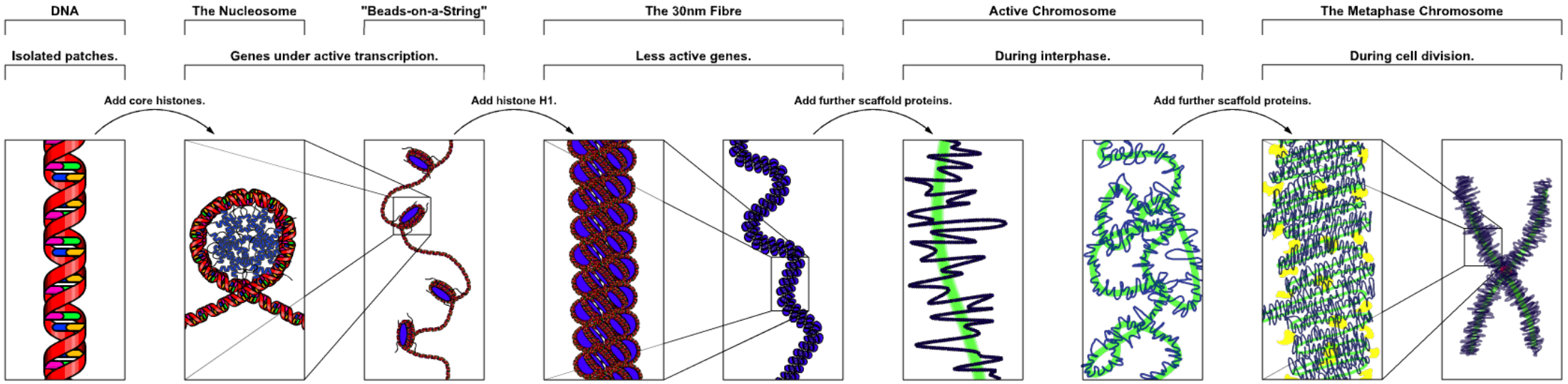
An Overview
on Experimental Subtypes and Variations
of ChIP-Seq

Alexei Lobanov, PhD, CCBR

The *DNA + histone = chromatin*

Chromatin's structure is currently poorly understood despite being subjected to intense investigation.



Nobel Prizes for chromatin research:

Year	Who	Award
1910	Albrecht Kossel (University of Heidelberg)	Nobel Prize in Physiology or Medicine for his discovery of the five nuclear bases: adenine, cytosine, guanine, thymine, and uracil.
1933	Thomas Hunt Morgan (California Institute of Technology)	Nobel Prize in Physiology or Medicine for his discoveries of the role played by the gene and chromosome in heredity, based on his studies of the white-eyed mutation in the fruit fly *Drosophila*.[29]
1962	Francis Crick, James Watson and Maurice Wilkins (MRC Laboratory of Molecular Biology, Harvard University and London University respectively)	Nobel Prize in Physiology or Medicine for their discoveries of the double helix structure of DNA and its significance for information transfer in living material.
1982	Aaron Klug (MRC Laboratory of Molecular Biology)	Nobel Prize in Chemistry "for his development of crystallographic electron microscopy and his structural elucidation of biologically important nucleic acid-protein complexes"
1993	Richard J. Roberts and Phillip A. Sharp	Nobel Prize in Physiology "for their independent discoveries of split genes," in which DNA sections called exons express proteins, and are interrupted by DNA sections called introns, which do not express proteins.
2006	Roger Kornberg (Stanford University)	Nobel Prize in Chemistry for his discovery of the mechanism by which DNA is transcribed into messenger RNA.

Regulatory DNA in humans:

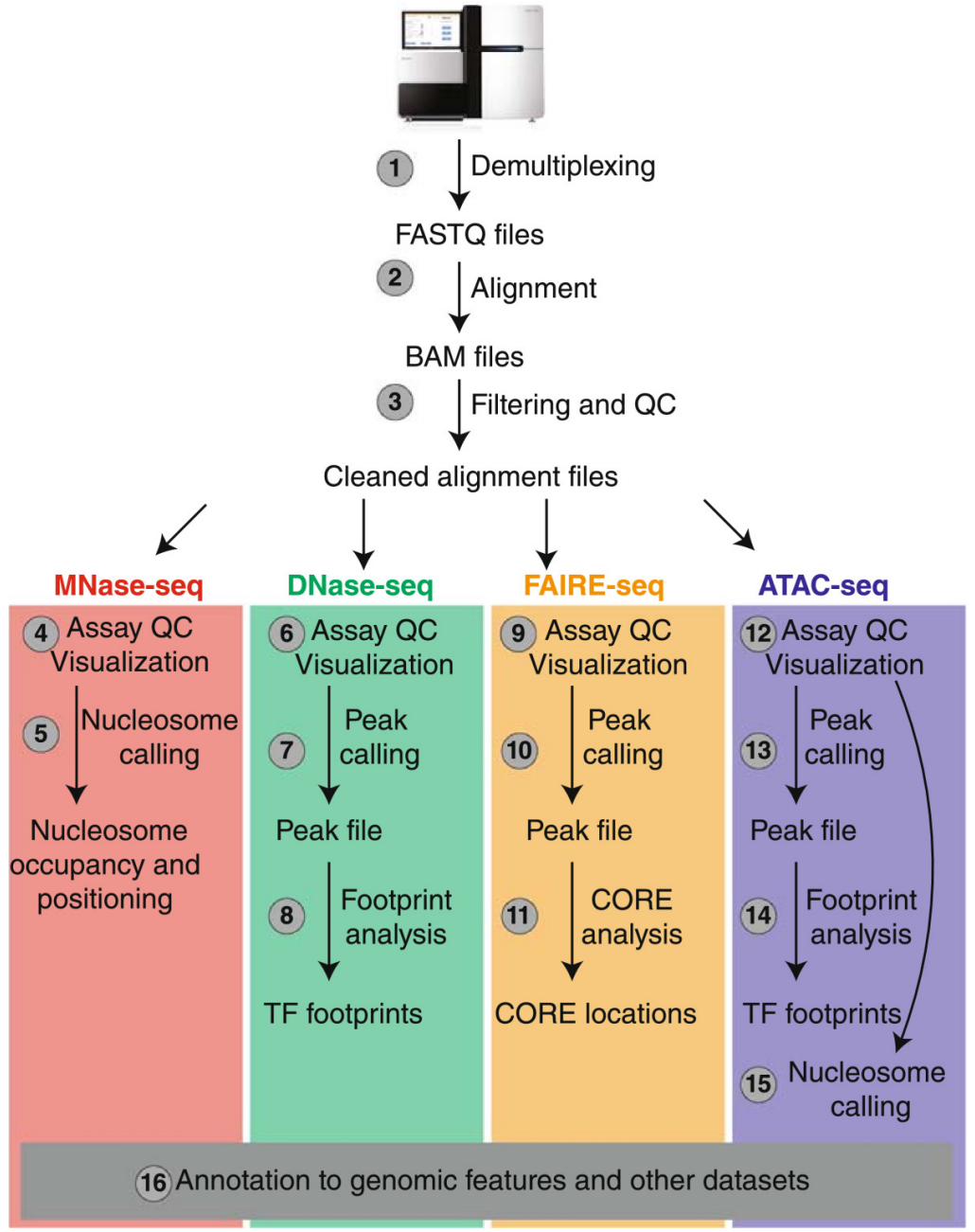
•**Transcription factor**: Using the ChIP-Seq technique, the binding sites to DNA in certain transcription factor groups are determined, and the DHS profiles are compared. The results confirm a high correlation, which show that the coordinated union of certain factors is implicated in the remodeling and accessibility of chromatin.
•**DNA methylation patterns**: CpG methylation has been closely linked with transcriptional silencing. This methylation causes a rearrangement of the chromatin, condensing and inactivating it transcriptionally. Methylated CpG falling within DHSs impedes the association of transcription factor to DNA, inhibiting the accessibility of chromatin. Data argue that methylation patterning paralleling cell-selective chromatin accessibility results from passive deposition after the vacation of transcription factors from regulatory DNA.
•**Promoter chromatin signature**: The H3K4me3 modification is related with transcriptional activity. This modification takes place in adjacent nucleosome to the transcription start site (TSS), relaxing the chromatin structure. This histone modification is used as a marker of promoters, using it to map these elements in the human genome.
•**Promoter/enhancer connections**: distal cis-regulatory elements, such as enhancers are in charge of modulating the activity of the promoters. In this way, the distal cis-regulatory elements are actively synchronized with their promoter in the cellular lines which is active the expression of the gene controlled. Using the DHS profiles, were looked for correlations between DHS to identify promoter/enhancer connections. Thus, it was able to create a map of candidate enhancers controlling specific genes.

Methods to investigate chromatin:

1. **ChIP-seq** (Chromatin immunoprecipitation sequencing), aimed against different histone modifications, can be used to identify chromatin states throughout the genome. Different modifications have been linked to various states of chromatin.
2. **DNase-seq** (DNase I hypersensitive sites Sequencing) uses the sensitivity of accessible regions in the genome to the DNase I enzyme to map open or accessible regions in the genome.
3. **FAIRE-seq** (Formaldehyde-Assisted Isolation of Regulatory Elements sequencing) uses the chemical properties of protein-bound DNA in a two-phase separation method to extract nucleosome depleted regions from the genome.
4. **ATAC-seq** (Assay for Transposable Accessible Chromatin sequencing) uses the Tn5 transposase to integrate (synthetic) transposons into accessible regions of the genome consequentially highlighting the localisation of nucleosomes and transcription factors across the genome.
5. **MNase-seq** (Micrococcal Nuclease sequencing) uses the micrococcal nuclease enzyme to identify nucleosome positioning throughout the genome.
6. **DNA footprinting** is a method aimed at identifying protein-bound DNA. It uses labeling and fragmentation coupled to gel electrophoresis to identify areas of the genome that have been bound by proteins.
7. **Chromosome conformation capture** determines the spatial organization of chromatin in the nucleus, by inferring genomic locations that physically interact.
8. **MACC profiling** (Micrococcal nuclease ACCessibility profiling) uses titration series of chromatin digests with micrococcal nuclease to identify chromatin accessibility as well as to map nucleosomes and non-histone DNA-binding proteins in both open and closed regions of the genome.

Outline:

- DNase-seq
- FAIRE-seq
- ATAC-seq
- MNase-seq



Tsompana and Buc, Epigenetics & Chromatin (2014)

The Popularity of Methods:

chip seq
Search term

dnase seq
Search term

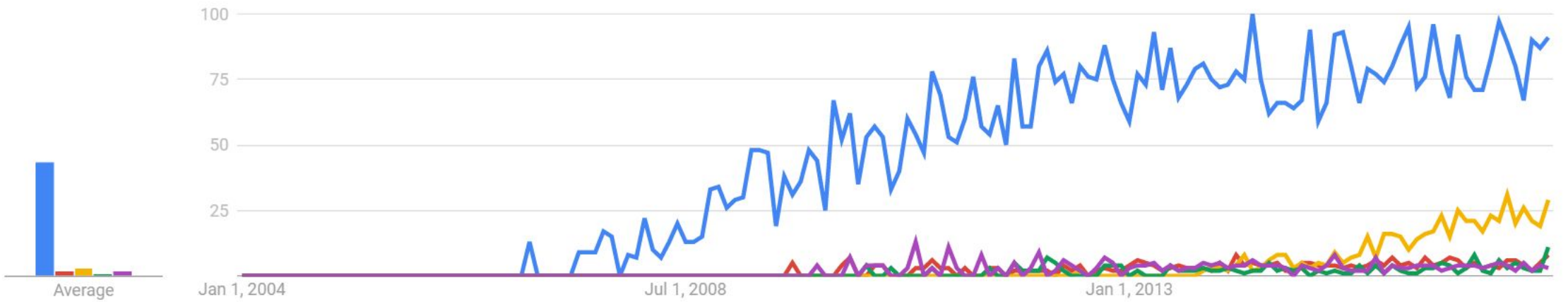
ATAC seq
Search term

MNase seq
Search term

FAIRE seq
Search term

Worldwide | 2004 - present | All categories | Web Search

Interest over time



DNase-seq

DNase I hypersensitive sites **seq**uencing

Cell

Volume 132, Issue 2, 25 January 2008, Pages 311–322

Resource

High-Resolution Mapping and Characterization of Open Chromatin across the Genome

Alan P. Boyle[1], Sean Davis[3], Hennady P. Shulha[2], Paul Meltzer[3], Elliott H. Margulies[4], Zhiping Weng[2], Terrence S. Furey[1], , , Gregory E. Crawford[1], , ,

