Supplementary Materials for "Alternative preprocessing of RNA-Sequencing data in The Cancer Genome Atlas leads to improved analysis results "

Supplementary Figures

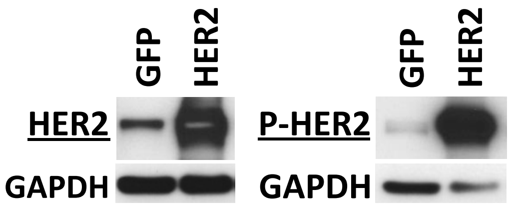

Figure S1: Western blots showing expression levels for HER2-activated and GFP-control cells. Lysates of HER2-adenovirus-vector (HER2) and green fluorescent protein (GFP) infected HMEC cells (18 hour infection) were generated, and expression of HER2 protein components were visualized by SDS-PAGE/Western blot. Western blots are shown for HER2 and phospho-Tyr1173-HER2 (P-HER2). GAPDH signal is used as an indication of loading equivalency.

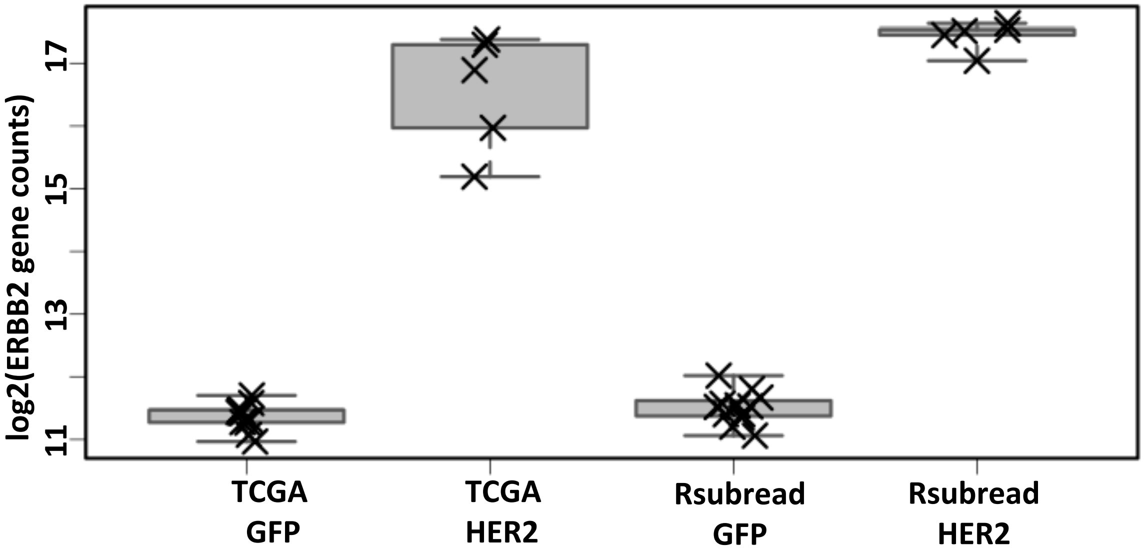

Figure S2: ERBB2 (HER2) raw gene counts produced using the TCGA and Rsubread pipelines. Log-transformed gene counts for the ERBB2 gene are shown for HER2-activated human mammary epithelial cells (n=5) and for GFP-treated control cells (n=12). For HER2-activated cells, the values were much more variable for the TCGA pipeline processed gene counts data (coefficient of variation = 0.53) than for the Rsubread data (coefficient of variation = 0.15). For the GFP-treated cells the coefficients of variation were similar for both methods (TCGA = 0.14, Rsubread = 0.18). In addition, the standardized mean difference between HER2-activated levels and GFP levels was greater for the Rsubread data (23.8) than for the TCGA data (10.0).

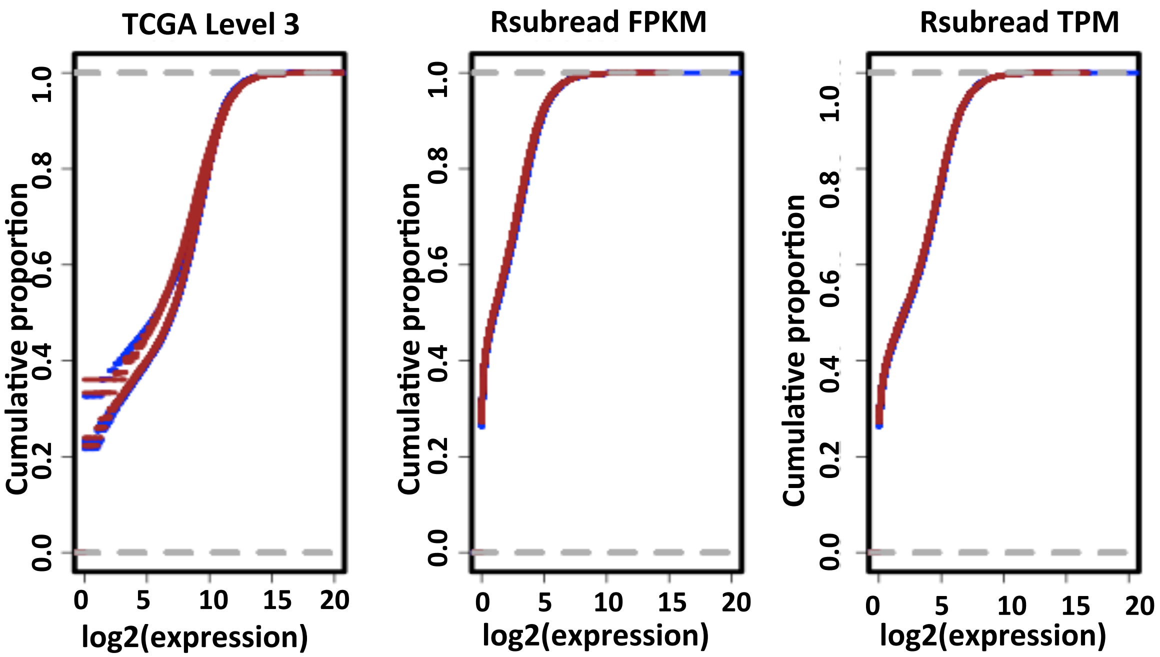

Figure S3: Empirical cumulative distribution of total mapped reads using normalized gene counts. In all cases, the cumulative distributions were more consistent for the Rsubread data than for the TCGA Level 3 data. The outlier samples for the TCGA Level 3 data are the same samples (GFP sample 4, HER2 samples 2 and 4) that showed visually different expression patterns in the heat maps (see Figure 1). GFP samples (n=12) are represented in blue and HER2 samples (n=5) are represented in brown color.

