Supporting Information

Maruyama et al. 10.1073/pnas.1010559107

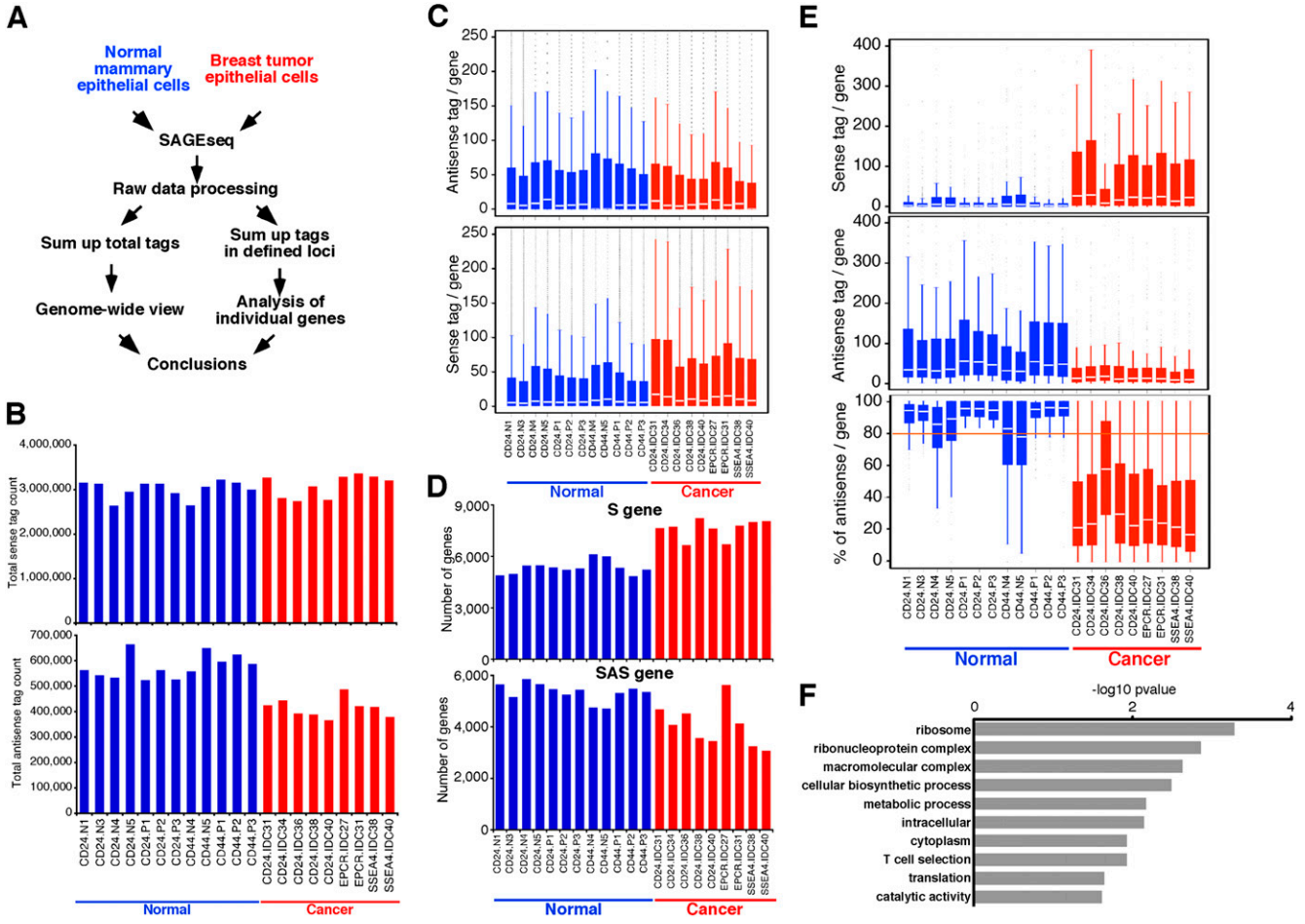

Fig. S1. Overall view of experimental outline, sense and antisense tag count distributions, and identity and function of 252 AS genes common among normal CD24.P samples. (*A*) Flow chart of SAGE-Seq experiment and data analysis. (*B*) Total counts of sense and antisense tags in each library. IDC, invasive ductal carcinoma; N, nulliparous; P, parous. CD24, CD44, EPCR, and SSEA4 indicate the cell surface markers used for the isolation of the epithelial cells (details presented in Table S1). (*C*) Box plots depicting sense and antisense tag counts per gene in each sample. The box indicates the 25th and 75th percentiles; the white bar indicates the median; the whiskers extend to the most extreme data point, which is no more than 1.5 times the interquartile range from the box; and outliers are plotted as small dots. (*D*) Number of S and SAS genes in each sample. (*E*) Box plots depicting sense and antisense tag counts per gene, and percentage of antisense in each sample. In box plots, the box indicates the 25th and 75th percentiles; the white bar indicates the median; the whiskers extend to the most extreme data point, which is no more than 1.5 times the interquartile range from the box; and outliers are plotted as small dots. (*F*) Functional annotation of common AS genes in CD24.P samples based on (GO) terms, including biological processes and cellular localization.

