Integrated Collection of Stem Cell Bank data, a data portal for standardized stem cell information

Authors

Ying Chen[1], Kunie Sakurai[1†], Sumihiro Maeda[2], Tohru Masui[3], Hideyuki Okano[2], Johannes Dewender[4], Stefanie Seltmann[4], Andreas Kurtz[4], Hiroshi Masuya[5], Yukio Nakamura[6], Michael Sheldon[7], Juliane Schneider[8], Glyn N. Stacey[9,10,11], Yulia Panina[1], and Wataru Fujibuchi[1,*]

Affiliations

[1] Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Sho-goin, Sakyo-ku, Kyoto 606-8507, Japan

[2] Department of Physiology, Keio University School of Medicine, Tokyo 160-8582, Japan

[3] National Center for Medical Genetics, Keio University School of Medicine, Tokyo 160-8582, Japan

[4] BIH Center for Regenerative Therapies (BCRT), Charité – Universitätsmedizin Berlin, Augustenburger Platz 1, 13353 Berlin, Germany

[5] Integrated Bioresource Information Division, RIKEN BioResource Research Center, Tsukuba, Ibaraki 305-0074, Japan

[6] Cell Engineering Division, RIKEN BioResource Research Center, Tsukuba, Ibaraki 305-0074, Japan

[7] Department of Genetics and Human Genetics Institute of New Jersey, Rutgers, The State University of New Jersey, Piscataway, New Jersey 08854, U.S.A.

[8] Harvard Catalyst | Clinical and Translational Science Center, Boston, Massachusetts 02215, U.S.A.

[9] International Stem Cell Banking Initiative, 2 High Street, Barley, Herts, SG88HZ, UK

[10] National Stem Cell Resource Centre, Institute of Zoology, Chinese Academy of Sciences, Beijing 100190, China

[11] Innovation Academy for Stem Cell and Regeneration, Chinese Academy of Sciences, Beijing 100101, China

Additional Footnotes

† Present address: Department of Cellular Biology and Pharmacology, Florida International University, Herbert Wertheim College of Medicine, Florida 33199, U.S.A.

Contact

Wataru Fujibuchi*

Address: Center for iPS Cell Research and Application (CiRA), Kyoto University, 53 Kawahara-cho, Sho-goin, Sakyo-ku, Kyoto 606-8507, Japan

E-mail: fujibuchi-g@cira.kyoto-u.ac.jp

Tel: +81-75-366-7012

SUMMARY

The last decade has witnessed an extremely rapid increase in the number of newly established stem cell lines worldwide. However, due to the lack of a standardized format, data exchange among stem cell line information resources has been challenging, and no system can search all stem cell lines across resources worldwide. To solve this problem, we have developed the Integrated Collection of Stem Cell Bank data (ICSCB) (http://icscb.stemcellinformatics.org/), a new and largest database search portal for stem cell line data from various resources, based on the standardized data items and terms of the MIACARM framework. Currently, ICSCB can retrieve 15,796 cell lines from four major data resources in Europe, Japan, and the U.S. ICSCB is automatically updated to provide the latest cell line information, and its integrative search engine helps users collect cell line information from donors with rare diseases worldwide, which has been a formidable task, thereby distinguishing itself from other database search portals.

INTRODUCTION

Since the first report of human induced pluripotent stem cells (iPSCs) (Takahashi et al., 2007), there has been a rapid rise in the number of iPSC lines and related information worldwide **(Table 1)**. This remarkable growth has not only accelerated studies of regenerative medicine but also provided opportunities to understand such pragmatic issues as the quality of pluripotent stem cells (Nishizawa et al., 2016) and the disease mechanisms (Sasaki et al., 2016). Stem cell banks and registries are expected to provide necessary data of individual stem cell lines. However, the exchange of data among different institutions is not a trivial matter, and scientific reproducibility based on available information is problematic for both basic studies and clinical applications (Yaffe et al., 2016; Isasi and Knoppers, 2011; Thirumala et al., 2009). Moreover, as technologies for the characterization of cell lines continue to progress, the addition of new quality standards as necessary data items is complicating and diversifying data formats among different stem cell banks and registries (Hug, 2009; Knoppers and Isasi, 2010). As an attempt to solve these problems, we previously reported MIACARM (Minimum Information About a Cellular Assay for Regenerative Medicine) guidelines in

2016 (Sakurai et al., 2016), which proposed the utilization of standardized data items and formats for all stem cell lines in regenerative medicine. Presently, MIACARM contains 258 items, covering such areas as stem cell production and materials (e.g., donor information, source cell information, and cell culture medium and substrate information), cell banking processes, cell characterization, sterility testing, and even ethical concerns. Later, a standardized nomenclature for pluripotent stem cells was introduced in 2018 with unification of cell line codification and minimization of information loss and confusion regarding cell lines as goals (Kurtz et al., 2018). Nevertheless, with the growing number of registered cell lines, existing data deposition formats have made it increasingly harder for not only data depositors but also users to seek and obtain cell lines collected under different projects, disease status, and privacy issues (Godard et al., 2003; Winickoff et al., 2009).

In this paper, as our next step towards the unification and utilization of stem cell line data in the world, we report our new database portal, Integrated Collection of Stem Cell Bank data (ICSCB), which was designed using MIACARM guideline items and formats. The main objectives of ICSCB are i) to establish an integrated stem cell database portal that can cover the majority of stem cell resources in the world, and ii) to offer users minimum but efficient access to information on stem cell lines based on MIACARM guidelines. Currently, ICSCB provides data of more than 15,000 stem cell lines registered in four major stem cell line databases: hPSCreg (Seltmann et al., 2015), SKIP (Kim et al., 2017), RIKEN BRC (Kobayashi et al., 2016), and eagle-i (Vasilevsky et al., 2012). ICSCB has a user-friendly search engine for stem cell lines and can be accessed directly at http://icscb.stemcellinformatics.org/, or as a slim version by removing cell line redundancy as much as possible through the SHOGoiN (Human Omics Database for the Generation of iPS and Normal Cells) homepage at http://shogoin.stemcellinformatics.org/.

RESULTS AND DISCUSSION

Web interface

ICSCB was designed for researchers searching for available cell lines to conduct various studies, such as regenerative medicine and disease analysis. Covering as many diverse cell lines as possible was the first priority when deciding which resources to include in

ICSCB. Sharing cell line information between different stem cell banks and registries is problematic due to different cell naming methods, different policies on cell assessment in different registries, unclear data sources, and so on. ICSCB is a collection of cell lines from four major and reliable cell line data resources based in Europe, Japan, and the United States. ICSCB updating is regularly performed for new SKIP and eagle-i stem cell line data as well as automatically performed for hPSCreg and RIKEN BRC data in a synchronized manner. Users can retrieve all related stem cell line information by using a free text search. Detailed information for a specific cell line can be accessed by clicking on the stem cell ID, which is linked to the information page in the original resources (**Figure 1**). The results can be further filtered according to users' requests. There may be several records for the same cell line if the cell line is included in multiple data resources. To provide users as much information as possible, the results page is designed to show cell lines with matching cell names as well as close descriptions.

