a
Data preprocessing and region mapping
NGS experiments generated ATAC-seq and ChIP-seq datasets
Next-generation sequencing
ATAC-seq
ChIP-seq
Accessible chromatin
Closed chromatin
Measured TR binding sites
Quality-control, Alignment,…
Peak calling
ATAC-seq Database
ChIP-seq Database
ATAC-seq database
Each ATAC-seq dataset profiles cancer-related chromatin accessibility
Binary vectors representing the chromatin accessibility
ChIP-seq database
Each ChIP-seq dataset profiles one TR's context-specific binding sites
Binary vectors representing the TR binding profiles
Matching each binarized ATAC-seq dataset versus all TR ChIP-seq datasets
Potential cases after matching datasets
Matching
Non-matching
Non-Informative (Dropped)
b
TRex hierarchical model
Observed
Parameter
Auxiliary
J
M
M^c
K_ij or K_i'j
c
Mean-field variational inference
Coordinate ascent mean-field variational inference
Mean-field variational family
Update the expectation of factors
Use coordinate ascent to find optimal:
Similarly for all factors
Minimize KL divergence:
Variational distribution
TR
Normal CDF
Score
TR profiles
ESR1
FOXA2
AR
PGR
0.067
0.023
0.011
0.003
for all collected ATAC-seq datasets
d
Integration results for TRex online portal
Integration of TR ranking results from same condition for downstream analysis
TR profiles
Aggregated based on different tissue of origin / subtypes / cell lines
for all collected cancer ATAC-seq datasets
Group 1
Group 2
Group …
TRex online portal for access to pan-cancer TR profiles and online analysis
TRex
BED
narrowPeak
broadPeak
SRX id
Cancer type
Cell line
User provided query region sets
Pan-cancer ATAC-seq data
TRex online computation
Output Retrieving
Visulization
TRex Online Portal
Bone
Bowel
Breast
CNS/Brain
Cervix
Esophagus /Stomach
Liver
Lung
Lymphoid
Myeloid
Pancreas
Peripheral Nervous System
Pleura
Prostate
Skin
Soft Tissue
Testis
Thyroid
Uterus