



nf-core/clipseq v2.0: a feature-filled DSL2 update

Charlotte Capitanchik^{1,2,3}, Klara Kuret^{4,5}, Chris M. Cheshire², Sam Ireland³, Alex Harston³, Marc D. Jones², Ira Iosub², Rupert Faraway², Oscar G. Wilkins^{2,6}, Charlotte West², Nicholas M. Luscombe⁷, Anob Chakrabarti^{2,8} and Jernej Ule^{1,2,4}.

1. UK DRI at King's College London, UK. 2. The Francis Crick Institute, UK. 3. Goodwright Ltd. 4. National Institute of Chemistry, Ljubljana, Slovenia. 5. Jozef Stefan International Postgraduate School, Jamova cesta 39, 1000, Ljubljana, Slovenia. 6. UCL Queen Square Motor Neuron Disease Centre, Department of Neuromuscular Diseases, UCL Queen Square Institute of Science and Technology, Japan. 8. UCL Respiratory, Division of Medicine, University College London, London, UK.

What is CLIP-Seq?

A high-throughput method to determine protein-RNA binding sites at single-nucleotide resolution. There are many variants with the most popular being iCLIP, eCLIP and PAR-CLIP.

| Preparation of cross-linked cell lysate | \rightarrow | Purification of specific cross-linked RNA fragments | \rightarrow | cDNA library sequencing and analysis |
|---|---------------|---|---------------|--------------------------------------|
|---|---------------|---|---------------|--------------------------------------|

Clippy: a new practical peak caller for CLIP analysis

| Howitzerke | | | | ● ● ● ■ Dash × ← → C ☆ ① 127.0.0.1:8050 | + | ද එ |
|--------------------------|---|---------------------|-----------------------|---|--|--|
| HUW II WUIKS. | | | Choosing parameters | III Apps 🖻 Crick Bookmarks 🔶 Google Scholar 😤 GEO - NCBI 🝐 Thesis - Google D 💁 Email - Charlotte 🗎 lab » 🗎 Other Boo Clippy Interactive Parameter Search | | |
| | | | for peak calling is | Run this on the command line with: | ./clip.py -i tests/data/crosslinkcounts.bed -a tests/data/a tests/data/genome.fa.fai -o OUTPUT_PREFIX -n 15 -up 0 -dowr | nnot.gff -g 0 -x 1 -hc 0.7 -mg 1 -mb 11 |
| Minimal annotation (gtf) | Crosslinks (bed) derived from any protocol | Relative height = 1 | always a pain. Clippy | Peak width Mean 문 | YJR094C; IME1 ; Total xlinks = 2043 | Controls |
| | | * Peak | allows you to easily | Prominence threshold Peak summits Yes | 15 | Server status: C Gene search |



Hafner, Markus, et al. "CLIP and complementary methods." Nature Reviews Methods Primers 1.1 (2021): 20.

Data analysis is analogous to ChIP-Seq, with several added complexities: variation in RNA abundance complicates peak calling; no clear consensus on what kind of input data to use, if at all; variety of diagnostic events and many variations on wet lab method with unclear impact on data quality.



idunch an interactive dashboard to help choose parameters, by visualising the impact of changes on multiple transcripts at once:

Clippy has better motif

recall than the recently

released "Skipper":





nf-core/clipseq v1.0

We released nf-core/clipseq v1.0 in April 2021. Since then, we have had >815 unique clones from our GitHub repo, suggesting substantial usership within the CLIP community.

New additions for v2.0

= feature added in dev; = feature dev in progress



- Transcriptomic output xlinks and peaks 날
- Positionally Enriched Kmer Analysis of peaks 🐲 \mathbf{X}
- Gene, biotype and transcript region summaries 🐲
- Metagene of xlinks around transcriptomic landmarks \star
- \star Replacing unitools with unicollapse for more efficient handling of large BAM files during UMI deduplication 💥
- ★ Clippy peak calling
- eCLIP mode for ENCODE eCLIPs 💥
- Code-free analysis & data-sharing via Flow.bio 🐲 \star
- \star Group analysis 🌱
- Peak calling with input data 🌱
- Human ncRNA-aware analysis 🌱
- ★ PAR-CLIP data handling ¥

We welcome input from labs using CLIP-Seq methods as we would like our pipeline to be of use to the widest possible community.



Code-free CLIP-Seq analysis and data-sharing on Flow.bio

Analyse your CLIP data for free online via our intuitive web interface and contribute to a growing community-driven database of CLIP and other data types (RNA-Seq, ChIP-Seq and scRNA-Seq also available).

Link to your flow.bio project in your manuscript to easily share your data and resulting analysis with the wider community.





Preserve relationships between data objects and the code and parameters that generated them - no more random peak files in old folders.

Search for available data to integrate with your own.



Check out our papers:

New Results

Follow this preprint

SOFTWARE TOOL ARTICLE

Flow: a web platform and open database to analyse, store, curate and share bioinformatics data at scale

D. Marc Jones, Charlotte Capitanchik, D Sam Ireland, Alex Harston, D Chris Cheshire, D Marc Jones, D Flora C.Y. Lee, D Igor Ruiz de los Mozos, D Ira A. Iosub, D Klara Kuret, D Rupert Faraway, Doscar G. Wilkins, D Rahul Arora, D Martina Hallegger, D Miha Modic, D Anob M. Chakrabarti, Dicholas M. Luscombe, Diernej Ule doi: https://doi.org/10.1101/2023.08.22.544179

nf-core/clipseq - a robust Nextflow pipeline for comprehensive CLIP data analysis [version 1; peer review: 1 approved, 2] approved with reservations]

Charlotte West, Charlotte Capitanchik, Chris Cheshire, Nicholas M. Luscombe, Anob Chakrabarti 🗠 💿, Jernej Ule 🗠 💿