How to Use Online Tools to Generate New Hypotheses for Mammary Gland Biology Research: A Case Study for *Wnt7b*

Yorick Bernardus Cornelis van de Grift[1] ⬤ · **Nika Heijmans**[1] ⬤ · **Renée van Amerongen**[1] ⬤

Abstract

An increasing number of '-omics' datasets, generated by labs all across the world, are becoming available. They contain a wealth of data that are largely unexplored. Not every scientist, however, will have access to the required resources and expertise to analyze such data from scratch. Fortunately, a growing number of investigators is dedicating their time and effort to the development of user friendly, online applications that allow researchers to use and investigate these datasets. Here, we will illustrate the usefulness of such an approach. Using regulation of *Wnt7b* expression as an example, we will highlight a selection of accessible tools and resources that are available to researchers in the area of mammary gland biology. We show how they can be used for in silico analyses of gene regulatory mechanisms, resulting in new hypotheses and providing leads for experimental follow up. We also call out to the mammary gland community to join forces in a coordinated effort to generate and share additional tissue-specific '-omics' datasets and thereby expand the in silico toolbox.

Keywords Wnt signaling · CTNNB1 · Beta-catenin · In silico analysis · Wnt7b · Gene regulation

Introduction

The experimental technology that allows genome wide analyses at the molecular level (genomics, epigenomics, transcriptomics, metabolomics and proteomics – hereafter combinedly referred to as 'omics' approaches) continues to evolve at breathtaking speed. Despite the fact that these techniques are becoming more affordable and therefore more widely available for scientists worldwide, they are still quite expensive – a prohibitory factor for those with limited financial resources. This is especially true for sophisticated approaches such as single-cell RNA sequencing (scRNAseq) and other-single cell approaches that are still being developed. Moreover, not everyone will have local access to the required infrastructure. Of course, scientific collaborations can offer a solution. Even then, it can be a challenge to integrate a variety of these technologies into one's research program [1].

As can be gleaned from the published literature, all too frequently only a few hits or top candidates are followed up in instances where genome-wide datasets are generated. As a consequence, a wealth of data remains unexplored. These datasets constitute a rich and valuable resource for the larger scientific community. As an example, we have previously used published microarray data to identify the most stably rather than the most differentially expressed genes, resulting in a new set of reference genes for qRT-PCR studies in the developing mouse mammary gland [2].

Most 'omics' datasets are deposited in public repositories such as the NCBI Gene Expression Omnibus (https://www.ncbi.nlm.nih.gov/geo/), either in raw format or in a more processed form. While this makes them available to all scientists in theory, in practice not everyone has the bioinformatics skills and expertise to analyze these data from scratch. Fortunately, multiple labs are dedicating their time and effort to the development of online tools that allow easy and intuitive access to these datasets, allowing researchers to explore them via a user friendly graphical interface.

Here we will highlight a selection of these online tools and demonstrate how they can be used to generate hypotheses and answer biological questions in the context of mammary gland biology. To illustrate this approach, we will build a case study around *Wnt7b*, a gene that has been implicated in

✉ Renée van Amerongen
  r.vanamerongen@uva.nl

1 Developmental, Stem Cell and Cancer Biology, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH Amsterdam, the Netherlands

mammary gland development and breast cancer, but whose precise activity and mode of regulation remain unknown.

We assume that the reader is familiar with the basic principles behind the different techniques (e.g. scRNAseq, snATACseq, Hi-C), as well as with the way in which these data are commonly presented (e.g. tSNE plots). Please note that for all figures we have kept the exact style and color schemes as generated by the different online tools to aid the reader in recognizing the output when they try out these tools for themselves.

WNT7B in Mammary Gland Development and Breast Cancer

WNT7B is expressed in human breast tissue and its expression has been reported to be altered in breast cancer [3, 4]. Its overexpression has been associated with a poor prognosis and reduced overall survival of breast cancer patients [5]. In breast cancer, *WNT7B* has not only been shown to be expressed by the tumor cells, but also by myeloid cells present in the local microenvironment. The latter promotes angiogenesis, invasion and metastasis [6].

Its murine counterpart, *Wnt7b*, is expressed in the ductal epithelium of the mouse mammary gland [7]. The levels of *Wnt7b* remain unaltered following ovariectomy, suggesting that regulation of *Wnt7b* expression is estrogen and progesterone independent [7]. During puberty, expression of *Wnt7b* is enriched in the terminal end bud epithelium, suggesting a role in branching morphogenesis [8]. *Wnt7b* has been reported to have mild transforming activities in vitro [9, 10] and in vivo [11] although not all studies agree on the extent of this effect [10, 12].

The precise role and regulation of *Wnt7b/WNT7B* in the mammary gland or breast remain unknown. So far, evidence that WNT7B protein can promote the activation of CTNNB1/TCF transcriptional complexes is lacking, despite the fact that *Wnt7b* is readily detected and shows prominent expression in luminal cells [13]. This is in contrast to other tissues, such as the skin, where the activities of WNT7B have been linked to CTNNB1/TCF driven processes [14].

Exploring Spatiotemporal Patterns of *Wnt7b* Expression Using scRNAseq Data

Public scRNAseq datasets are an ideal platform to start investigating spatiotemporal gene expression in the mammary gland [15, 16]. We want to highlight three user friendly scRNAseq tools that allow analysis of the in vivo expression patterns of a gene of interest in the postnatal stages of mouse mammary gland development (Box 1). Their combined use

reveals extensive details about the expression pattern of any given gene across different stages and cell populations.

Box 1: Online scRNA-seq Visualization Tools

https://marionilab.cruk.cam.ac.uk/mammaryGland/
    (Bach et al*., 2017, Nature Communications* [15]*).*
    scRNAseq dataset that contains EPCAM + sorted cells from multiple stages of the adult mouse mammary gland cycle: (nulliparous (8w), gestation (14.5d), lactation (6d) and post-involution (11d). Expression of a gene of interest can be investigated in the context of an inferred cell type or developmental stage. Results are visualized as a tSNE plot and a box plot, both illustrating gene expression by cluster. Gene expression can also be displayed along the (luminal) differentiation trajectory in pseudotime.
    https://tabula-muris.ds.czbiohub.org/
    (The Tabula Muris Consortium, 2018, Nature [16]*).*
    Large compendium of single cell transcriptome data from the mouse, which contains scRNAseq datasets from 20 adult tissues, including the mammary gland. This dataset explicitly includes stromal cells and other cell types from the supportive tissue (e.g. endothelial and immune cells). Of note, all tissues have been processed and analysed by two different protocols: cells were either FACS sorted, or single-cell sorted using microfluidic droplet-capture techniques (used for Fig. 1) and thus sequenced using two different methodologies, providing an innate technical validation of the data when using this tool. Fat pads 2,3 and 4 were processed from virgin mice ((10-15w), and subpopulations were separated by FACS by the following markers: Basal population (CD45⁻, CD31⁻, TER119⁻, $CD49f^{high-med}$, $CD24^{med-low}$), Luminal cells (CD45⁻, CD31⁻, TER119⁻, $CD49f^{med-low}$, $CD24^{high-med}$), mammary repopulating cells (CD45⁻, CD31⁻, TER119⁻, $CD49f^{high}$, $CD24^{med}$), and stromal cells (CD45⁻, CD31⁻, TER119⁻,CD49f⁻, CD24⁻).
    http://bis.zju.edu.cn/MCA/index.html
    (Han X, et al*., 2018, Cell* [18]
    The Mouse Cell Atlas is an online tool that compiles scRNAseq datasets from > 14 mouse tissues, including the mammary gland. This dataset explicitly includes mammary epithelial cells as well as cell types from the supportive tissues. The mammary gland has been sequenced at multiple stages of the adult mammary gland cycle: (virgin, pregnancy, lactation and involution). Cell clusters can be visualised by tSNE or UMAP, and gene expression within each cluster is illustrated as a bar graph. A human companion site is also available, see Table 1.

Wnt7b expression is absent (or at least below the limit of detection) in the fetal mammary gland (E18, Fig. 1a), but emerges postnatally (Fig. 1b-d, Fig. 2a,d, Fig. 3a-d). Its
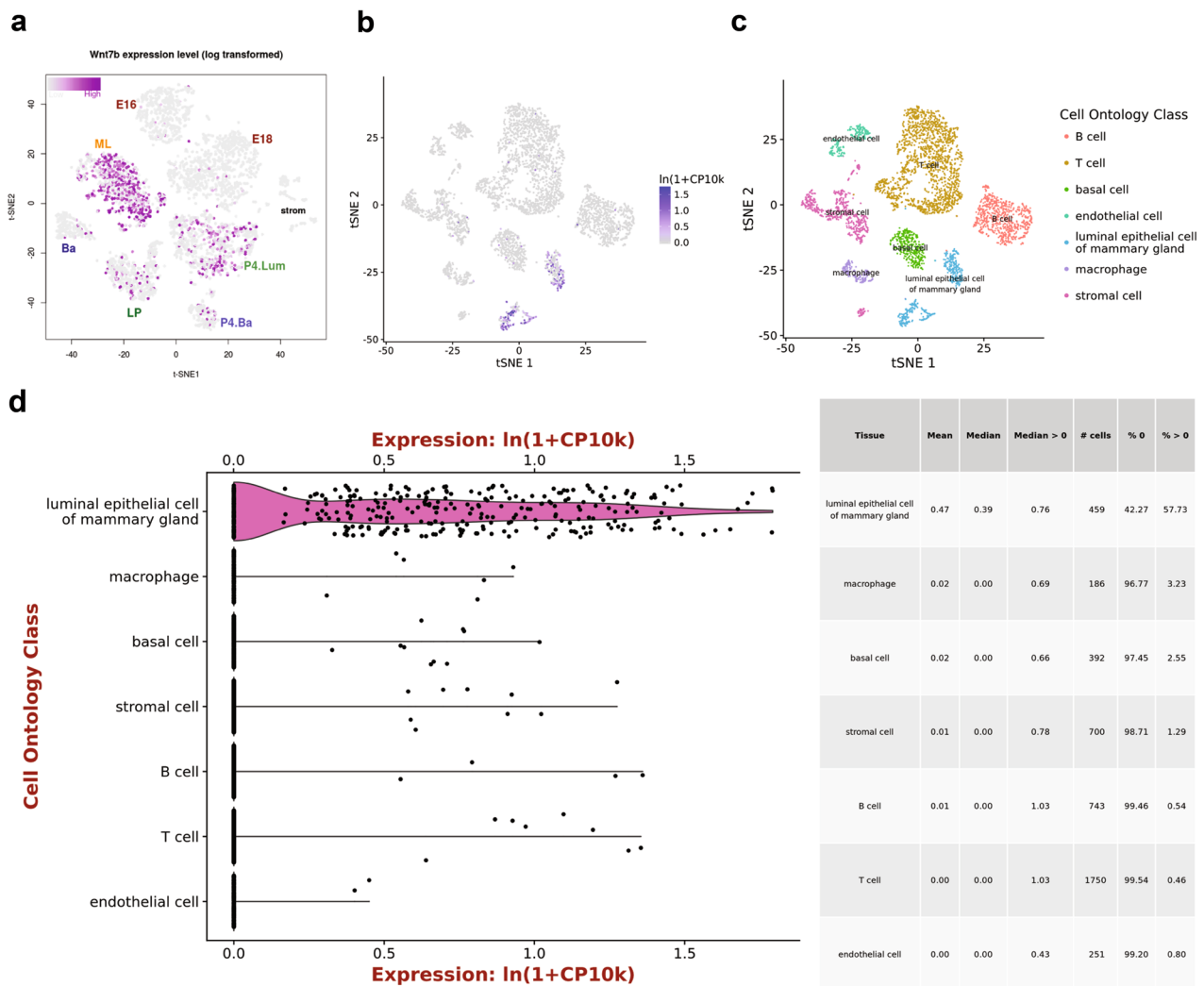
Fig. 1 Single cell RNAseq (scRNAseq) of *Wnt7b* gene expression for all cell types in the mammary gland. **a** tSNE plot of single cell *Wnt7b* expression from FACS-based scRNAseq data from [17]. **b** tSNE plot displaying single cell *Wnt7b* gene expression in virgin mice superimposed on pre-defined cell clusters. Gene expression is normalised to 10,000 counts per cell. **c** tSNE plot defining cell ontology of the cell clusters in **b**. **d** Violin plot of *Wnt7b* gene expression in individual cells in the clusters defined in **c**. Gene expression is normalised to 10.000 counts per cell. Further relevant statistical values for each subpopulation are displayed in a table format. Plots for **a** were generated at https://wahl-lab-salk.shinyapps.io/Mammary_snATAC/ and **b** – **d** at https://tabula-muris.ds.czbiohub.org

expression is cell type specific, displaying high gene expression in the luminal compartment, and low or absent expression in basal cells and supportive tissues (fat, endothelial, immune and stromal cells) (Fig. 1b-d, Fig. 2b-e, Fig. 3a-d).

