Genomics Part II

Applications of Sequencing Technology

Biomedical Data Science: Mining and Modeling
CB&B 752 • MB&B 452
Matt Simon
1/29/25

Overview

- Genomics I (Monday’s lecture): Focus on sequencing technology and genomes.

- Genomics II: (Today’s lecture): Focus on applications of sequencing technology.

    1. Annotation of the genome in chromatin
    2. RNA-seq methods and applications
    3. Topics suggested on Monday.

Workflow

1. Isolation of sample.

   e.g., Isolate DNA and shear.

2. Library preparation

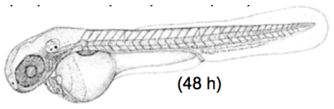
   e.g., Clean up and ligate Y-adaptors.

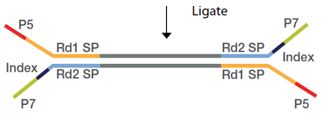
3. Sequencing

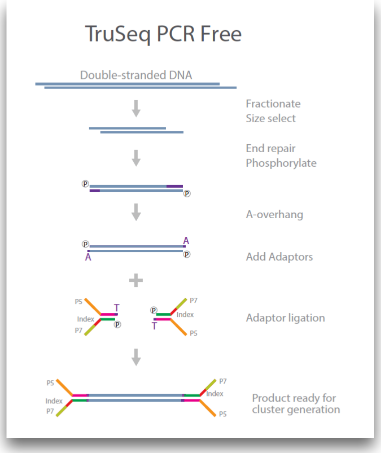
   e.g., Illumina HiSeq

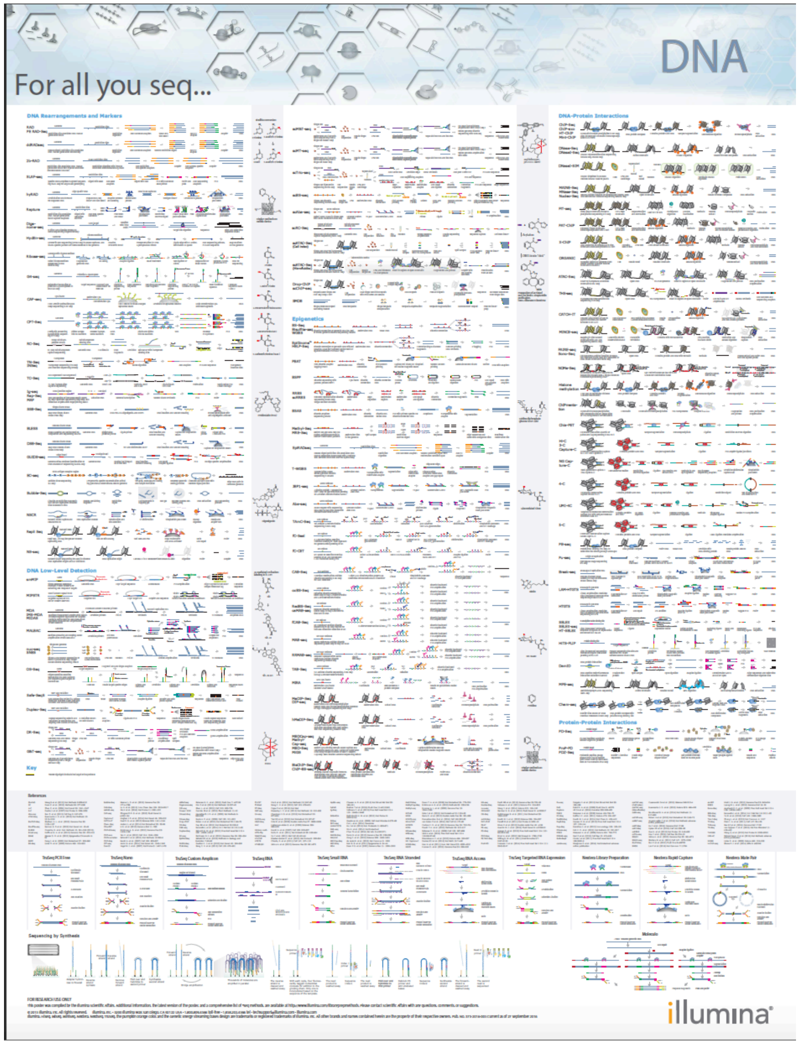
4. Analysis

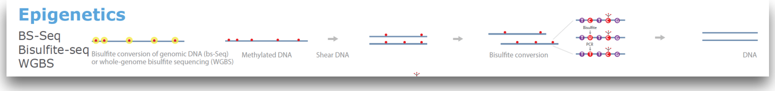
   e.g., Map to genome and interpret.

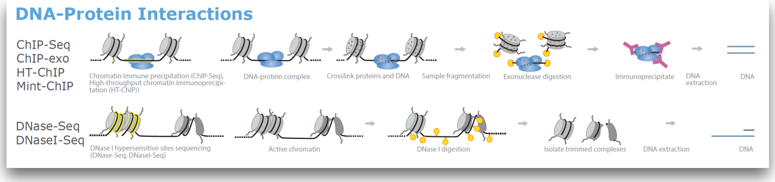


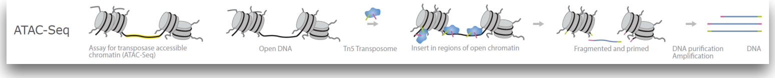


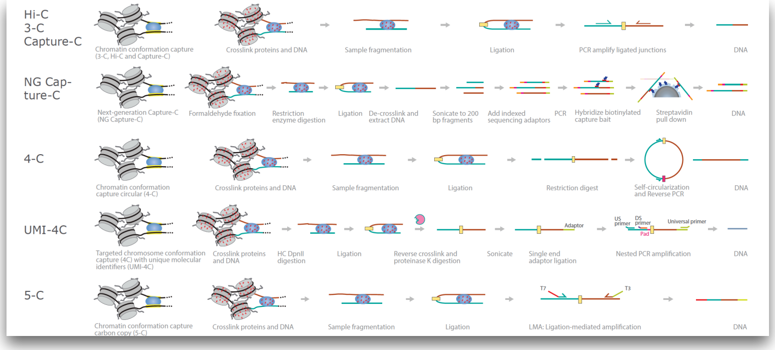


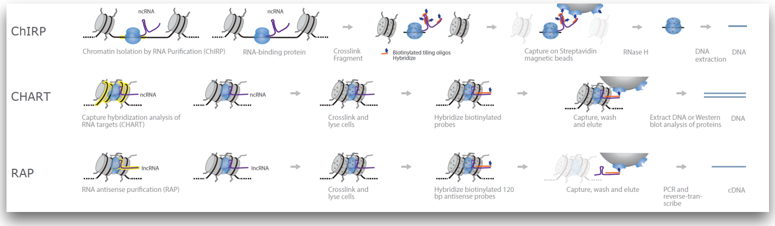


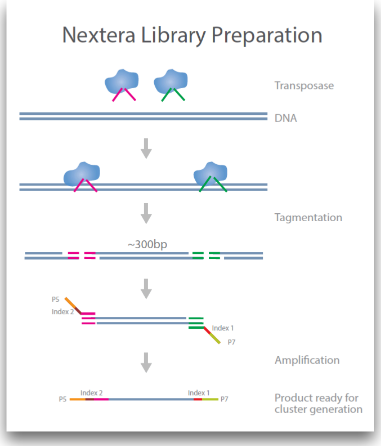

Epigenetics
BS-Seq
Bisulfite-seq
WGBS
Bisulfite conversion of genomic DNA (bs-Seq) or whole-genome bisulfite sequencing (WGBS)
Methylated DNA
Shear DNA
Bisulfite conversion
DNA




ATAC-Seq
Assay for transposase accessible chromatin (ATAC-Seq)
Open DNA
Tn5 Transposome
Insert in regions of open chromatin
Fragmented and primed
DNA purification Amplification
DNA


Hi-C
3-C
Capture-C
Chromatin conformation capture (3-C, Hi-C and Capture-C)
Crosslink proteins and DNA
Sample fragmentation
Ligation
PCR amplify ligated junctions
DNA
NG Capture-C
Next-generation Capture-C (NG Capture-C)
Formaldehyde fixation
Restriction enzyme digestion
Ligation
De-crosslink and extract DNA
Sonicate to 200 bp fragments
Add indexed sequencing adaptors
PCR
Hybridize biotinylated capture bait
Streptavidin pull down
DNA
4-C
Chromatin conformation capture circular (4-C)
Crosslink proteins and DNA
Sample fragmentation
Ligation
Restriction digest
Self-circularization and Reverse PCR
DNA
UMI-4C
Targeted chromosome conformation capture (4C) with unique molecular identifiers (UMI-4C)
Crosslink proteins and DNA
HC DpnII digestion
Ligation
Reverse crosslink and proteinase K digestion
Sonicate
Single end adaptor ligation
Adaptor
US primer
DS primer
Pad
Universal primer
Nested PCR amplification
DNA
5-C
Chromatin conformation capture carbon copy (5-C)
Crosslink proteins and DNA
Sample fragmentation
Ligation
T7
T3
LMA: Ligation-mediated amplification
DNA


ChIRP
ncRNA
Chromatin Isolation by RNA Purification (ChIRP)
RNA-binding protein
Crosslink Fragment
Biotinylated tiling oligos Hybridize
Capture on Streptavidin magnetic beads
RNase H
DNA extraction
DNA
CHART
ncRNA
Capture hybridization analysis of RNA targets (CHART)
Crosslink and lyse cells
Hybridize biotinylated probes
Capture, wash and elute
Extract DNA or Western blot analysis of proteins
DNA
RAP
lncRNA
RNA antisense purification (RAP)
Crosslink and lyse cells
Hybridize biotinylated 120 bp antisense probes
Capture, wash and elute
PCR and reverse-transcribe
cDNA


Nextera Library Preparation
Transposase
DNA
Tagmentation
~300bp
P5
Index 2
Index 1
P7
Amplification
Product ready for cluster generation

https://www.illumina.com/content/dam/illumina-marketing/documents/applications/ngs-library-prep/ForAllYouSeqMethods.pdf

What types of genomic annotation do we have/want?

