Molecular dynamics simulation-guided drug sensitivity prediction for lung cancer with rare EGFR mutations

Shinnosuke Ikemura\textsuperscript{a,b,c,1}, Hiroyuki Yasuda\textsuperscript{a,2}, Shingo Matsumoto\textsuperscript{a,4}, Mayumi Kamada\textsuperscript{a,1}, Junko Hamamoto\textsuperscript{a,1}, Keita Masuzawa\textsuperscript{a}, Keigo Kobayashi\textsuperscript{a}, Tadashi Manabe\textsuperscript{a}, Daisuke Araiz\textsuperscript{a}, Ichiro Nakachi\textsuperscript{a}, Ichiro Kawada\textsuperscript{a}, Kota Ishioka\textsuperscript{a,f}, Morio Nakamura\textsuperscript{a}, Ho Namkoong\textsuperscript{a}, Katsuhiko Nakaik\textsuperscript{a}, Fumie Ono\textsuperscript{a}, Mitsugu Araki\textsuperscript{a}, Ryo Kanada\textsuperscript{a}, Biao Ma\textsuperscript{a}, Yukihiro Hayashi\textsuperscript{a}, Sachiyo Mimaki\textsuperscript{a}, Kiyotaka Yoh\textsuperscript{a}, Susumu S. Kobayashi\textsuperscript{a,1}, Takashi Kohno\textsuperscript{a}, Yasushi Okuno\textsuperscript{a}, Koichi Goto\textsuperscript{a}, Katuya Tsuchihara\textsuperscript{a,c,2}, and Kenzo Soejima\textsuperscript{a}

\textsuperscript{a}Division of Pulmonary Medicine, Department of Medicine, Keio University, School of Medicine, Shinjuku-ku, 160-8582 Tokyo, Japan; \textsuperscript{b}Keio University School of Medicine, Shinjuku-ku, 160-8582 Tokyo, Japan; \textsuperscript{c}Division of Molecular Informatics, Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, Kashiwa, 277-8577 Chiba, Japan; \textsuperscript{d}Division of Thoracic Oncology, National Cancer Center Hospital East, Kashiwa, 277-8577 Chiba, Japan; \textsuperscript{e}Graduate School of Medicine, Kyoto University, Shogoin Sakyu-ku, 606-8507 Kyoto, Japan; \textsuperscript{f}Tokyo Saiseikai Central Hospital, Minato-ku, 108-0073 Tokyo, Japan; \textsuperscript{g}Compass to Healthy Life Research Complex Program, RIKEN, Kobe, 650-0047 Hyogo, Japan; \textsuperscript{h}Research and Development Group for In Silico Drug Discovery, Pro-Cluster Kobe, Foundation for Biomedical Research and Innovation, Kobe, 650-0047 Hyogo, Japan; \textsuperscript{i}Department of Pathology, Keio University School of Medicine, 160-8582 Tokyo, Japan; \textsuperscript{j}Division of Translational Genomics, Exploratory Oncology Research and Clinical Trial Center, National Cancer Center, Kashiwa, 277-8577 Chiba, Japan; \textsuperscript{k}Division of Hematology/Oncology, Beth Israel Deaconess Medical Center and Harvard Medical School, Boston, MA 02215; and \textsuperscript{l}Division of Genome Biology, National Cancer Center Research Institute, 104-0045 Tokyo, Japan

Edited by Tak W. Mak, The Campbell Family Institute for Breast Cancer Research at Princess Margaret Cancer Centre, University Health Network, Toronto, ON, Canada, and approved April 9, 2019 (received for review November 14, 2018)

Next generation sequencing (NGS)-based tumor profiling identified an overwhelming number of uncharacterized somatic mutations, also known as variants of unknown significance (VUS). The therapeutic significance of EGFR mutations outside mutational hotspots, consisting of >50 types, in nonsmall cell lung carcinoma (NSCLC) is largely unknown. In fact, our pan-nation screening of NSCLC without hotspot EGFR mutations (n = 3,779) revealed that the majority (>90%) of cases with rare EGFR mutations, accounting for 5.5% of the cohort subjects, did not receive EGFR-tyrosine kinase inhibitors (TKIs) as a first-line treatment. To tackle this problem, we applied a molecular dynamics simulation-based model to predict the sensitivity of rare EGFR mutants to EGFR-TKIs. The model successfully predicted the diverse in vitro and in vivo sensitivities of exon 20 insertion mutants, including a singleton, to osimertinib, a third-generation EGFR-TKI (R\textsuperscript{2} = 0.72, P = 0.0037). Additionally, our model showed a higher consistency with experimentally obtained sensitivity data than other prediction approaches, indicating its robustness in analyzing complex cancer mutations. Thus, the in silico prediction model will be a powerful tool in precision medicine for NSCLC patients carrying rare EGFR mutations in the clinical setting. Here, we propose an insight to overcome mutation diversity in lung cancer.

