

# Transcriptomic and epigenomic biomarkers associated with smoking habits in patients with lung or laryngeal cancer

Final project presentation

## **Group 5**

Mia Anscheit

Matanat Mammadli

Friederike Wohlfarth

# Changes in project topic

## Scientific question

Transcriptomic and epigenomic biomarkers  
associated with cancer subtypes



**Transcriptomic and epigenomic biomarkers  
associated with smoking habits**

## Cancer types

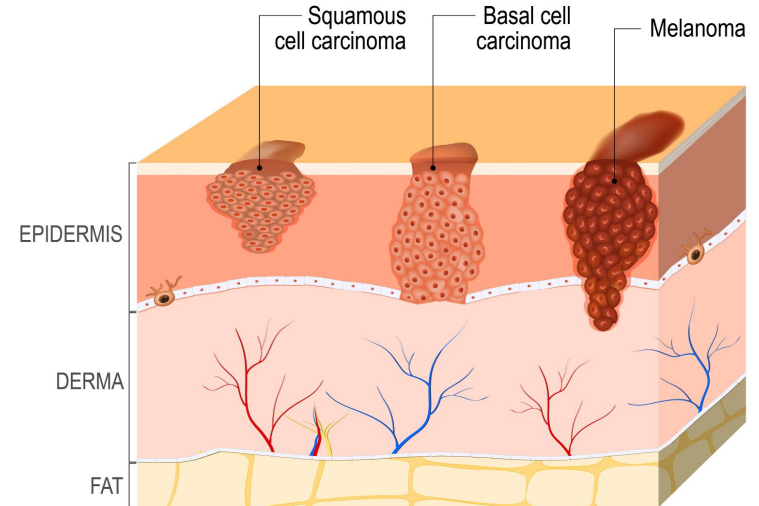
Kidney and laryngeal cancer



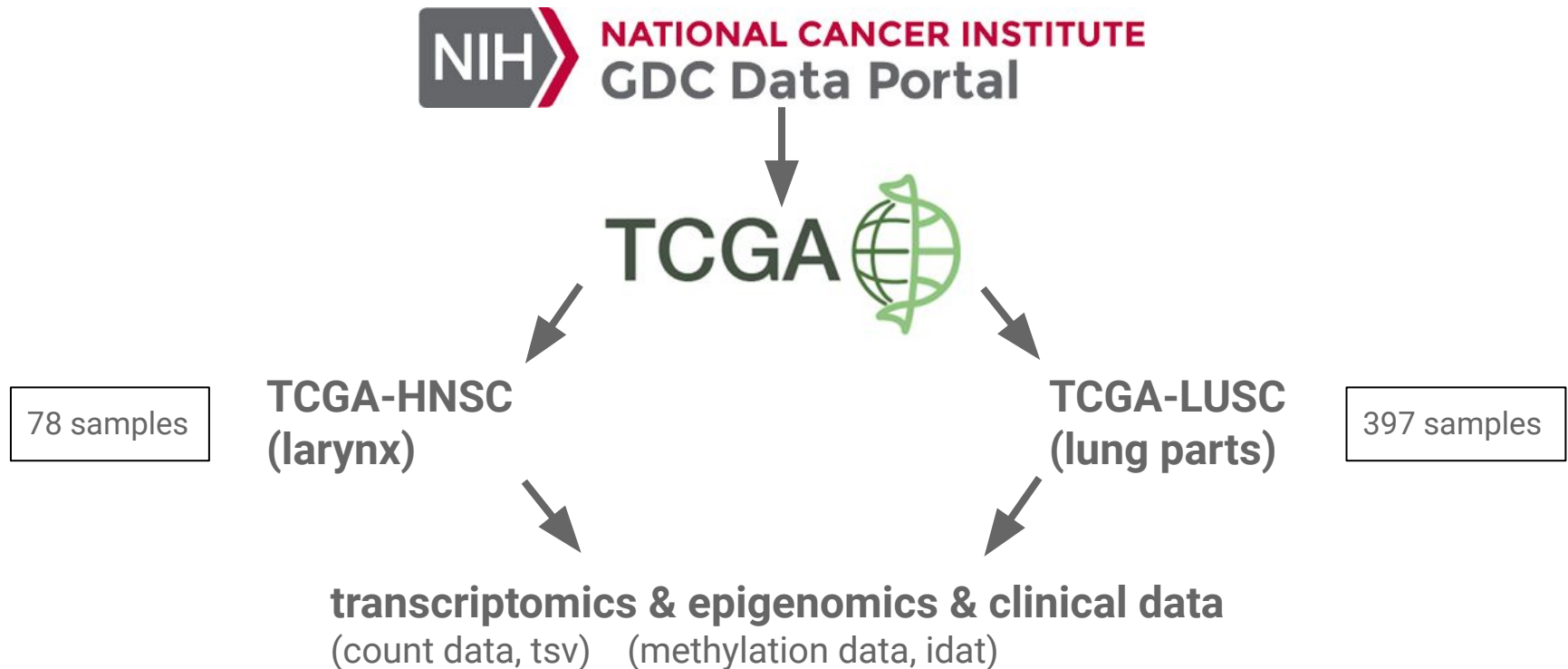
**Lung and laryngeal cancer**

# Squamous cell carcinoma

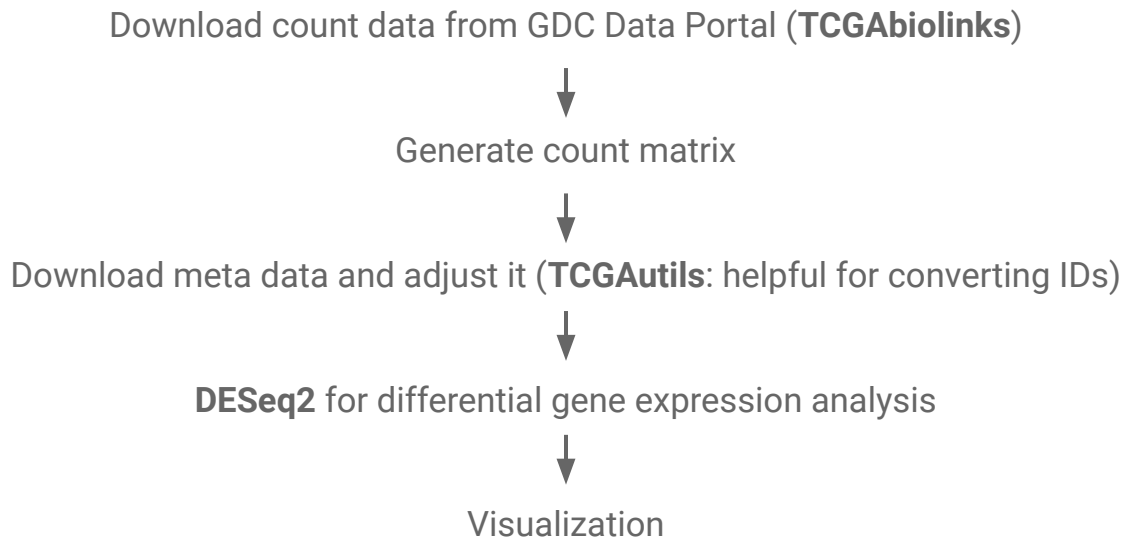
- squamous cells form surface of skin as well as interior surface of most organs
