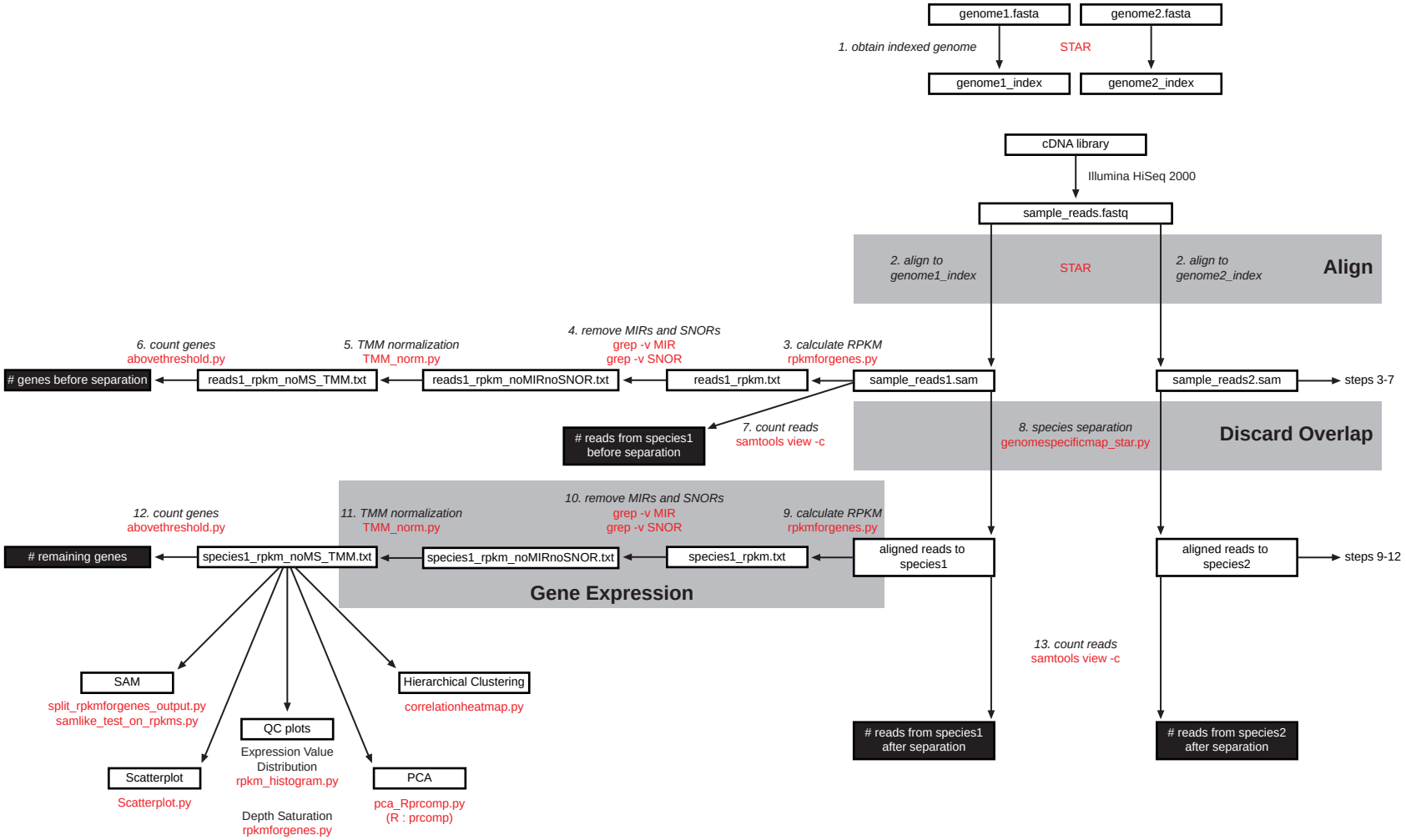


Figure S1. Pipeline of data processing.
This figure shows the steps used to produce the expression values (steps 1,2,8-11) as well as how to calculate numbers (black boxes) that are referenced in the formulas in Table S3. The grey boxes highlight the steps after cDNA library preparation and sequencing, indicated in Figure 1A.

