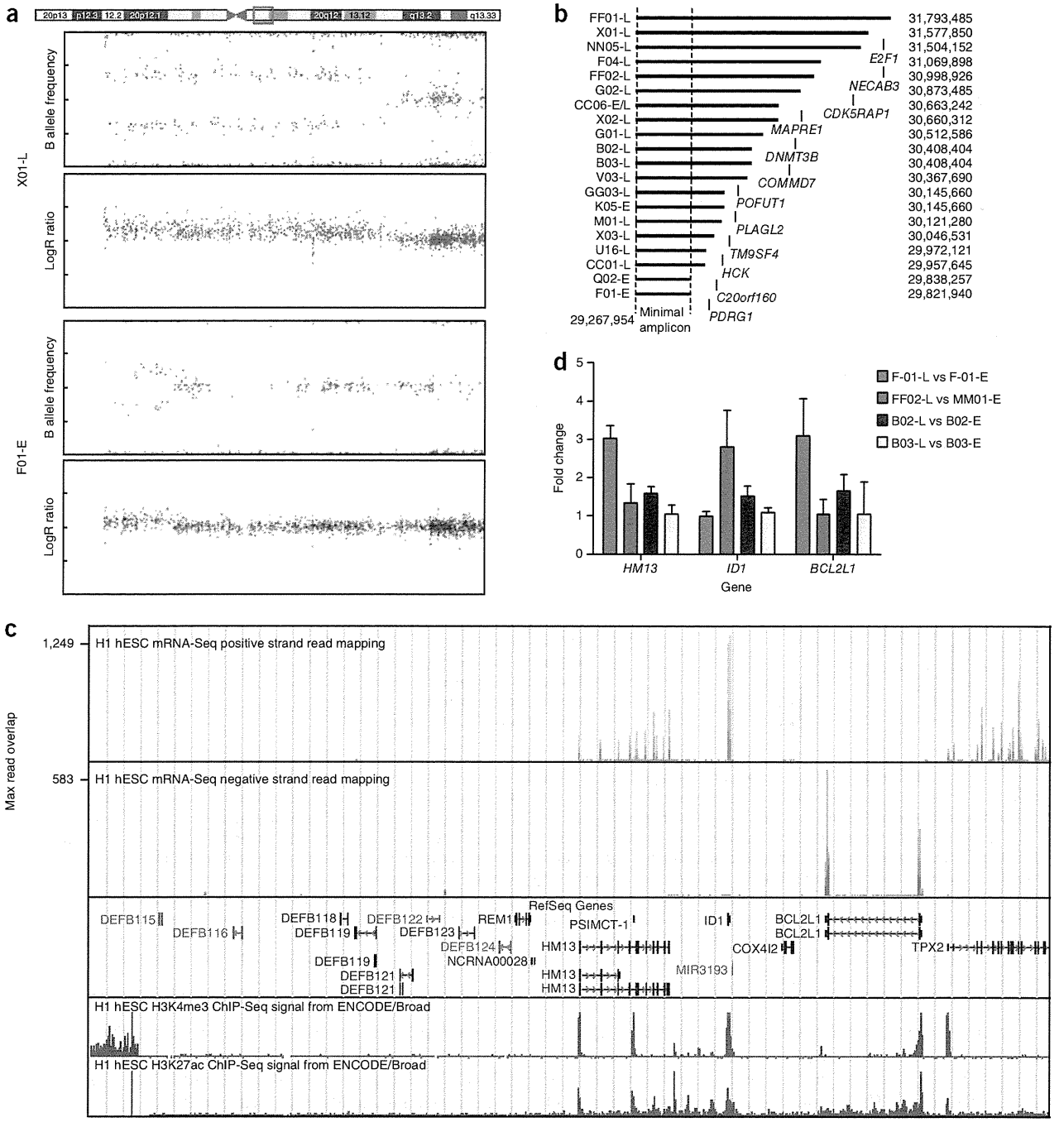


Figure 4 Copy number variation occurrence in human ES cell lines during prolonged passage. (**a**) 20q11.21 gain. The region on chromosome 20 frequently found to experience gain over extended human ES cell culture is indicated by the red boxed region in the chromosome ideogram. Also shown are the B allele frequency and logR ratio plots representing instances of one of the longest and one of the shortest 20q11.21 structural variants. (**b**) Length representation of all individual occurrences of gains in the 20q11.21 region. Samples from which the structural variant was derived are indicated on the left-hand column. The invariant 5′ region and the variable 3′ positions are indicated. Position of genes outside of the minimal amplicon that show greater than 20 RPKM level of expression in human ES cells are shown (RPKM = number of reads that map per kilobase of exon model per million mapped reads for each gene). (**c**) Expression, RefSeq gene, and regulation tracks in the minimal amplicon. Positive and negative strand mRNA-Seq data from H1 human ES cells indicating polyA RNA transcripts expressed within the minimal amplicon region (chr20:29,267,954-29,853,264) are shown together with H1 human ES cell ChIP-Seq data of histone modifications considered universal predictors of enhancer and promoter activity. (**d**) Comparison of expression levels of three genes (*HM13*, *ID1*, *BCL2L1*) contained within the identified minimal 20q11.2 amplicon between early- (normal) and late-passage (20q11.2 CNV carrying) samples. MM01 and FF02 are genetically identical sub-lines from two separate laboratories, MM01 has no amplification at 20q11.2, whereas FF02 possesses a copy number change at 20q11.2 that includes the identified minimal amplicon (**b**).

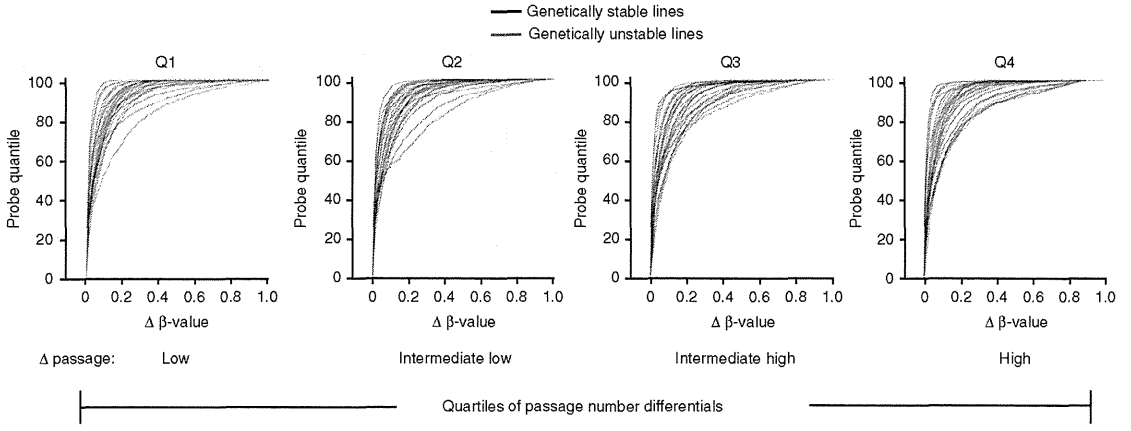


Figure 5 Cumulative distribution function of methylation changes in human ES cells in this study. The change in DNA methylation is represented by empirical CDF curves of the absolute difference in DNA methylation between early- and late-passage cell-line pairs for all 1,536 analyzed probes. The black curves denote genetically stable lines; the red curves denote genetically unstable lines. All analyzed lines were divided into quartiles based on the passage-number difference between the early and late member of each pair. The first quartile contains the lines with the lowest difference in passage number between the early and late sample (range 4 to 47), whereas the fourth quartile contains the lines with the highest difference in estimated population doublings (range 210 to 1,482).

over those without (**Supplementary Fig. 4**). It has also been recently reported that Bcl-X_L, the long, anti-apoptotic isoform encoded by the *BCL2L1* locus, can suppress apoptosis in human ES cells and increase their cloning efficiency[52]. Further, when we transfected MM01 ES cells with a constitutive vector encoding Bcl-X_L, the predominant isoform expressed in human ES cells, these cells showed a distinct growth advantage with respect to the parental cells (**Supplementary Fig. 4**).

DNA methylation analysis

To examine whether cell lines that are genetically unstable at the karyotype level tend to show higher levels of epigenetic instability, we analyzed DNA methylation patterns, focusing on developmentally relevant genes known to be targets of abnormal promoter DNA methylation in cancer[40], and thus most likely to be subjected to selection for altered expression during culture adaptation. For this we used a custom GoldenGate DNA methylation array developed to interrogate DNA methylation changes in known polycomb group protein (PcG) targets in human ES cells[53]. In general, the DNA methylation patterns of the human ES cells tended to be unstable, with both increases or decreases depending upon the locus (**Fig. 5** and **Supplementary Data Set 4**). **Table 2** summarizes those genes that were most frequently subject to gain or loss of methylation during passage, or that showed the least change. Overall, we did not observe any hot spots for DNA methylation at the ~1,500 loci interrogated in the array used in this study, and chromosomes 12, 17 and 20 were not any more methylated, on average, than the rest of the genome.

Table 2 The top 20 genes that were most frequently gained, lost or showed no change in DNA methylation levels in the 120 ES cell lines analyzed at early and late passage

Gained DNA methylation	Lost DNA methylation	No change in DNA methylation
GPC3	*CBLN4*	*NR4A3*
RAB9B	*HIST1H3C*	*EPHA4*
TCEAL4	*LY6H*	*COL12A1*
IL1RAPL2	*HIST1H4L*	*TIGD3*
ESX1	*ANKRD20B*	*SNX7*
TCEAL3	*HIST1H4F*	*PIP5K1B*
AMMECR1	*DMRT2*	*KCNJ2*
MGC39900	*TTLL7*	*T*
LRCH2	*FOXD4L1*	*ZBTB7A*
ZCCHC12	*FOXD4L2*	*IL20RA*
REPS2	*ONECUT1*	*GNAO1*
SOX3	*MAL*	*EPB41L4A*
RP13-360B22.2	*SYT6*	*VDR*
TSC22D3	*BHLHB4*	*HS6ST3*
NHS	*HIST1H3I*	*VGLL2*
TCEAL7	*XTP7*	*SIX1*
MGC4825	*NEUROG1*	*SFT2D2*
GPR50	*TFAP2D*	*BCAN*
BCL2L10	*DRD5*	*ELMOD1*
CDX4	*ASCL2*	*PTGER4*

GPC3 gained more than 5% DNA methylation (range: 98–5%) in over 70% of the samples analyzed, whereas *CBLN4* lost more than 5% DNA methylation (range: 70–5%) in over 60% of them. The genes listed in the "No change" column showed fluctuations in DNA methylation <1% in all samples profiled.

As shown by cumulative distribution function (CDF) curves, most cell lines underwent extensive DNA methylation changes during their time in culture (Online Methods). However, there was a marked difference between the cell lines. For example, in some cell lines there were few changes observed even if there was a large difference in passage level between the early- and late-passage samples (**Fig. 5** Q4 and **Supplementary Table 3**), whereas with other pairs there were large differences observed even when the passage-level difference between the samples was small (**Fig. 5** Q1 and **Supplementary Table 3**). However, the causes of the variation in methylation stability between the lines were not evident. There was no obvious laboratory effect, and the karyotypically abnormal cell lines were not any more unstable than their karyotypically normal counterparts. This suggests that genetic instability played little to no role in the epigenetic instability of the cell lines analyzed. In addition, the DNA methylation patterns of the sibling ES cell lines were as different between themselves as they were between unrelated lines (**Supplementary Data Set 4**), suggesting that the genetic background of human ES cells plays a minor role in the degree of their epigenetic instability.

DISCUSSION

The occurrence of genetic and epigenetic change in human ES cells on prolonged passage is clearly important with respect to their use in regenerative medicine. Understanding the key genes involved and the mechanisms that drive change is important, not only for minimizing the impact of such variants in applications of ES and iPS cells, but also

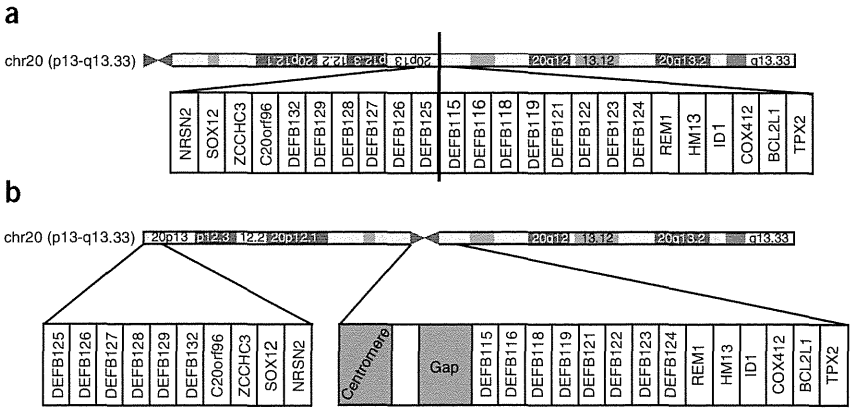


Figure 6 Recent pericentric inversion associated with 20q11.21 susceptibility to gain. (**a**) The ancestral condition of chromosome 20 before a pericentric inversion in the last common ancestor of the gorilla, chimp and human. (**b**) Structure of human chromosome 20 with the location of the gap indicated in which the proximal end of all 20q11.21 amplicons lie.

for exploring the mechanisms that control the fate decisions of pluripotent cells between self-renewal, death and differentiation. Nevertheless, given the scale of the present study, it is striking that most of the ES cell lines studied (79/120 pairs, 66%) remained karyotypically normal, even after many passages, whereas it was only with respect to chromosome 20 that evidence for structural variants in a specific region offering a strong selective advantage could be deduced. Among the small number of iPS cell samples studied, 3 out of 11 had abnormal karyotypes, with 1 of the 3 having the 20q11 gain in the late-passage sample.

Since the first reports of nonrandom chromosomal gain in human ES cells, many studies by standard karyology and by various molecular techniques, including CGH and SNP arrays, have found that, indeed, certain regions of the genome of both ES and, more recently, iPS cells are particularly subject to such genetic change upon prolonged passage in culture. Recently, it was also shown that iPS cells acquire mutations during their derivation, although many such mutations are lost on subsequent passaging[54]. It is commonly assumed that those genetic changes that repeatedly appear in pluripotent stem cells provide variant cells with a growth advantage, but the nature of the selective advantage is unclear. At the molecular karyotype level, it is difficult to disentangle changes that simply reflect variants existing in the human population from those acquired during culture. To address this, we explicitly sought to compare the genomes of a large set of human ES cell lines at two different passage levels and from as diverse a set as possible of the principal laboratories isolating these cells around the world. Although the number of human ES cell lines that have been derived worldwide is uncertain, the 125 ES cell lines analyzed in this study represent a substantial proportion of those commonly available. Notably, our data show that these lines include representatives of most major ethnic groups, reflecting far greater ethnic diversity than previously reported[55,56].

