Hdac2 is destroyed by valproic acid and is expressed at considerably higher levels in mouse fibroblasts than in human fibroblasts. Accordingly, reprogramming of mouse Hdac2$^{-/-}$ fibroblasts does not require valproic acid, and miR-302/367 alone can efficiently reprogram these cells.

What is the mechanism behind miRNA-mediated reprogramming? miRNAs usually target a large group of mRNAs simultaneously, and reprogramming is known to involve changes in the expression levels of hundreds of genes. That ESC miRNAs enhance reprogramming strongly suggests that they repress the expression of genes that would otherwise act to maintain the differentiated cell state and may therefore represent barriers to reprogramming. Using this rationale, Blelloch and colleagues[3] address the mechanism by which ESC miRNAs enhance reprogramming. They find that introducing synthetic mimics of miR-302b or miR-372 into human fibroblasts enhances the efficiency of both OSKM- and OSK-mediated reprogramming of human fibroblasts. Next, using their published data set[8], they choose a group of ~30 genes targeted by both miR-302b and miR-372 for further analysis. From this list they focus on a subset of 12 genes that respond to miRNA expression in the context of reprogramming. They use short interfering (si)RNAs to individually deplete each of these putative miRNA targets and measure the effect on reprogramming.

These experiments show that knockdown of 3 of the 12 genes (*RBl2*, *CDC2L6* and *RHOC*) enhances reprogramming both with OSKM and with OSK. Individual depletion of an additional three genes (*SMARCC2*, *MBD2* and *MECP2*) enhances reprogramming in one of the two conditions (either OSKM or OSK). As the gene encoding the TGFβ receptor TGFBR2 was among the selected gene set, the authors use a small-molecule inhibitor to confirm the involvement of this signaling pathway and show that the miRNAs directly repress *TGFBR2* mRNA. Another recent paper has described similar findings in mouse cells, namely, that mouse *Tgfbr2* mRNA is regulated by similar miRNAs and that siRNA-mediated knockdown of *Tgfbr2* enhances reprogramming of embryonic fibroblasts[9].

Blelloch and colleagues[3] provide several lines of evidence that miRNAs promote reprogramming of human cells by targeting genes in multiple downstream pathways (**Fig. 1**). First, reprogramming involves a mesenchymal-to-epithelial transition, as in mouse cells. Second, inhibition of any single gene leads to a modest enhancement in reprogramming compared with the effects of the miRNA itself. Finally, reprogramming efficiency is increased by simultaneous inhibition of multiple pathways.

But iPSC formation requires both miR-302-a/b/c/d and miR-367 (ref. 1). What genes are repressed by miR-367? This missing piece of the puzzle remains to be addressed. However, a closely related miRNA, miR-92b, represses expression of Cdkn1c (p57) and is important for cell-cycle progression of human ESCs[10]. Therefore, it is likely that miR-367 facilitates reprogramming at least in part by promoting cell proliferation.

An obvious next experiment is to try to eliminate viral delivery of the miRNA transgene. As Morrisey and colleagues[1] have shown that miR-367 is essential for miRNA-mediated reprogramming, it will be interesting to see whether adding this synthetic miRNA to those used by the Blelloch group[3] will enable the generation of iPSCs without the need for viruses or genome integration. miRNAs may prove particularly useful in this regard because they are small and easily synthesized. Moreover, it is relatively easy to transfect cells with them, and they may not stimulate an innate immune response in the host cell. Once incorporated into the relevant ribonucleoprotein complexes, they are relatively stable, with a reported half-life of several days in cells. These favorable properties, together with the breakthrough proof-of-concept experiments, suggest that miRNAs may be the ideal modality for reprogramming human cells for therapeutic applications.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Anokye-Danso, F. *et al. Cell Stem Cell* **8**, 376–388 (2011).
2. Gonzalez, F., Boue, S. & Izpisua Belmonte, J.C. *Nat. Rev. Genet.* **12**, 231–242 (2011).
3. Subramanyam, D. *et al. Nat. Biotechnol.* **29**, 443–448 (2011).
4. Takahashi, K. & Yamanaka, S. *Cell* **126**, 663–676 (2006).
5. Huangfu, D. *et al. Nat. Biotechnol.* **26**, 795–797 (2008).
6. Martinez, N.J. & Gregory, R.I. *Cell Stem Cell* **7**, 31–35 (2010).
7. Judson, R.L., Babiarz, J.E., Venere, M. & Blelloch, R. *Nat. Biotechnol.* **27**, 459–461 (2009).
8. Melton, C., Judson, R.L. & Blelloch, R. *Nature* **463**, 621–626 (2010).
9. Li, Z., Yang, C.S., Nakashima, K. & Rana, T.M. *EMBO J.* **30**, 823–834 (2011).
10. Sengupta, S. *et al. Stem Cells* **27**, 1524–1528 (2009).