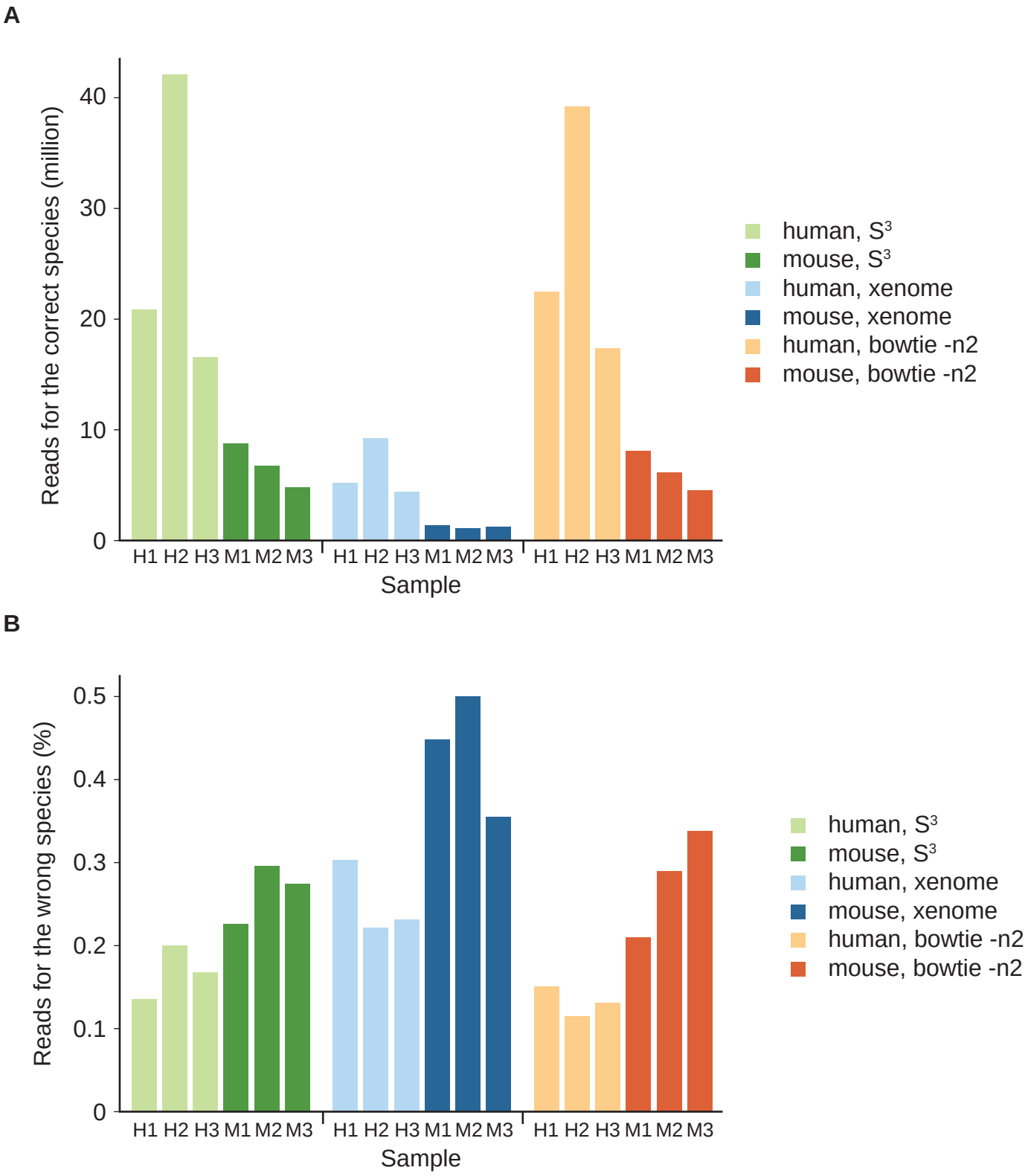


Figure S2. Comparison with other methods to separate human and mouse reads, both in terms of sensitivity (A) and specificity (B). We compared STAR-based S^3 to another algorithm of ours (genome+transcriptome bowtie alignment allowing two mismatches in the seed region) and to the species separation method Xenome (Conway et al. 2012), followed by STAR alignment with default parameters. M1-3 are mouse samples, H1-3 are human samples.

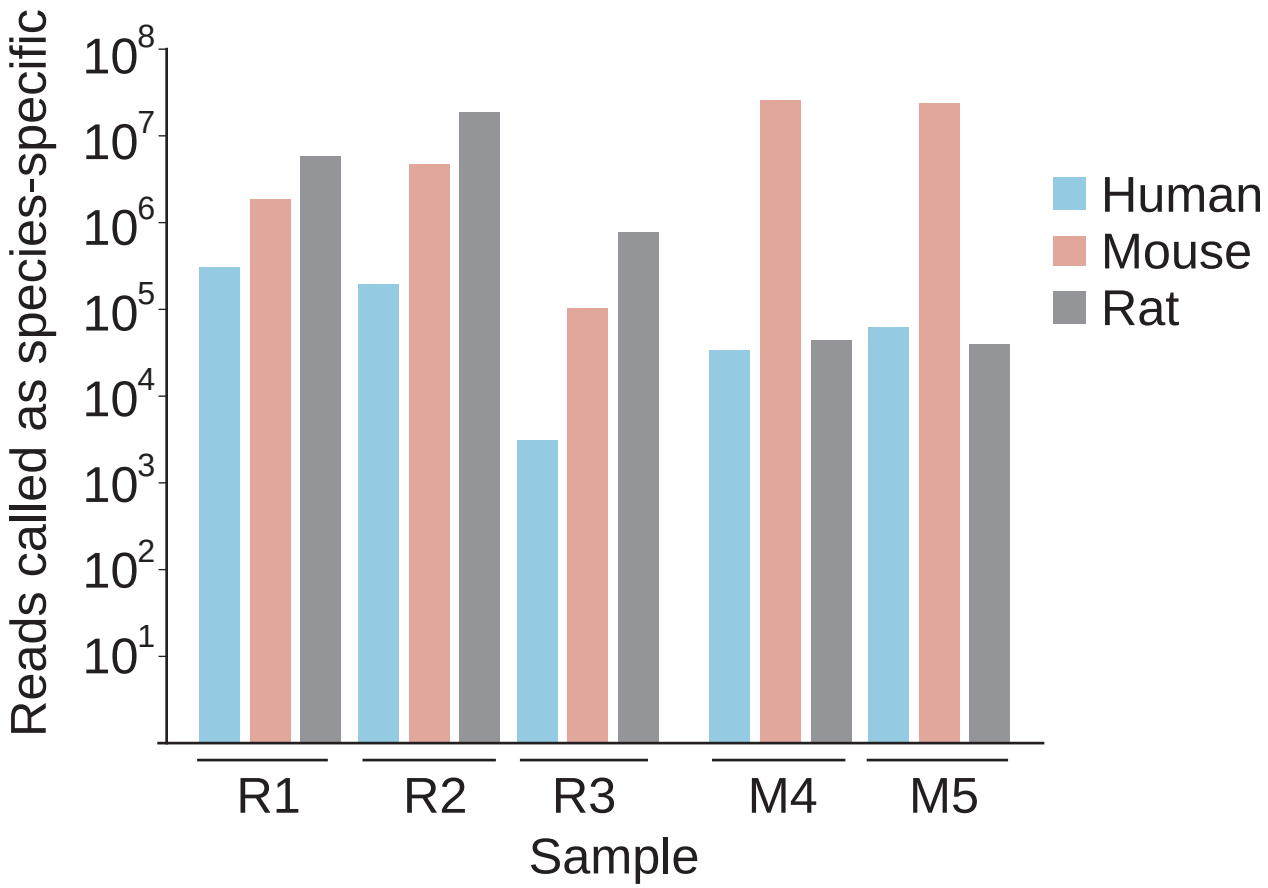


Figure S3. Comparison of three species (rat, mouse, human) separation of rat (R1-R3) and mouse (M4, M5) RNA-seq samples, similar to Figure 1F but with absolute number of reads on the y axis. The mouse samples have a 50 bp read length, similar to the 51 bp of R2.

