Supplementary Information

Functional heterogeneity of PyMT tumours is associated to lineage origin (cont).

Cluster 0 (C0) was consistent with undifferentiated cells or uncommitted progenitors (depleted Notch signalling). C0 was characterised by a high enrichment on proliferative and cell cycle associated hallmarks suggesting that these could be a cell of origin of a highly aggressive cancer. Interestingly C0 was enriched on inflammatory associated hallmarks, such as IL6/Stat3, Interferon a and g, inflammatory response, allograft rejection and ROS. This cluster is consistent with immunogenic tumours such as those of the TNBC subtype.

The luminal progenitor group was formed by clusters Cluster 2, 3 and 4. Cluster 2 was functionally similar in both genotypes with a slight cell enrichment in Elf5 tumours. C2 was characterised by a profound depletion on differentiation pathways such as Notch and TGFβ as well as proliferative and metabolic pathways, this suggests that C2 has a highly multipotent progenitor profile with high epithelial score (depletion of EMT), consistent with a transitory amplified progenitor. C3 was characterised by the expression of genes signatures associated to IFN α and γ, this suggests that through the differentiation to the specific lineages and, in particular, in the alveolar compartment, IFNα and g signalling gets progressively lost. For example, C0 presented enrichment for IFNα and g, this is conserved in the luminal progenitor cells C3 and only partially conserved in C1 (IFNα enriched and IFNγ depleted) to end up negatively both enriched in C6 and poorly enriched in C9. Interestingly, C3 as a luminal progenitor was depleted for signatures associated to hormone response (androgen, estrogen response early and late, IL2/Stat5) as well as hallmarks associated with TME communication (e.g. hypoxia, coagulation, angiogenesis) and inflammation (inflammatory response, ROS, complement) and a markedly low EMT, proliferative and differentiation pathways. This cluster was consistent with a non-committed luminal progenitor identity, with no clear commitment to either alveolar or the hormone sensing lineages. Albeit a slight enrichment in WT tumours, the functional activity of these cells is very similar between genotypes indicating a true luminal progenitor identity. Amongst the few differences that could be found between C3_WT and C3_Elf5 cells, it stood out an increase in the score for the signatures related to Notch and TGFβ, this could be consistent with a bias towards a transition to a more differentiated state in Elf5 tumours.

The functional annotation for Cluster 4 is consistent with the previous results of marker genes and trajectory analysis shown in Figs 3 and 4, placing this cluster as a luminal progenitor committed towards the hormone-sensing lineage. Cluster 4 was heavily enriched in WT tumours where Hormone Sensing differentiated cells (C9) are much more frequent. LPs are characteristically depleted in hormone signalling hallmarks, however C4 scores for hormone related hallmarks (ER_early [-0.110]; ER_late [-0.128]; Androgen [-0.017]) were comparatively higher than C2 (ER_early [-0.-0.323]; ER_late [-0.171]; Androgen [-0.164])and C3 (ER_early [-0.209]; ER_late [-0.168]; Androgen [-0.238]), consistent with their transition towards a Hs lineage but still retaining the functional characteristics of a luminal progenitor. Similarly to the other clusters, C4 Elf5 cells presented high Notch signalling enrichment score compared to C4 WT.

Consistently with previous studies, Elf5-driven cell clusters associated to the alveolar lineage, clusters 1 and 6 (C1 and C6), presented a negative enrichment on estrogen responsive genes, which is presumably associated to the opposing transcriptional force that Elf5 drives on ER signalling [37,38]. Similarly, a downregulation of genes involved in EMT were characteristics of these two clusters [38,58]. These clusters were also characterised by an enrichment of the IL6/Stat3 signalling pathway, a pathway associated with tumour progression and metastasis but also with induction of cancer immunetolerance [100], and cholesterol homeostasis, also associated with breast cancer progression, inflammation and metastasis. Interestingly aspects of innate immune related pathways (Hallmark Complement) were highly enriched in both alveolar clusters. Reduced cell cycle progression and cell proliferation has been associated to Elf5 expression [37] and tumour growth [38] this is consistent with the observed decline of Hallmarks related to cell cycle progression (G2M, E2F, mitotic spindle and Myc targets) and cell growth and proliferation (i.e Hedgehog). All these characteristics have been previously associated to Elf5 rich tumours including in the PyMT mouse model [38]. In addition, C6 was characterised by enriched genes in response to Hypoxia, coagulation and involved in angiogenesis, suggesting a strong sensitivity of these cells to signals of the extracellular ecosystem, as Elf5 tumours have been shown to exhibit strong angiogenesis, and vascular leakiness [38]. Similarly, cells from C6 seemed to be responding to oxidative stress (ROS and Xenobiotic signalling pathways), this is associated with the formation of chronic inflammation and recruitment of leukocytes, and the depletion on IFN signalling in this cluster might be associated to increase of immune tolerance and recruitment of immunosuppressive leukocytes, something that has been previously shown in this model associated with Elf5-high tumours [38].

The hormone sensing linage was represented by Cluster 9. This cluster was characterised by an enrichment of a number of hallmarks associated to hormonal response, Estrogen, (late and early), androgen and IL2/Stat5 pathways. Interestingly and similarly to what it was observed in C6, these more differentiated cells in both lineages were strongly associated to communication with the TME, including hypoxia, angiogenesis and inflammation associated hallmarks.

The basal/myoepithelial subtype was formed by clusters 7 and 8. Cluster 7 (C7) is consistent with cells from the myoepithelial lineage. These cells were heavily involved in inflammation-associated hallmarks (complement, allograft rejection, inflammatory response, etc.) and hallmarks associated with TME interactions such as hypoxia and angiogenesis. Such characteristics were also found in the undifferentiated cells from C0, and often undifferentiated multipotent cells and myoepithelial cells showed similar gene signatures. However, in direct contrast with cells from C0, C7 cells present a strong mesenchymal phenotype, characterised by EMT and a higher myogenesis score, and an enrichment in Notch signalling, indicating restriction of stem cell activity and commitment towards a more differentiated fate.

Cluster 8 was consistent with a basal/ myoepithelial lineage and cells of this lineage were more frequent in WT tumours, presumably by the luminal skew produced by the Elf5 differentiation force. Cells that belong to this cluster presented enrichment for Notch and TGFb pathways, indicating a differentiated stage, however these cells are rich in EMT markers, typical of a myoepithelial lineage. Amongst the differences by genotype, a negative

enrichment of innate immune inflammatory properties (complement, IFNg, and oxidative stress signals) as well as reduced proliferative hallmarks (mitotic spindle, G2M checkpoints and E2F targets) in Elf5 cells compared with WT cells were the major highlights (Supplementary Fig. 7, Heatmap).

Extended Methods

Batch effect analysis

Elf5 has two distinct effects in the mammary gland, it drives the formation of new cell types via differentiation and it alters the activity state of existing cell types. We thus performed Canonical Correlation Analysis (CCA)[41] to analyse the effect of the genotypes in the distribution of cancer epithelial cells shown in Figure 3 and to study whether this effect could correspond to an intrinsic property of the different genotypes or to the biological effects exerted by ELF5 in the mammary epithelial cells.

