iBidi dishes, or mounted in CUBIC Reagent 2 for imaging (Fig. 5c) (*see* **Note 33**).
15. Imaging considerations: Adjust laser power and gain manually to give optimal fluorescence intensity for each fluorophore with minimal photobleaching. Acquire *Z*-stacks using step-sizes and line averaging appropriate for the desired resolution. If performing tile scans, at least a 10% overlap is recommended for optimal stitching of tiled images into larger mosaics. Stitching can be performed using the microscope's acquisition software, or using ImageJ Plugins (e.g., the Grid Collection/Stitching plugin) (https://imagej.nih.gov/ij/).
16. If tissues were previously imaging by 4D intravital imaging, place cleared samples in the same orientation for imaging. To retrace regions of interest, some of the strategies suggested for longitudinal IVM may be used (*see* Subheading 3.2, **step 9**), e.g., using morphological landmarks or laser-induced autofluorescent photo-tattoos.
17. Process and analyze image stacks using ImageJ (https://imagej.nih.gov/ij/) and/or commercial software such as Imaris (http://www.bitplane.com/), depending on availability.

3.5.3 SeeDB Tissue Clearing and Immunostaining of Mammary Gland Tissues

1. Euthanize mice according to institutional/national guidelines.
2. Excise mammary glands and spread immediately on card (Tetra Pak) or glass slides (*see* **Note 29**).
3. Fix tissues by immersing in 10% neutral buffered formalin (NBF) for 6–9 h (according to tissue size/thickness) at room temperature (*see* **Note 30**).
4. Subsequently, wash tissues briefly in PBS with gentle agitation to remove residual NBF. Cut tissues into large (~15 × 15 × 2 mm) pieces if necessary (*see* **Note 31**).
5. If tissues are to be imaged for genetically encoded fluorescent proteins and do not require immunostaining, proceed to **step 11**. For timeline, *see* Fig. 6a.
6. *Optional immunostaining*: Immerse tissues in SeeDB blocking buffer overnight at 4 °C with gentle agitation.

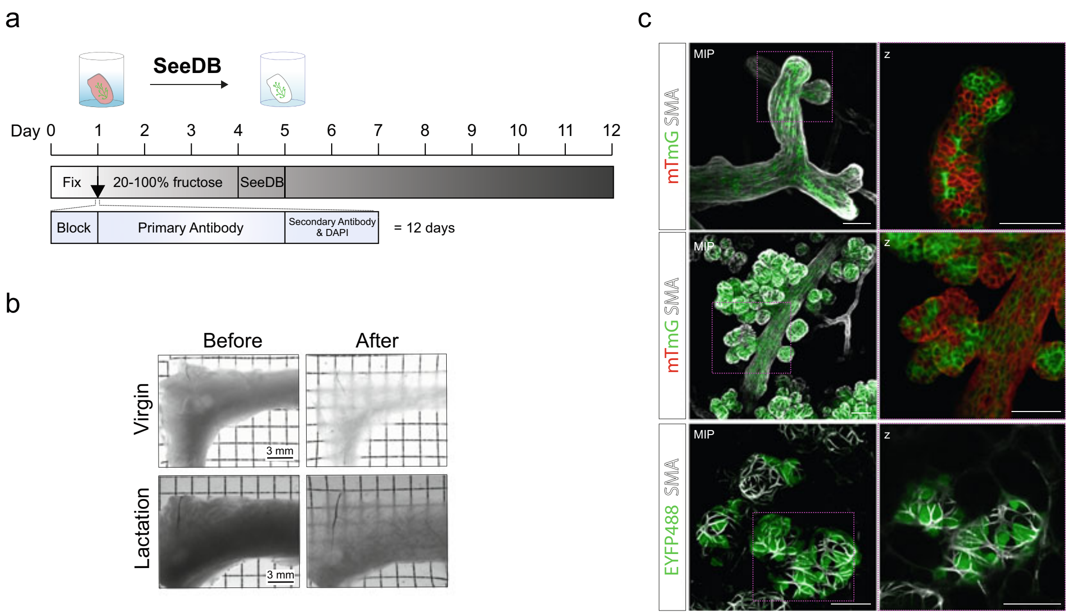


Fig. 6 See deep brain (SeeDB)-clearing and 3D imaging of mouse mammary tissues. (**a**) SeeDB tissue clearing and immunostaining protocol and timeline. Black arrow shows the stage at which (optional) immunostaining may be performed. The experimental timeline may be adapted depending on the desired degree of transparency, and the size and nature of the tissue. (**b**) Transmission images of whole abdominal (fourth) virgin and lactating mammary glands, before and after SeeDB clearing (adapted from [11] with permission from Springer under http:/creativecommons.org/licenses/by/4.0). (**c**) Example three-dimensional confocal images of SeeDB-cleared mammary tissues immunostained with a basal cell marker (SMA) showing compatibility of SeeDB clearing with high-resolution imaging of genetically-encoded fluorescent proteins. Top and middle panels show images of the mammary epithelium in virgin and pregnant *N1-Cre*^{ERT2}; *Rosa26-mTmG* reporter mice, respectively. Membrane labeling in recombined (mG, green) and unrecombined (mT, red) mammary luminal cells can be observed at high resolutions. Bottom panels show images of SeeDB-cleared mammary tissues from a lactating *Rosa26-YFP* mouse model immunostained with a GFP antibody. MIP, maximum intensity projection; *z*, single optical section. Scale bars: 50 μm

7. Incubate tissues with primary antibodies diluted in SeeDB blocking buffer at 4 °C for 4 days with gentle agitation. Incubation times may be optimized for specific antibodies.
8. Wash tissues 3 times in PBS (1 h each time) with gentle agitation at room temperature.
9. Incubate samples with fluorescent-dye conjugated secondary antibodies made up in PBS for 2 days at 4 °C with gentle agitation.
10. Wash tissues 3 times in PBS (1 h each time) with gentle agitation at room temperature.
11. Incubate with DAPI (10 μM) for 2–3 h at room temperature with gentle agitation.

12. Subsequently, serially incubate samples for 8–16 h (twice daily changes, i.e., at beginning and end of the day) at room temperature with gentle agitation in increasing fructose solutions: 20%, 40%, 60%, and 80% (w/v) fructose solutions containing freshly added 0.5% (v/v) α-thioglycerol to inhibit the Maillard reaction [30].
13. Incubate samples in 100% (w/v) fructose solution containing freshly added 0.5% (v/v) α-thioglycerol for 24 h at room temperature.
14. Incubate samples in SeeDB (115% (w/v)) fructose solution containing freshly added 0.5% (v/v) α-thioglycerol for 24 h at room temperature before imaging (Fig. 6b).
15. Image samples in SeeDB reagent by CLSM or multiphoton microscopy within 2 weeks (e.g., Zeiss LSM780/880 or Leica SP8 equipped with long-working distance objectives to ensure deep tissue imaging). Samples may be imaged on inverted microscopes using glass-bottomed iBidi dishes, or mounted in SeeDB reagent for imaging (Fig. 6c) (*see* **Note 33**).
16. Imaging considerations: Adjust laser power and gain manually to give optimal fluorescence intensity for each fluorophore with minimal photobleaching. Acquire *Z*-stacks using step-sizes and line averaging appropriate for the desired resolution. If performing tile scans, at least a 10% overlap is recommended for optimal stitching of tiled images into larger mosaics. Stitching can be performed using the microscope's acquisition software, or using ImageJ Plugins (e.g., the Grid Collection/Stitching plugin) (https://imagej.nih.gov/ij/).
17. If tissues were previously imaging by 4D intravital imaging, place cleared samples in the same orientation for imaging. To retrace regions of interest, some of the strategies suggested for serial IVM may be used, e.g., using morphological landmarks or laser-induced autofluorescent photo-tattoos.
18. Process and analyze *Z*-stack images using ImageJ (https://imagej.nih.gov/ij/) and/or commercial software such as Imaris (http://www.bitplane.com/), depending on availability.

4 Notes

1. This protocol describes a re-usable titanium MIW that is based on the design of the abdominal imaging window [17, 21, 22]. This consists of an upper and a lower titanium ring (outer diameter 14 mm), separated by a 0.9 mm width groove. Titanium possesses superior biocompatibility compared to other stainless alloys. More recent advances in imaging window designs include a titanium version with a replaceable lid [22],

and a silicone-based model consisting of a fully flexible, sutureless design that is ideally suited for long-term IVM of growing tissues and tumors, including the mammary gland [21].

