AWARD NUMBER: W81XWH-17-1-0013 (BC160959)

TITLE: Targeting Basal Breast Cancer

PRINCIPAL INVESTIGATOR: Dr. Pamela Cowin, Ph.D.

CONTRACTING ORGANIZATION:

New York University School of Medicine
550 First Avenue
New York, NY 10016

REPORT DATE: SEPTEMBER 2020

TYPE OF REPORT: Annual Year 3

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
SEPTEMBER 2020	Annual Year 3	09.01.19-08.31.20

4. TITLE AND SUBTITLE
Targeting Basal Breast Cancer

5a. CONTRACT NUMBER
W81XWH-17-1-0013

5b. GRANT NUMBER
BC160959

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Pamela Cowin PhD

E-Mail: cowinp01@nyumc.org

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
New York University School of Medicine
550 First Ave
New York 10016

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)

U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT

Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT: We set out to determine 1) if Gpr specifically identifies mammary stem/progenitors; and 2) the significance of Gpr expression in tumors and whether ablating Gpr+ cells would prevent or eradicate them. To test these hypotheses, we proposed to: a) identify, isolate and characterize Gpr+ cells, determine their potency by tracing their progeny, and monitor the effects of ablating them on mammary development; b) determine Gpr expression in human breast cancer, and test if ablating Gpr+ cells affects mammary tumorigenesis in mouse models. In this grant period we have 1) completed analysis of Gpr expression in the embryonic mammary gland; 2) consolidated scRNAseq and immunofluorescence analysis of Gpr;3) completed lineage tracing of the progeny of embryonic, pubertal and pregnancy-induced Gpr+ cells; 4) identified Gpr+ cells as migratory myoepithelial progenitors in other secretory organs; 5) confirmed Gpr association with poor outcome in basal-type breast cancer; 6) shown Gpr+ progenitors in MMTV-Wnt1 tumors retain embryonic mesenchymal features and potency. These results show that Gpr identifies mammary stem cells in the embryonic mammary gland and unipotent progenitors in perinatal and postnatal mammary gland. Gpr+ cells associated with early tumor onset are bipotent and share mesenchymal features with embryonic progenitors at invasive tips of the embryonic rudiment.

15. SUBJECT TERMS
Cell adhesion, G-protein coupled receptor, stem cell marker, basal breast cancer

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	124	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Page

1. INTRODUCTION

There is a significant unmet need for biomarkers that can distinguish aggressive from indolent breast cancers in order to spare women from unnecessary treatment with associated side-effects and to direct resources specifically to those who could derive benefit. 30% of women with basal-type breast cancer have very poor outcome whereas 70% have lower risk. Reproductive history is a significant risk factor for breast cancer as both ovulation and pregnancy stimulate proliferation of mammary stem cells (MaSCs) and progenitors and expose these long-lived cells to replicative error. To date there is no specific and universal hallmark of breast stem and progenitors. Several regenerative subpopulations have been described but surprisingly these label mutually exclusive populations at distinct locations. Moreover, most are not specific for epithelium compromising their utility as potential prognostic markers. Thus, our incomplete knowledge of the mammary hierarchy hampers our ability to comprehend the cellular origins of breast cancer subtypes. We proposed to validate an adhesion G-protein coupled receptor (Gpr) as a novel and specific marker of mammary stem/progenitors and cancer stem cells.

2. KEYWORDS

Cell Adhesion, G-protein coupled receptors, Stem cell marker, Basal Breast Cancer

3. ACCOMPLISHMENTS

The major goals of the project were:

Task 1) Determine the role of Gpr+ cells in mammary development. (95% complete)

Task 2) Determine the significance of Gpr expression in human and mouse breast cancers. (60% complete)

- ***What was accomplished under these goals:***

Here we report the results obtained this year. Our appended manuscripts that are currently under submission are a compilation of the results of the last three years effort.

Note on nomenclature: during the course of this proposal the official gene name has changed. Gpr refers to the Gpr125 protein encoded by the gene *Adgra3*. Mice referred to in the SOW as Gpr-lacZ are $Adgra3^{lz/+}$ and Gpr-DTR:EGFP.creER^{T2} are $Adgra3^{cre/+}$. Throughout this report we have used both terms

Specific Aim 1. Determine the role of Gpr+ cells in mammary development.

Subtask 1. Validate expression of Gpr in Gpr-DTR:EGFP-creER^{T2} and Gpr-lacZ mice by reporter expression, immunohistochemistry and FACS analysis. (1-12 months)

In our Year 1 report (Y1) we documented the generation of two reporter mice Gpr-lacZ ($Adgra3^{lz/lz}$) and Gpr-DTR:EGFP-creERT2($Adgra3^{cre/cre}$) (Y1 Figs. 1 and 4). The Gpr-lacZ reporter allowed us to determine the pattern of Gpr expression by staining with X-gal to detect beta-galactosidase fused to the Gpr ectodomain (Y1 Figs. 2, 3, 7). The Gpr-DTR:EGFP-creER^{T2} ($Adgra3^{cre/cre}$) knock out mouse showed an identical EGFP expression pattern (Y1 Fig. 5). Promoter activity was confirmed by analysis of mRNA by qPCR (Y1 Fig. 7). Colocalization of these reporters with lineage markers by immunofluorescence (Y1 Fig. 6) showed that Gpr is

expressed at predicted sites of embryonic, pubertal and pregnancy-induced stem/progenitor activity in a basal subpopulation. FACS analysis of surface markers (Y1 Fig. 8) indicated that Gpr+ cells express CD24 and integrin CD29 and CD49 characteristic of regenerative mammary repopulating units and lacked expression of Sca1 and showed low CD61, which are markers of more committed cell types. In our year 2 report (Y2) we investigated overlap of Gpr+ cells with other regenerative cell populations in pubertal mammary glands by FACS analysis (Y2 Fig.1) and demonstrated that Gpr+ cells clustered with cells expressing basal gene markers by mining the Tabula Muris and Bach et al scRNAseq dataset (Y2 Fig. 2, 3) comprising total adult mammary epithelial cells.

This year we expanded upon these findings as follows:

a) Completed expression analysis of Gpr expression in the embryonic mammary gland. Previously we reported that Gpr appears at E15 in whole embryos (Fig. 1A,B), a time when the mammary gland becomes specified and in females begins to proliferate. This year we sectioned E15 embryos and found that Gpr is concentrated towards the invading tip or the mammary sprout. Gpr+ cells colocalized with both basal keratin (K) K14 and luminal K18 markers as well as with PCNA and BrdU **(Fig. 1A'-F).** This demonstrated that at E14 Gpr+ cells are uncommitted to a particular mammary lineage and are proliferative.

Next, we sectioned E18 embryos in which the embryonic mammary tree has begun to branch and to form microlumen. At this stage, we observed the lineages had begun to segregate. Although K14 was expressed in most cells, K18 was upregulated in cells surrounding microlumen and p63 was becoming basally restricted. Gpr was concentrated in the nipple proximal duct in cells that still expressed both keratins **(Fig. 1 I-J).** However, it was also concentrated at the growing tips of the branches in a novel cell type that lacked all keratin expression **(Fig. 1G-K).** These data show, Gpr marks a previously undescribed cell-type with mesenchymal features at ductal tips that is poised for ductal invasion of the fat pad.

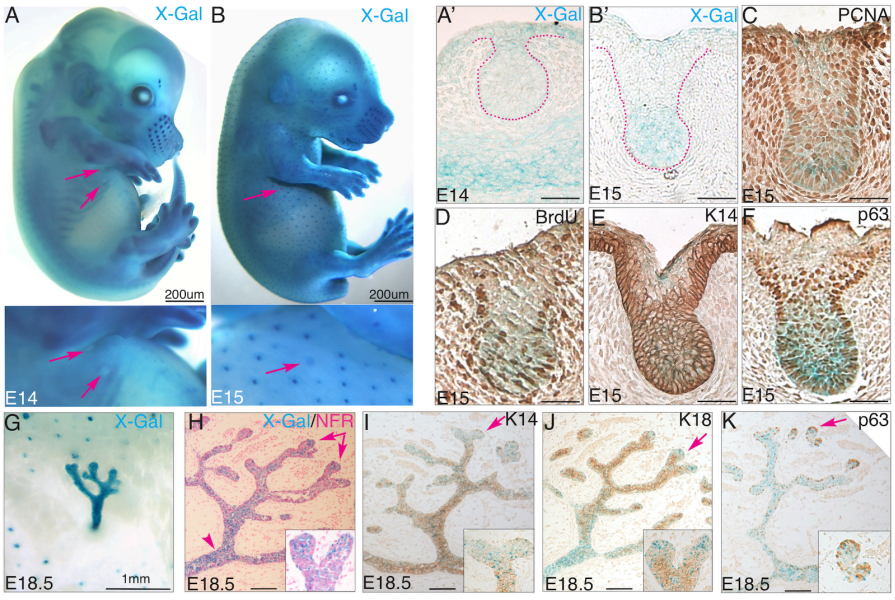


Fig. 1 Gpr expression in the embryonic mammary rudiment.
A, B) X-gal stained Gpr-lacZ (*Adgra3*$^{lz/+}$) embryos at embryonic day (E)14 **(A)** and E15 **(B)**. Arrows indicate mammary buds and sprout respectively magnified in boxes below. Scale bar 200µm. **A'-F)** Gpr-lacZ expression in sections of E14 buds **(A')** and E15 sprouts **(B'-F),** with immunolocalization for proliferative markers PCNA and BrdU **(C, D)**, and K14 and p63 **(E,F)**. Scale bar 50µm. **G)** X-gal stained skin whole mount showing Gpr-lacZ expression in the E18.5 tree and hair follicles encircling the developing nipple zone. Scale bar 1mm. **H-K)** Sections of the E18.5 rudiment tree stained with X-gal followed by NFR counterstain **(H)**; immunochemical detection of K14, K18, and p63 **(I-K)**. Boxed insets are higher magnification of branch tips regions indicated by arrows. Scale bar 50µm.

b) scRNAseq and immunofluorescence analysis of Gpr and other regenerative populations
Previously we analyzed of Gpr mRNA expression in adult mammary epithelial cell populations datasets. This year, we extended these scRNAseq analyses to datasets derived from embryonic and pubertal mammary epithelial cells and examined the relationship of Gpr cells with other regenerative subpopulations at all three stages (embryonic, pubertal and adult) of mammary development. Gpr was expressed in E14 Lgr5+ CD29+ embryonic cells and co-expressed both basal and luminal keratin K14, K18 as well as the transcription factors p63, Lef1 and Tcf1 **(Fig. 2A).** In contrast in datasets comprising pubertal **(Fig. 2B)** and adult **(Fig. 2C)** mammary cell types Gpr cells were found exclusively within a subset of the "basal" cluster characterized by K14 expression.

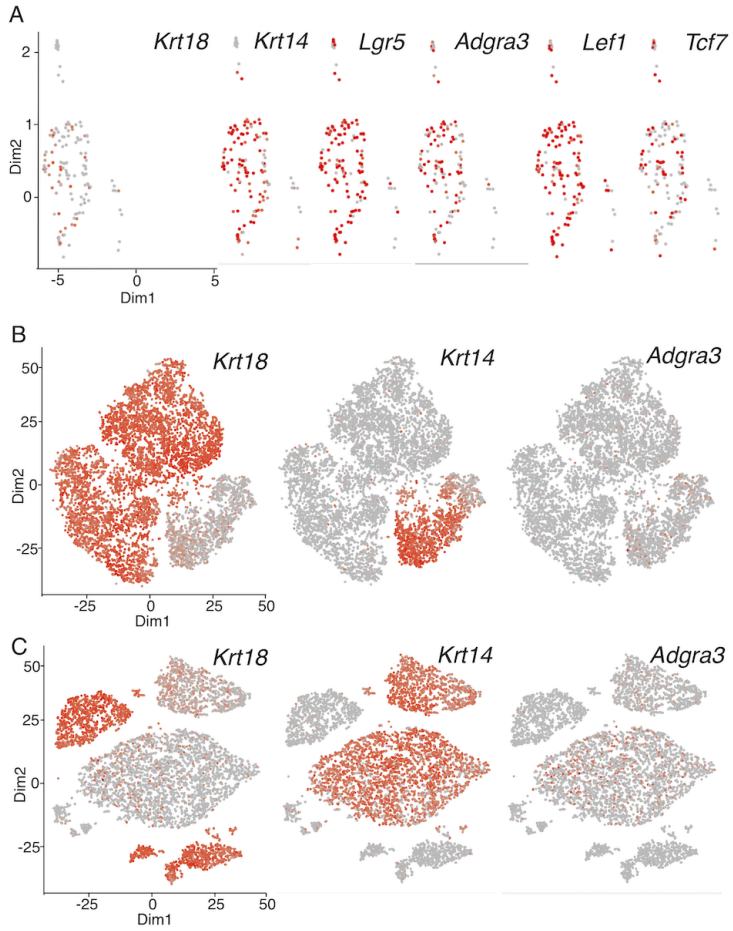


Fig. 2 Gpr/Adgra3 population is expressed embryonic cells expressing bilineage markers but confined to the basal cluster in pubertal and adult mice.

A) t-SNE plots of *Krt18, Krt14, Lgr5, Adgra3, Lef1 and Tcf7* in Lgr5+ CD29+ embryonic mammary cells. Data mined from dataset generated by Wuidart et al [2]

B) t-SNE plots of *Krt18*, *Krt14* and *Adgra3* in pubertal and adult mammary cells mined from dataset generated by Pal et al [3]

C) t-SNE plots of *Krt18*, *Krt14* and *Adgra3* in adult mammary cells mined from dataset generated by Pal et al [3]

In adult cells, Gpr/*Adgra3*, *Lgr5* and *Procr* expressing cells co-clustered together in the basal cell cluster **(Fig. 3).** To investigate the relationship of Gpr+ cells to sSHIP+ cells we compared reporter expression patterns in tissues from Gpr-lacZ and sSHIP-EGFP reporter mice. Both reporters were expressed throughout the cap cells of the terminal end buds **(Fig. 4).** Collectively, these data indicate that Gpr is expressed in Lgr5 cells, which are located in nipple proximal ducts, and in

sSHIP cells, which are located in the cap cell layer of TEB, and thus is a unifying hallmark of these otherwise disparate and mutually exclusive progenitor cell types.

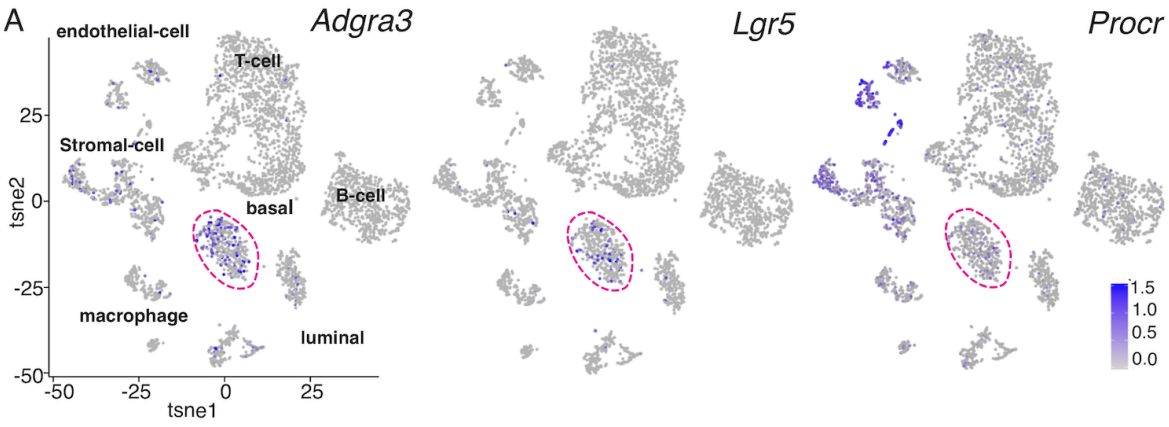


Figure 3. Gpr+ cells encompass distinct regenerative populations.
A) t-SNE plots of Gpr/*Adgra3*, *Lgr5* and *Procr* in total mammary cells mined from Tabula Muris dataset [1]. Nb *Adgra3*, *Lgr5* and *Procr* are all expressed in cells of the basal cluster, however *Procr* is additionally expressed in stromal cell types.

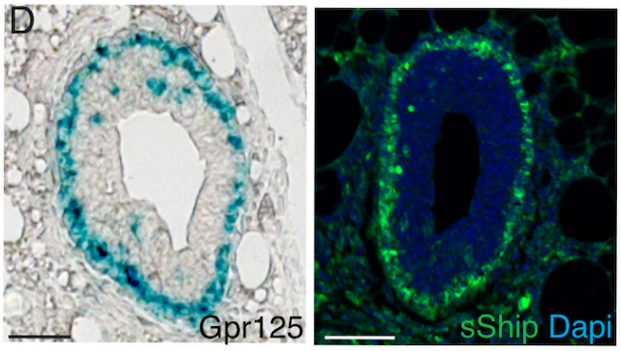


Figure 4. Co-expression of Gpr and sSHIP in the cap cell layer of TEB.
L**eft panel**) TEB sections show expression of Gpr-lacZ fusion protein (blue stain) in cap cell layer.
Right panel) TEB sections show expression of sShip-EGFP reporter (green) in cap cell layer.
Scale bar = 50µm.

c) Cell isolation and transplantation assays

In our original SOW, we had proposed to sort living Gpr+ cells from Gpr-DTR:EGFP-creER^{T2} mice by FACS using the EGFP reporter. Despite breeding the mice to homozygosity we were unable to detect the EGFP by this method, likely because the DTR:EGFP fusion protein is reported to be 100-fold less bright than free EGFP. As an alternative we attempted to use antibodies specific for human DTR, however these cross-reacted with mouse thus precluding our ability to specifically isolate Gpr+ cells and perform transplantation/regeneration assays.

Subtask 2 and 3. Cross Gpr-DTR:EGFP-creER^{T2} to Rosa26-STOP-reporter lines (1-18 months). Trace the Gpr lineage by inducing cre activity with TAM and detecting Rosa26R-reporter in Gpr descendants in conjunction with differentiation markers (18-30 months).

To address whether Gpr+ cells are stem cells or lineage-restricted progenitors during normal physiological development we carried out lineage tracing. In Y1 we crossed our Gpr-DTR:EGFP-creER^{T2} (*Adgra3$^{cre/cre}$*) mice to Rosa26R-STOP-tdTomato (tdT) lines. This transcriptional block is removed when tamoxifen (Tam) is administered, resulting in reporter expression in Gpr+ parental cells and in all cellular progeny. We established 3D confocal imaging of cleared tissue whole mounts (Y1 Fig 9 and 10) and showed labelled progeny in pubertal terminal end buds (Y1 Fig. 10) and ducts of pubertal mice (Y1 Fig. 11). In Y2 we traced the progeny of pubertal Gpr+ cells in mature ducts (Y2 Fig. 4), in alveoli and ducts during pregnancy (Y2 Fig. 5), in lactating glands (Y2 Fig 6.) and showed they colocalized with basal-specific markers by immunofluorescence (Y2 Fig 7) and lacked expression of luminal markers (Y2 Fig 8). We traced their progeny after 6 months in virgin mice and after completing two cycles of pregnancy and involution. (Y2 Fig. 9). These data showed that pubertal Gpr+ cells are long-lived unipotent basal stem/progenitors.

a) Further characterization of progeny of pubertal Gpr+ cells

We have extended analysis of the cell progeny of pubertal Gpr+ cells by analyzing the characteristics of their tdT-labeled progeny by flow cytometry. These analyses confirm that the progeny of pubertal Gpr+ cells show CD49f+ CD24 characteristics typical of basal myoepithelial cells (Fig. 5).

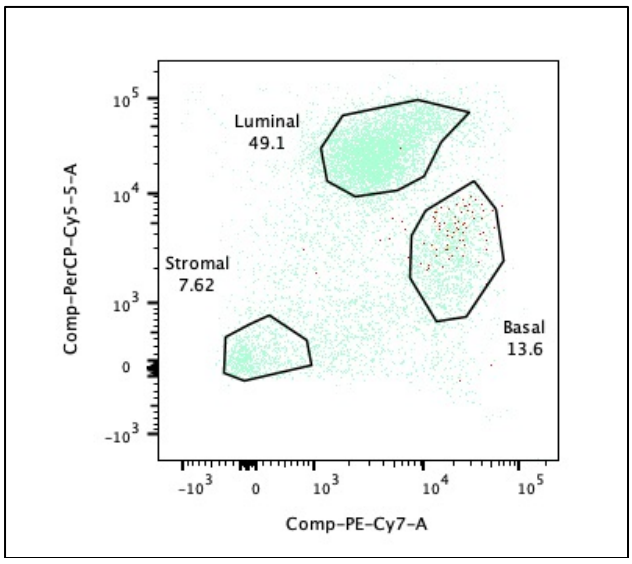


Fig. 5 Cell progeny of pubertal Gpr+ cells display basal characteristics on flow cytometry.

Lineage tracing was initiated in Rosa26 STOP tdT;Gpr-DTR:EGFP-creER^{T2} (*Adgra3$^{cre/cre}$*) mice by delivery of Tamoxifen at 5 weeks of age. Tissue was harvested at 12 weeks and analyzed by flow cytometry. Total MECs (green dots) stained with antibodies against CD24 and CD49f partitioned into stromal, luminal and basal populations as indicated. Cell progeny marked by expression of **tdT (red dots)** gated within the CD24$^{med/low}$/CD49f$^{+/hi}$ basal population.

b) Lineage tracing of the progeny of embryonic Gpr+ cells

This year, we also extended our lineage tracing analyses to tracing the progeny of embryonic Gpr+ cells and to the progeny of Gpr+ cells that appear at ductal tips during pregnancy. We successfully overcame the problem, noted in the previous report, of Tam-induced pup death and pup delivery problems by performing C-section on the dams just prior to term and fostering the pups (10,11). We labeled embryos at E14 and E19 and then traced their cellular progeny when the pups reached 8 weeks of age **(Fig. 6).** Our 3D-confocal imaging shows that E14 embryonic Gpr-expressing cells generate not only tdT+ basally located spindle-shaped cell progeny that colocalize with myoepithelial markers such as Sma **(Fig. 6M),** but also generate cuboidal cells contacting the lumen (Fig. 9N) that express luminal cell markers such as E-cadherin **(Fig. 6O).** In contrast, when embryos are labeled at E19, their tdT+ cell progeny are exclusively basal **(Fig. 6P).** These data

show that early E14 embryonic Gpr+ cells are bipotent stem cells whereas E19 Gpr+ cells have become lineage restricted and function from thereon as unipotent basal progenitors

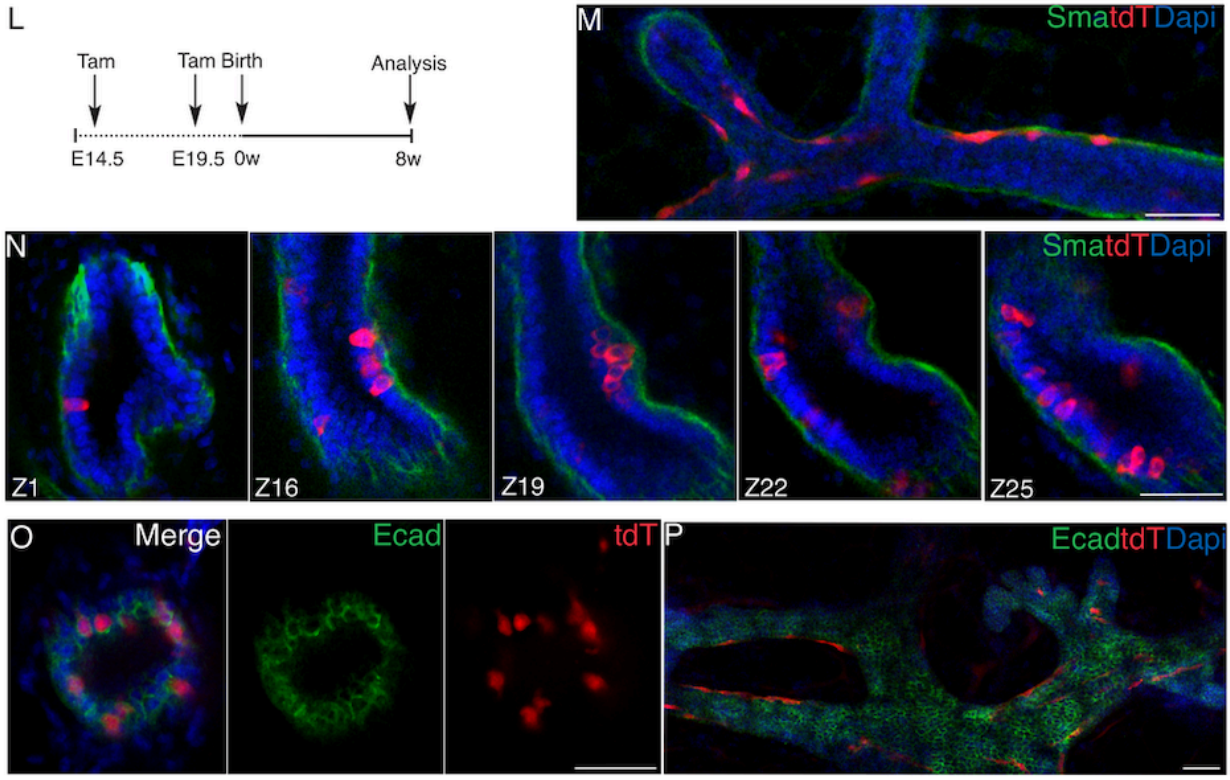


Fig. 6 Lineage tracing shows E14.5 Gpr+ cells are bipotent whereas E19.5 Gpr+ cells are unipotent progenitors
L) Tracing of Gpr+ cells in E14.5 or E19.5 embryos was initiated by delivering Tam in pregnant *Adgra3*$^{cre/cre}$ dams mated to tdT mice. Mammary tissue from the progeny were analyzed at 8 weeks of age. **M-O)** 3-D images showing representative regions of pubertal ducts from E14.5 labelled embryo containing clusters of basally located tdT+ cells that co-express the basal marker SMA **(M)** as well as tdT+ columnar luminal cells lacking SMA **(N)** and expressing luminal marker Ecad **(O)**. **P)** Glands from progeny labeled at E19.5 show basally located tdT+ cells devoid of Ecad. Dapi=nuclear staining. Scale bar 50µm. Two glands from each of 3 mice were analyzed/stage.

c) Lineage tracing of the progeny of pregnancy-induced Gpr+ cells
We also analyzed the progeny of cells that expressed Gpr+ cells during pregnancy. We initiated tracing in pregnant mice at P13.5 when Gpr is expressed at the tips of elongating side-branches and analyzed their glands just prior to birth and during lactation (Fig. 7). These cells gave rise to tdT+ cell progeny that exclusively displayed myoepithelial characteristics (Fig. 7). Thus Gpr+ cells present during pregnancy function as unipotent basal progenitors.

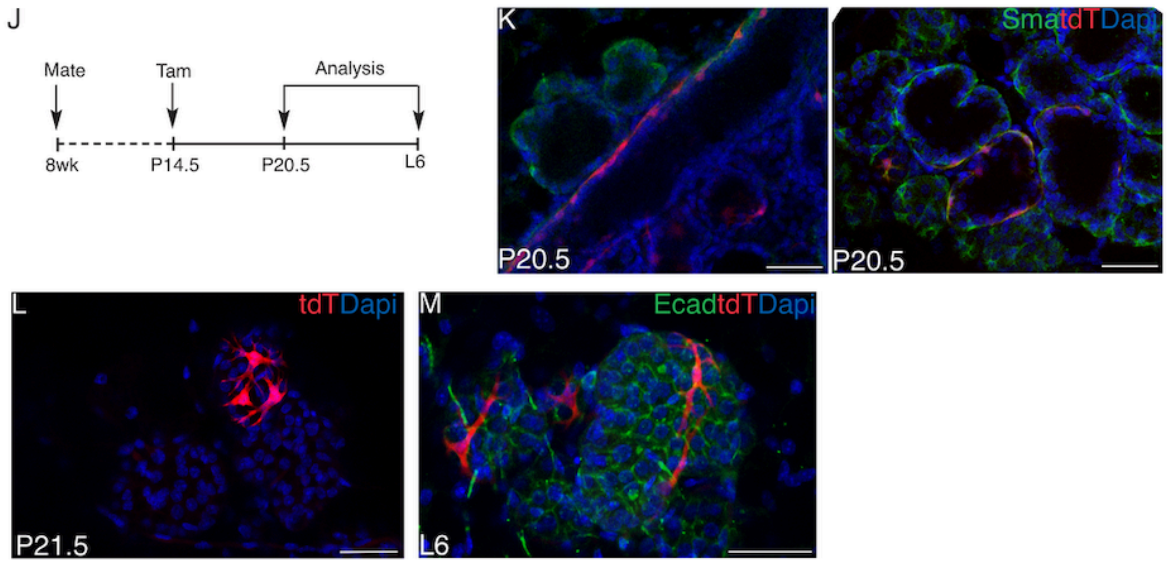


Fig. 7 Cells expressing Gpr on leading tips of branches during pregnancy are unipotent basal progenitors. J) Schematic illustrating strategy to trace the lineage of cells that express Gpr during early pregnancy by delivering Tam at p14.5 and harvesting glands at P20.5 and L6 mice. **K)** tdT+ Sma+ cells were found in ducts and alveoli at p20.5 and at birth **(L)**. **M)** Fully differentiated tdT+ cells were devoid of luminal Ecad marker at L6. Scale bar 50µm.

Subtask 4 and 5. Ablate Gpr+ cells in Gpr-DTR:EGFP-creER^{T2} (*Adgra3$^{cre/cre}$*) mice by administration of DT, validate cell death by loss of Gpr expression and monitor the effect on mammary development.

Initially, we proposed to kill cells expressing Gpr by administering Diphtheria Toxin (DT) to 6-week old mice expressing Gpr-DTR:EGFP-creER^{T2} (*Adgra3$^{cre/cre}$*). Unfortunately, as reported last year, these mice deteriorated rapidly due to Gpr expression in other essential organs such as salivary glands, parotids, testis, kidney, ear and brain (Y2 Fig. 10). To overcome this obstacle, during this year we have carried out pilot experiments transplanting mammary glands from Gpr-DTR:EGFP-creER^{T2} (*Adgra3$^{cre/cre}$*) mice to immunocompromised Nude mice and then delivering DT. We have shown that we can carry out the transplantations successfully, that the Nude mice tolerate the DT without adverse effects, and that Gpr-lacZ expression is impaired or eliminated by DT. Our preliminary analyses suggest impairment of ductal outgrowth in the transplanted glands **(Fig. 8).**

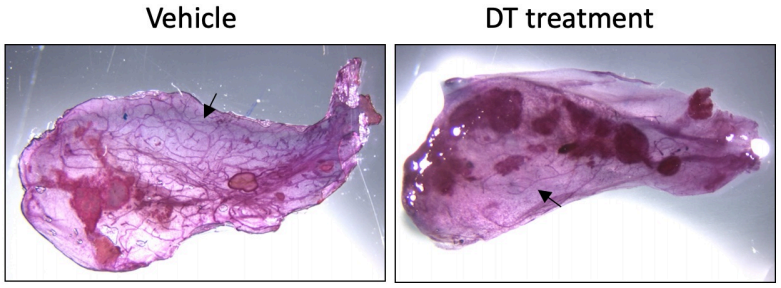


Fig. 8 DT impairs ductal outgrowth
Wholemount mammary glands from Gpr-DTR:EGFP-creER^{T2} (*Adgra3$^{cre/cre}$*)transplanted into Nude mice treated with vehicle or DT as indicated stained with carmine.

However, we encountered some problems with the suturing, which is required to secure the transplanted glands to the body wall of host mice, obscuring and interfering with the assay. We continue to try to optimize this experiment and scale up to obtain statistically sound data. To address the problem created by the suturing we are also attempting to carry out this experiment in a modified way by harvesting mammary epithelial cells and organoids from Gpr-DTR:EGFP-creER^{T2} mice and treating them with DT before and after transplantation directly into the host fatpad cleared of endogenous epithelium.

Gpr+ cells are located in sites of migration and invasion and function as myoepithelial progenitors in other secretory organs

In Y1 we documented the phenotypic effects of Gpr loss on mammary ductal elongation (Y1 Fig. 12), male sterility and the overt eye phenotype seen in homozygous Gpr-DTR:EGFP-creER^{T2} (*Adgra3*$^{cre/cre}$) and Gpr-lacZ (*Adgra3*$^{lz/lz}$)(mice (Y1 Fig. 13, 14). In Y2 we showed inflammatory infiltration of CD4, CD8 and macrophages into their lacrimal glands (Y2 Fig. 11). This year we completed these analyses by examining other clinical parameters, including measuring lacrimation by Shirmer's tear test, ocular pressure, corneal abrasion and histological examination of meibomian and goblet cells in eyelids and found these mice display many features that are associated with "dry eye" .

We examined Gpr expression and its lacrimal lineage in lacrimal glands taken from the same mice for our mammary experiments described above. Our results show that embryonic Gpr+ cells found at the tips of elongating lacrimal ducts generate the lacrimal myoepithelium **(Fig.9).** These data show that Gpr demarcates cells that serves a similar function in multiple secretory organs. This reinforces our studies on the mammary gland by showing that Gpr cells involved in guiding cell migration and tissue invasion in multiple secretory organs are myoepithelial progenitors. The role in directing migration and invasion of progenitors has considerable import for understanding the characteristics and hallmarks of breast cancer cells. (see appendix for BioRXiv preprint).

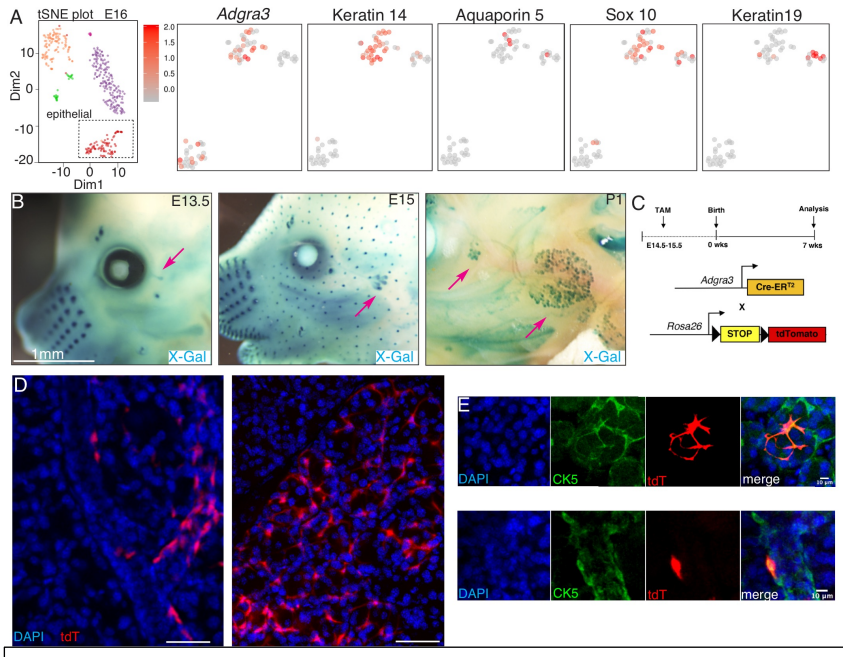


Fig. 9 Gpr+ cells, located at ductal tips during development, function as lacrimal myoepithelial progenitors.

