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		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

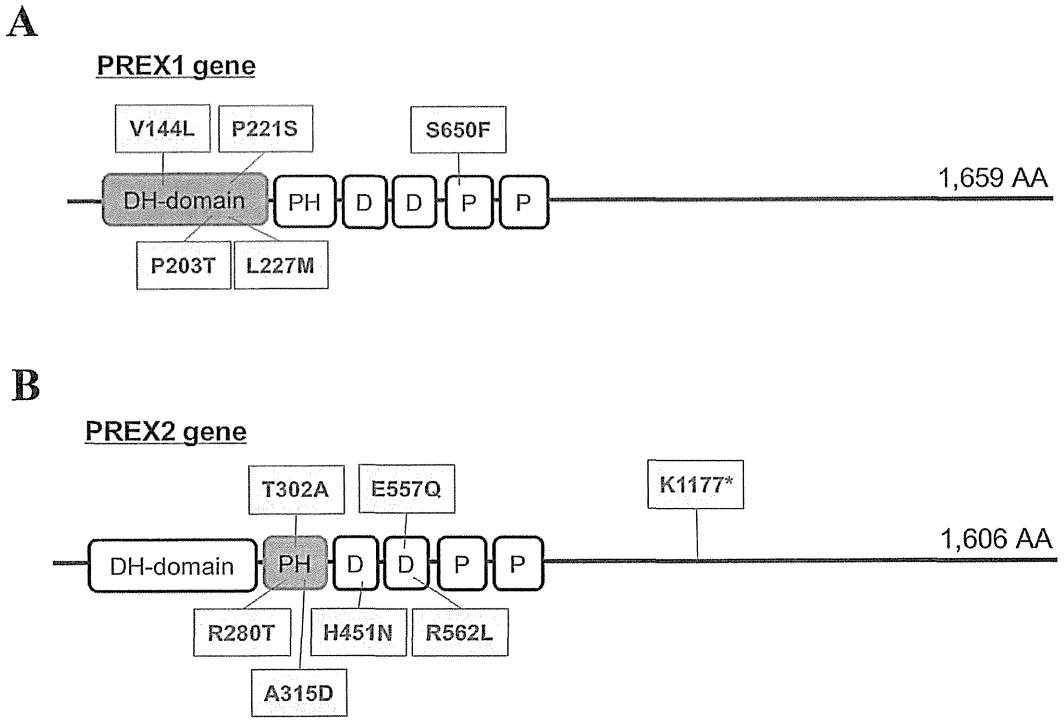

Fig. 1

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

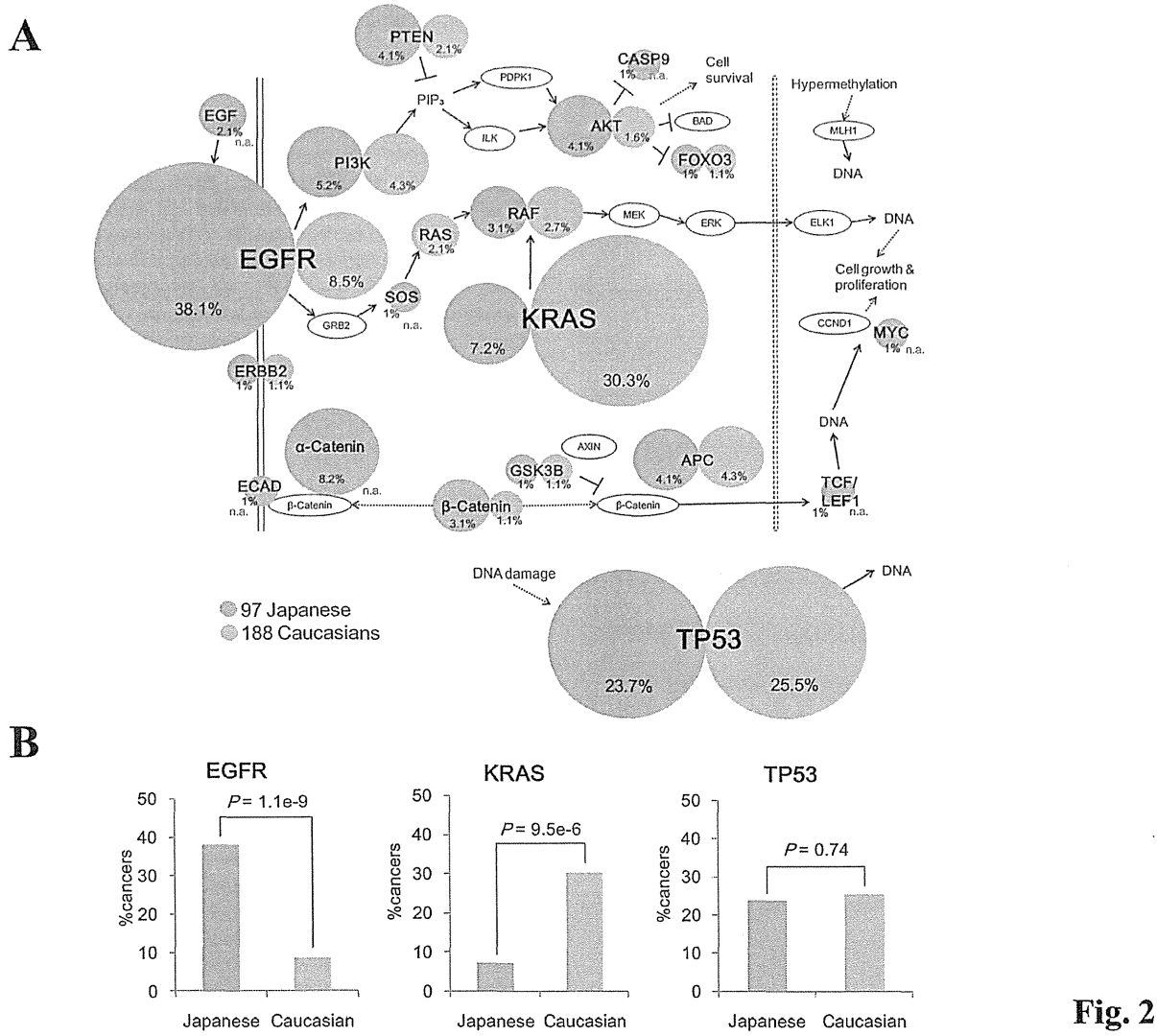

Fig. 2

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

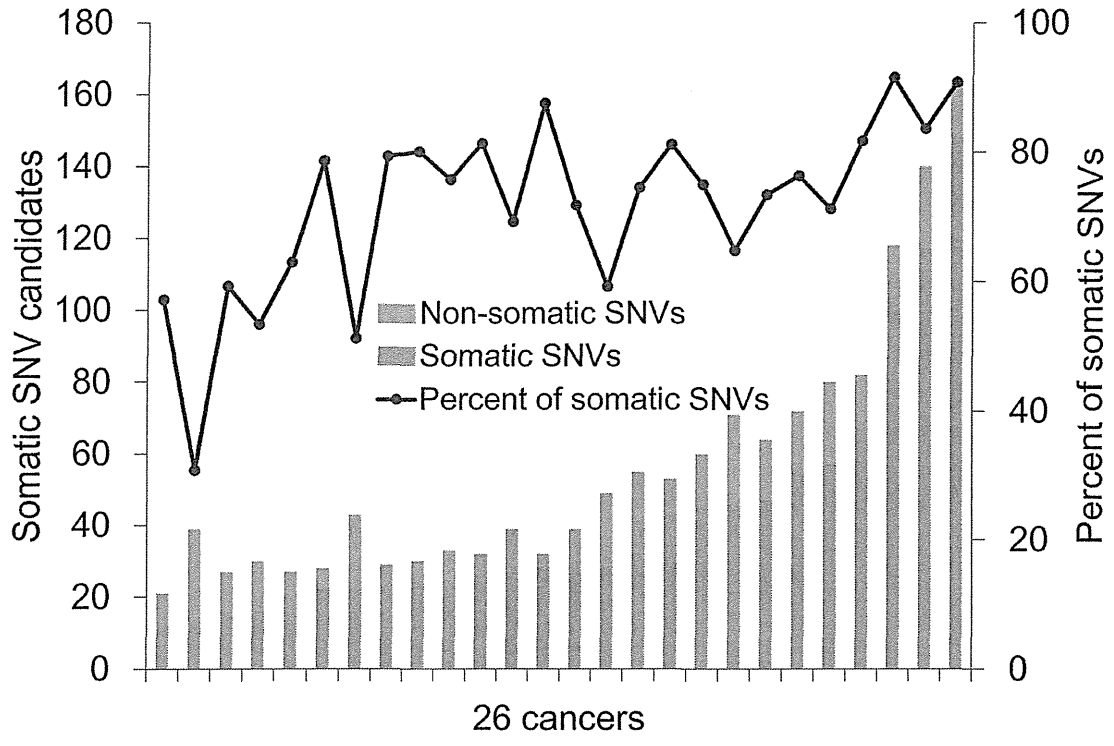

Fig. 3

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.

