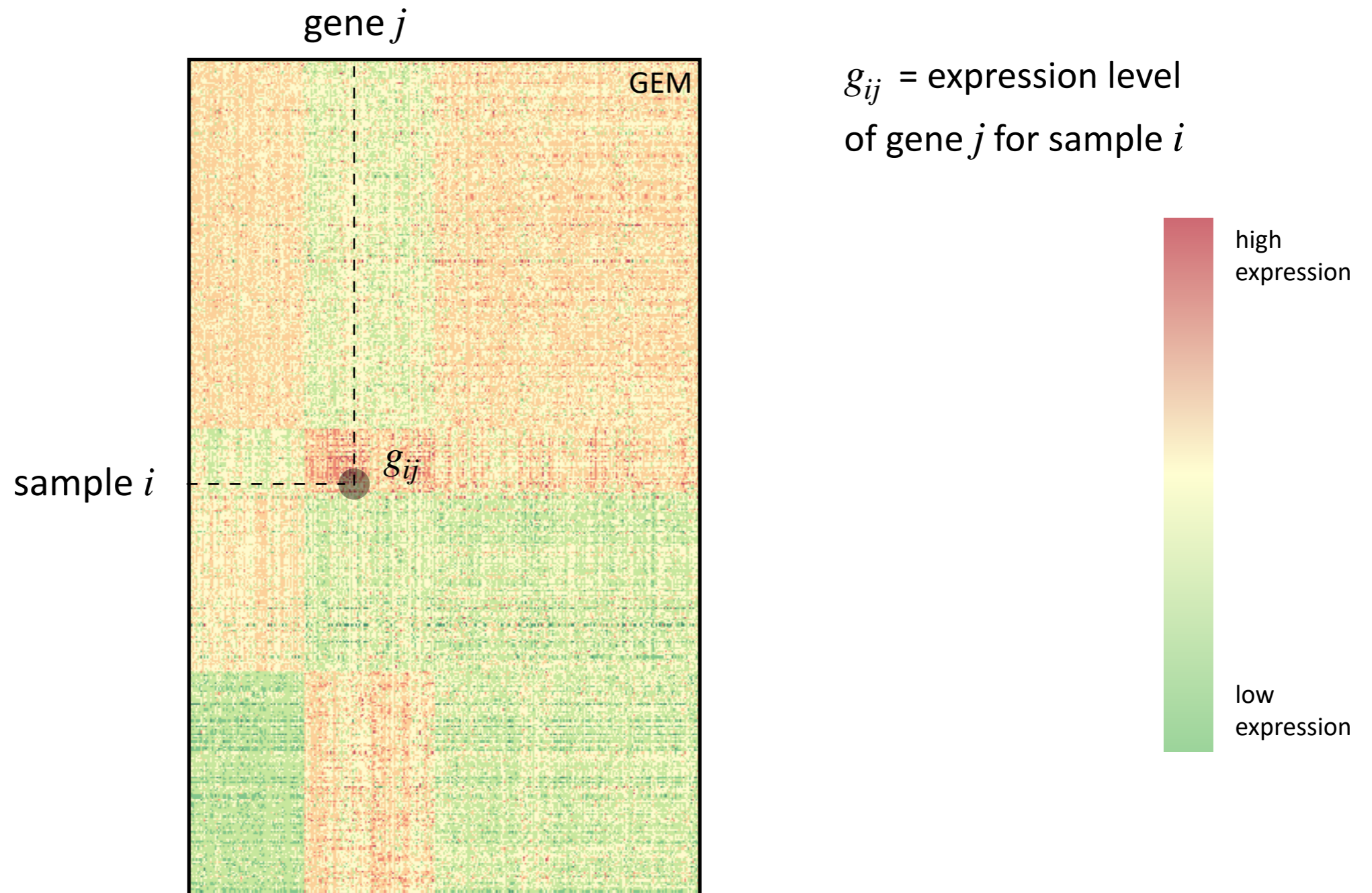


Interactive dual projections for gene expression analysis

Ignacio Díaz¹, José M. Enguita¹, Diego García¹, Ana González¹, Abel A. Cuadrado¹
M^o Dolores Chiara², Nuria Valdés³

Gene Expression Matrix (GEM)



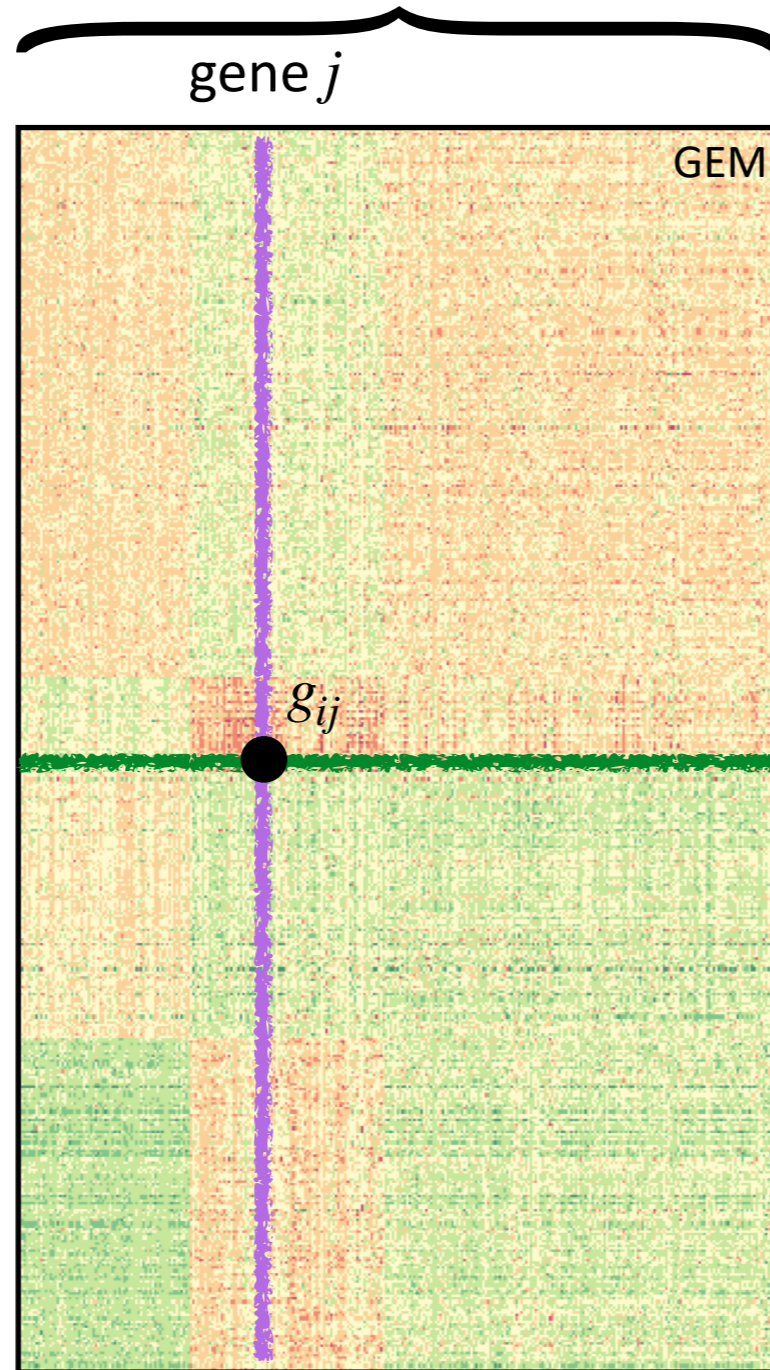
Gene Expression Matrix (GEM)

genes

their expressions depend on the biological condition groups of genes may be similarly expressed for certain conditions

samples
(biological conditions)

different cancer types and subtypes, normal samples, etc.



g_{ij} = expression level of gene j for sample i

$\mathbf{x}_i^s = i^{\text{th}}$ sample vector → describes the joint expression pattern of the genes for the biological condition of sample i

describes the joint expression pattern of gene j across all the biological conditions (samples)

$\mathbf{x}_j^g = j^{\text{th}}$ gene vector →

Gene Expression Matrix (GEM)

