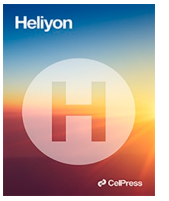
Heliyon 9 (2023) e17842

Contents lists available at ScienceDirect

Heliyon

journal homepage: www.cell.com/heliyon



# Cell type and stage specific transcriptional, chromatin and cell-cell communication landscapes in the mammary gland

Yura Song [a,1], Marco Fioramonti [a,1], Gaëlle Bouvencourt [a], Christine Dubois [a], Cédric Blanpain [a,b], Alexandra Van Keymeulen [a,*]

[a] Laboratory of Stem Cells and Cancer, Université Libre de Bruxelles (ULB), Brussels, Belgium
[b] WELBIO, Université Libre de Bruxelles (ULB), Brussels, Belgium

ARTICLE INFO

*Keywords:*
Mammary gland
Cell lineage
Cell identity
Gene regulatory network
Life stages

ABSTRACT

The mammary gland (MG) is composed of three main epithelial lineages, the basal cells (BC), the estrogen receptor (ER) positive luminal cells (ER+ LC), and the ER negative LC (ER− LC). Defining the cell identity of each lineage and how it is modulated throughout the different stages of life is important to understand how these cells function and communicate throughout life. Here, we used transgenic mice specifically labelling ER+ LC combined to cell surface markers to isolate with high purity the 3 distinct cell lineages of the mammary gland and defined their expression profiles and chromatin landscapes by performing bulk RNAseq and ATACseq of these isolated populations in puberty, adulthood and mid-pregnancy. Our analysis identified conserved genes, ligands and transcription factor (TF) associated with a specific lineage throughout life as well as genes, ligands and TFs specific for a particular stage of the MG. In summary, our study identified genes and TF network associated with the identity, function and cell-cell communication of the different epithelial lineages of the MG at different stages of life.

## 1. Introduction

The MG is essential for the survival of young mammalian offspring by the secretion of the milk that provide them fluids and nutriments during the first part of their life. The MG is composed of an epithelium organized as a ductal tree of basal cells (BC) surrounding luminal cells (LC). LC contained ER+ LC, which respond to the estrogen and progesterone signalling and ER− LC that are responsible for milk production during pregnancy and lactation. Mammary epithelial cells are surrounded by a unique microenvironment, described as the mammary fat pad, which is composed of adipocytes, fibroblasts, blood vessels and immune cells.

The MG development begins early during embryonic development and is the first skin appendage that occurs during morphogenesis. Depending on the species, one or multiple pairs of mammary placodes are specified in the ventral part of the skin that further develop until birth into a primitive rudimentary tree. At the onset of puberty, the MG expands until it completely fills the mammary fat pad. During pregnancy, the MG further expands and ER− LC differentiate into alveolar structures that will produce milk. BC have contractile properties which help to expulse the milk throughout the ductal tree and are called for this reason myoepithelial cells. At the end of lactation, due to the drop of prolactin concentration, the MG involutes and regresses to an appearance similar to the one

* Corresponding author.
*E-mail address:* alexandra.van.keymeulen@ulb.be (A. Van Keymeulen).
[1] These authors contributed equally.

https://doi.org/10.1016/j.heliyon.2023.e17842
Received 2 March 2023; Received in revised form 20 June 2023; Accepted 29 June 2023
Available online 30 June 2023

before pregnancy [1].

The estrogen and progesterone ovarian hormones are key regulators of the development and function of the mammary gland. Levels of estrogen and progesterone increase during puberty and promote MG outgrowth. In adulthood, hormonal levels control the menstrual cycle. Burst of proliferation and formation of side branches are thought to be triggered by progesterone during the estrus cycles. Estrogen induces modest proliferation in adulthood, but is permissive for the strong mitogenic effects of progesterone. Upon pregnancy, estrogen and progesterone induce a massive proliferation and morphological changes to produce milk-producing alveoli [2–4]. ER+ LCs are regulated by hormones secreted by the ovary and the pituitary gland, and relay the signals to ER− LC and BC through paracrine signals like amphiregulin or RANKL [5]. Crosstalk between hormone sensing cells and ER-PR-cells is crucial to coordinate the development of a functional organ but is still not fully understood.

Genetic lineage tracing and clonal analysis of BC, LC, ER− LC, ER+ LC and neutral genetic labelling in the MG demonstrated that BC are multipotent during embryonic development, giving rise to BC and LC, but that after birth, during puberty, adulthood and pregnancy cycles, BC, ER− LC and ER+ LC are lineage-restricted and self-sustained lineages [6–12].

The cellular heterogeneity and dynamics of mammary gland cell types during different stages of life have been previously studied by single cell RNA sequencing studies in mice and in human [8,13–21]. While single cell RNA sequencing is the gold standard technique to study the cellular heterogeneity within an organ, more robust expression profiles can be achieved with bulk RNA sequencing, pending cell population isolation is possible. Several transcriptome analysis have been previously performed in the MG. Transcriptome analysis have previously described the differences between BC and LC in mice [22–25] or human [26–28] but these studies did not separate the 2 subpopulations of LC, ER+ LC and ER− LC, which have very distinct functions. Expression profiles of BC, progenitors LC and mature LC, as defined by their stem cell properties, in mice and/or human were also performed during adulthood only [29–33]. Expression profiles of BC, ER+ LC and ER− LC, isolated based on cell surface markers, was performed during adulthood only [32], and expression profiles of ER+ LC were performed in virgin or early pregnancy, but were not compared to the expression profiles of ER− LC and BC [34]. Therefore, the transcriptome of the 3 cell subpopulations in parallel have only been investigated in adulthood by bulk sequencing, leaving open questions on how the transcriptome and chromatin landscape of these different lineages are regulated at different stages of life, what are their conserved signatures throughout life and how the different cell types differ during puberty, adulthood and pregnancy.

In this study, we used transgenic mice specifically labelling ER+ LC, previously fully characterized [9], combined with cell surface markers, to isolate with high purity the 3 distinct cell lineages of the mammary gland and defined their expression profiles and chromatin landscapes by performing bulk RNAseq and ATACseq of these isolated populations during puberty, adulthood and pregnancy. We established conserved and stage specific transcriptional and chromatin signatures of the different epithelial mammary cell types. We defined conserved or temporal specific genes, ligands and TF networks within a given epithelial lineage across puberty, adulthood or pregnancy. Our approach allowed to profile the rare ER+ LC during pregnancy and uncover genes associated with this specific population during pregnancy. Altogether, our study constitutes a useful resource for researchers working on mammary gland development and stem cell regulation that allows to define new candidate genes, cell-cell communications, and gene regulatory network (GRN) that are controlling cell identity and function of the different MG epithelial lineages across different stages of life.

## 2. Methods

**Mice.** RosaYFP [74] mice were obtained from the Jackson Laboratory. TetOCre [75] mice were provided by A. Nagy. ER-rtTA generation was previously described [9]. Mice colonies from mixed background were maintained in a certified animal facility in accordance with European guidelines. These experiments were approved by the local ethical committee (CEBEA).

**Targeting YFP expression.** For puberty, 3 weeks old ER-rtTA/TetOCre/RosaYFP female mice were induced with doxycycline diluted in drinking water (2 g/l, AG Scientific) and 3 intraperitoneal injections per week (200 µl of 10 mg/ml doxycycline diluted in PBS) during 14 days and analyzed at the end of induction, at 5 weeks old. For adult and pregnancy, 4 weeks old ER-rtTA/TetOCre/RosaYFP female mice were induced with doxycycline diluted in drinking water (2 g/l, AG Scientific) and 3 intraperitoneal injections per week (200 µl of 10 mg/ml doxycycline diluted in PBS) during 28 days. Mice for adult condition were processed at 10 weeks old. Mice for pregnancy condition were put in mating when 10 weeks old and processed when 12 days pregnant.

**Mammary cell preparation.** MGs were dissected and the lymph nodes removed before processing. 12 to 22 mice were used per sample for puberty, 1 mice was used per sample for adult, and 3 to 4 mice were used per sample for pregnancy. Samples were washed in HBSS and cut in pieces of 1 $mm^3$ with scissors. Samples were digested for 2 h at 37 °C under shaking in 300U/ml collagenase (Sigma)/ 300 µg/ml hyaluronidase (Sigma) in HBSS. EDTA at a final concentration of 5 mM was added for 10 min to the resultant organoid suspension, followed by 0.25% Trypsin/EGTA for 2 min. Samples were then filtrated through 40-µm mesh and rinsed in 2% FBS/PBS.

**Cell labelling and flow cytometry.** All steps of cell labelling were performed in PBS supplemented with 2% bovine serum. Cells were incubated in 250 µl per million cells in primary antibody dilution for 30 min on ice, with shaking every 10 min. Primary antibodies were washed and cells incubated with secondary antibodies, with shaking every 10 min. Secondary antibodies were washed and cells were resuspended in 2.5 µg/ml DAPI (Invitrogen) before analysis.

Primary antibodies used were: APC-conjugated anti-CD45 (1/100, clone 30-F11, 17-0451, eBiosciences), APC-conjugated anti-CD31 (1/100, clone 390, 17-0311, eBiosciences), APC-conjugated anti-CD104a (1/100, clone APA5, 17-1401, eBiosciences), PECy7-conjugated anti-CD24 (1/100, clone M1/69, 560535, BD Biosciences), AlexaFluor700-conjugated anti-CD29 (1/100, clone HMβ1-1, 102218, Biolegend), PE-conjugated anti-CD49b (1/100, clone DX5, 553858, BD Biosciences), PerCP/Cy5.5-conjugated anti-Sca1 (1/100, clone D7, 108124, Biolegend), Biotin-conjugated anti-CD133 (1/100, clone 13A4, 13-1331, eBiosciences). Secondary antibody used was: APC-Cy7-conjugated streptavidin (1/400, 554063, BD Biosciences). Cell sorting was performed on a FACS Aria III using the

FACS DiVa software (BD Biosciences).

**RNA sequencing.** RNA extraction from 20000 to 100000 FACS isolated cells was performed using RNeasy micro kit (QIAGEN) according to the manufacturer's recommendations. Prior to sequencing the quality of RNA was evaluated by Bioanalyzer 2100 (Agilent). Indexed cDNA libraries were obtained using the Ovation Solo RNA-seq Systems (NuGen) following manufacturer's recommendations. The multiplexed libraries (11pM/18pM) were loaded on flow cells and sequences were produced using a HiSeq PE Cluster Kit v4 and TruSeq SBS Kit v3-HS (250 cycles) on a Novaseq 6000 (Illumina). Reads were mapped against the mouse reference genome (Grcm38/mm10) using STAR software [76] to generate read alignments for each sample. Annotations Grcm38.87 was obtained from ftp.Ensembl.org. After transcripts assembling, gene level counts were obtained using HTseq-count [77] using options -f bam -r name -s no –nonunique all and normalized to 20-millions of aligned reads. Average expression for each gene for the different cell populations was computed based on the 2 biological replicates.

### *2.1. RNAseq analysis and signature derivations*

The average of the 2 replicates for each condition were used to define the expression fold change between 2 conditions. Upregulated genes were defined as genes upregulated for at least 2-fold changes, and normalized expression value is at least 20 in the upregulated condition. When normalized gene expression value was 0 in the condition to compare with, it was manually changed to 0.1, to avoid dividing by zero error when generating fold change table, therefore generating an artificially high fold change instead of an error message of dividing by zero.

Cell type specific signatures were defined as compared to the 2 other cell subtypes and were derived in puberty, in adult and in pregnancy. In puberty, 833 genes were upregulated both in ER+ LC compared to ER− LC and in ER+ LC compared to BC and constitute the ER+ LC specific gene signature in puberty. Similarly, ER− LC specific gene signature and BC specific signature in puberty were derived and were composed respectively of 1184 and 2260 genes. In adult, the ER+ LC specific gene signature was composed of the 751 genes upregulated both in ER+ LC compared to ER− LC and in ER+ LC compared to BC, while the ER− LC and BC specific gene signatures were composed of 828 and 2815 genes respectively. In pregnancy, specific gene signatures were derived similarly and were composed of 652, 471 and 1766 genes respectively for ER+ LC, ER− LC and BC.

