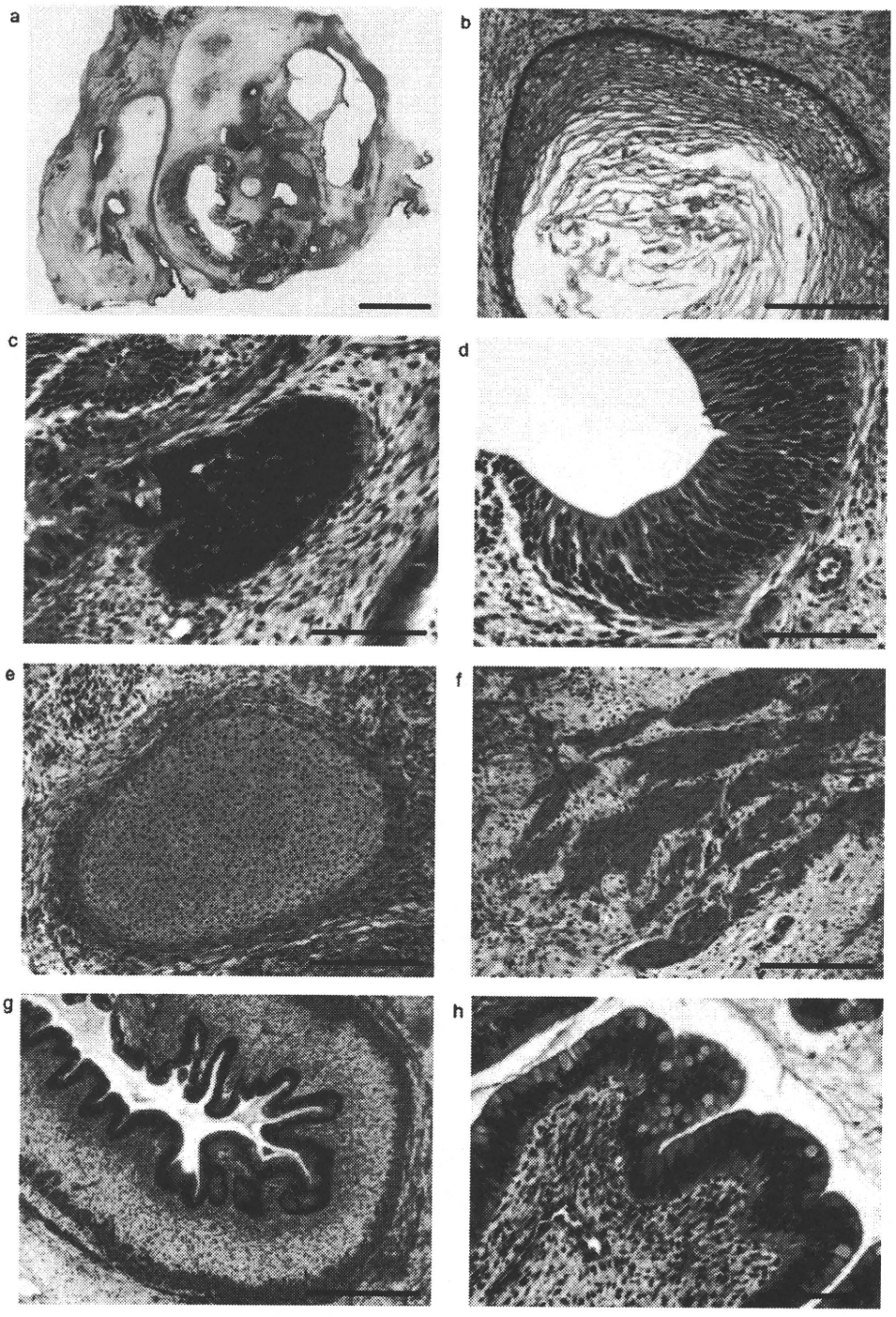


Figure 2 Teratomas derived from human induced pluripotent stem (iPS) cells. Cells from clone HiPS-RIKEN-1A were transplanted into the sub-capsular space of the testis of immunodeficient mice. A hematoxylin and eosin stained section from a resulting tumor is shown. (a) Whole teratoma. (b) Epithelial tissue. (c) Pigmented cells. (d) Neural epithelium. (e) Cartilage. (f) Muscle. (g) Intestinal tract. (h) Intestinal epithelium.

the mother. Thus, the iPS cell lines produced from HFM-1 were derived from both neonatal and maternal cells (Table 1).

DISCUSSION

In contrast to iPS cell lines produced using inbred mouse strains, human iPS cell lines are derived from individuals who possess different genetic backgrounds. This factor needs to be taken into account when considering differences in characteristics among human iPS cell lines. Additionally, it should be noted that the human cells used to generate iPS cell lines may have acquired genetic mutations and these might also influence the characteristics of the cell lines.

Acquired mutations are of interest, for example in an oncology laboratory, when iPS cell lines are derived from a patient's cells to analyze aspects of their pathology. By contrast, acquired genetic mutations are an obstacle to the clinical use of iPS cell lines for regenerative medicine. It will therefore be necessary to establish validated "clinical grade iPS cell lines", possibly using cells with a low risk of carrying acquired genetic mutations, for example, neonatal tissue cells.

However, the use of neonatal tissue cells is not free of potential problems. Thus, for example, we cannot exclude the possibility of late-onset diseases that may occur in the neonate's future. The development of the technology for whole genome sequencing may resolve this issue by enabling the prior estimation of the risk of disease.

ACKNOWLEDGMENTS

This work was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology in Japan.

REFERENCES

1 Takahashi K, Tanabe K, Ohnuki M *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* 2007; **131**: 861–72.
2 Wernig M, Meissner A, Foreman R *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* 2007; **448**: 318–24.
3 Yu J, Vodyanik MA, Smuga-Otto K *et al.* Induced pluripotent stem cell lines derived from human somatic cells. *Science* 2007; **318**: 1917–20.
4 Park IH, Zhao R, West JA *et al.* Reprogramming of human somatic cells to pluripotency with defined factors. *Nature* 2008; **451**: 141–6.
5 Okita K, Nakagawa M, Hyenjong H, Ichisaka T, Yamanaka S. Generation of mouse induced pluripotent stem cells without viral vectors. *Science* 2008; **322**: 949–53.
6 Fusaki N, Ban H, Nishiyama A, Saeki K, Hasegawa M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci* 2009; **85**: 348–62.
7 Yu J, Hu K, Smuga-Otto K *et al.* Human induced pluripotent stem cells free of vector and transgene sequences. *Science* 2009; **324**: 797–801.
8 Hiroyama T, Sudo K, Aoki N *et al.* Human umbilical cord-derived cells can often serve as feeder cells to maintain primate embryonic stem cells in a state capable of producing hematopoietic cells. *Cell Biol Int* 2008; **32**: 1–7.
9 Suemori H, Yasuchika K, Hasegawa K, Fujioka T, Tsuneyoshi N, Nakatsuji N. Efficient establishment of human embryonic stem cell lines and long term maintenance with stable karyotype by enzymatic bulk passage. *Biochem Biophys Res Commun* 2006; **345**: 926–32.
10 Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 2006; **126**: 663–76.
11 Yoshino K, Iimura E, Saijo K *et al.* Essential role for gene profiling analysis in the authentication of human cell lines. *Hum Cell* 2006; **19**: 43–8.

Full Report

ibc Interdisciplinary Bio Central

Open Access, Open Review Journal

Development of a Simple Method to Determine the Mouse Strain from which Cultured Cell Lines Originated

Kaori Yoshino[1], Kaoru Saijo[1], Chikako Noro[2] and Yukio Nakamura[1,*]

[1]Cell Engineering Division, RIKEN BioResource Center, Tsukuba, Ibaraki 305-0074, Japan
[2]Nihon University School of Medicine, Itabashi-ku, Tokyo 173-8610, Japan

Subject areas: Resources

Author contribution: K.Y., K.S. and C.N. performed experiments and analyzed data; Y.N. analyzed data and wrote the manuscript.

***Correspondence** and requests for materials should be addressed to Y.N. (yukionak@brc.riken.jp).

Reviewer: Atsushi Yoshiki, RIKEN BioResource Center, Japan; Arihiro Kohara, National Institute of Biomedical Innovation, Japan

Editor: Yeonhee Lee, Seoul Women's University, Republic of Korea

Received December 14, 2010;
Accepted December 17, 2010;
Published December 29, 2010

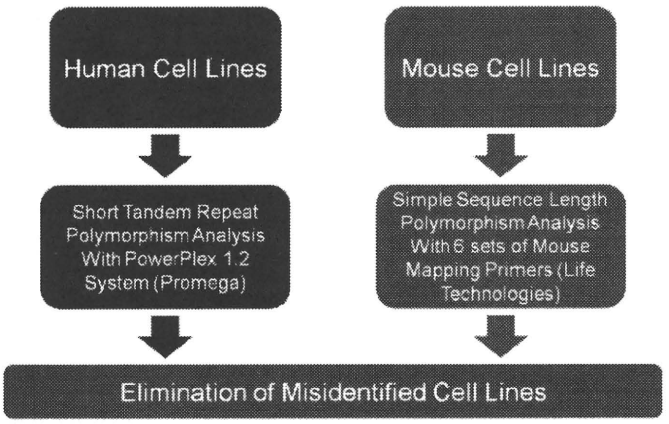
Citation: Nakamura, Y., et al. Development of a Simple Method to Determine the Mouse Strain from which Cultured Cell Lines Originated. IBC 2010, 2:14, 1-6.
doi: 10.4051/ibc.2010.2.4.0014

Funding: This work was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

Competing interest: All authors declare no financial or personal conflict that could inappropriately bias their experiments or writing.

SYNOPSIS

Misidentification of cultured cell lines results in the generation of erroneous scientific data. Hence, it is very important to identify and eliminate cell lines with a different origin from that being claimed. Various methods, such as karyotyping and isozyme analysis, can be used to detect inter-species misidentification. However, these methods have proved of little value for identifying intra-species misidentification, and it will only be through the development and application of molecular biological approaches that this will become practical. Recently, the profiling of microsatellite variants has been validated as a means of detecting gene polymorphisms and has proved to be a simple and reliable method for identifying individual cell lines. Currently, the human cell lines provided by cell banks around the world are routinely authenticated by microsatellite polymorphism profiling. Unfortunately, this practice has not been widely adopted for mouse cells lines. Here we show that the profiling of microsatellite variants can be also applied to distinguish the commonly used mouse inbred strains and to determine the strain of origin of cultured cell lines. We found that approximately 4.2% of mouse cell lines have been misidentified; this is a similar rate of misidentification as detected in human cell lines. Although this approach cannot detect intra-strain misidentification, the profiling of microsatellite variants should be routinely carried out for all mouse cell lines to eliminate inter-strain misidentification.



An SSLP analysis using the 6 MIT markers described in this study was sufficient to distinguish the common and popular inbred mouse strains such as C57BL/6, BALB/c, C3H/He, 129/Sv and DBA/2. Although this approach cannot detect intra-strain misidentification, it should be routinely carried out for all mouse cell lines to eliminate inter-strain misidentification.

Keywords: cell bank, cross-contamination, microsatellite polymorphism, misidentification, quality control, short tandem repeat polymorphism, simple sequence length polymorphism

Introduction

Cultured cell lines have proved a valuable resource in all fields of the life sciences and have been utilized in many types of biological study. Currently, however, stringent analyses to check the identity of a cell culture are not always included as part of the culture protocol routine. This has led to misidentification or cross-contamination of cell lines going undetected. As a result, the published literature contains a number of reports that are based on wrongly identified cell lines[1]. Despite numerous publications warning of inter- and intra-species misidentification of cell lines[2-6], the problem of misidentification continues to occur at an extremely high rate[7-10]. Therefore, articles pointing out misidentification of cell lines continue to be published[11-14].

