Tech News

Histone Code-breakers: The Technologies of an Epigenetic Enigma

Nearly 10 years ago, C. David Allis and Brian Strahl, then at the University of Virginia Health Science Center, proposed a novel way of thinking about histones.

Histones, the authors surmised, were not merely the scaffolds about which nuclear DNA winds. Rather, they were active participants in gene expression, their many post-translational modifications serving as complex semaphores to control DNA and regulate nearby genes.

“We will refer to the hypothesis—that multiple histone modifications, acting in a combinatorial or sequential fashion on one or multiple histone tails, specify unique downstream functions—as the histone code hypothesis,” the authors wrote (1).

The suggestion was controversial. Scientists already knew that histones (the building blocks of nucleosomes and in turn the building blocks of chromatin) could be extensively modified, their N-terminal tails bristling with acetyl, methyl, and phosphate groups, ADP-ribosylation, and ubiquitin. What they didn’t know was *why*.

Conventional thinking at the time was that the modifications served to strengthen or loosen the nucleosomes’ grip on DNA, altering gene expression accordingly. But Allis and Strahl proposed instead that these epigenetic modifications represented “a histone ‘language’” that other proteins could read, write, erase, and modify. If deciphered, this histone code could predict events such as transcription, chromatin remodeling, and silencing.

Though all agree that histones are extensively modified and that those modifications have functional consequences, not everyone believes these modifications rise to the strict definition of a “code.”

“It’s like the Bible,” says Neil Kelleher, professor of chemistry at the University of Illinois at Urbana-Champaign. “There are strict constructionists—they are more literal, more bound by dogma. And then there are those who say, yes there’s a God, but it’s a bit more loose.”

Epigenetic theology aside, the idea of a histone code has legs: Allis and Strahl’s paper has been cited nearly 2500 times, according to the ISI Web of Science. The concept “is really quite exciting,” says Bryan

Turner of the University of Birmingham in the UK, a self-described histone-code agnostic. “To try and crack that code is just an enormous challenge, much more difficult than the genetic code—and we forget how hard that was to crack.”

Learning the Language

Kelleher is a mass spectrometry expert who focuses on the histone modification lexicon using top-down proteomics.

Most proteomics studies are “bottom-up,” first digesting protein samples to peptides

becomes incredibly difficult to quantify these things going up and down,” says Kelleher.

Top-down proteomics, on the other hand, is a protein cataloging strategy in which intact proteins, complete with modifications, are broken into smaller and smaller pieces until the whole has been fully characterized. Because it starts with intact protein, the approach ensures that all detected modifications coexist.

Kelleher is something of a top-down evangelist, but the approach isn’t for everyone: excellent chromatography and top-of-the-line hardware, such as with Fourier transform mass spectrometry (FT-MS), is necessary to achieve the resolution and mass accuracy required to distinguish large proteins that differ by just a few Daltons.

In fact, histones don’t even need the full top-down treatment because most modifications are confined to the proteins’ N-terminal tails. So Kelleher and then-postdoc Benjamin Garcia applied a stripped-down “middle-down” approach—or, in Kelleher’s words, “Goldilocks proteomics—these peptides are too small, those proteins are too big, but these fragments are just right.”

Kelleher and Garcia digested histone preparations to release their N-termini, fractionated those peptides by both their acetyl and methyl content, and applied their top-down technique to the resulting pieces. The survey identified more than 170 different forms of the histone H3 variant H3.2, an observation that Kelleher suggests probably caps histone diversity at around 10,000 combinations.

Garcia, now an assistant professor of molecular biology at Princeton University, recently refined the technique to make it both higher-throughput and more broadly applicable. He migrated from FT-MS to faster ion traps, introduced online nanoflow liquid chromatography, reduced sample requirements, and simplified back-end data analysis. “What used to take 100 hours of MS acquisition time in 2007, we can do now in a three-hour run,” he says. Garcia has identified some 300 unique histone forms with the technique, including the highly modified histone,

C. David Allis (above), along with Brian Strahl, proposed a new way of thinking about histone modifications. Courtesy of The Rockefeller University.

prior to fractionation and sequencing. The technique is simple and works well if only protein identification is desired. But if a researcher is concerned with complex post-translational modifications, and how those patterns change under different conditions—say, during oncogenesis or stem cell development—a problem emerges. Since the proteins are diced into bite-sized pieces, it’s impossible to tell if two modifications are physically linked, or if they are, instead, mutually exclusive.

