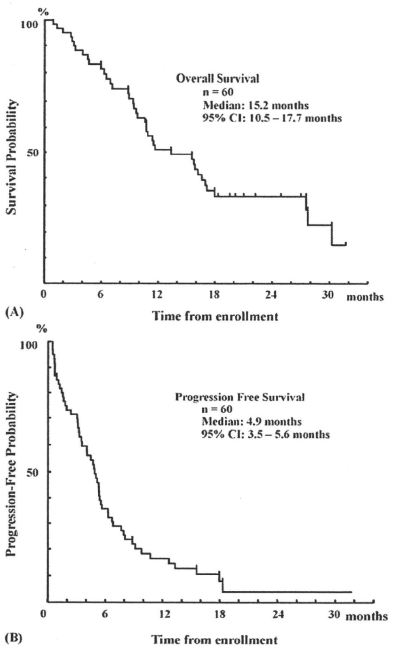

Fig. 1. Cumulative probability of overall (A) and progression-free survivals (PFS) (B), estimated by the Kaplan–Meier method in 60 patients with non-small cell carcinoma who were treated in phase II of the present study. At a median follow-up period of 10.6 months, the mean survival time and median PFS were 15.2 (95% CI: 10.5–17.7) and 4.9 (95% CI: 3.5–5.6) months, respectively. The vertical lines indicate censored cases.

DLT definition similar to ours, proposed the recommended doses as 40 mg/m^2 on day 1 for docetaxel and 40 mg/m^2 for S-1 administered twice daily on days 1–14, and repeated every 4 weeks [15]. Here, DLT of infection or grade 4 neutropenia lasting for 5 days or longer was observed in all the three patients at 50 mg/m^2 docetaxel. The discrepancy of the recommended doses proposed by the present and previous studies may be due to the small patient population in these phase I studies. Most recently, an additional phase I study for patients with previously treated NSCLC thereafter was published, and this study once again determined the recommended doses of docetaxel as 40 mg/m^2 on day 1 and S-1 as 40 mg/m^2 administered twice daily on days 1–14, and repeated every 3 weeks [23]. Here, DLT included grade 4 neutropenia of any duration, and two out of the three patients administered 50 mg/m^2 docetaxel experienced this toxicity. Since this study was for the second-line chemotherapy setting, it is quite reasonable that the recommended dose was lower than that for the first-line treatment of patients with untreated NSCLC. The fact that 30% of the patients in the present phase II study eventually experienced DLT might have confirmed that the optimal dose of docetaxel was 60 mg/m^2. Importantly, the subsequent phase II study showed the safety of the recommended dose proposed by the present phase I study. Although eligibility criteria allowed the inclusion of patients with a PS of 2 because of the possible use of non-platinum regimen for patients with marginally poor PS, there was no patient with a PS of 2 enrolled. Therefore, it is important to notice that the present study was for patients with good PS.

On the other hand, the present phase II study involving 60 patients showed promise with an ORR of 30%, a median PFS of 4.9 months and a median OS of 15.2 months, including a well-tolerated toxicity. Since a phase II study [16] of S-1 in patients with chemotherapy-naïve NSCLC yielded an ORR of 22.0%, a median response duration of 3.4 months and a median OS time of 10.2 months, docetaxel addition possibly enhanced the tumor control activity in the present study. Another phase II study of docetaxel and S-1 but at different doses of 40 mg/m^2 for docetaxel and 40 mg/m^2 for S-1 administered twice daily with the same schedule in 29 patients with previously treated NSCLC resulted in an ORR of 24.1% (95% CI: 10.3–44.8), and a median PFS and OS time of 3.9 and 11.8 months, respectively [23]. These supporting data on the second-line therapy for NSCLC also make this regimen promising. Results of a recently conducted clinical study have been published and demonstrated that a newly developed antifolate, pemetrexed, was effective for patients with adenocarcinoma, but not for those with squamous cell carcinoma of the lung. The authors of this recent report proposed one possible explanation related to a different TS activity in the tumor according to histological type. That is, its expression is lower in adenocarcinoma than in squamous cell carcinoma, and pemetrexed is more effective when such expression is lower in tumors [24]. If this were the case, S-1 may have similar characteristics to pemetrexed, because preclinical data also revealed the relationship between higher efficacy of 5-FU and lower expression of TS [20]. In the present study, however, superior ORR for adenocarcinoma over squamous cell carcinoma was not observed, although the small sample size prohibited definitive conclusion.

In addition, ORR, median PFS and OS times of the present study were comparable to those of platinum-based combination chemotherapy evaluated in large-scale phase III studies. Although direct comparison of the present results with others evaluated in phase III studies does not draw any conclusion, such comparison would clarify some characteristics of the present regimen. As summarized in Table 3, the present study yielded a similarly high ORR, and substantially excellent PFS and OS. Regarding toxicity, neutropenia, febrile neutropenia and infection appear to be more frequent in the present study than in other studies using a platinum-based regimen. Contrarily, nausea and appetite loss appear to be less frequent in the former than in the latter. Aside from this, grade 4 interstitial pneumonia was observed in two among the 60 patients. Pulmonary toxicity is relatively frequently encountered in clinical studies of chemotherapy for lung cancer regardless of agents, particularly when conducted in Japan [25–28]. Although its occurrence was more infrequent than that reported in these studies, special attention should be paid to this toxicity even with the docetaxel and S-1 combination. Therefore, the present non-platinum regimen consisting of docetaxel and S-1 has a therapeutic potential with comparable ORR and survival data to well-established platinum-based regimens with a different toxicity profile, and some patients may benefit from this regimen.

5. Conclusion

The present phase I/II study of combined docetaxel and S-1 has established the recommended dose for each drug, and this combination chemotherapy for patients with chemotherapy-naïve NSCLC showed promise in terms of tumor control and safety. These findings warrant further elucidation of the clinical relevance of the

regimen in the treatment of NSCLC, including prospective comparison of the regimen with combination chemotherapy consisting of a platinum and a taxane.

Conflict of interest

G. A. and A. Y. have consultancies with Taiho Pharmaceutical Co., Ltd.

Acknowledgments

The authors are grateful for the external review by Dr. Kiyoshi Mori, Tochigi Cancer Center. Data management and analysis by Niphix Co., Ltd. and Fulcrumpharma, KK are also appreciated. This research was conducted and funded by the NPO Tokyo Cooperative Oncology Group.

References

[1] Scagliotti GV, De Marinis F, Rinaldi M, Crino L, Gridelli C, Ricci S, et al. Phase III randomized trial comparing three platinum-based doublets in advanced non-small-cell lung cancer. J Clin Oncol 2002;20:4285–91.
[2] Schiller JH, Harrington D, Belani CP, Langer C, Sandler A, Krook J, et al. Comparison of four chemotherapy regimens for advanced non-small-cell lung cancer. N Engl J Med 2002;346:92–8.
[3] Kubota K, Watanabe K, Kunitoh H, Noda K, Ichinose Y, Katakami N, et al. Phase III randomized trial of docetaxel plus cisplatin versus vindesine plus cisplatin in patients with stage IV non-small-cell lung cancer: The Japanese Taxotere Lung Cancer Study Group. J Clin Oncol 2004;22:254–61.
[4] Ohe Y, Ohashi Y, Kubota K, Tamura T, Nakagawa K, Negoro S, et al. Randomized phase III study of cisplatin plus irinotecan versus carboplatin plus paclitaxel, cisplatin plus gemcitabine, and cisplatin plus vinorelbine for advanced non-small-cell lung cancer: Four-Arm Cooperative Study in Japan. Ann Oncol 2007;18:317–23.
[5] Sandler A, Gray R, Perry MC, Brahmer J, Schiller JH, Dowlati A, et al. Paclitaxel-carboplatin alone or with bevacizumab for non-small-cell lung cancer. N Engl J Med 2006;355:2542–50.
[6] Pirker R, Szczesna A, von Pawel J, Krzakowski M, Ramlau R, Park K, et al. FLEX: A randomized, multicenter, phase III study of cetuximab in combination with cisplatin/vinorelbine (CV) versus CV alone in the first-line treatment of patients with advanced non-small cell lung cancer (NSCLC). J Clin Oncol 2008;26(Suppl.) [abstract 3].
[7] D'Addario G, Pintilie M, Leighl NB, Feld R, Cerny T, Shepherd FA. Platinum-based versus non-platinum-based chemotherapy in advanced non-small-cell lung cancer: A meta-analysis of the published literature. J Clin Oncol 2005;23:2926–36.
[8] Shirasaka T, Shimamato Y, Ohshimo H, Yamaguchi M, Kato T, Yonekura K, et al. Development of a novel form of an oral 5-fluorouracil derivative (S-1) directed to the potentiation of the tumor selective cytotoxicity of 5-fluorouracil by two biochemical modulators. Anti-Cancer Drugs 1996;7:548–57.
[9] Giller SA, Zhuk RA, Lidak M. [Analogs of pyrimidine nucleosides. I. N1-(alpha-furanidyl) derivatives of natural pyrimidine bases and their antimetabolities]. Doklady Akademii Nauk SSSR 1967;176:332–5.
[10] Shirasaka T, Shimamoto Y, Fukushima M. Inhibition by oxonic acid of gastrointestinal toxicity of 5-fluorouracil without loss of its antitumor activity in rats. Cancer Res 1993;53:4004–9.
[11] Tatsumi K, Fukushima M, Shirasaka T, Fujii S. Inhibitory effects of pyrimidine, barbituric acid and pyridine derivatives on 5-fluorouracil degradation in rat liver extracts. Jpn J Cancer Res 1987;78:748–55.
[12] Koizumi W, Tanabe S, Saigenji K, Ohtsu A, Boku N, Nagashima F, et al. Phase I/II study of S-1 combined with cisplatin in patients with advanced gastric cancer. Br J Cancer 2003;89:2207–12.
[13] Ajani JA, Lee FC, Singh DA, Haller DG, Lenz HJ, Benson 3rd AB, et al. Multicenter phase II trial of S-1 plus cisplatin in patients with untreated advanced gastric or gastroesophageal junction adenocarcinoma. J Clin Oncol 2006;24:663–7.
[14] Inokuchi M, Yamashita T, Yamada H, Kojima K, Ichikawa W, Nihei Z, et al. Phase I/II study of S-1 combined with irinotecan for metastatic advanced gastric cancer. Br J Cancer 2006;94:1130–5.
[15] Yamaguchi K, Shimamura T, Hyodo I, Koizumi W, Doi T, Narahara H, et al. Phase I/II study of docetaxel and S-1 in patients with advanced gastric cancer. Br J Cancer 2006;94:1803–8.
[16] Kawahara M, Furuse K, Segawa Y, Yoshimori K, Matsui K, Kudoh S, et al. Phase II study of S-1, a novel oral fluorouracil, in advanced non-small-cell lung cancer. Br J Cancer 2001;85:939–43.
[17] Takahashi I, Emi Y, Kakeji Y, Uchida J, Fukushima M, Maehara Y. Increased antitumor activity in combined treatment TS-1 and docetaxel. A preclinical study using gastric cancer xenografts. Oncology 2005;68:130–7.
[18] Kodera Y, Fujiwara M, Yokoyama H, Ohashi N, Miura S, Ito Y, et al. Combination of oral fluoropyrimidine and docetaxel: reappraisal of synergistic effect against gastric carcinoma xenografts. In Vivo 2005;19:861–6.
[19] Wada Y, Yoshida K, Suzuki T, Mizuiri H, Konishi K, Ukon K, et al. Synergistic effects of docetaxel and S-1 by modulating the expression of metabolic enzymes of 5-fluorouracil in human gastric cancer cell lines. Int J Cancer 2006;119:783–91.
[20] Ichikawa W, Uetake H, Shirota Y, Yamada H, Nishi N, Nihei Z, et al. Combination of dihydropyrimidine dehydrogenase and thymidylate synthase gene expressions in primary tumors as predictive parameters for the efficacy of fluoropyrimidine-based chemotherapy for metastatic colorectal cancer. Clin Cancer Res 2003;9:786–91.
[21] Ichikawa W, Uetake H, Shirota Y, Yamada H, Takahashi T, Nihei Z, et al. Both gene expression for orotate phosphoribosyltransferase and its ratio to dihydropyrimidine dehydrogenase influence outcome following fluoropyrimidine-based chemotherapy for metastatic colorectal cancer. Br J Cancer 2003;89:1486–92.
[22] Yoshida K, Hirabayashi N, Takiyama W, Ninomiya M, Takakura N, Sakamoto J, et al. Phase I study of combination therapy with S-1 and docetaxel (TXT) for advanced or recurrent gastric cancer. Anticancer Res 2004;24:1843–51.
[23] Atagi S, Kawahara M, Kusunoki Y, Takada M, Kawaguchi T, Okishio K, et al. Phase I/II study of docetaxel and S-1 in patients with previously treated non-small cell lung cancer. J Thorac Oncol 2008;3:1012–7.
[24] Scagliotti GV, Parikh P, von Pawel J, Biesma B, Vansteenkiste J, Manegold C, et al. Phase III study comparing cisplatin plus gemcitabine with cisplatin plus pemetrexed in chemotherapy-naive patients with advanced-stage non-small-cell lung cancer. J Clin Oncol 2008;26:3543–51.
[25] Koo LC, Clark JA, Quesenberry CP, Higenbottam T, Nyberg F, Wolf MK, et al. National differences in reporting 'pneumonia' and 'pneumonia interstitial': an analysis of the WHO International Drug Monitoring Database on 15 drugs in nine countries for seven pulmonary conditions. Pharmacoepidemiol Drug Saf 2005;14:775–87.
[26] Katakami N, Takiguchi Y, Yoshimori K, Isobe H, Bessho A, Yoshimura A, et al. Docetaxel in combination with either cisplatin or gemcitabine in unresectable non-small cell lung carcinoma: a randomized phase II study by the Japan Lung Cancer Cooperative Clinical Study Group. J Thorac Oncol 2006;1:447–53.
[27] Hotta K, Kiura K, Tabata M, Harita S, Gemba K, Yonei T, et al. Interstitial lung disease in Japanese patients with non-small cell lung cancer receiving gefitinib: an analysis of risk factors and treatment outcomes in Okayama Lung Cancer Study Group. Cancer J 2005;11:417–24.
[28] Ando M, Okamoto I, Yamamoto N, Takeda K, Tamura K, Seto T, et al. Predictive factors for interstitial lung disease, antitumor response, and survival in non-small-cell lung cancer patients treated with gefitinib. J Clin Oncol 2006;24:2549–56.

