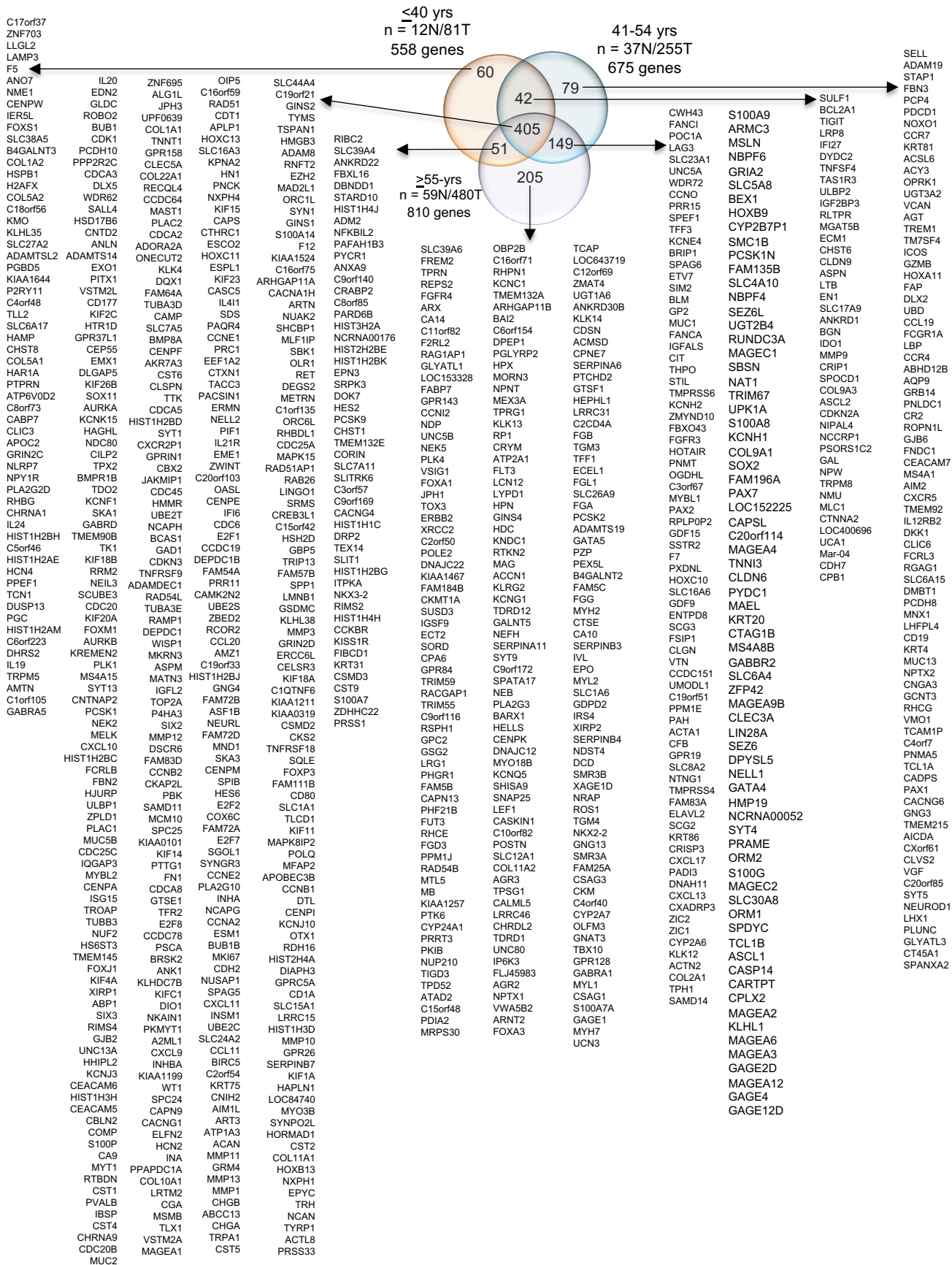
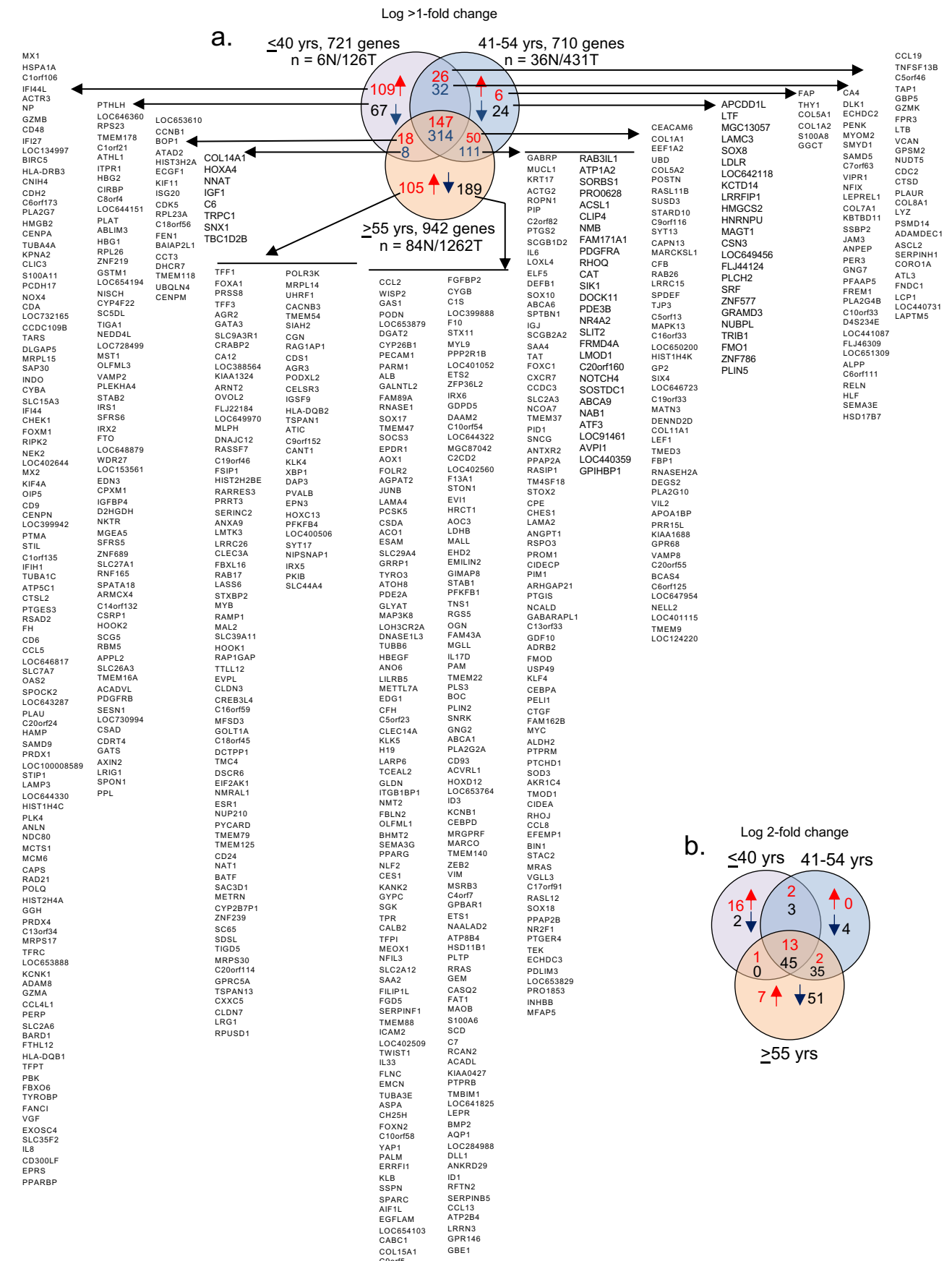

SUPPLEMENTARY FILE

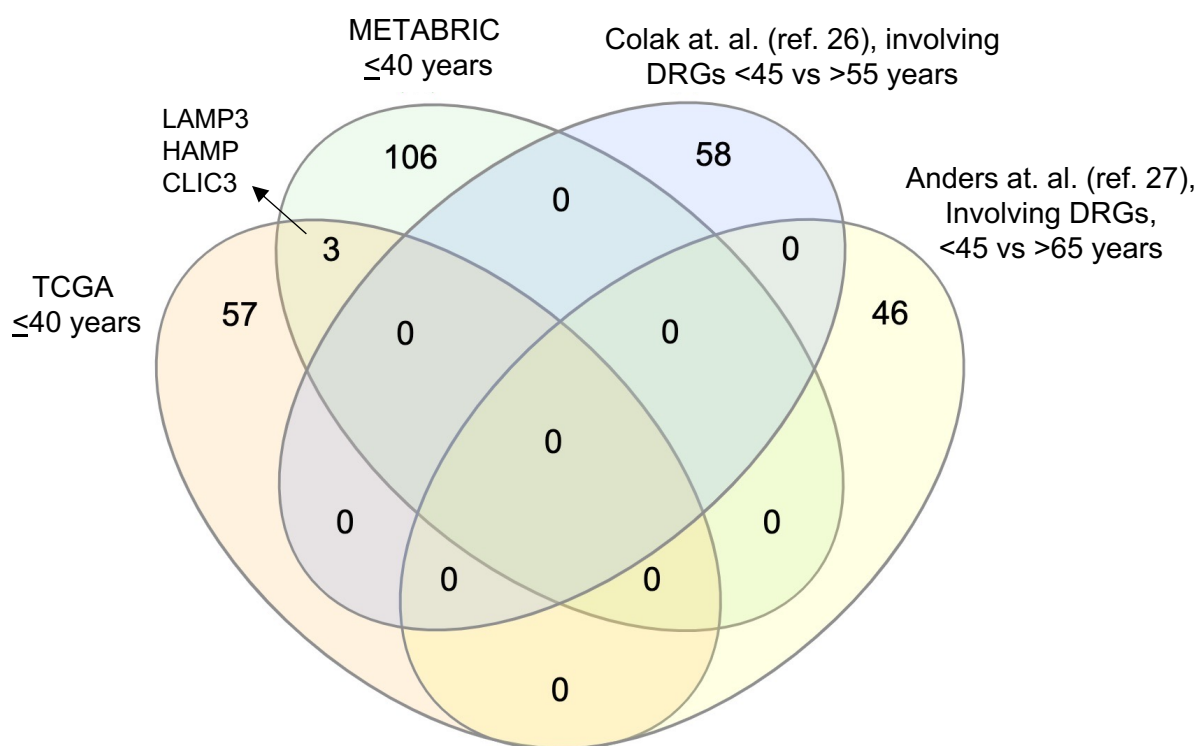
Delineation of Pathogenomic Insights of Breast Cancer in Young Women

**Aswathy Mary Paul, Bijesh George, Sunil Saini, Madhavan Radhakrishna Pillai,
Masakazu Toi, Luis Costa and Rakesh Kumar**

