Research Highlights

Papers from the literature selected by the Nature Biotechnology editors. (Follow us on witter, @NatureBiotech #nbtHighlight)

Discovery and saturation analysis of cancer genes across 21 tumour types
Lawrence, M.S. *et al. Nature* doi:10.1038/nature12912 (5 January 2014)

Rewiring yeast sugar transporter preference through modifying a conserved protein motif
Young, E.M. *et al. PNAS* **111**, 131–136 (2014)

Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas sysem
Chen, B. *et al. Cell* **155**, 1479–1491 (2013)

Genome-scale CRISPR-Cas9 knockout screening in human cells
Shalem, O. *et al. Science* **343**, 84–87 (2014)

Genetic screens in human cells using the CRISPR-Cas9 system
Wang, T. *et al. Science* **343**, 80–84 (2014)

PIQ-ing into chromatin architecture

Sebastian Rieck & Christopher Wright

Decrypting the logic of gene regulation is aided by a breakthrough method for analyzing chromatin-accessibility data.

The gene expression program of any cell depends on complex, dynamic patterns of transcription factor binding. Yet, extracting the biological meaning of such patterns has been limited by a dearth of technologies that can identify exactly where individual transcription factors are bound across an entire genome. In this issue, Sherwood *et al.*[1] describe notable progress toward this goal with a new computational method, called protein interaction quantification (PIQ), which from a single experiment infers the effects on chromatin architecture produced by hundreds of transcription factors. PIQ creates a cumulative catalog of the distinctive imprints caused by each DNA-binding transcription factor, in combination with their associated coregulators, on the patterns of DNA fragments produced by digesting chromatin with the enzyme DNase I. It uses these distinctive, individual DNase-sensitivity profiles to interpret how multiple transcription factors contribute to the genome-wide DNase hypersensitivity (DNase-seq) profile.

The study presents several important advances. First, using PIQ to reveal the bound sites of hundreds of transcription factors could, in some cases, allow a single DNase-seq experiment to replace hundreds of ChIP-seq experiments, each of which detects the binding of a single transcription factor. Second, compared with previous methods, PIQ allows a more sensible comparison of results from multiple experiments—for example, time points along a cell-differentiation process. The authors apply PIQ to a time series of DNase-sensitivity information taken from a cell differentiation model to enable a systematic categorization of transcription factors into the following functional subdivisions: pioneers, which are capable of opening closed chromatin; settlers, which require open chromatin to bind; and migrants, which appear to require more than open chromatin to bind, such as the presence of coregulators (**Fig. 1**). Third, the analysis seems to greatly expand the list of potential pioneers, adding ~120 new ones to the handful identified until now. Finally, the authors identify a new type of pioneer—the 'directional pioneer'—which induces chromatin to open preferentially in one direction.

Ordered, collaborative interactions between transcription factors and their coregulators affect the sequential building of patterns of chromatin and thereby gene activity across the genome. The experimental design and computational output emerging from Sherwood *et al.*[1] provide a template for unraveling the gene regulatory networks that dictate complex phenomena such as the state of cell stemness or differentiation. The authors also speculate that biologically critical gene-regulatory interactions, which depend upon both the distance and relative chromosomal location of motifs that bind directional pioneers and settlers (**Fig. 1b**,**c**), could be a significant component affecting the manner in which an organism's overall genome sequence may evolve over time.

Only a few transcription factors (such as Foxa, GATA, Pu.1 and glucocorticoid receptor) are known to have the ability to 'crack open' condensed chromatin. These pioneers function as gatekeepers: without them, other regulatory proteins cannot access their target sites (**Fig. 1a**). Sometimes even within the same cell, pioneer transcription factors activate or repress different genes; regulated toggling between such dual functions would be a potent influence, for instance, in how cells choose between separate pathways of cell differentiation[2]. Despite huge amounts of work on individual factors, and the invention of methods that attempt to describe chromosome-wide DNA-protein interactions, a systematic decoding of how chromatin architecture is progressively built during differentiation is still far from complete. One hurdle is that genome-wide methods to assay chromatin accessibility are technically challenging and inherently noisy.

The method of Sherwood *et al.*[1] (**Fig. 2**) for discovering bound transcription-factor sites and their particular effects on chromatin structure seems to offer a large improvement over previous approaches for analyzing DNase-seq data sets. DNase-seq involves subjecting carefully prepared chromatin to DNase-I cleavage, and using high-throughput techniques for en masse sequencing of the released DNA fragments. Accurately aligning the DNA fragment ends on the genome produces a DNase-seq profile indicating those regions most sensitive to DNase I and therefore accessible to protein binding. Transcription-factor binding can cause substantial loosening of nearby chromatin structure, increasing the sensitivity of DNA to DNase attack, but directly at the protein-DNA interface there is often a relative protection of DNA against DNase cleavage—a so-called 'footprint'. The end result is a pattern of characteristic local contours in the total DNase-sensitivity profile. The identity of proteins bound to the protected sequences, and thus their local chromatin effects, can be determined *de novo*[3] or conversely by referring to databases of the DNA motifs recognized by transcription factors[4].

Sherwood *et al.*[1] (**Fig. 2**) take the latter approach—using known transcription factor binding motifs—but made three key refinements to improve interpretation of DNase profiles[3,4]. First, PIQ smooths raw sequence read information, decreasing the noise inherent to DNase-seq[5]. In this step, the algorithm can also incorporate data from additional DNase-seq experiments to reduce noise even further, which strengthens the ability to carry

Sebastian Rieck and Christopher Wright are at the Department of Cell and Developmental Biology, Center for Stem Cell Biology and Program in Developmental Biology, Vanderbilt University, Nashville, Tennessee, USA.
e-mail: chris.wright@vanderbilt.edu