- main reason for squamous cell carcinoma in respiratory organs: smoking



# Data



# Differential gene expression analysis



# Type of analyses

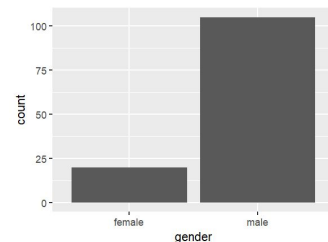
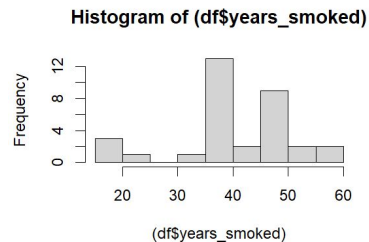
Differential gene expression between:

- healthy and cancer cells in lung/larynx tissue separately
- lung and larynx cancer
- different smoking habits (cigarettes/day or packages/year) per cancer type

# Exploratory data analysis – datasets

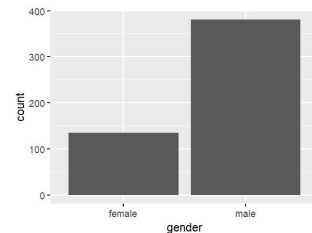
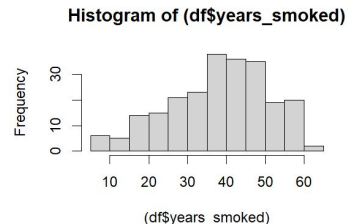
## TCGA-LUSC - lung

- matrix of 516 samples, 60.660 transcripts
- correcting for outliers (none) and normalization as well as filtering -> 48 normal tissue samples, 468 tumor samples in count matrix
- DEA using tcgabiolinks found 9619 DEGs

