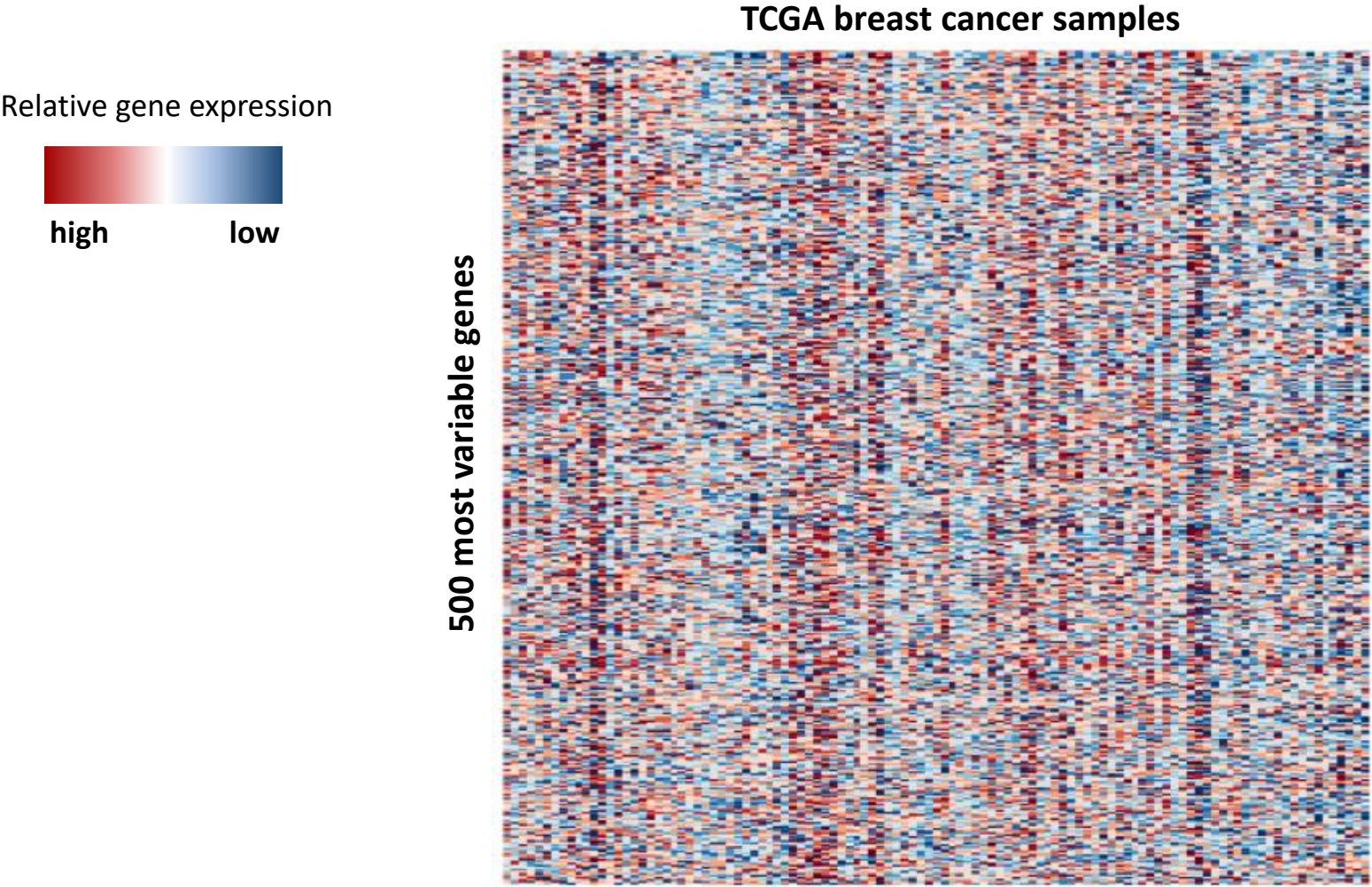
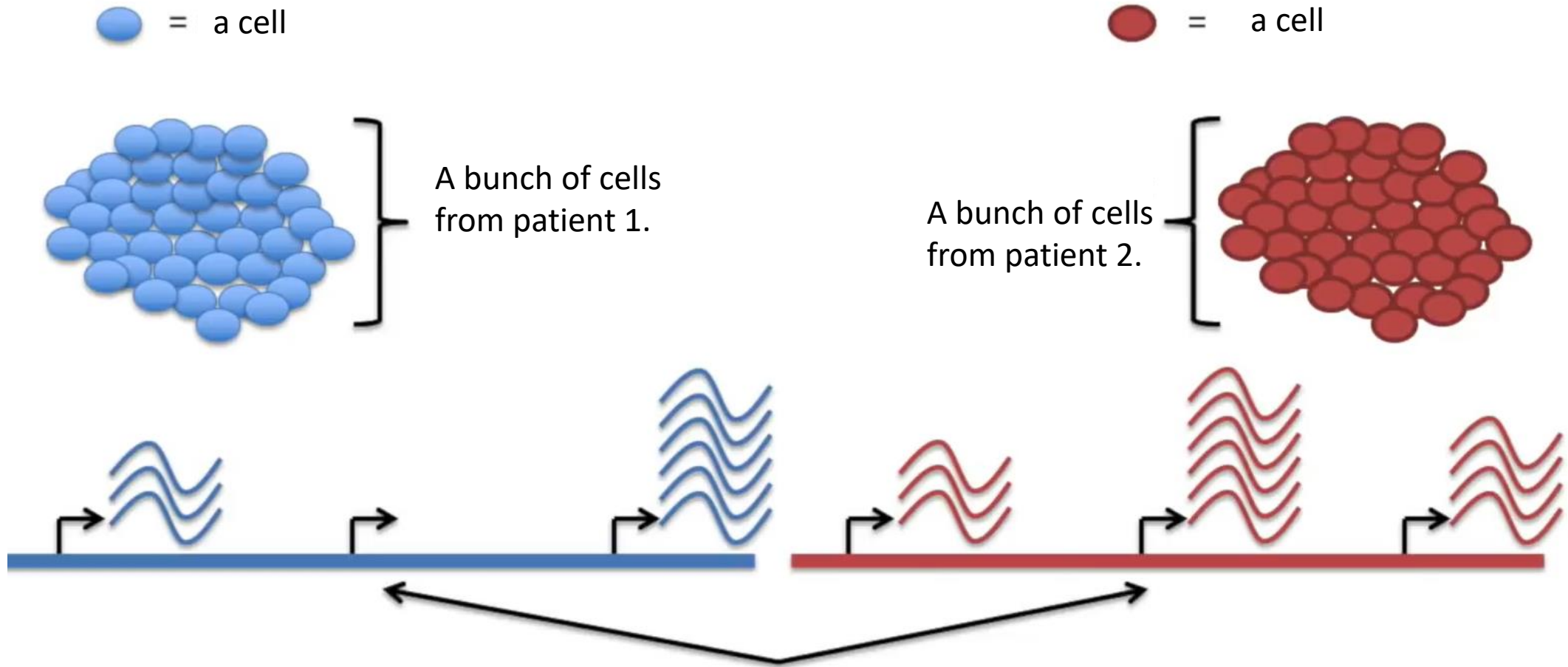


Heat map representation of gene expression levels across samples

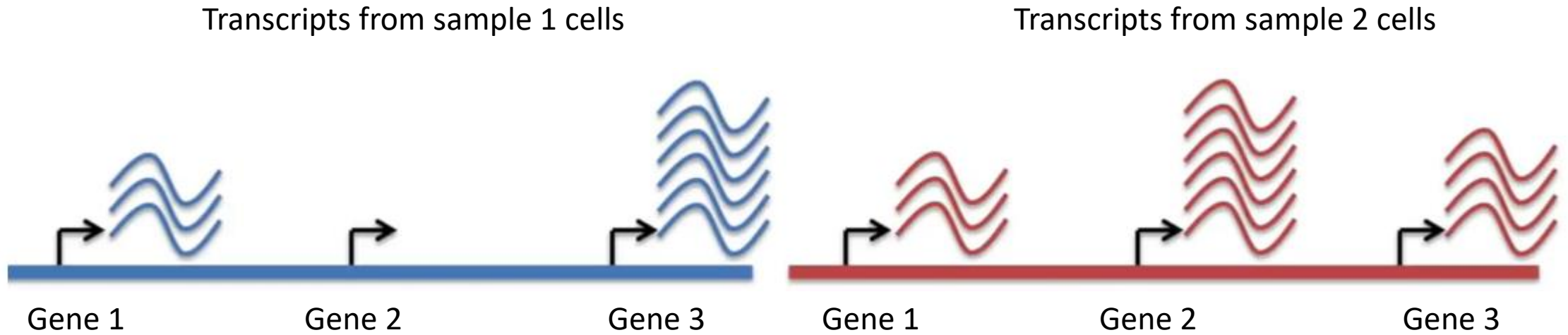


Gene expression data tells us how many copies of each gene in a sample (bunch of cells).

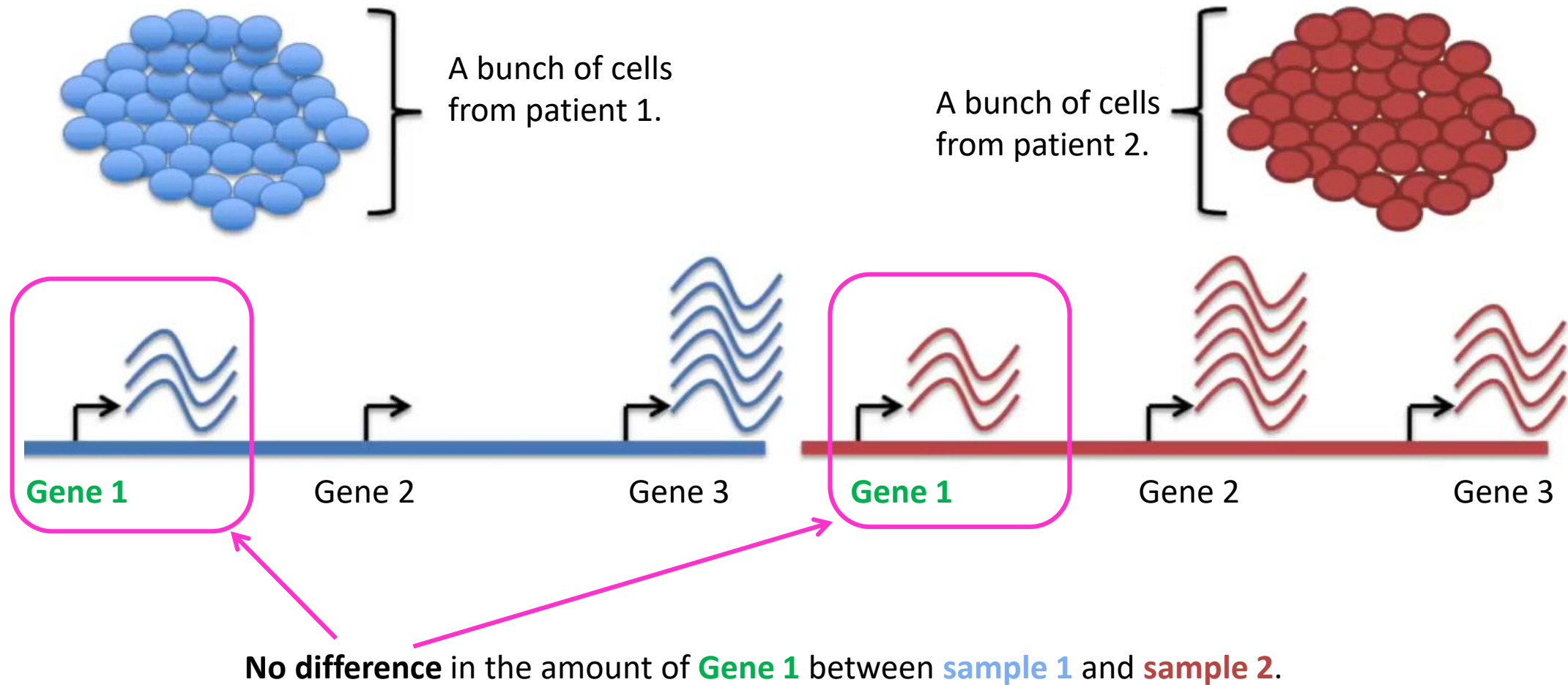


We compare different patient samples and figure out what's different in the two samples.