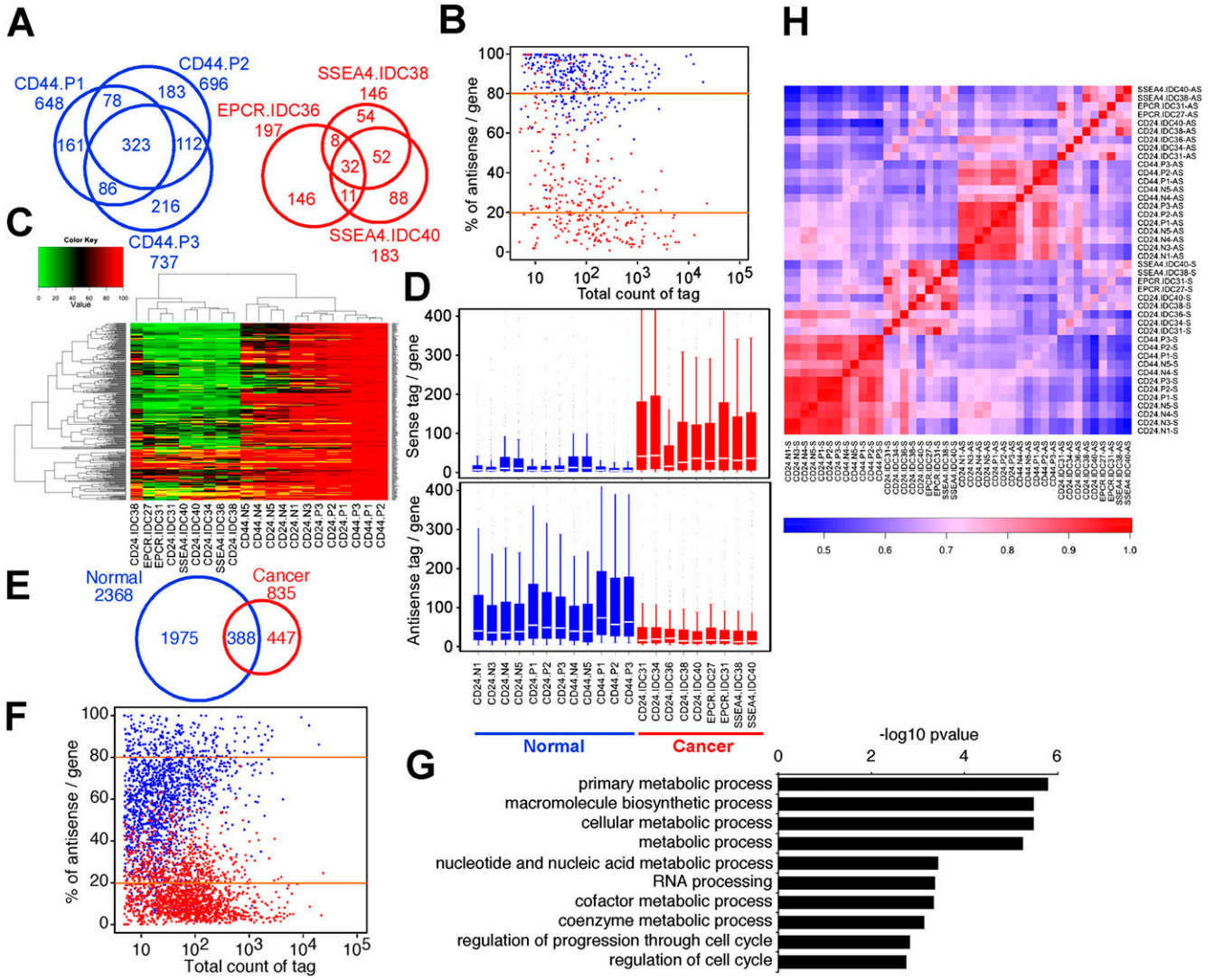

Fig. S2. Characterization of AS genes. (*A*) Venn diagrams depicting the number of overlapping AS genes among CD44$^+$ cell samples. (*B*) Scatterplot depicting percentage of antisense tags per gene relative to total tag counts in 323 AS genes common among CD44P cells in normal (blue) and cancer (red) samples. Orange lines mark the 80% and 20% values, which were used as criteria for gene classification into AS, S, and SAS groups. A list of these genes is presented in Dataset S2. (*C*) Hierarchical cluster analysis of all samples based on 323 AS genes common among CD44.P cells. The color scale indicates percentage of antisense tag counts in each gene. (*D*) Box plots depicting sense and antisense tag counts derived from 323 AS genes in each sample. The box indicates the 25th and 75th percentiles; the white bar indicates median; the whiskers extend to the most extreme data point, which is no more than 1.5 times the interquartile range from the box; and outliers are plotted as small dots. (*E*) Venn diagram shows the number of union of AS genes in all normal samples (blue) and union of all cancer samples (red). A total of 1,975 AS genes are specific for normal samples. (*F*) Scatterplots depicting the percentage of antisense tags per gene for 1,975 normal-specific AS genes in normal (*Left*) and cancer (*Right*) samples. (*G*) GO term enrichment analysis of 1,975 normal-specific AS genes. The bar graph represents the $-\log_{10}$ *P* value for the enrichment of the indicated GO term. The top 10 GO terms are shown in the graph. (*H*) Overall correlation between antisense and sense tag counts in all samples. A heat map of S and AS gene expression profiles for all the samples is shown. The $\log_2$ ratio is applied to the normalized count of all expression profiles before calculation of the cross-correlation. Red indicates higher correlation, and blue indicates lower correlation. Four subgroups are clearly shown, with higher correlations along the diagonal line, that correspond to: sense normal, sense cancer, antisense normal, and antisense cancer (*Lower Left* to *Upper Right*).

