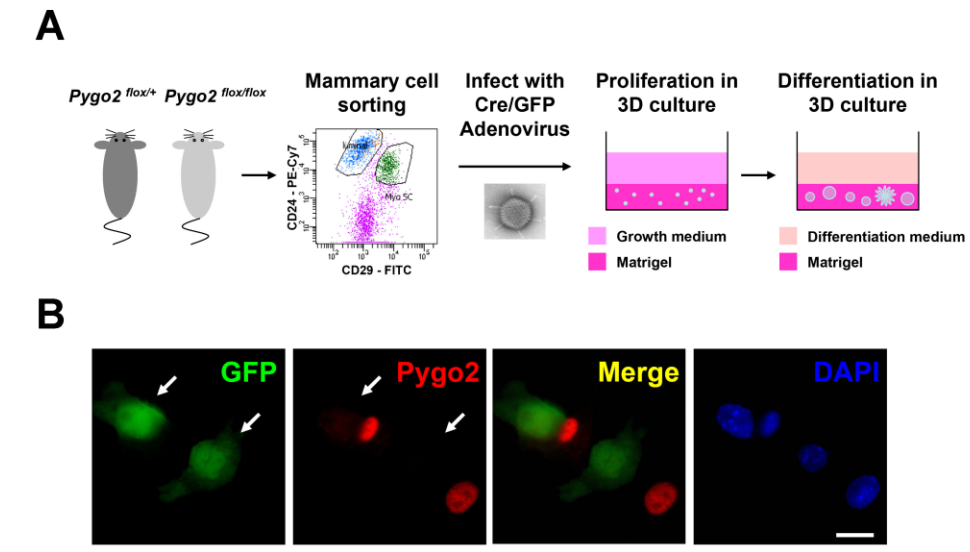
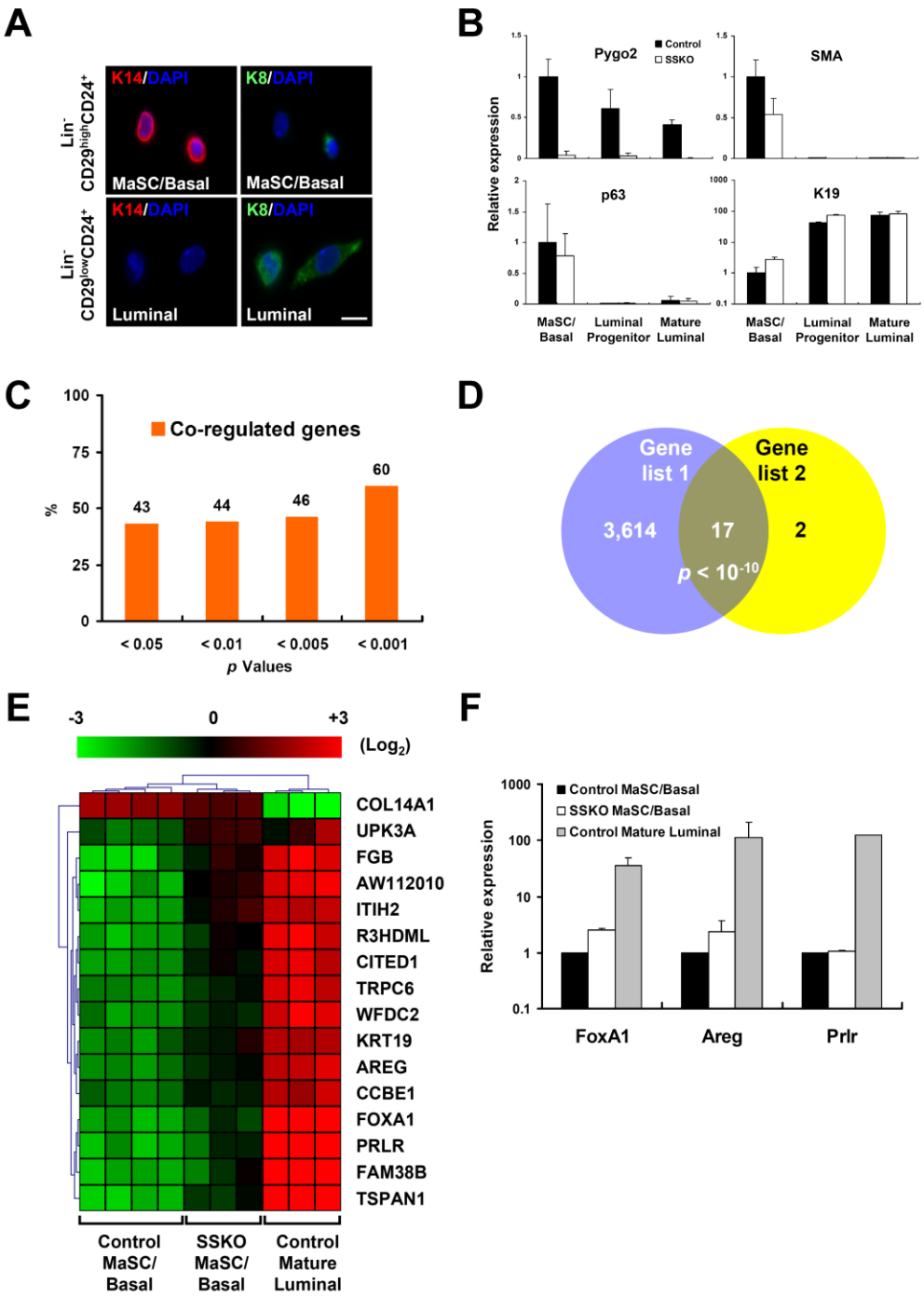
Gu et al. Figure S1.