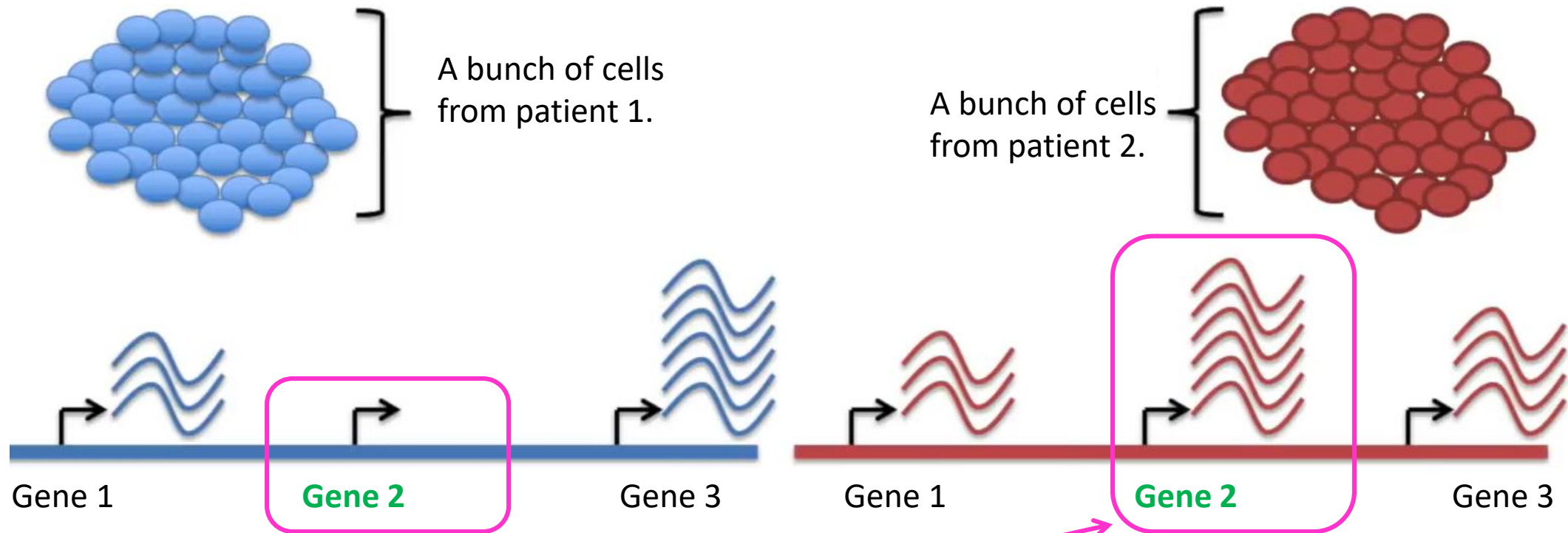
Gene expression data tells us how many copies of each gene in a sample (bunch of cells).



Gene expression data tells us how many copies of each gene in a sample (bunch of cells).

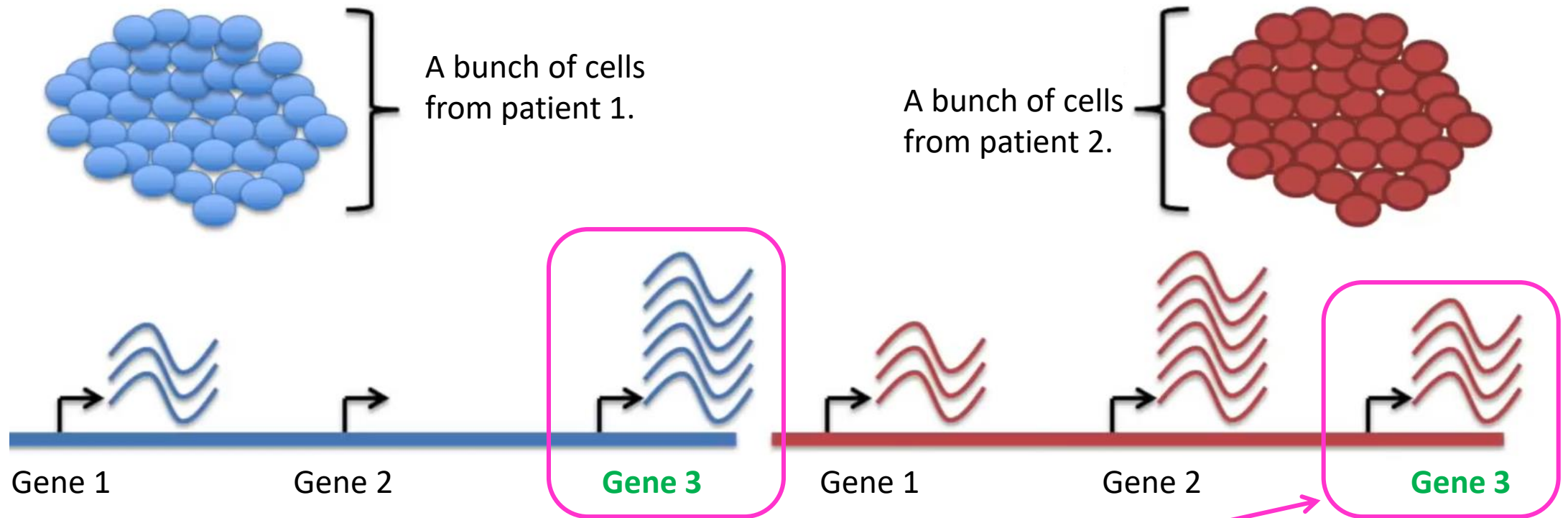


Gene expression data tells us how many copies of each gene in a sample (bunch of cells).



A **BIG difference** in the amount of Gene 2 between **sample 1** and **sample 2**.

Gene expression data tells us how many copies of each gene in a sample (bunch of cells).



A subtle difference in the amount of **Gene 3** between **sample 1** and **sample 2**.

Heat map representation of gene expression levels across samples

Expression relative to the
median across samples

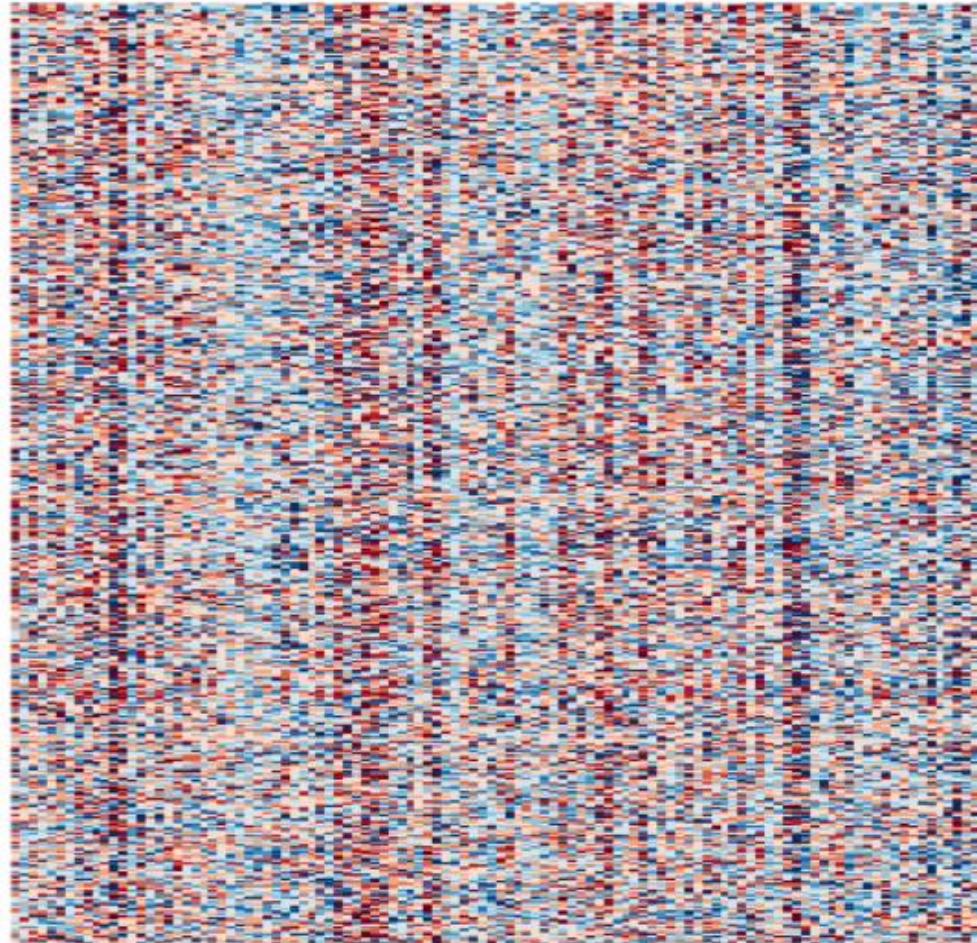


high

low

500 most variable genes

TCGA breast cancer samples



Clustering rearranges the
rows and columns according
to similarity.

Heat map representation of gene expression levels across samples

Expression relative to the
median across samples

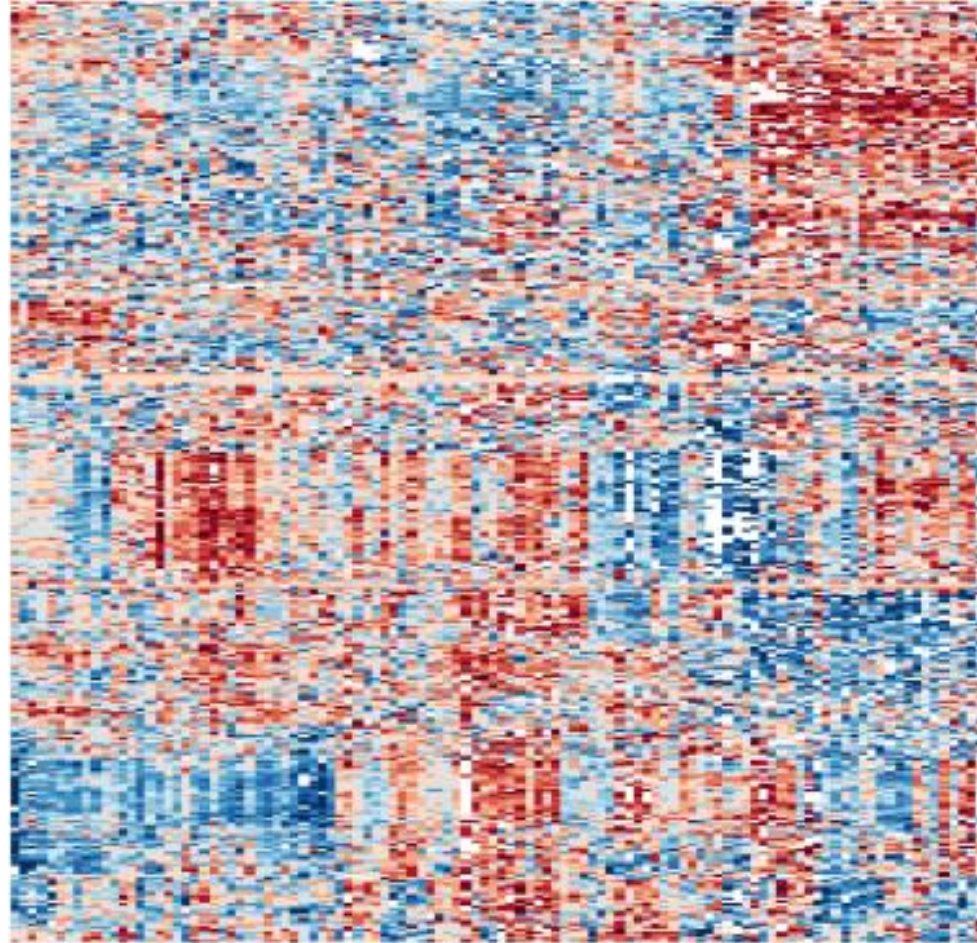


high

low

500 most variable genes

TCGA breast cancer samples



Clustering rearranges the rows and columns according to similarity.

It easy to see patterns in the clustered data.

