

Figure 1 | Aberrant silencing of the *Dlk1-Dio3* gene cluster in mouse iPSCs. **a**, Strategy for comparing genetically matched ES cells and iPSCs generated with the doxycycline-controllable collagen-OKSM system. **b**, Morphology of collagen-OKSM ES cells and iPSCs. **c**, Unsupervised clustering of four ES cell and six derivative iPSC lines based on microarray expression data. **d**, Scatter plot of microarray data comparing iPSCs and ES cells with differentially expressed genes highlighted in green (twofold, $P < 0.05$, t -test with

Benjamini-Hochberg correction). **e**, Heat map showing relative expression levels of selected mRNAs in ES cells and iPSCs. **f**, Schematic representation of the *Dlk1-Dio3* gene cluster with maternally and paternally expressed transcripts shown in red and blue, respectively. **g**, Heat map showing miRNAs that are differentially expressed between ES cells and iPSCs (twofold, $P < 0.01$, t -test).

We first compared the abilities of parental collagen-OKSM ES cells and iPSCs derived from mouse embryonic fibroblasts (MEFs) to support the development of all-iPSC mice using 4n embryo complementation. The two tested ES cell lines gave rise to viable mice at expected frequencies (13–20%)¹¹, demonstrating that the OKSM transgene per se does not adversely affect development (Supplementary Table 1). In contrast, all four tested iPSC lines repeatedly failed to support the development of all-iPSC mice, indicating qualitative differences between these iPSCs and ES cells (Supplementary Table 1).

We reasoned that a transcriptional comparison of the iPSC lines that failed 4n complementation with matched 4n-complementation-competent ES cell lines might reveal molecular changes that explain the developmental deficits of iPSCs. Global mRNA profiling showed marked similarities in the overall transcriptional patterns of four collagen-OKSM ES cells and six derivative iPSCs and did not separate these cell lines using unsupervised clustering or principal component analysis (Fig. 1c and data not shown). In fact, only two transcripts were identified as differentially expressed (>2-fold difference, t -test,

$P < 0.05$) between ES cells and iPSCs. These were the non-coding RNA *Gtl2* (also known as *Meg3*) and the small nucleolar RNA *Rian* (Fig. 1d, e).

Repression of *Dlk1-Dio3* transcripts in iPSCs

Gtl2 and *Rian* localize to the imprinted *Dlk1-Dio3* gene cluster on mouse chromosome 12qF1 and are maternally expressed in mammals (Fig. 1f)²⁷. Both genes were strongly repressed in iPSC clones compared to ES cell clones, whereas expression of pluripotency and housekeeping genes remained unaffected (Fig. 1e). Quantitative PCR (qPCR) analysis of *Gtl2*, *Rian* and *Mirg*, another maternally expressed imprinted gene in the *Dlk1-Dio3* cluster, confirmed transcriptional silencing in iPSCs (Supplementary Fig. 1a). Expression of other imprinted genes showed clone-to-clone variations, as was previously seen for ES cells²⁸, but no consistent differences between ES cells and iPSCs (Fig. 1e and Supplementary Table 2). This shows that imprinted gene silencing is not a genome-wide phenomenon. Of note, we failed to detect differences with the collagen-OKSM system

in the expression of ~300 transcripts that have recently been reported to be differentially regulated in a comparison of mouse and human iPSCs with ES cells¹⁵ (Supplementary Fig. 2a). However, silencing of *Gtl2* was also evident in the murine-specific iPSC data set used in that study (Supplementary Fig. 2c). Together, these data indicate that a relatively small set of transcripts distinguishes genetically matched iPSCs and ES cells, and suggest that many of the previously seen differences are probably due to variations in genetic background or viral transgene insertions.

Imprinting of the *Dlk1-Dio3* locus is accompanied by differential expression of about 50 miRNAs (Fig. 1f)^{29,30}. We therefore performed genome-wide miRNA profiling on the same samples as analysed for mRNA expression. Of 336 miRNAs detected, 21 (6.3%) were differentially expressed between all ES cell and iPSC clones analysed (Fig. 1g and Supplementary Table 3). All of these miRNAs localized to chromosome 12qF1 and were silenced in iPSCs, thus corroborating the notion that most iPSCs show aberrant silencing of this major imprinting domain.

To determine the generality of *Gtl2* silencing in iPSCs, we measured its expression in 61 additional iPSC lines derived from haematopoietic

stem cells (11 lines), granulocyte-macrophage progenitors (11 lines), granulocytes (9 lines), peritoneal fibroblasts (6 lines), tail-tip fibroblasts (6 lines) and keratinocytes (18 lines). Only four of these lines (5.8%), originating from either peritoneal or tail-tip fibroblasts, showed *Gtl2* expression levels similar to those of ES cells (termed '*Gtl2*^{on} clones') (Fig. 2a and Supplementary Fig. 1b, c). The finding that the vast majority of iPSC clones showed transcriptional suppression of *Gtl2* (termed '*Gtl2*^{off} clones') demonstrates that silencing of this locus occurs in iPSCs derived from different cell types at various stages of differentiation. Analysis of published microarray data sets comparing ES cells and iPSCs generated from mouse fibroblasts, neural and bone marrow cells also showed repression of maternally expressed 12qF1 transcripts (Supplementary Fig. 2b–e), supporting the notion that silencing of this cluster is a common outcome upon factor-mediated reprogramming.

Developmental effects of *Dlk1-Dio3* silencing

Dysregulation of genes within the *Dlk1-Dio3* cluster can be detrimental during pre- and postnatal mouse development^{31–35}. To assess whether the expression status of *Gtl2* and its associated transcripts

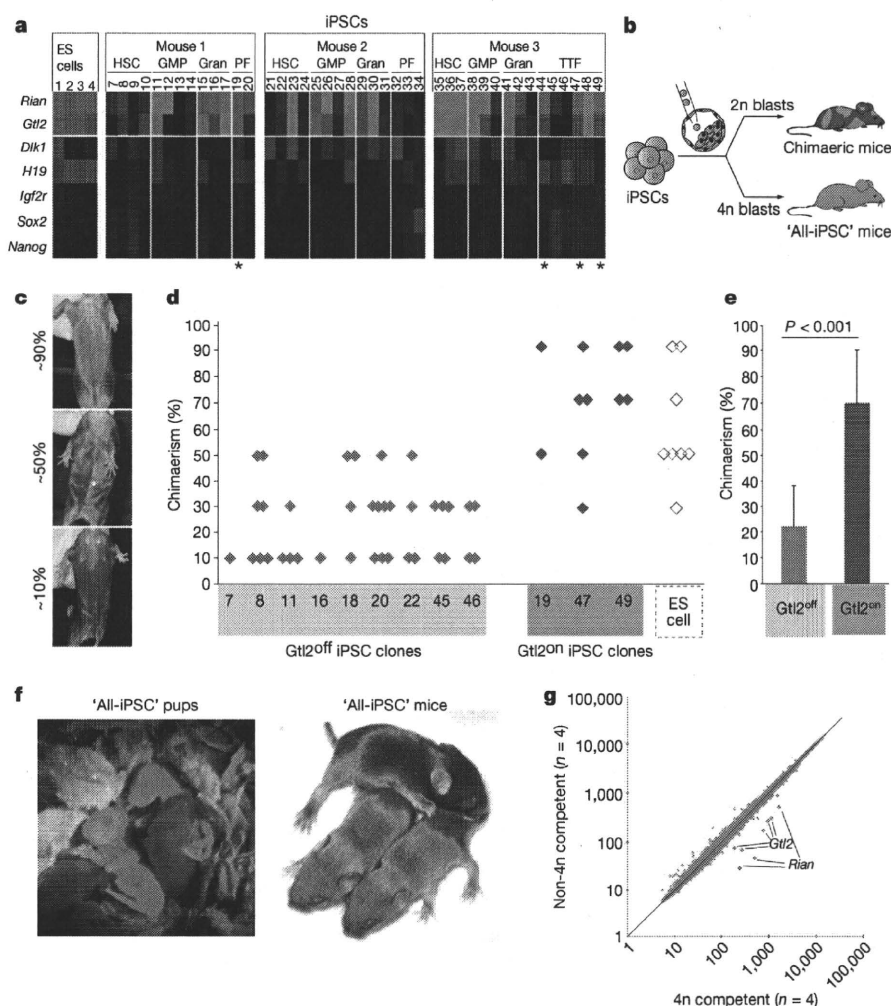


Figure 2 | Developmental consequences of *Dlk1-Dio3* silencing. **a**, Heat map showing relative expression levels of *Gtl2*, *Rian* and other select genes in ES cells and iPSCs derived from haematopoietic stem cells (HSC), granulocyte-macrophage progenitors (GMP), granulocytes (Gran), peritoneal fibroblasts (PF) and tail-tip fibroblasts (TTF). Four iPSC clones expressing ES-cell-like levels of *Gtl2* and *Rian* were identified (highlighted by asterisks) (iPSC clone number 18 was analysed only by qPCR; see Supplementary Fig. 1b). **b**, Strategy for assessing the developmental potential of iPSC clones by injection into diploid (2n) and tetraploid (4n) blastocysts to produce chimaeric or all-iPSC mice, respectively. **c**, Images of representative chimaeras with agouti coat colour indicating iPSC origin. **d**, Quantification of

coat colour chimaerism in mice derived from indicated *Gtl2*^{off} clones (green diamonds), *Gtl2*^{on} iPSC clones (red diamonds) and ES cells (open diamonds). **e**, Statistical analysis of coat colour chimaerism in mice derived from *Gtl2*^{off} and *Gtl2*^{on} iPSC clones. Error bars indicate standard deviations, *n* = 38 for *Gtl2*^{off} clones and *n* = 11 for *Gtl2*^{on} clones, *P* < 0.001. **f**, Images of two GFP⁺ all-iPSC neonates (left) and two agouti all-iPSC mice (right). **g**, Scatter plot showing intensity levels of all probe sets covered by microarray analysis; highlighted in green are those probe sets that were significantly different between 4n complementation-competent iPSCs (clones 19, 44, 47 and 49) and non-4n complementation-competent iPSCs (clones 18, 20, 45 and 48) (twofold, *P* < 0.05, *t*-test with Benjamini–Hochberg correction).

correlates with the developmental potential of iPSCs, we injected nine *Gtl2*^{off} clones into diploid blastocysts. This resulted in adult chimaeras that exhibited low-to-medium degree (10–50%) coat colour chimaerism (Fig. 2b–e and Supplementary Table 4). In contrast, injection of three *Gtl2*^{on} iPSC clones yielded adult mice with a coat colour chimaerism ranging from 30% to 100%, similar to the chimaerism seen with ES cells (Fig. 2d, e and Supplementary Table 4). Importantly, four out of four *Gtl2*^{on} iPSC clones supported the development of neonatal all-iPSC mice upon injection into 4n blastocysts at efficiencies comparable to those observed with ES cells (7–19% for iPSCs compared with 13–20% for ES cells) (Supplementary Table 1). We confirmed that these mice were entirely iPSC-derived by PCR for strain-specific polymorphisms (Supplementary Fig. 3), by detection of homogenous GFP fluorescence of all-iPSC neonates, originating from a *ROSA26-EGFP* allele introduced into the parental ES cells, and by uniform agouti coat colour of adolescent all-iPSC mice (Fig. 2f). To our knowledge, this is the first demonstration of animals produced entirely from adult-derived iPSCs.

In contrast to *Gtl2*^{on} iPSC clones, injection of ten *Gtl2*^{off} iPSC clones into 4n blastocysts consistently failed to produce all-iPSC pups but instead resulted in resorptions (Supplementary Table 1). Thus, the expression status of *Gtl2* in these iPSCs predicts their developmental potential into chimaeric and all-iPSC mice. The conclusion that the activation status of maternally expressed genes on chromosome 12qF1 is a strong indicator of the developmental potential of iPSCs was

further supported by the analysis of two published array data sets, which showed that *Gtl2* was expressed in 4n complementation-competent ES cell and iPSC lines but was downregulated in non-4n complementation-competent iPSC lines^{8,9} (Supplementary Fig. 4).

To test whether *Gtl2*^{on} and *Gtl2*^{off} iPSCs can be distinguished by the expression of genes outside of 12qF1, we performed global mRNA and miRNA expression profiling of four non-4n complementation-competent and four 4n complementation-competent iPSC lines (all derived from fibroblasts). This analysis identified only *Gtl2*, *Rian* and a total of 26 miRNAs, which all localize to the *Dlk1–Dio3* cluster, as differentially expressed, suggesting that the dysregulation of 12qF1 transcripts alone is responsible for the failure of *Gtl2*^{off} iPSCs to support the development of all-iPSC mice (Fig. 2g and Supplementary Table 5).

Epigenetic mechanism of *Dlk1–Dio3* repression

Imprinting of the *Dlk1–Dio3* cluster is regulated by differentially methylated regions (DMRs) that become epigenetically modified in the germ line. These include an intergenic DMR (IG-DMR), located between the *Dlk1* and *Gtl2* genes¹⁹, and a DMR spanning the *Gtl2* promoter (*Gtl2* DMR)¹⁵. To determine whether aberrant DNA methylation might be responsible for the transcriptional silencing seen in *Gtl2*^{off} iPSC lines, we compared the methylation status of IG-DMR, *Gtl2* DMR and three other CpG-rich regions on chromosome 12qF1 in ES cells, iPSCs and their parental tail-tip fibroblasts (Fig. 3a). As expected for germline-imprinted regions,

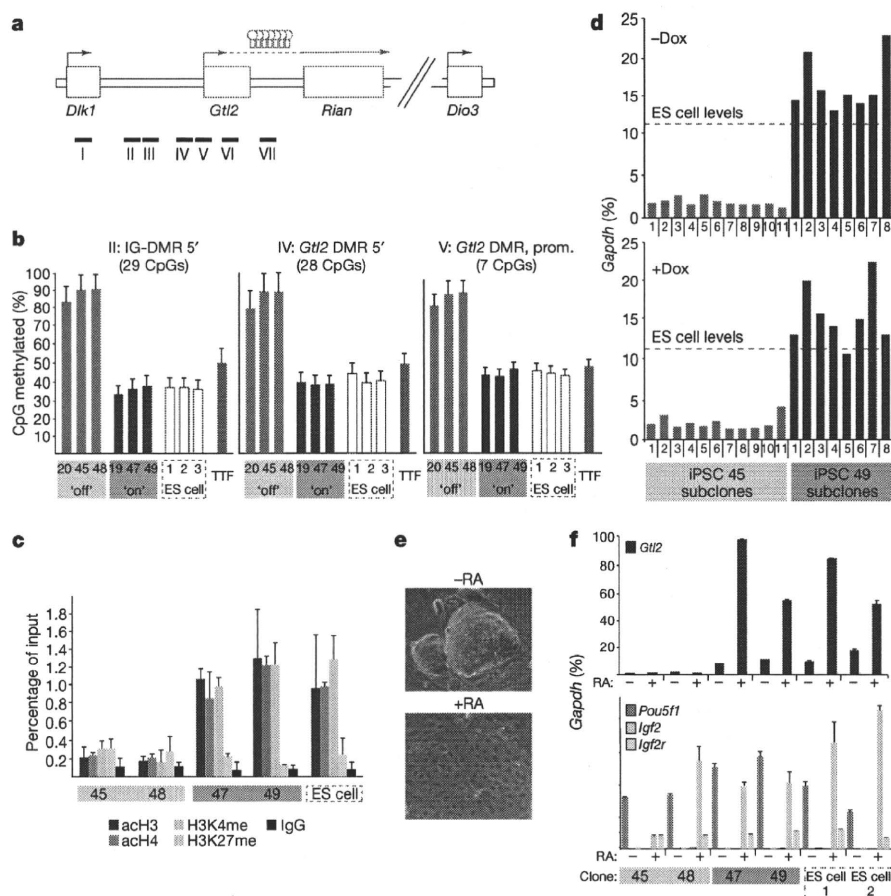


Figure 3 | Epigenetic silencing of the *Gtl2* locus in iPSCs. a, Structure of the *Dlk1–Dio3* locus with the position of the genomic regions (I–VII) analysed by pyrosequencing indicated by black bars. **b**, Degree of DNA methylation at IG-DMR and *Gtl2* DMR in three *Gtl2*^{off} iPSC clones (green bars), three *Gtl2*^{on} iPSC clones (red bars), three ES cell clones (red open bars), as well as the parental tail-tip fibroblasts (TTF, grey bars). Analysis of the other regions is shown in Supplementary Fig. 5. **c**, Prevalence of activation-associated (acH3, acH4 and H3K4me) and repression-associated (H3K27me) chromatin marks at the *Gtl2* promoter in two *Gtl2*^{off} iPSC

clones, two *Gtl2*^{on} iPSC clones and ES cells. **d**, *Gtl2* expression levels as measured by qPCR in subclones derived from *Gtl2*^{off} clone 45 and *Gtl2*^{on} clone 49 in the absence (–) or presence (+) of doxycycline (Dox). **e**, Bright-field images of iPSC culture in the absence or presence of all-trans retinoic acid (RA). **f**, Expression levels of *Gtl2*, other imprinted genes (*Igf2*, *Igf2r*), and the pluripotency marker *Pou5f1* in cells cultured with (+) or without (–) RA. All error bars indicate standard deviations with $n =$ number of CpGs within the corresponding region in **b** and $n = 3$ in **c** and **f**.

approximately 50% of CpGs within IG-DMR and *Gtl2* DMR were methylated in fibroblasts, ES cells and *Gtl2*^{on} iPSCs. In contrast, close to 100% of these CpGs were methylated in *Gtl2*^{off} iPSC lines (Fig. 3b and Supplementary Fig. 5). The other analysed CpG-rich regions remained unaffected (Supplementary Fig. 5). Imprinting of the *Dlk1–Dio3* cluster is also regulated by histone acetylation³⁶, and chromatin immunoprecipitation experiments indeed revealed a significant decrease in acetylated H3 and H4 as well as in methylated H3K4, another activation mark, in *Gtl2*^{off} iPSC lines compared with *Gtl2*^{on} iPSC lines and ES cells (Fig. 3c). Together, these observations demonstrate that the normally expressed maternal *Gtl2* allele has acquired an aberrant paternal-like silent state in *Gtl2*^{off} iPSC clones.

