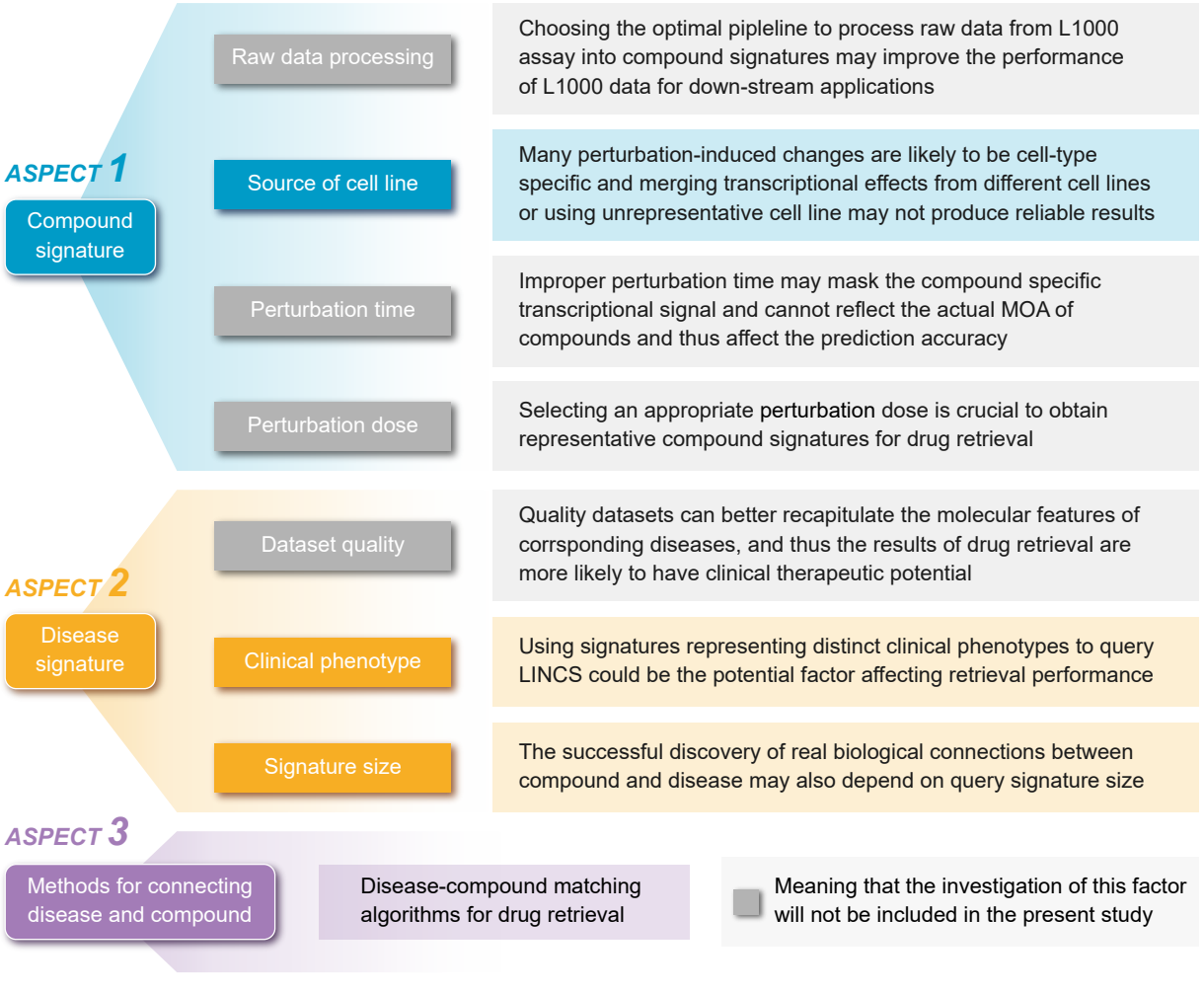
ASPECT 1
Compound signature
Raw data processing
Choosing the optimal pipleline to process raw data from L1000 assay into compound signatures may improve the performance of L1000 data for down-stream applications
Source of cell line
Many perturbation-induced changes are likely to be cell-type specific and merging transcriptional effects from different cell lines or using unrepresentative cell line may not produce reliable results
Perturbation time
Improper perturbation time may mask the compound specific transcriptional signal and cannot reflect the actual MOA of compounds and thus affect the prediction accuracy
Perturbation dose
Selecting an appropriate perturbation dose is crucial to obtain representative compound signatures for drug retrieval
ASPECT 2
Disease signature
Dataset quality
Quality datasets can better recapitulate the molecular features of corrsponding diseases, and thus the results of drug retrieval are more likely to have clinical therapeutic potential
Clinical phenotype
Using signatures representing distinct clinical phenotypes to query LINCS could be the potential factor affecting retrieval performance
Signature size
The successful discovery of real biological connections between compound and disease may also depend on query signature size
ASPECT 3
Methods for connecting disease and compound
Disease-compound matching algorithms for drug retrieval
Meaning that the investigation of this factor will not be included in the present study