

Bringing single-cell data analysis on GARR Cloud Platform

Borsista: Alessandrì Luca
Tutor: Prof. Beccuti Marco

Unito, Dipartimento di
informatica. Bando 1/19



GIORNATA DI INCONTRO
BORSE DI STUDIO GARR
“ORIO CARLINI”
ROMA





Indice

- Single-cell Analysis, RNA-seq

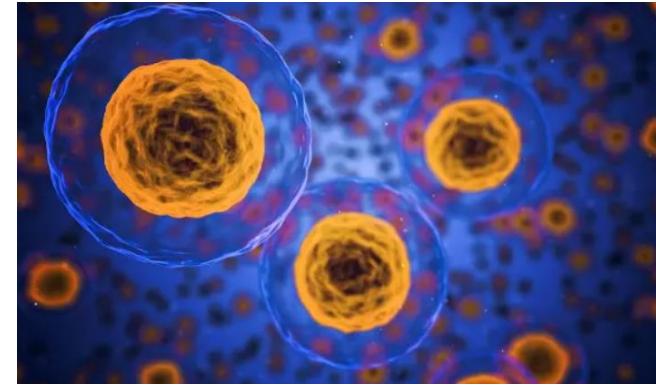
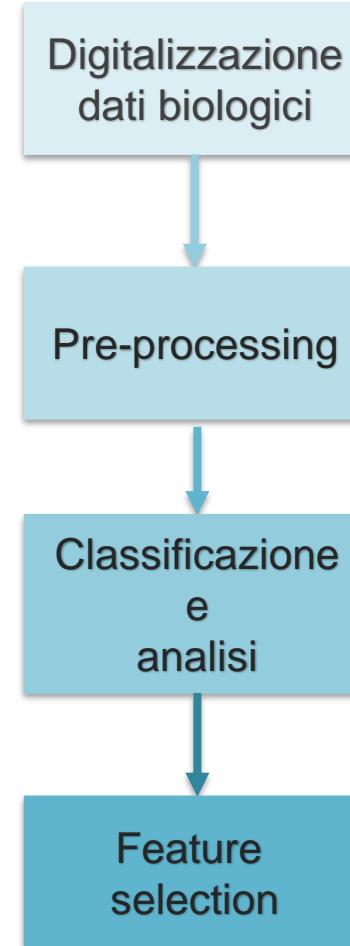


Immagine:<https://www.technologynetworks.com/genomics/blog/overcoming-challenges-in-single-cell-analysis-296313>



Indice

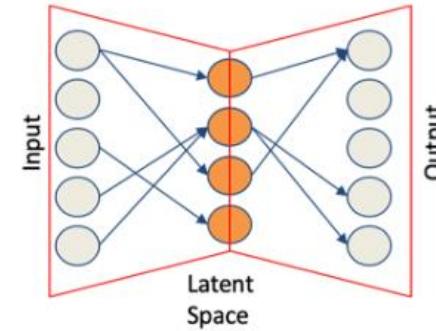
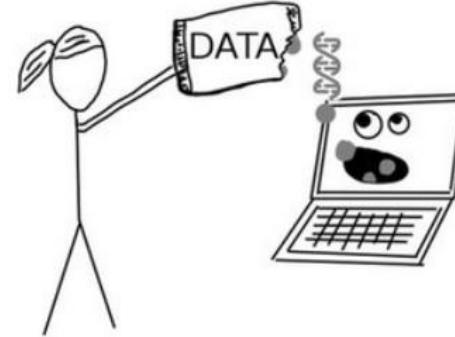
- Single-cell Analysis, RNA-seq
- rCASC workflow





Indice

- Single-cell Analysis, RNA-seq
- rCASC workflow
- Autoencoder for Data mining





Indice

- Single-cell Analysis, RNA-seq
- rCASC workflow
- Autoencoder for Data mining
- rCASC Galaxy Implementation





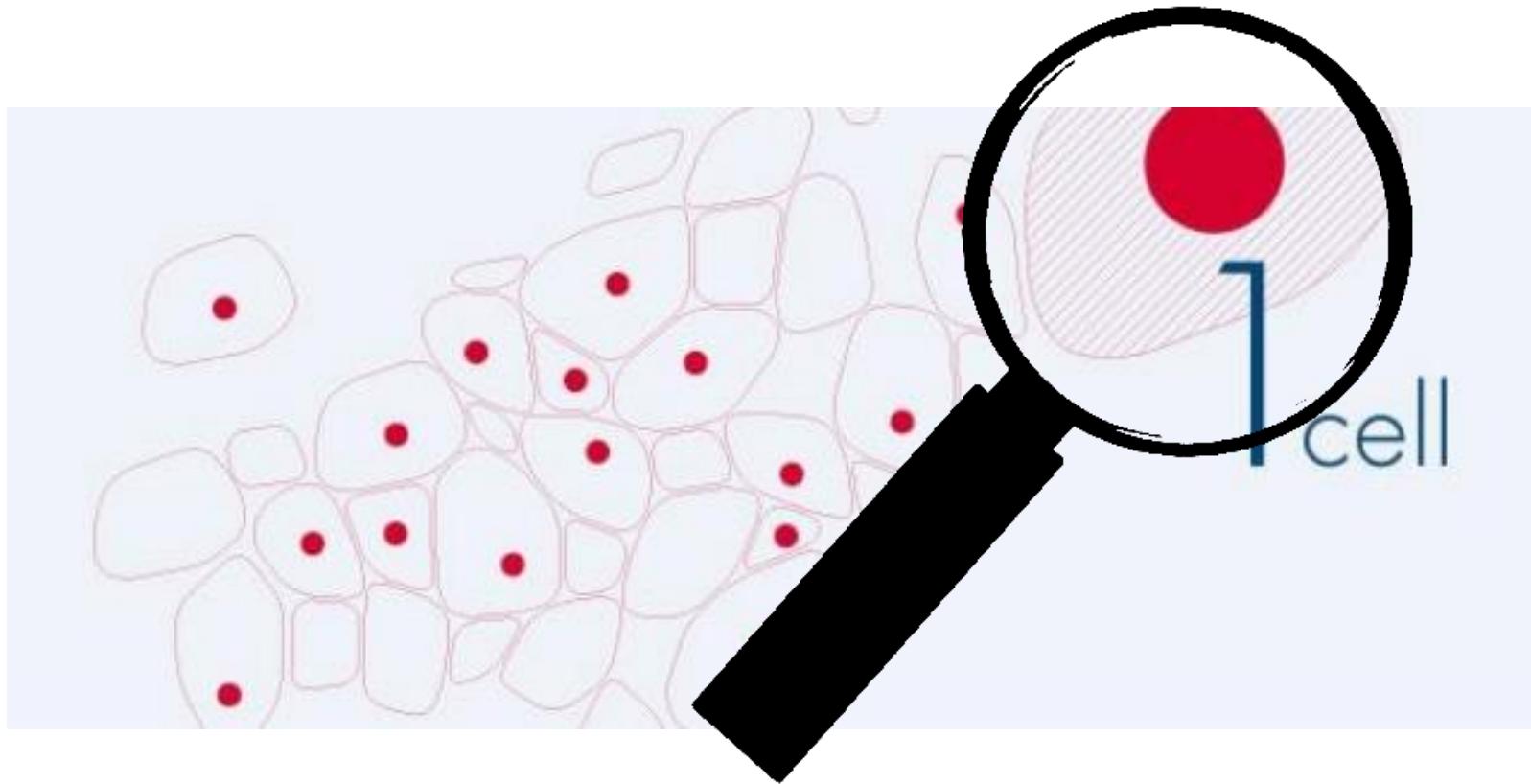
Indice

- Single-cell Analysis, RNA-seq
- rCASC workflow
- Autoencoder for Data mining
- rCASC Galaxy Implementation
- Proposal Project 2° year



