produce 'bona fide' iPS cell lines, as confirmed by expression of multiple pluripotency genes, demethylation of the *OCT4* and *NANOG* promoters and the ability to form well-differentiated teratomas when injected into immunocompromised mice.

The authors call these colonies "type III colonies." Two other colony types that are morphologically similar to the type III colonies come out of their experiments (**Fig. 1**). Like type III colonies, type I and type II colonies are expandable, but neither silences GFP. Furthermore, type I colonies do not activate SSEA-4 or TRA-1-60, and type II colonies express SSEA-4 but not TRA-1-60. Molecular and functional analysis of type I and II colonies show that they are not fully reprogrammed. For example, neither produces the well-differentiated teratomas one normally sees with human embryonic stem cells.

The authors conclude that type I and II cells are stuck in incompletely reprogrammed states, although type II cells do rarely spontaneously convert to type III cells. Why do these cells get locked into incompletely reprogrammed states? Hints come from previous work done in the mouse system. Mikkelsen *et al.*[6] found that incompletely reprogrammed mouse colonies continue to express lineage-specific genes and show abnormal DNA methylation of the promoters of key pluripotency genes. Manipulation of ongoing lineage-specific gene expression (by siRNA) and of DNA promoter methylation (by the DNA methyltransferase inhibitor 5-azaC) reverted some of the incompletely reprogrammed cells to 'true' iPS cells. Interestingly, Chan *et al.*[1] also observe inappropriate hypermethylation of the endogenous promoters of *OCT4* and *NANOG*, two key pluripotency transcription-factor genes, in their type I and II colonies. Therefore, inhibition of DNA methyltransferases might similarly help convert these colonies to type III colonies.

What enables incompletely reprogrammed cells to self-renew (that is, grow indefinitely while maintaining their identity)? One culprit is the exogenously introduced c-MYC. c-MYC is a well-known oncogene that can induce many cell types to inappropriately self-renew. Indeed, removal of c-MYC from the cocktail of reprogramming factors dramatically decreases the number of incompletely reprogrammed colonies in the mouse and human systems[7–9]. Therefore, it might be expected that the relative number of type I and II colonies would also be significantly reduced in the absence of c-MYC.

The ability to use combinations of markers to identify cells at different stages along the road to pluripotency is a valuable resource in efforts to understand reprogramming at the molecular level. Although we know something about the beginning and end of reprogramming, the process in between remains largely a black box. Using specific combinations of markers, one can potentially sort out cells at various intermediate stages and determine their transcriptional and epigenetic profiles. Such studies will no doubt shed light on the molecules and pathways involved and help identify the barriers that must be overcome in the making of an iPS cell.

1. Chan, E.M. *et al. Nat. Biotechnol.* **27**, 1033–1037 (2009).
2. Hochedlinger, K. & Plath, K. *Development* **136**, 509–523 (2009).
3. Brambrink, T. *et al. Cell Stem Cell* **2**, 151–159 (2008).
4. Stadtfeld, M., Maherali, N., Breault, D.T. & Hochedlinger, K. *Cell Stem Cell* **2**, 230–240 (2008).
5. Lowry, W.E. *et al. Proc. Natl. Acad. Sci. USA* **105**, 2883–2888 (2008).
6. Mikkelsen, T.S. *et al. Nature* **454**, 49–55 (2008).
7. Nakagawa, M. *et al. Nat. Biotechnol.* **26**, 101–106 (2008).
8. Marson, A. *et al. Cell Stem Cell* **3**, 132–135 (2008).
9. Judson, R.L., Babiarz, J.E., Venere, M. & Blelloch, R. *Nat. Biotechnol.* **27**, 459–461 (2009).

Targeted sequencing with microfluidics

Ewen F Kirkness

Combining microdroplet PCR with flow-cell technologies provides a novel approach for sequencing thousands of genomic targets.

