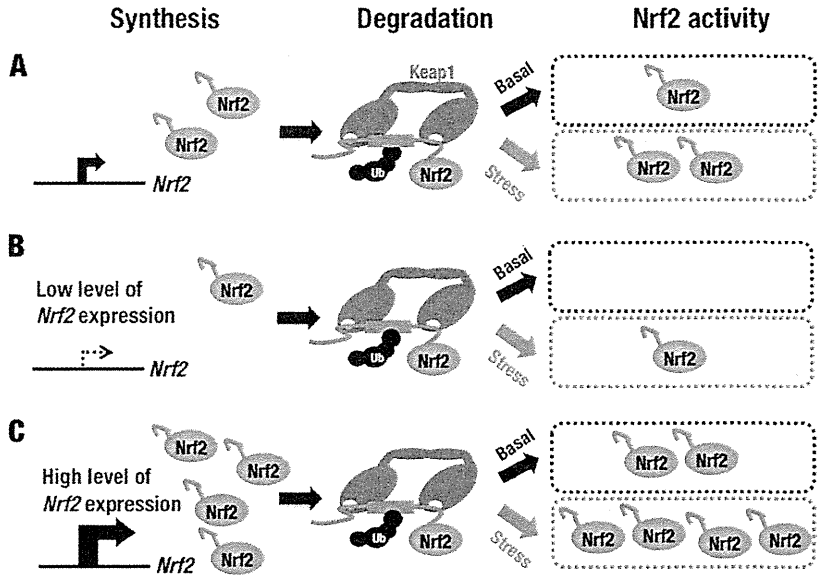

FIG 5 Dual regulation of Nrf2 activity by synthesis and degradation. (A) When an adequate level of the *Nrf2* transcript is supplied, the Nrf2 protein is maintained at low levels by Keap1-mediated degradation under the basal condition, and a relatively high level of the Nrf2 protein is accumulated after the inactivation of Keap1. (B) When the *Nrf2* transcription level is low, the Nrf2 protein is decreased in both the basal and induced states. (C) When a high *Nrf2* transcript level is achieved, a large amount of Nrf2 is produced in both the basal and induced states.

In addition to NRF2 rSNP-617, there appear to be other causes that result in a reduction of Nrf2 synthesis. Indeed, *Nrf2* expression is decreased in aged smokers and patients with chronic obstructive pulmonary diseases (35). We showed that transgenic overexpression of Nrf2 significantly increases the Nrf2 protein level and makes cells resistant to oxidative stress. These observations suggest that induction of the Nrf2 transcript is an effective approach for enhancing the activity of Nrf2. To date, several factors have been found to increase the Nrf2 transcript level, including Nrf2 itself by autoactivation (19), the aryl hydrocarbon receptor (36), and Jun (37).

Increasing numbers of studies have revealed that NRF2 is also involved in the malignant progression of various human cancers (13–15). Constitutive activation of NRF2 gives a strong advantage to cancer cells by conferring chemo- and radioresistance and accelerating proliferation (38–40). A recent study demonstrated that NRF2 is constitutively activated in lung cancer cells bearing *EGFR* mutations (41). Transcriptional activation of *NRF2* by the RAS oncogenic pathway also substantially contributes to the enhancement of NRF2 activity in cancer (42). Accordingly, we expect that *EGFR* and *KRAS* mutations are associated with NRF2 rSNP-617. We found that there are associations of the minor A/A homozygotes with the risk for developing lung cancers with *EGFR* and *KRAS* mutations (Table 3). One plausible explanation for this association is that *EGFR* and *KRAS* mutations may compensate for the compromised transcription of *NRF2*, allowing the A/A homozygous cancer cells to achieve a sufficiently high level of NRF2 activity.

In summary, Nrf2 has been considered a transcription factor which is mainly regulated by a posttranslational derepression mechanism. In this study, we found that weakened transcription of the *Nrf2* gene provides a basis for the development of lung cancers, possibly through the reduced expression of cytoprotective enzymes. This observation leads to the notion that in addition to control over proteasomal degradation and derepression from degradation/repression, transcriptional regulation of the *Nrf2* gene in response to various signals/insults is an important pathway in determining cellular Nrf2 levels. Collectively, the contribution of NRF2 to these physiological and pathological processes is regulated at various nodes, and the delineation of these mechanisms is a critical step toward a better understanding of our defense machinery.

ACKNOWLEDGMENTS

We thank Eriko Naganuma, Hiromi Suda, Yukie Kawatani, Naoko Okada, and Akiko Kokubu for technical assistance. We also thank Liam Baird for discussion and advice.

This work was supported in part by Grants-in-Aid for Creative Scientific Research and Scientific Research from JSPS, the Target Protein Program from MEXT, the Tohoku University Global COE Program for Conquest of Signal Transduction Diseases with Network Medicine, CREST from JST, the NAITO Foundation, the Takeda Foundation, and grants-in-aid from the Ministry of Health, Labor, and Welfare for the 3rd-Term Comprehensive 10-Year Strategy for Cancer Control and Research on Applying Health Technology.

REFERENCES

1. Motohashi H, Yamamoto M. 2004. Nrf2-Keap1 defines a physiologically important stress response mechanism. Trends Mol. Med. 10:549–557.
2. Friling RS, Bensimon A, Tichauer Y, Daniel V. 1990. Xenobiotic-inducible expression of murine glutathione S-transferase Ya subunit gene is controlled by an electrophile-responsive element. Proc. Natl. Acad. Sci. U. S. A. 87:6258–6262.
3. Rushmore TH, Morton MR, Pickett CB. 1991. The antioxidant responsive element. Activation by oxidative stress and identification of the DNA consensus sequence required for functional activity. J. Biol. Chem. 266: 11632–11639.
4. Kobayashi M, Yamamoto M. 2006. Nrf2-Keap1 regulation of cellular defense mechanisms against electrophiles and reactive oxygen species. Adv. Enzyme Regul. 46:113–140.
5. Ishii T, Itoh K, Takahashi S, Sato H, Yanagawa T, Katoh Y, Bannai S,

Yamamoto M. 2000. Transcription factor Nrf2 coordinately regulates a group of oxidative stress-inducible genes in macrophages. J. Biol. Chem. 275:16023–16029.
6. Itoh K, Chiba T, Takahashi S, Ishii T, Igarashi K, Katoh Y, Oyake T, Hayashi N, Satoh K, Hatayama I, Yamamoto M, Nabeshima Y. 1997. An Nrf2/small Maf heterodimer mediates the induction of phase II detoxifying enzyme genes through antioxidant response elements. Biochem. Biophys. Res. Commun. 236:313–322.
7. Aoki Y, Sato H, Nishimura N, Takahashi S, Itoh K, Yamamoto M. 2001. Accelerated DNA adduct formation in the lung of the Nrf2 knockout mouse exposed to diesel exhaust. Toxicol. Appl. Pharmacol. 173:154–160.
8. Cho HY, Jedlicka AE, Reddy SP, Kensler TW, Yamamoto M, Zhang LY, Kleeberger SR. 2002. Role of NRF2 in protection against hyperoxic lung injury in mice. Am. J. Respir. Cell Mol. Biol. 26:175–182.
9. Enomoto A, Itoh K, Nagayoshi E, Haruta J, Kimura T, O'Connor T, Harada T, Yamamoto M. 2001. High sensitivity of Nrf2 knockout mice to acetaminophen hepatotoxicity associated with decreased expression of ARE-regulated drug metabolizing enzymes and antioxidant genes. Toxicol. Sci. 59:169–177.
10. Iida K, Itoh K, Kumagai Y, Oyasu R, Hattori K, Kawai K, Shimazui T, Akaza H, Yamamoto M. 2004. Nrf2 is essential for the chemopreventive efficacy of oltipraz against urinary bladder carcinogenesis. Cancer Res. 64:6424–6431.
11. Ramos-Gomez M, Kwak MK, Dolan PM, Itoh K, Yamamoto M, Talalay P, Kensler TW. 2001. Sensitivity to carcinogenesis is increased and chemoprotective efficacy of enzyme inducers is lost in nrf2 transcription factor-deficient mice. Proc. Natl. Acad. Sci. U. S. A. 98:3410–3415.
12. Xu C, Huang MT, Shen G, Yuan X, Lin W, Khor TO, Conney AH, Kong AN. 2006. Inhibition of 7,12-dimethylbenz(a)anthracene-induced skin tumorigenesis in C57BL/6 mice by sulforaphane is mediated by nuclear factor E2-related factor 2. Cancer Res. 66:8293–8296.
13. Padmanabhan B, Tong KI, Ohta T, Nakamura Y, Scharlock M, Ohtsuji M, Kang MI, Kobayashi A, Yokoyama S, Yamamoto M. 2006. Structural basis for defects of Keap1 activity provoked by its point mutations in lung cancer. Mol. Cell 21:689–700.
14. Shibata T, Ohta T, Tong KI, Kokubu A, Odogawa R, Tsuta K, Asamura H, Yamamoto M, Hirohashi S. 2008. Cancer related mutations in NRF2 impair its recognition by Keap1-Cul3 E3 ligase and promote malignancy. Proc. Natl. Acad. Sci. U. S. A. 105:13568–13573.
15. Singh A, Misra V, Thimmulappa RK, Lee H, Ames S, Hoque MO, Herman JG, Baylin SB, Sidransky D, Gabrielson E, Brock MV, Biswal S. 2006. Dysfunctional KEAP1-NRF2 interaction in non-small-cell lung cancer. PLoS Med. 3:e420. doi:10.1371/journal.pmed.0030420.
16. Itoh K, Wakabayashi N, Katoh Y, Ishii T, Igarashi K, Engel JD, Yamamoto M. 1999. Keap1 represses nuclear activation of antioxidant responsive elements by Nrf2 through binding to the amino-terminal Neh2 domain. Genes Dev. 13:76–86.
17. Kobayashi A, Kang MI, Watai Y, Tong KI, Shibata T, Uchida K, Yamamoto M. 2006. Oxidative and electrophilic stresses activate Nrf2 through inhibition of ubiquitination activity of Keap1. Mol. Cell. Biol. 26:221–229.
18. Itoh K, Wakabayashi N, Katoh Y, Ishii T, O'Connor T, Yamamoto M. 2003. Keap1 regulates both cytoplasmic-nuclear shuttling and degradation of Nrf2 in response to electrophiles. Genes Cells 8:379–391.
19. Kwak MK, Itoh K, Yamamoto M, Kensler TW. 2002. Enhanced expression of the transcription factor Nrf2 by cancer chemopreventive agents: role of antioxidant response element-like sequences in the nrf2 promoter. Mol. Cell. Biol. 22:2883–2892.
20. Yamamoto T, Yoh K, Kobayashi A, Ishii Y, Kure S, Koyama A, Sakamoto T, Sekizawa K, Motohashi H, Yamamoto M. 2004. Identification of polymorphisms in the promoter region of the human NRF2 gene. Biochem. Biophys. Res. Commun. 321:72–79.
21. Marzec JM, Christie JD, Reddy SP, Jedlicka AE, Vuong H, Lanken PN, Aplenc R, Yamamoto T, Yamamoto M, Cho HY, Kleeberger SR. 2007. Functional polymorphisms in the transcription factor NRF2 in humans increase the risk of acute lung injury. FASEB J. 21:2237–2246.
22. Bouligand J, Cabaret O, Canonico M, Verstuyft C, Dubert L, Becquemont L, Guiochon-Mantel A, Scarabin PY, Estrogen and Thromboembolism Risk (ESTHER) Study Group. 2011. Effect of NFE2L2 genetic polymorphism on the association between oral estrogen therapy and the risk of venous thromboembolism in postmenopausal women. Clin. Pharmacol. Ther. 89:60–64.
23. Masuko H, Sakamoto T, Kaneko Y, Iijima H, Naito T, Noguchi E, Hirota T, Tamari M, Hizawa N. 2011. An interaction between Nrf2 polymorphisms and smoking status affects annual decline in FEV1: a longitudinal retrospective cohort study. BMC Med. Genet. 12:97. doi:10.1186/1471-2350-12-97.
24. Marczak ED, Marzec J, Zeldin DC, Kleeberger SR, Brown NJ, Pretorius M, Lee CR. 2012. Polymorphisms in the transcription factor NRF2 and forearm vasodilator responses in humans. Pharmacogenet. Genomics 22:620–628.
25. Wakabayashi N, Itoh K, Wakabayashi J, Motohashi H, Noda S, Takahashi S, Imakado S, Kotsuji T, Otsuka F, Roop DR, Harada T, Engel JD, Yamamoto M. 2003. Keap1-null mutation leads to postnatal lethality due to constitutive Nrf2 activation. Nat. Genet. 35:238–245.
26. Yamamoto T, Suzuki T, Kobayashi A, Wakabayashi J, Maher J, Motohashi H, Yamamoto M. 2008. Physiological significance of reactive cysteine residues of Keap1 in determining Nrf2 activity. Mol. Cell. Biol. 28:2758–2770.
27. Maruyama A, Tsukamoto S, Nishikawa K, Yoshida A, Harada N, Motojima K, Ishii T, Nakane A, Yamamoto M, Itoh K. 2008. Nrf2 regulates the alternative first exons of CD36 in macrophages through specific antioxidant response elements. Arch. Biochem. Biophys. 477:139–145.
28. Travis WD. 2004. Pathology and genetics: tumours of the lung, pleura, thymus and heart. International Agency for Research on Cancer Press, Lyon, France.
29. Takano T, Ohe Y, Sakamoto H, Tsuta K, Matsuno Y, Tateishi U, Yamamoto S, Nokihara H, Yamamoto N, Sekine I, Kunitoh H, Shibata T, Sakiyama T, Yoshida T, Tamura T. 2005. Epidermal growth factor receptor gene mutations and increased copy numbers predict gefitinib sensitivity in patients with recurrent non-small-cell lung cancer. J. Clin. Oncol. 23:6829–6837.
30. Higgins LG, Kelleher MO, Eggleston IM, Itoh K, Yamamoto M, Hayes JD. 2009. Transcription factor Nrf2 mediates an adaptive response to sulforaphane that protects fibroblasts in vitro against the cytotoxic effects of electrophiles, peroxides and redox-cycling agents. Toxicol. Appl. Pharmacol. 237:267–280.
31. Kosaka T, Yatabe Y, Endoh H, Kuwano H, Takahashi T, Mitsudomi T. 2004. Mutations of the epidermal growth factor receptor gene in lung cancer: biological and clinical implications. Cancer Res. 64:8919–8923.
32. McKay JD, Hung RJ, Gaborieau V, Boffetta P, Chabrier A, Byrnes G, Zaridze D, Mukeria A, Szeszenia-Dabrowska N, Lissowska J, Rudnai P, Fabianova E, Mates D, Bencko V, Foretova L, Janout V, McLaughlin J, Shepherd F, Montpetit A, Narod S, Krokan HE, Skorpen F, Elvestad MB, Vatten L, Njølstad I, Axelsson T, Chen C, Goodman G, Barnett M, Loomis MM, Lubiński J, Matyjasik J, Lener M, Oszutowska D, Field J, Liloglou T, Xinarianos G, Cassidy A, Vineis P, Clavel-Chapelon F, Palli D, Tumino R, Krogh V, Panico S, González CA, Ramón Quirós J, Martínez C, Navarro C, Ardanaz E, Larrañaga N, et al. 2008. Lung cancer susceptibility locus at 5p15.33. Nat. Genet. 40:1404–1406.
33. Wang Y, Broderick P, Webb E, Wu X, Vijayakrishnan J, Matakidou A, Qureshi M, Dong Q, Gu X, Chen WV, Spitz MR, Eisen T, Amos CI, Houlston RS. 2008. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. Nat. Genet. 40:1407–1409.
34. Rotunno M, Yu K, Lubin JH, Consonni D, Pesatori AC, Goldstein AM, Goldin LR, Wacholder S, Welch R, Burdette L, Chanock SJ, Bertazzi PA, Tucker MA, Caporaso NE, Chatterjee N, Bergen AW, Landi MT. 2009. Phase I metabolic genes and risk of lung cancer: multiple polymorphisms and mRNA expression. PLoS One 4:e5652. doi:10.1371/journal.pone.0005652.
35. Suzuki M, Betsuyaku T, Ito Y, Nagai K, Nasuhara Y, Kaga K, Kondo S, Nishimura M. 2008. Down-regulated NF-E2-related factor 2 in pulmonary macrophages of aged smokers and patients with chronic obstructive pulmonary disease. Am. J. Respir. Cell Mol. Biol. 39:673–682.
36. Miao W, Hu L, Scrivens PJ, Batist G. 2005. Transcriptional regulation of NF-E2 p45-related factor (NRF2) expression by the aryl hydrocarbon receptor-xenobiotic response element signaling pathway: direct cross-talk between phase I and II drug-metabolizing enzymes. J. Biol. Chem. 280:20340–20348.
37. Meixner A, Karreth F, Kenner L, Penninger JM, Wagner EF. 2010. Jun and JunD-dependent functions in cell proliferation and stress response. Cell Death Differ. 17:1409–1419.
38. Mitsuishi Y, Taguchi K, Kawatani Y, Shibata T, Nukiwa T, Aburatani

