What is the point of large-scale collections of human induced pluripotent stem cells?

To the Editor:

Human induced pluripotent stem cells (hiPSCs) are the focus of intense research because of their potential to provide patient-specific cell therapies and to model human disease. Small numbers of control and disease-specific hiPSC lines are publicly available, but they rarely have full data sets that include genomic, epigenomic and detailed patient phenotype data (Table 1).

With the global thrust to generate and exploit hiPSCs, several initiatives are emerging that aim to generate collections of hundreds to thousands of cell lines and to address the associated scientific, technical and financial challenges (Table 2). In light of these efforts, we consider whether such large collections are worthwhile, highlight some of the potential problems associated with them and suggest some solutions.

Large collections are important in three broad areas: disease modeling, understanding how normal genetic variation affects cell behavior, and drug development. We discuss each of these in turn below.

Disease modeling with hiPSCs is predicated on the ability to differentiate the cells to appropriate lineages. In some instances, researchers can robustly differentiate hiPSCs in vitro to cells that closely resemble the fully functional cell types in vivo, such as retinal pigment epithelium or sensory neurons. Some types of fully differentiated cells, particularly cardiac myocytes and hepatocytes, are becoming more readily available from commercial companies (e.g., GE Life Sciences (Pittsburgh, PA, USA), Cellular Dynamics (Madison, WI, USA), Life Technologies (Carlsbad, CA, USA) and Cellectis (Paris)).

hiPSCs are of particular interest in the study of diseases for which access to human tissue is difficult (e.g., neuronal disorders), that may have a developmental component or that are inherited. More than 6,000 disorders are inherited, with many caused by single gene defects (Online Mendelian Inheritance in Man; http://www.omim.org/). Although geneticists are rapidly identifying the genes involved, understanding the biological mechanisms frequently requires extensive in vitro and in vivo studies. What appears to be the same disease can be caused by mutations in many different genes (e.g., retinitis pigmentosa). Alternatively, many different mutations can occur in the same gene, producing clinical consequences that vary across patients (e.g., cystic fibrosis). Having access to a compendium of good cell systems with well-defined mutations would be ideal for mechanistic studies.

Many laboratories are already creating hiPSCs from patients with rare genetic disorders. Even a small number of lines can be highly informative. Two lines were enough to illustrate some potential features of schizophrenia, and a few cell lines have been sufficient to make useful models to explore Alzheimer's disease. Even so, to understand the biology underlying any one disease, a larger number of hiPSC lines will be required.

Although some diseases will be difficult to model in cell culture, it is likely that cellular models can provide valuable insights in many instances. Do cells grow, divide and differentiate normally? Can they carry out normal metabolic functions? It is possible that simple assays, such as measuring the proportion of cells that die or divide in response to defined in vitro stimuli, will give important clues as to disease mechanism.

Table 1  Some existing sources of hiPSC lines

| Bank                                      | Location         | Ownership        | hiPSC lines banked and available | Other information/services                                      |
|-------------------------------------------|------------------|------------------|---------------------------------|----------------------------------------------------------------|
| Coriell Institute for Medical Research    | Camden, NJ, USA  | Nonprofit        | 47 lines                        | Banking of NIH-derived lines                                   |
| Cellartis (owned by Cellectis)            | Göteborg, Sweden | Private company  | 30                              |                                                                  |
| NIH Center for Regenerative Medicine      | Bethesda, MD, USA| NIH funded       | 15, plus other types of lines listed | Provides cells, protocols and services                        |
| Boston University Medical Campus          | Boston           | University owned | 21, plus iPSC mouse lines       |                                                                  |
| Center for Regenerative Medicine          |                  |                  |                                 |                                                                  |
| Harvard Stem Cell Institute (HSCI)        | Cambridge, MA, USA| University owned | 20 lines                        | Not a bank; sends out lines from HSCI laboratories             |
| WiCell Research Institute                 | Madison, WI, USA | University owned | 17 lines                        | Both embryonic stem cells and hiPSCs available                 |
| Rutgers University Cell and DNA Repository| Piscataway, NJ, USA| University owned | 10 lines                        | Partnered with NIMH                                             |
| American Type Cell Collection             | Manassas, VA, USA| Nonprofit        | 7 lines                         | Major distributor of cell lines                                |
| Massachusetts Stem Cell Bank              |                  | State government owned | 6 lines | Closed and hiPSC lines reverted to Harvard                       |
| RIKEN Bioresource Center                  | Tsukuba-shi, Japan| Government owned | 200 plus mouse iPSC lines       |                                                                  |

NIH, National Institutes of Health; NIMH, National Institute of Mental Health.
Greater disease insights should be gained from comparing lines from multiple patients exhibiting the same disorder driven by different gene defects.

