phase, we focus on describing a task dependency in a rakefile as shown below:

```
1: rule '.dedup.bam.intervals' => [ suffix_proc(".bam.intervals" => ".bam") ] do |t|
2:     RakefileInvoke::Gatk.gatk_realigner_target_creater t
3: end
```

Next, in the parameter adjustment phase, we focus on describing command-line parameters for invoking external tools in the rakefile.invoke such as the following:

```
1: def gatk_realigner_target_creater(t)
2: sh [Java,
3:   "-Xmx#{JavaMemory}",
4:   "-Djava.io.tmpdir = #{JavaTempFile}",
5:   "-jar #{GATK_JAR}",
6:   "-T RealignerTargetCreator",
7:   "-R #{REFERENCE}",
8:   "-o #{t.name}",
9:   "-I #{t.source}",
10:  "-D #{DBSNP}",
11:  RakefileInvoke::Gatk::INTERVAL_OPTION,
12:  " > #{t.name}.log 2 > &1",
13:  ].join(" ")
14: end
```

Note that all constants with names starting with uppercase letters are defined at the top of the file, `rakefile.invoke`. The next iteration starts with the workflow definition phase again to extend the workflow. Modification or optimization after the workflow has completed can be achieved by iterating the same two phases using two distinct files. Separating the rakefiles simplifies finding files and places to be modified.

Procedure to describe new workflows

As a summary of the agile workflow development, the general procedure for describing new workflows in Pwrake is given below.

1) Workflow definition phase. Describe file dependencies in `Rakefile`.

```
1: task "output.dat" => "input.dat" do |t|
2:   RakefileInvoke::generate_target t
3: end
```

2) Parameter adjustment phase: Define the `RakefileInvoke::generate_target` method in `Rake.invoke`.

```
1: module RakefileInvoke
2:   def generate_target(t)
3:     sh "command-line #{t.prerequisite} > #{t.name}"
4:   end
5: end
```

3) Iteration of phases. Parameter adjustments require modifications to `Rakefile.invoke` only. Similarly, changes in file dependencies require modification to `Rakefile` only.

Discussion

Advantages in workflow execution

Workflows involving actively developed software packages, such as GATK, require frequent updates of details, such as combinations of data and programs, recommended parameters, and command-line options. Thus, well-organized workflow management helps GATK users to follow updates and process their data in improved workflows. A GATK workflow consists of multiple steps and takes a relatively longer time to finish. Pwrake has advantages of continuous execution of workflow tasks and selective task execution to ignore already executed tasks. Such ignorable tasks can be obtained from unexpected workflow suspension. Thus far, Pwrake cannot automatically remove output files containing partial results; such files have to be removed manually prior to restarting the workflow.

For the Dindel workflows, the parallelism offered by Pwrake improved performance. The parallelization model of Pwrake is process-based. Parallel programs based on technologies such as message passing interface (MPI) [40] enable efficient parallelization with fine granularity. However, scientists implementing bioinformatics software often focus not on parallelization, but on the novel implementation methodology. Therefore, process-based parallelization using non-parallel programs is a realistic solution and still has the advantage [41]. Furthermore, process-based parallelization can be efficient enough for embarrassingly parallel problems that can easily be separated into independent tasks and executed in parallel. For example, a stage in the Dindel workflow creates multiple intermediate files. Processes using these files as input are independent and do not need to communicate with each other. This stage is a typical embarrassingly parallel problem. Although the GATK framework supports the functional programming concept of MapReduce [42] and parallelism in the GATK framework is expected to improve its performance, it has only been supported to a limited extent by GATK components to date. Therefore, Pwrake still has the advantage with respect to parallelism.

Workflow description flexibility

One of the advantages of using an internal DSL is that the power of the host language is also available in the DSL scripts. The rakefile description is an internal DSL in Ruby, which is a programming language with a shallow learning curve for biologists [43]. Thus, rakefiles can make full use of the control flow features of Ruby, as well as the rich libraries for text processing, file manipulation,

network access, and so on. In particular, the BioRuby [44] library offers highly abstracted data processing methods for bioinformatics.

Sharing workflows

One of the key characteristics of agile software development is strong collaboration among all the people involved in the project. This can be accomplished naturally in projects in small laboratories. However, the nature of science is a global collaboration. Indeed, efforts to share and reuse workflows in the science community, such as the myExperiment project [45] and Wf4Ever [46], have already been started. From this point of view, the simplicity and readability of the rakefile DSL are advantageous, and improvement of helper methods to standardize the scripting style on the "Do not Repeat Yourself (DRY)" principle may enhance the advantages.

Conclusions

We have shown an appreciation of Pwrake as an agile parallel workflow system suitable for the bioinformatics domain using examples of GATK and Dindel workflows. Pwrake is able to invoke command-line tools without any "shims", define tasks dynamically during the workflow execution, and invoke tasks automatically in parallel. Separating a rakefile into two files for the workflow definition phase and the parameter adjustment phase increases the efficiency of the iterative workflow development. The nature of scientific projects is explorative and iterative. This is also a characteristic of agile software development. Another aspect of agile development, the reliance on the strong collaboration, may be enhanced by sharing and reusing workflows among the scientific community by taking advantage of the simplicity, readability and maintainability of rakefiles.

Availability and requirements

Project name: Workflows
Project home page: http://github.com/misshie/Workflows
Operating system(s): Platform independent
Programming language: Ruby 1.9.1 or higher
Other requirement: Pwrake or Rake
License: the MIT license
Any restrictions for use by non-academics: none

Availability of supporting data

Sample short read data for workflow evaluation: http://trace.ddbj.nig.ac.jp/DRASearch/experiment?acc=DRX000358

List of abbreviations used

HPC: high-performance computing; DSL: domain specific language; GATK: Genome Analysis Toolkit; SNV: single nucleotide variant; BWA: Burrows-Wheeler Alignment tool; MPI: message passing interface; DRY: do not repeat yourself.

Acknowledgements

The authors would like to thank members of the BioRuby mailing list for their informative discussions. HM is supported by the MEXT Grant-in-Aid for Young Scientists (B) 21791566 and 23791230. OT is supported by the MEXT Grant-in-Aid for Scientific Research on Priority Areas 21013005. OT and MT are supported by the MEXT Promotion of Research for Next Generation IT Infrastructure "Resources Linkage for e-Science (RENKEI)", and JST CREST "Development of System Software Technologies for post-Peta Scale High Performance Computing". KY is supported by grants from the Ministry of Health, Labour and Welfare, Grant-in-Aid for Scientific Research (B) 21390100 and the Takeda Scientific Foundation.

Author details

[1]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki, Nagasaki, Japan. [2]Nagasaki University Global Center of Excellence Program, 1-12-4 Sakamoto, Nagasaki, Nagasaki, Japan. [3]Center for Computational Sciences, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan. [4]Core Research for Evolutional Science and Technology, Japan Science and Technology Agency, 4-1-8 Honcho, Kawaguchi, Saitama, Japan. [5]Departmentent of Computer Science, Graduate School of Systems and Information Engineering, University of Tsukuba, 1-1-1 Tennodai, Tsukuba, Ibaraki, Japan.

Authors' contributions

HM conceived the study, implemented the workflows, and co-authored the manuscript. KS implemented the workflows. MT and OT developed Pwrake and evaluated the details of the workflows and the computational performance. KY conceived the study and co-authored the manuscript. All authors read and approved the final manuscript

Competing interests

The authors declare that they have no competing interests.

Received: 27 May 2011 Accepted: 8 September 2011
Published: 8 September 2011

References

1. Van der Aalst WMP, Ter Hofstede AHM, Kiepuszewski B, Barros AP: **Workflow patterns.** *Distrib Parallel Dat* 2003, **14**:5-51.
2. Zhao Y, Raicu I, Foster I: **Scientific Workflow Systems for 21st Century, New Bottle or New Wine?** *2008 IEEE Congress on Services - Part I* Honolulu, HI, USA; 2008, 467-471.
3. Hoon S, Ratnapu KK, J-ming Chia, Kumarasamy B, Juguang X, Clamp M, Stabenau A, Potter S, Clarke L, Stupka E: **Biopipe: A Flexible Framework for Protocol-Based Bioinformatics Analysis.** *Genome Res* 2003, **13**:1904-1915.
4. Deelman E, Blythe J, Gil Y, Baker C, Mehta G, Vahi K, Blackburn K, Lazzarini A, Arbree A, Cavanaugh R: **Mapping complex scientific workflows onto distributed systems.** *J Grid Comp* 2003, **1**:25-39.
5. Eker J, Janneck JW, Lee EA, Liu J, Liu X, Lidvig J, Neuendorffer S, Sachs S, Xiong Y: **Taming heterogeneity - the Ptolemy approach.** *Proc IEEE* 2003, **91**:127-144.
6. Oinn T, Addis M, Ferris J, Marvin D, Senger M, Greenwood M, Carver T, Glover K, Pocock MR, Wipat A, Li P: **Taverna: a tool for the composition and enactment of bioinformatics workflows.** *Bioinformatics* 2004, **20**:3045-3054.
7. Shah S, He D, Sawkins J, Druce J, Quon G, Lett D, Zheng G, Xu T, Ouellette BF: **Pegasys: software for executing and integrating analyses of biological sequences.** *BMC Bioinformatics* 2004, **5**:40.
8. Ludäscher B, Altintas I, Berkley C, Higgins D, Jaeger E, Jones M, Lee EA, Tao J, Zhao Y: **Scientific workflow management and the Kepler system.** *Concurrency Computat Pract Exper* 2006, **18**:1039-1065.
9. Churches D, Gombas G, Harrison A, Maassen J, Robinson C, Shields M, Taylor I, Wang I: **Programming scientific and distributed workflow with Triana services.** *Concurrency Computat Pract Exper* 2006, **18**:1021-1037.
10. Romano P, Bartocci E, Bertolini G, De Paoli F, Marra D, Mauri G, Merelli E, Milanesi L: **Biowep: a workflow enactment portal for bioinformatics applications.** *BMC Bioinformatics* 2007, **8**:S19.