DNase-seq

GOAL:

identify the location of regulatory regions,
based on the genome-wide sequencing of regions sensitive to cleavage by DNase I

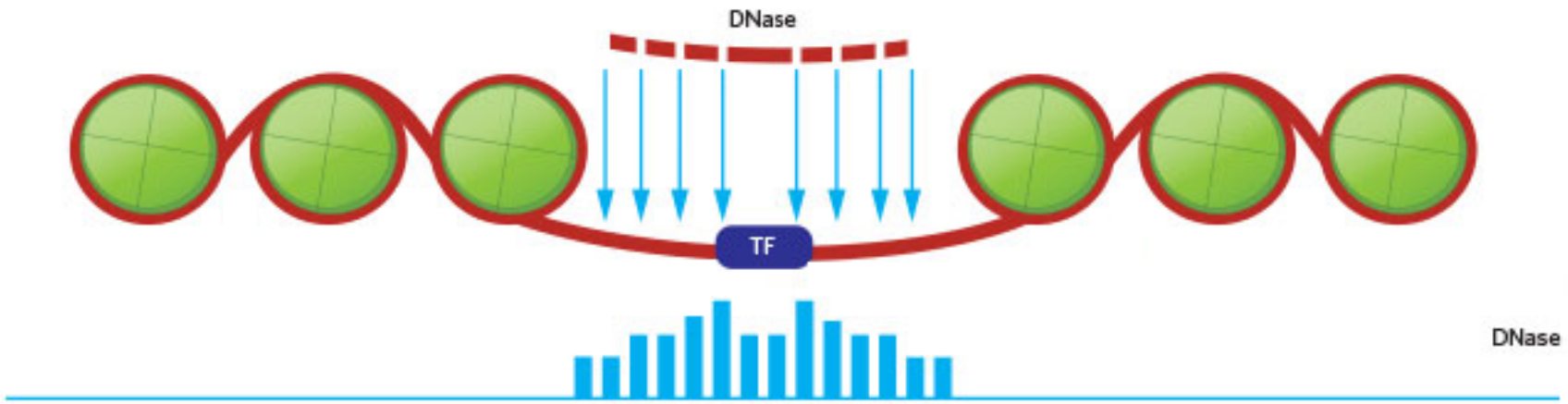
(since the discovery of such regions 30 years ago, they have been used as markers of regulatory DNA regions)

DNase-seq Footprinting:

Segmentation-based methods (HINT, Boyle method and Neph method):
use of sliding window to segment the genome into open/closed chromatin region.

Site-centric methods (CENTIPEDE, Cuellar-Partida method):
find footprints given the open chromatin profile around motif-predicted binding sites,
i.e., regulatory regions predicted using DNA-protein sequence information (Position weight matrix).

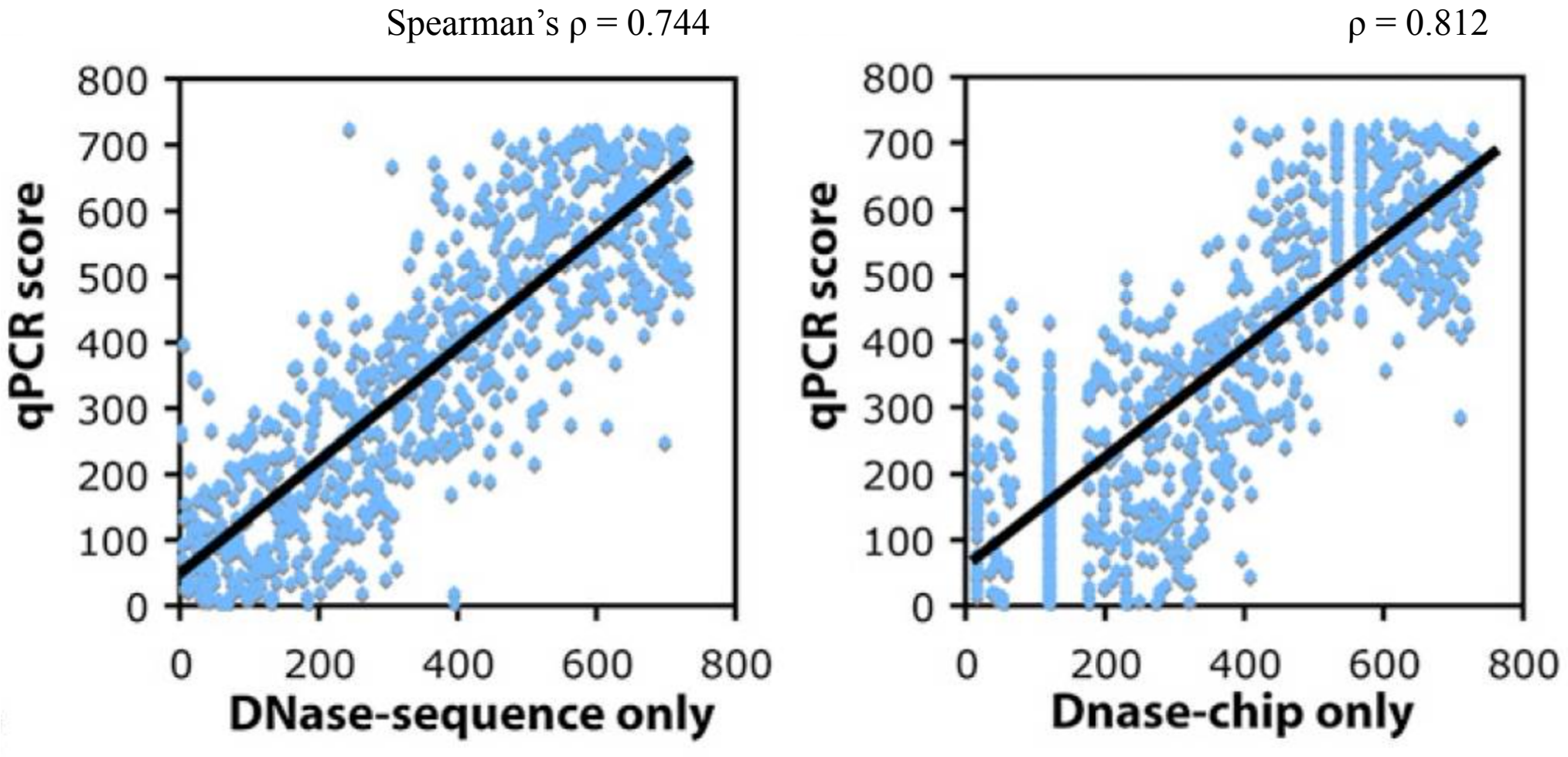
DNase-seq



How it works: DNase-seq takes advantage of the fact that exposed regions of the genome are naturally more prone to degradation by DNases. The method employs the enzyme DNase I to cleave DNA at sites along the genome that are not wrapped around nucleosomes, which become displaced by the binding of transcription factors. These small fragments, which are thought to infer the presence of transcription factors, are then sequenced and mapped to the genome.

Modified from: www.the-scientist.com

DNase-seq and DNase-chip results correlate with qPCR



Boyle AP *et al*, Cell, 2008

DNase-seq potential limitations

- Fresh samples only!

- The efficiency of DNase-seq in identifying TF binding sites is highly dependent on fragment size;

- Requires careful calibration of digestion conditions; finding the optimal digestion conditions for a given cell type and number could be challenging.

- Also requires greater sequencing depth (fragments longer than 147 bp saturate at much higher levels than shorter reads), and thus it may be challenging to analyze rare patient samples.

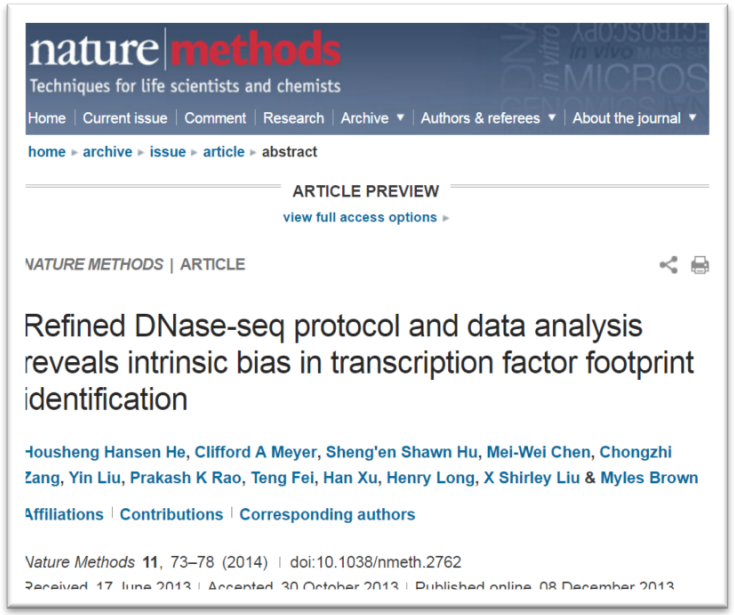
Problems and solutions:

- *What if you don't have recommended amount (50 million) of cells?*
There are reports of successful experiments with lower numbers. Experimental conditions need to be adjusted empirically, depending on cell type and quantity.

- *What if cells were lysed insufficiently (or excessive)?*
Different cell lines show different sensitivity. Test different concentrations on a smaller amount of cells (5 million) first.

- *What if DNA wasn't cut properly?*
The same as above: experimental conditions needs to be adjusted (enzyme concentration, cell amount, time)

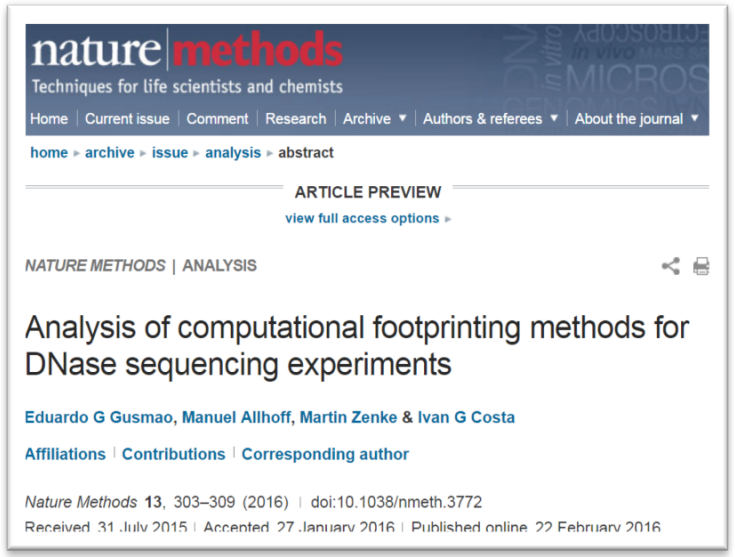
Footprinting: considerations



DNase I's cutting bias may limit the method's usefulness for the identification of DNA footprints.

(Analysis of binding of 36 different transcription factors revealed that DNase-seq data were not useful for illuminating footprints for many of them).

"If you see a pattern in the naked DNA, then you know it's got nothing to do with transcription-factor binding" © Clifford Meyer.



More recent analysis of DNase-seq data (> 80 TFs) indicated that a few advanced digital footprinting methods were not affected by cleveage bias.

"Our message is:
Don't look for footprints in raw DNAse-seq data, rather use state of art computational footpriting tools as HINT, DNAse2TF or PIQ"

Summary

DNase-seq has historically been a valuable tool for identifying all different types of **regulatory elements**, including promoters, enhancers, silencers, insulators and locus control regions.

DNase-seq is better established than any of the other chromatin accessibility methods.
It was applied to a wide range of cell types and species, including **plants**.

Its cutting **bias** is better understood.

