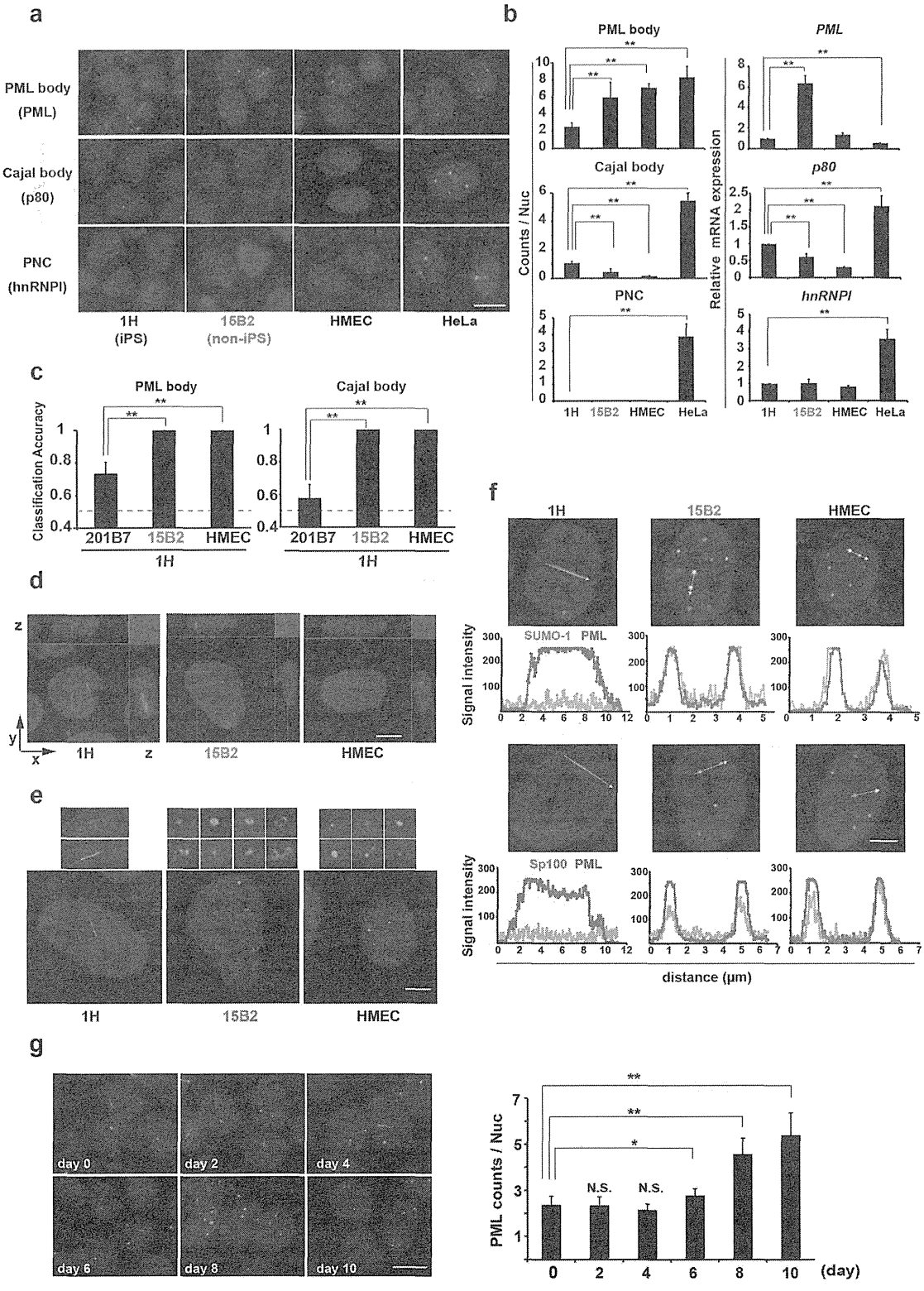

Figure 2 | Quantitative assessment of nuclear structures in completely and incompletely reprogrammed human iPSCs. (a) Identification of nuclear structure characteristics of iPSCs. Immunostaining was performed to identify the PML body (PML), Cajal body (p80 coilin), and perinucleolar compartment (PNC) (hnRNP I) (green). Nuclei were stained with DAPI (blue). (b) Quantification of nuclear structure formation (n>200, left) and the mRNA levels of the corresponding components in the structures (n=3, right). (c) wndchrm classifications against iPSCs (1H) using immunofluorescence images of PML and Cajal bodies (n=10). (d) Detection of linear PML structures by three-dimensional confocal microscopy. (e) Detection of PML structural variation by structured illumination microscopy (100 nm resolution). Enlarged images of PML structures are shown in the upper boxes. (f) Lack of SUMO-1 and Sp100 in the linear PML structures of bona fide iPSCs. The signal intensity along the arrow is shown below. PML, red; SUMO-1 and Sp100, green. (g) Transition of PML structures from linear to round during differentiation. The number of PML structures is shown at the right (n>300). Values are the means and s.d. *, $p<0.05$; **, $p<0.01$. Scale bars, 5 μm.

nucleus in 15B2 and 2B7 cells) and somatic cells (HMECs and IMR90 fibroblasts) was less frequent than that in iPSCs (1 ± 0.15 per nucleus in 1H, 201B7, and 253G1 cells), implying that hypoplasia of Cajal bodies is a feature of non-iPSCs. The PNC was only observed in HeLa cells (mean number = 4 ± 0.77), which is consistent with its specific appearance in cancerous cells[16] and high expression of one of its components, hnRNPI, in HeLa cells (Fig. 2a–b).

Using the wndchrm image libraries constructed from immunofluorescence of the PML and Cajal bodies (Supplementary Fig. S8), bona fide iPSCs (1H cells) were discriminated from non-iPSCs (15B2 cells) by extremely high CA values (~1.0) (Fig. 2c).

PML-defined structures in iPSCs were especially striking. Linear PML structures were found uniquely in bona fide iPSCs (Fig. 2a and Supplementary Fig. S7a). Three-dimensional imaging by confocal microscopy revealed approximately straight, rod-like PML structures traversing within the nuclei of iPSCs (Fig. 2d). In addition, more detailed structures visualized by structured illumination microscopy showed at least three classes of PML structures: linear and connected bead-like in iPSCs (1H), irregular ring-like in non-iPSCs (15B2), and normal spheres in HMECs (Fig. 2e). In terms of protein composition, the linear PML structure in bona fide iPSCs was distinct from that in somatic cells (Fig. 2f). In somatic cells such as HMECs, the PML protein and its SUMO modification are required for PML body formation and colocalisation with other components such as Sp100[26]. In contrast, the linear PML structure in iPSCs evidently lacked enrichment of SUMO-1 and Sp100. Finally, we found that the PML-defined structure in iPSCs transited to a somatic sphere PML body under differentiation conditions (day 6–10) in parallel with an increased number of the bodies (Fig. 2g). The resulting PML bodies coexisted with SUMO-1 and Sp100 on day 10 (Supplementary Fig. S9). Thus, the PML structure is dynamically regulated in iPSCs during their differentiation, indicating that the linear form of the PML body is one of the hallmarks of fully reprogrammed iPSCs.

In summary, we report the morphometric characteristics of human iPSCs by quantitative assessment of colony and nuclear structures.

Discussion

In the present study, we developed a new non-invasive method to distinguish nascent reprogrammed iPSC and non-iPSC colonies based on their morphologies. Previously, mouse reprogramming studies have used reporters integrated into the genomic loci of pluripotency genes Fbx15, Oct4, or Nanog[28–30] to identify reprogrammed cells. However, there have been no methods to reliably identify human iPSCs in a population of fibroblasts and imperfectly reprogrammed cells without cell labelling[4]. Our analysis using a collection of cell lines including standard iPSC lines (201B7 and 253G1)[2,3], newly generated iPSC lines (1H–4H), non-iPSC lines (15B2 and 2B7), and somatic cells (human mammary epithelial cells, HMECs) demonstrated that wndchrm analysis is effective and objective for discrimination of iPSC and non-iPSC colonies.

Quantitative measurement of morphological differences can be very complex, and it is sometimes difficult to analyse only predefined features[6,7,9]. Therefore, we used a supervised machine learning system, wndchrm, which has been developed to automatically mine for morphological similarities in a wide variety of objects[7,9]. As the first step, wndchrm computes a large number of features of both raw images and ones processed for transformations including Fourier, Wavelet and Chebyshev. Total calculated feature numbers are 2873 at maximum, which are expected to cover general image features. They include polynominal decompositions, high contrast features, pixel statistics and textures, and so on[7,19]. Next, wndchrm selects an informative feature by rejecting noisy feature by using Fisher Linear Discriminant algorithm. The Fischer scores are also used as feature weights, so that less discriminating features have a reduced effect on the classifier. The classifier used is WND5 which is related to the k-nearest neighbors type, except that it uses a negative exponential in a weighted feature space rather than a simple linear distance equation. While the classification system is composed of several linear elements, it is not justified to characterize this system as somehow limited in the complexity of the feature relationships it can exploit. In fact, wndchrm outperforms over a dozen state of the art classifiers which are constructed as either general, problem specific, linear or non-linear programming[7,31]. Because the method used here is versatile and not limited to any particular type of cell or image, it would be applicable to classify cells in various states, providing that a sufficient numbers of images and their retrospective metadata is available.

In this study, we found image features that discriminate between iPSCs and non-iPSC (Fig. 1e) within colony components. It is interesting that the value of morphological similarity does not depend on the original cell types or reprogramming procedures. For example, 201B7, 253G1, and 15B2 cells were all derived from human fibroblasts, However, 201B7 cells were closely clustered with 1H–4H cells that were derived from HMECs, suggesting that the cells have lost their lineage identities with morphological features during reprogramming. It is also interesting that the morphologies of the two non-iPSC lines (15B2 and 2B7) were not only distinct from iPSCs, but also from each other, implying that they have failed reprogramming at different stages or for different reasons (Supplementary Fig. S4d).

Prior to the colony classifications, we analyzed the cell lines for their undifferentiated states and differentiation abilities to find that they formed relatively homogeneous colonies (Supplementary Figs S1 and S2). However, reprogrammed cells generally could form a colony that is composed of different types of cells. A question of how the classification technology works on the heterogeneous colonies remains to be investigated. It is of interest that tiling an image was effective to determine the source of the classification signal in cultured colonies (Fig. 1e) and cells[10]. It was also successfully applied to determine the location of the predictive osteoarthritis signal in X-Rays[32].

We found that nuclear structures changed during reprogramming dynamically and specifically. Using immunofluorescence images of the PML and Cajal bodies, bona fide iPSCs (1H cells) were discriminated from non-iPSCs (15B2 cells) by extremely high accuracy, almost 100% accuracy, with wndchrm (Fig. 2c). This result suggests that one can use the information on the different nuclear bodies to discriminate between cell types effectively.

Among the nuclear structures, the linear PML-defined structure is an attractive indicator that represents the reprogramming state of iPSCs. The linear form of PML bodies has been observed in human embryonic stem (ES) cells[33], but not in mouse ES cells, possibly because human ES cells correspond to mouse-derived epiblast stem cells[34,35]. The ring-shaped PML body found in human non-iPSCs may represent a transition state from somatic cells to iPSCs. Collectively, the linear PML structure is likely to be characteristic of the pluripotent state of human cells.

In conclusion, the present study indicates that the reprogramming states of live human iPSCs can be evaluated precisely by machine learning technologies for image analyses. This versatile method is applicable to other cell types, and may be valuable for quality control of cells intended for regenerative medicine, as well as for basic research. Our findings will also significantly advance our knowledge of the nuclear landscape of iPSCs.

Methods

Generation and maintenance of iPSC lines. To generate human iPSC lines, four reprogramming factors (Oct3/4, Sox2, Klf4, and c-Myc) were introduced into HMECs and primary human fibroblasts using SeV vectors according to the manufacturer's protocol (Cyto tune-iPS DNAVEC). Briefly, 2×10^5 cells were plated and infected with SeV vectors, and then cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% foetal bovine serum. After ES cell-like

colonies appeared, the medium was changed to primate ES cell medium supplemented with 5 ng/ml basic fibroblast growth factor (bFGF) (ReproCELL). Growing colonies were picked up mechanically, expanded, and maintained on mouse embryonic fibroblasts (MEFs) to avoid spontaneous differentiation. Standard human iPSC lines, 201B7 and 253G1, were provided by Kyoto University and RIKEN BioResource Center, Japan, respectively. These cell lines were cultured on MEFs in Repro Stem medium supplemented with 5 ng/ml bFGF and penicillin/streptomycin.

***In vitro* differentiation of human iPSCs.** For embryoid body (EB) formation, iPSCs were treated with a dissociation solution (CTK, ReproCELL), and clumps of cells were cultured in DMEM/F12 containing 20% knockout serum replacement (Invitrogen) supplemented with 9.2 mM L-glutamine, 1×10^{-4} M non-essential amino acids, 1×10^{-4} M 2-mercaptoethanol, and penicillin/streptomycin. The medium was changed every other day. After 8 days of floating culture, EBs were transferred to a gelatin-coated plate and cultured in the same medium for another 8 days.

