单个基因的重要性分析

1. 基因重要性排序

根据RNA-seq报告中与CGI显著相关且有差异表达基因信息表，即与CGI显著相关和差异表达基因重合的基因集，根据其相关的显著性排序。

Table 1: 与CGI显著负相关且有差异表达基因信息表

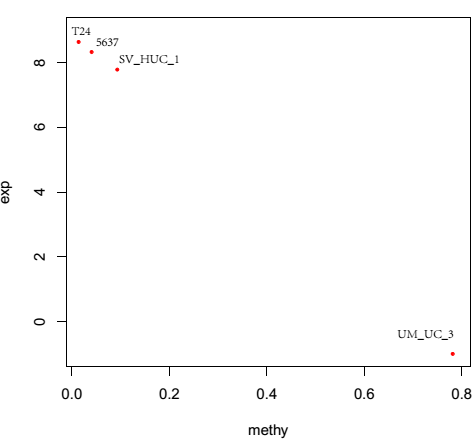
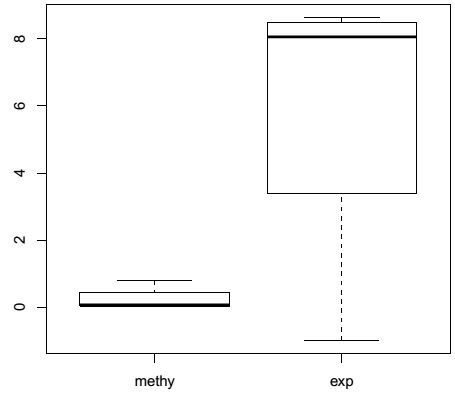
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **chr** | **CGIstart** | **CGIend** | **gene\_symbol** | **location** | **PCC** | **P.value** |
| chrX | 7927036 | 7927567 | PNPLA4 | 3'UTR | -0.9999 | 0.0001 |
| chr8 | 54457868 | 54459965 | SOX17 | 5'UTR | -0.99987 | 0.00013 |
| chr15 | 84657965 | 84658487 | NMB | 3'UTR | -0.99986 | 0.000136 |
| chr18 | 54959286 | 54959618 | CCDC68 | 3'UTR | -0.99985 | 0.000146 |
| chr9 | 112174874 | 112175544 | SUSD1 | 3'UTR | -0.99979 | 0.000207 |
| chr2 | 38959636 | 38960827 | LOC375196 | 3'UTR | -0.99978 | 0.00022 |
| chr19 | 12801117 | 12802104 | PRDX2 | 3'UTR | -0.99976 | 0.000236 |
| chr1 | 159780626 | 159781563 | DUSP23 | 5'UTR | -0.99966 | 0.000343 |
| chr20 | 58692471 | 58693364 | NPEPL1 | 5'UTR | -0.99958 | 0.000416 |
| chr10 | 22345071 | 22345933 | SPAG6 | 5'UTR | -0.99946 | 0.000536 |
| chr14\_KI270847v1\_alt | 37494 | 38116 | CHGA | 5'UTR | -0.99934 | 0.000658 |
| chr1 | 40316548 | 40317473 | COL9A2 | 3'UTR | -0.99926 | 0.000744 |
| chr9 | 98192004 | 98192311 | CORO2A | 3'UTR | -0.99924 | 0.000756 |
| chr4 | 154743667 | 154744681 | LRAT | 5'UTR | -0.99923 | 0.00077 |
| chr18 | 59272392 | 59274308 | RAX | 3'UTR | -0.9992 | 0.000804 |
| chr6 | 73394702 | 73395155 | DDX43 | 5'UTR | -0.99917 | 0.000827 |
| chrX | 51743318 | 51743914 | GSPT2 | 5'UTR | -0.9991 | 0.000902 |
| chr11 | 75561905 | 75562571 | SERPINH1 | 5'UTR | -0.99901 | 0.000994 |

