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

17 [4] BIH Center for Regenerative Therapies (BCRT), Charité – Universitätsmedizin Berlin,

18 Augustenburger Platz 1, 13353 Berlin, Germany

19 [5] Integrated Bioresource Information Division, RIKEN BioResource Research Center,

20 Tsukuba, Ibaraki 305-0074, Japan

21 [6] Cell Engineering Division, RIKEN BioResource Research Center, Tsukuba, Ibaraki

22 305-0074, Japan

23 [7] Department of Genetics and Human Genetics Institute of New Jersey, Rutgers, The

24 State University of New Jersey, Piscataway, New Jersey 08854, U.S.A.

25 [8] Harvard Catalyst | Clinical and Translational Science Center, Boston, Massachusetts

26 02215, U.S.A.

27 [9] International Stem Cell Banking Initiative, 2 High Street, Barley, Herts, SG88HZ, UK

28 [10] National Stem Cell Resource Centre, Institute of Zoology, Chinese Academy of

29 Sciences, Beijing 100190, China

30 [11] Innovation Academy for Stem Cell and Regeneration, Chinese Academy of Sciences,

31 Beijing 100101, China

32

33 **Additional Footnotes**

34 † Present address: Department of Cellular Biology and Pharmacology, Florida

35 International University, Herbert Wertheim College of Medicine, Florida 33199, U.S.A.

36

37 **Contact**

38 Wataru Fujibuchi*

39 Address: Center for iPS Cell Research and Application (CiRA), Kyoto University, 53

40 Kawahara-cho, Sho-goin, Sakyo-ku, Kyoto 606-8507, Japan

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

42 Tel: +81-75-366-7012

43

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.

58 **INTRODUCTION**

59 Since the first report of human induced pluripotent stem cells (iPSCs) (Takahashi et al.,
60 2007), there has been a rapid rise in the number of iPSC lines and related information
61 worldwide **(Table 1)**. This remarkable growth has not only accelerated studies of
62 regenerative medicine but also provided opportunities to understand such pragmatic
63 issues as the quality of pluripotent stem cells (Nishizawa et al., 2016) and the disease
64 mechanisms (Sasaki et al., 2016). Stem cell banks and registries are expected to provide
65 necessary data of individual stem cell lines. However, the exchange of data among
66 different institutions is not a trivial matter, and scientific reproducibility based on
67 available information is problematic for both basic studies and clinical applications
68 (Yaffe et al., 2016; Isasi and Knoppers, 2011; Thirumala et al., 2009). Moreover, as
69 technologies for the characterization of cell lines continue to progress, the addition of
70 new quality standards as necessary data items is complicating and diversifying data
71 formats among different stem cell banks and registries (Hug, 2009; Knoppers and Isasi,
72 2010). As an attempt to solve these problems, we previously reported MIACARM
73 (Minimum Information About a Cellular Assay for Regenerative Medicine) guidelines in

74 2016 (Sakurai et al., 2016), which proposed the utilization of standardized data items and
75 formats for all stem cell lines in regenerative medicine. Presently, MIACARM contains
76 258 items, covering such areas as stem cell production and materials (e.g., donor
77 information, source cell information, and cell culture medium and substrate information),
78 cell banking processes, cell characterization, sterility testing, and even ethical concerns.
79 Later, a standardized nomenclature for pluripotent stem cells was introduced in 2018 with
80 unification of cell line codification and minimization of information loss and confusion
81 regarding cell lines as goals (Kurtz et al., 2018). Nevertheless, with the growing number
82 of registered cell lines, existing data deposition formats have made it increasingly harder
83 for not only data depositors but also users to seek and obtain cell lines collected under
84 different projects, disease status, and privacy issues (Godard et al., 2003; Winickoff et al.,
85 2009).

86 In this paper, as our next step towards the unification and utilization of stem
87 cell line data in the world, we report our new database portal, Integrated Collection of
88 Stem Cell Bank data (ICSCB), which was designed using MIACARM guideline items
89 and formats. The main objectives of ICSCB are i) to establish an integrated stem cell

90 database portal that can cover the majority of stem cell resources in the world, and ii) to
91 offer users minimum but efficient access to information on stem cell lines based on
92 MIACARM guidelines. Currently, ICSCB provides data of more than 15,000 stem cell
93 lines registered in four major stem cell line databases: hPSCreg (Seltmann et al., 2015),
94 SKIP (Kim et al., 2017), RIKEN BRC (Kobayashi et al., 2016), and eagle-i (Vasilevsky et
95 al., 2012). ICSCB has a user-friendly search engine for stem cell lines and can be
96 accessed directly at http://icscb.stemcellinformatics.org/, or as a slim version by
97 removing cell line redundancy as much as possible through the SHOGoiN (Human
98 Omics Database for the Generation of iPS and Normal Cells) homepage at
99 http://shogoin.stemcellinformatics.org/.
100

