

Supplementary Material

dreamBase: DNA Modification, RNA Regulation and Protein Binding of Expressed Pseudogenes in Human Health and Disease

Authors:

Ling-Ling Zheng^{1#*}, Ke-Ren Zhou^{1#}, Shun Liu¹, Ding-Yao Zhang¹, Ze-Lin Wang¹, Zhi-Rong Chen¹, Jian-Hua Yang^{1.*} and Liang-Hu Qu^{1.*}

Affiliation:

¹Key Laboratory of Gene Engineering of the Ministry of Education, State Key Laboratory for Biocontrol, Sun Yat-sen University, Guangzhou 510275, P. R. China

These authors contributed equally to this work

***Corresponding author:**

Mailing address: Biotechnology Research Center, Sun Yat-sen University, Guangzhou 510275, P. R. China

Phone: 86-20-84112399

Fax: 86-20-84036551

Ling-Ling Zheng, Email: zhengl33@mail.sysu.edu.cn

Jian-Hua Yang, Email: yangjh7@mail.sysu.edu.cn

Liang-Hu Qu, Email: lssqlh@mail.sysu.edu.cn

Supplementary table and figures

| | |
|-------------------------|--|
| Supplementary Table S1 | Number of RNA-Seq samples in 32 types of carcinomas |
| Supplementary Table S2 | Number of RNA-Seq samples in 31 types of normal tissues |
| Supplementary Table S3 | The number of pseudogenes regulated by RNA binding proteins |
| Supplementary Figure S1 | The ChIP-seq evidence around ABCA17P gene locus |
| Supplementary Figure S2 | The web interface of the search module |
| Supplementary Figure S3 | The web interface of the expression module. |
| Supplementary Figure S4 | The scatter plot (A) and boxplot (B) of co-expression patterns between ABCC6P1 pseudogene and its parent gene ABCC6 in adrenal gland tissue. |
| Supplementary Figure S5 | The ceRNA network composed by CTNNA1P1, CTNNA1 and microRNAs in multiple cancer types. |
| Supplementary Figure S6 | The detailed information of ceRNA network (A) and microRNA binding site on pseudogene (B). |
| Supplementary Figure S7 | The dreamBase genome browser (A) and lists (B) of RBPs binding sites distribution on pseudogenes. |

Table Legends

Table S1. Number of RNA-Seq samples in 32 types of carcinomas

| Source | Abbreviation | Sample Number |
|---|--------------|---------------|
| Cholangiocarcinoma | CHOL | 45 |
| Uterine Carcinosarcoma | UCS | 57 |
| Diffuse Large B-cell Lymphoma | DLBC | 47 |
| Adrenocortical Cancer | ACC | 77 |
| Uveal Melanoma | UVM | 79 |
| Kidney Chromophobe | KICH | 91 |
| Mesothelioma | MESO | 87 |
| Rectum Adenocarcinoma | READ | 103 |
| Thymoma | THYM | 121 |
| Glioblastoma Multiforme | GBM | 171 |
| Testicular Germ Cell Tumor | TGCT | 154 |
| Pancreatic Adenocarcinoma | PAAD | 183 |
| Pheochromocytoma & Paraganglioma | PCPG | 185 |

| | | |
|--|------|------|
| Uterine Corpus Endometrioid Carcinoma | UCEC | 204 |
| Esophageal Carcinoma | ESCA | 195 |
| Sarcoma | SARC | 264 |
| Cervical & Endocervical Cancer | CESC | 309 |
| Colon Adenocarcinoma | COAD | 331 |
| Kidney Papillary Cell Carcinoma | KIRP | 321 |
| Liver Hepatocellular Carcinoma | LIHC | 421 |
| Bladder Urothelial Carcinoma | BLCA | 426 |
| Stomach Adenocarcinoma | STAD | 450 |
| Ovarian Serous Cystadenocarcinoma | OV | 425 |
| Skin Cutaneous Melanoma | SKCM | 470 |
| Brain Lower Grade Glioma | LGG | 523 |
| Lung Squamous Cell Carcinoma | LUSC | 548 |
| Prostate Adenocarcinoma | PRAD | 548 |
| Head & Neck Squamous Cell Carcinoma | HNSC | 564 |
| Thyroid Carcinoma | THCA | 571 |
| Breast Invasive Carcinoma | BRCA | 1212 |
| Kidney Clear Cell Carcinoma | KIRC | 603 |
| Lung Adenocarcinoma | LUAD | 574 |