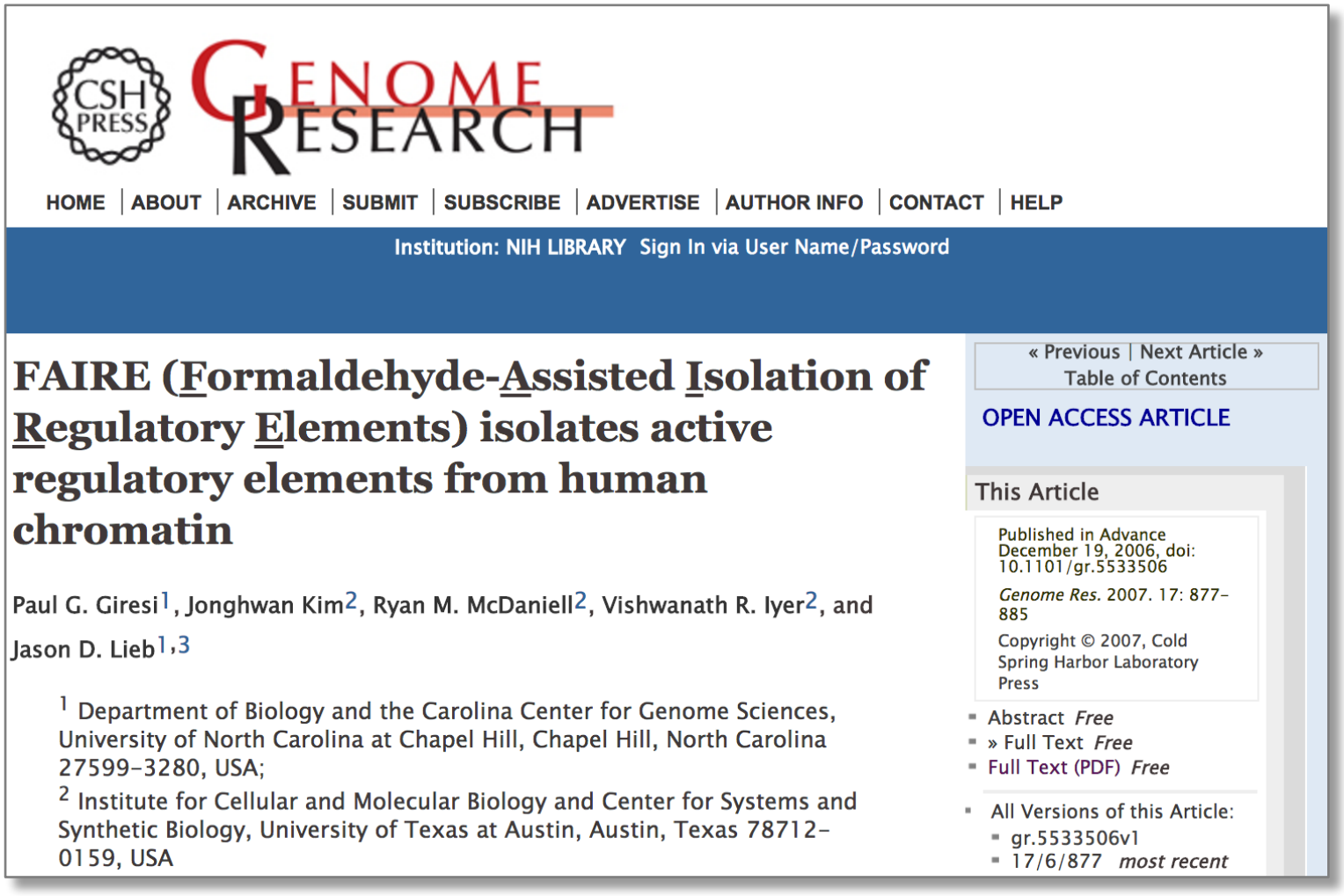
DNase-seq could be used for analysis of protected regions of the genome (DNase **footprinting**)

However, it does not directly explain **biological functions** of these elements. Therefore, it has to be supplemented by other methods.

The question if DNase-seq can replace **ChIP-seq** (and to which degree) is a subject of further investigations.

FAIRE-seq

Formaldehyde-Assisted Isolation of Regulatory Elements sequencing



CSH PRESS GENOME RESEARCH

HOME | ABOUT | ARCHIVE | SUBMIT | SUBSCRIBE | ADVERTISE | AUTHOR INFO | CONTACT | HELP

Institution: NIH LIBRARY Sign In via User Name/Password

« Previous | Next Article »
Table of Contents

OPEN ACCESS ARTICLE

FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin

Paul G. Giresi[1], Jonghwan Kim[2], Ryan M. McDaniell[2], Vishwanath R. Iyer[2], and Jason D. Lieb[1,3]

[1] Department of Biology and the Carolina Center for Genome Sciences, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3280, USA;
[2] Institute for Cellular and Molecular Biology and Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas 78712-0159, USA

This Article

Published in Advance December 19, 2006, doi: 10.1101/gr.5533506

Genome Res. 2007. 17: 877-885

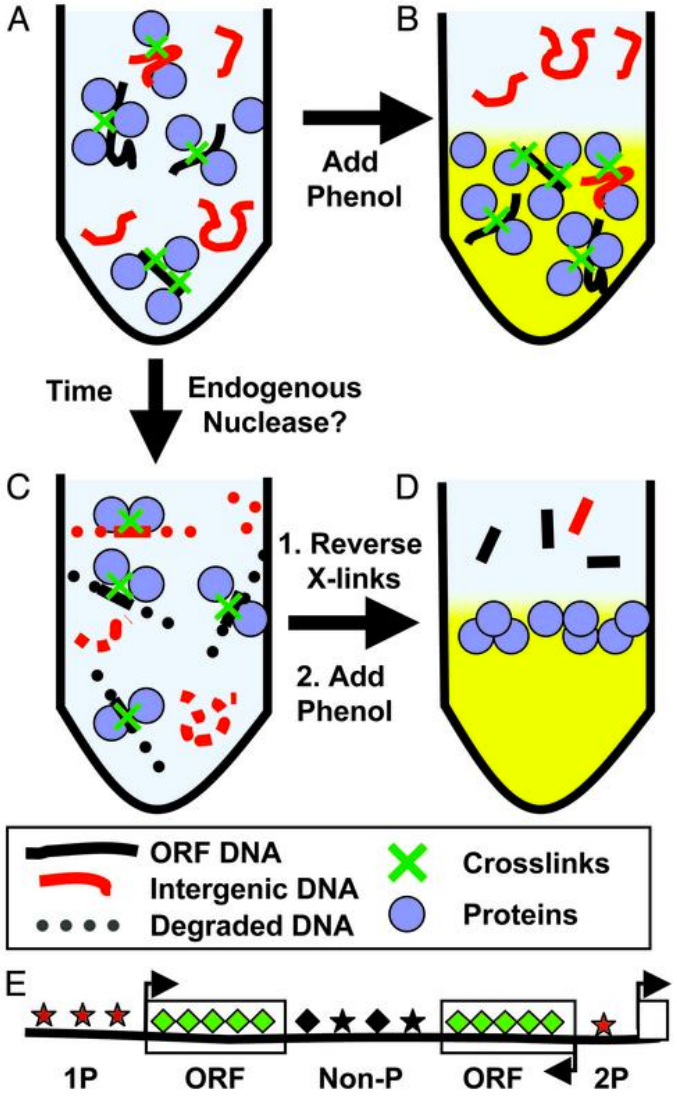
Copyright © 2007, Cold Spring Harbor Laboratory Press

- Abstract *Free*
- » Full Text *Free*
- Full Text (PDF) *Free*
- All Versions of this Article:
  - gr.5533506v1
  - 17/6/877 *most recent*

FAIRE

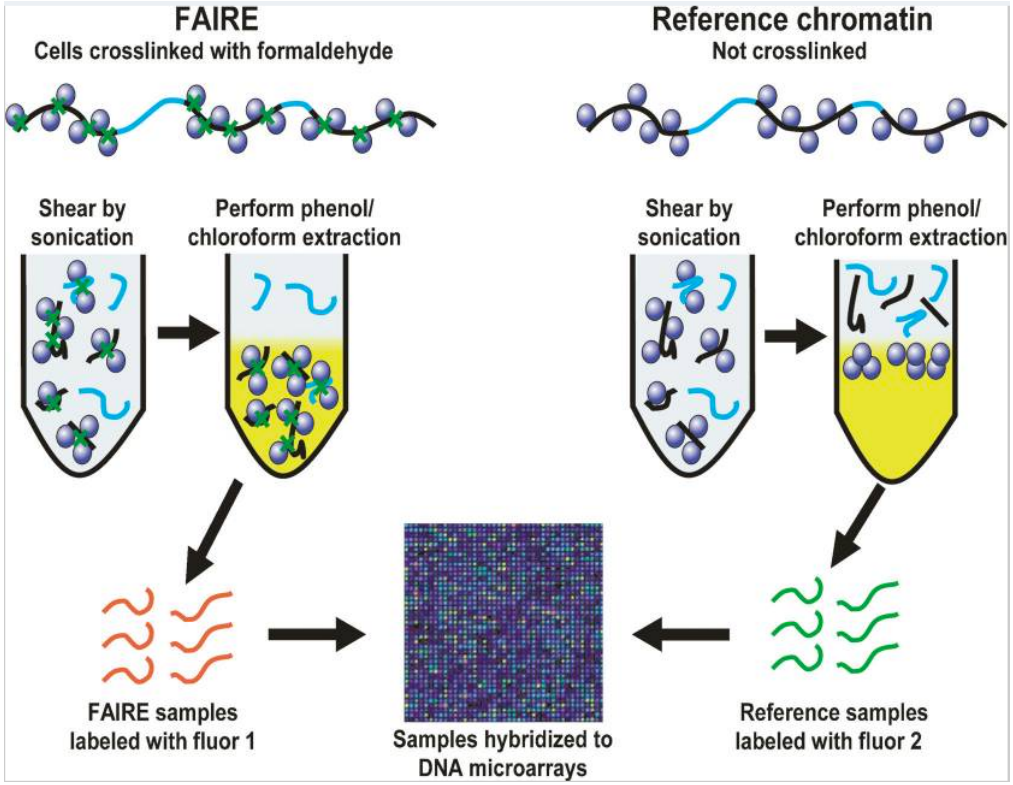
"We report here a genomewide approach for fractionating yeast chromatin into two functionally distinct parts, one containing RNA polymerase II transcribed sequences, and the other comprising noncoding sequences and genes transcribed by RNA polymerases I and III"

Nagy *et al*, PNAS USA (2003)



FAIRE-seq

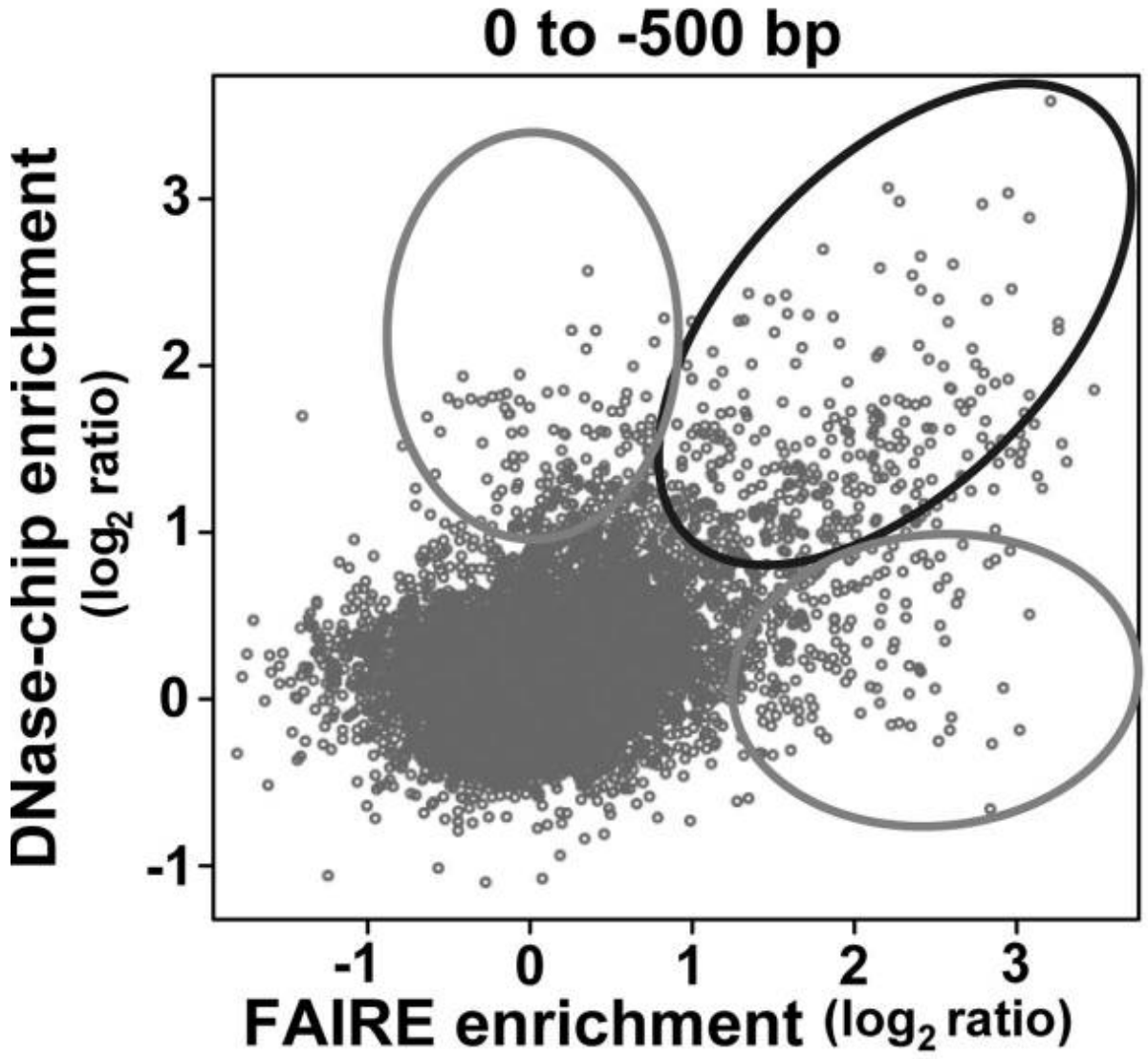
GOAL:
determine the sequences of DNA regions
in the genome associated with regulatory activity