101 ## RESULTS AND DISCUSSION

102 ### Web interface

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

106 ICSCB. Sharing cell line information between different stem cell banks and registries is
107 problematic due to different cell naming methods, different policies on cell assessment in
108 different registries, unclear data sources, and so on. ICSCB is a collection of cell lines
109 from four major and reliable cell line data resources based in Europe, Japan, and the
110 United States. ICSCB updating is regularly performed for new SKIP and eagle-i stem cell
111 line data as well as automatically performed for hPSCreg and RIKEN BRC data in a
112 synchronized manner. Users can retrieve all related stem cell line information by using a
113 free text search. Detailed information for a specific cell line can be accessed by clicking
114 on the stem cell ID, which is linked to the information page in the original resources
115 (**Figure 1**). The results can be further filtered according to users' requests. There may be
116 several records for the same cell line if the cell line is included in multiple data resources.
117 To provide users as much information as possible, the results page is designed to show
118 cell lines with matching cell names as well as close descriptions.
119
120 **Data coverage**

121 ICSCB covers more data than any other stem cell line repository available. The
122 integration of all major data resources allows us to check the current state of stem cell
123 research in the world **(Figure 2)**. Although we recognize redundancies in the data,
124 according to our statistics, the number of iPSC lines constitutes more than 80% of all cell
125 lines and the ratio of healthy to diseased donors is approximately 3 to 2 **(Figure 2A,B)**.
126 The total number of countries from which cell lines can be retrieved is 36 (as of March
127 26, 2020), of which the top 9 countries identified in SKIP and hPSCreg are (in descending
128 order) the United Kingdom, United States, Japan, Germany, China, Spain, Sweden,
129 Denmark, and Italy **(Figure 2C)**. In addition, as the recent number of iPSC lines
130 generated from patient donors is growing, ICSCB supports disease-oriented search to
131 help users find all disease-related stem cell lines by using disease names. The distribution
132 of disease and disorder types is shown in **Figure 2D**.
133

134 **Easy search interface on SHOGoiN homepage**

135 ICSCB also has a quick and easy search module on the SHOGoiN homepage
136 (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

153 lack of standardized data formats has caused problems for researchers who must usually
154 search several websites to find the stem cell lines they are looking for (Wells et al., 2013).
155 In the present work, we developed ICSCB, an integrated data distribution system that
156 provides stem cell line information from major stem cell banks and registries all over the
157 world. ICSCB adopts a standardized information format based on the "Source Cell"
158 module of MIACARM to integrate different data resources while keeping important
159 information.

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

167

168 **EXPERIMENTAL PROCEDURES**

169 **Data resources**

170 ICSCB resources were selected from existing major stem cell registries that collect cell
171 line information in Europe, Japan, and the U.S., and stem cell banks that provide cell lines
172 with information of the attributes. We checked the number of registered cell lines and the
173 criteria for registration in these registries and banks to decide to what extent their cell line
174 data can fulfill MIACARM guidelines for inclusion in ICSCB. Considering the size,
175 accessibility, and diversity of the different databases, three stem cell registries and one
176 stem cell bank were included: (1) SKIP (5,615 cell lines), (2) hPSCreg (3,099 cell lines),
177 (3) RIKEN BRC (3,534 cell lines), and (4) eagle-i (3,548 cell lines) (as of March 26,
178 2020). These data resources were selected because they had the highest number of
179 registered cell lines and large diversity, which would provide a good regional balance of
180 cell sources to reduce redundancies in cell line entries. RIKEN BRC basically collected
181 cell lines from Japanese institutions, SKIP contained data mostly from other Japanese and
182 Asian institutions, hPSCreg collected data mainly from European institutions, and eagle-i
183 collected data mostly from the United States. Details of the data sources are listed in
184 **Figure 3**.

185

186 **Data integration**

187 Since our previous research on the listed stem cell banks (Sakurai et al., 2016), the
188 number of registered cell lines had skyrocketed from 1,483 to approximately 8,000 in the
189 past three years. As a result, stem cell registries are facing the demand to collect
190 information on the rapidly increasing number of new cell lines and register the cell lines
191 into their databases as quickly as possible. However, because the stem cell banks and
192 registries are using their own formats for data entry, the integration of the data into a
193 centralized collection system is an extraordinary challenge. To solve this problem, we
194 used a decentralized or distributed database system (Fujibuchi et al., 1998) by adopting
195 items of different database formats into 12 attributes, or terms, from three MIACARM
196 modules: stem cell general identification, donor identification, and source cell
197 identification (**Table 2**). To practically integrate the data from the four data resources
198 (SKIP, hPSCreg, RIKEN BRC, and eagle-i), we adopted a mechanism of cross-reference
199 tables that allow users to conduct a search using MIACARM terms that are translated into
200 the corresponding terms in the individual data resources to implement the search. For

201 example, the term “Stem cell ID” in MIACARM was translated into the terms “stem cell
202 id” (hPSCreg), “stem cell id” (SKIP), “CellID” (RIKEN BRC), and “cell line label”
203 (eagle-i) for the search implementation. Thus, ICSCB submits search requests to each
204 data resource with its own (translated) terms and integrates all retrieved results by
205 common MIACARM terms, thereby achieving a standardized data format at the level of
206 display (**Figure 4**).
207
208 **ICSCB workflow and search engine updating**
209 In order to provide fast and easy access to the latest and accurate cell line information, we
210 built an automatic updating system that adds newly released cell lines to ICSCB as soon
211 as they become available in any of the four data resources. Data from eagle-i and SKIP
212 are directly collected and stored in the MySQL database with the terms required for the
213 MIACARM modules. Data from hPSCreg and RIKEN BRC are collected on the fly per
214 request using a web application programming interface (API) provided by the respective
215 sites. RIKEN BRC also uses SPARQL language for data retrieval requests (Kim et al.,
216 2017; Kobayashi et al., 2016).