Significance

A variety of rare mutations account for 10–20% of EGFR mutations in nonsmall cell lung cancer. However, due to high diversity, proper medication for patients with such mutations is impossible in daily clinic. To appropriately treat lung cancer patients harboring such rare EGFR mutations, a robust prediction model to predict sensitivities of rare EGFR mutants to existing drugs is strongly needed. Using molecular dynamics simulation-based model, we successfully predicted diverse sensitivities of EGFR exon 20 insertion mutants to existing inhibitors. The findings suggest the usefulness of in silico simulation to overcome mutation diversity at a clinically relevant level. The present in silico model will help in selecting effective drugs for these patients.

Author contributions: H.Y., M.K., Y.O., K.G., K.T., and K.S. designed research; S.I., S. Matsumoto, M.K., and J.H. contributed equally to this work. The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

This open access article is distributed under Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 (CC BY-NC-ND).

1\textsuperscript{1}S.I., S. Matsumoto, M.K., and J.H. contributed equally to this work.

2To whom correspondence may be addressed. Email: hiroyukiyasuda@a8.keio.jp or ktsuchihara@east.ncc.go.jp.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1819430116/-/DCSupplemental.

Published online May 1, 2019.

www.pnas.org/cgi/doi/10.1073/pnas.1819430116

PNAS | May 14, 2019 | vol. 116 | no. 20 | 10025–10030
several reports that *EGFR* exon 20 insertion mutants are resistant to EGFR-TKIs (7, 12, 22–24). NSCLC patients with these mutations are not administered EGFR-TKIs as the first-line treatment. However, we previously revealed that an *EGFR* exon 20 insertion mutant, A763_Y764insFQEA, is sensitive to the first- and second-generation EGFR-TKIs (23). Therefore, it is possible that a fraction of patients with *EGFR* exon 20 insertion mutations might benefit from therapy using some EGFR-TKIs. However, the high diversity of these mutations as well as the presence of many singleton mutations prevents the comprehensive characterization of the presently known mutants. Furthermore, the number of novel *EGFR* mutations is increasing owing to the use of NGS-based tests in lung cancer clinics. Thus, a rapid and robust method to accurately predict the sensitivity of *EGFR* rare mutants to existing TKIs in the clinical setting is necessary to tackle the problem that NSCLC patients with rare *EGFR* mutations often lose the chance of being treated with appropriate EGFR-TKIs.

Recently, computational structural modeling and molecular dynamics (MD) simulations have helped us clarify the activation mechanism of EGFR at the atomic level (25–27). In addition, predictions of sensitivity of *EGFR* mutants to EGFR tyrosine kinase inhibitors were performed for several *EGFR* mutations using binding free energy calculated with MD simulation (28, 29) and fitness scores calculated by molecular docking simulation (30). However, there is still room for discussion on the prediction accuracy and robustness of these models. Also, whether these methods can be applied to predict the sensitivity of various rare *EGFR* mutants to existing TKIs at a clinically relevant level remains elusive.

We have previously developed the supercomputer-based binding free energy calculation model utilizing MD simulation (31, 32) and applied our model to secondary ALK and RET mutants, which appeared during therapy using TKIs (33, 34). Based on our previous work, we hypothesized that our supercomputer-based model would allow us to predict the rare *EGFR* mutants to EGFR-TKIs at a clinically relevant level. To this end, we performed an interdisciplinary study, where computer science, cancer biology, and clinical oncology approaches were applied.

