Table 1. Transplantation efficacy demonstrated by the percentage of human nuclei in the transplanted muscles.

		KhES1	Average	SD		253G4	Average	SD	
*	4 weeks #1–1	15.1%				17.3%			
*	4 weeks #1–2	13.0%				18.6%			
*	4 weeks #2–1	8.1%				7.4%			
*	4 weeks #2–2	16.1%				6.1%			
**	4 weeks #3–1	12.6%				7.2%			
**	4 weeks #3–2	15.8%	13.5%	2.7%		10.3%	11.1%	5.0%	
*	12 weeks #1–1	26.5%				11.7%			
*	12 weeks #1–2	29.3%				10.3%			
*	12 weeks #2–1	13.5%				†			
*	12 weeks #2–2	16.5%				9.5%			
**	12 weeks #3–1	17.6%				11.7%			
**	12 weeks #3–2	ND	20.7%	6.1%		11.5%	10.9%	0.9%	
*	24 weeks #1–1	16.2%				12.2%			
*	24 weeks #1–2	13.9%				9.3%			
*	24 weeks #2–1	16.3%				9.6%			
*	24 weeks #2–2	20.9%				8.1%			
**	24 weeks #3–1	16.3%				11.4%			
**	24 weeks #3–2	19.6%	17.2%	2.3%		11.1%	10.3%	1.4%	
**	4 weeks #4–1	12.3%				6.9%			
**	4 weeks #4–2	14.6%				11.9%			
**	4 weeks #4–3	10.4%	12.5%	1.7%		15.7%	11.5%	3.6%	
**	3+1 weeks #4–1	30.4%				ND			
**	3+1 weeks #4–2	25.8%				13.0%			
**	3+1 weeks #4–3	40.7%	32.3%	6.3%	$P<0.05$	24.7%	18.9%	5.8%	NS

The number of cells of human origin was divided by the number of total nuclei stained by DAPI. The result also appears in Figures 4C and 4D. SD: standard deviation. ND: not detected. *, ** indicate the cell numbers transplanted at the site: * 5.0×10^5 cells/site, ** 1.0×10^5 cells/site. †: One mouse transplanted with 253G4-derived cells died accidentally before analysis. NS: statistically not significant.

doi:10.1371/journal.pone.0051638.t001

after transplantation and subjected to immunohistological analyses. In some experiments, the transplanted mice were re-injured by CTX injection at 3 weeks after transplantation and analyzed 1 week later to determine the regenerative capacity of the engrafted cells. For the immunohistological assay, the isolated muscles were frozen in isopentane cooled in liquid nitrogen. The frozen specimens were sectioned with a cryostat (CM1850; Leica Microsystems, Wetzlar, Germany) and analyzed as described above. All animal handling procedures were followed according to the Guide for the Care and Use of Laboratory Animals published by the U.S. National Institutes of Health (NIH Publication No. 85–23, revised 1996) and the Guidelines of the Animal Research Committee of the Graduate School of Medicine, Kyoto University. This work was approved by the Animal Research Committee of the Graduate School of Medicine, Kyoto University. Mice were sacrificed by cervical dislocation. All painful procedures including cervical dislocation were performed under anesthesia. In immunofluorescence analyses, the percentage of human nuclei in total nuclei was counted.

Statistics

Statistical analyses were performed using the unpaired Student's *t* test, and a value of $P<0.05$ was considered to be statistically significant.

Results

Myogenic induction by human ES and iPS cells in EB outgrowth culture

To develop an efficient differentiation protocol for inducing skeletal myogenesis from human ES and iPS cells, we adapted our previously established EB-based method that induces myogenic differentiation from murine ES and iPS cells [31,32]. EBs were formed by suspension in hESM for 7 days and then plated onto 0.1% gelatin-coated tissue culture plates. Attached EBs were cultured in serum-free medium for an additional 14 days and then in differentiation medium containing 10% FCS and 5% HS until day 112 (7+14+91) of differentiation. When EBs attached to the plates, the cells migrating out from the EBs formed a single layer, which we termed an EB-outgrowth (EB-OG). Immunostaining for PAX3 and PAX7, which mark early myogenic progenitors, showed that clusters of PAX3- and PAX7-positive cells were randomly distributed at day 21 (Figure 2A). Skeletal myosin (MYH)-positive multinucleated myofibers had appeared within most of the attached EBs at day 63 (Figure 2C) and were randomly distributed in the crowded EB-OG. Some attached EBs simultaneously contained TUJ1-positive neural cells (Figure 2B) and colonies of beating cardiomyocytes (Figure 2B), as previously observed in studies on murine EB differentiation [20,21].

We then analyzed the expression of genes associated with skeletal myogenesis, including *PAX3*, *PAX7*, desmin (*DES*), and the

muscle regulatory factors *MYF5*, *MYOD*, and *MYOG* (Figure 2D). The expression of *PAX3* was first detected on day 7, increased at day 21 (7+14), and decreased thereafter. *PAX7*, which is essential for the specification of muscle progenitors [3-5,35,36], was detected at day 21 (7+14), but its expression was highly variable. *MYOD* and *DES* were first detected at day 21 (7+14), and their expression increased thereafter. *MYF5* was detected at day 7, and its expression was sustained throughout differentiation. The expression of *MYOG* was not detected until day 49 (7+42) but increased thereafter. In contrast, the pluripotent stem marker genes *OCT3/4*, *NANOG*, and *REX1* were down regulated during differentiation. Thus, the hierarchical expression of myogenic genes in this system was similar to the patterns observed during murine embryogenesis [37] and to those of the EB differentiation of murine ES and iPS cells [20,31,32].

To verify whether this protocol was applicable to hiPS cells, we used 4 iPS cell lines expressing either 4 (*OCT3/4*, *SOX2*, *NANOG*, and *MYC*) [15] or 3 (lacking *MYC*) retrovirally introduced transcription factors [16]. The EBs from all examined hiPS cell lines contained MYH-positive myofibers, confirming the myogenic potential of hiPS cells (Figure 2E). The expression pattern of myogenic markers was quite similar to that of ES cells (Figure 2F), although the differentiation efficacy was different among the cell lines.

Thus, EB-OG successfully induced myogenesis from both human ES and iPS cells, thereby simulating the temporal patterning of embryogenesis. However, the EB-OGs were so heterogeneous that we could not distinguish the differentiated muscle lineage cells from other lineage cells. Most of the skeletal muscle lineage cells appeared as clusters in crowded EB-OGs and never appeared as a homogeneous population.

Selective expansion of myogenic mesenchymal cells from dissociated EB outgrowth culture

Because mature myofibers could not clearly be distinguished in the overcrowded EB-OG cells, we decided to purify the myogenic cells from the other EB-OG cells. We hypothesized that the separation of the myogenic cells might allow efficient myogenesis if these other lineage cells were disturbing skeletal myogenesis in the EBs. Hence, the differentiated EB-OG cells were dissociated into single cells and re-cultured on collagen type I-coated plates on day 21 (7+14) when the clusters of PAX3- and PAX7-positive myogenic progenitors were detected. When the dissociated cells were replaced onto collagen type I-coated plates, cells with a uniform spindle, fibroblastic morphology developed selectively during the next 4 weeks (Figure 3A). FCM analysis showed that the generated cells exclusively expressed the mesenchymal markers CD73, CD105, CD166, and CD29, whereas CD34 and CD106 were hardly detected. CD56 (NCAM; neural cell adhesion molecule) was also highly expressed (Figure 3B, Figure S1). A low-density culture of dissociated cells on collagen type I was crucial for selective development of myogenic cells; neural and cardiac cells were not detected in this dissociation protocol (Figure S2).

Quantification of the expression of the myogenic progenitor markers MYOD, MYF5, and PAX7 increased during cell proliferation and peaked around day 49 (7+14+28) in this dissociation protocol. Since serum deprivation is known to enhance further maturation of myogenic cells, we switched from serum-containing SkIM to serum-free ITS medium to induce terminal muscular differentiation. Following serum deprivation, the proportion of myogenic progenitor marker-positive nuclei decreased immediately, whereas that of MYOG-positive nuclei dramatically increased (Figure 3C–3E). The increased expression of skeletal myosin-positive fibers and the increased prevalence of MYOG-positive nuclei were maintained up to 3 weeks (Figure 3C, 3D).

Both hES (KhES1) and hiPS (253G4) cell-derived cells fused together and produced multinucleated myofibers that were positive for skeletal myosin and dystrophin (Figure 3F, 3G). These myofibers occasionally showed spontaneous contraction, indicating the maturation of skeletal muscle cells. Hence, the results showed that this multi-step culture method selectively isolated myogenic mesenchymal population from dissociated EB culture, and that these myogenic mesenchymal cell populations could terminally differentiate into functionally mature myofibers *in vitro*.

Transplantation in immunodeficient NOG mice

Next, we examined the *in vivo* myogenic potential of human ES and iPS cell-derived myogenic mesenchymal cells by transplanting them into the injured muscles of NOG mice. As myogenic progenitor markers were strongly expressed at day 49 (7+14+28), the cells were trypsinized again and used as donor cells. To avoid immunological rejection, we used immunodeficient NOG mice as the hosts.

Four weeks after transplant, human-specific lamin A/C-positive nuclei were detected in TA muscles (Figure 4A). Human lamin A/C-positive and lamin A/C-negative nuclei were co-localized in the same muscle fascicle and had formed hybrid myotubes (Table 1, Figure 4A). Furthermore, the detection of human-specific merosin (laminin alpha 2) proved that the transplanted cells had not only survived within muscle fibers but had also produced human protein around the muscle fibers into which they had integrated (Figure 4B). Thus, the myogenic mesenchymal cells derived here engrafted and functioned to repair damaged muscles.

We also investigated the log-term outcome of the myogenic mesenchymal cell transplant. Human lamin A/C-positive nuclei were continuously observed in the damaged muscles until 24 weeks after transplantation (Table 1, Figure 4C). Moreover, the presence of typical central nuclei of human origin strongly suggested that transplanted cell-related muscular regeneration occurred even at 24 weeks after transplantation (Figure 4E). Although most human cells were detected within muscle fascicles, some of them resided within the lamina rara, which is consistent with the location of satellite cells (Figure 4E). These human nuclei were also positive for PAX7, confirming the contribution of the transplanted cells to the satellite cell fraction (Figure 4F). Under macroscopic and histological examination, teratoma formation was not observed up to 24 weeks.

Finally, to assess whether the engrafted myogenic mesenchymal cells could proliferate in response to subsequent damage, we employed a second injury model, as previously reported [31,32]. Three weeks after transplantation, CTX was directly injected into the transplanted muscles. The proportion of human nuclei was higher in re-injured mice than that in non-re-injured mice at 4 weeks after transplantation in the experiment using hES cells, whereas no statistical significance was observed in the hiPS cells. (Table 1, Figure 4D).

Collectively, our results indicated that human ES and iPS cell-derived myogenic mesenchymal cells engrafted effectively, integrated into the satellite cell fraction, and contributed to the long-term reconstruction of the damaged muscle tissue without exhaustion.

Discussion

Recent remarkable advances in human ES and iPS cell technology have increased the demand for the clinical application

of these cells and their use in basic research. The excellent work by Barberi et al. clearly demonstrated the myogenic capacity of hES cells both *in vitro* and *in vivo* [26]; however, their strategy to obtain myogenic populations included repetitive cell sorting procedures, which would hamper practical use. The other approach to induce myogenesis from hES cells is to form EBs; however, the EB-based methods inevitably contain cells from other lineages. Ideally, homogeneous populations of the desired cell types should be generated using a simple method that does not involve complicated procedures.

In this report, we modified our murine ES and iPS cell culture system to develop a new EB-based culture method that can successfully induce mature skeletal muscle cells from hES cells. An advantage of this system is that selective expansion of myogenic mesenchymal cells can be accomplished by simple manipulation.

Moreover, this culture system is applicable to hiPS cells. Darabi and his colleagues recently reported successful skeletal muscle induction from hiPS cells by *PAX7* overexpression [30]. However, skeletal myogenesis from hiPS cells without genetic manipulation has not been previously reported. Our results facilitate the further use of hiPS cells in muscle regenerative medicine, which extends the possibilities of hES technology to clinical applications, such as patient-oriented research and autologous transplantation [10,17–19].

