

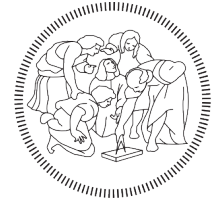


Michele Leone

PhD candidate, 3rd year.

Thesis submission: October 2020

Advisor: Marco Masseroli



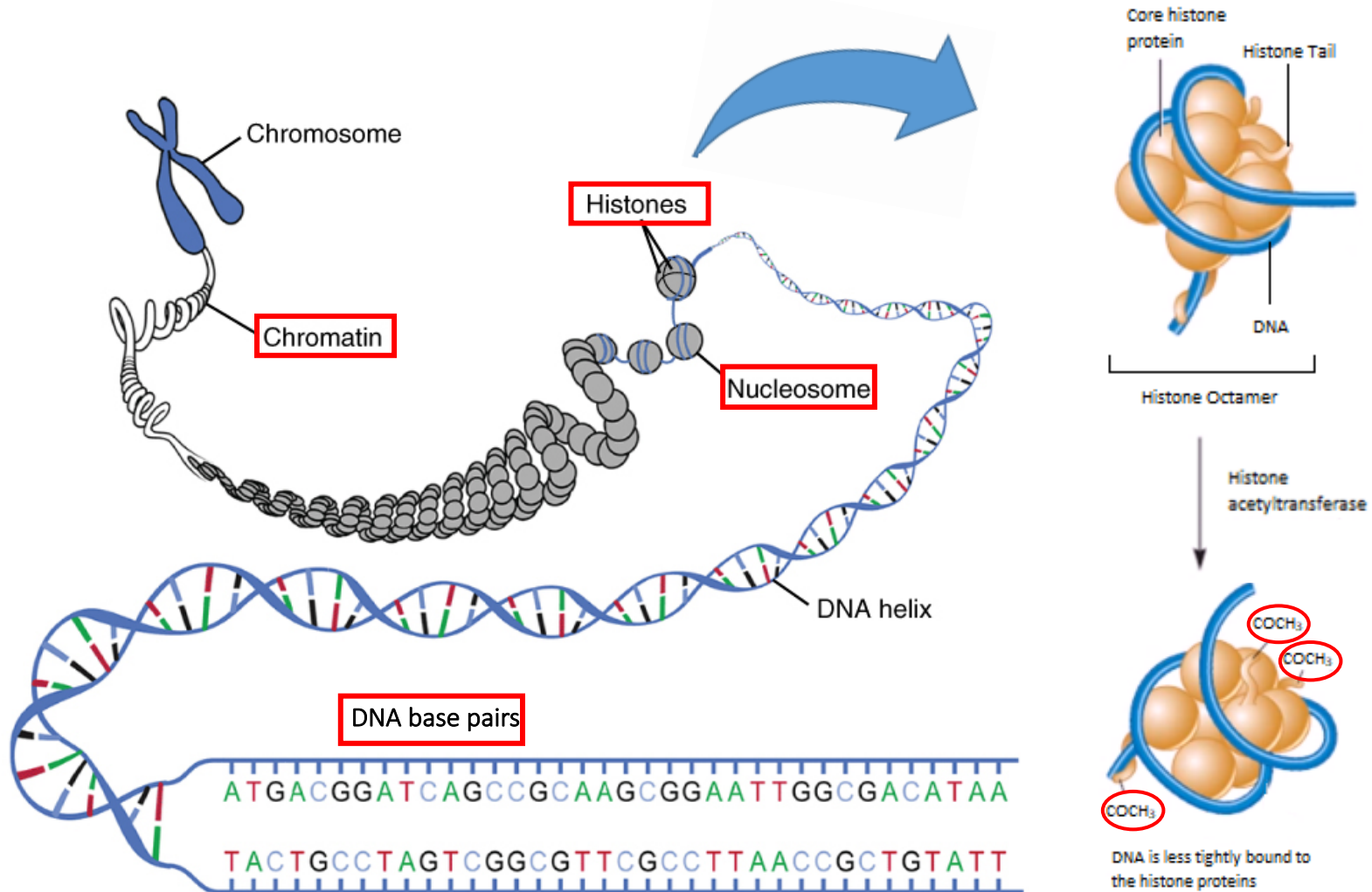
POLITECNICO
MILANO 1863

Identification, semantic annotation and comparison of chromatin regulatory states combinations in multiple biological conditions