CCA analysis of the epithelial cell compartment resulted in a clear alignment of the two genotypes (SuppFig. 5A), allowing the mapping of common cell types between WT and ELF5 expressing samples and so defining new cell types produced by increased ELF5 expression. A tSNE visualisation of the K-means clustering produced after CCA alignement is shown in SuppFig. 5B, defining 10 clusters. CCA-aligned clusters 2, 6, 7 and 9 were similar for both genotypes while all other clusters were clearly polarised upon ELF5 expression, similarly to the version before CCA alignement (Fig. 3). This result indicates that the polarisation is biological and not due to batch effects. Contour plots on SuppFig. 5C shows the distribution of the CCA-aligned cell clusters that were not influenced by Elf5 expression. We then matched the clusters defined by variable genes before CCA alignment and visualised them in the context of CCA distribution (SuppFig. 5D), showing that the clusters that were manly present only in ELF5 tumours still sit in one of the extremes of the transcriptional space opposite to the mainly WT populated clusters. The distribution of CCA aligned cells clusters in the mammary epithelial hierarchy also confirmed the results shown in SuppFig 5, but also defined the clusters that are not different between the two genotypes: CCA-aligned cluster 2 greatly overlapped with cluster 5, populated by cycling cells; CCA-6 corresponded with cluster 8, a basal/myoepithelial cluster; CCA-7 correlated with cluster 9, a hormone sensing cluster, but also contained some other cells polarised by genotype, suggesting that still portion of the hormone sensing cells are influenced by ELF5 expression in the mammary epithelium; finally, CCA-9 cells was populated by a small number of cells without a clear cluster pattern likely identified as normal cells as described in Figure 3. Interestingly, the other cell clusters that presented a polarised distribution by genotype correspond to those on a clear axis through the multipotent cells along the luminal progenitor lineage. This is consistent with the reported effect of ELF5 in the mammary epithelium.

Taken together these results show that the polarised distribution by genotype is not an effect of the analysis but responds to the biological effects exerted by ELF5 in the mammary epithelial cells.

Supplementary Figure legends.

Supplementary Figure 1. A) Relative amount of mitochondrial genes/ nuclear genes (percent.mito) in each tumour sample. **B)** Distribution of the cells with high mitochondrial gene content. **C)** Clustering analysis highlighting the cells with high mitochondrial gene content. Right hand plots show the same location of high mitochondrial content genes in each genotype. The colour gradient shows the cells with highest mitochondrial genes (dark red) to the lowest (light red). **D)** Violin plots of the relative amount of mitochondrial genes in each cell cluster define by defined by k-means clustering analysis, cluster 0 defines the high mitochondrial gene content cluster. **E)** Principal component analysis of the identified clusters. **F)** Distribution of the cells with high mitochondrial gene content after applying a 15% maximum mitochondrial content threshold.

Supplementary Figure 2. A) Distribution of UMI and number of genes (nGene) per sample and genotype. **B)** Correlation of UMI and Gene identities. **C)** Elbow plot showing the threshold used on principal components (PC) for subsequent clustering analysis. **D)** Heatmaps of the main gene drivers for each principal component. **E)** Principal component analysis by genotype (left panels) and by sample (right panels).

Supplementary Figure 3. A) tSNE plot showing the distribution of cells per replicate. **B)** tSNE plot showing the distribution of cells per genotype.

Supplementary Figure 4. A) Cluster tree of the tumour cell diversity. At each fork, embedded pie charts show the proportion of cells that belong to each genotype (Elf5 white and WT light blue) and a list of the top differential genes. **B)** Cluster tree of the cancer epithelial cell diversity. At each fork, embedded pie charts show the proportion of cells that belong to each genotype (Elf5 white and WT light blue). Table shows the top ten genes in each cluster at the defined resolution (0.7).

Supplementary Figure 5. A) Comparison of the principal component analysis before (left panel) and after (right panel) CCA alignment. **B)** tSNE plot of the K-means clustering of the CCA aligned data by CCA-cluster identity, and grouped by genotype (mid and right panels). **C)** Contour plot of the k-means clustering highlighting the distribution of the non-polarised clusters. **D)** Overlay of the colour palette of the cluster identities from figure 3 in the tSNE coordinates after CCA alignment.

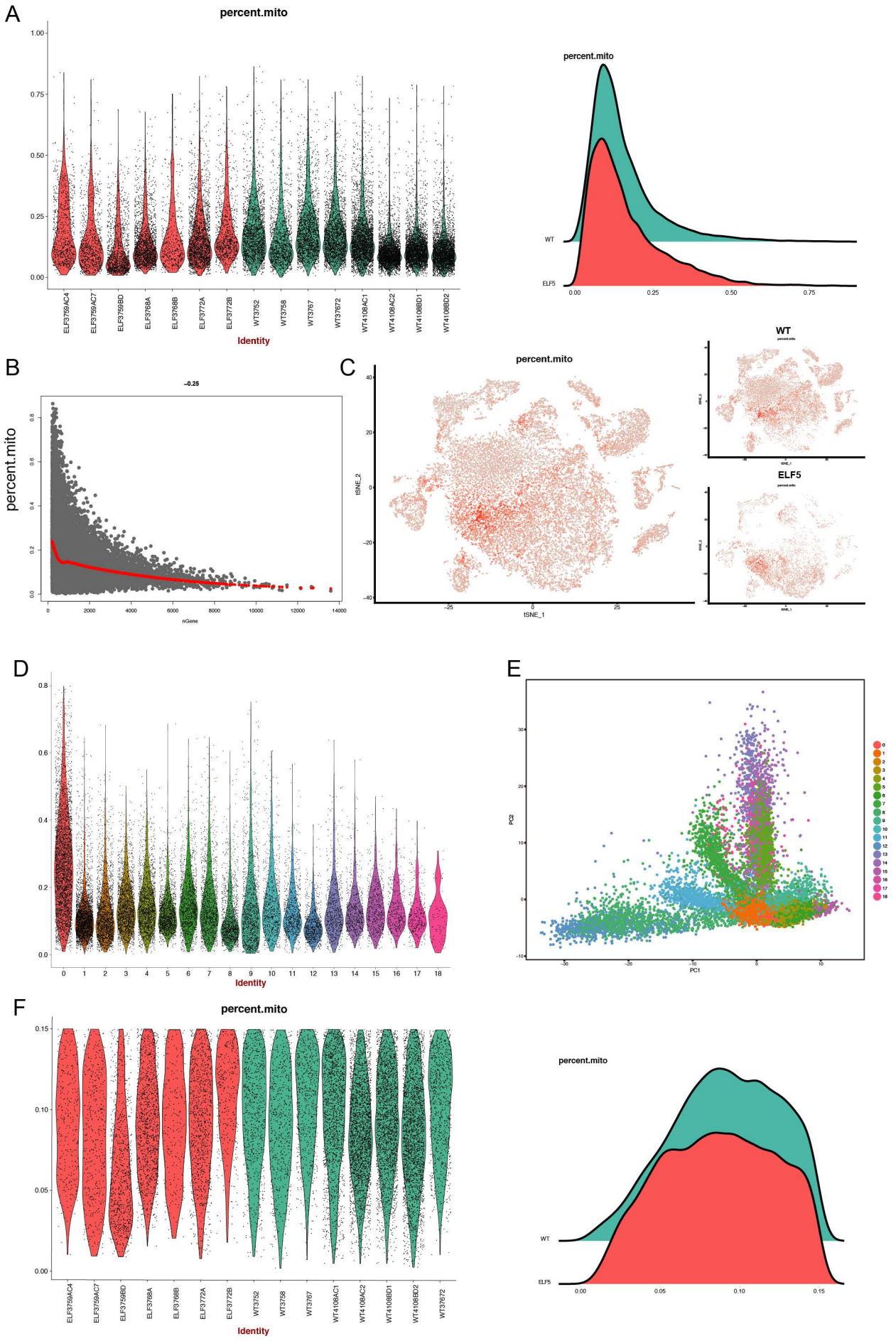
Supplementary Figure 6. A) Ternary plots using the gene signatures of the mammary epithelial hierarchy as in Pal et al. 2017. The colour palette correspond with the defined cluster identity in Figure 3A. **B)** Overlay of the monocle coordinates with the cluster identities for each genotype. **C)** Overlay of the lineage identities (colour palette) (left panel) and the cell distribution of each lineage (right panels) in the tSNE structure of the epithelial cancer cells after CCA alignment. Numbers correspond with the cluster identities. **D)** Overlay of the lineage identities (colour palette) (left panel) and the cell distribution of each lineage