Table S2. Number of RNA-Seq samples in 31 types of normal tissues

| Source | Abbreviation | Sample Number |
|----------------------------------|---------------------|----------------------|
| Adipose Mixed Tissue | Adipose_Tissue | 515 |
| Adrenal Gland | Adrenal_Gland | 127 |
| Bladder Tissue | Bladder | 9 |
| Blood Tissue | Blood | 444 |
| Blood Vessel Mixed Tissue | Blood_Vessel | 604 |
| Bone Marrow (CML) | Bone_Marrow | 70 |
| Brain Mixed Tissue | Brain | 1146 |
| Breast Tissue | Breast | 179 |
| Cervix Uteri Mixed Tissue | Cervix_Uteri | 10 |
| Colon Tissue | Colon | 304 |
| Esophagus Mixed Tissue | Esophagus | 651 |
| Fallopian Tube | Fallopian_Tube | 5 |
| Heart Mixed Tissue | Heart | 376 |
| Kidney Tissue | Kidney | 27 |
| Liver Tissue | Liver | 110 |
| Lung Tissue | Lung | 287 |
| Muscle Tissue | Muscle | 396 |
| Nerve Tissue | Nerve | 278 |
| Ovary Tissue | Ovary | 88 |
| Pancreas Tissue | Pancreas | 165 |
| Pituitary Tissue | Pituitary | 107 |
| Prostate Tissue | Prostate | 100 |
| Salivary Gland | Salivary_Gland | 55 |
| Skin Mixed Tissue | Skin | 813 |
| Small Intestine | Small_Intestine | 92 |

| | | |
|-----------------------|---------|-----|
| Spleen Tissue | Spleen | 99 |
| Stomach Tissue | Stomach | 173 |
| Testis Tissue | Testis | 165 |
| Thyroid Tissue | Thyroid | 277 |
| Uterus Tissue | Uterus | 78 |
| Vagina Tissue | Vagina | 84 |

Table S3. The number of pseudogenes regulated by RNA binding proteins

| Name | Number | Name | Number | Name | Number |
|----------------|---------------|-----------------|---------------|---------------------|---------------|
| FUS | 5106 | CSTF2T | 500 | EIF4G2 | 166 |
| RBM6 | 4275 | U2AF1 | 499 | CAPRIN1 | 153 |
| SRSF1 | 3996 | PRPF8 | 497 | m6A-antibody | 150 |
| RBM10 | 3908 | HNRNPA1 | 495 | AUH | 135 |
| WDR33 | 3626 | RBM47 | 444 | SF3A3 | 131 |
| nELAVL | 3475 | SND1 | 437 | SF3B4 | 130 |
| TAF15 | 3427 | KHDRBS1 | 420 | ZC3H7B | 126 |
| CPSF2 | 3023 | Ro60 | 385 | DDX42 | 118 |
| RBFOX2 | 2643 | FAM120A | 380 | TNRC6A | 116 |
| UPF1 | 2631 | LARP4B | 362 | DICER1 | 115 |
| NOP58 | 2583 | DDX3X | 362 | RBM27 | 113 |
| FBL | 2500 | LIN28A | 360 | WTAP | 111 |
| PTBP1 | 2392 | CPSF6 | 356 | PCBP2 | 108 |
| ELAVL1 | 1656 | EWSR1 | 322 | GNL3 | 96 |
| EIF4A3 | 1620 | HNRNPC | 321 | TARBP2 | 84 |
| SRSF3 | 1300 | SMNDC1 | 316 | LARP7 | 73 |
| U2AF2 | 1216 | FXR1 | 307 | NPM1 | 70 |
| SRSF7 | 1214 | HNRNPUL1 | 305 | HNRNPD | 53 |
| SRSF9 | 1138 | TARDBP | 291 | HNRNPL | 48 |
| SRSF10 | 1135 | GTF2F1 | 284 | RBM22 | 46 |
| FMR1 | 976 | YWHAG | 277 | FKBP4 | 38 |
| MOV10 | 939 | ADAR | 267 | SLBP | 34 |
| DGCR8 | 938 | FXR2 | 253 | MSI2 | 32 |
| LIN28B | 937 | ILF3 | 251 | RC3H1 | 29 |
| CNBP | 915 | KHSRP | 249 | SBDS | 25 |
| NOP56 | 809 | LSM11 | 239 | EIF4G1 | 24 |
| IGF2BP3 | 793 | BCCIP | 228 | QKI | 19 |
| IGF2BP1 | 789 | TROVE2 | 214 | NCBP3 | 15 |
| SAFB2 | 766 | HNRNPK | 210 | ORF57 | 13 |
| DKC1 | 683 | METTL14 | 210 | ALKBH5 | 6 |
| IGF2BP2 | 645 | YTHDF1 | 207 | PUM2 | 4 |
| RBM5 | 611 | TIAL1 | 204 | YTHDF2 | 4 |
| HNRNPU | 598 | NONO | 198 | ZFP36 | 2 |
| SLTM | 592 | HNRNPM | 187 | PAPD5 | 1 |
| YTHDC1 | 580 | TIA1 | 186 | EIF3A | 1 |
| MSI1 | 546 | XRN2 | 184 | EIF3D | 1 |
| TRA2A | 530 | METTL3 | 177 | | |
| BUD13 | 522 | RTCB | 171 | | |