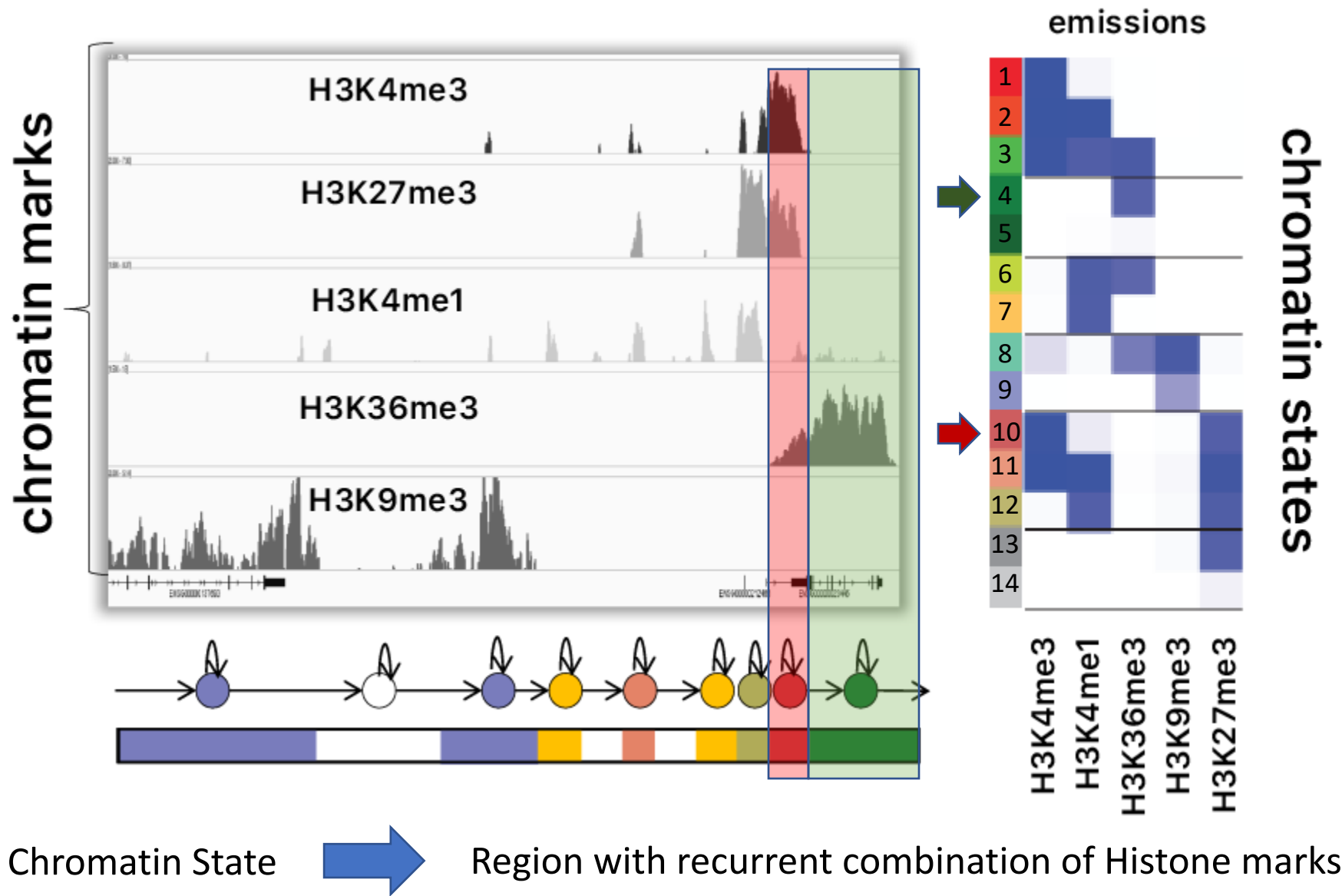
Relevant publications/submissions:

- **Leone M.**, Galeota E., Ceri S, Masseroli M., Pelizzola M. Identification, semantic annotation and comparison of chromatin regulatory states combinations in multiple biological conditions. Soon to be submitted to Bioinformatics journal.
- Cannizzaro G. & **Leone M.**, Bernasconi A., Canakoglu A, Carman M.J. Automated Integration of Genomics Metadata with Sequence-to-Sequence Models. Soon to be submitted to ECML-PKDD 2020 Conference.
- **Leone M.**, Galvani M., Masseroli M. De novo sequence-based method for ncRPI prediction using structural information. BIBE2019: 19th IEEE International Conference on Bioinformatics and Bioengineering; October 28-30, 2019; Athens, GR.
- Martano G., **Leone M.**, D'Oro P., Matafora V., Cattaneo A., Masseroli M., Bachi A. SMfinder: Small Molecules Finder for Metabolomics and Lipidomics analysis. Analytical Chemistry (submitted).

Biological Background: Chromatin Organization



Biological Background: Chromatin State

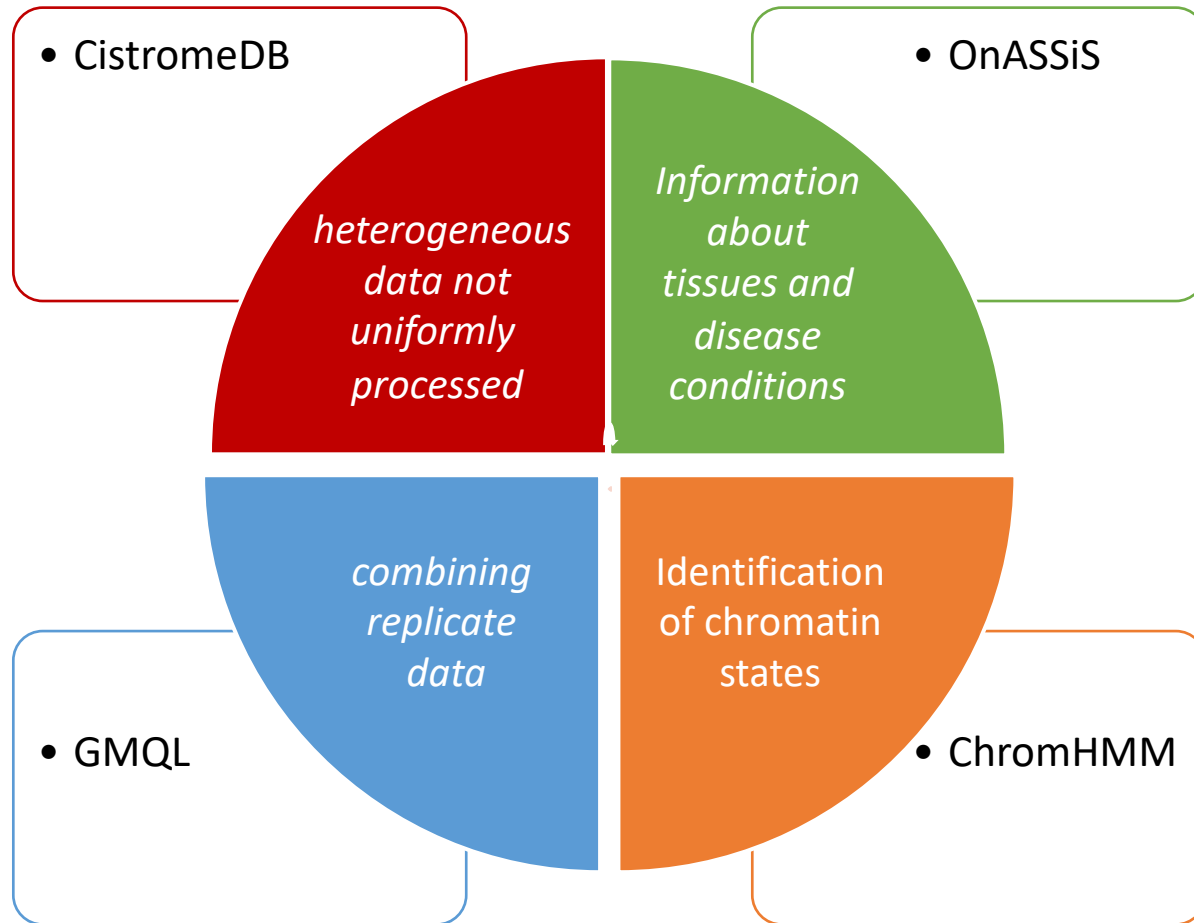


Project Goal

Create an efficient **computational method** able to extend the concept of chromatin states, and create a framework that starting from a set of functional elements:

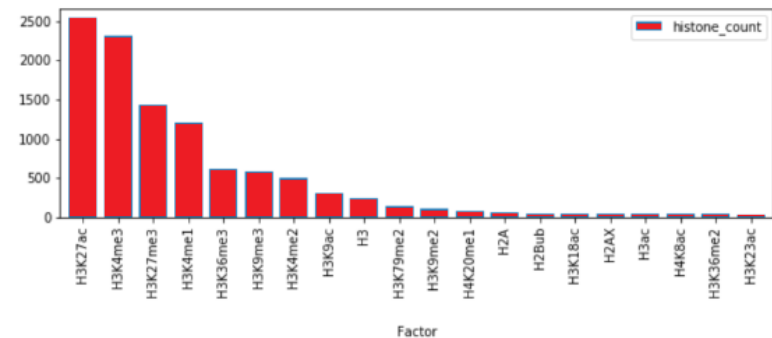
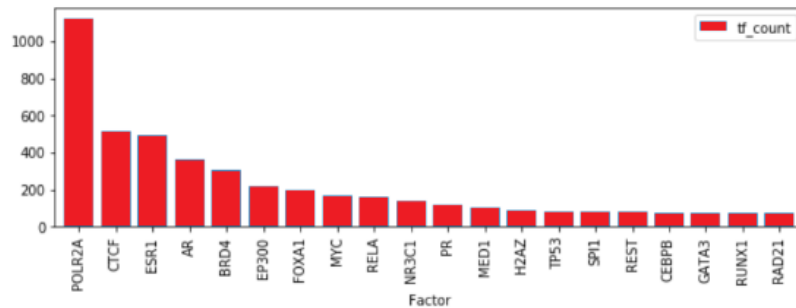
- Integrate and take advantage of the valuable and numerous, but **heterogeneous**, data publicly available in well-known big data repositories.
- Process them to make them **comparable** and **homogeneously characterized**
- Identify **altered chromatin states** in presence of various pathologies, compared to the state of each specific tissue in healthy patients.
- Generate combinations of **chromatin states** of many **cell types**, using not only **well-categorized data** specifically produced for a specific purpose through costly and **time-consuming experiments**.
- **Bi-clustering** of regions and samples, identification of **genome clusters** and **Gene-set enrichment analysis**

