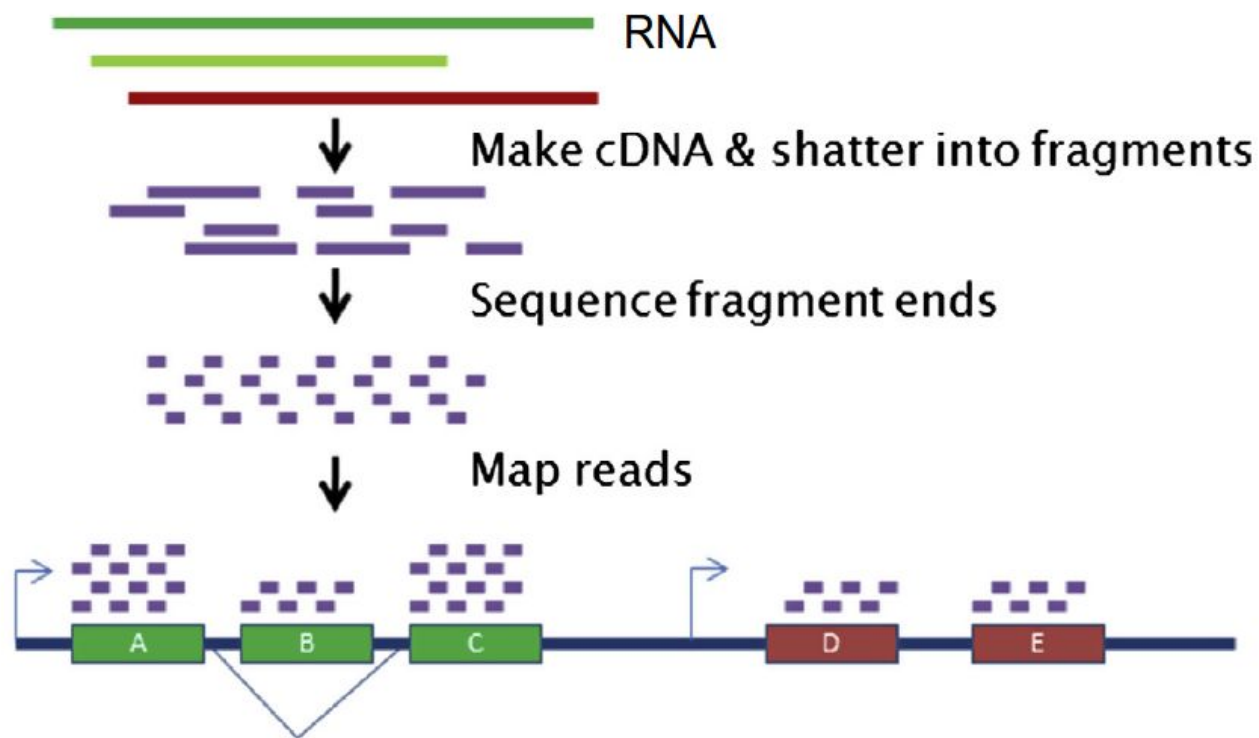


Анализ данных RNA-seq



Важно учесть общее количество ридов

sample



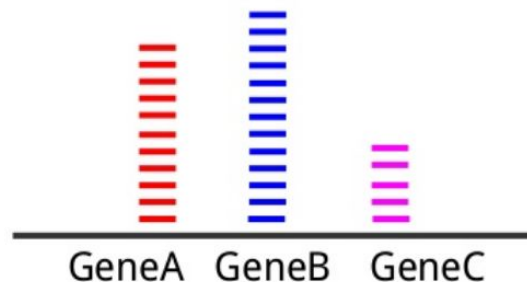
RNA-seq



sample



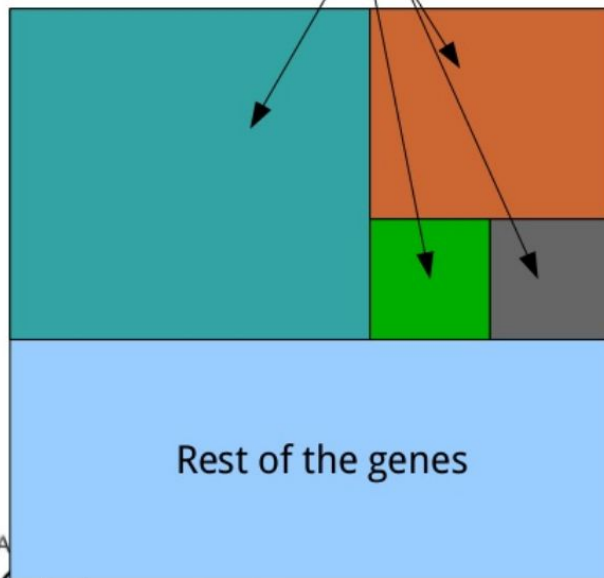
More RNA-seq



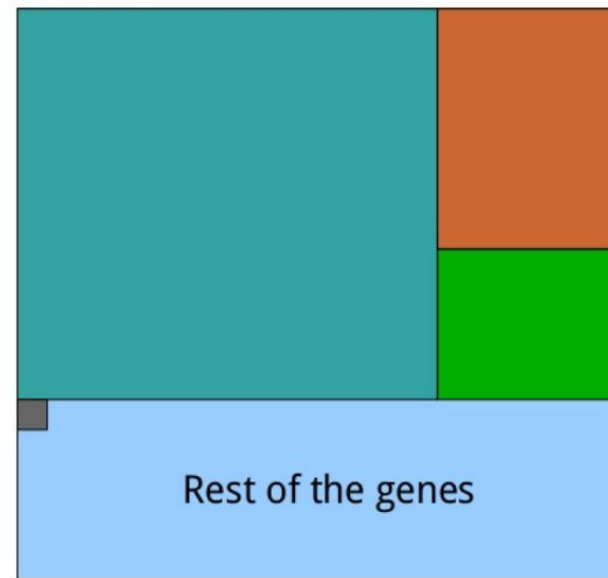
The **number of reads** is dependent on the total number of reads generated. If one library is sequenced to 20M reads, and another one to 40M, most genes will ~double their counts.

Schematically: when normalized on library size (square represent number of reads).