Conserved ER+ LC signature throughout the 3 developmental stages was then derived from the intersection of the puberty, adult and pregnancy ER+ LC specific gene signatures. This signature corresponds to the genes upregulated in ER+ LC compared to the 2 other cell types in the 3 stages of life and was composed of 215 genes. Conserved ER− LC and BC signature throughout the 3 stages were derived in a similar way and were composed respectively of 58 and 915 genes.

The puberty specific signature in ER+ LC was defined as genes commonly upregulated in ER+ LC puberty compared to ER+ LC adult, in ER+ LC puberty compared to ER+ LC pregnancy as well as in ER+ LC puberty compared to ER− LC puberty and ER+ LC puberty compared to BC puberty and comprised 146 genes. The adult specific signature in ER+ LC and the pregnancy specific signature in ER+ LC were derived similarly and included 45 and 206 genes respectively. For ER− LC, the puberty, adult and pregnancy signatures comprised 191, 218 and 163 genes respectively, while for BC, the puberty, adult and pregnancy signatures comprised 333, 718 and 160 genes respectively.

Genes signature of each population were tested for enrichment in each gene ontology (GO) class and pathway analysis using the EnrichR [78,79]. The categories used for analysis are GO Biological Process 2018 and KEGG 2019 mouse. Statistically significant enrichment corresponding to those presenting a $P$-value $< 0.05$, although some terms and genes involved in non-statistically significant classes were included for their biological relevance.

The heatmaps for visualization of gene expression were plotted with heatmapper [80]. Average linkage and pearson distance were applied for clustering of genes based on the gene expression.

**Human dataset conversion and comparison**. Transcriptomic datasets of human breast were obtained from previous studies [30,33]. Human genes were converted to equivalent mouse genes with Ensembl Biomart (Ensembl Genes 109) [81].

Among the signatures of each cell type described in *Lim* et al. [33], genes with positive average log fold-change were considered. In the transcriptomic data of *Pellacani* et al. [30], gene signatures of each cell type were defined further for comparison to our mouse datasets. BC signature was defined as the genes up-regulated in both BC vs luminal progenitor (LP) and BC vs LC. LP signature was defined as the genes both down-regulated in BC vs LP and LC vs LP. LC signature was defined as the genes down-regulated on BC vs LC and up-regulated on LC vs LP.

Basal/stem progenitor from *Lim* et al. and BC from *Pellacani* et al. were compared to conserved BC signature. Luminal progenitor signatures from both datasets were compared to conserved ER− LC. Mature luminal from *Lim* et al. and LC from *Pellacani* et al. were compared to conserved ER+ LC signature.

**ATAC-sequencing.** For ATAC sequencing, 100000 sorted cells from different cell lineages were collected in 1 ml of PBS supplemented with 3% FBS at 4 °C. Cells were centrifuged and cell pellets were resuspended in 100 µl of lysis buffer (TrisHCl 10 mM, NaCl 10 mM, MgCl2 3 mM, Igepal 0.1%) and centrifuged at 500 g for 25 min at 4 °C. Supernatant was carefully discarded and nuclei were resuspended in 50 µl of reaction buffer (Tn5 transposase 2.5 µl, TD buffer 22.5 µl, from Nextera DNA sample preparation kit, Illumina, and 25 µl $H_20$). The reaction was performed at 37 °C for 30 min and was stopped by addition of 5 µl of clean up buffer (NaCl 900 mM, EDTA 300 mM). DNA was purified using the MiniElute purification kit (QIAGEN) following manufacturer protocol. DNA libraries were PCR amplified (Nextera DNA Sample Preparation Kit, Illumina), and size selected from 200 to 800 bp (BluePippin, Sage Sciences), following manufacturer's recommendations.

**ATAC-sequencing analysis** Before starting the alignment and downstream analysis, quality check was done by FastQC (https://tumor.bioinformatics.babraham.ac.uk/projects/fastqc/). Adaptors sequences and low-quality regions were removed with

Trimmomatic [82] paired-end mode using options "HEADCROP:10 CROP:75 ILLUMINACLIP:adaptor.file:2:30:10 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:60". Trimmed reads were then aligned to mouse reference genome mm10 using Bowtie2 [83] (version 2.2.6) using options "-X 2000 –fr –very-sensitive –no-discordant –no-unal –no-mixed –non-deterministic". Mitochondrial reads, reads from unmapped or random contigs and reads with a mapping quality less than 20 were removed using samtools Duplicate reads were removed by Picard tools (http://broadinstitute.github.io/picard/).

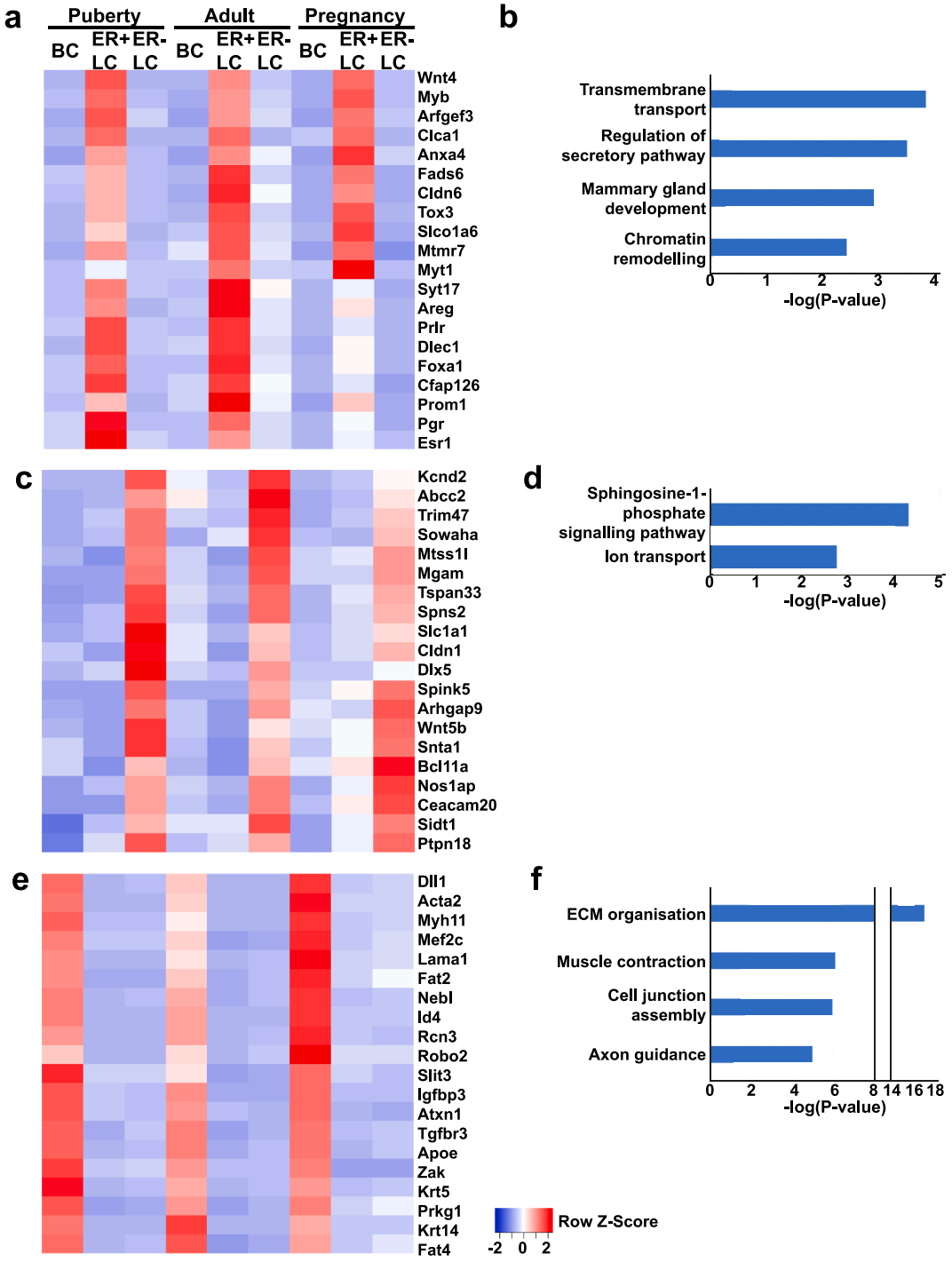



**Fig. 1.** Cell subpopulations of the mammary gland show conserved gene signatures throughout different stages of life. Heatmaps of gene expression levels of examples of genes from conserved signatures throughout different stages of life (a, c, e) and EnrichR gene ontology analysis on conserved signature (b, d, f) for ER+ LC (a, b), ER− LC (c, d) and BC (e, f).

Peak calling was performed on each individual sample by macs2 [84] (version 2.1.0.20151222) using options "-f BAMPE -g mm -q 0.01 –shift 0 –nomodel". Among the peaks from the different subpopulations, the blacklisted peaks were excluded by using bedtools [85] and then merged for downstream analysis.

Reads counts of merged peaks for each individual sample were calculated by HTSeq-count using options "-f bam -r pos -m intersection-nonempty". These counts were globally normalized by using trimmed mean of M values (TMM) [86] in merged peaks and fold-change were calculated between the subpopulations. Peaks were associated to genes with GREAT version 3.0.0 [87] software with the following parameters: 5.0 kb in proximal upstream, 1.0 kb in proximal downstream and 100.0 kb in distal. For most of the analysis, only the peaks annotated to at least one gene were kept. Differential peaks were defined as peaks having at least 2-fold change between two different subpopulations and being called as peaks in the subpopulations which are higher.

De novo motif search was performed using findMotifsGenome.pl program in HOMER [37] (version 4.10) software using parameters "-size −250,250 -bg background.bed -S 15 -len 6,8,10,12." The background peaks are defined as all peaks except regulated peaks. Specific motif research was performed using annotatePeaks.pl program in HOMER software using parameters "-size -250,250" and the motif matrices were obtained via HOMER motif database.

## 3. Results

### 3.1. Common gene signature of MG lineages across the different stages of life

To define the transcriptional profile of ER+ LC, ER− LC and BC at different time points, doxycycline treated ERrtTA/TetOCre/rosaYFP mice, in which most ER+ LC were specifically labelled with YFP, as previously described in details [9], were used to FACS isolate the different epithelial cell subpopulations of the MG at different stages of life, based on the YFP expression combined with expression markers discriminating between the cell types. BC were isolated based on their $Lin^{-}CD24^{+}CD29^{hi}$ profile [6,35]. ER+ LC were isolated based on the $Lin^{-}YFP^{+}CD24^{+}CD29^{lo}Sca1^{+}CD133^{+}$ profile while ER− LC were isolated based on $Lin^{-}YFP^{-}CD24^{+}CD29^{lo}Sca1^{-}CD133^{-}$ profile [9,36]. While Sca1 and CD133 markers enrich for ER+ LC cells, a small fraction of ER− LC is also $Sca1^{+}CD133^{+10}$. Therefore, combining the YFP expression of ER + traced cells with Sca1 and CD133 markers allowed to isolate ER+ LC with a higher purity than previously achieved in mice in puberty (5w old) or in adulthood (10w old). Since Sca1 and CD133 expression is low upon pregnancy [34], the YFP expression was essential to isolate ER + derived LC (YFP+) and ER− LC (YFP−) populations from mice in mid-pregnancy (12 d pregnant) which were respectively isolated based on their $Lin^{-}YFP^{+}CD24^{+}CD29^{lo}$ and $Lin^{-}YFP^{-}CD24^{+}CD29^{lo}$ expression profiles (Supplementary Fig. S1). Transcriptional profiling of the 3 FACS isolated epithelial cell subtypes of the MG were performed at puberty, adulthood and mid pregnancy to derive cell type specific and stage specific gene expression signatures, which were defined as genes upregulated for at least 2-fold changes compared to the 2 other cell subtypes at each time points (Supplementary Tables 1 and 2). The epithelial lineage specific signatures conserved throughout the 3 developmental stages were defined as the intersection of the puberty, adult and pregnancy specific gene signatures (Supplementary Fig. S2).