H, Yamamoto M, Motohashi H. 2012. Nrf2 redirects glucose and glutamine into anabolic pathways in metabolic reprogramming. Cancer Cell 22:66–79.
39. Shibata T, Kokubu A, Saito S, Narisawa-Saito M, Sasaki H, Aoyagi K, Yoshimatsu Y, Tachimori Y, Kushima R, Kiyono T, Yamamoto M. 2011. NRF2 mutation confers malignant potential and resistance to chemoradiation therapy in advanced esophageal squamous cancer. Neoplasia 13:864–873.
40. Singh A, Boldin-Adamsky S, Thimmulappa RK, Rath SK, Ashush H, Coulter J, Blackford A, Goodman SN, Bunz F, Watson WH, Gabrielson E, Feinstein E, Biswal S. 2008. RNAi-mediated silencing of nuclear factor erythroid-2-related factor 2 gene expression in non-small cell lung cancer inhibits tumor growth and increases efficacy of chemotherapy. Cancer Res. 68:7975–7984.
41. Yamadori T, Ishii Y, Homma S, Morishima Y, Kurishima K, Itoh K, Yamamoto M, Minami Y, Noguchi M, Hizawa N. 2012. Molecular mechanisms for the regulation of Nrf2-mediated cell proliferation in non-small-cell lung cancers. Oncogene 31:4768–4777.
42. DeNicola GM, Karreth FA, Humpton TJ, Gopinathan A, Wei C, Frese K, Mangal D, Yu KH, Yeo CJ, Calhoun ES, Scrimieri F, Winter JM, Hruban RH, Iacobuzio-Donahue C, Kern SE, Blair IA, Tuveson DA. 2011. Oncogene-induced Nrf2 transcription promotes ROS detoxification and tumorigenesis. Nature 475:106–109.
43. Hayes JD, McMahon M, Chowdhry S, Dinkova-Kostova AT. 2010. Cancer chemoprevention mechanisms mediated through the Keap1-Nrf2 pathway. Antioxid. Redox Signal. 13:1713–1748.
44. Kobayashi A, Kang MI, Okawa H, Ohtsuji M, Zenke Y, Chiba T, Igarashi K, Yamamoto M. 2004. Oxidative stress sensor Keap1 functions as an adaptor for Cul3-based E3 ligase to regulate proteasomal degradation of Nrf2. Mol. Cell. Biol. 24:7130–7139.

Cancer Science
The official journal of the Japanese Cancer Association JCA

Identification of a lung adenocarcinoma cell line with CCDC6-RET fusion gene and the effect of RET inhibitors *in vitro* and *in vivo*

Makito Suzuki,[1,2,11] Hideki Makinoshima,[1,11] Shingo Matsumoto,[1,3] Ayako Suzuki,[4] Sachiyo Mimaki,[1] Koutatsu Matsushima,[1,2] Kiyotaka Yoh,[3] Koichi Goto,[3] Yutaka Suzuki,[4] Genichiro Ishii,[5] Atsushi Ochiai,[5] Koji Tsuta,[6] Tatsuhiro Shibata,[7] Takashi Kohno,[8] Hiroyasu Esumi[9] and Katsuya Tsuchihara[1,2,10]

[1]Division of Translational Research, Research Center for Innovative Oncology, National Cancer Center Hospital East, Kashiwa, Chiba; [2]Department of Integrated Biosciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba; [3]Division of Thoracic Oncology, National Cancer Center Hospital East, Kashiwa, Chiba; [4]Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Kashiwa, Chiba; [5]Division of Pathology, Research Center for Innovative Oncology, National Cancer Center Hospital East, Kashiwa, Chiba; [6]Division of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo; Divisions of [7]Cancer Genomics; [8]Genome Biology, National Cancer Center Research Institute, Tokyo; [9]National Cancer Center Hospital East, Kashiwa, Chiba, Japan

(Received December 4, 2012/Revised April 1, 2013/Accepted April 1, 2013/Accepted manuscript online April 11, 2013/Article first published online May 12, 2013)

Rearrangements of the proto-oncogene *RET* are newly identified potential driver mutations in lung adenocarcinoma (LAD). However, the absence of cell lines harboring *RET* fusion genes has hampered the investigation of the biological relevance of RET and the development of RET-targeted therapy. Thus, we aimed to identify a *RET* fusion positive LAD cell line. Eleven LAD cell lines were screened for *RET* fusion transcripts by reverse transcription-polymerase chain reaction. The biological relevance of the *CCDC6-RET* gene products was assessed by cell growth, survival and phosphorylation of ERK1/2 and AKT with or without the suppression of RET expression using RNA interference. The efficacy of RET inhibitors was evaluated *in vitro* using a culture system and in an *in vivo* xenograft model. Expression of the *CCDC6-RET* fusion gene in LC-2/ad cells was demonstrated by the mRNA and protein levels, and the genomic break-point was confirmed by genomic DNA sequencing. Mutations in *KRAS* and *EGFR* were not observed in the LC-2/ad cells. CCDC6-RET was constitutively active, and the introduction of a siRNA targeting the *RET 3'* region decreased cell proliferation by downregulating RET and ERK1/2 phosphorylation. Moreover, treatment with RET-inhibitors, including vandetanib, reduced cell viability, which was accompanied by the downregulation of the AKT and ERK1/2 signaling pathways. Vandetanib exhibited anti-tumor effects in the xenograft model. Endogenously expressing CCDC6-RET contributed to cell growth. The inhibition of kinase activity could be an effective treatment strategy for LAD. LC-2/ad is a useful model for developing fusion RET-targeted therapy. (*Cancer Sci* 2013; 104: 896–903)

Lung cancer is the most common cause of cancer death worldwide.[1] The identification of oncogenic driver genes is to select the increasing number of small molecule inhibitors targeting these gene products.[2,3] In particular, in lung adenocarcinoma (LAD), the most dominant histological subtype of lung cancer, the application of kinase inhibitors for cases with specific gene alterations has been successful, that is, gefitinib and erlotinib for *EGFR* mutation-positive cases and crizotinib for *ALK* fusion-positive cases.[4–7] Furthermore, accumulating evidence has demonstrated somatic mutations and rearrangements of potential oncogenes, including *BRAF, ERBB2* and *ROS1*, in LAD.[8–10]

RET is one of the newest LAD driver genes.[11–15] *RET* gene is located on chromosome 10 and encodes a receptor tyrosine kinase,[16,17] and the oncogenic potential of this gene product has been suggested in several tumors, including thyroid cancer.[18–20] Recently, five independent groups identified aberrant fusion genes, *KIF5B-RET* and *CCDC6-RET* in clinical samples of LAD.[11–15] Ectopically expressed RET fusion products afforded NIH3T3 cells with anchorage-independent growth and tumorigenicity in nude mice.[11,14] Furthermore, KIF5B-RET-expressing H1299 cells exhibited growth factor-independent growth.[11] These findings strongly suggest the oncogenic activity of RET fusion products and also suggest the potential therapeutic efficacy of multi-kinase inhibitor targeting of RET using the abovementioned cells. However, LAD-derived cell lines harboring *RET* fusion genes had not been identified. Recently, Matsubara *et al.*[21] screened LAD cell lines that were sensitive to a RET inhibitor vandetanib and found a *CCDC6-RET* fusion gene-harboring cell line, LC-2/ad.

We have independently screened cell lines established from Japanese LAD samples by RT-PCR and found that LC-2/ad cells expressed the *CCDC6-RET* fusion gene product. We further examined whether LC-2/ad cells depend on RET fusion-mediated signaling. In addition, the antitumor effect of RET inhibitors in LC-2/ad cells was evaluated *in vitro* and *in vivo*.

Materials and Methods

Complete materials and methods were described in the supplementary information (Data S1. Materials and Methods).

Purchased materials. Cell lines were purchased from RIKEN Bio Resource Center, the Immuno-Biological Laboratories (Fujioka, Japan) and American Type Culture Collection. Procedures for western blotting was previously described.[22] Primary antibodies specific for RET and phospho-RET Tyr-905 were purchased from Epitomics (Burlingame, CA, USA) and Cell Signaling Technologies (Danvers, MA, USA), respectively. RET-targeting siRNA was purchased from Life Technologies (Carlsbad, CA, USA). Gefitinib, sunitinib malate and sorafenib were purchased from Santa Cruz Biotechnology (Dallas, TX, USA), Sigma-Aldrich (St. Louis, MO, USA) and Toronto Research Chemicals (Toronto, ON, Canada),

[10]To whom correspondence should be addressed.
E-mail: ktsuchih@east.ncc.go.jp
[11]These authors contributed equally to this work.

doi: 10.1111/cas.12175

respectively. Vandetanib, AZD6244 and BEZ235 were purchased from Selleck (Houston, TX, USA).

Multiplex RT-PCR. Reported *KIF5B/CCDC6-RET* fusion variants were detected by multiplex RT-PCR according to the procedures described elsewhere.[11,14]

Genomic DNA sequencing. LC-2/ad DNA was captured with custom hybridization probes targeting *CCDC6* intron 1 and *RET* whole gene (Agilent) followed by parallel sequencing on the MiSeq system (Illumina).

Real-time RT-PCR. Procedures for real-time RT-PCR was previously described.[22] The PCR primers used in the present study are shown in Table S1.

***In vivo* studies.** LC2/ad cells at 5.0×10^6 were subcutaneously inoculated to 8-week-old athymic nude mice (Clea Japan).[23] Vandetanib was administered once daily as a homogeneous suspension by oral gavage at a dosage of 50 mg/kg body weight.[24] The tumor volume was calculated as the product of a scaling factor ($\pi/6$) and the tumor length, width and height.[22] The study was approved by the Institutional Ethics Review Committee for animal experiments at the National Cancer Center.

Immunohistochemical analysis. The procedure for hematoxylin eosin staining and immunohistochemical (IHC) was previously described.[22,25]

Microarray analysis. Background information of clinical samples was described in a previous report.[26] The study was approved by the Institutional Review Boards of the National Cancer Center. Total RNA was analyzed using Affymetrix (Santa Clara, CA, USA) U133Plus2.0 arrays. The data were processed by the MAS5 algorithm, and the mean expression level of a total of 54 675 probes was adjusted to 1000 for each sample.