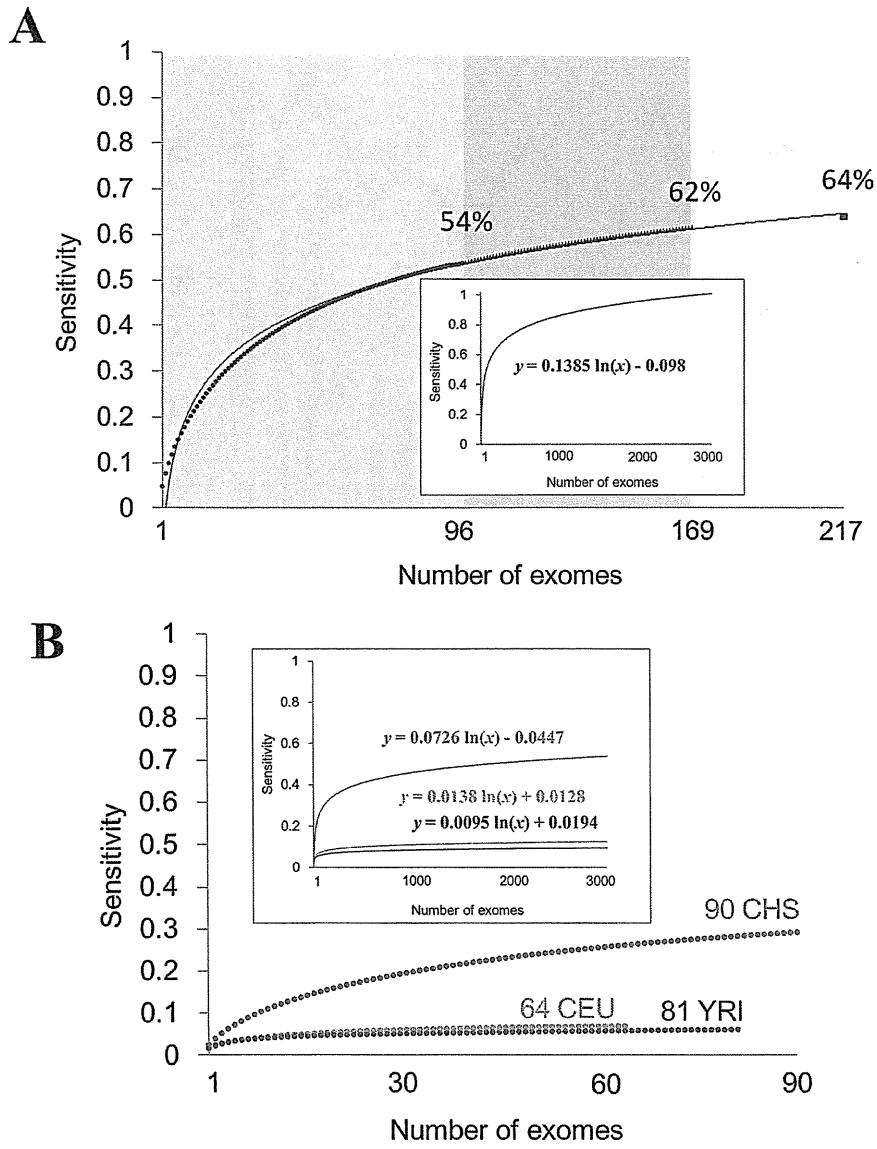

Fig. 4

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

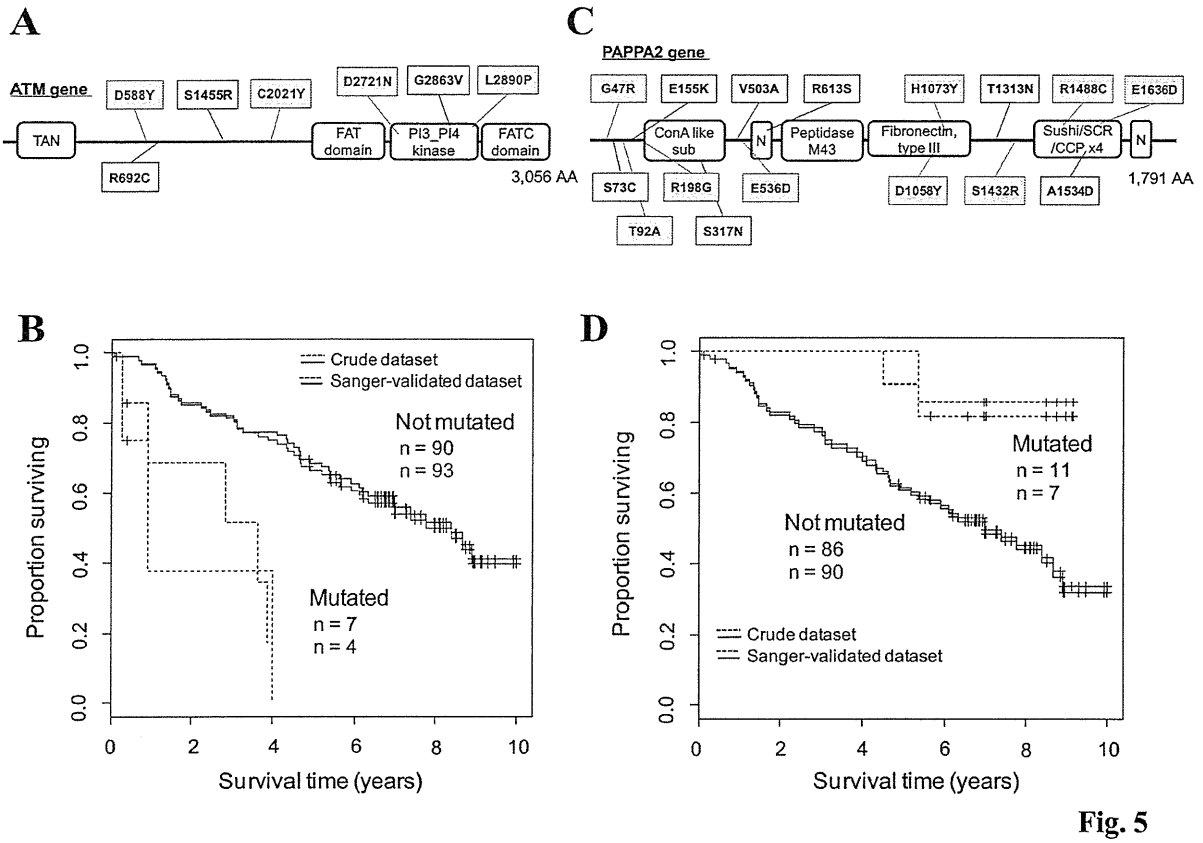

Fig. 5

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).