We did not employ cell sorting in this method. Several reports have demonstrated that multipotent mesenchymal cells can be generated by long term EB culturing or repetitive passaging [38,39]. Meanwhile, some subpopulations of mesenchymal stem cells (MSCs) are considered to possess myogenic potential. We re-plated dissociated EBs on several coating materials (Figure S2), and derived the immunophenotypically homogeneous mesenchymal cell population. Surface marker analyses demonstrated that the generated cells were uniformly positive for mesenchymal stem cell (MSC)-related surface antigens, including CD73, CD105, CD166, and CD29. However, the generated cells were different from MSCs because of the high expression of CD56. CD56 is known to be expressed in embryonic skeletal muscles and satellite cells [9,11], and was used for the prospective isolation of myogenic progenitors from hES cell-derived MSCs in a previous report [26]. These reports also supported the existence of the myogenic mesenchymal cells generated here. Further study will be necessary to characterize the derived cells in terms of their multipotency, including their osteogenic, chondrogenic, and adipogenic activities.

The homogenous mesenchymal population with myogenic potential developed in our culture system; these can be used for both *in vitro* and *in vivo* studies. By depleting serum, the transition from myogenic progenitors to mature myotubes can easily be observed during culture. In combination with the recent advancements in creating disease-specific hiPS cell lines, this culture system will be a powerful experimental tool for disease studies and the screening of drugs for various types of muscle diseases, both of which have been difficult to conduct because of the significant challenge in obtaining sufficient materials.

Furthermore, we demonstrated the *in vivo* myogenic capacity of derived cells. When transplanted in NOG mice, the engrafted cells contributed to satellite cell fraction and reconstructed the injured muscles up to at least 6 months without rejection. Moreover, the integrated cells showed a more robust proliferation in response to subsequent injury. Myogenic progenitors with such re-proliferation capacity *in vivo* have not been reported in previous studies [22–26,29,30], and can be a feasible cell source for muscle regenerative medicine, since long-lasting muscle repair will be required for patients with genetic muscle diseases. The risk of teratoma formation was eliminated by selective culture conditions, which is also suitable for clinical settings.

The underlying mechanisms of the selective proliferation of myogenic cells in our culture system remain unresolved. The initial EB-OG culture simultaneously contained neuronal and cardiac cells in addition to muscle lineage cells. Given that re-plating the dissociated cells at low cell density was a crucial step for effective myogenic differentiation, direct cell-to-cell contact or soluble factors from other cells or cell lineages might have negatively affected the selective development of muscle lineage cells. Otherwise, simple cell selection through the capacity to adhere to collagen type I-coated plates or differences in proliferation potential enriched the myogenic mesenchymal population in the residual cells (Figure S2).

In conclusion, we have reported a novel culture system that induces homogenous myogenic mesenchymal cells from both human ES and iPS cells. This method is a modification of the widely used EB method and is accomplished using a simple procedure. To date, there still are few reports describing human ES and iPS cell myogenic differentiation. This system encourages further utilization of human ES and iPS cells in skeletal muscle research, which could not otherwise be conducted in humans, and will provide a promising therapeutic approach for various muscle diseases.

Supporting Information

Figure S1 Surface marker changes with or without EB dissociation. Surface marker of the differentiating EBs without dissociation (left) and the cells after EB dissociation and re-plating (right). The mesenchymal stem cell marker, CD73, increased during continuous EB culture, but much homogeneous population was obtained at days 7+14+28 by EB dissociation. Representative histograms are shown. The experiments were conducted 3 times, and mean scores and standard deviations were calculated. SD: standard deviation.
(TIF)

Figure S2 Elimination of other lineage cells by low-cell-density culture on collagen type I. The results demonstrated that low-density culture on collagen type I eliminated neural and cardiac cells. (A) RT-PCR for TUBB3, TNNI3, and Desmin expression at different cell densities. RT-PCR result indicated contamination of neural and cardiac cells at high-density culture at variable levels. Laminin and poly-D-lysin preferentially captured many more neurogenic cells, whereas others did not. At low-density culture, 0.1% gelatin, laminin, and poly-D-lysin could not support cell proliferation. A heterogeneous non-mesenchymal population was obtained on collagen type IV. When cultured on Matrigel®, the results obtained were similar to that obtained using collagen type I. (B) TUJ1- or GFAP-positive cells observed at high-cell-density culture on collagen type I. At high cell density, the cells reached confluence at several days after re-plating. Upper: 7 days after re-plating. Lower: 28 days after re-plating. No neural cells were detected if cultured at a low-density.
(TIF)

Method S1 Selection of myogenic cells by low-cell-density culture on collagen type I EBs were dissociated at day 21 (7+14) of differentiation and re-plated onto different coating materials. We used 0.1% gelatin, laminin, poly-D-lysin, collagen type I, collagen type IV, and Matrigel® (All from BD Bioscience) for the experiments. The cells from dissociated EBs were seeded at low- or high-cell-density (3,000 and 30,000 cells/cm^2, respectively) and cultured for up to 28 days.

Cells were analyzed at day 7 and 28 after re-plating by RT-PCR or immunostaining.
(DOCX)

Table S1 The primer settings used for RT-PCR.
(TIF)

Acknowledgments

We wish to thank Prof. Norio Nakatsuji (Kyoto University, Kyoto, Japan) and Prof. Shinya Yamanaka (Kyoto University, Kyoto, Japan) for providing the ES and the iPS cell lines, respectively.

Author Contributions

Conceived and designed the experiments: TA TK TN TH. Performed the experiments: TA YM HC. Analyzed the data: TA TK YM HC AN KU TN TH. Wrote the paper: TA TK KU TH.

References

1. Wallace GQ, McNally EM (2009) Mechanisms of muscle degeneration, regeneration, and repair in the muscular dystrophies. Annu Rev Physiol 71: 37–57.
2. Mauro A (1961) Satellite cell of skeletal muscle fibers. J Biophys Biochem Cytol 9: 493–495.
3. Wang YX, Rudnicki MA (2012) Satellite cells, the engines of muscle repair. Nat Rev Mol Cell Biol 13: 127–133.
4. Scharner J, Zammit PS (2011) The muscle satellite cell at 50: the formative years. Skeletal Muscle 1: 28.
5. Cerletti M, Shadrach J, Jurga S, Sherwood R, Wagers A (2008) Regulation and Function of Skeletal Muscle Stem Cells. Cold Spring Harb Symp Quant Biol 73: 317–322.
6. Jejurikar SS, Kuzon WM Jr (2003) Satellite cell depletion in degenerative skeletal muscle. Apoptosis 8: 573–578.
7. Bulfield G, Siller WG, Wight PA, Moore KJ (1984) X chromosome-linked muscular dystrophy (mdx) in the mouse. Proc Natl Acad Sci USA 81: 1189–1192.
8. Dangain J, Vrbova G (1984) Muscle development in mdx mutant mice. Muscle Nerve 7: 700–704.
9. Péault B, Rudnicki M, Torrente Y, Cossu G, Tremblay JP, et al. (2007) Stem and Progenitor Cells in Skeletal Muscle Development, Maintenance, and Therapy. Mol Ther 15: 867–877.
10. Meregalli M, Farini A, Parolini D, Maciotta S, Torrente Y (2010) Stem cell therapies to treat muscular dystrophy: progress to date. BioDrugs 24: 237–247.
11. Tedesco FS, Dellavalle A, Diaz-Manera J, Messina G, Cossu G (2010) Repairing skeletal muscle: regenerative potential of skeletal muscle stem cells. J Clin Invest 120: 11–19.
12. Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. Nature 292: 154–156.
13. Thomson JA (1998) Embryonic Stem Cell Lines Derived from Human Blastocysts. Science 282: 1145–1147.
14. Takahashi K, Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126: 663–676.
15. Takahashi K, Tanabe K, Ohnuki M, Narita M, Ichisaka T, et al. (2007) Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell 131: 861–872.
16. Nakagawa M, Koyanagi M, Tanabe K, Takahashi K, Ichisaka T, et al. (2008) Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat Biotechnol 26: 101–106.
17. Grskovic M, Javaherian A, Strulovici B, Daley GQ (2011) Induced pluripotent stem cells – opportunities for disease modelling and drug discovery. Nat Rev Drug Discov 10: 915–929.
18. Bellin M, Marchetto MC, Gage FH, Mummery CL (2012) Induced pluripotent stem cells: the new patient? Nat Rev Mol Cell Biol. doi:10.1038/nrm3448.
19. Salani S, Donadoni C, Rizzo F, Bresolin N, Comi GP, et al. (2012) Generation of skeletal muscle cells from embryonic and induced pluripotent stem cells as an in vitro model and for therapy of muscular dystrophies. J Cellular Mol Med 16: 1353–1364.
20. Rohwedel J, Maltsev J, Bober E, Arnold HH, Hescheler J, et al. (1994) Muscle cell differentiation of embryonic stem cells reflects myogenesis in vivo: Developmentally regulated expression of myogenic determination genes and functional expression of ionic currents. Dev Biol 184: 87–101.
21. Guan K, Rohwedel J, Wobus AM (1999) Embryonic stem cell differentiation models: cardiogenesis, myogenesis, neurogenesis, epithelial and vascular smooth muscle cell differentiation in vitro. Cytotechnology 30: 211–226.
22. Zheng JK, Wang Y, Karandikar A, Wang Q, Gai H, et al. (2006) Skeletal myogenesis by human embryonic stem cells. Cell Res 16: 713–722.
23. Mahmood A, Harkness L, Schrøder HD, Abdallah BM, Kassem M (2010) Enhanced differentiation of human embryonic stem cells to mesenchymal progenitors by inhibition of TGF-beta/activin/nodal signaling using SB-431542. J Bone Miner Res 25: 1216–1233.
24. Ryan T, Liu J, Chu A, Wang L, Blais A, et al. (2011) Retinoic Acid Enhances Skeletal Myogenesis in Human Embryonic Stem Cells by Expanding the Premyogenic Progenitor Population. Stem Cell Rev 8: 482–493.
25. Barberi T, Willis LM, Socci ND, Studer L (2005) Derivation of multipotent mesenchymal precursors from human embryonic stem cells. PLoS Med 2: e161.
26. Barberi T, Bradbury M, Dincer Z, Panagiotakos G, Socci ND, et al. (2007) Derivation of engraftable skeletal myoblasts from human embryonic stem cells. Nat Med 13: 642–648.
27. Tapscott SJ, Davis RL, Thayer MJ, Cheng PF, Weintraub H, et al. (1988) MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science 242: 405–411.
28. Ozasa S, Kimura S, Ito K, Ueno H, Ikezawa M, et al. (2007) Efficient conversion of ES cells into myogenic lineage using the gene-inducible system. Biochem Biophys Res Commun 357: 957–963.
29. Darabi R, Gehlbach K, Bachoo RM, Kamath S, Osawa M, et al. (2008) Functional skeletal muscle regeneration from differentiating embryonic stem cells. Nat Med 14: 134–143.
30. Darabi R, Arpke RW, Irion S, Dimos JT, Grskovic M, et al. (2012) Human ES- and iPS-Derived Myogenic Progenitors Restore DYSTROPHIN and Improve Contractility upon Transplantation in Dystrophic Mice. Cell Stem Cell 10: 610–619.
31. Chang H, Yoshimoto M, Umeda K, Iwasa T, Mizuno Y, et al. (2009) Generation of transplantable, functional satellite-like cells from mouse embryonic stem cells. FASEB J 23: 1907–1919.
32. Mizuno Y, Chang H, Umeda K, Niwa A, Iwasa T, et al. (2010) Generation of skeletal muscle stem/progenitor cells from murine induced pluripotent stem cells. FASEB J 24: 2245–2253.
33. Suemori H, Yasuchika K, Hasegawa K, Fujioka T, Tsuneyoshi N, et al. (2006) Efficient establishment of human embryonic stem cell lines and long-term maintenance with stable karyotype by enzymatic bulk passage. Biochem Biophys Res Commun 345: 926–932.
34. Ito M, Hiramatsu H, Kobayashi K, Suzue K, Kawahata M, et al. (2002) NOD/SCID/gamma(c) (null) mouse: an excellent recipient mouse model for engraftment of human cells. Blood 100: 3175–3182.
35. Buckingham M, Relaix F (2007) The role of Pax genes in the development of tissues and organs: Pax3 and Pax7 regulate muscle progenitor cell functions. Annu Rev Cell Dev Biol 23: 645–673.
36. Bismuth K, Relaix F (2010) Genetic regulation of skeletal muscle development. Exp Cell Res 316: 3081–3086.
37. Bentzinger CF, Wang YX, Rudnicki MA (2012) Building muscle: molecular regulation of myogenesis. Cold Spring Harb Perspect Biol 4.
38. Olivier EN, Rybicki AC, Bouhassira EE (2006) Differentiation of human embryonic stem cells into bipotent mesenchymal stem cells. Stem Cells 24: 1914–1922.
39. Trivedi P, Hematti P (2008) Derivation and immunological characterization of mesenchymal stromal cells from human embryonic stem cells. Experimental Hematology 36: 350–359.