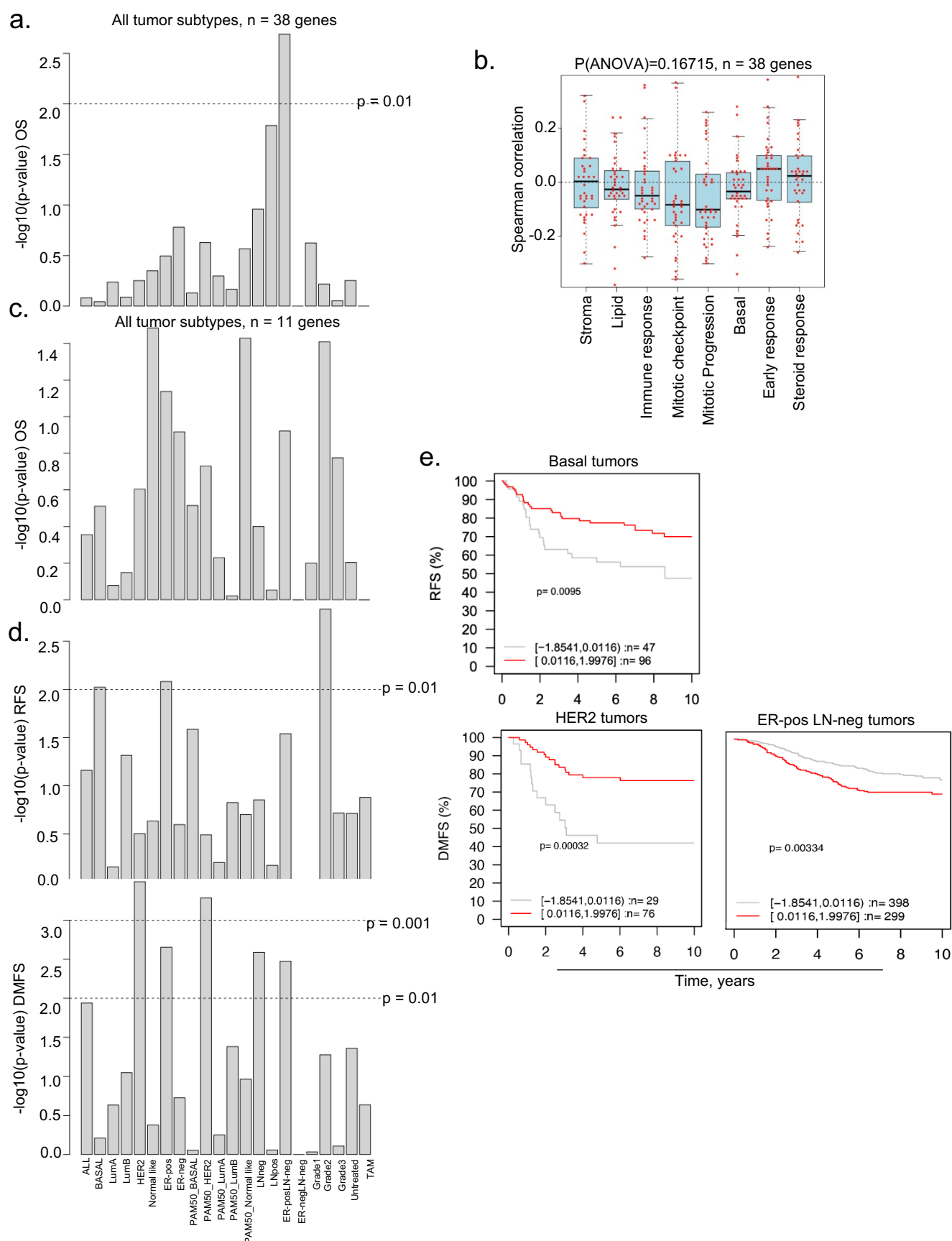




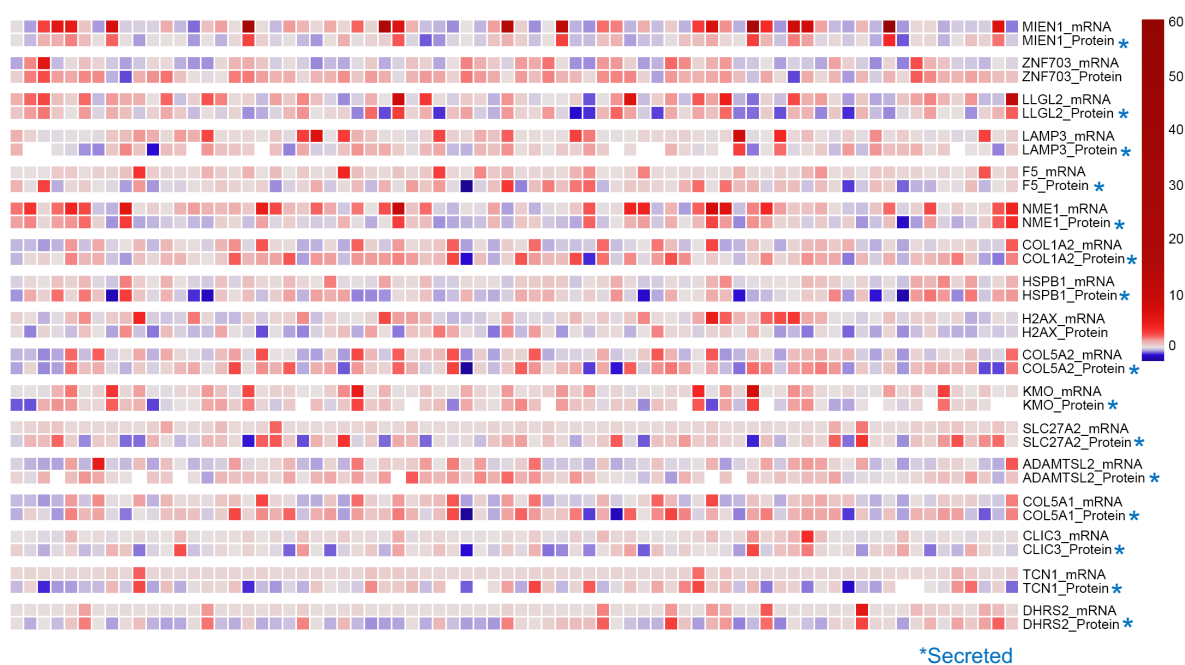




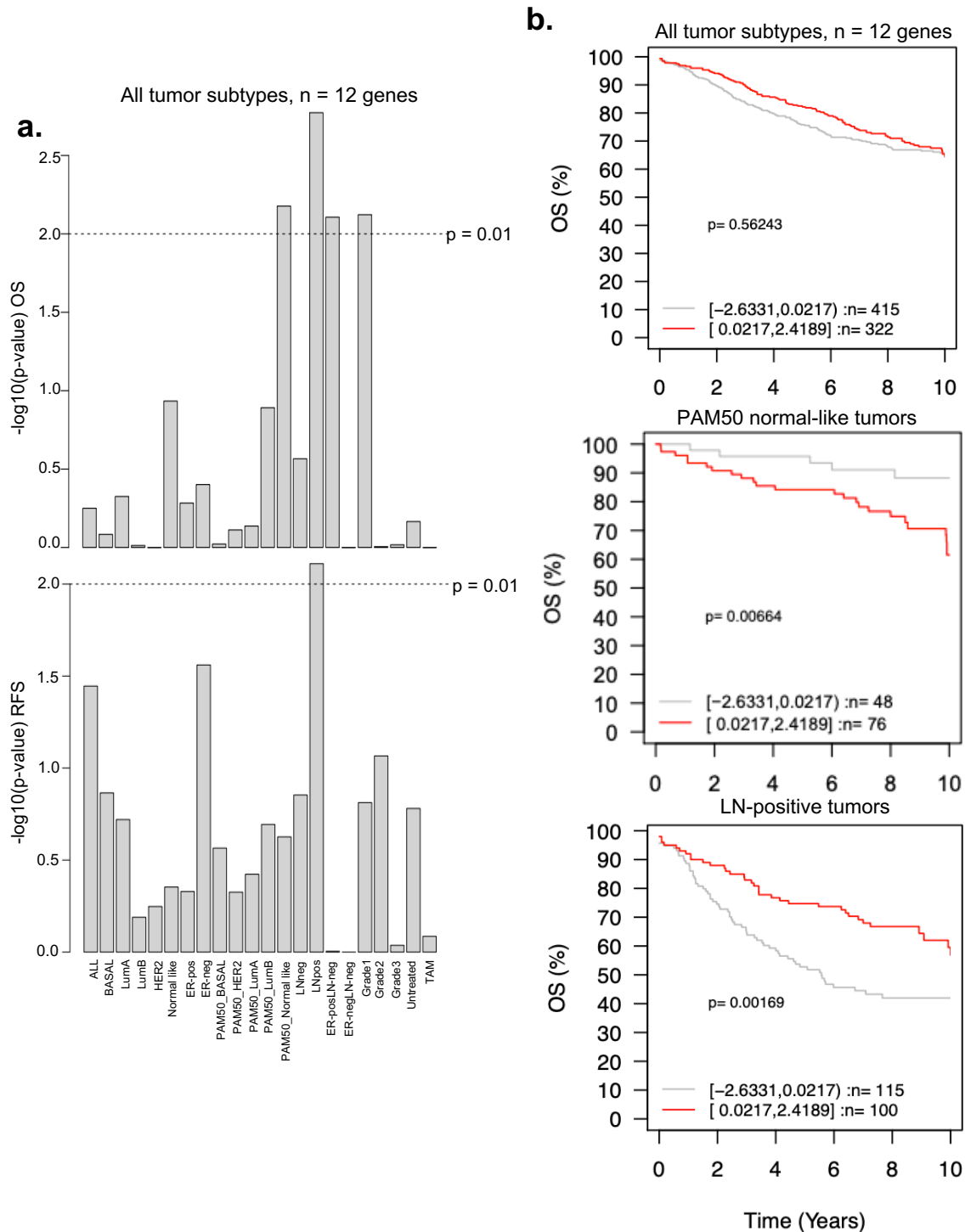
Supplementary Figure S4: Overlap of upregulated genes among four sets of differentially expressed genes reported in young women breast cancer.



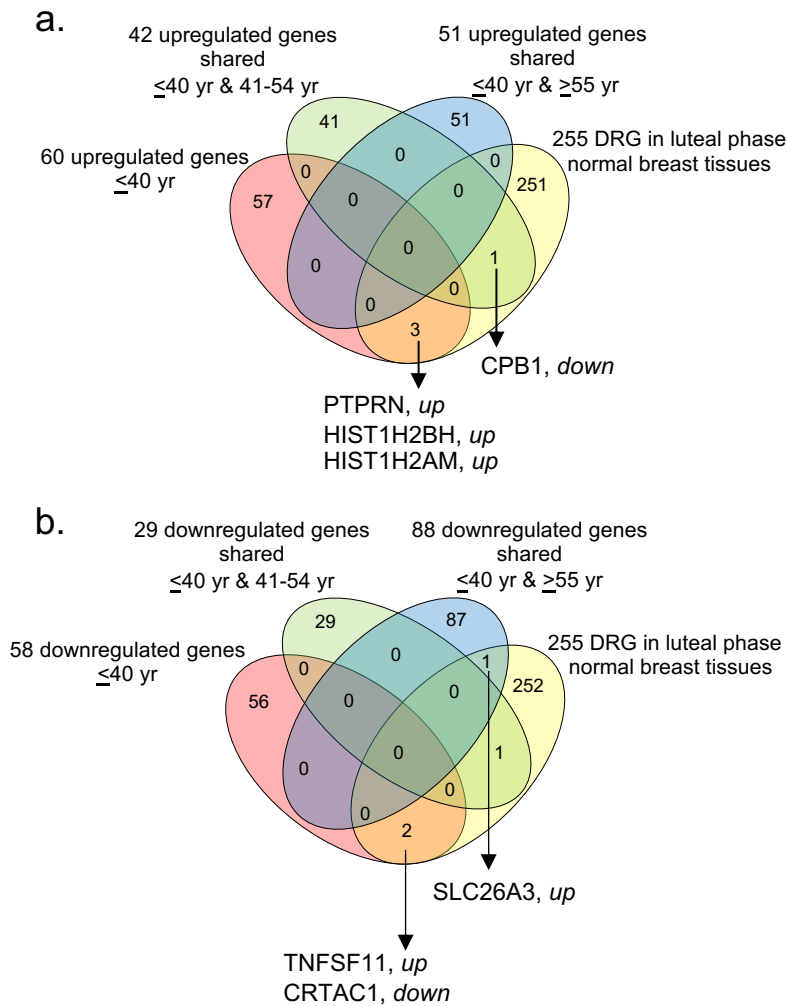
Supplementary Figure S5: Clinical significance of a subset of upregulated genes specific to ≤ 40 years with no overlaps with other age-groups. **a**, Kaplan-Meier OS summary of 38 out of 60 upregulated genes mapped in GOBO database with 1881 breast tumors; **b**, spearman correlation analysis of 38 genes in the context of co-expressed 8 gene modules; **c** and **d**, Kaplan-Meier OS, RFS and DMFS summaries of 11 out of 17 shortlisted genes in Fig. 1e, mapped in the GOBO database with 1881 breast tumors; **e**, RFS and DMFS curves using the high (red) and low (gray) expression of 11 genes for indicated breast tumor subtypes. All survival, correlation analysis performed using the online GOBO tools.



