

lnCAR: a comprehensive resource for lncRNAs from Cancer Arrays

Yueyuan Zheng^{1,#}, Qingxian Xu^{1,#}, Mengni Liu^{1,#}, Huanjing Hu^{1,#}, Yubin Xie¹, Zhixiang Zuo^{1,*}, Jian Ren^{1,*}

1. State Key Laboratory of Oncology in South China, Cancer Center, Collaborative Innovation Center for Cancer Medicine, School of Life Sciences, Sun Yat-sen University, Guangzhou 510060, China

[#] These authors contributed equally to this work

^{*} Correspondence to: Jian Ren (renjian@sysucc.org.cn); Zhixiang Zuo (zuozhx@sysucc.org.cn).

SUPPLEMENTARY METHODS

Integration of lncRNA annotations

All of the gene annotations from the Ensembl project and RefSeq database were combined using the Cuffcompare program in the Cufflinks suite (1-3). We obtained lncRNAs and protein-coding genes according to the following protocol steps (**Supplementary Fig. S3**). First, we merged the fully matched (class code “=”) or contained transcripts (class code “c”). Second, the gene-transcript mapping files in different databases were downloaded from the Ensembl and Refseq database to correct and supplement the results, respectively. Third, the transcripts were divided into protein-coding and non-coding transcripts according to the “Transcript biotype”. Fourth, small RNA and pseudogene transcripts were filtered out. Only transcripts longer than 200nt were retained as candidate lncRNA transcripts. We then predicted the coding ability of the lncRNA transcripts with the CPAT (4) and CNCI (5) programs and removed those transcripts which both predicted as protein-coding. Finally, the protein-coding and lncRNA transcripts were separately combined according to their gene names and genomic coordinates.

Probe re-annotation

The following steps were taken to re-annotate the probe sequences (**Supplementary Fig. S3**). First, all of the probe sequences from the different platforms were aligned to the human genome (GRCh38) using a BLAST-like alignment tool (BLAT) (6). Only the alignments with no more than one mismatch, no gaps and a similarity score larger than 90 were preserved. In general, the probes (50 to 60 nucleotides) from Agilent, Illumina and other companies were designed to locate target genes or transcripts. The microarrays from Affymetrix utilized a probe set containing a group of 25mer probes to represent a gene or transcript. Thus, for the Affymetrix data we combined probes that specifically corresponded to the same probe set and ensured there were at least 3 perfectly matching and adjacent probes in each probe set. Second, we mapped the probes to coding genes or lncRNAs according to their genomic coordinates. If any probes were targeted to both coding genes and lncRNAs, we only preserved the annotation of the coding genes.

Differential expression analysis

We categorized differential expression with one of the following conditions: cancer versus normal controls, high stage versus low stage, high grade versus low grade, metastasis versus primary, drug-treatment versus control, and “other features”. Other features referred to certain cancer-specific features, such as estrogen receptor, progesterone receptor and *HER2* status in breast cancer as well as the smoking and drinking status in lung cancer and esophageal cancer. The differential expression analysis under each condition was performed with the Limma

package (7). We utilized a robust rank aggregation algorithm to integrate the lncRNA profiles in an unbiased manner (8). The aggregation rank score (AR score) represents the integrated rank from the meta-analysis of the fold-change in the different microarray studies. All results were scaled by cancer subtype and presented in the heatmaps, allowing users to explore the expression of lncRNAs of interest in the different conditions.

Survival analysis

All of the data we collected contained four types of survival data: overall survival, relapse-free survival, metastasis-free survival and progression-free survival. For each study with survival information, we accessed the associations between lncRNA expression and survival via a univariate Cox regression analysis (R 'survival' package v. 3.4) (9). Specifically, genes with hazard ratios >1 and a p value <0.05 were considered as poorly prognostic lncRNAs, whereas genes with hazard ratios of $0-1$ and a p value < 0.05 were lncRNAs with good prognosis value. To check the prognostic landscape of lncRNAs in each cancer type, and to reduce the differences between the different studies, we utilized the Z scores in the Cox regression to perform a meta-analysis. Z scores were directly associated with P values, and were independent of different follow-up time scales and the range of input variables, allowing a direct comparison across multiple studies and platforms. We calculated the prognostic meta Z scores of lncRNAs in each cancer type by using Lipták's weighted meta Z test, in which weights refer to the square roots of the sample sizes (10,11). lncRNAs with positive meta Z scores indicated poor prognosis value in most studies; otherwise they were a bad prognosis. The results were shown in heatmaps in order to facilitate comparative analysis.

Advanced annotations and other analysis

Advanced annotations contained the transcript information, structure, coding potential and conservative score. Briefly, all transcript information was extracted from NCBI and Ensembl, and the transcript structure was predicted by RNAfold (12). CPAT and CNCI were used to assess the coding potential of the transcripts. The conservative scores of the transcripts were calculated from Phastcon100 (13). In addition, cancer-related lncRNAs validated by low-throughput experiments were manually collected from Lnc2Cancer, LncRNADisease, EVLncRNAs and MNDR v2.0. These data were integrated into lncCAR as validation datasets.

Moreover, our database also included co-expression analysis, pathway enrichment analysis and ceRNA analysis modules to help researchers further explore the function of lncRNAs of interest. To determine the co-expression patterns of each lncRNA and protein-coding genes in a study, we calculated pairwise expression correlations across all of the protein-coding genes. The first 200 lncRNA-coding gene pairs with strong correlations (correlation coefficient $|r|>0.3$) were selected to represent the co-expression network. Users are able to further determine

the potential function of such lncRNA through the pathway enrichment analysis of co-expressed protein-coding genes. In addition, we obtained the lncRNA-miRNA and miRNA-coding gene interaction pairs from starBase2 and constructed a ceRNA interaction network based on the lncRNAs and protein-coding genes having positive expression correlations (14).

Determination of cancer subtypes by lncRNA expression

Initially, to normalize the cross-platform datasets, some filters were employed on each dataset: i) lncRNAs shared by 2-3 platforms with the largest samples were selected as the standard and other expression platforms were merged subsequently. ii) For each study, batch effects were minimized using the Combat function (15). iii) MergeMaid was applied to remove lncRNAs of low reliability across platforms and the preserved data was scaled by mean centering on genes and samples, respectively (16). iv) 500 lncRNAs with the highest expression median difference were screened for the clustering process. Secondly, we exploited the Consensus Clustering method to classify the cancer subtypes based on lncRNA expression (17). The number of clusters in each cancer was determined by the relative change in area under the CDF curve (the delta area < 0.05). Then the differentially expressed lncRNAs in each cluster were identified using a moderated F-test, and these lncRNAs were merged for the next iteration. The iteration stopped when the lncRNAs were unchanged from the previous run. The resulting lncRNAs were regarded as the gene signatures for each cancer subtype.

Web interface implementation

In order to visualize the analysis results, multiple statistical diagrams were embedded in the web server. For example, the heatmaps were constructed by DataTables, the secondary structures of transcripts were shown using FornaContainer (18), the expression network were presented by Highcharts, the boxplot together with the bubble chart were demonstrated by Echarts, and the circos plots were displayed through BioCircos.js (19). Furthermore, all of the analyses in the lncCAR website were performed in R.

In addition, we allowed the genomic coordinates of the regions of interest as an input to search the expression and prognosis of any user-defined lncRNA. For each cancer, we filtered probes that targeted protein-coding transcripts in each platform and merged all the remaining probes as a candidate lncRNA library. Then the regions submitted by users were transformed into “bed” format and intersected with the candidate lncRNA probes by applying the BEDTools’ intersectBed (v. 2.26.0) (20). The differential expression analysis and survival analysis of each targeted probes were presented in the result pages.

Reference

1. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature biotechnology* **2010**;28:511-5
2. Zerbino DR, Achuthan P, Akanni W, Amode MR, Barrell D, Bhai J, *et al.* Ensembl 2018. *Nucleic acids research* **2018**;46:D754-D61
3. Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, *et al.* RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic acids research* **2018**;46:D851-D60
4. Wang L, Park HJ, Dasari S, Wang S, Kocher JP, Li W. CPAT: Coding-Potential Assessment Tool using an alignment-free logistic regression model. *Nucleic acids research* **2013**;41:e74
5. Sun L, Luo H, Bu D, Zhao G, Yu K, Zhang C, *et al.* Utilizing sequence intrinsic composition to classify protein-coding and long non-coding transcripts. *Nucleic acids research* **2013**;41:e166
6. Kent WJ. BLAT--the BLAST-like alignment tool. *Genome research* **2002**;12:656-64
7. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic acids research* **2015**;43:e47
8. Kolde R, Laur S, Adler P, Vilo J. Robust rank aggregation for gene list integration and meta-analysis. *Bioinformatics* **2012**;28:573-80
9. Therneau TM, Grambsch PM. Modeling survival data: extending the Cox model. Springer Science & Business Media; 2013.
10. Zaykin DV. Optimally weighted Z-test is a powerful method for combining probabilities in meta-analysis. *Journal of evolutionary biology* **2011**;24:1836-41
11. Gentles AJ, Newman AM, Liu CL, Bratman SV, Feng W, Kim D, *et al.* The prognostic landscape of genes and infiltrating immune cells across human cancers. *Nature medicine* **2015**;21:938-45
12. Lorenz R, Hofacker IL, Stadler PF. RNA folding with hard and soft constraints. *Algorithms for molecular biology : AMB* **2016**;11:8
13. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome research* **2005**;15:1034-50
14. Li JH, Liu S, Zhou H, Qu LH, Yang JH. starBase v2.0: decoding miRNA-ceRNA, miRNA-ncRNA and protein-RNA interaction networks from large-scale CLIP-Seq data. *Nucleic acids research* **2014**;42:D92-7
15. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **2012**;28:882-3
16. Cope L, Zhong X, Garrett E, Parmigiani G. MergeMaid: R tools for merging and cross-study validation of gene expression data. *Statistical applications in genetics and molecular biology* **2004**;3:Article29
17. Wilkerson MD, Hayes DN. ConsensusClusterPlus: a class discovery tool with confidence assessments and item tracking. *Bioinformatics* **2010**;26:1572-3
18. Kerpedjiev P, Hammer S, Hofacker IL. Forna (force-directed RNA): Simple and effective