“Once you cleave with a protease between two sites of modification, it

Histone tails are rife with potential modification sites; their combinatorial complexity—and the variety of biological outcomes they can produce—appears to some to represent a biological “code.” Courtesy of C. David Allis.

H3.2-K4me3K9acK14acK18acK23-acK27acK36me3.

Garcia’s team is now using the method to address questions such as how modification patterns change during stem cell differentiation or when a particular “writer” (histone-modifying enzyme) is knocked out. “This allows us to try to understand the impact of changes in the combinatorial code on biological processes,” says Garcia. “That is the question that is driving us.”

The Trouble with Chip

For all its power, MS offers little insight into the whereabouts of specific histone modifications. But biologists want to correlate modification patterns with gene expression—that is, where in the genome these chemical marks are located, and under what biological conditions. The epigeneticists’ tool to answer those questions is chromatin immunoprecipitation (ChIP).

In ChIP, chromatin is crosslinked, fragmented, and immunoprecipitated with an antibody to a particular DNA binding protein or epigenetic modification. The crosslinking is then reversed to release the DNA, at which point the enriched DNA populations can be queried. Though traditionally accomplished using PCR, researchers have more recently embraced microarrays and next-generation sequencing (dubbed ChIP-chip

and ChIP-seq, respectively) to profile on a genomic scale.

ChIP-seq, in particular, now enables genome-wide characterization of histone modifications at nucleosome resolution, says Peter Park, a bioinformatician and assistant professor of pediatrics at Children’s Hospital Boston. ChIP-seq also has fewer artifacts than ChIP-chip and can be cost-effective, especially for large genomes. But it isn’t easy.

“Doing the experiment itself is not that difficult, but the computational aspect is very difficult,” notes Park, who authored a recent review on ChIP-seq (2). A ChIP-seq run typically generates tens of millions of reads, each which must be aligned to the genome. The data must be crunched to identify regions of enriched read counts (relative to a control sample), which are then further scrutinized to draw biological insights.

There are other computational issues, too. “Most datasets out there are under-sampled, as sequencing still is a bit expensive for most investigators,” Park says. They’re also often generated using suboptimal reagents.

ChIP requires high-quality antibodies, Park says, because histone modifications are subtle. “You need an antibody that can recognize, say, dimethylation of a particular histone residue, but not trimethylation,” he explains. “It’s not easy to produce an antibody with such specificity.” Plus, adds Bryan Turner, nearby

Free Subscription

BioTechniques is the first peer-reviewed journal with open access to the whole of the life science community.

100% life science
100% peer-reviewed methods
100% “everyday” practical information

Sign up for free at
www.BioTechniques.com/subscribe

ChIP-seq data from fly (*Drosophila melanogaster*, chromosome 2L), showing a CTCF insulator position separating a cluster of active genes on the right (enriched for H4K16ac modification and Chromator protein), from a silent region on the left, marked by the H3K9me2 histone modification. Courtesy of Peter Kharchenko.

modifications may unexpectedly impact antibody binding.

Rigorous antibody validation is therefore essential, and one of the goals of the NIH-directed ENCODE and modENCODE projects. Park, who works on *Drosophila* modENCODE, says the team has performed many validation experiments for antibodies in addition to some 700 ChIP-chip experiments. Forty-four antibodies to histone modifications and 53 to chromosomal proteins have been tested so far, he says, and only 26 and 37 worked, respectively. “I think most people would say one-third of antibodies don’t work as advertised,” he says. “Some would say as many as half.”

Turner is addressing another technical limitation of ChIP. Because it generally requires at least a million cells per assay, ChIP has been effectively restricted to cultured cells. But Turner wanted to apply the technique to embryos. “If a histone code exists,” he explains, “then development and differentiation is where you will see it working.”

His strategy, called “Carrier ChIP,” requires just 100 cells, and works by spiking the sample with *Drosophila* genomic DNA, which acts as a carrier in the purification process. Enriched fragments are detected using species-specific PCR primers, a technique he recently used to probe histone modification changes at selected loci throughout mouse embryogenesis (3).

Turner is working to make the technique compatible with ChIP-chip and

ChIP-seq. “If we can refine the technique,” he says, “that will be a big step forward.”

