# Introduction to Machine Learning

Nov 01 2025

Phuc-Loi Luu, PhD

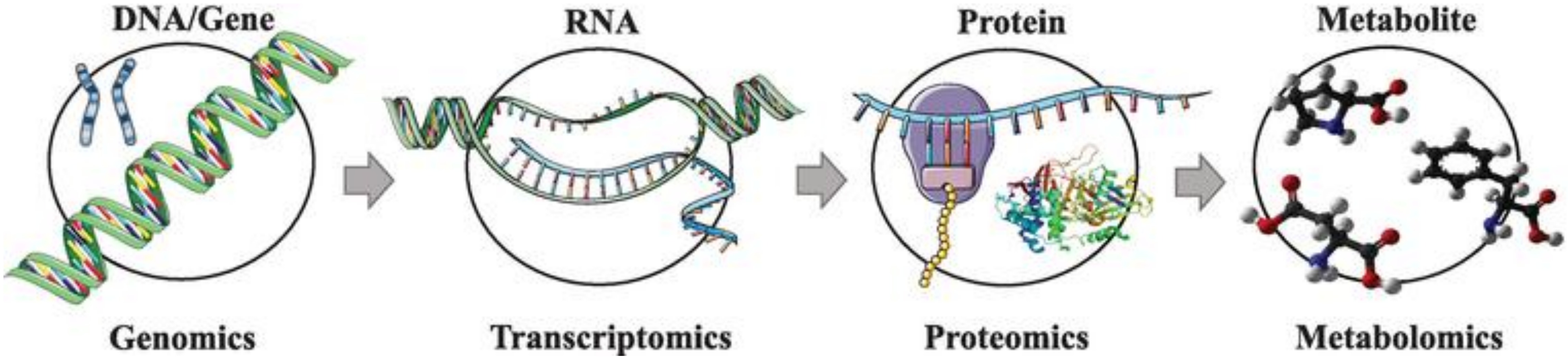
Email: luu.p.loi@googlemail.com

Zalo: 0901802182

# Content

- What are data?
- What do we attend this course for?
- What are we going to learn?
- How are we going to learn?