- online RNA secondary structure diagrams. *Bioinformatics* **2015**;31:3377-9
19. Cui Y, Chen X, Luo H, Fan Z, Luo J, He S, *et al.* BioCircos.js: an interactive Circos JavaScript library for biological data visualization on web applications. *Bioinformatics* **2016**;32:1740-2
 20. Quinlan AR. BEDTools: The Swiss-Army Tool for Genome Feature Analysis. *Current protocols in bioinformatics* **2014**;47:11 2 1-34

SUPPLEMENTARY FIGURES

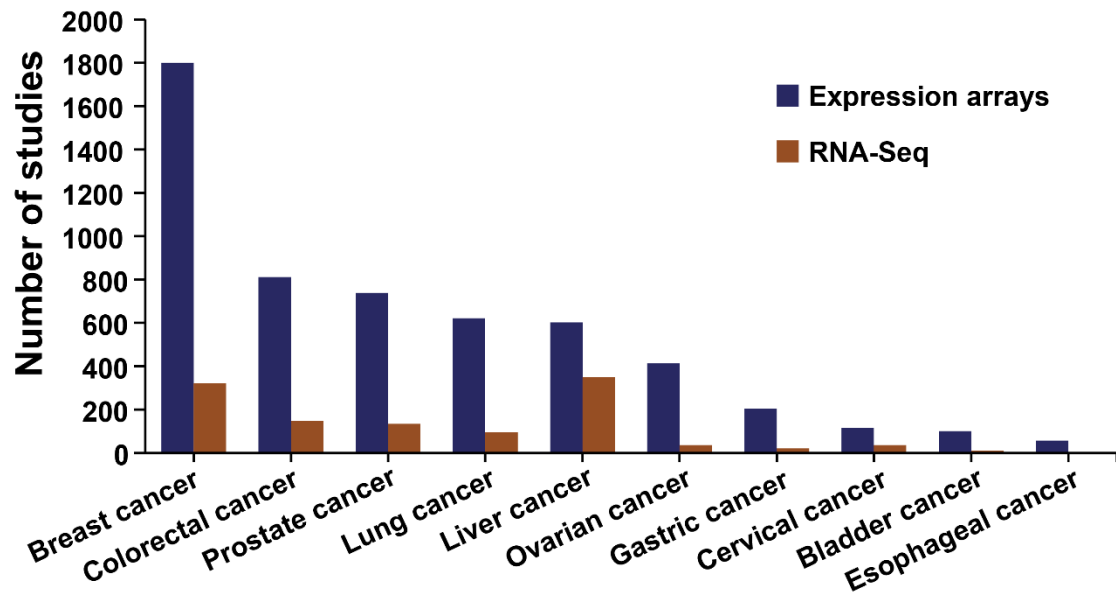


Fig. S1 – Comparison of the number of published gene expression datasets and RNA-Seq datasets from GEO database in cancer . All the datasets in GEO database were limited to human studies published before September 2017.

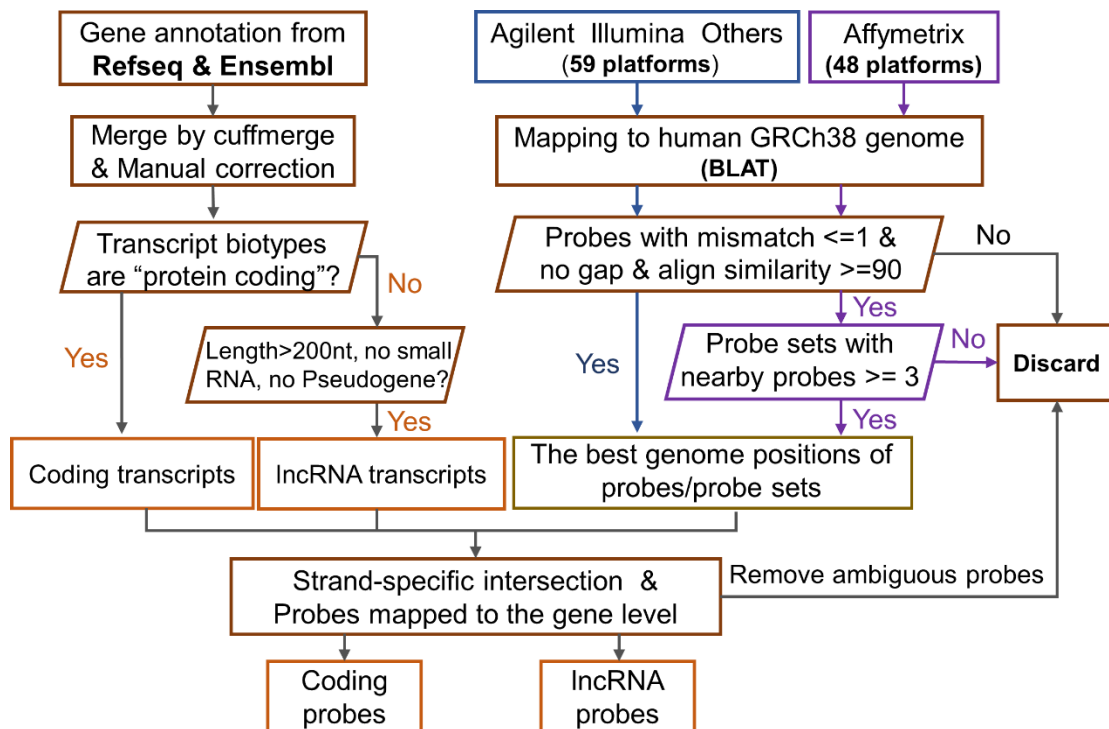


Fig. S2 - Computational pipeline for re-annotation of microarray probes

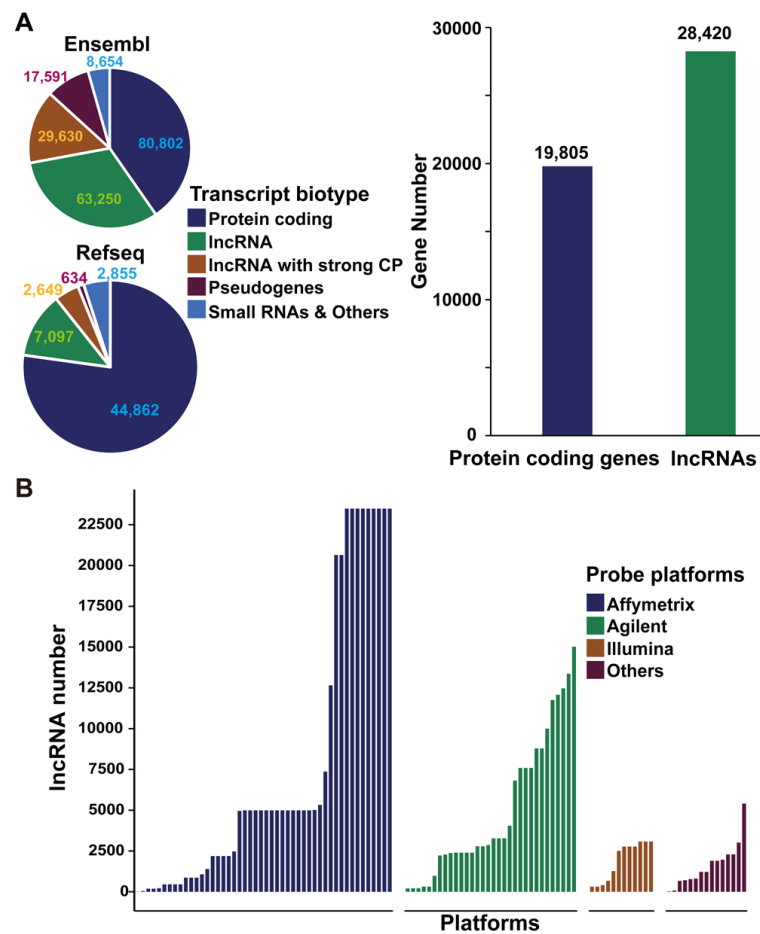


Fig. S3 - Re-annotation of the array probes. (A) The percentage of the transcript type distribution in protein-coding gene, lncRNA, pseudogenes, small RNAs and others for the transcripts from the Ensembl and RefSeq annotations (left panel), and the number of protein coding genes and lncRNAs in Ensembl and RefSeq annotations (right panel). (B) The number of lncRNAs probes reannotated from different array platforms including Affymetrix, Agilent and Illumina.