One feature of the human genome emphasized by the current study is that some regions are especially dynamic, particularly but not exclusively those including repetitive elements. In the current panel of ES cells, many regions showed gains or losses between the passage levels, but with no consistency, suggesting that there is no common selection pressure driving the copy number changes. That such dynamically variable regions were readily detected suggests that human ES cell cultures may go through population size restrictions more often than appreciated. Indeed, the cell cycle time of human ES cells is about 18–20 h, but common culture practice involves splitting cultures at low split ratios every 4–5 d or longer. This implies a very large proportion of undifferentiated cells, maybe as many as 90%, are lost between passages of stock cultures[33].

Likewise, the DNA methylation status of the ES cell lines also appeared to change dynamically. Although there was a marked increase in differential DNA methylation with time, indicated by the greater number of DNA methylation changes in the cell lines with the highest differences in passage number, there was also a substantial variation between lines that had undergone similar differences in passage numbers. Thus, human ES cells change not only genetically, but also epigenetically in culture. This conclusion is consistent with several other smaller scale studies that have interrogated human ES cells with respect to either general DNA methylation[25], or imprinting[29,31]. These studies all found DNA methylation and imprinting changes that appeared to be variable between lines and were locus dependent. However, we could not identify specific recurring regions subject to methylation in the genome and there was no observed correlation between DNA methylation changes and chromosomal abnormalities. This suggests that, in general, changes in DNA methylation may be a dynamic process and not necessarily associated with adaptation as such. This point is reinforced by the observation that DNA methylation is markedly different between sibling lines.

In addition to these apparently stochastic and dynamic changes in the genome and epigenome, we did detect marked nonrandom changes in certain parts of the genome. The karyotypic changes seen in the current study match well with other published reports (**Supplementary Fig. 5**)[1]. Gains of chromosomes 1, 12, 17 and 20 and losses of chromosomes 10p and 18q are common in both data sets, and it is only gains of chromosomes 12, 17 and 20 that are often seen as a sole karyotypic change. However, recurrent deletion of chromosome 22q is a novel finding. On the other hand, the gain of chromosome X is a relatively common finding in published studies, whereas only two instances of gain and three instances of loss were observed in the present study. In the light of their relatively frequent occurrence, the minimal amplicons 1q21-qter, 12p11-pter, 17q25-qter and 20q11.2, and perhaps minimal deletions 10p13-pter, 18q21-qter and 22q13-qter deserve special attention as being likely to harbor genes of particular importance for the culture adaptation of human ES cells.

The frequent nonrandom gain of chromosomes 1, 12, 17 and 20 suggests that these chromosomes include a gene(s) that, when overexpressed, confers a growth advantage. Yet, it is striking that in our current extensive study, as in previous studies, structural variant analysis did not point to any frequent repetitive minimal amplicon occurring on chromosomes 1, 12 and 17. Obvious candidate genes are located on these chromosomes—for example, *NANOG* on chromosome 12—but none seems to be more subject to structural variants than other genes on these chromosomes in the absence of karyotypic change. We did see gains spanning the neighboring *SLC2A3/NANOGP1* region described in a recent study[46] but this is just as prevalent, if not more so, within our reference samples and spread across most major ethnic groups, suggesting it is a common structural variant in the human population rather than specific to human ES cells. Together, these observations suggest that the selective advantage attributable to the

gain of chromosomes 1, 12 and 17 may depend upon overexpression of genes or genetic elements at multiple, spatially separated loci, or upon the combination of a structural gene with a long range *cis*-acting regulatory element such that both units must be amplified together to yield an increased function. Alternatively, the appearance of gains within smaller regions may be restricted by chromosomal structure less susceptible to this form of mutation.

By contrast, and in agreement with other studies[5,10,11,23,46,57], our karyotypic and structural variant data point to a region (20q11.21) that, when amplified, apparently drives selection. In this study, because of the much larger number of cell lines and our ability to compare early and late passage, we were able to map the gain to a specific region. Other studies have also reported that gains in this region are associated with enhanced growth characteristics[23], and at least some of the lines in the present study were reported by their contributors to have increased population growth rates (data not shown). The frequency of this gain (25% of the karyotypically normal cell lines), combined with the enrichment in late-passage samples, clearly indicates its selective advantage in human ES cell culture. The mechanism for the selective advantage presumably lies in the minimal region shared by all 22 affected lines, a region containing 13 genes, only three of which are known to be expressed in human ES cells: *HM13*, *ID1* and *BCL2L1*.

A recent genome-wide RNA interference (RNAi) screen highlights the functional importance of *BCL2L1*, an anti-apoptotic factor, in human ES cell biology[58]. This RNAi screen ranked *BCL2L1* twenty-second of 21,121 genes in reducing proliferation after knockdown, whereas *HM13* and *ID1* were ranked 6,679th and 4,224th, respectively[58]. Additionally, a recent structural variant screen of >3,000 specimens from two dozen cancer cell types similarly identified a reoccurring gain on 20q11.21 in which *BCL2L1* was also contained within the minimal amplicon, and knockdown experiments indicated a role for *BCL2L1* in cancer cell proliferation[59]. Recently, it has also been reported that overexpression of the related anti-apoptotic gene, *BCL2*, enhances the survival of human ES cells[60], although *BCL2* is encoded with the region of chromosome 18 subject to recurrent loss in the current data set. Taken together, these observations suggest that similar mutations shared between ES and cancer cells lead to a selective advantage during clonal evolution. The temporal component of our study, where we see[17] instances of early/normal to late/mutated transitions, provides additional support for the notion that the 20q11.21 mutation is the driver mutation in the clonal evolution of these adapted stem cells. Although a role for *ID1* (ref. 61) and *HM13* cannot be excluded, enhanced cell survival due to elevated expression levels of *BCL2L1* offers the most likely mechanism.

The repeated appearance of a structural variant across multiple lines requires both a selective advantage for the variant (e.g., increased expression of *BCL2L1*), and a predisposition for the respective mutation to occur. It is noteworthy that the proximal end of all human ES cell 20q11.21 gains lies within a gap region of the current human assembly[62]. The presumption is that the highly repetitive sequence within this gap predisposes the region to structural rearrangement. With the link between genome rearrangements, primate evolution and disease association[63], it is notable that this gap coincides with a recent chromosomal rearrangement, a pericentric inversion[64], occurring in the last common ancestor of gorilla, chimp and human (**Fig. 6**). The gap region, possibly a centromeric remnant of a tandem duplication[62], introduces the repetitive sequence creating 20q11.21 rearrangement (or amplification) susceptibility. The frequency of appearance that is created by this combination of mutability and the decreased apoptosis warrants routine surveillance similar to that now done in karyotypic analysis.

The identification of genes that drive both cancer progression of EC cells in germ cell tumors and the progressive culture adaptation of ES cells has been a goal since the first clear recognition that gain of sections of the short arm of chromosome 12 is an invariant feature of EC cells[14]. The commonality of the changes in the tumors and in the ES cell in culture suggests common underlying mechanisms. However, the identification of a specific driver gene on chromosomes 1, 12 and 17 has been elusive, suggesting that more than one gene may be involved in the growth advantage of the aneuploid cells. Our present results now point to a specific gene subject to gain, most likely the anti-apoptotic gene, *BCL2L1*, on chromosome 20, that may promote the survival of ES cells *in vitro* and EC cells *in vivo*, thereby providing a strong growth advantage, whether in cancers or *in vitro*.

METHODS

Methods and any associated references are available in the online version of the paper at http://www.nature.com/naturebiotechnology/.

Note: Supplementary information is available on the Nature Biotechnology website.

ACKNOWLEDGMENTS

The International Stem Cell Initiative is funded by The International Stem Cell Forum. The authors would like to acknowledge the following: Medical Research Council, UK (P.W.A., H.M.); Mohammad Pakzad & Adeleh Taei, Royan Institute (H.B., G.H.S.); California Institute for Regenerative Medicine (CIRM) (E.C., P.W.L.); Institute of Medical Biology, A*STAR, Singapore (J.M.C.); Ministry of Education, Youth and Sports of the Czech Republic (P.D., A.H.); Stem Cell Research Center of the 21st Century Frontier Research Program, Ministry of Education, Science & Technology, Republic of Korea (SC-1140) (D.R.L., S.K.O.); Ministry of Science and Technology of China (863 program 2006AA02A102) (L.G.); Swedish Research Council, Cellartis (O.H.); Department of Biotechnology, Government of India, UK-India Education and Research Initiative and the Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India (M.I.); Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation, Leading Project of the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Funding Program for World-Leading Innovative R&D on Science and Technology (FIRST Program) of the Japan Society for the Promotion of Science (JSPS), Grants in-Aid for Scientific Research of JSPS and MEXT (T.I., S.Y., K.T.); Swiss National Science Foundation (grant no. 4046-114410) (M.J.); Shanghai Science and Technology Developmental Foundation (06DJ14001), Chinese Ministry of Science and Technology (2007CB948004) (Y.J.); funding from the North West Science Fund, UK (S.K.); One North East Regional Developmental Agency, Medical Research Council, UK, Newcastle University (M.L.); research funding from the Australian Stem Cell Centre (A.L.L.); The Netherlands Proteomics Consortium grant T4-3 (C.M.); Stem Cell Network, Canada (A.N.); National BioResource Project, MEXT, Japan (N.N.); Singapore Stem Cell Consortium (SSCC) & the Agency for Science Technology and Research (A*STAR) (S.K.W.O., P.R.) and the Genome Institute of Singapore Core Genotyping Lab (P.R.); Academy of Finland, Sigrid Juselius Foundation (T.O.); Conselho Nacional de Desenvolvimento Científico e Tecnológico/Departamento de Ciência e Tecnologia do Ministério da Saúde (CNPq/MS/DECIT), and Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) (L.V.P.); supported by the kind donation of Judy and Sidney Swartz (B.R.); financial support from the Faculty of Medicine, University of New South Wales (UNSW) and the National Health and Medical Research Council (NHMRC) Program Grant no. 568969 (Perminder Sachdev), South Eastern Sydney and Illawarra Area Health Service (SEIAHS) for making hES cell line Endeavour-2 available for this study, and H. Chung and J. Kim for their help in preparing the samples (K.S.); Academy of Finland (grant 218050), the Competitive Research Funding of the Tampere University Hospital (grant 9F217) (H. Skottman).

AUTHOR CONTRIBUTIONS

Project coordination: P.W.A. **Cytogenetic analyses**: D.B., A.D., E.M., K.D.M. and T.G.-L. **Molecular karyotyping by SNP BeadChip**: P.R. **DNA methylation arrays**: R.M.B. and P.W.L. **Administration and data curation**: A. Ford and P.J.G. **Data analysis and manuscript drafting**: P.W.A., S.A., D.B., N.B., R.M.B., P.J.G., K.H., L.H., B.B.K., Y. Mayshar, S.K.W.O., M.F.P. and P.R. **The scientific management of the ISCI project was provided by a steering committee comprising:** P.W.A., N.B., B.B.K., S.K.W.O., M.F.P., J.R. and G.N.S. **Sample contribution**: A. Colman, A. Robins, A. Hampl, A. Bosman, A.M. Fraga, A. Nagy, A.B.H. Choo, A.L. Laslett, A. Feki, A. Kuliev, A. Kresentia Irwanto, B. Reubinoff, B. Sun, C. Denning,

C. Mummery, C. Li, C. Olson, C. Spits, D. Ben-Yosef, D. Collins, D.J. Weisenberger, D. Ryul Lee, D. Ward-van Oostwaard, E. Chiao, E. Sherrer, Fei Pan, F. Holm, G. Anyfantis, G.Q. Daley, G.H. Salekdeh, G. Selva Raj, G. Caisander, H. Gourabi, H. Moore, H. Skottman, H. Suemori, H. Baharvand, H. Shen, I. Mateizel, In-Hyun Park, J. Sheik Mohamed, J. Downie, J. Eun Lee, J.M. Crook, J. Chen, J. Hyllner, J.-C. Biancotti, J. Baker, K. Sermon, K. Amps, K. Narwani, K. Takahashi, K. Sidhu, L. Ge, L.S. Lim, L. Young , Q. Zhou, L. Guangxiu, L.V. Pereira, L. Armstrong, M. Lako, M.S. Inamdar, M.A. Lagarkova, M.B. Munoz, M. Mileikovsky, M.V. Camarasa, M. Jaconi, M. Gropp, N. Lavon, N. Strelchenko, N. Nakatsuji, O. Kopper, O. Hovatta, O. Qi, P. Venu, P.A. De Sousa, P. Dvorak, R. Strehl, R. Suuronen, S. Kiselev, S. Yong Moon, S. Yamanaka, S. Sivarajah, S. Beil, S.L. Minger, S.K.W. Oh, S. Pells, S. Kyung Oh, S. Kimber, T. Miyazaki, T.E. Ludwig, T. Ishii, T.C. Schulz, T. Otonkoski, T. Tuuri, T. Frumkin, V. Kukharenko, V. Fox, W. Herath, Y. Jin, Y. Min Choi, Y. Ma, Y. Wu and Y. Verlinsky.

COMPETING FINANCIAL INTERESTS
The authors declare competing financial interests: details accompany the full-text HTML version of the paper at http://www.nature.com/nbt/index.html.

Published online at http://www.nature.com/nbt/index.html.
Reprints and permissions information is available online at http://www.nature.com/reprints/index.html.

