

STEM CELLS

Lineage tracing with *Polylox* barcodes

A recombinase-based barcoding approach enables fate mapping in the hematopoietic system under physiological conditions.

Determining the developmental potential of stem cells is interesting not only in the context of organismal development, but also under steady-state conditions in adult stem cell contexts such as the hematopoietic system. Conceptually, different strategies are available to study this potential. Fluorescent-reporter-based systems (such as Brainbow or Confetti) and molecular barcoding approaches have been harnessed for lineage tracing.

Brainbow and Confetti have limitations, as only few reporters are available, and are mainly useful for tissues that do not exhibit much cell migration. On the other hand, molecular barcoding based on viral insertion sites or virally delivered barcodes provides access to many barcodes but typically requires the cells to be manipulated *ex vivo*. In addition, viral or transposon insertion sites may lead to adverse effects in cells.

To overcome these limitations, Hans-Reimer Rodewald from the DKFZ in Heidelberg and his collaborators developed the *Polylox* system. The *Polylox* locus consists of an array of unique DNA sequences that are interspersed by *loxP* sites. Barcodes are created *in vivo* through Cre-dependent recombination.

“In the beginning, nobody believed that this would ever work because Cre was thought to completely react on all available *loxP* sites on such a locus,” says Rodewald. Fortunately, the team’s instinct was correct, and they found that Cre-mediated recombination at the *Polylox* locus is incomplete when Cre is weakly or briefly expressed and can potentially lead to the generation of 1.8 million unique barcodes. However, the utility of individual barcodes varies according to their frequency. “You get some barcodes more frequently than others,” explains Rodewald. Thomas Höfer’s group at the University of Heidelberg, together with Rodewald and colleagues, addressed this issue both computationally and experimentally. They determined that certain barcodes have a low probability of being generated, and these are typically

PROTEOMICS

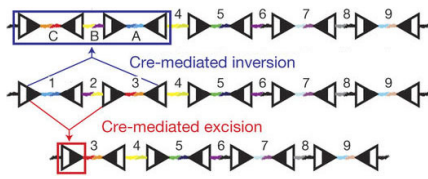
THE PROOF OF SPLICING IS IN THE PROTEOME

An integrated workflow allows the quantitative interrogation of the impact of alternative splicing at the proteome level.

It is generally assumed that changes in mRNA splicing patterns will impact the expression of the corresponding protein products. Yet, an experimental measure of the effect of alternative splicing on proteome diversity remains lacking. By manipulating splicing on a genome-wide level and developing a strategy that integrates RNA sequencing with a quantitative data-independent acquisition mass spectrometry approach for measuring protein abundance, Ruedi Aebersold from the ETH Zurich, Ashok Venkitaraman from the University of Cambridge, Vihandha Wickramasinghe from The Peter MacCallum Cancer Centre in Melbourne and their colleagues aimed to get a handle on the global effect of splicing on proteome composition.

In order to alter splicing of a large number of mRNAs, the researchers used RNA interference to deplete PRPF8—the core spliceosome U5 small nuclear ribonucleoprotein (snRNP) component—which they had previously shown altered the splicing pattern of over 3,000 transcripts through alternative exon usage and/or intron retention. Using a data-independent acquisition method, SWATH mass spectrometry (SWATH-MS), and the OpenSWATH software, the team reproducibly measured the abundance of ~14,000 peptides corresponding to ~2,800 unique protein-coding genes. Of those proteins, about 1,500 exhibited altered abundance of at least one peptide following PRPF8 knockdown.

Integrating the transcriptomic and proteomic data sets in order to decipher the effect of splicing on the proteome is problematic. This is because exon usage does



Polylox is a recombination-based fate-mapping tool. Figure adapted from Pei, W. *et al.*, *Nature Research*.

not shared across different experiments and are thus highly informative in fate-mapping experiments.

Rodewald and colleagues applied *Polylox* fate mapping to the mouse hematopoietic system. “I had tried for many years to get closer to the physiology,” says Rodewald. He thinks that currently used assays involving *ex vivo* manipulation of hematopoietic stem cells and subsequent transplantation into

irradiated mice are quite artificial. With “the transplantation approach, you empty the system... and then you have to start over again,” says Rodewald. Using the *Polylox* system, the fate of these stem cells can be studied under steady-state conditions, leading to different results from those in transplantation experiments. For instance, under steady-state conditions, the Rodewald and Höfer teams found in an earlier study that many stem cells contribute to the replenishment of the system, but individual contributions are small.

The *Polylox* system shares some features with the recently described CRISPR–Cas9-based lineage-tracing approaches, which also allow barcode generation *in vivo*. Yet Rodewald sees advantages in the *Polylox* system. Each of the individual DNA blocks is about 170 base pairs long, and it is necessary to sequence across the whole locus to obtain the full barcode information. “Even if you have poor sequencing or PCR problems or whatever, you will always recognize these blocks,” Rodewald points out. He thinks that the CRISPR–Cas9-based systems in which barcodes can differ in just single nucleotides may be more prone to PCR and sequencing artifacts than the *Polylox* system is.

While Rodewald and his team have applied the *Polylox* technology for fate mapping to the hematopoietic system, it should be possible to use this technology in other tissues or to answer other questions, as long as suitable Cre driver lines are available.

Nina Vogt

RESEARCH PAPERS

Pei, W. *et al.* *Polylox* barcoding reveals haematopoietic stem cell fates realized *in vivo*. *Nature* **548**, 456–460 (2017).

not inform on individual transcript expression levels, and differentially used exons (and the associated peptides) can map to several transcripts from the same gene. To overcome this difficulty, the researchers developed an approach that considered transcripts as whole units, first estimating expression levels with the existing MMSEQ and MMDIFF tools to define genes with differential transcript usage (DTU). DTU events are called when there is a change in the relative abundances of transcripts between experimental conditions. The researchers observed 338 DTU events upon PRPF8 depletion, while about 2,000 genes displayed only changes in expression. While several alternative splicing events had a detectable effect on proteome composition, overall the results confirmed the prevailing view that protein abundance is generally correlated to transcript levels. Intron retention is a form of alternative splicing that is believed to play important regulatory roles in gene expression, and the researchers also observed that intron retention tended to result in reduced expression of the corresponding protein.

Although several studies have aimed at establishing the contribution of mRNA isoform abundance to protein expression level, it has remained difficult to reach general conclusions. With this integrated approach, the researchers have now demonstrated that it is feasible to combine transcriptomics and quantitative proteomics data to obtain information from peptides that map to more than one transcript of the same gene. This methodology may be useful in future efforts aimed at gaining a more granular yet global view of how variation in splicing alters cell physiology.

Stéphane Larochelle

RESEARCH PAPERS

Liu, Y. *et al.* Impact of alternative splicing on the human proteome. *Cell Rep.* **20**, 1229–1241 (2017).