217 To simplify the search process, ICSCB provides an easy-to-use and
218 mobile-friendly web application. The goal of the application is to help users find the
219 desired stem cell lines as quickly as possible. The interface of the search engine is
220 designed with the 12 MIACARM terms (**Table 2**) except the term "Stem cell ID". Users
221 receive result pages with all the matching results listed in a table that includes all the basic
222 attributes under the structure of MIACARM. To ensure a more specific search with a
223 wide variety of attributes, ICSCB is designed to accommodate searches not only by
224 standardized terms from MIACARM but also by terms specific to each of the four data
225 resources, such as "age" or "country" (**Figure 5A**). When user queries are submitted,
226 ICSCB simultaneously retrieves MIACARM standardized data and resource-specific
227 data so as not to miss any relevant entries. If a keyword entered by a user in a general
228 keyword search does not exist in MIACARM terms but is included in data specific to any
229 of the four data resources, the user will get detailed descriptions of the matching data in
230 the results page. For example, even if the standardized MIACARM terms do not contain
231 "transgene", it is still possible to enter a gene name into the keyword field (e.g., Sox2)
232 such that the results page will display relevant entries by showing the indicated keyword

233 in the extra field below **(Figure 5B)**. Furthermore, the user can filter the results by data
234 resource and detailed keywords from "Searching options" box inside the results page to
235 narrow down the results list. In addition, all the results can be easily downloaded as a
236 table directly from the results page.

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

243

244

245 **SUPPLEMENTAL INFORMATION**

246 Supplemental Information includes Supplemental Experimental Procedures, three

247 figures, and five tables, and can be found with this article online at:

248

249 **AUTHOR CONTRIBUTIONS**

250 YC and YP drafted the manuscript. SM, TM, HO, JD, SS, AK, HM, YN, MS, JS, and

251 WF provided and facilitated the stem cell data. WF conceptualized the research. AK,

252 GS, and WF led the project.

253

254 **ACKNOWLEDGEMENTS**

255 The authors deeply appreciate Dr. Peter Karagiannis for kindly reviewing the

256 manuscript. This work was partially supported by the Core Center for iPS Cell

257 Research, Research Center Network for Realization of Regenerative Medicine

258 (16bm0104001h0004), and the Formulation of Regenerative Medicine National

259 Consortium which Renders Nation-wide Assistance to Clinical Researches, Project to

260 Build Foundation for Promoting Clinical Research of Regenerative Medicine

261 (19bk0204001h0004, 19bk0204001s0104), Japan Agency for Medical Research and
262 Development, AMED, and German Academic Exchange Service (DAAD) PPP grant.
263

264 **COMPETING INTERESTS**

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

266

267

268 **REFERENCES**

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

273

274 Fujibuchi, W., Goto, S., Migimatsu, H., Uchiyama, I., Ogiwara, A., Akiyama, Y.,
275 Kanehisa, M. (1998). DBGET/LinkDB: an integrated database retrieval system. In Pac.
276 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.

307 Sakurai, K., Kurtz, A., Stacey, G., Sheldon, M., and Fujibuchi, W. (2016). First proposal
308 of minimum information about a cellular assay for regenerative medicine. Stem Cells
309 Transl. Med. *5*, 1345–1361.
310
311 Sasaki, K., Makiyama, T., Yoshida, Y., Wuriyanghai, Y., Kamakura, T., Nishiuchi, S.,
312 Hayano, M., Harita, T., Yamamoto, Y., Hirose, S., et al. (2016). Patient-specific human
313 induced pluripotent stem cell model assessed with electrical pacing validates S107 as a
314 potential therapeutic agent for catecholaminergic polymorphic ventricular tachycardia.
315 PLoS One *11*. e0164795.
316
317 Seltmann, S., Lekschas, F., Müller, R., Stachelscheid, H., Bittner, M.S., Zhang, W.,
318 Luam, K., Anna, S., Anna, V., Stacey, G.N., et al. (2015). hPSCreg—the human
319 pluripotent stem cell registry. Nucleic Acids Res. *44*, D757–D763.
320

321 Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and
322 Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by
323 defined factors. Cell *31*, 861–872.
324
325 Thirumala, S., Goebel, W.S., Woods, E.J. (2009). Clinical grade adult stem cell banking.
326 Organogenesis *5*, 143–154.
327
328 Vasilevsky, N., Johnson, T., Corday, K., Torniai, C., Brush, M., Segerdell, E., Wilson,
329 M., Shaffer, C., Robinson, D., Haendel, M. (2012). Research resources: curating the
330 new eagle-i discovery system. Database *2012*, bar067.
331
332 Wells, C.A., Mosbergen, R., Korn, O., Choi, J., Seidenman, N., Matigian, N.A., Vitale,
333 A.M., Shepherd, J. (2013). Stemformatics: visualisation and sharing of stem cell gene
334 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.