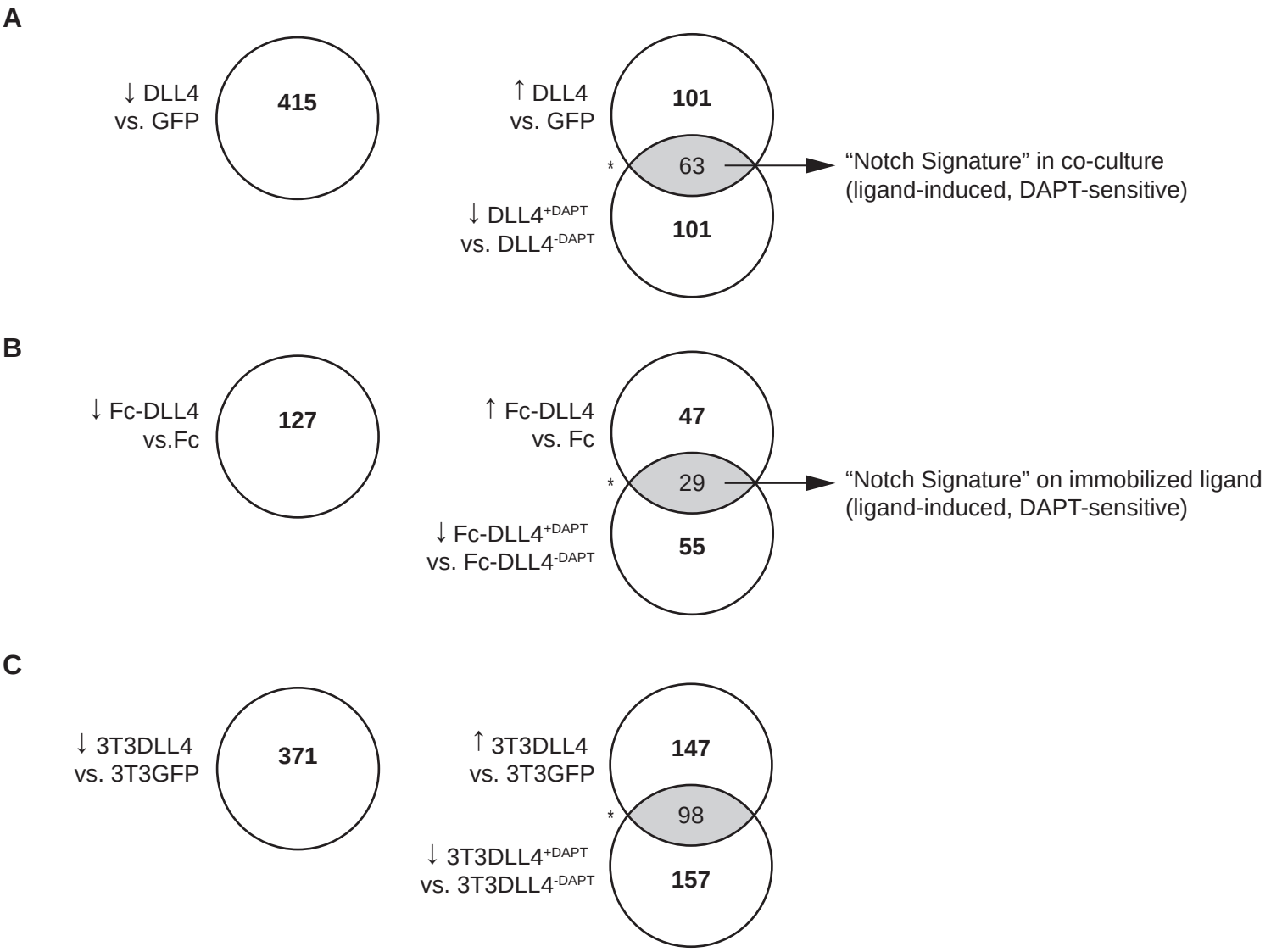
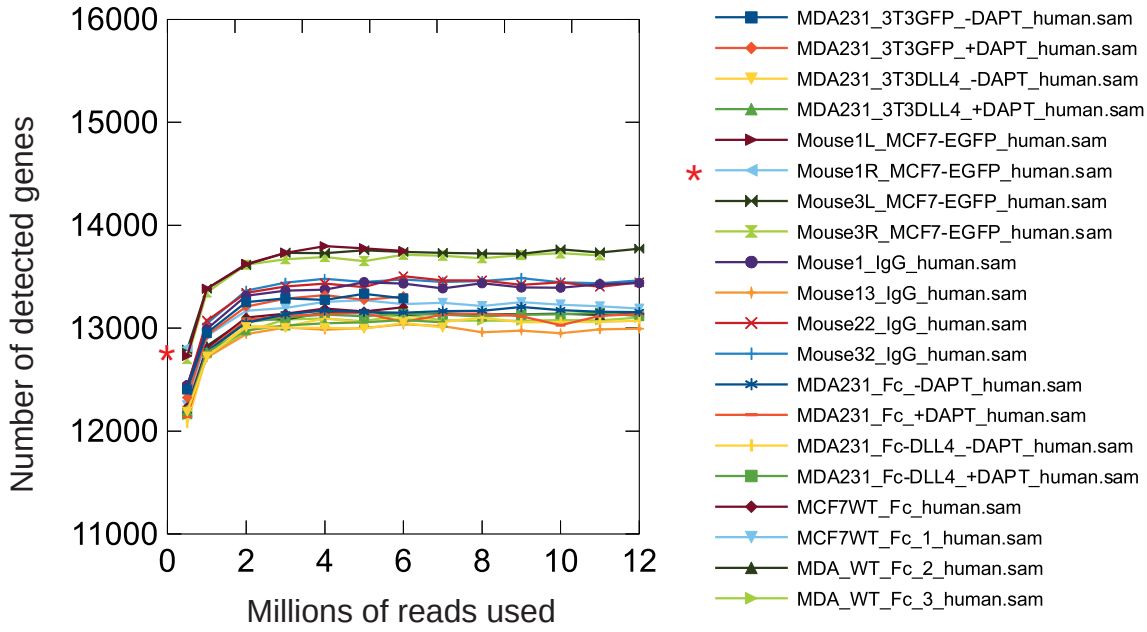


Figure S4. Comparison of the number of genes up- and downregulated in (A) MDA-MB-231 cells co-cultured with 3T3-L1 cells expressing DLL4 or GFP, (B) MDA-MB-231 cells on immobilized Fc-DLL4 or Fc, and (C) 3T3-L1 cells, expressing DLL4 or GFP, co-cultured with MDA-MB-231 cells. In (A) and (B), a "Notch Signature" is defined as genes upregulated through ligand induction and downregulated after DAPT treatment. Up- and downregulation is defined as FC>2. * signifies $p<0.05$ (Fisher's exact test).