DNA Research 19, 143–152, (2012)
Advance Access Publication on 24 January 2012
doi:10.1093/dnares/dsr047

Detection of Base Substitution-Type Somatic Mosaicism of the *NLRP3* Gene with >99.9% Statistical Confidence by Massively Parallel Sequencing

Kazushi Izawa[1,†], Atsushi Hijikata[2,†], Naoko Tanaka[1], Tomoki Kawai[1], Megumu K Saito[3], Raphaela Goldbach-Mansky[4], Ivona Aksentijevich[5], Takahiro Yasumi[1], Tatsutoshi Nakahata[3], Toshio Heike[1], Ryuta Nishikomori[1,*], and Osamu Ohara[2,6,*]

Department of Pediatrics, Kyoto University Graduate School of Medicine, 54 Shogoin Sakyo, Kyoto 606-8507, Japan[1]; Laboratory for Immunogenomics, RIKEN Research Center for Allergy and Immunology, RIKEN Yokohama Institute, 1-7-22 Suehiro-cho Tarumi-ku, Yokohama, Kanagawa 230-0045, Japan[2]; Clinical Application Department, Center for iPS Cell Research and Application (CiRA), Kyoto University, Kyoto, Japan[3]; Translational Autoinflammatory Disease Section NIH/NIAMS, Bethesda, MD, USA[4]; The National Human Genome Research Institute, Bethesda, MD, USA[5] and Department of Human Genome Research, Kazusa DNA Research Institute, 2-6-7 Kazusa-kamatari, Kisarazu, Chiba 292-0818, Japan[6]

*To whom correspondence should be addressed. Tel. +81 75-751-3291 (R.N.); +81 438-52-3913/+81 45-503-9696 (O.O.). Fax. +81 75-752-2361 (R.N.); +81 438-52-3914/+81 45-503-9694 (O.O.).
Email: rnishiko@kuhp.kyoto-u.ac.jp (R.N.); ohara@kazusa.or.jp/oosamu@rcai.riken.jp (O.O.)

Edited by Mitsuo Oshimura
(Received 18 November 2011; accepted 26 December 2011)

Abstract

Chronic infantile neurological cutaneous and articular syndrome (CINCA), also known as neonatal-onset multisystem inflammatory disease (NOMID), is a dominantly inherited systemic autoinflammatory disease and is caused by a heterozygous germline gain-of-function mutation in the *NLRP3* gene. We recently found a high incidence of *NLRP3* somatic mosaicism in apparently mutation-negative CINCA/NOMID patients using subcloning and subsequent capillary DNA sequencing. It is important to rapidly diagnose somatic *NLRP3* mosaicism to ensure proper treatment. However, this approach requires large investments of time, cost, and labour that prevent routine genetic diagnosis of low-level somatic *NLRP3* mosaicism. We developed a routine pipeline to detect even a low-level allele of *NLRP3* with statistical significance using massively parallel DNA sequencing. To address the critical concern of discriminating a low-level allele from sequencing errors, we first constructed error rate maps of 14 polymerase chain reaction products covering the entire coding *NLRP3* exons on a Roche 454 GS-FLX sequencer from 50 control samples without mosaicism. Based on these results, we formulated a statistical confidence value for each sequence variation in each strand to discriminate sequencing errors from real genetic variation even in a low-level allele, and thereby detected base substitutions at an allele frequency as low as 1% with 99.9% or higher confidence.

Key words: next generation sequencing; mosaicism; DNA diagnosis; chronic infantile neurological cutaneous and articular syndrome

1. Introduction

Chronic infantile neurological cutaneous and articular syndrome (CINCA; MIM #607115), also known as neonatal-onset multisystem inflammatory disease (NOMID), is a dominantly inherited autoinflammatory disease that is characterized by neonatal onset and a triad of symptoms, including an urticarial-like skin rash, neurological manifestations, and arthritis/arthropathy.[1–3] Patients often experience

† These authors contributed equally to this work.

recurrent fever and systemic inflammation. CINCA/NOMID is the most severe clinical phenotype in the spectrum of cryopyrin-associated periodic syndromes (CAPS), which also include two less severe but phenotypically similar syndromes, familial cold autoinflammatory syndrome (FCAS; MIM #120100), and Muckle–Wells syndrome (MWS; MIM #191900). CAPS are caused by mutations in the *NLRP3* gene, which is a member of the Nod-like receptor (NLR) family of the innate immune system.[4–6]

Approximately 60% of CINCA/NOMID patients carry heterozygous germline missense mutations in the *NLRP3* coding region (mutation-positive patients).[7] More than 80 different disease-causing mutations have been reported to date.[8] However, the remaining clinically diagnosed CINCA/NOMID patients (~40%) show no heterozygous germline *NLRP3* mutation based on conventional DNA sequencing-based genetic analyses (mutation-negative patients). In a previous international collaborative study, we found that there was a high incidence of somatic *NLRP3* mosaicism in mutation-negative CINCA/NOMID patients worldwide.[9] The level of mosaicism ranges from 4.2 to 35.8% (median = 10.2%). Rapidly diagnosing somatic *NLRP3* mosaicism is important to ensure proper treatment. However, the conventional approach used to identify somatic mosaicism of the *NLRP3* gene is time and labour intensive due to the subcloning of the *NLRP3* exon polymerase chain reaction (PCR) products, hereafter designated as amplicons, followed by capillary DNA sequencing of more than 100 subclones for each patient. Thus, this approach is not suitable to routinely diagnose somatic mosaicism of the *NLRP3* gene and additional labour and time will be required to reliably identify somatic mosaicism that occurs at a lower rate. The aim of the present study was to establish a new method that can be used to reliably diagnose somatic mosaicism using the *NLRP3* gene as a model. Massively parallel DNA sequencing (MPS) technology is an obvious method of choice to identify somatic mosaicism, and this approach has been already reported by other groups.[10–12] However, a well-known caveat of MPS is the high rate of sequencing errors, which cannot be disregarded when identifying low-level somatic mosaicism. To our knowledge, there have been no reports of a reliable method to discriminate MPS sequencing errors from somatic mosaicism with statistical confidence.

In this study, we first analysed the patterns of sequencing errors in *NLRP3* coding exons at a single-residue resolution by MPS using a Roche 454 GS-FLX sequencer and then constructed an error rate map for each base position in the *NLRP3* exons. Based on the error rate map, we could formulate a discrimination pipeline of somatic mosaicism from sequencing errors and thereby detect new somatic mosaicism in mutation-negative CINCA/NOMID patients, whose somatic mutations were subsequently confirmed by subcloning and Sanger sequencing. This approach can also be generally used to identify low-level somatic mosaicism in other genes.

2. Patients and methods

2.1. *Patients and DNA materials*

Patients were clinically diagnosed with CAPS by their referring physicians and the *NLRP3* gene was examined using the conventional Sanger sequencing method. DNA samples were obtained from Japanese *NLRP3* somatic mosaic patients ($n = 5$) who have been previously described,[9,13] CAPS patients ($n = 5$) with heterozygous *NLRP3* mutations, and healthy donors ($n = 50$). Genomic DNA samples from mutation-negative CINCA/NOMID patients ($n = 10$) were obtained from the National Institute of Health, Bethesda, USA. To generate DNA samples with no mosaicism, we constructed a set of subcloned plasmids containing each exon and its flanking intronic regions in the *NLRP3* gene from healthy donor genomic DNA using a Topo TA cloning kit (Invitrogen, San Diego, CA, USA). The cloned plasmids containing each exon and the flanking regions were validated by Sanger sequencing. Written informed consent was obtained from all the patients and their families. The study was approved by the ethical committees of Kyoto University and Kazusa DNA Research Institute and was conducted in accordance with the Helsinki Declaration.

2.2. *MPS of* NLRP3 *gene amplicons*

Genomic DNA samples were extracted from whole blood or peripheral blood mononuclear cells as previously described. We used a two-step PCR assay and pooled sample libraries for MPS. To cover the entire *NLRP3* coding exonic regions and flanking intronic regions, 14 amplicons were designed to be as long as an average read length for a 454 GS-FLX sequencer (up to 450 bases) and then amplified from each genomic DNA sample (Fig. 1A). The sequences of the PCR primers that were used to generate these 14 amplicons are provided in Supplementary Table S1. The upper and lower amplicon-specific primer sequences were flanked by common 15-base adapter sequences (TGTAAAACGACGGCC and GGAAACAGCTATGAC for the upper and lower primers, respectively) at the 5′ end in order to fuse the primer-binding sequence for MPS in the second-step PCR. The first PCR amplifications were performed in 50-μl reactions using 30 ng of genomic DNA, 1× PrimeSTAR GXL buffer, 0.2 mM of each dNTP,

