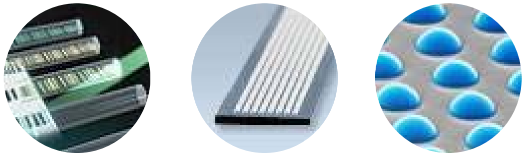
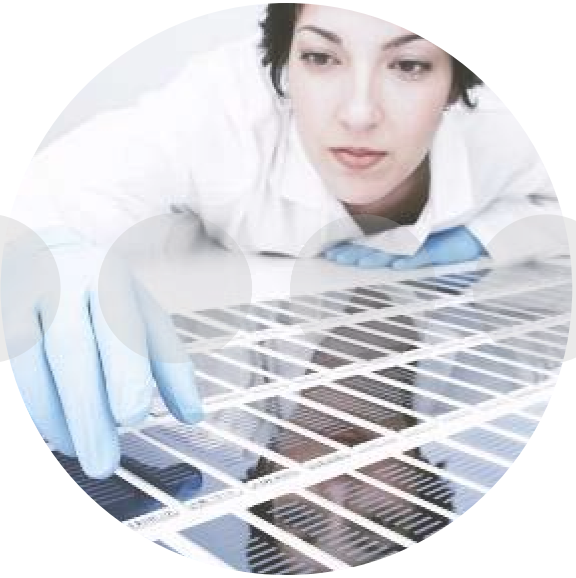
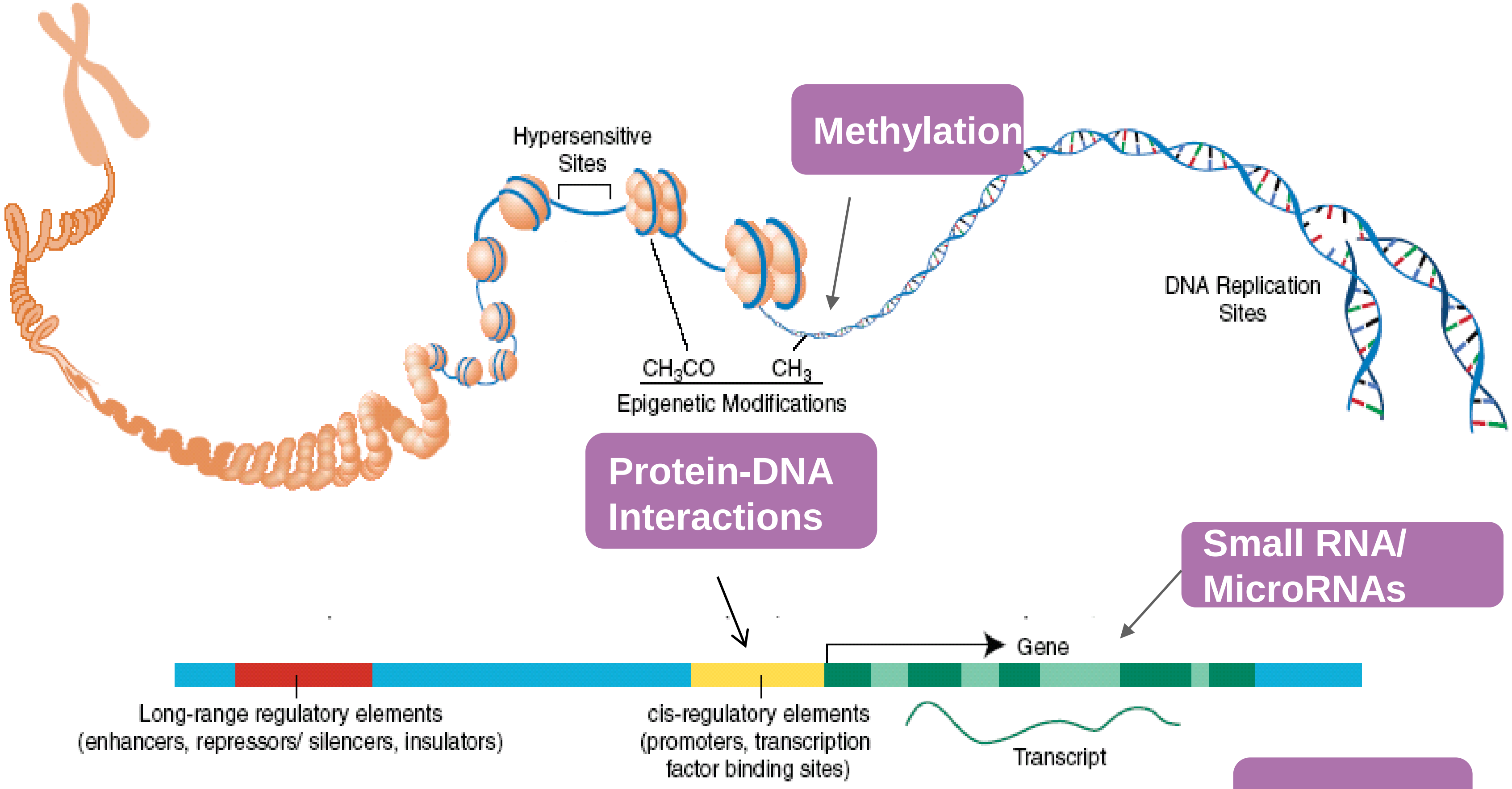