Vol 465|13 May 2010|doi:10.1038/nature08971

LETTERS

Mutations of optineurin in amyotrophic lateral sclerosis

Hirofumi Maruyama[1], Hiroyuki Morino[1], Hidefumi Ito[2]†, Yuishin Izumi[3], Hidemasa Kato[4], Yasuhito Watanabe[5], Yoshimi Kinoshita[2], Masaki Kamada[1,3], Hiroyuki Nodera[3], Hidenori Suzuki[6], Osamu Komure[7], Shinya Matsuura[8], Keitaro Kobatake[9], Nobutoshi Morimoto[10], Koji Abe[10], Naoki Suzuki[11], Masashi Aoki[11], Akihiro Kawata[12], Takeshi Hirai[12], Takeo Kato[13], Kazumasa Ogasawara[14], Asao Hirano[15], Toru Takumi[5], Hirofumi Kusaka[2], Koichi Hagiwara[16], Ryuji Kaji[3] & Hideshi Kawakami[1]

Amyotrophic lateral sclerosis (ALS) has its onset in middle age and is a progressive disorder characterized by degeneration of motor neurons of the primary motor cortex, brainstem and spinal cord[1]. Most cases of ALS are sporadic, but about 10% are familial. Genes known to cause classic familial ALS (FALS) are superoxide dismutase 1 (*SOD1*)[2], *ANG* encoding angiogenin[3], *TARDP* encoding transactive response (TAR) DNA-binding protein TDP-43 (ref. 4) and fused in sarcoma/translated in liposarcoma (*FUS*, also known as *TLS*)[5,6]. However, these genetic defects occur in only about 20–30% of cases of FALS, and most genes causing FALS are unknown. Here we show that there are mutations in the gene encoding optineurin (*OPTN*), earlier reported to be a causative gene of primary open-angle glaucoma (POAG)[7], in patients with ALS. We found three types of mutation of *OPTN*: a homozygous deletion of exon 5, a homozygous Q398X nonsense mutation and a heterozygous E478G missense mutation within its ubiquitin-binding domain. Analysis of cell transfection showed that the nonsense and missense mutations of *OPTN* abolished the inhibition of activation of nuclear factor kappa B (NF-κB), and the E478G mutation revealed a cytoplasmic distribution different from that of the wild type or a POAG mutation. A case with the E478G mutation showed OPTN-immunoreactive cytoplasmic inclusions. Furthermore, TDP-43- or SOD1-positive inclusions of sporadic and *SOD1* cases of ALS were also noticeably immunolabelled by anti-OPTN antibodies. Our findings strongly suggest that OPTN is involved in the pathogenesis of ALS. They also indicate that NF-κB inhibitors could be used to treat ALS and that transgenic mice bearing various mutations of *OPTN* will be relevant in developing new drugs for this disorder.

We analysed six Japanese individuals from consanguineous marriages who had ALS; two of them were siblings, the others were from independent families. We used homozygosity mapping, which has been shown to identify a locus of a disease-causing gene from as few as three individuals[8]. We performed a genome-wide scan of single nucleotide polymorphisms (SNPs) by using the GeneChip Human Mapping 500K Array Set (Affymetrix), and selected for the run of homozygous SNPs (RHSs) more than 3 centimorgans in length. Under this condition, the RHSs are able to retrieve more than 98% of the entire length of the autozygous segments created as a result of a first-cousin or second-cousin marriage (Supplementary Information)[8]. We extracted RHSs of six individuals (Supplementary Fig. 1a). A region (hg18: 12,644,480–15,110,539) in chromosome 10, which was an overlap among four subjects, was chosen as the primary candidate region (Supplementary Fig. 1b). Assuming that subjects ii, iii, v and vi had the same disease gene, the chance that the overlap had the disease gene was $P_{ii+iii+v+vi} = 0.935$ (Supplementary Information). We listed up to 17 candidate genes in the region and sequenced their exons (Supplementary Fig. 1c). We detected a deletion of exon 5 in the *OPTN* (also known as FIP-2 (ref. 9)) gene in two siblings (Fig. 1a, family 1, subjects 1 and 2). PCR with a forward primer of exon 4 and a reverse primer of intron 5 revealed a 2.5-kilobase (kb) band in the control, V-3 and IV-1, and a 0.7-kb band in IV-1, subject 1 and subject 2 (Fig. 1b). Direct sequence analysis of the short band showed the joining of the 5′ part of AluJb in intron 4 and the 3′ part of AluSx in intron 5 with 12-base-pair (bp) microhomology (Fig. 1c). Thus, the deletion resulted from Alu-mediated recombination. We also found a homozygous nonsense c.1502C>T mutation (Q398X, exon 12) in the gene in one individual with ALS (Fig. 1d, e, family 2, subject 3). For the other three subjects, we found neither mutations nor copy number changes in the *OPTN* gene, although we did not completely exclude the possibility of mutations in introns or intergenic regions in the gene. We extended our analysis of *OPTN* to ten additional individuals from consanguineous marriages who had ALS, 76 individuals with familial ALS and 597 individuals with sporadic ALS (SALS). We found the Q398X mutation in a sporadic individual (subject 4, family 3; Fig. 1d). Subjects 3 and 4, who were not related according to their family history, shared their haplotype for a 0.9-megabase (Mb) region (hg18: chr10: 12,973,261–13,879,735) containing the *OPTN* gene (Supplementary Table 1). We investigated a total of 170 copies of chromosome 10 from 85 Japanese subjects genotyped for the HapMap3 project, and found that the incidental length of haplotype sharing around *OPTN* gene was at most 320 kb. Given that a haplotype sharing of 0.9 Mb rarely occurs by chance, the mutation is likely to have been derived from a single ancestor (Supplementary Fig. 1d). Subjects 1 and 2 shared their haplotype for an 8.3-Mb

[1]Department of Epidemiology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima 734-8553, Japan. [2]Department of Neurology, Kansai Medical University, Moriguchi 570-8506, Japan. [3]Department of Clinical Neuroscience, University of Tokushima Graduate School, Tokushima 770-8503, Japan. [4]Division of Developmental Biology, Research Center for Genomic Medicine, Saitama Medical University, Saitama 350-1241, Japan. [5]Laboratory of Integrative Bioscience, Hiroshima University Graduate School of Biomedical Sciences, Hiroshima 734-8553, Japan. [6]Faculty of Human Science, Hiroshima Bunkyo Women's University, Hiroshima 731-0295, Japan. [7]South Osaka Neurosurgical Hospital, Osakasayama 589-0011, Japan. [8]Department of Genetics and Cell Biology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima 734-8553, Japan. [9]Department of Neurology, Kobatake Hospital, Fukuyama 720-1142, Japan. [10]Department of Neurology, Okayama University, Graduate School of Medicine, Dentistry, and Pharmaceutical Sciences, Okayama 700-8558, Japan. [11]Department of Neurology, Tohoku University School of Medicine, Sendai 980-8574, Japan. [12]Department of Neurology, Tokyo Metropolitan Neurological Hospital, Fuchu, Tokyo 183-0042, Japan. [13]Department of Neurology, Haematology, Metabolism, Endocrinology and Diabetology, Yamagata University Faculty of Medicine, Yamagata 990-9585, Japan. [14]Department of Pathology, School of Medicine, Shiga University of Medical Science, Ohtsu 520-2192, Japan. [15]Division of Neuropathology, Department of Pathology, Montefiore Medical Center, New York, New York 10467-2490, USA. [16]Department of Respiratory Medicine, Saitama Medical University, Saitama 350-0495, Japan. †Present address: Department of Neurology, Kyoto University Graduate School of Medicine, Kyoto 606-8507, Japan.

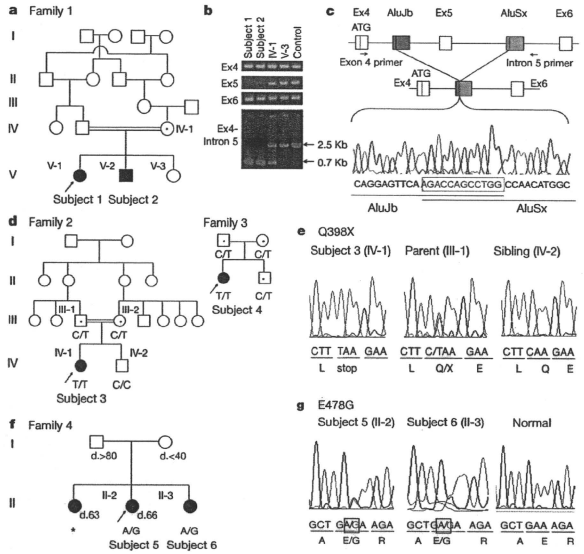

Figure 1 | Exon 5 deletion, nonsense and missense mutations of the *OPTN* gene. **a**, Family 1. The filled circle or square indicate the affected individuals; the arrows indicate the probands. **b**, Agarose gel electrophoretogram. Subject 1 (V-1) and subject 2 (V-2) showed lack of exon 5 PCR product and shortened product of exon 4 to intron 5. **c**, Chromatogram with OPTN deletion of exon 5 and schematic structure of deleted gene. **d**, Families 2 and 3. Dots indicate heterozygous carriers. **e**, Chromatograms from index subjects with OPTN mutation of c.1502 C>T. Homozygous mutation is in red, and the mutation is indicated by using the single-letter amino-acid code. **f**, Family 4. *DNA sample could not be obtained. Numerals show the age at death. **g**, Chromatograms from index subjects with the OPTN mutation of c.1743A>G. The heterozygous mutation is marked by the square.

region (hg18: chr10: 6,815,934–14,842,351), which contained the *OPTN* gene and was different from that in subjects 3 and 4 (Supplementary Table 1).