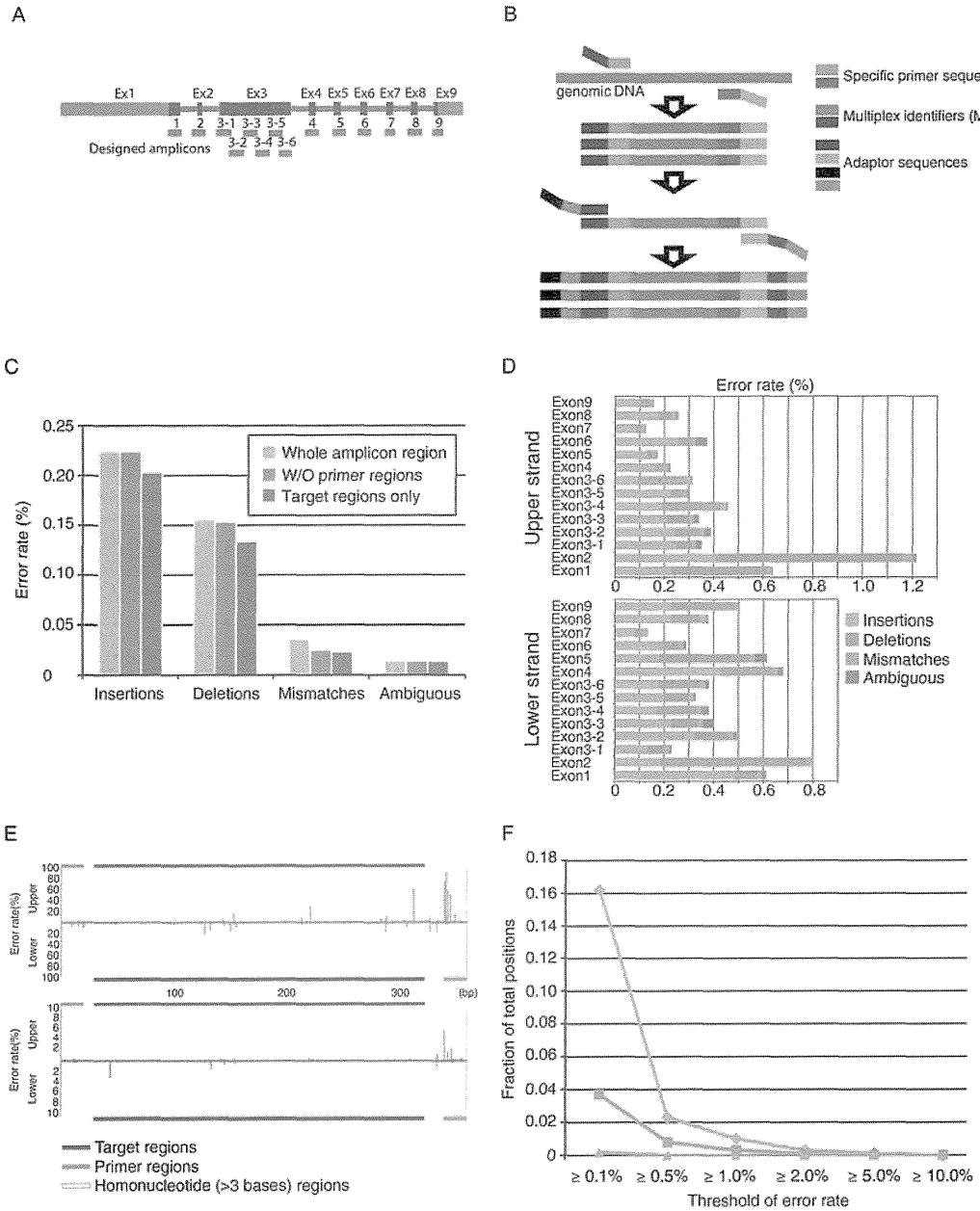

Figure 1. The amplicon analysis for *NLRP3* exons and its error rate. (A) Exon–intron structure of the *NLPR3* gene. Thick and thin rectangles depict exons and introns, respectively. Blue thick rectangles indicate the CDS region. The 14 designed amplicons (red) for nine exons are shown under the exon–intron structure. (B) Amplicon design schema. (C) Error rate for each error category in the region of entire amplicon (pale blue), that without designed primer regions (light blue), and the target regions (CDS + flanking intron; dark blue), respectively. (D) Strand-wise error rate for each amplicon. (E) Error rates along the amplicon sequence of exon 1 in each strand for insertions and deletions in the upper panel and mismatches and ambiguous base calls in the lower panel. The orange and blue lines depict the primer and target regions, respectively. The yellow shaded area depicts the homonucleotide ($n > 3$) region. The colour representation for the bars is the same as (D). (F) Co-occurrence error rate in both strands. The fraction of positions where a certain error occurred with the error rate for insertions, deletions, and mismatches. The colour representation is the same as in (D) and (E).

12.5 pmol of each forward and reverse primer, and 1.25 U of PrimeSTAR GXL DNA polymerase (Takara Bio, Shiga, Japan). The thermal cycling profile consisted of an initial denaturation step at 98°C for 1 min, followed by 28–32 cycles of 10 s denaturation at 98°C, 15 s of annealing at 60°C, and a 30 s extension at 68°C. The lengths of the PCR products ranged from 291 to 421 bp. The second PCR amplifications were performed using primers with adapter sequences at the 3′ end and Multiplex Identifier (MID) sequences at the 5′ end (Fig. 1B), which was used as a tag for each sample. The PCR reactions were performed in 50-μl volumes using 0.5 μl of the first PCR products, 1× PrimeSTAR GXL buffer, 0.2 mM of each dNTP, 12.5 pmol of each forward and reverse primer, and 1.25 U of PrimeSTAR GXL

DNA polymerase to attach the anchor sequences for MPS. The thermal cycling profile consisted of an initial denaturation step at 98°C for 20 s, followed by 20 cycles of 10 s denaturation at 98°C, 15 s of annealing at 60°C, and a 40 s extension at 68°C.

After confirming the amount and integrity of the PCR products by agarose gel electrophoresis, we mixed virtually equal amounts of the respective PCR amplicons that were generated using the same genomic DNA and applied the samples to a 454 Genome Sequencer (GS)-FLX system (Roche Diagnostics Corp., USA). All amplicons were amplified by emPCR and sequenced together in a multiplex fashion. MPS on this platform was performed as instructed by Roche. The sequencing reads from each of the pooled libraries were identified by their MID tags.

2.3. *Sequence data analysis*

The sequence read data were generated using GS RunProcessor ver.2.5.3 with default settings. Reads were sorted according to the MID tag sequences and were mapped to the reference amplicon sequences using the BLAT program[14] with the '-fine' option. In order to identify positions where the bases in a read differed from those in the reference sequence, each read was aligned to its reference sequence with the dpAlign module in the BioPerl package (http://www.bioperl.org/). The 454 pyrosequencing-related errors were categorized as insertions, deletions, mismatches, or ambiguous base calls. When aligning sequences, insertions/deletions are allocated based on the sequence context and strand orientation. To eliminate alignment artefacts due to insertion/deletion positions, the lower strand reads were converted to the reverse complement sequence, i.e. keeping the same strandness as the upper strand reads, when aligned with the reference sequence. A sequence error was defined as discordance in an equivalent position between the reference and control (from the 49 healthy individuals and a cloned plasmid vector). The error rate for a specified category was defined as the number of errors divided by the total number of bases in a read. The error rates of a base position on each strand were calculated from 50 control samples.

2.4. *Confirmation of somatic mosaicism of the* NLRP3 *gene by subcloning and subsequent capillary DNA sequencing*

To confirm the somatic mutational frequency that was identified based on the 454 sequencing data, we subcloned the PCR products and performed capillary DNA sequencing as previously described.[9] A Topo TA cloning kit (Invitrogen, San Diego, CA, USA) was used to subclone each of the 14 amplicons.

2.5. *Functional analysis*

To determine whether the identified *NLRP3* mutants are disease-causing, we assessed both ASC [apoptosis-associated speck-like protein containing a caspase recruitment domain; PYCARD, an approved symbol from the HUGO Gene Nomenclature Committee (HGNC) database]-dependent NF-κB activation in HEK293FT cells and transfection-induced cell death in THP-1 cells, a human monocytic cell line, as previously described.[9,13,15] cDNAs encoding carboxy-terminal green fluorescent protein (GFP)-tagged NLRP3 and its mutants were subcloned into pcDNA5/TO (Invitrogen). Before being introduced into THP-1 cells (10^6) using a Cell Line Nucleofector Kit V (Amaxa Biosystems, Cologne, Germany), phorbol myristate acetate (10 ng/ml) was added to enhance transient expression of *NLRP3* gene with minimizing spontaneous cell death.[15] Four hours after the introduction of plasmids (0.5 μg), cell death of GFP-positive THP-1 cells was measured by flow cytometry.

Expression plasmids for NLRP3 and ASC in the pEF-BOS vector background have been previously described.[13] HEK293FT cells (10^5) were transfected using TransIT-293 Transfection Reagent (Milus Bio, Madison, WI, USA) with an NF-κB reporter construct (pNF-κB-luc; 20 ng; BD Biosciences Clontech, Palo Alto, CA, USA), an internal control construct (pRL-TK; 5 ng; Toyo Ink, Tokyo, Japan), and wild-type or mutant NLRP3 expression plasmid (20 ng) in the presence or absence of ASC expression plasmid (20 ng). The amounts of total plasmid DNA used for transfection experiments were kept constant by adding pEF-BOS vector DNA. Twenty-four hours later, the transfected cells were harvested and subjected to dual luciferase assay by which the ability of each construct to induce NF-κB activation was assessed as previously described.[9]

3. Results

3.1. *Construction of base- and strand-specific error rate maps of* NLRP3 *exons from the MPS data of 50 control samples*

Errors in sequence reads generated by a Roche 454 GS-FLX sequencer are not randomly distributed along the sequence and depend on various factors.[16] Although this is a well-known characteristic of 454 sequencing, the occurrence pattern of these errors has not been explored in detail simply because these sequencing errors are considered noise that can be filtered out in most cases. However, it is highly critical to understand the occurrence pattern of sequencing errors on the MPS platform because low-level somatic mosaicism might appear at a rate close to that of sequencing errors. To address this, we collected

~1 million sequence reads using the 454 GS-FLX sequencer for 14 amplicons of *NLRP3* exons from 50 control samples that were thought to be free from somatic mosaicism, and ~94% of those reads were mapped to one of the reference *NLRP3* exon sequences. The number of sequencing depths for each amplicon of each sample on each strand was between 65 and 2139 (mean = 565.3, Supplementary Table S2). We found that the average error rate for each mutation category (insertion, deletion, mismatch, and ambiguous base calls) at each base position on each strand of the amplicons in the control samples was 0.22, 0.16, 0.036, and 0.014%, respectively (Fig. 1C). These values were consistent with those reported in a recent study on the error rates with 454 sequencing data.[16] The sequencing error in the 454 GS-FLX system tends to occur at the beginning and end of the reads,[11,16] and we confirmed this trend in our amplicon sequencing data (Supplementary Fig. S1). Moreover, after removing the end regions of the read sequences, we found that the error rates of the target regions for each category were 0.20, 0.134, 0.023, and 0.014%, respectively (Fig. 1C and Supplementary Table S3). When generating the amplicon sequences for the *NLRP3* exons, the target sequence (CDS region and flanking intron in 10-bp length) was designed to be 300–400 bp and not adjacent to primer sequences in order to obtain relatively low sequencing error rates (Fig. 1C). However, when the base- and strand-specific error rates of the respective amplicons were compared, we noticed that there were large variations in the error rate among amplicons in a strand-specific manner (Fig. 1D). We further examined the occurrence pattern of sequencing errors, as shown in Fig. 1E; the average sequencing error rates at each base in the 50 control amplicons are shown in a bar graph, where the bars in the upper or lower direction show the sequence error rates at the base position on the upper or lower strand of the amplicons, respectively. As evident in Fig. 1E, the error rates at most residues were low (<1%) with some hotspots for each type of error. Most of the insertion/deletion errors preferentially occurred at a homonucleotide region (yellow regions in Fig. 1E) as previously described,[17] but it was not always the case for all of homonucleotide regions. We could not find any tight relationship between other sequence patterns and the error rate. In addition, there was almost no position where sequencing errors occurred at a similar rate on both strands. This is more clearly shown in Fig. 1F, which indicates the numbers of positions with sequence variations in both strands that were higher than the threshold along the horizontal axis. These results indicate that the sequence errors can be discriminated from real genetic alterations when the sequence is read in both directions. However, it is important to keep in mind that PCR errors are not distinct from real genetic alterations. We did not observe any base substitution at a rate higher than 1% in our experiments (Fig. 1F), and the overall PCR error rate under MPS conditions was lower than 1% as long as a high-fidelity DNA polymerase was used to generate the amplicons.

Because Gilles *et al.*[16] recently reported that the occurrence of sequencing errors using the Roche 454 GS-FLX DNA sequencer depends on various factors, we first examined variations in the sequencing error rates of *NLRP3* exons among samples in the same run. For each mutation category, we found a similar trend in the error distribution rate in the amplicon sequences among the control samples (Supplementary Figs S2–S4). We confirmed that, for almost all residues, the error rate distributions among the 50 control samples fitted a Poisson distribution (data not shown). We next examined the run-to-run variation of the sequencing error rate for *NLRP3* exons. For this purpose, we performed an additional MPS run with seven amplicons (exons 3, 4, and 6) that were newly prepared and compared the number and positions of the sequencing errors between two independent sequencing runs. Out of 1993 base positions in the target regions, there was a low occurrence rate of mismatch errors in both runs and this seemed to fit a Poisson distribution. However, insertion/deletion errors (>1% error rate) were observed at ~100 base positions (<5% in the target regions) in each run, and only a half of these errors were shared between both runs (Table 1).

Table 1. Run-to-run variations in the error occurrence (>1% frequency)

Error category	Upper strand			Lower strand			All[a]
	First run	Second run	Overlap	First run	Second run	Overlap	
Insertions	63	73	42	76	96	52	10
Deletions	36	44	24	29	65	20	2
Mismatches	0	0	0	3	0	0	0
Ambiguous base calls	6	8	6	12	10	10	0

[a]The number of positions where the error rates in each category were commonly >1% for both strands in two independent runs.

This indicated that the occurrence of insertion/deletion errors was considerably affected by the run conditions (probably due to variations in the absolute signal strengths of pyrosequencing). Thus, as previously reported, the detection of insertion/deletion mutations by MPS on the 454 GS-FLX system was quite error-prone at least at a limited number of residues. However, the results also implied that false-positive mosaic mutations could be avoided by considering the sequencing data for both strands because these run-dependent insertion/deletion errors occur only in a single strand. Taken together, we conclude that the obtained sequence error map is stable and sufficiently robust to discriminate substitution sequencing errors from low-level mosaicism.