Heat map representation of gene expression levels across samples

Expression relative to the median across samples

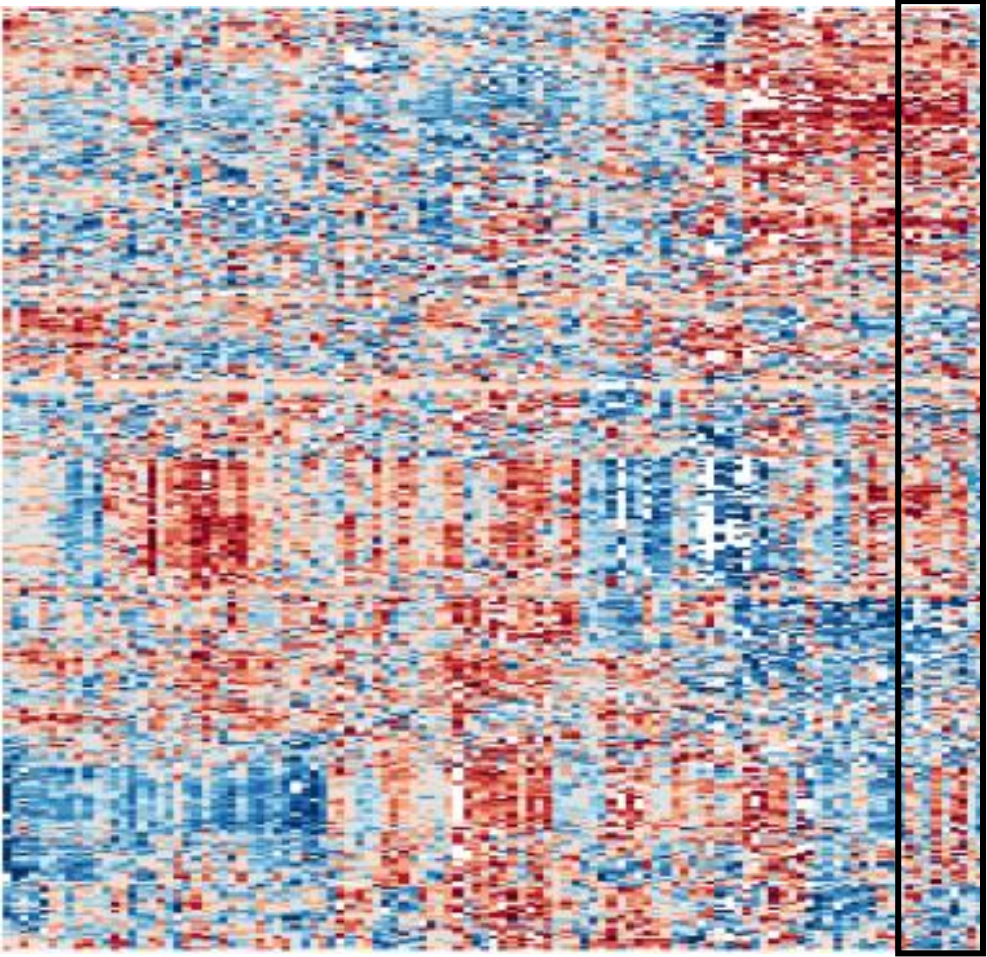


high

low

500 most variable genes

TCGA breast cancer samples



These samples express the similar genes.

Heat map representation of gene expression levels across samples

TCGA breast cancer samples

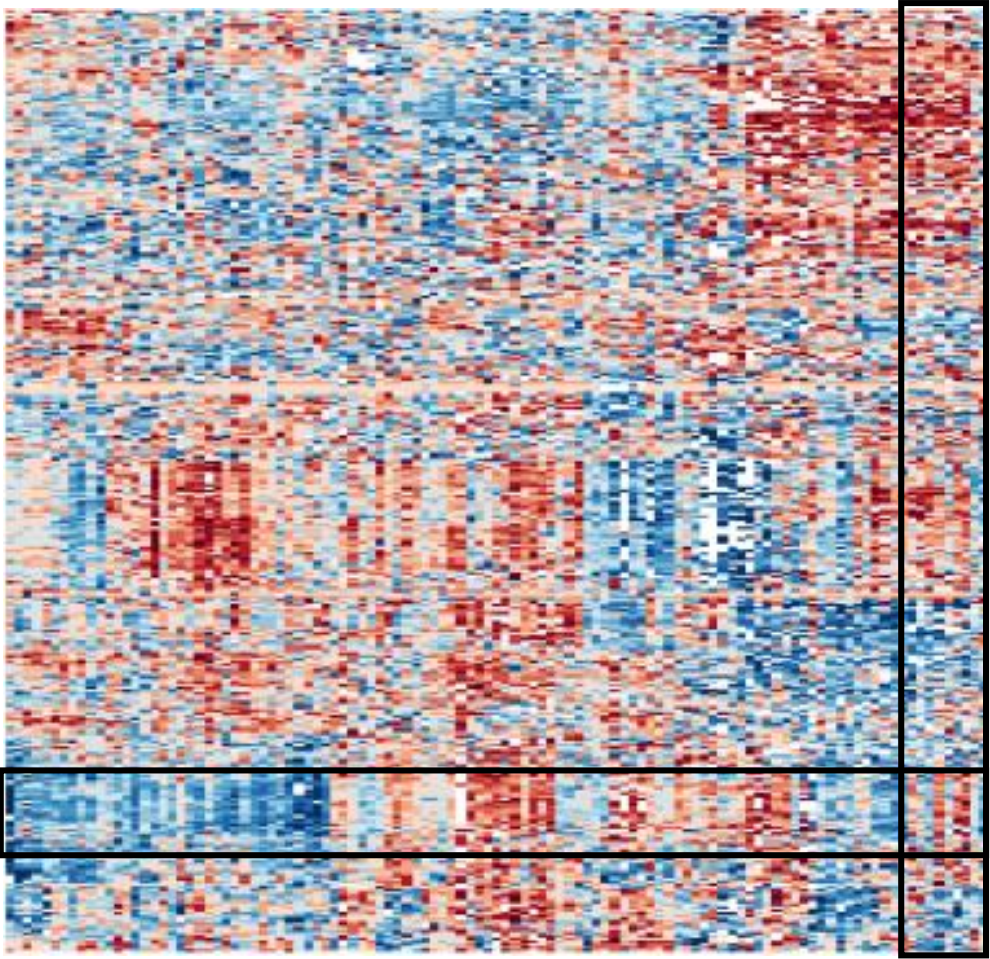
Expression relative to the median across samples



high

low

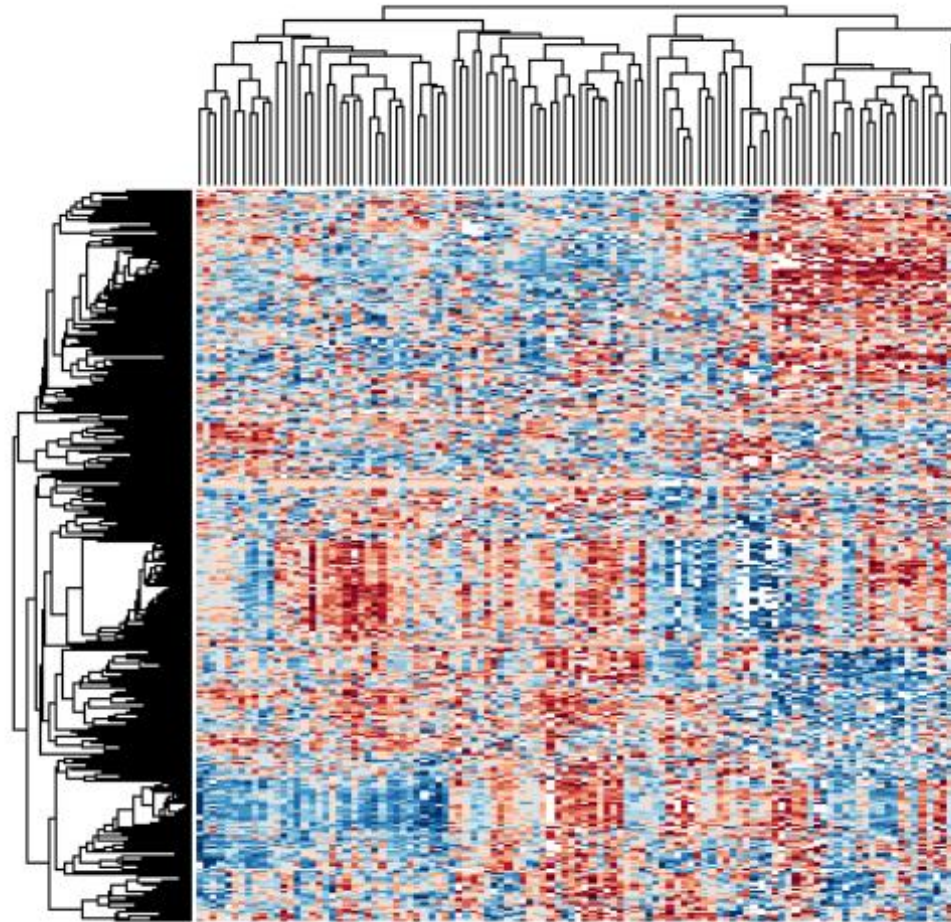
500 most variable genes



These samples express the same genes.

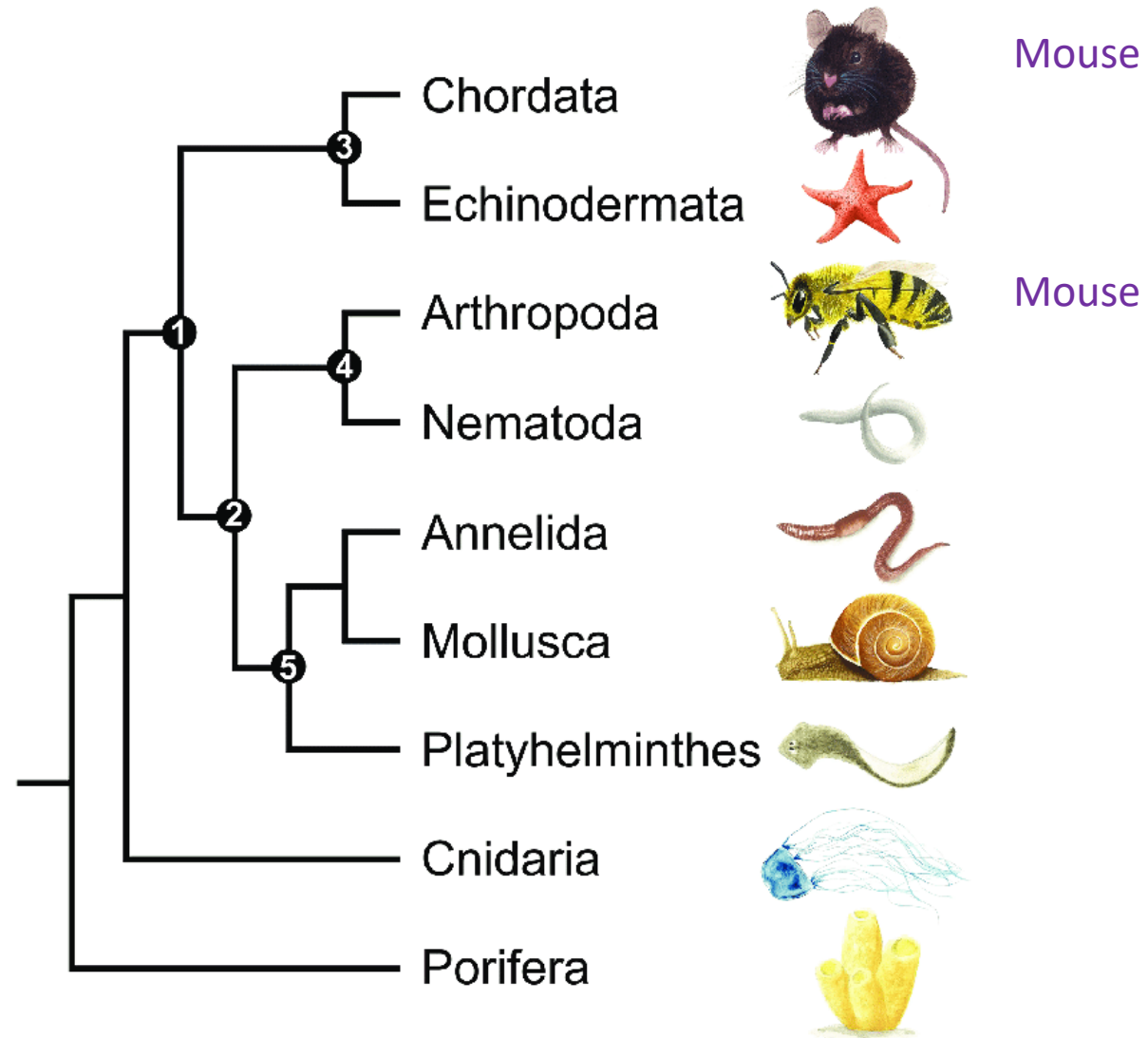
These genes behave similarly across the samples.