Overview of Illumina’s Expression and Regulation Portfolio

Brent Applegate
Associate Product Manager, Gene Expression & Regulation

ENCODE Project: Expression, Regulation and Epigenetics



Methylation

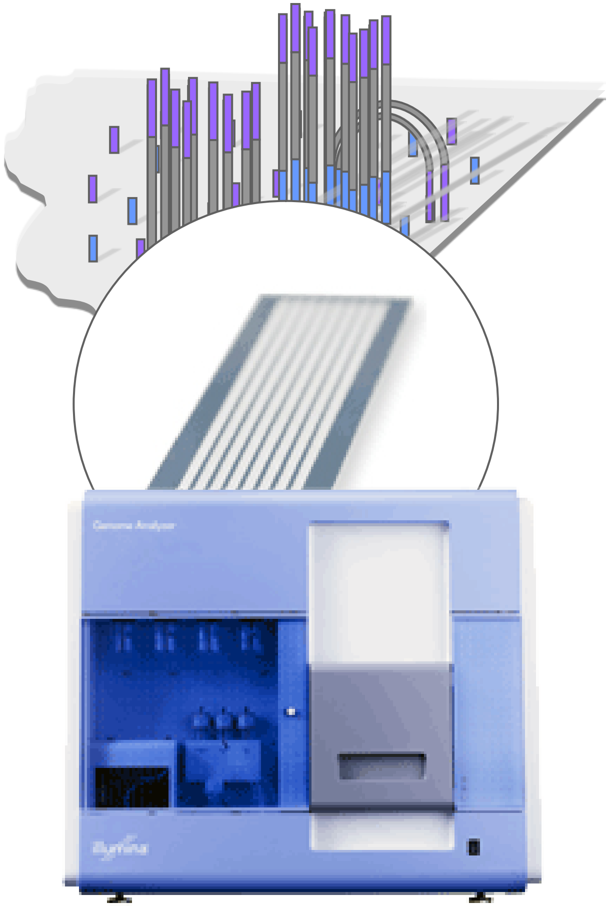
Protein-DNA Interactions

Small RNA/ MicroRNAs

Expression

Gene Expression and Regulation Platforms

Sequencing



BeadArray™

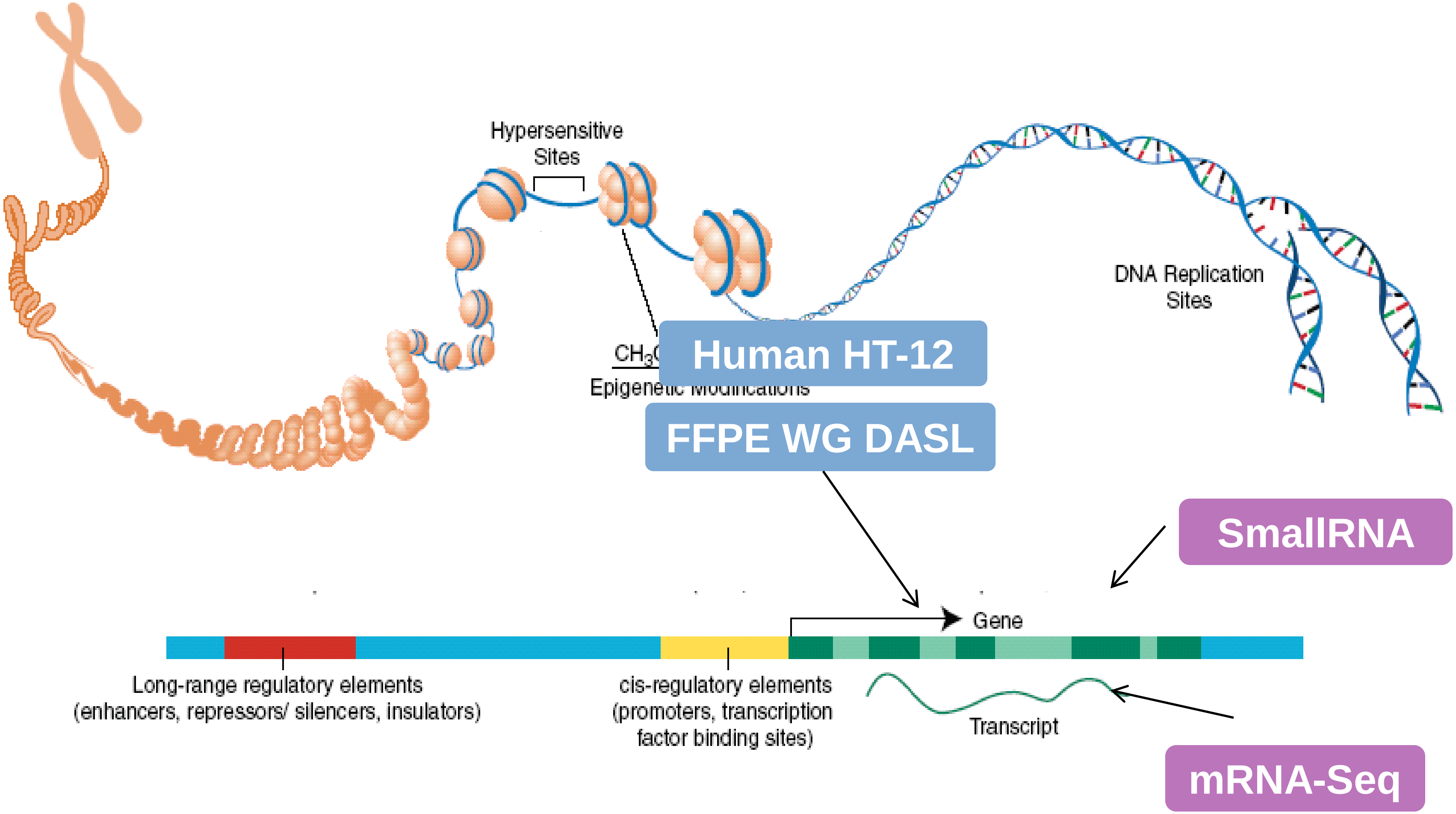
VeraCode™

Agenda

- Expression/Transcriptomics
  - Whole Genome Arrays
  - Whole Genome Arrays for Archival Samples
  - Transcriptome Analysis

- Regulation/Epigenomics
  - DNA Methylation Arrays
  - DNA Methylation Sequencing (BiS-Seq)
  - ChIP Sequencing (ChIP-Seq)

Gene Expression (Transcriptomics)



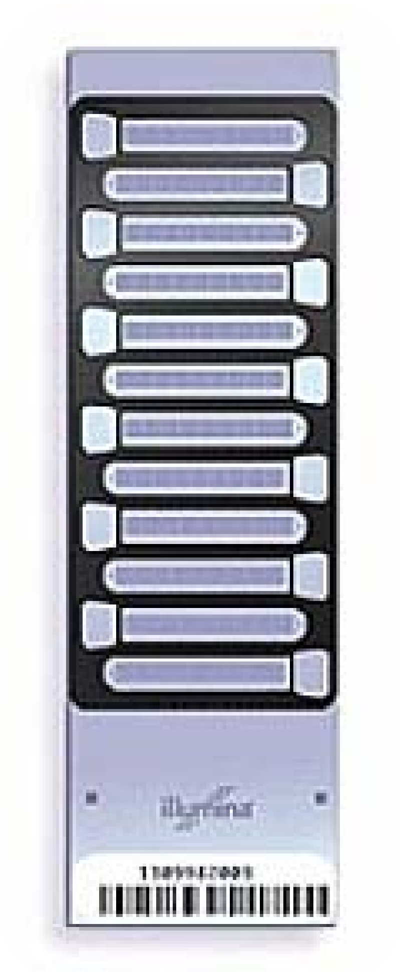
Human HT-12

FFPE WG DASL

SmallRNA

mRNA-Seq

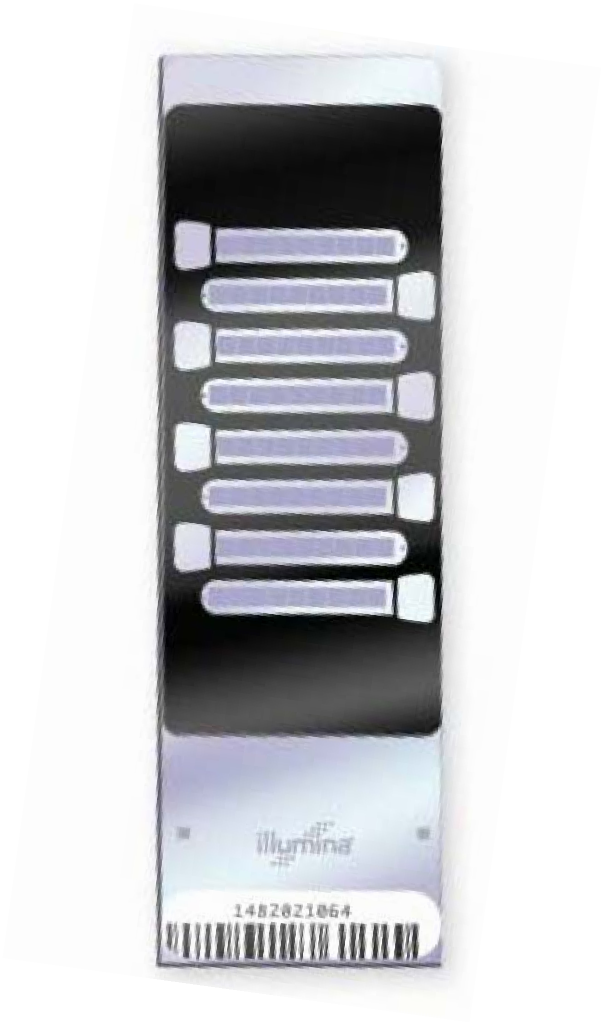
Whole Genome Expression Arrays



HumanHT-12 v3

- Ultra low cost
  - $75/sample (<$100 per sample with labeling)
- High throughput
  - 12 samples per slide
  - >100 samples/day without specialized equipment
- Up to date content
  - RefSeq release 22
  - UniGene build 199