Inter-species contamination can be detected by various methods, such as karyotyping and isozyme analysis. However, it was not possible to detect intra-species misidentification prior to the development of molecular biology techniques that make use of the genetic differences between cell lines to facilitate their identification. One such method makes use of microsatellite polymorphisms to develop diagnostic profiles for cell lines[15]. Microsatellite polymorphisms result from differences in the numbers of a repeating unit of 1-7 base pairs; these variants are also called short tandem repeat (STR) polymorphisms or simple sequence length polymorphisms (SSLPs). These polymorphisms have been extensively used in forensic science. Gene profiling using STR polymorphisms (STR profiling) has been shown to be an efficient and reliable means for identifying individual human cell lines[16,17] and is now performed routinely in the major cell banks around the world. What about mouse cell lines; do they also suffer from problems of misidentification? This question prompted us to establish a method to authenticate the identities of mouse cell lines.

Results and Discussion

The mouse genome possesses a huge number of microsatellite polymorphisms, similarly to the genomes of humans and other mammalian species. The so-called MIT markers developed at the Massachusetts Institute of Technology (MIT) Whitehead Institute[18] in the mouse have been extensively developed and utilized in various fields of research. Information on microsatellite polymorphisms, including data on MIT marker sizes in 47 mouse strains, is publicly available on the Center for Inherited Disease Research Web site (http://www.cidr.jhmi.edu/mouse/mouse_resources.html). It is possible to amplify a number of polymorphic microsatellite loci using commercially available sets of primers. The PCR products are analyzed simultaneously with size standards using automated fluorescent detection techniques. The result is a simple numerical code that corresponds to the lengths of the PCR products amplified at each locus, and is accurate to less than one base pair.

Our aim was to establish a simple method that can be used for routine analysis in our cell bank work. Thus, we sought to establish a method using the smallest practical number of polymorphic loci. First, we screened more than 500 microsatellite primers covering the autosomes and the X chromosome and selected the 24 MIT markers that exhibited the most distinct differences between inbred mouse strains (Figure 1, Table 1), mostly larger than 10 bp on electrophoretic patterns[19]. We then performed an SSLP analysis of 40 mouse strains using the 24 selected MIT markers (Table 1) and concluded that 6 MIT markers would be sufficient to distinguish the common and popular inbred strains such as C57BL/6, BALB/c, C3H/He, 129/Sv and DBA/2 (Table 2).

An SSLP analysis, using the 6 MIT markers described above, has now been adopted by the Cell Engineering Division of the RIKEN BioResource Center (RIKEN Cell Bank) to exclude misidentification among the cultured mouse cell lines that we currently provide. However, this analysis can detect inter-strain but not intra-strain misidentification.

The screening of the RIKEN Cell Bank indicated that 97.7% (334 lines out of 342 lines) of the mouse cell lines were derived from

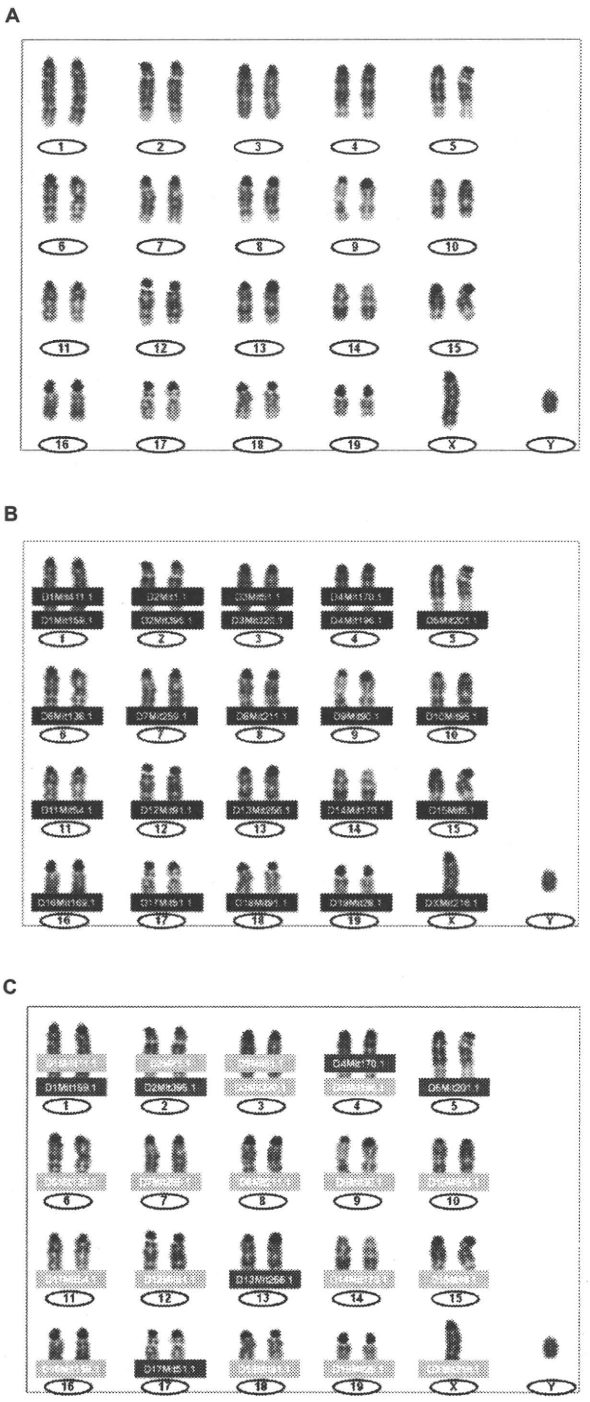


Figure 1. Distribution of the analyzed simple sequence length polymorphisms (SSLPs) loci in the mouse genome. (A) Karyotype of the mouse, 40XY. (B) The chromosomal distributions of the screened 24 loci are indicated. (C) The 6 loci selected as being sufficient for identification of common inbred mouse strains are indicated by the red background.

Table 1. Results of the SSLP analyses of 40 inbred mouse strains using the 24 loci indicated in Figure 1B

Strain Name	Origin	Mit159_1	Mit411_1	Mit1_1	Mit395_1	Mit51_1	Mit320_1	Mit170_1	Mit196_1	Mit201_1	Mit138_1	Mit259_1	Mit211_1	Mit90_1	Mit95_1	Mit54_1	Mit91_1	Mit256_1	Mit170_1	Mit6_1	Mit169_1	Mit51_1	Mit91_1	Mit28_1	Mit216_1
C57BL/6JMs	Ms	203	104	217	130	240	116	228	104	98	192	241	150	110	201	149	153	100	203	134	234	157	132	153	126
C57BL/6J	JAX	203	104	217	129	240	116	228	104	98	192	241	150	110	201	149	153	100	203	134	234	157	132	153	126
C57BL/6J Jcl	Jcl	203	104	217	129	240	116	228	104	98	192	241	150	110	201	149	153	100	203	134	234	157	132	153	126
C57BL/6N Jcl	Jcl	203	104	213	129	240	116	228	104	98	192	241	150	106	201	149	153	100	203	134	234	157	132	153	126
C57BL/6N Crl	Crl	203	104	213	129	240	116	228	104	98	192	241	150	101	201	149	153	100	203	134	234	157	132	153	126
C57BL/10SnJ	Ms	203	104	213	130	240	116	228	104	98	196	241	160	106	203	149	153	100	203	134	234	157	132	155	126
C57L/J	Ms	142	97	213	130	240	116	228	104	98	192	241	150	106	181	149	153	100	188	134	234	163	130	157	126
C58/J	Ms	142	104	213	123	240	116	228	104	98	211	241	160	106	181	149	142	100	188	134	234	153	132	157	126
DBA/J	Ms	185	97	213	123	240	101	236	104	92	211	246	162	98	181	119	151	76	203	136	238	140	132	162	126
DBA/1J	Ms	142	97	213	135	236	101	238	104	92	207	246	160	108	181	144	153	76	202	136	234	153	132	162	126
DBA/2J	JAX	142	97	213	135	236	105	242	104	92	207	246	160	108	181	144	153	76	192	136	234	155	132	162	126
BALB/cJ	JAX	142	97	213	135	254	103	242	116	94	211	225	162	106	181	147	151	88	206	134	234	155	130	162	132
BALB/cByJ	Ms	142	97	213	135	254	103	242	116	94	211	225	164	106	181	147	151	88	204	134	234	155	130	162	132
BKS/J	Ms	203	97	213	161	236	103	236	116	104	211	225	164	106	183	119	153	76	206	134	234	140	132	162	126
C3H/HeJ	JAX	185	104	213	124	240	103	236	116	92	215	246	160	106	181	146	151	76	206	127	232	140	130	162	126
C3H/HeNCrl	Crl	185	104	213	124	240	103	236	116	92	211	246	160	106	181	144	151	78	206	127	232	140	130	162	126
C/J	JAX	185	104	213	135	240	103	226	116	106	211	246	164	106	181	119	151	88	202	136	234	153	134	153	132
C/WySnJ	Ms	185	104	213	135	240	103	226	116	106	211	246	164	106	181	119	151	88	202	136	234	153	132	153	132
129/SvJ	Ms	191	104	215	157	246	103	228	116	94	211	239	162	106	181	147	142	76	202	108	234	163	132	153	132
BMJ	Ms	193	104	209	135	254	103	236	104	98	211	239	160	106	181	147	151	76	206	136	234	157	132	157	126
CWB/J	Jcl	191	104	224	161	254	108	234	104	102	211	244	160	112	181	149	150	88	200	134	260	142	124	162	126
DJ/J	JAX	203	104	219	161	254	108	242	104	106	211	246	160	108	181	147	149	88	202	134	234	153	130	153	126
DJ/Jms	Ms	203	104	219	161	254	108	242	104	106	211	246	160	108	181	147	149	88	202	134	234	153	130	153	126
DWB/J	Ms	203	97	213	161	256	103	232	104	104	215	244	162	106	181	149	150	92	203	131	262	142	130	162	126
EJ/J	Ms	185	97	215	161	236	103	242	116	106	215	246	164	98	181	151	145	76	192	131	232	140	132	162	132
ORS/A	AlchiCC	203	86	217	123	236	103	242	116	104	211	246	160	106	181	149	151	76	203	131	262	167	130	162	126
[illegible]	Nga	191	97	213	161	254	106	242	116	104	203	246	160	97	181	149	151	88	209	136	240	153	126	153	132
[illegible]	Nga	191	97	213	135	254	103	242	116	92	207	225	160	106	181	147	151	106	203	131	240	157	130	153	132
[illegible]	Stm	185	97	213	123	236	103	232	104	98	207	244	160	106	181	147	151	100	203	131	262	153	124	153	132
JF1	Ms	142	97	226	129	326	99	226	104	92	344	217	160	128	201	119	114	100	220	108	278	170	130	155	135
MOLF/Ei	Ms	168	111	229	138	232	91	247	106	102	236	227	182	107	187	135	114	90	174	108	311	165	142	160	136
NOD/Ms	Ms	144	97	230	140	288	105	228	106	108	249	219	156	112	201	123	114	92	204	108	286	172	227	176	118
NMRI/Ms	Ms	166	97	228	100	238	91	249	104	98	221	256	191	121	185	173	114	86	174	117	311	220	147	160	139
NZB/Ms	Ms	142	97	228	142		87	226	110	108	284	217	175	124	201	121	114	98	204	108	236	149	120	172	133
PL/MJS-BJL	Ms	176	106	224	142	230	91	226	106	100	280	217	141	84	197	139	114	90	197	106	266	151	138	172	121
SM/Ms	Ms	142	97	222	129	313	83	226	104	96	332	217	168	124	204	121	114	92	218	108	278	170	120	155	140
M.spretus	RPCI	176	97	179	122	252	132	279	91	114	247	221	160	88	134	157	118	90	201	110	205	134	128	178	134
SPRET/EI	JAX	180	97	190	124		106.5	269	91	114	266	213	185	95		151	116	90	205	110	204	134	115	170	139
WK	RPCI	142	111	217	142	281	91	242	106	102	280	217	164	88	199	138	114	100	203	106	274	164	132	185	123
ZRS/Ms	Ms	139	97	196	114		93	253	83	120	322	229	152	117		150	116	67	93	126		156	114	164	92

In each column (at each polymorphic locus), fragments of similar length are indicated by the same color.