# Central Dogma



PAH gene

Ref …ATCGAT…

P1 …AACGAT…

**NM_000277.3(PAH):c.971T>A**

PAH mRNA

Ref …AUCGAU…

P1 …AACGAU…

**NM_000277.3(PAH):c.971T>A**

PAH protein

Ref …**Ile**-Asp…

P1 …**Asn**-Asp…

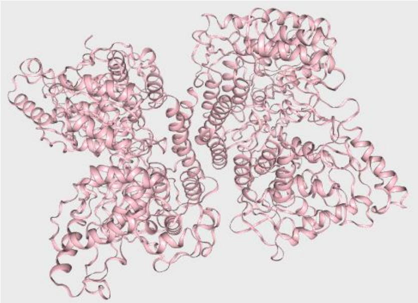
**NM_000277.3(PAH):p.Ile324Asn**

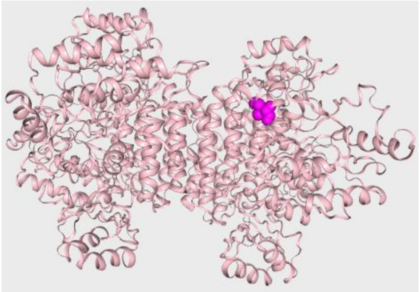
**PAH**

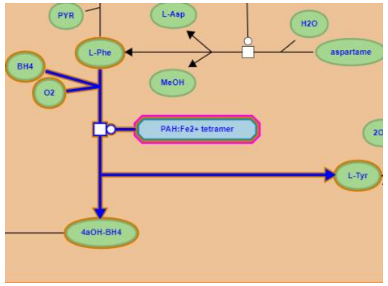
Ref Phe → Tyr

**PAH**

P1 Phe ↛ Tyr





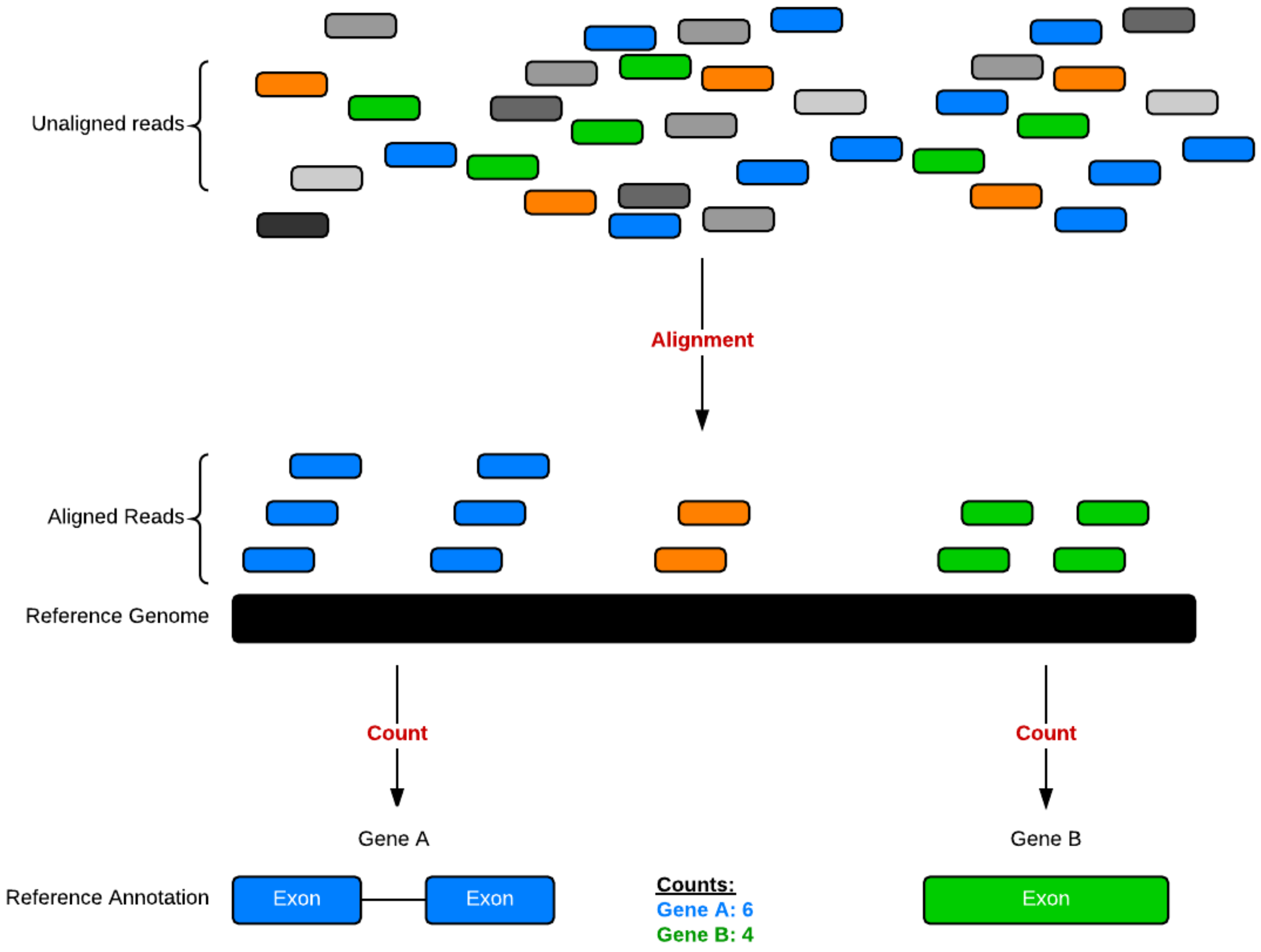


10.5772/intechopen.71769

# Genomic Data

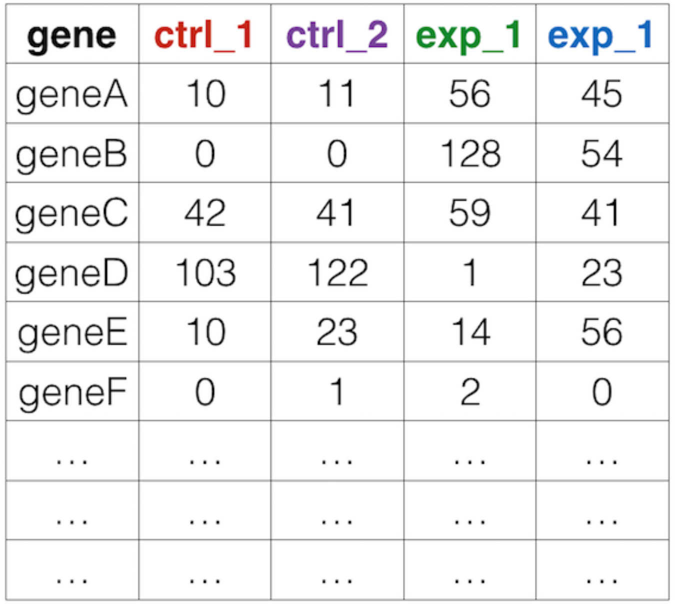
- **Genomics** (SNP microarray, CNV microarray and long or short read DNA-seq/WGS)
- **Transcriptomics** (microarray, bulk RNA-seq, single-cell RNA-seq and spatial transcriptomics)
- **Proteomics** (protein microarrays and mass spectrometry)
- **Metabolomics** (mass spectrometry and Nuclear Magnetic Resonance Spectroscopy)
- **Epigenomics**
- **Methylomics** (methylation microarray, WGBS, EM-seq and long-read sequencing)
- **Metagenomics** (long or short read DNA-seq)