11. Zhao Y, Hategan M, Clifford B, Foster I, Von Laszewski G, Nefedova V, Raicu I, Stef-Praun T, Wilde M: **Swift: Fast, reliable, loosely coupled parallel computation.** *Proceedings - 2007 IEEE Congress on Services, SERVICES 2007* 2007, 199-206.
12. Bartocci E, Corradini F, Merelli E, Scortichini L: **BioWMS: a web-based Workflow Management System for bioinformatics.** *BMC Bioinformatics* 2007, **8**:S2.
13. Fiers M, van der Burgt A, Datema E, de Groot J, van Ham R: **High-throughput bioinformatics with the Cyrille2 pipeline system.** *BMC Bioinformatics* 2008, **9**:96.
14. Berthold MR, Cebron N, Dill F, Gabriel TR, Kotter T, Meinl T, Thiel K, Wiswedel B: **KNIME - The Konstanz Information Miner.** *SIGKDD Explorations* 2009, **11**:26-31.
15. Orvis J, Crabtree J, Galens K, Gussman A, Inman JM, Lee E, Nampally S, Riley D, Sundaram JP, Felix V, Whitty B, Mahurkar A, Wortman J, White O, Angiuoli SV: **Ergatis: a web interface and scalable software system for bioinformatics workflows.** *Bioinformatics* 2010, **26**:1488-1492.
16. Goecks J, Nekrutenko A, Taylor J, Galaxy Team T: **Galaxy: a comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences.** *Genome Biol* 2010, **11**:R86.
17. Radetzki U, Leser U, Schulze-Rauschenbach SC, Zimmermann J, Lüssem J, Bode T, Cremers AB: **Adapters, shims, and glue–service interoperability for in silico experiments.** *Bioinformatics* 2006, **22**:1137-1143.
18. Lin C, Lu S, Fei X, Pai D, Hua J: **A Task Abstraction and Mapping Approach to the Shimming Problem in Scientific Workflows.** In *Services Computing, IEEE International Conference on. Volume 0.* Los Alamitos, CA, USA: IEEE Computer Society; 2009:284-291.
19. Kane D, Hohman M, Cerami E, McCormick M, Kuhlman K, Byrd J: **Agile methods in biomedical software development: a multi-site experience report.** *BMC Bioinformatics* 2006, **7**:273.
20. Taura K: **Grid Explorer: A Tool for Discovering, Selecting, and Using Distributed Resources Efficiently.** *IPSJ SIG Technical Report* 2004, **2004-HPC-099**:235-240.
21. **Python Programming Language.** [http://www.python.org/].
22. **Ruby Programming Language.** [http://www.ruby-lang.org/].
23. Goodstadt L: **Ruffus: a lightweight Python library for computational pipelines.** *Bioinformatics* 2010, **26**:2778-2779.
24. Cieslik M, Mura C: **A lightweight, flow-based toolkit for parallel and distributed bioinformatics pipelines.** *BMC Bioinformatics* 2011, **12**:61.
25. Cunningham HC: **A little language for surveys: Constructing an internal DSL in Ruby.** *Proceedings of the 46th Annual Southeast Regional Conference on XX, ACM-SE 46* 2008, 282-287.
26. **Rake.** [http://rake.rubyforge.org/].
27. **Pwrake.** [https://github.com/masa16/pwrake].
28. Tanaka M, Tatebe O: **Pwrake: a parallel and distributed flexible workflow management tool for wide-area data intensive computing.** *Proceedings of the 19th ACM International Symposium on High Performance Distributed Computing* New York, NY, USA: ACM; 2010, 356-359.
29. **Graphviz.** [http://graphviz.org/].
30. Tatebe O, Hiraga K: **Gfarm Grid File System.** *New Generat Comput* 2010, **28**:257-275.
31. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA: **The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data.** *Genome Res* 2010, **20**:1297-1303.
32. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ: **A framework for variation discovery and genotyping using next-generation DNA sequencing data.** *Nat Genet* 2011.
33. Albers CA, Lunter G, MacArthur DG, McVean G, Ouwehand WH, Durbin R: **Dindel: Accurate indel calls from short-read data.** *Genome Res* 2010.
34. The 1000 Genomes Project Consortium: **A map of human genome variation from population-scale sequencing.** *Nature* 2010, **467**:1061-1073.
35. Fujimoto A, Nakagawa H, Hosono N, Nakano K, Abe T, Boroevich KA, Nagasaki M, Yamaguchi R, Shibuya T, Kubo M, Miyano S, Nakamura Y, Tsunoda T: **Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing.** *Nat Genet* 2010, **42**:931-936.
36. The International HapMap Consortium: **A haplotype map of the human genome.** *Nature* 2005, **437**:1299-1320.
37. Li H, Durbin R: **Fast and accurate short read alignment with Burrows-Wheeler transform.** *Bioinformatics* 2009, **25**:1754-1760.
38. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup: **The Sequence Alignment/Map format and SAMtools.** *Bioinformatics* 2009, **25**:2078-2079.
39. **Picard.** [http://picard.sourceforge.net/].
40. Gropp W, Lusk E, Doss N, Skjellum A: **A high-performance, portable implementation of the MPI message passing interface standard.** *Parallel Comput* 1996, **22**:789-828.
41. Mishima H, Lidral AC, Ni J: **Application of the Linux cluster for exhaustive window haplotype analysis using the FBAT and Unphased programs.** *BMC Bioinformatics* 2008, **9(Suppl 6)**:S10.
42. Dean J, Ghemawat S: **MapReduce: simplified data processing on large clusters.** *Commun ACM* 2008, **51**:107-113.
43. Aerts J, Law A: **An introduction to scripting in Ruby for biologists.** *BMC Bioinformatics* 2009, **10**:221.
44. Goto N, Prins P, Nakao M, Bonnal R, Aerts J, Katayama T: **BioRuby: Bioinformatics software for the Ruby programming language.** *Bioinformatics* 2010, btq475.
45. Goble CA, Bhagat J, Aleksejevs S, Cruickshank D, Michaelides D, Newman D, Borkum M, Bechhofer S, Roos M, Li P, De Roure D: **myExperiment: a repository and social network for the sharing of bioinformatics workflows.** *Nucleic Acids Res* 2010, **38**:W677-W682.
46. **Wf4ever.** [http://www.wf4ever-project.org/].

doi:10.1186/1756-0500-4-331
Cite this article as: Mishima *et al.*: Agile parallel bioinformatics workflow management using Pwrake. *BMC Research Notes* 2011 **4**:331.

THE JOURNAL OF
Obstetrics and Gynaecology Research

doi:10.1111/j.1447-0756.2011.01601.x

J. Obstet. Gynaecol. Res. Vol. 37, No. 11: 1666–1670, November 2011

Pre-vaccination epidemiology of human papillomavirus infections in Japanese women with abnormal cytology

Kentaro Yamasaki[1], Kiyonori Miura[1], Takako Shimada[1], Rie Ikemoto[6], Shoko Miura[1], Makoto Murakami[7], Tetsuro Sameshima[2], Akira Fujishita[3], Kouhei Kotera[4], Akira Kinoshita[5], Koh-ichiro Yoshiura[5] and Hideaki Masuzaki[1]

[1]*Department of Obstetrics and Gynecology, School of Medicine, Nagasaki University,* [2]*Department of Obstetrics and Gynecology, The Japanese Red Cross Nagasaki Genbaku Hospital,* [3]*Department of Obstetrics and Gynecology, Saiseikai Hospital,* [4]*Department of Obstetrics and Gynecology, Nagasaki Municipal Hospital,* [5]*Department of Human Genetics, School of Medicine, Nagasaki University, Nagasaki,* [6]*SRL Corporation, Fukuoka, and* [7]*Department of Obstetrics and Gynecology, Sasebo Municipal Hospital, Sasebo, Japan*

Abstract

Aim: To investigate the pre-vaccination epidemiology of genital human papillomavirus (HPV) infections and genotypes in women with abnormal cytology in Nagasaki, Japan.

Material and Methods: We performed Pap smear tests, biopsies and HPV genotype testing in Nagasaki Prefecture from August 2007 through November 2009.

Results: During the study period, serial samples of uterine cervical specimens were obtained from 539 subjects with abnormal cytology and/or squamous intraepithelial lesions (SIL) confirmed previously, or with clinically suspected invasive cervical cancer. In 119 HPV-positive subjects with low-grade SIL, the three most prevalent high-risk HPV genotypes were HPV52 (21.8%; 26/119), HPV16 (20.2%; 24/119) and HPV56 (17.6%; 21/119). In 199 women, 127 HPV-positive subjects with high-grade SIL and 67 HPV-positive subjects with squamous cell carcinoma (SCC), the three most prevalent high-risk HPV genotypes were HPV16 (44.3%; 86/194), HPV52 (20.6%; 40/194) and HPV58 (16.0%; 31/194).

Conclusion: Compared with the distribution of high-risk HPV genotypes in other countries, HPV52 was a more common genotype in Nagasaki. With disease progression to SCC, the distribution of high-risk HPV56 belonging to the A6 HPV family decreased, while HPV16 and HPV52 belonging to the A9 HPV family persisted. Our data provide an important resource to address the case for vaccination against HPV genotypes other than HPV16 and HPV18 in Japan.

Key words: epidemiology, genotype, human papillomavirus, infection, uterine cervical neoplasia.

Introduction

Persistent infections with human papillomavirus (HPV) are recognized as a major cause of cervical cancer. Genital infections with HPV are very common, and these infections are transmitted by sexual contact.[1] However, HPV infections in most cases disappear naturally in a relatively short period, and induce little risk of developing disease.[2–4] We do not fully know the pathological mechanism that results in HPV infection developing into invasive cervical cancer (CC). During persistent infection, different viral characteristics along with HPV genotype may be important, such as the distribution of each type in the population and the ability to evade the host's immune system. Another important factor in persistent infection could be related

Received: November 2 2010.
Accepted: January 27 2011.
Reprint request to: Dr Kiyonori Miura, Department of Obstetrics and Gynecology, School of Medicine, Nagasaki University, 1-7-1 Sakamoto, Nagasaki 852-8501, Japan. Email: kiyonori@nagasaki-u.ac.jp

to the host, such as the host immune reaction against a specific HPV genotype, and sexual behavior.

The distribution of infectious high-risk HPV genotypes and the prevalence of CC in women varies worldwide. Clarification of the relation between clinical characteristics of CC and specific HPV genotypes in a local region may lead not only to implementation of a preventive strategy in that region, but also to an elucidation of the natural history of HPV infections compared with other regions in the world. In Japan, data on the distribution of HPV genotypes remains inadequate. To evaluate the possible effect of an HPV vaccine, we require knowledge of the pre-vaccination epidemiology of genital HPV infections. Thus, to determine the distribution and natural history of HPV infections in Nagasaki, Japan, we performed HPV genotype testing, cervical cytology and colposcopic biopsies.

Methods

Study population

The study included 625 subjects with abnormal cytology and/or histologically confirmed squamous intraepithelial lesions (SIL), or with clinically suspected invasive CC who required examination by colposcopy and directed biopsy. Cytology and HPV DNA test samples were collected in five hospitals in Nagasaki Prefecture from August 2007 through November 2009. Exclusion criteria were patients who had received therapeutic excisions previously or who had non-squamous neoplasms confirmed histologically. Thus, 86 subjects were excluded from the study.

The study protocol was approved by the Ethical Review Board of Nagasaki University and the other hospitals involved. All women were informed of the purpose of the study and gave their consent.

Sample collection and pathologic diagnoses

Specimens were collected using a Cervex Brush (Rovers Medical Devices, the Netherlands) and suspended in 10 mL of SurePath preservative fluid (Becton, Dickinson & Co., Franklin Lakes, NJ, USA). We used the samples from the same vial for cytology with the Bethesda III system (2001) and for HPV genotype testing. Cervical specimens for cytology and HPV genotyping were obtained at each visit from participants who received regular follow-up examinations. The cytologic diagnoses of the specimens were performed by experienced cytoscreeners in a commercial laboratory (SRL, Inc., Tokyo, Japan), and they were blinded from the HPV genotyping test. The histopathological review was performed by experienced pathologists of the Division of Pathology at Nagasaki University Hospital.

HPV genotyping test

Genotyping of HPV DNA in the SurePath preservative fluid after preparing glass slides was carried out using the Linear Array HPV Genotyping Test kit (Roche Molecular Systems, Indianapolis, IN, USA). The kit uses the PGMY09/PGMY11 primers[5] to amplify the L1 conserved region. Following polymerase chain reaction amplification, hybridization of the HPV amplicon was performed using an array of oligonucleotide probes that allowed independent identification of individual HPV genotypes. This kit can detect 37 HPV genotypes (6, 11, 16, 18, 26, 31, 33, 35, 39, 40, 42, 45, 51, 52, 53, 54, 55, 56, 58, 59, 61, 62, 64, 66, 67, 68, 69, 70, 71, 72, 73 (MM9), 81, 82 (MM4), 83 (MM7), 84 (MM8), IS39 and CP6108 (89)). For consistency with previous studies, we considered 16 HPV genotypes (16, 18, 31, 33, 35, 45, 51, 52, 53, 56, 58, 59, 66, 68, 73 and 82) as high-risk genotypes, which are related to CC based on previous reports.[6–8]

Results

Uterine cervical specimens for cytology and HPV DNA testing were collected from 539 women, with a mean age of 42 years in the age range of 19–94, with abnormal Pap smears and/or previously confirmed squamous intraepithelial lesions (SIL) or with clinically suspected invasive CC. In 154 subjects who were negative for intraepithelial lesion or malignancy (NILM), HPV was positive in 67 women (43.5%), with a mean age of 41 years at their initial HPV DNA test. The three most prevalent high-risk HPV genotypes in the NILM group were HPV 52 (23.9%; 16/67), HPV 16 (23.9%; 16/67) and HPV 71 (11.9%; 8/67) (Fig. 1a). In 125 subjects with cytologically low-grade SIL (LSIL), HPV was positive in 119 women (95.2%), with a mean age of 40 years at their initial test. The three most prevalent high-risk HPV genotypes were HPV 52 (21.8%; 26/119), HPV 16 (20.2%; 24/119) and HPV 56 (17.6%; 21/119) (Fig. 1b). In 128 participants diagnosed with cytologically high-grade SIL (HSIL), HPV infection was present in 127 women (99.2%), with a mean age of 41 years. In 71 women diagnosed cytologically with squamous cell carcinoma (SCC), HPV infection was positive in 67 women (94.4%), with a mean age of 56 years. In these latter two groups, the three most prevalent high-risk HPV genotypes were HPV 16