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
The official journal of the Japanese Cancer Association JCA

Review Article

RET fusion gene: Translation to personalized lung cancer therapy

Takashi Kohno,[1,2,5] Koji Tsuta,[3] Katsuya Tsuchihara,[1] Takashi Nakaoku,[2] Kiyotaka Yoh[4] and Koichi Goto[4]

[1]Division of Translational Research, Exploratory Oncology Research & Clinical Trial Center (EPOC), National Cancer Center, Tokyo; [2]Division of Genome Biology, National Cancer Center Research Institute, Tokyo; [3]Division of Pathology and Clinical Laboratories, National Cancer Center Hospital, Tokyo; [4]Division of Thoracic Oncology, National Cancer Center Hospital East, Kashiwa, Japan

(Received July 2, 2013/Revised July 27, 2013/Accepted August 21, 2013/Accepted manuscript online August 30, 2013/Article first published online October 1, 2013)

Development of lung adenocarcinoma (LADC), the most frequent histological type of lung cancer, depends in many cases on the activation of "driver" oncogenes such as *KRAS*, epidermal growth factor receptor (*EGFR*), and anaplastic lymphoma kinase (*ALK*). Inhibitors that target the EGFR and ALK tyrosine kinases show therapeutic effects against LADCs containing *EGFR* gene mutations and *ALK* gene fusions, respectively. Recently, we and others identified the *RET* fusion gene as a new targetable driver gene in LADC. The RET fusions occur in 1–2% of LADCs. Existing US Food and Drug Administration-approved inhibitors of RET tyrosine kinase show promising therapeutic effects both *in vitro* and *in vivo*, as well as in a few patients. Clinical trials are underway to investigate the therapeutic effects of RET tyrosine kinase inhibitors, such as vandetanib (ZD6474) and cabozantinib (XL184), in patients with *RET* fusion-positive non-small-cell lung cancer. (*Cancer Sci* 2013; 104: 1396–1400)

Personalized Therapy of LADC

Lung cancer is the leading cause of cancer-related mortality worldwide. Lung adenocarcinoma (LADC) is the most frequent type of lung cancer. LADC occurs both in smokers and non-smokers, and its incidence is increasing.[1] Genome analyses of LADC show that these tumors contain distinct genetic alterations that activate oncogenes.[2,3] Genetic alterations that result in the activation of several oncogenes are detected in a mutually exclusive manner (Fig. 1); of the hundreds of genes mutated in each case of LADC, these oncogenes are considered to be "driver genes".[4] Remarkably, molecular targeted therapy using inhibitory drugs against activated oncogene products has begun to replace conventional chemotherapy using cytotoxic drugs, even for first-line use.[2]

The epidermal growth factor receptor (*EGFR*) gene is activated by single amino acid substitution mutations or in-frame amino acid deletion mutations in 10–20% of LADC cases in the USA and in 30–40% of cases in East Asia.[2] Tumors harboring these *EGFR* mutations respond to EGFR tyrosine kinase inhibitors (TKIs) such as erlotinib and gefitinib, thereby improving progression-free survival and quality of life.[5,6] In addition, 3–5% of LADC harbor fusions that result in the activation of the anaplastic lymphoma kinase (*ALK*) gene; such mutations are mutually exclusive with *EGFR* mutations. Inhibitors, such as crizotinib, that target ALK tyrosine kinase show marked therapeutic effects against ALK fusion-positive LADCs.[7–9] These results indicate that personalized therapy for LADC using TKIs selected on the basis of somatic genetic alterations has been realized already; indeed, 20% of USA/European and 40% of Asian LADC patients benefit from such therapies.

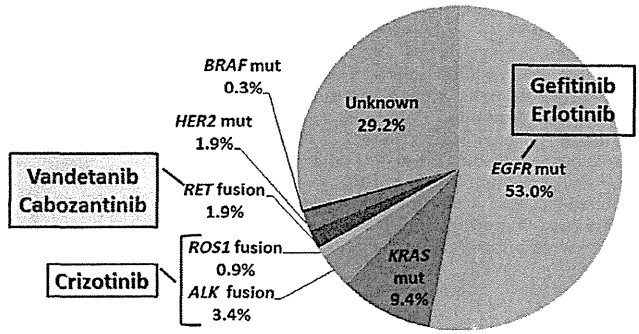

Fig. 1. Pie chart showing the fraction of Japanese lung adenocarcinoma patients that harbor "driver" gene mutations. Surgical specimens from 319 stage I–II lung adenocarcinomas deposited in the National Cancer Center Biobank (Japan) were subjected to analysis. The *EGFR*, *KRAS*, *BRAF*, and *HER2* mutations (mut) were examined using the high resolution melting method, whereas *ALK*, *ROS1* and *RET* fusions were examined by RT-PCR.[12,31] The protocol for this research project has been approved by the institutional review board of the National Cancer Center.

Discovery of the *RET* Fusion Gene as a New Targetable Driver Gene

In 2012, four studies, including one by our group, identified fusions of the *RET* (rearranged during transfection) oncogene[10–13] (Fig. 2). *RET* is a well-known driver oncogene kinase for thyroid cancer, and both activating mutations and fusions of this gene have been observed.[14,15] Germline gain-of-function mutations in *RET* predispose carriers to multiple endocrine neoplasia type 2, which is characterized by medullary thyroid cancer, pheochromocytoma, and hyperparathyroidism, and also to familial medullary thyroid carcinoma syndrome. Somatic gain-of-function *RET* mutations have been observed in 30–50% of sporadic medullary thyroid cancer, and somatic *RET* gene fusions have been observed in 30–50% of sporadic papillary thyroid cancer. The US Food and Drug Administration (FDA) have approved two inhibitory drugs, vandetanib (ZD6474) and cabozantinib (XL184), for the treatment of advanced medullary thyroid cancer. The molecular process for generating a *RET* fusion is similar to the mechanism underlying *ALK* fusion: the most frequent *RET*

[5]To whom correspondence should be addressed.
E-mail: tkkohno@ncc.go.jp

doi: 10.1111/cas.12275

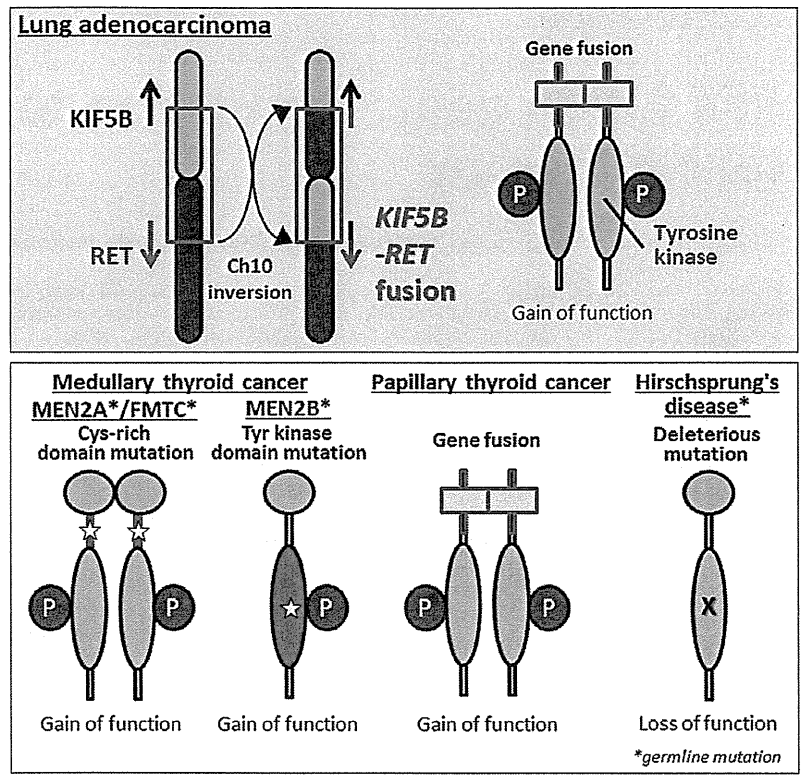

Fig. 2. Involvement of the *RET* gene in lung and thyroid carcinogenesis and in a developmental disorder. Upper panel, somatic inversion in chromosome 10 results in *KIF5B–RET* fusions. The RET fusion protein has constitutive tyrosine (Tyr) kinase activity, representing a gain-of-function alteration. Lower panel, *RET* alterations in other diseases. A germline gain-of-function mutation of *RET* drives thyroid carcinogenesis in patients with multiple endocrine neoplasia type 2 (MEN2). Somatic gain-of-function mutation and translocation of *RET* cause medullary and papillary thyroid cancers, respectively. Germline loss-of-function *RET* mutations cause Hirschsprung's disease, a hereditary disorder characterized by the absence of enteric ganglia in variable segments of intestine. FMTC, familial medullary thyroid carcinoma; P, phosphorylation; X, inactivating mutation.

fusion, *KIF5B–RET*, is generated by a pericentric inversion in chromosome 10, whereas the most frequent *ALK* fusion, *EML4–ALK*, is generated by a paracentric inversion in chromosome 2 (Fig. 2).