Whole Genome Expression with FFPE



Human WG-DASL

- Works with degraded samples
  - Based on the DASL assay
  - FFPE samples >20 years old
  - Brain samples
- Low sample input
  - As few as 100 cells
- Low cost
  - ~$150/sample (includes labeling)

mRNA-Seq: Sequencing Full Length mRNA

- **Quantitate levels of RNA expression**
- **Genome annotation**
  - Provide EST-like information to help annotate previously sequenced genomes
  - Discovery of novel transcripts
  - Characterize alternative splicing
- **Study diseases processes, such as cancer**
  - Find cSNPs and other point mutations
  - Discovery of disease-associated biomarkers
  - Quantitate aberrant expression in diseased tissue

mRNA-Seq: Truly Quantitative Gene Expression

FKBP8 Gene Expression

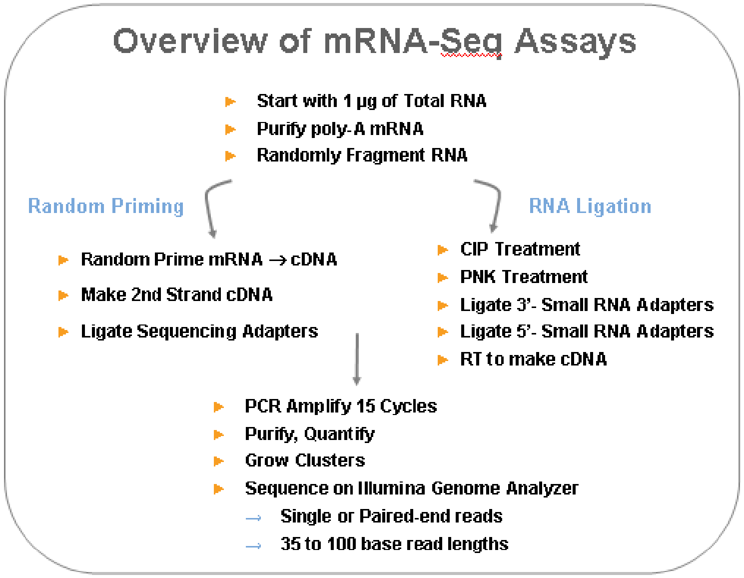
RPS3 Gene Expression

These screen shots were taken directly from the Illumina Genome Viewer (IGV), which is part of the GenomeStudio™ Software Suite

Comprehensive tools for transcriptome analysis

Two methods, complete answers

- mRNA-Seq
  - Most uniform coverage across transcripts.
  - Allows for more quantitative results.
  - Offers better coverage for cSNP discovery and calling.
  - Useful when strand specificity isn't needed
- Directional mRNA-Seq
  - Conserves strand specific information
  - Key for discovery & annotation of novel transcripts and isoforms



Overview of mRNA-Seq Assays

- Start with 1 µg of Total RNA
- Purify poly-A mRNA
- Randomly Fragment RNA

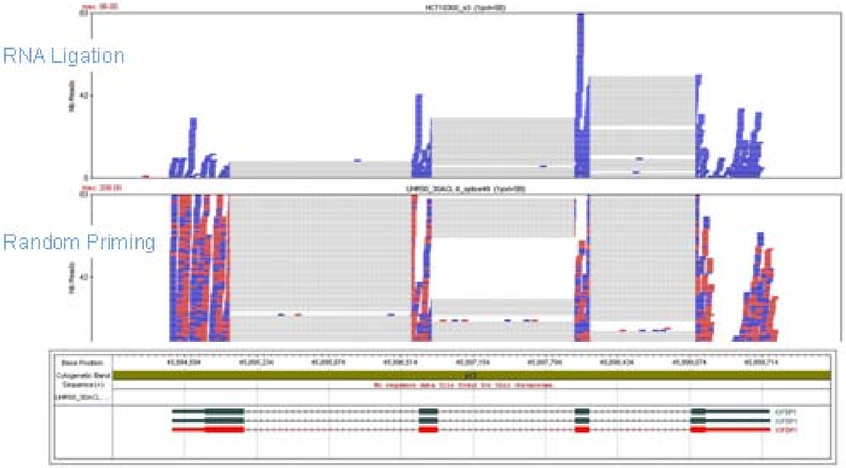
Random Priming

- Random Prime mRNA → cDNA
- Make 2nd Strand cDNA
- Ligate Sequencing Adapters

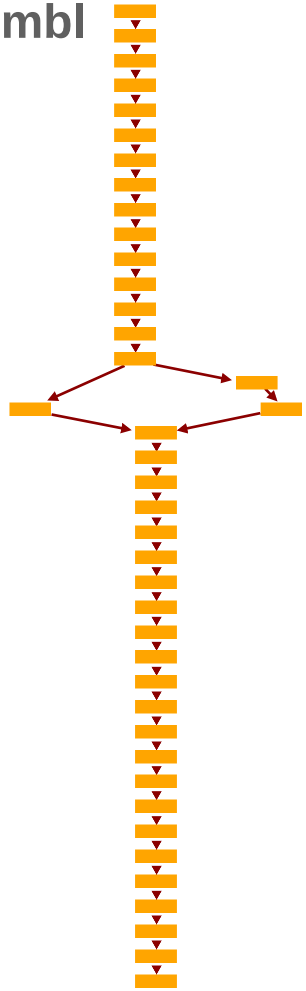
RNA Ligation

- CIP Treatment
- PNK Treatment
- Ligate 3'- Small RNA Adapters
- Ligate 5'- Small RNA Adapters
- RT to make cDNA

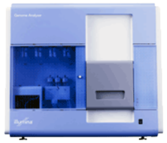
- PCR Amplify 15 Cycles
- Purify, Quantify
- Grow Clusters
- Sequence on Illumina Genome Analyzer
  - Single or Paired-end reads
  - 35 to 100 base read lengths