In contrast to DNase-seq, the FAIRE-seq protocol doesn't require the permeabilization of cells or isolation of nuclei, and can analyze any cell types

In a study of seven diverse human cell types, DNase-seq and FAIRE-seq produced strong cross-validation, with each cell type having 1-2% of the human genome as open chromatin.

Cell-type specific differences identified by FAIRE



Giresi et al, Genome Research (2007)

Advantages:

Antibody and enzyme independency. In contrast to ChIP, which is highly subject to antibody reliability and variability issues, FAIRE offers the consistency of a chemical-based isolation. Moreover, FAIRE does not require enzymes, such as DNase or MNase, which are commonly used in analogous methods for detecting nucleosome-free regions.

Enhancer detection. FAIRE may identify additional transcriptional enhancers and other distal regulatory elements in comparison to other methods such as DNase-seq.

Sequenced input control not required. A sequenced input control is generally not required for proper analysis of FAIRE-enriched regions. This reduces next-generation sequencing costs as well as the cost of reagents.

Applicability to tissue samples. As FAIRE does not require a single-cell suspension or nuclear isolation, it is easily adapted for use on tissue samples. The only additional step needed is pulverization of frozen tissue into a coarse powder before fixation.

Reproducibility. Avoiding the optimization and extra steps necessary for enzymatic processing or immunoprecipitations eliminates a major source of variation, and thus makes it a much more reliable and robust method. FAIRE is remarkably reproducible from experiment to experiment.

Limitations:

Promoter detection. Other methods, such as DNase-seq, may be better at identifying nucleosome-depleted promoters of highly expressed genes.

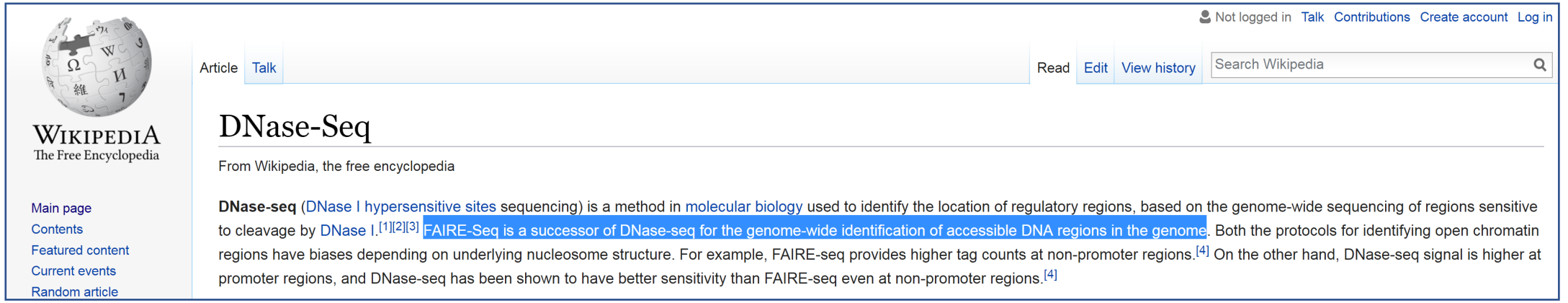
Analysis. Although FAIRE is relatively straightforward experimentally, an extensive amount of computational processing and analysis is required for comprehensive interpretation of genome-wide results. Quantification of FAIRE signal by qPCR or microarrays may be more straightforward.

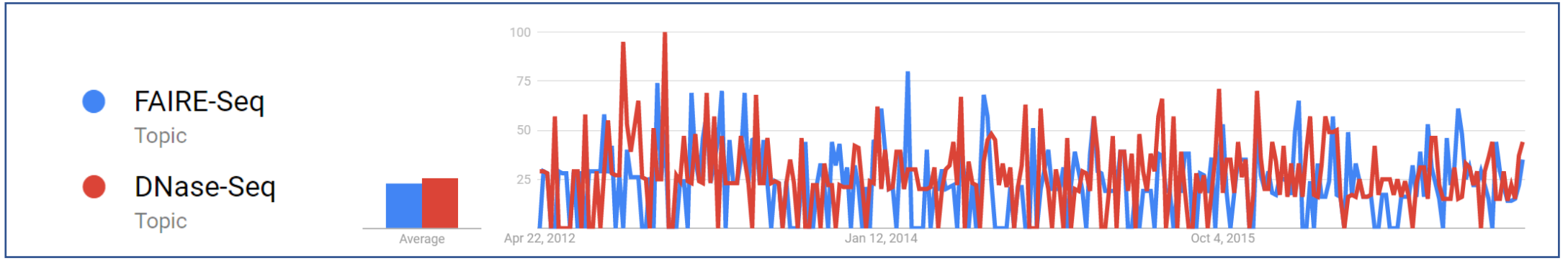
Absence of transcription factor footprinting. Transcription factor motifs can be identified in regions of open chromatin identified by FAIRE. However, the higher resolution and increased signal-to-noise of DNase-seq permits detection of specific transcription factor footprints in very deeply sequenced data.

Low signal-to-noise ratio. The sites detected by FAIRE can, at times, be only marginally enriched above the background signal. This leads to a reduced confidence in the sites identified.

Fixation variation among tissues. Fixation efficiency can vary considerably for many reasons, including differences in cellularity, permeability, purity, fat content and surface area. This variability can lead to inconsistent results; optimization is thus recommended.

Remarks

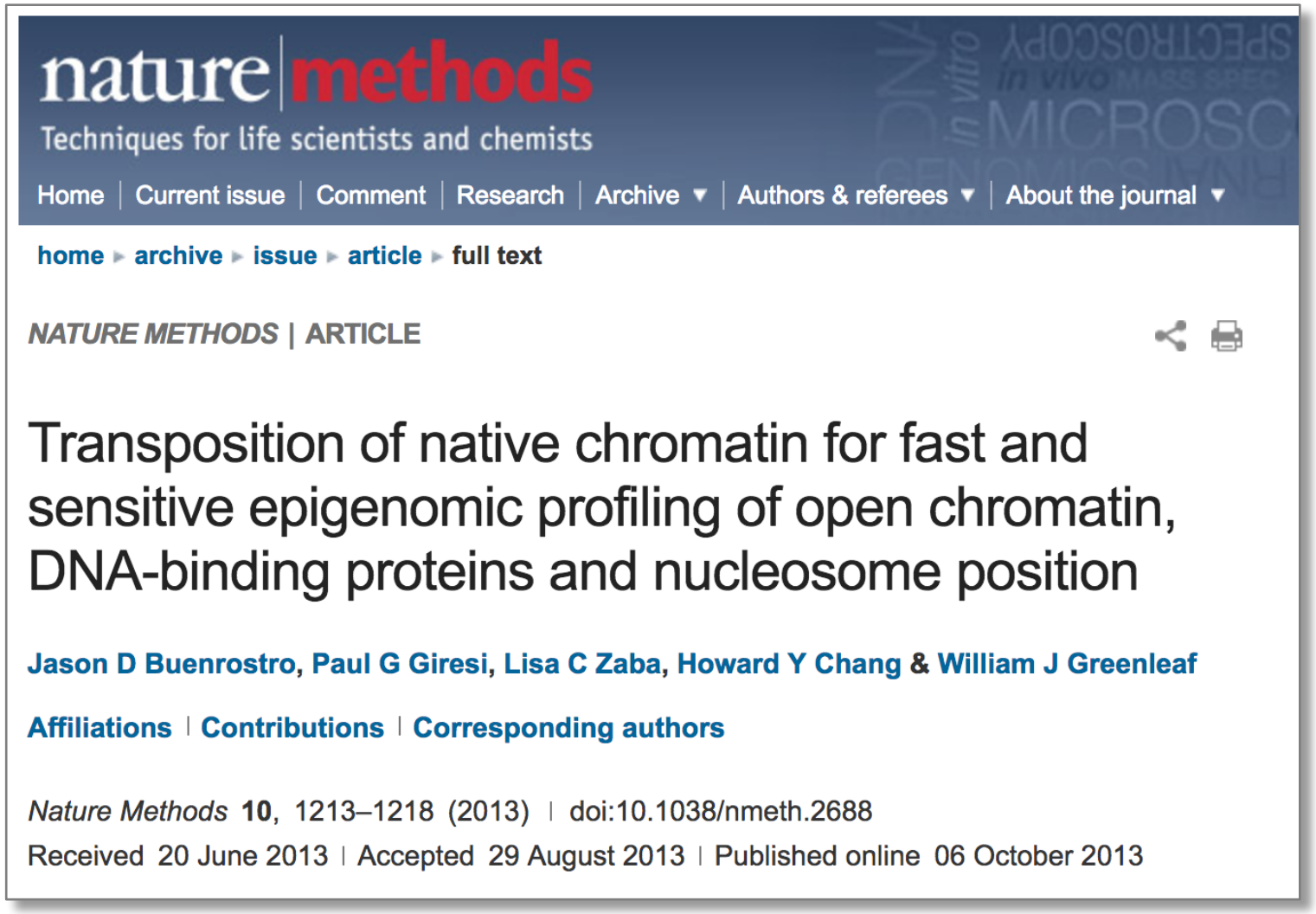




The University of Chicago's Jason Lieb, who developed FAIRE-seq several years ago, has mostly switched over to the ATAC-seq because it gives a better signal.