Human expression values:



Mouse expression values:

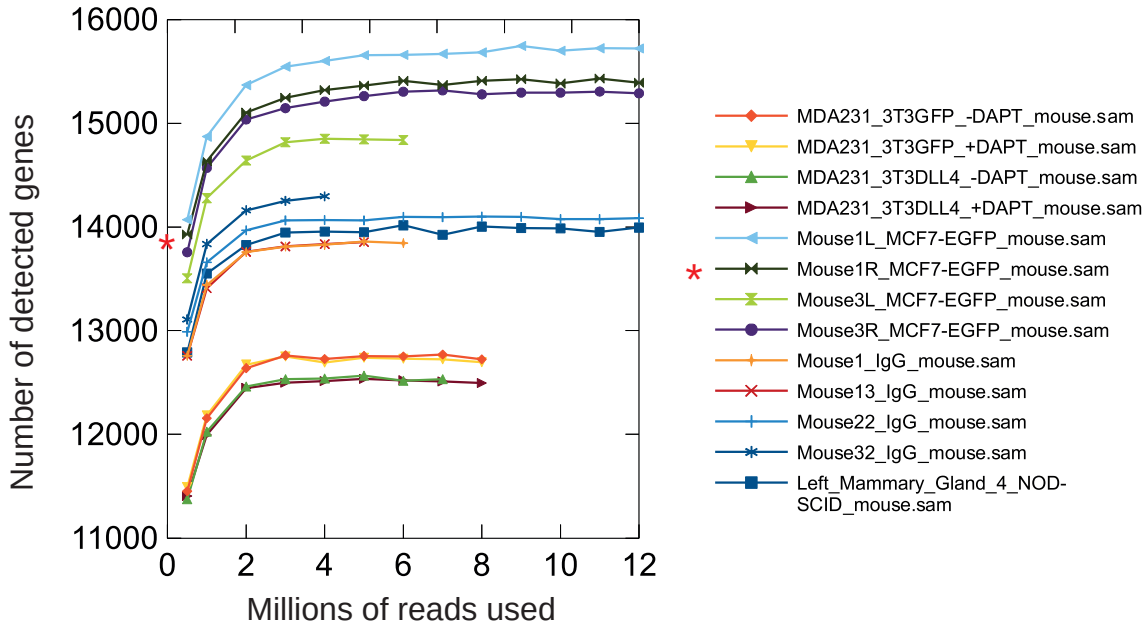
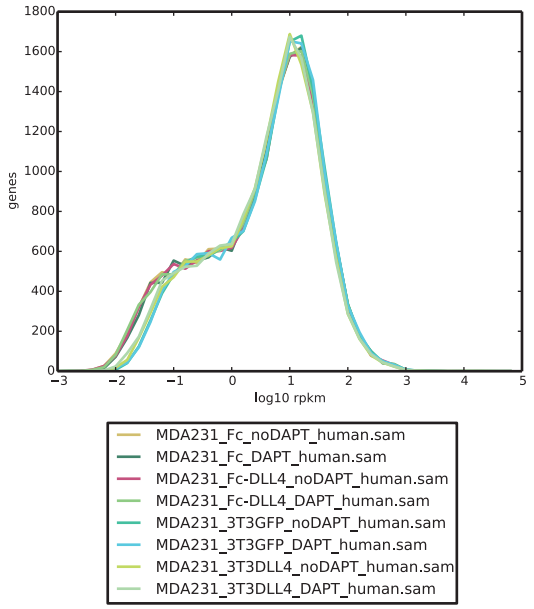
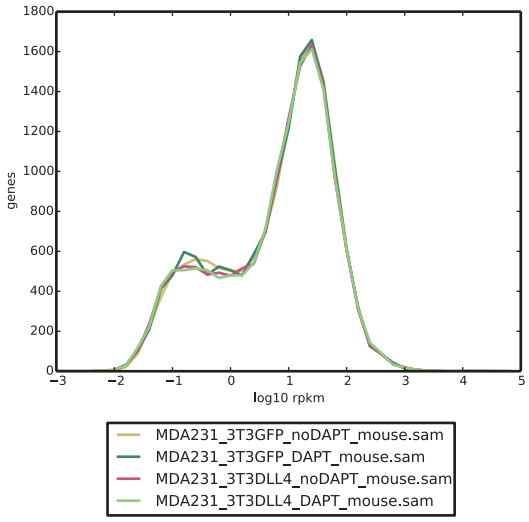


Figure S5. Whole-genome gene expression QC: Depth Saturation.
The graphs show the number of expressed genes (>0.3 RPKM) after downsampling to 0.5-12 million mapped reads. All samples generated in the study are shown, including one sample marked by a red asterisk which was excluded for failing quality control. The total number of reads for each sample is listed under “# reads” in Additional file 4: Table S3. Samples containing cell lines yielded fewer detected genes than samples containing mouse stroma, presumably because the latter is a more complex tissue containing many cell types. As seen above, gene detection levels out after 2-3 million reads, which is less then our sequencing depths (excluding the sample marked by *).

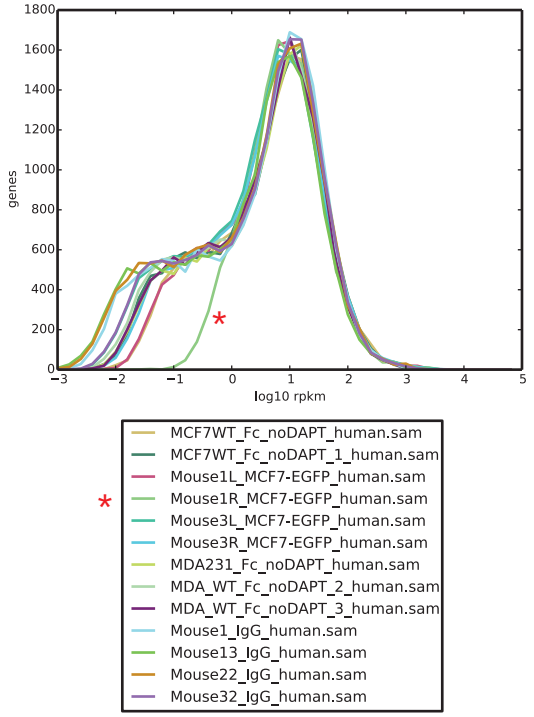
***in vitro* (human):**



***in vitro* (mouse):**



***in vivo* (human):**



***in vivo* (mouse):**

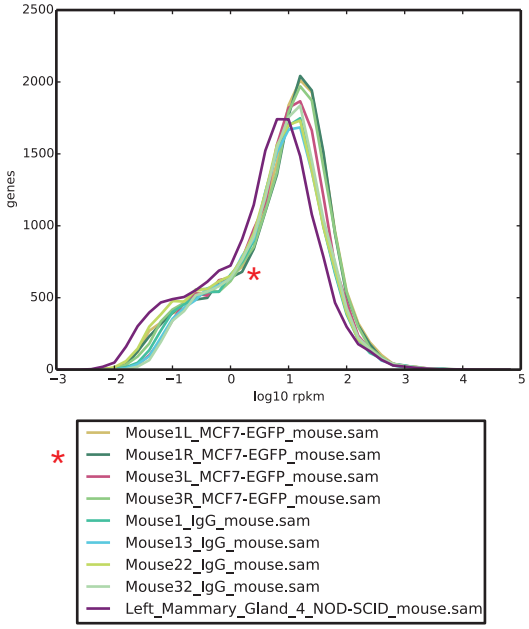


Figure S6. Whole-Genome Gene Expression QC: Density Plots after TMM Normalization. The plots show RPKM distributions for all samples generated in the study, including one sample marked by a red asterisk which was excluded for failing quality control (see also Figure S5).