A second area of interest is healthy controls. At first sight the case for making hiPSCs from many healthy individuals appears harder to make than the case for making large disease collections. The question ‘who is normal?’ is impossible to answer. In fact, we are all examples of the huge range of variation within the human genome—healthy at times but with myriad genetic variants that may predict disease at others. The only way to understand the heterogeneity within human biology is to look at lots of cells. By establishing a large enough bank of hiPSCs from normal individuals, it will be possible to acquire an in-depth understanding of the interindividual variability of specific cellular functions and provide a platform for genome-wide association genetics of genomic, proteomic and cellular traits. Data from 100 individuals would allow identification of common genetic variants that have strong effects, mainly with a cis-linkage to genomic traits, but data from 700 would allow identification of moderate effects and broader, trans-based effects. Furthermore, even in the case of well-characterized conditions resulting from the same mutation in the same gene, the disease can manifest itself to differing extents within a single family. Large collections of hiPSCs from normal individuals offer a means to make sense of data from ENCODE and other large-scale genomic efforts.

A final area of interest is the use of hiPSC lines in drug discovery and development. Three critical parts of the drug R&D process are drug screening, optimization for safety and patient stratification. Increasing numbers of hiPSC lines are needed at each stage. Once a hiPSC line has been produced that robustly recapitulates some features of a disorder, an obvious next step is to search for small molecules that reverse the phenotype. Differentiated hiPSCs much more closely recapitulate the human phenotype than many of the artificially engineered cell systems used previously. High-throughput screens have been carried out on differentiated embryonic stem cells, and, despite the additional time and cost, researchers are turning to hiPSCs to evaluate compounds and to validate new targets.

Although a large batch of a single, well-validated hiPSC line may suffice for initial drug screening, as the properties of a drug are optimized, additional cell lines are required. Two of the most common drug toxicities arise from either unwanted activity at cardiac ion channels or through substantive variation in liver metabolism leading to toxic metabolites or overdose. Panels of hiPSCs expressing a range of polymorphic channels can be differentiated into cardiac cells to predict whether new drugs are devoid of cardiotoxicity. Similarly, a panel of hiPSCs differentiated into hepatocytes that express a broad range of cytochrome P450 enzymes will be used to predict drug-induced liver injury. In both cases, tens of different cell lines will be required to cover the known major liabilities.

Interest continues to grow in patient stratification based on an understanding of which drug is best for each patient. Stratifying patients into subpopulations relies on phenotype or, increasingly, genotype. Rare pathogenic pain, for example, can arise from multiple genetic variants in the NaV1.7 channels that differentially affect the biophysics of sensory neurons, causing a variety of clinical symptoms with differing onset. Until NaV1.7 sequences from a large number of individuals were available (e.g., through the US National Institutes of Health 1,000 Genomes project), the extent to which some proteins are polymorphic was not appreciated. Furthermore, not all single-nucleotide polymorphisms (SNPs) have a physiological relevance.

Although some SNP variants have no effect on a gene product’s normal function, they can be highly relevant when considering the effects of a drug. Rare adverse responses to a drug can be derived from minor allelic variations in the way the human body handles the drug immunologically or metabolically. Minor variations in the enzymes responsible for metabolism and excretion can also substantially affect drug levels and therefore the therapeutic dose and maximal efficacy provided.
There are important classes of drugs, including analgesics, anticonvulsants and antidepressants, where not all patients benefit, and medicines are tried out sequentially or in combination. We now know that minor genetic variations in the drug target may also lead to interindividual variation in drug responses. A recent study has shown that an exploratory new drug differed by tenfold in affinity for its target, the P2X7 ion-channel, solely depending on two SNPs in the protein. Polymorphisms may be unrelated to known disease but determine which patients do and do not respond to a drug. For some drug targets, there are hundreds of variants. Having genetic sequences available that cover human diversity tells us the frequency of allelic variation in proteins. In vitro experiments are needed to know whether those variants affect drug responses. We are now in the realm of needing thousands of hiPSC lines.

Given that hiPSC lines have the potential to aid these important areas of research, what kinds of difficulties are associated with their use in large collections? With many laboratories across the world making hiPSC lines, there will inevitably be substantial heterogeneity in the cells produced. Sources of variation including different tissue sources of hiPSCs (e.g., hair, skin or blood), the donor’s age and state of health, and the conditions for making, selecting and maintaining the hiPSCs. Systematic understanding of the biological sources of such variation remains in its infancy. In such a fast-moving field, it will not be possible to standardize methodology in the near term, and a concerted effort will be required to assimilate best practices.

Rather than being too prescriptive, we should collect hiPSC lines with associated key information and learn what works and what doesn't from scientists using those lines. It is important to consolidate information on which lines prove most consistent and useful. Banks grow in value with the data deposited. Initially, some simple standard criteria should be applied to confirm that a cell is indeed an hiPSC, that it is free from mycoplasma or other contamination, and that its unique identity is verifiable, for example, by short tandem repeat fingerprinting. When using hiPSCs for experiments, three pieces of information should ideally be available: the clinical description of the patient, their genetic sequence and a differentiation protocol to produce the relevant cell type with all associated methodological data. Appropriate consent and donor anonymization are therefore critical.