Few genes with enormous counts

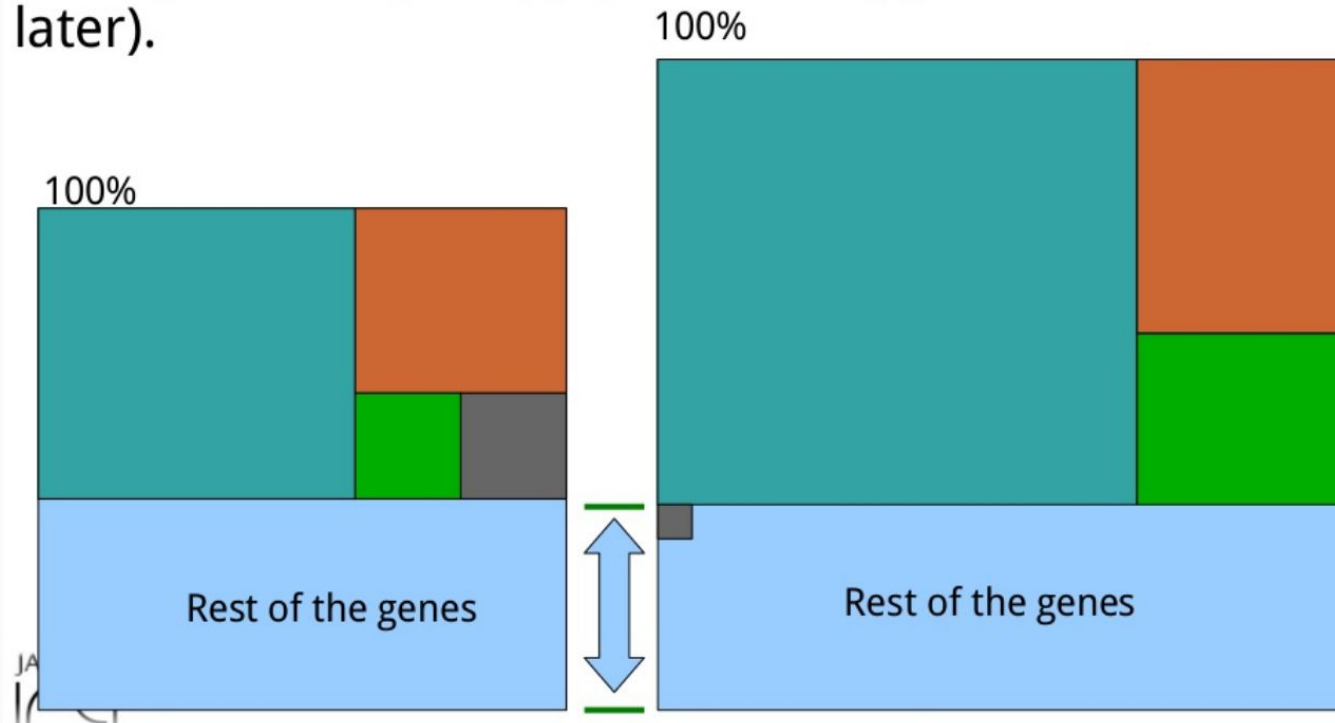


All counts for **library A**



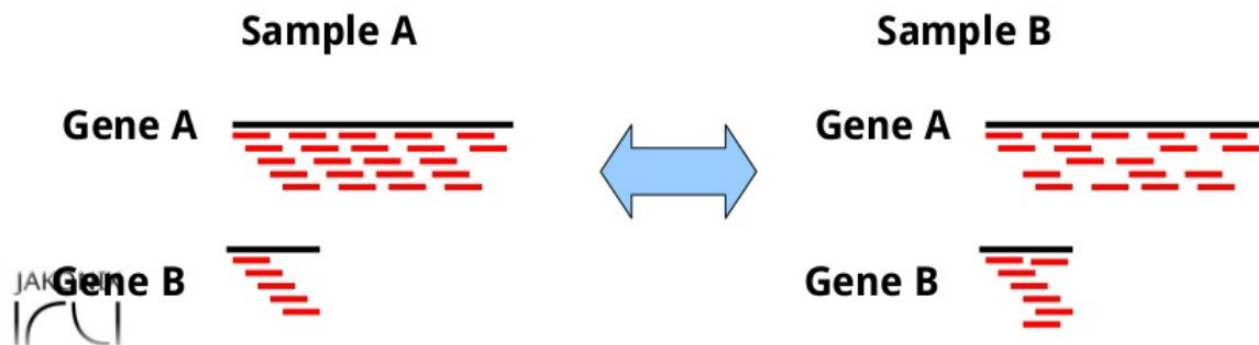
All counts for **library B**

Better normalization would be as shown below.
DESeq2 and EdgeR apply such an approach (see later).



“Longer transcripts generate more reads”

True! But the transcript **length does not differ between samples**. Since we are concerned with relative differences between samples, this needs no normalization (this story changes in case of absolute quantification).



Fragment per kilo base per million mapped reads (per upper-quartile)

$$\text{FPKM} = \frac{10^9 \times \text{number of reads mapped to the gene}}{\text{Number of reads mapped to all protein-coding genes} \times \text{length of the gene in base pairs}}$$

$$\text{FPKM} - \text{UQ} = \frac{10^9 \times \text{number of reads mapped to the gene}}{\text{The 75th percentile read count value for genes in the sample} \times \text{length of the gene in base pairs}}$$