Gu et al. Figure S2.



Gu et al. Figure S3.

A

Control MaSC/basal vs. SSKO MaSC/basal

	Fold	*p* Value
Notch3	-1.5	0.012
Dll4	-1.7	0.059
Hes1	-1.4	0.032

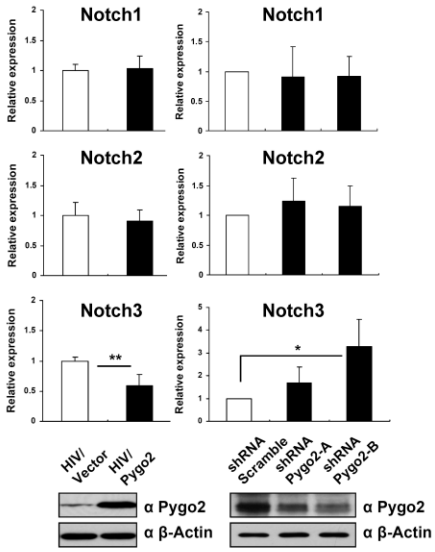
Control MaSC/basal vs. Control Luminal

Ligands	Fold	*p* Value
Jag1	3.7	$<10^{-10}$
Jag2	11.1	$<10^{-12}$
Dll1	30.8	$<10^{-13}$

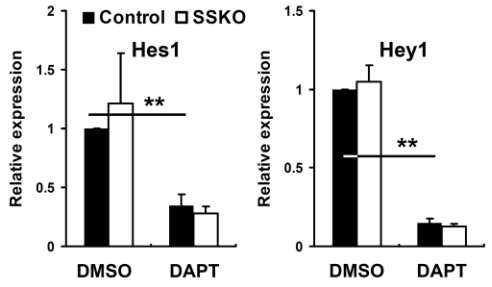
Receptors	Fold	*p* Value
Notch2	1.5	0.005
Notch3	-5.9	$<10^{-10}$

Targets	Fold	*p* Value
Hes1	-3.5	$<10^{-7}$
Hes6	-1.6	0.006
Hey1	2.0	$<10^{-3}$
Hey2	1.8	$<10^{-3}$