In the screening of ALS families, we identified a heterozygous missense mutation (c1743A>G, E478G, exon14, Fig. 1g) of *OPTN* in four individuals with ALS in two families with ALS. Subjects 5 and 6 were sisters, and the pedigree suggests that the mutation had an autosomal dominant trait with incomplete penetrance (Fig. 1f, family 4). Subjects 7 and 8 (family 5) were brothers. Although these families are not related according to their family history, subjects 5–8 shared their haplotype for 2.3 Mb (hg18: chr10: 11,460,985–13,703,017, Supplementary Table 3), again suggesting that the mutation was derived from a single ancestor. Indeed, the Q398X nonsense and E478G missense mutations were not observed in 781 healthy Japanese volunteers as well as in over 6,800 (including 1,728 Japanese) individuals in the glaucoma studies, where the entire coding region of the gene was investigated (Supplementary Table 2). Collectively, the mutation was absent over a total of 5,000 Japanese chromosomes. The deletion mutation was also absent in 200 Japanese, and not reported in the over 6,800 glaucoma individuals. The co-segregation of three different mutations of *OPTN* with the ALS phenotype strongly suggests that some mutations of *OPTN* cause ALS.

The eight individuals with mutations of *OPTN* showed onset from 30 to 60 years of age. Most of them showed a relatively slow progression and long duration before respiratory failure, although the clinical phenotypes were not homogeneous (see Supplementary Information).

The Q398X mutation causes a premature stop during translation, truncating the 577 amino-acid OPTN protein to one of 397 amino acids in length. This truncation results in a deletion of the coiled coil 2 domain[10], which is necessary for binding to ubiquitin[11], huntingtin[12] (htt), myosin VI[13] and the ubiquitinated receptor-interacting protein[14]. In the gene with the deletion of exon 5, if there was a transcript, the transcript splicing from exon 4 to exon 6 would cause a frame shift and make a stop codon (TGA in the ninth to eleventh codons in exon 6), which would be expected to translate a peptide 58 amino acids in length. The missense mutation (E478G) was located between coiled coil 2 domain and the leucine zipper domain. This glutamic acid is highly conserved among OPTN proteins of a wide range of species (Supplementary Fig. 2a), and is situated within the DFxxER motif, an ubiquitin-binding domain shared among OPTN, NF-κB essential molecule (NEMO), and A20 binding and inhibitor of NF-κB proteins (ABIN) (Supplementary Fig. 2b). The mutations in the DFxxER motif in ABIN reduce the binding to ubiquitin, which render them unable to inhibit NF-κB activation[11]. We investigated the ability of various mutations of *OPTN* to inhibit NF-κB-mediated transcriptional activation by performing a luciferase assay using NSC-34 cells (a mouse neuroblastoma and spinal-cord hybrid cell line) transfected with wild-type or mutant *OPTN*. E50K *OPTN*, which causes POAG[7], downregulated the NF-κB activity, as did the wild type. On the other hand, both Q398X and E478G had no ability to inhibit NF-κB activity (Tukey–Kramer, $P < 0.05$). These tendencies were retained after stimulation with tumour-necrosis factor (TNF)-α (Fig. 2A). We also examined the subcellular localization of overexpressed Flag-tagged wild-type OPTN (wild type) and its mutants in cells (Fig. 2B). Immunofluorescence staining was performed with their antibodies against Flag and the Golgi matrix marker GM130. Confocal images showed close apposition of granular signals of wild-type OPTN or E50K with GM130 (see g and i in Fig. 2B)[15,16]. E50K often shapes large granular structures near the Golgi apparatus. E478G rarely showed granular signals (see b in Fig. 2B); however, when closely observed, some of the signals were still closely localized to GM130 (see h in Fig. 2B). Western blotting using a lysate of transformed lymphoblasts showed that the 74-kDa band, corresponding to OPTN, was absent in subjects 3 and 4, but was present in the non-diseased mother and brother of subject 3 (Supplementary Fig. 3a). Quantitative PCR with reverse transcription

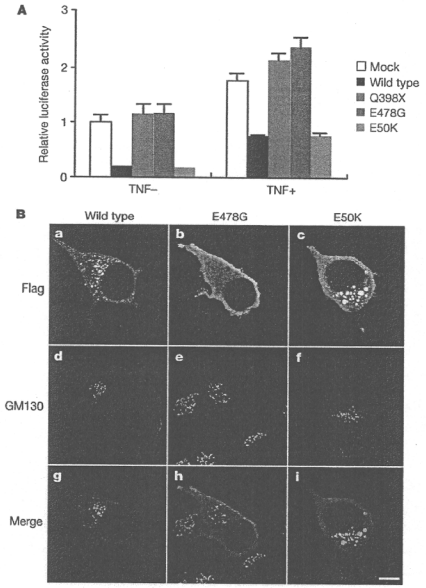

Figure 2 | Influence of OPTN mutations. A, Luciferase assay to assess the ability of various OPTNs to inhibit activation of NF-κB. The wild type and E50K have a similar NF-κB activation-inhibiting effect, whereas mock, Q398X and E478G types lack this effect. Error bars, standard deviations of triplicate assays. **B,** Localization of OPTN. Flag is the white signals in a–c and red signals in g–i. GM130 is the white signals in d–f and green signals in g–i. The wild type shows many fluorescent granules closely localized with the Golgi apparatus. E478G OPTN shows a reduced number of granules, and rarely co-localized with the Golgi apparatus. E50K OPTN granules have become large and closely localized with the Golgi apparatus. Scale bar, 10 μm.

revealed that the products were diminished to 58.0% in the heterozygote (III-2) and to 13.8% in the homozygote (subject 4) compared with the control levels (Supplementary Fig. 3b). In addition, cycloheximide recovered the decrease in the OPTN messenger RNA (mRNA) with the mutation (Supplementary Fig. 3c). Thus mRNA with this mutation, which bears a premature termination, might be degraded through nonsense-mediated mRNA decay in lymphoblasts.

The spinal cord from subject 5 with the E478G mutation revealed loss of myelin from the corticospinal tract and of the anterior horn cells (AHCs, Fig. 3a and Supplementary Fig. 4a, b). OPTN immunohistochemistry demonstrated increased staining intensity of the cytoplasm of the remaining AHCs and the neurites in the anterior horn (Supplementary Fig. 4c). Higher magnification of the motor neurons revealed intracytoplasmic eosinophilic inclusions (Fig. 3b, d). Intriguingly, these inclusions were distinctly immunopositive for OPTN (Fig. 3c, e). On the other hand, the cytoplasm of AHCs from control individuals was faintly labelled with anti-OPTN antibodies (Supplementary Fig. 5a, c), similar to the spinal-cord AHCs of mice (Supplementary Fig. 6b) and in contrast to the highly labelled sensory neurons in the dorsal root ganglia of mice (Supplementary Fig. 6d). In patients with sporadic ALS, the staining intensity for OPTN apparently increased not only in the cytoplasm of the remaining AHCs but also in their neurites (Supplementary Fig. 5b, d). In

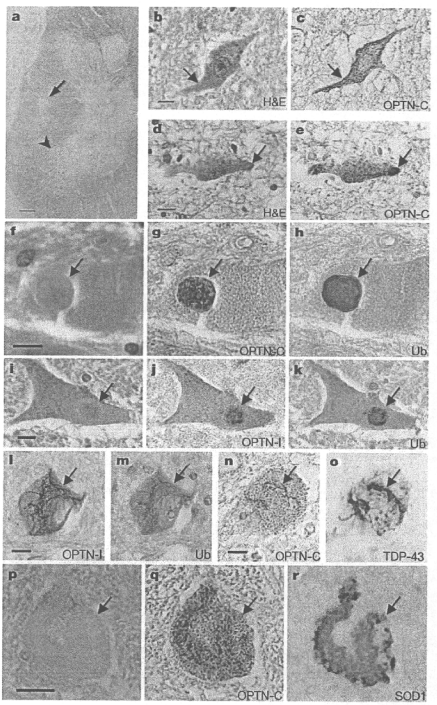

Figure 3 | Identification of OPTN in distinctive intracytoplasmic inclusions of subjects with ALS. a–e, Neuropathology of the lumbar spinal cord from subject 5. Klüver-Barrera (**a**) show loss of myelin from the corticospinal tract (arrow) and loss of motor neurons from the anterior horn (arrowhead). The cytoplasm of the remaining motor neurons contains an amorphous eosinophilic region (**b**, arrow). H&E, haematoxylin and eosin. The same neuron was re-stained with the anti-OPTN antibody (**c**, arrow). The eosinophilic retention occasionally appears to form a hyaline inclusion (**d**, arrow), which is intensely immunolabelled with the anti-OPTN antibody (**e**, arrow). **f–k,** Round hyaline inclusions of subjects with SALS (**f, i**) are immunolabelled with anti-OPTN-C and anti-OPTN-I antibodies (**g** and **j**, respectively). The sections were re-stained with anti-ubiquitin (Ub) antibodies (**h, k**). **l–o,** Skein-like inclusions of patients with SALS are reactive with the anti-OPTN-I and anti-OPTN-C antibodies (**l, n**). Re-staining of **l** with the anti-ubiquitin antibody (**m**) and **n** with anti-TDP-43 antibody (**o**). **p–r,** Lewy-body-like hyaline inclusion of a patient with FALS, stained with haematoxylin and eosin (**p**), anti-OPTN-C antibody (**q**) and SOD1 antibody (**r**). Scale bars, 200 μm (**a**), 20 μm (**b–p**).

addition, distinctive intracytoplasmic inclusions were also noticeably OPTN immunolabelled in cases of sporadic and familial ALS; eosinophilic round hyaline inclusions from patients with SALS were immunopositive for OPTN (Fig. 3f, g, i, j). Re-staining of the same sections for ubiquitin, a known constituent of many neurodegenerative inclusions, revealed that these inclusions were also positive and faithfully matched the distribution of OPTN immunoreactivity (Fig. 3h, k). The anti-OPTN antibodies also stained skein-like inclusions (Fig. 3l, n), which were again mirrored with the anti-ubiquitin antibodies (Fig. 3m) and with the anti-TDP-43

antibodies (Fig. 3o). The distinct OPTN immunoreactivity of ubiquitin- and TDP-43-positive intracytoplasmic inclusions was confirmed on serial sections from patients with SALS (Supplementary Fig. 7). Moreover, SOD1-immunopositive Lewy-body-like hyaline inclusions from cases with *SOD1* FALS were also immunopositive for OPTN (Fig. 3p–r). We found that OPTN antibody labelled both SOD1- and TDP-43-positive inclusions. As the staining of SOD1 and TDP-43 is generally mutually exclusive, OPTN staining appears to be a more general marker for inclusions in various types of ALS; therefore, the OPTN molecule might also be involved in a broader pathogenesis of ALS.

The mutations of the *OPTN* gene cause both recessive and dominant traits, and the mechanism causing the disease may be different between the two traits. The Q398X nonsense mutation and probably the exon 5 deletion mutation cause a decrease in OPTN expression resulting from nonsense-mediated mRNA decay of the transcript carrying the nonsense OPTN mutations. Therefore, the mutated OPTN protein by itself is unlikely to disturb cell function or to be included in the inclusion body in the motor neuron cells. The mechanism of recessive mutations causing ALS is expected to be simply loss of function, and the heterozygote for the Q398X mutation does not develop the ALS phenotype. On the other hand, the E478G missense mutation increased the immunoreactivity for OPTN in the cell body and the neurites. The increased amount and different distribution of the mutated protein would disturb neuronal functions, and may accelerate the inclusion body formation as well as the increase and the different distribution of OPTN immunoreactivity in sporadic ALS. Thus the heterozygote for the E478G mutation will develop the disease.

The different impact on NF-κB signalling and the different intracellular localization of ALS- and POAG-linked mutated protein may explain the phenotypic divergence between the two diseases. Subject 3 with homozygotic Q398X also showed POAG, whereas subject 4 with the same mutation, and subjects 1 and 2 with the exon 5 deletion, did not show it. The prevalence of POAG in the population older than 40 years is 3.9% in Japan[17].Considering this information, the ALS and glaucoma in subject 3 may accidentally coexist.

OPTN competes with NEMO for binding to the ubiquitinated receptor-interacting protein and negatively regulates TNF-α-induced activation of NF-κB[14], which mediates an upregulation of OPTN, creating a negative feedback loop[18]. ALS-related OPTN mutations lacked the inhibitory effect towards NEMO, and thus exaggerated NF-κB activation. In sporadic ALS, a previous report showed that NF-κB, which is classified as a 'cell death inhibitor', is upregulated in motor neurons[19]. The upregulated NF-κB may induce the overexpression of OPTN, and may also cause neuronal cell death[20]. Thus NF-κB is a major candidate target for treating this disease. Additionally OPTN plays an important role in the maintenance of the Golgi complex, in membrane trafficking, in exocytosis, through its interaction with myosin VI and Rab8 (ref. 13), and in post-Golgi trafficking to lysosomes dependent on the Rab8/OPTN/htt complex[21] (Supplementary Fig. 8). Interestingly, FUS/TLS has been reported to interact with myosin VI[22] as well as with myosin V[23]. Impairment of intracellular trafficking of the complex including OPTN and/or FUS/TLS may cause inclusions in this neurodegenerative disorder.