Data coverage

ICSCB covers more data than any other stem cell line repository available. The integration of all major data resources allows us to check the current state of stem cell research in the world **(Figure 2)**. Although we recognize redundancies in the data, according to our statistics, the number of iPSC lines constitutes more than 80% of all cell lines and the ratio of healthy to diseased donors is approximately 3 to 2 **(Figure 2A,B)**. The total number of countries from which cell lines can be retrieved is 36 (as of March 26, 2020), of which the top 9 countries identified in SKIP and hPSCreg are (in descending order) the United Kingdom, United States, Japan, Germany, China, Spain, Sweden, Denmark, and Italy **(Figure 2C)**. In addition, as the recent number of iPSC lines generated from patient donors is growing, ICSCB supports disease-oriented search to help users find all disease-related stem cell lines by using disease names. The distribution of disease and disorder types is shown in **Figure 2D**.

Easy search interface on SHOGoiN homepage

ICSCB also has a quick and easy search module on the SHOGoiN homepage (https://stemcellinformatics.org/). SHOGoiN is a repository for accumulating and integrating diverse human cell information to support a wide range of research using cell-related data. The database consists of several modules that store cell lineage maps, transcriptomes, methylomes, cell conversions, cell type markers, and cell images with morphology data curated from public as well as contracted resources based on sophisticated cell taxonomy. Collaboration between ICSCB and SHOGoiN makes it possible for users to directly use free text searches for stem cell line data on the SHOGoiN homepage. The ICSCB easy search module in SHOGoiN supports a simplified ICSCB search with keywords, and the advanced search is designed to redirect users to the ICSCB homepage with full functions. Results from the SHOGoiN homepage share the same structure with the ICSCB homepage.

Concluding remarks and future plan

So far, the registration and submission of newly established cell lines have been complicated by the lack of standardized data formats. Most data registries are currently limited by respective domestic policies and have adopted their information structures and validation processes independently (Andrews et al., 2015; Zarzeczny et al., 2009). The lack of standardized data formats has caused problems for researchers who must usually search several websites to find the stem cell lines they are looking for (Wells et al., 2013). In the present work, we developed ICSCB, an integrated data distribution system that provides stem cell line information from major stem cell banks and registries all over the world. ICSCB adopts a standardized information format based on the "Source Cell" module of MIACARM to integrate different data resources while keeping important information.

In the future, in order to respond to the rapid growth in the number of stem cell lines, we will include more data resources in ICSCB, including the Taiwan Human Disease iPSC Service Consortium and other recently developed stem cell banks, to make ICSCB more resource-abundant and usable. We also plan to add a detailed quality check to help users find stem cell lines of high quality. As the largest stem cell line information resource, we will support stem cell communities by improving the quality and increasing the scale of our database.

EXPERIMENTAL PROCEDURES

Data resources

ICSCB resources were selected from existing major stem cell registries that collect cell line information in Europe, Japan, and the U.S., and stem cell banks that provide cell lines with information of the attributes. We checked the number of registered cell lines and the criteria for registration in these registries and banks to decide to what extent their cell line data can fulfill MIACARM guidelines for inclusion in ICSCB. Considering the size, accessibility, and diversity of the different databases, three stem cell registries and one stem cell bank were included: (1) SKIP (5,615 cell lines), (2) hPSCreg (3,099 cell lines), (3) RIKEN BRC (3,534 cell lines), and (4) eagle-i (3,548 cell lines) (as of March 26, 2020). These data resources were selected because they had the highest number of registered cell lines and large diversity, which would provide a good regional balance of cell sources to reduce redundancies in cell line entries. RIKEN BRC basically collected cell lines from Japanese institutions, SKIP contained data mostly from other Japanese and Asian institutions, hPSCreg collected data mainly from European institutions, and eagle-i collected data mostly from the United States. Details of the data sources are listed in **Figure 3**.

Data integration

Since our previous research on the listed stem cell banks (Sakurai et al., 2016), the number of registered cell lines had skyrocketed from 1,483 to approximately 8,000 in the past three years. As a result, stem cell registries are facing the demand to collect information on the rapidly increasing number of new cell lines and register the cell lines into their databases as quickly as possible. However, because the stem cell banks and registries are using their own formats for data entry, the integration of the data into a centralized collection system is an extraordinary challenge. To solve this problem, we used a decentralized or distributed database system (Fujibuchi et al., 1998) by adopting items of different database formats into 12 attributes, or terms, from three MIACARM modules: stem cell general identification, donor identification, and source cell identification (**Table 2**). To practically integrate the data from the four data resources (SKIP, hPSCreg, RIKEN BRC, and eagle-i), we adopted a mechanism of cross-reference tables that allow users to conduct a search using MIACARM terms that are translated into the corresponding terms in the individual data resources to implement the search. For example, the term "Stem cell ID" in MIACARM was translated into the terms "stem cell id" (hPSCreg), "stem cell id" (SKIP), "CellID" (RIKEN BRC), and "cell line label" (eagle-i) for the search implementation. Thus, ICSCB submits search requests to each data resource with its own (translated) terms and integrates all retrieved results by common MIACARM terms, thereby achieving a standardized data format at the level of display (**Figure 4**).

ICSCB workflow and search engine updating

In order to provide fast and easy access to the latest and accurate cell line information, we built an automatic updating system that adds newly released cell lines to ICSCB as soon as they become available in any of the four data resources. Data from eagle-i and SKIP are directly collected and stored in the MySQL database with the terms required for the MIACARM modules. Data from hPSCreg and RIKEN BRC are collected on the fly per request using a web application programming interface (API) provided by the respective sites. RIKEN BRC also uses SPARQL language for data retrieval requests (Kim et al., 2017; Kobayashi et al., 2016).

To simplify the search process, ICSCB provides an easy-to-use and mobile-friendly web application. The goal of the application is to help users find the desired stem cell lines as quickly as possible. The interface of the search engine is designed with the 12 MIACARM terms (**Table 2**) except the term "Stem cell ID". Users receive result pages with all the matching results listed in a table that includes all the basic attributes under the structure of MIACARM. To ensure a more specific search with a wide variety of attributes, ICSCB is designed to accommodate searches not only by standardized terms from MIACARM but also by terms specific to each of the four data resources, such as "age" or "country" (**Figure 5A**). When user queries are submitted, ICSCB simultaneously retrieves MIACARM standardized data and resource-specific data so as not to miss any relevant entries. If a keyword entered by a user in a general keyword search does not exist in MIACARM terms but is included in data specific to any of the four data resources, the user will get detailed descriptions of the matching data in the results page. For example, even if the standardized MIACARM terms do not contain "transgene", it is still possible to enter a gene name into the keyword field (e.g., Sox2) such that the results page will display relevant entries by showing the indicated keyword in the extra field below **(Figure 5B)**. Furthermore, the user can filter the results by data resource and detailed keywords from “Searching options” box inside the results page to narrow down the results list. In addition, all the results can be easily downloaded as a table directly from the results page.