2. Suture material may depend on user preference. Non-absorbable 5-0 nylon and 6-0 silk sutures are also used for imaging window implantation.
3. Dissolve pancreatin from porcine pancreas (P3292-25G, Sigma) and NaCl in MilliQ water using a magnetic stirrer on ice for 3–4 h (or at 4 °C overnight). Centrifuge at approximately 3200 × *g* for 10 min and filter using suction prior to use.
4. Dissolve 0.225 g porcine trypsin (cat. no: 85450C-25G, Sigma) in 6 mL of ice-cold Thyrode's solution using a magnetic stirrer on ice. Once in solution add 1 mL of 10× pancreatin stock solution, in addition to 20 μL of PS. Using NaOH, adjust the pH of the solution to 7.4. Make up the solution to a final volume of 10 mL using additional ice-cold Thyrode's solution. Filter and prepare single-use aliquots.
5. CUBIC Reagent 1A is a modified, unpublished version of CUBIC Reagent 1 [31] available at http://cubic.riken.jp/. CUBIC Reagent 2 has been slightly modified from the original formulation 31 as described in Ref. [11] CUBIC Reagent 1 and 2 require intermittent cycles of heating (60–65 °C, 30 minutes) and agitation (using a magnetic stirring bar and stirrer) to dissolve (typically over 2–3 h).
6. *N*,*N*,*N*′,*N*′-tetrakis(2-hydroxypropyl)ethylenediamine is highly viscous. Dispense this reagent first and adjust other reagent amounts accordingly. In the original protocol [31], the authors state that a specific brand of Triton X-100 (Nacalai Tesque Inc., 25987-85, Japan) is required to avoid quenching endogenous EGFP fluorescence. However, in our hands, Triton X-100 purchased from VWR International is compatible with EGFP fluorescent protein in 10% NBF-fixed mammary gland tissues. If quenching is suspected, consider immunostaining tissues with a GFP antibody, or use an alternative optical tissue clearing method [2, 46].
7. Fructose solutions greater than 60% are difficult to get into solution. Use heat (60 °C) and agitation to dissolve.
8. To avoid contaminating the surgical station, prepare the mouse for surgery in a separate area before transferring to the sterilized operating station. Autoclaved tools should only come into contact with sterile gloves, surfaces, and the disinfected skin area undergoing surgery. To safeguard sterility throughout the procedure, the surgeon should only contact sterile tools and the prepared surgical field. Mouse preparation and

handling, adjustment of anesthesia flow rates, and similar tasks should be performed by an assistant.

9. For optimal results, it is essential that the seal between the titanium ring and coverslip is air and watertight.
10. PEG-based coating of the glass coverslip can be performed to prevent cell attachment and to improve its biocompatibility, limiting the risk of an inflammatory or immune reaction in response to implantation.
11. Ideally, a new set of sterilized tools should be used per mouse. If this is unfeasible, a bead sterilizer may be used to sterilize tools between limited numbers of mice.
12. Isoflurane—which requires oxygen as a carrier gas—is a well-tolerated inhalation anesthetic with minimal side-effects and a short recovery time. Isoflurane induction and maintenance doses may require optimization depending on animal parameters (including strain, age, and condition) and the available setup. It is advisable to use relatively high doses of isoflurane for rapid induction (up to 4%), reducing doses to maintenance levels as quickly as possible to minimize the time exposed to high concentrations [14]. Isoflurane should be handled in well-ventilated areas using systems equipped with a gas scavenger to minimize user exposure to exiting gases.
13. Maintaining the physiological body temperature of the mouse is critical for long-term survival under anesthesia. This can be achieved using adjustable heated anesthesia posts and induction boxes, heat pads, or environmental chambers during intravital imaging. Conversely, hyperthermia may amplify the effect of the inhalation anesthetic, depressing respiration rates. Monitor the mouse's body temperature throughout surgery and intravital imaging using a rectal probe.
14. Buprenorphine may cause mild respiratory depression. Multimodal regimes consisting of Buprenorphine and a NSAID (e.g., Carprofen at a 5 mg/kg) may be recommended by your local veterinarian. However, depending on the nature of the study, suppression of the immune system by NSAIDs may impact important experimental parameters.
15. Gently handle the skin at incision edges when using forceps to avoid compression-associated damage. Consider using non-serrated forceps to minimize this risk. For optimal implantation, the incision size should closely match the size of the window.
16. While the MIW can be implanted over inguinal (third) mammary glands, the fourth is less impacted by respiratory movements during imaging.
17. Post-operative buprenorphine (0.1 mg/kg for up to 3 days) may be administered to provide additional pain relief. NSAIDs

(e.g., Carprofen at a 5 mg/kg dose and/or ibuprofen administered in drinking water) may also be used if appropriate (*see* **Note 14**). Local application of a topical analgesic (e.g. 1% Xylocaine) at the surgical site can be used to provide additional pain relief [17].

18. Endogenous and injectable fluorescent probes may be used to study aspects of mammary gland/tumor cell biology and tissue morphology [10, 47]. For example, blood vessels can be labeled by injecting fluorescently conjugated dextrans into the circulation to investigate vascular flow and permeability, in addition to the invasion of mammary tumorigenic cells into nearby vessels. Vessel labeling can also aid identification of imaging regions in repeated IVM sessions, in addition to the registration of serially acquired image stacks.
19. A winged infusion set attached to a syringe can be used to administer saline during imaging. This can be performed manually, or by using a programmable syringe pump for controlled, continuous administration.
20. An inverted microscope is preferable as it provides better stabilization, reducing image distortions arising from respiration-induced tissue movement. It is important not to compress the underlying tissue and impair blood flow when immobilizing the MIW.
21. These anesthesia levels are optimal for long-term maintenance of mice in a nonresponsive state with a constant and non-forced breathing pattern. Irregular and abnormal breathing patterns are associated with persistent anesthesia greater than 1.5%, which can decrease survival times [14, 41].
22. Due to the limited tolerance of mice to repeated anesthesia, the duration and frequency of imaging sessions in longitudinal experiments should be adapted according to the study parameters under investigation, and in line with institutional/national ethical rules. For instance, if mice are to be anesthetized daily for imaging, this should be restricted to shorter study time periods, with imaging sessions kept as brief as possible to aid recovery.
23. Generating a tile scan of the imaging field using a low-magnification objective also aids retracing of mammary regions of interest.
24. Before E14, there are no obvious morphological differences between female and male mammary embryonic buds. To distinguish the embryos' sex, perform a PCR using the following primer sequences and cycling conditions:

    5′- TGGATGGTGTGGCCAATG -3′, 3′- CACCTG CACGTTGCCCTT-5′

    94 °C for 2 min; then 35 cycles of: 94 °C for 30 s, 58 °C for 30 s, 72 °C for 30 s; finally 72 °C for 5 min.

Run PCR products on a 2% agarose gel. One band is observed for female embryos and two bands for male embryos.

25. Avoid cutting the tissue into small pieces as this will make the next steps challenging.
26. To exchange the medium hereafter, use pipette tips with narrow orifices (e.g., gel loading tips) to carefully remove the medium without contacting mammary tissues.
27. Humidified conditions are required to minimize evaporation of the culture medium during imaging.
28. Of a number of tissue clearing techniques tested in the mammary gland, SeeDB and CUBIC-based protocols provided optimal results [11]. These protocols have subsequently been further developed [48, 49], although they have yet to be tested in mammary tissues. In the event of issues with penetration and staining performance of some antibodies, other tissue clearing protocols are available [46], including FUnGI, a method recently developed for the mammary gland [50].
29. Well-spread tissues are thinner and easier to render transparent. Fibers from card/paper can transfer to tissues after fixation, hampering imaging from that side. Using Tetra Pack card (e.g., milk carton card) overcomes this issue. Alternatively, foam biopsy pads may be used.
30. Fixation time depends on the thickness and size of tissues, and may require antibody-specific optimization. Thinner or smaller tissue pieces and embryonic tissues may be fixed after only 2–3 h at room temperature. A ~10:1 ratio v/v of NBF to tissue is required for adequate fixation. Alternatively, fresh solutions of 4% PFA (for 2–4 h at room temperature) may be used. In our hands, 10% NBF is compatible with several genetically encoded fluorescent proteins, including TdTomato, EGFP, YFP, RFP, and CFP in *Rosa26-mTmG*, *R26R-Confetti* and *Rosa26-tdTomato* reporter mouse strains [11, 32, 33]. If quenching of fluorescent proteins is suspected, consider using methanol-free formaldehyde solutions.
31. While best to process samples immediately after harvesting, fixed tissues can be stored at 4 °C in PBS containing 0.05% (w/v) sodium azide for up to 8 weeks. Cutting samples into large pieces may improve antibody penetration and immunostaining, in addition to allowing more immunostainings to be performed in tissues harvested from the same mouse.
32. CUBIC clearing is also compatible with wholemount immunohistochemistry using HRP-conjugated secondary antibodies and horseradish peroxidase-3,3-diaminobenzidine detection, in addition to the detection of β-glucosidase expression (a magenta histochemical stain) (*see* [11, 32]).

33. Tissues may be mounted between coverslips using iSpacer chambers or concave glass microscope slides. However, in these contexts, samples can be difficult to adjust or repositioned for optimal illumination, making working distance a limiting factor in image acquisition [11]. To mitigate this, sample thickness must be closely matched to the thickness of the iSpacer, or concave chamber, and/or specialized imaging objectives with long-working distances must be used. Alternative refractive index matching solutions may also be considered.

Acknowledgments

We are grateful to members of the Fre laboratory, particularly Dr Silvia Fre (Institut Curie), for support and constructive discussions. Protocols for optical tissue clearing and 3D imaging in the mammary gland were developed in collaboration with Dr Felicity Davis (Mater Research Institute, University of Queensland) and Professor Christine Watson (University of Cambridge). We thank Marie Irondelle and Lucie Sengmanivong (Institut Curie) for intravital imaging support. We are also grateful to the Institut Curie In Vivo Experimental Facility for help in the maintenance and care of mouse colonies. We acknowledge the Cell and Tissue Imaging Platform (PICT-IBiSA) of the Genetics and Developmental Biology Department (UMR3215/U934) and the Nikon Imaging Centre of Institut Curie, member of the French National Research Infrastructure France-BioImaging (ANR10-INBS-04). B.L-L is supported by the Academy of Medical Sciences and the University of Bristol (Vice-Chancellor's Research Fellowship). C.C is supported by funding from the European Union's Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No 666003.