**Results**

**High Diversity of Rare EGFR Mutations in NSCLC.** To obtain clinically relevant information regarding individual *EGFR* mutations and to assess their diversity, we led the Lung Cancer Genomic Screening Project for Individualized Medicine in Japan (LC-SCRUM-Japan), a prospective nationwide lung cancer clinical and genomic characterization network, in which 217 institutions in Japan participated as of May 2017 (35). In this project, NSCLC cases in which the major *EGFR* mutations (exon 19 deletions, L858R, G719X, or L861Q) were not detected by the routine clinical testing underwent NGS evaluation for possible somatic alterations using a panel of cancer-related genes. Within LC-SCRUM-Japan, 3,779 NSCLC patients were enrolled from February 2013 to March 2017. Of these patients, 201 from October 2013 to June 2014 (first cohort) and 1,963 from March 2015 to March 2017 (second cohort) were subjected to NGS. The NGS study revealed that major *EGFR* mutations (exon 19 deletions, L858R, G719X, or L861Q) were detected in 2.7% (53/1,963) of such patients in the second cohort (Fig. 1A), indicating false negative results of routine clinical tests. Rare *EGFR* mutations were detected in 108 (5.5%) patients. The frequency of rare *EGFR* mutations was higher than those of *ROS1* fusions (3.6%) and *RET* fusions (2.9%). In addition, the frequency of rare *EGFR* mutations comprised approximately one-third of *KRAS* mutation frequency (15.3%). Rare *EGFR* mutations were detected in both nonsquamous NSCLC and squamous cell carcinoma (SI Appendix, Fig. S1). These data indicate that rare *EGFR* mutations account for a significant proportion of NSCLC cases.

The distribution of rare mutations throughout the *EGFR* gene sequence is illustrated in Fig. 1B. Most mutations were found in the region encoding the tyrosine kinase domain, in exons 18–21, while some mutations outside the tyrosine kinase domain, particularly in exons 6, 7, 8, 12, 15, and 17, were also detected. The most frequent rare *EGFR* mutations were *EGFR* exon 20 insertion mutations (Fig. 1B and SI Appendix, Table S1). Of the 113 NSCLC cases with rare *EGFR* mutations, including the five cases in the first cohort and the 108 cases in the second cohort, 52 (46.0%) harbored *EGFR* exon 20 insertion mutations, indicating that such mutations comprised about a half of rare *EGFR* mutations. Of the identified 73 types of rare *EGFR* mutations, 68 types (93.1%) were found in only one or two cases. These data indicate a high diversity of *EGFR* mutations in NSCLC.

**Low Chance of EGFR-TKI Therapy for Rare EGFR Mutation Cases.**

Clinical data were available for 53, 47, and 61 NSCLC patients with *EGFR* major mutations (exon 19 deletions, L858R, G719X, or L861Q), exon 20 insertion mutations, and other rare mutations, respectively (SI Appendix, Table S2). Notably, the characteristics of NSCLC patients with exon 20 insertion mutations were similar to those of patients with major mutations. However, the characteristics of NSCLC patients with other rare mutations were slightly different by including more male and heavy smoking patients. The majority of rare *EGFR* mutations were mutually exclusive with other driver oncogene mutations, although coexistence with *ERBB2* mutation or amplification, *PIK3CA* mutation, *KRAS* mutation, and *MET* amplification was observed in a small subset of patients (Fig. 1C). Of the 113 patients with rare *EGFR* mutations, 82 patients, including 33 cases with exon 20 insertion mutations, were available for information on chemotherapy after gene testing (Table 1 and SI Appendix, Table S3). Of those 82 patients, 77 (93.9%) received cytotoxic chemotherapy as first-line treatment, while only five patients (6.1%)...
were treated with EGFR-TKIs as first-line treatment. Thirty-two (97.0%) patients with EGFR exon 20 insertion mutations were treated with cytotoxic chemotherapy as first-line treatment.

The overall response rate to EGFR-TKIs in patients with rare EGFR mutations was only 17.4% (4/23; Table 1). Notably, one of the four cases, who responded to afatinib, carried NSCLC with an exon 20 insertion mutation, A767_V769dupASV (SI Appendix, Table S3), and the result was consistent with our previous study (36). This result validates that while the overall response rate to first- and second-generation EGFR-TKIs for NSCLC cases with rare EGFR mutations is low, there is a responsive subgroup among them.