(right panels) in the tSNE structure of the epithelial cancer cells. Numbers correspond with the cluster identities.

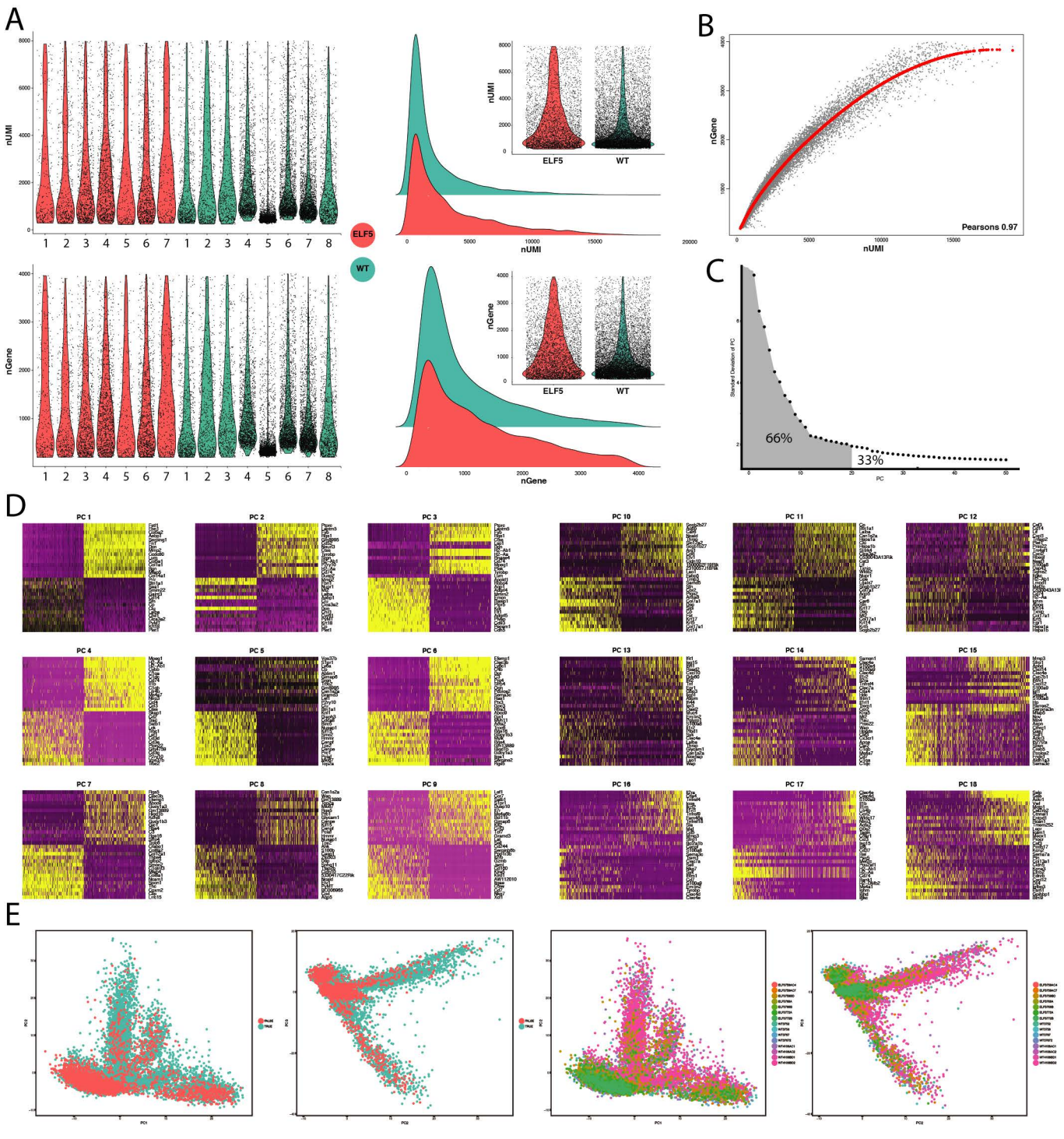
Supplementary Figure 7. GSVA analysis for the H Hallmarks from MSigDB in the cancer epithelial clusters. Waterfall plots are presented for clusters mainly formed by one genotype and a heatmap is presented in the clusters formed by the two genotypes (according to Figure 3B). The variation score shows the hallmarks enriched (light blue, positive values) and not enriched (dark blue, negative values) for each cluster.

Supplementary Figure 8. A) Cluster tree modelling the phylogenic relationship of the different clusters within the cancer-associated fibroblasts at different clustering resolutions. **B)** Gene-expression heatmap of the top expressed genes for each cancer-associated fibroblast cluster at resolution 0.1. **C)** Upper panel: tSNE representation of the cell cycle stages of the cancer-associated fibroblasts as defined by gene expression signatures using tSNE coordinates. Circled area shows the cycling cluster characterised by a total absence of G1 cells. Bottom panel: tSNE plot of the cancer-associated fibroblasts defined by k-means clustering analysis at resolution 0.4. The circle highlights the same area for cycling cells, which it is mostly populated by a single cell cluster, cluster #4. **D)** Violin plots displaying the classical marker genes that define CAFs for each of the three fibroblast clusters defined in resolution 0.1. **E)** tSNE representation of the CCA analysis of the CAF cell compartment revealing a similar polarisation between the two genotypes, MMT/PyMT (WT, blue) MMT/PyMT Elf5 (Elf5, red).