The conserved ER+ LC signature across the different stages comprised 215 genes including well known markers of ER+ LC like Esr1, Pgr, Prlr, Foxa1, Tbx3, Areg and Prom1 and gene ontology on this signature shows that *trans*-membrane transport and secretion genes were enriched in ER+ LC population. Several genes associated with nerves, like Myt1, Syt17 and Adra1a and genes associated with cilia like Dlec1, Ccdc40 and Cfap126 were upregulated in ER+ LC at all stages (Fig. 1a and b and Supplementary Table 3). Many ligands were expressed at the 3 life stages in ER+ LC, like Wnt4, Insl6, Tgfb1, Areg, Inha, Cxcl15, Npnt, Cmtm8 and Tgfa. While Wnt4 and Insl6 were highly expressed in pregnancy, the other ligands presented a decrease in their expression during pregnancy. This signature shows that a common program and function is conserved through all life stages for ER+ LC but that even if all these ligands are expressed throughout life, their expression profile differs in pregnancy compared to puberty and adulthood, suggesting a different paracrine function for ER+ LC during pregnancy.

The conserved ER− LC signature comprised only 58 genes. As illustrated by the Venn diagram in Fig. S2b, this small number is due to the weak number of genes in common between the pregnancy signature and the 2 others. The number of common genes between puberty and adult is much greater in ER− LC, comprising 324 genes. These numbers of genes in common illustrate that ER− LC are more similar in puberty and adulthood, and that upon pregnancy, ER− LC have a distinct expression program than in puberty and adulthood. Bcl11a and Dlx5 TFs were amongst the conserved ER− LC signature. Gene ontology of the conserved ER− LC signature shows enrichment of genes involved in Sphingosine-1-phosphate signalling pathway, like Spns2 and Spns3, and transport, like Kcnn4 and Snta1 (Fig. 1c and d and Supplementary Table 4). Ligands like Enpp3, Btn1a1, and Cxcl2 were high in ER− LC during the 3 life stages. In conclusion, a very small number of genes are conserved throughout all life stages in ER− LC, reflecting the drastic reprogramming of these cells as they differentiate into alveolar cells during pregnancy.

The conserved BC signature comprised a larger number of genes, 915 genes, reflecting BC are more distinct than the ER+ LC and ER− LC between them and that BC properties are conserved throughout the different stages of life. This signature included well known BC associated genes like Krt5, Krt14, Dll1, Acta2 and Myh11 and gene ontology shows that genes associated with muscle contraction and extracellular matrix organization were enriched in conserved BC signature, as expected. Axon guidance genes like Robo2, Slit2, Slit3 and several semaphoring genes were also specifically upregulated in BC at all stages (Fig. 1e and f and Supplementary Table 5). Trp73 transcription factor was also specifically upregulated in BC at all stages, though at much lower levels than Trp63, and could play a role in BC specification as well. A large number of ligands were also highly expressed in BC in the 3 life stages: Bdnf, Mdk, Plau, Lama3, Igfbp3, Ntng2, Lgals1, Wif1, Igsf10, Hgf, Vegfc, Lamb3, Pianp, Lamc2, Slit3, Angptl2, Pthlh, Nptx2, Ntf5, Sema3b, Nkd2, Fam19a5, Efnb3, Pgf, Adm, Sema3a, Pdgfb, Pdgfc and Csf1. This confirms that BC express a higher number of ligands than the other cell types and could be key mediators of paracrine signalling, as described previously [32].

We compared the conserved signatures we defined in ER+ LC, ER− LC and BC to the previously described ones defined in adulthood in which the 3 cell types were isolated based on Sca1 and CD24 expression [32] and to the ones described in mouse in which cells were separated as luminal progenitors, mature LC and MaSC based on CD61, CD24 and CD29 expression [29]. As expected, we observed an overlap between the signatures derived from our analysis and the previous ones described in adulthood only. For the ER+ LC conserved signature, 40% of the genes (86 out of 215) were common either to the signature described for ER+ LC [32] or to the signature of mature LC [29]. For ER− LC, 11 out of the 58 genes found in the conserved signature were described in at least one of these 2 studies, while 48% of the genes (441 out of 915) in the conserved BC signature were previously described as BC specific (Supplementary Fig. S3 and Supplementary Table 6). The partial overlap between our described conserved signatures and these studies is

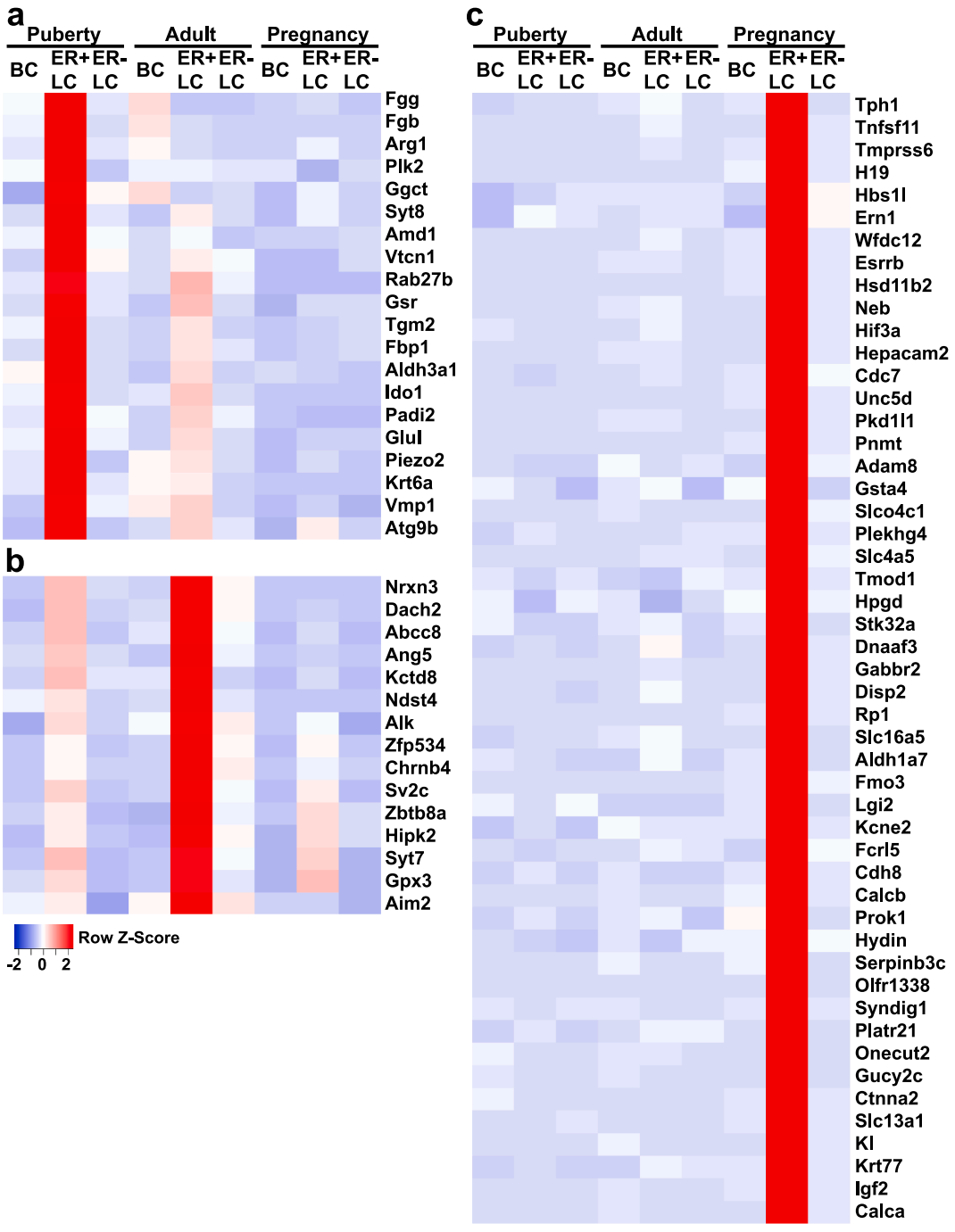



**Fig. 2.** ER + LC show stage specific signatures, in particular in pregnancy. Heatmaps of gene expression levels of examples of genes in puberty (a), adult (b), and pregnancy (c) specific signatures in ER+ LC.

probably be due to distinct experimental strategies. Indeed, the 2 previously published studies were performed using microarrays while our study was performed using RNAseq which allows full sequencing of the transcriptome and is therefore likely to identify more differentially modulated transcripts than microarrays. We also compared our conserved signatures to 2 human datasets. In both studies, adult human mature luminal, luminal progenitor and basal stem/progenitor were isolated based on Epcam and CD49f [30,33]. The overlap between our study and the 2 human datasets was smaller than in the comparison with 2 mice datasets. However, markers

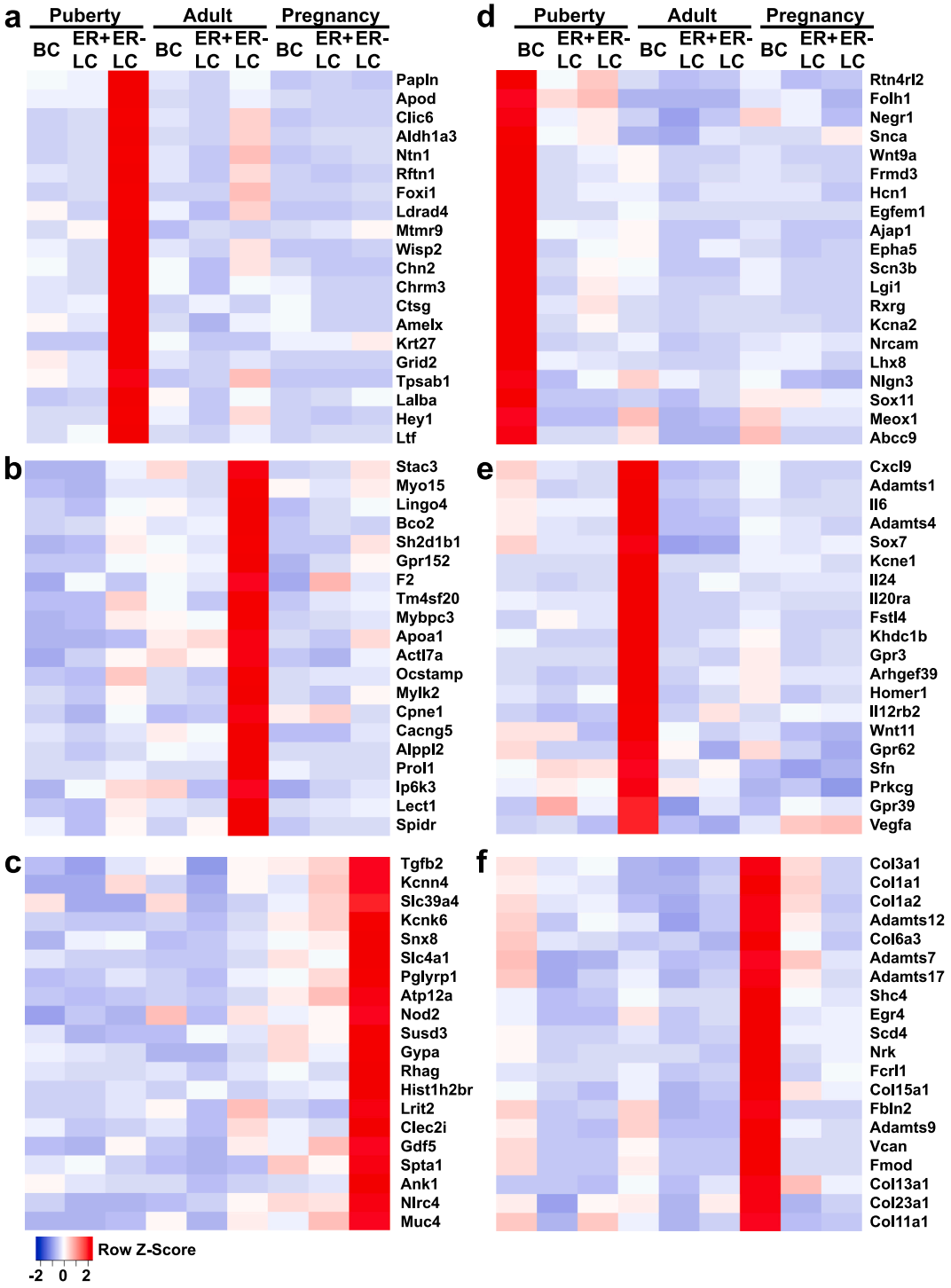



**Fig. 3.** ER− LC and BC show stage specific signatures. Heatmaps of gene expression levels of examples of genes in puberty (a, d), adult (b, e), and pregnancy (c, f) specific signatures in ER− LC (a–c) and BC (d–f).

of ER+ LC like Foxa1, Pgr, Prlr and Areg were conserved through the studies as well as markers of BC like Krt5, Krt14, Myh11 and Trp63, supporting the relevance and similarity of cell types in the mammary gland between mice and human (Fig. S4 and supplementary Supplementary Table S7). Altogether these data show that all 3 MG cell types have a conserved signature throughout the different stages of life and this comparison shows that our analysis uncovers many new genes preferentially expressed by one cell lineage that were not previously described and which could play important roles during MG development, homeostasis or pregnancy.