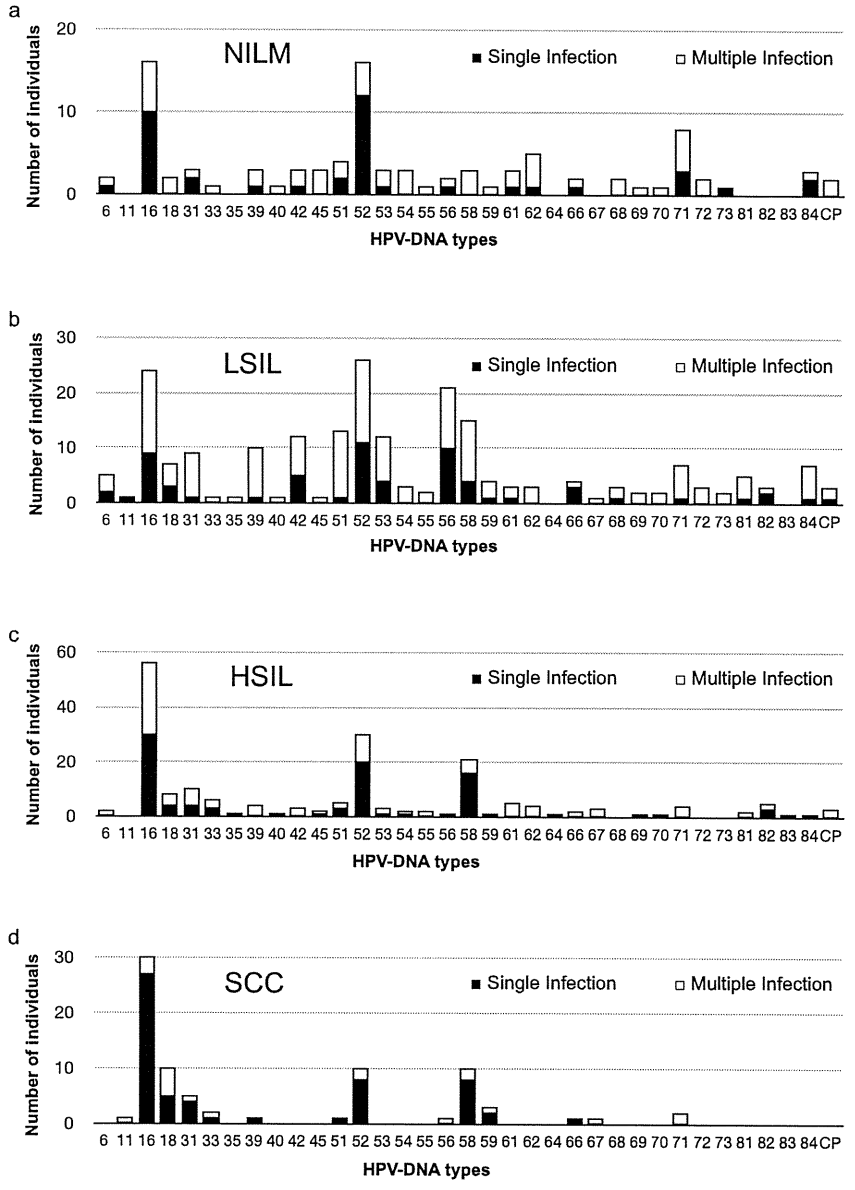


Figure 1 Prevalence of human papillomavirus (HPV) genotype among participants who were diagnosed with NILM (a), participants who were diagnosed with LSIL (b), participants who were diagnosed with HSIL (c) and with SCC (d). Closed boxes show infection with a single type of HPV DNA, and open boxes show multiple infections with two or more HPV-DNA types. NILM, negative for intraepithelial lesion or malignancy; LSIL, low-grade squamous intraepithelial lesion; HSIL, high-grade squamous intraepithelial lesion; SCC, squamous cell carcinoma; CP: CP6108 (HPV 89).

(44.3%; 86/194), HPV 52 (20.6%; 40/194) and HPV 58 (16.0%; 31/194) (Fig. 1c,d). The results of the other 61 subjects who were diagnosed with atypical squamous cells of undetermined significance (ASC-US) and atypical cells-cannot exclude HSIL (ASC-H) were not considered in this report. However, the HPV genotype distribution of the ASC-US group was similar to that of the LSIL group. The number of participants in the ASC-H group was too small to determine the distribution of HPV genotypes.

The Table 1 shows the multiple HPV infection status by cytological diagnosis. The percentage of single HPV infection was significantly higher in the SCC group (86.6%; 58/67) than in the LSIL group (53.8%; 64/119) ($P < 0.01$, Fisher's exact test).

Figure 2 shows a histogram of patient age distribution and the number of single infections of HPV 16, HPV 18 and HPV 52 in patients diagnosed histologically with cervical intraepithelial lesions grade 3 (CIN3) and invasive CC. The mean age of patients infected with HPV 52 tended to be higher than that of patients with HPV 16 infection but the difference was not statistically significant ($P = 0.07$, Student's *t*-test).

Table 1 Multiple human papillomavirus (HPV) infection status by cytologic diagnosis

Cytology	*n*	Single type	Two types	Three types	More than four types
LSIL	119	53.8% (64)	24.4% (29)	10.1% (12)	11.8% (14)
HSIL	127	67.7% (86)	22.8% (29)	5.5% (7)	3.9% (5)
SCC	67	86.6% (58)	10.4% (7)	3.0% (2)	0

The table shows the percentage and the number of participants who were infected with a single HPV genotype, and two, three or more than four HPV genotypes.

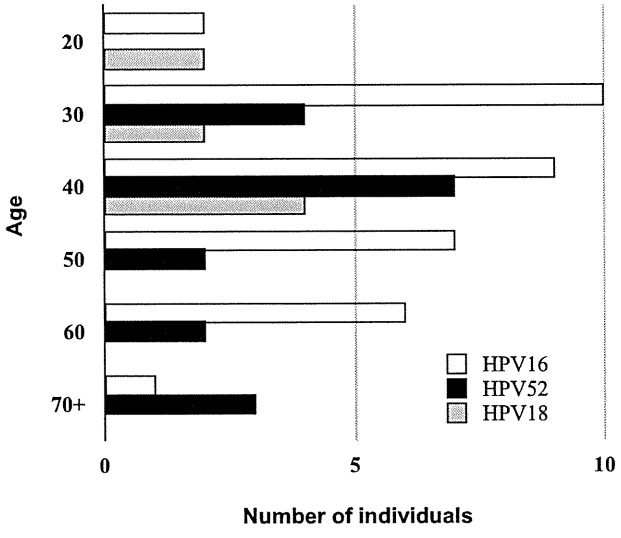


Figure 2 Distribution of ages in CIN3 and SCC patients who were infected with one HPV type, HPV 16, HPV 18 or HPV 52. Open boxes show the number of patients infected with only HPV 16, closed gray boxes show the number of patients infected with only HPV 18 and closed black boxes show the number of patients infected with only HPV 52. 20: 20–29 years; 30: 30–39 years; 40: 40–49 years; 50: 50–59 years; 60: 60–69 years; 70+: ≥70 years; CIN3: cervical intraepithelial lesions grade 3.

Discussion

The distribution of HPV genotypes in the LSIL group suggests that HPV 52 is the most frequently observed genotype among subjects with persistent HPV infections in Nagasaki. Other investigators also have reported that HPV 52 was dominant among women with normal cytology or cervical neoplastic lesions in Japan.[8–10] In the general population, the prevalence of HPV genotypes exhibit geographic differences in different countries, though HPV 16 is found to be most prevalent worldwide.[7,11] In pre-neoplastic and cancer cases, the geographic differences in prevalence of HPV genotype are diminished and HPV 16 tends to be the most dominant all over the world. HPV 18 and HPV 31 infections have also been reported to show higher prevalence in CIN and CC patients, but in the current study, there was a low prevalence of HPV 18 in the HSIL group. In the SCC group, the prevalence of HPV 18 was similar to that of HPV 52 and 58, although there was a lower number of single infection cases. Because four cases of multiple infections included HPV 18 as well as HPV 16 and 52 infections, the contribution of HPV 18 infection in the SCC group was difficult to evaluate.

In the LSIL group and the HSIL-SCC group, the distribution of HPV genotypes was different; the most marked differences between the HSIL-SCC group and LSIL group were a more than doubling of the HPV 16 genotype and the disappearance of HPV 56 infection in the former. HPV 16, HPV 52 and HPV 58 belong to the same alpha-papillomavirus species no. 9 family (A9 HPV family), which also includes HPV 31, 33, 35 and 67. However, HPV 56 belongs to the A6 HPV family, which also includes HPV 53 and 66.[12] The results indicated that the prevalence of the A6 HPV family was not small, especially in the Nagasaki LSIL group, but this HPV family was less likely to continue into persistent infection, and the observed prevalence of HPV 56 infection was found to be reduced in the HSIL and SCC groups.

Interestingly, the LSIL group had the lowest single infection rate of HPV (53.8%) (Table 1) and the rate of single infection was higher (67.7%) in the HSIL group. The SCC group showed the highest rate of single infection (86.6%). This finding has also been reported by other investigators and was suggested to support a monoclonal origin for cancer.[10]

We analyzed samples from patients with CIN3 and invasive SCC histologically and counted the number of patients in each age group, 20–29, 30–39, 40–49, 50–59, 60–69 and 70-plus. The most prevalent and dangerous HPV genotypes appeared to be HPV 16 and HPV 18, but the degree of risk of HPV 16/18 remained to be quantified. The histogram in Figure 2 shows the

number of individuals who had a single infection with HPV 16, HPV 18 or HPV 52. The mean age and distribution of ages among HPV 16 and 52 types was different (paired *t*-test, $P = 0.07$), but HPV 52 appeared to be associated with slower progression of carcinogenesis, and HPV 16 and 18 with faster progression. The differences in HPV genotypes may be related not only to development of persistent infection, but also to the speed of progression to SCC.

In Japan, one commercial CC vaccine became available in December 2009. Although this study has some limitations because we included data only from SIL/CC women, our pre-vaccination data on the distribution of genital HPV infections in a region where HPV 52 and 58 are prevalent is valuable to determine the potential usefulness of a bivalent HPV vaccine. Paavonen *et al.* reported an estimated cross-reactivity against CIN2+ lesions with non-vaccine oncogenic HPV types of 37–54%.[13] Further study of the distribution of HPV genotypes in a SIL/CC population and the transition of pathological changes in patients according to HPV genotype is warranted.

Acknowledgments

This study was supported by Nagasaki University President's Fund Grant. K. Yamasaki was supported by a Grant-in-Aid for Young Scientists (B) (No. 20791150) from the Ministry of Education, Sports, Culture, Science and Technology of Japan. K. Miura was supported by a Grant-in-Aid for Young Scientists (B) (No. 21791567) from the Ministry of Education, Sports, Culture, Science and Technology of Japan, Grants from a Grant for Child Health and Development (No. 20C-1) from the Ministry of Health, Labour and Welfare, Japan. K. Yoshiura was supported by Grants-in-Aid for Scientific Research from the Ministry of Health, Labour and Welfare, Japan.

References

1. Shimada T, Miyashita M, Miura S *et al.* Genital human papilloma virus infection in mentally-institutionalized virgins. *Gynecol Oncol* 2007; **106**: 488–489.
2. Moscicki AB, Palefsky J, Smith G, Siboshski S, Schoolnik G. Variability of human papillomavirus DNA testing in a longitudinal cohort of young women. *Obstet Gynecol* 1993; **82** (Pt 1): 578–585.
3. Woodman CB, Collins S, Winter H *et al.* Natural history of cervical human papillomavirus infection in young women: A longitudinal cohort study. *Lancet* 2001; **357**: 1831–1836.
4. Ho GY, Bierman R, Beardsley L, Chang CJ, Burk RD. Natural history of cervicovaginal papillomavirus infection in young women. *N Engl J Med* 1998; **338**: 423–428.
5. Gravitt PE, Peyton CL, Alessi TQ *et al.* Improved amplification of genital human papillomaviruses. *J Clin Microbiol* 2000; **38**: 357–361.
6. Walboomers JM, Jacobs MV, Manos MM *et al.* Human papillomavirus is a necessary cause of invasive cervical cancer worldwide. *J Pathol* 1999; **189**: 12–19.
7. Munoz N, Bosch FX, de Sanjose S *et al.* Epidemiologic classification of human papillomavirus types associated with cervical cancer. *N Engl J Med* 2003; **348**: 518–527.
8. Asato T, Maehama T, Nagai Y, Kanazawa K, Uezato H, Kariya K-I. A large case-control study of cervical cancer risk associated with human papillomavirus infection in Japan, by nucleotide sequencing-based genotyping. *J Infect Dis* 2004; **189**: 1829–1832.
9. Miura S, Matsumoto K, Oki A *et al.* Do we need a different strategy for HPV screening and vaccination in East Asia? *Int J Cancer* 2006; **119**: 2713–2715.
10. Inoue M, Sakaguchi J, Sasagawa T, Tango M. The evaluation of human papillomavirus DNA testing in primary screening for cervical lesions in a large Japanese population. *Int J Gynecol Cancer* 2006; **16**: 1007–1013.
11. de Sanjose S, Diaz M, Castellsague X *et al.* Worldwide prevalence and genotype distribution of cervical human papillomavirus DNA in women with normal cytology: A meta-analysis. *Lancet Infect Dis* 2007; **7**: 453–459.
12. de Villiers EM, Fauquet C, Broker TR, Bernard HU, zur Hausen H. Classification of papillomaviruses. *Virology* 2004; **324**: 17–27.
13. Paavonen J, Naud P, Salmeron J *et al.* Efficacy of human papillomavirus (HPV)-16/18 AS04-adjuvanted vaccine against cervical infection and precancer caused by oncogenic HPV types (PATRICIA): Final analysis of a double-blind, randomised study in young women. *Lancet* 2009; **374**: 301–314.