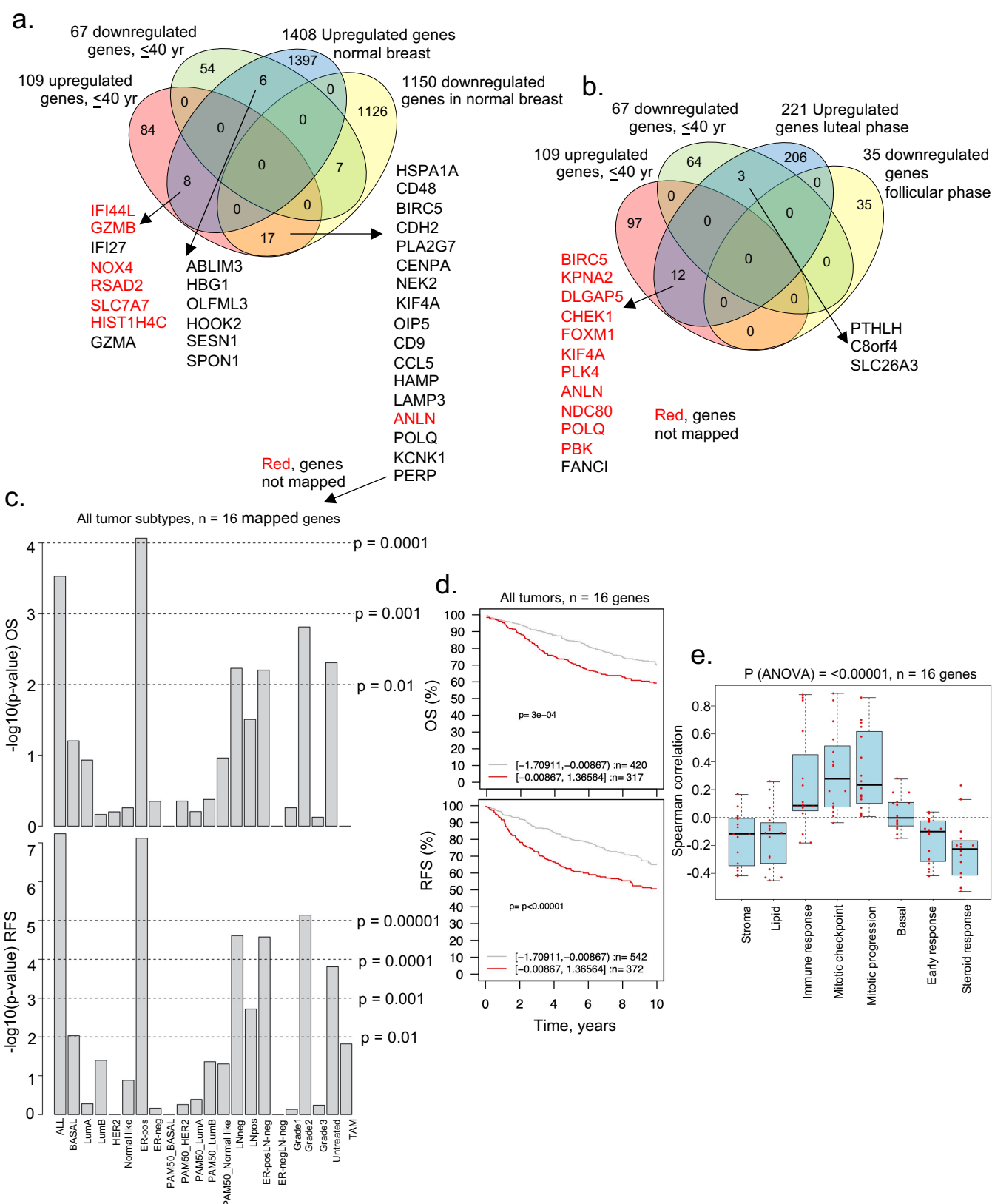
Supplementary Figure S6: Heatmap representing the expression of indicated at the levels of mRNA and protein in the same set of TCGA breast cancer samples, using the dataset curated from the cBioPortal platform.



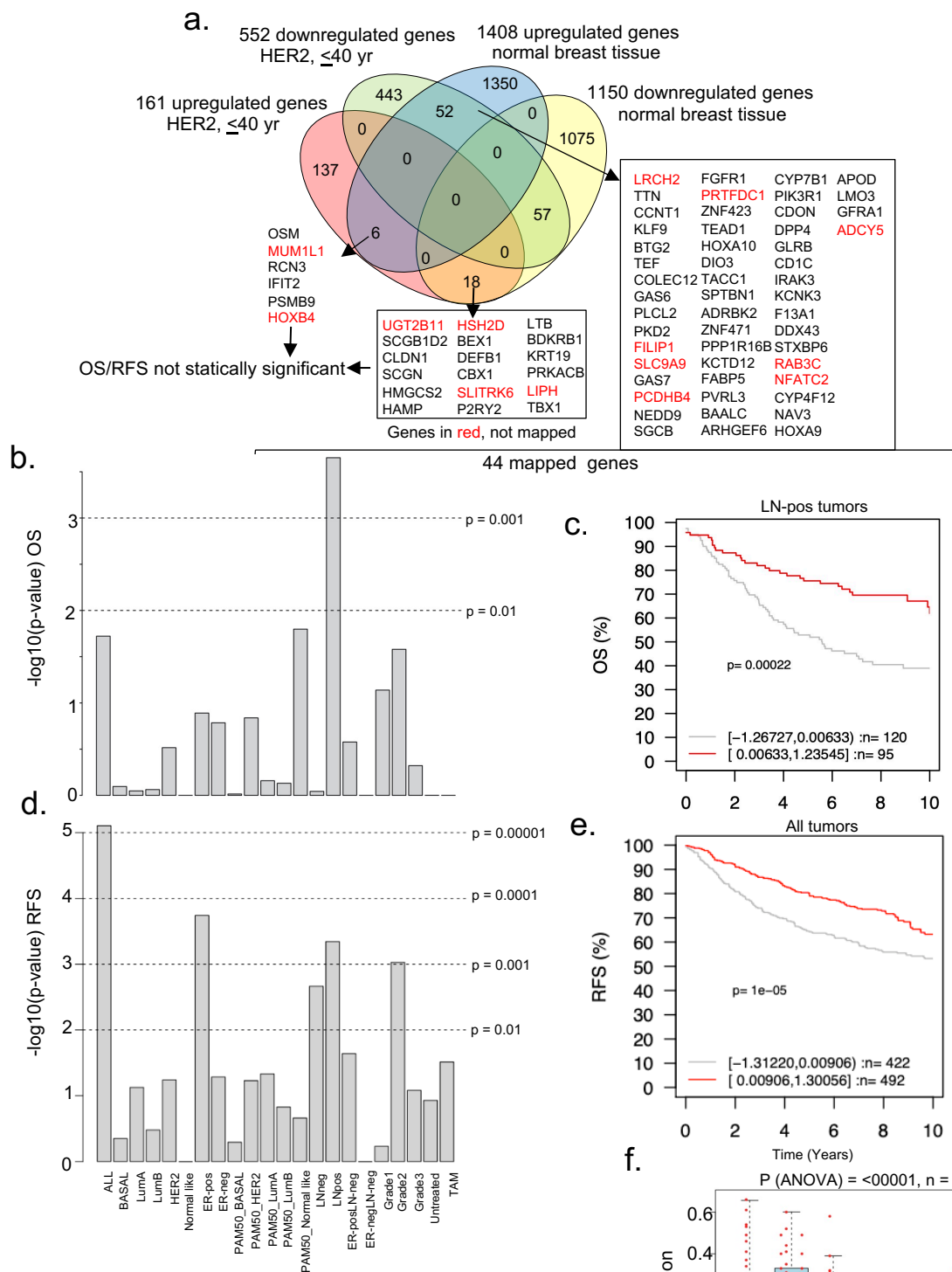
Supplementary Figure S7: Examples of genes upregulated in breast tumors but downregulated in adjacent normal breast tissues. **a**, Kaplan-Meier OS and RFS summaries of 12 out of 20 upregulated genes in Fig. 2b. **b**, OS curves using the high (red) and low (gray) expression of 12 genes for all tumor sub-types, PAM50 normal-like tumors and LN-positive tumors. All survival, correlation analysis performed using the online GOBO tools.



Supplementary Figure S8: Dysregulated genes in BCYW aged ≤ 40 years with DEGs specific to the menstrual cycle phases. **a**, overlap among DRGs in normal breast tissues during the luteal phase of the menstrual cycle and upregulated in breast tumors from women aged ≤ 40 years or shared with women between the ≤ 40 and 41 - 54 years or ≤ 40 and ≥ 55 years. **b**, overlap among DRGs in normal breast tissues during the luteal phase of the menstrual cycle and downregulated u in breast tumors from women aged ≤ 40 years or shared with women between the ≤ 40 and 41 - 54 years or ≤ 40 and ≥ 55 years.

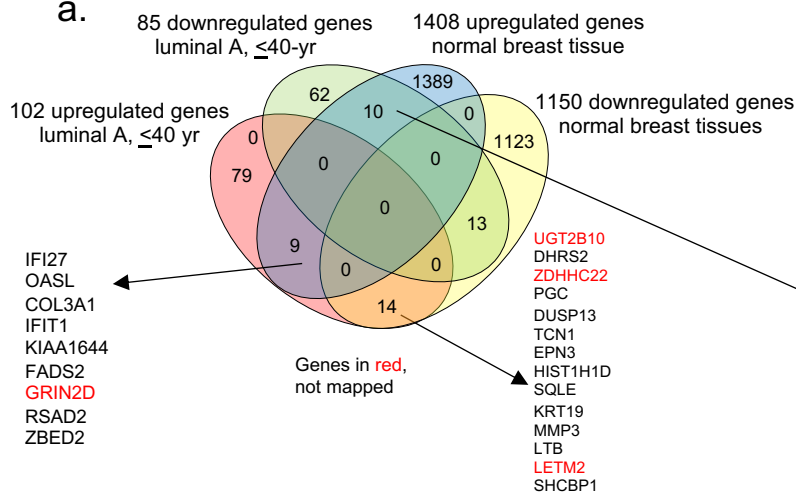


Supplementary Figure S9: Clinical significance of a subset of upregulated genes specific to ≤ 40 years in the METABRIC breast tumor dataset. **a**, Overlap of up- and down-regulated genes in patients aged ≤ 40 years and modulated genes in normal breast tissues; **b**, Overlap of up- and down-regulated genes in patients aged ≤ 40 years and DRGs in normal breast tissues from the luteal or follicular phases of the menstrual cycle; **c**, Kaplan-Meier OS and RFS summaries of 16 genes upregulated in breast tumors < 40 years but downregulated in normal breast tissues; **d**, OS and RFS curves using the high (red) and low (gray) expression of 16 genes for all tumor subtypes. **e**, Spearman correlation analysis of 16 genes in the context of co-expressed 8 gene modules. All survival, correlation analysis performed using the online GOBO tools.