去除了一些正相关的基因。

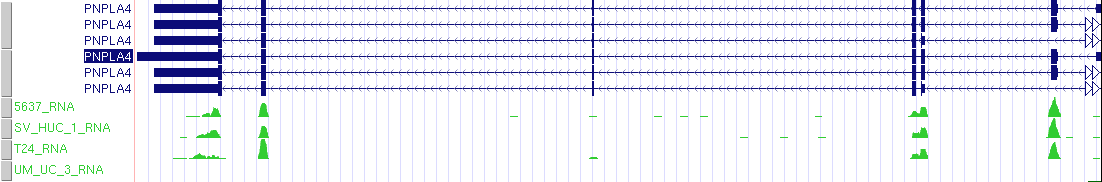
注：上表中只有标绿的CGI是非特异的，但是有差异的，其他CGI都是特异的。

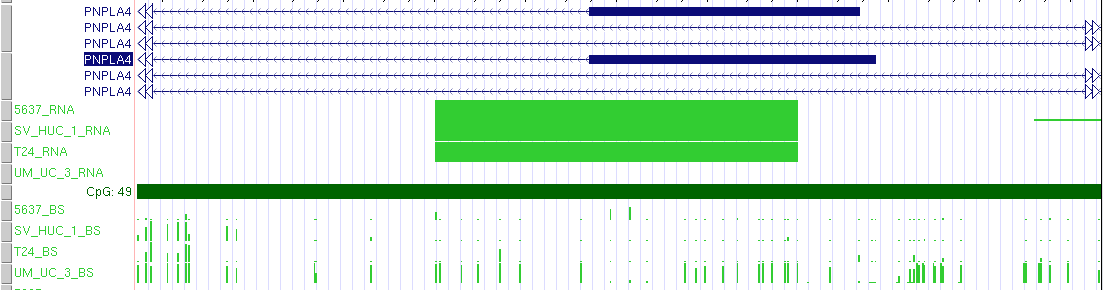
1. 甲基化值、表达值的图

以表中第一个基因-PNPLA4为例，在特异甲基化的CGI中查找发现，该CGI是UM\_UC\_3-Hyper的岛，对其对应的CGI（chrX：7927036-7927567）的甲基化值和基因的表达值分别作了箱式图和散点图(表达值经log2转换)。



可以看出UM\_UC\_3这个细胞系的CGI是特异高的，对应的基因表达是低的。





UCSC浏览器可视化界面图，其中UM\_UC\_3几乎没有表达值。而对应的CGI中UM\_UC\_3的甲基化值是特异高的。

箱式图和散点图都是用R绘制

可视化链接: <http://genome.ucsc.edu/cgi-bin/hgTracks?hgS_doOtherUser=submit&hgS_otherUserName=hongbo919&hgS_otherUserSessionName=BladderCancer_Methy_hg38_use_450K_combine>

1. 其他数据库中关于该基因的信息

在GeneCards数据库(<http://www.genecards.org/>)中查询PNPLA4的相关信息，有一段总结: PNPLA4 (Patatin Like Phospholipase Domain Containing 4) is a Protein Coding gene. Diseases associated with PNPLA4 include griscelli syndrome, type 2. Among its related pathways are Metabolism and Regulation of lipid metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha). GO annotations related to this gene include *triglyceride lipase activity*. An important paralog of this gene is PNPLA2.

其相关的两个通路 Metabolism : [http://pathcards.genecards.org/card/metabolism](%20http:/pathcards.genecards.org/card/metabolism)

Regulation of lipid metabolism by Peroxisome proliferator-activated receptor alpha (PPARalpha): <http://pathcards.genecards.org/card/regulation_of_lipid_metabolism_by_peroxisome_proliferator-activated_receptor_alpha_(pparalpha)>

另在pubmed中查找PNPLA2和膀胱癌有关的文献，但没有结果。

1. 方法及流程

箱式图、散点图及可视化图都是展示了一个基因的差异性、相关性，主要的方法就是在数据库中查询该基因相关的信息。

主要的数据库： GeneCards (<http://www.genecards.org/>)

miRBase (<http://www.mirbase.org/index.shtml>)