Four different strategies resulted in the discovery of the same *RET* fusion gene (Table 1, Fig. 3). We carried out whole-transcriptome sequencing using RNA from 30 snap-frozen surgical LDAC specimens to identify novel fusion-gene transcripts.[12] Ju *et al.*[13] analyzed the whole genome and transcriptome of a single young (33-year-old) LADC patient. Lipson *et al.*[11] carried out targeted-capture sequencing of 145 cancer-relevant genes from genomic DNA obtained from 24 formalin-fixed paraffin-embedded tumor samples to identify genes mutated or fused in LADC. Takeuchi *et al.*[10] carried out a FISH-based screen against known fusion kinase and partner genes to detect rearrangement of oncogenes in >1500 LADC cases.

To date, *RET* fusions have been identified that involve four fusion partners comprising nine subtypes of fusion variants: *KIF5B*, *CCDC6/PTC/H4*, *NCO4/PTC3/ELE1*, and *TRIM33 /PTC7*.[16] The latter three partners are also fused to *RET* in thyroid cancer, whereas *KIF5B* is not. The deduced features of the proteins encoded by all types of *RET* fusion gene are similar to those of ALK: coiled-coil domains in the N-terminal fusion partners cause the RET domains to dimerize, resulting in activation of RET tyrosine kinase in the absence of ligands (Fig. 2). The ligand-independent dimerization and constitutive activation of RET protein are also caused by gain-of-function mutations and translocations of *RET,* which have been detected in sporadic and hereditary thyroid cancers.[15] In fact, autophosphorylation of the KIF5B–RET fusion protein, representing RET protein activation, was observed in LADC tissues harboring the corresponding *RET* fusion gene,[12] as well as in cells cultured in the absence of serum. The transforming and signal-addictive activities of KIF5B–RET fusion proteins are suppressed by FDA-approved drugs (e.g., vandetanib, sorafenib, and sunitinib), which themselves suppress RET kinase.[10–12] In addition, the LADC cell line, LC-2/ad, which harbors a *CCDC6–RET* fusion, is sensitive to these drugs both *in vitro* and *in vivo*.[17,18] Unfortunately, these drugs are not approved for use as treatments for lung cancer; however, the existing data led us to investigate their therapeutic effects in clinical trials, as described below.

Prevalence and Characteristics of *RET* Fusion-Positive LADC

Several studies have validated the presence of *RET* fusion in a small subset of non-small-cell lung cancers (NSCLCs).[16,19–24] The total number of examined cases has reached approximately 5000 (Table 1). Most of the positive cases are LADC, but several cases involve other histological types of NSCLC, such as adenosquamous carcinoma.[19,20] The *RET* fusions are present in 1–2% of NSCLC/ADC of patients of both Asian and European descent. Several studies indicate that *RET* fusion occurs preferentially in young, never-smoker, and light-smoker patients.[10,12,20]

The LADCs harboring *KIF5B–RET* fusions are well or moderately differentiated, similar to LADCs harboring *EGFR* mutations. This is in contrast to *EML4–ALK* fusion-positive LADCs, which tend to show signet-ring and mucinous cribriform patterns.[10] Those LADCs harboring *CCDC6–RET* fusions show such histological features.[10,18]

In our previous study, we did not detect *RET* fusions in a screen of 234 squamous cell, 17 large cell, and 20 small-cell lung cancers.[12] Adenocarcinomas of other organs, such as colon (n = 200) and ovary (n = 100), were also negative for *RET* fusion. To date, whole-transcriptome analysis of other organs has not identified *RET* fusions in cancers outside the lung. Therefore, *RET* fusion may occur mainly in LADC and papillary thyroid cancer.

Table 1. Prevalence of *RET* gene fusion in non-small-cell lung cancer (NSCLC)

Institution	No. of cases examined	No. of *RET* fusion (+) cases	*RET* fusion%	Fusion type	Ref.
	NSCLC/lung adenocarcinoma				
National Cancer Center, Japan	704/433	7/7	1.0/1.6	*KIF5B–RET*: 7	12
Japan Foundation for Cancer Research, Japan	1482/1119	13/13	0.9/1.2	*KIF5B–RET*: 12 *CCDC6–RET*: 1	10
Foundation Med, USA	643/561	12/12	1.8/2.1	*KIF5B–RET*: 12	11
Seoul National University, Korea	21/21 (Driver mutation –)	3/3	14/14	*KIF5B–RET*: 3	13
Chinese Academy of Sciences, China	202/202 (Driver mutation –)	2/2	1.0/1.0	*CCDC6–RET*: 2	24
Nagoya City University, Japan	371/270	3/3	0.8/1.1	*KIF5B–RET*: 3	23
Memorial Sloan-Kettering Cancer Center, USA	69/69 (Driver mutation –)	1/1	1.4/1.4	*KIF5B–RET*: 1	21
Fudan University Shanghai Cancer Center, China	936/633	13/11	1.4/1.7	*KIF5B–RET*: 9 *CCDC6–RET*: 3 *NCOA4–RET*: 1	20
Tongji University School of Medicine, China	392/231	6/4	1.5/1.7	*KIF5B–RET*: 6	19
Korea Research Institute of Bioscience and Biotechnology, Korea	6/6 (Female non-smoker)	1/1	17/17	*CCDC6–RET*: 1	22
Memorial Sloan-Kettering Cancer Center, USA	31/31 (Driver mutation –)	5/5	16/16	*KIF5B–RET*: 2 *TRIM33–RET*: 1 (Unknown: 2)	16
Total	4857/3576	66/62	1.4/1.8	*KIF5B–RET*: 55 *CCDC6–RET*: 7 *NCOA4–RET*: 1 *TRIM33–RET*: 1	

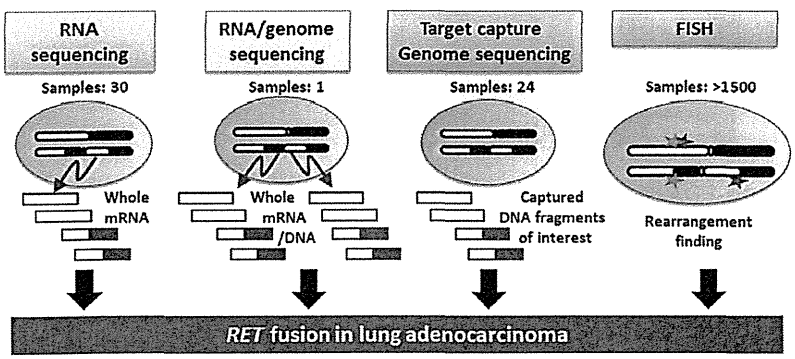

Fig. 3. Strategies used to identify *RET* fusion in lung adenocarcinoma. Four different methods were used to identify novel oncogenic fusions in lung adenocarcinomas.[10–13]

Therapeutic Effects of RET TKIs in Patients with *RET* Fusion-Positive NSCLC

In clinical trials, the ALK TKI, crizotinib, showed a dramatic therapeutic effect against NSCLCs harboring *ALK* gene fusions. Crizotinib was approved for use in the USA in August 2011 and for use in Japan in March 2012.[8] Considering that the *ALK* gene fusion was first identified in NSCLC in 2007, approval has been achieved extremely rapidly. Consequently, the discovery of the *RET* fusion has raised expectations that patients with NSCLCs harboring *RET* fusions will soon benefit from targeted therapy using existing RET TKIs.

Several commercially available multikinase inhibitors, such as vandetanib (ZD6474), cabozantinib (XL184), sorafenib, sunitinib, lenvatinib (E7080), and ponatinib (AP24534), have activity against the RET kinase; however, no selective RET inhibitors have yet been developed for clinical use. Several phase II clinical trials have been initiated to investigate the therapeutic effects of such multikinase inhibitors in patients with advanced *RET* fusion-positive NSCLC (Table 2). As for previous clinical trials of ALK TKIs, all of these trials have open-label and single-arm designs, with response rate as the primary endpoint. One study, carried out by Drilon *et al.* at the Memorial Sloan-Kettering Cancer Center (NCT01639508),

doi: 10.1111/cas.12275

Table 2. Ongoing phase II clinical trials of RET tyrosine kinase inhibitors in patients with *RET* fusion-positive non-small-cell lung carcinoma

Trial number†	Drug (pharmaceutical company)	Study design	Primary end-point	Enrolment no.	Study start
NCT01639508	Cabozantinib/XL184 (Exelixis)			25	July 2012
UMIN000010095	Vandetanib/ZD6474 (AstraZeneca)			17	Feb 2013
NCT01823068	Vandetanib/ZD6474 (AstraZeneca)	Open-label, single arm	Response rate	17	April 2013
NCT01877083	Lenvatinib/E7080 (Eisai)			20	April 2013
NCT01813734	Ponatinib/AP24534 (ARIAD)			20	June 2013

†Detailed information is available at http://clinicaltrials.gov/ or https://upload.umin.ac.jp.

