(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LAD samples identified one case of *CCDC6-RET* expressing LAD by RT-PCR and break-apart FISH (Tsuta *et al.*, 2012, unpublished data).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

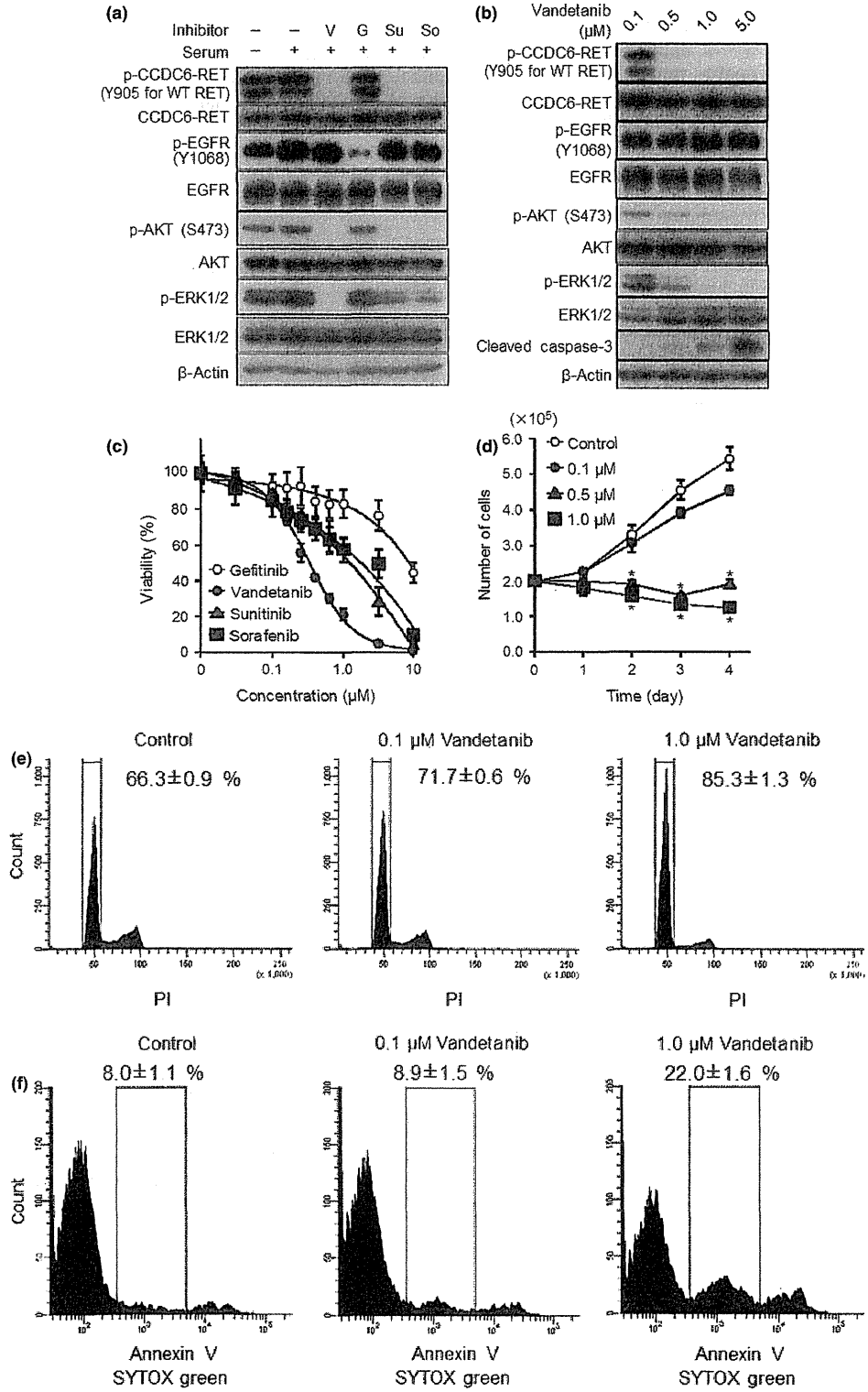

Fig. 3. Effect of RET inhibitors on LC-2/ad cells. (a) Western blot analysis of inhibitor-treated cells. The cells were incubated under serum-starved conditions for 22 h and treated with 1 μM of inhibitor or dimethylsulfoxide (DMSO) for 2 h. Prior to cell lysis, the cells were treated with 10% fetal 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 ($n = 4$).

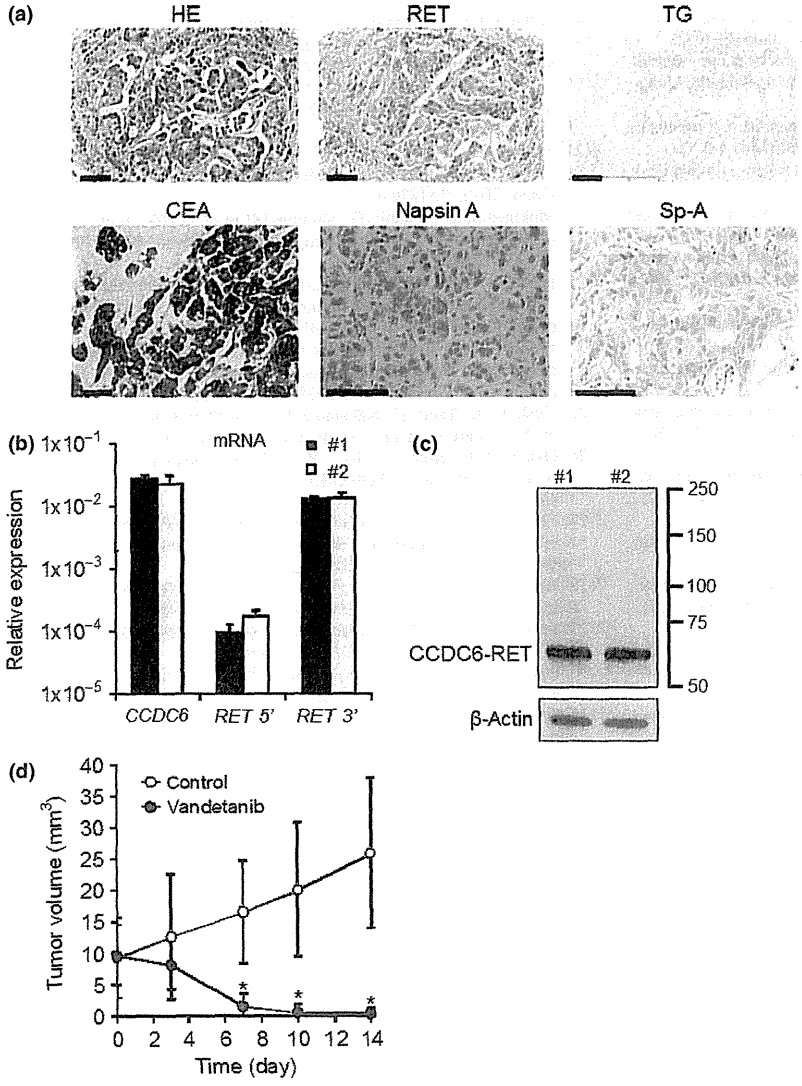

Fig. 4. Characterization of the LC-2/ad xenograft and anti-tumor effects of vandetanib. (a) Histological features of the xenograft. Hematoxylin and eosin staining and immunohistochemical staining with the indicated antibodies. Scale bars were 100 μm. Hematoxylin eosin (HE), RET, thyroglobulin (TG) and 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 ($P = 3.1e\text{-}15$, 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		
		Domain	Gene	P-value*
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