Spatiotemporal expression is dynamically regulated throughout the adult reproductive cycle (Fig. 2a,d, Fig. 3a-d). In nulliparous mice, *Wnt7b* is expressed in luminal progenitor cells, as well as in more differentiated, hormone-sensing luminal progeny (Fig. 2b-e). During gestation and lactation *Wnt7b* expression is (mostly) switched off, but it re-emerges post-involution (Fig. 2a,d, Fig. 3a-d). Thus, it is exclusively expressed in the 'resting' state, be it nulliparous or post-involution. Of note, although the luminal progenitor

population itself re-appears post-involution, *Wnt7b* expression is lost in this population, becoming restricted to the hormone-sensing luminal lineage post-pregnancy (Fig. 2b-f).

From these analyses we would conclude that *Wnt7b* is expressed exclusively in the luminal compartment of the nulliparous mammary gland, is lost during pregnancy, and is re-established post-involution (Fig. 4). Indeed, this is supported by other studies showing that *Wnt7b* is expressed in the virgin mammary gland, but drops at Preg12.5 of pregnancy to undetectable levels [7]. This underscores the validity of this approach and illustrates the usefulness of interactive in silico tools to determine spatiotemporal patterns of in vivo gene expression.

Table 1 Compilation of publicly available online tools that are outside the scope of the current case study. These tools are not specific for mammary gland biology and/or do not always include mammary gland datasets. Many more tools and resources are available as a stepping stone for those willing to invest the time in developing more advanced bioinformatics and data analysis skills, such as those maintained by the Broad Institute

Tool	Description	Reference
http://asntech.org/dbsuper/index.php	dbSuper is an interactive database containing more than 80,000 putative super enhancers for 25 mouse and >100 human tissues and cell lines. The database has migrated from its original reported location (http://bioinfo.au.tsinghua.edu.cn/dbsuper/) and while functional and highly intuitive, it is not clear whether it has been updated since 2017.	[19]
http://sea.edbc.org	SEA version 3.0 was updated in 2019 and promises to be a comprehensive resource that stores predicted super-enhancers and enhancers from 11 different species and more than 200 types of cells, tissues and diseases.	[20]
https://tabula-muris-senis.ds.czbiohub.org/	A large compendium of single cell transcriptome data from the model organism *Mus musculus* that contains scRNAseq datasets of 23 organs and tissues, including the mammary gland at 6 different timepoints (1 month, 3 months, 18 months, 21 months, 24 months, 30 months). This online dataset explicitly includes stromal cells and other cell types from the supportive tissue (e.g. endothelial and immune cells). Of note, all tissues have been processed and analysed by two different protocols: cells were either FACS sorted, or single-cell sorted using microfluidic droplet-capture techniques and thus sequenced using two different methodologies, providing an innate technical validation of the data when using this tool.	[21]
https://twc-stanford.shinyapps.io/maca/	Also part of the Tabula Muris Senis effort. Offers extensive statistical analysis and visualization of bulk RNA seq datasets from 17 organs of Mus musculus at 10 different timepoints.	[22]
https://www.kobic.kr/3div/	3DIV collects human Hi-C data from 80 cells lines or tissues (including HMEC, MCF7, MCF10A) and promoter capture Hi-C from 27 tissues. Chromatin conformation data from the locus of a gene or location of interest can be either displayed as a Hi-C heatmap and as a virtual 4C (with the location of interest as viewpoint). If applicable, it also predicts the boundaries of local TADs based on the provided datasets. 3DIV offers more flexibility to its users as it allows the user to select the algorithm used to predict TADs, define the cut-off for positive interactions in the virtual 4C and it is straightforward to extract the coordinates of positive hits.	[23, 24]
https://www.ebi.ac.uk/gxa/sc/home	Single Cell Expression Atlas & Gene Expression Atlas: A database that compiles and visualizes published RNA & scRNA-seq datasets from Human, Mouse & a wide variety of model organisms. Selected datasets are plotted as a tSNE, and a heatmap highlighting marker genes for each annotated cluster is displayed. The database can be searched by gene across species, experiments, tissues and cell lines to reveal where this gene is expressed.	[25]
http://bioinfo.vanderbilt.edu/AE/HACER/	HACER is an atlas of Human ACtive Enhancer to interpret Regulatory variants, which includes active, transcribed enhancers derived from GRO-seq, PRO-seq and CAGE data. HACER not only compiles cell type specific enhancers but also integrates transcription factor-enhancer binding prediction, validated chromatin interactions and links GWAS SNPs and eQTL variants to enhancer regions. The database includes the MCF10A and MCF7 cell lines.	[26]
https://www.spatialomics.org/SpatialDB/	An online database that compiles published spatial transcriptomic datasets and offers a web interface for spatially resolved transcriptomic data visualisation and comparison. Includes a human breast cancer dataset.	[27]
http://uofuhealth.utah.edu/huntsman/labs/spike/d3.php	tSNE visualisation of gene expression during mammary gland development: from E16 to Adult.	[17]
https://panglaodb.se/index.html	PanglaoDB is a database that collects and integrates scRNAseq data from human and mouse and presents them through an unified framework.	[28]
http://www.enhanceratlas.org/indexv2.php	The database provides enhancer annotation in nine species, including human (hg19), mouse (mm9), fly (dm3), worm (ce10), zebrafish (danRer10), rat (rn5), yeast (sacCer3), chicken (galGal4), and boar (susScr3). The consensus enhancers were predicted based on multiple high throughput experimental datasets (e.g. histone modification, CAGE, GRO-seq, transcription factor binding and DHS). This database includes the HMEC cell line.	[29, 30]
https://apps.kaessmannlab.org/evodevoapp/	A database visualized by an intuitive shiny app that allows for an interactive exploration of gene expression profiles across tissues, developmental stages and species. This does not only include protein coding genes but also putative LncRNAs. The mammary gland is not included in this dataset.	[31, 32]

Table 1 (continued)

| Tool	Description	Reference
https://amp.pharm.mssm.edu/archs4/data.html	ARCHS4 is a web resource that compiles the majority of RNA-seq data published from both human and mouse datasets and makes that available at the gene and transcript levels. It provides a web-interface that allows exploration of the processed data. Moreover, individual genes can be searched for their average expression across cell lines and tissues, top co-expressed genes, and predicted biological functions and protein-protein interactions.	[33]
http://cistrome.org/db/#/	Cistrome DB is a comprehensive database (~47.000 sets) for curated ChIP and DNase-seq data. It provides an uniform platform that contains manually curated information for each ChIP-seq and DNase-seq dataset, including species, factors, biological source, publication etc, the analysis results of each dataset from human and mouse, and comprehensive quality control checks across the complete database. By using the CistromeDB toolkit, epigenetic features or transcription factors that regulate your gene of interest can be predicted based on the datasets present in Cistrome DB.	[34, 35]
http://bis.zju.edu.cn/HCL/	The Human Cell Landscape offers a large compendium of human scRNA-seq data. Mammary gland tissue is not included in the original dataset, but scRNA-seq data from Nguyen et al. 2018 has been integrated in the online visualisation tool. Gene expression can be visualised superimposed on a tSNE plot.	[36]
https://www.medical-epigenomics.org/papers/krausgrubers2019/#data	A 'gene atlas for structural immunity'. This multi-omics dataset profiles the immuno-logical potential of epithelial cells, endothelial cells and fibroblasts from 12 different mouse tissues. The mammary gland is not included in this dataset. Aggregated ATAC-seq, ChIP-seq and RNAseq can be visualised in the UCSC genome browser.	[37]
https://www.cbioportal.org/	The cBioPortal for cancer genomics is an open-access resource for exploring and visualizing multidimensional cancer genomics datasets. cBioPortal compiles a wide variety of datasets, including TCGA, that can contain non-synonymous mutations, DNA copy-number variation, mRNA and microRNA expression data, protein-level and phosphoprotein level data, DNA methylation, and de-identified clinical data.	[38, 39]
https://kmplot.com/analysis/	The Kaplan-meier plotter is a tool that can be used to assess the effect of 54k genes (mRNA and protein levels) on survival across 21 cancer types, including breast, ovarian, lung and gastric cancer. Sources for these datasets include GEO, TCGA, and EGA. This is a valuable and easy to use tool for discovering and validating cancer survival biomarkers.	[81]
https://maayanlab.cloud/Enrichr/#	Enrichr is a comprehensive online tool or gene set enrichment analysis that includes over 30 gene-set libraries. It offers interactive and intuitive visualisation of the results via clustergrams. Note that while highly useful this tool requires predefined gene sets from e.g. RNA-seq and is not as useful for e.g. searching KEGG or GO terms for a gene of interest. For these kind of queries ARCHS4 is better suited.	[40, 41]

Identifying Putative Regulatory Elements

Little is known about the molecular signals and cis-regulatory elements that control mouse *Wnt7b* or human *WNT7B* gene expression. In ER-/HER2 + breast tumors, *WNT7B* was shown to be a direct transcriptional target of the androgen receptor (AR) [42] and predicted to be regulated by Nuclear respiratory factor 1 (NRF1) [43]. Although *Wnt7b* is also expressed in hormone-responsive cells (Fig. 2 and [13]), at present there is no experimental evidence to support that its expression is regulated by steroid hormones, in particular progesterone [44]. *Wnt7b* expression is not limited to the mammary gland, however. It is required for lung [45, 46], and kidney development [47] to name but a few and can therefore be regulated by a myriad of signals.

