Published in final edited form as:
*Nat Biotechnol*. 2013 October 8; 31(10): . doi:10.1038/nbt.2710.

# What is the point of large-scale collections of human iPS cells?

**Ruth Mckernan**[1] and **Fiona M. Watt**[2]

[1]Ruth McKernan is CSO of Neusentis, Pfizer, Granta Park, Cambridge, UK

[2]Fiona M. Watt is Director of the Centre for Stem Cells and Regenerative Medicine, King's College London, UK

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, including 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.

## Why do we need large collections of hiPSCs?

There are three broad answers to this question: for disease modeling, to understand how normal genetic variation affects cell behaviour and for use in drug development.

### Disease modeling

Disease modelling 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[1] or sensory neurons[2]. Some types of fully differentiated cells, particularly cardiac myocytes and hepatocytes, are becoming more readily available from commercial companies (e.g. GE, Cellular Dynamics International, Life technologies and Cellectis).

HiPSCs are of particular interest in the study of diseases for which access to human tissue is difficult (e.g. neuronal disorders[3,4]), that may have a developmental component[4], or that are inherited[5]. More than 6,000 disorders are inherited, with many caused by single gene defects (Online Mendelian Inheritance in Man, 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[6], and a few cell lines have been sufficient to make useful models to explore Alzheimer's disease[3,7]. However, to understand the biology underlying any one disease, a larger number of hiPSC lines will be required.


Europe PMC Funders Author Manuscripts

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. Greater disease insights should be gained from comparing lines from multiple patients exhibiting the same disorder but driven by different gene defects.

## Healthy controls

At first sight the case for making hIPSc 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 hIPSc from normal individuals, it will be possible to acquire an in-depth understanding of the inter-individual 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[8]. 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[9].

## Drug discovery

hiPSC lines are important new tools at many stages in drug discovery and development. Three critical stages are drug screening, optimization for safety, and patient stratification. Increasing numbers of cell lines are needed at each stage. Once an 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[10]. High-throughput screens have been carried out on differentiated embryonic stem cells[11], and, despite the additional time and cost, researchers are turning to hiPSCs to evaluate compounds and to validate new targets[12,13].

Although a large batch of a single well-validated iPSC 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[14]. 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 [15]. In both cases, tens of different cell lines will be required to cover the known major liabilities.

A recent development in medicine is 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[16]. Until NaV1.7 sequences from a large number of individuals were available (e.g., through the NIH

1,000 genome project), the extent to which some proteins are polymorphic was not appreciated. Furthermore, not all SNPs have a physiological relevance.

Although some 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[17]. Minor variations in the enzymes responsible for metabolism and excretion can also significantly affect drug levels and therefore the therapeutic dose and maximal efficacy provided[18]. 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 inter-individual variation in drug responses. A recent study showed that an exploratory new drug differed by 10 fold in affinity for its target, the P2X7 ion-channel, solely depending on two polymorphisms in the protein[19], and a single polymorphism in the TNF-alpha 1 receptor can predict an adverse effect of TNF antagonist treatment[20]. 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 iPSC lines.

## Problems of large collections and potential solutions

With many labs across the world making hiPSC lines, there will inevitably be substantial heterogeneity in the cells produced. Sources of variation including different tissue sources (such as hair, skin or blood), the donor's age and state of health and the conditions for making, selecting and maintaining the hiPSCs. A systematic understanding of the biologic sources of such variation is 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 practice.

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 STR 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 iPSCs 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 ‘catalogue’ 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) pre-competitively and collaboratively. Generating, validating and expanding iPSC lines is costly, with estimates of $10-20,000 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.