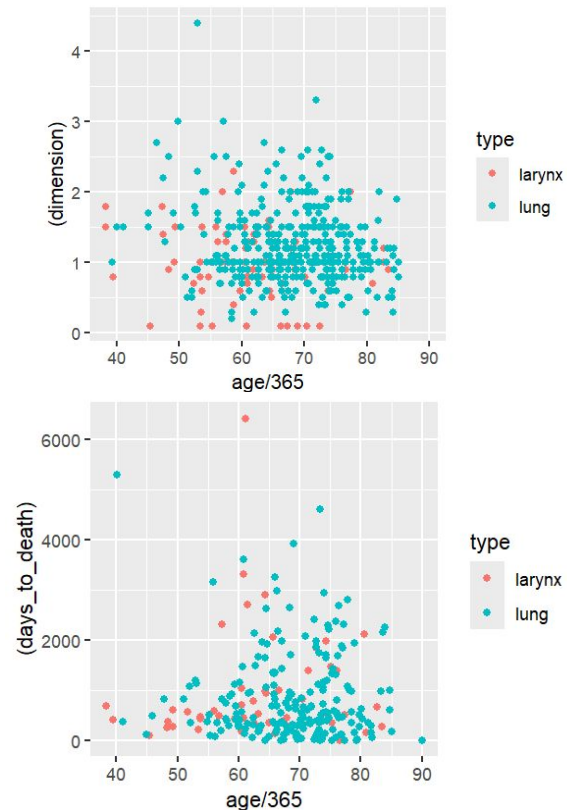
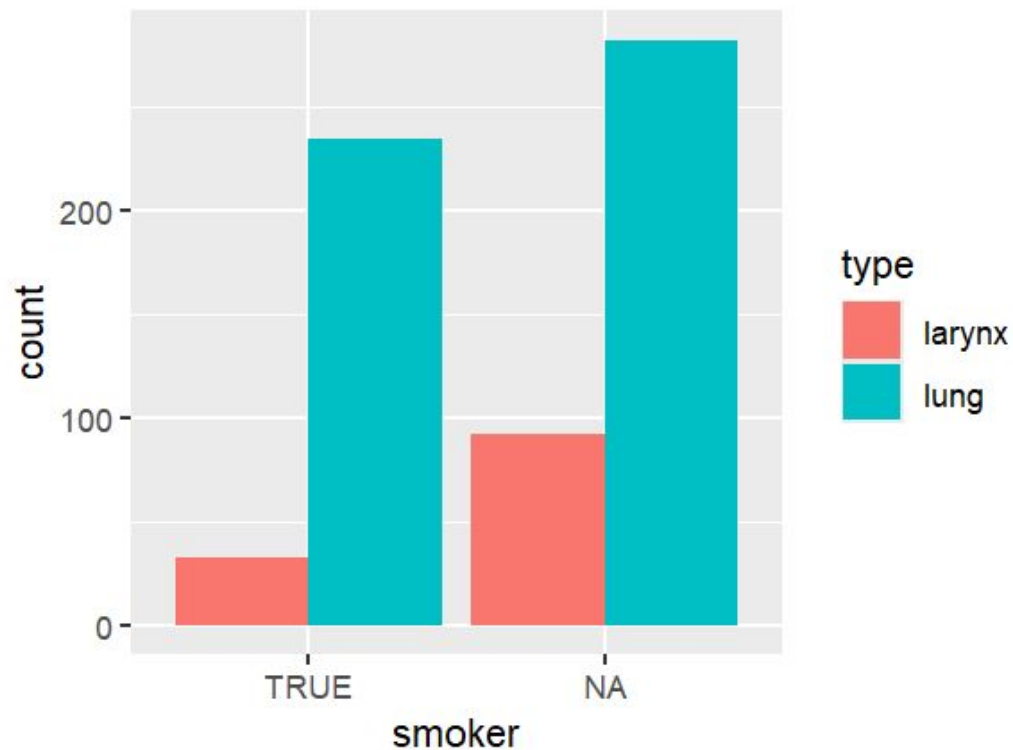


## TCGA-HNSC - larynx

- matrix of 125 samples, 60.660 transcripts
- 12 normal tissue samples, 113 tumor samples
- DEA found 6995 DEGs