Because imprinted gene expression is unstable in murine ES cells^{28,37}, we assessed *Gtl2* expression upon subcloning of iPSCs. *Gtl2* remained silent in subclones from *Gtl2*^{off} iPSCs and continued to be expressed in subclones from *Gtl2*^{on} iPSCs, demonstrating stability of its expression state in undifferentiated cells (Fig. 3d, top). This pattern was not altered if doxycycline was administered during the subcloning procedure (Fig. 3d, bottom), indicating that overexpression of the reprogramming factors in established iPSCs is insufficient to induce or revert silencing.

To assess if silencing of *Gtl2* is resolved during differentiation, we exposed iPSCs and ES cells to the differentiation-stimulating agent retinoic acid. Marked changes in cellular morphology and downregulation of *Pou5f1* indicated successful differentiation (Fig. 3e, f). Whereas retinoic-acid-treated *Gtl2*^{on} iPSCs and ES cells readily upregulated *Gtl2*

(Fig. 3f, top) and *Rian* (Supplementary Fig. 6), *Gtl2*^{off} iPSCs showed stable silencing of these genes, demonstrating that *in vitro* differentiation fails to reactivate maternally imprinted genes in the *Dlk1–Dio3* cluster. The expression of imprinted genes outside of chromosome 12qF1 was not affected (Fig. 3f, bottom, and Supplementary Fig. 6).

We next sought to gain an insight into the causes for the failure of *Gtl2*^{off} iPSCs to produce viable all-iPSC mice. To this end, we determined if these cells could autonomously support development into early embryos. Indeed, injection of both *Gtl2*^{off} and *Gtl2*^{on} iPSC clones into 4n blastocysts gave rise to normal-appearing mid-gestation (embryonic day (E) 11.5) embryos (Fig. 4a). However, the number of live embryos obtained from *Gtl2*^{off} clones was substantially reduced compared with *Gtl2*^{on} clones (Fig. 4b), suggesting that *Gtl2*^{off} mice die around this developmental stage. This phenotype resembles that of mice with paternal uniparental disomy of distal chromosome 12³⁸, which die before E16.5, but is distinct from the phenotype of maternal *Gtl2* knockout mice (*Gtl2*^{mKO}), which die perinatally³¹. The less severe phenotype of *Gtl2*^{mKO} embryos might be due to the comparably modest reduction in maternally expressed 12qF1 genes³¹. In agreement with this notion, we found low but detectable levels of *Rian* and *Mirg* transcripts in *Gtl2*^{mKO} MEFs (Fig. 4c), whereas these genes were almost completely silenced in MEFs and different tissues derived from *Gtl2*^{off} all-iPSC embryos (Fig. 4d, e).

Notably, expression of the *Dlk1* gene, which is reciprocally imprinted to *Gtl2* (ref. 39), was upregulated in *Gtl2*^{off} MEFs but not in *Gtl2*^{mKO} MEFs (Fig. 4c), further supporting the observation

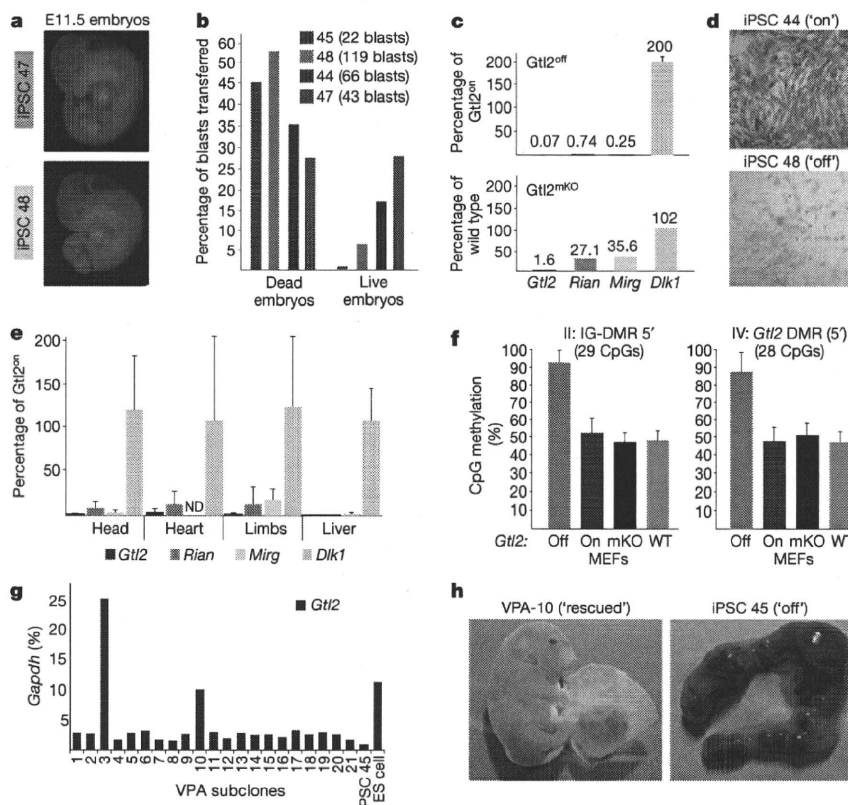


Figure 4 | Developmental defects in embryos derived from *Gtl2*^{off} iPSCs. **a**, Images of 'all-iPSC' E11.5 embryos obtained with *Gtl2*^{on} clone 47 and *Gtl2*^{off} clone 48, both of which express EGFP ubiquitously from the *Rosa26* locus. **b**, Frequency of dead and live E11.5 all-iPSC embryos obtained with two *Gtl2*^{on} (red bars) and two *Gtl2*^{off} (green bars) iPSC clones upon 4n blastocyst injection. Number of blastocysts transferred per clone is indicated in brackets. **c**, Expression of *Gtl2*, *Rian*, *Mirg* and the paternally expressed gene *Dlk1* in *Gtl2*^{off} MEFs relative to *Gtl2*^{on} MEFs (upper panel) as well as in *Gtl2*^{mKO} MEFs relative to MEFs isolated from wild-type embryos (lower panel). **d**, *In situ* hybridization for *Gtl2* mRNA in MEFs derived from all-iPSC embryos generated with either *Gtl2*^{on} clone 44 or *Gtl2*^{off} clone 48.

e, Expression levels of *Gtl2*, *Rian*, *Mirg* and *Dlk1* in the indicated tissues isolated from all-iPSC embryos produced with *Gtl2*^{off} iPSCs relative to the levels seen in tissues derived from *Gtl2*^{on} iPSCs. **f**, Degree of DNA methylation at the indicated *Dlk1–Dio3* regions in *Gtl2*^{off}, *Gtl2*^{on}, *Gtl2*^{mKO} and wild-type MEFs. **g**, *Gtl2* expression levels in iPSC lines derived by subcloning *Gtl2*^{off} clone 45 in the presence of valproic acid (VPA). **h**, Images of a neonatal stillborn pup (left) and a uterus filled with resorptions (right) derived after 4n blastocyst injections with either VPA-10 or the parental iPSC clone 45, respectively. All error bars indicate standard deviations with *n* = 3 in **c**, *n* = 5 in **e** and *n* = number of CpGs in **f**.

that the maternal *Dlk1–Dio3* cluster has acquired a paternal-like expression state in *Gtl2*^{off} iPSCs. Accordingly, IG-DMR and *Gtl2* DMR were hypermethylated in *Gtl2*^{off} MEFs but remained unaffected in *Gtl2*^{mkKO} MEFs (Fig. 4f). Together, these observations are in agreement with the notion that stable epigenetic silencing of the *Dlk1–Dio3* locus is the cause for the developmental failure of *Gtl2*^{off} all-iPSC embryos.

Reversal of epigenetic silencing in *Gtl2*^{off} iPSCs

ES cells derived from cloned embryos are transcriptionally indistinguishable from ES cells produced from fertilized embryos and support the development of all-ES-cell mice, regardless of donor cell identity^{21,22,40,41}, indicating that nuclear transfer generates faithfully reprogrammed pluripotent cells (Supplementary Fig. 7a). In agreement with this observation, we found that *Gtl2* was expressed in previously characterized 4n complementation-competent nuclear transfer ES cell lines derived from fibroblasts and haematopoietic cells²¹ (Supplementary Fig. 7b).

We next tested whether nuclear transfer could reverse the aberrant silencing of genes within the *Dlk1–Dio3* cluster in *Gtl2*^{off} iPSCs and rescue their ability to support the development of all-iPSC mice (Supplementary Fig. 7c). We derived nine nuclear transfer ES cell lines⁴² from *Gtl2*^{off} iPSCs that had been generated from tail-tip fibroblasts and fetal liver cultures using adenoviral vectors⁴³ or from haematopoietic stem cells and granulocytes using the collagen-OKSM system. Some of these iPSCs were germline competent⁴³, indicating that they were genetically normal, but failed to give rise to all-iPSC mice (Supplementary Table 6). Global transcriptome analysis showed no consistent differences in mRNA and miRNA expression between nuclear transfer ES cells and the donor iPSC clones. Importantly, *Gtl2* and *Rian* remained repressed in all lines (Supplementary Fig. 7d). Accordingly, these cells failed to generate all-iPSC mice (Supplementary Table 6), indicating that nuclear transfer cannot reset the faulty gene expression pattern acquired during iPSC generation. This notion is consistent with the previous finding that aberrant genomic imprints present in donor cells cannot be restored in cloned animals after nuclear transfer²⁸.

Given that *Gtl2*^{off} iPSC clones showed reduced histone acetylation at the *Gtl2* locus (Fig. 3c), we wondered whether their treatment with the histone deacetylase inhibitor valproic acid (VPA) could reactivate the silenced gene cluster. Indeed, 2 out of 21 subclones treated with VPA exhibited increased *Gtl2* expression with one iPSC clone showing expression levels comparable to ES cells (Fig. 4g). Consistent with transcriptional reactivation of the cluster, we found re-appearance of H3 acetylation and H3K4 methylation at the *Gtl2* locus (Supplementary Fig. 8). Injection of this rescued clone into 4n blastocysts gave rise to apparently normal E11.5 embryos at frequencies similar to those seen with *Gtl2*^{on} iPSC clones (Fig. 4b and Supplementary Fig. 9a). These embryos expressed *Gtl2*, *Rian* and *Mirg* (Supplementary Fig. 10a) and showed normal expression levels of tissue-specific marker genes such as *Ascl1* (also known as *Mash-1*) and *Hes5* that were repressed in *Gtl2*^{off} embryos (Supplementary Fig. 10b). Importantly, the rescued clone supported the development of several full-term all-iPSC pups, which was never seen with the parental iPSCs or any other *Gtl2*^{off} iPSC clone (Fig. 4h and Supplementary Fig. 9b). These pups were severely overgrown, however, and hence non-viable. We surmise that the elevated levels of *Dlk1* in the rescued clone (Supplementary Fig. 10a) is responsible for this phenotype, as over-expression of this gene causes neonatal lethality owing to fetal overgrowth³⁴. Alternatively, VPA treatment may have caused the dysregulation of other genes even though we found no aberrant expression of several candidate imprinted genes implicated in growth control (Supplementary Fig. 10c).

Discussion

Our data indicate that the expression of a surprisingly small number of transcripts, which localize to a single cluster in the genome,

distinguishes most mouse iPSCs from ES cells and is predictive for their developmental potential. Silencing of *Dlk1–Dio3* was seen in iPSCs derived from different cell types and using various reprogramming approaches, including the collagen-OKSM system as well as retroviral and adenoviral transduction, suggesting that aberrant repression of this locus is a general consequence of factor-mediated reprogramming in mouse cells. It remains to be tested, however, (1) if other consistent expression changes exist in iPSCs that are not detectable in the undifferentiated state or with current profiling technology; (2) whether iPSC-specific non-recurrent genetic and epigenetic aberrations are present in only some clones; and (3) if iPSCs derived from more defined cell populations than fibroblasts exhibit cell-of-origin-specific transcriptional and epigenetic patterns in addition to the observed silencing of *Gtl2*. It is noteworthy in this respect that all of the *Gtl2*^{on} iPSC clones described here or published previously were derived from fibroblast cultures, which express high levels of *Gtl2*, whereas the reprogramming of haematopoietic cells, which exhibit low *Gtl2* levels, uniformly yielded *Gtl2*^{off} iPSC clones. This observation raises the question of whether the expression status of *Gtl2* in the somatic cell of origin may affect silencing of the locus in resultant iPSCs and whether *Gtl2*^{on} iPSC clones can be derived from somatic cells other than fibroblasts. Another important point that needs to be addressed is whether similar expression abnormalities are seen in human iPSCs. Although a preliminary evaluation of published expression data⁴⁴ did not indicate aberrant regulation of the *Gtl2* homologue *MEG3* in human iPSCs compared with ES cells, our observations in mouse suggest that a meaningful answer to this question requires the establishment of genetically matched human ES cells and iPSCs.

Epigenetic analysis of the *Dlk1–Dio3* locus indicates that the stable repression of maternally expressed genes in *Gtl2*^{off} iPSC clones is due to DNA hypermethylation and hypoacetylation of key imprint control regions, leading to a 'paternalization' of the maternally inherited *Dlk1–Dio3* cluster. The homogeneous expression levels of *Gtl2* in fibroblasts and the observation that nuclear transfer ES cells derived from various cell types express normal levels of *Gtl2* rule out the possibility that aberrant silencing of the locus pre-existed in the somatic cell of origin. Rather, the iPSC reprogramming procedure itself seems to introduce these epigenetic changes by an as-yet-undefined mechanism that may point towards an involvement of one or more of the reprogramming factors in the control of *Gtl2* in germ cells, where imprints are established. Notably, the pluripotency factors Oct4 and Sox2, which are both expressed in germ cells, have been recently implicated in the regulation of the *Xist* and *Tsix* transcripts that control epigenetic silencing of the X chromosome in female embryos^{45,46}. Our findings also support the notion that nuclear transfer generates faithfully reprogrammed cells more frequently than factor-mediated reprogramming. A better understanding of the causes for the specific silencing of the *Dlk1–Dio3* cluster during iPSC formation will shed light on the molecular mechanisms of reprogramming as well as on the epigenetic regulation of this particular locus. These studies may also lead to improved reprogramming strategies, such as alternative factor combinations or modified culture conditions, that faithfully establish a fully pluripotent state in somatic cells.

