
Feng Lab - Data Science Toolbox and ChIP-Seq

Release 1.0

February 18, 2017

1	Getting started	3
1.1	Using this Web site	3
1.2	Preparation	3
1.3	Intro to technology	3
2	For today	5
2.1	ChIP-seq	5
3	Examples from the Wild 1: REGULAR EXPRESSIONS	7
3.1	NGS Analysis of ChIP-seq data with NUCwave	7
4	Examples from the Wild 2: IPython Notebooks	9
4.1	NGS Analysis of ChIP-seq data using IPython Notebooks to Explore	9
5	Examples from the Wild 3: Git, Github, and Gists	11
5.1	Git, Github, and Gists	11
6	Examples from the Wild 4: R, the Bioconductor Project for R, RStudio	13
6.1	R, the Bioconductor Project for R, RStudio	13
7	Sources	15
8	Going forward	17
8.1	Look into	17
9	Literature Selections for ChIP-seq	21
9.1	ChIP-Seq	21
9.2	Bias issues	22
9.3	Related	23
9.4	Motif identification	23
10	ACRONYMS	25
10.1	Partial List of the Acronyms in ChIP-related Work	25

The Data Scientist's Toolbox for ChIP-Seq and Beyond

Feng Lab Group meeting by Wayne in March.

Slides link.

A Google Doc for sharing today

<http://we.tl/TYWUi1aFRf>

3:40am-5:30pm, March 19th, 2015. (2nd session; see [first session info here](#); next session [hands-on here](#))

Getting started

Using this Web site

Reload to get the latest links!

A Google Doc for sharing today

Preparation

Read

Interactive notebooks: Sharing the code by Helen Shen. *Nature*. 2014 Nov 6;515(7525):151-2. doi: 10.1038/515151a. PMID: 25373681

Programming tools: Adventures with R by Sylvia Tippmann. *Nature*. 2015 Jan 1;517(7532):109-10. doi: 10.1038/517109a. PMID: 25557714

Tech prep

Be sure you have a modern, updated browser on your system. Preferably Chrome or Firefox.

Register and do the follow-up activation at [SourceLair](#).

Register for [Sagemath Cloud](#).

Be sure to have a good text editor on your computer. Sounds like you may have been using AquaMac in the past and so this shouldn't be a problem. I highly recommend [Sublime Text](#). However, for what we'll be doing Thursday, even TextWrangler on a Mac will be sufficient. For those not on a Mac, I'd recommend [Sublime Text](#) or [Notepad++](#) or [jEdit](#).

Intro to technology

We'll use as a group two technologies today.

- [SourceLair](#)
- [Sagemath Cloud](#)

The idea for using cloud-based tools is to make it easier upfront to get coding and then you can modify what you use as you develop your coding workflow preferences. (Sorry for needing two, but finding a good interface that has all the features desired and works on the Upstate network is not easy.)

For today

ChIP-seq

Background on ChIP-seq in preparation for running through an analysis workflow next session.

I'll sprinkle in some real world examples of using some available tools with at least one possibly being hands-on for those who wish to participate

Slides [link](#).

A Google Doc for sharing today, if needed

Examples from the Wild I: REGULAR EXPRESSIONS

NGS Analysis of ChIP-seq data with NUCwave

ChIP-Seq example at NUCwave site

S. cerevisiae reference genome was downloaded from SGD and FASTA headers for chromosome names were replaced with chrI-chrXVI.

Of course, there are only sixteen chromosomes in yeast, plus the mitochondrial genome, so this is not an overly difficult to do by hand. But it is tedious and offers a good place to utilize regular expressions.

Highly recommend the following combination for learning Regular Expressions, or `Regex` or `Regexp` as it is often called:

- Chapters 2 & 3 of Practical Computing for Biologists book by Haddock and Dunn. The related appendix #2 is freely available as part of tables of Appendices from Practical Computing for Biologists book by Haddock and Dunn
- Regular Expressions Primer
- Regular Expressions 101: online regex editor and debugger tool (This seems best with `g` global modifier on.)