346 **FIGURE LEGENDS**

347 **Fig. 1: Web interface of ICSCB.**

348 (A) The ICSCB search page. Any keyword related to cell lines (including cell line name,

349 disease name, gender, and so on) can be used to perform an instant search. (B) The

350 ICSCB results page. Matched or partially matched cell lines are listed according to

351 MIACARM terms. To check the details of the cell lines, the user can click on the stem cell

352 ID, which is linked to the original source of cell line information.

353

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

355 Cell line information is categorized as (A) stem cell type, (B) health/disease status of

356 donor, (C) country that established the cell lines, and (D) disease category.

357

358 **Fig. 3: Overview of ICSCB.**

359 ICSCB includes data from three stem cell registries and one cell bank in order to

360 maximize data coverage worldwide.

361

362 **Fig. 4: Workflow of ICSCB data integration.**

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

367

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

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

Fig1 Interface

A

Integrated Collection of Stem Cell Bank data by MIACARM

Searching fields (15796 cell lines)

Keyword			
Data source	☑ All ☑ SKIP(5615) ☑ RIKEN BRC (3534) ☑ eagle-i(3548) ☑ hPSCreg (3099) ☐ UKSCB (0) ☐ RUCDR (0)		
Stem cell name	201B7	**Stem cell type**	ES Cell, iPS Cell
Cell grade	Clinical, Research	**Produced by**	Yamanaka
Provider/ distributor	Kyoto University	**Reference publication**	PMID
Gender of donor	Male, Female	**Ethnicity of donor**	African, Asian, Caucasian, Hispanic, Latino
Health status of donor	Parkinson	**Source cell type**	Fibroblast
Organ/tissue of origin of source cell	Skin		
# of rows/page	100	Search	

What's New

2019.12.01 Cell lines from SKIP database has been updated.

2019.10.01 Japanese version of the ICSCB has been released.

B

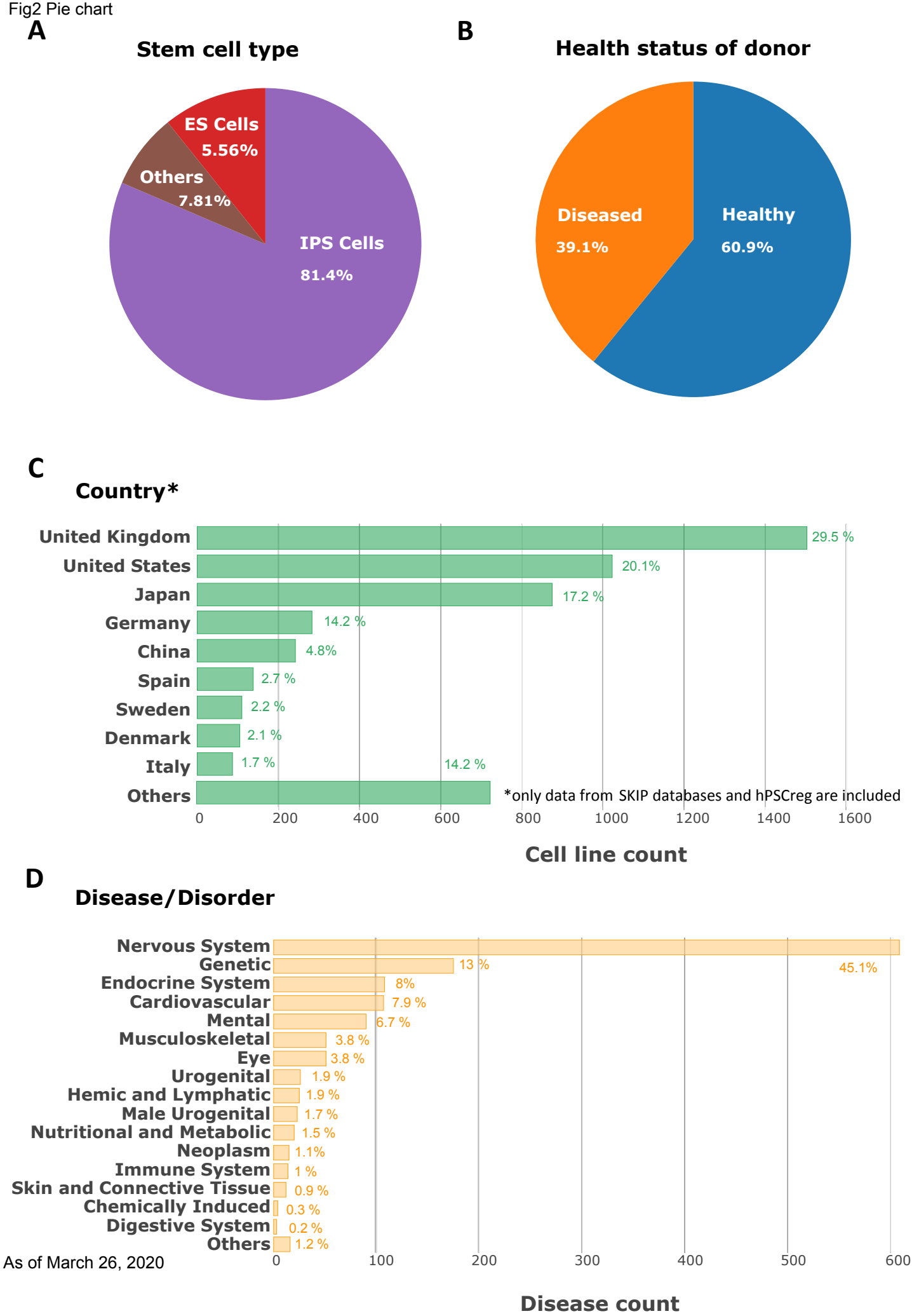
Integrated Collection of Stem Cell Bank data by MIACARM