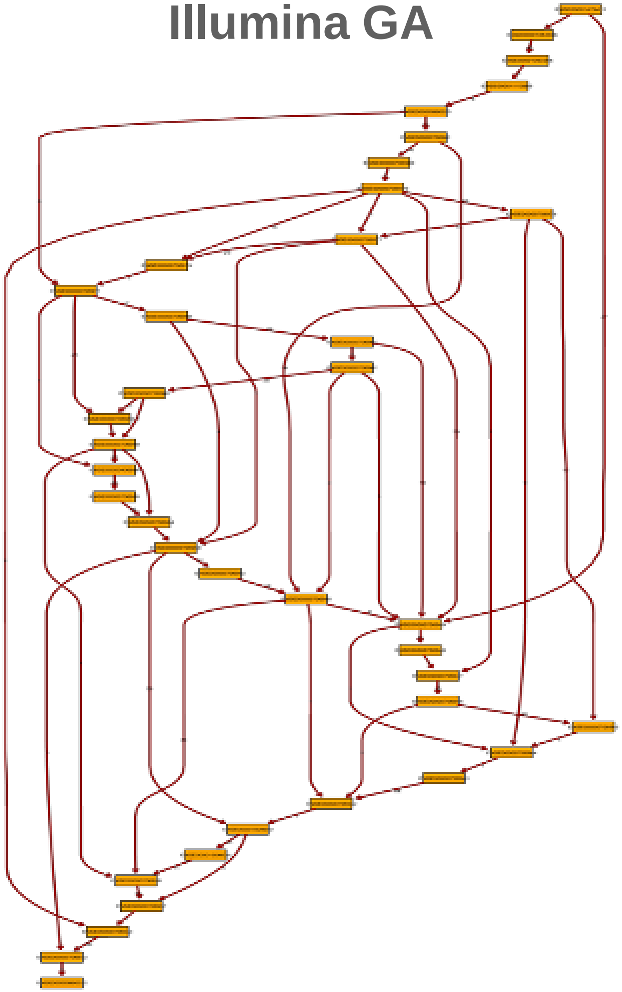


Ensembl

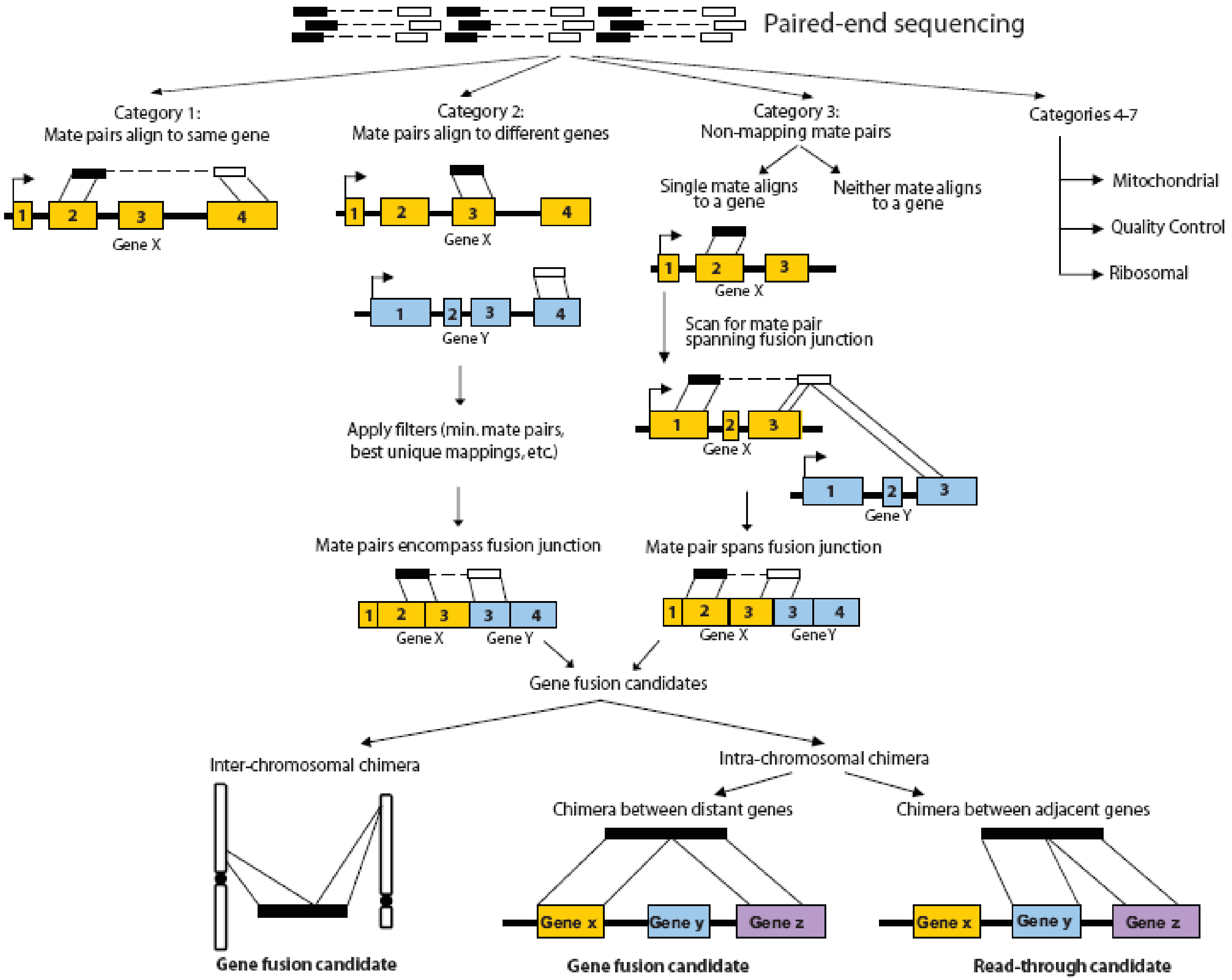


Illumina GA





Maher *et al*, **Chimeric transcript discovery by paired-end transcriptome sequencing** PNAS June 09



Illumina’s Small RNA v1.5 library prep protocol

Just 6 hr library prep starting from total RNA

- Enhanced workflow
  - Shortest time from library to sequence
  - 2.5 hrs hands on time (6 hr total)
  - Single gel excision
- NO need for small RNA purification!
  - Starts directly from total RNA
- Reduced RNA input requirements
  - 1.0 µg total RNA input
    - Titrated to 100 ng total
- As always, strand specific information
  - Discover & profile microRNA
  - Sequence other non-coding RNAs
- Highest quality data
  - Most accurate quantification of small RNAs
  - Superior results to RNA/DNA duplex method

