

## GENE EXPRESSION

# Next-generation protein binding

A next-generation sequencing instrument allows deep quantitative measurement of protein-DNA binding affinity.

There is no disputing the impact of next-generation sequencing technology as a consequence of its ability to rapidly sequence immense quantities of DNA. But the open architecture of the machines is also allowing researchers to apply the technology in unintended ways. In January 2011 *Nature Methods* published work using the Roche 454 instrument for synthesizing DNA for synthetic biology. Now Chris Burge and colleagues at the Massachusetts Institute of Technology (MIT) describe the use of an Illumina instrument for direct quantitative measurement of protein-DNA binding affinity.

In 2007, when MIT purchased an Illumina GA instrument, Burge recalls discussing over lunch whether the machine could do something besides sequencing. “We realized it was pretty much an open system. You can control what goes on a flow cell using simple scripts. The machine has good microfluidics and high-end illumination and imaging.” These attributes make it quite flexible.

Researchers in Burge’s laboratory are primarily interested in RNA-binding factors but chose to initially study DNA binding using the machine, which seemed more straightforward. “What if you put a fluorescently tagged protein directly into the flow cell?” Burge recalls. “If the fluorescent tag matches the characteristics of one of the fluorescent dNTPs used for sequencing, then you can tell the machine to image as if it were sequencing another base, and you would see which clusters of DNA the protein was binding to, and based on the strength of the signal, how strongly it bound.”

Burge’s team briefly took over the machine at MIT to test their approach using the yeast transcription factor GCN4, the master regulator of the amino acid starvation response. They made a DNA library of random 25-mers, put it in the instrument and performed a sequencing step that built a complementary strand of nonnative fluorescent dNTPs. They stripped off the nonnative strand by denaturation and synthesized clean DNA with a new primer and Klenow. Then they added GCN4 fused to monomeric (m)Orange, which conveniently matches the spectral characteristics of one of the fluorescent nucleotides used by Illumina, and imaged the bound protein.

Finally, they superimposed and registered the protein binding fluorescence image over the cluster sequencing images. Correction of the intensity for the size of the cluster allowed them to estimate intrinsic affinity for specific sequences. “We were excited when we saw that GCN4 was binding mostly to clusters that contain the consensus 7-mer,” says Burge.

But could they obtain dissociation constants from the data? Such quantitative measurements were not feasible using the state-of-the-art technology, protein-binding microarrays. These measurements required applying the protein to the flow cell at concentrations over a 600-fold range and were done in collaboration with Gary Schroth’s group at Illumina. The effort paid off with dissociation constants that they verified by gel shift assays and with technical replicates.

In addition to the ability to measure dissociation constants, the method has other advantages. Because of the low background afforded by the nonstick surface of the flow cell, the wash step before imaging could be limited to 2 minutes or eliminated altogether. Protein-binding microarrays, in contrast, require a 20-min wash step that removes many weak binding interactions.

But most impressive is the great increase in depth of measurements. They could obtain 100,000 measurements of protein-binding affinity for each 7-mer and 500 measurements each for 12-mers, all in different 25-mer contexts. Averaging over many contexts is arguably more physiologically relevant than using isolated 7- or 12-mers.

The achieved analysis depth enabled them to uncover complicated interdependencies between residues at different positions. In fact, the binding affinity was too complex to describe using a traditional weight matrix and instead they defined the binding using the full set of dissociation constants to all sequences. By capturing weaker and more complex interactions, the data allowed improved predictions of GCN4 binding and activity and revealed subsets of GCN4-regulated promoters with distinct expression kinetics.

**Daniel Evanko**

## RESEARCH PAPERS

Nutiu, R. *et al.* Direct measurement of DNA affinity landscapes on a high-throughput sequencing instrument. *Nat. Biotechnol.* **29**, 659–664 (2011).