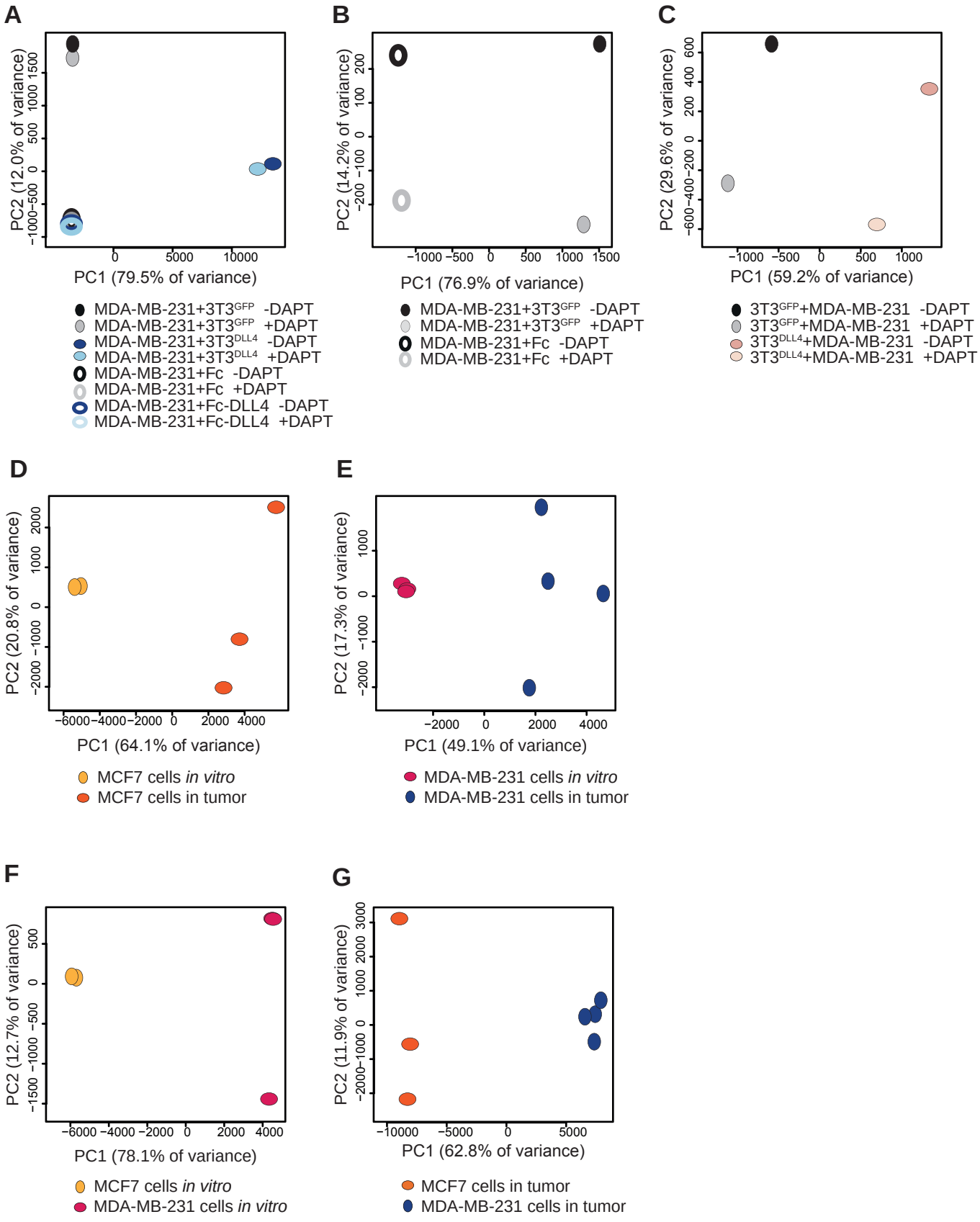


Figure S7: Additional principal component analyses (PCA). These use the expression values of all human or mouse genes and show the first and second principal components. Additional sample information can be found in Additional File 4: Table S3.

(A) MDA-MB-231 (DLL4-ligand-activated) cells cultured with 3T3-L1 cells expressing/not expressing DLL4 or on immobilized ligand and in the presence or absence of DAPT, as indicated.

(B) MDA-MB-231 cells cultured with 3T3-L1 cells or on Fc-fragment.