In addition, ICSCB provides a quality control panel based on MIACARM, supporting customized searches according to quality control results. At present, assays for teratoma formation, differentiation ability in vitro, morphology data, marker gene expression/surface antigen expression data, karyotyping assay results, copy number variation, residual exogene detection results, genome profiling, transcriptome profiling, and epigenome profiling data are accessible from ICSCB.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, three figures, and five tables, and can be found with this article online at:

AUTHOR CONTRIBUTIONS

YC and YP drafted the manuscript. SM, TM, HO, JD, SS, AK, HM, YN, MS, JS, and WF provided and facilitated the stem cell data. WF conceptualized the research. AK, GS, and WF led the project.

ACKNOWLEDGEMENTS

The authors deeply appreciate Dr. Peter Karagiannis for kindly reviewing the manuscript. This work was partially supported by the Core Center for iPS Cell Research, Research Center Network for Realization of Regenerative Medicine (16bm0104001h0004), and the Formulation of Regenerative Medicine National Consortium which Renders Nation-wide Assistance to Clinical Researches, Project to Build Foundation for Promoting Clinical Research of Regenerative Medicine

(19bk0204001h0004, 19bk0204001s0104), Japan Agency for Medical Research and Development, AMED, and German Academic Exchange Service (DAAD) PPP grant.

COMPETING INTERESTS

H.O. is a founding scientist of SanBio Co., Ltd. and K Pharma Inc.

REFERENCES

Andrews, P.W., Baker, D., Benvinisty, N., Miranda, B., Bruce, K., Brüstle, O., Choi, M., Choi, Y.-M., Crook, J.M., Dvorak, P. (2015). Points to consider in the development of seed stocks of pluripotent stem cells for clinical applications: International Stem Cell Banking Initiative (ISCBI). Regen. Med. *10*, 1–44.

Fujibuchi, W., Goto, S., Migimatsu, H., Uchiyama, I., Ogiwara, A., Akiyama, Y., Kanehisa, M. (1998). DBGET/LinkDB: an integrated database retrieval system. In Pac. Symp. Biocomput. *98*, 683–694.

Godard, B., Schmidtke, J., Cassiman, J.J., Aymé, S. (2003). Data storage and DNA banking for biomedical research: informed consent, confidentiality, quality issues, ownership, return of benefits. A professional perspective. Eur. J. Hum. Genet. *11*, S88–S122

Hug, K. Banks, repositories and registries of stem cell lines in Europe: regulatory and ethical aspects. Stem Cell Rev. Rep. *5*, 18–35.

Isasi, R., and Knoppers B.M. (2011). From banking to international governance: fostering innovation in stem cell research. Stem cells Int. 2011, 498132.

Kim, J.H., Kurtz, A., Yuan, B.Z., Zeng, F., Lomax, G., Loring, J.F., Crook, J., Ju, J.H., Clarke, L., Inamdar, M.S., et al. (2017). Report of the International Stem Cell Banking Initiative workshop activity: current hurdles and progress in seed-stock banking of human pluripotent stem cells. Stem Cells Transl. Med. *6*, 1956–1962.

Knoppers, B.M., and Isasi, R. (2010). Stem cell banking: between traceability and identifiability. Genome Med. *2*, 73

Kobayashi, N., Lenz, K., and Masuya, H. (2016). RIKEN MetaDatabase: a database platform as a microcosm of linked open data cloud in the life sciences. In Joint International Semantic Technology Conference. (Springer, Cham), pp. 99–115.

Kurtz, A., Seltmann, S., Bairoch, A., Bittner, M.S., Bruce, K., Capes-Davis, A., Laurence, D., Johannes, D., et al. (2018). A standard nomenclature for referencing and authentication of pluripotent stem cells. Stem Cell Rep. *10*, 1–6.

Nishizawa, M., Chonabayashi, K., Nomura, M., Tanaka, A., Nakamura, M., Inagaki, A., Nishikawa, M., Takei, I., Oishi, A., Tanabe, K., et al. (2016). Epigenetic variation between human induced pluripotent stem cell lines is an indicator of differentiation capacity. Cell Stem Cell *19*, 341–354.

Sakurai, K., Kurtz, A., Stacey, G., Sheldon, M., and Fujibuchi, W. (2016). First proposal of minimum information about a cellular assay for regenerative medicine. Stem Cells Transl. Med. *5*, 1345–1361.

Sasaki, K., Makiyama, T., Yoshida, Y., Wuriyanghai, Y., Kamakura, T., Nishiuchi, S., Hayano, M., Harita, T., Yamamoto, Y., Hirose, S., et al. (2016). Patient-specific human induced pluripotent stem cell model assessed with electrical pacing validates S107 as a potential therapeutic agent for catecholaminergic polymorphic ventricular tachycardia. PLoS One *11*. e0164795.

Seltmann, S., Lekschas, F., Müller, R., Stachelscheid, H., Bittner, M.S., Zhang, W., Luam, K., Anna, S., Anna, V., Stacey, G.N., et al. (2015). hPSCreg—the human pluripotent stem cell registry. Nucleic Acids Res. *44*, D757–D763.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *31*, 861–872.

Thirumala, S., Goebel, W.S., Woods, E.J. (2009). Clinical grade adult stem cell banking. Organogenesis *5*, 143–154.

Vasilevsky, N., Johnson, T., Corday, K., Torniai, C., Brush, M., Segerdell, E., Wilson, M., Shaffer, C., Robinson, D., Haendel, M. (2012). Research resources: curating the new eagle-i discovery system. Database *2012*, bar067.

Wells, C.A., Mosbergen, R., Korn, O., Choi, J., Seidenman, N., Matigian, N.A., Vitale, A.M., Shepherd, J. (2013). Stemformatics: visualisation and sharing of stem cell gene expression. Stem Cell Res. *10*, 387–395.

Winickoff, D.E., Saha, K., Graff, G.D. (2009). Opening stem cell research and development: A policy proposal for the management of data, intellectual property, and ethics. Yale J. Health Pol'y, L. & Ethics *9*, 52–127.

Yaffe, M.P., Noggle, S.A., and Solomon, S.L. (2016). Raising the standards of stem cell line quality. Nat. Cell Biol. *18*, 236.

Zarzeczny, A., Scott, C., Hyun, I., Bennett, J., Chandler, J., Chargé, S., Heine, H., Isasi, R., Kato, K., Lovell-Badge, R., et al. (2009). iPS cells: mapping the policy issues. Cell, *139*, 1032–1037.

FIGURE LEGENDS

Fig. 1: Web interface of ICSCB.

(A) The ICSCB search page. Any keyword related to cell lines (including cell line name, disease name, gender, and so on) can be used to perform an instant search. (B) The ICSCB results page. Matched or partially matched cell lines are listed according to MIACARM terms. To check the details of the cell lines, the user can click on the stem cell ID, which is linked to the original source of cell line information.

