by raising CO_2 levels around the leaf[3] in a manner conceptually similar to adding a CCM. There is currently increased focus on the second strategy — if a CCM could be introduced into crops, it might turbocharge photosynthetic CO_2 fixation. CCMs have evolved independently in cyanobacteria, microalgae and some plants (mostly those regarded by us as weeds). Although several types of CCM are being considered for introduction into crops, Lin and colleagues' work focuses on the cyanobacterial CCM.

This CCM involves a series of membrane-based pumps for CO_2 and bicarbonate (HCO_3^-), and special microcompartments called carboxysomes, which contain Rubisco[4]. HCO_3^- is pumped into the cell, then converted to CO_2 in the carboxysomes by the enzyme carbonic anhydrase (Fig. 1). The resulting high local CO_2 concentrations increase Rubisco efficiency, and so almost eliminate O_2 fixation[4]. Furthermore, thanks to the CCM, cyanobacteria have retained an ancient form of Rubisco that is almost three times as efficient as that found in most crops[5].

Lin *et al.* engineered tobacco plants to express a functional cyanobacterial form of Rubisco. This enzyme usually consists of a complex of eight large subunits and five to eight small subunits. The authors replaced DNA that encodes the large subunit of Rubisco in the tobacco plant with that encoding the cyanobacterial enzyme, ensuring that the photosynthesis and growth they observed occurred as a result of the introduced Rubisco, rather than the native version. This DNA is located in the cells' photosynthesizing factories, structures called chloroplasts.

Lin and colleagues' approach differs from those of earlier, unsuccessful efforts[6] in several ways; most notably, the authors co-expressed the cyanobacterial Rubisco with proteins that are involved in the enzyme's assembly. They found that co-expression of cyanobacterial Rubisco with either the RbcX chaperone protein (which helps protein folding) or a carboxysomal protein called CcmM35 (a Rubisco-organizing protein) were equally effective at forming functional Rubisco. However, the latter approach produced large complexes of Rubisco, which seemed to be related to those that form during the assembly of pre-carboxysomes in cyanobacteria. This is because CcmM35 mimics three of Rubisco's small subunits and so is incorporated into Rubisco. But the protein also crosslinks to other Rubisco complexes, producing enzyme aggregates[4].

The authors did not demonstrate whether the addition of CcmM35 or RbcX was the pivotal step in successfully expressing cyanobacterial Rubisco in tobacco, or whether other elements of the experimental design provided the crucial advantage. Earlier this year, the same group showed that co-expression of several carboxysomal shell proteins in chloroplasts can produce structures suggestive of carboxysome self-assembly[7]. Thus, prospects for building functional carboxysomes in tobacco-plant chloroplasts are now quite good. However, extending this to crops would be greatly aided by the development of technologies for altering the chloroplast genomes of key crop species.

In the past two years, the sequence of events required to build a cyanobacterial CCM in the chloroplast has been identified in detail[8,9]. Stand-alone addition of cyanobacterial Rubisco, or even of carboxysomes, to chloroplasts provides no obvious advantage. In fact, Lin *et al.* show that their modified plants survive only at high CO_2 concentrations. To provide an advantage, both CO_2 and HCO_3^- pumps are required, to elevate HCO_3^- levels in the chloroplast and so turbocharge CO_2 levels in the carboxysomes. And even when these remaining steps have been achieved in model plants such as tobacco, improved crops are still some way off. However, this work is a milestone on the road to boosting plant efficiency. The advance can be likened to having a new engine block in place in a high-performance car engine — now we just need the turbocharger fitted and tuned. ■

G. Dean Price *and* **Susan M. Howitt** *are in the Research School of Biology, Australian National University, Canberra, ACT 0200, Australia.*
e-mails: dean.price@anu.edu.au; susan.howitt@anu.edu.au

1. www.un.org/News/Press/docs/2009/gaef3242.doc.htm
2. Lin, M. T., Occhialini, A., Andralojc, P. J., Parry, M. A. J. & Hanson, M. R. *Nature* **513,** 547–550 (2014).
3. Long, S. P., Zhu, X.-G., Naidu, S. L. & Ort, D. R. *Plant Cell Environ.* **29,** 315–330 (2006).
4. Rae, B. D., Long, B. M., Badger, M. R. & Price, G. D. *Microbiol. Mol. Biol. Rev.* **77,** 357–379 (2013).
5. Tcherkez, G. G. B., Farquhar, G. D. & Andrews, T. J. *Proc. Natl Acad. Sci. USA* **103,** 7246–7251 (2006).
6. Kanevski, I., Maliga, P., Rhoades, D. F. & Gutteridge, S. *Plant Physiol.* **119,** 133–141 (1999).
7. Lin, M. T. *et al. Plant J.* **79,** 1–12 (2014).
8. Price, G. D. *et al. J. Exp. Bot.* **64,** 753–768 (2013).
9. McGrath, J. M. & Long, S. P. *Plant Physiol.* **164,** 2247–2261 (2014).