One way to gain understanding into tissue-specific gene expression, is to identify cis-acting enhancer elements. Using ChIPseq analysis, a recent study predicted 440 mammary-specific super-enhancers [48]. Super-enhancers can be classified as dense clusters of transcriptional enhancers that are likely to control genes important for cell type specification [47–49]. Only one of these was followed up in more detail in that particular study. However, a supplementary file listing all 440 of these putative regulatory elements is available. We were particularly intrigued by a sequence that spans more than 24 kb on chromosome 15 (published mm9 coordinates chr15: 85,475,778–85,500,063, mm10 coordinates chr15: 85,645,348–85,669,633), which was assigned as a putative regulator of the nearest gene: *Wnt7b* (Fig. 5a). While it is common to do so, linear proximity alone is not an accurate measure for functional interaction between an enhancer and its putative target gene [49, 50]. Other genes in this region – including two miRNAs (*Mirlet7c-2/Mirlet7b*) and a protein coding gene (*Ppara*) – might also be regulated by this particular super-enhancer. A region on the edge of this super-enhancer (mm9 coordinates

Fig. 2 Single cell RNAseq (scRNAseq) of *Wnt7b* gene expression throughout mammary gland development **a** tSNE plot displaying mammary gland developmental timepoints superimposed on pre-defined cell clusters. NP: Nulliparous, G: Gestation, L: Lactation, PI: Post-involution. **b** tSNE plot defining cell ontology (through known marker genes) of cell clusters depicted in in **a** & **d**. See **c** for cell type classification. **c** Dendrogram of clusters based on log transformed mean expression of 15 clusters. The tree was generated by Spearman's rank correlation with Ward linkage. **d** tSNE plot of single cell *Wnt7b* gene log transformed mean expression superimposed on pre-defined clusters. **e** Bargraphs of log transformed mean expression for each 15 clusters. **f** Pseudotime trajectory of the single cell *Wnt7b* log transformed mean expression in the luminal lineage, displaying both the average expression in the hormone sensing and secretory lineages. Each dot represents an individual cell, and the color its associated cluster. Plots were generated at https://marionilab.cruk.cam.ac.uk/mammaryGland/
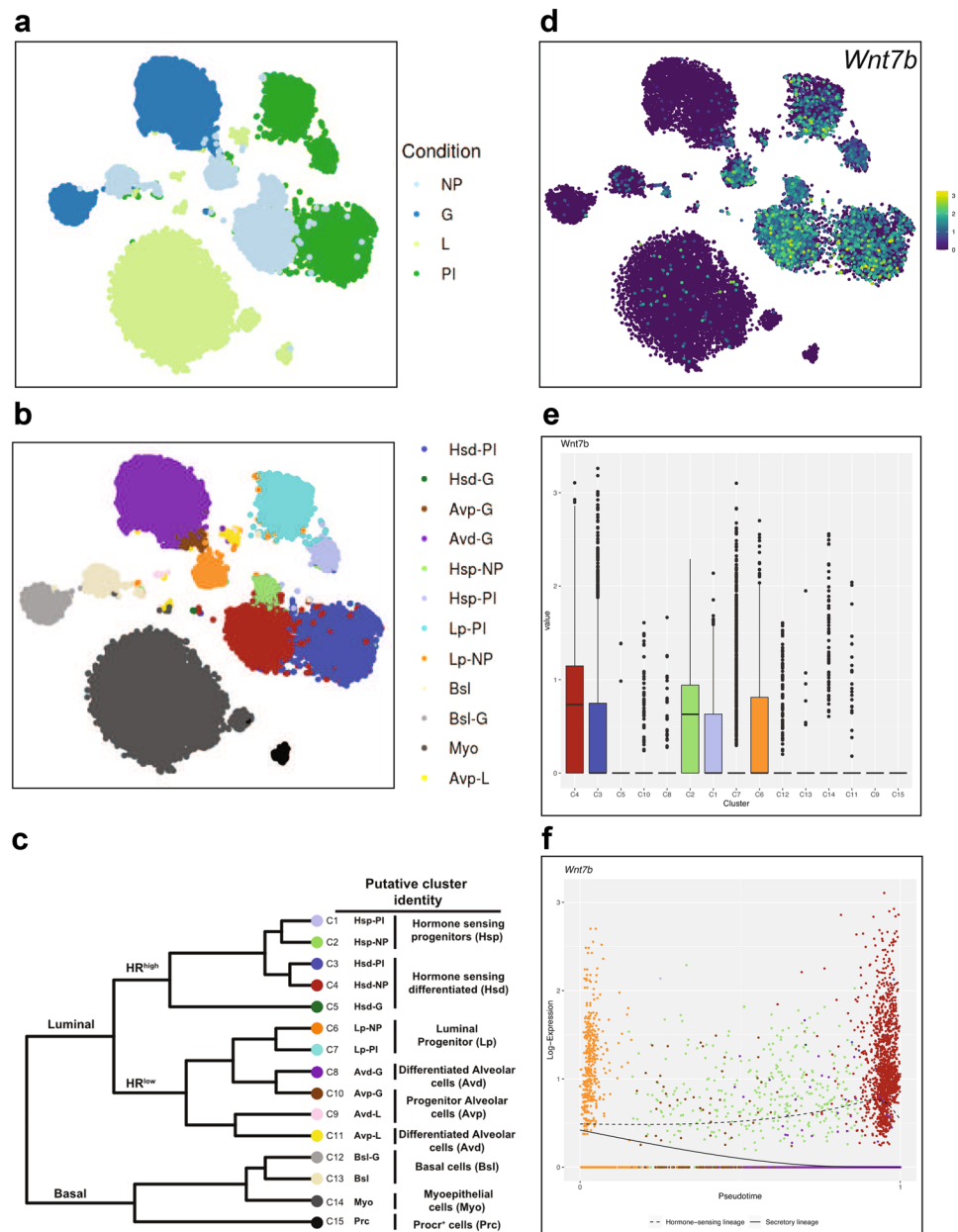

chr15:85,473,689–85,478,592, published mm10 coordinates chr15: 85,643,259–85,648,162) was recently indeed associated with *Wnt7b*, albeit not in the mammary gland but in a mouse model for hair-follicle derived skin tumors, and based on strain-specific polymorphisms rather than on having been shown to directly regulate *Wnt7b* expression [14]. These results show that association of this super-enhancer with *Wnt7b* in the mammary gland is worthy of follow-up analysis.

The term "super-enhancer" is used to define a larger chromatin area that contains clusters of smaller, individual enhancers and that is enriched for active chromatin marks (e.g. H3K27ac) or occupied by transcriptional activators (e.g. MED1) and master regulatory transcription factors (e.g. STAT5A) [48, 51, 52]. More than 80,000 super-enhancers (combined numbers for the mouse and human genome) can be accessed through the online dbSuper database [19]. An updated version of the Super Enhancer Archive (SEA 3.0) provides another entry point [20] (Table 1).

A first screen of the dbSuper database shows the tissue-specificity of super-enhancers: a putative *Wnt7b* super-enhancer has also been identified in the murine heart, lung and testis. However, this sequence does not overlap with the mammary-specific super-enhancer described by Shin et al. [48]. Instead, the dbSuper database predicts this particular location to contain two super-enhancers, identified in hair follicle stem cells, linked to *Mirlet7c-2/Mirlet7b* [53]. Additional super-enhancers in this

Fig. 3 ScRNAseq of *Wnt7b* gene expression in the mammary gland throughout the adult reproductive cycle (virgin, pregnancy, lactation, involution). **a** tSNE plot displaying all identified cell clusters in the virgin mammary gland (top) together with the Log2 *Wnt7b* expression as mean TPM per cluster (bottom). Cluster 10 corresponds to *Krt19* high luminal cells and cluster 17 to *Csn3* high Luminal cells. **b** Same as in **a**, but now during pregnancy. Cluster 1 corresponds to *Wfdc12* high luminal cells, cluster 3 to *Ly6d* high B cells, cluster 4 to *Csn1s1* high epithelial cells, and cluster 14 to *Calca* high luminal cells. **c** Same as in **a**, but now for the lactating mammary gland. *Wnt7b* is not expressed at this stage. **d** Same as in **a**, but now during involution. Cluster 1 corresponds to *Trf* high secretory alveolar cells, cluster 6 to *Pf4* high macrophages, cluster 9 to *Fxyd2* high luminal cells, and cluster 16 to *Csn2* high alveolar cells. All plots were generated at http://bis.zju.edu.cn/MCA/index.html
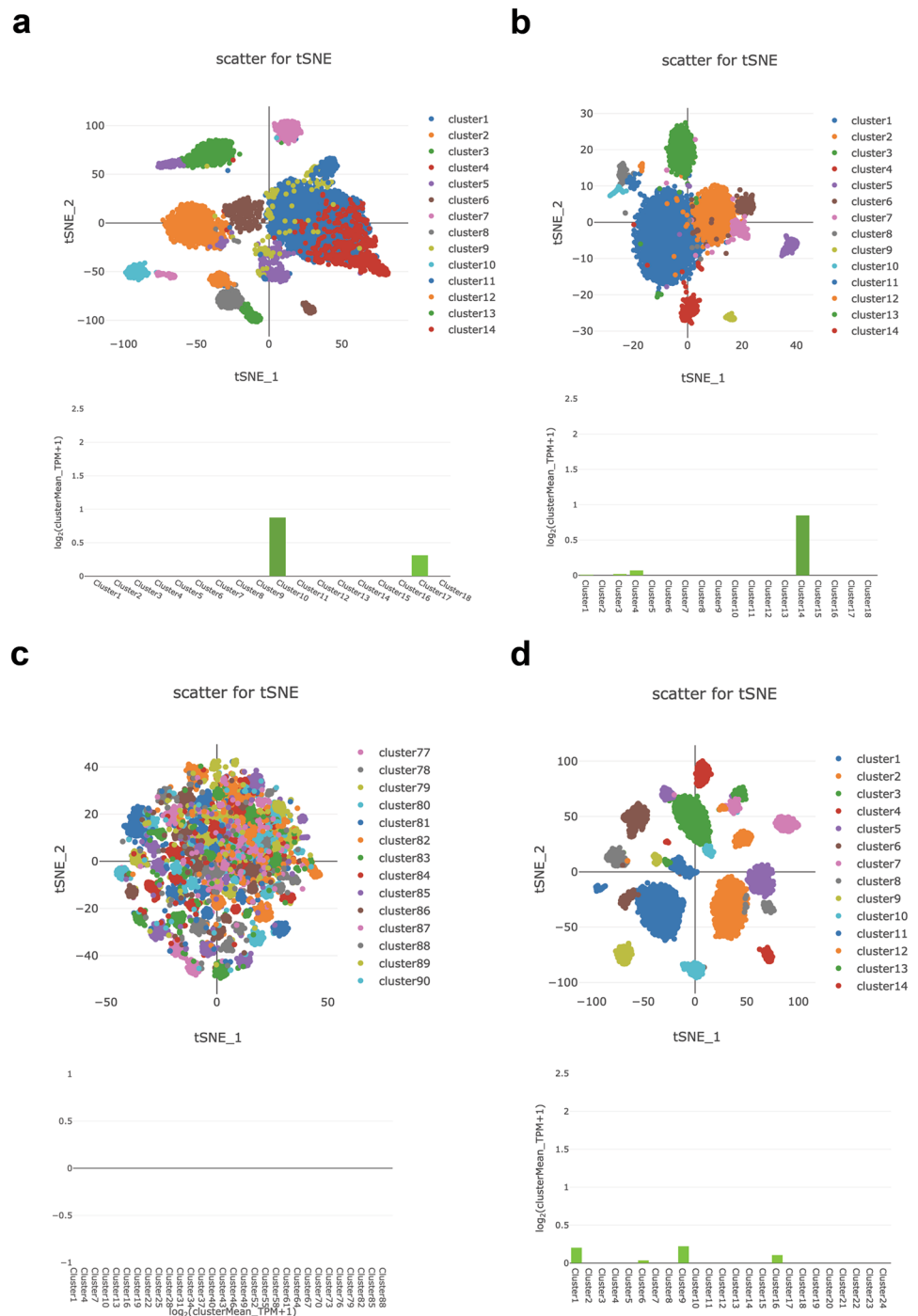

region, identified in the kidney and the liver, are tentatively associated with *Ppara* (Fig. 5b). It should be noted that also in dbSuper, super-enhancers and their associated genes are linked based on a simple proximity rule to the nearest transcriptional start site (TSS) [19]. Out of the genes located in this ~500 kb area on chromosome 15, only *Atxn10* and *Wnt7b* show prominent expression in one or more mammary gland cell subpopulations, although *Ppara*, *Mirlet7c-2/b* and a non-coding RNA, *Lncppara*, may be differentially expressed at low levels (Fig. 5c,d).

