## Where protein and RNAs meet

Researchers co-opt a DNA sequencer to quantify RNA-protein interactions on a massive scale.

A few years ago, a team led by Chris Burge at the Massachusetts Institute of Technology found a way to turn a 'next-generation' DNA sequencer into a high-throughput platform for biochemistry. In collaboration with Gary Schroth and his group at Illumina, they flooded a sequencing cell with a fluorescently labeled DNA-binding protein and used the sequencer's imaging capability to measure the interactions between this protein and millions of DNA molecules (Nutiu *et al.*, 2011). Inspired by this work, two groups have now expanded the platform to study RNAprotein interactions *en masse*.

RNA is a flexible molecule; its many secondary structures allow it to assume a range of important regulatory roles in the cell through specific protein interactions. Although high-throughput approaches exist to test for these interactions, the sequencerbased methods mark the first time that it is possible to make quantitative measurements of binding on a large scale.

Illumina sequencing works by imaging the incorporation of labeled nucleotides into clustered copies of immobilized DNA template. For studying RNA, the challenge was to make transcripts that stayed connected to the sequenced DNA cluster for subsequent protein binding measurements.

One solution came from the laboratory of William Greenleaf, working with researchers jointly advised by Howard Chang and Michael Snyder at Stanford University (Buenrostro *et al.*, 2014). They applied an idea from Greenleaf's single-molecule studies of RNA polymerase in graduate school. "We came up with this way of stalling the polymerase after it makes an interesting RNA structure and then holding onto that structure," says Greenleaf.

In their analysis of RNA on a massively parallel array (RNA-MaP) method, each DNA template includes a promoter to initiate transcription and ends with a biotin moiety that is bound by the bulky streptavidin protein, which halts RNA polymerase and causes it to act as a tether between RNA and template. They developed software to quantify fluorescent protein

binding from images from the sequencer. The researchers applied RNA-MaP to the MS2 viral coat protein, which binds to a short RNA hairpin. They measured equilibrium binding affinities of a comprehensive library of mutated target RNAs at different protein concentrations, as well as dissociation rates in the presence of unlabeled protein competitor to determine the permanence of binding. By determining how primary and secondary structure relates to binding energies, they could "breathe life" into the crystal structure, says Greenleaf. The data allowed them to reconstruct the order of mutations that RNA may undertake while maintaining affinity for MS2 during RNA-protein evolution.

Separately, the lab of John Lis at Cornell University, also in collaboration with the Schroth group, came up with a similar solution using their expertise in polymerase pausing (Tome *et al.*, 2014). "We had been agonizing over how to characterize our RNA aptamers at the time," says Lis. The Lis group uses aptamers—synthetic RNAs engineered for specific binding—as inducible and highly selective protein inhibitors in living systems. Rounds of selection and amplification can 'evolve' aptamers with binding specificity, but characterizing the results of selection is a bottleneck.

The Lis team's high-throughput sequencing-RNA affinity profiling (HiTS-RAP) method achieves tethering by including a site in the template that is bound by Tus protein, which causes the polymerase to stall. They used HiTS-RAP to probe libraries of mutated aptamers that bind GFP and NELF-E, a



Measuring RNA-protein interactions on a sequencer requires linking RNA to a DNA template.

protein required for RNA polymerase pausing, mapping binding affinities and residues critical for interaction. Two of the GFP aptamer variants had stronger binding than the strongest found by selection methods.

"The fact that you can take something that is very, very difficult to get in the first place and then improve on it is something we think will be generally useful," says Lis.

Both groups used the older GAIIx sequencer, which is easy to program and allowed them to query up to millions of sequences. But applications such as screening a mammalian transcriptome will require higher capacity, which could come from using newer sequencers.

The studies give just a taste for what is possible at this new scale. Greenleaf and Lis point out that the platforms can be used as tools to study cooperative binding with multiple labeled proteins, to profile basic RNA structures, and to screen entire libraries of RNA molecules that have been artificially selected for binding a target or enriched for binding to cellular proteins. **Tal Nawy** 

## iat Nawy

## **RESEARCH PAPERS**

Buenrostro, J.D. *et al.* Quantitative analysis of RNAprotein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* doi:10.1038/nbt.2880 (13 Apr 2014).

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

Tome, J.M. *et al.* Comprehensive analysis of RNAprotein interactions by high-throughput sequencing– RNA affinity profiling. *Nat. Methods* doi:10.1038/ nmeth.2970 (8 May 2014).