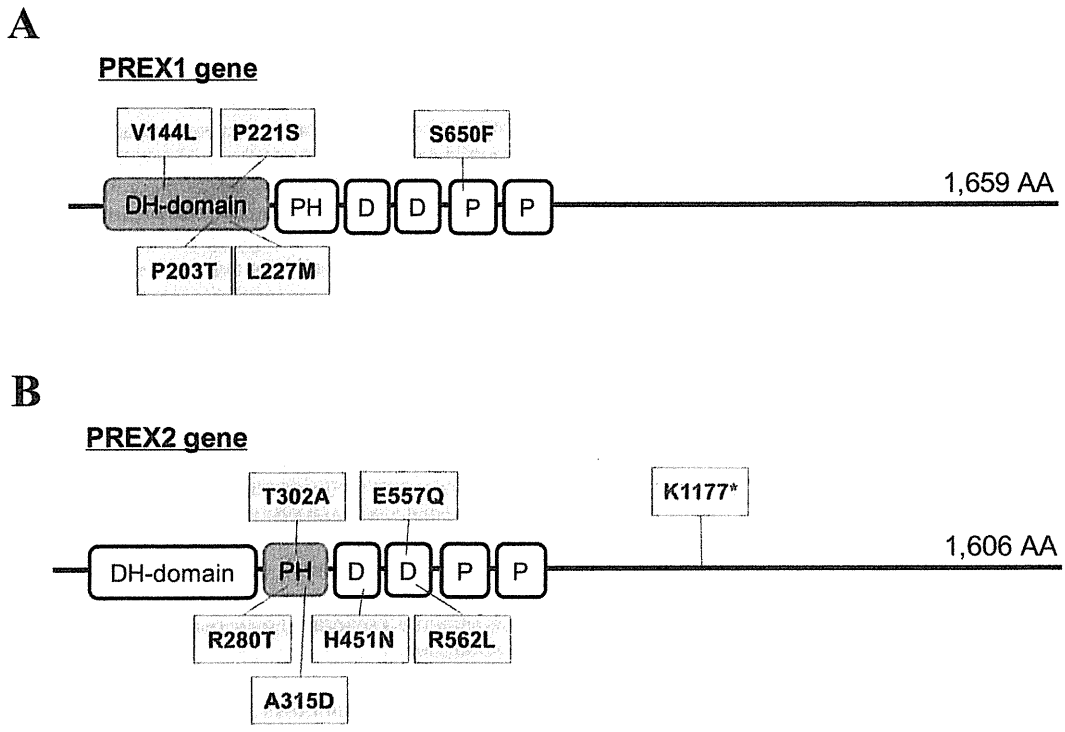

Fig. 1

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

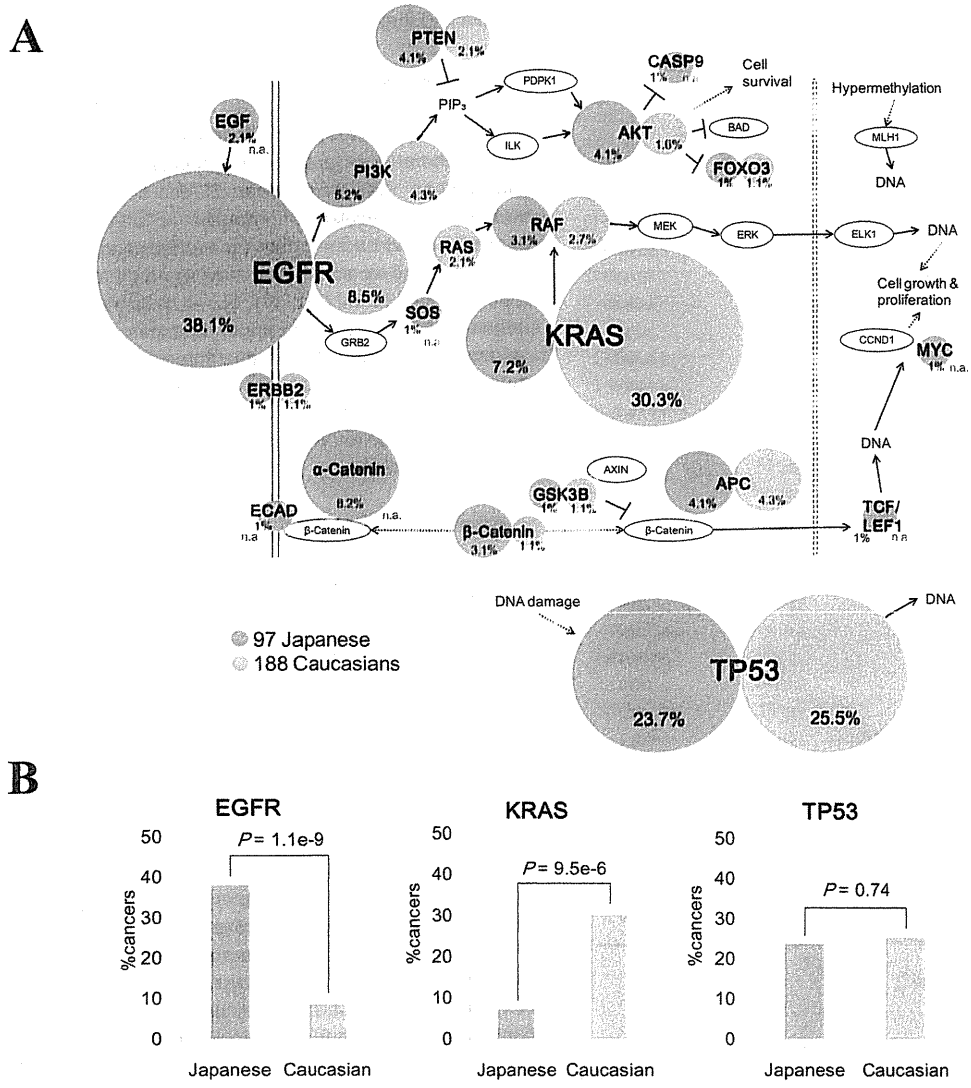

Fig. 2

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

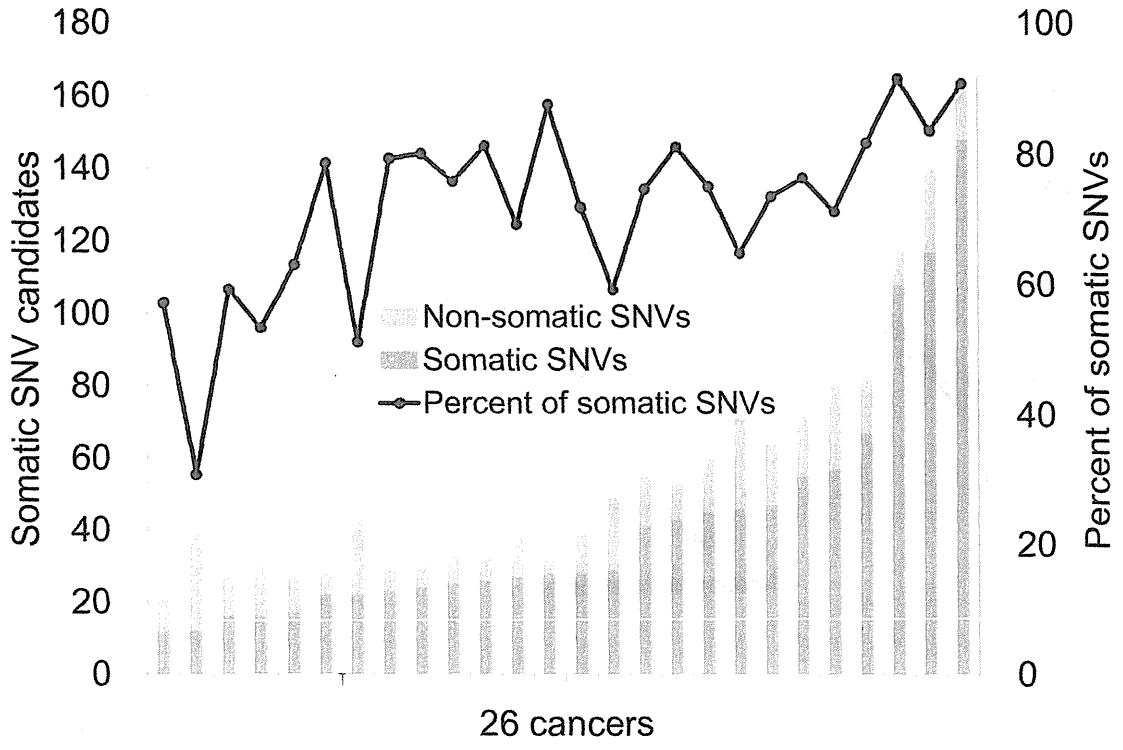

Fig. 3