First I'll demonstrate doing this with Sublime Text using the process I already worked out.

So what are Regular Expressions? See Exploring with Regular Expressions 101.

I'll demo wildcards, character sets, quantifiers and capturing.

Finally, we'll use Regular Expressions 101 to really follow what was going on in this example.

Examples from the Wild 2: IPython Notebooks

NGS Analysis of ChIP-seq data using IPython Notebooks to Explore

Determining Average ChIP-seq signal over promoters with Metaseq > This example demonstrates the use of `:mod:metaseq` for performing a common task when analyzing ChIP-seq data: what is the average signal over transcription start sites (TSS) throughout the genome?

That looks interesting but what framework is being used to make and host this?

So what are IPython Notebooks?

Allows you to code interactively in your browser and take advantage of all the aspects of HTML and other special web features, including sharing online

These are especially useful for exploring data and developing code or developing approaches to analyzing your data.

Titus Brown's screencast and associated notebook illustrates much of this.

I'll show two other notebooks I have made and show them interactively.

Q: You've run your notebook and populated the cells, now how can you share it with colleagues? A: If you follow these steps you or anyone else you share it with can see your notebook on the web.

- Upload your notebook code to somewhere. Github or simply even as a Gist will work fine.
- Place the URL [here](#) and click 'Go!'

Note that the notebooks shared in this form will not be interactive. You can though download them and run them locally.

The future ...

The IPython Notebooks concept goes beyond Python and now they are developing a language-agnostic version of the Notebook as the [Jupyter Project](#) project.

Back to Metaseq

The page [here](#) actually has a legend with some of the plots that describes additionally aspects of the exploratory analyses done on the page [Determining Average ChIP-seq signal over promoters with Metaseq](#).

Examples from the Wild 3: Git, Github, and Gists

Git, Github, and Gists

Quick tour of Github site and Gists since most useful for getting started in the world of using git for version control software.

See a section under ‘Going forward’ for additional resources.

Examples from the Wild 4: R, the Bioconductor Project for R, RStudio

R, the Bioconductor Project for R, RStudio

Quick tour of RStudio to simply show it is exceptional.

R is very much about the concept of [tidy data](#)

RStudio can easily be run on Amazon Web Services if computing power is an issue.

See a section under ‘Going forward’ for additional resources.

Sources

The sources for the information used today came from those linked throughout the content.

However, certain sources deserve special highlighting as they were particularly useful in developing this presentation, contain a wealth of related resources, or are especially pertinent at this stage.

- Practical Computing for Biologists book by Haddock and Dunn
- Interactive notebooks: Sharing the code by Helen Shen. Nature. 2014 Nov 6;515(7525):151-2. doi: 10.1038/515151a. PMID: 25373681
- Programming tools: Adventures with R by Sylvia Tippmann. Nature. 2015 Jan 1;517(7532):109-10. doi: 10.1038/517109a. PMID: 25557714
- Feng et al. 2012. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012 Sep; 7(9): 10.1038/nprot.2012.101.
- Titus Brown and Colleague's Next-Gen Sequence Analysis Workshops, most recent is [Next-Gen Sequence Analysis Workshop (2014)](<http://angus.readthedocs.org/en/2014/>) Particularly pertinent are the sections Istvan Albert's 2012 ChIP-Seq lecture, Day 7: ChIP-seq: Peak Predictions and Cis-regulatory Element Annotations, Using MEME to identify TF binding motif from ChIP-seq data and here.
- ChIP- and DNase-seq data analysis workshop 2014

