乳腺癌UCSC不同组学数据详细说明

UCSC网站中乳腺癌数据下载网址：

https://xenabrowser.net/datapages/?cohort=TCGA%20Breast%20Cancer%20(BRCA)&removeHub=https%3A%2F%2Fxena.treehouse.gi.ucsc.edu%3A443

不同组学数据详细说明如下表：

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| 数据类型 | 数据来源 | 数据规模 | 数据人群 | 数据格式 | 缺失值 | 前期是否有归一化或其他预处理 |
| 基因表达  (gene expression RNAseq) | UCSC网站：从TCGA DCC下载 | （行\*列）20,531个gene基因 \*1218个sample样本 | 乳腺浸润癌（BRCA）样本 | .TXT | 无处理说明 | 经过log2（x + 1）转化后Level\_3 RSEM标准化计数 |
| 甲基化  (HumanMethylation450) | 同上 | （行\*列）485,578个甲基化位点\*888个sample样本 | 同上 | 同上 | 同上 | Level\_3 |
| 拷贝数变异(copy number (gene-level)) | 同上 | （行\*列）24,777个基因\*1080个sample样本 | 同上 | 同上 | 同上 | 应用GISTIC2方法生成分段CNV数据，然后将其映射到基因以产生基因水平估计。使用UCSC xena HUGO probeMap将基因定位到人类基因组坐标上。参考GISTIC2方法PMID：21527027。 |
| 拷贝数变异-阈值型(copy number (gene-level)-threshod) | 同上 | （行\*列）24,777个基因\*1080个sample样本 | 同上 | 同上 | 同上 | 应用GISTIC2方法产生基因水平拷贝数估计。 GISTIC2进一步将估计值阈值化为-2，-1,0,1,2，表示纯合缺失，单拷贝缺失，二倍体正常拷贝，低水平拷贝数扩增或高水平拷贝数扩增。使用UCSC xena HUGO probeMap将基因定位到人类基因组坐标上。参考GISTIC2方法PMID：21527027。 |
| 体细胞非沉默突变（基因水平）(somatic non-silent mutation (gene-level)) | 同上 | （行\*列）46,246个基因\*776个sample样本 | 同上 | 同上 | 同上 | TCGA DCC下载.maf文件，在UCSC处理成基因样本矩阵 |
| 蛋白质表达(RPPA) | 同上 | （行\*列）282个蛋白质\*937个sample样本 | 同上 | 同上 | 同上 | 单位标准化后的level\_3 RPPA值 |
| miRNA成熟链表达(IlluminaHiseq RNAseq) | 同上 | （行\*列）2,239个miRNA\*832个sample样本 | 同上 | 同上 | 同上 | 对于每个样品，将相同miRNA成熟链的所有同种型表达加在一起， Level\_3，log2（total\_RPM +1）转化 |
| 通路活性(PARADIGM pathway activity  ) | 同上 | （行\*列）7,204个基因\*1093个sample样本 | 同上 | 同上 | 同上 | 使用PARADIGM方法仅对RNAseq数据推断TCGA乳腺浸润癌（BRCA）基因活性水平。PARADIGM（PMID 20529912）是通过结合策划的途径相互作用以及整合不同类型的基因组数据来推断肿瘤样品特异性遗传活性的途径分析方法。该分析中使用的途径来自NCI途径相互作用数据库。 |
| 转录因子调控影响(Transcription factor regulatory impact - HiSeqV2) | http://rabit.dfci.harvard.edu/download/ | （行\*列）1011个sample样本\*117个转录因子 | 同上 | 同上 | 同上 | 每种转录因子（TF）对由RABIT工具计算的肿瘤特异性基因表达模式的调节影响。 RABIT使用线性模型t值（系数/ stderr）来表示TF调节对每个肿瘤中基因表达的影响，并显示每个基质细胞中的t值。正t值表示TF上调其靶基因，而负值表明TF下调其靶基因。 |
| 临床变量(Phenotypes) | UCSC网站：来自TCGA数据 | （行\*列）1247个样本 samples \* 203 clinical variables临床变量 | 同上 | 同上 | 同上 | 无说明 |