**Calculation of Binding Energy of EGFR-TKIs for EGFR with Rare Mutations.** The above findings regarding EGFR mutation diversity and the efficacy of EGFR-TKIs for a subset of mutants prompted us to apply our in silico drug sensitivity prediction model (31, 32). We have previously indicated that a differential sensitivity of two representative rare EGFR mutants, exon 20 insertion mutants A763_Y764insFQEA and A767_V769dupASV, to two EGFR-TKIs, afatinib and osimertinib, revealed that A763_Y764insFQEA was more sensitive than A767_V769dupASV (36). We generated Ba/F3 cells harboring these EGFR mutations and evaluated the sensitivity by MTS cell proliferation assay. Here, we verified that the IC50 values of afatinib and osimertinib for the EGFR mutant harboring A763_Y764insFQEA were 10 times lower than those for the A767_V769dupASV mutant (11 vs. 171 nM and 33 vs. 321 nM, respectively). First, to evaluate whether our supercomputer-based binding free energy (ΔGbind) calculation model could predict the difference in sensitivity conferred by these mutations, we calculated ΔGbind values for the binding of afatinib to these EGFR mutants. In this model, the structures of EGFR molecules harboring rare mutations were built by homology modeling, and their binding affinities for EGFR-TKIs were evaluated using massively parallel computation of absolute binding free energy with a well-equilibrated system (the MPCAFEE method) (31, 32, 37). The modeled structure with bound afatinib is shown in SI Appendix, Fig. S2A. The calculated ΔGbind values were −23.3 and −18.5 for A763_Y764insFQEA and A767_V769dupASV, respectively (SI Appendix, Fig. S2B), indicating a more stable binding of afatinib to A763_Y764insFQEA than to A767_V769dupASV. Although a limited efficacy of afatinib for EGFR exon 20 insertion mutation-positive NSCLC patients was reported (38), the potential efficacy of osimertinib for EGFR exon 20 insertion mutation-positive NSCLC patients was reported (36, 39, 40), and several human clinical trials for evaluating the efficacy of osimertinib for EGFR exon 20 insertion mutations are ongoing. Therefore, we applied the model to osimertinib and nine recurrent or novel EGFR exon 20 insertion mutations found in our LC-SCRUM-Japan cohort including A763_Y764insFQEA, Y764_V765insHH, A767_S768insSVD, A767_V769dupASV, V769_D770insDPN, D770_N771insNPG, D770_N771insNPH, N771_P772insPGD, and P772_H773insHV. The modeled structures with bound osimertinib are shown in Fig. 2A and SI Appendix, Figs. S3 and S4. The calculated ΔGbind values are shown in SI Appendix, Table S4. Among the variable ΔGbind values, that of A763_Y764insFQEA was the lowest, −15.4 (kcal/mol), indicating the sensitivity of this mutation to osimertinib. In contrast, although the amino acid sequence was similar between D770_N771insNPG and D770_N771insNPH, the calculated ΔGbind values were distinctive [−13.2 (kcal/mol) for D770_N771insNPG and −10.0 (kcal/mol) for D770_N771insNPH]. These data may demonstrate structural complexity of osimertinib binding to mutant EGFR proteins. Finally, we found that the ΔGbind value for N771_P772insPGD, a mutation found in this cohort, was the lowest [−14.3 (kcal/mol)], except for that of A763_Y764insFQEA, among the mutations. These data indicate the potential sensitivity of this mutation to osimertinib.

**Sensitivity of EGFR with Rare Mutations to EGFR-TKIs and Its Correlation with Calculated Binding Free Energy Values.** To support our hypothesis that the ΔGbind values of rare EGFR mutations predict sensitivity to osimertinib, we generated a Ba/F3 library of rare EGFR mutations that were found in the LC-SCRUM-Japan cohort. EGFR transgenes bearing individual mutations were transduced into Ba/F3 cells, mouse pro-B cells (23). To examine the sensitivity of cells transduced with EGFR incorporating rare mutations to EGFR-TKIs, we performed MTS proliferation assays. We used the following EGFR-TKIs: erlotinib (first generation),

| Table 1. Treatment response of NSCLC harboring rare EGFR mutations | Median line: 2 (range 0–8 except Exon 20 ins range 0–6) |
|---------------------------------------------------------------|------------------------------------------------------------------------|
| **Chemotherapy** | **Mutations** | **First** | **Second** | **Third** | **Fourth and more** | **Total response rate, % (n)** |
| Cytotoxic agent | All rare mutations | 77 | 42 | 27 | 22 | — |
| | Exon 20 ins | 32 | 19 | 11 | 6 | — |
| | Other rares | 45 | 23 | 16 | 16 | — |
| EGFR-TKI | A) | A763_Y764insFQEA | 3 | 4 | 3 | 7 | 17.6 (3/17) |
| | | Exon 20 ins | 1 | 1 | 1 | 3 | 16.7 (1/6) |
| | | Other rares | 2 | 3 | 2 | 4 | 18.2 (2/11) |
| | B) | A767_V769dupASV | 1 | 1 | 0 | 1 | 33.3 (1/3) |
| | | Exon 20 ins | 0 | 1 | 0 | 0 | 0.0 (0/1) |
| | | Other rares | 1 | 0 | 0 | 1 | 50.0 (1/2) |
| | C) | A763_Y764insFQEA | 1 | 2 | 0 | 0 | 0.0 (0/3) |
| | | Exon 20 ins | 0 | 0 | 0 | 0 | — |
| | | Other rares | 1 | 2 | 0 | 0 | 0.0 (0/3) |
| ICI | Nivolumab | All rare mutations | 0 | 8 | 8 | 13 | 3.4 (2/29) |
| | | Exon 20 ins | 0 | 4 | 5 | 5 | 0.0 (0/14) |
| | | Other rares | 0 | 4 | 3 | 8 | 13.3 (2/15) |
| **Total number** | All rare mutations | 82 | 57 | 38 | 43 | — |
| | Exon 20 ins | 33 | 25 | 17 | 14 | — |
| | Other rares | 49 | 32 | 21 | 29 | — |