~3 billion bp

```
ACAATAAATCACATTAATTCCTTATCTCATGTGAAATTTCATATTTATGATTG
ATACCTTTAAATGTCATTTGTTGAAGGAAGATTATTCATTTTTTCATTCAAT
AAATATTTTTTAGAATAATAAGTCCCAGGCACAAGACCAGTATTATGTTCT
AGGCATTGGGGATACCATGTTCACAAGACAGACTATGATTTACAGGATC
AGATGTGGACTCTCAAATTCGACTGAGAATAAAACAGACACTAAACAAG
TAAATAAAGTTAATTTCAAGTTGTAATTGATGCTAGAAAGACAATGAAACA
GAGCCATGTGACCAATGAGAGAGATGAGGGTGGCAGCAGCCTGTTTTA
GATAAGGTACCTGATTGGTGGGATTGGAAGACCTCTCTGAGATTAGTGT
CTTCAGATATGCCTTAATGATATGAAAGAACCATTCATGGGAAGGCCTAG
CATTAAAAACCGTCTAGGCAGAATGAGCAGCAAGTGCAAGGGTCCTGG
ATAGGAATGAGCTGGATATACTCAAGGAAGAAAGAGAAACTATGGAAAA
ATGAAAATAGATTTTAAAACATGTTAATTCACGTTACTTTTTGTTAAATTTA
CTTTTCTTCTTTCACTTCTTACCTGTCAATGTTATTAATATTTTTAGGAACA
ATAAATCACATTAATTCCTTATCTCATGTGAAATTTCATATTTATGATTGATA
CCTTTAAATGTCATTTGTTGAAGGAAGATTATTCATTTTTTCATTCAATAAA
TATTTTTTAGAATAATAAGTCCCAGGCACAAGACCAGTATTATGTTCTAGG
CATTGGGGATACCATGTTCACAAGACAGACTATGATTTACAGGATCAGAT
GTGGACTCTCAAATTCGACTGAGAATAAAACAGACACTAAACAAGTAAAT
AAAGTTAATTTCAAGTTGTAATTGATGCTACTATGGAAAAATGAAAATAGA
TTTTAAAACATGTTAATTCACGTTACTTTTTGTTAAATTTACTTTTCTTCTTT
CACTTCTTACCTGTCAATGTTATTAATATTTTTAGGAACAATAAATCACATT
AATTCCTTATCTCATGTGAAATTTCATATTTATGATTGATACCTTTAAATGT
CATTTGTTGAAGGAAGATTATTCATTTTTTCATTCAATAAATATTTTTTAGA
ATAATAAGTCCCAGGCACAAGACCAGTATTATGTTCTAGGCATTGGGGAT
ACCATGTTCACAAGACAGACTATGATTTACAGGATCAGATGTGGACTCTC
AAATTCGACTGAGAATAAAACAGACACAAACAAGTAAATAAAGTTAATTT
CAAGTTGTAATTGATGCTATCCCAGGCACAAGACCA….
```

Genes:

- Coding, noncoding, miRNA, etc.
- Isoforms
- Expression

Genetic variation:

- SNPs and CNVs

Sequence conservation

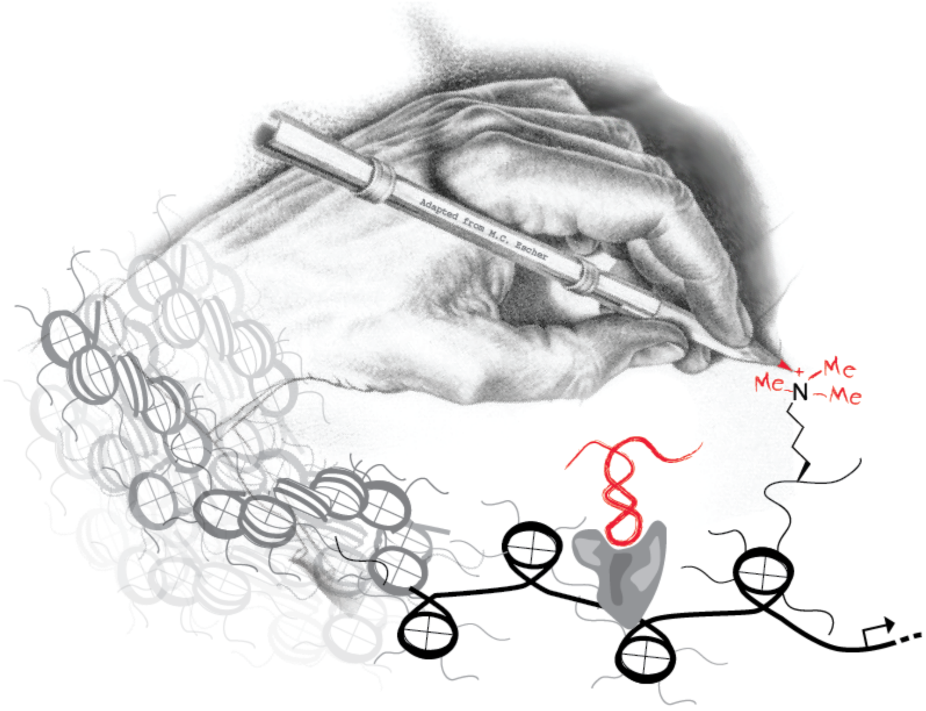
Regulatory sequences:

- Promoters
- Enhancers
- Insulators

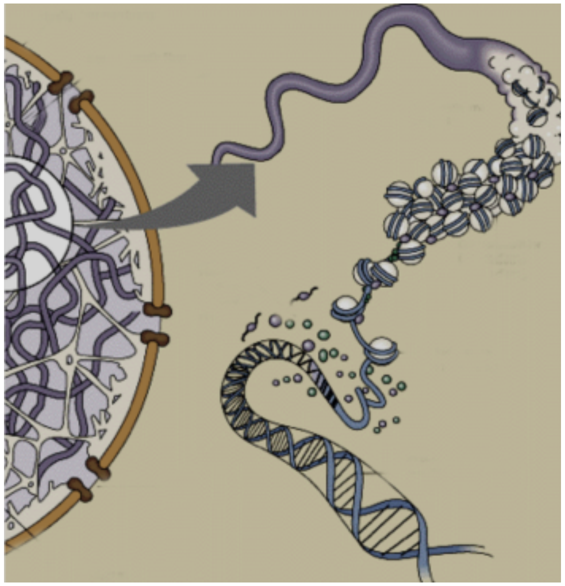
Epigenetics:

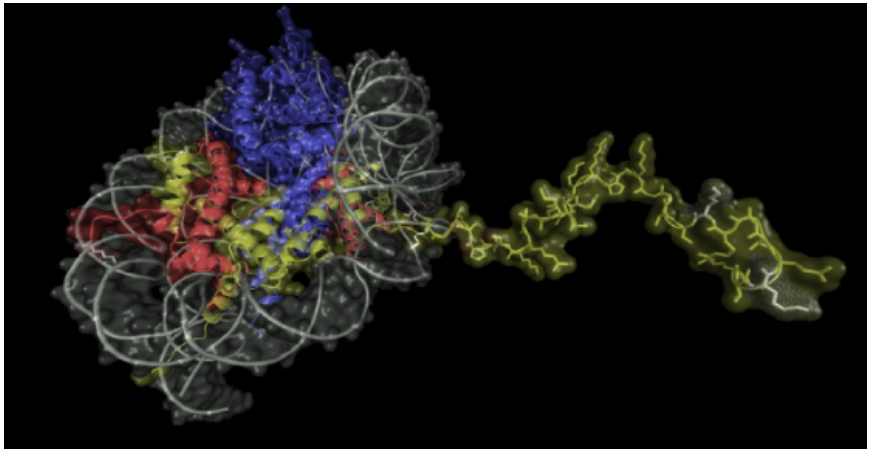
- DNA methylation
- Chromatin

Part 1. How do cells annotate their genomes?



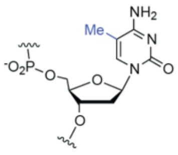
DNA in the cell is packaged into chromatin

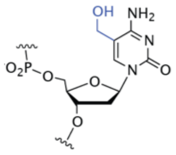


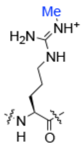


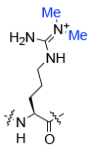
Modeled nucleosome based on Luger et al., *Nature* **1997** *389,* 251.

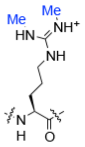
Summary and nomenclature of common covalent modifications.

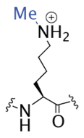

mC

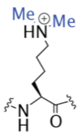

hmC

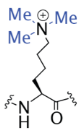

Rme1

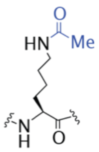

Rme2a

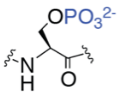

Rme2s

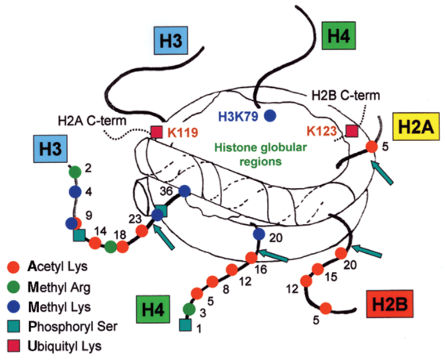

Kme1


Kme2


Kme3


Kac


Sph



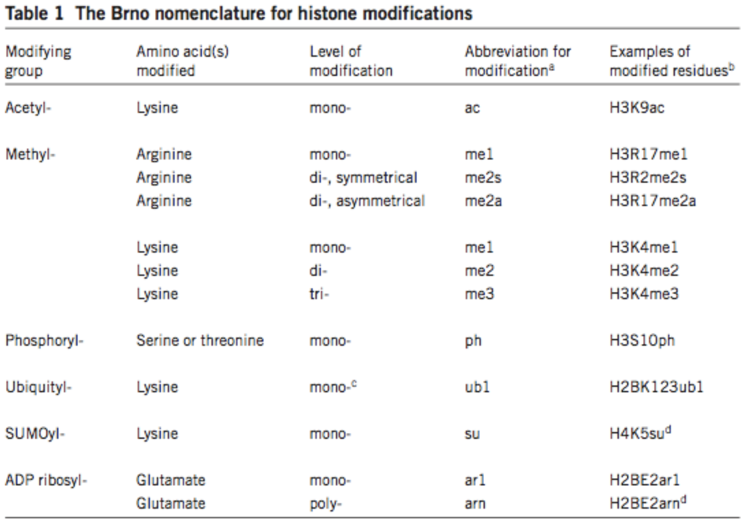
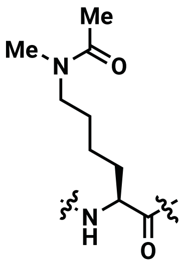


Table 1 The Brno nomenclature for histone modifications

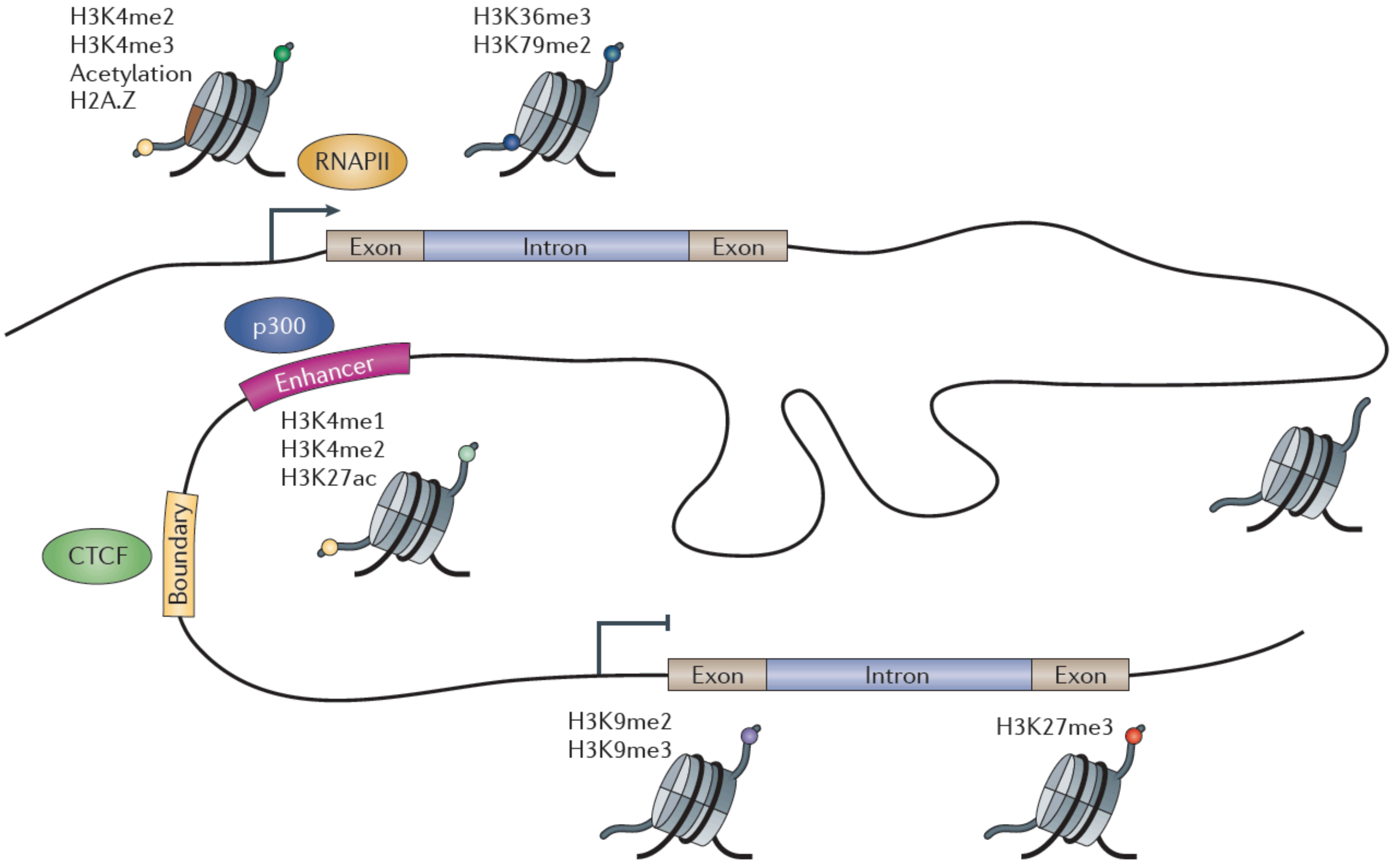
Modifying group	Amino acid(s) modified	Level of modification	Abbreviation for modification[a]	Examples of modified residues[b]
Acetyl-	Lysine	mono-	ac	H3K9ac
Methyl-	Arginine	mono-	me1	H3R17me1
	Arginine	di-, symmetrical	me2s	H3R2me2s
	Arginine	di-, asymmetrical	me2a	H3R17me2a
	Lysine	mono-	me1	H3K4me1
	Lysine	di-	me2	H3K4me2
	Lysine	tri-	me3	H3K4me3
Phosphoryl-	Serine or threonine	mono-	ph	H3S10ph
Ubiquityl-	Lysine	mono-[c]	ub1	H2BK123ub1
SUMOyl-	Lysine	mono-	su	H4K5su[d]
ADP ribosyl-	Glutamate	mono-	ar1	H2BE2ar1
	Glutamate	poly-	arn	H2BE2arn[d]