To be effective and most useful, a bank should have the following attributes:
1. Fully-informed donor consent supporting the donation of tissue to generate hiPSCs together with genetic information and relevant medical history. The ethical considerations here are not insignificant.
2. A process to anonymize donors and maintain a robust database.
3. Where donated cells and associated information are to be used for research, we must recognize that the cell lines made are not restricted to one group of researchers but are made broadly available to all researchers who can contribute to the understanding of disease and its treatment, including those from academia, biotech and pharma.
4. Standardized protocols for storage, retrieval, culture and differentiation, where known.
5. A mechanism to collect knowledge on any phenotypic abnormalities arising after differentiation and characteristics unique to particular cell types.
6. A searchable electronic 'catalog' where cells can be requested based on specific gene sequence or medical background, and a quick, easy way of shipping cells to scientists globally.

A future can be envisaged in which thousands of hiPSC lines with some fundamental elements of quality control are broadly available. The challenge is substantial, not least in terms of ethical review, data management, cost and logistics. The only economically viable path forward is to generate such a bank (or network of banks) precompetitively and collaboratively. Generating, validating and expanding hiPSC lines is costly, with estimates of $10,000–20,000 per line. It is also time consuming, requiring 4–6 months from tissue harvest to robust characterization of the expanded line. Yet the costs are surely outweighed by the benefits, as ensuring that hiPSCs become standardized, readily accessible, high-quality reagents will enable scientists to optimize time spent in understanding human biology and disease and in generating new therapies.

ACKNOWLEDGMENTS
We acknowledge the valuable contribution of Dr. Devyn Smith.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details are available in the online version of the paper (doi:10.1038/nbt.2710).

Ruth McKernan1 & Fiona M Watt2
1Neurosis, Pfizer, Granta Park, Cambridge, UK. 2Centre for Stem Cells and Regenerative Medicine, King’s College London, London, UK. e-mail: fiona.watt@kcl.ac.uk

A proposal to use gamete cycling in vitro to improve crops and livestock

To the Editor:
The grand challenge of producing enough food, fiber and fuel for an ever-growing global population has benefited tremendously from genetic improvements in agriculturally important plants and animals over the past century. These genetic modifications have enabled billions more people to meet basic needs while using less arable land and providing good returns on research investment (Supplementary Fig. 1). Yet despite reductions in malnutrition-driven stunting and wasting, many humans remain undernourished. Satisfying these basic needs becomes more challenging with climate variability, constraints on productive farmland and limited availability of off-farm inputs (e.g., water, pesticides, phosphorus). Here, we outline the potential implications of an in vitro approach (thus far demonstrated

1. Carr, A.J. et al. PLoS ONE 4, e8152 (2009).
2. Chambers, S.M. et al. Nat. Biotechnol. 30, 715–720 (2012).
3. Ming, G.L. et al. J. Neurosci. 31, 16070–16075 (2011).
4. Pajac, S.P. et al. Nat. Med. 17, 1657–1662 (2011).
5. Israel, M.A. et al. Nature 482, 216–220 (2012).
6. Brennand K.J. et al. Nature 473, 221–225 (2011).
7. Shi, Y. et al. Sci. Transl. Med. 4, 124ra429 (2012).
8. Nica, A.C. & Dermitzakis, E.T. Philos. Trans. R. Soc. Lond. B Biol. Sci. 368, 20120362 (2013).
9. ENCODE Project Consortium. Nature 489, 57–74 (2012).
10. Grisovic, M. et al. Nat. Rev. Drug Discov. 10, 915–929 (2011).
11. McNeish, J. et al. J. Biol. Chem. 285, 17209–17217 (2010).
12. Xu, X. et al. Stem Cell Res. (Amst.) 10, 213–227 (2013).
13. Lee, G. et al. Nat. Biotechnol. 30, 1244–1248 (2012).
14. Liang, F. et al. Circulation 127, 1677–1691 (2013).
15. Xia, R. et al. Br. J. Clin. Pharmacol. 75, 885–896 (2013).
16. Dib-Hajj, S.D. et al. Nat. Rev. Neurosci. 14, 49–62 (2013).
17. Illing, P.T. et al. Curr. Opin. Immunol. 25, 81–89 (2013).
18. Jorgensen, A.L. et al. PLoS ONE 7, e44064 (2012).
19. McHugh, S.M. et al. Br. J. Clin. Pharmacol. 74, 376–380 (2012).
20. Corrected after print 8 October 2013
Corrigendum: What is the point of large-scale collections of human induced pluripotent stem cells?

Ruth McKernan & Fiona M Watt

Nat. Biotechnol. 31, 875–877 (2013); published online 8 October 2013; corrected after print 8 October 2012

In the version of this article initially published, in the Acknowledgments, the name Devyn Smith should have been Devyn Smith. An email address for the corresponding author has also been added. The errors have been corrected in the HTML and PDF versions of the article.