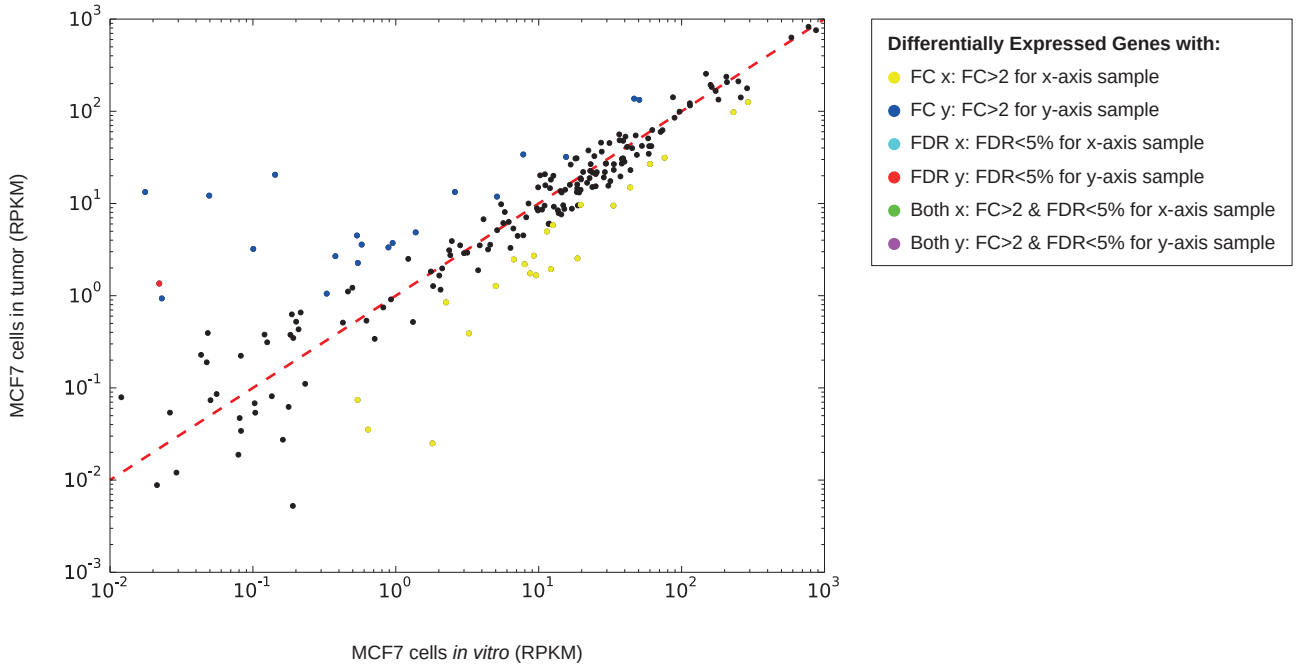
(C) 3T3-L1 (ligand-expressing) cells in co-culture.

(D) MCF7 cells *in vitro* (left to right: Fc_1, Fc) and in tumor (top to bottom: 1L, 3L, 3R).

(E) MDA-MB-231 cells *in vitro* (top to bottom: Fc, Fc_2, Fc_3) and in tumor (top to bottom: 22, 32, 13, 1).

(F) MCF7 (left to right: Fc, Fc_1) and MDA-MB-231 (top to bottom: Fc_2, Fc_3, Fc) cells *in vitro.*

(G) MCF7 (top to bottom: 1L, 3L, 3R) and MDA-MB-231 (top to bottom: 13, 32, 22, 1) cells in tumor.



Gene Set:

ADAM17	CTNNAL1	IKBKB	MMP9	PLCB4	SHC1	VEGFA
AKT1	CTNNBIP1	IKBKE	MNAR	PLCD1	SHC2	VEGFB
AKT2	CTNND1	IKBKG	MTA3	PLCD3	SHC3	VEGFC
AKT3	CTNND2	IL6	MYBBP1A	PLCD4	SHC4	ZEB2
AP1B1	DDX5	ITGA2	NCOA1	PLCE1	SMARCA1	
AP1G1	E2F1	JUN	NCOA2	PLCG1	SMARCA2	
AP1M1	E2F2	KAT3A	NCOA3	PLCG2	SMARCA4	
AP1M2	EBAG9	KRAS	NCOR1	PLCH1	SMARCA5	
AP1S1	EGF	LAMA5	NCOR2	PLCH2	SMARCAL1	
AP1S2	EGFR	LDLR	NEDD8	PLCXD1	SMARCB1	
AP1S3	ELK1	MAPK1	NFKB1	PLCXD2	SMARCC1	
ARAF	EMILIN1	MAPK10	NFKB2	PLCXD3	SMARCC2	
AREG	EMILIN2	MAPK11	NFKBIA	PLCZ1	SMARCD1	
AREGB	EMILIN3	MAPK12	NFYA	PRKACA	SMARCD2	
ATF1	EP300	MAPK14	NFYB	PRKACB	SMARCD3	
ATF2	EREG	MAPK3	NFYC	PRKACG	SMARCE1	
BCL2	ESR1	MAPK7	NOS1	PRKCA	SNAI1	
BCL2L1	ESR2	MAPK8	NOS2	PRKCB	SNAI2	
BMP7	ESRRA	MAPK9	NOS3	PRKCD	SNAI3	
BRAF	ESRRB	MED1	NR0B2	PRKCE	SNRK	
BTC	ESRRG	MMP1	NRAS	PRKCG	SOL4A6	
CARM1	FGF10	MMP10	NRIP1	PRKCH	SP1	
CAV1	FGFR1	MMP11	OCLN	PRKCI	SRC	
CAV2	FGFR2	MMP12	PELP1	PRKCQ	STAT1	
CAV3	FGFR3	MMP13	PIK3C2A	PRKCZ	STAT3	
CBP	FGFR4	MMP14	PIK3C2B	PRL	STAT5A	
CCND1	FOS	MMP15	PIK3C2G	PRMT1	STAT5B	
CD44	GJB1	MMP16	PIK3C3	RAF1	STRIP1	
CDH1	GPER1	MMP17	PIK3CA	RARA	STRIP2	
CDH3	GPR182	MMP19	PIK3CB	RARB	STRN	
COL4A1	GRB2	MMP2	PIK3CD	RARG	STRN3	
COL4A2	GSK3B	MMP20	PIK3CG	REL	STRN4	
COL4A3	HBEGF	MMP21	PIK3R1	RELA	TGFA	
COL4A4	HDAC4	MMP24	PIK3R2	RELB	TGFB1	
COL4A5	HSP90AA1	MMP25	PIK3R3	RSTS	TIMP2	
COL7A1	HSP90AB1	MMP26	PIK3R4	SAFB	TRAP	
CREBBP	HSPB1	MMP27	PIK3R5	SAFB2	TWIST1	
CSN2	HSPB2	MMP28	PIK3R6	SAP30	TWIST2	
CTNNA1	HSPB3	MMP3	PLCB1	SERPINB5	UCP1	
CTNNA2	HSPD1	MMP7	PLCB2	SERPINB6	UCP2	
CTNNA3	IGF1	MMP8	PLCB3	SERPINB7	UCP3	

Figure S8. Scatterplot of genes in Estrogen-related signaling, differentially expressed in MCF7 cells *in vitro* and MCF7 cells in tumor.