Sequencing targeted regions of DNA rather than complete genomes is likely to remain the preferred approach for most genomics applications, at least until the cost of whole-genome sequencing becomes competitive. In this issue, Tewhey *et al.*[1] introduce a new method for targeted sequencing that exploits the capacity of microfluidic technology for multiplexed analysis on small amounts of sample. By coupling second-generation sequencing with PCR conducted in microdroplets, they demonstrate parallel characterization of thousands of targeted sequences. The technique has sufficient specificity and sensitivity for accurate genotyping and represents an attractive alternative to current methods[2–7] for capturing genomic regions of interest.

A good DNA targeting technology must capture sequences both comprehensively and uniformly. Whereas comprehensiveness measures the fraction of targeted sequence that is captured, uniformity reports bias in the coverage of captured sequence. Uniformity is important because it determines how many times each sample of captured DNA must be sequenced to call variants reliably across most of the targets; if the capture is very biased, it may be necessary to sequence the entire sample to an average depth of 1,000× in order to obtain 10× read depth for the poorly captured targets. Biased capture therefore incurs higher sequencing costs.

Ewen F. Kirkness is at the J. Craig Venter Institute, Rockville, Maryland, USA.
e-mail: ekirkness@jcvi.org

The microdroplet target capture method developed by Tewhey *et al.*[1] uses a microfluidic chip to merge droplets (~20 pl each) containing ~3 pg of 2- to 4-kb biotinylated fragments of template DNA with droplets containing a pair of PCR primers designed to amplify a specific sequence. Each run of the microfluidic instrument yields more than a million of these merged droplets. After PCR in a single tube, the emulsion is broken, and the PCR products from all the droplets are combined for sequencing.

The authors generated ~4,000 distinct PCR products (combined length of 1.35 Mb), with the downstream sequencing performed on either the Illumina GA-II or Roche 454 sequencers. After sequencing was done to a depth of ~200× (Illumina), 98% of all amplicons had at least 95% of their bases covered by at least five sequence reads. They also confirmed 99.3% of 2,226 HapMap genotypes tested. These results show that microdroplet PCR demonstrates high uniformity of coverage and comprehensiveness.

How does microdroplet PCR compare with existing methods for target capture? The authors attempted a comparison with conventional PCR. First, they sequenced 457 exons after amplifying 172 kb of genomic DNA by either microdroplet PCR or conventional PCR. Sequencing of pooled products to a depth of ~200× on the Illumina GA platform enabled them to call variants at 97.5% of the loci previously identified as variant via HapMap genotyping. These variant calls were concordant at

Table 1 Comparison of studies demonstrating massively parallel capture and sequencing of targeted genomic regions

Capture approach	Targeted DNA (Mb)	Raw sequence (fold coverage)	Percentage of target passing criteria for variant detection	Reference
Microdroplet PCR	1.4	~270×	94%	1
Molecular inversion probes	1.4	~400×	75%	5
Solution hybridization	3.7	~230×	89%	6
Chip hybridization	26.6	~240×	96%	7

The microdroplet PCR approach of Tewhey *et al.*[1] was used to target a similar amount of genomic sequence as methods based on molecular inversion probes[5] and solution hybridization[6] (1–5 Mb), whereas a study that used chip hybridization[7] sequenced the majority of human protein-coding exons (~27 Mb). All data were obtained using the Illumina GA-II platform. The sequence coverage represents the generated sequence data that was used for mapping and was estimated from details provided in the references indicated.

99.1% of 2,390 comparisons, indicating a high level of accuracy for variant detection at this depth of sequencing. The authors conclude that the coverage and genotyping accuracy of microdroplet PCR is similar to that obtained using conventional PCR.

Comparisons of microdroplet PCR with other targeted capture methods are difficult, in part because previous studies used different targets (with different GC content and homologs in the genome) and different depths of sequencing coverage. Furthermore, it is not always known whether the probes and primers used for capture were chosen for optimal functionality. Nevertheless, the authors note that capture of off-target DNA and the accuracy of variant calling by their method are in the same range as those of enrichment strategies based on molecular inversion probes[5], solution hybridization[6] or chip hybridization[7] (**Table 1**). Uniformity of coverage for the microdroplet PCR method permits a high variant detection rate at a given level of sequence coverage (**Table 1**, column 4).