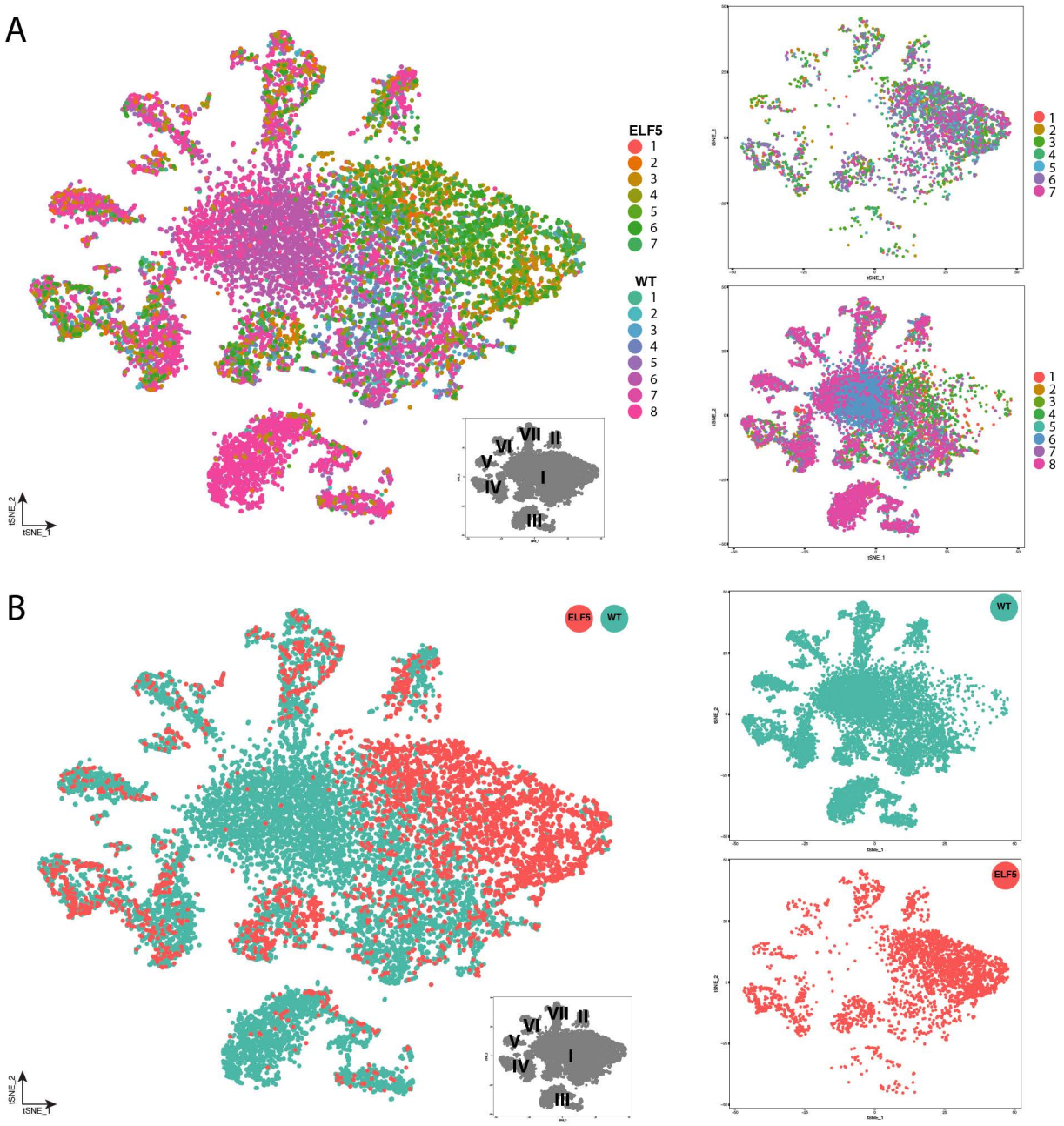
Supplementary Figure 9. A) Waterfall plot of the GSVA analysis for the H Hallmarks from MSigDB in the CAF cluster defined by resolution 0.1; Cluster 0 and 1 were defined as "CAFs" and Cluster 2 as "Myofibroblasts". The variation score shows the hallmarks enriched (light blue, positive values) and not enriched (dark blue, negative values) for each cluster. **B)** Waterfall plot of the GSVA analysis for the H Hallmarks from MSigDB in the involution CAFs clusters defined by resolution 1 (clusters #2 and 3). The variation score shows the hallmarks enriched (light blue, positive values) and not enriched (dark blue, negative values) for each cluster. **C)** tSNE plot showing the assignment of each cell type for the analysis of cell to cell interactions using CellphoneDB. The circles correspond to the cells from the Elf5 genotype and the triangles are the cells coming from the WT genotype.



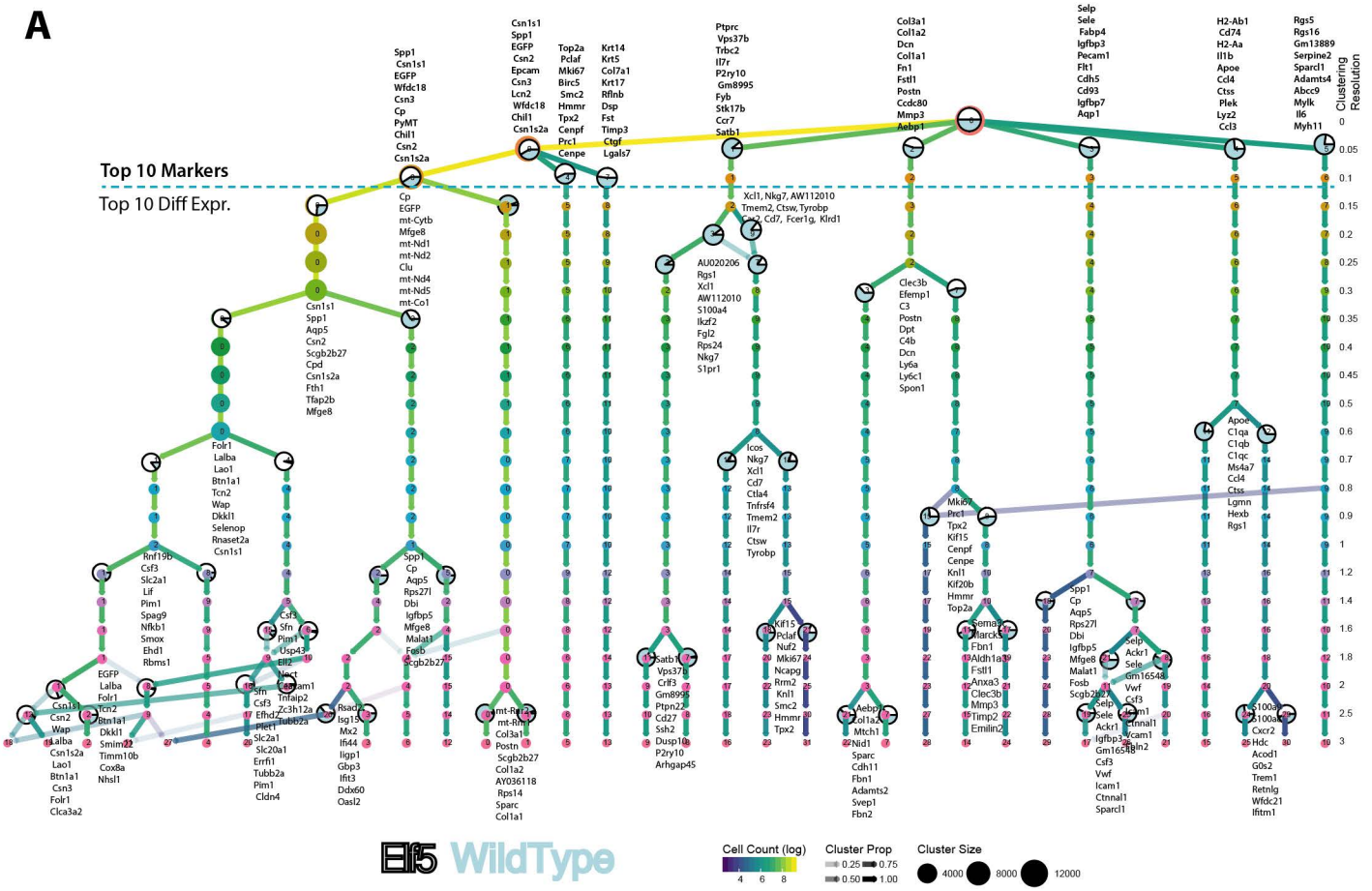
Supplementary Figure 1. Valdes-Mora et al. 2019