Kacme

$\underbrace{\text{H3}}_{\text{Histone}}$ $\underbrace{\text{K27}}_{\text{Residue}}$ $\underbrace{\text{ac}}_{\text{Modification}}$

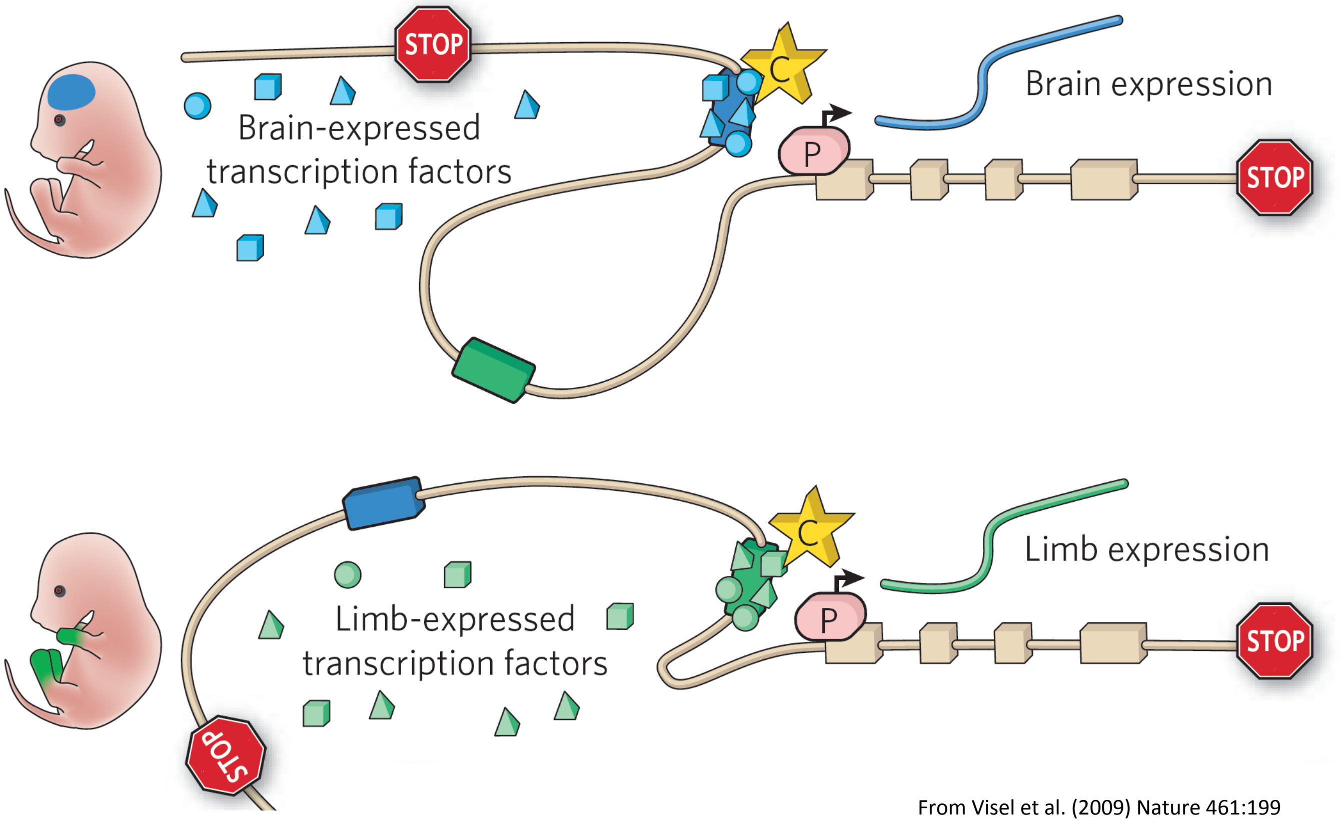
Turner, B. M. Reading signals on the nucleosome with a new nomenclature for modified histones. *Nat Struct Mol Biol* 12, 110–112 (2005).

Chromatin modifications correlate with different genomic functions.



Zhou *et al. Nat Rev Genet* 12:7 (2011)

Regulation is temporally and specially controlled



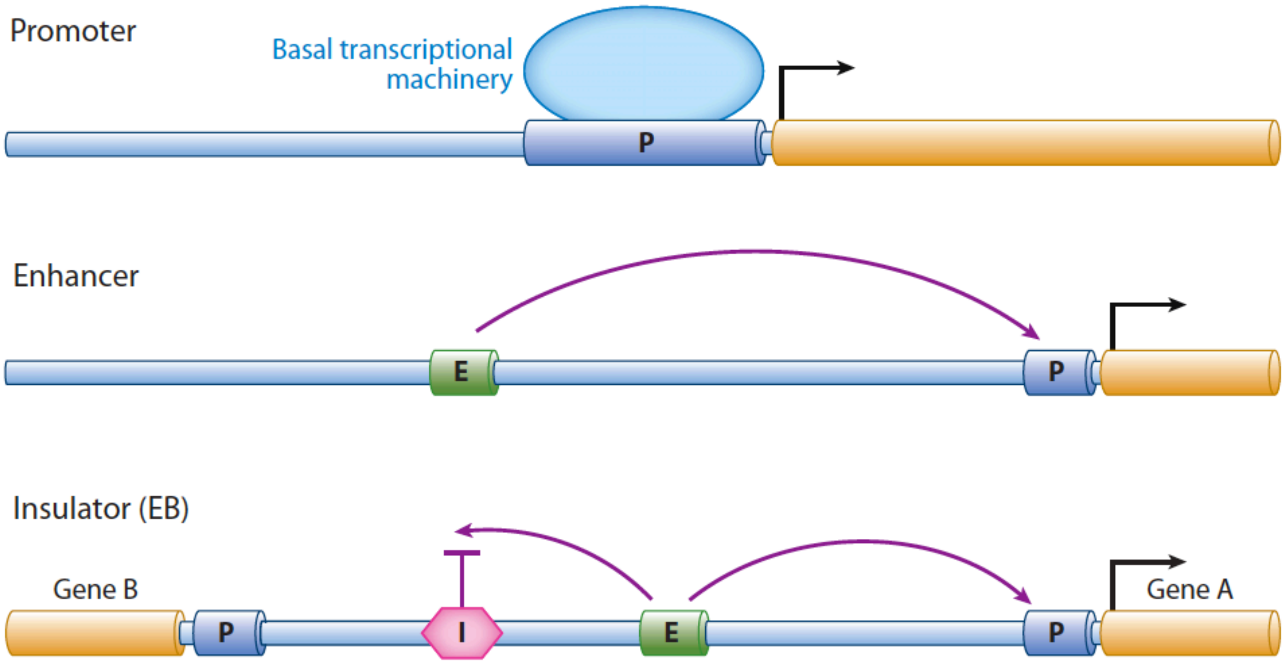
From Visel et al. (2009) Nature 461:199

Using sequencing to annotate the genome

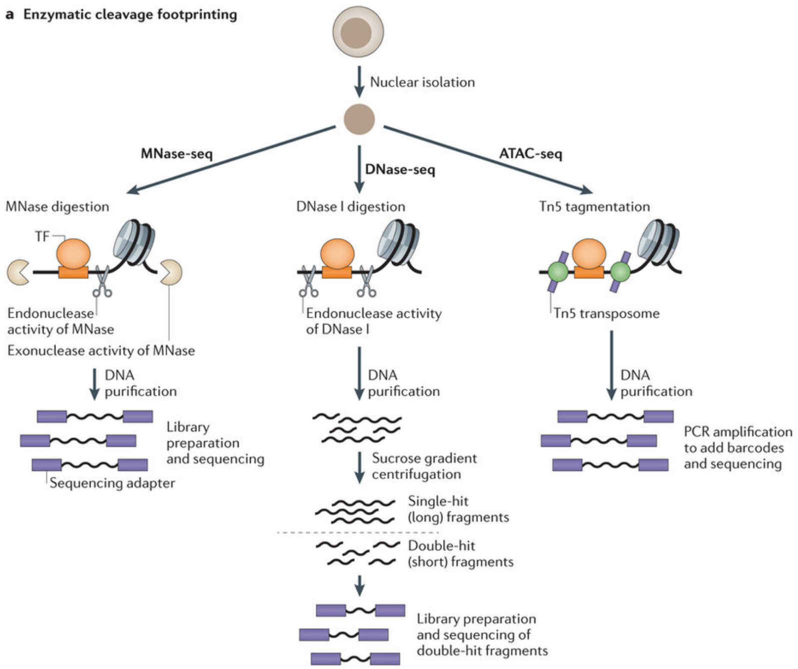
1. Where are the cis-acting regulatory elements in DNA?
   - A. DNase I hyper-sensitivity mapping (**DNase-Seq**).
   - B. **FAIRE** to map regulatory elements.
   - **C. ATAC-Seq to map regulatory elements.**
2. How does the chromatin composition vary across the genome?
   - **D. ChIP-seq of transcription factors (or in high res, ChIP-exo)**
   - **E. CUT&RUN and CUT&Tag for small scale/single cell analysis.**
3. Where is RNA polymerase transcribing?
   - F. **ChIP-Seq** of polymerase.
   - G. **GRO-Seq, PRO-Seq, TT-seq** and **NET-Seq** to measure RNA polymerase activity.
4. What sites are methylated in the genome?
   - H. **Bisulfite-Seq** to measure mC levels.
   - I. **Methyl-Seq** to measure mC levels.
5. How is the genome folded in the nucleus?
   - J. **Hi-C** to measure ligation/contact frequencies.
   - K. **3C/4C/5C** to measure looping at specific loci.

Targeted approaches v Global approaches

How do we identify regulatory elements in the genome?