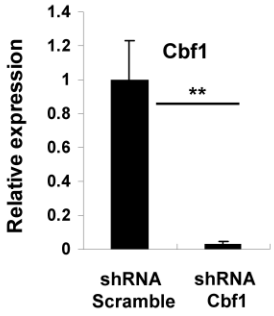
B



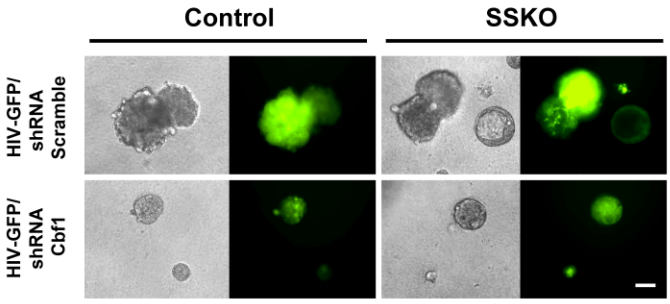
C



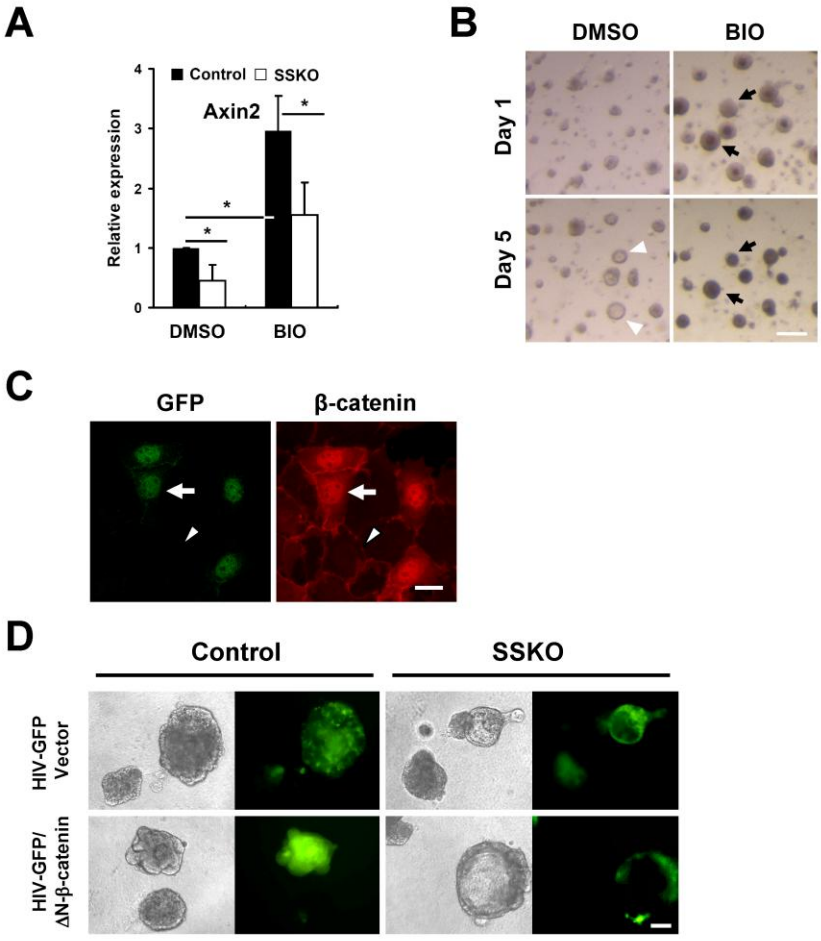
D



E



Gu et al. Figure S4.



SUPPLEMENTAL FIGURE LEGENDS

Figure S1, related to Figure 2. Ade-Cre directs acute deletion of *Pygo2* in sorted MaSC/basal cells. (A) Schematic representation of the experimental procedure. (B) Immunofluorescence analysis revealing lack of Pygo2 protein in GFP-positive, Cre-expressing cells (arrows). Bar = 10 μm.

Figure S2, related to Figure 3. Gene expression analysis of sorted mammary cell populations. (A) Immunofluorescence analysis of keratin markers in $Lin^-CD29^{high}CD24^+$ MaSC/basal and $Lin^-CD29^{low}CD24^+$ luminal cells, verifying the identity and purity of the sorted populations. Bar = 10 μm. (B) RT-qPCR analysis of MaSC/basal, luminal progenitor, and mature luminal cell populations from control and SSKO mice (n=2 each). (C) Percent of co-regulated genes in list 2 (Figure 3A) using different *p*-value cut-offs. (D-E) Venn diagram (D) and heat map (E) showing overlap between the two indicated comparisons using a cut-off of $p < 5x10^{-05}$ [FDR < $1.5x10^{-4}$ (list 1), and < 0.05 (list 2)]. (F) RT-qPCR confirms the elevated expression of *FoxA1* and *Areg* but not *Prlr* in *Pygo2* SSKO MaSC/basal comparing to control MaSC/basal cells (n=2 pairs of mice). Values in (B) and (F) are means ± standard deviations (SD).