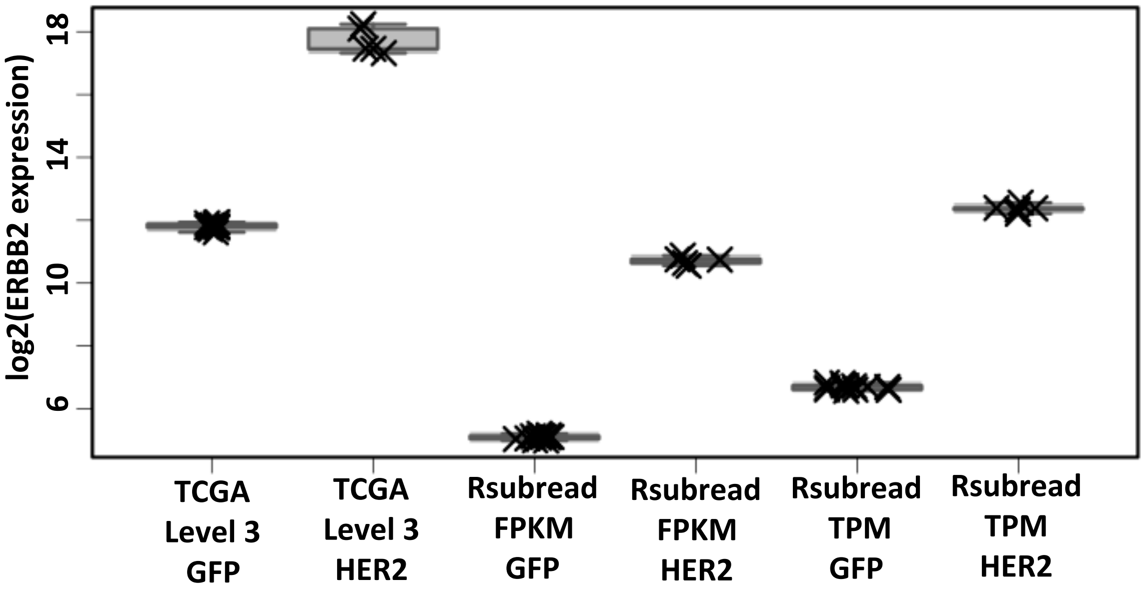

Figure S4: ERBB2 (HER2) normalized expression levels produced using the TCGA Level 3 and Rsubread pipelines. For HER2-activated cells, the values were more highly variable for the TCGA Level 3 data (coefficient of variation = 0.30) than for the Rsubread data (coefficient of variation for FPKM = 0.09, coefficient of variation for TPM = 0.06). In addition, the standardized mean difference between HER2-activated level and control levels was greater for the Rsubread data (FPKM = 66.9, TPM = 67.2) than for the TCGA Level data (25.8).

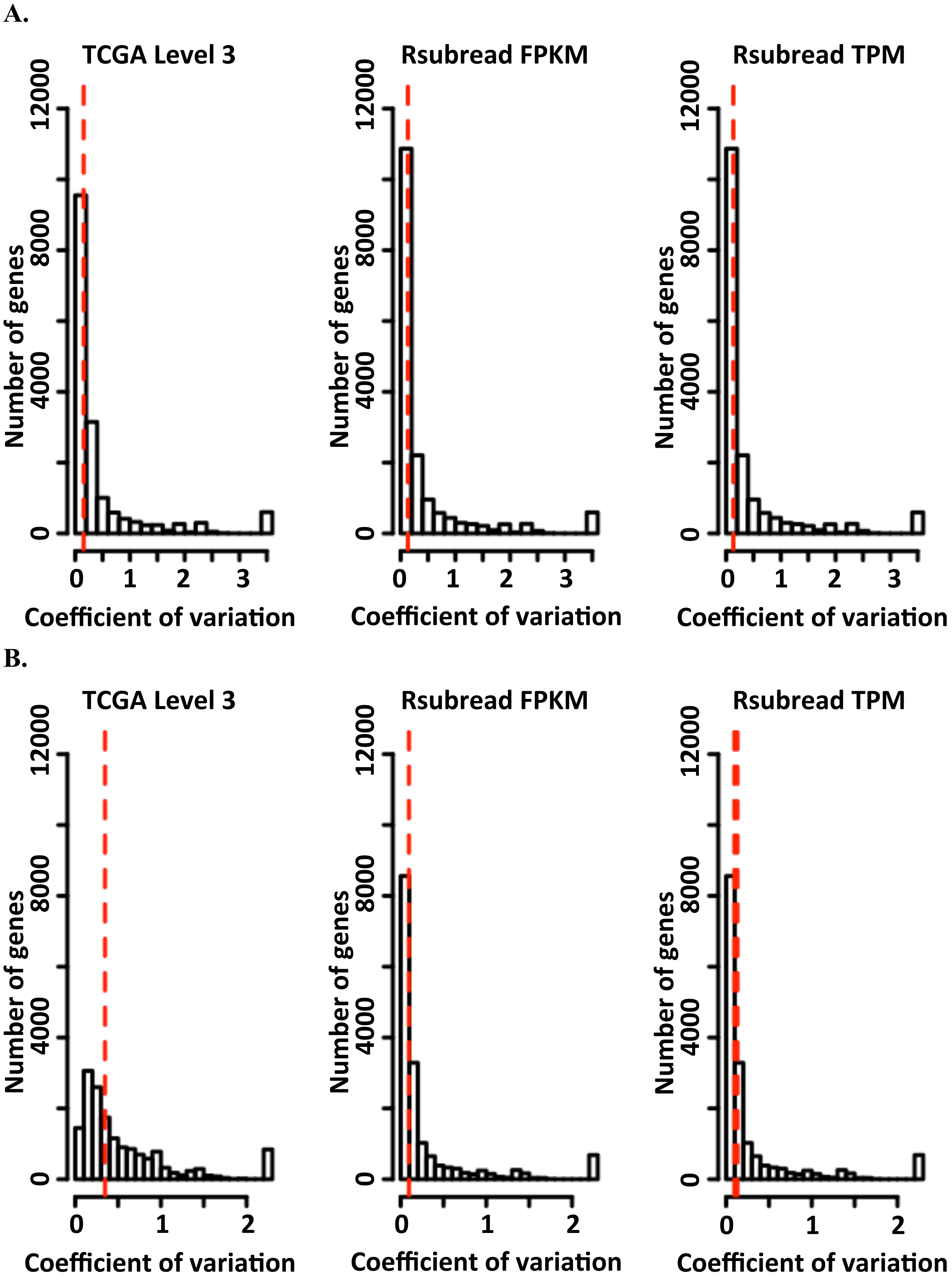

Figure S5: Histogram of coefficients of variation across (A) control and (B) HER2 overexpressed samples using 19584 common genes across the normalized gene expression datasets. In all cases there were some genes with high coefficient of variation in expression values. However, Rsubread FPKM and TPM normalized data had a higher number of genes and a lower median coefficient of variation than the TCGA Level 3 upper quartile normalized data.