Table 3. Response of lung adenocarcinoma patients to RET tyrosine kinase inhibitors

Patient	*RET* fusion gene	Inhibitor	Ethnicity	Sex	Age, years	Pathological diagnosis	Smoking history (pack-year)	Response (% decrease)	Reference
1	*TRIM33–RET*	Cabozantinib	Caucasian	Female	41	Papillary adenocarcinoma	Never-smoker	Partial response (66)	16
2	*KIF5B–RET*	Cabozantinib	African-American	Female	75	Poorly differentiated adenocarcinoma	Never-smoker	Partial response (32)	16
3	*KIF5B–RET*	Cabozantinib	Caucasian	Female	68	Mixed subtype adenocarcinoma	Never-smoker	Stable disease	16
4	*KIF5B–RET*	Vandetanib	Caucasian	Male	58	Poorly differentiated adenocarcinoma	Former smoker (5)	Decrease in size	26

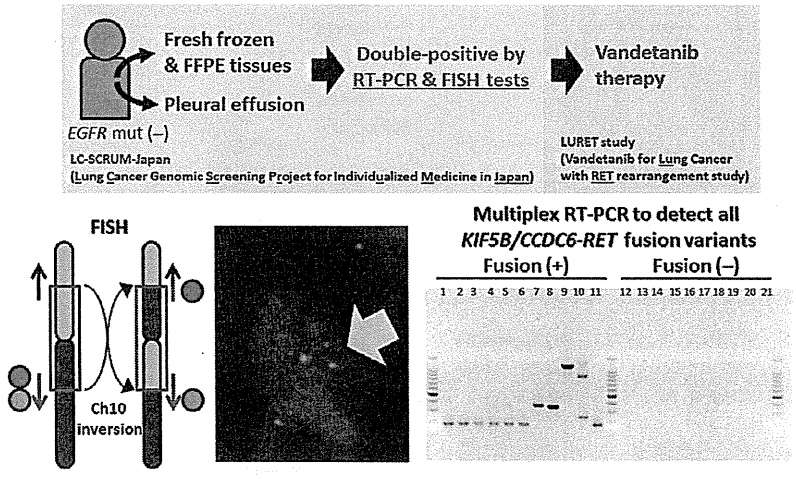

Fig. 4. Consolidated Standards of Reporting Trials diagram of the Lung Cancer Genomic Screening Project for Individualized Medicine in Japan (LC-SCRUM) and the Lung Cancer with *RET* rearrangement (LURET) study in Japan. The LC-SCRUM screen identified 17 *RET* fusion-positive cases from non-squamous non-small-cell lung carcinoma cases without epidermal growth factor receptor (*EGFR*) mutations (mut). The *RET* fusion-positive cases are defined as being positive in both RT-PCR and subsequent FISH tests. Representative pictures of these tests are shown. Fusion-positive cases were treated with vandetanib in the LURET study. Ch10, chromosome 10; FFPE, formalin-fixed paraffin-embedded.

is testing cabozantinib, a drug recently approved by the FDA for the treatment of thyroid cancer. The therapeutic responses of the first three patients to be treated with cabozantinib were reported to be promising (Table 3).[16]

The other phase II clinical trial was initiated by our own group in Japan (UMIN00001009). This trial, designated LURET (Lung Cancer with *RET* rearrangement study), is investigating the therapeutic effects of vandetanib in 17 patients with *RET* fusion-positive NSCLC (Table 2). Because vandetanib is a multikinase inhibitor that is effective against EGFR and vascular endothelial growth factor, this drug was previously examined for its therapeutic efficacy in advanced NSCLC patients in several "all-comer" clinical trials.[25] Those trials were carried out without considering gene alterations in determining eligibility, and the trials did not show significantly greater therapeutic effects than pre-existing therapeutic regimens. Therefore, only *RET* fusion-positive cases, which represent 1–2% of all NSCLCs, are eligible for the LURET study.

To evaluate eligibility for this study, we established a diagnostic method for detecting *RET* fusions using a combination of RT-PCR and FISH (Fig. 4). In this study, RNAs from frozen biopsy tissue or pleural effusion from patients with non-squamous NSCLCs without *EGFR* mutations are subjected to RT-PCR; this method enables us to detect all seven *KIF5B–RET* and *CCDC6–RET* variants identified to date.[16] The positive cases are then subjected to break-apart and fusion FISH to validate the RT-PCR results. Cases positive by both RT-PCR and FISH are eligible for the LURET study. The RT-PCR screening is being carried out in >100 hospitals throughout Japan by a consortium designated LC-SCRUM (Lung Cancer Genomic Screening Project for Individualized Medicine in Japan). The therapeutic results will be obtained within 2 years.

Notably, a recent study reported that one patient with LADC harboring a *KIF5B–RET* fusion responded to vandetanib (Table 3). The patient was Caucasian male and a former smoker. Tumor shrinkage was observed starting in the first week, and continued for 4 weeks.[26]

Perspective

The *RET* gene is predicted to be an additional therapeutic target for therapy against LADC. Three other oncogene kinases,

HER2 (activated by inflame insertion mutations), BRAF (activated by point mutation), and ROS1 (activated by gene fusion) are also promising targets for personalized therapy in addition to EGFR and ALK (Fig. 1). In fact, inhibition of these kinases has yielded therapeutic effects in several lung cancer patients. The LADCs harboring *HER2* mutations responded to therapy with anti-HER2 antibodies and HER2 TKIs.[27] One LADC case harboring a *BRAF* mutation responded to therapy with vemurafenib, an FDA-approved drug for the treatment of melanoma.[28] The ALK TKI, crizotinib, suppresses the activity of the ROS1 tyrosine kinase due to the high structural similarity between the ALK and ROS1 tyrosine kinase domains. Consistent with this, a significant portion of the LADC patients with *ROS1* fusions that were enrolled in a clinical trial responded to crizotinib.[29] Therefore, developing therapies that target RET and other kinases means that increasing numbers of LADC patients will benefit from personalized therapy (Fig. 1). Thus, LADC represents a type of cancer in which "precision cancer medicine"[30] based on somatic gene alterations will be realized.

Acquisition of drug resistance is a serious problem for therapies based on TKIs. The LADCs harboring ALK fusions become resistant to crizotinib by acquiring second-site mutations in the gatekeeper region of ALK tyrosine kinase.[7] Those LADCs harboring *ROS1* fusions also become resistant to crizotinib, in this case through second-site mutations in the gatekeeper region of ROS1.[29] Therefore, *RET* fusion-positive LADCs might also acquire resistance to RET TKIs through the same mechanism. Clinical trials of RET TKIs as a treatment for fusion-positive NSCLCs should be carried out carefully, and focus both on efficacy and the acquisition of resistance.

Acknowledgments

The authors thank all the collaborators in the National Cancer Center and the LC-SCRUM/LURET studies. This work was supported in part by: the Program for Promotion of Fundamental Studies in Health Sciences from the National Institute of Biomedical Innovation; Grants-in-Aid from the Ministry of Health, Labor, and Welfare for the Third-term Comprehensive 10-year Strategy for Cancer Control and for Research on New Drug and Medical Device Development; and the National Cancer Center Research and Development Fund. The National Cancer Center Biobank is supported by the National Cancer Center Research and Development Fund, Japan.

Disclosure Statement

The authors have no conflict of interest.

