

实验6、Pearson相关分析

年级：15级 专业：生信 学号：1513401013 姓名：郑磊

编号 一 二 三 四 总分 评阅人

得分

软硬件平台：

1. 硬件平台：（硬件配置）i5，2.9HZ处理器，16G内存，64位操作系统
2. 系统平台：（操作系统及其版本号）Windows10 企业版
3. 软件平台：（软件系统及其版本号，若是在线分析平台，还需要提供URL地址）R3.4.1 ， Rstudio

一、目的要求：

- 1、加深对相关性的理解；
- 2、熟悉并掌握相关性分析所涉及的R语言函数和脚本；
- 3、从统计学角度来理解不同基因转录表达之间的生物学关联。

二、实验内容：

1、不同基因表达水平之间相关性的分析

该环节需要大家提前准备好一个基因表达谱数据，如果没有，则有授课教师提供（gds4794）。

1.1、加载数据

加载GEOquery包，读取基因表达谱数据（gds4794）。

```
setwd("D:/RFile/实验六")
```

```
library(GEOquery)
```

```
gds4794 <- getGEO(filename='GDS4794.soft.gz')
```

1.2、提取数据表

从该基因表达谱数据中提取基因表达数据表。

```
data<-Table(gds4794)

rownames(data)<-data[,1]

row.name = rownames(data)
```

1.3、随机抽样

从该数据表中随机抽取一行数据，记录对应的探针ID和基因名称。

```
n = 1

set.seed(1)

sam.row.name = sample(row.name,n,replace=F)

sam.row.name

a <- unlist(data[sam.row.name,3:67])

gene_name_a <- as.character(data[sam.row.name,2])

gene_name_a
```

1.5、不同基因表达水平Pearson相关系数计算

遍历整个基因表达谱数据表，利用R语言中的Pearson相关系数检验函数，分析其他所有基因表达水平，与1.3步随机抽取的基因表达水平之间的相关性；记录p.value和相关系数值（cor）；同时为所有p.value和相关系数值（cor）关联基因名称。

```
p = NULL

r = NULL

for(i in 1:nrow(data)){
```

```

b <- unlist(data[i,3:67])

x <- cor.test(a,b, method="pearson")

p <- c(p,x$p.value)

r <- c(r,x$estimate)

}

names(p)<-data[,2]

names(r)<-data[,2]

```

1.6、高相关性基因筛选

设定p.value（至少小于0.05）和相关系数r的筛选阈值（至少大于0.5）；对1.5步计算结果进行筛选，保留符合条件的基因；对符合条件的p.value和相关系数r所关联的基因名称进行交集运算；查看交集运算结果中是否存在非法基因信息，如果有去除它。

```

p_value = 0.01

r_cutoff = 0.6

p2 <- p[p<p_value]

r2 <- r[r>r_cutoff]

length(p2)

length(r2)

genes <- intersect(names(p2),names(r2))

genes[-match("ARHGAP26",genes)]

length(genes)

```

```
tail(genes)

genes2 <- c(gene_name_a, genes[1:645])

out = "pearson-related-genes.txt"

write.table(genes2, out)
```

1.7、高相关性基因生物学意义的探讨

为什么有些基因的表达与你所选择的基因表达水平之间存在较高的相关性？

三、实验结果：

```
> sam.row.name
[1] "205069_s_at"
> gene_name_a
[1] "ARHGAP26"
> length(genes2)
[1] 646
```

结果中第一行为抽的样本基因共有645个基因与之相关。

四、讨论：

1.7 在生物体中一个性状往往是由多个基因共同调控的，甚至会影响到其他性状或代谢过程，还有一些反馈调节也可能会影响到基因的表达。还有一种可能，这些基因可能调控着同一条代谢通路。

在DAVID网站处理结果

Category	Term	Count	%	PValue	Genes
GOTERM_C_C_DIRECT	GO:0030054~cell junction	55	9.353741	3.46E-19	KCNC3, SRCIN1, SYT2, CLSTN1, GABBR1, GRIK5, SYT9, SYNGR1, SYP, ATP2B2, PRRT2, GRIN2C, HOMER3, DLG4, CDK16, SCAMP5,