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.

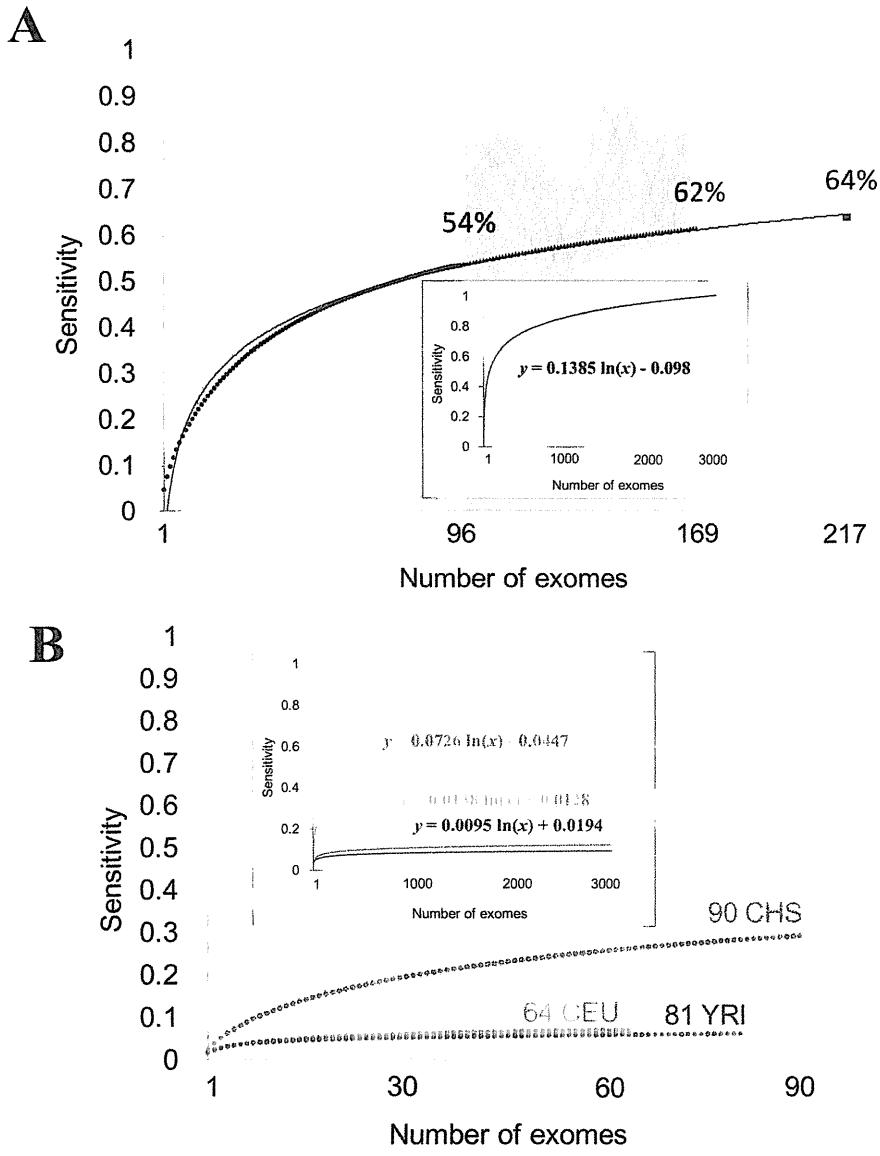

Fig. 4

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

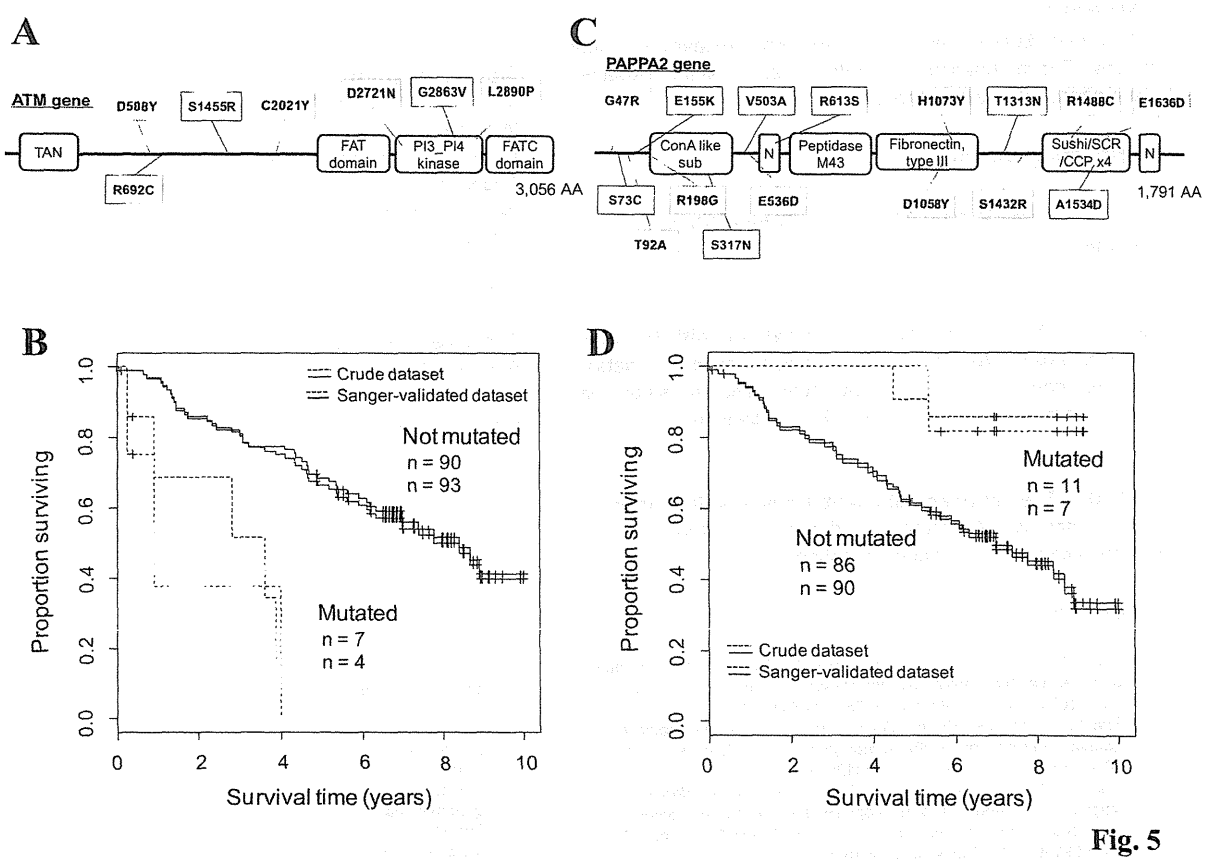

Fig. 5

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).