Table 2. SSLP analysis of 5 common and popular mouse strains using the 6 selected loci indicated in Figure 1C

Sample	D1 Mit159.1	D2 Mit395.1	D4 Mit170.1	D5 Mit201.1	D13 Mit256.1	D17 Mit51.1
C57BL/6	203.2	129.7	226.0	98.9	100.3	157.3
BALB/c	141.8	135.5	242.5	94.9	88.3	155.1
C3H/He	185.1	123.8	236.4	92.5	78.4	140.1
129/Sv	191.5	157.3	226.1	94.8	76.6	163.7
DBA/2	141.8	135.3	242.6	92.6	76.5	155.2

the common inbred strains (e.g., Table 3); the remaining 2.3% appeared to be derived from non-inbred mice. Cell lines derived from an F1 hybrid of two different strains show different alleles at each locus that correspond with those of their parental strains (Table 4). With regard to the cell line UV.CC3-11.1 (RCB2074), which was established from an F1 hybrid from the cross BALB/c x C3H/He, only the BALB/c allele at the D5Mit201.1 locus could be detected (Table 4). The C3H/He allele has been replaced by the BALB/c allele during culture, i.e., so-called loss of heterozygosity (LOH) appears to have occurred.

For some cell lines of uncertain provenance, we were able to identify the originating strains by SSLP analysis. For example, we found that LLC[20], PU5-18[21], and MBT-2[22] were derived from the C57BL/6, BALB/c, and C3H/He strains, respectively (Table 5). Interestingly, the LLC cell line had a deletion mutation at the locus D13Mit256.1 and only this mutated allele was detectable. Presumably, LOH had occurred subsequent to the deletion mutation.

Approximately 4.2% (14 lines out of 334 lines) of the mouse cell lines derived from common inbred strains were misidentified (Table 6), i.e., the strains were different from those claimed by the depositors of the cell lines. This rate of misidentification is similar to that reported for human cell lines[17]. As an example of misidentification, the cell line TSt-4 was registered as being C57BL/6-derived; however, SSLP analysis indicated that the cell line was derived from the BALB/c strain (Table 7). LOH following a deletion mutation was detected at the D17Mit51.1 locus of the TSt-4 cell line. Similarly, LOH following a deletion mutation was also detected in the MC3T3-E1[23] cell line at locus D1Mit159.2 (Table 8). In general, it is inevitable that aberrations such as point mutations, deletion mutations and LOHs accumulate in cell lines following long term culture. Thus, the optimum strategy is to culture the cell for as short period as possible not only in the cell bank but also for ordinary laboratory work.

Table 3. SSLP analysis of three C57BL/6-derived cell lines, B6mt-2, MEDEP-BRC5, and UV.B6-4.1 and three BALB/c-derived cell lines, RAW264, J774.1, and UV.BAL-7.1

Sample	D1 Mit159.1	D2 Mit395.1	D4 Mit170.1	D5 Mit201.1	D13 Mit256.1	D17 Mit51.1
C57BL/6	203.2	129.7	226.0	98.9	100.3	157.3
B6mt-2	203.2	129.8	226.0	98.7	100.3	157.0
MEDEP-BRC5	203.2	130.0	226.0	98.9	100.6	157.0
UV.B6-4.1	203.6	130.1	226.0	98.9	100.8	157.4
BALB/c	141.8	135.5	242.5	94.9	88.3	155.1
RAW264	141.8	135.3	242.5	94.6	88.3	155.0
J774.1	142.2	135.7	242.6	94.4	88.5	154.8
UV.BAL-7.1	141.8	135.3	242.4	94.6	88.2	155.1

The analyses authenticated the origins of the six cell lines.

Table 4. SSLP analysis of a mouse cell line derived from a (BALB/c x C3H/He) F_1 Mouse

Sample	D1 Mit159.1	D2 Mit395.1	D4 Mit170.1	D5 Mit201.1	D13 Mit256.1	D17 Mit51.1
BALB/c	141.8	135.5	242.5	94.9	88.3	155.1
C3H/He	185.1	123.8	236.4	92.5	78.4	140.1
UV.CC3 -11.1	142.0 185.2	135.5 123.9	242.6 236.4	94.8	88.4 78.5	155.2 140.2

At locus D5Mit201.1 only the allele corresponding to that of BALB/c was detected.

Table 5. SSLP analysis of the LLC, PU5-18, and MBT-2 mouse cell line

Sample	D1 Mit159.1	D2 Mit395.1	D4 Mit170.1	D5 Mit201.1	D13 Mit256.1	D17 Mit51.1
C57BL/6	203.2	129.7	226.0	98.9	100.3	157.3
LLC	203.4	129.7	226.1	98.4	98.5	157.0
BALB/c	141.8	135.5	242.5	94.9	88.3	155.1
PU5-18	141.5	135.0	242.5	94.2	87.6	155.0
C3H/He	185.1	123.8	236.4	92.5	78.4	140.1
MBT-2	184.7	123.2	236.3	92.2	77.8	139.8

The mouse strains from which the cell lines originated were not registered by the depositors. The SSLP analysis indicated that they were derived from C57BL/6, BALB/c, and C3H/He, respectively.

Table 6. List of cell lines that the strains were different from those claimed by the depositors of the cell lines

RCB No.	Cell Name	Registered Strain	Result	Comment
RCB0792	T88-M	DBA/2J	C3H/He	Providing
RCB1144	DA-3	BALB/c	DBA/2	Providing
RCB2116	TSt-4	C57BL/6	BALB/c	Providing
RCB2117	TSt-4/G	C57BL/6	BALB/c	Providing
RCB2118	TSt-4/G-DLL1	C57BL/6	BALB/c	Providing
RCB2119	TSt-4/N	C57BL/6	BALB/c	Providing
RCB2120	TSt-4/N-DLL1	C57BL/6	BALB/c	Providing
RCB2633	MM46 CEA-2	C3H/He	Unknow	Providing
RCB2634	MM46-APR-MUC1 cl.1	C3H	Swiss	Providing
RCB2195	FVB-2	Swiss FVB	129/Sv	Stopped provision
RCB2196	ICRmt-1	ICR	Unknow	Stopped provision
RCB2617	MM46	C3H	Swiss	Stopped provision
RCB2632	BALB/3T3aP R-MUC1 clone 16	BALB/c	Swiss	Stopped provision
RCB2647	BALB/3T3	BALB/c	Swiss	Stopped provision

Table 7. Misidentification of a cell line identified using SSLP analysis

Sample	D1 Mit159.1	D2 Mit395.1	D4 Mit170.1	D5 Mit201.1	D13 Mit256.1	D17 Mit51.1
C57BL/6	203.2	129.7	226.0	98.9	100.3	157.3
TSt-4	142.0	135.5	242.7	94.5	88.4	152.3
BALB/c	141.8	135.5	242.5	94.9	88.3	155.1

The mouse cell line TSt-4 was registered as a C57BL/6-derived cell line, but our SSLP analysis showed it was derived from the BALB/c strain.

Table 8. SSLP analysis of the mouse cell line MC3T3-E1

Sample	D1 Mit159.1	D2 Mit395.1	D4 Mit170.1	D5 Mit201.1	D13 Mit256.1	D17 Mit51.1
C57BL/6	203.2	129.7	226.0	98.9	100.3	157.3
MC3T3-E1	199.4	139.6	255.9	98.4	100.4	157.0

Loss of heterozygosity following a deletion mutation appears to have occurred at locus D1Mit59.1.

The C57BL/6N and C57BL/6J substrains were shown to have 11 single nucleotide polymorphisms (SNPs) in a study using mouse MD Linkage Panel 1449 SNPs (Illumina)[24]. It is therefore possible to distinguish cell lines derived from the C57BL/6N and C57BL/6J substrains using these 11 SNP.

The relatively high rate (4.2%) of misidentification of cell lines derived from common inbred mouse strains strongly suggests that intra-strain misidentification is also likely to have occurred. For example, although a cell line may be registered as being derived from colon cancer cells, an error may have occurred and the cell line was actually derived from another cancer (a similar phenomenon also applies to human cell lines). It is impossible to identify the originating tissue by microsatellite polymorphism analysis. In this context, a profiling analysis based on gene expression using many cell lines may be useful for authenticating the originating tissues of cultured cell lines.

Conclusion and Prospects

We have established a simple and reliable method to identify the common inbred mouse strains from which cultured mouse cell lines are derived. With respect to intra-strain misidentification, such as errors regarding the originating tissue type, it will be necessary to develop other analytic techniques, for example, gene expression profiling analysis. Other types of OMICS analysis, such as whole genome sequencing, will also be useful for authentication of cell lines. In this context, bioinformatics will become increasingly important for the quality control of cultured cell lines.

Materials and Methods

Mouse cell lines

All mouse cell lines that the Cell Engineering Division of the RIKEN BioResource Center (http://www.brc.riken.jp/lab/cell/english/) has collected, 342 cell lines in total, were subjected to SSLP analysis. We selected the latest preserved stock cells and those that were preserved immediately after deposition (token stock cells) for this analysis.

DNA preparation

DNA was prepared from approximately 2×10^6 cells using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany).

SSLP analysis

Multiplex PCR reactions for the SSLP analysis were carried out using the following fluorescent dye-linked primers (Mouse Mapping Primers, Life Technologies, Carlsbad, CA, USA): D1Mit159.1 (VIC) on chromosome 1, D2Mit395.1 (6-FAM) on chromosome 2, D4Mit170.1 (6-FAM) on chromosome 4, D5Mit201.1 (VIC) on chromosome 5, D13Mit256.1 (NED) on chromosome 13, and D17Mit51.1 (NED) on chromosome 17. VIC, 6-FAM, and NED are green, blue, and yellow fluorescent dyes, respectively.

PCR was performed with 2.4 ℓ of genomic DNA (25 ng/ ℓ) and 1.25 units AmpliTaq Gold (Applied Biosystems, Foster, CA, USA) in a 15 µl reaction volume using the GeneAmp PCR system 9700 (Applied Biosystems). Samples were amplified under the following conditions: an initial incubation at 95°C for 12 min was followed by 10 cycles of 94°C for 20 sec, 55°C for 20 sec, 72°C for 30 sec, and 20 cycles of 89°C for 20 sec, 55°C for 20 sec, 72°C for 30 sec, and finally incubation at 72°C for 10 min.

Labeled products were detected by electrophoretic size fractionation on an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). A size control PCR was performed and the products were subjected to electrophoretic size fractionation as an internal control with each analysis. The end result for each cell line was an electropherogram with each allele represented as one or two peaks. As expected, one peak was detected at each locus in cell lines derived from inbred mouse strains. Samples that failed to give measurable peaks at all loci were reanalyzed using a different concentration of DNA or using newly prepared DNA.

Acknowledgements

This work was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology of Japan.

References

1. Stacey, G.N. (2000). Cell contamination leads to inaccurate data: we must take action now. *Nature* 403, 356.
2. Harris, N.L., Gang, D.L., Quay, S.C., Poppema, S., Zamecnik, P.C., Nelson-Rees, W.A., and O'Brien, S.J. (1981). Contamination of Hodgkin's disease cell cultures. *Nature* 289, 228-230.
3. Lavappa, K.S. (1978). Survey of ATCC stocks of human cell lines for HeLa contamination. *In Vitro* 14, 469-475.
4. Nelson-Rees, W.A., Daniels, D.W., and Flandermeyer, R.R. (1981). Cross-contamination of cells in culture. *Science* 212, 446-452.
5. O'Toole, C.M., Povey, S., Hepburn, P., and Franks, L.M. (1983). Identity of some human bladder cancer cell lines. *Nature* 301, 429-430.
6. Povey, S., Hopkinson, D.A., Harris, H., and Franks, L.M. (1976). Characterisation of human cell lines and differentiation from HeLa by enzyme typing. *Nature* 264, 60-63.
7. Dirks, W.G., MacLeod, R.A., and Drexler, H.G. (1999). ECV304 (endothelial) is really T24 (bladder carcinoma): cell line cross-contamination at source. *In Vitro Cell Dev Biol Anim* 35, 558-559.
8. MacLeod, R.A., Dirks, W.G., Matsuo, Y., Kaufmann, M., Milch, H., and Drexler, H.G. (1999). Widespread intraspecies cross-contamination of human tumor cell lines arising at source. *Int J Cancer* 83, 555-563.
9. Markovic, O., and Markovic, N. (1998). Cell cross-contamination in cell cultures: the silent and neglected danger. *In Vitro Cell Dev Biol Anim* 34, 1-8.
10. Scudiero, D.A., Monks, A., and Sausville, E.A. (1998). Cell line designation change: multidrug-resistant cell line in the NCI anticancer screen. *J Natl Cancer Inst* 90, 862.
11. (2010). Cell line misidentification: the beginning of the end. *Nat Rev Cancer* 10, 441-448.
12. Chatterjee, R. (2007). Cell biology. Cases of mistaken identity. *Science* 315, 928-931.
13. Katsnelson, A. (2010). Biologists tackle cells' identity crisis. *Nature* 465, 537.
14. Phuchareon, J., Ohta, Y., Woo, J.M., Eisele, D.W., and Tetsu, O. (2009). Genetic profiling reveals cross-contamination and misidentification of 6 adenoid cystic carcinoma cell lines: ACC2, ACC3, ACCM, ACCNS, ACCS and CAC2. *PLoS One* 4, e6040.
15. Oldroyd, N.J., Urquhart, A.J., Kimpton, C.P., Millican, E.S., Watson, S.K., Downes, T., and Gill, P.D. (1995). A highly discriminating octoplex short tandem repeat polymerase chain reaction system suitable for human individual identification.