n = 82, 33, and 49 for All rare mutations, Exon 20 ins, and Other rares, respectively. ICI, immune checkpoint inhibitor.
The calculated docking scores did not show a correlation with the experimentally observed IC\textsubscript{50} values (R\textsuperscript{2}: 0.0054, P = 0.8508; SI Appendix, Fig. S5). The calculated docking scores using rDock are shown in SI Appendix, Table S6. These data indicate that the molecular docking method is not useful for sensitivity prediction of EGFR exon 20 insertion mutants.

Next, to compare the usefulness of other MD simulation-based methods with our model, we calculated binding free energies using MM-PBSA (Molecular Mechanics Poisson-Boltzmann Surface Area) and MM-GBSA (Molecular Mechanics Generalized Born Surface Area) with the MMTBSA.py module (41) in Amber Tools, which were employed in previous studies to investigate the sensitivity of EGFR mutants to TKIs (28, 29). Although the calculated binding free energies by MM-GGBSA and MM-PBSA showed statistically significant correlations (R\textsuperscript{2}: 0.5733, P = 0.018 for MM-GGBSA and R\textsuperscript{2}: 0.5744, P = 0.018 for MM-PBSA; SI Appendix, Fig. S6), the correlations were lower than those of our model. The calculated binding free energy values are shown in SI Appendix, Table S7. These results indicate that the sensitivity of cells expressing EGFR exon 20 insertion mutations to EGFR-TKIs can be predicted more accurately using our in silico prediction method based on MD simulation. In addition, to evaluate the robustness of our model, we calculated the binding energies or rDock for nonexon 20 insertion mutations to osimertinib. We selected representative single nucleotide variations (SNVs) and combinations of SNVs (SNV combinations) mutations. The calculated values are shown in SI Appendix, Tables S8 and S9. Interestingly, our model demonstrated higher correlations with the experimentally observed IC\textsubscript{50} values (R\textsuperscript{2}: 0.8392, P = 0.0288 for SNV and R\textsuperscript{2}: 0.8768, P = 0.0191 for SNV combinations) (SI Appendix, Fig. S7) than other models. These data indicate the robustness of our model for several types of EGFR mutations. In addition, to predict the therapeutic window of each EGFR-TKI, we have proposed that the selectivity index (SI), i.e., the ratio of log-transformed IC\textsubscript{50} values in Ba/F3 cells transduced with mutated and wild-type EGFR, should be used (36). SI values for EGFR-TKIs that inhibited the growth of cells with EGFR exon 20 insertions and other rare EGFR mutations are shown in SI Appendix, Figs. S8 and S9. These data indicate that osimertinib would affect a wide spectrum of lung cancers with rare EGFR mutations. However, sensitivity to osimertinib varied among the cells with rare EGFR mutations as there was a >100-fold difference between the lowest and the highest osimertinib IC\textsubscript{50} values. These data also indicate the variation in sensitivity of rare EGFR mutations, including EGFR exon 20 insertion mutations to osimertinib.

**Biological Confirmation of the Sensitivity of EGFR N771_P772insPGD to Osimertinib.** In this study, we have shown several previously unreported EGFR mutations. Of these EGFR mutations, we analyzed the EGFR exon 20 insertion mutation, N771_P772insPGD, found in a 45-y-old NSCLC patient diagnosed with lung adenocarcinoma in 2016. The positron emission tomography-computed tomography scan of the patient and the N771_P772insPGD sequences are shown in Fig. 3 A and B, respectively. Both in silico model and experimentally obtained IC\textsubscript{50} values predicted the sensitivity of this mutation to osimertinib. To validate the prediction that N771_P772insPGD is sensitive to osimertinib in vivo, we generated a patient-derived xenograft (PDX) model from pleural effusion of this patient. The histology of the PDX tumor in the mouse was also confirmed as adenocarcinoma (SI Appendix, Fig. S10). Consistent with the in silico prediction results, once-daily administration of 25 mg/kg (a dose that approximates the clinically approved 80-mg dose; ref. 39) or 50 mg/kg osimertinib induced a significant regression of the tumor (Fig. 3C and SI Appendix, Fig. S11). These data indicate that osimertinib could be effective in NSCLC cases with this mutation. These data support the applicability of the in silico prediction model for predicting the sensitivity of NSCLC with rare EGFR mutations to EGFR-TKIs.
Discussion