Going forward

Look into

- Practical Computing for Biologists book by Haddock and Dunn
- Interactive notebooks: Sharing the code by Helen Shen. *Nature*. 2014 Nov 6;515(7525):151-2. doi: 10.1038/515151a. PMID: 25373681
- Programming tools: Adventures with R by Sylvia Tippmann. *Nature*. 2015 Jan 1;517(7532):109-10. doi: 10.1038/517109a. PMID: 25557714
- a two year-old screencast intro of IPython notebook by Titus Brown (Skip to the three-minute mark since we aren't necessarily interested in running it on Amazon web services right now.) A non-interactive version of the notebook he demonstrates is [here](#).
- Another take on the wonders of the IPython notebook, [from a blog](#).
- Orchestrating high-throughput genomic analysis with Bioconductor. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M. *Nat Methods*. 2015 Jan 29;12(2):115-21. doi: 10.1038/nmeth.3252. PMID: 25633503
- Programming: Pick up Python by Jeffrey M. Perkel. *Nature*. 2015 February 5;518:125–6. doi:10.1038/518125a. PMID:25653001 [too](#)
- March 2015 blog post suggesting mandatory primer courses for basic skills for students in cellular & molecular biology, genetics, and related subfields
- A History of Bioinformatics (in the Year 2039), a presentation by Titus Brown encouraging good data practices in his unique way

Regular Expressions

- Chapters 2 & 3 of Practical Computing for Biologists book by Haddock and Dunn. The related appendix #2 is freely available as part of tables of Appendices from Practical Computing for Biologists book by Haddock and Dunn
- Regular Expressions Primer
- Regular Expressions 101: online regex editor and debugger tool (Seems best with `g` global modifier on.)
- RegExr v2.0: online tool to learn, build, & test Regular Expressions
- regex tester

- [Python Regular Expression Testing Tool](#)
- For using regular expressions in Sublime Text, you need to click on the box `. *` next to `Find What` when `Find` tool open to activate. See [here](#) for more information.
- In TextWrangler the trick to activate Regular Expressions is to toggle on `Grep` box under `Matching` in the `Find and Replace` panel. See [here](#) about 1 and half minutes into the video.
- [BBEdit-TextWrangler_RegEx_Cheat_Sheet.txt](#)
- See [here](#) under ‘Search and replace with special characters (regular expressions)’ for using Regular Expressions in AquaMacs.

IPython Notebook

- Interactive notebooks: Sharing the code by Helen Shen. *Nature*. 2014 Nov 6;515(7525):151-2. doi: 10.1038/515151a. PMID: 25373681
- a two year-old screencast intro of IPython notebook by Titus Brown (Skip to the three-minute mark since we aren’t necessarily interested in running it on Amazon web services right now.) A non-interactive version of the notebook he demonstrates is [here](#).
- Another take on the wonders of the IPython notebook, from a [blog](#)
- The future of the IPython Notebook is the Jupyter project
- Analyzing data with R in the IPython notebook

R and Bioconductor in general

- Programming tools: Adventures with R by Sylvia Tippmann. *Nature*. 2015 Jan 1;517(7532):109-10. doi: 10.1038/517109a. PMID: 25557714
- Orchestrating high-throughput genomic analysis with Bioconductor. Huber W, Carey VJ, Gentleman R, Anders S, Carlson M, Carvalho BS, Bravo HC, Davis S, Gatto L, Girke T, Gottardo R, Hahne F, Hansen KD, Irizarry RA, Lawrence M, Love MI, MacDonald J, Obenchain V, Oleś AK, Pagès H, Reyes A, Shannon P, Smyth GK, Tenenbaum D, Waldron L, Morgan M. *Nat Methods*. 2015 Jan 29;12(2):115-21. doi: 10.1038/nmeth.3252. PMID: 25633503