Items: 1 ~ 100 / 15656

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 of donor	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	18035408 23300777 27073925 27161380	Female	Caucasian			Skin
SKIP	SKIP000002	253G1	iPS Cell		Yamanaka Shinya	Center for iPS Cell Research and Application, Kyoto		Female	Caucasian			Skin
SKIP	SKIP000003	iPS-TIG107 3f1	iPS Cell		Yamanaka Shinya	Center for iPS Cell Research and Application, Kyoto	18035408	Female	Asian			Skin
SKIP	SKIP000004	iPS-TIG107 4f1	iPS Cell		Yamanaka Shinya	Center for iPS Cell Research and Application, Kyoto	18035408	Female	Asian			Skin

As of March 26, 2020

Fig2 Pie chart



As of March 26, 2020

Fig3 Overview

	Database	Registry/Bank(Country)	Countries from which data were acquired
ICSCB	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

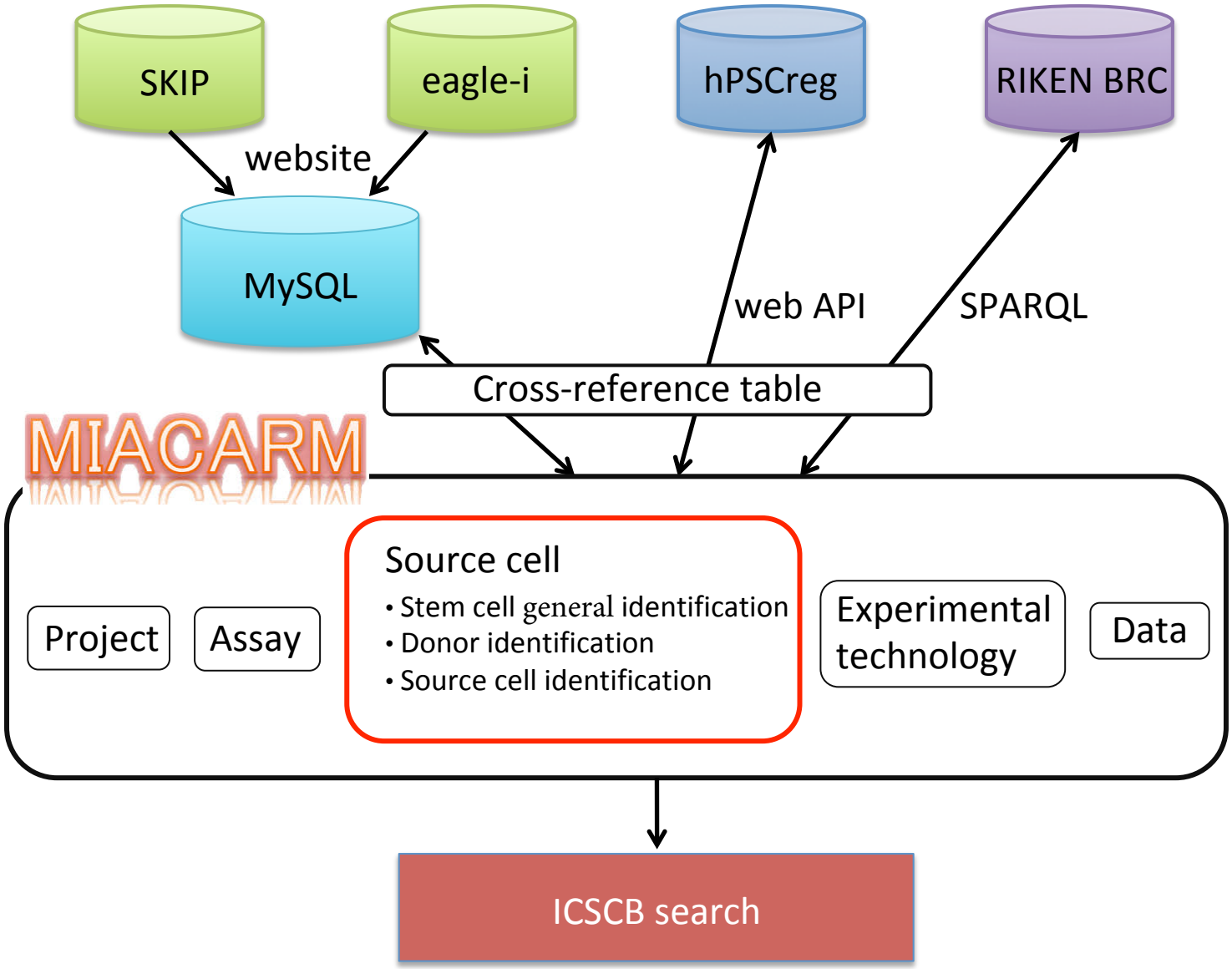


Fig5 Reference table

A

Keyword search

SKIP	eagle-i	hPSCreg	Riken BRC
cell id	cell id	source organism	cell name
stemcell_id	cell line label	race	cell ID
cell line name	cell line provider	cell lineage	description
research grade	sex	health status	original website
establisher name	ethnicity	shogoin cell id	originator
establisher organization	diagnosed disease	stem cell id	depositor
pubmed ID	cell line type	stem cell name	taxon
donor sex	cell line URL	produced by	depiction
donor race		distributor	common name
disease name		publication	cell grouping
cell type		gender	reference
organ/tissue of origin of source cell		race	derived from
in vivo differentiation assay		health status	gender
in vitro differentiation		source cell type	race
cell morphology		source cell description	country
pluripotent marker		origin of source cell	disease
karyotype assay			age
CNV detail			
remaining vector detection test assay			
whole genome detail			
stem cell transcriptome analysis detail			
epigenetics detail			

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

11 **SUPPLEMENTAL FIGURES**

12 **Fig S1. Details of cell line types collected by eagle-i, RIKEN BRC, SKIP, and**

13 **hPSCreg** (as of March 26, 2020)**.** (A) eagle-i, (B) RIKEN BRC, (C) SKIP, and (D)

14 hPSCreg

15

16 **Fig S2**. **Details of countries that have established cell lines in hPSCreg** (as of March

