

## RESEARCH ARTICLE

## Whole genome sequencing to identify predictive markers for the risk of drug-induced interstitial lung disease

Chihiro Udagawa<sup>1</sup>, Hidehito Horinouchi<sup>2</sup>, Kouya Shiraishi<sup>3</sup>, Takashi Kohno<sup>3</sup>, Takuji Okusaka<sup>4</sup>, Hideki Ueno<sup>4</sup>, Kenji Tamura<sup>5</sup>, Yuichiro Ohe<sup>2</sup>, Hitoshi Zembutsu<sup>1,6\*</sup>

**1** Division of Genetics, National Cancer Center Research Institute, Tokyo, Japan, **2** Department of Thoracic Oncology, National Cancer Center Hospital, Tokyo, Japan, **3** Division of Genome Biology, National Cancer Center Research Institute, Tokyo, Japan, **4** Department of Hepatobiliary and Pancreatic Oncology, National Cancer Center Hospital, Tokyo, Japan, **5** Department of Breast and Medical Oncology, National Cancer Center Hospital, Tokyo, Japan, **6** Project for Development of Liquid Biopsy Diagnosis, Japanese Foundation for Cancer Research, Research Institute, Tokyo, Japan

\* [hitoshi.zembutsu@jfc.or.jp](mailto:hitoshi.zembutsu@jfc.or.jp)

## OPEN ACCESS

**Citation:** Udagawa C, Horinouchi H, Shiraishi K, Kohno T, Okusaka T, Ueno H, et al. (2019) Whole genome sequencing to identify predictive markers for the risk of drug-induced interstitial lung disease. PLoS ONE 14(10): e0223371. <https://doi.org/10.1371/journal.pone.0223371>

**Editor:** Rania Mohamed Labib, Children's Cancer Hospital-57357-Egypt, EGYPT

**Received:** May 17, 2019

**Accepted:** September 19, 2019

**Published:** October 4, 2019

**Copyright:** © 2019 Udagawa et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** Although the authors performed next-generation sequencing (whole genome sequencing), they are unable to upload these data, because they did not receive an agreement from all patients to register their sequence data in a database. However, the authors have described allele count data of the candidate variants (variants associated with DIILD) in the paper and Supporting Information files. Interested researchers may send data requests to: Ms. Tomomi Kuroki ([tomomi.kuroki@jfc.or.jp](mailto:tomomi.kuroki@jfc.or.jp)).

## Abstract

Drug-induced interstitial lung disease (DIILD) is a serious side effect of chemotherapy in cancer patients with an extremely high mortality rate. In this study, to identify genetic variants with greater risk of DIILD, we carried out whole genome sequencing (WGS) of germline DNA samples from 26 patients who developed DIILD, and conducted a case-control association study between these 26 cases and general Japanese population controls registered in the integrative Japanese Genome Variation Database (iJGVD) as a screening study. The associations of 42 single nucleotide variants (SNVs) showing  $P < 0.0001$  were further validated using an independent cohort of 18 DIILD cases as a replication study. A further combined analysis of the screening and replication studies showed a possible association of two SNVs, rs35198919 in intron 1 of the chromosome 22 open reading frame 34 (*C22orf34*) and rs12625311 in intron 1 of the teashirt zinc finger homeobox 2 (*TSHZ2*), with DIILD ( $P_{combined} = 1.87 \times 10^{-5}$  and  $5.16 \times 10^{-5}$ , respectively). Furthermore, in a subgroup analysis of epidermal growth factor receptor (EGFR)-tyrosine kinase inhibitor (TKI)-induced interstitial lung disease (ILD), we observed seven candidate SNVs that were possibly associated with ILD ( $P < 0.00001$ ). This is the first study to identify genetic markers for the risk of DIILD using WGS. Collectively, our novel findings indicate that these SNVs may be applicable for predicting the risk of DIILD in patients receiving chemotherapy.

## Introduction

Drug-induced interstitial lung disease (DIILD) is a serious side effect of a wide range of drugs, including anticancer [1–3], antirheumatic [4–6], and cardiology drugs [7–9]. While the incidence rate of DIILD is approximately 0.1%–10% for anticancer drugs, DIILD is potentially life-threatening and represents a serious clinical problem [10]. Therefore, tools for predicting

**Funding:** This work was supported by AstraZeneca (HZ), <https://www.astrazeneca.com/>. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** Hitoshi Zembutsu received research funding from AstraZeneca as we described in the online submission form and had no other relationship with them or others. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

the risk of DIILD are important in order to avoid life-threatening drug toxicity. Although the precise mechanism underlying DIILD is not well understood, cumulative evidence suggests that direct cytotoxic effects related to the parent drug and/or its metabolites, along with certain immune factors, may be involved [11–13]. Serum levels of Krebs von der Lungen-6 (KL-6), surfactant protein (SP) -A and SP-D have commonly been used as biomarkers for ILD in clinical settings [14–16]. KL-6 is a high molecular weight glycoprotein secreted by proliferating or damaged type II alveolar pneumocytes and bronchial epithelial cells and has high diagnostic sensitivity (93.9%) and specificity (96.3%) for ILD [17]. Clinical risk factors for developing DIILD vary according to the disease or drug involved; however, certain risk factors such as increased age [18–20], preexisting ILD or idiopathic pulmonary fibrosis (IPF) [21, 22], smoking [19, 22], and poor performance status [19, 20] are associated with the development of the toxicity. Although human leukocyte antigen (HLA) alleles have been reported to be associated with methotrexate-induced ILD and gemcitabine plus erlotinib-induced ILD [23, 24], a clinically useful biomarker for predicting the risk of DIILD before treatment has yet to be developed.

In this study, we carried out whole genome sequencing (WGS) on genomic DNA from 26 patients who developed DIILD and identified two genomic regions, the chromosome 22 open reading frame 34 (*C22orf34*) and the tea shirt zinc finger homeobox 2 (*TSHZ2*) gene on chromosome 20, which may be associated with DIILD. Our results may provide insight into the underlying mechanism of DIILD and aid in the development of a diagnostic system for predicting the risk of this life-threatening adverse event.

## Materials and methods

### Patients

To identify genetic susceptibility loci for DIILD, we used 26 DIILD patients treated with epidermal growth factor receptor (EGFR)–tyrosine kinase inhibitor (TKI), nivolumab, gemcitabine, and trastuzumab for the screening study. Of the 26 patients, 12 who had EGFR-TKI-induced ILD between June 2002 and July 2008 at the National Cancer Center Hospital, and 1 patient who was registered in the BioBank Japan (<http://biobankjp.org/>) and satisfied the following criteria; a. EGFR-TKI was used, b. patient with DIILD during the therapy, were used for the screening study. In addition, from the registered samples in the National Cancer Center (NCC) biobank, 13 patients who were treated with nivolumab, gemcitabine, or trastuzumab, and were confirmed to have DIILD in clinical data as of May 2016, were used for the screening study (Table 1). In a replication study, 18 patients who were treated with nivolumab or trastuzumab and were newly confirmed to have DIILD in updated clinical data as of November 2018, were used from the registered samples in the NCC biobank (Table 1). DIILD was diagnosed by clinical course, objective findings, and independent computed tomography (CT) findings by radiologists or respiratory oncologists.