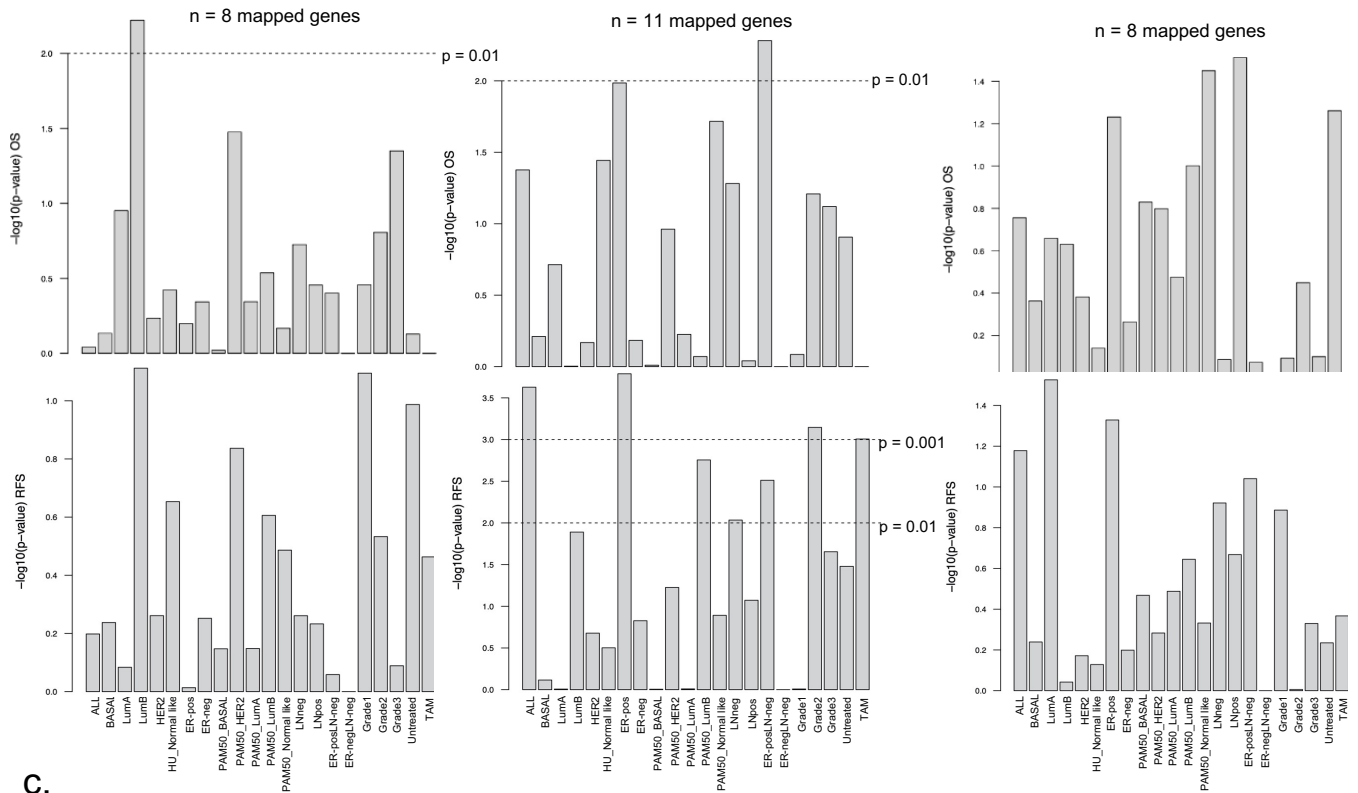


Supplementary Figure S10: Status of BCYW genes in intrinsic HER2 breast tumor subtype. a, Overlap of up- and down-regulated genes in HER2 breast tumors from patients aged ≤ 40 year with upregulated or downregulated genes in normal breast tissues; **b,** Kaplan-Meier OS summary of 44 genes downregulated in HER2 tumors but upregulated in normal breast tissues mapped in GOBO database with 1881 breast tumors; **c,** OS curve using the high (red) and low (gray) expression of 44 genes in LN-pos breast tumors; **d,** Kaplan-Meier RFS summary of 44 genes mapped in GOBO database with 1881 breast tumors; **e,** RFS curve using the high (red) and low (gray) expression of 44 genes in all tumor subtypes; and **f,** Spearman correlation analysis of 44 genes in the context of co-expressed 8 gene modules. All survival, correlation analysis performed using the the GOBO online tools.

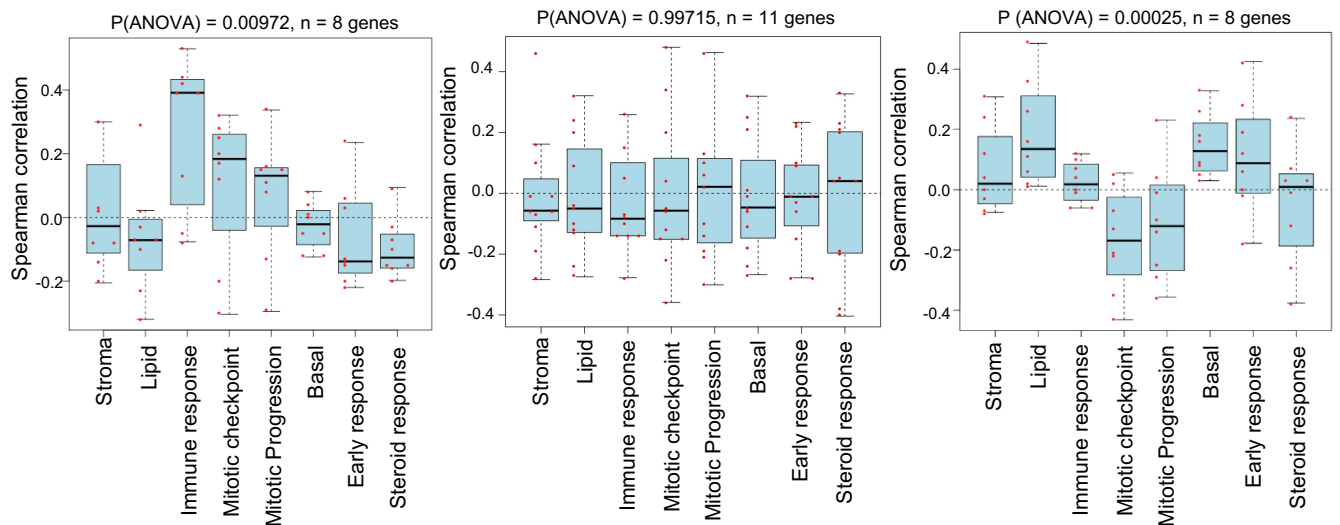
a.



b.

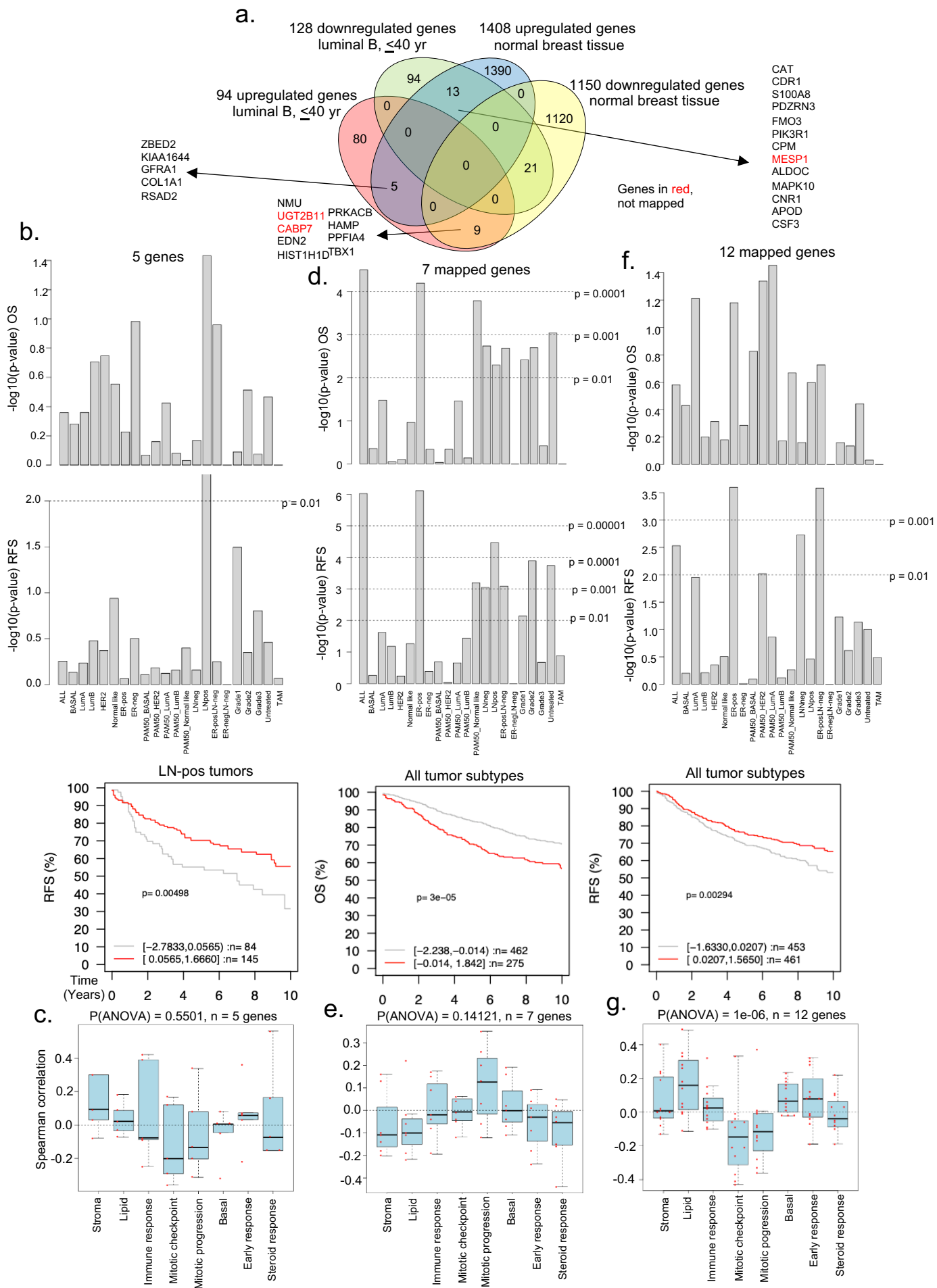


c.