### 3.2. Stage specific signatures of the MG lineages

As a general tendency, ER+ LC and ER− LC signatures were more similar when comparing puberty and adult than when comparing puberty and pregnancy or adult and pregnancy. This suggests that upon pregnancy, ER+ LC and ER− LC undergo a profound reprogramming. While this observation is not surprising for ER− LC, which are known to differentiate into alveolar cells during pregnancy, the function of ER+ LC during pregnancy is poorly known and a profound reprogramming not described, mainly due to the scarcity of these cells upon pregnancy. In contrast, BC signatures were not drastically different in one stage compared to the others, suggesting that these cells do not undergo major reprogramming upon pregnancy (Supplementary Fig. S2).

To more precisely define how the transcriptional program of the 3 different MG epithelial lineages changes during puberty, adulthood and pregnancy, we defined stage-specific signatures of each cell type during puberty, adulthood and pregnancy.

The puberty specific signature was defined as genes specifically upregulated in each lineage (as compared to the other lineages) at puberty compared to adult and pregnancy. The other stage signatures were derived following the same strategy.

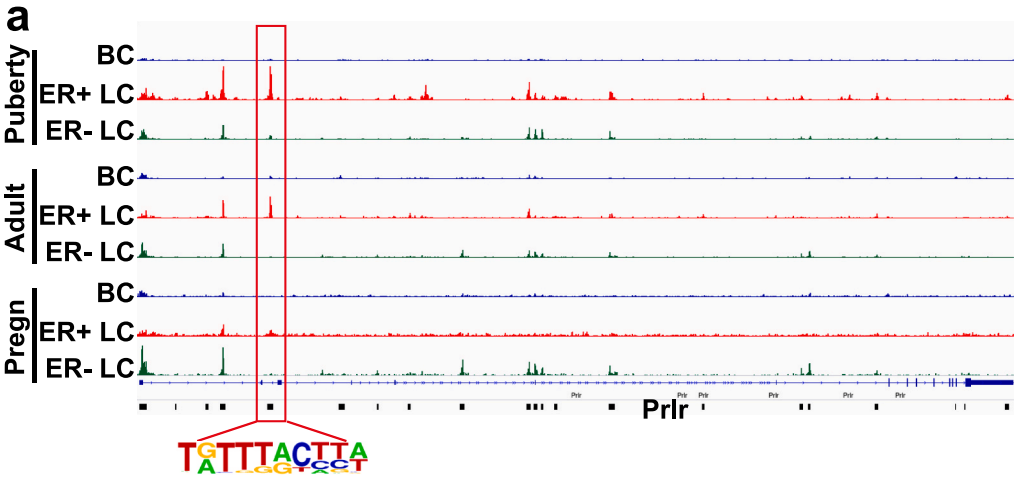
The puberty specific signature in ER+ LC comprised 146 genes upregulated in ER+ LC in puberty compared to the other cell types and compared to ER+ LC at other cell stages. Gene ontology on the 146 genes of this signature showed an enrichment in genes linked to metabolism, like Arg1, GluI, Amd1, Aldh3a1, Ido1, Padi2 and Ggct for example. Genes related to autophagy, like Atg9b and Vmp1 were also more expressed by ER+ LC in puberty. This suggests that ER+ LC have specific functions during puberty, with potentially a higher metabolism and autophagy, which could be related to their increased cycling behaviour during puberty. Several ligands, like Nrg4, Cxcl17, Fgf13, Epgn and Inhba were upregulated in puberty in ER+ LC suggesting a specific role in pubertal development (Fig. 2a, Supplementary Fig. S5 and Supplementary Tables 3 and 8).

The adult specific signature in ER+ LC comprised only 45 genes, suggesting no major specific role of ER+ LC only in adulthood. GO terms and a few examples of genes comprised in this signature are illustrated in Fig. 2 and Supplementary Fig. S5. No ligand were specifically upregulated in ER+ LC only in adult, but several ligands were expressed in puberty and adult and not during pregnancy in ER+ LC: Ptn, Col8a1, Il33, Stc2, Enho, Wnt7b and Wnt5a, again supporting that ER+ LC do not have a specific role in adulthood that they would not have in puberty as well (Fig. 2b, Supplementary Fig. S5 and Supplementary Tables 3 and 8).

The pregnancy specific signature in ER+ LC comprised 206 genes. A striking observation was that many of those genes were very highly expressed in LC ER+ in pregnancy compared to the other stages and cell types (Fig. 2c). This suggests a unique and highly specific cell expression program for the rare ER+ LC cells during pregnancy, supporting a unique role for these very rare cells during pregnancy. Gene ontology of this signature revealed an enrichment in genes involved in sodium transport, MAPK cascade regulation, regulation of kinase B signalling and response to estradiol. Esrrb, Gabbr2 and Unc5d receptors were very highly expressed in ER+ LC in mid pregnancy, suggesting a major role of these receptors in signalling in ER+ LC only in pregnancy (Fig. 2c, Supplementary Fig. S5 and Supplementary Table 3). Defb1, Bmp3, Igf2, Calca, Tnfsf11, Penk and Il34 ligands were mainly associated with pregnancy, supporting a specific paracrine expression profile in ER+ LC in pregnancy (Supplementary Table 8). The exact role of these cells specifically during pregnancy and the cell-cell communications they are involved in would certainly deserve more attention.

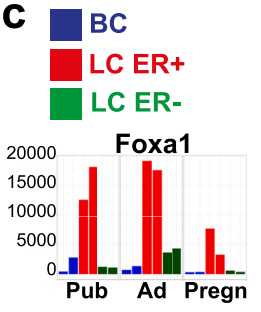
In ER− LC, the puberty, adult and pregnancy specific signatures comprised 191, 218 and 163 genes respectively and were associated with different GO terms depending on the life stage (Supplementary Fig. S5). Foxi1 and Hey1 TFs were only upregulated in puberty in ER-LC. The paracrine function of ER− LC seemed life-stage dependent as different ligands were expressed at different stages. Many ligands were expressed higher in ER− LC in puberty and to a lesser extend in adult stage: Inhbb, Cgref1, Dmkn, Cxcl1, Eml5, Ntn1, Lrg1, Tgfb3, Angpt1, Rnpep, Lcn2, Igfbp5 and Cp. Sbsn, Lif, Cxcl16, Tnf, Esm1 and Lect1 ligands were higher in ER− LC adult and to a lesser extend in puberty and pregnancy. Some ligands were also associated with puberty and adult in ER− LC like Cma1 and Hapln3. In pregnancy, Chrdl2, Apln, Tnc, Spp1, Slit1, Cspg5, Cel and Tgfb2 ligands were highly expressed in ER− LC, while Scube1 and Il23a were higher in adult and pregnancy (Fig. 3a–c, Supplementary Fig. S5 and Supplementary Tables 4 and 8). These results show that ER− LC have different transcriptomes, and in particular ligand expression, depending on the life stage, likely regulating the stage specific functions of this cell lineage.

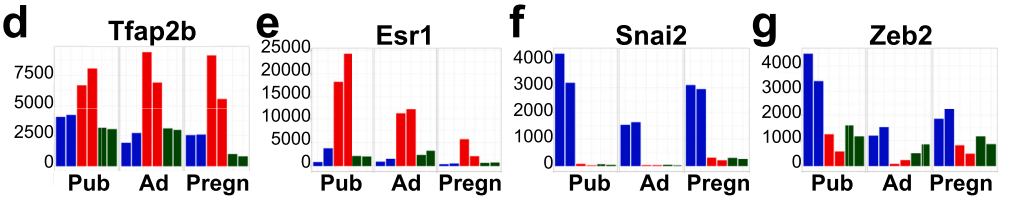
In BC, the puberty, adult and pregnancy specific signatures comprised 333, 718 and 160 genes respectively, and were also associated with different gene ontology terms depending on the life stage. While Sox11 was associated with puberty, Sox7 was associated with adulthood in BC, suggesting different TF of the same family could finetune the expression program of these cells depending on the life stage. Adult and pregnancy specific signatures were both associated with extracellular matrix remodelling terms, but different sets of genes were upregulated in the 2 stages (Adamts4, Adamts1, Sfn in adulthood and Adamts17, Adamts9, Adamts12, Adamts7, Col1a1, Col1a2, Col3a1, Col11a1, Col6a3, Col13a1, Col15a1, Col23a1 in pregnancy), suggesting BC are inducing a significant change in extracellular matrix organization upon pregnancy. Many ligands were expressed in BC and were differentially expressed depending on the life stage. Ligands expressed higher in puberty were Wnt9a, Chrdl1, Sfrp5, Dhh and Eda. Thbs1, Ccl7, Wnt10a, Fst, Ccl2, Tfpi, Ccl11, Srgn, Bmp2, Il6, Cxcl9, Brinp3, Il24 and Tnfsf10 ligands were higher in adult, while ligands higher in puberty and adult stages were C1qtnf9, Bmp6, Egfl7, Bmp4, Fgf2, Egfl6, Cxcl14, Pros1, Cxcl12, Fgf1, Fstl3, Edil3, Tenm2, Igfbp7, Scube3, Dkk3, Sfrp1, Slit2, Nbl1, Kitl, and Jag1. Many ligands showed higher expression in BC in pregnancy: Lamc3, Dlk2, Igfbp2, Ndp, Ereg, Nrg1, Nkd1, Fjx1, Lamb1, Sema3d, Igfbp4, Fstl1, Wnt6, Sostdc1, Ntf3, Ndnf, Lamc1, Lrr4c, Lama1, Il17b, Pdgfa, Jag2, Dll1, Bmp7, Sema4g and Chadl,




**b** Peaks in ER+ LC conserved at all stages

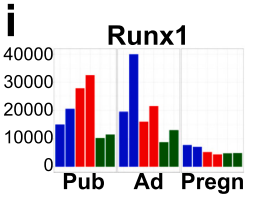
| Motif | Name | Pval | %Target | %Bg |
|---|---|---|---|---|
| | FOXA1(Forkhead) | 1E-162 | 77.66 | 45.23 |
| | AP-2gamma(AP2) | 1E-9 | 30.61 | 23.89 |
| | Slug(Zf) | 1E-9 | 16.61 | 11.63 |
| | Esrrb(NR) | 1E-6 | 20.15 | 15.49 |



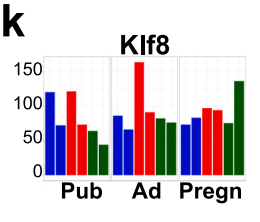

**h** In ER+ LC in puberty

| Motif | Name | Pval | %Target | %Bg |
|---|---|---|---|---|
| | FOXA1(Forkhead) | 1E-33 | 53.00 | 42.19 |
| | JunB(bZIP) | 1E-32 | 51.31 | 40.64 |
| | RUNX2(Runt) | 1E-12 | 32.12 | 26.32 |



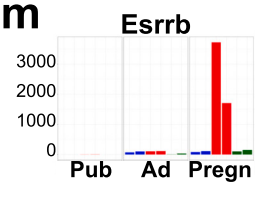

**j** In ER+ LC in adult

| Motif | Name | Pval | %Target | %Bg |
|---|---|---|---|---|
| | FOXA1(Forkhead) | 1E-209 | 59.81 | 43.65 |
| | KLF5(Zf) | 1E-37 | 29.53 | 23.63 |
| | Slug(Zf) | 1E-7 | 13.37 | 11.45 |
| | ERE(NR) | 1E-3 | 8.48 | 7.58 |




**l** In ER+ LC in pregnancy

| Motif | Name | Pval | %Target | %Bg |
|---|---|---|---|---|
| | FOXA1(Forkhead) | 1E-21 | 66.04 | 34.20 |
| | Slug(Zf) | 1E-3 | 22.17 | 13.06 |
| | Esrrb(NR) | 1E-3 | 24.53 | 15.56 |




*(caption on next page)*

**Fig. 4.** ER + LC show a conserved chromatin opening signature throughout different stages of life as well as stage specific chromatin remodelling and associated TFs. a, Representative example of chromatin opening associated with ER+ LC at all stages, as measured by ATAC-seq. b, HOMER motif enrichment analysis of peaks in ER+ LC conserved at all stages (n = 712), total background sequences = 103441. c-g, Expression levels of transcriptions factors associated with motifs enrichment analysis showing higher expression (c–e) or lower expression (f–g) in ER+ LC. h, HOMER motif enrichment analysis of peaks in ER+ LC in puberty (n = 3138 and total background sequences = 105277). i, Expression levels of Runx1 transcription factor associated with motifs enrichment analysis showing higher expression in ER+ LC in puberty. j, HOMER motif enrichment analysis of peaks in ER+ LC in adult (n = 9095 and total background sequences = 95095). k, Expression levels of Klf8 transcription factor associated with motifs enrichment analysis showing higher expression in ER+ LC in adult. l, HOMER motif enrichment analysis of peaks in ER+ LC in pregnancy (n = 1092 and total background sequences = 102832). m, Expression levels of Essrb transcription factor associated with motifs enrichment analysis showing higher expression in ER+ LC in pregnancy. P Values were calculated using a binomial test.

suggesting different cell-cell communication pathways between the different cell types during the different stages of life in the mammary gland (Fig. 3d–f, Supplementary Fig. S5 and Supplementary Tables 5 and 8).