Electrophoresis 16, 334-337.

16. Masters, J.R., Thomson, J.A., Daly-Burns, B., Reid, Y.A., Dirks, W.G., Packer, P., Toji, L.H., Ohno, T., Tanabe, H., Arlett, C.F., et al. (2001). Short tandem repeat profiling provides an international reference standard for human cell lines. *Proc Natl Acad Sci U S A* 98, 8012-8017.
17. Yoshino, K., Iimura, E., Saijo, K., Iwase, S., Fukami, K., Ohno, T., Obata, Y., and Nakamura, Y. (2006). Essential role for gene profiling analysis in the authentication of human cell lines. *Hum Cell* 19, 43-48.
18. Dietrich, W., Katz, H., Lincoln, S.E., Shin, H.S., Friedman, J., Dracopoli, N.C., and Lander, E.S. (1992). A genetic map of the mouse suitable for typing intraspecific crosses. *Genetics* 131, 423-447.
19. Moriwaki, K., Miyashita, N., Mita, A., Gotoh, H., Tsuchiya, K., Kato, H., Mekada, K., Noro, C., Oota, S., Yoshiki, A., et al. (2009). Unique inbred strain MSM/Ms established from the Japanese wild mouse. *Exp Anim* 58, 123-134.
20. Ralph, P., Moore, M.A., and Nilsson, K. (1976). Lysozyme synthesis by established human and murine histiocytic lymphoma cell lines. *J Exp Med* 143, 1528-1533.
21. Ralph, P., and Nakoinz, I. (1977). Direct toxic effects of immunopotentiators on monocytic, myelomonocytic, and histiocytic or macrophage tumor cells in culture. *Cancer Res* 37, 546-550.
22. Soloway, M.S. (1977). Intravesical and systemic chemotherapy of murine bladder cancer. *Cancer Res* 37, 2918-2929.
23. Sudo, H., Kodama, H.A., Amagai, Y., Yamamoto, S., and Kasai, S. (1983). In vitro differentiation and calcification in a new clonal osteogenic cell line derived from newborn mouse calvaria. *J Cell Biol* 96, 191-198.
24. Mekada, K., Abe, K., Murakami, A., Nakamura, S., Nakata, H., Moriwaki, K., Obata, Y., and Yoshiki, A. (2009). Genetic differences among C57BL/6 substrains. *Exp Anim* 58, 141-149.

Int J Hematol
DOI 10.1007/s12185-010-0742-2

Red blood cell production from immortalized progenitor cell line

Yukio Nakamura · Takashi Hiroyama · Kenichi Miharada · Ryo Kurita

Received: 21 November 2010 / Revised: 24 November 2010 / Accepted: 30 November 2010

Abstract The supply of transfusable red blood cells (RBCs) is not sufficient in many countries. If immortalized erythroid progenitor cell lines able to produce transfusable RBCs in vitro were established, they would be valuable resources. However, such cell lines have not been established. We have developed a robust method to establish immortalized erythroid progenitor cell lines following the induction of hematopoietic differentiation of mouse embryonic stem (ES) cells and have established many immortalized erythroid progenitor cell lines so far. Although their precise characteristics varied among cell lines, each of these lines could differentiate in vitro into more mature erythroid cells, including enucleated RBCs. Following transplantation of these erythroid cells into mice suffering from acute anemia, the cells proliferated transiently, subsequently differentiated into functional RBCs, and significantly ameliorated the acute anemia. Considering the number of human ES cell lines that have been established so far and the number of induced pluripotent stem cell lines that will be established in future, the intensive testing of a number of these lines for establishing immortalized erythroid progenitor cell lines may allow the establishment of such cell lines similar to the mouse erythroid progenitor cell lines.

Keywords Erythrocyte · Erythropoiesis · Red blood cell · Transfusion therapy

Y. Nakamura (✉) · T. Hiroyama · K. Miharada · R. Kurita
Cell Engineering Division, RIKEN BioResource Center,
Koyadai 3-1-1, Tsukuba, Ibaraki 305-0074, Japan
e-mail: yukionak@brc.riken.jp

1 Introduction

Organ and cell transplantation therapy is now a standard therapy. However, the supply of organs or cells is not necessarily sufficient all over the world. Hence, production of artificial organs and in vitro production of transplantable cells have been studied in earnest. At the moment, transfusion therapies involving red blood cells (RBCs), platelets, and neutrophils depend on the donation of these cells from healthy volunteers. Unpredictable adverse results derived from the current transfusion therapy system such as contamination of hazardous viruses and transfusion-related acute lung injury (TRALI) may be eliminated by the development of new technologies in the future. On the other hand, lack of supply of transfusable materials by the current system will become a severe problem in advanced nations including Japan, since in those countries the ratios of aged generations who have more opportunities to require transfusion therapies are increasing while the ratios of younger generations who can donate transfusable materials are decreasing. In this situation, research and development to produce transfusable blood materials is very important and should be carried out more earnestly. In particular, since RBC transfusion is now routine and indispensable for many clinical purposes, in vitro production of transfusable RBC is an urgent theme.

2 RBC production from hematopoietic stem cells

The hematopoietic stem cells present in bone marrow and umbilical cord blood are promising materials for in vitro production of RBCs and this has stimulated interest in the development of in vitro procedures for the generation of functional RBCs from these tissues [1–3]. Umbilical cord

Journal : Large 12185 | Dispatch : 6-12-2010 | Pages : 5
Article No. : 742 | LE | TYPESET
MS Code : IJHM-D-10-00708 | CP | DISK

blood cells are of particular interest as they are readily available but are usually discarded. Provided the mother of a neonate consents to use of the umbilical cord blood, this material can provide a useful resource without any further complicating critical or ethical concerns.

It was reported that human erythroid cells (nucleated cells) produced on a large scale ex vivo could differentiate in vivo into enucleated RBCs [1]. This study demonstrated that erythroid progenitor cells produced in vitro from hematopoietic stem and progenitor cells could have a clinical application as an alternative method for transfusing terminally differentiated RBCs. More recently, the same group described an ex vivo methodology for producing fully mature human RBCs from hematopoietic stem cells [2]. The enucleated RBCs produced by this approach are potentially even more valuable as they should be functional immediately after transfusion without requiring the long latency period for enucleation normally necessary for erythroid cells.

The mechanism of erythroblast enucleation, a critical step in RBC production, has not yet been fully elucidated [4, 5]. The role of interactions between erythroblasts and other cells, such as macrophages, in this process is a controversial topic [6–10]. Macrophages in retinoblastoma gene (Rb)-deficient embryos are unable to physically interact with erythroblasts and RBC production is impaired in these embryos [9]. In addition, in vitro production of enucleated RBCs from immature hematopoietic stem/progenitor cells proceeds efficiently in the presence [2] but not in the absence [1] of feeder cells.

Of note, however, enucleation can apparently be initiated in vitro in erythroblasts that have been induced to differentiate in vivo to a developmental stage that is competent for nuclear self-extrusion [10, 11]. Moreover, we have developed a method to produce enucleated RBCs efficiently in vitro without use of feeder cells [3]. The culture system has allowed erythroid cells to differentiate to a developmental stage competent for nuclear self-extrusion [3]. Taken together, although it has generally been thought that efficient enucleation of erythroblasts is largely dependent on signals mediated by cells in their local environment [6–9], the interaction of erythroblasts with other cells is not necessary for efficient erythroblast enucleation [3]. Signals mediated by humoral factors appear to be sufficient for the efficient autonomous completion of erythroblast enucleation [12].

Since culture without the use of feeder cells is technically easier and less expensive, the method we established [3] has the potential to be a cost-effective means of producing transfusable RBCs on a large scale from immature hematopoietic stem/progenitor cells. Currently, however, cost factors mean that it is not yet realistic to produce RBCs on a large scale, approximately 200 ml or more, using our in vitro culture system. In particular, patents on the growth factors used in the culture system are a major obstacle, because these growth factors are very expensive, at least at the moment. After the relevant patents expire, our in vitro culture system will become a more realistic scenario.

3 RBC production from embryonic stem (ES) cells

ES cells possess the potential to produce various differentiated cells able to function in vivo and thus represent another promising resource for RBC production. Furthermore, since ES cell lines are immortalized, they can be used repeatedly and have potential to produce abundant differentiated cells in the quantities required for clinical use. However, it will be important to carry out routing screening of the ES cell lines for de novo chromosomal aberrations and/or genetic mutations that may arise in vitro, before these long-term cell cultures are applied in the clinic. Unsurprisingly, there is now a widespread and enthusiastic debate on standardization of the characteristics of ES cells for regenerative medicine protocols that exploit these cell lines. In my opinion, since chromosomal aberrations and genetic mutations are inevitable in long-term cell cultures, only ES cell lines that have been cultured for a limited period, e.g., less than 30 passages, should be selected for clinical use.

Hematopoietic cells, including those in the erythroid lineage, have been generated from mouse ES cells [13–16], non-human primate ES cells [17–19], and human ES cells [20–25]. We have also established a long-term in vitro method for culturing hematopoietic cells derived from ES cells of the non-human primate, the common marmoset [26]. Recently, abundant production of enucleated RBCs from human ES cells was reported [27].

Taken together, we can now produce mature RBCs by in vitro culture of ES cells or the hematopoietic stem/progenitor cells present in umbilical cord blood. In practice, however, the efficiency of RBC generation varies with the quality of the ES cell line or the umbilical cord blood sample. Since ES cell lines can be utilized repeatedly, derivation of RBCs from ES cells appears to be more practical. However, even with optimal experimental procedures and the most appropriate ES cell line, the generation of abundant RBCs directly from primate ES cells is a costly and time-consuming process [26, 27]. If human erythroid progenitor cell lines can be established that have efficient production of mature RBCs, they would provide a much more useful resource than ES cell lines.