3.2. *Discrimination formula for detection of somatic mosaicism with statistical confidence*

We next examined known SNPs, known heterozygous mutations and somatic mosaic mutations of CAPS patients using MPS. All of these variations appeared on both strands at the expected allele frequencies as shown in Fig. 2, again indicating that filtering the strand-specific sequence variations is unlikely to eliminate real genetic variations.

Based on the experimentally observed sequencing errors with the 454 GS-FLX system described above, we established a discrimination formula to detect low-level somatic mosaicism as follows. In previous studies, the number of reads with the sequence error of a certain category in a sequence position was modelled based on the Poisson distribution with two parameters λ and k where the expected number of reads containing an error and the observed number of reads containing a sequence alteration, respectively, are as shown below[18]:

$$\mathrm{Pois}(k;\lambda) = \frac{\lambda^k e^{-\lambda}}{k!}. \qquad (1)$$

This model assumes that the error rate is constant across the different sequence regions but our data described above pointed out that the sequence error rate varies with the sequence content.[19] Thus, we introduced a position- and strand-specific error rate $q_{i,j,d}$ for a certain error category j in amplicon position i with strand d based on the sequencing data from 50 control samples. With the error rate $q_{i,j,d,}$ the upper probability (P) that the number of reads (R) with a certain sequence alteration of category j in position i is equal or greater than the number of observed reads r out of N reads with a sequenced direction d for an unknown sample was defined as:

$$P(R \geq r_{i,j,d}|\lambda_{i,j,d}) = 1 - \sum_{k=0}^{r-1} \frac{\lambda_{i,j,d}^k e^{-\lambda_{i,j,d}}}{k!}, \qquad (2)$$

where, $\lambda_{i,j,d} = N_{i,d} \times q_{i,j,d}$.

For the mismatch error rate, we did not consider the type of base substituted in an amplicon position in this study. We took $(1 - P)$ as a measure of the statistical confidence of the data and conventionally set a threshold of the statistical confidence to be 99.9%. In other words, if P-value was <0.001, the sequence alteration was considered to be a real sequence variation, not an error. For the final identification of real genetic variation with low-level somatic mosaicism, we determined that both of the P-values for the ith residue in the upper and lower strands must be smaller than the threshold.

To evaluate the lower detection limit for the allele frequencies of somatic mosaicism based on the statistical formulation shown above, we generated a series of known allele frequencies by diluting DNA from CAPS patients carrying heterozygous *NLRP3* mutations (c.1043C>T, c.1316C>T, and c.1985T>C) with DNA from normal donors carrying the wild-type *NLRP3* gene. In the dilution series, the mutant allele frequencies were adjusted to be 10, 5, 3, 2, 1, and 0.5% (Table 2). The data indicated that somatic mosaicism at these sites and at an allele frequency $\geq$1% could be convincingly detected with statistical significance ($P < 0.001$) if more than 350 reads for each strand were obtained for an amplicon. We also applied this statistical method to detect somatic mosaicism in patients with known low-level mosaic mutations described above and confirmed that all of

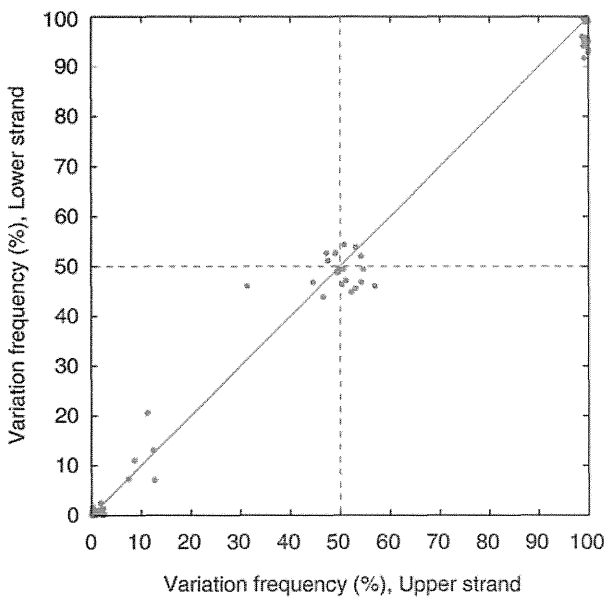

Figure 2. Scatter plot of the observed frequency variation in both strands. The colours depict known SNPs (green), heterozygous and mosaic mutations (orange) and errors (grey).

Table 2. Evaluation of the lower detection limit for mosaicism with three sets of dilution series

Mutation	Dilution (%)	Upper strand				Lower strand			
		Total reads	Mutant reads	%Mutant	*P*-value	Total reads	Mutant reads	%Mutant	*P*-value
c.1043C>T; p.Thr348Met	10.0	724	61	8.43	8.62E−130	520	57	10.96	1.73E−117
	5.0	453	24	5.30	2.86E−47	372	15	4.03	1.26E−25
	3.0	876	27	3.08	1.16E−46	757	21	2.77	6.83E−32
	2.0	737	10	1.36	1.05E−14	645	7	1.09	8.68E−09
	1.0	715	9	1.26	4.73E−13	624	4	0.64	1.11E−04
	0.5	1025	7	0.68	1.15E−14	756	3	0.40	3.22E−03[a]
c.1431C>A; p.Asn477Lys	10.0	542	65	11.99	1.22E−113	346	24	6.94	6.84E−49
	5.0	491	30	6.11	1.13E−44	356	17	4.78	2.42E−32
	3.0	487	21	4.31	1.26E−28	374	19	5.08	1.78E−36
	2.0	577	18	3.12	2.78E−22	495	9	1.82	4.57E−14
	1.0	491	4	0.82	9.17E−04	354	5	1.41	7.34E−08
	0.5	483	0	0	NA	424	3	0.71	NA
c.1985T>C; p.Met662Thr	10.0	658	79	12.01	1.13E−179	643	74	11.51	4.64E−167
	5.0	643	31	4.82	2.56E−59	608	33	5.43	9.96E−65
	3.0	777	27	3.48	4.65E−48	704	29	4.12	1.26E−53
	2.0	929	21	2.26	7.59E−34	835	15	1.80	3.92E−23
	1.0	735	17	1.09	2.74E−11	709	9	1.27	4.06E−13
	0.5	702	2	0.29	3.90E−03[a]	590	1	0.17	1.37E−01[a]

[a]Not significant.

Table 3. Potential mosaic mutations detected in patients with unknown mutations

Patient ID	Amplicon #	Variation		% Variation frequency		*P*-value		dbSNP	State
				Forward	Reverse	Forward	Reverse		
P1	Exon3_2	c.907G>C	p.Asp303His	7.12	11.56	3.0E−44	1.7E−84	rs121908153	Known
P2	Exon3_5	c.1699G>A	p.Glu567Lys	5.94	5.79	2.0E−69	8.9E−47	—	Known
P3	Exon3_5	c.1699G>A	p.Glu567Lys	18.28	15.33	0.0E+00	1.0E−312	—	Known
P4	Exon3_2	c.906C>A	p.Phe302Leu	9.78	9.70	1.7E−86	2.2E−122	—	Novel

the mutations could be detected with statistical significance without any false positives (data not shown).

3.3. *Detection and characterization of* NLRP3 *somatic mosaicism using the MPS platform*

To demonstrate the power of this approach in practice, we applied our new pipeline for 10 CINCA/NOMID patients in whom we failed to detect mutations in the *NLRP3* gene using a conventional direct DNA sequencing approach. The mutations detected by the analysis formulated using the MPS platform in this study are listed in Table 3. We successfully identified four out of the 10 patients with *NLRP3* somatic mosaicism, which was confirmed by subcloning and Sanger sequencing. The nucleotide substitutions were as follows (parentheses indicate the corresponding amino acid change): c.907G>C (p.Asp303His), c.1699G>A (p.Glu567Lys) in two patients, and c.906C>A (p.Phe302Leu). The frequencies of mosaicism identified in these patients by the MPS approach were consistent with those that were identified by the subcloning and subsequent capillary DNA sequencing method (data not shown). Both c.907G>C and c.1699G>A variants were reported as CINCA/NOMID-associated mutations in Infevers database (http://fmf.igh.cnrs.fr/ISSAID/infevers/) and in the dbSNP database (http://www.ncbi.nlm.nih.gov/projects/SNP/).[8]

Because the NLRP3 p.Phe302Leu mutation was novel and not detected in the 50 healthy controls, we performed an *in vitro* functional analysis to see the effect of p.Phe302Leu on the protein function. We used two different *in vitro* transfection experiments,

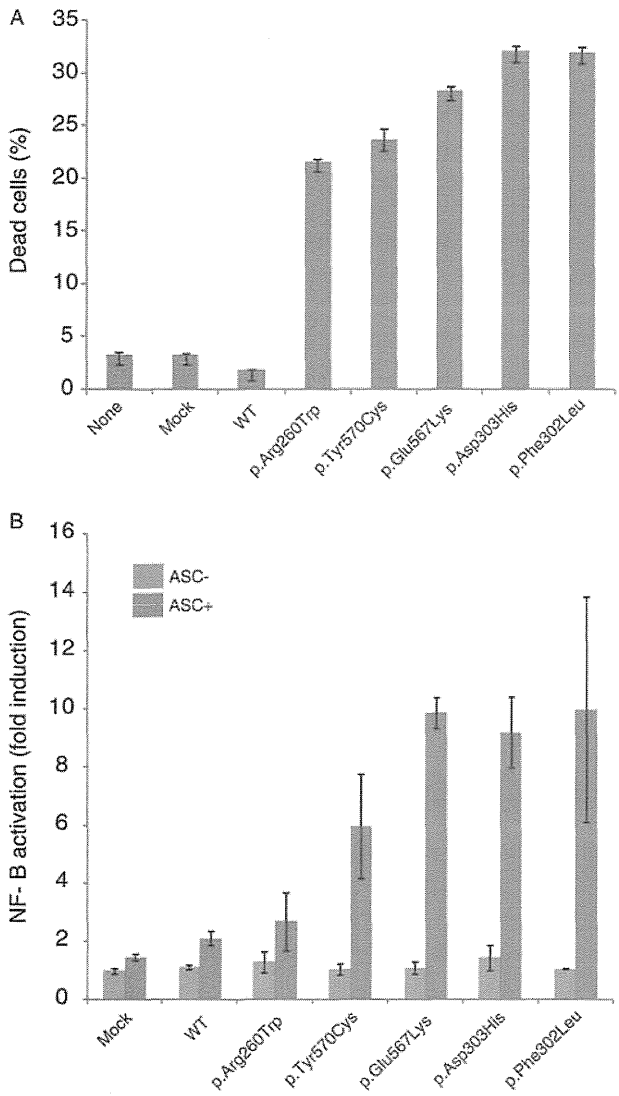

Figure 3. *In vitro* functional analysis of the identified *NLRP3* mosaic mutations. (A) Rapid cell death in transfected THP-1 cells. A GFP-fused wild-type or mutant *NLRP3* was transfected into THP-1 cells and incubated with PMA (10 ng/ml) for 4 h. The percentage of dead cells (7-amino-Actinomycin D [7-AAD]-positive) among the GFP-positive cells is shown. Data represent the means ± SD of triplicate experiments and are representative of two independent experiments. The data for previously reported mutations as well as the mutations found in this study are shown. (B) ACS-dependent NF-κB activation in transfected HEK293FT cells. HEK293FT cells were co-transfected with wild-type or mutant *NLRP3* in the presence or absence of ASC. NF-κB induction is shown as the fold-change compared with cells that were transfected with a control vector without ASC (set equal to one). Values are the means ± SD of triplicate experiments, and the data are representative of three independent experiments. The data for previously reported mutations (p.Arg260Trp and p.Tyr570Cys) and the mutations found in this study are shown. For each mutation, the data obtained in the presence and absence of ASC are shown. These findings identified p.Phe302Leu as a novel disease-causing mutation.

the rapid cell death in transfected THP-1 cells and the ASC-dependent NF-κB activation in transfected HEK293FT cells (Fig. 3A and B, respectively). Both assays clearly showed that p.Phe302Leu was a disease-causing mutation similar to known CINCA/NOMID-associated pathogenic mutations (p.Asp303His and p.Glu567Lys).[9]

4. Discussion

Although the somatic mutation rate at the nucleotide level *in vivo* was difficult to quantitatively measure due to the complexity of the genome and laborious molecular detection processes, recent advances in MPS technologies have allowed us to directly quantitate somatic mutations in human genome.[20–22] The current estimate for the somatic (*de novo*) mutation rate is $1-2 \times 10^{-8}$ residues/generation/haploid, and this estimate is sufficiently low that we would expect to never observe somatic mosaicism in the *NLRP3* gene by chance; although the error rate of the high-fidelity DNA polymerase used to produce the amplicons is two orders of magnitude larger than the somatic mutation rate,[23,24] we could not detect PCR-generated mosaicism higher than 1% in the 454 sequencing error maps. Based on the literature, the single base substitutions are the most frequent type of somatic mutations (~500 times more frequent than short insertions/deletions)[25] and protein-coding sequences are less mutagenic than sequences in non-coding regions, assuming that the somatic mutation spectrum in malignant cells is the same as in normal cells. Somatic mosaicism is thought to result from *de novo* gain-of-function-type mutations that are introduced at a very early and limited stage of development, and it is reasonable to focus our efforts on detecting base substitutions for somatic mosaicism in the *NLRP3* gene.