1. Baker, D.E. *et al.* Adaptation to culture of human embryonic stem cells and oncogenesis *in vivo*. *Nat. Biotechnol.* **25**, 207–215 (2007).
2. Draper, J.S. *et al.* Recurrent gain of chromosomes 17q and 12 in cultured human embryonic stem cells. *Nat. Biotechnol.* **22**, 53–54 (2004).
3. Mitalipova, M.M. *et al.* Preserving the genetic integrity of human embryonic stem cells. *Nat. Biotechnol.* **23**, 19–20 (2005).
4. Hoffman, L.M. & Carpenter, M.K. Characterization and culture of human embryonic stem cells. *Nat. Biotechnol.* **23**, 699–708 (2005).
5. Maitra, A. *et al.* Genomic alterations in cultured human embryonic stem cells. *Nat. Genet.* **37**, 1099–1103 (2005).
6. Buzzard, J.J., Gough, N.M., Crook, J.M. & Colman, A. Karyotype of human ES cells during extended culture. *Nat. Biotechnol.* **22**, 381–382, author reply 382 (2004).
7. Caisander, G. *et al.* Chromosomal integrity maintained in five human embryonic stem cell lines after prolonged in vitro culture. *Chromosome Res.* **14**, 131–137 (2006).
8. Inzunza, J. *et al.* Comparative genomic hybridization and karyotyping of human embryonic stem cells reveals the occurrence of an isodicentric X chromosome after long-term cultivation. *Mol. Hum. Reprod.* **10**, 461–466 (2004).
9. Rosler, E.S. *et al.* Long-term culture of human embryonic stem cells in feeder-free conditions. *Dev. Dyn.* **229**, 259–274 (2004).
10. Lefort, N. *et al.* Human embryonic stem cells reveal recurrent genomic instability at 20q11.21. *Nat. Biotechnol.* **26**, 1364–1366 (2008).
11. Spits, C. *et al.* Recurrent chromosomal abnormalities in human embryonic stem cells. *Nat. Biotechnol.* **26**, 1361–1363 (2008).
12. Mayshar, Y. *et al.* Identification and classification of chromosomal aberrations in human induced pluripotent stem cells. *Cell Stem Cell* **7**, 521–531 (2010).
13. Wang, N., Trend, B., Bronson, D.L. & Fraley, E.E. Nonrandom abnormalities in chromosome 1 in human testicular cancers. *Cancer Res.* **40**, 796–802 (1980).
14. Atkin, N.B. & Baker, M.C. Specific chromosome change, i(12p), in testicular tumours? *Lancet* **320**, 1349 (1982).
15. Rodriguez, E. *et al.* Molecular cytogenetic analysis of i(12p)-negative human male germ cell tumors. *Genes Chromosom. Cancer* **8**, 230–236 (1993).
16. Skotheim, R.I. *et al.* New insights into testicular germ cell tumorigenesis from gene expression profiling. *Cancer Res.* **62**, 2359–2364 (2002).
17. Mostert, M. *et al.* Comparative genomic and in situ hybridization of germ cell tumors of the infantile testis. *Lab. Invest.* **80**, 1055–1064 (2000).
18. Schneider, D.T. *et al.* Genetic analysis of childhood germ cell tumors with comparative genomic hybridization. *Klin. Padiatr.* **213**, 204–211 (2001).
19. Looijenga, L.H. *et al.* Comparative genomic hybridization of microdissected samples from different stages in the development of a seminoma and a non-seminoma. *J. Pathol.* **191**, 187–192 (2000).
20. Longo, L., Bygrave, A., Grosveld, F.G. & Pandolfi, P.P. The chromosome make-up of mouse embryonic stem cells is predictive of somatic and germ cell chimaerism. *Transgenic Res.* **6**, 321–328 (1997).
21. Liu, X. *et al.* Trisomy eight in ES cells is a common potential problem in gene targeting and interferes with germ line transmission. *Dev. Dyn.* **209**, 85–91 (1997).
22. Zody, M.C. *et al.* DNA sequence of human chromosome 17 and analysis of rearrangement in the human lineage. *Nature* **440**, 1045–1049 (2006).
23. Werbowetski-Ogilvie, T.E. *et al.* Characterization of human embryonic stem cells with features of neoplastic progression. *Nat. Biotechnol.* **27**, 91–97 (2009).
24. Narva, E. *et al.* High-resolution DNA analysis of human embryonic stem cell lines reveals culture-induced copy number changes and loss of heterozygosity. *Nat. Biotechnol.* **28**, 371–377 (2010).
25. Allegrucci, C. *et al.* Restriction landmark genome scanning identifies culture-induced DNA methylation instability in the human embryonic stem cell epigenome. *Hum. Mol. Genet.* **16**, 1253–1268 (2007).
26. Calvanese, V. *et al.* Cancer genes hypermethylated in human embryonic stem cells. *PLoS ONE* **3**, e3294 (2008).
27. Enver, T. *et al.* Cellular differentiation hierarchies in normal and culture-adapted human embryonic stem cells. *Hum. Mol. Genet.* **14**, 3129–3140 (2005).
28. Rugg-Gunn, P.J., Ferguson-Smith, A.C. & Pedersen, R.A. Epigenetic status of human embryonic stem cells. *Nat. Genet.* **37**, 585–587 (2005).
29. Adewumi, O. *et al.* Characterization of human embryonic stem cell lines by the International Stem Cell Initiative. *Nat. Biotechnol.* **25**, 803–816 (2007).
30. Rugg-Gunn, P.J., Ferguson-Smith, A.C. & Pedersen, R.A. Status of genomic imprinting in human embryonic stem cells as revealed by a large cohort of independently derived and maintained lines. *Hum. Mol. Genet.* **16**Spec No. 2, R243–R251 (2007).
31. Kim, K.P. *et al.* Gene-specific vulnerability to imprinting variability in human embryonic stem cell lines. *Genome Res.* **17**, 1731–1742 (2007).
32. Andrews, P.W. *et al.* The International Stem Cell Initiative: toward benchmarks for human embryonic stem cell research. *Nat. Biotechnol.* **23**, 795–797 (2005).
33. Olariu, V. *et al.* Modeling the evolution of culture-adapted human embryonic stem cells. *Stem Cell Res.* **4**, 50–56 (2010).
34. Martin, G.R. & Evans, M.J. The morphology and growth of a pluripotent teratocarcinoma cell line and its derivatives in tissue culture. *Cell* **2**, 163–172 (1974).
35. Andrews, P.W., Bronson, D.L., Benham, F., Strickland, S. & Knowles, B.B. A comparative study of eight cell lines derived from human testicular teratocarcinoma. *Int. J. Cancer* **26**, 269–280 (1980).
36. Chambers, I. *et al.* Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell* **113**, 643–655 (2003).
37. Mitsui, K. *et al.* The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. *Cell* **113**, 631–642 (2003).
38. Darr, H., Mayshar, Y. & Benvenisty, N. Overexpression of NANOG in human ES cells enables feeder-free growth while inducing primitive ectoderm features. *Development* **133**, 1193–1201 (2006).
39. Korkola, J.E. *et al.* Down-regulation of stem cell genes, including those in a 200-kb gene cluster at 12p13.31, is associated with in vivo differentiation of human male germ cell tumors. *Cancer Res.* **66**, 820–827 (2006).
40. Widschwendter, M. *et al.* Epigenetic stem cell signature in cancer. *Nat. Genet.* **39**, 157–158 (2007).
41. Frazer, K.A. *et al.* A second generation human haplotype map of over 3.1 million SNPs. *Nature* **449**, 851–861 (2007).
42. Li, J.Z. *et al.* Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104 (2008).
43. Abdulla, M.A. *et al.* Mapping human genetic diversity in Asia. *Science* **326**, 1541–1545 (2009).
44. Pritchard, J.K., Stephens, M. & Donnelly, P. Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959 (2000).
45. Novembre, J. *et al.* Genes mirror geography within Europe. *Nature* **456**, 98–101 (2008).
46. Laurent, L.C. *et al.* Dynamic changes in the copy number of pluripotency and cell proliferation genes in human ESCs and iPSCs during reprogramming and time in culture. *Cell Stem Cell* **8**, 106–118 (2011).
47. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res.* **17**, 1665–1674 (2007).
48. Assou, S. *et al.* A meta-analysis of human embryonic stem cells transcriptome integrated into a web-based expression atlas. *Stem Cells* **25**, 961–973 (2007).
49. Peng, J.C. *et al.* Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* **139**, 1290–1302 (2009).
50. Nottke, A., Colaiacovo, M.P. & Shi, Y. Developmental roles of the histone lysine demethylases. *Development* **136**, 879–889 (2009).
51. Morin, R.D. *et al.* Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. *Genome Res.* **18**, 610–621 (2008).
52. Bai, H. *et al.* Bcl-xL enhances single-cell survival and expansion of human embryonic stem cells without affecting self-renewal. *Stem Cell Res. (Amst.)* (in the press).
53. Lee, T.I. *et al.* Control of developmental regulators by Polycomb in human embryonic stem cells. *Cell* **125**, 301–313 (2006).
54. Hussein, S.M. *et al.* Copy number variation and selection during reprogramming to pluripotency. *Nature* **471**, 58–62 (2011).
55. Mosher, J.T. *et al.* Lack of population diversity in commonly used human embryonic stem-cell lines. *N. Engl. J. Med.* **362**, 183–185 (2010).
56. Laurent, L.C. *et al.* Restricted ethnic diversity in human embryonic stem cell lines. *Nat. Methods* **7**, 6–7 (2010).
57. Wu, H. *et al.* Copy number variant analysis of human embryonic stem cells. *Stem Cells* **26**, 1484–1489 (2008).
58. Chia, N.Y. *et al.* A genome-wide RNAi screen reveals determinants of human embryonic stem cell identity. *Nature* **468**, 316–320 (2010).
59. Beroukhim, R. *et al.* The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
60. Ardehali, R. *et al.* Overexpression of BCL2 enhances survival of human embryonic stem cells during stress and obviates the requirement for serum factors. *Proc. Natl. Acad. Sci. USA* **108**, 3282–3287 (2011).
61. Martins-Taylor, K. *et al.* Recurrent copy number variations in human induced pluripotent stem cells. *Nat. Biotechnol.* **29**, 488–491 (2011).
62. Deloukas, P. *et al.* The DNA sequence and comparative analysis of human chromosome 20. *Nature* **414**, 865–871 (2001).
63. Shaw, C.J. & Lupski, J.R. Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. *Hum. Mol. Genet.* **13** Spec No 1, R57–R64 (2004).
64. Misceo, D. *et al.* Evolutionary history of chromosome 20. *Mol. Biol. Evol.* **22**, 360–366 (2005).

Katherine Amps[1], Peter W Andrews[1], George Anyfantis[2], Lyle Armstrong[2], Stuart Avery[3], Hossein Baharvand[4], Julie Baker[5], Duncan Baker[6], Maria B Munoz[7], Stephen Beil[8], Nissim Benvenisty[9], Dalit Ben-Yosef[10,11], Juan-Carlos Biancotti[12], Alexis Bosman[13], Romulo Martin Brena[14], Daniel Brison[15], Gunilla Caisander[16], María V Camarasa[17], Jieming Chen[18], Eric Chiao[5,19], Young Min Choi[20], Andre B H Choo[21], Daniel Collins[22], Alan Colman[3,23], Jeremy M Crook[3,23–26], George Q Daley[27–30], Anne Dalton[6], Paul A De Sousa[22,31], Chris Denning[7], Janet Downie[22], Petr Dvorak[32], Karen D Montgomery[33], Anis Feki[34], Angela Ford[1], Victoria Fox[8], Ana M Fraga[35], Tzvia Frumkin[10], Lin Ge[36], Paul J Gokhale[1], Tamar Golan-Lev[9], Hamid Gourabi[4], Michal Gropp[37], Lu Guangxiu[36], Ales Hampl[38,39], Katie Harron[40], Lyn Healy[41], Wishva Herath[18], Frida Holm[42], Outi Hovatta[42], Johan Hyllner[16], Maneesha S Inamdar[43], Astrid Kresentia Irwanto[18], Tetsuya Ishii[44,73], Marisa Jaconi[13], Ying Jin[45], Susan Kimber[17], Sergey Kiselev[46,47], Barbara B Knowles[3], Oded Kopper[9], Valeri Kukharenko[48], Anver Kuliev[48], Maria A Lagarkova[47], Peter W Laird[14], Majlinda Lako[2], Andrew L Laslett[49,50], Neta Lavon[12], Dong Ryul Lee[51], Jeoung Eun Lee[52], Chunliang Li[53], Linda S Lim[18], Tenneille E Ludwig[33], Yu Ma[53], Edna Maltby[6], Ileana Mateizel[54], Yoav Mayshar[9], Maria Mileikovsky[55], Stephen L Minger[56,57], Takamichi Miyazaki[58], Shin Yong Moon[20], Harry Moore[1], Christine Mummery[59], Andras Nagy[55], Norio Nakatsuji[60], Kavita Narwani[12], Steve K W Oh[21], Sun Kyung Oh[61], Cia Olson[62], Timo Otonkoski[62,63], Fei Pan[14], In-Hyun Park[64], Steve Pells[31], Martin F Pera[65], Lygia V Pereira[35], Ouyang Qi[36], Grace Selva Raj[23], Benjamin Reubinoff[37], Alan Robins[66], Paul Robson[18], Janet Rossant[67], Ghasem H Salekdeh[68], Thomas C Schulz[66], Karen Sermon[54], Jameelah Sheik Mohamed[18], Hui Shen[14], Eric Sherrer[66], Kuldip Sidhu[69], Shirani Sivarajah[23–25], Heli Skottman[70], Claudia Spits[54], Glyn N Stacey[41], Raimund Strehl[16], Nick Strelchenko[48,73], Hirofumi Suemori[58], Bowen Sun[53], Riitta Suuronen[70], Kazutoshi Takahashi[44], Timo Tuuri[62], Parvathy Venu[43], Yuri Verlinsky[48,74], Dorien Ward-van Oostwaard[59], Daniel J Weisenberger[14], Yue Wu[56,57], Shinya Yamanaka[44,60,71,72], Lorraine Young[7] & Qi Zhou[49]