# How to generate genomic data: RNA-seq



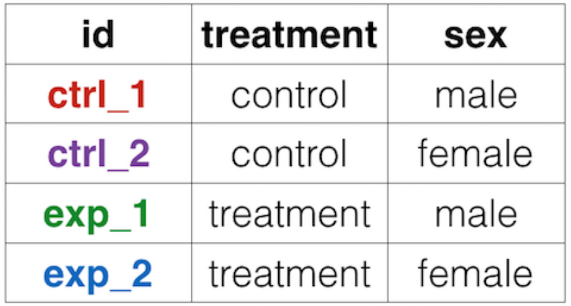
# RNA-seq count table

## countData



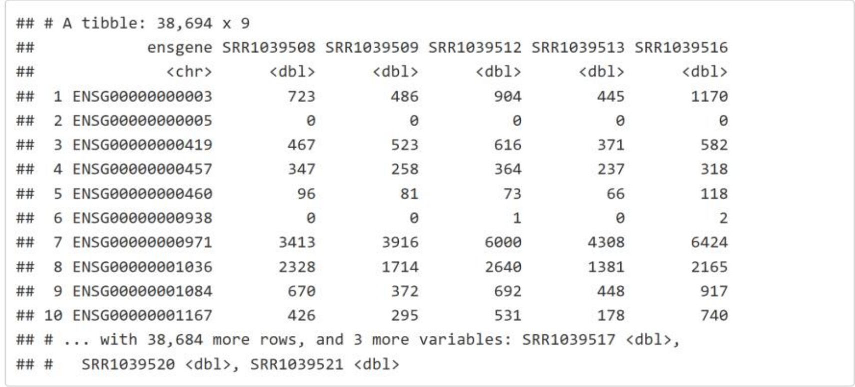
| gene | ctrl_1 | ctrl_2 | exp_1 | exp_1 |
|---|---|---|---|---|
| geneA | 10 | 11 | 56 | 45 |
| geneB | 0 | 0 | 128 | 54 |
| geneC | 42 | 41 | 59 | 41 |
| geneD | 103 | 122 | 1 | 23 |
| geneE | 10 | 23 | 14 | 56 |
| geneF | 0 | 1 | 2 | 0 |
| ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... |
| ... | ... | ... | ... | ... |

## colData



| id | treatment | sex |
|---|---|---|
| ctrl_1 | control | male |
| ctrl_2 | control | female |
| exp_1 | treatment | male |
| exp_2 | treatment | female |

Sample names:
ctrl_1, ctrl_2, exp_1, exp_2



```
## # A tibble: 38,694 x 9
##            ensgene SRR1039508 SRR1039509 SRR1039512 SRR1039513 SRR1039516
##              <chr>      <dbl>      <dbl>      <dbl>      <dbl>      <dbl>
##  1 ENSG00000000003        723        486        904        445       1170
##  2 ENSG00000000005          0          0          0          0          0
##  3 ENSG00000000419        467        523        616        371        582
##  4 ENSG00000000457        347        258        364        237        318
##  5 ENSG00000000460         96         81         73         66        118
##  6 ENSG00000000938          0          0          1          0          2
##  7 ENSG00000000971       3413       3916       6000       4308       6424
##  8 ENSG00000001036       2328       1714       2640       1381       2165
##  9 ENSG00000001084        670        372        692        448        917
## 10 ENSG00000001167        426        295        531        178        740
## # ... with 38,684 more rows, and 3 more variables: SRR1039517 <dbl>,
## #   SRR1039520 <dbl>, SRR1039521 <dbl>
```

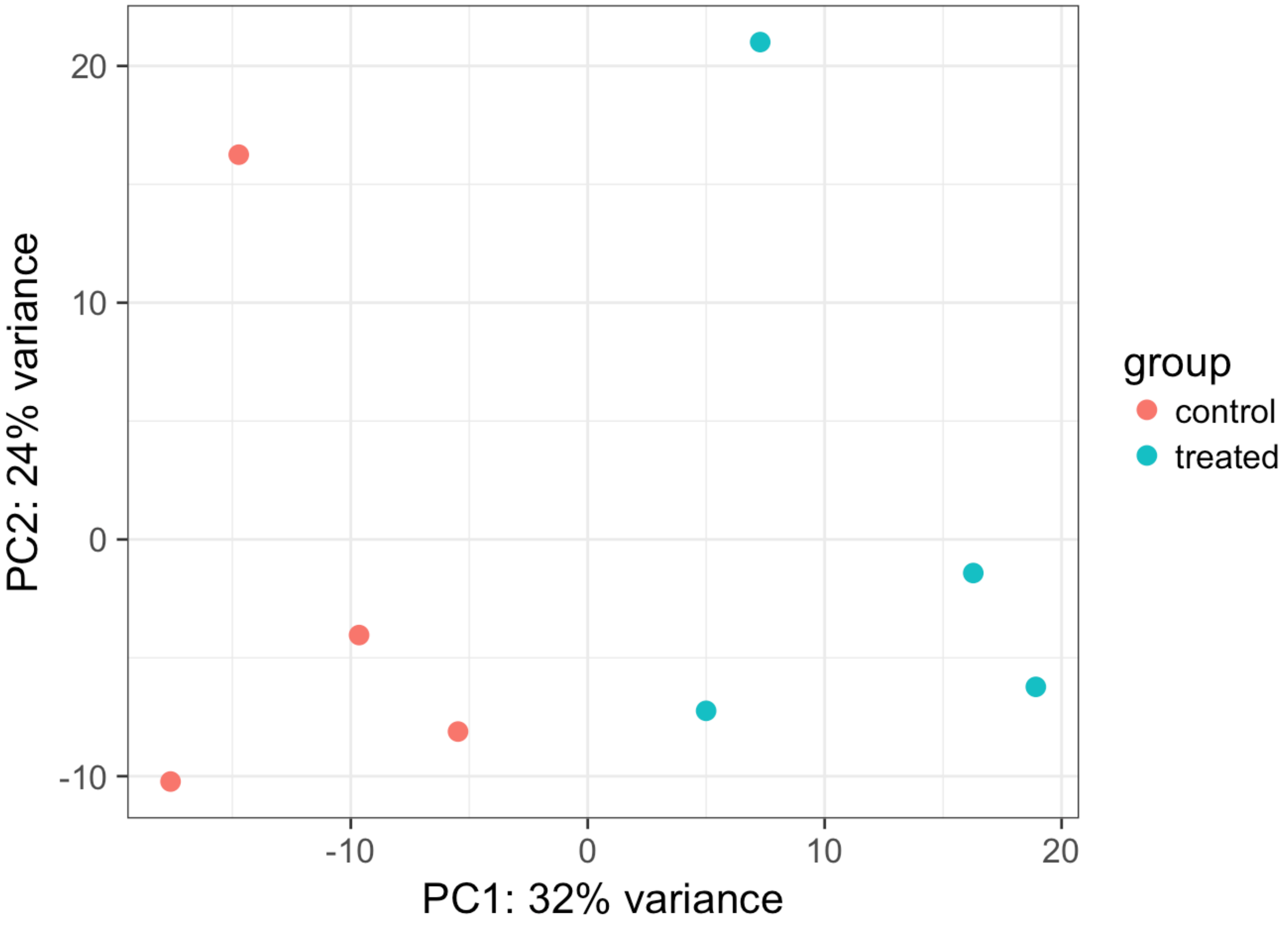
**countData** is the count matrix
(number of reads mapping to each gene for each sample)