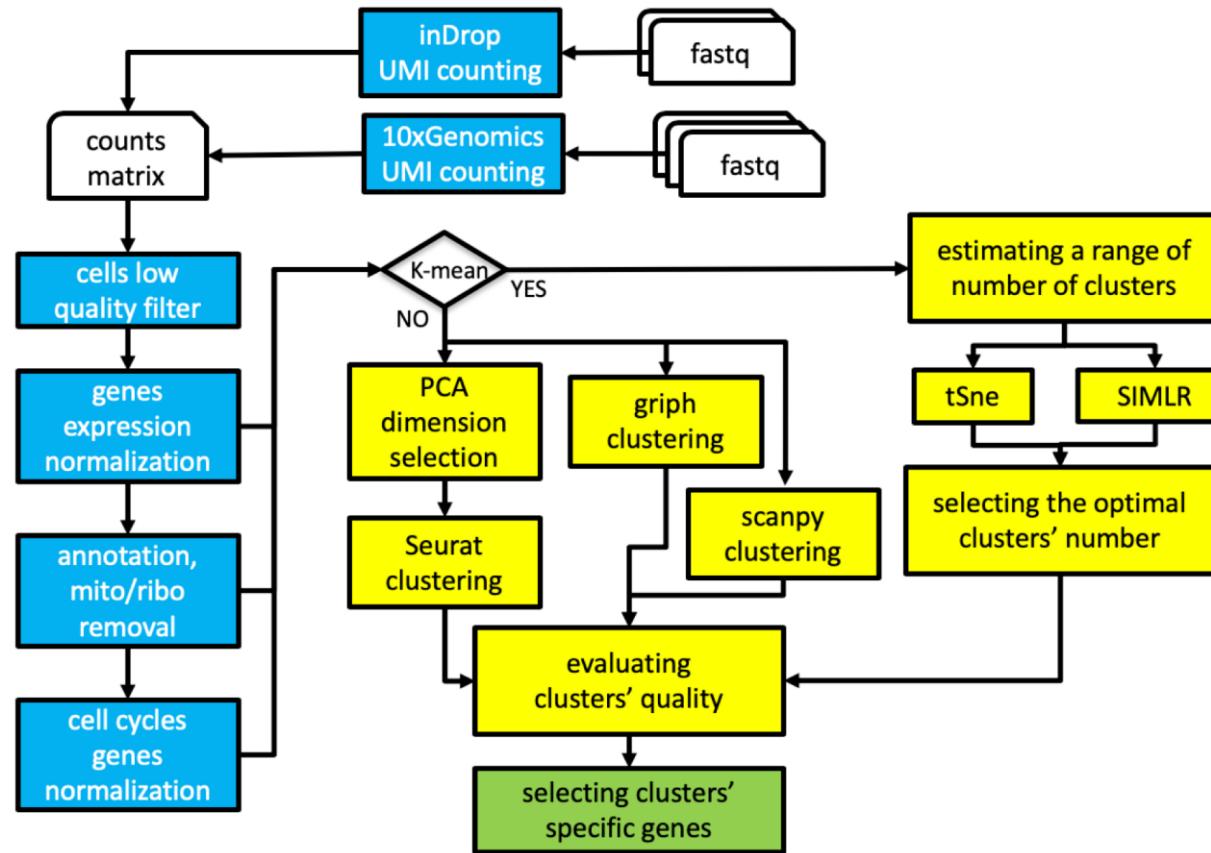


Single-cell Analysis, RNA-seq



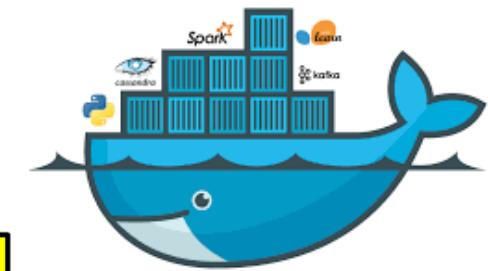
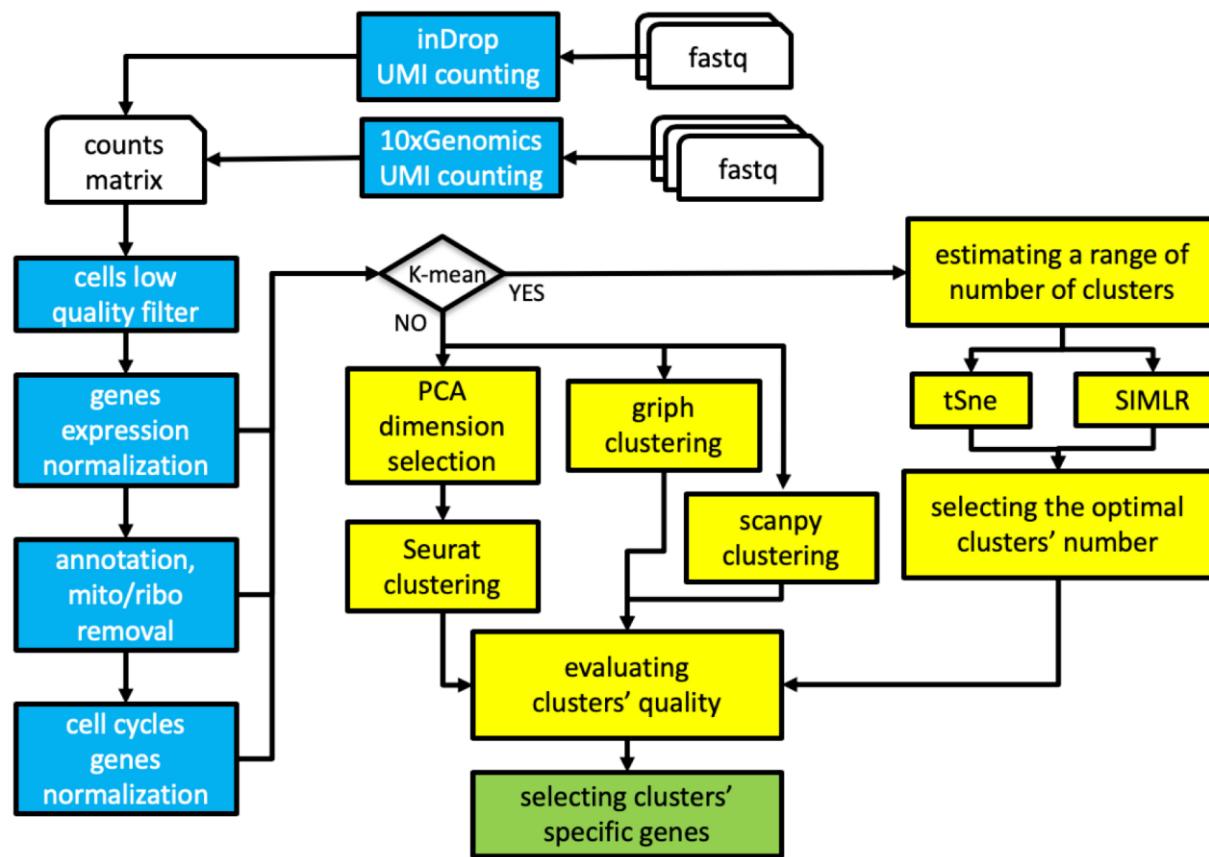


rCASC (reproducible classification and analysis for single-cell) workflow



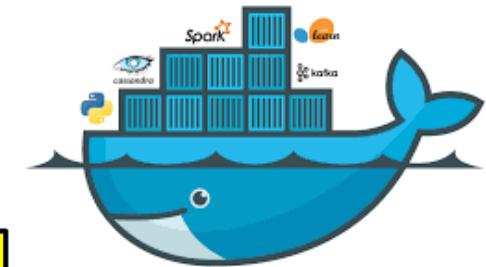
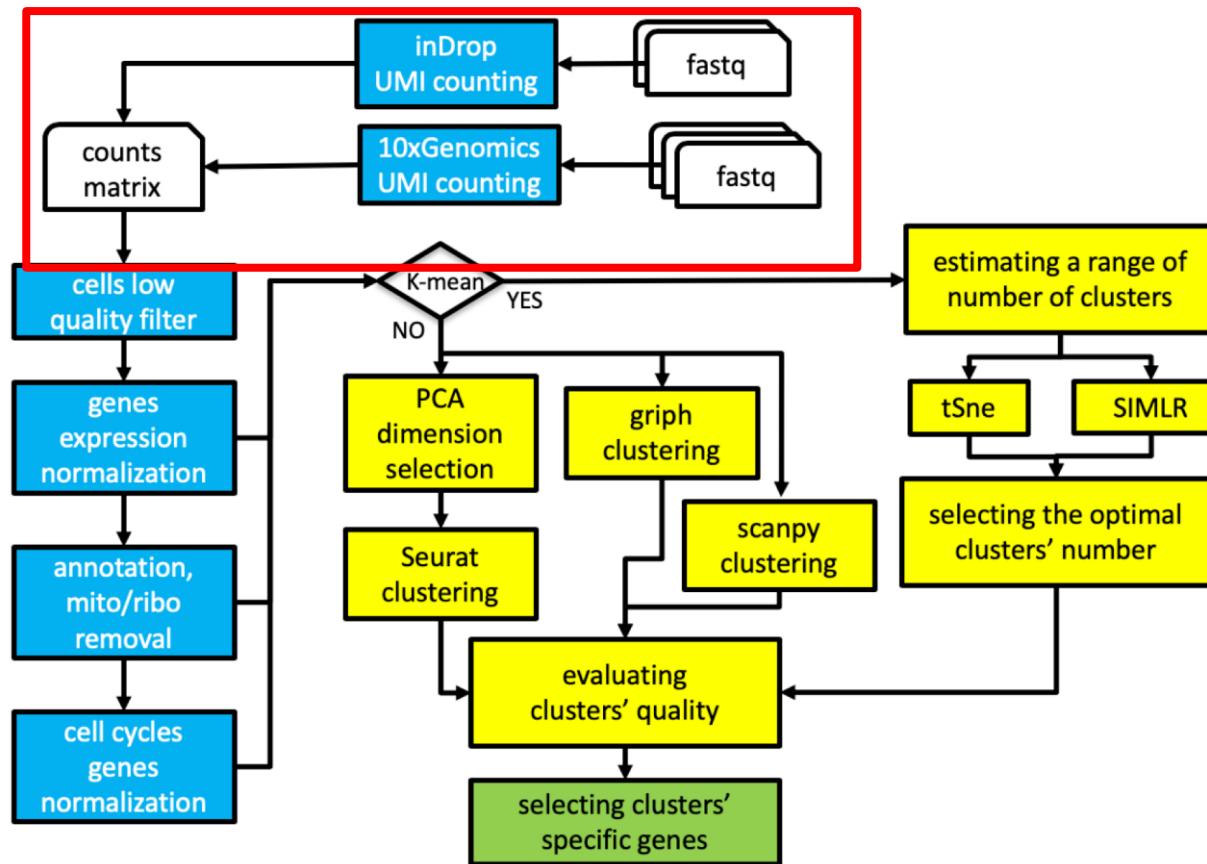


rCASC (reproducible classification and analysis for single-cell) workflow



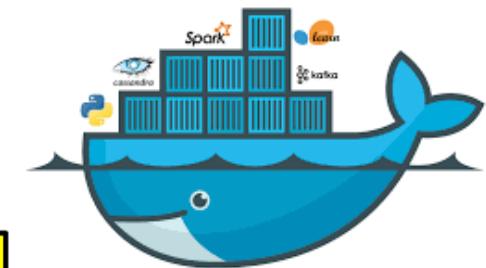
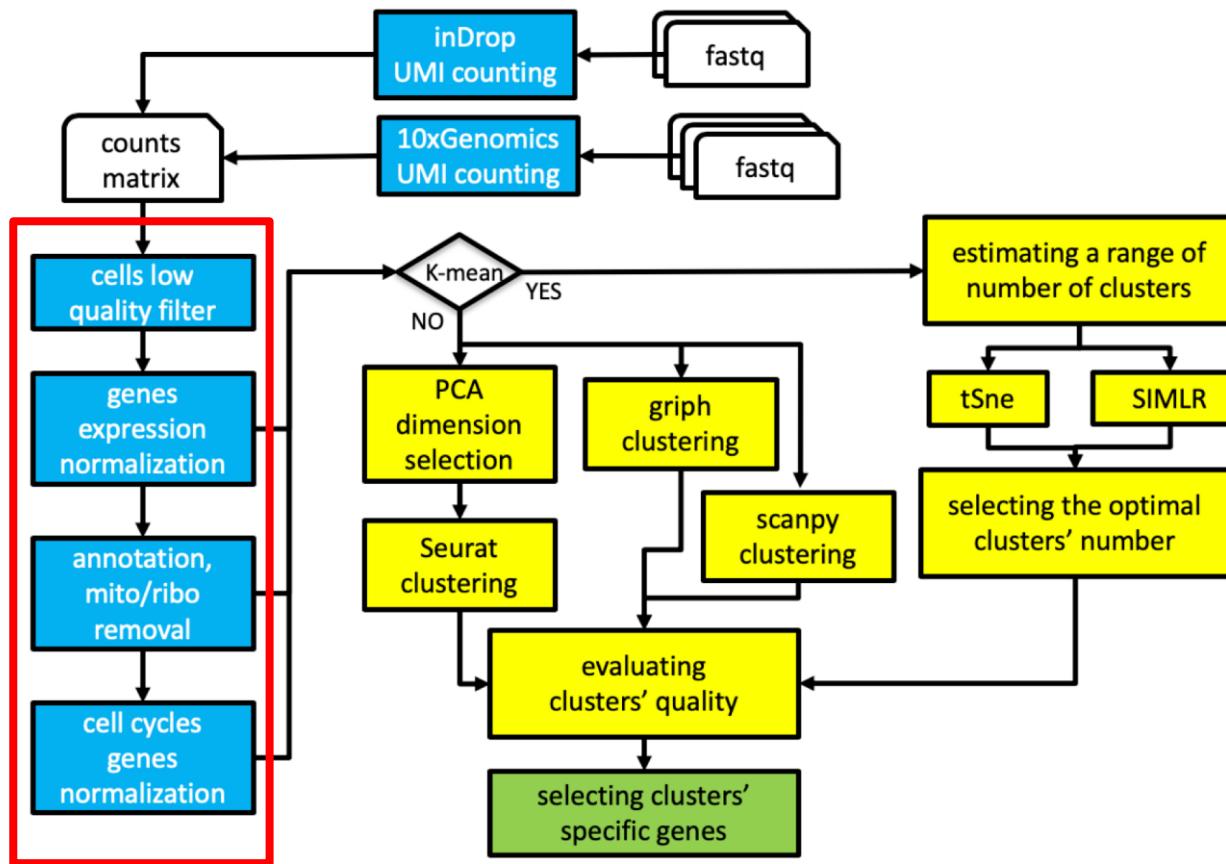


