What is…

ChIP-SEQ
ATAC-SEQ
BISULFITE-SEQ
HiC-SEQ

RNA-SEQ
smRNA-SEQ
PRO-SEQ
CLIP-SEQ

# "SURVEY OF GENOMIC TECHNOLOGIES FOR GENE REGULATION"

Jen Grenier Director, TREx Facility

# Announcements

- **New and Improved Project Submission Form**

  Available on our web site

- **New service: ATACseq**

  **A**ssay for **T**ransposase-**A**ccessible **C**hromatin by sequencing

  Identify promoters, enhancers, motifs enriched in open chromatin
  expressed genes, ‘poised’ genes (vs RNAseq)

  Researcher provides intact nuclei (preserving native state)

  Soft launch in January

  Interested? Contact us at *trex_info@cornell.edu*,
  or come to our next Tech Talk for more information!

# What is Next-Generation Sequencing?

**Massively parallel, high-throughput DNA sequencing**

Spatially separated, clonally amplified DNA templates on a flow cell

Illumina platform:

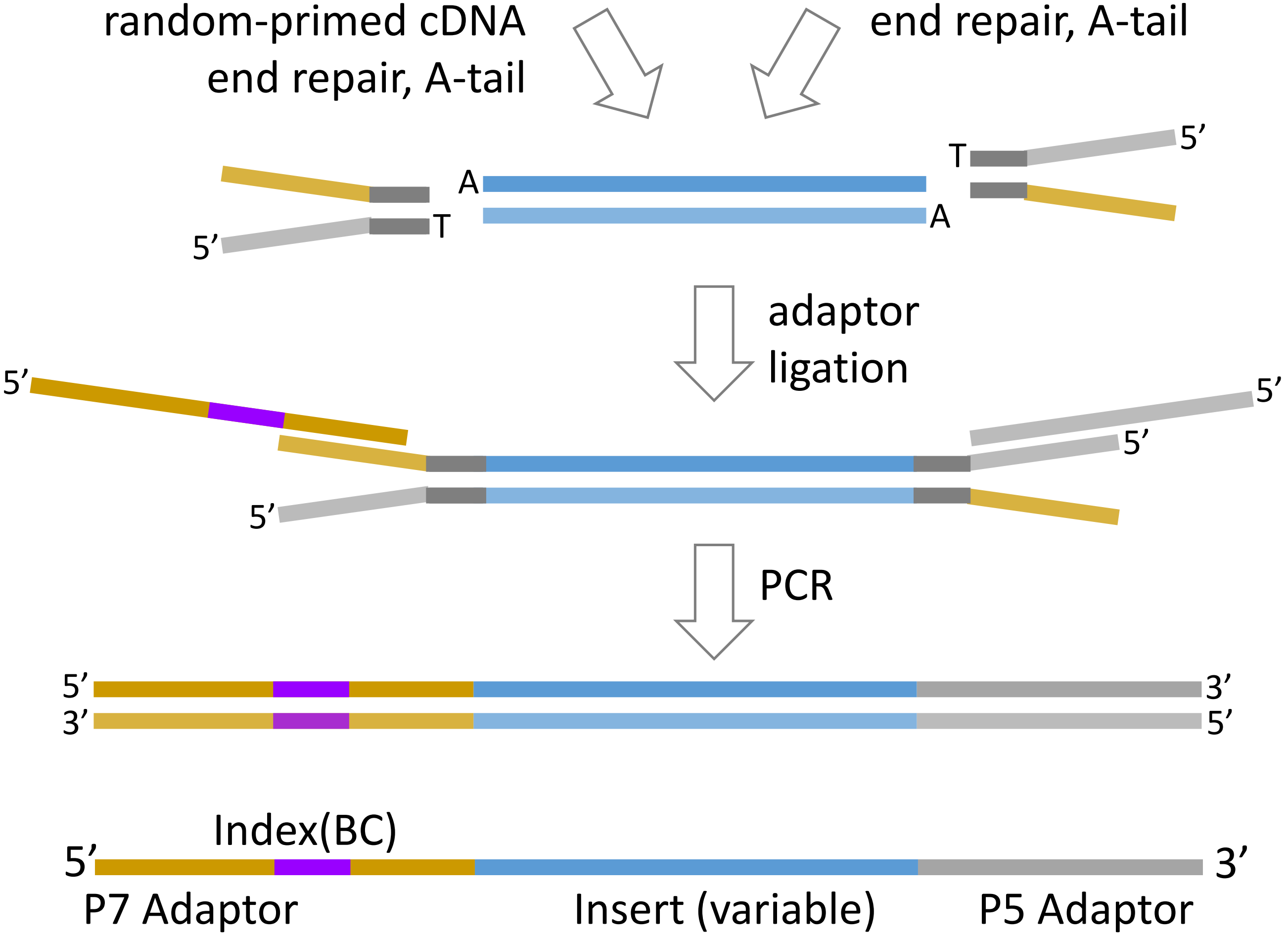
- library is captured by probes on the surface of the flow cell
- captured molecules form colonies with bridge amplification
- sequencing by synthesis generates fluorescent signal
- camera/optics reads signal for each base, each cycle (base)
- software converts images into text file (fastq format)
- up to 4 reads (with different primers) per cluster, per run