**Table 1. Demographics of the 44 patients with drug-induced interstitial lung disease (DIILD).**

Patient	Tumor	Drug	Smoking history/status	Preexisting ILD	Treatment for ILD	Study set
1	Lung	Gefitinib	Yes / current	No	DX + Oxygen	Screening
2	Lung	Gefitinib	Yes / former	Yes	PSL + Oxygen + CAZ	Screening
3	Lung	Erlotinib	No	Yes	MP + Oxygen + CFPM	Screening
4	Lung	Gefitinib	Yes / former	No	PSL + Oxygen + MEPM	Screening
5	Lung	Gefitinib	No	No	MP + Oxygen + MEPM + Warfarin	Screening

(Continued)

Table 1. (Continued)

Patient	Tumor	Drug	Smoking history/status	Preexisting ILD	Treatment for ILD	Study set
6	Lung	Gefitinib	Yes / former	Yes	PSL + Oxygen + CAZ	Screening
7	Lung	Erlotinib	Yes / former	Yes	DX	Screening
8	Lung	Gefitinib	Yes / former	No	PSL + Oxygen + BIPM	Screening
9	Lung	Gefitinib	No	No	MP + Oxygen + CAZ	Screening
10	Lung	Gefitinib	Yes / former	No	PSL + Oxygen + ST	Screening
11	Lung	Gefitinib	Yes / former	No	MP + Oxygen + PARM/BP	Screening
12	Lung	Gefitinib	Yes / former	Yes	PSL + Oxygen	Screening
13	Lung	Nivolumab	Yes	No	PSL + Oxygen	Screening
14	Lung	Nivolumab	Yes / former	No	MP + Oxygen + Carbocisteine + AZM + CAM	Screening
15	Lung	Nivolumab	Yes / former	No	Oxygen + Warfarin	Screening
16	Lung	Gefitinib	Unknown	Unknown	Unknown	Screening
17	Pancreas	Gemcitabine	Unknown	No	None	Screening
18	Pancreas	Gemcitabine	No	No	None	Screening
19	Pancreas	Gemcitabine	Yes	No	PSL + Ambroxol + ST + Cilostazol	Screening
20	Pancreas	Gemcitabine <sup>a</sup>	No	No	None	Screening
21	Pancreas	Gemcitabine + nab-PTX	No	No	PSL + CEZ	Screening
22	Pancreas	Gemcitabine + nab-PTX	No	No	PSL + Oxygen + TAZ/PIPC	Screening
23	Breast	Trastuzumab + paclitaxel	No	No	PSL + Oxygen + SBTPC + Warfarin	Screening
24	Breast	Trastuzumab <sup>b</sup>	No	No	ST	Screening
25	Breast	Trastuzumab + paclitaxel	No	No	None	Screening
26	Gastric	Trastuzumab + paclitaxel	Yes / former	No	PSL + Oxygen + ICS/LABA + ST	Screening
27	Lung	Nivolumab	Yes / unknown	No	PSL + CPFX	Replication
28	Lung	Nivolumab	Yes / unknown	No	PSL + ST	Replication
29	Lung	Nivolumab	Yes / former	No	PSL + ST	Replication
30	Lung	Nivolumab	Yes / unknown	No	PSL + Oxygen + ST	Replication
31	Lung	Nivolumab	Yes / former	No	PSL + ST	Replication
32	Lung	Nivolumab	Unknown	No	Eprazinone + AZM	Replication
33	Lung	Nivolumab <sup>c</sup>	Yes / unknown	No	MP + Oxygen + TAZ/PIPC + ST	Replication
34	Melanoma	Nivolumab	No	Yes	PSL + ST	Replication
35	Melanoma	Nivolumab	Yes / unknown	No	PSL	Replication
36	Melanoma	Nivolumab	No	No	PSL + Oxygen + ST	Replication
37	Melanoma	Nivolumab	Unknown	No	BMZ + Oxygen + ST	Replication
38	Melanoma	Nivolumab	No	Yes	PSL + ST	Replication
39	Breast	Trastuzumab + paclitaxel	No	No	PSL	Replication
40	Breast	Trastuzumab + paclitaxel	No	No	PSL	Replication
41	Breast	Trastuzumab + paclitaxel	No	No	PSL + CPFX	Replication
42	Breast	Trastuzumab + paclitaxel	No	No	PSL + Budesonide + ST	Replication
43	Gastric	Trastuzumab <sup>d</sup>	No	No	PSL + Oxygen + MEPM	Replication
44	Gastric	Trastuzumab emtansine	No	No	MP + Oxygen + TAZ/PIPC	Replication

DX, Dexamethasone; PSL, Prednisolone; CAZ, Ceftazidime; MP, Methylprednisolone; CFPM, Cefepime; MEPM, Meropenem; BIPM, Biapenem; ST, Sulfamethoxazole trimethoprim; PARM/BP, Panipenem/betamipron; AZM, Azithromycin; CAM, Clarithromycin; CEZ, Cefazolin; TAZ/PIPC, Tazobactam/piperacillin; SBTPC, Sultamicillin; ICS/LABA, Inhaled corticosteroid/long-acting  $\beta$ -agonists; CPFX, Ciprofloxacin; BMZ, Betamethasone.

<sup>a</sup> This patient had DIILD during FOLFIRINOX (a combination of 5-fluorouracil, leucovorin, irinotecan, and oxaliplatin) therapy following gemcitabine.

<sup>b</sup> This patient had DIILD during adriamycin + cyclophosphamide therapy following trastuzumab.

<sup>c</sup> This patient had DIILD during docetaxel therapy following nivolumab.

<sup>d</sup> This patient had DIILD during nanoparticle albumin-bound paclitaxel therapy following trastuzumab.

<https://doi.org/10.1371/journal.pone.0223371.t001>

We used 3,554 general Japanese population registered in the integrative Japanese Genome Variation Database (iJGVD) as study controls. Because DIILD is considered a rare adverse event in the population, using a large population control set (healthy individuals) increases the statistical power even while recognizing the (small) potential for misclassification. Population controls have been successfully used in this fashion in previous studies of rare adverse drug reactions and have yielded significant associations [25–28]. In addition, 1,138 patients undergoing chemotherapy but never developed DIILD were recruited as tolerant controls from the NCC biobank in a replication analysis for the two possible SNVs (rs35198919 and rs12625311). This study was approved by the Institutional Review Board of the National Cancer Center (Tokyo, Japan) and the Japanese Foundation for Cancer Research (Tokyo, Japan).