METHODS SUMMARY

ES cells and iPSCs were cultured in ES cell media containing 15% FBS and 1,000 U ml⁻¹ LIF. For the reprogramming of somatic cells the media was supplemented with 1 ng μl⁻¹ doxycycline and, for haematopoietic cell reprogramming, with cell-type-specific cytokines^{24,47}. For qPCR and microarray analyses, ES cells or iPSCs were trypsinized and pre-plated onto cell culture flasks for 45 min to remove feeder cells. Total RNA was extracted from the non-adherent cell suspension using the miRNeasy kit (Qiagen). Microarrays were done using 430 2.0 and 430A 2.0 chips (Affymetrix) followed by analysis with GeneSifter software for mRNAs or using the miRCURY LNA array platform (Exiqon). Injections into diploid and tetraploid BDF1 blastocysts were done as described previously¹¹.

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

Received 6 September 2009; accepted 17 March 2010.
Published online 25 April 2010.

- Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
- Hochedlinger, K. & Plath, K. Epigenetic reprogramming and induced pluripotency. *Development* **136**, 509–523 (2009).
- Maherali, N. *et al.* A high-efficiency system for the generation and study of human induced pluripotent stem cells. *Cell Stem Cell* **3**, 340–345 (2008).
- Mikkelsen, T. S. *et al.* Dissecting direct reprogramming through integrative genomic analysis. *Nature* **454**, 49–55 (2008).
- Wernig, M. *et al.* *In vitro* reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
- Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
- Boland, M. J. *et al.* Adult mice generated from induced pluripotent stem cells. *Nature* **461**, 91–94 (2009).
- Zhao, X. Y. *et al.* iPS cells produce viable mice through tetraploid complementation. *Nature* **461**, 86–90 (2009).
- Kang, L., Wang, J., Zhang, Y., Kou, Z. & Gao, S. iPS cells can support full-term development of tetraploid blastocyst-complemented embryos. *Cell Stem Cell* **5**, 135–138 (2009).
- Nagy, A. *et al.* Embryonic stem cells alone are able to support fetal development in the mouse. *Development* **110**, 815–821 (1990).
- Eggan, K. *et al.* Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl Acad. Sci. USA* **98**, 6209–6214 (2001).
- Kim, J. B. *et al.* Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature* **454**, 646–650 (2008).
- Hanna, J. *et al.* Direct reprogramming of terminally differentiated mature B lymphocytes to pluripotency. *Cell* **133**, 250–264 (2008).
- Meissner, A., Wernig, M. & Jaenisch, R. Direct reprogramming of genetically unmodified fibroblasts into pluripotent stem cells. *Nature Biotechnol.* **25**, 1177–1181 (2007).
- Chin, M. H. *et al.* Induced pluripotent stem cells and embryonic stem cells are distinguished by gene expression signatures. *Cell Stem Cell* **5**, 111–123 (2009).
- Marchetto, M. C. *et al.* Transcriptional signature and memory retention of human-induced pluripotent stem cells. *PLoS One* **4**, e7076 (2009).
- Wilson, K. D. *et al.* MicroRNA profiling of human-induced pluripotent stem cells. *Stem Cells Dev.* **18**, 749–758 (2009).
- Pick, M. *et al.* Clone- and gene-specific aberrations of parental imprinting in human induced pluripotent stem cells. *Stem Cells* **27**, 2686–2690 (2009).
- Doi, A. *et al.* Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nature Genet.* **41**, 1350–1353 (2009).
- Deng, J. *et al.* Targeted bisulfite sequencing reveals changes in DNA methylation associated with nuclear reprogramming. *Nature Biotechnol.* **27**, 353–360 (2009).
- Brambrink, T., Hochedlinger, K., Bell, G. & Jaenisch, R. ES cells derived from cloned and fertilized blastocysts are transcriptionally and functionally indistinguishable. *Proc. Natl Acad. Sci. USA* **103**, 933–938 (2006).
- Wakayama, S. *et al.* Equivalency of nuclear transfer-derived embryonic stem cells to those derived from fertilized mouse blastocysts. *Stem Cells* **24**, 2023–2033 (2006).
- Soldner, F. *et al.* Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. *Cell* **136**, 964–977 (2009).
- Stadtfield, M., Maherali, N., Borkent, M. & Hochedlinger, K. A reprogrammable mouse strain from gene-targeted embryonic stem cells. *Nature Methods* **7**, 53–55 (2009).
- Sommer, C. A. *et al.* Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. *Stem Cells* **27**, 543–549 (2009).
- Beard, C., Hochedlinger, K., Plath, K., Wutz, A. & Jaenisch, R. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**, 23–28 (2006).
- da Rocha, S. T., Edwards, C. A., Ito, M., Ogata, T. & Ferguson-Smith, A. C. Genomic imprinting at the mammalian *Dlk1-Dio3* domain. *Trends Genet.* **24**, 306–316 (2008).
- Humpherys, D. *et al.* Epigenetic instability in ES cells and cloned mice. *Science* **293**, 95–97 (2001).
- Seitz, H. *et al.* Imprinted microRNA genes transcribed antisense to a reciprocally imprinted retrotransposon-like gene. *Nature Genet.* **34**, 261–262 (2003).
- Seitz, H. *et al.* A large imprinted microRNA gene cluster at the mouse *Dlk1-Gtl2* domain. *Genome Res.* **14**, 1741–1748 (2004).
- Takahashi, N. *et al.* Deletion of *Gtl2*, imprinted non-coding RNA, with its differentially methylated region induces lethal parent-origin-dependent defects in mice. *Hum. Mol. Genet.* **18**, 1879–1888 (2009).
- Lin, S. P. *et al.* Differential regulation of imprinting in the murine embryo and placenta by the *Dlk1-Dio3* imprinting control region. *Development* **134**, 417–426 (2007).
- Steshina, E. Y. *et al.* Loss of imprinting at the *Dlk1-Gtl2* locus caused by insertional mutagenesis in the *Gtl2* 5' region. *BMC Genet.* **7**, 44 (2006).
- da Rocha, S. T. *et al.* Gene dosage effects of the imprinted delta-like homologue 1 (*dlk1/pref1*) in development: implications for the evolution of imprinting. *PLoS Genet.* **5**, e1000392 (2009).
- Lin, S. P. *et al.* Asymmetric regulation of imprinting on the maternal and paternal chromosomes at the *Dlk1-Gtl2* imprinted cluster on mouse chromosome 12. *Nature Genet.* **35**, 97–102 (2003).
- Carr, M. S., Yevtdiyenko, A., Schmidt, C. L. & Schmidt, J. V. Allele-specific histone modifications regulate expression of the *Dlk1-Gtl2* imprinted domain. *Genomics* **89**, 280–290 (2007).
- Dean, W. *et al.* Altered imprinted gene methylation and expression in completely ES cell-derived mouse fetuses: association with aberrant phenotypes. *Development* **125**, 2273–2282 (1998).
- Tevendale, M., Watkins, M., Rasberry, C., Cattanch, B. & Ferguson-Smith, A. C. Analysis of mouse conceptuses with uniparental duplication/deficiency for distal chromosome 12: comparison with chromosome 12 uniparental disomy and implications for genomic imprinting. *Cytogenet. Genome Res.* **113**, 215–222 (2006).
- Schmidt, J. V., Matteson, P. G., Jones, B. K., Guan, X. J. & Tilghman, S. M. The *Dlk1* and *Gtl2* genes are linked and reciprocally imprinted. *Genes Dev.* **14**, 1997–2002 (2000).
- Eggan, K. *et al.* Mice cloned from olfactory sensory neurons. *Nature* **428**, 44–49 (2004).
- Li, J., Ishii, T., Feinstein, P. & Mombaerts, P. Odorant receptor gene choice is reset by nuclear transfer from mouse olfactory sensory neurons. *Nature* **428**, 393–399 (2004).
- Ono, Y. & Kono, T. Irreversible barrier to the reprogramming of donor cells in cloning with mouse embryos and embryonic stem cells. *Biol. Reprod.* **75**, 210–216 (2006).
- Stadtfield, M., Nagaya, M., Utikal, J., Weir, G. & Hochedlinger, K. Induced pluripotent stem cells generated without viral integration. *Science* **322**, 945–949 (2008).
- Yu, J. *et al.* Human induced pluripotent stem cells free of vector and transgene sequences. *Science* **324**, 797–801 (2009).
- Navarro, P. *et al.* Molecular coupling of *Xist* regulation and pluripotency. *Science* **321**, 1693–1695 (2008).
- Donohoe, M. E., Silva, S. S., Pinter, S. F., Xu, N. & Lee, J. T. The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* **460**, 128–132 (2009).
- Eminli, S. *et al.* Differentiation stage determines potential of hematopoietic cells for reprogramming into induced pluripotent stem cells. *Nature Genet.* **41**, 968–976 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to H. Arnold for assistance with GeneSifter analysis; K. Coser, K. Claycomb and P. August for technical support on Affymetrix expression profiling; S. Sato and M. Machida for technical assistance; V. Greco for advice on keratinocyte isolation; and S. Schubert for advice on miRNA isolation. We thank members of the Hochedlinger laboratory for helpful suggestions and A. Umezawa for discussions and support. M.S. was supported by a postdoctoral fellowship from the Schering Foundation, E.A. was supported by a Jane Coffin Childs postdoctoral fellowship and K.H. was supported by a NIH Director's Innovator Award and by funds provided by the Harvard Stem Cell Institute, MGH and HHMI.

Author Contributions M.S., E.A. and K.H. conceived the ideas for this study, designed and analysed experiments and wrote the manuscript. M.S. derived iPSC lines, conducted *in vitro* differentiation assays and performed expression array analysis. E.A. conducted qPCR analyses, *in situ* hybridizations and chromatin immunoprecipitations. H.A. and A.F. performed nuclear transfer experiments. P.F. did blastocyst injections. T.S. performed microarray experiments and analyses. S.N. and T.K. provided important study materials.

Author Information The mRNA profiling data discussed in this paper have been deposited in NCBI's Gene Expression Omnibus and are accessible through GEO series accession number GSE20576. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.H. (khochedlinger@helix.mgh.harvard.edu).

METHODS

Cell culture. ES cells and iPSCs were cultured in ES cell medium (DMEM with 15% FBS, L-glutamin, penicillin-streptomycin, non-essential amino acids, β -mercaptoethanol and $1,000 \text{ U ml}^{-1}$ LIF) on irradiated feeder cells. Mouse embryonic fibroblasts (MEFs) were isolated by trypsin digestion of mid-gestation (E14.5) ES-cell-chimaeric embryos followed by culture in fibroblast medium (DMEM with 10% FBS, L-glutamin, penicillin-streptomycin, non-essential amino acids and β -mercaptoethanol). A total of $1 \mu\text{g ml}^{-1}$ puromycin was added to these cultures for 5 days to select for ES-cell-derived cells. Tail-tip fibroblast (TTF) cultures were established by trypsin digestion of tail-tip biopsies taken from newborn (3–8 days of age) chimaeric mice derived after blastocyst injection of ROSA26-EGFP targeted ES cells. ES-cell-derived cells were isolated based on GFP expression and maintained in fibroblast medium. For the establishment of peritoneal fibroblast cultures, adult mice were killed and $1\text{--}2 \text{ cm}^2$ of peritoneal body wall isolated and chopped into small pieces in 0.25% trypsin/EDTA in a 35-mm cell culture vessel. After 5 min of incubation at 37°C , 6 ml fibroblast medium was added and the tissue re-suspended several times through a pipette. Peritoneal fibroblast cultures were maintained and propagated like MEF and TTF cultures. To derive iPSCs from fibroblast cultures, cells were trypsinized and seeded on top of irradiated feeder layers in fibroblast medium. The next day, ES cell medium containing $1 \mu\text{g ml}^{-1}$ doxycycline was added and replenished every 3 days. Doxycycline was withdrawn after 10–20 days and colonies picked that maintained ES cell morphology after another 4 days of culture in regular ES cell medium. Haematopoietic cells were isolated from bone marrow as previously described⁴⁷. Briefly, freshly isolated mononucleated bone marrow cells were separated into the desired subfractions by FACS using the following surface marker combinations: $\text{CD150}^+\text{CD48}^-\text{ckit}^+\text{Sca-1}^-\text{lineage}^-$ for haematopoietic stem cells, $\text{Fc}\gamma\text{R}^+\text{CD34}^+\text{ckit}^+\text{Sca-1}^-\text{lineage}^-$ for granulocyte-macrophage progenitors and $\text{CD11b}^{\text{high}}\text{Gr-1}^{\text{high}}\text{ckit}^-$ for granulocytes. Sorted cells were plated on top of irradiated feeder layers in ES cell medium containing doxycycline. For haematopoietic stem cells and granulocyte-macrophage progenitors, the medium was in some cases supplemented with Flt3 ligand ($10 \text{ ng } \mu\text{l}^{-1}$), SCF ($10 \text{ ng } \mu\text{l}^{-1}$) and TPO ($10 \text{ ng } \mu\text{l}^{-1}$). Doxycycline was withdrawn after 2 weeks and colonies with ES cell morphology picked after another 4 days of culture in regular ES cell medium. For the establishment of keratinocyte cultures the back skin of newborn mice was placed onto 0.25% trypsin (epidermis facing up) and cultured overnight at 4°C . Using fine forceps, the epidermis was then separated from the dermis and incubated in 0.25% trypsin at 37°C for 15 min, followed by mechanical dissociation using a pipette. Cells were then filtered using a $40 \mu\text{m}$ strainer and cultured in CnT-07 medium (CELLnTEC). After 2 weeks of culture, keratinocytes were further purified by FACS sorting using the surface marker combination $\text{CD49}^+\text{CD90}^-$. Sorted cells were seeded on top of irradiated feeder layers in CnT-07 medium containing doxycycline, which was replaced with ES cell medium plus doxycycline after 3 days. Doxycycline was withdrawn after a total of 14 days and colonies picked on day 20.

Generation of OKSM ES cells. A polycistronic cassette encoding Oct4, Klf4, Sox2 and c-Myc (previously called STEMCCA²⁵) was cloned into the shuttle plasmid pBS31 using NotI/ClaI digestion. The resulting plasmid was electroporated into KH2 ES cells²⁶ together with a plasmid driving expression of Flp recombinase. Correctly targeted clones were isolated by hygromycin selection and confirmed by Southern blot analysis as previously described²⁶. Individual OKSM ES cell subclones were gene targeted with ROSA26-EGFP as has been described before⁴⁸ to facilitate tracking of ES-cell-derived cells after blastocyst injection. OKSM ES cells and derivative mice are described in detail elsewhere²⁴.

Quantitative PCR. ES cells and iPSCs grown on 35-mm dishes were harvested when they reached about 50% confluency and pre-plated on non-gelatinized T25 flasks for 45 min to remove feeder cells. Cells were spun down and the pellet used for isolation of total RNA using the miRNeasy Mini Kit (Qiagen) without DNase digestion. RNA was eluted from the columns using $50 \mu\text{l}$ RNase-free water or Tris-EDTA (TE) buffer, pH 7.5 (10 mM Tris-HCl and 0.1 mM EDTA) and quantified using a NanoDrop spectrophotometer (Thermo Scientific). cDNA was produced with the First Strand cDNA Synthesis Kit (Roche) using $1 \mu\text{g}$ of total RNA input. Real-time quantitative PCR reactions were set up in triplicate using $5 \mu\text{l}$ of cDNA (1:100 dilution) with the Brilliant II SYBR Green QPCR Master Mix (Stratagene) and run on a Mx3000P QPCR System (Stratagene). Primer sequences are listed in Supplementary Table 7.