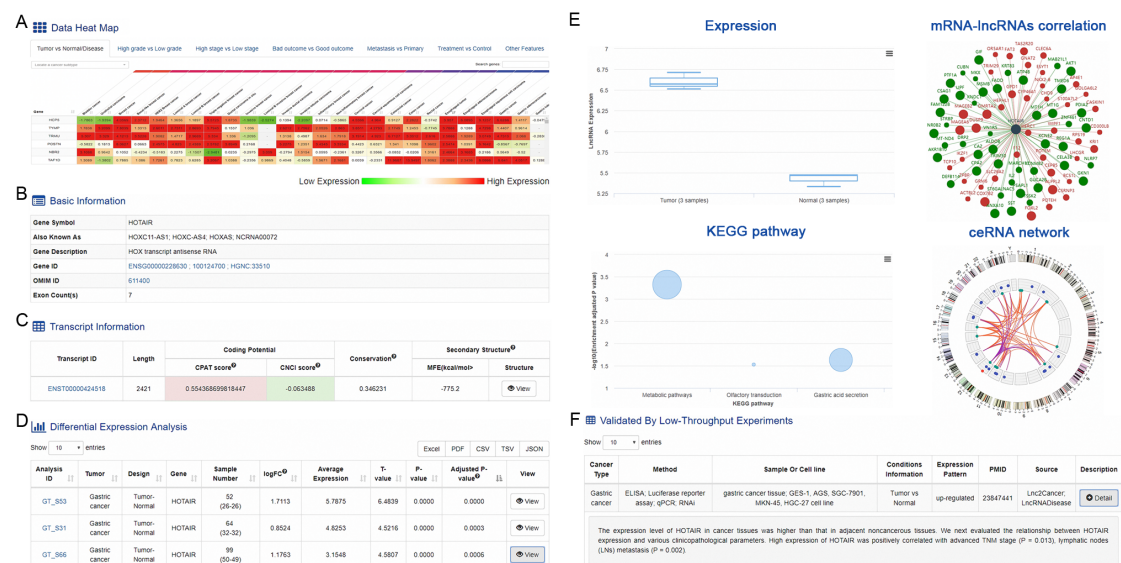


Fig. S4 - A snapshot of the example for differential expression analysis in lncCAR. (A) The lncRNA expression characteristics in different cancer subtypes under the same conditions. Red cells indicate increased expression while green cells indicate decreased expression. (B) The basic information on lncRNA gene. (C) The information on the lncRNA transcripts, including the coding potential, degree of conservation and secondary structure. (D-E) The expression of lncRNA in different studies of gastric cancer compared to normal tissues. Click the “view” button to get the expression pattern of lncRNAs in different tissues, the related coding genes, the potential biological pathways and related ceRNA network. (F) The validated results in the low-throughput experiments.

differential expression analysis in different studies. (C) Prognostic impact of target probes by survival analysis.

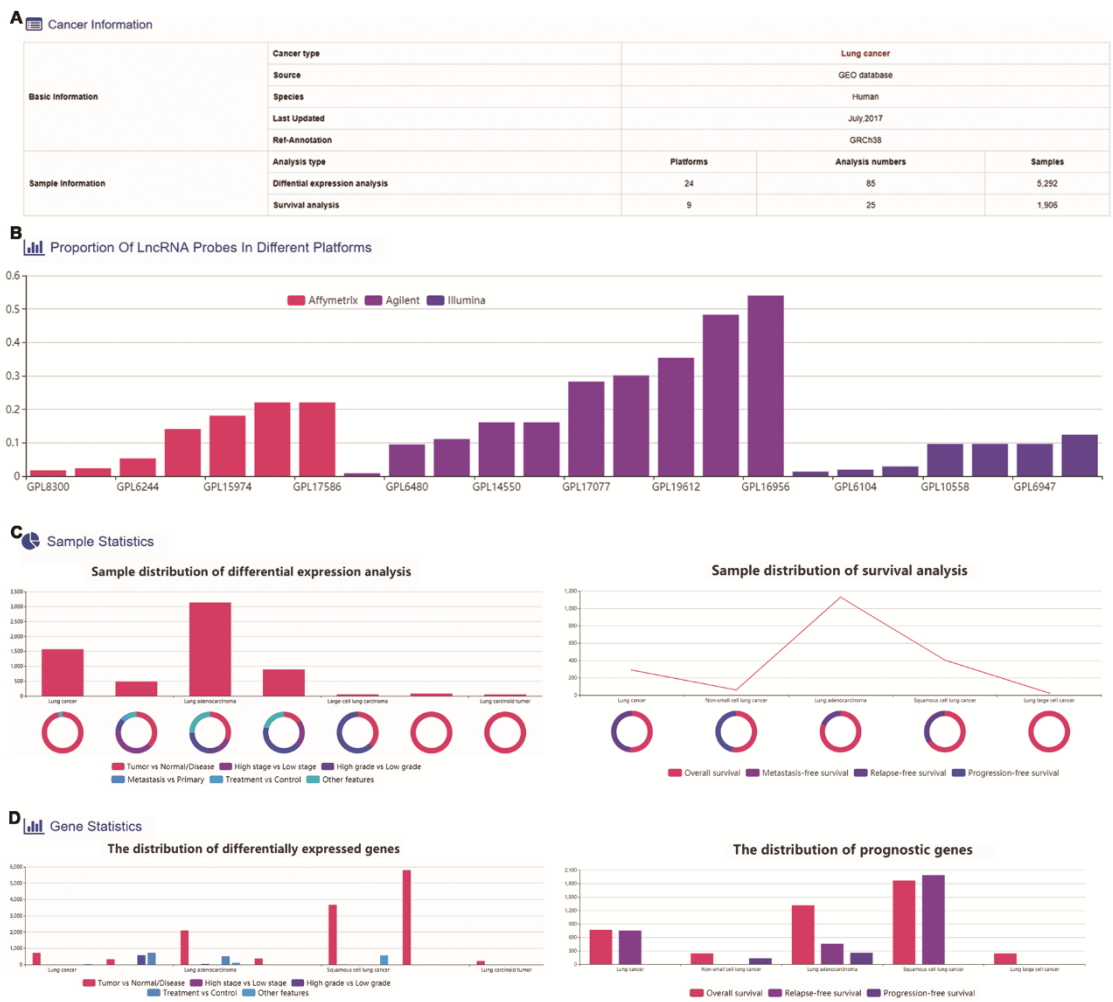


Fig. S7 - A snapshot of the browse interface. Take Lung cancer as an example. (A) Summary information of the cancer. (B) The proportion of lncRNA probes in different platforms. (C) The sample distribution of differential expression analysis and survival analysis. (D) The distribution of differentially expressed genes and prognostic genes.

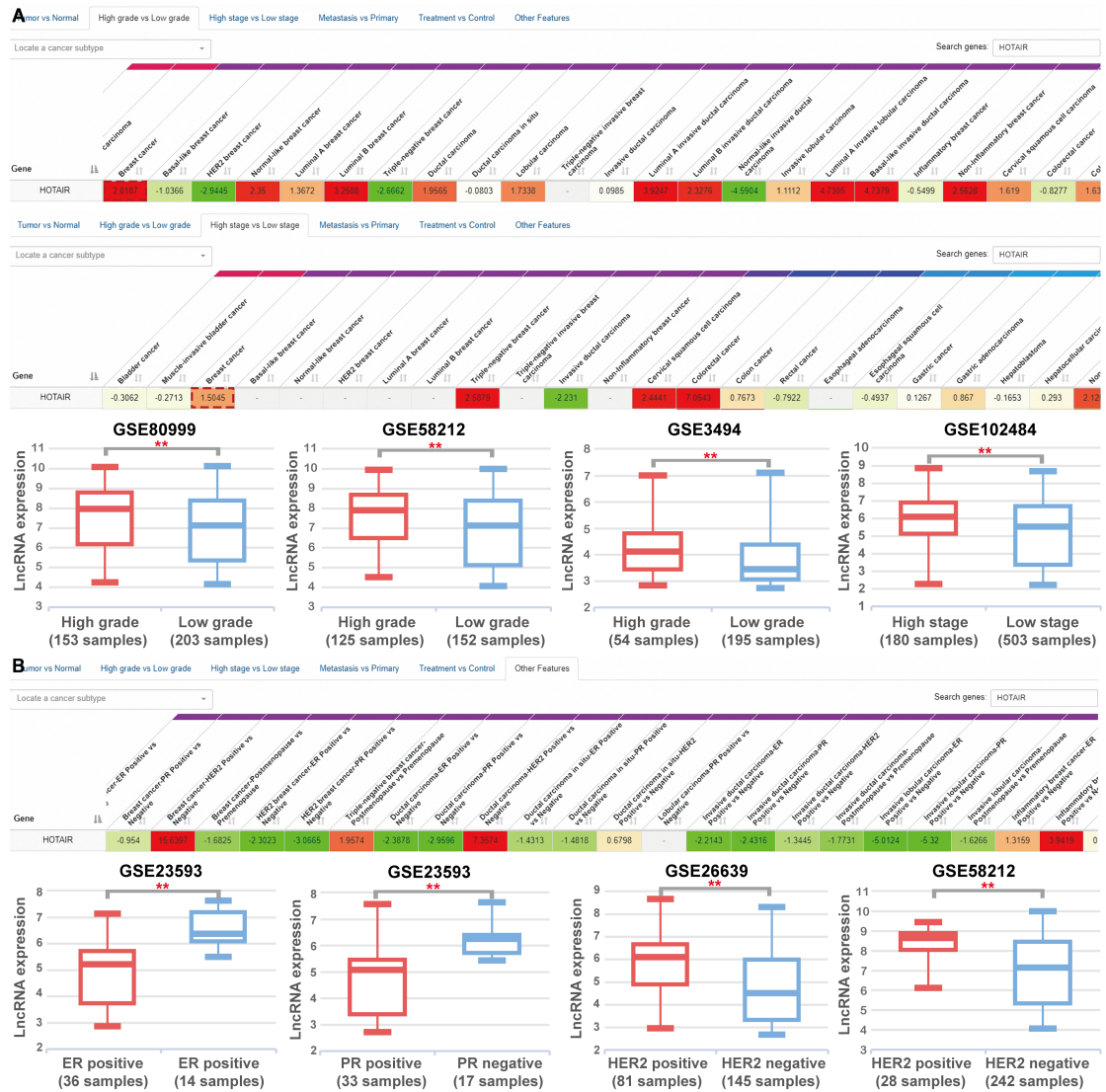


Fig S8 – Differential expression of lncRNA *HOTAIR* between different breast cancer clinical factors. (A) Increased *HOTAIR* expression is related to advanced tumor stage and higher tumor grade (B) *HOTAIR* appears to be overexpressed in ER-negative, PR-negative and HER2-positive patients.