**colData** describes metadata about the *columns* of countData

**First column of colData must match column names of countData (-1st)**

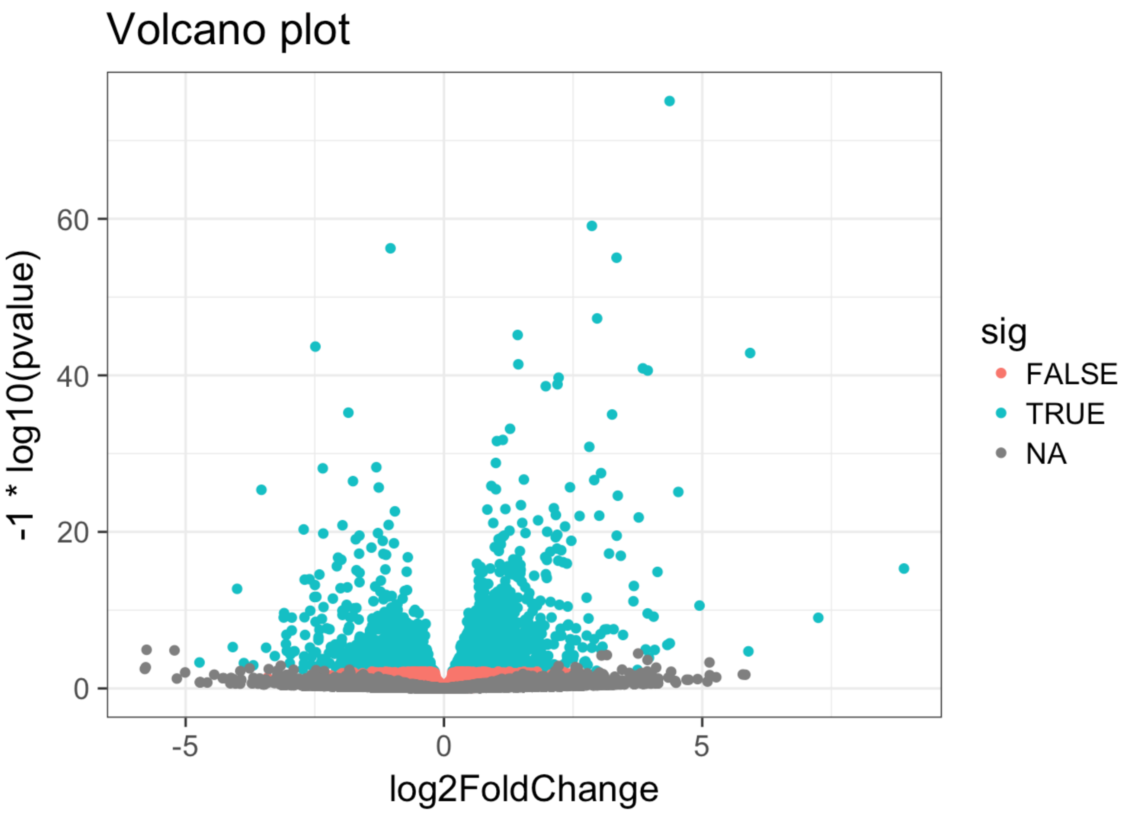
https://bioconnector.github.io/workshops/r-rnaseq-airway.html