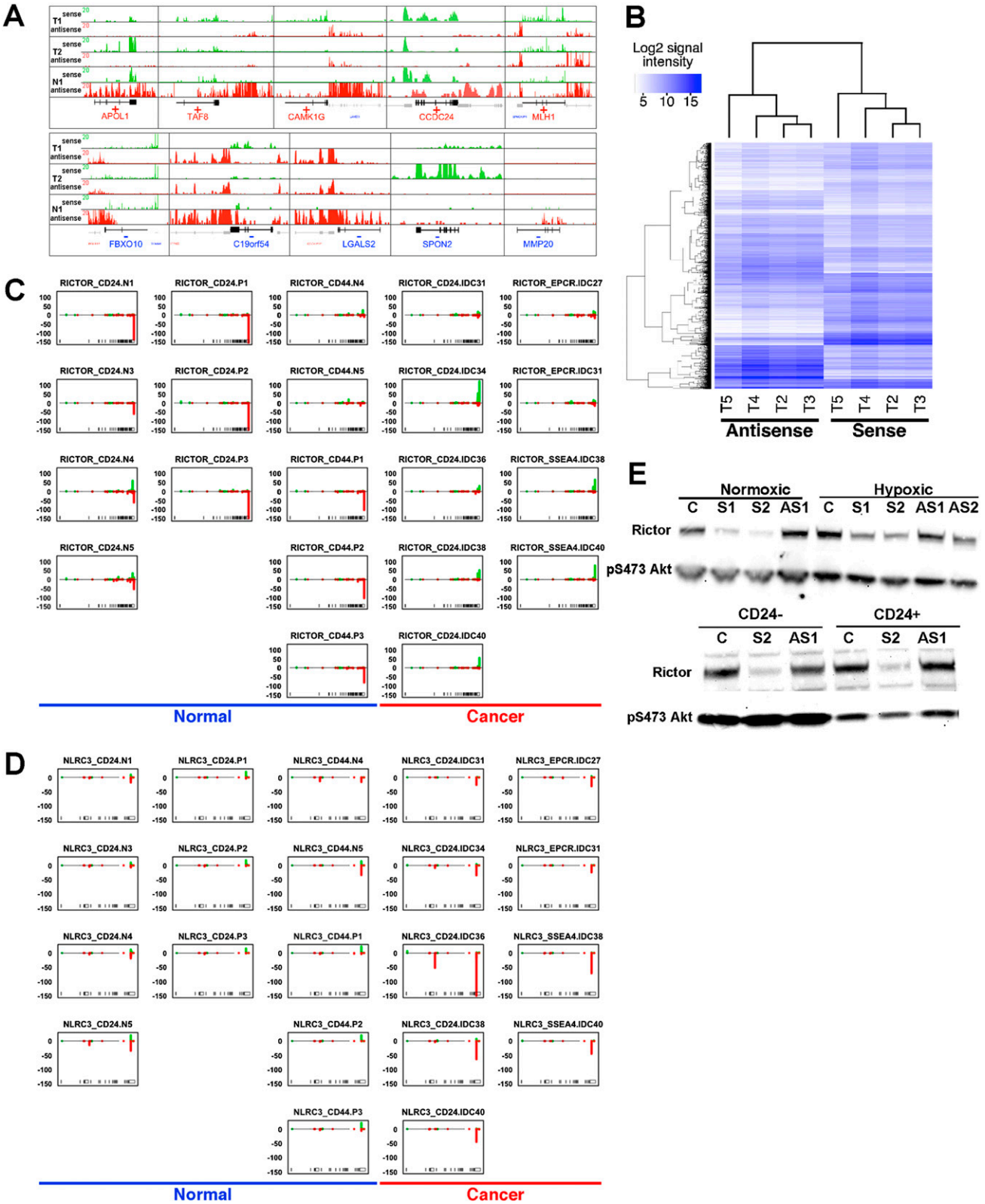

Fig. S3. Validation of AS genes using various approaches. (*A*) Representative example of tag distributions in selected genes based on ASSAGE. ASSAGE tag density from two cancer samples (T1 and T2) and one normal sample (N1) are plotted along with gene structures. Green and red mark tags derived from plus and minus strands, respectively. Gene structure and gene names are depicted under the graph; the box represents an exon; and the line represents an intron. Red and blue colors of the gene names indicate that the gene is a plus- and minus-strand gene, respectively. ASSAGE tag counts are normalized to 5 million per sample, and tag density in 10-bp windows is calculated assuming that each fragment size is 300 bp. The scale of the y axis is fixed, whereas the scale of the x axis is variable and correlates with the size of the genes. In the first example, APOL1 is a plus-strand gene, and only sense tags (plus-strand tags in this case) are