ER+ LC show a conserved chromatin opening signature throughout different stages of life as well as cell stage specific chromatin remodelling and associated TFs.

To identify the chromatin landscapes associated with the different cell lineages of the MG, we performed ATACseq to define the open chromatin regions associated with the three different epithelial cell types of the MG in puberty, adulthood and pregnancy. Chromatin-open regions (ATAC-seq peaks) associated with a given lineage in a specific stage were defined as the ATAC-seq peaks up-regulated for at least 2-fold, compared to the two other lineages at each developmental stage (Supplementary Tables 9 and 10).

To define the TF associated with chromatin remodelling, we performed motif discovery analyses on the chromatin-remodelled regions using homer [37]. The peaks remodelled specifically in each population were compared to the corresponding backgrounds, to obtain enriched TF motifs in each population.

Motif prediction was first performed on open-chromatin regions that are open in each lineage across the different developmental stages. As expected, Esr1 and Foxa1 were enriched in ER+ LC population [38]. Zeb2, and Snai2, which have been associated with stem cell properties of BC [39], were also enriched in ER+ LC population. However, both Snai2 and Zeb2 motifs were enriched in peaks related to genes downregulated in ER+ LC, consistent with a repressor activity of these TFs. Additionally, Tfap2b (AP-2β) was enriched in ER+ LC population (Fig. 4a–g).

Motif prediction was also performed on the different stage specific open-chromatin regions. In ER+ LC, the highest enriched motif was Foxa1 in all stages. Slug and ERE motifs were also enriched in ER+ LC in adult and pregnancy suggesting that Snai2, which is expressed only in BC, represses the chromatin regions opened preferentially in ER+ LC. ERE motif was also enriched in ER+ LC in adult and pregnancy, consistent with the importance of ER signalling. Jun-AP1 and Runx motifs were enriched exclusively in ER+ LC in puberty, with no clear Jun-AP1 TF upregulated in puberty ER+ LC while Runx1 showed higher expression in puberty ER+ LC. Klf motifs were enriched in adult ER+ LC, and among the Klf TFs, Klf8 showed higher expression, while in pregnancy ER+ LC, NR motifs were enriched and Essrb TF was highly enriched (Fig. 4 h-m and Supplementary Fig. S6). The motif discovery analysis in ER+ LC confirms the well-known role of Esr1 and FoxA1 being the highest enriched motifs at all stages in ER+ LC. In addition, this analysis also uncovers stage specific open chromatin regions in ER+ LC that are associated with Jun-AP1 and Runx TF motifs in puberty and Essrb motif in pregnancy, suggesting that while FoxA1 and Esr1 are the main TF responsible for ER+ LC specification, other TF, like Essrb in pregnancy, could control the stage specific program of ER+ LC.

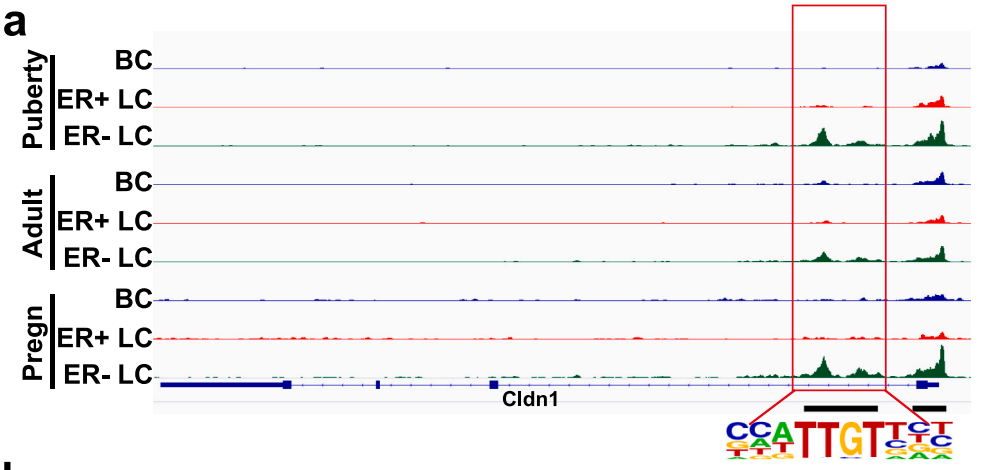
ER− LC show a conserved chromatin opening signature as well as puberty and pregnancy specific chromatin remodelling and associated TFs.

Within the peaks conserved in the 3 different developmental stages in ER− LC, ETS family (EHF and ELF5) were enriched with the highest rank, as previously described [30]. Sox TF motif was also enriched in ER− LC, and among Sox TFs, Sox6 [32] and Sox13 showed higher expression in ER− LC. NFkB motif was enriched in ER− LC to a lower extend, and only Nfkb2 showed a slight increase of expression in ER− LC (Fig. 5a–g).

ETS, Sox and NfkB TF motifs were enriched in ER− LC across different developmental stages. Jun-AP1 motif enrichment, along with Maff and Mafk TF higher expression, and CEBP motif enrichment, along with Cebpd higher expression were observed in ER− LC only at puberty. No motif was enriched specifically in ER− LC in adulthood. CP2 motif together with a higher expression of Grhl1, and AP2 motif were enriched in ER− LC during pregnancy (Fig. 5h-m and Supplementary Fig. S7). This data set suggests that, in ER− LC, beside the well-known ETS TF found to be the most enriched motif at all stages, Sox and Nfkb motifs are also enriched at all stages, while Jun-AP1, CEBP, Grhl1 and AP2 are likely to control the ER− LC transcriptional and chromatin landscape in puberty and pregnancy.

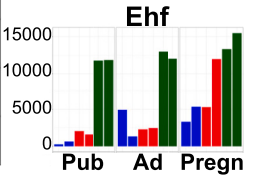
### *3.3. BC identity and function are mainly controlled by the same set of TFs throughout the different stages of life*

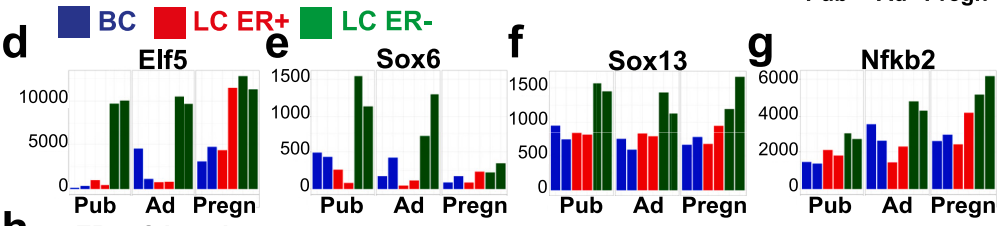
Conserved TF motifs enriched in the open chromatin regions of BC across the 3 developmental stages, showed that p63 (p53) motif was the most enriched motif and Trp63 mRNA was highly upregulated specifically in BC, as previously described [26]. Of note, p73 was also found to be upregulated in BC, although at a lower level, and could also play a role in the regulation of BC identity. Tead motifs were enriched in BC, with Tead1 as the only member of this TF family to be more specifically upregulated in BC, although Tead2 and Tead3 were also expressed. MADS motifs were enriched in BC, with only the Mef2c isoform being upregulated specifically in BC. Sox motifs were also enriched in BC, and within all Sox members, only Sox8 showed upregulation in BC at all stages, while Sox6, Sox7 and Sox13 were also expressed. Zinc finger motifs were enriched in BC as previously described [30] and EGR2 enriched in BC at all stages. Finally, Gata motifs were enriched in genes downregulated in BC, consistent with the repression of these genes and the closing




**b** Peaks in ER- LC conserved at all stages

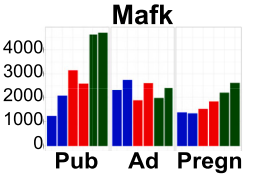
| Motif | Name | Pval | %Target | %Bg |
|---|---|---|---|---|
| | EHF(ETS) | 1E-110 | 73.25 | 39.76 |
| | Sox17(HMG) | 1E-10 | 30.79 | 22.23 |
| | NFkB-p65(RHD) | 1E-7 | 21.23 | 15.25 |

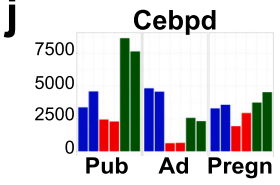



**h** In ER- LC in puberty

| Motif | Name | Pval | %Target | %Bg |
|---|---|---|---|---|
| | EHF(ETS) | 1E-35 | 49.50 | 38.05 |
| | JunB(bZIP) | 1E-11 | 45.34 | 39.11 |
| | Sox9(HMG) | 1E-8 | 31.03 | 26.25 |
| | CEBP(bZIP) | 1E-7 | 22.26 | 18.39 |




**k** In ER- LC in adult

| Motif | Name | Pval | %Target | %Bg |
|---|---|---|---|---|
| | EHF(ETS) | 1E-320 | 64.75 | 37.89 |
| | NFkB-p65(RHD) | 1E-10 | 28.55 | 15.19 |
| | Sox2(HMG) | 1E-7 | 29.69 | 26.18 |

**l** In ER- LC in pregnancy

| Motif | Name | Pval | %Target | %Bg |
|---|---|---|---|---|
| | GRHL2(CP2) | 1E-34 | 26.68 | 13.67 |
| | EHF(ETS) | 1E-25 | 53.13 | 38.15 |
| | AP-2gamma(AP2) | 1E-10 | 39.44 | 30.83 |
| | NFkB-p65(RHD) | 1E-3 | 4.49 | 2.78 |

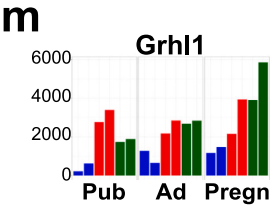



**Fig. 5.** ER− LC show a conserved chromatin opening signature throughout different stages of life as well as cell stage specific chromatin remodelling and associated TFs. a, Representative example of chromatin opening associated with ER− LC at all stages, as measured by ATAC-seq. b, HOMER motif enrichment analysis of peaks in ER− LC conserved at all stages (n = 1088), total background sequences = 102496. c-g, Expression levels of transcription factors associated with motifs enrichment analysis showing higher expression in ER− LC. h, HOMER motif enrichment

analysis of peaks in ER− LC in puberty (n = 2929 and total background sequences = 106632). i,j, Expression levels of transcription factors associated with motifs enrichment analysis showing higher expression in ER− LC in puberty. k, HOMER motif enrichment analysis of peaks in ER− LC in adult (n = 4981 and total background sequences = 104722). l, HOMER motif enrichment analysis of peaks in ER− LC in pregnancy (n = 1294 and total background sequences = 108791). m, Expression levels of Grhl1 transcription factor associated with motifs enrichment analysis showing higher expression in ER− LC in pregnancy. P Values were calculated using a binomial test.