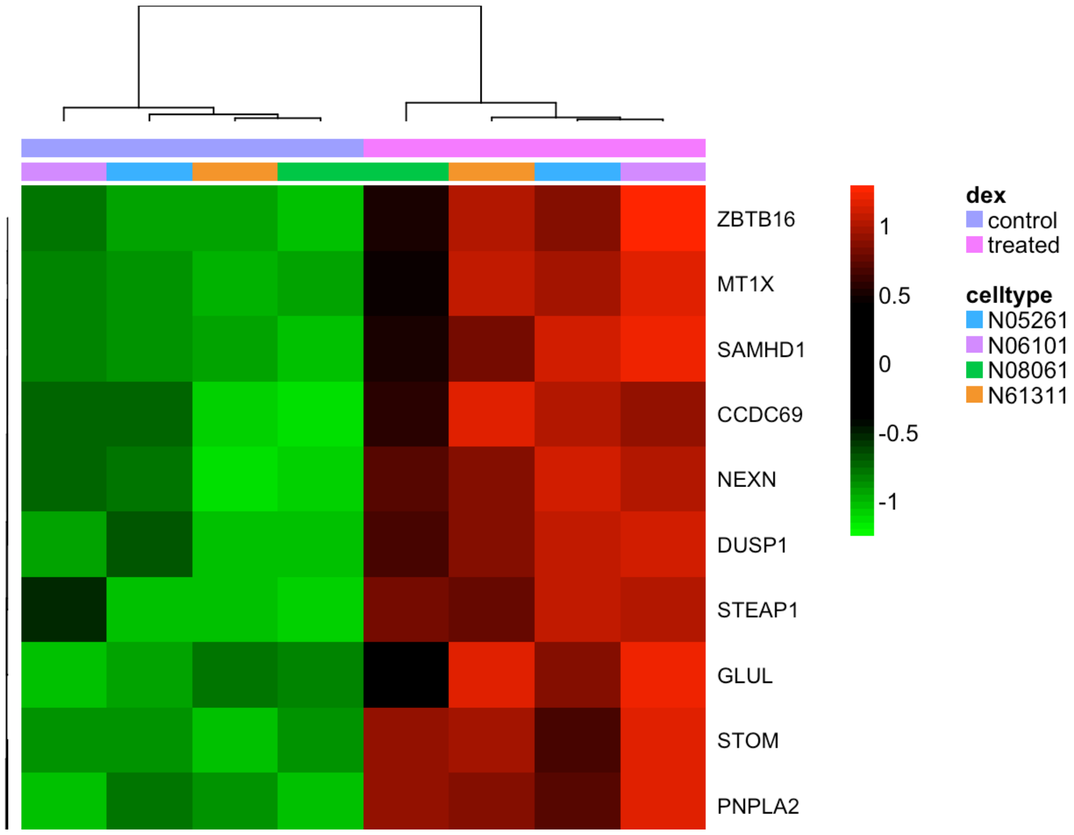
# RNA-seq Downstream Analysis



https://bioconnector.github.io/workshops/r-rnaseq-airway.html

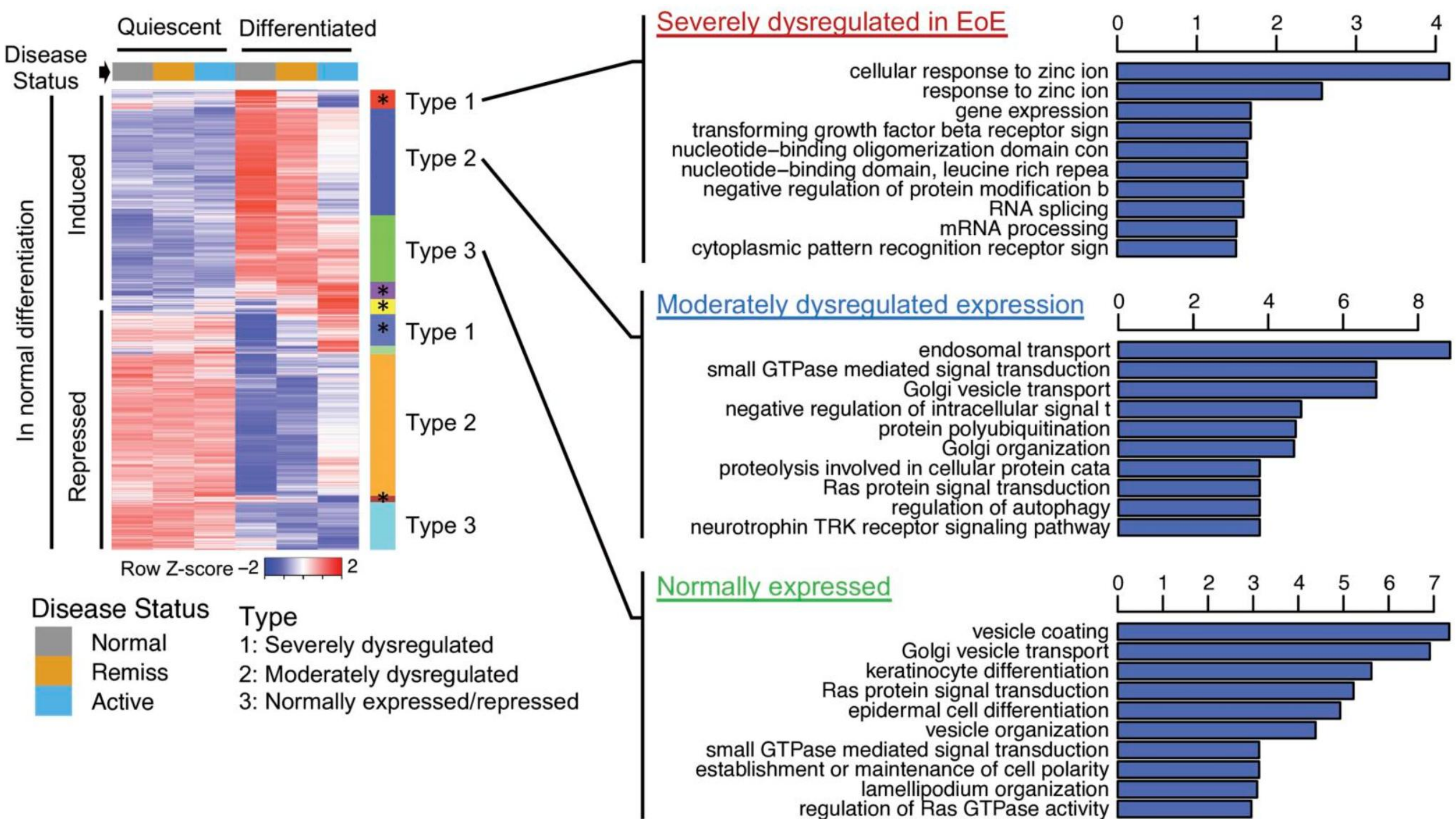
# RNA-seq Downstream Analysis





https://bioconnector.github.io/workshops/r-rnaseq-airway.html

# RNA-seq Downstream Analysis

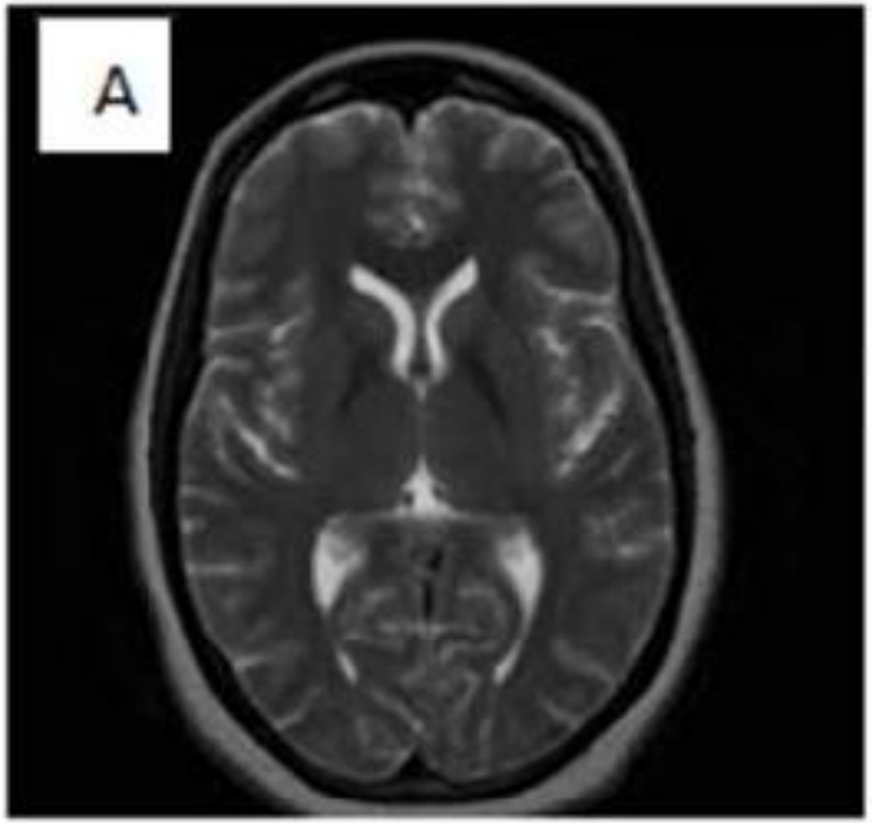