Illumina video

# What is an Illumina library?

**RNAseq**
mRNA isolation, fragmentation
random-primed cDNA
end repair, A-tail

**WGS**
gDNA fragmentation
end repair, A-tail



adaptor ligation

PCR

Index(BC)

P7 Adaptor

Insert (variable)

P5 Adaptor

# What is Gene Regulation?

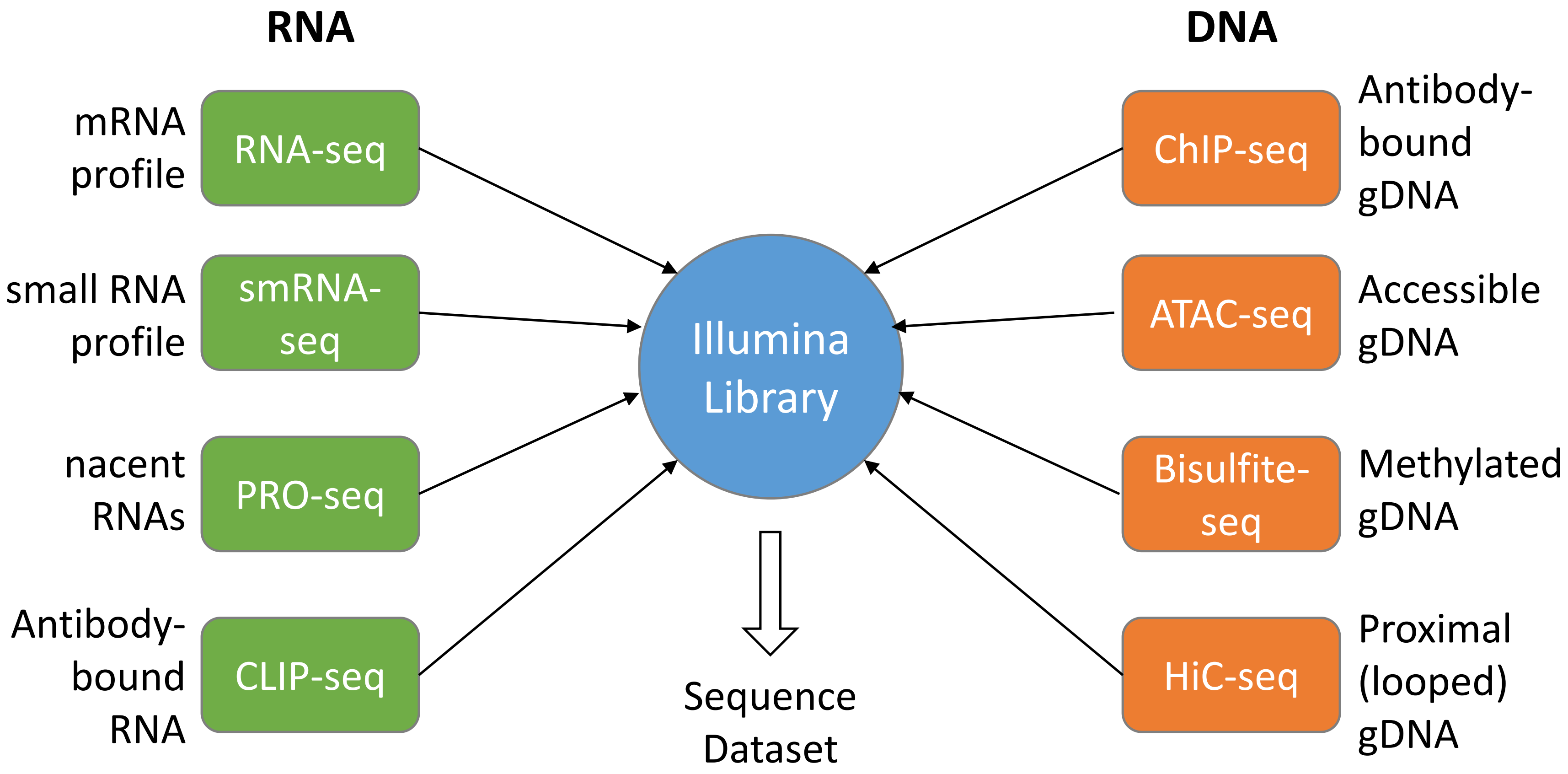
**A wide range of mechanisms that control the production of specific gene products**

- What genes are expressed under different conditions?
  expression profiling

- How is transcription regulated?
  chromatin state, transcription factor occupancy, DNA methylation

- What about post-transcriptional regulation?
  RNA binding proteins, microRNA regulation