Human peptidome display

William H Robinson & Lawrence Steinman

A peptide library representing the entire human proteome is applied to the discovery of autoantigens.

Autoimmune diseases arise when the immune system misrecognizes the 'self' as foreign and attacks the body's own tissues. In most cases, the target molecules that trigger these self-destructive immune reactions are unknown, impeding the development of therapies and diagnostic tests. In this issue, Larman *et al.*[1] describe a powerful new approach to autoantigen discovery that is based on the first synthetic representation of a complete human proteome. The authors identify novel autoantigens in individuals with a neurological autoimmune disease and also show how their method can identify protein-protein interactions unrelated to antibodies. Knowledge of the specificity of autoantibodies is important as it may allow the development of antigen-specific treatments or provide mechanistic biomarkers to streamline clinical trial design.

A major obstacle to antigen discovery is the paucity of technologies that enable large-scale, unbiased surveys of the vast universe of targets that antibodies can recognize. These targets include small domains on proteins (including post-translational modifications with carbohydrates and lipids) as well as carbohydrates and lipids that are not associated with proteins. One of the least biased technologies for autoantigen discovery is mass spectrometry. In this approach, lysates are generated from the diseased tissue, immunoblotting or immunoprecipitation is carried out with the patient's serum as the source of antibodies, and mass spectrometry identifies the bound autoantigens. Autoantigens targeted in systemic

William H. Robinson is in the Department of Medicine, Division of Immunology and Rheumatology, Stanford University, Stanford, California, USA and the VA Palo Alto Health Care System, Palo Alto, California, USA, and Lawrence Steinman is in the Department of Neurology and Pediatrics, Stanford University, Stanford, California, USA.
e-mail: wrobins@stanford.edu or steinman@stanford.edu

© 2011 Nature America, Inc. All rights reserved.

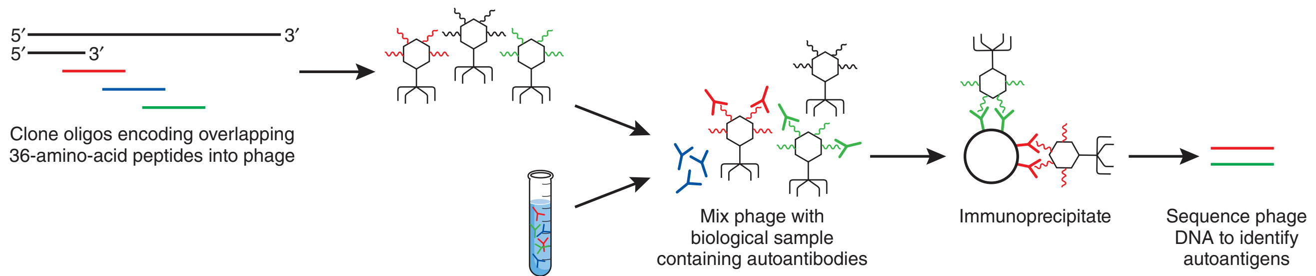

Figure 1 Phage immunoprecipitation sequencing for autoantigen discovery. Larman *et al.*[1] construct a T7-phage peptide library to express 413,611 DNA sequences that encode overlapping 36-amino-acid peptides spanning the 24,239 polypeptides encoded by all of the open reading frames in the human genome. The phage are then mixed with autoantibody-containing biological fluid from patients with an autoimmune disease. Phage expressing a 36-amino-acid peptide bound by a patient autoantibody are immunoprecipitated using beads coated with Protein A and Protein G. Finally, high-throughput sequencing of the human DNA carried by the immunoprecipitated phage identifies the 36-amino-acid peptide and, thus, the putative autoantigen.