Figures

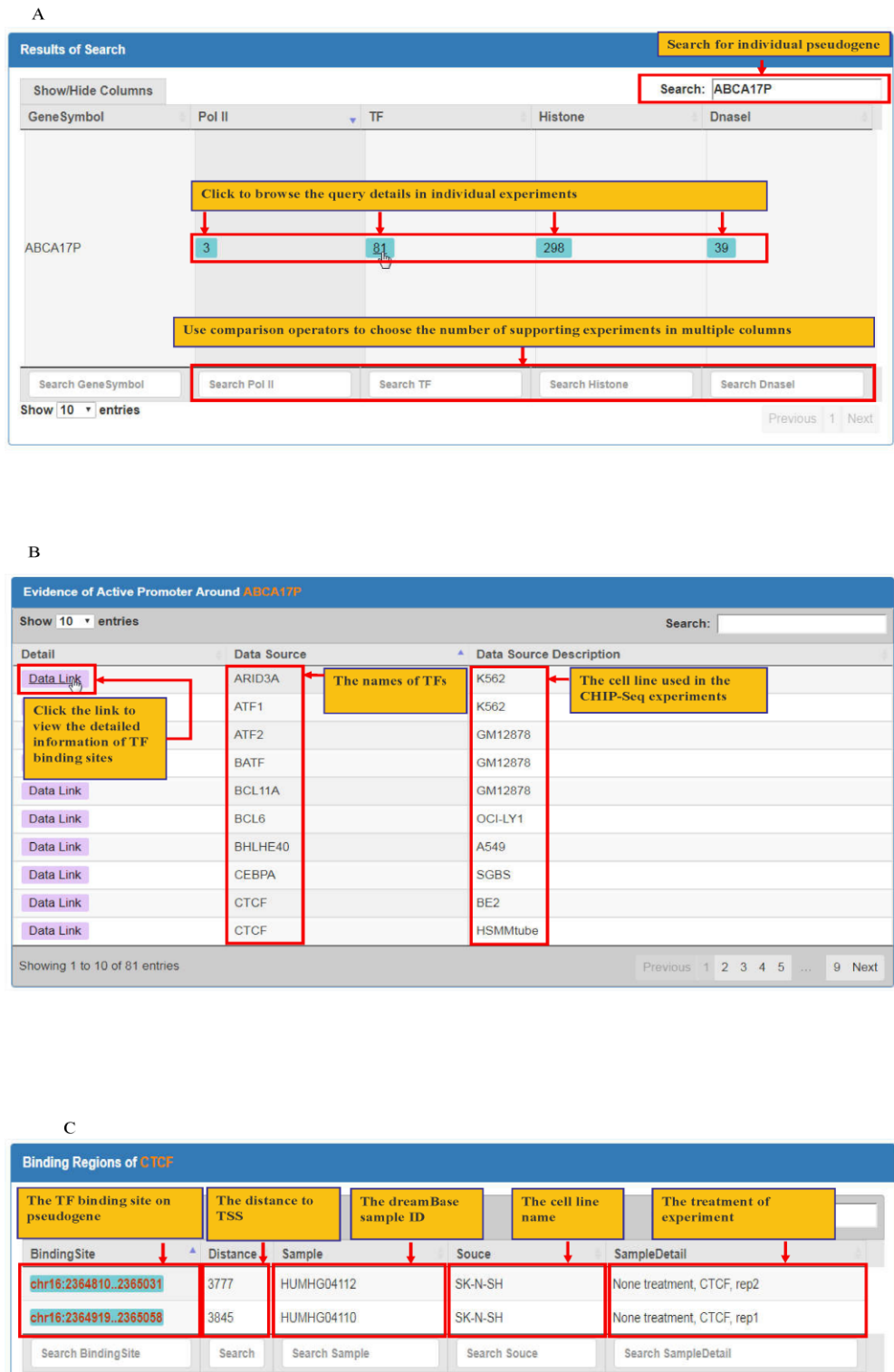


Figure S1. The ChIP-seq evidence around ABCA17P gene locus. (A) Number of support ChIP-seq experiments of ABCA17P. (B) Evidence of active promoters found around ABCA17P. (C) Detailed binding regions of CTCF around ABCA17P.