It is challenging but highly important in many areas of research, such as cancer, to detect low-level somatic mutations, which we designated as somatic mosaicism in this study, from apparently mutation-negative samples by conventional sequencing. Subcloning followed by the capillary DNA sequencing has been a *de facto* standard to identify somatic mosaicism, but this is not the method of choice for routine diagnostics because it is laborious, time consuming, and costly. Thus, it is reasonable for us to explore MPS as a new tool for this purpose. Although previous studies have used MPS technology to detect somatic mosaicism, it was unclear how sensitive this method is to detect a low-level somatic mosaicism using the MPS platform because this platform is generally error-prone. To address this challenge, we developed a new pipeline to detect low-level somatic mosaicism with statistical confidence using base position- and strand-specific error rate maps for the *NLRP3* amplicons to be studied. Whereas the

detection limit of somatic mosaicism depends on the base position and the read depth of the amplicons, the limit of detection could be as low as 1% allele frequency with no false positives for substitutions (the precision is higher than 99.9%). Our error map shows that 98.1% of base positions (3343 out of 3407 target positions) in the *NLRP3* exonic amplicons can be detected with ~1% mosaicism when more than ~350 reads were accumulated for each strand. Although the remaining region (64 base positions out of 3407 target positions) was too error-prone (the error rate ranged from 0.1 to 1.7% in either the upper or lower strand) to detect low-level mosaicism by MPS, medium-level mosaicism (5% or high) could be identified in all base positions in the target region with the same significance level. Based on this pipeline, we successfully identified four cases of somatic mosaicism among 10 apparently mutation-negative CINCA/NOMID patients. These results were subsequently confirmed by functional analysis and subcloning followed by capillary DNA sequencing method.

As described above, we revealed that a read depth of ~350 for each strand of each amplicon would be sufficient to detect somatic mosaicism as low as 1% with statistical confidence. This means that an analysis of somatic mosaicism (detection limit of 1% allele frequency) of the *NLRP3* gene for one sample requires $350 \times 2 \times 14 = 9800$ reads with the 454 GS-FLX sequencer, which has a capacity to obtain 1 000 000 reads per run. Thus, we could analyse ~100 patient samples with a single run (~10 h) using this MPS platform. For this purpose, a miniaturized 454 sequencer might be more convenient because it could analyse 10 patient samples at once with a reasonably reduced running cost.

The approach used to detect somatic mosaicism is very similar to that for low-frequency alleles in pooled DNA samples, for which MPS applications have been reported by many groups.[18,26,27] However, the main aim of these previous studies was to screen for a rare allele in a population. Thus, the discovery phase on the MPS platform must be followed by an evaluation phase using conventional methods. Therefore, when diagnosing somatic mosaicism of the *NLRP3* gene based solely on the MPS platform, we could not use the same approach to detect rare alleles in a population due to its low accuracy. The sequencing error rate on the Roche MPS platform was sufficiently stable and low enough as shown in this study. Using our pipeline, we were able to detect 1% somatic mosaicism in the *NLRP3* gene with 99.9% confidence. Although another research group recently used a similar approach with a short-read MPS,[28] the Roche long-read MPS is more suitable as a diagnostic tool mainly because of the short run time. If we could diagnose somatic mosaicism of the *NLRP3* gene within a reasonable time with low labour and costs as shown in this study, the success rate of CINCA/NOMID genetic diagnosis will increase from 60 to 80% or higher,[9] which will greatly advance the health and care of these patients and prevent irreversible bone and neurological complications of disease.

This pipeline would also be efficient to detect somatic mosaicism in mutation-negative patients with other diseases, including cancer. The error rate map for a given gene should be constructed from authentic plasmids, and used to detect somatic mosaicism of other genes as well as rare alleles in various populations.

Supplementary data: Supplementary data are available at www.dnaresearch.oxfordjournals.org.

Funding

This study was supported by the Japanese Ministry of Education, Science, Sports, and Culture, and the Japanese Ministry of Health, Labor, and Welfare.

Acknowledgements: We thank all patients who participated in the study. We are grateful to Ms. Yuki Takaoka at the Department of Pediatrics, Kyoto University Graduate School of Medicine and Mr. Takashi Watanabe at the Department of Human Genome Research, Kazusa DNA Research Institute for their technical assistance.

References

1. Prieur, A.M. and Griscelli, C. 1981, Arthropathy with rash, chronic meningitis, eye lesions, and mental retardation, *J. Pediatr.*, **99**, 79–83.
2. Hassink, S.G. and Goldsmith, D.P. 1983, Neonatal onset multisystem inflammatory disease, *Arthritis Rheum.*, **26**, 668–73.
3. Torbiak, R.P., Dent, P.B. and Cockshott, W.P. 1989, NOMID—a neonatal syndrome of multisystem inflammation, *Skeletal Radiol.*, **18**, 359–64.
4. Feldmann, J., Prieur, A.M., Quartier, P., et al. 2002, Chronic infantile neurological cutaneous and articular syndrome is caused by mutations in CIAS1, a gene highly expressed in polymorphonuclear cells and chondrocytes, *Am. J. Hum. Genet.* United States, 198–203.
5. Aksentijevich, I., Nowak, M., Mallah, M., et al. 2002, De novo CIAS1 mutations, cytokine activation, and evidence for genetic heterogeneity in patients with neonatal-onset multisystem inflammatory disease (NOMID)—a new member of the expanding family of pyrin-associated autoinflammatory diseases, *Arthritis Rheum.*, **46**, 3340–8.

6. Hoffman, H.M., Mueller, J.L., Broide, D.H., Wanderer, A.A. and Kolodner, R.D. 2001, Mutation of a new gene encoding a putative pyrin-like protein causes familial cold autoinflammatory syndrome and Muckle-Wells syndrome, *Nat. Genet.*, **29**, 301–5.
7. Goldbach-Mansky, R. 2011, Current status of understanding the pathogenesis and management of patients with NOMID/CINCA, *Curr. Rheumatol. Rep.*, **13**, 123–31.
8. Milhavet, F., Cuisset, L., Hoffman, H.M., et al. 2008, The infevers autoinflammatory mutation online registry: Update with new genes and functions, *Hum. Mutat.*, **29**, 803–8.
9. Tanaka, N., Izawa, K., Saito, M.K., et al. 2011, High incidence of NLRP3 somatic mosaicism in patients with chronic infantile neurologic, cutaneous, articular syndrome: results of an International Multicenter Collaborative Study, *Arthritis Rheum.*, **63**, 3625–32.
10. Qin, W., Kozlowski, P., Taillon, B.E., et al. 2010, Ultra deep sequencing detects a low rate of mosaic mutations in tuberous sclerosis complex, *Hum. Genet.*, **127**, 573–82.
11. Campbell, P.J., Pleasance, E.D., Stephens, P.J., et al. 2008, Subclonal phylogenetic structures in cancer revealed by ultra-deep sequencing, *Proc. Natl. Acad. Sci. USA*, **105**, 13081–6.
12. Rohlin, A., Wernersson, J., Engwall, Y., Wiklund, L., Bjoerk, J. and Nordling, M. 2009, Parallel sequencing used in detection of mosaic mutations: Comparison with four diagnostic DNA screening techniques, *Hum. Mutat.*, **30**, 1012–20.
13. Saito, M., Nishikomori, R., Kambe, N., et al. 2008, Disease-associated CIAS1 mutations induce monocyte death, revealing low-level mosaicism in mutation-negative cryopyrin-associated periodic syndrome patients, *Blood*, **111**, 2132–41.
14. Kent, W.J. 2002, BLAT—the BLAST-like alignment tool, *Genome Res.*, **12**, 656–64.
15. Fujisawa, A., Kambe, N., Saito, M., et al. 2007, Disease-associated mutations in CIAS1 induce cathepsin B-dependent rapid cell death of human THP-1 monocytic cells, *Blood*, **109**, 2903–11.
16. Gilles, A., Meglecz, E., Pech, N., Ferreira, S., Malausa, T. and Martin, J.F. 2011, Accuracy and quality assessment of 454 GS-FLX Titanium pyrosequencing, *BMC Genomics*, **12**, 245.
17. Margulies, M., Egholm, M., Altman, W.E., et al. 2005, Genome sequencing in microfabricated high-density picolitre reactors, *Nature*, **437**, 376–80.
18. Altmann, A., Weber, P., Quast, C., Rex-Haffner, M., Binder, E.B. and Muller-Myhsok, B. 2011, vipR: variant identification in pooled DNA using R, *Bioinformatics*, **27**, i77–84.
19. Dohm, J.C., Lottaz, C., Borodina, T. and Himmelbauer, H. 2008, Substantial biases in ultra-short read data sets from high-throughput DNA sequencing, *Nucleic Acids Res.*, **36**, e105.
20. Lee, W., Jiang, Z., Liu, J., et al. 2010, The mutation spectrum revealed by paired genome sequences from a lung cancer patient, *Nature*, **465**, 473–7.
21. Awadalla, P., Gauthier, J., Myers, R.A., et al. 2010, Direct Measure of the de novo mutation rate in Autism and Schizophrenia Cohorts, *Am. J. Hum. Genet.*, **87**, 316–24.
22. Conrad, D.F., Keebler, J.E.M., DePristo, M.A., et al. 2011, Variation in genome-wide mutation rates within and between human families, *Nat. Genet.*, **43**, 712–4.
23. Cha, R.S. and Thilly, W.G. 1993, Specificity, efficiency, and fidelity of PCR, *PCR Methods Appl.*, **3**, S18–29.
24. Vandenbroucke, I., Marck, H.V., Verhasselt, P., et al. 2011, Minor variant detection in amplicons using 454 massive parallel pyrosequencing: experiences and considerations for successful applications, *BioTechniques*, **51**, 167–77.
25. Pleasance, E.D., Cheetham, R.K., Stephens, P.J., et al. 2010, A comprehensive catalogue of somatic mutations from a human cancer genome, *Nature*, **463**, 191–6.
26. Fakhrai-Rad, H., Zheng, J.B., Willis, T.D., et al. 2004, SNP discovery in pooled samples with mismatch repair detection, *Genome Res.*, **14**, 1404–12.
27. Bansal, V. 2010, A statistical method for the detection of variants from next-generation resequencing of DNA pools, *Bioinformatics*, **26**, i318–24.
28. Flaherty, P., Natsoulis, G., Muralidharan, O., et al. 2012, Ultrasensitive detection of rare mutations using next-generation targeted resequencing, *Nucleic Acids Res.*, **40**, e2.

