

GTRD – an integrated view on transcription regulation

Fedor A. Kolpakov
Institute of Computational
Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
fkolpakov@gmail.com

Semyon K. Kolmykov
FRC Institute of Cytology and Genetics
SB RAS
Institute of Computational
Technologies SB RAS
Novosibirsk, Russia
kolmykovsk@gmail.com

Mikhail A. Kulyashov
Novosibirsk State University
Institute of Computational
Technologies SB RAS
Novosibirsk, Russia
m.kulyashov@mail.ru

Ivan S. Evshin
Institute of Computational
Technologies SB RAS
BIOSOFT.RU, LLC
Novosibirsk, Russia
ivan@biosoft.ru

Yury V. Kondrakhin
Institute of Computational
Technologies SB RAS
Novosibirsk, Russia
yvkondrat@mail.ru

Ruslan N. Sharipov
Novosibirsk State University Institute
of Computational BIOSOFT.RU, LLC
Novosibirsk, Russia
shrus79@biosoft.ru

Abstract — GTRD - Gene Transcription Regulation Database (<http://gtrd.biouml.org/>) contains uniformly annotated and processed NGS data (ChIP-seq, ChIP-exo, DNase-seq, MNase-seq, ATAC-seq and RNA-seq) related with transcription regulation. All cell types (cell lines and tissues) presented in the GTRD were arranged into a dictionary, linked with specialised ontologies (BRENDA, Cell ontology, Uberon, Celosaurus, Experimental factor ontology) and further linked with experiments in specialised databases on transcription regulation: FANTOM 5, ENCODE and GTEx. For integrated view in transcription regulation we have developed web-interface that provides browsing, advanced search possibilities, an integrated genome browser and integrated reports by given cell type, transcription factor or gene.

Keywords — GTRD, transcription regulation, NGS, uniformed analysis, data integration

Introduction

Transcription regulation is a complex process that involves a lot of participants. There is a lot of experimental data of different types obtained for different cell types and different experimental conditions which are stored in different databases.

An important task is integration of such data for a given cell type and experimental condition to get an integrated view on main events related with transcription regulation. This is one of the main goals of GTRD database [1].

Results

The suggested approach is shown on figure 1:

1) we are collecting and uniformly process[1] following NGS data:

- ChIP, Chip-exo - to identify binding regions for transcription factors (TFs) and their coactivators, as well as to identify regions with corresponding histone marks.

- DNase-seq, MNase-seq and ATAC-seq - to identify genome regions with open chromatin and footprints for TF binding sites;

- RNA-seq data where gene expression was changed by TFs activation or knockout to identify genes regulated by these TFs.

2) all cell types (cell lines and tissues) presented in the GTRD were arranged into a dictionary, linked with specialised ontologies (BRENDA, Cell ontology, Uberon, Celosaurus, Experimental factor ontology) and further linked with experiments in specialised databases on transcription regulation: FANTOM 5, ENCODE and GTEx [2].

3) for integrated view in transcription regulation we have developed web-interface:

- based on BioUML platform[3] - it provides browsing and displaying information, advanced search possibilities and an integrated genome browser;

- based on BeanExplorer technology[4] - it provides integrated reports by given cell type, TF or gene.

ACKNOWLEDGMENT

This study was supported by the Russian Science Foundation, grant No. 19-14-00295.

REFERENCES

- [1] Yevshin I, Sharipov R, Kolmykov S, Kondrakhin Y, Kolpakov F. GTRD: a database on gene transcription regulation - 2019 update. *Nucleic acids research*. 2019 Jan 8;47(D1):D100-5.
- [2] Kulyashov M., Kolmykov S., Yevshin I., Kolpakov F. "Description, Characteristic And Algorithm For Creation Of A Dictionary Of Cell Types And Tissues In The Gtrd Database", *CEUR-Workshop proceedings*, 2020, Vol-2569, p.13-18.
- [3] Kolpakov F, Akberdin I, Kashapov T, Kiselev L, Kolmykov S, Kondrakhin Y, Kutumova E, Mandrik N, Pintus S, Ryabova A, Sharipov R. BioUML: an integrated environment for systems biology and collaborative analysis of biomedical data. *Nucleic acids research*. 2019 Jul 2;47(W1):W225-33.
- [4] Bean Explorer <https://github.com/DevelopmentOnTheEdge/be5>