Ethics Statement *All animal experimentation were carried out in accordance with the Animal (Scientific Procedures) Act 1986/609 and with local ethics committee approval.*

References

1. Neville MC (2009) Classic studies of mammary development and milk secretion: 1945-1980. J Mammary Gland Biol Neoplasia 14:193–197
2. Lloyd-Lewis B (2020) Multidimensional imaging of mammary gland development: a window into breast form and function. Front Cell Dev Biol 8:203
3. Follain G, Mercier L, Osmani NL et al (2017) Seeing is believing-multi-scale spatio-temporal imaging towards in vivo cell biology. J Cell Sci 130:23. https://doi.org/10.1242/jcs.189001
4. Conchello JA, Lichtman JW (2005) Optical sectioning microscopy. Nat Methods 2:920–931
5. Helmchen F, Denk W (2005) Deep tissue two-photon microscopy. Nat Methods 2:932–940

6. Dunn KW, Young PA (2006) Principles of multiphoton microscopy. Nephron Exp Nephrol 103:e33–e40. https://doi.org/10.1159/000090614
7. Lloyd-Lewis B, Harris OB, Watson CJ, Davis FM (2017) Mammary stem cells: premise, properties, and perspectives. Trends Cell Biol 27:556
8. Fumagalli A, Bruens L, Scheele CLGJ, van Rheenen J (2019) Capturing stem cell behavior using intravital and live cell microscopy. Cold Spring Harb Perspect Biol 12:a035949. https://doi.org/10.1101/cshperspect.a035949
9. Zomer A, Ellenbroek SIJ, Ritsma L et al (2013) Brief report: intravital imaging of cancer stem cell plasticity in mammary tumors. Stem Cells 31:602–606. https://doi.org/10.1002/stem.1296
10. Ellenbroek SIJ, Van Rheenen J (2014) Imaging hallmarks of cancer in living mice. Nat Rev Cancer 14:406–418
11. Lloyd-Lewis B, Davis FM, Harris OB et al (2016) Imaging the mammary gland and mammary tumours in 3D: optical tissue clearing and immunofluorescence methods. Breast Cancer Res 18:127. https://doi.org/10.1186/s13058-016-0754-9
12. Suijkerbuijk SJE, van Rheenen J (2017) From good to bad: intravital imaging of the hijack of physiological processes by cancer cells. Dev Biol 428:328–337
13. Pittet MJ, Weissleder R (2011) Intravital imaging. Cell 147:983–991. https://doi.org/10.1016/j.cell.2011.11.004
14. Ewald AJ, Werb Z, Egeblad M (2011) Preparation of mice for long-term intravital imaging of the mammary gland. Cold Spring Harb Protoc 6:pdb.prot5562. https://doi.org/10.1101/pdb.prot5562
15. Ewald AJ, Werb Z, Egeblad M (2011) Dynamic, long-term in vivo imaging of tumor-stroma interactions in mouse models of breast cancer using spinning-disk confocal microscopy. Cold Spring Harb Protoc 6:pdb.top97. https://doi.org/10.1101/pdb.top97
16. Kedrin D, Gligorijevic B, Wyckoff J et al (2008) Intravital imaging of metastatic behavior through a mammary imaging window. Nat Methods 5:1019–1021. https://doi.org/10.1038/nmeth.1269
17. Alieva M, Ritsma L, Giedt RJ et al (2014) Imaging windows for long-term intravital imaging: general overview and technical insights. Intravital 3:e29917. https://doi.org/10.4161/intv.29917
18. Gligorijevic B, Kedrin D, Segall JE et al (2009) Dendra2 photoswitching through the mammary imaging window. J Vis Exp. https://doi.org/10.3791/1278
19. Jacquemin G, Benavente-Diaz M, Djaber S, Bore A, Dangles-Marie V, Surdez D, Tajbakhsh S, Fre S, Lloyd-Lewis B (2021) Longitudinal high-resolution imaging through a flexible intravital imaging window. Science Advances 7(25) https://doi.org/10.1126/sciadv.abg7663
20. Messal HA, van Rheenen J, Scheele CLGJ (2021) An Intravital Microscopy Toolbox to Study Mammary Gland Dynamics from Cellular Level to Organ Scale. J Mammary Gland Biol Neoplasia 26, 9–27
21. Ritsma L, Steller EJA, Ellenbroek SIJ et al (2013) Surgical implantation of an abdominal imaging window for intravital microscopy. Nat Protoc 8:583. https://doi.org/10.1038/nprot.2013.026
22. Ritsma L, Steller EJA, Beerling E et al (2012) Intravital microscopy through an abdominal imaging window reveals a pre-micrometastasis stage during liver metastasis. Sci Transl Med 4:158ra145. https://doi.org/10.1126/scitranslmed.3004394
23. Kratochwil K (1969) Organ specificity in mesenchymal induction demonstrated in the embryonic development of the mammary gland of the mouse. Dev Biol 20:46–71. https://doi.org/10.1016/0012-1606(69)90004-9
24. Hens JR, Dann P, Zhang JP et al (2007) BMP4 and PTHrP interact to stimulate ductal outgrowth during embryonic mammary development and to inhibit hair follicle induction. Development 134:1221–1230. https://doi.org/10.1242/dev.000182
25. Voutilainen M, Lindfors PH, Mikkola ML (2013) Protocol: ex vivo culture of mouse embryonic mammary buds. J Mammary Gland Biol Neoplasia 18:239–245. https://doi.org/10.1007/s10911-013-9288-2
26. Scheele CLGJ, Hannezo E, Muraro MJ et al (2017) Identity and dynamics of mammary stem cells during branching morphogenesis. Nature 542:313–317. https://doi.org/10.1038/nature21046
27. Wuidart A, Ousset M, Rulands S et al (2016) Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. Genes Dev 30:1261–1277. https://doi.org/10.1101/gad.280057.116
28. Lilja AM, Rodilla V, Huyghe M et al (2018) Clonal analysis of Notch1-expressing cells reveals the existence of unipotent stem cells

that retain long-term plasticity in the embryonic mammary gland. Nat Cell Biol 20: 677–687. https://doi.org/10.1038/s41556-018-0108-1

29. Wuidart A, Sifrim A, Fioramonti M et al (2018) Early lineage segregation of multipotent embryonic mammary gland progenitors. Nat Cell Biol 20:666–676. https://doi.org/10.1038/s41556-018-0095-2
30. Ke M-T, Fujimoto S, Imai T (2013) SeeDB: a simple and morphology-preserving optical clearing agent for neuronal circuit reconstruction. Nat Neurosci 16:1154–1161. https://doi.org/10.1038/nn.3447
31. Susaki EA, Tainaka K, Perrin D et al (2014) Whole-brain imaging with single-cell resolution using chemical cocktails and computational analysis. Cell 157:726–739. https://doi.org/10.1016/j.cell.2014.03.042
32. Davis FM, Lloyd-Lewis B, Harris OB et al (2016) Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. Nat Commun 7:13053. https://doi.org/10.1038/ncomms13053
33. Lloyd-Lewis B, Davis FM, Harris OB et al (2018) Neutral lineage tracing of proliferative embryonic and adult mammary stem/progenitor cells. Development 145:dev164079. https://doi.org/10.1242/dev.164079
34. Elias S, Morgan MA, Bikoff EK, Robertson EJ (2017) Long-lived unipotent Blimp1-positive luminal stem cells drive mammary gland organogenesis throughout adult life. Nat Commun 8:1714. https://doi.org/10.1038/s41467-017-01971-w
35. Hitchcock JR, Hughes K, Harris OB, Watson CJ (2019) Dynamic architectural interplay between leucocytes and mammary epithelial cells. FEBS J 287:250. https://doi.org/10.1111/febs.15126
36. Stewart TA, Hughes K, Hume DA, Davis FM (2019) Developmental stage-specific distribution of macrophages in mouse mammary gland. Front Cell Dev Biol 7:250. https://doi.org/10.3389/fcell.2019.00250
37. Stevenson AJ, Vanwalleghem G, Stewart TA et al (2019) Multiscale activity imaging in the mammary gland reveals how oxytocin enables lactation. bioRxiv:657510. https://doi.org/10.1101/657510
38. Barker N, Van Es JH, Kuipers J et al (2007) Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature 449: 1003–1007. https://doi.org/10.1038/nature06196
39. Muzumdar MD, Tasic B, Miyamichi K et al (2007) A global double-fluorescent Cre reporter mouse. Genesis 45:593–605. https://doi.org/10.1002/dvg.20335
40. Masedunskas A, Milberg O, Porat-Shliom N et al (2012) Intravital microscopy: a practical guide on imaging intracellular structures in live animals. BioArchitecture 2:143–157. https://doi.org/10.4161/bioa.21758
41. Ewald AJ, Werb Z, Egeblad M (2011) Monitoring of vital signs for long-term survival of mice under anesthesia. Cold Spring Harb Protoc 6:pdb.prot.5563. https://doi.org/10.1101/pdb.prot5563
42. Snippert HJ, van der Flier LG, Sato T et al (2010) Intestinal crypt homeostasis results from neutral competition between symmetrically dividing Lgr5 stem cells. Cell 143: 134–144. https://doi.org/10.1016/j.cell.2010.09.016
43. Ritsma L, Vrisekoop N, van Rheenen J (2013) In vivo imaging and histochemistry are combined in the cryosection labelling and intravital microscopy technique. Nat Commun 4:2366. https://doi.org/10.1038/ncomms3366
44. Watson CJ, Khaled WT (2008) Mammary development in the embryo and adult: a journey of morphogenesis and commitment. Development 135:995–1003. https://doi.org/10.1242/dev.005439
45. Panchal H, Wansbury O, Howard BA (2010) Embryonic mammary anlagen analysis using immunolabelling of whole mounts. Methods Mol Biol 585:261–270. https://doi.org/10.1007/978-1-60761-380-0_18
46. Matryba P, Kaczmarek L, Gołąb J (2019) Advances in ex situ tissue optical clearing. Laser Photonics Rev 13:1800292. https://doi.org/10.1002/lpor.201800292
47. Perrin L, Bayarmagnai B, Gligorijevic B (2019) Frontiers in intravital multiphoton microscopy of cancer. Cancer Rep 3:1192. https://doi.org/10.1002/cnr2.1192
48. Tainaka K, Murakami TC, Susaki EA et al (2018) Chemical landscape for tissue clearing based on hydrophilic reagents. Cell Rep 24: 2196–2210.e9. https://doi.org/10.1016/j.celrep.2018.07.056
49. Ke MT, Nakai Y, Fujimoto S et al (2016) Super-resolution mapping of neuronal circuitry with an index-optimized clearing agent. Cell Rep 14:2718–2732. https://doi.org/10.1016/j.celrep.2016.02.057
50. Rios AC, Capaldo BD, Vaillant F et al (2019) Intraclonal plasticity in mammary tumors revealed through large-scale single-cell resolution 3D imaging. Cancer Cell 35:618–632.e6. https://doi.org/10.1016/J.CCELL.2019.02.010