Legend continued on following page

observed in two cancer samples, whereas in the normal sample, high levels of antisense tags (minus-strand tags) are also observed. (*B*) Validation of antisense transcripts using custom oligonucleotide arrays. Processed signals for each probe were obtained from Agilent Technologies Feature Extraction software and converted to a $\log_2$ scale. One probe showing the highest average signal was selected from five probes for each sense and antisense transcript. In the heat map, each row corresponds to each gene, which has both sense and antisense transcripts. Signals for antisense transcripts were observed together with signals for sense in four cancer samples. The blue color indicates the intensity in the $\log_2$ scale. (*C*) Sense and antisense tag distribution patterns in RICTOR. Sense (green) and antisense (red) tag positions are shown across the gene structure. The left-to-right direction of the x axis corresponds to the 5′- to 3′-end of the indicated gene. The height of the bar represents normalized tag counts for each tag. (*D*) Sense and antisense tag distribution patterns in NLRC3. Sense (green) and antisense (red) tag positions are shown across the gene structure. The left-to-right direction of the x axis corresponds to the 5′- to 3′-end of the indicated gene. The height of the bar represents normalized tag counts for each tag. (*E*) Effect of antisense lncRNA on Rictor protein levels. (*Upper*) Immunoblot analysis of Rictor protein and phospho-Ser 473 Akt levels in human mammary epithelial cells cultured at 20% O_2 or 1% O_2 transfected with control (C) and Rictor sense (S1 and S2) and antisense (AS1 and AS2) targeting siRNAs. (*Lower*) Immunoblot analysis of Rictor protein and phospho-Ser 473 Akt levels in $CD24^-$ and $CD24^+$ human microvascular endothelial cells cultured at 1% O_2 transfected with control (C) or Rictor sense (S2) and antisense (AS1) targeting siRNAs.

Table S1. Summary of tissue samples used for the indicated experiments

SAGE-Seq	ASSAGE	Array	Tissue	Cell type	Differentiation status	Histology	Grade	ER	PR	HER2
EPCR.IDC27			IDC27	EPCR	Progenitor-like	Invasive ductal carcinoma, male		Pos	Pos	Neg
EPCR.IDC31			IDC31	EPCR	Progenitor-like	Inflammatory cancer	H	Neg	Neg	Pos
CD24.IDC31				CD24	Luminal epithelial					
CD24.IDC34			IDC34	CD24	Luminal epithelial	Invasive lobular carcinoma	IM	Pos	Neg	Neg
CD24.IDC36			IDC36	CD24	Luminal epithelial	Invasive ductal carcinoma	IM/H	Pos	Pos	Pos
CD24.IDC38			IDC38	CD24	Luminal epithelial	Invasive ductal carcinoma	IM	Pos	Pos	Neg
SSEA4.IDC38				SSEA4	Progenitor-like					
CD24.IDC40	T1		IDC40	CD24	Luminal epithelial	Invasive ductal carcinoma	H	Neg	Neg	Pos
SSEA4.IDC40				SSEA4	Progenitor-like					
	T2	T2	PE18	CD24	Luminal epithelial	Pleural effusion		Neg	Neg	Pos
CD24.N1			N33	CD24	Luminal epithelial	Normal, nulliparous	NA	NA	NA	NA
CD24.N3			N35	CD24	Luminal epithelial	Normal, nulliparous	NA	NA	NA	NA
CD24.N4			N47	CD24	Luminal epithelial	Normal, nulliparous	NA	NA	NA	NA
CD44.N4				CD44	Progenitor-like	Normal, nulliparous				
CD24.N5			N48	CD24	Luminal epithelial	Normal, nulliparous	NA	NA	NA	NA
CD44.N5				CD44	Progenitor-like	Normal, nulliparous				
CD24.P1			N37	CD24	Luminal epithelial	Normal, parous	NA	NA	NA	NA
CD44.P1				CD44	Progenitor-like					
CD24.P2			N39	CD24	Luminal epithelial	Normal, parous	NA	NA	NA	NA
CD44.P2				CD44	Progenitor-like					
CD24.P3			N40	CD24	Luminal epithelial	Normal, parous	NA	NA	NA	NA
CD44.P3				CD44	Progenitor-like					
	N1		N58	CD24	Luminal epithelial	Normal, nulliparous	NA	NA	NA	NA
		T3	IDC42	Bulk	NA	Invasive ductal carcinoma	H	Pos	Pos	Pos
		T4	IDC50	ESA	Luminal epithelial	Invasive ductal carcinoma	H	Neg	Neg	Neg
		T5	PE2	Bulk	NA	Pleural effusion	NA	NA	NA	NA