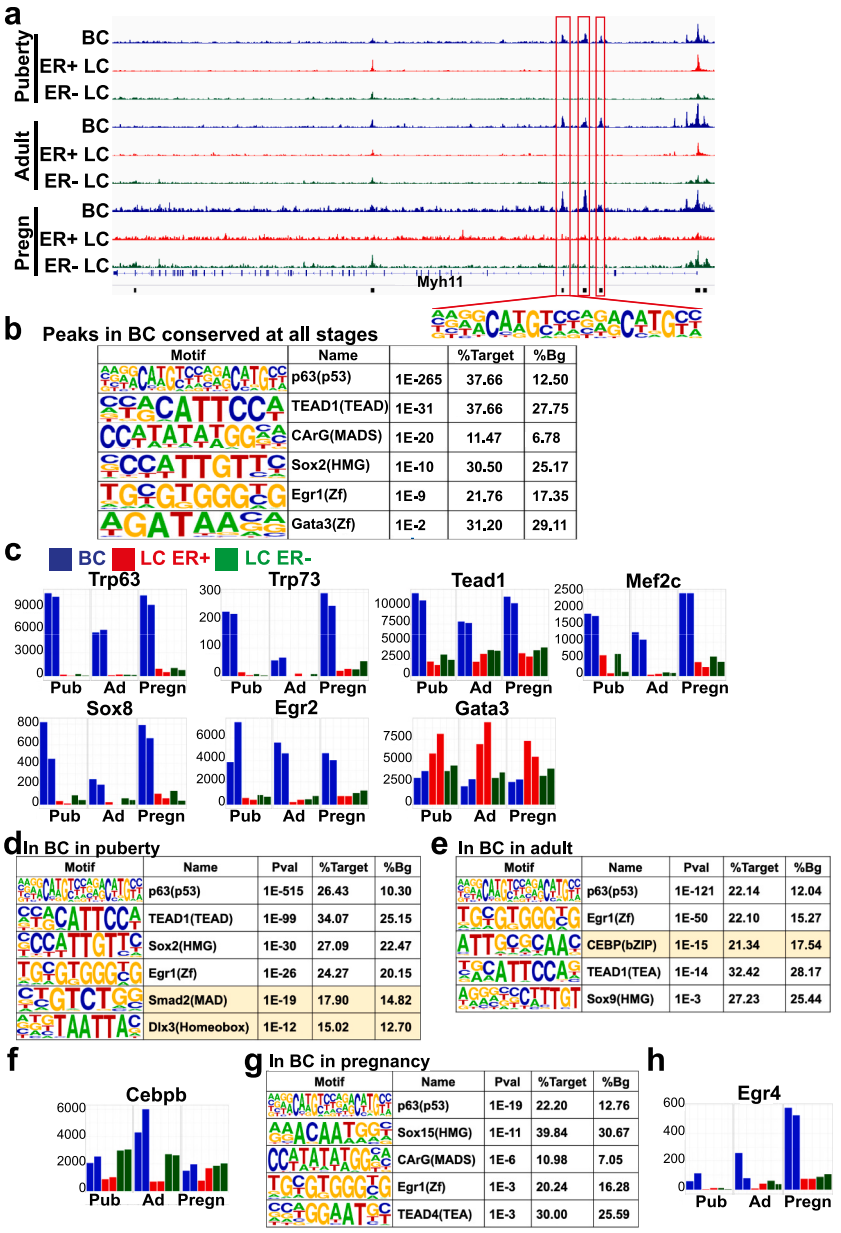



**Fig. 6.** BC show a conserved chromatin opening signature throughout different stages of life as well as cell stage specific chromatin remodelling and associated TFs. a, Representative example of chromatin opening associated with BC at all stages, as measured by ATAC-seq. b, HOMER motif enrichment analysis of peaks in BC conserved at all stages (n = 2974), total background sequences = 104997. c, Expression levels of transcription factors associated with motifs enrichment analysis showing higher or lower (Gata3) expression in BC. d, HOMER motif enrichment analysis of peaks in BC in puberty (n = 11445 and total background sequences = 124724). e, HOMER motif enrichment analysis of peaks in BC in adult (n = 6942 and total background sequences = 101488). f, Expression levels of Cebpb transcription factors associated with motifs enrichment analysis showing higher expression in BC in adult. g, HOMER motif enrichment analysis of peaks in BC in pregnancy (n = 2973 and total background sequences = 106145). h, Expression levels of Egr4 transcription factors associated with motifs enrichment analysis showing higher expression in BC in pregnancy. P Values were calculated using a binomial test.

of their chromatin regions by Gata3, a major regulator of luminal identity [40,41] (Fig. 6a–c).

When focusing on the chromatin regions only opened in BC in a specific developmental stage, most of the enriched TF motifs were identical to the ones found from open chromatin regions in the 3 life stages, suggesting that BC identity and function are controlled by the same set of TFs irrespective of their developmental stage. However, Mad and Dlx motifs were enriched specifically during puberty, with no clear specific upregulation of TF of these families. bZip motif was enriched in the adult stage, along with higher expression of Cebpb TF. No motif was enriched only during pregnancy, but higher expression of Egr4 was observed (Fig. 6d–h and Supplementary Fig. S8).

Overall, the TF motif prediction analysis on the chromatin open regions clearly shows that while well-known cell type specific TF are associated with the different cell types at all stages of life, other TF are associated with specific developmental stages, and could potentially play a role in regulating the function of each lineage at a particular developmental stage.

## 4. Discussion

### *4.1. Resource dataset of conserved and stage-specific transcriptional, chromatin opening and cell-cell communication landscapes in the MG*

Here, we describe the transcriptional and chromatin opening landscapes of the different epithelial cell lineages of the MG at different developmental stages. While single cell RNA sequencing is the gold standard technique to study the cellular heterogeneity within an organ, more robust expression profiles can be achieved with bulk RNA sequencing, pending cell population isolation is possible. In this study, the use of ER+ LC YFP tracing combined with cell surface markers allowed us to isolate the different cell types of the MG with a higher purity than previously achieved in adult stage and allowed, for the first time, the isolation of ER+ LC at mid pregnancy based on the YFP expression since expression of surface markers associated to ER (Sca1 and CD133) is decreased upon pregnancy [34]. RNAseq and ATACseq, performed at different stages of life allowed the definition of conserved and stage-specific gene regulatory networks in the different MG epithelial cells. Compared to previous studies performed by microarray technology only in adult [29,32], the use of RNAseq in this current study allowed full sequencing of the transcriptome and was therefore likely to identify more differentially modulated transcripts than microarrays. Indeed, while a clear overlap of genes were observed, a maximum of 48% (in BC) of the genes found in our conserved signature were previously described in one of these two studies. Another study profiled the ER+ LC in early pregnancy (3 and 7 days) and in virgin mice by microarray [34]. While this study provided important insights on genes differentially regulated upon pregnancy in ER+ LC, it failed to put in light genes differentially regulated upon pregnancy only in ER+ LC, and as ER+ LC were isolated based on surface markers being lost upon pregnancy, later time points in pregnancy could not be achieved. Our study allowed the profiling of ER+ LC at mid pregnancy (day 12), and the identification of genes highly upregulated only in ER+ LC in pregnancy. Some of the genes put in light in our study were previously described to be highly expressed in early pregnancy in ER+ LC, like Igf2, Calca and Tnfsf11 [34]. However, genes previously described to be upregulated in ER+ LC in early pregnancy were not included in our ER+ LC pregnancy signature, as they were also upregulated in BC, for example Mdk and Emb, in ER− LC, like Gldc or both in BC and ER− LC, like Nrep [34]. In conclusion, our analysis allows the identification of genes, TF and ligands conserved or stage-specific specifically for each cell type of the MG.

While bulk sequencing of isolated populations allows deeper sequencing than single cell sequencing, the drawback of this technique is that subpopulations characteristics are lost. Several single cell sequencing studies in mice and human have confirmed that the MG is composed of the three main cell types we adressed in our study [13,14,19,21]. However, we have to acknowledge that two distinct clusters have been described within the ER− LC adult population [13,17], as well as heterogeneity within the BC, in which some subtypes could correspond to BC with higher potential to reconstitute gland upon transplantation [13], and that our study did not allow to address the gene signatures of the different subtypes.

This study was performed in transgenic mice expressing YFP in ER+ LC, allowing cell isolation with high purity that would not be possible from human samples. Similarities in the gene networks and gene signatures of the 3 cell types of the MG have been described previously [29,32] and were confirmed in our study since an overlap between our conserved signatures was found with human corresponding ones, and since major TFs associated with the different cell types are similarly found in human samples [30]. These data support the relevance of our study for human biology.

### *4.2. Conserved and stage-specific gene regulatory networks*

The MG morphology is very dynamic after birth, with an extensive expansion and ductal branching during puberty, cycles of growth and regression with each estrus cycle in adult life, and major differentiation program to form lobuloalveolar structures upon each pregnancy until the end of lactation [1]. Hormonal cues are responsible for these morphological changes of the MG throughout life stages. In particular, estrogen, progesterone and prolactin have been described to be crucial for MG development and morphological changes. Hormone sensing cells (ER+ LC) are thought to sense these cues and to relay the information to neighbouring cells though paracrine signals [3]. Therefore, stage-specific signatures are expected for each cell type of the MG, to reflect the different hormonal environment and to induce the morphological changes associated with each stage. Here, we uncovered common gene signature for each cell lineage across the different stages of life, as well as stage-specific gene signature. We compared cell type specific gene signature at different stages of life, and these data with the analysis of the corresponding chromatin landscape, allowing to define the common GRN of each lineage as well as stage-specific GRN (Supplementary Figs. S9–S11).

In ER+ LC, this analysis confirmed that FoxA1 and Esr1 motifs were the most enriched TF, at all life stages, while Zeb2 and Slug motif, which are expressed in BC exclusively, were enriched in peaks related to genes downregulated in ER+ LC, consistent with a

repressor activity of Zeb2 and Slug. Tfap2b (AP-2β) was enriched in ER+ LC population at all stages as well. While its expression has not been previously associated with ER+ LC, its expression has been shown to be increased in lobular carcinoma [42]. This analysis also shows stage specific motifs in ER+ LC like Jun-AP1 and Runt motifs in puberty, the Klf motif, associated with high expression of Klf8, in adulthood, and the highly enriched Essrb motif in pregnancy. While Runx1 and AP1 role in the mammary gland has been described [43,44], a role for Klf8 and Esrrb has not been described in the mammary gland. Esrrb, which is an orphan receptor, is of particular interest since its expression is highly specific and very strong in ER+ LC specifically in pregnancy. This is associated with the expression a very specific gene expression program in ER+ LC in pregnancy. The potential role of these rare cells could be to sense and relay signals even during pregnancy and lactation, through paracrine signalling cues different than the ones in non-pregnant state and to orchestrate the pregnancy development and involution.

In ER− LC, our analysis not only confirms the central role of Ehf and Elf5 TF, but also confirms a role for Sox6 and suggests a role for Sox13 and Nfkb2 in ER− LC identity. The role of Sox6 in maintaining luminal cell identity has been previously described as Sox6 overexpression maintains luminal differentiation in MG in vitro and in vivo [32], while a putative role of Sox13 in the mammary gland has not been described. While NFkB role in the mammary gland has been described to be required for proper organogenesis in the MG and lactation [45], only Nfkb2 showed a slight increase of expression in ER− LC. Motifs and TF were also enriched in stage specific manner, Maff, Mafk and Cebpd showed higher expression in puberty in ER− LC. While Cebpd has been described to regulate pro-apoptotic gene expression during mammary gland involution [46], its role during puberty remains unclear. Overall, several TF have been shown to be stage specific in ER− LC and could have a role to finetune the gene regulatory network of ER− LC specified by conserved TF at all stages.