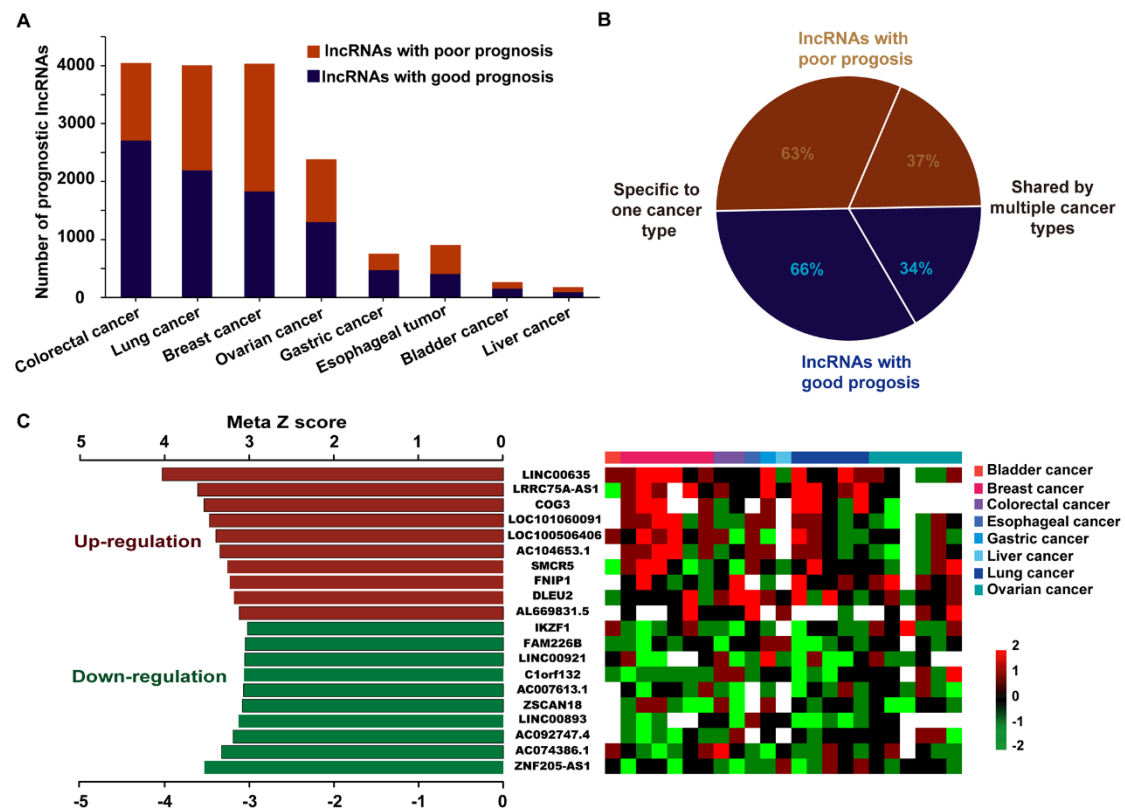


Fig. S9 - A large number of lncRNAs with potential clinical significance in various cancer types. (A) Bar plot showing the number of prognostic lncRNAs in different cancer types. (B) The proportion of prognostic lncRNAs specific to one cancer type and shared by multiple cancer types for poor prognosis and good prognosis. (C) The top 10 lncRNAs that were significantly associated with poor prognosis or good prognosis. Left panel is showing the Meta Z score of the P values from the survival analysis of different datasets of different cancer types. Right panel is showing the P values from the survival analysis of different datasets.

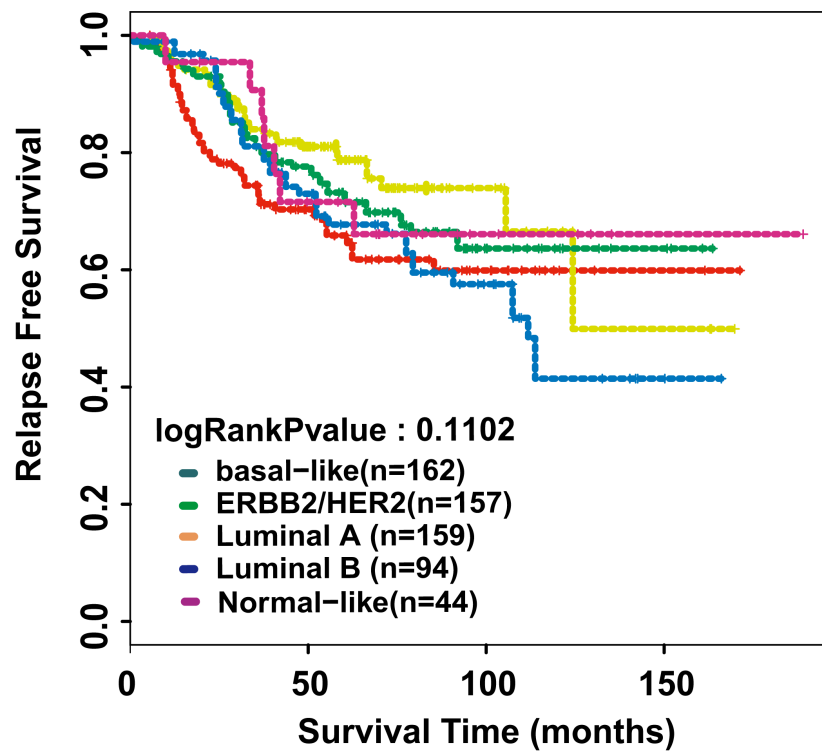


Fig. S10 Relapse-free survival curves for four groups defined by known molecular subtypes of breast cancer

Table S1 Search keywords for different cancer types

Cancer	Search keywords
Bladder cancer	(bladder cancer) OR (urothelial carcinoma of the bladder) OR (transitional cell carcinoma of urinary bladder) OR (Squamous cell carcinoma of urinary bladder) OR (adenocarcinoma of urinary bladder) OR (Small-cell carcinoma of urinary bladder) OR (bladder sarcoma)
Breast cancer	(breast cancer) OR (breast carcinoma) OR (breast ductal carcinoma)
Cervical cancer	(Uterine cancer) OR (cervical cancer) OR (endometrial carcinoma) OR (endometrial stromal sarcoma) OR (Squamous Cell Carcinoma of the Cervix) OR (Cervical Adenocarcinoma) OR (adenosquamous carcinoma of the cervix) OR (Small cell cervical cancer) OR (Small cell neuroendocrine cervical carcinoma) OR (Glass cell carcinoma of the cervix) OR (Villoglandular adenocarcinoma of the cervix)
Colorectal cancer	(rectum cancer) OR (rectal carcinoma) OR (carcinoma of rectum) OR (rectal cancer)OR (colon cancer) OR (colorectal cancer) OR (colorectal carcinoma) OR (carcinoma of colon)
Esophagus cancer	(esophageal cancer) OR (esophageal adenocarcinoma) OR (esophageal squamous cell carcinoma)
Gastric cancer	(gastric cancer) OR (Stomach cancer) OR (gastric adenocarcinoma) OR (gastrointestinal stromal tumor)
Liver cancer	(Hepatocellular carcinoma) OR (Liver cancer) OR (hepatic cancer) OR (intrahepatic cholangiocarcinoma) OR (Hepatoblastoma) OR (hepatoma)
Lung cancer	(lung carcinoma) OR (small-cell lung carcinoma) OR (non-small-cell lung carcinoma) OR (Lung cancer) OR (lung Adenocarcinoma) OR (lung tumor) OR (lung squamous cell carcinoma)
ovarian cancer	(ovarian cancer) OR (ovarian carcinoma) OR (ovarian epithelial carcinoma) OR (Malignant mixed müllerian tumor) OR (Mucinous tumors) OR (ovarian epithelial carcinoma) OR (Serous ovarian carcinoma) OR (Clear-cell ovarian carcinomas) OR (Clear-cell ovarian adenocarcinomas) OR (ovarian squamous cell carcinomas) OR (ovarian borderline tumor)
prostate cancer	(prostate cancer) OR (prostate carcinoma)

Table S2 Clustering datasets with known cancer subtypes of colorectal cancer, bladder cancer and breast cancer

[illegible]

[illegible]

[illegible]

