A

Basal Markers

Krt5

Krt14

Tp63

C1D-bi C1D-B C1D-L

Luminal Markers

Krt8

Krt18

CD14

Epcam

C1D-bi C1D-B C1D-L

B

Ly6a

C1D-bi C1D-B C1D-L

Ly6e

C1D-bi C1D-B C1D-L

C

pseudotime

C1D-bi-stem   C1D-L   C1D-B

Fig. S1. Single-cell analysis of Comma-1D cells. A) Expression level of luminal and basal markers in the Comma-1D populations. Notice that this figure is complementary to Fig. 1C. B) Expression level of SCA1 genes (*Ly6a/e*) in the Comma-1D populations. C) The violin graph shows the distribution of pseudotime values in Comma-1D populations. Notice that this graph complements Fig. 1D.
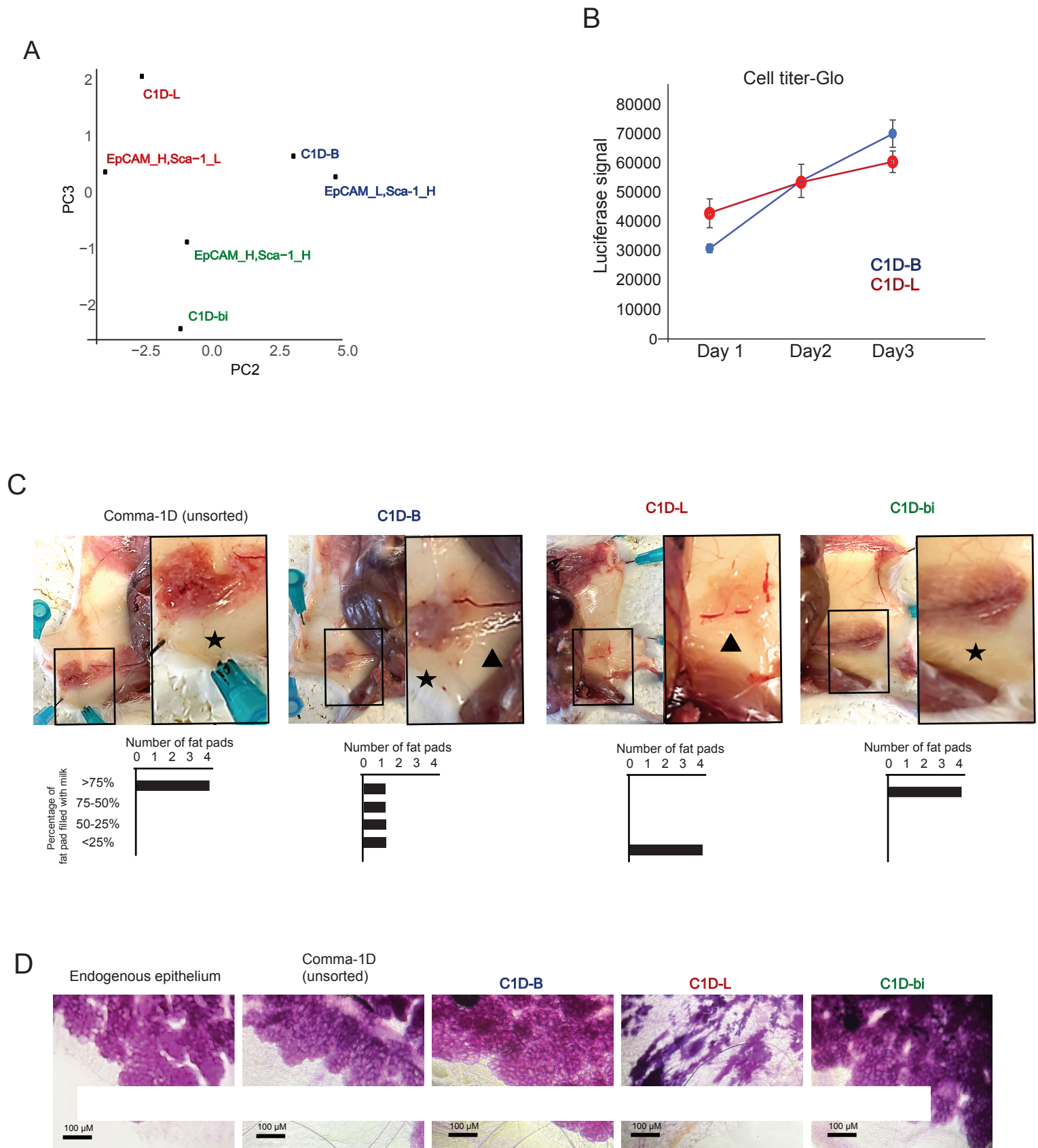
Fig. S2. Characteristics of purified Comma-1D populations. A) Principal component analysis comparing the expression profiles of FACS