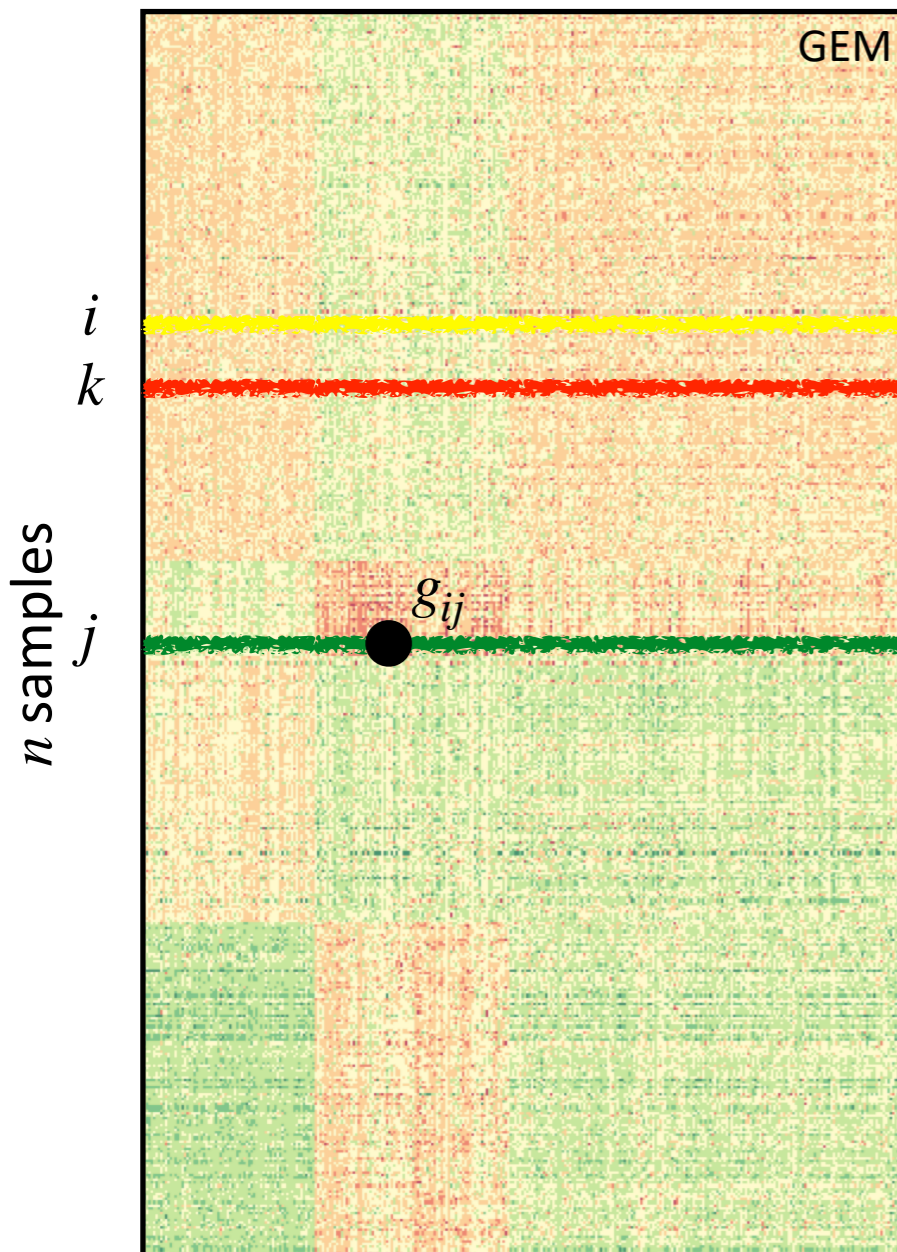
the sample space

samples become grouped according to their similarity in their expression patterns



groups reveal different biological conditions

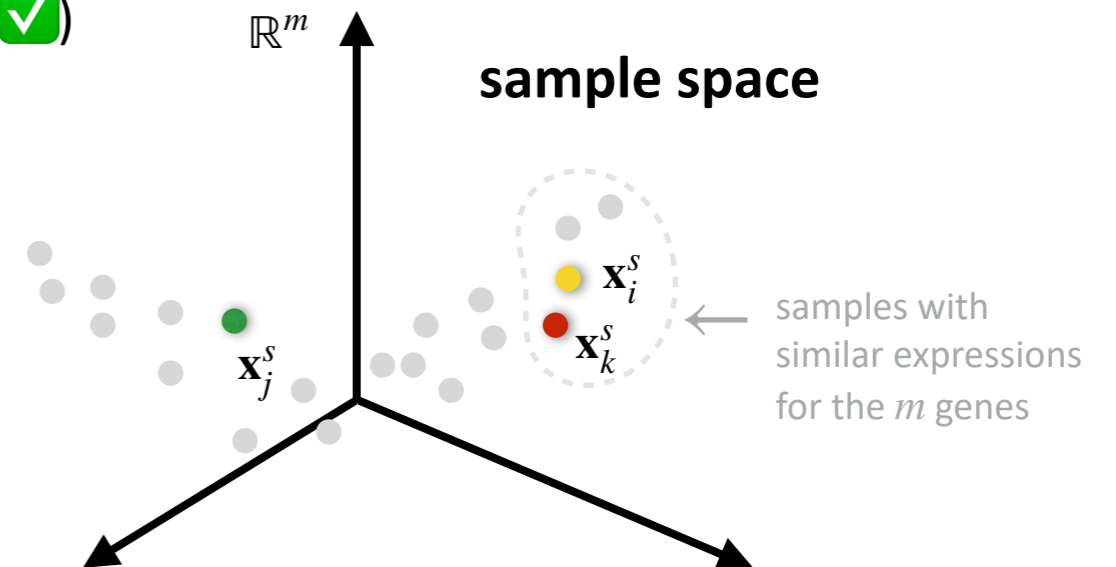
m genes



... but not all biological conditions are interesting:
(tissue **X** vs. cancer **✓**)

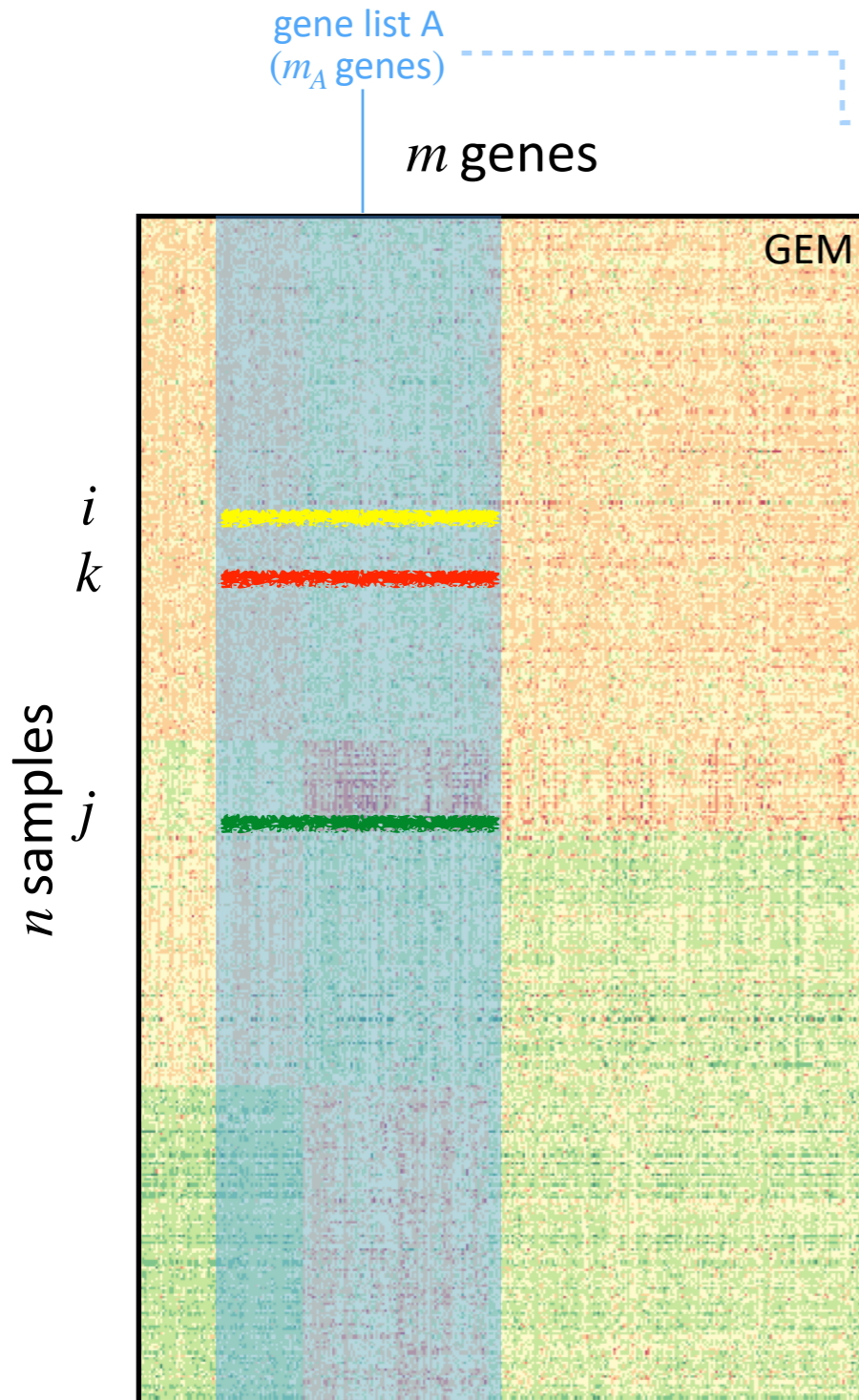


The m genes might be more sensitive to certain biological conditions (e.g. tissue differentiation) and less for others (e.g. cancer)