J Clin Immunol (2011) 31:968–976
DOI 10.1007/s10875-011-9594-7

Nationwide Survey of Patients with Primary Immunodeficiency Diseases in Japan

Masataka Ishimura · Hidetoshi Takada · Takehiko Doi · Kousuke Imai · Yoji Sasahara · Hirokazu Kanegane · Ryuta Nishikomori · Tomohiro Morio · Toshio Heike · Masao Kobayashi · Tadashi Ariga · Shigeru Tsuchiya · Shigeaki Nonoyama · Toshio Miyawaki · Toshiro Hara

Received: 7 August 2011 / Accepted: 11 September 2011 / Published online: 29 September 2011

Abstract To determine the prevalence and clinical characteristics of patients with in Japan, we conducted a nationwide survey of primary immunodeficiency disease (PID) patients for the first time in 30 years. Questionnaires were sent to 1,224 pediatric departments and 1,670 internal medicine departments of Japanese hospitals. A total of 1,240 patients were registered. The estimated number of patients with PID was 2,900 with a prevalence of 2.3 per 100,000 people and homogenous regional distribution in Japan. The male-to-female ratio was 2.3:1 with a median age of 12.8 years. Adolescents or adults constituted 42.8% of the patients. A number of 25 (2.7%) and 78 (8.5%) patients developed malignant disorders and immune-related diseases, respectively, as complications of primary immunodeficiency disease. Close monitoring and appropriate management for these complications in addition to prevention of infectious diseases is important for improving the quality of life of PID patients.

Keywords Primary immunodeficiency disease · epidemiology · nationwide survey · Japan

Abbreviations

APECED	Autoimmune polyendocrinopathy with candidiasis and ectodermal dystrophy
BTK	Bruton's tyrosine kinase
CGD	Chronic granulomatous disease
CID	Combined T and B cell immunodeficiency
CVID	Common variable immunodeficiency disease
FMF	Familial Mediterranean fever
IPEX	Immune dysregulation polyendocrinopathy enteropathy X-linked
NEMO	Nuclear factor kappa B essential modulator
PID	Primary immunodeficiency disease
SIgAD	Selective IgA deficiency
SLE	Systemic lupus erythematosus

M. Ishimura (✉) · H. Takada · T. Doi · T. Hara
Department of Pediatrics, Graduate School of Medical Sciences, Kyushu University,
3-1-1 Maidashi, Higashi-ku,
Fukuoka 812-8582, Japan
e-mail: ischii@pediatr.med.kyushu-u.ac.jp

K. Imai · S. Nonoyama
Department of Pediatrics, National Defense Medical College, Tokorozawa, Japan

Y. Sasahara · S. Tsuchiya
Department of Pediatrics, Tohoku University School of Medicine, Sendai, Japan

H. Kanegane · T. Miyawaki
Department of Pediatrics, Graduate School of Medicine and Pharmaceutical Science, University of Toyama, Toyama, Japan

R. Nishikomori · T. Heike
Department of Pediatrics, Kyoto University Graduate School of Medicine, Kyoto, Japan

T. Morio
Department of Pediatrics, Tokyo Medical and Dental University Graduate School, Bunkyo-ku, Tokyo, Japan

M. Kobayashi
Department of Pediatrics, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima, Japan

T. Ariga
Department of Pediatrics, Graduate School of Medicine, Hokkaido University, Sapporo, Japan

TRAPS	Tumor necrosis factor receptor-associated periodic syndrome
WAS	Wiskott–Aldrich syndrome
WHIM	Warts hypogammaglobulinemia, infections, and myelokathexis

Introduction

Patients with primary immunodeficiency disease (PID) show susceptibility to infections due to congenital immune system defects. These patients are also associated with noninfectious complications including autoimmune diseases and malignant disorders. Recent studies have revealed the causes of many PIDs to be mutations in various genes encoding molecules involved in the host defense mechanisms [1]. In addition, various new PIDs including defects in innate immunity and autoinflammatory disorders were identified under the recent progress in immunology and molecular genetics [2]. PID classification has been revised according to the identification of new PIDs and on the basis of new findings in PID pathophysiology. For a more precise clinical analysis, data should be obtained in accordance with the latest PID classifications.

The first nationwide survey of patients with PID in Japan was conducted between 1974 and 1979, which included 497 registered cases [3]. By 2007, a total of 1,297 patients were cataloged by a small number of PID specialists into a registration system [4]. The approximate prevalence of PID patients in Japan in the first nationwide survey was 1.0 in 100,000 people, which was much lower than that in other countries [5–7]. This difference in PID prevalence between Japan and other countries suggested that some PID patients in Japan remained unregistered. To determine the prevalence and clinical characteristics of patients with PID in Japan on the basis of the recent international classification system for PID, we conducted a nationwide survey of PID for the first time in 30 years.

Methods

This study was performed according to the nationwide epidemiological survey manual of patients with intractable diseases (2nd edition 2006, Ministry of Health, Labour, and Welfare of Japan) as described previously [8]. PID classification was based on the International Union of Immunological Societies Primary Immunodeficiency Diseases Classification Committee in 2007 [2]. Patients with chronic benign neutropenia and syndrome of periodic fever, aphthous stomatitis, pharyngitis, and cervical adenitis were excluded because these were considered to be acquired diseases. The survey was conducted on PID patients who were alive on December 1, 2008 and those who were newly diagnosed and dead between December 1, 2007 and November 30, 2008 in Japan. Among the 2,291 pediatric departments and 8,026 internal medicine departments in Japan, hospitals participating in the survey were randomly selected after setting the selection ratio according to the number of beds (overall selection rate: 53.4% for pediatric departments, 20.8% for internal medicine departments; Table I). University hospitals and pediatric training hospitals, where many PID patients were considered to be treated, were stratified separately (Table I). Primary questionnaires regarding the number of patients and disease names based on PID classification were sent to the selected hospitals. Secondary questionnaires regarding age, gender, clinical manifestations, and complications of individual PID patients were sent to respondents who answered that they observed at least one PID patient with characteristics listed in the primary questionnaires.

Results

Questionnaires were distributed to 1,224 pediatric departments and 1,670 internal medicine departments of hospitals in Japan, and the response rate was 55.0% and 20.1%, respectively (Table I). A total of 1,240 patients (1,146 patients from pediatric departments and 94 patients from internal medicine departments) were registered (Table I). The estimated number of patients with PIDs in Japan was 2,900 (95% confidence interval: 2,300–3,500), and the prevalence was 2.3 per 100,000 inhabitants. We also determined the regional distribution on the basis of the patients' addresses. The estimated regional prevalence ranged from 1.7 to 4.0 per 100,000 inhabitants, and no significant differences were observed between different regions in Japan (Fig. 1). The most common form of PID was predominantly antibody deficiencies (40%), followed by congenital defects of phagocyte number, function, or both (19%) and other well-defined immunodeficiency syndromes (16%; Table II). Autoinflammatory disorders were observed in 108 cases (9%). The most common PID was Bruton's tyrosine kinase (BTK) deficiency (182 cases, 14.7%), followed by chronic granulomatous disease (CGD; 147 cases, 11.9%). However, common variable immunodeficiency disease (CVID) and selective IgA deficiency (SIgAD) were observed only in 136 (11.0%) and 49 cases (4.0%), respectively. Among patients registered from internal medicine departments, antibody deficiencies were the most common disorder (71%).

In the secondary survey, 923 cases were registered. The male-to-female ratio was 2.3:1 (n=914, unanswered: 9 cases) with a median age of 12.8 years (range: 0 to 75 years; n= 897, unanswered: 26 cases). The number of adolescent or

Table I Stratification and selection of hospitals and the survey results

	Stratification	Departments in Japan	Departments selected	Selection rate (%)	Return[a]	Response	Response rate (%)	PID Patient	Patients per department	Patients estimated
Pediatrics	University hospital	118	118	100	0	80	67.8	661	8.3	975
	Training hospital	402	402	100	4	242	60.8	376	1.6	618
	≥500 beds	92	92	100	5	48	55.2	24	0.5	44
	400–499 beds	118	118	100	3	63	54.8	42	0.7	77
	300–399 beds	287	230	80.1	4	122	54.0	31	0.3	72
	200–299 beds	289	116	40.1	4	53	47.3	6	0.1	32
	100–199 beds	486	98	20.2	0	44	44.9	4	0.1	44
	<99 beds	499	50	10.0	1	10	20.4	2	0.2	100
	Subtotal	2,291	1,224	53.4	21	662	55.0	1,146	1.7	1,961
Internal medicine	University hospital	156	156	100	1	47	30.3	37	0.8	122
	≥500 beds	374	374	100	1	86	23.1	35	0.4	152
	400–499 beds	328	263	80	1	54	20.6	6	0.1	36
	300–399 beds	692	278	40.2	6	49	18.0	10	0.2	140
	200–299 beds	1,008	202	20.0	0	36	17.8	2	0.1	56
	100–199 beds	2,460	246	10.0	1	36	14.7	1	0.0	68
	<99 beds	3,008	151	5.0	6	24	16.6	3	0.1	375
	Subtotal	8,026	1,670	20.8	16	332	20.1	94	0.3	950
Total		10,317	2,894	28.1	37	994	34.8	1,240		2,911

[a] Due to the closure of departments

adult cases (≥15 years) was 384 (42.8%; Fig. 2a). The male-to-female ratio of the younger generation (<15 years) was 2.7:1, while that of the older generation (≥15 years) was 2.0:1. Combined T and B cell immunodeficiencies (CIDs) were predominantly observed in the younger generation, while antibody deficiencies were more common with

Region	Reported PID Patients	Population ($x10^5$)	Estimated prevalence per 10^5 (95% C.I.)
Hokkaido	73	55.4	4.0 (0.2-7.8)
Tohoku	81	94.3	2.1 (0.9-3.3)
Kanto	387	419.8	1.8 (1.2-2.5)
Chubu	236	236.9	2.3 (1.4-3.1)
Kinki	158	208.4	2.0 (1.2-2.8)
Chugoku/Shikoku	105	116.1	1.7 (1.0-2.3)
Kyushu/Okinawa	200	146.0	2.5 (1.7-3.3)
Total	1240	1276.9	2.3 (1.8-2.7)

