# Анализируем результаты RNA-seq анализа 4 регионов мозга человека

Цель - выявить дифференциальноэкспрессирующиеся гены, проанализировать обагощение по GO

```
In [1]: import pandas as pd
```

## Готовим таблицу с описанием образцов

```
In [18]: samples = pd.read_csv(r'/Users/igorfesenko/project.samples.tsv', sep='\t', nam готовом таблицу для загрузки в R

In [19]: samples['tissue'] = samples['tissue'].map(lambda x: x.split(' ', maxsplit=1)[1]

In [20]: samples['sample'] = samples['R1'].map(lambda x: x.split('_', maxsplit=1)[0]) samples.drop(columns=['R1','R2'],inplace=True)

In [24]: samples['sample'] = samples['sample'].map(lambda x: x.replace('-','.'))
```

#### samples.to\_csv(r'/Users/igorfesenko/samples\_brain\_R.csv', sep='\t')

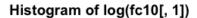
```
In [27]: samples
```

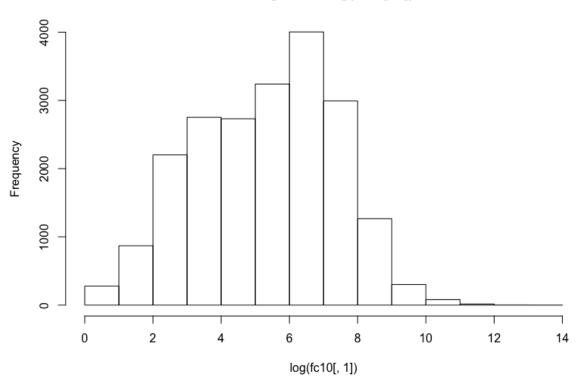
Out[27]:		tissue	sample
	HA74	Prefrontal (BA10)	R18001261.BM1.BM1.11
	HB74	Prefrontal (BA10)	R18001263.BM3.BM3.2
	HC74	Prefrontal (BA10)	R18001262.BM2.BM2.20
	HD74	Prefrontal (BA10)	R18001262.BM2.BM2.13
	HA32	Hypothalamus	R18001261.BM1.BM1.20
	HB32	Hypothalamus	R18001261.BM1.BM1.14
	HC32	Hypothalamus	R18001262.BM2.BM2.2
	HD32	Hypothalamus	R18001263.BM3.BM3.16
	HA71	Cerebellar White Matter	R18001265.BM5.BM5.13
	HB71	Cerebellar White Matter	R18001261.BM1.BM1.2
	HC71	Cerebellar White Matter	R18001263.BM3.BM3.27

	tissue	sample
HD71	Cerebellar White Matter	R18001263.BM3.BM3.8
HA72	Cerebellar Grey Matter	R18001264.BM4.BM4.23
HB72	Cerebellar Grey Matter	R18001264.BM4.BM4.18
HC72	Cerebellar Grey Matter	R18001264.BM4.BM4.14
HD72	Cerebellar Grey Matter	R18001263.BM3.BM3.11

## Анализ вариабельности образцов в R (RNAseq\_human.R)

Посмотрим на распределение каунтов в первом образце

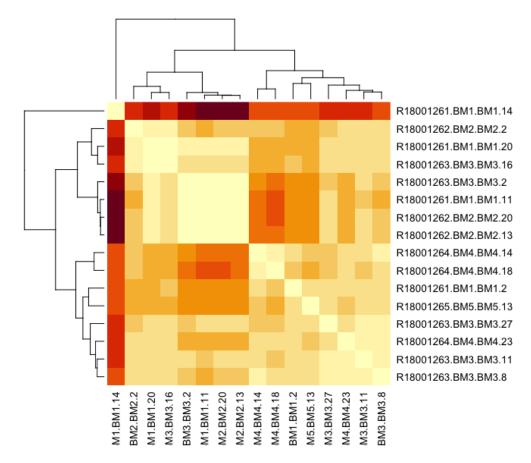




Распределение похоже на лог-нормальное

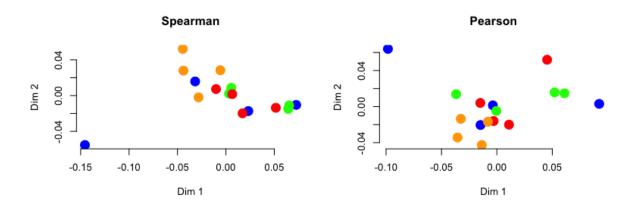
Сделаем отсечку - не менее 10 ридов на образец. В таком случае общее число генов составит - 20890

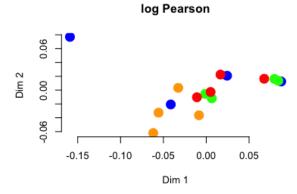
Оценим корреляцию между образцами (Спирмен)



Мы видим, что образцы из региона Prefrontal лучше всего скоррелированы между собой.