Name of the SAGE-Seq and ASSAGE libraries, tissue code, cell surface marker used for isolation, differentiation status of the cells, histology, tumor grade (H, high; IM, intermediate; L, low), estrogen receptor (ER) and progesterone receptor (PR), and HER2 status (Pos, positive; Neg, negative) are listed. ESA, epithelial surface antigen; IDC, invasive ductal carcinoma; NA, nonapplicable.

Table S2. Summary of differentially expressed genes

Transcript	Differential pattern	No. genes
Sense	Normal < Cancer	5,017
	Normal > Cancer	2,291
Antisense	Normal < Cancer	1,322
	Normal > Cancer	2,044

	Differential pattern		
	Sense	Antisense	No. genes
1	N < C	N < C	810
2	N < C	N > C	404
3	N < C	N = C	3,803
4	N > C	N > C	688
5	N > C	N < C	74
6	N > C	N = C	1,529
7	N = C	N < C	438
8	N = C	N > C	952

Differentially expressed genes were identified based on the $\log_2$ ratio of normalized gene tag counts applying the significance analysis of microarray algorithm. Genes with differential sense or antisense counts between normal and cancer samples were identified using a 1% FDR as the cutoff for significance.

Table S3. Comparison of ASSAGE and SAGE-Seq data: Summary of ASSAGE read counts

Library name	Raw reads	Aligned reads	% Aligned	+ Strand	− Strand
T1	10,711,873	6,130,300.00	57.2	3,145,196	2,985,104
T2	13,047,359	8,799,516.00	67.4	4,370,183	4,429,333
N1	12,752,300	1,832,358.00	14.4	915,515	916,843

T1 and T2 are $CD24^+$ cells purified from two different breast carcinomas, and N1 is normal $CD24^+$ cells.

Table S4. Comparison of ASSAGE and SAGE-Seq data: Common AS genes detected in both normal and cancer samples by SAGE-Seq