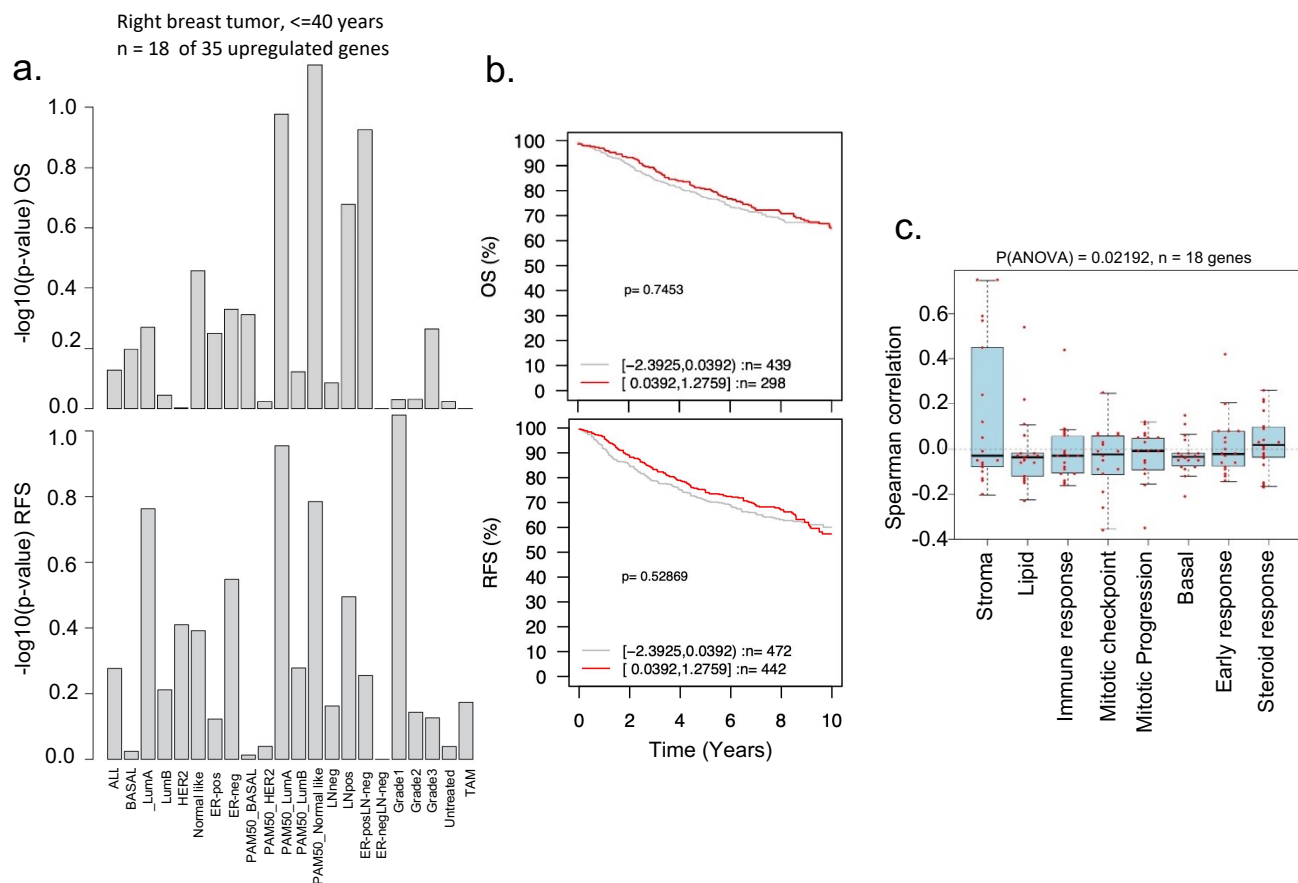
Supplementary Figure S11: Status of BCYW genes in intrinsic luminal A breast tumor subtype.

a, Overlap of up- and down-regulated genes in luminal A breast tumors from patients aged ≤ 40 years with upregulated or downregulated genes in normal breast tissues; **b**, Kaplan-Meier OS and RFS summaries of indicated 8 genes, 11 genes and 8 genes, respectively; **c**, Spearman correlation analysis of indicated 8 genes, 11 genes and 8 genes, respectively. All survival, correlation analysis performed using the online GOBO tools.

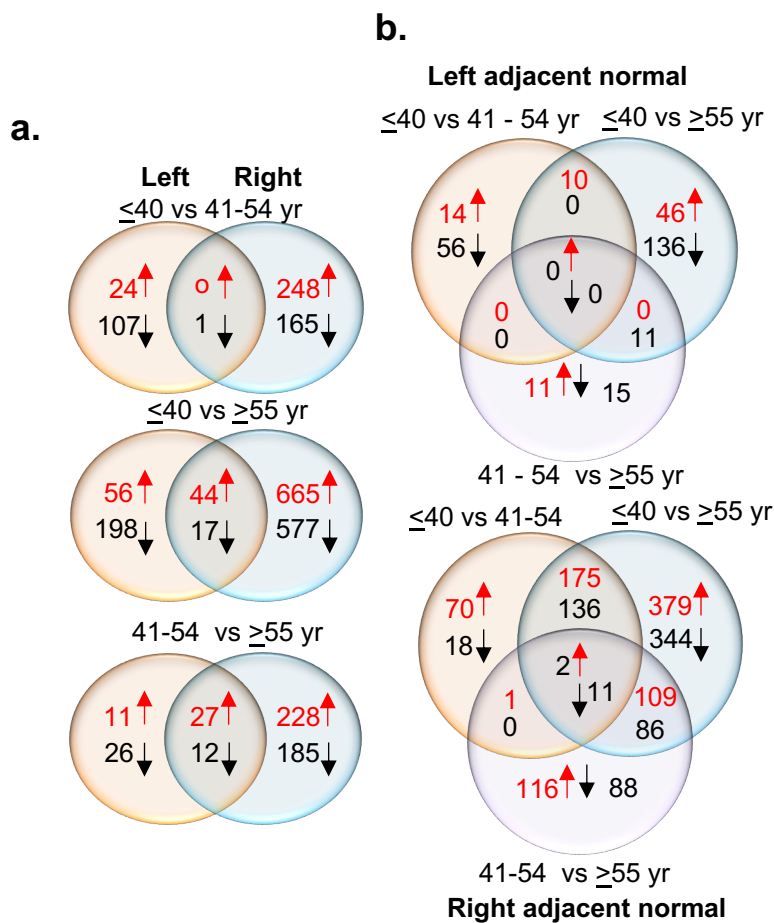


Supplementary Figure S12: Status of BCYW genes in intrinsic luminal B breast tumor subtype.

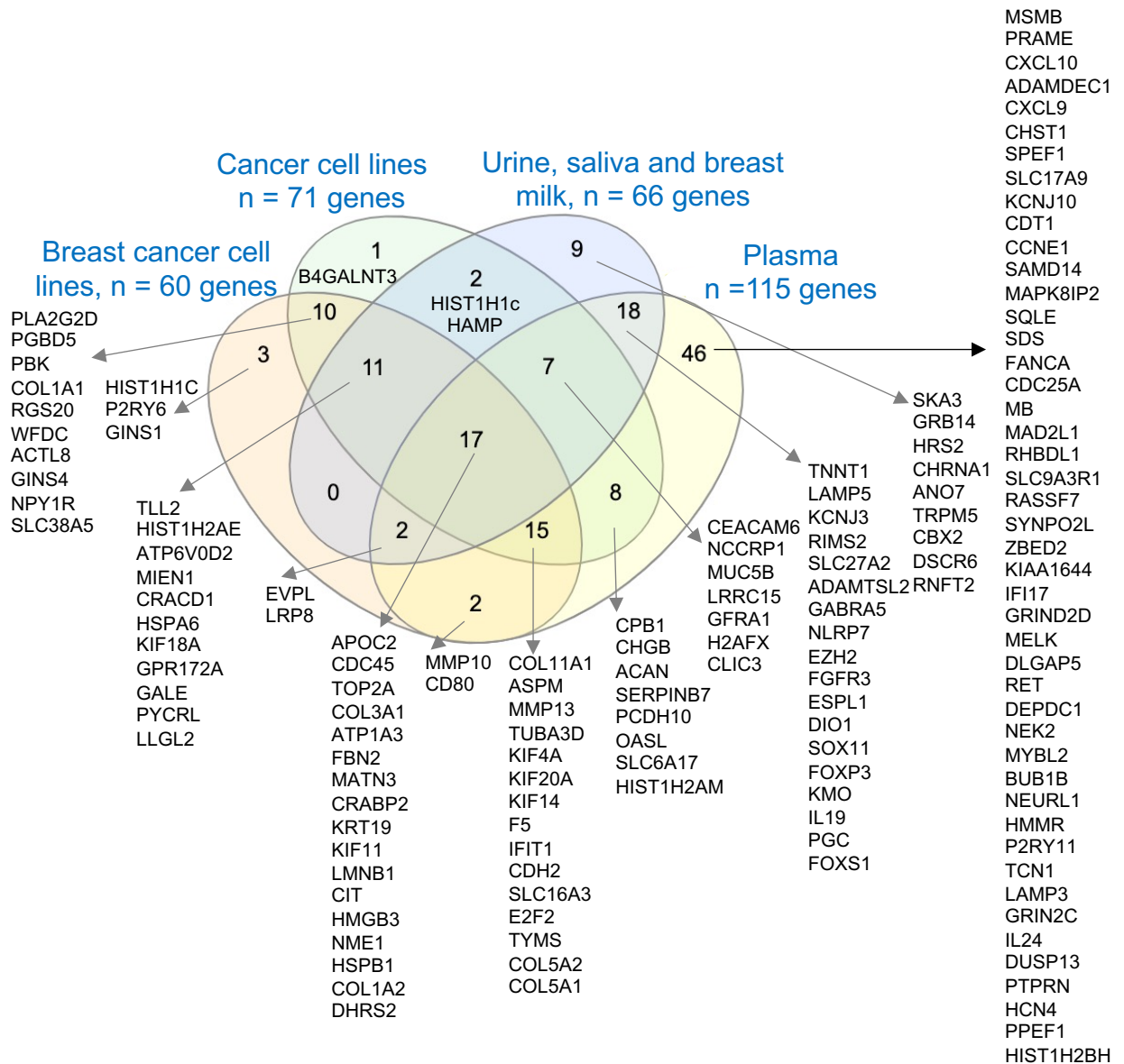
a, Overlap of up- and down-regulated genes in luminal B breast tumors from patients aged ≤ 40 year with upregulated or downregulated genes in normal breast tissues; **b**, Kaplan-Meier OS and RFS summaries of 5 genes, upregulated in luminal B breast tumors as well as in normal breast tissues; RFS curve using the high (red) and low (gray) expression of 5 genes in LN-pos breast tumors; **c**, Spearman correlation analysis of 5 genes in the context of co-expressed 8 gene modules; **d**, Kaplan-Meier OS and RFS summaries of 7 genes, upregulated in luminal B breast tumors but downregulated in normal breast tissues; OS curve using the high (red) and low (gray) expression of 7 genes in all tumor subtypes; **e**, Spearman correlation analysis of 7 genes in the context of co-expressed 8 gene modules; **f**, Kaplan-Meier OS and RFS summaries of 12 genes, downregulated in luminal B breast tumors but upregulated in normal breast tissues; RFS curve using the high (red) and low (gray) expression of 12 genes in all tumor subtypes; and **g**, Spearman correlation analysis of 12 genes in the context of co-expressed 8 gene modules. All survival, correlation analysis performed using the online GOBO tools.



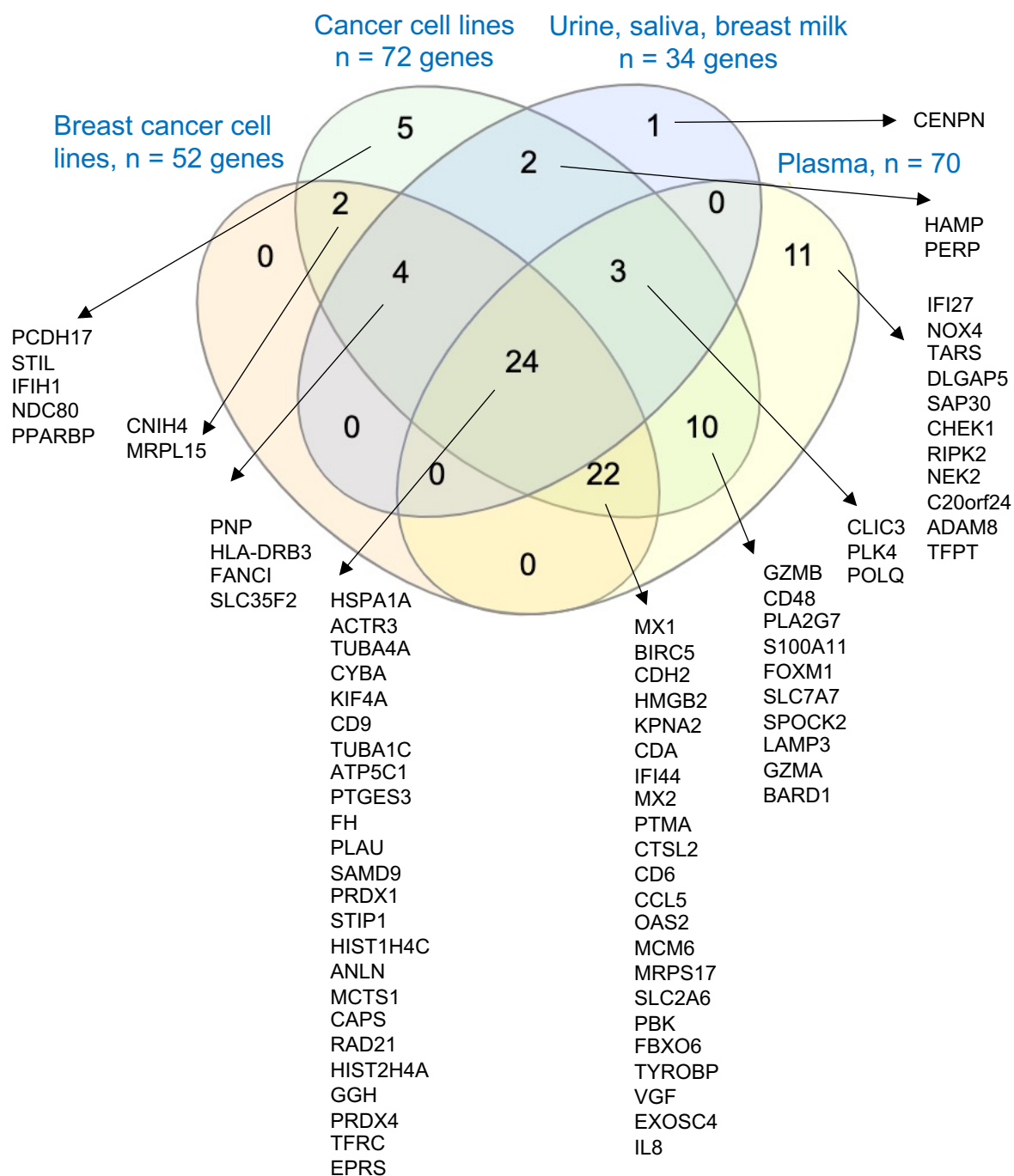
Supplementary Figure S13: Breast cancer genes in right breast tumors aged ≤ 40 years. **a**, Kaplan-Meier OS and RFS summaries of 18 of 35 upregulated genes in the right breast tumors from patients aged ≤ 40 years; **b**, OS and RFS curves using the high (red) and low (gray) expression of 18 genes in all breast tumor subtypes; and **c**, Spearman correlation analysis of 18 genes in the context of co-expressed 8 gene modules.