Figure S3, related to Figure 4. Expression analysis and experimental manipulation of the Notch signaling pathway. (A) mRNA expression of Notch signaling components in sorted mammary epithelial populations by cDNA microarray analysis. For all, n>3. (B) RT-qPCR analysis (top) of *Notch1-3* expression in MCF10A cells following Pygo2 overexpression (left, n=4) or knockdown (right, n=3), which was confirmed by Western blot analysis (bottom). (C) RT-qPCR analysis of *Hes1* and *Hey1* expression in Matrigel colonies produced by MaSC/basal

cells from control and *Pygo2* SSKO mice (n=3 each), cultured in the presence of DMSO or DAPT. Note that DAPT was effective at inhibiting Notch target gene expression. (D) RT-qPCR analysis of *Cbf-1* expression, showing that the Cbf-1 shRNA lentiviruses were effective at efficiently depleting Cbf-1. HC11 mammary epithelial cells were infected with pLKO-GFP/shScramble or pLKO-GFP/shCbf-1, and RNA harvested 4 days after infection for analysis. β-actin was used for normalization. (E) Morphology of, and GFP expression in, colonies produced by control (left) or *Pygo2* SSKO (right) MaSC/basal cells infected with control scramble shRNA (top) or Cbf-1 shRNA (bottom) lentiviruses. GFP marks infected cells. Bar = 50 μm. Values are means ± SD; * $p < 0.05$; ** $p < 0.01$.

Figure S4, related to Figure 5. Expression analysis and experimental manipulation of the Wnt/β-catenin signaling pathway. (A) RT-qPCR analysis of *Axin2* expression in Matrigel colonies produced by control or *Pygo2* SSKO (n=3 each) MaSC/basal cells in the presence of DMSO or BIO. Values are means ± SD; * $p < 0.05$; ** $p < 0.01$. (B) Morphology of the Matrigel colonies derived from control MaSC/basal cells in the presence of DMSO or BIO (0.5 μM) at 1 and 5 days after HIP induction. Arrow and arrowhead indicate solid colony; and acinus, respectively. (C) Immunofluorescence analysis revealing nuclear β-catenin (red) in GFP-positive (arrow) HC11 cells infected with GFP/ΔN-β-catenin-expressing lentiviruses. Arrowhead indicates an uninfected surrounding cell that lacks nuclear β-catenin. (D) Morphology of, and GFP expression in, Matrigel colonies produced by control (left) or *Pygo2* SSKO (right) MaSC/basal cells infected with lentiviruses expressing GFP (top) or GFP/ΔN-β-catenin (bottom). GFP marks infected cells. Bar = 300 μm in (B), 10 μm in (C), and 50 μm in (D).

SUPPLEMENTAL TABLE LEGENDS

Table S1, related to Figure 3. Gene sets used for GSEA analysis.

Table S2, related to Figures 2, 4, 5, and 6. Primers used in the study.

EXTENDED EXPERIMENTAL PROCEDURES

Mouse strains

Pygo2 SSKO female mice were generated in congenic C57BL/6 background by crossing *K14-cre*; *Pygo2*$^{+/-}$ males with *Pygo2*$^{flox/flox}$ females as previously described (Gu et al., 2009). *Pygo2*$^{flox/flox}$ and *Pygo2*$^{flox/+}$ female littermates were generated by crossing *Pygo2*$^{flox/+}$ males with *Pygo2*$^{flox/flox}$ females. All experiments have been approved by and conform to the regulatory guidelines of the International Animal Care and Use Committee of the University of California, Irvine.