rCASC (reproducible classification and analysis for single-cell) workflow



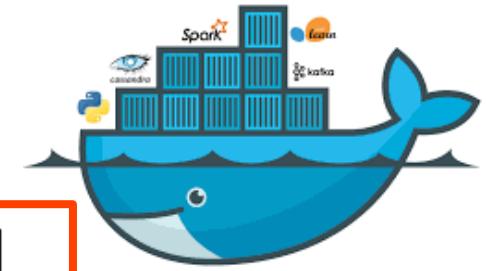
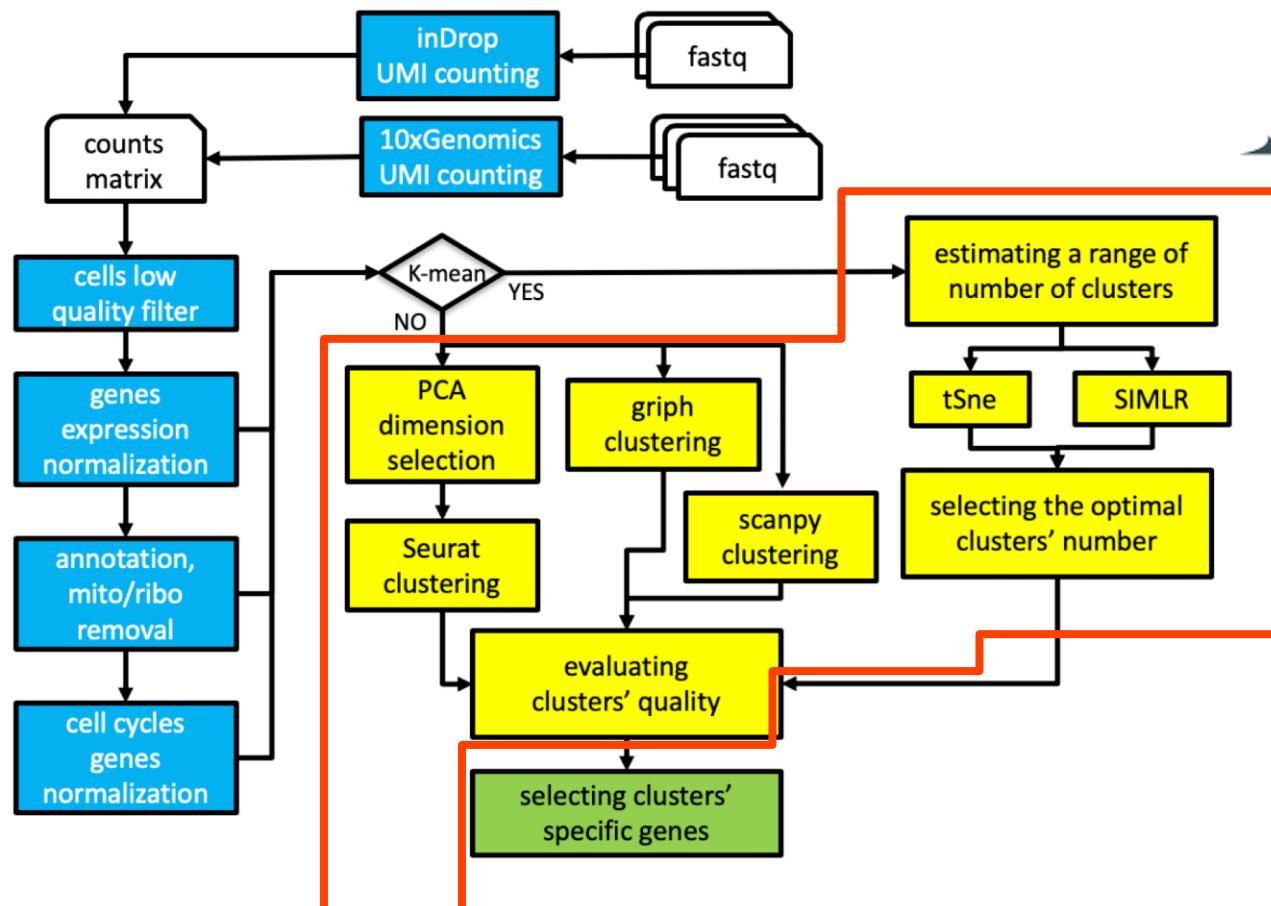


rCASC (reproducible classification and analysis for single-cell) workflow



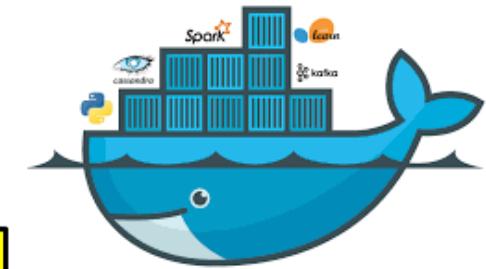
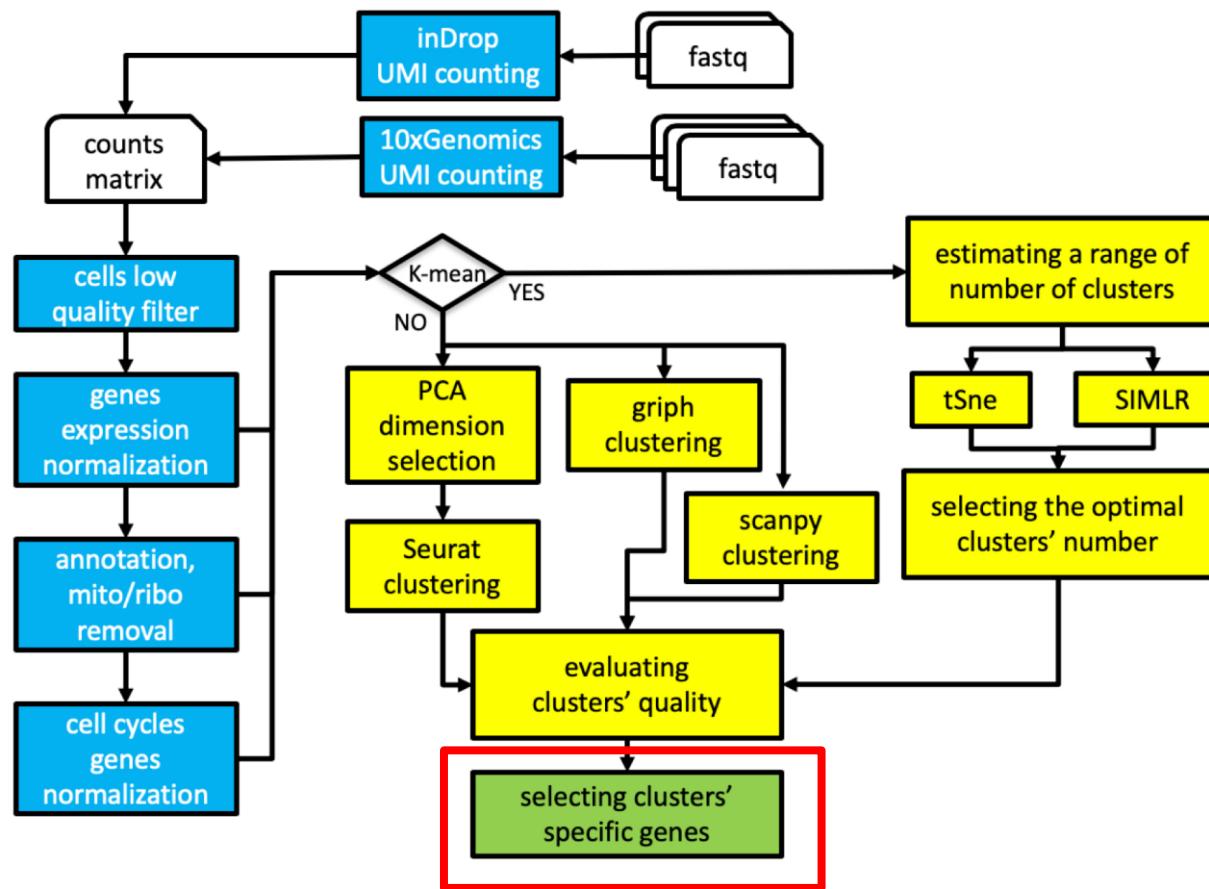


rCASC (reproducible classification and analysis for single-cell) workflow



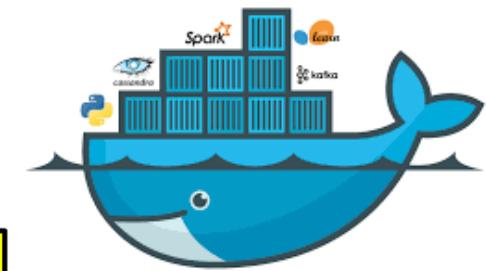
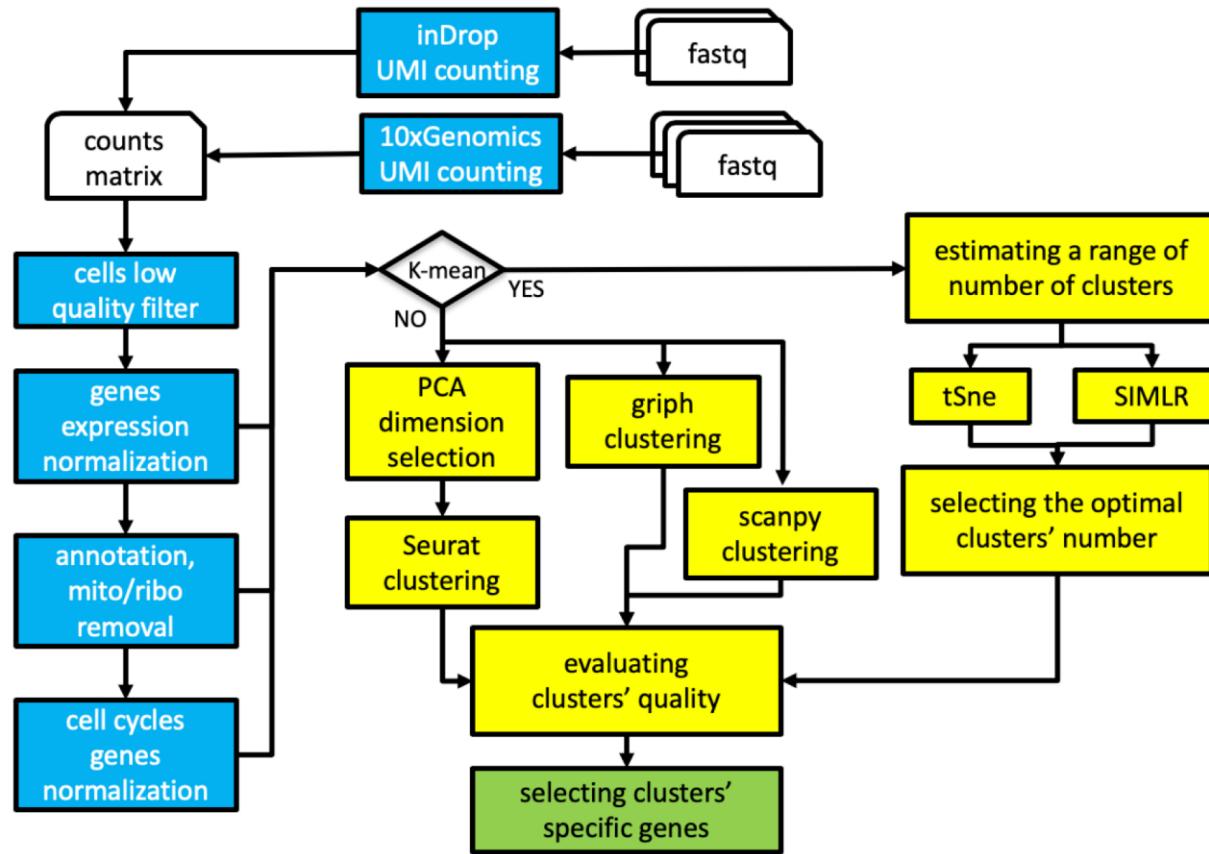


rCASC (reproducible classification and analysis for single-cell) workflow



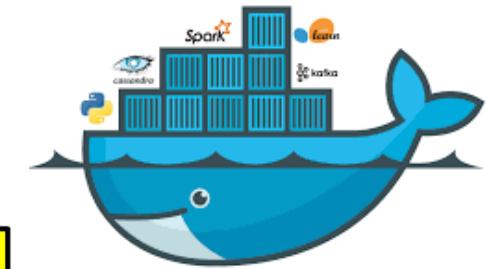
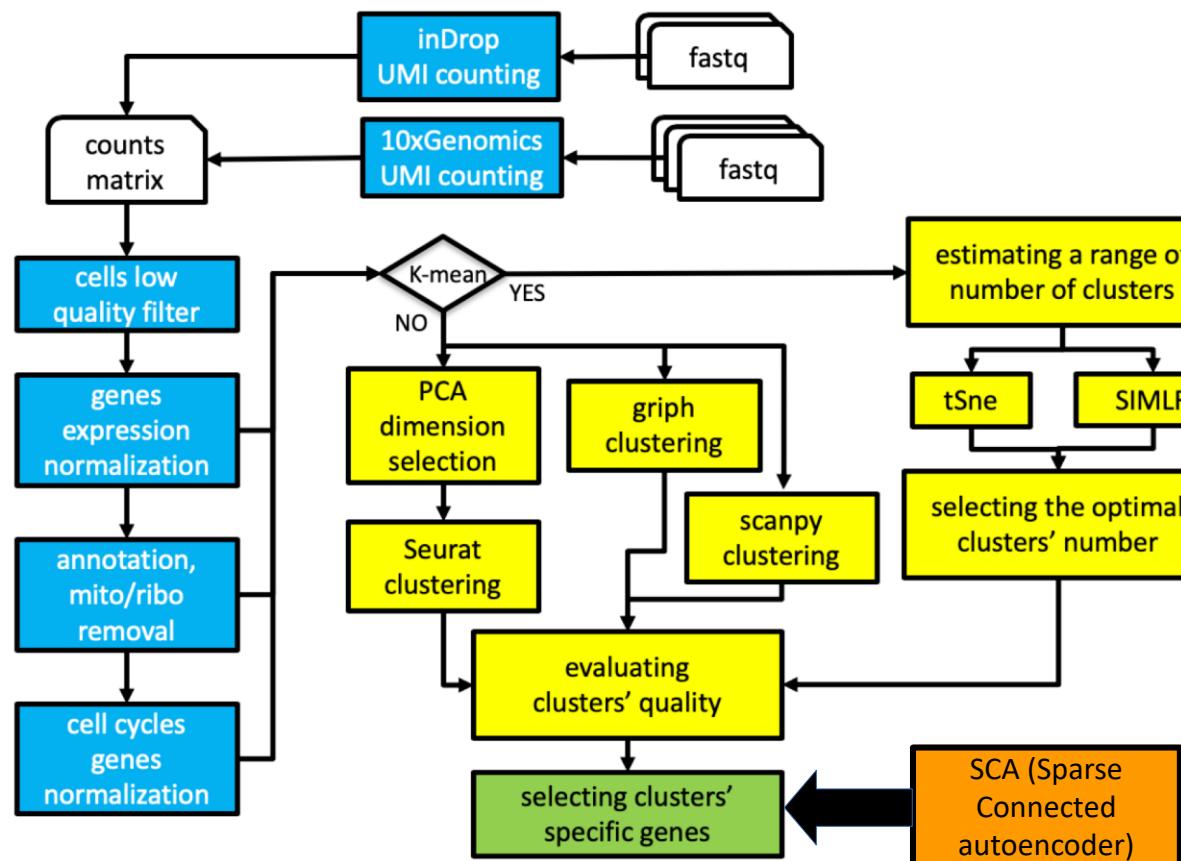