Eur Respir J 2011; 38: 861–869
DOI: 10.1183/09031936.00143610

Surfactant protein C G100S mutation causes familial pulmonary fibrosis in Japanese kindred

S. Ono[*,#,§§], T. Tanaka[¶,§§], M. Ishida[¶], A. Kinoshita[*], J. Fukuoka[+], M. Takaki[¶], N. Sakamoto[§], Y. Ishimatsu[§], S. Kohno[§], T. Hayashi[f], M. Senba[**], M. Yasunami[##], Y. Kubo[¶¶], L.M. Yoshida[¶], H. Kubo[++], K. Ariyoshi[¶], K. Yoshiura[*] and K. Morimoto[¶]

ABSTRACT: **Several mutations in the surfactant protein C (SP-C) gene (*SFTPC*) have been reported as causing familial pulmonary fibrosis (FPF). However, the genetic background and clinical features of FPF are still not fully understood.**

We identified one Japanese kindred, in which at least six individuals over three generations were diagnosed with pulmonary fibrosis. We examined the patients radiologically and histopathologically and sequenced their *SFTPC* and *ABCA3* genes. We also established a cell line stably expressing the mutant gene.

All the patients had similar radiological and histopathological characteristics. Their histopathological pattern was that of usual interstitial pneumonia, showing numerous fibroblastic foci even in areas without abnormal radiological findings on chest high-resolution computed tomography. No child had respiratory symptoms in the kindred. Sequencing of *SFTPC* showed a novel heterozygous mutation, c.298G>A (G100S), in the BRICHOS domain of proSP-C, which co-segregated with the disease. However, in the *ABCA3* gene, no mutation was found. *In vitro* expression of the mutant gene revealed that several endoplasmic reticulum stress-related proteins were strongly expressed.

The mutation increases endoplasmic reticulum stress and induces apoptotic cell death compared with wild-type SP-C in alveolar type II cells, supporting the significance of this mutation in the pathogenesis of pulmonary fibrosis.

KEYWORDS: **Endoplasmic reticulum stress, familial pulmonary fibrosis, mutation, surfactant protein C**

Familial pulmonary fibrosis (FPF) is characterised by cases of idiopathic interstitial pneumonia in two or more first-degree relatives [1]. MARSHALL *et al.* [1] estimated that familial cases account for 0.5–2.2% of all individuals with idiopathic pulmonary fibrosis (IPF). Several kindreds with FPF have been reported, and the familial form is likely to be transmitted in an autosomal dominant inheritance mode [1–3]. Recent studies have revealed that several cases of FPF are associated with mutations in *SFTPC* [4, 5]. *SFTPC*, located at 8p21.3, has six exons and encodes the hydrophobic peptide surfactant protein C (SP-C). The first reported *SFTPC* mutation, IVS4+1G>A, located at the first base of intron 4, disrupted the donor splice site and resulted in the skipping of exon 4 and the deletion of 37 amino acids from the C-terminal region of the proprotein of SP-C (proSP-C) [6]. 26 *SFTPC* mutations have since been identified [6–19] (online supplementary table 1), all of which are heterozygous mutations in affected individuals. However, only a few reports have described familial cases including several affected individuals (table 1) [6–11].

SP-C is synthesised as a 197-amino acid proSP-C, which undergoes multiple processing steps to form mature SP-C. It is finally released into the alveoli in association with other surfactant proteins and phospholipids [4, 5, 20]. Mature SP-C, consisting of 35 amino acids corresponding to Phe24–Leu58 of proSP-C, is encoded within exon 2 of *SFTPC* and is stored in the lamellar body, from where it is secreted into the alveolar space. In the

AFFILIATIONS
[*]Depts of Human Genetics,
[#]Psychiatry, Nagasaki University Graduate School of Biomedical Sciences,
[¶]Depts of Clinical Medicine,
[**]Pathology,
[##]Immunogenetics, and
[¶¶]Preventive Medicine and AIDS Research, Institute of Tropical Medicine, Nagasaki University,
[§]Second Dept of Internal Medicine, Nagasaki University School of Medicine,
[f]Dept of Pathology, Nagasaki University Hospital, Nagasaki,
[+]Dept of Surgical Pathology, Toyama University Hospital, Toyama, and
[++]Dept of Advanced Preventive Medicine for Infectious Disease, Tohoku University Graduate School of Medicine, Sendai, Japan.
[§§]These authors contributed equally to this study.

CORRESPONDENCE
K. Morimoto
1-12-4 Sakamoto
Nagasaki 852-8523
Japan
E-mail: komorimo@nagasaki-u.ac.jp

Received:
Sept 09 2010
Accepted after revision:
March 10 2011
First published online:
Aug 04 2011

European Respiratory Journal
Print ISSN 0903-1936
Online ISSN 1399-3003

This article has supplementary material accessible from www.erj.ersjournals.com

TABLE 1 Published surfactant protein C mutations found in large families

Mutation	First author [ref.]	Families with ILD n	Pathology
Met71Val	VAN MOORSEL [11]	1	UIP/DIP adults
Ile73Thr	CAMERON [8]	3	NSIP
	ABOU TAAM [10]	1	Unspecified
	VAN MOORSEL [11]	3	UIP adults
			NSIP/DIP child
IVS4+1,G>A	NOGEE [6]	1	NSIP child
			DIP/UIP adults
IVS4+2,T>C	VAN MOORSEL [11]	1	NSIP/DIP children
			UIP adults
Leu188Gln	THOMAS [7]	1	NSIP children
			DIP/UIP adults
Cys189Tyr	GUILLOT [9]	1	NSIP
Leu194Pro	GUILLOT [9]	1	NSIP

ILD: interstitial lung disease; UIP: usual interstitial pneumonia; DIP: desquamative interstitial pneumonia; NSIP: nonspecific interstitial pneumonia.

lung, proSP-C is expressed only in alveolar type II epithelial cells. The N-terminus of proSP-C is in the cytosol, with the mature SP-C domain anchoring it in the membrane [4, 5, 20]. Furthermore, proSP-C contains a domain known as BRICHOS, which is thought to be involved in proteolytic processing and protecting the peptide from aggregation [21], corresponding to residues Phe94–Ile197 in the C-terminal domain of proSP-C. About three-quarters of all mutations that have been reported in *SFPTC* from interstitial lung diseases are in the BRICHOS domain. It has been reported that a BRICHOS mutant protein increased the amount of insoluble aggregates and resulted in apoptosis following an ER stress response [22].

The current study investigated the clinical features of one Japanese FPF kindred with a heterozygous mutation, G100S, in the BRICHOS domain of proSP-C (SP-C^{G100S}).

MATERIALS AND METHODS

Subjects

Pedigree and DNA samples

The family we encountered is shown in figure 1a. Patient IV-1 was the proband, a Japanese female who was referred to our hospital at the age of 18 yrs for further assessment of an abnormal shadow in the lung field that was noticed at a school medical health check. Patient IV-2 is a younger brother of the proband (fig. 1a). In a routine preoperative chest radiological examination (orthopaedic surgery for congenital dysplasia of the femur), abnormal chest shadows were noticed, and further analysis was performed in our department after the operation at the age of 16 yrs. Patient IV-3 is a younger sister of the proband. After the family history of the proband was taken, we assessed patient IV-3's chest by radiographic examination at the age of 14 yrs, in accordance with her and her father's requests. The three individuals in generation IV were delivered without any problem and showed normal development. None of them had histories of coughing, shortness of breath or environmental exposures, and all were free from other respiratory symptoms. Patient III-3 was the proband's mother. She had been diagnosed with IPF at age 34 yrs; she died at age 41 yrs from lung fibrosis. Patient II-4 is a grandmother of the proband who developed a cough at age 63 yrs and was diagnosed with interstitial pneumonia. Patient III-1 is an aunt of the proband. She had no respiratory symptoms. After the family history was taken, we performed a chest radiological examination on patient III-1, at her request. It was supposed that three more family members, I-1, II-1 and II-2, died from lung disease at ages between 35 and 45 yrs following a few years of illness.

Written informed consent was obtained from the patients and their family members before they participated in this study. Genetic counselling was given to patients before and after genetic analyses. Genomic DNA was extracted from individuals' peripheral blood (II-3, II-4, III-1, III-2, IV-1, IV-2 and IV-3) or from formalin-fixed paraffin-embedded lung tissue (III-3) using a QIAamp DNA Mini Kit (Qiagen, Hilden, Germany). This study was approved by the institutional review boards of Nagasaki University (Nagasaki, Japan).

Lung biopsy and lung histopathology

For histopathological diagnosis, a lung biopsy was performed by video-assisted thoracic surgery (VATS) under general anaesthesia. Tissue sections were prepared from formalin-fixed paraffin-embedded samples. Haematoxylin and eosin-stained sections were prepared following conventional procedures. Pathology slides were observed by two trained pulmonary pathologists.

Mutation analysis

We performed PCR-based mutation analysis of *SFTPC* (National Center for Biotechnology Information (NCBI) Reference Sequence NM_003018.3) from eight specimens, composed of six affected individuals (II-4, III-1, III-3, IV-1, IV-2 and IV-3) and two unaffected individuals (II-3 and III-2). Subsequently, we also sequenced ATP-binding cassette, sub-family A member 3 (*ABCA3*; NM_001089), which was postulated to be a gene that modifies the disease severity of FPF caused by *SFTPC* mutations [23]. All exons and intron–exon boundaries of the two genes were sequenced on a 3130xl automated sequencer (Applied Biosystems, Foster City, CA, USA) using BigDye Terminator version 3.1 (Applied Biosystems). DNA sequences were analysed using Variant Reporter and Sequencing Analysis (Applied Biosystems). Genomic sequences were obtained from the University of California, Santa Cruz (UCSC) genome browser (http://genome.ucsc.edu/; assembly: March 2006; NCBI36/hg18). PCR primers were designed with the assistance of Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3.cgi). Primer sequences are available from the authors on request.

In silico *analysis*

The determination of whether an amino acid substitution is a recognised polymorphism was carried out using the dbSNP database (www.ncbi.nlm.nih.gov/SNP/). Predicted protein functions caused by an amino acid substitution were examined using PolyPhen (http://genetics.bwh.harvard.edu/pph/) and SIFT (http://sift.jcvi.org/). Comparisons of genomic alignments of human and other species were accessed using online software, the Evolutionary Conserved Regions (ECR) browser (http://ecrbrowser.dcode.org/) and the UCSC genome browser.