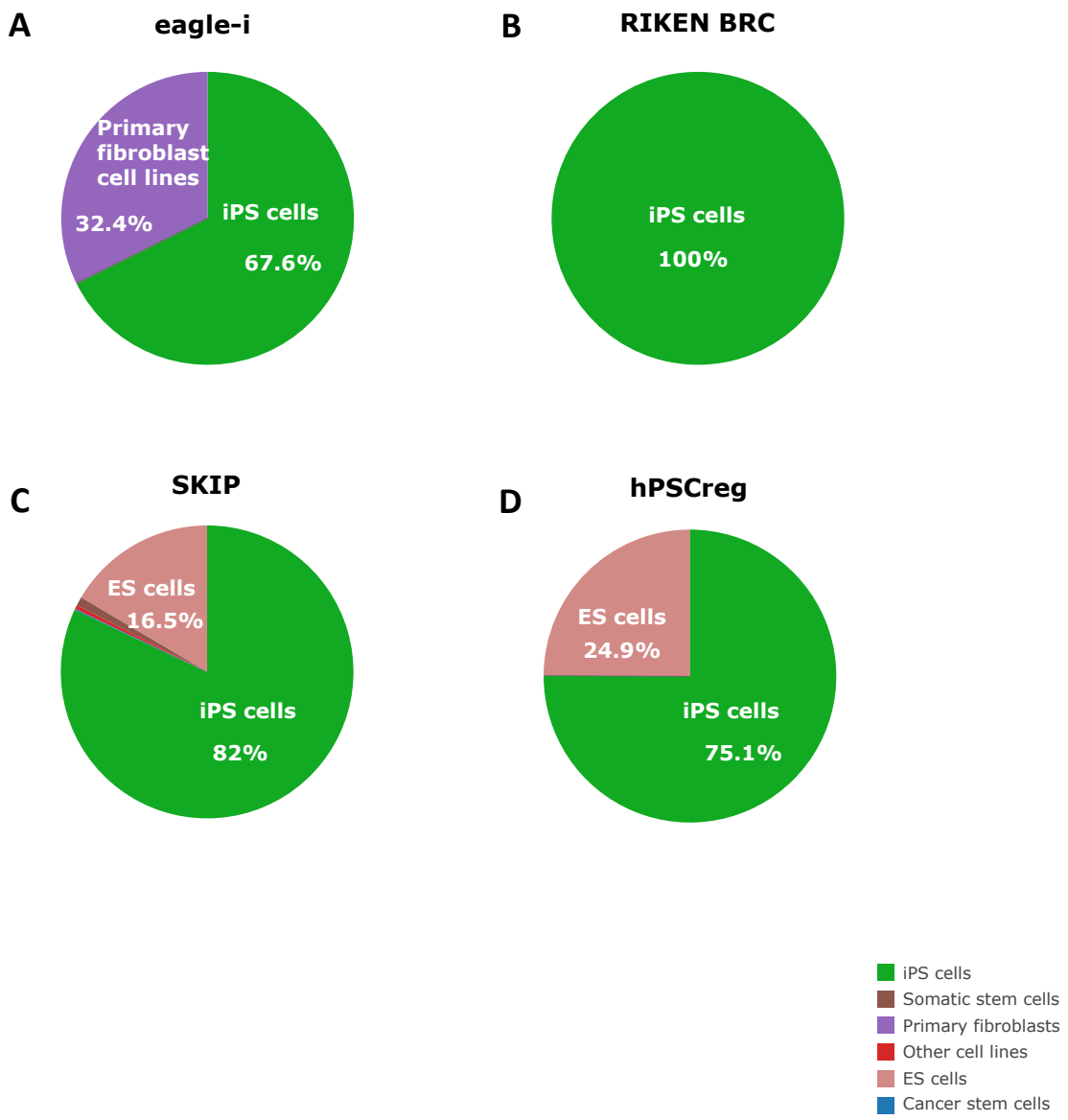
17 26, 2020)**.**

18

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

20 2020)**.**