Данные об экспрессии генов в здоровых и опухолевых тканях молочной железы

Gene Symbol	TCGA-A7-A0CE_tumor	TCGA-A7-A0CE_normal	TCGA-A7-A0CH_tumor	TCGA-A7-A0CH_normal
5S_rRNA	12.5504824437442	11.4060402970475	13.4971943828243	0
5_8S_rRNA	0	0	0	0
7SK	0	0	0	0
A1BG	9.91358818573394	11.1159316938797	11.2074527627992	11.2789351734531
A1BG-AS1	14.0711226965584	14.7048443946543	14.7778668198846	14.2980055549285
A1CF	8.65435384578248	8.37289043722157	5.58360491069903	0
A2M	20.6609983718851	22.6921921905997	19.8441845999705	22.4231170004858
A2M-AS1	13.7984092619639	13.8361513315943	12.4920451205339	15.062581364992
A2ML1	12.0624039151596	12.2747809480081	10.6495364489508	12.5062517976428
A2ML1-AS1	0	0	10.9630157769265	0
A2ML1-AS2	0	9.68516976997711	0	0
A2MP1	9.04550925781317	10.2617886786604	8.92219172832542	11.9156386084584
A3GALT2	10.7447853199172	9.27960604127287	0	10.5482348095854
A4GALT	15.5593427987835	17.9069925350423	15.6645692375837	18.4709098939268
A4GNT	11.7181492713697	10.2940292795217	7.99777860960778	8.758847216342
AA06	0	0	0	0
AAAS	17.6814381395663	17.8352819570118	17.6881135823049	17.8974716378576
AACS	15.8898567053666	16.2558527122921	16.0406788368626	17.213010047673
AACSP1	10.7860792274975	8.87806338500476	0	0

<https://yadi.sk/d/XHQjWo4t1aNpaA>

Данные об экспрессии микроРНК

isomiR	TCGA-A7-A0CE_normal	TCGA-A7-A0CE_tumor	TCGA-A7-A0D9_normal	TCGA-A7-A0D9_tumor
hsa-let-7a-2-3p 0 0	1.01810851413196	4.32503037322536	3.54761125098042	2.83742060675561
hsa-let-7a-2-3p 0 +1	0	0	0	0
hsa-let-7a-2-3p 0 +2	0	0	0	0
hsa-let-7a-2-3p 0 -1	0.597060063214157	3.8522141309232	0.593788934248154	0.822466083889712
hsa-let-7a-2-3p 0 -2	0.329198798682205	2.12450639661508	0	0.822466083889712
hsa-let-7a-2-3p 0 -3	0	0	0	0
hsa-let-7a-2-3p 0 -4	0	0	0	0
hsa-let-7a-2-3p 0 -5	0	0	0	0
hsa-let-7a-2-3p +1 0	0	0	0	0
hsa-let-7a-2-3p +1 -2	0	0	0	0
hsa-let-7a-2-3p +2 0	0	0	0	0
hsa-let-7a-2-3p -1 0	0	0	0	0
hsa-let-7a-2-3p -1 -1	0	1.08418647166452	0	0
hsa-let-7a-2-3p -1 -2	0	0	0	0
hsa-let-7a-2-3p -1 -3	0	0	0	0
hsa-let-7a-2-3p -1 -4	0	0	0	0
hsa-let-7a-2-3p -2 -1	0	0	0	0
hsa-let-7a-2-3p -2 -2	0	0	0	0
hsa-let-7a-2-3p -2 -5	0	0	0	0

<https://yadi.sk/d/DyMOjnlgQNSwBg>

- Пробегаясь по всем парам генов;
- Для каждой пары смотрим на долю здоровых образцов, в которых экспрессия первого гена больше второго. То же самое среди опухолевых образцов;
- Если первая доля большая, а вторая малая (или наоборот), то нашлась пара генов, меняющих знак неравенства;
- Содержательно, это переход к внутриобразцовым рангам;
- Можно попробовать ввести score, учитывающий величину разницы в экспрессии;
- Из полученных генов строим граф взаимодействий.

Спасибо за внимание

<https://www.slideshare.net/jakonix/part-1-of-rnaseq-for-de-defining-the-goal>