Learning R

- See the box in the article [Programming tools: Adventures with R](#) by Sylvia Tippmann. *Nature*. 2015 Jan 1;517(7532):109-10. doi: 10.1038/517109a. PMID: 25557714.
- The Coursera courses in Johns Hopkins’ [Data Science Specialization](#), in particular the [R Programming](#) and [Getting and Cleaning Data](#) courses. If you are brand new to this and don’t yet know how to use Github, [The Data Scientist’s Toolbox](#) would probably be helpful as a starting point.
- You do get a bit of flavor for the use of R in data analysis in the Coursera courses [Bioinformatic Methods I](#) and [Bioinformatic Methods II](#)
- [Comparing Python and R for Data Science](#)
- [How to Transition from Excel to R: An Intro to R for Microsoft Excel Users](#)

ChIP-seq data analysis

- Titus Brown and Colleague's Next-Gen Sequence Analysis Workshops, most recent is [Next-Gen Sequence Analysis Workshop (2014)](<http://angus.readthedocs.org/en/2014/>) Particularly pertinent are the sections Istvan Albert's 2012 ChIP-Seq lecture, Day 7: ChIP-seq: Peak Predictions and Cis-regulatory Element Annotations, Using MEME to identify TF binding motif from ChIP-seq data and here.
- ChIP- and DNase-seq data analysis workshop 2014
- Cis-regulatory Element Annotation System by Hyunjin Shin and Tao Liu from Xiaole Shirley Liu's Lab
- ab initio motif finder MEME and the related MEME suite
- MEME-LaB wraps the popular ab initio motif finder in a web tool
- Motif enrichment tool. Blatti C, Sinha S. Nucleic Acids Res. 2014 Jul;42(Web Server issue):W20-5. doi: 10.1093/nar/gku456. Epub 2014 May 23. PMID: 24860165
- Motif-based analysis of large nucleotide data sets using MEME-ChIP

Plus see the 'literature' page in this collection of pages from the session.

Python and ChIP-Seq and related analyses

- Feng et al. 2012. Identifying ChIP-seq enrichment using MACS. Nat Protoc. 2012 Sep; 7(9): 10.1038/nprot.2012.101.
- Metaseq is a Python module and covers Determining Average ChIP-seq signal over promoters with Metaseq.
- NUCwave as described in Comparative analysis of methods for genome-wide nucleosome cartography. Luis Quintales, Enrique Vázquez and Francisco Antequera. Briefings in Bioinformatics 2014. doi: 10.1093/bib/bbu037
- pyDNase - a library for analyzing DNase-seq data as described in Piper et al. 2013. Wellington: A novel method for the accurate identification of digital genomic footprints from DNase-seq data, Nucleic Acids Research 2013; doi: 10.1093/nar/gkt850

Many people currently analyzing DNase-seq data are using tools designed for ChIP-seq work, but may be inappropriate for DNase-seq data where one is less interested in the overlaps of sequenced fragments, but the site at which the cut occurs (the 5' most end of the aligned sequence fragment). We have developed pyDNase to interface with a sorted and indexed BAM file from a DNase-seq experiment, allowing efficient and easy random access of DNase-seq cut data from any genomic location, e.g. Piper et al. 2013. Wellington: A novel method for the accurate identification of digital genomic footprints from DNase-seq data, Nucleic Acids Research 2013; doi: 10.1093/nar/gkt850

R and ChIP-seq

- ChIPpeakAnno: a Bioconductor package to annotate ChIP-seq and ChIP-chip data. Lihua J Zhu, Claude Gazin, Nathan D Lawson, Hervé Pagès, Simon M Lin, David S Lapointe and Michael R Green.

I need to add the other main ones I saw here still.

Git and Github

- Github bootcamp
- main Git resource. You can even learn git right in your browser

- [Simple guide to git](#)
- [About Gists](#)

Questions

- Try Google, probably will lead you to one of my listed resources or...
- [Biostars](#)
- Stackoverflow for general scripting and computing
- [SEQanswers](#) - a high throughput sequencing community
- Try Twitter - for example [this](#)

Literature Selections for ChIP-seq

ChIP-Seq

- Lefrançois P1, Euskirchen GM, Auerbach RK, Rozowsky J, Gibson T, Yellman CM, Gerstein M, Snyder M. Efficient yeast ChIP-Seq using multiplex short-read DNA sequencing. *BMC Genomics*. 2009 Jan 21;10:37. doi: 10.1186/1471-2164-10-37.