# How can we use NGS to study gene regulation?



**RNA**

mRNA profile — RNA-seq

small RNA profile — smRNA-seq

nacent RNAs — PRO-seq

Antibody-bound RNA — CLIP-seq

**DNA**

ChIP-seq — Antibody-bound gDNA

ATAC-seq — Accessible gDNA

Bisulfite-seq — Methylated gDNA

HiC-seq — Proximal (looped) gDNA

Illumina Library

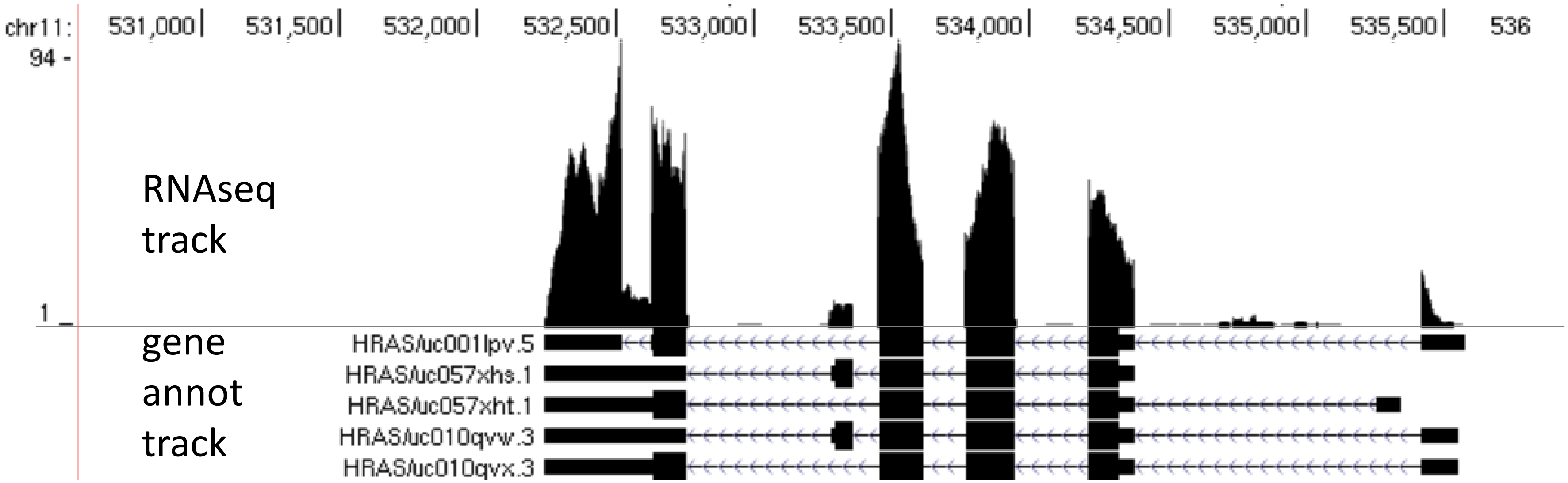
Sequence Dataset

# What is RNA-seq?

RNA
↓
Enrichment
↓
Fragmentation
↓
cDNA
↓
Adapter ligation
↓
PCR amplification

| | |
|---|---|
| **Applications**: | gene expression profiling<br>transcript annotation/assembly<br>*pathogen identification*<br>*variant discovery/identification* |
| **Input**: | total RNA |
| **Enrichment**: | polyA$^+$ or rRNA-depletion |
| **Reads**: | map to exons |
| **Analysis**: | differential gene expression<br>transcript assembly/annotation |
| **Variations:** | 3’ RNA-seq, targeted RNA-seq |

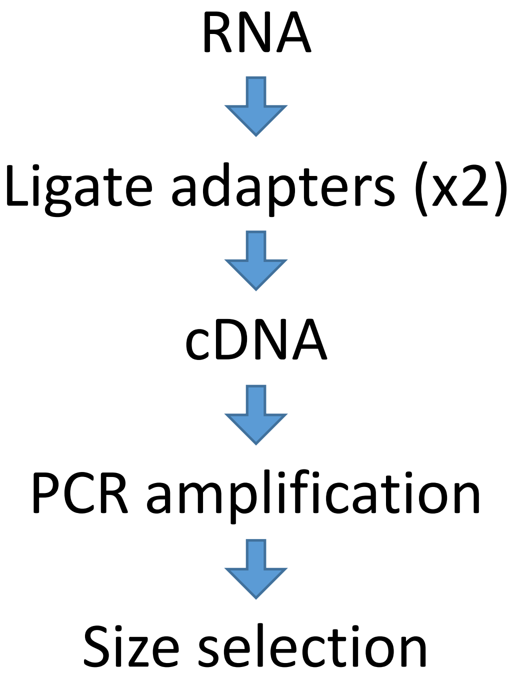
# What is RNA-seq?



RNAseq
track

gene
annot
track

# What is smRNA-seq?



RNA

Ligate adapters (x2)