Results

Identification of the *CCDC6-RET* fusion gene in a Japanese LAD cell line. To identify *RET* fusion-derived mRNA expression in human LAD cell lines, all reported *KIF5B-RET* and *CCDC6-RET* gene products were screened by multiplex RT-PCR in 11 cell lines derived from Japanese patients. LC-2/ad cells were found to express *CCDC6-RET* mRNA at significantly higher levels, whereas the other cell lines did not exhibit any fusion gene products (Fig. 1a). The expressed fusion *RET* product was sequenced, and an in-frame fusion of *CCDC6* exon 1 and *RET* exon 12, which was identical to the previously reported *CCDC6-RET* fusion products, was identified (Fig. 1b).[14] We then identified a breakpoint of chromosome 10 by retrieving genomic DNA fragments, including the entire *RET* gene and intron 1 of *CCDC6*, by target capture system followed by parallel sequencing. The identified break-point between *CCDC6* intron 1 and *RET* exon 11 was confirmed by Sanger sequencing (Fig. 1b). Quantitative RT-PCR revealed that the expression of 3′ end of *RET* was increased comparable to that of *CCDC6*, whereas the transcript level of the 5′ end of *RET* was significantly lower (Fig. 1c). Consistent with the amount of transcript, western blotting using an antibody recognizing the C-terminus of RET isoform 2 detected a 60-kDa specific band equivalent to

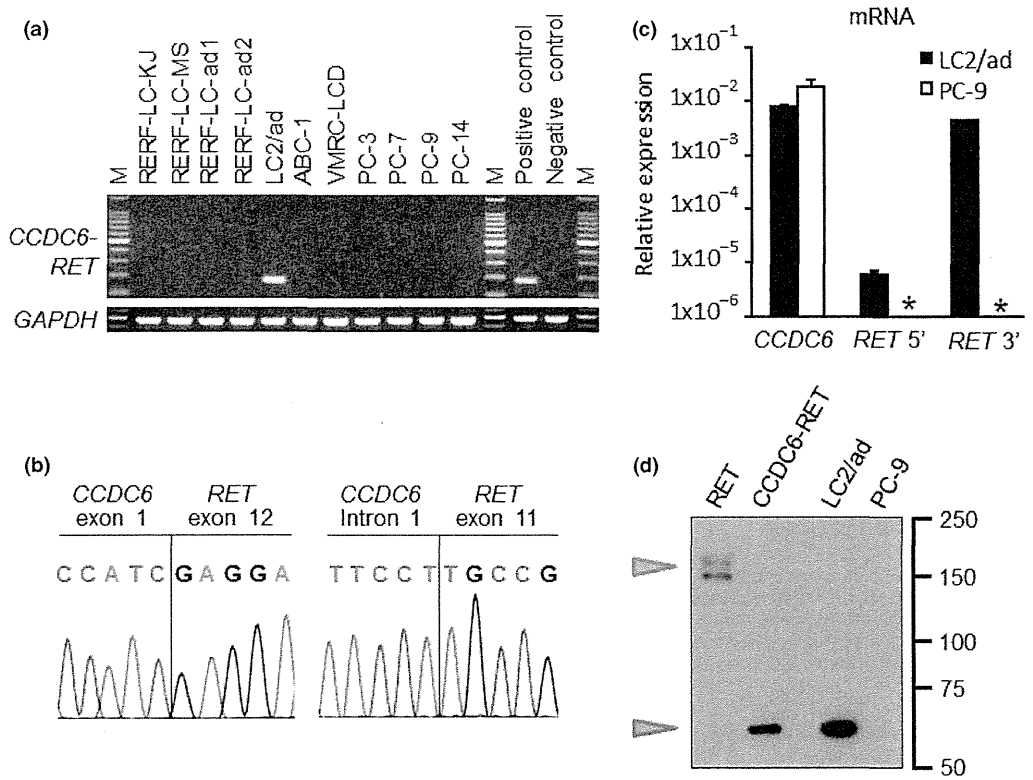

Fig. 1. Identification of the *CCDC6-RET* fusion gene. (a) Detection of *RET* fusion transcripts in lung adenocarcinoma (LAD) cell lines by multiplex reverse transcription-polymerase chain reaction (RT-PCR). (b) Sanger sequencing around the fusion point of the cDNA (left) and the breakpoint of the genomic DNA (right) of *CCDC6-RET* in LC-2/ad cells. (c) 3′ region-specific expression of *RET* mRNA in LC-2/ad cells. The 5′ or 3′ region of *RET* and *CCDC6* cDNA level was normalized to glyceraldehyde 3-phosphate dehydrogenase (*GAPDH*) expression. The data are shown as the mean ± standard deviation (SD) ($n = 3$). Asterisks indicate that mRNA expression were below the level of detection. (d) Specific expression of the CCDC6-RET fusion protein. Whole-cell lysates of LC2/ad and PC-9 cells and HEK293 cells transfected with wild-type RET (RET) or CCDC6-RET expression plasmids were subjected to western blot analysis to detect RET protein isoform 2. The LC-2/ad cells showed an approximately 60-kDa (red arrowhead) but not 170-kDa (blue arrowhead) band.

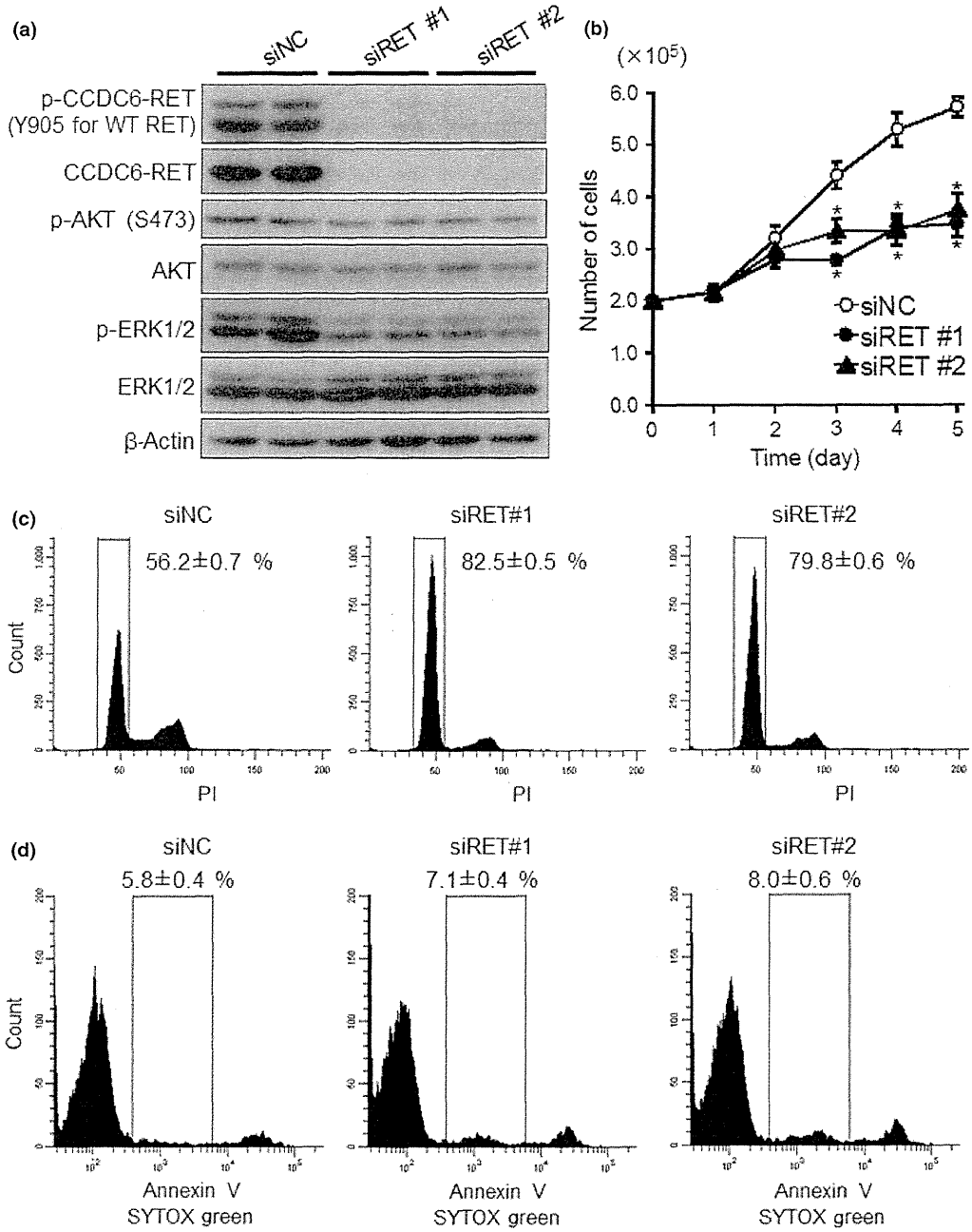

Fig. 2. Suppression of CCDC-RET expression by siRNA in LC-2/ad cells. (a) Western blot analysis of siRET-treated LC-2/ad cells. The siRNA transfected cell lysates were applied to the western blotting. (b) Involvement of RET suppression in cell growth inhibition. LC-2/ad cells transfected with siRNAs were incubated for the indicated times. The data are shown as the mean ± standard deviation (SD) ($n = 4$). *$P < 0.01$ (Student's *t*-test). (c,d) The DNA ploidy (c) and Annexin V-positive population (d) of siRET-transfected LC-2/ad cells. After 72 h of siRNA transfection, the cells were subjected to DNA ploidy analysis and Annexin V staining. The data are shown as the mean ± SD ($n = 4$).

the estimated size of the fusion protein composed of 503 amino acids (GeneBank BAM36435), whereas no significant signal was detected that approximated the size of wild-type RET, 170-kDa (Fig. 1d).[11] Taken together, we concluded that LC-2/ad cells express *CCDC6-RET* fusion gene products. *KRAS* exon 2 and *EGFR* exon 19 and 21 were examined by Sanger sequencing, but no obvious mutation was confirmed (Fig. S1).

CCDC6-RET-dependent ERK1/2 phosphorylation and the proliferation of LC-2/ad cells. We suppressed *RET* expression by RNAi to characterize the function of CCDC6-RET in LC-2/ad cells. For avoiding off-target siRNA effects, two different sequences of siRNA directed against the 3′ region of *RET* (siRET#1 and #2) and a nontargeting siRNA (siNC) were used. When compared to siNC, a significant reduction in mRNA expression was observed by quantitative RT-PCR detecting the 3′ end of the *RET* mRNA: 66.5% for siRET#1 and 94.2% for siRET#2 (Fig. S2). Western blot analyses also revealed significant decreases in the expression of CCDC6-RET protein (60-kDa) upon the introduction of siRET#1 and #2 compared to the control siNC in the LC-2/ad cells

doi: 10.1111/cas.12175

(Fig. 2a). To examine whether the downstream signaling pathway was altered by the introduction of siRNA, the phosphorylation of ERK1/2 and AKT was examined. The phospho-ERK1/2 signal was significantly decreased by the suppression of CCDC6-RET expression, whereas the decrease of AKT phosphorylation was marginal (Fig. 2a). The involvement of RET fusion in LC-2/ad cell proliferation was then examined. The number of live CCDC6-RET-suppressed cells decreased throughout the experiment, and the difference became significant at day 3 and thereafter (Fig. 2b). To address the growth suppression further, the cell cycle of the siRNA-treated cells were assessed by the DNA ploidy pattern. The LC-2/ad cells treated with siRET exhibited significant increases in the percent of cells arrested in the G1 phase relative to the cells treated with siNC (Fig. 2c). However, the apoptotic cells, as assessed by Annexin V positivity, was not significantly increased by the suppression of RET expression (Fig. 2d).

RET-dependent transcriptome profile in LC-2/ad cell. To characterize the transcriptome profile, which is regulated by CCDC6-RET and its downstream signaling pathway, siRET#2 and siNC treated LC-2/ad cells were subjected to genome-wide expression profiling using Affymetrix U133Plus2.0 arrays. A total of 243 genes, evaluated with 285 probes were selected as those preferentially suppressed by less than half in siRET-treated cells.As well, 566 genes with 661 probes were expressed more than twice in siRET-treated cells (Table S2 and Fig. S3). The *RET* gene itself (probe ID = 211421_s_at) showed the highest fold-difference of 19.6 between siNC- and siRET#2-treated cells. Following *RET*, previously identified Gene Ontology-annotated Ras-MAPK downstream genes like *DUSP6* was preferentially suppressed in the siRET-treated cells. In addition, cell cycle regulation-related genes like *EREG, CDC6, MCM10, MAD2L1, CHEK1* and *PLK4* were expressed <0.5-fold in siRET-treated cells (Table 1).

RET fusion gene screening of 300 consecutive surgically resected LAD samples identified one case of *CCDC6-RET* expressing LAD by RT-PCR and break-apart FISH (Tsuta *et al.*, 2012, unpublished data). We checked the expression level of potential CCDC6-RET-driven genes identified above in the clinical sample. Among 285 preferentially expressed probes, 81 probes were also upregulated more than twofold in the *CCDC6-RET* positive LAD tissue compared to the surrounding non-cancerous tissue (Table 1 and Table S2).

RET inhibitor-induced cell cycle arrest and apoptosis in LC-2/ad cells. The phosphorylation status of the tyrosine 905 residue of RET isoforms 2 and 4 was high in the LC-2/ad cells, regardless of the presence or absence of serum in the culture medium, whereas the total amount of RET isoform 2 was not significantly altered. Similarly, the phosphorylation status of AKT and ERK1/2 was high under serum-starved conditions, and the enhanced phosphorylation of these molecules was slight with serum stimulation, suggesting that the fusion RET kinase was constitutively active and activated its downstream signaling pathways (Fig. 3a).

Next, the effects of kinase inhibitors, which inhibit spectrum including RET were applied to evaluate their effects on the signaling pathways in the LC-2/ad cells. We treated the cells with RET inhibitors vandetanib, sunitinib and sorafenib at a final concentration of 10 μM, which was 10–30 times higher than the *in vitro* half maximal inhibitory concentration (IC_{50}) for RET kinase activity of each compound. Gefitinib, another small molecule inhibitor targeting EGFR but not RET,[13] was also examined. All the inhibitors except gefitinib significantly suppressed the phosphorylation of RET, AKT and ERK1/2. Although vandetanib, sunitinib and sorafenib equivalently suppressed RET phosphorylation, vandetanib most significantly suppressed the phosphorylation of ERK1/2 (Fig. 3a). The inhibitory effect of vandetanib on RET, AKT and ERK1/2 phosphorylation exhibited concentration dependency (Fig. 3b). Gefitinib significantly suppressed EGFR phosphorylation while total EGFR protein level was not altered. Meanwhile, gefitinib did not alter the phosphorylation status of AKT and ERK1/2 (Fig. 3a). Meanwhile, vandetanib suppressed EGFR as well as AKT and ERK1/2 in *EGFR*-mutant PC-9 cells (Fig. S4).