					SVOP, BCR, PTPRR, BSN, MINK1, CBLN3, CBLN1, VAMP2, UNC13C, FAIM2, UNC13A, PRKCZ, ACHE, BRSK1, RIMS1, RIMS4, SYN1, ATN1, ARHGAP44, LRFN1, GABRD, PARD6A, DLGAP3, GABRA6, SYT12, NLGN2, SLC6A17, GRIA4, IGSF9B, KCNK1, SHANK3, SLC17A7, NMT1, MAP1S, GTF2F1, CYFIP2, SPTBN2, TPRG1L, ADGRL1
GOTERM_BP_DIRECT	GO:0007269~neurotransmitter secretion	15	2.55 102	1.0 7E-10	PPFIA3, RAB3A, PPFIA4, CPLX1, NRXN2, NRXN3, SYT2, STXBP1, BRSK1, RIMS1, GRM4, SYN1, DNAJC5, VAMP2, UNC13A
GOTERM_CC_DIRECT	GO:0048786~presynaptic active zone	12	2.04 0816	1.8 0E-10	SYP, SLC17A7, PPFIA3, PPFIA4, SYN1, BSN, SV2A, UNC13C, ADORA1, RIMS1, RIMS4, UNC13A
GOTERM_BP_DIRECT	GO:0007268~chemical synaptic transmission	29	4.93 1973	3.2 5E-10	KCNC4, SCN1B, CPLX1, SNCB, SYN1, GRIN2C, SLC1A6, LRFN1, DLG4, APBA2, KCNQ2, NPFF, GABRD, GPR176, SCN2B, NRXN2, SLC12A5, ASIC2, MINK1, BSN, PRKCG, GRIA4, GRM1, GRM4, CBLN1, CACNA1G, RIT2, UNC13C, UNC13A
GOTERM_CC_DIRECT	GO:0014069~postsynaptic density	25	4.25 1701	4.5 3E-10	ALS2, BCR, DLGAP3, SRCIN1, CLSTN3, BAIAP2, CLSTN1, GRIK5, BSN, MINK1, PRKCG, IGSF9B, GRM1, ADORA1, SHANK3, PPP1R9B, SYN1, GRIN2C, PTK2B, ARRB1, HOMER3, MAPT, LRFN1, DLG4, ADD2
GOTERM_CC_DIRECT	GO:0045211~postsynaptic membrane	26	4.42 1769	1.5 2E-09	SRCIN1, CLSTN3, CLSTN1, GRIK5, GABBR1, ADORA1, ANK1, HOMER3, GRIN2C, LRFN1, DLG4, IQSEC3, GABRD, BCR, DLGAP3, COL13A1, GABRA6, NLGN2, MINK1, GRIA4, IGSF9B, SHANK3, CBLN1, ARRB1, APBB1, FAIM2
GOTERM_BP_DIRECT	GO:0016079~synaptic vesicle exocytosis	10	1.70 068	4.4 6E-09	RAB3A, CPLX2, CPLX1, PIP5K1C, DNAJC5, ADGRL1, VAMP2, UNC13C, RIMS1, UNC13A
GOTERM_CC_DIRECT	GO:0043195~terminal bouton	14	2.38 0952	1.4 8E-	SYP, RAB3A, CPLX2, CPLX1, SYN1, PVALB, GRIK5, STXBP1, DNAJC5,

				08	SV2A, UNC13C, SYNGR1, ADORA1, UNC13A
GOTERM_C C_DIRECT	GO:0008021~synaptic vesicle	16	2.72 1088	4.2 3E-08	SYP, RAB3A, SVOP, SYN1, SYT2, DLG4, APBA2, DNAJC5, TPRG1L, SLC6A17, SV2A, BRSK1, VAMP2, CDK16, STX1B, DNMT1
GOTERM_B P_DIRECT	GO:0048172~regulation of short-term neuronal synaptic plasticity	8	1.36 0544	4.5 0E-08	SYP, SHISA6, RAB3A, PPFIA3, SHISA8, SLC8A2, SYNGR1, UNC13A
GOTERM_C C_DIRECT	GO:0043025~neuronal cell body	29	4.93 1973	9.4 2E-08	ALS2, CPLX2, CPLX1, SRCIN1, FKBP4, ADORA1, ATP2B2, SEZ6L2, PVALB, PTK2B, TIAM1, PDE1A, KNDC1, TUBB4A, DAB2IP, KCND1, BAIAP2, SPTBN4, SLC12A5, RTN4R, ASIC2, GPD5, IGSF9B, PPP1R9B, MAP1S, SPTBN2, FBXO31, APBB1, SEZ6
GOTERM_C C_DIRECT	GO:0030425~dendrite	30	5.10 2041	9.9 3E-08	ALS2, KCNC1, CPLX2, KCNC3, CPLX1, SRCIN1, GABBR1, GNG13, GRIK5, TRIM3, SYN1, PTK2B, MAPT, LMTK3, KNDC1, NPFF, KCND3, DAB2IP, KCND1, MINK1, BSN, PRKCG, GRIA4, IGSF9B, KCNK1, GRM1, AMIGO1, PPP1R9B, MAP1S, MAPK8IP3
GOTERM_C C_DIRECT	GO:0042734~presynaptic membrane	13	2.21 0884	1.3 6E-07	KCNC3, NRXN2, GRIK5, GABBR1, NLGN2, ADORA1, RIMS1, SYP, GRM4, ADGRL1, UNC13C, APBB1, UNC13A
GOTERM_C C_DIRECT	GO:0045202~synapse	21	3.57 1429	2.1 3E-07	GABRD, NMNAT2, PPFIA4, ACHE, SRCIN1, SNCB, GABRA6, NLGN2, ASIC1, KCNK1, SHISA6, CBLN3, ATP2B2, SHISA8, PRRT2, MAP1S, CYFIP2, DLG4, VAMP2, ADGRL1, APBB1
GOTERM_B P_DIRECT	GO:0014047~glutamate secretion	9	1.53 0612	8.2 8E-07	SLC17A7, RAB3A, PPFIA3, PPFIA4, CPLX1, SLC1A6, STXBP1, VAMP2, RIMS1
GOTERM_C C_DIRECT	GO:0005886~plasma membrane	164	27.8 9116	1.0 8E-06	SLC22A17, KCNC1, CADM3, KCNC4, ADCY1, KCNC3, ATP1B2, SYT2, SYT3, TSPAN5, GRIK5, SYT9,

