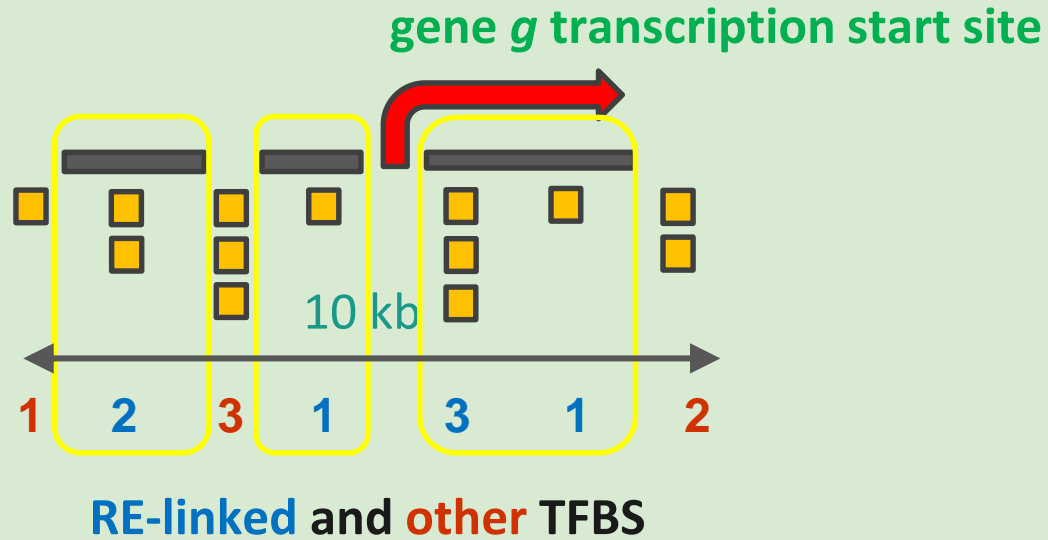


- **Quantization of RE-linked TFBS
impacts on genes and
molecular pathways**

—

GRE and NGRE scores (gene level of data analysis)



$$TES_g = \sum \text{TFBS in REs in gene } g = 7$$

$$TTS_g = \sum \text{TFBS} + \text{TFBS in gene } g = 7 + 6$$

Gene RE-linked TFBS Enrichment (GRE)

$$GRE_g = \frac{TES_g}{TES_m} \quad \begin{array}{l} \text{(TES for gene } g) \\ \text{(mean TES for all genes)} \end{array}$$

$$GTE_g = \frac{TTS_g}{TTS_m} \quad \begin{array}{l} \text{(TTS for gene } g) \\ \text{(mean TTS for all genes)} \end{array}$$

Normalized GRE (NGRE)

$$NGRE_g = \frac{GRE_g}{GTE_g} \quad \begin{array}{l} \text{(GRE for gene } g) \\ \text{(GTE for gene } g) \end{array}$$

High GRE = High *number* of RE-linked TFBS in a gene

High $NGRE$ = High *proportion* of RE-linked TFBS in a gene

PII and NPII scores (pathway level of data analysis)

Pathway Involvement Index (PII)

Normalized Pathway Involvement Index (NPII)

$$PII_p = \frac{\sum_{i=1}^n GRE_i}{n}$$

(sum of **GRE** for all pathway genes)
(number of genes in a pathway)

$$PGI_p = \frac{\sum_{i=1}^n GTE_i}{n}$$

(sum of **GTE** for all pathway genes)
(number of genes in a pathway)

Normalized PII (NPII)