We further examined the effect of the above inhibitors on the growth of the LC-2/ad cells using the WST-8 assay. Consistent with the effects of the inhibitors on the RET signaling pathway, vandetanib suppressed cell growth most significantly (IC_{50} = 0.32 μM), followed by sunitinib and sorafenib, whereas gefitinib only exhibited an apparent suppression at its highest dose (Fig. 3c). However, the effects of these inhibitors on *KRAS*-mutant A549 cells were much lower (Fig. S5). Gefitinib and vandetanib, both of which inhibit EGFR, suppressed *EGFR*-mutant PC-9 cells, whereas sunitinib and sorafenib had less effect (Fig. S5). Evaluating the number of live cells by trypan blue staining under the treatment of several doses of

Table 1. Up- or downregulated genes associated with mitogen-activated protein kinase (MAPK) cascade or cell cycle

Gene symbol	Probe set ID	siNC/siRET	Tumor/Non-tumor
Upregulated			
RET	211421_s_at	19.63	19.52
	205879_x_at	3.76	5.03
	215771_x_at	2.37	4.72
DUSP6	208892_s_at	4.45	5.22
	208893_s_at	4.17	6.34
	208891_at	4.17	3.56
EREG	1569583_at	3.68	1.60
	205767_at	2.93	5.69
CDC6	203967_at	2.42	4.82
	203968_s_at	1.95	5.32
MCM10	220651_s_at	2.30	4.83
	223570_at	1.72	1.71
MAD2L1	203362_s_at	2.28	5.91
	1554768_a_at	1.91	4.34
CHEK1	205394_at	2.17	9.03
	205393_s_at	2.14	6.87
PLK4	204886_at	2.07	4.38
	204887_s_at	1.56	4.08
Downregulated			
MEF2C	209200_at	0.21	0.46
	209199_s_at	0.26	0.65
GAB1	214987_at	0.23	0.42
	229114_at	0.53	0.65
	225998_at	0.62	0.68
	226002_at	0.64	0.76
CDKN1C	216894_x_at	0.26	0.41
	213348_at	0.32	0.23
	213183_s_at	0.35	0.30
	219534_x_at	0.42	0.27
	213182_x_at	0.44	0.21
PTEN	233314_at	0.33	0.27
	225363_at	0.77	0.47
TIMP2	231579_s_at	0.34	0.33
	224560_at	0.37	0.27
ID2	201566_x_at	0.35	0.31
	201565_s_at	0.40	0.39
	213931_at	0.52	0.31
CCNL2	232274_at	0.35	0.42
	222999_s_at	0.79	0.52
RPS6KA2	212912_at	0.41	0.34
	204906_at	0.59	0.49

vandetanib suggested a dose-dependent suppression in the LC-2/ad cells. Furthermore, the number of cells treated with 0.5 and 1.0 μM vandetanib was apparently reduced to less than the starting amount, strongly suggesting that vandetanib induced both cell death and the suppression of cell proliferation (Fig. 3d). An assessment of the DNA ploidy revealed that vandetanib arrested the cell cycle in G1 phase in a dose-dependent manner (Fig. 3e), and an increased concentration of vandetanib induced an Annexin V-positive apoptotic cell population (Fig. 3f). The proapoptotic effect of vandetanib was confirmed by the detection of cleaved caspase-3 by western blotting (Fig. 3b). Meanwhile, 1.0 μM sunitinib and sorafenib induced cell cycle arrest but induction of apoptosis was marginal (Figs S6 and S7).

To further evaluate the contribution of Ras-ERK and AKT axes to cell survival, LC-2/ad cells were treated with MEK1/2 inhibitor AZD6244 or PI3K/mTOR inhibitor BEZ235. Cytotoxic effect of AKT-inhibiting BEZ235 was more than that of ERK-inhibiting AZD6244. However, both inhibitors did not completely reduce the cell survival even their maximal dose (Figs S8 and S9).

Anti-tumor effect of vandetanib in an LC-2/ad xenograft model. Subcutaneously transplanted LC-2/ad tumors exhibited typical adenocarcinoma morphology. These tumors were positive for SFTPA, Napsin A and carcinoembryonic antigen (CEA) but thyroid marker thyroglobulin negative using immunohistochemistry (IHC). Furthermore, using an antibody cross-reacting with both human and mouse RET protein, IHC revealed that RET was highly expressed specifically in the tumor cells but not in the interstitial cells (Fig. 4a). The overexpression of RET in these tumors was confirmed using quantitative RT-PCR and Western blotting. Similar to the results from cultured LC-2/ad cells, much more mRNA of the 3′ end of *RET* was detected than that of the 5′ end (Fig. 4b), and a specific band equivalent to the size of the CCDC6-RET fusion protein was detected (Fig. 4c). Vandetanib (50 mg/kg) was orally administrated to the mice harboring the LC-2/ad xenograft, and the daily administration of vandetanib significantly reduced the tumor size. Although the tumors were diminished at day 14 of the treatment, the body weight of the treated mice was not significantly reduced (Fig. 4d and Fig. S10). Sorafenib (30 mg/kg) and sunitinib (40 mg/kg) did not reduce the body weight, either (Fig. S10). Sorafenib reduced but not diminished the tumors at day14. Anti-tumor effect of sunitinib was not significant (Fig. S11).

Discussion

Previous reports suggest that the incidence of *RET*-fusion-positive cases in LAD is 1–2% and that these cases are concentrated in the *EGFR* mutation-, *KRAS* mutation-, and *ALK*-fusion-negative population.[10,27] To identify cell lines expressing endogenous *RET*-fusion genes, we selected 11 cell lines that were derived from pathologically identified Japanese LAD cases. Among them, activating EGFR mutations have been reported in PC-3 and PC-9 cells.[28] However, the mutation status of known driver genes of other cell lines was not well investigated. The LC-2/ad cells were originally derived from pleural effusion of LAD in a patient who had received combined chemotherapy (endoxan, Adriamycin, Cisplatin and mitomycin C)[23]; the cancer was diagnosed by cytological examination of the patient's sputum and pleural effusion. The original report indicated that the LC-2/ad cells were positive for an adenocarcinoma marker, cytokeratin 18.[23] In addition, we detected surfactant protein, an aspartate proteinase, Napsin A, and CEA expression in the xenograft tumor (Fig. 4a). These findings support the origin of LC-2/ad as lung adenocarcinoma. The modal chromosome number described in the original report was 53–56, though an apparent translocation between the chromosomes was not reported, consistent with the fact that the inversion of chromosome 10 was not obvious in the conventional chromosome counts.

The Sanger sequencing in this study and the whole-transcriptome sequencing (Tsuchihara, 2012, unpublished data) revealed no driver mutations of *KRAS*, *EGFR* and known genes other than the *CCDC6-RET* fusion in the LC-2/ad cells, highly suggesting that the CCDC6-RET fusion protein plays pivotal roles in the proliferation of these cells. The autophosphorylation of CCDC6-RET was clearly observed in a serum-independent manner, accompanied with a constitutive elevation of ERK1/2 phosphorylation. The suppression of CCDC6-RET expression induced a decrease in ERK1/2 phosphorylation, accompanied with a decrease in the expression of the genes that regulate the cell cycle. As a result, the CCDC6-RET-suppressed cells exhibited significant growth retardation.

Recently, a Japanese group independently reported the CCDC6-RET fusion in LC2/ad cells.[21] However, the efficacy of RET inhibitors to the RET and downstream pathways and *in vivo* anti-tumor effects have been partially described.[21] Vandetanib, sorafenib and sunitinib suppress the activities of multiple kinases, including RET, and have been approved for several cancers.[29–31] In *in vitro* analyses, these compounds effectively suppressed the phosphorylation of CCDC6-RET and suppressed proliferation and induced death in LC-2/ad cells. It should be noted that the IC_{50} value for the growth suppression of these compounds was equivalent to the dose suggested in a previous study using culture cells expressing ectopic *KIF5B-RET* cDNA.[13] These effects were most likely dependent on RET inhibition. Sunitinib and sorafenib did not affect PC-9 and A549 cells, which have activating mutations of *EGFR* and *KRAS*, respectively. Vandetanib presumably suppressed the growth of PC-9 cells, as EGFR is included in its inhibitory spectrum. Meanwhile, gefitinib, which targets EGFR but not RET, did not significantly suppress the growth of LC-2/ad cells. Interestingly, gefitinib did not alter the phosphorylation of AKT and ERK1/2 in LC-2/ad cells albeit equivalently suppressing EGFR phosphorylation as vandetanib. Although precise molecular mechanisms should be further examined, LC-2/ad cells might not depend on EGFR for transducing downstream signaling.

Vandetanib exhibited apparent anti-tumor effects in the xenograft model in this study. Recently, efficacy of vandetanib on thyroid cancer cells harboring *RET*-fusion gene was also reported.[32] These findings strongly suggest that RET inhibition is a plausible therapeutic strategy for RET-fusion-positive tumors.

We noticed a discrepancy between the effects of RNA interference and inhibitor treatment on RET. Though RET suppression/inhibition equivalently reduced the level of phosphorylated RET and induced cell cycle arrest, obvious apoptosis was not found in the cells treated with siRNA. A possible explanation is that CCDC6-RET is mainly involved in the RAS-ERK pathway to regulate cell proliferation, whereas the anti-apoptotic signaling pathway mediated by AKT could be regulated by other signaling molecules inhibited by the multi-kinase inhibitors. A recent study using a Drosophila *in vivo* screening system suggested that the antitumor effects and toxicity of RET inhibitors were dependent on the profile of the "off-target" inhibition of multiple kinases in addition to the specific inhibition of RET.[33] Further investigation elucidating the molecules and signaling pathways relevant to the cytotoxic effect of vandetanib in LC-2/ad cells is anticipated.

Whether LC-2/ad-based models adequately represent clinical *RET* fusion-positive LAD cases is another challenging question. Takeuchi stated that clinically identified *CCDC6-RET*-positive LAD exhibited a histologically cribriform pattern.[14]

doi: 10.1111/cas.12175

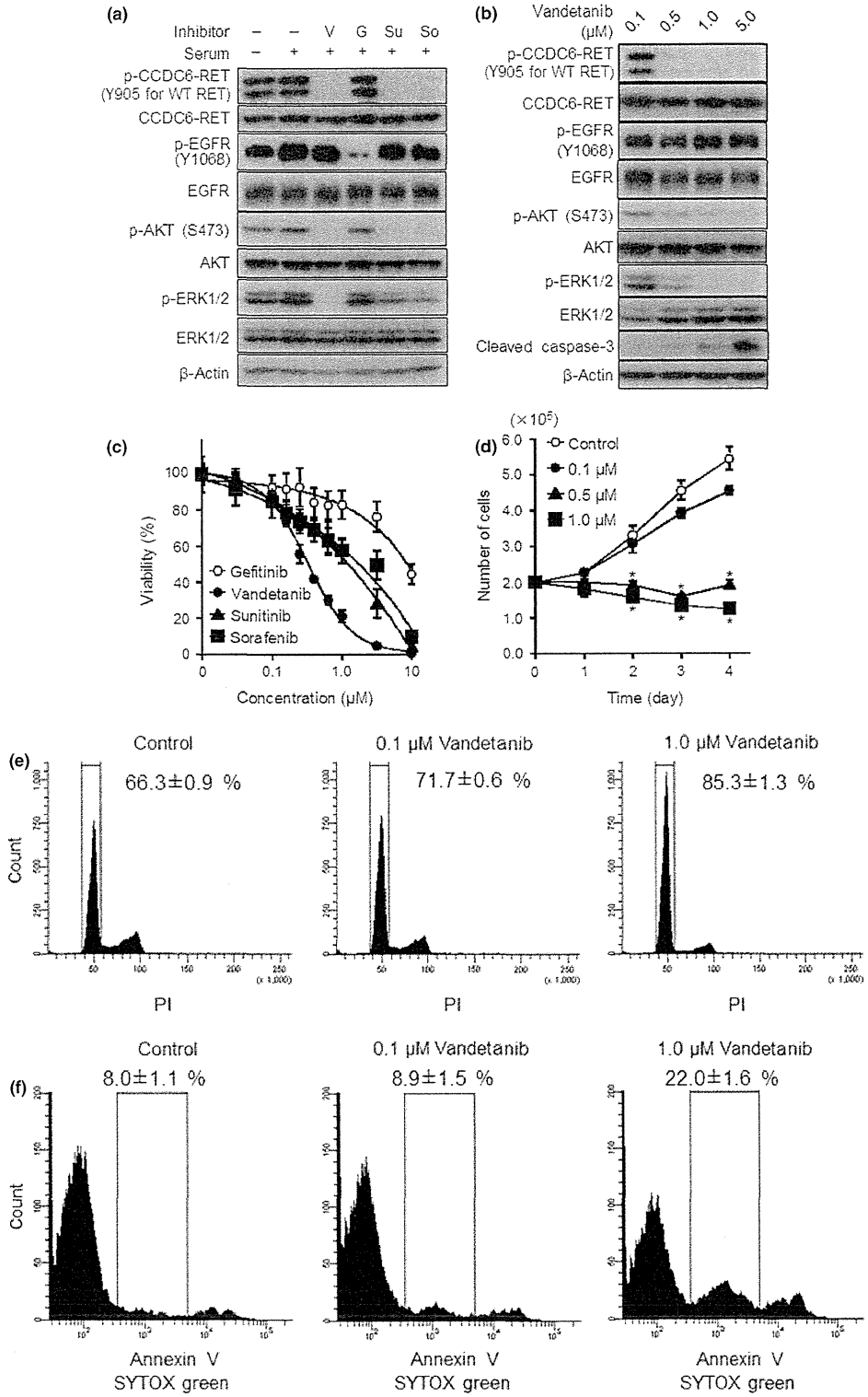

Fig. 3. Effect of RET inhibitors on LC-2/ad cells. (a) Western blot analysis of inhibitor-treated cells. The cells were incubated under serum-starved conditions for 22 h and treated with 1 μM of inhibitor or dimethylsulfoxide (DMSO) for 2 h. Prior to cell lysis, the cells were treated with 10% fetal bovine serum (FBS) for 10 min. Whole-cell lysates were subjected to western blot analysis to detect the indicated proteins. G, gefitinib; So, sorafenib; Su, sunitinib; V, vandetanib. (b) Dose-dependent effect of vandetanib. Cells were treated with the indicated concentration of vandetanib for 12 h, and western blotting was used to detect the indicated proteins. (c) WST-8 assay with kinase inhibitors. Cells were treated with the indicated inhibitors for 72 h, and the viability was assessed using the WST-8 assay. The data are shown as the mean ± standard deviation (SD) (n = 6). (d) Effect of vandetanib for growth inhibition. Cells were treated with vandetanib and incubated for the indicated time. The data are shown as the mean ± SD (n = 3). *P < 0.01 (Student's t test). (e,f) DNA ploidy (e) and Annexin V-positive population (f) of the cells treated with vandetanib for 48 h. The data are shown as the mean ± SD (n = 4).