Using differences in biochemical properties of regulatory elements to identify them by Seq

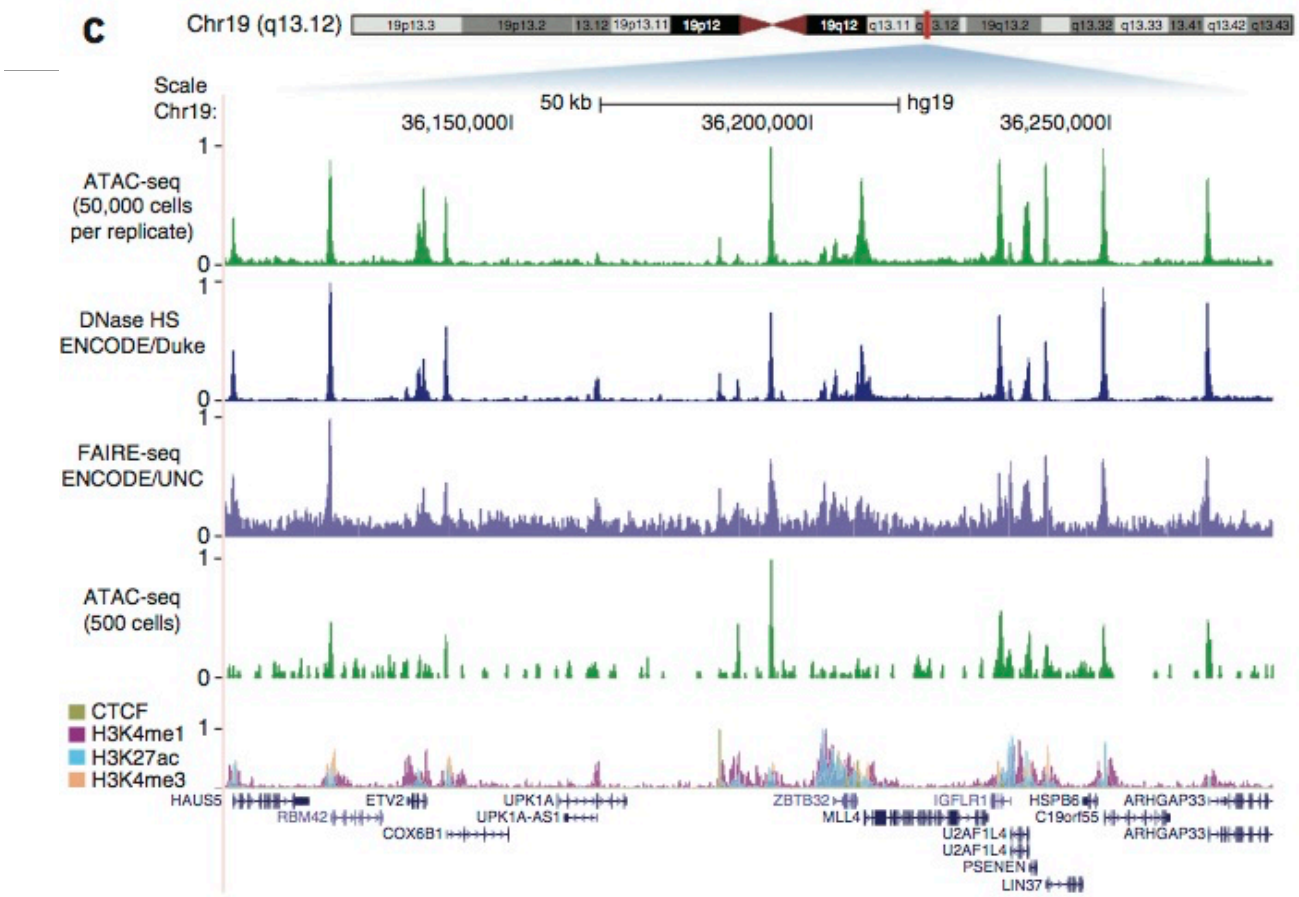

a Enzymatic cleavage footprinting
Nuclear isolation
MNase-seq
DNase-seq
ATAC-seq
MNase digestion
TF
Endonuclease activity of MNase
Exonuclease activity of MNase
DNA purification
Library preparation and sequencing
Sequencing adapter
DNase I digestion
Endonuclease activity of DNase I
DNA purification
Sucrose gradient centrifugation
Single-hit (long) fragments
Double-hit (short) fragments
Library preparation and sequencing of double-hit fragments
Tn5 tagmentation
Tn5 transposome
DNA purification
PCR amplification to add barcodes and sequencing

Zentner GE, Henikoff S. High-resolution digital profiling of the epigenome. Nat Rev Genet. 2014;15: 814–827. doi:10.1038/nrg3798

Transcription factor binding frequently deforms the B-form DNA, making it hypersensitive to DNase I and transposases.

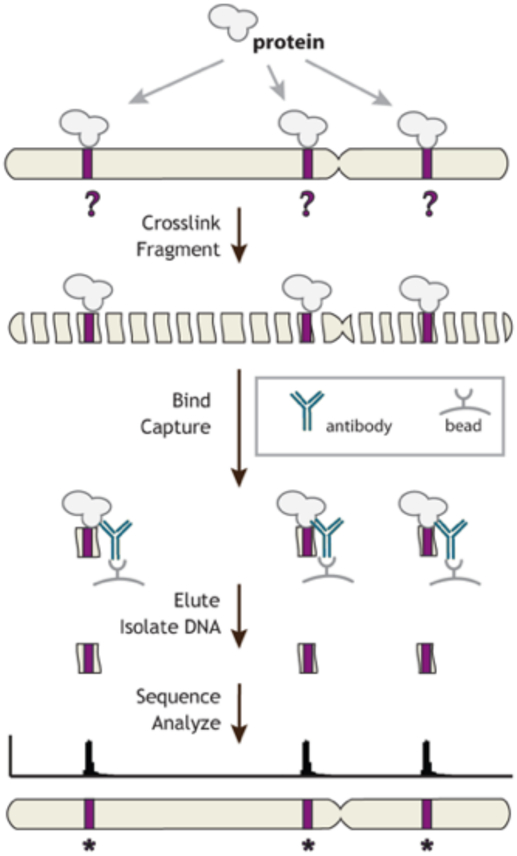
Changes in **accessibility of chromatin** can provide information about regulation

-ATAC-seq (shown)
-MNase-Seq (shown).
-DNase-Seq (shown).
-FAIRE-Seq (not shown).



Buenrostro JD, Giresi PG, Zaba LC, Chang HY, and Greenleaf WJ. (2013) "Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position." *Nature Methods*

Localization of *specific proteins* in the genome with chromatin immunoprecipitation (ChIP-Seq)



1. **Crosslink** the cells with formaldehyde to “fix” factors in place.

   Exception: Native ChIP with histone antibodies.

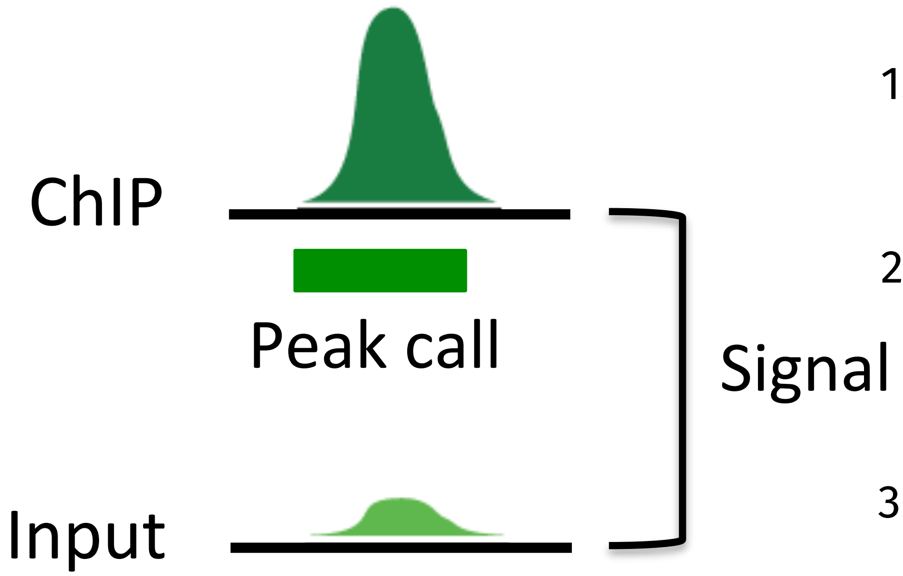
2. **Shear chromatin** to smaller pieces.

   Shear size determines resolution.
   Note: ChIP-exo uses an exonuclease at a later step to increase resolution.

3. **Enrich** target using an antibody.

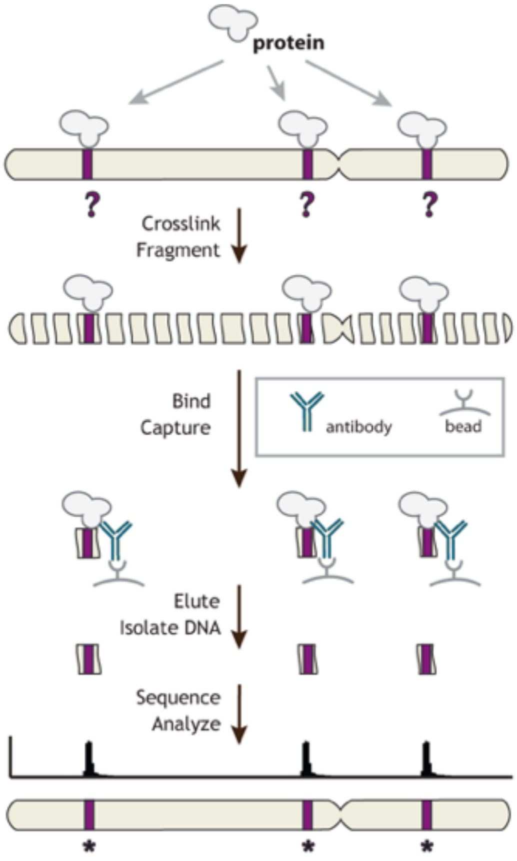
   Enrichment is only as good as the antibody.

Determining sites of enrichment from ChIP-Seq



1. **Align** reads to the genome.
2. **Compare to input** to look for enrichment.
   Input coverage is not even.
3. **Call peaks** to determine statistically significant sites of enrichment.

Limitations of ChIP-Seq



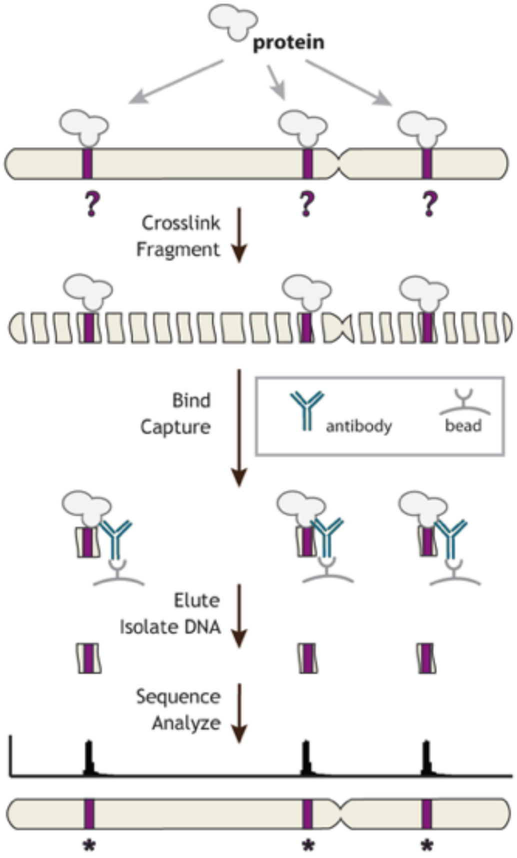
1. **Cross linking** efficiency is not necessarily uniform.

2. Enrichment is dependent on the **quality of antibody.**
   e.g., Site and degree of histone modifications.

3. Enrichment is dependent on the **accessibility of the epitope.**
   Comparing different sites to each other in the genome can be problematic.