We used this method to map the binding sites for Cse4, Ste12 and Pol II throughout the yeast genome and we found 148 binding targets for Cse4, 823 targets for Ste12 and 2508 targets for PolII. Cse4 was strongly bound to all yeast centromeres as expected and the remaining non-centromeric t

- Feng J, Liu T, Qin B, Zhang Y, Liu XS. 2012. Identifying ChIP-seq enrichment using MACS. *Nat Protoc*. 2012 Sep; 7(9): 10.1038/nprot.2012.101.

MACS is coded in Python, an increasingly popular programming language in bioinformatics, which is pre-loaded with the majority of UNIX, Linux, or Mac OS installations. MACS works in Python version 2.6 or 2.7, and version 2.6.5 is recommended. To run MACS in a 64-bit environment, Python for the 64-bit CPU should be installed.

- Use model-based Analysis of ChIP-Seq (MACS) to analyze short reads generated by sequencing protein-DNA interactions in embryonic stem cells. Liu T. *Methods Mol Biol*. 2014;1150:81-95. doi: 10.1007/978-1-4939-0512-6_4.
- ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. Landt SG, Marinov GK, Kundaje A, Kheradpour P, Pauli F, Batzoglou S, Bernstein BE, Bickel P, Brown JB, Cayting P, Chen Y, DeSalvo G, Epstein C, Fisher-Aylor KI, Euskirchen G, Gerstein M, Gertz J, Hartemink AJ, Hoffman MM, Iyer VR, Jung YL, Karmakar S, Kellis M, Kharchenko PV, Li Q, Liu T, Liu XS, Ma L, Milosavljevic A, Myers RM, Park PJ, Pazin MJ, Perry MD, Raha D, Reddy TE, Rozowsky J, Shores N, Sidow A, Slattery M, Stamatoyannopoulos JA, Tolstorukov MY, White KP, Xi S, Farnham PJ, Lieb JD, Wold BJ, Snyder M. *Genome Res*. 2012 Sep;22(9):1813-31. doi: 10.1101/gr.136184.111. PMID: 2295599
- Bailey T, Krajewski P, Ladunga I, Lefebvre C, Li Q, et al. (2013) Practical Guidelines for the Comprehensive Analysis of ChIP-seq Data. *PLoS Comput Biol* 9(11): e1003326. doi:10.1371/journal.pcbi.1003326
- ChIP-Seq: technical considerations for obtaining high-quality data. Kidder BL, Hu G, Zhao K. *Nat Immunol*. 2011 Sep 20;12(10):918-22. doi: 10.1038/ni.2117. PMID: 21934668
- Integrative analysis of public ChIP-seq experiments reveals a complex multi-cell regulatory landscape. *Nucleic Acids Res*. 2015 Feb 27;43(4):e27. doi: 10.1093/nar/gku1280. Epub 2014 Dec 3. Griffon A, Barbier Q1, Dalino J, van Helden J, Spicuglia S, Ballester B. PMID: 25477382
- In silico pooling of ChIP-seq control experiments. Sun G, Srinivasan R, Lopez-Anido C, Hung HA, Svaren J, Keleş S. *PLoS One*. 2014 Nov 7;9(11):e109691. doi: 10.1371/journal.pone.0109691. eCollection 2014. PMID: 25380244