Supplementary Figure S14: Comparative analysis of dysregulated genes in matching normal specimens. **a**, Status of dysregulated DRG genes in the left and right adjacent normal tissues from three age groups; **b**, Overlap of dysregulated genes among ≤ 40 vs 41 - 54 years, ≤ 40 vs ≥ 55 years and 41 - 54 vs ≥ 55 -years age groups of adjacent matching normal breast tissue samples. Red, upregulated, and blue, downregulated genes.



Supplementary Figure S15: Summary of exemplified comparative analyses to highlight the noticed overlap among secreted proteins by breast cancer cell lines, cancer cell lines, and three human body fluids such as urine, saliva and breast milk and plasma.



Supplementary Fig. S16: Summary and overlap of secreted proteins, identified in the METABRIC dataset for breast tumors aged ≤ 40 years, by breast cancer cell lines, cancer cell lines, human body fluids such as urine, saliva and breast milk, and plasma.

Supplementary Table S1: Summary of clinical features of the breast tumor specimens in the TCGA Firehose legacy breast cancer study.

Age Groups	≤40 yrs.	41-54 yrs.	≥55 yrs.
Sample Numbers	81	255	480
PAM50 Subtype Classifications			
Basal	16	53	70
HER2	3	25	39
LumA	42	108	267
LumB	18	50	120
AJCC Stage			
Stage I	25	13	40
Stage IA	6	20	38
Stage IB	1	2	8
Stage II	20	18	18
Stage IIA	20	71	158
Stage IIB	18	60	79
Stage III	4	12	2
Stage IIIA	17	24	61
Stage IIIB	0	4	19
Stage IIIC	3	9	20
Stage IV	0	5	10
Stage X	2	2	6
Anatomic Neoplasm Subdivision			
Right	9	52	93
Left	17	52	100
Right Upper	22	59	128
Right Lower	4	15	46
Left Upper	16	87	141
Left Lower	3	20	32
Adjacent Matching Normal Samples			
	12	37	59

Supplementary Table S2: Gene Ontology biological function analysis of 45 mapped genes out of 60 upregulated genes in Figure 1a.

Analysis: Biological process									
Dataset name: Group4									
Number of genes in the dataset: 60									
Not recognized for the analysis : 15				IER5L;B4GALNT3;H2AFX;C18orf56;KLHL35;KIAA1644;C4orf48;HAR1A;MROH6;CABP7;HIST1H2BH;HIST1H2AE;HIST1H2AM;C6orf223;C1orf105					
Number of genes in the dataset (which are available in Biological process database) :45				COL5A2;COL5A1;COL1A2;TLL2;LLGL2;SLC6A17;FOXS1;SLC38A5;ANO7;CLIC3;CHRNA1;GABRA5;ADAMTSL2;KMO;HAMP;NLRP7;IL24;P2RY11;GRIN2C;TRPM5;HCN4;APOC2;PTPRN;PGBD5;PLA2G2D;LAMP3;F5;PPEF1;ZNF703;SLC27A2;RHBG;PGC;NPY1R;HSPB1;AMTN;DHRS2;ATP6V0D2;TCN1;CENPW;IL19;NME1;CHST8;MIEN1;DUSP13;C5orf46;					
Number of genes in background: 17857									
Is analysis quantitative: No									
Corrected p-values									
	No. of genes in the dataset	No. of genes in the background dataset	Percentage of genes	Fold enrichment	P-value (Hypergeometric test)	Bonferroni method	Q-value (Storey-Tibshirani method)	genes mapped (from input data set)	
Biological process									
GO:1903225	2	2	4.44	396.73	0.00	0.08	0.07	COL5A2; COL5A1;	
collagen fibril organization	4	56	8.89	28.40	0.00	0.15	0.07	COL1A2; COL5A2; TLL2; COL5A1;	
eye morphogenesis	2	4	4.44	198.86	0.00	0.46	0.15	COL5A2; COL5A1;	
leucine transport	2	5	4.44	159.17	0.00	0.77	0.19	LLGL2; SLC6A17;	
blood vessel development	3	41	6.67	29.12	0.00	1.00	0.37	FOXS1; COL1A2; COL5A1;	
glycine transport	2	9	4.44	88.51	0.00	1.00	0.46	SLC38A5; SLC6A17;	
ion transmembrane transport	4	142	8.89	11.20	0.00	1.00	0.78	ANO7; CLIC3; CHRNA1; GABRA5;	
extracellular matrix organization	4	156	8.89	10.20	0.00	1.00	0.96	COL1A2; COL5A2; ADAMTSL2; COL5A1;	
cellular response to lipopolysaccharide	4	160	8.89	9.94	0.00	1.00	0.96	KMO; HAMP; NLRP7; IL24;	
calcium-mediated signaling	3	81	6.67	14.74	0.00	1.00	1.00	P2RY11; GRIN2C; TRPM5;	

Supplementary Table S3: Gene Ontology biological function analysis of 14 mapped genes out of highly upregulated 17 genes in Figure 1e.

Analysis: Biological process									
Dataset name: Group4									
Number of genes in the dataset: 17									
Not recognized for the analysis : 4KLHL35;HAR1A;MROH6;C1ORF105									
Number of genes in the dataset (which are available in Biological process database) :13PGBD5;LAMP3;KMO;ZNF703;F5;RHBG;LLGL2;GRIN2C;IL24;ATP6V0D2;IL19;MIEN1;NME1;									
Number of genes in background: 17857									
Is analysis quantitative: No									
Corrected p-values									
	No. of genes in the dataset	No. of genes in the background dataset	Percentage of genes	Fold enrichment	P-value (Hypergeometric test)	Bonferroni method	Q-value (Storey-BH method)	Tibshirani method	genes mapped (from input data set)
Biological process									
GO:0098038	1	1	7.69	1372.56	0.00	1	1	1	PGBD5;
GO:1903900	1	1	7.69	1372.56	0.00	1	1	1	LAMP3;
kynurenic acid biosynthetic process	1	2	7.69	689.69	0.00	1	1	1	KMO;
GO:1903296	1	2	7.69	689.69	0.00	1	1	1	KMO;
negative regulation of homotypic cell-cell adhesion	1	2	7.69	689.69	0.00	1	1	1	ZNF703;
quinolinate biosynthetic process	1	3	7.69	460.56	0.00	1	1	1	KMO;
anthranilate metabolic process	1	3	7.69	460.56	0.00	1	1	1	KMO;
L-kynurenine metabolic process	1	3	7.69	460.56	0.00	1	1	1	KMO;
response to vitamin K	1	3	7.69	460.56	0.00	1	1	1	F5;
transepithelial ammonium transport	1	3	7.69	460.56	0.00	1	1	1	RHBG;

Supplementary Table S4: Gene Ontology biological function analysis of 19 mapped genes out of 22 upregulated genes in Figures 2b and 2c.

Analysis: Biological process									
Dataset name: Group1									
Number of genes in the dataset: 22									
Not recognized for the analysis : 3									
NXPH1;MUC5B;C6orf223									
FIBCD1;ACAN;PCDH10;CST9;SERPINB7;CPB1;HAMP;FOXS1;RIMS2;P2RY11;NCCRP1;DHRS2;CHGB;NPY1R;SLC6A17;LRRC15;TCN1;SYNPO2L;HS6ST3									
Number of genes in the dataset (which are available in Biological process database) :19									
Number of genes in background: 17857									
Is analysis quantitative: No									
Corrected p-values									
Biological process	No. of genes in the dataset	No. of genes in the background dataset	Percent age of genes	Fold enrichment	P-value (Hypergeometric test)	Bonferroni method	BH method	Q-value (Storey-Tibshiran method)	genes mapped (from input data set)
response to erythropoietin	1	1	5.26	939.35	0.001064	1.00	1.00	1.00	HAMP;
GO:1904479	1	1	5.26	939.35	0.001064	1.00	1.00	1.00	HAMP;
positive regulation of platelet-derived growth factor production	1	2	5.26	472.01	0.002127	1.00	1.00	1.00	SERPINB7 ;
negative regulation of iron ion transmembrane transport	1	2	5.26	472.01	0.002127	1.00	1.00	1.00	HAMP;
GO:1990641	1	3	5.26	315.20	0.003189	1.00	1.00	1.00	HAMP;
positive regulation of glomerular mesangial cell proliferation	1	5	5.26	189.37	0.005309	1.00	1.00	1.00	SERPINB7 ;
positive regulation of cell growth involved in cardiac muscle cell development	1	5	5.26	189.37	0.005309	1.00	1.00	1.00	HAMP;
leucine transport	1	5	5.26	189.37	0.005309	1.00	1.00	1.00	SLC6A17;
GO:1903413	1	5	5.26	189.37	0.005309	1.00	1.00	1.00	HAMP;
alanine transport	1	5	5.26	189.37	0.005309	1.00	1.00	1.00	SLC6A17;

Supplementary Table S5: Gene Ontology biological function analysis of 8 mapped genes out of 12 downregulated genes in Figure 2g.