4. Output is **descriptive**.
   Hard to infer function without more experimentation.

Extensions of ChIP



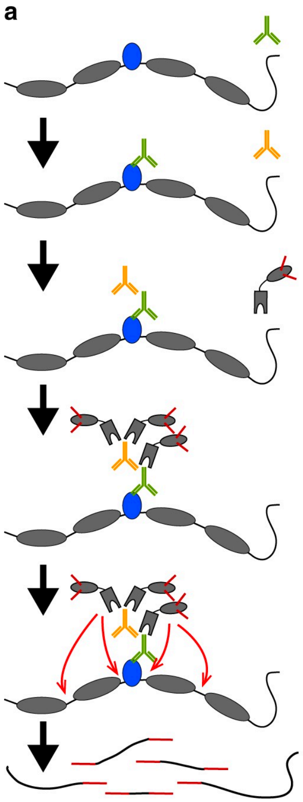
1. Using a nuclease to achieve **higher resolution** (ChIP-exo).

2. Make more quantitative using **spike-in normalization**.

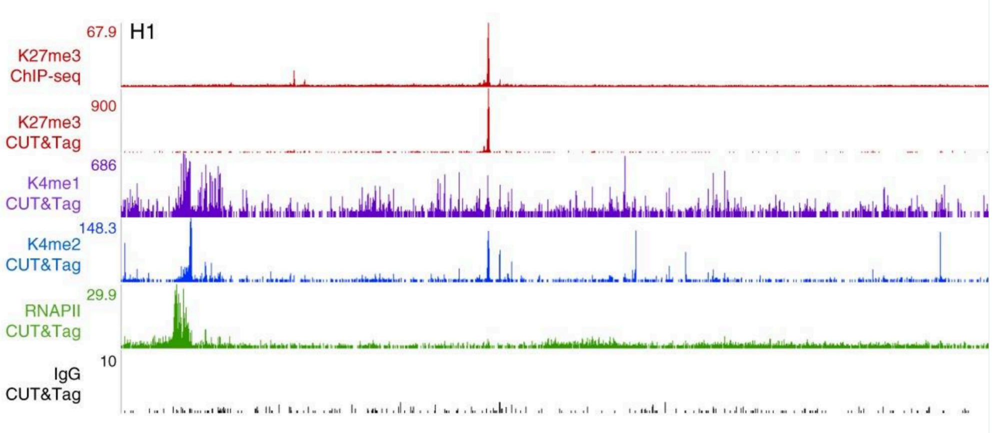
3. Analysis of **small samples or single cells** (CUT&RUN or CUT&Tag).

4. Extension to **RNA factors**.

CUT&Tag



Concept: Use factor-specific antibodies to target a transposes to direct the addition of DNA tags.



Kaya-Okur…& Henikoff (2019) CUT&Tag for efficient epigenomic profiling of small samples and single cells. *Nat Commun*

All of the following are advantages of CUT & Tag EXCEPT:

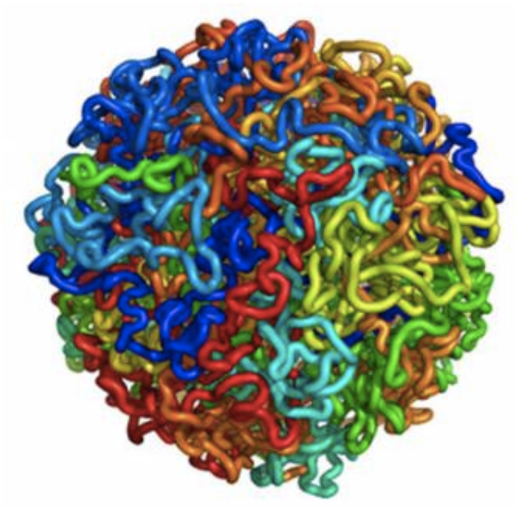
(A) Using a transpose simplifies library preparation.

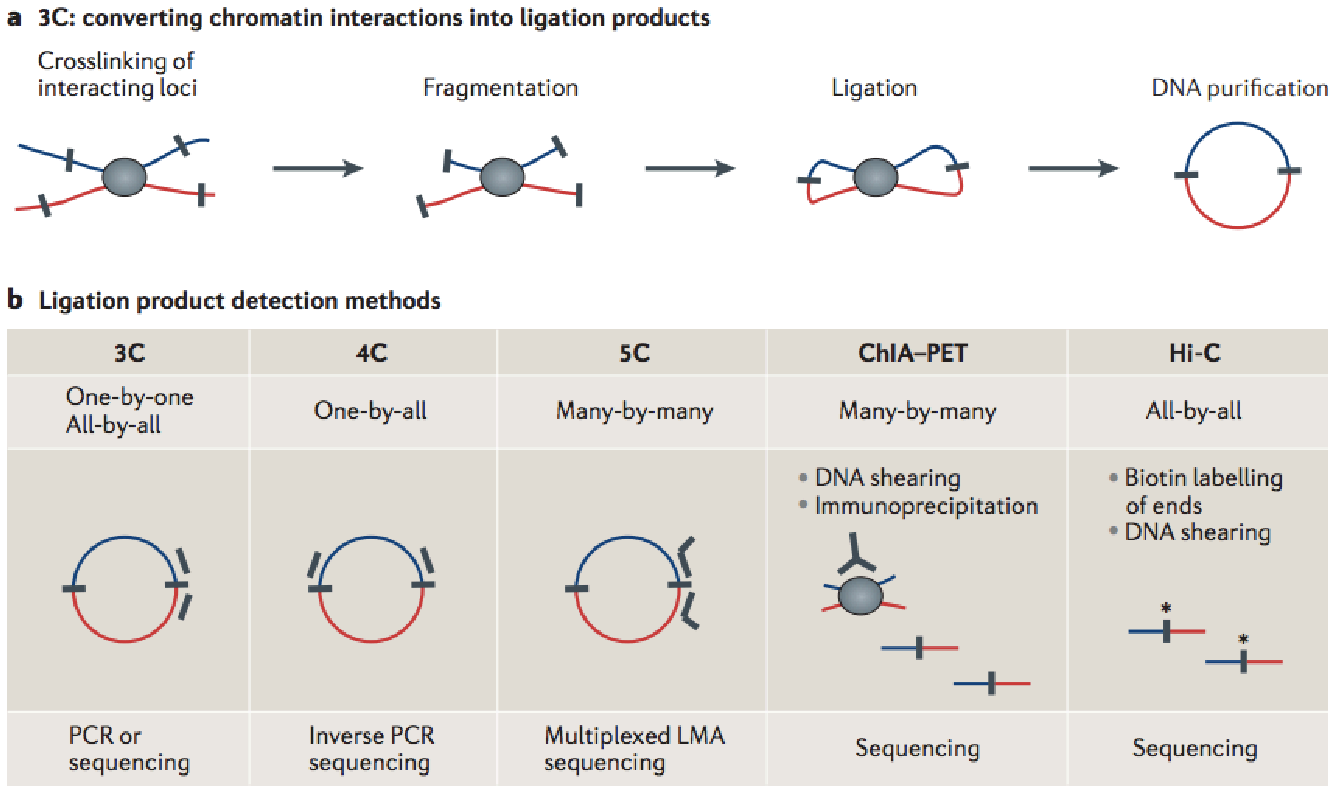
(B) CUT & Tag allows analysis of endogenous proteins without high affinity antibodies.

(C) Can be performed with very few cells.

(D) Avoids artifacts from chromatin sheering.

Mapping genome folding (and rearrangements)





a 3C: converting chromatin interactions into ligation products

b Ligation product detection methods

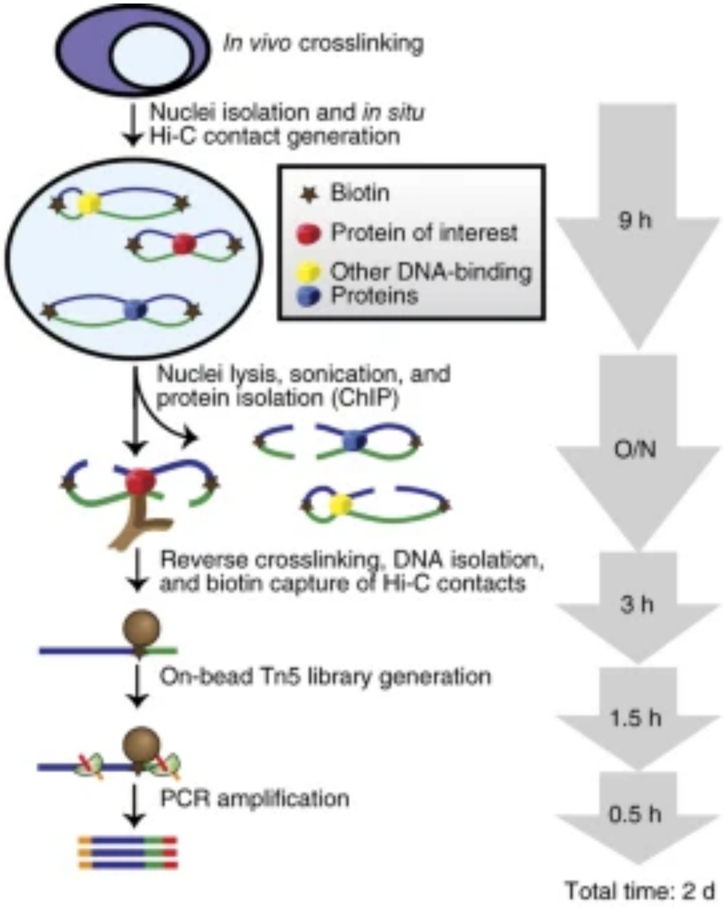
3C	4C	5C	ChIA–PET	Hi-C
One-by-one All-by-all	One-by-all	Many-by-many	Many-by-many	All-by-all
			• DNA shearing • Immunoprecipitation	• Biotin labelling of ends • DNA shearing
PCR or sequencing	Inverse PCR sequencing	Multiplexed LMA sequencing	Sequencing	Sequencing

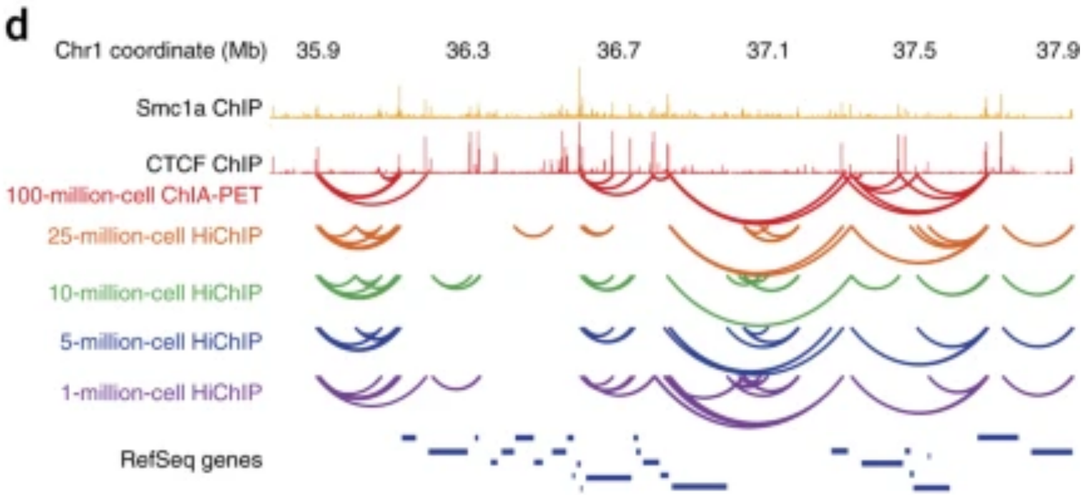
Dekker, J., Marti-Renom, M. A. & Mirny, L. A.. Nat Rev Genet 14, 390–403 (2013).