Journal : Large 12185 | Dispatch : 6-12-2010 | Pages : 5
Article No. : 742 | LE | TYPESET
MS Code : IJHM-D-10-00708 | CP | DISK

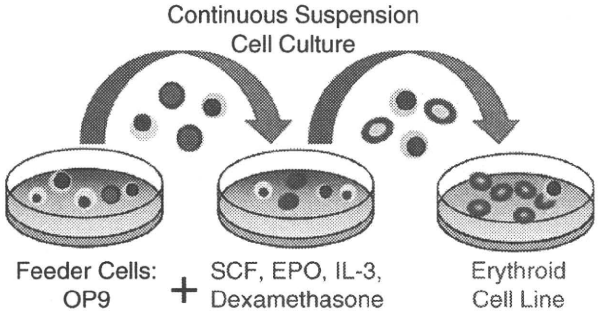


Fig. 1 Culture protocol to establish immortalized erythroid progenitor cell lines. Hematopoietic cells derived from ES or iPS cells are cultured on feeder cells, OP9, in the presence of stem cell factor (SCF), erythropoietin (EPO), interleukin-3 (IL-3) and dexamethasone

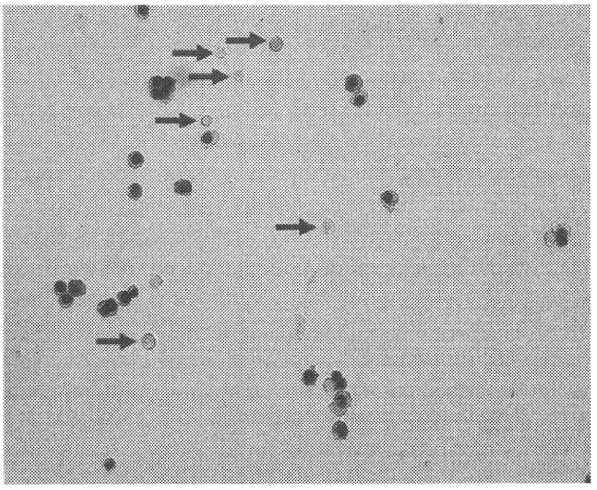


Fig. 2 Red blood cells (RBCs) produced from immortalized mouse progenitor cell line. *Arrows* indicate enucleated RBCs

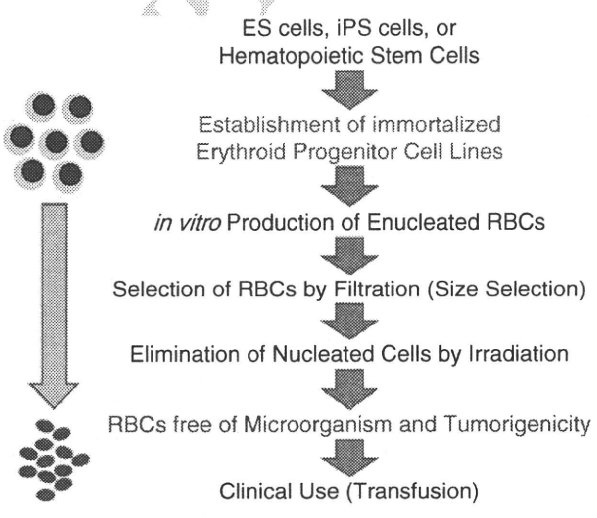


Fig. 3 Strategy to utilize red blood cells (RBCs) produced in vitro from immortalized erythroid progenitor cell lines in the clinic

4 Establishment of mouse RBC progenitor cell lines able to produce transfusable RBCs

Several mouse and human erythroid cell lines have been established. However, to the best of our knowledge, there is no cell line that can efficiently differentiate into enucleated RBCs. It is generally difficult to establish hematopoietic cell lines from adult hematopoietic stem and progenitor cells as both are sensitive to DNA damage and are unable to maintain the lengths of telomere repeats on serial passage [28]. In contrast, ES cells are relatively resistant to DNA damage and maintain telomere lengths on serial passage [28]. Therefore, these characteristics of ES cells may be advantageous for the establishment of cell lines since differentiated cells derived from ES cells may retain them.

Recently, we developed a robust method to obtain differentiated cell lines following the induction of hematopoietic differentiation of mouse ES cells (Fig. 1), and established five independent hematopoietic cell lines using this method [29]. Three of these lines exhibited characteristics of erythroid cells, and they were designated mouse ES cell-derived erythroid progenitor (MEDEP) cell lines. Although their precise characteristics varied, each of the MEDEP lines could differentiate in vitro into more mature erythroid cells, including enucleated RBCs (Fig. 2). Following transplantation into mice suffering from acute anemia, MEDEP cells proliferated transiently and subsequently differentiated into functional RBCs. Treated mice showed a significant amelioration of acute anemia. In addition, MEDEP cells did not form tumors following transplantation into mice. This report was the first to demonstrate the feasibility of establishing erythroid cell lines able to produce mature RBCs [29]. At present, the mechanism underlying the establishment of differentiated cell lines from ES cells has not been elucidated. Nevertheless, our data clearly indicate that useful erythroid cell lines can be reproducibly obtained from mouse ES cells.

5 Establishment of RBC progenitor cell lines from human ES cells or human induced pluripotent stem (iPS) cells

The reproducible establishment of MEDEP cell lines described above strongly suggests that similar erythroid cell lines could also be established from human ES cells. We, therefore, sought to establish human erythroid progenitor cell lines. The methods used to induce hematopoietic cells from ES cells and to culture the induced hematopoietic cells are similar to those established for MEDEP cell lines, with the exception that the corresponding human factors were applied and IL-3 was not used at all. Exclusion of IL-3 was based on our finding that

Journal : Large 12185 | Dispatch : 6-12-2010 | Pages : 5
Article No. : 742 | LE | TYPESET
MS Code : IJHM-D-10-00708 | CP | DISK

the compound was not necessary for establishment of MEDEP cell lines [29].

Initially, we used three human ES cell lines, KhES-1, KhES-2 and KhES-3, that had been established in Japan. Although we were able to induce hematopoietic cells from all three lines, the efficiency of production of hematopoietic cells was extremely low compared to the cases of mouse ES cells. As a result, we have yet been successful in establishing immortalized cell lines from the three human ES cell lines.

During the course of the experiments using these human ES cell lines, a breakthrough discovery in the field of regenerative medicine was reported, namely, the establishment of human iPS cells [30] following that of mouse iPS cells [31]. This discovery prompted us to establish human iPS cells, since the characteristics of pluripotent stem cells, such as ES cells, differ among cell lines. In other words, we speculated that we could obtain iPS cell lines that could have the ability to differentiate into hematopoietic cells efficiently. We were able to establish a number of human iPS cell lines using fibroblast-like cells derived from neonatal tissues [32]. Fortunately, we were able to induce abundant numbers of hematopoietic cells from some of these iPS cell lines and also to establish immortalized hematopoietic cell lines from the induced hematopoietic cells (data not shown). Currently, we are investigating the characteristics of these immortalized hematopoietic cell lines. Some seem to be erythroid cell lines.

6 Clinical application of human RBC progenitor cell lines

We reported that MEDEP cells did not exhibit tumorigenicity in vivo [29]. Nevertheless, the tumorigenic potential of any human erythroid cell line will need to be thoroughly analyzed prior to clinical use [33, 34]. In general, immortalized cell lines are not necessarily homogenous in karyotype, even after cloning. The emergence of cells possessing abnormal karyotypes is often observed following continuous culture of immortalized cell lines. Indeed, although the vast majority of the MEDEP cells in each cell line could differentiate into mature erythroid cells and transplantation of these cells significantly ameliorated anemia, the MEDEP lines included many cells possessing abnormal karyotypes [29]. Hence, it may be advisable to engineer the cells in such a way that they are eliminated if a malignant phenotype arises for any reason [35].

Alternatively, the use of terminally differentiated cells that no longer have the capability of proliferating should allow clinical applications of ES cell derivatives without the associated risk of tumorigenicity. Thus, e.g., RBCs lack nuclei following terminal differentiation, and are highly unlikely to exhibit tumorigenicity in vivo. As such, even if the original ES/iPS cells or the immortalized erythroid progenitor cell lines derived from them possessed abnormal karyotypes and/or genetic mutations, they might nonetheless be useful for clinical applications, provided that they can produce enucleated RBCs. Since enucleated RBCs are much smaller than normal nucleated cells, enucleated RBCs produced in vitro could be selected by size prior to use in the clinic so as to exclude nucleated cells, e.g., by filtration. In addition, X-ray irradiation might be useful for eradicating any contaminating nucleated cells without affecting the enucleated RBCs (Fig. 3).

Another potential obstacle to the clinical use of ES cell derivatives is that of immunogenicity [36, 37]. Transplanted MEDEP cells could not ameliorate acute anemia in mouse strains other than those from which each individual cell line was derived or in immunodeficient mice [29], suggesting immunological rejection in heterologous strains. Hence, the direct clinical application of immortalized erythroid cell lines will require use of many cell lines that express different major histocompatibility (MHC) antigens. However, in vitro-generated mature RBCs need to be compatible with ABO and RhD antigens alone (8 types in total). Moreover, RBCs lacking all of A, B and RhD antigens can be transfused into the vast majority of patients around the world. Hence, the establishment of a human erythroid cell line lacking the genes to produce A, B and RhD antigens would be a very useful resource for clinical application.

7 Conclusion

We propose that by utilizing ES cells or iPS cells it will be possible to establish human erythroid progenitor cell lines able to produce enucleated RBCs. RBCs produced by in vitro culture of such erythroid cell lines could be applied in the clinic following size selection and elimination of nucleated cells by irradiation (Fig. 3).

Acknowledgments This work was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology in Japan.

References

1. Neildez-Nguyen TM, Wajcman H, Marden MC, Bensidhoum M, Moncollin V, Giarratana MC, Kobari L, Thierry D, Douay L. Human erythroid cells produced ex vivo at large scale differentiate into red blood cells in vivo. Nat Biotechnol. 2002;20: 467–72.
2. Giarratana MC, Kobari L, Lapillonne H, Chalmers D, Kiger L, Cynober T, et al. Ex vivo generation of fully mature human red

Journal : Large 12185 | Dispatch : 6-12-2010 | Pages : 5
Article No. : 742 | LE | TYPESET
MS Code : IJHM-D-10-00708 | CP | DISK