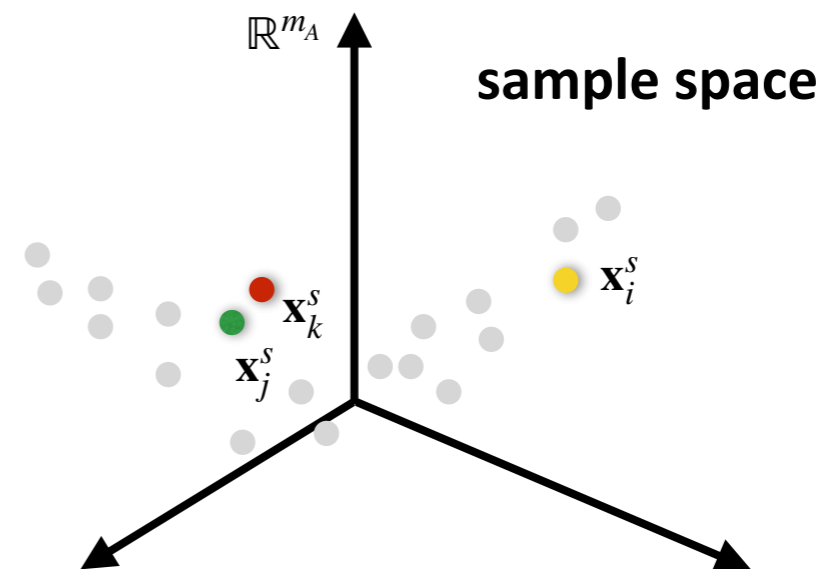
Gene Expression Matrix (GEM)

conditional grouping of samples



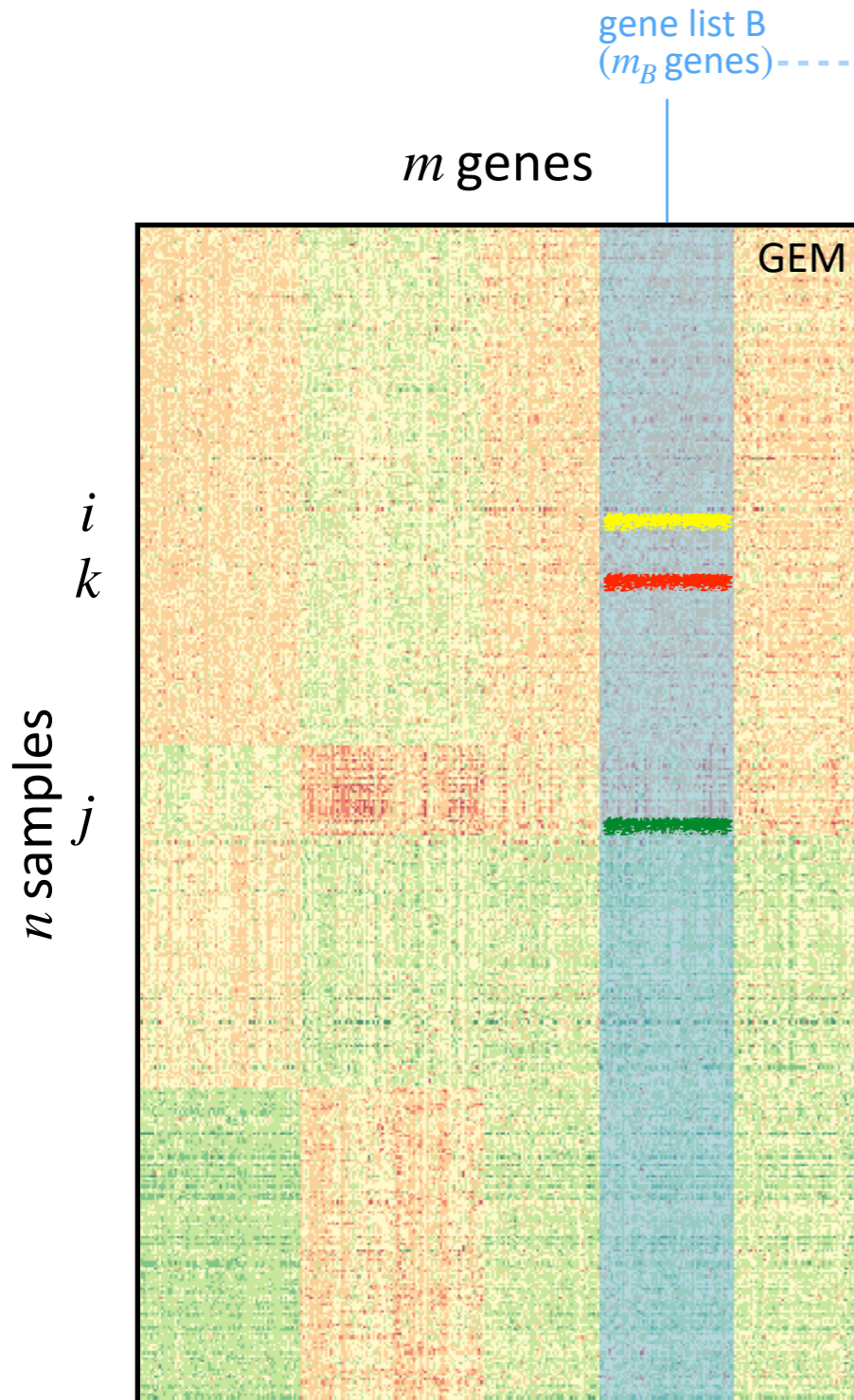
changing the set of genes,
the groupings change
in the sample space ...

... since different groups of genes
may be sensitive
to other biological conditions



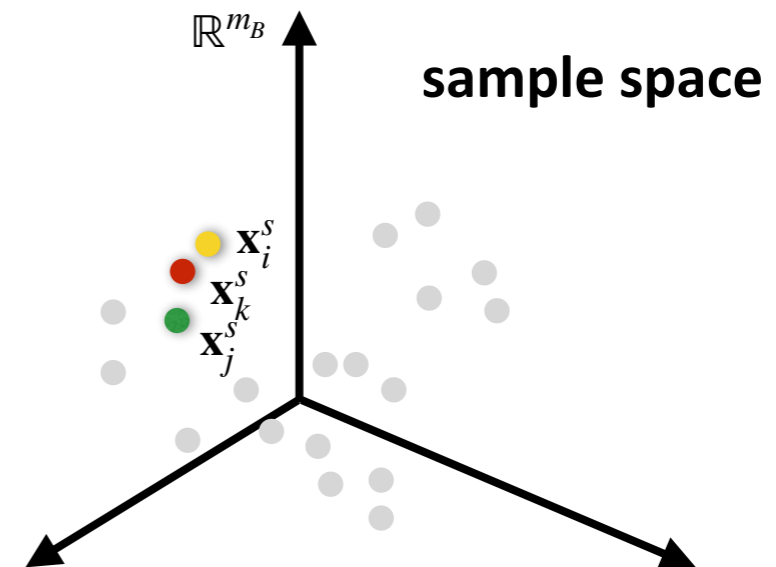
Gene Expression Matrix (GEM)

