



# Databases and collaboration require standards for human stem cell research

Nadia K. Litterman<sup>1</sup> and Sean Ekins<sup>1,2</sup>



<sup>1</sup> Collaborative Drug Discovery, 1633 Bayshore Highway, Suite 342, Burlingame, CA 94010, USA

<sup>2</sup> Collaborations in Chemistry, 5616 Hilltop Needmore Road, Fuquay Varina, NC 27526, USA

**Stem cell research is at an important juncture: despite significant potential for human health and several countries with key initiatives to expedite commercialization, there are gaps in capturing and exploiting the results of past and current research. Here, we propose a concerted plan that could be taken to foster a more collaborative approach and ensure that all research efforts can be leveraged across the community. The creation of a definitive centralized database repository, or at least harmonized data repositories, for stem cell groups in academia and industry, enabling secure selective sharing of data when needed, could provide the core structure that is sought globally and protect intellectual property. The development of minimum information about stem cell experiments (MIASCE) could be key to this development.**

## Introduction

It is widely accepted that human stem cell research has tremendous potential for the development of new therapies for disease treatment through regenerative medicine and disease modeling [1–4]. Stem cells harbor both the potential for self-renewal and ability to differentiate into any cell type of the body, allowing for the promise of cures to intractable diseases. Research initially focused on leveraging human embryonic stem (ES) cells and adult stem cells. In 2007, Shinya Yamanaka discovered a method for generating induced pluripotent stem cells (iPSC) [5]. The development of iPSCs allowed the derivation of stem cells from patients for the first time, and propelled the field of stem cell translational medicine forward in two major areas with implications for human health [1]. First, human stem cells could be directed to differentiate into any required cell type to replenish diseased tissue. Importantly, iPSCs can be generated from immune-matched healthy individuals and applied for regenerative purposes, a strategy that would avoid many of the serious problems associated with immune rejection of tissue or organ transplants [6–8]. Second, patient iPSCs can be differentiated into relevant cell types *in vitro* and, thus, serve as unique physiologically relevant models for mechanistic insight, toxicity screening, and drug discovery high-throughput screening (HTS) [1,9,10]. Continual research

advances make these revolutionary applications of stem cells appear increasingly likely to come to fruition.

Understanding of pluripotency factors that determine the capacity of a cell to retain stem-like qualities have led to the ability to generate and maintain an increasing number of human ES and iPSC lines. To date, more than 1500 different human stem cell lines have been created and are listed on the International Stem Cell Registry (<http://www.umassmed.edu/iscr/>). Given that iPSCs can be generated from anyone across the population, including patients with diseases or clinical disorders, those prone to certain adverse effects or toxicity, or healthy individuals, they offer the unparalleled opportunity to explore aspects of human health in early-phase research programs [1,4,10]. There have also been advances in differentiation technology, with success in the areas of cardiac, hepatic, pancreatic and neuronal cells [11–18]. This already has important industrial applications, with iPSC-derived cardiomyocytes and hepatocytes being used for toxicity screening [9,19–21]. One can imagine that, as this technology grows, it will be possible to evaluate toxicity and efficacy across patient populations before initiating clinical trials, thus avoiding idiosyncratic results [22]. Additionally, small-molecule HTS on differentiated iPSCs has the unique advantage of initiating a drug discovery project on the disease-relevant, human cell type, thereby avoiding artifacts typically associated with common artificial overexpression systems. Although animal models are important for drug

Corresponding author: Ekins, S. ([ekinssean@yahoo.com](mailto:ekinssean@yahoo.com))

development, there are crucial biological differences between, for example, mice and humans, and iPSCs might be able to fill this gap [1,23,24]. The use of iPSCs to predict differences in efficacy across a patient population before clinical trials has been suggested as another way to capitalize on the potential of iPSCs as the experimental arm of personalized medicine [1,25]. Given that the major causes of drug development failure are lack of efficacy and unforeseen toxicity, which in turn impact cost [26,27], the use of stem cells in drug development will be important for reducing these risks and associated costs.

The additional use of iPSCs for modeling genetically defined diseases has led to breakthroughs in our understanding of disease mechanisms and has important implications for therapeutic development. In the first example, iPSCs from a patient with spinal muscular atrophy (SMA) were differentiated into motor neurons and displayed disease-specific defects [28]. More recently, electrophysiological changes in motor neurons from patients with amyotrophic lateral sclerosis (ALS) [29], aberrant protein aggregation in dopaminergic neurons from patients with Parkinson's disease [30], and prolongation of action potential duration in cardiomyocytes from patients with long QT syndrome [31], have been observed. iPSCs are well suited to shed light on disorders caused by a single or few genetic mutations, and have become invaluable tools for numerous rare diseases, which often have limited funding resources. Additionally, new gene-editing technologies are enabling researchers to make hypothesis-driven modifications to iPSCs to allow for investigation of cellular phenotypes in more complex genetic disorders [32].

### Impediments to progress in stem cell research

There have been significant advances in the use of iPSCs, such as the first clinical trial with iPSCs in the world, which was run by RIKEN (<http://www.riken.jp>) and focused on macular degeneration [33,34]. Still, considerable hurdles to the realization of the full potential of stem cell-derived therapies remain (Table 1). For example, there is large variability in differentiation potential, efficiency, and phenotypic output between iPSC lines, which has been attributed to incomplete reprogramming and genetic

or epigenetic factors [1,35–39]. Additionally, even the most advanced differentiation protocols and strategies remain unstandardized, complex, and expensive. Furthermore, differentiation is never complete or synchronized, and the resulting heterogeneous cell population is difficult to define. In practical terms, this means that reproducibility and the ability to scale up production are important challenges that must be faced for translation of stem cell research to the clinic, either in the form of implantation or HTS [23,40–42]. For clinical application of regenerative therapy, iPSCs must be able to be robustly tracked, similar to blood, tissue, and organ transplantation, including information on donor features and consent, selection, immune compatibility, and good manufacturing practice (GMP) [6]. Furthermore, genetic and epigenetic changes to iPSCs could alter their efficacy and tumorigenicity, which is an issue that requires further investigation and will require diligent cataloguing [43]. In addition, there are ethical and cultural hurdles that need to be overcome if stem cell research is to reach its full potential (Table 1). The complexity of these problems suggests that their solution will require the concerted effort of the community of stem cell researchers, rather than individual piecemeal solutions. As seen with other technologies relevant to healthcare and drug discovery, bottlenecks that can hamper progress can be caused by gaps in collaboration or software, which can be readily addressed [44,45].