Immunofluorescence analysis. Cells were fixed with 4% paraformaldehyde in PBS for 10 min at room temperature, washed, and permeabilized with PBS containing 0.5% Triton X-100 for 5 min on ice. The cells were then incubated with primary antibodies for 1 h, followed by secondary antibodies for 1 h. Images were obtained under a microscope (IX-71; Olympus) equipped with a 60 × NA1.0 Plan Apo objective lens and a cooled charged-coupled device camera (Hamamatsu). Alternatively, images were captured with a confocal laser-scanning microscope (LSM 710, Carl Zeiss) with a 63 ×/1.4 Plan-Apochromat objective lens and a cooled charged-coupled device camera (Carl Zeiss). For immunofluorescences of the nuclear structures (PML body, Cajal body and PNC), image stacks containing three-dimensional datasets were collected at 1.0 μm intervals through the z axis, and projected onto two dimensions using imaging software (Lumina Vision; Mitani Corp). For structured illumination microscopic analyses, we used a microscope (Ti-E; Nikon) equipped with a 100 × NA1.49 CFI Apo TIRF objective lens, an electron multiplying charged-coupled device camera (iXon Em-CCD, Andor), and image acquisition software (Nikon).

Image quantification. For image classifications, we used wndchrm ver. 1.31[6,7]. Images used were; phase contrast of colonies (1024 × 767 pixels) (Supplementary Fig. S3), and immunofluorescences of nuclear structures (1280 × 1024 pixels) (Supplementary Fig. S8). No further segmentation was done prior to wndchrm analysis. The numbers of training/test images were 60/6 (Fig. 1), 40/8 (PML body in Fig. 2), 35/7 (Cajal body in Fig. 2), and 26/7 (Supplementary Fig. S5). The options used were a larger feature set of 2873 (-l), tiling of an image into 4 (-t4) for Fig. 1b–1d, and (-l, -t8) for Fig. 1e. Fisher scores were automatically computed for each feature in the following groups: ChF, Chebyshev-Fourier Statistics; Ch, Chebyshev Statistics; Com, Combined First Fourier Moments; Fra, Edge and Fractal Statistics; Har, Haralick Texture; MsH, Multiscale Histogram; Zer, Zernike Moments (Fig. 1c and Supplementary Fig. S4c)[7]. Pairwise class similarity values in Supplementary Table S2 were computed from the average of the marginal probabilities of all of the test images in each class. The per-class marginal probabilities were used as coordinates in a marginal probability space, where pairwise inter-class distances were computed using the Euclidean distance formula. Morphological similarity is the inverse of morphological distance between classes. Phylogenies were computed using the Fitch-Margoliash method implemented in the PHYLIP package, which is based on pairwise class similarity values reported by wndchrm ver 1.3[9,36]. For imaging cytometry analyses, CELAVIEW RS100 (Olympus) was used to capture images, and quantitative analyses was done using CELAVIEW analysis software as following. For automated quantification of PML body, Cajal body and PNC, each fluorescent image was segmented in CELAVIEW (Olympus) using the DAPI channel to define the nuclei. The nuclear bodies were automatically segmented based on their intensity and area using CELAVIEW as previously reported[37].

RNA isolation and quantitative PCR analysis. Total RNA was isolated using TRIzol (Invitrogen). For cDNA synthesis, 1 μg of total RNA was reverse transcribed with a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Quantitative PCR of target cDNAs was performed using Power SYBR Green PCR Master Mix (Applied Biosystems). Each experiment was performed at least three times. Relative fold changes were quantified by normalisation to *β-actin* expression. Primer sequences are listed in Supplementary Table S3.

Antibodies. The primary antibodies used were: mouse PML (1 : 500, SC-966, Santa Cruz), p80 coilin (1 : 300, #612074, BD Biosciences), hnRNPI (1 : 300, sc-16547, Santa Cruz), lamin A/C (1 : 300, sc-7292, Santa Cruz), Sp1 (1 : 300, sc-16547, Santa Cruz), and lamin B (1 : 300, sc-6216, Santa Cruz). The following secondary antibodies were used: Alexa 488-conjugated donkey anti-mouse IgG (1 : 250), Alexa 488-conjugated donkey anti-rabbit IgG (1 : 250) (Molecular Probes), Cy3-conjugated donkey anti-mouse IgG (1 : 1000), Cy3-conjugated donkey anti-rat IgG (1 : 1000), and Cy3-conjugated donkey anti-rabbit IgG (1 : 1000) (Jackson ImmunoResearch).

1. Yamanaka, S. & Blau, H. M. Nuclear reprogramming to a pluripotent state by three approaches. *Nature* **465**, 704–712 (2010).
2. Takahashi, K. *et al.* Induction of pluripotent stem cells from adult human fibroblasts by defined factors. *Cell* **131**, 861–872 (2007).
3. Nakagawa, M. *et al.* Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nat Biotechnol* **26**, 101–106 (2008).
4. Chan, E. M. *et al.* Live cell imaging distinguishes bona fide human iPS cells from partially reprogrammed cells. *Nat Biotechnol* **27**, 1033–1037 (2009).
5. Meissner, A. Epigenetic modifications in pluripotent and differentiated cells. *Nat Biotechnol* **28**, 1079–1088 (2010).
6. Eliceiri, K. W. *et al.* Biological imaging software tools. *Nat Methods* **9**, 697–710 (2012).
7. Orlov, N. *et al.* WND-CHARM: Multi-purpose image classification using compound image transforms. *Pattern Recog Lett* **29**, 1684–1693 (2008).
8. Grisendi, S. *et al.* Role of nucleophosmin in embryonic development and tumorigenesis. *Nature* **437**, 147–153 (2005).
9. Johnston, J., Iser, W. B., Chow, D. K., Goldberg, I. G. & Wolkow, C. A. Quantitative image analysis reveals distinct structural transitions during aging in Caenorhabditis elegans tissues. *PLoS One* **3**, e2821 (2008).
10. Shamir, L., Orlov, N., Mark Eckley, D., Macura, T. J. & Goldberg, I. G. IICBU 2008: a proposed benchmark suite for biological image analysis. *Med Biol Eng Comput* **46**, 943–947 (2008).
11. Scaglioni, P. P. *et al.* Translation-dependent mechanisms lead to PML upregulation and mediate oncogenic K-RAS-induced cellular senescence. *EMBO Mol Med* **4**, 594–602 (2012).
12. Zhao, R., Bodnar, M. S. & Spector, D. L. Nuclear neighborhoods and gene expression. *Curr Opin Genet Dev* **19**, 172–179 (2009).
13. Bernardi, R. & Pandolfi, P. P. Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies. *Nat Rev Mol Cell Biol* **8**, 1006–1016 (2007).
14. Morris, G. E. The Cajal body. *Biochim Biophys Acta* **1783**, 2108–2115 (2008).
15. Ghetti, A., Pinol-Roma, S., Michael, W. M., Morandi, C. & Dreyfuss, G. hnRNP I, the polypyrimidine tract-binding protein: distinct nuclear localization and association with hnRNAs. *Nucleic Acids Res* **20**, 3671–3678 (1992).
16. Norton, J. T. & Huang, S. The perinucleolar compartment: RNA metabolism and cancer. *Cancer Treat Res* **158**, 139–152 (2013).
17. Fussner, E. *et al.* Constitutive heterochromatin reorganization during somatic cell reprogramming. *The EMBO Jl* **30**, 1778–1789 (2011).
18. Fusaki, N., Ban, H., Nishiyama, A., Saeki, K. & Hasegawa, M. Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome. *Proc Jpn Acad Ser B Phys Biol Sci* **85**, 348–362 (2009).
19. Shamir, L. *et al.* Wndchrm - an open source utility for biological image analysis. *Source Code Biol Medicine* **3**, 13 (2008).
20. Shamir, L., Delaney, J. D., Orlov, N., Eckley, D. M. & Goldberg, I. G. Pattern recognition software and techniques for biological image analysis. *PLoS Comp Biol* **6**, e1000974 (2010).
21. Gasser, S. M. Visualizing chromatin dynamics in interphase nuclei. *Science* **296**, 1412–1416 (2002).
22. Meshorer, E. & Misteli, T. Chromatin in pluripotent embryonic stem cells and differentiation. *Nature reviews. Molec Cell Biol* **7**, 540–546 (2006).
23. Aoto, T., Saitoh, N., Ichimura, T., Niwa, H. & Nakao, M. Nuclear and chromatin reorganization in the MHC-Oct3/4 locus at developmental phases of embryonic stem cell differentiation. *Dev Biol* **298**, 354–367 (2006).
24. Constantinescu, D., Gray, H. L., Sammak, P. J., Schatten, G. P. & Csoka, A. B. Lamin A/C expression is a marker of mouse and human embryonic stem cell differentiation. *Stem Cells* **24**, 177–185 (2006).
25. Safe, S. & Abdelrahim, M. Sp transcription factor family and its role in cancer. *Eur J Cancer* **41**, 2438–2448 (2005).
26. Zhong, S., Salomoni, P. & Pandolfi, P. P. The transcriptional role of PML and the nuclear body. *Nat Cell Biol* **2**, E85–90 (2000).
27. Mao, Y. S., Zhang, B. & Spector, D. L. Biogenesis and function of nuclear bodies. *Trends Genet* **27**, 295–306 (2011).
28. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**, 663–676 (2006).
29. Okita, K., Ichisaka, T. & Yamanaka, S. Generation of germline-competent induced pluripotent stem cells. *Nature* **448**, 313–317 (2007).
30. Wernig, M. *et al.* In vitro reprogramming of fibroblasts into a pluripotent ES-cell-like state. *Nature* **448**, 318–324 (2007).
31. Orlov, N. V., Eckley, D. M., Shamir, L. & Goldberg, I. G. Improving class separability using extended pixel planes: a comparative study. *Machine Vis Applic* **23**, 1047–1058 (2012).
32. Shamir, L. *et al.* Early detection of radiographic knee osteoarthritis using computer-aided analysis. *Osteoarth Cartilage/OARS, Osteoarth Res Society* **17**, 1307–1312 (2009).
33. Butler, J. T., Hall, L. L., Smith, K. P. & Lawrence, J. B. Changing nuclear landscape and unique PML structures during early epigenetic transitions of human embryonic stem cells. *J Cell Biochem* **107**, 609–621 (2009).
34. Brons, I. G. *et al.* Derivation of pluripotent epiblast stem cells from mammalian embryos. *Nature* **448**, 191–195 (2007).
35. Tesar, P. J. *et al.* New cell lines from mouse epiblast share defining features with human embryonic stem cells. *Nature* **448**, 196–199 (2007).
36. Felsenstein, J. PHYLIP—Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164–166 (1989).

37. Saitoh, N. *et al.* The distribution of phosphorylated SR proteins and alternative splicing are regulated by RANBP2. *Mol Biol Cell* **23**, 1115–1128 (2012).

Acknowledgments

We thank Dr. Tamiyo Kobayashi (Olympus, Japan) for technical assistance. This work was supported by grants from the Ministry of Education, Culture, Sports, Science and Technology of Japan and the Japan Science and Technology Agency (CREST) (M.N. and N.S.).

Author contributions

K.T., N.S. and M.N. designed and performed the experiments, and prepared the manuscript, together with I.G.G. for the wndchrm analyses, Y.Y. and S.Y. for the images of iPSC colonies, and C.S. and Y.Y. for technical assistance.

Additional information

Supplementary information accompanies this paper at http://www.nature.com/scientificreports

Competing financial interests: S.Y. is a member without salary of the scientific advisory board of iPS Academia Japan. Y.Y. is a founder of iPS Portal.

How to cite this article: Tokunaga, K. *et al.* Computational image analysis of colony and nuclear morphology to evaluate human induced pluripotent stem cells. *Sci. Rep.* **4**, 6996; DOI:10.1038/srep06996 (2014).

The Review of DIABETIC STUDIES

REVIEW

Profiling of Embryonic Stem Cell Differentiation

Nobuaki Shiraki, Soichiro Ogaki, and Shoen Kume

Department of Stem Cell Biology, Institute of Molecular Embryology and Genetics, Kumamoto University, Honjo 2-2-1, Kumamoto 860-0811, Japan. Address correspondence to: Shoen Kume, e-mail: skume@kumamoto-u.ac.jp

Manuscript submitted MARCH 22, 2013; resubmitted MAY 6, 2013; accepted JULY 3, 2013

Abstract

Embryonic stem (ES) cells have been shown to recapitulate normal developmental stages. They are therefore a highly useful tool in the study of developmental biology. Profiling of ES cell-derived cells has yielded important information about the characteristics of differentiated cells, and allowed the identification of novel marker genes and pathways of differentiation. In this review, we focus on recent results from profiling studies of mouse embryos, human islets, and human ES cell-derived differentiated cells from several research groups. Global gene expression data from mouse embryos have been used to identify novel genes or pathways involved in the developmental process, and to search for transcription factors that regulate direct reprogramming. We introduce gene expression databases of human pancreas cells (Beta Cell Gene Atlas, EuroDia database), and summarize profiling studies of islet- or human ES cell-derived pancreatic cells, with a focus on gene expression, microRNAs, epigenetics, and protein expression. Then, we describe our gene expression profile analyses and our search for novel endoderm, or pancreatic, progenitor marker genes. We differentiated mouse ES cells into mesendoderm, definitive endoderm (DE), mesoderm, ectoderm, and Pdx1-expressing pancreatic lineages, and performed DNA microarray analyses. Genes specifically expressed in DE, and/or in Pdx1-expressing cells, were extracted and their expression patterns in normal embryonic development were studied by *in situ* hybridization. Out of 54 genes examined, 27 were expressed in the DE of E8.5 mouse embryos, and 15 genes were expressed in distinct domains in the pancreatic buds of E14.5 mouse embryos. *Akr1c19*, *Aebp2*, *Pbxip1*, and *Creb3l1* were all novel, and none has been described as being expressed, either in the DE, or in the pancreas. By introducing the profiling results of ES cell-derived cells, the benefits of using ES cells to study early embryonic development will be discussed.