METHODS SUMMARY

Genotyping and extraction of candidate regions. The genotype of the GeneChip Human Mapping 500K Array Set (Affymetrix) was performed by AROS Applied Biotechnology. Computer analyses of the SNPs were performed by a homozygosity mapping algorithm accommodated to the whole-genome SNP scan data (Supplementary Information). To investigate the existence of a large insertion or deletion in this region, we analysed the copy number using Affymetrix Genotyping Console version 4.0 for the Affymetrix Mapping 500K data.

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

Received 17 August 2009; accepted 2 March 2010.
Published online 28 April 2010.

1. Leigh, P. N. in *Handbook of Clinical Neurology: Amyotrophic Lateral Sclerosis* Vol. 82 (eds Eisen, A. A. & Shaw, P. J.) 249–278 (Elsevier, 2007).
2. Rosen, D. R. *et al.* Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* **362,** 59–62 (1993).
3. Greenway, M. J. *et al. ANG* mutations segregate with familial and 'sporadic' amyotrophic lateral sclerosis. *Nature Genet.* **38,** 411–413 (2006).
4. Sreedharan, J. *et al.* TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis. *Science* **319,** 1668–1672 (2008).
5. Kwiatkowski, T. J. *et al.* Mutations in the *FUS/TLS* gene on chromosome 16 cause familial amyotrophic lateral sclerosis. *Science* **323,** 1205–1208 (2009).
6. Vance, C. *et al.* Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6. *Science* **323,** 1208–1211 (2009).
7. Rezaie, T. *et al.* Adult-onset primary open-angle glaucoma caused by mutations in optineurin. *Science* **295,** 1077–1079 (2002).
8. Huqun, et a. l. Mutations in the SLC34A2 gene are associated with pulmonary alveolar microlithiasis. *Am. J. Respir. Crit. Care Med.* **175,** 263–268 (2007).
9. Li, Y., Kang, J. & Horwitz, M. S. Interaction of an adenovirus E3 14.7-kilodalton protein with a novel tumor necrosis factor alpha-inducible cellular protein containing leucine zipper domains. *Mol. Cell. Biol.* **18,** 1601–1610 (1998).
10. Schwamborn, K., Weil, R., Courtois, G., Whiteside, S. T. & Israël, A. Phorbol esters and cytokines regulate the expression of the NEMO-related protein, a molecule involved in a NF-κB-independent pathway. *J. Biol. Chem.* **275,** 22780–22789 (2000).
11. Wagner, S. *et al.* Ubiquitin binding mediates the NF-κB inhibitory potential of ABIN proteins. *Oncogene* **27,** 3739–3745 (2008).
12. Hattula, K. & Peränen, J. FIP-2, a coiled-coil protein, links huntingtin to Rab8 and modulates cellular morphogenesis. *Curr. Biol.* **10,** 1603–1606 (2000).
13. Sahlender, D. A. *et al.* Optineurin links myosin VI to the Golgi complex and is involved in Golgi organization and exocytosis. *J. Cell Biol.* **169,** 285–295 (2005).
14. Zhu, G., Wu, C. J. & Ashwell, J. D. Optineurin negatively regulates TNFα-induced NF-κB activation by competing with NEMO for ubiquitinated RIP. *Curr. Biol.* **17,** 1438–1443 (2007).
15. De Marco, N., Buono, M., Troise, F. & Diez-Roux, G. Optineurin increases cell survival and translocates to the nucleus in a Rab8-dependent manner upon an apoptotic stimulus. *J. Biol. Chem.* **281,** 16147–16156 (2006).
16. Chalasani, M. L., Balasubramanian, D. & Swarup, G. Focus on molecules: optineurin. *Exp. Eye Res.* **87,** 1–2 (2008).
17. Iwase, A. *et al.* The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study. *Ophthalmology* **111,** 1641–1648 (2004).
18. Mrowka, R., Blüthgen, N. & Fähling, M. Seed-based systematic discovery of specific transcription factor target genes. *FEBS J.* **275,** 3178–3192 (2008).
19. Jiang, Y. M. *et al.* Gene expression profile of spinal motor neurons in sporadic amyotrophic lateral sclerosis. *Ann. Neurol.* **57,** 236–251 (2005).
20. Pizzi, M. & Spano, P. Distinct roles of diverse nuclear factor-κB complexes in neuropathological mechanisms. *Eur. J. Pharmacol.* **545,** 22–28 (2006).
21. Toro, D. *et al.* Mutant huntingtin impairs post-Golgi trafficking to lysozomes by delocalizing optineurin/Rab8 complex from the Golgi apparatus. *Mol. Biol. Cell* **20,** 1478–1492 (2009).
22. Takarada, T. *et al.* A protein–protein interaction of stress-responsive myosin VI endowed to inhibit neural progenitor self-replication with RNA binding protein, TLS, in murine hippocampus. *J. Neurochem.* **110,** 1457–1468 (2009).
23. Yoshimura, A. *et al.* Myosin-Va facilitates the accumulation of mRNA/protein complex in dendritic spines. *Curr. Biol.* **16,** 2345–2351 (2006).

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

Acknowledgements This work was supported in part by grants-in-aid from the Ministry of Education, Science, and Culture of Japan, by a grant from the Smoking Research Foundation to H. Kawakami, and by the Japan Science and Technology Agency, Core Research for Evolutional Science & Technology to T.T. We thank E. Nakajima for technical support, K. Nakayama, H. W. Shin, M. Omi and H. Nakamura for conducting some of the experiments, and T. Miki and K. Noda for providing some DNA samples and clinical information. This paper is dedicated to the patients and families who contributed to this project.

Author Contributions H. Kawakami designed and supervised the study. H.Mo. and K.H. extracted candidate genes. H.Ma. and M.K. performed sequencing analysis. H.Ma., H.Mo., Y.W., T.T., S.M., H. Kawakami and H.S. conducted molecular biological analysis. H.I., Y.K., H. Ku., H. Kato, K.O. and A.H. performed pathological analysis and provided pathological samples. Y.I., H.N., R.K., O.K., N.M., K.A., A.K., T.H, T.K., M.A., N.S. and K.K. collected clinical information and samples. H. Kawakami, H.Ma., H.I. and K.H. wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Correspondence and requests for materials should be addressed to H. Kawakami (hkawakam@hiroshima-u.ac.jp).

METHODS

Ethical considerations. The study was approved by the institutional review boards of the participating institutions. All examinations were performed after having obtained informed consent from all subjects or their families.

Subjects. Neurologists performed the clinical diagnosis. The mean age at onset of subjects with ALS was 59.9 years (range 10–85 years, including 14 cases confirmed by autopsy). The possibility of mutation of SOD1 was excluded.

Screening for the mutation of *OPTN*. A list of PCR primer pairs used to amplify individual *OPTN* in the regulatory regions (~1,000 bases upstream from transcription start sites), non-coding exons, coding exons and the surrounding sequences (50–100 bases) of the exons or intron 4 and 5 is provided in Supplementary Table 4. Deletion of Exon 5 was checked by using exon 4 forward and intron 5 reverse primer pairs. Direct sequence of the joining part was performed by using intron 4-2 forward primer or intron 5-6 reverse primer. Screening for the c.1502C>T mutation was performed by analysing restriction-fragment length polymorphism or direct sequencing on 781 healthy control subjects (mean age 62.3 years; range 30–100 years). Exon 12 was amplified and then restricted with MseI, and thereafter the products were electrophoresed in 2% agarose gel. The wild type was digested into 204-, 106-, 14- and 12-bp fragments, and the mutant type (204 bp) into 169 + 35-bp fragments. The c.1743A>G mutation was determined by direct sequencing. In the Affymetrix Mapping 500K, there were 11 SNPs in the *OPTN* gene. However, there are no SNP markers between exon 2 and exon 12 of *OPTN*, and additional quantitative PCR analysis of all exons of the *OPTN* gene was performed.

Luciferase assay. We investigated the activity of NF-κB by using the luciferase assay. Four types of complementary DNA (cDNA) from OPTN were inserted into separate pDNR (Clontech). These were wild (IMAGE clone 3831267), Q398X (recessive), E478G (dominant) and E50K (which causes glaucoma) types. pDNR vector was used as mock. NSC-34 cells were co-transfected with NF-κB reporter ($(Igk)_3$ conaluc plasmid) (a gift from S. Yamaoka) and pDNR-OPTN by using Lipofectamine 2000 (Invitrogen). Luciferase activity was measured 5 h after either PBS or TNF-α (10 ng ml^{-1}, R&D) stimulation by using a Dual-Luciferase Reporter Assay System (Promega). Consistent results were obtained by conducting three independent experiments.

Localization of OPTN. We investigated the localization of OPTN by using a 3×Flag tag. This was inserted into pcDNA3 (Invitrogen), and three types of OPTN cDNA (wild, E478G, E50K) were inserted after the 3×Flag tag. These plasmids were used to transfect NSC-34 cells with the aid of Lipofectamine 2000 (Invitrogen). GM130 (BD Transduction Laboratories) was used as a marker of the Golgi apparatus.

Immunofluorescence microscopy. Cells were grown on glass-bottomed glass dishes (Matsunami) coated with poly-L-lysine and laminin (Sigma Aldrich) and transfected by Lipofectamine 2000 (Invitrogen) according to the manufacturer's protocol; 24–48 h after transfection, the cells were fixed, blocked with normal serum and incubated with primary antibody at 4 °C overnight. Confocal images were acquired with an Olympus FV300 by using a ×100 oil immersion lens with a sequential-acquisition setting at a resolution of 512 pixels × 512 pixels with threefold magnification. Each cellular picture was generated by combining multiple optical images (10–15 slices, *z*-spacing of 0.2 μm) spanning 2–3 μm along the *z*-axis. Subcellular localization of Flag-tagged optineurin was verified by at least three independent experiments. More than 100 cells were photographed for each optineurin construct. The following antibodies were used: mouse monoclonal anti-GM130 (BD Transduction Laboratories, 1:1,000) and affinity-purified rabbit polyclonal anti-Flag (Sigma, 1:1,000).

Western blotting. We investigated the expression of OPTN by western blotting. Cell lysates were prepared from Epstein-Barr-virus immortalized B lymphocytes from subject 3, her brother and mother, and subject 4 by using standard protocols. Polyclonal antibodies recognizing the carboxy (C)-terminal part of OPTN (Cayman Chemical) and anti-rabbit IgG-HRP antibody (R&D Systems) were used. For the internal control, we used glyceraldehyde-3-phosphate dehydrogenase polyclonal antibody (IMGENEX).

Quantitative PCR with reverse transcription. Quantitative PCR with reverse transcription was performed by using THUNDERBIRD SYBR qPCR Mix (TOYOBO) and ABI 7900HT Fast Real Time PCR system (Applied Biosystems). Epstein-Barr-virus immortalized B lymphocytes were treated with cycloheximide (Sigma, 100 μg ml^{-1}) for 2 h before RNA extraction.

Immunohistochemistry of mouse nervous tissue. Several antibodies were tested for their use in detecting mouse OPTN in tissue sections (data not shown). Among them, rabbit polyclonal antibodies raised against various peptides of human/mouse OPTN origin gave consistent and reasonable results. One such antibody was OPTN-C raised against the C-terminal part of OPTN, which is identical between human and mouse (amino acids 575–591; Cayman Chemical). Immunohistochemistry was performed on adult DBA/2 mouse. Mice were transcardially fixed with 4% paraformaldehyde in PBS, post-fixed in the same fixative overnight, and then dehydrated in 30% sucrose in PBS overnight. Frozen sections were obtained by using a cryostat and mounted onto 3-triethoxysilylpropylamine (TESPA)-coated glass slides. After air-drying, the slides were washed in PBS and blocked for 2 h at room temperature in 5% BSA/0.3% Triton X-100 containing PBS. The sections were then incubated overnight at 4 °C with primary antibodies against OPTN diluted in 1% BSA/1% normal goat serum/0.3% Triton X-100/PBS. After several washes in PBS, Alexa-594-conjugated secondary antibody (Invitrogen) in PBS was applied. Pictures were taken with a camera attached to a fluorescence microscope (BIOREVO BZ-9000; Keyence).