A) t-SNE plot of cells clusters within E16 lacrimal glands. Zoomed images of E16 epithelial compartment (boxed region) show cells expressing Adgra3 mRNA also express myoepithelial markers, Keratin14 and Sox10 but not luminal markers Keratin19 or Aquaporin5. **B)** Gpr-lacZ expression in embryos (arrows point to invading tips of lacrimal glands. **C)** Strategy for tracing the lineage of Gpr-positive cells in E14.5-E15.5 embryos. 3D-confocal images of lacrimal glands from mice at **(D)** 7 weeks; and **(E)** 6 months showing tdT expression in elongated myoepithelial cells along the basal border of ducts and stellate cells enmeshing acini colocalized with myoepithelial marker (CK5). Scale bar 50µm.n=3.

Specific Aim 2. To determine the significance of Gpr expression in breast cancers.
Task 1. Screen breast cancer cell lines for Gpr expression by qPCR and mine bioinformatics datasets of human breast cancers.

In Y1 we documented Gpr/Adgra3 mRNA expression in human breast cancer cell lines (Y1 Fig 15) and mined public datasets using the kmplotter algorithm to show that high expression of Gpr predicted for poor prognosis in relapse free survival exclusively in the basal breast cancer subtype in human breast cancers. This year we extended these analyses to access further datasets and found high GPR/ADGRA3 levels within this subtype correlated with worse patient outcome in terms of with disease-free survival using the breastmark algorithm (Figure 10, blue line).

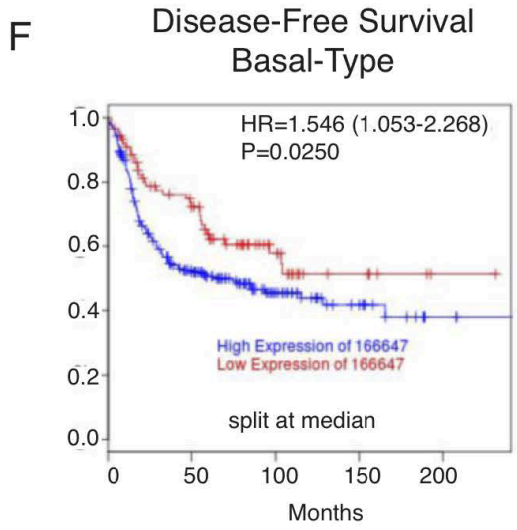


Fig. 10 High GPR/ADGRA3 is predictive of poor outcome.

Disease-free survival utilizing BreastMark ssp 2003/2006 datasets and splitting patients at the median level for expression (http://glados.ucd.ie/BreastMark/) into low (red line) and high (blue line) Gpr expression groups

Task 2 Test whether eradicating Gpr+ cells affects tumorigenesis and/or regression
Subtask 1: Generate bi-transgenic Gpr-DTR:EGFP-creER^{T2}/MMTV-Wnt1 mice. As reported in Y2 at the beginning of this report year we had begun to cross homozygous Gpr-DTR:EGFP-creER^{T2}(*Adgra3$^{cre/cre}$)* to MMTV-Wnt1 and Gpr-lacZ (*Adgra3$^{lz/lz}$)(* mice to generate the first experimental cohorts of compound Adgra3$^{cre/lz}$;MMTV-Wnt1 female progeny. MMTV-Wnt1 tumors arise with an average latency of 7 months. As expected these mice born in September 2019 began to produce MMTV-Wnt1 tumors in March 2020. Unfortunately, this period coincided with a mandatory shutdown of all experiments on March 17 due to the COVID-19 outbreak in New York. The majority of this cohort had to be euthanized as their tumors formed during the lock down period without us being able to harvest the tissue. Dr. Spina volunteered for the very first pilot group that returned to work at the end of May 2020. She set up the breeding again however the loss of this cohort of experimental animals has set back our work by more than 8 months. Once this second cohort develops tumors we will, as planned, harvest the cells, treat them with and without DT, monitor for ablation of Gpr+ cells by loss of Gpr-lacZ expression, and transplant them into Foxn1 nu mice to monitor the effects of ablating Gpr+ cells on tumor propagation. Tumors arising from cells transplanted in this manner generally arise within 1 month.

Subtask 2: Ablate Gpr cells by DT administration and assess effects on tumor onset, progression, regression and histology.

In Y2 we reported our first pilot lineage tracing experiments to identify the progeny of Gpr+ cells present in the tumor setting (Y2 Fig. 12). We showed that when TAM was administered to Adgra3$^{cre/lz}$;MMTV-Wnt1 females at 5 weeks of age and their hyperplastic glands were analyzed at 12 weeks of age that in contrast to the unipotency of Gpr+ cells in the normal gland, in the context of the MMTV-Wnt1 hyperplasia Gpr+ cells acquire bipotency. We have repeated these experiments in more animals and confirmed this observation. These data show that Gpr+ cells generate cells expressing K14 and other expressing K18 which comprise a substantial percentage of the MMTV-Wnt1 tumor bulk.

Gpr+ progenitors are expanded in MMTV-Wnt1 tumors and retain embryonic features. In characterizing the Gpr+ cells further we have found them to comprise a third novel undifferentiated cell type that is greatly expanded in tumors with early onset. Our results show this cell type lacks both basal (K14) and luminal (K18) differentiation markers and also SMA **(Fig. 11).** It expresses p63 and Tcf1 transcription factors and shows intermediate levels of integrin expression as detected by FACS analyses **(Fig. 12).** Intriguingly, these features mirror those we have described for cells at the invading tips of embryonic ducts. We propose that the poor outcome associated with Gpr expansion could relate to reactivation of the normal invasive functions of this cell type during embryonic development. (see appendix for manuscript)

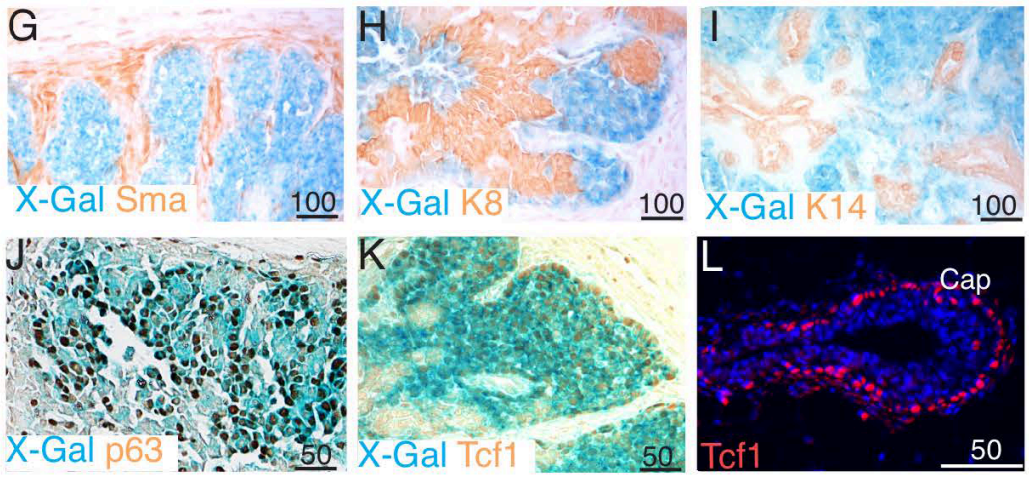


Figure 11. Gpr+ progenitors are expanded in MMTV-Wnt1 tumors and retain embryonic features.
X-gal stained sections of *MMTV-Wnt1;Adgra3*$^{lz/+}$ tumor (**G-K)** showing Gpr+ cells (blue stain) devoid of immunolocalization for SMA, K14 or K8, but expressing p63 and Tcf1. **L)** Tcf1 expression in the cap cells of normal TEB.

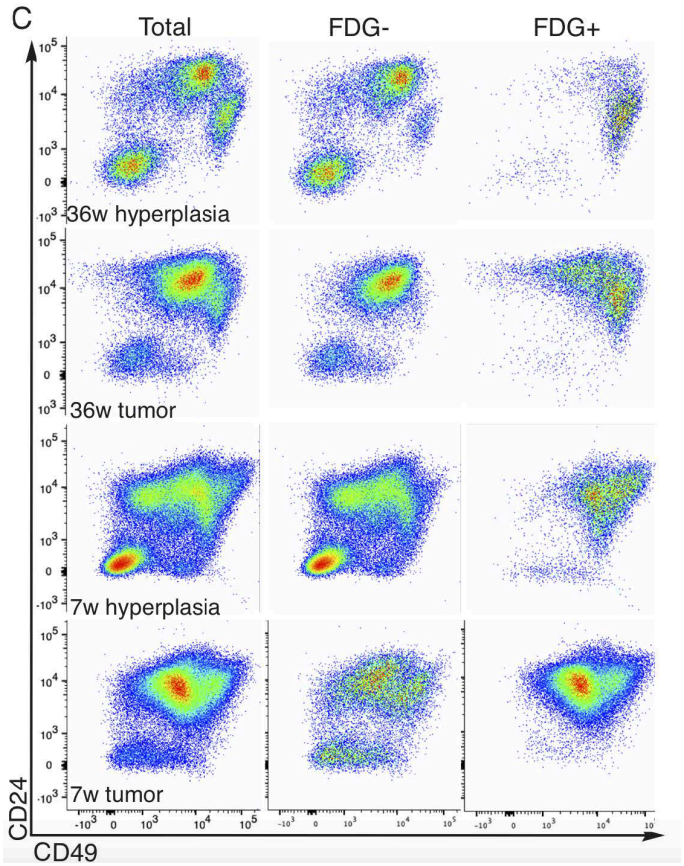


Fig. 12 Gpr+ progenitors expanded in MMTV-Wnt1 tumors with short latency express embryonic integrin expression features.

C) Representative flow cytometry dot plots of total MECs from *MMTV-Wnt1;Adgra3*$^{lz/+}$ hyperplastic uninvolved glands and their associated short and long latency tumors, stained with CD24 and CD49f. Plots are representative of three independent experiments. **Bottom right panel** Note the vast expansion of a central population in tumors with short latency (7 week).

- **What opportunities for training and professional development has the project provided.**

  - **Dr. Cowin** acquired skills in tissue clearing and 3D-confocal imaging from colleagues in the Pathology Department, Cambridge who have pioneered this technique in the mammary gland (8).
  - **Dr. Spina** is a fully trained pharmacist and Ph.D. During the course of this project she has augmented her existing skills in breast cancer cell culture by acquiring new skills in a) mouse genetics b) histological analysis, c) tissue clearing and 3D-confocal imaging d) *in vivo* survival surgery and fat-pad clearing and transplantation. e) lineage tracing f) organoid isolation and mammosphere culture g) analysis of scRNAseq datasets
  - **Dr. Spina** took courses in ethics, animal welfare and professional career development. She attends the "works in progress" presentations of the stem cell and pharmacology and molecular oncology training programs as well as the breast cancer group within the NYU Cancer Center. She has presented in these venues as well as at the departmental retreat. She has met with her committee comprising: Dr. Mayumi Ito Ph.D. an expert in the lineage tracing of hair follicle and nail ectodermal appendages; Dr. Dimitris Placontonakis MD, an expert on adhesion-GPCR 133 in glioblastomas and Dr. Konstantin Itchenko, an expert in Adhesion-GPCR signal transduction.
  - **Dr. Spina and Cowin** have written two manuscripts together and a review on Embryonic Mammary development for Trends in Cell and Developmental Biology (see appendix).

- **How were the results disseminated to communities of interest?**
  - **Dr. Cowin** was scheduled to present these findings as an invited speaker at the Gordon Conference on Mammary Gland Biology in April 2020 (canceled due to COVID travel restrictions)
  - **Dr. Cowin** presented these findings to the Skirball Institute, NYUSOM Oct 9, 2019
  - **Dr. Cowin** presented these findings to the Dept of Opthalmology, NYUSOM on November 5, 2019
  - **Dr. Spina** was to have presented this work to the 10th Adhesion GPCR Workshop in Copenhagen (canceled due to COVID travel restrictions)
  - **Dr. Spina** presented this work to the Annual Skirball Institute Retreat in Oct 2019
  - **Dr. Spina** presented this work as a selected speaker for the Postdoctoral Association Research Day NYUSOM on September 23, 2019
  - **Drs. Spina and Cowin** have submitted two manuscripts and a review (see appendix)

- **What do you plan to accomplish during the next reporting period to accomplish the goals and objectives?**
  - Our first goal will be to treat glands, MECS and organoids with DT and examine the effects on ductal outgrowth and tumor propagation after transplantation into host mice

- Our second focus will be to meet any requirements requested by reviewers to bring the attached manuscripts to publication (see appendix).

4. IMPACT

- **What was the impact on the development of the principal discipline of the project?**

  - Our expression and linage tracing studies show that Gpr identifies mammary stem/bipotent progenitors in the embryonic mammary gland and unipotent basal progenitors in the perinatal pubertal and adult mammary gland.
  - Our results in Y1 showed that mice lacking this adhesion GPCR show delayed mammary ductal elongation and that the presence of the cytoplasmic signaling domain of this orphan receptor is essential. In Y2 we have shown that Gpr and its signaling domain are required for glandular development more generally and serve a common function as progenitors of the myoepithelium in multiple secretory organs.
  - Our results in Y1 showed that high levels of Gpr occur in aggressive forms of basal positive breast cancer, and that patients with higher levels within these groups have particularly poor outcome. These support the concept that Gpr expression has value as a prognostic indicator of patient outcome in basal type breast cancer. This has pioneered a new field in breast cancer research since there are no studies besides our own on this Gpr. In Y2 and Y3 we have addressed the mechanism by showing that Gpr+ cells acquire bipotency in the tumor situation. We have also shown that Gpr identifies a novel cell type that is expanded in tumors with reduced latency and displays mesenchymal undifferentiated features reminiscent of embryonic progenitors found at invasive tips in embryonic glands

- **What was the impact on other disciplines?**
  - The homozygous Gpr-DTR:EGFP-creER^{T2} mice, which have a pronounced eye phenotype display several features of "Dry Eye Syndrome" and could have utility as a model for this human inflammatory disease.
  - Our studies show that the pattern of Gpr expression is consistent with those affected by Mumps virus.
  - The expression of Gpr in several secretory glandular structures that share common ectodermal origin suggests it may play in ductal branching and secretory differentiation. Its presence in the stem cell compartments of several other ectodermal appendages including the bulge and secondary germ compartments of hair follicles (see Y1 report) suggest it may be an indicator of a more generalized stem cell function.

- **What was the impact on technology transfer?**
  - Nothing to report

- **What was the impact on society beyond science and technology?**
  - Linking breast development factor to breast cancer risk opens the door to preventative strategies linked to reproductive history.

5. CHANGES/PROBLEMS

- **Changes in approach and reasons for change**:
  - As we were unable to detect the EGFP reporter even in homozygous mice we have had to abandon the part of the project examining lineage plasticity in a transplantation setting. However, we have discovered lineage plasticity in the context of MMTV-Wnt1 expression and so have focused on this instead.

- **Actual or anticipated problems or delays and actions or plans to resolve them**
  - **Subtask 1: Generate bi-transgenic Gpr-DTR:EGFP-creER^{T2}/MMTV-Wnt1 (**Adgra3$^{cre/lz}$;MMTV-Wnt1) **mice.** As reported in Y2 at the beginning of this report year we had begun to cross homozygous Gpr-DTR:EGFP-creER^{T2} (Adgra3$^{cre/cre}$) to MMTV-Wnt1 and Gpr-lacZ (Adgra3$^{lz/lz}$) mice to generate the first experimental cohorts of compound Gpr-DTR:EGFP-creER^{T2}/MMTV-Wnt1 **(**Adgra3$^{cre/lz}$;MMTV-Wnt1) female progeny. MMTV-Wnt1 tumors arise with an average latency of 7 months. As expected these mice born in September 2019 began to produce MMTV-Wnt1 tumors in March 2020. Unfortunately, this period coincided with a mandatory shutdown of all experiments on March 17 due to the COVID-19 outbreak in New York. The majority of this cohort had to be euthanized as their tumors formed during the lock down period without us being able to harvest the tissue. Dr. Spina volunteered for the very first pilot group that returned to work at the end of May 2020. She set up the breeding again however the loss of this cohort of experimental animals has set back our work by more than 8 months. Once this second cohort develops tumors we will, as planned, harvest the cells, treat them with and without DT, monitor for ablation of Gpr+ cells by loss of Gpr-lacZ expression, and transplant them into Foxn1 nu mice to monitor the effects of ablating Gpr+ cells on tumor propagation. Tumors arising from cells transplanted in this manner generally arise within 1 month.

- **Changes that had a significant impact on expenditures –**
  - Mouse costs during COVID lock down. We have incurred significant extra animal husbandry costs due to the loss of the mouse cohort involved in Task 2 as their tumors came during the lockdown period as detailed above. We have had to incur duplicate costs to regenerate these animals and to house a second set of progeny for over 7 months until they generate their tumors.

- **Significant changes in use or care of human subjects**
  - Nothing to report
- **Significant changes in use or care of vertebrate animals**
  - Due to COVID-19 lock down we lost a cohort of animals that developed tumors during this period and have had to repeat the breeding of significant numbers of mice
- **Significant changes in use of biohazard or select agants**

- Nothing to report

6. PRODUCTS

- **Publications, conference papers and presentations**

- ***Manuscripts:*** see attached manuscripts submitted
  - **Spina E**, Simundza J, Incassati A, Chandramouli A, Watson CJ, Cowin P. Gpr125 Identifies Mammary Stem and Progenitor Cells and is Associated with Reduced Tumor Latency. Submitted to Nature Cell Bio.
  - **Spina E**, Cowin P. Embryonic Mammary development. Review. Seminar in Cell and Development Biology.
  - **Spina E**, Handlin R, Simundza J, Incassati A,Faiq M, Sainulabdeen A, Chan KC, Cowin P. Gpr125 Plays Critical Roles in Lacrimal Myoepithelia and Tear Film. bioRxiv. doi: https://doi.org/10.1101/2020.09.15.296749
- ***Books etc:*** Nothing to report

- ***Other publications, conference papers and presentations***
  Acknowledgement of Federal Support:

- **Website(s) or other internet site (s):** Nothing to report

- **Technologies or Techniques:** Nothing to report

- **Inventions, patent applications and/or licenses:** Nothing to report

- **Other products:**
  - **Research material:** Generation of Adgra3$^{cre/cre}$ knock out mouse model

7. PARTICIPANTS AND OTHER COLLABORATING ORGANIZATIONS

Individuals working on the project:

Name	Pamela Cowin
Project Role	P.I.
Research Identifier	
Nearest person month worked	12
Contribution to project	Directed research
Funding Support	DOD BC160959 40%

Name	Elena Spina
Project Role	Postdoctoral fellow
Research Identifier	
Nearest person month worked	12
Contribution to project	Performed work on Aim 1
Funding Support	DOD BC160959 100%

Has there been a change in the active support of the PI or senior key personnel since the last reporting period

P.I. Dr. Pamela Cowin – No change
Postdoc Dr. Elena Spina – No change

What other organizations were involved as partners?

Nothing to report

8. SPECIAL REPORTING REQUIREMENTS N/A

9. APPENDICES:

References cited.

1. Tabula Muris, C. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372 (2018).
2. Wuidart, A. *et al.* Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat Cell Biol* **20**, 666-676 (2018).
3. Pal, B. *et al.* Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nat Commun* **8**, 1627 (2017).
4. Bach, K. *et al.* Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. *Nat Commun* **8**, 2128 (2017).

Manuscripts submitted:

- **Spina E**, Simundza J, Incassati A, Chandramouli A, Watson CJ, Cowin P. Gpr125 Identifies Mammary Stem and Progenitor Cells and is Associated with Reduced Tumor Latency. Submitted to Nature Cell Bio.
- **Spina E**, Cowin P. Embryonic Mammary development. Review. Seminar in Cell and Development Biology.
- **Spina E**, Handlin R, Simundza J, Incassati A,Faiq M, Sainulabdeen A, Chan KC, Cowin P. Gpr125 Plays Critical Roles in Lacrimal Myoepithelia and Tear Film. bioRxiv. doi: https://doi.org/10.1101/2020.09.15.296749

Gpr125 Identifies Mammary Stem and Progenitor Cells and is Associated with Reduced Tumor Latency.

Departments of Cell Biology[1] and Dermatology[2] New York University School of Medicine. Department of Pathology, University of Cambridge[3].

Elena Spina[1], Julia Simundza[1], Angela Incassati[1], Anupama Chandramouli[1], Christine J Watson[3], Pamela Cowin[1,2]*

*Address for correspondence:

Dr. Pamela Cowin

New York University Medical Center

MSB 621, 550 1st Ave

New York, NY 10016

Conflict of interest statement: The authors declare that no conflict of interest exists.

Abstract

Gpr125, encoded by *Adgra3* (adhesion G protein-coupled receptor A3), is an orphan G-protein coupled receptor implicated in planar polarity that has homology to cell adhesion and axonal guidance factors. Using a variety of approaches including genetic reporters, flow cytometry, 3-dimensional imaging and lineage tracing, we show that Gpr125 expression identifies bipotent embryonic mammary progenitors and unipotent basal progenitors at multiple locations in perinatal and postnatal glands. Gpr125 progenitors that retain embryonic bipotency and undifferentiated (keratin18/14-negative) features are greatly expanded in MMTV-Wnt1 tumors and predictive of reduced tumor latency. Furthermore, Gpr125 is expressed in basal breast cancers, and higher expression within this subtype identifies patients that have a particularly poor outcome. The presence of Gpr125 mammary progenitors at sites of collective cell migration, and massed at pushing margins of tumors, highlights the mechanistic parallels between intrinsic developmental properties and cancer.

Introduction

G-protein coupled receptors (GPCRs) characterized by a seven-pass transmembrane (7TM) domain fall into five categories: Glutamate, Rhodopsin, Adhesion, Frizzled and Secretin (GRAFS)[1]. Adhesion GPCRs form the second largest group, yet remain the least understood[2]. Gpr125 contains a domain resembling immunoglobulin-like cell adhesion molecules (Ig-CAM) and has leucine rich repeats (LRR) found in Slit1, an axonal guidance factor, and LRIG, a marker of hair follicle progenitors [3, 4]. In keeping with functions suggested by these motifs, we recently discovered Gpr125-expressing myoepithelial progenitors at the migrating tips of embryonic lacrimal ducts, and others have reported its expression in spermatogonial progenitors [5, 6].

The ligand for Gpr125 is unknown and it remains unclear whether it signals through G-proteins. Indeed, several studies suggest it acts via non-canonical routes [7-10]. Gpr125 interacts with several PDZ proteins typically involved in cell junctions, polarity, directional movement and morphogenesis. For example, in zebrafish Gpr125 clusters Disheveled (Dsh) into membrane subdomains and modulates planar cell polarity complexes directing convergent extension cell movements and facial motor neuron migration [7]. Gpr125 also binds Discs large (Dlg), a tumor suppressor member of the ZO-1 protein family [8, 11]. This interaction implicates Gpr125 in cancer, where high expression has been correlated with good outcome in colon cancer and poor outcome in myeloid leukemia [10, 12]. When introduced into cultured cells, Gpr125 is internalized constitutively to endosomes suggesting roles in receptor recycling[9]. Collectively, these studies lead us to hypothesize

that Gpr125 demarcates cells with stem/progenitor potency that participate in cell polarity and adhesive events linked to directed cell migration, Wnt signaling and cancer. To test this concept further we set out to study Gpr125 in the mammary gland and mammary cancers.

Mammary glands provide an ideal system to study developmental processes, such as cell growth, adhesion, migration, and differentiation *in vivo*. Between embryonic days 10-12 (E10-12) ectodermal cells rearrange into placodes that become committed to a mammary fate and sprout towards the mammary fat pad ~E15, then invade and branch to form a small bi-layered tree [13]. Mammary ducts comprise an internal luminal layer of hormone receptor positive and negative cells that express keratin (K) K8 and K18 encapsulated by a basal layer expressing smooth muscle actin (SMA), K14 and K5[14, 15]. Development continues during puberty, when the permanent ductal system is extended by hormone-induced proliferation in terminal end buds (TEBs) [14, 16-18]. During pregnancy, side-branches emerge followed by two waves of alveolar proliferation and differentiation to produce a functional lactating gland at birth [15, 19-21]. Upon weaning, the gland involutes, to remove these temporary, and now redundant, branches and alveoli while retaining the permanent ductal system and regenerating the fat pad [22]. This cycle of development and destruction is repeated with each pregnancy. Thus, three functional types of stem/progenitor cells are required to support the natural life-cycle of the mammary gland: embryonic, pubertal and long-lived adult progenitors [18, 23].

The existence of regenerative mammary cells dispersed throughout the mature ductal system was established by seminal experiments demonstrating that fragments from any part of the mammary epithelium could regenerate an entire gland when transplanted into a cleared mammary fat pad [24]. Serial passage of fragments and barcoded mammary cells provided the first evidence of a mammary hierarchy, with fully potent stem cells at the apex giving rise to more restricted ductal and alveolar progenitors [25, 26]. These pioneering studies paved the way for analyses of cell subpopulations, defined by their surface marker expression and regenerative properties as mammary repopulating units (MRU) or mammary stem cells (MaSCs) [27, 28]. In concert with lineage tracing and single cell RNA sequencing, these analyses produced a working model of the mammary hierarchy [14, 23, 27-36].

There is general agreement that the hierarchy begins with multipotent ectodermal stem cells, which give rise to bipotent embryonic mammary stem cells and long-lived unipotent luminal- and basal-restricted mammary progenitors. However, the developmental timing of potency restriction remains a source of debate [14, 23, 30-32, 36-39]. A major gap in our knowledge concerns the location of putative stem and progenitor populations. Attempts to address this problem have identified mutually exclusive cell populations at disparate sites [23, 40-45]. For example, cells that express Lgr5 are restricted to the nipple zone, whereas cells expressing s-SHIP are confined to ductal tips and branches, and cells expressing Procr and Bcl11b are dispersed along ductal borders [42-45]. Moreover, many of these proteins are expressed in additional mammary populations therein compromising

their specificity as stem/progenitor markers. To date no specific unifying progenitor hallmark has been identified. Here we show that Gpr125 identifies long-lived progenitors at multiple sites and stages of mammary development. These cells are engaged in collective cell migration and their expansion is associated with reduced tumor latency and poor outcome in cancer.

Results

Gpr125 is expressed in nipple-proximal regions, terminal end buds and cells dispersed along mature ducts.

As nothing is known about Gpr125 in the mammary gland we began by analyzing mRNA levels in whole mammary glands by qPCR over the course of mammary development (**Figure S1A**). Gpr125 mRNA was high during puberty, decreasing as mice reached maturity (12 weeks) and was higher during earlier stages of pregnancy than later. This temporal pattern indicates that Gpr125 mRNA levels are elevated during ductal elongation and branching, and decrease as alveoli expand and differentiate.

Next, we determined where Gpr125 protein is expressed by X-gal staining tissue from *Adgra3*$^{lz/+}$ mice in which β-galactosidase is fused to the first transmembrane region of Gpr125 (**Figure 1A, S1B, D**). *Adgra3*$^{lz/+}$ mice are viable, fertile and indistinguishable from wildtypes (**Figure S1E**). *Adgra3*$^{lz/lz}$ mice show an identical expression pattern but have delayed ductal elongation that rectifies by the end of puberty (**Figure S1F**). Gpr125-β-gal was expressed throughout the dormant pre-pubescent mammary tree (**Figure 1B, C**). As pubertal ductal elongation began, the X-gal staining pattern partitioned. Weak staining was retained in nipple-proximal ducts (**Figure 1D-F**) and persisted there throughout postnatal development. In contrast, robust Gpr125-β-gal expression appeared in proliferative TEBs (**Figure 1E-G**), and became reduced to an intense dot at ductal tips as the TEB reached the edges of the fat pad and regressed (arrows **Figure 1E**). In histological sections, Gpr125-β-gal was expressed in cap cells (**Figure 1G**) and also in

single cells dispersed among the basal layer of mature ducts (**Figure 1G**, inset). Gpr125-β-gal colocalized with SMA, p63 and low K14 (**Figure 1H-J**) but was absent from cells expressing E-cadherin (Ecad), estrogen receptor (ER) and progesterone receptor (PR) (**Figure 1K-M**). Co-expression of Gpr125-β-gal with PCNA and exclusion of p27 indicated their proliferative potential (**Figure 1N, O**).

Pubertal Gpr125+ cells display a marker profile of regenerative cell types.

To characterize Gpr125+ cells further we mined the Tabula Muris scRNAseq dataset [46]. Gpr125 mRNA was detected within a subset of the "basal" epithelial cell cluster and some stromal cells (**Figure 2A**). We confirmed Gpr125 protein expression within a subset of basal cells by flow cytometry using a fluorogenic β-galactosidase (β-gal) substrate: Fluorescein di-β-D-galactopyranoside (FDG). Gating for $Gpr125^{+}/FDG^{+}$ cells in suspensions of total mammary epithelial cells (MECs) from 6-week old pubertal *Adgra3*$^{lz/+}$ mice produced a clear enrichment for cells within the basal ($CD24^{med/low}CD49f^{+/Hi}$) population; FDG^{-} cells were concomitantly depleted within this gate (**Figure 2B**). Importantly, within the basal population, $Gpr125^{+}/FDG^{+}$ cells displayed the highest level of integrins α6 and β1 (CD49f and CD29) (**Figure 2C**), which are hallmarks of regenerative MRU/MaSCs [27, 28]. Conversely, $Gpr125^{+}/FDG^{+}$ cells were low for the luminal progenitor marker CD61 (integrin β3) and negative for Sca-1 (**Figure 2C**) which is expressed on more committed cell types [27, 47, 48]. Gpr125-β-gal and s-SHIP-EGFP were both expressed in all cap cells surrounding the TEB (**Figure 2D**). By scRNAseq, cells expressing Gpr125 mRNA co-clustered with Lgr5 populations in datasets of embryonic,

pubertal and adult mammary epithelial cells (**Figures 2A, S2**)[34, 35, 37]. Collectively, these data show that Gpr125+ cells are found at multiple sites of predicted stem/progenitor activity, display a surface profile consistent with cell types high within the putative mammary hierarchy, encompass disparate cell populations with documented regenerative capacity and lack markers associated with lineage commitment.

Pubertal Gpr125+ cells are long-lived unipotent basal progenitors.

To position pubertal Gpr125+ cells within the mammary hierarchy, we carried out lineage tracing to determine their physiological potency. We generated a mouse strain harboring a *creER^{T2}* module inserted after the endogenous *Adgra3* promoter (**Figures 3A and S1B,C**), crossed them to the Tomato (tdT) lineage reporter strain (B6.Cg-Gt(ROSA)26Sor$^{tm14(CAG-tdTomato)Hze/J}$) and initiated tracing in 5-week-old female progeny by delivering Tamoxifen (Tam) via IP (**Figure 3B**). 48 hours later few cells expressed tdT (data not shown). After two weeks we observed clusters of tdT+ cells within the basal layer of ducts (Figure 3) that were most abundant in the nipple-proximal region (Figure 3C). tdT+ cells displayed the characteristic bipolar shape of myoepithelial cells and by immunofluorescence, co-localized with the basal markers K5, K14, SMA and p63 (**Figure 3D-G**) and were devoid of the luminal markers Ecad and K8 (**Figure 3H,I**). tdT was also found in cap cells of TEB (**Figure 3J-L**) and in strips of cells extending down the basal surface of subtending ducts.

Next, we mated mice in which tracing had been initiated during puberty and analyzed their glands during pregnancy. Again, tdT+ cells were basally restricted. At p15.5. clusters of tdT+ cells were observed colocalizing with a subset of K5-expressing cells (**Figure 3M**) and surrounding immature Ecad+ alveoli (**Figure 3N**). At lactation day 6 (L6), tdT+ cells displayed the typical basket-like features of contractile myoepithelial cells, enmeshing fully differentiated Ecad+ alveoli (**Figure 3O, S3 movie**). We addressed the longevity of the Gpr125^{+} progenitors labeled during puberty by tracing their tdT+ progeny in both aged nulliparous mice and multiparous mice. In both, clusters and extensive strips of elongated tdT+ cells were found along basal ductal borders (**Figure 3P-R**). Collectively, these data show Gpr125 identifies long-lived unipotent basal progenitors.

Gpr125+ cells are located on the leading tips of side branches during pregnancy.

During early pregnancy (p12-13.5) Gpr125-β-gal appeared focally where side-branches emerge from ducts (**Figure 4A-E**) and became concentrated at branch tips (**Figure 4F**). It persisted along the basal layer of permanent ducts during late pregnancy but was absent from differentiating alveoli (**Figure 4G-I**) with the exception of rare Gpr125-β-gal+ cells that likely represent the branch tip of each alveolar cluster (**Figure 4H arrowheads**).

To interrogate the potency of cells expressing Gpr125+ cells during pregnancy we initiated tracing in pregnant mice at p13.5 and analyzed their glands just prior to birth and during lactation (**Figure 4I**). Again, TdT+ cells exclusively displayed myoepithelial characteristics (**Figure 4J-N**).

Gpr125 identifies bipotent and unipotent progenitors during early embryogenesis. Extending our Gpr125 studies to embryonic mammary development, we found that early *Adgra3*$^{lz/+}$ embryos showed diffuse ectodermal Gpr125-β-gal expression that by E14.5 concentrated into ectodermal appendages, such as whisker follicles but was absent from the mammary line, placodes and buds (**arrows Figure 5A, A'**). However, at E15 Gpr125-β-gal appeared in the mammary sprout (**arrows Figure 5B, B'**), coincident with the onset of proliferation, indicated by nuclear PCNA and BrdU (**Figure 5C,D**), and was concentrated towards the leading tip together with K14 and p63 (**Figure 5E, F**). At this stage Gpr125-β-gal also became concentrated in the "bulge" stem cell compartment of hair follicles and whiskers (**Figure S4**). At E18.5 Gpr125-β-gal was strongly expressed in the rudimentary tree (**Figure 5G**) and concentrated in the lactiferous duct and the multilayered branch tips (**Figure 5H**). Although K14 expression was present in most cells (**Figure 5I**) and K18 was increased in central cells (**Figure 5J**), Gpr125+ branch tips lacked both keratins (**Figure 5I,J arrow and insets**) but expressed p63 (**Figure 5K**).

To test the potency of embryonic Gpr125+ cells we crossed *Adgra3*$^{cre/cre}$ to tdT mice and administered Tam to the pregnant dams to label embryos at E14.5 and at E19 (**Figure 5L**) then analyzed these pups at 8 weeks of age. Glands from mice labeled at E14.5 revealed tdT not only in basally-located bipolar cells co-expressing SMA (**Figure 5M**), but also in columnar cells situated above the SMA+ basal layer (**Figure 5N**) that expressed

the luminal marker Ecad (**Figure 5O**). By contrast, when tracing was initiated at E19, tdT+ cells were exclusively basal, bipolar and SMA+ (**Figure 5P**).

These data show that Gpr125 appears at the onset of directed growth in a bipotent progenitor population expressing markers of both lineages (K14/K18). However, before birth Gpr125 cells become lineage-restricted and from thereon function as unipotent basal progenitors. Of note, Gpr125 expression concentrates at this stage in undifferentiated $p63^{+}K14\text{-}/K18^{-}$ cells at branch tips poised for ductal extension.

Gpr125+ cells retaining embryonic undifferentiated characteristics and bipotency are expanded in MMTV-Wnt1 glands.