lupus erythematosus and rheumatoid arthritis were discovered in this way[2]. Although mass spectrometric methods allow *de novo* discovery of a wide range of autoantigens, including those that are post-translationally modified, they are somewhat limited by biases in selection and preparation of tissue samples.

In array-based approaches, the reactivity of antibodies in patient samples is tested against panels of previously identified autoantigens arrayed on surfaces[3]. Antigen arrays also allow for screening of autoantibodies and other proteins for reactivity against polypeptides expressed using phage, bacterial or mammalian cell-based cDNA and peptide expression libraries[4,5]. Recently, this approach was extended to characterize reactivity against libraries of random synthetic peptoids, *N*-substituted glycines whose side chains are appended to the nitrogen atom of the peptide backbone, enabling identification of antigenic shapes[6]. Although powerful and less biased than antigen arrays, peptoid arrays require substantial follow-on experimentation to identify the actual autoantigens.

Another traditional approach to autoantigen identification relies on expression cloning with cDNA libraries. Polypeptides encoded by the cDNAs are screened for reactivity with patient-derived autoantibodies. Although several autoantigens have been identified using this technique[7,8], current methods are limited by low rates of expression of in-frame coding sequences and inefficiencies in analyzing binding interactions.

The new technology developed by Larman *et al.*[1], called phage immunoprecipitation sequencing (PhIP-Seq), combines phage display of a synthetic human peptidome with immunoprecipitation and high-throughput sequencing (**Fig. 1**). The authors begin by synthesizing >400,000 oligonucleotides encoding 36-amino-acid peptides that cover all open reading frames in the human genome. The oligos are then packaged in phage for display on the phage surface. Phage expressing candidate autoantigens are immunoprecipitated using antibodies present in patient samples and identified by high-throughput DNA sequencing.

Using this system, the authors analyze autoantibodies in the spinal fluid of three patients with paraneoplastic neurological autoimmune disease. Paraneoplastic syndromes are unique diseases at the interface of autoimmunity and tumor immunity. In these rare syndromes, the body produces autoantibodies that recognize molecules present both in a tumor and in normal brain tissue[7], leading to an autoimmune reaction. Larman *et al.*[1] identify previously described autoantigens and new candidate neurologic autoantigens in two patients who had clinical features of paraneoplastic disease but tested negative for antibodies against a panel of established paraneoplastic disease–associated antigens. Naturally, further testing and validation of the new putative autoantigens will be necessary to assess their role in pathogenesis and clinical utility. In principle, however, validated autoantigens could be used to design therapeutic blocking antibodies against domains in the antigens that are expressed in the tumor but not in normal tissue.

Detection of autoantibodies represents a central component of the diagnostic criteria for rheumatoid arthritis, systemic lupus erythematosus, myasthenia gravis and Graves' disease, and elucidation of the autoantigens in additional autoimmune diseases is expected to enable development of tests to detect autoantibodies with diagnostic utility. Although multiple programs are attempting to develop antigen-specific therapeutics, to date no therapy using antigen-specific tolerance is approved for the treatment of any human autoimmune disease.

In addition to identifying antigens, PhIP-Seq can also be used to characterize a variety of biomolecular interactions. The authors show this by identifying proteins that interact with protein replication factor 2. In another variation of the approach, it should be possible to screen for small molecules that modulate protein-protein interactions. Finally, by designing a different peptidome based on a microbial genome, one might adapt the approach to identify immunodominant epitopes present in microbes.