https://insight.jci.org/articles/view/159093/figure/5

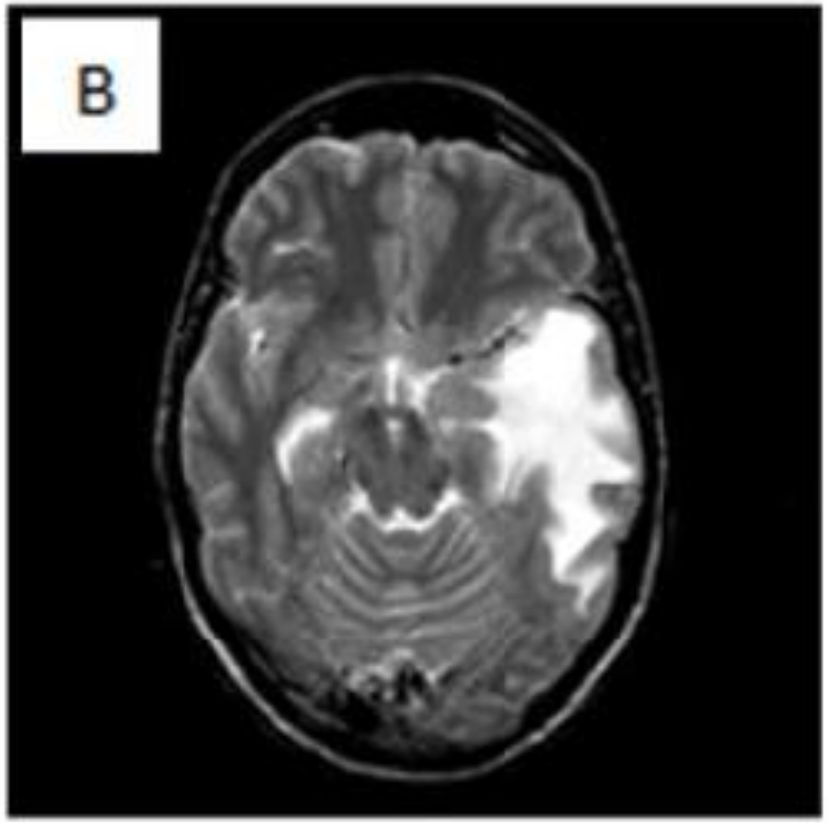
# Medical Data

- Text (Patient Medical Record or Health Record)
- Image (Hematoxylin Eosin (HE), Immunohistochemistry (IHC) CT nd MRI)
- Video (Medical Ultrasonography and Endoscopy)
- Signal (Electrocardiogram (ECG) and Electroencephalogram (EEG))