Heat maps often come with *dendrograms*.

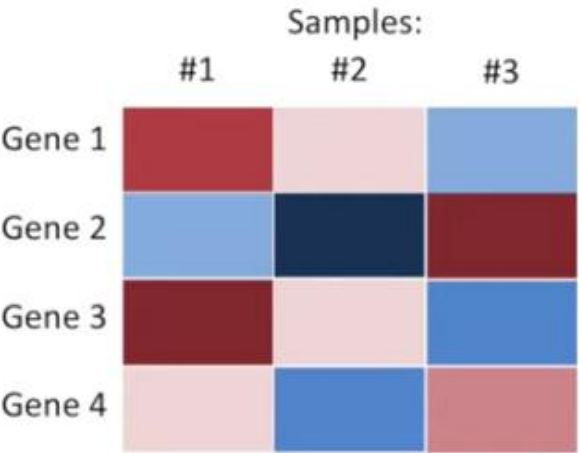


Dendrograms are **tree structures** that show how similar things (samples or genes) are to each other.

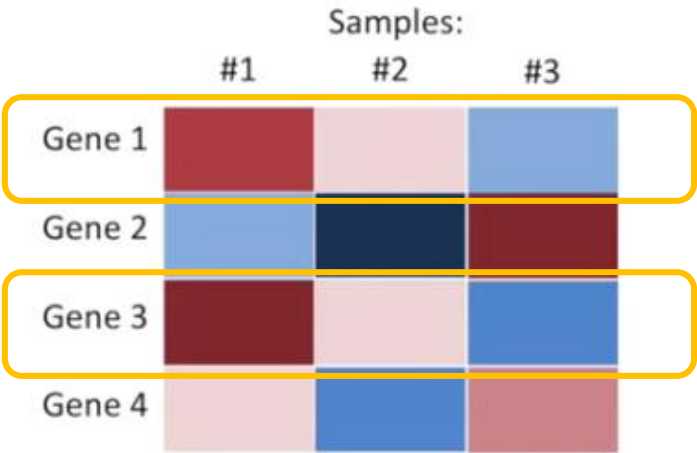
A “phylogenetic tree” is an example of a dendrogram.



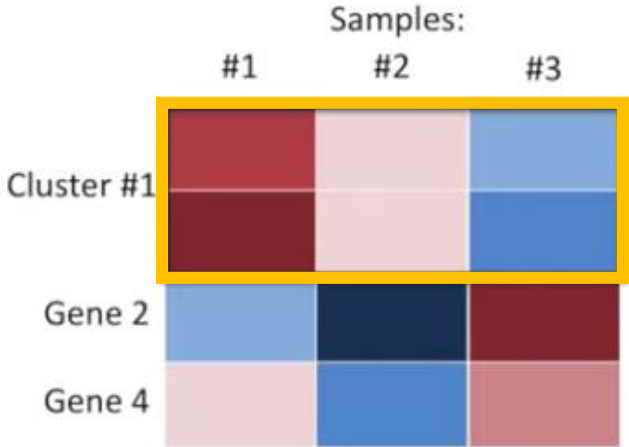
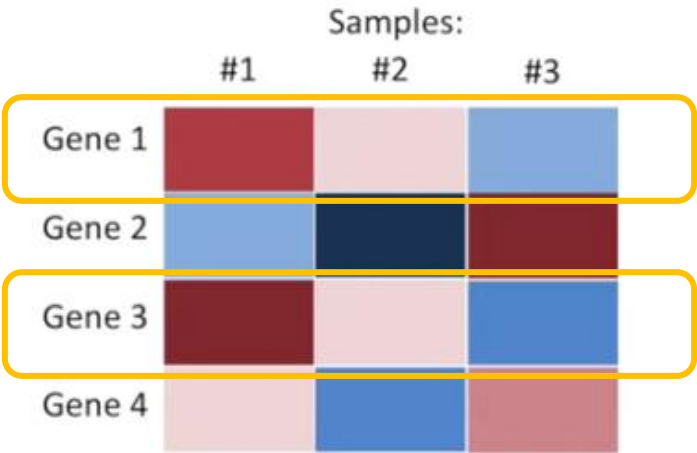
Heat map for expression values of four genes in three samples.



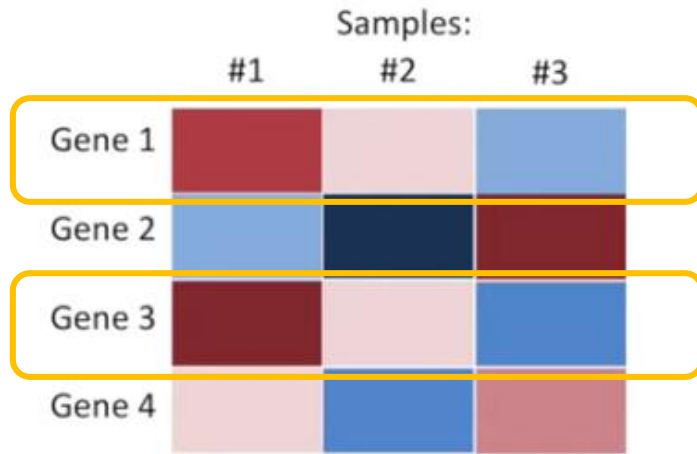
Gene 1 and Gene 3 are most similar to each other.



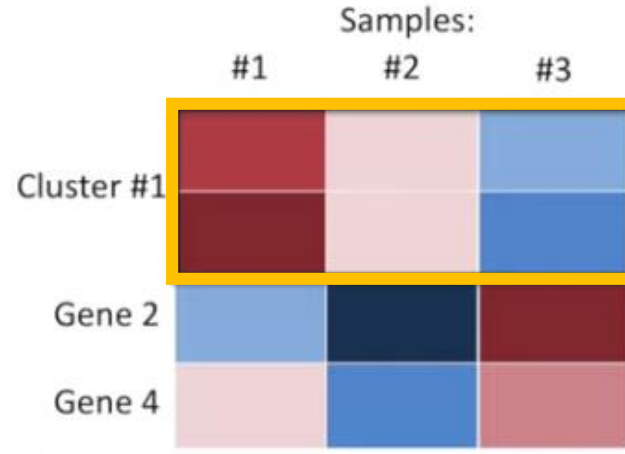
Gene 1 and Gene 3 become **Cluster #1**.



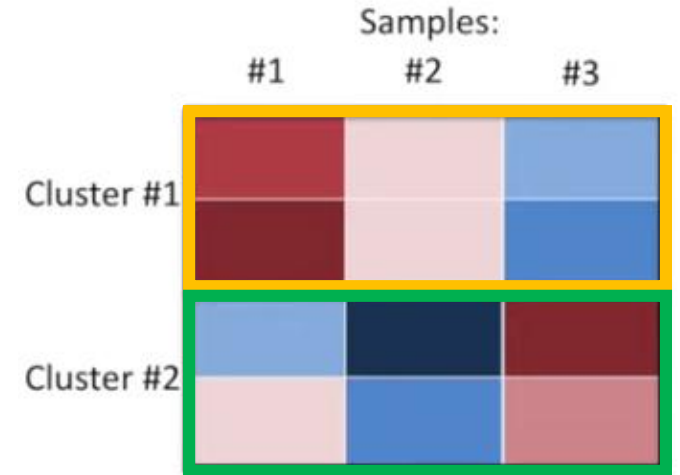
Genes 1 and 3 are most similar.



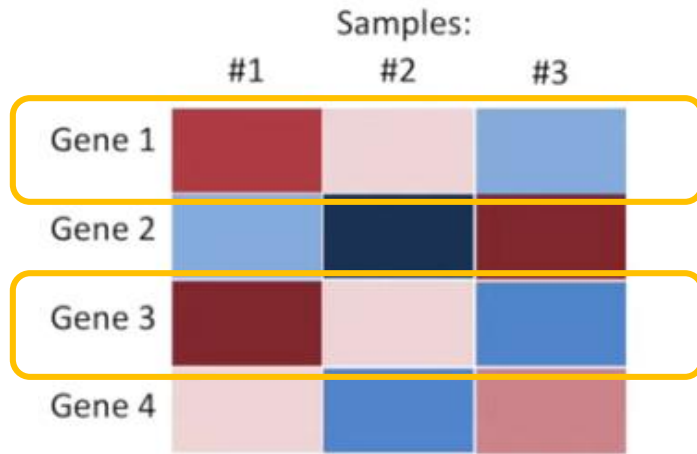
Genes 1 and 3 become **Cluster #1**.



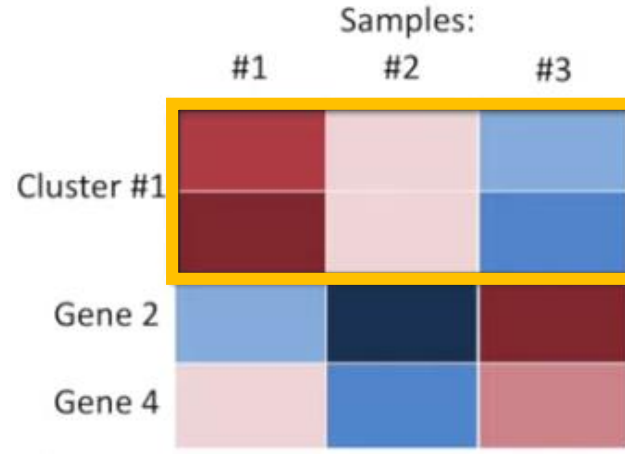
Genes 2 and 4 become **Cluster #2**.



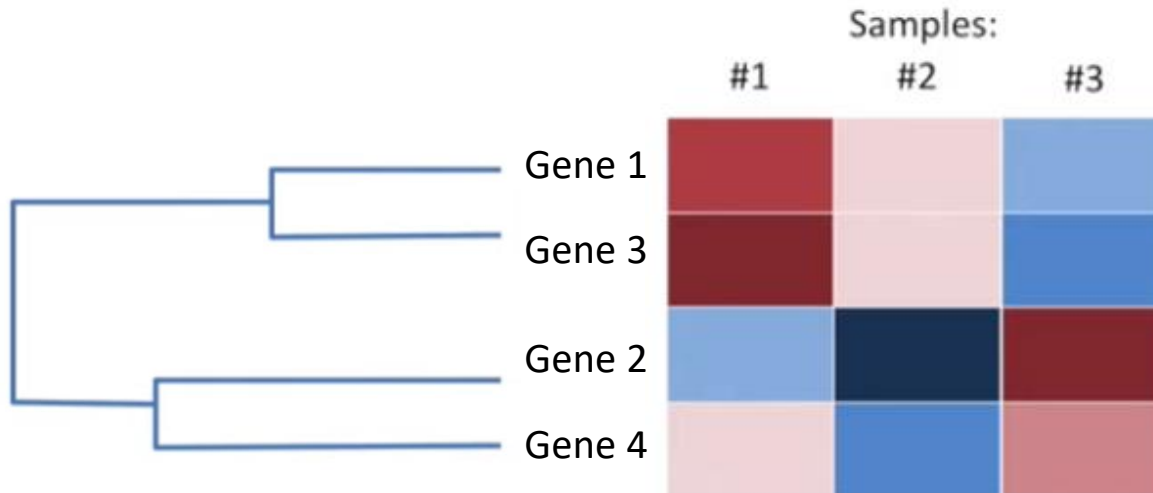
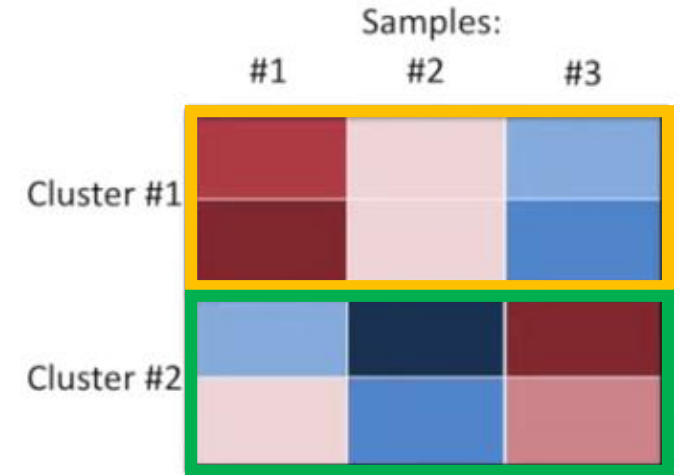
Genes 1 and 3 are most similar.