Next we investigated Gpr125 in mouse breast cancer models. Gpr125 showed the highest mRNA expression in MMTV-Wnt1 mice (**Table S5A**) which develop mixed lineage tumors enriched in cells with MaSCs profiles and show transcriptomic resemblance to basal-type breast cancer [49-54]. We therefore generated *MMTV-Wnt1*; *Adgra3*$^{lz/+}$ mice and analyzed glands over the course of tumor progression. In 8-week old mice, Gpr125-β-gal was expressed exclusively in basally-located cells within the nipple proximal hyperplasia and on hyperbranched tips (**Figure 6A-D**). Robust expression was seen in tumors, where Gpr+ cells formed large homogenous regions and concentrated at pushing margins (**Figure E-H**). Although Wnt-1 tumors display regions composed of K14+ and K18+ bilayers these populations rarely overlapped with Gpr125. Gpr125+ cells also lacked both SMA and K8 and although some expressed K14, the majority lacked both keratins

(**Figure 6G-I)**. They did, however, express p63 as well as Tcf1 (**Figure 6J-K**). In normal gland the latter is found exclusively in a subset of embryonic cells and TEB cap cells (**Figure S2 6L**). To determine the effect of Wnt expression on the potency of Gpr125 expressing cells, we generated *MMTV-Wnt1;Adgra3*$^{cre/+}$*;tdT* mice and performed lineage tracing by delivering Tam to 5-week old pubertal mice and harvesting hyperplastic glands from 12-week old mice (**Figure 6M**). tdT was found in both SMA+ and Ecad+ cells (**Figure 6N, O**). These data indicate that in the context of MMTV-Wnt1 transformation pubertal Gpr125+ cells retain the undifferentiated characteristics and bipotency of embryonic progenitors.

High Gpr125 expression is indicative of early tumor onset in mice and poor outcome in human basal-type breast cancer.

We noted that Gpr125-β-gal expression was consistently more extensive in tumors arising with shorter latency (**Figure 7A**). MMTV-Wnt1 tumors have been separated into two subtypes with distinct gene expression (ex) profiles: Wnt1-early(ex) and Wnt1-late(ex), that correlate with early (average 6.5 weeks) and late (average 22.5 weeks) tumor onset respectively [49-51]. Mining these microarray data, we found Wnt1-early(ex) tumors have twice the level of Gpr125 mRNA as Wnt1-late(ex) tumors (**Table S5A**) and confirmed this by qPCR analysis ($p=0.0052$ $n=6$) (**Figure 7B**). To investigate the course of Gpr125+ cell expansion we carried out flow cytometry. In hyperplastic glands, the Gpr125^{+}/FDG^{+} population localized within the traditional basal gate but in tumors it was expressed in a new population with intermediate CD49 levels (**Figure 7C**) that was more pronounced in

the uninvolved glands and tumors that arose early. Collectively, these data show that expansion of the Gpr125 tumor population correlates with early tumor onset in mice. We investigated GPR125 expression in human breast cancer and found it was highest within the basal-type (**Table S5B**). Interrogating publicly available datasets using the km plotter algorithm[55] showed high GPR125 levels within this subtype correlated with worse patient outcome in terms of relapse-free survival p=0.0054 and distant metastasis free survival (p=0.0043) (**Figure 7D,E red line**), and with disease-free survival using the breastmark algorithm[56] (**Figure 7F, blue line**).

Discussion

Here, we present the first report of Gpr125 expression and function in the developing mammary gland and breast cancer. Our results demonstrate the powerful ability of Gpr125 to specifically localize progenitors at multiple sites and stages of mammary development (see graphical abstract). We show that Gpr125 cells are concentrated at the tips of migrating structures during normal development and massed at pushing margins in tumors, where their expansion correlates with early tumor onset in mice and poor outcome in humans.

Gpr125 marks bipotent embryonic and unipotent perinatal mammary progenitors.

Our analyses show that in the embryo Gpr125 is expressed in stem cell compartments of other ectodermal appendages such as hair follicles and whiskers (**Figure S4**). This, together with its expression in lacrimal and spermatogonial progenitors, indicates that Gpr125 has important value in demarcating early progenitors in multiple tissues [6]. In the embryonic mammary gland, we found that E15 Gpr125+ cells are bipotent. Whether they are mammary stem cells or comprise a mixture of pre-committed unipotent basal and luminal progenitors will further require clonal analysis. In support of the former scenario, scRNAseq and molecular profiling studies could find no evidence of lineage bias among early embryonic progenitors [36, 37, 57]. However, by E19, Gpr125-expressing cells clearly become committed unipotent basal progenitors consistent with recent studies by Fre et al. that have indicated that lineage-restriction begins early and is completed before birth [32, 38].

Gpr125 Identifies unipotent basal progenitors in multiple locations in postnatal glands.

During postnatal mammary development Gpr125 identifies long-lived unipotent basal progenitors at multiple locations. Gpr125 expression in the nipple proximal zone resembles that of Lgr5 [43, 58, 59]. Gpr125 and s-SHIP are both present in all cap cells of the TEB [42, 60]. Gpr125 expression in progenitors dispersed throughout mature ductal system is consistent with findings from the earliest mammary transplantation studies and similar to that of Procr and Bcl11b [24-26 44, 45]. While the precise relationship of Gpr125+ cells to these diverse cell types requires further delineation, Gpr125 coexpression in Lgr5+ and s-SHIP+ cells shows it to be a common link between disparate progenitors. Thus whereas mammary progenitor heterogeneity likely reflects the need to respond to niche and developmental stage-specific signals, we propose Gpr125 must serve a more universal role in progenitor biology. It is possible that the presence of Gpr125 in these progenitors relates to its involvement in Wnt signaling. Wnt signaling is critical at all stages of mammary development and has been shown to sustain progenitor potency in vitro [18, 33, 52, 53, 61, 62]. Intriguingly, the closely related protein, Gpr124, selectively promotes canonical signaling by specific Wnt ligands[63]. In contrast, Gpr125, has been implicated in non-canonical Wnt signaling and detected in biochemical complexes with Ryk[7, 12 64].

Gpr125 marks sites of collective cell migration

It is striking that in several tissues Gpr125 is concentrated in a graded fashion at sites of directed migration [5]. These include the tips of the mammary sprout, rudimentary tree, pubertal TEB and side-branches. Its expression in these locations suggest an involvement of Gpr125 in directing progenitor migration. This concept is supported by previous reports showing Gpr125 levels influence facial motor neuronal precursor migration in zebrafish[7]. Indeed, regulation of directed cell movement may be a common function of the ADGRA family as Gpr124 is required for tip cell function in endothelia during angiogenesis [63].

High Gpr125 expression is indicative of early tumor onset in mice and poor outcome in human basal-type breast cancer.

A role in cell migration has considerable clinical significance in the setting of breast cancer, a disease where metastasis is the primary cause of death. Preliminary in silico analyses suggest Gpr125 could have prognostic utility in basal type breast cancer, where there is a great unmet need for markers that can stratify risk. In mice, Gpr125 mRNA is elevated and Gpr125+ cells, lacking lineage-specific keratins are greatly expanded in MMTV-Wnt1 tumors arising with short latency [49]. This Gpr125+ tumor cell-type shares features with Gpr125 progenitors found the multilayered tips of the embryonic mammary tree and pubertal TEB suggesting its pathological contribution to reduced latency may relate to its acquisition of intrinsic migratory and invasive properties of embryonic and pubertal progenitors.

Figure Legends 1-7

Figure 1. Gpr125 is expressed as predicted sites of stem/progenitor activity during pubertal mammary development.

A) Left: Schematic of Gpr125 protein, **Right**: Gpr125-β-gal fusion protein. N-terminus (N), leucine rich repeats (LRR), Immunoglobulin-like domain (Ig), hormone binding domain (HBD), GPCR autoproteolytic-inducing (GAIN) domain, transmembrane region (TM) and cytoplasmic region (C). **B-F)** Gpr125-β-gal expression in mammary whole mounts from pre-pubertal (3w) and pubertal (5w and 7w) nulliparous mice. Scale bar 2mm. **G)** section of X-gal treated whole mount counterstained with nuclear fast red (NFR), shows Gpr125 expression in the cap layer of terminal end buds (TEB) and in cells dispersed along the basal layer of subtending ducts. The inset box is a higher magnification of area indicated by arrows. **H-M)** X-gal stained sections of TEB with immunolocalization of (**H-J**) basal markers: smooth muscle actin (SMA), p63, Keratin (K14); Note the occasional cells expressing Gpr125 within the body layer all express basal cap cell markers. (**K-M**) luminal markers: E-cadherin (Ecad), estrogen receptor (ER), progesterone receptor (PR); (**N,O)** markers indicating proliferative status: proliferating nuclear cell antigen (PCNA) and p27. Scale bar in sections are 50μm.

Figure 2. Gpr125+ cells have an MaSC/MRU profile and encompass distinct regenerative populations.

A) t-SNE plots of *Adgra3*, *Lgr5* and *Procr* in total mammary cells mined from Tabula Muris dataset [46]. **B)** Representative FACS dot plots of total MECs isolated from 6-week *Adgra3*$^{lz/+}$ pubertal mice stained with antibodies against CD24 and CD49f. Nb. Gpr125^{+}/FDG^{+} cells gated within the CD24$^{med/low}$/CD49f$^{+/hi}$ basal population and the basal population is depleted in FDG- populations. **C)** Histograms of expression of CD49f, CD29, Sca1 and CD61 in Gpr125^{+}/FDG^{+} (red line), Gpr125^{-}/FDG^{-}(black line) and total MECs (grey line). **D)** TEB sections show expression of Gpr125-β-gal (**left**) and sShip-EGFP cells in the cap cell layer of TEB (**right panel**). Scale bar = 50μm.

Figure 3. Pubertal Gpr125+ cells are long-lived unipotent basal progenitors.

A) Genetic strategy used to target Tomato expression to Gpr125-expressing cells and their progeny. **B)** Schematic illustrating strategy to trace the lineage of cells that express Gpr125 during early puberty by delivering Tamoxifen (Tam) to (5w) nulliparous mice and harvesting glands during late puberty (7w), mid-pregnancy (P15); lactation(L6) and in aged mice. **C)** Representative 3D-imaging of mammary gland nipple-proximal region of *Adgra3-CreER*^{T2}*;tdT* mice analyzed 2 weeks after labeling. **D-G)** tdT+ cells with immunolocalization of basal markers K5, K14, SMA, and p63, and exclusion of luminal markers Ecad, K8 **(H,I)**. **J)** tdT+ cells in the cap cell layer of TEB and subtending duct with immunolocalization of basal K5 **(K)** and exclusion of luminal Ecad **(L)**. **M)** p15 Immature alveoli show basally located tdT+ cells that express K5, but lack K8 expression

(N). O) tdT+ cells form characteristic basket-like morphology of mature myoepithelial cells enmeshing an Ecad+ alveoli in lactating mammary gland. **P-R)** Extensive strips of tdT-positive cells along the outer basal layer of ducts in aged mice **(P,Q)** and after multiple pregnancies **(R).** Scale bar 50μm. Dapi = nuclear staining. Two glands from each of 5 mice were analyzed/stage.

Figure 4. Cells expressing Gpr125 on leading tips of branches during pregnancy are unipotent basal progenitors.

X-gal stained mammary whole mounts from pregnant mice showing Gpr125-β-gal expression at: **A)** sites of emerging side-branches at 12.5 days of pregnancy (p12.5), **F)** tips of elongating side-branches at p13.5 and **G)** in ducts but not alveoli at p15.5. Scale bar 500μm. **B, C, H, I)** X-gal stained sections counterstained with NFR. Boxed insets are higher magnification of regions indicated by arrows. **D, E)** X-gal stained sections of a permanent duct and side branch with immunolocalization of basal marker p63 and K14. Scale bar 20μm. **J)** Schematic illustrating strategy to trace the lineage of cells that express Gpr125 during early pregnancy by delivering Tam at p14.5 and harvesting glands at p20.5 and L6 mice. **K)** Tdt+ Sma+ cells were found in ducts and alveoli at p20.5 and at birth **(L)**. **M)** Fully differentiated tdt+ cells were devoid of luminal Ecad marker at L6. Scale bar 50μm. 1 gland from each of 5 mice were analyzed.

Figure 5. Gpr125 identifies an early bipotent and later unipotent basal progenitor population during embryogenesis.

A, B) X-gal stained *Adgra3*$^{lz/+}$ embryos at embryonic day (E)14 **(A)** and E15 **(B)**. Arrows indicate mammary buds and sprout respectively magnified in boxes below. Scale bar 200μm. **A'-F)** Gpr125-β-gal expression in sections of E14 buds **(A')** and E15 sprouts**(B'-F),** with immunolocalization for proliferative markers PCNA and BrdU **(C, D**), and K14 and p63 **(E,F)**. Scale bar 50μm. **G)** X-gal stained skin whole mount showing Gpr125-β-gal expression in the E18.5 tree and hair follicles encircling the developing nipple zone. Scale bar 1mm. **H-K)** Sections of the E18.5 rudiment tree stained with X-gal followed by NFR counterstain**(H)**; immunochemical detection of K14, K18, and p63 **(I-K)**. Boxed insets are higher magnification of branch tips regions indicated by arrows. Scale bar 50μm. **L)** Tracing of Gpr125+ cells in E14.5 or E19.5 embryos was initiated by delivering Tam in pregnant *Adgra3*$^{cre/cre}$ dams mated to Tdt mice. Mammary tissue from the progeny were analyzed at 8 weeks of age. **M-O)** 3-D images showing representative regions of pubertal ducts from E14.5 labelled embryo containing clusters of basally located tdT+ cells that co-express the basal marker SMA **(M)** as well as tdT+ columnar luminal cells lacking SMA **(N)** and expressing luminal marker Ecad **(O)**. **P)** Glands from progeny labeled at E19.5 show basally located tdT+ cells devoid of Ecad. Dapi=nuclear staining. Scale bar 50μm. Two glands from each of 3 mice were analyzed/stage.

Figure 6. Gpr125+ progenitors are expanded in MMTV-Wnt1 tumors and retain embryonic features and bipotency.

A-C) X-gal stained mammary whole mounts from 12-week old *MMTV-Wnt1;Adgra3$^{lz/+}$* mice show robust Gpr125-β-gal expression in hyperbranched ductal tips **(C)** and nipple-proximal zones compared to **(B)** control *Adgra3$^{lz/+}$* littermate. **D)** X-gal section of hyperbranched ductal tip counterstained with NFR presenting basal restriction of Gpr125-β-gal expression in 12-week *MMTV-Wnt1;Adgra3$^{lz/+}$* mice. **E-H)** X-gal stained whole mount and sections counterstained with NFR of *MMTV-Wnt1;Adgra3$^{lz/+}$* tumor (**G-K)** showing Gpr125 cells devoid of immunolocalization for SMA, K14 or K8, but expressing p63 and Tcf1. **L)** Tcf1 expression in the cap cells of normal TEB. **M-O)** Lineage tracing strategy in 5w *MMTV-Wnt1;Adgra3$^{lz/+}$* produced both **(N)** K5+ and **(O)** Ecad+ Tdt cells in hyperplastic glands. 1 gland from each of 5 mice were analyzed.

Figure 7. High Gpr125 is predictive of poor outcome.

A) X-gal/NFR stained sections showing higher Gpr125-β-gal expression in a short (7w) latency tumor *MMTV-Wnt1;Adgra3$^{lz/+}$* tumor (Wnt1-early) versus a long (36w) latency tumor (Wnt1-late) Scale bar 100μm and 50μm. **B)** Gpr125 mRNA levels in early and long latency tumors *MMTV-Wnt1;Adgra3$^{lz/+}$* tumors (n=3 samples for each subtype, p=0.0052,**). **C)** Representative FACS dot plot of total MECs from *MMTV-Wnt1;Adgra3$^{lz/+}$* hyperplastic uninvolved glands and associated short and long latency tumors stained with CD24 and CD49f. Plots are representative of 3 independent experiments. **D,E)** Relapse-free survival and Distant Metastasis Free Survival for high

(**red lines**) and low Gpr125 mRNA expression (**black lines**) in human basal-like breast cancer subtype and autobestfit cutoff to divide patients into high and low expression (indicated in beehive plots) sourced from KMplotter (https://kmplot.com/analysis/index.php?p=service&cancer=breast)[55]. logrank P-value and Hazard Ratio (HR) indicate a significant association between high expression and poor prognosis. **F**) Disease-free survival utilizing BreastMark [56] ssp 2003/2006 datasets and splitting patients at the median level for expression (http://glados.ucd.ie/BreastMark/) into low (red line) and high (blue line) Gpr125 expression groups

Figure S1. *Adgra3*$^{lz/+}$ and *Adgra3*$^{cre/+}$ mouse strains

A) Bar plot of Gpr125 relative mRNA levels at different stages of mammary development normalized to male mammary glands; wV = weeks of age virgin nulliparous mice, dP = days of pregnancy. Each bar represents the mean ±SEM on 2-5 mice/stage with 4 technical replicates each**.** **B)** Schematic of *Adgra3* gene. **C)** *Adgra3*$^{cre/+}$ mice were generated by replacement of 502 bp after the first codon with a cassette containing c*reER*^{T2}. **D)** *Adgra3*$^{lz/+}$ mice were generated by deletion of 10 kb sequence downstream of the first TM and replacement by *lacZ* inserted in frame using Regeneron Velocigene technology [6]. **E)** Carmine stained mammary whole mounts from 4-week old and 13.5 days pregnant *Adgra3*$^{lz/+}$ and *Adgra3*$^{lz/lz}$ mice **F)** Mild impairment in ductal elongation in mammary wholemounts from pubertal *Adgra3*$^{lz/lz}$ females compared to littermates. Ducts from 7w old mice *Adgra3*$^{lz/+}$ mice have elongated further than those of *Adgra3*$^{lz/lz}$ littermates with respect to the midpoint of the inguinal lymph node marked by the dashed line. **G)** Quantitation of ductal elongation in glands from *Adgra3*$^{lz/lz}$ compared with pooled glands from *Adgra3*$^{lz/+}$ littermates ($p<0.0001$)****, n=6.

Figure S2. Gpr125 population is confined to the basal cluster in virgin and adult mice.

A) t-SNE plots of *Krt18, Krt14, Lgr5, Adgra3, Lef1 and Tcf7* in embryonic mammary cells mined from dataset generated by Wuidart et al [37]

B,C) t-SNE plots of *Krt18*, *Krt14* and *Adgra3* in pubertal and adults mammary cells mined from dataset generated by Pal et al [35]

Figure S3.video

Video of 3D confocal imaging of tdT+ cells in glands from *Adgra3*$^{cre/+;}$*tdT* mice treated with Tam during puberty. Glands were harvested at lactation day 6 (L6) stained with DAPI and antibodies against Ecad.

Figure S4. Gpr125 expression in other ectodermal appendages

A,B) Skin from E18 embryo stained with X-gal shows Gpr125-β-gal expression in the five pairs of mammary trees and developing hair follicles. **C)** Whole mount of E14 embryo showing robust Gpr125-β-gal expression in the whisker-pad. **D)** Sections of E14 embryos show Gpr125-β-gal expression is concentrated in the "bulge" stem cell compartment of E14 whisker follicles and bulge and bulb compartments of P2 hair follicles during anagen **(E)** immunolocalized with p63 as well as in the secondary germ of P18 hair follicles during telogen.

Figure S5. Gpr125 expression in murine and human breast cancer subtypes

A) Gpr125 mRNA fold change (from high to low mRNA level) in several mouse tumor models of breast cancers showing greater expression in Wnt1-Early^{Ex}, p53null-Basal^{Ex} and Wnt1-Late^{Ex} tumors. **B)** Gpr125 mRNA fold change higher in Human Basal like. Gpr125 mRNA fold change was obtained mining microarray data available at the Gene Expression Omnibus under the series GSE3165 [54].

Methods

Ethics statement. All animal protocols were approved by the Institutional Animal Care and Use Committee (IACUC) at NYU School of Medicine.

Mice. Mice were constructed by Ingenious Technologies, Ronkonkoma, NY as follows. A cassette containing *CreER^{T2}* followed by a 3' polyadenylation signal, harboring SV40-driven Neo flanked by FRT sites inserted in a central intron, was recombined into a bacterial artificial chromosome (BAC) to place *CreER^{T2}* under the control of the *Adgra3* promoter, excising 502 bp encompassing 221 bp of exon 1 and part of the following intron 1-2 of *Adgra3*. Mice generated from these ES cells were selected for germline transmission by PCR, verified by southern analysis and sequencing then bred to a Flp deleter strain to remove Neo. *Adgra3$^{lz/+}$* mice were generated by Regeneron using VelociGene methods [65] to modify a bacterial artificial chromosome (BAC) clone carrying the mouse *Adgra3* gene by replacement of sequence encompassing exons 16-19 with *lacZ* to produce expression of fusion protein comprising the N-terminal extracellular domain, the first transmembrane domain, and part of the first intracellular loop of Gpr125 fused to β-galactosidase (Figure 1A) [6].

Lineage Tracing: For lineage tracing experiments, *Adgra3-CreER^{T2}* mice were crossed to the fluorescent Rosa26R-lox.STOP.lox-tdTomato (tdT) lineage reporter strain (Stock No. 007909) Jackson laboratory. The transcriptional STOP was deleted by cre recombination by delivering tamoxifen. For tracing at mid-puberty tamoxifen was delivered ip at low dose: 1.5mg, and high dose: 5-15mg (delivered in 2.5mg aliquots every

other day). For tracing during pregnancy and embryogenesis 2 doses of 2.5mg Tam was given by oral gavage to *Adgar3*$^{cre/cre}$ pregnant dams over a 24hr time period. Pups were delivered at E19.5-E20.5 by caesarian section to avoid Tam-induced problems with delivery and fostered by SWR/J mice. Tissue was harvested at the indicated intervals over course of mammary gland development. For each experiment, two mammary glands of at least three mice were analysed. No fluorescence was observed in non-induced mice.

Tissue clearing and 3-D imaging: Mammary glands were excised and fixed overnight in 4% PFA then processed using a CUBIC protocol optimized for mammary gland [32, 66]. Tissue was incubated in CUBIC Reagent 1A(10wt% Triton,5wt% N,N,N',N'-tetrakis (2-HP)ethylenediamine, 10wt% Urea, NaCl 25mM) clearing solution for 4 days, rinsed 3X in PBS, then incubated at 4C for 4 days in primary antibodies diluted in PBST containing 10% serum, rinsed again, then incubated in secondary antibody for 2 days, rinsed 3X, then cleared in CUBIC Reagent 2 (50w/v% Sucrose, 25w/v% Urea, 10w/v% Triethanolamine, 0.1w/v% Triton) for 24hrs. Primary antibodies: rabbit anti-K5 (Covance, PRB160P, 1:100); rat anti-K8 (Developmental Studies Hybridoma Bank, TROMA-I, 1:50); mouse anti-SMA (Dako, M0851, 1:100); rabbit anti-E-cadherin (Cell Signaling, 3195S, 1:100); rabbit anti-p63 (Abcam, ab124762,1:100); rabbit anti K14 (Abcam,Ab181595). Alexa Fluor-conjugated secondary antibodies Thermo Fisher Scientific, diluted 1:500: goat anti-mouse 647 (A21237); goat anti-rat 647 (A21247); goat anti-rabbit 647 (A21245). Cleared mammary tissues were imaged using a Zeiss 880 Laser Scanning inverted confocal microscope with 10X, 20X air Plan-Apochromat N.A. 0.8 M27 objective lenses.

X-gal staining: Embryos and mammary glands were fixed in 4% paraformaldehyde (PFA) (Sigma-Aldrich) at room temperature (RT) for 30-60 min, rinsed 3X in X-gal rinse buffer (2 mM MgCl2, 0.1% Sodium deoxycholate, and 0.2% NP-40 in PBS) at RT, then incubated in X-gal staining solution (50 mg/ml 5-bromo-4-chloro-3-indolyl-β Dgalactopyranoside in rinse buffer containing 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide) (Applichem, Cheshire, CT) at RT overnight. After staining, glands were rinsed in PBS, post-fixed in 4% PFA overnight then prepared for whole mount analysis or processed for paraffin embedding, sectioning and histological analysis [67, 68].

Mammary gland whole mounts. X-gal stained whole mounts were post-fixed in 4% PFA, washed twice with 1X PBS, dehydrated through an increasing ethanol gradient, cleared of lipids in Carnoy's Fixative (60% Ethanol, 30% Chloroform, 10% Glacial Acetic Acid) for two hours, and further cleared in Citrisolv (Fisher Scientific, Suwanne, GA) for two hours. Glands were pressed flat between the slide and coverslip under a heavy weight for 30 minutes, and imaged on a Leica dissecting microscope Model WILD M3Z (Leica Microsystems, Bannockburn, IL) with an Optronics digital camera Model 60800 (Goleta, CA). The glands were then re-hydrated through a decreasing ethanol gradient and counterstained with Carmine alum (500mL distilled water containing 1g Carmine and 2.5g aluminum potassium sulfate; Sigma Aldrich, St Louis, MO) diluted 1:4 in distilled water. Glands were once again dehydrated in ethanol, cleared in Carnoy's Fixative and Citrisolv,

and pressed flat before mounting under a coverslip with Cytoseal (VWR, West Chester PA) then re-photographed [69].

Immunohistochemistry: Mammary glands were removed from mice and fixed with either 10% neutral buffered formalin or 4% paraformaldehyde (PFA) and embedded in paraffin. Slides containing paraffin-embedded tissue sections were cleared of paraffin by incubation in a 60°C oven for one hour, and submerged in Citrisolv for 10 minutes. Sections were rehydrated through an ethanol gradient and rinsed in distilled water for 10 minutes. Antigen retrieval was performed by microwaving at 900 watts for 30 minutes in 10mM Citric Acid buffered to pH 6. From this point forward, the slides were washed thrice with 1X PBS between each step. Endogenous peroxidase activity was quenched by treating slides with 3% Hydrogen Peroxide (Sigma Aldrich) for 15 minutes at room temperature. Slides were blocked with 20% Normal Goat Serum for 30 minutes to reduce background signal. Primary antibodies were diluted in 2% bovine serum albumin (BSA, Sigma-Aldrich) in 1X PBS, and incubated at 4°C overnight. Primary antibodies: rabbit anti-K14 (Covance PRB-155P 1:500); mouse anti-K8 undiluted (Progen 65138); mouse anti-p63 (Neomarkers MS-1081-P1 1:300); mouse anti E-cadherin (BD 610182 1:100); mouse anti-PCNA (Dako M0987 1:500); rabbit anti TCF1 (Cell Signaling 22035 1:100);. Biotinlyated secondary antibodies were diluted in 2% BSA/PBS for 1 hour at room temperature, followed by HRP-conjugated Streptavidin (Vector Labs, Burlingame CA) for 30 minutes at room temperature. Colorimetric signal was developed using the DAB substrate (Vector Labs).

Flow cytometry. To detect Gpr125- β-gal expression, cells were labelled with fluorescein di-V-galactoside (FDG) according to manufacturer's protocol (Molecular Probes, Eugene, Oregon). Briefly, cells were resuspended at 107/mL in HF, and the samples were pre-warmed at 37°C for 10 minutes. FDG loading was performed by adding an equal volume of pre-warmed 2mM FDG (diluted in distilled water) to the cell suspension for exactly 1 minute at 37°C, then immediately quenched by adding 2mL ice-cold HF. The FDG-loaded cells were then centrifuged and stained with surface antibodies. The following antibodies were used to label cells for flow cytometry: biotinylated- TER119 (BD 553672,1:200), biotinylated-CD31 (BD 558737,1:200), biotinylated-CD45 (BD 553077,1:200), biotinylated-CD140a (eBioscience 12-1401-80,1:200), CD24-PE (BD 553262,1:400), CD49f-PerCP-Cy5.5 (Biolegend 313617,1:200,), CD24-FITC (BD553261,1:100), CD49f-PE (1:100), Streptavidin-AlexaFluor647 (Molecular Probes S21374,1:600), CD61-APC (Caltag,1:200), Sca1-PE-Cy7 (eBioscience 25-5981-81,1:600), CD29-Pacific Blue (Biolegend 102224,1:200). Cells were incubated with conjugated antibodies diluted in HF, for 30 minutes on ice in a dark container, washed with 2mL of HF, and resuspended in 250μL HF for analysis. Cell viability was assessed by adding 4',6-Diamidino-2-phenylindole (DAPI, Sigma-Aldrich) to the final suspension at a concentration of 1μg/mL. Data collection for flow cytometry was done on a Beckton-Dickinson (East Rutherford, NJ) LSRII analyser. Analyses were done using FlowJo software version 9.

Reverse transcription and quantitative PCR. Total mRNA was generated from snap-frozen tissues (30-50mg) with a Qiagen RNA-Easy Mini column (Qiagen, Venio, Limburg). Tissues were homogenized in 1mL Trizol (Life Technologies) using a Polytron PT 1200C homogenizer (Kinematica, Bohemia, NY), mixed with 200µL Chloroform, and centrifuged at 12500rpm for 10 minutes in a 4°C minicentrifuge. The aqueous layer was decanted, mixed with an equal volume of 70% Ethanol, and added to the Qiagen spin column. The remaining preparation was performed according to manufacturer's protocol, and RNA was eluted in a final volume of 30-50µL RNase-free distilled water. RNA concentration and quality were assessed with a Nanodrop 2000 (Thermo Scientific, Waltham, MA) and stored at -80°C. cDNA was made from 100µg of RNA with a Verso cDNA Synthesis Kit 2000 (Thermo Scientific) and stored at -20°C. Quantitative PCR reactions were performed with three replicates of 10ng cDNA on a Bio-rad CFX96 Detection System (Bio-rad, Hercules, CA) using SYBR Green Real-Time PCR Master Mix (Life Technologies). Data was analyzed by the ΔΔCt method.

scRNA-seq analysis: scRNA-seq analysis were generated using available data from Tabula Muris (https://tabula-muris.ds.czbiohub.org/). We also conducted an analysis with two other single cell RNA-seq mammary gland datasets [34, 35]. We processed the dataset using iCellR, Single (i) Cell R package, an interactive R package to work with high-throughput single cell sequencing technologies with the help of NYU Langone's Applied BioinformaticsLaboratories (https://www.biorxiv.org/content/10.1101/2020.03.31.019109v1).

Microarray analysis. Gpr125 mRNA expression in murine and human samples of breast cancer was obtained by using microarray data available at the Gene Expression Omnibus under the series GSE3165 (https://www.ncbi.nlm.nih.gov/geo/). Analysis was conducted using GEO2R. Version info: R 3.2.3, Biobase 2.30.0, GEOquery 2.40.0, limma 3.26.8. Analysis of Gpr125 in human breast cancer was carried out using kmplotter[55] https://kmplot.com/analysis/index.php?p=service&cancer=breast with criteria: Gpr125 affy ID 210473_s_at, Auto select best cutoff, excluding biased arrays and selecting for basal-type breast cancer RFS: N=618; DMSF: N=232 and using BreastMark[56]: http://glados.ucd.ie/BreastMark/mRNA_custom.html DFS, median cut off was selected for ssp2003 basal-type (N=318) [70] or ssp2006 basal-type (N=366) [71] datasets .

Statistics and reproducibility. Statistical significance was determined using GraphPad Prims software. Normal distribution of data was assessed using Shapiro-Wilk normality tests. Unpaired Student's t-test was performed. Data are always expressed as mean ± SEM. Each experiment was repeated independently at least 3 times. P values and N of repeat are indicated in the figure legends.

Acknowledgements:

This work was supported by Department of Defense W81XWH-17-1-0013 BC160959 (PC), NIH-R21-CA129905 and NYS Peter Rowley Award C02857 (PC), NIH training grants: T32GM066704 (JS), T32CA009161-44 (AI) and The Susan G Komen Foundation For The Cure (AI). We thank Dr. Aris Economides, Regeneron, Tarrytown NY for advice on *Adgra3*cre-ER^{T2} construction. We thank Drs. Alireza Khodadadi-Jamayran, Cynthia Loomis, Michael Cammer, Peter Lopez for assistance provided by the Applied Bioinformatics, Experimental Pathology, Microscope Laboratory and Flow Cytometry Research Cores: Grants P30CA016087, S10OD021747 and P30CA016087, and Drs. Alan Frey and David Frendewey for reading the manuscript.

1. Fredriksson, R., Lagerstrom, M.C., Lundin, L.G. & Schioth, H.B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol Pharmacol* **63**, 1256-1272 (2003).
2. Simundza, J. & Cowin, P. Adhesion G-protein-coupled receptors: elusive hybrids come of age. *Cell Commun Adhes* **20**, 213-226 (2013).
3. Fredriksson, R., Gloriam, D.E., Hoglund, P.J., Lagerstrom, M.C. & Schioth, H.B. There exist at least 30 human G-protein-coupled receptors with long Ser/Thr-rich N-termini. *Biochem Biophys Res Commun* **301**, 725-734 (2003).
4. Pickering, C. *et al.* The Adhesion GPCR GPR125 is specifically expressed in the choroid plexus and is upregulated following brain injury. *BMC Neurosci* **9**, 97 (2008).
5. Spina, E.H.R.S., J.; Incassati, A.; Faiq, M.; Singh, A.; Cowin, P. Role of Gpr125 in Tear Film and Dry Eye Disease *manuscript submitted* (2020).
6. Seandel, M. *et al.* Generation of functional multipotent adult stem cells from GPR125+ germline progenitors. *Nature* **449**, 346-350 (2007).
7. Li, X. *et al.* Gpr125 modulates Dishevelled distribution and planar cell polarity signaling. *Development* **140**, 3028-3039 (2013).
8. Yamamoto, Y. *et al.* Direct binding of the human homologue of the Drosophila disc large tumor suppressor gene to seven-pass transmembrane proteins, tumor endothelial marker 5 (TEM5), and a novel TEM5-like protein. *Oncogene* **23**, 3889-3897 (2004).
9. Spiess, K. *et al.* Arrestin-independent constitutive endocytosis of GPR125/ADGRA3. *Ann N Y Acad Sci* **1456**, 186-199 (2019).
10. Wu, Y. *et al.* Elevated G-Protein Receptor 125 (GPR125) Expression Predicts Good Outcomes in Colorectal Cancer and Inhibits Wnt/beta-Catenin Signaling Pathway. *Med Sci Monit* **24**, 6608-6616 (2018).
11. Woods, D.F. & Bryant, P.J. The discs-large tumor suppressor gene of Drosophila encodes a guanylate kinase homolog localized at septate junctions. *Cell* **66**, 451-464 (1991).
12. Fu, J.F. *et al.* Involvement of Gpr125 in the myeloid sarcoma formation induced by cooperating MLL/AF10(OM-LZ) and oncogenic KRAS in a mouse bone marrow transplantation model. *Int J Cancer* **133**, 1792-1802 (2013).
13. Cowin, P. & Wysolmerski, J. Molecular mechanisms guiding embryonic mammary gland development. *Cold Spring Harb Perspect Biol* **2**, a003251 (2010).
14. Davis, F.M. *et al.* Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. *Nat Commun* **7**, 13053 (2016).
15. Watson, C.J. & Khaled, W.T. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development* **135**, 995-1003 (2008).
16. Rosen, J.M. Hormone receptor patterning plays a critical role in normal lobuloalveolar development and breast cancer progression. *Breast Dis* **18**, 3-9 (2003).
17. Daniel, C.W. & Silberstein, G.B. Postnatal development of the rodent mammary gland., in *The Mammary Gland*. (eds. M.C. Neville & C.W. Daniel) 3-31 (Plenum, New York; 1987).
18. Sreekumar, A. *et al.* WNT-Mediated Regulation of FOXO1 Constitutes a Critical Axis Maintaining Pubertal Mammary Stem Cell Homeostasis. *Dev Cell* **43**, 436-448 e436 (2017).