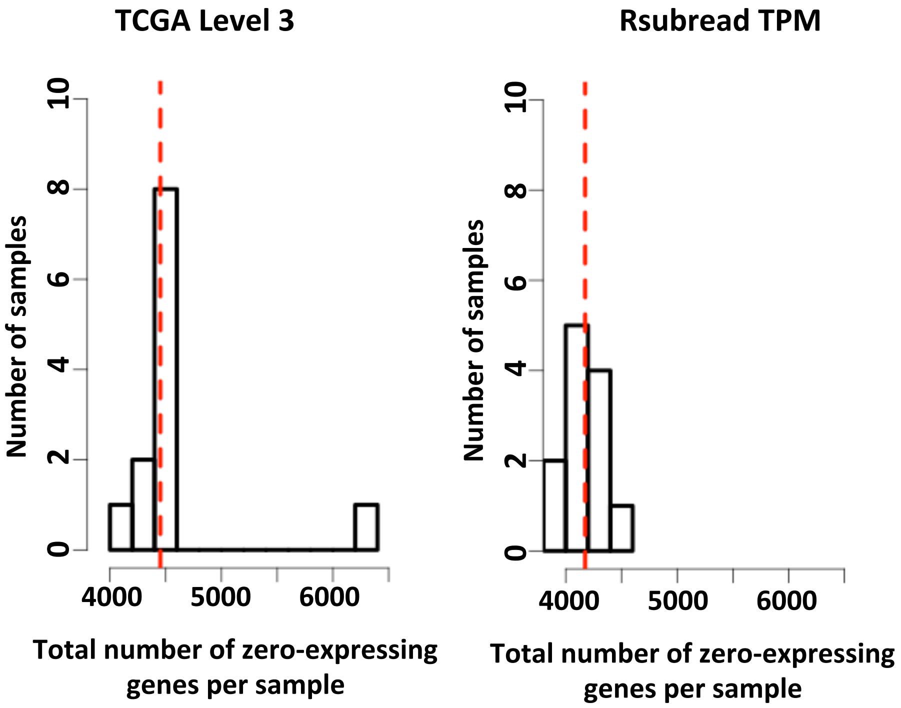

Figure S6: Distribution of number of zero expressing genes per HMEC GFP sample (n=12) for the TCGA Level 3 (median: 4452) and Rsubread TPM (median: 4174) methods.

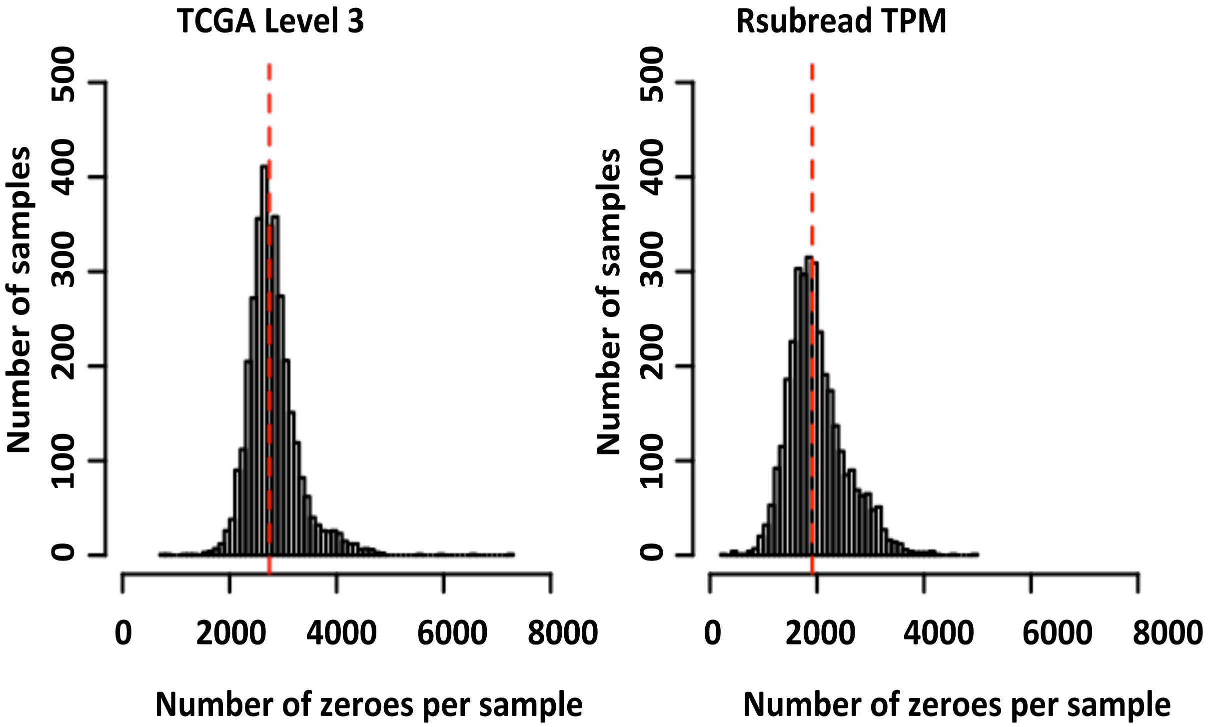

Figure S7: The number of genes per sample that each pipeline determined to have zero expression. We limited this analysis to the TCGA tumor samples (n=3380) and genes (n=19,584) that were common between the TCGA Pan-Cancer 12 dataset and our *Rsubread* processed dataset. The TCGA Level 3 samples had a higher number of zeroes per sample than the *Rsubread* samples (p-value<0.001). Vertical lines show the median value for each pipeline (TCGA Level 3 = 2742.5, *Rsubread* TPM =1910). In addition, the TCGA Level 3 data contained more extreme outliers.

