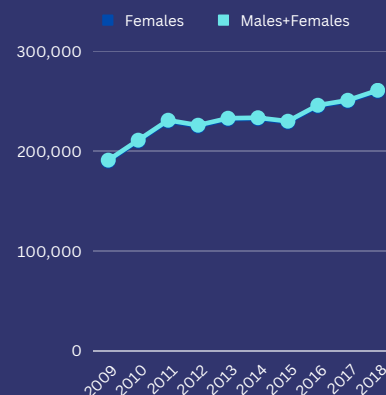Sarthak Ray
200106059
Bachelor Thesis Project

Differential Gene Expression Analysis
(Breast Invasive Carcinoma)

Background
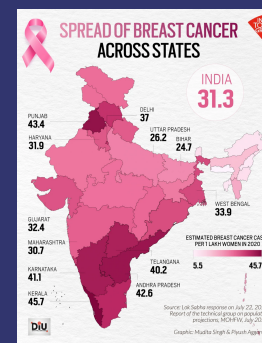
- Evolving silently, breast cancer's prevalence surges; 25% of all female cancers.

- Alarming trends indicate a steady rise in breast cancer cases over the years.

- DGE analysis unveils molecular signatures, paving the way for personalized interventions.

Females / Males+Females

Data courtesy: WHO

Data courtesy: WHO
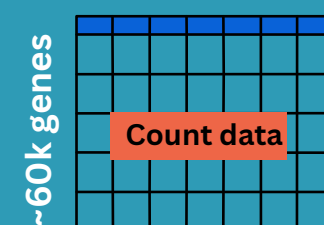
SPREAD OF BREAST CANCER ACROSS STATES
INDIA 31.3
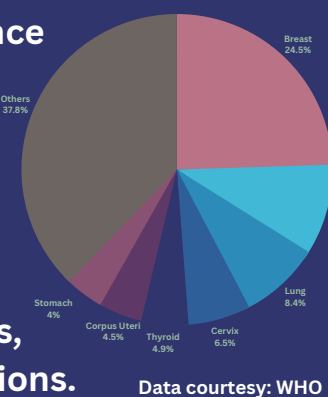
Graphic: India Today

GOAL: Empower precise prognostication in breast cancer to enhance treatment strategies.

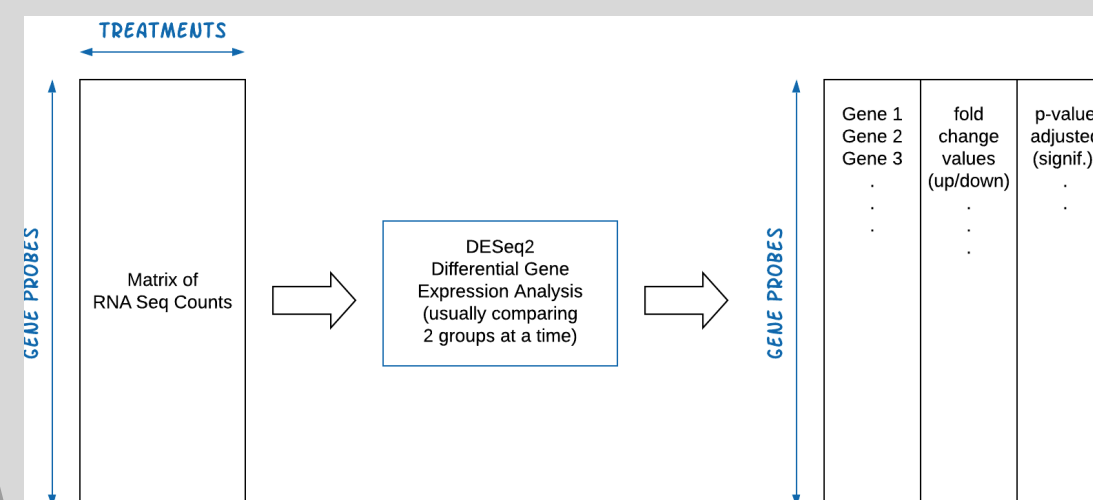### Input

~1200 samples
~60k genes
Count data
Clinical data

- We obtain the data from The Cancer Genome Atlas (TCGA) repository.

- The entire processing is done in RStudio using the package DESEQ2.

RStudio

Data & Methods

Post filtration of data:

	TCGA-D8-A1XO-01A	TCGA-AN-A0FN-01A	TCGA-BH-A18M-01A
ENSG00000000003.15	4370	2443	1635
ENSG00000000005.6	7	144	101
ENSG00000000419.13	2625	2322	1565
ENSG00000000457.14	3005	1466	1183

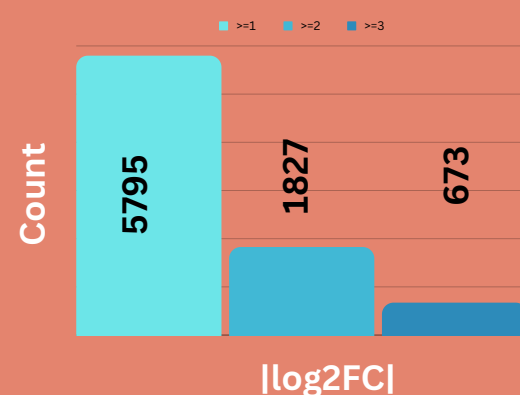|	Sample_Submitter_ID	ER_status
701	TCGA-D8-A1XO-01A	Positive
970	TCGA-AN-A0FN-01A	Positive
686	TCGA-BH-A18M-01A	Positive
304	TCGA-E2-A15T-01A	Positive
284	TCGA-AR-A0TU-01A	Negative
313	TCGA-EW-A1IX-01A	Positive

TREATMENTS

GENE PROBES — Matrix of RNA Seq Counts → DESeq2 Differential Gene Expression Analysis (usually comparing 2 groups at a time) → GENE PROBES

	fold change values (up/down)	p-value adjusted (signif.).
Gene 1		
Gene 2		
Gene 3		

1) Modelling raw counts for each gene

2) Shrinking log2 fold changes

3) Testing for differential expression

DE genes selection

>=1 / >=2 / >=3

Count

5795
1827
673

|log2FC|

|log2FC| >=3

Volcano Plot
-log10(Adjusted p-value)
Log2 Fold Change
Genes: Not Significant / Significant
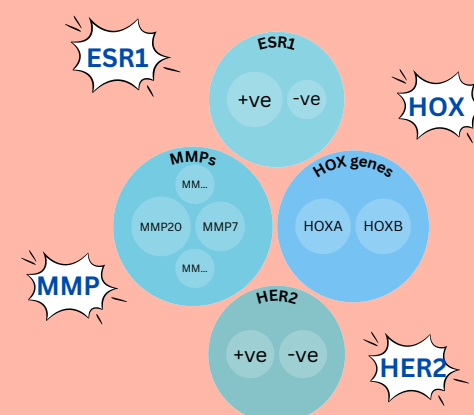
- We select differentially expressed genes that show at least 8-fold change.

- We have **673 genes** in our hand.

- We can now analyze the association of prognosis of breast cancer with these genes.

Genes of importance

ESR1
ESR1 +ve -ve
HOX
MMPs
MM... MMP20 MMP7 MM...
HOX genes
HOXA HOXB
MMP
HER2 +ve -ve
HER2

Further Work

- We analyze particular genes, on samples gathered from local hospitals.

- We figure out correlation between hormone expression vs. tumour prevalence.

- We can use endocrine therapy for suitable patients.