Описание Project\_rnaseq1  
Два региона

**Цель:** Выявление дифференциально экспрессирующийхся генов в двух регионах мозга человека: Префронтальная кора и белое вещество мозжечка. Определение перепредставленных GO-функций среди генов со статистически значимыми различиями в экспрессии. Кластеризовать значимые гены по похожести изменения экспрессии. Сравнить GO-функции между кластерами.

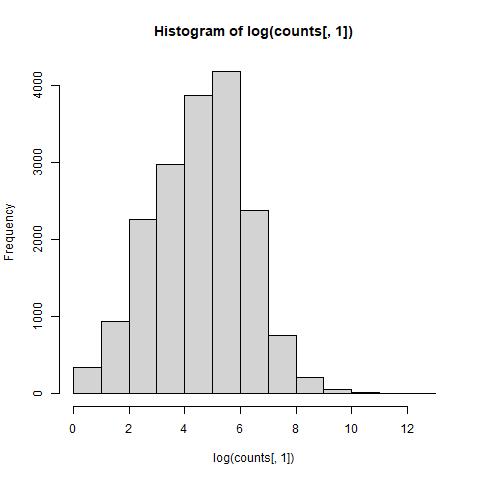
|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Status | SRR8750487.bam | SRR8750488.bam | SRR8750489.bam | SRR8750490.bam | SRR8750495.bam | SRR8750496.bam | SRR8750497.bam | SRR8750498.bam |
| 1 | Assigned | 7274109 | 7908369 | 9347947 | 8176398 | 8376566 | 8484980 | 9097226 | 10682805 |
| 2 | Unassigned\_Unmapped | 4238295 | 4933488 | 6713551 | 6095526 | 4312120 | 4770489 | 4210784 | 4948310 |
| 3 | Unassigned\_Read\_Type | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 4 | Unassigned\_Singleton | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 5 | Unassigned\_MappingQuality | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | Unassigned\_Chimera | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | Unassigned\_FragmentLength | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 8 | Unassigned\_Duplicate | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 9 | Unassigned\_MultiMapping | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 10 | Unassigned\_Secondary | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 11 | Unassigned\_NonSplit | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 12 | Unassigned\_NoFeatures | 2341835 | 3298199 | 4875343 | 6766937 | 3560521 | 3610096 | 3319902 | 2212683 |
| 13 | Unassigned\_Overlapping\_Length | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 14 | Unassigned\_Ambiguity | 843676 | 1090610 | 1122771 | 1099443 | 950696 | 852614 | 819562 | 1075187 |

**Этапы:**1. Подготовка библиотек1.1 Картирование полученных ридов на референсный геном с аннотацией с помощью Hisat2. Результат: Уровень выравнивания 60-75% - низкий, но так как образцы были получены из тканей мертвого донора – ок  
1.2 Подсчет ридов в R.   
Результат: Присвоено 40-55% ридов. (?)

1.3 Создаем таблицу с метаданными.  
Результат:

|  |  |
| --- | --- |
|  | region |
| Cwhite\_1 | Cwhite |
| Cwhite\_2 | Cwhite |
| Cwhite\_3 | Cwhite |
| Cwhite\_4 | Cwhite |
| Prefro\_5 | Prefro |
| Prefro\_6 | Prefro |
| Prefro\_7 | Prefro |
| Prefro\_8 | Prefro |

1.4 Отфильтровать гены по покрытию. Условие – минимум 20 ридов  
Результат: Из 65616 осталось 15561. Достаточное кол-во, но для поиска GO – увеличить min глуб

1.5 Контроль качества.   
Распределение каунтов в логарифмической шкале – нормальное

2.0 Поиск дифференциально экспрессирующихся генов

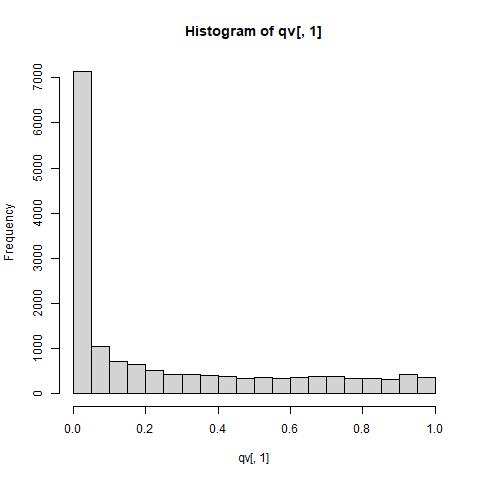
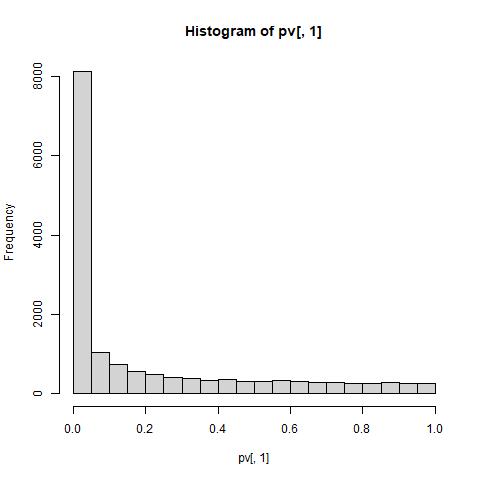
2.1 RLE нормировка.  
Результат: Все RLE коэффициенты равны 1±0.1 – сильных отклонений в размерах библиотек нет. **(??)**