References

1 Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. *CA Cancer J Clin* 2010 (Sep–Oct); **60**: 277–300.
2 Oxnard GR, Binder A, Janne PA. New targetable oncogenes in non-small-cell lung cancer. *J Clin Oncol* 2013 (Mar); **31**: 1097–104.
3 Pao W, Hutchinson KE. Chipping away at the lung cancer genome. *Nat Med* 2012 (Mar); **18**: 349–51.
4 Imielinski M. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 2012; **150**: 1107–20.
5 Maemondo M, Inoue A, Kobayashi K *et al.* Gefitinib or chemotherapy for non-small-cell lung cancer with mutated EGFR. *N Engl J Med* 2010 (Jun 24); **362**: 2380–8.
6 Oizumi S, Kobayashi K, Inoue A *et al.* Quality of life with gefitinib in patients with EGFR-mutated non-small cell lung cancer: quality of life analysis of North East Japan Study Group 002 Trial. *Oncologist* 2012; **17**: 863–70.
7 Shaw AT, Engelman JA. ALK in lung cancer: past, present, and future. *J Clin Oncol* 2013 (Mar 10); **31**: 1105–11.
8 Mano H. ALKoma: a cancer subtype with a shared target. *Cancer Discov* 2012 (Jun); **2**: 495–502.
9 Sakamoto H, Tsukaguchi T, Hiroshima S *et al.* CH5424802, a selective ALK inhibitor capable of blocking the resistant gatekeeper mutant. *Cancer Cell* 2011 (May 17); **19**: 679–90.
10 Takeuchi K, Soda M, Togashi Y *et al.* RET, ROS1 and ALK fusions in lung cancer. *Nat Med* 2012 (Mar); **18**: 378–81.
11 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 (Mar); **18**: 382–4.
12 Kohno T, Ichikawa H, Totoki Y *et al.* KIF5B-RET fusions in lung adenocarcinoma. *Nat Med* 2012 (Mar); **18**: 375–7.
13 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 (Mar); **22**: 436–45.
14 Gild ML, Bullock M, Robinson BG, Clifton-Bligh R. Multikinase inhibitors: a new option for the treatment of thyroid cancer. *Nat Rev Endocrinol* 2011 (Oct); **7**: 617–24.
15 Borrello MG, Ardini E, Locati LD, Greco A, Licitra L, Pierotti MA. RET inhibition: implications in cancer therapy. *Expert Opin Ther Targets* 2013 (Apr); **17**: 403–19.
16 Drilon A, Wang L, Hasanovic A *et al.* Response to cabozantinib in patients with RET fusion-positive lung adenocarcinomas. *Cancer Discov* 2013 (Jun); **3**: 630–5.
17 Suzuki M, Makinoshima H, Matsumoto S *et al.* Identification of a lung adenocarcinoma cell line with CCDC6-RET fusion gene and the effect of RET inhibitors in vitro and in vivo. *Cancer Sci* 2013 (Apr 11); **104**: 896–903.
18 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 (Dec); **7**: 1872–6.
19 Cai W, Su C, Li X *et al.* KIF5B-RET fusions in Chinese patients with non-small cell lung cancer. *Cancer* 2013 (Apr 15); **119**: 1486–94.
20 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 (Dec 10); **30**: 4352–9.
21 Suehara Y, Arcila M, Wang L *et al.* Identification of KIF5B-RET and GOPC-ROS1 fusions in lung adenocarcinomas through a comprehensive mRNA-based screen for tyrosine kinase fusions. *Clin Cancer Res* 2012 (Dec 15); **18**: 6599–608.
22 Kim SC, Jung Y, Park J *et al.* A high-dimensional, deep-sequencing study of lung adenocarcinoma in female never-smokers. *PLoS ONE* 2013; **8**: e55596.
23 Yokota K, Sasaki H, Okuda K *et al.* KIF5B/RET fusion gene in surgically-treated adenocarcinoma of the lung. *Oncol Rep* 2012 (Oct); **28**: 1187–92.
24 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 (May); **22**: 928–31.
25 Chu CT, Sada YH, Kim ES. Vandetanib for the treatment of lung cancer. *Expert Opin Investig Drugs* 2012 (Aug); **21**: 1211–21.
26 Gautschi O, Zander T, Keller FA *et al.* A patient with lung adenocarcinoma and RET fusion treated with vandetanib. *J Thorac Oncol* 2013 (May); **8**(5): e43–4.
27 Mazieres J, Peters S, Lepage B *et al.* Lung cancer that Harbors an HER2 mutation: epidemiologic characteristics and therapeutic perspectives. *J Clin Oncol* 2013 (Jun 1); **31**: 1997–2003.
28 Gautschi O, Pauli C, Strobel K *et al.* A patient with BRAF V600E lung adenocarcinoma responding to vemurafenib. *J Thorac Oncol* 2012 (Oct); **7**(10): e23–4.
29 Awad MM, Katayama R, McTigue M *et al.* Acquired resistance to crizotinib from a mutation in CD74-ROS1. *N Engl J Med* 2013 (Jun 1); **368**: 2395–401.
30 Mendelsohn J. Personalizing oncology: perspectives and prospects. *J Clin Oncol* 2013 (May 20); **31**: 1904–11.
31 Yoshida A, Kohno T, Tsuta K *et al.* ROS1-rearranged lung cancer: a clinicopathologic and molecular study of 15 surgical cases. *Am J Surg Pathol* 2013 (Apr); **37**: 554–62.

doi: 10.1111/cas.12275

PLOS ONE

Mouse Model for ROS1-Rearranged Lung Cancer

Yasuhito Arai[1☯], Yasushi Totoki[1☯], Hiroyuki Takahashi[1], Hiromi Nakamura[1], Natsuko Hama[1], Takashi Kohno[2], Koji Tsuta[3], Akihiko Yoshida[3], Hisao Asamura[4], Michihiro Mutoh[5], Fumie Hosoda[1], Hitoshi Tsuda[3], Tatsuhiro Shibata[1]*

1 Division of Cancer Genomics, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan, **2** Division of Genome Biology, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan, **3** Division of Pathology and Clinical Laboratories, National Cancer Center Hospital, Chuo-ku, Tokyo, Japan, **4** Thoracic Surgery Division, National Cancer Center Hospital, Chuo-ku, Tokyo, Japan, **5** Division of Cancer Prevention Research, National Cancer Center Research Institute, Chuo-ku, Tokyo, Japan

Abstract

Genetic rearrangement of the *ROS1* receptor tyrosine kinase was recently identified as a distinct molecular signature for human non-small cell lung cancer (NSCLC). However, direct evidence of lung carcinogenesis induced by *ROS1* fusion genes remains to be verified. The present study shows that *EZR-ROS1* plays an essential role in the oncogenesis of NSCLC harboring the fusion gene. *EZR-ROS1* was identified in four female patients of lung adenocarcinoma. Three of them were never smokers. Interstitial deletion of 6q22–q25 resulted in gene fusion. Expression of the fusion kinase in NIH3T3 cells induced anchorage-independent growth *in vitro*, and subcutaneous tumors in nude mice. This transforming ability was attributable to its kinase activity. The ALK/MET/ROS1 kinase inhibitor, crizotinib, suppressed fusion-induced anchorage-independent growth of NIH3T3 cells. Most importantly, established transgenic mouse lines specifically expressing EZR-ROS1 in lung alveolar epithelial cells developed multiple adenocarcinoma nodules in both lungs at an early age. These data suggest that the *EZR-ROS1* is a pivotal oncogene in human NSCLC, and that this animal model could be valuable for exploring therapeutic agents against *ROS1*-rearranged lung cancer.

Citation: Arai Y, Totoki Y, Takahashi H, Nakamura H, Hama N, et al. (2013) Mouse Model for ROS1-Rearranged Lung Cancer. PLoS ONE 8(2): e56010. doi:10.1371/journal.pone.0056010

Editor: John D. Minna, Univesity of Texas Southwestern Medical Center at Dallas, United States of America

Received October 3, 2012; **Accepted** January 4, 2013; **Published** February 13, 2013

Funding: This study was supported by the Program for Promotion of Fundamental Studies in Health Sciences from the National Institute of Biomedical Innovation, National Cancer Center Research and Development Funds (23-A-7 and 23-B-28), Research Grant of the Princess Takamatsu Cancer Research Fund and Grants-in-Aid from the Ministry of Health, Labour and Welfare for the 3rd-term Comprehensive 10-year Strategy for Cancer Control. National Cancer Center Biobank is supported by the National Cancer Center Research and Development Fund, Japan. 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: tashibat@ncc.go.jp

☯ These authors contributed equally to this work.