cDNA

PCR amplification

Size selection

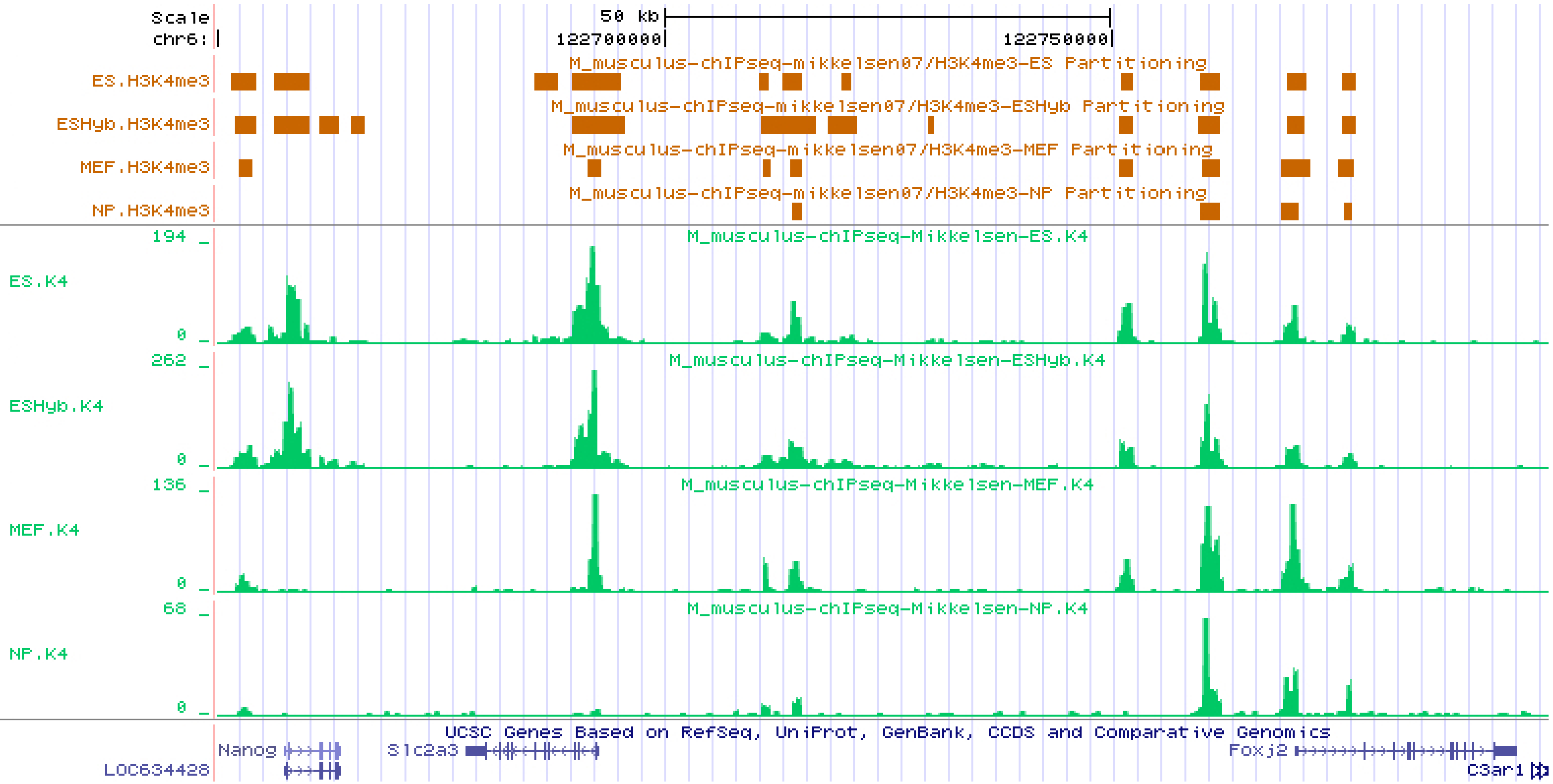
| | |
|---|---|
| **Applications**: | microRNA profiling |
| **Input**: | total RNA<br>cell-free RNA |
| **Enrichment**: | ligation to 5’-monoP, 3’-OH<br>post-PCR size selection |
| **Reads:** | map to mature miRNAs |
| **Analysis**: | differential microRNA counts<br>microRNA discovery, processing<br>biomarker ID/profiling |
| **Variations:** | poly-adenylation/RT<br>circularization |

# What is ChIP-seq?

Cells/tissue
↓
Crosslink DNA+Ptn
↓
Lyse/sonicate
↓
Immunoprecipitate
↓
Reverse crosslink
↓
Adapter ligation
↓
PCR amplification

| | |
|---|---|
| **Applications**: | chromatin mark distribution<br>transcription factor occupancy<br>DNA binding protein sites |
| **Input**: | cells/tissue (native gDNA) |
| **Enrichment**: | antibody immunoprecipitation |
| **Reads**: | map to small intervals (‘peaks’) |
| **Analysis**: | peak identification<br>differential peak representation<br>motif enrichment |
| **Variations:** | ChIP-exo, CUT&RUN, CUT&Tag |

# What is ChIP-seq?



Peak
Calls

Mapped
Reads

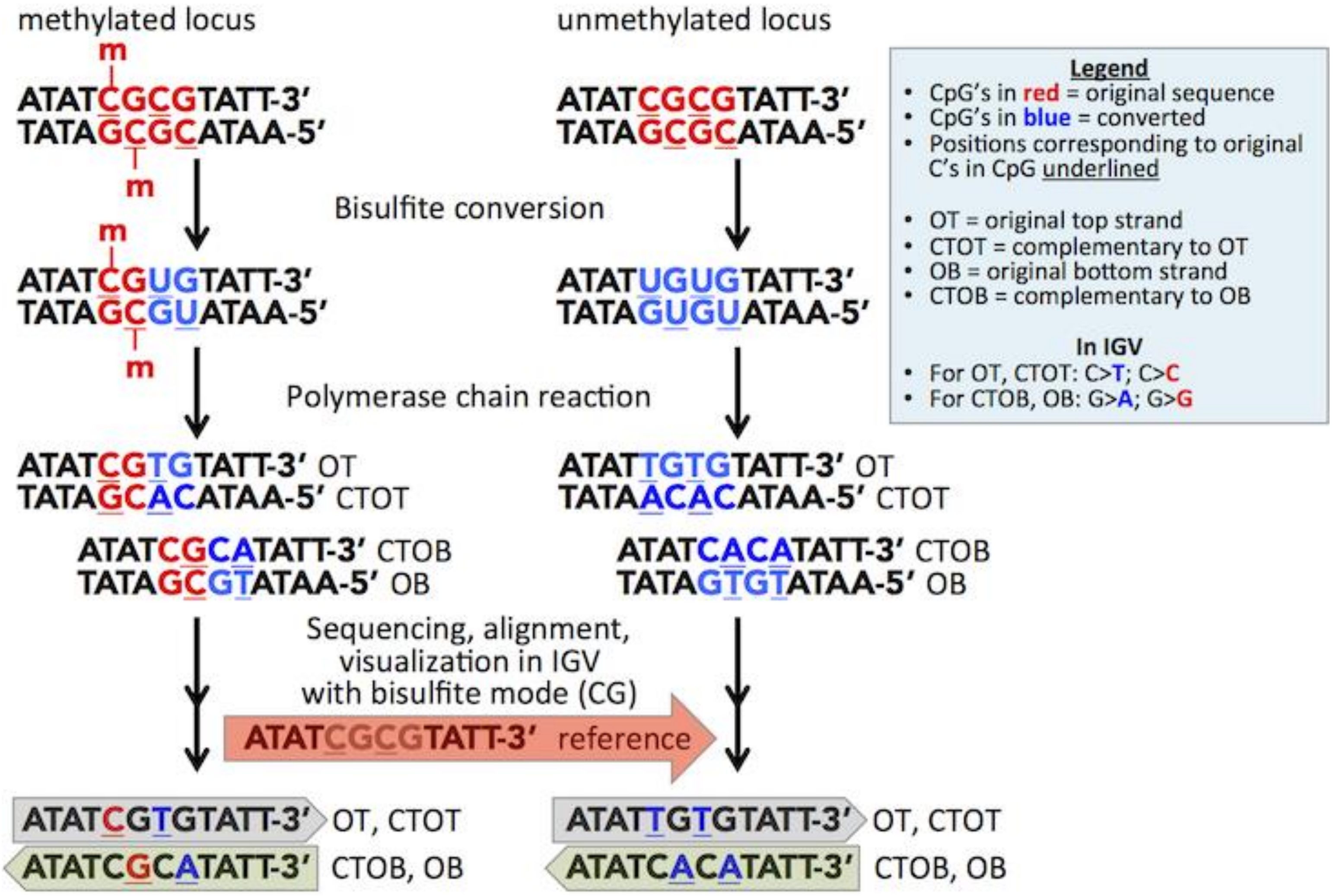
Gene
annot.

# What is Bisulfite-seq?

gDNA
↓
Fragment
↓
Me-Adapter ligation
↓
Bisulfite conversion
↓
PCR amplfication

| | |
|---|---|
| **Applications**: | identify methylated CpG |
| **Input**: | gDNA |
| **Enrichment**: | *optional (RRBS, targeted)* |
| **Reads**: | C (reference)→T (read)<br>indicates unmethylated-C |
| **Analysis**: | location, frequency of me-C<br>differential methylation |
| **Variations:** | RRBS, targeted, TAB-seq |

# What is Bisulfite-seq?



**Legend**

- CpG's in **red** = original sequence
- CpG's in **blue** = converted
- Positions corresponding to original C's in CpG underlined

- OT = original top strand
- CTOT = complementary to OT
- OB = original bottom strand
- CTOB = complementary to OB

**In IGV**

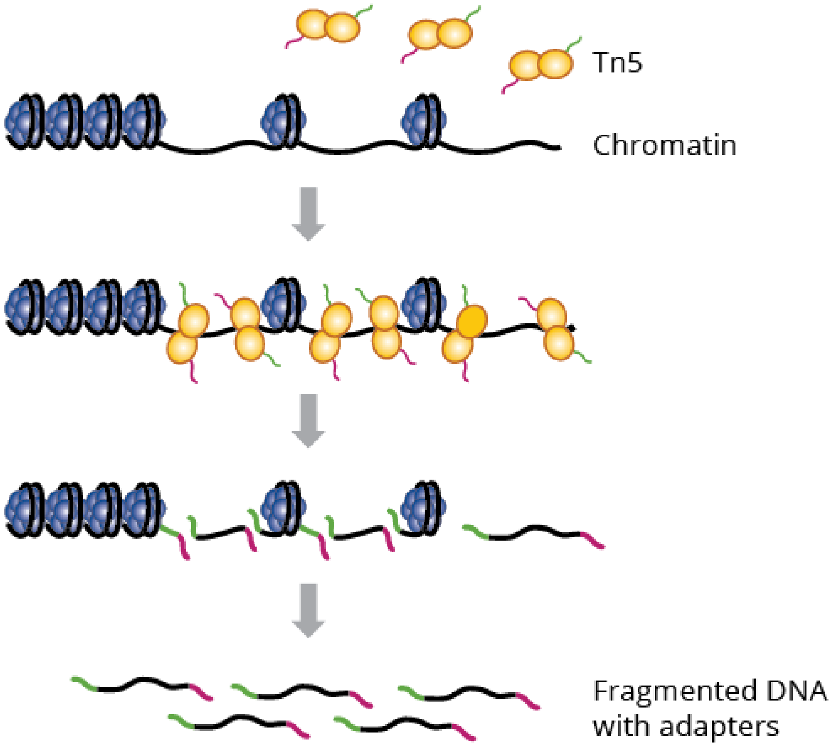
- For OT, CTOT: C>T; C>C
- For CTOB, OB: G>A; G>G

# What is ATAC-seq?

Cells/tissue
↓
Permeabilize nuclei
↓
Tagment
↓
PCR amplification

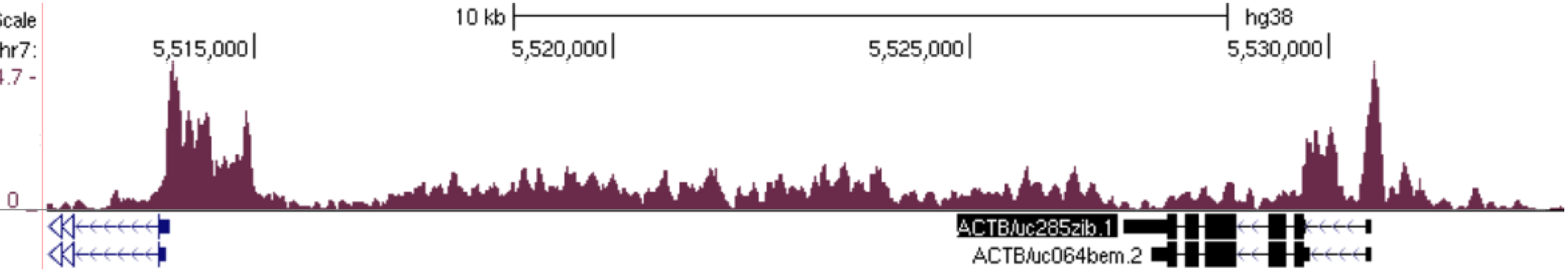
| | |
|---|---|
| **Applications**: | chromatin accessibility assay<br>enhancer identification<br>‘poised’ genes (open but off) |
| **Input**: | cells/tissue (native nuclei) |
| **Enrichment**: | accessible chromatin |
| **Reads**: | map to small intervals (‘peaks’) |
| **Analysis**: | peak identification<br>differential peak representation<br>motif enrichment |
| **Variations:** | |

# What is ATAC-seq?



Mapped Reads

Gene annot.



# What is CLIP-seq?

Cells/tissue
↓
Crosslink **RNA**+Ptn
↓
Lyse cells
↓
Immunoprecipitate
↓
Ligate adapters (x2)
↓
cDNA
↓
PCR amplification

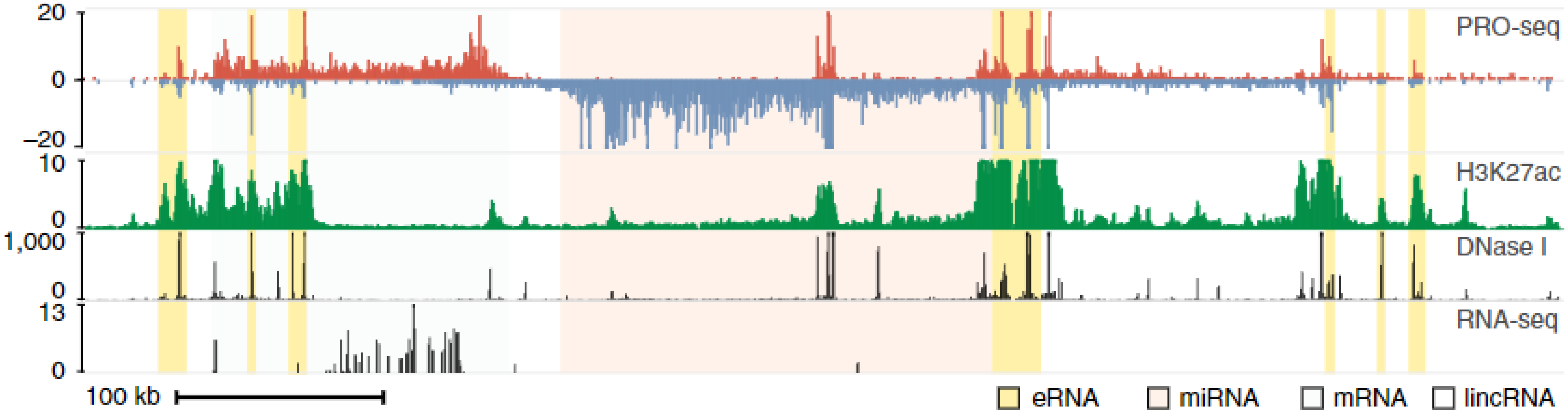
| | |
|---|---|
| **Applications**: | identify RBP binding sites |
| **Input**: | cells/tissue (native RNA) |
| **Enrichment**: | antibody immunoprecipitation |
| **Reads**: | map to mRNAs (binding sites) |
| **Analysis**: | RBP binding site identification<br>differential binding<br>RNA motif enrichment |
| **Variations:** | RIP-seq, PAR-CLIP, … |

# What is PRO-seq?

Cells
↓
Permeabilize nuclei
↓
Run-on (bio-dXTP)
↓
Biotin capture
↓
Ligate adapters (x2)
↓
cDNA
↓
PCR amplification

| | |
|---|---|
| **Applications**: | identify nacent RNA transcripts |
| **Input**: | cells (native nuclei) |
| **Enrichment**: | Biotin (run-on incorporation) |
| **Reads**: | map to exons, promoters,<br>enhancers (eRNAs) |
| **Analysis**: | gene expression profiling,<br>enhancer identification,<br>pol II localization (pausing),<br>transcription rate,… |
| **Variations:** | GRO-seq, ChRO-seq, PRO-cap,… |

# What is PRO-seq?

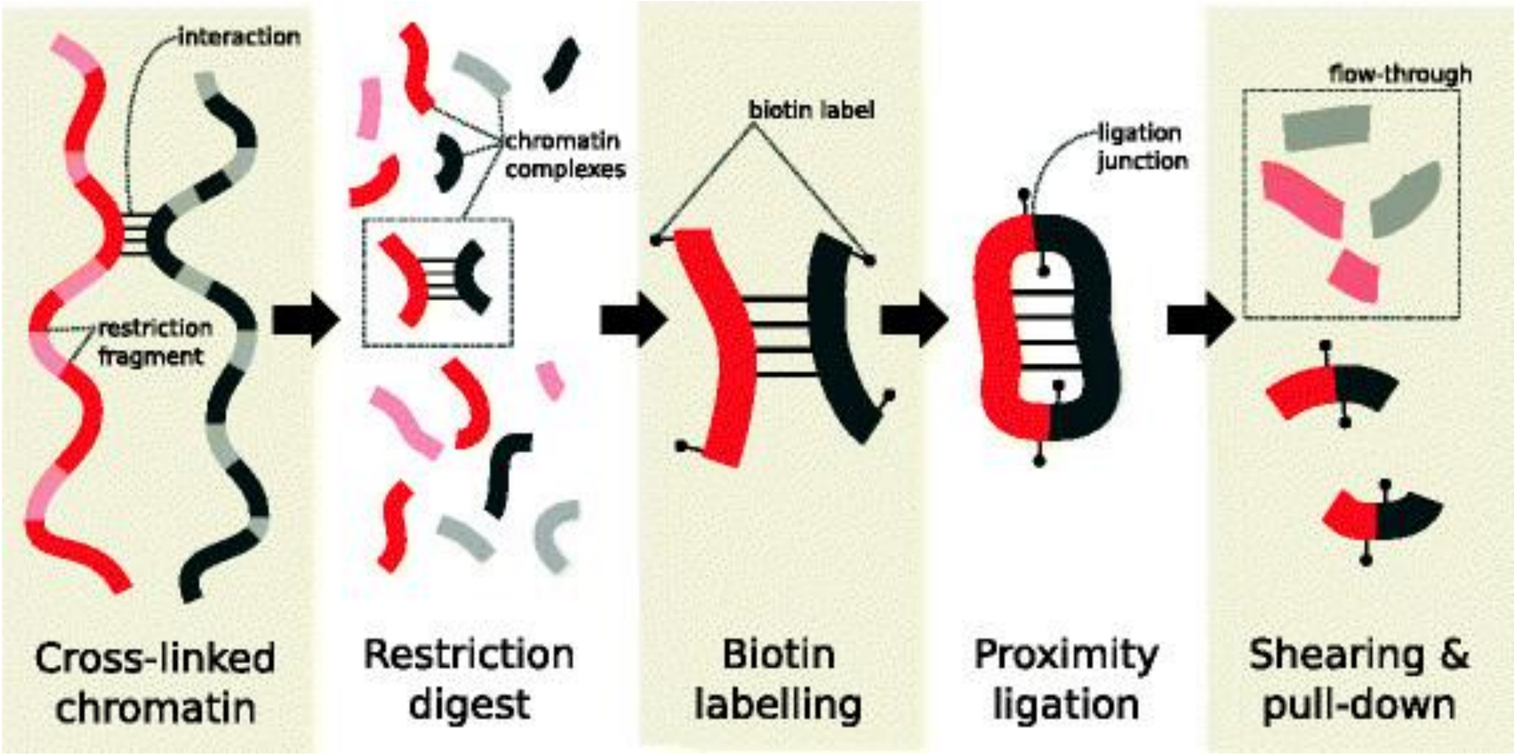


# What is HiC-seq?

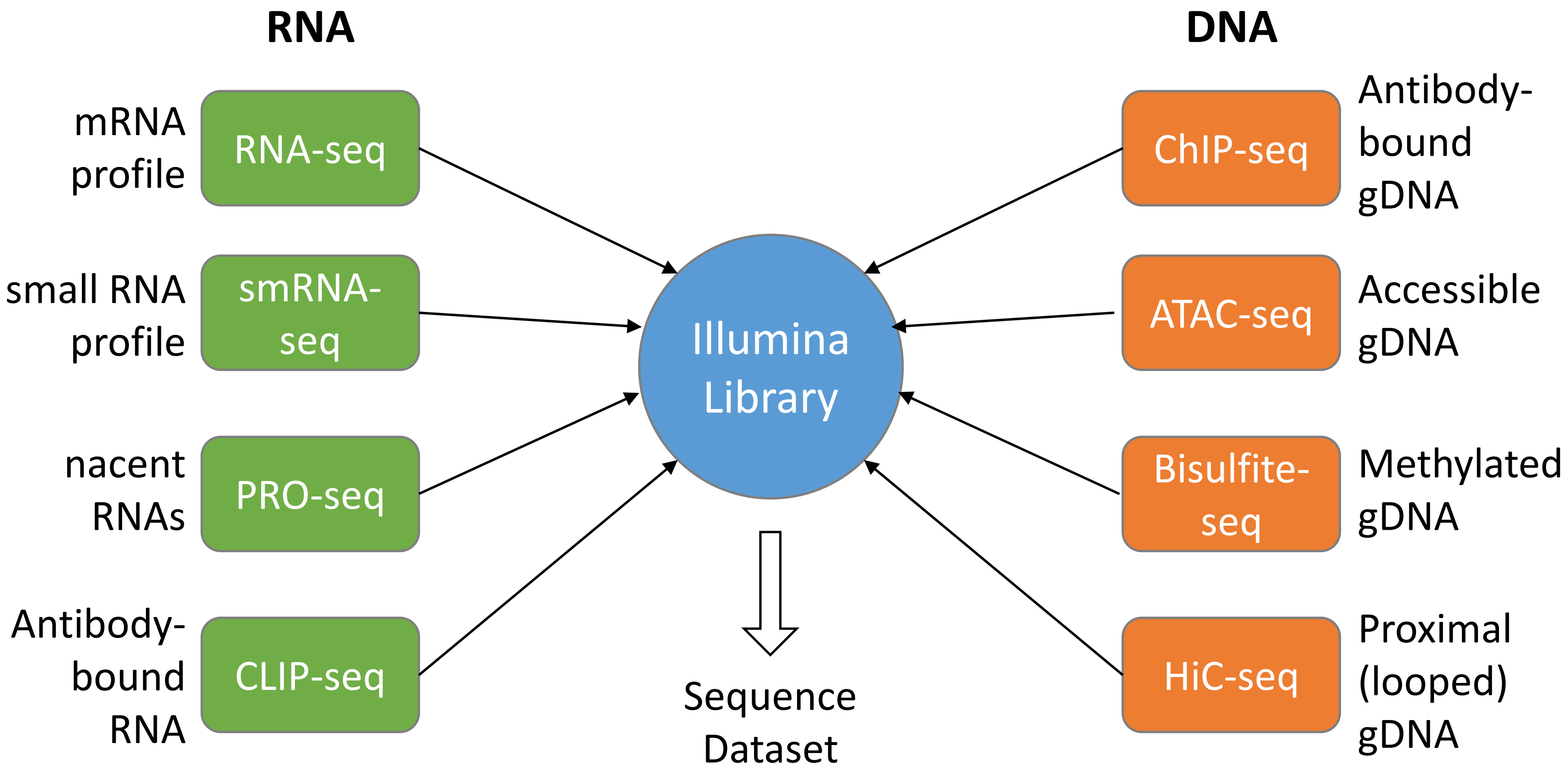
Cells/tissue
↓
Crosslink **gDNA**
↓
Digest
↓
Biotin end-label
↓
Proximity ligation
↓
Shear, Biotin capture
↓
Adapter ligation, PCR

| | |
|---|---|
| **Applications**: | chromatin interaction (looping)<br>genome configuration |
| **Input**: | cells/tissue (native gDNA) |
| **Enrichment**: | Biotin (proximity ligation) |
| **Reads**: | map to genome |
| **Analysis**: | paired-end read positions = proximity sites |
| **Variations:** | C3, ChIA-PET, Hi-cap |

# What is HiC-seq?



# How can we use NGS to study gene regulation?



# Important Considerations

**Experimental Design**

| | |
|---|---|
| Controls | relative quantification |
| Replicates | statistical power |

**Quality Control Checks**

| | |
|---|---|
| Input material | RNA integrity, Ab quality, … |
| Library | size distribution, concentration |
| Sequencing data | base quality, mapping quality, … |

# Transcriptional Regulation and Expression Facility

***trex_info@cornell.edu***

*Faraz Ahmed*
*Christine Butler*
*Ann Tate*

## Sign up for our List-Serv!

*Send an email message to
TREX-GENEREG-L-request@cornell.edu
with **“join”** as the subject