rCASC (reproducible classification and analysis for single-cell) workflow





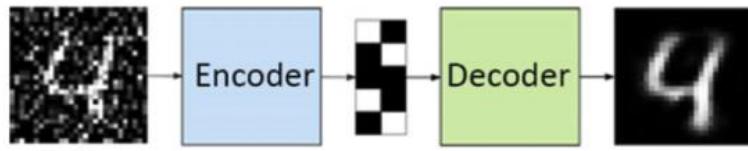
rCASC (reproducible classification and analysis for single-cell) workflow



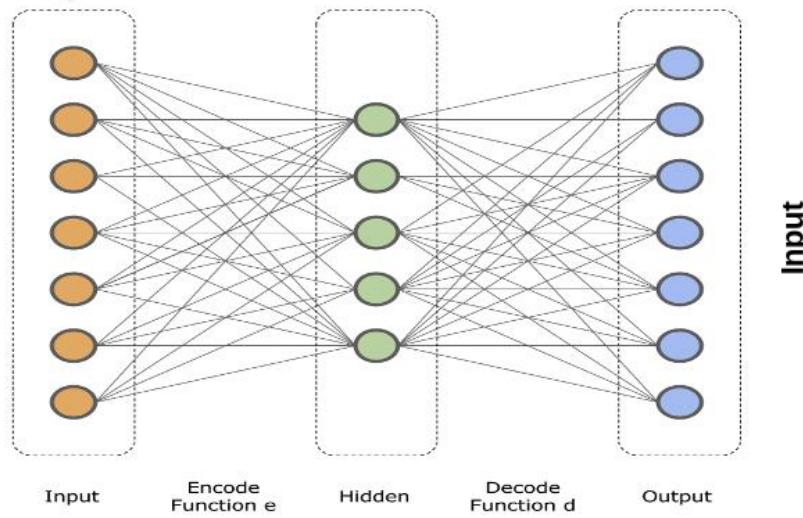


Autoencoder

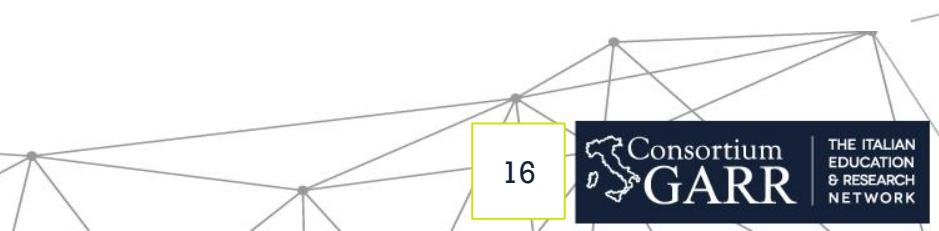
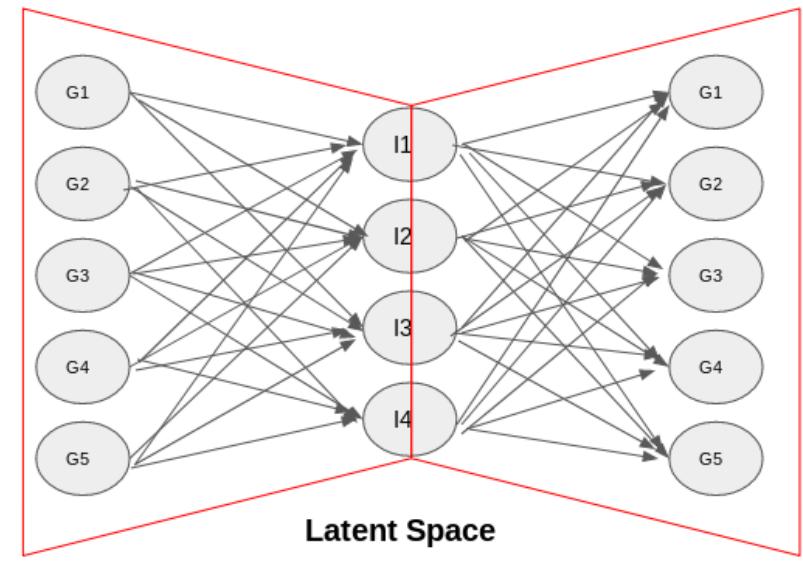
a)



b)

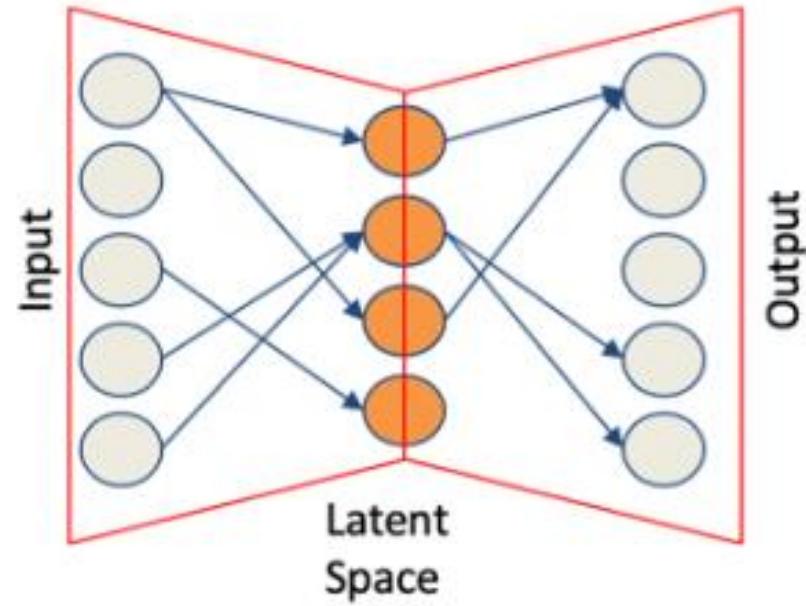
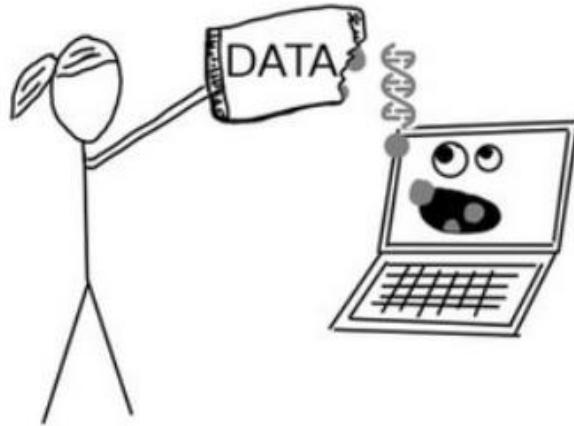


Input





Sparse Autoencoder

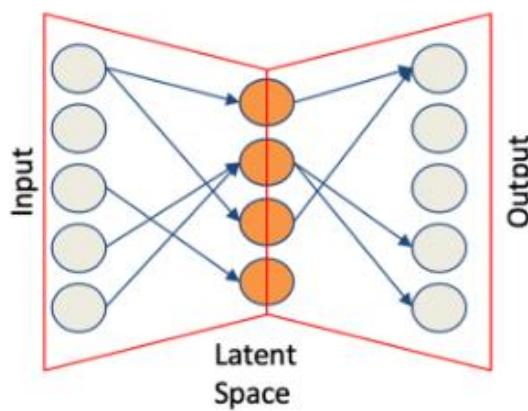
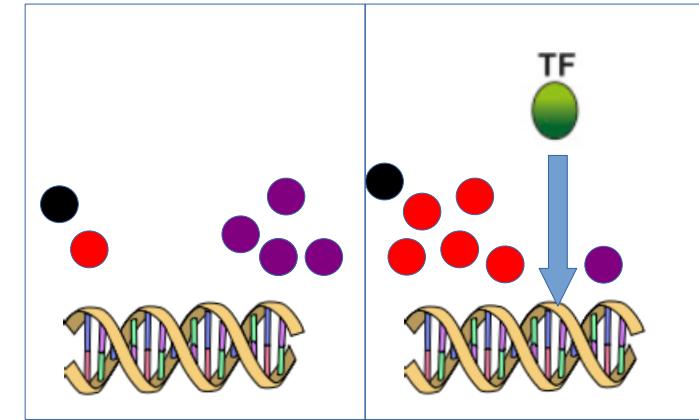
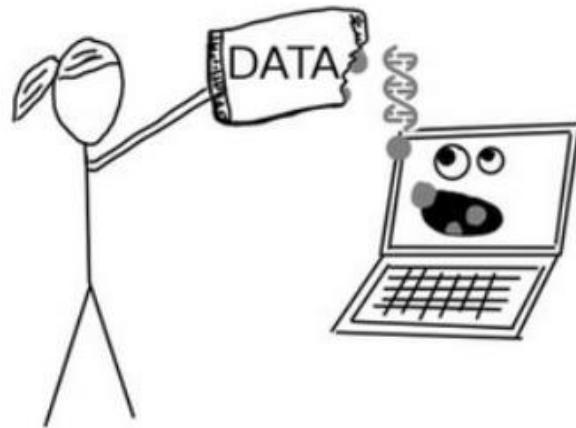


Input = Espressione del gene (**feature**) per ogni cellula (**campione**)

Latent Space = Informazione biologica nota



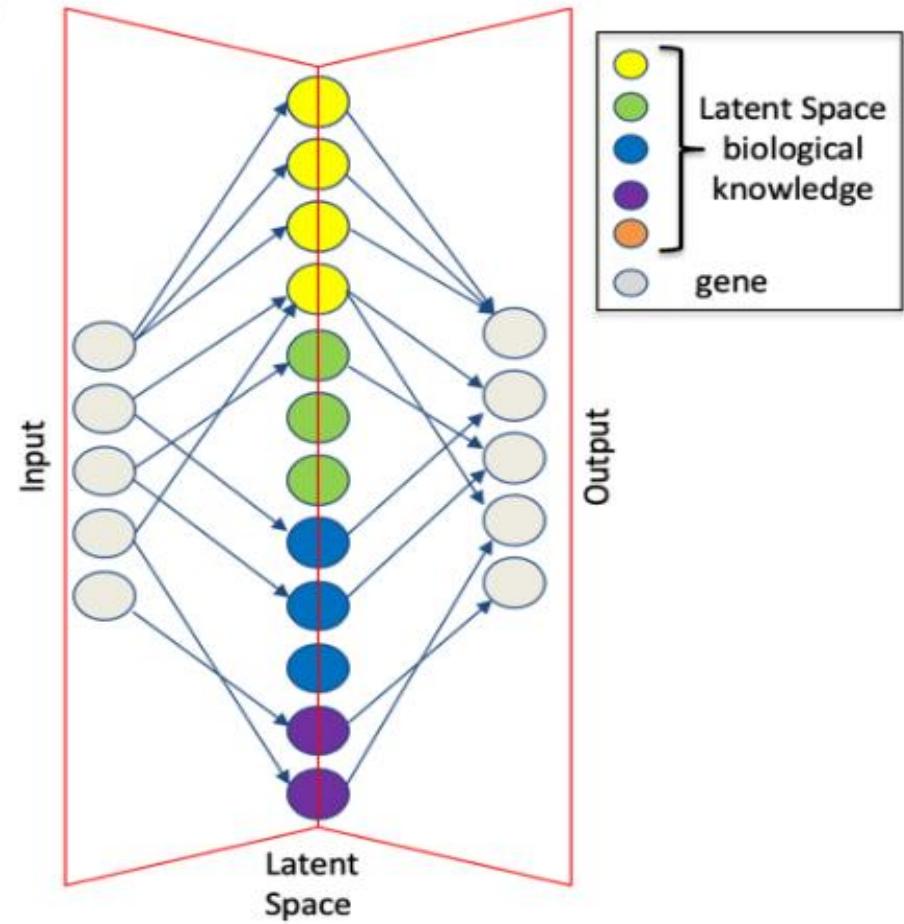
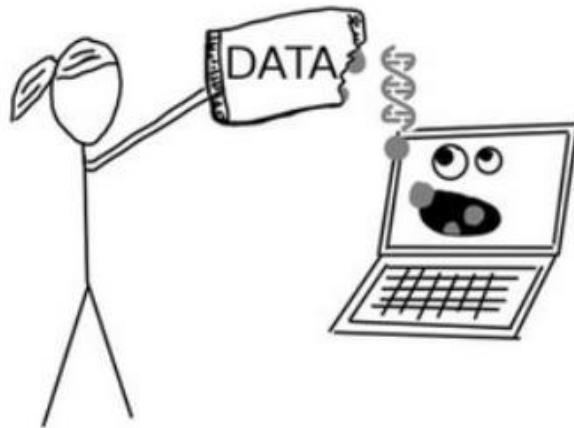
Sparse Autoencoder Relation Matrix