21 **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

22

Fig. S2

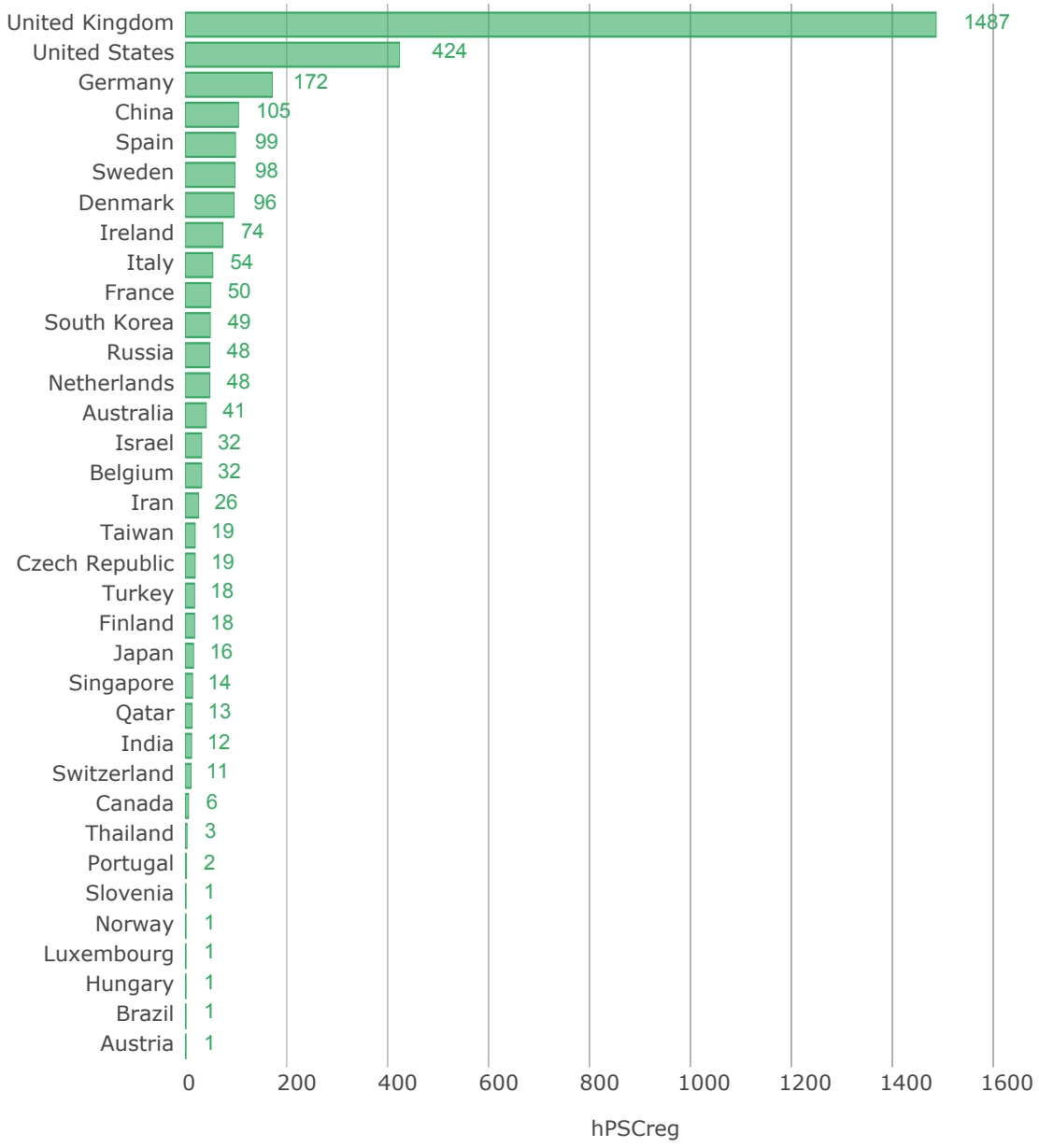
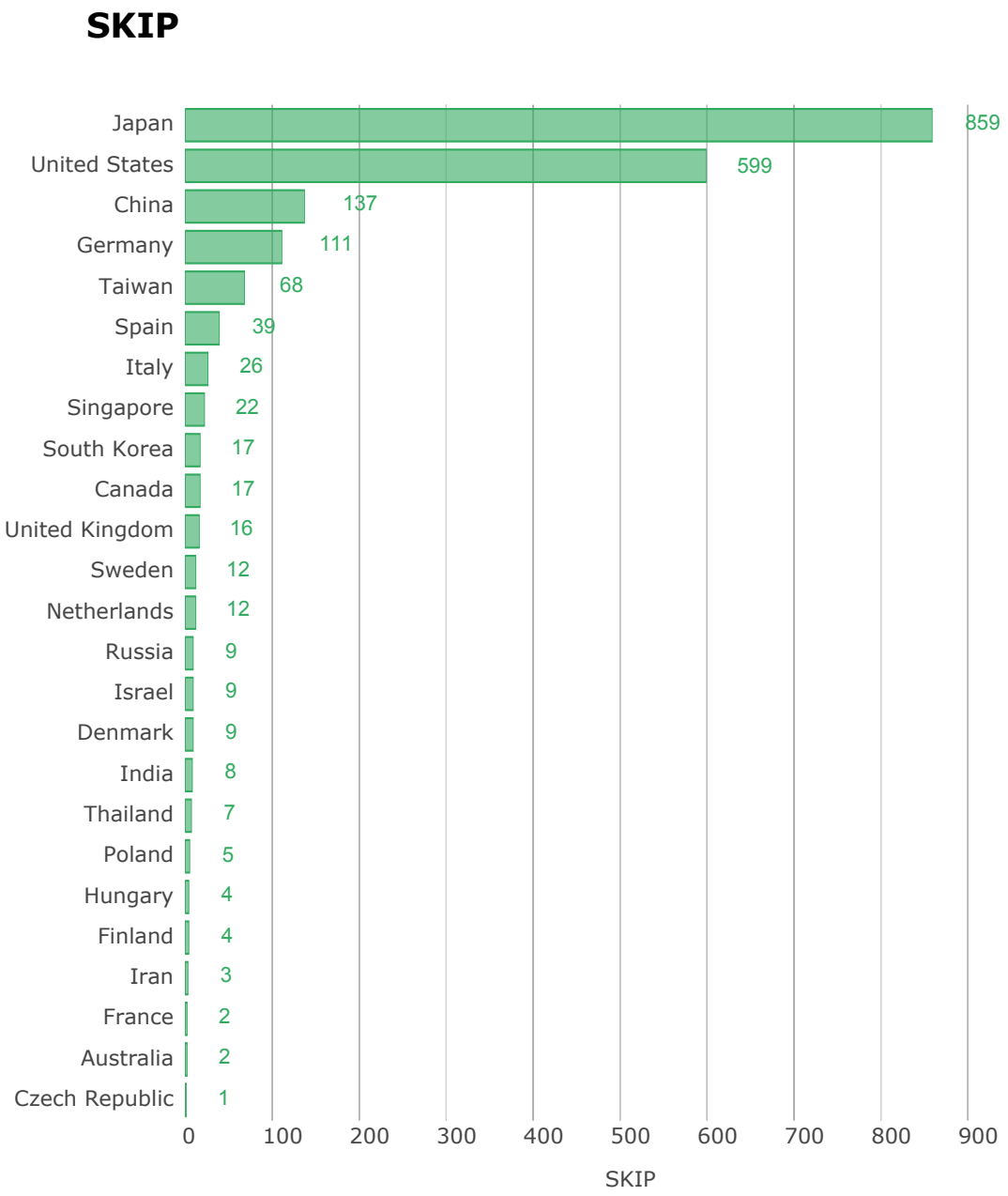