Keywords: diabetes · embryonic stem cell · differentiation · beta-cell · Pdx1 · Ngn3 · Sox · gene profiling · microRNA

1. Introduction

Endoderm gives rise to respiratory and digestive organs, such as pancreas, liver, lung, stomach, and intestine. Multipotent endoderm has the potential to be used in tissue repair. However, despite the importance of definitive endoderm (DE)-derived tissues, not much is known about how they emerge from the primary gut tube. Fate mapping studies suggest that the DE fate begins to segregate at embryonic day 6-6.5 (E6-E6.5), and that the progenitors fated to become specific tissues of the gut tube appear shortly after the completion of gastrulation [1, 2]. The expression of the region-specific transcription factors has provided clues as to how the endoderm is patterned into different organ domains. *Pancreatic and duodenal homeobox gene 1* (*Pdx1*) expression is the first clear sign of pancreatic differentiation, and is detected at E8.5 in the dorsal endoderm of the gut. *Pdx1* is expressed before the buds become evident, and is required for the progression of pancreatic and rostral duodenal development [3]. Genetic lineage tracing studies have shown that Pdx1-expressing cells give rise to all three cell lineages in the pancreas: endocrine, exocrine, and duct cells [4].

Recent advances in the analysis and identification of early endodermal or pancreatic genes has been remarkable [5-9]. Several reports have dem-

onstrated the identification of novel endodermal genes using early embryos. Progress in embryonic stem (ES) cell studies has demonstrated that ES cells provide a good system for studying developmental biology. In particular, the human ES cell differentiation system is a useful tool to study the molecular mechanisms of human embryonic pancreas development, most notably from germ layer specification to pancreatic endocrine cell differentiation [10, 11].

Here, we first summarize gene profiling studies using mouse embryo and mouse/human ES cell-derived cells. Then, we describe our *in vitro* differentiation method and gene expression profile analyses of mouse ES cell-derived DE and *Pdx1*-expressing cells.

2. Profiling studies of mouse early embryos

Analyses of individual genes have defined critical stages in the development of the endocrine pancreas. Global gene expression analyses provide fundamental information on the processes that regulate the normal development of the endocrine pancreas.

2.1 Finding genes that regulate pancreatic development

Gene profiling analyses of E7.5 endoderm and mesectoderm, E10.5 GFP+ and GFP- cells from *Pdx1*/GFP-transgenic mice, E13.5 GFP+ and GFP- cells from *Neurogenin3* (*Ngn3*)/GFP-transgenic mice, and adult islets, have been described [12]. In this study, *Myelin transcription factor 1* (*Myt1*) was identified as a candidate gene expressed in E13.5 *Ngn3*/GFP-positive cells, and an investigation of its loss-of-function revealed that Myt1 is a downstream effector of NGN3 [12]. The function of *Myt1* was further confirmed by a mutant mouse study [13]. Sherwood and coworkers carried out gene expression analysis of the E8.5 DE and visceral endoderm using *Sox17*/GFP-transgenic mice and cell surface markers, such as epithelial cell adhesion molecule (EpCAM) and dolichos biflorus agglutinin (DBA) [7]. By developing an early endoderm gene expression signature, they characterized the transcriptional similarities and differences between DE and visceral endoderm [7]. Also, they performed profiling analyses on several early endodermal organ domains, such as the mouse E11.5 esophageal, lung, distal tracheal, stomach, hepatic, and pancreatic regions, using cell surface markers, such as EpCAM, Liv2, and Rae [8]. An

Abbreviations:

Aebp2 - AE-binding protein 2
AFP - alpha-fetoprotein
Akr1c19 - aldo-keto reductase family 1 member C19
bFGF - basic fibroblast growth factor
BIO - 6-bromoindirubin-3'-oxime
BMP - bone morphogenetic protein
CALB1 - calbindin 1
cAMP - cyclic adenosine monophosphate
CHIP-Seq - chromatin immunoprecipitation DNA-sequencing
Creb3l1- cAMP responsive element binding protein 3-like 1
CTCF - CCCTC-binding factor
Cxcr4 - chemokine (C-X-C motif) receptor 4
DBA - dolichos biflorus agglutinin
DE - definitive endoderm
Dex - dexamethasone
E - embryonic day
ECT - ectoderm
EpCAM - epithelial cell adhesion molecule
ES - embryonic stem
FAIRE - formaldehyde-assisted isolation of regulatory elements
GCNT2 - glucosaminyl (N-acetyl) transferase 2
GFP - green fluorescence protein
GLP1 - glucagon-like peptid 1
GPR50 - G protein-coupled receptor 50
GSK - glycogen synthase kinase
H3K4me3 - histone 3 lysine 4 trimethylation
HGF - hepatocyte growth factor
Hnf1beta - hepatocyte nuclear factor 1beta
IL-6 - interleukin 6
iPS - induced pluripotent stem
KDM5B - lysine (K)-specific demethylase 5B
LPM - lateral plate mesoderm
Mafa - v-maf musculoaponeurotic fibrosarcoma oncogene homolog a
MES - mesendoderm
MPSS - massively parallel signature sequencing
mRNA - messenger ribonucleic acid
Myt1 - myelin transcription factor 1
NEUROG3 - neurogenin 3
Ngn3 - neurogenin 3
PAM - paraxial mesoderm
PAK6 - p21-activated kinase 6
Pbxip1 - pre B cell leukemia transcription factor interacting protein 1
PcG - polycomb
Pdx1 - pancreatic and duodenal homeobox 1
Plekhh1 - pleckstrin homology domain-containing h1
PRDM1 - positive regulatory domain I-binding factor 1
Ptf1a - pancreas transcription factor 1 subunit alpha
Rae - RNA export factor
Rbm47 - RNA-binding motif protein 47
RBPMS2 - RNA-binding protein with multiple splicing 2
SAGE - serial analysis of gene expression
SCID - severe combined immunodeficiency
SOCS3 - suppressor of cytokine signaling 3
Sox17 - sex-determining region Y (Sry) box 17
SSEA1 - stage-specific embryonic antigen 1
STAT3 - signal transducer and activator of transcription 3
TROP2 - tumor-associated calcium signal transducer 2
UCN3 - urocortin 3
Wnt - wingless-type MMTV integration site family

endoderm transcription factor map at E9.5 was constructed, and anterior-posterior patterning dynamics were revealed [8]. Hoffman *et al.* performed serial analysis of gene expression (SAGE) of E10.5-E18.5 pancreas, adult duct, and islets. After extracting candidate genes by SAGE, the GenePaint database was used to validate their results [9]. Selective isolation of cells is necessary for profiling analysis of specific cell types. Fagman *et al.* employed laser capture microdissection and microarray analysis, to define genes expressed in the mouse E10.5 thyroid and lung. They found a regulatory pathway involving the anti-apoptotic gene *Bcl2* that controls cell survival in early thyroid development [14]. These studies indicated that global gene expression analyses of the mouse embryo are useful at the molecular level to characterize the similarities and differences between the various developmental domains, stages, or lineages, and to identify novel genes or pathways involved in developmental processes.

2.2 Finding genes related to reprogramming

Gene expression profiling is also useful to identify candidate genes that regulate reprogramming. Zhou *et al.* performed a genome-wide transcription-factor expression analysis of mouse E14.5 pancreas. The expression pattern of 1,100 mouse transcription factors was confirmed by whole mouse *in situ* hybridization. There are at least 20 transcription factors expressed in mature β-cells and their precursors or endocrine progenitors. Mutagenesis of 9 of these genes resulted in β-cell developmental phenotypes [5]. Reprograming from exocrine cells to pancreatic β-cells was attempted by overexpressing these 9 genes, including 3 transcription factors (*Ngn3, Pdx1, and Mafa)* found to reprogram pancreatic exocrine cells into cells closely resembling pancreatic β-cells [15]. This report demonstrated the possibility of direct cell reprograming into other lineages [16].

3. Profiling studies of human islets

Profiling studies of mature human pancreatic β-cells and islets were performed, with a focus on gene expression [17, 18], microRNAs [19-21], epigenetics [22-24], and protein expression [25].

3.1 Gene expression profiling of human islets

The Beta Cell Gene Atlas (http://www.t1dbase.org/page/AtlasHome) is a useful resource that contains detailed information on the gene expression profiles of pancreatic β-cells, islets, and insulin-producing cell lines. A 'massively parallel signature sequencing (MPSS) analysis' of human pancreatic islet samples and microarray analyses were performed with purified rat pancreatic β-cells, α-cells, and INS-1 cells. The results were compared with array data available in literature [17]. Another database, EuroDia database (http://eurodia.vital-it.ch), was established to build a unique collection of gene expression analyses performed on β-cells of the human, mouse, and rat. The EuroDia database is now available to the entire diabetes research community to ensure continuous access to this valuable data collection after the formal end of the project [18].

3.2 microRNA profiling of human islets

Joglekar and coworkers performed microRNA profiling of human pancreatic islet cells [20]. Islets were isolated from 55 human fetal pancreata during 8-37 weeks of gestation, from neonatal pancreas and human fetal liver. Skin and muscle tissues were used for comparison. It showed that miR-375 might be involved in human pancreatic islet development. Cell sorting based on immunostaining with antibodies against intracellular molecules is possible because microRNAs remain stable after fixation. Based on this technique, microRNA profiling of glucagon+ α-cells, and C-peptide+ β-cells from a human pancreas, were performed [21].

3.3 Epigenetics of human islets

Analyses, using human pancreatic islets, were performed to characterize epigenetic regulation. DNase I hypersensitive sites, histone H3 lysine methylation modifications (H3K4me1, H3K4me3, and H3K79me2), and CCCTC factor (CTCF) binding in human islet cells were examined [24]. Another group mapped the genome-wide location of 4 histone marks (H3K4me1, H3K4me2, H3K4me3, and H3K27me3) [23]. Gaulton and colleagues profiled chromatin, using formaldehyde-assisted isolation of regulatory elements coupled with high-throughput sequencing (FAIRE-seq), to identify regulatory DNA sites active in human pancreatic islets. A comparison of FAIRE-seq data from human islets with data from 5 non-islet cell lines revealed ~3,300 physically linked clusters of islet-selective open chromatin sites, which typically encompassed genes that showed islet-specific expression [22].

These data provided insight into pancreatic β-cell function and the molecular mechanisms causing diabetes. The results are also useful for the validation of ES cell-derived pancreatic β-cells.

4. Human ES cells

Profiling studies of human ES cell-derived cells have been performed to examine the characteristics of the differentiated cells, identify novel marker genes, and understand embryonic development.

4.1 Gene profiling of human ES cell-derived cells

By microarray analysis, *Cxcr4* was identified as a gene that encodes CXCR4, which can be used as a cell surface marker specifically expressed in DE, but not in visceral endoderm [11]. Wang *et al.* established a *Sox17*/GFP knock-in human ES cell line, and carried out gene expression analysis of *Sox17*/GFP+ cells that were differentiated based on the procedure established by the D'Amour group. The results of their gene expression analysis, *in vitro* differentiation, and transplantation-based assays showed that CD49e+CD141+CD238+ cells are primitive gut tube endoderm cells [26]. Human ES cell lines were established, with a *Sox17*/GFP or *Pdx1*/GFP transgene introduced via BAC vectors. These cells are useful resources for the identification of novel cell surface markers. G protein-coupled receptor 50 (GPR50) and tumor-associated calcium signal transducer 2 (TROP2) were identified as cell surface proteins that were highly enriched in pancreatic progenitor cells [27]. The identification of cell surface marker genes enabled the isolation of DE [11, 28], primitive gut tube endoderm [26], and pancreatic progenitor cells [27], without genetic manipulation of ES cells. This method represents a powerful tool for future characterization of similar cell populations.

Recently, a pancreatic differentiation protocol was developed by sequentially exposing human ES cells to different growth factors and small molecules. However, the resultant differentiated cells are immature, and are mostly polyhormonal cells [10, 29]. Using a modified procedure, Basford *et al.* established *Insulin*/GFP knock-in human ES cells for prospective isolation and the study of gene expression profiles by microarray analysis to characterize human ES cell-derived pancreatic cells, both functionally and molecularly [30]. Genomic analyses revealed that *Insulin*/GFP+ cells collectively resemble immature endocrine cells [31]. These findings suggest that additional effort is required to derive fully mature β-cells from human ES cells.