Peptidome phage display has several limitations. First, it cannot include post-translational modifications. Certain post-translational modifications are known to be critical targets of autoimmune responses, including protein citrullination in rheumatoid arthritis and protein cleavage in systemic lupus erythematosus[9]. Second, 36-amino-acid peptides expressed on the surface of phage cannot display conformational structures that exist only in native, full-length proteins. Certain tertiary and quaternary protein structures have been identified as the targets of autoantibodies that mediate autoimmune tissue injury and of neutralizing antimicrobial antibodies. For example, in the autoimmune neurologic condition known as stiff-person syndrome, the autoantibody response recognizes conformational tertiary epitopes on glutamic acid decarboxylase[10]. Peptide sequences expressed on the phage surface in nonnative conformations have the potential to bind antibodies in a nonphysiological manner. Finally, alternative splicing variants could act as autoantigens, and could be incorporated in the next generation of peptidome phage-display libraries.

Despite these limitations, PhIP-Seq is a comprehensive, powerful approach for autoantigen discovery as it combines a synthetic representation of the human proteome with high-throughput sequencing. We anticipate that it will be fruitful to apply this approach to investigate a spectrum of autoimmune diseases, including multiple sclerosis, psoriasis and Crohn's disease.

COMPETING FINANCIAL INTERESTS
The authors declare no competing financial interests.

1. Larman, H.B. *et al. Nat. Biotechnol.* **29**, 535–541 (2011).

© 2011 Nature America, Inc. All rights reserved. npg

2. Wu, T. & Mohan, C. *Autoimmun. Rev.* **8**, 595–598 (2009).
3. Robinson, W.H. *et al. Nat. Med.* **8**, 295–301 (2002).
4. Auger, I. *et al. Ann. Rheum. Dis.* **68**, 591–594 (2009).
5. Zhu, H. *et al. Science* **293**, 2101–2105 (2001).
6. Reddy, M.M. *et al. Cell* **144**, 132–142 (2011).
7. Albert, M.L. & Darnell, R.B. *Nat. Rev. Cancer* **4**, 36–44 (2004).
8. Wang, X. *et al. N. Engl. J. Med.* **353**, 1224–1235 (2005).
9. Doyle, H.A. & Mamula, M.J. *Ann. NY Acad. Sci.* **1050**, 1–9 (2005).
10. Kim, J. *et al. J. Exp. Med.* **180**, 595–606 (1994).

Molecular evolution picks up the PACE

Adam J Meyer & Andrew D Ellington

Protein variants with improved properties can be rapidly generated by a phage-based system that enables continuous directed evolution.

© 2011 Nature America, Inc. All rights reserved. npg

As a recent US Secretary of Defense memorably noted, one can distinguish between "known knowns," "known unknowns" and "unknown unknowns, the [things] we don't know we don't know." Like many scientific investigations, efforts to improve the properties of proteins by directed evolution are an excursion into the unknown. A recent paper in *Nature* by Esvelt *et al.*[1] opens the way for easier exploration of unknown sequence space—or at least the "known unknown" made up of stepwise, single-residue changes. Where current methods require weeks or months to complete a handful of rounds of optimization, the new approach, called phage-assisted continuous evolution (PACE), permits dozens of rounds of evolution in one day without the need for human intervention.

At present, directed evolution of a protein usually involves generating a library of heritable variants, selecting or screening for desirable variants, amplifying the survivors and repeating this cycle until the desired functionality is obtained[2]. In PACE, *Escherichia coli* host cells are passaged through a fixed-volume vessel containing a replicating population of phage encoding the protein of interest. Uninfected *E. coli* are continuously supplied to the system and removed; a constant rate of efflux means that phage are eliminated unless they propagate themselves in sufficient numbers by infecting new cells. Esvelt *et al.*[1] refer to this system as a 'lagoon', although the principle can also be explained by likening it to a conveyor belt that becomes progressively enriched with cells propagating phage that express an optimized protein or function (**Fig. 1**).