hPSCreg
United Kingdom 1487
United States 424
Germany 172
China 105
Spain 99
Sweden 98
Denmark 96
Ireland 74
Italy 54
France 50
South Korea 49
Russia 48
Netherlands 48
Australia 41
Israel 32
Belgium 32
Iran 26
Taiwan 19
Czech Republic 19
Turkey 18
Finland 18
Japan 16
Singapore 14
Qatar 13
India 12
Switzerland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

Fig. S3


SKIP
Japan 859
United States 599
China 137
Germany 111
Taiwan 68
Spain 39
Italy 26
Singapore 22
South Korea 17
Canada 17
United Kingdom 16
Sweden 12
Netherlands 12
Russia 9
Israel 9
Denmark 9
India 8
Thailand 7
Poland 5
Hungary 4
Finland 4
Iran 3
France 2
Australia 2
Czech Republic 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)**.**

42 **EXPERIMENTAL PROCEDURES**

43 **Generation of Fig. 1**

44 Among all the databases, SKIP and hPSCreg provided details of countries from which

45 data were acquired. For SKIP, this information was provided on its homepage

46 (skip.stemcellinformatics.org/en/). For hPSCreg, country information for every cell line

47 could be accessed from its homepage (https://hpscreg.eu/) by clicking "find by

48 location".

49

50 **Generation of Fig. 5**

51 Full information data were directly downloaded from ICSCB results page (**Table S1**)

52 and filtered according to the following criteria: (A) stem cell type (**Table S2**); (B)

53 health/disease status (**Table S3**); (C) country (**Table S4**); and (D) disease (**Table S5**).

54 Disease categories were determined by search results with keywords under the "Disease

55 Category" in NCBI MeSH page (https://www.ncbi.nlm.nih.gov/mesh). For example,

56 searching with keywords of "Parkinson disease" will lead to the MeSH term "Nervous

57 System Diseases" under the "Disease Category". Pie charts and bar graphs were

58 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