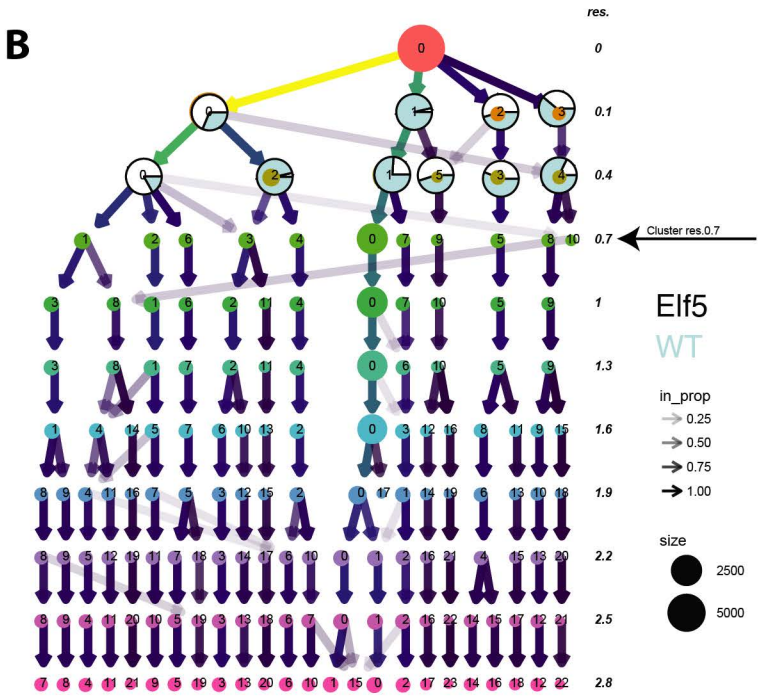


Supplementary Figure 2. Valdes-Mora et al. 2019



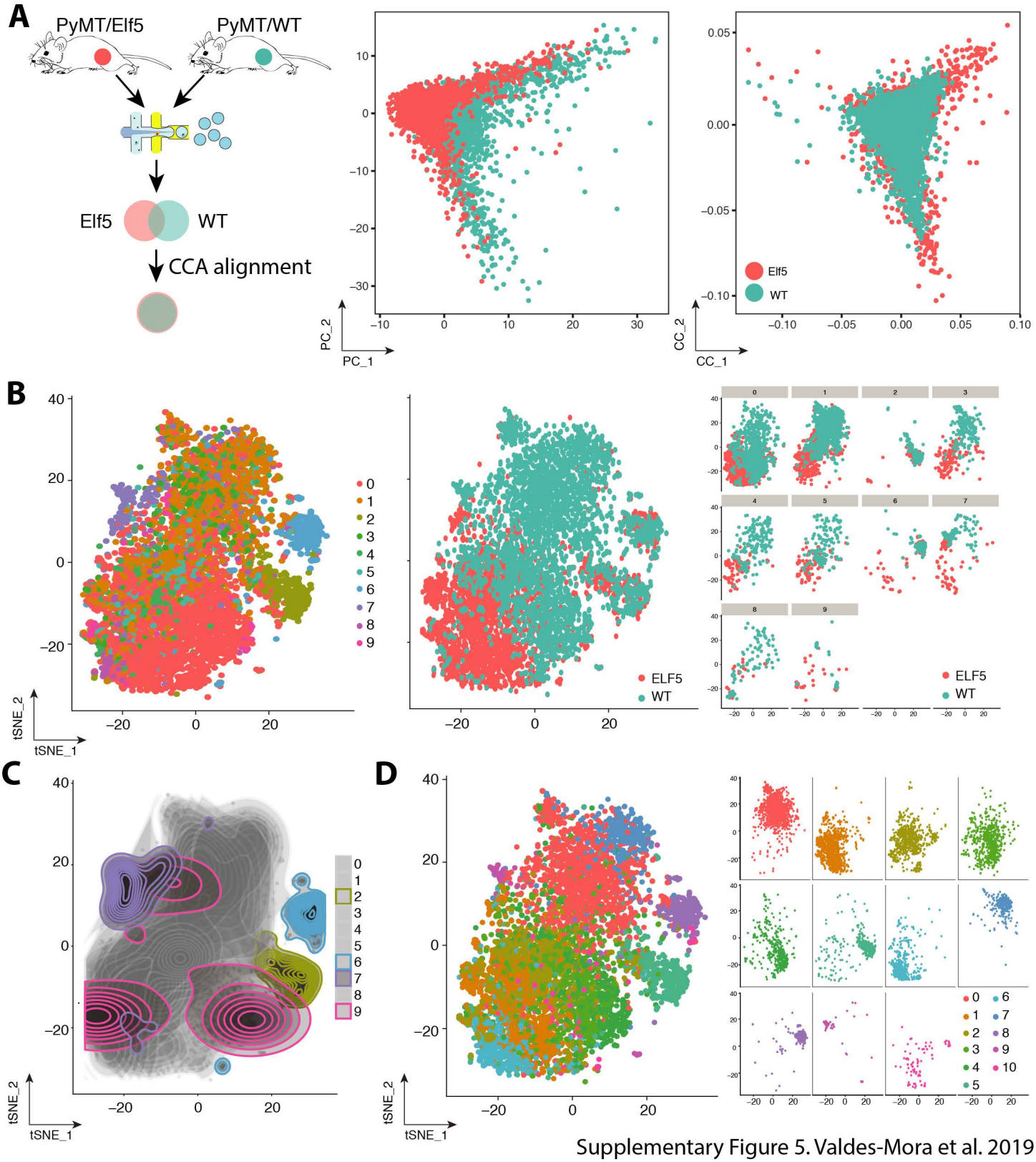
Supplementary Figure 3. Valdes-Mora et al. 2019



