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
  Assay for Transposase-Accessible Chromatin 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
- Antibody-bound gDNA
  - ChIP-seq
- Accessible gDNA
  - ATAC-seq
- Methylated gDNA
  - Bisulfite-seq
- Proximal (looped) gDNA
  - HiC-seq
What is RNA-seq?

**Applications:**
- gene expression profiling
- transcript annotation/assembly
- pathogen identification
- variant discovery/identification

**Input:**
- total RNA

**Enrichment:**
- polyA+ or rRNA-depletion

**Reads:**
- map to exons

**Analysis:**
- differential gene expression
- transcript assembly/annotation

**Variations:**
- 3’ RNA-seq, targeted RNA-seq
What is RNA-seq?

RNAseq track

gene annot track
What is smRNA-seq?

**Applications:** microRNA profiling

**Input:** total RNA
cell-free RNA

**Enrichment:** ligation to 5’-monoP, 3’-OH
post-PCR size selection

**Reads:** map to mature miRNAs

**Analysis:** differential microRNA counts
microRNA discovery, processing
biomarker ID/profiling

**Variations:** poly-adenylation/RT
circularization

**Diagram:**
1. RNA
2. Ligate adapters (x2)
3. cDNA
4. PCR amplification
5. Size selection
What is ChIP-seq?

**Applications:**
- chromatin mark distribution
- transcription factor occupancy
- DNA binding protein sites

**Input:**
- cells/tissue (native gDNA)

**Enrichment:**
- antibody immunoprecipitation

**Reads:**
- map to small intervals (‘peaks’)

**Analysis:**
- peak identification
- differential peak representation
- motif enrichment

**Variations:**
- ChIP-exo, CUT&RUN, CUT&Tag

---

Cells/tissue → Crosslink DNA+Ptn → Lyse/sonicate → Immunoprecipitate → Reverse crosslink → Adapter ligation → PCR amplification
What is ChIP-seq?

Peak Calls

Mapped Reads

Gene annot.
What is Bisulfite-seq?

**Applications:** identify methylated CpG

**Input:** gDNA

**Enrichment:** optional *(RRBS, targeted)*

**Reads:** C (reference)→T (read) indicates unmethylated-C

**Analysis:** location, frequency of me-C 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

---

**methylated locus**

```
ATATCGGTATT-3'
TATAgcGcATAA-5'
```

**unmethylated locus**

```
ATATCGGTATT-3'
TATAgcGcATAA-5'
```

**Bisulfite conversion**

```
ATATCGUGTATT-3'
TATAgcGuATAA-5'
```

**Polymerase chain reaction**

```
ATATcgTGtATT-3'
TATAgcAcATAA-5'
```

```
ATATUGUGTATT-3'
TATAgUGuATAA-5'
```

**Sequencing**

```
ATATCGGTATT-3'
```

**Reference**

```
ATATCGGTATT-3'
```

```
ATATCGCATATT-3'
```

```
ATATCGCATATT-3'
```

```
ATATCGCATATT-3'
```

```
ATATCGCATATT-3'
```

```
ATATCGCATATT-3'
```
What is ATAC-seq?

**Applications**: chromatin accessibility assay enhancer identification ‘poised’ genes (open but off)

**Input**: cells/tissue (native nuclei)

**Enrichment**: accessible chromatin

**Reads**: map to small intervals (‘peaks’)

**Analysis**: peak identification differential peak representation motif enrichment

**Variations**:
What is ATAC-seq?

- Mapped Reads
- Gene annot.
What is CLIP-seq?

**Applications:** identify RBP binding sites

**Input:** cells/tissue (native RNA)

**Enrichment:** antibody immunoprecipitation

**Reads:** map to mRNAs (binding sites)

**Analysis:** RBP binding site identification
differential binding
RNA motif enrichment

**Variations:** RIP-seq, PAR-CLIP, ...

**Cells/tissue**

Crosslink RNA+Ptn

Lyse cells

Immunoprecipitate

Ligate adapters (x2)

cDNA

PCR amplification
What is PRO-seq?

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

**Cells**

- Permeabilize nuclei
- Run-on (bio-dXTP)
- Biotin capture
- Ligate adapters (x2)
- cDNA
- PCR amplification
What is PRO-seq?
What is HiC-seq?

**Applications:** chromatin interaction (looping) 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

Cells/tissue

↓

Crosslink gDNA

↓

Digest

↓

Biotin end-label

↓

Proximity ligation

↓

Shear, Biotin capture

↓

Adapter ligation, PCR
What is HiC-seq?

Cross-linked chromatin → Restriction digest → Biotin labelling → Proximity ligation → Shearing & pull-down
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**
- Antibody-bound gDNA
  - ChIP-seq
- Accessible gDNA
  - ATAC-seq
- Methylated gDNA
  - Bisulfite-seq
- Proximal (looped) gDNA
  - HiC-seq

**Illumina Library**

**Sequence Dataset**
Important Considerations

Experimental Design

- Controls
- Replicates
  - relative quantification
  - statistical power

Quality Control Checks

- Input material
- Library
- Sequencing data
  - RNA integrity, Ab quality, ...
  - size distribution, concentration
  - base quality, mapping quality, ...
Transcriptional Regulation and Expression Facility
trex_info@cornell.edu

Sign up for our List-Serv!
*Send an email message to TREX-GENEREG-L-request@cornell.edu with “join” as the subject

Faraz Ahmed
Christine Butler
Ann Tate