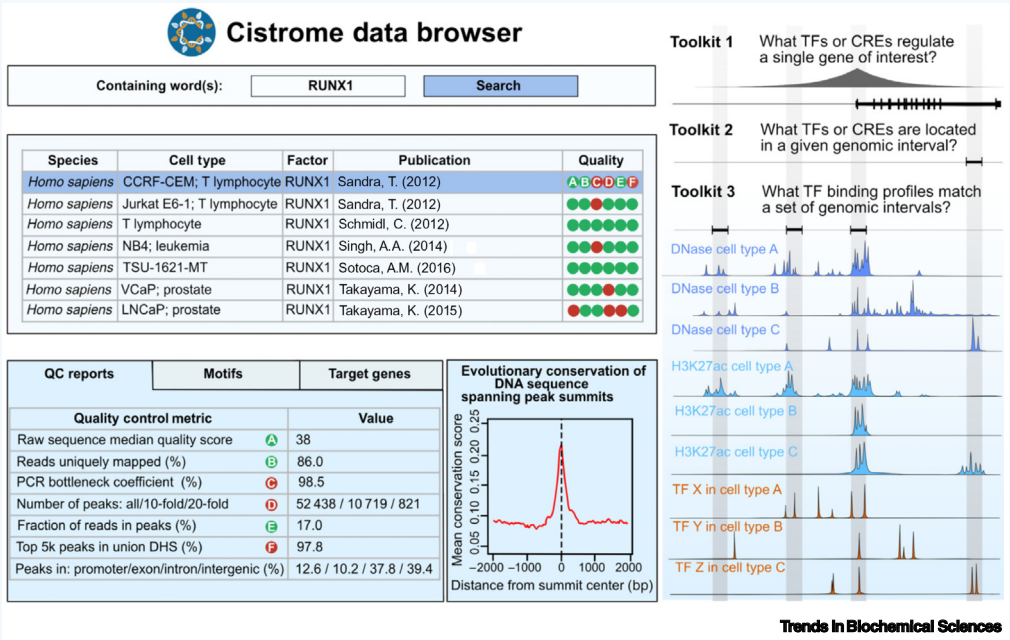
Trends in Biochemical Sciences | Technology of the Month

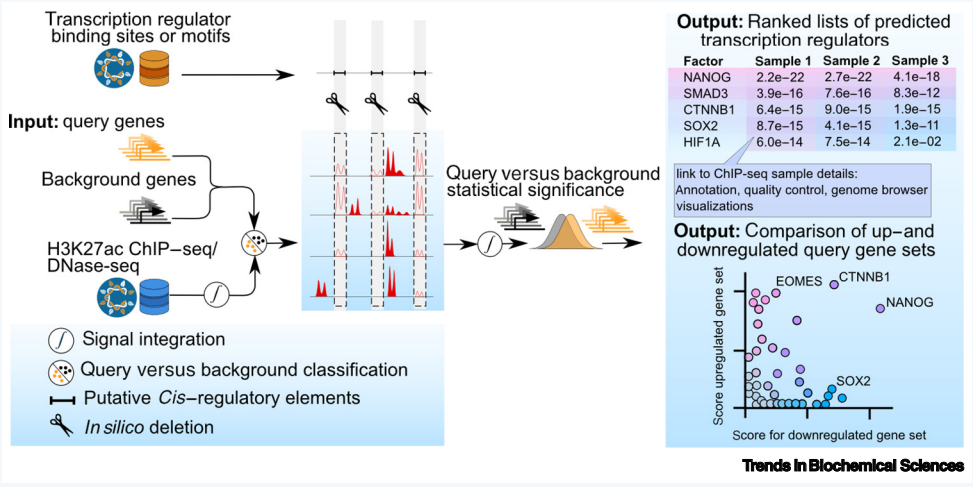
Computational Approaches to Modeling Transcription Factor Activity and Gene Regulation

Clifford A. Meyer[1,*] **and X. Shirley Liu**[1]

[1]Department of Data Science, Dana-Farber Cancer Institute, Harvard T.H. Chan School of Public Health, Boston, MA 02215, USA



The Cistrome Data Browser (DB) (www.cistrome.org/db) contains a comprehensive collection of publicly available ChIP-seq, DNase-seq, and ATAC-seq profiles in human and mouse, processed using consistent workflows. Cistrome DB functions help investigators find relevant profiles and answer common questions on transcriptional or epigenetic gene regulation. Cistrome DB Toolkit functions allow users to query the database by gene, genomic interval, or set of genomic intervals.



The Lisa (lisa.cistrome.org) algorithm predicts the transcription regulators (TRs) of query gene sets, for example, from differential expression. Lisa uses compendia of H3K27ac ChIP-seq and DNase-seq profiles to construct an epigenetic model of *cis*-regulatory activity related to the query gene set, then probes this model with TR cistromes to determine a ranked list of TRs that induce the most significant perturbations in the query gene set.

ADVANTAGES:

Cistrome DB keyword searches allow users to find transcription factor (TF) and histone modification ChIP-seq and chromatin accessibility profiles of interest. Users can then examine their quality, antibody reagent, and explore enriched motifs and target genes.

In well-studied cell types, relevant TFs and *cis*-regulatory elements (CREs) for a gene of interest are readily identifiable using the Cistrome DB Toolkit function 1 (dbtoolkit.cistrome.org).

Querying genomic intervals with Toolkit function 2 provides insights into CREs and noncoding variants, including SNPs and somatic mutations.

Using Cistrome DB Toolkit function 3, sets of differential ATAC-seq genomic regions can be compared with the Cistrome DB compendium to find TFs that may be responsible for the observed chromatin accessibility changes.

TFs that regulate a set of query genes can be identified using the Lisa algorithm (lisa.cistrome.org).

The Cistrome DB allows experimental investigators to find publicly available data relevant to their own systems. These data can be used to plan informative experiments that complement the available Cistrome DB data.

CHALLENGES:

ChIP-seq is not available for every relevant TF and cell type combination.

Cistrome data generated by different groups varies in quality and can be influenced by different batch effects.

Gene regulation is complex and current models of *cis*-regulation, although informative, are incomplete.

*Correspondence:
cliff_meyer@ds.dfci.harvard.edu (C.A. Meyer).

  https://doi.org/10.1016/j.tibs.2020.09.005

Acknowledgements
This work was supported by a grant from the NIH (U24 CA237617).

Literature

1. Zheng, R. *et al.* (2019) Cistrome Data Browser: expanded datasets and new tools for gene regulatory analysis. *Nucleic Acids Res.* 47, 729–735
2. Qin, Q. *et al.* (2020) Lisa: inferring transcriptional regulators through integrative modeling of public chromatin accessibility and ChIP-seq data. *Genome Biol.* 21, 1–14
3. Wang, S. *et al.* (2016) Modeling *cis*-regulation with a compendium of genome-wide histone H3K27ac profiles. *Genome Res.* 26, 1417–1429
4. Chen, C.H. *et al.* (2020) Determinants of transcription factor regulatory range. *Nat. Commun.* 11, 1–15
5. Layer, R.M. *et al.* (2018) GIGGLE: a search engine for large-scale integrated genome analysis. *Nat. Methods* 15, 123–126
6. Myers, R.M. *et al.* (2011) A user's guide to the Encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 9, e1001046
7. Li, S. *et al.* (2019) Cistrome-GO: a web server for functional enrichment analysis of transcription factor ChIP-seq peaks. *Nucleic Acids Res.* 47, W206–W211

  https://doi.org/10.1016/j.tibs.2020.09.005