19. Fata, J.E. *et al.* The osteoclast differentiation factor osteoprotegerin-ligand is essential for mammary gland development. *Cell* **103**, 41-50 (2000).
20. Brisken, C. *et al.* Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling. *Genes and Development* **14**, 650-654 (2000).
21. Beleut, M. *et al.* Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland. *Proc Natl Acad Sci U S A* **107**, 2989-2994 (2010).
22. Lloyd-Lewis, B. *et al.* Analysis of the Involuting Mouse Mammary Gland: An In Vivo Model for Cell Death. *Methods Mol Biol* **1501**, 165-186 (2017).
23. Lloyd-Lewis, B., Harris, O.B., Watson, C.J. & Davis, F.M. Mammary Stem Cells: Premise, Properties, and Perspectives. *Trends Cell Biol* **27**, 556-567 (2017).
24. Smith, G.H. & Medina, D. A morphologically distinct candidate for an epithelial stem cell in mouse mammary gland. *J Cell Sci* **90 (Pt 1)**, 173-183 (1988).
25. Kordon, E.C. & Smith, G.H. An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**, 1921-1930 (1998).
26. Smith, G.H. Experimental mammary epithelial morphogenesis in an in vivo model: evidence for distinct cellular progenitors of the ductal and lobular phenotype. *Breast Cancer Res Treat* **39**, 21-31 (1996).
27. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84-88 (2006).
28. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993-997 (2006).
29. Van Keymeulen, A. *et al.* Lineage-Restricted Mammary Stem Cells Sustain the Development, Homeostasis, and Regeneration of the Estrogen Receptor Positive Lineage. *Cell Rep* **20**, 1525-1532 (2017).
30. Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189-193 (2011).
31. Rios, A.C., Fu, N.Y., Lindeman, G.J. & Visvader, J.E. In situ identification of bipotent stem cells in the mammary gland. *Nature* **506**, 322-327 (2014).
32. Lloyd-Lewis, B., Davis, F.M., Harris, O.B., Hitchcock, J.R. & Watson, C.J. Neutral lineage tracing of proliferative embryonic and adult mammary stem/progenitor cells. *Development* **145** (2018).
33. van Amerongen, R., Bowman, A.N. & Nusse, R. Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland. *Cell Stem Cell* **11**, 387-400 (2012).
34. Bach, K. *et al.* Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. *Nat Commun* **8**, 2128 (2017).
35. Pal, B. *et al.* Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling. *Nat Commun* **8**, 1627 (2017).
36. Giraddi, R.R. *et al.* Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in Early Mammary Gland Development. *Cell Rep* **24**, 1653-1666 e1657 (2018).
37. Wuidart, A. *et al.* Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat Cell Biol* **20**, 666-676 (2018).

38. Lilja, A.M. *et al.* 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 (2018).
39. Ying, Z. & Beronja, S. Embryonic Barcoding of Equipotent Mammary Progenitors Functionally Identifies Breast Cancer Drivers. *Cell Stem Cell* **26**, 403-419 e404 (2020).
40. Fu, N.Y. *et al.* Identification of quiescent and spatially restricted mammary stem cells that are hormone responsive. *Nat Cell Biol* **19**, 164-176 (2017).
41. Huo, Y. & Macara, I.G. The Par3-like polarity protein Par3L is essential for mammary stem cell maintenance. *Nat Cell Biol* **16**, 529-537 (2014).
42. Bai, L. & Rohrschneider, L.R. s-SHIP promoter expression marks activated stem cells in developing mouse mammary tissue. *Genes Dev* **24**, 1882-1892 (2010).
43. Plaks, V. *et al.* Lgr5-expressing cells are sufficient and necessary for postnatal mammary gland organogenesis. *Cell Rep* **3**, 70-78 (2013).
44. Wang, D. *et al.* Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* **517**, 81-84 (2015).
45. Cai, S. *et al.* A Quiescent Bcl11b High Stem Cell Population Is Required for Maintenance of the Mammary Gland. *Cell Stem Cell* **20**, 247-260 e245 (2017).
46. Tabula Muris, C. *et al.* Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature* **562**, 367-372 (2018).
47. Vaillant, F. *et al.* The mammary progenitor marker CD61/beta3 integrin identifies cancer stem cells in mouse models of mammary tumorigenesis. *Cancer Res* **68**, 7711-7717 (2008).
48. Welm, B.E. *et al.* Sca-1(pos) cells in the mouse mammary gland represent an enriched progenitor cell population. *Dev Biol* **245**, 42-56 (2002).
49. Pfefferle, A.D. *et al.* The MMTV-Wnt1 murine model produces two phenotypically distinct subtypes of mammary tumors with unique therapeutic responses to an EGFR inhibitor. *Dis Model Mech* **12** (2019).
50. Pfefferle, A.D. *et al.* Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biol* **14**, R125 (2013).
51. Cleary, A.S., Leonard, T.L., Gestl, S.A. & Gunther, E.J. Tumour cell heterogeneity maintained by cooperating subclones in Wnt-driven mammary cancers. *Nature* **508**, 113-117 (2014).
52. Roarty, K. & Rosen, J.M. Wnt and mammary stem cells: hormones cannot fly wingless. *Curr Opin Pharmacol* **10**, 643-649 (2010).
53. Hatsell, S., Rowlands, T., Hiremath, M. & Cowin, P. Beta-catenin and Tcfs in mammary development and cancer. *J Mammary Gland Biol Neoplasia* **8**, 145-158 (2003).
54. Herschkowitz, J.I. *et al.* Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* **8**, R76 (2007).
55. Gyorffy, B., Surowiak, P., Budczies, J. & Lanczky, A. Online survival analysis software to assess the prognostic value of biomarkers using transcriptomic data in non-small-cell lung cancer. *PLoS One* **8**, e82241 (2013).

56. Madden, S.F. *et al.* BreastMark: an integrated approach to mining publicly available transcriptomic datasets relating to breast cancer outcome. *Breast Cancer Res* **15**, R52 (2013).
57. Spike, B.T. *et al.* A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell* **10**, 183-197 (2012).
58. Trejo, C.L., Luna, G., Dravis, C., Spike, B.T. & Wahl, G.M. Lgr5 is a marker for fetal mammary stem cells, but is not essential for stem cell activity or tumorigenesis. *NPJ Breast Cancer* **3**, 16 (2017).
59. de Visser, K.E. *et al.* Developmental stage-specific contribution of LGR5(+) cells to basal and luminal epithelial lineages in the postnatal mammary gland. *J Pathol* **228**, 300-309 (2012).
60. Rohrschneider, L.R., Custodio, J.M., Anderson, T.A., Miller, C.P. & Gu, H. The intron 5/6 promoter region of the ship1 gene regulates expression in stem/progenitor cells of the mouse embryo. *Dev Biol* **283**, 503-521 (2005).
61. Roarty, K. & Serra, R. Wnt5a is required for proper mammary gland development and TGF-beta-mediated inhibition of ductal growth. *Development* **134**, 3929-3939 (2007).
62. Zeng, Y.A. & Nusse, R. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* **6**, 568-577 (2010).
63. Vanhollebeke, B. *et al.* Tip cell-specific requirement for an atypical Gpr124- and Reck-dependent Wnt/beta-catenin pathway during brain angiogenesis. *Elife* **4** (2015).
64. Berndt, J.D. *et al.* Mindbomb 1, an E3 ubiquitin ligase, forms a complex with RYK to activate Wnt/beta-catenin signaling. *J Cell Biol* **194**, 737-750 (2011).
65. Valenzuela, D.M. *et al.* High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* **21**, 652-659 (2003).
66. Lloyd-Lewis, B. *et al.* Imaging the mammary gland and mammary tumours in 3D: optical tissue clearing and immunofluorescence methods. *Breast Cancer Res* **18**, 127 (2016).
67. Hatsell, S.J. & Cowin, P. Gli3-mediated repression of Hedgehog targets is required for normal mammary development. *Development* **133**, 3661-3670 (2006).
68. Chandramouli, A. *et al.* Ltbp1L is focally induced in embryonic mammary mesenchyme, demarcates the ductal luminal lineage and is upregulated during involution. *Breast Cancer Res* **15**, R111 (2013).
69. Imbert, A., Eelkema, R., Jordan, S., Feiner, H. & Cowin, P. ΔN89ß-catenin induces precocious development, differentiation, and neoplasia in mammary gland. *J. Cell Biology* **153**, 555-568 (2001).
70. Sorlie, T. *et al.* Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100**, 8418-8423 (2003).
71. Hu, Z. *et al.* The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7**, 96 (2006).

Figure 1

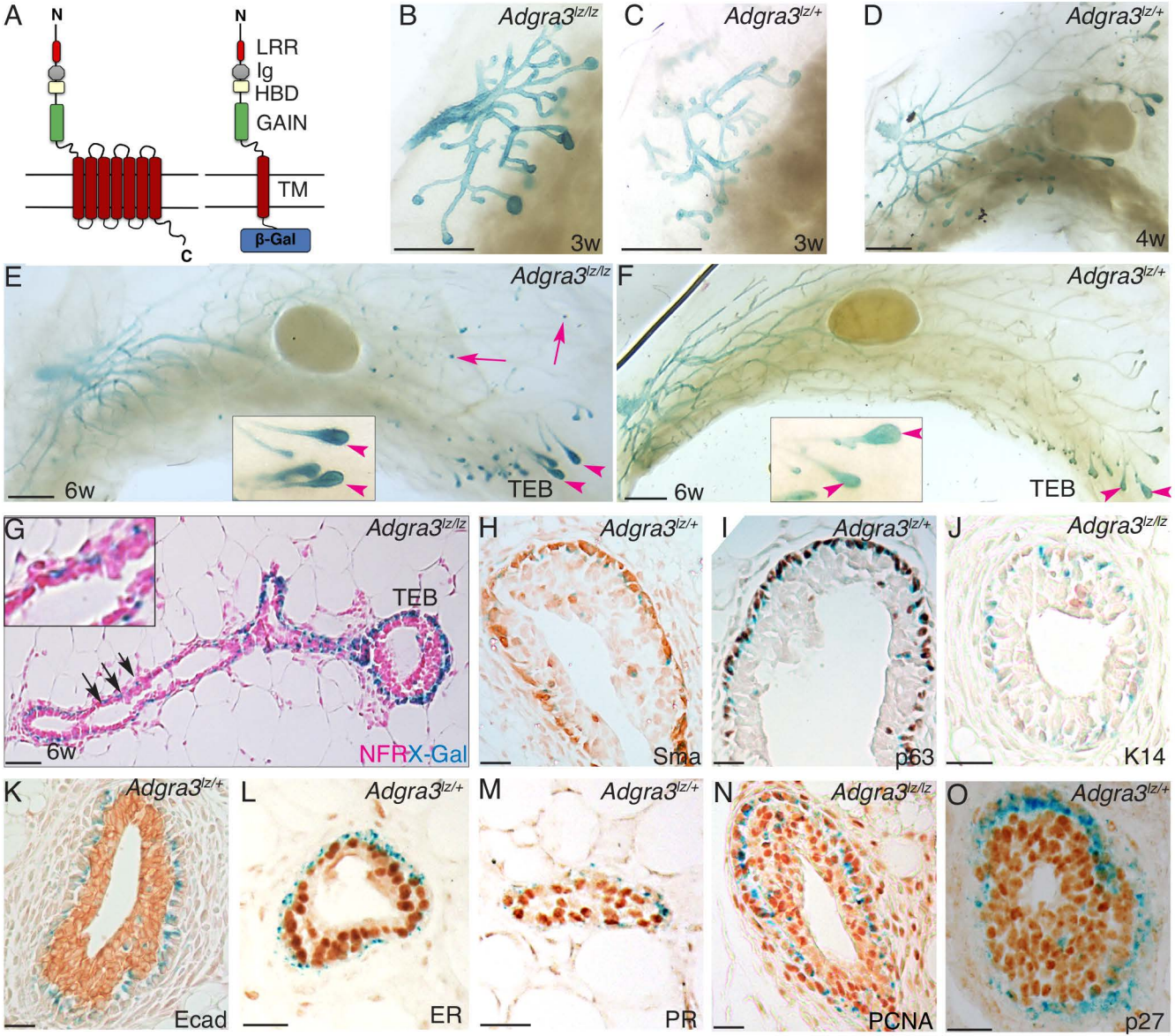

A
N
LRR
Ig
HBD
GAIN
TM
β-Gal
C
B
Adgra3lz/lz
3w
C
Adgra3lz/+
3w
D
Adgra3lz/+
4w
E
Adgra3lz/lz
6w
TEB
F
Adgra3lz/+
6w
TEB
G
Adgra3lz/lz
TEB
6w
NFRX-Gal
H
Adgra3lz/+
Sma
I
Adgra3lz/+
p63
J
Adgra3lz/lz
K14
K
Adgra3lz/+
Ecad
L
Adgra3lz/+
ER
M
Adgra3lz/+
PR
N
Adgra3lz/lz
PCNA
O
Adgra3lz/+
p27

Figure 2

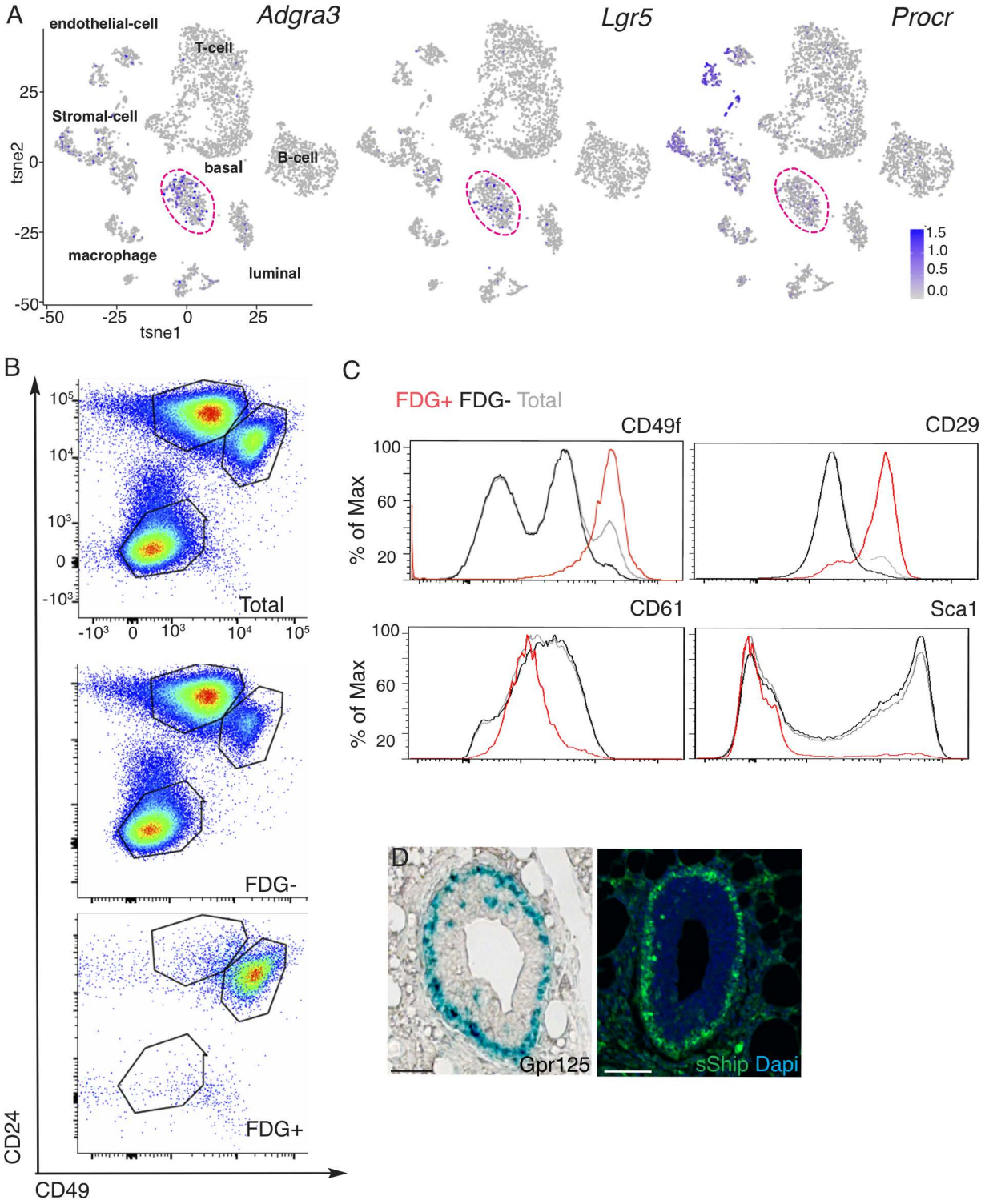

A
Adgra3
Lgr5
Procr
endothelial-cell
T-cell
Stromal-cell
B-cell
basal
macrophage
luminal
tsne2
tsne1
1.5
1.0
0.5
0.0
B
Total
FDG-
FDG+
CD24
CD49
C
FDG+ FDG- Total
CD49f
CD29
CD61
Sca1
% of Max
D
Gpr125
sShip Dapi

Figure 3

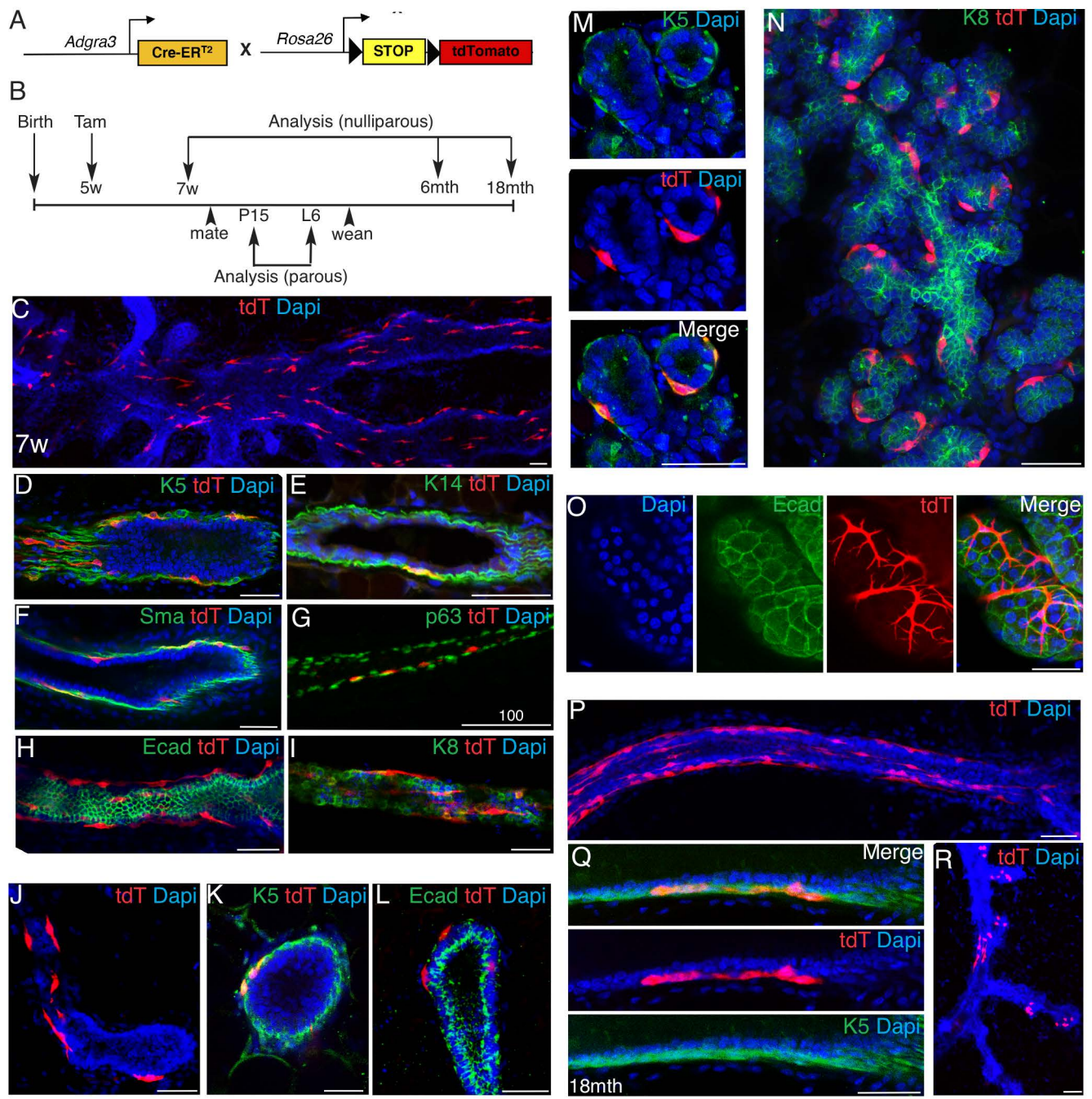

A
Adgra3
Cre-ER^{T2}
X
Rosa26
STOP
tdTomato
B
Birth
Tam
5w
Analysis (nulliparous)
7w
6mth
18mth
mate
P15
L6
wean
Analysis (parous)
C
tdT Dapi
7w
D
K5 tdT Dapi
E
K14 tdT Dapi
F
Sma tdT Dapi
G
p63 tdT Dapi
100
H
Ecad tdT Dapi
I
K8 tdT Dapi
J
tdT Dapi
K
K5 tdT Dapi
L
Ecad tdT Dapi
M
K5 Dapi
tdT Dapi
Merge
N
K8 tdT Dapi
O
Dapi
Ecad
tdT
Merge
P
tdT Dapi
Q
Merge
tdT Dapi
K5 Dapi
18mth
R
tdT Dapi

Figure 4

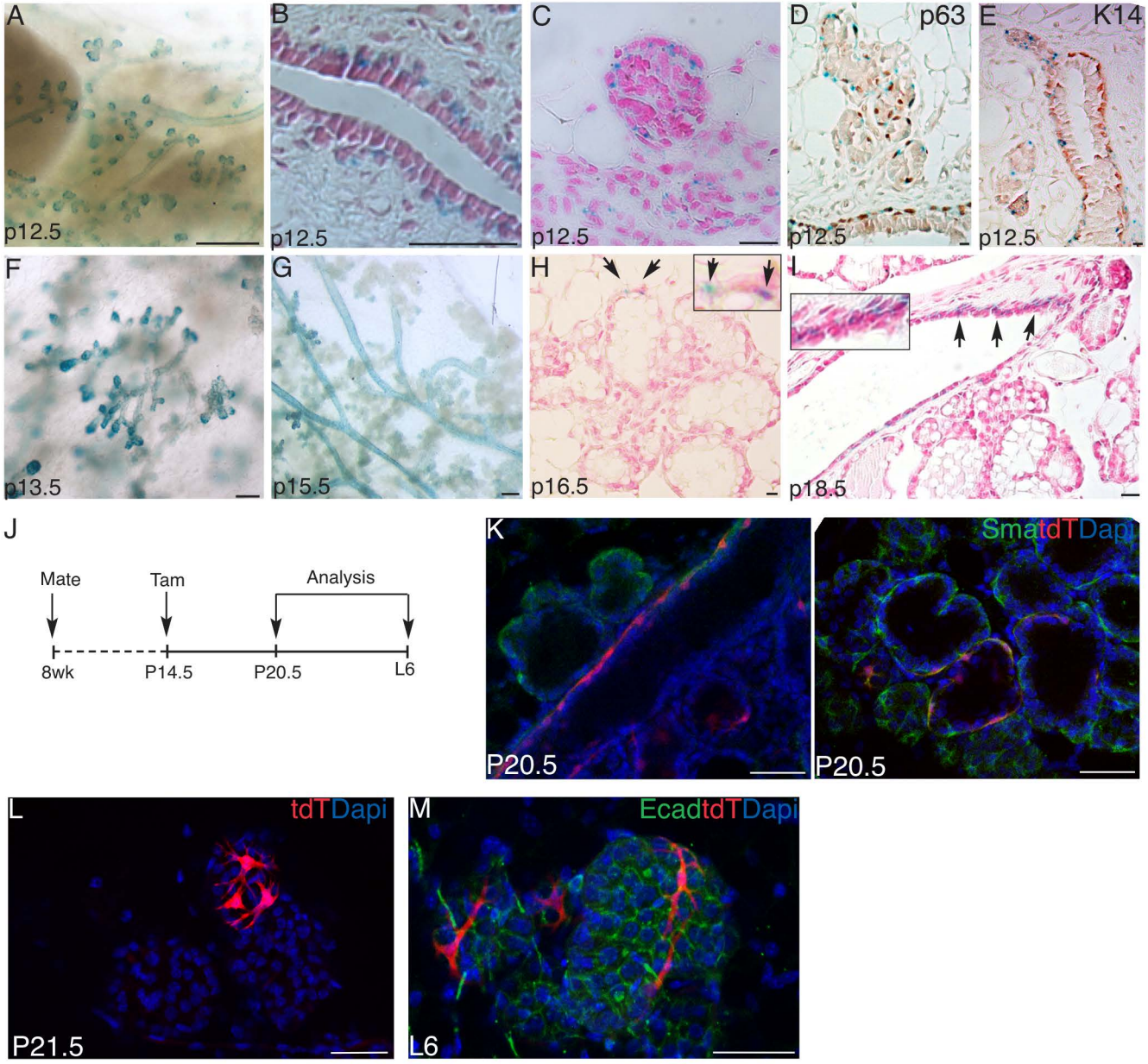

A
p12.5
B
p12.5
C
p12.5
D
p63
p12.5
E
K14
p12.5
F
p13.5
G
p15.5
H
p16.5
I
p18.5
J
Mate
Tam
Analysis
8wk
P14.5
P20.5
L6
K
P20.5
SmatdTDapi
P20.5
L
tdTDapi
P21.5
M
EcadtdTDapi
L6

Figure 5

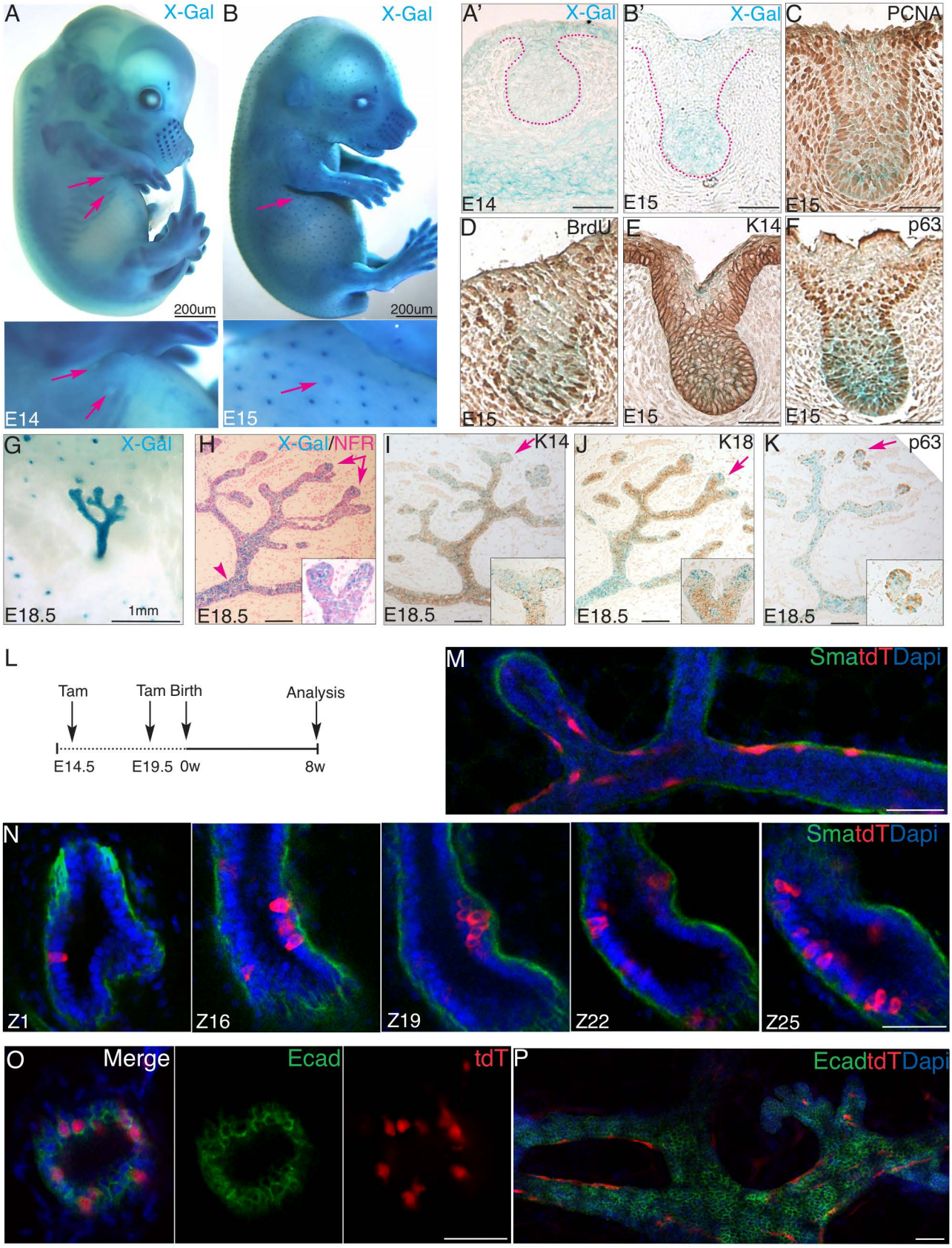

A
X-Gal
200um
E14
B
X-Gal
200um
E15
A'
X-Gal
E14
B'
X-Gal
E15
C
PCNA
E15
D
BrdU
E15
E
K14
E15
F
p63
E15
G
X-Gal
E18.5
1mm
H
X-Gal/NFR
E18.5
I
K14
E18.5
J
K18
E18.5
K
p63
E18.5
L
Tam
Tam
Birth
Analysis
E14.5
E19.5
0w
8w
M
SmatdTDapi
N
SmatdTDapi
Z1
Z16
Z19
Z22
Z25
O
Merge
Ecad
tdT
P
EcadtdTDapi

Figure 6

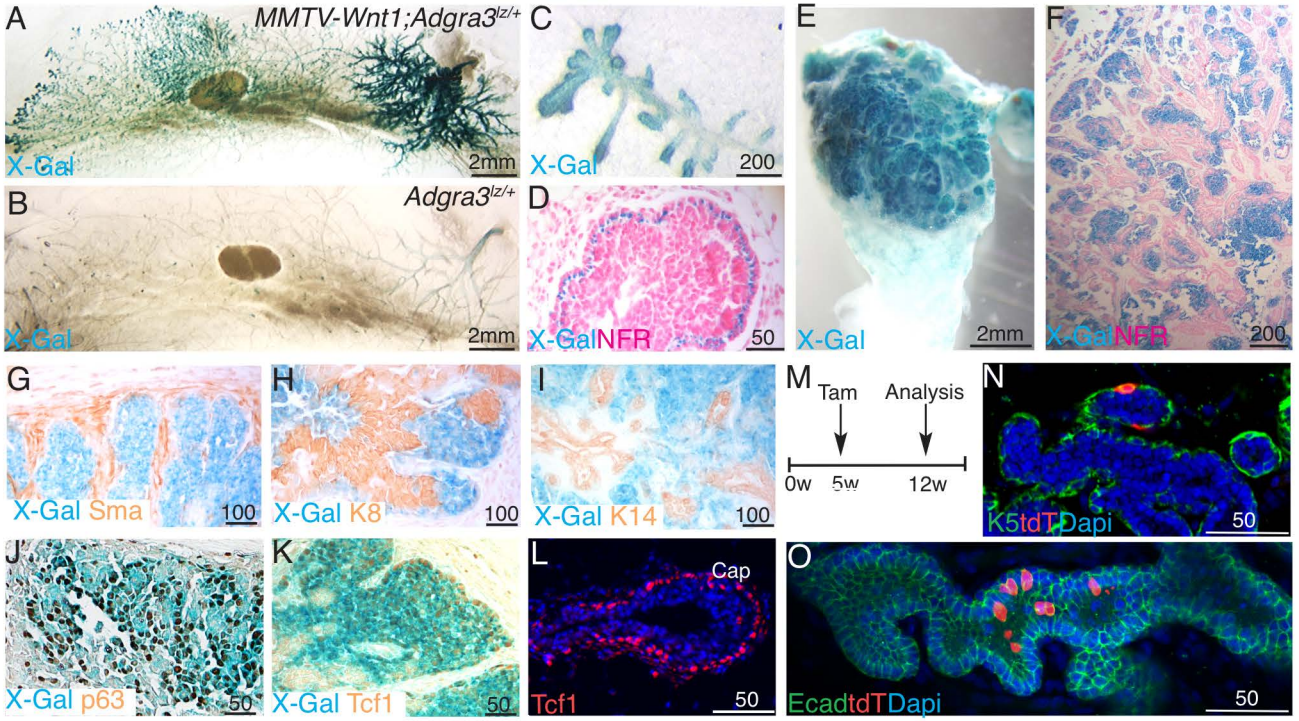
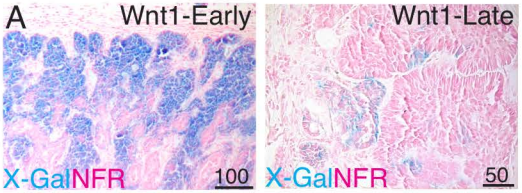
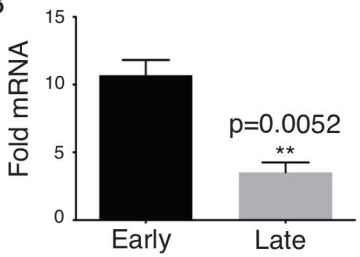
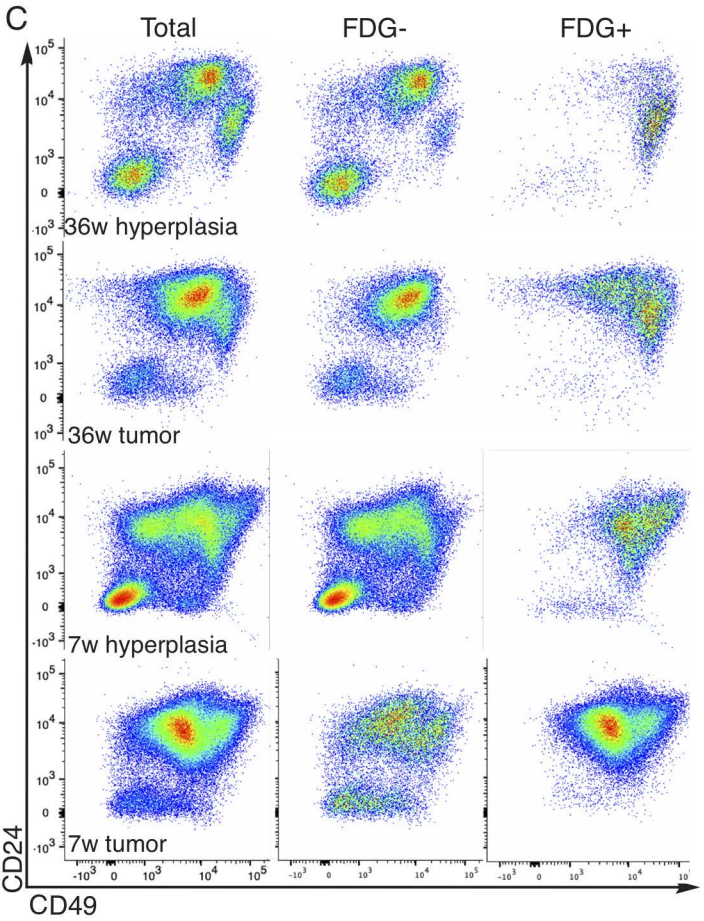
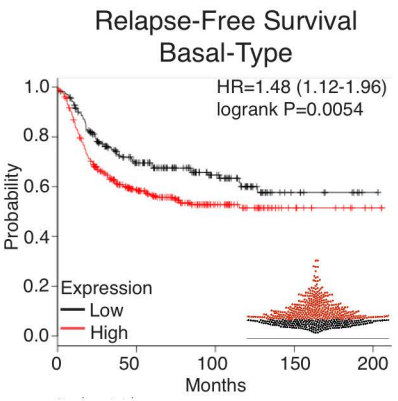