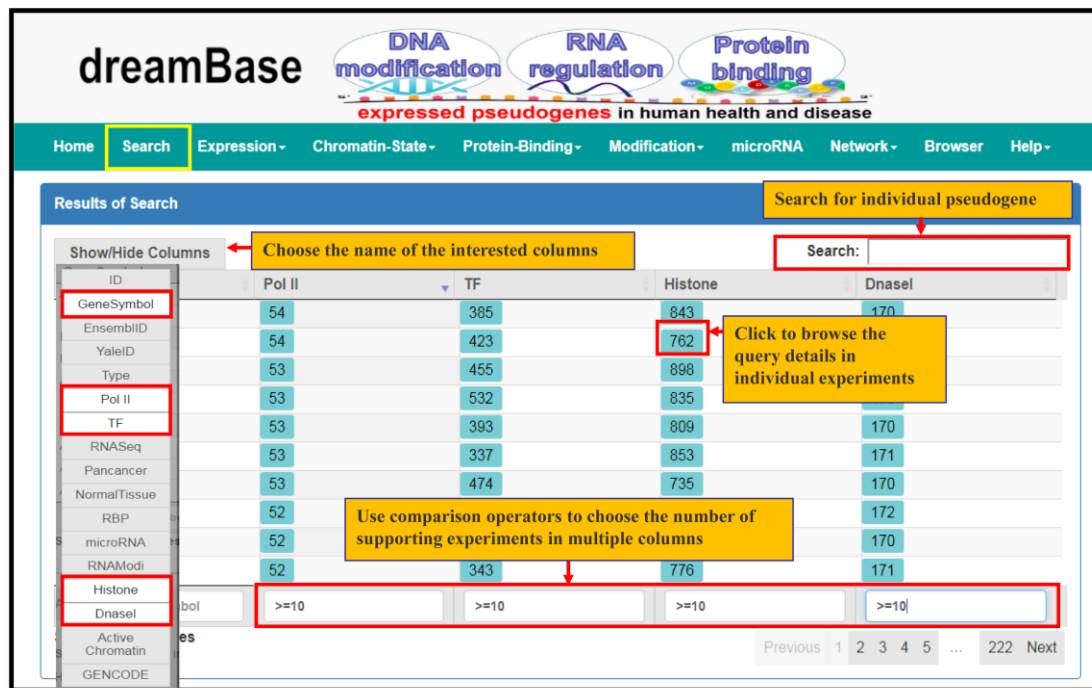


Figure S2. The web interface of the search module

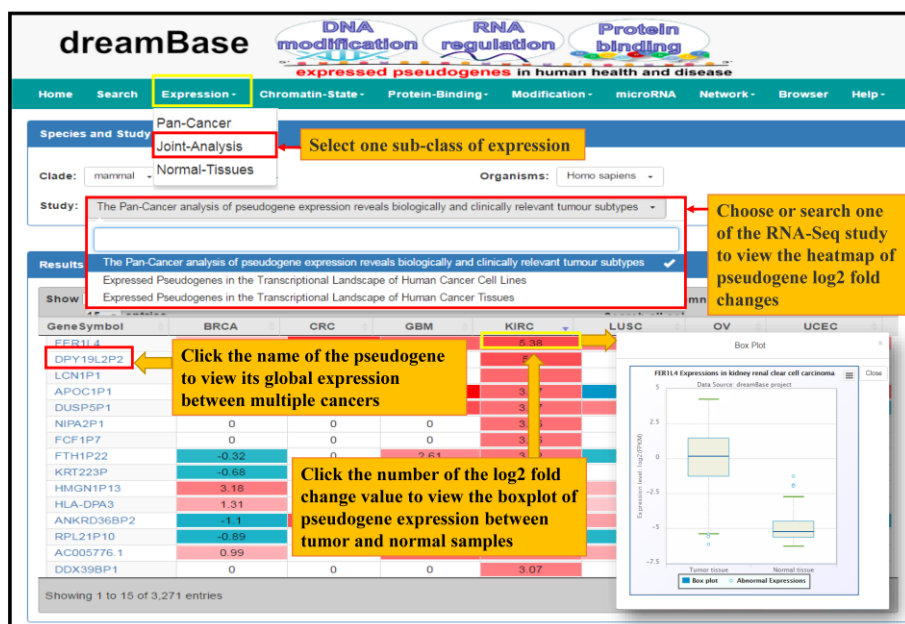


Figure S3. The web interface of the expression module.