Data access

Full raw datasets will be shared with researchers upon request. The information of somatic mutations at the respective genomic coordinates has been provided in Table **S2**.

Supporting Information

File S1. Figures S1 to S12 and Tables S3 to S11 are included.
(PDF)

Table S1. The comparison of our dataset with the other different study. We provided the comparison of our dataset with the genes identified in the other different study with transcriptome and epigenome data in lung cancers.
(XLSX)

Table S2. The list of somatic mutations identified from the refined dataset. All mutations described in this table are somatic and non-synonymous mutations.
(XLSX)

Acknowledgements

We thank W. Makałowski for constructive comments and suggestions for this manuscript. We are also grateful to F. Todokoro and F. Iguchi for their technical assistance with the data processing and N. Ogasawara for the Sanger sequencing validation.

Author Contributions

Conceived and designed the experiments: KT YS HE KG SS. Performed the experiments: SM YY AK KM MS. Analyzed the data: AS YS KT. Contributed reagents/materials/analysis tools: KG KT. Wrote the manuscript: AS KT YS.

References

1. The 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061-1073. doi:10.1038/nature09534. PubMed: 20981092.
2. The 1000 Genomes Project Consortium (2012) An integrated map of genetic variation from 1,092 human genomes. Nature 491: 56-65. doi: 10.1038/nature11632. PubMed: 23128226.
3. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR et al. (2009) Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A 106: 19096-19101. doi:10.1073/pnas.0910672106. PubMed: 19861545.
4. Ng SB, Turner EH, Robertson PD, Flygare SD, Bigham AW et al. (2009) Targeted capture and massively parallel sequencing of 12 human exomes. Nature 461: 272-276. doi:10.1038/nature08250. PubMed: 19684571.
5. Clark MJ, Chen R, Lam HY, Karczewski KJ, Chen R et al. (2011) Performance comparison of exome DNA sequencing technologies. Nat Biotechnol 29: 908-914. doi:10.1038/nbt.1975. PubMed: 21947028.
6. Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ et al. (2011) Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12: 745-755. doi:10.1038/nrg3031. PubMed: 21946919.
7. Ng SB, Buckingham KJ, Lee C, Bigham AW, Tabor HK et al. (2010) Exome sequencing identifies the cause of a mendelian disorder. Nat Genet 42: 30-35. doi:10.1038/ng.499. PubMed: 19915526.
8. Biesecker LG (2010) Exome sequencing makes medical genomics a reality. Nat Genet 42: 13-14. doi:10.1038/ng0110-13. PubMed: 20037612.
9. Louis-Dit-Picard H, Barc J, Trujillano D, Miserey-Lenkei S, Bouatia-Naji N et al. (2012) KLHL3 mutations cause familial hyperkalemic hypertension by impairing ion transport in the distal nephron. Nat Genet.
10. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L et al. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29: 308-311. doi:10.1093/nar/29.1.308. PubMed: 11125122.
11. Futreal PA, Coin L, Marshall M, Down T, Hubbard T et al. (2004) A census of human cancer genes. Nat Rev Cancer 4: 177-183. doi: 10.1038/nrc1299. PubMed: 14993899.
12. International Cancer Genome Consortium (2010) International network of cancer genome projects. Nature 464: 993-998. doi:10.1038/nature08987. PubMed: 20393554.
13. Totoki Y, Tatsuno K, Yamamoto S, Arai Y, Hosoda F et al. (2011) High-resolution characterization of a hepatocellular carcinoma genome. Nat Genet 43: 464-469. doi:10.1038/ng.804. PubMed: 21499249.
14. Jones DT, Jäger N, Kool M, Zichner T, Hutter B et al. (2012) Dissecting the genomic complexity underlying medulloblastoma. Nature 488: 100-105. doi:10.1038/nature11284. PubMed: 22832583.
15. The Cancer Genome Atlas Research Network (2011) Integrated genomic analyses of ovarian carcinoma. Nature 474: 609-615. doi: 10.1038/nature10166. PubMed: 21720365.
16. Stephens PJ, Tarpey PS, Davies H, Van Loo P, Greenman C et al. (2012) The landscape of cancer genes and mutational processes in breast cancer. Nature 486: 400-404. PubMed: 22722201.
17. The Cancer Genome Atlas Research Network (2012) Comprehensive genomic characterization of squamous cell lung cancers. Nature.
18. The Cancer Genome Atlas Research Network (2012) Comprehensive molecular characterization of human colon and rectal cancer. Nature 487: 330-337. doi:10.1038/nature11252. PubMed: 22810696.
19. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ et al. (2012) Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell 150: 1107-1120. doi:10.1016/j.cell.2012.08.029. PubMed: 22980975.
20. Davies H, Hunter C, Smith R, Stephens P, Greenman C et al. (2005) Somatic mutations of the protein kinase gene family in human lung cancer. Cancer Res 65: 7591-7595. PubMed: 16140923.
21. Ding L, Getz G, Wheeler DA, Mardis ER, McLellan MD et al. (2008) Somatic mutations affect key pathways in lung adenocarcinoma. Nature 455: 1069-1075. doi:10.1038/nature07423. PubMed: 18948947.
22. Kan Z, Jaiswal BS, Stinson J, Janakiraman V, Bhatt D et al. (2010) Diverse somatic mutation patterns and pathway alterations in human cancers. Nature 466: 869-873. doi:10.1038/nature09208. PubMed: 20668451.
23. Forbes SA, Bhamra G, Bamford S, Dawson E, Kok C et al. (2008) The Catalogue of Somatic Mutations in Cancer (COSMIC). Curr Protoc Hum Genet Chapter 10: Unit 10.11: PubMed: 18428421
24. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY et al. (2011) COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic Acids Res 39: D945-D950. doi: 10.1093/nar/gkq929. PubMed: 20952405.
25. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760. doi: 10.1093/bioinformatics/btp324. PubMed: 19451168.
26. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K et al. (2010) The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 20: 1297-1303. doi:10.1101/gr.107524.110. PubMed: 20644199.
27. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43: 491-498. doi: 10.1038/ng.806. PubMed: 21478889.
28. Welch HC, Coadwell WJ, Ellson CD, Ferguson GJ, Andrews SR et al. (2002) P-Rex1, a PtdIns(3,4,5)P3- and Gbetagamma-regulated guanine-nucleotide exchange factor for Rac. Cell 108: 809-821. doi: 10.1016/S0092-8674(02)00663-3. PubMed: 11955434.