Transcriptome profiling. Analysis of mRNA expression by Affymetrix GeneChip 430 2.0 microarray (used for comparing nuclear transfer ES cells and parental iPSCs) was performed similarly to the method described previously⁴⁹. Briefly, total RNA samples were subjected to Lab-on-a-Chip RNA analysis to determine their RNA integrity number (RIN). Biotinylated antisense cRNA probes were synthesized for RNA samples with 9.0 or greater RIN using Affymetrix One-Cycle target labelling and control reagents (for cDNA synthesis)

and Enzo Bioarray high-yield RNA transcript labelling kit (for cRNA synthesis and labelling). Labelled probes were hybridized to microarrays overnight. Microarray chips were then stained with avidin-conjugated phycoerythrin fluorescence dye and washed using Affymetrix Fluidics Station 450 (protocol: EukGE-WS2v5_450). Arrays were scanned by an Affymetrix 3000-7G scanner, and .CHP data files were generated by GCOS using the statistical algorithm and scaling (target intensity value = 150). Hierarchical clustering was performed using Cluster software and visualized by TreeView software⁵⁰. For analysis using Affymetrix 430A 2.0 arrays (all other samples), RNA was amplified using the Nugen Ovation amplification system using the manufacturer's protocol. Amplified samples were normalized to $17 \text{ ng } \mu\text{l}^{-1}$ and $1.5 \mu\text{g}$ of total cDNA was hybridized to the arrays for 16 h at 45°C . Arrays were washed and scanned using the HT Array Plate Scanner according to the manufacturer's protocol. GeneSifter microarray data analysis server (Geospiza) was used for statistical analyses. For miRNA analysis, total RNA was subjected to quality control consisting of RNA measurement on the NanoDrop ($A_{260/230 \text{ nm}}$ and $A_{260/280 \text{ nm}}$ had to be greater than 1.8) and a run on the Agilent Bioanalyser 2100 (RIN values had to be higher than 7). The samples were then labelled using the miRCURY Hy3/Hy5 power labelling kit (Exiqon) and hybridized on the miRCURY LNA Array (v.11.0) (Exiqon). Labelling was determined to be successful when all capture probes for the control spike-in oligonucleotides produced signals in the expected range. The quantified signals were normalized using the global Lowess (locally weighted scatterplot smoothing) regression algorithm.

Blastocyst injections. 2n and 4n blastocyst injections were performed as described before¹¹. Briefly, female BDF1 mice were superovulated by intraperitoneal injection of PMS and hCG and mated to BDF1 stud males. Zygotes were isolated from females with a vaginal plug 24 h after hCG injection. Zygotes for 2n injections were *in vitro* cultured for 3 days *in vitro* in KSOM media, blastocysts were identified, injected with ES cells or iPSCs and transferred into pseudopregnant recipient females. For 4n injections, zygotes were cultured overnight until they reached the 2-cell stage, at which point they were electrofused. One hour later, 1-cell embryos were carefully identified and separated from embryos that had failed to fuse, cultured in KSOM for another 2 days and then injected.

Nuclear transfer. Nuclear transfer was performed as previously described⁴². Briefly, donor iPSCs were cultured in collagen-coated dishes without a feeder layer for 3 days in standard ES cell medium. To synchronize cells at metaphase, the cultures were cultured for 2 h in a medium containing $0.4 \mu\text{g ml}^{-1}$ nocodazole (Sigma-Aldrich), a microtubule polymerization inhibitor. Cells floating in the medium were collected. While being sucked into a transfer pipette, only the cells arrested at metaphase were selected and used as nuclear donors. The recipient oocytes were collected from mature B6CBF1 female mice. Micromanipulations were performed in M2 medium containing $5 \mu\text{g ml}^{-1}$ cytochalasin B (Sigma) and $1 \mu\text{g ml}^{-1}$ nocodazole in a micromanipulation chamber. Explantation of cloned blastocysts and ES-cell derivation was done as described previously⁴².

Chromatin immunoprecipitation. 20 million iPSCs, ES cells or MEFs were fixed with 1% formaldehyde for 10 min at 25°C and then lysed in 1 ml lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS, protease inhibitors) for 20 min on ice. The lysate was split into three tubes and sonicated using Bioruptor for five times 5 min at high intensity, 30 s on –30 s off. After 10 min centrifugation, the supernatant was pre-cleared for 1 h at 4°C with agarose beads pre-blocked with BSA (1 mg BSA for 10 ml beads) in IP buffer (50 mM Tris-HCl, pH 8, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS, protease inhibitors). A total of 100 μl of pre-cleared chromatin per reaction diluted in 1 ml IP buffer in the presence of $2 \mu\text{g}$ antibody was used for each immunoprecipitation reaction according to the manufacturer's protocol. The antibodies used for this study were: anti-acH3 (06-599 Millipore), anti-acH4 (06-866, Millipore), anti-dimethyl K4 of H3 (07-030, Millipore), anti-trimethyl K27 of H3 (ab6002, Abcam) and normal rabbit IgG (Millipore). The precipitate was purified using Qiaquick PCR purification kit and was analysed by qPCR using Brilliant II SYBR Green qPCR Master Mix (600828, Agilent Technologies) using the sequence specific primer sets. *Glt2*, 5'-AGCCCTGACTGATGT TCTG-3' (forward) and 5'-TGGAAGGGCGATTGGTAGAC-3' (reverse); and *Pou5f1*, 5'-GGAGGTGCAATGGCTGTCTGTGCC-3' (forward) and 5'-CTGCCTTGGGTACCTTACACCTCAC-3' (reverse).

In situ hybridization. MEFs grown on coverslips were fixed with 4% formaldehyde/5% acetic acid in PBS for 15 min at 25°C . After extensive PBS washes, they were dehydrated in 70% ethanol and left overnight at 4°C . The next day, they were rehydrated in a series of ethanol dilutions and incubated in hybridization buffer (50% formamide-5X SSC-RNase inhibitors) for 1 h at 65°C . The hybridization was done overnight in a humidified chamber using 400 ng of sense or antisense *Glt2*-specific probe per ml of hybridization buffer. The sense and antisense probes were synthesized by *in vitro* transcription with DIG RNA labelling mix (Roche) and SP6 and T7 polymerase, respectively, using *Glt2* cDNA

amplified with the primers 5'-CTCTCGGGACTCCTGGCTCCAC-3' (forward) and 5'-GGGTCCAGCATGTCCCACAGGA-3' (reverse). The cells were serially washed and stained with an anti-DIG AP conjugated Fab fragment (1:2,000 in blocking buffer) for 1 h at room temperature. The detection was performed with NBT/BCIP reagent.

Pyrosequencing. Genomic DNA was isolated using the DNeasy Blood and Tissue Kit (Qiagen). ES cells and iPSCs were pre-plated onto cell culture vessels for 45 min after harvesting to remove feeder cells. Genomic DNA was bisulphite-converted using the EpiTect Bisulfite Kit (Qiagen) with 400 ng of input DNA. DNA was eluted with 10 ml and 1 ml of it was used for PCR. PCR products were sequenced using the Pyrosequencing PSQ96 HS System (Biotage AB) following

the manufacturer's instructions. The methylation status of each locus was analysed using QCpG software (Biotage).

48. Hochedlinger, K., Yamada, Y., Beard, C. & Jaenisch, R. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121**, 465–477 (2005).
49. Coser, K. R. *et al.* Global analysis of ligand sensitivity of estrogen inducible and suppressible genes in MCF7/BUS breast cancer cells by DNA microarray. *Proc. Natl Acad. Sci. USA* **100**, 13994–13999 (2003).
50. Eisen, M. B., Spellman, P. T., Brown, P. O. & Botstein, D. Cluster analysis and display of genome-wide expression patterns. *Proc. Natl Acad. Sci. USA* **95**, 14863–14868 (1998).

Production of a Non-Triple Helical Collagen α Chain in Transgenic Silkworms and Its Evaluation as a Gelatin Substitute for Cell Culture

Takahiro Adachi,^{1,2} Xiaobiao Wang,¹ Tomoko Murata,¹ Masanobu Obara,³ Hidenori Akutsu,⁴ Masakazu Machida,⁴ Akirihito Umezawa,⁴ Masahiro Tomita¹

¹Neosilk Co., Ltd., 3-13-26 Kagamiyama, Higashihiroshima, Hiroshima 739-0046, Japan; telephone: +81 82 431 0652; fax: +81 82 431 0654; e-mail: mtomita@neosilk.co.jp

²Hiroshima Prefectural Institute of Industrial Science and Technology, Higashihiroshima, Hiroshima, Japan

³Developmental Biology Laboratory, Department of Biological Science, Graduate School of Science, Hiroshima University, Higashihiroshima, Hiroshima, Japan

⁴National Center for Child Health and Development, Setagaya-ku, Tokyo, Japan

Received 16 December 2009; revision received 25 March 2010; accepted 29 March 2010

Published online 8 April 2010 in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/bit.22752

ABSTRACT: We generated transgenic silkworms that synthesized human type I collagen $\alpha 1$ chain [$\alpha 1(I)$ chain] in the middle silk glands and secreted it into cocoons. The initial content of the recombinant $\alpha 1(I)$ chain in the cocoons of the transgenic silkworms was 0.8%. The IE1 gene, a trans-activator from the baculovirus, was introduced into the transgenic silkworm to increase the content of the chain. We also generated silkworms homozygous for the transgenes. These manipulations increased the $\alpha 1(I)$ chain content to 8.0% (4.24 mg per cocoon). The $\alpha 1(I)$ chain was extracted and purified from the cocoons using a very simple method. The $\alpha 1(I)$ chain contained no hydroxyprolines due to the absence of prolyl-hydroxylase activity in the silk glands. Circular dichroism analysis showed that the secondary structure of the $\alpha 1(I)$ chain is similar to that of denatured type I collagen, demonstrating the absence of the triple helical structure. Human skin fibroblasts were seeded on the $\alpha 1(I)$ chain-coated dishes. The cells attached and spread, although at decreased chain concentrations the spreading rate was lower than that of the collagen and gelatin. Cynomolgus monkey embryonic stem cells cultured on the $\alpha 1(I)$ chain-coated dishes maintained an undifferentiated state after 30 passages, and their pluripotency was confirmed by teratoma formation in severe combined immunodeficient mice. These results show that the recombinant human $\alpha 1(I)$ chain is a promising candidate biomaterial as a high-quality and safe gelatin substitute for cell culture.

Biotechnol. Bioeng. 2010;106: 860–870.

© 2010 Wiley Periodicals, Inc.

KEYWORDS: transgenic silkworm; cocoon; collagen; gelatin; cell culture

Introduction

Collagen, a protein with a triple helical structure composed of three α chains, is used as a biomaterial for a variety of medical and cosmetic applications (Lee et al., 2001; Miyata et al., 1992). Gelatin, the denatured and fragmented form of collagen, is applied in medicine as a material for oral drug delivery and parenteral formulations (Gill and Feinberg, 2001; Tabata and Ikada, 1998). These materials come mainly from animal sources, which carry a risk of pathogen contamination and can also cause allergic reactions (Bradley, 1993; Mullins et al., 1996). One promising approach to overcome such problems is to produce them by recombinant means using appropriate hosts. For example, collagens have been produced successfully using host cells possessing prolyl-hydroxylase activity, which is required for the formation of the triple helical structure (Berg and Prockop, 1973; Fichard et al., 1997; Geddis and Prockop, 1993). Collagens have also been produced in host cells that did not possess sufficient prolyl-hydroxylase activity for triple helix formation but were transfected with prolyl-hydroxylase genes (John et al., 1999; Lamberg et al., 1996; Merle et al., 2002; Vuorela et al., 1997). In contrast, cells lacking hydroxylase activity have been utilized as hosts for the synthesis of recombinant gelatins (Olsen et al., 2003; Werten et al., 1999).

We previously generated transgenic silkworms that synthesized a recombinant fusion protein containing the collagen sequence in their silk glands and secreted it into cocoons. Due to the absence of prolyl-hydroxylase activity in the glands, prolyl residues in the collagen sequence were not hydroxylated (Tomita et al., 2003, 2005). To produce

prolyl-hydroxylated protein, prolyl-hydroxylase genes were simultaneously introduced to the silkworms along with the recombinant protein gene. The resulting silkworms produced a protein containing a prolyl-hydroxylated collagen sequence in their cocoons (Adachi et al., 2006). While these studies demonstrated the possibility of mass production of recombinant collagens in transgenic silkworms, we recognized an important issue that needed to be addressed: given that the recombinant protein was incorporated into insoluble silk fibers, the protein could not be extracted without using strong chaotropic reagents.

Silk fibers are composed mainly of two types of protein: fibroin and sericin. The former comprises 70–80% of all silk proteins, constitutes the core of insoluble silk fibers, and is synthesized in the posterior silk glands (PSGs). Sericin, which accounts for 20–30% of silk proteins, refers to a group of hydrophilic glue proteins that surround the fibroin core and are synthesized in the middle silk glands (MSGs; Garel et al., 1997; Grzelak, 1995). The above-described recombinant fusion protein containing the collagen sequence was expressed in the PSGs, resulting in production of the recombinant protein into the insoluble fibroin core (Tomita et al., 2005). Recently, we also developed a sericin promoter-driven recombinant protein expression system, in which recombinant proteins are expressed in the MSGs and secreted into hydrophilic sericin layers of silk fibers. Green fluorescent protein (Tomita et al., 2007), human serum albumin (Ogawa et al., 2007), and mouse IgG (Iizuka et al., 2009) were successfully produced in cocoons. These proteins were extractable from cocoons by soaking them in mild neutral aqueous solutions such as phosphate-buffered saline (PBS) or Tris-buffered saline.

In this study, we generated transgenic silkworms that expressed the full-length triple helical region of the human type I collagen $\alpha 1$ chain [$\alpha 1(I)$ chain] in the MSGs. Type I collagen is the most abundant fibril-forming collagen in the human body. A common form of type I collagen is a heterotrimer composed of two $\alpha 1(I)$ chains and one $\alpha 2(I)$ chain. It is also known that $\alpha 1(I)$ chains are able to form a homotrimer in the absence of $\alpha 2(I)$ chain (Nicholls et al., 1979). The recombinant $\alpha 1(I)$ chain expressed in the MSGs was efficiently secreted into cocoons and was easily recoverable. Given that the silk glands had no prolyl-hydroxylase activity, the chains contained no hydroxyproline residues. In addition, the recombinant $\alpha 1(I)$ chain expressed in this study lacks the telopeptide and propeptide promoting triple helix formation (Doege and Fessler, 1986; Olsen et al., 2001; Rosenbloom et al., 1976). Therefore, the $\alpha 1(I)$ chains did not form the homotrimer with the triple helical structure, and their physio-chemical properties were similar to those of gelatin rather than collagen. We tested the possibility of using the recombinant $\alpha 1(I)$ chain as a cell culture scaffold, and found that cells could be cultured on the chain as on marketed gelatin. As the recombinant $\alpha 1(I)$ chain has uniform molecular weight and no risks of animal-derived pathogen contamination, the recombinant chain

may be useful as a high-quality and safe substitute for marketed gelatin.

Materials and Methods

Experimental Animals

A strain of silkworm, pnd-w1, was obtained from the National Institute of Agrobiological Science (Tsukuba, Japan). Larvae were reared at 25°C on an artificial diet (Nosan Corporation, Yokohama, Japan).

Vector Construction for Transgenic Silkworms

The vector COL1A1/pMSG for transgenic silkworms was constructed using a full-length cDNA coding for the pro $\alpha 1$ chain of human type I procollagen (GenBank Accession No. Z74615) obtained by RT-PCR from human placenta total RNA (Clontech, Palo Alto, CA). The PCR-amplified product was cloned into the pCR4-TOPO vector (Invitrogen, Carlsbad, CA) using a TOPO cloning system (Invitrogen), which yielded pCR4COL1A1. To obtain the DNA fragment containing the baculovirus polyhedrin 5'-untranslated region (UTR; nt 140–189; GenBank Accession No. M30925; Iizuka et al., 2008), the signal sequence of human calreticulin (nt 109–159; GenBank Accession No. M84739), and the cDNA for the triple helical region of the $\alpha 1(I)$ chain (nt 654–3695; Fig. 1A), PCR was performed using pCR4COL1A1 as a template. A set of primers, calSP/COL1A1-F (5'-ATGCTGCTATCCGTGCCGTTGCTGCTCGCCTCCTCGGCCGTGGCCGTCGCCGCCCCATGGGTCCCTCT-3') and *NruI*/STP/COL1A1-R (5'-TCGCGATTAGGGGAGGACCAGGG-GGACC-3'), was used for the first PCR. A second PCR was performed using the product of the first PCR as a template and *NruI*/Bm5UTR/calSP-F (5'-TCGCGAAAGTATTTTACTGTTTTCTGTAACAGTTTTGTAATAAAAAAACCTATAAATATGCTGCTATCCGTGCCG-3') and *NruI*/STP/COL1A1-R as a primer set. The amplified product was cloned into the pCR4-TOPO vector to generate pCR4BmUTRcalSPCOL1A1. The DNA fragment was re-excised from pCR4BmUTRcalSPCOL1A1 by digestion with *NruI*, and inserted between the *Bombyx mori* sericin 1 promoter and the fibroin light chain 3'-flanking sequence of pMSG1.1MG (Iizuka et al., 2009), giving rise to COL1A1/pMSG (Fig. 1B).