Antibodies

α-Pygo2 rabbit polyclonal antibody was described previously (Li et al., 2007). α-K14 was a gift from J. Segre (National Institutes of Health, Bethesda, MD). Other antibodies were obtained commercially: α-SMA (Sigma-Aldrich); α-K8 and α-K19 (Developmental Studies Hybridoma Bank); α-β-actin (Abcam); α-Pygo2 [(Gu et al., 2009); GeneTex, GTX119726]; normal rabbit IgG and α-β-catenin (Santa Cruz Biotechnology, H-102, sc-7199); α-H3, α-H3K4me3, α-H3K9me3, and α-H3K27me3 (Cell Signaling Technology); α-Notch3-PE

(eBioscience); α-CD29-FITC and IgG-FITC (BioLegend); α-CD31-APC, α-CD45-APC, α-TER119-APC, IgG-APC, α-CD24-PE-Cy7, IgG-PE-Cy7, α-CD61-PE, IgG-PE, and α-CD49f-FITC (BD Biosciences).

Cell culture and generation of recombinant lentiviruses

MCF10A cells, 293T, and HC11 cells were cultured as previously described (Ball et al., 1988; Gu et al., 2009).

Recombinant lentiviruses expressing Pygo2 were generated following the instruction provided by Addgene and as previously reported (Gu et al., 2012). Three days after infection, cells were lysed in high salt buffer and prepared for Western blot as previously described (Gu et al., 2012). shRNA against Cbf-1 was generated based on the previously reported shRNA target sequence (Bouras et al., 2008). Briefly, annealed oligos (forward: 5'-CCGGGCACAGAAGTCTTACGGAAATCTCGAGATTTCCGTAAGACTTCTGTGCTTTTTG-3'; reverse: 5'-AATTCAAAAAGCACAGAAGTCTTACGGAAATCTCGAGATTTCCGTAAGACTTCTGTGC-3') were cloned into the AgeI/EcoRI site of pLKO.1-ZsGreen which has been generated from pLKO.1 (Addgene) by replacing the Puro R cassette with ZsGreen. The ΔN-βcatenin lentiviral construct was generated by cloning a Myc-tagged, N-terminal truncated form (lacking the first 89 amino acids) of human β-catenin into the NotI (blunt-ended)/XbaI site of the pHIV-Zsgreen construct (Welm et al., 2008). Production of lentiviruses was carried out as previously described (Gu et al., 2012). Transduction unit (TU) of viral solution was estimated by measuring GFP-positive population using FACS analysis.

Mammary cell preparation, flow cytometry, and sorting

Mammary glands were dissected from 8-12 weeks old virgin mice. Mammary cells were isolated as described (Gu et al., 2009) except that collagenase digestion was performed for 1.5 hr digestion at 37ºC with shaking. Single cell suspension was immuno-labeled with specific antibodies (CD31-APC, CD45-APC, TER119-APC, CD24-PE-Cy7, CD29-FTIC, CD49f-FITC, CD61-PE or Notch3-PE) in PBS containing 2% FBS for 25 minutes at room temperature as described (Shackleton et al., 2006; Stingl et al., 2006). After washing once with PBS/2% FBS, cells were analyzed by LSRII (BD Biosciences) or sorted into specific subpopulations by FACSAriaII (BD Biosciences). Gating was set to exclude cells labeled with isotype-matched control IgG conjugated with the corresponding fluorochromes. Immediate post-sorting analysis confirmed that the cell populations were at least 98% pure. Data analysis was performed by Flowjo 7.6.1.

3D-Matrigel differentiation assay

Mammary glands were collected from 8-12-week old control and SSKO virgin mice and digested into dispersed cells. After labeling and sorting, single cells of the MaSC/basal population were gently resuspended in 100% chilled growth factor reduced Matrigel (BD Biosciences) and plated onto 8-well chamber slide (Thermo Fisher Scientific) at $2x10^4$ cells/50 μL Matrigel/well. After set, the gel was covered with 400 μL mammary epithelial growth medium [EpiCult-B basal medium containing 1x proliferation supplements (Stem Cell Technologies), 10 ng/mL epidermal growth factor (Millipore), 10 ng/mL basic fibroblast growth factor (Peprotech), and 4 μg/mL heparin (Stem Cell Technologies)], which was replaced every 3 days. After a week, the medium was switched to DMEM/F12 (Invitrogen) containing HIP [500

ng/mL hydrocortisone (Calbiochem), 5 μg/mL insulin (Sigma-Aldrich), 5 μg/mL Prolactin (Sigma-Aldrich)], and 1% FBS to induce luminal/alveolar differentiation as previously reported (Shackleton et al., 2006).