29. Rosenfeldt H, Vázquez-Prado J, Gutkind JS (2004) P-REX2, a novel PI-3-kinase sensitive Rac exchange factor. FEBS Lett 572: 167-171. doi:10.1016/j.febslet.2004.06.097. PubMed: 15304342.
30. Das B, Shu X, Day GJ, Han J, Krishna UM et al. (2000) Control of intramolecular interactions between the pleckstrin homology and Dbl homology domains of Vav and Sos1 regulates Rac binding. J Biol Chem 275: 15074-15081. doi:10.1074/jbc.M907269199. PubMed: 10748082.
31. Hill K, Krugmann S, Andrews SR, Coadwell WJ, Finan P et al. (2005) Regulation of P-Rex1 by phosphatidylinositol (3,4,5)-trisphosphate and Gbetagamma subunits. J Biol Chem 280: 4166-4173. PubMed: 15545267.
32. Chhatriwala MK, Betts L, Worthylake DK, Sondek J (2007) The DH and PH domains of Trio coordinately engage Rho GTPases for their efficient activation. J Mol Biol 368: 1307-1320. doi:10.1016/j.jmb.2007.02.060. PubMed: 17391702.
33. Sosa MS, Lopez-Haber C, Yang C, Wang H, Lemmon MA et al. (2010) Identification of the Rac-GEF P-Rex1 as an essential mediator of ErbB signaling in breast cancer. Mol Cell 40: 877-892. doi:10.1016/j.molcel.2010.11.029. PubMed: 21172654.
34. Kanehisa M, Goto S, Furumichi M, Tanabe M, Hirakawa M (2010) KEGG for representation and analysis of molecular networks involving diseases and drugs. Nucleic Acids Res 38: D355-D360. doi:10.1093/nar/gkp896. PubMed: 19880382.
35. Hecht JL, Mutter GL (2006) Molecular and pathologic aspects of endometrial carcinogenesis. J Clin Oncol 24: 4783-4791. doi:10.1200/JCO.2006.06.7173. PubMed: 17028294.
36. Li C, Fang R, Sun Y, Han X, Li F et al. (2011) Spectrum of oncogenic driver mutations in lung adenocarcinomas from East Asian never smokers. PLOS ONE 6: e28204. doi:10.1371/journal.pone.0028204. PubMed: 22140546.
37. Shigematsu H, Lin L, Takahashi T, Nomura M, Suzuki M et al. (2005) Clinical and biological features associated with epidermal growth factor receptor gene mutations in lung cancers. J Natl Cancer Inst 97: 339-346. doi:10.1093/jnci/dji055. PubMed: 15741570.
38. Sharma SV, Bell DW, Settleman J, Haber DA (2007) Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer 7: 169-181. doi:10.1038/nrc2088. PubMed: 17318210.
39. Bos JL (1989) ras oncogenes in human cancer: a review. Cancer Res 49: 4682-4689. PubMed: 2547513.
40. Degen M, Brellier F, Kain R, Ruiz C, Terracciano L et al. (2007) Tenascin-W is a novel marker for activated tumor stroma in low-grade human breast cancer and influences cell behavior. Cancer Res 67: 9169-9179. doi:10.1158/0008-5472.CAN-07-0666. PubMed: 17909022.
41. Degen M, Brellier F, Schenk S, Driscoll R, Zaman K et al. (2008) Tenascin-W, a new marker of cancer stroma, is elevated in sera of colon and breast cancer patients. Int J Cancer 122: 2454-2461. doi:10.1002/ijc.23417. PubMed: 18306355.
42. Brellier F, Martina E, Degen M, Heuzé-Vourc'h N, Petit A et al. (2012) Tenascin-W is a better cancer biomarker than tenascin-C for most human solid tumors. BMC Clin Pathol 12: 14. doi:10.1186/1472-6890-12-14. PubMed: 22947174.
43. Dasika GK, Lin SC, Zhao S, Sung P, Tomkinson A et al. (1999) DNA damage-induced cell cycle checkpoints and DNA strand break repair in development and tumorigenesis. Oncogene 18: 7883-7899. PubMed: 10630641.
44. Bartkova J, Horejsi Z, Koed K, Kramer A, Tort F et al. (2005) DNA damage response as a candidate anti-cancer barrier in early human tumorigenesis. Nature 434: 864-870. doi:10.1038/nature03482. PubMed: 15829956.
45. Yan X, Baxter RC, Firth SM (2010) Involvement of pregnancy-associated plasma protein-A2 in insulin-like growth factor (IGF) binding protein-5 proteolysis during pregnancy: a potential mechanism for increasing IGF bioavailability. J Clin Endocrinol Metab 95: 1412-1420. doi:10.1210/jc.2009-2277. PubMed: 20103653.
46. Overgaard MT, Boldt HB, Laursen LS, Sottrup-Jensen L, Conover CA et al. (2001) Pregnancy-associated plasma protein-A2 (PAPP-A2), a novel insulin-like growth factor-binding protein-5 proteinase. J Biol Chem 276: 21849-21853. doi:10.1074/jbc.M102191200. PubMed: 11264294.
47. Shimasaki S, Shimonaka M, Zhang HP, Ling N (1991) Identification of five different insulin-like growth factor binding proteins (IGFBPs) from adult rat serum and molecular cloning of a novel IGFBP-5 in rat and human. J Biol Chem 266: 10646-10653. PubMed: 1709938.
48. Su Y, Wagner ER, Luo Q, Huang J, Chen L et al. (2011) Insulin-like growth factor binding protein 5 suppresses tumor growth and metastasis of human osteosarcoma. Oncogene 30: 3907-3917. doi:10.1038/onc.2011.97. PubMed: 21460855.
49. Huang T, Jiang M, Kong X, Cai YD (2012) Dysfunctions associated with methylation, microRNA expression and gene expression in lung cancer. PLOS ONE 7: e43441. doi:10.1371/journal.pone.0043441. PubMed: 22912875.
50. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J et al. (2009) The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079. doi:10.1093/bioinformatics/btp352. PubMed: 19505943.
51. Fujimoto A, Nakagawa H, Hosono N, Nakano K, Abe T et al. (2010) Whole-genome sequencing and comprehensive variant analysis of a Japanese individual using massively parallel sequencing. Nat Genet 42: 931-936. doi:10.1038/ng.691. PubMed: 20972442.
52. Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N et al. (2005) InterProScan: protein domains identifier. Nucleic Acids Res 33: W116-W120. doi:10.1093/nar/gni118. PubMed: 15980438.

Cancer Science
Japanese Cancer Association JCA Open Access

Effect of a poly(ADP-ribose) polymerase-1 inhibitor against esophageal squamous cell carcinoma cell lines

Tomomitsu Nasuno,[1,2] Sachiyo Mimaki,[1] Makito Okamoto,[2] Hiroyasu Esumi[1,3] and Katsuya Tsuchihara[1]

[1]Division of Translational Research, Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, Kashiwa, Chiba; [2]Department of Otorhinolaryngology, University of Kitasato Hospital, Minami-ku, Sagamihara, Kanagawa; [3]Research Institute for Biological Sciences, Tokyo University of Science, Noda, Chiba, Japan

Key words

DNA repair, esophageal cancer, poly(ADP-ribose) polymerase inhibitor, RNF8, γ-H2AX

Correspondence

Katsuya Tsuchihara, Division of Translational Research, Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, 6-5-1 Kashiwanoha, Kashiwa, Chiba 277-8577, Japan.
Tel: +81-4-7133-1111; Fax: +81-4-7134-8786;
E-mail: ktsuchih@east.ncc.go.jp

Funding information

National Cancer Center Research and Development Fund (23-A-8, 15). Third Term Comprehensive 10-year Strategy for Cancer Control from the Ministry of Health, Labour and Welfare (H22-033).