## WGS

WGS was conducted for 26 DNA samples from patients with DIILD. Genomic DNA was extracted from the whole blood or normal tissues, sheared into approximately 350 bp fragments, and used to make a library with the TruSeq Nano DNA Sample Prep Kit (Illumina, San Diego, CA, USA). Sequencing was performed on an Illumina HiSeq X platform using a paired-end 150 bp configuration.

## Mapping, variant calling, filtering, and Sanger sequencing

Adapter sequences were removed by Cutadapt (v1.2.1) [29]. After quality control, reads were mapped to the reference human genome (hg19) using Burrows–Wheeler Aligner (BWA) (ver.0.7.10) [30]. Mapping results were corrected using Picard (ver.1.73) (<http://broadinstitute.github.io/picard/>) for removing duplicates and Genome Analysis Toolkit (GATK, ver.1.6–13) for local alignment and quality score recalibration. SNV calls were performed with multi-sample calling using GATK [31] and filtered to coordinate with Variant Quality Score Recalibration (passed) and the variant call quality score ( $\geq 30$ ). Annotations of the variants were based on dbSNP149, CCDS (NCBI, Release 15), RefSeq (UCSC Genome Browser, Feb 2017), Gencode (UCSC Genome Browser, ver. 19), and 1000 Genomes (phase3 release v5). SNVs were further filtered according to the following criteria: (i) A genotype quality score  $\geq 20$ ; (ii) Variants in segmental duplication, homopolymer, or repeat regions identified by Repeat Masker or Tandem Repeats Finder were excluded; (iii) SNVs showing significantly different allele frequencies ( $P \leq 0.05$ ) between 104 Japanese subjects (JPT) in the 1000 Genomes Project and 3,554 Japanese subjects in the iJGVD were excluded. Forty-two candidate SNVs identified by the WGS screening study were verified by Sanger sequencing. In the replication study, genotyping was performed by Sanger sequencing as previously described [32].

## Statistical analysis

In the screening (WGS) and replication study, a case (patients with DIILD)-control (3,554 general Japanese individuals in iJGVD) [33] association study was performed using the Fisher's exact test in an allele frequency model. The significance levels after Bonferroni correction for multiple testing were  $P = 7.60 \times 10^{-8}$  (0.05/658,234) in the screening study and  $P = 1.47 \times 10^{-3}$  (0.05/34) in the replication study because eight SNVs were highly linked ( $r^2 = 1$ ) in the 42 candidate SNVs. The odds ratios (ORs) and confidence intervals (CIs) were calculated using the non-risk allele as a reference. All statistical analyses were carried out using R statistical environment version 3.3.1 (<http://www.r-project.org/>) or the BellCurve for Excel (Social Survey Research Information Co., Ltd., Tokyo, Japan).

## Results

### Clinical characteristics of patients with drug-induced ILD

The clinical characteristics of the 44 patients who developed DIILD are summarized in [Table 1](#). The median age was 62 years (range: 36–80 years) and 25 (57%) were female. Seven (16%) and 21 (48%) patients had a history of ILD and smoking, respectively. We enrolled patients with lung cancer (23 patients), breast cancer (7 patients), pancreatic cancer (6 patients), melanoma (5 patients), and gastric cancer (3 patients). The causative drugs of ILD in the 44 patients were found to be nivolumab (14 patients), EGFR-TKI (gefitinib or erlotinib; 13 patients), trastuzumab plus paclitaxel (7 patients), gemcitabine (3 patients), gemcitabine plus nanoparticle albumin-bound paclitaxel (2 patients), FOLFIRINOX (a combination of 5-fluorouracil, leucovorin, irinotecan, and oxaliplatin; 1 patient), adriamycin plus cyclophosphamide (1 patient), trastuzumab emtastine (1 patient), docetaxel (1 patient), and nanoparticle albumin-bound paclitaxel (1 patient). In all patients with DIILD, the causative drugs were withdrawn after the diagnosis of ILD, and most received treatment for ILD as shown in [Table 1](#).

### WGS and identification of candidate genetic biomarkers for the risk of DIILD

To identify genetic susceptibility loci for DIILD, we performed WGS using the germline DNA of 26 patients who developed DIILD during anticancer treatment ([Table 1](#)). The mean number of SNVs per patient was 4,042,041 with a 35× average depth. Variants were filtered using an in-house program as described in the Materials and methods; a total of 658,234 SNVs were analyzed in the 26 patients. We next conducted a case-control association study between the 26 cases and the 3,554 (maximum) general Japanese population controls registered in the iJGVD to identify SNVs associated with DIILD as a screening study. Although no SNVs reached a significance level of  $P < 7.60 \times 10^{-8}$  (see [Materials and methods](#)), we observed 42 SNVs showing  $P < 0.0001$  ( $3.31 \times 10^{-6} < P < 9.94 \times 10^{-5}$ ) as shown in [S1 Table](#).

### Replication and combined study using additional patients

The associations of 42 SNVs showing  $P < 0.0001$  in the screening study were further investigated using an independent cohort consisted of 18 DIILD cases ([Table 1](#), [S1 Table](#)). We performed genotyping of the 42 SNVs by the Sanger sequencing method using the above 18 cases; however, we could not observe the SNV that reached the significance level of  $P = 1.47 \times 10^{-3}$  (see [Materials and methods](#)) as shown in [S1 Table](#). However, in the combined analysis of the screening and replication studies for the 42 SNVs, we found that the associations of two SNVs, rs35198919 in intron 1 of *C22orf34* and rs12625311 in intron 1 of *TSHZ2*, with DIILD were stronger than those in the screening study ( $P_{\text{combined}} = 1.87 \times 10^{-5}$ , odds ratio (OR) = 4.15, 95% CI; 2.36–7.31, and  $P_{\text{combined}} = 5.16 \times 10^{-5}$ , OR = 2.43, 95% CI; 1.59–3.71, respectively, [Table 2](#), [Fig 1](#)).

### Subgroup analysis for the identification of genetic markers for EGFR-TKI-induced ILD

To identify the genetic marker(s) for EGFR-TKI-induced ILD, we next performed an association study using the 13 patients with EGFR-TKI-induced ILD (cases) and the abovementioned controls from the iJGVD as a subgroup analysis. We observed that seven SNVs were possibly associated with EGFR-TKI-induced ILD, with  $P < 0.00001$  as shown in [Table 3](#), however, they showed no or weak associations in the 13 patients with nonEGFR-TKI-induced ILD ( $P = 0.25$ –1).