Annex 2

Cell Tracking in 3D using deep learning segmentations

Varun Kapoor‡∗, Claudia Carabaña‡

◆

Abstract—Live-cell imaging is a highly used technique to study cell migration and dynamics over time. Although many computational tools have been developed during the past years to automatically detect and track cells, they are optimized to detect cell nuclei with similar shapes and/or cells not clustering together. These existing tools are challenged when tracking fluorescently labelled membranes of cells due to cell's irregular shape, variability in size and dynamic movement across Z planes making it difficult to detect and track them. Here we introduce a detailed analysis pipeline to perform segmentation with accurate shape information, combined with BTrackmate, a customized codebase of popular ImageJ/Fiji software Trackmate, to perform cell tracking inside the tissue of interest. We developed VollSeg, a new segmentation method able to detect membrane-labelled cells with low signal-to-noise ratio and dense packing. Finally, we also created an interface in Napari, an Euler angle based viewer, to visualize the tracks along a chosen view making it possible to follow a cell along the plane of motion. Importantly, we provide a detailed protocol to implement this pipeline in a new dataset, together with the required Jupyter notebooks. Our codes are open source available at [Git].

Index Terms—3D segmentation, cell tracking, deep learning, irregular shaped cells, fluorescent microscopy.

Introduction

Live-cell imaging is a highly used technique to study cell migration and dynamics over time. The image analysis workflow of volumetric (3D) imaging of cells via fluorescence microscopy starts with an accurate detection and segmentation of cells followed by cell tracking and track analysis. Broadly speaking the task of segmentation can be separated into semantic segmentation (classifying pixels as background or pixels belonging to the cell) or instance segmentation (classifying pixels belonging to individual cells by assigning a unique label to each cell). Segmentation is complicated due to presence of multiple objects in the image, overlapping object pixels and non-homogeneous intensity distribution. Several methods have been proposed for such automated detection and segmentation tasks such as the traditional intensity based thresholding, watershed transform [BM18] and of recent machine learning methods based on random-forest classifiers and support vector machines [BKK+19]. It was shown in [RHH20] that conventional computer vision and machine learning based techniques alone will almost always lead to sub-optimal segmentation and that methods based on deep learning have improved the accuracy of segmentation for natural and biomedical images alike. For the purpose of semantic segmentation U-Net [RFB15] has emerged as the most widely used network for biological applications. This network also forms the backbone of another successful network to do cell nuclei segmentation in 3D, Stardist [SWBM18] [WSH+20]. Stardist directly predicts a shape representation as star-convex polygons for cell nuclei in 2D and 3D. However, cell membrane segmentation is especially challenging as opposed to nuclei segmentation due to fewer boundary pixels and the need to separate touching cells. To predict cell contours together with cell centroids, Eschweiler et al. proposed a 3D U-Net network using centroids as seeds for watershed in 3D confocal microscopy images [ESC+18]. The drawback of this approach is misclassification due to sub-optimal seeding. Another approach proposed by Wolny et al., is to directly predict the cell boundaries using a 3D U-Net followed by a volume partitioning algorithm to segment each cell based on the boundary prediction [WCV+20]. This approach requires well defined cell boundaries, which may create segmentation errors in low signal-to-noise imaging conditions.

To address the issues with existing segmentation algorithms just described, we developed Vollseg. In brief we use Stardist in 3D to obtain a star convex shape approximation for the cells and extract the cell centroids from these polygons. We also train a 3D U-Net model to obtain a semantic segmentation map of the cells. We then perform a marker controlled watershed on the probability map of Stardist using the U-Net segmentation as a mask image to prevent the overflow of segmentation regions. To avoid the error of sub-optimal seeding we developed a seed pooling approach taking advantage of strength of both the Stardist and U-Net networks. We benchmark our segmentation result on a challenging dataset comprised of epithelial cells of mouse embryonic mammary glands with membrane labelling. These cells are highly irregular in shape and have a low signal-to-noise ratio to obtain an accurate segmentation only based on the boundary information. Using this dataset, we obtain different metrics showing that our approach is able to obtain shape approximation for the overlapping cells that go beyond the star convex shape. The complete segmentation pipeline is illustrated in Figure 1.

For analysis of the cell migration behavior we need to reliably track the cells and obtain certain attributes such as signal intensity or changes over time of the distance between the cells and tissue boundary. Cell tracking is challenging due to erratic volumetric motion, occlusion and cell divisions. Tracking using only the

---

∗ *Corresponding author: varun.kapoor@curie.fr*
‡ *Institut Curie, PSL Research University, Sorbonne University, CNRSUMR 3215, INSERM U934, Genetics and Developemental Biology, Paris, France.*

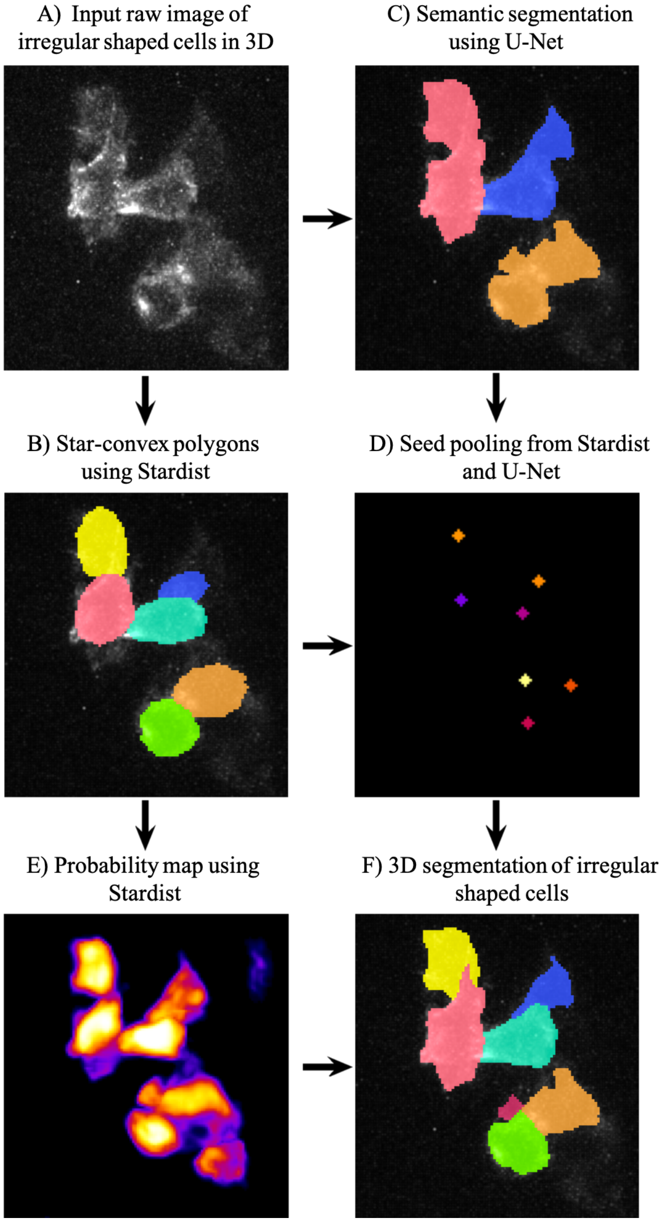


Fig. 1: *Schematic representation showing the segmentation approach used in VollSeg. First, we input the raw fluorescent image in 3D (A) and pre-process it to remove noise. Next, we obtain the star convex approximation to the cells using Stardist (B) and the U-Net prediction labelled via connected components (C). We then obtain seeds from the centroids of labelled image in B, for each labelled region of C in order to create bounding boxes and centroids. If there is no seed from B in the bounding box region from U-Net, we add the new centroid (shown in yellow) to the seed pool (D). Finally, we do a marker controlled watershed in 3D using skimage implementation on the probability map shown in (E) to obtain the final cell segmentation result (F). All images are displayed in Napari viewer with 3D display view.*

centroid information may lead to wrong cell assigements, hence we need to include other cell attributes such as the shape and intensity information while making the links between the cells in successive time frames. Trackmate is a popular tracking software that uses customizable cost matrix for solving the linear assignment problem and uses Jaqman linker as a second step to link segments of dividing and merging cells [TPS^{+}17]. In this paper, we introduce BTrackmate, a Fiji/ImageJ plugin to track the previously segmented cells. The major advantage of BTrackmate is the ability to track the cells inside a tissue. It allows the input of the cell and tissue segmentation image files and/or a csv file of the cell attributes. Furthermore, we also add some biological context in the tracking process where after segment linking is done a track inspector removes segments that are shorter than a user defined time length. Such short segments are unlikely to be true division events if they are too short and manually removing them can be tedious when many tracks are present. The users can choose this parameter in time units and can set it to 0 if removing such short segments is not required.