A
MMTV-Wnt1;Adgra3lz/+
X-Gal
2mm
B
Adgra3lz/+
X-Gal
2mm
C
X-Gal
200
D
X-GalNFR
50
E
X-Gal
2mm
F
X-GalNFR
200
G
X-Gal Sma
100
H
X-Gal K8
100
I
X-Gal K14
100
M
Tam
Analysis
0w
5w
12w
N
K5tdTDapi
50
J
X-Gal p63
50
K
X-Gal Tcf1
50
L
Cap
Tcf1
50
O
EcadtdTDapi
50

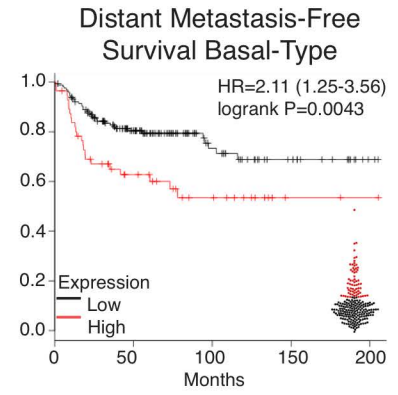
Figure7


A
Wnt1-Early
Wnt1-Late
X-GalNFR
100
X-GalNFR
50


B
Fold mRNA
15
10
5
0
p=0.0052
**
Early
Late


C
Total
FDG-
FDG+
36w hyperplasia
36w tumor
7w hyperplasia
7w tumor
CD24
CD49


D
Relapse-Free Survival
Basal-Type
HR=1.48 (1.12-1.96)
logrank P=0.0054
Probability
Expression
Low
High
Months


E
Distant Metastasis-Free
Survival Basal-Type
HR=2.11 (1.25-3.56)
logrank P=0.0043
Expression
Low
High
Months

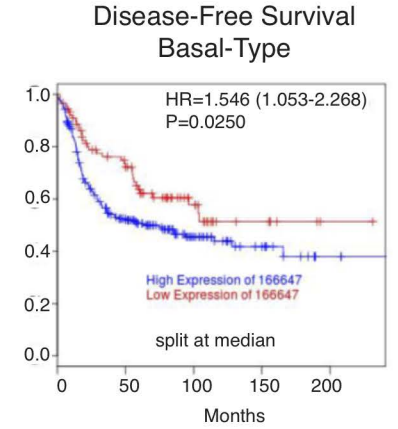
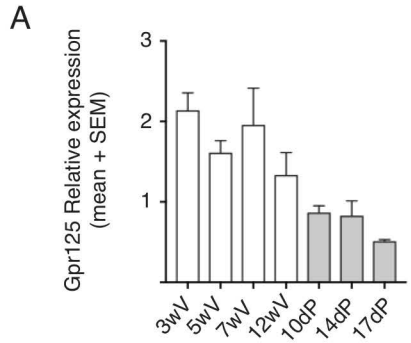
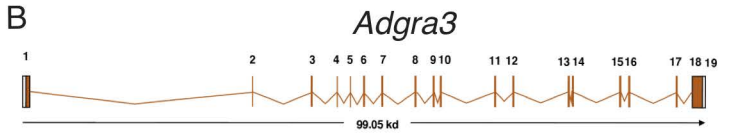
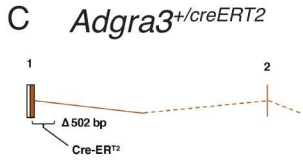
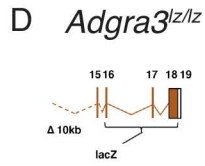

F
Disease-Free Survival
Basal-Type
HR=1.546 (1.053-2.268)
P=0.0250
High Expression of 166647
Low Expression of 166647
split at median
Months

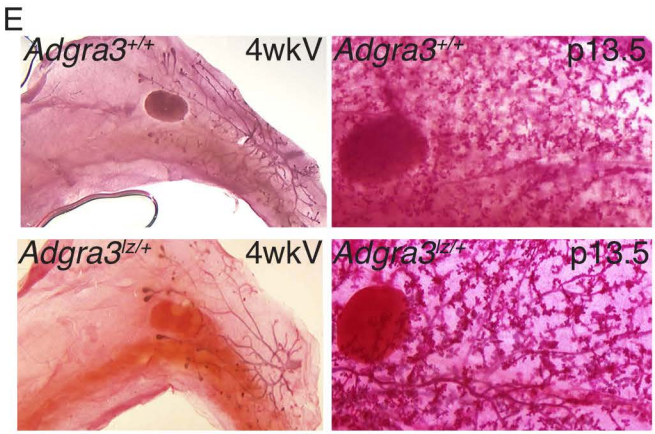
Figure S1

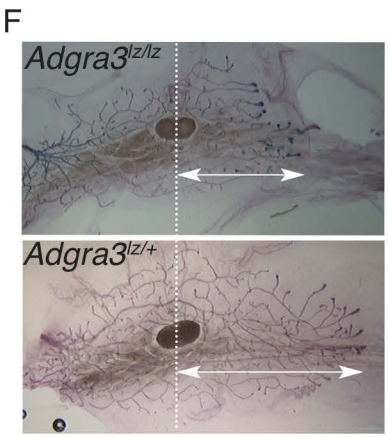

A
Gpr125 Relative expression
(mean + SEM)
3
2
1
3wV
5wV
7wV
12wV
10dP
14dP
17dP


B
Adgra3
1
2
3
4 5 6
7
8
9 10
11 12
13 14
15 16
17
18 19
99.05 kd


C
Adgra3+/creERT2
1
2
Δ 502 bp
Cre-ERT2


D
Adgra3lz/lz
15 16
17
18 19
Δ 10kb
lacZ


E
Adgra3+/+
4wkV
Adgra3+/+
p13.5
Adgra3lz/+
4wkV
Adgra3lz/+
p13.5


F
Adgra3lz/lz
Adgra3lz/+

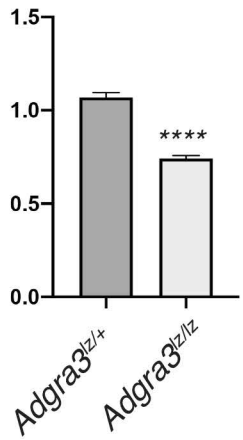

G
1.5
1.0
0.5
0.0
****
Adgra3lz/+
Adgra3lz/lz

Figure S2

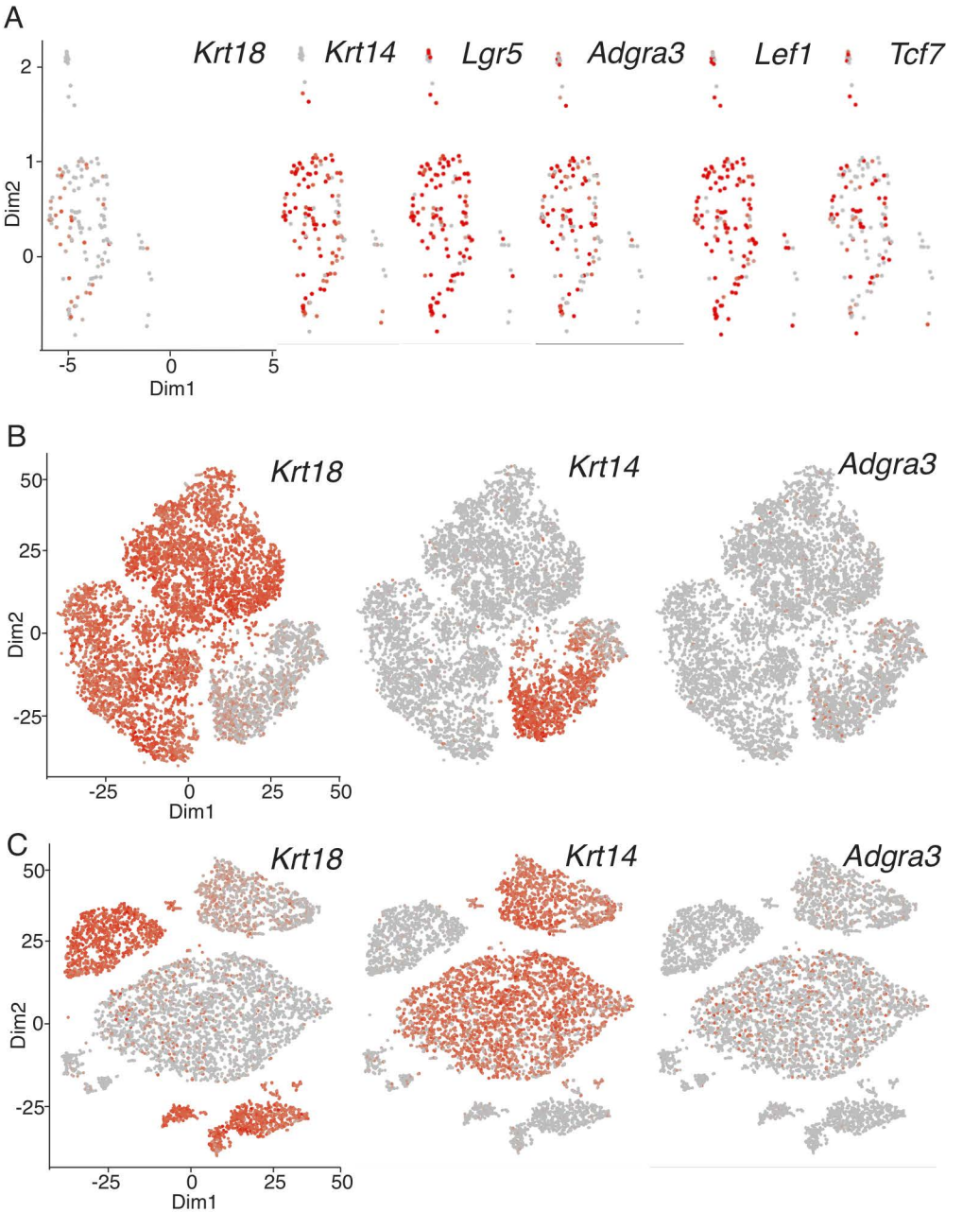

A
Krt18
Krt14
Lgr5
Adgra3
Lef1
Tcf7
Dim2
Dim1
B
Krt18
Krt14
Adgra3
Dim2
Dim1
C
Krt18
Krt14
Adgra3
Dim2
Dim1

Figure S4

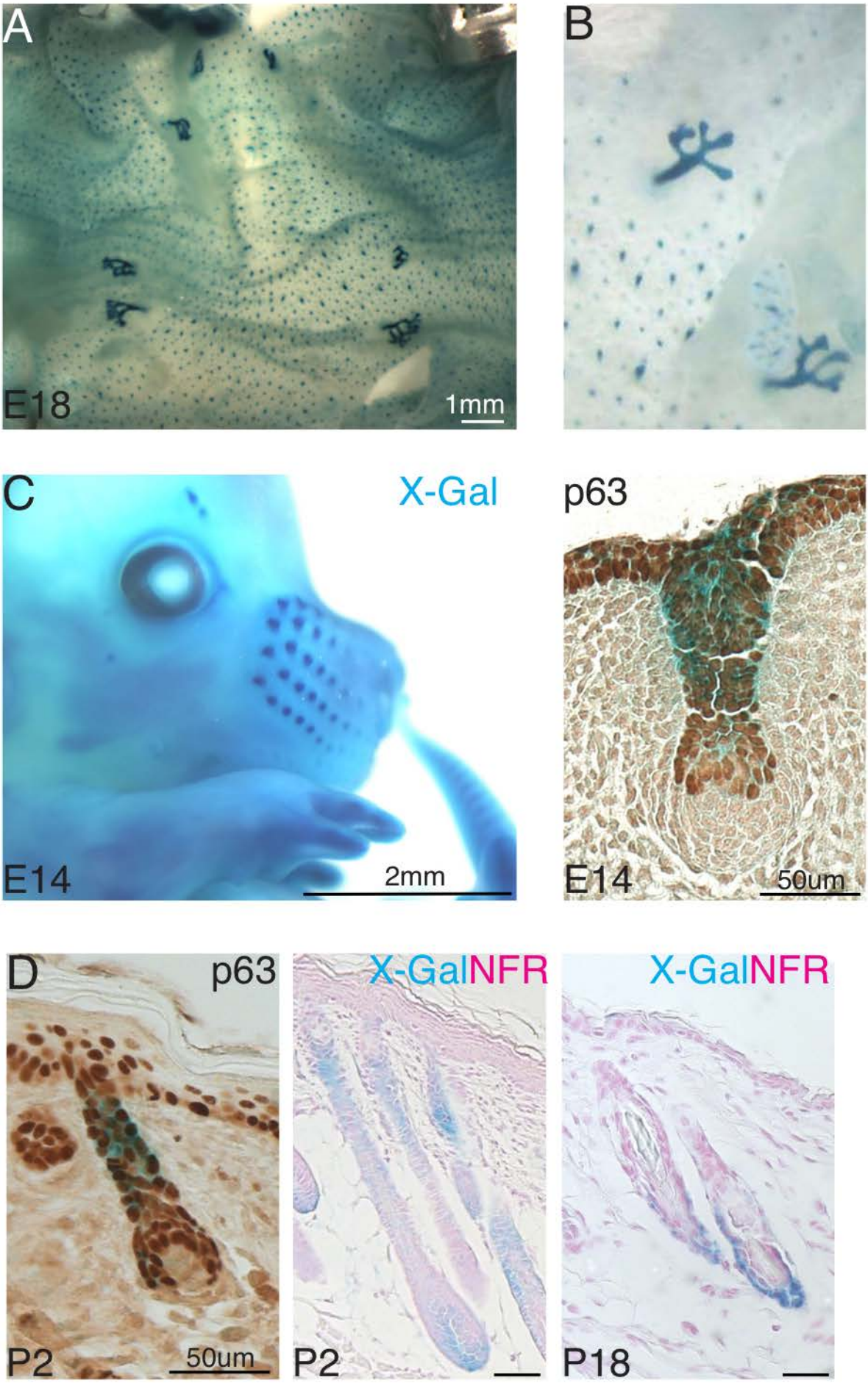

A
E18
1mm
B
C
X-Gal
E14
2mm
p63
E14
50um
D
p63
P2
50um
X-GalNFR
P2
X-GalNFR
P18

Table S5

A

Mouse Tumor Model	Fold change
Wnt1-Early^{Ex}	3.17
p53null-Basal^{Ex}	1.76
Wnt1-Late^{Ex}	1.7
Class14^{Ex}	1.31
Class8^{Ex}	1.2
p53null-Luminal^{Ex}	0.95
Squamous-like^{Ex}	0.9
Neu^{Ex}	0.9
Erbb2-like^{Ex}	0.9
PyMT^{Ex}	0.8
Class3^{Ex}	0.78
Stat1^{Ex}	0.77
C3Tag^{Ex}	0.76
Myc^{Ex}	0.75
Normal^{Ex}	0.7
ClaudinLow^{Ex}	0.68
Wapint3^{Ex}	0.6

B

Human Breast Cancer subtype	Fold change
Human Basal-like	2.03
Human Normal-like	1.33
Human LumB	0.78
Human LumA	0.77
Human Her2-Enriched	0.75
Human Claudin-low	0.8

Embryonic Mammary Development

Departments of Cell Biology[1] and Dermatology[2]

New York University School of Medicine

Elena Spina[1], Pamela Cowin[1,2]*

*Address for correspondence:

Dr. Pamela Cowin

New York University Medical Center

MSB 621, 550 1st Ave

New York, NY 10016

Conflict of interest statement: The authors declare that no conflict of interest exists.

Keywords

Embryonic mammary development; Breast cancer; Wnt; Hedgehog; Ltbp1

Abstract

Embryonic mammary development provides exquisite examples of epithelial-mesenchymal reciprocal signaling during the formation of the ectodermal placode, invagination of a flask-shaped bud and development of a miniature bi-layered ductal tree. Here we provide an overview of the highly conserved pathways that shape the embryonic mammary gland. Currently there is a good understanding of the major factors that contribute to specification, commitment, movement and potency. This knowledge has illuminated not only embryonic mammary development but also highlighted parallels processes in breast cancer. Better understanding of the balance between epithelial-mesenchymal signaling during normal mammary development holds considerable promise to advance recent attempts to eliminate cancer by restoring differentiative signals.

Introduction

Evolutionary importance and variety of mammary gland development

Mammary glands are the defining hallmark of mammals. Their emergence provided the significant evolutionary advantage of enabling mothers to remain mobile while providing nutrition to their offspring. The importance of this advance led Darwin to dedicate an entire chapter to their description.

Mammary glands are specializations of the embryonic ectoderm and comparative anatomy of living species, suggests they evolved from ancestral hair-associated apocrine glands (for primary refs see [1, 2]). For example, the duck-billed platypus secretes milk from an abdominal patch comprising 100-200 compound mammo-lobular-pilo-sebaceous units and is wicked towards the ends of specialized hairs. This secretion served as an adhesive to bind the eggs to the mother and to nourish the hatchlings. Koala bears form vestigial mammary hairs that regress as their nipples develop. Similar structures are formed and retained in horses. Squirrels develop bilateral sensory vibrissae and nipples from the same epidermal anlage. Molecular evidence suggests that lactation arose from cutaneous secretions that prevented desiccation of parchment eggs which evolved to enable rapid growth of offspring by providing immediate inter-generational transfer of ~10% of maternal skeletal calcium [1, 2]. With this fact in mind, it is not surprising that many proteins that regulate mammary development serve dual roles in bone metabolism.

In this review we describe the basic stages of embryonic mammary development, highlight the key regulatory factors, and discuss how our understanding of this process has contributed to biomolecular medicine.

1. Position and Number of Mammary Glands

The first decisive step in mammary gland development pertains to their position and different species have solved this problem in a great variety of intriguing ways [1, 2]. Manatees form axilliary mammary glands, elephants and primates have a single pectoral pair, ungulates develop inguinal glands that are sometimes amalgamated to form an udder. Other species, for example pigs and the multi-mammate rat display bilateral rows each with a dozen glands. Marsupials have central abdominal glands that open into their pouch, and in the case of the kangaroo,

secrete distinct types of milk from two types of nipple to feed offspring of different developmental stages. Whales and seals have retractable mammary glands to streamline their swimming, bats and possums make additional use their nipples as attachments sites to facilitate carrying their young.

Setting aside such comparative studies and attempts to maximize milk production and animal husbandry in farm animals, the majority of our knowledge on the embryonic stage of mammary development comes mostly from studies on mice and rabbits. Genes associated with mammary development are highly conserved. Thus, even though information on human embryonic mammary development is limited, a number of genes responsible for human developmental syndromes have been identified and studied further through the generation of genetically engineered mice with similar phenotypes (**Table 1**). It has also become apparent that genes regulating murine embryonic mammary morphogenic processes involving directed collective cell migration, invasion, proliferation, cell potency and differentiation are frequently reactivated in human breast cancer [3-9].

2. The basic developmental process

Murine mammary development begins mid-gestation ~ embryonic day 10.5 (E10.5) with the definition of bilateral milk lines. ~E11-E12 ectodermal cells coalesce into five pairs of mammary placodes. By E13.5 these cell movements have formed raised hillocks that quickly sink below the dermis to form flask-shaped buds surrounded by a compacted highly specialized mammary mesenchyme ~E14. ~E15.5 the mammary primordium is destroyed in males but in females begins to proliferate and sprout towards a nearby fat pad. After invading the fat pad, it branches and forms a lumen, producing a hollow ductal tree by birth (**Fig. 1**). Mammary development, like all ectodermal appendages, is regulated by highly conserved local developmental patterning pathways engaged in epithelial-mesenchymal cross-talk. We will now briefly discuss the role of these regulatory paths at different stages of mammary development.

2.1 BMP4:TBX3 antagonism defines the mammary line

Murine mammary glands form in arcs on either side of the body that roughly correspond to the dorsolateral-ventral (D/V) boundary of the underlying mesenchyme [10]. It has been proposed that the mammary line is defined by mutual antagonism between ventral expression of the bone morphogenetic protein (Bmp4) and dorsal expression of the transcription factor, Tbx3 [10] (**Table 2, Fig. 2a**). Morphologically, this line is visible in rabbit embryos as a prominent ectodermal ridge [11]. Counterparts have been described in many species, however, the corresponding structure in mice is more nuanced and discernable only in histological sections, as a bilayered row of columnar cells within the presumptive region of the third mammary rudiment (MR)[12]. Molecularly, the field in which the mammary line will develop can be detected ~E10.5 by a band of mesenchymal Tbx3 mRNA expression. At ~E11.5 a line of Wnt10b expression appear in the ectoderm between the fore and hind limbs, accompanied by two other streaks around the limbs where axillary and inguinal glands will later form [13].

.

2.2 The essential role of early Wnt signaling along mammary line and limb streaks

Intriguingly, expression of TOP-gal, a transgenic reporter of canonical Wnt (Tcf/Lef) signaling precedes that of Wnt10b [14]. Its appearance one day earlier at E10.5 suggests the involvement of additional Wnts. Wnt 3 and 6, are present in the ectoderm and Wnt5a and 11 in the mesenchyme at this time, and may contribute to boundary formation by antagonistic canonical and non-canonical Wnt signaling. An absolute requirement for Wnt function in forming the mammary line is demonstrated by the finding that experimental expression of DKK, a secreted Wnt inhibitor, abolishes the earliest signs (expression TOP-gal, Wnt10b and Tbx3). Deletion of Lef1, a downstream transcriptional mediator of Wnt signaling, also impairs all other ectodermal appendages [15]. However, the effects of its loss on mammary rudiments are more attenuated that those produced by DKK, suggesting partial redundancy with other members of the Tcf/Lef family. Wnt signaling is critical at all subsequent stages of mammary formation [14](**Table 2, Fig. 3**).

2.3 Migration into Placodes

The mammary line and limb streaks are transient entities and in less than a day Tbx3, TOP-gal and Wnt10b expression becomes confined to elliptical placodal thickenings of the ectoderm [13, 14, 16, 17]. Placodes arise by cell rearrangement rather than proliferation and form in a rather surprising temporal order [11]. Placode MR#3 forms first, followed swiftly by MR#4 through coalescence of cells at either end of the central mammary line. MR#1 emerges (sometimes preceding MR#4) together with MR#5 by movement of cells along the streaks encircling the fore and hind limbs respectively. The last placode to develop, MR#2, is thought to receive cells from the forelimb streak as well as the mammary line. The suggestion that cells actively migrate along the central mammary line arose from early studies on rabbit embryos that tagged cells with carbon and tracked their location over time as well as from scanning electron micrographs that showed cell processes polarized towards the placodes[18] [11]. It has been proposed that Wnt expression may stimulate this migration in a manner similar to its actions in propelling cells out of intestinal crypts. The significance of the mammary line as a migratory path, however, has been challenged [12]. Placode MR#3 has been proposed to result instead from centripetal aggregation of ectodermal cells towards an Fgf10 attractant, emanating from the tips of underlying somites [19, 20] (see below). In this model cells are proposed to sustain Wnt signaling only as they enter the placode and Wnt signaling is extinguished elsewhere. Nrg3 has also emerged as a candidate that promotes migration of mammary progenitors into MR#3 (see below) [21, 22] . In this ongoing debate, the movement of cells around the limbs has been relatively neglected. At these sites, the earliest marker appearing ~E11 within the mesenchyme is latent TGFβ-binding protein 1 (Ltbp1) [23](**Fig. 4**). Ltbp1 functions to tether and position latent forms of TGFβ, along fibronectin and elastin fibers. Importantly, it plays an essential role in integrin-mediated stretch activation of this promigratory cytokine [24]. Thus, Ltbp1 is ideally placed in time and space to focally stimulate and guide movement of overlying ectodermal cells into MR# 1, 2, and 5.

3. Site-specific factors and regulatory genes

Analyses of mutant and genetically engineered mice has revealed the surprising fact that each placode pair is independently governed by its own unique complement of regulatory genes

(**Table 2**). Placode MR#3 has most intensively studied whereas molecular mechanisms guiding formation of axillary and inguinal MR are more obscure. Attempts have been made to establish a genetic hierarchy by analyzing of the effects of loss of one gene on the expression of others. A working model of their potential epistatic actions is presented in Fig. 2. This model proposes genes act collectively, but with site specific requirement and redundancy, either to direct cell migration and/or to augment and/or sustain canonical Wnt signaling, which must reach a critical threshold within a narrow temporal window for placodes to form and be maintained. Below is a brief summary on studies on each of these regulatory genes.

3.1 Gli3 repression of Hedgehog signaling is required for mammary fate of MR#2-5

Gli3 is currently considered to be the most upstream regulator of MR#2-5. Gli3 is a transcription factor in the Hedgehog pathway that can have activator or repressor functions depending on the cell context. Genetic analyses have shown hedgehog signaling must be repressed for most mammary rudiments to form properly [25]. This is in stark contrast to other ectodermal appendages, such as hair and teeth, which require positive Hedgehog signaling. Mice lacking Gli3 repressor function (*Gli3*$^{xt/xt}$) show abnormalities in early patterning of the mammary field, displaying inappropriate dorsal expansion of Bmp4, constriction of the band of Tbx3 and failure of Tbx3 to concentrate into the presumptive MR#3 placodal region [26]. *Gli3*$^{xt/xt}$ mice and others in which Hedgehog signaling is misactivated (*Gli2*$^{1ki/1ki}$) show identical phenotypes: failure to form MR#3 and #MR5 and severe impairment of MR#2, MR#4 [25, 26] (**Table 2**). Molecular analysis of residual MR#2 and MR#4 revealed misactivation of Hedgehog signaling, loss of epithelial Wnt signaling and failure to maintain mammary mesenchyme specification, which in turn compromised bud invagination and lead to loss of sexual dimorphism [25, 26] . Intriguingly, when the balance of Gli3A to Gli3R was tipped towards pathway activation then hair follicles, which are normally suppressed in the presumptive nipple zone around MR, developed within the mammary field (**Fig. 5**) [26]. Consistent with these findings the converse phenotype was produced by abrogating Hedgehog signaling through elimination of Smoothened, the Hedgehog signal transducer. These mice displayed mammary features in appendages that would otherwise become hair follicles [27]. Collectively, these studies support the concept that Gli3R specifies

mammary fate by blocking Hedgehog-mediated hair follicle differentiation [27]. The targets of Gli3 repression remain to be determined. However, the finding that loss of Gli3R alters the patterning of Bmp4 and Tbx3, coupled with the presence of Gli3 binding sites are present in the murine Bmp4 promoter suggests that Bmp4 could be a target[28]. In this scenario Gli3R would function to derepress Tbx3 by sculpting the zone of Bmp4 expression [26].

3.2 Tbx3 is required for placodal induction and maintenance of MR#1,3,4,5

As discussed above, mesenchymal expression of another transcriptional repressor Tbx3, precedes mammary line formation. However, Tbx3 is also one of the earliest markers expressed in the ectodermal placode appearing ~E10.5 [16, 17](**Fig. 2b**). In Tbx3 null mice, Wnt10b and Lef1 fail to accumulate, FGF signaling is abrogated, and placodes 1,3,4,5 do not form. Thus, Tbx3 acts downstream of Gli3R and upstream of Wnt and FGF signaling. However, it is, in turn, augmented by these pathways in a positive feedback loop. Tbx3 levels are critical: as demonstrated by the finding that Tbx3 haploinsufficiency also leads to frequent aplasia of MR#1-3 ~E13.5 and reduced ductal branching in retained glands [29]. Thus, Tbx3 is required not only for placodal induction but its level is critical for placodal maintenance. In other tissues, Tbx3 has been implicated in regulating cell proliferation, however as proliferation is not a prominent driver of placodal formation it is more likely in this situation that Tbx3's ability to repress E-cadherin and promote migration are more relevant[29]. Tbx3 is highly relevant for human mammary development. C-terminal mutants, which show impaired ability to repress their targets, result in Ulna-mammary -syndrome (UMS)[30] (**Table 1**), which involves defects in the ulna, hypoplastic or aplastic breasts in both genders, supernumerary, inverted or absent nipples and defective lactation.

3.3 Fgf10/Fgfr2b vertically augment placodal Wnt10b

Fibroblast growth factor 10 (Fgf10) is essential for MR#1,2,3,5 and its receptor, Fgfr2b, is required to maintain all MR [19]. Fgf10 expression in the dermomyotome of the somites ~ E10.5 onwards has been proposed to activate its receptor, Fgfr2b, in the ectoderm [20]. Classical transplantation studies have shown that mesenchyme alone can induce mammary fate within ectoderm independent of instruction from underlying tissue and the insolubility of Fgf10 make it unlikely

to diffuse over such a distance. Delamination and migration of cells expressing Fgf10 has been proposed as a solution [20]. Somitic Fgf10 expression is unaffected by Wnt inhibitors, whereas Fgf10 hypomorphs fail to express TOP-gal and Wnt10b, placing Fgf10 signals upstream of Wnt10b and mammary line stratification. Of note, TOP-gal and Wnt10b expression aligns with the segmentation pattern of the thoracic somites and Pax3 mutants, which fail to extend the thoracic somites show dorsalization of the mammary line[20]. Moreover, humans with Poland syndrome, which is characterized by hypoplasia of structures derived from thoracic somites also affect breasts [31](**Table 1**). Homeobox genes have recently been proposed to act upstream of Fgf10. Hoxc8, for example, although normally expressed in the E10.5 ectoderm, when mis-expressed induces somatic Fgf10 and ectopic placodes[32].

3.4 Ngr3 stimulates aggregation of mammary progenitors into MR#3

Neuregulin 3 (Nrg3) presents another example of growth factor required for embryonic mammary development[22]. Nrg3 is a ligand of receptor tyrosine-protein-kinase ErbB4, implicated in chemoattraction, directed cell migration, adhesion, proliferation, and stratification [22]. Nrg3 appears ~E11 within the dermal mesenchyme underneath MR#3 but by ~E12 is restricted to the epithelial placode, together with its receptor [33]. Its requirement for embryonic mammary development became evident from studies of *Nrg3^{ska}* (scaramanga) mutants which express reduced levels of Nrg3. *Nrg3^{ska}* mice lack MR#3[34] (**Table 2**). By contrast, mice overexpressing K14-Nrg3 or those exposed to beads coated in recombinant Nrg3 produce a supernumerary MR between MR3 and 4. Collectively these experiments show Nrg3 promotes mammary cell fate. Intriguingly further ectopic glands appear in adult *K14-Nrg3* mice, suggesting that Nrg3 may continue to prime or recruit cells for additional de novo appendage development. *Nrg3^{ska}* mice show reduced Lef1 expression whereas *K14-Nrg3* mice induce ectopic Lef1 expression. However, this is due to Nrg3 stimulating aggregation of cells engaged in Wnt signaling into placodes rather than by inducing Lef1. Nrg3 mutants show changes in Tenacin C, a matrix protein associated with stem cell niches and a ligand for integrins found on mammary progenitor cells. Howard and colleagues elegantly visualized the effect of Nrg3 on progenitor recruitment into placodes by crossing *Nrg3^{ska}* mice to mice expressing s-SHIP-GFP, a marker of mammary progenitors [21]. s-SHIP-GFP positive progenitors remained dorsally dispersed in

Nrg3^{ska} mutants and the few cells that that were present in Nrg3^{ska} hypoplastic placodes failed to adopt placodal cell arrangement and shape characteristics. These data lead them to conclude that Nrg3 regulates progenitor aggregation into, and rearrangement within, placodes (**Fig 2**).

3.5 Eda/Edar patterns placodal and inter-placodal domains by sculpting Wnt signaling

A further route that serves to regulate placodal definition and spacing is the ectodysplasin (Eda) signaling pathway. Eda, a tumor necrosis factor-like ligand, binds to its receptor, Edar, and signals through an adaptor protein, Edaradd, to activate NFkB [35, 36]. This pathway is an important regulator of mammary gland development in humans where mutations in *EDA*, *EDAR* or *EDARADD,* as well as in *IKK, the* inhibitor of NFkB, cause hypohidrotic ectodermal dysplasia (HED)[37-41] (**Table 1**). This condition involves defects in hair, teeth, sweat, salivary and mammary glands. The mammary phenotypes include loss of nipples and gain (polythelia), compromised ability to breast feed and occasional loss of breasts altogether. Although Eda is dispensable for endogenous MR formation, K14-Eda expression, induces 2-3 extra MR along the central mammary line and others in the neck region [42-45]. This suggests that Eda acts downstream of mammary line specification to regulate mammary placodal fate along this extended arc. Microarray analysis revealed that both activators (Wnt10a,10b) and inhibitors (DKK, Sostdc1) of Wnt signaling are transcriptional targets of Eda signaling and suppression of canonical Wnt signaling was found to reduce supernumerary placode induction by K14-Eda in a dose-dependent manner [42, 43, 46, 47]. Eda is proposed to pattern placodes by stimulating expression of a short-range insoluble activator (eg Wnt ligands) and defines interplacodal regions, by simultaneously stimulating more diffusible Wnt inhibitors (eg DKK) (**Fig. 2c**) [42]. A similar reaction-diffusion mechanism has been proposed to pattern intestinal crypts where expression of short-range Wnt/noggin creates the stem cell niche at the crypt base and long-range diffusion of HH/Bmp stimulates BMPR1 expressed in the villi to inhibit crypt expansion [48, 49]. This model provides an elegant explanation for how varying Eda expression can produce both loss and gain of MR. Changes in the Eda/Edar signaling have been linked to adaptive changes in other skin appendages for example sweat gland density in modern human populations. It is hypothesized that Eda/Edar may in addition to causing HED also be responsible for the prevalence of polythelia in humans [50]. In addition to its manifest importance for human

appendage development is has been suggested that Eda patterning could be the molecular mechanism responsible for evolutionary diversity in the number and position of mammary glands seen in different species [51]. In addition to modulating Wnt signaling, Eda has been shown to modulate expression of several other key factors including Fgf20 and PthrP and thus may have a wide repertoire of ways to modulate mammary development[43] [52].