Streamlined Workflow

3’ RNA Adaptor Ligation

↓

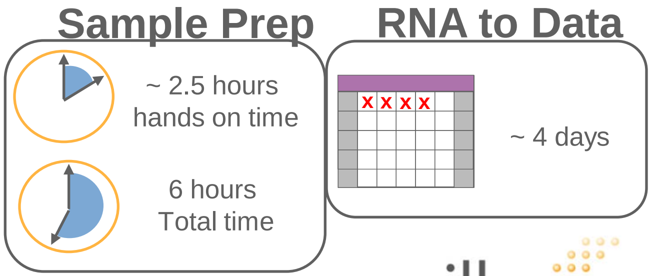
5’ RNA Adapter Ligation

↓

Perform RT-PCR Amplification

↓

Purify Small RNA Library



Sample Prep

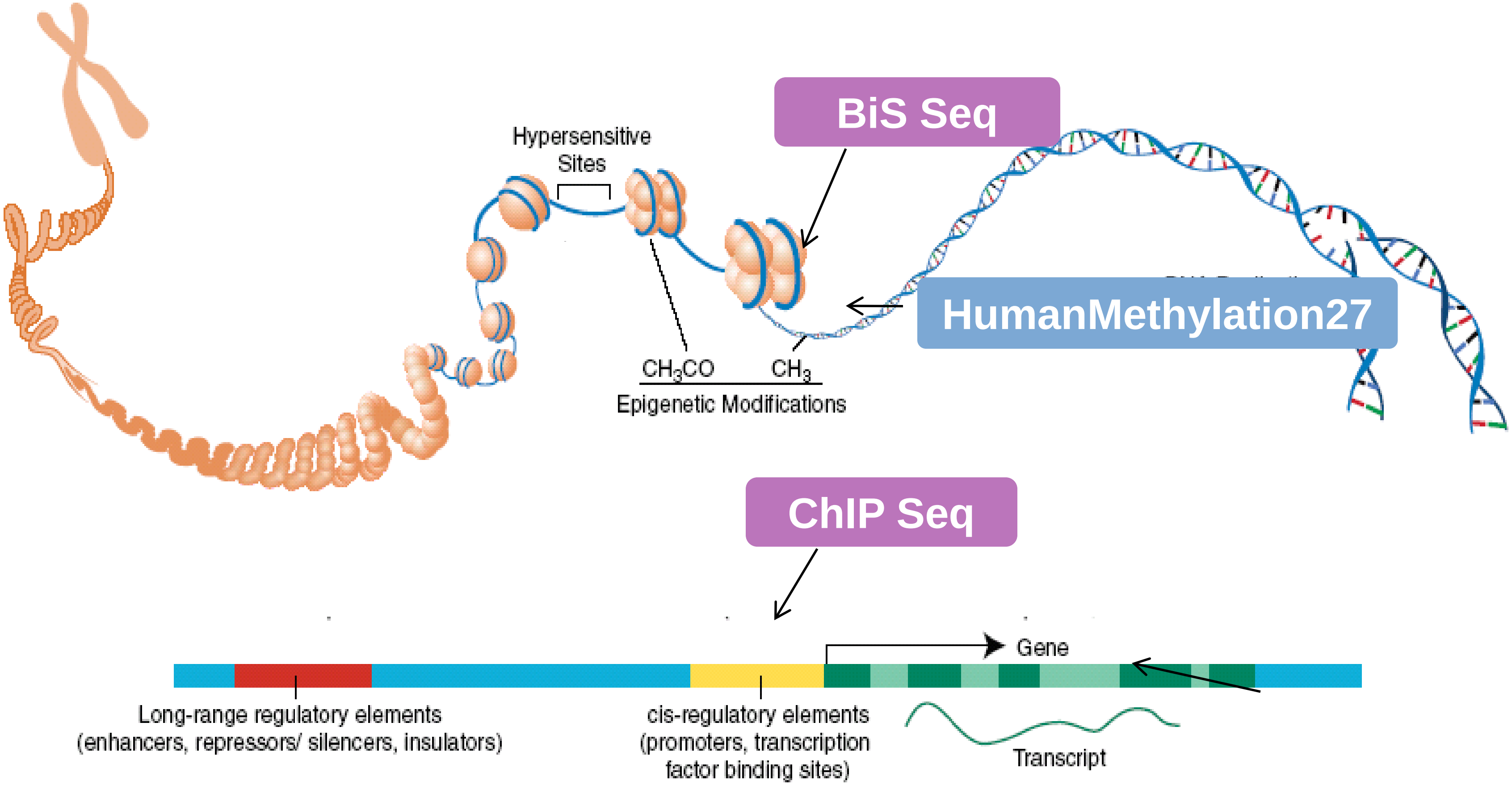
~ 2.5 hours hands on time

6 hours Total time

RNA to Data

~ 4 days

Gene Regulation (Epigenomics)



BiS Seq

HumanMethylation27

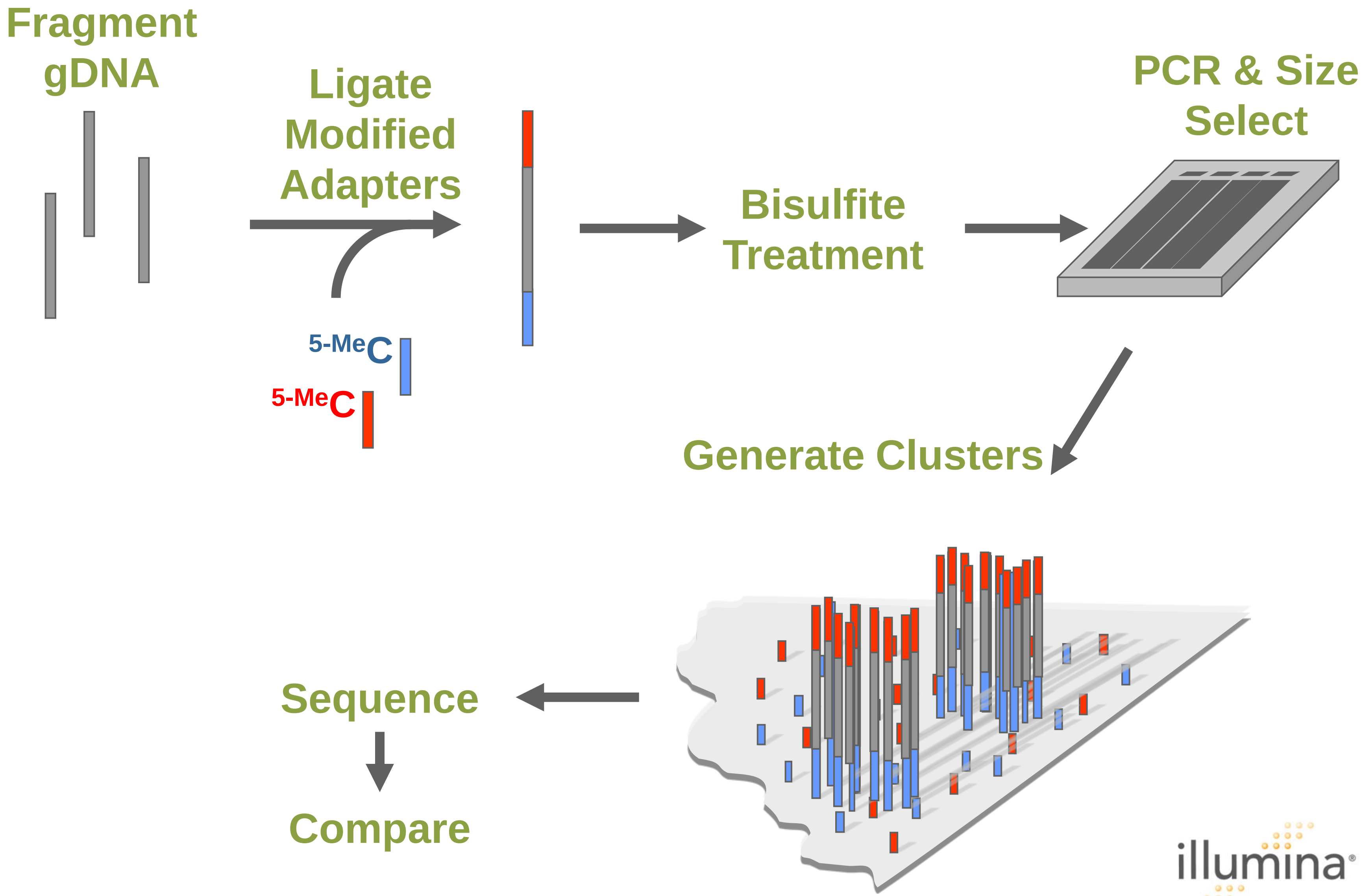
ChIP Seq

Whole Genome Methylation

HumanMethylation27

- High resolution
  - Based on the Infinium assay
  - Single CpG site resolution
  - Over 27,000 sites
- Low sample input
  - Works with samples from LCM, FACS
- Intelligent content
  - well-annotated genes
  - cancer targets
  - methylation hotspots