Fig 2. **Details of cell lines collected by ICSCB**.

Cell line information is categorized as (A) stem cell type, (B) health/disease status of donor, (C) country that established the cell lines, and (D) disease category.

Fig. 3: Overview of ICSCB.

ICSCB includes data from three stem cell registries and one cell bank in order to maximize data coverage worldwide.

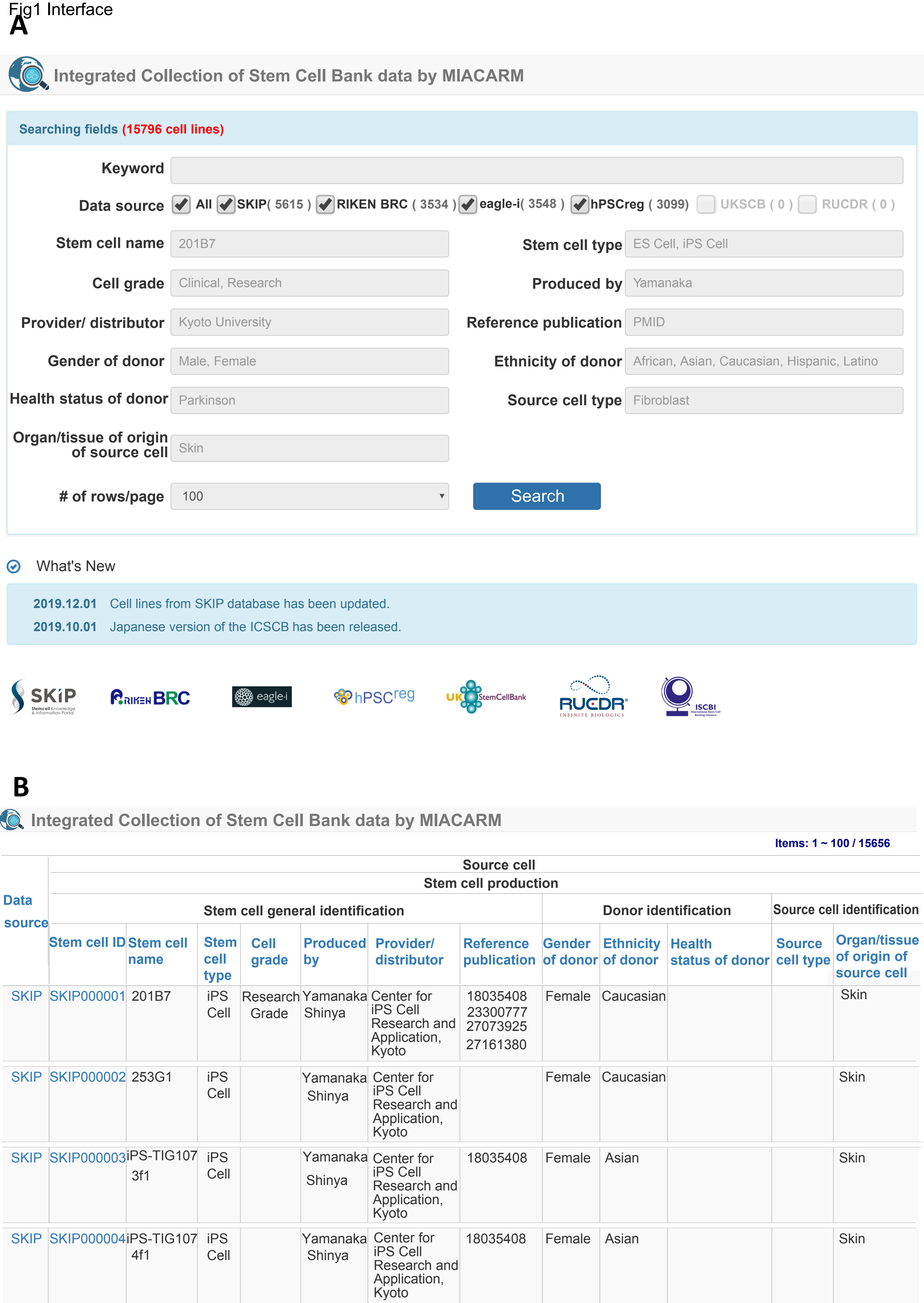
Fig. 4: Workflow of ICSCB data integration.

The SKIP and eagle-i databases were fully replicated from websites and imported to MySQL (even when updating ICSCB), whereas hPSCreg and RIKEN BRC used a web API and SPARQL for data collection. Cross-reference tables (Table 3) were used when ICSCB integrated and standardized cell line data.

Fig. 5: Keyword search is automatically extended to all terms provided by the four data resources even if a keyword is not included in standardized MIACARM terms.

(A) Terms specific to each of the four data resources. (B) Even if the standardized MIACARM terms do not contain, for example, “transgene”, it is still possible to enter a gene name into the keyword field (e.g., Sox2), which will lead users to results from the four data resources with relevant information. The results of the match will be shown in another row below the standardized fields.