Determining the Boundaries of the *Wnt7b* Regulatory Domain

In recent years, it has become generally accepted that regulatory elements control target gene expression within the confines of larger, structurally ordered regions of the chromatin known as topologically associating domains (TADs) [54]. Specific DNA sequences (i.e. regulatory elements and their target genes) are much more likely to

Fig. 4 Graphic summary of mouse *Wnt7b* expression dynamics based on the scR-NAseq data from Figs. 1, 2 and 3. Drawn by the authors
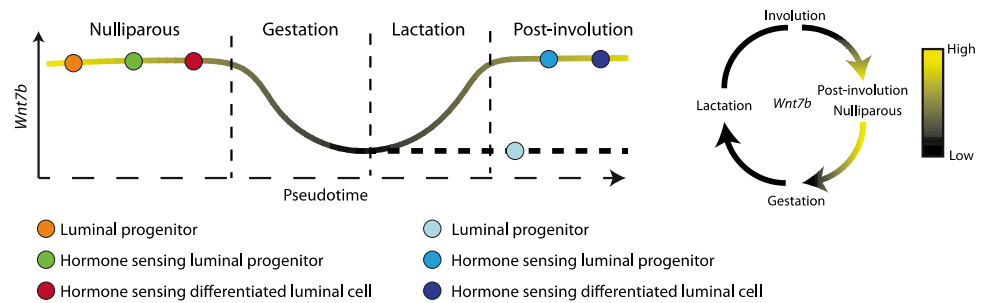
interact within a TAD, than across a TAD boundary. A logical next step in exploring the potential regulation of *Wnt7b* by the aforementioned mammary-specific super-enhancer would therefore be to determine the boundaries of the *Wnt7b* TAD.

We used the 3D Genome Browser (Box 2) to visualize TAD predictions of the *Wnt7b* locus using publicly available Hi-C datasets [55]. In this browser, TAD boundary predictions are calculated according to the so-called directionality index, which is a method that looks at the degree of up- and downstream interaction bias for DNA regions [56]. It was noted that DNA regions at the periphery of TADs are highly biased in their direction
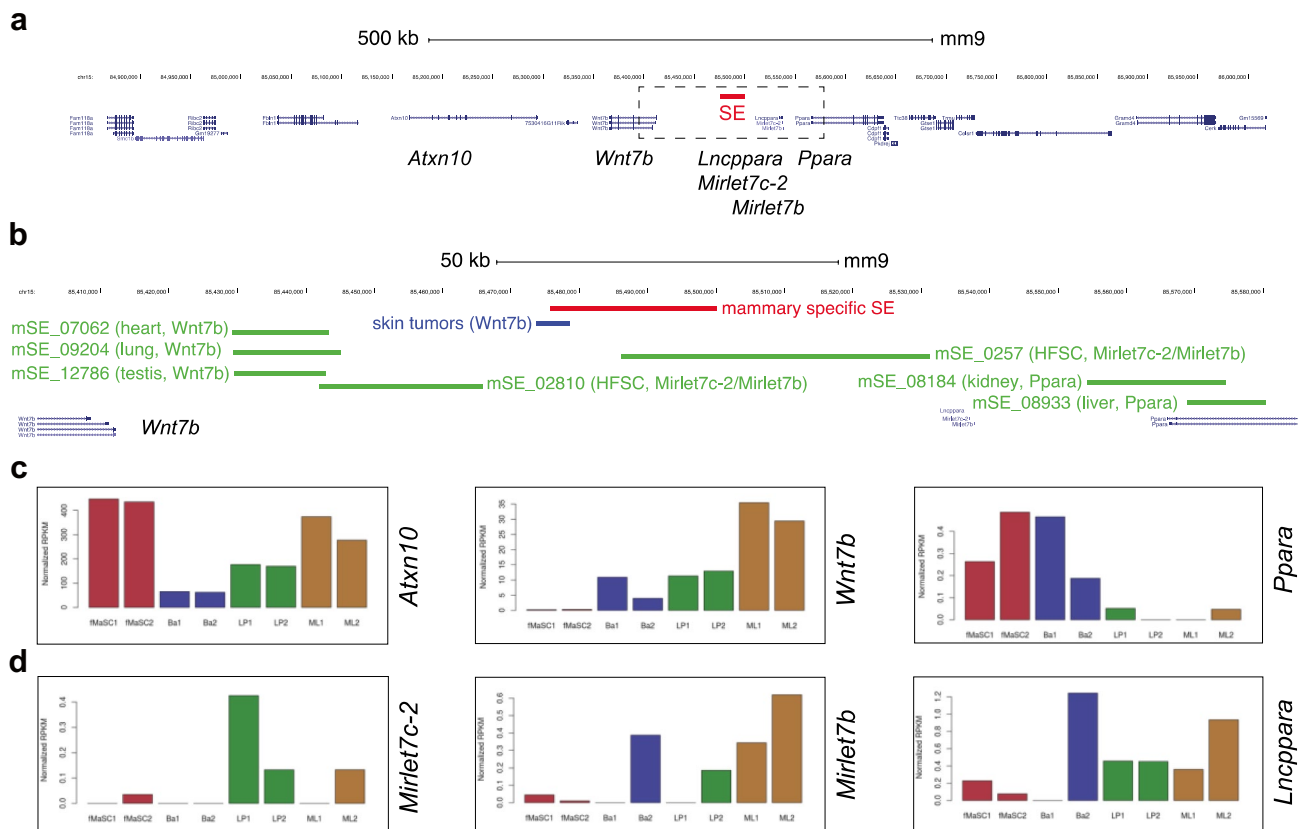

Fig. 5 Overview of super-enhancers assigned to genes in the vicinity of *Wnt7b*. **a** Location of a mammary-specific super-enhancer (SE, in red) on mouse chromosome 15. Scale bar is 500 kb. **b** Close up of the region boxed in **a**. The mammary specific SE is shown in red. A *Wnt7b* associated regulatory region in skin tumors is highlighted in blue. Other super-enhancers in this genomic region, listed in Superdb, are depicted in green. The tissue of origin in which they were identified and the genes to which they have been associated based on proximity rules are indicated. HFSC = hair follicle stem cells. Scale bar is 50 kb. **c** Gene expression of annotated genes in the vicinity of the mammary gland specific SE for all epithelial mammary gland subpopulations. **d** Expression of putative non-coding RNAs in the vicinity of the mammary gland specific SE for all epithelial mammary gland subpopulations. **c-d** show normalized RPKM values. Plots for **c** and **d** were generated at https://wahl-lab-salk.shinyapps.io/Mammary_snATAC/

of interaction. Upstream regions in a TAD are highly biased towards interacting with downstream regions and vice versa. Using this directional bias, the boundaries of adjacent TADs can be predicted. Their coordinates are provided by the 3D Genome Browser, which also includes an intuitive visual reference (Fig. 6).

Box 2: Chromatin Conformation Capture Hi-C Data

http://3dgenome.fsm.northwestern.edu/

(built by the Yue lab, described in Wang et al. (2018), The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions [55]).

The 3D Genome Browser compiles published Hi-C and capture Hi-C datasets from both mouse and human cell lines or tissues, including HMEC. Chromatin conformation data from the locus of a gene or location of interest can either be displayed as a Hi-C heatmap or as a virtual 4C (with the location of interest as viewpoint). Where applicable, it will predict the boundaries of local TADs based on the provided dataset. In the upper menu bar are several options for visualizing data: In "HiC" you can visualize the data from different papers/datasets. In "Compare HiC" you can compare TADs from two different datasets. The coordinates of different TADs can also be downloaded in text file format for hg19, hg38, mm9 and mm10.


Only one mammary-specific Hi-C dataset is currently available, derived from human mammary epithelial cells (HMEC) [57]. However, TADs have been reported to be stable across cell types and even species [56, 60]. Although not all TAD boundaries are equally stable [61], TAD organization can therefore also be investigated using Hi-C datasets generated from a different tissue as input.

According to this analysis, the *Wnt7b* TAD boundary lies immediately upstream of the *Wnt7b* TSS in both HMECs and mouse lymphoma cells (Fig. 6a,b). This would imply that the mammary-specific super-enhancer identified by Shin et al. lies outside of the predicted *Wnt7b* TAD, which makes it less likely that this particular super-enhancer directly regulates the expression of *Wnt7b*. However, in other Hi-C datasets this TAD boundary is less well defined (Fig. 6c,d).

