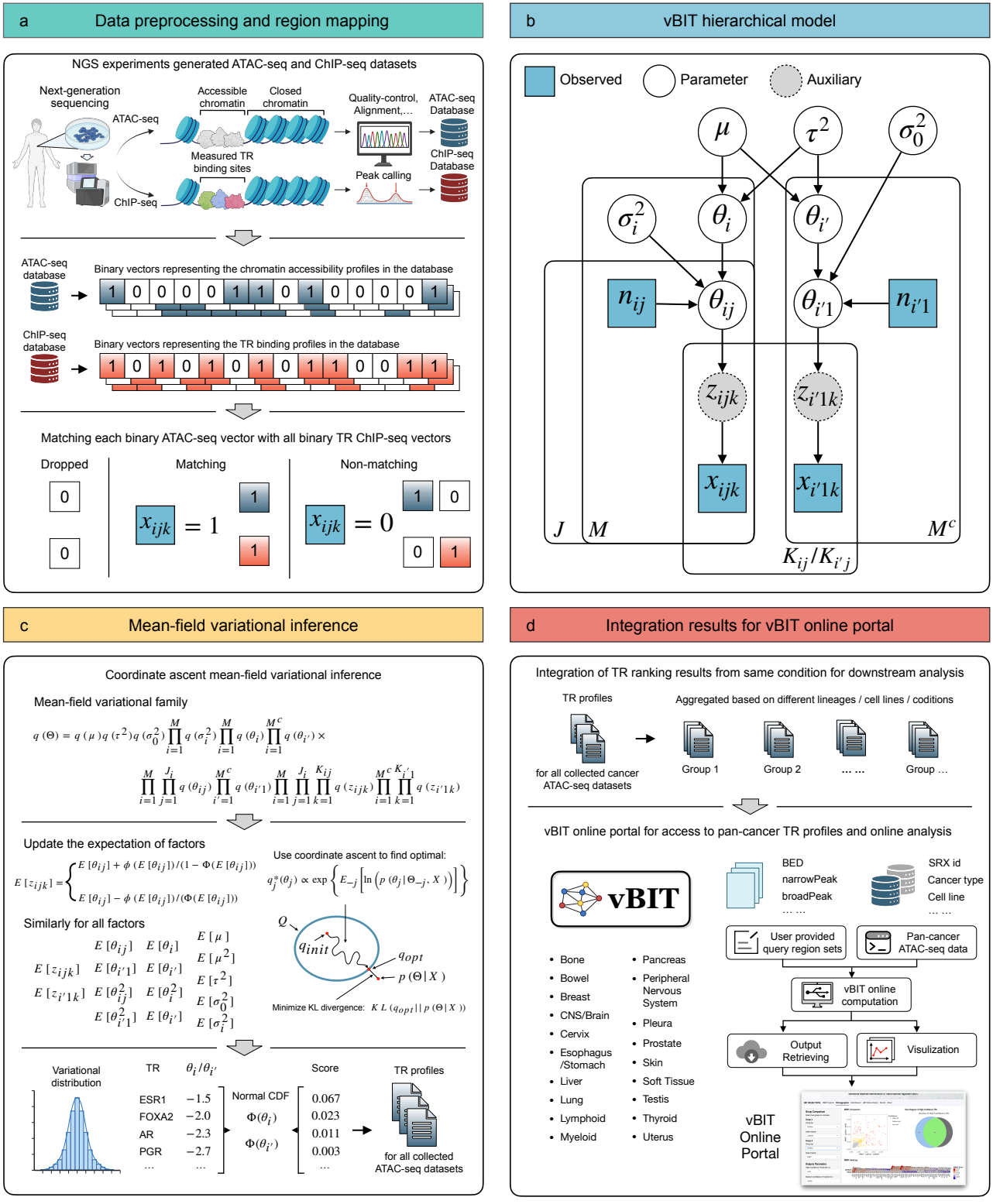
a
Data preprocessing and region mapping
NGS experiments generated ATAC-seq and ChIP-seq datasets
Next-generation sequencing
ATAC-seq
ChIP-seq
Accessible chromatin
Closed chromatin
Measured TR binding sites
Quality-control, Alignment,…
Peak calling
ATAC-seq Database
ChIP-seq Database
ATAC-seq database
Binary vectors representing the chromatin accessibility profiles in the database
ChIP-seq database
Binary vectors representing the TR binding profiles in the database
Matching each binary ATAC-seq vector with all binary TR ChIP-seq vectors
Dropped
Matching
Non-matching
b
vBIT hierarchical model
Observed
Parameter
Auxiliary
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
for all collected ATAC-seq datasets
d
Integration results for vBIT online portal
Integration of TR ranking results from same condition for downstream analysis
TR profiles
Aggregated based on different lineages / cell lines / coditions
for all collected cancer ATAC-seq datasets
Group 1
Group 2
Group …
vBIT online portal for access to pan-cancer TR profiles and online analysis
vBIT
BED
narrowPeak
broadPeak
SRX id
Cancer type
Cell line
User provided query region sets
Pan-cancer ATAC-seq data
vBIT online computation
Output Retrieving
Visulization
Bone
Bowel
Breast
CNS/Brain
Cervix
Esophagus /Stomach
Liver
Lung
Lymphoid
Myeloid
Pancreas
Peripheral Nervous System
Pleura
Prostate
Skin
Soft Tissue
Testis
Thyroid
Uterus
vBIT Online Portal