purified (RNA-seq) and sc-seq Comma-1D populations. B) Growth of C1D-L and C1D-B in 2D culture. Notice that C1D-bi cannot be evaluated because they quickly differentiate. C) The pictures show the generation of milk-producing epithelium in fat pads transplanted with C-ID subpopulations during pregnancy (4th mammary fat pads). Notice the clearly distinct white color of the mammary glands transplanted with Comma-1 (unsorted), C1D-B and C1D-bi indicating the high content in milk. The endogenous (3rd mammary fat pads) are also included in the picture and can be used as internal control for comparison. The bar graph indicates the percentage of fat pad filled with milk (n=4). The start indicates the milk-filled portion of the mammary gland while the triangle shows areas lacking milk generating epithelium. Notice the 3rd mammary gland in all pictures showing endogenous milk-producing epithelium. D) Carmine red stainings of the mammary glands shown in C. For C1D-B a portion of the mammary gland filled with epithelium is shown.
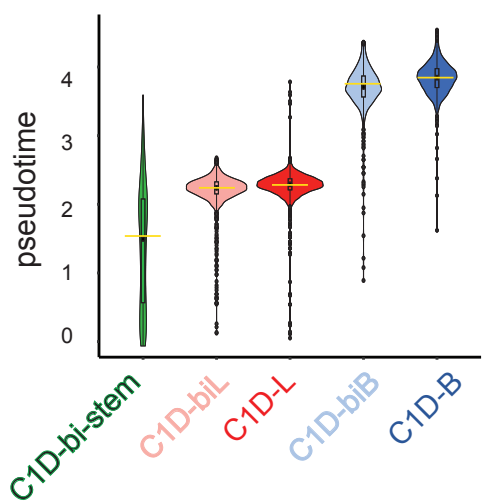
Fig S3. Pseudotime analysis of FACS purified Comma-1D subpopulations. The violin graph shows the distribution of pseudotime values in FACS purified Comma-1D subpopulations. Notice that this graph complements Fig. 3B.
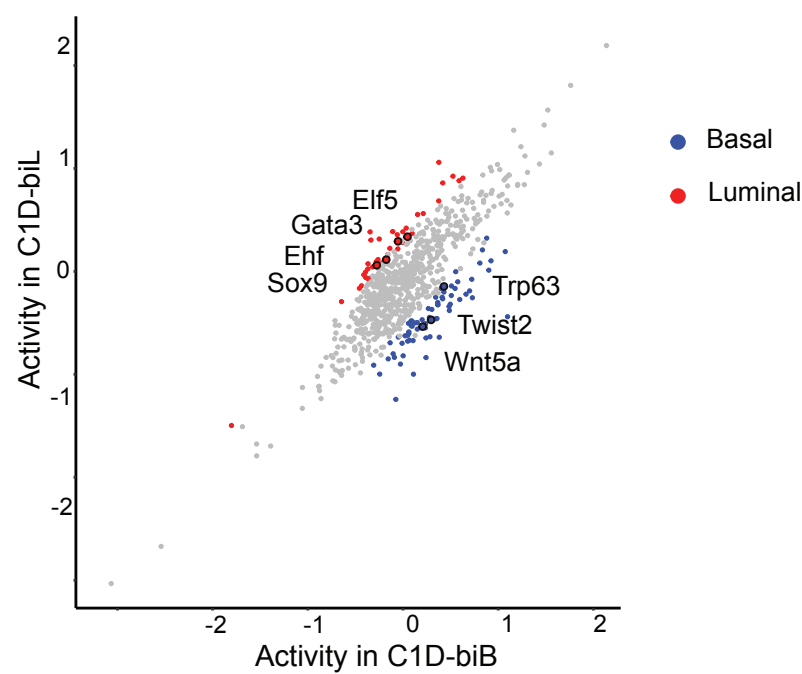
Fig S4. Transcription factor activity in Comma-1D populations. The dot-plot shows the activity of key transcription factors during basal and luminal differentiation in Comma-1D cells.
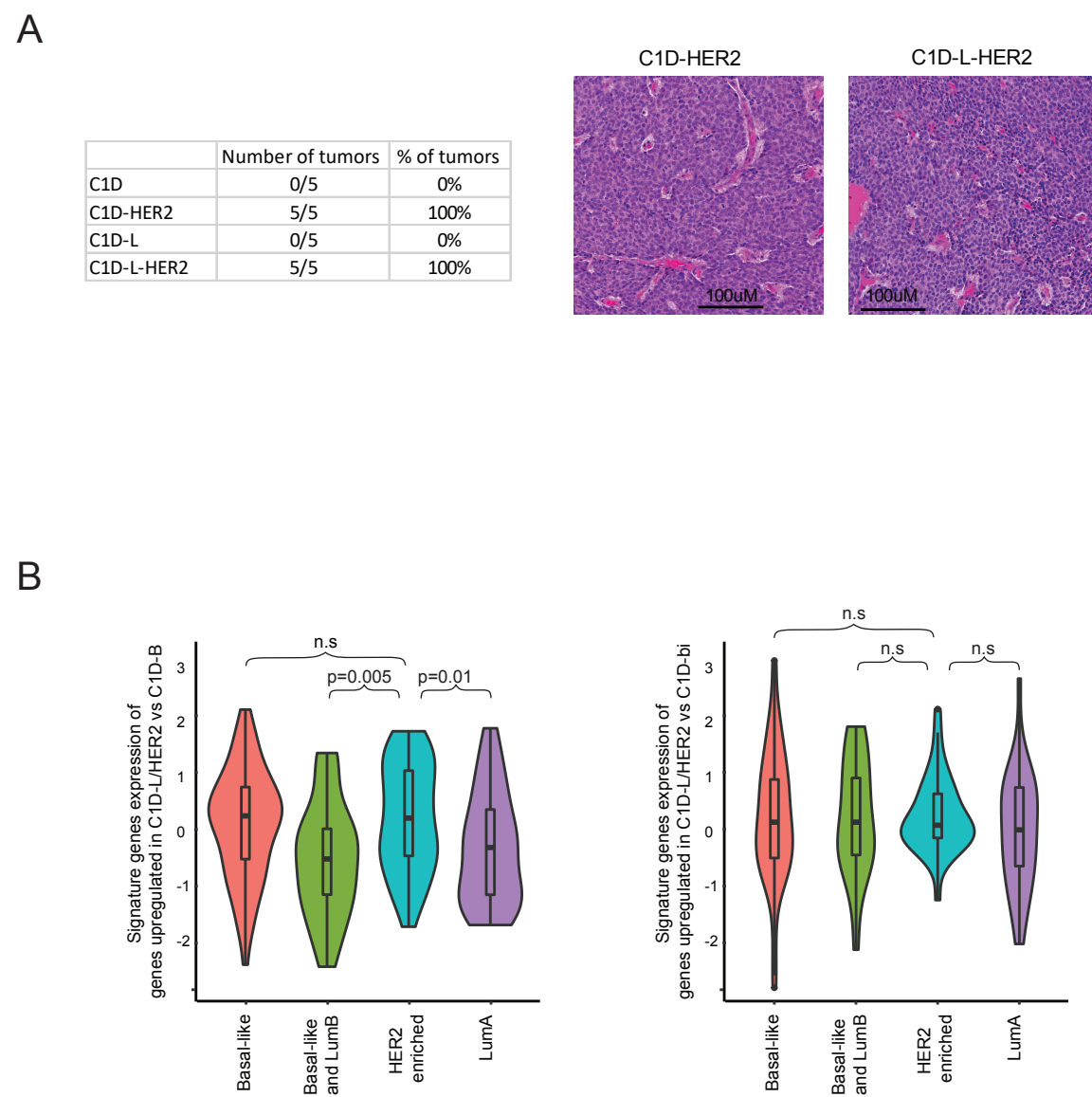
A