Gene Set Sources:

Marino M, et al. Estrogen Signaling Multiple Pathways to Impact Gene Transcription. Curr Genomics. 2006 Nov; 7(8): 497–508.
Heldring N , et al. Estrogen Receptors: How Do They Signal and What Are Their Targets. Physiological Reviews. 1 July 2007; 87(3): 905-931
Tocris Bioscience. Estrogen Signaling Pathway. 2015. http://www.tocris.com/pathways/estrogenPathway.php#.VTZRkq2qpBc

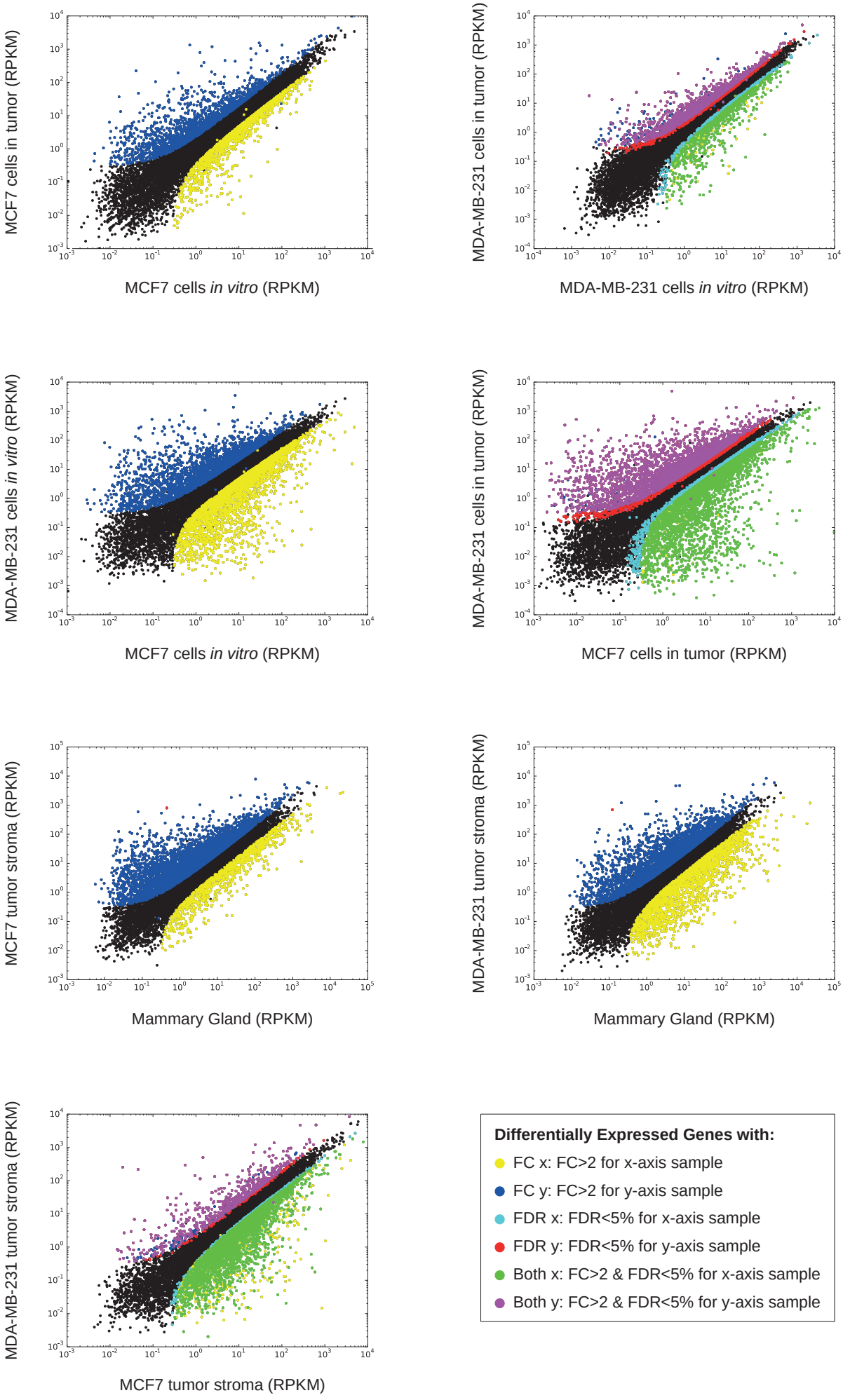
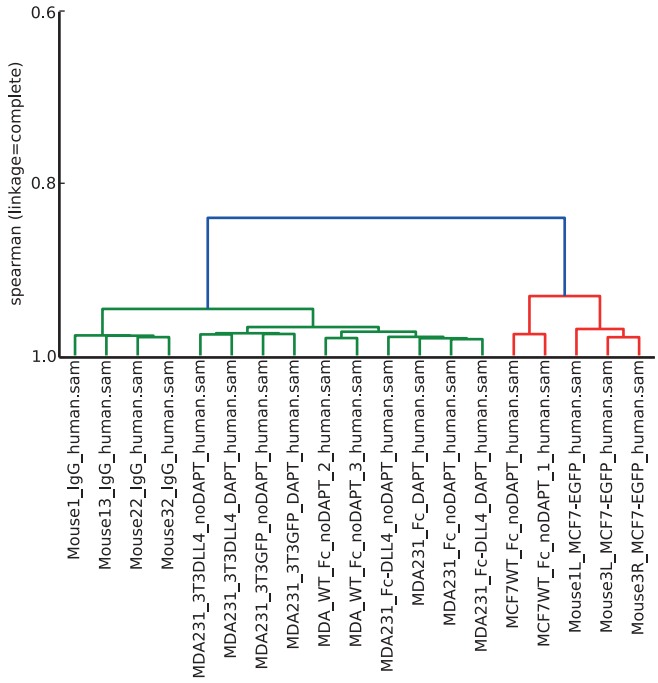


Figure S9. Scatterplots of Table S5 comparison groups.