A vector carrying the gene of baculovirus trans-activator IE1 (Tomita et al., 2007) was prepared as follows. To obtain the DNA fragment consisting of the polyhedrin 5'-UTR and the IE1 gene, PCR was performed using pIE1 (Tomita et al., 2007) as a template. A set of primers, *EcoRV*/Bm5UTR/IE1-F (5'-GATATCAAGTATTTTACTGTTTTCTGTAACAGTTTTGTAATAAAAAAACCTATAAATATGACGCAAATTAATTTTAAACGCGTCCG-3') and *Bgl*/IE1-R (5'-AGATCTTTAATTTAAATTCAAATTTTTTATATTTACAA-TTTAG-3'), was used for the PCR. The amplified product

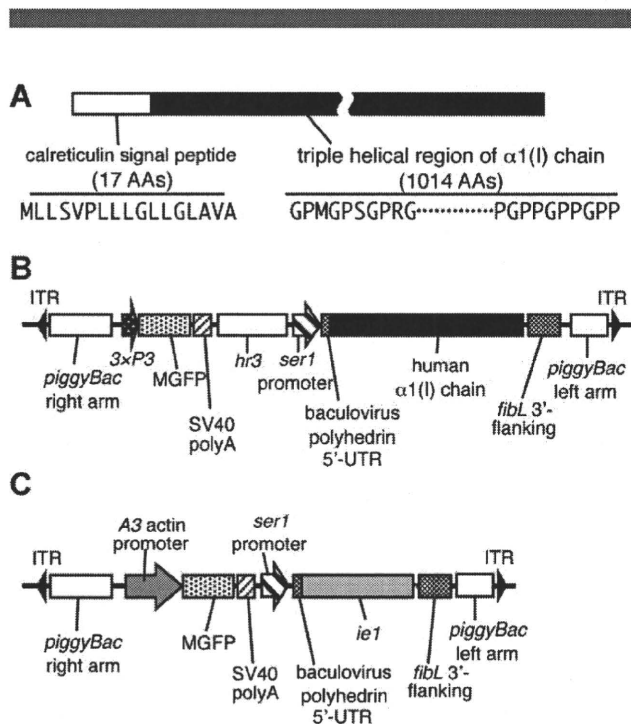


Figure 1. The structure of the transformation vector. **A:** The structure of the amino acid sequence of the $\alpha 1(I)$ chain. The sequence of the $\alpha 1(I)$ chain consists of a 17 amino acid (AA)-long calreticulin signal peptide (open box) and a 1,014 AA-long triple helical region of the $\alpha 1(I)$ chain (filled box). **B:** The structure of the transformation vector COL1A1/pMSG. COL1A1/pMSG contained two expression units between the right and left arms of *piggyBac* as follows: *3xP3* promoter-driven MGFP cDNA with the SV40 polyA signal sequence (SV40 polyA) and the baculovirus-derived enhancer *hr3*-linked sericin 1 (*ser1*) promoter-driven $\alpha 1(I)$ chain cDNA with the fibroin light chain gene polyA signal sequence (*fibL* 3'-flanking). ITR, inverted terminal repeat. **C:** Transformation vector pIM1. pIM1 contained two expression units between the right and left arms of *piggyBac* as follows: *Bombyx mori* A3 actin promoter-driven MGFP cDNA with the SV40 polyA and *ser1* promoter-driven *ie1* gene with the *fibL* 3'-flanking.

was cloned into pCRII-TOPO vector (Invitrogen) to generate pCRBmUTRie1. The DNA fragment was released from pCRBmUTRie1 by digestion with *EcoRV*, and inserted between the *B. mori* sericin1 promoter and the fibroin light chain 3'-flanking sequence of pMSG1.1A3-MG, which had been created by replacing the *3xP3* promoter with the *B. mori* A3 actin promoter (nt 1757–2595, GenBank Accession No. U49854) in pMSG1.1MG. The resulting vector was termed pIM1 (Fig. 1C).

Generation of Transgenic Silkworms

COL1A1/pMSG was injected with the helper vector pHA3PIG (Tamura et al., 2000) into pre-blastoderm embryos as described previously (Tomita et al., 2003). Hatched G_0 larvae were reared at 25°C to the moth stage. G_1 embryos obtained by sibling mating were screened for the expression of Monster Green Fluorescent Protein (MGFP; Promega, San Luis Obispo, CA) in the eyes and nervous tissues to obtain transgenic silkworms.

pIM1 was injected into eggs, and transgenic silkworms were created as described above, except that screening was performed by observing MGFP fluorescence in the larval body. The resulting transgenic silkworm (IM1) was crossed with silkworms carrying the $\alpha 1(I)$ chain to obtain silkworms hemizygous for both the IE1 and the $\alpha 1(I)$ chain genes.

To generate silkworms homozygous for the IE1 and the $\alpha 1(I)$ chain genes, the above hemizygous silkworms were crossed with one another. From the next generations, homozygous silkworms for both genes were screened by genomic PCR using DNA extracted from blood cells as a template. Primers used in this PCR were designed from the genomic sequences flanking the transgene insertions, which were determined with an APA Transgene Locator Kit (BIO S&T Inc., Montreal, QC, Canada) according to the manufacturer's instructions.

SDS-PAGE and Western Blotting

Proteins in the sericin layer of silk fibers were extracted by incubating cocoons at 80°C for 5 min in a buffer consisting of 8 M urea, 2% 2-mercaptoethanol, and 50 mM Tris-HCl, pH 8.0. After centrifugation at 23,500g for 5 min, the obtained supernatant was electrophoresed on 0.1% SDS/5–20% polyacrylamide gradient gels (Atto, Tokyo, Japan) in running buffer (0.1% SDS, 12.5 mM Tris, and 125 mM glycine). The gels were stained with CBB-R250, and densitometric analyses were performed using the image-processing program, ImageJ (<http://rsb.info.nih.gov/ij/>). For Western blotting, the proteins on the gels were transferred onto PVDF membranes (Immobilon-P; Millipore, Billerica, MA), reacted with antihuman/bovine type I collagen antibodies (LB-1190; Cosmo Bio, Tokyo, Japan) and then with horseradish peroxidase (HRP)-linked anti-(rabbit IgG) antibodies (Cell Signaling Technology, Danvers, MA). The antigen-antibody complexes were visualized using the ECL Western Blotting Detection System (GE Healthcare, Little Chalfont, Buckinghamshire, UK).

Purification of the Recombinant $\alpha 1(I)$ Chain From Transgenic Silkworm Cocoons

Cocoons were crushed into powder using a mill, suspended in a solution of 8 M urea and 0.5 M CH_3COOH , pH 2.0, at a concentration of 10 mg powdered cocoons/mL, and incubated at 4°C for 2 days with stirring. The resulting solution was filtered through filter paper and 70- μm nylon mesh. The solution was concentrated approximately threefold with an ultrafiltration membrane (Millipore) and urea was removed from the solution by adding 0.5 M CH_3COOH , pH 2.0, and subsequent ultrafiltration. NaCl was then added to the solution at a concentration of 0.5 M, and precipitated sericin was removed by centrifugation. Next, the recombinant $\alpha 1(I)$ chain in the solution was collected by the addition of 3.5 M NaCl. The precipitate was

solubilized in 0.5 M CH₃COOH and subjected to dialysis against water.

N-Terminal Sequencing and Measurement of Amino Acid Composition

The N-terminal sequences of the purified $\alpha 1(I)$ chain were determined with a G1000A protein sequencer (Hewlett Packard, Palo Alto, CA) using the standard protocol of Edman degradation. For the degradation reaction, 36 pmol of the chain were loaded onto the sequencer and the reaction was carried out for five cycles.

The purified $\alpha 1(I)$ chain was hydrolyzed in 6 N HCl for 22 h at 110°C under vacuum, and analyzed for amino acid composition using a Hitachi 835 amino acid analyzer (Hitachi, Tokyo, Japan).

Determination of Melting and Gelation Points

The purified $\alpha 1(I)$ chain was dissolved in distilled water at a concentration of 50 mg/mL, and the solution was subjected to determination of gelation and melting points as follows. The $\alpha 1(I)$ chain solution was gradually cooled from 35 to -5°C at a rate of 1°C/min with a thermal cycler (Atto). The gelation point was determined by reading the temperature of the sample when its fluidity disappeared. For the analysis of the melting point, the $\alpha 1(I)$ chain solution was incubated in icy water for 30 min to create a gel. The $\alpha 1(I)$ chain gel was gradually heated from 0 to 40°C at a rate of 1°C/min with the thermal cycler. The melting point was determined by reading the temperature of the sample when the bubble at the bottom of the tube reached the surface of the sample solution.

Measurement of Circular Dichroism (CD) Spectra

CD spectra were recorded for the recombinant $\alpha 1(I)$ chain using a spectropolarimeter J-820 (Jasco, Tokyo, Japan). The purified $\alpha 1(I)$ chain was dissolved in 50 mM CH₃COOH at a concentration of 100 μ g/mL for measurement in far ultraviolet (190–240 nm) regions, and the solution was placed in a cuvette with 2-mm path length. The temperature was kept at 4°C during the procedures. Measurements were performed three times and the averaged values were plotted. Thermal transition curves were recorded by measuring molar ellipticity at 224 nm between 5 and 60°C at a rate of 30°C/h. The denaturing temperature was calculated with J-820 software (Jasco).

Endotoxin Test

An endotoxin level of the recombinant $\alpha 1(I)$ chain was quantified using the limulus amoebocyte lysate (LAL) as per the manufacturer's instructions (Endospecy ES-50M and Toxicolor DIA-MP; Seikagaku Biobusiness, Tokyo, Japan).

Briefly, 50 μ L standards or samples diluted with endotoxin-free water were mixed with LAL and incubated at 37°C for 30 min. After the substrate solution was added, the absorbance at 545 nm was measured. A standard curve was constructed by using the standards in the range of 0.02–0.1 EU/mL, and the concentration of endotoxin in each sample was determined by plotting the absorbance to the standard curve.

Quantifying the Spread of Human Dermal Skin Fibroblasts

The cell adhesivity to the recombinant $\alpha 1(I)$ chain was analyzed by a quantitative cell-spreading assay using human dermal skin fibroblasts (HSFs; Kurabo, Osaka, Japan) as described previously (Yamada and Kennedy, 1984). In brief, a 96-well tissue culture plate (Becton Dickinson Labware, Franklin Lakes, NJ) was incubated with the recombinant $\alpha 1(I)$ chain at various concentrations (0.31, 0.62, 1.25, 2.5, and 5.0 μ g/mL) at 37°C for 1 h, treated with heat-denatured 0.5% bovine serum albumin to block the direct interaction between cells and the plate surface, and extensively washed with PBS. HSFs were seeded on the wells at a density of 1×10^4 cells per well, cultured in serum-free Dulbecco's modified Eagle's medium for 1 h, fixed in 4% paraformaldehyde, and viewed through a phase-contrast microscope (Nikon, Tokyo, Japan) to calculate the ratio of fully spreading cells to all cells observed in five randomly selected fields. HSFs were also cultured on the 96-well tissue culture plate coated with 10 μ g/mL of the $\alpha 1(I)$ chain as above to observe the cell morphology.

Culture of Cynomolgus Monkey Embryonic Stem (ES) Cells on Dishes Coated With the Recombinant $\alpha 1(I)$ Chain

Cynomolgus monkey ES cells were cultured on murine embryonic fibroblast (MEF) feeder cells in Petri dishes (90 mm in diameter) according to previously established protocols (Cameron et al., 2006). Briefly, dishes were coated with the recombinant $\alpha 1(I)$ chain by incubation with a solution of 1 mg/mL $\alpha 1(I)$ chain for 30 min at room temperature. γ -Irradiated MEF cells were then cultured on the coated dishes as feeder cells for 1 day. Subsequently, monkey ES cells were seeded on MEF feeder layers and maintained by changing the medium. Immunostaining of monkey ES cell colonies was performed using NANOG, TRA1-81, SSEA-4, SOX2, and OCT4-specific primary antibodies (Millipore) according to the manufacturer's protocol. Nuclei were visualized by DAPI staining (Lin et al., 1976).

Monkey ES cells were injected subcutaneously into the hind leg of severe combined immunodeficient (SCID) mice (Suemori et al., 2001). Two months after injection, the mice were killed to remove teratomas. The teratomas

were then fixed with 4% paraformaldehyde in PBS, paraffin-embedded, sectioned, and histologically analyzed following staining with hematoxylin and eosin.

Results

Generation of Transgenic Silkworms

COL1A1/pMSG was injected into 9,834 pre-blastoderm embryos, and the hatched 5,282 G_0 larvae were allowed to develop to moths. G_1 embryos from the G_0 moths were screened for MGFP expression in the eyes and nervous tissues to obtain transgenic silkworms. Genomic Southern blot analysis of the transgenic silkworms demonstrated the existence of 41 independent transgenic lines with respect to the chromosomal insertion positions and copy numbers of the transgenes. Among them, 34 lines of transgenic silkworms with a single-copy transgene were selected, and the cocoon proteins of the lines were analyzed by SDS-PAGE. The line with the highest level of $\alpha 1(I)$ chain expression was crossed with wild-type silkworms, and offspring hemizygous for the $\alpha 1(I)$ chain gene were used in the following experiments as the COL249 line.

The worms of the COL249 line spun cocoons that were normal in appearance, size, and weight. Proteins in the sericin layer of the silk fibers of COL249 and wild-type silkworms were separated by SDS-PAGE and stained with CBB (Fig. 2, lanes 1–2). A band with an apparent molecular weight of 120 kDa was seen only in the cocoon proteins of COL249 (Fig. 2, lane 1). This band reacted with antihuman/bovine type I collagen antibody (Fig. 2, lane 6), indicating that this was the recombinant product from the human $\alpha 1(I)$ chain gene. The band intensity of the recombinant $\alpha 1(I)$ chain on the CBB-stained gel was quantified by densitometry. The content of the $\alpha 1(I)$ chain was estimated to be 0.8% of the dried cocoon.

To increase the $\alpha 1(I)$ chain content in the cocoon, a COL249 moth was crossed with an IM1 moth bearing the gene of the baculovirus-derived trans-activator IE1. Approximately 25% of the offspring carried both the $\alpha 1(I)$ chain and IE1 genes hemizygously. The $\alpha 1(I)$ chain/IE1 hemizygous silkworms (COL249/IM1) were further crossed with each other and silkworms homozygous for both the $\alpha 1(I)$ chain and IE1 genes (COL249/IM)² were screened from the offspring by genomic PCR. Proteins in the cocoon extracts of COL249, COL249/IM, and (COL249/IM)² were separated by SDS-PAGE and stained with CBB (Fig. 3). By measuring the band intensity, the contents of the $\alpha 1(I)$ chain in cocoons of the silkworm lines COL249, COL249/IM, and (COL249/IM)² were estimated to be 0.8%, 4.8%, and 8.0%, respectively. The average weights of cocoons in the COL249, COL249/IM, and (COL249/IM)² lines were 72, 65, and 53 mg, respectively. Although the cocoon weight decreased slightly with increased transgene copy numbers, the synthesis of the recombinant $\alpha 1(I)$ chain per larvae was improved by this procedure; the amounts of the chain per

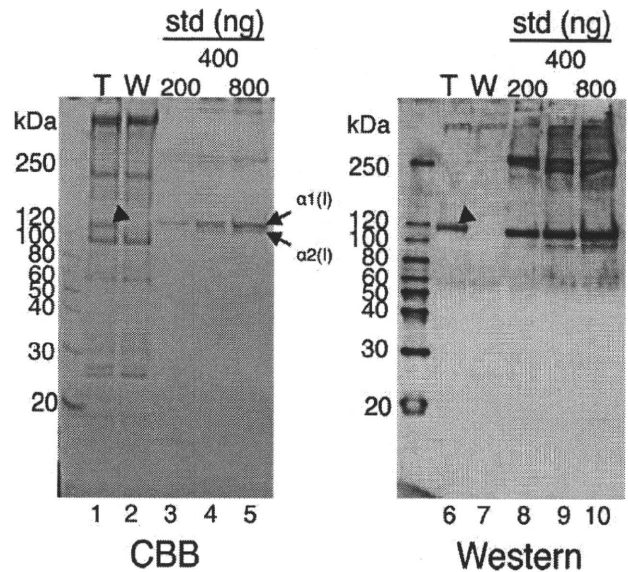


Figure 2. Expression of the $\alpha 1(I)$ chain in cocoons of transgenic silkworms. The proteins in the cocoons of COL249 (T) and wild-type (W) silkworms were extracted with 8 M urea, 2% 2-mercaptoethanol, and 50 mM Tris-HCl, pH 8.0, separated by SDS-PAGE, and stained with CBB (left panel). Aliquots of the cocoon extracts were also assessed by Western blotting with antihuman/bovine type I collagen antibodies (right panel). Bovine pepsinized type I collagen in the amounts indicated was analyzed by CBB staining and Western blotting as a standard (std). The arrowheads in lanes 1 and 6 point to the band of the recombinant $\alpha 1(I)$ chain. The arrows in lane 5 point to $\alpha 1(I)$ and $\alpha 2(I)$ chains of bovine type I collagen. Arabic numerals at the left side are molecular masses in kDa.

cocoon of the COL249, COL249/IM, and (COL249/IM)² lines were 0.58, 3.12, and 4.24 mg, respectively.

