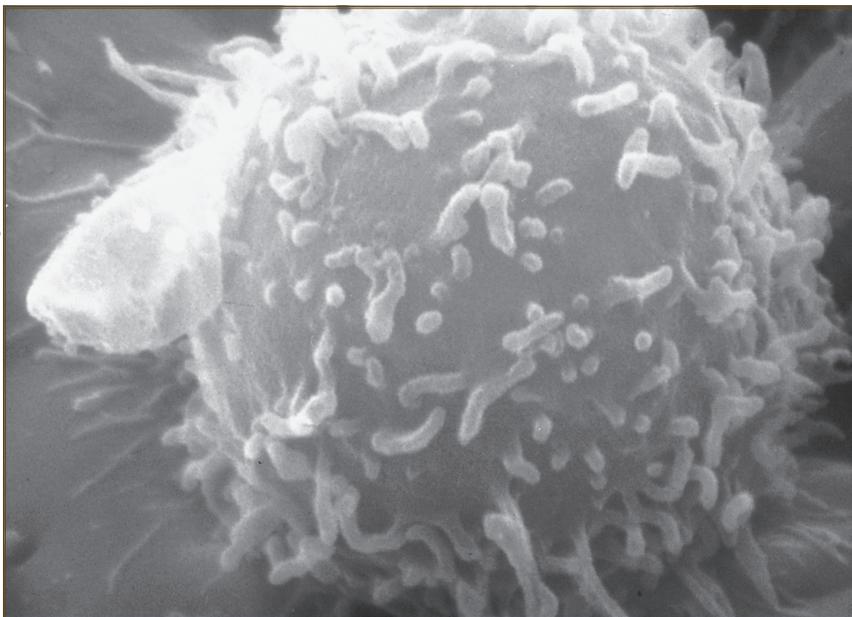


Global Amplification: A New Look at Transcriptional Regulation

Biologists often identify and describe cells based on the molecules they express. The expression of certain genes can be used to separate a stem cell from a neuron, a cancerous cell from a healthy one. However, as cells grow and adapt, they change not just the types of genes they express but the amount of that expression. David Levens, M.D., Ph.D., Senior Investigator in CCR's Laboratory of Pathology, and Rafael Casellas, Ph.D., Senior Investigator with the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) and Adjunct Investigator in CCR's Laboratory of Cancer Biology and Genetics, came together to study genome-wide transcriptional regulation in B cells from two very different vantage points. Together, they are discovering that global transcriptional amplification is a uniquely regulated process.

(Image: National Cancer Institute)



Electron microscopic image of human lymphocyte, which consists of T cells and B cells

Over the last 20 years, David Levens and his colleagues have learned a lot about the regulation of a single gene: *myc*. But, *myc*'s well-known importance as a critical regulator of cellular growth, both during development and in cancer, was not Levens' chief concern. "I stayed out of studying *myc* function for a long time," said Levens. "I was more interested in the mechanisms

of gene regulation. *Myc* seems to have one of the most complicated promoters around. And no one had a comprehensive model of how a cell decides how much *myc* to make and when."

As a result of their investigations, Levens' laboratory has put together a fascinating and unusual story of transcriptional regulation. Among the many regulators of *myc*

expression, they found proteins that were not binding to classical double-stranded segments of DNA (dsDNA). Instead, they traced the action of these factors to a single-stranded DNA (ssDNA) element in the promoter region, the Far Upstream Sequence Element (FUSE).

Molecular forces predispose DNA to adopt the famous Watson-Crick double helix when its bases are appropriately paired: adenine (A) with thymine (T), guanine (G) with cytosine (C). But that seemingly static picture changes with transcription. As DNA is screwed through the active site of enzymes that travel along it—DNA polymerases, RNA polymerases, helicases—rotational forces are transmitted through the DNA. "We saw that at particular sites on the DNA, the dsDNA would essentially buckle, popping open like the threads of a rope unfraying at particular sites." This turned out to be important for *myc* regulation.

Levens hypothesized that FUSE regulation was occurring once other transcriptional initiation events had been precipitated. "The cell can sense how much *myc* is

being made—not how much has already been made,” said Levens. “The element senses in real time how much Myc is being made and allows the cell to respond to that.” This mechanism differs from feedback regulation, which relies on the actions of the end-product feeding back on the production process and which occurs widely in cellular signaling pathways, but suffers from inevitable delays. “For the FUSE mechanism to evolve, rapid fluctuations in Myc must be deleterious to the cell,” said Levens.

Based on their work with the *myc* promoter, Levens began working on a way to assess how broadly such changes in DNA structure occur. “I don’t really believe there are principles that apply to a single gene,” said Levens. “When Nature develops a useful trick, she uses it over and over again, modifies it, plays with it, and finds new ways to exploit it.”

Mechanism Meets Action

“One of our lab’s main interests is to understand B-cell activation during the immune response. When I joined the NIH 10 years ago I decided to approach this problem from a nuclear standpoint,” said Rafael Casellas. Casellas’ laboratory uses genomics approaches to study mouse B-cell development.