Fig1 Interface



As of March 26, 2020

Fig2 Pie chart

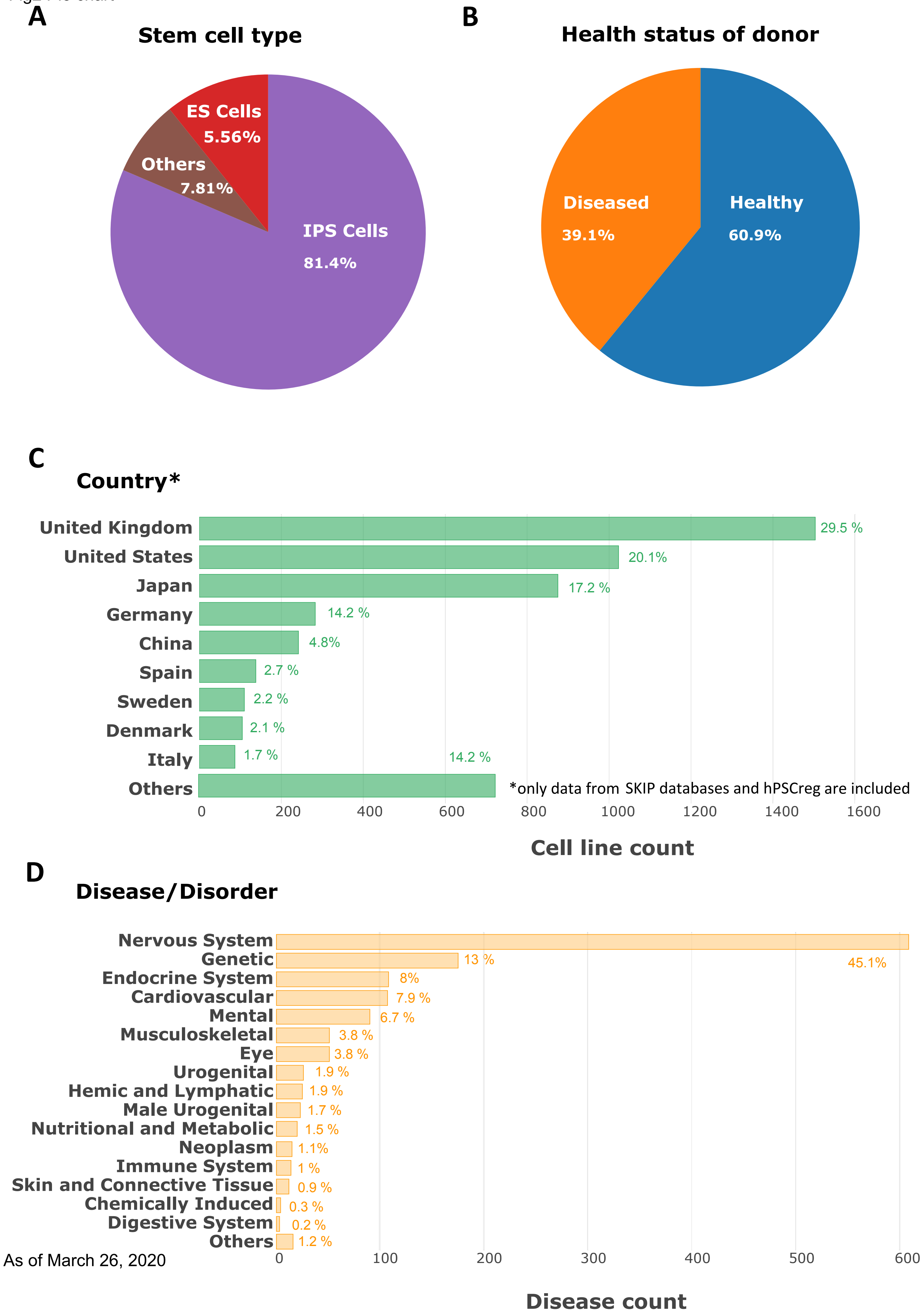


Fig3 Overview

ICSCB

Database	Registry/Bank(Country)	Countries from which data were acquired
SKIP	Registry(Japan) skip.stemcellinformatics.org/en/	Australia; Canada; China; Czech Republic; Denmark; Finland; Germany; Hungary; India; Iran; Israel; Italy; Japan; Netherlands; Poland; Russia; Singapore; Slovenia; South Korea; Spain; Sweden; Switzerland; Taiwan; Thailand; United Kingdom; United States.
RIKEN	Bank(Japan) www.riken.jp/en/	Japan*1
eagle-i	Registry(U.S.) www.eagle-i.net	United States*2
hPSCreg	Registry(EU) hpscreg.eu	Australia; Austria; Belgium; Brazil; Canada; China; Czech Republic; Denmark; Finland; France; Germany; Hungary; India; Iran; Ireland; Israel; Italy; Japan; Luxembourg, Netherlands; Portugal; Norway; Qatar; Russia; Singapore; Slovenia; South Korea; Spain; Sweden; Switzerland; Taiwan; Thailand; Turkey; United Kingdom; United States.

1. Related organization list: https://www.amed.go.jp/content/000043772.pdf
2. Participating institutions: https://www.eagle-i.net/about/participating-institutions/

As of March 26, 2020

Fig4 Workflow

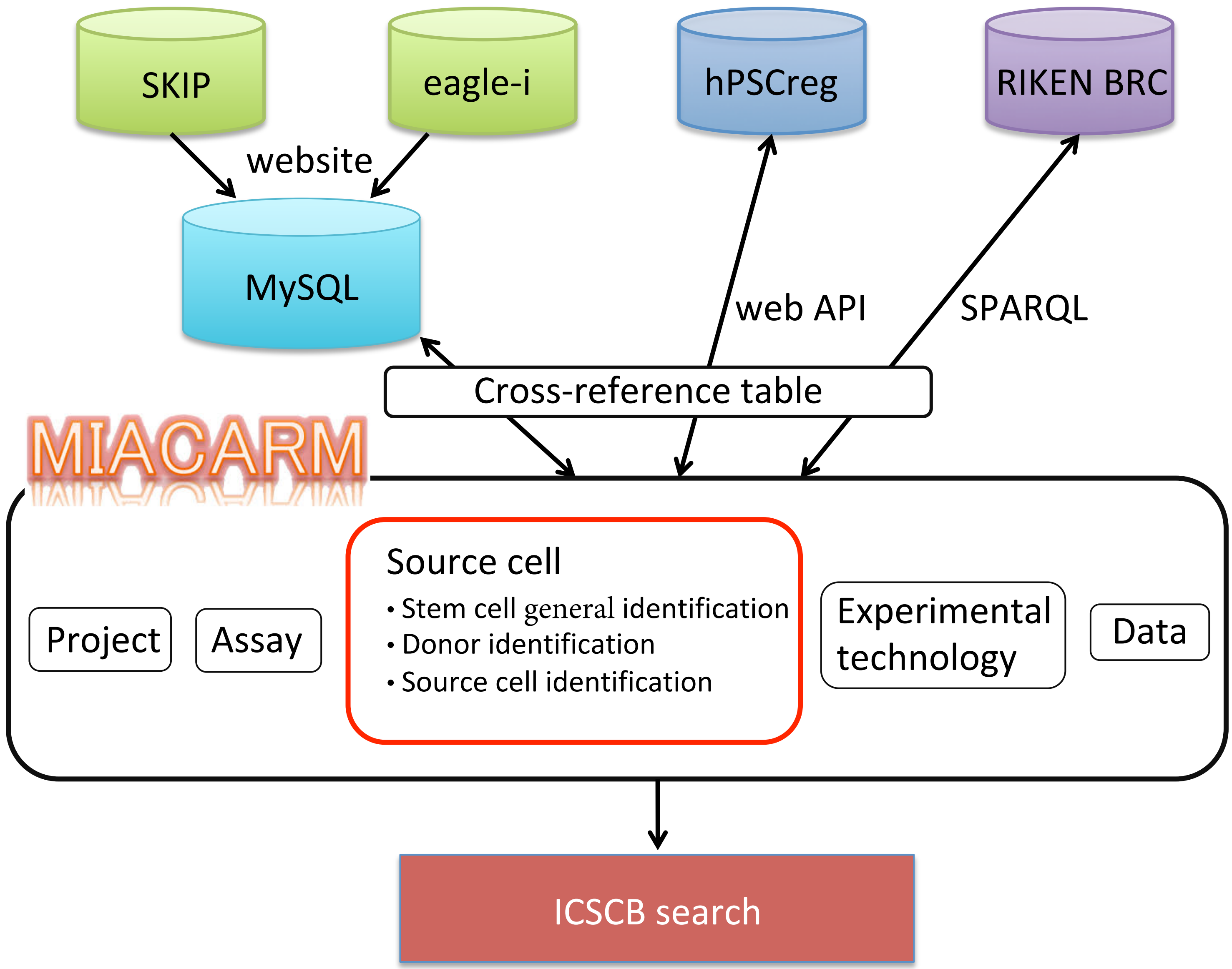


Fig5 Reference table

A

Keyword search

SKIP

- cell id
- stemcell_id
- cell line name
- research grade
- establisher name
- establisher organization
- pubmed ID
- donor sex
- donor race
- disease name
- cell type
- organ/tissue of origin of source cell
- in vivo differentiation assay
- in vitro differentiation
- cell morphology
- pluripotent marker
- karyotype assay
- CNV detail
- remaining vector detection test assay
- whole genome detail
- stem cell transcriptome analysis detail
- epigenetics detail

eagle-i

- cell id
- cell line label
- cell line provider
- sex
- ethnicity
- diagnosed disease
- cell line type
- cell line URL

hPSCreg

- source organism
- race
- cell lineage
- health status
- shogoin cell id
- stem cell id
- stem cell name
- produced by
- distributor
- publication
- gender
- race
- health status
- source cell type
- source cell description
- origin of source cell