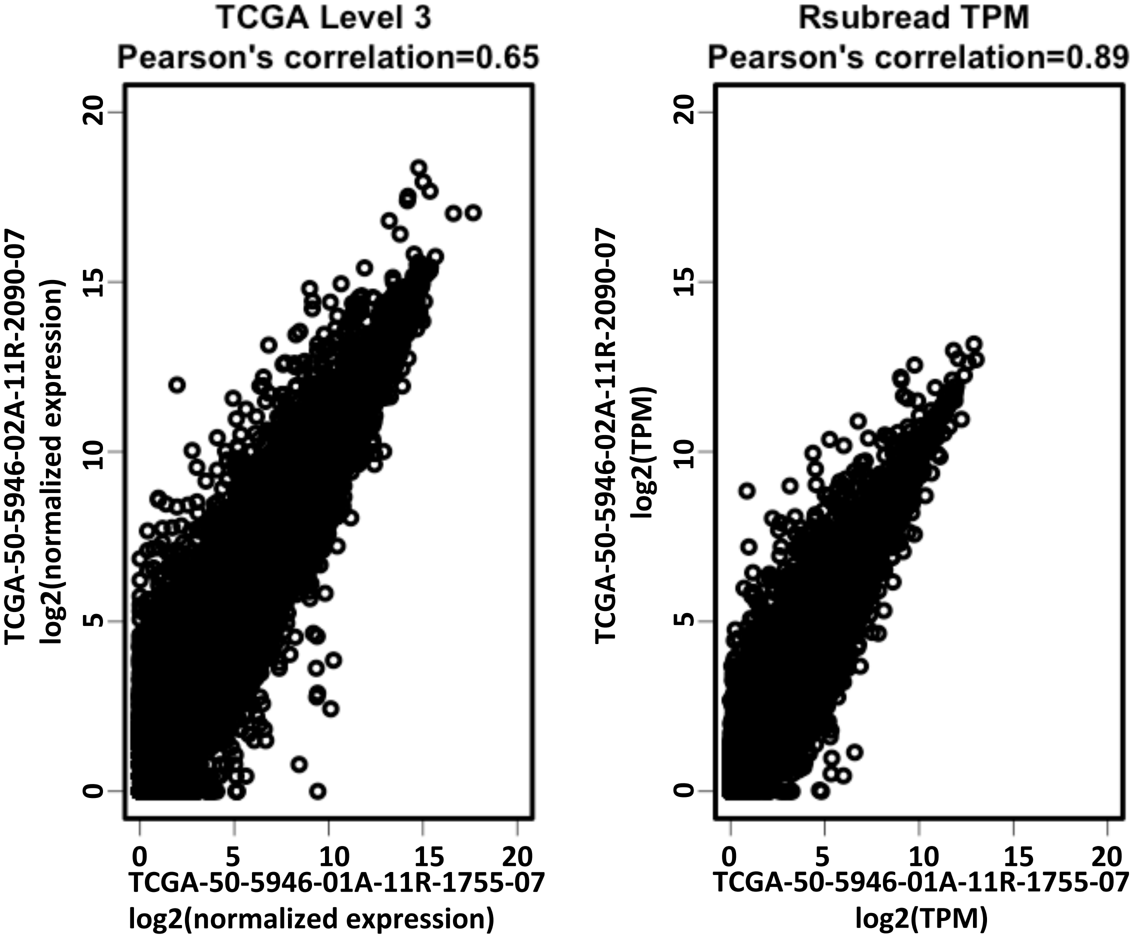

Figure S8: Scatter plots for two biological samples from patient TCGA-50-5946.

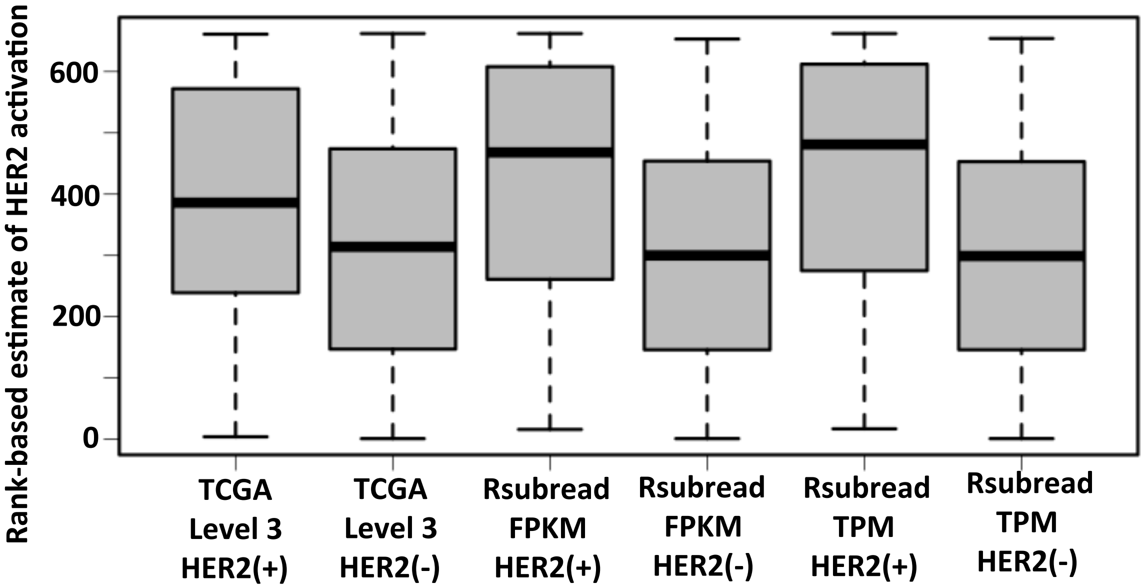

Figure S9: Signature-based estimates of HER2 activation in TCGA breast-cancer samples (n = 662). We compared samples that had been identified via immunohistochemistry as either HER2 positive or negative. The standardized mean difference between $HER2^+$ and $HER2^-$ samples was higher for the Rsubread processed data (FPKM = 0.52, TPM = 0.59) than for the TCGA Level 3 data (0.44). For visual consistency across the comparisons, we converted the signature predictions to rank-based values (a higher rank indicates that a given sample was more likely to be HER2 positive).