conditional grouping of samples



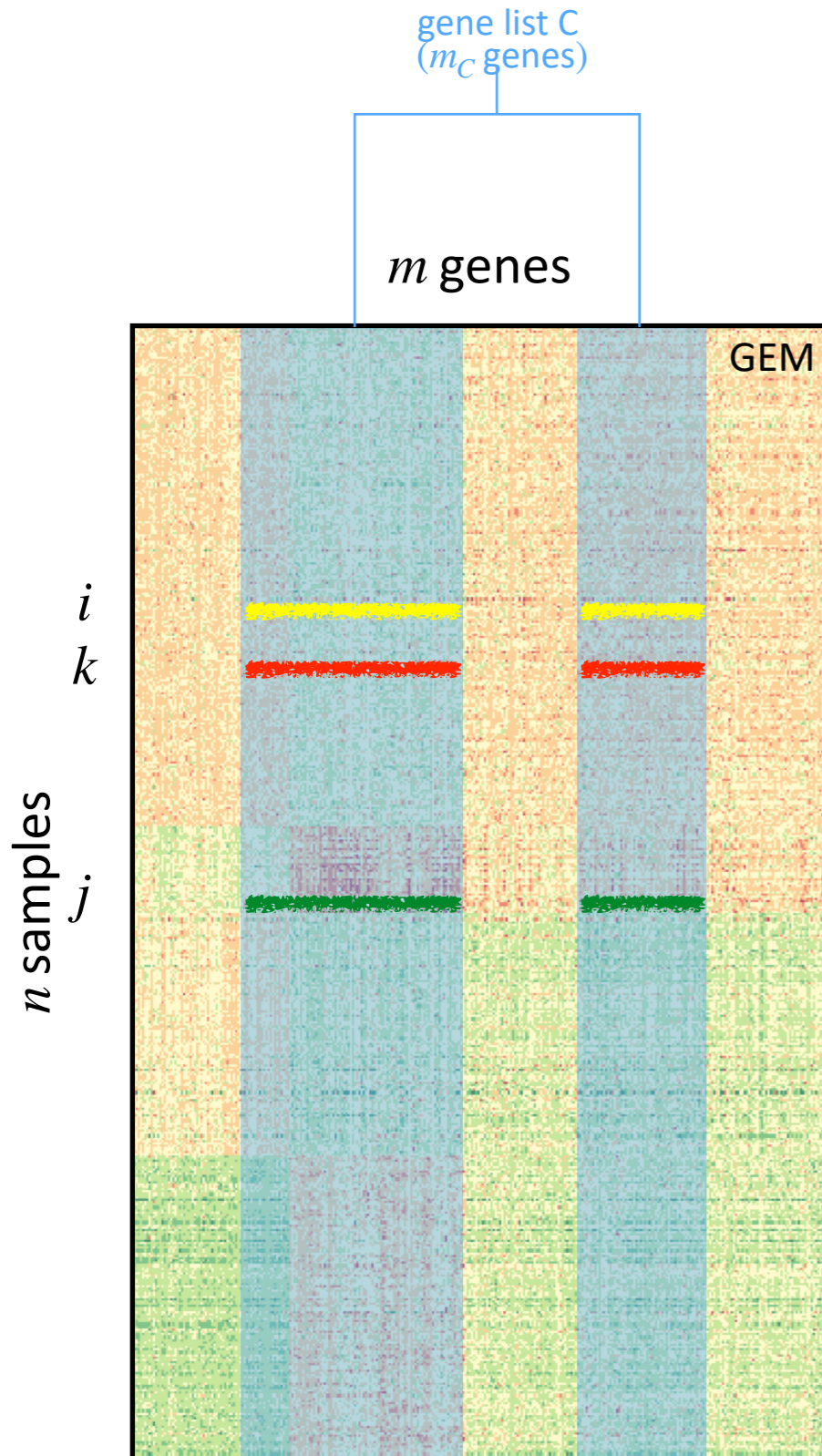
changing the set of genes,
the groupings change
in the sample space ...

... since different groups of genes
may be sensitive
to other biological conditions

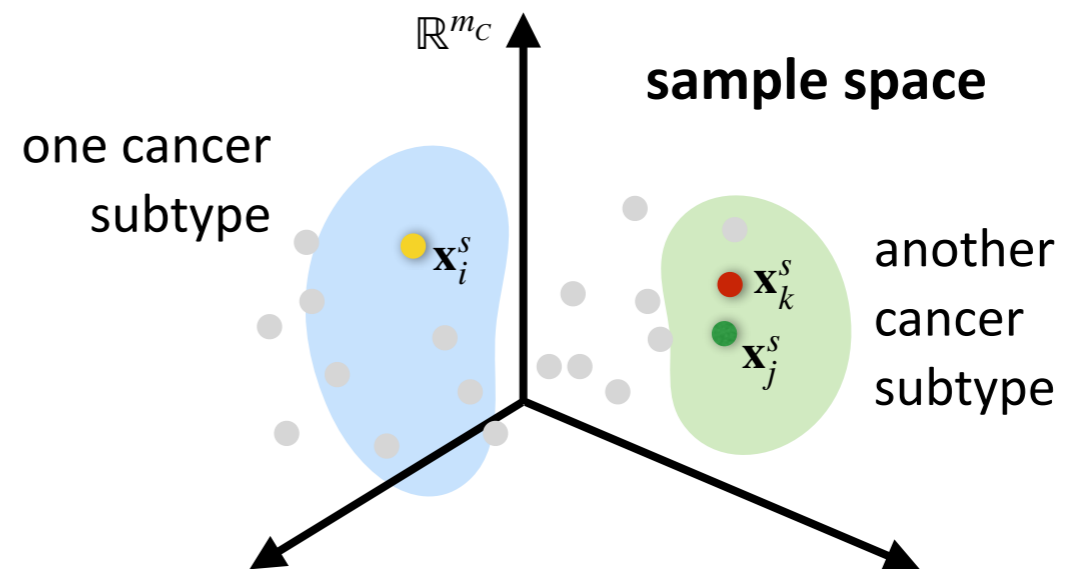


Gene Expression Matrix (GEM)

conditional grouping of samples



... and some selections of genes could be **relevant** !

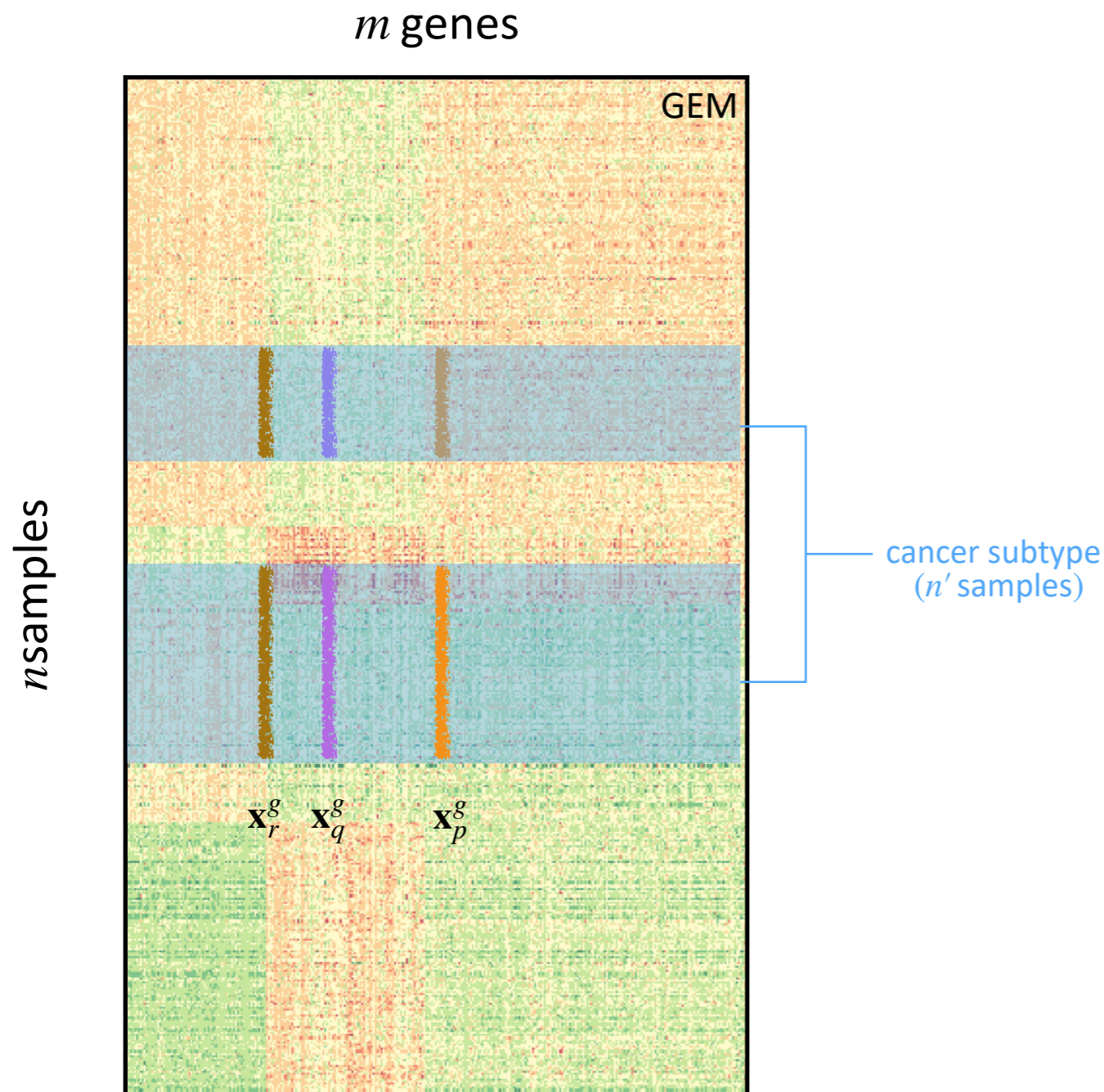


gene list C

- is sensitive to certain cancer condition
- can suggest biomarkers
- the gene list C itself is already useful knowledge !