[1]Centre for Stem Cell Biology, Department of Biomedical Science, The University of Sheffield, Sheffield, UK. [2]North East England Stem Cell Institute at Life, International Centre for Life, Newcastle upon Tyne, UK. [3]Institute of Medical Biology, A-STAR, Immunos, Singapore. [4]Royan Institute for Reproductive Biomedicine, Department of Genetics, Tehran, Islamic Republic of Iran. [5]Stanford University, Stanford, California, USA. [6]Sheffield Diagnostic Genetic Services, Sheffield Children's NHS Trust, Sheffield, UK. [7]Wolfson Centre for Stem Cells, Tissue Engineering & Modelling (STEM), Centre for Biomolecular Sciences, University of Nottingham, UK. [8]USC Stem Cell Core Facility, The Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research at USC, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [9]Stem Cell Unit, Department of Genetics, Silberman Institute of Life Sciences, The Hebrew University of Jerusalem, Edmond J. Safra Campus, Givat Ram, Jerusalem, Israel. [10]Racine IVF Unit, Lis Maternity Hospital, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel. [11]Department of Cell Developmental Biology, Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel. [12]Regenerative Medicine Institute, Cedars-Sinai Medical Institute, Los Angeles, California, USA. [13]Department of Pathology and Immunology, Faculty of Medicine, Geneva University, Geneva, Switzerland. [14]USC Epigenome Center, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [15]Department of Reproductive Medicine, St. Marys's Hospital, Central Manchester NHS Foundation Trust, Manchester Academic Health Sciences Centre, Manchester, UK. [16]Cellartis AB, Goteborg, Sweden. [17]Faculty of Life Sciences, University of Manchester, Manchester, UK. [18]Genome Institute of Singapore, Singapore. [19]Hoffmann-LaRoche, Nutley, New Jersey, USA. [20]Department of Obstetrics & Gynaecology, Seoul National University College of Medicine, Seoul, Republic of Korea. [21]Bioprocessing Technology Institute, Singapore. [22]Roslin Cells Ltd., Roslin Biocentre, Roslin, Midlothian, UK. [23]Singapore Stem Cell Consortium, A-STAR, Singapore. [24]Centre for Neural Engineering, The University of Melbourne, Parkville, Australia. [25]Optics and Nanoelectronics Research Group, NICTA Victorian Research Laboratory, The University of Melbourne, Parkville, Australia. [26]Department of Surgery, St. Vincent's Hospital, The University of Melbourne, Fitzroy, Australia. [27]Stem Cell Transplantation Program, Division of Pediatric Hematology/Oncology, Manton Center for Orphan Disease Research, Howard Hughes Medical Institute, Children's Hospital Boston and Dana-Farber Cancer Institute, Boston, Massachusetts, USA. [28]Division of Hematology, Brigham and Women's Hospital, Boston, Massachusetts, USA. [29]Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA. [30]Harvard Stem Cell Institute, Boston, Massachusetts, USA. [31]Medical Research Council Centre for Regenerative Medicine, University of Edinburgh, Edinburgh, UK. [32]Department of Biology, Faculty of Medicine, Masaryk University, Brno, Czech Republic. [33]WiCell Research Institute, Madison, Wisconsin, USA. [34]Department of Obstetrics and Gynecology, Hopital Cantonal Fribourgois, Freibourg, Switzerland. [35]National Laboratory for Embryonic Stem Cell Research (LaNCE), Department of Genetics and Evolutionary Biology, University of São Paulo, São Paulo, Brazil. [36]Institute of Reproductive & Stem Cell Engineering, Central South University, Reproductive & Genetic Hospital CITIC-XIANGYA, Changsha, Hunan, People's Republic of China. [37]The Hadassah Human Embryonic Stem Cell Research Center, The Goldyne Savad Institute of Gene Therapy, Hadassah University Medical Center, Jerusalem, Israel. [38]Department of Histology and Embryology, Faculty of Medicine, Masaryk University, Brno, Czech Republic. [39]Institute of Experimental Medicine ASCR, Prague, Czech Republic. [40]MRC Centre of Epidemiology for Child Health, Institute of Child Health, University College London, London, UK. [41]UK Stem Cell Bank, Division of Cell Biology and Imaging, National Institute for Biological Standards and Control, South Mimms, Herts, UK. [42]Department of Clinical Science, Intervention and Technology, Karolinska Institutet, Karolinska University Hospital Huddinge, Stockholm, Sweden. [43]Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India. [44]Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan. [45]Key Laboratory of Stem Cell Biology, Institute of Health Sciences, Shanghai Institutes of Biological Sciences, CAS/Shanghai JiaoTong University School of Medicine, Shanghai, People's Republic of China. [46]Stem Cell Department, NRC Kurchatov Institute, Moscow, Russia. [47]Vavilov Institute of General Genetics, Moscow, Russia. [48]Reproductive Genetics Institute, Chicago, Illinois, USA. [49]CSIRO Material Science and Engineering, Clayton, Australia. [50]Department of Anatomy and Developmental Biology, Monash University, Clayton, Australia. [51]Department of Biomedical Science, CHA Stem Cell Institute, CHA University, Gangnam-gu, Seoul, Republic of Korea. [52]CHA Stem Cell Institute, CHA University, Gangnam-gu, Seoul, Republic of Korea. [53]Shanghai Stem Cell Institute, Shanghai JiaoTong University School of Medicine, Shanghai, People's Republic of China. [54]Department of Embryology and Genetics, Vrije Universiteit Brussel, Brussels, Belgium. [55]Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario, Canada. [56]Wolfson Centre for Age-Related Diseases, King's College London, London, UK. [57]GE Healthcare, Cardiff, UK. [58]Laboratory of Embryonic Stem Cell Research, Stem Cell Research Center, Institute for Frontier Medical Sciences, Kyoto University, Kyoto, Japan. [59]Department of Anatomy & Embryology, Leiden University Medical Center, Leiden, The Netherlands. [60]Institute for Integrated Cell-Material Sciences, Kyoto University, Ushinomiya-cho, Yoshida, Sakyo-ku, Kyoto, Japan. [61]Institute of Reproductive Medicine & Population, Medical Research Center, Seoul National University, Seoul, Republic of Korea. [62]Research Programs Unit, Molecular Neurology, Biomedicum Stem Cell Centre, University of Helsinki, Finland. [63]Children's Hospital, University of Helsinki and Helsinki University Central Hospital, Finland. [64]Department of

Genetics, Yale Stem Cell Center, Yale School of Medicine, New Haven, Connecticut, USA. [65]The Eli and Edythe Broad Center for Regenerative Medicine and Stem Cell Research at USC, Keck School of Medicine, University of Southern California, Los Angeles, California, USA. [66]Viacyte, Athens, Georgia, USA. [67]Program for Developmental Biology, The Hospital for Sick Children, Toronto, Ontario, Canada. [68]Department of Molecular Systems Biology, Cell Science Research Center, Royan Institute for Stem Cell Biology and Technology, ACECR, Tehran, Islamic Republic of Iran. [69]Stem Cell Laboratory, Faculty of Medicine, University of New South Wales, Australia. [70]Institute for Regenerative Medicine, University of Tampere, Tampere, Finland. [71]Yamanaka iPS Cell Special Project, Japan Science and Technology Agency, Kawaguchi, Japan. [72]Gladstone Institute of Cardiovascular Disease, San Francisco, California, USA. [73]Present addresses: Japan Science and Technology Agency, Tokyo, Japan (T. Ishii) and Department of Obstetrics and Gynecology, New York University Langone Medical Center, New York, New York, USA (N. Strelchenko). [74]Deceased.

ONLINE METHODS

Design of study. Laboratories that agreed to contribute cell lines to the study were asked to provide four preparations of DNA from a culture of each cell line to be analyzed, two preparations from as early a passage as possible and two from as late a passage as possible (**Supplementary Note 1**, protocol for DNA preparation). The samples were coded and shipped directly, one pair of early- and late-passage DNA for SNP analysis to the Genome Institute, A*STAR, Singapore, and one pair of samples for DNA methylation analysis to the University of Southern California. The laboratories were also asked to prepare fixed hypotonic-treated samples of cells at both passage levels, following culture in colcemid, according to a specified protocol (**Supplementary Note 2**). These samples were shipped to the University of Sheffield for preparation and karyotypic analysis of metaphase spreads. Ideally the laboratories were asked to provide the material for both DNA analyses and the cytogenetic analysis from the same cultures of the cells. In a few cases this could not be done, for example, when a first sample did not meet quality control criteria and additional samples had to be prepared, but all samples were provided from as close a passage as possible. The full details of each cell line included in the study, the passage levels of each sample and culture details, including split ratios, culture time between passages, and medium and subculture methodology, are provided in **Supplementary Table 1**.

Karyology. Chromosome analysis was carried out in accord with the general principles developed by the International Stem Cell Banking Initiative[65]. Briefly, cells were to be cultured in the presence of 0.1 μg/ml colcemid (Invitrogen) for up to 4 h, followed by dissociation with 0.25% trypsin/versene (Invitrogen). The cells were pelleted by centrifugation and resuspended in prewarmed 0.0375 M KCl hypotonic solution and incubated for 10 min. After centrifugation the cells were resuspended in fixative (3:1 methanol/acetic acid) and sent to the central karyotyping facility, where metaphase spreads were prepared on glass microscope slides and G-banded by brief exposure to trypsin and stained with 4:1 Gurr's/Leishmann's stain (Sigma-Aldrich). (See **Supplementary Note 2** for detailed protocols.) A minimum of 10 metaphase spreads were analyzed in detail and a further 20 counted and scored from both the early- and late-passage cultures. A 30-cell examination can exclude mosaicism at the 10% level with 95% confidence[66]. Analysis was performed by a Health Professionals Council registered Clinical Cytogeneticist in a CPA accredited laboratory (Duncan Baker DipRCPath, Sheffield Diagnostic Genetic Services, Sheffield Children's NHS Foundation Trust and the CSCB, University of Sheffield). A representative image of each cell line was captured using Applied Imaging's Cytovision system. All abnormal karyotype results were confirmed by two other experienced analysts (Tamar Golan-Lev, Hebrew University, Jerusalem, and Karen Montgomery, WiCell).

Cell lines were described as abnormal if at least two cells were found with the same chromosome aberration. Fluorescent *in situ* hybridization (FISH) was used to characterize abnormalities when appropriate. Abnormalities confined to a single cell were recorded (**Supplementary Table 2**). For those in which the single cell abnormality involved chromosomes 1, 12, 17 or 20, interphase FISH analysis with appropriate probes (**Supplementary Table 2**) was performed to confirm or exclude low level mosaicism. FISH analysis was also performed, on the early-passage cultures of those cell lines with normal early cultures and abnormal late cultures to confirm or exclude low-level mosaicism in the early passage.

SNP array analysis. To analyze genomic structural variations below the resolution of standard chromosomal banding analysis, we chose the Illumina 1M Quad SNP array technology. In addition to providing the ability to detect structural variants by virtue of appropriately spaced invariant genetic features, the SNP features allowed characterization of the population structure between all human ES cell lines analyzed. The samples received were run on the Illumina 1M Quad platform and the data were subjected to structural variant quality control (QC) assessment in which the minimum number SNP/structural variant = 10, the minimum length of structural variant = 1 kb, the minimum confidence score for structural variant call = 10, the Log R Ratio s.d. < 0.35, the B allele freq s.d. < 0.06 and the waviness factor −0.04 < WF < 0.04.

PLINK[67] was used to detect genetically related samples (early- and late-passage paired samples and other sibling relationships) and the related samples with the lower call-rate were removed. QC was performed on these samples, by excluding structural variant probes, nonautosomal SNPs, and samples and SNPs with call-rates <95%. Thereafter, 982,351 SNPs from 114 unrelated samples were merged with three publicly available data sets: HGDP, Hapmap Phase 2 (release 23) and the Pan-Asian SNP Initiative (PASNPI). Additional QC was then performed on the merged data set by exclusion of samples and SNPs with call-rates <95% and monomorphic SNPs. Furthermore, we excluded the extremely diverse southeast-Asian samples in PASNPI, to focus the analyses on human ES cell origins in Africa, Europe, Central-south Asia and East Asia. Ultimately, the merged data set contained 1,967 samples and 11,279 SNPs. Subsequently, a series of PC analyses (PCAs) were performed. The PCAs were performed using smartpca from the software EIGENSTRAT[68] (found in EIGENSOFT, http://genepath.med.harvard.edu/~reich/Software.htm).

The first PCA was performed on the entire final merged data set. The scatter plot of PC2 against PC1 of this PCA was then dissected into four sectors, based primarily on the global regions of Africa (PC2 < −0.04), Middle East-Europe-Central-South Asia (PC1 > 0.005 and PC2 > 0.008), Oceania-America-Central-South Asia (−0.015 < PC1 < 0.005) and East Asia (PC1 < −0.015). (**Fig. 1a**). PCAs were then reperformed on the samples found in these four sectors. Notably, most of the human ES cell samples were concentrated in the European sector (with the HGDP-Europe and HapMap European cluster).

Hence, to further elucidate the origins of the human ES cell lines in Europe, we removed 16 human ES cell European outliers in the Central-south Asia-Europe-Middle East sector (**Fig. 1d**); these are considered to be of dubious ancestry, suspected to originate from two groups: Central-south Asian or Southern European or mixed from both; Middle Eastern or Eastern European or mixed from both. Lack of reference populations and insufficient SNPs could be possible reasons for this uncertainty in determination of ancestry. The remaining human ES cell samples (found in the European cluster) were then merged with 1,385 samples and 168,352 SNPs from the POPRES European data set[45]. (The sample and SNP lists were obtained through correspondence with the author.) This second merged data set between human ES cell and POPRES European data yielded 1,448 samples and 55,972 SNPs, after excluding samples and SNPs with call rates <95% and monomorphic SNPs. PCA was then performed again on this data set (**Fig. 1b**). Color coding was deliberately implemented in a similar manner to that of reference [45].

The human ES cell lines were classified into broad categories of European, Asian, African and other ancestries and then further subclassified into ethnicities listed in **Table 1**. Human ES cell samples that are full siblings of the 114 samples were included, resulting in a total of 120 human ES cell lines.

Frequency and mapping analysis of SNP data. The PennCNV[47] Hidden-Markov model algorithm was used to identify structural variants in the ISCI samples with human hg18 as the reference genome. High-quality structural variant calls were filtered as follows: first, samples were checked for overall quality using the following criteria from the PennCNV output: 0.01 < BAF_drift < 0.01; −0.05 < WF < 0.05; LRR_SD < 0.35. Second, individual structural variant calls for samples passing QC were assessed as follows: minimum SNPs per structural variant > 10; structural variant Length > 1K; PennCNV confidence threshold > 10. Most samples that failed quality control either exhibited extensive karyotypic abnormalities and/or bad SNP call rates, both of which could contribute to difficulties in structural variant detection. The samples that did pass quality control identified 39,926 deletions of an average size of 23.1 kb and 14,351 duplications of an average size of 117.4 kb in length. These size and total number differences between deletions and duplications is consistent with previous structural variant studies of human populations[47]. The sensitivity of detection was in the 5–10% range based on CNV calls in regions identified as amplified by karyotype analysis.

DNA methylation analysis. 1 μg of genomic DNA was treated with sodium bisulfite using the Zymo EZ-96 DNA Methylation Kit (Zymo Research) following the manufacturer's recommended protocol. DNA methylation measurements were generated using the Illumina GoldenGate DNA methylation platform as previously described[69], at the University of Southern California Epigenome Center production facility. The sequences assayed were previously identified[53] as targets of SUZ12, a subunit of the Polycomb Repressive Complex 2, in the nonrepetitive portion of the genome in human

ES cells (**Supplementary Table 4** array loci, sequences, etc). DNA methylation measurements (β-values) were generated for 1,452 autosome and 84 X-linked loci.

To obtain a global view of the DNA methylation changes undergone by each cell line pair from the early to the late passage, CDF curves were generated. CDF curves were chosen because they are able to concomitantly capture and visually represent the changes detected in all loci analyzed in a single line. To test whether increased passage number correlated with increase in the number of loci gaining DNA methylation, we calculated the difference in passage number between the early and the late member of each cell line pair, ranked these differences and divided the samples into quartiles based on passage-number difference. To investigate whether sibling lines behaved similarly in terms of DNA methylation changes, we calculated the number of loci with an absolute DNA methylation difference ≥10% between the early and late member of each pair of sibling lines. We then compared whether the total number of loci with the specified absolute change in DNA methylation differed significantly from the number of loci undergoing a similar change in pairs of unrelated lines using a Wilcoxon signed-rank test. The data were analyzed using the R statistical programming package version 2.12.1 (http://www.r-project.org/).

Cell line availability. Supplementary Table 5 provides contact information and conditions of availability for the cell lines described in this study.

65. The International Stem Cell Banking Initiative. Consensus guidance for banking and supply of human embryonic stem cell lines for research purposes. *Stem Cell Rev.* **5**, 301–314 (2009).
66. Hook, E.B. Exclusion of chromosomal mosaicism: tables of 90%, 95% and 99% confidence limits and comments on use. *Am. J. Hum. Genet.* **29**, 94–97 (1977).
67. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
68. Price, A.L. *et al.* Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
69. Bibikova, M. *et al.* High-throughput DNA methylation profiling using universal bead arrays. *Genome Res.* **16**, 383–393 (2006).