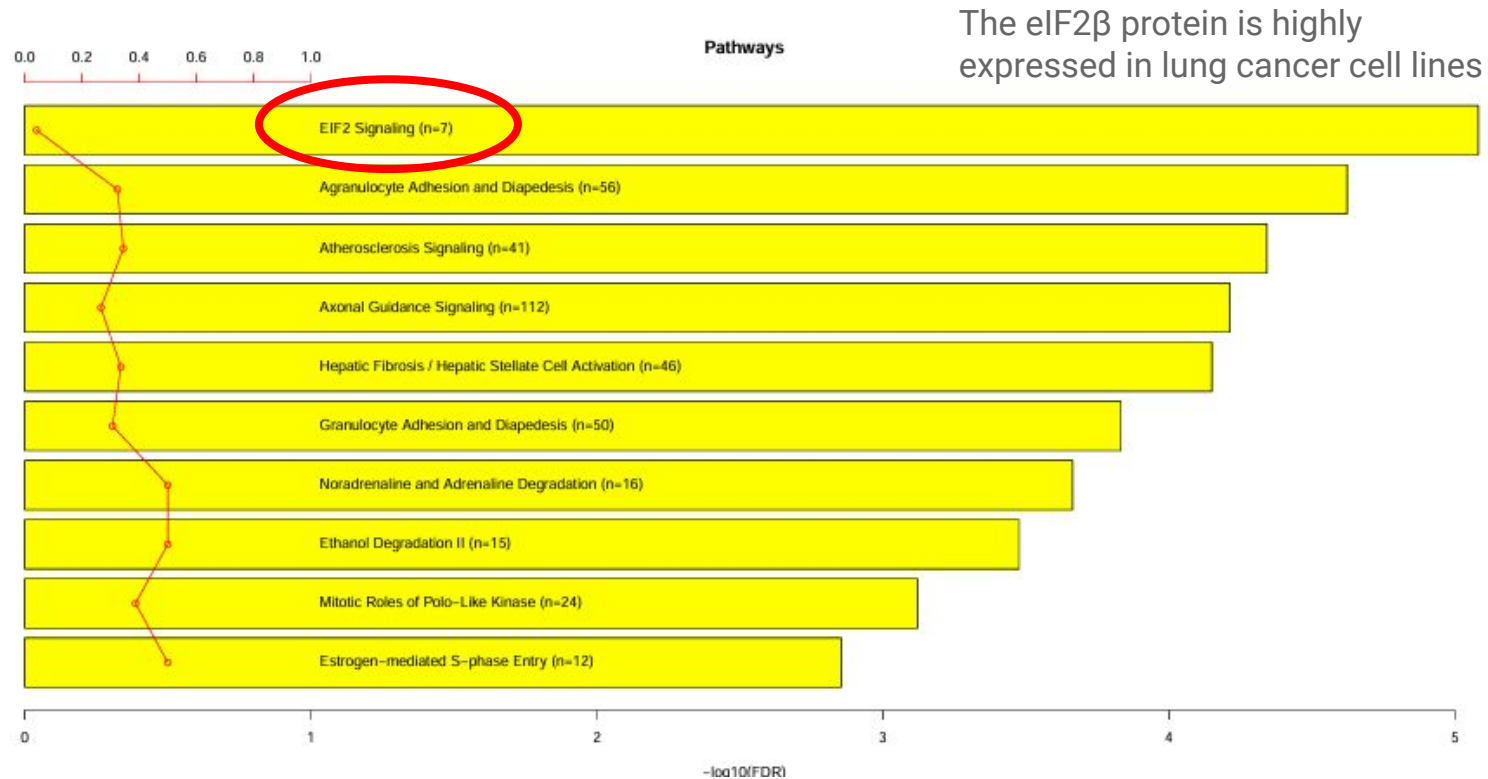


# Exploratory data analysis – datasets

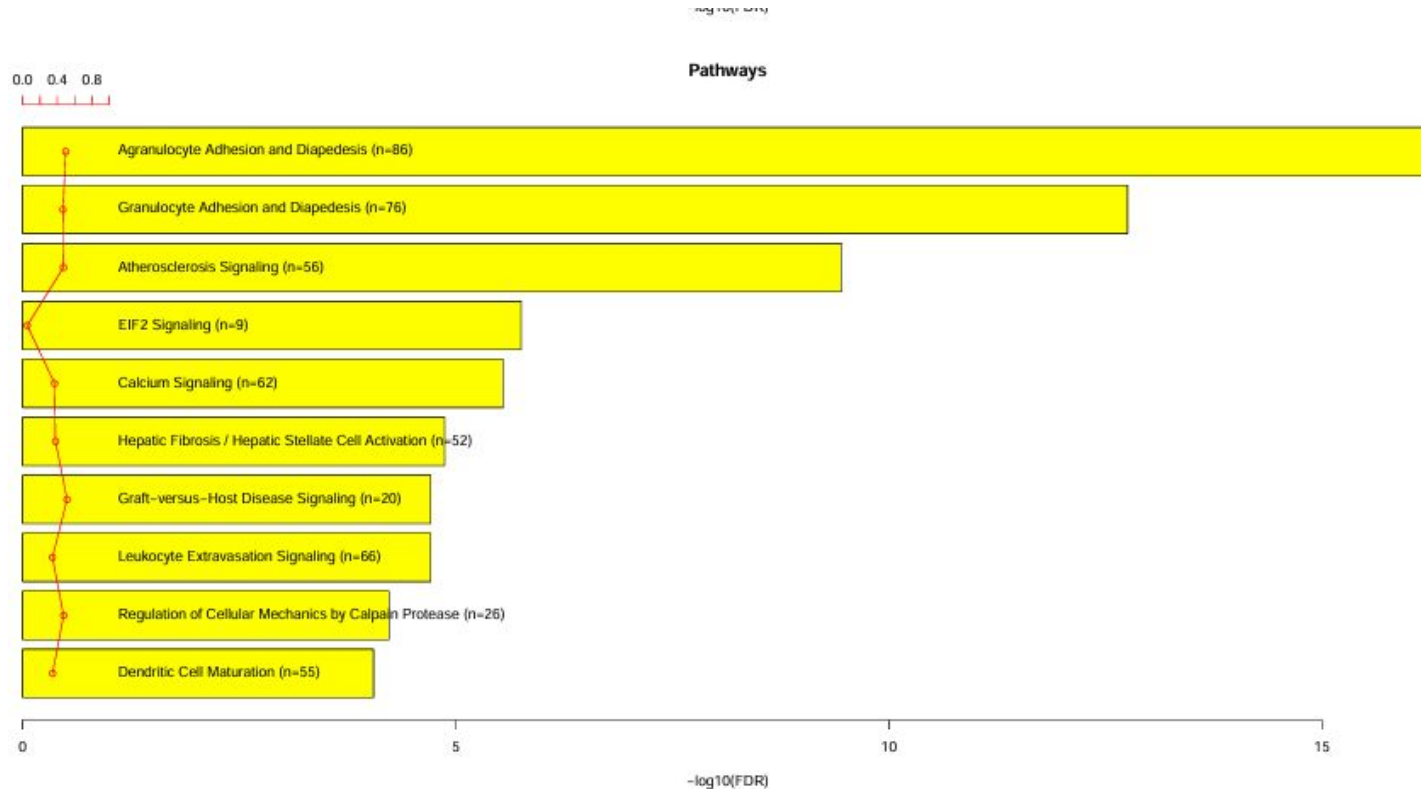




# Exploratory data analysis – lung dataset



# Exploratory data analysis – larynx dataset

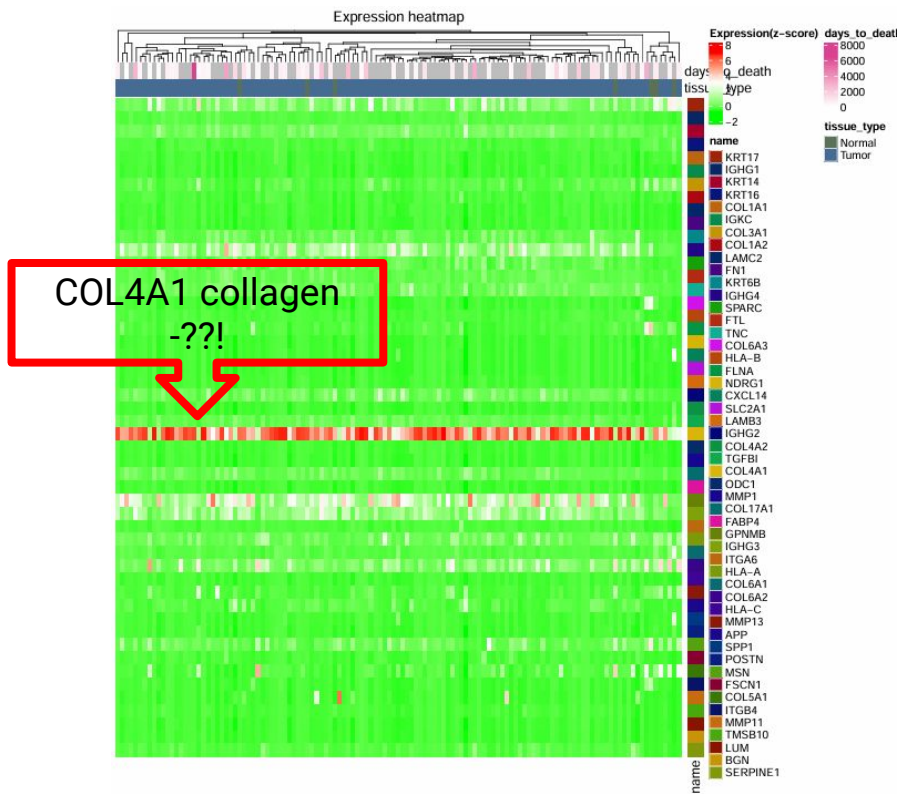


# Exploratory data analysis – lung dataset



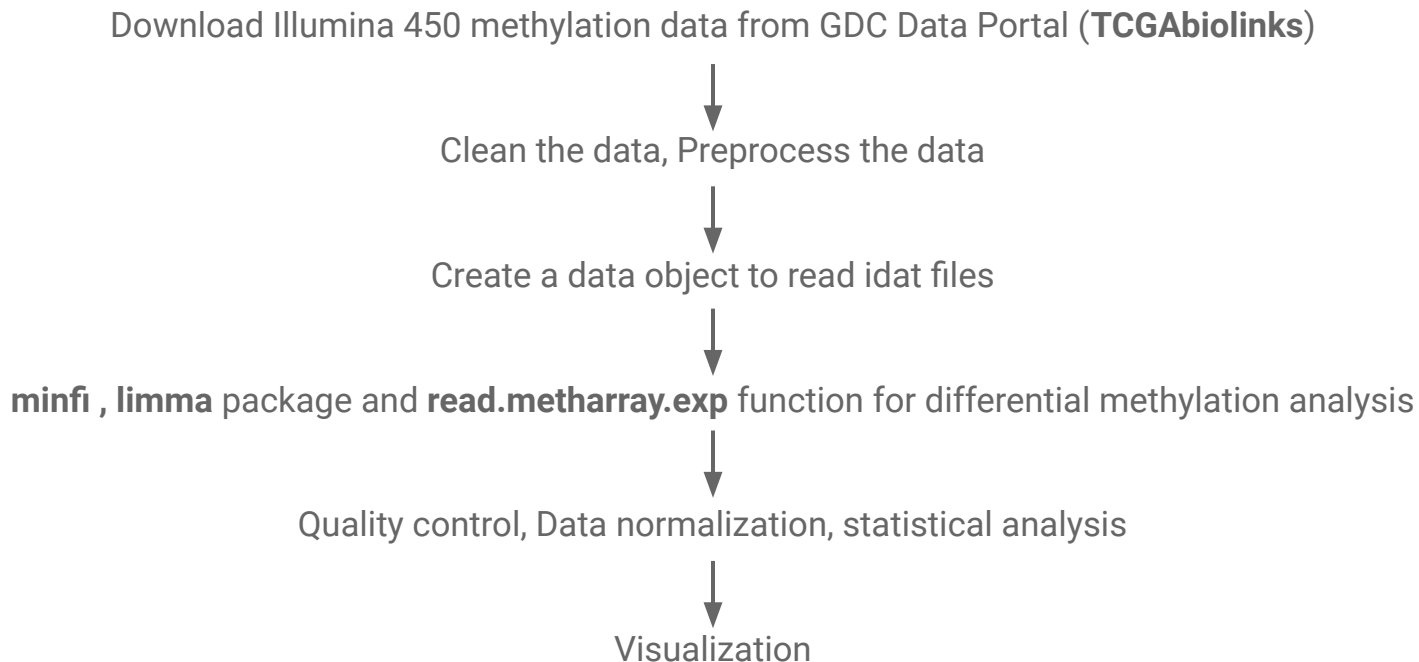
- Top 5:
  - KRT5
  - KRT6A
  - KRT17
  - IGHG1
  - IGRC
- Keratins are used as tumor markers in lung cancer
- KRT5 is overexpressed in lung cancer cells
- immunoglobulines specific for squamous lung cancer

# Exploratory data analysis – larnynx dataset



- Top 5:
  - KRT17
  - IGHG1
  - KRT14
  - KRT16
  - COL1A1

# Differential methylation analysis



# Differential methylation analysis

Data was too large, we downloaded 1160 idat files (red and green) for **larynx cancer**