Received July 8, 2013; Revised November 5, 2013; Accepted November 11, 2013

Cancer Sci **105** (2014) 202–210

doi: 10.1111/cas.12322

Effective molecular target drugs that improve therapeutic efficacy with fewer adverse effects for esophageal cancer are highly anticipated. Poly(ADP-ribose) polymerase (PARP) inhibitors have been proposed as low-toxicity agents to treat double strand break (DSB)-repair defective tumors. Several findings imply the potential relevance of DSB repair defects in the tumorigenesis of esophageal squamous cell carcinoma (ESCC). We evaluated the effect of a PARP Inhibitor (AZD2281) on the TE-series ESCC cell lines. Of these eight cell lines, the clonogenic survival of one (TE-6) was reduced by AZD2281 to the level of DSB repair-defective Capan-1 and HCC1937 cells. AZD2281-induced DNA damage was implied by increases in γ-H2AX and cell cycle arrest at G2/M phase. The impairment of DSB repair in TE-6 cells was suggested by a sustained increase in γ-H2AX levels and the tail moment calculated from a neutral comet assay after X-ray irradiation. Because the formation of nuclear DSB repair protein foci was impaired in TE-6 cells, whole-exome sequencing of these cells was performed to explore the gene mutations that might be responsible. A novel mutation in RNF8, an E3 ligase targeting γ-H2AX was identified. Consistent with this, polyubiquitination of γ-H2AX after irradiation was impaired in TE-6 cells. Thus, AZD2281 induced growth retardation of the DSB repair-impaired TE-6 cells. Interestingly, a strong correlation between basal expression levels of γ-H2AX and sensitivity to AZD2281was observed in the TE-series cells (R^2 = 0.5345). Because the assessment of basal DSB status could serve as a biomarker for selecting PARP inhibitor-tractable tumors, further investigation is warranted.

Esophageal carcinoma is the sixth most common cause of cancer-related deaths worldwide and is associated with a poor prognosis.[1] Surgical therapies of resectable esophageal cancer exhibit a 5-year survival rate ranging from 20% to 27%.[2–4] Less invasive therapies that preserve the esophagus have also been introduced. Endoscopic therapies, such as endoscopic mucosal resection (EMR) and endoscopic submucosal dissection (ESD), have been adopted for early esophageal cancer and have achieved favorable outcomes, but postoperative esophageal stricture frequently occurs after these treatments. Furthermore, intensive follow-up is necessary to manage new heterochronous lesions.[5–7] Chemoradiotherapy (CRT), which combines radiation, 5-fluorouracil (5-FU) and cisplatin (CDDP), is a promising therapeutic alternative to esophagectomy with a survival rate equivalent to that of surgical therapies.[8,9] However, the acute and late adverse effects of chemoradiotherapy, including pancytopenia and pneumonitis, still require consideration. There is a demand for effective molecular target drugs for esophageal cancer that combine an improved therapeutic efficacy with fewer adverse effects.

Poly(ADP-ribose) polymerase (PARP) inhibitors induce the accumulation of DNA single-strand breaks (SSB), which cause the formation of DNA double-strand breaks (DSB) after the stalling and collapse of progressing DNA replication folks.[10] Though DSBs are repaired by the error-free homologous recombination repair (HRR) pathway in non-tumor cells, they remain unrepaired and induce lethality in HRR-defective tumor cells.[11,12] Based on this mechanism, PARP inhibitors have been proposed as low toxicity agents for HRR-defective tumors. BRCA1 and BRCA2 are key components of the HRR machinery, and the abnormality of these genes is known to cause sporadic and hereditary breast and ovarian cancers.[13] Consistent with this, PARP inhibitors have been developed for breast and ovarian cancers. In addition, an increasing number of biomarker candidates that predict the sensitivity of a tumor to PARP inhibitors have been reported.[14–18]

Esophageal carcinoma is histologically classified into squamous cell carcinoma (ESCC) and adenocarcinoma; the former is common in East Asia. Although the direct relevance has not been well investigated, several findings suggest that a defect in the HRR pathway contributes to the tumorigenesis of ESCC. The risk of esophageal and head and neck squamous carcinoma is increased among Fanconi anemia (FA)

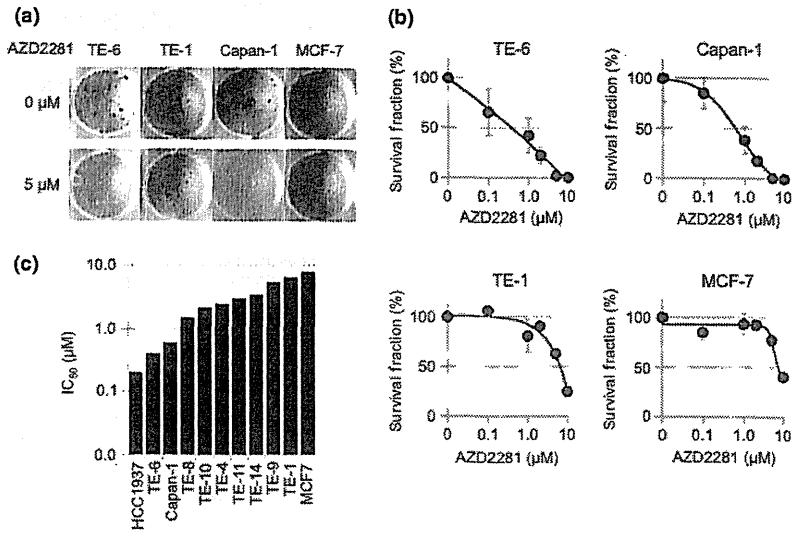

Fig. 1. Sensitivity of TE-series cell lines to a poly (ADP-ribose) polymerase (PARP) inhibitor, AZD2281. (a) TE-1, TE-6, Capan-1 (BRCA2-deficient) and MCF7 (wild-type BRCA) cells were treated with or without AZD2281 at the indicated concentrations for at least seven doubling times. Cells were fixed and stained with crystal violet and the number of colonies was counted. (b) Sensitivity to AZD2281 was evaluated by clonogenic assay. Colonies consisting of more than 64 cells were counted and the survival fraction was estimated. Three independent experiments were carried out. The data represent the average and standard deviations. (c) IC_{50} (μM) value of AZD2281 in eight TE series, MCF7, HCC1937 and Capan-1 cells measured by the clonogenic assays.

patients whose HRR pathway was disturbed due to FA-predisposing gene alterations.[19,20] In addition, recently reported whole-exome sequencing data from 74 head and neck SCCs revealed that more than half of SCC cases harbored mutations in genes involved in DNA repair.[21] Therefore, we assume that some fraction of ESCCs harbor DSB repair defects and might be favorable targets of PARP inhibitors. The aim of this study was to examine the efficacy of a potent PARP-1 inhibitor in a series of ESCC cell lines established from Japanese patients.

Materials and Methods

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

Purchased materials. A PARP inhibitor, AZD2281 (Olaparib) and BSI-201 (Iniparib) were purchased from Selleck Chemicals (Houston, TX, USA). The TE-1, TE-4, TE-6, TE-8, TE-9, TE-10, TE-11 and TE-14 cell lines were purchased from the Riken BioResource Center (Tsukuba, Japan). The Capan-1, HCC1937, MDA-MB-436 and MCF-7 cell lines were purchased from the American Type Culture Collection (ATCC, Manassas, VA, USA).

Clonogenic assays. A total of 500–2000 cells were cultured with AZD2281- or vehicle-containing media. After 10–16 days, cells were fixed and stained with crystal violet. Colonies consisting of more than 64 cells were subsequently counted.