Adam J. Meyer and Andrew D. Ellington are at the Center for Systems and Synthetic Biology, University of Texas at Austin, Austin, Texas, USA. e-mail: andy.ellington@mail.utexas.edu

How is phage propagation linked to protein evolution? Initially, *E. coli* are infected with M13 phage that lack a gene essential for their replication, which encodes the pIII protein. The *E. coli* themselves contain the pIII gene, albeit in an inert form that requires activation. A clever feedback loop between host and pathogen makes activation of the pIII gene dependent upon the phage. For example, in the system described by Esvelt *et al.*[1], the phage carries a T7 RNA polymerase gene, and the endogenous pIII gene is under the control of the T7 RNA polymerase promoter. The infecting phage can create infectious offspring only if it carries an active T7 RNA polymerase variant. Mutations can be generated at the normal background rate. However, there is no substantial price to be paid for accelerating the process by using chemical or enzymatic mutagenesis, as the survival of the constantly replenished host cells is not an issue. Because uninfected *E. coli* are continuously fed into and removed from the system, only those phage that express active enzyme can efficiently infect new hosts. Esvelt *et al.*[1] used this system to select for polymerases with remarkably changed promoter specificities.

The two key advantages of PACE are that it promotes evolution in the context of cells and that it is continuous. Although the directed

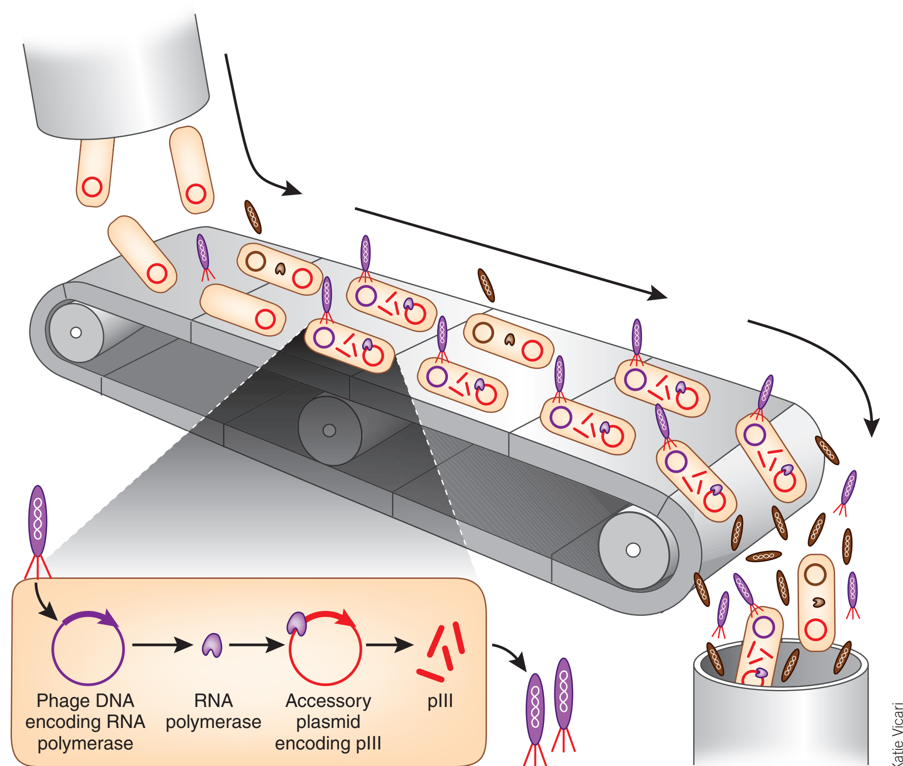

Figure 1 Phage-assisted continuous evolution. The 'lagoon' described by Esvelt *et al.*[1] is represented as a conveyor belt that allows a continuous influx and efflux of host cells. M13 phage require pIII protein (red bars) for their propagation. Ablation of pIII activity in M13 phage and expression of the gene encoding pIII from accessory plasmids (red circles) is used to couple pIII production, and thus phage fitness, with an activity of interest (in this case, the promoter specificity of an RNA polymerase). Host cells entering the system are infected by phage selected through their ability to express pIII from the accessory plasmid in the bacterial cell. Phage that produce fewer active polymerase mutants (brown) are progressively depleted from the lagoon owing to their reduced capacity for infection. Phage with beneficial mutations (purple) accumulate in the 'lagoon'. Not shown is a mutagenesis plasmid that can be used to increase the error rate during phage replication.