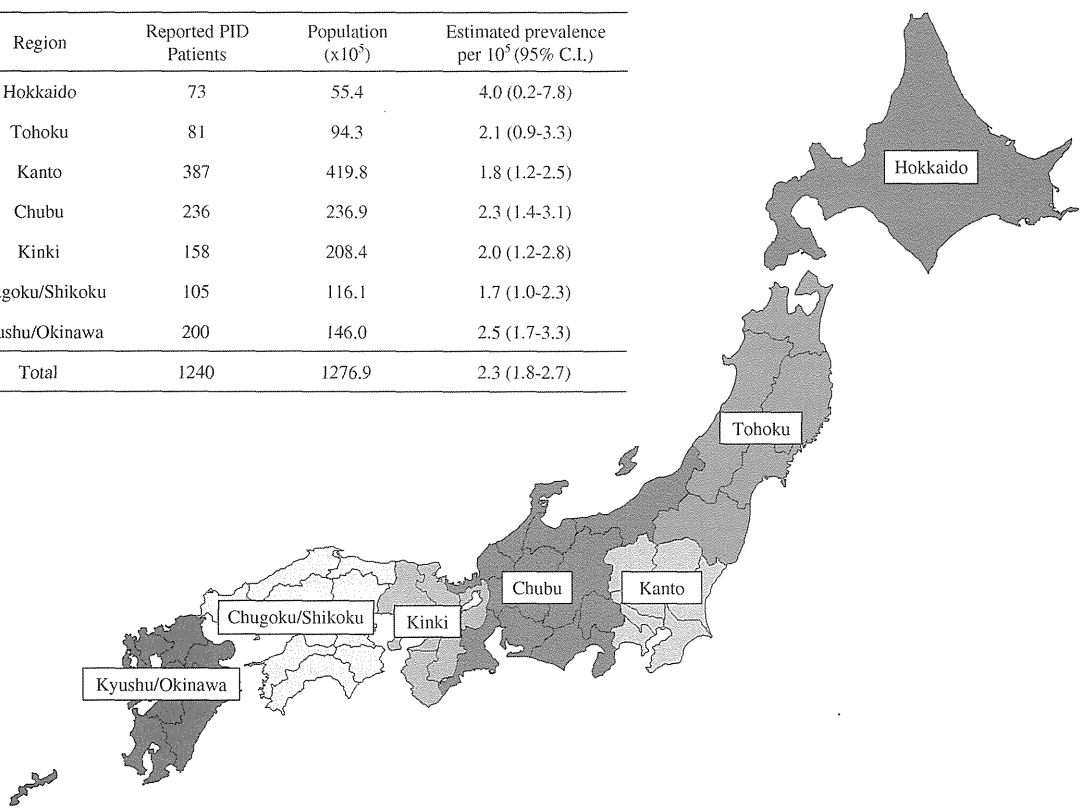

Fig. 1 Regional distribution of PID patients. *CI* Confidence interval

Table II Reported number of PID

Category	Total number	Pediatric department	Internal medicine department
I. Combined T and B cell immunodeficiencies	93 (7%)	93 (8%)	0 (0%)
γc deficiency	47	47	0
Adenosine deaminase deficiency	9	9	0
Omenn syndrome	4	4	0
Others	23	23	0
Untested or undetermined	10	10	0
II. Predominantly antibody deficiencies	501 (40%)	434 (38%)	67 (71%)
BTK deficiency	182	173	9
Common variable immunodeficiency disorders	136	107	29
Selective IgG subclass deficiency	66	58	8
Selective IgA deficiency	49	34	15
Hyper IgM syndrome	34	34	0
Transient hypogammaglobulinemia of infancy	7	7	0
Others	11	7	4
Untested or undetermined	16	14	2
III. Other well-defined immunodeficiency syndromes	194 (16%)	189 (17%)	5 (5%)
Wiskott–Aldrich syndrome	60	60	0
DNA repair defects (other than those in category I)	15	15	0
DiGeorge anomaly	38	38	0
Hyper-IgE syndrome	56	52	4
Chronic mucocutaneous candidiasis	17	16	1
Others	5	5	0
Untested or undetermined	3	3	0
IV. Diseases of immune dysregulation	49 (4%)	48 (4%)	1 (1%)
Chediak–Higashi syndrome	9	8	1
Familial hemophagocytic lymphohistiocytosis syndrome	5	5	0
X-linked lymphoproliferative syndrome	8	8	0
Autoimmune lymphoproliferative syndrome	8	8	0
APECED	4	4	0
IPEX syndrome	7	7	0
Others	2	2	0
Untested or undetermined	6	6	0
V. Congenital defects of phagocyte number, function, or both	230 (19%)	223 (19%)	7 (8%)
Severe congenital neutropenia	44	42	2
Cyclic neutropenia	19	17	2
Chronic granulomatous disease	147	144	3
Mendelian susceptibility to mycobacterial disease	5	5	0
Others	9	9	0
Untested or undetermined	6	6	0
VI. Defects in innate immunity	15 (1%)	15 (1%)	0
Anhidrotic ectodermal dysplasia with immunodeficiency	7	7	0
Interleukin-1 receptor-associated kinase 4 deficiency	2	2	0
Others	5	5	0
Untested or undetermined	1	1	0
VII. Autoinflammatory disorders	108 (9%)	101 (9%)	7 (8%)
Familial Mediterranean fever	44	40	4
TNF receptor-associated periodic syndrome	13	12	1
Hyper IgD syndrome	4	4	0
Cryopyrin-associated periodic syndrome	22	22	0

Table II (continued)

Category	Total number	Pediatric department	Internal medicine department
Others	3	3	0
Untested or undetermined	22	20	2
VIII. Complement deficiencies	32 (3%)	29 (3%)	3 (3%)
IX. Undetermined	18 (1%)	14 (1%)	4 (4%)
Total	1,240	1,146	94

APECED Autoimmune polyendocrinopathy with candidiasis and ectodermal dystrophy, *IPEX* immune dysregulation, polyendocrinopathy, enteropathy, X-linked

increasing age (Fig. 2b). The median age of CID, BTK deficiency, CVID, and CGD patients was 5.2, 12.8, 25.1, and 14.7 years, respectively.

It is well known that PID patients are susceptible to many pathogens and experience community-acquired or opportunistic infections. In this study, we focused on noninfectious complications of PID because they have been less well studied on a large scale and may provide important information for improving the quality of life of PID patients. Twenty-five PID patients developed malignant disorders (2.7%; Table III). Lymphoma, in particular, Epstein–Barr virus-related, and leukemia were dominant, while there were no patients with gastric carcinoma. CVID, Wiskott–Aldrich syndrome (WAS), and ataxia telangiectasia were more frequently associated with malignant diseases among PID patients. A case of Mendelian susceptibility

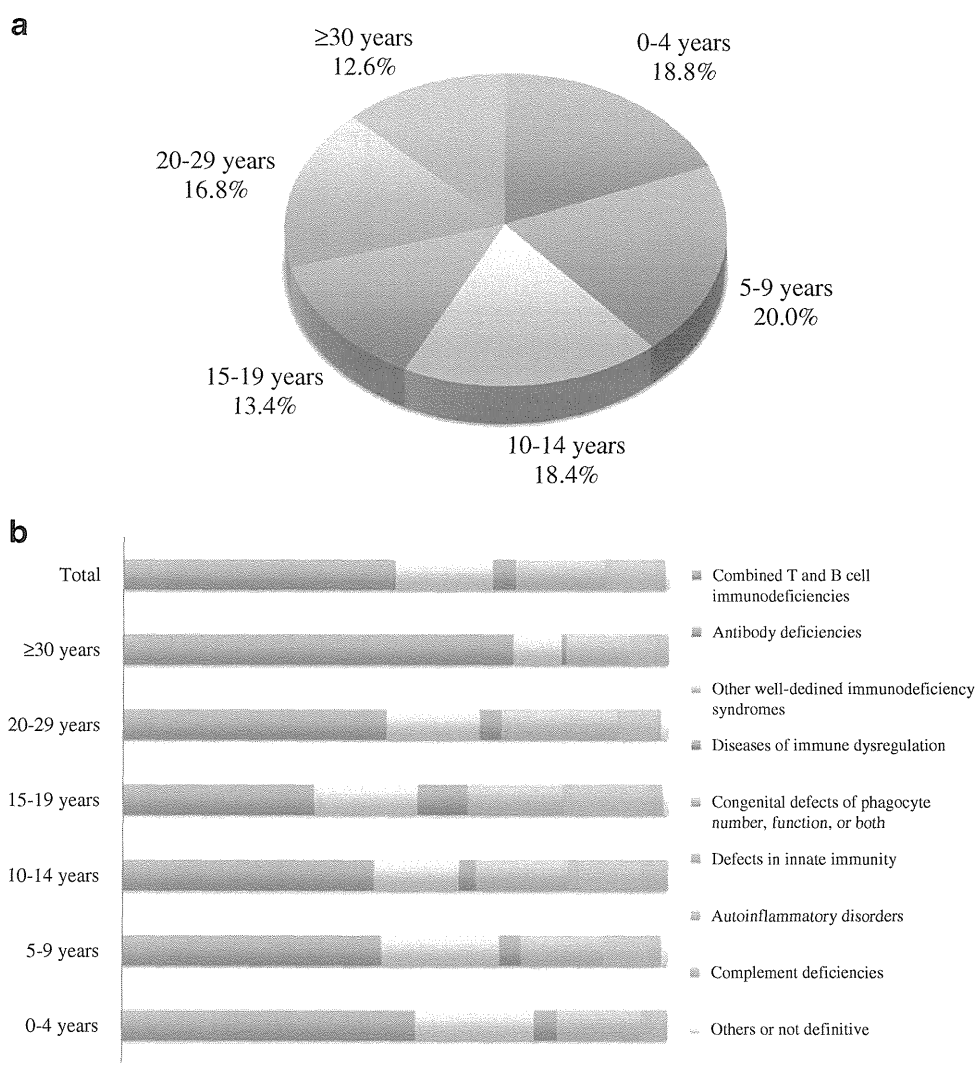

Fig. 2 **a** Age distribution of PID patients. **b** Distribution of PID in each age group

to mycobacterial disease with squamous cell carcinoma was also observed [9] (Table III).

Seventy-eight PID patients had immune-related (autoimmune) diseases (8.5%; Table IVa). Autoimmune lymphoproliferative syndrome, immune dysregulation, polyendocrinopathy, enteropathy X-linked (IPEX) syndrome, and nuclear factor kappa B essential modulator (NEMO) deficiency were associated with immune-related diseases at a very high incidence. In addition, immune-related diseases were relatively common in CGD and CVID patients (Table IVa). The most commonly observed immune-related disease was inflammatory bowel disease (33 cases), which was most frequently observed in CGD patients, followed by immune thrombocytopenic purpura (13 cases), autoimmune hemolytic anemia (8 cases), and systemic lupus erythematosus (SLE; 8 cases; Table IVa and b). Kawasaki disease occurred in WAS and CGD patients. In addition, this is the first report of Kawasaki disease in patients with complement deficiency (C9) and familial Mediterranean fever (FMF). A patient with warts, hypogammaglobulinemia, infections, and myelokathexis (WHIM) syndrome and a patient with tumor necrosis factor receptor-associated periodic syndrome (TRAPS) were first reported as cases of type 1 diabetes mellitus and SLE, respectively [10, 11].

Discussion

We conducted a nationwide survey of PID for the first time in 30 years and report the prevalence of PID in Japan. We registered 1,240 PID patients and found that the estimated prevalence of PID (2.3/100,000) is higher than that previously reported (1.0/100,000) in Japan. Our results are equivalent to those reported in Singapore (2.7/100,000) and Taiwan (0.77–2.17/100,000) [12–14]. However, our values are lower than those reported in Middle Eastern countries such as Kuwait (11.98/100,000) or in European countries such as France (4.4/100,000) [5–7, 15]. The high rate of consanguinity may be a cause of the high prevalence rate of PID reported in Middle Eastern countries [6, 15]. There may has been sample selection bias in this study because some asymptomatic cases (SIgAD, etc.), clinically recovered cases (transient hypogammaglobulinemia of infancy, etc.), and cases in which patients were deceased were not registered. In addition, lack of recognition of PID in internal medicine departments, not just the low response rate, might also have influenced the estimated prevalence of PID as well as the age and disease distribution. The regional prevalence of PIDs in Japan was homogenous, unlike in other countries in which a higher prevalence was

Table III Malignancies in PID patients

Primary immunodeficiency	Total	*n*	Malignancy
I. Combined T and B cell immunodeficiencies	75	2	(2.7%)
Ommen syndrome	3	1	NHL (EBV+) 1[a]
Adenosine deaminase deficiency	4	1	Breast carcinoma 1
II. Predominantly antibody deficiencies	378	8	(2.1%)
Common variable immunodeficiency disorders	93	7	HL 2, ML 2, ALL 1, Basal cell carcinoma 1, Cervical carcinoma 1
Good syndrome	4	1	Double primary carcinoma of breast and colon 1
III. Other well-defined immunodeficiency syndromes	165	7	(4.2%)
Wiskott–Aldrich syndrome	57	5	NHL 3, NHL/HL 1, LPD (EBV-) 1
Ataxia telangiectasia	13	2	T-ALL 1, MDS 1
IV. Diseases of immune dysregulation	38	4	(10.5%)
X-linked lymphoproliferative syndrome	5	2	Burkitt lymphoma 2
Autoimmune lymphoproliferative syndrome	6	2	HL (EBV+) 1, Brain tumor 1
V. Congenital defects of phagocyte number, function, or both	153	4	(2.6%)
Severe congenital neutropenia	35	3	MDS 3 (including 2 cases with monosomy 7)
MSMD	7	1	Squamous cell carcinoma of finger 1
VI. Defects in innate immunity	12	0	(0%)
VII. Autoinflammatory disorders	74	0	(0%)
VIII. Complement deficiencies	23	0	(0%)
IX. Undetermined	5	0	(0%)
Total	923	25	(2.7%)

n Number of PID patients who had malignant disorders, *ALL* acute lymphoblastic leukemia, *EBV* Epstein-Barr virus, *HL* Hodgkin lymphoma, *LPD* lymphoproliferative disease, *MDS* myelodysplastic syndrome, *ML* malignant lymphoma, *MSMD* Mendelian susceptibility to mycobacterial disease, *NHL* non-Hodgkin lymphoma

[a] The number of patients