Project Pipeline



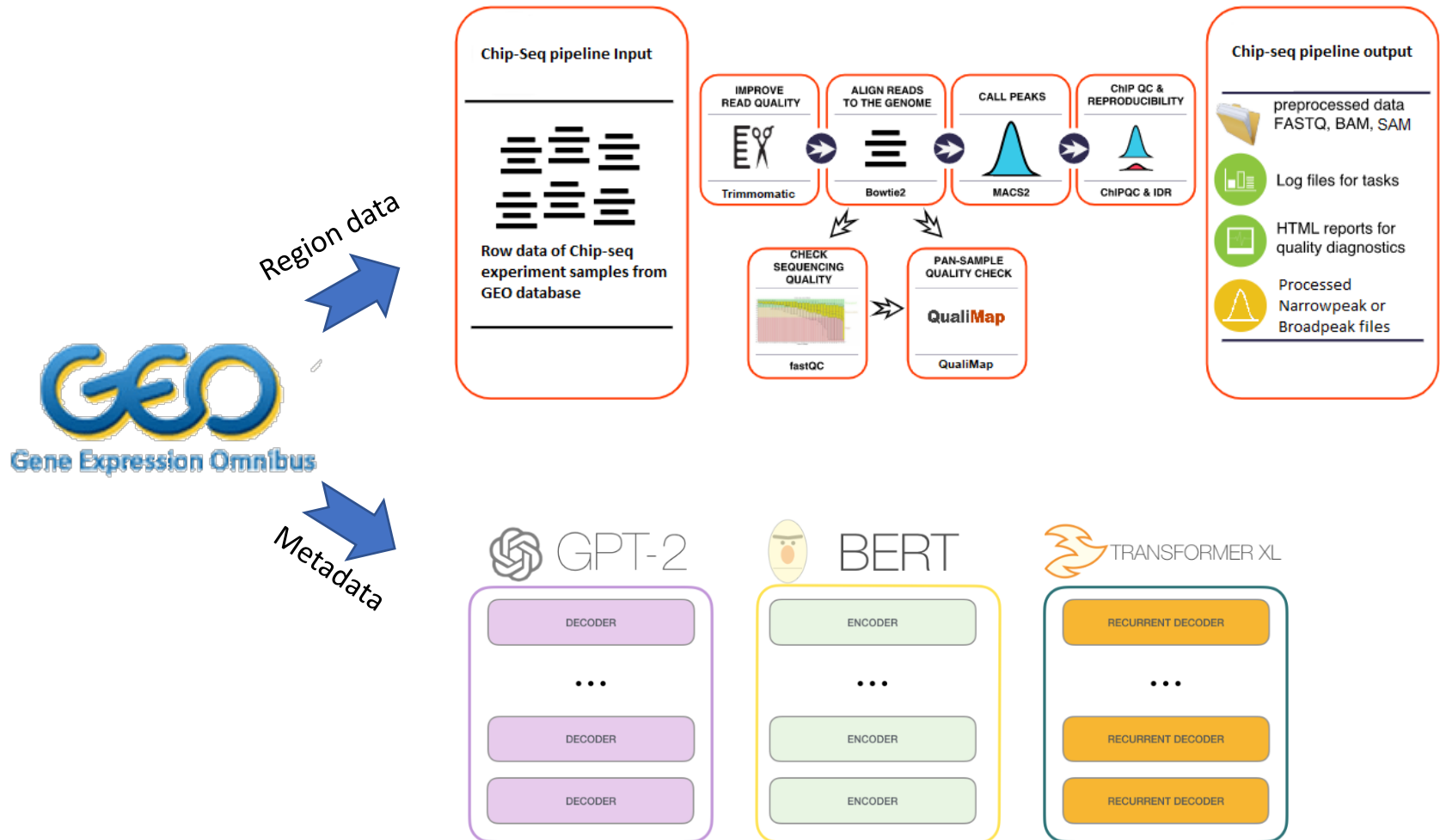
Overcome heterogeneity of processing data: Cistrome

Identify a set of genomic features that can accurately characterize chromatin states and cover a wide range of **tissues/conditions**



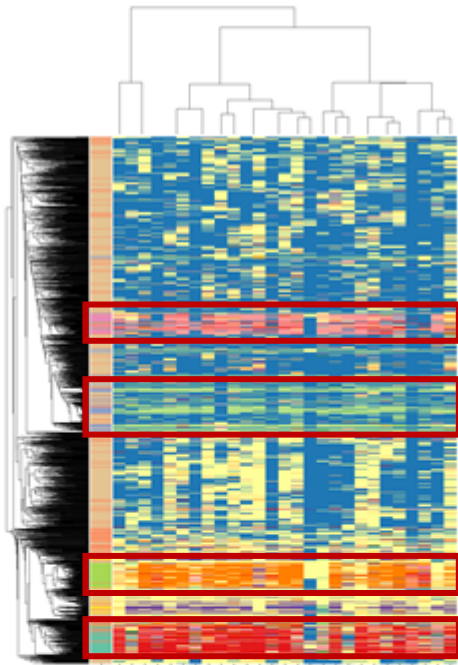
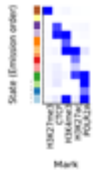
Factor	Type	Functional association
H3k4me3	Histone Mark	Active Promoters
H3K4me1	Histone Mark	Active Enhancers
H3K27ac	Histone Mark	Active Promoters and enhancers
H3K27me3	Histone Mark	Inactive chromatin
H3K36me3	Histone Mark	Active transcribed gene bodies
CTCF	Transcription Factor	Transcriptional regulation
POLR2A	Transcription Factor	Synthesize mRNA in eukaryotes
MYC	Transcription Factor	Activate expression of many pro-proliferative genes