In this study, we clarified the diversity and driver roles of rare EGFR mutations in a large prospective Japanese NSCLC cohort and revealed the limited efficacy of precision medicine approaches for NSCLC patients with rare EGFR mutations. Among the 73 types of rare EGFR mutations detected, 68 (93.1%) were found in only one or two patients; the frequency of each rare EGFR mutation was less than 1%. In addition, sensitivity to EGFR-TKIs was quite diverse even if the mutation sequences were similar, e.g., for EGFR with D770_N771insNPV and D770_N771insNPH mutations. These data point to the diversity of rare EGFR mutations in their structural interaction with EGFR-TKIs. In vitro and/or in vivo experimental evaluation of drug sensitivity is effective in the evaluation of a few EGFR exon 20 mutants as shown by recent studies (42, 43). However, considering the extremely low frequency of each mutation and the continual appearance of novel mutations, such methods are not realistic in the clinical setting.

To overcome the problem of such mutation diversity, we, here, applied a supercomputer-based in silico prediction model that can be used to promptly estimate sensitivity to existing TKIs, obviating the need for time-consuming “wet” experiments. In this model, ΔGbind values that are highly correlated with sensitivity to drugs can be obtained in approximately 1 wk. Although several reports have proposed the availability of the in silico approach, by employing MD simulation, for predicting drug sensitivity of several EGFR mutations (28, 29), our model demonstrated the highest prediction accuracy based on comparisons with experimentally obtained sensitivity values.

These results indicate the usefulness and robustness of our model and show the potential to overcome mutation diversity in cancer. Here, in this study, we evaluated our model based on experimentally observed values for binding affinity. Since osimertinib is one of the targeted covalent inhibitors, further discussion on both the binding affinity and the rate of the subsequent bond-forming reaction will be needed. Of course, the clinical utility of our method should be evaluated in humans since it has been evaluated only in the PDX model. In this study, we indicated the efficacy of osimertinib for several EGFR exon 20 mutants. Thus, we have launched a clinical trial to evaluate the efficacy of osimertinib for cases with EGFR exon 20 insertion mutations [University Hospital Medical Information Network (UMIN) 000031929]. Prospective in silico and in vivo studies of these patients will prove the utility of our prediction method. Recent large-scale genomic characterization programs as well as rapid technological advancements have enabled the application of NGS in the clinical setting to identify numerous mutations in a variety of genes in lung and other cancers. Nonhot spot EGFR mutations have been discovered in small fractions of several cancers other than NSCLC. Application of our method might improve the prognosis of cancer patients by guiding drug development or promoting drug repositioning based on information on VUS in cancers.

Methods

Patients. LC-SCRUM-Japan is a prospective, nationwide clinical and genomic screen of lung cancer [UMIN ID: UMIN000010234]. In this study, a total of 3,779 NSCLC patients were enrolled from February 2013 to March 2017. Cases with NSCLC stage II or more advanced stage [tumor, nodes, metastasis (TNM) classification version 7], which were confirmed to have no major EGFR mutations (exon 19 deletions, L858R, G719X, or L861Q) by local hospitals, were included. The methods of EGFR mutation detection performed in local hospitals included PNA-LNA PCR clamp, Scorpion-ARMS, Cyclone PCR, PCR-invader, or Cobas EGFR mutation assay v.2. All patients provided written informed consent for the entry to the LC-SCRUM-Japan study.

DNA Extraction and Next Generation Sequencing. In this study, DNA samples were extracted from fresh frozen specimens or pleural effusion. From October 2013 to June 2014 (first cohort), DNA samples from 201 cases were analyzed by a targeted NGS assay, the Ion AmpliSeq Cancer Hotspot Panel (Thermo Fisher Scientific). From March 2015 to March 2017 (second cohort), DNA samples from 1,963 cases were analyzed by another targeted NGS assay, the Oncomine Comprehensive Assay (OCA v.1) (Thermo Fisher Scientific). Sequencing of the paired normal tissues or blood was not performed in this cohort. We selected potential somatic mutations those are registered as “confirmed somatic” in the COSMIC (Catalogue Of Somatic Mutations In Cancer) database. The mutations confirmed as somatic mutations by the COSMIC database are indicated with * in SI Appendix, Table S1. The present study was approved by the institutional Ethics Committee of all 217 institutions (SI Appendix, Table S10) that participated in the LC-SCRUM-Japan cohort. All patients provided written informed consent for the molecular analysis of their samples. All analyses were done at SRL, Inc.