**Table 2. Association results of two SNVs with smaller *P* values in the combined analysis of screening and replication studies than those in the screening study.**

Chr.	SNP ID	Position <sup>a</sup>	Gene	Allele (1/2) <sup>b</sup>	Location	Stage	Risk allele	Case					Control <sup>c</sup>					<i>P</i> value	Odds ratio [95% CI]
								Allele		Allele Frequency		RAF	Allele		Allele Frequency		RAF		
								1	2	1	2		1	2	1	2			
22	rs35198919	50,034,703	<i>C22orf34</i>	T/C	Intron	Screening	C	41	11	0.79	0.21	0.21	6,725	333	0.95	0.05	0.05	2.93 x 10 <sup>-5</sup>	5.42 [2.76–10.64]
								32	4	0.89	0.11	0.11						8.94 x 10 <sup>-2</sup>	2.52 [0.89–7.18]
								73	15	0.83	0.17	0.17						1.87 x 10 <sup>-5</sup>	4.15 [2.36–7.31]
20	rs12625311	51,836,793	<i>TSHZ2</i>	T/C	Intron	Screening	C	21	31	0.40	0.60	0.60	4,755	2,349	0.67	0.33	0.33	9.59 x 10 <sup>-5</sup>	2.99 [1.71–5.21]
								19	17	0.53	0.47	0.47						7.75 x 10 <sup>-2</sup>	1.81 [0.94–3.49]
								40	48	0.45	0.55	0.55						5.16 x 10 <sup>-5</sup>	2.43 [1.59–3.71]

Chr., chromosome; RAF, risk allele frequency; CI, confidence interval.

<sup>a</sup> Based on GRCh37 genome assembly.

<sup>b</sup> Reference allele (GRCh37) was defined as allele 1.

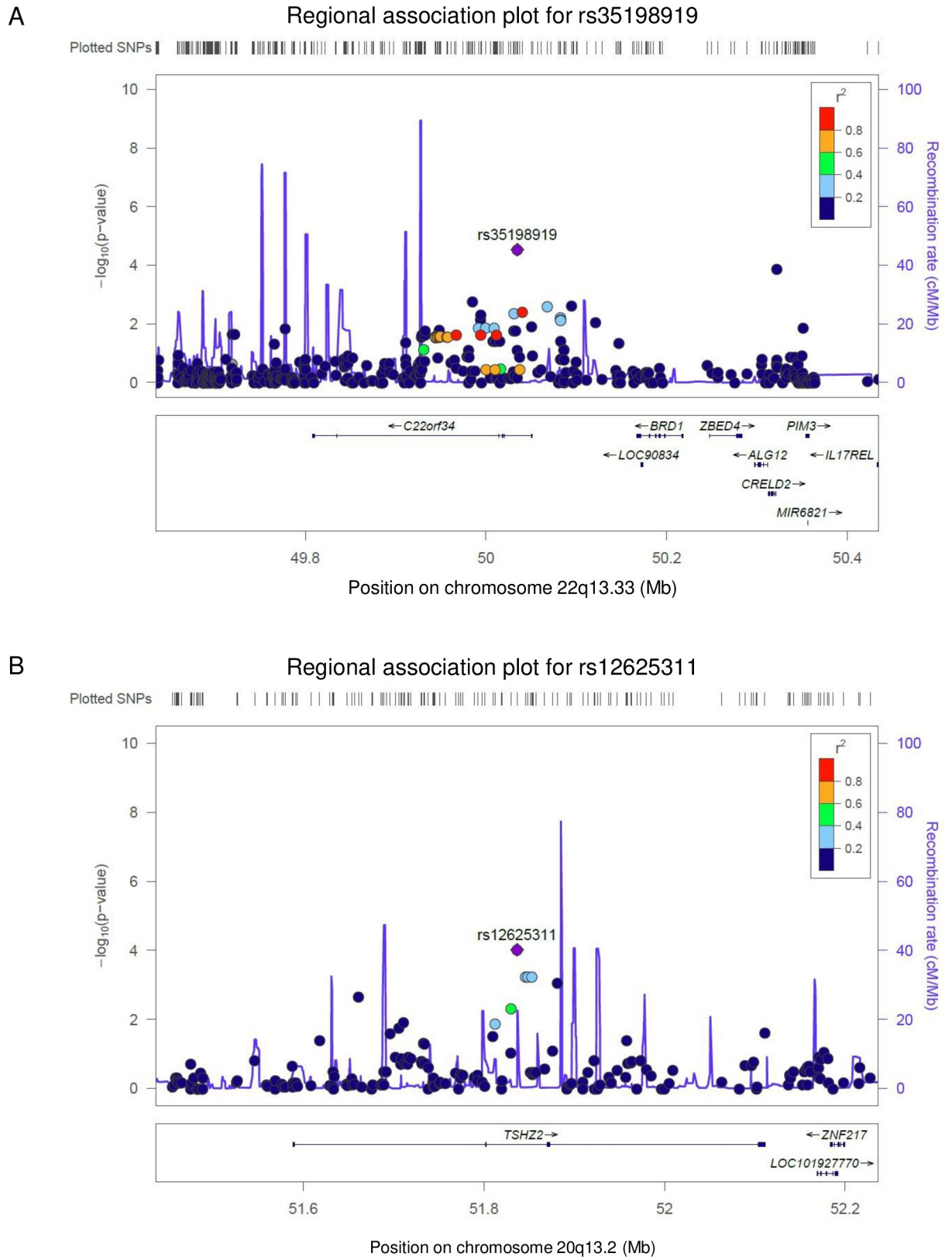
<sup>c</sup> The same controls were used in screening, replication, and combined analysis.

<https://doi.org/10.1371/journal.pone.0223371.t002>

## Discussion

The development of predictive marker(s) for the risk of life-threatening toxicities related to cancer treatment, including ILD, based on germline mutation or variation is important in order to provide safe treatment regimens for patients with cancer. This study is the first WGS that attempts to identify common or rare genetic variants associated with DIILD in the Japanese population. The SNV strongly associated with DIILD was not identified in this study; however, we identified two independent loci showing possible associations with DIILD (Table 2, Fig 1). Moreover, subgroup analysis using patients with EGFR-TKI-induced ILD suggested a possible association of seven SNVs with EGFR-TKI-induced ILD (Table 3). To enlarge the sample size, we further added patients undergoing chemotherapy after cancer diagnosis but never developed DIILD as a control for a replication cohort and analyzed for rs35198919 and rs12625311 (S2 Table). The association results of the replication and combined studies were comparable with those obtained from the original control set (general Japanese population) as shown in Table 2 and S2 Table.

The most strongly associated SNV from the combined results of the screening and replication studies was rs35198919 ( $P_{combined} = 1.87 \times 10^{-5}$ , OR = 4.15), located in *C22orf34* on chromosome 22q13.33. Although the function of *C22orf34* remains to be clarified, the gene expression of *C22orf34* in human lung is higher compared with other human tissues according to the Genotype-Tissue Expression (GTEx) Analysis Release V7 [34]. To effectively clarify the mechanism of DIILD, we performed expression quantitative trait locus (cis-eQTL) analysis for rs35198919 and observed a cis-regulatory effect on *C22orf34* expression in whole blood and



**Fig 1. Regional association plots of two SNVs associated with DIILD.** Upper panel;  $P$  values of the SNVs are plotted (as  $-\log_{10}P$  values) against their physical location on chromosome 22 (A) and 20 (B). The genetic recombination rates estimated from the 1,000 Genomes samples

(Japanese in Tokyo, JPT + Chinese in Beijing, CHB) are shown with a blue line. SNV color indicates linkage disequilibrium (LD) with rs35198919 (A) and rs12625311 (B) according to a scale from  $r^2 = 0$  to 1 based on pairwise  $r^2$  values from the Asian (ASN) data from the 1,000 Genomes Project. Lower panel; gene annotations from the University of California Santa Cruz genome browser.