Gene expression profiling of ES cell-derived cells was also performed for other lineages, including neural cells [32], intestinal cells [33], adipocytes [34], or myoblasts [35]. For hepatocyte differentiation, expression profiling was performed to estimate the maturation state of ES cell-derived hepatic cells in comparison with adult hepatocytes [36, 37]. Similarly, ES cells bearing fluorescent reporter genes were used for microarray analyses of hepatic differentiation. Chiao *et al.* used a lentiviral vector containing the alpha fetoprotein promoter to drive enhanced green fluorescent protein expression (AFP:eGFP) [38], and our group established *albumin*/mKO1 knock-in human ES/iPS cells [39].

4.2 MicroRNA profiling of human ES cell-derived cells

MicroRNAs are endogenous small non-coding RNAs that play important roles in embryogenesis, cell fate, growth control, and apoptosis, and are also targets of profiling studies. Human ES cell-derived pancreatic islet-like clusters showed very high expression of the microRNAs miR-186, miR-199a, and miR-339, which downregulate the expression of *LIN28*, *PRDM1*, *CALB1*, *GCNT2*, *RBM47*, *PLEKHH1*, *RBPMS2*, and *PAK6* [40]. Wei *et al.* also reported that miR-34a was expressed during pancreatic progenitor differentiation from endoderm cells, and that miR-146a, miR-7, and miR-375 were specifically expressed during differentiation from pancreatic progenitors to insulin-positive cells. Overexpression of miR-375 downregulated *Hnf1β* and *Sox9* expression [41]. Although these microRNA approaches are promising, further research is required to utilize microRNA for *in vitro* maturation of ES cell-derived pancreatic cells.

4.3 Epigenetics of human ES cell-derived cells

As described above, embryonic development and ES cell differentiation are characterized by dynamic changes in genome-wide gene expression. Yet, the roles of epigenetic modifications remain elusive in these events. Recently, two groups reported the profiling of histone modifications using ES cell-derived pancreatic cells. Gutteridege performed 3 types of genome-wide profiling (mRNA expression, microRNA expression, and histone 3 lysine 4 trimethylation (H3K4me3)), to identify novel pancreatic endocrine maturation pathways.

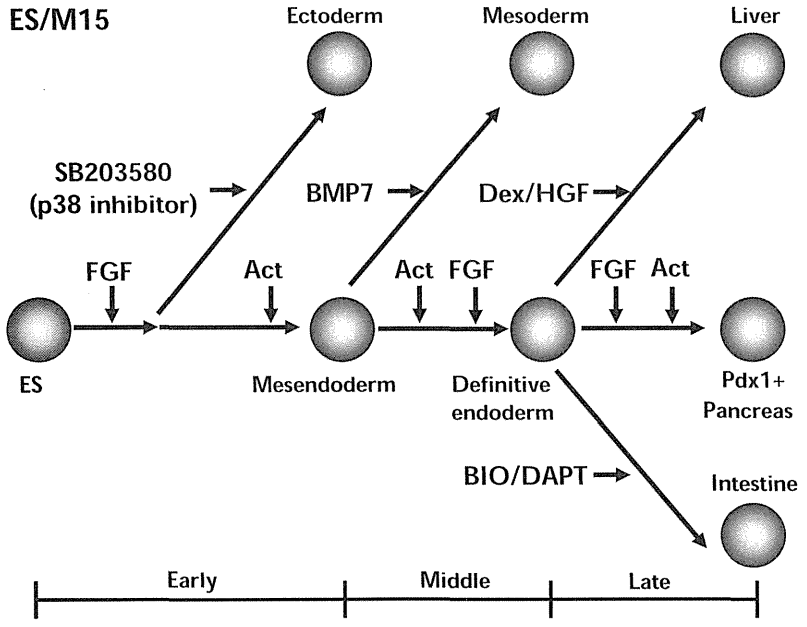

Figure 1. A schematic drawing of M15 cell-mediated signaling events. Signaling molecules involved in the *in vitro* differentiation process mediated by M15 are shown. *Abbreviations*: BIO - 6-bromoindirubin-3'-oxime; BMP7 - bone morphogenetic protein 7; DAPT - N-(N-(3,5-difluorophenacetyl)-L-alanyl)-S-phenylglycine t-butyl ester; Dex - dexamethasone; ES - embryonic stem; FGF - fibroblast growth factor; HGF - hepatocyte growth factor; Pdx1 - pancreatic and duodenal homeobox 1.

H3K4me3 is found at all active transcriptional start sites. Undifferentiated ES (day 0), mesendoderm (day 1), DE (day 2), primitive foregut (day 5), pancreatic progenitor (day 8), and pancreatic endocrine (day 11) cells were used for this profiling study. Data analysis suggested the involvement of novel gene networks, such as NEUROG3/E2F1/KDM5B and SOCS3/STAT3/IL-6, in endocrine cell differentiation. Finally, they showed that the addition of IL-6 increased Nkx2.2 and NEUROG3 expression [42].

Other groups performed RNA-seq and CHIP-seq profiling to identify the gene targets for H3K27me3 and H3K4me3 in ES cell-derived cells. H3K27me3 is enriched in genes that are repressed by polycomb (PcG) proteins. Cells differentiated *in vitro* (gut tube, posterior foregut, pancreatic endoderm, and polyhormonal cells) and functional endocrine cells produced by further differentiation *in vivo* in mice were used for these analyses. They demonstrated that *in vivo*, but not *in vitro*, differentiated endocrine cells exhibit close similarity to human islet and endocrine cells produced *in vitro*, but that they do not fully eliminate the PcG-mediated repression of endocrine-specific genes, such as *insulin*, *GLP1*, and *UCN3*, which are thought to contribute to maturation [43]. Epigenetic profiling of ES cell-derived cells to date has suggested future strategies for manipulating epigenetic signatures to improve cell differentiation *in vitro*.

5. Mouse ES cells

In the next sections, we describe the *in vitro* differentiation method and gene expression profile analysis of mouse ES cell-derived DE and *Pdx1*-expressing cells performed by our group.

5.1 Pancreatic differentiation of mouse ES cells using M15 cells

The embryonic endoderm requires signals from the adjacent germ layers for subsequent regionalization into specific endoderm organs [44]. The requirement to induce signals from the mesoderm led to the idea that coculture of ES cells with a feeder cell line would induce the ES cells to differentiate into DE cells. This led to the discovery of M15, a mesonephros-derived cell line, which has been shown to be an excellent endoderm inductive source [45]. The M15 system efficiently and reproducibly supports ES cells to give rise to the DE and *Pdx1*-expressing cells. The use of a *Pdx1*/GFP-expressing ES cell line (SK7), cultured on M15 cells, has allowed a close examination of the differentiation processes. The differentiation of ES cells to *Pdx1*/GFP-expressing cells is a multistep process. In the early phase, ES cells are first differentiated into mesendoderm (MES) or ectoderm (ECT) cells. In the next phase, the bipotential mesendoderm differentiates into mesoderm or DE. Finally, in the late phase, DE gives rise to region-specific tissue of the endoderm. The molecular bases of the signaling events involved in each step of the process are summarized in **Figure 1**.

Activin and basic fibroblast growth factor (bFGF) both promote ES cell differentiation at all phases of induction. Therefore, activin and/or

bFGF were added throughout the entire process of ES differentiation. The simultaneous treatment of activin and bFGF resulted in a dramatic increase of *Pdx1*/GFP+ cells, from 2% to 31%. When ES cell-derived *Pdx1*/GFP cells were grafted under the kidney capsule of mice with SCID (severe combined immunodeficiency), they differentiated into all 3 pancreatic lineages: endocrine, exocrine, and duct cells. Therefore, the ES cell-derived *Pdx1*/GFP+ cells we obtained had the potential to differentiate similarly into embryonic *Pdx1*/GFP cells.

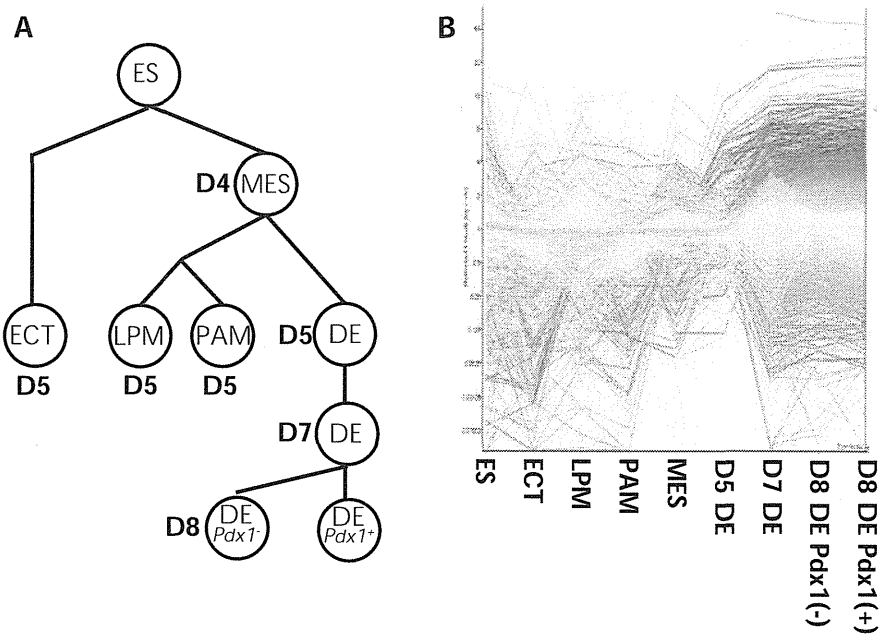

Figure 2. Microarray analyses of ES cell-derived cells. (**A**) ES cells and ES cell-derived cells were isolated based on the expression of cell surface antigens, as previously described [48]. We isolated ES cells (ES), ectoderm (ECT), mesendoderm (MES), lateral plate mesoderm (LPM), paraxial mesoderm (PAM), and DE at day 5 (D5), day 7 (D7), and day 8 (D8, DE *Pdx1*-, DE *Pdx1*+). (**B**) Clustering of gene expression in ES, ECT, LPM, PAM, MES, D5 DE, D7 DE, D8 DE *Pdx1*-, and D8 DE *Pdx1*+ cell lineages. Each line indicates an individual gene. Red lines indicate genes with high expression and green lines indicate genes with low expression in the DE lineages. The y-axis represents normalized values of the expression levels. *Abbreviations*: D - day; DE - definitive endoderm; ES - embryonic stem; ECT - ectoderm; LPM - lateral plate mesoderm; MES - mesendoderm; PAM - paraxial mesoderm; Pdx1 - pancreatic and duodenal homeobox 1.

5.2 Non-pancreatic endoderm, ectoderm, and mesoderm differentiation of mouse ES cells using M15 cells

The M15 cell line was later proved as an inducing source for pancreatic differentiation, for hepatic and intestinal differentiation [46, 47], and for induction of the ectoderm and mesoderm cell lineages [48]. Pancreatic differentiation was at the expense of hepatic differentiation. The withdrawal of activin and bFGF induced alpha-fetoprotein (AFP) expression. The addition of hepatocyte growth factor (HGF) and dexamethasone (Dex) promoted hepatic differentiation [46]. ES cells were differentiated into DE and challenged with various growth factors or chemicals that affect certain signaling pathways at a late stage (**Figure 1**) to establish optimal conditions for differentiation into intestinal cell lineages. Among these tested growth factors and chemicals, we found that intestinal differentiation was efficiently induced through:

1. Activation of the Wnt/β-catenin and inhibition of the Notch signaling pathways.
2. Simultaneous application of 6-bromoindirubin-3'-oxime (BIO), a glycogen synthase kinase (GSK)-3β inhibitor, and DAPT, a known γ-secretase inhibitor [47].

SB203580, a p38 MAPK inhibitor, increased the neuroectodermal population (**Figure 1**). These SB203580-treated cells were multipotent neuronal progenitors able to give rise to astrocytes, oligodendrocytes, neurons, and dopaminergic neurons [48]. Bone morphogenetic protein (BMP) antagonized activin and resulted in the potentiation of mesodermal differentiation (**Figure 1**). Further differentiation into lineage-specific cells was achieved by subjecting ES cell-derived mesodermal cells to adipogenic or osteogenic differentiation conditions. Differentiation into Alizarin red S-

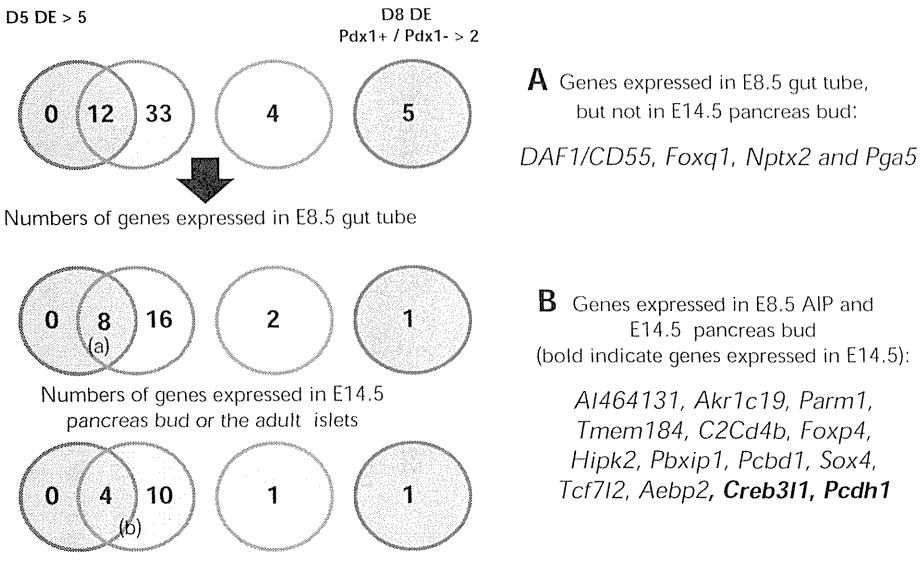

Figure 3. Numbers of endoderm-specific candidate genes. Summary of the numbers of genes selected for further analyses by whole mount *in situ* hybridization (top). The numbers of genes expressed in gut endoderm at E8.5 (middle) or the pancreatic bud at E14.5 (bottom) are shown. Blue circle: genes expressed at >5-fold in D5 DE. Red circle: genes expressed at >5-fold in D8 DE *Pdx1*+. Green circle: genes expressed at >5-fold in D7 DE versus D5 DE. Brown circle, genes expressed at >2-fold in D8 DE *Pdx1*+ versus D8 DE *Pdx1*-. (**A**, **B**) The names of genes expressed only in E8.5 gut tube, but not in E14.5 pancreas bud are listed in (**A**). The names of genes expressed in both E8.5 AIP and E14.5 pancreas bud are listed in (**B**). More information is detailed in **Tables 1** and **2**.

positive osteogenic cells or oil red O-positive adipogenic cells was observed at day 20 [48].