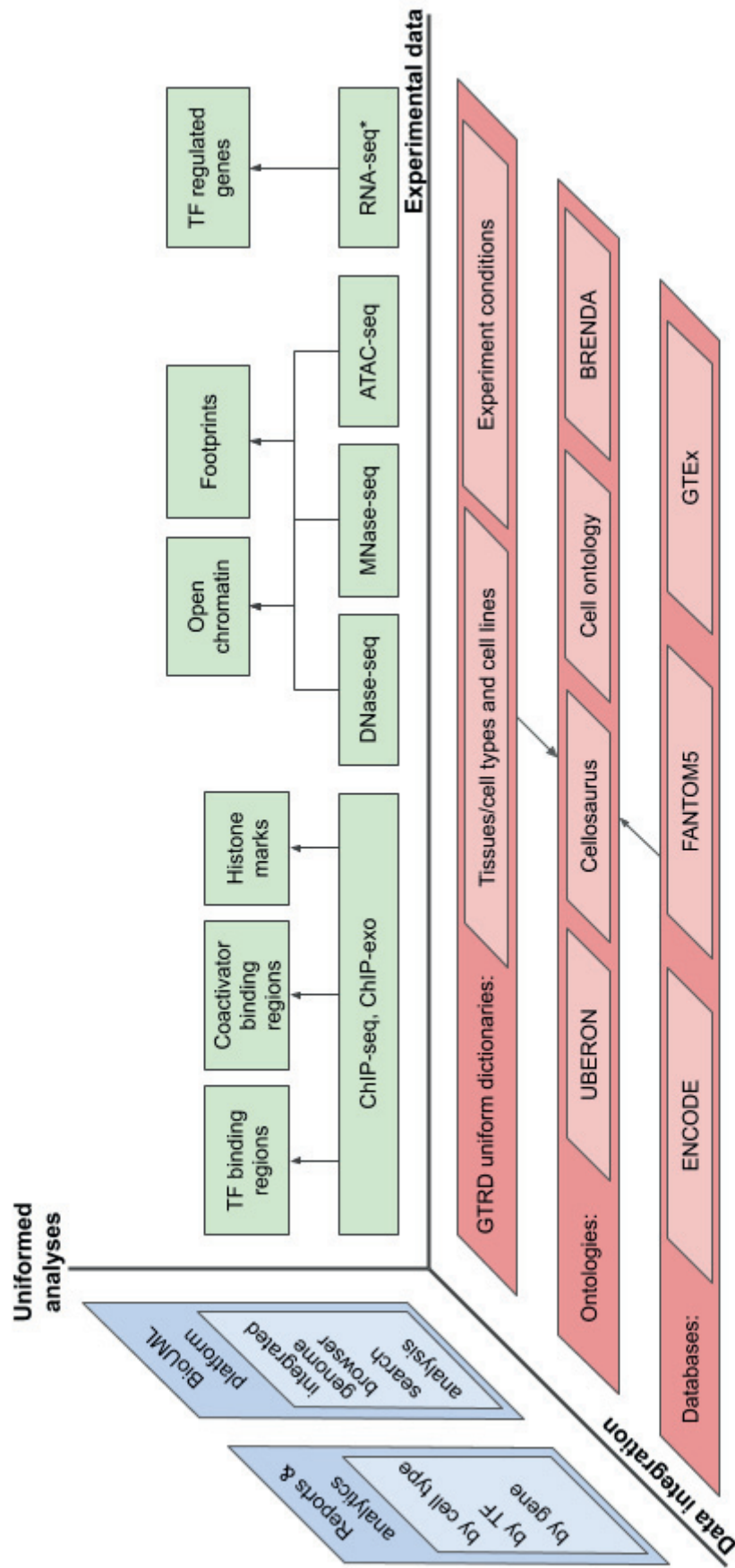


Fig. 1. GTRD - an integrated view on transcription regulation. * RNA-seq data for different TFs activation or knockout.