### The need for a centralized stem cell database

Leaders in the stem cell community have called for initiatives for improving the ability of scientists to collaborate to work toward stem cell research goals and for the benefit of patients. For example, a workshop held at the 2010 International Society for Stem Cell Research (ISSCR) conference noted that the lack of a comprehensive database, nomenclature, or minimum standards has led to confusion and inefficiencies in traceability, interpretation and repetition of results [46]. Recently, Marc Turner and colleagues suggested creating a network of stem cell banks for tracking key features such as donor consent, selection, immune compatibility and GMP [6]. The European Bank for Induced Pluripotent Stem Cells (EBiSC), was recently established as a central storage and distribution facility for human iPSCs, which could help share these resources. However, there are few coordinated sources of information, which would also be helpful. Shinya Yamanaka also announced the establishment of a blood-derived iPSC bank in Japan, which would require a database [47]. One valuable aspect of such a network of stem cell banks would be the ability to match for immune compatibility across the globe. For a relatively homogenous population, it has been estimated the screening 15,000 individuals could lead to 30 cell lines with a likelihood of matching 82.2% of the population [48]. For more genetically diverse populations, more individuals would need to be screened to cover a higher percentage of the population [49]. Taking these concerns into consideration, a comprehensive, mineable, freely available database, or decentralized harmonized data repositories, to store and share data associated with the derivation, differentiation, and storage of stem cells, is necessary for the growth of stem cell research and its application to human disease and associated research.

Researchers have previously proposed the need for interoperable bioscience data and used the Stem Cell Discovery Engine of the

TABLE 1

#### Hurdles to overcome for stem cell research to realize its full potential

Hurdle	Further details
<b>Scientific</b>	Lack of standardization and understanding of differences in methods for reprogramming of somatic cells and differentiation of stem cells; unknowns surrounding epigenetic memory, genetic background, and cell culture artifacts; heterogeneity of cell types in differentiated cultures and lack of ways to analyze these cell types, such as single-cell sequencing
<b>Ethical</b>	Patient privacy issues and how to apply IRB protocols properly and their variable conditions on information sharing; risks of de-identification in public databases; types of consent required for using major histocompatibility-matched lines for regenerative purposes in other individuals
<b>Cultural</b>	Delineation of proper protections for intellectual property and publications. Safeguards installed should incentivize collaboration; researchers engaged in collaboration should receive benefits in terms of scientific advancement

TABLE 2

**A selection of databases and data sources containing stem cell research information**

Name	URL	Description and references
NIH Human ES Cell Registry	<a href="http://grants.nih.gov/stem_cells/registry/current.htm">http://grants.nih.gov/stem_cells/registry/current.htm</a>	NIH registry where many ES lines are housed, with information on who to contact and the status regarding eligibility
Stem Cell Database (SCDb)	<a href="http://stemcell.mssm.edu/v2/">http://stemcell.mssm.edu/v2/</a>	DNA/RNA sequencing and microarray database for hematopoietic stem cells at multiple stages during differentiation [73]
Stemdb	<a href="https://www.stemdb.org/stemdb/">https://www.stemdb.org/stemdb/</a>	Database for gene expression, antibody, cell line data with a subset publicly available; funded by EuroStemCell and Eurosystem
Stem Cell Lineage database	<a href="http://sclcd.mcb.uconn.edu/SCLCD/">http://sclcd.mcb.uconn.edu/SCLCD/</a>	Lineage maps for endogenous and directed differentiation for multiple cell types (mouse and human), focusing mainly on immunology and hematopoietic lineages
Stem Cell-Omics Repository	<a href="http://scor.chem.wisc.edu/">http://scor.chem.wisc.edu/</a>	Stem cell-omics library: proteomic and microarray data with a limited number of data sets
International Stem Cell Registry at UMass Medical	<a href="http://www.umassmed.edu/isocr/index.aspx">http://www.umassmed.edu/isocr/index.aspx</a>	Largest database with basic information on ES cell and iPSC lines
StemBook	<a href="http://www.stembook.org/about">http://www.stembook.org/about</a>	An open-access repository for protocols and searching stem cell reports in other journals
Embryonic Stem Cell Atlas of Pluripotency Evidence (ESCAPE)	<a href="http://www.maayanlab.net/ESCAPE/">http://www.maayanlab.net/ESCAPE/</a>	Mouse and Human ES cell analysis using genome-wide technologies, including RNA-seq, Chip-seq, and RNAi screens
Stem Cell Matrix	Jeanne Loring (not web accessible)	Genomic and epigenomic data for more than 5000 samples and quality control for clinical applications
HSC-Explorer	<a href="http://mips.helmholtz-muenchen.de/HSC/">http://mips.helmholtz-muenchen.de/HSC/</a>	Publicly available, integrative database [74] containing detailed information about the early steps of hematopoiesis; contains more than 7000 experimentally validated interactions between molecules, bioprocesses, and environmental factors
Stem Cell Discovery Engine (SCDE)	<a href="http://discovery.hsci.harvard.edu">http://discovery.hsci.harvard.edu</a>	A database of curated cancer stem cell experiments coupled to the Galaxy analytical framework [51]; it allows users to describe, share, and compare data at the gene and pathway level; focused on tissue and cancer stem cell-related experiments from blood, intestine, and brain
CELLPEDIA	<a href="http://cellpedia.cbrc.jp/">http://cellpedia.cbrc.jp/</a>	A repository for current knowledge about human cells [75]. It contains various types of information, such as cell morphologies, gene expression and literature references
Functional Genomics in Embryonic Stem Cells' consortium (FunGenES)	<a href="http://biit.cs.ut.ee/fungenes/">http://biit.cs.ut.ee/fungenes/</a>	A consortium analyzed the transcriptome of mouse ES cells in eleven diverse settings representing sixty-seven experimental conditions [76]. The results are organized in an interactive database with a number of features and tools
ToxCast	<a href="http://www.epa.gov/ncct/toxcast/">http://www.epa.gov/ncct/toxcast/</a>	High-throughput screening (HTS) of chemical libraries across more than 650 <i>in vitro</i> assays including biochemical assays, human cells and cell lines, and alternative models such as mouse embryonic stem cells and zebrafish embryo development [77]
LifeMap Discovery	<a href="http://discovery.lifemapsc.com/">http://discovery.lifemapsc.com/</a>	Embryonic Development and Stem Cell compendium has collected data from the literature and high-throughput data sources surrounding stem cells and regenerative medicine [78]