ATAC-seq

Assay for
Transposase-
Accessible
Chromatin
with high throughput
sequencing



nature methods
Techniques for life scientists and chemists

NATURE METHODS | ARTICLE

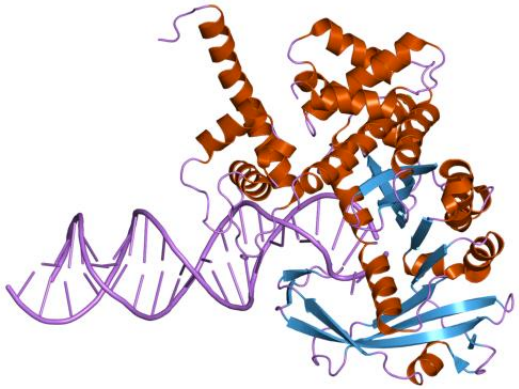
Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position

Jason D Buenrostro, Paul G Giresi, Lisa C Zaba, Howard Y Chang & William J Greenleaf

Affiliations | Contributions | Corresponding authors

Nature Methods **10**, 1213–1218 (2013) | doi:10.1038/nmeth.2688
Received 20 June 2013 | Accepted 29 August 2013 | Published online 06 October 2013

ATAC-seq

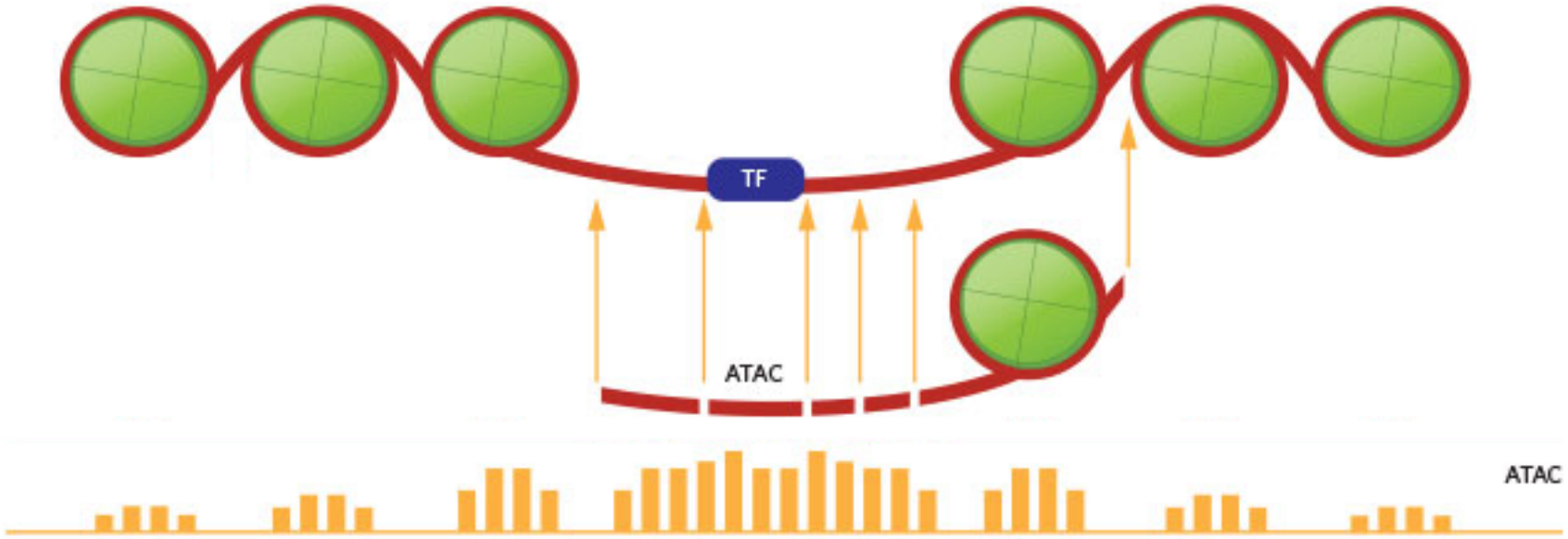


GOAL:
identify accessible DNA regions
(equivalent to Dnase I hypersensitive sites)

The key part of the ATAC-seq procedure is the action of the **transposase Tn5**
on the genomic DNA of the sample
(transposases are enzymes catalyzing the movement of transposons to other parts in the genome)

ATAC-seq employs a mutated hyperactive transposase
(naturally occurring transposases have a low level of activity)

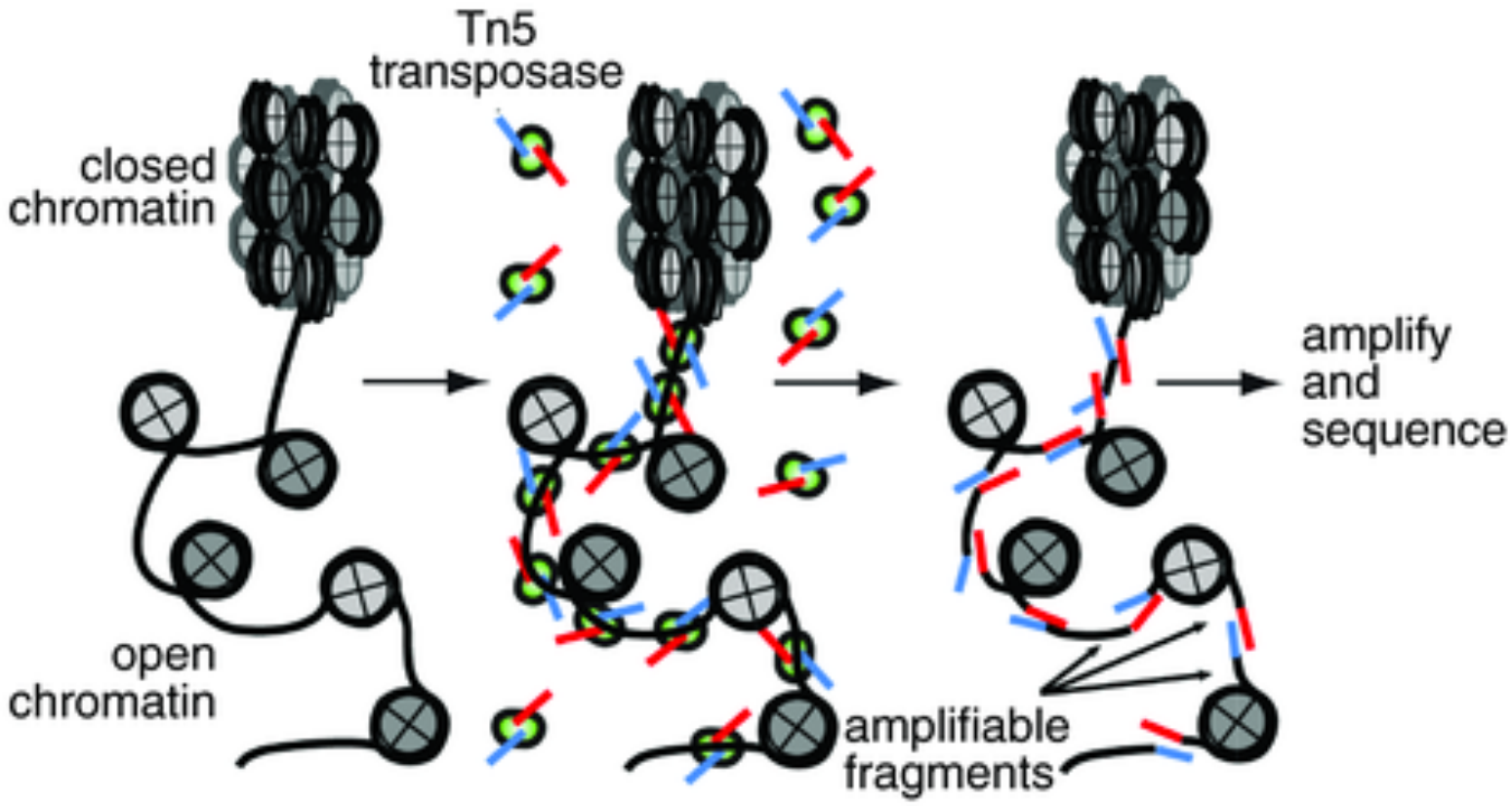
ATAC-seq



How it works: ATAC-seq inserts sequencing adapters directly into accessible DNA using the enzyme Tn5 transposase. The bits captured between the adapters are then amplified with qPCR and sequenced.

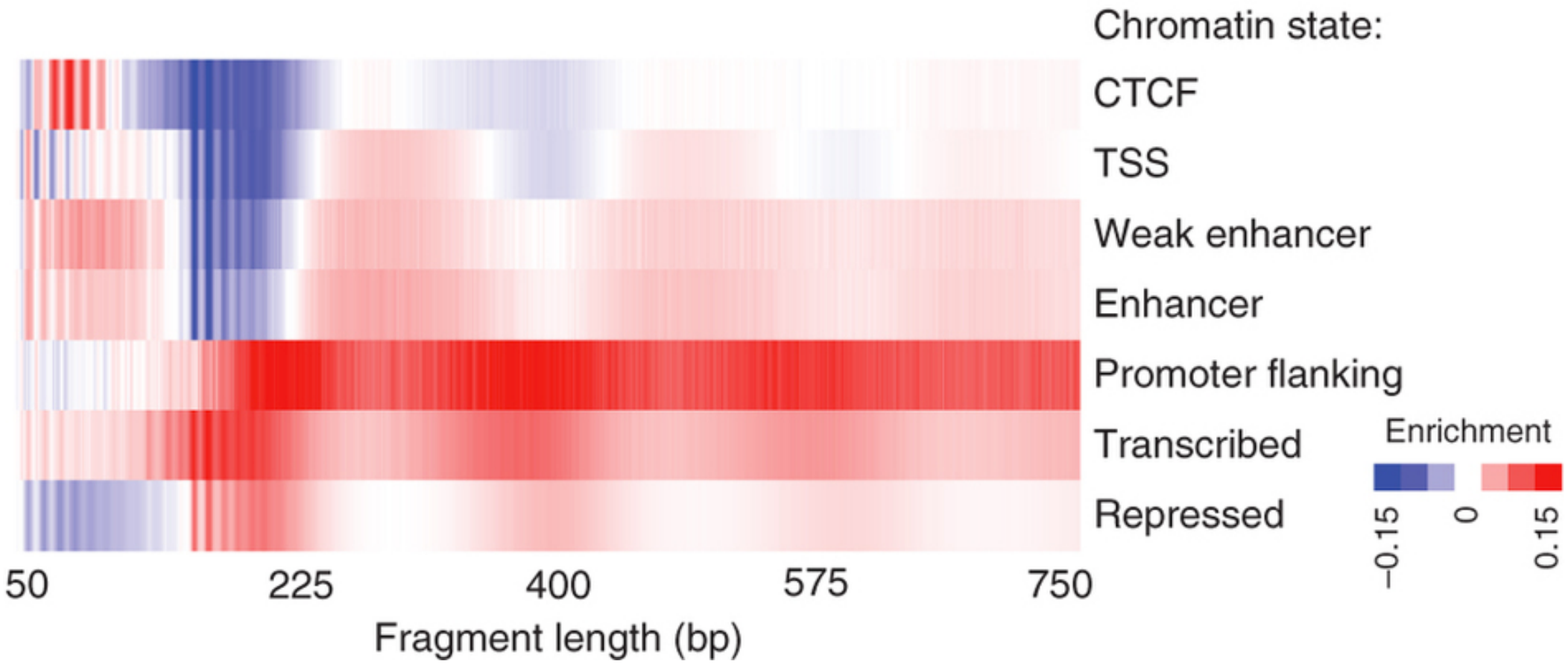
Modified from: www.the-scientist.com

ATAC-seq



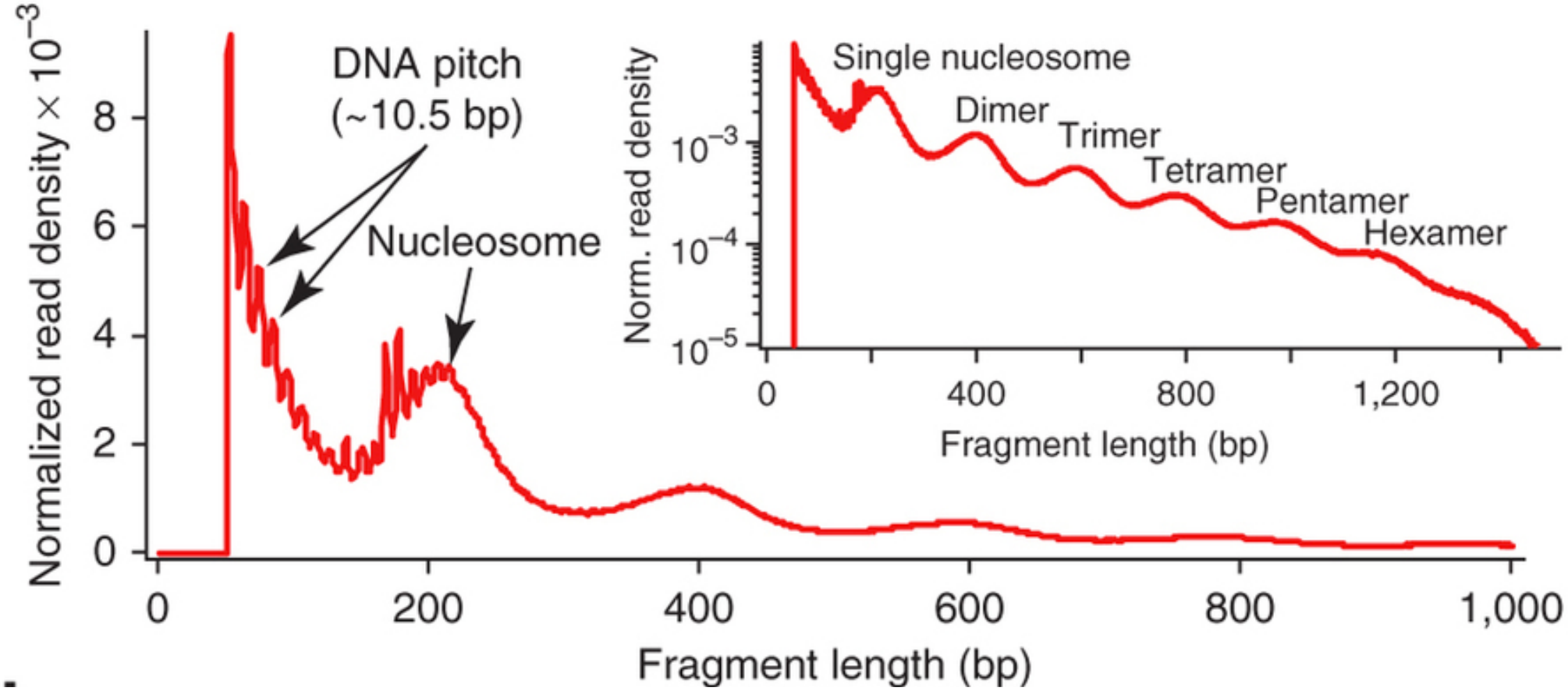
Each sequencing read points to a position on the genome where one transposition (or cutting) event took place during the experiment