Discovering Novel Regulatory Interactions

To gain a better understanding of how the spatiotemporal expression of *Wnt7b* is regulated in the adult mammary gland, we can start by probing the epigenetic state of the *Wnt7b* locus in an R shiny app published by the Wahl lab (Box 3). This tool not only allows chromatin accessibility and relevant histone modifications to be examined, but also can be used to make predictions about specific promoters and their regulatory sequences of interest. An attractive

Fig. 6 TAD boundary prediction using the 3D Genome Browser. In the depicted triangle, the physical interaction frequency of DNA regions is represented by the color intensity. A dark blue spot can be observed that connects the TAD boundaries of the predicted *Wnt7b* TAD, indicating that these genomic regions were found to frequently interact in this Hi-C dataset. Alternating beige and grey blocks (depicted in between the chromosome coordinates and the genes) depict individual TAD predictions. **a** Hi-C data from human mammary epithelial cells, HMEC [57]. **b** Hi-C data from mouse lymphoma cells, CH12 [57]. **c** Hi-C data from mouse embryonic stem cells, mESC [58]. **d** Hi-C data from mouse myoblasts [59]. Plots were generated at http://3dgenome.fsm.northwestern.edu/
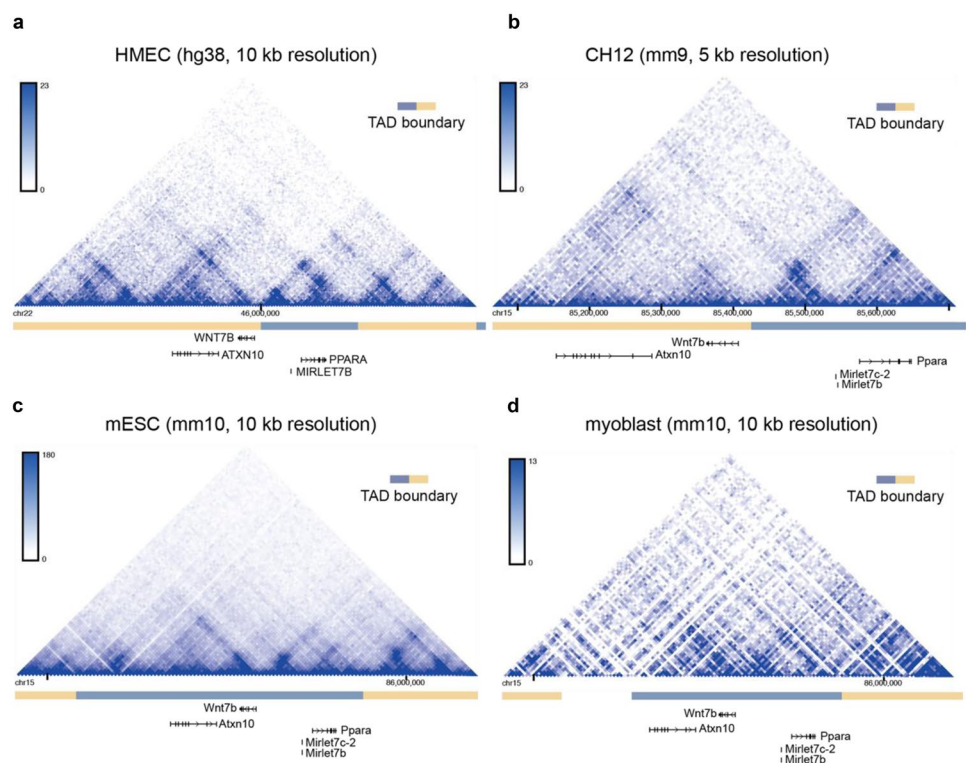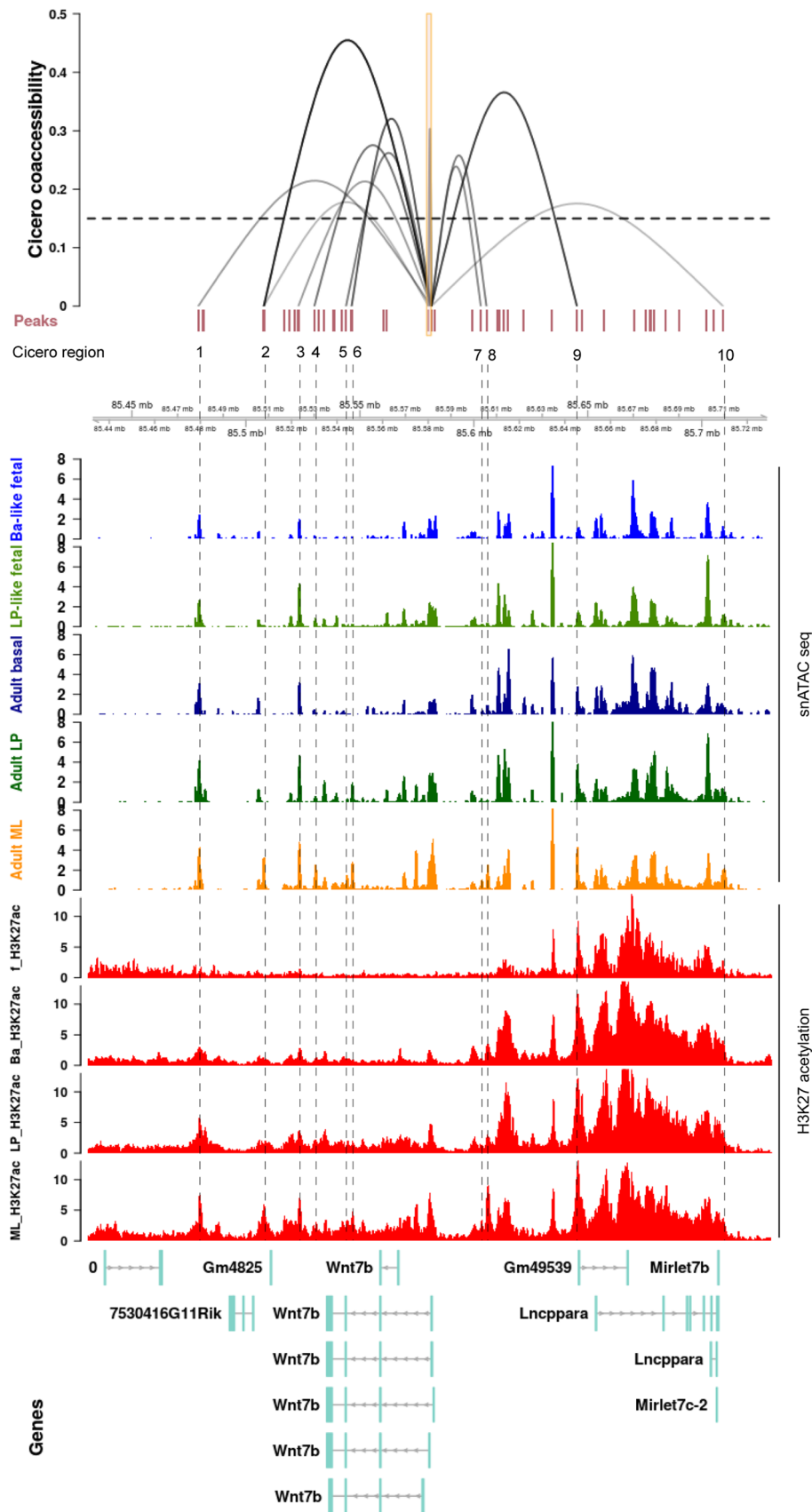
◀**Fig. 7** Overview of chromatin accessibility, genomic interactions and the epigenetic status of the *Wnt7b* locus. Top) Regions identified by Cicero as having a higher co-accessibility score than 0.15 are displayed as interacting loops with the *Wnt7b* promoter. The height of the loop indicated the corresponding Cicero score. The viewpoint size at the promoter is a stretch of 1000 bp. Middle) The 5 snATAC tracks display the aggregated snATAC signal from [62] at the *Wnt7b* locus for each epithelial mammary gland subpopulation. Ba-like fetal: Basal-like fetal, LP-like fetal: Luminal Progenitor-like fetal, Ba-like: Basal-like, LP: Luminal Progenitor, ML: Mature Luminal. Bottom). The 4 H3K27 acetylation tracks display bulk ChIPseq of FACS sorted cells from [63] at the *Wnt7b* locus for each epithelial mammary gland subpopulation. ML_H3K27ac: Mature Luminal, LP_H3K27ac: Luminal Progenitor, Ba_H3K27ac: Basal, f_H3K27ac: fetal. All the data is aligned to mm10 and displayed in a window size of 300 kb. Note that *Wnt7b* is oriented in the reverse orientation (i.e. expressed from the minus strand). Regulatory regions 1–6 to the left of *Wnt7b* are therefore downstream of the promoter and regions 7–10 to the right of *Wnt7b* are upstream. Distal elements that are predicted to interact with the *Wnt7b* promoter and alter their epigenetic status in accordance to *Wnt7b* expression can have the potential to be involved in the spatial temporal regulation of *Wnt7b* in the mammary gland and therefore warrant further investigation. The shiny app offers an intuitive and interactive visual tool to quickly compare numerous epigenetic features, and identify novel regions of interest. It should be noted that no statistical analysis or specific coordinates are provided, although these are available in supplementary data and the GEO accession file. Hence, it serves as an excellent hypothesis generating tool that requires further validation either by in silico analysis or experimentation

graphical interface allows intuitive interpretation of the data (Fig. 7).

Box 3: Probing Chromatin Accessibility and Epigenetic Interactions

https://wahl-lab-salk.shinyapps.io/Mammary_snATAC/
(Chung et al*., 2019, Cell Reports* [62]*).*

The R shiny app published by the Wahl lab combines bulk RNAseq and H3K27 acetylation ChIPseq data from [63] with single-nucleus ATACseq (snATACseq) from [62] and scRNAseq data from [17] in an online web interface that allows its users to investigate numerous (epi)-genetic feature in fetal mammary stem cells (E18 fMaSCs), basal, luminal progenitor and mature luminal cells. This allows researchers to investigate single cell expression & chromatin state (accessibility in the case of snATACseq and active enhancer marks in the case of H3K27Ac ChIPseq) of their gene of interest, and to follow expression of the gene along a pseudotime trajectory. Moreover, if this gene is a transcription factor, its activity can be predicted for each subpopulation based on motif enrichment in open chromatin regions from snATACseq data. Lastly, based on co-accessibility of distal sites and promoter regions in single cells

promoter-enhancer interactions for the gene of interest can be predicted using the so-called Cicero algorithm [64], and concurrently displayed with chromatin accessibility scores and H3K27Ac from aggregate snATACseq and bulk ChIPseq data. In the online tool, Cicero makes predictions in a region of max. 300 kb (with 150 kb upstream and 150 kb downstream of the viewpoints).

If we focus our attention on the *Wnt7b* promoter and gene region (i.e. the center portion of Fig. 7), snATACseq reveals that the chromatin is relatively accessible in all mammary cell type subpopulations irrespective of *Wnt7b* gene expression levels (Fig. 7, top 5 rows). In contrast, H3K27ac of the *Wnt7b* promoter and gene region is exclusively enriched in the luminal compartment (Fig. 7, bottom 4 rows in red). This suggests that *Wnt7b* is 'primed' and open in all epithelial cells in the mammary gland, but its potential for increased gene expression is only realized in the luminal compartment where the chromatin displays the proper histone acetylation marks.

Combining the Cicero algorithm (see Box 3) with snATACseq data, this online tool can also be used to infer co-accessibility of distal sites and the promoter of their putative genes in individual cells. In this manner, Cicero can predict cis-regulatory elements that would be able to interact with the *Wnt7b* promoter in vivo. At a co-accessibility threshold of 0.15, Cicero identifies 10 regions within 150 kb up- or downstream of the viewpoint that interact with the promoter of *Wnt7b*. Of these, 4 are located upstream of *Wnt7b* in an area dense with H3K27ac that encompasses, but extends beyond, the super-enhancer region, and 6 are located downstream of *Wnt7b* (Fig. 7).

The interacting regions depicted to the left of the *Wnt7b* promoter (regions 1–6, located 3′ distal to the TSS) all fall within in the predicted *Wnt7b* TAD (compare Fig. 6,7). These distal sites are either somewhat enriched for chromatin accessibility or H3K27ac, or a combination of both epigenetic features, in adult luminal progenitor and mature luminal cells compared to the adult basal subpopulation (Fig. 7,8). The 4 regions downstream of *Wnt7b* (7–10) do not display evident changes in chromatin accessibility or H3K27ac when luminal cells are compared to the basal compartment, except for region 8 (Fig. 7,8). Note that the distance between region 9 and 10 spans more than 60 kb, which is considerably larger than the reported size of the mammary-specific super-enhancer. This entire stretch of 60 kb shows characteristic marks of active and open chromatin, suggesting that a much larger collection of regulatory elements may exist in this area (Fig. 7).