Breast Cancer	GSM781122	Luminal B	Subtype 1	1	721
Breast Cancer	GSM781123	Luminal B	Subtype 1	0	1249
Breast Cancer	GSM781124	Luminal B	Subtype 1	0	4277
Breast Cancer	GSM781130	Luminal B	Subtype 1	0	1289
Breast Cancer	GSM781131	Luminal B	Subtype 1	0	4062
Breast Cancer	GSM781142	Luminal B	Subtype 1	0	4062
Breast Cancer	GSM781144	Luminal B	Subtype 1	0	3979
Breast Cancer	GSM781151	Luminal B	Subtype 1	0	4206
Breast Cancer	GSM781156	Luminal B	Subtype 1	0	2961
Breast Cancer	GSM781162	Luminal B	Subtype 1	0	1336
Breast Cancer	GSM781163	Luminal B	Subtype 1	0	4311
Breast Cancer	GSM781176	Luminal B	Subtype 1	1	1569
Breast Cancer	GSM781179	Luminal B	Subtype 1	1	785
Breast Cancer	GSM781182	Luminal B	Subtype 1	1	945
Breast Cancer	GSM781196	Luminal B	Subtype 2	0	931
Breast Cancer	GSM781410	Luminal B	Subtype 1	1	851
Breast Cancer	GSM781411	Luminal B	Subtype 2	1	722
Breast Cancer	GSM781433	Luminal B	Subtype 4	1	1178
Breast Cancer	GSM781435	Luminal B	Subtype 4	0	3862
Breast Cancer	GSM781440	Luminal B	Subtype 2	0	2249
Breast Cancer	GSM781445	Luminal B	Subtype 2	0	1544
Breast Cancer	GSM781446	Luminal B	Subtype 2	0	2564
Breast Cancer	GSM781480	Luminal B	Subtype 4	0	1646
Breast Cancer	GSM781481	Luminal B	Subtype 4	0	1646
Breast Cancer	GSM781484	Luminal B	Subtype 4	0	1531
Breast Cancer	GSM781489	Luminal B	Subtype 1	0	1774
Breast Cancer	GSM781492	Luminal B	Subtype 1	0	679
Breast Cancer	GSM781497	Luminal B	Subtype 2	0	1641
Breast Cancer	GSM781510	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781511	Luminal B	Subtype 2	0	1774
Breast Cancer	GSM781515	Luminal B	Subtype 2	0	1774
Breast Cancer	GSM781518	Luminal B	Subtype 2	0	1774
Breast Cancer	GSM781523	Luminal B	Subtype 1	0	588
Breast Cancer	GSM781526	Luminal B	Subtype 1	0	588
Breast Cancer	GSM781532	Luminal B	Subtype 4	0	3418
Breast Cancer	GSM781538	Luminal B	Subtype 2	1	1126
Breast Cancer	GSM781544	Luminal B	Subtype 2	1	3223
Breast Cancer	GSM781548	Luminal B	Subtype 2	1	1278
Breast Cancer	GSM781551	Luminal B	Subtype 1	1	2573
Breast Cancer	GSM781555	Luminal B	Subtype 4	1	373
Breast Cancer	GSM781739	Luminal B	Subtype 2	0	3866
Breast Cancer	GSM781770	Luminal B	Subtype 2	0	3219
Breast Cancer	GSM781771	Luminal B	Subtype 2	0	3296
Breast Cancer	GSM781772	Luminal B	Subtype 2	0	3278
Breast Cancer	GSM781773	Luminal B	Subtype 2	0	2779
Breast Cancer	GSM781775	Luminal B	Subtype 2	1	1662
Breast Cancer	GSM781786	Luminal B	Subtype 1	0	3227
Breast Cancer	GSM781799	Luminal B	Subtype 2	0	2931
Breast Cancer	GSM781842	Luminal B	Subtype 2	0	2068
Breast Cancer	GSM781943	Luminal B	Subtype 2	0	2716
Breast Cancer	GSM781944	Luminal B	Subtype 4	0	1613
Breast Cancer	GSM7819746	Luminal B	Subtype 4	1	748
Breast Cancer	GSM7819750	Luminal B	Subtype 2	0	3862
Breast Cancer	GSM7819752	Luminal B	Subtype 1	0	111
Breast Cancer	GSM7819754	Luminal B	Subtype 1	0	1424
Breast Cancer	GSM7819758	Luminal B	Subtype 4	0	2449
Breast Cancer	GSM7819763	Luminal B	Subtype 4	0	1672
Breast Cancer	GSM7819769	Luminal B	Subtype 4	0	1672
Breast Cancer	GSM7819771	Luminal B	Subtype 2	1	2723
Breast Cancer	GSM7819773	Luminal B	Subtype 1	1	617
Breast Cancer	GSM7819783	Luminal B	Subtype 4	1	2161
Breast Cancer	GSM7819786	Luminal B	Subtype 2	1	3354
Breast Cancer	GSM781809	Luminal B	Subtype 1	1	721
Breast Cancer	GSM781810	Luminal B	Subtype 1	0	1249
Breast Cancer	GSM781816	Luminal B	Subtype 4	0	4277
Breast Cancer	GSM781820	Luminal B	Subtype 4	0	4062
Breast Cancer	GSM781828	Luminal B	Subtype 4	0	3979
Breast Cancer	GSM781831	Luminal B	Subtype 4	0	4206
Breast Cancer	GSM781839	Luminal B	Subtype 4	0	1336
Breast Cancer	GSM781842	Luminal B	Subtype 4	1	2961
Breast Cancer	GSM781849	Luminal B	Subtype 4	0	1531
Breast Cancer	GSM781852	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781859	Luminal B	Subtype 4	0	679
Breast Cancer	GSM781869	Luminal B	Subtype 2	0	1641
Breast Cancer	GSM781870	Luminal B	Subtype 2	0	588
Breast Cancer	GSM781806	Luminal B	Subtype 2	0	1774
Breast Cancer	GSM781811	Luminal B	Subtype 2	0	1774
Breast Cancer	GSM781816	Luminal B	Subtype 1	0	588
Breast Cancer	GSM781820	Luminal B	Subtype 4	0	4277
Breast Cancer	GSM781826	Luminal B	Subtype 2	1	3413
Breast Cancer	GSM781832	Luminal B	Subtype 2	1	1126
Breast Cancer	GSM781833	Luminal B	Subtype 2	1	3223
Breast Cancer	GSM781836	Luminal B	Subtype 2	1	2162
Breast Cancer	GSM781841	Luminal B	Subtype 4	0	1278
Breast Cancer	GSM781842	Luminal B	Subtype 4	0	3474
Breast Cancer	GSM781843	Luminal B	Subtype 4	1	373
Breast Cancer	GSM781848	Luminal B	Subtype 1	0	1249
Breast Cancer	GSM781853	Luminal B	Subtype 4	0	4277
Breast Cancer	GSM781856	Luminal B	Subtype 4	0	4062
Breast Cancer	GSM781858	Luminal B	Subtype 4	0	3979
Breast Cancer	GSM781861	Luminal B	Subtype 4	0	4206
Breast Cancer	GSM781862	Luminal B	Subtype 4	0	1336
Breast Cancer	GSM781863	Luminal B	Subtype 4	0	2961
Breast Cancer	GSM781864	Luminal B	Subtype 4	0	1531
Breast Cancer	GSM781865	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781866	Luminal B	Subtype 4	0	679
Breast Cancer	GSM781867	Luminal B	Subtype 4	0	1641
Breast Cancer	GSM781868	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781869	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781870	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781871	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781872	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781873	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781874	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781875	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781876	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781877	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781878	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781879	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781880	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781881	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781882	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781883	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781884	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781885	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781886	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781887	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781888	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781889	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781890	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781891	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781892	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781893	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781894	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781895	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781896	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781897	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781898	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781899	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781900	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781901	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781902	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781903	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781904	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781905	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781906	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781907	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781908	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781909	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781910	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781911	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781912	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781913	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781914	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781915	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781916	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781917	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781918	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781919	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781920	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781921	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781922	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781923	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781924	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781925	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781926	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781927	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781928	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781929	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781930	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781931	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781932	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781933	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781934	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781935	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781936	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781937	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781938	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781939	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781940	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781941	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781942	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781943	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781944	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781945	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781946	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781947	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781948	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781949	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781950	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781951	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781952	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781953	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781954	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781955	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781956	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781957	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781958	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781959	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781960	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781961	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781962	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781963	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781964	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781965	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781966	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781967	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781968	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781969	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781970	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781971	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781972	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781973	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781974	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781975	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781976	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781977	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781978	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781979	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781980	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781981	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781982	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781983	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781984	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781985	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781986	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781987	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781988	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781989	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781990	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM781991	Luminal B	Subtype 4	0	1774
Breast Cancer	GSM7				