Riken BRC

- cell name
- cell ID
- description
- original website
- originator
- depositor
- taxon
- depiction
- common name
- cell grouping
- reference
- derived from
- gender
- race
- country
- disease
- age

B

Integrated Collection of Stem Cell Bank data by MIACARM

Data source	Source cell											
	Stem cell production											
	Stem cell general identification							Donor identification			Source cell identification	
	Stem cell ID	Stem cell name	Stem cell type	Cell grade	Produced by	Provider/ distributor	Reference publication	Gender of donor	Ethnicity of donor	Health status ofdonor	Source cell type	Organ/ tissue of origin of source cell
SKIP	SKIP000001	201B7	iPS Cell	Research Grade	Yamanaka Shinya.	Center for iPS Cell Research and Application, Kyoto University	18035408 -- 23300777 -- 27073925 -- 27161380	Female	Caucasian			Skin

CellLine:Transgene:Retrovirus, pMXs-Oct3/4, -Sox2, -Klf4, -c-Myc, pLenti6/UbC/V5-DEST, mouse Slc7a1

Table1

B/R*	Stem cell bank or registry	Country	Website	Number of cell lines
B	BLCB	Spain	http://www.cmrb.eu/	79
R	hPSCreg	Germany	https://hpscreg.eu/	3099
R	HipSci	United Kingdom	http://www.hipsci.org/	799
B	U.K. Stem Cell Bank	United Kingdom	https://www.nibsc.org/	36
B	EBiSC	Germany	https://ebisc.org/	897
B/R	CIRM / FUJIFILM	United States	https://fujifilmcdi.com/the-cirm-ipsc-bank/	1545
B	Harvard Stem Cell Institute	United States	http://stemcelldistribution.harvard.edu/	40
B	NYSCF	United States	https://nyscf.org/	63
B	NINDS Human Cell and Data Repository	United States	https://bioq.nindsgenetics.org/	377
B	WiCell Research Institute	United States	https://www.wicell.org/	1505
B/R	eagle-i	United States	https://www.eagle-i.net/	3548
B/R	RIKEN BRC	Japan	https://en.brc.riken.jp/	3545
R	SKIP	Japan	https://skip.stemcellinformatics.org/	5615
B	JCRB	Japan	https://cellbank.nibiohn.go.jp/	16
B	Taiwan Human Disease iPSC Service Consortium	Taiwan	https://catalog.bcrc.firdi.org.tw/Welcome/	89
B	National Stem Cell Bank of Korea	Korea	http://kscr.nih.go.kr/nscb/en/kscr/index.do/	172

* B, bank; R, registry

Table 1. Stem cell banks and registries worldwide (as of March 26, 2020)

Table2

MIACARM module	ICSCB term	hPSCreg	SKIP	RIKEN BRC	eagle-i
Stem cell general identification	Stem cell ID	stem cell id	stem cell id	CellID	cell line label
	Stem cell name	stem cell name	cell line name	CellName	cell line label
	Stem cell type	N.A.	cell type	cell grouping	cell line type
	Cell grade	N.A.	research grade	N.A.	N.A.
	Produced by	produced by	establisher name	originator	N.A.
	Provider/ distributor	distributor	establisher organization	depositor	cell line provider
	Reference publication	publication	pubmed ID	reference	N.A.
Donor identification	Gender of donor	gender	donor sex	gender	sex
	Ethnicity of donor	race	donor race	race	ethnicity
	Health status	health status	disease name	disease	diagnosed disease
Source cell identification	Source cell type	source cell type	N.A.	N.A.	N.A.
	Organ/tissue of origin of source cell	origin of source cell	organ/tissue of origin of source cell	N.A.	N.A.

N.A., not available.

Table 2. Cross-reference table for integration of four databases according to MIACARM module (as of March 26, 2020).

SUPPLEMENTAL INFORMATION

Integrated Collection of Stem Cell Bank data, a data portal for standardized stem cell information

Authors

Ying Chen, Kunie Sakurai, Sumihiro Maeda, Tohru Masui, Hideyuki Okano, Johannes Dewender, Stefanie Seltmann, Andreas Kurtz, Hiroshi Masuya, Yukio Nakamura, Michael Sheldon, Juliane Schneider, Glyn N. Stacey, Yulia Panina, and Wataru Fujibuchi