Harvard Stem Cell Institute to illustrate the investigation, study, assay (ISA) framework as the backbone for discovery, exchange, and integration [50,51]. This has not been broadly applied to date and neither has there been the development of a definitive database of stem cell information. Several niche databases exist (Table 2), and have been established to serve as tools for the community, but represent data silos because they are not integrated. For example, the National Institutes of Health (NIH) has compiled a national registry with information on ES cell lines that are eligible for NIH-funded research. Academic groups have created searchable databases on a lab-by-lab basis, offering genome-wide data for stem cell-derived samples (Table 2). These would serve as useful building blocks for integration once a

complete stem cell database is established. The existence of these resources for data sharing suggest that the stem cell research community is aware of the need and willing to participate, but needs some organization. A fully functional database, or multiple linked databases adhering to the same standards, that would serve the needs of all stem cell researchers would require significant sponsorship, worldwide support, active participation, and continual maintenance. We suggest that there needs to be active engagement with the California Institute of Regenerative Medicine, the New York Stem Cell Foundation, the European Commission, the Japan Science and Technology Agency, and other potential global groups that could ensure standardization and provide financial support.

## The need for increased collaboration

To understand the value, feasibility, and impact of collaboration and a centralized database for secure selective sharing, it is perhaps useful to learn from a different field. Although the biological challenges are different, neglected disease research often suffers from a similar lack of coordination, collaboration, and data sharing as stem cell research [45,52]. In particular, TB research mainly comprised disjointed, individual efforts from academic and non-profit laboratories distributed across the globe. However, program funding from the Bill and Melinda Gates Foundation and The European Commission have enabled the TB Drug Accelerator and the More Medicines For Tuberculosis, respectively as large-scale (over 20 and seven groups, respectively) collaborations between academia, research institutes, and industry. These collaborations have focused TB research and promoted the selective sharing of related data in a secure environment among collaborators. The main platform for sharing and mining TB data in this community has been the Collaborative Drug Discovery (CDD; <https://www.collaborativedrug.com>) Vault [45], which is one of several cloud-based technologies developed for researchers to collaborate and share their data securely (Table 3). Publicly available data on compounds screened for activity against *Mycobacterium tuberculosis* are also provided to the community as part of this database, allowing researchers to leverage the existing literature to capture and screen already validated compounds. Given that private and public data are housed together in one place, researchers are able to compare data sets to identify important properties. Notably, because chemical structure information in the large-scale selective collaborations can be kept private, it remains patentable and, thus, concerns about intellectual property are not an issue.

The real value of this type of centralized database even goes beyond helping researchers find and access resources; it is prioritizing the next compounds for testing. Prior knowledge captured in the database can be effectively leveraged through careful curation of data and the building of validated computational models. The implementation of cheminformatics for TB drug discovery has led to significant savings in time and effort, limiting the number of compounds that need to be tested to find novel inhibitors [53,54]. In fact, recent cheminformatics models achieved hit rates of greater than 20%, whereas typical HTS rates range from 0.1% to 1%; such models had the added benefit of focusing only on molecules with low or no cytotoxicity in mammals [55,56]. Thus, organization of data in a central database can be used for machine-learning models that can spur discoveries and prioritize research goals for researchers using stem cells for drug discovery.

## Spurring collaboration with computational models

We propose that the field of stem cell research could follow the lead of TB research. It is not too difficult to imagine a single repository that could provide access to detailed information on cells, their associated preparation protocols, and HTS data, that could then be merged with published studies as well as private data (Fig. 1). Fostering collaboration and sharing data would directly address issues of reproducibility, which is a pervasive problem in this field as well as others [57]. Instead of haphazard, unintended duplication of experiments that cannot be directly compared, open sharing and centralization of results would enable researchers to coordinate their efforts to increase confidence in their methodologies. For commercialization, this will provide the necessary information for enticing investment and follow-up investigation of protocols for scaling up production. In this case, secure, selective data sharing could be used to avoid public disclosure and protect claims of intellectual property.

Capturing which laboratories work with stem cells harboring specific disease mutations will be useful to identify expertise as well as to compare cellular phenotypes across a broad array of genotypes. For example, for understanding complex disease pathology, such as in Parkinson's disease, it might be useful to compare iPSC-derived dopaminergic neurons from patients with Parkinson's disease to those of patients with ALS or other neurodegenerative diseases. If such studies became common, it might also help untangle disease-specific phenotypes and lead to better disease characterization.