blood cells from hematopoietic stem cells. Nat Biotechnol. 2005; 23:69–74.
3. Miharada K, Hiroyama T, Sudo K, Nagasawa T, Nakamura Y. Efficient enucleation of erythroblasts differentiated in vitro from hematopoietic stem and progenitor cells. Nat Biotechnol. 2006; 24:1255–6.
4. Lee JC, Gimm JA, Lo AJ, Koury MJ, Krauss SW, Mohandas N, et al. Mechanism of protein sorting during erythroblast enucleation: role of cytoskeletal connectivity. Blood. 2004;103:1912–9.
5. Kingsley PD, Malik J, Fantauzzo KA, Palis J. Yolk sac-derived primitive erythroblasts enucleate during mammalian embryogenesis. Blood. 2004;104:19–25.
6. Ohneda O, Bautch VL. Murine endothelial cells support fetal liver erythropoiesis and myelopoiesis via distinct interactions. Brit J Haematol. 1997;98:798–808.
7. Yanai N, Sato Y, Obinata M. A new type-II membrane protein in erythropoietic organs enhances erythropoiesis. Leukemia. 1997; 11:484–5.
8. Hanspal M, Smockova Y, Uong Q. Molecular identification and functional characterization of a novel protein that mediates the attachment of erythroblasts to macrophages. Blood. 1998;92: 2940–50.
9. Iavarone A, King ER, Dai XM, Leone G, Stanley ER, Lasorella A. Retinoblastoma promotes definitive erythropoiesis by repressing Id2 in fetal liver macrophages. Nature. 2004;432:1040–5.
10. Spike BT, Dirlam A, Dibling BC, Marvin J, Williams BO, Jacks T, et al. The Rb tumor suppressor is required for stress erythropoiesis. EMBO J. 2004;23:4319–29.
11. Yoshida H, Kawane K, Koike M, Mori Y, Uchiyama Y, Nagata S. Phosphatidylserine-dependent engulfment by macrophages of nuclei from erythroid precursor cells. Nature. 2005;437:754–8.
12. Hebiguchi M, Hirokawa M, Guo YM, Saito K, Wakui H, Komatsuda A, et al. Dynamics of human erythroblast enucleation. Int J Hematol. 2008;88:498–507.
13. Keller G, Kennedy M, Papayannopoulou T, Wiles MV. Hematopoietic commitment during embryonic stem cell differentiation in culture. Mol Cell Biol. 1993;13:473–86.
14. Nakano T, Kodama H, Honjo T. Generation of lymphohematopoietic cells from embryonic stem cells in culture. Science. 1994;265:1098–101.
15. Nakano T, Kodama H, Honjo T. In vitro development of primitive and definitive erythrocytes from different precursors. Science. 1996;272:722–4.
16. Carotta S, Pilat S, Mairhofer A, Schmidt U, Dolznig H, Steinlein P, et al. Directed differentiation and mass cultivation of pure erythroid progenitors from mouse embryonic stem cells. Blood. 2004;104:1873–80.
17. Li F, Lu S, Vida L, Thomson JA, Honig GR. Bone morphogenetic protein 4 induces efficient hematopoietic differentiation of rhesus monkey embryonic stem cells in vitro. Blood. 2001;98:335–42.
18. Umeda K, Heike T, Yoshimoto M, Shiota M, Suemori H, Luo HY, et al. Development of primitive and definitive hematopoiesis from nonhuman primate embryonic stem cells in vitro. Development. 2004;131:1869–79.
19. Kurita R, Sasaki E, Yokoo T, Hiroyama T, Takasugi K, Imoto H, et al. Tal1/Scl gene transduction using a lentiviral vector stimulates highly efficient hematopoietic cell differentiation from common marmoset (*Callithrix jacchus*) embryonic stem cells. Stem Cells. 2006;24:2014–22.
20. Kaufman DS, Hanson ET, Lewis RL, Auerbach R, Thomson JA. Hematopoietic colony-forming cells derived from human embryonic stem cells. Proc Natl Acad Sci USA. 2001;98:10716–21.
21. Chadwick K, Wang L, Li L, Menendez P, Murdoch B, Rouleau A, et al. Cytokines and BMP-4 promote hematopoietic differentiation of human embryonic stem cells. Blood. 2003;102:906–15.
22. Cerdan C, Rouleau A, Bhatia M. VEGF-A165 augments erythropoietic development from human embryonic stem cells. Blood. 2004;103:2504–12.
23. Vodyanik MA, Bork JA, Thomson JA, Slukvin II. Human embryonic stem cell-derived CD34+ cells: efficient production in the coculture with OP9 stromal cells and analysis of lymphohematopoietic potential. Blood. 2005;105:617–26.
24. Wang L, Li L, Menendez P, Cerdan C, Bhatia M. Human embryonic stem cells maintained in the absence of mouse embryonic fibroblasts or conditioned media are capable of hematopoietic development. Blood. 2005;105:4598–603.
25. Olivier EN, Qiu C, Velho M, Hirsch RE, Bouhassira EE. Large-scale production of embryonic red blood cells from human embryonic stem cells. Exp Hematol. 2006;34:1635–42.
26. Hiroyama T, Miharada K, Aoki N, Fujioka T, Sudo K, Danjo I, et al. Long lasting in vitro hematopoiesis derived from primate embryonic stem cells. Exp Hematol. 2006;34:760–9.
27. Lu SJ, Feng Q, Park JS, Vida L, Lee BS, Strausbauch M, et al. Biologic properties and enucleation of red blood cells from human embryonic stem cells. Blood. 2008;112:4475–84.
28. Lansdorp PM. Role of telomerase in hematopoietic stem cells. Ann NY Acad Sci. 2005;1044:220–7.
29. Hiroyama T, Miharada K, Sudo K, Danjo I, Aoki N, Nakamura Y. Establishment of mouse embryonic stem cell-derived erythroid progenitor cell lines able to produce functional red blood cells. PLoS One. 2008;3:e1544.
30. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, Tomoda K, et al. Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell. 2007;131:861–72.
31. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell. 2006;126:663–76.
32. Fujioka T, Shimizu N, Yoshino K, Miyoshi H, Nakamura Y. Establishment of induced pluripotent stem cells from human neonatal tissues. Hum Cell. 2010;23:113–8.
33. Vogel G. Ready or not? Human ES cells head toward the clinic. Science. 2005;308:1534–8.
34. Hentze H, Graichen R, Colman A. Cell therapy and the safety of embryonic stem cell-derived grafts. Trends Biotechnol. 2007;25: 24–32.
35. Schuldiner M, Itskovitz-Eldor J, Benvenisty N. Selective ablation of human embryonic stem cells expressing a "suicide" gene. Stem Cells. 2003;21:257–65.
36. Drukker M, Benvenisty N. The immunogenicity of human embryonic stem-derived cells. Trends Biotechnol. 2004;22:136–41.
37. Boyd AS, Higashi Y, Wood KJ. Transplanting stem cells: potential targets for immune attack. Modulating the immune response against embryonic stem cell transplantation. Adv Drug Deliv Rev. 2005;57:1944–69.

Journal : Large 12185 | Dispatch : 6-12-2010 | Pages : 5
Article No. : 742 | LE | TYPESET
MS Code : IJHM-D-10-00708 | CP | DISK

Published online 13 November 2010 *Nucleic Acids Research, 2011, Vol. 39, Database issue* ***D861–D870***
doi:10.1093/nar/gkq1078

The RIKEN integrated database of mammals

Hiroshi Masuya[1,*], Yuko Makita[2], Norio Kobayashi[2], Koro Nishikata[2], Yuko Yoshida[2], Yoshiki Mochizuki[2], Koji Doi[2], Terue Takatsuki[1], Kazunori Waki[1], Nobuhiko Tanaka[1], Manabu Ishii[2], Akihiro Matsushima[2], Satoshi Takahashi[2], Atsushi Hijikata[3], Kouji Kozaki[4], Teiichi Furuichi[5], Hideya Kawaji[6], Shigeharu Wakana[1], Yukio Nakamura[1], Atsushi Yoshiki[1], Takehide Murata[1], Kaoru Fukami-Kobayashi[1], Sujatha Mohan[3], Osamu Ohara[3], Yoshihide Hayashizaki[6], Riichiro Mizoguchi[4], Yuichi Obata[1] and Tetsuro Toyoda[2,*]

[1]RIKEN BioResource Center (BRC), Tsukuba, [2]RIKEN Bioinformatics and Systems Engineering (BASE) division, Yokohama, [3]RIKEN Research Center for Allergy and Immunology (RCAI), Yokohama, [4]Department of Knowledge Systems, The Institute of Scientific and Industrial Research (ISIR), Osaka University, Ibaraki, [5]Laboratory of Molecular Neurogenesis, RIKEN Brain Science Institute (BSI), Wako and [6]RIKEN Omics Science Center (OSC),
Yokohama, Japan

Received September 15, 2010; Revised and Accepted October 14, 2010

ABSTRACT

The RIKEN integrated database of mammals (http://scinets.org/db/mammal) is the official undertaking to integrate its mammalian databases produced from multiple large-scale programs that have been promoted by the institute. The database integrates not only RIKEN's original databases, such as FANTOM, the ENU mutagenesis program, the RIKEN Cerebellar Development Transcriptome Database and the Bioresource Database, but also imported data from public databases, such as Ensembl, MGI and biomedical ontologies. Our integrated database has been implemented on the infrastructure of publication medium for databases, termed SciNetS/SciNeS, or the Scientists' Networking System, where the data and metadata are structured as a semantic web and are downloadable in various standardized formats. The top-level ontology-based implementation of mammal-related data directly integrates the representative knowledge and individual data records in existing databases to ensure advanced cross-database searches and reduced unevenness of the data management operations. Through the development of this database, we propose a novel methodology for the development of standardized comprehensive management of heterogeneous data sets in multiple databases to improve the sustainability, accessibility, utility and publicity of the data of biomedical information.

INTRODUCTION

Securing the sustainability of databases is one of the most important issues for research institutes, funding agencies and research communities, because the accumulated cost of maintenance becomes a serious burden on the responsible institutes and communities (1). Moreover, the development of technology for biomedical analyses has brought about a dramatic increase in the amount and variety of data and information. The outdating of isolated data is also a serious problem. The association with public data records broadly used in the research community is crucially important to improve the usability and accessibility of data. If data are isolated in the application software without updates from external data, then the data will become increasingly difficult to retrieve by external retrieval systems and will become useless, unnecessarily occupying the storage resource. On the contrary, the integration of a datum with external data will generally increase its usability and value, often promoting unexpected uses and knowledge discovery. In the community of mammalian research, authoritative data are provided by the Mouse Genome Informatics Database (MGI), HUGO

*To whom correspondence should be addressed. Tel: +81 45 503 9553; Fax: +81 45 503 9553; Email: toyoda@base.riken.jp
Correspondence may also be addressed to Hiroshi Masuya. Tel: +81 29 836 9013; Fax: +81 29 836 9017; Email: hmasuya@brc.riken.jp

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

Gene Nomenclature Committee (HGNC) and Rat Genome Database (RGD) with nomenclature activities for genes, alleles and strains for each species (2–4). Data from the National Center for Biotechnology Information (NCBI) and Ensembl are also broadly used across species (5,6). The Open Biomedical Ontology (OBO) Consortium, an umbrella body for the developers of life-science ontologies, also provides ontologies developed with the aim of comprehensive annotation of biological information (7,8).

In the mouse genetical research community, these issues have been discussed by international consortia. The Mouse Phenotype Database Integration Consortium (InterPhenome) (http://www.interphenome.org/) and the Coordination and Sustainability of International Mouse Informatics Resources (CASIMIR) (http://www.casimir.org.uk/) have discussed broad issues regarding the integration, coordination, interoperability and sustainability of databases, such as methodologies to integrate phenotype information, the association of phenotype with human disease, models for long-term and financial sustainability for databases and legal issues of data accessibility (9–11). A complete solution to satisfy these multiple and broad requirements at once is desired to ensure the sustainability of databases.

One effective way to reduce the management cost of databases is to share common fundamental infrastructures such as the hardware and application software used in their implementations. Recently, such common operations have been effectively implemented through 'cloud computing', which is a type of internet-based computing whereby shared resources, software and information are provided on demand. Cloud computing is often economically beneficial for the facility in terms of the running costs of space, electricity, cooling and staff support (12). If data are properly and continuously managed and integrated with the public data records that are regarded as the *de facto* standard in the biomedical community, then a common infrastructure could be one of the best ways to achieve cost effectiveness and advanced usability. On the other hand, the 'semantic web' offers a series of methods and technologies to develop extensions of the current World Wide Web (WWW) in which information is given well-defined meanings and integrated (13). These technologies include the Resource Description Framework (RDF), a variety of data interchange formats (e.g. RDF/XML, N3, Turtle and N-Triples), and notations such as the RDF Schema (RDFS) and the Web Ontology Language (OWL), all of which are intended to provide a formal description of concepts, terms and relationships within a given knowledge domain. The semantic web is regarded as an integrator across different content and information applications and systems and provides mechanisms for the realization of a common information system. It is also useful for the dissemination of data, providing a standardized framework to describe metadata recommended by the WWW consortium that aids the automated (and also manual) processing of disseminated data to derive meaning from the data. The dissemination of data with standardized metadata risks the extinction of the data and creates the opportunity to promote the discovery of new knowledge. Consequently, the semantic web seems to be suitable as a fundamental technology to implement the common infrastructure.

In this study, we developed a new database, the RIKEN integrated database of mammals, as an official undertaking in RIKEN to integrate heterogeneous mammal-related data in multiple individual databases. This database was constructed on the Scientists' Networking System (SciNetS: http://www.riken.jp/engn/r-world/info/release/press/2009/090331_2/), a general fundamental system that applies the semantic web technology to provide massive data management, supported by Japan's national database integration project. In this system, we achieved the top-level ontology-based re-organization of imported data to integrate the typical and instructive knowledge with individual data records. The RIKEN integrated database of mammals is complementary to the original databases. For example, the FANTOM web resource aims to present data on the dynamic behavior of transcription and its regulation in the expanding fields of the transcriptome, epigenome and transcriptional networks (14,15). By contrast, this integrated database attaches greater importance to the standardization of data for better distribution, metadata-level integration and cross-database retrieval.