This article was published online on 17 September 2014.

NETWORK BIOLOGY

A compass for stem-cell differentiation

The development of CellNet — network-biology software that determines how cell types generated *in vitro* relate to their naturally occurring counterparts — could improve our ability to produce desirable cells in culture.

FRANZ-JOSEF MÜLLER & JEANNE F. LORING

Over the past few years, websites such as Facebook and Google have attained an uncanny ability to understand us and to predict our behaviour, even before we have consciously decided what to do. This predictive power is achieved through the systematic application of statistical 'inference algorithms' to the vast numbers of connections and links that users establish when browsing the Internet — making up a 'social graph' that can be exploited to characterize distinct groups of Internet users. It would be wonderful to have such a graph to characterize distinct groups of cells. This could then be used in regenerative medicine to overcome the challenge of coercing stem cells to become the cell type needed for a particular therapy. Writing in *Cell*, Cahan *et al.*[1] and Morris *et al.*[2] describe a network-biology platform, CellNet, that takes a first step in this direction.

The most popular representation of the differentiation of cells from immature precursors to mature cell types was, for many years, the 'epigenetic landscape' diagram conceived by the biologist Conrad Hal Waddington[3,4]. This diagram evokes a set of one-way paths down which immature cells roll along defined routes to more-differentiated cellular states. But over the past decade, this simple model has morphed into the concept of a multidirectional cell-identity transfer hub.

In 2007, Yamanaka and colleagues[5] reprogrammed ordinary human skin cells called dermal fibroblasts into induced pluripotent stem (iPS) cells using transcription factors that are highly expressed in embryonic stem (ES) cells, an equivalent cell type that is derived from early embryos. Both iPS cells and ES cells are pluripotent — they can, given the correct molecular cues, differentiate into almost any cell in the body, forming any one of hundreds of different cell types. Each of these mature cell types is characterized by distinct networks of highly expressed transcription factors, which regulate the expression of large sets of genes. Researchers have used transcription-factor cocktails specific to cell types of interest to try to directly convert one cell type, such as a

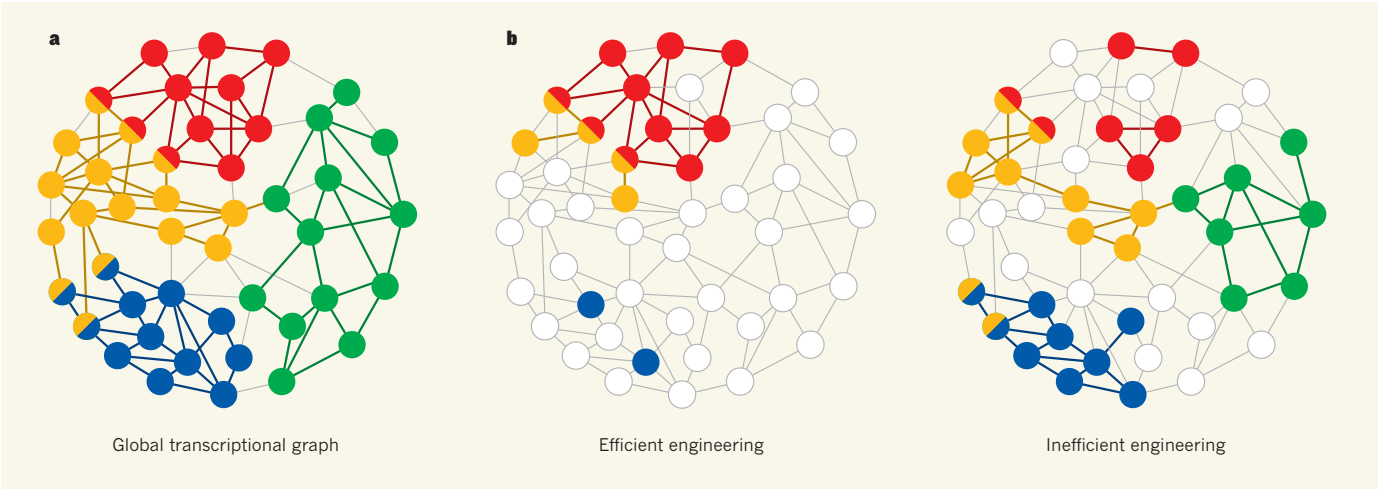

Figure 1 | Improving cellular engineering. Two papers[1,2] report the development of a network-biology platform called CellNet. **a**, A simplified schematic of the global transcriptional graph generated by CellNet, in which each circle represents a gene and each line links genes that are co-regulated. CellNet defines gene-regulatory networks — groups of highly interconnected and coordinately regulated genes — that are specific to distinct cell types. Gene-regulatory networks that are characteristic of four cell types are indicated by different colours. **b**, Two examples of gene-expression patterns in engineered cells. CellNet compares the gene-expression networks of these experimentally derived cells with the global transcriptional graph to determine whether the cultured cells resemble a naturally occurring cell type (as in the example on the left) or not (right). Colourless circles indicate genes that are not expressed in the cell type being analysed.

fibroblast, into another, such as a neuron[6] or a liver cell[7] (known as a hepatocyte). The abundance of reports suggests that there are almost no limits to the number of possible cellular transformations.