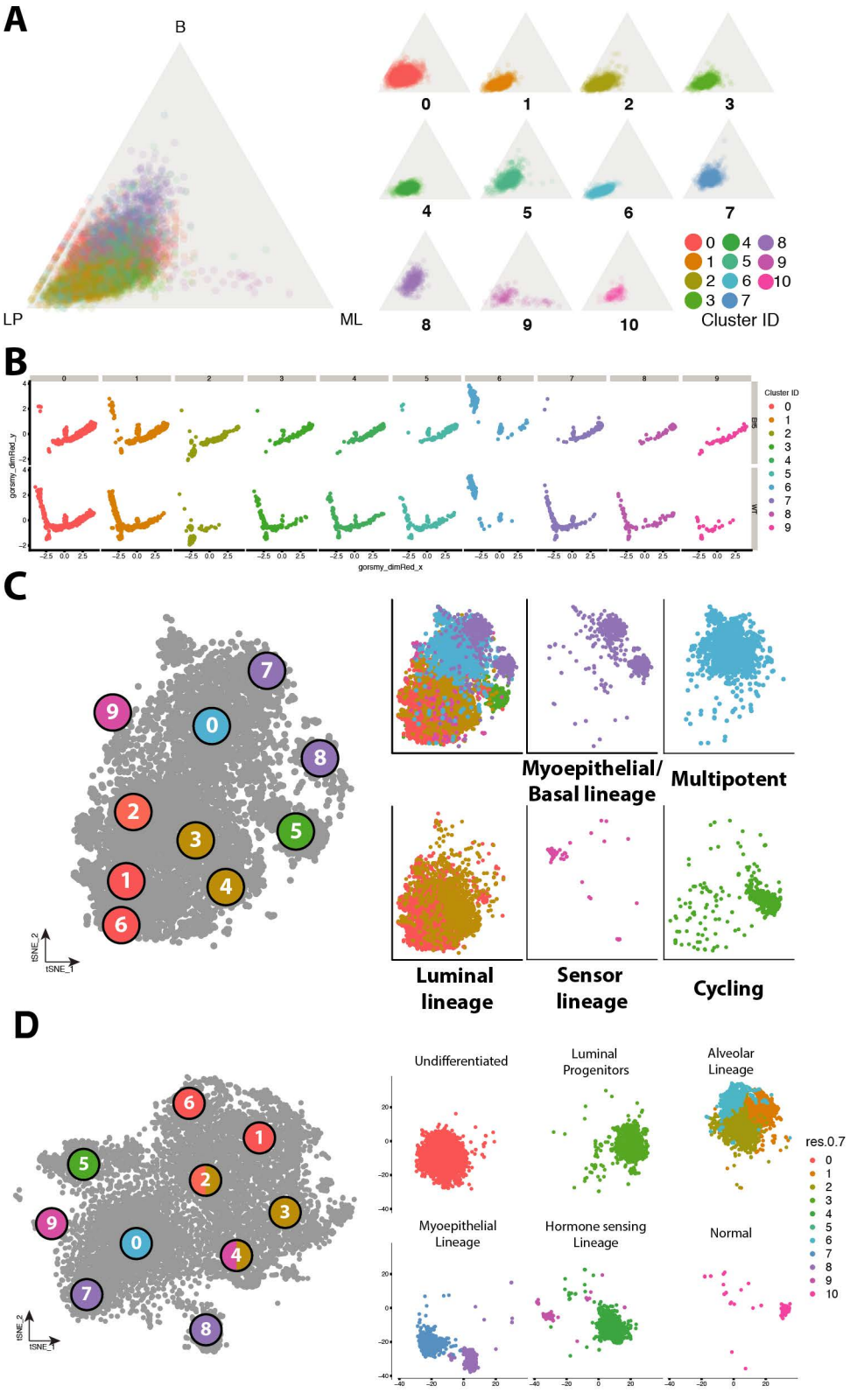


top n by logFC	Cluster res.0.7										
	0	1	2	3	4	5	6	7	8	9	10
1	Gm42418	Cp	Plet1	Hspa1b	Aqp5	Top2a	Wap	Col1a2	Krt14	Gm10801	Aldh1a3
2	Scgb2b27	Spp1	Epcam	Hspa1a	Cpd	Mki67	Lao1	Col3a1	Krt5	Gm26870	Krt7
3	Col1a2	Mfge8	Csf3	Fos	Pgp	Cenpf	Csn1s1	Gm42418	Col7a1	Gm10800	Actn1
4	Scgb1b27	Csn1s1	Efhd2	Fosb	Socs2	Smc2	Csn1s2a	Sparc	Krt17	Gm10717	Tnc
5	Sparc	Srgn	Cd14	Jun	Car6	Prc1	Csn2	Col1a1	Fst	Col3a1	S100a6
6	Col1a1	Cst3	Ceacam1	Egr1	Slc12a2	Hmgb2	Btn1a1	Bgn	Ctgf	Col1a1	Msn
7	Col3a1	Slpi	Tm4sf1	Ier2	PyMT	Pclaf	Glycam1	Fstl1	Rflnb	Neat1	Cd44
8	Mgp	Chil1	Neat1	Atf3	Scgb2b27	Birc5	Lalba	Col5a2	Postn	Krt19	Rgs16
9	AY036118	Lgals3	Map2k3	Cyr61	Srebf1	Cenpe	Ly6a	Rgs5	Timp3	Xist	Crip1
10	Cald1	Ptgs2	Slc2a1	Ltf	Aldoa	Smc4	Ltf	Col4a1	Apoe	Nek7	Ptgs2

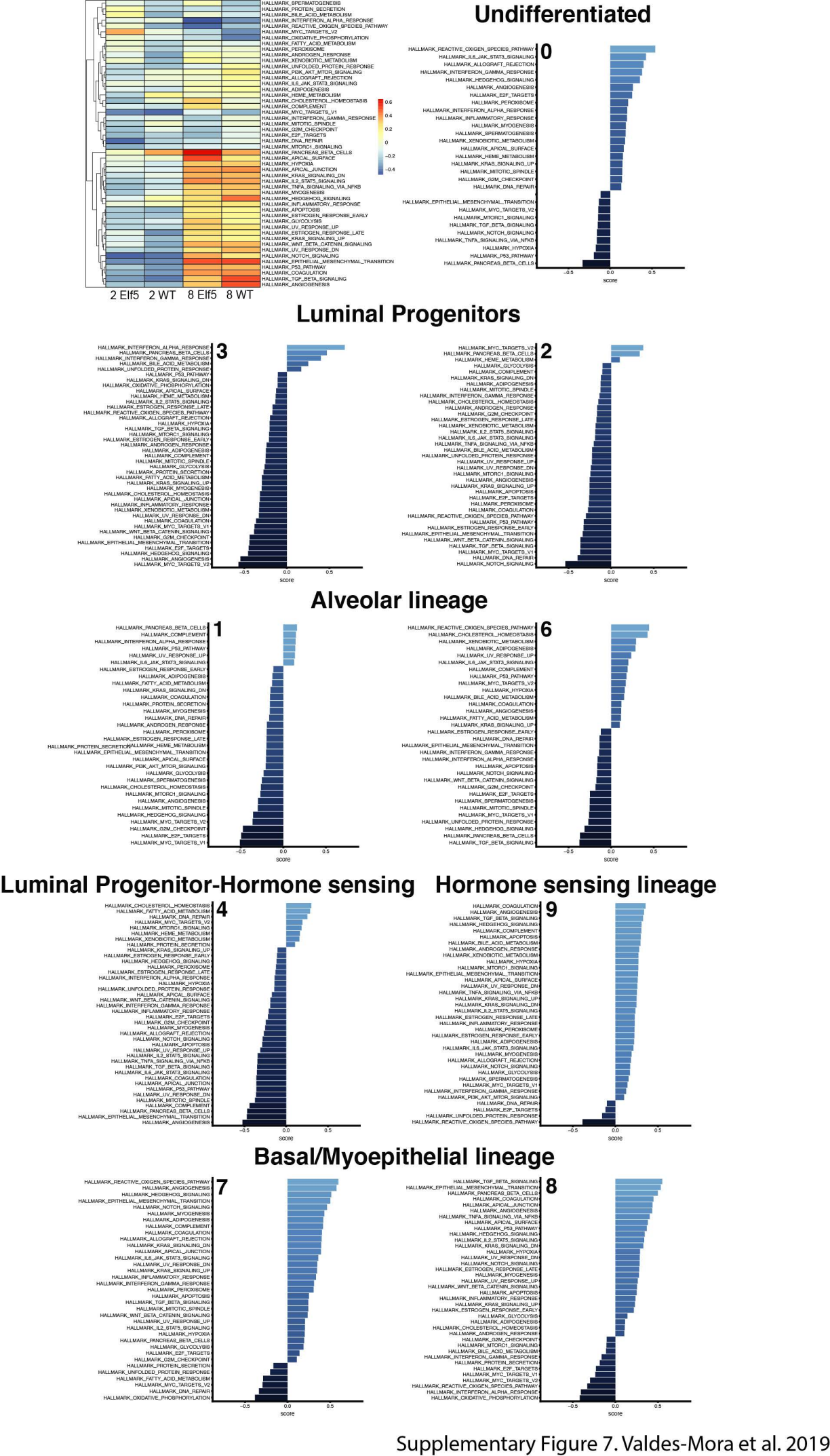
Supplementary Figure 4. Valdes-Mora et al. 2019