DATABASES TO BE INTEGRATED

In RIKEN, there are a number of databases related to mammalian research resources. In the primary development of the integrated database, we integrated six database projects: the Functional Annotation of the Mammalian Genome 4 (FANTOM 4: http://fantom.gsc.riken.jp) (14–16), the RIKEN Cerebellar Development Transcriptome Database (CDT-DB: http://www.cdtdb.brain.riken.jp/CDT/Top.jsp) (17,18), the resource database from the RIKEN BioResource Center (BRC) (19–21) including mutant resources produced by the ENU mutagenesis program (22,23) and the Resource of Asian Primary Immunodeficiency Diseases (RAPID) (24), the RIKEN Structural Genomics/Proteomics Initiative (RSGI) and two data repositories for the Reference Database of Immune Cells (RefDIC) (25) and the RIKEN Expression Array Database (READ) (26), all of which are produced from individual research projects in the human and mouse. Each database project has its original data schema to represent a variety of data ranging from research resources, such as biological strains, cell lines and DNA clones, to experimental data, such as gene expression and phenotypic analyses. There are no relationships defined among original data tables, which are described by various data formats such as text, images and movies. However, as is usual for most databases, they are compiled in a main data table to represent the objects of the database and related information (Table 1). In the discussions of InterPhenome and CASIMIR, it was recommended that the equivalences or relationships among records from the MGI database for genes and alleles, the International Mouse Strain

Table 1. Imported databases in RIKEN (as for September 2010)

Database (URL)	Contents	Project URL in SciNetS
FANTOM4 (http://fantom.gsc.riken.jp/4/)	Monitoring of the dynamics of transcription start site (TSS) usage during a time course of monocytic differentiation in the acute myeloid leukemia cell line THP-1.	http://scinets.org/item/ria187i/
Bio-resource catalog (http://www.brc.riken.jp/)	The online catalog of bioresources including mammalian laboratory strains (mouse), cells and DNA clones in the RIKEN BioResouce Center (BRC).	http://scinets.org/item/ria256i/
RIKEN ENU Mouse Lines (http://www.brc.riken.jp/lab/gsc/mouse/)	Phenotype information of mutant mouse lines generated from large-scale ENU mutagenesis as a resource of the RIKEN BRC.	http://scinets.org/item/rib190i/
Pheno-Pub (http://www.brc.riken.jp/lab/jmc/mouse_clinic/en/m-strain_en.html)	Phenotype data from the standardized phenotyping platform of the Japan Mouse Clinic (JMC) project in the RIKEN BRC.	http://scinets.org/item/ria110i/
Cerebellar Development Transcriptome Database (CDT-DB: http://www.cdtdb.brain.riken.jp/CDT/Top.jsp)	The spatio-temporal gene expression profile of the postnatal development of the mouse cerebellum,	http://scinets.org/item/cria237u1i/
Resource of Asian Primary Immunodeficiency Diseases (RAPID: http://rapid.rcai.riken.jp/RAPID)	A web-based compendium of molecular alterations in primary immunodeficiency diseases.	http://scinets.org/item/cria271u1i/
Systems and Structural Biology Center (SSBC) database (http://www.rsgi.riken.jp/rsgi_e/index.html)	The crystal structures of proteins and the protein–protein interactions in living cells analyzed with the expansion of the genetic code.	http://scinets.org/item/ria46i/
Reference Database of Immune Cells (RefDIC: http://refdic.rcai.riken.jp/welcome.cgi)	An open-access database of quantitative mRNA and protein profiles specifically for immune cells and tissues.	http://scinets.org/item/crib225s27rib225s7i/
RIKEN Expression Array Database (READ: http://read.gsc.riken.jp/)	An integrated system for microarray data that works like 'glue' in post-sequence and post-hybridization analyses.	http://scinets.org/item/crib225s27rib225s8i/

Resource (IMSR) for experimental strain (27) and terms of OBO ontologies be specified. To show the association between the institute's data and the public data broadly used in the research community, we constructed an association between RIKEN's data and public data (Supplementary Table S1).

THE FUNDAMENTALS OF THE INTEGRATED DATABASE: SEMANTIC WEB-BASED CLOUD SYSTEM 'SCINETS'

We have implemented the integrated database on the data-hosting system, SciNetS, which is a fully web-based common platform that ensures cloud computing in the scientific community on the basis of semantic web technologies (Figure 1). It has multiple features useful for data integration:

(1) The system is designed to support sharing of academic information with secure, and to handle databases for sharing, collaborating or publication. Database developers can set multiple levels of accessibility within user groups or the public for each 'project' (private workspace). The user can also declare the copyright licenses for their digital content with Creative Commons (CC) or GNU to indicate the availability for secondary use.
(2) In the project, database developers can also design the semantics with elements and the equivalent methodology to RDF and OWL-Full (i.e. the definition of the semantic links between class and subclass, class and instance, property and sub-property and so on) with graphic user interfaces (GUI) for ontology editors such as Protégé (28). The system assigns the Uniform Resource Identifier (URI) to each data element.
(3) The structured data and metadata can be placed in the public directories of the SciNetS with various standardized data formats, such as RDF, OWL or tab-delimited text file, for downloading or direct connection from external systems and application software such as Protégé.
(4) The system provides the tracking back function making automatically reverse links for RDF relationships across projects. It ensures automatic integration of distributed effort of annotation and curation.
(5) The system is designed to handle a large number of databases simultaneously and is scalable for increased data with the distributed processing technologies on databases and query functions (29).

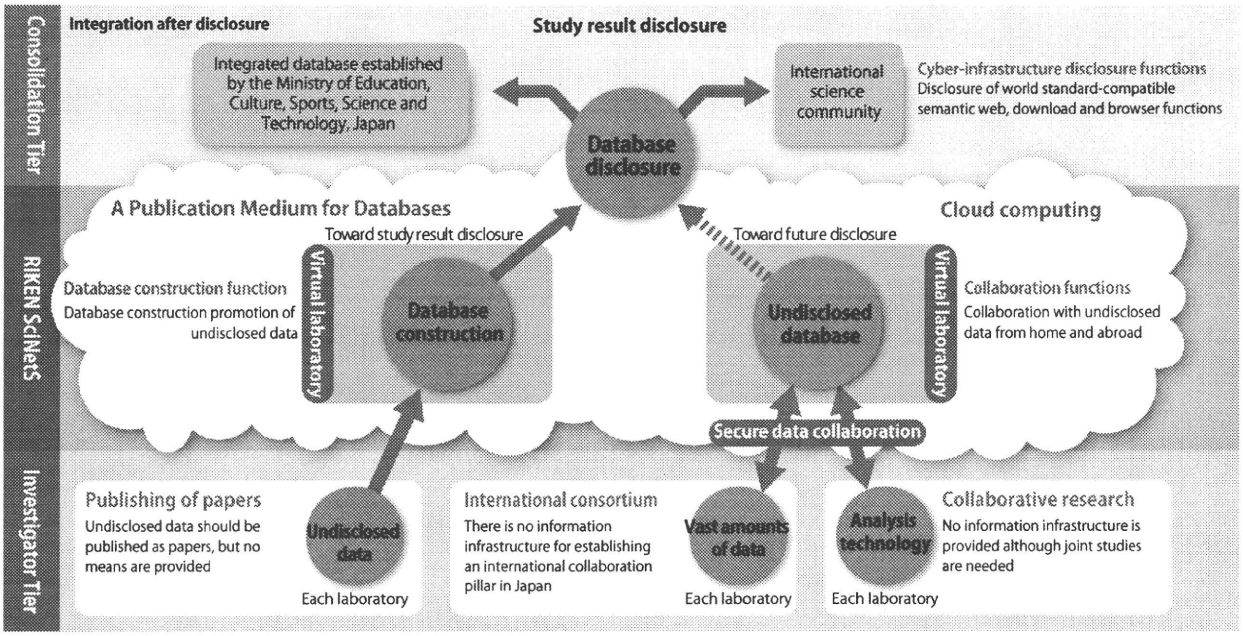


Figure 1. Schematic diagram showing the concept of SciNetS. SciNetS provides incubation functions from database construction to the integration of databases in computing clouds or a group of large-scale servers, and discloses databases using interfaces compatible with international standards, thus contributing to the establishment of cyber-infrastructure for integrating worldwide databases [reprinted with the courtesy from Tetsuro Toyoda, 'Synthetic biology—creating biological resources from information resources' RIKEN RESEARCH 5(10) 13–16, 2010 (http://www.rikenresearch.riken.jp/eng/frontline/6397)].

The high-speed retrieval of semantic content is implemented with the General and Rapid Association Study Engine (GRASE), which enables semantic Boolean-based deduction and statistical evaluation of RDF resources (29).

IMPLEMENTATION OF MAMMALIAN DATA IN SCINETS

The overview of the implementation of this database is presented in Figure 2. The mammalian data and public data shown in Table 1 and Supplementary Table S1, respectively, were imported to SciNetS as individual database projects such that their intact data schema were reflected fully or partially. According to the forms of the original data sources, the databases were imported as three distinct types of projects implemented in SciNetS. First, in the database-type project, a replication of the original database elements, the database table and a data record, is represented with a class and an instance, respectively. Second, the ontology-type project is a replication of the ontology with the OWL methodology. Upon the import of OBO ontologies, ontology files in the OWL format are downloaded from the OBO Foundry website (http://www.obofoundry.org/). Then, the ontology is directly imported into SciNetS. Third, in repository projects, the complete data from a database are stored as single or multiple files. As a result, 27 projects (17 for database, nine for ontology and one for repository) composed of 108 396 classes and 777 319 instances were defined as for September in 2010. These projects are updated monthly in average from constituent databases and ontologies.

Then, we examined the contents and semantics (not the data format or syntax) of 41 classes of imported projects, which play the principal roles in each project. To ensure the consistent classification of the content, we used a top-middle level ontology, YAMATO-GXO Lite (http://scinets.org/item/rib23i/), which is the lightened version of the middle-level ontology, Genetics Ontology (GXO) (30) (http://www.brc.riken.jp/lab/bpmp/ontology/ontology_gxo.html), to bridge between the experimental genetics domain and the latest top-level ontology, Yet Another More Advanced Top-level Ontology (YAMATO) (31) (http://www.ei.sanken.osaka-u.ac.jp/hozo/onto_library/upperOnto.htm). YAMATO-GXO Lite was developed with the ontology editor in SciNetS (paper in preparation). As a result, 41 classes conveying the key information from each project are classified under the fifteen upper classes as follows: 'Genome segment and gene in mammal', 'Allele in mammal', 'Transcript in mammal', 'Protein in mammal', 'Strain resource in mammal', 'Cell line resource in mammal', 'Disease', 'Experimental data with mammalian sample' and 'Mammalian Orthologous group' (Figure 3). The RIKEN Integrated Database of Mammals is implemented as a project to define these classes as a root (http://SciNetS.org/db/mammal). The ontology-based classification of contents was embodied with rdf:subclassOf links, which can be applied across multiple projects in SciNetS. To integrate across species

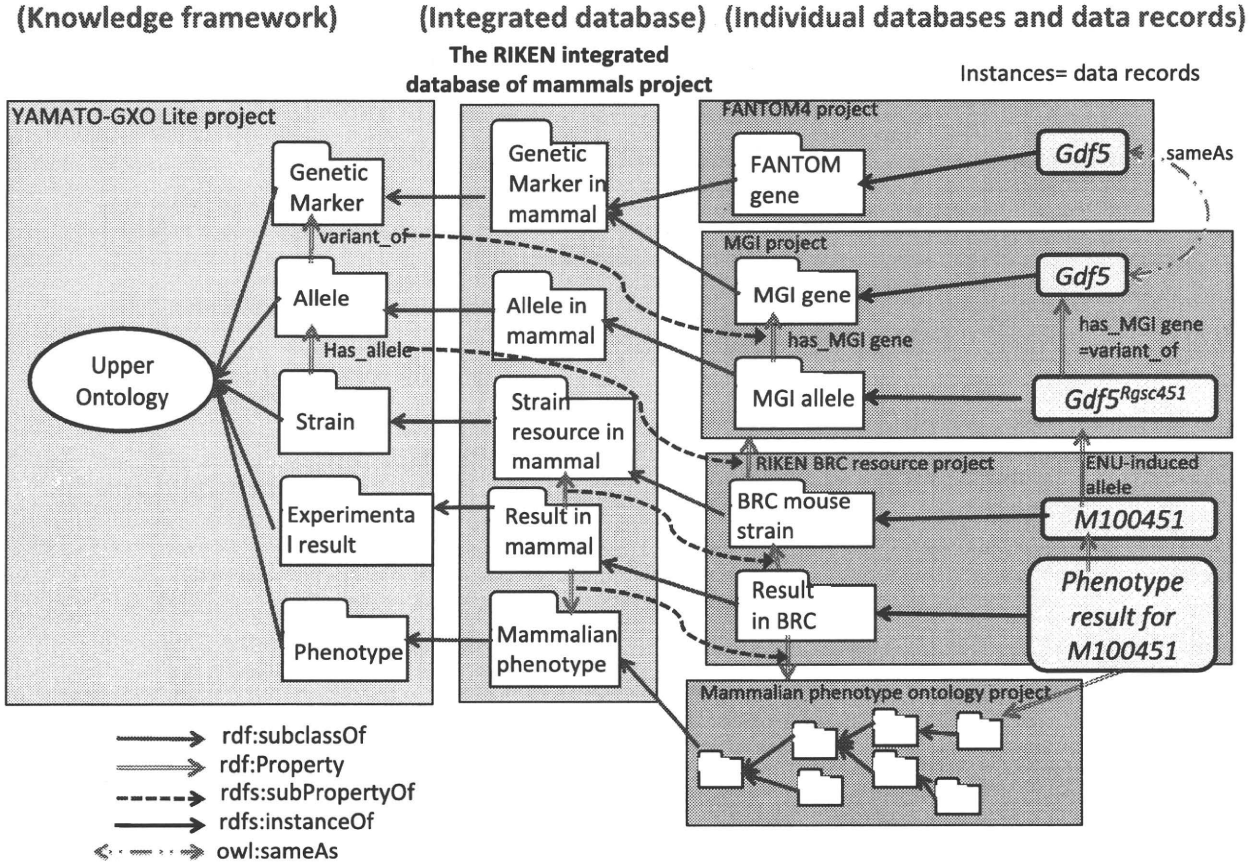


Figure 2. The implementation of data in the RIKEN integrated database of mammals to ensure direct integration between ontologies and databases based on the semantic web technology. Individual public and RIKEN databases are imported as individual projects, and their main contents are reviewed to classify them into the lower classes to root the 15 classes of the integrated database, such as gene, transcript, experimental data, strain and so on. The classification follows the top-level ontology and is directly linked to the equivalents of rdfs:subclassOf across projects. Property links also organized with rdfs:subPropertyOf to represent the logical definition of upper classes are inherited to lower classes or instances.