## References

1. Carr AJ, et al. PLoS One. 2009; 4:e8152. [PubMed: 19997644]
2. Chambers SM, et al. Nat. Biotechnol. 2012; 30:715–720. [PubMed: 22750882]
3. Ming GL, et al. J Neurosci. 2011; 31:16070–16075. [PubMed: 22072658]
4. Paşca SP, et al. Nat Med. 2011; 17:1657–1662. [PubMed: 22120178]
5. Israel MA, et al. Nature. 2012; 482:216–220. [PubMed: 22278060]
6. Brennand KJ, et al. Nature. 2011; 473:221–225. [PubMed: 21490598]
7. Shi Y, et al. Sci Transl Med. 2012; 4:124ra29.
8. Nica AC, Dermitzakis ET. Philos Trans R Soc Lond B Biol Sci. **doi**: 10.1098/rstb.2012.0362(2013).
9. ENCODE Project Consortium. Nature. 2012; 489:57–74. [PubMed: 22955616]
10. Grskovic I, et al. Nature Reviews Drug Discovery. 2011; 10:915–929.
11. McNeish J, et al. J. Biol Chem. 2010; 285:17209–17. [PubMed: 20212047]
12. Xu X, et al. Stem Cell Res. 2013; 10:213–227. [PubMed: 23305945]
13. Lee G, et al. Nat Biotechnol. 2012; 30:1244–1248. [PubMed: 23159879]
14. Liang P, et al. Circulation. 2013; 127:1677–1691. [PubMed: 23519760]
15. Kia R, et al. Br. J. Clin. Pharm. 2013; 75:885–896.
16. Dib-Hajj SD, et al. Nat. Rev. Neurosci. 2012; 14:49–56. [PubMed: 23232607]
17. Illing PT, et al. Curr. Opin. Immunol. 2013; 25:81–89. [PubMed: 23141566]
18. Jorgensen AL, et al. PLoS One. 2012; 7:e44064. [PubMed: 22952875]
19. McHugh, et al. Br. J. Clin. Pharmacol. 2012; 74:376–380. [PubMed: 22295949]
20. Gregory AP, et al. Nature. 2012; 488:508–511. [PubMed: 22801493]

Europe PMC Funders Author Manuscripts

## Table 1

Some existing sources of hiPSC lines

| Bank | Location | Ownership | hiPS Lines Banked & Available | Other |
|---|---|---|---|---|
| Coriell Institute for Medical Research | New Jersey, USA | Non-profit | 47 lines | Banking of NIH derived lines |
| Cellartis (owned by Cellectis) | Sweden | Private company | 30 | |
| NIH Centre for Regenerative Medicine | Maryland, USA | NIH Funded | 15 plus other types of lines listed | Provides cells, protocols and services |
| Boston University Center for Reg. Med. | Massachusetts, USA | University owned | 21 plus iPS mouse lines | |
| Harvard Stem Cell Institute | Massachusetts, USA | University owned | 20 lines | Not a bank but sends out lines from HSCI Labs |
| WiCell | Wisconsin, USA | University Owned | 17 lines | Both ESCs and iPSC available |
| Rutgers University Cell and DNA Repository | New Jersey, USA | University owned | 10 lines | Partnered with NIMH |
| ATCC | Virginia, USA | Non-profit | 7 lines | Major distributor of cell lines |
| Massachusetts Stem Cell Bank | Massachusetts, USA | Government owned | 6 lines | Closed and iPSC lines reverted to Harvard |
| RIKEN Bioresource Center | Japan | Government owned | 5 plus mouse iPS lines | |

## Table 2

Large-scale iPS cell banks under development.

| Bank | Location | Ownership | hiPS Lines Banked & Available | Comment |
|---|---|---|---|---|
| IMI EU Centralized iPS Repository | EU | Public private partnership | Not started, but goal of 10,000+ lines combining cells from all IMI initiatives and other EU collections | Consortium to be announced in 2013 |
| New York Stem Cell Foundation | New York | Not for profit | Building a repository of 2,500 iPSC lines representing the diversity of US | Open access for global benefit |
| CIRM | California, USA | Government owned | $32M effort to collect and make store 9,000 cell lines from 3,000 people across 11 disease areas. | Most cells to be made by CDI and banked by the Coriell Institute for Medical Research at the Buck Institute in Novato |
| Kyoto U. Center for iPS Cell Research & Application | Japan | Government owned | Just starting but 75 lines by 2020 | Cells to be GMP rather than research grade. |
| Progenicyte | Florida, USA | Private company | Not started | In “near future” will bank |
| HiPScI | UK | Wellcome Trust, MRC | Initial phase is 700 normal, 100 disease | Open access data and cells |
| STEMBANCC (STEM cells for Biological Assays of Novel drugs and prediCtive toxicology; stembancc.org) | EU | Public Private Partnership | Lines from 500 individuals to study neurological disorders, diabetes drug safety | Will bank |
| National Institute for Mental Health (nimhstemcells.org) | USA | NIH funded | Lines to study a variety of mental Health disorders | Will bank and distribute |