5.3 Microarray analysis of mouse ES cell-derived cells

We isolated mouse ES cell-derived differentiated cells for microarray analysis using the procedure described above, by tracking the expression of specific cell surface antigens using flow cytometry. The cell types and cell surface markers (or GFP) used for prospective cell isolation were: SSEA1-Flk1-PDGFRα- (ES cell-derived ectoderm (ECT)), E-cadherin+PDGFRα+ (mesendoderm (MES)), E-cadherin-PDGFRα+Flk1+ (paraxial mesoderm (PAM)), E-cadherin-PDGFRα-Flk1+ (lateral plate mesoderm, LPM), and E-cadherin+CXCR4+ (DE) populations (**Figure 2**). DE cells at D5, D7, or D8 were collected. DE at D8 was further subdivided into *Pdx1*/GFP-negative and -positive populations (D8 DE *Pdx1*- and *Pdx1*+). A remarkable transition in the gene expression profile was observed from D5 to D7 DE, and thereafter (**Figure 2B**). Comparison between ES cell-derived cells and embryonic tissue, such as E7.5 embryonic endoderm [12], E8.25 endoderm [7], and E10.5 *Pdx1*+ cells [12], suggested that gene expression profiles in D5, D7, and D8 DE were similar to E7.5, E8.25, or E10.5 embryonic *Pdx1*+ cells, respectively. These analyses show that ES cell-derived DE cells or *Pdx1*+ cells mimic cells in normal developmental processes.

5.4 Identification of DE-specific genes in ES cell-differentiation

Gene expression profiles of undifferentiated ES cells and ES cell-derived differentiated cells of the 3 germ layers (ECT, LPM, PAM, MES, D5 DE, D7 DE, *Pdx1*- D8 DE, and *Pdx1*+ D8 DE) were compared. **Figure 3** and **Tables 1** and **2** show the summary of the numbers of genes analyzed and indicate the genes expressed in the gut endoderm at E8.5 and/or the pancreatic bud at E14.5. Thus, these results indicate that ES cell-derived differentiated cells served as a good model cellular system for studying the gene expression of normal developmental stages.

Decay accelerating factor (DAF1/CD55), a gene found to be highly expressed in ES cell-derived DE, was detected in the DE and mesoderm in early embryos at E8.5 [49]. Flow cytometry analysis of ES cell-derived differentiated cells revealed that DAF1+ cells also expressed CXCR4 on the cell surface. Moreover, DAF1 expression is maintained until differentiation day 12 in ES cell-derived DE cells. Analysis of the *Pdx1*/GFP+ cells in E9.5 embryos and ES cell-derived cells with anti-DAF1 revealed that most *Pdx1*/GFP+ cells expressed DAF1. These results suggest that DAF1, when

Table 1. Summary of genes upregulated in definitive endoderm at days 5, 7, and 8

Microarray analysis	Gene	Genbank	*In situ* hybridization			Publication on gut or pancreas	
			E8.5 endoderm	E14.5 pancreas		Expression or function	Reference
d5 DE > 5 and d8 DE > 5, compared with ES, ECT, LPM, PAM, MES, d5 DE	AI464131	BG063189	Whole gut	Epithelium	Mesenchyme	-	-
	Akr1c19	BG073853	AIP	Epithelium		-	-
	DAF1/CD55	NM_010016	AIP, lateral gut			Endoderm	Shiraki *et al.* 2010 [49]
	Foxq1	NM_008239	AIP			Stomach	Verzi *et al.* 2008 [50]
	Nptx2	NM_016789	Whole gut			Pancreatic cancer	Brune *et al.* 2008 [60]
	Parm1	NM_145562	Anterior endoderm	Tip		E10.5 pancreas	Svensson *et al.* 2007 [6]
	Pga5	NM_021453	lateral gut			-	-
	Tmem184a	BC019731	AIP, lateral gut	Epithelium		E12.5, pancreas exocrine	Best and Adams 2009 [58]
d7 DE / d5 DE > 5	Aebp2	BB667191	whole gut	Epithelium		-	-
	Barhl2	NM_001005477	lateral gut			-	-
d8 DE, Pdx1(GFP)+ /Pdx1(GFP)- > 2	Kiss1r	NM_053244	lateral gut			Mouse islets	Hauge-Evans *et al.* 2006 [70]

Legend: Enlisted are genes upregulated (>5-fold) in d5 DE or d7 DE, or upregulated (>2-fold) in d8 DE, and their expression patterns observed in E8.5 endoderm or E14.5 pancreatic buds. Data include gene, genbank number, expression patterns in E8.5 endoderm and E14.5 pancreatic buds, and original publication. Table created based on [80]. *Abbreviations*: Aebp2 - adipocyte enhancer-binding protein 2; AI464131 - expressed sequence AI464131; AIP - anterior intestinal portal; Akr1c19 - aldo-keto reductase family 1, member C19; d - day; Barhl2 - Barh-like homeobox 2 (Drosophila); DAF1 - decay-accelerating factor 1 (aka CD55); DE - definite endoderm; E - embryonic day; ECT - ectoderm; ES - embryonic stem cell; Foxq1 - forkhead box transcription factor Q1; GFP green fluorescent protein; Kiss1r - Kiss1 receptor; LPM - lateral plate mesoderm; MES - mesendoderm; Nptx2 - neuronal pentraxin 2; PAM - paraxial mesoderm; Parm1 - prostate androgen-regulated mucin-like protein 1 (Riken cDNA 9130213B05 gene); Pdx1 - pancreatic and duodenal homeobox 1; PGA5 - pepsinogen 5, group I (pepsinogen A); Tmem184a - transmembrane protein 184A.

used in combination with E-cadherin, is useful for the prospective identification of DE cells.

Among the genes whose expression is increased in the ES cell-derived DE population, *Foxq1* [50], *CpM* [51-53], *Foxp4* [54, 55], *Pcdh1* [56], and *Zmiz1* [57] were found to be expressed in the foregut, hindgut, or whole gut at E8.5. *Parm1* [6], *Tmem184* [58], *HIPK-2* [59], *Nptx2* [60, 61], *Tcf7l2* [62-65], *C2Cd4b* [66], *Sox4* [67-69], and *Kiss1r* [70-72] were revealed for the first time to be expressed at this early stage of E8.5 and E14.5. *Hipk2* was co-expressed with glucagon, but not insulin, implicating that it might be associated with β-cell differentiation (**Figure 4**). *C2cd4b*, a gene expressed in the trunk, was co-expressed with insulin, but not glucagon, implicating its function in endocrine β-cell differentiation (**Figure 4**). It is of interest that genes responsible for β-cell maturation are expressed at early stages of development. Future studies examining the functions of these genes should reveal their role in β-cell replication or differentiation of the pancreas.

We found for the first time that the following 4 genes are expressed in the E8.5 endoderm or E14.5 pancreas:

1. *Aldo-keto reductase family 1 member C19 (Akr1c19)*
2. *AE binding protein 2 (Aebp2)*
3. *Pre B cell leukemia transcription factor interacting protein 1 (Pbxip1)*
4. *cAMP responsive element binding protein 3-like 1 (Creb3l1)*

Akr1c19 was reported to be highly expressed in the liver and gastrointestinal tract [73]. *Aebp2* encodes a zinc finger protein that interacts with the mammalian polycomb repression complex 2 (PRC2) [74]. Its *Drosophila* homolog, *jing*, is a zinc-finger transcription factor that interacts with the fly polycomb group (PcG) protein complexes, and plays an essential role in controlling CNS midline and tracheal cell differentiation [75]. *Pbxip1* is a PBX interacting protein, also known as HPIP,

Table 2. Summary of genes upregulated in definitive endoderm at day 8

Microarray analysis	Gene	Genbank	*In situ* hybridization			Publication on gut or pancreas	
			E8.5 endoderm	E14.5 pancreas		Expression or function	Reference
d8 DE, Pdx1 (GFP)+ > 5, compared with ES, ECT, LPM, PAM, MES, d5 DE	ApoE	AK019319	Visceral endoderm		Vascular	-	-
	C2cd4b	AK014341	AIP, posterior gut	Trunk		Associated with β-cell function	Boesgaard *et al.* 2010 [66]
	Chi3l1	BC005611	AIP			-	-
	CpM	AK004327	AIP, lateral gut			Lung	Nagae *et al.* 1993 [51]
	Creb3l1	BC016447	-	Epithelium	Mesenchyme	-	-
	Fam188b	BB667136	AIP			-	-
	Fhl2	NM_010212	AIP, anterior gut			-	-
	Foxp4	BQ286886	AIP, lateral gut	Epithelium		E9.5~, pulmonary, gut	Lu *et al.* 2002 [54]
	Hipk2	NM_010433	AIP	Epithelium		E12.5~, pancreas	Boucher *et al.* 2009 [59]
	Irf6	NM_016851	Anterior gut, hind-gut			-	-
	Lbh	NM_029999	AIP			-	-
	Palld	NM_001081390	Dorsal gut			-	-
	Pbxip1	AV220340	AIP	Trunk		-	-
	Pcbd1	NM_025273	AIP	Epithelium			-
	Pcdh1	AK008111	-	Tip	Mesenchyme	E12.5 blood vessels of the gut	Redies *et al.* 2008 [56]
	Sox4	AI428101	AIP, lateral gut	Epithelium		E12.4~, pancreas	Lioubinski *et al.* 2003 [67]
	Tcf7l2	BM218908	AIP, lateral gut	Epithelium	Mesenchyme	diabetes risk gene	Grant *et al.* 2006 [62]
	Zmiz1	NM_183208	AIP, lateral gut				-

Legend: Enlisted are genes upregulated (>5-fold) at d8 DE, and their expression patterns observed in E8.5 endoderm or E14.5 pancreatic buds. Data include gene, genbank number, expression patterns in E8.5 endoderm and E14.5 pancreatic buds, and original publication. Table created based on [80]. *Abbreviations*: AIP - anterior intestinal portal; ApoE - apolipoprotein E; C2cd4b - C2 calcium-dependent domain containing 4B; cAMP - cyclic adenosine monophosphate; Chi3l1 - chitinase 3-like 1; CpM - carboxypeptidase M; Creb3l1 - cAMP-responsive element-binding protein 3-like 1 (aka OASIS); ECT - ectoderm; ES - embryonic stem cell; Fam188b - family with sequence similarity 188, member B (RIKEN cDNA C330043M08 gene); Fhl2 - four and a half LIM domains 2; Foxp4 - forkhead box P4; Hipk2 - homeodomain-interacting protein kinase 2; HMG-box - high mobility group box; Irf6 - interferon-regulatory factor 6; MES - mesendoderm; Lbh - limb-bud-and-heart; LIM - Lin-11, Isl-1, Mac-3; LPM - lateral plate mesoderm; Palld - palladin, cytoskeletal associated protein (2410003B16Rik, immunoglobulin domain paladin); PAM - paraxial mesoderm; Pbxip1 - pre-B-cell leukemia transcription factor interacting protein 1; Pcbd1 - Pterin 4 alpha carbinolamine dehydratase/dimerization cofactor of hepatocyte nuclear factor 1 (TCF1); Pcdh1 - PUrotocadherin 1; Sox4 - SRY-box-containing gene 4; Tcf7l2 - transcription factor 7-like 2, T-cell-specific, HMG-box; Zmiz1 - zinc finger, MIZ-type-containing 1.

which inhibits the binding of Pbx1-Hox complexes to DNA [76]. *Creb3l1*, also known as *OASIS*, is a ZIP (basic leucine zipper) transcription factor, which is a member of the CREB/ATF family and has been identified as an ER stress transducer [77].