					L1CAM, FLCN, ADORA1, RELL2, CDH22, ATP2B2, SEZ6L2, DIRAS1, ANK1, PRRT2, GRIN2C, HOMER3, TIAM1, VPS4A, DNAJC5, SLC4A5, SCAMP5, DAB2IP, CRTAM, KCND3, KCND1, MADD, MDGA1, BAIAP2, TTC7B, PTPRR, SOCS7, TMEM25, CRHR1, ARRB1, CLIP3, VAMP2, UNC13C, ADD2, NCR3LG1, GLG1, ARFGAP2, ACHE, SCN1B, ARHGEF25, PANX2, RUND3A, NKAIN4, RIMS1, KCNJ3, AMER3, PLCH2, TTYH1, LRFN1, ENTPD6, CDC42EP4, GABRD, PARD6A, MLC1, TBC1D10B, GABRA6, SYT12, SPTBN4, KCNK1, RNF103-CHMP3, ANKH, SHANK3, NMT2, NMT1, KCNJ9, CACNA1G, TNK2, CACNA1F, RIT2, GPR83, GPR162, CLTB, BCAR1, GABBR1, PPIP5K1, MTSS1L, MCF2L, PCDHGA1, ST6GALNAC6, AP2B1, NISCH, MGRN1, DYNLL2, SLC1A6, MAPT, DLG4, CALN1, SV2A, KCNQ2, CDK16, DPP6, TRPM3, ADAM11, GPR176, ZP2, NRXN2, MICAL3, STXBP1, PRKCG, CACNG2, GRM1, STX1B, GPR37L1, GRM4, MAST1, CD99L2, GNB5, SEZ6, MAP3K12, EPN1, PLPP4, RAB3A, PRKCZ, PTOV1, ABTB1, GNG13, RHBG, DISP2, STUB1, TPCN1, RGS11, TSPYL2, FAT2, ENO2, APBA2, CAMK2B, EHD1, INPP5A, DGKQ, SLC8A2, NF2, COL13A1, SLC12A5, ASIC2, RTN4R, TMEM266, GRIA4, ASIC1, SLC17A7, PPP1R9B, GBA2, AP2A2, CDH15, TENM1, JMJD6, SLC6A8, CDH18, ATP6V0A1, DGKZ, ADGRL1, APBB1, DNM1
GOTERM_C C_DIRECT	GO:0030672~synaptic vesicle membrane	11	1.87 0748	2.7 8E-06	SLC17A7, SVOP, SYN1, SYT2, SYT12, SYT9, SV2A, VAMP2, SLC6A17, SYNGR1, SCAMP5
GOTERM_B	GO:0006887~ex	13	2.21	4.2	ANK1, CPLX1, SRCIN1, ARHGAP44,

P_DIRECT	ocytosis		0884	5E-06	MICAL3, BRSK2, SCRNI1, DNAJC5, VAMP2, CDK16, UNC13C, RIMS1, SCAMP5
GOTERM_B P_DIRECT	GO:0060079~ex citatory postsynaptic potential	8	1.36 0544	1.0 9E-05	SLC17A7, GRIN2C, MECP2, GRIK5, NPAS4, ADORA1, SEZ6, NPFF

其中详细说明了基因之间的关系与联系。