Defining the Readers

Enumerating histone modifications is only part of the histone code challenge; the proteins that recognize and act on those modifications (the “readers”) must also be identified.

Reader proteins such as HP1 [which recognizes methylated lysine-9 on histone H3 (H3K9me)] and WDR5 [dimethyl-H3K4 (H3K4me2)], target specific modifications via well-characterized binding “modules,” including PHD domains, bromodomains, and chromodomains. Sequence homology suggests the genome encodes hundreds of such proteins. The trick is to match them with their cognate modification signals.

Researchers typically use *in vitro* approaches employing modified histone peptides to capture potential modification-specific binding proteins. But these techniques are performed under non-physiological conditions and are, according to Judd Rice of the University of Southern California Keck School of Medicine, “a gigantic pain.”

To overcome these limitations, Rice (a former Allis postdoc) expanded on an approach developed by Min-Hao Kuo at Michigan State University to identify yeast histone modification-specific binding proteins *in vivo*. The new approach, called mammalian tethered catalysis (MTcC), utilizes a fusion protein containing

A schematic diagram of mammalian tethered catalysis (MTcC). Courtesy of Tanya Spektor and Judd Rice.

affinity tags in tandem with a histone peptide sequence and the catalytic domain of a particular modifying enzyme. When expressed in cells, the enzyme modifies the histone peptide, serving as the bait to capture, isolate, and identify putative binding proteins in an *in vivo* context.

Rice says the approach is yielding results. “It’s too early to say for sure, but our preliminary data suggest this method is working quite well,” he says, noting that some “very exciting” and “unexpected” proteins have been identified.

Histone code co-pioneer Strahl (now at the University of North Carolina School of Medicine) has developed an alternative approach to matching readers with modifications: custom peptide microarrays. Strahl says that the approach, which probes arrays of different modification variants with purified proteins, will ultimately be guided by catalogers like Kelleher and

Garcia. Initially, though, he will focus on one modification at a time.

"We are not trying to 'crack the code,'" he says, "but we are trying to use this technique to help define how proteins respond to and interact with one or more modifications."

For his part, Allis (now at The Rockefeller University) also continues working on the histone code, but on a different scale. Along with chemist Tom Muir at Rockefeller, Allis is using peptide ligation to build "designer" nucleosomes with defined modifications that can probe crosstalk within or across nucleosomes. In one case, Allis' team probed the interaction of nucleosomes harboring H3K4me3 and H4K16ac with an engineered dual-module reader containing domains that individually could read either modification alone. The results suggest "there may be a combinatorial aspect to this," he says.

"I think that [the technique] is going to be unbelievably powerful," Allis says. "And you can add them to any DNA. That starts to build up di- or trinucleosomes, and then you can start to get into things that are really fascinating, like higher-order chromatin."

Everyone agrees that such tools will likely have applications beyond chromatin. Given the extensive modification of proteins, such as transcription factors and signaling molecules, it is possible there even exists what Garcia calls "a general protein code," with the histone code being a special case.

But whatever the general implications, it is clear that chromatin biologists have had a remarkably productive decade. And for Strahl, that alone makes the histone code hypothesis worthwhile. "Regardless of the outcome, it's had a major impact on people's thinking," he says. "It's driven a lot of exciting science, and that to me is the most rewarding thing."

References

1. **Strahl, B.D. and C.D. Allis.** 2000. The language of covalent histone modifications. *Nature* 403:41-45.
2. **Park, P.J.** 2009. ChIP-seq: Advantages and challenges of a maturing technology. *Nat. Rev. Genet.* 10:669-680.
3. **VerMilyea, M.D., L.P. O'Neill, and B.M. Turner.** 2009. Transcription-independent heritability of induced histone modifications in the mouse preimplantation embryo. *PLoS One* 4:e6086.

Written by Jeffrey M. Perkel, Ph.D.

BioTechniques 48:185-191 (March 2010)
doi 10.2144/000113380

- Assay Development
- Nucleic Acid Purification
- Genotyping
- Microarray Technology
- qRT-PCR
- Toxicogenomics
- Bioinformatics
- Tumor Profiling
- Sample Management
- Biostorage

THE Pharmacogenomics Company

Revolutionizing Medicine Through Pharmacogenomics[®]

For more information, contact us at 919.653.5532 or visit our website at www.gentris.com