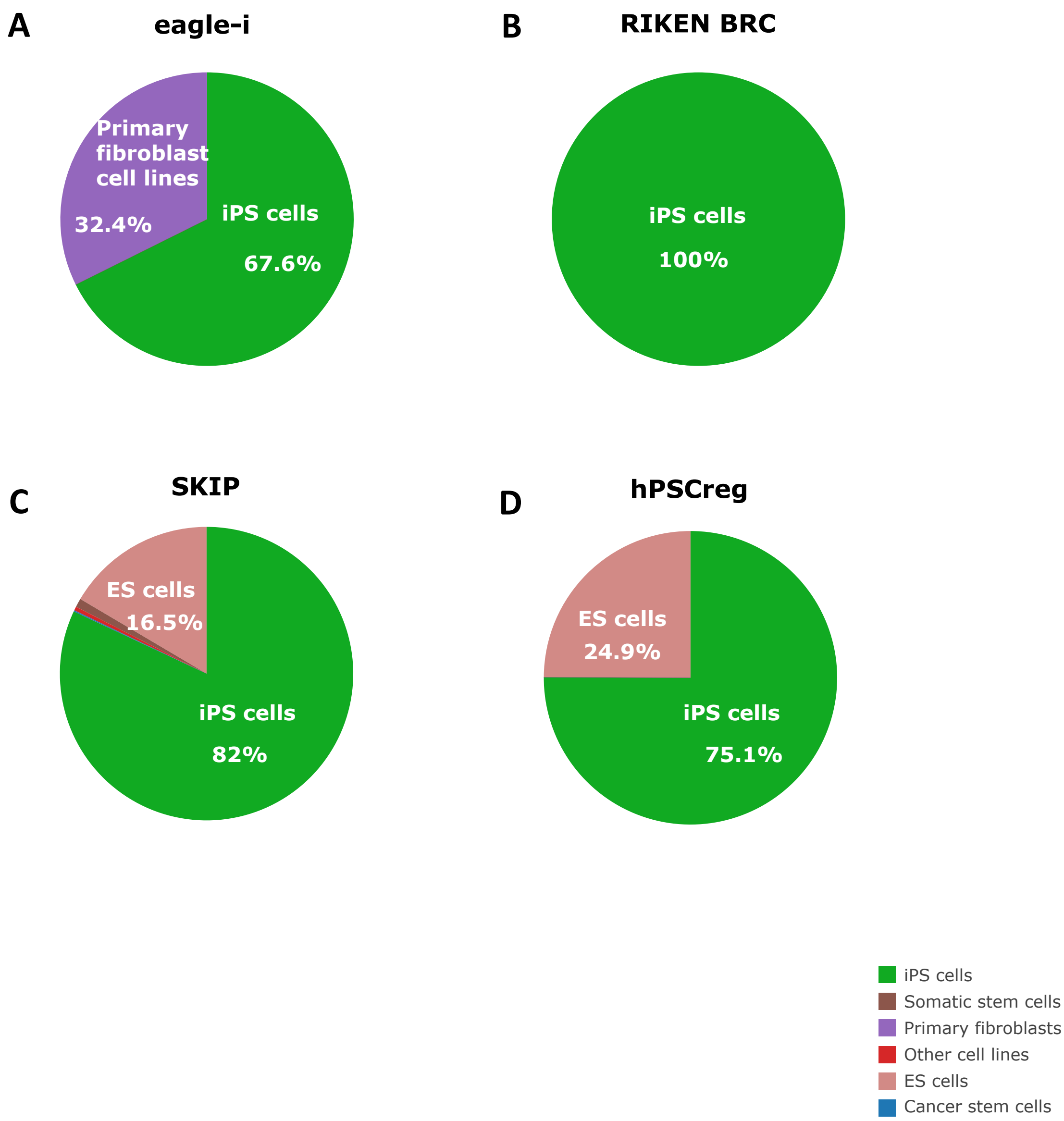
SUPPLEMENTAL FIGURES

Fig S1. Details of cell line types collected by eagle-i, RIKEN BRC, SKIP, and hPSCreg (as of March 26, 2020)**.** (A) eagle-i, (B) RIKEN BRC, (C) SKIP, and (D) hPSCreg

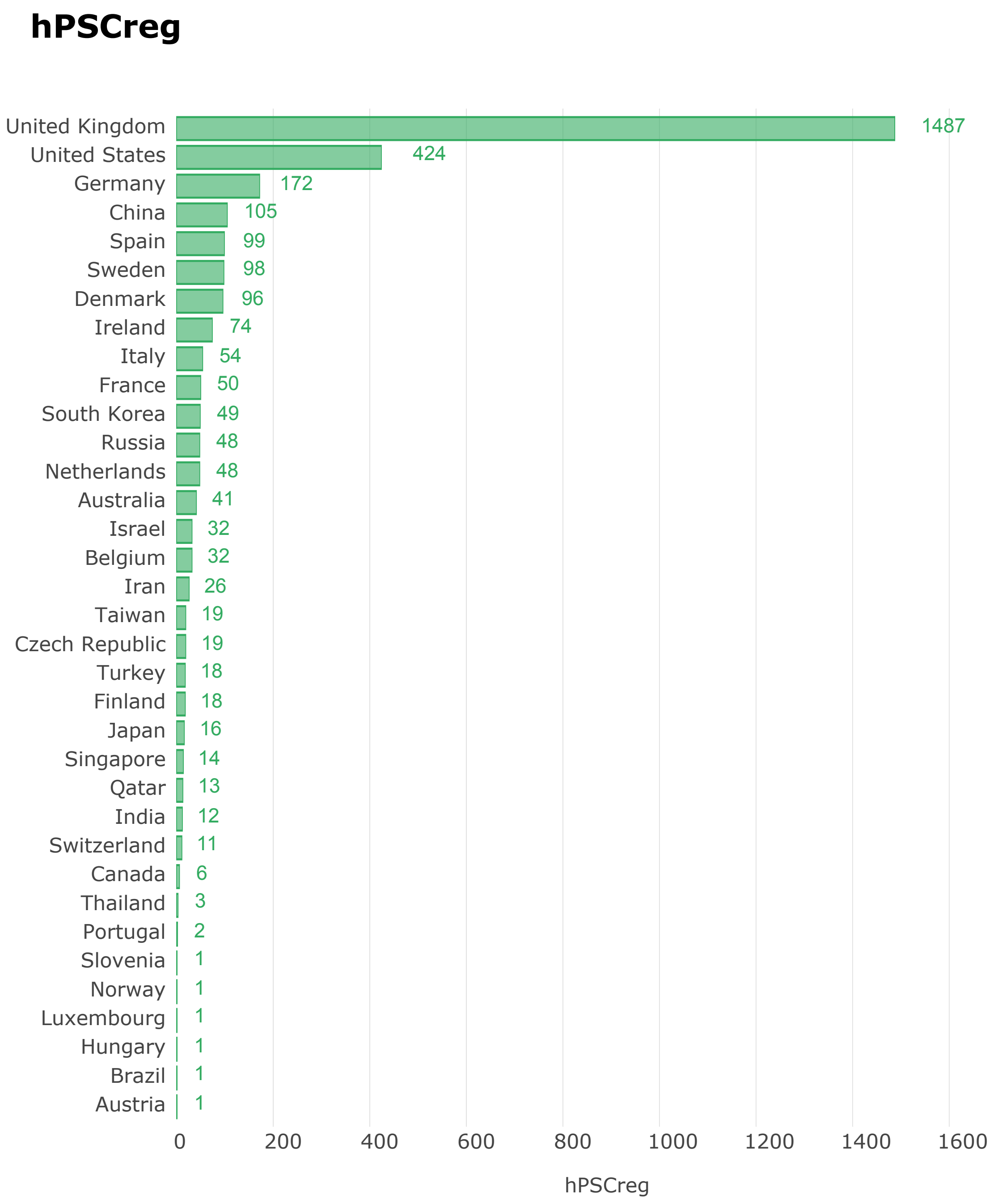
Fig S2. **Details of countries that have established cell lines in hPSCreg** (as of March 26, 2020)**.**

Fig S3. Details of countries that have established cell lines in SKIP (as of March 26, 2020)**.**