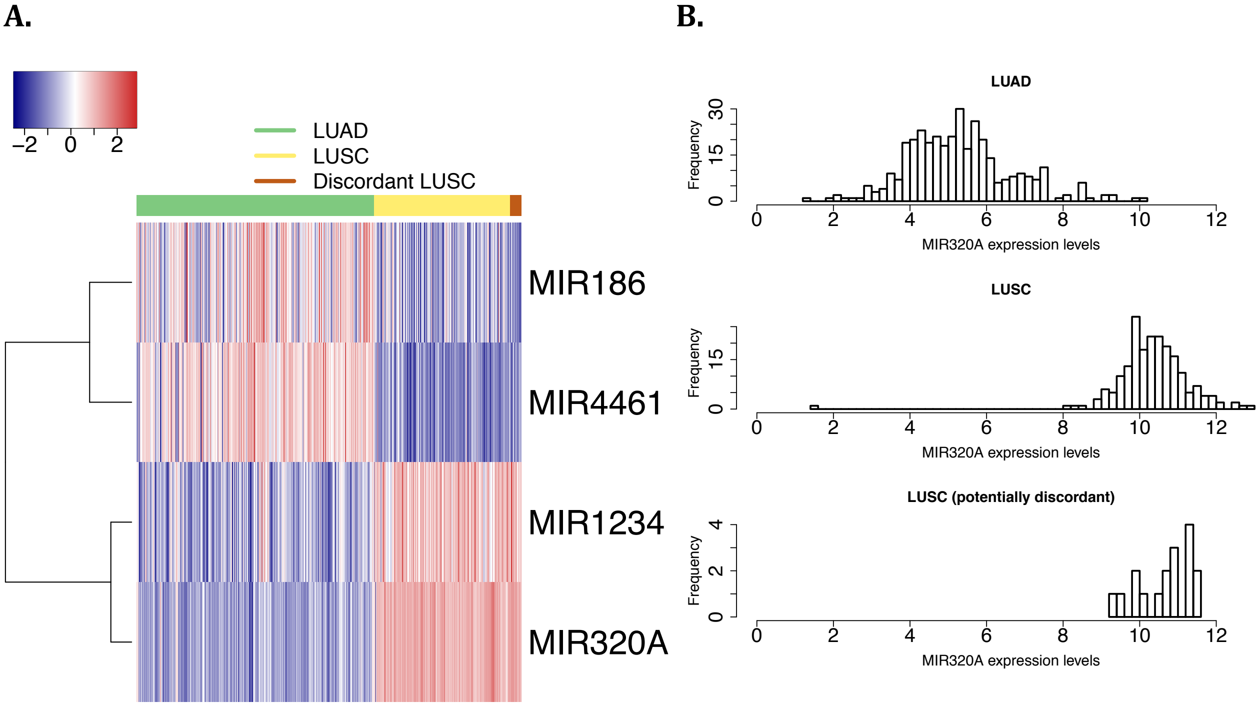

Figure S10: A. Gene expression patterns for four genes (non-coding RNA) that are consistent with LUAD and LUSC histological classification. The “Discordant LUSC” samples were identified by Cline et al. as exhibiting LUAD-like properties; however, expression levels for these genes, which are not included in the TCGA Level 3 data, are consistent with histological classification. B. Histograms showing expression levels for MIR320A gene in LUAD, LUSC and discordant LUSC samples. Expression levels for “LUSC Discordant” samples are highly concordant with LUSC samples.

1. Supplementary Tables

Table S1: Analyses scenarios, datasets and number of samples used in comparing TCGA Level 3 and Rsubread FPKM/TPM pipeline.

Analysis Name	**Goal**	**Datasets used**	**Number of samples used**
Gene counts and normalized expression	To compare gene level differences before and after normalization for the HER2 gene	Our experimental HMEC dataset	17
Effect of upper quartile normalization	To compare the number of zero-expressed genes in the dataset with common genes and samples	Common samples between TCGA PanCan 12 Level 3 and Rsubread TPM dataset	3380
HER2 gene expression signature	To compare gene expression based signatures with 200 genes	Our experimental HMEC dataset	17
HER2 status prediction using HER2 signatures	To predict HER2 status in TCGA BRCA samples where the HER2 status is known from immunohistochemistry	TCGA BRCA dataset and clinical dataset	662
Classifying TCGA lung samples	To compare accuracy in classifying gene expression based lung adeno (LUAD) versus lung squamous carcinoma (LUSC) samples	TCGA LUAD and LUSC RNA-Seq datasets	575

Table S2: Comparison of standardized means

Comparison of Hedge's standardized mean differences with all HMEC samples and with 2 HMEC outlier samples removed. For the *Rsubread* data, we used TPM values.

	All samples included [GFP n=12 and HER2 n=5]		**Outlier samples removed [GFP n=12 and HER2 n=3]**	
	TCGA Level 3	Rsubread	TCGA Level 3	Rsubread
Normalized HER2 expression	25.8	67.2	64.77	81.86
HER2 predictions	0.44	0.59	0.40	0.55

Table S3: Comparison of Pearson's correlation coefficients for biological replicates

Pearson correlation coefficients for 13 samples that had been profiled twice with RNA-Seq in our data set and in the PANCAN12 data set.

Replicate_1	Replicate_2	TCGA Level 3	Rsubread TPM
TCGA-06-0125-01A-01R-1849-01	TCGA-06-0125-02A-11R-2005-01	0.89	0.88
TCGA-06-0190-01A-01R-1849-01	TCGA-06-0190-02A-01R-2005-01	0.72	0.88
TCGA-06-0210-01A-01R-1849-01	TCGA-06-0210-02A-01R-2005-01	0.79	0.83
TCGA-06-0211-01B-01R-1849-01	TCGA-06-0211-02A-02R-2005-01	0.89	0.88
TCGA-14-1034-01A-01R-1849-01	TCGA-14-1034-02B-01R-2005-01	0.75	0.78
TCGA-19-4065-01A-01R-2005-01	TCGA-19-4065-02A-11R-2005-01	0.63	0.82
TCGA-50-5066-01A-01R-1628-07	TCGA-50-5066-02A-11R-2090-07	0.68	0.80
TCGA-50-5946-01A-11R-1755-07	TCGA-50-5946-02A-11R-2090-07	0.65	0.89
TCGA-BH-A18V-01A-11R-A12D-07	TCGA-BH-A18V-06A-11R-A213-07	0.80	0.89
TCGA-BH-A1FE-01A-11R-A13Q-07	TCGA-BH-A1FE-06A-11R-A213-07	0.69	0.65
TCGA-E2-A15A-01A-11R-A12D-07	TCGA-E2-A15A-06A-11R-A12D-07	0.90	0.93
TCGA-E2-A15E-01A-11R-A12D-07	TCGA-E2-A15E-06A-11R-A12D-07	0.83	0.86
TCGA-E2-A15K-01A-11R-A12P-07	TCGA-E2-A15K-06A-11R-A12P-07	0.79	0.85