CHEMICAL REVIEWS

Review

pubs.acs.org/CR

Biomimetic Cell Culture Proteins as Extracellular Matrices for Stem Cell Differentiation

Akon Higuchi,*[,†,‡,§] Qing-Dong Ling,[§,∥] Shih-Tien Hsu,[⊥] and Akihiro Umezawa[‡]

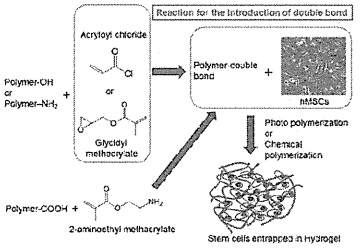
[†]Department of Chemical and Materials Engineering, National Central University, Jhongli, Taoyuan, 32001 Taiwan

[‡]Department of Reproductive Biology, National Research Institute for Child Health and Development, 2-10-1 Okura, Setagaya-ku, Tokyo 157-8535, Japan

[§]Cathay Medical Research Institute, Cathay General Hospital, No. 32, Ln 160, Jian-Cheng Road, Hsi-Chi City, Taipei, 221 Taiwan

[∥]Institute of Systems Biology and Bioinformatics, National Central University, No. 300, Jhongda RD., Jhongli, Taoyuan, 32001 Taiwan

[⊥]Taiwan Landseed Hospital, 77 Kuangtai Road, Pingjen City, Tao-Yuan County, 32405 Taiwan



CONTENTS

1. INTRODUCTION

Each year, millions of people suffer loss or damage to organs and tissues due to accidents, birth defects, and disease. Stem cells are an attractive prospect for tissue engineering and regenerative medicine because of their unique biological properties. Embryonic stem cells (ESCs) derived from

Received: January 14, 2012
Published: May 23, 2012

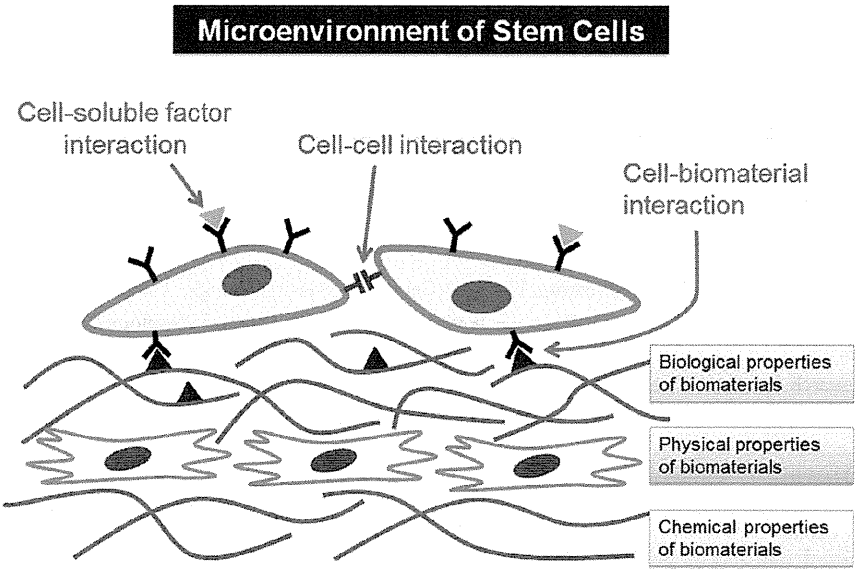


Figure 1. Schematic representation of the microenvironment and niches of stem cells and their regulation by the following factors: (a) soluble factors, such as growth factors or cytokines, nutrients, and bioactive molecules; (b) cell–cell interactions; (c) cell–biomaterial interactions. Biological, physical, and chemical properties of biomaterials also regulate stem cell fate.

preimplantation embryos have the potential to differentiate into any cell type derived from the three germ layers—the ectoderm (epidermal tissues and nerves), mesoderm (muscle, bone, and blood), and endoderm (liver, pancreas, gastrointestinal tract, and lungs).[1] The basis of pluripotency lies in conserved regulatory networks composed of numerous transcription factors and multiple signaling cascades. Together, these regulatory networks maintain human ESCs (hESCs) in a pluripotent and undifferentiated state, and alterations in the stoichiometry of these signals promote differentiation. hESCs have been shown to generate multipotent stem and progenitor cells in vitro and are capable of differentiating into a limited number of cell fates, and thus they have great potential for use in transplantation of cells and tissues into patients.[2]

Although hESCs are promising donor sources for cell transplantation therapies,[1] they face immune rejection after transplantation. Furthermore, ethical issues regarding human embryos hinder their widespread usage. These concerns can be circumvented if pluripotent stem cells can be derived directly from patients' own somatic cells.[3] Recently, pluripotent stem cells similar to ESCs, known as induced pluripotent stem cells (iPSC's), were derived from adult somatic cells by inducing a "forced" expression of certain pluripotent (stem cell) genes[4−6] such as Oct3/4, Sox2, (c-myc), and klf-4, or certain miRNAs[7] or proteins (piPS).[8] iPSC's are believed to be similar to ESCs in many respects, including the expression of certain stem cell genes and proteins, chromatin methylation patterns, doubling time, embryoid body formation, teratoma formation, viable chimera formation, pluripotency, and differentiability.

The pluripotent nature of iPSC's opens many avenues for potential stem cell-based regenerative therapies and for development of drug-discovery platforms.[9,10] The nearest-term therapeutic uses of iPSC's may exist in the transplantation of differentiated nerve cells or β-cells for treatment of Parkinson's Disease and diabetes, respectively, which arise from disorders of single cell types. However, there are several barriers to the clinical application of iPSC's, such as the use of viral vectors, cultivation using xeno-derived materials [e.g., mouse embryonic fibroblasts (MEFs)], and the extremely low efficiency of iPSC generation.[11]

Stem cells have also been isolated from a variety of somatic tissues, including hematopoietic stem cells (HSCs) derived from umbilical cord blood and mesenchymal stem cells (MSCs) derived from bone marrow, umbilical cord blood, umbilical cord, dental pulp, and tissues such as fat. There have been no reports to date of MSCs or fetal stem cells differentiating into tumors, unlike ESCs and iPSC's. Consequently, HSCs, MSCs, and fetal stem cells are the most promising sources of cells for tissue engineering and cell therapies. Currently, MSCs are thought to be the most widely available autologous source of stem cells for practical and clinical applications. Fetal stem cells derived from amniotic fluid are pluripotent cells capable of differentiating into multiple lineages, including cell types of the three embryonic germ layers. Bone marrow MSCs, adipose-derived stem cells (ADSCs), and amniotic fluid stem cells may be more suitable sources of stem cells in regenerative medicine and tissue engineering than ESCs and iPSC's because of ethical concerns regarding their use and concerns about xenogenic contamination arising from the use of mouse embryonic fibroblasts (MEFs) as a feeder layer for ESC and iPSC culture.[11]

Stem cell characteristics, such as proper differentiation and maintenance of pluripotency, are regulated not only by the stem cells themselves but also by the microenvironment. Therefore, mimicking stem cell microenvironments and niches using biopolymers will facilitate the production of large numbers of stem cells and specifically differentiated cells needed for in vitro regenerative medicine. Several factors in the microenvironment and niches of stem cells influence their fate: (i) soluble factors, such as growth factors or cytokines, nutrients, and bioactive molecules; (ii) cell–cell interactions; (iii) cell–biomacromolecule (or biomaterial) interactions; and (iv) physical factors, such as the rigidity of the environment (Figure 1). Some excellent review articles addressing the

engineering of stem cell microenvironments and niches using natural and synthetic biopolymers are listed in Table 1.[11−22]

Table 1. Key Review and Articles Dealing with Biopolymers for Culture and Differentiation of Stem and Progenitor Cells

author	contents	ref (year)
Lee and Mooney	hydrogels for tissue engineering	12 (2001)
Little et al.	biomaterials for neural stem cell microenvironments	13 (2008)
Higuchi et al.	polymeric materials for ex vivo expansion of HSCs	16 (2009)
Mei et al.	combinatorial development of biomaterials for clonal growth of human pluripotent stem cells	17 (2010)
Melkoumian et al.	synthetic peptide-acrylate surfaces for long-term self-renewal of hESCs	18 (2010)
G. J. Delcroix et al.	adult cell therapy for brain neuronal damages and the role of tissue engineering	22 (2010)
Higuchi et al.	biomaterials for the feeder-free culture of hESCs and human iPSC's	11 (2011)
Balakrishnam and Banerjee	biopolymer-based hydrogels for cartilage tissue engineering	14 (2011)
Kim et al.	design of artificial extracellular matrices for tissue engineering	15 (2011)
Engler et al.	matrix elasticity directs stem cell lineage	19 (2006)
Gilbert et al.	substrate elasticity regulates skeletal muscle stem cell self-renewal	20 (2010)
Huebsch et al.	harnessing traction-mediated manipulation of the cell/matrix interface to control stem-cell fate	21 (2010)

These articles focus on biopolymers employed for maintenance of pluripotency of hESCs, iPSC's, or hematopoietic stem cells (HSCs),[16−18] and for specific differentiation lineages such as chondrocytes (cartilage), muscle cells, and neural cells.[13,14,20] There have been no review articles specifically describing extracellular matrix (ECM) scaffolds (ECM in 3D) or ECM-immobilized dish coatings (ECM in 2D) that guide stem cell fates and differentiation. Therefore, this review focuses on the chemical, physical, and biological characteristics of natural biopolymers, especially ECM proteins, which are the major functional biopolymers, and deals with the ability of these biopolymers to guide differentiation of MSCs into osteogenic, chondrogenic, adipogenic, cardiomyogenic, and neural cell lineages.

2. CELL SOURCES AND ANALYSIS OF DIFFERENTIATION LINEAGES OF MSCS

2.1. Cell Sources

Human MSCs (hMSCs), including fetal stem cells, are one of the most widely available autologous sources of stem cells for clinical applications. hMSCs can be obtained from bone marrow,[23,24] adipose tissue,[25,26] dental pulp,[27] and urine,[28] among other sources. Fetal stem cells can be obtained from amniotic fluid,[29−31] umbilical cord,[32−34] menstrual blood,[35,36] umbilical cord blood,[25,34,37] and placenta.[38,39] hMSCs derived from bone marrow and fat are primarily used for biomaterials research on stem cell culture and differentiation because bone marrow MSCs and ADSCs are easily accessible and can be obtained in large quantities. Bone marrow MSCs (BMSCs) are now commercially available from several companies. Stem cell research is facilitated with these stem cell sources because it is not necessary to obtain permission from ethics committees of the Institutional Review Board (IRB) for use of commercially available MSCs. Otherwise, informed consent from donors and permission from the IRB must be obtained.

2.2. Analysis of Differentiation Lineages

MSCs are multipotent stem cells that can be differentiated into various mesodermal lineages, including osteoblasts, chondrocytes (cartilages), adipocytes, myocytes, and cardiomyocytes.[19,40,41] MSCs are also reported to be able to differentiate into ectodermal lineages (e.g., neuron, oligodendrocyte, astrocyte, neural stem cells, and dopamine-secreting cells)[22,42−45] and endodermal lineages (hepatocytes and β-cells),[31,46−52] although with lower probability than mesoderm lineages. Table 2 summarizes methods for characterizing specific differentiated cells from MSCs.[11,34,46,48,51−87]

MSCs differentiate into an osteogenic phenotype in vitro when supplements such as ascorbic acid, β-glycerophosphate, dexamethasone, and/or bone morphogenic protein 2 (BMP-2) are added to the culture medium. Figure 2 shows the expression of several genes and proteins, as well as mineral deposition, by MSCs upon osteogenic differentiation. Runt-related transcription factor 2 (Runx2, also known as Cbfa1, Pebp2αA, and AML3) is a master regulator of osteogenic gene expression and osteoblast differentiation, and it is an early marker of osteogenesis.[88−90] Runx2 activity is stimulated by mitogen-activated protein kinase (MAPK) signaling and is negatively regulated by thrombin-like enzyme 2 (TLE2). Alkaline phosphatase (ALP) activity is an early osteogenic marker, and osteopontin and osteocalcin are late osteogenic markers.[88] Mineral deposition is generated in the late stage of osteogenic differentiation and is detected by Alizarin Red staining (calcium deposition) and von Kossa staining (calcium phosphate deposition).[57,60,62]

MSCs commit to a chondrogenic phenotype when supplied with transforming growth factor-β1 (TGF-β1). Chondrogenic differentiation of MSCs is typically determined by immunostaining for specific proteins, such as collagen type II and Sox9, dye labeling of glycosamino glycans, and evaluation of expression of chondrogenic proteins or transcription factors (such as collagen type II and type X, cartilage oligomeric protein, aggrecan, and Sox9) (Table 2).[63,64,67,70,91] Sulfated glycosaminoglycans (sGAG's) are visualized by staining with Alcian blue.[91] Accumulation of sulfated proteoglycans are also visualized by Safranin O staining.[72]

Only a few groups have investigated adipogenic differentiation of MSCs cultured on natural and artificial biomaterials[53,62,70,74,75,92] because adipose tissue is in less demand in clinical usage than osteoblasts and cartilage cells. Adipogenic differentiation is also analyzed by immunostaining for specific proteins (vimentin), dye staining of oil droplets, and measuring expression of transcription factors or other marker proteins, such as peroxisome proliferator-activated receptorγ [PPARγ] and adipocyte Protein 2 (aP-2).[53,61,62,74,75,92] aP-2 is a carrier protein for fatty acids that is primarily expressed in adipocytes.[93] Preadipocytes and mature adipocytes contain multiple or single lipids in cell bodies, respectively. Therefore, Oil Red O or Nile red staining of preadipocytes and mature adipocytes is frequently used for the detection of lipids.