Immunoblotting analysis. The treated cell lysates were separated by 15% SDS-PAGE and the blot was hybridized with the

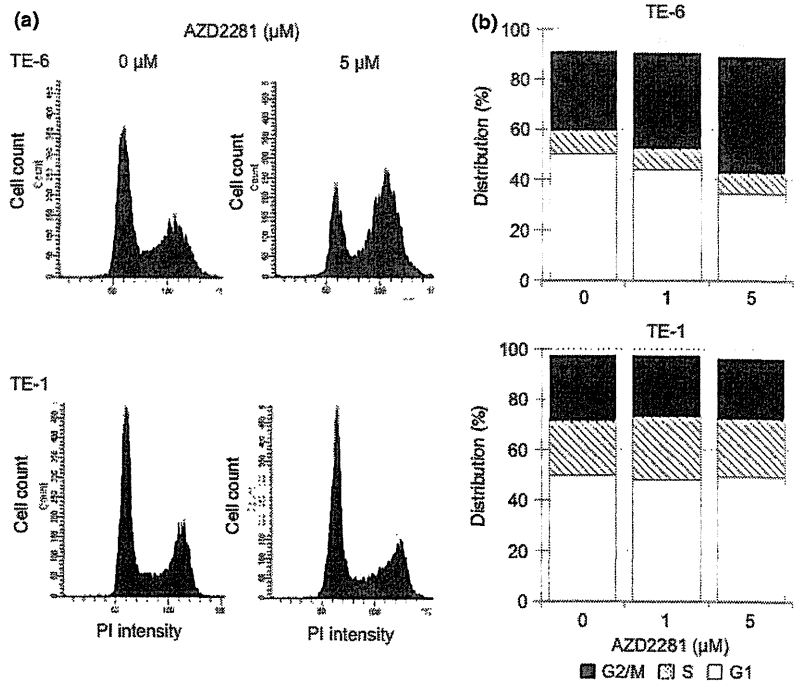

Fig. 2. AZD2281-induced G2/M arrest in TE-6 cells. (a) TE-6 and TE-1 cells were cultured with or without AZD2281 for 24 h. DNA ploidy was assessed by propidium iodide (PI) staining and flow cytometry. (b) The proportion of estimated cell-cycle phases in TE-6 and TE-1 cells treated with or without AZD2281. The data represent the average of three independent experiments.

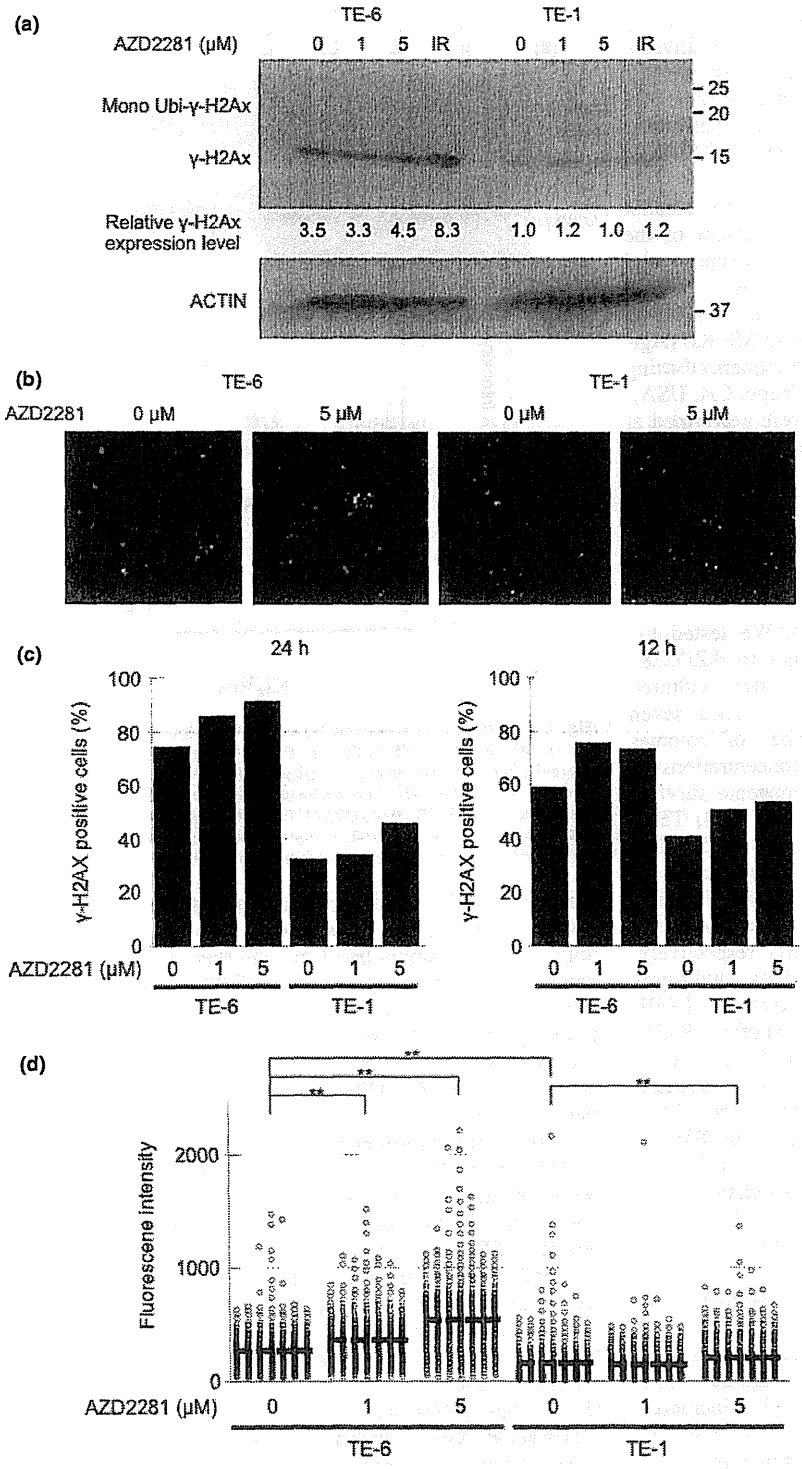

Fig. 3. Increase in double strand breaks (DSBs) in TE-6 cells treated with AZD2281. (a) TE-6 and TE-1 cells were treated with AZD2281 for 24 h and with 5 Gy X-ray irradiation, and γ-H2AX was assessed using Western blotting. The anti-γ-H2AX antibody detected both unubiquitinated (15 kD) and mono-ubiquitinated (23.6 kD) γ-H2AX. (b) TE-6 and TE-1 cells were treated with AZD2281 for 24 h and γ-H2AX was assessed by immunofluorescence. DAPI (blue) and γ-H2AX (red) images were superimposed. (c) Number of the γ-H2AX-positive TE-6 and TE-1 cells treated with or without AZD2281 at the indicated concentrations for 24 h. (d) Scatter diagrams show the fluorescence intensity of individual TE-6 and TE-1 cells treated with or without AZD2281 at the indicated concentrations for 24 h. The lines shown indicated the averages of the data plotted. The data were obtained from at least 500 cells for each condition. $^{**}P < 0.01$ (Student's *t*-test).

phospho-Histone H2A.X (Ser139) (20E3) rabbit monoclonal antibody (1:1000; Cell Signaling Technology, Danvers, MA, USA) and then with a HRP-conjugated secondary antibody (1:50000; Santa Cruz Biotechnology, Dallas, TX, USA). Densitometric analysis was performed with Image-J software (http://imagej.nih.gov/ij/).