Table S4: Coefficients for HER2 signature genes

This table lists the 200 HER2-signature genes, along with coefficients identified using the two pipelines. Among these genes, 91-92 (~46%) genes were common between the TCGA Level 3 pipeline and *Rsubread* processed (FPKM and TPM) datasets, and 188 (94%) were common between FPKM and TPM data processed by *Rsubread.*

		Rsubread			
TCGA RNA-Seq Level 3		**FPKM**		**TPM**	
Name	Coefficient	Name	Coefficient	Name	Coefficient
Intercept	4.524853	Intercept	0.168851	Intercept	-0.504928
ERBB2	0.164782	ERBB2	0.257577	ERBB2	0.305527
HSPA7	-0.125612	HSPA7	-0.187866	HSPA6	-0.15878
GDF6	-0.111343	HSPA6	-0.136333	HSPA7	-0.151412
HSPA6	-0.097087	GDF6	0.09874	CCL2	-0.106984
CCL2	-0.093873	DNAJA4	-0.080598	DNAJA4	-0.09334
CXCL10	-0.092074	KPRP	0.074612	TNFAIP2	-0.075825
LOC338651	0.079326	EEF1A2	0.069003	HSPA1A	-0.073306
TNFSF14	-0.07371	TNFAIP2	-0.06772	EEF1A2	0.07144
CD248	-0.059249	PDGFB	0.066514	PDGFB	0.06787
IFIT1	-0.057644	TSPAN18	0.066512	EPGN	-0.067303
DNAJA4	-0.053322	HSPA1A	-0.062749	HSPA1B	-0.066745
GNAO1	-0.050292	ATP6V0A4	0.058443	ATP6V0A4	0.062446
CRHR1	0.048706	CFB	-0.058034	CFB	-0.060075
EEF1A2	0.045896	HSPA1B	-0.057605	CALB2	0.05829
HSPA1B	-0.045632	EPGN	-0.057545	CRYAB	-0.054796
CCL20	-0.044527	CALB2	0.054193	SAA2	-0.050794
TNFAIP2	-0.04433	PNMA2	0.048449	PNMA2	0.0504
LOC91948	0.042751	SAA2	-0.047311	KRT80	0.050203
ATP6V0A4	0.038768	CRYAB	-0.046179	TNFRSF11B	0.048283
CFB	-0.03783	KRT80	0.045195	UCA1	0.046302
CALB2	0.036782	SRMS	0.043627	CXCL5	-0.045923
PADI1	0.035659	GPR1	-0.04332	ANGPTL7	-0.04499
PDGFB	0.034971	UCA1	0.041757	KPRP	0.044522
LOC285629	-0.034876	TNFRSF11B	0.041583	SOD2	-0.044234
CRYAB	-0.032468	FAM83A	0.040141	SYTL5	0.043949
GABRA2	0.030593	EPHA3	-0.039923	KRT19	0.043441

		Rsubread			
TCGA RNA-Seq Level 3		**FPKM**		**TPM**	
SOD2	-0.028653	CXCL5	-0.039762	AKAP12	0.043351
ULBP1	-0.028346	RGS2	-0.039724	SRMS	0.042485
KRT18	0.028246	DDAH1	0.039198	PADI1	0.042177
GPR1	-0.027639	ULBP1	-0.038466	GPR1	-0.041418
CXCL5	-0.027617	AKAP12	0.038418	RGS2	-0.041195
EPHA3	-0.026868	SOD2	-0.037183	MYADM	0.040819
IL8	-0.025943	KRT19	0.036641	SHC4	0.04055
EPHA4	-0.025735	TLR3	-0.035985	BST2	-0.039644
TLR3	-0.025646	SHC4	0.035642	EPHA3	-0.0395
HSPB8	-0.025054	PPP1R3C	-0.035295	KLK6	0.038871
RPSAP52	0.02498	PTK6	0.034658	KRT18	0.038599
RGS2	-0.024874	SPON1	0.034473	SAA1	-0.038474
SLC2A12	-0.024861	MYADM	0.034361	SPON1	0.038178
KRT19	0.024626	BST2	-0.034136	HSP90AA1	-0.038082
TRANK1	-0.024277	GRAMD2	-0.034067	TSPAN18	0.037454
MGP	0.023918	SAA1	-0.033523	EPHA4	-0.037243
SAA1	-0.023534	HSP90AA1	-0.032999	ANGPTL4	0.036491
SHC4	0.022446	KRT18	0.032801	PAQR7	-0.036256
KITLG	-0.022152	EPHA4	-0.032767	ULBP1	-0.035505
KRT8	0.022084	PIK3C2B	-0.032631	HSPH1	-0.035296
CGNL1	-0.021984	KLK6	0.032407	PGM2L1	0.035069
MYCL1	-0.021942	CXCR1	0.031954	CRHR1	0.034918
ANGPTL4	0.02165	PGM2L1	0.031133	SERPINB13	-0.03484
PARP9	-0.021303	ANGPTL4	0.031075	PIK3C2B	-0.034825
DNAJB4	-0.021262	PAQR7	-0.031038	PTK6	0.034722
SPON1	0.021236	DAPK1	-0.030705	CXCR1	0.034384
PIK3C2B	-0.021143	FAM198B	-0.03023	FAM198B	-0.034254
PARP14	-0.021042	SERPINB13	-0.030208	GRAMD2	-0.034033
SERPINB1	0.020839	GBP6	-0.030003	DDAH1	0.033964
CXCL2	-0.020713	VWA1	0.029805	GPRC5A	0.033659
SERPINB13	-0.020613	SLC1A1	0.029764	DAPK1	-0.03362
SNX9	0.020262	HSPH1	-0.029464	SLC1A1	0.033565
TRIM22	-0.020121	KITLG	-0.028275	VWA1	0.033251
DNAJB1	-0.019926	GPRC5A	0.027836	DNAJA1	-0.032433