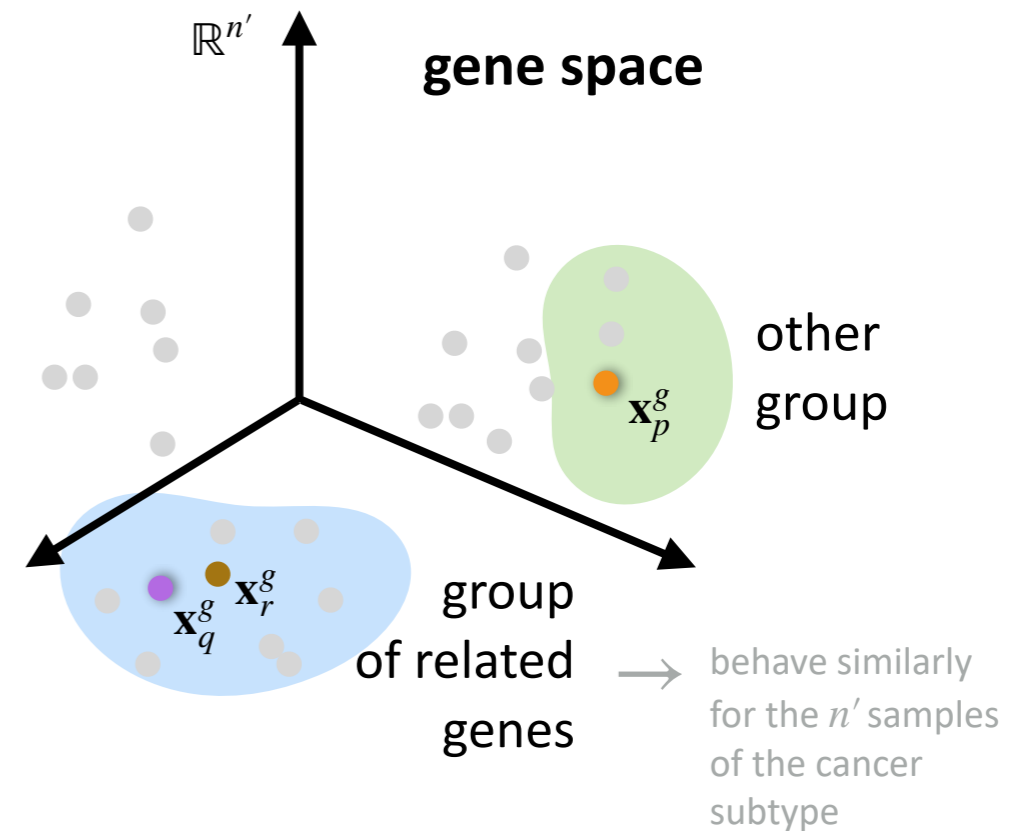
Gene Expression Matrix (GEM)

conditional grouping of samples



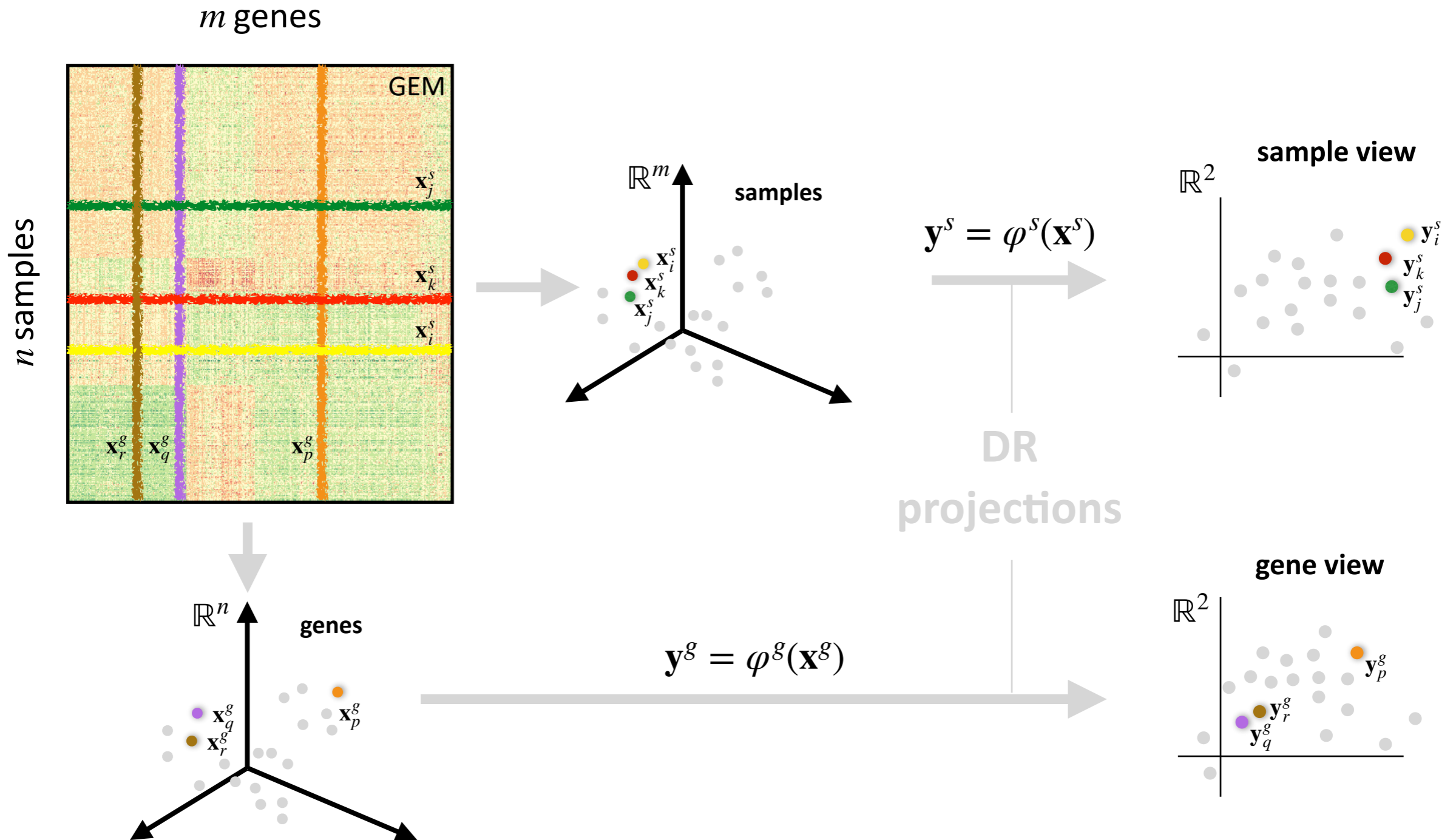
Dual problem

In the **gene space** genes become grouped according to their expression behavior for a given biological condition (e.g. a cancer subtype)