B cells first arise in the bone marrow. They then migrate to the periphery, where they remain quiescent until encountering antigens, at which point they rapidly proliferate and differentiate into cells with more specialized immune functions to respond to the threat.

Major changes in the quantity and quality of gene expression occur during these different phases.

“Rafael is a very courageous scientist. He is unafraid to take big steps into new areas,” said Levens. “When he heard that we were developing a method to study DNA structure on a genome-wide level, he wanted to apply it.”

The method they developed was based on potassium permanganate (KMnO₄), which oxidizes nucleotides if they are not base-paired, disrupting the DNA structure so it cannot refold. In theory, therefore, it could be used to signal the presence of alternative DNA structures. “The problem is that the average base pair in DNA is flipping in and out of a double helix about 100 times per second, for only a fraction of a microsecond. That’s enough for permanganate to react with them,” said Levens. So his team introduced an enzymatic step to the process in which only DNA that had multiple bases oxidized within a small region would be cut across both strands, reducing the impact of random events.

Getting this method to work required a thorough appreciation of the fundamentals of nucleic acid chemistry, biochemistry, and biophysics. “I’ve had some great teachers for nucleic acid structures and chemistry, and I’ve had a lot of experience—when I was in grad school, there were no kits,” said Levens. He and his colleagues were confident in their success in developing the technique, but they needed to test it. “Because there had been so few studies of ssDNA conformations in living cells that



Rafael Casellas, Ph.D.

had been both characterized and accepted, we needed an unimpeachable gold standard to test whether the new method was working. We picked transcription bubbles.”

Transcriptional Amplification

Transcription bubbles form when RNA synthesis is initiated. After initiation, RNA polymerase begins the process of elongation, traveling along the DNA and locally unwinding or “melting” it to allow focal RNA hybridization within the enzyme. Levens and his colleagues found that they could detect this melting in a Burkitt’s lymphoma cell line with their assay.

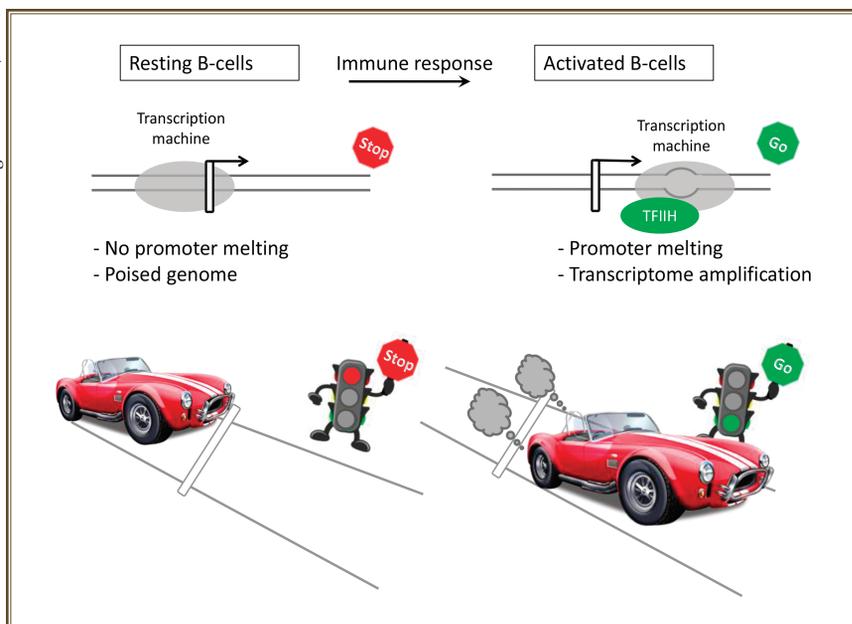
“David’s laboratory had put together a biochemical assay to measure promoter DNA melting in live cells,” explained Casellas. “Conversely, our laboratory had used deep-sequencing protocols to create genome-wide maps of more than 40 chromatin modifications, polymerase recruitment, and RNA synthesis, so we decided to complement these datasets by mapping ssDNA in the entire genome.”

Moving away from their initially encouraging results in Burkitt’s lymphoma cells, Casellas wanted to test resting B cells. To their disappointment, they saw practically no promoter melting whatsoever.

“...we needed an unimpeachable gold standard to test whether the new method was working.

We picked transcription bubbles.”

Image: F. Kozzine, CCR



Unmelted promoters help to limit transcription in resting B cells (left), whereas in activated B cells promoters are melted and transcription has progressed further downstream to support higher levels of expression (right).

“That shouldn’t happen,” said Levens. “We thought it hadn’t worked.” But repetition confirmed the results, as did a comparison with activated B cells.

In a paper published in *Cell* earlier this year, Levens and Casellas described their overall findings. They found that resting B cells have very low basal gene expression, but are poised for a massive increase in gene expression, dependent on promoter melting. Approximately 90 percent of the promoters for genes that will be expressed once the B cell is activated are already loaded with RNA polymerase, but unmelted. Concurrently, these same promoters lack virtually all subunits of the transcription factor IIH (TFIIH) complex, which spurs promoter melting and transcriptional elongation. Basal gene expression levels therefore remain low until the cell is activated, at which point the TFIIH complex is recruited to the gene promoters and gene expression increases 10 to 15 fold.