The value of uniformity of coverage is determined largely by the relative costs of the capture and sequencing technologies. In other words, poor uniformity in the yield of captured DNA can often be salvaged by deeper sequencing if the sequencing costs are a minor component of the overall budget. Given that the capture efficiency and variant calling accuracy appear so similar for the different capture technologies, it is likely that the choice of one platform over another will largely be determined by their relative operational costs. The potential of microdroplet PCR to be highly automated should reduce labor costs. However, the relative costs of reagents for the different capture methods are at present unclear, particularly as some systems may be more amenable to multiplexing.

Preliminary data indicate that future runs of the microfluidic instrument might be scalable to 20,000 PCR products by including multiple primer pairs in each droplet. However, even this level of throughput is approximately one-tenth of that required for amplification of the human exome. The potential to increase the length of amplified products using long-range PCR is limited by the current requirement to shear the template DNA into 2- to 4-kb fragments, making the present version of the microdroplet method best suited for amplification of exon subsets (e.g., CAN genes, which are frequently mutated in tumors[2]) or specific genomic loci in the megabase size range.

In addition, although microdroplet PCR appears to capture less off-target DNA and hence yields less unwanted sequence data than other target capture strategies[5–7], it is limited by the need to amplify and sequence the flanking sequences of any target of interest. Indeed, the targeting of 457 exons for validation of the process yielded products that consisted mainly of flanking intron sequences. Consequently, in designing primers one must compromise between the genomic locations that are optimal for PCR priming and those that are closest to the target sequence.

Finally, some questions remain regarding the ability of microdroplet PCR to amplify genomic loci with very high or very low GC content. It is unclear whether it will be possible to identify one set of reaction conditions that can be applied to all targets or whether it will be necessary to conduct multiple amplification reactions that would be combined before sequencing.

With the explosive development of second-generation sequencing technologies, rapid analysis of entire exomes or subsets of exomes is now feasible. Until the costs of whole-genome sequencing fall substantially, the study of genome sequences in large cohorts will require optimized methods for targeted DNA capture. The high uniformity of coverage and comprehensiveness provided by microdroplet PCR offers a valuable new approach for focused genome analysis.

1. Tewhey, R. *et al. Nat. Biotechnol.* **27**, 1025–1031 (2009).
2. Sjoblom, T. *et al. Science* **314**, 268–274 (2006).
3. Hodges, E. *et al. Nat. Genet.* **39**, 1522–1527 (2007).
4. Albert, T.J. *et al. Nat. Methods* **4**, 903–905 (2007).
5. Turner, E.H. *et al. Nat. Methods* **6**, 315–316 (2009).
6. Gnirke, A. *et al. Nat. Biotechnol.* **27**, 182–189 (2009).
7. Ng, S.B. *et al. Nature* **461**, 272–276 (2009).

A systems view of host defense

Daniel E Zak & Alan Aderem

Large-scale perturbations unravel the complex networks of activated dendritic cells.

The body's first line of defense against infection is the innate immune system, which recognizes conserved molecular patterns on microbes via receptors such as toll-like receptors (TLRs)[1]. TLRs transduce the information into pathogen-specific immune responses involving networks comprising ~2,000 genes[2] (**Fig. 1**). A new study by Amit *et al.*[3] in *Science* describes an important advance in elucidating these networks. By combining expression profiling and large-scale perturbations, the authors discover many novel regulators and interactions that might control the physiological processes induced by TLRs. These network components represent novel candidates for detailed analysis and potential targets for the development of vaccines and antimicrobial or anti-inflammatory drugs.

Addressing the complexity of innate immunity requires the large-scale approaches of systems biology. Previous work has focused primarily on the responses of the transcriptome to TLR activation. Our group has studied these responses in macrophages, integrating transcription factor binding site analysis and dynamic computational modeling to identify small novel regulatory networks

Daniel E. Zak and Alan Aderem are at the Institute for Systems Biology, Seattle, Washington, USA.
e-mail: dzak@systemsbiology.org