Region	Cicero score	snATAC >2				H3K27ac >5			
		Fetal	Basal	Luminal Progenitor	Mature Luminal	Fetal	Basal	Luminal Progenitor	Mature Luminal
1	> 0.2	+	+	+	+	-	-	+	+
2	> 0.4	-	-	-	+	-	-	-	+
3	> 0.2	+	+	+	+	-	-	-	+
4	> 0.25	-	-	-	+	-	-	-	-
5	> 0.25	-	-	-	-	-	-	-	-
6	> 0.3	-	-	+	+	-	-	-	-
7	> 0.2	-	-	-	-	-	-	-	-
8	> 0.25	-	-	-	+	-	-	-	+
9	> 0.35	-	+	+	+	+	+	+	+
10	> 0.15	-	-	-	+	-	-	-	-

Fig. 8 Summary of the epigenetic features of each region identified by Cicero as depicted in Fig. 7. Somewhat high(er) levels of snATAC seq signal are defined above a cut-off of 2 and H3K27 acetylation above a cutoff of 5. Note that the tool does not offer any statistical analysis, and therefore cut-offs were user-defined compared to the total signal in the 300 kb window. They should thus be considered reasonable, but relatively arbitrary and worthy of more in-depth investigation

---

Box 4: Looking for Evolutionary Conservation

http://ecrbrowser.dcode.org

(Ovcharenko et al.*, 2004, Nucleic Acids Research* [65]*).*

This web-based tool enables access to pairwise alignments for the genomes of 13 species and visualizes evolutionary conserved regions (ECRs) in a graphical interface. Users can set their own parameters to select regions with a desired cut-off (in Fig. 9: > 85% sequence identity over > 200 bp). Sequences that are conserved within these chosen parameters are represented as colored peaks. Conservation between species is shown relative to a base genome of choice (in Fig. 9: Hg19). Sequence information from the UCSC Genome Browser (http://genome.ucsc.edu/) [66] can be extracted when selecting the DNA region of interest.

Genome assemblies used in the ECR browser: human: Hg19, Tetraodon: tetNig1, frog: xenTro3, fugu: fr3, zebrafish: danRer7, chicken: galGal3, opossum: monDom5, rat: rn4, mouse: mm10, cow: bosTau6, dog: canFam2, chimpanzee: panTro3, rhesus macaque: rheMac2.

Exploring Conservation of Putative Regulatory Enhancer Sequences

In previous studies, highly conserved sequences were associated with developmental and transcriptional regulators [67–73]. Given the fundamental role of Wnt signaling not only in vertebrate development [74], but also specifically in mammary gland development and maintenance [75–77], focusing on conserved sequences could be another criteria for the selection of candidate *Wnt7b* enhancers. To identify conserved regions in the vicinity of *Wnt7b*, we used the evolutionary conserved region (ECR) browser (Box 4).

Often, conservation is scored across vertebrate species. However, in an attempt to identify regions that are specifically conserved in mammals, we specifically selected candidate sequences in a region of ~100 kb up- and downstream of the *Wnt7b* TSS that are conserved across mammalian, but not necessarily in non-mammalian vertebrate species available in the ECR browser (Fig. 9).

A Working Model for Follow-Up Studies

Of course, none of these approaches (sequence conservation, histone modification, transcription factor ChIPseq), either by themselves or in combination, are sufficient to definitively link any of these putative regulatory elements to *Wnt7b*. This requires further experimental validation and specific follow up. However, as a prediction tool these combined analyses provide an excellent starting point for dissecting this super-enhancer in more detail. If we put all of the different pieces of information together (Fig. 10), we can draft some hypotheses regarding the regulation of *Wnt7b* expression in the mammary gland.
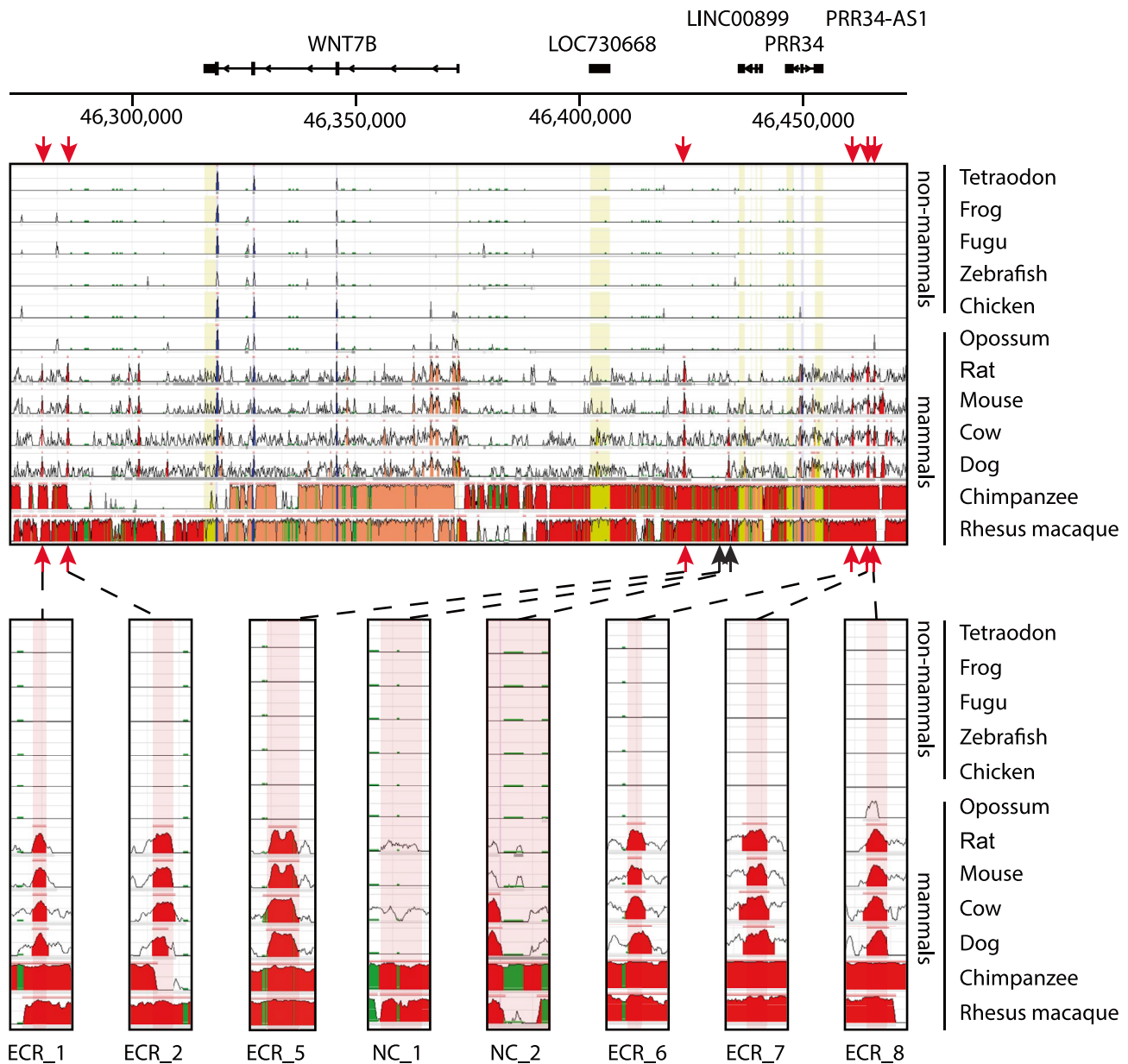
Hg19, chr22:46,273,008-46,473,008

Fig. 9 Selection of candidate enhancers based on sequence conservation. A 200 kb region of the human Hg19 genome assembly, 100 kb up- and downstream of *WNT7B* TSS. In each track the sequence conservation between Hg19 and one of 12 vertebrates is shown. Different colors indicate the following: Red = intergenic, salmon = intragenic, yellow = UTRs, blue = coding sequences, green = transposons and simple repeats. Locations of putative candidate *Wnt7b* enhancers are indicated by red arrows. Lower panels show zoomed in regions of 1000 bp where the candidate enhancers are located. The red shade over the tracks represent the chosen candidate enhancer region. Examples of sequences that are conserved in mammals, but not in non-mammalian vertebrates are shown (ECR_1, ECR_2, ECR_5, ECR_6, ECR_7 and ECR_8) alongside two examples of a non-conserved region (NC_1, NC_2). Parameters used: > 85% sequence identity over > 200 bp

First, we propose that in mammary epithelial cells the proposed TAD boundary immediately upstream of *Wnt7b* (Fig. 6) is not very stable, given that the Cicero algorithm predicts four interactions between the *Wnt7b* promoter and regions to the right of this presumed TAD boundary (i.e. regions 7–10 in Fig. 7). Of note, two of these interactions (Cicero regions 7 and 8) occur in the direct vicinity of this presumed TAD boundary. The other two interactions (Cicero regions 9 and 10) border a large area of active chromatin, which extends beyond
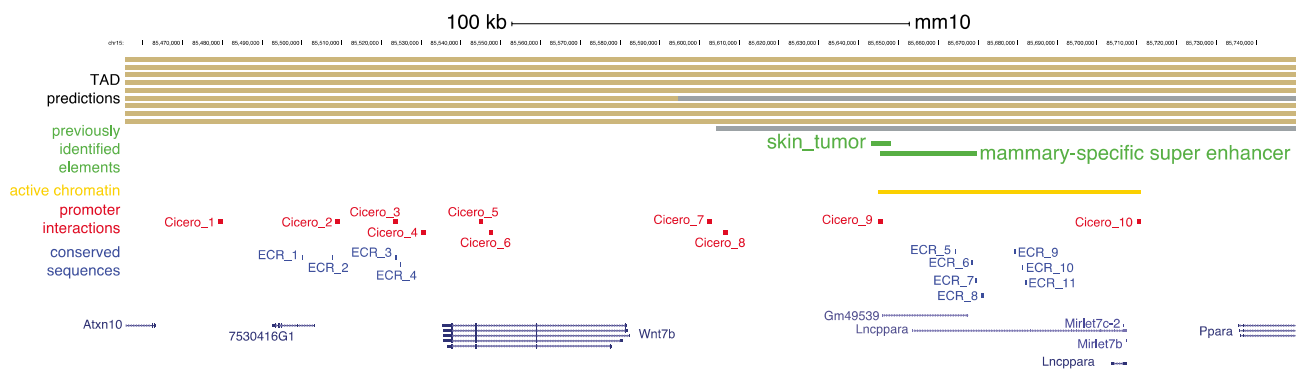
Fig. 10 Integration of the insights obtained from the different analyses. Beige and grey bars at the top represent neighboring TADs as predicted in cortex (liftover from mm9), mESC (liftover from mm9), neurons (mm10), mESC (mm10), NPC (mm10), CH12 (liftover from mm9), cortical neurons (mm10), myoblast (mm10), G1E-ER4 (mm10) and HMEC (liftover from human). Previously identified (super) enhancer elements are depicted in green, the region of active chromatin identified in Figure 7 is depicted in yellow, promoter interactions predicted by Cicero are depicted in red and conserved sequences identified in the ECR browser are depicted in blue. Coding and non-coding genes are shown at the bottom for reference

the super-enhancer region previously identified by Shin et al. [48]. This 60 kb area harbors an annotated lncRNA (*Lncppara*) and two microRNAs, *MirLet7b/MirLet7c-2*, which are broadly expressed and implicated in cancer formation [78–80]. Moreover, this region also contains multiple conserved sequences that could represent functional enhancer elements (including ECR_6, ECR_7 and ECR_8 from Fig. 9).