There are genes whose expression we could not detect in Pdx1+ cells during normal pancreatic development. This might be due to their low expression levels and/or technical limitations of our experimental setup. In addition, some of the genes show expression patterns that are difficult to be catalogued at E14.5, since pancreatic differentiation undergoes a secondary transition at this stage, and many genes show a dramatic change in their expression patterns after this transition.

6. Conclusions

We reviewed gene expression profiling studies using mouse embryo, islets, and ES cell-derived cells, and described our *in vitro* differentiation method that used feeder cells and growth factors. Then, we described our gene expression profile

analyses. These analyses revealed that ES cell-derived cells mimic cells that arise during normal development. Profiling of ES cell-derived cells yielded important information about the characteristics of differentiated cells, identified novel marker genes, and revealed novel pathways of differentiation. Currently, several groups have reported the generation of pancreatic β-cell like cells. Although these cells were immature human pancreatic progenitor cells, they matured into functional β-cell after transplantation [10, 29].

Multi-level genome-wide profiling assessing gene expression, microRNAs expression, proteome composition, metabolome makeup, DNA methylation patterns, and histone modifications might provide us with useful information to induce *in vitro* maturation of ES cell-derived pancreatic cells. During the last decade, high-throughput techniques have been developed, including microarray and next-generation sequencing, together with public databases, such as Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/), the RIKEN FANTOM project (http://fantom.gsc.riken.jp/), Genepaint [78, 79], and the Mouse Atlas Website (http://www.mouseatlas.org/). In the near future, profiling studies using the aforementioned new technologies will lead to the identification of novel signaling molecules which may promote pancreatic development, and which may offer novel targets for the treatment of diabetes.

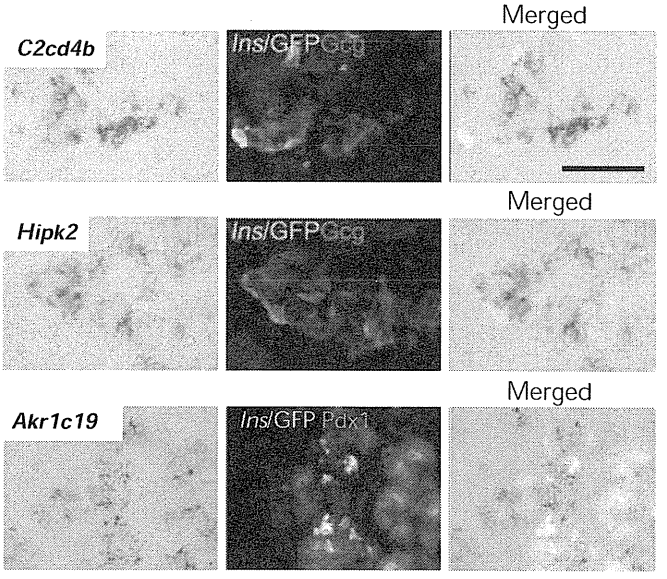

Figure 4. The co-expression of candidate genes with insulin, glucagon, or Pdx1 in the E14.5 pancreatic bud. *C2cd4b* was co-expressed with insulin, but not glucagon, in the trunk (upper panels). *Hipk2* was co-expressed with glucagon, but not insulin, in the epithelium. *Akr1c19* was co-expressed with Pdx1 or insulin in the epithelium. Scale bar: 100 μm.

Disclosures: The authors do not have any conflict of interests. There are no financial, institutional, corporate, or patent-licensing arrangements associated with this paper.

References

1. **Tremblay KD, Zaret KS.** Distinct populations of endoderm cells converge to generate the embryonic liver bud and ventral foregut tissues. *Dev Biol* 2005. 280(1):87-99.
2. **Lawson KA, Meneses JJ, Pedersen RA.** Clonal analysis of epiblast fate during germ layer formation in the mouse embryo. *Development* 1991. 113(3):891-911.
3. **Offield MF, Jetton TL, Labosky PA, Ray M, Stein RW, Magnuson MA, Hogan BL, Wright CV.** PDX-1 is required for pancreatic outgrowth and differentiation of the rostral duodenum. *Development* 1996. 122(3):983-995.
4. **Gu G, Brown JR, Melton DA.** Direct lineage tracing reveals the ontogeny of pancreatic cell fates during mouse embryogenesis. *Mech Dev* 2003. 120(1):35-43.
5. **Zhou Q, Law AC, Rajagopal J, Anderson WJ, Gray PA, Melton DA.** A multipotent progenitor domain guides pancreatic organogenesis. *Dev Cell* 2007. 13(1):103-114.
6. **Svensson P, Williams C, Lundeberg J, Ryden P, Bergqvist I, Edlund H.** Gene array identification of Ipf1/Pdx1-/- regulated genes in pancreatic progenitor cells. *BMC Dev Biol* 2007. 7:129.
7. **Sherwood RI, Jitianu C, Cleaver O, Shaywitz DA, Lamenzo JO, Chen AE, Golub TR, Melton DA.** Prospective isolation and global gene expression analysis of definitive and visceral endoderm. *Dev Biol* 2007. 304(2):541-555.
8. **Sherwood RI, Chen TY, Melton DA.** Transcriptional dynamics of endodermal organ formation. *Dev Dyn* 2009. 238(1):29-42.
9. **Hoffman BG, Zavaglia B, Witzsche J, Ruiz de Algara T, Beach M, Hoodless PA, Jones SJ, Marra MA, Helgason CD.** Identification of transcripts with enriched expression in the developing and adult pancreas. *Genome Biol* 2008. 9(6):R99.
10. **D'Amour KA, Bang AG, Eliazer S, Kelly OG, Agulnick AD, Smart NG, Moorman MA, Kroon E, Carpenter MK, Baetge EE.** Production of pancreatic hormone-expressing endocrine cells from human embryonic stem cells. *Nat Biotechnol* 2006. 24(11):1392-1401.
11. **D'Amour KA, Agulnick AD, Eliazer S, Kelly OG, Kroon E, Baetge EE.** Efficient differentiation of human embryonic stem cells to definitive endoderm. *Nat Biotechnol* 2005. 23(12):1534-1541.
12. **Gu G, Wells JM, Dombkowski D, Preffer F, Aronow B, Melton DA.** Global expression analysis of gene regula-

tory pathways during endocrine pancreatic development. *Development* 2004. 131(1):165-179.
13. **Wang S, Zhang J, Zhao A, Hipkens S, Magnuson MA, Gu G.** Loss of Myt1 function partially compromises endocrine islet cell differentiation and pancreatic physiological function in the mouse. *Mech Dev* 2007. 124(11-12):898-910.
14. **Fagman H, Amendola E, Parrillo L, Zoppoli P, Marotta P, Scarfo M, De Luca P, de Carvalho DP, Ceccarelli M, De Felice M, Di Lauro R.** Gene expression profiling at early organogenesis reveals both common and diverse mechanisms in foregut patterning. *Dev Biol* 2011. 359(2):163-175.
15. **Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA.** In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* 2008. 455(7213):627-632.
16. **Vierbuchen T, Wernig M.** Direct lineage conversions: unnatural but useful? *Nat Biotechnol* 2011. 29(10):892-907.
17. **Kutlu B, Burdick D, Baxter D, Rasschaert J, Flamez D, Eizirik DL, Welsh N, Goodman N, Hood L.** Detailed transcriptome atlas of the pancreatic beta cell. *BMC Med Genomics* 2009. 2:3.
18. **Liechti R, Csardi G, Bergmann S, Schutz F, Sengstag T, Boj SF, Servitja JM, Ferrer J, Van Lommel L, Schuit F, et al.** EuroDia: a beta-cell gene expression resource. *Database (Oxford)* 2010. 2010:baq024.
19. **Joglekar MV, Parekh VS, Mehta S, Bhonde RR, Hardikar AA.** MicroRNA profiling of developing and regenerating pancreas reveal post-transcriptional regulation of neurogenin3. *Dev Biol* 2007. 311(2):603-612.
20. **Joglekar MV, Joglekar VM, Hardikar AA.** Expression of islet-specific microRNAs during human pancreatic development. *Gene Expr Patterns* 2009. 9(2):109-113.
21. **Klein D, Misawa R, Bravo-Egana V, Vargas N, Rosero S, Piroso J, Ichii H, Umland O, Zhijie J, Tsinoremas N, et al.** MicroRNA Expression in Alpha and Beta Cells of Human Pancreatic Islets. *PLoS One* 2013. 8(1):e55064.
22. **Gaulton KJ, Nammo T, Pasquali L, Simon JM, Giresi PG, Fogarty MP, Panhuis TM, Mieczkowski P, Secchi A, Bosco D, et al.** A map of open chromatin in human pancreatic islets. *Nat Genet* 2010. 42(3):255-259.
23. **Bhandare R, Schug J, Le Lay J, Fox A, Smirnova O, Liu C, Naji A, Kaestner KH.** Genome-wide analysis of histone modifications in human pancreatic islets. *Genome Res* 2010. 20(4):428-433.
24. **Stitzel ML, Sethupathy P, Pearson DS, Chines PS, Song L, Erdos MR, Welch R, Parker SC, Boyle AP, Scott LJ, et al.** Global epigenomic analysis of primary human pancreatic islets provides insights into type 2 diabetes susceptibility loci. *Cell Metab* 2010. 12(5):443-455.
25. **Waanders LF, Chwalek K, Monetti M, Kumar C, Lammert E, Mann M.** Quantitative proteomic analysis of single pancreatic islets. *Proc Natl Acad Sci U S A* 2009. 106(45):18902-18907.
26. **Wang P, Rodriguez RT, Wang J, Ghodasara A, Kim SK.** Targeting SOX17 in human embryonic stem cells creates unique strategies for isolating and analyzing developing endoderm. *Cell Stem Cell* 2011. 8(3):335-346.
27. **Fishman B, Segev H, Kopper O, Nissenbaum J, Schulman M, Benvenisty N, Itskovitz-Eldor J, Kitsberg D.** Targeting pancreatic progenitor cells in human embryonic stem cell differentiation for the identification of novel cell surface markers. *Stem Cell Rev* 2012. 8(3):792-802.
28. **Yasunaga M, Tada S, Torikai-Nishikawa S, Nakano Y, Okada M, Jakt LM, Nishikawa S, Chiba T, Era T, Nishikawa S.** Induction and monitoring of definitive and visceral endoderm differentiation of mouse ES cells. *Nat Biotechnol* 2005. 23(12):1542-1550.
29. **Kroon E, Martinson LA, Kadoya K, Bang AG, Kelly OG, Eliazer S, Young H, Richardson M, Smart NG, Cunningham J, et al.** Pancreatic endoderm derived from human embryonic stem cells generates glucose-responsive insulin-secreting cells in vivo. *Nat Biotechnol* 2008. 26(4):443-452.
30. **Nostro MC, Sarangi F, Ogawa S, Holtzinger A, Corneo B, Li X, Micallef SJ, Park IH, Basford C, Wheeler MB, et al.** Stage-specific signaling through TGFbeta family members and WNT regulates patterning and pancreatic specification of human pluripotent stem cells. *Development* 2011. 138(5):861-871.
31. **Basford CL, Prentice KJ, Hardy AB, Sarangi F, Micallef SJ, Li X, Guo Q, Elefanty AG, Stanley EG, Keller G, et al.** The functional and molecular characterisation of human embryonic stem cell-derived insulin-positive cells compared with adult pancreatic beta cells. *Diabetologia* 2012. 55(2):358-371.
32. **Li W, Sun W, Zhang Y, Wei W, Ambasudhan R, Xia P, Talantova M, Lin T, Kim J, Wang X, et al.** Rapid induction and long-term self-renewal of primitive neural precursors from human embryonic stem cells by small molecule inhibitors. *Proc Natl Acad Sci U S A* 2011. 108(20):8299-8304.
33. **Spence JR, Mayhew CN, Rankin SA, Kuhar MF, Vallance JE, Tolle K, Hoskins EE, Kalinichenko VV, Wells SI, Zorn AM, et al.** Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro. *Nature* 2011. 470(7332):105-109.
34. **Billon N, Kolde R, Reimand J, Monteiro MC, Kull M, Peterson H, Tretyakov K, Adler P, Wdziekonski B, Vilo J, Dani C.** Comprehensive transcriptome analysis of mouse embryonic stem cell adipogenesis unravels new processes of adipocyte development. *Genome Biol* 2010. 11(8):R80.
35. **Barberi T, Bradbury M, Dincer Z, Panagiotakos G, Socci ND, Studer L.** Derivation of engraftable skeletal myoblasts from human embryonic stem cells. *Nat Med* 2007. 13(5):642-648.
36. **DeLaForest A, Nagaoka M, Si-Tayeb K, Noto FK, Konopka G, Battle MA, Duncan SA.** HNF4A is essential for specification of hepatic progenitors from human pluripotent stem cells. *Development* 2011. 138(19):4143-4153.
37. **Si-Tayeb K, Noto FK, Nagaoka M, Li J, Battle MA, Duris C, North PE, Dalton S, Duncan SA.** Highly efficient generation of human hepatocyte-like cells from induced pluripotent stem cells. *Hepatology* 2010. 51(1):297-305.
38. **Chiao E, Elazar M, Xing Y, Xiong A, Kmet M, Millan MT, Glenn JS, Wong WH, Baker J.** Isolation and transcriptional profiling of purified hepatic cells derived from human embryonic stem cells. *Stem Cells* 2008. 26(8):2032-2041.
39. **Umeda K, Suzuki K, Yamazoe T, Shiraki N, Higuchi Y, Tokieda K, Kume K, Mitani K, Kume S.** Albumin

gene targeting in human embryonic stem cells and induced pluripotent stem cells with helper-dependent adenoviral vector to monitor hepatic differentiation. *Stem Cell Res* 2013. 10(2):179-194.