Histochemistry. Post-mortem material from one of the OPTN mutant cases (subject 5) was available. Sections (6 μm) of formalin-fixed, paraffin-embedded spinal cord were examined with Klüver–Barrera and haematoxylin and eosin staining. Some sections stained with haematoxylin and eosin were photographed, decolourized and immunostained with OPTN-C (mouse monoclonal, 1: 50,000) or OPTN-1 (rabbit polyclonal, Cayman Chemical, 1:400). In addition, lumbar spinal cord tissue was obtained from clinically and neuropathologically proven cases of SALS (seven cases) and familial ALS with the A4V *SOD1* mutation (FALS, three cases). Six age-matched normal individuals served as controls. After confirmation of complete removal of the OPTN antibody, we immunostained the same sections with the anti-ubiquitin antibodies (mouse monoclonal, Santa Cruz Biotechnology, 1:400; rabbit polyclonal, Sigma, 1:600), anti-TDP-43 antibodies (mouse monoclonal, Abnova, 1:1,000; rabbit polyclonal, Proteintech Group, 1:4,000) or anti-SOD1 antibodies (mouse monoclonal, Lab Vision Corporation, 1:50; rabbit polyclonal, Stressgen Biotechnologies, 1:2,000).

Huqun et al. BMC Bioinformatics 2010, 11(Suppl 7):S5
http://www.biomedcentral.com/1471-2105/11/S7/S5

BMC Bioinformatics

PROCEEDINGS Open Access

A quantitatively-modeled homozygosity mapping algorithm, qHomozygosityMapping, utilizing whole genome single nucleotide polymorphism genotyping data

Huqun[1,2], Shun-ichiro Fukuyama[1†], Hiroyuki Morino[3], Hiroshi Miyazawa[1], Tomoaki Tanaka[1], Tomoko Suzuki[1], Masakazu Kohda[4], Hideshi Kawakami[3], Yasushi Okazaki[4], Kuniaki Seyama[5], Koichi Hagiwara[1*†]

From Asia Pacific Bioinformatics Network (APBioNet) Ninth International Conference on Bioinformatics (InCoB2010)
Tokyo, Japan. 26-28 September 2010

Abstract

Homozygosity mapping is a powerful procedure that is capable of detecting recessive disease-causing genes in a few patients from families with a history of inbreeding. We report here a homozygosity mapping algorithm for high-density single nucleotide polymorphism arrays that is able to (i) correct genotyping errors, (ii) search for autozygous segments genome-wide through regions with runs of homozygous SNPs, (iii) check the validity of the inbreeding history, and (iv) calculate the probability of the disease-causing gene being located in the regions identified. The genotyping error correction restored an average of 94.2% of the total length of all regions with run of homozygous SNPs, and 99.9% of the total length of them that were longer than 2 cM. At the end of the analysis, we would know the probability that regions identified contain a disease-causing gene, and we would be able to determine how much effort should be devoted to scrutinizing the regions. We confirmed the power of this algorithm using 6 patients with Siiyama-type α1-antitrypsin deficiency, a rare autosomal recessive disease in Japan. Our procedure will accelerate the identification of disease-causing genes using high-density SNP array data.

Background

Identification of the genetic factors underlying disease causation provides crucial information for disease prevention and treatment. Nevertheless, genetic factors have not yet been elucidated for many diseases [1,2].

Homozygosity mapping [3] enables the detection of recessive disease-causing genes in a few patients from families with a history of inbreeding; this mapping technique is especially useful for the detection of rare genes. With this technique, chromosomal segments in which all polymorphic markers are homozyogous are considered autozygous segment (AS) [4]. If a patient's coefficient of consanguinity is F, and the frequency of the disease-causing gene in the population is p, then the chance that the recessive disease-causing gene is located in an AS (P_{AS}) is.

$$P_{AS} = \frac{F}{(1-F)p+F} \quad (1)$$

[3]

If a patient is from an inbred family (i.e., F is large) and the disease is rare (i.e., p is small), then $P_{AS} \approx 1$, indicating that the gene is located in an AS. There are implementations that utilize single-nucleotide polymorphism (SNP) genotyping data obtained by high-density arrays [5,6]. The usable implementation should (i) correct genotyping errors because thousands of SNPs are mistyped per high-density SNP array, adversely

* Correspondence: hagiwark@saitama-med.ac.jp
† Contributed equally
[1]Department of Respiratory Medicine, Saitama Medical University, 38 Morohongo, Moroyama, Saitama 350-0495, Japan
Full list of author information is available at the end of the article

affecting the homozygosity mapping analysis; (ii) search for ASs genome-wide; (iii) check the validity of the inbreeding history, which is vital for homozygosity mapping but is often erroneous, and (iv) calculate the probability of the disease-causing gene being located in the regions identified. At the end of the analysis, we would know the probability that regions identified contain a disease-causing gene, and we would be able to determine how much effort should be devoted to scrutinizing the regions.

In the current study, we present an algorithm that implements the capabilities described in the above paragraph. We confirmed the power of this algorithm using 6 patients with Siiyama-type α1-antitrypsin deficiency, a rare autosomal recessive disease in Japan [7,8]. The preliminary version of the algorithm described here has been used to prove that the *SLC34A2* gene is responsible for pulmonary alveolar microlithiasis [9]; the current version has been used to show that the *OPTN* gene is responsible for amyotrophic lateral sclerosis [10].

Implementation

Crossover model

We used the Haldane's Poisson process model for the occurrence of crossovers and performed all calculations based on this model [11]. Information on SNPs used by Affymetrix's Genome-Wide Human SNP Array 6.0 (hereafter referred to as SNP Array 6.0) was summarized in the annotation file, [12], in which the genetic distance from the telomere of the short arm of a chromosome to each SNP was obtained by interpolation using the sex-averaged data published by deCODE Genetics [13]. We restricted our analysis to a total of 890,625 autosomal SNPs with assigned dbSNP refIDs [14].

Monte Carlo simulation

The average number, the average length, and the maximal length of the ASs derived from a common ancestor were calculated for a range of m + n values (Figure 1A) using a Monte Carlo simulation. The trial was repeated until we observed 100,000 events in which at least 1 AS appeared in the autosomal region.

The length of AS

The subject is removed from the common ancestor m generations on the paternal side and n generations on the maternal side (Figure 1A). Assuming that the length of each autosome is infinite, the length of AS conforms to an exponential distribution with a probability density function of

$$f(x) = \lambda e^{-\lambda x}\lambda = \frac{m+n}{100}\left(cM^{-1}\right). \quad (2)$$

In actuality, the autosomes have finite length; however, **equation 2** provides a good approximation when the length of an AS is much shorter than the length of an autosome.

RHS (run of homozygous SNPs), false negative, type A false positive and type B false positive

An RHS is defined as a run of homozygous SNPs with a genetic length greater than the RHS cutoff value (Figure 1B). All SNPs in an AS are homozygous, and therefore an RHS suggests the presence of an AS. We defined 3 types of errors. False negatives are ASs that are not contained in RHSs. Type A false positives are RHSs that do not contain ASs. Type B false positives are the spaces within an RHS that do not contain an AS. The false negative rate ($R_{false\ negative}$) is the ratio of false negatives to the total length of the AS. The false positive rate ($R_{false\ positive}$) is the ratio of false positives (the type A false positives plus the type B false positives) to the total length of the autosomes.

(1) $R_{false\ negative}$, the ratio of the total length of false negatives to the total length of the AS

According to the **equation 2**,

$$R_{false\ negative} = \frac{\int_0^c xf(x)dx}{\int_0^\infty xf(x)dx} = 1 - e^{-\lambda c}(1+\lambda c). \quad (3)$$

(2) $R_{Type\ A\ false\ positive}$, the ratio of the total length of type A false positives to the total length of the autosomes

Given that N_{SNP} is the total number of SNPs on a genotyping array, and P_n and Q_n are the frequencies of the major and minor alleles for the nth SNP, respectively, then the average frequencies of the major alleles ($\bar{F}_{major\ allele}$) and the minor alleles ($\bar{F}_{minor\ allele}$) are

$$\bar{F}_{major\ allele} = \frac{\sum_{n=1}^{N_{SNP}} P_n}{N_{SNP}} \text{ and } \bar{F}_{minor\ allele} = \frac{\sum_{n=1}^{N_{SNP}} Q_n}{N_{SNP}},$$

respectively. The numbers of homozygous SNPs ($N_{homozygousSNP}$) and heterozygous SNPs ($N_{heterozygousSNP}$) are approximated by

$$N_{homozygousSNP} \approx \left(\bar{F}_{major\ allele}\right)^2 N_{Pt} + \left(\bar{F}_{minor\ allele}\right)^2 N_{Pt}, \text{ and}$$

$$N_{heterozygousSNP} \approx 2\left(\bar{F}_{major\ allele}\right)\left(\bar{F}_{minor\ allele}\right) N_{Pt},$$

where N_{pt} is the number of SNPs successfully genotyped. Assuming that heterozygous SNPs are randomly

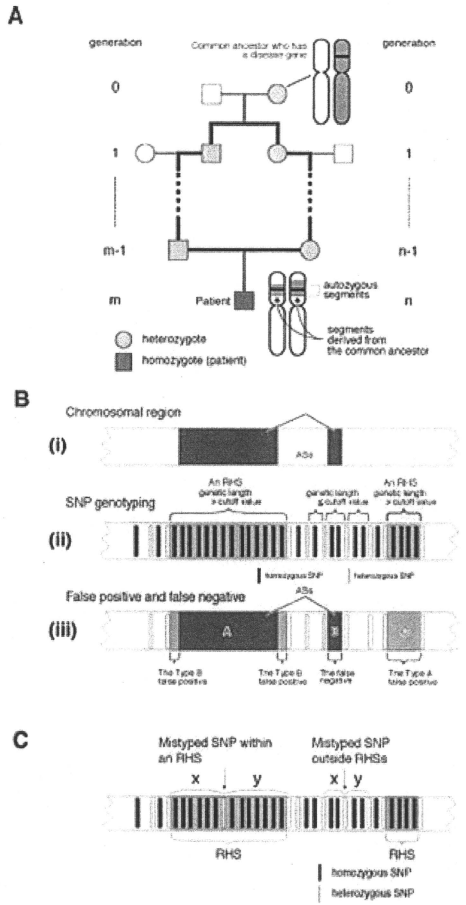

Figure 1 Connections between AS, RHS, false negative, type A false positive, and type B false positive values. (A) In a family with a consanguineous marriage, a loop is formed in the pedigree (bold lines). A chromosomal segment that is separately inherited through both sides of the arc becomes homozygous in an offspring and forms an autozygous segment. (B) (i) a chromosomal region with 2 ASs (dark gray boxes). (ii) An RHS is a region whose genetic length greater than the cutoff value. (iii) Relationship of an RHS and an AS. ASs are shown by dark gray boxes, and RHSs are shown by light gray boxes. Three types of errors are defined: false negative, type A false positive, and type B false positive. (C) Principle used for the genotyping error correction. If a homozygous SNP in an RHS is mistyped and becomes heterozygous, it is likely to have a greater distance (i.e. $x + y$) from the adjacent heterozygous SNPs than a heterozygous SNP that exists in another part of the autosomes. Therefore, heterozygous SNPs with a large $x + y$ are likely to be mistyped.

located, then the length between 2 heterozygous SNPs conforms to an exponential distribution with a probability density function of

$$f(x) = \lambda x \quad \lambda = \frac{N_{heterozygousSNP}}{L_{autosome}}(cM^{-1}), \tag{4}$$

where $L_{autosome}$ is the entire length of the autosomes. Therefore, at a cutoff value of c cM,

$$R_{Type\ A\ false\ positive} = \frac{\int_{c}^{\infty} xf(x)dx}{\int_{0}^{\infty} xf(x)dx} = (1+\lambda c)e^{-\lambda c}. \tag{5}$$

(3) $R_{Type\ B\ false\ positive}$, the ratio of the total length of type B false positives to the total length of the autosomes