Second, if we do take the TAD boundary prediction into account, it may be wise to prioritize the interactions that occur between *Wnt7b* and more downstream sequences (i.e. regions 1–6 in Fig. 7). Although the coordinates from the Cicero prediction algorithm deserve further scrutiny of the original datasets, these downstream interacting regions also lie in close vicinity to conserved sequence elements.

Third, in combination with the expression data analysis (Fig. 1-3), the published literature and the active enhancer marks (Fig. 7,8), we can make a further prioritization of putative *Wnt7b* enhancer sequences that are worthy of experimental validation and follow up. In this case, region 2 is particularly interesting as is has the highest Cicero score and displays both differential chromatin accessibility and H3K27 acetylation in the luminal compartment.

To summarize, by using publicly available online tools we assessed the genomic conformation of the *Wnt7b* locus, and how this relates to the previously identified putative *Wnt7b* super enhancer. By examining the epigenetic status of the *Wnt7b* locus more closely, we noticed that although the *Wnt7b* promoter is predicted to interact with the super-enhancer region, this is likely not cell type specific as both chromatin accessibility and H3K27ac do not change between the basal and luminal lineages in

this region. However, regions downstream of *Wnt7b* do change their epigenetic status in accordance to *Wnt7b* gene expression and are also predicted to interact with the *Wnt7b* promoter. This entire area would be worthy of experimental follow up to definitively associate specific regulatory elements with *Wnt7b* and/or other nearby genes – in particular the miRNAs and *Ppara*.

Discussion

Using publicly available genome wide datasets and accessible online tools, we have identified several regions that might play a role in the regulation of spatiotemporal expression of *Wnt7b* in the mouse mammary gland. Our main goal was to show the reader how these findings provide additional information for future investigations. However, we also want to use this opportunity to highlight and stress the added value of making large datasets available to a wide audience through interactive online tools. We thank our colleagues who invest their resources to do so.

At the same time, we call for joint efforts from our community to ensure that the repertoire of tools as well as of accessible datasets continues to grow and remains of high quality and value to investigators worldwide. As others have undoubtedly noticed, mammary gland and breast tissue datasets are often notoriously absent from public, large-scale -omics efforts. Generating and curating additional genome wide datasets (e.g. Hi-C and others) for both epithelial and stromal cells of multiple species, including mouse and human, would be a tremendous resource for our community as a whole. The

careful generation of such datasets in combination with user-friendly online tools provide a valuable resource for researchers, and could in the long run also help to reduce animal experimentation. Certain features will enhance the user experience and promote the wide use of such tools, including the ability to export high resolution graphs (ideally allowing further customization, e.g. PDF format as offered by [15, 55]) and the ability to easily download specific sequences or genome coordinates (as offered by [55, 65]). Given the challenges associated with keeping these databases up to date and operating smoothly, international and consortium efforts that provide sufficient support infrastructure may, in the long term, prove to be essential in this regard.

Here we have shown how the combined use of different online tools can be applied to generate novel hypotheses. Of course, the same tools can also be used to complement existing projects by providing additional data. Ideally, in the not too near future, researchers will have a broad compendium of resources available to them that are of such high quality that they will allow in vivo analyses to be performed in silico, thereby bringing such genome-wide analyses within reach of all scientists. This will only be possible, however, if sufficient tissue-specific datasets can be accessed. Especially in the case of the mammary gland, great care should be taken to include different timepoints to cover both embryonic and postnatal developmental stages, as well as the entire gestational cycle. Here, biological and computational expertise will continually need to go hand in hand to ensure that such online tools can meet the demands of the scientific questions that are being asked.

References

1. Misra BB, Langefeld C, Olivier M, Cox LA. Integrated omics: Tools, advances and future approaches. J Mol Endocrinol [Internet]. 2019 [cited 2020 Jul 23];62:R21–45. Available from: https://jme.bioscientifica.com/view/journals/jme/62/1/JME-18-0055.xml.

2. Van De Moosdijk AAA, Van Amerongen R. Identification of reliable reference genes for qRT-PCR studies of the developing mouse mammary gland. Sci Rep [Internet]. 2016;6:35595. Available from: http://www.nature.com/articles/srep35595.

3. Huguet EL, McMahon JA, McMahon AP, Bicknell R, Harris AL. Differential Expression of Human Wnt Genes 2, 3, 4, and 7B in Human Breast Cell Lines and Normal and Disease States of Human Breast Tissue. Cancer Res. 1994;54:2615–21.

4. Milovanovic T, Planutis K, Nguyen A, Marsh JL, Lin F, Hope C. Expression of Wnt genes and frizzled 1 and 2 receptors in normal breast epithelium and infiltrating breast carcinoma. Int J Oncol. 2004;25:1337–42.

5. Chen J, Liu T-Y, Peng H-T, Wu Y-Q, Zhang L-L, Lin X-H. Up-regulation of Wnt7b rather than Wnt1, Wnt7a, and Wnt9a indicates poor prognosis in breast cancer. Int J Clin Exp Pathol. 2018;11:4552–61.

6. Yeo EJ, Cassetta L, Qian BZ, Lewkowich I, Li JF, Stefater JA. Myeloid precursor 7b mediates the angiogenic switch and metastasis in breast cancer. Cancer Res. 2014;74:2962–73.

7. Weber-Hall SJ, Phippard DJ, Niemeyer CC, Dale TC. Developmental and hormonal regulation of Wnt gene expression in the mouse mammary gland. Differentiation [Internet]. Elsevier; 1994 [cited 2020 Jul 7];57:205–14. Available from: https://www.sciencedirect.com/science/article/pii/S0301468111601618?via%3Dihub.

8. Kouros-Mehr H, Werb Z. Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis. Dev Dyn. 2006;235:3404–12.

9. Wong GT, Gavin BJ, McMahon AP. Differential transformation of mammary epithelial cells by Wnt genes. Mol Cell Biol. 1994;14:6278–86.

10. Naylor S, Smalley MJ, Robertson D, Gusterson BA, Edwards PAW, Dale TC. Retroviral expression of Wnt-1 and Wnt-7b produces different effects in mouse mammary epithelium. J Cell Sci [Internet]. 2000 [cited 2020 Jul 7];113:2129–38.

11. Roarty K, Shore AN, Creighton CJ, Rosen JM. Ror2 regulates branching, differentiation, and actincytoskeletal dynamics within the mammary epithelium. J Cell Biol [Internet]. 2015;208:351–66. Available from: http://www.jcb.org/lookup/doi/10.1083/jcb.201408058.

12. Shimizu H, Julius MA, Giarré M, Zheng Z, Brown AMC, Kitajewski J. Transformation by wnt family proteins correlates with regulation of β- catenin. Cell Growth Differ. 1997;8:1349–58.

13. Cai C, Yu QC, Jiang W, Liu W, Song W, Yu H. R-spondin1 is a novel hormone mediator for mammary stem cell self-renewal. Genes Dev. 2014;28:2205–18.

14. Krimpenfort P, Snoek M, Lambooij J-P, Song J-Y, van der Weide R, Bhaskaran R, et al. A natural WNT signaling variant potently synergizes with Cdkn2ab loss in skin carcinogenesis. Nat Commun [Internet]. 2019;10:1425. Available from: http://www.nature.com/articles/s41467-019-09321-8.

15. Bach K, Pensa S, Grzelak M, Hadfield J, Adams DJ, Marioni JC, et al. Differentiation dynamics of mammary epithelial cells revealed by single-cell RNA sequencing. Nat Commun [Internet]. 2017;8:2128. Available from: http://www.nature.com/articles/s41467-017-02001-5.

16. Schaum N, Karkanias J, Neff NF, May AP, Quake SR, Wyss-Coray T. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. Nature. 2018;562:367–72.

17. Giraddi RR, Chung C-Y, Heinz RE, Perou CM, Wahl GM, Spike BT. Single-Cell Transcriptomes Distinguish Stem Cell State

Changes and Lineage Specification Programs in Early Mammary Gland Development. CellReports [Internet]. 2018 [cited 2020 Jul 15];24:1653–1666.e7. Available from: https://doi.org/10.1016/j.celrep.2018.07.025

18. Han X, Wang R, Zhou Y, Fei L, Sun H, Lai S, Saadatpour A, Zhou Z, Chen H, Ye F, Huang D, Xu Y, Huang W, Jiang M, Jiang X, Mao J, Chen Y, Lu C, Xie J, Fang Q, Wang Y, Yue R, Li T, Huang H, Orkin SH, Yuan G-C, Chen M, Guo G. Mapping the Mouse Cell Atlas by Microwell-Seq. Cell. 2018;172(5):1091-1107.e17

19. Khan A, Zhang X. DbSUPER: A database of Super-enhancers in mouse and human genome. Nucleic Acids Res [Internet]. 2016 [cited 2020 Jul 25];44:D164–71.

20. Chen C, Zhou D, Gu Y, Wang C, Zhang M, Lin X, et al. SEA version 3.0: a comprehensive extension and update of the Super-Enhancer archive. Nucleic Acids Res [Internet]. 2020 [cited 2020 Jul 25];48. Available from: https://academic.oup.com/nar/article-abstract/48/D1/D198/5610346.

21. Almanzar N, Antony J, Baghel AS, Bakerman I, Bansal I, Barres BA. A single-cell transcriptomic atlas characterizes ageing tissues in the mouse. Nature. 2020;583:590–5.

22. Schaum N, Lehallier B, Hahn O, Hosseinzadeh S, Lee SE, Sit R, et al. The murine transcriptome reveals global aging nodes with organ-specific phase and amplitude. bioRxiv. 2019;

23. Yang D, Jang I, Choi J, Kim M-S, Lee AJ, Kim H, et al. 3DIV: A 3D-genome Interaction Viewer and database. Nucleic Acids Res [Internet]. 2018;46:D52–7. Available from: http://academic.oup.com/nar/article/46/D1/D52/4584622.

24. Jung I, Schmitt A, Diao Y, Lee AJ, Liu T, Yang D. A compendium of promoter-centered long-range chromatin interactions in the human genome. Nat Genet. 2019;51:1442–9.

25. Papatheodorou I, Moreno P, Manning J, Fuentes AMP, George N, Fexova S. Expression Atlas update: From tissues to single cells. Nucleic Acids Res. 2020;48:D77-83.

26. Wang J, Dai X, Berry LD, Cogan JD, Liu Q, Shyr Y. HACER: An atlas of human active enhancers to interpret regulatory variants. Nucleic Acids Res. 2019;47:D106-12.

27. Fan Z, Chen R, Chen X. SpatialDB: A database for spatially resolved transcriptomes. Nucleic Acids Res. 2020;48:D233-7.

28. Franzén O, Gan L-M, Björkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database [Internet]. 2019;2019. Available from: https://academic.oup.com/database/article/doi/10.1093/database/baz046/5427041.

29. Gao T, He B, Liu S, Zhu H, Tan K, Qian J. EnhancerAtlas: A resource for enhancer annotation and analysis in 105 human cell/tissue types. Bioinformatics. 2016;32:3543–51.