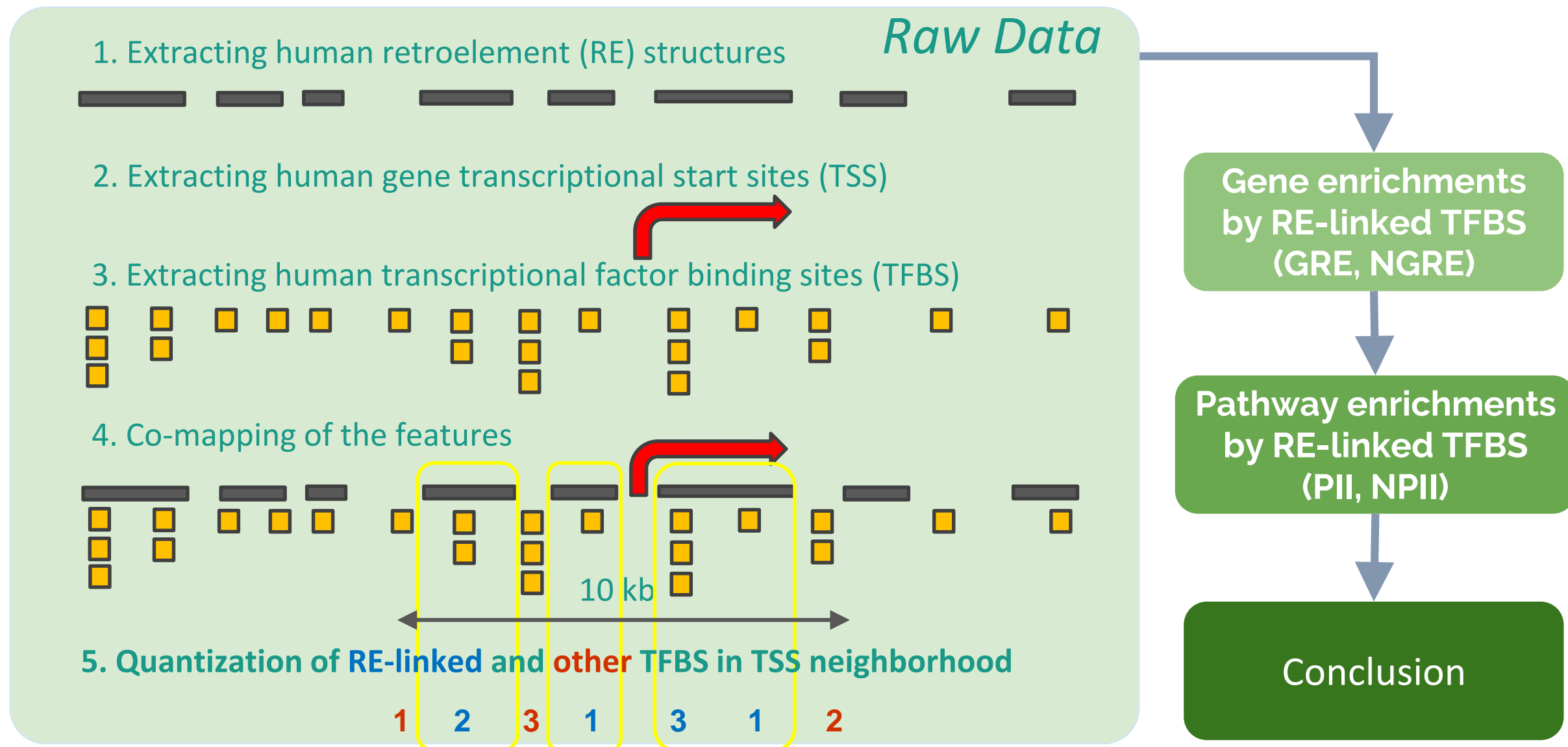
$$NPII_p = \frac{PII_p}{PGI_p}$$

(PII for pathway **p**)
(PGI for pathway **p**)

High **PII** = High *number* of RE-linked TFBS for pathway components

High **NPII** = High *proportion* of RE-linked TFBS for pathway components

Data analysis pipeline



Data analysis platform

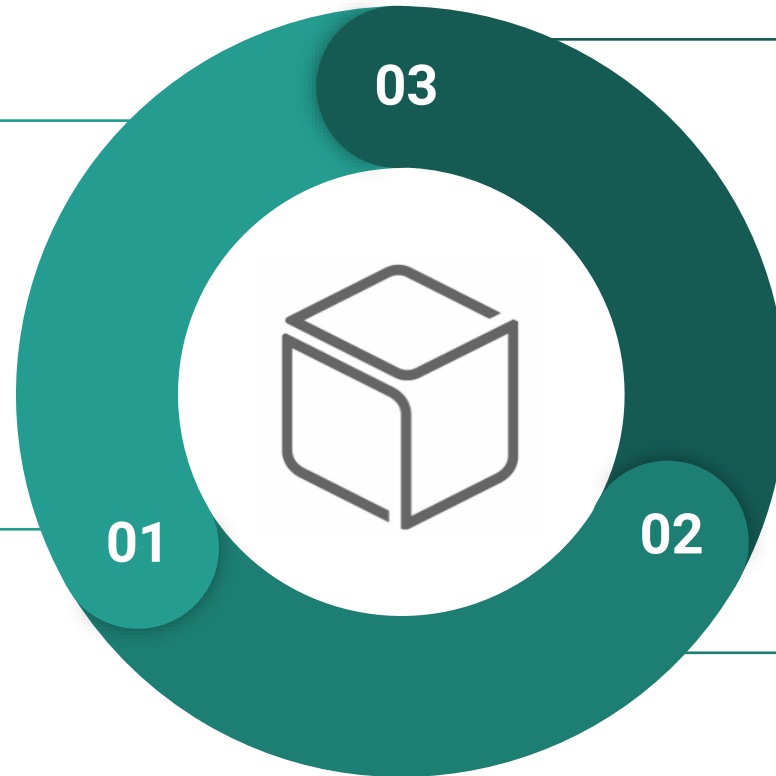
Analytic Algorithm

Structure Databank

- ✓ 3,124 molecular pathways for >8,000 connected human genes
- ✓ >24,000 human gene structures

Experimental Database

- ✓ ENCODE project experimental data
- ✓ Transcription Factor Binding Site (TFBS) profiles for 563 transcription factor proteins
- ✓ Totally ~21 million human TFBS



✓ Calculation of GRE scores for gene transcriptional start neighborhood enrichments by RE-linked TFBS

✓ Calculation of NGRE scores for *normalized* gene enrichments

✓ Calculation of PII scores for pathway-aggregated enrichments by RE-linked TFBS

✓ Calculation of NPPI scores for *normalized* pathway enrichments

Features Mapping

✓ Mapping all gene transcriptional start 10-kb neighborhoods

✓ Mapping all human retrotransposons

✓ Mapping all human TFBS