milestone

Pulling Down Answers to Biological Mysteries with Chromatin Immunoprecipitation

BY MAGGIE CHEN

The genomic revolution allowed scientists to uncover specific genes that associate with various functions. As more insights into the genome emerged, curiosities arose as to what regulated these genes. Scientists understood many relationships between proteins, DNA, and RNA, but untangling the implications of those relationships, particularly in the context of gene expression, led to the development of chromatin immunoprecipitation (ChIP).

Through many innovations over the past 40 years, ChIP rapidly evolved to become a powerful tool for understanding the epigenome — the modifications that alter gene activity without changing the DNA sequence. ChIP allowed scientists to uncover chromatin states in cell differentiation, discover important transcription factors, and map the epigenome in health and disease.

1980s

ChIP emerges

As scientists learned different ways to manipulate and study the genome, biochemist John Lis at Cornell University became inspired by the DNA cloning and gene identification experiments recently completed by David Hogness and others at Stanford University. With the desire to motivate a new graduate student, David Gilmour (now at Pennsylvania State University), Lis discussed the potential for isolating and studying individual genes. Gilmour countered with a proposal for studying the proteins involved in regulating genes. Lis went home that day pondering Gilmour's ideas and wondered what they could learn about proteins.

"I wanted to push the scope because I knew there was a bottleneck," Lis said. "Now you get genes, and what do you do with them? You can characterize them. But how do you go about understanding how gene regulation works?"

Back in the laboratory, Gilmour and Lis decided to focus their efforts on studying protein-DNA interactions within living cells. Their original idea was to use crosslinking methods to "trap" proteins with associated segments of DNA and then "fish out DNA sequences," which would hypothetically bring down the corresponding proteins.

After more than three years of failed experiments, the turning point came during a seminar where a scientist described precipitating proteins and then probing for DNA — essentially the reverse of Gilmour and Lis's idea. "I remember looking across the auditorium at [Gilmour], and he was looking at me, and I said, 'this is it,'" Lis recalled.

The seminar speaker described crosslinking *E. coli* RNA polymerase protein to DNA. "I decided that might be a nice model system to work with," Gilmour said.

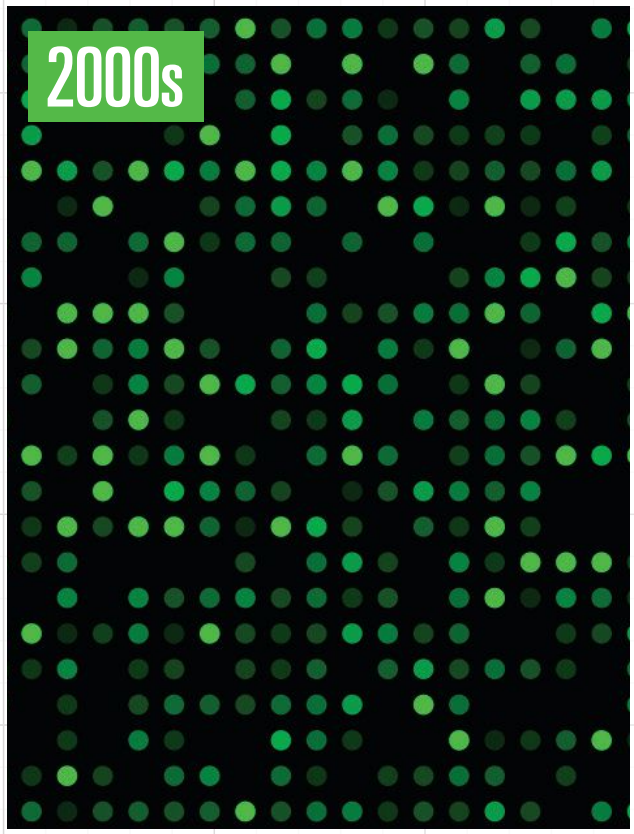
Back in the lab, the pair crosslinked protein to DNA with UV irradiation in *E. coli* and then immunoprecipitated RNA polymerase along with any DNA attached to it using an antibody donated by a colleague. To study the DNA sequences attached to the polymerase, the pair used hybridization assays to bind specific DNA fragments. Gilmour and Lis reported these first ChIP experiments in *Proceedings of the National Academy of Sciences* in 1984 (1).