Nine rules of thumb about sequencing methodology

1. **Global approaches** can be (mostly) comprehensive, but require more sequencing.
2. **Targeted approaches** can provide better coverage of features of interest but require prior information.
3. **Biochemical enrichment** provides many opportunities, but generally requires more starting material.
4. **Enzymes** can often provide more sensitive approaches to target specific types of nucleic acids, but are limited by the efficiency and specificity of the enzymes that are available.
5. Single nucleotide information can be revealed through **mutations** or the **location of the end** of a read.
6. **Chemistry** can often be used to reveal latent information in a sequencing experiment.
7. Unique molecular identifiers (**UMIs**) can provide **additional information** about each read.
8. Many sequencing methods are **modular** and can be combined with one another.
9. Each **sequencing platform** has unique advantages and challenges.

Example of combined methods: Hi-ChIP

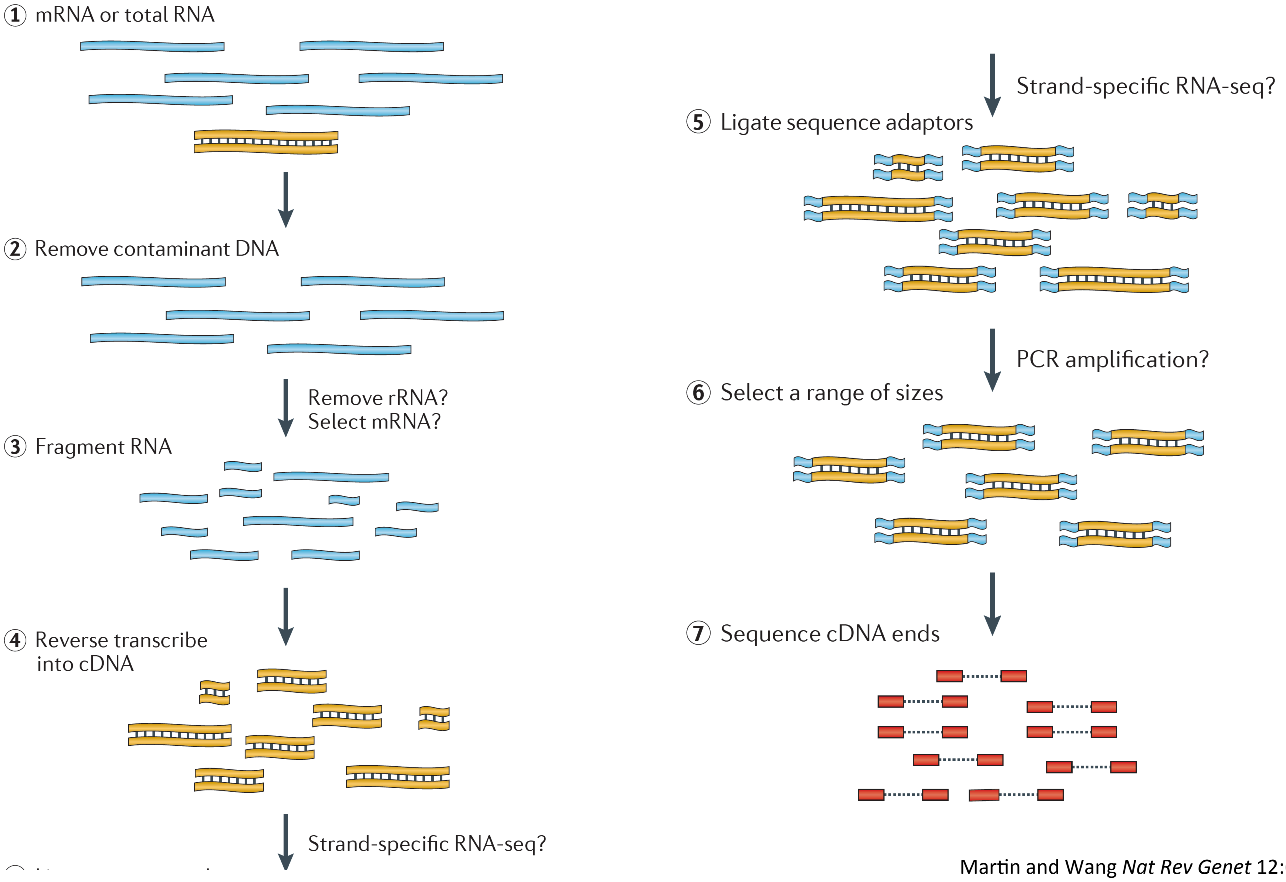

In vivo crosslinking
Nuclei isolation and in situ Hi-C contact generation
Biotin
Protein of interest
Other DNA-binding Proteins
Nuclei lysis, sonication, and protein isolation (ChIP)
Reverse crosslinking, DNA isolation, and biotin capture of Hi-C contacts
On-bead Tn5 library generation
PCR amplification
9 h
O/N
3 h
1.5 h
0.5 h
Total time: 2 d



Mumbach MR, et al. HiChIP: efficient and sensitive analysis of protein-directed genome architecture. Nat Methods.

Part 2: RNA-Seq and applications of RNA-Seq

Example of an RNA-Seq workflow



① mRNA or total RNA

② Remove contaminant DNA

Remove rRNA?
Select mRNA?

③ Fragment RNA

④ Reverse transcribe
into cDNA

Strand-specific RNA-seq?

Strand-specific RNA-seq?

⑤ Ligate sequence adaptors

PCR amplification?

⑥ Select a range of sizes

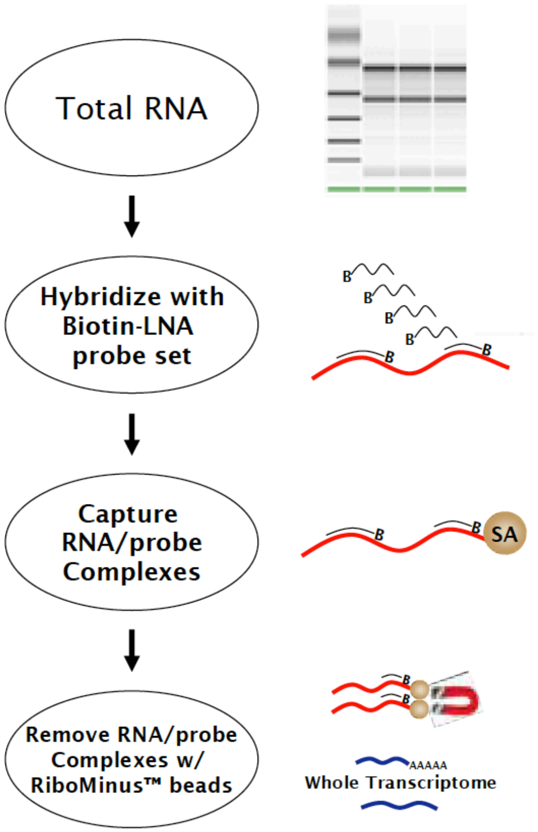
⑦ Sequence cDNA ends

Martin and Wang *Nat Rev Genet* 12:671 (2011)

How is RNA-Seq different from standard DNA-seq?

- Wide dynamic range of RNA concentrations.
- RNA is strand specific (unlike dsDNA)
- RNA degrades easily (RNase and spontaneous chemical hydrolysis)
- RNA is processed (e.g., capped, spliced, polyA)
- RNA can have modifications that can block RT or be invisible (e.g., tRNAs).
- There are a wide range of sizes or RNAs and specialized protocols are necessary for studying shorter RNAs (e.g., miRNA, short capped RNAs)
- RNA has secondary structure (possible blocks to reverse transcriptase).

Ribosomal RNA will dominate the sequenced reads unless removed or avoided



Total RNA

Hybridize with Biotin-LNA probe set

Capture RNA/probe Complexes

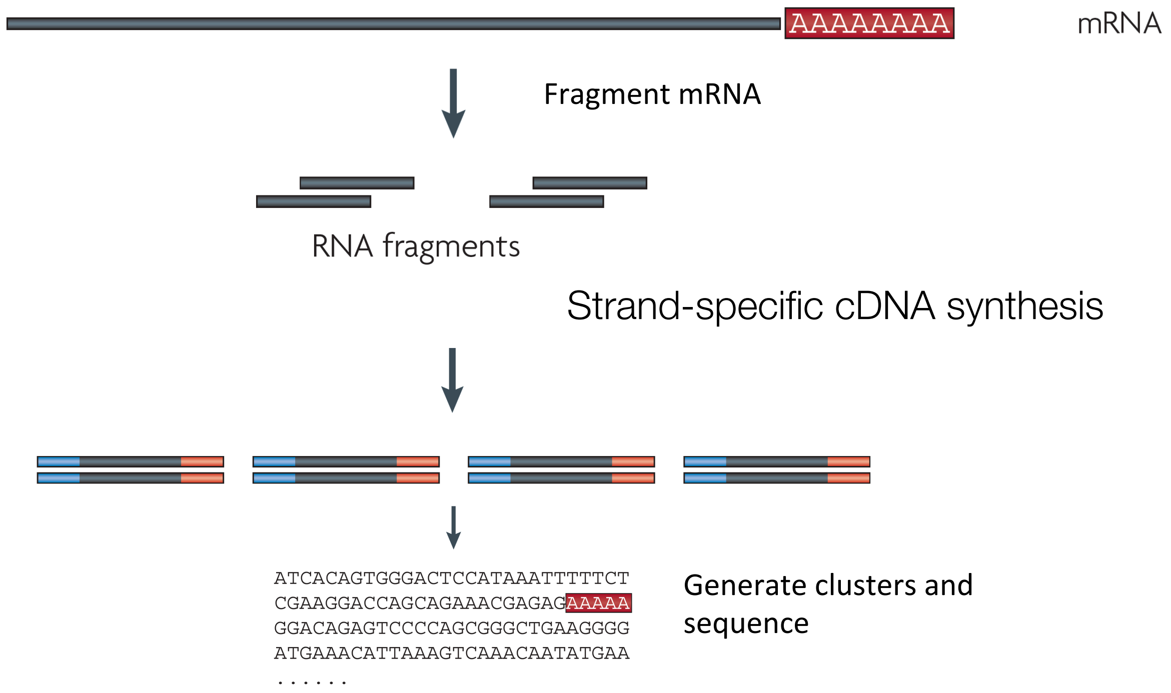
Remove RNA/probe Complexes w/ RiboMinus™ beads

RiboMinus

polyA-based RNA-seq workflow

Capture poly-A RNA with poly-T oligo attached beads (100 ng total) (2x)

- RNA quality must be high – degradation produces 3’ bias
- Non-poly-A RNAs are not recovered