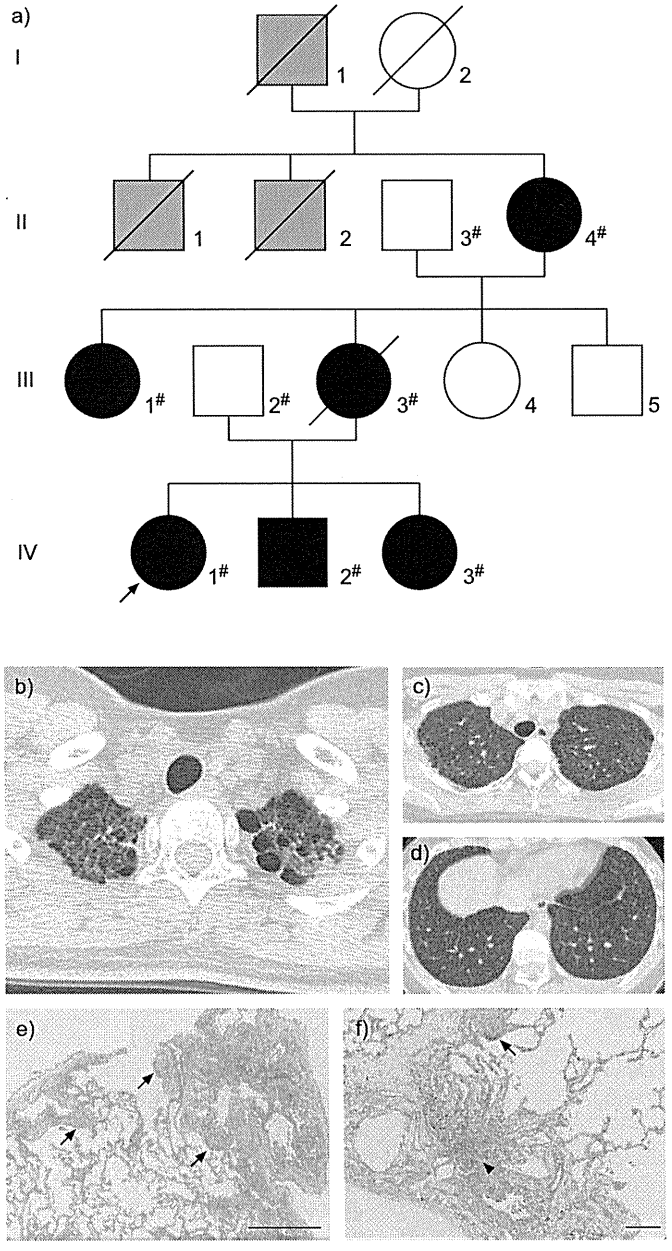


FIGURE 1. Pedigree of a family with familial pulmonary fibrosis (FPF), and radiological and histopathological findings of the proband. a) Pedigree of the family with FPF. Squares: males; circles: females; black: individuals diagnosed with pulmonary fibrosis; grey: individuals who had died due to respiratory failure, but about whom detailed information was not available; #: individuals whose DNA was available and used in direct sequencing; arrow: proband. b) High-resolution computed tomography (HRCT) image of the proband. Reticulonodular opacity, predominantly in both upper lung fields, and intralobar opacity in the subpleural area were observed in the HRCT image from the proband. No honeycombing lesions could be seen. c, d) Magnified CT images of the proband. e and f) Haematoxylin and eosin-stained tissue samples from the proband (right lung S8). Haematoxylin and eosin staining revealed a usual interstitial pneumonia pattern, including patchy peripheral accentuated fibrosis, marked fibroblastic foci (arrows), smooth muscle hyperplasia (arrow head) and abrupt changes to adjacent normal lung areas. Scale bars: e) 1 mm; f) 200 μm. Biopsies were performed from right lung S2 and S8. Pathological findings were similar in S2 and S8.

Functional analysis of mutant protein

SP-C cDNA constructs

A cDNA encoding the full-length human SP-C ($SP\text{-}C^{1-197}$) was cloned into the pcDNA3.1 vector (Invitrogen, Carlsbad, CA) to generate $SP\text{-}C^{1-197}$/pcDNA3.1. A QuikChange® II Site-Directed Mutagenesis Kit (Stratagene, Santa Clara, CA, USA) was used to generate mutant $SP\text{-}C^{G100S}$ in a single PCR with two primers: 5′-ATCGGCTCCACTAGCCTCGTGGTGT-3′ (forward) and 5′-ACACCACGAGGCTAGTGGAGCCGAT-3′ (reverse). The mutation site is underlined.

Cell culture and transfection

A human embryonal kidney (HEK) 293T cell line was obtained from the American Type Culture Collection (ATCC; Manassas, VA, USA) and cultured in Dulbecco's modified Eagle's medium (Gibco, Carlsbad, CA, USA) at 37 °C in 5% CO_2. The culture media were supplemented with 10% fetal bovine serum (Biofluids, Rockville, MD, USA). A549 cells (ATCC) over-expressing wild-type ($SP\text{-}C^{WT}$) or mutant proSP-C were constructed as follows. HEK293T cells were transiently transfected with murine leukaemia virus gag–pol (2 μg) (TaKaRa Bio, Shiga, Japan), proSP-C-encoding retroviral vector (2 μg), and VSV-G expression plasmids (2 μg), which were obtained from L. Chang through the AIDS Research and Reference Reagent Program (Division of AIDS, National Institute of Allergy and Infectious Diseases, Bethesda, MD, USA) [24] using the FuGene HD reagent (30 μL) (Roche Applied Science, Mannheim, Germany). The cells were washed 24 h after transfection and cultured for 24 h in fresh medium. Culture supernatant of the transfected cells was inoculated into A549 cells. The inoculated cells were selected by puromycin (2.5 $\mu g \cdot mL^{-1}$). The puromycin-resistant cell pool was utilised in this study.

RNA isolation and real-time RT-PCR

Total RNA from the stably transfected A549 cells was isolated using a FastPure RNA Kit (TaKaRa Bio) and reverse transcribed into cDNA using a PrimeScript RT Reagent Kit with gDNA Eraser (TaKaRa Bio). We performed real-time quantitative RT-PCR using Thunderbird SYBR qPCR Mix reagent (Toyobo, Osaka, Japan). PCR amplification was run on a LightCycler 480 Real-Time PCR system (Roche Diagnostics, Mannheim, Germany). All samples were measured in triplicate.

Western blot analysis

Cells were solubilised in RIPA buffer with PhosSTOP Phosphatase inhibitor cocktail (Roche Applied Science). Cells treated with proteasome inhibitor MG-132 (Merck Ltd, Lutterworth, UK) for 16 h were also solubilised in the same manner. Total protein extracts were separated by 5–15% Tris-HCl gel (BioRad Laboratories, Richmond, CA, USA) electrophoresis and transferred to polyvinylidene fluoride membranes. The membranes were blocked in blocking buffer (1×PBS, 0.1% Tween-20 with 5% weight/volume nonfat dry milk) for 1 h at room temperature and incubated with primary antibodies at 4°C overnight. After washing in 1×PBS with 0.1% w/v Tween-20, membranes were incubated with horseradish peroxidase-linked secondary antibodies for 1 h at room temperature. Detection was performed by enhanced chemiluminescence with ECL-Plus (GE Health Care, Little Chalfont, UK). Primary antibodies to BiP, IRE1α and cleaved caspase-3 were purchased from Cell Signaling Technology (Danvers, MA, USA). Anti-phosphorylated PERK

(phospho-PERK) antibody was purchased from Santa Cruz Biotechnology (Santa Cruz, CA, USA). β-actin was measured as a loading control for each sample using anti-β-actin antibodies (Santa Cruz Biotechnology).

Statistics

Data are presented as mean±SE. The *t*-statistic was used to determine significant differences between two groups. One-way ANOVA was used to determine significant differences among groups.

RESULTS

Clinical presentation of patients

High-resolution computed tomography (HRCT) findings of the proband, patient IV-1, revealed a reticulonodular shadow and intralobular fine linear opacity predominantly in both upper lung fields. Centrilobular micronodule lesions were observed mainly in subpleural lesions (fig. 1b–d). The HRCT findings of patients II-4, III-1, III-3, IV-2 and IV-3 are presented in online supplementary figure 1a and are similar to those found in the proband. All affected individuals showed similar radiological findings, *i.e.* upper lung field dominant shadow. Additionally, IV-1, IV-2 and IV-3 showed moderate cystic changes, mainly in the upper lobes, as shown in a previous report of adult FPF [11].

VATS lung biopsy was performed for diagnosis and pathological assessment. Haematoxylin and eosin-stained samples from the proband showed features of the usual interstitial pneumonia (UIP) pattern with marked fibroblastic foci and mild infiltration of lymphoid cells (fig. 1e and f). In addition, mild-to-moderate airway-centred fibrosis/inflammation, along with peribronchiolar metaplasia, were observed. No granulomas were seen. Interestingly, all histological samples from the patients (III-3, IV-2 and IV-3) showed a similar UIP pattern (online supplementary fig. 2).

The clinical findings and information are summarised in table 2. Briefly, for the proband and her siblings, serum biomarkers, pulmonary function and respiratory condition were almost normal, and no airway inflammation was observed in their bronchoalveolar lavage fluid (BALF). Because they had kept pet birds in their home, we measured serum antibodies to avian antigen, but those were negative for these siblings. Thus, chronic hypersensitivity pneumonitis was clinically ruled out. Based on radiopathological findings and family history, familial interstitial pneumonia was diagnosed.

Mutation analysis and in silico analysis

Two genes, *SFTPC* and *ABCA3*, were analysed. We detected a base alteration, c.298G>A, in exon 3 of *SFTPC* causing a GGC-to-AGC change that results in a glycine-to-serine change at codon 100 (fig. 2a). This variant segregated with the disease in this family (fig. 2b) and was not present among 576 ethnically matched control alleles. The ECR browser and the UCSC genome browser indicated that codon 100 of *SFTPC* is conserved among mammals (fig. 2c). Furthermore, *in silico* analysis using SIFT and Polyphen predicted a damaging effect on the protein by this one amino acid change (position-specific independent counts score 1.722; SIFT score 0.03).

Expression of proSP-C in A549 cells

To prove comparable expression of proSP-C, we performed western blotting of cell lysates of SP-C^{WT} and SP-C^{G100S} stably expressed A549 cells. The amount of proSP-C was increased in A549 cells stably expressing SP-C^{G100S} compared with those stable expressing SP-C^{WT} (fig. 3a, b). However, the expression levels of SP-C mRNA from these two cell pools assessed by real-time quantitative RT-PCR were equivalent (fig. 3d).

SP-C^{G100S} causes endoplasmic reticulum stress, resulting in apoptosis

We performed western blotting analysis to detect the expression of proSP-C, BiP, phospho-PERK, IRE1α and cleaved caspase-3 to determine whether the expression of the SP-C^{G100S} induces endoplasmic reticulum (ER) stress in epithelial cells compared with SP-C^{WT}. The activations of BiP, IRE1α and cleaved caspase-3 were increased in mutant cells compared with wild-type cells (fig. 3c and e). After MG-132 treatment, A549 cells stably expressing SP-C^{G100S} showed increases in expression BiP, phospho-PERK, IRE1 and cleaved caspase-3 that significantly exceeded the increases seen in A549 cells stably expressing SP-C^{WT} (fig. 3c and f).

DISCUSSION

In the present study, we have described a novel pathogenic *SFTPC* variant, which is associated with FPF in a Japanese kindred who had abnormal HRCT findings at ages ranging from the mid-second to the fifth decade of life. This pedigree included six individuals with similar radiological findings and histopathological characteristics of the UIP pattern. Notably, all the patients were asymptomatic until they were age at least 15 yrs, and there was no child with respiratory symptoms. Furthermore, we also verified that expression of the mutant protein, SP-C^{G100S}, resulted in caspase-3 activation following the induction of ER stress.

Glycine at codon 100 of *SFTPC*, which was mutated to serine in this kindred, is in the BRICHOS domain of proSP-C. This mutation is novel and is the first reported pathogenic mutation of *SFTPC* in an Asian kindred, proving that pulmonary fibrosis caused by *SFTPC* mutations is a worldwide phenomenon. Recent reports showed that the BRICHOS domain of proSP-C has chaperone-like properties that prevent the transmembrane region of proSP-C from aggregating. Mutations of this region in proSP-C triggered induction of intracellular aggregate formation, ER stress and accumulation in endosomal–lysosomal compartments [22, 25, 26]. To further characterise the mutant protein SP-C^{G100S}, we showed that unfolded protein response (UPR) proteins, including BiP (chaperone proteins), phospho-PERK and IRE1α (proximal sensor for UPR), were upregulated in A549 cells stably transformed with SP-C^{G100S}, eventually resulting in apoptotic cell death. These results are consistent with previous observations in several studies of other mutations in the BRICHOS domain, including SP-$C^{\Delta exon4}$ and SP-C^{L188Q} [22, 26, 27, 28]. Recently, SISSON *et al.* [29] reported that targeted injury of type II alveolar epithelial cells induced pulmonary fibrosis in mice [29]. Collectively, these observations lead us to conclude that SP-C^{G100S} is a pathogenic mutation leading to cell death, which leads to pulmonary fibrosis. Categorising *SFTPC* mutations inducing lung fibrosis by functional analysis of the mutant protein might help in tailoring treatment for IPF patients. ROSEN and WALTZ [16] have reported that hydroxychloroquine was

TABLE 2 Patient profile and laboratory data

	Patient					
	II-4	III-1	III-3#	IV-1	IV-2	IV-3
Sex	Female	Female	Female	Female	Male	Female
Current age yrs	68	46	41(died)	18	16	14
Age of diagnosis yrs	66	46	34	18	16	14
Age of first evidence yrs	57	44	34	18	16	14
Age of first symptoms yrs	63	None	34	None	None	None
Serum biomarker						
KL-6 $U \cdot mL^{-1}$ (normal range <500)	1560	386	NA	245	309	332
LDH $IU \cdot L^{-1}$ (normal range 119–229)	303	147	166	161	141	144
Pulmonary function tests						
VC % pred (normal range >80)	42.5	101.9	65.3	72.2	85	96.6
FEV_1 % pred (normal range >70)	92.9	92.8	83.3	84.1	90.3	85
D_{LCO} % pred (normal range >80)	38.5	72.2	NA	69.3	NA	65.2
Blood gas analysis (room air)						
P_{a,O_2} Torr (normal range 75–100)	68.3¶	83.7	90.6	113	109	111
$P_{(A\text{-}a),O_2}$ Torr (normal range <10)	24.2¶	16.2	6.1	-8.5	-14	-10.8
BAL						
Cell count $10^5 \cdot mL^{-1}$	1.21	NA	3.85	2.4	2	1.4
Alveolar macrophages %	54.2	NA	80	90	86	91
Lymphocytes %	10.1	NA	17.3	7.5	12	5.8
Neutrophils %	34.5	NA	1.1	2.5	1	2.4
Eosinophils %	1.2	NA	1.6	0	1	0.8
CD4/CD8 ratio	0.25	NA	0.6	1.7	1.6	1.5
Histological pattern	UIP	NA	UIP	UIP	UIP	UIP

LDH: lactate dehydrogenase; VC: vital capacity; % pred: % predicted; FEV_1: forced expiratory volume in 1 s; D_{LCO}: diffusing capacity of the lung for carbon monoxide; P_{a,O_2}: arterial oxygen tension; $P_{(A\text{-}a),O_2}$: alveolar–arterial oxygen tension difference; BAL: bronchoalveolar lavage; NA: not available: UIP: usual interstitial pneumonia. #: data from patient III-3 are based on those from the first diagnosis; ¶: data from the time of first diagnosis.