A. Human expression values:



B. Mouse expression values:

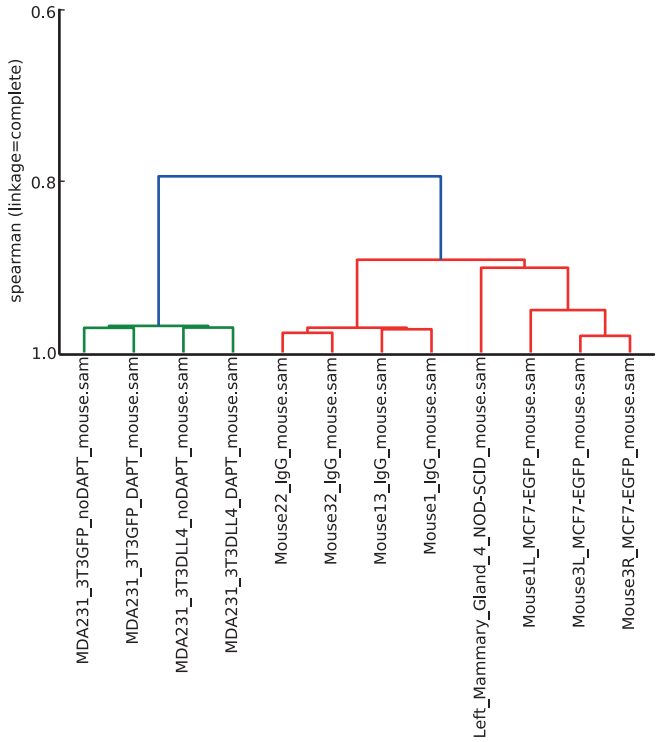


Figure S10. Hierarchical Clustering.
(A) shows human expression values, (B) shows mouse expression values. Human-only samples are only included in (A), mouse-only samples are only included in (B). Clustering is done by complete linkage neighbor-joining of 1-Spearman correlation coefficients. For human expression values, samples containing MDA-MB-231 cells cluster in green and samples containing MCF7 cells cluster in red. For mouse expression values, *in vitro* samples cluster in green and *in vivo* samples cluster in red. The mammary gland sample clusters more closely with MCF7 tumor stroma than MDA-MB-231 tumor stroma.

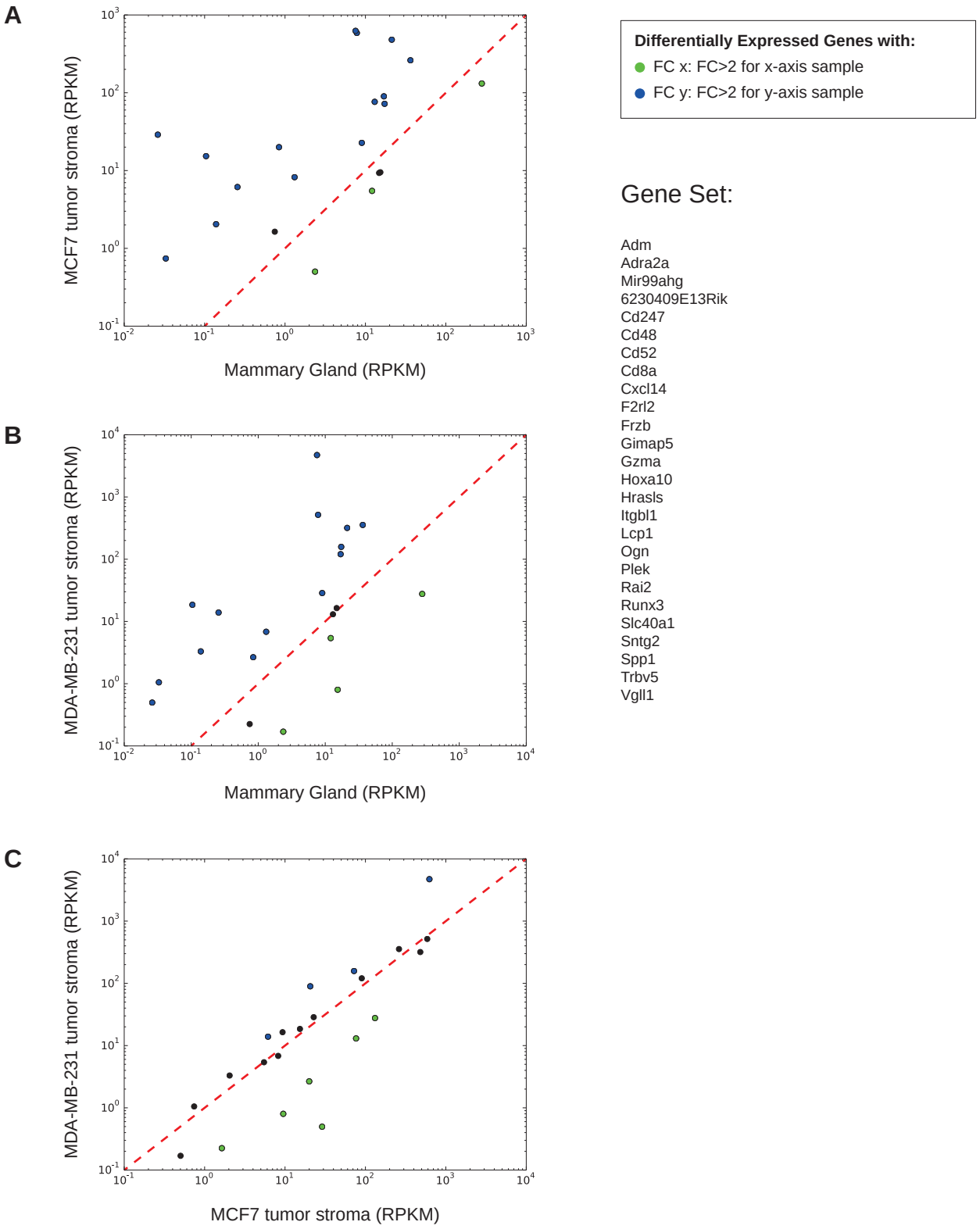


Figure S11. FINAK_BREAST_CANCER_SDPP_SIGNATURE: Scatterplot of genes in the stroma-derived prognostic predictor of breast cancer disease outcome (Finak et al, 2008), differentially expressed in (A) Mammary Gland (MG) compared to MCF7 tumor stroma, (B) MG compared to MDA-MB-231 tumor stroma, and MCF7 tumor stroma compared to MDA-MB-231 tumor stroma.
The Finak_SDPP human gene names were converted to mouse official gene symbols for this study through the use of GeneCards.org (Weizmann Institute of Science), Jax MGI, and NCBI AceView.

Gene Set Sources:
Finak G, et al. Stromal gene expression predicts clinical outcome in breast cancer. Nature Medicine 14, 518 - 527 (2008)
Broad Institute Gene Set Enrichment Analysis, Gene Set: FINAK_BREAST_CANCER_SDPP_SIGNATURE
Mouse Genome Informatics (MGI). The Jackson Laboratory. Mir99ahg. http://www.informatics.jax.org/marker/MGI:1919929
Mouse Genome Informatics (MGI). The Jackson Laboratory. Trbv5. http://www.informatics.jax.org/marker/MGI:98583
AceView. National Center for Biotechnology Information (NCBI). C6orf168. http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/av.cgi?db=human&c=Gene&l=C6orf168