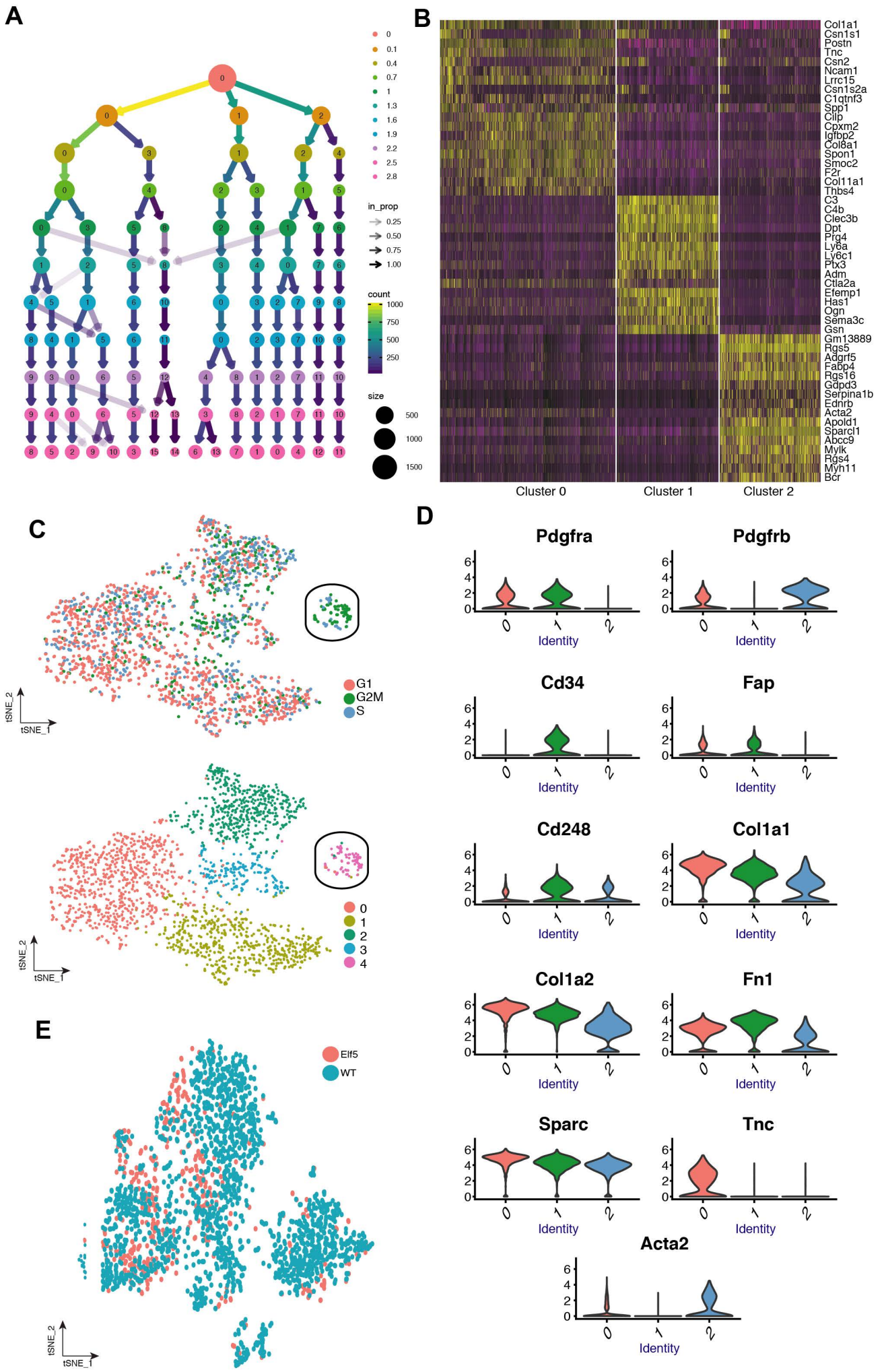
Supplementary Figure 5. Valdes-Mora et al. 2019