Neural differentiation of MSCs is primarily analyzed by observing characteristic morphologies of neurons, astrocytes, oligodendrocytes, and microglia. Neuronal progenitor cells and early-stage neurons are also identified by Sox1, Sox2, and CD133 gene expression and by nestin and β-tubulin-III

dx.doi.org/10.1021/cr3000169 | Chem. Rev. 2012, 112, 4507−4540

Table 2. Characterization of Differentiation of MSCs into Specific Lineages [Osteoblasts and Chondrocyte (Cartilages)]

differentiation lineage	characterization	specification	ref (example)
1. Osteoblast	morphology	spread shape tends to differentiate into osteoblasts, bonelike nodule formation	53−55
	protein level (immunostaining)	**collagen I**, osteocalcin, osteonectin	56, 57
	surface marker analysis and immunostaining	osteopontin, bisphosphonate [2-(2-pyridinyl)ethylidene-BP] (PEBP), alkaline phosphatase (ALP)	34, 58
	enzyme activity	**alkaline phosphatase**	
	gene level	**runt-related transcription factor 2** [Runx2 or core binding protein A-1 (CBFA-1)], **osterix** (OSX), **osteocalcin** (OCN), **osteopontin** (OPN), **bone sialoprotein** (BSP), alkaline phosphatase, integrin-binding sialoprotein (IBSP), bone γ-carboxyglutamate protein (BGLAP)	34, 58−61
	dye staining	**Alizarin Red staining** (calcium)	62
	mineral deposition	**von Kossa staining** (calcium phosphate)	57, 60
2. Chondrocytes	protein level (immunostaining)	**collagen type II** (Col II), **collagen type X** (Col X), **aggrecan** (AGN), Sox-9, chondroitin-4-sulfate, chondroitin-6-sulfate, sulphated glycosaminoglycans	56, 57, 63−68
	glycosaminoglycan assay	glycosaminoglycan content	
	dimethylmethylene blue (DMMB) assay	proteoglycan (PG) content	69
	hydroxyproline assay	collagen content	65
	gene level	**collagen II**, collagen IX (Col IX), **collagen X**, collagen XI (Col XI), **aggrecan**, Sox 5, Sox 6, **Sox 9**, **cartilage oligomeric protein** (COMP), xylosyltransferase I (XT-1), α-4-*N*-acetylhexosaminyltransferase (EXTL2), β-1,4-*N*-acetylgalactosaminyltransferase (GalNAcT), glucuronyl C5 epimerase (GlcACSE)	63, 64, 67, 70−73
	dye staining	**Safanin O staining (proteoglycan)**, **Alcian blue staining (proteoglycan)**, EVG-staining, Masson's trichrome staining	34, 62, 64, 67, 70, 72
3. Adipocytes	morphology	round shape cells tends to differentiated into adipocytes	53, 54
	protein level	vimentin, adipocyte lipid-binding protein (ALBP)	53, 74
	enzyme activity	glycerol-3-phosphate dehydrogenase activity	75
	gene level	PPARγ, aP-2	61
	staining	Oil red O and Nile red staining for lipid droplet	62
4. Neural cells	morphology	neuronal-like cells having long neurites	76
	protein level	nestin, neuron-specific class III β-tubulin (TuJ1), galactosylceramidase (GalC), glial fibrillary acidic protein (GFAP), β-tubulin-III, microtubule-associated protein 2 (MAP2), O4, tyrosine hydroxylase (TH), neurofibromatosis (NFM), neurone-specific enolase (NSE)	76−81
	gene level	nestin, Musashi 1, neuron-specific class III β-tubulin (TuJ1), glial fibrillary acidic protein, microtubule-associated protein 2, Sox1, Sox2, CD133, tyrosine hydroxylase, neurofibromatosis, Nurr1, dopamine transporter (DAT), dihydropyrimidinase-related protein 2 (DRP-2), purine-sensitive aminopeptidase (PSA)	11, 61, 76, 81, 82
5. Cardiomyocytes	morphology	contractile cells	
	protein level	**cardiac troponin T (*cTnT*)**, **desmin**, *myosin light chain (MLC)*, *myosin heavy chain (MHC)*	81
	gene level	Nkx2.5, GATA-4, MYH-6, TNNT2, TBX-5, myosin light chain (Mlc2a, MLC-2 V), tropomyosin, cTnI, ANP, **desmin, myosin heavy chain** (α-MHC, **β-MHC**), **cardiac troponin T**, Isl-1, and Mef2c	11
	electrocardiogram	electrocardiogram	
6. Smooth muscle cells	protein level	α-smooth muscle actin (ASMA), h1-calponin (CALP), SM2	83
	gene level	α-smooth muscle actin, h1-calponin, caldesmon, Smemb, SM22α, SM1, SM2	83
7. Epidermis	protein level	keratin 10 (early marker), filaggrin (intermediate marker), involucrin (late marker)	84
	gene level	keratin 10 (early marker), filaggrin (intermediate marker), involucrin (late marker)	84
8. Hepatocyte	morphology	oval cell morphology, small round cell morphology	46

dx.doi.org/10.1021/cr3000169 | Chem. Rev. 2012, 112, 4507−4540

Table 2. continued

differentiation lineage	characterization	specification	ref (example)
	protein level	CXCR4 (endoderm), α-fetoprotein (AFP), albumin (ALB), asialoglycoprotein receptor (ASGPR), cytochrome P450 (CYP_1A_1), hepatocyte nuclear factor-1α (HNF-1α), hepatocyte nuclear factor-3β (HNF-3β), hepatocyte nuclear factor-4α (HNF-4α), CCAAT-enhancer binding protein α (C/EBPα), cytokeratin-18 (CK18), cytokeratin-19 (CK19), low-density lipoprotein (LDL), GATA4	46, 51, 52, 86, 87, 113
	gene level	Sox17 (endoderm), Foxa2 (endoderm), Gata6 (endoderm), α-fetoprotein, albumin, hepatocyte nuclear factor-1α, hepatocyte nuclear factor-3β, hepatocyte nuclear factor-4α, cytokeratin 18, cytokeratin-19, asialoglycoprotein receptor, tryptophan oxygenase (TO), cytochrome P450 (CYP1A1, CYP2B6), CCAAT-enhancer binding protein α, glucose 6-phosphate (G6P), GATA4	46, 51, 52, 86, 87, 113
	urea assay	urea production	46, 51, 113
	albumin assay	albumin production	52, 86, 113
	glycogen assay	glycogen production	46, 52, 113
	α-fetoprotein assay	α-fetoprotein production	52, 86
	pentoxyresorufin (PROD) assay	cytochrome P450 activity	113
	staining	periodic acid–Schiff (PAS) staining for glycogen storage	46, 113

immunostaining. Mature neurons express neuron-specific class III β-tubulin (Tuj1), microtubule-associated protein 2 (MAP2), neuron-specific enolase (NSE), and purine-sensitive aminopeptidase (PSA). Oligodendrocytes express galactosylceramidase (GalC) and O4. Dopaminergic neurons express tyrosine hydroxylase (TH), neurofibromatosis (NFM), and dopamine transporter (DAT). Nerve cells are electrically excitable cells that transmit information by electrical and chemical signaling. Therefore, electrical and action potentials in nerve cells can be monitored using electrodes.

3. PREPARATION OF CULTURE MATRIX

Biomimetic stem cell cultures can be categorized as two-dimensional (2D) or three-dimensional (3D). 2D culture is useful for basic research to investigate the fundamental interactions between cells and immobilized nanosegments on dishes, but 3D culture of stem cells in biomaterials is essential for clinical applications. Figure 3 shows some examples of biomaterial designs for carrying stem cells, as well as direct injection of biomaterials without cells. The injection of hydrogels or scaffolds containing stem cells is categorized as 3D cultures. Cell sheets prepared on a surface-grafting polymer having low critical solution temperature (LCST), such as poly(*N*-isopropylacrylamide) (poly(NIPAM)), can be prepared on 2D dishes.[94,95] Recently, patch sheets of immobilized antibodies or ligands targeting specific stem cells, which recruit the stem cells from the patient's body, are reported to be effective in gathering autologous stem cells at sites of injury.[40] The following sections describe methods for (a) surface immobilization of ECM proteins and ECM-mimicking peptides on 2D culture dishes and (b) preparing hydrogels or scaffolds containing ECM proteins and ECM-mimicking peptides for 3D culture of stem cells.

3.1. ECM Immobilization on 2D Dishes

Typically, 2D cell culture dishes are coated with ECM proteins or ECM-mimicking peptides. Tables 3 and 4 show examples of the ECM proteins and ECM-mimicking peptides used to coat culture dishes and their binding sites on stem cells.[16,18,53,58,71,83,91,96−118] Collagen types I, II, and IV, gelatin, laminin, laminin-1, laminin-5, vitronectin, and fibronectin are typically used as coating materials.[58,71,83,91,96−98,100−102] ECM-mimicking peptides (e.g., RGD, DGEA, YIGSR, IKVAV, KRSR, P15, and GFOGER) are commonly used as coating or grafting materials.[16,18,53,97,103−118] Covalent binding is preferable for long-term effects in culture, but noncovalent coating is the simplest method for the preparation of dishes with immobilized ECM proteins or ECM-mimicking peptides. Figure 4 summarizes typical surface reactions for the covalent immobilization of ECM proteins and peptides on dishes. Proteins and ECM-mimicking peptides should be used in aqueous solution, as they are unstable biomolecules. Reactions between amino groups and between amino groups and carboxylic acids can be used to bind ECM proteins and ECM-mimicking peptides to plastic dishes. These plastic surfaces should therefore have amino groups, carboxylic acid groups, or hydroxyl groups to bind and immobilize ECM proteins or peptides. For dishes made of polyesters, such as poly(ε-caprolactone) (PCL), poly(glycolic acid) (PGA), poly(lactic acid) (PLA), or poly(lactic acid-*co*-glycolic acid) (PLGA), treatment with a diamine, such as hexamethylene diamine, generates amino groups on the surface by an aminolysis reaction. Then, ECM proteins and ECM-mimicking

dx.doi.org/10.1021/cr3000169 | *Chem. Rev.* 2012, 112, 4507–4540

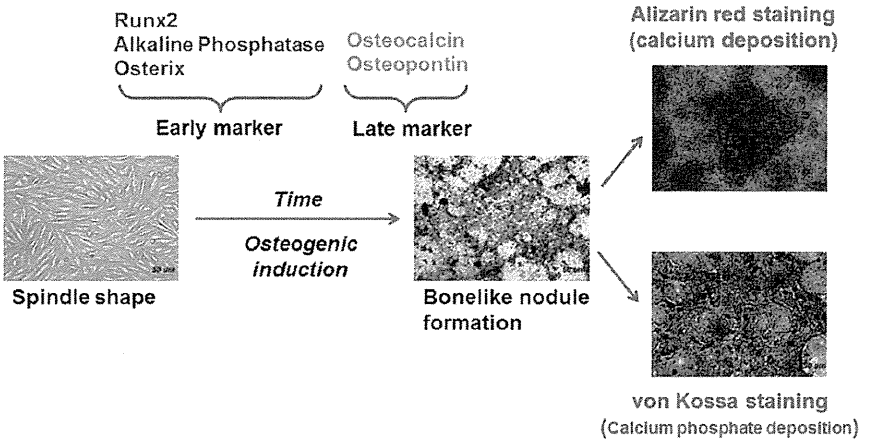


Figure 2. Osteogenic differentiation of MSCs, gene expression, and mineral deposition at early and late stages.

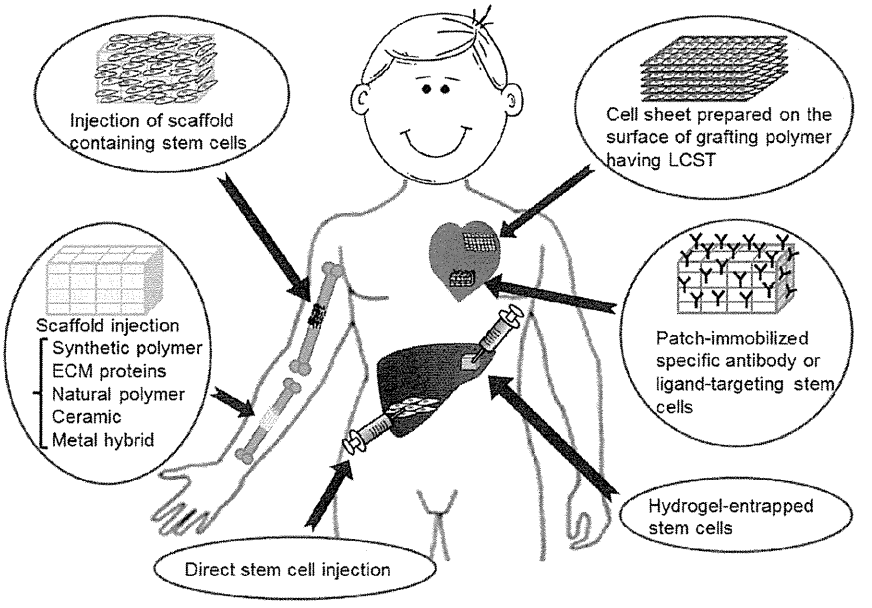


Figure 3. Some examples of biomaterial designs with and without stem cells for the injection of biomaterials in clinical applications: (a) injection of scaffold containing stem cells, (b) injection of scaffold without cells, (c) direct stem cell injection, (d) injection of cell sheets, (e) injection of patch-immobilized specific antibody or ligand-targeting stem cells, and (f) injection of hydrogel-entrapped stem cells.

peptides can be covalently immobilized using hexamethylene diisocyanate (HMDIC), 1,6-dimethyl suberimidate dihydrochloride (DMS),[119] or NHS/EDC reagent,[18] where NHS is *N*-hydroxysuccinimide and EDC is *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide (Figure 4). EDC is a water-soluble carbodiimide that is generally used in the 4.0–6.0 pH range. Therefore, it is possible to immobilize ECM proteins and ECM-mimicking peptides in aqueous solution using NHS/EDC reagents. The covalent bonding between amino groups can be reacted with aqueous DMS.[119]

Genipin is generally used to cross-link proteins, such as collagen and gelatin, and chitosan via amino groups.[120,121] Genipin can also be used for the immobilization of ECM proteins and peptides on the surface of culture dishes with amino groups (Figure 4). NHS/EDC, DMS, and genipin are the recommended reagents to covalently immobilize ECM proteins and ECM-mimicking peptides on culture dishes.

3.2. 3D Culture in Hydrogels

Hydrogels are physically or chemically cross-linked polymer networks that are able to absorb large amounts of water. Injectable hydrogels containing stem cells can be delivered to sites of damage in patients with minimal invasiveness, and the hydrogels ensure that stem cells remain localized to the damaged sites more effectively than injected cells alone. Physical cross-linking is performed on ECM proteins with thermosensitive properties of lower critical solution temperature (LCST) or upper critical solution temperature (UCST), such as collagen and gelatin. Collagen can be dissolved in aqueous solutions at low temperature and forms gels at ~37 °C because of its LCST characteristics, and gelatin can be dissolved in aqueous solution at high temperatures and forms gels at room temperature because of its UCST. Therefore, stem cells can be dissolved in ECM protein solutions and efficiently entrapped in ECM gels at 20–37 °C. However, most ECM

Table 3. ECM Immobilized on Dishes for Adhesion, Differentiation, And Proliferation of Stem Cells and Some Examples of the Literature

ECM	binding site of cells	ref
collagen I	integrin ($\alpha V\beta 3$, $\alpha 2\beta 1$)	58, 96
collagen I	integrin ($\alpha 1\beta 1$)	97
collagen I	integrin ($\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 3\beta 1$)	71
collagen II	integrin ($\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 10\beta 1$)	71, 91
collagen IV	integrin ($\alpha 2\beta 1$, CD44)	98
gelatin		99
fibronectin	integrin ($\alpha 4\beta 1$, $\alpha 5\beta 1$, $\alpha V\beta 3$, $\alpha IIb\beta 3$, $\alpha V\beta 6$, $\alpha V\beta 5$)	58, 96
laminin	integrin ($\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 3\beta 1$, $\alpha 6\beta 1$, $\alpha 6\beta 4$)	100
laminin-1 (laminin 111)	integrin ($\alpha 1\beta 1$, $\alpha 2\beta 1$, $\alpha 6\beta 1$, $\alpha 7\beta 1$, $\alpha 9\beta 1$), α-dystroglycan, suifade, and heparan sulfate proteoglycan	83, 101
laminin-5 (laminin 332)	integrin ($\alpha 2\beta 1$, $\alpha 3\beta 1$, $\alpha 6\beta 1$, $\alpha 6\beta 4$)	102
laminin-10/11	integrin ($\alpha 3\beta 1$, $\alpha 6\beta 1$, $\alpha 6\beta 4$)	100
vitronectin	integrin ($\alpha V\beta 3$, $\alpha V\beta 5$)	58, 96

proteins and ECM-derived oligopeptides (ECM peptides) need other forms of cross-linking to trap stem cells and generate hydrogels. Typically, photocross-linking and chemical cross-linking of ECM proteins and ECM peptides are used. There are several excellent reviews that discuss hydrogel preparation and reaction in detail.[12,14] Therefore, this section deals briefly with the preparation of ECM hydrogels using photocross-linking and chemical cross-linking with cross-linking agents. The application of ECM hydrogels containing stem cells is discussed in section 5 for specific ECM proteins and ECM peptides.