Ensemble ID	Symbol	PBMC	HCT116	Jurkat	Jurkat2	Miapaca2	MRC5	MRC5 2	T1	T2	N1	Gene class in Normal	Gene class in Cancer	% Normal	% Cancer	Sense Normal	Sense Cancer	Antisense Normal	Antisense Cancer
ENSG00000184788	SATL1	NA	100.0	NA	91.7	100.0	NA	NA	100.0	94.7	NA	AS	AS	97.8	94.2	0	1	6	10
ENSG00000196407	THEM5	100.0	NA	NA	NA	NA	NA	NA	100.0	NA	100.0	AS	AS	93.4	91.5	1	1	9	8
ENSG00000145721	LIX1	100.0	NA	NA	NA	NA	NA	NA	NA	NA	100.0	AS	AS	100.0	100.0	0	0	13	7
ENSG00000128617	OPN1SW	NA	100.0	NA	NA	100.0	100.0	100.0	100.0	100.0	100.0	AS	SAS	99.9	70.8	0	36	161	87
ENSG00000080644	CHRNA3	NA	NA	0.0	5.4	NA	NA	NA	NA	77.8	100.0	AS	AS	96.9	98.9	1	1	16	71
ENSG00000205923	CEMP1	95.8	91.7	100.0	96.6	100.0	100.0	92.9	100.0	100.0	100.0	AS	SAS	99.9	78.9	0	2	104	8
ENSG00000204352	C9orf129	NA	NA	NA	NA	NA	NA	NA	63.3	NA	100.0	AS	AS	96.2	83.6	4	2	91	9
ENSG00000183260	C20orf135	100.0	NA	100.0	100.0	100.0	NA	100.0	100.0	100.0	100.0	SAS	AS	79.9	93.2	6	2	23	26
ENSG00000128346	C22orf23	NA	NA	NA	100.0	NA	NA	100.0	98.6	98.8	97.2	AS	AS	94.8	85.7	12	5	213	33
ENSG00000137674	MMP20	NA	NA	NA	NA	NA	NA	NA	NA	NA	93.2	AS	AS	98.1	84.7	1	1	34	5
ENSG00000188493	C19orf54	6.7	8.3	0.0	3.6	0.0	0.0	11.8	22.7	17.1	81.6	AS	SAS	88.8	77.3	6	5	44	18
ENSG00000168280	KIF5C	0.0	0.0	NA	NA	NA	NA	NA	25.0	12.1	77.4	AS	AS	100.0	97.8	5	20	11223	906
ENSG00000196860	TOMM20L	NA	NA	NA	NA	NA	NA	NA	31.3	73.3	33.3	AS	AS	95.0	89.7	0	1	9	11
ENSG00000166035	LIPC	36.4	80.0	NA	41.7	51.9	NA	33.3	54.5	72.0	NA	AS	AS	88.6	92.7	1	1	7	7
ENSG00000165521	EML5	28.6	20.0	3.7	1.4	NA	NA	NA	12.1	64.0	6.8	SAS	AS	83.1	91.1	5	3	23	30
ENSG00000100095	SEZ6L	9.4	NA	NA	66.7	NA	NA	NA	58.8	54.5	NA	AS	AS	97.7	93.3	1	1	26	7
ENSG00000135144	DTX1	22.2	NA	3.2	5.3	NA	NA	NA	57.1	5.4	33.3	AS	SAS	93.0	75.1	5	3	71	9
ENSG00000165490	C11orf82	NA	0.0	0.0	0.0	0.0	NA	NA	0.0	NA	50.0	AS	AS	99.9	89.4	0	10	91	83
