

TECHNIQUE

SMiLE-seq illuminates transcription factor motifs

A study published recently in *Nature Methods* describes a new technique, called selective microfluidics-based ligand enrichment followed by sequencing (SMiLE-seq), which enables the rapid identification of the binding motifs of single and dimeric transcription factors (TFs).

Existing methods that probe protein–DNA interactions have characterized less than 50% of single human TFs. The search for specific TF motifs is complicated by the fact that many TFs can form heterodimer configurations, in addition to forming homodimers, and may bind DNA only when forming a dimer with a specific partner. Now, Isakova *et al.* have developed a novel microfluidic platform, SMiLE-seq, that can quickly identify the binding specificities of TF homodimers and heterodimers.

SMiLE-seq involves capillary loading of *in vitro*-transcribed and -translated bait TF, target double-stranded DNA and the synthetic polymer poly-dIdC (which acts as a competitor to inhibit nonspecific interactions) into a microfluidic device. Following this, target TF–DNA complexes are

immunotrapped on a chip surface. Bound DNA is then eluted, amplified and subjected to high-throughput sequencing and a hidden Markov model (HMM)-based TF motif discovery pipeline.

To validate the new method, the authors tested 58 well-characterized human, mouse and *Drosophila melanogaster* TFs from distinct structural classes in their system. Importantly, the team was able to derive quantitative DNA-binding models for all tested proteins; these models corresponded with the previously described DNA binding motif for each protein. Additional experiments on previously characterized TFs established that SMiLE-seq does not exhibit bias towards TFs that bind to DNA with strong affinity.

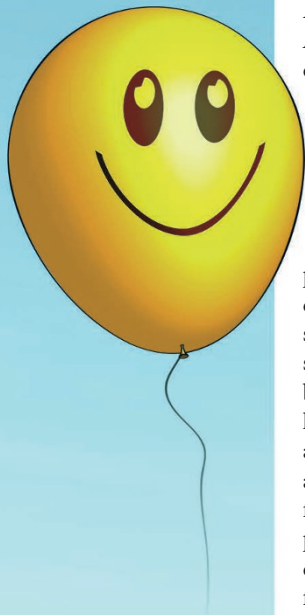
To assess the ability of SMiLE-seq to identify the DNA binding motifs of TF heterodimers, on-chip dimer formation was visualized and monitored using multiple fluorescent protein fusions. Motifs obtained for well-characterized heterodimer pairs, including mouse PPAR γ –RXR α and human ARNTL–CLOCK, closely matched known binding sites, confirming that this technology can

be used successfully to investigate TF heterodimers. Furthermore, when SMiLE-seq was used to investigate nuclear receptor heterodimers, the authors discovered novel binding motifs that may be specific to particular nuclear receptor heterodimer combinations. Lastly, this technology was used to examine DNA binding motifs for uncharacterized Krüppel-associated box (KRAB) containing C2H2 zinc-finger proteins (ZFPs). The majority of newly discovered binding motifs for ZFPs matched those that had been computationally predicted, and ChIP-exo (a modified version of chromatin immunoprecipitation sequencing) validated three out of four of the newly detected motifs. These findings confirm the value of SMiLE-seq for exploring previously unknown protein–DNA interactions.

In summary, this study presents a new sequencing method, SMiLE-seq, which can be used to perform *de novo* discovery of the binding motifs of TF homodimers and heterodimers. In the future, this technology could be used to identify uncharacterized protein–DNA interactions and may also be adapted to explore the binding motifs of RNA binding proteins.

Shimona Starling

ORIGINAL ARTICLE Isakova, A. *et al.* SMiLE-seq identifies binding motifs of single and dimeric transcription factors. *Nat. Methods* <http://dx.doi.org/10.1038/nmeth.4143> (2017)



“this technology could be used to identify uncharacterized protein–DNA interactions”

P. Morgan/Macmillan Publishers Limited