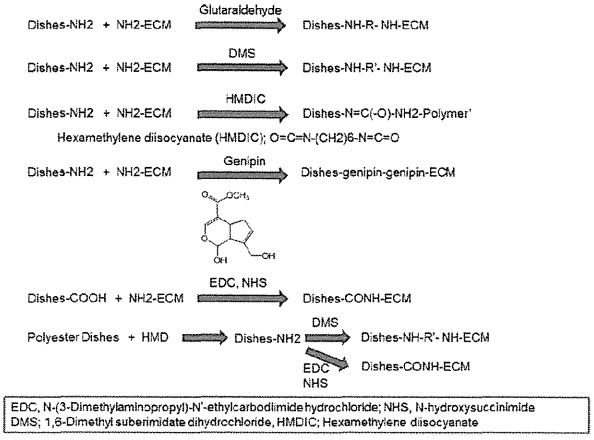


Figure 4. Surface reactions of covalent immobilization of ECM proteins and ECM-mimicking peptides on dishes.

3.2.1. Photocross-Linking of ECM Proteins and ECM Peptides. Hydrogels containing stem cells can be easily prepared by UV irradiation of ECM proteins and ECM–peptide solutions. These preparations can be used as injectable hydrogels via photocross-linking. However, it is first necessary to introduce double bonds into ECM proteins and ECM peptides for photocross-linking. ECM proteins and ECM peptides have $-OH$, $-NH_2$, and $-COOH$ functional groups. Double bonds can be introduced into ECM proteins and ECM

Table 4. ECM-Mimicking Peptides Immobilized on Dishes for Adhesion, Differentiation, And Proliferation of Stem Cells

ECM-mimicking peptide	ECM proteins for mimicking	binding site of cells	ref
DGEA	collagen I	integrin ($\alpha 2\beta 1$)	103–105
GTPGPQGIAGQRGVV (P15)	collagen I	integrin ($\alpha 2\beta 1$)	103, 106
$(RADA)_4$GG**DGEA**	collagen I	integrin ($\alpha 2\beta 1$)	116
$(RADA)_4$GGFPGERGVEGPGP	collagen I		116
GFOGER	collagen	integrin ($\alpha 2\beta 1$)	103, 107, 108
MNYYSNS	collagen IV		109
RGD	collagen I	integrin ($\alpha V\beta 3$)	97, 110
ELIDVPST (CS-1)	fibronectin	integrin ($\alpha 4\beta 1$); VLA-4	16, 111
FN-40	fibronectin	integrin ($\alpha 4\beta 1$, VLA-4)	16, 112
FN-120	fibronectin	integrin ($\alpha 5\beta 1$); VLA-5	16, 112
FN–CH296	fibronectin	integrin ($\alpha 4\beta 1$, $\alpha 5\beta 1$)	16, 112
KGGAVTG**RGD**SPASS	fibronectin	integrin ($\alpha 5\beta 1$); VLA-5	18, 113
G**RGD**SPK	fibronectin	integrin ($\alpha 5\beta 1$); VLA-5	18, 113
KNNQKSEPLIGRKKT	fibronectin	heparin-binding domain	53
RGDS	fibronectin		109
PHSRN	fibronectin		109
KYGAASI**KVAV**SADR	laminin		18, 114
YIGSR	laminin		109
IKVAV	laminin		115
PPFLMLLKGSTR	laminin-5 (laminin332)	integrin ($\alpha 3\beta 1$)	
(RADA)4-GG**PDSGR**	laminin		116
(RADA)4-GGSDPG**YIGSR**	laminin		116
(RADA)4-GGIKVAV	laminin		116
KGGPQVT**RGD**VFTMP	vitronectin	integrin ($\alpha V\beta 5$)	18, 117
KGGNGEP**RGD**TYRAY	bone sialoprotein (BSP)		18, 118
PEO4-NGEP**RGD**TYRAY	BSP-linker		18, 118
RGD	osteopontin	integrin ($\alpha V\beta 3$)	97

dx.doi.org/10.1021/cr3000169 | Chem. Rev. 2012, 112, 4507–4540

peptides by the reactions of acryloyl chloride,[122] glycidyl methacrylate,[12,123] and 2-aminoethylmethacrylate[12,124] (Figure 5). Figure 5 also shows a schematic for preparation method of

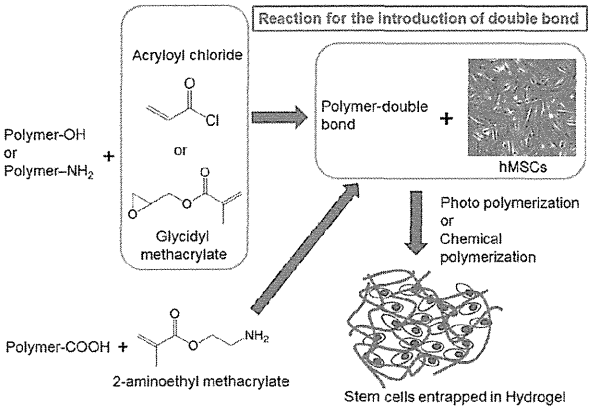


Figure 5. Schematic of the preparation method of hydrogels with entrapped stem cells by photopolymerization.

hydrogels with entrapped stem cells by photopolymerization. Aqueous solutions containing stem cells and macromers of ECM proteins and ECM peptides are irradiated with UV light to generate hydrogels with entrapped stem cells.

Poly(ethylene glycol)diacrylate (PEODA) is typically added to the reaction solution to generate optimal hydrogels.[65,125–129] Yang et al. prepared PEODA hydrogels incorporating RGD adhesive peptides and goat BMSCs by photopolymerization. They found that RGD-conjugated PEODA hydrogels promoted the osteogenic differentiation of BMSCs, and RGD enhanced differentiation in a dosage-dependent manner, with the highest concentration (2.5 mM) in the reaction solution being optimal in their study.[125]

3.2.2. Chemical Cross-Linking of Hydrogels. Hydrogels of ECM proteins can also be prepared by chemical cross-linking. Similar to ECM protein immobilization on 2D dishes, as discussed in section 3.1, NHS/EDC, DMS, HMDIC, and genipin are typically used as cross-linking agents. Glutaraldehyde is not commonly used for the preparation of hydrogels in tissue engineering because it is relatively toxic to stem cells. DMS, HMDIC, and genipin allow cross-linking between amino groups, whereas NHS/EDC leads to cross-linking between carboxylic acids and amino groups in ECM proteins.

Chang et al. compared gelatin hydrogels cross-linked with genipin and gelatin hydrogels cross-linked with glutaraldehyde.[120] They found that the degree of inflammatory reaction in wounds treated with the genipin-cross-linked gelatin was significantly less severe than those covered with the glutaraldehyde-cross-linked gelatin in vivo.[120] In addition, the healing rates of wounds treated with the genipin-cross-linked gelatin were notably faster than those with glutaraldehyde-cross-linked hydrogels.[120]

3.3. 3D Culture in Scaffolds

Scaffolds seeded with stem cells can support 3D tissue formation artificially. It is optimal for scaffolds (a) to allow cell attachment and migration, (b) to allow diffusion of nutrients, growth factors, and waste secreted by cells, and (c) to have mechanical properties similar to the natural tissue. Most of the scaffolds have high porosity (>80%) and large pore size (200–800 μm), which allow diffusion of nutrients, growth factors, and waste, but these properties also lead to weak mechanical properties. Biodegradability of scaffolds is often required because scaffolds should be absorbed by the surrounding tissues without the necessity of surgical removal. It is preferable that the degradation rate of scaffolds should be matched to the speed of tissue formation. The degradation speed of scaffolds can be regulated by the degree of cross-linking. Scaffolds prepared from ECM proteins and ECM peptides are desirable because of their biodegradable characteristics. ECM proteins used for the preparation of scaffolds are typically collagen type I, collagen type II, gelatin, fibronectin, laminin, and vitronectin. ECM proteins can be used as (a) coating materials, (b) blending materials, and (c) main materials of scaffolds.

3.3.1. Preparation of Scaffolds. There are several methods used to prepare scaffolds for tissue engineering and 3D culture of stem cells, including (a) freeze-drying, (b) salt leaching, (c) porogen leaching, (d) use of nonwoven fabric or mesh, (e) nanotopography, and (f) electrospinning. In the freeze-drying method, ECM proteins are dissolved in a buffer solution. The ECM solution is frozen at −20 or −80 °C and then lyophilized in a freeze-dryer before being washed and stored (Figure 6). If necessary, the scaffolds are also cross-linked.

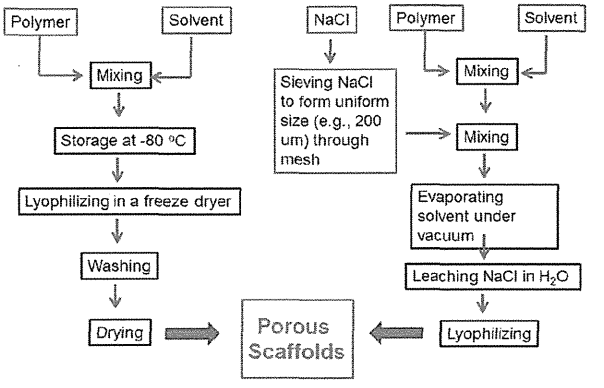


Figure 6. Typical preparation method of porous scaffolds by freeze-drying (a) and salt leaching (b).

The salt-leaching method is as follows. Biopolymers and/or ECM proteins are dissolved in a solvent. Salt, typically NaCl, is sieved to generate a uniform distribution of size using filtration through mesh and added into the solution. The solvent of the solution is vaporized under vacuum to generate dry scaffolds. Salt is then leached from the scaffolds by immersion in water after drying the scaffolds (Figure 6). The porogen-leaching method is a similar method to the salt-leaching method, but other uniformly sized particles, such as polymeric particles, are used instead of salt.

3.4. 3D Culture in Nanofibers

Peptide amphiphiles (PAs), which have a hydrophilic domain and a hydrophobic domain, are known to spontaneously generate self-assembled nanofibers above critical micelle concentrations.[109,116,130] MSC differentiation on self-assembled nanofibers using ECM peptides is discussed in section 5.8.1.

A typical method to create nanofibers is electrospinning. Electrospun scaffolds can support cell adhesion and growth and

promote differentiation of stem cells.[131] Nanofibers can be generated from a spinning nozzle when high voltage is applied between the spinning nozzle and a flat metal collector. Typical electrospinning products are flat and highly interconnected scaffolds with a nonwoven fabric sheetlike morphology. These characteristics hinder cell infiltration and growth throughout the scaffolds. Blakeney et al. have developed a three-dimensional cotton ball-like electrospun scaffold that consists of low-density, uncompressed nanofibers.[131] A grounded spherical dish and an array of needle-like probes were used instead of a traditional flat-plate collector to create a cotton ball-like scaffold. Scanning electron microscopy showed that the cotton ball-like scaffold consisted of electrospun nanofibers with a similar diameter, but with larger pores and less dense structures than traditional electrospun scaffolds.[131] The cotton-ball like scaffolds prepared from ECM proteins by electrospinning will be interesting for use as scaffolds for guiding specific lineages of stem cell differentiation.

4. PHYSICAL PROPERTIES OF BIOPOLYMERS (BIOMATERIALS) GUIDE STEM CELL DIFFERENTIATION FATE (LINEAGE)

The interactions between MSCs and ECM proteins are classified as physical, chemical, and biological. It has recently been recognized that stem cell differentiation is directed by physical properties of culture materials as well as by biochemical responses to growth factors and ECM proteins.[19,20,132] Cells in bone, muscle, liver, and brain tissues reside in different environments that have diverse physical properties.[133] The matrix stiffness for differentiated cells is known to influence focal-adhesion structure and the cytoskeleton.[134−139] Engler et al. reported that soft materials, with similar stiffness to the brain, tend to differentiate MSCs into neurogenic cells, whereas stiffer materials that mimic muscle guide MSCs into myogenic cells and rigid materials similar to collagenous bone induce osteogenic differentiation (Figure 7).[19] However, this work was performed on a 2D surface of hydrogels coated with collagen. The effect of stiffness in 3D culture may produce different results than in 2D culture.

Gilbert et al. also reported that the elasticity of culture materials regulates self-renewal of skeletal muscle stem cells.[20]

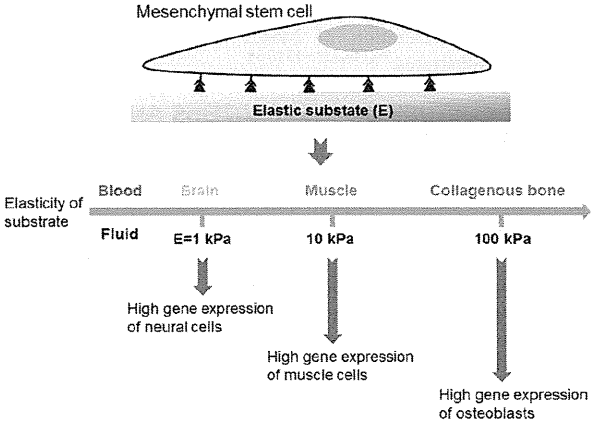


Figure 7. Physical properties decide the fate of stem cell cultured on biomaterials with different elasticity. Modified with permission from ref 19. Copyright 2006 Elsevier Inc.

Muscle stem cells (MuSC's) exhibit robust regenerative capacity in vivo, but this capacity is rapidly lost in culture. They showed that the elasticity of culture materials was a potent regulator of MuSC fate. MuSC's cultured on soft hydrogel substrates that mimicked the elasticity of muscle (12 kPa) self-renew in vitro and contributed extensively to muscle regeneration when transplanted into mice, unlike MuSC's grown on rigid plastic dishes (~106 kPa), as shown by histology and bioluminescence imaging. These studies provide evidence that propagation of adult muscle stem cells is possible by recapitulating physiological tissue rigidity.[20] This finding may contribute to future cell-based therapies for muscle-wasting diseases.

The effect of physical interactions between MSCs and culture materials on stem cell fate is discussed in several articles.[19,20,61,133,140−154] Some landmark findings are summarized in Table 5, and some examples of physical effects on differentiation of MSCs cultured on ECM proteins are reviewed here.