30. Gao T, Qian J. EnhancerAtlas 2.0: An updated resource with enhancer annotation in 586 tissue/cell types across nine species. Nucleic Acids Res. 2020;48:D58–64.

31. Cardoso-Moreira M, Halbert J, Valloton D, Velten B, Chen C, Shao Y. Gene expression across mammalian organ development. Nature. 2019;571:505–9.

32. Sarropoulos I, Marin R, Cardoso-Moreira M, Kaessmann H. Developmental dynamics of lncRNAs across mammalian organs and species. Nature. 2019;571:510–4.

33. Lachmann A, Torre D, Keenan AB, Jagodnik KM, Lee HJ, Wang L, et al. Massive mining of publicly available RNA-seq data from human and mouse. Nat Commun [Internet]. 2018;9:1366. Available from: http://www.ncbi.nlm.nih.gov/pubmed/29636450.

34. Mei S, Qin Q, Wu Q, Sun H, Zheng R, Zang C. Cistrome Data Browser: A data portal for ChIP-Seq and chromatin accessibility data in human and mouse. Nucleic Acids Res. 2017;45:D658-62.

35. Zheng R, Wan C, Mei S, Qin Q, Wu Q, Sun H. Cistrome Data Browser: Expanded datasets and new tools for gene regulatory analysis. Nucleic Acids Res. 2019;47:D729-35.

36. Han X, Zhou Z, Fei L, Sun H, Wang R, Chen Y. Construction of a human cell landscape at single-cell level. Nature. 2020;581:303–9.

37. Krausgruber T, Fortelny N, Fife-Gernedl V, Senekowitsch M, Schuster LC, Lercher A, et al. Structural cells are key regulators of organ-specific immune responses. Nature [Internet]. 2020;583:296–302. Available from: http://www.nature.com/articles/s41586-020-2424-4.

38. Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA. The cBio Cancer Genomics Portal: An open platform for exploring multidimensional cancer genomics data. Cancer Discov. 2012;2:401–4.

39. Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal. Sci Signal [Internet]. 2013;6:pl1–pl1. Available from: https://stke.sciencemag.org/lookup/doi/10.1126/scisignal.2004088.

40. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles G, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics [Internet]. 2013;14:128. Available from: http://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-128.

41. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z. Enrichr: A comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016;44:W90-7.

42. Ni M, Chen Y, Lim E, Wimberly H, Bailey STT, Imai Y, et al. Targeting Androgen Receptor in Estrogen Receptor-Negative Breast Cancer. Cancer Cell [Internet]. Cell Press; 2011 [cited 2020 Jul 7];20:119–31. Available from: https://www.sciencedirect.com/science/article/pii/S1535610811001966?via%3Dihub#fig3.

43. Ramos J, Das J, Felty Q, Yoo C, Poppiti R, Murrell D. NRF1 motif sequence-enriched genes involved in ER/PR −ve HER2 +ve breast cancer signaling pathways Breast Cancer. Res Treat. 2018;172:469–81.

44. Fernandez-Valdivia R, Mukherjee A, Creighton CJ, Buser AC, DeMayo FJ, Edwards DP. Transcriptional response of the murine mammary gland to acute progesterone exposure. Endocrinology. 2008;149:6236–50.

45. Shu W, Jiang YQ, Lu MM, Morrisey EE. Wnt7b regulates mesenchymal proliferation and vascular development in the lung. Development. 2002;129:4831–42.

46. Rajagopal J, Carroll TJ, Guseh JS, Bores SA, Blank LJ, Anderson WJ. Wnt7b stimulates embryonic lung growth by coordinately increasing the replication of epithelium and mesenchyme. Development. 2008;135:1625–34.

47. Yu J, Carroll TJ, Rajagopal J, Kobayashi A, Ren Q, McMahon AP. A Wnt7b-dependent pathway regulates the orientation of epithelial cell division and establishes the cortico-medullary axis of the mammalian kidney. Development. 2009;136:161–71.

48. Shin HY, Willi M, Yoo KH, Zeng X, Wang C, Metser G, et al. Hierarchy within the mammary STAT5-driven Wap super-enhancer. Nat Genet [Internet]. 2016;48:904–11. Available from: http://www.nature.com/articles/ng.3606.

49. Dowen JM, Fan ZP, Hnisz D, Ren G, Abraham BJ, Zhang LN, et al. Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell [Internet]. 2014;159:374–87.

50. Novo CL, Javierre BM, Cairns J, Segonds-Pichon A, Wingett SW, Freire-Pritchett P. Long-Range Enhancer Interactions Are Prevalent in Mouse Embryonic Stem Cells and Are Reorganized upon Pluripotent State Transition. Cell Rep. 2018;22:2615–27.

51. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153:307–19.

52. Pott S, Lieb JD. What are super-enhancers? Nat. Genet. 2015. p. 8–12.

53. Adam RC, Yang H, Rockowitz S, Larsen SB, Nikolova M, Oristian DS. Pioneer factors govern super-enhancer dynamics in stem cell plasticity and lineage choice. Nature. 2015;521:366–70.

54. Lupiáñez DG, Kraft K, Heinrich V, Krawitz P, Brancati F, Klopocki E, et al. Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. Cell [Internet]. 2015;161:1012–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/25959774.

55. Wang Y, Song F, Zhang B, Zhang L, Xu J, Kuang D, et al. The 3D Genome Browser: A web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol Genome Biology. 2018;19:1–12.

56. Dixon JR, Selvaraj S, Yue F, Kim A, Li Y, Shen Y, et al. Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature Nature Publishing Group. 2012;485:376–80.

57. Rao SSP, Huntley MH, Durand NC, Stamenova EK, Bochkov ID, Robinson JT, et al. A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell. Elsevier Inc.; 2014;159:1665–80.

58. Bonev B, Mendelson Cohen N, Szabo Q, Fritsch L, Papadopoulos GL, Lubling Y. Multiscale 3D Genome Rewiring during Mouse Neural Development. Cell. 2017;171:557–72.

59. Doynova MD, Markworth JF, Cameron-Smith D, Vickers MH, O'Sullivan JM. Linkages between changes in the 3D organization of the genome and transcription during myotube differentiation in vitro. Skelet Muscle [Internet]. 2017;7:5. Available from: http://skeletalmusclejournal.biomedcentral.com/articles/10.1186/s13395-017-0122-1.

60. Krefting J, Andrade-Navarro MA, Ibn-Salem J. Evolutionary stability of topologically associating domains is associated with conserved gene regulation. BMC Biol [Internet]. 2018;16:87. Available from: https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-018-0556-x.

61. McArthur E, Capra J. Topologically associating domain (TAD) boundaries stable across diverse cell types are evolutionarily constrained and enriched for heritability. bioRxiv. 2020;

62. Chung CY, Ma Z, Dravis C, Preissl S, Poirion O, Luna G. Single-Cell Chromatin Analysis of Mammary Gland Development Reveals Cell-State Transcriptional Regulators and Lineage Relationships. Cell Rep. 2019;29:495–510.

63. Dravis C, Chung CY, Lytle NK, Herrera-Valdez J, Luna G, Trejo CL, et al. Epigenetic and Transcriptomic Profiling of Mammary Gland Development and Tumor Models Disclose Regulators of Cell State Plasticity. Cancer Cell [Internet]. 2018;34:466–482.e6. Available from: https://linkinghub.elsevier.com/retrieve/pii/S1535610818303210.

64. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D. Cicero Predicts cis-Regulatory DNA Interactions from Single-Cell Chromatin Accessibility Data. Mol Cell. 2018;71:858–71.

65. Ovcharenko I, Nobrega MA, Loots GG, Stubbs L. ECR Browser: A tool for visualizing and accessing data from comparisons of multiple vertebrate genomes. Nucleic Acids Res [Internet]. 2004;32:280–6. Available from: https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkh355.

66. James Kent W, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM. The human genome browser at UCSC. Genome Res. 2002;12:996–1006.

67. Ahituv N, Prabhakar S, Poulin F, Rubin EM, Couronne O. Mapping cis-regulatory domains in the human genome using multi-species conservation of synteny. Hum Mol Genet. 2005;14:3057–63.

68. Boffelli D, Nobrega MA, Rubin EM. Comparative genomics at the vertebrate extremes. Nat Rev Genet. 2004;5:456–65.

69. Woolfe A, Goodson M, Goode DK, Snell P, McEwen GK, Vavouri T, et al. Highly Conserved Non-Coding Sequences Are Associated with Vertebrate Development. Sean Eddy, editor. PLoS Biol [Internet]. 2004;3:e7. Available from: https://dx.plos.org/10.1371/journal.pbio.0030007.

70. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, et al. Ultraconserved elements in the human genome. Science (80-). 2004;304:1321–5.

71. Pennacchio LA, Ahituv N, Moses AM, Prabhakar S, Nobrega MA, Shoukry M. In vivo enhancer analysis of human conserved non-coding sequences. Nature. 2006;444:499–502.

72. Chen J, Sun M, Hurst LD, Carmichael GG, Rowley JD. Human antisense genes have unusually short introns: Evidence for selection for rapid transcription. Trends Genet. 2005;21:203–7.

73. Ahituv N. Exploiting human-fish genome comparisons for deciphering gene regulation. Hum Mol Genet [Internet]. 2004;13:R261–6. Available from: https://academic.oup.com/hmg/article-lookup/doi/10.1093/hmg/ddh229.

74. Clevers H. Wnt/β-Catenin Signaling in Development and Disease. Cell. 2006;127:469–80.

75. Roarty K, Rosen JM. Wnt and mammary stem cells: Hormones cannot fly wingless. Curr Opin Pharmacol. 2010;10:643–9.

76. Wend P, Holland JD, Ziebold U, Birchmeier W. Wnt signaling in stem and cancer stem cells. Semin Cell Dev Biol. 2010;21:855–63.

77. Incassati A, Chandramouli A, Eelkema R, Cowin P. Key signaling nodes in mammary gland development and cancer: β-catenin. Breast Cancer Res. 2010;2:1–14.

78. Sauvageau M, Goff LA, Lodato S, Bonev B, Groff AF, Gerhardinger C, et al. Multiple knockout mouse models reveal lincRNAs are required for life and brain development. Elife [Internet]. 2013;2:e01749. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24381249.

79. Lai K-MV, Gong G, Atanasio A, Rojas J, Quispe J, Posca J, et al. Diverse Phenotypes and Specific Transcription Patterns in Twenty Mouse Lines with Ablated LincRNAs. Wu Q, editor. PLoS One [Internet]. 2015;10:e0125522. Available from: https://dx.plos.org/10.1371/journal.pone.0125522.

80. Madison BB, Jeganathan AN, Mizuno R, Winslow MM, Castells A, Cuatrecasas M, et al. Let-7 Represses Carcinogenesis and a Stem Cell Phenotype in the Intestine via Regulation of Hmga2. Gartenhaus RB, editor. PLOS Genet [Internet]. 2015;11:e1005408. Available from: https://dx.plos.org/10.1371/journal.pgen.1005408.

81. Nagy Á, Lánczky A, Menyhárt O, Győrffy B. Validation of miRNA prognostic power in hepatocellular carcinoma using expression data of independent datasets. Sci Rep [Internet]. 2018;8:9227. Available from: http://www.nature.com/articles/s41598-018-27521-y.