Gene Nero	NULL
Gene Rosso	TF Verde
Gene Viola	TF Verde

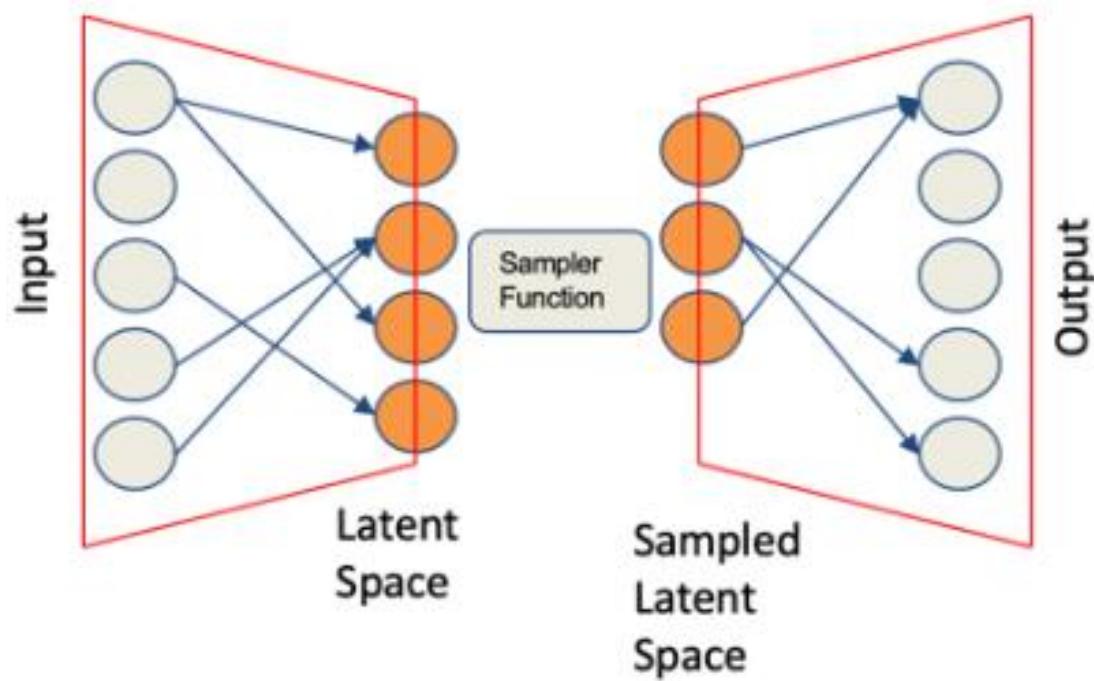
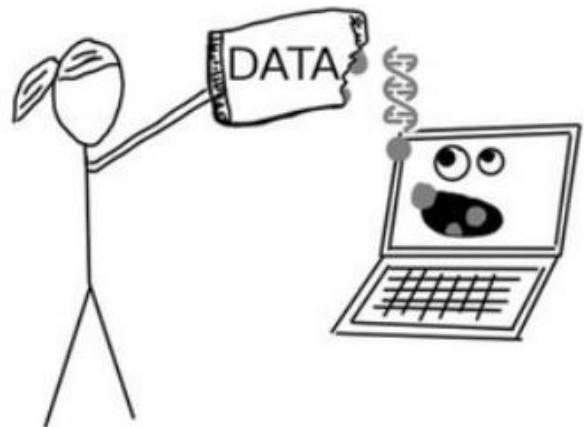


Sparse Autoencoder with multiple biological knowledge



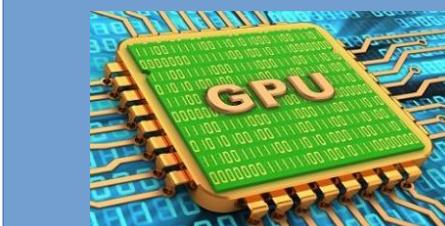
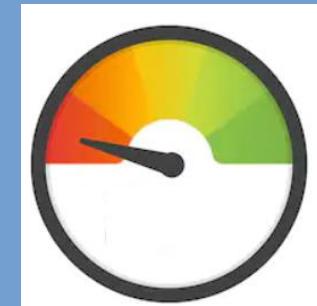
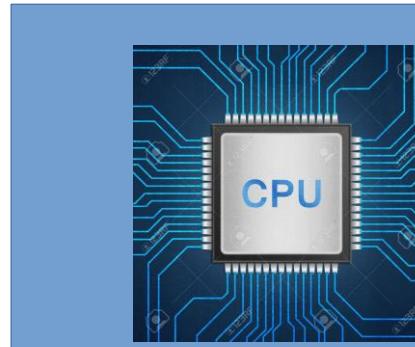
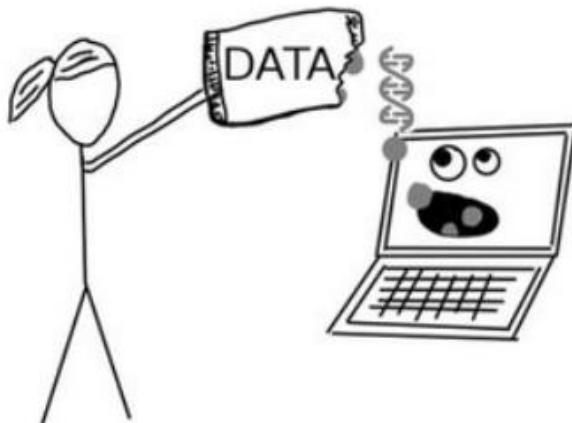


Variational Sparse Autoencoder





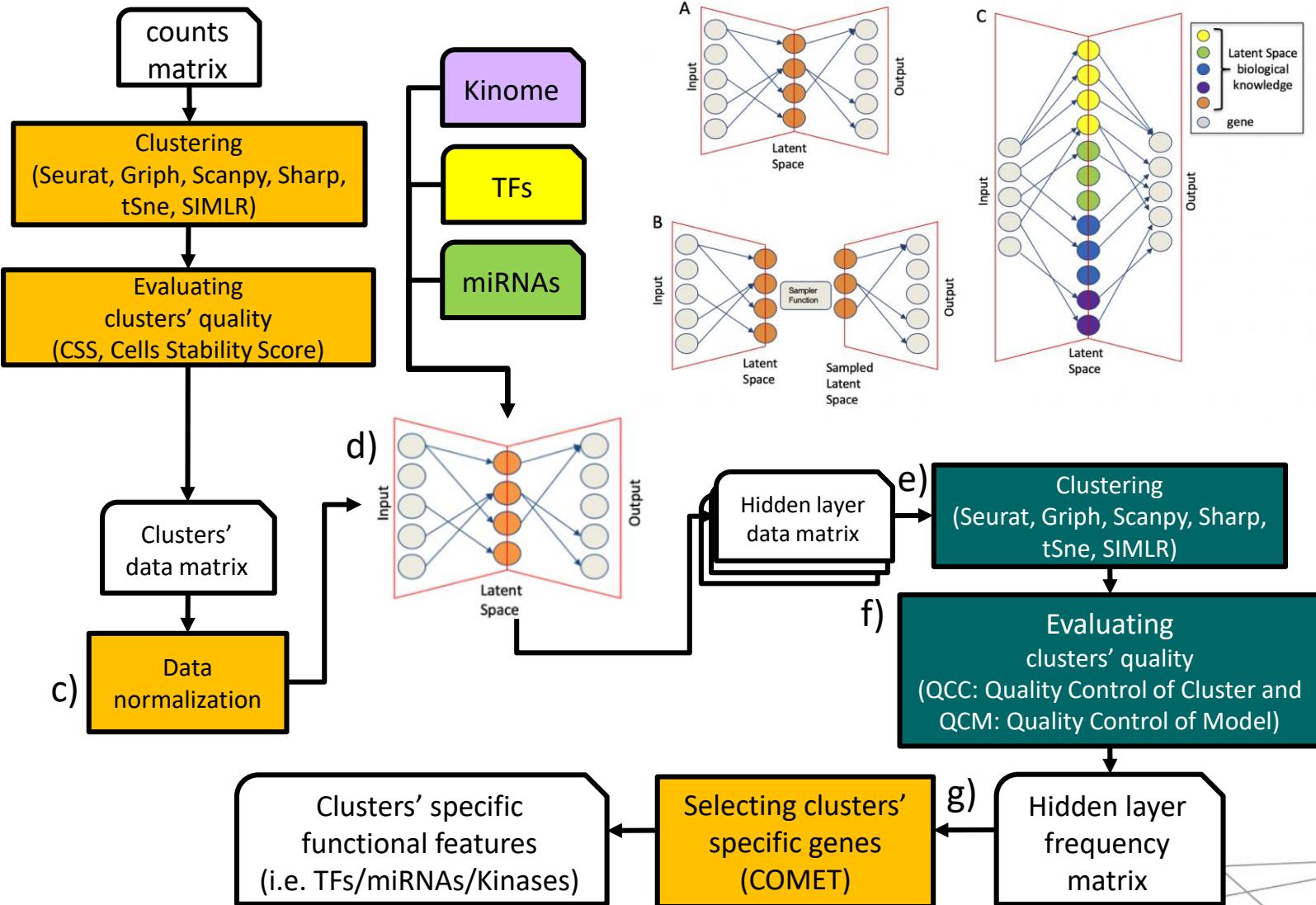
Sparse Autoencoder GPU vs CPU



GPU picture from <https://www.hpcwire.com/2018/03/27/nvidia-riding-high-as-gpu-workloads-and-capabilities-soar/>
CPU picture https://it.123rf.com/photo_27838598_isolato-carta-tagliata-di-chip-della-cpu-sul-circuito-blu-è-la-tecnologia-di-processore-centrale-del-c.html