Fig. S1

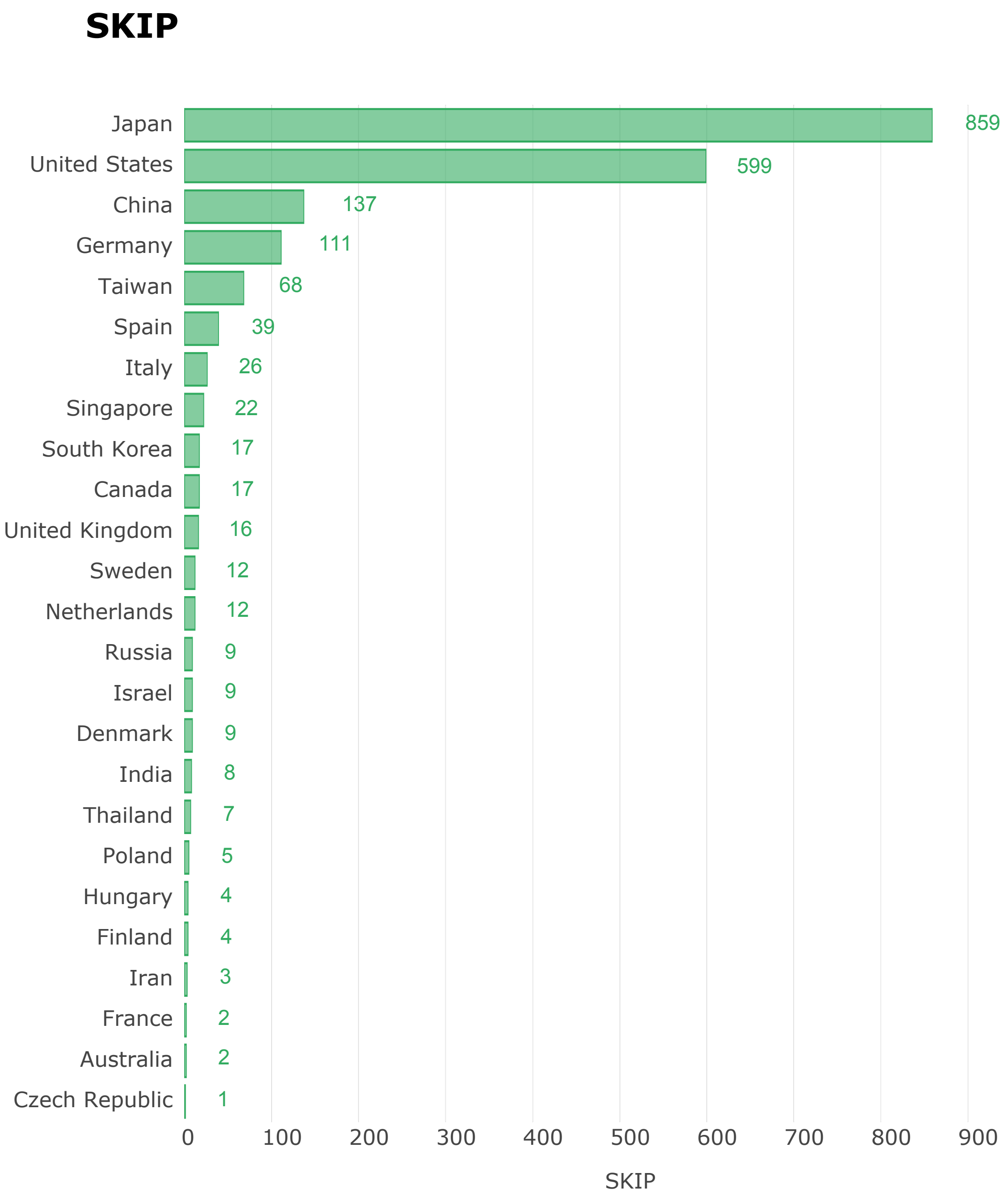

A
eagle-i
Primary fibroblast cell lines
32.4%
iPS cells
67.6%
B
RIKEN BRC
iPS cells
100%
C
SKIP
ES cells
16.5%
iPS cells
82%
D
hPSCreg
ES cells
24.9%
iPS cells
75.1%
iPS cells
Somatic stem cells
Primary fibroblasts
Other cell lines
ES cells
Cancer stem cells

23 **Fig. S2**


hPSCreg
United Kingdom
United States
Germany
China
Spain
Sweden
Denmark
Ireland
Italy
France
South Korea
Russia
Netherlands
Australia
Israel
Belgium
Iran
Taiwan
Czech Republic
Turkey
Finland
Japan
Singapore
Qatar
India
Switzerland
Canada
Thailand
Portugal
Slovenia
Norway
Luxembourg
Hungary
Brazil
Austria
1487
424
172
105
99
98
96
74
54
50
49
48
48
41
32
32
26
19
19
18
18
16
14
13
12
11
6
3
2
1
1
1
1
1
1
0
200
400
600
800
1000
1200
1400
1600
hPSCreg

24

Fig. S3


SKIP
Japan
United States
China
Germany
Taiwan
Spain
Italy
Singapore
South Korea
Canada
United Kingdom
Sweden
Netherlands
Russia
Israel
Denmark
India
Thailand
Poland
Hungary
Finland
Iran
France
Australia
Czech Republic
859
599
137
111
68
39
26
22
17
17
16
12
12
9
9
9
8
7
5
4
4
3
2
2
1
0
100
200
300
400
500
600
700
800
900
SKIP

SUPPLEMENTAL TABLES

Table S1. Related to Figure 5: List of full information of ICSCB as of March 26, 2020.

Table S2. Related to Figure 5A and Figure S1: List of cell line types across all four databases (as of March 26, 2020)**.**

Table S3. Related to Figure 5B: Statistics of healthy/diseased cell lines in ICSCB (as of March 26, 2020)**.**

Table S4. Related to Figure 5C: Statistics of cell line types based on country (as of March 26, 2020)**.**

Table S5. Related to Figure 5D: Statistics of diseased cell lines based on disease category (as of March 26, 2020)**.**

EXPERIMENTAL PROCEDURES

Generation of Fig. 1

Among all the databases, SKIP and hPSCreg provided details of countries from which data were acquired. For SKIP, this information was provided on its homepage (skip.stemcellinformatics.org/en/). For hPSCreg, country information for every cell line could be accessed from its homepage (https://hpscreg.eu/) by clicking "find by location".

Generation of Fig. 5

Full information data were directly downloaded from ICSCB results page (**Table S1**) and filtered according to the following criteria: (A) stem cell type (**Table S2**); (B) health/disease status (**Table S3**); (C) country (**Table S4**); and (D) disease (**Table S5**). Disease categories were determined by search results with keywords under the "Disease Category" in NCBI MeSH page (https://www.ncbi.nlm.nih.gov/mesh). For example, searching with keywords of "Parkinson disease" will lead to the MeSH term "Nervous

System Diseases" under the "Disease Category". Pie charts and bar graphs were produced by R (graph.r) using the package "plotly".

Table S1

Click here to access/download
Supplemental Movies and Spreadsheets
TableS1.200407.xlsx

Table S2

Click here to access/download
Supplemental Movies and Spreadsheets
TableS2.200407.xlsx

Table S3

Click here to access/download

Supplemental Movies and Spreadsheets

TableS3.200407.xlsx

Table S4

Click here to access/download
Supplemental Movies and Spreadsheets
TableS4.200407.xlsx

Table S5

Click here to access/download
Supplemental Movies and Spreadsheets
TableS5.200407.xlsx