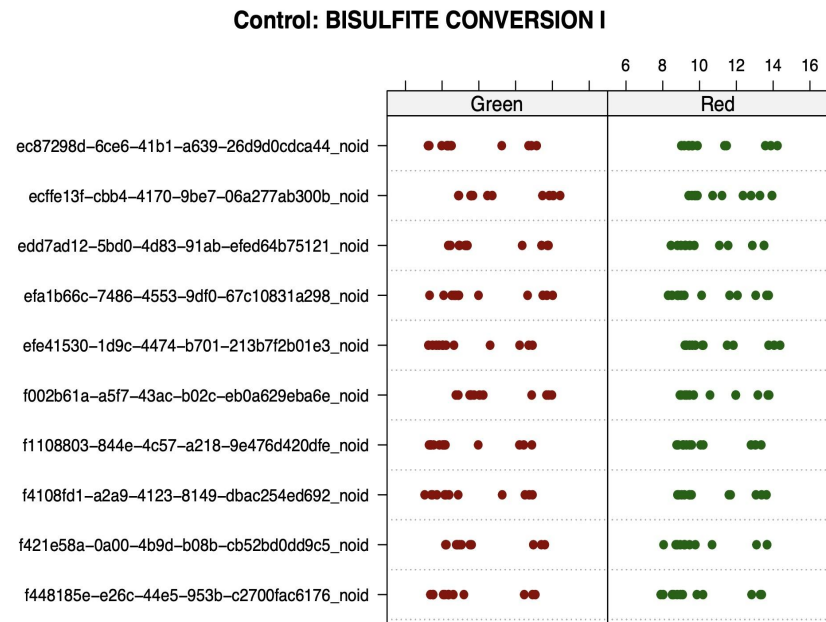
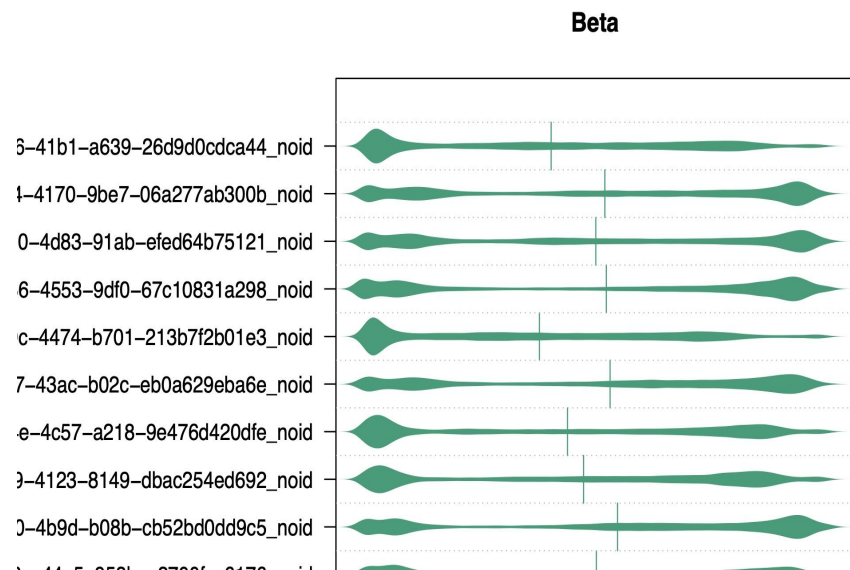


what we did: we divided data into **green** and **red** channel idat pairs, and from 580 pairs we randomly selected and kept 290 (half of it).

We did same steps for **lung cancer** data, we kept 191 pairs at the end (from initial 764 red and green idat files).

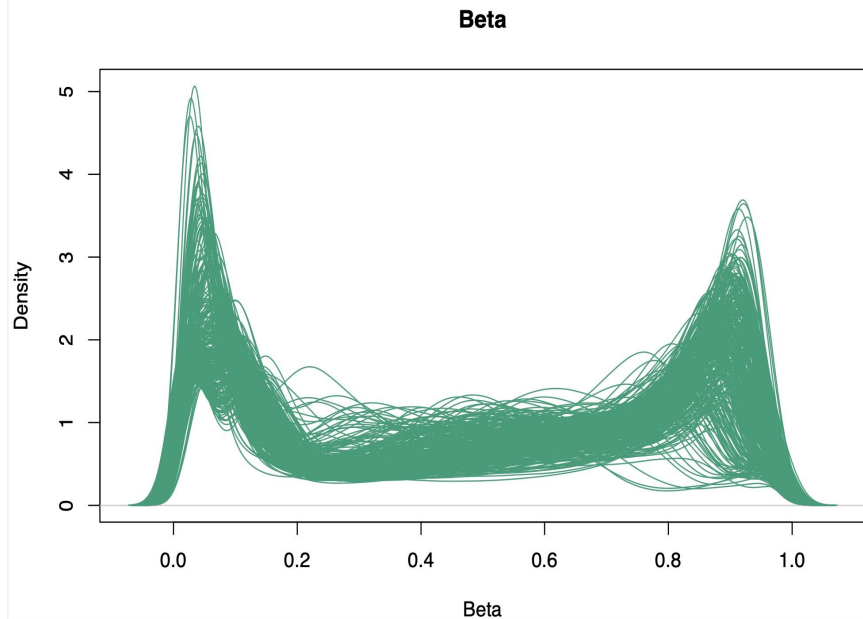
# Differential methylation analysis

Quality control (qcReport(methylation\_data)) results (larynx)