Importantly, a centralized database would enable advanced data analytics and computational models to be applied to stem cell research. Given that the application of computational modeling to stem cell research is not yet commonplace, it will probably be useful to start with a limited question that can be used as a proof-of-concept to demonstrate the value of the approach to the field at large. Topics that might lend themselves well to computational modeling include optimization of differentiation protocols or the analysis of high-content image-based screens, which might help to address the issues associated with culture heterogeneity. Furthermore, because the number of compounds tested in human stem cell-based assays is typically small (a few thousand compounds) [58–62] compared with target-based biochemical screens or phenotypic screens (often more than 100,000 compounds), the application of a cheminformatics modeling approach might be especially important for using active and inactive compounds to help prioritize future molecules to test. Rather than screen entire libraries, it is more feasible to test compounds iteratively based on machine learning and other ligand-based computational

TABLE 3

### Commercial software enabling collaborations, storing cells and data

Software	Vendor	Information or website
Accelrys Biological Registration	Accelrys	<a href="http://accelrys.com/products/datasheets/biological-registration.pdf">http://accelrys.com/products/datasheets/biological-registration.pdf</a>
Science Cloud	Biovia (formerly Accelrys)	<a href="https://www.sciencecloud.com/">https://www.sciencecloud.com/</a>
CDD Vault	Collaborative Drug Discovery, Inc.	<a href="https://www.collaborativedrug.com/">https://www.collaborativedrug.com/</a>
Bio Register	Dotmatics	<a href="http://www.dotmatics.com/products/bio-register/">http://www.dotmatics.com/products/bio-register/</a>
Gateway	Dotmatics	<a href="http://www.dotmatics.com/products/gateway/">http://www.dotmatics.com/products/gateway/</a>

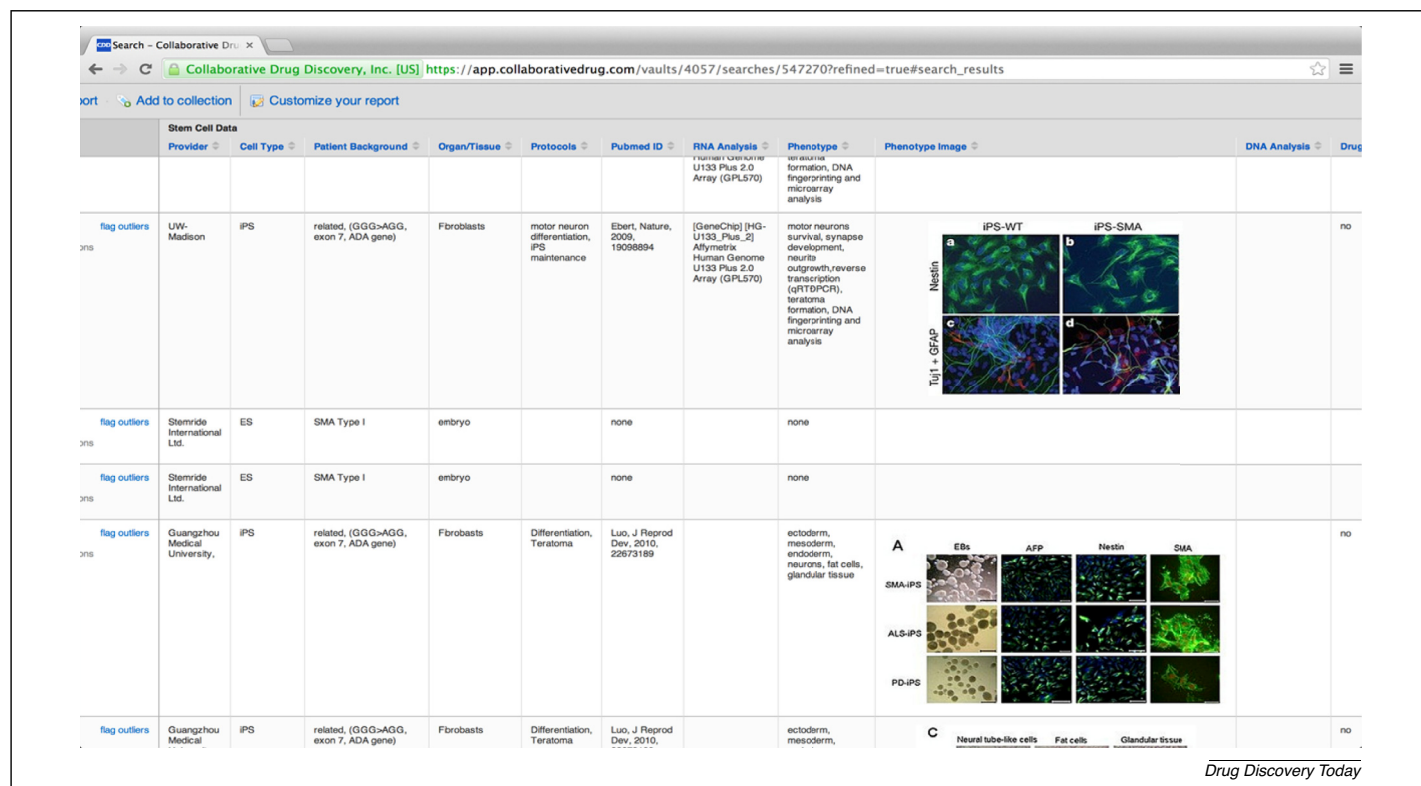


FIGURE 1

An example of human stem cell data stored in the Collaborative Drug Discovery Vault to illustrate how a comprehensive database could be created and used for collaborations. A potential search for 'SMA' might return these results [28,72], including all embryonic stem (ES) and induced pluripotent (iPS) cell lines generated, the patient background, protocols used, and data available.