3.6 PTHrP: the master regulator of mammary mesenchyme differentiation

Once placodes have been defined the next step involves their rearrangement into hillocks ~E12 and invagination to form a flask shaped bud (**Fig. 6**). The mature bud comprises a sphere of concentrically oriented epithelial cells connected to the skin surface by a stalk surrounded by 2-3 layers of condensed, elongated mesenchymal cells. In males, androgens cause the mesenchyme to enlarge, constricting the stalk and leading to severance and apoptosis [53]. In females, buds are quiescent at E14, become specified ~E15 then begin proliferation and sprout towards the fat pad ~E16. Parathyroid hormone-related protein (PTHrP) was discovered as a cause of humoral hypercalcemia of malignancy in cancer patients. This likely arises due to misactivation of its physiological role in mobilizing skeletal calcium stores for milk production during lactation. During embryogenesis, however, PTHrP functions locally, and is considered to be the master epithelial inducer of mesenchymal differentiation. As such, it plays a critical role in mammary fate commitment. PTHrP is expressed within the mammary line and placodes but is not required until ~E15. Eda-Edar-NFκB signaling upregulates PTHrP mRNA within embryonic mammary buds [43]. PTHrP protein it is secreted by the epithelial bud and acts in a paracrine fashion to activate the G-protein-coupled receptor, PTHR1, within the surrounding mesenchyme [54, 55]. Disruption of either gene impairs mammary development at in four ways. Buds form, but the mesenchymal cells that condense around them lack all hallmarks of mammary mesenchyme specification (Androgen Receptor (AR), Estrogen Receptor (ER), Lef1, cadherin 11 and tenascin C) [53, 54, 56]. The absence of AR prevents the mesenchyme from responding to androgens and so male buds are retained [53]. The undifferentiated immature mesenchyme fails to send reciprocal differentiative signals to the bud. As a result, female buds lose their mammary identity, fail to sprout, and the presumptive nipple reverts to an epidermal fate [56, 57]. A converse phenotype

occurs in response to PTHrP overexpression: ectopic mammary mesenchymal markers become expressed resulting in suppression of hair follicles and conversion of ventral epidermis to nipple-sheath. The importance of this pathway for human mammary development is demonstrated by human fetuses with Blomstrand chondrodysplasia (**Table 1**) that have homozygous null mutations in the PTHR1 gene and lack breast duct development similar to *PTHrP*$^{-/-}$ mice[58].

PTHrP mediated epithelial-mesenchymal cross talk regulates the expression of many different molecules. Among these, BmpR1, Lef1, β-catenin and Msx2 play have received the most attention[59, 60]. β-catenin plays an essential role downstream of PTHrP signaling in specifying the mammary mesenchyme [61]. Loss of PTHrP inhibits mesenchymal Lef1 and β-catenin expression whereas overexpression of PTHrP upregulates them ectopically in the ventral mesenchyme [53, 56]. Deletion of mesenchymal β-catenin phenocopies loss of PTHrP and blocks the ectopic induction of mammary mesenchymal markers by K14-PTHrP[61]. Further experiments demonstrated that PTHrP requires Wnt/β-catenin within the mesenchyme to induce mammary mesenchyme specification. While the mechanism by which PTHrP activates Wnt signaling remain obscure it has been proposed to likely result from induction of Wnt11 and Rspo1 [59, 61]. Intriguingly, very similar phenotypes involving male bud retention, failure to maintain mammary mesenchyme markers, and loss of ductal downgrowth, are produced by misactivation of Hedgehog signaling (*Gli3*$^{xt/xt}$ and *Gli2*$^{1ki/1ki}$) [26]. In this case, however the phenotype is accompanied by loss of epithelial Wnt signaling. While further work is required to decipher the connections analyses on Gli3R, Eda and PTHrP suggest that Wnt signaling plays critical roles both upstream and downstream of PTHrP in different compartments. Additional downstream effectors of PTHrP signaling include BMPs and Msx2 [62]. Addition of Bmp4 to cultured *PTHrP*$^{-/-}$ mammary buds has been shown to rescue their arrested sprouting. PTHrP induction of Msx2 is proposed to suppress hair follicles formation in the nipple region [62]. Supporting this concept, loss of hair follicles seen in ventral skin in K14-PTHrP mice is rescued by Msx2 deletion[62, 63].

4. Ductal morphogenesis, branching and lumen formation: Wnts, Fgf10, Eda and Tgfβ pathways

Once the mammary rudiment has been specified, the last stages involve sprouting, branching and lumen formation. The onset of proliferation ~E15.5 coincides with resumption of Wnt reporter activity within the mammary sprout [14]. A requirement for Wnt activity at this stage is demonstrated by the fact that ablation of Wnt co-receptor, Lrp6, or pygopus, a canonical Wnt modulator delays and stunts their outgrowth (**Table 2**) [64-66]. Branching begins ~E16. Deficiency in Eda, Edar, or NFκB also result in smaller ductal trees with less branches, and the converse is seen in mice overexpressing Eda and Edar [36]. Fgf10 is also implicated in ductal branching, $Fgf10^{-/-}$ glands fail to ramify and remain as sprouts, however FGF also promotes adipogenesis in the fat pad via CEBPβ [67]. Lumen formation in the embryonic mammary gland is thought to involve both apoptosis of cells at the center of the solid ductal chords as well as changes in cell adhesion. Of note LTBP1, which in the adult gland is a highly specific marker of ductal luminal cells becomes expressed coincident with the first appearance of micro lumen and is one of the earliest reliable marker of luminal cell specification [23](**Fig. 1**). This expression pattern suggests involvement of TGFβ pathways in generating polarity and cavitation.

5. Embryonic mammary stem cells and progenitors

Mammary stem cells (MaSCs) are specified and committed during embryonic development. Classical transplantation studies demonstrated that whole buds from as early as E13 could successfully regenerate a tree in adult fat pads [68]. In contrast Spike et al found that isolated mammary epithelial cells, termed fetal mammary stem cells (fMaSCs), did not gain this capability until E15.5 [5]. Collectively this suggests that until E15.5 mammary epithelial cells require signals form the underlying mesenchyme in order to retain a mammary identity and only after that time are truly committed to mammary fate. E15.5 fMaSCs increase markedly over the course of embryonic development reaching their peak ~E18.5 where they are more prevalent than MASCs in adult glands. They express levels of CD24 and CD49f intermediate between those that define luminal and basal progenitor populations in adult glands [5]. Several other mammary stem/progenitor markers are expressed at this time. For example, s-SHIP is expressed strongly at E13.5 and as discussed above has been utilized to follow cells migrating into mammary placodes [21, 69, 70]. As might be predicted from the requirement for Wnt signaling at all stages

of mammary development, several Wnt target genes, such as Lgr5 and Axin2, are expressed at this time and have been used to identify or isolate embryonic mammary stem/progenitor populations [71, 72].

Early lineage tracing studies showed that K14-expressing cells labelled at E17 gave rise to both basal and luminal lineages in adult mice, and thus at the population level were bipotent[73]. However, more recent scRNAseq and lineage tracing experiments have demonstrated that embryonic cells are initially bipotent cells, but become unipotent lineage restricted progenitors prior to birth [3]. Wuidart et al showed by single cell sequencing that E14.5 Lgr5/CD49f^{hi} embryonic progenitors express both basal and luminal genes. However, by E17, as microlumen appear and branching occurs, lineage segregation begins [3]. Cells in the inner layer upregulate K8 expression and cells in the outer layer express higher p63 levels. They further demonstrated that ectopic expression of p63 in adult luminal cells could promote a basal-like state [3]. Similar conclusions were made using Notch1 lineage tracing. Lilja et al, showed that embryonic Notch1-positive cells also express both luminal and basal markers [74]. Using a multicolor lineage tracing approach with a Notch1-creERT2 mouse line, they found that cells labelled at E13.5 and E15.5 gave rise to both basal and luminal cells. However, their mathematical model demonstrated that individual Notch1 cells were already unipotent committed progenitors at E12.5 despite their undifferentiated phenotype [74]. Finally, they showed that when Notch1 was ectopically expressed in basal cells it could switch them to luminal ER alpha negative cell fate [74] [75]. Lineage restriction in late embryogenesis has also been assessed by neutral lineage tracing approach using a R26-CreERT2;R26-Confetti mouse line that avoids the bias inherent to the use of gene promoters. Using this approach Lloyd-Lewis et al. also found evidence that lineage restriction embryonic mammary progenitors had already occurred at E16.5-E17.5[75]. Collectively, these studies show that although very early embryonic mammary populations are multipotent, lineage commitment at the individual cell level occurs early during embryonic mammary development and identify key roles for Notch1 and p63 in driving luminal or basal fate [3, 74-78].

Conclusion

In summary, repression of Hedgehog signaling plays a central role in appendage fate and is the critical point of divergence between hair and mammary development. Mesenchymal patterning by Bmp4 and Tbx3 designates the mammary zone and is influenced by homeotic regulation of somites and limbs. Nrg3 promotes cell migration into placodes and Eda defines their radius and spacing. All pathways intersect in site-specific ways to promote critical thresholds of placodal Tbx3 and Wnt, which sustain the potency of undifferentiated embryonic mammary cells. PTHrP induces mammary mesenchymal differentiation, which acts via Wnt and BMP signaling to unleash a suite of factors that instruct the bud and epithelium to commit to mammary and nipple identity. Subsequent actions of Eda, FGF and TGFβ, promote lumen formation and ductal branching and are temporally linked to potency restriction and lineage segregation driven by Notch and p63. Thus, embryonic patterning pathways play two key roles regulating potency and migration. These features frequently become revived in the context of breast cancer as evidenced by the preponderance of embryonic stem/progenitor signatures in breast tumors and reactivation of cell movement leading to metastasis [3-9]. Therefore, a greater understanding of how these pathways collaborate to regulate embryonic mammary development is necessary to advance in our ability to thwart their collusion in the pathological setting. Classical studies on appendage development pointed to the significant role of epithelial and mesenchymal reciprocity [79]. Recent studies are now revealing a glimmer of hope that it may be possible by rebalancing signals between these compartments to control some types of cancer by promoting differentiation [80, 81].

Acknowledgments

This article was supported by Department of Defense W81XWH-17-1-0013 BC160959 (PC).

References

1. Oftedal, O.T. (2002) The mammary gland and its origin during synapsid evolution. J Mammary Gland Biol Neoplasia 7 (3), 225-52.
2. Oftedal, O.T. and Dhouailly, D. (2013) Evo-devo of the mammary gland. J Mammary Gland Biol Neoplasia 18 (2), 105-20.
3. Wuidart, A. et al. (2018) Early lineage segregation of multipotent embryonic mammary gland progenitors. Nat Cell Biol 20 (6), 666-676.
4. Van Keymeulen, A. et al. (2015) Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity. Nature 525 (7567), 119-23.
5. Spike, B.T. et al. (2012) A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. Cell Stem Cell 10 (2), 183-97.
6. Eyre, R. et al. (2019) Microenvironmental IL1beta promotes breast cancer metastatic colonisation in the bone via activation of Wnt signalling. Nat Commun 10 (1), 5016.
7. Giraddi, R.R. et al. (2018) Single-Cell Transcriptomes Distinguish Stem Cell State Changes and Lineage Specification Programs in Early Mammary Gland Development. Cell Rep 24 (6), 1653-1666 e7.
8. Dravis, C. et al. (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.
9. Dravis, C. et al. (2015) Sox10 Regulates Stem/Progenitor and Mesenchymal Cell States in Mammary Epithelial Cells. Cell Rep 12 (12), 2035-48.
10. Cho, K.W. et al. (2006) Molecular interactions between Tbx3 and Bmp4 and a model for dorsoventral positioning of mammary gland development. Proc Natl Acad Sci U S A 103 (45), 16788-93.
11. Propper, A.Y. et al. (2013) Prenatal morphogenesis of mammary glands in mouse and rabbit. J Mammary Gland Biol Neoplasia 18 (2), 93-104.
12. Lee, M.Y. et al. (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), e26242.
13. Veltmaat, J.M. et al. (2004) Identification of the mammary line in mouse by Wnt10b expression. Dev Dyn 229 (2), 349-56.
14. Chu, E.Y. et al. (2004) Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis. Development 131 (19), 4819-29.
15. van Genderen, C. et al. (1994) Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in Lef-1 deficient mice. Genes and Development 8, 2691-2704.
16. Davenport, T.G. et al. (2003) Mammary gland, limb and yolk sac defects in mice lacking Tbx3, the gene mutated in human ulnar mammary syndrome. Development 130 (10), 2263-73.
17. Douglas, N.C. and Papaioannou, V.E. (2013) The T-box transcription factors TBX2 and TBX3 in mammary gland development and breast cancer. J Mammary Gland Biol Neoplasia 18 (2), 143-7.

18. Propper, A.Y. (1978) Wandering epithelial cells in the rabbit embryo milk line. A preliminary scanning electron microscope study. Dev Biol 67 (1), 225-31.
19. Mailleux, A.A. et al. (2002) Role of FGF10/FGFR2b signaling during mammary gland development in the mouse embryo. Development 129 (1), 53-60.
20. Veltmaat, J.M. et al. (2006) Gli3-mediated somitic Fgf10 expression gradients are required for the induction and patterning of mammary epithelium along the embryonic axes. Development 133 (12), 2325-35.
21. Kogata, N. et al. (2014) Neuregulin-3 regulates epithelial progenitor cell positioning and specifies mammary phenotype. Stem Cells Dev 23 (22), 2758-70.
22. Kogata, N. et al. (2013) Neuregulin 3 and erbb signalling networks in embryonic mammary gland development. J Mammary Gland Biol Neoplasia 18 (2), 149-54.
23. Chandramouli, A. et al. (2013) Ltbp1L is focally induced in embryonic mammary mesenchyme, demarcates the ductal luminal lineage and is upregulated during involution. Breast Cancer Res 15 (6), R111.
24. Chandramouli, A. et al. (2011) Choreographing metastasis to the tune of LTBP. J Mammary Gland Biol Neoplasia 16 (2), 67-80.
25. Hatsell, S.J. and Cowin, P. (2006) Gli3-mediated repression of Hedgehog targets is required for normal mammary development. Development 133 (18), 3661-70.
26. Chandramouli, A. et al. (2013) Gli activity is critical at multiple stages of embryonic mammary and nipple development. PLoS One 8 (11), e79845.
27. Gritli-Linde, A. et al. (2007) Abnormal hair development and apparent follicular transformation to mammary gland in the absence of hedgehog signaling. Dev Cell 12 (1), 99-112.
28. Kawai, S. and Sugiura, T. (2001) Characterization of human bone morphogenetic protein (BMP)-4 and -7 gene promoters: activation of BMP promoters by Gli, a sonic hedgehog mediator. Bone 29 (1), 54-61.
29. Jerome-Majewska, L.A. et al. (2005) Tbx3, the ulnar-mammary syndrome gene, and Tbx2 interact in mammary gland development through a p19Arf/p53-independent pathway. Dev Dyn 234 (4), 922-33.
30. Bamshad, M. et al. (1997) Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. Nat Genet 16 (3), 311-5.
31. Baban, A. et al. (2009) Poland syndrome with bilateral features: case description with review of the literature. Am J Med Genet A 149A (7), 1597-602.
32. Carroll, L.S. and Capecchi, M.R. (2015) Hoxc8 initiates an ectopic mammary program by regulating Fgf10 and Tbx3 expression and Wnt/beta-catenin signaling. Development 142 (23), 4056-67.
33. Wansbury, O. et al. (2008) Dynamic expression of Erbb pathway members during early mammary gland morphogenesis. J Invest Dermatol 128 (4), 1009-21.
34. Howard, B. et al. (2005) Identification of the scaramanga gene implicates Neuregulin3 in mammary gland specification. Genes Dev 19 (17), 2078-90.
35. Lindfors, P.H. et al. (2013) Ectodysplasin/NF-kappaB signaling in embryonic mammary gland development. J Mammary Gland Biol Neoplasia 18 (2), 165-9.
36. Myllymaki, S.M. and Mikkola, M.L. (2019) Inductive signals in branching morphogenesis - lessons from mammary and salivary glands. Curr Opin Cell Biol 61, 72-78.

37. Clarke, A. et al. (1987) Clinical aspects of X-linked hypohidrotic ectodermal dysplasia. Arch Dis Child 62 (10), 989-96.
38. Megarbane, H. et al. (2008) Unusual presentation of a severe autosomal recessive anhydrotic ectodermal dysplasia with a novel mutation in the EDAR gene. Am J Med Genet A 146A (20), 2657-62.
39. Cui, C.Y. et al. (2005) X-linked anhidrotic ectodermal dysplasia disruption yields a mouse model for ocular surface disease and resultant blindness. Am J Pathol 167 (1), 89-95.
40. Haghighi, A. et al. (2013) Whole-exome sequencing identifies a novel missense mutation in EDAR causing autosomal recessive hypohidrotic ectodermal dysplasia with bilateral amastia and palmoplantar hyperkeratosis. Br J Dermatol 168 (6), 1353-6.
41. Mikkola, M.L. (2009) Molecular aspects of hypohidrotic ectodermal dysplasia. Am J Med Genet A 149A (9), 2031-6.
42. Voutilainen, M. et al. (2015) Ectodysplasin/NF-kappaB Promotes Mammary Cell Fate via Wnt/beta-catenin Pathway. PLoS Genet 11 (11), e1005676.
43. Voutilainen, M. et al. (2012) Ectodysplasin regulates hormone-independent mammary ductal morphogenesis via NF-kappaB. Proc Natl Acad Sci U S A 109 (15), 5744-9.
44. Mustonen, T. et al. (2004) Ectodysplasin A1 promotes placodal cell fate during early morphogenesis of ectodermal appendages. Development 131 (20), 4907-19.
45. Pispa, J. et al. (2003) Ectodysplasin, Edar and TNFRSF19 are expressed in complementary and overlapping patterns during mouse embryogenesis. Gene Expr Patterns 3 (5), 675-9.
46. Narhi, K. et al. (2012) Sostdc1 defines the size and number of skin appendage placodes. Dev Biol 364 (2), 149-61.
47. Fliniaux, I. et al. (2008) Identification of dkk4 as a target of Eda-A1/Edar pathway reveals an unexpected role of ectodysplasin as inhibitor of Wnt signalling in ectodermal placodes. Dev Biol 320 (1), 60-71.
48. Gierer, A. and Meinhardt, H. (1972) A theory of biological pattern formation. Kybernetik 12 (1), 30-9.
49. Meinhardt, H. and Gierer, A. (2000) Pattern formation by local self-activation and lateral inhibition. Bioessays 22 (8), 753-60.
50. Loukas, M. et al. (2007) Accessory breasts: a historical and current perspective. Am Surg 73 (5), 525-8.
51. Kamberov, Y.G. et al. (2013) Modeling recent human evolution in mice by expression of a selected EDAR variant. Cell 152 (4), 691-702.
52. Elo, T. et al. (2017) Ectodysplasin target gene Fgf20 regulates mammary bud growth and ductal invasion and branching during puberty. Sci Rep 7 (1), 5049.
53. Dunbar, M.E. et al. (1999) Parathyroid hormone-related protein signaling is necessary for sexual dimorphism during embryonic mammary development. Development 126 (16), 3485-93.
54. Wysolmerski, J.J. et al. (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), 1285-94.
55. Dunbar, M.E. et al. (1998) Stromal cells are critical targets in the regulation of mammary ductal morphogenesis by parathyroid hormone-related protein. Dev Biol 203 (1), 75-89.

56. Foley, J. et al. (2001) Parathyroid hormone-related protein maintains mammary epithelial fate and triggers nipple skin differentiation during embryonic breast development. Development 128 (4), 513-25.
57. Dunbar, M.E. et al. (2001) Temporally regulated overexpression of parathyroid hormone-related protein in the mammary gland reveals distinct fetal and pubertal phenotypes. J Endocrinol 171 (3), 403-16.
58. Jobert, A.S. et al. (1998) Absence of functional receptors for parathyroid hormone and parathyroid hormone-related peptide in Blomstrand chondrodysplasia. J Clin Invest 102 (1), 34-40.
59. Hens, J. et al. (2009) Analysis of gene expression in PTHrP-/- mammary buds supports a role for BMP signaling and MMP2 in the initiation of ductal morphogenesis. Dev Dyn 238 (11), 2713-24.
60. Hiremath, M. and Wysolmerski, J. (2013) Parathyroid hormone-related protein specifies the mammary mesenchyme and regulates embryonic mammary development. J Mammary Gland Biol Neoplasia 18 (2), 171-7.
61. Hiremath, M. et al. (2012) Parathyroid hormone-related protein activates Wnt signaling to specify the embryonic mammary mesenchyme. Development 139 (22), 4239-49.
62. Hens, J.R. et al. (2007) BMP4 and PTHrP interact to stimulate ductal outgrowth during embryonic mammary development and to inhibit hair follicle induction. Development 134 (6), 1221-30.
63. Satokata, I. et al. (2000) Msx2 deficiency in mice causes pleiotropic defects in bone growth and ectodermal organ formation. Nat Genet 24 (4), 391-5.
64. Lindvall, C. et al. (2006) The Wnt signaling receptor Lrp5 is required for mammary ductal stem cell activity and Wnt1-induced tumorigenesis. J Biol Chem 281 (46), 35081-7.
65. Gu, B. et al. (2009) Pygo2 expands mammary progenitor cells by facilitating histone H3 K4 methylation. J Cell Biol 185 (5), 811-26.
66. Lindvall, C. et al. (2009) The Wnt co-receptor Lrp6 is required for normal mouse mammary gland development. PLoS One 4 (6), e5813.
67. Rivetti, S. et al. (2020) Fgf10/Fgfr2b Signaling in Mammary Gland Development, Homeostasis, and Cancer. Front Cell Dev Biol 8, 415.
68. Sakakura, T. et al. (1979) Capacity of mammary fat pads of adult C3H/HeMs mice to interact morphogenetically with fetal mammary epithelium. J Natl Cancer Inst 63 (3), 733-6.
69. Bai, L. and Rohrschneider, L.R. (2010) s-SHIP promoter expression marks activated stem cells in developing mouse mammary tissue. Genes Dev 24 (17), 1882-92.
70. Tian, L. et al. (2019) s-SHIP Promoter Expression Identifies Mouse Mammary Cancer Stem Cells. Stem Cell Reports 13 (1), 10-20.
71. Trejo, C.L. et al. (2017) Lgr5 is a marker for fetal mammary stem cells, but is not essential for stem cell activity or tumorigenesis. NPJ Breast Cancer 3, 16.
72. van Amerongen, R. et al. (2012) Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland. Cell Stem Cell 11 (3), 387-400.
73. Van Keymeulen, A. et al. (2011) Distinct stem cells contribute to mammary gland development and maintenance. Nature 479 (7372), 189-93.

74. Lilja, A.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 (6), 677-687.
75. Lloyd-Lewis, B. et al. (2018) Neutral lineage tracing of proliferative embryonic and adult mammary stem/progenitor cells. Development 145 (14).
76. Davis, F.M. et al. (2016) Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. Nat Commun 7, 13053.
77. Lloyd-Lewis, B. et al. (2017) Mammary Stem Cells: Premise, Properties, and Perspectives. Trends Cell Biol 27 (8), 556-567.
78. Bland, P. and Howard, B.A. (2018) Mammary lineage restriction in development. Nat Cell Biol 20 (6), 637-639.
79. Sakakura, T. (1987) Mammary Embryogenesis. In The Mammary Gland (Neville, M.C. and Daniel, C.W. eds), pp. 37-63, Plenum.
80. Sun, X. et al. (2020) Coordinated hedgehog signaling induces new hair follicles in adult skin. Elife 9.
81. Dow, L.E. et al. (2015) Apc Restoration Promotes Cellular Differentiation and Reestablishes Crypt Homeostasis in Colorectal Cancer. Cell 161 (7), 1539-1552.
82. Andl, T. et al. (2002) WNT signals are required for the initiation of hair follicle development. Dev Cell 2 (5), 643-53.
83. Ahn, Y. et al. (2013) Lrp4 and Wise interplay controls the formation and patterning of mammary and other skin appendage placodes by modulating Wnt signaling. Development 140 (3), 583-93.

Table 1. Table of Human syndromes involving major regulators of embryonic mammary development.

Table 2. Effects of loss and gain of function of key regulators of embryonic mammary development in mice.

Figure 1. Murine embryonic mammary development

Left to right. Mammary development begins in the mouse embryo with thickening of the ectoderm to generate 5 pairs of raised elliptical placodes. At E13.5 placodes invaginate to form flask-shaped buds surrounded by a specialized mammary mesenchyme. Once committed to mammary cell fate they proliferate and extend toward the underlying fat pad (E16.5). At E18.5 a phase of branching morphogenesis ensues and microlumen develop to form a hollow bi-layered rudimentary tree prior the birth. Histological sections show from left to right an E11.5 placode stained for keratin 14(K14); E13.5 mammary bud stained for K14 in brown and surrounded by mesenchymal Ltbp1L-*lacZ*; E16.5 mammary sprout at from Gli2-*lacZ* mouse; E18.5 rudimental tree expressing Ltbp1L-*lacZ*.

Figure 2. Schematic of signaling pathways regulating mammary development

a. The mammary line is specified by antagonism between dorsal Tbx3 and ventral Bmp4. Somitic cells expressing Fgf10 translocate to activate Fgfr2b in the ectoderm. Tbx3, Wnt10b and Lef1 become focally upregulated along the mammary line.

b. Placode formation. Nrg3 and its receptor become restricted and stimulate migration of progenitor cells into the placode. Tbx3 and Wnt10b are essential for the formation of most placodes and must be maintained at a critical threshold.

c. Eda/Edar signaling defines placodal and interplacodal regions by simultaneously stimulating expression of activators (e.g. Wnt10b and PTHrP) and inhibitors (e.g. DKK, Sostdc1) of mammary fate with different diffusion characteristics.

Figure 3. Wnt signaling is important at every stage of embryonic mammary development

A) Schematic of canonical Wnt signaling. **B-D**) X-gal stained sections of embryos show expression of the Wnt signaling reporter conductin-gal ~E12.5 within the placode, ~E14.5 in the bud and mammary mesenchyme, and ~E16.5 within the proliferating mammary sprout. Wnt signaling sustains placodal identity, plays a critical role in specifying the mammary mesenchyme, is important for ductal proliferation and elongation and regulates stem cell potency.

Figure 4. Ltbp1 is an early marker of mammary mesenchyme

X-gal stained section shows Ltbp1L-*lacZ* reporter expression in **A)** a mesenchymal axillary streak (white arrow) oriented towards mammary gland 1 (MG1) ~E12. Scale bars 200um. **B)** in mammary mesenchyme of ~E14.5 bud with immunolocalization of p63 Scale bars, 100um. From Chandramouli A, et al.2013.

Figure 5. Gli is expressed in hair follicles around mammary sprout.

A) Gli1-*lacZ* reporter of positive hedgehog is expressed in hair placodes but is repressed in mammary buds (box). Buds #3 and #5 are lost when Gli activators exceed Gli repressors and **B)** the remaining buds evaginate and Gli1-lacZ positive hair follicles form within the mammary field (arrow) Chandramouli A, et al.2013.

Figure 6. PTHrP specifies the mammary mesenchyme

A) PTHrP is secreted from the mammary epithelium (pink) and acts on the mammary mesenchyme (green) to activate its receptor PTHR1. Acting through Wnt and BmpR1 pathways this induces AR which activates mesenchymal constriction and apoptosis in male buds(**B**). (**C-E**),

Mammary mesenchyme markers AR, ER, Lef1 Tenascin C and many others signal to the bud and overlying epithelium to stimulates ductal sprouting in females, and nipple differentiation

Table 1. Table of Human defects/syndromes involving major regulators of embryonic mammary development.

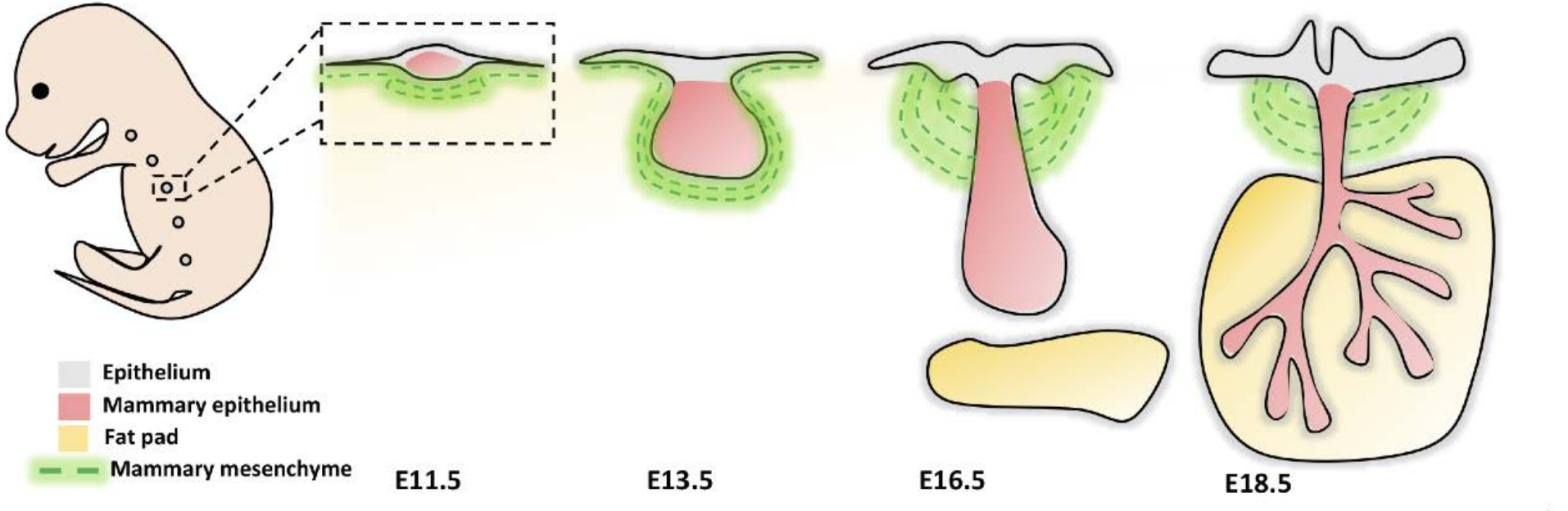
Genes	Human defects/syndrome
TBX3	Ulnar Mammary Syndrome (UMS)
FGF10/FGFR2	Poland syndrome
EDA/EDAR	Hypohidrotic Ectodermal Dysplasia (HED)
PTHLH	Blomstrand chondrodysplasia

Table 2. Effects of loss and gain of function of key regulators of embryonic mammary development in mice

Genes	Mouse model	Phenotype	Ref
Gli2A ***Gli3R***	*Gli3$^{xt/xt}$* *Gli2$^{1nki/1nki}$; Gli3$^{xt/+}$*	Loss MR3, MR5 Impaired mm specification MR2, MR4: Bud evagination Inappropriate hair follicle formation Loss of sexual dimorphism	25 26
Tbx3	*Tbx3$^{-/-}$*	Loss MR1,3,4,5	16
	Tbx3$^{-/+}$	Loss MR1-3 hypoplastic branching in others	17
Nrg3	*Nrg3^{ska}/Nrg3^{ska}* *K14-Nrg3*	MR3 loss MR4 duplication (more in adults)	34
Fgfr2b	*Fgfr2b$^{-/-}$*	Loss of all MR	19
Fgf10	*Fgf10$^{-/-}$*	Loss of MR1,2,3,5	19
Eda	*K14-Eda*	Supernumerary between MR3-4 & in neck	41, 44
Pthlh ***Pth1r***	*Pthlh$^{-/-}$* *Pth1r$^{-/-}$*	Arrest at late bud stage Impaired mm specification Loss of sexual dimorphism	53, 54
Dkk1	*K14-PthrP*	Gain of mm differentiation and loss of hair follicles in ventral epidermis develops nipple characteristics	14, 56
	K5-rtTA;teO-Dkk1	No placodes	82
Sostdc1	*Sostdc1$^{-/-}$*	Enlarged placodes, MR2 and 3 fused	46
	K14-rTA;tetO-WISE	Small MR	83
Lpr4	*Lrp4$^{mdit/mdit}$* *Lpr4$^{mitt/mitt}$*	Fusion of MR2 & 3	83
	Lpr4$^{mdit/mdit}$;Lpr5$^{-/-}$	Rescued buds 2 and 3 fusion	
	Lpr4$^{mdit/mdit}$;Lpr6$^{-/-}$		
	Lpr4$^{mdit/mdit}$;Lpr5+/Lpr6$^{+/-}$		
	Lpr4$^{mdit/mitt}$;K14cre;Bcatenin$^{fx/-}$	Small placodes, no bud downgrowth Inhibit ectopic Wnt signaling between MR2 and 3	
Lpr5	*Lpr5$^{-/-}$*	Small placodes-slow development	64
Lpr6	*Lpr6$^{-/-}$*	Impaired ductal outgrowth and adipogenesis	66
	MMTV-Wnt1;Lpr6$^{-/-}$	Rescue MR	
β-catenin	*K14cre; β-catenin$^{fx/-}$*	Small MR	82
	Dermo1-cre; β-catenin$^{lox/lox}$	Impaired mm specification Loss of sexual dimorphism	61
Lef1	*Lef1$^{-/-}$*	Loss MR2, MR3 Arrest MR1,4,5 at bud stage & demise	15
Pygo2	Pygo2$^{-/-}$	E18.5 MR loss or impaired elongation and branching	65

mm, mammary mesenchyme; MR, mammary rudiment.