But are these engineered cells genuine copies of cells that exist in the body? There are well-established gene-expression tools for determining whether a stem cell is pluripotent[8,9]. But remarkably, given the flurry of research into directing stem cells to take on a particular identity, or fate, there is no commonly accepted way to determine whether a differentiating cell is moving towards the right developmental destination. Cahan *et al.* designed the CellNet software to give researchers an idea of how closely matched a cultured cell type is to its presumed counterpart in the body. The program applies a sequence of statistical inference algorithms to create a global transcriptional graph, which resembles the social graphs used by websites such as Facebook (Fig. 1a).

To construct this transcriptional graph, the authors used publicly available gene-expression data from tissues and cells, and information from genome-wide transcription-factor binding studies performed by the ENCODE consortium. The CellNet program identifies gene-regulatory networks (GRNs) for specific cell types in the body, such as neurons and hepatocytes. GRNs are groups of genes that are coordinately regulated in distinct cell types and that are more highly interconnected with one another than with other genes. The program then compares these cell-type-specific GRNs with those from experimentally derived cultured cells to determine how accurately the derived cells mimic the 'real' cell type (Fig. 1b). In addition, CellNet suggests transcription factors that could be modulated to shift an *in vitro* cell type closer to its *in vivo* correlate.

Cahan and colleagues compared two strategies for producing mature cell types: differentiation from pluripotent stem cells and direct conversion from another cell type. Testing engineered neurons and hepatocytes, CellNet analysis revealed differences between the two approaches. Cell types derived from pluripotent stem cells were similar to the naturally occurring cell types, but directly converted cells could be abnormal. For example, neurons directly converted from fibroblasts retained substantial fibroblastic identity and expressed GRNs that were characteristic of cells from the heart and pancreas.

How can CellNet improve the quality of engineered cells? An inkling of its future utility comes from Morris and colleagues' study of directly converted hepatocytes, called induced hepatocytes (iHeps). A CellNet comparison of iHeps and actual hepatocytes revealed that iHeps did express GRNs that were characteristic of hepatocytes, but they also activated illicit, developmentally immature transcriptional programs. Following up on this observation, Morris *et al.* found that the iHeps would be better described as induced endoderm progenitors (iEPs). These endodermal precursors can give rise to many cell types that arise from the endoderm (an embryonic cell layer), including cells of the colon, liver and pancreas. Indeed, when the authors used CellNet as a guide to modify the transcription-factor cocktail used for direct conversion, they generated iEPs that could differentiate into mature colon cells when transplanted into mice, and could repair damaged colons.

The major limitation of predictive programs such as CellNet is a shortage of data. In the modern world of 'big data', more data always lead to better predictions. Although Google can profile hundreds of millions of search users and follow their behaviour over several years, the CellNet team were limited to published data generated from a few thousand genome-wide analyses of gene expression. Because of the lack of large data sets from human cells and tissues, the current version of CellNet is practical only for experimental studies of mouse cells.

The lack of large, high-quality genome-wide transcriptional profiles for normal human cell types is a major bottleneck in the development of stem-cell-based therapies and drug screens. We need to learn how to robustly define and mechanistically understand the molecular coordinates for differentiated cell types before we can give human stem cells the directions they need to arrive at the right fate. As the baseball player Yogi Berra once said, "If you don't know where you're going, you'll wind up someplace else". ■

Franz-Josef Müller *is at the Zentrum für Integrative Psychiatrie Kiel, Universitätsklinikum Schleswig-Holstein, 24105 Kiel, Germany.* **Jeanne F. Loring** *is in the Department of Chemical Physiology, Center for Regenerative Medicine, The Scripps Research Institute, California 92037, USA.*
e-mails: fj.mueller@zip-kiel.de; jloring@scripps.edu

1. Cahan, P. *et al. Cell* **158,** 903–915 (2014).
2. Morris, S. A. *et al. Cell* **158,** 889–902 (2014).
3. Slack, J. M. *Nature Rev. Genet.* **3,** 889–895 (2002).
4. Waddington, C. H. *The Strategy of the Genes* (George Allen & Unwin, 1957).
5. Takahashi, K. *et al. Cell* **131,** 861–872 (2007).
6. Vierbuchen, T. *et al. Nature* **463,** 1035–1041 (2010).
7. Sekiya, S. & Suzuki, A. *Nature* **475,** 390–393 (2011).
8. Müller, F. J. *et al. Nature* **455**, 401–405 (2008).
9. Müller, F. J. *et al. Nature Methods* **8,** 315–317 (2011).