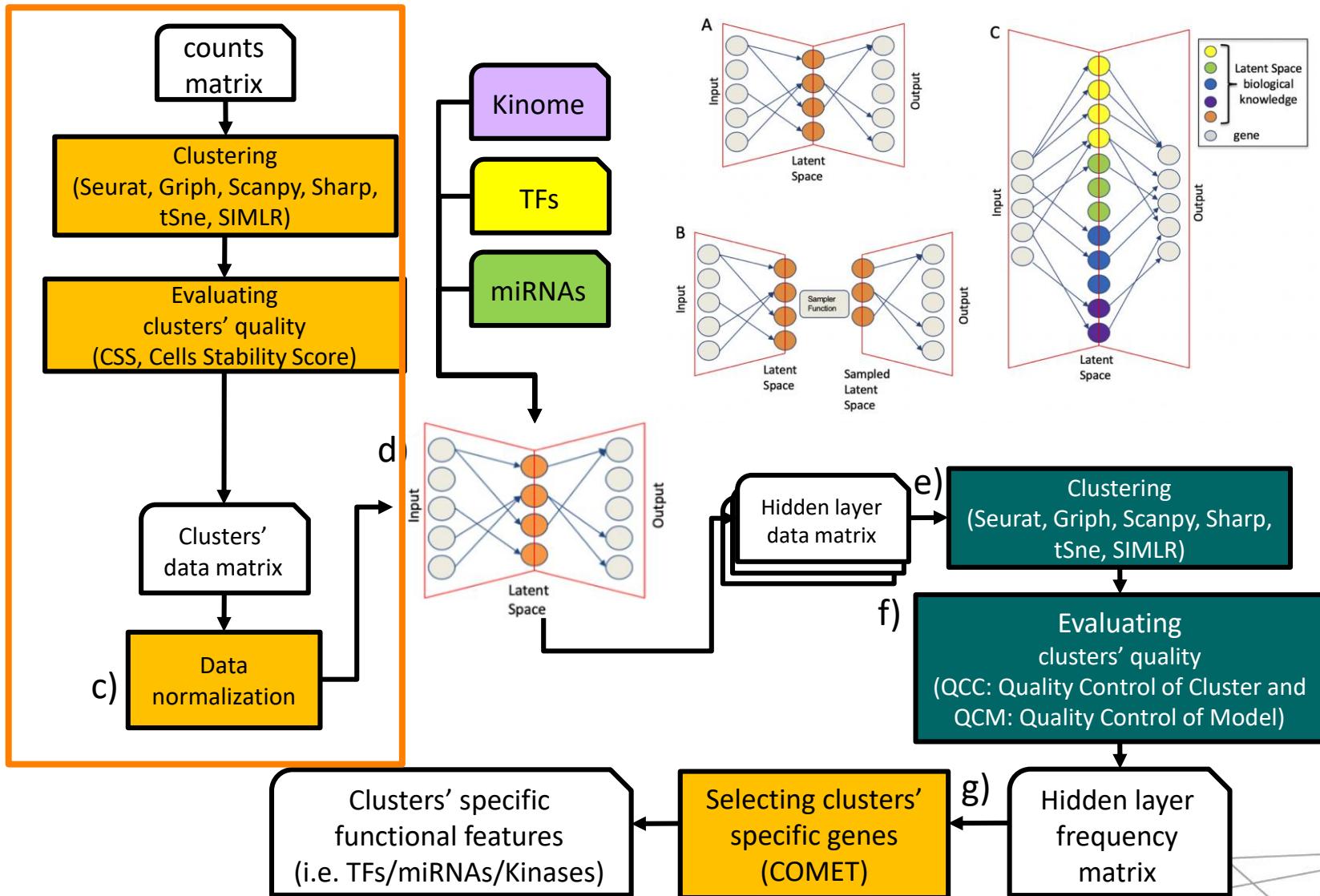


GIORNATA DI INCONTRO BORSE DI STUDIO GARR “ORIO CARLINI” BORSISTI DAY 2020



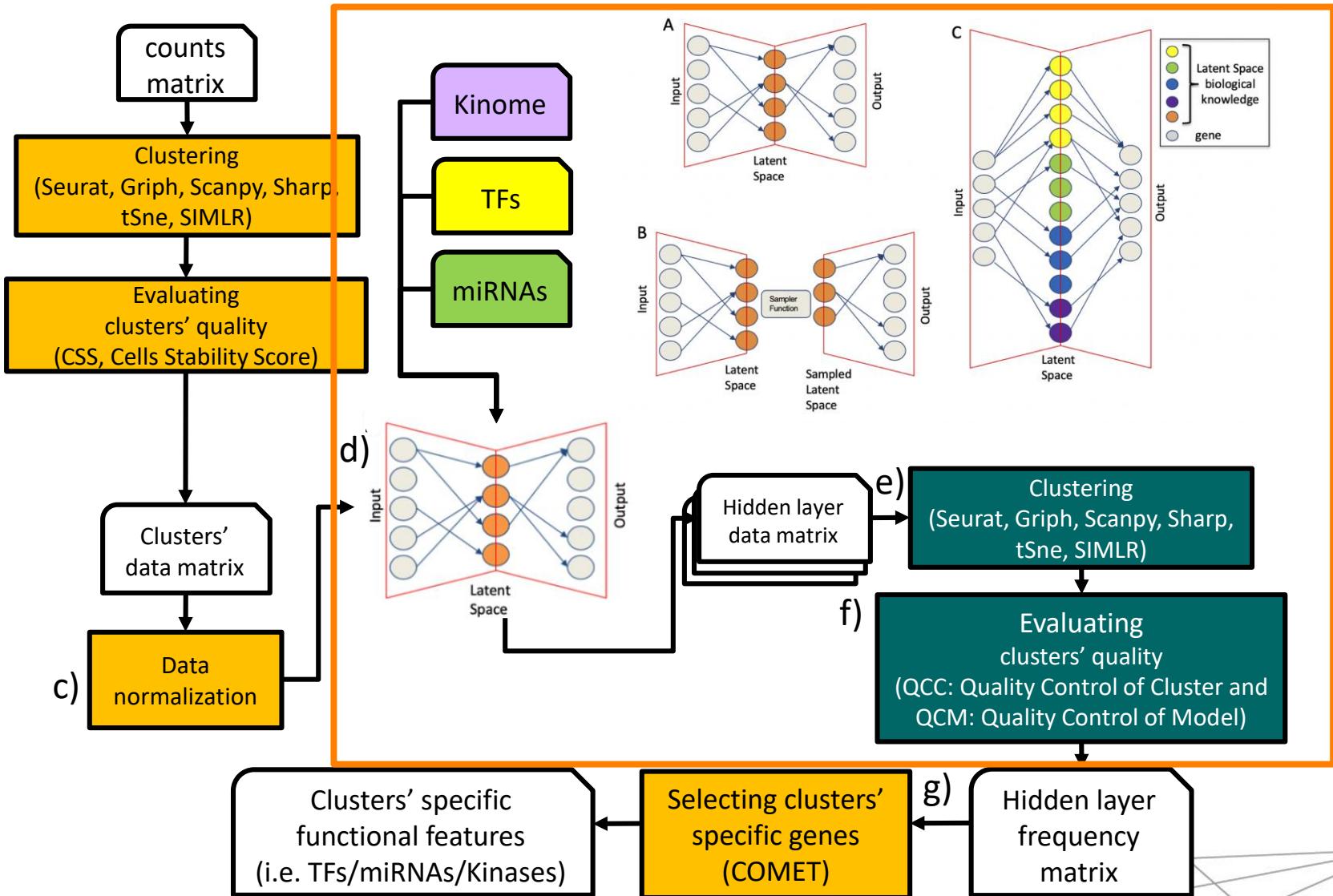


GIORNATA DI INCONTRO BORSE DI STUDIO GARR “ORIO CARLINI” BORSISTI DAY 2020



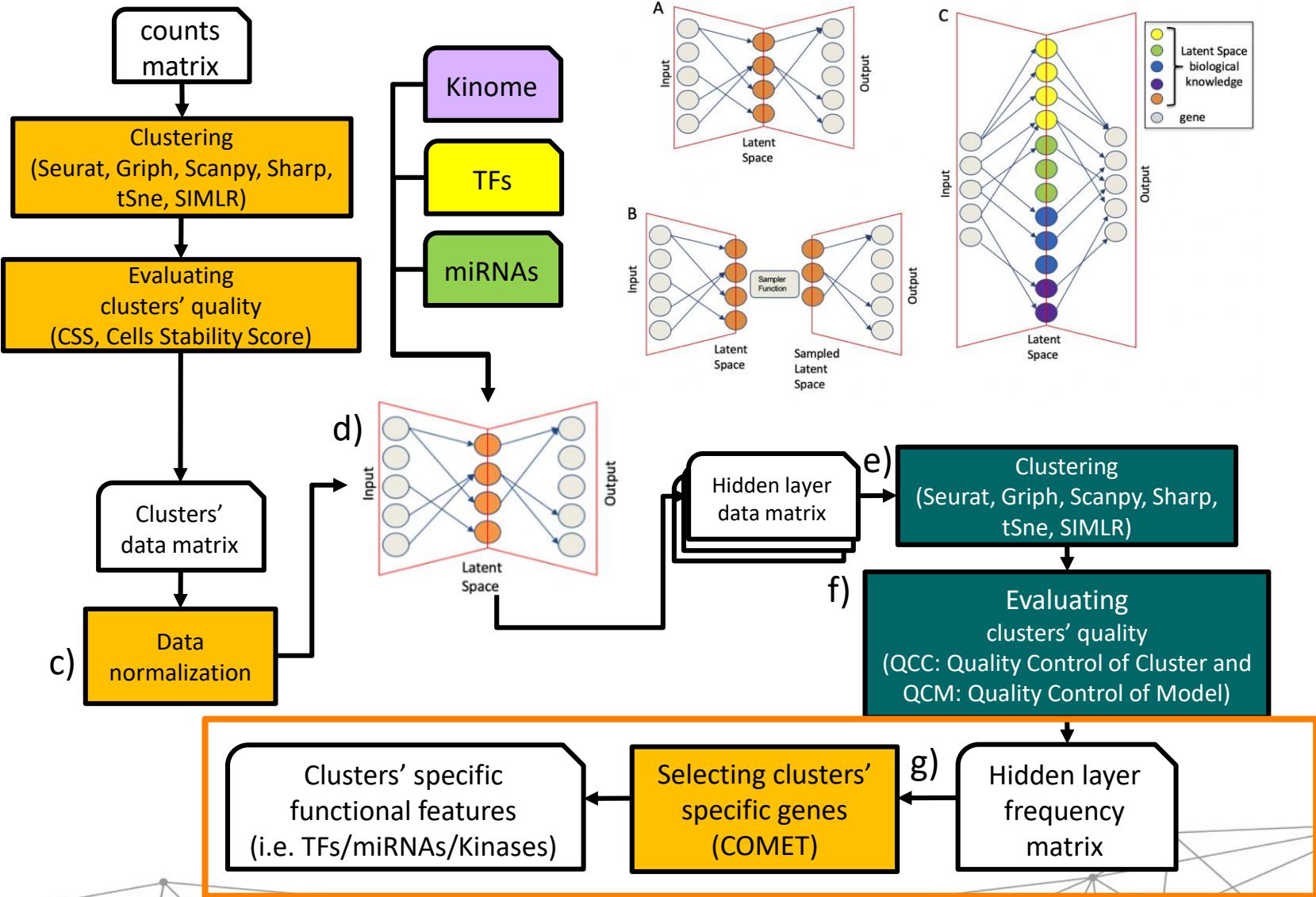


GIORNATA DI INCONTRO BORSE DI STUDIO GARR “ORIO CARLINI” BORSISTI DAY 2020



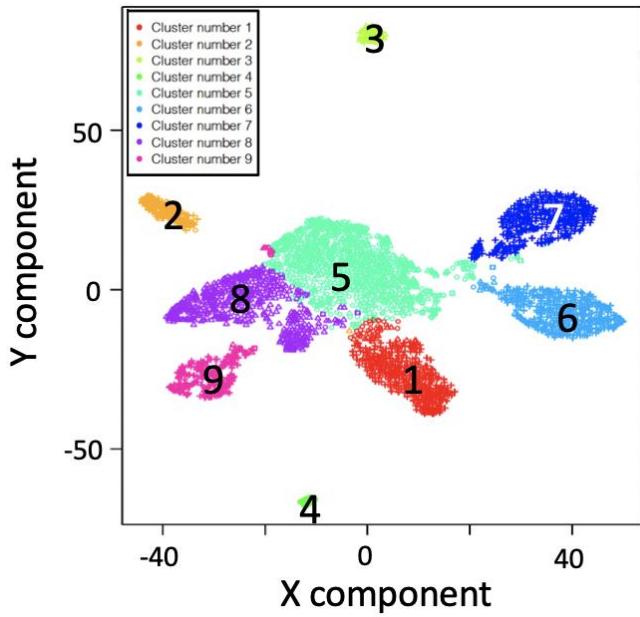


GIORNATA DI INCONTRO BORSE DI STUDIO GARR “ORIO CARLINI” BORSISTI DAY 2020



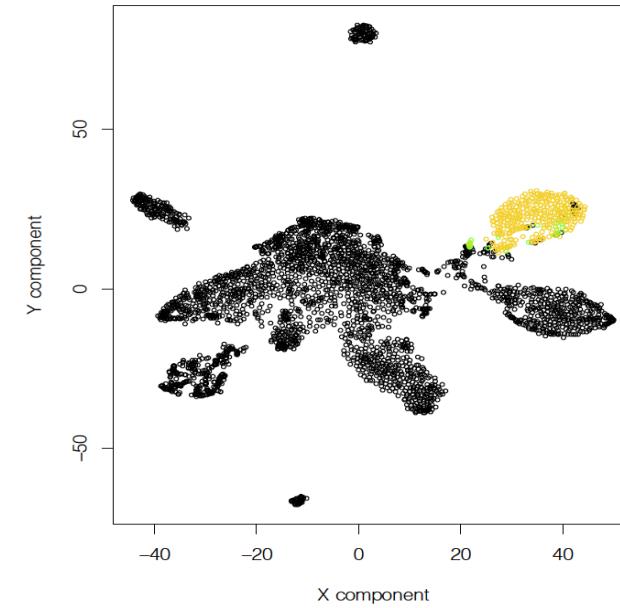
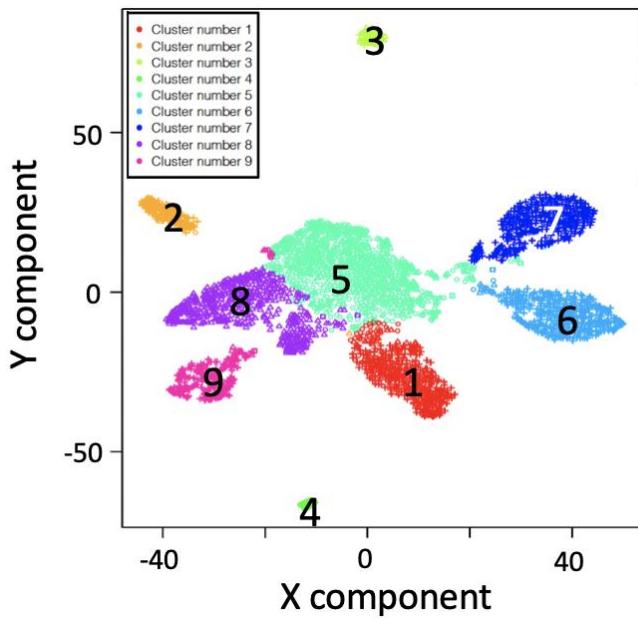