Finally, the tracking results obtained with BTrackmate are saved as an xml file that can be re-opened in an Euler angle based viewer in python called Napari, allowing volumetric viewing of the tracked cells using the track layer feature [UVCL20]. We made a python package called napatrackmater to export the track xml file as tracks layer in Napari for dividing and non-dividing tracks. We provide a customized Napari widget to view selected tracks and obtain their cell migration attributes.

Material and Methods

Preparation of the dataset

We used fluorescent microscopy images of mouse embryonic mammary glands stabilized in an ex vivo culture previously collected in the laboratory of Dr. S. Fre at Institut Curie. All images were acquired with an inverted confocal laser scanning microscope (e.g. Zeiss LSM780/880) equipped with long-working distance objectives to acquire high-resolution 3D image stacks. We acquired images of pixel size (22, 512, 512) with calibration of (3, 0.52, 0.52) micrometer. The quality at which these images are acquired is determined by the spatial resolution of the used optical device, desired temporal resolution, duration of the experiment and depth of the acquired Z-stacks. We perform unsupervised image denoising [KBJ19] on our dataset, an algorithm we chose based on its performance compared to other methods [Ric72], [Luc74]. Post-restoration of the 3D images, we developed a method to perform the segmentation of the cells using deep learning techniques. We created a training dataset with hand drawn segmentation of 14 Z-stacks. We performed data augmentation on the microscopy images by denoising, adding Poisson and Gaussian noise, random rotations and flips to create 700 Z-stacks. We chose a patch size of (16, 128, 128) and created 11,264 patches for training Stardist and U-Net network. For the Stardist network we chose 192 rays to have a better shape resolution for the irregular shaped cells.

Parameter Setting

Stardist predicts object instances based on probability threshold and non maximal suppression threshold to merge overlapping predictions. These parameters can be automatically determined using the optimize threshold program that we provide with the segmentation package. Higher values of the probability threshold yield fewer object instances, but avoids false positives. Higher values of the overlap threshold would lead to oversegmentation. We used 32 Z-stacks to determine the optimal parameters of probability threshold of 0.76 and non maximal suppression threshold of 0.3.

Segmentation

As illustrated in Figure 1, we first obtain the centroids of the star convex approximated cell shapes and create a seed pool with these centroid locations. Even with the optimized threshold values we find that the seeds can be sub-optimal as many cells instances with low signal are missed. In order to make the seed pool optimal we use the U-Net prediction to obtain a binary image of semantic segmentation, perform connected component analysis to label the image and obtain bounding boxes (computed using scikit-image [vdWSN+14], version 0.18.x) for each label in 3D. For each bounding box we search for a seed from the Stardist predicted seed pool. If a Stardist seed is found inside the bounding box, the centroid of the U-Net predicted bounding box is rejected else the centroid is added to the seed pool to make a complete set of seeds that we use to start a watershed process in 3D. We use the probability map of Stardist to start the watershed process to obtain a better shape approximation for the irregular shaped cells that goes beyond the star convex shape.

The code for the merging of U-Net and Stardist seeds is the following:

```
def iou3D(box_unet, centroid_star):

  ndim = len(centroid_star)
  inside = False

  Condition = [Conditioncheck(centroid_star, box_unet,
   p, ndim)
   for p in range(0,ndim)]

  inside = all(Condition)

  return inside

def Conditioncheck(centroid_centroid, box_unet,
       p, ndim):

  condition = False

  if centroid_star[p] >= box_unet[p]
  and centroid_star[p] <= box_unet[p + ndim]:

       condition = True

  return condition
```

The code for doing watershed in 3D using the complete set of seeds on the probability map of Stardist is the following:

```
def WatershedwithMask3D(Image, Label, mask, grid):
  #Image = ProbabilityMap of Stardist
  #Label = Label segmentation image of Stardist
  #Mask = U-Net predicted image post binarization
  properties = measure.regionprops(Label, Image)
  binaryproperties =
  measure.regionprops(label(mask), Image)
  Coordinates = [prop.centroid for prop in properties]
  BinaryCoordinates = [prop.centroid for
  prop in binaryproperties]
  Binarybbox =
  [prop.bbox for prop in binaryproperties]
  Coordinates = sorted(Coordinates ,
  key=lambda k: [k[0], k[1], k[2]])

  if len(Binarybbox) > 0:
       for i in range(0, len(Binarybbox)):

          box = Binarybbox[i]
          inside = [iou3D(box, star)
          for star in Coordinates]

          if not any(inside) :
                Coordinates.append(BinaryCoordinates[i])

  Coordinates.append((0,0,0))
  Coordinates = np.asarray(Coordinates)
  coordinates_int = np.round(Coordinates).astype(int)

  markers_raw = np.zeros_like(Image)
  markers_raw[tuple(coordinates_int.T)] = 1
  + np.arange(len(Coordinates))
  markers = morphology.dilation(
  markers_raw.astype('uint16'), morphology.ball(2))

  watershedImage = watershed(-Image, markers,
  mask = mask.copy())
  return watershedImage, markers
```

Performance Metrics

Accuracy of segmentation results is assesed by comparing the obtained labels to the ground truth (GT) labels. The most commonly used metric is to compute intersection over union (IOU) score between the predicted and the GT label image. We define GT, labels and IOU score as:

$GT = \{gt\}$, $SEG = \{seg\}$ are two sets of segmented objects.

$IOU(a,b)$ is the value of the IOU operation between two segmented objects a and b.

A threshold score value $\tau \in [0,1]$ is used to determine the true positive (TP), false positives (FP) and false negatives (FN) defined as:

$$TP = \{seg \in SEG, \exists\ gt \in GT,\ IOU(gt,seg) > \tau\}$$

$$FP = \{seg \in SEG, \forall\ gt \in GT,\ IOU(gt,set) < \tau\}$$

$$FN = \{gt \in GT, \forall\ seg \in SEG,\ IOU(gt,seg) < \tau\}$$

We use the Stardist implementation to compute accuracy scores which uses the hungarian method (scipy implementation) [Kuh55] to compute an optimal matching to do a one to one assingement of predicted label to GT labels. This implementation avoids finding multiple TP for a given instance of GT. We also compute precision (TP/(TP + FP)), recall (TP / (TP + FN)), F1 score (geometric mean of precision and recall) and accuracy score $AP_\tau = \frac{TP_\tau}{TP_\tau + FP_\tau + FN_\tau}$. To evaluate the accuracy of our method in resolving the shape of the cells we compute the mean squared error (MSE) and structural similarity index measurment (SSIM) between the GT and obtained segmentation images post-binarization operation on the obtained instance segmentation maps. MSE shows a low score if the image is structurally closer to GT. SSIM score is higher if the two images are structurally more similar to each other.

Detailed Procedure

The software package we provide comes with training and prediction notebooks for training the base U-Net and Stardist networks on your own dataset. We provide jupyter notebooks to do so on local GPU servers and also on Google Colab.

Network Training: In the first Jupyter notebook we create the dataset for U-Net and Stardist training. In the first cell of the notebook the required parameters are the path to your data that contains the folder of Raw and Segmentation images to create training pairs. Also to be specified is the name of the generated npz file along with the model directory to store the h5 files of the trained model and the model name.

```
Data_dir = '/data/'
NPZ_filename = 'VollSeg'
Model_dir = '/data/'
Model_Name = 'VollSeg'
```

The model parameters are specified in the next notebook cell. These parameters are described as follows:

1) NetworkDepth = Depth of the network, with each increasing depth the image is downsampled by 2 hence the XYZ dimension of the data / 2^depth has to be greater than 1.
2) Epochs: training for longer epochs ensures a well converged network and requires longer GPU runtimes.
3) Learning rate is the parameter which controls the step size used in the optimization process and it should not be greater than 0.001 at the start of the training.
4) Batch size controls the number of images used for doing stochastic gradient descent and is a parameter limited by the GPU memory available, batch size < 10 should be optimal.
5) Patch X, Y, Z is the size used for making patches out of the image data. The original image is broken down into patches for training. Patch size is chosen based on having enough context for the network to learn the details at different scales.
6) Kernel is the receptive field of the neural network, usual choices are 3, 5 or 7. This is the size of the convolutional kernel used in the network.
7) n_patches_per_image is the number of patches sampled for each image to create the npz file, choose an optimal value so that the file fits in the RAM memory.
8) Rays stand for the number of rays used to learn the distance map, low rays decreases the spatial resolution and high rays are able to resolve the shape better.
9) use_gpu_opencl is a boolean parameter that is set true if you want to do some opencl computations on the GPU, this requires GPU tools python package.
10) Before starting the U-Net training an npz file containing the paried Raw and Binary segmentation images needs to be created, by setting GenerateNPZ = True such a file is created.
11) If there are multiple GPU's available, the training of U-Net and Stardist can be split between the GPU's. Set TrainUNET = True for training a U-Net network, create a copy of the notebook and only set TrainSTAR = True for training a Stardist network. If there are no multiple GPU's available, set all of these parameters in 10) and 11) to be True to create and train both the networks in a single notebook run.