		Rsubread			
TCGA RNA-Seq Level 3		**FPKM**		**TPM**	
KANK4	-0.019885	HSPB8	-0.027616	SNX9	0.032379
GBP6	-0.019667	SNX9	0.027574	KITLG	-0.032252
MLPH	0.019478	DNAJA1	-0.026591	HSPB8	-0.032155
APOL6	-0.019334	C10orf10	0.026544	GBP6	-0.031284
OAS3	-0.019302	SREK1IP1	0.026213	C10orf10	0.030517
HSP90AA1	-0.019165	GM2A	-0.026028	CCNA1	0.03031
KRT81	0.019156	C8orf84	0.025904	GM2A	-0.030108
GM2A	-0.019126	CCNA1	0.025808	C8orf84	0.029972
ENGASE	-0.017973	TRIM22	-0.025731	ALDH1A3	0.02968
KRT75	0.017856	APOL6	-0.025483	TRIM22	-0.029548
CBLC	0.017765	KRT8	0.025158	SREK1IP1	0.029351
CCNA1	0.017623	DNAJB4	-0.025018	KRT8	0.029074
FERMT2	0.017321	TCF4	-0.024505	NOTCH1	-0.028721
CEACAM1	0.01713	NOTCH1	-0.024433	DNAJB4	-0.028676
SLC13A5	0.017066	ALDH1A3	0.024322	FERMT2	0.027438
MTSS1L	-0.017003	MAFF	0.023981	EMP1	0.027141
TCF4	-0.016884	PARP14	-0.023917	MAFF	0.026901
PLAUR	0.016528	FERMT2	0.023615	TCF4	-0.02667
GPR110	0.01633	IL7R	-0.023182	DNAJB1	-0.02646
TP53AIP1	-0.016244	LOC644961	0.023169	PARP14	-0.026319
APAF1	0.016161	KHDRBS3	0.022993	PLAUR	0.026168
HSPH1	-0.016115	EMP1	0.022449	LOC644961	0.026082
RAB6B	0.016005	KMO	-0.022438	KHDRBS3	0.02565
LOXL4	0.015594	PLAUR	0.022023	PLAU	0.025228
OSBP2	0.015384	DNAJB1	-0.022019	KANK4	-0.02509
HSPA8	-0.015298	IFIT5	-0.021954	ESR1	-0.02467
UNC5B	-0.015048	RAPH1	0.02169	APOL6	-0.024617
RASA3	0.014898	KANK4	-0.021458	KCNN4	0.024463
KCNN4	0.014783	DUSP10	0.020861	IGFL3	-0.024452
ANPEP	0.014734	SMO	-0.020834	MTSS1L	-0.02421
AMACR	-0.01448	DFNB31	-0.020759	RAPH1	0.024168
ZC3HAV1	-0.01428	MTSS1L	-0.020665	IFIT5	-0.024094
COBLL1	-0.014277	PLAU	0.020509	DUSP10	0.024043
ECT2	0.014259	KCNN4	0.020505	PMP22	0.023801

		Rsubread			
TCGA RNA-Seq Level 3		**FPKM**		**TPM**	
SMURF2	0.014218	PMP22	0.02033	VASP	0.023373
CBR1	-0.014049	STX2	0.020322	ARRDC4	-0.023118
TUFT1	0.013455	VASP	0.02023	SMO	-0.023104
C1R	-0.013313	IGFL3	-0.020208	FAM176A	0.022803
SESN2	-0.013303	POU2F1	0.020096	CBR1	-0.022764
TWF2	0.013165	WWTR1	0.01976	WWTR1	0.022599
INPP4B	0.013134	FAM176A	0.019732	PGF	0.022576
SMO	-0.013129	PGF	0.019637	STX2	0.022286
ITGB3	0.013106	ARRDC4	-0.019625	ZPLD1	0.022175
CAST	0.013084	TNS3	-0.019394	KMO	-0.022123
FBXW7	-0.013061	CBR1	-0.019365	FAM214B	0.021843
VASP	0.012979	RASA3	0.019126	TUFT1	0.021717
SASH1	-0.012828	APAF1	0.01874	TNS3	-0.021558
MT2A	0.012725	HERC3	0.018697	MAP6	0.021499
NAV3	0.012684	HMGB3	0.018691	ST3GAL4	0.021422
NET1	0.012572	ZXDB	0.01865	HMGB3	0.021401
CGN	0.012481	ST3GAL4	0.018588	HS6ST1	-0.021304
SYTL2	-0.01244	HS6ST1	-0.018541	DLC1	-0.021275
CYBASC3	-0.012341	IGF2BP3	0.018523	POU2F1	0.021216
ST3GAL4	0.012295	TUFT1	0.018493	APAF1	0.021057
TNS3	-0.012073	FAM214B	0.018467	STOX2	-0.020845
BCAR3	0.011678	NET1	0.017866	RASA3	0.020767
SEC24D	0.011623	XPC	-0.017726	HERC3	0.020487
DTX4	-0.011553	FBXO22	-0.017678	DFNB31	-0.020337
PYGB	0.011389	MR1	-0.017472	FBXO22	-0.02015
MYO1E	0.011297	CYBASC3	-0.017218	BRMS1	-0.020097
PTPRE	0.011089	KCNJ5	-0.017167	IER3	0.020017
GFPT1	0.011087	IER3	0.017056	NET1	0.019989
ACTB	0.011033	NME7	0.016958	CYBASC3	-0.019984
STIM2	-0.011012	PYGB	0.016808	PYGB	0.01983
XPC	-0.011008	NAV3	0.016742	XPC	-0.019811
MFI2	0.01095	BRMS1	-0.016648	BCAR3	0.019647
NFATC3	-0.010879	ARV1	-0.016434	ZXDB	0.019586
C19orf66	-0.010511	BCAR3	0.016403	CELF2	0.019402