Normalized read enrichments for seven classes of chromatin state

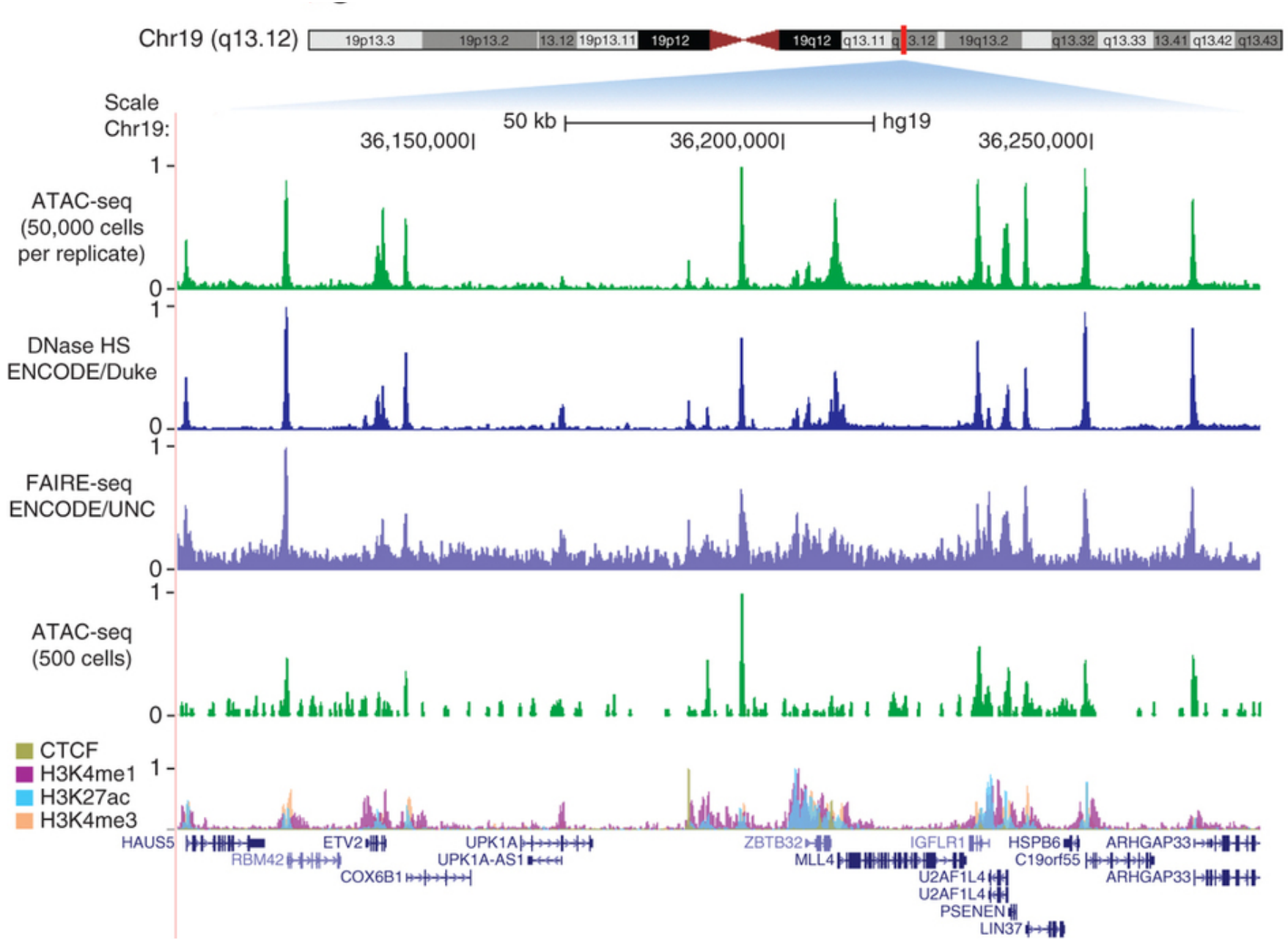


Buenrostro *et al*, Nature Methods (2013)

ATAC-seq



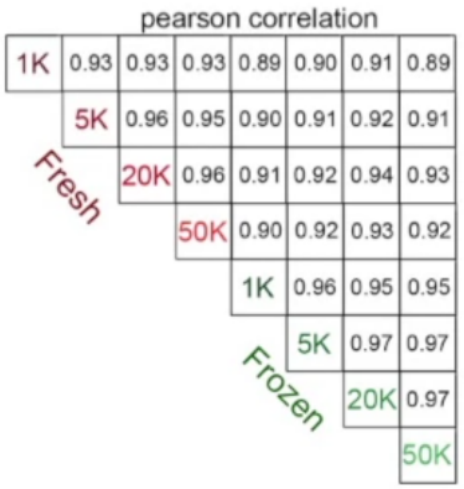
Buenrostro *et al*, Nature Methods (2013)

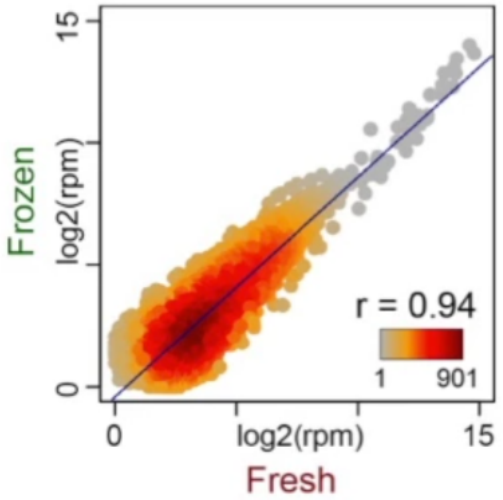


Buenrostro *et al*, Nature Methods (2013)

ATAC-seq works for frozen samples

Biobanked and Freshly Processed Samples Identify Similar Sets of Peaks



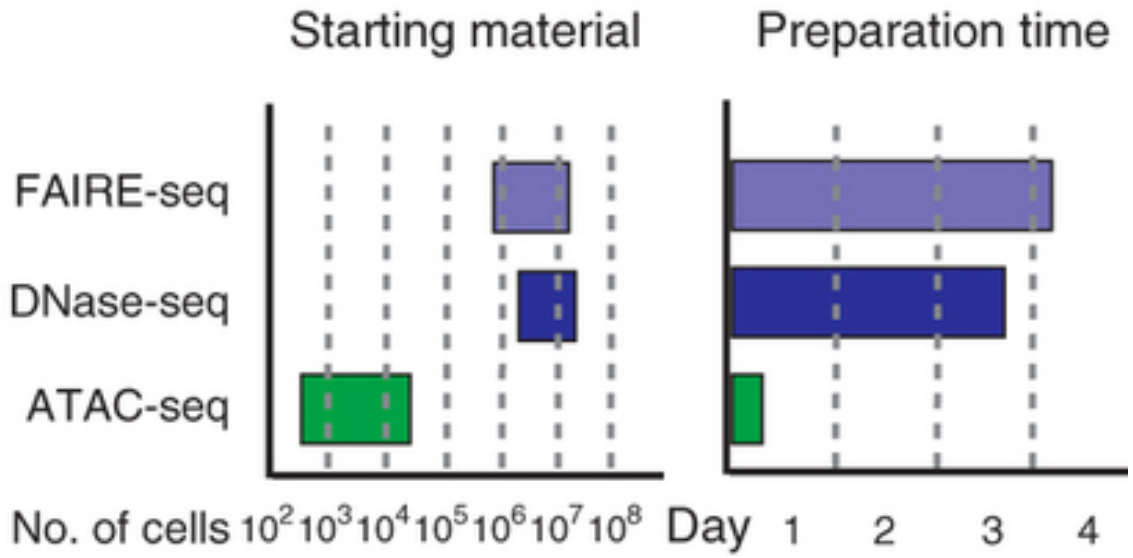


Dr Chris Scharer, Emory University

http://www.selectscience.net

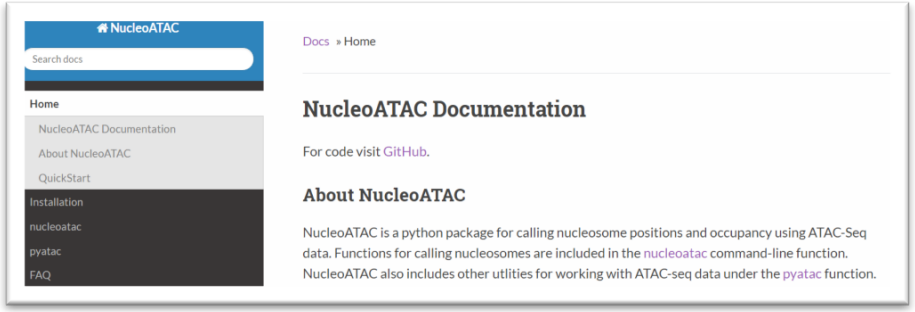
Advantages

- Requires fewer cells:
  50,000 cells are sufficient
- The protocol is the easiest of any of the accessibility methods
- the signal-to-noise ratio is fantastic
- Speed:
  less experimental calibration
  (the protocol requires 3 hours)



Disadvantages

- Bias characteristics are not well understood
- Starting materials are slightly more expensive
- There's not much precedent for ATAC-seq footprinting yet



MNase-seq

Micrococcal
Nuclease
sequencing

Cell

Volume 144, Issue 2, 21 January 2011, Pages 175–186

Primer

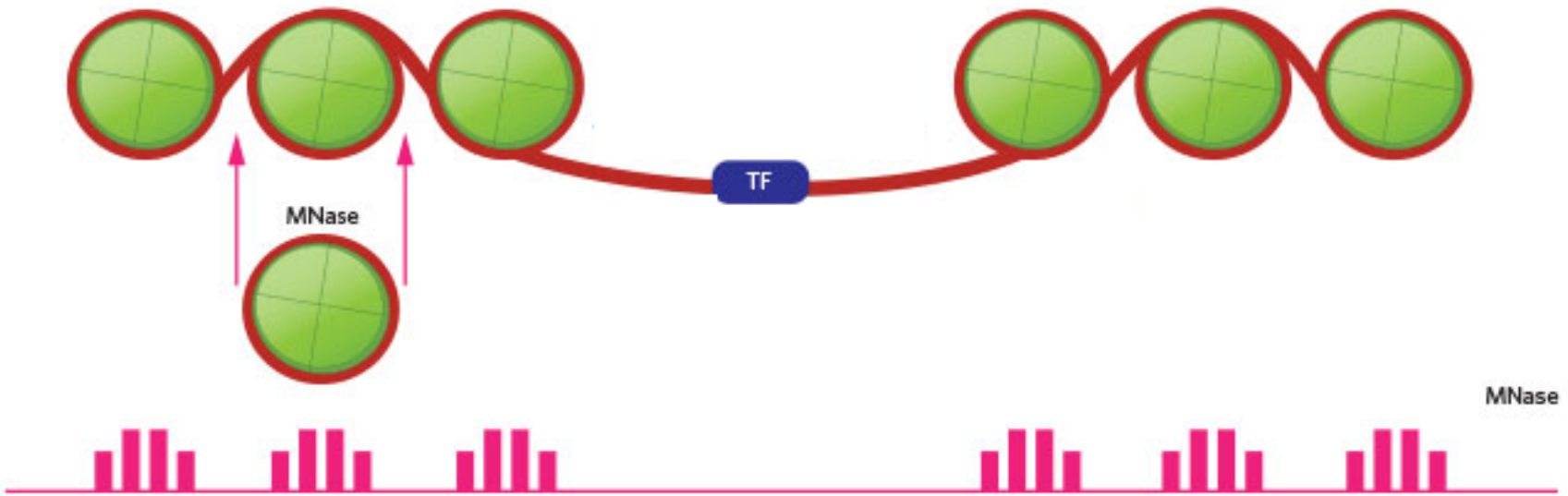
High-Resolution Genome-wide Mapping of the Primary Structure of Chromatin

Zhenhai Zhang[1], B. Franklin Pugh[1],

[1] Center for Comparative Genomics and Bioinformatics, Center for Eukaryotic Gene Regulation, Department of Biochemistry and Molecular Biology, The Pennsylvania State University, University Park, PA 16802, USA

Available online 20 January 2011

MNase-seq



How it works: MNase works by chewing up exposed stretches of the genome; the DNA associated with nucleosomes is recovered and sequenced. That makes MNase-seq the inverse of ATAC-seq and DNase-seq, at least conceptually.