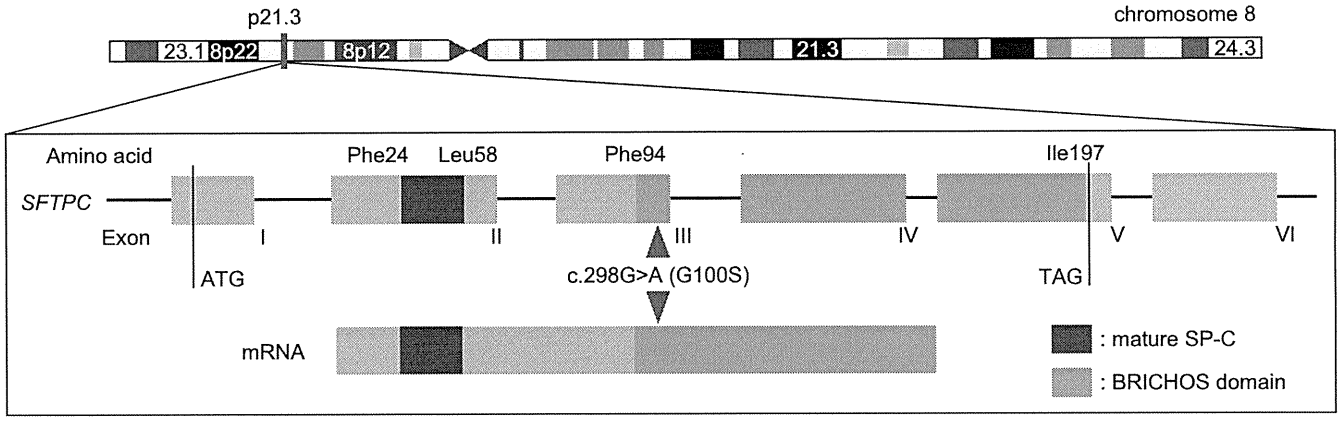
useful in treating a case of interstitial lung disease in a child with an *SFTPC* mutation in the BRICHOS domain [16]. They predicted that hydroxychloroquine caused inhibition of the intracellular processing of proSP-C, thereby reducing the dominant negative effect elicited by mutant proSP-C. It is possible that the suitability of a treatment for interstitial lung diseases with *SFTPC* mutations depends upon the location of the mutation. Hydroxychloroquine might be a suitable treatment for our cases with $SP\text{-}C^{G100S}$.

Intriguingly, A549 cells transfected with $SP\text{-}C^{G100S}$ contained more proSP-C protein than cells expressing $SP\text{-}C^{WT}$, despite the SP-C mRNA levels being equivalent. This result was inconsistent with the previous report by BRIDGES *et al.* [27], which showed that the mutant protein of $SP\text{-}C^{\Delta exon4}$ was barely detectable in contrast to the wild-type protein in the stably expressing HEK293 cell lines [27]. We also confirmed the minimal accumulation of $proSP\text{-}C^{G100S}$ when HEK293 cells were transfected with $SP\text{-}C^{G100S}$ (data not shown). Therefore, the observed difference is likely to be due to the difference in cell origins, not the features of the mutations. Our experiments also showed that the expression of the 26-kDa isoform of the mutant $SP\text{-}C^{G100S}$ was weaker than that of wild type in A549 cells. Formation of the 26-kDa isoform requires palmitoylation of proSP-C [30] and a 21-kDa isoform is considered to be the proprotein of pre-proteolytic processing [5, 25]. Taken together, we speculate that the palmitoylation process in the mutant $proSP\text{-}C^{G100S}$ was impaired and unpalmitoylated proprotein accumulated in human alveolar epithelial cells (A549). We believe that the slow degradation of unpalmitoylated proprotein in A549 cells is a better reflection of the process actually taking place in the patients presented in this report.

To date, more than 20 mutations have been described in *SFTPC*. Although studies of *SFTPC* mutations have focused on cases of children with interstitial lung diseases, there have been a few studies focusing on pedigrees with adult FPF [11, 31]. They found five kindreds with *SFTPC* mutations, including two new mutations, M71V and IVS4+2T>C, in adult FPF patients. They showed histopathological patterns of UIP and non-classifiable HRCT patterns with reticulonodular opacity and multiple lung cysts in combination with ground-glass opacities or diffuse lung involvement on chest HRCT. The present study, similarly focusing on a pedigree with adult FPF, highlighted some outstanding characteristics of this kindred with $SP\text{-}C^{G100S}$. Our patients presented with a histopathological pattern of UIP and the HRCT findings had features of reticulonodular opacity and multiple lung cysts. These findings, however, were seen predominantly in the upper lobes. In particular a small number of lung cysts were present only at the apex, a feature that was inconsistent with the above report.

Interestingly, the age of phenotypic appearance (*i.e.* the appearance of positive radiological and histopathological findings, even

a)



b)

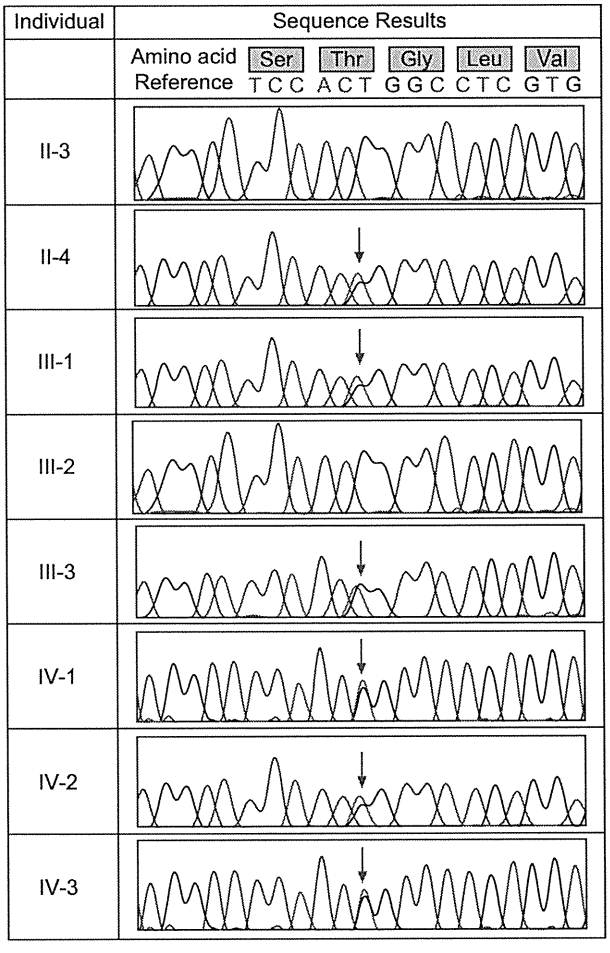


FIGURE 2. Amino acid substitution identified in surfactant protein (SP)-C in individuals with familial pulmonary fibrosis. a) Location of *SFTPC*, the gene encoding SP-C. Red triangle indicates the location of the c.298G>A (G100S) mutation of *SFTPC*, which is in the BRICHOS domain. b) Results of direct DNA sequencing in eight individuals. Red arrows indicate the location of the nonsynonymous substitution (c.298 G>A). c) The highly conserved orthologous protein sequences of SP-C across eight species of mammal. The area surrounded by the red line indicates the location of codon 100 of *SFTPC*.

c)

DNA sequence (human)	A T C	G G C	T C C	A C T	G G C	C T C	G T G	G T G	T A T
Human	I	G	S	T	G	L	V	V	Y
Chimpanzee	I	G	S	T	G	L	V	V	Y
Gorilla	I	G	S	T	G	L	V	V	Y
Mouse	I	G	S	T	G	I	V	V	Y
Rat	I	G	S	T	G	I	V	L	Y
Cow	I	G	S	T	G	T	V	V	Y
Dog	I	G	S	T	G	I	V	V	Y
Opossum	I	G	S	S	G	T	V	V	Y

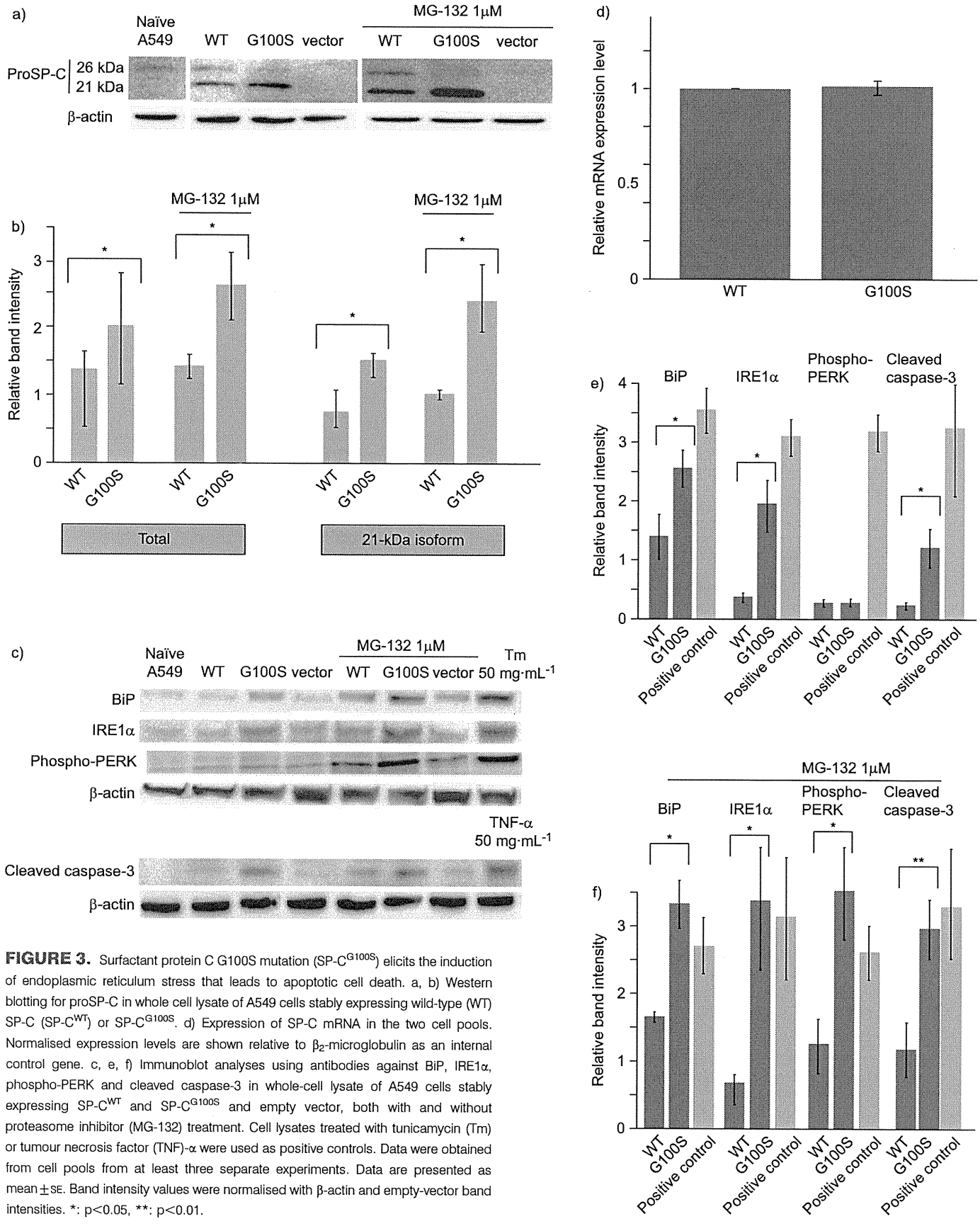


FIGURE 3. Surfactant protein C G100S mutation (SP-C^{G100S}) elicits the induction of endoplasmic reticulum stress that leads to apoptotic cell death. a, b) Western blotting for proSP-C in whole cell lysate of A549 cells stably expressing wild-type (WT) SP-C (SP-C^{WT}) or SP-C^{G100S}. d) Expression of SP-C mRNA in the two cell pools. Normalised expression levels are shown relative to β_2-microglobulin as an internal control gene. c, e, f) Immunoblot analyses using antibodies against BiP, IRE1α, phospho-PERK and cleaved caspase-3 in whole-cell lysate of A549 cells stably expressing SP-C^{WT} and SP-C^{G100S} and empty vector, both with and without proteasome inhibitor (MG-132) treatment. Cell lysates treated with tunicamycin (Tm) or tumour necrosis factor (TNF)-α were used as positive controls. Data were obtained from cell pools from at least three separate experiments. Data are presented as mean±SE. Band intensity values were normalised with β-actin and empty-vector band intensities. *: $p<0.05$, **: $p<0.01$.