Introduction

Lung cancer is the leading cause of cancer death around the world [1]. Lung adenocarcinoma (LADC), the most common form of non-small-cell lung cancer (NSCLC), comprises several different genomic subsets defined by unique oncogenic alterations, and a considerable proportion of LADC cases harbor driver alterations in the *EGFR*, *KRAS* and *ALK* genes at the mutually exclusive manner with rare exceptions [2–5]. Understanding the molecular basis of cancer allows us to develop therapeutic agents that target genetic druggable aberrations identified in cancer genomes. Tyrosine kinase inhibitors (TKIs) that target the EGFR and ALK proteins are particularly effective in the treatment of LADC carrying *EGFR* mutations and *ALK* fusions, respectively [2–6]. However, the development of an effective TKI requires experimental validation of the genetic aberrations as actionable and druggable. Transgenic mouse models harboring *EGFR* mutations or *EML4-ALK* gene fusions have successfully demonstrated the oncogenic potential of the alterations and the efficacy of TKI therapy [7,8]. Genetic rearrangement of the *ROS1* was recently identified as a distinct molecular signature for human LADC [9–16]. In the present study, we established a mouse model of *ROS1* fusion, and showed that *EZR-ROS1* as an essential driver oncogene in lung carcinogenesis.

Results

Identification of EZR-ROS1 Fusion Gene in LADC of Never-smokers

Whole transcriptome high-throughput sequencing of tumor specimens is one of the most effective methods for identifying fusion oncogenes [17]. Analysis of five LADC cases of never-smokers without *EGFR/KRAS/ALK* alterations using transcriptome sequencing identified 56 reads overriding the in-frame *EZR-ROS1* gene fusion point connecting *EZR* exon 10 to *ROS1* exon 34 in one tumor. RT-PCR analysis of matched non-cancerous tissues confirmed tumor-specific expression of the fusion transcript (Figure 1A). In addition, transcriptome sequencing clearly demonstrated a specific increase in the expression of the fused 3′ portion of *ROS1* (exons 34 to 43) after the breakpoint, suggesting that the *EZR-ROS1* fusion transcript causes aberrant overexpression of *ROS1* tyrosine kinase domain along with the 5′ portion of *EZR* (Figure 1B). SNP array comparative genomic hybridization (array CGH) data showed that this fusion gene was generated by a large interstitial deletion spanning ~41.5 Mb on chromosome 6q22-q25 (Figure 1C). Genomic PCR and sequencing analysis also revealed the deletion of 41.5 Mb causing somatic fusions of the

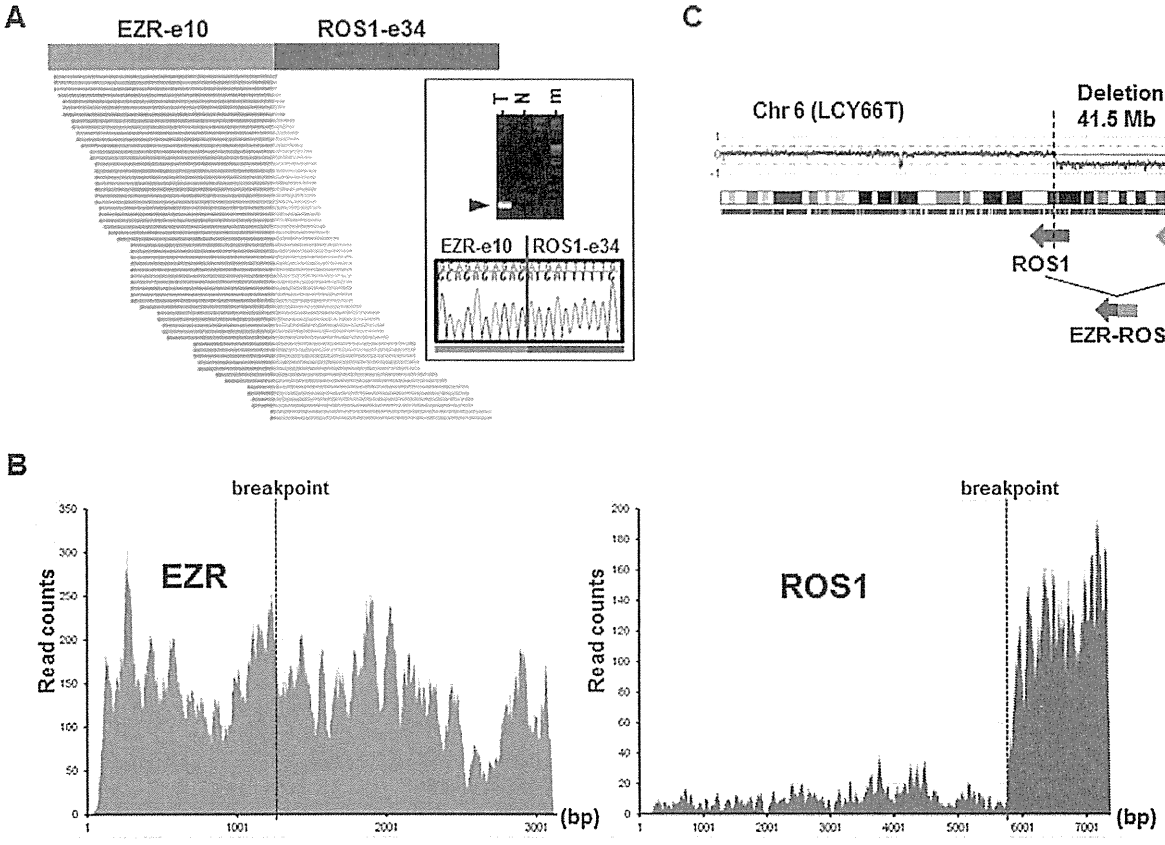

Figure 1. Identification of the *EZR-ROS1* fusion. (A) Junction reads representing *EZR-ROS1* fusion transcripts in LCY66T sample (left). Sanger sequencing of the RT-PCR product validated tumor-specific in-frame fusion transcript (right). m: molecular marker. (B) Expression profiles of *EZR* and *ROS1* in LCY66T. Active expression of the *ROS1* gene was observed after the fusion point. (C) SNP array CGH analysis of the LCY66T. Copy number throughout chromosome 6 is plotted as the log2 ratio.
doi:10.1371/journal.pone.0056010.g001

EZR intron 10 at 6q25 with the *ROS1* intron 33 at 6q22 (Figure S1).

RT-PCR and Sanger sequencing analysis of 569 LADC specimens from Japanese individuals, including the above-mentioned cases (343 cases with early pathological stage and 226 cases with advanced stage), identified four cases harboring this fusion transcript (Figure S2). All four *EZR-ROS1* fusion-positive cases were female, and harbored neither *EGFR/KRAS/HER2* mutations nor *EML4-ALK/KIF5B-RET* fusions. Three cases were poorly differentiated adenocarcinomas of never smokers, and the other was a moderately differentiated adenocarcinoma of a smoker.

Transforming Activity of EZR-ROS1

EZR-ROS1 cDNA isolated from the tumor specimen encoded a protein of 858 amino acids (Figure 2A; GenBank/DDBJ accession number AB698667). The protein connects the FERM domain [18] of ezrin (EZR) with the transmembrane and kinase domains of ROS1, but lacks most of the coiled-coil domain of EZR.

To examine the oncogenic activity of the *EZR-ROS1* fusion *in vitro*, we established stable NIH3T3 clones expressing wild-type EZR-ROS1 and kinase-dead mutant EZR-ROS1 (KD), in which the ATP-binding lysine residue was mutated to methionine (K491M), as well as mutants with serially deleted amino-terminal FERM domains (DL1, DL2 and DL3; Figure 2A). Autophosphorylation of specific tyrosine residues is a crucial event in the activation of distinct signal transduction pathways, and Tyr-2274 of ROS1 is a specific autophosphorylation site essential to induce kinase activity for transformation [19]. In transformation assays, phosphorylation of the Tyr-2274 (corresponding to Tyr-785 in wild type EZR-ROS1 fusion) was observed in a wild-type EZR-ROS1-expressing clone, but was not detected in kinase-dead (KD) and deleted (DL) mutants; this implies that the amino-terminal portion of FERM (1–88 amino acids) is necessary for ROS1 kinase activation (Figure 2B). Wild-type *EZR-ROS1* but not KD/DL mutants specifically induced activation of STAT3 for downstream signaling, and produced significantly anchorage-independent growth (Figure 2C, D). The anchorage-independent growth induced by *EZR-ROS1* was suppressed by treatment with crizotinib, a TKI against ALK/MET/ROS1, whereas the growth induced by another oncogene of lung, *CCDC6-RET* [11] was not (Figure 2E). On the contrary, vandetanib, a TKI against RET/EGFR/VEGFR was effective in inhibiting the colony formation of CCDC6-RET expressing cells, but not in the EZR-ROS1 expressing cells. As shown in Figure 2C, crizotinib treatment suppressed phosphorylation of EZR-ROS1, and inhibit the activation of STAT3.