- A comparison of control samples for ChIP-seq of histone modifications. Flensburg C, Kinkel SA, Keniry A, Blewitt ME, Oshlack A. *Front Genet.* 2014 Sep 25;5:329. doi: 10.3389/fgene.2014.00329. eCollection 2014. PMID: 25309581
- ChIP-Enrich: gene set enrichment testing for ChIP-seq data. Welch RP, Lee C, Imbriano PM, Patil S, Weymouth TE, Smith RA, Scott LJ, Sartor MA. *Nucleic Acids Res.* 2014 Jul;42(13):e105. doi: 10.1093/nar/gku463. Epub 2014 May 30. PMID: 24878920

Emphasizes adjusting for gene locus length and that two commonly used gene set enrichment methods, Fisher's exact test and the binomial test implemented in Genomic Regions Enrichment of Annotations Tool (GREAT), can have highly inflated type I error rates and biases in ranking.

Bias issues

- *Proc Natl Acad Sci U S A.* 2013 Nov 12;110(46):18602-7. doi: 10.1073/pnas.1316064110. Epub 2013 Oct 30. Highly expressed loci are vulnerable to misleading ChIP localization of multiple unrelated proteins. Teytelman LI, Thurtle DM, Rine J, van Oudenaarden A.

We analyzed ChIP-Seq peaks of the Sir2, Sir3, and Sir4 silencing proteins and discovered 238 unexpected euchromatic loci that exhibited enrichment of all three. Surprisingly, published ChIP-Seq datasets for the Ste12 transcription factor and the centromeric Cse4 protein indicated that these proteins were also enriched in the same euchromatic regions with the high Sir protein levels. The 238 loci, termed "hyper-ChIPable", were in highly expressed regions with strong polymerase II and polymerase III enrichment signals, and the correlation between transcription level and ChIP enrichment was not limited to these 238 loci but extended genome-wide. ... Whereas ChIP is a broadly valuable technique, some published conclusions based upon ChIP procedures may merit reevaluation in light of these findings.

- Insightful post on PubPeer related to this

Oh, also, if you are not aware of it, Vishy Iyer's recent PLOS One paper finds the exact same artifact as we do.

<http://www.plosone.org/article/authors/info%3Adoi%2F10.1371%2Fjournal.pone.0083506;jsessionid=F590D75E9C265BA>

And related to this discussion: <http://www.biomedcentral.com/1471-2164/11/414>

<http://www.biomedcentral.com/1471-2164/14/254/abstrac>

<http://www.biomedcentral.com/1471-2164/14/638> And this paper from Kevin Struhl:

<http://www.plosone.org/article/related/info%3Adoi%2F10.1371%2Fjournal.pone.0005029;jsessionid=93B6A4A5F2062E6>

Below are other publications reporting the expression-associated ChIP artifact. Fan X, 2009 "Where Does Mediator Bind In Vivo?" (Work in *S. cerevisiae* questioning reports of pervasive genome-wide binding of the Mediator complex.) Waldminghaus T, 2010 "ChIP on Chip: surprising results are often artifacts" (Work in *E. coli*; also see arising correspondence Schindler D, 2013.) Park D, 2013 "Widespread Misinterpretable ChIP-seq Bias in Yeast" (Analysis very similar to ours.) Kasinathan S, 2014 "High-resolution mapping of transcription factor binding sites on native chromatin" (Questions specificity of standard ChIP in *S. cerevisiae* and at HOT regions of *Drosophila*. This work possibly provides a solution to the artifact with a modification of the ChIP technique.) http://www.ncbi.nlm.nih.gov/pubmed/24173036#cm24173036_3919

- Widespread misinterpretable ChIP-seq bias in yeast. Park D, Lee Y, Bhupindersingh G, Iyer VR. *PLoS One.* 2013 Dec 9;8(12):e83506. doi: 10.1371/journal.pone.0083506. eCollection 2013. PMID: 24349523
- Fan X and Struhl, 2009 "Where Does Mediator Bind In Vivo?" *PLOS One.* Published: April 03, 2009 DOI: 10.1371/journal.pone.0005029
- Waldminghaus and Skarstad, 2010 "ChIP on Chip: surprising results are often artifacts" *BMC Genomics.* 2010 Jul 5;11:414. doi: 10.1186/1471-2164-11-414.