		Rsubread			
TCGA RNA-Seq Level 3		**FPKM**		**TPM**	
PDZD2	-0.010452	ARHGAP12	0.016383	IGF2BP3	0.019325
ARHGEF2	0.010354	PPP3CC	0.016377	TIMP1	-0.019048
TRIOBP	0.010316	PODXL2	0.016365	ARHGAP12	0.01901
SLC34A2	-0.010288	PDZD2	-0.016253	NME7	0.018951
FRMD4A	-0.010219	TWF2	0.016132	ARV1	-0.018928
MAP3K2	-0.010081	RBMS2	0.016093	CASP1	-0.018873
NPAS2	0.010074	CASP1	-0.015992	MR1	-0.018826
IGFL3	-0.009956	TIMP1	-0.015829	KCNJ5	-0.018762
ARHGAP12	0.009927	LRRC8C	0.015828	LRRC8C	0.018716
SH2D3A	0.009911	SH3KBP1	0.015714	TWF2	0.018592
NAV2	-0.009866	CAST	0.015525	PPP3CC	0.018547
SMOC1	0.009764	TP53AIP1	-0.0153	ANKRD33B	-0.018542
HERPUD1	0.009567	DAB2	0.015248	CAST	0.018294
WDR1	0.009562	FGFR2	-0.01521	SH3KBP1	0.017947
RASA1	0.009529	INPP4B	0.015146	PODXL2	0.017847
MBD4	-0.009337	HMGN3	-0.01512	INPP4B	0.017676
PLEK2	0.009276	SESN1	-0.014994	TNS4	0.01766
BCAP29	0.00927	TRIOBP	0.01497	DAB2	0.017551
ATG16L1	0.009237	GFPT1	0.014771	MFI2	0.01754
LDB1	-0.009222	ARHGEF2	0.014671	RBMS2	0.017501
NCDN	-0.009177	TNS4	0.014658	FGFR2	-0.017469
NEK9	-0.009083	MFI2	0.014631	GFPT1	0.017427
CSGALNACT2	0.009018	CROT	-0.014554	TP53AIP1	-0.017304
ATP1B1	-0.008895	KIAA1671	-0.013946	NAV3	0.017121
APBB2	-0.008881	ZNFX1	-0.013815	ARHGEF2	0.017063
CAPN2	0.00888	DNAJB9	0.013602	SESN1	-0.016845
CALM2	0.008674	NFE2L1	-0.013277	DNAJB9	0.016278
TRAFD1	-0.008589	PIK3R1	-0.013264	NFE2L1	-0.016229
PGM1	0.008555	FBXW2	-0.013023	TRIOBP	0.016197
FGFR2	-0.008354	RASSF1	0.012832	KIAA1671	-0.016057
DOPEY1	-0.008331	MICALCL	0.01279	ZNFX1	-0.015835
NISCH	-0.008191	SLC20A2	0.012767	CROT	-0.015664
PI4KB	-0.008141	LDB1	-0.012706	SLC20A2	0.015334
TOR3A	-0.007819	IGFBP4	-0.012603	B2M	-0.015314

		Rsubread			
TCGA RNA-Seq Level 3		**FPKM**		**TPM**	
LRIG3	0.007766	SEC24D	0.012592	UBB	-0.015001
POLR2A	-0.007749	B2M	-0.012511	FBXW2	-0.014918
NEU1	-0.007665	CCDC50	0.012451	LDB1	-0.014863
KPNA4	0.007656	SLC41A1	-0.012315	SEC24D	0.014746
PIK3CD	0.007606	TOR3A	-0.01228	MICALCL	0.014702
ANKRD13A	-0.007496	HERPUD1	0.012254	MYO1E	0.014521
TBRG1	-0.007462	TRAFD1	-0.012195	RASSF1	0.014486
EPS15	0.007458	MYO1E	0.012108	TOR3A	-0.01446
TRIM5	-0.007361	MEF2D	0.012092	PIK3R1	-0.014459
PCSK7	-0.007332	FRMD4A	-0.011928	TRAFD1	-0.014282
ANKFY1	-0.00732	LRRFIP1	0.011781	ANKRD13A	-0.014195
C20orf194	0.007244	ANKRD13A	-0.011763	SLC41A1	-0.014065
C19orf42	-0.007162	PI4KB	-0.011583	MEF2D	0.013983
ITGA5	0.007095	PRRC1	0.011518	PI4KB	-0.013683
ARHGEF12	-0.006996	UBB	-0.011513	LRRFIP1	0.013638
STK40	-0.006932	FAM129B	0.011441	PRRC1	0.013535
MLLT6	-0.006786	PNMAL1	-0.010498	FRMD4A	-0.012667
C1orf85	-0.006767	LPP	0.010416	PNMAL1	-0.012235
PTPN12	0.00648	APBB2	-0.010189	LPP	0.011861
MAP2K4	-0.006351	PRDM4	-0.010085	CAPN2	0.011646
ZNF532	-0.006134	ADAR	-0.010018	ADAR	-0.011625
AFAP1L2	0.006103	SEC14L1	0.009938	PRDM4	-0.011432
ARID1B	-0.005924	CAPN2	0.009793	APBB2	-0.01135
SEC14L1	0.005811	ASAP2	0.009678	SEC14L1	0.011315
PLEKHA6	-0.005776	PPP2R5B	0.00955	UBP1	-0.010824
ELOVL1	0.005764	NFATC3	-0.009429	ASAP2	0.010731
CLASP1	-0.005727	PRPSAP2	-0.009416	PRPSAP2	-0.010671
SMEK1	-0.005478	DCAF7	0.009216	PPP2R5B	0.010646
NUMA1	-0.005168	MEX3C	0.009174	NFATC3	-0.010535
ZMYND8	0.005151	AFAP1	0.009148	AFAP1	0.010482
PDXK	-0.005071	UBP1	-0.008794	DCAF7	0.010296
MYO10	0.004929	ARHGEF12	-0.008606	MYL12A	0.009901
UBP1	-0.00478	SDC1	0.008466	ARHGEF12	-0.009895
RCC2	0.004742	ADCY9	-0.008152	STAT3	-0.009518

		Rsubread			
TCGA RNA-Seq Level 3		**FPKM**		**TPM**	
SGK1	0.004731	STAT3	-0.008103	ANKRD27	0.008986
RFWD3	-0.004666	ANKRD27	0.007958	IFFO2	0.008553
C20orf3	-0.004354	IFFO2	0.007081	GTF2I	-0.008151
WDR91	-0.004333	GTF2I	-0.006848	CYB561	0.00765

Table S5: Coefficients of variation for HER status predictions in TCGA breast cancer samples.

HER2 status	**Method used**	**Coefficient of variation**
HER2 (-)	TCGA	0.62
	Rsubread FPKM	0.21
	Rsubread TPM	0.30
HER2 (+)	TCGA	0.72
	Rsubread FPKM	0.14
	Rsubread TPM	0.20