modeling approaches. Also, by sharing models based on multiple assays, it would be possible to optimize compounds with desired features, such as promoting axon growth with no effect in other non-neuronal cell types. For example, data generated with iPSCs treated with various drugs to address cardiotoxicity or hepatotoxicity could be collated for building predictive machine-learning models, which could then be used to select compounds for further testing *in vitro*. These models could also be shared if generated with open-source algorithms and molecular descriptors [63]. Additionally, bioinformatics comparison of the effects of the same compounds on cells derived from different iPSCs could reveal genotype-specific differences in response and toxicity that would not otherwise be revealed. In this way, HTS data might become directly relevant to clinical practice for different patient populations, and could provide an important aspect of moving the drug discovery process toward personalized medicine. A central database might also result in better understanding of immune compatibility and enhanced methods for matching patient to donor iPSCs. Of course, this would need to comply with institutional review board (IRB) protocols (Table 1). Although computational models of stem cell data are not meant to replace *in vitro* experimental results, they would increase efficiency and the chance of success, given limited resources. We would propose using some of the examples and data sets described earlier as proof of concept. Once it has been demonstrated that there is a scientific advantage to selective data sharing for collaboration and computational approaches, it is likely to gain traction in the stem cell community.

### Minimum information about stem cell experiments

If we are to move toward more collaboration in stem cell research, then standards should be at the forefront and this has been proposed elsewhere [46,64–66]. This is important to ensure that other scientists can reproduce one another's efforts and will enable informed comparison of data, perhaps through automated data mining. This would also build confidence in the field as it moves forward toward translational goals. Minimum information standards are available in bioscience, although it appears that there is none specifically for stem cell research ([http://en.wikipedia.org/wiki/Minimum\\_Information\\_Standards](http://en.wikipedia.org/wiki/Minimum_Information_Standards)). If MIASCE are to be created, this will require consultation with the research community. Examples of the types of data that could be included in the MIASCE include patient background, drug screening, differentiation protocols, genomic, proteomic and phenotypic results. It will also be useful to share specific reagent details and sources to enhance reproducibility. MIASCE would be invaluable for creating the structure of a definitive stem cell database. However, developing this will require investment and the concerted effort of journals and funding agencies to ensure that data are reported according to this standard. Again, this will require coordination of the research community and all sides agreeing on what is best for the field to move ahead.

Once MIASCE are determined, technological tools can be developed to make it simple for scientists to adhere to such standards. For example, the BioAssay Ontology (BAO) is a standardized framework for describing HTS assays, and allows for in-depth

comparison, organization, and aggregation of protocol methods [67]. One major issue that prevents the wide adoption of the BAO is that it requires substantial time and ontology domain expertise to annotate each assay correctly. To address this issue, efforts are currently underway to develop hybrid technology that coordinates automatic and manual demarcation of annotations, which could improve the current situation [68]. Applying technologies that reduce the effort required for researchers to associate their experiments easily with the MIASCE would be useful for its adoption.

### Stem cell research at a crossroads

Currently, stem cell research is at a crucial juncture with the recent introduction of the Regenerative Medicine Promotion Act of 2014 in the US Senate, which seeks a national strategy and coordination on the topic [69]. Although other countries such as Japan are investing heavily in stem cell research, it will be important that the rest of the world is not left behind, and this will require considerable financial commitment. Despite its clear therapeutic promise, stem cell technology remains a costly and risky investment, with

few therapies having reached US Food and Drug Administration (FDA), European Medicines Agency (EMA) or other regulatory agency approval. Several major pharmaceutical companies are investing peripherally in stem cell research, including Genentech, Biogen, Pfizer, GlaxoSmithKline and others, often through external collaboration with academic groups [70]. This essentially represents waiting on the sidelines, expecting other smaller companies to shoulder the burden of risk in developing stem cell therapeutics. There are many smaller companies in the sector that have very early-stage pipelines, which vary widely in therapeutic areas and approaches (Table 4). Cellular Dynamics International (<http://www.cellulardynamics.com/>) is a successful model of a company that is providing iPSC-derived differentiated cells to customers under standardized optimized conditions. iPierian (now owned by Bristol-Myers-Squibb), founded on the premise of screening on iPSC models, has since changed strategies to focus on a monoclonal antibody program. Today's stem cell researcher relies on ordering kits and premade formulations for experiments, instead of making buffers from individual components. Although