Extraction and Purification of the Recombinant $\alpha 1(I)$ Chain From Cocoons

The extraction efficiency of the recombinant $\alpha 1(I)$ chain from cocoons was examined. The powder of (COL249/IM)² cocoons was suspended in either PBS (Fig. 4A, lane 3), 0.5 M CH_3COOH , pH 3.0 (Fig. 4A, lane 4), 0.5 M CH_3COOH , pH 2.0 (Fig. 4A, lane 5), or 8 M urea and 0.5 M CH_3COOH , pH 2.0 (Fig. 4A, lane 6), at 4°C for 16 h, and the extracted proteins were analyzed by SDS-PAGE. Total proteins in the sericin layer of silk fibers of wild-type and transgenic silkworms were extracted by incubating cocoons at 80°C for 5 min in a buffer consisting of 8 M urea, 2% 2-mercaptoethanol, and 50 mM Tris-HCl, pH 8.0 (Fig. 4A, lanes 1 and 2). The ratios of the $\alpha 1(I)$ chain extracted with the solutions to the total $\alpha 1(I)$ chain in the sericin layer were calculated by quantifying the band intensities of the CBB-stained $\alpha 1(I)$ chain. The efficiency of extraction was estimated to be 10% for PBS, 30% for 0.5 M CH_3COOH , pH 3.0, 50% for 0.5 M CH_3COOH , pH 2.0, and 80% for 8 M urea and 0.5 M CH_3COOH , pH 2.0. Thus, the use of CH_3COOH -solutions was more effective than PBS for extracting the recombinant $\alpha 1(I)$ chain from

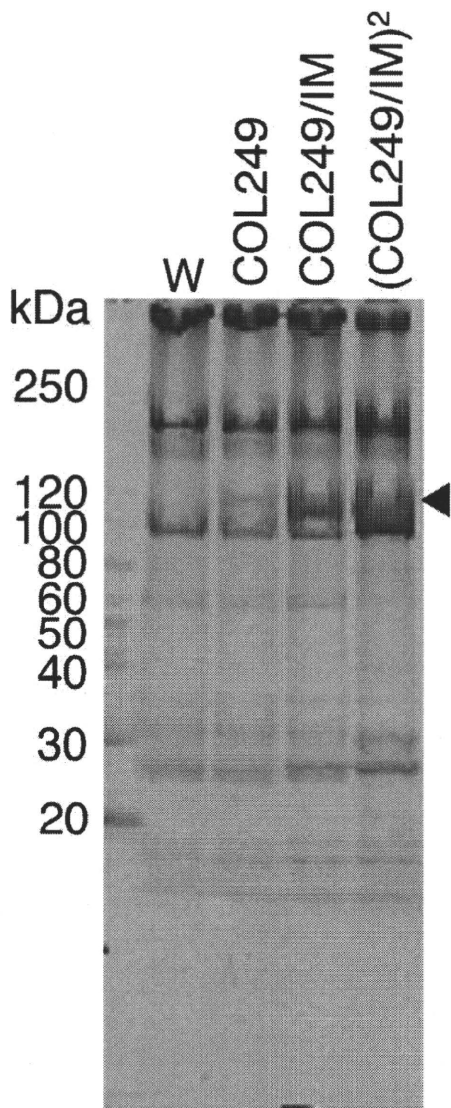


Figure 3. Increase in the $\alpha 1(I)$ chain content in cocoons of transgenic silkworms. Transgenic silkworms were genetically modified to increase the amount of the $\alpha 1(I)$ chain. Proteins were extracted from cocoons produced by wild-type (W), COL249, COL249/IM, and (COL249/IM)² silkworms, separated by SDS-PAGE, and stained with CBB. The arrowhead points to the band of the recombinant $\alpha 1(I)$ chain. Arabic numerals at the left side are molecular masses in kDa.

cocoons, and a major part of the $\alpha 1(I)$ chain was extractive with 8 M urea and 0.5 M CH₃COOH, pH 2.0.

For purification of the $\alpha 1(I)$ chain, 30 g of (COL249/IM)² cocoon powder, which was estimated to contain 2.4 g of the $\alpha 1(I)$ chain, were suspended in 8 M urea and 0.5 M CH₃COOH, pH 2.0. The extracted $\alpha 1(I)$ chain (Fig. 4B, lane 2) was concentrated by ultrafiltration (Fig. 4B, lane 3). The urea in the solution was removed by adding 0.5 M CH₃COOH, pH 2.0, and subsequent ultrafiltration. In this process, the endogenous sericin in the extract formed an insoluble aggregate, increasing the $\alpha 1(I)$ chain content in the extract (Fig. 4B, lane 4). Small amounts of contaminated

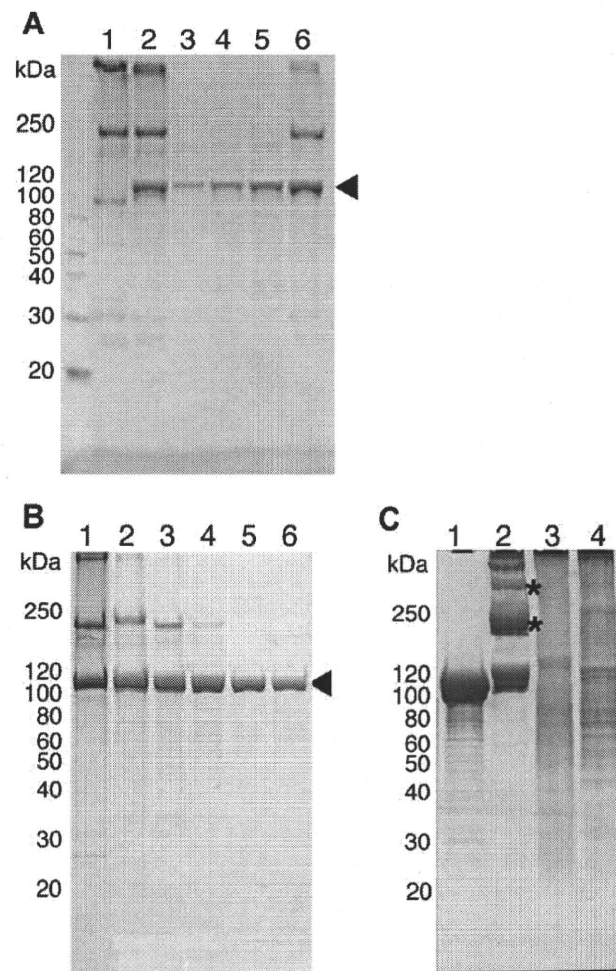


Figure 4. Extraction and purification of the $\alpha 1(I)$ chain from cocoons of (COL249/IM)² silkworms. **A:** Extraction of the $\alpha 1(I)$ chain from cocoons. Cocoon proteins were extracted with PBS (lane 1), 0.5 M CH₃COOH, pH 3.0 (lane 2), 0.5 M CH₃COOH, pH 2.0 (lane 3), and 8 M urea and 0.5 M CH₃COOH, pH 2.0 (lane 4). The extracted proteins were analyzed by SDS-PAGE. Lanes 1 and 2 represent total proteins in the sericin layer of the silk fibers of wild-type and transgenic silkworms, respectively. The arrowhead points to the $\alpha 1(I)$ chain. **B:** Analysis of the $\alpha 1(I)$ chain in the purification processes. Cocoon proteins were extracted with 8 M urea and 0.5 M CH₃COOH, pH 2.0 (lane 1). The extracted proteins were concentrated by ultrafiltration (lane 2), and urea in the extract was removed by adding 0.5 M CH₃COOH, pH 2.0, and subsequent ultrafiltration (lane 3). After removing sericin by 0.5 M NaCl precipitation (lane 4), the $\alpha 1(I)$ chain in the extract was collected by 4 M NaCl precipitation (lane 5). The proteins in each purification step were analyzed by SDS-PAGE and CBB staining. Lane 1 represents total proteins in the sericin layer of the silk fibers. The arrowhead points to the $\alpha 1(I)$ chain. **C:** Analysis of the purified $\alpha 1(I)$ chain. Aliquots of 15 μ g of proteins were electrophoresed and stained with CBB. Lane 1, the recombinant $\alpha 1(I)$ chain; lane 2, bovine type I collagen; lane 3, alkaline-treated bovine gelatin; lane 4, acid-treated porcine gelatin. The asterisks in lane 2 point to the dimer and trimer composed of the two and three α chains, respectively. Arabic numerals at the left side are molecular masses in kDa.

sericin were removed by 0.5 M NaCl precipitation (Fig. 4B, lane 5), and the $\alpha 1(I)$ chain in the supernatant was then collected by precipitation with 4 M NaCl. The collected $\alpha 1(I)$ chain was dissolved in 0.5 M CH₃COOH, pH 2.0, again, and the $\alpha 1(I)$ chain solution was finally dialyzed against water (Fig. 4B, lane 6). The proteins in each

purification step and total proteins in the sericin layer (Fig. 4B, lane 1) were analyzed by SDS-PAGE, demonstrating that this simple purification process is sufficient to purify the $\alpha 1(I)$ chain to apparent homogeneity. As a result, 990 mg of the $\alpha 1(I)$ chain were purified from 30 g of cocoons; the recovery rate was estimated to be approximately 41%.

Biochemical Characterization

The purified recombinant $\alpha 1(I)$ chain was analyzed by SDS-PAGE. Although small amounts of degradation products were found, the purified recombinant chain was composed of the polypeptide with a uniform molecular weight. The molecular weight of the chain was slightly smaller than the standard bovine $\alpha 1(I)$ chain (Fig. 4C, lanes 1 and 2), indicating the possibility of insufficient prolyl-hydroxylation in the recombinant chain. The dimer (β chain) and trimer (γ chain) of the α chain, which were present in the standard collagen, were not detected from the purified recombinant chain, suggesting the absence of covalent cross-linking among the $\alpha 1(I)$ chains. The molecular weight distribution of the recombinant α chain was quite different from that of the alkali-treated bovine (Fig. 4C, lane 3) or acid-treated porcine gelatins (Fig. 4C, lane 4). The gelatins gave broad molecular weight distributions because they were hydrolyzed products of collagens.

The $\alpha 1(I)$ chain was subjected to an amino acid sequencer with five cycles of Edman degradation. The N-terminal amino acid sequencing of the $\alpha 1(I)$ chain detected major and minor amino acid peaks in each cycle as shown in Table I. The sequence deduced from the minor peaks (GPM) was consistent with that of the predicted signal peptide cleavage (Fig. 1A) although peaks were not detected in the fourth and fifth cycles. The sequence from the major peaks (MGPSG) was probably derived from a cleavage at two amino acids downstream of the predicted site.

The amino acid composition of the purified $\alpha 1(I)$ chain was determined after acid hydrolysis using a Hitachi L835 automated analyzer (Table II). The determined values were very similar to the predicted ones, except for the absence of hydroxyprolines and hydroxylysines.

The endotoxin levels of the $\alpha 1(I)$ chain and the porcine gelatin were measured. The endotoxin level of the $\alpha 1(I)$ chain was much lower (26 EU/g) than the gelatin (6,400 EU/g).

Structural Characterization

Far-ultraviolet (190–240 nm) CD spectra were recorded for the recombinant $\alpha 1(I)$ chain, the native bovine type I

Table I. N-terminal sequencing.

Amino acid number	1	2	3	4	5
Major peaks	M	G	P	S	G
Minor peaks	G	P	M	—	—

Table II. Measurement of amino acid composition.

Amino Acid	Composition (mol%)	
	Recombinant $\alpha 1(I)$ chain	Human $\alpha 1(I)$ chain (predicted)
Aspartic acid	4.48	4.14
Threonine	1.56	1.58
Serine	2.91	3.35
Glutamic acid	7.71	7.30
Glycine	34.23	33.63
Alanine	11.72	11.74
Valine	1.81	1.87
Cystein	0.00	0.00
Methionine	0.33	0.69
Isoleucine	0.70	0.59
Leucine	2.12	1.87
Tyrosine	0.00	0.00
Phenylalanine	1.27	1.18
Hydroxylysine	0.00	2.37*
Lysine	3.66	1.18
Histidine	0.00	0.20
Arginine	4.71	5.03
Tryptophan	0.00	0.00
Hydroxyproline	0.00	11.44*
Proline	22.80	11.83
Total	100.00	100.00

*Assuming that all of lysine and proline residues in Y-position are hydroxylated.

collagen, the heat-denatured bovine type I collagen, and the porcine gelatin (Fig. 5A). The positive peak at 224 nm that is characteristic of the triple helical structure of collagen (Miller and Gay, 1982) was observed in the type I collagen. The gelatin exhibited a positive low peak at this wavelength, suggesting that the gelatin contained a partly formed triple helical structure in the molecule. In contrast, the recombinant $\alpha 1(I)$ chain did not show a positive peak at 224 nm. The peak of the type I collagen disappeared when it was heat-denatured. The spectra of the denatured collagen were almost identical to those of the recombinant $\alpha 1(I)$ chain. These results suggest that the $\alpha 1(I)$ chain contained no triple helical structure. A negative peak at 198 nm represents the triple helical structure (Miller and Gay, 1982). The peak intensity at this wavelength of the recombinant $\alpha 1(I)$ chain was similar to that of the heat-denatured collagen rather than that of gelatin, confirming the absence of the triple helical structure in the recombinant chain.

The 224-nm spectra were recorded for the recombinant chain, the native collagen, the denatured collagen and the gelatin at temperatures from 4 to 60°C (Fig. 5B). Apparent structural transition of the native collagen was observed in the range 39–46°C, which is in accordance with a report that the denatured temperature of bovine type I collagen is 42.8°C (Peltonen et al., 1980). In contrast, the recombinant $\alpha 1(I)$ chain, the denatured collagen and the gelatin showed slight structural changes in the range 25–45°C.