Modified from: www.the-scientist.com

Advantages

- Can be used to study regulatory factors that bind to **nucleosomes** (in combination with ChIP-seq);
- MNase-seq has been used on the cells of **many species**, from yeast to humans.

Disadvantages

- MNase-seq requires 10–20 million cells;
- Most enzymes used in chromatin accessibility assays have sequence-specific biases; MNase likes to cut in AT-rich regions of the genome;
- For reasons that are not always clear, certain regions of the genome are more sensitive than others to MNase digestion;
- There’s no publications for single cell applications yet.

MNase-seq

Sequencing	Cell type/Number	Traditional approach	Genomic target	Experimental considerations
Single End and Paired End	Any cell type 1 to 10 million cells	MNase digests unprotected DNA	Maps the total nucleosome population in a qualitative and quantitative manner	1. Requires many cells.
				2. Laborious enzyme titrations.
				3. Probes total nucleosomal population, not active regulatory regions only.
				4. Degrades active regulatory regions, making their detection possible only *indirectly*.
				5. Requires 150 to 200 million reads for standard accessibility studies of the human genome.

Remarks

MNase digestion has been applied to study chromatin structure in a low-throughput manner since the early **1970s**;

MNase-seq probes chromatin accessibility ***indirectly,*** by unveiling the areas of the genome occupied by nucleosomes and other regulatory factors;

MNase digestion produces DNA fragments with ends that correspond to the ends of nucleosomes and, thus, produces maps with very high **resolution**;

DNase-seq and MNase-seq are not perfect opposites:
DNase-seq and ATAC-seq = exposed regions of DNA,
MNase-seq = regions protected by nucleosomes.
But because the methods provide **snapshots** of a dynamic process that is averaged across many thousands of cells, DNase- and ATAC-seq do not provide data that perfectly complement those of MNase-seq;

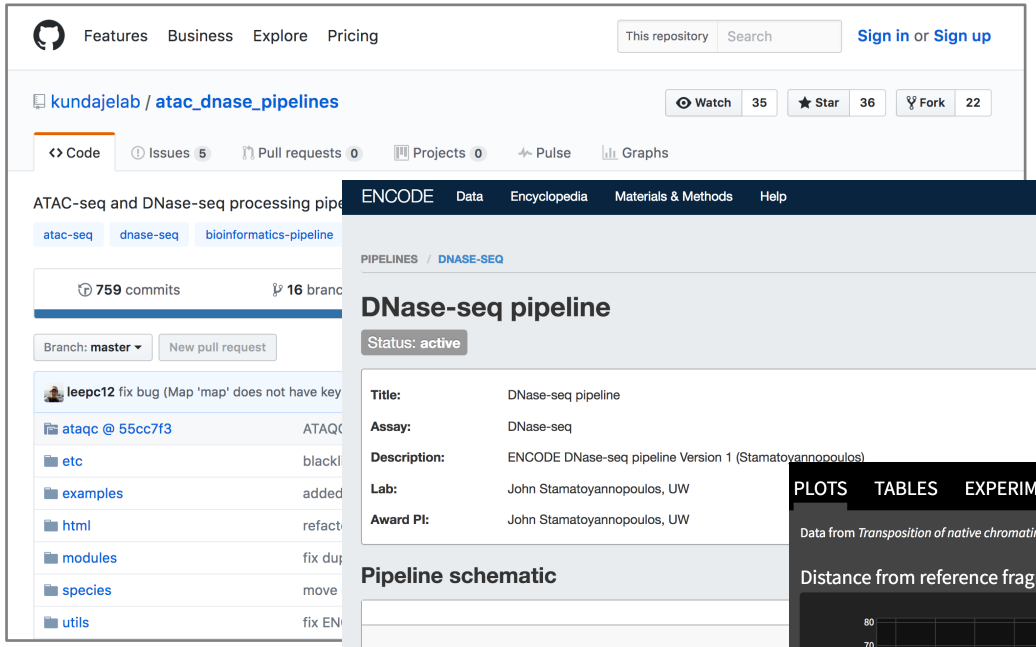
Although nucleosome size is 147 bp in higher eukaryotes, the real size of DNA fragments after MNase digestion can **vary** from ~120 bp to 170 bp;

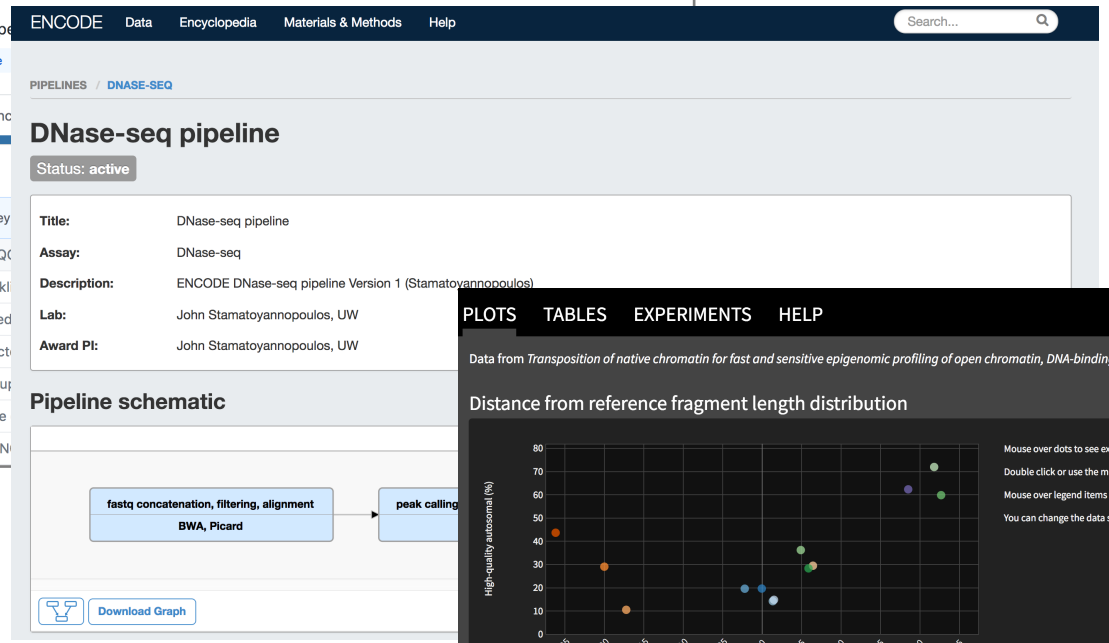
Chromatin accessibility high-throughput sequence data analysis

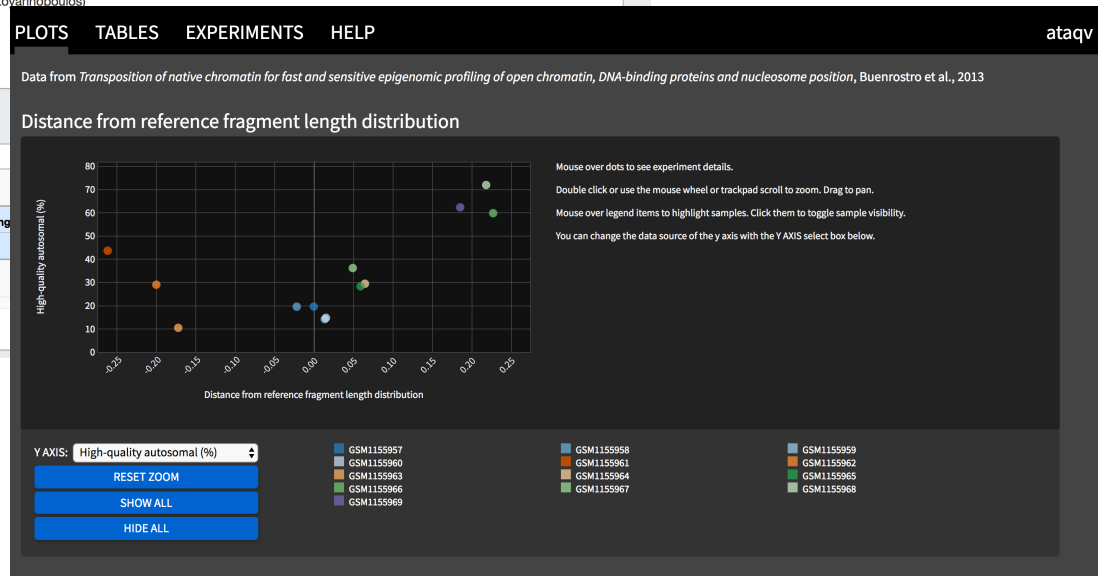
	Detection of enriched regions	Estimation of nucleosome organization and TF occupancy metrics
MNase-seq	1. GeneTrack	1. Nucleosome positioning algorithms
	2. Template filtering algorithm	2. Nucleosome occupancy algorithms
	3. DANPOS	3. V-plots for TF occupancy
	4. iNPS	
DNase-seq	1. F-Seq	1. Digital genomic footprinting algorithms
	2. Hotspot, DNase2Hotspots	2. Nucleosome and TF occupancy algorithms
	3. ZINBA	3. CENTIPEDE
	4. MACS	
FAIRE-seq	1. MACS2	Not available
	2. ZINBA	
ATAC-seq	1. ZINBA	1. Digital genomic footprinting algorithms
	2. MACS2	2. CENTIPEDE
	3. Hotspot, DNase2Hotspots	