TABLE 4

#### Examples of companies with stem cell therapies or targeting treatments in their pipelines<sup>a</sup>

Company	Pipeline	Details	Website
<b>Stem Cell Therapeutics</b>	Tigecycline in phase I trials for acute myeloid leukemia (AML): it kills AML tumor cells and leukemic stem cells	Anticancer approach targeting stem cells	<a href="http://www.stemcellthera.com/Technology.aspx?section=about">http://www.stemcellthera.com/Technology.aspx?section=about</a>
<b>California Stem Cell</b>	Supports cancer clinical development programs and optimizes and streamlines the clinical and commercial manufacturing process	Production and supply of clinically relevant human cell populations, and their therapeutic application to human disease and injury	<a href="http://www.californiastemcell.com/pipeline-programs/">http://www.californiastemcell.com/pipeline-programs/</a>
<b>Brain Storm Cell Therapeutics</b>	In phase II trials for ALS	NurOwn is an autologous, adult stem cell therapy technology in which autologous mesenchymal stem cells secreting neurotrophic factors (MSC-NTF) are administered via combined intramuscular and intrathecal injections to differentiate bone marrow-derived MSC into MSC-NTF cells	<a href="http://www.brainstorm-cell.com/">http://www.brainstorm-cell.com/</a>
<b>Cytomedix</b>	Phase 2 trial of ALD-401 in stroke underway; ALD-301 completed phase 1/2 study in critical limb ischemia; ALD-201 completed a phase I trial for ischemic heart failure	Develops therapies derived from, and delivered to, the patient	<a href="http://www.cytomedix.com/pipeline/">http://www.cytomedix.com/pipeline/</a>
<b>Advanced Cell Technologies</b>	Two phase I clinical trials underway using human ES cell retinal pigment epithelium for AMD and Stargardt's macular dystrophy	The company's intellectual property portfolio includes PSC platforms, both human and induced, and other cell therapy research programs	<a href="http://advancedcell.com/our-technology/act-stem-cell-related-research-pipeline/">http://advancedcell.com/our-technology/act-stem-cell-related-research-pipeline/</a>
<b>Fate Therapeutics</b>	PROHEMA <sup>®</sup> , is a pharmacologically modulated hematopoietic stem cell (HSC) therapeutic derived from umbilical cord blood that is currently in phase II clinical trials	HSC modulation platform focuses on <i>ex vivo</i> pharmacologic optimization of HSCs	<a href="http://fatetherapeutics.com/therapeutic-programs/">http://fatetherapeutics.com/therapeutic-programs/</a>
<b>Celgene</b>	PDA-001 for Crohn's disease and PDA-002 for diabetic foot ulcers	Placenta-derived adherent cells (PDAC <sup>®</sup> cells)	<a href="https://www.celgene.com/content/uploads/2014/03/2014.3-Q1-Celgene-IR-Combined-Pipeline-Otezla-Approval-FINAL.pdf">https://www.celgene.com/content/uploads/2014/03/2014.3-Q1-Celgene-IR-Combined-Pipeline-Otezla-Approval-FINAL.pdf</a>

<sup>a</sup> Additional companies can be found at <http://www.stem-cell-companies.com/>.

many of the commercially available kits work well, they contain proprietary formulations or supplements. These and other suppliers have considerable power to shape the field, setting their own standards. For commercialization, it is clear that working together might be a way for companies and academics to address precompetitive issues, share the risk, and enable the high-value products to be derived more quickly. It is likely that, as with other technologies, such as RNA interference (RNAi) therapeutics [44], stem cells could benefit from the development of informatics tools that combine existing databases and data sources (Table 2) and complement these with best practices for collaboration [71].

### The way forward

The preceding discussion sets the stage for proposing a way forward. Taking multiple strategies in parallel will be necessary to promote successful collaboration in the field. First, we need to engage all the stakeholders and ensure that funding organizations are supportive of this proposal. In addition, to capitalize on the vast untapped potential, stem cell researchers must collaborate to overcome the common problems faced. This will include the creation of stem cell standards, such as MIASCE. Preferably a centralized stem cell database, or decentralized harmonized data repositories, could be created for storing information on patient

background, drug screening, differentiation protocols, genomic, proteomic and phenotypic results. Finally, applying computational approaches for specific projects to demonstrate the value of sharing and learning from shared data will be needed to convince researchers of the value of the collaborative approach. Once these steps are taken, we might be able to provide a stronger foundation for the field. At the very least, it will be important that, as politicians and scientific administrators develop a long-term strategy for stem cell research, they consider the central importance of creating such standards, a database, and a collaborative environment.

### Conflicts of interest

N.L. is an employee of, and S.E. is a consultant for, CDD Inc. S.E. owns stock in Advanced Cell Technologies (ACTC).

### Acknowledgements

The CDD TB database has been developed thanks to funding from the Bill and Melinda Gates Foundation (Grant # 49852 ‘Collaborative drug discovery for TB through a novel database of SAR data optimized to promote data archiving and sharing’), The European Commission (More Medicines for TB), and the NIH (Award Numbers R43 LM011152-01, 2R42AI088893-02). We kindly acknowledge Dr. Myles Axton for preliminary discussion.