Next, the NIH3T3 cells were subcutaneously injected into immune-compromised mice. Wild-type EZR-ROS1-expressing clones invariably produced tumors (6/6), while none of the KD

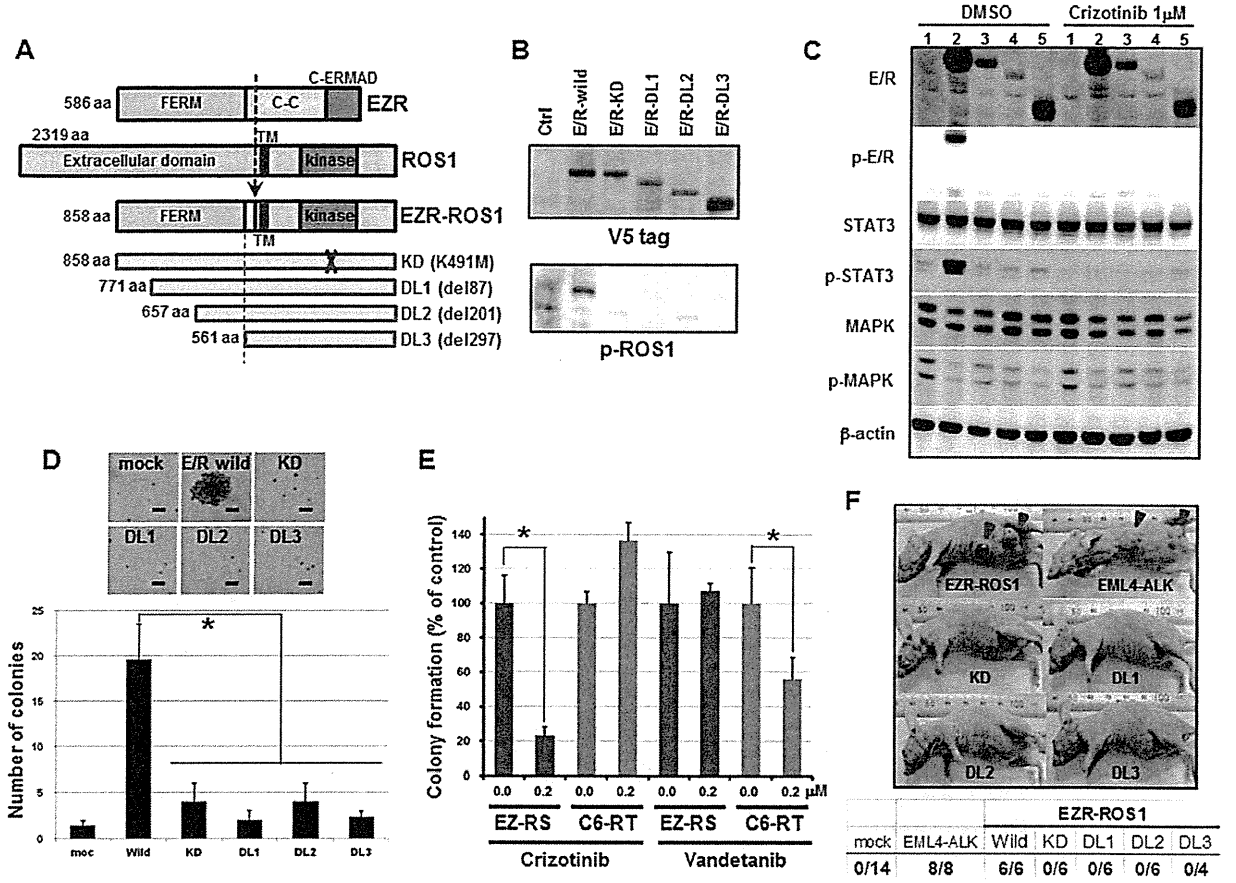

Figure 2. Oncogenic activity of the *EZR-ROS1* fusion gene. (A) Schematic representation of EZR, ROS1, EZR-ROS1, and deletions/mutations of EZR-ROS1 genes. The domain organization is shown. C-C: coiled-coil domain; TM: transmembrane; C-ERMAD: C-terminal ERM associated domain. (B) ROS1 phosphorylation in wild-type and mutant EZR-ROS1 (E/R)-expressing NIH3T3 clones. Cell lysates from each clone were immunoblotted with anti-V5-tag (top) and anti-phosphorylated ROS1 (Tyr-2274, bottom) antibodies. (C) Suppression of ROS 1 kinase activity of EZR-ROS1 by crizotinib inhibits STAT3 activation. NIH3T3 cells transfected with 1: empty vector, 2: wild-type EZR-ROS1, 3: KD 4: DL1, 5: DL3 were serum starved and treated for 2 hr with DMSO or 1 μM of crizotinib, and immunoblotted with the relevant antibodies. β-actin was used as a loading control. E/R: EZR-ROS1, p-E/R: phosphorylated EZR-ROS1 detected with an anti-phosphotyrosine-2274 antibody of ROS1. (D) Soft agar colony formation of wild-type and mutant EZR-ROS1 expressing NIH3T3 clones. A representative picture of colony formation for each clone is plotted at the top (scale bar, 100 μm). The number of colonies obtained for each clone is plotted at the bottom. *P<0.05. (E) Crizotinib-induced suppression of anchorage-independent growth of NIH3T3 cells expressing EZR-ROS1. Bar graph showing the percentage of NIH3T3 colonies induced by *EZR-ROS1* or *CCDC6-RET* after treatment with 200 nM of crizotinib or vandetanib with respect to those formed by DMSO-treated cells. EZ-ROS: EZR-ROS1, C6-RET: CCDC6-RET. *P<0.05. (F) Representative pictures of mice subcutaneously transplanted with NIH3T3 cells expressing wild-type, kinase domain-mutated, or amino-terminal-deleted EZR-ROS1. An EML4-ALK-expressing NIH3T3 clone was used as a positive control. The number of tumors per injection in each transfectant is shown below the photographs.
doi:10.1371/journal.pone.0056010.g002

and DL mutants-expressing clones produced tumors (Figure 2F), confirming that *in vivo* tumorigenic activity of *EZR-ROS1* requires ROS1 kinase activity.

Development of LADC in EZR-ROS1 Transgenic Mice

To further evaluate the role of *EZR-ROS1* in lung carcinogenesis, we generated transgenic mice expressing the fusion gene under the control of a type 2 alveolar epithelium-specific surfactant C gene promoter [20] (Figure 3A). We obtained four independent lines (TgA, B, C and D) with different copy number of the transgene (Figure S3) and detected lung adenocarcinoma nodules in all lines examined except TgD. Analysis of fusion protein expression level among them revealed no expression in TgD (Figure S4). The birth rate of transgene-positive progenies was low in TgC (Transgene-positive F1 progeny number : total F1 number; 1:3), and we failed to keep up a TgC line, then we mainly analyzed one line (TgA), which harbors approximately four copies of the transgene. RT-PCR and immunoblot analysis verified lung-specific *EZR-ROS1* mRNA and protein expression, and indicated phosphorylation of the EZR-ROS1 fusion protein (Figure 3B). Although endogenous *Ezrin* was ubiquitously expressed in many tissues, endogenous *Ros1*-transcript was detected only in stomach, kidney and lung. Protein expression levels of endogenous ROS1 were very weak compared with the levels of the fusion gene in the transgenic mice (Figure S4). Even at the four-week-old, multiple lesions over 1 mm in diameter were detected in the transgenic mice, and tumors occupied over 40% of sectioned surface of lung (Figure 3C and Figure S5). Computed tomography examination