“If you are a little cell and you want to become big rapidly, how do you do that?” asked Levens. “What’s the switch you have to throw to make everything bigger? When you look at a resting cell, it expresses largely the same genes as fully active cells.” TFIIH is part of that story, but only one part.

Back to *myc*

In a separate line of investigation, Levens’ laboratory was venturing into functional studies of *myc*. “The literature made it sound as if *myc* was some kind of master decision maker, setting precise levels of gene expression. These studies weren’t

“We thought it hadn’t worked.”
But repetition confirmed the results, as did a comparison with activated B cells.”

dealing with the integration of *myc* regulatory mechanisms with its function,” said Levens. Most of the literature was based on work in which *myc* was overexpressed stably or induced at highly unphysiological levels for prolonged periods of time. Levens’ team created knock-in mice in which normally regulated Myc was tagged with the enhanced green fluorescent protein (EGFP). Casellas offered his expertise in activating *myc* in specific cell populations.

“The two projects advanced in parallel originally; we didn’t realize they would reinforce the same point about transcriptome amplification,” said Levens.

Given the vast literature on the topic, it may be surprising that *myc* targets are not well enumerated. Myc is a basic helix-loop-helix leucine zipper (bHLH-Zip) transcription factor, which conventionally means that it dimerizes with a partner Max to bind preferentially to DNA sequences known as E-box motifs.

Working with Keji Zhao, Ph.D., Senior Investigator in Laboratory of Epigenome Biology of the National Heart Lung and Blood Institute (NHLBI) and the inventor of the ChIP-Seq method for genome-wide analysis of chromatin modifications, Levens and his colleagues analyzed genome-wide binding of EGFP-tagged Myc. Their results were consistent with reports that Myc prefers specific to nonspecific binding sites by 200:1. “That sounds like it could be reasonably specific, until you realize that the lac repressor has up to a million-fold preference for specific versus nonspecific binding. 200x is enough to bias, but not to determine targeting,” said Levens.

In activated and resting B cells, they compared the genome-wide distributions of *myc*, gene expression levels, RNA polymerase II binding, and chromatin modifications.

(Photo: R. Baer)



Fedor Kouzine, Ph.D., David Levens, M.D., Ph.D., and Zuquin Nie, Ph.D.

“Putting everything together, it seemed to me that you could make almost all the problems of the *myc* literature go away by positing that it is amplifying expression, not determining gene expression. When you turn *myc* on, everything goes up,” said Levens.

“I think some scientists still have reservations about Myc playing an amplifying role in transcription. One would expect such a reaction when a long-established idea is displaced by a new finding,” said Casellas. “At the same time, because the new model explains the data better, a large fraction of the community has accepted the idea.”

One reason that gene expression amplification may have gone unnoticed in previous studies is the way experiments are conducted. Gene expression is usually normalized when comparing cell populations, meaning that the same amounts of mRNA are typically compared between, for example, resting and activated B cells.

Going Forward

“The most important unanswered questions in biology have not changed that much in the past 25 years,” said Casellas. “What has changed is how we approach them and bioinformatics has definitely revolutionized the way we do it. At the same time, the greatest challenge now is to obtain a holistic view of the cell.”

Bioinformatics has played a critical role in these studies, particularly in analyzing the ssDNA assay data. “In our first data sets, Fedor Kouzine spent three days looking at chromosome 22—the smallest chromosome—by hand. We could see interesting features in the data, but we didn’t have dense enough sampling. It became a computational problem to identify DNA structures,” said Levens.

Levens was co-chairing a trans-NIH search committee for the Tenure-Track Earl Stadtman Investigators program, when he met committee member, Teresa Przytycka, Ph.D.,

Senior Investigator in the National Center for Biotechnology Information. “One day I turned to her and asked if she knew anyone who might be interested in helping us and she said, ‘How about me?’ I was delighted,” said Levens.

Having succeeded with transcription bubbles, the collaborators are now venturing further into uncharted waters to survey other alternative DNA structures. A vast biophysical literature provides numerous examples of structures that do not conform to the Watson-Crick double helix, including the left-handed double helix (Z DNA) and quadruplex structures, in which one DNA strand folds up on itself. Many are likely not to be biologically significant. “The best-understood examples have occurred in bacteria; the literature is less cohesive, although considerable in mammalian cells,” said Levens.

“When we started this work, certainly back when we started working on alternative DNA structures—there was no funding agency in the world that would have not considered this too outside the box to fund. Here, we were able to get together and share resources without petitioning for money up front. The hardest part was the three of us coming together,” said Levens.

To learn more about Dr. Levens’ research, please visit his CCR Web site at <http://ccr.cancer.gov/staff/staff.asp?name=levens>.

To learn more about Dr. Casellas’ research, please visit his CCR Web site at <http://ccr.cancer.gov/staff/staff.asp?name=rcasellas>.

“...because the new model explains the data better, a large fraction of the community has accepted the idea.”