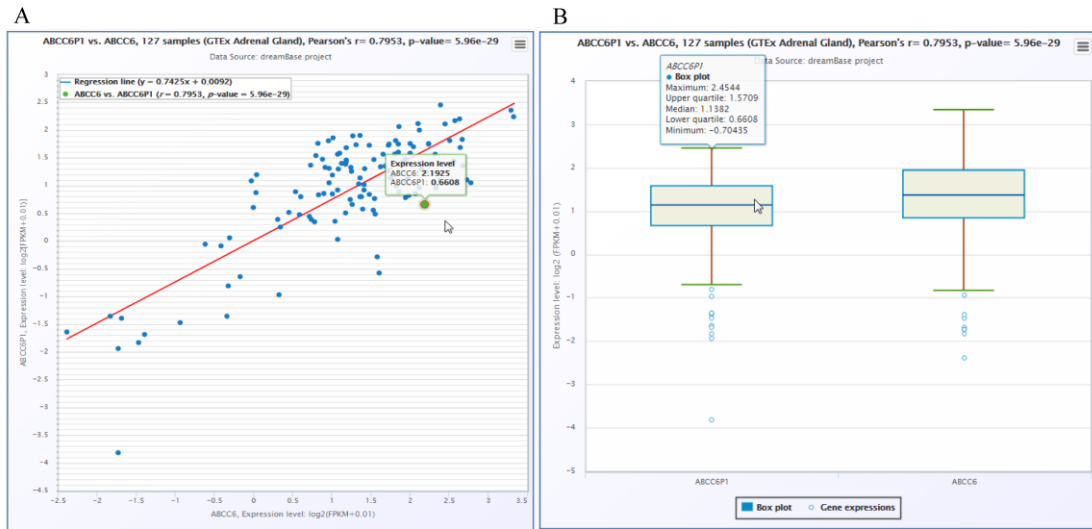


Figure S4. The scatter plot (A) and boxplot (B) of co-expression pattern between ABCC6P1 pseudogene and its parent gene ABCC6 in adrenal gland tissue.