Genes 1 and 3 become **Cluster #1**.



Genes 2 and 4 become **Cluster #2**.



The dendrogram tree shows how the close the genes and clusters are to each other.

Heat map representation of gene expression levels across samples

TCGA breast cancer samples

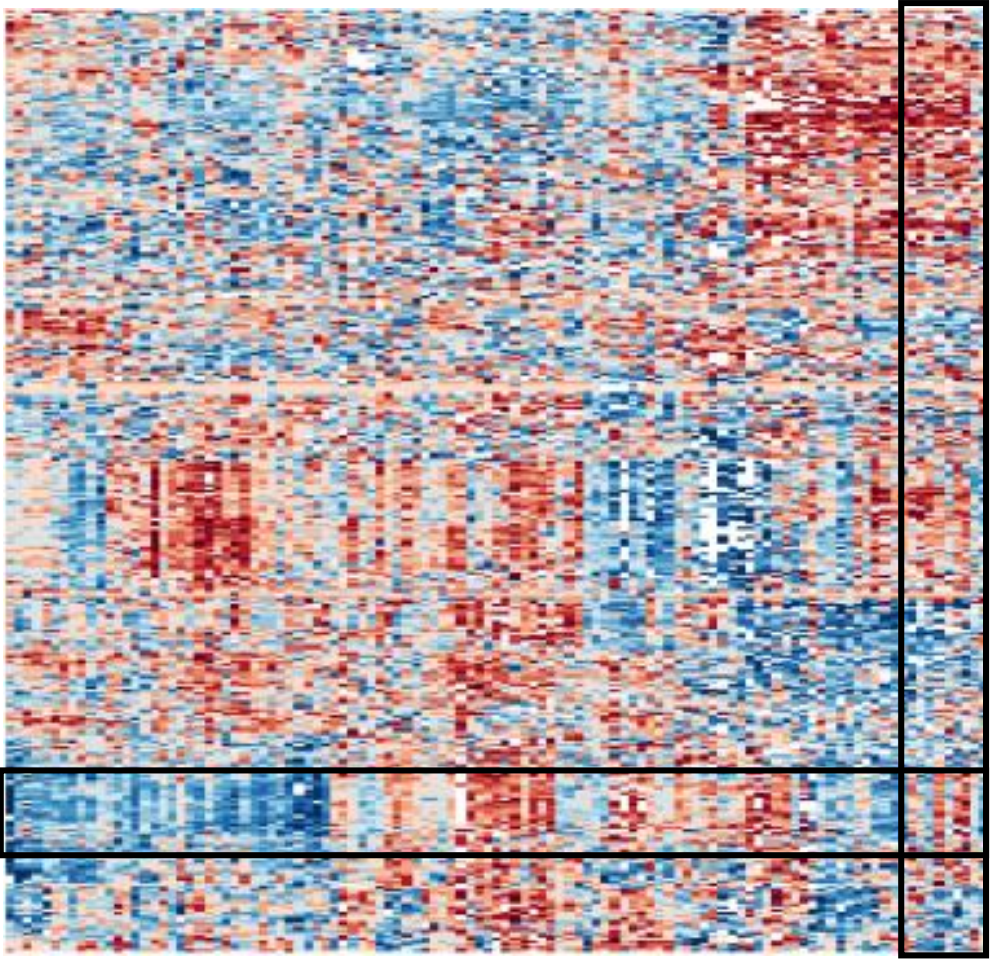
Expression relative to the median across samples



high

low

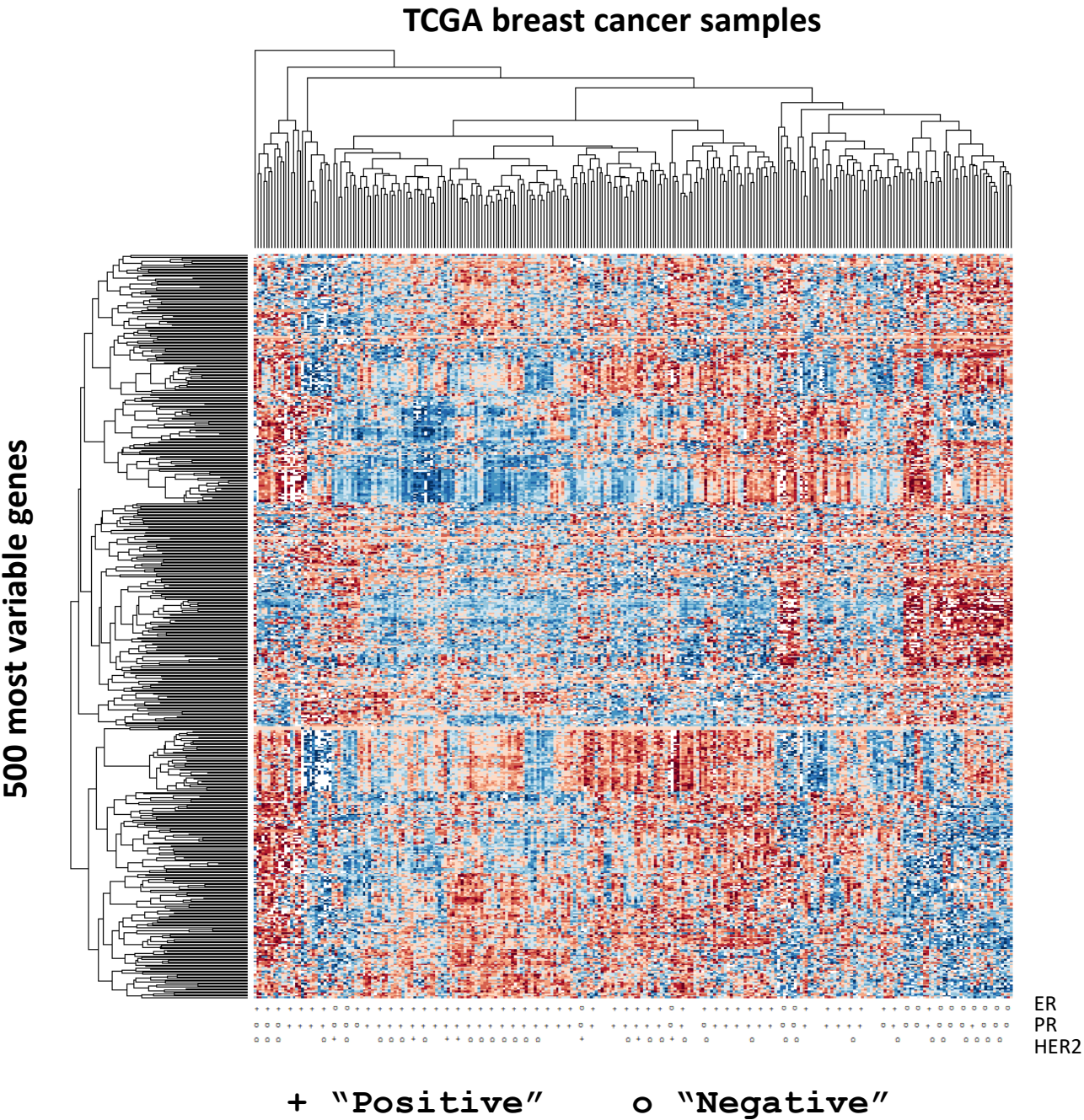
500 most variable genes



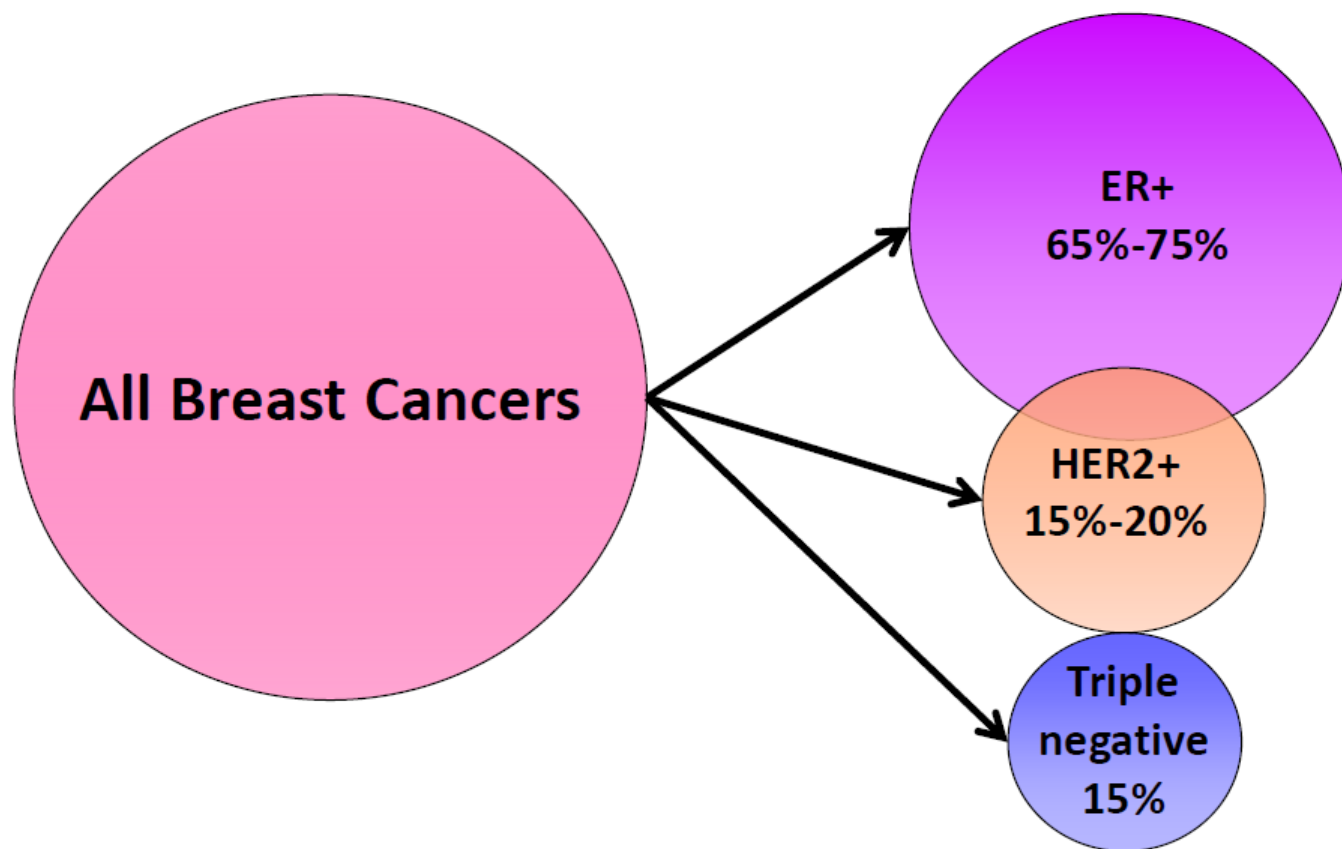
These samples express the same genes.

These genes behave similarly across the samples.