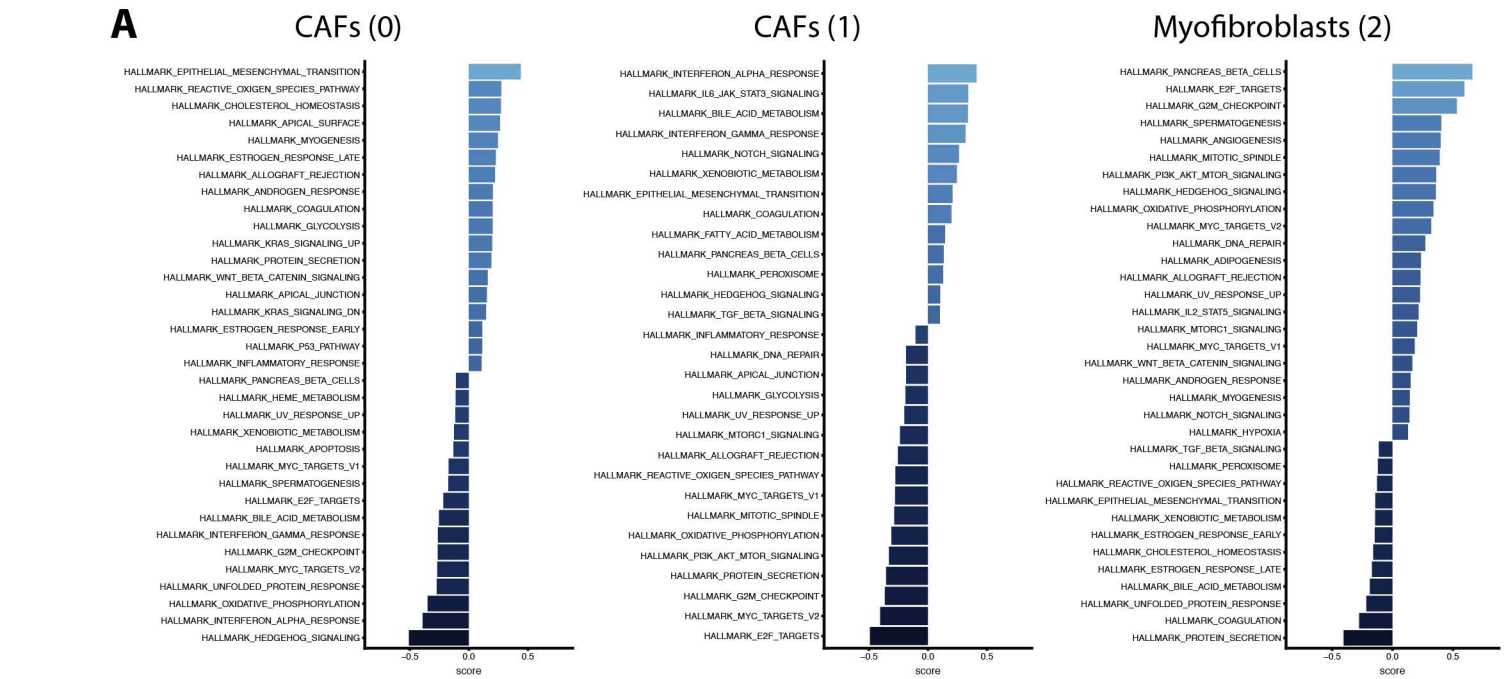
Supplementary Figure 6. Valdes-Mora et al. 2019

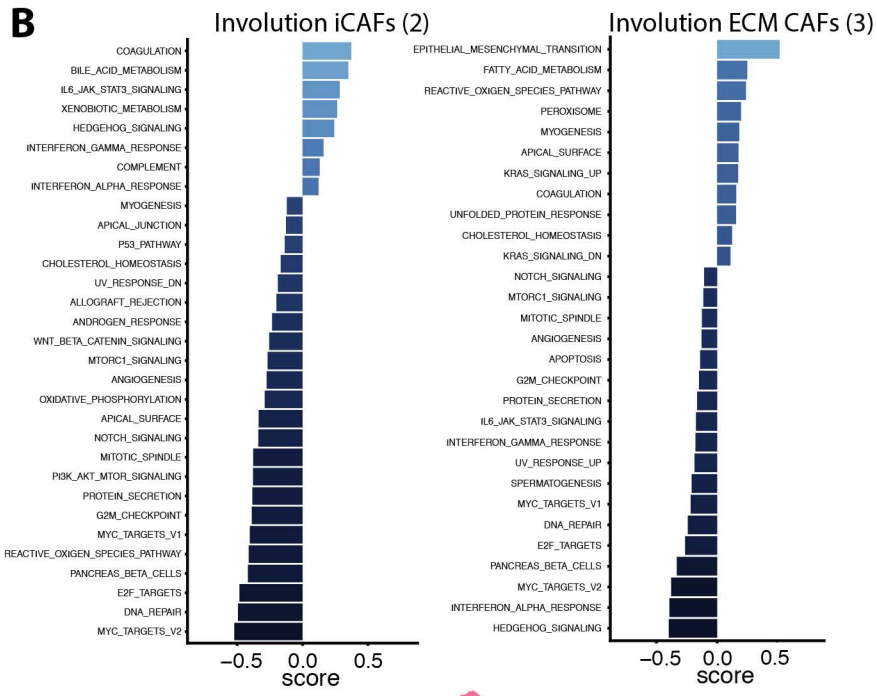


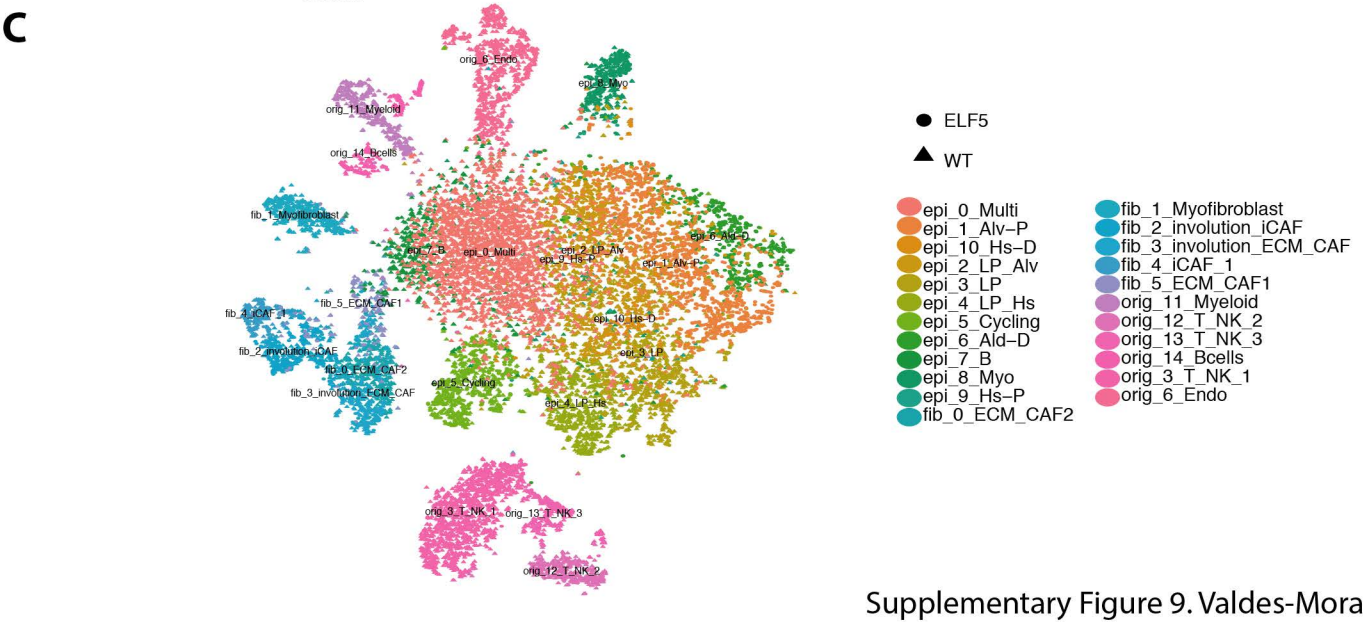
Supplementary Figure 7. Valdes-Mora et al. 2019



Supplementary Figure 8. Valdes-Mora et al. 2019







Supplementary Figure 9. Valdes-Mora et al. 2019