Data Extraction and Processing



Data-driven analysis

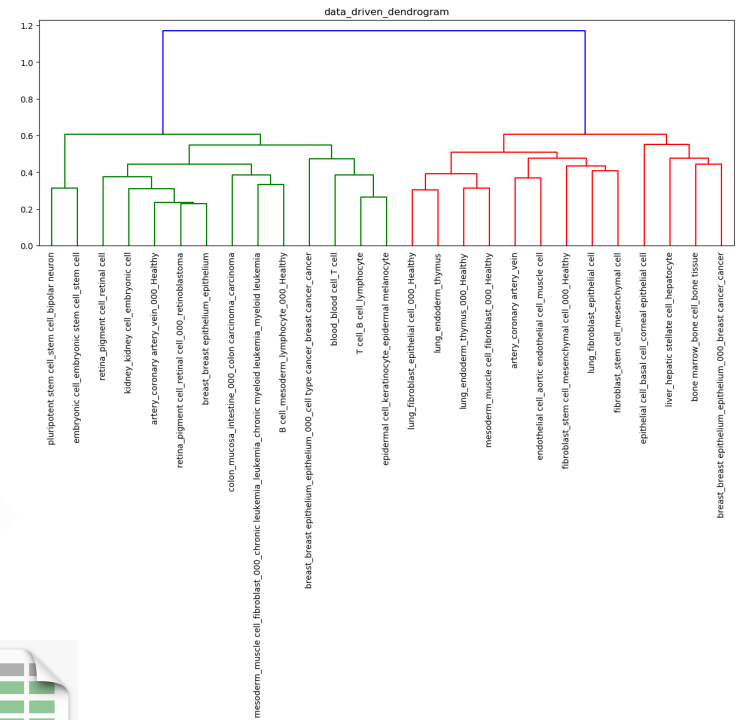
Emission Parameters



colon_mucosa_intestine_000_colon_carcinoma_carcinoma
 breast_breast_epithelium_000_cell_type_cancer_breast_cancer_cancer
 breast_breast_epithelium_000_breast_cancer_cancer
 breast_breast_epithelium_000_breast_cancer_cancer
 lung_endoderm_thymus_000_healthy
 artery_coronary_artery_000_healthy
 artery_coronary_artery_000_healthy
 bone_marrow_bone_cell_bone_tissue
 endothelial_cell_000_healthy
 lung_fibroblast_epithelial_cell_000_healthy
 lung_fibroblast_epithelial_cell_000_healthy
 mesoderm_muscle_cell_mesoderm_muscle_cell_000_healthy
 mesoderm_muscle_cell_mesoderm_muscle_cell_000_healthy
 kidney_kidney_cell_embryonic_cell
 T_cell_T_cell_lymphocyte
 B_cell_mesoderm_lymphocyte_000_healthy
 blood_blood_cell_T_cell
 blood_blood_cell_T_cell
 fibroblast_stem_cell_mesenchymal_cell
 fibroblast_stem_cell_mesenchymal_cell
 pluripotent_stem_cell_stem_cell_bipolar_neuron
 embryonic_cell_embryonic_stem_cell_stem_cell
 retina_pigment_cell_retinal_cell_000_mesoderm
 retina_pigment_cell_retinal_cell_000_mesoderm
 epithelial_cell_keratinocyte_epithelial_melanocyte
 epithelial_cell_basal_cell_cornal_epithelial_cell



Gene set Enrichment Analysis



Thank You for your attention