To acutely delete *Pygo2*, sorted *Pygo2*$^{flox/flox}$ and control *Pygo2*$^{flox/+}$ MaSC/basal cells were infected in suspension with Ade-Cre (Vector Biolabs) at 50 multiplicity of infection in DMEM/F12 medium containing 2% FBS at room temperature for 90 minutes before embedding in Matrigel. To modulate signaling pathways, growing cells/colonies were pretreated with DMSO, DAPT (Calbiochem) or BIO (Sigma-Aldrich) for 24 hours in growth medium before switching to differentiation medium (referred to as Day 0) containing the same agent. Images were taken using a MZFLIII fluorescent dissecting scope (Leica) equipped with a Spot Insight camera (Diagnostic Instruments).

Cleared fat pad transplantation

Sorted MaSC/basal populations from control or *Pygo2* SSKO mice were prepared as described above. For limiting dilution transplantation, cells of each genotype were counted and resuspended in a 1:1 solution of PBS/Matrigel (BD Biosciences) at desired concentrations (10,000, 5,000, 2,500, 1,250, 500, and 250 cells/10μL). For shCbf-1 knockdown experiment, $2x10^4$ sorted MaSC/basal cells were transduced with 10^5 TU of lentiviruses (multiplicity of infection = 5) via centrifugation at 1,800 rpm for 90 minutes at room temperature in a volume of 200 μL. Transduced cells were washed with MEC isolation medium as described previously (Gu et al., 2009) and resuspended in 1:1 PBS/Matrigel solution at a concentration of 5,000 cells/10 μL. Cell/Matrigel solutions were then injected in a 10-μL volume into contralateral

cleared fat pads of inguinal mammary glands of 3-week old female C57B/6 mice using a custom-ordered 26-G needle attached to a 50-μL Hamilton glass syringe.

Outgrowths were analyzed eight weeks after transplantation. For GFP imaging, surgically dissected glands were compressed between two glass slides and fluorescent ducts were visualized with a MZFLIII fluorescent dissecting scope (Leica) or Nikon E600 fluorescent microscope equipped with a Spot Insight camera (Diagnostic Instruments). A NeuronJ package for ImageJ software was used to quantify the total ductal length in a mammary transplant. Whole-mount staining was described previously (Gu et al., 2009). Statistical analysis of the take rate was performed using the ELDA Web-based tool (http://bioinf.wehi.edu.au/software/elda/), and the statistical difference in filled fat pad area was estimated by one-way ANOVA.

RNA isolation and RT-qPCR

Cells were freshly sorted from mouse mammary gland or recovered from 3D-Matrigel culture with 5 mg/mL dispase (Stem Cell Technologies) at 37 ºC for 20 minutes. Total RNA was purified using RNeasy Mini Kit with on-column DNase treatment according to manufacturer's protocol (QIAGEN). Total cDNA was synthesized from RNA using High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) according to manufacturer's instructions. qPCR was performed using SYBR Green Supermix (Biorad Laboratories) on iCycler iQ Multicolor Real Time PCR Detection System (Biorad Laboratories) following manufacturer's instructions. GAPDH, β-actin, or 18S served as an internal control. Primers for amplified loci were described in Table S2. ΔΔCt method was used to determine the fold change and converted values were normalized to the average of control samples. PCR specificity was monitored by running melting curve for each primer/reaction.

Indirect immunofluorescence and confocal imaging

The detection of SMA, K19, K14, and K8 in mammary ductal sections or sorted cells were performed as previously described (Gu et al., 2009). For quantification of SMA- and K19-positive cells in sections of control and SSKO mammary ducts, about 10 sections per individual mouse were analyzed, and an average value of $K19^{+}/SMA^{+}$ ratio was calculated for each mouse. *p* values were calculated from multiple mice using two-tailed *t*-test.

