

Tutorial for Build database

The following is a step-by-step tutorial on building a database, covering the execution logic and key operations of four core scripts

In this case, we used a subset of three types from MISAR to run the analysis. If you want to process all types, please load the file **cell_type_gt.csv**.

Step1: Detecting the interaction between peak regions (step1_detect_TFs_for_peaks.R)

Step2: Fine grained detection based on ArchR (step2_detect_ITs_for_peaks_upon_ArchR.R)

Now, we get TF-peak links

Step3: Generate TF-TG relationship pairs (step3_generate_tf_tg.py)

```
python step3_generate_tf_tg.py
```

```
Loading data...
Data loading done.
Extracting TF lists...
TF lists extraction done.
Building TF peak dictionary...
TF peak dictionary built.
Saving TF peak union...
TF peak union saved.
Extracting marker regions...
Marker regions extracted and saved.
Building TG peak dictionary...
100%|██████████████████████████████████████████████████████████████████████████| 523/523 [00:29<00:00, 17.52it/s] TG peak dictionary built.
Building TF-TG dictionary...
100%|██████████████████████████████████████████████████████████████████████████| 112/112 [00:04<00:00, 22.67it/s] TF-TG dictionary built and saved.
```

Step4: Build the final database (step4_build_database.py)

```
python step4 build database.py
```

100% | ██████████

	Percentage
Yes	78%
No	22%

| 129/129 [00:05<00:00, 23.49it/s]

Filtering RE interactions...

[illegible]

Filtering TFRE pairs...

[illegible]

Building final database...

Database construction completed.

Final database shape: (66769, 4)

Final database saved to: /home/nas2/biod/zhencaiwei/RegChat-main/RegChat_Real_Datasets/MISAR/output/database/intersection_all_DB.csv