Breast Cancer	GSM761441	Luminal A	Subtype 2	0	2026
Breast Cancer	GSM761442	Luminal A	Subtype 2	0	1541
Breast Cancer	GSM761443	Luminal A	Subtype 2	0	2836
Breast Cancer	GSM761444	Luminal A	Subtype 2	0	1814
Breast Cancer	GSM761445	Luminal A	Subtype 2	0	2663
Breast Cancer	GSM761446	Luminal A	Subtype 2	0	3008
Breast Cancer	GSM761447	Luminal A	Subtype 2	0	2251
Breast Cancer	GSM761448	Luminal A	Subtype 2	0	2832
Breast Cancer	GSM761449	Luminal A	Subtype 2	0	1201
Breast Cancer	GSM761450	Luminal A	Subtype 2	1	1601
Breast Cancer	GSM761451	Luminal A	Subtype 2	1	1107
Breast Cancer	GSM761452	Luminal A	Subtype 2	0	1306
Breast Cancer	GSM761453	Luminal A	Subtype 2	0	2010
Breast Cancer	GSM761454	Luminal A	Subtype 2	0	1619
Breast Cancer	GSM761455	Luminal A	Subtype 2	0	1786
Breast Cancer	GSM761456	Luminal A	Subtype 2	0	1759
Breast Cancer	GSM761457	Luminal A	Subtype 2	0	1288
Breast Cancer	GSM761458	Luminal A	Subtype 2	0	1063
Breast Cancer	GSM761459	Luminal A	Subtype 2	0	
Breast Cancer	GSM761460	Luminal A	Subtype 2	0	
Breast Cancer	GSM761461	Luminal A	Subtype 2	0	221
Breast Cancer	GSM761462	Luminal A	Subtype 2	1	985
Breast Cancer	GSM761463	Luminal A	Subtype 2	1	842
Breast Cancer	GSM761464	Luminal A	Subtype 2	1	2758
Breast Cancer	GSM761465	Luminal A	Subtype 2	1	1607
Breast Cancer	GSM761466	Luminal A	Subtype 2	1	799
Breast Cancer	GSM761467	Luminal A	Subtype 2	1	864
Breast Cancer	GSM761468	Luminal A	Subtype 2	1	1526
Breast Cancer	GSM761469	Luminal A	Subtype 2	1	1221
Breast Cancer	GSM761470	Luminal A	Subtype 4	1	822
Breast Cancer	GSM761471	Luminal A	Subtype 4	1	1311
Breast Cancer	GSM761472	Luminal A	Subtype 1	1	101
Breast Cancer	GSM761473	Luminal A	Subtype 1	1	923
Breast Cancer	GSM761474	Luminal A	Subtype 2	0	2216
Breast Cancer	GSM761475	Luminal A	Subtype 2	0	2429
Breast Cancer	GSM761476	Luminal A	Subtype 2	0	2172
Breast Cancer	GSM761477	Luminal A	Subtype 2	0	2151
Breast Cancer	GSM761478	Luminal A	Subtype 2	0	2039
Breast Cancer	GSM761479	Luminal A	Subtype 2	0	2639
Breast Cancer	GSM761480	Luminal A	Subtype 2	0	2639
Breast Cancer	GSM761481	Luminal A	Subtype 2	0	1956
Breast Cancer	GSM761482	Luminal A	Subtype 2	0	2365
Breast Cancer	GSM761483	Luminal A	Subtype 2	0	3646
Breast Cancer	GSM761484	Luminal A	Subtype 2	0	1389
Breast Cancer	GSM761485	Luminal A	Subtype 2	0	2230
Breast Cancer	GSM761486	Luminal A	Subtype 2	0	4
Breast Cancer	GSM761487	Luminal A	Subtype 2	1	899
Breast Cancer	GSM761488	Luminal A	Subtype 2	1	1623
Breast Cancer	GSM761489	Luminal A	Subtype 4	1	442
Breast Cancer	GSM761490	Luminal A	Subtype 4	1	913
Breast Cancer	GSM761491	Luminal A	Subtype 2	0	4839
Breast Cancer	GSM761492	Luminal A	Subtype 2	0	4523
Breast Cancer	GSM761493	Luminal A	Subtype 2	0	4023
Breast Cancer	GSM761494	Luminal A	Subtype 2	0	2152
Breast Cancer	GSM761495	Luminal A	Subtype 2	0	2108
Breast Cancer	GSM761496	Luminal A	Subtype 2	0	2108
Breast Cancer	GSM761497	Luminal A	Subtype 2	0	3931
Breast Cancer	GSM761498	Luminal A	Subtype 2	0	2608
Breast Cancer	GSM761499	Luminal A	Subtype 2	1	1087
Breast Cancer	GSM761500	Luminal A	Subtype 2	1	122
Breast Cancer	GSM761501	Luminal A	Subtype 2	0	2587
Breast Cancer	GSM761502	Luminal A	Subtype 2	0	839
Breast Cancer	GSM761503	Luminal A	Subtype 2	0	3644
Breast Cancer	GSM761504	Luminal A	Subtype 2	0	3623
Breast Cancer	GSM761505	Luminal A	Subtype 2	0	1274
Breast Cancer	GSM761506	Luminal A	Subtype 2	0	2514
Breast Cancer	GSM761507	Luminal A	Subtype 2	1	1663
Breast Cancer	GSM761508	Luminal A	Subtype 2	1	3668
Breast Cancer	GSM761509	Luminal A	Subtype 2	0	3578
Breast Cancer	GSM761510	Luminal A	Subtype 2	0	642
Breast Cancer	GSM761511	Luminal A	Subtype 2	0	2237
Breast Cancer	GSM761512	Luminal A	Subtype 2	0	2023
Breast Cancer	GSM761513	Luminal A	Subtype 2	0	2621
Breast Cancer	GSM761514	Luminal A	Subtype 2	0	2692
Breast Cancer	GSM761515	Luminal A	Subtype 2	0	2972
Breast Cancer	GSM761516	Luminal A	Subtype 2	1	1053
Breast Cancer	GSM761517	Luminal A	Subtype 2	0	862
Breast Cancer	GSM761518	Luminal A	Subtype 2	0	1999
Breast Cancer	GSM761519	Luminal A	Subtype 2	0	2117
Breast Cancer	GSM761520	Luminal A	Subtype 2	0	3346
Breast Cancer	GSM761521	Luminal A	Subtype 2	0	757
Breast Cancer	GSM761522	Luminal A	Subtype 2	0	3004
Breast Cancer	GSM761523	Luminal A	Subtype 2	0	2668
Breast Cancer	GSM761524	Luminal A	Subtype 2	0	1682
Breast Cancer	GSM761525	Luminal A	Subtype 2	0	2112
Breast Cancer	GSM761526	Luminal A	Subtype 2	0	2926
Breast Cancer	GSM761527	Luminal A	Subtype 2	0	1511
Breast Cancer	GSM761528	Luminal A	Subtype 2	0	2830
Breast Cancer	GSM761529	Luminal A	Subtype 2	0	1414
Breast Cancer	GSM761530	Luminal A	Subtype 2	0	2063
Breast Cancer	GSM761531	Luminal A	Subtype 2	0	1010
Breast Cancer	GSM761532	Luminal A	Subtype 2	0	368
Breast Cancer	GSM761533	Luminal A	Subtype 2	0	2988
Breast Cancer	GSM761534	Luminal A	Subtype 2	0	2623
Breast Cancer	GSM761535	Luminal A	Subtype 2	1	2281
Breast Cancer	GSM761536	Luminal A	Subtype 2	1	1693
Breast Cancer	GSM761537	Luminal A	Subtype 1	1	1107
Breast Cancer	GSM761538	Luminal A	Subtype 1	1	1396
Breast Cancer	GSM761539	Luminal A	Subtype 2	0	2010
Breast Cancer	GSM761540	Luminal A	Subtype 2	0	1619
Breast Cancer	GSM761541	Luminal A	Subtype 2	0	1786
Breast Cancer	GSM761542	Luminal A	Subtype 2	0	1784
Breast Cancer	GSM761543	Luminal A	Subtype 1	0	1759
Breast Cancer	GSM761544	Luminal A	Subtype 1	0	1288
Breast Cancer	GSM761545	Luminal A	Subtype 2	0	1063
Breast Cancer	GSM761546	Luminal A	Subtype 2	0	
Breast Cancer	GSM761547	Luminal A	Subtype 2	0	
Breast Cancer	GSM761548	Luminal A	Subtype 2	0	221
Breast Cancer	GSM761549	Luminal A	Subtype 4	1	985
Breast Cancer	GSM761550	Luminal A	Subtype 4	1	842
Breast Cancer	GSM761551	Luminal A	Subtype 4	1	2758