Clustered heat map
and receptor status



Invasive Breast Cancer Subsets Defined by IHC

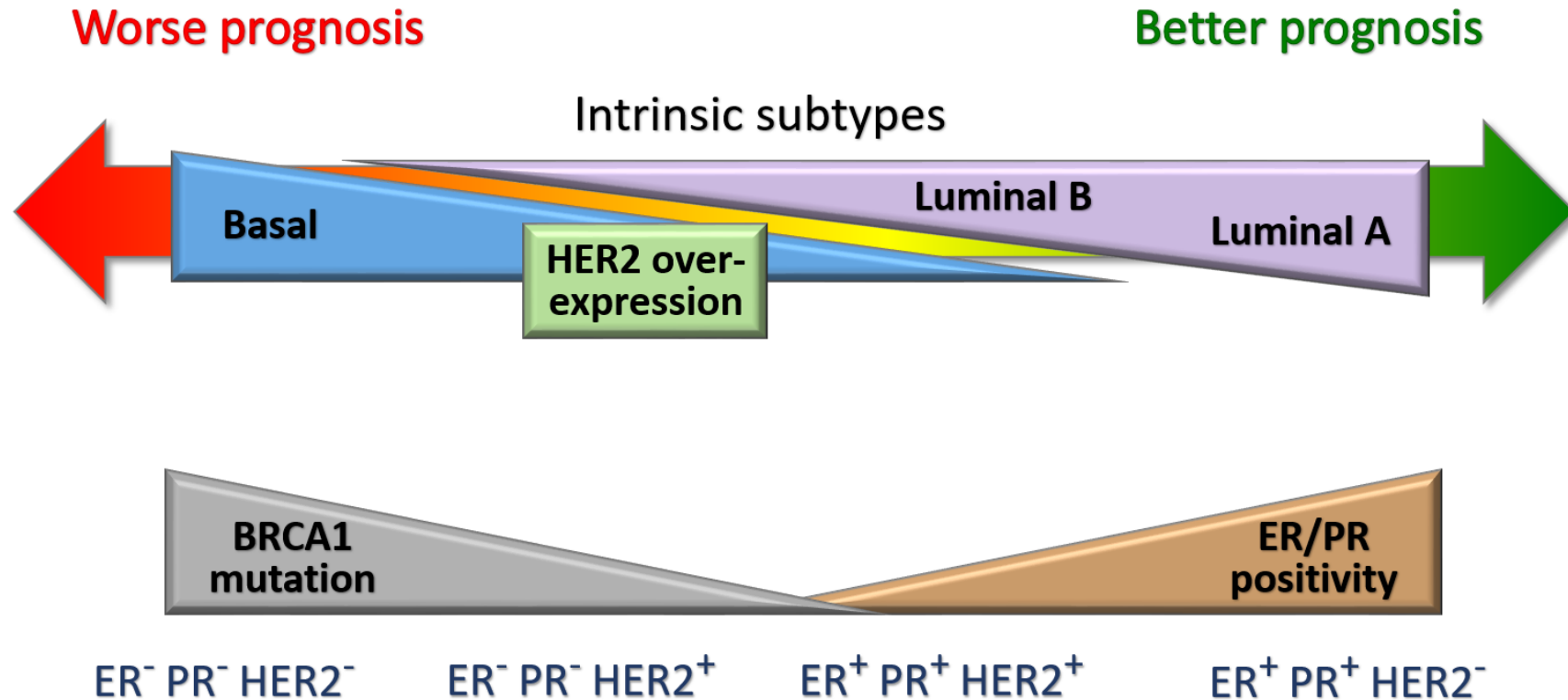


Patient outcomes based on breast cancer intrinsic (molecular) subtypes

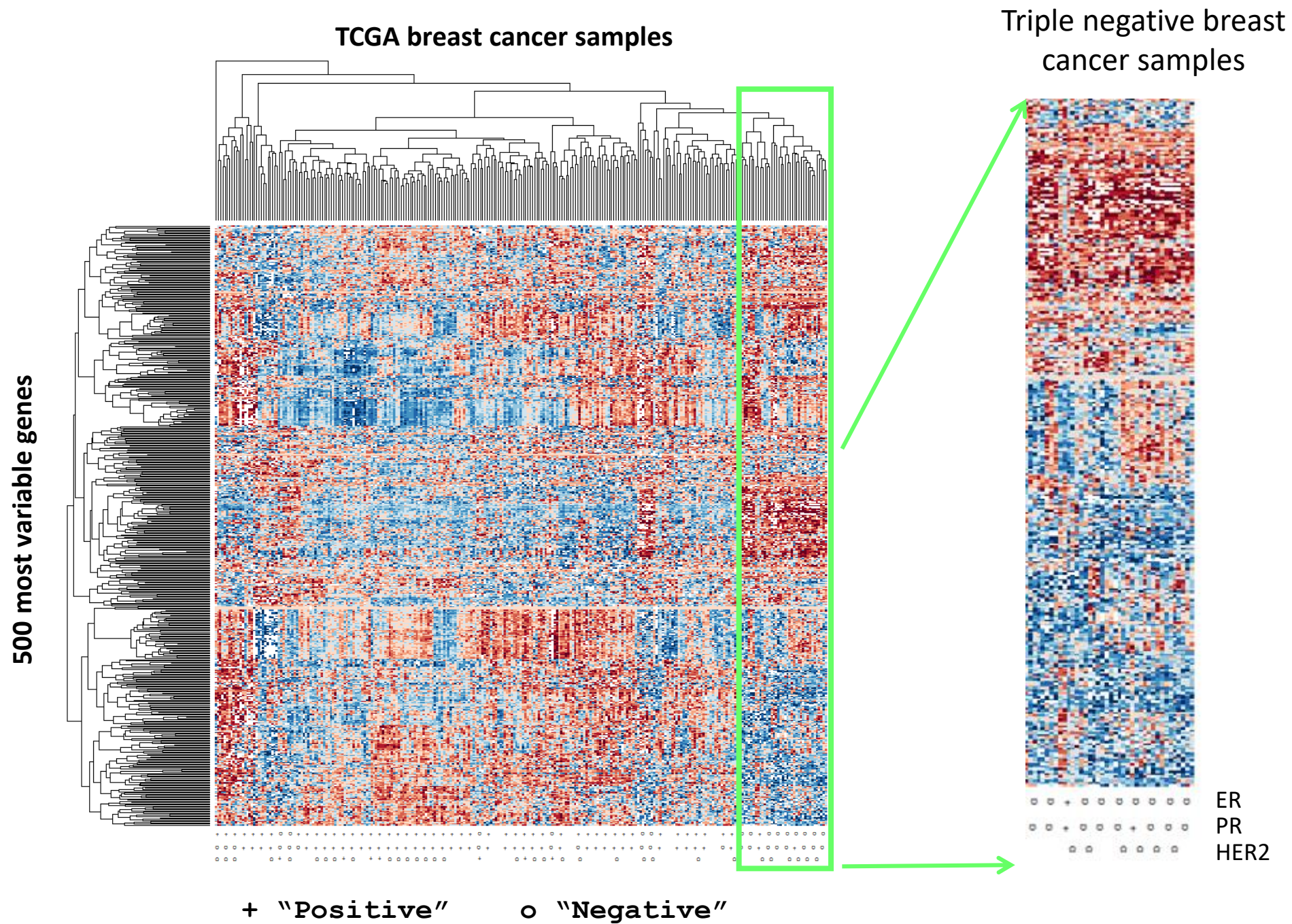
Expand information used to stratify patients for prognosis:

Classical clinical parameters age, node status, tumor size, histologic grade.

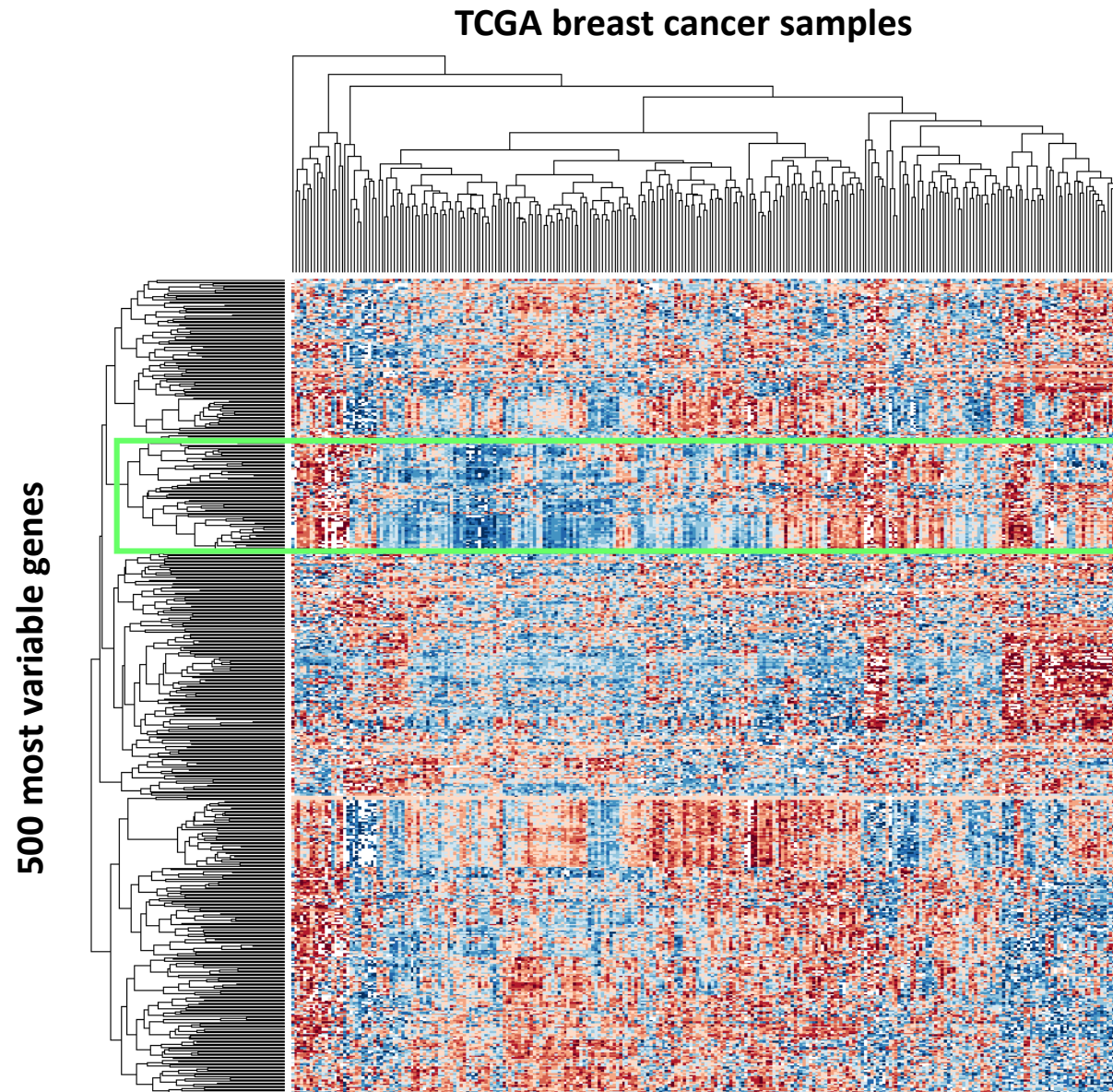
Pathologic markers ER, PR, and HER2.