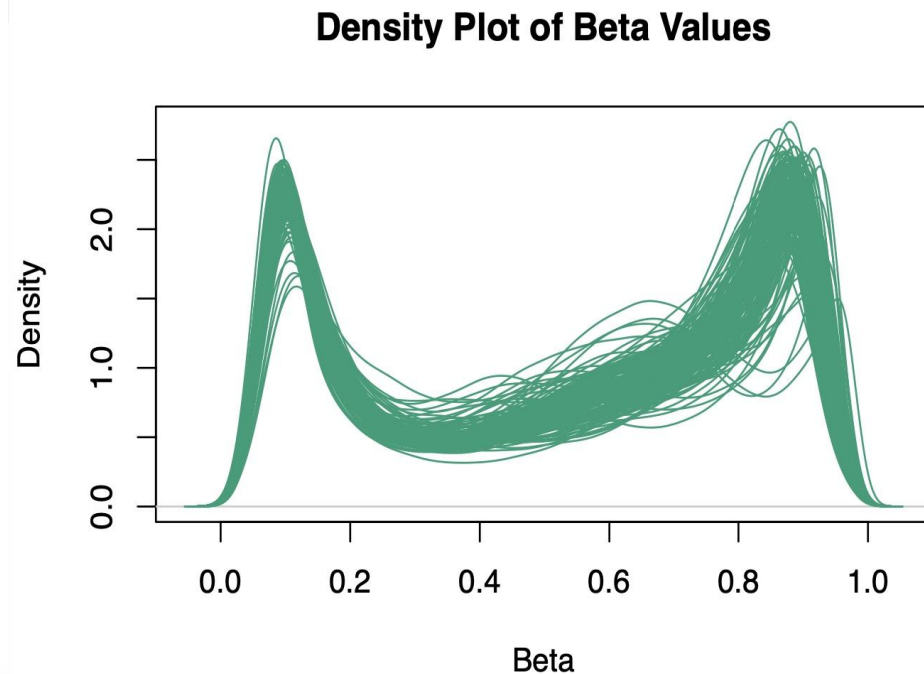


# Differential methylation analysis

for larynx cancer



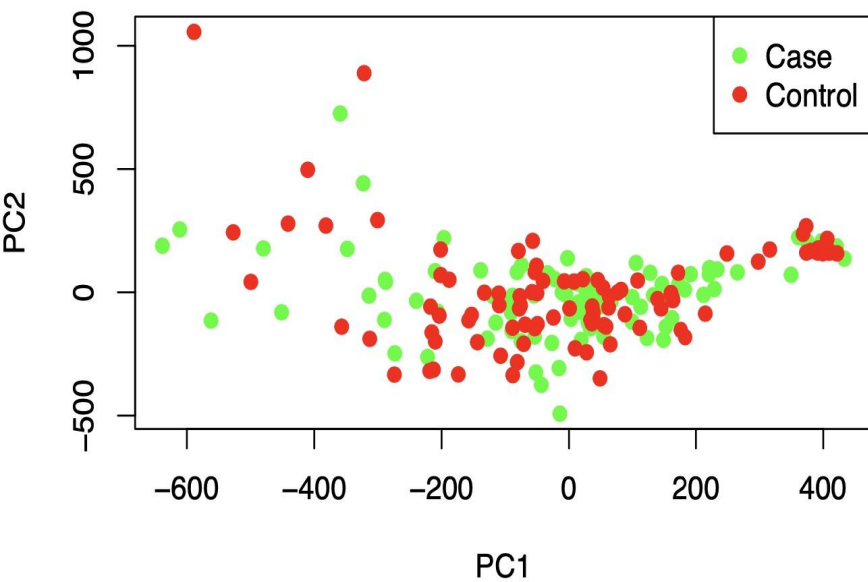
for lung cancer





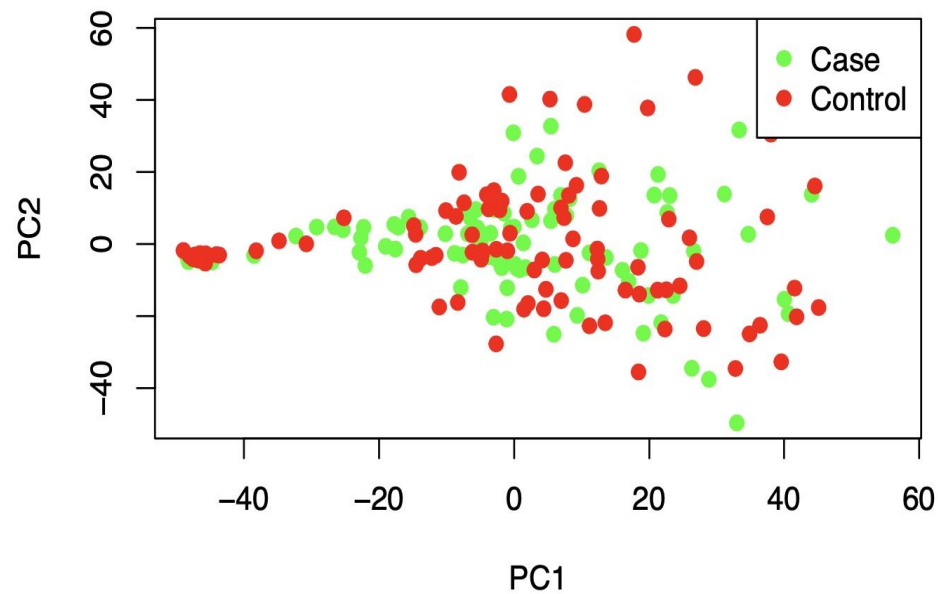
for larynx cancer

### PCA of Beta Values



for lung cancer

### PCA of Beta Values



# Differential methylation analysis

## Data normalization steps (preprocessQuantile()):

- Mapping to Genome
- Fixing Outliers
- Quantile Normalizing

```
methylation_data_normalized
class: GenomicRatioSet
dim: 485512 290
metadata(0):
assays(2): M CN
rownames(485512): cg13869341 cg14008030 ... cg08265308
cg14273923
rowData names(0):
colnames(290): 01d6d8c3-7c94-4f96-aa7b-a2955d84ff36_noid
01dd514a-fef7-4b86-ab98-ca690970cd95_noid ...
ff8494b7-2651-44d9-b162-0bbc16e87598_noid
ff93e589-ca51-43f3-b988-b2bfacf6b4b9_noid
colData names(3): xMed yMed predictedSex
Annotation
array: IlluminaHumanMethylation450k
annotation: ilmn12.hg19
Preprocessing
Method: Raw (no normalization or bg correction)
minfi version: 1.46.0
Manifest version: 0.4.0
```

# Differential methylation analysis

```
betas <- getBeta(methylation_data_normalized)  
for lung cancer
```

99c03e8f-0511-45fd-97e8-a9108c6d95ae\_noid

cg13869341 0.85418382

cg14008030 0.63133197

cg12045430 0.21507860

cg20826792 0.50569133

cg00381604 0.21111381

99e0dd91-2bd6-4889-a4ea-0c50e974cd7a\_noid

cg13869341 0.86718232

cg14008030 0.74298201

cg12045430 0.23023874

cg20826792 0.41454710