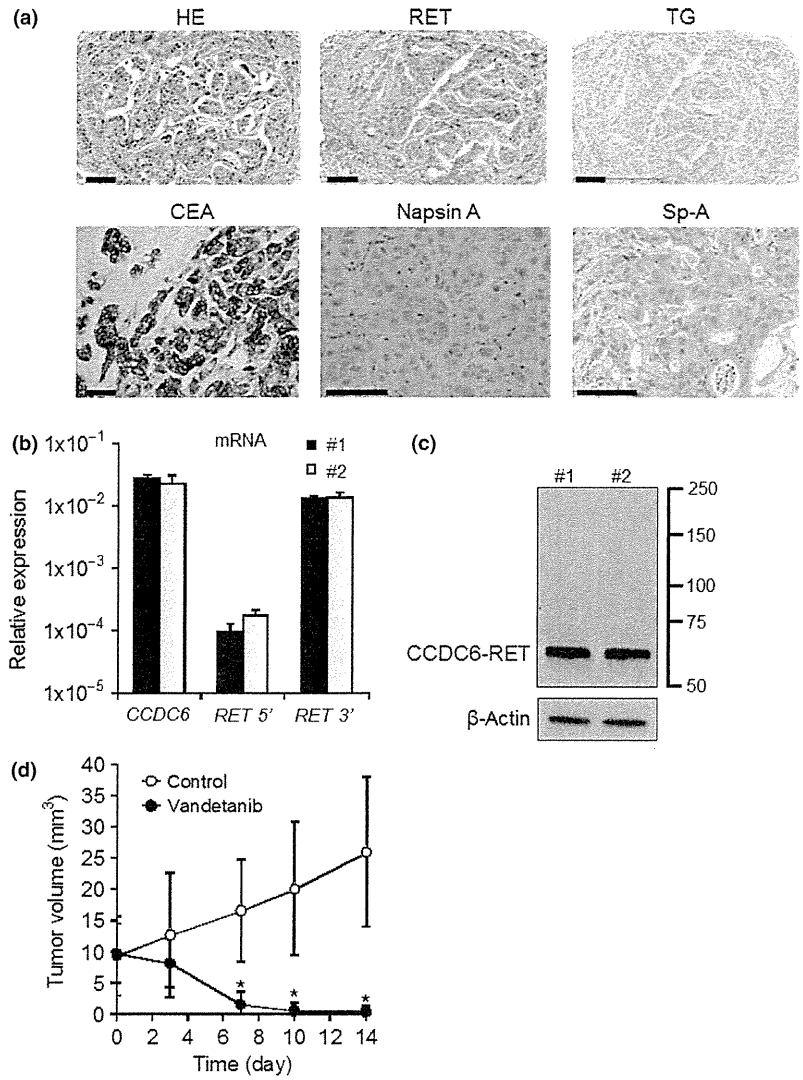

Fig. 4. Characterization of the LC-2/ad xenograft and anti-tumor effects of vandetanib. (a) Histological features of the xenograft. Hematoxylin and eosin staining and immunohistochemical staining with the indicated antibodies. Scale bars were 100 μm. Hematoxylin eosin (HE), RET, thyroglobulin (TG) and carcinoembryonic antigen (CEA) (×20); Napsin A and Sp-A (×40). (b) 3′ region-specific expression of *RET* mRNA in the xenograft. Total RNA extracted from tumors was subjected to real-time reverse transcription-polymerase chain reaction (RT-PCR) analysis with the primer sets designed for the 5′ or 3′ region of the *RET* and *CCDC6* cDNA. The data are shown as the mean ± standard deviation (SD) ($n = 3$). (c) Expression of the CCDC6-RET protein in mice xenografts. Whole-cell lysates of tumors were subjected to western blot analysis. (d) Anti-tumor effect of vandetanib *in vivo*. Vandetanib was administered once a day at a dosage of 50 mg/kg. The data are shown as the mean ± SD ($n = 9$). *$P < 0.01$ (control *vs* sorafenib; Student's *t* test).

Because the cribriform structure was presumably developed from normal alveolar architecture, this specific morphology was not observed in the subcutaneously transplanted LC-2/ad tumors. We assume that the comparison of the transcriptome profile between the LC-2/ad cells and clinically identified LAD tissue samples may provide clues. Approximately one-third of the genes suppressed by RNA interference directed at *RET* overlapped with the genes preferentially expressed in the clinical tumor sample. Because we have had only one example of paired data, it is difficult to estimate the similarity between the cell line and clinical samples. However, the above overlap appears promising, and we will continue to screen both cell lines and clinical samples to accumulate comprehensive data.

In this study, the screening of Japanese LAD cell lines was effective for the identification of *RET* fusion-positive cancer cells, representing a clinically rare subpopulation. LC-2/ad cells might be useful in the development of RET-targeted therapies, that is, new compound screening, clarifying the pharmacological mechanisms and investigating the mechanisms for acquired resistance.

Acknowledgments

We thank Drs Hiroki Sasaki and Kazuhiko Aoyagi and the National Cancer Center Research Core Facility for the microarray analyses. The Core Facility was supported by National Cancer Center Research and Development Fund (23-A-7). This work was supported by National Cancer Center Research and Development Fund (23-A-8, 15, 24-A-1) and JSPS KAKENHI Grant number 24300345.

Disclosure Statement

The authors have no conflict of interest.

References

1 Jemal A, Bray F, Center MM, Ferlay J, Ward E, Forman D. Global cancer statistics. *CA Cancer J Clin* 2011; **61**: 69–90.

2 Pao W, Girard N. New driver mutations in non-small-cell lung cancer. *Lancet Oncol* 2011; **12**: 175–80.

3 Pao W, Chmielecki J. Rational, biologically based treatment of EGFR-mutant non-small-cell lung cancer. *Nat Rev Cancer* 2010; **10**: 760–74.

doi: 10.1111/cas.12175

4 Kwak EL, Bang YJ, Camidge DR *et al.* Anaplastic lymphoma kinase inhibition in non-small-cell lung cancer. *N Engl J Med* 2010; **363**: 1693–703.
5 Lynch TJ, Bell DW, Sordella R *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N Engl J Med* 2004; **350**: 2129–39.
6 Paez JG, Janne PA, Lee JC *et al.* EGFR mutations in lung cancer: correlation with clinical response to gefitinib therapy. *Science* 2004; **304**: 1497–500.
7 Tsao MS, Sakurada A, Cutz JC *et al.* Erlotinib in lung cancer – molecular and clinical predictors of outcome. *N Engl J Med* 2005; **353**: 133–44.
8 Govindan R, Ding L, Griffith M *et al.* Genomic landscape of non-small cell lung cancer in smokers and never-smokers. *Cell* 2012; **150**: 1121–34.
9 Imielinski M, Berger AH, Hammerman PS *et al.* Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012; **150**: 1107–20.
10 Seo JS, Ju YS, Lee WC *et al.* The transcriptional landscape and mutational profile of lung adenocarcinoma. *Genome Res* 2012; **22**: 2109–19.
11 Kohno T, Ichikawa H, Totoki Y *et al.* KIF5B-RET fusions in lung adenocarcinoma. *Nat Med* 2012; **18**: 375–7.
12 Li F, Feng Y, Fang R *et al.* Identification of *RET* gene fusion by exon array analyses in "pan-negative" lung cancer from never smokers. *Cell Res* 2012; **22**: 928–31.
13 Lipson D, Capelletti M, Yelensky R *et al.* Identification of new *ALK* and *RET* gene fusions from colorectal and lung cancer biopsies. *Nat Med* 2012; **18**: 382–4.
14 Takeuchi K, Soda M, Togashi Y *et al.* RET, ROS1 and ALK fusions in lung cancer. *Nat Med* 2012; **18**: 378–81.
15 Wang R, Hu H, Pan Y *et al.* RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. *J Clin Oncol* 2012; **30**: 4352–9.
16 Gardner E, Papi L, Easton DF *et al.* Genetic linkage studies map the multiple endocrine neoplasia type 2 loci to a small interval on chromosome 10q11.2. *Hum Mol Genet* 1993; **2**: 241–6.
17 Mole SE, Mulligan LM, Healey CS, Ponder BA, Tunnacliffe A. Localisation of the gene for multiple endocrine neoplasia type 2A to a 480 kb region in chromosome band 10q11.2. *Hum Mol Genet* 1993; **2**: 247–52.
18 Mulligan LM, Kwok JB, Healey CS *et al.* Germ-line mutations of the RET proto-oncogene in multiple endocrine neoplasia type 2A. *Nature* 1993; **363**: 458–60.
19 Hofstra RM, Landsvater RM, Ceccherini I *et al.* A mutation in the RET proto-oncogene associated with multiple endocrine neoplasia type 2B and sporadic medullary thyroid carcinoma. *Nature* 1994; **367**: 375–6.
20 Grieco M, Cerrato A, Santoro M, Fusco A, Melillo RM, Vecchio G. Cloning and characterization of H4 (D10S170), a gene involved in RET rearrangements *in vivo*. *Oncogene* 1994; **9**: 2531–5.
21 Matsubara D, Kanai Y, Ishikawa S *et al.* Identification of CCDC6-RET fusion in the human lung adenocarcinoma cell line, LC-2/ad. *J Thorac Oncol* 2012; **7**: 1872–6.
22 Makinoshima H, Ishii G, Kojima M *et al.* PTPRZ1 regulates calmodulin phosphorylation and tumor progression in small-cell lung carcinoma. *BMC Cancer* 2012; **12**: 537.
23 Kataoka H, Itoh H, Seguchi K, Koono M. Establishment and characterization of a human lung adenocarcinoma cell line (LC-2/ad) producing alpha 1-antitrypsin *in vitro*. *Acta Pathol Jpn* 1993; **43**: 566–73.
24 Wedge SR, Ogilvie DJ, Dukes M *et al.* ZD6474 inhibits vascular endothelial growth factor signaling, angiogenesis, and tumor growth following oral administration. *Cancer Res* 2002; **62**: 4645–55.
25 Yoshida A, Tsuta K, Nakamura H *et al.* Comprehensive histologic analysis of ALK-rearranged lung carcinomas. *Am J Surg Pathol* 2011; **35**: 1226–34.
26 Okayama H, Kohno T, Ishii Y *et al.* Identification of genes upregulated in ALK-positive and EGFR/KRAS/ALK-negative lung adenocarcinomas. *Cancer Res* 2012; **72**: 100–11.
27 Ju YS, Lee WC, Shin JY *et al.* A transforming *KIF5B* and *RET* gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res* 2012; **22**: 436–45.
28 Blanco R, Iwakawa R, Tang M *et al.* A gene-alteration profile of human lung cancer cell lines. *Hum Mutat* 2009; **30**: 1199–206.
29 Phay JE, Shah MH. Targeting RET receptor tyrosine kinase activation in cancer. *Clin Cancer Res* 2010; **16**: 5936–41.
30 Antonelli A, Fallahi P, Ferrari SM *et al.* RET TKI: potential role in thyroid cancers. *Curr Oncol Rep* 2012; **14**: 97–104.
31 Scagliotti GV. Potential role of multi-targeted tyrosine kinase inhibitors in non-small-cell lung cancer. *Ann Oncol* 2007; **18**(Suppl 10): x32–41.
32 Couto JP, Almeida A, Daly L *et al.* AZD1480 blocks growth and tumorigenesis of RET- activated thyroid cancer cell lines. *PLoS ONE* 2012; **7**: e46869.
33 Dar AC, Das TK, Shokat KM, Cagan RL. Chemical genetic discovery of targets and anti-targets for cancer polypharmacology. *Nature* 2012; **486**: 80–4.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Data S1. Materials and methods.

Fig. S1. The absence of the known driver mutations.

Fig. S2. Suppression of *RET* mRNA in siRET-treated cells.

Fig. S3. RET-dependent transcriptome profile in LC-2/ad cells.

Fig. S4. Dose-dependent effect of vandetanib in PC-9 cells.

Fig. S5. WST-8 assay with various kinase inhibitors.

Fig. S6. Effect of sunitinib and sorafenib on G1 phase population of LC-2/ad cells.

Fig. S7. Effect of sunitinib and sorafenib on apoptosis of LC-2/ad cells.

Fig. S8. Dose-dependent effect of AZD6244 and BEZ235 in LC-2/ad cells.

Fig. S9. WST-8 assay of LC-2/ad cells treated with AZD6244 and BEZ235.

Fig. S10. Body weight of the vandetanib-, sunitinib-, sorafenib- and vehicle-treated mice.

Fig. S11. Effect of sunitinib and sorafenib *in vivo*.

Table S1. Polymerase chain reaction primers.

Table S2. Summary of the microarray data.

