# Brief Introduction of Epigenetics Analyses

We categorized Epigenetics Analyses briefly into three major topics: [Regulation discovery](file:///\\gsc.stjude.org\pages\viewpage.action%3fpageId=53319309#BriefIntroductionofEpigeneticsAnalyses-regdis), [Chromatin state](file:///\\gsc.stjude.org\pages\viewpage.action%3fpageId=53319309#BriefIntroductionofEpigeneticsAnalyses-chrsta) and [3D Genome](file:///\\gsc.stjude.org\pages\viewpage.action%3fpageId=53319309#BriefIntroductionofEpigeneticsAnalyses-3dgen)

# [Regulation discovery](https://wiki.stjude.org/display/CAB/Regulation+discovery)

(↑click to see analyses offered)

Transcription is one of the key stage of ["Central Dogma"](https://en.wikipedia.org/wiki/Central_dogma_of_molecular_biology#Transcription). Regulation of transcription is frequently achieved by [transcription factors(TFs) and enhancers](https://en.wikipedia.org/wiki/Transcriptional_regulation#Through_transcription_factors_and_enhancers). Thus, in order for researchers to identify important implications of TF or epigenetics modifier, the first thing they usually want know if where does that TF binds or how does the epigenetic profiles change if they knockout/knockdown the epigenetic modifier. Common methods for determining this include Chromatin immunoprecipitation (ChIP) followed by massively parallel DNA sequencing (ChIPSeq) and a variety of other methods (such as [Cut-and-Run](https://www.ncbi.nlm.nih.gov/pubmed/29651053), [ULI-NChIP-Seq](https://www.ncbi.nlm.nih.gov/pubmed/25607992)).

Despite ChIPSeq's widespread use, availability of ChIP-grade antibodies and experimental practices could affect the data quality greatly. So before we starting answering biological questions using the ChIPSeq data, it’s essential to make sure the data quality are good enough to carry out the analysis needed. Thus we following [ENCODE guideline](https://doi.org/10.1101/gr.136184.111) for [quality control(QC) of ChIPSeq data](https://wiki.stjude.org/display/CAB/ChIPseq+QC+and+peak+calling). After QC, if the investigator requests and provides a meta data table (sample clinical information), we will pair IP samples and INPUT samples per the investigator’s instruction and then call peaks using [MACS2](https://doi.org/10.1186/gb-2008-9-9-r137) (for point-source factors, narrow peaks) or [SICER](https://doi.org/10.1007/978-1-4939-0512-6_5) (for broad markers, see [known broad markers](https://wiki.stjude.org/display/CAB/ChIPseq+QC+and+peak+calling#ChIPseqQCandpeakcalling-2.2Whichpeakcallertochoose?) for a list).

For most TFs, peaks having a nearby regulated gene upon knockout/knockdown/overexpression of the TF could roughly be considered as the key TF that is directly targeting regulatory elements. However, since ChIP might also pull down sequences binding to additional proteins that interact with your target TF, people usually look for peaks containing that TF's specific motif (could from motif database such as [JASPAR](http://jaspar.genereg.net/)/[TRANSFAC](http://genexplain.com/transfac/) or from [de novo motif discovery by the peaks called](https://en.wikipedia.org/wiki/Sequence_motif#De_novo_motif_discovery)) as TF's directly binding sites.

For the investigation of enhancers/[super-enhancers](https://www.ncbi.nlm.nih.gov/pubmed/25547603), people usually choose H3K27ac (active enhancer marker if distal to gene TSS) or H3K4me1 (poised enhancer marker if distal and no H3K27ac). For example, BRD4 were highly [correlated with H3K27ac](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5508517/figure/F3/). [Med1 also](https://www.ncbi.nlm.nih.gov/pubmed/23582323) have been demonstrated for super-enhancer calling.

[Chromatin state](https://wiki.stjude.org/display/CAB/Chromatin+state)

(↑click to see analyses offered)

Modifications of chromatin structure is another [key epigenetic mechanism used to regulate gene expression](https://www.nature.com/articles/s41556-018-0258-1). At large scale, chromatin can be either densely packed, in the form of heterochromatin, which is largely inaccessible to transcriptional machinery and hence mostly inactive genes, or as open and accessible euchromatin, which contains greater numbers of active genes. They can be regulated as needed via various [epigenetic mechanisms](https://www.nature.com/articles/nrm.2017.119). As a results of regulation, [chromatin accessibility](https://www.nature.com/articles/s41576-018-0089-8) can be measure by method such as [ATAC-seq](https://www.nature.com/articles/nmeth.4396), [DNase-seq](https://www.encodeproject.org/data-standards/dnase-seq/). Together with motif database, transcription factors (TFs) likely to associated with differential accessible regions can be identified.

On the other hand, if you don't have specific targets, profiling histone post-translational modification (PTM) might help you generating hypothesis. However, because histones could have [many type of modifications](https://www.sciencedirect.com/science/article/pii/S0303264717303684#fig0010) and the number of [known PTM are still increasing](https://doi.org/10.1042/EBC20180061), it's nearly impossible to study all their combinations, thus frequently chromatin segmentation computational method such as [ChromHMM](https://www.nature.com/articles/nprot.2017.124) could be applied to learn the  relationship between different PTM and try to summarize the combination of PTM into "chromatin states". This would make the downstream analysis and visualization easier and more clear.

# [3D Genome](file:///\\gsc.stjude.org\display\CAB\3D+Genome)

(↑click to see analyses offered)

Thanks to [Chromosome conformation capture](https://en.wikipedia.org/wiki/Chromosome_conformation_capture)(3C) technologies or 3C-based methods, people are able to study the [3D structure of chromatin](https://www.nature.com/collections/rsxlmsyslk). The most common concepts and their biological implications include:

1. [Topological association domains (TADs)](https://en.wikipedia.org/wiki/Topologically_associating_domain) are megabase long regions that the physically interact within the region are more frequently than with sequences outside the region. They are structural and functional units usually considered as boundaries for enhancers to keep them from target random genes.
2. [Compartments](https://en.wikipedia.org/wiki/Nuclear_organization#A/B_compartments) have first been noticed on Hi-C profiles that shown "checkerboard-like" interaction pattern at large scale ([DOI: 10.1016/j.ymeth.2014.10.031](https://www.sciencedirect.com/science/article/pii/S1046202314003582?via%3Dihub#f0040)).  They could be briefly considered as group of TADs that either be type A (more euchromatin) or type B (more heterochromatin). They usually only shown up for mature cells but not for earlier stages such as embryonic stem cells.
3. [Chromatin loops](https://en.wikipedia.org/wiki/Nuclear_organization#DNA_looping) are stable region to region interaction highly possible involved in enhancer-promoter regulation. Well studied(still not all known yet) proteins important for loops regulation include [CTCF](https://www.ncbi.nlm.nih.gov/pubmed/19563753)(interesting CTCF motif between anchors of loop shown [convergent orientation](https://www.sciencedirect.com/science/article/pii/S0092867414014974?via%3Dihub#fig6)), [YY1](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5785279/), [WAPL](https://www.ncbi.nlm.nih.gov/pubmed/28475897), [NIPBL](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5687303/), [Cohesins](https://www.ncbi.nlm.nih.gov/pubmed/20720539)