Statistical Analysis. IBM SPSS Statistics 24 was used for data management and statistical analyses. For descriptive analysis, quantitative variables are expressed as the median and range. Categorical variables are expressed as the number of cases and percentage. Two-sided Student’s t tests were used for pairwise comparisons. The Pearson’s correlation test was performed to calculate R and P values. All statistical analyses were conducted with a significance level of α = 0.05 (P < 0.05). All P values are two-sided.

Other methods are described in SI Appendix.

ACKNOWLEDGMENTS. We thank Ms. Mikiko Shibuya for her excellent technical assistance, Dr. Matthew Meyerson (Dana–Farber Cancer Institute) for the important comments, and the Collaborative Research Resources at the Keio University School of Medicine for assistance with cell sorting. This research used computational resources of the K computer provided by the RIKEN Advanced Institute for Computational Science through High-Performance Computing Infrastructure (HPCI) System Research Project ID hp160213 and hp170275 and the NIG supercomputer at Research Organization of Information and Systems (R O I S) National Institute of Genetics. This work was supported in part by the Research Complex Promotion Program; Japan Society for the Promotion of Science Grants 22590870 (to K.S.) and 17K09667 (to H.Y.).
1. Cancer Genome Atlas Research Network (2014) Comprehensive molecular profiling of lung adenocarcinoma. Nature 511:543–550.

2. Cancer Genome Atlas Research Network (2012) Comprehensive genomic characterization of squamous cell lung cancers. Nature 489:519–525.

3. Kohsaka S, et al. (2017) A method of high-throughput functional evaluation of EGFR gene variants of unknown significance in cancer. Sci Transl Med 9eaam566.

4. Kris MG, et al. (2014) Using multiplexed assays of oncoviruses in lung cancers to select targeted drugs. JAMA 311:1998–2006.

5. Sharma SV, Bell DW, Settleman J, Haber DA (2007) Epidermal growth factor receptor mutations in lung cancer. Nat Rev Cancer 7:169–181.

6. Kosaka T, et al. (2004) Mutations of the epidermal growth factor receptor gene in lung cancer: Biological and clinical implications. Cancer Res 64:8919–8923.

7. Yasuda H, Kobayashi S, Costa DB (2012) EGFR exon 20 insertion mutations in non-small-cell lung cancer: Preatlinal data and clinical implications. Lancet Oncol 13:e23–e31.

8. Yun CH, et al. (2007) Structures of lung cancer-derived EGFR mutants and inhibitor complexes: Mechanism of activation and insights into differential inhibitor sensitivity. Cancer Cell 11:217–227.

9. Jura N, et al. (2011) Catalytic control in the EGFR receptor and its connection to general kinase regulatory mechanisms. Mol Cell 42:9–22.

10. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell 125:1137–1149.

11. Janne PA, Johnson BE (2006) Effect of epidermal growth factor receptor tyrosine kinase domain mutations on the outcome of patients with non-small cell lung cancer treated with epidermal growth factor receptor tyrosine kinase inhibitors. Clin Cancer Res 12:4416–4426.

12. Ponzoni M, Yatabe Y (2007) Mutations of the epidermal growth factor receptor gene and related genes as determinants of epidermal growth factor receptor tyrosine kinase inhibitors sensitivity in lung cancer. Cancer Sci 98:1817–1824.

13. Sequist LV, et al. (2013) Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. J Clin Oncol 31:3327–3334.

14. Cross DA, et al. (2014) AZD9291, an irreversible EGFR TKI, overcomes T790M-mediated resistance to covalent EGFR and HER2 inhibitors. Proc Natl Acad Sci USA 111:E6115–E6120.

15. Sequist LV, et al. (2013) Phase III study of afatinib or cisplatin plus pemetrexed in patients with metastatic lung adenocarcinoma with EGFR mutations. J Clin Oncol 31:3327–3334.

16. Ruan Z, Kannan N (2018) Altered conformational landscape and dimerization dependency underpins the activation of EGFR by ic50–ic200 loop insertion mutations. Proc Natl Acad Sci USA 115:E1675–E1685.

17. Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE (2013) Transitions to catalytically inactive conformations in EGFR kinase. Proc Natl Acad Sci USA 110:7270–7275.

18. Nakaoku T, et al. (2018) A secondary RET mutation in the activation loop conferring resistance to vandetanib. Nat Commun 9:625.

19. Katayama R, et al. (2014) Novel ALK mutations mediate acquired resistance to the next-generation ALK inhibitor alectinib. Clin Cancer Res 20:5686–5696.

20. Blankin AJ, Piantadosi S, Hollingsworth SJ (2015) Patient-centric trials for therapeutic personalized prediction of EGFR mutation-induced drug resistance in lung cancer. Sci Rep 5:1500.

21. Gryuch H, et al. (2006) Oncogenic transformation by inhibitor-sensitive and -resistant EGFR mutants. PLoS Med 3:2193.

22. Kondo T, et al. (2013) Structural, biochemical, and clinical characterization of epidermal growth factor receptor (EGFR) exon 20 insertion mutations in lung cancer. Sci Transl Med 5:216ra177.

23. Vizeta Y, et al. (2007) Allele-dependent variation in the relative cellular potency of distinct EGFR inhibitors. Cancer Biol Ther 6:661–667.

24. Zhang X, Gureasko J, Shen K, Cole PA, Kuriyan J (2006) An allosteric mechanism for activation of the kinase domain of epidermal growth factor receptor. Cell 125:1137–1149.

25. Shan Y, Arkhipov A, Kim ET, Pan AC, Shaw DE (2013) Transitions to catalytically inactive conformations in EGFR kinase. Proc Natl Acad Sci USA 110:7270–7275.

26. Ruan Z, Katayama R, Kannan N (2017) Computational and experimental characterization of patient derived mutations reveal an unusual mode of regulatory spine assembly and drug sensitivity in EGFR kinase. Biochemistry 56:22–32.

27. Wang DD, Zhou W, Yan H, Wong M, Lee V (2013) Personalized prediction of EGFR mutation-induced drug resistance in lung cancer. Sci Rep 3:2855.

28. Ma L, et al. (2015) EGFR mutant structural database. Computationally predicted 3D structures and the corresponding binding free energies with gefitinib and erlotinib. BMC Bioinformatics 16:85.

29. Ruiz-Carmena S, et al. (2014) Dox: A fast, versatile and open source program for docking ligands to proteins and nucleic acids. PLoS Comput Biol 10:e1003571.

30. Araki M, et al. (2016) The effect of conformational flexibility on binding free energy estimation between kinases and their inhibitors. J Chem Inf Model 56:2445–2456.

31. Brown JB, Nakatsui M, Okuno Y (2014) Constructing a foundational platform driven by Japan’s K supercomputer for next-generation drug design. Mol Inform 33:732–741.

32. Nakaoku T, et al. (2018) A secondary RET mutation in the activation loop conferring resistance to vandetanib. Nat Commun 9:625.

33. Fujitani H, Tanida Y, Matsuura A (2009) Massively parallel computation of absolute binding free energy with well-equilibrated states. Phys Rev E Stat Nonlin Soft Matter Phys 79:021914.

34. Yang J, et al. (2015) Clinical activity of afatinib in patients with advanced non-small-cell lung cancer harbouring uncommon EGFR mutations: A combined post-hoc analysis of LUX-lung 2, LUX-lung 3, and LUX-lung 6. Lancet Oncol 16:830–838.

35. Feuch N, et al. (2018) Antitumor activity of osimertinib, an irreversible mutant-selective EGFR tyrosine kinase inhibitor, in NSCLC harboring EGFR exon 20 insertions. Mol Cancer Ther 17:885–896.

36. Masuwa K, et al. (2017) Characterization of the efficacies of osimertinib and naxartinib against cells expressing clinically relevant epidermal growth factor receptor mutations. Oncotarget 8:105479–105491.

37. Miller BR, 3rd, et al. (2012) MMPBSA.py: An efficient program for end-state free energy calculations. J Chem Theory Comput 8:3314–3321.

38. Robichaux JP, et al. (2018) Mechanisms and clinical activity of an EGFR and HER2 exon 20-selective kinase inhibitor in non-small cell lung cancer. Nat Med 24:639–646.

39. Nakaoku T, et al. (2018) TAS6617, a novel epidermal growth factor receptor inhibitor targeting exon 20 insertion mutations. Mol Cancer Ther 17:1648–1658.