<https://doi.org/10.1371/journal.pone.0223371.g001>

thyroid ( $P = 5.80 \times 10^{-6}$  and  $6.50 \times 10^{-6}$ , respectively, S1 Fig). However, further studies are required to clarify the functional associations of *C22orf34* with DIILD.

rs12625311 demonstrated a stronger association in the combined study and is located in intron 1 of *TSHZ2*, which encodes a transcriptional repressor that suppresses the transcriptional activity of genes involved in tumor development and growth [35, 36]. Single nucleotide polymorphisms (SNPs) in *TSHZ2* are known to be involved in Stevens-Johnson syndrome/toxic epidermal necrolysis with severe ocular complications (SJS/TEN with SOCs) in the Japanese population [37]. ILD has been reported to be one of the complications in patients with SJS [38, 39], suggesting a common mechanism between SJS and DIILD involving a pathway regulated by *TSHZ2*. However, further validation studies and functional analyses are needed to evaluate the true association. DIILD, including EGFR-TKI-induced ILD, is more common in the Japanese population [40, 41]. The frequency of the risk allele (C) in rs12625311 is higher in the Japanese population (36%) compared with other ethnic populations (South Asian: 19%, African: 18%, American: 3% and European: 2% in the 1000 Genomes Project), suggesting that the difference in the above allele frequency may cause interethnic differences in the frequency of DIILD.

**Table 3. Association results of seven SNVs with  $P < 0.00001$  in the EGFR-TKI-induced ILD subgroup analysis.**

Chr.	SNP ID	Position <sup>a</sup>	Gene	Allele (1/2) <sup>b</sup>	Location	Risk allele	EGFR-TKI-induced ILD										nonEGFR-TKI-induced ILD			
							Case					Control					P value	Odds ratio [95% CI]	P value	Odds ratio [95% CI]
							Allele		Allele Frequency		RAF	Allele		Allele Frequency		RAF				
							1	2	1	2		1	2	1	2					
5	rs75399069	24,664,646	Intergenic	A/C		C	19	7	0.73	0.27	0.27	6,921	171	0.98	0.02	0.02	2.39 x 10 <sup>-6</sup>	14.91 [6.19–35.94]	4.71 x 10 <sup>-1</sup>	1.62 [0.22–12.02]
5	rs417168	135,216,860	<i>SLC25A48</i>	T/C	Intron	C	23	3	0.88	0.12	0.12	7,086	6	0.999	0.001	0.001	3.58 x 10 <sup>-6</sup>	154.04 [36.31–653.49]	1.00	0
5	rs442281	135,216,861	<i>SLC25A48</i>	G/A	Intron	A	23	3	0.88	0.12	0.12	7,086	6	0.999	0.001	0.001	3.58 x 10 <sup>-6</sup>	154.04 [36.31–653.49]	1.00	0
20	rs17690253	16,894,293	Intergenic	T/G		G	19	7	0.73	0.27	0.27	6,892	200	0.97	0.03	0.03	6.53 x 10 <sup>-6</sup>	12.70 [5.28–30.54]	1.00	0
6	rs184448987	103,169,971	Intergenic	C/A		A	21	5	0.81	0.19	0.19	7,023	73	0.99	0.01	0.01	7.61 x 10 <sup>-6</sup>	22.91 [8.41–62.40]	1.00	0
2	rs10165147	154,697,039	Intergenic	C/G		G	9	17	0.35	0.65	0.65	5,412	1,686	0.76	0.24	0.24	8.22 x 10 <sup>-6</sup>	6.06 [2.70–13.63]	2.45 x 10 <sup>-1</sup>	1.70 [0.76–3.82]
2	rs1348851	178,418,677	Intergenic	A/G		A	6	20	0.23	0.77	0.23	134	6,968	0.02	0.98	0.02	8.59 x 10 <sup>-6</sup>	15.60 [6.17–39.47]	9.99 x 10 <sup>-1</sup>	0

Chr., chromosome; RAF, risk allele frequency; CI, confidence interval.

<sup>a</sup> Based on GRCh37 genome assembly.

<sup>b</sup> Reference allele (GRCh37) was defined as allele 1.

<https://doi.org/10.1371/journal.pone.0223371.t003>



We next examined the association of the two SNVs (rs35198919 and rs12625311) in patients with a history of ILD and patients that had never had ILD before. The above two SNVs showed larger odds ratios in patients with a history of ILD (rs35198919: OR = 5.51, rs12625311; OR = 7.42) compared with those that had never had ILD (rs35198919: OR = 3.91, rs12625311; OR = 2.02). These findings suggested that these two SNVs may be strong predictors for DIILD in patients with a history of ILD (S3 Table). We further carried out a subgroup analysis of the tumor type for the two possible SNVs (rs35198919 and rs12625311). Although none of the tumor types showed smaller *P* values compared with those in the original analysis, lung and pancreatic cancers showed larger ORs in rs35198919 and rs12625311, and gastric cancer also showed larger ORs in rs12625311 (S4 Table), suggesting that these SNVs may be specific predictors for the risk of DIILD in patients with lung, pancreatic, and gastric cancers.

In the subgroup analysis of EGFR-TKI-induced ILD, we identified seven candidate SNVs that could be predictive markers for the risk of EGFR-TKI-induced ILD (Table 3). Although none of these SNVs showed a significant association with nonEGFR-TKI-induced ILD, four (rs417168, rs442281, rs75399069, and rs10165147) were included in the 42 candidate SNVs identified in the screening study using the 26 patients with DIILD (S1 Table). Moreover, the associations of these four SNVs in the 13 patients with EGFR-TKI-induced ILD were stronger compared with those of the 26 patients with ILD induced by either EGFR-TKI or nonEGFR-TKI. Hence, these four SNVs may be candidate predictive markers for the risk of EGFR-TKI-induced ILD rather than common predictive markers for the risk of DIILD. Of these four SNVs, rs10165147 is located 31 kb upstream of the polypeptide N-acetylgalactosaminyltransferase 13 (*GALNT13*) (S2 Fig). *GALNT13* is a member of the UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase family, which initiate the O-linked glycosylation of mucins [42]. Mucin 5B, oligomeric mucus/gel-forming (*MUC5B*) has been reported to be one of the candidate genetic markers for ILD [43, 44]. Although further analysis is required to clarify the functional importance of *GALNT13* in DIILD, the difference in efficiency of the O-linked glycosylation of mucins by *GALNT13* may cause interindividual differences in the susceptibility to DIILD. Furthermore, we investigated the association between *MUC5B* and DIILD and found that rs565375327 showed the lowest *P* value [screening cohort (26 patients);  $P = 3.58 \times 10^{-3}$ , EGFR-TKI-induced ILD subgroup (13 patients);  $P = 4.74 \times 10^{-4}$ ] of the SNVs in *MUC5B*, suggesting that *MUC5B* may be a promising predictive marker for the risk of DIILD, particularly for EGFR-TKI-induced ILD as well as *GALNT13*.