"It took an awful long time to get to this point," Gilmour said. "And ChIP was almost fortuitous."

CREDIT: JOHN LIS AND DAVID GILMOUR

John Lis (bottom) and David Gilmour (top) conducted the first ChIP experiments using UV crosslinking.

2000s



DNA microarrays helped scale up the ChIP technique.

Speeding and scaling up

Following the publication of Gilmour and Lis' UV crosslinking method, biochemist Alexander Varshavsky and others at the California Institute of Technology developed a formaldehyde-based crosslinking method in 1988 (2). Using low concentrations of formaldehyde enabled reversible crosslinking. This formaldehyde-based technique caught on and became the widely used crosslinking method, giving rise to a suite of newer ChIP methods.

Even after the introduction of formaldehyde-based crosslinking, the ChIP method only permitted scientists to assess one (or a few) proteins for DNA binding activity per sample. Additionally, scientists found it difficult to read the DNA sequences attached to the proteins. But in the early 2000s, scientists introduced DNA microarrays, which enabled them to probe a wide array of sequences simultaneously. Combining traditional ChIP with microarrays led to ChIP-chip, which allowed for genome-wide analysis of specific transcriptional regulator binding activities in yeast (3,4).

The original ChIP methods also required approximately three days to complete one experiment. "I had a graduate student who was very tall spending all of his days under the fume hood," Karol Bomsztyk, physician-scientist at the University of Washington, said. "I said to him, 'this is no way to make a living or to spend a weekend.'"

Inspired by marine biologists using the compound Chelex to purify DNA from fossils, Bomsztyk and colleagues (including the tall graduate student, Joel Nelson) applied Chelex to the ChIP beads used to pull down chromatin, which vastly decreased the time needed for DNA isolation. Improving the ChIP method led to Fast-ChIP, which Nelson, Bomsztyk, and others published in 2006 (5). Instead of needing three days and six samples, Fast-ChIP took one day and assessed 24 samples.

"[Chelex] was made by BioRad, I believe," Bomsztyk said. "After we published, BioRad ran out of Chelex!"

2007–2008

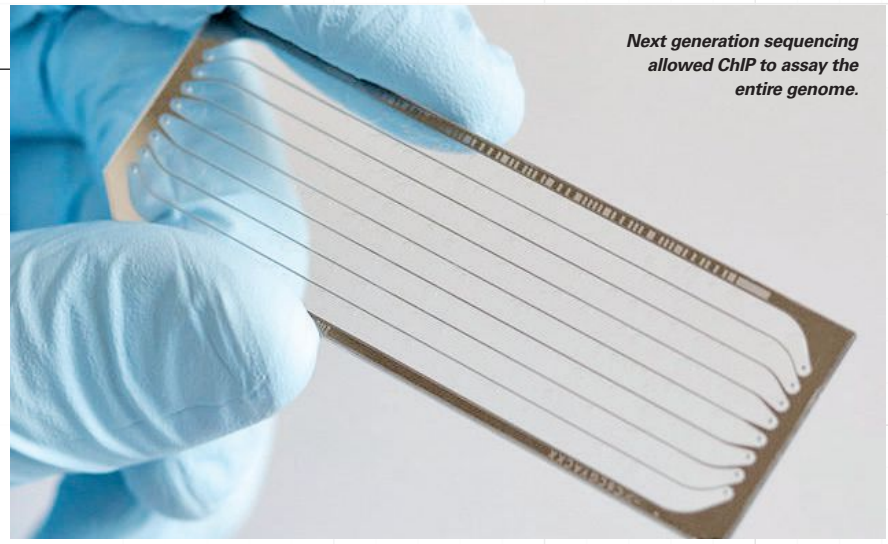
ChIP meets sequencing

While DNA microarrays allowed for some degree of scalability when reading out DNA sequences, they were still limited by cost and bias. In 2005 and 2006, Solexa, Illumina, and 454 Life Sciences released next generation sequencers, enabling scientists to rapidly read genomes in a high-throughput and less biased manner.