ENSG00000213311	C11orf82	NA	NA	NA	NA	NA	NA	NA	0.0	NA	50.0	AS	AS	99.9	89.4	0	10	91	83
ENSG00000160050	CCDC28B	32.0	37.5	8.0	11.8	5.9	52.9	38.5	46.7	4.3	20.0	AS	SAS	88.3	79.7	1	1	11	5
ENSG00000101695	RNF125	0.9	0.0	8.3	0.0	0.0	NA	NA	42.9	14.9	10.0	SAS	SAS	70.5	71.6	74	24	177	61
ENSG00000196787	HIST1H2AG	0.0	1.0	0.0	1.9	0.0	0.0	0.0	0.0	29.4	0.0	AS	SAS	83.3	70.5	1	6	7	14
ENSG00000136531	SCN2A	0.0	NA	NA	NA	NA	0.0	25.0	NA	21.4	NA	AS	AS	98.6	84.8	0	11	8	59
ENSG00000147003	TMEM27	NA	NA	NA	NA	NA	NA	NA	13.3	11.1	0.0	AS	AS	93.6	96.9	5	1	68	26
ENSG00000105926	MPP6	3.4	1.6	5.3	0.0	0.0	0.0	1.5	4.7	13.2	1.7	SAS	AS	83.3	86.1	22	12	110	76
ENSG00000173457	PPP1R14B	NA	25.0	NA	3.7	20.0	16.7	33.3	0.0	10.0	2.5	SAS	AS	73.3	83.9	18	3	48	18
ENSG00000116133	DHCR24	26.3	0.3	0.4	0.0	0.4	0.0	1.1	0.3	8.2	0.0	SAS	SAS	77.1	81.8	10	17	33	74
ENSG00000170088	TMEM192	0.0	9.1	NA	11.1	NA	NA	NA	3.9	4.5	0.0	SAS	SAS	76.2	72.6	10	23	32	60
ENSG00000048471	SNX29	0.9	12.5	3.6	2.7	26.7	3.1	2.3	1.0	4.4	2.1	AS	SAS	94.8	71.7	41	39	753	99
ENSG00000006075	CCL3	NA	NA	NA	NA	NA	NA	NA	NA	2.5	NA	AS	AS	99.9	99.9	0	1	79	607
ENSG00000206262	C3orf72	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	AS	AS	100.0	100.0	0	0	20	6
ENSG00000106436	MYLC2PL	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	AS	AS	100.0	100.0	0	0	145	16
ENSG00000185960	SHOX	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	AS	AS	100.0	100.0	0	0	10	9
ENSG00000185823	C15orf2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	AS	AS	89.1	94.5	1	0	12	5
ENSG00000188000	OR7D2	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	AS	SAS	99.0	71.7	0	5	13	12
ENSG00000161905	ALOX15	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	AS	AS	100.0	96.2	0	1	27	16
ENSG00000105996	HOXA2	NA	NA	NA	NA	NA	40.0	46.2	NA	NA	0.0	AS	SAS	97.2	70.8	0	2	8	4