if presymptomatic) of all six patients was school-age or older, not at the neonatal or infancy stage, as is commonly reported for other *SFTPC* mutations. These observations caused us to make two speculations. Firstly, SP-C^{G100S} is directly involved in the severity of the disease; the late onset and slow progress of respiratory symptoms might be unique to this mutation. However, the *SFTPC* mutation may have pleiotropic effects across different families, so other families with SP-C^{G100S} mutation need to be investigated carefully to confirm more characteristics of this mutation. IL-8 production in IPF patients is increased [32], but BALF findings of the proband and her siblings showed no inflammatory cell response (table 2) and no IL-8 response (data not shown). This could be related to the relatively modest radiological change and late onset. SP-C^{G100S} might lead to chronic cell death, but it does not induce acute inflammation, eventually resulting in respiratory symptoms and progression to lung fibrosis. Second, any genetic modifier shared with their patients might suppress the progression of the disease caused by SP-C^{G100S}, indicating an indirect involvement of SP-C^{G100S} in the severity of the disease. BULLARD *et al.* [23] implied that *ABCA3* mutations modified the severity of lung disease associated with *SFTPC* mutations. In the present study, we detected no mutations in *ABCA3*. However, considering the late onset of our patients through three generations, it is likely that inherited genetic and epigenetic factors might have homogeneously and moderately suppressed the cytotoxicity induced by SP-C^{G100S}.

Despite the fact that intralobular reticular opacities were barely observed in the lower lobe on chest CT, histopathological findings of fibrotic changes were found, similar to the findings in the upper lobe, where both radiological and histopathological abnormalities were seen. Recent studies have suggested that fibrotic changes might be present in family members with *SFTPC* mutations who have little evidence of disease [31, 33]. In the present study, histopathological examination revealed a UIP pattern in the lower lobe in which no radiological finding was observed. Several reports have shown that individuals carrying other *SFTPC* mutations, including I73T, have not developed symptoms even in adulthood [8, 34]. These observations suggest that individuals with no clinical symptoms, no radiological findings and no phenotypic appearance, but who carry *SFTPC* mutations, might have pathologically recognisable fibrosis, and their lesions might be progressing slowly.

In conclusion, we have detected a new pathogenic mutation in *SFTPC*. The functional analyses in this study suggest that this mutant protein, SP-C^{G100S}, elicits ER stress leading to apoptotic cell death. Our study indicates that this mutation is pathogenic and caused slow progression of pulmonary fibrosis in this kindred. We could not confirm the reason for this slow progression; it might be a characteristic of SP-C^{G100S} or it might be due to the influence of other genes or epigenetic modifications. Functional understanding of the misfolded SP-C protein is important to determine treatment approaches for FPF, which might help in tailor-made treatment based on genotype.

STATEMENT OF INTEREST

None declared.

ACKNOWLEDGEMENTS

We are grateful to the family members for their participation in this study. We especially thank M. Kitaichi (Dept of Pathology, National Hospital Organisation Kinki-Chuo Chest Medical Centre, Osaka, Japan) for reviewing the pathological diagnosis.

REFERENCES

1 Marshall RP, Puddincombe A, Cookson WO, *et al.* Adult familial cryptogenic fibrosing alveolitis in the UK. *Thorax* 2000; 55: 143–146.
2 Bitterman PB, Rennard SI, Keogh BA, *et al.* Familial idiopathic pulmonary fibrosis. Evidence of lung inflammation in unaffected family members. *N Engl J Med* 1986; 314: 1343–1347.
3 Marney A, Lane KB, Phillips JA, *et al.* Idiopathic pulmonary fibrosis can be an autosomal dominant trait in some families. *Chest* 2001; 120: 1 Suppl., 56S.
4 Nogee LM. Alterations in SP-B and SP-C expression in neonatal lung disease. *Annu Rev Physiol* 2004; 66: 601–623.
5 Beers MF, Mulgeta S. Surfactant protein C biosynthesis and its emerging role in conformational lung disease. *Annu Rev Physiol* 2005; 67: 663–696.
6 Nogee LM, Dunbar AE, Wert SE, *et al.* A mutation in the surfactant protein C gene associated with familial interstitial lung disease. *N Engl J Med* 2001; 344: 573–579.
7 Thomas AQ, Lane K, Phillips J, *et al.* Heterozygosity for a surfactant protein C gene mutation associated with usual interstitial pneumonitis and cellular nonspecific interstitial pneumonitis in one kindred. *Am J Respir Crit Care Med* 2002; 165: 1322–1328.
8 Cameron HS, Somaschini M, Carrera P, *et al.* A common mutation in the surfactant protein C gene associated with lung disease. *J Pediatr* 2005; 146: 370–375.
9 Guillot L, Epaud R, Thouvenin G, *et al.* New surfactant protein C gene mutations associated with diffuse lung disease. *J Med Genet* 2009; 46: 490–494.
10 Abou Taam R, Jaubert F, Emond S, *et al.* Familial interstitial disease with I73T mutation: A mid- and long-term study. *Pediatr Pulmonol* 2009; 44: 167–175.
11 Van Moorsel CH, van Oosterhout MF, Barlo NP, *et al.* SFTPC mutations are the basis of a significant portion of adult familial pulmonary fibrosis in a Dutch cohort. *Am J Respir Crit Care Med* 2010; 182: 1419–1425.
12 Nogee LM, Dunbar AE, Wert S, *et al.* Mutations in the surfactant protein C gene associated with interstitial lung disease. *Chest* 2002; 121: Suppl. 3, 20S–21S.
13 Brasch F, Griese M, Tredano M, *et al.* Interstitial lung disease in a baby with a *de novo* mutation in the SFTPC gene. *Eur Respir J* 2004; 24: 30–39.
14 Hamvas A, Nogee LM, White FV, *et al.* Progressive lung disease and surfactant dysfunction with a deletion in surfactant protein C gene. *Am J Respir Cell Mol Biol* 2004; 30: 771–776.
15 Tredano M, Griese M, Brasch F, *et al.* Mutation of SFTPC in infantile pulmonary alveolar proteinosis with or without fibrosing lung disease. *Am J Med Genet A* 2004; 126A: 18–26.
16 Rosen DM, Waltz DA Hydroxychloroquine, surfactant protein C deficiency. *N Engl J Med* 2005; 352: 207–208.
17 Stevens PA, Pettenazzo A, Brasch F, *et al.* Nonspecific interstitial pneumonia, alveolar proteinosis, and abnormal proprotein trafficking resulting from a spontaneous mutation in the surfactant protein C gene. *Pediatr Res* 2005; 57: 89–98.
18 Poterjoy BS, Vibert Y, Sola-Visner M, *et al.* Neonatal respiratory failure due to a novel mutation in the surfactant protein C gene. *Perinatol* 2010; 30: 151–153.
19 Soraisham AS, Tierney AJ, Amin HJ. Neonatal respiratory failure associated with mutation in the surfactant protein C gene. *Perinatol* 2006; 26: 67–70.
20 Keller A, Eistetter HR, Voss T, *et al.* The pulmonary surfactant protein C (SP-C) precursor is a type II transmembrane protein. *Biochem J* 1991; 277: 493–499.

21 Sanchez-Pulido L, Devos D, Valencia A. BRICHOS: a conserved domain in proteins associated with dementia, respiratory distress and cancer. *Trends Biochem Sci* 2002; 27: 329–332.

22 Mulugeta S, Nguyen V, Russo SJ, *et al.* A surfactant protein C precursor protein BRICHOS domain mutation causes endoplasmic reticulum stress, proteasome dysfunction, and caspase 3 activation. *Am J Respir Cell Mol Biol* 2005; 32: 521–530.

23 Bullard JE, Nogee LM. Heterozygosity for ABCA3 mutations modifies the severity of lung disease associated with a surfactant protein C gene (SFTPC) mutation. *Pediatr Res* 2007; 62: 176–179.

24 Chang LJ, Urlacher V, Iwakuma T, *et al.* Efficacy and safety analyses of a recombinant human immunodeficiency virus type 1 derived vector system. *Gene Ther* 1999; 6: 715–728.

25 Ten Brinke A, van Golde LM, Batenburg JJ. Palmitoylation and processing of the lipopeptide surfactant protein C. *Biochim Biophys Acta* 2002; 1583: 253–265.

26 Nerelius C, Martin E, Peng S, *et al.* Mutations linked to interstitial lung disease can abrogate anti-amyloid function of prosurfactant protein C. *Biochem J* 2008; 416: 201–209.

27 Bridges JP, Xu Y, Na CL, *et al.* Adaptation and increased susceptibility to infection associated with constitutive expression of misfolded SP-C. *J Cell Biol* 2006; 172: 395–407.

28 Mulugeta S, Maguire JA, Newitt JL, *et al.* Misfolded BRICHOS SP-C mutant proteins induce apoptosis *via* caspase-4- and cytochrome c-related mechanisms. *Am J Physiol Lung Cell Mol Physiol* 2007; 293: L720–L729.

29 Sisson TH, Mendez M, Choi K, *et al.* Targeted injury of type II alveolar epithelial cells induces pulmonary fibrosis. *Am J Respir Crit Care Med* 2010; 181: 254–263.

30 Vorbroker DK, Dey C, Weaver TE, *et al.* Surfactant protein C precursor is palmitoylated and associates with subcellular membranes. *Biochim Biophys Acta* 1992; 1105: 161–169.

31 Crossno PF, Polosukhin VV, Blackwell TS, *et al.* Identification of early interstitial lung disease in an individual with genetic variations in ABCA3 and SFTPC. *Chest* 2010; 137: 969–973.

32 Keane MP, Arenberg DA, Lynch JP, *et al.* The CXC chemokines, IL-8 and IP-10, regulate angiogenic activity in idiopathic pulmonary fibrosis. *J Immunol* 1997; 159: 1437–1443.

33 Rosas IO, Ren P, Avila NA, *et al.* Early interstitial lung disease in familial pulmonary fibrosis. *Am J Respir Crit Care Med* 2007; 176: 698–705.

34 Percopo S, Cameron HS, Nogee LM, *et al.* Variable phenotype associated with SP-C gene mutations: fatal case with the I73T mutation. *Eur Respir J* 2004; 24: 1072–1073.