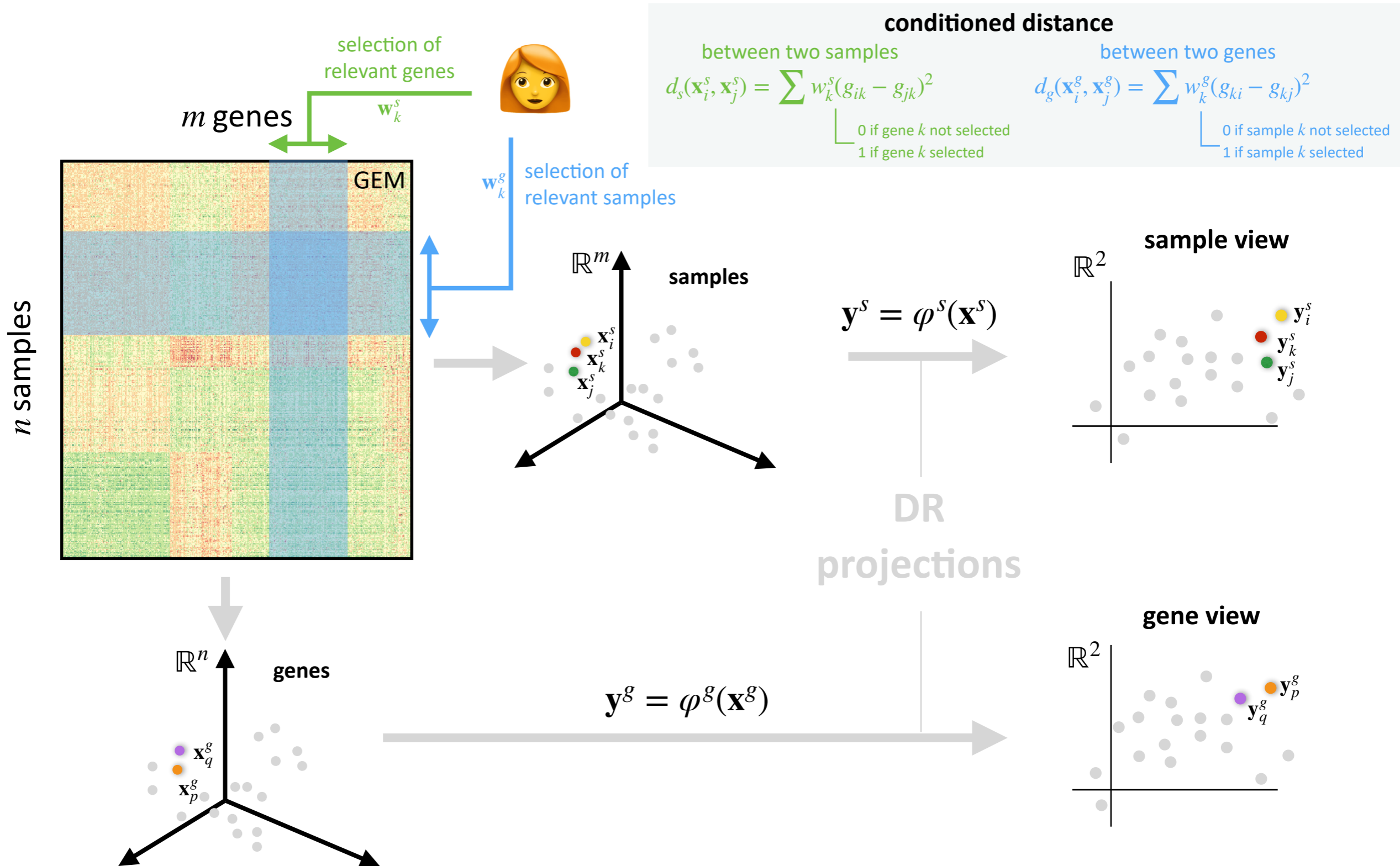
DR projections of samples and genes

visualization of the sample and gene space



DR projections of samples and genes

interactive visualization by selection of genes and samples



DR algorithm as a dynamical system

An interactive DR (e.g. tSNE) algorithm can be seen as a *state-space* dynamical system in a configuration space [3,4]

$\mathbf{y}(k), \mathbf{u}(k)$

configuration vectors

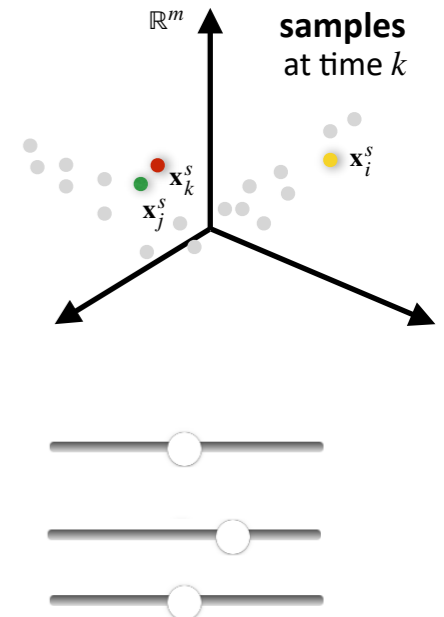
flattened vectors with all the coordinates of a set of points or elements

e.g., $\mathbf{y} := (y_{11}, y_{12}, y_{21}, \dots, y_{n1}, y_{n2})$

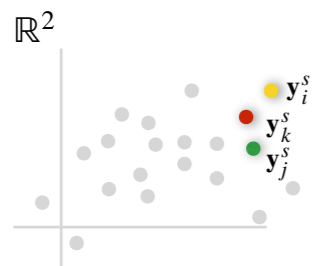
$$\mathbf{y}(k + 1) = f(\mathbf{y}(k), \mathbf{u}(k))$$

$\{\mathbf{x}_i^s\}$ set of all input data points at time k

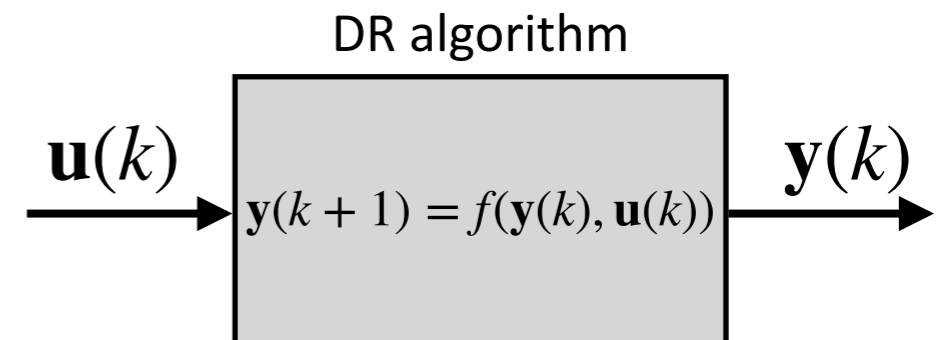
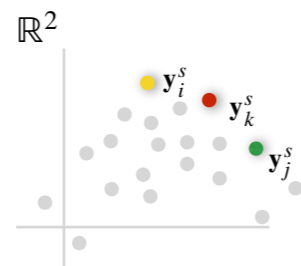
$\mathbf{p}(k)$ algorithm parameters (perplexity, weights on attributes, ...) at time k