### References

- Inoue, H. *et al.* (2014) iPS cells: a game changer for future medicine. *EMBO J.* 33, 409–417
- Yamanaka, S. (2012) Induced pluripotent stem cells: past, present, and future. *Cell Stem Cell* 10, 678–684
- Surani, M.A. (2012) Cellular reprogramming in pursuit of immortality. *Cell Stem Cell* 11, 748–750
- Wu, S.M. and Hochedlinger, K. (2011) Harnessing the potential of induced pluripotent stem cells for regenerative medicine. *Nat. Cell Biol.* 13, 497–505
- Takahashi, K. *et al.* (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 131, 861–872
- Turner, M. *et al.* (2013) Toward the development of a global induced pluripotent stem cell library. *Cell Stem Cell* 13, 382–384
- Taylor, C.J. *et al.* (2012) Generating an iPSC bank for HLA-matched tissue transplantation based on known donor and recipient HLA types. *Cell Stem Cell* 11, 147–152
- Rao, M. *et al.* (2012) Concise review: cord blood banking, transplantation and induced pluripotent stem cell: success and opportunities. *Stem Cells* 30, 55–60
- Deshmukh, R.S. *et al.* (2012) Drug discovery models and toxicity testing using embryonic and induced pluripotent stem-cell-derived cardiac and neuronal cells. *Stem Cells Int.* 2012, 379569
- Kiskinis, E. and Eggen, K. (2010) Progress toward the clinical application of patient-specific pluripotent stem cells. *J. Clin. Invest.* 120, 51–59
- Chambers, S.M. *et al.* (2009) Highly efficient neural conversion of human ES and iPSC cells by dual inhibition of SMAD signaling. *Nat. Biotechnol.* 27, 275–280
- Maroof, A.M. *et al.* (2013) Directed differentiation and functional maturation of cortical interneurons from human embryonic stem cells. *Cell Stem Cell* 12, 559–572
- Studer, L. (2012) Derivation of dopaminergic neurons from pluripotent stem cells. *Prog. Brain Res.* 200, 243–263
- Song, Z. *et al.* (2009) Efficient generation of hepatocyte-like cells from human induced pluripotent stem cells. *Cell Res.* 19, 1233–1242
- Si-Tayeb, K. *et al.* (2010) Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* 51, 297–305
- Di Pasquale, E. *et al.* (2013) Generation of human cardiomyocytes: a differentiation protocol from feeder-free human induced pluripotent stem cells. *J. Vis. Exp.* 76, e50429
- Carpenter, L. *et al.* (2012) Efficient differentiation of human induced pluripotent stem cells generates cardiac cells that provide protection following myocardial infarction in the rat. *Stem Cells Dev.* 21, 977–986
- Zhang, D. *et al.* (2009) Highly efficient differentiation of human ES cells and iPSC cells into mature pancreatic insulin-producing cells. *Cell Res.* 19, 429–438
- Davilla, J.C. *et al.* (2009) Stem cell technology for embryotoxicity, cardiotoxicity and hepatotoxicity evaluation. In *Drug Efficacy, Safety, and Biologics Discovery: Emerging Technologies and Tools* (Ekins, S. and Xu, J.J., eds), pp. 175–213, John Wiley & Sons
- Guo, L. *et al.* (2011) Estimating the risk of drug-induced proarrhythmia using human induced pluripotent stem cell-derived cardiomyocytes. *Toxicol. Sci.* 123, 281–289
- Medine, C.N. *et al.* (2013) Developing high-fidelity hepatotoxicity models from pluripotent stem cells. *Stem Cells Transl. Med.* 2, 505–509
- Ekins, S. *et al.* (2010) A predictive ligand-based bayesian model for human drug induced liver injury. *Drug Metab. Dispos.* 38, 2302–2308
- Rowntree, R.K. and McNeish, J.D. (2010) Induced pluripotent stem cells: opportunities as research and development tools in 21st century drug discovery. *Regen. Med.* 5, 557–568
- Inoue, H. and Yamanaka, S. (2011) The use of induced pluripotent stem cells in drug development. *Clin. Pharmacol. Ther.* 89, 655–661
- Egashira, T. *et al.* (2013) Novel insights into disease modeling using induced pluripotent stem cells. *Biol. Pharm. Bull.* 36, 182–188
- Munos, B. (2009) Lessons from 60 years of pharmaceutical innovation. *Nat. Rev. Drug Discov.* 8, 959–968
- Paul, S.M. *et al.* (2010) How to improve R&D productivity: the pharmaceutical industry’s grand challenge. *Nat. Rev. Drug Discov.* 9, 203–214
- Ebert, A.D. *et al.* (2009) Induced pluripotent stem cells from a spinal muscular atrophy patient. *Nature* 457, 277–280
- Kiskinis, E. *et al.* (2014) Pathways disrupted in human ALS motor neurons identified through genetic correction of mutant SOD1. *Cell Stem Cell* 14, 781–795
- Chung, C.Y. *et al.* (2013) Identification and rescue of alpha-synuclein toxicity in Parkinson patient-derived neurons. *Science* 342, 983–987
- Itzhaki, I. *et al.* (2011) Modelling the long QT syndrome with induced pluripotent stem cells. *Nature* 471, 225–229
- Merkle, F.T. and Eggen, K. (2013) Modeling human disease with pluripotent stem cells: from genome association to function. *Cell Stem Cell* 12, 656–668
- Kamao, H. *et al.* (2014) Characterization of human induced pluripotent stem cell-derived retinal pigment epithelium cell sheets aiming for clinical application. *Stem Cell Reports* 2, 205–218
- Kanemura, H. *et al.* (2014) Tumorigenicity studies of induced pluripotent stem cell (iPSC)-derived retinal pigment epithelium (RPE) for the treatment of age-related macular degeneration. *PLoS ONE* 9, e85336