Breast Cancer	GSM761552	Luminal A	Subtype 4	1	1607
Breast Cancer	GSM761553	Luminal A	Subtype 4	1	799
Breast Cancer	GSM761554	Luminal A	Subtype 4	1	864
Breast Cancer	GSM761555	Luminal A	Subtype 4	1	1526
Breast Cancer	GSM761556	Luminal A	Subtype 4	1	1221
Breast Cancer	GSM761557	Luminal A	Subtype 4	1	822
Breast Cancer	GSM761558	Luminal A	Subtype 4	1	1311
Breast Cancer	GSM761559	Luminal A	Subtype 1	1	101
Breast Cancer	GSM761560	Luminal A	Subtype 1	1	923
Breast Cancer	GSM761561	Luminal A	Subtype 2	0	2216
Breast Cancer	GSM761562	Luminal A	Subtype 2	0	2429
Breast Cancer	GSM761563	Luminal A	Subtype 2	0	2172
Breast Cancer	GSM761564	Luminal A	Subtype 2	0	2151
Breast Cancer	GSM761565	Luminal A	Subtype 2	0	2039
Breast Cancer	GSM761566	Luminal A	Subtype 2	0	2639
Breast Cancer	GSM761567	Luminal A	Subtype 2	0	2639
Breast Cancer	GSM761568	Luminal A	Subtype 2	0	1956
Breast Cancer	GSM761569	Luminal A	Subtype 2	0	2365
Breast Cancer	GSM761570	Luminal A	Subtype 2	0	3646
Breast Cancer	GSM761571	Luminal A	Subtype 2	0	1389
Breast Cancer	GSM761572	Luminal A	Subtype 2	0	2230
Breast Cancer	GSM761573	Luminal A	Subtype 2	0	4
Breast Cancer	GSM761574	Luminal A	Subtype 2	1	899
Breast Cancer	GSM761575	Luminal A	Subtype 2	1	1623
Breast Cancer	GSM761576	Luminal A	Subtype 4	1	442
Breast Cancer	GSM761577	Luminal A	Subtype 4	1	913
Breast Cancer	GSM761578	Luminal A	Subtype 2	0	4839
Breast Cancer	GSM761579	Luminal A	Subtype 2	0	4523
Breast Cancer	GSM761580	Luminal A	Subtype 2	0	4023
Breast Cancer	GSM761581	Luminal A	Subtype 2	0	2152
Breast Cancer	GSM761582	Luminal A	Subtype 2	0	2108
Breast Cancer	GSM761583	Luminal A	Subtype 2	0	2108
Breast Cancer	GSM761584	Luminal A	Subtype 2	0	3931
Breast Cancer	GSM761585	Luminal A	Subtype 2	0	2608
Breast Cancer	GSM761586	Luminal A	Subtype 2	1	1087
Breast Cancer	GSM761587	Luminal A	Subtype 2	1	122
Breast Cancer	GSM761588	Luminal A	Subtype 2	0	2587
Breast Cancer	GSM761589	Luminal A	Subtype 2	0	839
Breast Cancer	GSM761590	Luminal A	Subtype 2	0	3644
Breast Cancer	GSM761591	Luminal A	Subtype 2	0	3623
Breast Cancer	GSM761592	Luminal A	Subtype 2	0	1274
Breast Cancer	GSM761593	Luminal A	Subtype 2	0	2514
Breast Cancer	GSM761594	Luminal A	Subtype 2	1	1663
Breast Cancer	GSM761595	Luminal A	Subtype 2	1	3668
Breast Cancer	GSM761596	Luminal A	Subtype 2	0	3578
Breast Cancer	GSM761597	Luminal A	Subtype 2	0	642
Breast Cancer	GSM761598	Luminal A	Subtype 2	0	2237
Breast Cancer	GSM761599	Luminal A	Subtype 2	0	2023
Breast Cancer	GSM761600	Luminal A	Subtype 2	0	2621
Breast Cancer	GSM761601	Luminal A	Subtype 2	0	2692
Breast Cancer	GSM761602	Luminal A	Subtype 2	0	2972
Breast Cancer	GSM761603	Luminal A	Subtype 2	1	1053
Breast Cancer	GSM761604	Luminal A	Subtype 2	0	862
Breast Cancer	GSM761605	Luminal A	Subtype 2	0	1999
Breast Cancer	GSM761606	Luminal A	Subtype 2	0	2117
Breast Cancer	GSM761607	Luminal A	Subtype 2	0	3346
Breast Cancer	GSM761608	Luminal A	Subtype 2	0	757
Breast Cancer	GSM761609	Luminal A	Subtype 2	0	3004
Breast Cancer	GSM761610	Luminal A	Subtype 2	0	2668
Breast Cancer	GSM761611	Luminal A	Subtype 2	0	1682
Breast Cancer	GSM761612	Luminal A	Subtype 2	0	2112
Breast Cancer	GSM761613	Luminal A	Subtype 2	0	2926
Breast Cancer	GSM761614	Luminal A	Subtype 2	0	1511
Breast Cancer	GSM761615	Luminal A	Subtype 2	0	2830
Breast Cancer	GSM761616	Luminal A	Subtype 2	0	1414
Breast Cancer	GSM761617	Luminal A	Subtype 2	0	2063
Breast Cancer	GSM761618	Luminal A	Subtype 2	0	1010
Breast Cancer	GSM761619	Luminal A	Subtype 2	0	368
Breast Cancer	GSM761620	Luminal A	Subtype 2	0	2988
Breast Cancer	GSM761621	Luminal A	Subtype 2	0	2623
Breast Cancer	GSM761622	Luminal A	Subtype 2	1	2281
Breast Cancer	GSM761623	Luminal A	Subtype 2	1	1693
Breast Cancer	GSM761624	Luminal A	Subtype 1	1	1107
Breast Cancer	GSM761625	Luminal A	Subtype 1	1	1396
Breast Cancer	GSM761626	Luminal A	Subtype 2	0	2010
Breast Cancer	GSM761627	Luminal A	Subtype 2	0	1619
Breast Cancer	GSM761628	Luminal A	Subtype 2	0	1786
Breast Cancer	GSM761629	Luminal A	Subtype 2	0	1784
Breast Cancer	GSM761630	Luminal A	Subtype 1	0	1759
Breast Cancer	GSM761631	Luminal A	Subtype 1	0	1288
Breast Cancer	GSM761632	Luminal A	Subtype 2	0	1063
Breast Cancer	GSM761633	Luminal A	Subtype 2	0	
Breast Cancer	GSM761634	Luminal A	Subtype 2	0	
Breast Cancer	GSM761635	Luminal A	Subtype 2	0	221
Breast Cancer	GSM761636	Luminal A	Subtype 4	1	985
Breast Cancer	GSM761637	Luminal A	Subtype 4	1	842
Breast Cancer	GSM761638	Luminal A	Subtype 4	1	2758
Breast Cancer	GSM761639	Luminal A	Subtype 4	1	1607
Breast Cancer	GSM761640	Luminal A	Subtype 4	1	799
Breast Cancer	GSM761641	Luminal A	Subtype 4	1	864
Breast Cancer	GSM761642	Luminal A	Subtype 4	1	1526
Breast Cancer	GSM761643	Luminal A	Subtype 4	1	1221
Breast Cancer	GSM761644	Luminal A	Subtype 4	1	822
Breast Cancer	GSM761645	Luminal A	Subtype 4	1	1311
Breast Cancer	GSM761646	Luminal A	Subtype 1	1	101
Breast Cancer	GSM761647	Luminal A	Subtype 1	1	923
Breast Cancer	GSM761648	Luminal A	Subtype 2	0	2216
Breast Cancer	GSM761649	Luminal A	Subtype 2	0	2429
Breast Cancer	GSM761650	Luminal A	Subtype 2	0	2172
Breast Cancer	GSM761651	Luminal A	Subtype 2	0	2151
Breast Cancer	GSM761652	Luminal A	Subtype 2	0	2039
Breast Cancer	GSM761653	Luminal A	Subtype 2	0	2639
Breast Cancer	GSM761654	Luminal A	Subtype 2	0	2639
Breast Cancer	GSM761655	Luminal A	Subtype 2	0	1956
Breast Cancer	GSM761656	Luminal A	Subtype 2	0	2365
Breast Cancer	GSM761657	Luminal A	Subtype 2	0	3646
Breast Cancer	GSM761658	Luminal A	Subtype 2	0	1389
Breast Cancer	GSM761659	Luminal A	Subtype 2	0	2230
Breast Cancer	GSM761660	Luminal A	Subtype 2	0	4
Breast Cancer	GSM761661	Luminal A	Subtype 2	1	899
Breast Cancer	GSM761662	Luminal A	Subtype 2	1	1623
Breast Cancer	GSM761663	Luminal A			