The code to set the parameters is the following:

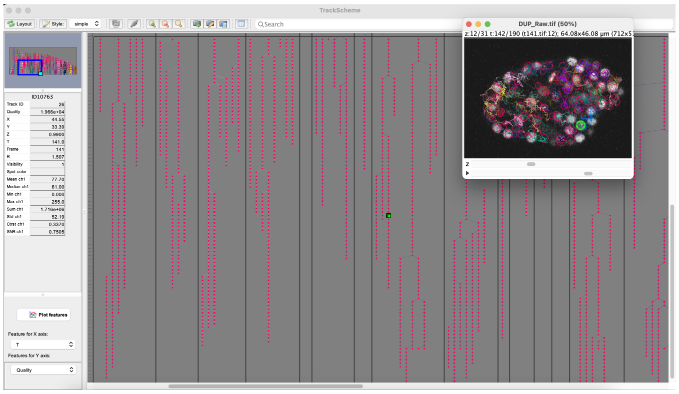
```
#Network training parameters
NetworkDepth = 3
Epochs = 100
LearningRate = 1.0E-4
batch_size = 5
PatchX = 128
PatchY = 128
PatchZ = 16
Kernel = 3
n_patches_per_image = 16
Rays = 192
startfilter = 48
use_gpu_opencl = True
GenerateNPZ = True
TrainUNET = False
TrainSTAR = False
```

After the network has been trained it will save the configuration files of the training for both the networks along with the weight vector file as h5 files that will be used by the prediction notebook. For running the network prediction on XYZ shape images use the prediction notebook either locally or on Colab. In this notebook you only have to specify the path to the image and the model directory. The only two parameters to be set here are the number of tiles (for creating image patches to fit in the GPU memory) and min_size in pixel units to discard segmented objects below that size. We perform the watershed operation on the probability map as a default. However, this operation can also be changed to use the distance map coming out of Stardist prediction instead by setting 'UseProbability' variable to false. The code below operates on a directory of XYZ shape images:

```
ImageDir = 'data/tiffiles/'
Model_Dir = 'data/'
SaveDir = ImageDir + 'Results/'
UNETModelName = 'UNETVollSeg'
StarModelName = 'VollSeg'
NoiseModelName = 'NoiseVoid'

UnetModel = CARE(config = None,
name = UNETModelName,
basedir = Model_Dir)
StarModel = StarDist3D(config = None,
name = StarModelName,
basedir = Model_Dir)
NoiseModel = N2V(config=None,
name=NoiseModelName,
basedir=Model_Dir)

Raw_path =
os.path.join(ImageDir, '*.tif')
filesRaw =
glob.glob(Raw_path)
filesRaw.sort
min_size = 50
n_tiles = (1,1,1)
for fname in filesRaw:

     SmartSeedPrediction3D(ImageDir,
     SaveDir, fname,
     UnetModel, StarModel, NoiseModel,
     min_size = min_size,
     n_tiles = n_tiles,
     UseProbability = False)
```

Tracking

After we obtain the segmentation using VollSeg, we create a csv file of the cell attributes that include their location, size and volume inside a region of interest. For large datasets memory usage could be of concern while loading the images into memory, hence inputs via csv could prove helpful. Tracking is performed in ImageJ/Fiji, an image processing package. We developed our code over the existing tracking solution called Trackmate [TPS+17]. Trackmate uses linear assignment problem (LAP) algorithm to do linking of the cells and uses Jaqman linker for linking the segments for dividing and merging trajectories. It also provides other trackers such as the Kalman filter to do tracking of non-dividing cells.



DUP_Raw.tif (50%)
z:12/31 t:142/190 (t141.tif:12); 64.08x46.08 μm (712x5

ID10763	
Track ID	26
Quality	1.966e+04
X	44.55
Y	33.39
Z	0.9900
T	141.0
Frame	141
R	1.507
Visibility	1
Spot color	
Mean ch1	77.70
Median ch1	61.00
Min ch1	0.000
Max ch1	255.0
Sum ch1	1.716e+06
Std ch1	52.19
Ctrst ch1	0.3370
SNR ch1	0.7505

Fig. 2: *Trackscheme display for the C. elegans dataset.*

Trackmate comes with a fully interactive track editing interface with graph listener to show the selected cell in the trackscheme and vice versa, to click on the graph and have the selected cell being highlighted in the image, making the process of track editing interactive. Post-editing the tracks are saved as an xml file which can then be loaded back into the program to do more track editing if needed. When a cell divides, the track is splitted up in two tracklets. In order to aid in track editing, we introduced a new parameter of minimum tracklet length to remove tracklets in a track that are short in the time dimension. This introduces a biological context of not having very short trajectories, reducing the track editing effort to correct for the linking mistakes made by the program. For testing our tracking program we used a freely available dataset from the cell tracking challenge of a developing C. elegans embryo [Cel] [MBB$^+$08]. Using our software we can remove cells from tracking which do not fit certain criteria such as being too small (hence most likely a segmentation mistake) or being low in intensity or outside the region of interest such as when we want to track cells only inside a tissue. For this dataset we kept 12,000 cells and after filtering short tracks kept about 50 tracks with and without division events.

For this dataset the track scheme along with overlayed tracks is shown in Figure 2. Selected node in the trackscheme is highlighted in green and vice versa. Extensive manual for using the track editing is available on ImageJ/Fiji wiki [Tin].

Results

Quantitative Comparisons between Segmentation Methods

We compare our proposed VollSeg segmentation approach to two commonly used methods for cell segmentation of fluorescent microscopy images, 3D Stardist [SWBM18] [WSH$^+$20] and 3D U-Net [RFB15]. A 3D cell rendering using all analyzed segmentation methods is shown in the Figure 3. Stardist in 3D was previously compared to other classical method, the IFT watershed, and it was shown to perform better than the classical method, hence we use Stardist as a baseline for comparison. To assess the performance of our segmentation, we compute the metrics described in material and methods section. VollSeg and Stardist methods perform at comparable accuracy, but higher than U-Net, as shown in Figure 4 A. This is expected, as U-Net can not perform instance segmentation of overlapping cells. In addition, when quantifying the F1-score in Figure 4 B, U-Net obtains the lowest score because it detects less TP segmented pixels in comparision to VollSeg and Stardist as shown in Figure 4 C. However, Stardist has the highest mean squared error as it is unable to detect the irregular shape while U-Net and Vollseg have similar performance, as shown in Figure 5 A. This result can also be seen from structural similarity index measurement, shown in Figure 5 B. In conclusion, VollSeg is able to strength the shape accuracy from U-Net and the ability to separate the overlapping instances from Stardist.

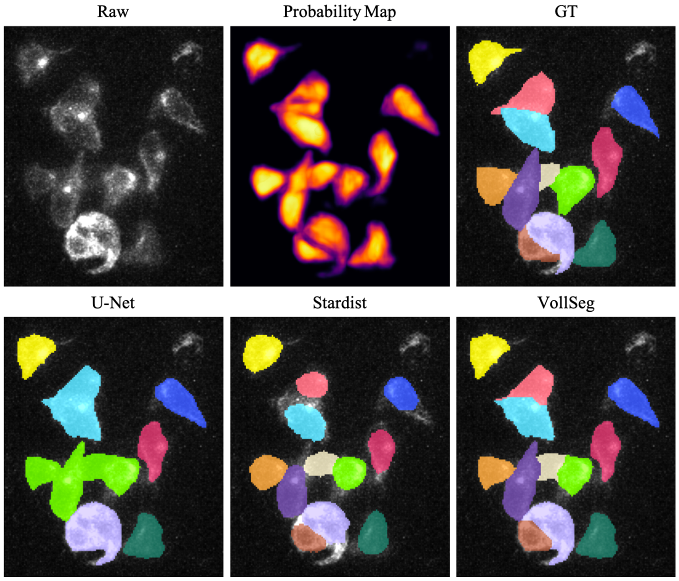


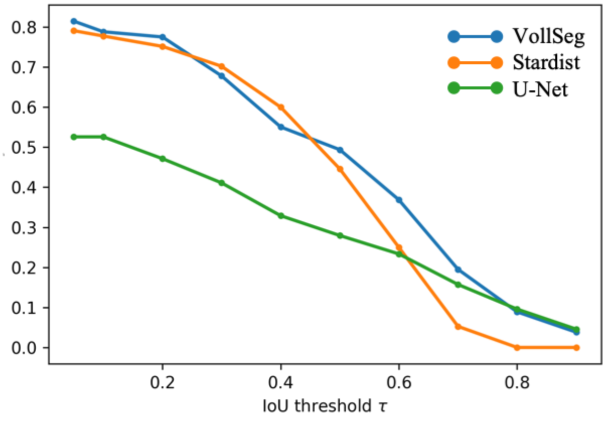
Fig. 3: *Visual 3D segmentation comparison between the Ground truth (GT) image, Stardist, U-Net and VollSeg results. The images are displayed in Napari viewer with 3D display view.*

