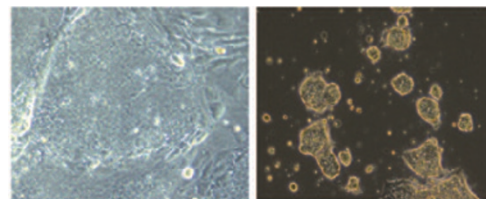


STEM CELLS

Naive human pluripotent stem cells

Researchers define conditions in which human stem cells can be maintained in the ground state of pluripotency.

Embryonic stem cells are derived from a transient population of cells in the early embryo, the inner cell mass (ICM), which gives rise to the various lineages of the developing organism. Cells of the ICM are not meant to stay pluripotent for very long. But this is precisely what researchers would like embryonic stem cells in the culture dish to do, in order to both study them as an *in vitro* model of development



Primed (left) and naive (right) human pluripotent stem cells. Image courtesy of J. Hanna.

and use them as a source of other cell types in a myriad of applications in the laboratory and clinic. The trick, however, is in working out what the best culture conditions might be.

Work in the mouse in recent years has shown that culture in LIF-2i conditions (a combination of leukemia inhibitory factor and inhibitors of GSK-3 β and ERK1/2) yields cells that are in a developmentally naive 'ground state' of pluripotency. These cells, it is thought, are the closest *in vitro* equivalent to the *in vivo* ICM. Conventional human embryonic stem cells (hESCs), however, are morphologically, molecularly and functionally quite distinct from these naive mouse cells. The question has remained open whether conditions can be found in which human cells can be maintained stably in a naive-like state.

In a recent paper, Jacob Hanna and colleagues at the Weizmann Institute in Israel describe such conditions. They started not with embryonic stem cells but with cells they had reprogrammed to pluripotency from fibroblasts, using doxycycline-inducible versions of the

SEQUENCING

END-TO-END RNA SEQUENCING

Long-read sequencing uncovers transcript features missed by short-read methods.

As a pioneer of deep RNA sequencing, Mike Snyder surveyed what he called the transcriptional landscape of yeast in 2008. By sequencing RNA messages many times over, it was possible to map the boundaries of expressed sequences across the yeast genome. Five years on, he and his team at Stanford University are still applying the technique to improve transcript annotation, using a technology that in some ways gives a more complete picture of transcription.

In the original study, each sequenced fragment, or 'read', contributed around 30 nucleotides of useable sequence. Read lengths have improved many fold since then but still fall well short of full-length transcripts and thus have to be pieced together computationally. "As Hagen Tilgner in my lab put it, 'the way we do RNA-seq now ... is you take the transcriptome, you blow it up into pieces and then you try to figure out how they all go back together again'," says Snyder. "If you think about it, it's kind of a crazy way to do things."

Reconstruction from small sequence fragments is complicated by the fact that modular expressed elements—the exons—are spliced together in the cell, often in complex permutations, to create the final transcript isoforms. Mapping short fragments onto this complexity in the midst of sequencing errors and biases is challenging.

The alternative? "The way to get a really accurate picture of the transcriptome," says Snyder, "is to get full-length molecule sequence." His team used the long-read capability of the Pacific Biosciences platform to sequence RNA from 20 human tissues and organs. The technology produces reads of around 7 kilobases but with a relatively high error rate. Circularizing templates can remove error by allowing the sequencing enzyme to traverse each molecule repeatedly so that errors can be canceled out in a final consensus sequence.

standard Yamanaka reprogramming factors. Previously they had reported that, in the continued presence of doxycycline, which maintains expression of reprogramming factors, they could grow naive-like human pluripotent cells in LIF-2i (conventional human pluripotent cells, by contrast, differentiate in LIF-2i). Although the researchers identified small molecules that could keep the cells naive even upon withdrawal of doxycycline, they could not stably maintain these without also ectopically expressing at least two reprogramming factors. In their present work, Hanna and colleagues searched for conditions that maintain naive-like human cells indefinitely in LIF-2i, in the absence of additional genetic manipulation.

After an initial screen of 16 factors that modulate a variety of signaling pathways, followed by much combinatorial optimization, the researchers settled on a cocktail they term naive human stem cell medium (NHSM). The resulting cells, molecularly characterized in fine detail and reported to be very similar to naive mouse cells, can be derived by reprogramming from somatic cells, from already established hESC lines and directly from blastocysts. The naive human cells have gene expression profiles more similar to the ICM than do conventional human pluripotent stem cells and are less heterogenous in their gene expression as well.

Perhaps most notably, when the human cells are injected into mouse morulae and the chimeric embryos allowed to develop for up to 8 days in the mouse *in vivo*, human cells can still be detected in the embryos by confocal microscopy at embryonic day 10.5, in contrast to previous reports with conventional hESCs. To what extent these cells actually contribute to developing tissues in this interspecies system remains unknown.

“One way that people deal with heterogeneity [in stem cells] is to accept it, to say let’s profile the cells and predict their bias” says Hanna. “But there’s another approach, which I prefer, and that’s to say let’s make them more uniform.” He predicts that naive human cells will prove more adept at differentiating along various lineages. The experiments are undoubtedly being done, and time will tell.

Natalie de Souza

RESEARCH PAPERS

Gafni, O. *et al.* Derivation of novel human ground state naive pluripotent stem cells. *Nature* doi:10.1038/nature12745 (30 October 2013).

From nearly half a million of these circular consensus sequence reads, the group characterized ~25,500 spliced transcripts from close to 14,000 genes; around one-tenth of the transcripts had not been annotated previously. Sequences under 1.5 kilobases (close to the average length of human transcripts) were nearly all complete transcripts. The fraction of full-length sequences dropped with transcript length, especially after 2.5 kilobases, potentially reflecting technical issues with performance of the reverse transcriptase, efficiency of circularization or effects of sample handling on longer sequences. They detected many new long noncoding RNA sequences with confidence.

The platform is unique in that, given enough starting material, it does not require fragmentation or amplification, which can introduce substantial biases in the sequence output. The main bottleneck is throughput. “That’s the only limitation right now, you get something like 75,000 reads per run,” says Snyder. “But the runs are fast,” he adds. As throughput improves, large-scale quantitative transcriptome measurement will become more practical.

Analyses of a broad range of tools to address the spliced-alignment problem (p. 1185) and transcript reconstruction (p. 1177) presented in this issue show that even sophisticated computation currently cannot be used to identify or assemble a sizeable fraction of full-length transcript isoforms. Improved experimental methods may be needed to improve on this. Snyder is confident that others will embrace long reads for sequencing RNA. “This is clearly the way the field is going to move,” he says.

Tal Nawy

RESEARCH PAPERS

Sharon, D. *et al.* A single-molecule long-read survey of the human transcriptome. *Nat. Biotechnol.* doi:10.1038/nbt.2705 (13 October 2013).

Engström, P.G. *et al.* Systematic evaluation of spliced alignment programs for RNA-seq data. *Nat. Methods* **10**, 1185–1191 (2013).

Steijger, T. *et al.* Assessment of transcript reconstruction methods for RNA-seq. *Nat. Methods* **10**, 1177–1184 (2013).