Clustered heat map
and receptor status



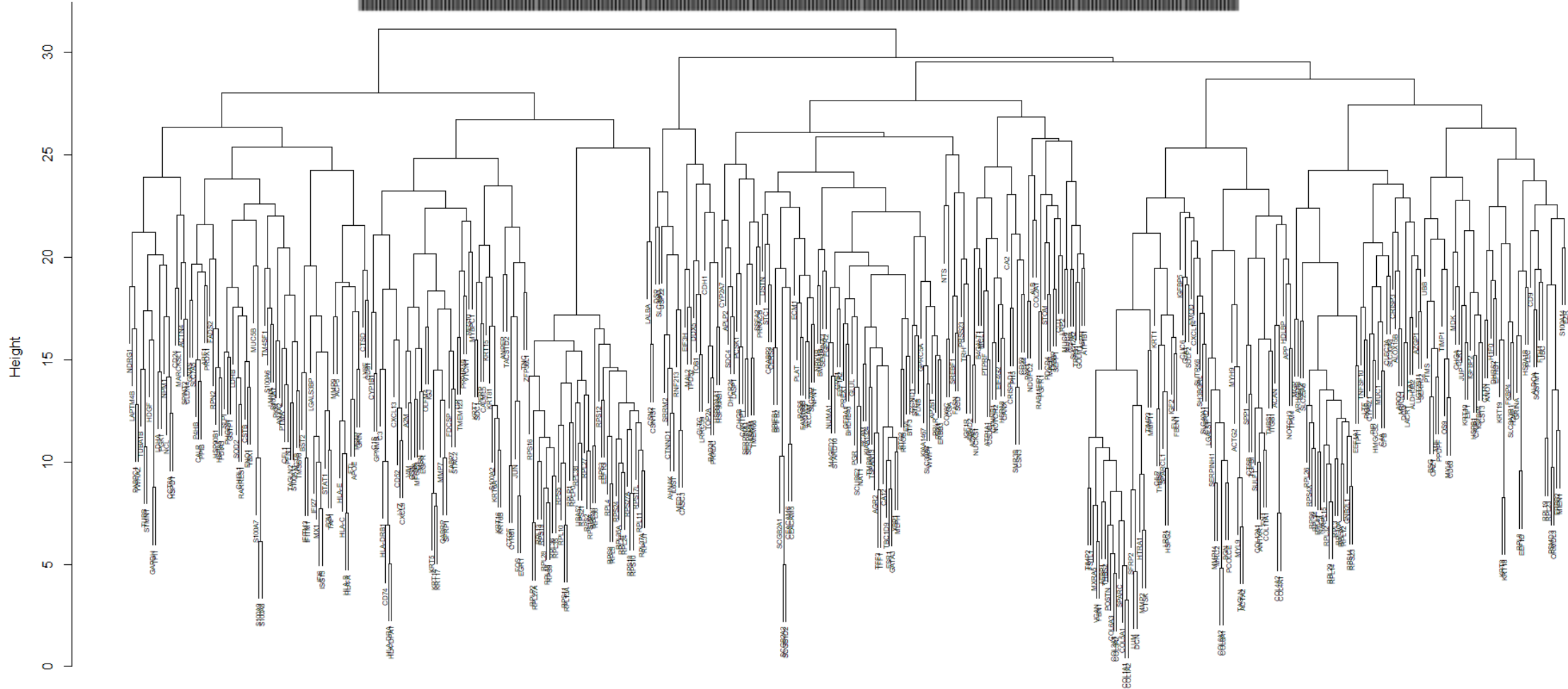
Clustered heat map
and gene clusters



These genes behave similarly across
the patient samples.

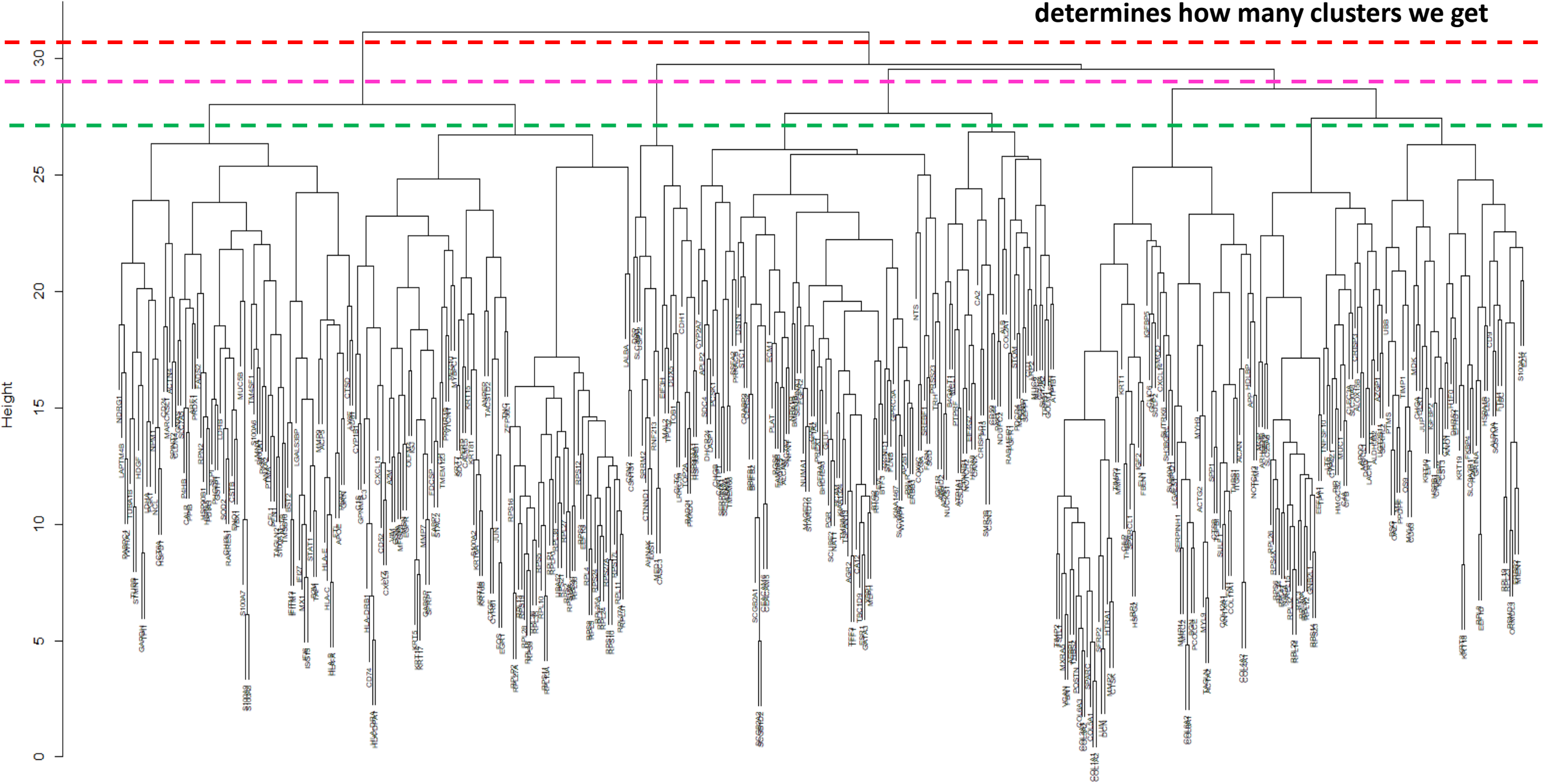
Do they perform similar functions in
breast cancer cells??

500 most variable genes



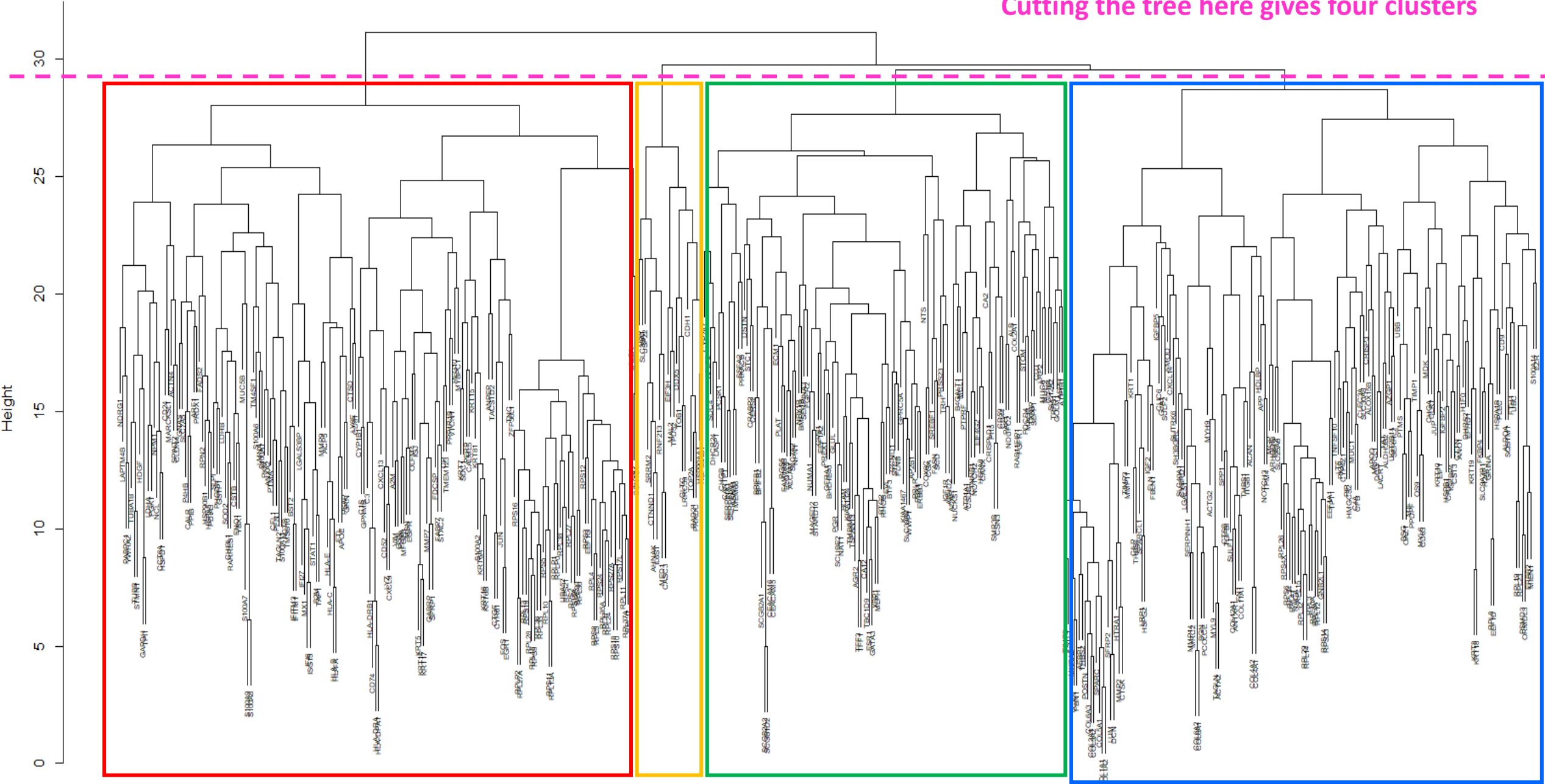
500 most variable genes

Where we “cut” the braches of the tree
determines how many clusters we get

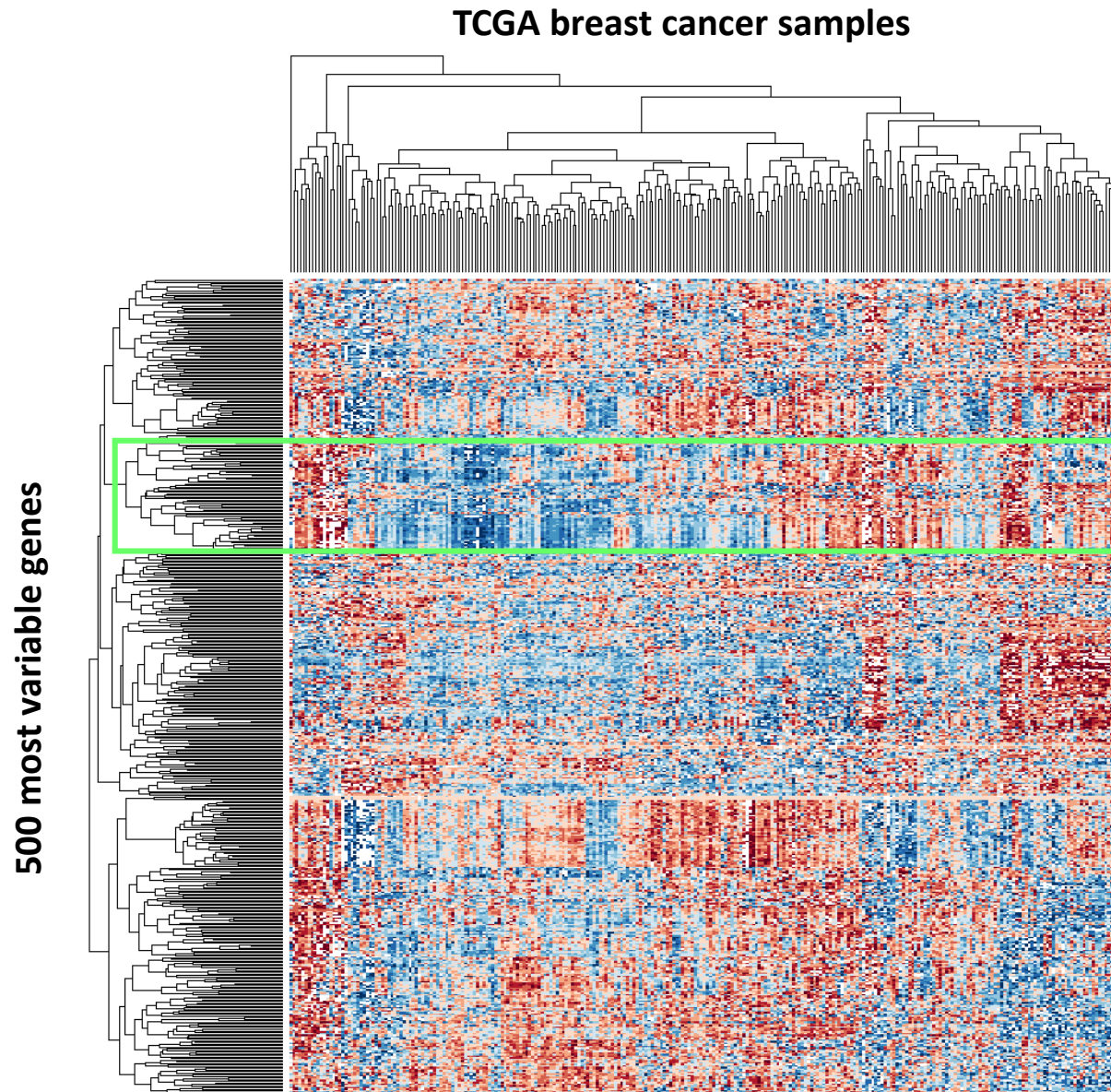


500 most variable genes

Cutting the tree here gives four clusters




Clustered heat map
and gene clusters



These genes behave similarly across
the patient samples.