# MRI



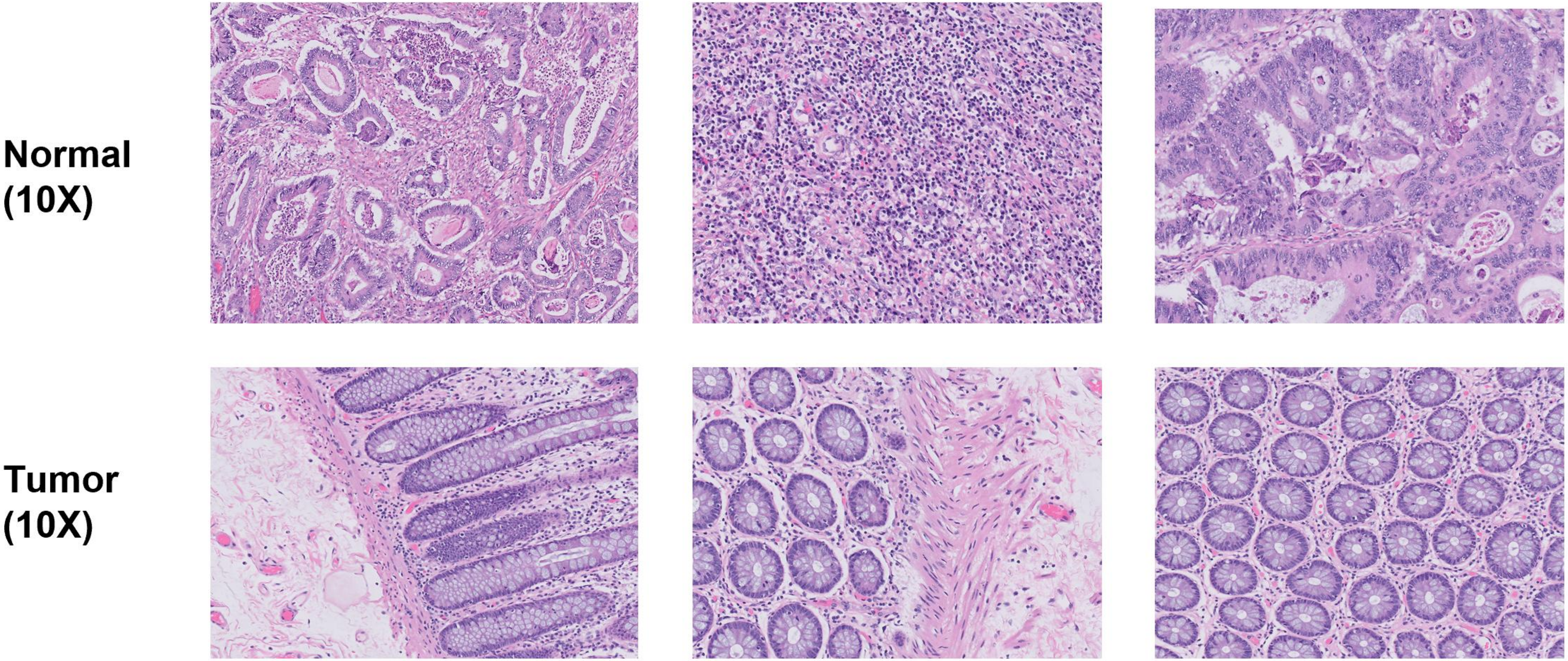
(A)-Normal brain



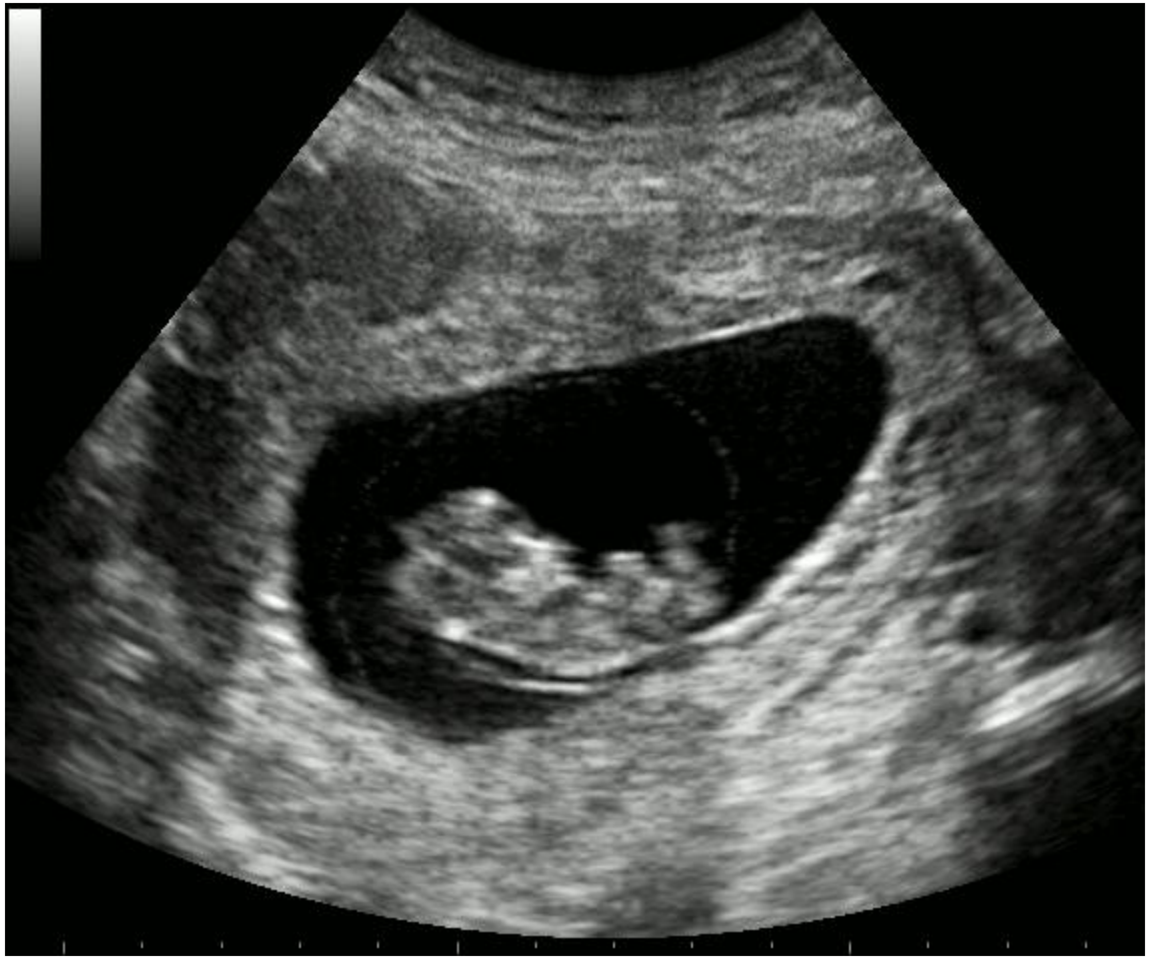
(B)-Tumor Brain

DOI:10.1109/ASONAM.2012.142

# Hematoxylin Eosin

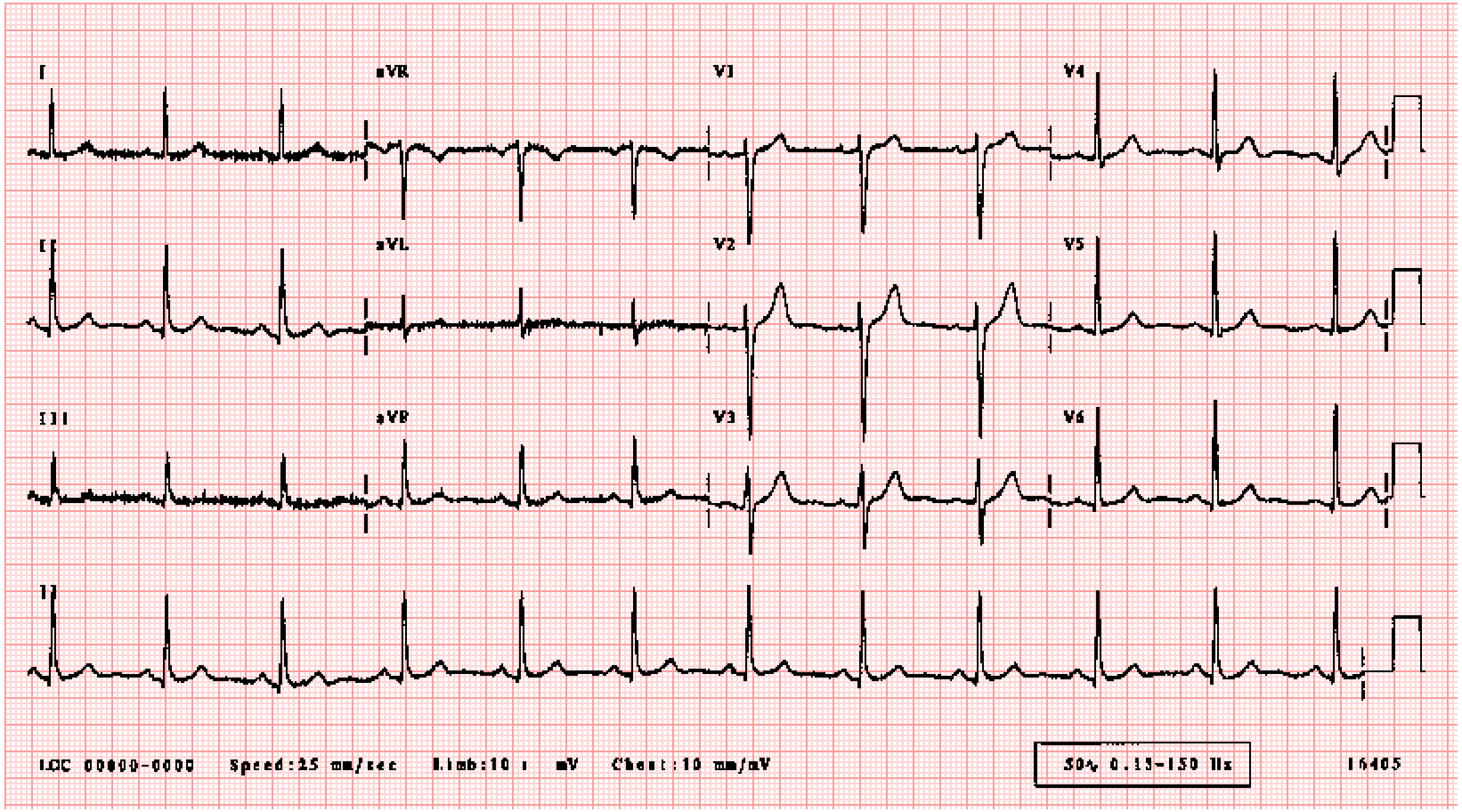


# Medical Ultrasonography



https://commons.wikimedia.org/wiki/File:Ultrasound_Scan_ND_0105143736_1439320.gif

# ECG



https://ecglibrary.com/norm.php

# Thank you!