AAAAAAAA

mRNA

Fragment mRNA

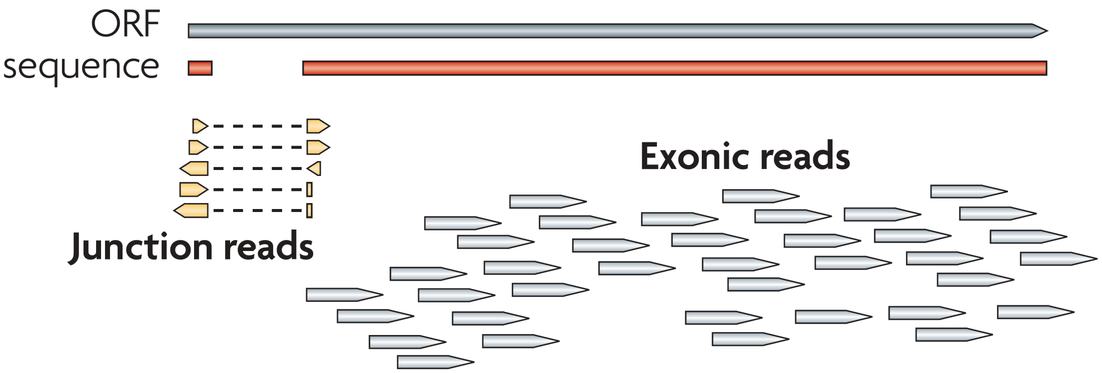
RNA fragments

Strand-specific cDNA synthesis

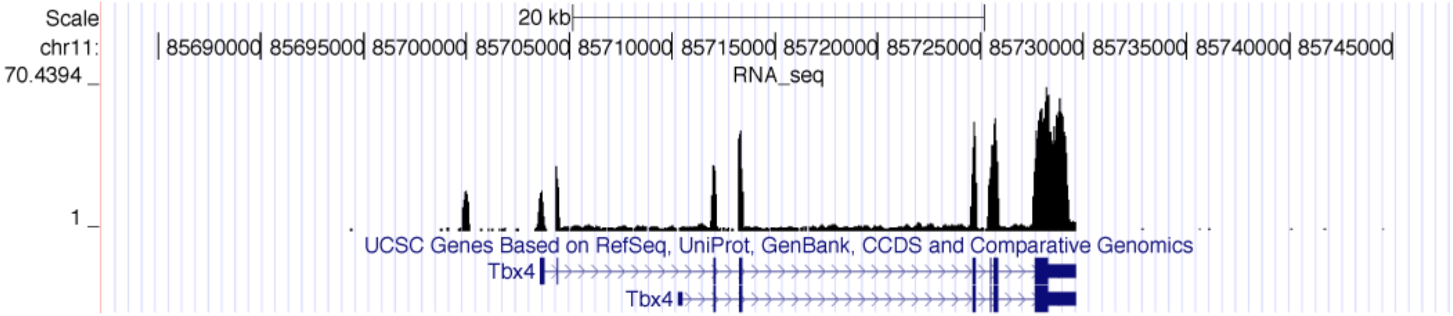
```
ATCACAGTGGGACTCCATAAATTTTTCT
CGAAGGACCAGCAGAAACGAGAGAAAAA
GGACAGAGTCCCCAGCGGGCTGAAGGGG
ATGAAACATTAAAGTCAAACAATATGAA
......
```

Generate clusters and sequence

RNA-Seq reads map mostly to exons



Martin and Wang *Nat Rev Genet* 12:671 (2011)



How does one analyze RNA levels from RNA-Seq?

Use existing gene annotation:

Align to genome plus annotated splices
Depends on high-quality gene annotation
Which annotation to use: RefSeq, GENCODE, UCSC?
Isoform quantification?
Identifying novel transcripts?

Reference-guided alignments:

Align to genome sequence
Infer splice events from reads
Allows transcriptome analyses of genomes with poor gene annotation

De novo transcript assembly:

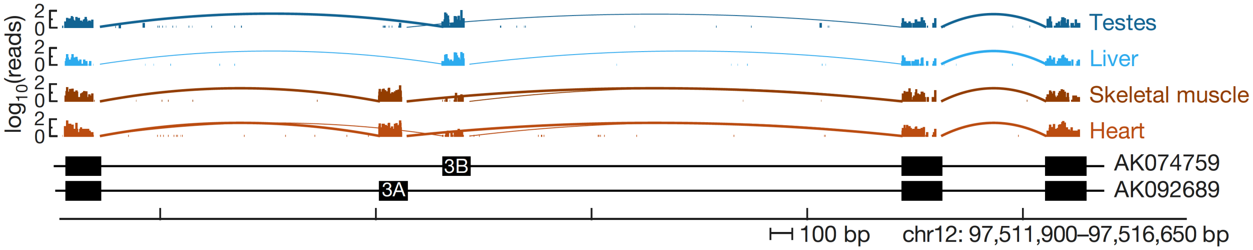
Assemble transcripts directly from reads
Allows transcriptome analyses of species without reference genomes

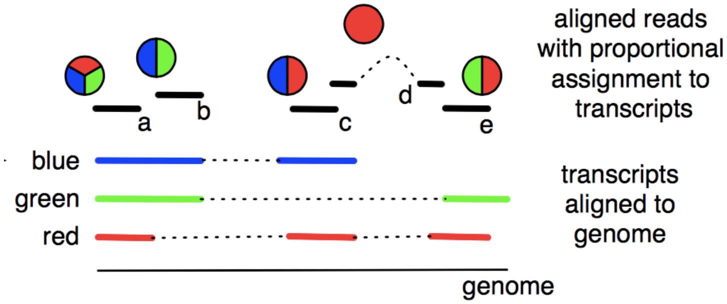
RNA-seq reads contain information about the abundance of different transcript isoforms

Normalization :

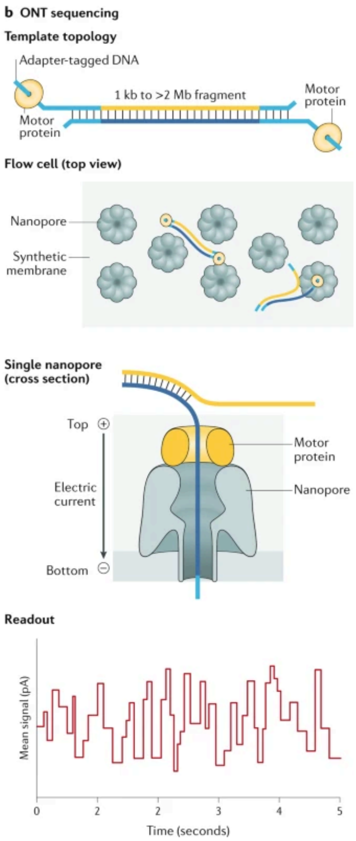
Internal: *Reads or Fragments* per kilobase of feature length per million mapped reads (RPKM or FPKM)

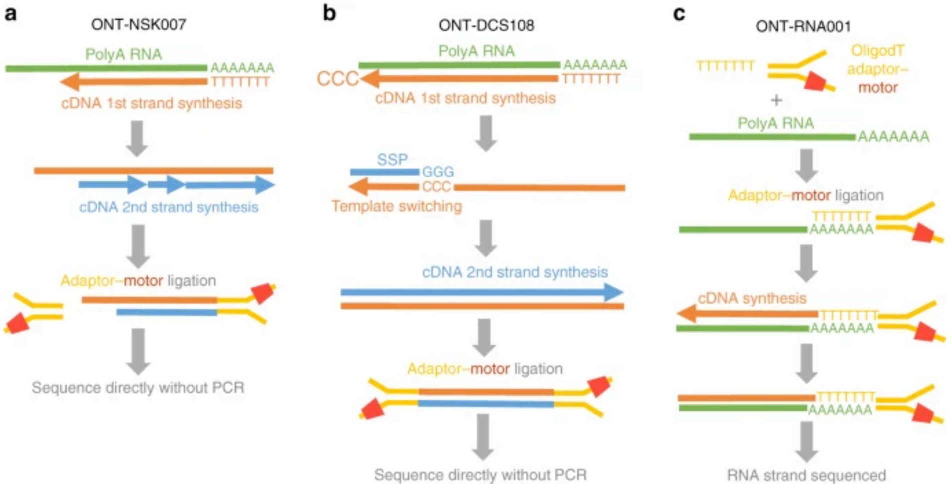
External: Reads relative to a standard “spike”




aligned reads with proportional assignment to transcripts
a
b
c
d
e
blue
green
red
transcripts aligned to genome
genome

Direct RNA sequencing using ONT

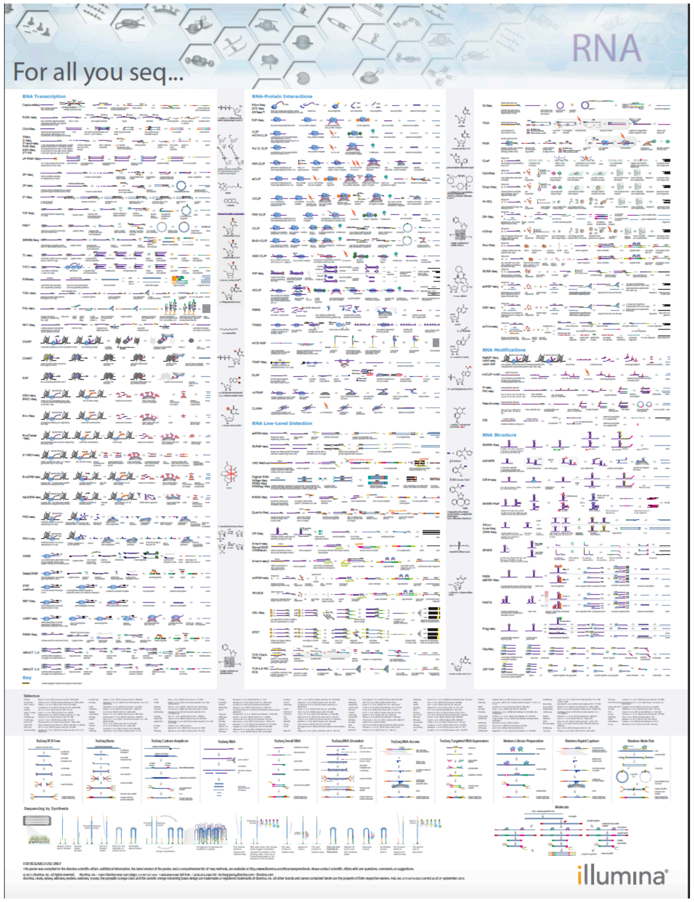




Soneson C, et al. Nat Commun. 2019 Jul 31;10(1):3359. doi: 10.1038/s41467-019-11272-z. PMID: 31366910; PMCID: PMC6668388.

Long reads identify each RNA transcript isoform.
Potential to identify RNA modifications directly.
Can measure the length of RNA polyA tail.
In principle avoids challenges/biases of library preparation.

Examples of applications of RNA-seq



- Characterizing transcriptome complexity
  - Alternative splicing
  - RNA modifications
  - RNA structures
- Differential expression analysis
  - Gene- and isoform-level expression comparisons
- Novel RNA species
  - lncRNAs and eRNAs
  - Pervasive transcription
- Translation
  - Ribosome profiling
- Allele-specific expression
- Measuring RNA half-lives and decay
- Examining protein-RNA interactions
- Effect of genetic variation on gene expression
  - Imprinting
  - RNA editing
  - Novel events

Nine rules of thumb about sequencing methodology

1. **Global approaches** can be (mostly) comprehensive, but require more sequencing.
2. **Targeted approaches** can provide better coverage of features of interest but require prior information.
3. **Biochemical enrichment** provides many opportunities, but generally requires more starting material.
4. **Enzymes** can often provide more sensitive approaches to target specific types of nucleic acids, but are limited by the efficiency and specificity of the enzymes that are available.
5. Single nucleotide information can be revealed through **mutations** or the **location of the end** of a read.
6. **Chemistry** can often be used to reveal latent information in a sequencing experiment.
7. Unique molecular identifiers (**UMIs**) can provide **additional information** about each read.
8. Many sequencing methods are **modular** and can be combined with one another.
9. Each **sequencing platform** has unique advantages and challenges.