Bisulfite Sequencing



Fragment gDNA

Ligate Modified Adapters

Bisulfite Treatment

PCR & Size Select

Generate Clusters

Sequence

Compare

Bisulfite Sequencing

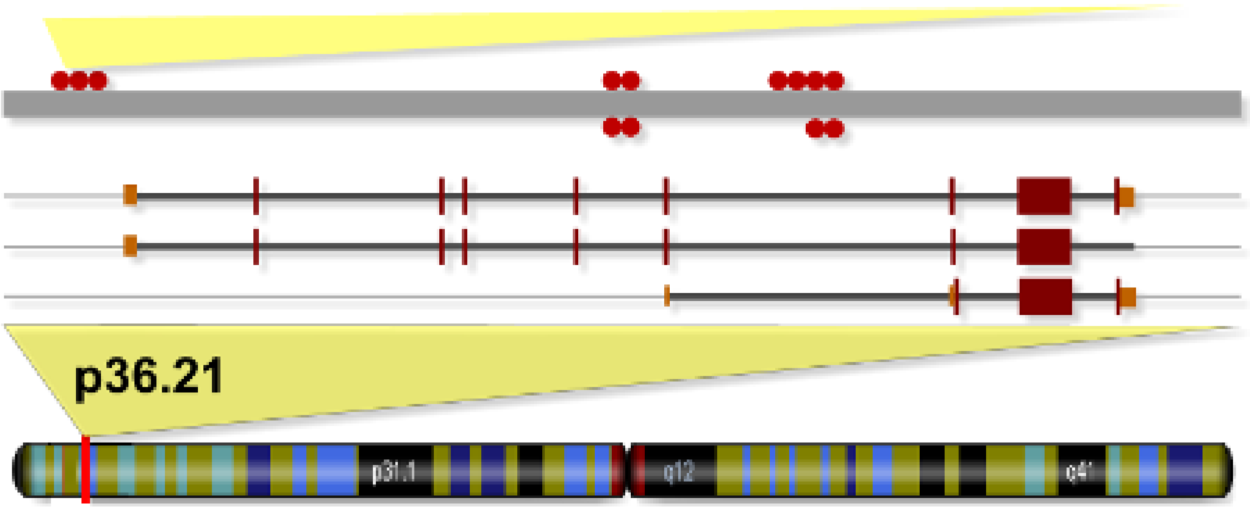
gctgaattacttcttggcacagcttgtctgaaatatgtaaagctaagtgc

Cell line 1

gCtgaattaTttTtggCaTagCttgtTtgaaatatgtaaagCtaagtgC

Cell line 2

gTtgaattaTttTtggTaTagTttgtTtgaaatatgtaaagTtaagtgT



Whole Genome Bisulfite Sequencing

Cell Lister *et al* '08

Highly Integrated Single-Base Resolution Maps of the Epigenome in *Arabidopsis*

nature Cokus *et al* '08

Shotgun bisulphite sequencing of the *Arabidopsis* genome reveals DNA methylation patterning

nature Meissner *et al* '09

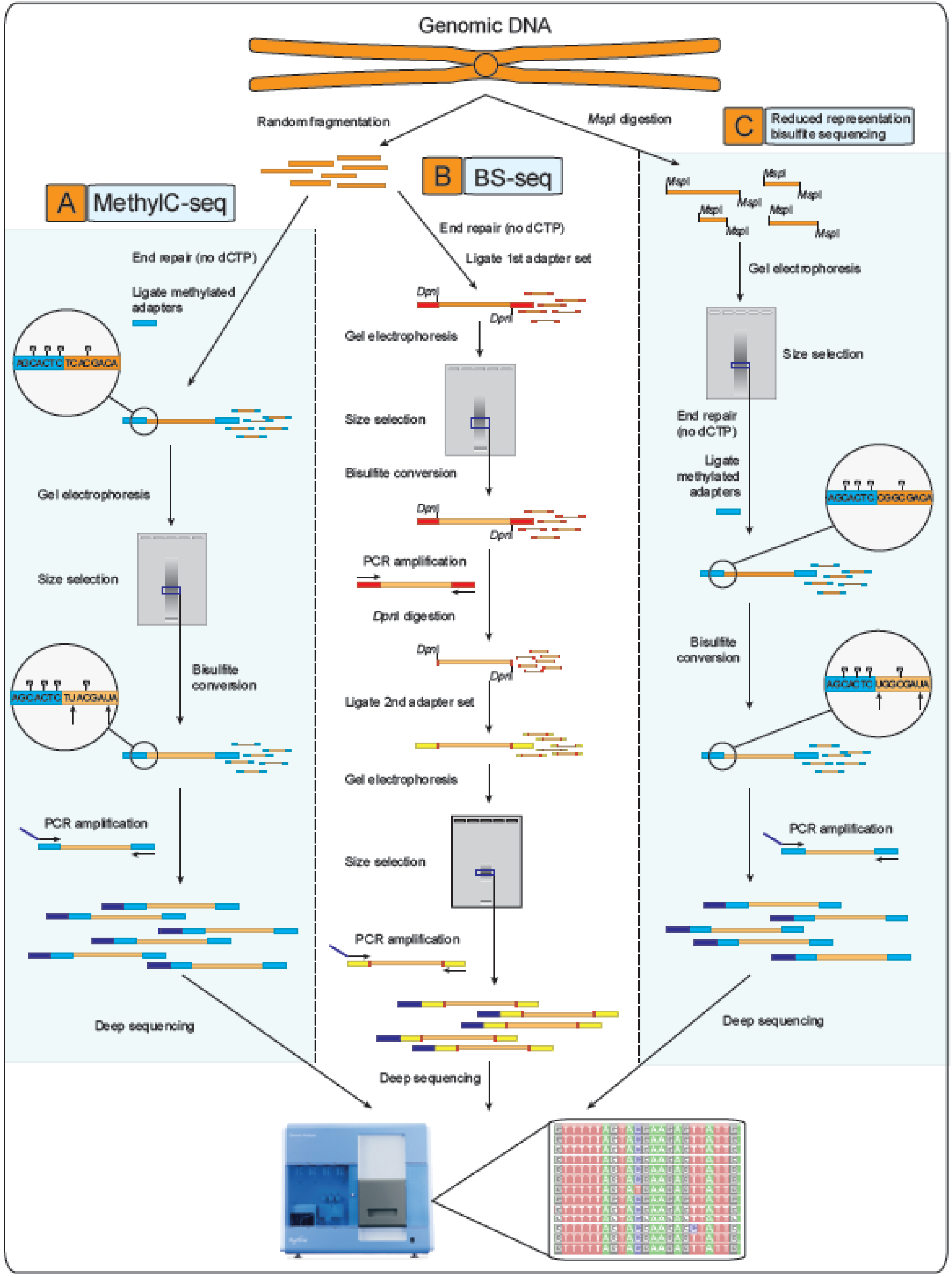
Genome-scale DNA methylation maps of pluripotent and differentiated cells

CSH PRESS GENOME RESEARCH Brunner *et al* '09*

Distinct DNA methylation patterns characterize differentiated human embryonic stem cells and developing human fetal liver

CSH PRESS GENOME RESEARCH Lister, Ecker '09

Finding the fifth base: Genome-wide sequencing of cytosine methylation



*not shown in figure

ChIP-Seq
(*Genome Wide Characterization of Transcription Factor Binding Sites and Epigenetic Modifications*)

- Superior Performance
  - High sensitivity (millions of tags per sample)
  - Very low background and false positive rate
- Excellent Coverage
  - Genome-wide coverage
  - High specificity/fine resolution
- Low Sample Input
  - 1 to 10 nanograms of DNA