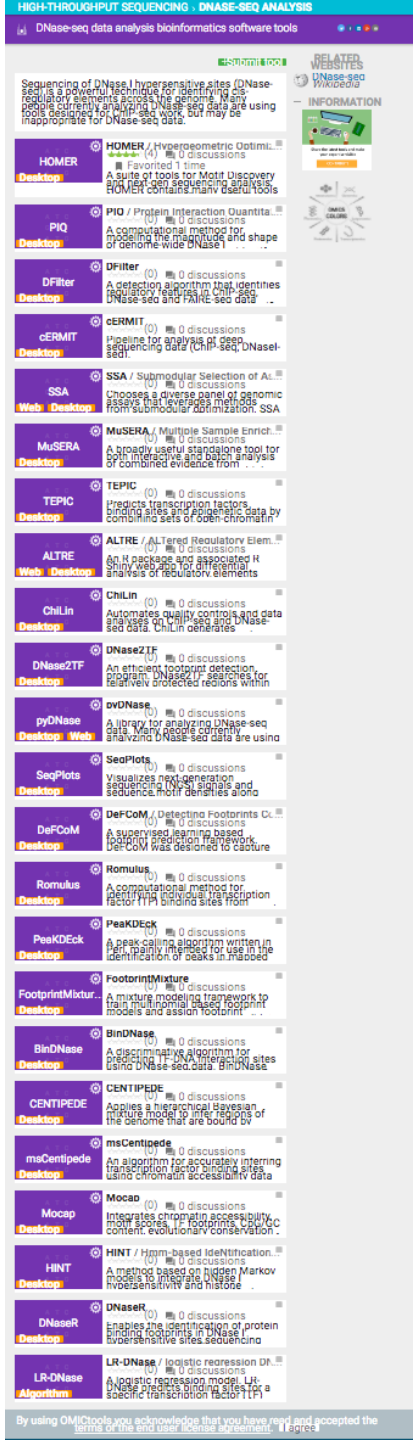
Maria Tsompana and Michael J Buck

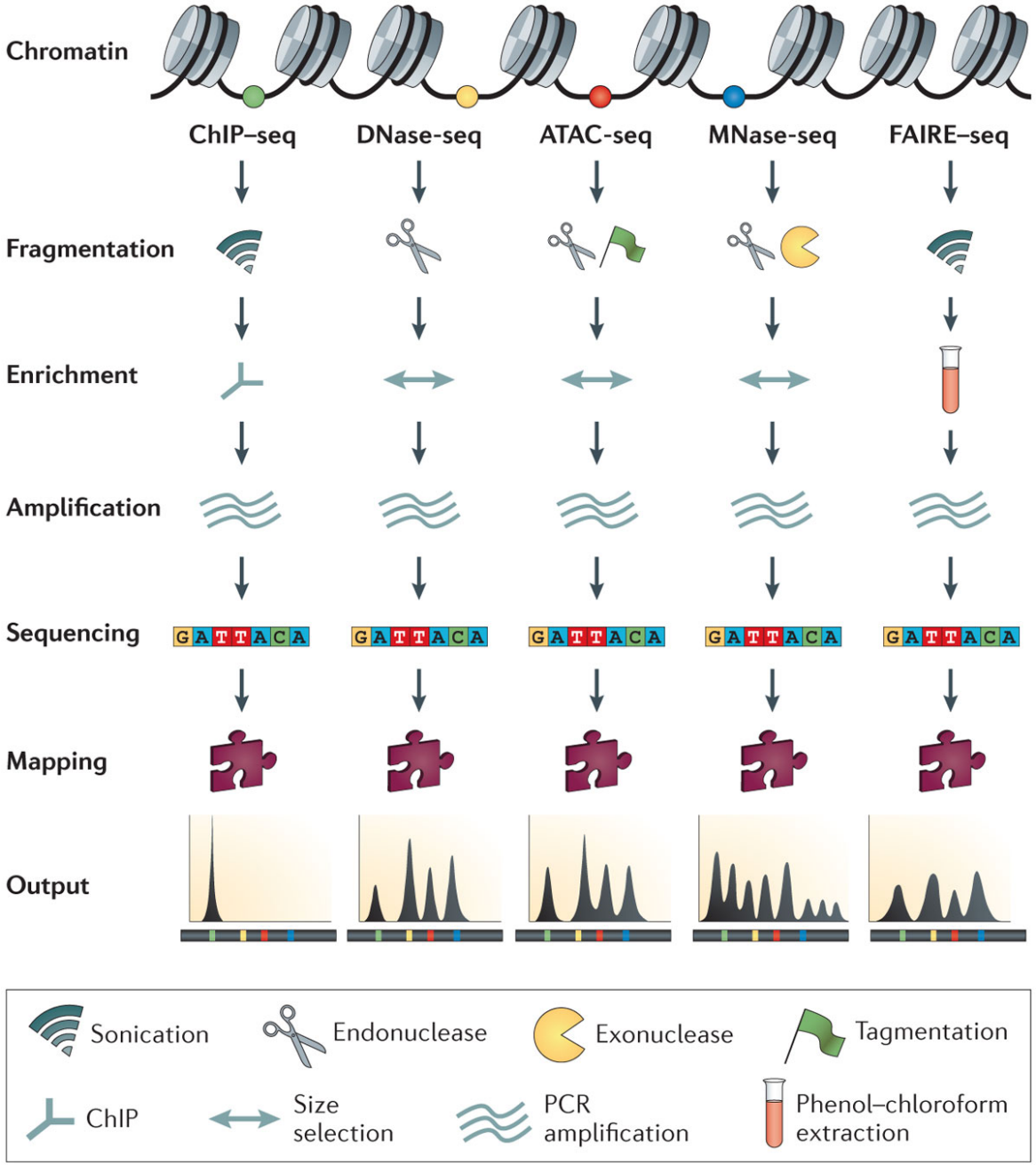
Pipelines and software:





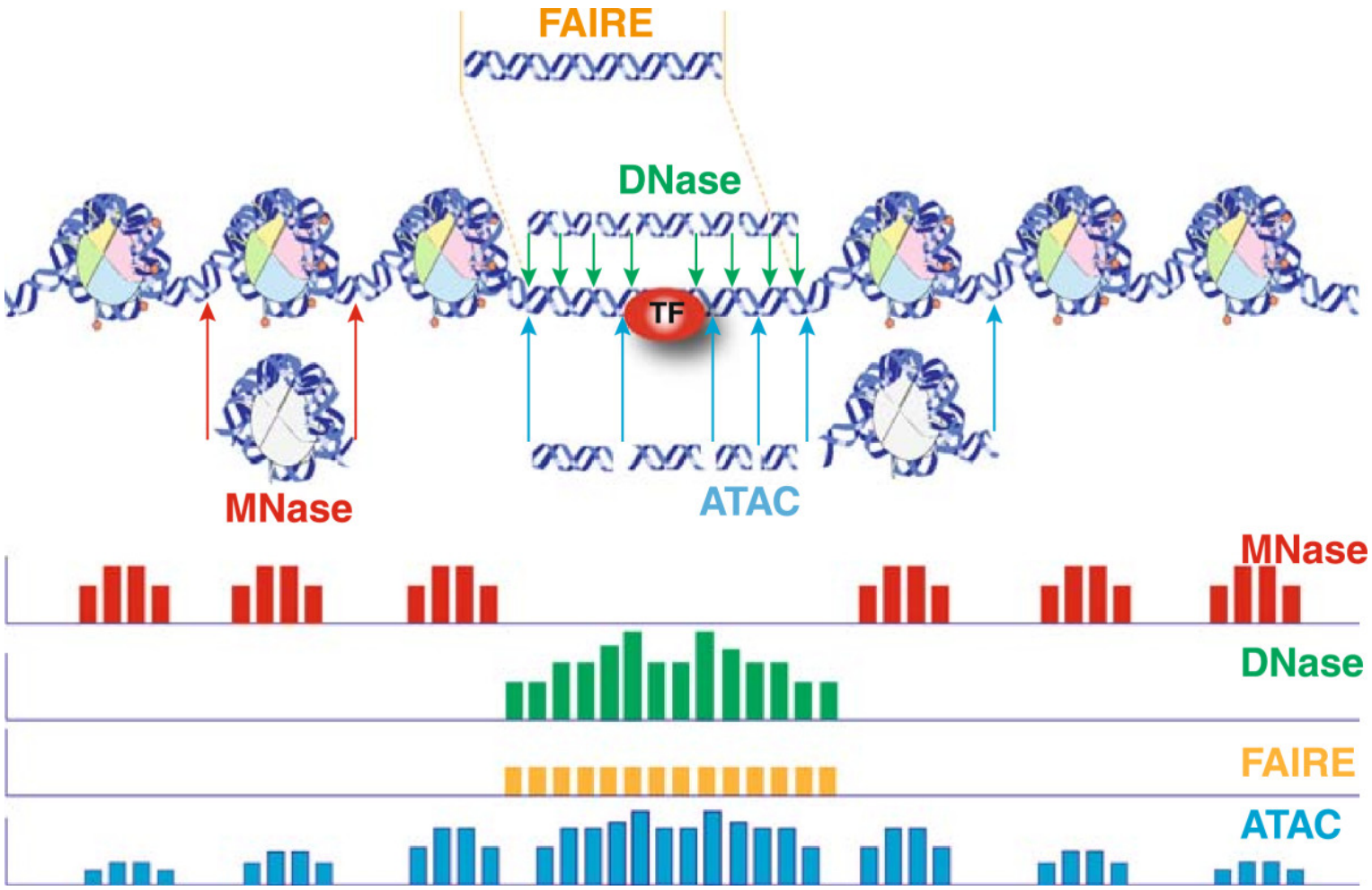






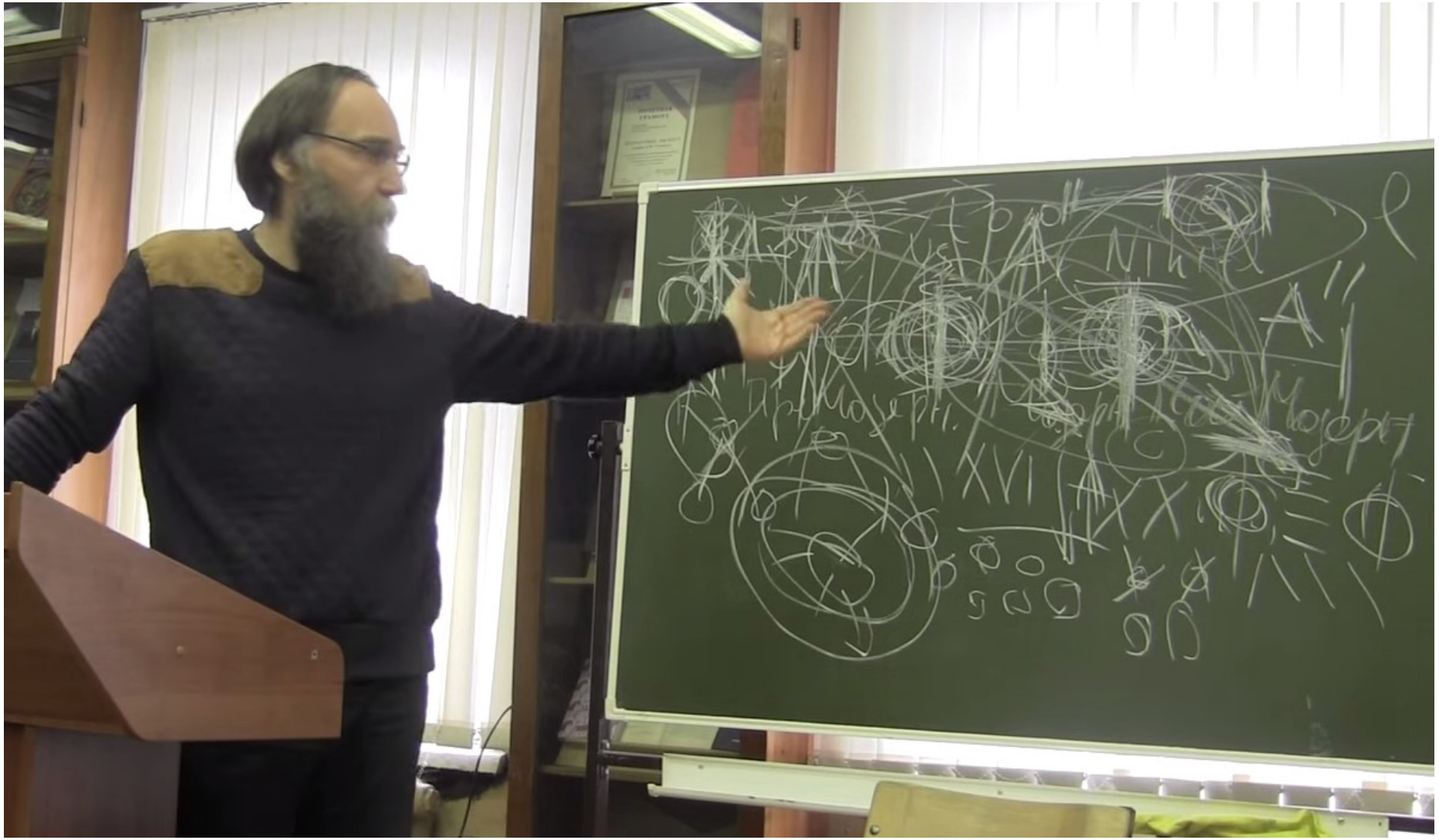
Meyer & Liu, Nature Reviews Genetics (2014)

“Take Home” message:



Maria Tsompana and Michael J Buck,
Epigenetics & Chromatin (2014)

It is very simple…



Dr. Peter FitzGerald,
Dr. Anand Merchant,
Dr. Bong-Hyun Kim,

and **CCBR Team**