Identification and Characterization of Cancer Mutations in Japanese Lung Adenocarcinoma without Sequencing of Normal Tissue Counterparts

Ayako Suzuki[1☯], Sachiyo Mimaki[2☯], Yuki Yamane[3], Akikazu Kawase[3], Koutatsu Matsushima[2], Makito Suzuki[2], Koichi Goto[3], Sumio Sugano[1], Hiroyasu Esumi[2], Yutaka Suzuki[1*], Katsuya Tsuchihara[2]

1 Department of Medical Genome Sciences, Graduate School of Frontier Sciences, The University of Tokyo, Chiba, Japan, **2** Division of Translational Research, Research Center for Innovative Oncology, National Cancer Center Hospital East, Chiba, Japan, **3** Thoracic Oncology Division, National Cancer Center Hospital East, Chiba, Japan

Abstract

We analyzed whole-exome sequencing data from 97 Japanese lung adenocarcinoma patients and identified several putative cancer-related genes and pathways. Particularly, we observed that cancer-related mutation patterns were significantly different between different ethnic groups. As previously reported, mutations in the EGFR gene were characteristic to Japanese, while those in the KRAS gene were more frequent in Caucasians. Furthermore, during the course of this analysis, we found that cancer-specific somatic mutations can be detected without sequencing normal tissue counterparts. 64% of the germline variants could be excluded using a total of 217 external Japanese exome datasets. We also show that a similar approach may be used for other three ethnic groups, although the discriminative power depends on the ethnic group. We demonstrate that the ATM gene and the PAPPA2 gene could be identified as cancer prognosis related genes. By bypassing the sequencing of normal tissue counterparts, this approach provides a useful means of not only reducing the time and cost of sequencing but also analyzing archive samples, for which normal tissue counterparts are not available.

Citation: Suzuki A, Mimaki S, Yamane Y, Kawase A, Matsushima K, et al. (2013) Identification and Characterization of Cancer Mutations in Japanese Lung Adenocarcinoma without Sequencing of Normal Tissue Counterparts. PLoS ONE 8(9): e73484. doi:10.1371/journal.pone.0073484

Editor: H. Sunny Sun, Institute of Molecular Medicine, Taiwan

Received March 22, 2013; **Accepted** July 19, 2013; **Published** September 12, 2013

Funding: This work was supported by JSPS KAKENHI Grant number 24300345. This work was also supported by MEXT KAKENHI Grant Number 221S0002. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

* E-mail: ysuzuki@hgc.jp

☯ These authors contributed equally to this work.

Introduction

The advent of next generation sequencing technology has greatly facilitated the detection and characterization of genetic variations in the human genome. Most remarkably, this type of study has driven the 1000 Genomes Project [1,2], which aims to provide a comprehensive map of human genetic variants across various ethnic backgrounds. However, because whole-genome sequencing is still costly, the sequencing of whole exon regions using hybridization capture methods (exome sequencing) [3-5] is widely used to screen for genes that are related to hereditary diseases. By sequencing exomes from healthy and diseased individuals and comparing them, genes that are responsible for many diseases have been identified [6], including Miller syndrome [7,8] and familial hyperkalemic hypertension [9]. Along with the progress that has been made in exome sequencing, the volume of germline single nucleotide polymorphism (SNP) data that has been registered in dbSNP is rapidly expanding for various populations [10].

Exome sequencing provides a powerful tool for cancer studies as well. Indeed, a number of papers have been published describing the identification and characterization of single nucleotide variants (SNVs) that somatically occur in cancers and are suspected to be responsible for carcinogenesis and disease development [11]. The International Cancer Genome Consortium (ICGC) has been collecting exome data for somatic SNVs that are present in more than 50 types of cancers as a part of an international collaborative effort [12-14]. The Cancer Genome Atlas (TCGA) has developed a large genomic dataset, including exomes for high-grade ovarian carcinoma, that has been used to detect significantly mutated genes, including TP53, BRCA1 and BRCA2 [15]. They have also identified various genomic aberrations and deregulated pathways that may act as therapeutic targets.

In most ongoing cancer exome studies, normal tissue counterparts have been sequenced in parallel with cancer tissue [15-19]. This is assumed to be necessary because germline variants must be excluded from the full set of SNVs to detect the somatic SNVs that are unique to cancers. However, the sequencing of normal tissue counterparts increases the cost and time of the analysis. Also, in some cases, it is difficult to obtain normal tissue counterparts. In addition, it remains unclear how accurately germline SNVs can be excluded using normal tissue exomes. To conservatively exclude germline SNVs, their sequence depths and accuracies may need to be greater than those that are obtained from the cancer exomes.

In this study, we generated and analyzed 97 cancer exomes from Japanese lung adenocarcinoma patients. We also demonstrate that somatic SNVs can be enriched to a level that is sufficient for further statistical analyses even in the absence of the sequencing of normal tissue counterparts. To separate the germline from the somatic SNVs, we first compared the variation patterns between a cancer exome with the 96 other patients' normal tissue exomes. We also attempted to conduct a similar mutual comparison solely utilizing cancer exomes, without the consideration of exomes of normal tissue counterparts. It is true that if we completely omitted normal tissue sequencing, we would tentatively disregard of somatic mutations that occurs at exactly the same genomic position in multiple cancers. However, recent papers have elucidated that such shared SNVs are very rare [15,20-22]. Moreover, many of these recursively mutations have been registered in the cancer somatic mutation databases such as Sanger COSMIC [23,24], and those recurrent SNVs can be recovered by follow-up studies partially using the data from the normal tissues. To understand the unique nature of each cancer, a statistical analysis of the distinct SNVs is presumed to be essential in addition to the analysis of the common SNVs.

In this study, we demonstrate that it is possible to identify the first candidates for cancer-related genes and pathways, even without the sequencing of a normal tissue counterpart. We show that this approach is useful not only to reduce the cost of the sequencing but also to improve the fidelity of the data. It should be also useful for analyzing old archive samples, for which normal tissue counterparts are not always available. Here, we describe a practical and cost-effective method to expedite cancer exome sequencing.

Results and Discussion

Characterization of SNVs using the 97 exome dataset

Firstly, we generated and analyzed whole-exome sequences from 97 Japanese lung adenocarcinoma patients. Exome data were collected from both cancer and normal-tissue counterparts, separated by laser capture microdissection. We purified the exonic DNA (exomes) and generated 76-base paired-end reads using the illumina GAIIx platform. Approximately 30 million mapped sequences were obtained from each sample, providing 74× coverage of the target regions; 93% of the target regions had 5× coverage (Figure S1 in File **S1**). Burrows-Wheeler Aligner (BWA) [25] and the Genome Analysis Toolkit (GATK) [26,27] were used to identify SNVs (Figure S2 in File **S1**). Only SNVs that were detected in cancer tissues and showed no evidence of variation in normal tissues were selected for further analysis.

The obtained dataset was used to characterize the cancer-specific mutation patterns (Table S3 in File **S1**). We calculated the enrichment of the SNVs within particular genes, protein domains, functional categories, and pathways. We searched for genes with somatic SNVs significantly enriched in Japanese lung adenocarcinoma. As shown in Table S4 in File **S1**, several genes were identified as significantly mutated. In particular, we searched for domains that are enriched with SNVs and harbor known cancer-related mutations in the COSMIC database. In total, 11 genes were identified ($P < 0.02$, Table **1**). For example, the Dbl homology (DH) domain of PREX1 gene [28] was enriched with SNVs ($P = 0.00071$). However, in the PREX2 gene [29], the Pleckstrin homology (PH) domain was enriched with SNVs ($P = 0.011$) (Figure 1A and B). Both the PREX1 and the PREX2 genes activate the exchange of GDP to GTP for the Rho family of GTPases and the DH/PH domains are indispensable for nucleotide exchange of GTPases and its regulation [30-32]. In addition, we analyzed the expression patterns of these genes using a cancer gene expression database, GeneLogic (Figure S3 in File **S1**). Expression levels of PREX1 and PREX2 were not enhanced in lung adenocarcinoma but were enhanced in wide variety of cancers, which is partly indicated in previous studies [33]. The SNVs in the PREX1 and PREX2 genes, which were concentrated at its pivotal signaling domains, might enhance activities in these genes, and thereby functionally mimics the increased expressions of this gene in some different types of cancers. The cancer-related gene candidates identified from this dataset are listed in Table **1**.

Similarly, pathway enrichment analyses using the KEGG database [34] also detected several putative cancer-related pathways. The identified pathways are listed in Table **2**. Interestingly, the endometrial cancer pathway [35] was detected in this enrichment analysis ($P = 3.1e-15$, Figure **2A**). This pathway includes major cancer-related pathways, for example, the MAPK signaling pathway and the PI3K/AKT pathway. For this pathway, we compared mutation patterns between our Japanese data and those of the previous study of lung adenocarcinoma in Caucasians [21]. We found that the SNVs in the EGFR gene were four times more frequent in the Japanese population than among Caucasian populations (Figure **2B**, left panel). EGFR mutations were frequently occurring in non-smoker, female and Asian patients of lung adenocarcinoma [36], which is a molecular target of anti-cancer drug, *gefitinib* [20,37,38]. Conversely, KRAS mutations, which are also well-known cancer-related mutations [39], were more than four times frequent among Caucasians (Figure **2B**, center panel). However not all mutational patterns are different between populations. For instance, TP53 harbored mutations in both datasets with similar frequency (Figure **2B**, right panel).

Ambiguity in SNV identification of normal tissue counterparts

In the aforementioned analysis, we discriminated germline variants using the normal tissue counterparts. A number of

Table 1. List of the identified possible cancer-related genes.

Gene	Domain	Number of SNVs		P-value*
		Domain	Gene	
EGFR†	IPR001245:Serine-threonine/tyrosine-protein kinase	34	37	4.4e-21
KRAS†	IPR001806:Ras GTPase	6	7	8.0e-6
TNN	IPR003961:Fibronectin, type III	4	5	5.2e-5
TP53†	IPR008967:p53-like transcription factor, DNA-binding	20	23	9.5e-5
PREX1	IPR000219:Dbl homology (DH) domain	4	5	0.00071
DNAH7	IPR004273:Dynein heavy chain	5	7	0.0025
FSTL5	IPR011044:Quinoprotein amine dehydrogenase, beta chain-like	7	7	0.0043
NRXN3	IPR008985:Concanavalin A-like lectin/ glucanase	5	7	0.0063
PREX2	IPR001849:Pleckstrin homology	3	7	0.011
FER1L6	IPR008973:C2 calcium/lipid-binding domain, CaLB	3	6	0.013
COL22A	IPR008985:Concanavalin A-like lectin/ glucanase	3	6	0.015

* $P < 0.02$

† Reported in the Cancer Gene Census [11]. Note that the genes atop the list are previously reported to be associated with this cancer type, while most of them are novel possible cancer-related genes.

doi: 10.1371/journal.pone.0073484.t001

SNVs initially identified as somatic were also found to be present in normal tissues, thus, were false positive calls under the validations by visual inspection of the mapped sequences and Sanger sequencing. To examine the cause of this problem, we inspected the errors in randomly selected 26 cancers and their normal tissues. On average in each cancer, twenty-five percent of somatic SNV candidates were found to be false positive (Figure **3**). In these cases, the sequence coverage and quality of the normal counterpart were not sufficient. Indeed, the sequences supporting each SNV and these qualities were significantly diverged between the cancer and normal tissues. Although we increased the total number of reads in the normal tissues, it was difficult in practice to cover all of the genomic positions (Figure S4 in File **S1**). A summary of the germline SNV validations is shown in Table S5 in File **S1**.

However, we noticed that some were correctly identified as germline SNVs in external reference exomes. Twenty-five exomes allowed us to exclude eight false positive calls in each cancer. This raised the possibility that the SNVs from the other patients may be used as surrogates to increase the depth and quality of the sequencing.

Excluding germline SNVs by considering mutual overlaps of other persons' exomes

To further test this possibility, we examined whether cancer exome analyses would be possible without sequencing of the normal tissue counterpart of each cancer. First, we evaluated the extent to which the germline SNVs could be discriminated using external exomes. For this purpose, we used the 97 paired cancer-normal exome datasets for the validation dataset. We found that we could detect 54% of the germline SNVs by using the 96 normal tissue exomes from the external reference (Figure **4A**). We further expanded the filtration dataset using the externally available 73 Japanese exome data and 48 in-house Japanese exome datasets. Altogether, we were able to remove 64% of the germline SNVs, using a total of 217 Japanese exome datasets from other individuals, without sequencing each cancer's normal counterpart (Figure **4A**). The extrapolation of the graph also indicated that 1,350 and 2,000 samples would be required to remove 90% and 95% of the germline SNVs, respectively. We expect that such a sample size will be available in near future considering current rapid expansion of the exome analysis.

We further evaluated if the same filtration could be done by solely using cancer exomes. We obtained essentially the same results (Figure S5 in File **S1**). Obvious caveat of this approach is that this would disregard about 3% of somatic SNVs recurrently occurring (Figure S5 in File **S1**, blue). However, as aforementioned, we found that those recurrent SNVs were very rare [15,19] and most of them were derived from dubious somatic SNVs, which were overlooked in the normal tissues. We also consider that most of those recurrent SNVs, if any, can be analyzed separately by sequencing a limited number of normal tissues.

Filtering out germline SNVs by considering mutual overlaps for different ethnic groups and for rare SNPs

We examined whether SNVs in other ethnic backgrounds could be used as external datasets for the filtration. We obtained exome data from individuals of various ethnic backgrounds from the 1000 Genome Project. We used these exome datasets to exclude the germline SNVs that were identified in the Japanese cancers. We found that the discriminative power was significantly lower compared with exomes from Japanese populations. Therefore, these datasets were not suitable for this purpose (Figure **4B**). We also examined and found that the exomes in each ethnic group were useful to discriminate the germline SNVs in the corresponding group (Figure S6, S7 and Table S6 in File **S1**).

We, then, examined to what extent minor germline variants could be covered with this approach in the Japanese population. We evaluated the sensitivity of the filtration process for the SNVs in the 97 cancers (Figure S8 in File **S1**). We found that 88% of the germline SNVs occurring in more than five percent of the 97 exomes could be detected using the 73 external Japanese datasets. For the SNVs occurring in 1% of the 97 cancers, 19% could be excluded.