Immunofluorescence analysis. To evaluate the formation of DSBs, cells grown on 96-well plates were treated with an anti-γ-H2AX rabbit monoclonal antibody (Cell Signaling Technology) followed by a goat anti-Mouse IgG (H + L) DyLight 549-conjugated secondary antibody (Thermo Scientific, Waltham, MA, USA). γ-H2AX was observed under an ArrayScan HCS System (Thermo Scientific). To evaluate the formation of 53BP1 and RAD51 nuclear foci, cells grown on μ-Dish$^{35\ mm,\ low}$ (ibidi) were treated with an anti-53BP1 rabbit polyclonal antibody (Abcam, Cambridge, UK) or an anti-RAD51 rabbit

polyclonal antibody (Santa Cruz Biotechnology) followed by an Alexa Flour 488 donkey anti-rabbit IgG (H + L) (Invitrogen, Carlsbad, CA, USA). These foci were observed under a BIOREVO BZ-9000 microscope (Keyence, Osaka, Japan).

Neutral comet assays. Neutral comet assays were performed using the CometAssay kit (Trevigen, Gaithersburg, MD, USA) according to the manufacturer's instructions.

Cell cycle analysis. The cells were fixed with 70% ethanol and stained with propidium iodide. The DNA content of the cells was evaluated using a FACS CantoII flow cytometer and FACS Diva software (BD, Franklin Lakes, NJ, USA).

Whole-exome sequencing. Targeted enrichment was performed using the SureSelect Human All Exon 50 Mb Kit (Agilent Technologies, Santa Clara, CA, USA) and sequenced using Illumina Genome Analyzer IIx (Illumina, San Diego, CA, USA).

Statistical analysis. Individual experiments were performed at least in triplicate. The statistical significance of observed differences was analyzed using Student's *t*-test. One asterisk (*) indicates a *P*-value smaller than 0.05. Two asterisks (**) indicate a *P*-value smaller than 0.01.

Results

Sensitivity of TE series cell lines to AZD2281. We tested the sensitivity of the TE series of ESCC cell lines to AZD2281 using clonogenic assays. All of the cell lines were cultured with various concentrations of AZD2281 for at least seven doubling times (10–16 days), and the number of colonies with more than 64 cells was counted. The concentrations of AZD2281 that caused a 50% reduction in clonogenic survival (IC_{50}) in the TE-6, TE-8, TE-10, TE-4, TE-11, TE-14, TE-9, and TE-1 cells were 0.4, 1.5, 2.2, 2.5, 3.0, 3.5, 5.5 and 6.5 μM, respectively (Figs 1a–c, S1). The IC_{50} values for HCC1937 and Capan-1 cells, which have deletions or mutations in the BRCA genes and are reported to be sensitive to PARP inhibitors,[22,23] were 0.2 and 0.6 μM, respectively. The IC_{50} values for MCF7 cells, which have wild-type BRCA genes and are known to be resistant to PARP inhibitors, was 8.0 μM. We further tested another PARP inhibitor, BSI-201. The IC_{50} values for TE-6 and TE- 1 cells were 9.6 and 22.0 μM, respectively (Fig. S2). Because AZD2281 and BSI-201 suppressed the growth of the TE-6 cells as efficiently as it suppressed the growth of the BRCA-deficient, PARP-inhibitor-sensitive cell lines and failed to suppress the growth of the TE-1 cells; we designated TE-6 as a PARP inhibitor-sensitive ESCC cell line and TE-1 as a PARP inhibitor-resistant cell line. We selected these cells for further analyses.

AZD2281-induced G2/M arrest in TE-6 cells. To further study the mechanism of growth retardation of TE-6 cells by AZD2281, the status of the cell cycle in these cells was assessed by analyzing the DNA ploidy pattern (Figs 2a,b, S3). Treatment with 1 and 5 μM of AZD2281 for 12 h increased the population with 4n DNA content from 36.7% to 40.7% and 40.6%, respectively. Treatment for 24 h further increased the population with 4n DNA content from 31.6% to 37.9% and 46.3%, respectively. This suggested an increase in the G2 /M or M arrested population after AZD2281 treatment. On the other hand, no significant increase of tetraploid cells was observed in the TE-1 cells.

Increase of DSBs in TE-6 cells treated with AZD2281. To determine whether DSBs are formed after treatment with AZD2281 for 24 h, we assessed the amount of γ-H2AX as a marker of DSBs. Western blotting revealed that the level of γ-H2AX staining in TE-6 cells increased in a dose-dependent manner. However, no such increase was observed in TE-1 cells (Fig. 3a). The same trend was observed by immunofluorescence. Both the percentage of γ-H2AX positive cells determined by visual inspection and the average fluorescence intensity of γ-H2AX staining per cell increased significantly in a dose-dependent manner in TE-6 cells, but not in TE-1 cells, suggesting that AZD2281 induced an accumulation of DNA damage in TE-6 cells (Fig. 3b–d).

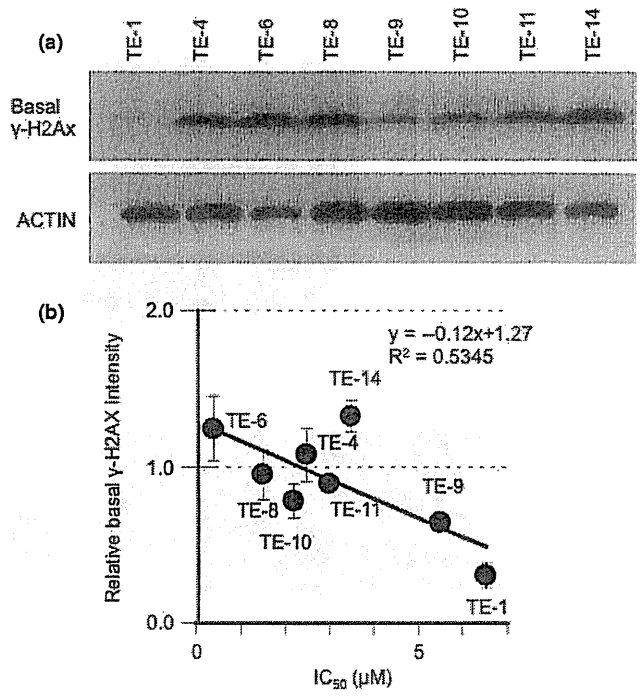

Fig. 4. A strong correlation between base-level γ-H2AX and sensitivity to AZD2281 of TE cells. (a) Eight non-treated TE-series cell lines were subjected to Western blot analysis with antibodies against γ-H2AX and actin. (b) The correlation between basal γ-H2AX expression levels and IC50 of AZD2281. The average intensity of γ-H2AX was standardized with actin. The data represent the averages and standard deviations of three independent experiments.

A strong correlation between base-level γ-H2AX and sensitivity to AZD2281 of TE cells. The Western blotting and immunofluorescence data also suggested that the baseline γ-H2AX level was higher in the AZD2281-sensitive TE-6 cells than in the resistant TE-1 cells (Fig. 3a). Because the increased amount of γ-H2AX may suggest the accumulation of DSBs, we evaluated the correlation between the basal expression levels of γ-H2AX and sensitivity to AZD2281 among the eight TE cell lines. A significant correlation between the basal levels of γ-H2AX and the IC_{50} of AZD2281 was observed (R^2 = 0.5345) (Fig. 4a,b).

Sustained X-ray irradiation-induced DSBs in TE-6 cells. To assess whether the impairment of DSB repair is relevant to the sensitivity of TE-series cells to AZD2281, we evaluated X-ray

Table 1. Tail moment of the X-ray-irradiated TE-6 and TE-1 cells

	Time after irradiation			
	0 h	15 min	2 h	6 h
TE-6	2.43 ± 3.19	5.27 ± 4.35	9.78 ± 6.27	13.7 ± 8.20
TE-1	3.69 ± 4.14	5.56 ± 3.97	5.20 ± 4.22	3.20 ± 4.71