Analysis: Biological process									
Dataset name: Group3									
Number of genes in the dataset: 12									
Not recognized for the analysis : 4					VAT1L;COLCA1;SPHKAP;EMILIN3				
Number of genes in the dataset (which are available in Biological process database) :8					TNFSF11;CSRNP3;PLPPR1;RSPO1;SLC26A3;SLCO1A2;SYCE1;IL22RA2				
Number of genes in background: 17857									
Is analysis quantitative: No									
Corrected p-values									
	No. of genes in the dataset	No. of genes in the background dataset	Percentage of genes	Fold enrichment	P-value (Hypergeometric test)	Bonferroni method	BH method	Q-value (Storey-Tibshirani method)	genes mapped (from input data set)
Biological process									
positive regulation of corticotropin-releasing hormone secretion	1	1	12.5	2229.34	0.000448	1	1	1	TNFSF11;
GO:1904616	1	1	12.5	2229.34	0.000448	1	1	1	TNFSF11;
positive regulation of fever generation by positive regulation of prostaglandin secretion	1	2	12.5	1120.215	0.000896	1	1	1	TNFSF11;
positive regulation of ERK1 and ERK2 cascade via TNFSF11-mediated signaling	1	2	12.5	1120.215	0.000896	1	1	1	TNFSF11;
cytokine-mediated signaling pathway	2	145	25	30.90113	0.001776	1	1	1	TNFSF11; IL22RA2;
osteoclast proliferation	1	4	12.5	561.5045	0.001791	1	1	1	TNFSF11;
positive regulation of osteoclast development	1	4	12.5	561.5045	0.001791	1	1	1	TNFSF11;
intracellular pH elevation	1	4	12.5	561.5045	0.001791	1	1	1	SLC26A3;
positive regulation of homotypic cell-cell adhesion	1	5	12.5	449.4277	0.002238	1	1	1	TNFSF11;
TNFSF11-mediated signaling pathway	1	5	12.5	449.4277	0.002238	1	1	1	TNFSF11;

Supplementary Table S6: Gene Ontology biological function analysis of 26 mapped genes out of 27 upregulated genes in Figure 3b.

Analysis: Biological process									
Dataset name: Group2									
Number of genes in the dataset: 27									
Not recognized for the analysis: 1 HMMR									
Number of genes in the dataset SGO1;DLGAP5;NEK2;TOP2A;SKA3;TUBA3D;PBK;MYBL2;ESPL1;KIF4A;KIF20A;BUB1B;KIF (which are available in Biological 14;MELK;RET;TNNT1;ASPM;CDC45;KCNJ3;COL11A1;CRACD;NEURL1;MMP13;LAMP5;S process database) :26 YT13;DEPDC1									
Number of genes in background: 17857									
Is analysis quantitative: No									
Corrected p-values									
	No. of genes in the dataset	No. of genes in the background dataset	Perc enta ge of gene s	Fold enrich ment	P-value (Hyper geome tric test)	Bonfer roni metho d	BH metho d	Q-value (Storey-Tibshirani method)	genes mapped (from input data set)
Biological process									
chromosome segregation	5	73	19.2 3	47.11	0.00	0.00	0.00	0.01	SGO1; DLGAP5; NEK2; TOP2A; SKA3;
mitotic cell cycle	5	132	19.2 3	26.06	0.00	0.01	0.01	0.05	TUBA3D; PBK; NEK2; MYBL2; SKA3;
mitotic sister chromatid segregation	3	33	11.5 4	62.60	0.00	0.18	0.06	0.41	SGO1; NEK2; ESPL1;
cytokinesis after mitosis	3	50	11.5 4	41.32	0.00	0.64	0.14	0.96	KIF4A; KIF20A; ESPL1;
homologous chromosome segregation	2	8	7.69	172.28	0.00	0.70	0.14	0.96	SGO1; ESPL1;
cell division	5	348	19.2 3	9.88	0.00	1.00	0.24	1.00	SGO1; NEK2; BUB1B; KIF14; SKA3;
positive regulation of mitotic metaphase/anaphase transition	2	12	7.69	114.90	0.00	1.00	0.24	1.00	DLGAP5; ESPL1;
peptidyl-tyrosine phosphorylation	3	130	11.5 4	15.89	0.00	1.00	1.00	1.00	MELK; RET; PBK;
GO:1905463	1	1	3.85	686.54	0.00	1.00	1.00	1.00	TOP2A;
negative regulation of muscle contraction	1	1	3.85	686.54	0.00	1.00	1.00	1.00	TNNT1;

Supplementary Table S7: Gene Ontology biological function analysis of 95 mapped genes out of 124 upregulated genes in Figure 7b (left breast cancer).

Analysis: Biological process

Dataset name: Left<40 124Genes

Not recognized for the analysis : 124

Number of genes in the dataset: 29 TRY6;LOC400696;C2orf54;CNTD2;LINGO1;FAM64A;RDM1;C19orf33;RLTPR;DSCR6;NCRNA00176;C15orf42;MLF1IP;HIST1H2BE;C9orf140;NFKBIL2;ORC1L;KIAA1524;KIAA1211;C16orf75;PRR19;C3orf67;C20orf151;C18orf56;FCGR1C;GPR172A;PYCRL;C14orf80;RAG1AP1

Number of genes in the dataset (which are available in Biological process database) :95	IL4I1;FCGR1A;SRMS;IRF7;HMGB3;ADAMDEC1;HAMP;CD80;ESPL1;SAMD14;EZH2;CXCL10;LLGL2;RCC1;TMEM145;CXCL9;RAMP1;ADORA2A;P2RY6;HSPB1;TLX1;ZNF695;E2F2;FOXP3;FOXS1;SIX4;EVPL;PRAME;CEACAM6;SOX11;FGFR3;TNFRSF18;NUAK2;MAD2L1;SAMD11;CXCL11;MMP10;FAM111B;LRP8;RHBDL1;CDH2;CCNE1;MAPK8IP2;CRABP2;SLC9A3R1;TYMS;HSH2D;METRN;RASSF7;CAPS;CIT;JAKMIP1;KIF18A;POC1A;RNFT2;FBXO43;CBX2;CDC25A;MFSD2B;NME1;VSTM2L;MSMB;CHST1;SDS;LGALS9C;MND1;CDCA2;CDT1;KIF11;MATN3;MB;RAD51AP1;PIF1;SBK1;SPEF1;APOC2;SQLE;FANCA;PRR11;KCNJ10;PPM1J;ATP1A3;SLC17A9;KRT19;FBN2;ACTL8;IL21R;DIO1;LMNB1;GINS1;GINS4;IL17REL;SLC16A3;AIFM3;GALE
---	---

Number of genes in background: 17857

Is analysis quantitative: No

Corrected p-values

Biological process	No. of genes in the dataset	No. of genes in the background dataset	Percent age of genes	Fold enrichment	P-value (Hypergeometric test)	Selected p-values			
						Bonferroni method	BH method	Q-value (Storey-Tibshirani method)	genes mapped (from input data set)
regulation of mitosis	3	23	3.16	24.59	0.000239	1.00	1.00	1.00	CDCA2; FBXO43; RCC1;
GO:1903934	2	5	2.11	75.40	0.000277	1.00	1.00	1.00	GIN51; GINS4;
adenylate cyclase-activating G-protein coupled receptor signaling pathway	5	121	5.26	7.78	0.000468	1.00	1.00	1.00	CXCL11; CXCL10; CXCL9; RAMP1; ADORA2A;
positive regulation of release of sequestered calcium ion into cytosol	3	29	3.16	19.50	0.000482	1.00	1.00	1.00	CXCL11; CXCL10; CXCL9;
GO:0097421	3	29	3.16	19.50	0.000482	1.00	1.00	1.00	HAMP; EZH2; TYMS;
cell division	8	348	8.42	4.33	0.000535	1.00	1.00	1.00	CDCA2; CDT1; CCNE1; KIF11; CDC25A; MAD2L1; LLGL2; RCC1;
positive regulation of phospholipase activity	2	7	2.11	53.89	0.000578	1.00	1.00	1.00	FGFR3; APOC2;
mitotic sister chromatid segregation	3	33	3.16	17.14	0.000709	1.00	1.00	1.00	ESPL1; KIF18A; MAD2L1;
neuroepithelial cell differentiation	2	8	2.11	47.16	0.000768	1.00	1.00	1.00	SOX11; CDH2;
negative regulation of retinoic acid receptor signaling pathway	2	9	2.11	41.93	0.000984	1.00	1.00	1.00	PRAME; EZH2;

Supplementary Table S8: Gene Ontology biological function analysis of 11 highly upregulated genes in Figure 7f.

Analysis: Biological process									
Number of genes in the dataset: 11									
Number of genes in the dataset (which are available in Biological process database) :11									
SLC50A1;KCNJ10;FCGR1A;CRABP2;SLC52A2;PYCR3;FBXO43;SQLE;TONSL;NUAK2;ZNF695									
Number of genes in background: 17857									
Is analysis quantitative: No									
Corrected p-values									
	No. of genes in the dataset	No. of genes in the background dataset	Percentage of genes	Fold enrichment	P-value (Hypergeometric test)	Bonferroni method	BH method	Q-value (Storey-Tibshirani method)	genes mapped (from input dataset)
Biological process									
cellular aromatic compound metabolic process	1	1	9.09	1621.89	0.000616	1.00	1.00	1.00	SQLE;
glucoside transport	1	1	9.09	1621.89	0.000616	1.00	1.00	1.00	SLC50A1;
glutamate uptake involved in synaptic transmission	1	2	9.09	814.98	0.001232	1.00	1.00	1.00	KCNJ10;
antibody-dependent cellular cytotoxicity	1	3	9.09	544.22	0.001847	1.00	1.00	1.00	FCGR1A;
retinoic acid biosynthetic process	1	5	9.09	326.97	0.003077	1.00	1.00	1.00	CRABP2;
riboflavin metabolic process	1	5	9.09	326.97	0.003077	1.00	1.00	1.00	SLC52A2;
L-proline biosynthetic process	1	6	9.09	272.56	0.003691	1.00	1.00	1.00	PYCR3;
negative regulation of meiosis	1	6	9.09	272.56	0.003691	1.00	1.00	1.00	FBXO43;
riboflavin transport	1	6	9.09	272.56	0.003691	1.00	1.00	1.00	SLC52A2;
positive regulation of collateral sprouting	1	8	9.09	204.51	0.004918	1.00	1.00	1.00	CRABP2;

Supplementary Table S9: Examples of secreted gene products relevant to the biology of BCYW, aged ≤ 40 years.

A. Upregulated genes in BC ≤ 40 years, Fig. 1a

Breast cancer cells	LLGL2, F5, NME1, COL1A2, HSPB1, COL5A2, PGBD5, TLL2, COL5A1, NPY1R, PLA2G2D, HIST1H2AE, DHRS2, SLC38A5, ATP6V0D2, APOC2
Cancer cell lines	LLGL2, F5, NME1, COL1A2, HSPB1, COL5A2, PGBD5, TLL2, COL5A1, NPY1R, PLA2G2D, HIST1H2AE, DHRS2, H2AFX, HAMP, CLIC3, HIST1H2AM, SLC38A5, ATP6V0D2, APOC2, B4GALNT3, SLC6A17
Urine	LLGL2, NME1, COL1A2, HSPB1, H2AFX, KMO, SLC27A2, ADAMTSL2, HAMP, CLIC3, CHRNA1, HIST1H2AE, PGC, DHRS2, GABRA5, ANO7, FOXS1, ATP6V0D2, NLRP7
Saliva	NME1, HSPB1, HIST1H2AE
Breast milk	LLGL2, NME1, HSPB1, TLL2, HAMP, IL19
Proteome	LAMP3, F5, NME1, COL1A2, HSPB1, H2AFX, COL5A2, KMO, SLC27A2, ADAMTSL2, P2RY11, COL5A1, PTPRN, CLIC3, GRIN2C, IL24, HCN4, PPEF1, TCN1, DUSP13, PGC, HIST1H2AM, DHRS2, GABRA5, FOXS1, SLC6A17, APOC2, NLRP7, HIST1H2BH

B. Subset of highly upregulated genes in BC ≤ 40 years, Fig. 1e

Breast cancer cells	MIEN1, LLGL2, F5, NME1, PGBD5, ATP6V0D2
Cancer cell lines	MIEN1, LLGL2, F5, NME1, PGBD5, ATP6V0D2
Urine	MIEN1, LLGL2, NME1, KMO, ATP6V0D2
Saliva	NME1
Breast milk	IL19, MIEN1, NME1, LLGL2
Proteome	LAMP3, F5, NME1, KMO, GRIN2C, IL24, IL19