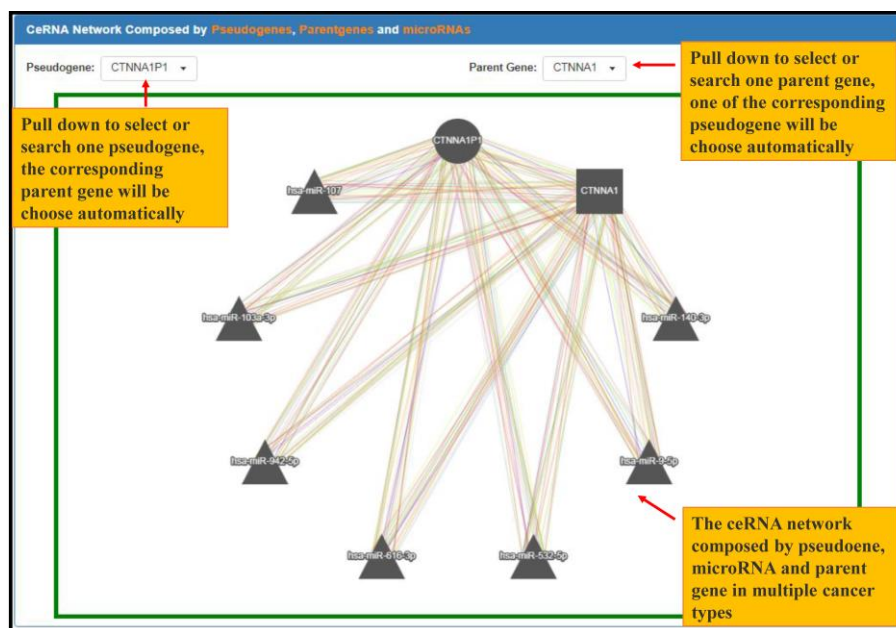


Figure S5. The ceRNA network composed by CTNNA1P1, CTNNA1 and microRNAs in multiple cancer types. Each node represents a gene, the circle node represents pseudogene, the square node represents parent gene and the triangle node represent microRNAs. Each edge represents a co-expression relationship of the two nodes in one cancer. Different edge color represents different types of cancers.