[illegible]

Table S3 Summary of the re-annotated probes in different platforms

Platform	Platform type	Platform title	coding gene probes	coding genes	lncRNA probes	lncRNAs
GPL11097	Affymetrix	[Glyco_v4_Hs] Custom Affymetrix Glyco v4 GeneChip	4,774	1,003	21	9
GPL201	Affymetrix	[HG-Focus] Affymetrix Human HG-Focus Target Array	8,455	8,560	41	46
GPL91	Affymetrix	[HG_U95A] Affymetrix Human Genome U95A Array	11,224	8,984	199	188
GPL8300	Affymetrix	[HG_U95Av2] Affymetrix Human Genome U95 Version 2 Array	11,224	8,985	199	188
GPL11060	Affymetrix	[CG_AGPas20460F] Affymetrix Adenoma Biomarker 46K Gene	3,763	1,789	255	218
GPL96	Affymetrix	[HG_U133A] Affymetrix Human Genome U133A Array	19,091	12,762	462	454
GPL3921	Affymetrix	[HT_HG-U133A] Affymetrix HT Human Genome U133A Array	19,091	12,762	462	454
GPL571	Affymetrix	[HG-U133A_2] Affymetrix Human Genome U133A 2.0 Array	19,092	12,768	462	454
GPL10553	Affymetrix	Affymetrix GeneChip Human Genome HG-U133A Custom CDF	19,091	12,762	462	454
GPL13667	Affymetrix	[HG-U219] Affymetrix Human Genome U219 Array	46,022	19,065	1,025	866
GPL15207	Affymetrix	[PrimeView] Affymetrix Human Gene Expression Array	46,022	19,065	1,025	866
GPL16043	Affymetrix	GeneChip® PrimeView™ Human Gene Expression Array (with E	46,022	19,065	1,025	866
GPL92	Affymetrix	[HG_U95B] Affymetrix Human Genome U95B Array	7,794	6,064	1,099	1,066
GPL93	Affymetrix	[HG_U95C] Affymetrix Human Genome U95C Array	5,755	4,528	1,480	1,393
GPL97	Affymetrix	[HG-U133B] Affymetrix Human Genome U133B Array	12,055	8,234	2,654	2,472
GPL16786	Affymetrix	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [HuGene	65,681	17,987	3,666	2,188
GPL14010	Affymetrix	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [hugene1	65,681	17,987	3,666	2,188
GPL10739	Affymetrix	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [probe se	65,681	17,987	3,666	2,188
GPL6244	Affymetrix	[HuGene-1_0-st] Affymetrix Human Gene 1.0 ST Array [transcrip	65,681	17,987	3,666	2,188
GPL15676	Affymetrix	[HG-U133_Plus_2] Affymetrix GeneChip Human Genome U133	34,728	18,174	5,701	4,987
GPL570	Affymetrix	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Ai	34,728	18,174	5,701	4,987
GPL13916	Affymetrix	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Ai	34,728	18,174	5,701	4,987
GPL16311	Affymetrix	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Ai	34,728	18,174	5,701	4,987
GPL16356	Affymetrix	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Ai	34,728	18,174	5,701	4,987
GPL11084	Affymetrix	Affymetrix GeneChip Human Genome U133 Plus 2.0 Array [CDF	34,728	18,174	5,701	4,987
GPL6791	Affymetrix	Affymetrix GeneChip Human Genome U133 Plus 2.0 Array [CDF	34,728	18,174	5,701	4,987
GPL9419	Affymetrix	Affymetrix GeneChip Human Genome U133 Plus 2.0 Array [CDF	34,728	18,174	5,701	4,987
GPL18756	Affymetrix	Affymetrix HG-U133 PLUS_2[CDF: HGU133Plus2_Hs_ENTRI	34,728	18,174	5,701	4,987
GPL14877	Affymetrix	Affymetrix Human Genome U133 Plus 2.0 Array [Brainarray Ver	34,728	18,174	5,701	4,987
GPL9101	Affymetrix	Affymetrix GeneChip Human Genome U133 Plus 2.0 Array (MBI	34,728	18,174	5,701	4,987
GPL7869	Affymetrix	Affymetrix GeneChip Human Genome U133 Plus 2.0 Array [CDF	34,759	18,174	5,701	4,987
GPL9828	Affymetrix	[HG-U133_Plus_2] Affymetrix Human Genome U133 Plus 2.0 Ai	34,728	18,174	5,701	4,987
GPL13158	Affymetrix	[HT_HG-U133_Plus_PM] Affymetrix HT HG-U133+ PM Array I	34,590	18,172	5,721	5,009
GPL1352	Affymetrix	[U133_X3P] Affymetrix Human X3P Array	37,714	18,155	6,329	4,955
GPL22462	Affymetrix	[bMxHERV3b520850] Affymetrix HERV3 Array	10,693	4,025	9,045	5,321
GPL18990	Affymetrix	[Xcel] Affymetrix Human Almac Xcel Array	56,142	19,406	10,723	7,365
GPL17692	Affymetrix	[HuGene-2_1-st] Affymetrix Human Gene 2.1 ST Array [transcrip	38,843	9,776	32,403	12,652
GPL15236	Affymetrix	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [CDF: Brai	356,276	19,770	78,641	23,484
GPL10520	Affymetrix	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [HuEx-1_0_	356,276	19,770	78,641	23,484
GPL17279	Affymetrix	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [HuEx-1_0_	356,276	19,770	78,641	23,484
GPL5188	Affymetrix	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [probe set (	356,276	19,770	78,641	23,484
GPL22995	Affymetrix	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (	356,276	19,770	78,641	23,484
GPL5175	Affymetrix	[HuEx-1_0-st] Affymetrix Human Exon 1.0 ST Array [transcript (	356,276	19,770	78,641	23,484
GPL10264	Affymetrix	Affymetrix Human Exon 1.0 ST Array [CDF: HuEx_1_0_st_v2_n	356,276	19,770	78,641	23,484
GPL15974	Affymetrix	Human Exon 1.0 ST Array [CDF: Brainarray Version 9.0.1, HsE:	356,276	19,770	78,641	23,484
GPL9793	Affymetrix	Affymetrix Human Exon 1.0 ST Array [CDF: HuEx-1_0-st-v2,co	356,276	19,770	78,641	23,484
GPL17585	Affymetrix	[HTA-2_0] Affymetrix Human Transcriptome Array 2.0 [probe se	375,203	19,722	105,989	20,643
GPL17586	Affymetrix	[HTA-2_0] Affymetrix Human Transcriptome Array 2.0 [transcrip	375,203	19,722	105,989	20,643
GPL7260	Agilent	Agilent-011521 Human 1A Microarray G4110A (Probe Name ver	13,224	13,127	205	215
GPL885	Agilent	Agilent-011521 Human 1A Microarray G4110A (Feature Numbe	13,935	13,127	214	215
GPL9053	Agilent	Agilent-UNC-custom-4X44K	25,821	15,699	228	209
GPL7264	Agilent	Agilent-012097 Human 1A Microarray (V2) G4110B (Probe Nam	15,640	14,611	293	311
GPL887	Agilent	Agilent-012097 Human 1A Microarray (V2) G4110B (Feature Nu	15,640	14,611	293	311
GPL5325	Agilent	Agilent UNC Perou Lab Homo sapiens 1X44K Custom Array	30,069	16,773	1,065	977
GPL16280	Agilent	Agilent Homo sapiens 37K DiscoverPrint_19742	23,769	16,529	2,363	2,283
GPL18348	Agilent	Agilent-021441 NCode Human Long Non-coding RNA microarra	3,570	3,021	2,372	2,223
GPL16233	Agilent	Agilent_human_DiscoverPrint_15746	24,394	16,588	2,464	2,371
GPL6848	Agilent	Agilent-012391 Whole Human Genome Oligo Microarray G4112.	23,634	16,542	2,475	2,396
GPL1708	Agilent	Agilent-012391 Whole Human Genome Oligo Microarray G4112.	23,634	16,542	2,475	2,396
GPL6480	Agilent	Agilent-014850 Whole Human Genome Microarray 4x44K G4112.	23,634	16,542	2,475	2,396
GPL4133	Agilent	Agilent-014850 Whole Human Genome Microarray 4x44K G4112.	23,634	16,542	2,475	2,396
GPL10332	Agilent	Agilent-026652 Whole Human Genome Microarray 4x44K v2 (Fe	23,320	17,924	2,907	2,790
GPL13497	Agilent	Agilent-026652 Whole Human Genome Microarray 4x44K v2 (Pr	23,320	17,924	2,907	2,790
GPL21047	Agilent	Agilent-074348 Human LncRNA v6 4X180K [Probe Name Versio	2,563	2,526	3,010	2,862
GPL13607	Agilent	Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (Fea	23,721	17,950	4,548	3,276
GPL14550	Agilent	Agilent-028004 SurePrint G3 Human GE 8x60K Microarray (Prol	23,721	17,950	4,548	3,276
GPL15931	Agilent	Agilent-029949 Custom SurePrint G3 Human GE 8x60K Microar	23,960	17,970	4,554	3,278
GPL17077	Agilent	Agilent-039494 SurePrint G3 Human GE v2 8x60K Microarray 0.	24,198	18,084	9,533	6,816
GPL15314	Agilent	Arraystar Human LncRNA microarray V2.0 (Agilent_033010 Pro	22,980	18,462	9,623	8,790
GPL13825	Agilent	Arraystar Human LncRNA microarray V2.0 (Agilent-033010)	22,980	18,462	9,623	8,790
GPL13648	Agilent	Agilent-021412 nONCochip_1.0 021253	47,404	17,192	10,006	4,044
GPL16956	Agilent	Agilent-045997 Arraystar human lncRNA microarray V3 (Probe N	11,085	10,119	13,030	9,997
GPL19243	Agilent	Agilent SurePrint G3 CBC lncRNA+mRNA microarray V2.0 (cor	30,917	18,367	13,325	7,593
GPL18109	Agilent	Agilent-038314 CBC Homo sapiens lncRNA + mRNA microarray	30,917	18,367	13,325	7,593
GPL19748	Agilent	Agilent-038314 CBC Homo sapiens lncRNA + mRNA microarray	30,917	18,367	13,325	7,593
GPL21185	Agilent	Agilent-072363 SurePrint G3 Human GE v3 8x60K Microarray 0.	24,430	18,101	16,714	12,074
GPL19072	Agilent	Agilent-052909 CBC_lncRNAmRNA_V3 (Probe Name version)	25,195	18,093	21,544	12,465
GPL19612	Agilent	Agilent-062918 OE Human lncRNA Microarray V4.0 028004 (Pr	39,627	18,903	21,718	15,024
GPL17843	Agilent	Agilent-042818 Human lncRNA Microarray 8_24_v2 [Probe Nam	28,048	18,279	22,210	11,750
GPL20115	Agilent	Agilent-067406 Human CBC lncRNA + mRNA microarray V4.0 (	25,253	18,079	23,585	13,358
GPL6883	Illumina	Illumina HumanRef-8 v3.0 expression beadchip	21,703	16,829	300	311
GPL8432	Illumina	Illumina HumanRef-8 WG-DASL v3.0	21,703	16,829	300	311
GPL6104	Illumina	Illumina humanRef-8 v2.0 expression beadchip	19,444	16,510	386	406
GPL14951	Illumina	Illumina HumanHT-12 WG-DASL V4.0 R2 expression beadchip	24,699	17,496	736	670
GPL6098	Illumina	Illumina humanRef-8 v1.0 expression beadchip	19,071	16,811	1,186	1,267
GPL6947	Illumina	Illumina HumanHT-12 V3.0 expression beadchip	26,460	17,971	2,836	2,771
GPL17518	Illumina	Illumina HumanHT-12_V3_0_R1_Bioconductor_2.11	26,459	17,971	2,836	2,771
GPL6884	Illumina	Illumina HumanWG-6 v3.0 expression beadchip	26,460	17,971	2,836	2,771
GPL10558	Illumina	Illumina HumanHT-12 V4.0 expression beadchip	28,179	18,414	3,013	2,510
GPL6102	Illumina	Illumina human-6 v2.0 expression beadchip	22,021	17,585	3,120	3,083
GPL13376	Illumina	Illumina HumanWG-6 v2.0 expression beadchip	22,020	17,585	3,120	3,083
GPL6370	Illumina	Illumina human-6 v2.0 expression beadchip (extended)	22,020	17,585	3,120	3,083
GPL8068	Others	RIKEN BRC Human 11K Microarray Platform for Sense-AFAS T	444	437	23	22
GPL13128	Others	State Key Lab Homo sapien 2.6K	1,780	1,708	71	75
GPL21494	Others	State Key Laboratory Human lncRNA array 2412	311	268	651	668
GPL8070	Others	RIKEN BRC Human 11k Microarray Platform for Sense-Antisens	2,560	2,306	702	706
GPL3991	Others	Human 3.0 A1	18,161	15,831	753	779
GPL13287	Others	Phalanx Human OneArray [Annotation HOA5 release 1.0]	23,561	18,317	840	813
GPL5049	Others	MLRG Human 21K V12.0	13,322	12,518	1,187	1,216
GPL8926	Others	PC Human Operon 21k v2	13,324	12,520	1,187	1,216
GPL3676	Others	NKI-CMF Homo sapiens 35k oligo array	23,871	16,358	1,952	1,965
GPL13152	Others	Human SMD Print_1360 (hok)	31,807	18,523	2,223	1,899
GPL6699	Others	Print_1437	31,807	18,523	2,223	1,899
GPL5936	Others	HEEBO Human oligo array	32,137	18,537	2,652	2,292
GPL13151	Others	Human SMD Print_1355 (hoq)	32,183	18,561	2,653	2,293
GPL19407	Others	Phalanx Human OneArray (30968 probes)	19,611	17,302	2,984	3,022
GPL22109	Others	NimbleGen Human 100309 AS human 100426 pz array [MATC	65,210	16,787	16,223	5,409