***Do genes in the same cluster perform
similar biological functions??***

 Gene Ontology Consortium

HomeDocumentation ▾Downloads ▾Tools ▾About ▾Contact us

Enrichment analysis

Your gene IDs here...

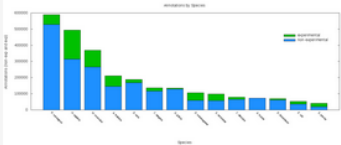
biological process ▾

Homo sapiens ▾

Submit

[Help](#)
Powered by [PANTHER](#)

Statistics



Other GOC tools

Gene Ontology Consortium

Search GO data

Search for terms and gene products...

Search

Ontology

[Filter classes](#)
[Download ontology](#)

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

molecular function
molecular activities of gene products

cellular component
where gene products are active

biological process
pathways and larger processes made up of the activities of multiple gene products.

[more](#)


Annotations

[Download annotations](#) (standard files)
[Filter and download](#) (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. [more](#)

The mission of the GO Consortium is to develop an up-to-date, comprehensive, **computational model of biological systems**, from the molecular level to larger pathways, cellular and organism-level systems. [more](#)

Search documentation

Search 

What is the Gene Ontology?

- [An introduction to the Gene Ontology](#)
- [What are annotations?](#)
- [Enrichment analysis](#)
- [Downloads](#)



on·tol·o·gy

/än'täləjē/

noun

1. the branch of metaphysics dealing with the nature of being.
2. a set of concepts and categories in a subject area or domain that shows their properties and the relations between them.

"what's new about our ontology is that it is created automatically from large datasets"



GENEONTOLOGY
Unifying Biology

THE GENE ONTOLOGY RESOURCE

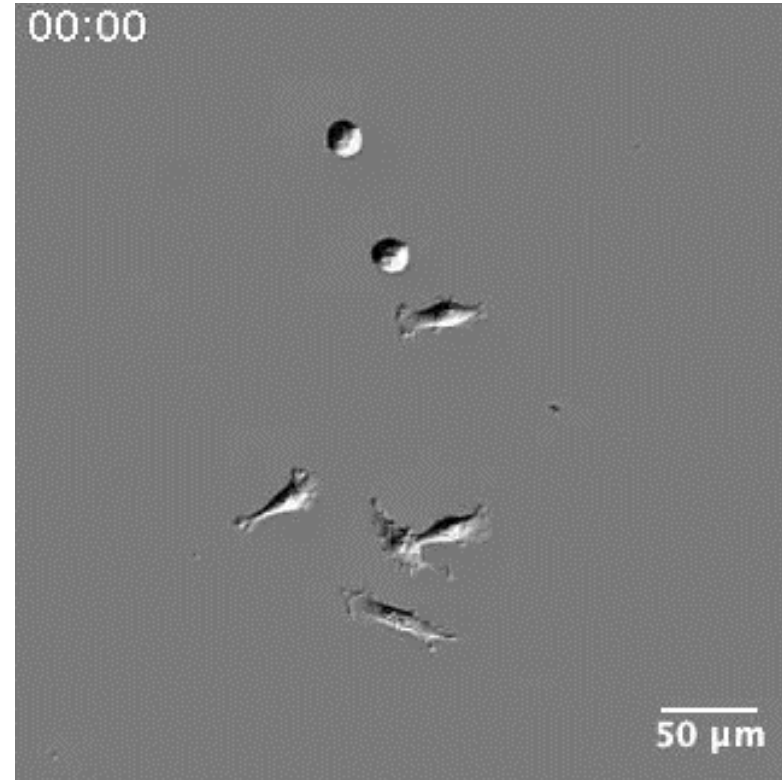
The mission of the GO Consortium is to develop a comprehensive, **computational model of biological systems**, ranging from the molecular to the organism level, across the multiplicity of species in the tree of life.

The Gene Ontology (GO) knowledgebase is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research.

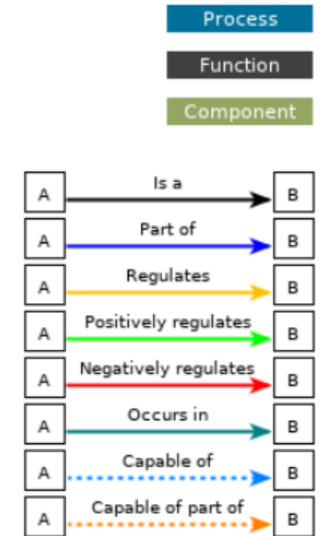
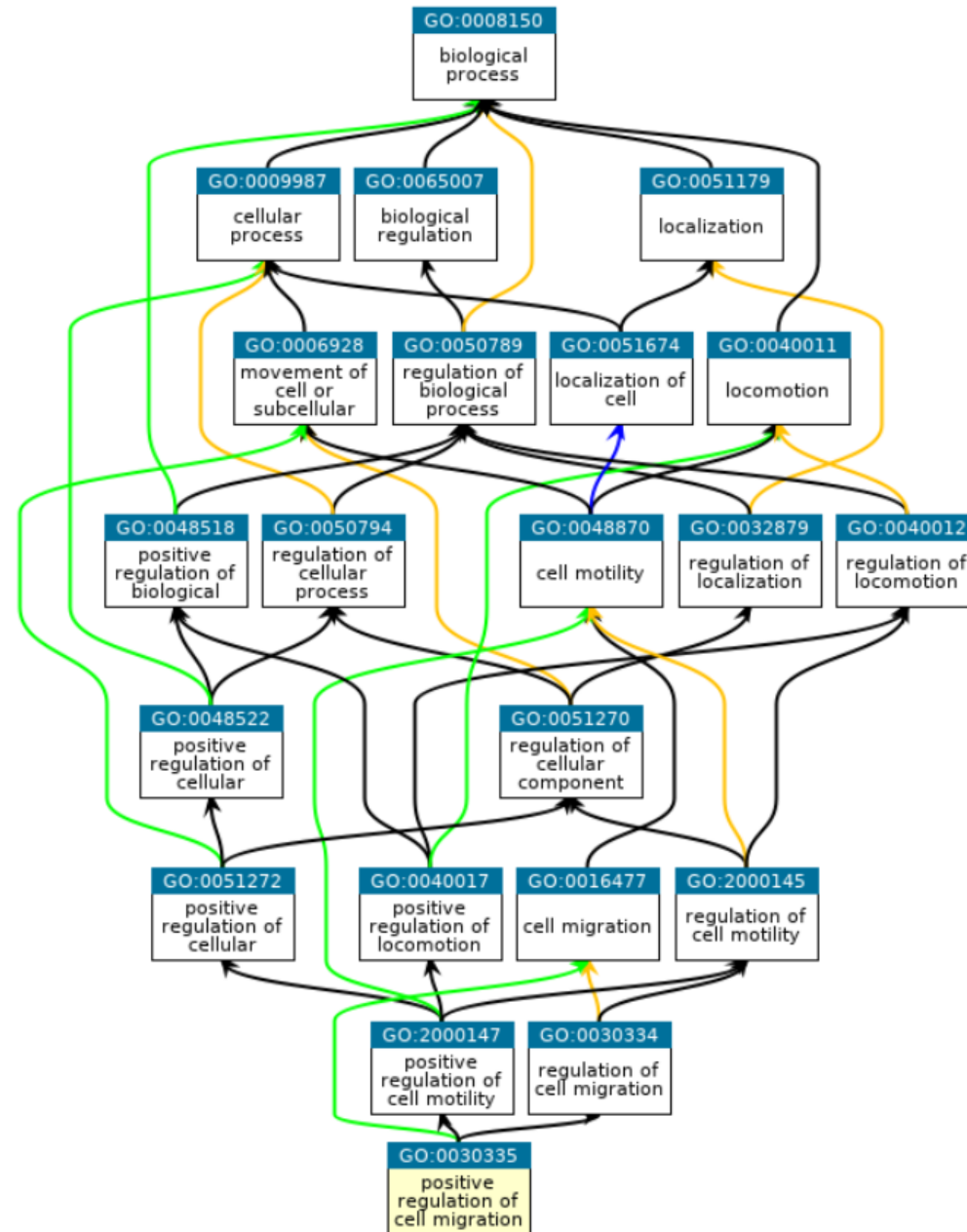


<https://www.youtube.com/watch?v=bdWRZd19swg>

MDA-MB-231 cell migration































Terms for biological processes such as *positive regulation of cell migration* are carefully organized into an ontology that describes biology.




Each gene has Gene Ontology (GO) annotations associated with it.

MIEN1: Migration and invasion enhancer 1

Symbol	GO Term	Evidence	Reference
MIEN1	GO:0010269    response to selenium ion	ECO:0000318  IBA	PMID:21873635
MIEN1	GO:0043066    negative regulation of apoptotic process	ECO:0000318  IBA	PMID:21873635
MIEN1	GO:0051491    positive regulation of filopodium assembly	ECO:0000318  IBA	PMID:21873635
MIEN1	GO:0006915    apoptotic process	ECO:0000322  IEA	GO_REF:0000043
MIEN1	GO:0051491    positive regulation of filopodium assembly	ECO:0000314  IDA	PMID:21628459
MIEN1	GO:0043066    negative regulation of apoptotic process	ECO:0000314  IDA	PMID:21068479
MIEN1	GO:0030335    positive regulation of cell migration	ECO:0000314  IDA	PMID:21628459

<http://www.geneontology.org>

 Gene Ontology Consortium

HomeDocumentation ▾Downloads ▾Tools ▾About ▾Contact us

Enrichment analysis

Your gene IDs here...

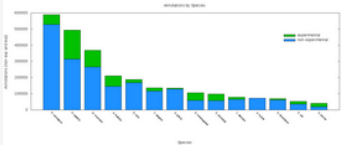
biological process ▾

Homo sapiens ▾

Submit

[Help](#)
Powered by [PANTHER](#)

Statistics



Other GOC tools

Gene Ontology Consortium

Search GO data

Search for terms and gene products...

Search

Ontology

[Filter classes](#)
[Download ontology](#)

Gene Ontology: the framework for the model of biology. The GO defines concepts/classes used to describe gene function, and relationships between these concepts. It classifies functions along three aspects:

molecular function
molecular activities of gene products

cellular component
where gene products are active

biological process
pathways and larger processes made up of the activities of multiple gene products.
[more](#)


Annotations

[Download annotations](#) (standard files)
[Filter and download](#) (customizable files <100k lines)

GO annotations: the model of biology. Annotations are statements describing the functions of specific genes, using concepts in the Gene Ontology. The simplest and most common annotation links one gene to one function, e.g. FZD4 + Wnt signaling pathway. Each statement is based on a specified piece of evidence. [more](#)

The mission of the GO Consortium is to develop an up-to-date, comprehensive, **computational model of biological systems**, from the molecular level to larger pathways, cellular and organism-level systems. [more](#)

Search documentation

Search 

What is the Gene Ontology?

- [An introduction to the Gene Ontology](#)
- [What are annotations?](#)
- [Enrichment analysis](#)
- [Downloads](#)