Surfactant protein A1 (*SFTPA1*) and surfactant protein A2 (*SFTPA2*) have been suggested as susceptibility genes for IPF [45, 46]. In the current study, SNVs in these candidate genes showed no or weak associations with DIILD ( $P = 2.29 \times 10^{-1}$ – $9.99 \times 10^{-1}$ ). It is reported that gefitinib inhibits cyclin G-associated kinase (GAK), as well as EGFR signaling [47], and kinase deficiency of GAK causes the neonatal death of mice due to respiratory dysfunction [48]. However, no SNVs in *GAK* showed a significant association with EGFR-TKI-induced ILD (EGFR-TKI-induced ILD subgroup;  $P = 5.76 \times 10^{-2}$ – $9.99 \times 10^{-1}$ ) in our study. In addition, although the expression level of Mucin 4, cell surface associated (*MUC4*) has been reported to be increased in IPF [49], we observed no association between *MUC4* and EGFR-TKI-induced ILD ( $P = 1.74 \times 10^{-1}$ – $9.99 \times 10^{-1}$ ). Further validation analysis using a larger number of patients is required to clarify the effects of these variants on the risk of DIILD.

Previous candidate gene approaches have reported the positive association of HLA alleles with DIILD [23, 24]. However, these results are unlikely to be fully replicated, and some of the studies failed to confirm the positive association of DIILD with these HLA alleles. Further validation studies using independent cohorts are needed to confirm the true association of the HLA alleles with DIILD. On the other hand, in our study, we carried out a WGS approach to identify novel candidate variants that might be involved in the unknown mechanism

underlying DIILD. Because the sample size used in our study was not large enough, we could not identify the variants that reached the significance level after Bonferroni correction. Similar to the above candidate genes, further validation studies using larger sample size cohorts are needed to confirm the true association of the variants identified in this study with DIILD. To establish a clinically useful prediction system for the risk of DIILD, the integration of the results from the candidate gene and WGS approaches is important. Furthermore, combination analysis of variants identified using the WGS approach with previously identified variants using candidate gene approach, including HLA alleles, may prove to have cumulative effects on the risk of DIILD.

In conclusion, using a WGS approach, we identified two SNVs as possible genetic susceptibility factors for the risk of DIILD. Although the mechanism underlying DIILD should be further investigated by using a larger number of samples and molecular analysis, our results provided a possibility of prediction of the risk of DIILD that could lead to better prognosis and quality of life for patients with cancer.

## Supporting information

**S1 Fig. Results of eQTL analysis for rs35198919.** Expression quantitative trait locus (eQTL) analysis using GTEx Portal (<https://gtexportal.org/home/>) indicated that the C allele (risk allele) of rs35198919 was significantly associated with decreased expression of *C22orf34* in whole blood ( $P = 5.80 \times 10^{-6}$ ) (A) and thyroid ( $P = 6.50 \times 10^{-6}$ ) (B).  
(TIF)

**S2 Fig. Regional association plot for rs10165147.** Upper panel;  $P$  values of SNVs are plotted ( $-\log_{10}P$  values) against their physical location on chromosome 2q23.3-q24.1. The genetic recombination rates estimated from 1,000 Genomes samples (Japanese in Tokyo, JPT + Chinese in Beijing, CHB) are shown with a blue line. SNV color indicates the linkage disequilibrium (LD) with rs10165147 according to a scale from  $r^2 = 0$  to 1 based on pairwise  $r^2$  values from the Asian (ASN) data in the 1,000 Genomes Project. Lower panel; gene annotations from the University of California Santa Cruz genome browser.  
(TIF)

**S1 Table. Association results of screening, replication, and combined studies for the 42 SNVs with  $P < 0.0001$  in the screening study.**  
(XLSX)

**S2 Table. Replication analysis of rs35198919 and rs12625311 using tolerant controls.**  
(XLSX)

**S3 Table. Subgroup analysis of rs35198919 and rs12625311 according to the history of ILD.**  
(XLSX)

**S4 Table. Subgroup analysis of rs35198919 and rs12625311 by cancer type.**  
(XLSX)

## Acknowledgments

We express our heartfelt gratitude to all the study participants. We thank Ms. Aya Imai, Mr. Satoshi Kasahara, Ms. Marie Muramatsu and Ms. Emi Shigino for their contribution to the collection of clinical information, and all staff at the National Cancer Center Hospital, and National Cancer Center Biobank for their contribution to sample collection.

## Author Contributions

**Conceptualization:** Hitoshi Zembutsu.

**Data curation:** Hidehito Horinouchi, Kouya Shiraishi, Takashi Kohno, Takuji Okusaka, Hideki Ueno, Kenji Tamura, Yuichiro Ohe.