40. **Chen BZ, Yu SL, Singh S, Kao LP, Tsai ZY, Yang PC, Chen BH, Shoei-Lung Li S.** Identification of microRNAs expressed highly in pancreatic islet-like cell clusters differentiated from human embryonic stem cells. *Cell Biol Int* 2011. 35(1):29-37.
41. **Wei R, Yang J, Liu GQ, Gao MJ, Hou WF, Zhang L, Gao HW, Liu Y, Chen GA, Hong TP.** Dynamic expression of microRNAs during the differentiation of human embryonic stem cells into insulin-producing cells. *Gene* 2013.
42. **Gutteridge A, Rukstalis JM, Ziemek D, Tie M, Ji L, Ramos-Zayas R, Nardone NA, Norquay LD, Brenner MB, Tang K, et al.** Novel pancreatic endocrine maturation pathways identified by genomic profiling and causal reasoning. *Plos One* 2013. 8(2):e56024.
43. **Xie R, Everett LJ, Lim HW, Patel NA, Schug J, Kroon E, Kelly OG, Wang A, D'Amour KA, Robins AJ, et al.** Dynamic chromatin remodeling mediated by polycomb proteins orchestrates pancreatic differentiation of human embryonic stem cells. *Cell Stem Cell* 2013. 12(2):224-237.
44. **Wells JM, Melton DA.** Early mouse endoderm is patterned by soluble factors from adjacent germ layers. *Development* 2000. 127(8):1563-1572.
45. **Shiraki N, Yoshida T, Araki K, Umezawa A, Higuchi Y, Goto H, Kume K, Kume S.** Guided differentiation of embryonic stem cells into Pdx1-expressing regional-specific definitive endoderm. *Stem Cells* 2008. 26(4):874-885.
46. **Yoshida T, Shiraki N, Baba H, Goto M, Fujiwara S, Kume K, Kume S.** Expression patterns of epiplakin1 in pancreas, pancreatic cancer and regenerating pancreas. *Genes Cells* 2008. 13(7):667-678.
47. **Ogaki S, Shiraki N, Kume K, Kume S.** Wnt and notch signals guide embryonic stem cell differentiation into the intestinal lineages. *Stem Cells* 2013. 31(6):1086-1096.
48. **Shiraki N, Higuchi Y, Harada S, Umeda K, Isagawa T, Aburatani H, Kume K, Kume S.** Differentiation and characterization of embryonic stem cells into three germ layers. *Biochem Biophys Res Commun* 2009. 381(4):694-699.
49. **Shiraki N, Harada S, Ogaki S, Kume K, Kume S.** Identification of DAF1/CD55, a novel definitive endoderm marker. *Cell Struct Funct* 2010. 35(2):73-80.
50. **Verzi MP, Khan AH, Ito S, Shivdasani RA.** Transcription factor foxq1 controls mucin gene expression and granule content in mouse stomach surface mucous cells. *Gastroenterology* 2008. 135(2):591-600.
51. **Nagae A, Abe M, Becker RP, Deddish PA, Skidgel RA, Erdos EG.** High concentration of carboxypeptidase M in lungs: presence of the enzyme in alveolar type I cells. *Am J Respir Cell Mol Biol* 1993. 9(2):221-229.
52. **Fujiwara N, Ikeda M, Hirabayashi S, Mori H, Shirasawa M, Kansaku A, Sunamori M, Hata Y.** Monoclonal antibody 7F9 recognizes rat protein homologous to human carboxypeptidase-M in developing and adult rat lung. *Respirology* 2007. 12(1):54-62.
53. **Tamplin OJ, Kinzel D, Cox BJ, Bell CE, Rossant J, Lickert H.** Microarray analysis of Foxa2 mutant mouse embryos reveals novel gene expression and inductive roles for the gastrula organizer and its derivatives. *BMC Genomics* 2008. 9:511.
54. **Lu MM, Li S, Yang H, Morrisey EE.** Foxp4: a novel member of the Foxp subfamily of winged-helix genes co-expressed with Foxp1 and Foxp2 in pulmonary and gut tissues. *Mech Dev* 2002. 119(Suppl 1):S197-S202.
55. **Chokas AL, Trivedi CM, Lu MM, Tucker PW, Li S, Epstein JA, Morrisey EE.** Foxp1/2/4-NuRD interactions regulate gene expression and epithelial injury response in the lung via regulation of interleukin-6. *J Biol Chem.* 285(17):13304-13313.
56. **Redies C, Heyder J, Kohoutek T, Staes K, Van Roy F.** Expression of protocadherin-1 (Pcdh1) during mouse development. *Dev Dyn* 2008. 237(9):2496-2505.
57. **Rodriguez-Magadan H, Merino E, Schnabel D, Ramirez L, Lomeli H.** Spatial and temporal expression of Zimp7 and Zimp10 PIAS-like proteins in the developing mouse embryo. *Gene Expr Patterns* 2008. 8(3):206-213.
58. **Best D, Adams IR.** Sdmg1 is a component of secretory granules in mouse secretory exocrine tissues. *Dev Dyn* 2009. 238(1):223-231.
59. **Boucher MJ, Simoneau M, Edlund H.** The homeodomain-interacting protein kinase 2 regulates insulin promoter factor-1/pancreatic duodenal homeobox-1 transcriptional activity. *Endocrinology* 2009. 150(1):87-97.
60. **Brune K, Hong SM, Li A, Yachida S, Abe T, Griffith M, Yang D, Omura N, Eshleman J, Canto M, et al.** Genetic and epigenetic alterations of familial pancreatic cancers. *Cancer Epidemiol Biomarkers Prev* 2008. 17(12):3536-3542.
61. **Zhang L, Gao J, Li L, Li Z, Du Y, Gong Y.** The neuronal pentraxin II gene (NPTX2) inhibit proliferation and invasion of pancreatic cancer cells in vitro. *Mol Biol Rep* 2011. 38(8):4903-4911.
62. **Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, Stefansson H, Emilsson V, Helgadottir A, et al.** Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. *Nat Genet* 2006. 38(3):320-323.
63. **Jin T, Liu L.** The Wnt signaling pathway effector TCF7L2 and type 2 diabetes mellitus. *Mol Endocrinol* 2008. 22(11):2383-2392.
64. **Shu L, Sauter NS, Schulthess FT, Matveyenko AV, Oberholzer J, Maedler K.** Transcription factor 7-like 2 regulates beta-cell survival and function in human pancreatic islets. *Diabetes* 2008. 57(3):645-653.
65. **da Silva Xavier G, Loder MK, McDonald A, Tarasov AI, Carzaniga R, Kronenberger K, Barg S, Rutter GA.** TCF7L2 regulates late events in insulin secretion from pancreatic islet beta-cells. *Diabetes* 2009. 58(4):894-905.
66. **Boesgaard TW, Grarup N, Jorgensen T, Borch-Johnsen K, Hansen T, Pedersen O.** Variants at DGKB/TMEM195, ADRA2A, GLIS3 and C2CD4B loci are associated with reduced glucose-stimulated beta cell function in middle-aged Danish people. *Diabetologia* 2010. 53(8):1647-1655.
67. **Lioubinski O, Muller M, Wegner M, Sander M.** Expression of Sox transcription factors in the developing mouse pancreas. *Dev Dyn* 2003. 227(3):402-408.
68. **Ragvin A, Moro E, Fredman D, Navratilova P, Drivenes O, Engstrom PG, Alonso ME, de la Calle Mustienes E, Gomez Skarmeta JL, Tavares MJ, et al.** Long-range gene regulation links genomic type 2 diabetes and obesity risk regions to HHEX, SOX4, and IRX3. *Proc*

Natl Acad Sci U S A 2010. 107(2):775-780.
69. **Goldsworthy M, Hugill A, Freeman H, Horner E, Shimomura K, Bogani D, Pieles G, Mijat V, Arkell R, Bhattacharya S, et al.** Role of the transcription factor sox4 in insulin secretion and impaired glucose tolerance. *Diabetes* 2008. 57(8):2234-2244.
70. **Hauge-Evans AC, Richardson CC, Milne HM, Christie MR, Persaud SJ, Jones PM.** A role for kisspeptin in islet function. *Diabetologia* 2006. 49(9):2131-2135.
71. **Funes S, Hedrick JA, Vassileva G, Markowitz L, Abbondanzo S, Golovko A, Yang S, Monsma FJ, Gustafson EL.** The KiSS-1 receptor GPR54 is essential for the development of the murine reproductive system. *Biochem Biophys Res Commun* 2003. 312(4):1357-1363.
72. **Bowe JE, King AJ, Kinsey-Jones JS, Foot VL, Li XF, O'Byrne KT, Persaud SJ, Jones PM.** Kisspeptin stimulation of insulin secretion: mechanisms of action in mouse islets and rats. *Diabetologia* 2009. 52(5):855-862.
73. **Ishikura S, Horie K, Sanai M, Matsumoto K, Hara A.** Enzymatic properties of a member (AKR1C19) of the aldo-keto reductase family. *Biol Pharm Bull* 2005. 28(6):1075-1078.
74. **Kim H, Kang K, Kim J.** AEBP2 as a potential targeting protein for Polycomb Repression Complex PRC2. *Nucleic Acids Res* 2009. 37(9):2940-2950.
75. **Sedaghat Y, Miranda WF, Sonnenfeld MJ.** The jing Zn-finger transcription factor is a mediator of cellular differentiation in the Drosophila CNS midline and trachea. *Development* 2002. 129(11):2591-2606.
76. **Abramovich C, Shen WF, Pineault N, Imren S, Montpetit B, Largman C, Humphries RK.** Functional cloning and characterization of a novel nonhomeodomain protein that inhibits the binding of PBX1-HOX complexes to DNA. *J Biol Chem* 2000. 275(34):26172-26177.
77. **Murakami T, Saito A, Hino S, Kondo S, Kanemoto S, Chihara K, Sekiya H, Tsumagari K, Ochiai K, Yoshinaga K, et al.** Signalling mediated by the endoplasmic reticulum stress transducer OASIS is involved in bone formation. *Nat Cell Biol* 2009. 11(10):1205-1211.
78. **Visel A, Thaller C, Eichele G.** GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res* 2004. 32(Database issue):D552-556.
79. **Visel A, Carson J, Oldekamp J, Warnecke M, Jakubcakova V, Zhou X, Shaw CA, Alvarez-Bolado G, Eichele G.** Regulatory pathway analysis by high-throughput in situ hybridization. *Plos Genet* 2007. 3(10):1867-1883.
80. **Ogaki S, Harada S, Shiraki N, Kume K, Kume S.** An expression profile analysis of ES cell-derived definitive endodermal cells and Pdx1-expressing cells. *BMC Dev Biol* 2011. 11:13.

PEDIATRICS INTERNATIONAL Official Journal of the Japan Pediatric Society

Pediatrics International (2015) **57**, e1–e3 doi: 10.1111/ped.12457

Patient Report

Early intervention for late-onset ornithine transcarbamylase deficiency

Daisuke Fujisawa,[1] Hiroshi Mitsubuchi,[1] Shirou Matsumoto,[1] Masanori Iwai,[1] Kimitoshi Nakamura,[1] Ryuji Hoshide,[1] Nawomi Harada,[2] Makoto Yoshino[3] and Fumio Endo[1]

[1]*Department of Pediatrics, Kumamoto University Graduate School of Medical Sciences, Kumamoto,* [2]*Department of Pediatrics and Child Health, Kurume University of Medicine and* [3]*Division of Gene Therapy and Regenerative Medicine, Cognitive and Molecular Research Institute of Brain Diseases, Kurume University, Kurume, Japan*

Abstract We report the case of a family with late-onset ornithine transcarbamylase deficiency (OTCD). Several family members had died from OTCD, and the c.221G>A, p.Lys221Lys mutation was detected at the 3′ end of exon 6 of *OTC* in the X-chromosome of some members. We provided genetic counseling on pregnancy, delivery, and neonate management to a 4th-generation female carrier and decided on metabolic management of her child from birth. Two male patients were diagnosed with late-onset OTCD on the basis of blood amino acid and genetic analysis, and they received arginine supplementation from the asymptomatic, early neonatal period. These children grew and developed normally, without decompensation. Patients with late-onset OTCD can and should be diagnosed and treated in the early neonatal period, especially those from families already diagnosed with late-onset OTCD, and family members must be provided with genetic counseling.

Key words arginine supplementation, blood amino acid analysis, genetic analysis, genetic counseling, late-onset ornithine transcarbamylase deficiency.