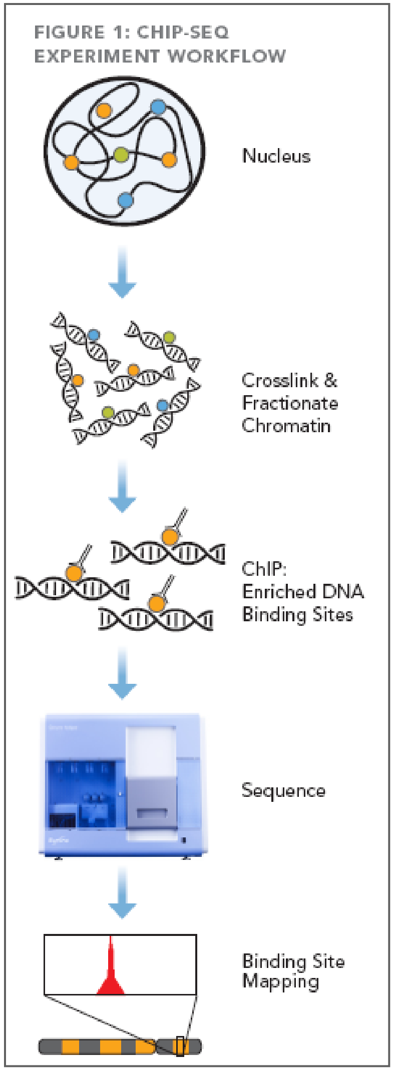
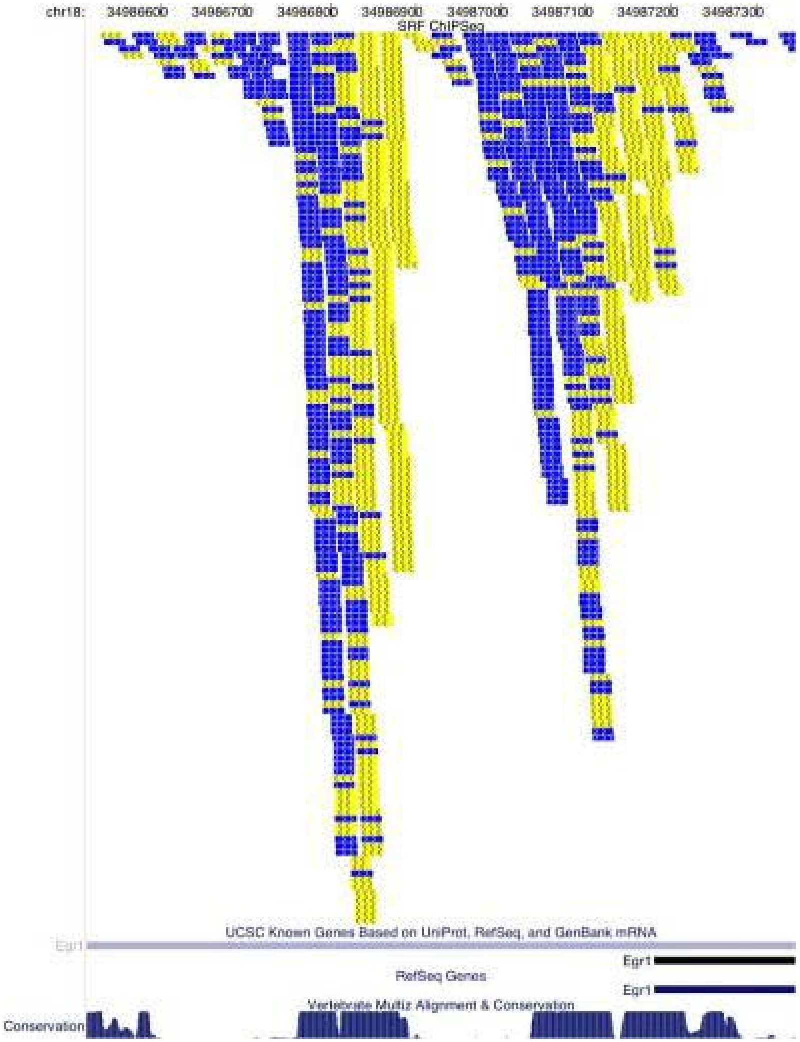


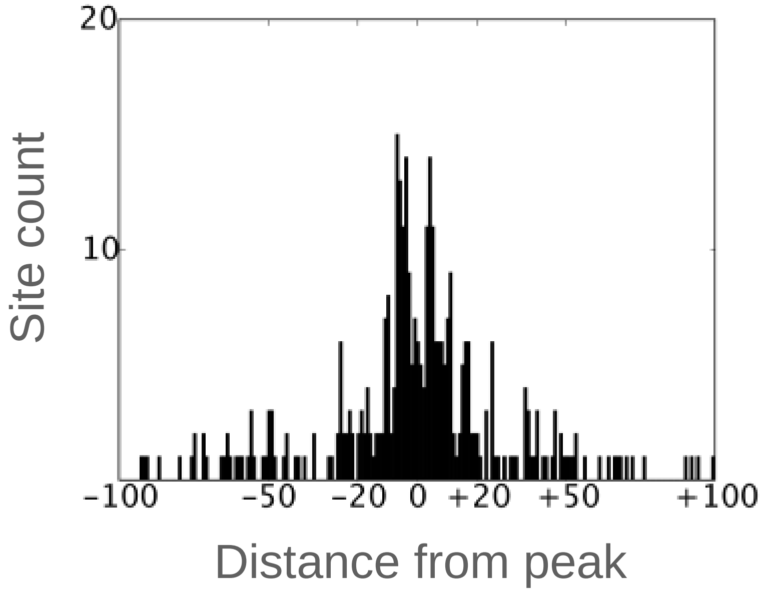
FIGURE 1: CHIP-SEQ EXPERIMENT WORKFLOW

ChIP-Seq:

Serum Response Factor (SRF) in Mouse C2C12 Cells



CarG Boxes in SRF-enriched Regions



4.7 M reads

Wold Lab

Renaissance in the analysis of regulatory elements

- High resolution mapping of regulatory elements including promoters, enhancers, control regions

- Nucleosome positioning
- Histone modification
- Transcription Factor binding
- Chromatin structure
- RNA-binding proteins

nature Vol 457|12 February 2009|doi:10.1038/nature07730

Please cite this article in press as: Boyle et al., High-Resolution Mapping and Characterization of Open Chromatin across the Genome, Cell (2008), doi:10.1016/j.cell.2007.12.014

ARTICLES

nature structural & molecular biology

An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells

Gene W Yeo[1,2,4], Nicole G Coufal[2], Tiffany Y Liang[1,3,4], Grace E Peng[2], Xiang-Dong Fu[3] & Fred H Gage[2]

The elucidation of a code for regulated splicing has been a long-standing goal in understanding the control of post-transcriptional gene expression events that are crucial for cell survival, differentiation and development. We decoded functional RNA elements *in vivo* by constructing an RNA map for the cell type–specific splicing regulator FOX2 (also known as RBM9) via cross-linking immunoprecipitation coupled with high-throughput sequencing (CLIP-seq) in human embryonic stem cells. The map identified a large cohort of specific FOX2 targets, many of which are themselves splicing regulators, and comparison between the FOX2 binding profile and validated splicing events revealed a general rule for FOX2-regulated exon inclusion or skipping in a position-dependent manner. These findings suggest that FOX2 functions as a critical regulator of a splicing network, and we further show that FOX2 is important for the survival of human embryonic stem cells.

Understanding regulated gene expression is vital to providing insights into disease and development. Whereas much effort has been placed on deciphering transcriptional regulation though interactions with functional DNA elements by the more than a thousand transcription factors encoded in mammalian genomes, little is known about an equally sizable number of RNA binding proteins and their involvement in diverse aspects of RNA metabolism. A dominant function of these RNA binding proteins is to regulate alternative splicing, a major form of post-transcriptional regulation of gene expression that is thought to contribute to the structural and functional diversity of the cellular proteome[1]. One of the ultimate goals in the RNA field is to deduce a set of rules that govern the control of splice-site selection to produce the 'splicing code'. This goal can now be approached due to recent advances in functional genomics and high-throughput sequencing.

Human embryonic stem cells (hESCs) are pluripotent cells that propagate perpetually in culture as undifferentiated cells and can be readily induced to differentiate into various cell types both in vitro and *in vivo*[2]. As hESCs can theoretically generate most if not all of the cell types that constitute a human, they serve as an excellent model for understanding early embryonic development. Furthermore, hESCs are a nearly infinite source for generating specialized cells such as neurons and glia for potential therapeutic purposes or for screening small molecules to intervene with specific biological processes[3,4]. Therefore, there has been intense interest in identifying the molecular changes that are important for the survival of hESCs, maintenance of pluripotency and promotion of cell differentiation.