C. Upregulated genes in BC but downregulated in adjacent normal, Fig. 2b,c

Breast cancer cells	DHRS2, NCCRP1, CHGB, NPY1R, ACAN, MUC5B, LRRC15
Cancer cell lines	DHRS2, SLC6A17, CPB1, NCCRP1, CHGB, NPY1R, ACAN, SERPINB7, HAMP, MUC5B, LRRC15, PCDH10
Urine	HRS2, NCCRP1, RIMS2, HAMP, FOXS1, MUC5B, LRRC15
Saliva	NCCRP1, MUC5B
Breast milk	NCCRP1, MUC5B, HAMP
Proteome	DHRS2, TCN1, SLC6A17, CPB1, NCCRP1, RIMS2, CHGB, SYNPO2L, ACAN, SERPINB7, FOXS1, MUC5B, LRRC15, P2RY11, PCDH10

D. Upregulated genes in BC 3-age groups and adjacent normal DRG ≤ 40 vs >55 years, Fig. 3a

Breast cancer cells	COL11A1, ASPM, MMP13, CDC45, TUBA3D, PBK, CRACD1, KIF4A, KIF20A, TOP2A, KIF14
Cancer cell lines	COL11A1, ASPM, MMP13, CDC45, TUBA3D, PBK, CRACD1, KIF4A, KIF20A, TOP2A, KIF14, CDC45, TUBA3D
Urine	TNNT1, LAMP5, CDC45, CRACD1, TOP2A, KCNJ3, SKA3
Proteome	TNNT1, COL11A1, LAMP5, ASPM, MMP13, MELK, DLGAP5, RET, CDC45, TUBA3D, DEPDC1, KIF4A, NEK2, MYBL2, KIF20A, BUB1B, NEURL1, TOP2A, HMMR, KCNJ3, KIF14, ESPL1

E. Upregulated genes in BC ≤ 40 year and normal breast tissues, Fig. 4a

Breast cancer cells	COL1A2, COL5A2, COL5A1, NPY1R
Cancer cell lines	COL1A2, COL5A2, COL5A1, NPY1R
Urine	COL1A2
Proteome	COL1A2, COL5A2, COL5A1

F. Upregulated genes in BC ≤ 40 year but downregulated in normal breast tissues, Fig. 4b

Breast cancer cells	F5, SLC38A5, DHRS2
Cancer cell lines	F5, SLC38A5, DHRS2, HAMP
Urine	HAMP, PGC, DHRS2
Breast milk	HAMP
Proteome	LAMP3, F5, TCN1, DUSP13, PGC, DHRS2

G. Upregulated genes in basal BC ≤40 years and in normal breast tissues, Fig. 6c		
Breast cancer cells	HSPA6, RGS20, WFDC, HIST1H1c	
Cancer cell lines	HSPA6, RGS20, WFDC, HIST1H1C, GRB14	
Urine	HSPA6, GRB14, HIST1H1C	
Saliva	HSPA6	
Breast milk	HIST1H1c	
H. Upregulated genes in LumA and LumB BC ≤40 years and normal breast tissues, Supplementary Fig. 10 and 11		
Breast cancer cells	COL1A1, COL3A1, IFIT1	
Cancer cell lines	GFRA1, COL1A1, OASL, COL3A1, IFIT1	
Urine	GFRA1, COL1A1,	
Saliva	COL3A1	
Proteome	ZBED2, KIAA1644, GFRA1, IFI17, OASL, COL3A1, IFIT1GRIND2D, ZBED2	
I. Downregulated in HER2 BC ≤40 years and upregulated in normal breast tissues, Supplementary Fig. 9a		
Breast cancer cells	TTN, GAS6, GAS7, SGCB, SPTBN1, KCTD12, FABP5, PVRL3, ARHGEF6, PIK3R1, DPP4, CD1C, F13A1, STXBP6, APOD	
Cancer cell lines	TTN, COLEC12, GAS6, PKD2, GAS7, NEDD9, SGCB, FGFR1, HOXA10, TACC1, SPTBN1, PPP1R16B, KCTD12, FABP5, PVRL3, ARHGEF6, PIK3R1, DPP4, CD1C, KCNK3, F13A1, STXBP6, APOD, GFRA1	
Urine	TTN, COLEC12, GAS6, PKD2, GAS7, NEDD9, SGCB, FGFR1, HOXA10, TACC1, SPTBN1, PPP1R16B, KCTD12, FABP5, PVRL3, ARHGEF6, PIK3R1, DPP4, CD1C, KCNK3, F13A1, STXBP6, APOD, GFRA1	
Saliva	FGFR1, DPP4	
Breast milk	FGFR1, SPTBN1, FABP5, APOD	
Proteome	TTN, CCNT1, KLF9, BTG2, TEF, COLEC12, GAS6, PLCL2, PKD2, GAS7, NEDD9, SGCB, FGFR1, ZNF423, TEAD1, HOXA10, DIO3, TACC1, SPTBN1, ADRBK2, PPP1R16B, KCTD12, FABP5, PVRL3, ARHGEF6, CYP7B1, PIK3R1, DPP4, GLRB, IRAK3, KCNK3, F13A1, CYP4F12, NAV3, HOXA9, APOD, GFRA1	
J. Upregulated genes in left BC ≤40 years, Fig. 7b left		
Breast cancer cells	ACTL8, CEACAM6, ATP1A3, FBN2, MATN3, CDH2, CRABP2, SLC16A3, E2F2, KRT19, KIF18A, KIF11, LMNB1, HSPB1, CIT, GINS4, LLGL2, HMGB3, GPR172A, GALE, PYCRL, TYMS, NME1	
Cancer cell lines	ACTL8, MMP10, ATP1A3, FBN2, MATN3, CDH2, CRABP2, CD80, SLC16A3, E2F2, KRT19, KIF18A, KIF11, LMNB1, P2RY6, EVPL, HSPB1, GINS1, LRP8, CIT, GINS4, LLGL2, HMGB3, GPR172A, GALE, PYCRL, TYMS, NME1	
Urine	CBX2, FGFR3, ESPL1, DSCR6, CRABP2, KRT19, KIF18A, KIF11, RNFT2, LMNB1, EVPL, HSPB1, EZH2, CIT, LLGL2, HMGB3, GALE, NME1, CEACAM6, DIO1, SOX11, ATP1A3, FBN2, MATN3	
Saliva	CEACAM6, KRT19, EVPL, NME1	
Breast milk	MATN3, CRABP2, KRT19, EVPL, HSPB1, LLGL2, GPR172A, GALE, PYCRL, NME1	
Proteome	TTN, CCNT1, KLF9, BTG2, TEF, COLEC12, GAS6, PLCL2, PKD2, GAS7, NEDD9, SGCB, FGFR1, ZNF423, TEAD1, HOXA10, DIO3, TACC1, SPTBN1, ADRBK2, PPP1R16B, KCTD12, FABP5, PVRL3, ARHGEF6, CYP7B1, PIK3R1, DPP4, GLRB, IRAK3, KCNK3, F13A1, CYP4F12, NAV3, HOXA9, APOD, GFRA1	

Supplementary Table S10: Summary of key observations in the present study. Empty blocks in OS and RFS, no significant p-value; ND, not determined; NS, not significant; pos, positive correlation; Neg, negative correlation.

	Upregulated genes BC≤40 yr	Highly Upregulated ≤40 yr	Up BC ≤40 years Down in adjacent normal	Up BC Up adjacent normal ≤40 vs ≥55 yr	Up BC ≤40 yr common Up normal breast	Up BC ≤40 yr common Down normal breast	Down BC <40 yr common Up normal breast	Up basal ≤40 yr common Up normal breast	Down basal ≤40 yr common Up normal breast	Up HER2 ≤40 yr common Up normal breast	Down HER2 ≤40 yr common Up normal breast	Up LumA ≤40 yr common Up normal breast	Down LumA ≤40 yr common Up normal breast	Up LumB ≤40 yr common Up normal breast	Down LumB ≤40 yr common Up normal breast	Up ≤40 yr Left Tumors
GENESET	60	60	22	27	60	60	58	132	194	161	552	102	85	94	94	124
Subset gene number		17			5	28	10	11	22	6	52	9	10	5	9	13
Genes mapped	38	11	12	19	5	17	9	10	21	4	44	8	8	5	7	76
OVERALL SURVIVAL																
p-value significant	Yes	No	Yes	Yes	Yes	Yes	No	Yes	Yes	No	Yes	No	No	No	Yes	Yes
All Tumors	0.82859	No	0.56243	<0.00001	0.00072	0.02739		0.14248	0.00116		0.01896				3.00E-05	1.00E-05
Basal								0.0512	0.47058							
Lum A				0.00141				0.00624								
Lum B																
ER+				<0.00001	0.0056	0.00497		0.00431	0.00424						6.00E-05	<0.00001
LN-				<0.00001											0.00183	0.00029
LN+			0.00169	4.00E-05					0.00056		2.20E-04				0.00508	
ER-posLN-neg	0.00204		0.00783	<0.00001		0.00825		0.00436							0.00207	4.00E-05
Pam50 normal-like			0.00664			0.00335									0.00016	
Grade 1			0.00754												0.00384	
Grade 2				<0.00001	0.00497										0.00202	9.00E-05
Grade 3																
RELAPSE FREE SURVIVAL																
p-value significant	No	Yes	Yes	Yes	Yes	Yes	Yes	No	Yes	No	Yes	No	No	Yes	Yes	Yes
All tumors	No	0.06891	0.03584	<0.00001	<0.00001	0.00323	0.00013	0.36986	<0.00001		1.00E-05			No	<0.00001	0.00294
Basal		0.0095				0.01217			0.25768							
Lum A				0.0031	0.00031											
Lum B				0.00131		0.00583										0.00698
ER+		0.00828		<0.00001	6.00E-05	4.00E-05	0.00024		1.00E-05		0.00029				<0.00001	<0.00001
LN-				<0.00001	0.00094		0.00137		<0.00001		0.00196				9.00E-04	0.00188
LN+			0.00775	1.00E-05	0.00015				<0.00001		5.00E-04			0.00498	3.00E-05	0.01095
ER-posLN-neg				<0.00001	0.00488	0.00452	0.00342		1.00E-05						0.00081	0.00026
Pam50 Normal-like															0.00064	
Grade 1				0.00099	0.00957										0.00718	
Grade 2		0.00147		<0.00001	0.00178		0.00689				0.00023				0.00013	0.00111
Grade 3									0.00321							
CORRELATION with 8 GENE MODULES																
P(ANOVA)	0.16715	0.10505	1.00E-05	<0.00001	<0.00001	0.10291	<0.00001	0.05231	<0.00001	No	<0.00001	0.00972	0.00025	0.5501	No	<0.00001
Stroma	NS	NS	pos	neg	pos	NS	pos	neg	pos	NS	pos	neg	NS	pos	NS	neg
Lipid	NS	NS	pos	neg	pos	NS	pos	neg	pos	NS	pos	neg	pos	NS	NS	neg
Immune	NS	NS	neg	pos	neg	NS	neg	pos	neg	NS	NS	pos	NS	neg	NS	pos
Mitotic	NS	NS	neg	pos	neg	NS	neg	NS	neg	NS	neg	pos	neg	neg	NS	pos
Basal	NS	NS	neg	neg	NS	NS	pos	pos	pos	NS	pos	neg	pos	NS	NS	neg
Early	NS	NS	pos	neg	pos	NS	pos	neg	pos	NS	pos	neg	pos	NS	NS	neg
Steroid	NS	NS	pos	neg	pos	NS	pos	neg	pos	NS	pos	neg	pos	NS	NS	neg
SECRETION	SECRETION	SECRETION	SECRETION	SECRETION	SECRETION	SECRETION	ND	SECRETION	ND	ND	ND	SECRETION	ND	ND	ND	SECRETION
Breast cancer cells	16	6	7	11	4	3		4				3				23
Cancer cells	22	6	12	13	4	4		5				5				28
Urine	19	5	7	7	1	3		3				2				24
Saliva	3	1	2	none	none	none		1				1				4
Breast milk	6	4	3	none	none	1		1				none				10
Plasma	29	7	15	22	3	6		none				8				47