illumina®

Whole-Genome Chromatin IP Sequencing (ChIP-Seq)

Illumina ChIP-Seq combines chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing to identify binding sites of DNA-associated proteins. Illumina ChIP-Seq technology precisely and cost-effectively maps global binding sites for a protein of interest.

Highlights of Illumina ChIP-Seq

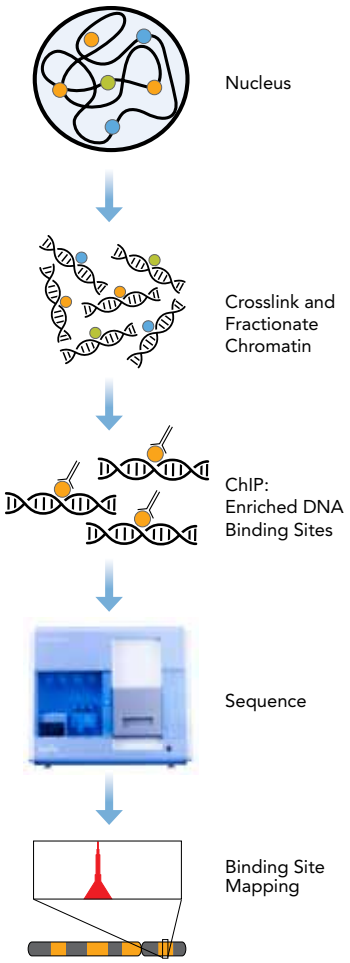
- **High Quality Data:**
  Positional precision of mapped binding sites ± 50bp
- **Wide Dynamic Range:**
  Robust quantification for determining binding specificities of varying strengths
- **High Signal-to-Noise Ratio:**
  Lower background than ChIP-chip, no cross hybridization
- **Genome-Wide Analysis:**
  Identifies any binding sites, not limited to array features or candidate sequences
- **Low Starting Material Requirement:**
  Robust output from as little as 10ng of precious input

Introduction

Transcription factors and other chromatin-associated proteins are essential phenotype-influencing mechanisms. Determining how proteins interact with DNA to regulate gene expression is essential for fully understanding many biological processes and disease states. This epigenetic information is complimentary to genotype and expression analysis. Traditional methods have successfully identified transcription factor binding sites and specific DNA-associated protein modifications and their roles in regulating specific genes, but these experiments are limited in scale and resolution. The powerful Illumina Whole-Genome Chromatin IP Sequencing (ChIP-Seq) application allows researchers to easily expand the scale of their studies to identify binding sites across the entire genome simultaneously with high resolution and without constraints.

Specific DNA sites in direct physical interaction with transcription factors and other proteins can be isolated by chromatin immunoprecipitation (ChIP). ChIP produces a library of target DNA sites that a given factor was bound to in vivo. The revolutionary Solexa Sequencing technology is an ideal method to identify isolated DNA sites from ChIP. This massively parallel sequence analysis in the context of easy access to whole-genome sequence databases has made analyzing the interaction pattern of any protein with DNA[1,2], or the pattern of any epigenetic chromatin modifications[3,4], across the entire genome fast and cost-effective. The Illumina Genome Analyzer determines the sequences of ChIP-isolated DNA fragments to identify and quantify the sites bound by a protein of interest. ChIP-Seq technology supports virtually unconstrained selection of any ChIP-able protein and modifications to be studied[5], such as transcription factors, polymerases and transcriptional machinery, structural proteins, protein modifications, and DNA modifications (Table 1).