$R_{Type\ B\ false\ positive}$ is not calculated mathematically but is calculated according to the actual data. An RHS containing an AS is expected to have type B false positives with an average length of $\frac{1}{2}\times\frac{L_{autosome}}{N_{heterozygousSNP}}$ on each end. It is impossible to distinguish RHSs that contain ASs from those that do not. We calculated $R_{Type\ B\ false\ positive}$ under the assumption that every RHS contains an AS. Therefore, the $R_{Type\ B\ false\ positive}$ calculation results in an overestimation, which we consider better than an underestimation for determination of the appropriate RHS cutoff. Therefore,

$$R_{Type\ B\ false\ positive} = \frac{number\ of\ RHS}{2}\times\frac{L_{autosome}}{N_{heterozygousSNP}}. \tag{6}$$

(4) $R_{false\ positive}$, the ratio of the total length of false positives to the total length of the autosomes

$$\begin{aligned} R_{false\ positive} &= R_{Type\ A\ false\ positive} + R_{Type\ B\ false\ positive} \\ &= (1+\frac{N_{heterozygousSNP}}{L_{autosome}}c)e^{-\frac{N_{heterozygousSNP}}{L_{autosome}}c} \\ &+ \frac{number\ of\ RHS}{2}\times\frac{L_{autosome}}{N_{heterozygousSNP}}. \end{aligned} \tag{7}$$

Probability that a disease-causing gene is contained in RHSs, or the overlap of RHSs

The probability that RHSs obtained contains a disease-causing gene is calculated using **equation 1**.

$$\begin{aligned} P_{GeneIsInRHS} &= (1-R_{false\ negative})\times P_{AS} \\ &= (1-R_{false\ negative})\times\frac{F}{(1-F)p+F}. \end{aligned} \tag{8}$$

Here, F is the coefficient of consanguinity and is calculated by

$$F \approx \frac{\text{total length of RHSs}}{\text{total length of the autosomes}}. \tag{9}$$

The probability that the overlap of RHSs among multiple patients contain the gene is calculated by

$$P_{GeneIsInOverlap} = \prod_{All\ patients} P_{GeneIsInRHS}. \tag{10}$$

Human Subjects and genotyping

This study was approved by the Institutional Review Boards of Saitama Medical University and Juntendo University. After obtaining written informed consent, DNA samples from 6 patients with α1-antitrypsin deficiency were purified from peripheral blood. These patients were not related and lived in different areas of Japan. Patients 1-5 were from families with a history of inbreeding because their parents were first cousins. Patient 6 did not have any family history of inbreeding. These 6 patients were genotyped using the SNP Array 6.0. The genotyping data for 86 HapMap JPT were available in the HapMap3 draft release 2 http://www.hapmap.org, and were downloaded from the Wellcome Trust Sanger Institute web site http://www.sanger.ac.uk/humgen/hapmap3/. The genotyping data for NA18987, a subject in HapMap JPT, was also distributed from Affymetrix and was used in the current study.

Genotyping error correction

Genotyping errors may convert homozygous SNPs to heterozygous SNPs and erroneously terminate an RHS, resulting in the failure to detect a portion of an RHS. According to Affymetrix, SNP Array 6.0 has an accuracy of > 0.997, implying that the genotyping error rate ($P_{genotypingError}$) may be 0.003 at maximum. A mistyped heterozygous SNP occurring in an RHS is separated by a large distance from neighboring heterozygous SNPs (Figure 1C). Therefore, if a heterozygous SNP is separated from neighboring SNPs by a distance that is rarely observed by chance, we speculated that the SNP was mistyped. Using **equation 4**, we calculated the probability of a heterozygous SNP being separated from neighboring SNPs at the observed distance ($P_{distanceOccurredByChance}$). A SNP with

$P_{distanceOccreceByChance} < 0.01$ was considered a mistyped SNP and these data were removed. This algorithm may erroneously remove 20 correctly genotyped heterozygous SNPs ($N_{homozygousSNP} \times P_{genotypingError} \times 0.01$) from a single SNP array analysis data, which we considered acceptable.

Statistical analysis

The number of patients and controls who shared an RHS at each SNP position was compared. The assumption was made that

$$u = \frac{\hat{P}_1^* - \hat{P}_2^*}{\sqrt{\hat{P}^*(1-\hat{P}^*)\left(\frac{1}{n_1}+\frac{1}{n_1}\right)}}$$

has a standard normal distribution, where $\hat{P}_2^* = \frac{x_2+0.5}{n_2+1}$, $\hat{P}_2^* = \frac{x_2+0.5}{n_2+1}$, $\hat{P}^* = \frac{x_1+x_2+0.5}{n_1+n_2+1}$. Here, x_1 and x_2 represent the numbers of patients and controls sharing RHSs, respectively, and n_1 and n_2 represent the total numbers of patients and controls, respectively. The P value was calculated by

$$P = \int_{u_0}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

Computer program

The computer program was written in the ANSI standard C programming language. The program was compiled by the GNU C compiler 4.2 and run on a MacBook Pro (CPU: 2.53 GHz Intel Core 2 Duo, 4 GB RAM) computer. The command line programs and the programs equipped by graphic user interface are both available from our web site at http://www.hhanalysis.com.

Result

Strategy

Our aim was to establish an algorithm for homozygosity mapping that uses SNP genotyping data obtained by high-density arrays, is equipped by a powerful genotyping error correction algorithm, detects ASs genome-wide, allows investigation into the family inbreeding history, and is able to calculate the probability that the identified regions contain the target gene.

The algorithm searches for the ASs (Figure 1A, B(i)) through runs of homozygous SNPs, or RHSs, that are formed by consecutively homozygous SNPs and are longer than the RHS cutoff value (Figure 1B(ii)). RHSs are presumably the autozygous segments (ASs). Three types of errors were defined; false negative, type A false positive, and type B false positive (Figure 1B(iii)). The main determinants of the false negative rate ($R_{false\ negative}$), which is the ratio of the total length of false negatives to the total length of ASs, are the number of SNPs investigated and the genotyping error rate. The main determinants of the false positive rate ($R_{false\ positive}$), which is the ratio of the total length of type A false positives plus type B false positives to the entire length of the autosomes, are the positioning of SNPs, local haplotype block structure [15], and population substructure [16].

To attain the aims stated above while avoiding the influence of these errors, our algorithm had the following steps: **Step (a)** determine an appropriate RHS cutoff value based on the Haldane's recombination model; **Step (b)** perform genotyping error correction; **Step (c)** detect RHSs; **Step (d)** obtain the overlaps of RHSs among patients; and **Step (e)** correct false positives by a case-control approach. The validity of the family history is checked at **Step (c)**. We used 5 patients with Siiyama-type α1-antitrypsin deficiency, a rare disease in Japan, to verify our strategy. Analyses performed in the Result section can be reproduced using the program contained in additional file 1 according to the tutorial also contained in the additional file 1.

Determination of the RHS cutoff

The expected false negative and false positive rates for the SNP Array 6.0 from the Haldane's model were calculated by using **equation 3** and **7 [Step (a)]** (Figure 2A). We gave the priority to reducing the false positive rate than to reducing the false negative rate, because we empirically determined that it simplified the analysis. We chose 0.6 cM as the RHS cutoff value, at which the false negative rate was 0.0006 and the false positive rate was 0.0029. The probability that the RHSs contained the disease-causing gene ($P_{GeneIsInRHS}$) at this condition was calculated using **equation 8** (Figure 2B).

Genotyping error correction

The power of the genotyping error correction algorithm was investigated using genotyping data for subject NA18987 (female) from HapMap JPT. The subject was independently genotyped in HapMap draft 3 and by Affymetrix, and data were made public from both sources. A comparison of these 2 datasets revealed that the genotyping results for 701,753 SNPs matched between these 2 sources, and they were therefore considered highly accurate. Using the matched data, RHSs were obtained with an RHS cutoff value of 0.6 cM (Figure 3A). The presence of a long RHS (36.2 cM at

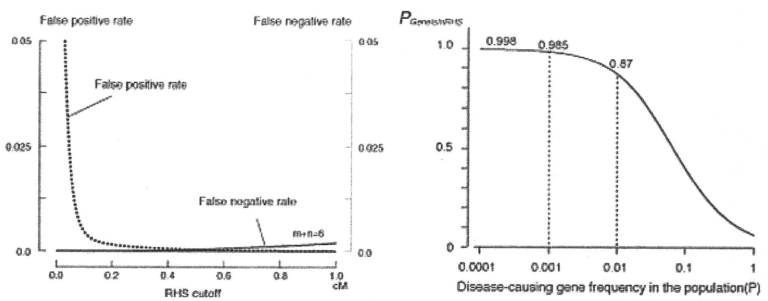

Figure 2 Determination of the RHS cutoff and the probability that the disease-causing gene is contained in RHSs. (A) The false negative rate ($R_{false\ negative}$) and the false positive rate ($R_{false\ positive}$) were calculated using **equations 3** and **7** using the genotyping data for 5 α1-antitrypsin deficiency patients. The false negative rate shown is for a child from a first-cousin marriage (m + n = 6). (B) The probability that RHSs contain the disease gene ($P_{GenelsInRHS}$) calculated for a child from a first-cousin marriage. The coefficient of consanguinity (*F*) used was 1/16, which was calculated according to Wright {Wright, S. Systems of Mating. V. General Considerations *Genetics* 1921: **6**:167-178}. *F* can be more precisely calculated as the total length of RHSs divided by the total length of the autosomes for the actual calculation (**equation 9**). $P_{GenelsInRHS}$ varies depending on the frequency of the gene in the population.

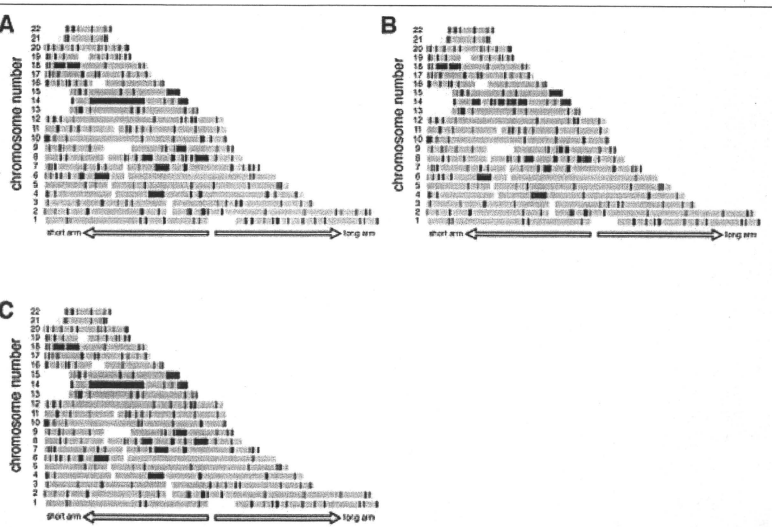

Figure 3 Genotyping error corrections. (A) RHSs for NA18987. (B) RHSs detected after introducing genotyping errors to 2,105 SNPs. (C) RHSs after the genotyping error correction algorithm was applied.

maximum) suggested that she had a family history of inbreeding, as described later. Considering the fact that the manufacturer (Affymetrix) claimed that the genotyping error rate for the SNP Array 6.0 is less than 0.003, we randomly introduced errors into selected 2,105 SNPs (701,753 SNPs × 0.003) and obtained RHSs. These error hampered the detection of RHSs, especially the long ones (Figure 3B). Following application of the genotyping error correction algorithm (Figure 1C), RHSs were restored (Figure 3C). The same trial repeated 100 times revealed that the genotyping error correction restored an average of 94.2% of the total length of all RHSs, and 99.9% of the total length of RHSs that were longer than 2 cM. This indicated that 99.9% of the total length of ASs resulting from first- or second cousin marriages would be correctly detected as RHSs after the correction. The total length of the regions that were erroneously detected as RHSs amounted to only 0.2% of the total length of the autosomes. These results indicated that the performance of the genotyping error correction algorithm was excellent.

RHSs in the patients

We applied the genotyping error correction algorithm to the data for 5 patients with Siiyama-type α1-antitrypsin deficiency [**Step (b)**], and then obtained RHSs [**Step (c)**] (Figure 4A-E). All patients had long RHSs, which were likely to be the result of first-cousin marriages.