Comparison of our ASSAGE (red, tumor; blue, normal) and SAGE-Seq (black) data as well as previously reported ASSAGE (green) data (1). A total of 9,550 genes common between the two datasets (our data and the Johns Hopkins University data) were used for the comparison. The percentage of antisense for each gene was calculated the same way in ASSAGE data. NA, total (sense + antisense) tag count is <5. Gene class, percentage of antisense, and sense and antisense tag counts for each gene in averaged normal and cancer data are shown. About 30% of common AS genes in SAGE-Seq also showed a high percentage of AS genes in ASSAGE data, but about 50% of them were not detected by ASSAGE, likely attributable to lower sequencing depths.

1. Saha S, et al. (2002) Using the transcriptome to annotate the genome. *Nat Biotechnol* 20:508–512.

Dataset S1. Summary of SAGE-Seq tag counts in each sample and number of genes classified as S (sense predominant), AS (antisense predominant), and SAS (both sense and antisense) in each sample

Dataset S1

Dataset S2. Identity of 252 AS genes common among three CD24.P (blue) samples and 323 AS genes common among three CD44.P (red) samples

Dataset S2

A list of genes with the gene symbol, the GenBank accession number, and a description is provided.

Dataset S3. Function of 252 AS genes common among three CD24.P (blue) samples, 323 AS genes common among three CD44.P (red) samples, and a union of all AS genes (1,975 genes in total) detected only in normal samples (black)

Dataset S3

Functional categories of genes are based on GO term analysis. GO term category, term, number (count), and percentage (%) of genes in category as well as P values are listed. Only GO terms with $P < 0.05$ are shown.

Dataset S4. Dataset S4.List of differentially expressed genes between normal and cancer samples

Dataset S4

(*A*) List of genes that show higher sense tag counts in cancer compared with normal and higher antisense tag counts in normal compared with cancer samples. (*B*) List of genes that show higher sense tag counts in cancer compared with normal and higher antisense tag counts in cancer compared with normal samples. Blue and red colors indicate normal and cancer samples, respectively.