Ornithine transcarbamylase deficiency (OTCD) is one of the most common urea cycle disorders, with an estimated prevalence of 1 per 80 000 births. It is transmitted as an X-linked trait and the gene responsible for it, namely, *OTC*, is located on locus Xp21.1.[1] A remarkable feature of OTCD is that the phenotype is extremely heterogenous.[2,3]

Male neonates with OTCD often suffer from ammonia toxicity, protein intolerance, and death within 1 week after birth. Those with late-onset OTCD (age at onset, ≥28 days), however, have several, wide-ranging symptoms, from asymptomatic phenotype to hyperammonemia, coma, and death.[1,4,5] In heterozygous female neonates, one allele of *OTC* on the X-chromosome is silenced because of X-inactivation, and the phenotypes are varied, similar to those in male neonates with late-onset OTCD.

No reports are available on the management of asymptomatic male patients with late-onset OTCD who are diagnosed and treated early in the neonatal period. Here, we report a large pedigree involving late-onset OTCD in which male neonates were definitively diagnosed using gene and amino acid analysis soon after birth. The patients received arginine supplementation treatment, and they grew and developed well.

Correspondence: Hiroshi Mitsubuchi, MD PhD, Department of Pediatrics, Kumamoto University Graduate School of Medical Sciences, 1-1-1 Honjo, Kumamoto 860-8556, Japan. Email: mitsubuchi @fc.kuh.kumamoto-u.ac.jp

Received 27 June 2013; revised 3 July 2014; accepted 14 July 2014.

Case report

The proband (IV-19) was admitted (Fig. 1) at 7 years of age for hyperammonemia (blood ammonia, 178 μmol/L) with vomiting and coma, although his condition stabilized with intensive care. After OTC activity in the liver tissue was found to be low (0.047 μmol/mg protein/min; control, 0.96), he was prescribed oral arginine supplementation for long-term management. Because of the proband's low OTC activity, his sister underwent allopurinol test and was found to be a heterozygous carrier of the OTCD mutation. After the proband was discharged from hospital, a c.221G>A, p.Lys221Lys mutation was detected at the 3′ end of exon 6 of *OTC* in his X-chromosome, which was judged to cause a splicing abnormality (Fig. 2). Based on this finding, the proband was diagnosed with late-onset OTCD. Eight years later, the proband's 4-year-old male second cousin (IV-18) died of hyperammonemic coma (blood ammonia, >1953 μmol/L). The diagnosis of late-onset OTCD was confirmed on detection of the same mutation in the cousin.[6]Subsequently, the sisters of the deceased cousin were examined for the mutation, and were found to be heterozygous carriers.

Eleven years after the onset of OTCD in the proband, a female carrier (IV-16) of OTCD – the sister of patient IV-18 – requested genetic counseling for pregnancy and delivery, considering the family history of late-onset OTCD. The information provided included postpartum management of neonates. The carrier was informed that the mutation would manifest late-onset OTCD phenotypes in the family. We planned to analyze metabolic profiles from the early neonatal period for all this family's children born henceforth and to sequence the *OTC* gene simultaneously in these individuals.

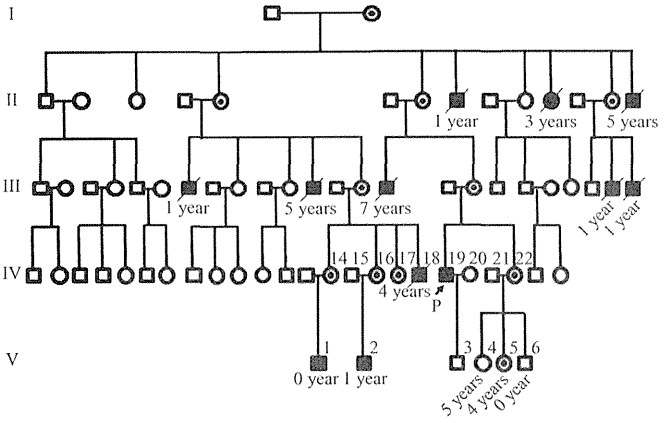

Fig. 1 Pedigree of patients with late-onset ornithine transcarbamylase (OTC) deficiency. *OTC* was analyzed in the proband (P), his two nieces (V-4, V-5) and nephew (V-6), the male member in the fourth generation who died at age 4 years with his three sisters, and two fifth-generation male infants (V-1 and V-2). The proband's sister was diagnosed on allopurinol test.

The clinical progress of patient IV-16's pregnancy was stable; the fetus was determined to be male (V-2). The baby was born at term with an Apgar score of 8/9. He started breast-feeding with his own mother's milk on the day of birth. At 3 days of age, his blood ammonium gradually increased to 91.8 μmol/L; therefore, he was started on oral arginine supplementation. At the age of 3 days the infant's blood amino acid was analyzed before arginine supplementation was started (Table 1), and slight changes were observed: glutamine increased, citrulline and arginine decreased, and ornithine remained unchanged. Urinary orotic acid was not elevated at hyperammonemia onset. The patient's blood ammonium decreased to the normal range within a few days after initiation of arginine supplementation.

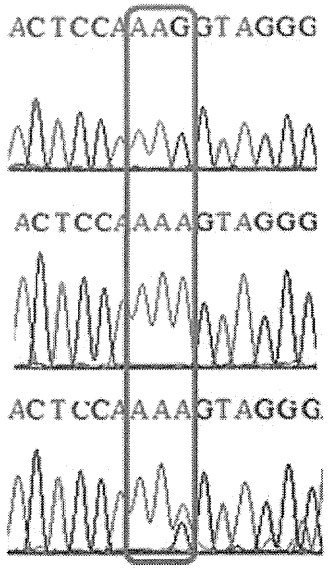

Fig. 2 *OTC* gene analysis at codon 221, exon 6 in the X-chromosome. Red line, codon 221 at the 3′ end of exon 6 of the X-chromosome. Wild-type codon 221 (upper) is AAG. In the patients, the mutation from AAG to AAA was either homozygous (middle) or heterozygous (lower).

Gene analysis conducted in the early neonatal period showed that patient V-2 (Fig. 1) had the aforementioned mutation in exon 6 of *OTC*. After 1 year, the same metabolic profile management and gene analysis were conducted in the early neonatal period for two other male infants – namely, V-1 and V-6 – and the genetic profile of the female infants – namely, V-4 and V-5 – was examined. Patient V-1 had the mutation and showed changes in blood amino acid level, similar to patient V-2. Therefore, for metabolic balance management, he also received arginine supplementation. At this time, however, the increase in urinary orotic acid could not be detected in this patient. Because low citrulline was found in the early neonatal period, the patients also received citrulline supplementation. At 2 months of age, arginine and citrulline levels improved (Table 1). The male infant V-6 did not have the mutation or the abnormal blood amino acid levels. The female infant V-5, however, was diagnosed as a carrier on genetic analysis (Fig. 1). Her blood amino acid levels were not analyzed at that time, given that she looked asymptomatic and stable.

Patients V-1 and V-2 grew well, and were followed up for a period of 1 and 2 years, respectively. They developed without decompensation during childhood and had a normal balance of arginine and glutamine, although the citrulline in V-2 remained slightly low (Table 1).

Discussion

Late-onset OTCD involves various hyperammonemic symptoms of wide-ranging severity, including vomiting, lethargy, seizures, coma, and even death. The aim of long-term management of late-onset OTCD is to improve the survival rate and maintain normal blood ammonia level and growth.[5] Patients may develop mental retardation or die when the blood ammonia exceeds 360 μmol/L.[7] Arginine is an indispensable amino acid for increasing the production of urea. Its shortage causes rapid onset of life-threatening hyperammonemia. It was found that long-term management with arginine supplementation significantly decreased the frequency of hyperammonemic attacks and improved physical growth of OTCD patients.[8] Given that it is important to prevent hyperammonemic decompensation, metabolic balance must be achieved before onset.

In addition to examination of metabolic profile, such as blood amino acid and ammonia and urinary orotic acid, mutation analysis is a very useful diagnostic tool in cases of female heterozygotes. While it is possible to predict the severity of the disease for hemizygous male patients from gene mutation analysis and/or residual enzyme activities in liver tissue, clinical features vary among female heterozygotes, because their condition is affected by not only random X-chromosome inactivation but also factors such as puberty, pregnancy, labor, delivery, and the environment.[5]We planned genetic examination of V-4, V-5 for the following reasons. First, some female carriers develop symptoms in childhood. Additionally, arginine supplementation is effective not only in the chronic phase but also in the acute phase in this large family. Therefore it is important for these female family

Table 1 Blood amino acid analysis

Amino acid (nmol/mL)	V-1†		V-2†		V-6	Reference range
	At birth	After intervention‡	At birth	After intervention‡		
Asp	3.9	6.3	6.4	11.3	3.8	≤2.4
Asn	66.3	51.6	77.4	55.3	42.2	44.7–96.8
Glu	60.2	82.6	60.1	88.3	72.9	12.6–62.5
Gln	1007.2	664.6	827	696	758.9	422.1–703.8
Ala	274.5	292.8	355.5	308	205.3	208.7–522.7
Pro	234.5	160.3	323.5	242	154.1	77.8–272.7
Val	140.1	148.8	181.2	171	117.6	147.8–307
Ile	42.8	60.6	55	73	37	43–112.8
Leu	96.4	110.1	119.1	125	82	76.6–171.3
Ornithine	68.6	115.9	91.3	103	48.4	31.3–104.7
Citrulline	4.6	22.5	5.6	7.1	20.7	17.1–42.6
Arg	38.4	121.4	37.1	108	47.6	53.6–133.6

†Ornithine transcarbamylase deficiency. ‡At 2 months of age.

members to be diagnosed as a carrier or not. Second, genetic examination is more useful than profiling of plasma amino acids for definite diagnosis of female carriers, especially when the specific mutation has been determined to be the cause of OTCD in a family. We discussed the benefits and disadvantages of the genetic diagnosis of V-4, V-5 as a definite diagnosis. Consequently, we admitted the opinion of their parents that V-4 and V-5 were examined by gene mutation analysis. The two male neonatal patients (V-1 and V-2) in this study did not have increased urinary orotic acid on gas chromatography/mass spectrometry in the early neonatal period. Blood amino acid analysis conducted simultaneously, however, provided significant information: a slight increase in glutamine and a decrease in arginine and citrulline. The high nitrogen burden causes a blood amino acid imbalance in OTCD patients. Previously, glutamine was found to have the most prominent reaction.[5]Early treatment of patient V-2 improved the amino acid balance and prevented metabolic crisis, because management was started when blood amino acid was only slightly changed and plasma ammonia was only slightly elevated. Moreover, because citrulline remained slightly low despite arginine supplementation (V-2), citrulline supplementation plays an important role in the amino acid balance (V-1). In this large pedigree, most female carriers did not show any symptoms of OTCD. Therefore, intervention with supplementation may not be necessary for some female carriers, but the outcome of these female carriers is difficult to predict. For this reason, as part of the management of the female carrier, regular profiling of her plasma amino acids is planned.

When a patient is first diagnosed with hereditary late-onset OTCD in a family, monitoring of male patients and female carriers is important so that genetic counseling can be provided, whereby it may be possible to prevent onset among family members.[4]It is vital to diagnose the condition during the asymptomatic period to prevent severe symptoms at onset.

Conclusion

Late-onset OTCD was diagnosed in asymptomatic male patients and their condition managed to prevent decompensation via treatment with arginine supplementation during the early neonatal period. The patients grew and developed well without any metabolic crises. Thus, blood amino acid and gene analysis during the early neonatal period is important for the management of families with late-onset OTCD.

References

1 Lindgren V, de Martinville B, Horwich AL, Rosenberg LE, Francke U. Human ornithine transcarbamylase locus mapped to band Xp21.1 near the Duchenne muscular dystrophy locus. *Science* 1984; **226**: 698–700.

2 Balasubramaniam S, Rudduck C, Bennetts B, Peters G, Wilcken B, Ellaway C. Contiguous gene deletion syndrome in a female with ornithine transcarbamylase deficiency. *Mol. Genet. Metab.* 2010; **99**: 34–41.

3 McCullough BA, Yudkoff M, Batshaw ML, Wilson JM, Raper SE, Tuchman M. Genotype spectrum of ornithine transcarbamylase deficiency: Correlation with the clinical and biochemical phenotype. *Am. J. Med. Genet.* 2000; **93**: 313–19.

4 Ausems MG, Bakker E, Berger R *et al.* Asymptomatic and late-onset ornithine transcarbamylase deficiency caused by a A208T mutation: Clinical, biochemical and DNA analyses in a four-generation family. *Am. J. Med. Genet.* 1997; **68**: 236–9.

5 Berry GT, Steiner RD. Long-term management of patients with urea cycle disorders. *J. Pediatr.* 2001; **138**: S56–S61.

6 Shimadzu M, Matsumoto H, Matsuura T *et al.* Ten novel mutations of the ornithine transcarbamylase (OTC) gene in OTC deficiency. *Hum. Mutat.* 1998; Suppl. 1: S5–S7.

7 Kido J, Nakamura K, Mitsubuchi H *et al.* Long-term outcome and intervention of urea cycle disorders in Japan. *J. Inherit. Metab. Dis.* 2012; **35**: 777–85.

8 Nagasaka H, Yorifuji T, Murayama K *et al.* Effects of arginine treatment on nutrition, growth and urea cycle function in seven Japanese boys with late-onset ornithine transcarbamylase deficiency. *Eur. J. Pediatr.* 2006; **165**: 618–24.