The melting and gelation points of the 5% $\alpha 1(I)$ chain or the gelatin solution were measured as described in the Materials and Methods Section. The melting and gelation

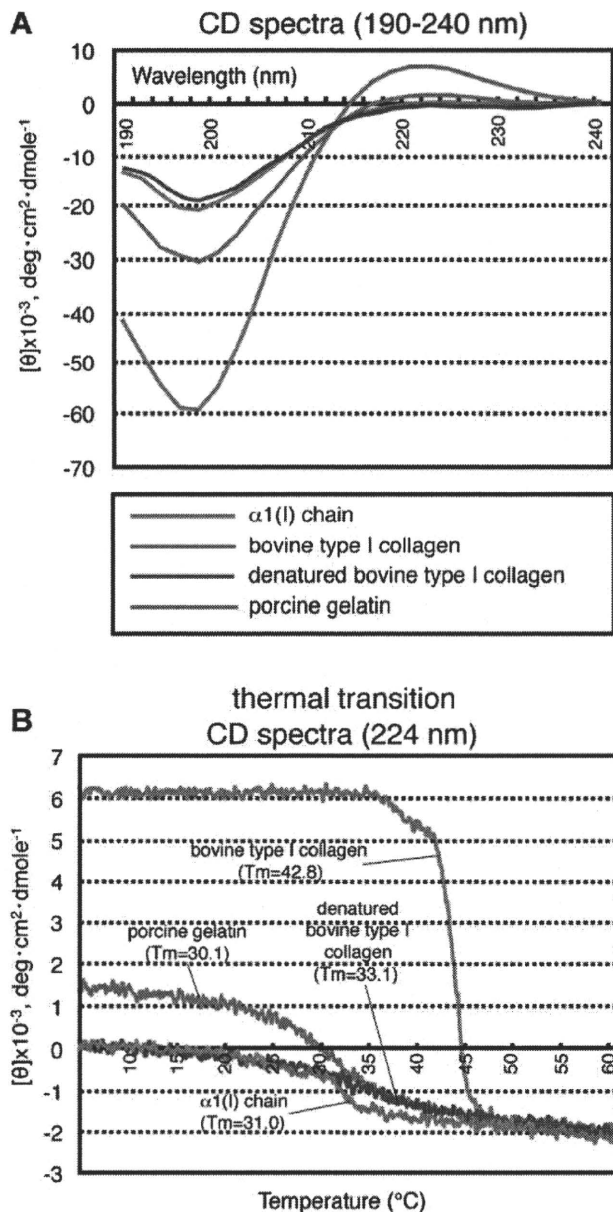


Figure 5. CD spectra of the $\alpha 1(I)$ chain. **A:** Measurement of CD spectra of the $\alpha 1(I)$ chain. Far-ultraviolet (190–240 nm) CD spectra were recorded for the recombinant $\alpha 1(I)$ chain (red line), bovine type I collagen (pink line), denatured bovine type I collagen (blue line), and porcine gelatin (green line) at a concentration of 100 $\mu\text{g}/\text{mL}$. **B:** Thermal transition curves of the $\alpha 1(I)$ chain. The CD spectra at 224 nm of the $\alpha 1(I)$ chain, bovine type I collagen, denatured bovine type I collagen and porcine gelatin were monitored at temperatures from 4 to 60 $^{\circ}\text{C}$.

points of the $\alpha 1(I)$ chain were 17 and 10 $^{\circ}\text{C}$, respectively, while the melting and gelation points of the gelatin were 30 and 26 $^{\circ}\text{C}$, respectively. Thus, both the melting and gelation points of the $\alpha 1(I)$ chain were lower than the respective points of the gelatin, which may support the result from the measurement of CD spectra showing an absence of triple helical structures in the $\alpha 1(I)$ chain. We also analyzed whether the $\alpha 1(I)$ chain formed collagen fibrils under the

physiological conditions as native collagen (Michalopoulos and Pitot, 1975) and found that the $\alpha 1(I)$ chain did not form the fibrils (data not shown).

Cell Biological Properties

To investigate the cell biological properties of the recombinant $\alpha 1(I)$ chain, HSFs were cultured on dishes coated with the $\alpha 1(I)$ chain, the native collagen, or gelatin at various concentrations, and cell spreading on the materials was analyzed as described in the Materials and Methods Section (Fig. 6A). HSFs spread on all coating materials in a concentration-dependent manner. On collagen at concentrations of more than 0.63 $\mu\text{g}/\text{mL}$, HSFs spread at a rate of 100%. More than 80% of the cells spread on gelatin at concentrations >0.63 $\mu\text{g}/\text{mL}$, but 100% cell spreading was never observed even at the highest concentration tested (5.0 $\mu\text{g}/\text{mL}$). Cell spreading rates for the $\alpha 1(I)$ chain at concentrations of <2.5 $\mu\text{g}/\text{mL}$ were slightly low compared to those on gelatin at the same concentrations. However, HSFs spread on the $\alpha 1(I)$ chain at a similar rate to on gelatin when inoculated at a concentration of 5.0 $\mu\text{g}/\text{mL}$. No differences were observed among cell morphologies when the cells were cultured on the $\alpha 1(I)$ chain, the native collagen, or the gelatin at a concentration of 10 $\mu\text{g}/\text{mL}$ (Fig. 6B, panels a–c). Cell-spreading was not observed on the uncoated dishes (Fig. 6B, panel d).

Cynomolgus monkey ES cells were cultured on feeder cells that had been cultured on dishes coated with the $\alpha 1(I)$ chain or porcine gelatin. The ES cells cultured on dishes coated with the $\alpha 1(I)$ chain formed tightly packed and flattened colonies (Fig. 7A, panel a). This morphology was the same as that of ES cell colonies cultured on dishes coated with porcine gelatin (Fig. 7A, panel b). Immunocytochemical studies confirmed that the monkey ES cell colonies on the $\alpha 1(I)$ chain expressed the ES cell marker proteins NANOG, TRA1-81, SSEA-4, SOX2, and OCT4 (Fig. 7B). When the ES cells were subcutaneously injected into SCID mice after the passages on the $\alpha 1(I)$ chain, the cells formed teratomas in the mouse tissues. Histological analyses of the teratomas showed formation of pigment epithelium, gastrointestinal epithelium, and cartilage (Fig. 7C). Thus, the $\alpha 1(I)$ chain was confirmed to be useful for the maintenance of monkey ES cells.

Discussion

We generated transgenic silkworms that secreted the recombinant human $\alpha 1(I)$ chain into the sericin layer of silk fibers. The content of the $\alpha 1(I)$ chain in the cocoons of the established line COL249 was estimated to be 0.8%. By introducing the gene of the trans-activator IE1 into the silkworm as in our previous studies (Ogawa et al., 2007; Tomita et al., 2007), the expression of the $\alpha 1(I)$ chain was enhanced to 4.8%. We then generated silkworms (COL249/

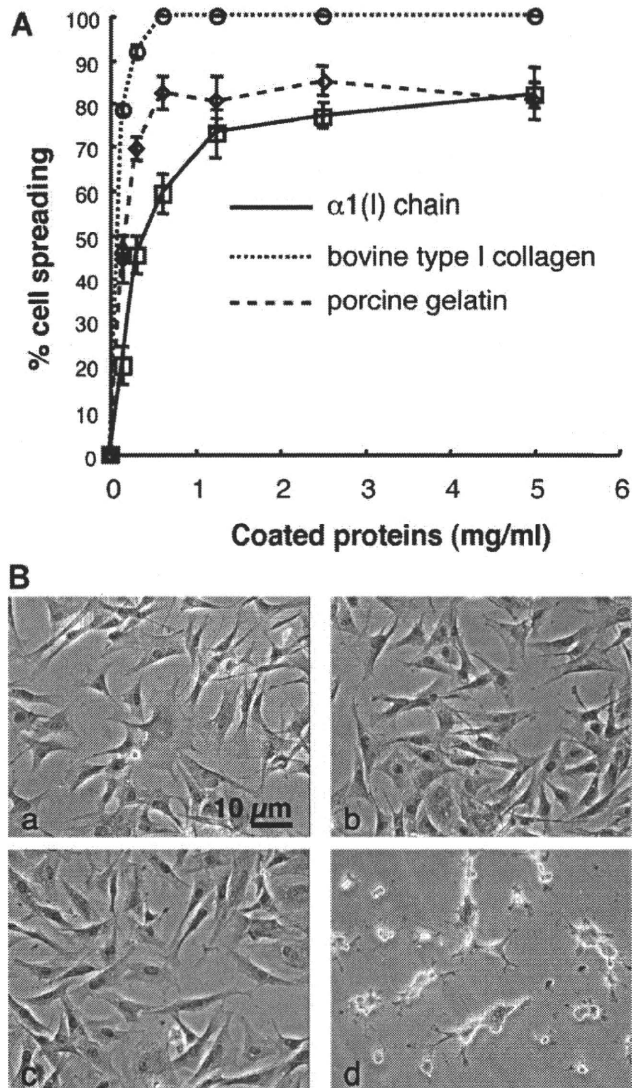


Figure 6. Spreading of HSFs on the $\alpha 1(I)$ chain-coated dishes. **A:** Cell spreading assay using HSFs. The wells of tissue culture plates were coated with the $\alpha 1(I)$ chain (black line), bovine type I collagen (dotted line), and porcine gelatin (dashed line) at various concentrations, and treated with heat-denatured bovine serum albumin to block the direct interaction between cells and the plate. HSFs were seeded on these wells and cultured for 1 h. The cells were then fixed, and the ratio of spreading cells to all cells in observed fields was calculated. **B:** Cell morphology of HSFs cultured on the $\alpha 1(I)$ chain. HSFs were cultured on dishes coated with 10 $\mu\text{g}/\text{mL}$ of the $\alpha 1(I)$ chain (a), bovine type I collagen (b), porcine gelatin (c). The cells were also cultured on the uncoated but the albumin-treated dish (d). Scale bar, 10 μm .

IM)² homozygous for both the $\alpha 1(I)$ chain and IE1 genes. This manipulation increased the $\alpha 1(I)$ chain content up to 8.0%. Given that the average weight of a (COL249/IM)² cocoon was 53 mg, the $\alpha 1(I)$ chain content per cocoon was calculated to be 4.24 mg. If 1,000 (COL249/IM)² silkworms were reared, we could produce 4.24 g of the $\alpha 1(I)$ chain. The (COL249/IM1)² silkworm was generated from the pnd-w1 strain, which produces small cocoons (50–70 mg). Our preliminary experiment revealed that the cocoon weight could be increased to approximately 150 mg by crossing it

with typical silkworm strains that produce 300- to 500-mg cocoons, leading to elevation of the $\alpha 1(I)$ content to more than 10 mg per cocoon (data not shown). We also demonstrated the superiority of the transgenic silkworm system for the purification of the recombinant $\alpha 1(I)$ chain. The $\alpha 1(I)$ chain was highly purified from the cocoon extract by a simple method consisting of ultrafiltration and salt precipitation. Thus, this study offers experimental evidence for the viability of using transgenic silkworms in the production of the human recombinant $\alpha 1(I)$ chain on an industrial scale.

Our previous study demonstrated that prolyl-hydroxylase activity is absent in silk glands (Adachi et al., 2005), and the recombinant fusion protein containing the collagen sequence expressed in the glands includes no hydroxyprolines (Tomita et al., 2005). The recombinant $\alpha 1(I)$ chain produced in this study also contained no hydroxyprolines as predicted. The presence of hydroxyprolines is a prerequisite for forming the stable collagen triple helix (Berg and Prockop, 1973). In addition, the $\alpha 1(I)$ chain did not contain the C-telopeptide and C-propeptide, which are known to promote triple helix formation (Doege and Fessler, 1986; Rosenbloom et al., 1976). Therefore, we postulated that the $\alpha 1(I)$ chain is not capable of forming the triple helix. Indeed, CD spectra of the chain showed a complete absence of the triple helical structure. The importance of the telopeptide and propeptide for the triple helix formation was also shown in the previous studies. Unhydroxylated type I collagen with the telopeptide, and unhydroxylated $\alpha 1(I)$ chain with the telopeptide and propeptide were synthesized as correctly folded triple helices in yeast (Olsen et al., 2001) and tobacco (Ruggiero et al., 2000), respectively. On the other hand, this study revealed that the animal-derived gelatin contained a partially folded triple helix, suggesting the significance of hydroxyprolines in the stability of triple helix. To further clarify the difference of physiological properties among the recombinant $\alpha 1(I)$ chain, the gelatin and the collagen, we investigated the gelation and fibril-forming properties of the $\alpha 1(I)$ chain. Unlike the collagen, the $\alpha 1(I)$ chain did not form collagen fibrils under the physiological conditions examined. In contrast, the $\alpha 1(I)$ chain, as well as the gelatin, gelled at lower temperatures than physiological ones. The melting and gelation points of the $\alpha 1(I)$ chain were lower than those of gelatin. Thus, the physico-chemical properties of the recombinant $\alpha 1(I)$ chain were similar to gelatin rather than collagen. However, due to the complete absence of the triple helical structure, the properties of the $\alpha 1(I)$ chain differed slightly from those of gelatin.

The $\alpha 1(I)$ chain promoted cell attachment and the spread of HSFs, but the cell-spreading rates for the $\alpha 1(I)$ chain as well as gelatin were lower than those for collagen at all concentrations tested. At decreased concentrations of the materials, fewer cells spread on the $\alpha 1(I)$ chain than on gelatin. Thus, HSFs were likely able to distinguish among these three materials. The cell–collagen interaction is mediated via integrins. Integrins $\alpha 1\beta 1$ and $\alpha 2\beta 1$ recognize collagens as collagen receptors (Hynes, 2002), and integrin

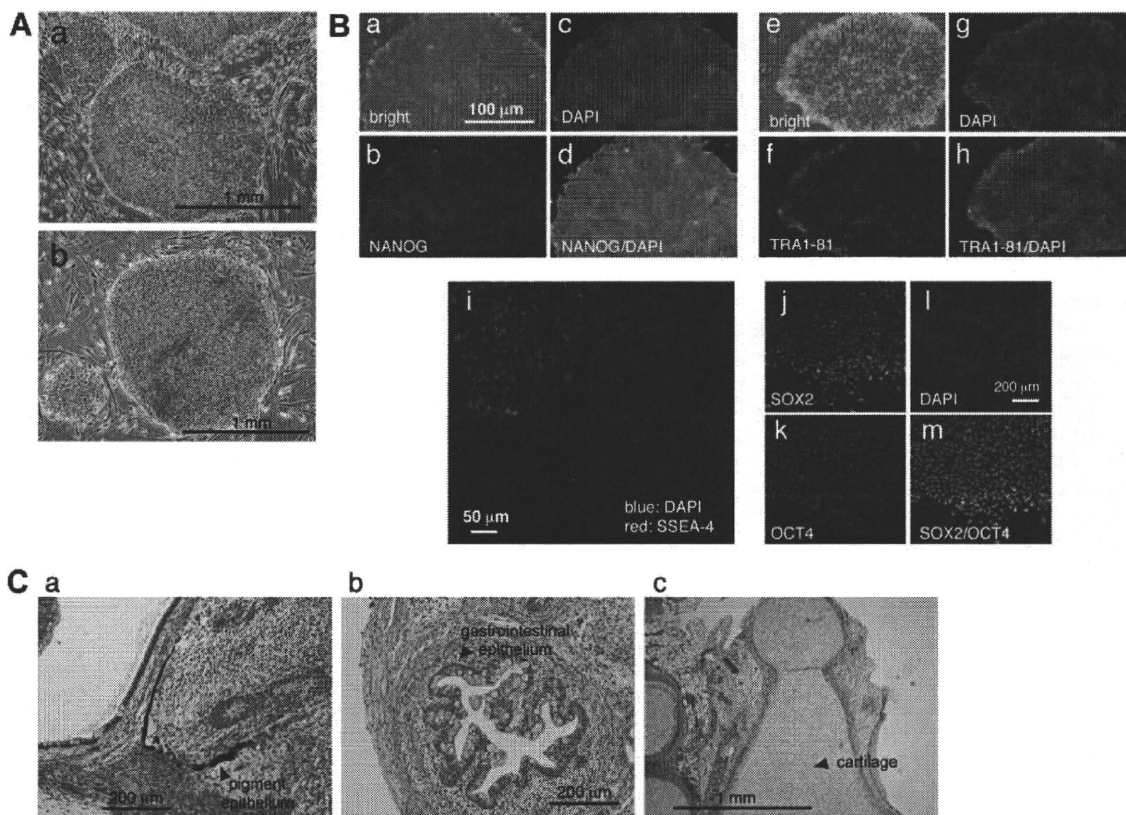


Figure 7. Culture of cynomolgus monkey ES cells on dishes coated with the $\alpha 1(I)$ chain. **A:** Appearance of monkey ES cells. Cynomolgus monkey ES cells were cultured on feeder cells that had been cultured on dishes coated with the $\alpha 1(I)$ chain (a) or porcine gelatin (b). Scale bar, 1 mm. **B:** Immunostain of the ES cell colonies. ES cell colonies were analyzed for the expression of marker proteins. **Panels (a–d and e–h)** show colonies immunostained with anti-NANOG and TRA1-81 antibodies, respectively. **Panels (a) and (e):** appearances in bright fields; **panels (b) and (f):** immunostained signals; **panels (c) and (g):** DAPI-stained signals; **panels (d) and (h):** mergers of immunostained and DAPI-stained signals. **Panel (i)** represents a colony stained with the anti-SSEA-4 antibody and DAPI. **Panels (j–m)** show a colony double-stained with anti-SOX2 and OCT4 antibodies. **Panel (j):** immunostained signal with anti-SOX2 antibody; **panel (k):** immunostained signal with anti-OCT4 antibody; **panel (l):** DAPI-stained signal; **panel (m):** merger of signals with anti-SOX2 and those with OCT4 antibodies. Scale bars, (a): 100 μm ; (i): 50 μm ; (l): 200 μm . **C:** Teratoma formation in SCID mice. The ES cells were injected subcutaneously into the hind leg of SCID mice. Sections of the teratomas that formed were stained with hematoxylin and eosin. Arrowheads in (a–c) point to the pigment epithelium, gastrointestinal epithelium, and cartilage, respectively. Scale bars, (a) and (b): 200 μm ; (c): 1 mm.