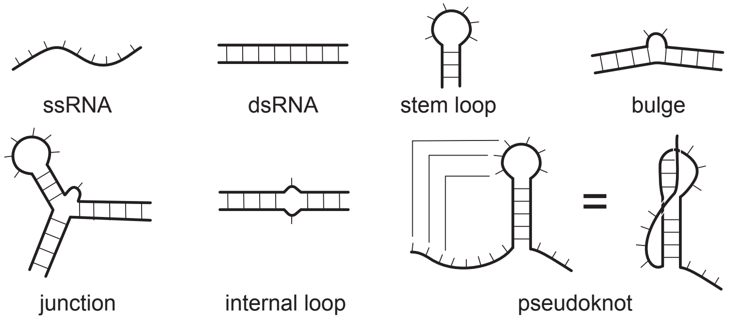
Examples of how to target sub-populations of RNAs

1. Deplete unwanted RNAs
   A. Ribo-minus etc. to remove rRNA.
   B. Enzymatic removal using targeted enzymes (e.g., RNase H, Cas9)
   C. Globally degrade unwanted RNAs (e.g., uncapped RNAs with a 5'-to-3' exonuclease.
2. Enrich/amplify specific transcripts.
   D. Use targeted RT primers.
   E. Biochemically enrich RNAs/cDNAs using capture hybridization.
   F. Use knowledge of 5' and 3' modifications (e.g., miRNA with 5'-phosphate and 3'-hydroxyl)
3. Select newly made RNAs
   G. Fractionate chromatin-associated RNAs.
   H. Only consider intron-containing RNAs.
   I. Immunoprecipitate RNA PolII engaged RNAs.
   J. Metabolic labeling with short pulses.
4. Select modified RNAs
   K. Immunoprecipitate with a modification-specific antibody.
   L. Use chemistries that induce RT-stops or mutations.
5. Select RNAs from specific cells.
   M. Microdissect cells or FACS sort cells of interest.
   N. TU-tagging (targeted metabolic labeling of RNAs in certain cells)
   O. Single-cell RNA-seq (scRNA-seq)

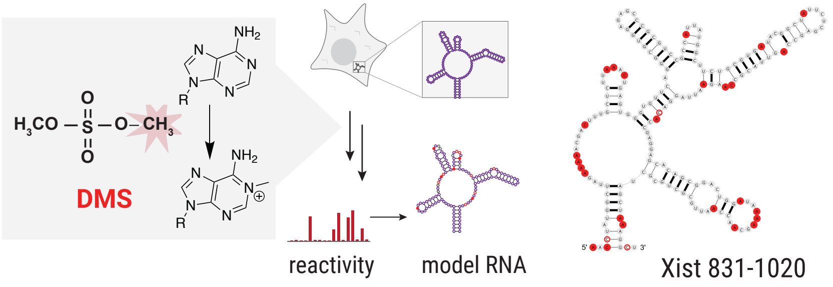
Biochemical

Computational

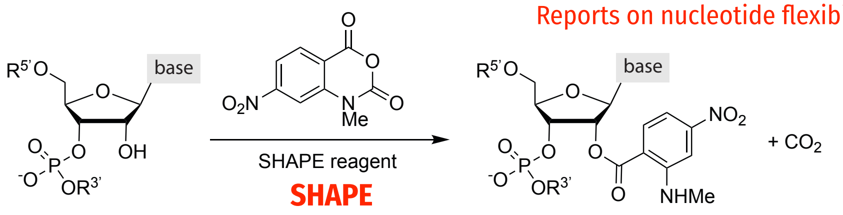
Examining RNA structure with chemical probing



Reports on A/C base accessibility



Reports on nucleotide flexibility



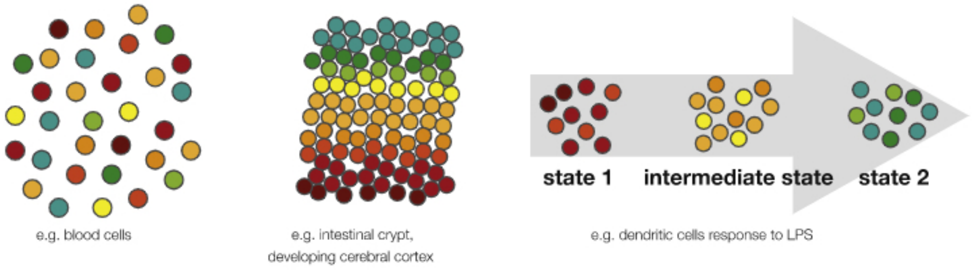
RNA can fold into elaborate structures.

Accessible nucleotides (e.g., those in ssRNA) are often more reactive than base-paired nucleotides to chemical reagents.

Chemical modifications cause reverse transcription termination or mutations that can be read out using sequencing.

Reviewed in: Strobel, E. J., Yu, A. M. & Lucks, J. B. High-throughput determination of RNA structures. *Nat Rev Genet* **19**, 615-634 (2018).

Examining cell heterogeneity with scRNA-seq

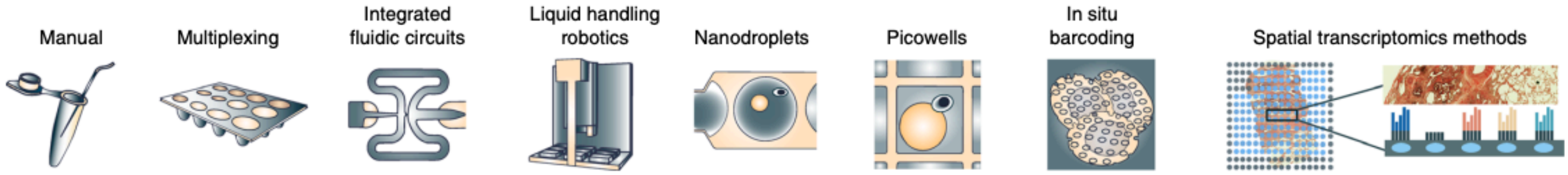

state 1
intermediate state
state 2
e.g. blood cells
e.g. intestinal crypt, developing cerebral cortex
e.g. dendritic cells response to LPS

Kolodziejczyk … & Teichmann (2015). Mol Cell

Bulk RNA-seq averages over the RNA content of many cells masking differences.

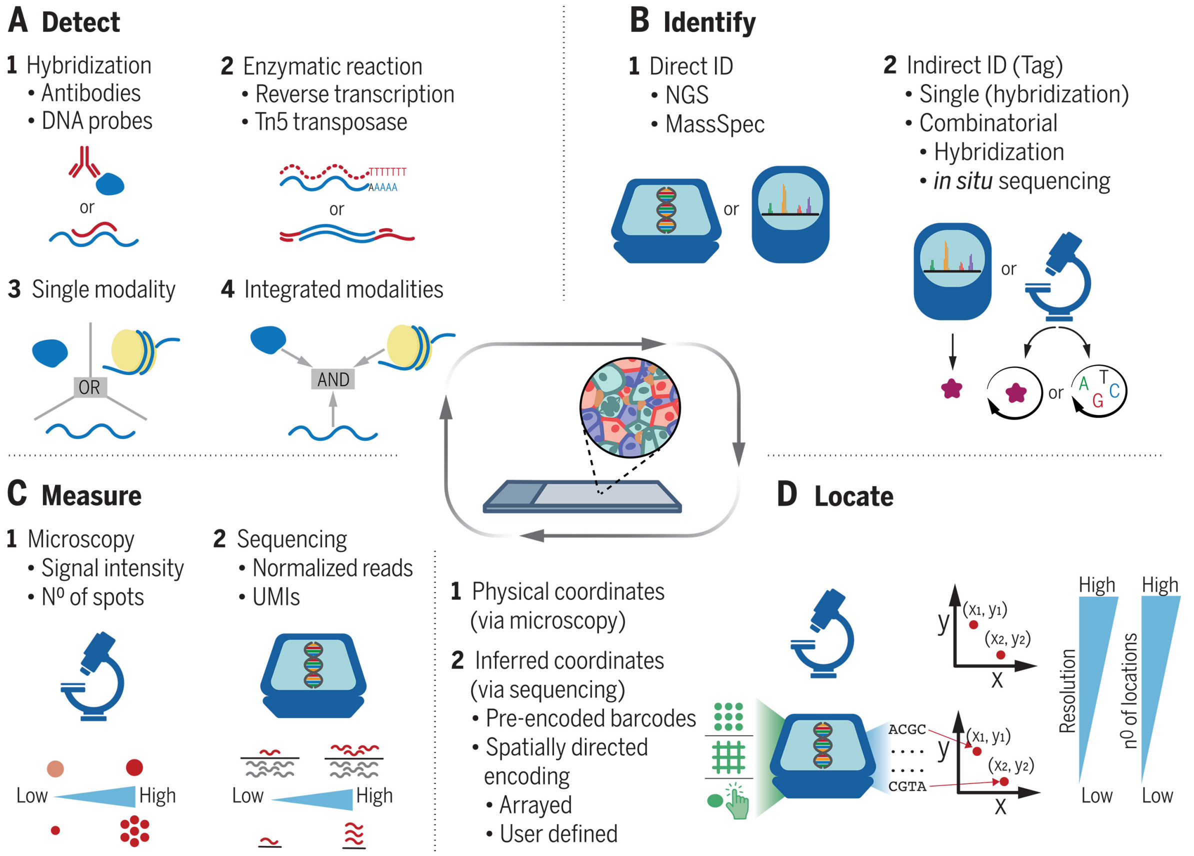
These differences can be revealed by sequencing the RNA from individual cells using single cell RNA-seq (scRNA-seq)

Analysis of RNA transcripts in individual cells can reveal rare cell populations and lineage trajectories.



Aldridge S, Teichmann SA. Single cell transcriptomics comes of age. Nat Commun. 2020 Aug 27;11(1):4307. doi: 10.1038/s41467-020-18158-5. PMID: 32855414; PMCID: PMC7453005.

Overview of spatial sequencing methods



A Detect

1 Hybridization
- Antibodies
- DNA probes

2 Enzymatic reaction
- Reverse transcription
- Tn5 transposase

3 Single modality

4 Integrated modalities

B Identify

1 Direct ID
- NGS
- MassSpec

2 Indirect ID (Tag)
- Single (hybridization)
- Combinatorial
  - Hybridization
  - *in situ* sequencing

C Measure

1 Microscopy
- Signal intensity
- Nº of spots

2 Sequencing
- Normalized reads
- UMIs

D Locate

1 Physical coordinates (via microscopy)

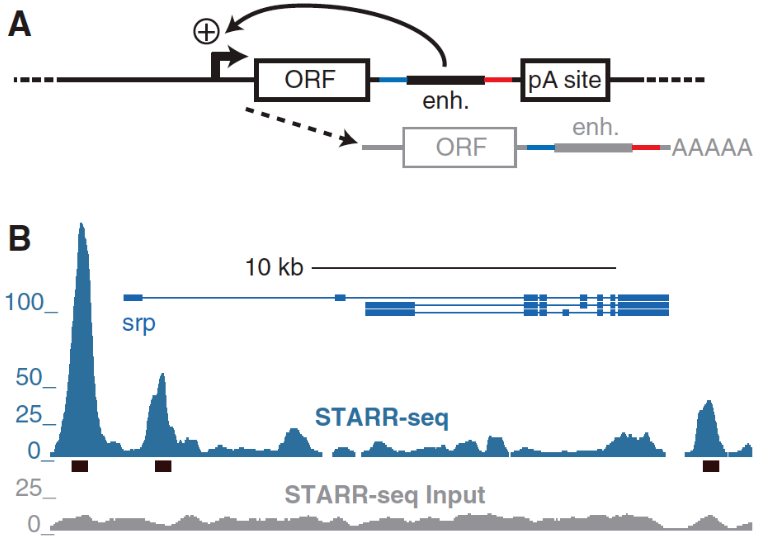
2 Inferred coordinates (via sequencing)
- Pre-encoded barcodes
- Spatially directed encoding
  - Arrayed
  - User defined

Dario Bressan et al. ,The dawn of spatial omics.Science381,eabq4964(2023).DOI:10.1126/science.abq4964

Almost any assay can be adapted to sequencing!

- CRISPR screens
- Massively parallel reporter assays (MPRA)

Fig. 1. STARR-seq genome-wide quantitative enhancer discovery. (**A**) STARR-seq reporter setup [enh., enhancer candidate; ORF, open-reading frame (here: GFP); pA site, polyadenylation site; +, transcriptional activation]. (**B**) STARR-seq (blue) and input (gray) fragment densities in the *srp* locus. Black boxes denote predicted enhancers ("peaks"). (**C**) STARR-seq and luciferase signals are linearly correlated: R^2, coefficient of determination and Pearson correlation coefficient (PCC



Arnold CD, Gerlach D, Stelzer C, Boryń ŁM, Rath M, Stark A. Science. 2013 Mar 1;339(6123):1074-7. PMID: 23328393.

Summary

- Genomics I: Deep sequencing gives us access to information on a genomic level.

- Genomics II: These approaches provide a diverse set of tools to study life at a genomic scale.

✴ Sophisticated use of data from genomics requires an integrated understanding of the biological experiment, sample preparation and down stream computational analyses of the data.