"ChIP was a highly localized technique," computational biologist Tarjei Mikkelsen, now at Arsenal-Bio explained. "You had to sort of really know where you were looking before you did the experiment, which was obviously limiting because a lot of the questions we wanted to ask were like 'where's the interesting stuff that we didn't know about?'"

To overcome this limitation, Mikkelsen and other scientists at the Broad Institute of Harvard and Massachusetts Institute of Technology combined next generation sequencing with ChIP to create ChIP-seq in 2007 (6,7). By sequencing the DNA resulting from ChIP experiments, scientists could fully map out their sequences. This allowed for discovery experiments, where scientists could uncover on a whole genome scale which DNA regions were most interesting in terms of protein-DNA binding and where they were located. Mikkelsen and his colleagues applied ChIP-seq to determine methylation and histone acetylation marks in pluripotent versus lineage-committed mammalian cells (7); biologist Barbara Wold's group at the California Institute of Technology looked at interactions between neuronal factors and cell differentiation (6).

For Mikkelsen, ChIP-seq was an exciting innovation that allowed for a deeper look into the genome and epigenome. "We were surprised once we started seeing the richness of the data and how much information [ChIP-seq] provided — just being able to annotate the genome in various ways, finding promoters we didn't know about, finding all this evidence of non-coding regulation on coding RNA — that just hadn't come out in the past," he said.



Next generation sequencing allowed ChIP to assay the entire genome.

2010s

Using exonucleases (shown in this structure) to digest DNA brought higher resolution to ChIP-seq.

Bigger and better

With large-scale sequencing methods in place, scientists worked to further optimize ChIP in a myriad of ways. Biologist B. Franklin Pugh, now at Cornell University, described the introduction of lambda exonuclease as paramount for the specificity and resolution of ChIP-exo, the new method that he introduced in 2012.

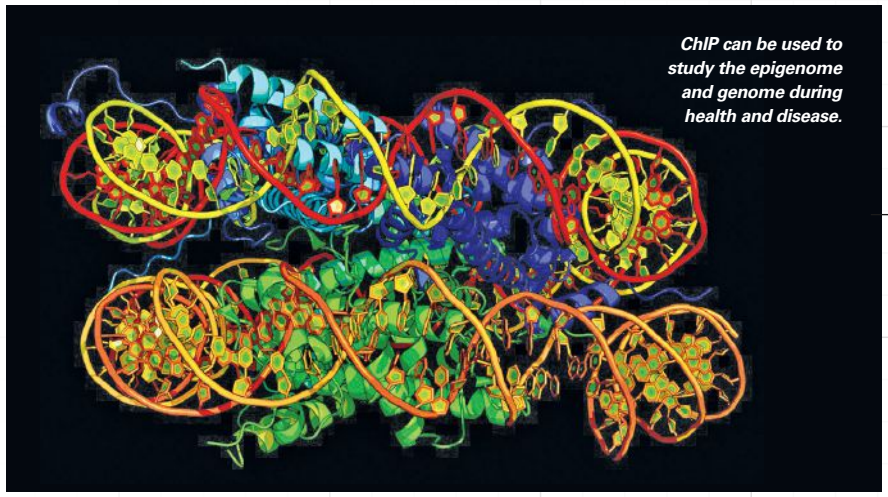
Because lambda exonuclease chews DNA in a specific direction, Pugh's team hypothesized that proteins bound to DNA would block the exonuclease from chewing on that specific region. This led to ChIP-exo, where lambda exonuclease digests DNA resulting from ChIP to reveal almost single-nucleotide resolution of binding sites after high-throughput sequencing (8).

"It'll stop precisely from where the protein is bound, plus or minus. It's really concentrated to a single base pair; there's a little bit of wiggle room," Pugh explained.

Conducting high-throughput DNA sequencing and downstream analysis of the exonuclease-digested ChIP DNA precisely mapped the protein binding locations in the genome at a high resolution. ChIP-exo joined even more ChIP-seq adjacent technologies like HiChIP (9), which explored chromatin conformation, and Matrix (microplate-based)-ChIP/PIXUL-ChIP, which scaled the number of proteins that could be assessed simultaneously and further standardized sample preparation (10,11).