In BC, the combined analysis shows many motifs enriched at all stages. p63 (p53) motif and Trp63 was highly upregulated specifically in BC, as previously described [26]. The role of Trp63 in BC specification has been demonstrated first in vitro and upon transplantation [47], and then in vivo, in which sustained p63 expression in embryonic MG lead to BC only, while sustained p63 expression in adult LC lead to a reprogramming of LC into BC [8]. p73 was also found to be upregulated in BC, though at a much lower extend, and could potentially also have a role in the mammary gland. Tead motif, with higher expression of Tead1, MADS motifs with higher expression of Mef2c, Sox motifs with higher expression of Sox8 in BC, and Zinc finger motifs with higher expression of EGR2, were also enriched in BC at all stages in our analysis. While a direct role of Tead1 has not been described in the mammary gland, it is known to interact with YAP/TAZ [48], and YAP/TAZ have been described to play a role in stem cell properties of the mammary epithelial cells, as transient expression of YAP/TAZ in differentiated LC turns them into mammary stem cells [49]. Zinc finger motifs enrichment has been previously described in human BC [30]. Mef2 has previously been associated with a negative regulation proliferation in mammary epithelial cells [50], and with poor prognosis in ER + breast cancer [51]. Sox8 has been shown to act as an oncogene and to be associated with poor prognosis in triple-negative breast cancer [52]. Gata motifs were enriched in BC but these motifs were enriched only in genes downregulated in BC, consistent with a repressor activity of Gata3, a major regulator of luminal identity [40,41]. Few motifs were found to be stage specific in BC. The putative role of the stage specific TF found in BC, like Cepbp in adult and Egr4 in pregnancy, has not been previously described. Overall, BC gene regulatory networks are regulated by a set of TF conserved across the different stages of life, but other stage specific TFs could finetune the BC transcriptional program at different stages of life.

The role of TFs identified in this study, either conserved throughout the stages or specific to one developmental stage will need to be functionally assessed with gain and loss of function experiments to confirm their putative role in MG. Such functional validations have previously demonstrated that, for example, Gata3 and Foxa1 TFs are crucial for ER+ LC specification [40,41,53], that Elf5, Stat5, Runx2 and Sox9 TFs play a crucial role in ER− LC and alveologenesis specification [54–58], and that p63 is a master regulator of basal specification [8,47].

### *4.3. Conserved and stage-specific ligands constitute potential paracrine signals*

Our study also identified many potential ligands, specifically expressed by one cell type, often modulated during life stages, that could be involved in the heterotopic communication between cell types. While the role of some ligands like Areg [59] and Tnfsf11 [5] in relaying hormonal signals from ER+ LC to neighbouring cells have been extensively studied in the MG and the expression or role of other ligands has been previously described as well, like the role of Wnt family in MG development and ductal growth [60–62], the role of Tgfb family in cell proliferation and differentiation [63,64], the diverse role of ligands for EGF receptor [65,66] and the negative effect of Ptn ligand on MG differentiation [67], other ligand functions in the MG have not been elucidated. For example, a role for Bmp3 in the mammary gland has not been described but it has been shown to be expressed in breast cancers and was associated with a better survival [68]. Stc2 expression is dependent on ER signalling and low expression was associated with poor prognosis in breast cancer [69,70]. In vitro, Stc2 expression reduced cell growth, motility and viability of breast cancer cell lines [71] but its physiological role in the mammary gland has not been investigated. Npnt has been shown to promote metastasis in breast cancer cell lines and is correlated with a poor prognosis in breast cancer patients [72], but its role in the mammary gland is not clear. Calca, which we show is highly expressed only by ER+ LCs in mid-pregnancy, has been previously described to be expressed in epithelial cells during pregnancy, and is known to be responsive to progesterone. Since its expression correlates with the gland proliferation and remodelling during pregnancy, it has been suggested that it may be involved in these processes but this has not been investigated further [34,73]. These few examples show that ligands expressed specifically by one cell type of the MG could have putative roles in cell-cell communication and that there are still a lot of candidates which would deserve more attention. All these ligands constitute potential autocrine, paracrine or longer range signalling cues between the different cell types of the MG. While some ligands are very specific for one cell type and are highly expressed at all stages of life, other ligands are associated with puberty, adulthood or pregnancy,

suggesting that different signalling pathways could be used for cell-cell communication in the MG depending on the life stage.

The functional role of the ligands put in light in this study will need to be further studied to demonstrate their putative role as signalling cues between the MG cells. Loss of function and overexpression assays will need to be performed to pinpoint their function in MG, as it has extensively been done for Areg [59] or, as another example of a less known signaling cue in the MG, Cxcl12 [24].

## 5. Conclusions

This resource will be useful to define the TF networks responsible for the specification and maintenance of the different MG epithelial lineages as well as to understand how perturbations in these GRN affect MG development and function at the different stages of life and whether they play a role during breast tumorigenesis.

## Author contribution statement

Yura Song: Analyzed and interpreted the data; Wrote the paper.
Marco Fioramonti: Analyzed and interpreted the data.
Gaëlle Bouvencourt, Christine Dubois: Performed the experiments.
Cédric Blanpain: Conceived and designed the experiments; Analyzed and interpreted the data; Wrote the paper.
Alexandra Van Keymeulen: Conceived and designed the experiments; Performed the experiments; Analyzed and interpreted the data; Wrote the paper.

## Funding statement

Dr Alexandra Van Keymeulen was supported by Fonds De La Recherche Scientifique - FNRS.

## Data availability statement

RNAseq datasets have been deposited in the Gene Expression Omnibus under accession number GSE222485.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper

## Acknowledgements

We thank the animal house facility, the FACS facility and the Brightcore Genomics core facility of ULB (Erasme campus). Y.S. and G. B. were supported by the TELEVIE. M.F. was a post-doctoral researcher of the FNRS. C.B. is an investigator of WELBIO. A.V.K. is senior research associate from the FNRS. This work was supported by the FNRS and TELEVIE.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.heliyon.2023.e17842.

## References

[1] C.J. Watson, W.T. Khaled, Mammary development in the embryo and adult: a journey of morphogenesis and commitment, Development 135 (2008) 995–1003, https://doi.org/10.1242/dev.005439.
[2] J.L. Fendrick, A.M. Raafat, S.Z. Haslam, Mammary gland growth and development from the postnatal period to postmenopause: ovarian steroid receptor ontogeny and regulation in the mouse, J. Mammary Gland Biol. Neoplasia 3 (1998) 7–22, https://doi.org/10.1023/a:1018766000275.
[3] C. Brisken, B. O'Malley, Hormone action in the mammary gland, Cold Spring Harbor Perspect. Biol. 2 (2010) a003178, https://doi.org/10.1101/cshperspect.a003178.
[4] G. Söderqvist, et al., Proliferation of breast epithelial cells in healthy women during the menstrual cycle, Am. J. Obstet. Gynecol. 176 (1997) 123–128, https://doi.org/10.1016/s0002-9378(97)80024-5.
[5] L.M. Arendt, C. Kuperwasser, Form and function: how estrogen and progesterone regulate the mammary epithelial hierarchy, J. Mammary Gland Biol. Neoplasia 20 (2015) 9–25, https://doi.org/10.1007/s10911-015-9337-0.
[6] A. Van Keymeulen, et al., Distinct stem cells contribute to mammary gland development and maintenance, Nature 479 (2011) 189–193, https://doi.org/10.1038/nature10573.
[7] A. Wuidart, et al., Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells, Genes Dev. 30 (2016) 1261–1277, https://doi.org/10.1101/gad.280057.116.
[8] A. Wuidart, et al., Early lineage segregation of multipotent embryonic mammary gland progenitors, Nat. Cell Biol. 20 (2018) 666–676, https://doi.org/10.1038/s41556-018-0095-2.
[9] A. Van Keymeulen, et al., Lineage-restricted mammary stem cells sustain the development, homeostasis, and regeneration of the estrogen receptor positive lineage, Cell Rep. 20 (2017) 1525–1532, https://doi.org/10.1016/j.celrep.2017.07.066.

[10] V. Rodilla, et al., Luminal progenitors restrict their lineage potential during mammary gland development, PLoS Biol. 13 (2015), e1002069, https://doi.org/10.1371/journal.pbio.1002069.
[11] B. Lloyd-Lewis, F.M. Davis, O.B. Harris, J.R. Hitchcock, C.J. Watson, Neutral lineage tracing of proliferative embryonic and adult mammary stem/progenitor cells, Development 145 (2018), https://doi.org/10.1242/dev.164079.
[12] L. Tao, M.P. van Bragt, E. Laudadio, Z. Li, Lineage tracing of mammary epithelial cells using cell-type-specific cre-expressing adenoviruses, Stem Cell Rep. 2 (2014) 770–779, https://doi.org/10.1016/j.stemcr.2014.04.004.
[13] K. Bach, et al., Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing, Nat. Commun. 8 (2017) 2128, https://doi.org/10.1038/s41467-017-02001-5.
[14] B. Pal, et al., Construction of developmental lineage relationships in the mouse mammary gland by single-cell RNA profiling, Nat. Commun. 8 (2017) 1627, https://doi.org/10.1038/s41467-017-01560-x.
[15] C.M. Li, et al., Aging-associated alterations in mammary epithelia and stroma revealed by single-cell RNA sequencing, Cell Rep. 33 (2020), 108566, https://doi.org/10.1016/j.celrep.2020.108566.
[16] H. Sun, et al., Single-cell RNA-Seq reveals cell heterogeneity and hierarchy within mouse mammary epithelia, J. Biol. Chem. 293 (2018) 8315–8329, https://doi.org/10.1074/jbc.RA118.002297.
[17] N. Pervolarakis, et al., Integrated single-cell transcriptomics and chromatin accessibility analysis reveals regulators of mammary epithelial cell identity, Cell Rep. 33 (2020), 108273, https://doi.org/10.1016/j.celrep.2020.108273.
[18] N. Kanaya, et al., Single-cell RNA-sequencing analysis of estrogen- and endocrine-disrupting chemical-induced reorganization of mouse mammary gland, Commun Biol 2 (2019) 406, https://doi.org/10.1038/s42003-019-0618-9.
[19] R.R. Giraddi, et al., Single-cell transcriptomes distinguish stem cell state changes and lineage specification programs in early mammary gland development, Cell Rep. 24 (2018) 1653–1666.e1657, https://doi.org/10.1016/j.celrep.2018.07.025.
[20] B. Pal, et al., Single cell transcriptome atlas of mouse mammary epithelial cells across development, Breast Cancer Res. 23 (2021) 69, https://doi.org/10.1186/s13058-021-01445-4.
[21] Q.H. Nguyen, et al., Profiling human breast epithelial cells using single cell RNA sequencing identifies cell diversity, Nat. Commun. 9 (2018) 2028, https://doi.org/10.1038/s41467-018-04334-1.
[22] J. Stingl, et al., Purification and unique properties of mammary epithelial stem cells, Nature 439 (2006) 993–997, https://doi.org/10.1038/nature04496.
[23] A.E. Casey, et al., Mammary molecular portraits reveal lineage-specific features and progenitor cell vulnerabilities, J. Cell Biol. 217 (2018) 2951–2974, https://doi.org/10.1083/jcb.201804042.
[24] Y.J. Shiah, et al., A progesterone-CXCR4 Axis controls mammary progenitor cell fate in the adult gland, Stem Cell Rep. 4 (2015) 313–322, https://doi.org/10.1016/j.stemcr.2015.01.011.
[25] A. Van Keymeulen, et al., Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity, Nature 525 (2015) 119–123, https://doi.org/10.1038/nature14665.
[26] N. Forster, et al., Basal cell signaling by p63 controls luminal progenitor function and lactation via NRG1, Dev. Cell 28 (2014) 147–160, https://doi.org/10.1016/j.devcel.2013.11.019.
[27] M. Shipitsin, et al., Molecular definition of breast tumor heterogeneity, Cancer Cell 11 (2007) 259–273, https://doi.org/10.1016/j.ccr.2007.01.013.
[28] P. Gascard, et al., Epigenetic and transcriptional determinants of the human breast, Nat. Commun. 6 (2015) 6351, https://doi.org/10.1038/ncomms7351.
[29] E. Lim, et al., Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways, Breast Cancer Res. 12 (2010) R21, https://doi.org/10.1186/bcr2560.
[30] D. Pellacani, et al., Analysis of normal human mammary epigenomes reveals cell-specific active enhancer states and associated transcription factor networks, Cell Rep. 17 (2016) 2060–2074, https://doi.org/10.1016/j.celrep.2016.10.058.
[31] J.A. Colacino, et al., Heterogeneity of human breast stem and progenitor cells as revealed by transcriptional profiling, Stem Cell Rep. 10 (2018) 1596–1609, https://doi.org/10.1016/j.stemcr.2018.03.001.
[32] H. Kendrick, et al., Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate, BMC Genom. 9 (2008) 591, https://doi.org/10.1186/1471-2164-9-591.
[33] E. Lim, et al., Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers, Nat. Med. 15 (2009) 907–913, https://doi.org/10.1038/nm.2000.
[34] D. De Silva, K. Kunasegaran, S. Ghosh, A.M. Pietersen, Transcriptome analysis of the hormone-sensing cells in mammary epithelial reveals dynamic changes in early pregnancy, BMC Dev. Biol. 15 (2015) 7, https://doi.org/10.1186/s12861-015-0058-9.
[35] M. Shackleton, et al., Generation of a functional mammary gland from a single stem cell, Nature 439 (2006) 84–88, https://doi.org/10.1038/nature04372.
[36] K.E. Sleeman, et al., Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland, J. Cell Biol. 176 (2007) 19–26, https://doi.org/10.1083/jcb.200604065.
[37] S. Heinz, et al., Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities, Mol. Cell. 38 (2010) 576–589, https://doi.org/10.1016/j.molcel.2010.05.004.
[38] A. Hurtado, K.A. Holmes, C.S. Ross-Innes, D. Schmidt, J.S. Carroll, FOXA1 is a key determinant of estrogen receptor function and endocrine response, Nat. Genet. 43 (2011) 27–33, https://doi.org/10.1038/ng.730.
[39] X. Ye, et al., Distinct EMT programs control normal mammary stem cells and tumour-initiating cells, Nature 525 (2015) 256–260, https://doi.org/10.1038/nature14897.
[40] M.L. Asselin-Labat, et al., Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation, Nat. Cell Biol. 9 (2007) 201–209, https://doi.org/10.1038/ncb1530.
[41] H. Kouros-Mehr, E.M. Slorach, M.D. Sternlicht, Z. Werb, GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland, Cell 127 (2006) 1041–1055, https://doi.org/10.1016/j.cell.2006.09.048.
[42] M. Raap, et al., Lobular carcinoma in situ and invasive lobular breast cancer are characterized by enhanced expression of transcription factor AP-2β, Lab. Invest. 98 (2018) 117–129, https://doi.org/10.1038/labinvest.2017.106.
[43] D. Hong, et al., Runx1 stabilizes the mammary epithelial cell phenotype and prevents epithelial to mesenchymal transition, Oncotarget 8 (2017) 17610–17627, https://doi.org/10.18632/oncotarget.15381.
[44] Q. Shen, et al., The AP-1 transcription factor regulates postnatal mammary gland development, Dev. Biol. 295 (2006) 589–603, https://doi.org/10.1016/j.ydbio.2006.03.042.
[45] Y. Cao, M. Karin, NF-kappaB in mammary gland development and breast cancer, J. Mammary Gland Biol. Neoplasia 8 (2003) 215–223, https://doi.org/10.1023/a:1025905008934.
[46] M. Thangaraju, et al., C/EBPdelta is a crucial regulator of pro-apoptotic gene expression during mammary gland involution, Development 132 (2005) 4675–4685, https://doi.org/10.1242/dev.02050.
[47] O. Yalcin-Ozuysal, et al., Antagonistic roles of Notch and p63 in controlling mammary epithelial cell fates, Cell Death Differ. 17 (2010) 1600–1612, https://doi.org/10.1038/cdd.2010.37.
[48] L. Currey, S. Thor, M. Piper, TEAD family transcription factors in development and disease, Development 148 (2021), https://doi.org/10.1242/dev.196675.
[49] T. Panciera, et al., Induction of expandable tissue-specific stem/progenitor cells through transient expression of YAP/TAZ, Cell Stem Cell 19 (2016) 725–737, https://doi.org/10.1016/j.stem.2016.08.009.
[50] A. Clocchiatti, et al., The MEF2-HDAC axis controls proliferation of mammary epithelial cells and acini formation in vitro, J. Cell Sci. 128 (2015) 3961–3976, https://doi.org/10.1242/jcs.170357.
[51] A. Clocchiatti, et al., Class IIa HDACs repressive activities on MEF2-depedent transcription are associated with poor prognosis of ER⁺ breast tumors, Faseb. J. 27 (2013) 942–954, https://doi.org/10.1096/fj.12-209346.