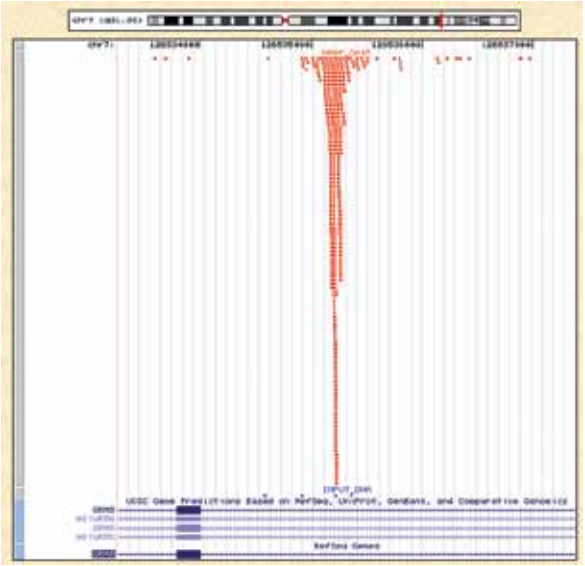
Figure 1: ChIP-Seq Experiment Workflow



Protein-chromatin interactions are first crosslinked *in situ*. Specific DNA fragments are co-immunoprecipitated and sequenced to identify genome-wide sites associated with a factor or modification of interest.

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA

Figure 2: Mapping Binding Sites onto the Genome



Sequence tags (red) from DNA ChIP-enriched for Neuron-restrictive Silencer Factor (NRSF) binding sites are shown in the UCSC genome browser aligned to a region on chromosome 7. There is a clear binding site in the GRM8b gene, and almost no non-specific signal (blue, input DNA). These data were kindly provided by Dr. Barbara Wold.

Whole Genome ChIP-Seq

Chromatin immunoprecipitation is a powerful method to selectively enrich for DNA sequences bound by a particular protein in living cells. However, the widespread use of this method has been limited by the lack of a sufficiently robust method to identify all of the enriched DNA sequences. The ChIP process enriches specific crosslinked DNA-protein complexes using an antibody against a protein of interest (Figure 1). Oligonucleotide adapters are then added to the small stretches of DNA that were bound to the protein of interest to enable massively parallel sequencing. After size selection, all the resulting ChIP DNA fragments are sequenced simultaneously using the Genome Analyzer and Solexa® Sequencing technology. A single sequencing run can scan for genome-wide associations with high resolution, as opposed to large sets of tiling arrays required for lower resolution ChIP-chip.

Illumina DNA Sequencing technology uses a unique process to generate high-density, high-throughput sequencing runs. The fully automated Illumina Cluster Station amplifies adapter-ligated ChIP DNA fragments on a solid flow cell substrate to create clusters of approximately 1000 clonal copies each. The resulting high density array of template clusters on the flow cell surface is sequenced by the fully automated Illumina Genome Analyzer. Each template cluster undergoes sequencing by synthesis in parallel using novel fluorescently labeled reversible terminator nucleotides. Templates are sequenced base-by-base during each read. Then, the data collection and analysis software aligns sample sequences to known genomic sequence to identify the ChIP DNA fragments (Figure 2).

High Quality Data

A large number of short individual sequence reads are produced by the Illumina Genome Analyzer. Sensitivity and signal-to-noise ratios are very high since three to five million individual reads are typically produced in each run. Since the system has the capacity for high oversampling and redundancy, signals are readily detectable above background. Additionally, sensitivity and statistical certainty can be tuned by adjusting the total number of sequence reads to provide an even wider dynamic range and greater ability to detect rare DNA-protein interaction sites. DNA sequence reads are aligned to a reference genome sequence, allowing determination of all of the binding sites for a factor of interest. Sequence read lengths of only 25–32 bases are sufficient to accurately align and identify millions of fragments per run. Unlike microarray-based ChIP methods, the accuracy of the ChIP-Seq assay is not limited by the spacing of predetermined probes. By integrating a large number of short reads, highly precise (± 50bp) binding site localization is obtained. Binding affinities of a protein to different DNA sites can be compared by quantifying the number of appearances of a given sequence[1].

The Genome Analyzer, powered by Solexa Sequencing technology, yields outstanding base call accuracy (greater than 98.5%). This accuracy and read length enables true whole-genome ChIP profiling.