CLINICAL GENETICS An International Journal of Genetics, Molecular and Personalized Medicine

Clin Genet 2011: 80: 478–483
Printed in Singapore. All rights reserved

CLINICAL GENETICS
doi: 10.1111/j.1399-0004.2010.01599.x

Short Report

Maternal uniparental isodisomy and heterodisomy on chromosome 6 encompassing a *CUL7* gene mutation causing 3M syndrome

Sasaki K, Okamoto N, Kosaki K, Yorifuji T, Shimokawa O, Mishima H, Yoshiura K-i, Harada N. Maternal uniparental isodisomy and heterodisomy on chromosome 6 encompassing a *CUL7* gene mutation causing 3M syndrome.
Clin Genet 2011: 80: 478–483. © John Wiley & Sons A/S, 2010

We report a case of segmental uniparental maternal hetero- and isodisomy involving the whole of chromosome 6 (mat-hUPD6 and mat-iUPD6) and a cullin 7 (*CUL7*) gene mutation in a Japanese patient with 3M syndrome. 3M syndrome is a rare autosomal recessive disorder characterized by severe pre- and postnatal growth retardation that was recently reported to involve mutations in the *CUL7* or obscurin-like 1 (*OBSL1*) genes. We encountered a patient with severe growth retardation, an inverted triangular gloomy face, an inverted triangle-shaped head, slender long bones, inguinal hernia, hydrocele testis, mild ventricular enlargement, and mild mental retardation. Sequence analysis of the *CUL7* gene of the patient revealed a homozygous missense mutation, c.2975G>C. Genotype analysis using a single nucleotide polymorphism array revealed two mat-hUPD and two mat-iUPD regions involving the whole of chromosome 6 and encompassing *CUL7*. 3M syndrome caused by complete paternal iUPD of chromosome 6 involving a *CUL7* mutation has been reported, but there have been no reports describing 3M syndrome with maternal UPD of chromosome 6. Our results represent a combination of iUPDs and hUPDs from maternal chromosome 6 involving a *CUL7* mutation causing 3M syndrome.

Conflict of interest

None of the authors of this paper declares a conflict of interest.

K Sasaki[a], N Okamoto[b], K Kosaki[c], T Yorifuji[d], O Shimokawa[e], H Mishima[a], K-i Yoshiura[a] and N Harada[e]

[a]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan, [b]Osaka Medical Center and Research Institute for Maternal and Child Health, Osaka, Japan, [c]Department of Pediatrics, Keio University School of Medicine, Tokyo, Japan, [d]Department of Pediatric Endocrinology and Metabolism, Children's Medical Center, Osaka City General Hospital, Osaka, Japan, and [e]Cytogenetic Testing Group B, Advanced Medical Science Research Center, Mitsubishi Chemical Medience Corporation, Nagasaki, Japan

Key words: 3M syndrome – cullin 7 (*CUL7*) – Genome-Wide Human SNP Array 6.0 (SNP6.0) – maternal uniparental disomy of chromosome 6 (matUPD6)

Corresponding author: Dr Koh-ichio Yoshiura, Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Sakamoto 1-12-4, Nagasaki 852-8523, Japan.
Tel.: +81 95 819 7118;
fax: +81 95 849 7121;
e-mail: kyoshi@nagasaki-u.ac.jp

Received 10 September 2010, revised and accepted for publication 12 November 2010

3M syndrome is a rare inherited autosomal recessive disorder characterized by pre- and postnatal growth retardation, characteristic facial features, and skeletal anomalies. Clinical features of 3M syndrome include large head circumference, broad forehead, a triangular facial outline, dolichocephaly, long philtrum, short stature, short thorax and neck, tall vertebral bodies, and slender

long bones and ribs. Males with 3M syndrome occasionally have hypogonadism and hypospadias (1–9). However, intelligence is unaffected and karyotype is normal on conventional chromosome analysis.

In patients with 3M syndrome, disease-causing mutations have been identified in the cullin 7 (*CUL7*, MIM *609577) and obscurin-like 1 (*OBSL1*, MIM *610991) genes (7–9). Mutations of *CUL7* are the major cause of 3M syndrome, accounting for 80% of previously reported cases, whereas *OBSL1* accounts for 20% of cases (8, 10).

Uniparental disomy (UPD) is the transmission pattern of either two copies of the identical chromosome (uniparental isodisomy; iUPD) or of both homologous chromosomes (uniparental heterodisomy; hUPD) from one parent with no contribution from the other parent (11). Phenotypes that are clinically associated with paternal UPD of chromosome 6 (patUPD6) and genomic imprinting have been established, but because of the rarity of maternal UPD of chromosome 6 (matUPD6), clinical features have not yet been established. Here, we report a patient with a homozygous mutation in *CUL7* due to a maternal iUPD of chromosome 6 (mat-iUPD6).

Materials and methods

Clinical report

A Japanese male patient with 3M syndrome was examined in this study. The patient was delivered by caesarean section at 36 weeks of gestation without a family history of 3M syndrome (Fig. 1a). His birth weight was 1000 g (−4.8 SD), length 33.0 cm (−6.8 SD), head circumference 30.2 cm (−1.5 SD), and Apgar score 7/9. Feeding difficulty was noted during the neonatal period. He remained in a neonatal intensive care unit for 2 months and was referred to our group because of developmental delay and muscle hypotonia at 4 months. The patient displayed anomalies including hypospadias, inguinal hernia, hydrocele testis, inverted triangular gloomy face, malar hypoplasia, long eyelashes, epicanthal folds, short nose, anteverted nares, full lips, long philtrum, pointed chin, short chest, grooved lower anterior thorax, hypermobility of joints, and slender long bones (Fig. 1a,b). Mild ventricular enlargement was observed by neuroradiological studies. His growth was severely retarded.

At 2 years 9 months, his height, weight, and head circumference were 69.3 cm (−4.6 SD), 6.8 kg (−6.7 SD), and 48 cm (−1.2 SD), respectively. His head size was disproportionately large compared to his height. Thus the patient was diagnosed as suffering from 3M syndrome. He could understand simple sentences, but could not speak nor sit alone. Partial growth hormone (GH) deficiency was noted. GH replacement therapy was started from 2 years. GH was effective without side effects. At 5 years, his height and weight were 84.8 cm (−5.9 SD) and 10 kg (−3 SD), respectively. He was moderately mentally retarded.

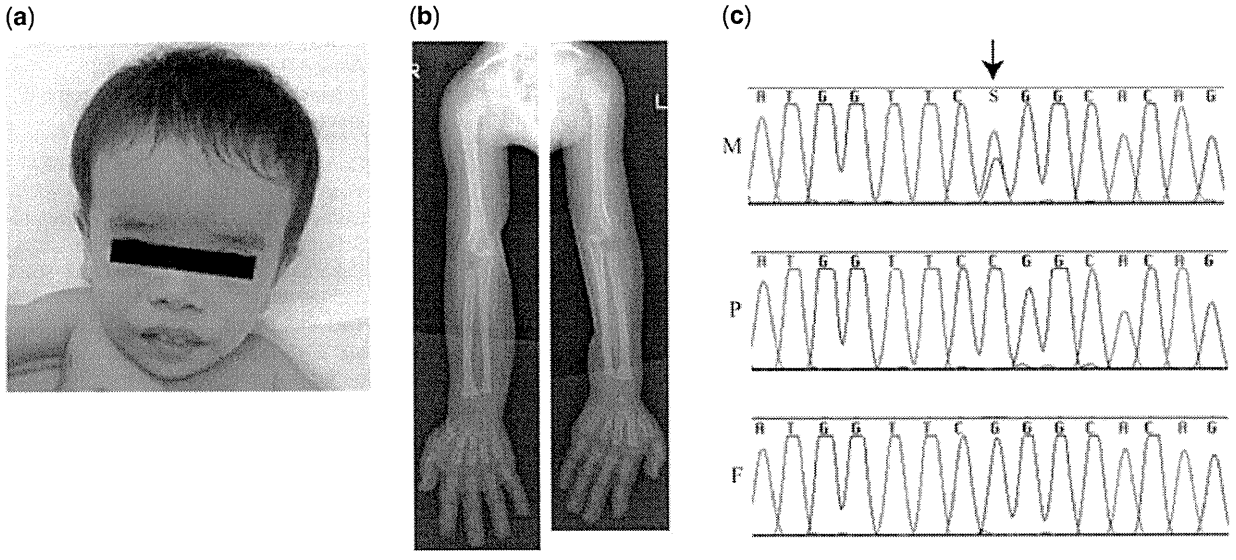


Fig. 1. Facial and skeletal features of the patient at 2 years 7 months of age. (**a**) Note the inverted triangular gloomy face, short nose, full lips, and long philtrum. (**b**) Note the slender long bones. (**c**) Electropherograms of the patient and parents. DNA sequence showing a single base change substituting cytosine for guanine, which results in p.R992P, in the patient. M, mother; P, patient; and F, father.

Conventional cytogenetic studies and FISH analysis

We obtained blood samples under written informed consent for participation in this study. Conventional cytogenetic examination of G-banded chromosomes from peripheral blood lymphocytes was performed. We also performed fluorescence *in situ* hybridization (FISH) analysis on lymphocyte metaphase spreads from the patient using two Bacterial Artificial Chromosome (BAC) clones containing *CUL7*, RP11-628J2 and RP11-653G5, as probes.

Genomic sequencing

Genomic DNA was extracted from peripheral blood following standard protocols. For mutation analysis, we designed primers to amplify all the coding exons of *CUL7* and the flanking intron sequences. Direct sequencing was carried out using a BigDye Terminator v3.1 Cycle sequencing Kit™ and separated on a Genetic Analyzer 3130*xl* (Applied Biosystems Inc., Foster City, CA). Sequence electropherograms were aligned with SEQUENCHER™ software (Gendcodes, Ann Arbor, MI) to visually inspect base alterations.

Microarray analysis

We performed genome-wide single nucleotide polymorphism (SNP) genotyping using Genome-Wide Human SNP Array 6.0 (SNP6.0) following the manufacturer's instructions (Affymetrix, Santa Clara, CA, http://www.affymetrix.com/index.affx). The data generated from Genotyping Console (GTC) 4.0 were loaded into CHROMOSOME ANALYSIS SUITE (CHAS) 1.0.1 software to display the results. We carried out UPD analyses of the patient using genotype data in trio. Genomic positions of SNPs corresponded to the March 2006 human genome (hg18).

Results

Genomic sequencing

We sequenced all 26 coding exons and flanking intronic regions of the *CUL7* gene, which spans a genomic region of approximately 16.3 kb, in the family. In the patient, we detected a homozygous missense mutation (c.2975G>C) in exon 15, which resulted in the substitution of proline for arginine at amino acid residue 992 (p.R992P) (Fig. 1c). The mother was a heterozygous carrier of the mutation, whereas the father was homozygous for the wild-type allele (Fig. 1c). The p.R992P mutation was not detected in 100 unrelated control individuals.

Conventional and molecular cytogenetic analyses

G-banding and FISH analysis at the *CUL7* locus showed a normal karyotype in the patient and the parents with no microdeletion at *CUL7* locus in the patient (data not shown).

Microarray analysis

To confirm paternity, and to find a small size deletion, we performed SNP6.0 analysis. However, no copy number variations (CNVs) were identified in the region containing both *CUL7* and *OBSL1* genes (Fig. 2a). The other variants overlap with reported regions of CNVs in the Database of Genomic Variants (http://projects.tcag.ca/variation) or were transmitted from the parents (data not shown).

To confirm matUPD6 in the patient, we examined the genotypes of the patient/father/mother trio. The results using informative markers indicated that there were two maternal heterodisomic regions (hUPD6-1 and hUPD6-2) and two maternal isodisomic regions (iUDP6-1 and iUPD6-2) in chromosome 6, respectively (Fig. 2 and Table 1). The results indicated that the patient had inherited two alleles from his mother, but none from his father, in chromosome 6. The final karyotype of this patient was 46,XY,upd(6)mat and arr 6p25.3p22.3(110,391–16,287,166)×2 htz mat,6p 22.3q12(16,290,223–65,796,893)×2 hmz mat,6q 12q25.1(65,799,990–150,517,779)×2 htz mat,6q 25.1q27(150,518,012–170,759,956)×2 hmz mat.

Discussion

We identified a causative homozygous mutation in *CUL7* in a patient with 3M syndrome. The results clearly indicate that mat-iUPD6 involving a mutant allele of the *CUL7* gene caused 3M syndrome in the patient.

matUPD6 is relatively rare and seven cases have been reported. The first case was a renal transplant patient who showed growth retardation at birth and mat-iUPD6 (12). The second case was a patient with congenital adrenal hyperplasia (CAH) resulting from a homozygous mutation in the 21-hydroxylase gene (*CYP21*), and had intrauterine growth retardation (IUGR) and mat-iUPD6 (13). The third case was a macerated male fetus from a pregnancy terminated at 23 weeks of gestation because of intrauterine death. The patient showed a mosaic trisomy 6 (14). The fourth case was a male patient with two clinical phenotypes, Klinefelter's syndrome and CAH. His karyotype was mosaic 48,XXY, +mar[30]/47,XXY[20] and