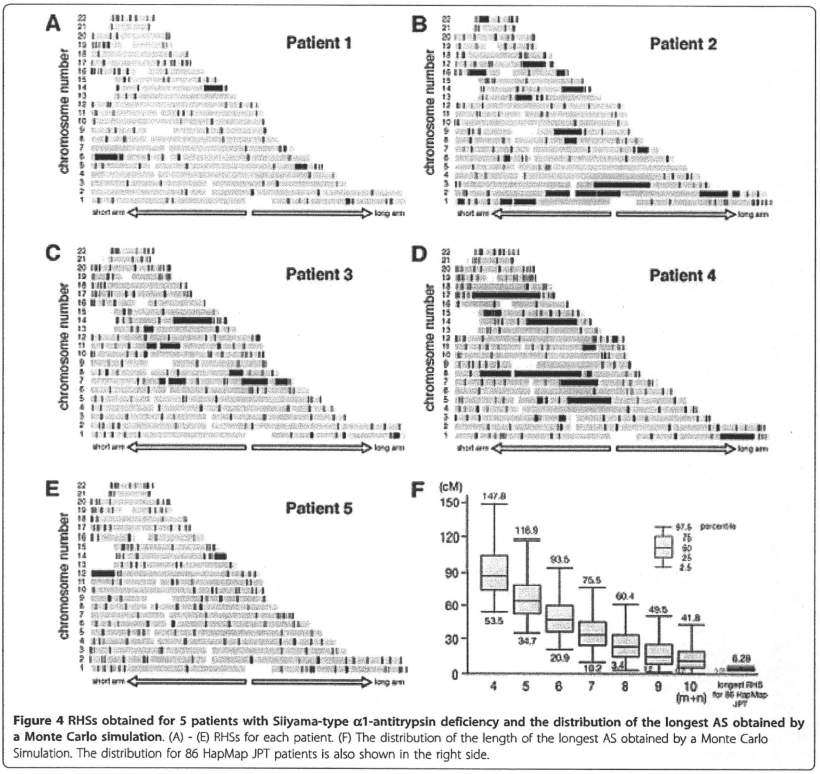

Figure 4 RHSs obtained for 5 patients with Siiyama-type α1-antitrypsin deficiency and the distribution of the longest AS obtained by a Monte Carlo simulation. (A) - (E) RHSs for each patient. (F) The distribution of the length of the longest AS obtained by a Monte Carlo Simulation. The distribution for 86 HapMap JPT patients is also shown in the right side.

Table 1 Size of the longest RHS for each patient

	Length of the longest RHS (cM)
Patient 1	36.2
Patient 2	39.6
Patient 3	22.1
Patient 4	40.3
Patient 5	30.2

Statistics of AS

We investigated whether the RHSs obtained for each patient were consistent with family history [**Step (d)**]. We focused on the size of the longest AS because they are an index of the most recent occurrence of inbreeding in the patient's family (**equation 2**). The distribution of the length of the longest AS is calculated by a Monte Carlo simulation (Figure 4F). From this distribution we are able to say that the family history of a first cousin marriage (m + n = 6) is unlikely when the longest RHS is less than 20.9 cM. The size of the longest RHS for Patients 1-5 were consistent with what expected from their family histories (Table 1).

Overlap of RHSs

We then obtained the overlaps of the RHSs for Patients 1-5 whose parents were first cousins [**Step (d)**] (Figure 5A). The probability that these regions contained the disease-causing gene ($P_{GeneIsInOverlap}$) was calculated by **equation 10** and is shown in Figure 5B. The prevalence of Siiyama-type α1-antitrypsin deficiency is less than 1 in a million in Japan, and the frequency of the gene is suspected to be less than 0.001 in the general population, indicating that the overlaps likely contained the disease-causing gene.

Some of the autosomal regions are prone to type A or type B false positives, and thus are likely to appear as an overlap [**Step (e)**]. To prioritize regions for in-depth analysis, we performed a case-control study using 86 HapMap JPT subjects as controls. One overlap had the largest $-\log_{10}(P)$ value (16.47) and was considered to be the candidate region (Figure 5C). This region (between

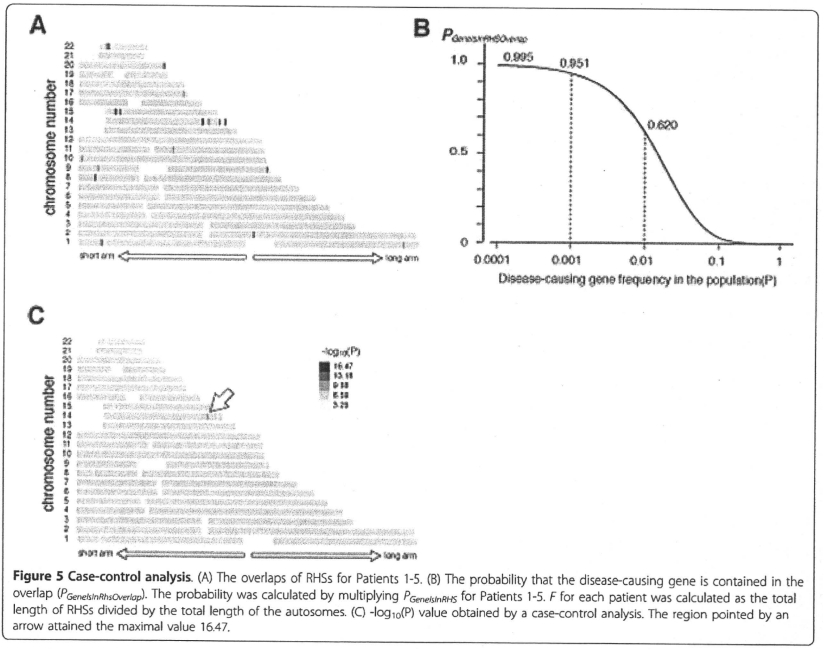

Figure 5 Case-control analysis. (A) The overlaps of RHSs for Patients 1-5. (B) The probability that the disease-causing gene is contained in the overlap ($P_{GeneIsInRhsOverlap}$). The probability was calculated by multiplying $P_{GeneIsInRHS}$ for Patients 1-5. *F* for each patient was calculated as the total length of RHSs divided by the total length of the autosomes. (C) $-\log_{10}(P)$ value obtained by a case-control analysis. The region pointed by an arrow attained the maximal value 16.47.

Table 2 Genes present in the candidate RHS overlap

C14orf48	chromosome 14 open reading frame 48
OTUB2	OTU domain, ubiquitin aldehyde binding 2
DDX24	DEAD (Asp-Glu-Ala-Asp) box polypeptide 24
IFI27L1	interferon, alpha-inducible protein 27-like 1
IFI27	interferon, alpha-inducible protein 27
IFI27L2	interferon, alpha-inducible protein 27-like 2
PPP4R4	protein phosphatase 4, regulatory subunit 4
SERPINA10	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 10
SERPINA6	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 6
LOC10028	Description: hypothetical protein LOC100287997
SERPINA2	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 2
SERPINA1	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 1
SERPINA11	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 11
SERPINA9	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 9
SERPINA12	serpin peptidase inhibitor, clade A (alpha-1 antiproteinase, antitrypsin), member 12

rs10134551 and rs910349) had a genetic length of 1.44 cM, and contained 15 genes (Table 2), one of which was the disease-causing gene for Siiyama-type α1-antitrypsin deficiency, SERPIN1.

A patient without family history of inbreeding

We occasionally encounter patients who do not have a family history of inbreeding while searching for a recessive disease-causing gene. Data from such patients are not used in the main analysis, but these data may be used for prioritizing the overlaps of RHSs as obtained in Figure 5 for an in-depth search. Patient 6 had Siiyama-type α1-antitrypsin deficiency but did not have a family history of inbreeding. The length of the longest RHS (6.8 cM, Figure 6A) was outside of the 95% range for the Japanese population (Figure 4F, **rightmost bar and**

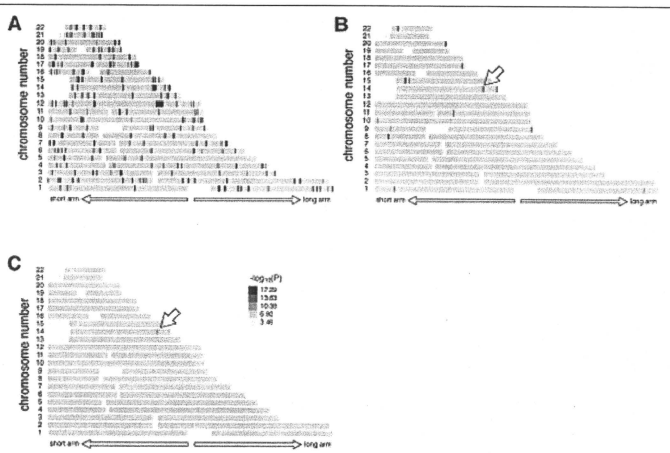

Figure 6 Subject without family history of inbreeding. (A) RHSs obtained for a patient without a family history of inbreeding (Patient 6). (B) RHS overlaps for Patients 1-6. Addition of data for Patient 6 further narrowed the overlapped regions (compare with **Figure 5**). The disease-causing gene was contained in the region indicated by a white arrow. (C) $-\log_{10}(P)$ value obtained by a case-control analysis. The region pointed by an arrow attained the maximal value 17.29.

whisker). We reasoned that the patient's family might have had forgotten inbreeding history, and that the RHSs for the patient may have a high probability of containing the disease-causing gene. This was indeed the case; addition of the data from Patient 6 excluded several overlapped regions (Figure 6B, **compare with** Figure 5A) and increased $-\log_{10}(P)$ (Figure 6C, **compare with** Figure 5C), although the list of the genes was the same as Table 2. If the length of the longest RHS suggested a hidden inbreeding history, the data for subjects without an inbreeding history could be used to prioritize some RHS overlaps for an in-depth search.

Discussion

In the current report, we described the quantitatively-modeled homozygosity mapping algorithm that uses high density array SNP genotyping data.

Homozygosity mapping is simple in principle, but many pitfalls were discovered when it was actually applied. Problems that included (i) unexpected allelic heterogeneity, (ii) identification of a homozygous identical-by-descent (IBD) region to the disease locus, (iii) underestimation of the extent of inbreeding, were pointed out in the analyses using microsatellite markers [17] and are still observed in the analyses using SNPs. Moreover, use of high-density SNP arrays introduced a novel problem, (iv) a large number of mistyped SNPs. Although the genotyping error rate is low for high-density arrays, the huge number of SNPs in these arrays inevitably produces a large number of mistyped SNPs. Even a single mistyped SNP erroneously terminates an RHS, making the detection of large RHSs difficult. Our algorithm has overcome all these problems: problem (i) is solved by using high-density SNP arrays, problem (ii) by case-control approach, problem (iii) by identifying ASs as RHSs and calculating F by the total length of RHSs divided by the total length of the autosomes, and problem (iv) by applying genotyping error correction algorithm.

As stated as Problem (ii) above, we observed some autosomal regions had a high probability of having RHSs. This may be caused by SNP positioning, local haplotype block structure, or population substructure. The effect of them was eliminated by using a case-control approach, which is performed in the order that (a) obtain overlap of RHS among patients, and (b) perform a case-control analysis targeting obtained overlaps.

Homozygosity mapping has power to identify a disease-causing gene in as few as 3 patients, and we have indeed identified the *SLC34A2* gene in pulmonary alveolar microlithiasis and the *OPTN* gene in the amyotrophic lateral sclerosis both in 3 patients [9,10]. Amyotrophic lateral sclerosis has multiple causative genes. In the latter report, we were able to identify one of the genes by investigating each combination of 3 patients from 7 patients with a history of inbreeding, seeking for 3 patients harboring the same disease-causing gene. Our algorithm worked fine in this approach. During the process, it was quite helpful that the algorithm provided the probability that the identified regions contain the disease-causing gene, which determined how much effort should be further devoted. To our knowledge, the algorithm presented in the current study is the first to provide this information.

Conclusions

We described an algorithm that enables homozygosity mapping to be performed based on a quantitative model using SNP genotyping data. Our procedure will accelerate the identification of disease-causing genes using high-density SNP array data.

Availability and requirements

Project name: qHomozygosityMapping
Project home page: http://www.hhanalysis.com
Operating system(s): Mac, Linux and Windows.
Programming language: C
License: GNU GPL.
Any restrictions to use by non-academics: The software is for academic purpose only.