Track Analysis

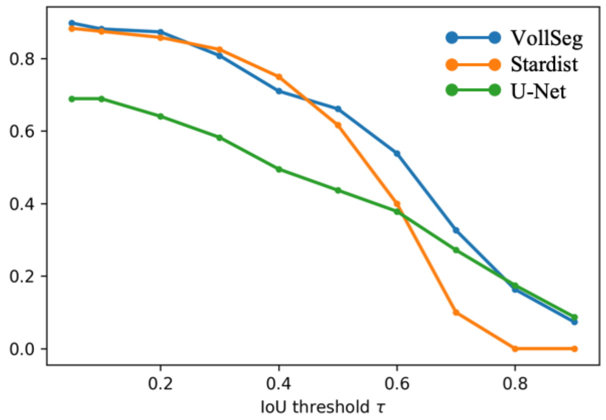
After obtaining the tracks from BTrackmate, we save them as Trackmate xml file, which contains the information about all the cells in a track. Since the cells can be highly erratic in their volumetric motions, we use Napari, an Euler angle based viewer, to visualize such tracks from different reference positions. We made a python package to export the xml files previously saved in ImageJ/Fiji and convert them into the tracks layer of Napari. We made a customised widget based graphic user interface (GUI) to view selected tracks, display the track information and save the cell track along user selected view, as shown in Figure 6 A. On the top left panel, the image and tracks layer properties are displayed and can be changed (1). In the bottom left, there is a dropdown menu enlisting all the tracks (2). Users can select the track to be displayed in the central window and it can be switched between the hyperstack and the 3D view (3). The user can also choose to view all the tracks at once and then toggle the visibilty of the tracks using the eye icon next to the image and tracks layer (4). On the top right panel, we show two plots displaying the track information (5). The 3D central view can be rotated and translated to view the tracks along the plane of motion of the cells and the selected view can be saved as an animation using the bottom right animation panel (6). For the cells that divide we show the intensity variation and associated fast fourier transform for each tracklet.

We provide two example jupyter notebooks with the package. In the first one we compute the cell distance from the tissue boundary change over time for dividing and non-dividing trajectories. The user selects a track of interest and it displays two plots next to the track view that show the distance change over time for the whole track (non-dividing trajectory) and the starting and end

A) Accuracy



B) F1 score



C) True Positive

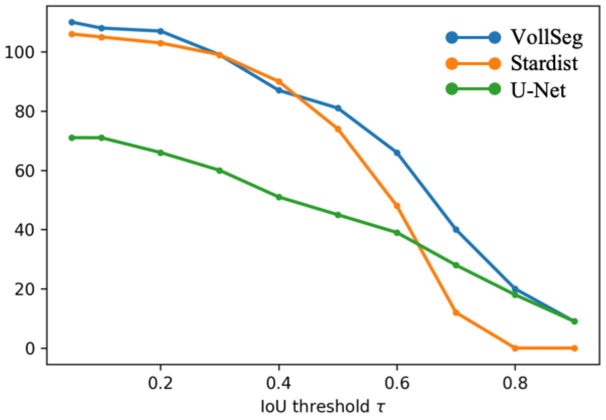
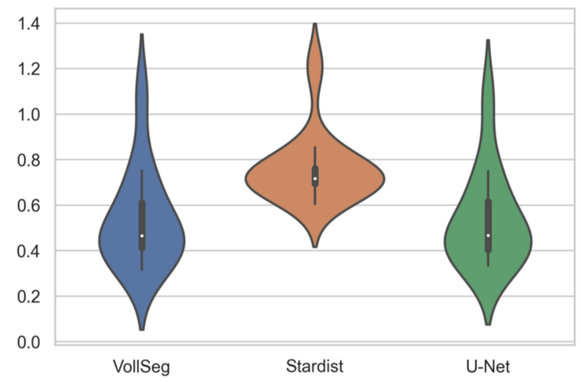


Fig. 4: *Segmentation comparision metrics between VollSeg (in blue), Stardist (in orange) and U-Net (in green). We plot (A) accuracy (as percentage), (B) F1 score (as percentage) and (C) true positive rates (as number of pixels) for all the networks.*

A) Mean Squared error



B) Structural similarity index measurement

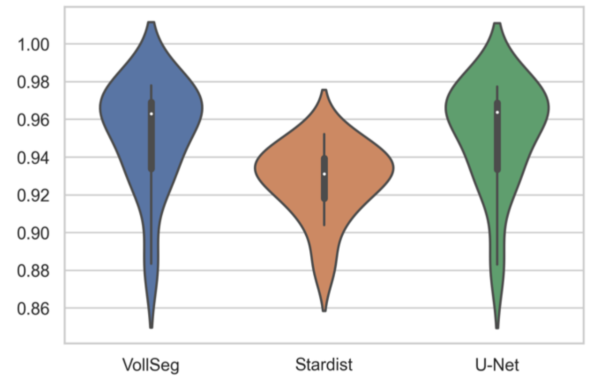
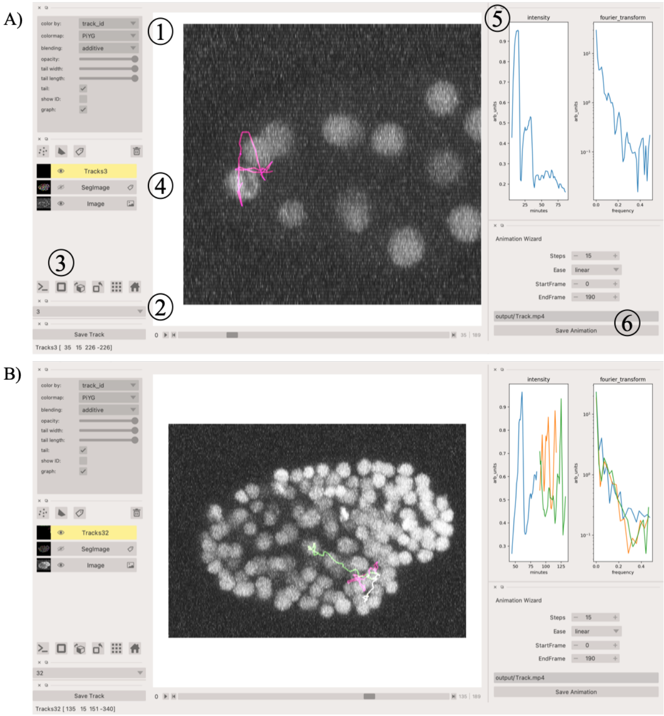


Fig. 5: *We plot Mean Squared error (MSE) (A) and Structural similarity index measurement (SSIM) (B) comparing between VollSeg (in blue), Stardist (in orange) and U-Net (in green).*

location of the cells, as shown in Figure 7. For the tracks with multiple events of cell division we show the distance change over time of each tracklet. In the localization plot the parent tracklet start and end location is shown in green while all the daughter cells start and end locations are shown in red. In the second example notebook, the plots show intensity change in the track over time along with the associated frequency of intensity oscillation present in each tracklet. The frequency associated with each tracklet is computed using the scipy implementation of fast fourier transform. The results of track analysis can be saved as plots, mp4 files of the track animation or csv files.

Conclusions

We have presented a workflow to do segmentation, tracking and track analysis of cells in 3D with irregular shape and intensity distribution. For performing segmentation we developed VollSeg, a jupyter notebook based python package that combines the strengths of semantic and instance deep learning segmentation methods. Post-segmentation we create a csv file containing the information about the cells inside a region of interest which serves as an input to Btrackmate, the ImageJ/Fiji plugin we created for doing the tracking. The tracking software uses existing track editing interface of Trackmate and saves the track information as an xml file. To view and analyze such volumetric tracks we created napatrackmater, a python package to export such trajectories as



A)

color by: track_id
colormap: PiYG
blending: additive
opacity:
tail width:
tail length:
tail: ☑
show ID: ☐
graph: ☑

Tracks3
SegImage
Image

3

Save Track

Tracks3 [35 15 226 -226]

Animation Wizard

Steps 15
Ease linear
StartFrame 0
EndFrame 190

output/Track.mp4

Save Animation

B)

color by: track_id
colormap: PiYG
blending: additive
opacity:
tail width:
tail length:
tail: ☑
show ID: ☐
graph: ☑

Tracks32
SegImage
Image

32

Save Track

Tracks32 [135 15 151 -340]

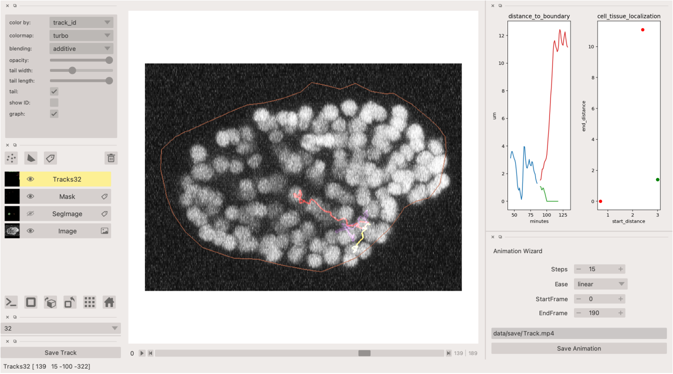
Animation Wizard

Steps 15
Ease linear
StartFrame 0
EndFrame 190

output/Track.mp4

Save Animation

Fig. 6: *Napari widget to view tracks and plot track information in non-dividing trajectories (A) and dividing trajecrtories (B). For the selected track we see the intensity change over time and its associated fast Fourier transform.*



color by: track_id
colormap: turbo
blending: additive
opacity:
tail width:
tail length:
tail: ☑
show ID: ☐
graph: ☑

Tracks32
Mask
SegImage
Image

32

Save Track

Tracks32 [139 15 -100 -322]

Animation Wizard

Steps 15
Ease linear
StartFrame 0
EndFrame 190

data/save/Track.mp4

Save Animation

Fig. 7: *Napari widget to analyze the distance of the cell to the boundary. The left plot displays the distance of the daughter cells to the boundary, while the right plot shows the start and end distance localization of the mother cell (in green) and daughter cells (in red).*

track layer of Napari and we provide jupyter notebook based enviornment for track analysis with two example notebooks.