Autoencoder Data mining Statistics



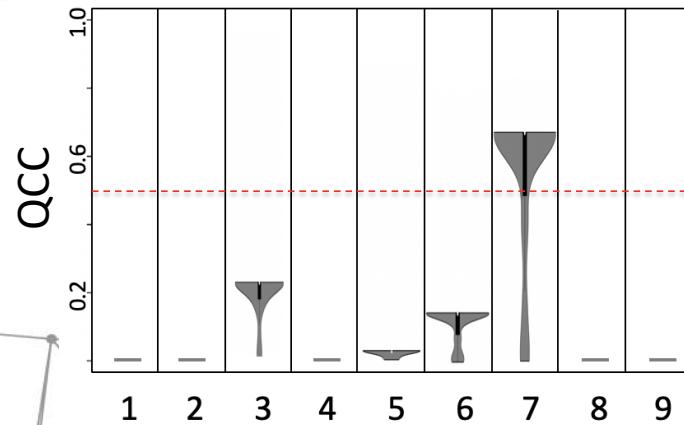
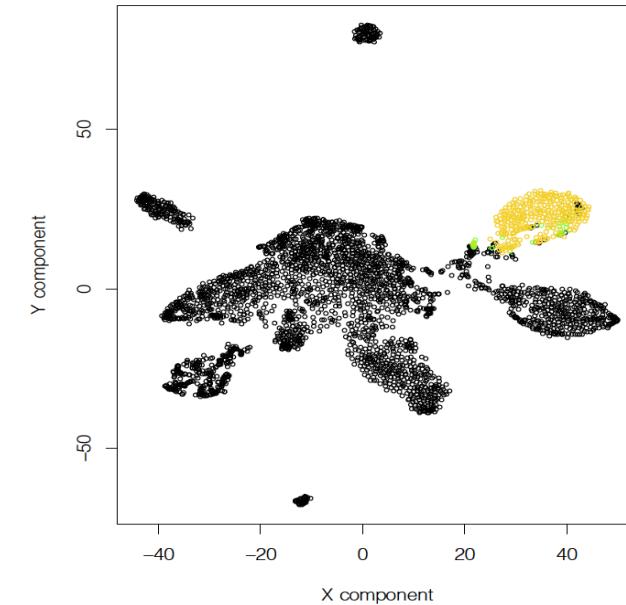
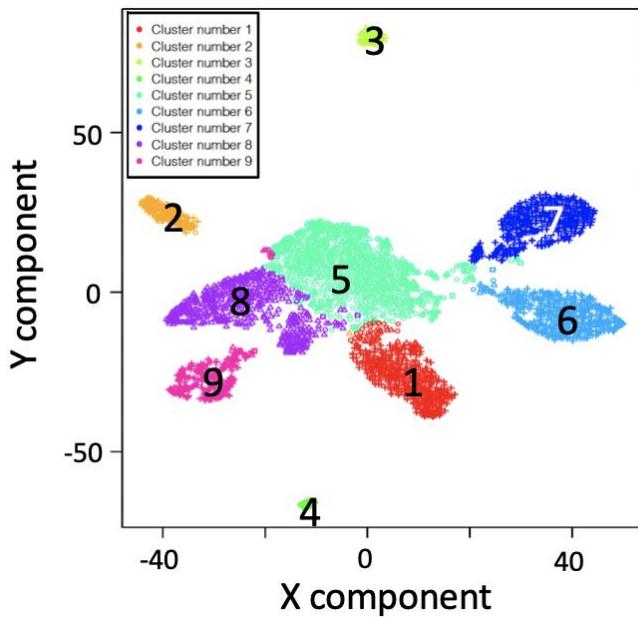


Autoencoder Data mining Statistics QCC



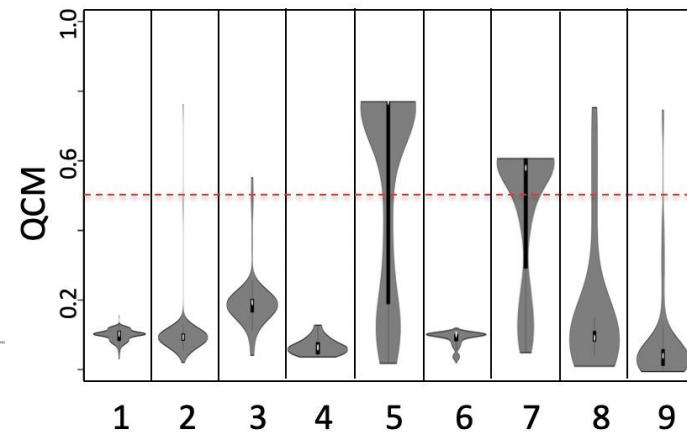
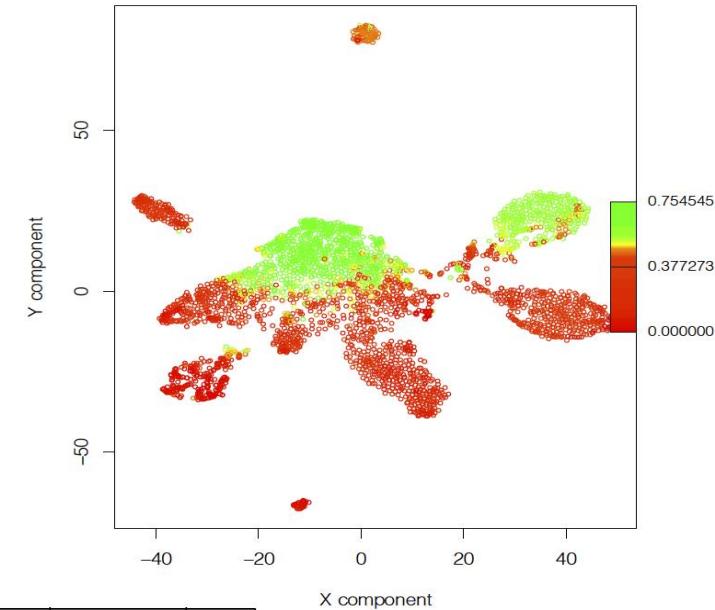
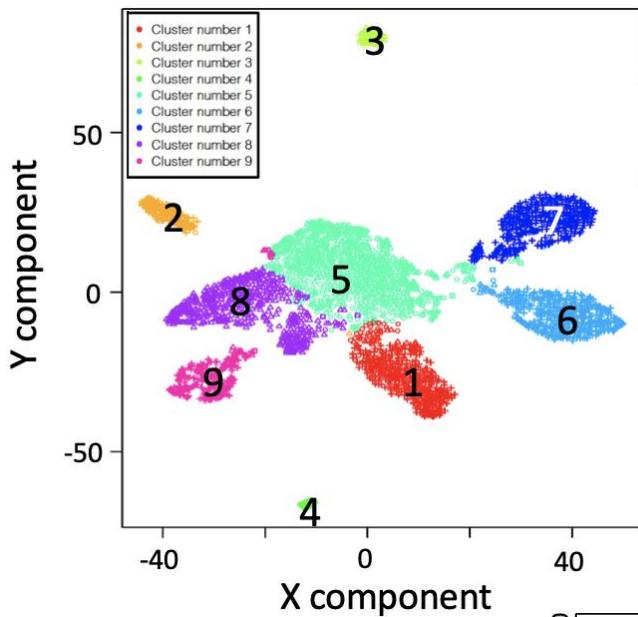


Autoencoder Data mining Statistics QCC



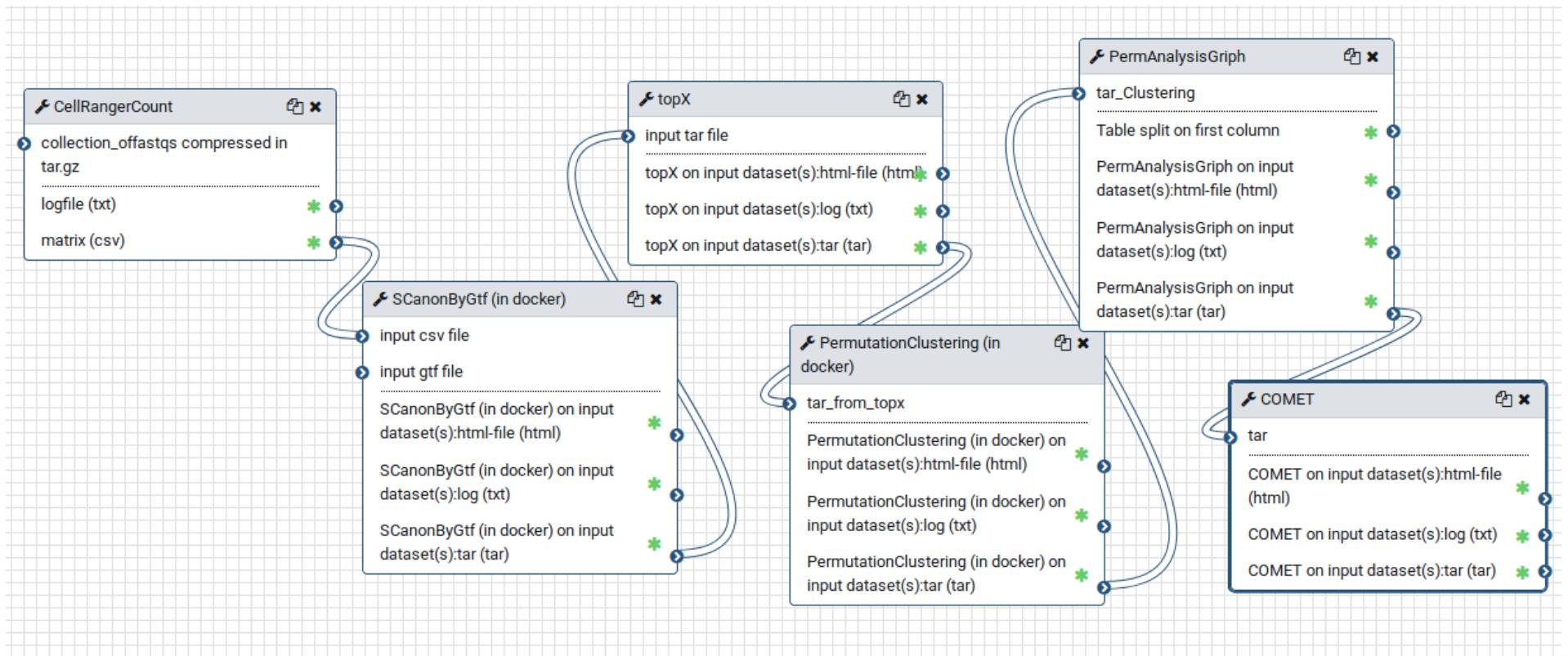


Autoencoder Data mining Statistics QCM





rCASC Workflow in Galaxy





Results in Html

Output COMET

[matrix.csv](#)
[filtered.annotated_matrix.csv](#)
[annotated_matrix.csv](#)