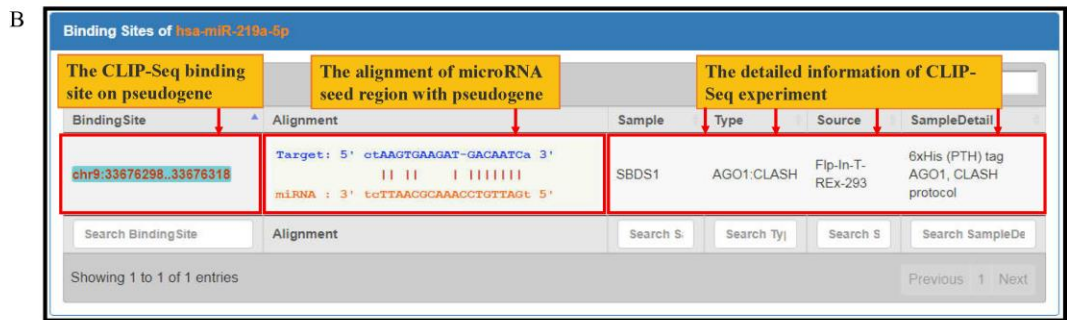
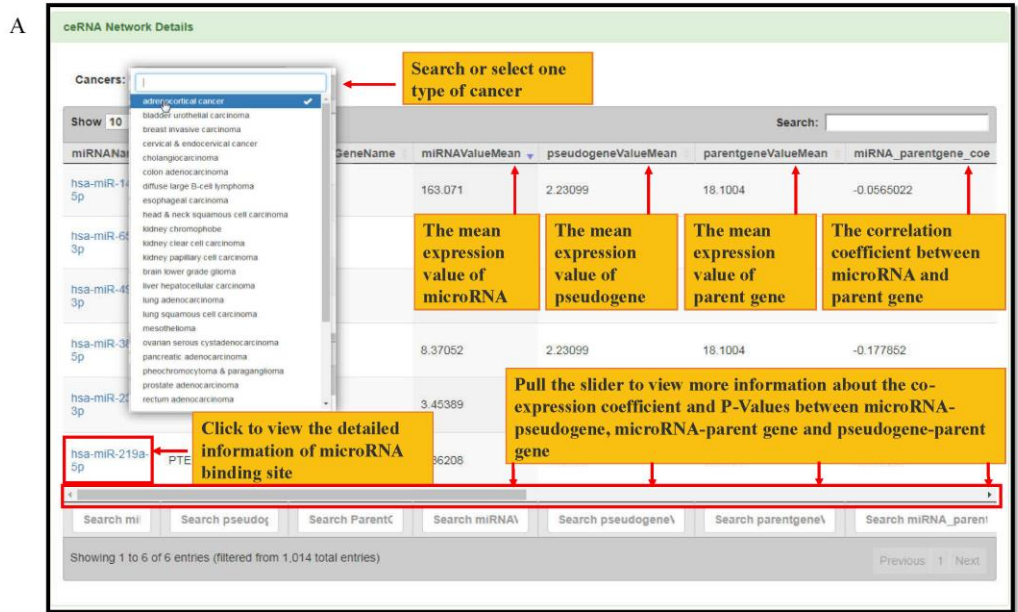
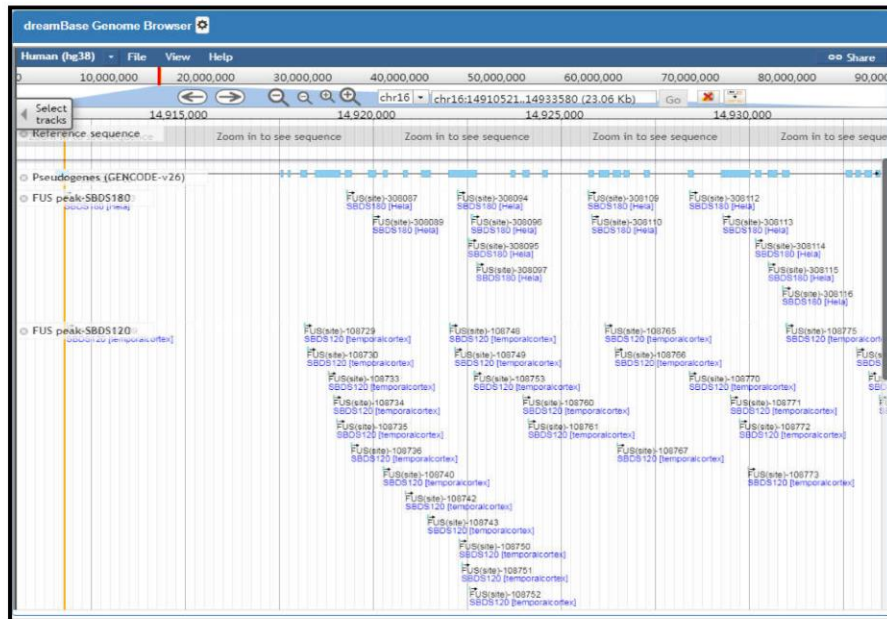


Figure S6. The detailed information of ceRNA network (A) and microRNA binding site on pseudogene (B).

A



B

Results of RBP-Binding

Download the results → Download: EXCEL | TXT

Show 10 entries Search:

| RBP Name | Gene Symbol | Supported Experiments | Binding Sites |
|----------|----------------|-----------------------|---------------|
| FUS | RP11-958N24.1 | 4 | 94 |
| FUS | PKD1P5 | 3 | 76 |
| FUS | RP11-958N24.2 | 3 | 70 |
| FUS | PKD1P1 | 4 | 68 |
| FUS | RP11-1212A22.1 | 3 | 52 |
| FUS | PKD1P6 | 4 | 41 |
| FUS | GTF2IP4 | 4 | 31 |
| FUS | HERC2P3 | 4 | 29 |
| FUS | SMG1P3 | 4 | 28 |
| FUS | EEF1A1P5 | 4 | 26 |

The number of supported CLIP-Seq experiments

The number of distinct binding sites on pseudogene

Showing 1 to 10 of 5,106 entries (filtered from 79,976 total entries) Previous 1 2 3 4 5 ... 511 Next

Figure S7. The dreamBase genome browser (A) and the distribution (B) of binding sites of RBPs on pseudogenes.