- 35 Soldner, F. and Jaenisch, R. (2012) Medicine. iPSC disease modeling. *Science* 338, 1155–1156
- 36 Kim, K. *et al.* (2010) Epigenetic memory in induced pluripotent stem cells. *Nature* 467, 285–290
- 37 Bock, C. *et al.* (2011) Reference maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. *Cell* 144, 439–452
- 38 Boulting, G.L. *et al.* (2011) A functionally characterized test set of human induced pluripotent stem cells. *Nat. Biotechnol.* 29, 279–286
- 39 Osafune, K. *et al.* (2008) Marked differences in differentiation propensity among human embryonic stem cell lines. *Nat. Biotechnol.* 26, 313–315
- 40 Gorba, T. and Allsopp, T.E. (2003) Pharmacological potential of embryonic stem cells. *Pharmacol. Res.* 47, 269–278
- 41 Pouton, C.W. and Haynes, J.M. (2007) Embryonic stem cells as a source of models for drug discovery. *Nat. Rev. Drug Discov.* 6, 605–616
- 42 McNeish, J.D. (2007) Stem cells as screening tools in drug discovery. *Curr. Opin. Pharmacol.* 7, 515–520
- 43 Barrilleaux, B. and Knoepfler, P.S. (2011) Inducing iPSCs to escape the dish. *Cell Stem Cell* 9, 103–111
- 44 Anderson, J.W. *et al.* (2013) Novel diaryl ureas with efficacy in a mouse model of malaria. *Bioorg. Med. Chem. Lett.* 23, 1022–1025
- 45 Hohman, M. *et al.* (2009) Novel web-based tools combining chemistry informatics, biology and social networks for drug discovery. *Drug Discov. Today* 14, 261–270
- 46 Luong, M.X. *et al.* (2011) A call for standardized naming and reporting of human ESC and iPSC lines. *Cell Stem Cell* 8, 357–359
- 47 Anon. (2014) *Yamanaka announces plan to establish global iPSC cell bank.* Asahi Shimbun 17 January
- 48 Nakatsuji, N. *et al.* (2008) HLA-haplotype banking and iPSC cells. *Nat. Biotechnol.* 26, 739–740
- 49 Gourraud, P.A. *et al.* (2012) The role of human leukocyte antigen matching in the development of multiethnic ‘haplobank’ of induced pluripotent stem cell lines. *Stem Cells* 30, 180–186
- 50 Sansone, S.A. *et al.* (2012) Toward interoperable bioscience data. *Nat. Genet.* 44, 121–126
- 51 Ho Sui, S.J. *et al.* (2012) The stem cell discovery engine: an integrated repository and analysis system for cancer stem cell comparisons. *Nucleic Acids Res.* 40, D984–D991
- 52 Bunin, B.A. and Ekins, S. (2011) Alternative business models for drug discovery. *Drug Discov. Today* 16, 643–645
- 53 Ekins, S. *et al.* (2010) A collaborative database and computational models for tuberculosis drug discovery. *Mol. Biosyst.* 6, 840–851
- 54 Ekins, S. *et al.* (2010) Analysis and hit filtering of a very large library of compounds screened against *Mycobacterium tuberculosis*. *Mol. Biosyst.* 6, 2316–2324
- 55 Ekins, S. *et al.* (2013) Enhancing hit identification in *Mycobacterium tuberculosis* drug discovery using validated dual-event Bayesian models. *PLoS ONE* 8, e63240
- 56 Ekins, S. *et al.* (2013) Bayesian models leveraging bioactivity and cytotoxicity information for drug discovery. *Chem. Biol.* 20, 370–378
- 57 Ioannidis, J.P. *et al.* (2012) Reproducibility concerns. *Nat. Med.* 18, 1736–1737
- 58 Lee, G. *et al.* (2012) Large-scale screening using familial dysautonomia induced pluripotent stem cells identifies compounds that rescue IKBKAP expression. *Nat. Biotechnol.* 30, 1244–1248
- 59 Engle, S.J. and Vincent, F. (2014) Small molecule screening in human induced pluripotent stem cell-derived terminal cell types. *J. Biol. Chem.* 289, 4562–4570
- 60 Xu, X. *et al.* (2013) Prevention of beta-amyloid induced toxicity in human iPSC cell-derived neurons by inhibition of cyclin-dependent kinases and associated cell cycle events. *Stem Cell Res.* 10, 213–227
- 61 Burkhardt, M.F. *et al.* (2013) A cellular model for sporadic ALS using patient-derived induced pluripotent stem cells. *Mol. Cell. Neurosci.* 56, 355–364
- 62 Charbord, J. *et al.* (2013) High throughput screening for inhibitors of REST in neural derivatives of human embryonic stem cells reveals a chemical compound that promotes expression of neuronal genes. *Stem Cells* 31, 1816–1828
- 63 Gupta, R.R. *et al.* (2010) Using open source computational tools for predicting human metabolic stability and additional ADME/TOX properties. *Drug Metab. Dispos.* 38, 2083–2090
- 64 Bersenev, A. (2014) Standards and reproducibility in stem cell research. *Open Sci.* 13 February
- 65 The Hinxtongroup, (2011) Statement on Policies and Practices Governing Data and Materials Sharing and Intellectual Property in Stem Cell Science. [http://hinxtongroup.wordpress.com/2011/01/21/consensus\\_statement/](http://hinxtongroup.wordpress.com/2011/01/21/consensus_statement/).
- 66 Brivanlou, A.H. *et al.* (2003) Stem cells. Setting standards for human embryonic stem cells. *Science* 300, 913–916
- 67 Visser, U. *et al.* (2011) BioAssay Ontology (BAO): a semantic description of bioassays and high-throughput screening results. *BMC Bioinformatics* 12, 257
- 68 Clark, A.M. *et al.* (2014) Fast and accurate semantic annotation of bioassays exploiting a hybrid of machine learning and user confirmation. *PeerJ.* 2, e524
- 69 Klips, A. (2014) *Senators Support Regenerative Medicine Promotion Act of 2014.* [secfilings.com](http://secfilings.com) 28 April
- 70 Doudement, E. and Uppal, H. (2014) New opportunities for stem cells in drug discovery and development. *Drug Discov. Dev.* 1 July
- 71 Ekins, S. *et al.* (2011) *Collaborative Computational Technologies for Biomedical Research.* Wiley
- 72 Luo, Y. *et al.* (2012) Generation of induced pluripotent stem cells from Asian patients with chronic neurodegenerative diseases. *J. Reprod. Dev.* 58, 515–521
- 73 Hackney, J.A. and Moore, K.A. (2005) A functional genomics approach to hematopoietic stem cell regulation. *Methods Mol. Med.* 105, 439–452
- 74 Montrone, C. *et al.* (2013) HSC-explorer: a curated database for hematopoietic stem cells. *PLoS ONE* 8, e70348
- 75 Hatano, A. *et al.* (2011) CELLPEDIA: a repository for human cell information for cell studies and differentiation analyses. *Database* 2011, bar046
- 76 Schulz, H. *et al.* (2009) The FunGenES database: a genomics resource for mouse embryonic stem cell differentiation. *PLoS ONE* 4, e6804
- 77 Knudsen, T. *et al.* (2013) Predictive models and computational toxicology. *Methods Mol. Biol.* 947, 343–374
- 78 Edgar, R. *et al.* (2013) LifeMap discovery: the embryonic development, stem cells, and regenerative medicine research portal. *PLoS ONE* 8, e66629