databases, we applied the 'query-class', which dynamically refers only to specific instances from another class. For example, the diffraction data class in the SSBC project includes the diffraction data from mammal and non-mammal proteins. To extract only mammal data, we implemented the query-class, which is an expanded use of the owl:oneOf element to define a class by enumerating its elements. With these operations, the project for the integrated database works as the bridge to connect the YAMATO-GXO Lite and imported projects, in which the imported classes are defined as lower concepts of the top-level ontology as shown in Figure 2.

In the next step, to ensure further semantic integration of the imported data, we examined the equivalencies of property links (semantic links) between the upper ontology and lower classes in imported projects. For example, the 'Allele' class in YAMATO-GXO Lite has a property named 'variant_of' that takes its value from the range of the 'Genome segment' class. It is the logical representation of one of the features of an allele that the 'allele is a variant of a genome segment'. The examination of properties in lower classes reveals that the 'MGI allele' class has the 'MGI gene' property range of 'MGI gene', which is equivalent to 'variant_of'. Consequently, we defined the 'MGI gene' property as a specified type of (rdfs:subPropertyOf) 'variant_of' to show that *Gdf5*^{Rgsc451}, an instance of the MGI allele class, is a variant of *Gdf5*, an instance of the MGI gene class. With this equivalence mapping of properties between YAMATO-GXO Lite and lower imported database classes, we built the ontology-based information structure so that information defined in the upper classes is instantiated in lower database classes and instances.

In addition, regarding the import of external and internal data records, multiple overlaps of records (instances) were collapsed to represent a single identical entity in the real world (i.e. instances of a gene in the Ensembl, MGI and FANTOM projects). We also examined such equality between instances in lower classes that belong to a single upper class. We related identical data items with a semantic link that is equivalent to owl:sameAs.

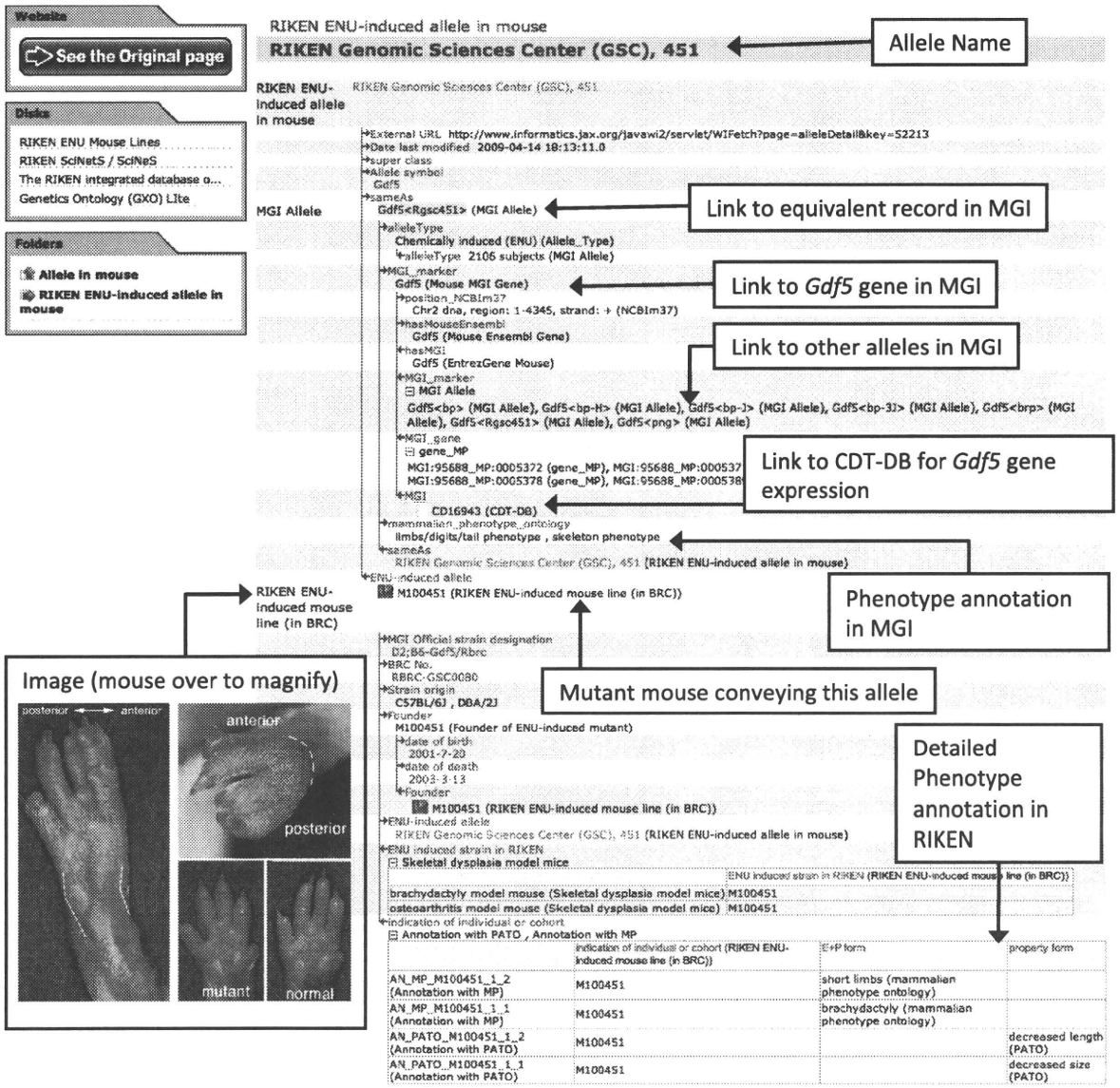


Figure 3. The instance page to illustrate the graphical representation of detailed semantic links of the RIKEN integrated database of mammals. Explanations are given by open squares with an arrow.

USER INTERFACE

At the top page of this integrated database, users can overlook all the classes of integrated databases and those data sizes shown in Supplementary Table S2. The overview of the data structure is presented on the 'data folder' page, where users can navigate down the class hierarchy across database or ontology projects by clicking on the folder icons that represent classes. On the page of each project, detailed explanations of the projects and URL links to the original database websites are shown. On the class and instance pages, detailed explanations, a table view of instances, a graphic representation of semantic links and links to original data records are displayed (Figure 3).

SciNetS implements two kinds of search function, the internal-search and the cross-search. When users search with 'Search' button, SciNetS executes internal-search to retrieve queries within the accessing project and related projects and shows the number of query hits on each folder icon of the accessing page (Figure 4). For the cross-search from whole SciNetS data, users can access from 'Search All' button. SciNetS replies search results in descending order according to traffic. From this cross-search, users can jump to the Positional Medline (PosMed) search, allows the user to retrieve various information (i.e. gene, phenotypes or diseases) correlated with a genomic position by jumping the PosMed database for a full-document search of various contents: scientific literature, genome annotations, phenome

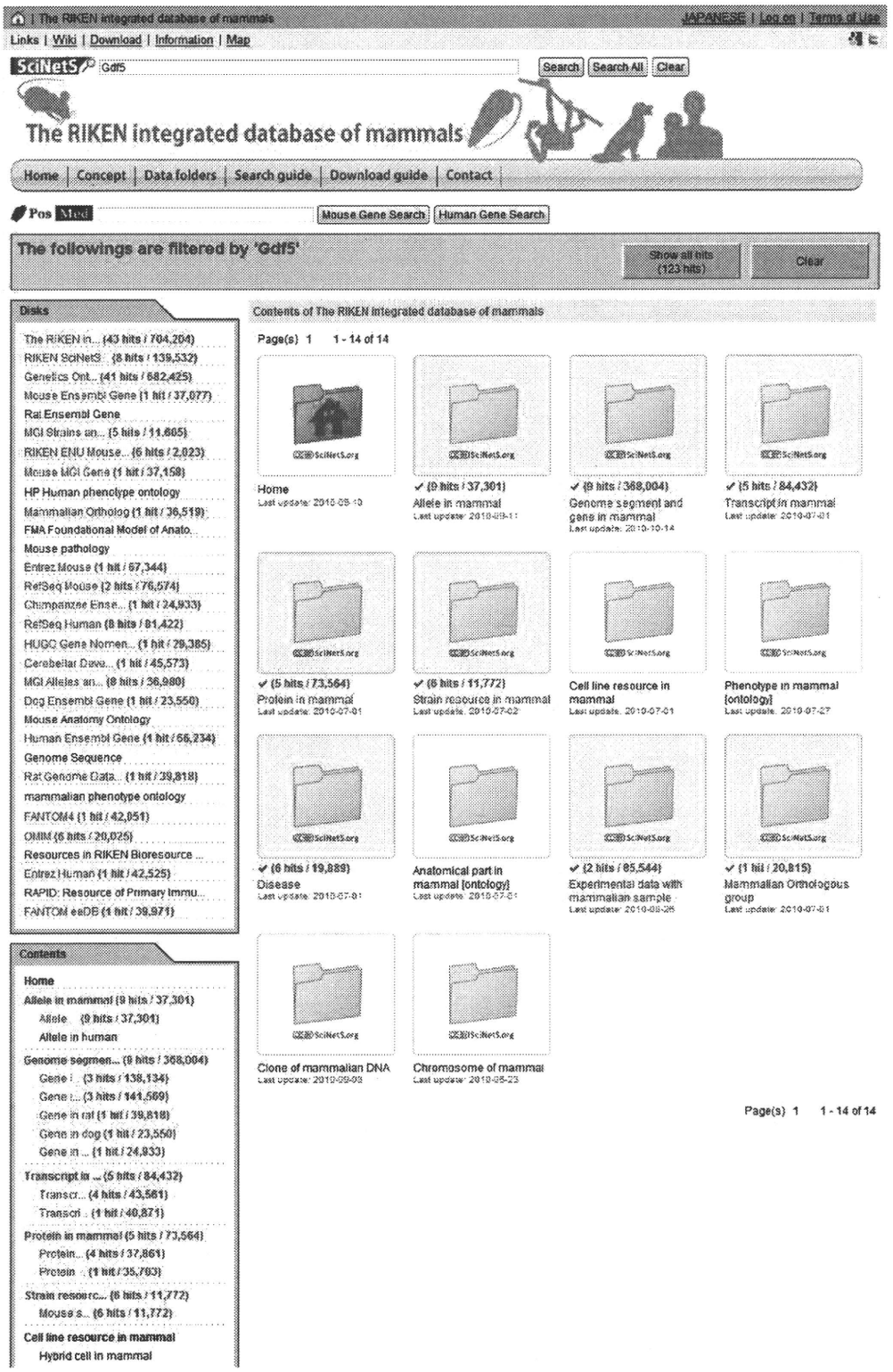


Figure 4. The representation of filtering (search) result of the RIKEN integrated database of mammals. The number of query hits is represented on each disk or folder icon with red letters.