**Formal analysis:** Chihiro Udagawa, Kouya Shiraishi, Takashi Kohno, Hitoshi Zembutsu.

**Resources:** Hidehito Horinouchi, Takuji Okusaka, Hideki Ueno, Kenji Tamura, Yuichiro Ohe.

**Writing – original draft:** Chihiro Udagawa, Hitoshi Zembutsu.

## References

1. Bellamy EA, Husband JE, Blaquiére RM, Law MR. Bleomycin-related lung damage: CT evidence. *Radiology*. 1985; 156(1):155–8. <https://doi.org/10.1148/radiology.156.1.2408293> PMID: 2408293
2. Gurjal A, An T, Valdivieso M, Kalemkerian GP. Etoposide-induced pulmonary toxicity. *Lung cancer (Amsterdam, Netherlands)*. 1999; 26(2):109–12.
3. Yague XH, Soy E, Merino BQ, Puig J, Fabregat MB, Colomer R. Interstitial pneumonitis after oxaliplatin treatment in colorectal cancer. *Clinical & translational oncology: official publication of the Federation of Spanish Oncology Societies and of the National Cancer Institute of Mexico*. 2005; 7(11):515–7.
4. Sawada T, Inokuma S, Sato T, Otsuka T, Saeki Y, Takeuchi T, et al. Leflunomide-induced interstitial lung disease: prevalence and risk factors in Japanese patients with rheumatoid arthritis. *Rheumatology (Oxford, England)*. 2009; 48(9):1069–72.
5. Perez-Alvarez R, Perez-de-Lis M, Diaz-Lagares C, Pego-Reigosa JM, Retamozo S, Bove A, et al. Interstitial lung disease induced or exacerbated by TNF-targeted therapies: analysis of 122 cases. *Seminars in arthritis and rheumatism*. 2011; 41(2):256–64. <https://doi.org/10.1016/j.semarthrit.2010.11.002> PMID: 21277618
6. Koike T, Harigai M, Ishiguro N, Inokuma S, Takei S, Takeuchi T, et al. Safety and effectiveness of adalimumab in Japanese rheumatoid arthritis patients: postmarketing surveillance report of 7740 patients. *Modern rheumatology*. 2014; 24(3):390–8. <https://doi.org/10.3109/14397595.2013.843760> PMID: 24252049
7. Fernandez AB, Karas RH, Alsheikh-Ali AA, Thompson PD. Statins and interstitial lung disease: a systematic review of the literature and of food and drug administration adverse event reports. *Chest*. 2008; 134(4):824–30. <https://doi.org/10.1378/chest.08-0943> PMID: 18689579
8. Yamasaki K, Yatera K, Noguchi S, Oda K, Akata K, Nishida C, et al. Incidence and outcomes of bepridil-induced interstitial pneumonia. *Respiratory medicine*. 2013; 107(12):2088–91. <https://doi.org/10.1016/j.rmed.2013.09.023> PMID: 24140111
9. Kang IS, Kim KJ, Kim Y, Park SH. The diagnostic utility of chest computed tomography scoring for the assessment of amiodarone-induced pulmonary toxicity. *The Korean journal of internal medicine*. 2014; 29(6):746–53. <https://doi.org/10.3904/kjim.2014.29.6.746> PMID: 25378973
10. Skeoch S, Weatherley N, Swift AJ, Oldroyd A, Johns C, Hayton C, et al. Drug-Induced Interstitial Lung Disease: A Systematic Review. *Journal of clinical medicine*. 2018; 7(10).
11. Furukawa H, Oka S, Shimada K, Tsuchiya N, Tohma S. Genetics of Interstitial Lung Disease: Vol de Nuit (Night Flight). *Clinical medicine insights Circulatory, respiratory and pulmonary medicine*. 2015; 9 (Suppl 1):1–7. <https://doi.org/10.4137/CCRPM.S23283> PMID: 26056507
12. Schwaiblmair M, Behr W, Haeckel T, Markl B, Foerg W, Berghaus T. Drug induced interstitial lung disease. *The open respiratory medicine journal*. 2012; 6:63–74. <https://doi.org/10.2174/1874306401206010063> PMID: 22896776
13. Ripley B, Kelil T, Gill R. Deciphering drug-induced interstitial lung disease: A mechanistic approach 2016. 9–18 p.
14. Ishikawa N, Hattori N, Yokoyama A, Tanaka S, Nishino R, Yoshioka K, et al. Usefulness of monitoring the circulating Krebs von den Lungen-6 levels to predict the clinical outcome of patients with advanced nonsmall cell lung cancer treated with epidermal growth factor receptor tyrosine kinase inhibitors. *International journal of cancer*. 2008; 122(11):2612–20. <https://doi.org/10.1002/ijc.23411> PMID: 18324627
15. Qin H, Xu XP, Zou J, Zhao XJ, Wu HW, Zha QF, et al. Krebs von den Lungen-6 associated with chest high-resolution CT score in evaluation severity of patients with interstitial lung disease. *Pulmonology*. 2018.

16. Nakamura K, Kato M, Shukuya T, Mori K, Sekimoto Y, Ihara H, et al. Surfactant protein-D predicts prognosis of interstitial lung disease induced by anticancer agents in advanced lung cancer: a case control study. *BMC cancer*. 2017; 17(1):302. <https://doi.org/10.1186/s12885-017-3285-6> PMID: 28464801
17. Ohnishi H, Yokoyama A, Kondo K, Hamada H, Abe M, Nishimura K, et al. Comparative study of KL-6, surfactant protein-A, surfactant protein-D, and monocyte chemoattractant protein-1 as serum markers for interstitial lung diseases. *American journal of respiratory and critical care medicine*. 2002; 165(3):378–81. <https://doi.org/10.1164/ajrccm.165.3.2107134> PMID: 11818324
18. Minami-Shimmyo Y, Ohe Y, Yamamoto S, Sumi M, Nokihara H, Horinouchi H, et al. Risk factors for treatment-related death associated with chemotherapy and thoracic radiotherapy for lung cancer. *Journal of thoracic oncology: official publication of the International Association for the Study of Lung Cancer*. 2012; 7(1):177–82.
19. Osawa M, Kudoh S, Sakai F, Endo M, Hamaguchi T, Ogino Y, et al. Clinical features and risk factors of panitumumab-induced interstitial lung disease: a postmarketing all-case surveillance study. *International journal of clinical oncology*. 2015; 20(6):1063–71. <https://doi.org/10.1007/s10147-015-0834-3> PMID: 25967287
20. Hamada T, Yasunaga H, Nakai Y, Isayama H, Matsui H, Fushimi K, et al. Interstitial lung disease associated with gemcitabine: A Japanese retrospective cohort study. *Respirology (Carlton, Vic)*. 2016; 21(2):338–43.
21. Tamura M, Saraya T, Fujiwara M, Hiraoka S, Yokoyama T, Yano K, et al. High-resolution computed tomography findings for patients with drug-induced pulmonary toxicity, with special reference to hypersensitivity pneumonitis-like patterns in gemcitabine-induced cases. *The oncologist*. 2013; 18(4):454–9. <https://doi.org/10.1634/theoncologist.2012-0248> PMID: 23404815
22. Johkoh T, Sakai F, Kusumoto M, Arakawa H, Harada R, Ueda M, et al. Association between baseline pulmonary status and interstitial lung disease in patients with non-small-cell lung cancer treated with erlotinib—a cohort study. *Clinical lung cancer*. 2014; 15(6):448–54. <https://doi.org/10.1016/j.clcc.2014.06.003> PMID: 25043209
23. Furukawa H, Oka S, Shimada K, Tsuchiya N, Tohma S. HLA-A\*31:01 and methotrexate-induced interstitial lung disease in Japanese rheumatoid arthritis patients: a multidrug hypersensitivity marker? *Annals of the rheumatic diseases*. 2013; 72(1):153–5. <https://doi.org/10.1136/annrheumdis-2012-201944> PMID: 22887846
24. Nishimura M, Toyoda M, Takenaka K, Imamura Y, Chayahara N, Kiyota N, et al. The combination of HLA-B\*15:01 and DRB1\*15:01 is associated with gemcitabine plus erlotinib-induced interstitial lung disease in patients with advanced pancreatic cancer. *Cancer chemotherapy and pharmacology*. 2016; 77(6):1165–70. <https://doi.org/10.1007/s00280-016-3026-6> PMID: 27100735
25. Daly AK, Donaldson PT, Bhatnagar P, Shen Y, Pe'er I, Floratos A, et al. HLA-B\*5701 genotype is a major determinant of drug-induced liver injury due to flucloxacillin. *Nature genetics*. 2009; 41(7):816–9. <https://doi.org/10.1038/ng.379> PMID: 19483685
26. Lucena MI, Molokhia M, Shen Y, Urban TJ, Aithal GP, Andrade RJ, et al. Susceptibility to amoxicillin-clavulanate-induced liver injury is influenced by multiple HLA class I and II alleles. *Gastroenterology*. 2011; 141(1):338–47. <https://doi.org/10.1053/j.gastro.2011.04.001> PMID: 21570397
27. Behr ER, Ritchie MD, Tanaka T, Kaab S, Crawford DC, Nicoletti P, et al. Genome wide analysis of drug-induced torsades de pointes: lack of common variants with large effect sizes. *PloS one*. 2013; 8(11): e78511. <https://doi.org/10.1371/journal.pone.0078511> PMID: 24223155
28. Nicoletti P, Aithal GP, Bjornsson ES, Andrade RJ, Sawle A, Arrese M, et al. Association of Liver Injury From Specific Drugs, or Groups of Drugs, With Polymorphisms in HLA and Other Genes in a Genome-Wide Association Study. *Gastroenterology*. 2017; 152(5):1078–89. <https://doi.org/10.1053/j.gastro.2016.12.016> PMID: 28043905
29. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. 2011. 2011; 17(1):3.
30. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics (Oxford, England)*. 2009; 25(14):1754–60.
31. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernysky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome research*. 2010; 20(9):1297–303. <https://doi.org/10.1101/gr.107524.110> PMID: 20644199
32. Udagawa C, Nakamura H, Ohnishi H, Tamura K, Shimoi T, Yoshida M, et al. Whole exome sequencing to identify genetic markers for trastuzumab-induced cardiotoxicity. *Cancer science*. 2018; 109(2):446–52. <https://doi.org/10.1111/cas.13471> PMID: 29247589
33. Nagasaki M, Yasuda J, Katsuoka F, Nariai N, Kojima K, Kawai Y, et al. Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. *Nature communications*. 2015; 6:8018. <https://doi.org/10.1038/ncomms9018> PMID: 26292667

34. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature*. 2017; 550(7675):204–13. <https://doi.org/10.1038/nature24277> PMID: 29022597
35. Inaguma S, Riku M, Ito H, Tsunoda T, Ikeda H, Kasai K. GLI1 orchestrates CXCR4/CXCR7 signaling to enhance migration and metastasis of breast cancer cells. *Oncotarget*. 2015; 6(32):33648–57. <https://doi.org/10.18632/oncotarget.5203> PMID: 26413813
36. Riku M, Inaguma S, Ito H, Tsunoda T, Ikeda H, Kasai K. Down-regulation of the zinc-finger homeobox protein TSHZ2 releases GLI1 from the nuclear repressor complex to restore its transcriptional activity during mammary tumorigenesis. *Oncotarget*. 2016; 7(5):5690–701. <https://doi.org/10.18632/oncotarget.6788> PMID: 26744317
37. Ueta M. Cold medicine-related Stevens-Johnson syndrome/toxic epidermal necrolysis with severe ocular complications-phenotypes and genetic predispositions. *Taiwan journal of ophthalmology*. 2016; 6(3):108–18. <https://doi.org/10.1016/j.tjo.2016.06.001> PMID: 29018724
38. Kamata T, Sakamaki F, Fujita H, Urano T, Mori M, Yamaguchi K, et al. Toxic epidermal necrolysis with tracheobronchial and pulmonary complications. *Internal medicine (Tokyo, Japan)*. 1994; 33(4):252–5.
39. Lee T, Bae YJ, Park SK, Park HJ, Kim SH, Cho YS, et al. Severe pneumonia caused by combined infection with *Pneumocystis jirovecii*, parainfluenza virus type 3, cytomegalovirus, and *Aspergillus fumigatus* in a patient with Stevens-Johnson syndrome/toxic epidermal necrolysis. *Acta dermato-venereologica*. 2010; 90(6):625–9. <https://doi.org/10.2340/00015555-0977> PMID: 21057748
40. Cohen MH, Williams GA, Sridhara R, Chen G, McGuinn WD Jr., Morse D, et al. United States Food and Drug Administration Drug Approval summary: Gefitinib (ZD1839; Iressa) tablets. *Clinical cancer research: an official journal of the American Association for Cancer Research*. 2004; 10(4):1212–8.
41. Gemma A, Kudoh S, Ando M, Ohe Y, Nakagawa K, Johkoh T, et al. Final safety and efficacy of erlotinib in the phase 4 POLARSTAR surveillance study of 10 708 Japanese patients with non-small-cell lung cancer. *Cancer science*. 2014; 105(12):1584–90. <https://doi.org/10.1111/cas.12550> PMID: 25287435
42. Zhang Y, Iwasaki H, Wang H, Kudo T, Kalka TB, Hennet T, et al. Cloning and characterization of a new human UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase, designated pp-GalNAc-T13, that is specifically expressed in neurons and synthesizes GalNAc alpha-serine/threonine antigen. *The Journal of biological chemistry*. 2003; 278(1):573–84. <https://doi.org/10.1074/jbc.M203094200> PMID: 12407114
43. Juge PA, Lee JS, Epstein E, Furukawa H, Dobrinskikh E, Gazal S, et al. MUC5B Promoter Variant and Rheumatoid Arthritis with Interstitial Lung Disease. *The New England journal of medicine*. 2018; 379(23):2209–19. <https://doi.org/10.1056/NEJMoa1801562> PMID: 30345907
44. Seibold MA, Wise AL, Speer MC, Steele MP, Brown KK, Loyd JE, et al. A common MUC5B promoter polymorphism and pulmonary fibrosis. *The New England journal of medicine*. 2011; 364(16):1503–12. <https://doi.org/10.1056/NEJMoa1013660> PMID: 21506741
45. Selman M, Lin HM, Montano M, Jenkins AL, Estrada A, Lin Z, et al. Surfactant protein A and B genetic variants predispose to idiopathic pulmonary fibrosis. *Human genetics*. 2003; 113(6):542–50. <https://doi.org/10.1007/s00439-003-1015-4> PMID: 13680361
46. Wang Y, Kuan PJ, Xing C, Cronkhite JT, Torres F, Rosenblatt RL, et al. Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *American journal of human genetics*. 2009; 84(1):52–9. <https://doi.org/10.1016/j.ajhg.2008.11.010> PMID: 19100526
47. Brehmer D, Greff Z, Godl K, Blencke S, Kurtenbach A, Weber M, et al. Cellular targets of gefitinib. *Cancer research*. 2005; 65(2):379–82. PMID: 15695376
48. Tabara H, Naito Y, Ito A, Katsuma A, Sakurai MA, Ohno S, et al. Neonatal lethality in knockout mice expressing the kinase-dead form of the gefitinib target GAK is caused by pulmonary dysfunction. *PLoS one*. 2011; 6(10):e26034. <https://doi.org/10.1371/journal.pone.0026034> PMID: 22022498
49. Ballester B, Milara J, Guizarro R, Morcillo E, Cortijo J. Role of MUC4 in idiopathic pulmonary fibrosis. *European Respiratory Journal*. 2018; 52(suppl 62):PA4798.