- Non-canonical protein-DNA interactions identified by ChIP are not artifacts. Bonocora RP, Fitzgerald DM, Stringer AM, Wade JT. BMC Genomics. 2013 Apr 15;14:254. doi: 10.1186/1471-2164-14-254. (Concerns the *E. coli* data.)
- High-resolution mapping of transcription factor binding sites on native chromatin. Sivakanthan Kasi-nathan,Guillermo A Orsi, Gabriel E Zentner, Kami Ahmad & Steven Henikoff. Nature Methods 11, 203–209 (2014) doi:10.1038/nmeth.2766

The resulting occupied regions of genomes from affinity-purified naturally isolated chromatin (ORGANIC) profiles of *Saccharomyces cerevisiae* Abf1 and Reb1 provide high-resolution maps that are accurate, as defined by the presence of known TF consensus motifs in identified binding sites, that are not biased toward accessible chromatin and that do not require input normalization.

Related

- ChIA- PET - Genome-wide map of regulatory interactions in the human genome. Heidari N, Phanstiel DH, He C, Grubert F, Jahanbani F, Kasowski M, Zhang MQ, Snyder MP. Genome Res. 2014 Dec;24(12):1905-17. doi: 10.1101/gr.176586.114. Epub 2014 Sep 16. PMID: 25228660

Motif identification

- Cis-regulatory Element Annotation System by Hyunjin Shin and Tao Liu from Xiaole Shirley Liu's Lab
A tool designed to characterize genome-wide protein-DNA interaction patterns from ChIP-chip and ChIP-Seq of both sharp and broad binding factors. As a stand-alone extension of our web application CEAS (Cis-regulatory Element Annotation System), it provides statistics on ChIP enrichment at important genome features such as specific chromosome, promoters, gene bodies, or exons, and infers genes most likely to be regulated by a binding factor. CEAS also enables biologists to visualize the average ChIP enrichment signals over specific genomic features, allowing continuous and broad ChIP enrichment to be perceived which might be too subtle to detect from ChIP peaks alone.
- [ab initio motif finder MEME](#) and the related [MEME suite](#)
- [MEME-LaB](#) wraps the popular [ab initio motif finder](#) in a web tool
- Motif enrichment tool. Blatti C, Sinha S. Nucleic Acids Res. 2014 Jul;42(Web Server issue):W20-5. doi: 10.1093/nar/gku456. Epub 2014 May 23. PMID: 24860165
- Motif-based analysis of large nucleotide data sets using [MEME-ChIP](#)

ACRONYMS

Partial List of the Acronyms in ChIP-related Work

ChIP-seq : chromatin immunoprecipitation (“ChIP”) coupled with high throughput sequencing

ChIP-chip or ChIP-on-chip : chromatin immunoprecipitation (“ChIP”) with DNA microarray (“chip”).

MeDip : Methylated DNA immunoprecipitation

MNase-seq : Chromatin digestion with micrococcal nuclease coupled with high throughput sequencing

DNase-seq : high throughput sequencing of DNase I-digested DNA

CC-seq : Chemical cleavage coupled with high throughput sequencing

PE : Paired-end

SR : Single read

ChIA-PET : Chromatin Interaction Analysis by Paired-End Tag Sequencing

R-ChIP : robotic ChIP

Native ChIP (N-ChIP) vs. X-ChIP (typical ChIP where crosslinked)

ChIP-exo : chromatin immunoprecipitation (ChIP) combined with lambda exonuclease digestion followed by high-throughput sequencing

ORGANIC : high throughput sequencing of Occupied Regions of Genomes from Affinity-purified Naturally Isolated Chromatin

Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target sites

Bar-ChIP : chromatin immuno-precipitation followed by sequencing (ChIP-Seq), relying on the direct ligation of molecular barcodes to chromatin fragments