Figure 1


Epithelium
Mammary epithelium
Fat pad
Mammary mesenchyme
E11.5
E13.5
E16.5
E18.5

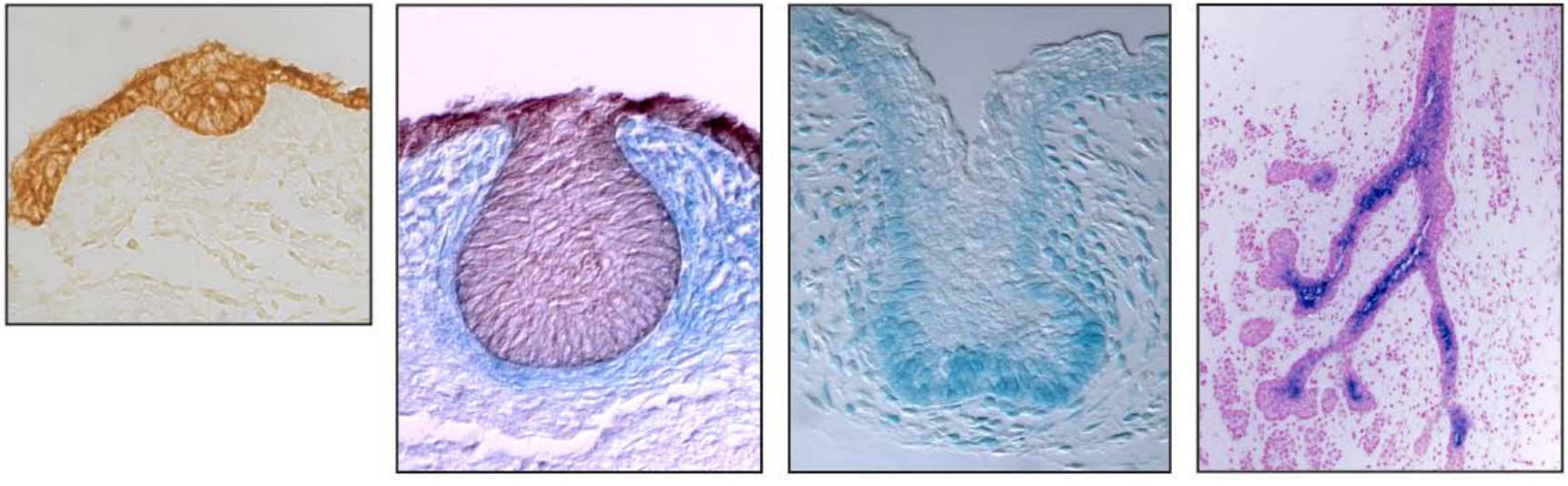


Figure 2

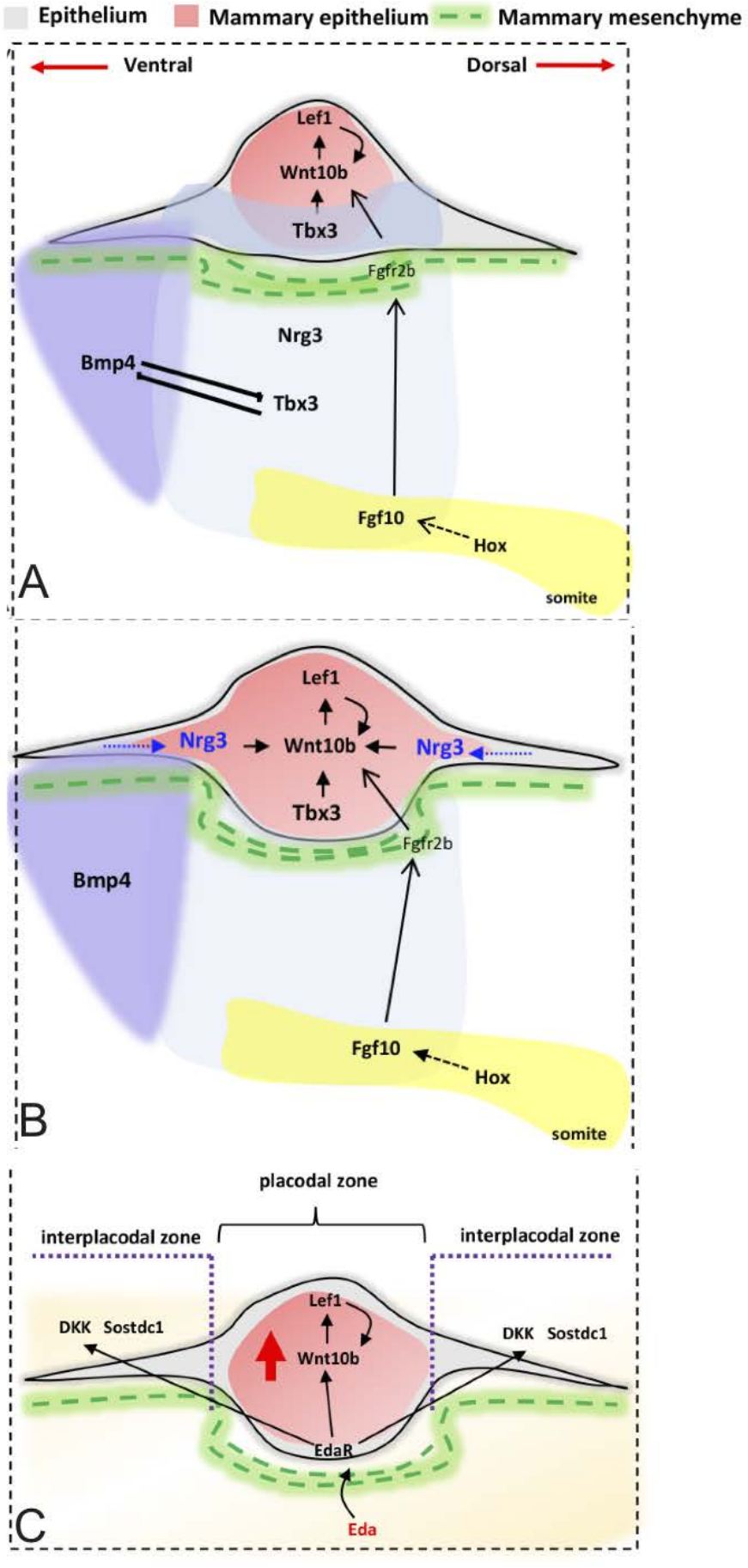
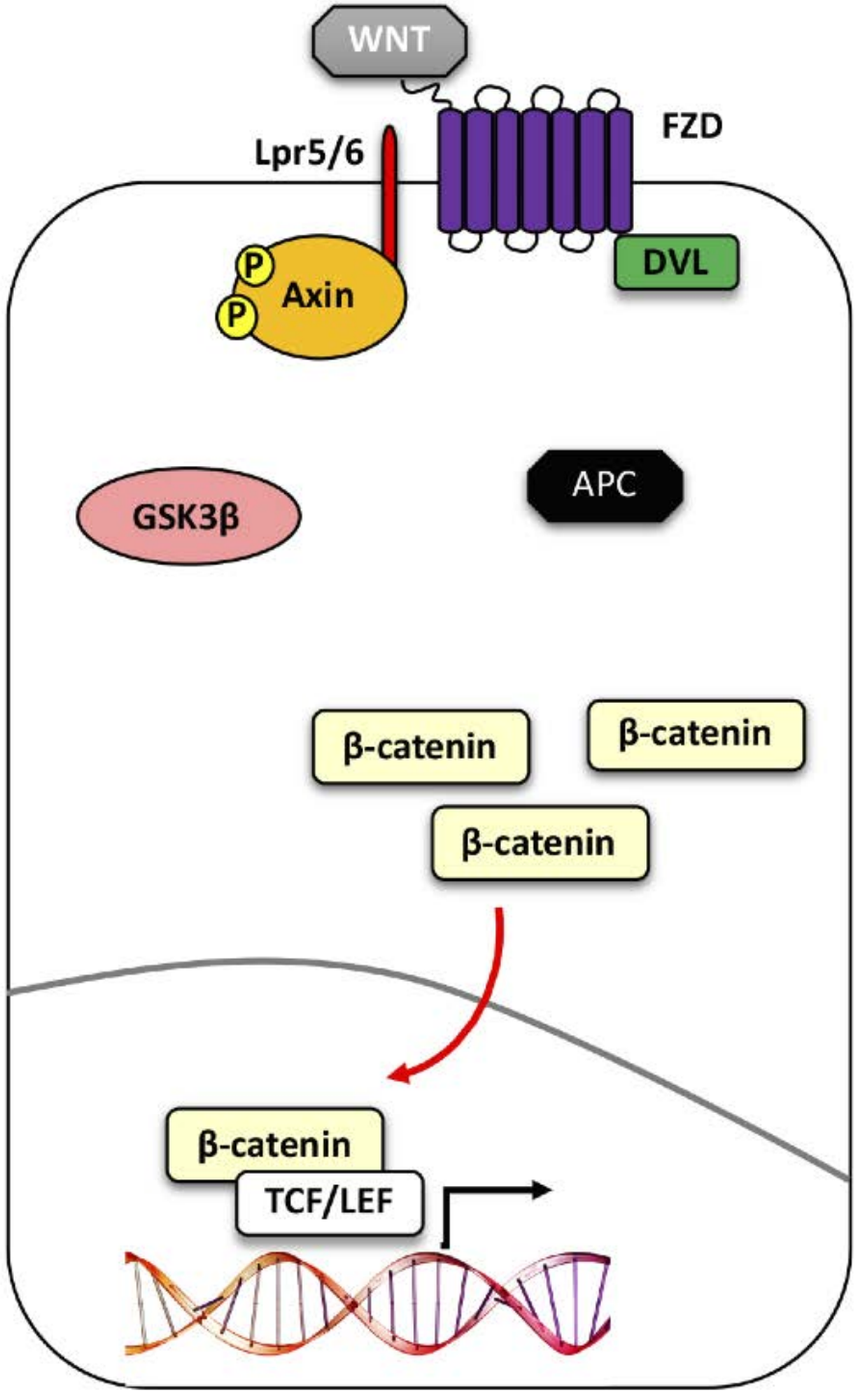
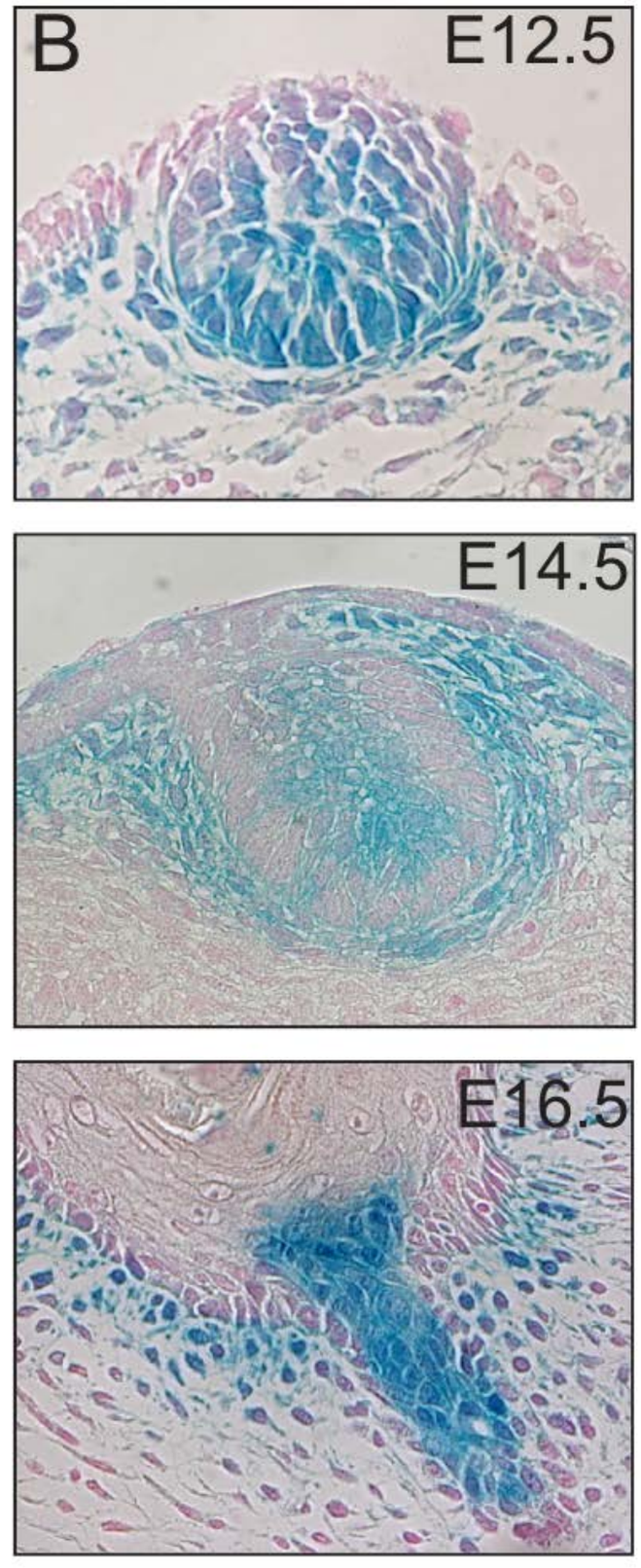

Epithelium
Mammary epithelium
Mammary mesenchyme
Ventral
Dorsal
Lef1
Wnt10b
Tbx3
Fgfr2b
Nrg3
Bmp4
Tbx3
Fgf10
Hox
somite
A
Lef1
Nrg3
Wnt10b
Nrg3
Tbx3
Fgfr2b
Bmp4
Fgf10
Hox
somite
B
placodal zone
interplacodal zone
interplacodal zone
DKK Sostdc1
Lef1
Wnt10b
DKK Sostdc1
EdaR
Eda
C

Figure 3

A Canonical Wnt pathway





NFR Conductin-gal

Figure 4

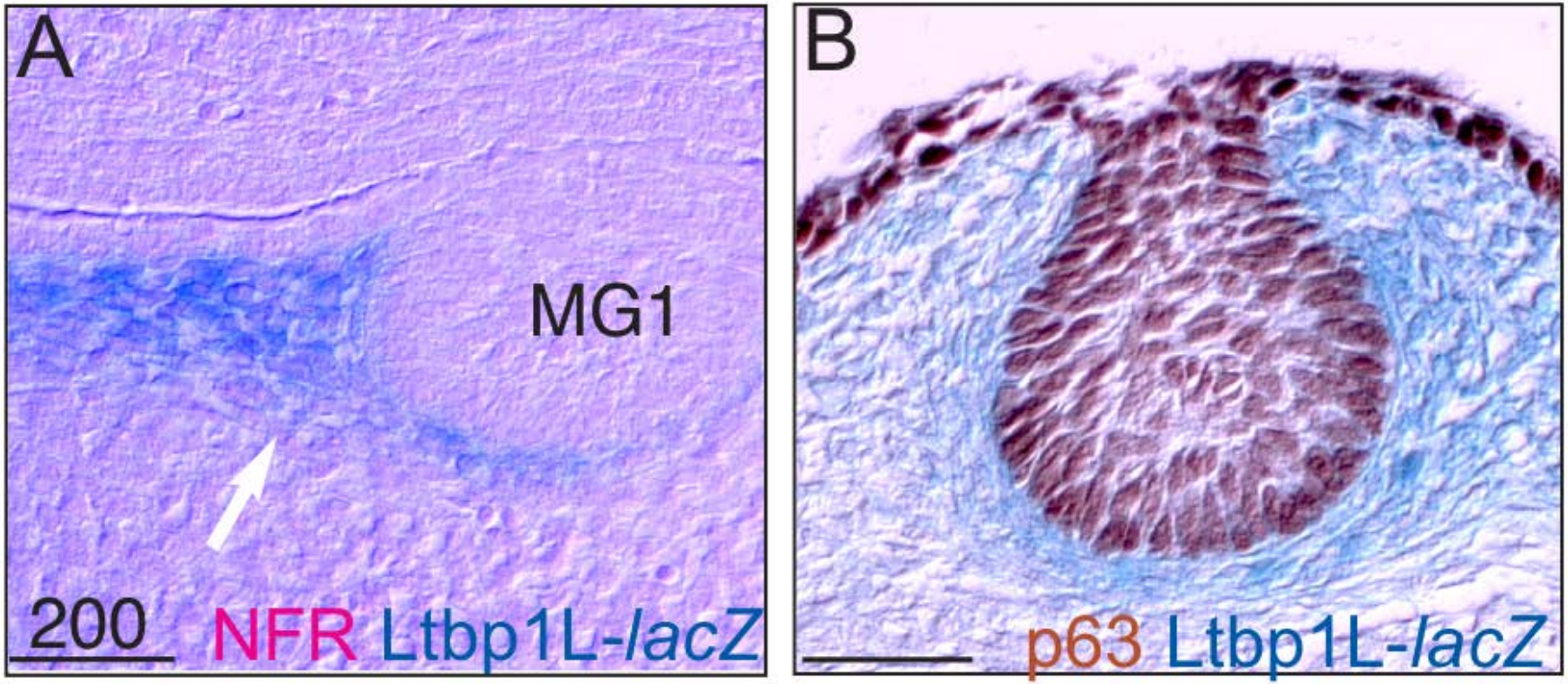


Figure 5

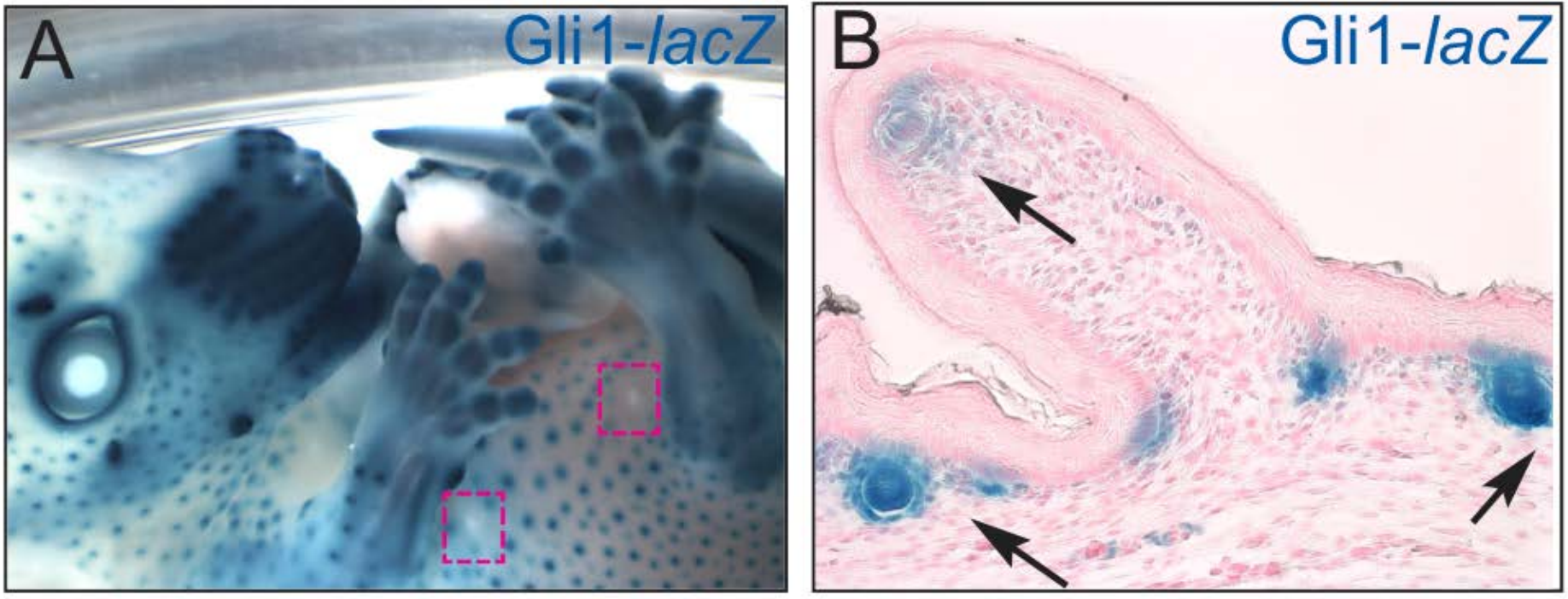
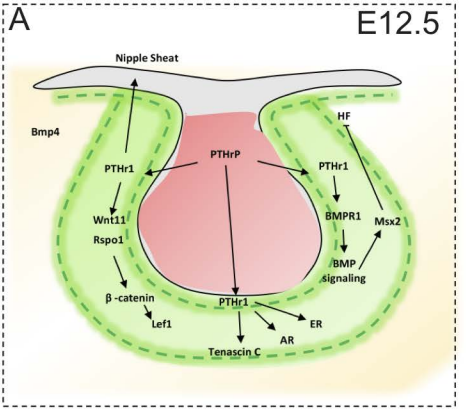
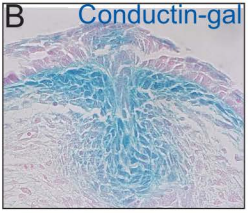
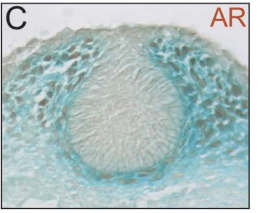
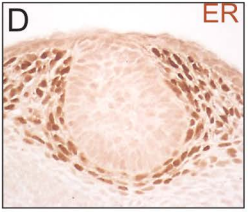


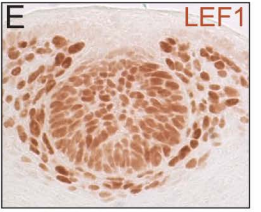
Figure 6


A
E12.5
Nipple Sheat
Bmp4
PTHr1
PTHrP
PTHr1
HF
Wnt11
Rspo1
BMPR1
Msx2
BMP
signaling
β -catenin
Lef1
PTHr1
ER
AR
Tenascin C
Epithelium
Mammary epithelium
Mammary mesenchyme


B
Conductin-gal


C
AR


D
ER


E
LEF1

Gpr125 Plays Critical Roles in Lacrimal Myoepithelia and Tear Film

Departments of Cell Biology[1] Dermatology[2] and Ophthalmology[3]

New York University School of Medicine

Elena Spina[1], Rebecca Handlin[1], Julia Simundza[1], Angela Incassati[1], Muneeb Faiq[3], Anoop Sainulabdeen[3], Kevin C Chan[3], Pamela Cowin[1,2]*

*Address for correspondence:

Dr. Pamela Cowin

New York University Medical Center

MSB 621, 550 1st Ave

New York, NY 10016

E-mail: pamela.cowin@nyulangone.org

Conflict of interest statement: The authors declare that no conflict of interest exists.

Abstract

Gpr125, encoded by *Adgra3*, is an orphan adhesion G-protein coupled receptor (aGPCR) implicated in Wnt signaling and planar polarity. Here we establish both physiological and pathological roles for Gpr125. We show that mice lacking Gpr125 display an ocular phenotype with many hallmarks of human dry eye disease. These include squinting, abnormal lacrimation, mucus accumulation, swollen eyelids and inflammatory infiltration of lacrimal and meibomian glands. We further demonstrate that mice expressing Gpr125 lacking six transmembrane and the cytoplasmic domain recapitulate the null phenotype, indicating that downstream signaling is essential. Utilizing a Gpr125-β-gal reporter and scRNAseq, we identify Gpr125 expression in a discrete population of embryonic myoepithelial cells located at the tips of developing lacrimal ducts. By lineage tracing we show these cells function as progenitors of the adult lacrimal myoepithelium. Beyond defining an essential role for Gpr125 in tear film and identifying its utility as a marker of lacrimal progenitors, this study implicates Gpr125 in the etiology of dry eye disease, and defines novel animal models of this common malady.

Introduction

Tears are required to lubricate corneal and conjunctival surfaces and to prevent eyes from desiccation (1). They also function to protect eyes from microbial infection and preserve visual acuity. Dry Eye Disease (DED) is a significant health problem that is particularly prevalent in the elderly and women (2-4). It affects ~5% of the population overall and is a central feature of Sjogren's syndrome, the third most common auto-immune disease (2-4).

Tear film is composed of three layers, each secreted from a different source. Goblet cells, clustered along the conjunctival rim, provide the inner mucus layer that spreads tear film evenly over the ocular surface (5). Meibomian glands, found between eyelash follicles on the inner surface of eyelids, produce the outer lipid layer that prevents evaporation (6). Lacrimal glands, secrete the central aqueous component that contains water-soluble immuno-active and antibacterial proteins, as well as glucose, urea, and salts (7). Defects in the volume or composition of any layer destabilizes tear film and induces DED.

Like other ectodermal appendages, lacrimal and meibomian glands develop during mid-embryogenesis. Murine lacrimal glands emerge around embryonic day 13 (E13) as a bulbous outgrowth of the conjunctival epithelium, which by E15 has elongated as a bi-layered hollow duct with 4-5 bulbous tips. Subsequent branching produces a compact mass of secretory acini, enmeshed by contractile myoepithelial cells, which fully differentiate after birth (7-10). Meibomian glands emerge ~E18 as a row of placodes along the margins of the inner eyelids (11). These invaginate as solid chords, form lumen

around postnatal day 3 (P3), then branch ~P5 and complete acinar differentiation ~P8-15 (11). Goblet cells appear in the conjunctival rim coincident with eye opening (5, 12).

Analysis of knockout (KO) mice has revealed a suite of critical regulators of tear film (5, 7-9, 13-20). However, the involvement of GPCRs in this process has not been studied. Here we uncover an essential physiological role for Gpr125 in tear film, and a pathological role in the etiology of DED. Gpr125 was discovered through homology searches of the human genome database and is an orphan aGPCR (21). Like other members of this family, Gpr125 has a large extracellular domain with sequence similarity to cell adhesion molecules (**Figure 1A**). It remains unclear, however, if it functions in cell adhesion or signals via G-proteins (21, 22). Previous studies have highlighted Gpr125 as a marker of undifferentiated murine spermatogonial progenitors (23), documented its elevation in the choroid plexus following injury and correlated its high expression with both good and poor outcome in cancer (24-26). Ectopic expression of Gpr125 has shown that in zebrafish it modulates Wnt/planar cell polarity processes by interacting with the cytoplasmic adaptor Disheveled (Dsh), and in cultured cells undergoes constitutive clathrin-mediated internalization to endosomes (27, 28). However, the physiological function of native Gpr125 has remained elusive.

Combined Results & Discussion

Mice lacking Gpr125 display blepharitis, blepharedema and mucoid accumulation. To address the role of native Gpr125, we developed mice that permit Gpr125 expression to be ablated and the lineage of cells normally expressing it to be traced by inserting a creER^{T2} cassette downstream of the *Adgra3* promoter (**Figure 1B**). Mice lacking Gpr125 expression (*Adgra3*$^{cre/cre}$) display a prominent eye phenotype (**Figure 1C**) comprising several hallmarks of DED. *Adgra3*$^{cre/cre}$ mice squint as soon as their eyes open; whereas, heterozygous *Adgra3*$^{cre/+}$ are indistinguishable from wild-type littermates (**Figure 1C**). As *Adgra3*$^{cre/cre}$ mice mature, this early blepharitis progresses to blepharedema (swollen balding eyelids) and mucus precipitation (**Figure 1D**). The phenotype is constant in males but in females oscillates with reproductive status, becoming pronounced during pregnancy and lactation. During these stages, mice develop proptosis (bulging eyes) that resolves during weaning. The eye phenotype in *Adgra3*$^{cre/cre}$ mice is 100% penetrant on all strain backgrounds examined (C57B6/CH3, FVBN, and mixed). To dissect the role of Gpr125's adhesion ectodomain from its internal signaling functions we examined a second strain, *Adgra3*$^{lz/lz}$, which expresses the Gpr125 ectodomain and 1st transmembrane domain fused in frame to β-galactosidase and lacks regions required for signaling/adaptor functions (**Figure 1E**) (23). Homozygous *Adgra3*$^{lz/lz}$ mice recapitulate the *Adgra3*$^{cre/cre}$ null phenotype (**Figure 1F**), whereas, *Adgra3*$^{lz/+}$ mice are normal. Collectively, these data demonstrate that Gpr125 has an essential physiological role in normal eye development and indicate that signaling downstream of the receptor is required. Our genetic analyses also reveal that loss of Gpr125 protein or Gpr125 signaling

is sufficient to trigger several common eye pathologies, such as blepharitis, blepharedema and mucoid accumulation.

Gpr125 in ciliary body and iris

Next we examined adult eye globes by X-gal staining and found strong Gpr125-β-gal expression in the ciliary body, which secretes aqueous humor, and in the inner layer of the iris of *Adgra3*$^{lz/+}$ mice (**Figure 2A, B**). As abnormal aqueous humor dynamics can alter intraocular pressure (IOP), which is a major risk factor for glaucoma, we measured IOP, but found no significant difference between wildtype and mutant genotypes (**Figure 2C**).

Goblet cell and meibomian pathologies

We submitted both strains for evaluation by a veterinary ophthalmologist. Examination of the lens and retina by slit lamp revealed well-documented characteristics of control B6 and FVBN mice, but no abnormality specifically linked to the *Adgra3*$^{cre/cre}$ or *Adgra3*$^{lz/lz}$ genotypes. Fluorescein staining found no evidence for corneal abrasion, but highlighted the presence of large mucoid precipitates around the eyelids of homozygous mutants (**Figure 2D**). This feature pointed towards abnormal tear film composition (**Figure 2E**), prompting us to investigate the tear glands in more detail. As *Adgra3* mutant mice had swollen eyelids we looked for changes in goblet cells and meibomian glands. Histological sections of eyelids stained with Alcian blue revealed goblet cell in *Adgra3*$^{cre/cre}$ mice (**Figure 2F**) but with greater variation in number (average =70/eyelid; range 7-210 n=23) compared to controls (average of 65 goblet cells/eyelid; range 45-94; n=23): some

showed epithelial and goblet cell desquamation next to swathes of mucus; others showed clusters of goblet cells. Meibomian glands displayed inflammatory infiltration by T-cells and macrophages (**Figure 2G,H**). Surprisingly, X-gal staining of eyelids indicated that goblet cells and meibomian glands were devoid of Gpr125 expression (**Figure 2I**), though eyelash follicles stained prominently, serving as internal procedural controls. These data show that changes in goblet and meibomian glands, reminiscent of symptoms accompanying human DED, occur in *Adgra3* mutants. However, they indicate that these changes are not the initiating event. Rather, they are the secondary abrasive consequences of loss of Gpr125 elsewhere.

Adgra3$^{cre/cre}$ and *Adgra3*$^{lz/lz}$ mice have abnormal lacrimation.

Next we tested whether Gpr125 loss affected the lacrimal function by measuring tear volume. *Adgra3*$^{cre/cre}$ and *Adgra3*$^{lz/lz}$ mice produced two to three-fold more tears than heterozygous or wild-type controls (**Figure 3A**). Tear volume was greater in female than male mice. *Adgra3*$^{cre/cre}$ and *Adgra3*$^{lz/lz}$ mice often presented (**Figure 3H**) with a mild phenotype in one eye (squint only) and a severe phenotype (blepharedema and or mucus) in the other. When we separated eyes into these two categories according to photographic assignment taken prior to measurement and reanalyzed the data, we found tear volume for the mild phenotypic category were similar, and sometimes lower, than those of wildtypes. In contrast, those in the severe category showed high values indicative of excessive tearing (**Figure 3A**). Thus, our mice recapitulated the paradoxical

phenomenon documented in human patients where individuals with tear film abnormality originating from initial ocular dryness respond with secondary hyper-lacrimation.

Adgra3$^{cre/cre}$ and *Adgra3*$^{lz/lz}$ mice show inflammatory infiltration of lacrimal glands

Inflammatory infiltration of lacrimal glands plays a significant role in the pathological mechanism of Sjogren's syndrome and is a frequent feature of sporadic DED (2-4, 29). To test if *Adgra3* mutant mice modeled this we examined histological sections of lacrimal glands and found foci, composed of small round cells, in both *Adgra3*$^{cre/cre}$ (**Figure 3B, S1**) and *Adgra3*$^{lz/lz}$ mice. Lack of immunostaining for cytokeratin5 (CK5) indicated that myoepithelial cells were lost in these regions (**Figures 3C,D**). The foci were surrounded by F480-positive macrophages (**Figures 3C,E**), and filled with cells recognized by CD4 and CD8 antibodies (**Figures 3C, F,G, S1**) indicating infiltration by T-helper cells and cytoxic T-cells. Histological analysis of females with facial swelling (**Figure 3H**) and proptosis during pregnancy and lactation revealed enlarged lacrimal glands (**Figure 3I,J**) and swathes of macrophages around large areas of acinar loss (**Figure 3K**). Thus, loss of Gpr125 predisposes the lacrimal gland to lymphocytic infiltration, replicating the inflammatory infiltration seen in human DED and the epidemiological characteristics seen in humans with respect to gender and reproductive status.

Gpr125-expressing cells are located at the leading tips of ducts during lacrimal development and function as progenitors of the lacrimal myoepithelium.

Given the significant effect of *Adgra3* mutation and loss on lacrimal structure and function, we sought to identify cell types that express Gpr125 over the course of lacrimal development. We began by mining scRNAseq data (10). Gpr125 mRNA was detected in a small cell population that co-expressed myoepithelial mRNAs: keratin 14, smooth muscle actin, and Sox10 (**Figure 4A**). This population was present during ductal elongation (E16) but diminished by P4 (data not shown) as acinar differentiation ensued. Next, we stained embryos with X-gal to look for Gpr125 expression (**Figure 4B**). Gpr125-β-gal appeared in the lacrimal bud as it emerged from the conjunctival rim ~E14, but by E15.5 it was restricted to a discrete population of cells located at the leading tips of lacrimal ducts and by P1 at the front of lacrimal branches. Intriguingly, during the course of these experiments we noted that Gpr125-β-gal was also expressed within a well-characterized "bulge" stem cell compartment of hair follicles and whiskers (**Figure 4B**) (30). This prompted us to ask if the embryonic Gpr125-positive cells functioned similarly as lacrimal progenitors. To test this, we performed lineage tracing by using the *creER*^{T2} cassette present in *Adgra3*^{cre} mice to activate expression of a lineage reporter. We labeled embryos harboring the ROSA-lox-STOP-lox-tdTomato reporters at E13-E15 by delivering tamoxifen to *Adgra3*$^{cre/cre}$ dams during mid-pregnancy. Lacrimal glands were harvested at 7 weeks and 6 months of age, cleared and analyzed by 3-D immunofluorescence confocal microscopy (**Figure 4C**). At 7 weeks after birth, we found tdTomato (tdT)-labeled cells with an elongated shape along the basal borders of ducts and with stellate

morphology enmeshing acini. These characteristics, together with their expression of CK5, identified them as contractile myoepithelial cells. A similar pattern was seen in glands from mice harvested at 6 months, indicating significant longevity of the original progenitor population (**Figures 4D,E**).

Collectively, these data support the concept that Gpr125 is expressed in cells at migrating tips of embryonic lacrimal ducts that function as long-lived unipotent progenitors of the ductal and acinar myoepithelium. This finding, taken together with previous reports where Gpr125 expression was used to isolate spermatogonial progenitors, suggests that Gpr125 could have utility in isolating lacrimal progenitors.

Gpr125 is essential for the generation of a functional tear film

Our results establish a physiological role for Gpr125 in the eye and show that loss of this aGPCR induces many features of DED (**Figure 1**). Human DED results from tear film instability exposing the eye to irritation and desiccation, prompting compensatory excessive tearing and immune response, which leads to further lacrimal destruction. Our *Adgra3* mutants provide a new model that reproduces this complex spectrum of symptoms, from early eye discomfort to blepharedema (**Figure 1**), mucus accumulation **(Figures 1, 2D),** compensatory hyper-lacrimation, inflammatory infiltration of meibomian **(Figures 2H)** and lacrimal glands **(Figure 3B-G),** and goblet cells desquamation **(Figure 2F).** Moreover, *Adgra3* mutants show worsening of their eye phenotype during pregnancy

and lactation, recapitulating the hormonal/gender epidemiology of DED, which is more prevalent in women and exacerbated by pregnancy and post-menopause.

Many genetic mouse models of DED arise from immune dysregulation or defects in matrix inhibition of immune activation and thus recapitulate late stages of DED. In contrast, our mice identify an initiating event during lacrimal development that predisposes mice to the full pathophysiological progression of DED. Our data reveal that a discrete myoepithelial subpopulation of embryonic cells located at tips of ducts and branches express Gpr125, and function as unipotent progenitors of the lacrimal myoepithelium (**Figure 4**). Importantly, we show that in the absence of Gpr125, focal areas of the lacrimal gland become devoid of myoepithelium **(Figure 3D)** and infiltrated by lymphocytes and macrophages (**Figure 3B-G**). Our results support a critical role for myoepithelial cells in DED, and complement previous studies, which have shown that myoepithelial differentiation and contractile function is altered in Sjogren's patients (31, 32). It will be important to determine whether the phenotypes of *Adgra3*$^{cre/cre}$ and *Adgra3*$^{lz/lz}$ mice arise from defective myoepithelial adhesion, premature progenitor exhaustion, or impaired myoepithelial differentiation and matrix secretion.

Our mouse genetic results raise the question whether Gpr125 and its downstream pathways are involved in the etiology of human DED. If so, then *Adgra3*-KO mice will provide a powerful tool to study the pathological mechanism and progression of DED and to test therapeutic strategies to cure or manage DED. GPCRs are attractive candidates

for drug development and currently are targets of approximately 34% of drugs approved by the US Food and Drug Administration. Though less well understood, recent discoveries of involvement of aGPCRs in human disease has produced a growing interest in their therapeutic potential (33). Although Gpr125 is an orphan receptor evidence has emerged recently suggesting a role in receptor trafficking and endosomal signaling (28). Deciphering Gpr125 signaling pathways holds promise to uncover novel targets for therapeutic intervention in this common condition.

Methods: Refer to Supplemental Methods for details.

Study Approval: Animal experiments were approved by NYUMC institutional animal care and use committee and conformed to American Association for Accreditation of Laboratory Animal Care guidelines.

Statistics: Experimental data are presented as mean ± SEM. P values for experiments comparing two groups were calculated using student's t test. For experiments comparing more than two groups, an Ordinary one-way ANOVA was used with multiple comparisons test. $P<0.05$ was considered statistically significant.

Author Contributions: PC conceived the study, designed experiments and wrote the paper with input from ES and RB; ES RB JS AI MF and AS conducted the experiments.