	Number of tumors	% of tumors
C1D	0/5	0%
C1D-HER2	5/5	100%
C1D-L	0/5	0%
C1D-L-HER2	5/5	100%

B

Fig S5. Effect of oncogenic HER2 overexpression in the transcriptome of Comma-1D populations. A) The pictures show H&E staining of tumors emerging after orthotopic transplantation of $2.5*10^5$ unsorted comma-1D and sorted C1D-L overexpressing oncogenic HER2 cells. The table indicates the number of mice and tumor outgrowth compared with control counterparts. B) Upregulation of genes associated with different molecular subtypes of mammary tumors from mouse transgenic models in C1D-B and C1D-bi overexpressing oncogenic HER2.
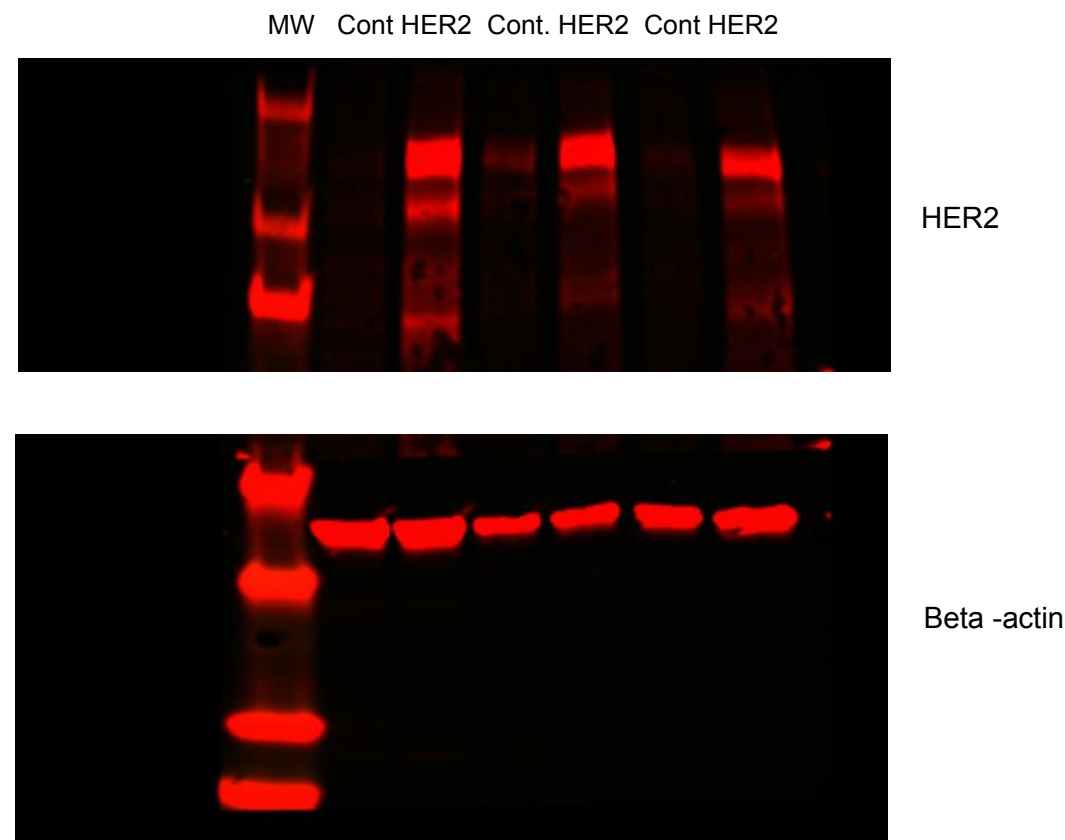
Figure S.6 Unprocessed blots

Fig. 5A

MW   Cont HER2 Cont. HER2 Cont HER2

HER2

Beta -actin

Fig**.** S6**.** Unprocessed WT-bl ot. Unprocessed blot corresponding to figure 5

Table S1**.** Selected lineage specific markers. Gene   list with the specific basal and luminal epithelial genes used to calculate the luminal and basal scores.

Click here to download Table S1

Table S2**.** Transcription Factor (TF) Activity. The table indicates the differential activity of TFs in the different comma-1D populations.

Click here to download Table S2