cg00381604 0.21548686

9c05da65-6e66-47f7-83b1-c83db4a7fe94\_noid

cg13869341 0.875499954

cg14008030 0.674002816

cg12045430 0.215711182

cg20826792 0.409132893

cg00381604 0.119837133

# Differential methylation analysis

Statistical analysis results (with limma, fit, lmFit, eBayes etc.)

results

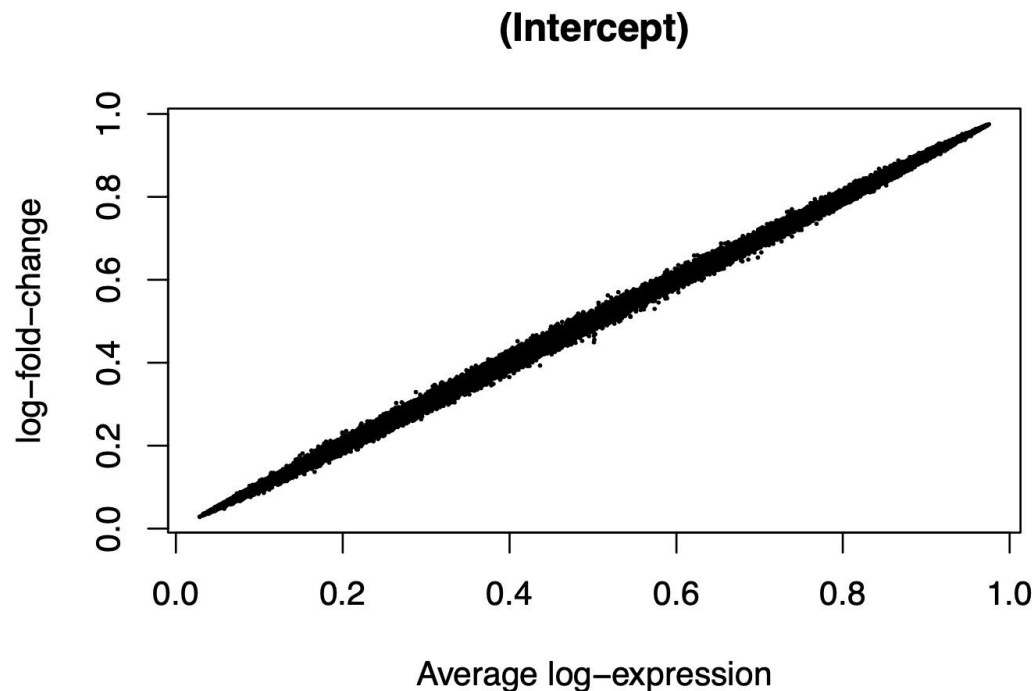
	logFC	AveExpr	t	P.Value	adj.P.Val	B
cg09869144	0.9689744	0.9685472	1322.4769	0	0	1253.037
cg03522766	0.9713235	0.9715661	1311.5113	0	0	1250.647
cg06096175	0.9750521	0.9751767	1304.7310	0	0	1249.160
cg06884679	0.9729082	0.9733130	1295.3970	0	0	1247.099
cg16385941	0.9725823	0.9731205	1204.5779	0	0	1226.214
cg22342925	0.9730781	0.9736369	1198.6808	0	0	1224.803
cg20306425	0.9670741	0.9668850	1190.3632	0	0	1222.800
cg06671654	0.9727307	0.9720465	1185.6003	0	0	1221.647
cg02008951	0.9666592	0.9671985	1174.1436	0	0	1218.854

.....

# Differential methylation analysis

## Visualization results

for larynx cancer



# Next steps

- detailed DEA with Deseq2 package with MA and volcano plots
- perform clustering with **TCGAanalyze\_Clustering** and annotate cluster
- perform survival analysis for cluster
- plot heatmaps with annotated cluster information, DEA for pairs of clusters?
- clusters in accordance with tissue site or other annotation?
- Principal Component Analysis plot for differentially expressed genes
- **TCGAvisualize\_starburst**: Integration of gene expression and DNA methylation data