[scratch type: DIRECTORY](#)
[filtered_expression_annotated_matrix.csv](#)

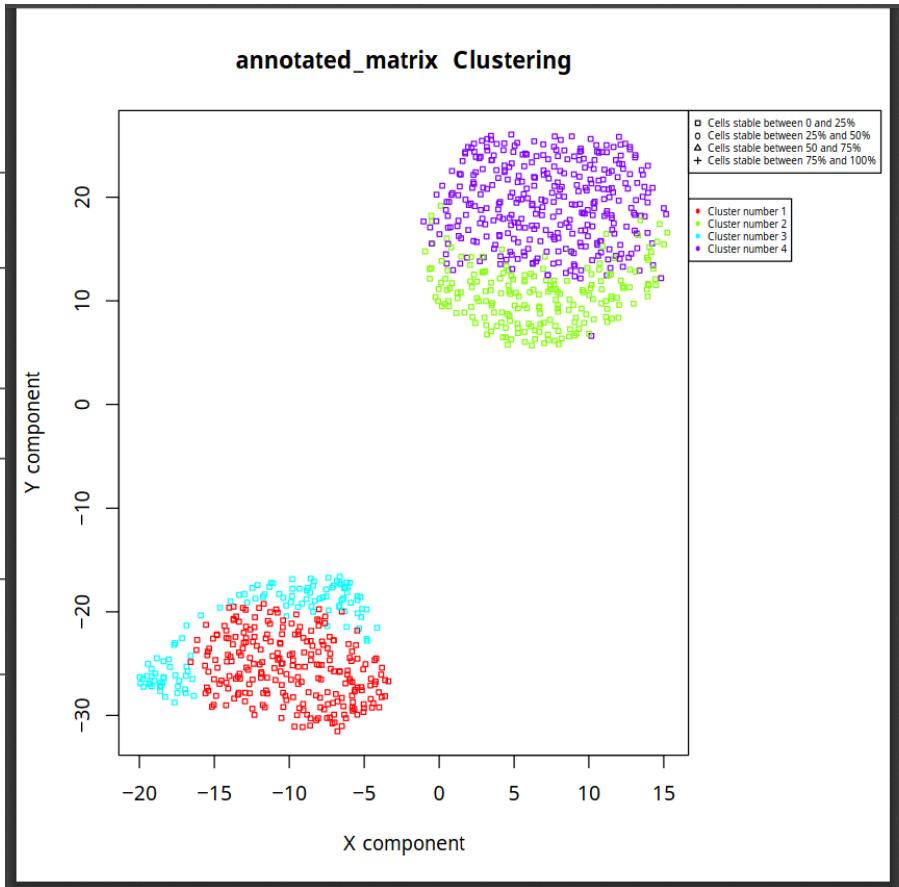
[annotated_matrix type: DIRECTORY](#)
[annotated_matrix_violplot.pdf](#)
[annotated_matrix.csv](#)

[4 type: DIRECTORY](#)

[outputdata type: DIRECTORY](#)
[hist.pdf](#)
[annotated_matrix.csv](#)

[4 type: DIRECTORY](#)
[log.txt](#)

[outputvis type: DIRECTORY](#)
[annotated_matrix_4_clusterP.csv](#)
[annotated_matrix_clustering.output.csv](#)
[annotated_matrix_Stability_Plot.pdf](#)
[cluster.txt](#)
[annotated_matrix_4_killedCell.csv](#)
[vis.txt](#)
[outputmetadata.txt](#)



search datasets

TEST wf

31 shown, 131 deleted
24.94 GB

162: COMET on data 159: tar

161: COMET on data 159: log

160: COMET on data 15 9:html-file
format: html, database: ?

/matrix.csv
/filtered.annotated_matrix.csv
/annotated_matrix.csv
/scratch/
/filtered_expression_annotated_matrix.csv
/annotated_matrix/
/annotated_matrix
/annotated_matrix_violplot.pdf
/annotated_matrix
/annotated_matrix.csv
/annotated_matrix/4/

HTML file

159: PermAnalysisGrph on data 156:tar

158: PermAnalysisGrph on data 156:log

157: PermAnalysisGrph



Proposta estensione borsa : Integrazione di dati di trascrizione e dati genomici

- Integrazione di **rCASC** in **Galaxy** 
- Sviluppo del tool **SCA** (Sparse Connected Autoencoder), paper accettato dalla rivista *NPJ Systems Biology and Applications* 



Proposta estensione borsa : Integrazione di dati di trascrizione e dati genomici

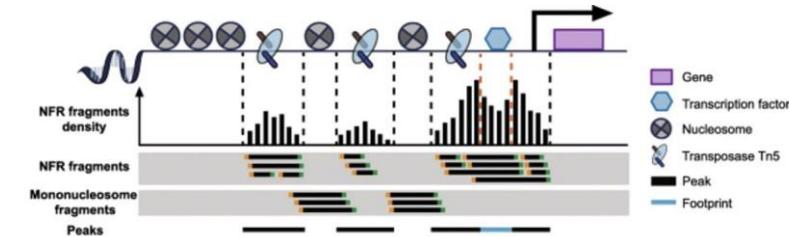
- Integrazione di **rCASC** in **Galaxy**
- Sviluppo del tool **SCA** (Sparse Connected Autoencoder), paper accettato dalla rivista *NPJ Systems Biology and Applications*
- Implementazione **ATAC-seq** Workflow





Proposta estensione borsa : Integrazione di dati di trascrizione e dati genomici

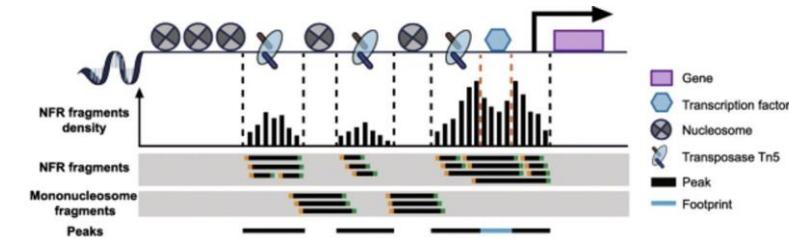
- Integrazione di **rCASC** in **Galaxy**
- Sviluppo del tool **SCA** (Sparse Connected Autoencoder), paper accettato dalla rivista *NPJ Systems Biology and Applications*
- Implementazione **ATAC-seq** Workflow





Proposta estensione borsa : Integrazione di dati di trascrizione e dati genomici

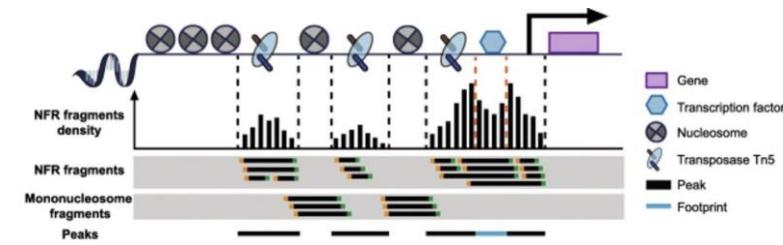
- Integrazione di **rCASC** in **Galaxy**
- Sviluppo del tool **SCA** (Sparse Connected Autoencoder), paper accettato dalla rivista *NPJ Systems Biology and Applications*
- Implementazione **ATAC-seq Workflow**
- Implementazione di data mining tools per l'identificazione di motivi di legame di fattori trascrizionali associati a picchi di ATAC-seq
 - **Meme suite**
 - **Shape motif**
 - **Gimmemotifs**





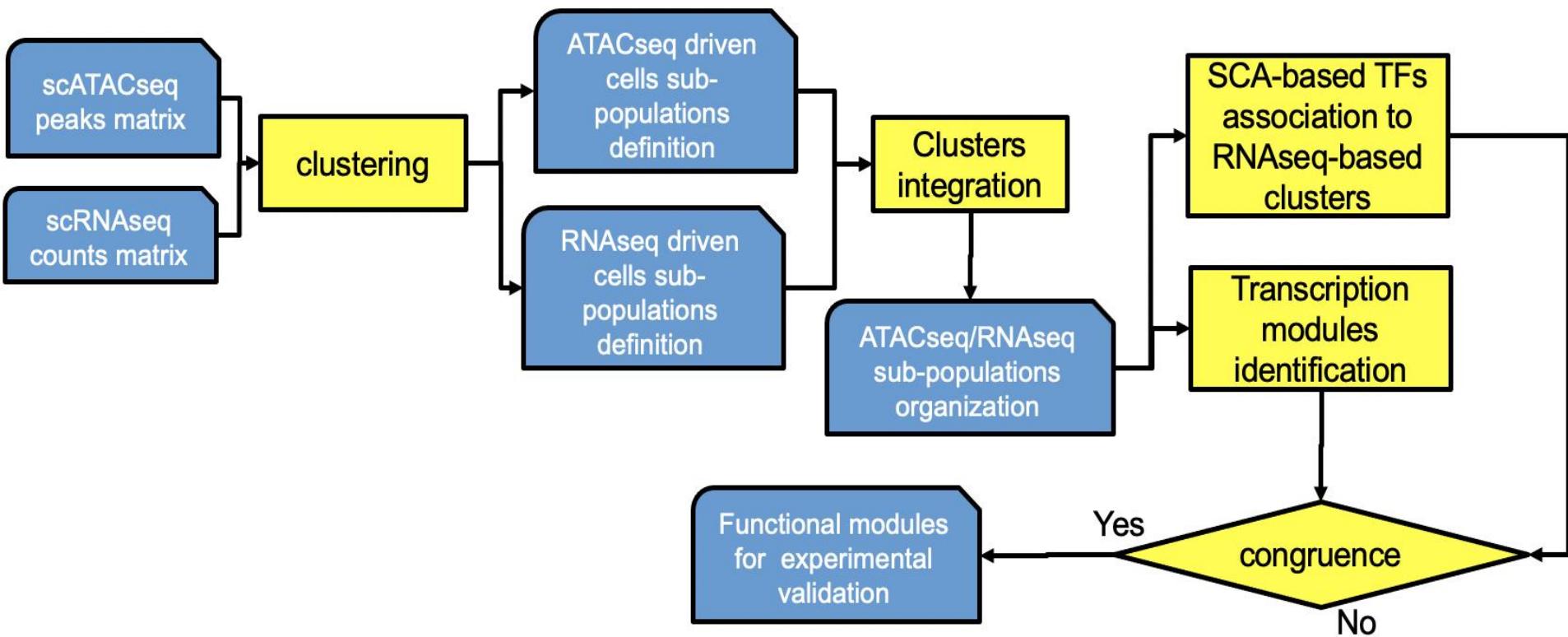
Proposta estensione borsa : Integrazione di dati di trascrizione e dati genomici

- Integrazione di **rCASC** in **Galaxy**
- Sviluppo del tool **SCA** (Sparse Connected Autoencoder), paper accettato dalla rivista *NPJ Systems Biology and Applications*
- Implementazione **ATAC-seq Workflow**
- Implementazione di data mining tools per l'identificazione di motivi di legame di fattori trascrizionali associati a picchi di ATAC-seq
 - **Meme suite**
 - **Shape motif**
 - **Gimmemotifs**
- **Integrazione** di dati ATAC-seq ed RNA-seq per l'identificazione dei moduli regolativi di controllo dell'espressione genica associati a sottopopolazioni cellulari





Proposta estensione borsa : Integrazione di dati di trascrizione e dati genomici



- **Integrazione** di dati ATAC-seq ed RNA-seq per l'identificazione dei moduli regolativi di controllo dell'espressione genica associati a sottopopolazioni cellulari



Grazie per l'attenzione