2.2 Вычисление диф генов с помощью двух методов: t-test (exactTest) и линейной модели (glm)  
Результат:

|  |  |  |
| --- | --- | --- |
|  | t-test | glm |
| Кол-во генов pv < 0.5 | 8046 | 8117 |
| Кол-во генов qv <0.5 после поправки | 7049 | 7132 |

После поправки:

До поправки:



3.0 Поиск перепредставленных GO – категорий среди значимых генов

Результат: среди значимых генов 572 перепредставленных группы по сравнению с незначимыми, покрытие ридами которых >50.

**ТОП 10 ПЕРЕПРЕДСТАВЛЕННЫХ GO – функций среди значимых генов**

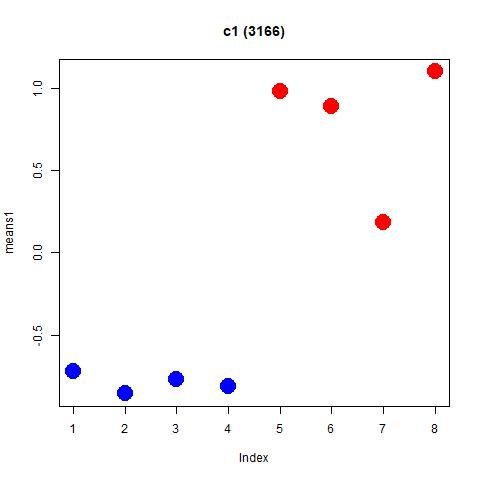
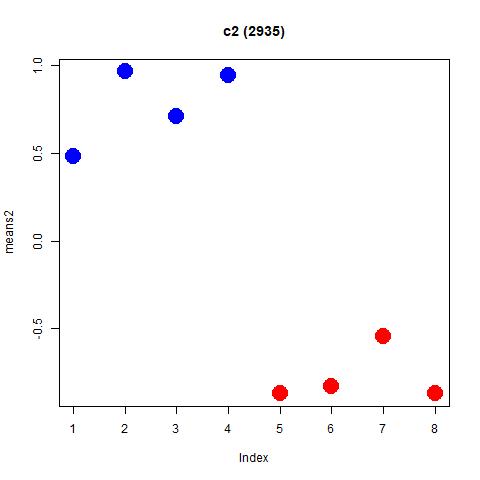
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | GO.ID | Term | Annotated | Significant | Expected | result1 |
| 1 | GO:0007268 | chemical synaptic transmission | 542 | 402 | 253,36 | < 1e-30 |
| 2 | GO:0098916 | anterograde trans-synaptic signaling | 542 | 402 | 253,36 | < 1e-30 |
| 3 | GO:0099537 | trans-synaptic signaling | 550 | 405 | 257,1 | < 1e-30 |
| 4 | GO:0099536 | synaptic signaling | 569 | 415 | 265,98 | < 1e-30 |
| 5 | GO:0007399 | nervous system development | 1847 | 1099 | 863,38 | < 1e-30 |
| 6 | GO:0022008 | neurogenesis | 1280 | 791 | 598,33 | < 1e-30 |
| 7 | GO:0048699 | generation of neurons | 1198 | 747 | 560 | 1.1e-30 |
| 8 | GO:0030182 | neuron differentiation | 1079 | 678 | 504,38 | 4.0e-29 |
| 9 | GO:0048666 | neuron development | 910 | 582 | 425,38 | 9.4e-28 |
| 10 | GO:0031175 | neuron projection development | 820 | 531 | 383,31 | 3.3e-27 |
| 11 | GO:0050804 | modulation of chemical synaptic transmis... | 373 | 275 | 174,36 | 4.5e-27 |
| 12 | GO:0099177 | regulation of trans-synaptic signaling | 374 | 275 | 174,83 | 9.2e-27 |
| 13 | GO:0051960 | regulation of nervous system development | 747 | 487 | 349,18 | 6.3e-26 |
| 14 | GO:0007267 | cell-cell signaling | 1188 | 722 | 555,33 | 6.8e-25 |
| 15 | GO:0048468 | cell development | 1557 | 910 | 727,82 | 1.2e-23 |
| 16 | GO:0050767 | regulation of neurogenesis | 667 | 436 | 311,79 | 1.4e-23 |
| 17 | GO:0050877 | nervous system process | 620 | 409 | 289,82 | 2.7e-23 |
| 18 | GO:0048812 | neuron projection morphogenesis | 547 | 366 | 255,69 | 1.5e-22 |
| 19 | GO:0120039 | plasma membrane bounded cell projection ... | 560 | 372 | 261,77 | 4.8e-22 |
| 20 | GO:0048858 | cell projection morphogenesis | 564 | 374 | 263,64 | 6.0e-22 |