Using the crude dataset to characterize cancer related SNVs and pathways

Taken together, with 217 Japanese exomes used for filtration, 36% of the germline SNVs remained unfiltered. Nevertheless, we considered that it may be still possible to use the crude SNV dataset as a first approximation for identifying and analyzing cancer-related genes and pathway candidates. To validate this idea, we compared the results of enrichment analyses between the crude dataset and the refined somatic

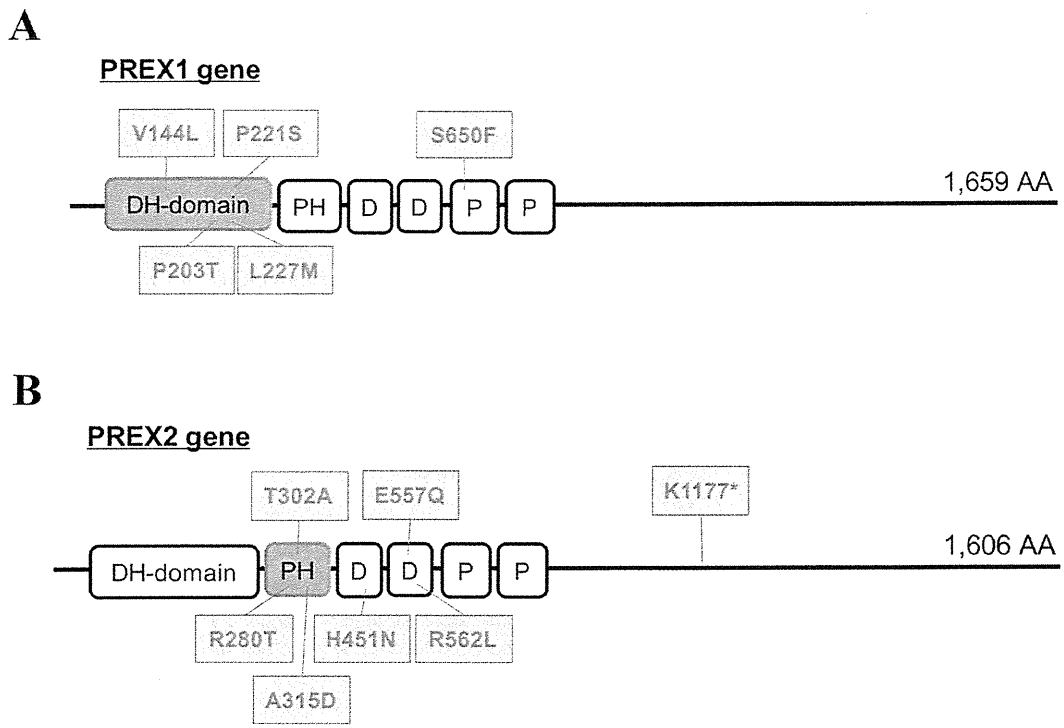

Fig. 1

Figure 1. Identification and characterization of the putative cancer-related genes using 97 cancer exomes. SNVs in the PREX1 (**A**) and PREX2 (**B**) genes are represented in the boxes. The protein domains in which the enrichments of the SNVs were statistically significant are represented in orange boxes (also see Materials and Method). DH-domain: Dbl homology (DH) domain; PH: Pleckstrin homology domain; D: DEP domain; P: PDZ/DHR/GLGF.

doi: 10.1371/journal.pone.0073484.g001

SNV datasets, which were generated from the paired cancer-normal exomes.

Most of the putative cancer-related genes and pathways that were identified from the refined dataset were also present in the crude dataset (Tables S7 and S8 in File **S1**). The example of the TNN gene, which was reported as a marker of tumor stroma [40-42], is shown in Figure S9 in File **S1**. In this case, even with the germline SNVs, which were unfiltered in the crude dataset (indicated by black in Figure S9 in File **S1**), the enrichment of somatic SNVs in this domain was statistically significant. In total, nine genes which identified as possessing cancer-related SNVs from the refined dataset were also detected in the crude dataset. On the other hand, two genes from the refined dataset were not represented in the crude dataset. In the pathway analysis, we identified 26 cancer-related pathways which were identified from the refined dataset. In addition, 19 pathways were also represented in the crude dataset as well as the refined dataset. The overlap between the datasets is summarized in Table **3**. It should be noted that statistically enrichment analyses were possible even at the current coverage of the filter dataset. With the expanded external dataset, it would be more practical to subject the candidates to the results of Sanger sequencing validations as well as removing remaining germline SNVs.

Identification of prognosis related genes by using the crude dataset

As one of the most important objectives of the cancer exome studies, we investigated whether mutations affecting cancer prognoses can be identified by using crude dataset (Table S9 and Figure S10 in File **S1**). In the Kaplan-Meier analysis, seven patients who carried SNVs in the ATM gene (Figure **5A**)

Table 2. List of the identified possible cancer-related pathways.

KEGG ID	Pathway definition	Number of cancers with SNVs	P-value*
hsa05213	Endometrial cancer	72	3.1e-15
hsa04320	Dorso-ventral axis formation	48	4.4e-15
hsa05219	Bladder cancer	62	4.9e-14
hsa05223	Non-small cell lung cancer	66	7.1e-12
hsa05214	Glioma	70	6.5e-11
hsa05218	Melanoma	70	1.3e-9
hsa05212	Pancreatic cancer	68	6.9e-9
hsa05215	Prostate cancer	71	4.3e-7
hsa05216	Thyroid cancer	36	1.1e-6
hsa04520	Adherens junction	59	3.7e-6
hsa05210	Colorectal cancer	53	1.8e-5
hsa04012	ErbB signaling pathway	64	2.6e-5
hsa05120	Epithelial cell signaling in *Helicobacter pylori* infection	53	4.8e-5
hsa04540	Gap junction	60	0.00024
hsa04912	GnRH signaling pathway	61	0.0011
hsa05217	Basal cell carcinoma	41	0.0020
hsa05222	Small cell lung cancer	52	0.0069
hsa05220	Chronic myeloid leukemia	46	0.010
hsa05160	Hepatitis C	67	0.012
hsa05014	Amyotrophic lateral sclerosis (ALS)	36	0.014
hsa04977	Vitamin digestion and absorption	20	0.015
hsa05416	Viral myocarditis	40	0.028
hsa04512	ECM-receptor interaction	47	0.034
hsa02010	ABC transporters	29	0.035
hsa04510	Focal adhesion	78	0.037
hsa05412	Arrhythmogenic right ventricular cardiomyopathy (ARVC)	40	0.039

* $P < 0.05$

doi: 10.1371/journal.pone.0073484.t002

showed statistically significant poor prognoses (P = 9.6e-6, Figure **5B**). Three SNVs in the ATM gene were significantly enriched in the the phosphatidylinositol 3-/4-kinase catalytic domain (P = 0.014). ATM senses DNA damage and phosphorylates TP53, which, in turn, invokes various cellular responses, such as DNA repair, growth arrest and apoptosis, and collectively prevents cancer progression (Figure S11 in File **S1**) [43,44].

We also examined whether other frequently mutated genes were associated with better or worse prognoses. We found that patients with PAPPA2 mutations showed prolonged survival times (P = 0.026, Figure 5C and D). PAPPA2 proteolyzes IGFBP5 [45,46], which is an inhibitory factor for IGFs [47]. Mutations in the PAPPA2 gene may result in the accumulation of IGFBP5, and the resulting decrease in IGF signaling may impair the proliferation of cancer cells [48]. Again, it should be noted that for both the ATM and PAPPA2 genes, the statistical significance of the prognostic difference persisted both before (black line) and after (red line) the remaining germline mutations were removed, which was validated by Sanger sequencing (Figure 5B, D and Table S10 in File **S1**).

Conclusions

We have identified and characterized the SNVs in lung adenocarcinoma in a Japanese population. Further biological evaluations of the discovered SNVs will be described elsewhere. In particular, information of transcriptome and epigenome should be important for further analyses of cancer genomes, as they would shed new lights on the cancer biology (Table **S1**) [49]. In this study, we also presented a useful approach for the analysis of cancer exomes, without the need to sequence the normal tissue counterpart. We believe that the approach not only lowers the barriers in cost, time and data fidelity in the exome analysis, but also enables exome analysis of archive samples, for which normal tissue counterparts are not always available.

Materials and Methods

Ethics statement

All of the samples were collected by following the protocol (and written informed consent) which were approved by Ethical Committee in National Cancer Center, Japan (Correspondence to: Katsuya Tsuchihara; ktsuchih@east.ncc.go.jp).

Case selection and DNA preparation

All of the tissue materials were obtained from Japanese lung adenocarcinoma patients with the appropriate informed consent. Surgically resected primary lung adenocarcinoma samples with lengthwise dimensions in excess of 3 cm were selected. Data on the 52 patients who had relapses and other clinical information about the 97 cases are shown in Table S11 in File **S1**. All 97 cancer and normal tissues were extracted from methanol-fixed samples by laser capture microdissection. DNA purification was performed using an EZ1 Advanced XL Robotic workstation with EZ1 DNA Tissue Kits (Qiagen).

Whole-exome sequencing

Using 1 μg of isolated DNA, we prepared exome-sequencing libraries using the SureSelect Target Enrichment System (Agilent Technologies) according to the manufacturer's protocol. The captured DNA was sequenced by the illumina Genome Analyzer IIx platform (Illumina), yielding 76-base paired-end reads.

Somatic SNV detection

The methods that were used to detect the SNVs, including BWA, SAMtools [50] and GATK, are shown in Figure S2 in File S1. Using data from NCBI dbSNP build 132 and one Japanese genome [51], major germline SNVs were excluded. In addition, rare germline SNVs were discarded using 97 exomes from normal tissue counterparts, 73 Japanese exomes provided from the 1000 Genomes Project (the phase1 exome data, 20110521) and 48 in-house Japanese exomes. We also validated a portion of the SNV datasets by the Sanger

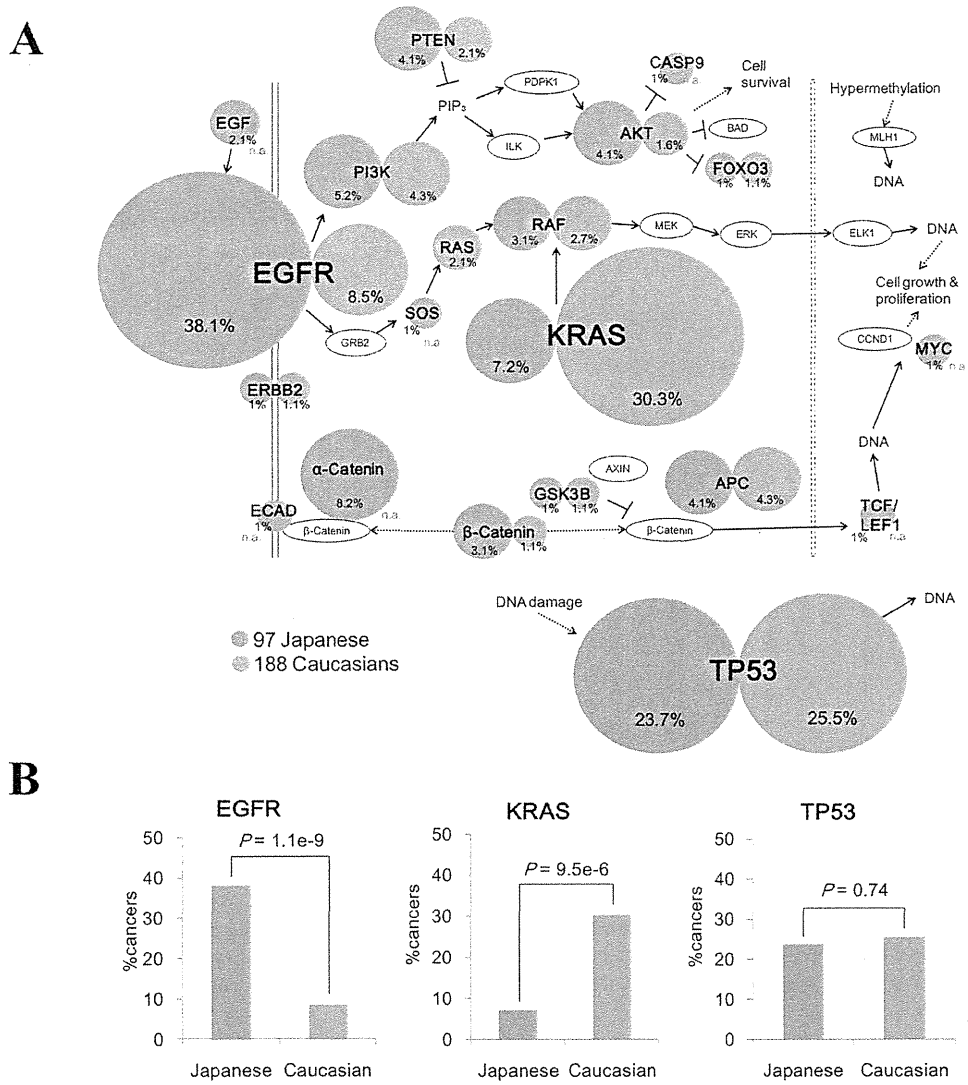

Fig. 2

Figure 2. The EGFR/Ras pathways in Japanese and Caucasian populations. (**A**) Mutation patterns in the endometrial cancer pathway that was detected in the enrichment analysis are shown. The size of the circle represents the population of the cancers harboring the SNVs in the corresponding gene (percentage is also shown in the margin). SNVs in this study and the external dataset in Caucasian populations are shown in red and blue circles, respectively. n.a.: mutation frequencies were not available. (**B**) Comparison of mutation ratio of EGFR, KRAS and TP53 genes among both datasets. The p-values were calculated by two-sample test for equality of proportions.

doi: 10.1371/journal.pone.0073484.g002

sequencing of cancer tissues and their normal tissue counterparts (Figure S12 in File **S1**).