The tools that we present here can also be useful for segmentation of cells coming from other organisms or imaging modalities (transmitted light and light sheet imaging) as our method can be applied to segment cells that go beyond the star convex polyhedra.

Acknowledgements

We acknowledge the Cell and Tissue Imaging Platform (PICT-IBiSA) of the Genetics and Developmental Biology Department (UMR3215/U934) at Institut Curie, member of the French National Reserch infrastructure France-Bioimaging (ANR-10-INBS-04). We thank specially Olivier Renaud for supporting the software development. We are grateful to Dr Silvia Fre for support and constructive discussions. We thank Leo Guginard for insightful comments about the manuscript. V.K is supported by Labex DEEP at Institut Curie (ANR-11- LBX0044 grant). C.C is supported by funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 666003.

Author Contributions

V.K wrote the code; C.C performed the image acquisition of the used dataset and created labelled training dataset in 3D; V.K and C.C wrote the manuscript.

REFERENCES

[BKK+19] Stuart Berg, Dominik Kutra, Thorben Kroeger, Christoph N. Straehle, Bernhard X. Kausler, Carsten Haubold, Martin Schiegg, Janez Ales, Thorsten Beier, Markus Rudy, Kemal Eren, Jaime I. Cervantes, Buote Xu, Fynn Beuttenmueller, Adrian Wolny, Chong Zhang, Ullrich Koethe, Fred A. Hamprecht, and Anna Kreshuk. ilastik: interactive machine learning for (bio)image analysis. *Nature Methods*, September 2019. doi:10.1038/s41592-019-0582-9.

[BM18] S. Beucher and F. Meyer. The morphological approach to segmentation: The watershed transformation. 2018. doi:10.1201/9781482277234-12.

[Cel] Waterston lab, university of washington, seattle, wa, usa. http://celltrackingchallenge.net/3d-datasets/.

[ESC+18] Dennis Eschweiler, Thiago V. Spina, Rohan C. Choudhury, Elliot Meyerowitz, Alexandre Cunha, and Johannes Stegmaier. Cnn-based preprocessing to optimize watershed-based cell segmentation in 3d confocal microscopy images, 2018. arXiv:1810.06933, doi:10.1109/isbi.2019.8759242.

[Git] https://github.com/kapoorlab/vollseg,
https://github.com/kapoorlab/napatrackmater,
https://github.com/kapoorlab/btrackmate.

[KBJ19] Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2129–2137, 2019. doi:10.1109/cvpr.2019.00223.

[Kuh55] H. W. Kuhn. The hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97, 1955. arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/nav.3800020109, doi:10.1002/nav.3800020109.

[Luc74] L.B. Lucy. An iterative technique for the rectification of observed distributions. *The Astronomical Journal 7*, 79:745, 1974. doi:10.1086/111605.

[MBB+08] John Isaac Murray, Zhirong Bao, Thomas J Boyle, Max E Boeck, Barbara L Mericle, Thomas J Nicholas, Zhongying Zhao, Matthew J Sandel, and Robert H Waterston. Automated analysis of embryonic gene expression with cellular resolution in c. elegans. *Nature Methods*, 5(8):703–709, 2008. doi:10.1038/nmeth.1228.

[RFB15] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In Nassir Navab, Joachim Hornegger, William M. Wells, and Alejandro F. Frangi, editors, *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, pages 234–241, Cham, 2015. Springer International Publishing.

[RHH20] Tobias M. Rasse, Réka Hollandi, and Peter Horvath. Opsef: Open source python framework for collaborative instance segmentation of bioimages. *Frontiers in Bioengineering and Biotechnology*, 8:1171, 2020. doi:10.3389/fbioe.2020.558880.

[Ric72] William Hadley Richardson. Bayesian-based iterative method of image restoration∗. *J. Opt. Soc. Am.*, 62(1):55–59, Jan 1972. doi:10.1364/JOSA.62.000055.

[SWBM18] Uwe Schmidt, Martin Weigert, Coleman Broaddus, and Gene Myers. Cell detection with star-convex polygons. In *Medical Image Computing and Computer Assisted Intervention - MICCAI 2018 - 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part II*, pages 265–273, 2018. doi:10.1007/978-3-030-00934-2_30.

[Tin] Jean-Yves Tinevez. Trackmate manual. https://imagej.net/TrackMate.

[TPS+17] Jean-Yves Tinevez, Nick Perry, Johannes Schindelin, Genevieve M. Hoopes, Gregory D. Reynolds, Emmanuel Laplantine, Sebastian Y. Bednarek, Spencer L. Shorte, and Kevin W. Eliceiri. Trackmate: An open and extensible platform for single-particle tracking. *Methods*, 115:80–90, 2017. Image Processing for Biologists. doi:10.1016/j.ymeth.2016.09.016.

[UVCL20] Kristina Ulicna, Giulia Vallardi, Guillaume Charras, and Alan R. Lowe. Automated deep lineage tree analysis using a bayesian single cell tracking approach. *bioRxiv*, 2020. arXiv:https://www.biorxiv.org/content/early/2020/09/10/2020.09.10.276980.full.pdf, doi:10.1101/2020.09.10.276980.

[vdWSN+14] Stéfan van der Walt, Johannes L. Schönberger, Juan Nunez-Iglesias, François Boulogne, Joshua D. Warner, Neil Yager, Emmanuelle Gouillart, Tony Yu, and the scikit-image contributors. scikit-image: image processing in Python. *PeerJ*, 2:e453, 6 2014. doi:10.7717/peerj.453.

[WCV+20] Adrian Wolny, Lorenzo Cerrone, Athul Vijayan, Rachele Tofanelli, Amaya Vilches Barro, Marion Louveaux, Christian Wenzl, Susanne Steigleder, Constantin Pape, Alberto Bailoni, Salva Duran-Nebreda, George Bassel, Jan U. Lohmann, Fred A. Hamprecht, Kay Schneitz, Alexis Maizel, and Anna Kreshuk. Accurate and versatile 3d segmentation of plant tissues at cellular resolution. *bioRxiv*, 2020. arXiv:https://www.biorxiv.org/content/early/2020/01/18/2020.01.17.910562.full.pdf, doi:10.1101/2020.01.17.910562.

[WSH+20] Martin Weigert, Uwe Schmidt, Robert Haase, Ko Sugawara, and Gene Myers. Star-convex polyhedra for 3d object detection and segmentation in microscopy. In *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, March 2020. doi:10.1109/WACV45572.2020.9093435.

RÉSUMÉ

L'un des mécanismes par lequel le cancer se développe implique la dérégulation des cellules souches. Ainsi, l'étude des mécanismes de développement normal des cellules souches nous permet de comprendre comment ces cellules peuvent contribuer à la formation d'un cancer.

Dans le tissu mammaire, des cellules souches multipotentes, capables de générer plusieurs types cellulaires, existent seulement pendant le développement embryonnaire. Après la naissance, des progéniteurs qui ne peuvent se différencier qu'en un seul type de cellule remplacent les cellules souches.

Au cours de mes études doctorales, j'ai identifié de nouveaux marqueurs pour les cellules souches mammaires, ce qui nous a permis d'obtenir un atlas spatio-temporel de la différenciation des cellules souches au cours du développement mammaire embryonnaire. Nos résultats fournissent également des informations importantes sur la hiérarchie des cellules mammaires stromales, qui jouent un rôle clé dans la morphogenèse du tissu mammaire.

MOTS CLÉS

Cellules souches mammaires, Traçage du lignage, Développement mammaire, Imagerie en temps réel, Séquençage de cellules uniques.

ABSTRACT

Breast cancer is the most prevalent cancer among women in developed countries. One mechanism whereby cancer is initiated is through the dysregulation of stem cell programs, thus understanding the physiological behaviour of mammary stem cells will uncover how and why these cells lose control and participate to cancer initiation. Multipotent stem cells are capable of generating many specialised cell types and in the breast, they are found only during embryonic development. After birth, unipotent progenitors that can only give rise to one cell type ensure tissue homeostasis.

During my PhD studies, I have identified new markers for specific mammary epithelial cells, which allowed us to obtain a spatiotemporal atlas of stem cell differentiation during embryonic mammary morphogenesis. Our results deliver also fundamental insights into the hierarchy of embryonic mammary stromal cell populations, that play key roles in directing epithelial branching morphogenesis.

KEYWORDS

Mammary stem cells, Lineage tracing, Mammary gland development, Time-lapse imaging, Single cell RNA sequencing.