Acknowledgements: This work was supported by Department of Defense W81XWH-17-1-0013 BC160959 (PC), NIH training grants: T32GM066704 (JS), T32CA009161-44 (AI) and The Susan G Komen Foundation For The Cure (AI). We thanks Applied Bioinformatics, Experimental Pathology Research and Microscopy Laboratory Cores for the service provided: Grants P30CA016087, S10 OD021747 and P30CA016087. We thank Dr. Aris Economides, Regeneron, Tarrytown NY for advice on *Adgra3*creER^{T2} construction and Dr. Michael Brown DVM (Oradell Veterinary Center, NJ) for ophthalmological evaluation.

REFERENCES

1. Botelho SY. Tears and the Lacrimal Gland. *Sci Am.* 1964;211:78-86.
2. Schaumberg DA, Sullivan DA, and Dana MR. Epidemiology of dry eye syndrome. *Adv Exp Med Biol.* 2002;506(Pt B):989-98.
3. Schaumberg DA, Sullivan DA, Buring JE, and Dana MR. Prevalence of dry eye syndrome among US women. *Am J Ophthalmol.* 2003;136(2):318-26.
4. Pflugfelder SC, and de Paiva CS. The Pathophysiology of Dry Eye Disease: What We Know and Future Directions for Research. *Ophthalmology.* 2017;124(11S):S4-S13.
5. Gipson IK. Goblet cells of the conjunctiva: A review of recent findings. *Prog Retin Eye Res.* 2016;54:49-63.
6. Bron AJ, and Tiffany JM. The meibomian glands and tear film lipids. Structure, function, and control. *Adv Exp Med Biol.* 1998;438:281-95.
7. Makarenkova HP, Ito M, Govindarajan V, Faber SC, Sun L, McMahon G, et al. FGF10 is an inducer and Pax6 a competence factor for lacrimal gland development. *Development.* 2000;127(12):2563-72.
8. Dean C, Ito M, Makarenkova HP, Faber SC, and Lang RA. Bmp7 regulates branching morphogenesis of the lacrimal gland by promoting mesenchymal proliferation and condensation. *Development.* 2004;131(17):4155-65.
9. Dean CH, Miller LA, Smith AN, Dufort D, Lang RA, and Niswander LA. Canonical Wnt signaling negatively regulates branching morphogenesis of the lung and lacrimal gland. *Dev Biol.* 2005;286(1):270-86.
10. Farmer DT, Nathan S, Finley JK, Shengyang Yu K, Emmerson E, Byrnes LE, et al. Defining epithelial cell dynamics and lineage relationships in the developing lacrimal gland. *Development.* 2017;144(13):2517-28.
11. Nien CJ, Massei S, Lin G, Liu H, Paugh JR, Liu CY, et al. The development of meibomian glands in mice. *Mol Vis.* 2010;16:1132-40.
12. Rios JD, Forde K, Diebold Y, Lightman J, Zieske JD, and Dartt DA. Development of conjunctival goblet cells and their neuroreceptor subtype expression. *Invest Ophthalmol Vis Sci.* 2000;41(8):2127-37.
13. Tsau C, Ito M, Gromova A, Hoffman MP, Meech R, and Makarenkova HP. Barx2 and Fgf10 regulate ocular glands branching morphogenesis by controlling extracellular matrix remodeling. *Development.* 2011;138(15):3307-17.
14. Chen Z, Huang J, Liu Y, Dattilo LK, Huh SH, Ornitz D, et al. FGF signaling activates a Sox9-Sox10 pathway for the formation and branching morphogenesis of mouse ocular glands. *Development.* 2014;141(13):2691-701.
15. Kenchegowda D, Swamynathan S, Gupta D, Wan H, Whitsett J, and Swamynathan SK. Conditional disruption of mouse Klf5 results in defective eyelids with malformed meibomian glands, abnormal cornea and loss of conjunctival goblet cells. *Dev Biol.* 2011;356(1):5-18.
16. Marko CK, Menon BB, Chen G, Whitsett JA, Clevers H, and Gipson IK. Spdef null mice lack conjunctival goblet cells and provide a model of dry eye. *Am J Pathol.* 2013;183(1):35-48.

17. Tong L, and Gupta PK. Need for Animal Models of Meibomian Gland Dysfunction. *Ophthalmol Ther.* 2016;5(2):129-34.
18. Plikus M, Wang WP, Liu J, Wang X, Jiang TX, and Chuong CM. Morpho-regulation of ectodermal organs: integument pathology and phenotypic variations in K14-Noggin engineered mice through modulation of bone morphogenic protein pathway. *Am J Pathol.* 2004;164(3):1099-114.
19. Cui CY, Smith JA, Schlessinger D, and Chan CC. X-linked anhidrotic ectodermal dysplasia disruption yields a mouse model for ocular surface disease and resultant blindness. *Am J Pathol.* 2005;167(1):89-95.
20. McMahon A, Lu H, and Butovich IA. A role for ELOVL4 in the mouse meibomian gland and sebocyte cell biology. *Invest Ophthalmol Vis Sci.* 2014;55(5):2832-40.
21. Bjarnadottir TK, Gloriam DE, Hellstrand SH, Kristiansson H, Fredriksson R, and Schioth HB. Comprehensive repertoire and phylogenetic analysis of the G protein-coupled receptors in human and mouse. *Genomics.* 2006;88(3):263-73.
22. Simundza J, and Cowin P. Adhesion G-protein-coupled receptors: elusive hybrids come of age. *Cell Commun Adhes.* 2013;20(6):213-26.
23. Seandel M, James D, Shmelkov SV, Falciatori I, Kim J, Chavala S, et al. Generation of functional multipotent adult stem cells from GPR125+ germline progenitors. *Nature.* 2007;449(7160):346-50.
24. Pickering C, Hagglund M, Szmydynger-Chodobska J, Marques F, Palha JA, Waller L, et al. The Adhesion GPCR GPR125 is specifically expressed in the choroid plexus and is upregulated following brain injury. *BMC Neurosci.* 2008;9:97.
25. Wu Y, Chen W, Gong L, Ke C, Wang H, and Cai Y. Elevated G-Protein Receptor 125 (GPR125) Expression Predicts Good Outcomes in Colorectal Cancer and Inhibits Wnt/beta-Catenin Signaling Pathway. *Med Sci Monit.* 2018;24:6608-16.
26. Fu JF, Yen TH, Chen Y, Huang YJ, Hsu CL, Liang DC, et al. Involvement of Gpr125 in the myeloid sarcoma formation induced by cooperating MLL/AF10(OM-LZ) and oncogenic KRAS in a mouse bone marrow transplantation model. *Int J Cancer.* 2013;133(8):1792-802.
27. Li X, Roszko I, Sepich DS, Ni M, Hamm HE, Marlow FL, et al. Gpr125 modulates Dishevelled distribution and planar cell polarity signaling. *Development.* 2013;140(14):3028-39.
28. Spiess K, Bagger SO, Torz LJ, Jensen KHR, Walser AL, Kvam JM, et al. Arrestin-independent constitutive endocytosis of GPR125/ADGRA3. *Ann N Y Acad Sci.* 2019;1456(1):186-99.
29. Pflugfelder SC, Bian F, Gumus K, Farley W, Stern ME, and De Paiva CS. Severity of Sjogren's Syndrome Keratoconjunctivitis Sicca Increases with Increased Percentage of Conjunctival Antigen-Presenting Cells. *Int J Mol Sci.* 2018;19(9).
30. Cotsarelis G, Sun TT, and Lavker RM. Label-retaining cells reside in the bulge of the pilosebaceous unit: implications for follicular stem cells, hair cycle, and skin carcinogenesis. *Cell.* 1990;61:1329-37.
31. Hawley D, Tang X, Zyrianova T, Shah M, Janga S, Letourneau A, et al. Myoepithelial cell-driven acini contraction in response to oxytocin receptor stimulation is impaired in lacrimal glands of Sjogren's syndrome animal models. *Sci Rep.* 2018;8(1):9919.

32. Makarenkova HP, and Dartt DA. Myoepithelial Cells: Their Origin and Function in Lacrimal Gland Morphogenesis, Homeostasis, and Repair. *Curr Mol Biol Rep.* 2015;1(3):115-23.
33. Bassilana F, Nash M, and Ludwig MG. Adhesion G protein-coupled receptors: opportunities for drug discovery. *Nat Rev Drug Discov.* 2019;18(11):869-84.

SUPPLEMENT

MATERIALS AND METHODS

Mice: Mice were constructed by Ingenious Technologies, Ronkonkoma, NY as follows. A cassette containing *CreER*^{T2} followed by a 3' polyadenylation signal, harboring SV40-driven Neo flanked by FRT sites inserted in a central intron, was recombined into a bacterial artificial chromosome (BAC) to place *CreER*^{T2} under the control of the *Adgra3* promoter, excising 502 bp encompassing 221 bp of exon 1 and part of the following intron 1-2 of *Adgra3*. Mice generated from these ES cells were selected for germline transmission by PCR, verified by southern analysis and sequencing then bred to a Flp deleter strain to remove Neo. *Adgra3*$^{lz/+}$ mice were generated by Regeneron using VelociGene methods (Valenzuela, D.M. et al. High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. Nat Biotechnol 21, 652-659 (2003) to modify a bacterial artificial chromosome (BAC) clone carrying the mouse *Adgra3* gene by replacement of sequence encompassing exons 16-19 with *lacZ* to produce expression of fusion protein comprising the N-terminal extracellular domain, the first transmembrane domain, and part of the first intracellular loop of Gpr125 fused to β-galactosidase (Figure 1A) (23).

Ophthalmologic examination: Standard ophthalmic examination was performed by a trained veterinary ophthalmology consultant (Dr. Michael Brown, Oradell Veterinary Center, New Jersey). Slit lamp biomicroscopy was used to assess the cornea, anterior

chamber, iris, lens, and vitreous humor. Mydriasis was induced with tropicamide and the retina was examined via indirect ophthalmoscopy.

Corneal fluorescein staining was performed by applying sodium fluorescein (1%), for 3 minutes to the cornea of mice. Excess fluorescein was removed by flushing with sterile phosphate buffered saline (PBS) and corneal staining was evaluated and photographed with a slit lamp biomicroscope (Humphrey-Zeiss, Dublin, CA) using a cobalt bluelight. Punctate staining was recorded using a standardized National Eye Institute grading system of 0 to 3 for each of the five areas of the cornea.

Schirmer Tear Test: Tear production was measured via a modified Schirmer Tear Test. Briefly, 35mm x 5mm wide commercial Schirmer Tear Test standardized sterile strips (Schirmer Tear Test; Merck Animal Health) were transected with into two 15mm x 2.5mm strips, with the top notch removed. Individual strips were placed under the lower eyelid using forceps and removed after 15 seconds. The length of dye migration and wetting of the strip was measured in millimeters under a dissecting microscope.

Intraocular pressure measurement

Mice were anesthetized and maintained on isoflurane through a nose cone. IOPs were measured using a TonoLab rebound tonometer (Icare, Finland) within 5 min after isoflurane gas anesthesia induction. For every 6 valid measurements, the highest and lowest IOP values were automatically excluded by the device, and the average of the remaining 4 IOP values was displayed along with the deviation. For quality control, only

averages with slight deviation of less than 2.5 mmHg were considered acceptable readings. This procedure was repeated at least 3 times for each eye, and the acceptable readings were averaged IOP was measured eighteen times for each eye, and the average value was used for final analysis.

Histological Analysis: The exorbital lacrimal gland, salivary and parotid glands, and whole globes were removed from mice and fixed with either 10% neutral buffered formalin or 4% paraformaldehyde (PFA) and embedded in paraffin. For general histological assessment, sections were stained with hematoxylin and eosin (H&E), or with periodic acid-Schiff (PAS) and Alcian Blue to visualize conjunctival goblet cells. Goblet cells in the bulbar and palpebral conjunctiva were quantified by two separate readers. Serial sections of tissues were stained with antibodies for anti-CD4, anti-CD8, anti-F480, anti-cytokeratin CK5 optimized by the Experimental Histology Core, NYUMC for analysis by Akoya/PerkinElmer Vectra® multispectral imaging system then counterstained with Dapi.

Lineage Tracing: For lineage tracing experiments, *Adgra3*^{creERT2} mice were crossed to the fluorescent Rosa26R-lox.STOP.lox-tdTomato lineage reporter strain (Stock No. 007909) Jackson laboratory. The transcriptional STOP was deleted by cre recombination during embryonic development (E14.5-E15.5) by delivering tamoxifen (5mg per mouse-2 doses of 2.5mg) by oral gavage to *Adgar3*$^{cre/cre}$ dams with during mid-pregnancy. Pups were delivered at E19.5-E20.5 by caesarian section to avoid problems with delivery caused by Tamoxifen and fostered by SWR/J mice. Their

tissues were harvested at 7 weeks and at 6 months to test for progenitor potency and longevity.

Tissue clearing and 3-D imaging: Lacrimal glands were excised and fixed overnight in 4% PFA then processed using a modified CUBIC (Reagent 1A) protocol Single-cell lineage tracing in the mammary gland reveals stochastic clonal dispersion of stem/progenitor cell progeny. (Davis FM et al. 2016 doi: 10.1038/ncomms13053. PMID:27779190.) Tissue was incubated in CUBIC Reagent 1A clearing solution for 4 days, rinsed 3X in PBS then immunostained for 4 days at 4C in PBST containing 10% rabbit serum and rabbit anti-K5 (Covance, PRB160P, 1:100), rinsed again then 2 days in goat anti-rabbit AlexaFluor (AF) 647 (Thermo Fisher Scientific, A21245, lot number 1805235, 1:500), rinsed 3X then cleared in CUBIC R2 for 24hrs. Cleared lacrimal tissues were imaged using a Zeiss 880 Laser Scanning inverted confocal microscope with 20X air Plan-Apochromat N.A. 0.8 M27 objective lenses.

X-gal staining: Embryos, Eyes and lacrimal glands were fixed in 4% paraformaldehyde (PFA) (Sigma-Aldrich) at room temperature (RT) for 30-60 min, rinsed 3X in X-gal rinse buffer (2 mM MgCl2, 0.1% Sodium deoxycholate, and 0.2% NP-40 in PBS) at RT, then incubated in X-gal staining solution (50 mg/ml 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside in rinse buffer containing 5 mM potassium ferricyanide, 5 mM potassium ferrocyanide) (Applichem, Cheshire, CT) at RT overnight. After staining, glands were rinsed in PBS, post-fixed in 4% PFA overnight then prepared for whole mount analysis or processed for paraffin embedding and sectioned for histological analysis.

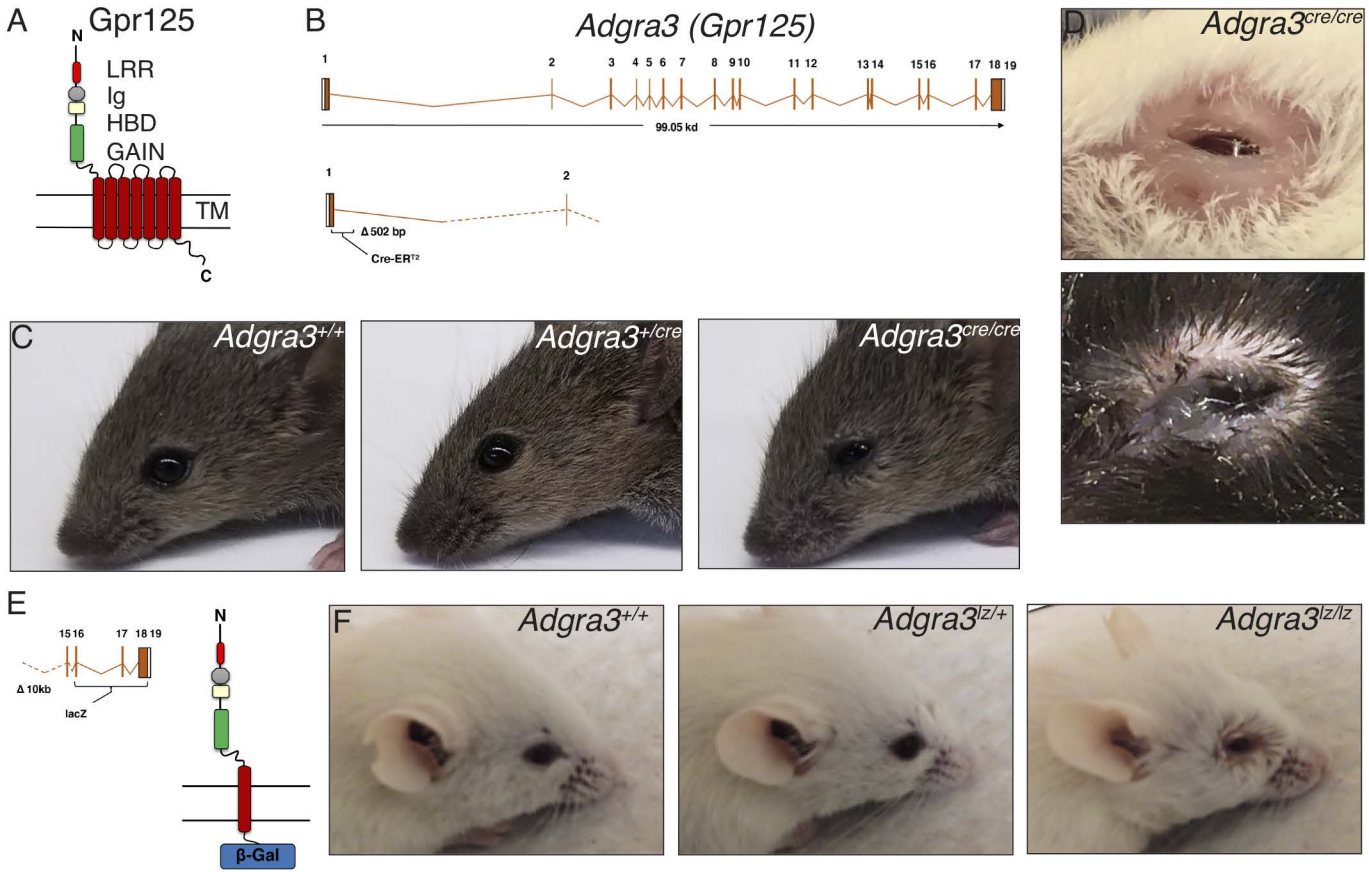


Figure 1. Gpr125 loss induces blepharedema, blepharitis and mucus accumulation.
(A) Schematic of Gpr125 protein comprising N-terminus (N), leucine rich repeats (LRR), Immunoglobulin-like domain (Ig), hormone binding domain (HBD), GPCR autoproteolyis-inducing (GAIN) domain, 7-pass transmembrane region (TM) and cytoplasmic region (C). **(B)** Schematic of *Adgra3*. *Adgra3*$^{cre/cre}$ mice were generated by replacement of 502bp after the first codon with a Cre-ER^{T2} module. **(C)** Eye phenotype of *Adgra3*$^{cre/cre}$ and **(F)** *Adgra3*$^{lz/lz}$ mice compared to controls. **(D)** Examples of blepharedema and mucus accumulation in *Adgra3*$^{cre/cre}$ mice. **(E)** Schematic of the Gpr125-beta-gal protein generated by deletion of 10 kb sequence downstream of the first TM and replacement by *lacZ*.

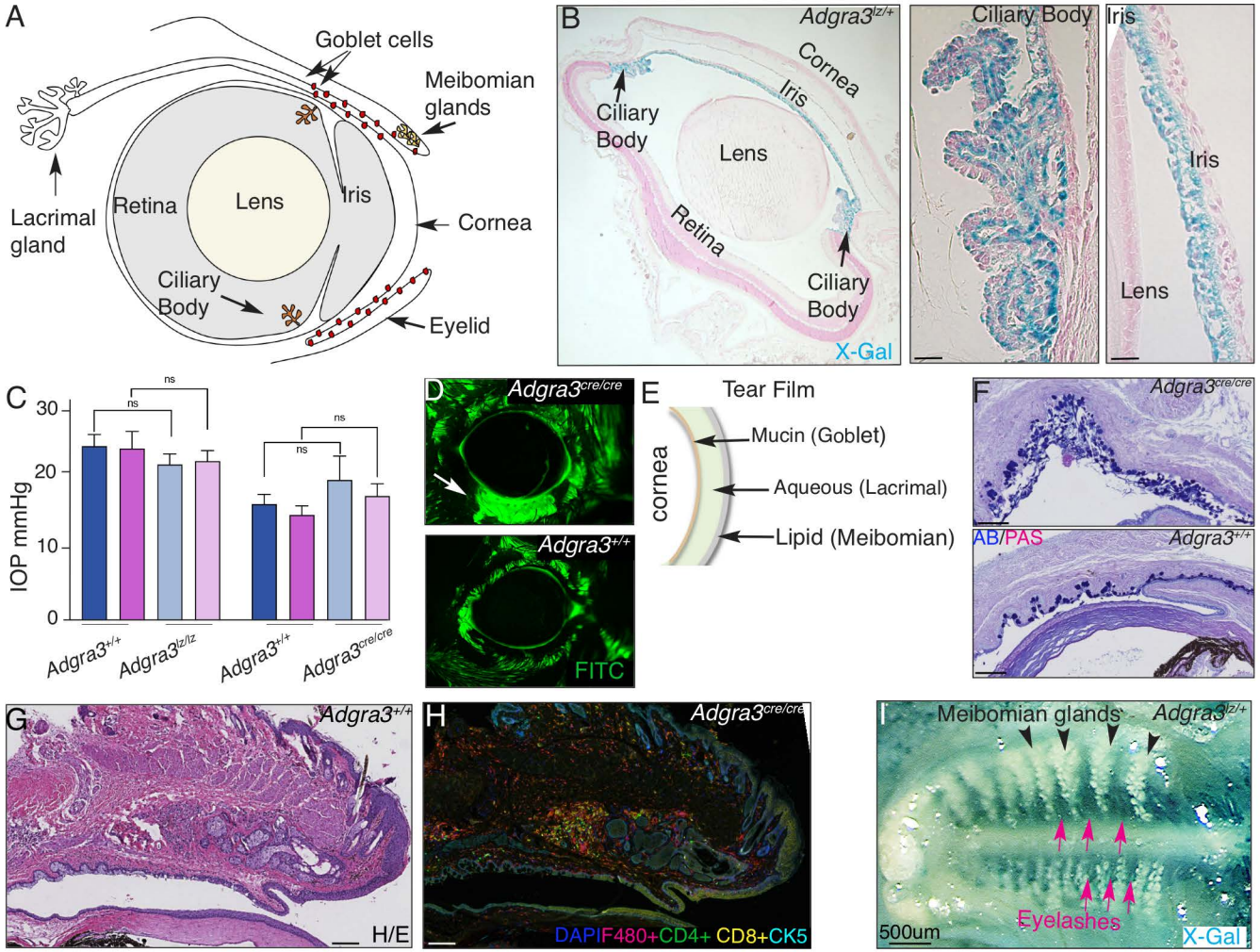


Figure 2. Gpr125 is expressed in eyes and eyelids.

(A) Diagram of murine eye. (**B**) Section of X-gal stained *Adgra3*$^{lz/lz}$ eye shows Gpr125-beta-gal expression in the ciliary body and iris. **(C)** Intraocular pressure (IOP) (mmHg) in male (blue) and female (pink) *Adgra3*$^{lz/lz}$ and *Adgra3*$^{cre/cre}$ mice compared to their respective FVBN and B6 controls. Each bar represents the mean ± SEM on 6-14 mice/group. ns, not significant. **(D)** Fluorescein stained corneas in *Adgra3*$^{cre/cre}$ and control mice.n=3 (**E**) Schematic of tear film. **(F)** Eyelid sections stained with alcian blue/PAS show goblet cells in *Adgra3*$^{cre/cre}$ and control mice.n=23. **(G-H)** H/E show meibomian glands and **(H)** immunostained with antibodies: F480, CD4,CD8, CK5 and DAPI to detect macrophages, T-helper, cytotoxic T cells, cytokeratin 5 and nuclei respectively in *Adgra3*$^{cre/cre}$ mice. Control *Adgra3*$^{+/+}$ in S1A. **I)** X-Gal stained whole mounts of P10 eyelids from *Adgra3*$^{lz/lz}$ mice show meibomian glands (arrowheads) are devoid of Gpr125. Scale bar.100μm. n=3.

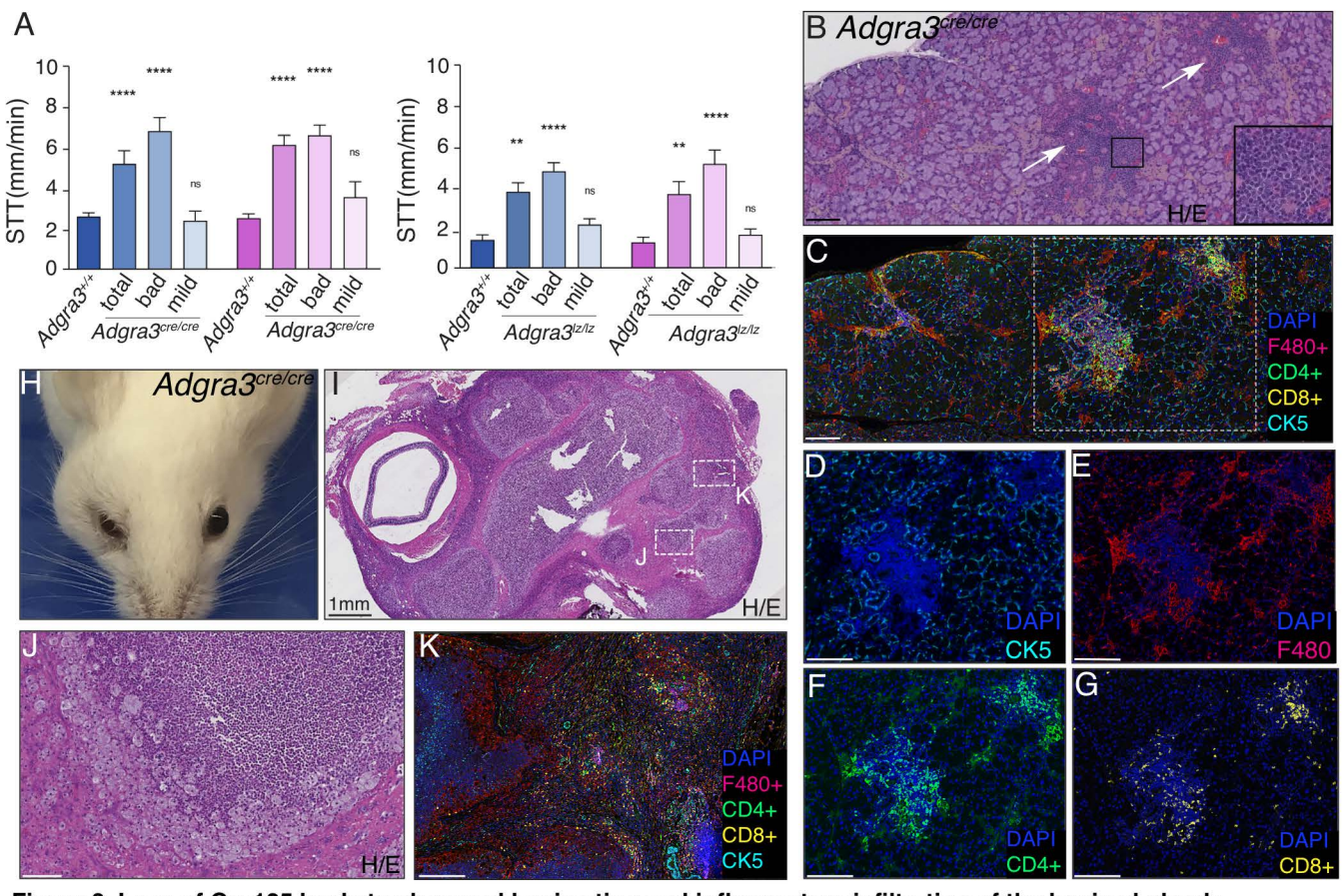


Figure 3. Loss of Gpr125 leads to abnormal lacrimation and inflammatory infiltration of the lacrimal glands

(A) Increased tear production observed in *Adgra3*$^{cre/cre}$ and *Adgra3*$^{lz/lz}$ male (blue) and female (pink) mice compared to controls. Each bar represents the mean ± SEM on 6-32 mice. **** $p<0.0001$, **, $p<0.05$ value significant; ns, not significant. **(B)** H/E section of lacrimal gland from *Adgra3*$^{cre/cre}$ mice shows foci of infiltration (arrows). Control *Adgra3*$^{+/+}$ in S1C. **(C-G)** Immunofluorescence of lacrimal gland co-stained for **(D)** CK5, **(E)** macrophages, **(F)** T-helper, **(G)** cytotoxic T cells. **(H)** *Adgra3*$^{cre/cre}$ female with lacrimal mass H/E stained in **(I,J)**.**(K)** Immunofluorescence analysis of boxed region in **I**. Scale bar 100μm.

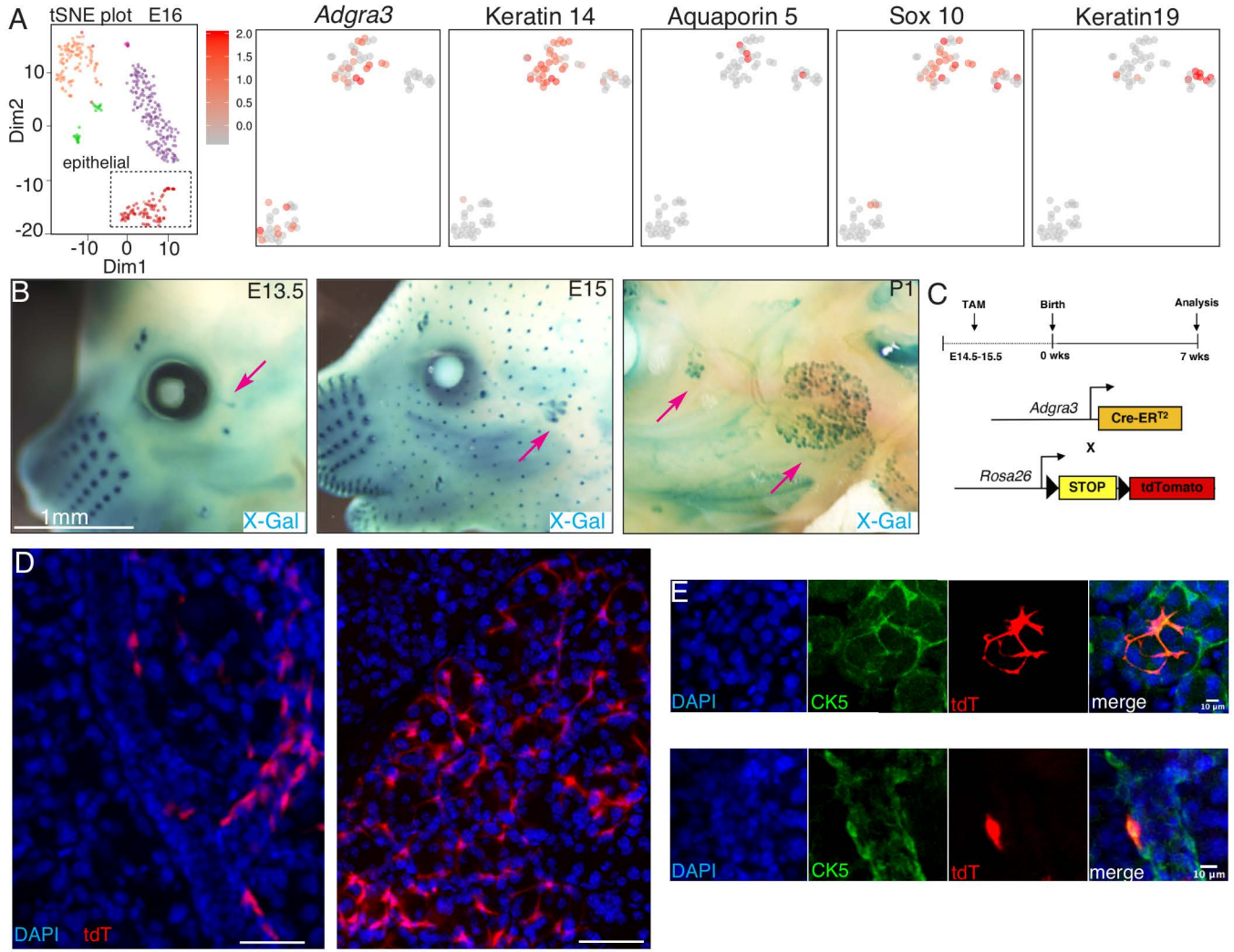


Figure 4. Gpr125 cells, located at ductal tips during development, function as lacrimal myoepithelial progenitors.
(A) t-SNE plot of cells clusters within E16 lacrimal glands (10). Zoomed images of E16 epithelial compartment (boxed region) show cells expressing *Adgra3* mRNA also express myoepithelial markers, Keratin14 and Sox10 but not luminal markers Keratin19 or Aquaporin5. **(B)** Gpr125-beta-gal expression in embryos. **(C)** Strategy for tracing the lineage of Gpr125-positive cells in E14.5-E15.5 embryos carrying the *Rosa26*.lox.STOP.lox.TdTomato reporter by tamoxifen injection of pregnant *Adgra3*$^{cre/cre}$ dams. 3D-confocal images of lacrimal glands from mice at **(D)** 7 weeks and **(E)** 6 months showing tdT expression in elongated myoepithelial cells along the basal border of ducts and stellate cells enmeshing acini colocalized with myoepithelial marker (CK5). Scale bar 50μm.n=3.

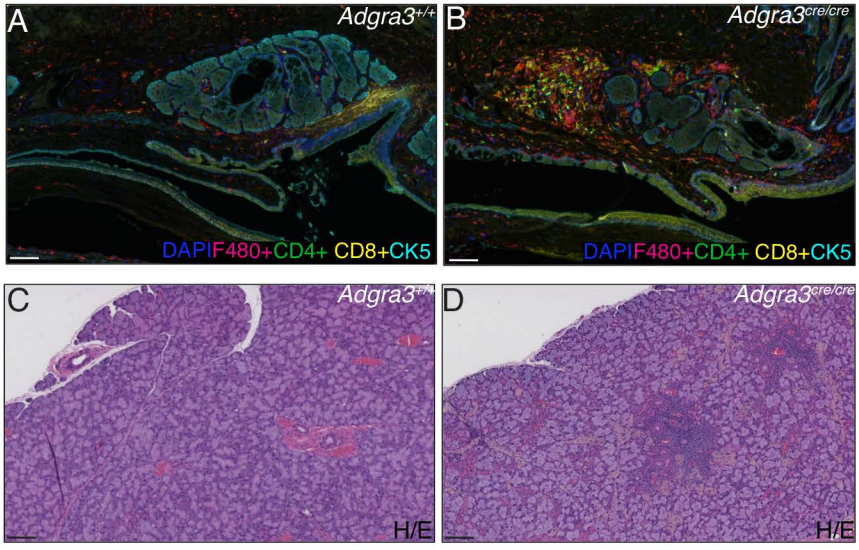


Figure S1.

(A,B) Meibomian gland immunostained with antibodies: F480, CD4, CD8, CK5 and DAPI to detect macrophages, T-helper, cytotoxic T cells, cytokeratin 5 and nuclei respectively in control *Adgra3*$^{+/+}$ and *Adgra3*$^{cre/cre}$ mice.

(C,D) H/E section of lacrimal gland from control *Adgra3*$^{+/+}$ and *Adgra3*$^{cre/cre}$ mice. Scale bar 100μm.