Summary

Illumina ChIP-Seq achieves unparalleled data density with highly accurate and precise results enabling comprehensive whole-genome mapping of DNA-binding sites. Researchers can study any immunoprecipitates from virtually any sequenced organism using a single system. Low sample input requirements minimize tedious immunoprecipitations. Truly comprehensive mapping of *in vivo* binding sites across an entire genome is achieved for significantly lower cost than genome-wide ChIP-chip. Illumina ChIP-Seq does not require iterative probe design and validation, making it much more efficient than other ChIP methods for studying a variety of organisms. Plus, most transcription factor binding sites can be mapped using data generated in a single lane from one eight-lane flow cell.

Illumina Sequencing Solutions

For Illumina ChIP-Seq, the standard Genome Analyzer and Cluster Station are required. Only minor changes to the sample preparation protocol are required to use ChIP-isolated DNA. A full assay manual describing the ChIP-Seq application is available from Illumina. A ChIP-Seq Data Analysis Technical Note describes some third-party software packages for downstream analysis recommended by Illumina.

Illumina's Genome Analyzer system enables much more than ChIP-Seq analysis. Many applications are enabled with just the single capital investment, and training on just a single technology. For example, sample preparation kits are also available for resequencing and de novo sequencing with paired-end reads, small RNA identification, and digital gene expression analysis.

Table1: ChIP-Seq Compared to ChIP-ChIP Analysis

	ChIP-Seq	ChIP-chip	ChIP-Seq Advantage
Starting material	Low: Down to 10 ng	4 µg	Hundreds-fold lower DNA input requirements means fewer IP reactions
Flexibility	Yes: Genome-wide assay of any sequenced organism	Limited: Dependent on available products	Not limited to content available on arrays
Positional resolution	± 50bp	± 500–1000bp	Site mapping can be an order of magnitude more precise
Sensitivity	Widely tunable: Increase counts to increase sensitivity	Poor: Based on hybridization and ratios	Simply increase the number of counts to obtain desired sensitivity
Cross-hybridization	None: Each DNA is individually sequenced	Significant	Higher quality data even in complex genomes

REFERENCES

1. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316: 1497-1502.
2. Robertson G, Hirst M, Bainbridge M, Bilenky M, Zhao Y et al. (2007) Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. Nature Methods 4: 651-657.
3. Barski A, Cuddapah S, Cui K, Roh TY, Schones DE et al. (2007) High-resolution profiling of histone methylations in the human genome. Cell 129: 823-837.
4. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553-560.
5. Fields S (2007) Molecular biology. Site-seeing by sequencing. Science 316: 1441-1442.

ADDITIONAL INFORMATION

Visit our website or contact us at the address below to learn more about Illumina DNA Sequencing Applications using Solexa Sequencing technology.

Illumina, Inc. • 9885 Towne Centre Drive, San Diego, CA 92121 USA • 1.800.809.4566 toll-free • 1.858.202.4566 tel • techsupport@illumina.com • illumina.com

FOR RESEARCH USE ONLY

© 2010 Illumina, Inc. All rights reserved.
Illumina, illumina*Dx*, Solexa, Making Sense Out of Life, Oligator, Sentrix, GoldenGate, GoldenGate Indexing, DASL, BeadArray, Array of Arrays, Infinium, BeadXpress, VeraCode, IntelliHyb, iSelect, CSPro, GenomeStudio, Genetic Energy, HiSeq, and HiScan are registered trademarks or trademarks of Illumina, Inc. All other brands and names contained herein are the property of their respective owners. Pub. No. 770-2007-007 Current as of 26 November 2007

illumina®

AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
TCAACGTACCGTAACGAACGTATCATTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTT
CGACGAAAAGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACC
AACGTACCATTAAGAGCTACCGTGCAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGACGAAAAGAATGATAA
AGAATGATAACAGTAACACACTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAAC
GATTACTTGATCCACTGATTCAACGTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGCTTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAGCAACGACG
CGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGTCTGTTAACCTTAAGATTACTTGATCCACTGATTCAACGTACCGTAACGAACGTATCAATTGAGACTAAATATTAACGTACCATTAAGAGCTACCGTGCAACGAAAAGAATGATAACAGTA