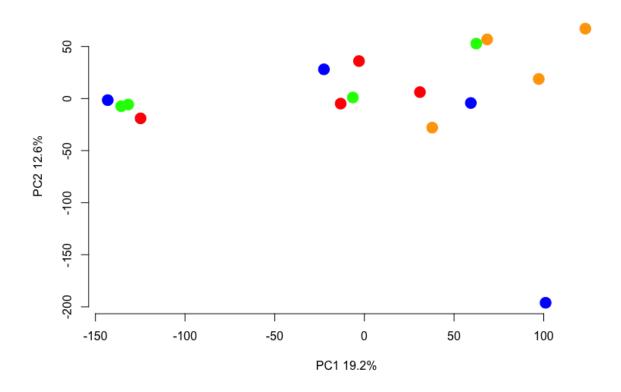
#### Multidimensional scaling (MDS)





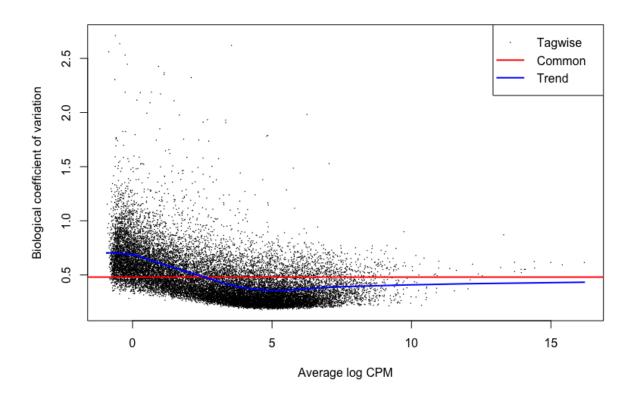
Мы видим, что самосогласованность между образцами слабая. Можно выделить регион Cerebellar Grey Matter (orange)

PCA



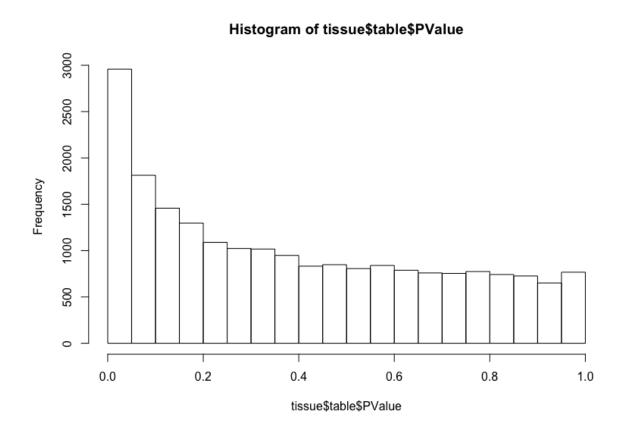
### Поиск DE генов

Зависимость вариабельности от средней экспресии (edgeR)



Строим GLM модель и считаем p-value

распределение p-value



Есть пик около 0, значит есть различающиеся гены

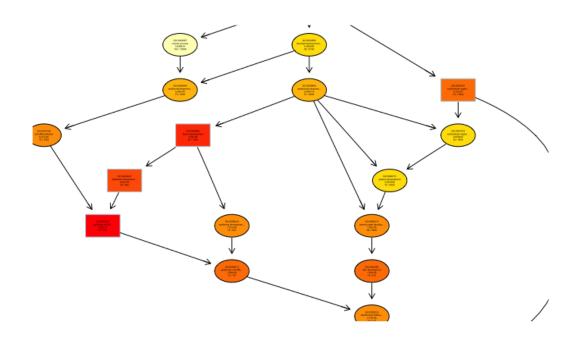
С четом поправки на множественное сравнение получается 255 гена, которые значимо меняются в разных регионах мозга

10 наиболее значимо меняющихся гена по всем типам тканей: ENSG00000185499 ENSG00000205221 ENSG00000205959 ENSG00000105976 ENSG00000185101 ENSG00000279058 ENSG00000136352 ENSG00000130701 ENSG00000124107 ENSG00000130176

### Посчитаем GO enrichment

Используем пакет topGO

GO.ID	Term	Annotated	Significant	Expected	result1
GO:0030855	epithelial cell differentiation	506	29	6.94	5.8e-11
GO:0009888	tissue development	1588	50	21.78	1.0e-08
GO:0060429	epithelium development	962	36	13.20	2.8e-08
GO:0031424	keratinization	45	8	0.62	1.5e-07
GO:0032501	multicellular organismal process	5846	114	80.20	5.7e-07
GO:0070268	cornification	38	7	0.52	7.2e-07
GO:0043588	skin development	214	14	2.94	1.5e-06
GO:0009887	animal organ morphogenesis	893	31	12.25	1.5e-06
GO:0009913	epidermal cell differentiation	157	12	2.15	1.7e-06
GO:0030216	keratinocyte differentiation	112	10	1.54	3.1e-06



Странный результат, нет GO terms связанных с нервной тканью