In our previous *Affymetrix* exon-tiling array analysis, we demonstrated that the FOX binding motif GCAUG was enriched proximal to a set of exons that are alternatively spliced in hESCs, suggesting that FOX splicing factors may have a vital role in the biology of hESCs[5]. Thus, we selected the RNA binding protein FOX2 to identify the functional RNA elements in the human genome in hESCs by deep sequencing. FOX2, a member of the FOX family of RNA binding proteins, was initially identified as a factor involved in dosage compensation in *Caenorhabditis elegans* and was later found to be evolutionarily conserved across mammalian genomes[6]. FOX2 is best known for its tissue-specific expression in muscle and neuronal cells and for its activity in regulated splicing in those highly differentiated cell types[6,7]. Unexpectedly, we found that FOX2 is expressed abundantly in the hESC lines HUES6 and H9, which are positive for the pluripotency markers OCT4, SOX2, NANOG and SSEA4 (Fig. 1a and Supplementary Fig. 1 online). In contrast, FOX1 (also known as A2BP1) is not expressed in any hESCs examined. Consistent with their tissue-specific expression in cells of the neural lineage, both FOX1 and FOX2 are expressed in neural progenitors.

RESULTS

CLIP-seq for mapping functional RNA elements

We began to address the function of FOX2 in hESCs by developing a high-throughput experimental approach to large-scale identification of FOX2 targets *in vivo*, by coupling a modified CLIP technology[8] with high-throughput sequencing, a method we refer to as CLIP-seq (Fig. 1b). Key features of CLIP include: stabilization of *in vivo*

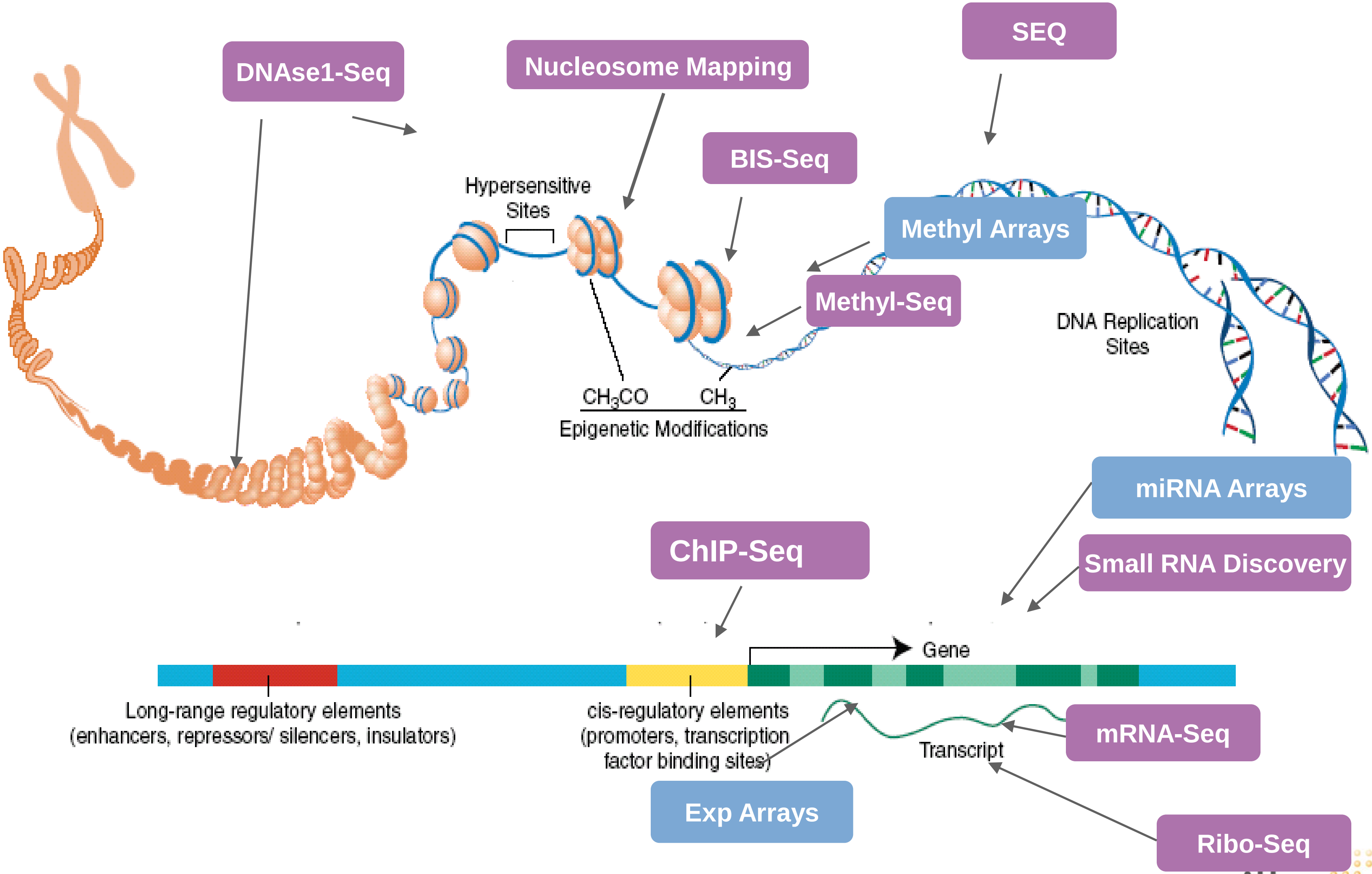
npg

---

[1] Crick-Jacobs Center for Theoretical and Computational Biology, Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [2] Laboratory of Genetics, Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA. [3] Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-5004, USA. [4] Present address: Stem Cell Program, Department of Cellular and Molecular Medicine, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093-5004, USA. Correspondence should be addressed to G.W.Y. (geneyeo@ucsd.edu), F.H.G. (gage@salk.edu) or X.-D.F. (xdfu@ucsd.edu).

Received 27 October 2008; accepted 11 December 2008; published online 11 January 2009; doi:10.1038/nsmb.1545

Ilumina's Complete Expression and Regulation Solution



DNAse1-Seq

Nucleosome Mapping

SEQ

BIS-Seq

Methyl Arrays

Methyl-Seq

miRNA Arrays

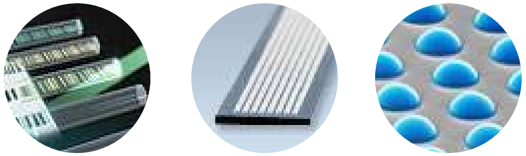
Small RNA Discovery

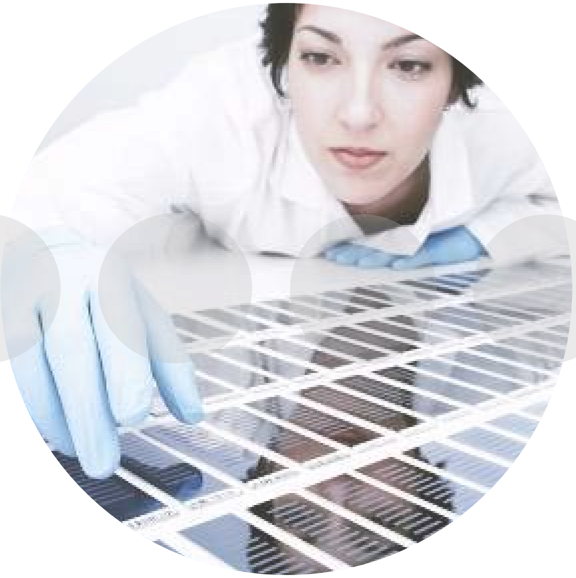
ChIP-Seq

mRNA-Seq

Exp Arrays

Ribo-Seq





Thank You!