[52] H. Tang, et al., SOX8 acts as a prognostic factor and mediator to regulate the progression of triple-negative breast cancer, Carcinogenesis 40 (2019) 1278–1287, https://doi.org/10.1093/carcin/bgz034.
[53] G.M. Bernardo, et al., FOXA1 is an essential determinant of ERalpha expression and mammary ductal morphogenesis, Development 137 (2010) 2045–2054, https://doi.org/10.1242/dev.043299.
[54] J. Zhou, et al., Elf5 is essential for early embryogenesis and mammary gland development during pregnancy and lactation, EMBO J. 24 (2005) 635–644, https://doi.org/10.1038/sj.emboj.7600538.
[55] S.R. Oakes, et al., The Ets transcription factor Elf5 specifies mammary alveolar cell fate, Genes Dev. 22 (2008) 581–586, https://doi.org/10.1101/gad.1614608.
[56] Y. Cui, et al., Inactivation of Stat5 in mouse mammary epithelium during pregnancy reveals distinct functions in cell proliferation, survival, and differentiation, Mol. Cell Biol. 24 (2004) 8037–8047, https://doi.org/10.1128/mcb.24.18.8037-8047.2004.
[57] T.W. Owens, et al., Runx2 is a novel regulator of mammary epithelial cell fate in development and breast cancer, Cancer Res. 74 (2014) 5277–5286, https://doi.org/10.1158/0008-5472.Can-14-0053.
[58] G.K. Malhotra, et al., The role of Sox9 in mouse mammary gland development and maintenance of mammary stem and luminal progenitor cells, BMC Dev. Biol. 14 (2014) 47, https://doi.org/10.1186/s12861-014-0047-4.
[59] L. Ciarloni, S. Mallepell, C. Brisken, Amphiregulin is an essential mediator of estrogen receptor alpha function in mammary gland development, Proc. Natl. Acad. Sci. U. S. A. 104 (2007) 5455–5460, https://doi.org/10.1073/pnas.0611647104.
[60] S. Naylor, et al., Retroviral expression of Wnt-1 and Wnt-7b produces different effects in mouse mammary epithelium, J. Cell Sci. 113 (Pt 12) (2000) 2129–2138, https://doi.org/10.1242/jcs.113.12.2129.
[61] C. Brisken, et al., Essential function of Wnt-4 in mammary gland development downstream of progesterone signaling, Genes Dev. 14 (2000) 650–654.
[62] K. Roarty, R. Serra, Wnt5a is required for proper mammary gland development and TGF-beta-mediated inhibition of ductal growth, Development 134 (2007) 3929–3939, https://doi.org/10.1242/dev.008250.
[63] W.V. Ingman, S.A. Robertson, Mammary gland development in transforming growth factor beta1 null mutant mice: systemic and epithelial effects, Biol. Reprod. 79 (2008) 711–717, https://doi.org/10.1095/biolreprod.107.067272.
[64] A.E. Gorska, et al., Transgenic mice expressing a dominant-negative mutant type II transforming growth factor-beta receptor exhibit impaired mammary development and enhanced mammary tumor formation, Am. J. Pathol. 163 (2003) 1539–1549, https://doi.org/10.1016/s0002-9440(10)63510-9.
[65] C. Mukhopadhyay, X. Zhao, D. Maroni, V. Band, M. Naramura, Distinct effects of EGFR ligands on human mammary epithelial cell differentiation, PLoS One 8 (2013), e75907, https://doi.org/10.1371/journal.pone.0075907.
[66] M. Voutilainen, et al., Ectodysplasin regulates hormone-independent mammary ductal morphogenesis via NF-κB, Proc. Natl. Acad. Sci. U. S. A. 109 (2012) 5744–5749, https://doi.org/10.1073/pnas.1110627109.
[67] S.M. Rosenfield, et al., Pleiotrophin (PTN) expression and function and in the mouse mammary gland and mammary epithelial cells, PLoS One 7 (2012), e47876, https://doi.org/10.1371/journal.pone.0047876.
[68] E. Katsuta, A.A. Maawy, L. Yan, K. Takabe, High expression of bone morphogenetic protein (BMP) 6 and BMP7 are associated with higher immune cell infiltration and better survival in estrogen receptor-positive breast cancer, Oncol. Rep. 42 (2019) 1413–1421, https://doi.org/10.3892/or.2019.7275.
[69] A.H. Charpentier, et al., Effects of estrogen on global gene expression: identification of novel targets of estrogen action, Cancer Res. 60 (2000) 5977–5983.
[70] K. Iwao, et al., Molecular classification of primary breast tumors possessing distinct prognostic properties, Hum. Mol. Genet. 11 (2002) 199–206, https://doi.org/10.1093/hmg/11.2.199.
[71] S. Raulic, Y. Ramos-Valdes, G.E. DiMattia, Stanniocalcin 2 expression is regulated by hormone signalling and negatively affects breast cancer cell viability in vitro, J. Endocrinol. 197 (2008) 517–529, https://doi.org/10.1677/joe-08-0043.
[72] T.S. Steigedal, et al., Nephronectin is correlated with poor prognosis in breast cancer and promotes metastasis via its integrin-binding motifs, Neoplasia 20 (2018) 387–400, https://doi.org/10.1016/j.neo.2018.02.008.
[73] P.M. Ismail, et al., Progesterone involvement in breast development and tumorigenesis–as revealed by progesterone receptor "knockout" and "knockin" mouse models, Steroids 68 (2003) 779–787, https://doi.org/10.1016/s0039-128x(03)00133-8.
[74] S. Srinivas, et al., Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus, BMC Dev. Biol. 1 (2001) 4, https://doi.org/10.1186/1471-213x-1-4.
[75] A.K. Perl, S.E. Wert, A. Nagy, C.G. Lobe, J.A. Whitsett, Early restriction of peripheral and proximal cell lineages during formation of the lung, Proc. Natl. Acad. Sci. U. S. A. 99 (2002) 10482–10487, https://doi.org/10.1073/pnas.152238499.
[76] A. Dobin, et al., STAR: ultrafast universal RNA-seq aligner, Bioinformatics 29 (2013) 15–21, https://doi.org/10.1093/bioinformatics/bts635.
[77] S. Anders, P.T. Pyl, W. Huber, HTSeq–a Python framework to work with high-throughput sequencing data, Bioinformatics 31 (2015) 166–169, https://doi.org/10.1093/bioinformatics/btu638.
[78] E.Y. Chen, et al., Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, BMC Bioinf. 14 (2013) 128, https://doi.org/10.1186/1471-2105-14-128.
[79] M.V. Kuleshov, et al., Enrichr: a comprehensive gene set enrichment analysis web server 2016 update, Nucleic Acids Res. 44 (2016) W90–W97, https://doi.org/10.1093/nar/gkw377.
[80] S. Babicki, et al., Heatmapper: web-enabled heat mapping for all, Nucleic Acids Res. 44 (2016) W147–W153, https://doi.org/10.1093/nar/gkw419.
[81] F. Cunningham, et al., Ensembl 2022, Nucleic Acids Res. 50 (2022) D988–d995, https://doi.org/10.1093/nar/gkab1049.
[82] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, Bioinformatics 30 (2014) 2114–2120, https://doi.org/10.1093/bioinformatics/btu170.
[83] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (2012) 357–359, https://doi.org/10.1038/nmeth.1923.
[84] Y. Zhang, et al., Model-based analysis of ChIP-seq (MACS), Genome Biol. 9 (2008) R137, https://doi.org/10.1186/gb-2008-9-9-r137.
[85] A.R. Quinlan, I.M. Hall, BEDTools: a flexible suite of utilities for comparing genomic features, Bioinformatics 26 (2010) 841–842, https://doi.org/10.1093/bioinformatics/btq033.
[86] M.D. Robinson, A. Oshlack, A scaling normalization method for differential expression analysis of RNA-seq data, Genome Biol. 11 (2010) R25, https://doi.org/10.1186/gb-2010-11-3-r25.
[87] C.Y. McLean, et al., GREAT improves functional interpretation of cis-regulatory regions, Nat. Biotechnol. 28 (2010) 495–501, https://doi.org/10.1038/nbt.1630.