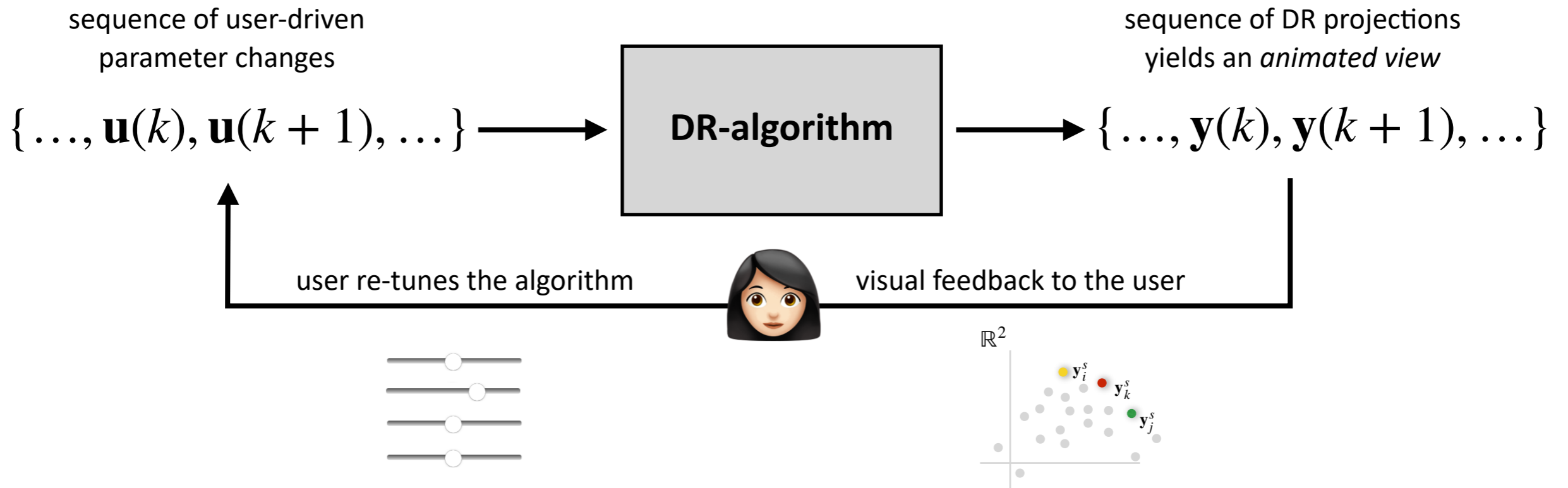


2D projections at sample $k + 1$



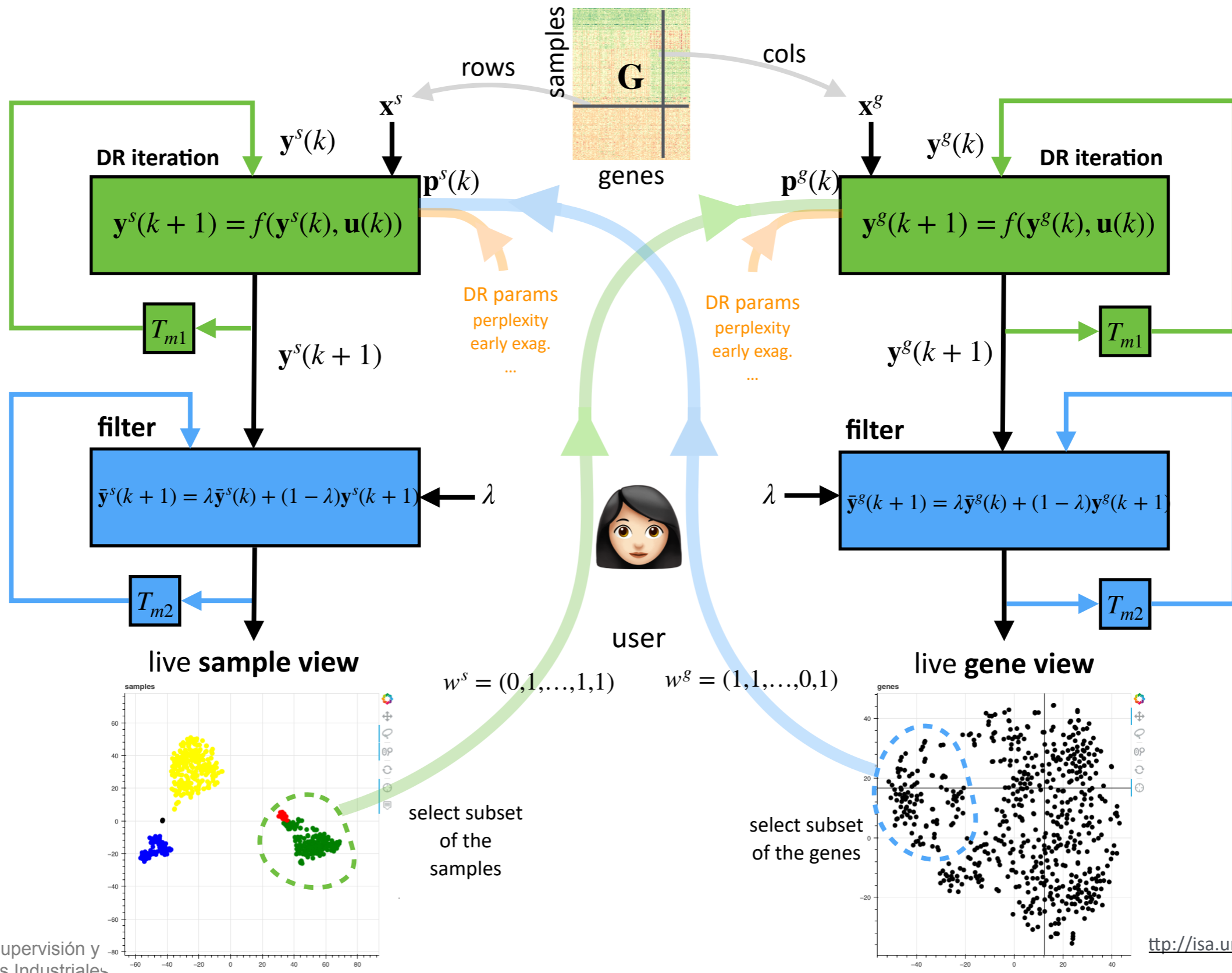
2D projections at sample k





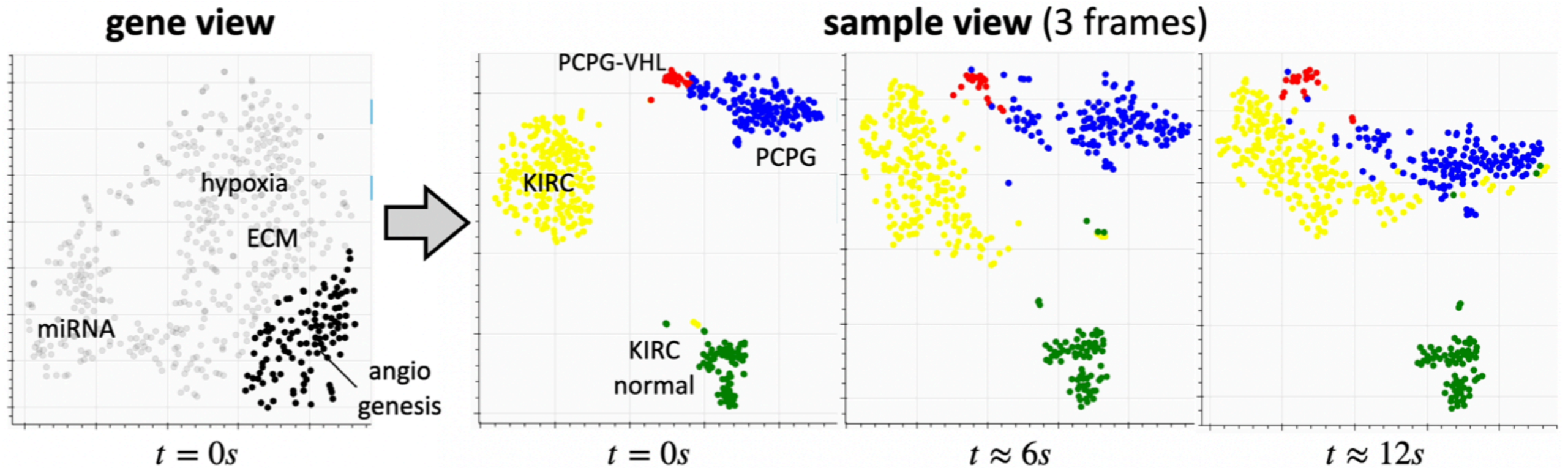
Interactive Dimensionality Reduction

interactive dual projections for gene expression analysis



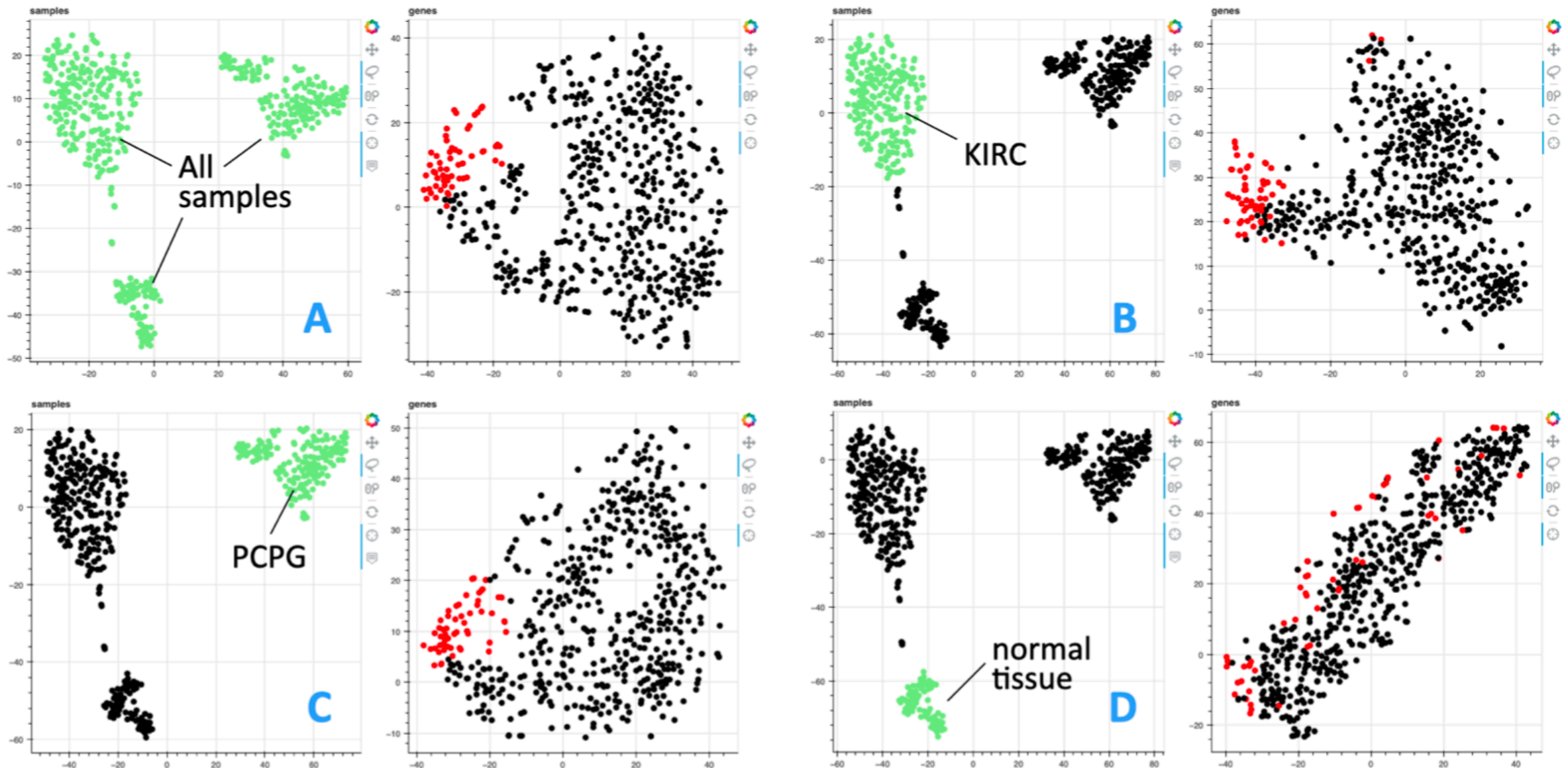
Results: case 1

Impact of angiogenesis genes in characterization of KIRC/PCPG



angiogenesis genes undergo clearly different expression activity for PCPG and KIRC w.r.t. normal

Different behavior of miRNA group for cancer subtypes



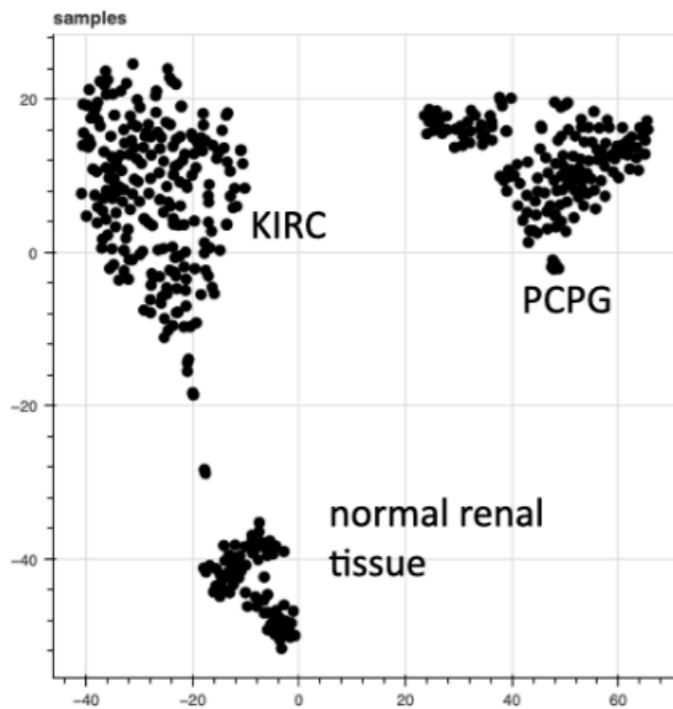
expression behavior of the miRNA set is clearly different for normal tissue

Results: case 3

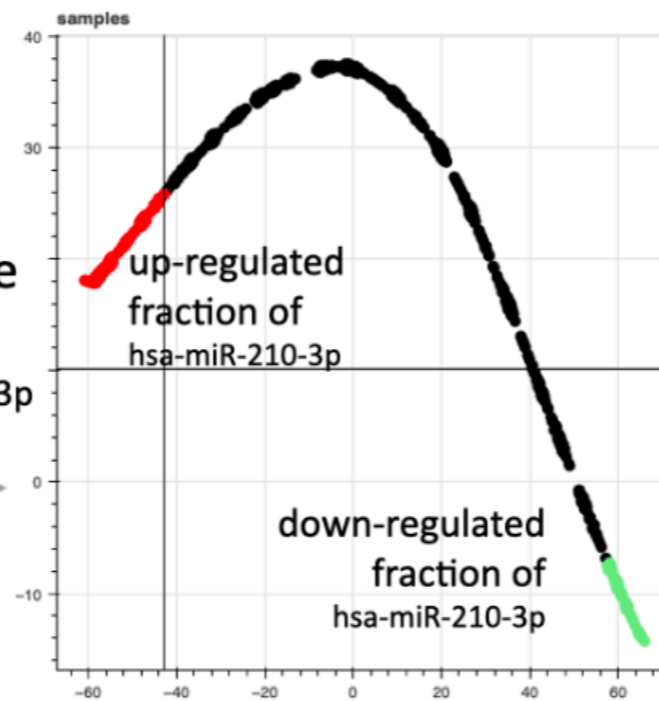
revealing location of up/down regulated hsa-miR-210p

mark the
upper and lower tails

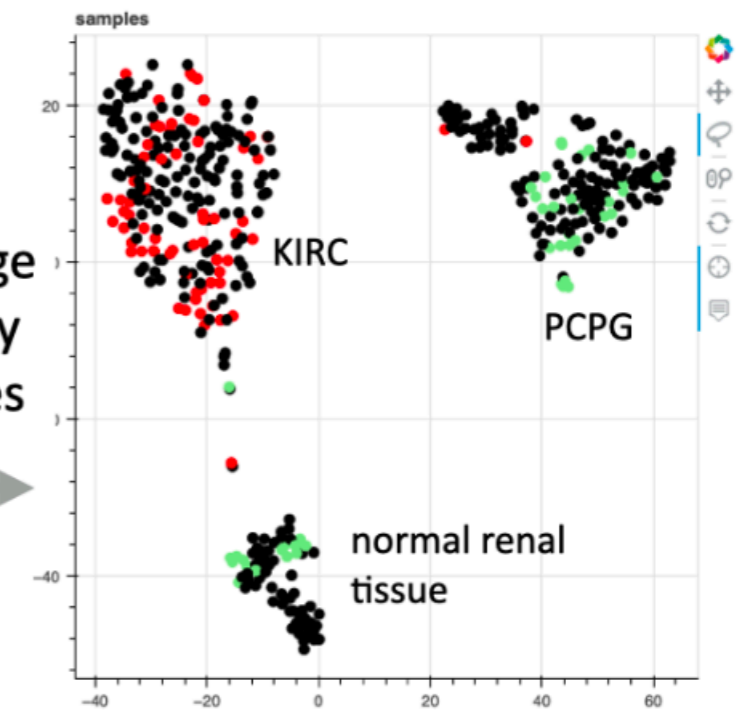
track the positions of
up-regulated and down-regulated
samples



rearrange
by
hsa-miR-210-3p



rearrange
again by
all genes



💡 miRNA 210
differentially
expressed
for KIRC

dual

- visual *map* of genes
- visual *map* of samples

conditional

- constrain to a subset of biological conditions
- constrain to a subset of genes
- ... other kinds of constraints are possible

animated transitions

- track interesting items (genes or samples) upon changing conditions
- immediate feedback in the analysis allowing to get fast & continuous *question-answer* cycle

potential for biomedical knowledge discovery

- discover sets of genes that best differentiate biological conditions → biomarkers
- discover sets of genes behaving similarly for a given condition → discovering roles of genes → pathways
- focus on a single gene/sample or sets of them → analyze within context of other genes/samples