Funding

This work is supported in part by the grant-in-aid for scientific research (No. 18390242) from the Japan Society of Promotion of Science, and in part by the grants-in-aid for Health and Labor Science (Nos. H22-Nanchi-Ippan-005 and H20-Nanchi-Ippan-023) from the Ministry of Health, labor and Welfare, Japan.

Additional material

Additional file 1: This file is a zipped package that contains programs and tutorial for Linux, MacOS X and Windows platforms.

Acknowledgements

The authors thank Ms. Tomoko Hirata for her technical assistance.
This article has been published as part of BMC Bioinformatics Volume 11 Supplement 7, 2010: Ninth International Conference on Bioinformatics (InCoB2010): Bioinformatics. The full contents of the supplement are available online at http://www.biomedcentral.com/1471-2105/11?issue=S7.

Author details

[1]Department of Respiratory Medicine, Saitama Medical University, 38 Morohongo, Moroyama, Saitama 350-0495, Japan. [2]Department of Medical Oncology, The Affiliated Hospital of Inner Mongolia Medical College, Tong Dao Bei Jie, 010050 Hohhot, China. [3]Department of Epidemiology, Research Institute for Radiation Biology and Medicine, Hiroshima University, Hiroshima 734-8553, Japan. [4]Division of Functional Genomics and Systems Medicine, Research Center for Genomic Medicine, Saitama Medical University, 1397-1 Yamane, Hidaka City, Saitama 350-1241, Japan. [5]Department of Respiratory Medicine, Juntendo University, School of Medicine, 2-1-1 Hongo, Bunkyo-ku, Tokyo 113-8421, Japan.

Authors' contribution
Huqun, S.F., H.M., T.T., T.S., M.K., H.K., Y.O., and K.S. tested the programs, did genetic analyses and provided ideas to improve the program. K.S. collected the patients' samples. K.H. provided basic ideas, wrote the program, and prepared manuscript.

Competing interests
The authors declare that they have no competing interests.

Published: 15 October 2010

References

1. McKusick VA: **Mendelian Inheritance in Man and its online version, OMIM.** *Am J Hum Genet* 2007, **80**:, 588-604.
2. **OMIM-Online Mendelian Inheritance in Man.** [http://www.ncbi.nlm.nih.gov/Omim/mimstats.html].
3. Lander ES, Botstein D: **Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children.** *Science* 1987, **236**:1567-1570.
4. Clark AG: **The size distribution of homozygous segments in the human genome.** *Am J Hum Genet* 1999, **65**:1489-1492.
5. Woods CG, Valente EM, Bond J, Roberts E: **A new method for autozygosity mapping using single nucleotide polymorphisms (SNPs) and EXCLUDEAR.** *J Med Genet* 2004, **41**:e101.
6. Seelow D, Schuelke M, Hildebrandt F, Nurnberg P: **HomozygosityMapper–an interactive approach to homozygosity mapping.** *Nucleic Acids Res* 2009, **37**:W593-599.
7. Seyama K: **State of alpha1-antitrypsin deficiency in Japan.** *Respirology* 2001, **6(Suppl)**:S35-38.
8. Seyama K, Nukiwa T, Souma S, Shimizu K, Kira S: **Alpha 1-antitrypsin-deficient variant Siiyama (Ser53[TCC] to Phe53[TTC]) is prevalent in Japan. Status of alpha 1-antitrypsin deficiency in Japan.** *Am J Respir Crit Care Med* 1995, **152**:2119-2126.
9. Izumi S, Miyazawa H, Ishii K, Uchiyama B, Ishida T, Tanaka S, Tazawa R, Fukuyama S, Tanaka T, Nagai Y, Yokote A, Takahashi H, Fukushima T, Kobayashi K, Chiba H, Nagata M, Sakamoto S, Nakata K, Takebayashi Y, Shimizu Y, Kaneko K, Shimizu M, Kanazawa M, Abe S, Inoue Y, Takenoshita S, Yoshimura K, Kudo K, Tachibana T, Nukiwa T, Hagiwara K: **Mutations in the SLC34A2 gene are associated with pulmonary alveolar microlithiasis.** *Am J Respir Crit Care Med* 2007, **175**:263-268.
10. Maruyama H, Morino H, Ito H, Izumi Y, Kato H, Watanabe Y, Kinoshita Y, Kamada M, Nodera H, Suzuki H, Komure O, Matsuura S, Kobatake K, Morimoto N, Abe K, Suzuki N, Aoki M, Kawata A, Hirai T, Kato T, Ogasawara K, Hirano A, Takemi T, Kusaka H, Hagiwara K, Kaji R, Kawakami H: **Mutations of optineurin in amyotrophic lateral sclerosis.** *Nature* 2010, **465**:223-226.
11. Haldane J: **The combination of linkage values, and the calculation of distances between the loci of linked factors.** *J Genet* 1919, **8**:299-309.
12. **Affymetrix - Home.** [http://www.affymetrix.com/index.affx].
13. Kong A, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K: **A high-resolution recombination map of the human genome.** *Nat Genet* 2002, **31**:241-247.
14. **National Center for Biotechnology Information.** [http://www.ncbi.nlm.nih.gov].
15. International HapMap Consortium: **The International HapMap Project.** *Nature* 2003, **426**:789-796.
16. Overall AD, Nichols RA: **A method for distinguishing consanguinity and population substructure using multilocus genotype data.** *Mol Biol Evol* 2001, **18**:2048-2056.
17. Miano MG, Jacobson SG, Carothers A, Hanson I, Teague P, Lovell J, Cideciyan AV, Haider N, Stone EM, Sheffield VC, Wright AF: **Pitfalls in homozygosity mapping.** *Am J Hum Genet* 2000, **67**:1348-1351.

doi:10.1186/1471-2105-11-S7-S5
Cite this article as: Huqun *et al.*: **A quantitatively-modeled homozygosity mapping algorithm, qHomozygosityMapping, utilizing whole genome single nucleotide polymorphism genotyping data.** *BMC Bioinformatics* 2010 **11**(Suppl 7):S5.

肺の稀少疾患とその医学的意義

The contribution of rare lung diseases to our knowledge in lung biology

埼玉医科大学呼吸器内科教授　**萩原　弘一**　*Koichi Hagiwara*

Key words

疾患遺伝子，単遺伝子疾患，GWAS，人類アフリカ起源説

Summary

肺の稀少疾患では，肺の発生，恒常性維持に必須の生化学経路が障害され，特異な症状を発症するものが多い。肺稀少疾患のかなりの部分は単遺伝子疾患であり，これらの遺伝子は肺の生理学に関して多くの情報を与えている。現在多因子疾患と考えられている肺疾患でも，関与している遺伝子が多数なのか少数なのか明確ではない。アフリカ以外の人類は，5～10万年前にアフリカを出発した少数の人類の子孫であるが，いくつかの疾患はその移動の道筋を辿るように分布する。稀少疾患の分布は，その移動の末端の道筋をみているといえる。稀少疾患で得られた知識は，より頻度の高い疾患の本態の研究に，大きな役割を演じると考えられる。

Ⅰ　肺の稀少疾患と肺の生理学

肺の稀少疾患といわれる一群の疾患がある。多くの医師は，一般に稀少疾患の一部しか経験しない。公衆衛生上のインパクトが小さい珍しい疾患が，なぜその患者数以上に重要視されるのか。

その理由は単純である。稀少疾患は，その成り立ちを追求することにより，肺の生理学について大きな情報を与えてくれるからである。1つの疾患単位が教えてくれる情報量には患者数の差ほどの開きはない。肺の稀少疾患は，肺生理を理解するうえで，よくある疾患に負けない貴重な情報を与えてくれる。現時点で知られている肺の稀少疾患から，肺に関連する遺伝子，シグナル伝達系，特定の免疫反応の役割に関して，重要な情報が得られている。そしてこれらは，より頻度の高い疾患を解決していくうえで，必要不可欠な知識となるだろう。

Ⅱ　肺の稀少疾患はなぜ稀少なのか

肺の稀少疾患はなぜ稀少なのか。理由の1つは，症状が生存に不利になるため集団から遺伝子が除去されやすく，患者数が少数に留まること，もう1つは，少数例でも発見できるほど症状が「目立つ」疾患であることである。稀少疾患のうち，疾患遺伝子が関与する疾患では家族内発生も明確である。突然変異で新たな異常遺伝子が生

じたり，先祖からの遺伝子が表現型として現れた場合に，少数の患者が発生しただけで診断が可能なことが，稀少疾患が稀少なままみつかる大きな原因と考えられる。

症状が特異ということは，その原因も特異ということであり，そのため肺の生理学に大きな情報を与えてくれる。肺の稀少疾患のかなりのものは遺伝疾患であるが，逆は成り立たない。遺伝疾患であるから稀少とはいえない。遺伝疾患でも，生殖年齢を超えて生存する疾患では自然選択が働きにくい。高齢で発症する疾患の遺伝子は排除されるどころか，集団内に広がる場合もある。その結果，ある地域での「稀少疾患」が，他の地域ではごく普通にみられ，時にあまりに頻度が高いため「体質」と考えられる場合さえある。日本では稀少疾患であるα1アンチトリプシン欠損症は，欧米では普通の疾患である。お酒を飲めない「下戸」の人はアセトアルデヒド脱水素酵素2(ALDH2)に変異があり，日本では「体質」と思われるほど頻繁にみられるが，西欧では「稀少疾患」である。

III 肺の稀少疾患と単遺伝子疾患

肺の稀少疾患のうち，かなりのものは単遺伝子疾患である。遺伝子の疾患発生における強さは「浸透率」で表される(**図1**)[1]。遺伝子異常を有する人のうち，実際に疾患が発生する割合を浸透率という。肺稀少疾患の原因遺伝子の浸透率は1に近いものが多い。

浸透率が1に近いと，疾患はメンデルの遺伝実験をみているかのように家族内で発生する。そのために気付かれやすく，遺伝子の同定も進む。浸透率が1に近いということは，その遺伝子異常が特定の生体システムの中核的部分にあり，他の遺伝子が欠損を補助することができず，1つの遺伝子異常のみで疾患が明確になってしまうことを示している。近年，基礎医学分野で特定の遺伝子機能のみを喪失させた「ノックアウトマウス」を作製し，それぞれの遺伝子機能を調べる実験が盛んだが，これはマウスで単遺伝子疾患家系を作製しているのと同じである。重要不可欠な遺伝子のノックアウトマウスは，特異な表現型を示し，その表現型はメンデルの遺伝法則に従って遺伝する。

しかし，遺伝子機能には種差があることに注意する必要がある。ヒトの単遺伝子疾患の原因遺伝子と同一の遺伝子をマウスで欠損させた場合に，同様の疾患が起こる場合もあるし，起こらない場合もある。顆粒球マクロファージコロニー刺激因子(GM-CSF)受容体遺伝子の欠損では，ヒトでもマウスでも肺胞蛋白症が起こる。しかしながら，IIb型ナトリウム依存性リン運搬遺伝子(SLC34A2)の欠損では，ヒトでは肺胞微石症が起こるものの患者は中年以降まで無症状で生存するが，マウスでは子宮内で発生のごく初期に死亡してしまう。ヒトの疾患原因，肺の生理学はヒトでしかわからない場合も多い。肺の稀少疾患は，マウスではなく，ヒトの肺で何が大事なのかを教えてくれる。

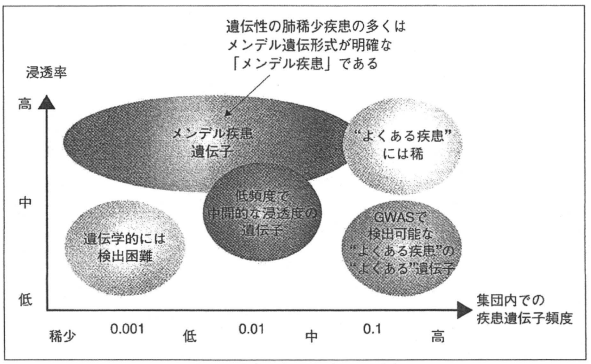

図1．疾患遺伝子頻度と浸透率

「メンデル疾患の疾患遺伝子」は家系解析で，「よくある疾患のよくある遺伝子」はGWASで検出可能である。「低頻度で中間的な浸透度の遺伝子」は，われわれが日常よくみる疾患に関与していると考えられるが，適切な検出手法が確立されていない。

(文献1より引用・改変)