Table 5. Some Articles Discussing Physical Effect of Substrates on Differentiation of MSCs Cultured on the Substrates

authors	contents	ref (year)
J. R. Mauney et al.	**mechanical stimulation** promotes osteogenic differentiation of hBMSCs	140 (2004)
J. S. Park et al.	differential effects of **equiaxial and uniaxial strain** on MSCs	141 (2004)
V. E. Meyers et al.	**microgravity** disrupts collagen I/integrin signaling during osteogenic differentiation of hMSCs	142 (2004)
V. I. Sikavitsas et al.	**flow perfusion** enhances the calcified matrix deposition of marrow stromal cells in scaffolds	143 (2005)
H. Hosseinkhani et al.	**perfusion culture** enhances osteogenic differentiation of MSCs	144 (2005)
A. J. Engler et al.	**matrix elasticity** directs stem cell lineage specification	19 (2006)
R. D. Sumanasinghe et al.	osteogenic differentiation of hMSCs in collagen matrices: effect of **uniaxial cyclic tensile strain**	145 (2006)
D. F. Ward et al.	**mechanical strain** promotes osteogenic differentiation of hMSCs	61 (2007)
E. K. F. Yim et al.	**nanostructures** inducing differentiation of hMSCs into neurinal lihneage	154 (2007)
B. Lanfer et al.	growth and differentiation of MSCs on **aligned collagen** matrices	146 (2009)
Q. Li et al.	ECM with the **rigidity** of adipose tissue helps adipocytes maintain insulin responsiveness	147 (2009)
M. Zscharnack et al.	**low O_2** expansion improves subsequent chondrogenesis of BMSCs in hydrogel	148 (2009)
C. H. Huang et al.	interactive effects of **mechanical stretching** and ECM proteins on initiating osteogenic differentiation of hMSCs	149 (2009)
P. M. Gilbert et al.	substrate **elasticity** regulates skeletal muscle stem cell self-renewal in culture	20 (2010)
G. C. Reilly and A. J. Engler	intrinsic ECM properties regulate stem cell differentiation (**mechanobiology**)	150 (2010)
J. M. Kemppainen and S. J. Hollister	differential effects of **designed scaffold** permeability on chondrogenesis by BMSCs	151 (2010)
E. K. F. Yim et al.	**nanotopography**-induced changes in focal adhesions, cytoskeletal organization, and mechanical properties of hMSCs	152 (2010)
J. Tang et al.	regulation of stem cell differentiation by cell−cell contact on **micropatterned** material surfaces	153 (2010)
P. A. Janmey and R. T. Miller	mechanisms of **mechanical** signaling in development and disease	133 (2011)

dx.doi.org/10.1021/cr3000169 | *Chem. Rev.* 2012, 112, 4507−4540

4.1. Mechanical Stretching Effect of Culture Surface-Coated with ECM Proteins

Mechanical strain and ECM proteins play important roles in the osteogenic differentiation of hMSCs.[61,140,145,149] Several studies have shown that mechanical strain can promote osteogenic or other lineage differentiation in cells cultured on ECM proteins even in the absence of osteogenic supplements in the culture medium.[61,145,149]

Park et al. reported that mechanical strain regulated the expression of vascular smooth muscle cell (SMC) markers in MSCs (Figure 8).[141] Cyclic equiaxial strain downregulated smooth muscle (SM) α-actin and SM-22α in MSCs on collagen- or elastin-coated membranes after one day and decreased the level of α-actin in stress fibers. In contrast, cyclic uniaxial strain transiently increased the expression of SM α-actin and SM-22α after one day, which subsequently returned to basal levels after the cells aligned in the direction perpendicular to the strain.[141] In addition, uniaxial but not equiaxial strain induced a transient increase in collagen type I expression. DNA microarray experiments showed that uniaxial strain increased SMC markers and regulated the expression of matrix molecules without significantly changing the expression of differentiation markers (e.g., ALP and collagen type II) in other cell types.[141] Their results suggest that uniaxial strain, which better mimics the type of mechanical strain experienced by SMCs, could promote MSC differentiation into SMCs if cell orientation is controlled.[141]

Ward et al. showed that application of a 3−5% tensile strain to a collagen type I substrate stimulated osteogenesis in attached hMSCs through gene focusing via a MAPK signaling pathway.[61] They found that mechanical strain led to an increase in the expression of osteogenic marker genes while simultaneously reducing expression of marker genes from three alternate lineages (chondrogenic, adipogenic, and neurogenic).[61] Mechanical strain also increased matrix mineralization (a hallmark of osteogenic differentiation) and activation of extracellular signal-related kinase 1/2 (ERK).[61] These results demonstrated that mechanical strain enhanced collagen type I-induced gene focusing and osteogenic differentiation in hMSCs through the ERK/MAPK signaling pathway.[61]

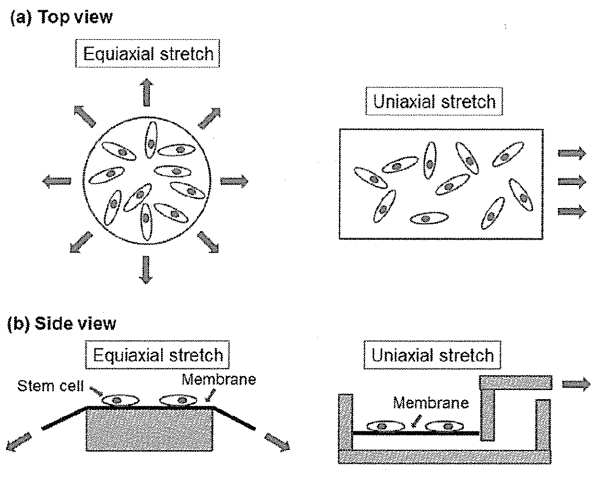


Figure 8. Schematic model of the apparatus that can apply equiaxial (a) and uniaxial (b) strain to MSCs. Modified with permission from ref 141. Copyright 2004 Wiley Periodicals.

Huang et al. investigated the combined effects of ECM proteins and mechanical factors (cyclic stretching) in driving hMSCs toward osteogenic differentiation.[149] hMSCs cultured in regular medium were grown on substrates coated with various ECM proteins (collagen type I, vitronectin, fibronectin, and laminin) and subjected to cyclic mechanical stretching.[149] All of the ECM proteins tested supported hMSC differentiation into osteogenic phenotypes in the absence of osteogenic supplements.[149] Cyclic mechanical stretching activated the phosphorylation of focal adhesion kinase (FAK), induced upregulation of the transcription and phosphorylation of Runx2, and subsequently increased ALP activity and mineralized matrix deposition.[149] Fibronectin and laminin exhibited greater effects of supporting stretching-induced osteogenic differentiation than did collagen type I and vitronectin.[149] It was suggested that the ability of ECM proteins and mechanical stretching to regulate osteogenesis in hMSCs may be exploited in bone tissue engineering by appropriate matrix design and by mechanical stimulation.[149]

4.2. Low Oxygen Expansion Promotes Differentiation of MSCs

Several groups have reported the effects of low oxygen tension on the differentiation of MSCs, especially in chondrogenic differentiation of MSCs cultured on ECM substrates.[148,155] Zscharnack et al. investigated the effect of low oxygen tension (5%) during the expansion of ovine MSCs on colony-forming unit-fibroblast (CFU-F) formation and chondrogenesis in pellet culture and in collagen type I hydrogels.[148] MSCs expanded in 5% O_2 showed a 2-fold higher CFU-F potential, and chondrogenic differentiation was enhanced in both pellet culture and collagen type I hydrogels. It was demonstrated that physiologically low oxygen tension during monolayer expansion of ovine MSCs was advantageous to improving cartilage tissue engineering in a sheep model.[148]

4.3. Other Physical Effect Affecting Differentiation of MSCs

There are several other physical effects that promote differentiation of MSCs on ECM protein surfaces. (i) Perfusion culture promotes osteogenic differentiation of MSCs cultured on ECM protein surface.[143,144] (ii) Microgravity disrupts collagen type I/integrin signaling during osteoblastic differentiation of hMSCs.[142] (iii) The mechanical properties of ECM proteins guide specific lineage differentiation of MSCs.[147,150,156,157] (iv) The topography of ECM proteins promotes differentiation of MSCs cultured on aligned or patterned substrates.[74,146,151−154,158]

5. MSC CULTURE ON ECM PROTEINS AND NATURAL BIOPOLYMERS

The ECM is the extracellular component of animal tissues that provides structural support for the cells, in addition to stimulating various important biological functions. ECM proteins are able to dictate whether cells will proliferate or undergo growth arrest, migrate or remain stationary, and thrive or undergo apoptotic death.[159] Therefore, the ECM proteins are an important factor in reproducing the biological niches of cells in vitro, which guides MSCs to differentiate into different lineages such as osteoblasts, chondrocytes, adipocytes, cardiomyocytes, neural cells, hepatocytes, and β-cells. The differentiation of MSCs in culture systems depends on the components, structure (morphology), origin, and quantity of ECM proteins that are used. Because ECM proteins are used as scaffolds for the organization of cells in tissues, ECM proteins

are the main cell culture materials used to control the proliferation and differentiation of MSCs in tissue engineering and regenerative medicine, both in vitro and in vivo. Therefore, this review focuses mainly on the differentiation of MSCs cultured on biomaterials made of specific ECM proteins and on the biological and chemical interactions between these cells and proteins.

5.1. Chemical and Biological Interactions of ECM Proteins and Stem Cells

ECM proteins have chemical functional groups of carboxylic acid, amine, phosphate, and/or sulfonic acid. They also have aspects of polyelectrolytes and characteristic isoelectric points (IEPs).[160−175] Table 6 shows the IEPs of some ECM proteins,

Table 6. Isoelectric Points of Some ECM Proteins, Growth Factors, And Polymers

materials	isoelectric point	ref
	ECM	
collagen type I	4.7, 6.4, 6.78, 7.02, and 8.26 depending on preparation conditions	172−174
gelatin sol	7.8, temp > 40, or increasing pH	344
gelatin gel	4.7, temp < 15, or decreasing pH	344
fibronectin	5.5−6.0	160
vitronectin	4.75−5.25	161
laminin	5.87, 4.89, and 5.08	162
heparin	3.4	163
hyaluronic acid	2.5	170
	growth factor	
FGF-1 (aFGF)	5.6	169
FGF-2 (bFGF)	9.6	169
rhBMP-2	9	171
insulin	5.3	168
PDGF	9.8	165
EGF	4.0−5.0	
TGF-β1	9.5	164
	polymer	
agarose	5.5	166
alginate	5.4	175
poly(lactic-*co*-glycolic acid) (PLGA)	2.75	163
poly(L-lysine)	9.5	163
chitosan	8.7	167
polyacrylamide	5.7	166

natural biopolymers, and growth factors.[160−172,174,175] IEPs are as follows: gelatin gel and collagen type I 4.7−8.3,[172,174] fibronectin 5.5−6.0,[160] laminin 4.9−5.9,[162] vitronectin 4.8−5.3,[161] heparin 4.7,[163] hyaluronic acid 2.5,[170] agarose 5.5,[166] and alginate 5.4.[175] Most ECM proteins and natural biopolymers are negatively charged under physiological conditions. The IEP of some growth factors is <7 (e.g., 5.6 for FGF-1[169] and 5.3 for insulin[168]), whereas for other growth factors, it is >7 (e.g., 9.6 for FGF-2,[169] 9.0 for BMP-2,[171] 9.8 for PDGF,[165] and 9.5 for TGF-β1[164]). Some binding between ECM proteins and growth factors (e.g., collagen type I and BMP-2) is mainly due to electrochemical interactions.

The binding of ECM proteins to cells is mainly mediated by integrin receptors. Integrins comprise a large family of cell-surface receptors that bind and mediate adhesion to ECM components, organize the cytoskeleton, and activate intracellular signaling pathways.[159] Each integrin consists of two type-1 transmembrane subunits: α and β. In mammals, 18 α- and 8 β-subunits associate in various combinations to form 24 integrin dimers that can bind to distinct subsets of ECM ligands.[176,177]

Most ECM proteins have molecular weights of 10−1000 kDa but only a few integrin-binding domains. These integrin-binding domains have specific sequences of a few amino acids (3−10), e.g., RGD, DGEA, YIGSR, IKVAV, and GFOGER. Table 4 summarizes the integrin receptors and amino acid sequences that mediate cell−ECM associations that are important for MSC proliferation and differentiation, as well as normal cell culture.

Many members of the integrin family, including $\alpha5\beta1$, $\alpha8\beta1$, αIIb$\beta3$, $\alpha V\beta3$, $\alpha V\beta5$, $\alpha V\beta6$, and $\alpha V\beta8$, recognize an Arg-Gly Asp (RGD) motif within fibronectin,[16,18,109] fibrinogen,[109] vitronectin,[18] von Willebrand factor, and other large glycoproteins. Collagen type I has a cell-binding domain of DGEA, which binds to integrin $\alpha2\beta1$.[103] Collagen type I is also bound by integrins $\alpha1\beta1$, $\alpha3\beta1$, and $\alpha V\beta3$.[58,97] RGD in collagen type I is reported to associate with integrin $\alpha V\beta3$.[97] The large size of ECM proteins, compared to the small integrin-binding motifs, provides not only structural support but also conformational regulation of the cell-binding domains. The differences in conformation of the cell-binding domains lead to different associations with specific integrin receptors.[178,179] MSC differentiation on culture materials composed of specific ECM and natural biopolymers is discussed in the following sections.

5.2. Collagen

Collagen is a typical ECM protein used in the culturing of MSCs, which is found in all animals, especially in the flesh and connective tissues of mammals.[180] Collagen is the main component of connective tissue and the most abundant protein in mammals,[181] making up ~25−35% of the whole-body protein content. Elongated collagen fibrils are found in fibrous tissues, including skin, ligaments, and tendons. Collagen is also abundant in the cornea, cartilage, bone, blood vessels, gut, and intervertebral discs. Because of its abundance, collagen, especially collagen type I, is relatively inexpensive compared to other ECM proteins such as laminin, vitronectin, and fibronectin, which allows us to use it in large quantities to make scaffolds and hydrogels for stem cell culture.[144,182−185]

To date, 29 types of collagen have been identified and described. The five most common types are (i) collagen type I (genes; COL1A1, COL1A2), which is the main component of bone and also found in skin and tendons; (ii) collagen type II (gene; COL2A), which is the main component of cartilage; (iii) collagen type III (gene; COL3A), which is the main component of reticular fibers; (iv) collagen type IV (genes; COL4A1, COL4A2, COL4A3, COL4A4, COL4A5, and COL4A6), which is found in basement membranes;[186] and (v) collagen type V (COL5A1, COL5A2, and COL5A3), which is found in placenta and hair.[187]

Collagen undergoes many post-translational modifications, including extensive cross-linking. Defective cross-linking has been implicated in human syndromes (e.g., osteogenesis imperfecta and Ehlers-Danlos syndrome).[188] However, it was reported that the inhibition of cross-linking of collagen was not required for osteogenic differentiation of hMSCs, as shown by the expression of ALP and genome-wide gene-expression analysis, but it did enhance matrix mineralization.[188] Specific characteristics of collagen, such as stiffness, elasticity, degree of cross-linking, and origin (i.e., cow-, pig-, or fish-derived collagen

dx.doi.org/10.1021/cr3000169 | Chem. Rev. 2012, 112, 4507−4540