For Pugh, these ChIP technologies illuminate the mammalian epigenome and other organismal epigenomes. "We've been mapping all the proteins that interact with the yeast genome for starters as a sort of proof of principle, but it's been really informative," he said. "And, now we're working in human systems."

2012–Present



ChIP can be used to study the epigenome and genome during health and disease.

Unraveling the epigenome and genome even further

As ChIP evolved, the variety of tools available for studying the epigenome also grew in scope. Assay for Transposase Accessible Chromatin (ATAC)-seq joined ChIP in 2013 as a method for uncovering open regions of chromatin.

"ATAC-seq is a very specific signal," Mikkelsen said. "It tells you where accessible regions in the genome are. It doesn't directly tell you why those regions are accessible."

Mikkelsen explained that combining ChIP-seq with techniques like ATAC-seq could reveal which histone modifications, transcription factors, or other protein binding components are responsible for different chromatin states.

The emergence of these techniques also allowed for conception of the Encyclopedia of DNA Elements (ENCODE), a public research project founded in 2003 that aimed to map out all functional elements in the human and mouse genomes (12). "Using this technique to annotate the genome ended up being one of the major focuses of the ENCODE projects," Mikkelsen said.

Application of ChIP and other sequencing technologies for exploring the epigenome can provide insights into how the epigenome and genome change during disease for the sake of therapeutic development. One of Bomsztyk's current projects uses ChIP techniques to explore the methylation profiles of a biomarker promoter to stratify patients for glioblastoma treatments.

"Epigenetic therapy is one of the fastest growing areas of therapeutics," Bomsztyk said. "The ultimate validation of technologies like ChIP will be with the patients. If you help a patient, then you can know that you did something nobody will question."

REFERENCES

1. Gilmour, D. S. & Lis, J. T. Detecting protein-DNA interactions in vivo: distribution of RNA polymerase on specific bacterial genes. *Proc Natl Acad Sci U S A* 81, 4275–4279 (1984).

2. Solomon, M. J., Larsen, P. L., & Varshavsky, A. Mapping proteinDNA interactions in vivo with formaldehyde: Evidence that histone H4 is retained on a highly transcribed gene. *Cell* 53, 937–947 (1988).

3. Lee, T. I. *et al.* Transcriptional Regulatory Networks in Saccharomyces cerevisiae. *Science* 298, 799–804 (2002).

4. Ren, B. *et al.* Genome-Wide Location and Function of DNA Binding Proteins. *Science* 290, 2306–2309 (2000).

5. Nelson, J. D., Denisenko, O., & Bomsztyk, K. Protocol for the fast chromatin immunoprecipitation (ChIP) method. *Nat Protoc* 1, 179–185 (2006).

6. Johnson, D. S., Mortazavi, A., Myers, R. M., & Wold, B. Genome-wide mapping of in vivo protein-DNA interactions. *Science* 316, 1497–1502 (2007).

7. Mikkelsen, T. S. *et al.* Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560 (2007).

8. Rhee, H. S. & Pugh, B. F. ChIP-exo: A Method to Identify Genomic Location of DNA-binding proteins at Near Single Nucleotide Accuracy. *Curr Protoc Mol Biol* 0 21, 10.1002/0471142727.mb2124s100 (2012).

9. Mumbach, M. R. *et al.* HiChIP: efficient and sensitive analysis of protein-directed genome architecture. *Nat Methods* 13, 919–922 (2016).

10. Flanagin, S., Nelson, J. D., Castner, D. G., Denisenko, O., & Bomsztyk, K. Microplate-based chromatin immunoprecipitation method, Matrix ChIP: a platform to study signaling of complex genomic events. *Nucleic Acids Res* 36, e17 (2008).

11. Bomsztyk, K. *et al.* PIXUL-ChIP: integrated high-throughput sample preparation and analytical platform for epigenetic studies. *Nucleic Acids Research* 47, e69 (2019).

12. Dunham, I. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74 (2012).