Identification of highly mutated genes

We detected genes which were significantly enriched with SNVs by calculating the expected number of cancers with SNVs in the gene. The length of total CDS regions was represented in N (approximately 30.8 M bases). When one patient harbored total of m SNVs, the probability that the patient harbors SNVs in the gene t (length: n) was calculated as P:

$$P_{m,t,n} = 1 - \left(1 - \frac{m}{N}\right)^n$$

The sum of P in 97 cancers was represented in the expected number of cancers with SNVs in the gene t. The p-values of the

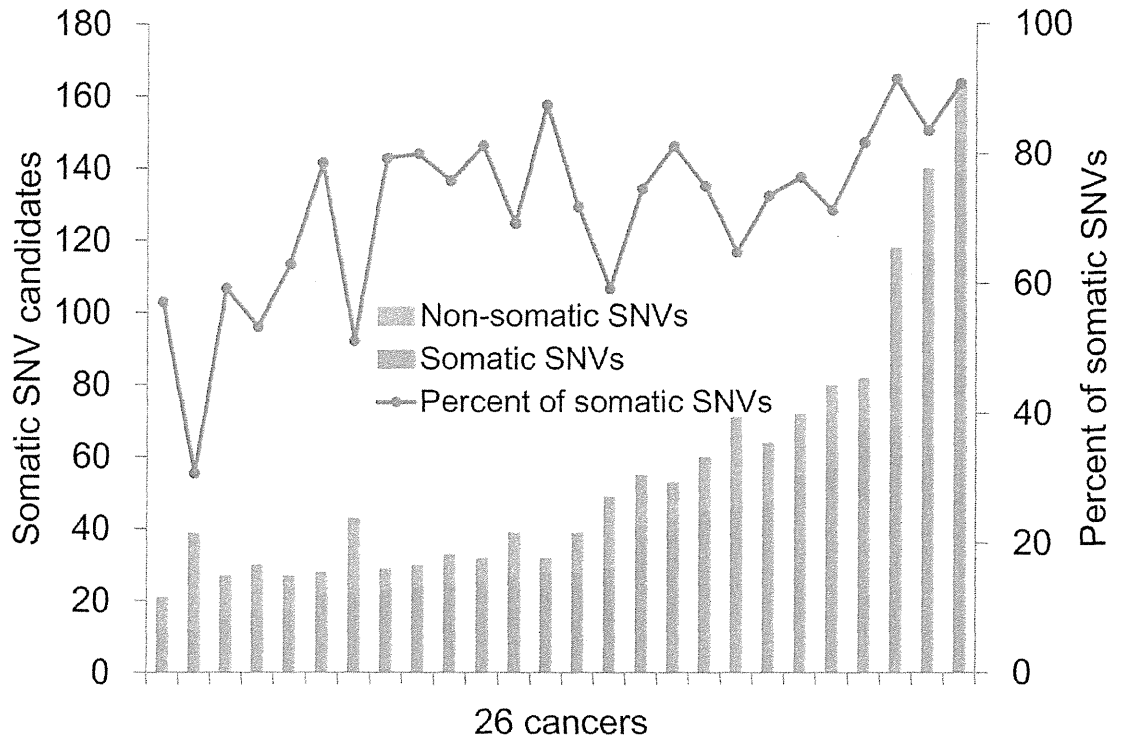

Fig. 3

Figure 3. Fidelity of the germline SNV detection in cancer exome analysis. Somatic SNV candidates were identified by using 26 cancer exomes and each normal counterpart. Correct somatic SNVs and false positives were shown in pink and blue bars, respectively. The 26 cancers used for the analysis were sorted by the increasing total number of SNVs (x-axis).

doi: 10.1371/journal.pone.0073484.g003

observed number were calculated by the Poisson probability function using R ppois.

Statistical approach to enrichment analyses

To examine the enrichment of mutations in functional protein domains, we mapped the SNVs to domains using InterProScan [52] and assigned them to the Catalogue of Somatic Mutations in Cancer (COSMIC). We analyzed the enrichment of the SNVs in the same domains as the mutations that were provided by the COSMIC. The p-values for the observed mutations in these domains were calculated using their hypergeometric distributions (R phyper). Briefly, the domains in which the SNVs were enriched statistically significantly than the expected number of SNVs in the given length of the domain were selected. For estimating the expected number, the total number of the SNVs belonging to the gene was divided by the gene length. For this analysis, we used genes harboring five or more SNVs in the coding region and three or more SNVs in the domain.

We assigned SNVs to pathways as described by the Kyoto Encyclopedia of Genes and Genomes (KEGG) and calculated the enrichments of the SNVs in the pathways. The mutation rate M represented the ratio of the average number of mutated genes to the total number of genes (17,175) that were used in our study. The expected value for the number of cancers with SNVs in pathway t was designated λ and calculated from the mutation rate M and the number of genes in the pathway n as follows:

$$\lambda_{t,n} = \left\{1-(1-M)^n\right\} \times 97$$

The p-value for the observed number of cancers with SNVs in pathway t was calculated by the Poisson probability function using R ppois.

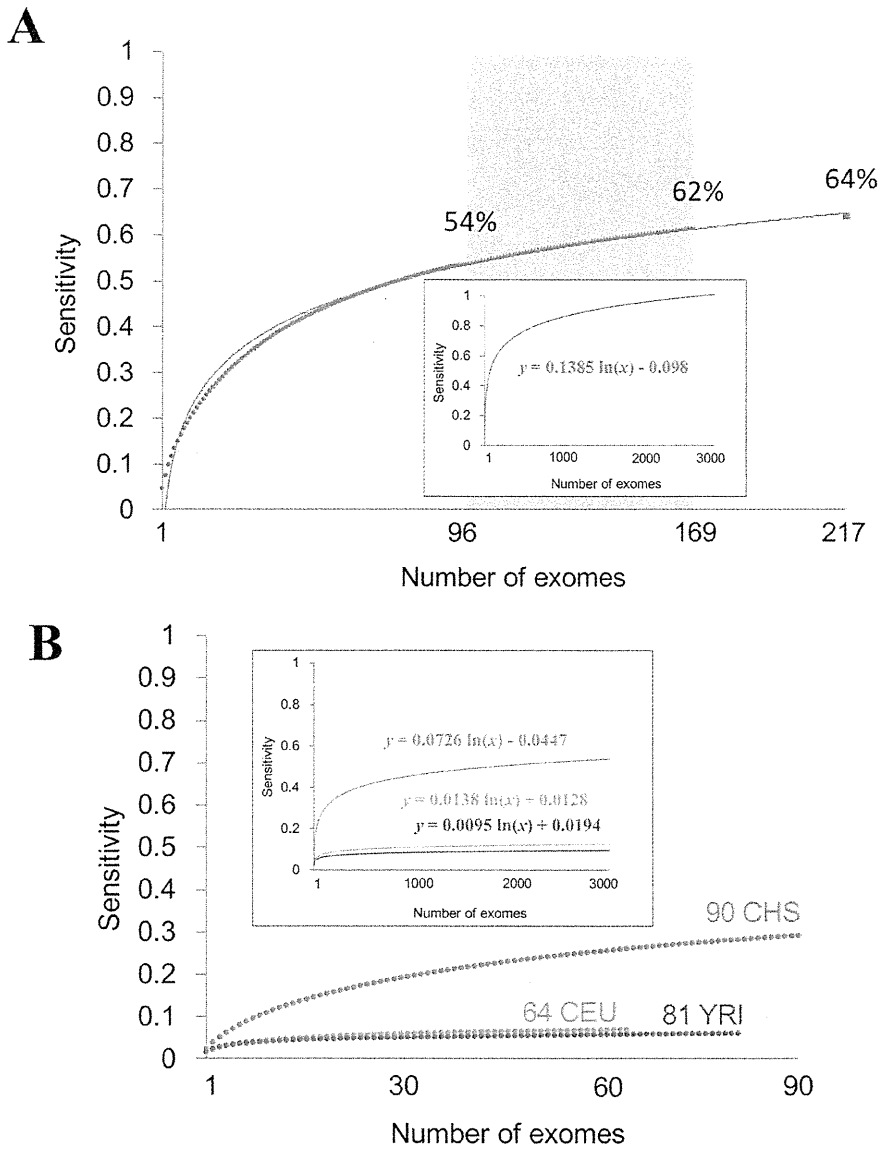

Fig. 4

Figure 4. Discriminative powers of detecting germline SNVs using external references. (**A**) The power of detecting germline SNVs considering mutual overlap between other Japanese individuals. Sensitivity represents the proportion of germline SNVs correctly detected. The datasets used to exclude the germline SNVs are shown on the x axis. The inset represents the extrapolation of the graph. Fitting curve of the graph is also shown. (**B**) Discriminative powers of three different ethnic groups for the germline SNVs in 97 Japanese cancers. Sensitivities for detecting germline SNVs are shown by the following colors; green: Chinese; purple: Yoruba; orange: Caucasian.

doi: 10.1371/journal.pone.0073484.g004

Estimate of discriminative power for exclusion of germline SNVs by considering mutual overlaps

We estimated the discriminative power for the exclusion of germline SNVs by considering those from other non-cancerous exomes. Germline SNVs from 97 paired tumor-normal exomes were used as reference datasets. Up to 217 samples (96 normal tissue exomes from others and 121 additional Japanese exomes) were randomly selected, and their sensitivities and specificities for detecting the germline SNVs were detected by taking the averages of either all of the combinations or a subset of approximately 10,000 combinations. We also estimated the discriminative power with

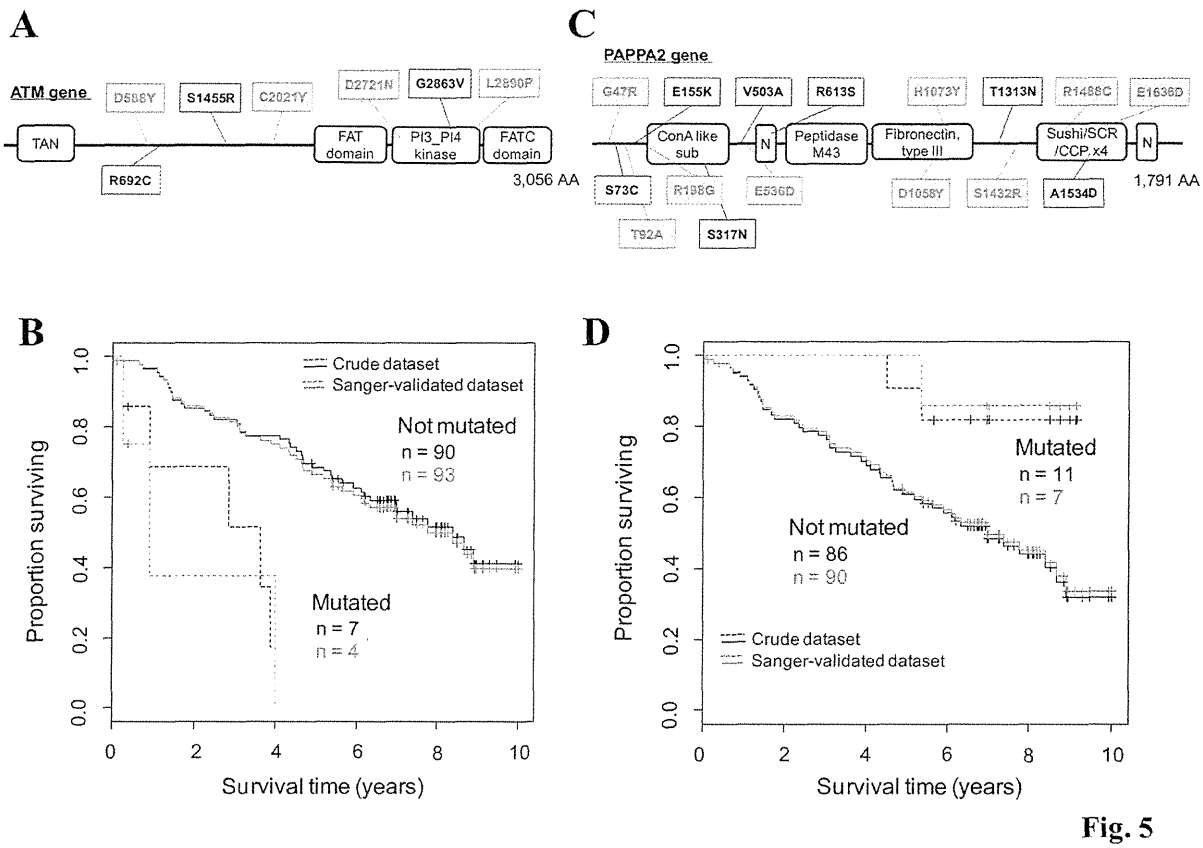

Fig. 5

Figure 5. Identification of the putative prognosis-related genes. (**A**) SNVs in the ATM gene. The SNVs that were identified in the initial screening and those remaining after the Sanger sequencing validation of the normal-tissue counterpart were shown in black and red, respectively. TAN: Telomere-length maintenance and DNA damage repair; PI3_PI4 kinase: Phosphatidylinositol 3-/4-kinase, catalytic. (**B**) Survival analysis of patients with and without ATM SNVs. The datasets before and after the Sanger sequencing validation are represented by black and red lines, respectively. Statistical significance was calculated using a log-rank test ($P < 0.05$). Note that the survival differences for individuals with SNVs in the non-Sanger-validated dataset were significant before the Sanger validation. (**C**, **D**) Results of a similar analysis as that described in A and B for the PAPPA2 gene. In this case, the patients with the SNVs showed better prognoses. ConA like sub: Concanavalin A-like lectin/glucanase, subgroup; N: Notch dimain; Peptidase M43: Peptidase M43, pregnancy-associated plasma-A.

doi: 10.1371/journal.pone.0073484.g005

Table 3. Comparison of the results in the enrichment analyses between the crude and refined dataset.

	Number of identified genes/pathways		
	Crude*	Refined†	Overlap‡
Genes	16	11	9
Pathways	23	26	19

* Identified using the crude dataset.

† Identified using the refined dataset.

‡ Significant in both crude and refined datasets.

doi: 10.1371/journal.pone.0073484.t003

data from the 1000 Genomes Project for four ethnic groups (73 JPT, 90 CHS, 81 YRI and 64 CEU) using similar trials. Whole-exome sequences (the phase1 exome data, 20110521) were obtained from the ftp site in the 1000 Genomes Project.

Kaplan-Meier curves

The Kaplan-Meier method was used to test the relations of the observed mutations to survival time, and calculations were performed using the R software package. Changes in survival rates that were correlated with SNVs were examined using the log-rank test (R survdiff).