4.0 Кластеризация значимых генов.

Результаты: Два кластера: группа генов, увеличивших свою экспрессию в префронтальной коре, и гены, уменьшившие

Увеличили экспрессию – группа 1

Уменьшили экспрессию – группа 2



red – Prefrontal

blue – Cerebellar

4.1 Поиск перепредставленных GO -категорий внутри группы 1 по сравнению с группой 2

Результаты: GO-категории перепредставленные в группе 1(уменьшивших экспрессию в префронатльной коре)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | GO.ID | Term | Annotated | Significant | Expected | result1 |
| 1 | GO:0016071 | mRNA metabolic process | 319 | 240 | 147,56 | 1.5e-27 |
| 2 | GO:0010467 | gene expression | 1628 | 926 | 753,04 | 1.7e-25 |
| 3 | GO:0006396 | RNA processing | 312 | 231 | 144,32 | 8.1e-25 |
| 4 | GO:1901360 | organic cyclic compound metabolic proces... | 1730 | 971 | 800,22 | 3.4e-24 |
| 5 | GO:0034641 | cellular nitrogen compound metabolic pro... | 1819 | 1014 | 841,39 | 3.8e-24 |
| 6 | GO:0006614 | SRP-dependent cotranslational protein ta... | 74 | 73 | 34,23 | 8.0e-24 |
| 7 | GO:0045047 | protein targeting to ER | 78 | 76 | 36,08 | 1.7e-23 |
| 8 | GO:0006613 | cotranslational protein targeting to mem... | 77 | 75 | 35,62 | 3.6e-23 |
| 9 | GO:0019083 | viral transcription | 104 | 95 | 48,11 | 7.3e-23 |
| 10 | GO:0019080 | viral gene expression | 110 | 99 | 50,88 | 1.4e-22 |

GO-категории перепредствленные в группе 1 (генов, увеличивших уровень экспрессии в префронтальной коре)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | GO.ID | Term | Annotated | Significant | Expected | result1 |
| 1 | GO:0099536 | synaptic signaling | 415 | 356 | 223,04 | < 1e-30 |
| 2 | GO:0007268 | chemical synaptic transmission | 402 | 346 | 216,05 | < 1e-30 |
| 3 | GO:0098916 | anterograde trans-synaptic signaling | 402 | 346 | 216,05 | < 1e-30 |
| 4 | GO:0099537 | trans-synaptic signaling | 405 | 347 | 217,66 | < 1e-30 |
| 5 | GO:0007267 | cell-cell signaling | 722 | 549 | 388,03 | < 1e-30 |
| 6 | GO:0050804 | modulation of chemical synaptic transmis... | 275 | 237 | 147,8 | < 1e-30 |
| 7 | GO:0099177 | regulation of trans-synaptic signaling | 275 | 237 | 147,8 | < 1e-30 |
| 8 | GO:0007399 | nervous system development | 1099 | 740 | 590,65 | 5.3e-25 |
| 9 | GO:0050808 | synapse organization | 259 | 212 | 139,2 | 2.3e-22 |
| 10 | GO:0023052 | signaling | 2137 | 1316 | 1148,52 | 1.0e-21 |

ИТОГИ: Выбранные два региона мозга имеют ~7100 дифференциально экспрессирующихся генов, среди которых ~3100 имеет повышенный уровень экспрессии в префронтальной коре по сравнению с белым веществом мозжечка. Среди генов с большим уровнем экспрессии в префронтальной коре перепредставлены GO-категории связанные с синаптической передачей. GO-категории генов, имеющих меньший уровень экспрессии в префронтальной коре, являются не специфичными для нервной ткани.

ВЫВОДЫ: