ChIP-chip (or ChIP-on-chip)

- ChIP-chip uses microarray technology to determine the identity of DNA fragments produced by ChIP.
- Typically a control sample (genomic DNA without going through ChIP) is used to properly define relative enrichment of specific sequences in the ChIP DNA.
- It is the dominant high-throughput technique before the arrival of ChIPSeq.
ChIPSeq vs. ChIP-chip

• The experimental design of ChIPSeq is considerably simpler.

• ChIPSeq typically can achieve higher genomic coverage than ChIP-chip (also depends on read length vs. probe length).

• The data from ChIPSeq is arguably cleaner and easier to process.

• Costs are comparable (?).
ChIP-chip maps the physical locations of proteins bound to DNA (or RNA)
• **High throughput sequencing – a new paradigm**
  – Solexa
  – SOLiD
  – 454/Roche Genome Sequencer
  – Ion Torrent

• **Applications**
  – Genome sequencing
  – microRNA screening
  – Gene expression
  – ChIP-seq
What is ChIP- Sequencing?

• ChIP- Sequencing is a frontier technology to analyze protein interactions with DNA.

• ChIP-Seq
  – Combination of chromatin immunoprecipitation (ChIP) with ultra high-throughput massively parallel sequencing
  – Allow mapping of protein–DNA interactions in-vivo on a genome scale
Workflow of ChIP-Seq

1. Cross-link whole cells with formaldehyde
2. Isolate genomic DNA
3. Add protein-specific antibody
4. Sonicate DNA to produce sheared, soluble chromatin
5. Immunoprecipitate and purify immunocomplexes
6. Reverse cross-links, purify DNA and prepare for sequencing

Workflow of ChIP-Seq

Sequence

Map to genome
ChIP-seq

Challenges:
• Millions of segments
• Mapping to genome
• Visualization
• Peak detection
• Data normalization
• …

1. Prepare gDNA Library
   Duration: 3 hr hands-on (6 hrs total)
   ▶ Fragment genomic DNA
   ▶ Repair ends
   ▶ Phosphorylate & add A-overhang
   ▶ Ligate adapters
   ▶ Purify library

2. Generate Clusters
   Duration: < 1 hr hands-on (5 hrs total)
   ▶ Place reagents into Cluster Station
   ▶ Place cell onto Cluster Station
   ▶ Load DNA Libraries
   ▶ Press “Start”

3. Sequence Clusters
   Duration: < 1 hr hands-on (2.5 days single read)
   ▶ Place reagents into the Genome Analyzer
   ▶ Place flow cell into Genome Analyzer
   ▶ Start run

4. Data Analysis
   Duration: 1-2 days
   ▶ Image processing
   ▶ Real-time reporting
   ▶ Base calling
   ▶ Alignment
Genome-Wide Mapping of in Vivo Protein-DNA Interactions

David S. Johnson, Ali Mortazavi, Richard M. Myers, Barbara Wold

In vivo protein-DNA interactions connect each transcription factor with its direct targets to form a gene network scaffold. To map these protein-DNA interactions comprehensively across entire mammalian genomes, we developed a large-scale chromatin immunoprecipitation assay (ChIPSeq) based on direct ultrahigh-throughput DNA sequencing. This sequence census method was then used to map in vivo binding of the neuron-restrictive silencer factor (NRSF; also known as REST, for repressor element–1 silencing transcription factor) to 1946 locations in the human genome. The data display sharp resolution of binding position [±50 base pairs (bp)], which facilitated our finding motifs and allowed us to identify noncanonical NRSF-binding motifs. These ChIPSeq data also have high sensitivity and specificity [ROC (receiver operator characteristic) area ≥ 0.96] and statistical confidence (P < 10^{-4}), properties that were important for inferring new candidate interactions. These include key transcription factors in the gene network that regulates pancreatic islet cell development.
Johnson et al, 2007

- ChIP-Seq technology is used to understand in vivo binding of the neuron-restrictive silencer factor (NRSF)
- Results are compared to known binding sites
  - ChIP-Seq signals are strongly agree with the existing knowledge
- Sharp resolution of binding position
- New NRSF binding motifs are identified
Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing

Gordon Robertson¹, Martin Hirst¹, Matthew Bainbridge¹, Misha Bilenky¹, Yongjun Zhao¹, Thomas Zeng¹, Ghia Euskirchen², Bridget Bernier¹, Richard Varhol¹, Allen Delaney¹, Nina Thiessen¹, Obi L Griffith¹, Ann He¹, Marco Marra¹, Michael Snyder² & Steven Jones¹

We developed a method, ChIP-sequencing (ChIP-seq), combining chromatin immunoprecipitation (ChIP) and massively parallel sequencing to identify mammalian DNA sequences bound by transcription factors in vivo. We used ChIP-seq to map STAT1 targets in interferon-γ (IFN-γ)–stimulated and unstimulated human HeLa S3 cells, and compared the method's performance to ChIP-PCR and to ChIP-chip for four chromosomes. By ChIP-seq, using 15.1 and 12.9 million uniquely mapped sequence reads, and an estimated false discovery rate of less than 0.001, we identified 41,582 and 11,004 putative STAT1-binding regions in stimulated and unstimulated cells, respectively. Of the 34 loci known to contain STAT1 interferon-responsive binding sites, ChIP-seq found 24 (71%). ChIP-seq targets were enriched in sequences similar to known STAT1 binding motifs. Comparisons with two ChIP-PCR data sets suggested that ChIP-seq sensitivity was between 70% and 92% and specificity was at least 95%.

single-end tags (SETs), which are simpler to prepare than PETs, may be effective for profiling mammalian protein-DNA interactions. Thus we appraised the 1G system as a platform for ChIP with tag sequencing.

As a test system, we selected the mammalian transcription factor STAT1, whose cellular biology is relatively well characterized, and whose use permits a comparison of unstimulated and stimulated cellular states. In both resting and stimulated cells, STAT proteins shuttle continuously between cytoplasm and nucleus. Signaling by several cytokines, growth factors and hormone receptors leads to activation of receptor-associated JAK family kinases that phosphorylate a substantial fraction of cytoplasmic STAT1 proteins. Phosphorylated STAT1 forms specific homodimers, heterodimers and heterotrimers that bind DNA with high affinity, and thus accumulate in the nucleus. STAT1 complexes activate or repress transcription primarily by the homodimer binding to IFN-γ activation site (GAS) elements, but also to interferon-stimulated response elements (ISREs). The regulatory activity of STAT1
Robertson *et al*, 2007

- ChIP-Seq technology used to study genome-wide profiles of STAT1 DNA association
- STAT1 targets in interferon-γ-stimulated and unstimulated human HeLa S3 cells are compared
- The performance of ChIP-Seq is compared to the alternative protein-DNA interaction methods of ChIP-PCR and ChIP-chip.
- 41,582 and 11,004 putative STAT-1 binding regions are identified in stimulated and unstimulated cells respectively.
Why ChIP-Seqencing?

- Current microarray and ChIP-ChIP designs require knowing sequence of interest as a promoter, enhancer, or RNA-coding domain.
- Lower cost
- Less work in ChIP-Seq
- Higher accuracy
- Alterations in transcription-factor binding in response to environmental stimuli can be evaluated for the entire genome in a single experiment.
Cross-link whole cells with formaldehyde

Isolate genomic DNA

Add protein-specific antibody

Sonicate DNA to produce sheared, soluble chromatin

Immunoprecipitate and purify immunocomplexes

Reverse cross-links, purify DNA and prepare for sequencing

Sequence

Bioinformatics

Map to genome
Array CGH

• Array based comparative genome hybridization (CGH)
• Measures amount of DNA
• Comparison between two samples
  – ‘Test’ sample
  – ‘Reference’ sample
Comparative Genomic Hybridization

Control DNA  Patient DNA
Array CGH

Reference DNA (cy5)
Mix and hybridize to array

Test DNA (cy3)

Array containing probes corresponding to genomic DNA

Scan and analyze image

Feuk et al.
Nature Reviews Genetics 2006
Array CGH

- Detecting genomic rearrangements found in cancer (tumor genome vs normal genome)
- Study of genomic copy number variation
Experimental Procedure

• Label test and reference DNA samples with different fluorophores

• Mix and hybridize to array
  – Array contains probes corresponding to genome sequence
    • Early arrays consisted of spotted BACs (n=~30,000 for human)
    • Current arrays contain 100,000-2 million short (45-80 nucleotide) oligonucleotide probes

• Scan array, recording intensity of each fluorophore for each probe

• A second experiment, using the same samples but with the fluorophores swapped, can be performed as a control