$\alpha 5\beta 1$ indirectly recognizes it via collagen-bound fibronectin (Mould et al., 1997). Although the detailed mechanism is unknown, the complete absence of hydroxyprolines or the triple helical structure in the $\alpha 1(I)$ chain may be responsible for this discrimination. At concentrations $>5.0 \mu\text{g}/\text{mL}$, however, HSFs spread on the $\alpha 1(I)$ chain at a similar rate to that of gelatin. At a concentration of $10 \mu\text{g}/\text{mL}$, the cell morphology was indistinguishable from that on collagen or gelatin. These results emphasize the practical utility of the recombinant $\alpha 1(I)$ chain as a cell scaffold.

To demonstrate the practicality of the $\alpha 1(I)$ chain, monkey ES cells were cultured on chain-coated dishes. After 30 passages, the monkey ES cell colonies maintained excellent morphology and the expression of several marker proteins for ES cells. The pluripotency of the cells was also confirmed by the formation of teratomas in SCID mice.

Gelatins are generally used for culturing ES or iPS cells. However, most marketed gelatins are derived from bovine or porcine bone, and therefore there is a risk of

contamination with animal-derived pathogens, including viruses. In contrast, the recombinant $\alpha 1(I)$ chain developed in this study does not pose such a risk because the chain is extracted from silk cocoons without using animal-derived materials. In addition, the $\alpha 1(I)$ chain is composed of human sequences with constant molecular weight. Unlike the animal-derived gelatin extracted by hydrolyzing tissue collagens, the quality of the chain can be easily controlled with lot-to-lot consistency. The endotoxin level of the $\alpha 1(I)$ chain was much lower than marketed gelatins. The recombinant $\alpha 1(I)$ chain is a promising candidate material for use as a high-quality gelatin substitute for tissue engineering, drug delivery, and other applications.

References

- Adachi T, Tomita M, Yoshizato K. 2005. Synthesis of prolyl 4-hydroxylase α subunit and type IV collagen in hemocytic granular cells of silkworm,

- Bombyx mori*: Involvement of type IV collagen in self-defense reaction and metamorphosis. *Matrix Biol* 24:136–154.
- Adachi T, Tomita M, Shimizu K, Ogawa S, Yoshizato K. 2006. Generation of hybrid transgenic silkworms that express *Bombyx mori* prolyl-hydroxylase alpha-subunits and human collagens in posterior silk glands: Production of cocoons that contained collagens with hydroxylated proline residues. *J Biotechnol* 126:205–219.
- Berg RA, Prockop DJ. 1973. The thermal transition of a non-hydroxylated form of collagen. Evidence for a role for hydroxyproline in stabilizing the triple-helix of collagen. *Biochem Biophys Res Commun* 52:115–120.
- Bradley R. 1993. The research programme on transmissible spongiform encephalopathies in Britain with special reference to bovine spongiform encephalopathy. *Dev Biol Stand* 80:157–170.
- Cameron CM, Hu WS, Kaufman DS. 2006. Improved development of human embryonic stem cell-derived embryoid bodies by stirred vessel cultivation. *Biotechnol Bioeng* 94:938–948.
- Doerge KJ, Fessler JH. 1986. Folding of carboxyl domain and assembly of procollagen I. *J Biol Chem* 261:8924–8935.
- Fichard A, Tillet E, Delacoux F, Garron R, Ruggiero F. 1997. Human recombinant alpha1(V) collagen chain. Homotrimeric assembly and subsequent processing. *J Biol Chem* 272:30083–30087.
- Garel A, Deleage G, Prudhomme JC. 1997. Structure and organization of the *Bombyx mori* sricin 1 gene and of the sericins 1 deduced from the sequence of the Ser 1B cDNA. *Insect Biochem Mol Biol* 27:469–477.
- Geddis AE, Prockop DJ. 1993. Expression of human COL1A1 gene in stably transfected HT1080 cells: The production of a thermostable homotrimer of type I collagen in a recombinant system. *Matrix* 13:399–405.
- Gill J, Feinberg J. 2001. Saquinavir soft gelatin capsule: A comparative safety review. *Drug Saf* 24:223–232.
- Grzelak K. 1995. Control of expression of silk protein genes. *Comp Biochem Physiol* 110:671–681.
- Hynes RO. 2002. Integrins: Bidirectional, allosteric signaling machines. *Cell* 110:673–687.
- Iizuka M, Tomita M, Shimizu K, Kikuchi Y, Yoshizato K. 2008. Translational enhancement of recombinant protein synthesis in transgenic silkworms by a 5'-untranslated region of polyhedrin gene of *Bombyx mori* Nucleopolyhedrovirus. *J Biosci Bioeng* 105:595–603.
- Iizuka M, Ogawa S, Takeuchi A, Nakakita S, Kubo Y, Miyawaki Y, Hirabayashi J, Tomita M. 2009. Production of a recombinant mouse monoclonal antibody in transgenic silkworm cocoons. *FEBS J* 276:5806–5820.
- John DC, Watson R, Kind AJ, Scott AR, Kadler KE, Bulleid NJ. 1999. Expression of an engineered form of recombinant procollagen in mouse milk. *Nat Biotechnol* 17:385–389.
- Lamberg A, Helaakoski T, Myllyharju J, Peltonen S, Notbohm H, Pihlajaniemi T, Kivirikko KI. 1996. Characterization of human type III collagen expressed in a baculovirus system. Production of a protein with a stable triple helix requires coexpression with the two types of recombinant prolyl 4-hydroxylase subunit. *J Biol Chem* 271:11988–11995.
- Lee CH, Singla A, Lee Y. 2001. Biomedical applications of collagen. *Int J Pharm* 221:1–22.
- Lin MS, Alfi OS, Donnell GN. 1976. Differential fluorescence of sister chromatids with 4'-6-diamidino-2-phenylindole. *Can J Genet Cytol* 18:545–547.
- Merle C, Perret S, Lacour T, Jonval V, Hudaverdian S, Garrone R, Ruggiero F, Theisen M. 2002. Hydroxylated human homotrimeric collagen I in *Agrobacterium tumefaciens*-mediated transient expression and in transgenic tobacco plant. *FEBS Lett* 515:114–118.
- Michalopoulos G, Pitot HC. 1975. Primary culture of parenchymal liver cells on collagen membranes. Morphological and biochemical observations. *Exp Cell Res* 94:70–78.
- Miller EJ, Gay S. 1982. The collagens: An overview and update. *Methods Enzymol* 82:3–32.
- Miyata T, Taira T, Noishiki Y. 1992. Collagen engineering for biomaterial use. *Clin Mater* 9:139–148.
- Mould AP, Askari JA, Aota S, Yamada KM, Irie A, Takada Y, Mardon HJ, Humphries MJ. 1997. Defining the topology of integrin alpha5 beta1-fibronectin interactions using inhibitory anti-alpha5 and anti-beta1 monoclonal antibodies. Evidence that the synergy sequence of fibronectin is recognized by the amino-terminal repeats of the alpha5 subunit. *J Biol Chem* 272:17283–17292.
- Mullins RJ, Richards C, Walker T. 1996. Allergic reactions to oral, surgical and topical bovine collagen. Anaphylactic risk for surgeons. *Aust NZ J Ophthalmol* 24:257–260.
- Nicholls AC, Pope FM, Schloon H. 1979. Biochemical heterogeneity of osteogenesis imperfecta: New variant. *Lancet* 1:1193.
- Ogawa S, Tomita M, Shimizu K, Yoshizato K. 2007. Generation of a transgenic silkworm that secretes recombinant proteins in the sericin layer of cocoon: Production of recombinant human serum albumin. *J Biotechnol* 128:531–544.
- Olsen DR, Leigh SD, Chang R, McMullin H, Ong W, Tai E, Chisholm G, Birk DE, Berg RA, Hitzeman RA, Toman PD. 2001. Production of human type I collagen in yeast reveals unexpected new insights into the molecular assembly of collagen trimers. *J Biol Chem* 276:24038–24043.
- Olsen D, Yang C, Bodo M, Chang R, Leigh S, Baez J, Carmichael D, Perala M, Hamalainen ER, Jarvinen M, Polarek J. 2003. Recombinant collagen and gelatin for drug delivery. *Adv Drug Del Rev* 55:1547–1567.
- Peltonen L, Palotie A, Hayashi T, Prockop DJ. 1980. Thermal stability of type I and type III procollagens from normal human fibroblasts and from a patient with osteogenesis imperfecta. *Proc Natl Acad Sci U S A* 77:162–166.
- Rosenbloom J, Endo R, Harsch M. 1976. Termination of procollagen chain synthesis by puromycin. Evidence that assembly and secretion require a COOH-terminal extension. *J Biol Chem* 251:2070–2076.
- Ruggiero F, Exposito JY, Bournat P, Gruber V, Perret S, Comte J, Olgner B, Garrone R, Theisen M. 2000. Triple helix assembly and processing of human collagen produced in transgenic tobacco plants. *FEBS Lett* 469:132–136.
- Suemori H, Tada T, Torii R, Hosoi Y, Kobayashi K, Imahie H, Kondo Y, Iritani A, Nakatsuji N. 2001. Establishment of embryonic stem cell lines from cynomolgus monkey blastocysts produced by IVF or ICSI. *Dev Dyn* 222:273–279.
- Tabata Y, Ikada Y. 1998. Protein release from gelatin matrices. *Adv Drug Deliv Rev* 31:287–301.
- Tamura T, Thibert C, Royer C, Kanda T, Abraham E, Kamba M, Komoto N, Thomas JL, Mauchamp B, Chavancy G, Shirk P, Fraser M, Prudhomme JC, Couble P. 2000. Germline transformation of the silkworm *Bombyx mori* L. using a *piggyBac* transposon-derived vector. *Nat Biotechnol* 18:81–84.
- Tomita M, Munetsuna T, Adachi T, Hino R, Hayashi M, Shimizu K, Nakamura N, Tamura T, Yoshizato K. 2003. Transgenic silkworms produce recombinant human type III procollagen in cocoons. *Nat Biotechnol* 21:52–56.
- Tomita M, Shimizu K, Yoshizato K. 2005. Transgenic silkworms that weave recombinant human collagen in cocoons. In: Yoshizato K, editor. *Transgenic silkworms: Eureka Bioscience Collection*. Georgetown: Landes Bioscience. chapter 2657.
- Tomita M, Hino R, Ogawa S, Iizuka M, Adachi T, Shimizu K, Sotoshiro H, Yoshizato K. 2007. A germline transgenic silkworm that secretes recombinant proteins in the sericin layer of cocoon. *Transgenic Res* 16:449–465.
- Vuorela A, Myllyharju J, Nissi R, Pihlajaniemi T, Kivirikko KI. 1997. Assembly of human prolyl 4-hydroxylase and type III collagen in the yeast *Pichia pastoris*: Formation of a stable enzyme tetramer requires coexpression with collagen and assembly of a stable collagen requires coexpression with prolyl 4-hydroxylase. *EMBO J* 16:6702–6712.
- Werten MW, van den Bosch TJ, Wind RD, Mooibroek H, de Wolf FA. 1999. High-yield secretion of recombinant gelatins by *Pichia pastoris*. *Yeast* 15:1087–1096.
- Yamada KM, Kennedy DW. 1984. Dualistic nature of adhesion protein function: Fibronectin and its biologically active peptide fragments can autoinhibit fibronectin function. *J Cell Biol* 99:29–36.

Defining Hypo-Methylated Regions of Stem Cell-Specific Promoters in Human iPS Cells Derived from Extra-Embryonic Amnions and Lung Fibroblasts

Koichiro Nishino¹, Masashi Toyoda¹, Mayu Yamazaki-Inoue¹, Hatsune Makino¹, Yoshihiro Fukawatase¹, Emi Chikazawa¹, Yoriko Takahashi¹, Yoshitaka Miyagawa², Hajime Okita², Nobutaka Kiyokawa², Hidenori Akutsu¹, Akihiro Umezawa^{1*}

¹ Department of Reproductive Biology, National Institute for Child Health and Development, Tokyo, Japan, ² Department of Developmental Biology, National Institute for Child Health and Development, Tokyo, Japan

Abstract

Background: Human induced pluripotent stem (iPS) cells are currently used as powerful resources in regenerative medicine. During very early developmental stages, DNA methylation decreases to an overall low level at the blastocyst stage, from which embryonic stem cells are derived. Therefore, pluripotent stem cells, such as ES and iPS cells, are considered to have hypo-methylated status compared to differentiated cells. However, epigenetic mechanisms of “stemness” remain unknown in iPS cells derived from extra-embryonic and embryonic cells.

Methodology/Principal Findings: We examined genome-wide DNA methylation (24,949 CpG sites covering 1,3862 genes, mostly selected from promoter regions) with six human iPS cell lines derived from human amniotic cells and fetal lung fibroblasts as well as two human ES cell lines, and eight human differentiated cell lines using Illumina’s Infinium HumanMethylation27. A considerable fraction (807 sites) exhibited a distinct difference in the methylation level between the iPS/ES cells and differentiated cells, with 87.6% hyper-methylation seen in iPS/ES cells. However, a limited fraction of CpG sites with hypo-methylation was found in promoters of genes encoding transcription factors. Thus, a group of genes becomes active through a decrease of methylation in their promoters. Twenty-three genes including *SOX15*, *SALL4*, *TGDF1*, *PPP1R16B* and *SOX10* as well as *POU5F1* were defined as genes with hypo-methylated SS-DMR (Stem cell-Specific Differentially Methylated Region) and highly expression in iPS/ES cells.

Conclusions/Significance: We show that DNA methylation profile of human amniotic iPS cells as well as fibroblast iPS cells, and defined the SS-DMRs. Knowledge of epigenetic information across iPS cells derived from different cell types can be used as a signature for “stemness” and may allow us to screen for optimum iPS/ES cells and to validate and monitor iPS/ES cell derivatives for human therapeutic applications.

Citation: Nishino K, Toyoda M, Yamazaki-Inoue M, Makino H, Fukawatase Y, et al. (2010) Defining Hypo-Methylated Regions of Stem Cell-Specific Promoters in Human iPS Cells Derived from Extra-Embryonic Amnions and Lung Fibroblasts. PLoS ONE 5(9): e13017. doi:10.1371/journal.pone.0013017

Editor: Tadafumi Kato, RIKEN Brain Science Institution, Japan

Received: April 21, 2010; **Accepted:** September 6, 2010; **Published:** September 27, 2010

Copyright: © 2010 Nishino et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This study was supported by grants from the Ministry of Education, Culture, Sports, Science, and Technology (MEXT) of Japan; Ministry of Health, Labour and Welfare Sciences (MHLW) research grants; by a Research Grant on Health Science focusing on Drug Innovation from the Japan Health Science Foundation; by the program for the promotion of Fundamental Studies in Health Science of the Pharmaceuticals and Medical Devices Agency; by a Research Grant for Cardiovascular Disease from the MHLW; and by a Grant for Child Health and Development from the MHLW. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: umezawa@1985.jukuin.keio.ac.jp

Introduction

Human embryonic stem (ES) cells [1] and induced pluripotent stem (iPS) cells [2,3,4,5] are currently used as powerful resources in regenerative medicine. However, epigenetic mechanisms of “stemness” remain unknown. DNA methylation is known to be a key component in normal differentiation and development [6,7]. Tissue-specific genes, such as *OCT-4/3* [8], *Sy* (sex determining region on Y chromosome) [9] and *MyoD* [10], show tissue-specific demethylation corresponding to their expression during development. Furthermore, DNA methylation in cells specifically varies depending on cell lineage and tissue types [7]. Transformation to iPS cells from differentiated cells requires a process of epigenetic

reprogramming [11]. Understanding the epigenetic regulation in human pluripotent stem cells, therefore, enable us to elucidate “stemness” and to screen for optimum iPS/ES cells for human therapeutic applications. Human extra-embryonic amnion cells are a useful cell source for generation of iPS cells, because they can be collected without invasion and are conventionally freeze-storable. Recently, we generated iPS cells from human amnion cells as well as human fetal lung fibroblast cells [12,13]. Here, we show DNA methylation profiles of human pluripotent stem cells including iPS cells, which were derived from extra-embryonic amnion cells and fetal lung fibroblast cells, and human ES cells. We also defined another subset that may play a key practical role in maintaining the state of “stemness”.