For detection of K14 and K8 in 3D-Matrigel colonies, 18% and 30% sucrose in PBS was applied sequentially to colonies in Matrigel to maintain their structure. Immunostaining was then performed according to the protocol described previously (Fata et al., 2007). Images were acquired using Zeiss LSM 700 Laser Scanning Confocal Microscope (Carl Zeiss, Inc.). Creation of 3D reconstruction from multiple focal plains was carried out using Zen software (Carl Zeiss, Inc.) with maximum transparency mode.

Micro chromatin immunoprecipitation

MicroChIP was performed according to the protocol described previously (Dahl and Collas, 2008). Briefly, $5x10^{4}$ sorted cells were cross-linked with 1% formaldehyde at room temperature for 8 minutes. Sonication condition was optimized by both qPCR and agarose gel electrophoresis. Chromatin was sheared into 500-1500 bp fragments on ice using Sonicator 3000 (Misonix). Each lysate was diluted and divided into 8-tube strip of PCR tubes, and incubated at 4 ºC overnight with control IgG or specific antibodies. The immuno-complexes were precipitated by Dynabeads Protein A (Invitrogen) at 4 ºC for 2 hours, washed extensively, and eluted. After reversal of cross-linking, the immunoprecipitated DNA was recovered by Phenol/Chloroform

extraction followed by ethanol precipitation. The DNA was quantified by qPCR as described above. The primers for *Notch3* promoter were summarized in Table S2.

Microarray, GSEA, and statistics

Hybridization of arrays (GeneChip® Mouse Gene 1.0 ST Array, Affymetrix) was performed in triplicate using independent biological samples. Data were submitted to GEO repository (No. GSE45826). For each population analyzed, RNAs from two age-matched mice with the same genotype were pooled (1:1) to minimize individual variation. Genes with normalized expression levels over detection threshold were called and analyzed for differential expression using the Cyber-T program (Baldi and Long, 2001) (http://cybert.ics.uci.edu/). Cut-off *p*-value of *t*-test was less than 0.05 (Bayes.lnp) for both comparisons. False discovery rate (FDR) was calculated as 1- posterior probability of differential gene expression [PPDE(<p)]. The overlapped genes were used to generate the heat map by TMEV 4.5.1, using hierarchical clustering with the setting of Euclidean distance and average linkage clustering. Fisher's exact test (21,064 total genes) was used to calculate the *p* value of overlap between lists 1 and 2 in Figure S2D. GSEA was performed using published mammary basal/luminal gene sets (Table S1) as instructed from the Broad institute (http://www.broadinstitute.org/gsea/index.jsp), with 1000 permutations of samples. Two-tailed student *t*-test was used to calculate *p*-values between two groups of samples.

SUPPLEMENTAL REFERENCES

Baldi, P., and Long, A.D. (2001). A Bayesian framework for the analysis of microarray expression data: regularized t -test and statistical inferences of gene changes. Bioinformatics *17*, 509-519.

Ball, R.K., Friis, R.R., Schoenenberger, C.A., Doppler, W., and Groner, B. (1988). Prolactin regulation of beta-casein gene expression and of a cytosolic 120-kd protein in a cloned mouse mammary epithelial cell line. EMBO J *7*, 2089-2095.

Fata, J.E., Mori, H., Ewald, A.J., Zhang, H., Yao, E., Werb, Z., and Bissell, M.J. (2007). The MAPK(ERK-1,2) pathway integrates distinct and antagonistic signals from TGFalpha and FGF7 in morphogenesis of mouse mammary epithelium. Dev Biol *306*, 193-207.

Li, B., Rheaume, C., Teng, A., Bilanchone, V., Munguia, J.E., Hu, M., Jessen, S., Piccolo, S., Waterman, M.L., and Dai, X. (2007). Developmental phenotypes and reduced Wnt signaling in mice deficient for pygopus 2. Genesis *45*, 318-325.

Welm, B.E., Dijkgraaf, G.J., Bledau, A.S., Welm, A.L., and Werb, Z. (2008). Lentiviral transduction of mammary stem cells for analysis of gene function during development and cancer. Cell Stem Cell *2*, 90-102.