Journal of Human Genetics (2015) 60, 597–603
© 2015 The Japan Society of Human Genetics All rights reserved 1434-5161/15
www.nature.com/jhg

ORIGINAL ARTICLE

Germline mutations causing familial lung cancer

Koichi Tomoshige[1,2], Keitaro Matsumoto[1], Tomoshi Tsuchiya[1], Masahiro Oikawa[1], Takuro Miyazaki[1], Naoya Yamasaki[1], Hiroyuki Mishima[2], Akira Kinoshita[2], Toru Kubo[3], Kiyoyasu Fukushima[3], Koh-ichiro Yoshiura[2] and Takeshi Nagayasu[1]

Genetic factors are important in lung cancer, but as most lung cancers are sporadic, little is known about inherited genetic factors. We identified a three-generation family with suspected autosomal dominant inherited lung cancer susceptibility. Sixteen individuals in the family had lung cancer. To identify the gene(s) that cause lung cancer in this pedigree, we extracted DNA from the peripheral blood of three individuals and from the blood of one cancer-free control family member and performed whole-exome sequencing. We identified 41 alterations in 40 genes in all affected family members but not in the unaffected member. These were considered candidate mutations for familial lung cancer. Next, to identify somatic mutations and/or inherited alterations in these 40 genes among sporadic lung cancers, we performed exon target enrichment sequencing using 192 samples from sporadic lung cancer patients. We detected somatic 'candidate' mutations in multiple sporadic lung cancer samples; *MAST1*, *CENPE*, *CACNB2* and *LCT* were the most promising candidate genes. In addition, the *MAST1* gene was located in a putative cancer-linked locus in the pedigree. Our data suggest that several genes act as oncogenic drivers in this family, and that *MAST1* is most likely to cause lung cancer.
Journal of Human Genetics (2015) **60,** 597–603; doi:10.1038/jhg.2015.75; published online 16 July 2015

INTRODUCTION

Cancer is typically characterized as a genomic disorder. It is thought that somatic mutations accumulate with age, some of which drive cancer development, and that germline mutations could explain predisposition to cancer development. Familial lung cancers usually express an autosomal dominant form, and as such, cancer susceptibility is passed down through the generations and the disease has a relatively young onset.[1] Smoking is an environmental predisposing factor; exposure to tobacco smoke causes genetic alterations and is strongly associated with lung cancer.[2,3] So, genetic factors and environmental factors are closely linked with cancer. Familial cancer accounts for 15–20% of total cancers,[1] examples of which include Li–Fraumeni syndrome (OMIM: 151623), hereditary retinoblastoma (OMIM: 180200), familial breast cancer (OMIM: 604370) and Lynch Syndrome (OMIM: 120435). Familial lung cancers, however, are less common.

Previous studies have identified clues as to the genetic factors in lung cancer. One of the most common causative genes is epidermal growth factor receptor (*EGFR*), which is a therapeutically targetable driver mutation in non-small cell lung cancer.[4] Recently, driver mutations in Kirstine rat sarcoma viral oncogene homolog (*KRAS*), human epidermal growth factor receptor 2 (*HER2*) and the echinoderm microtubule-associated protein-like 4–anaplastic lymphoma receptor tyrosine kinase (*EML4-ALK*) fusion gene have been discovered.[5–7] In addition to these crucial mutations, genome-wide association studies have revealed inherited susceptibility variants on chromosome 15q24-25.1,[8,9] 6q23-25,[10–12] and 12q24.[10] In a familial lung cancer study, Liu *et al.*[13] identified that a combination of single-nucleotide polymorphisms in chromosomal regions 5p15.33, 6p21.33, 6q23-25/RGS17 and 15q24-25.1 conferred susceptibility to familial lung cancer. Wang *et al.*[14] suggested that heterozygous mutations in surfactant protein A2 were associated with lung cancer and pulmonary fibrosis in two pedigrees. However, many of the causative genes for familial lung cancer are yet to be identified.

Here, we attempted to identify a genetic factor in lung cancer by investigating a three-generation family with lung cancer susceptibility by whole-exome sequencing (WES). We identified 41 alterations in 40 genes linked to lung cancer development in the family. After somatic mutation screening in 192 sporadic lung cancers, we noted that 'deleterious' somatic mutations in *CENPE* or *MAST1* were also present in multiple lung cancer samples. After considering the nonaffected family members and other branches of the family, we believe that *MAST1* is most likely to be a familial lung cancer-related gene.

MATERIALS AND METHODS

Family

One family containing 17 members (Family N), 16 of whom were cognate and one who was a spouse, registered for this study. The 16 patients comprised nine men and seven women, with an average age of onset of 58 years. Pathological diagnoses of the 12 available cases, including double lung cancer cases, were 10

[1]Division of Surgical Oncology, Department of Translational Medical Sciences, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan; [2]Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, Nagasaki, Japan and [3]Department of Internal Medicine, The Japanese Red Cross Nagasaki Genbaku Isahaya Hospital, Nagasaki, Japan
Correspondence: Dr K Yoshiura, Department of Human Genetics, Nagasaki University Graduate School of Biomedical Sciences, 1-12-4 Sakamoto, Nagasaki 852-8523, Japan.
E-mail: kyoshi@nagasaki-u.ac.jp
Received 26 March 2015; revised 13 May 2015; accepted 2 June 2015; published online 16 July 2015

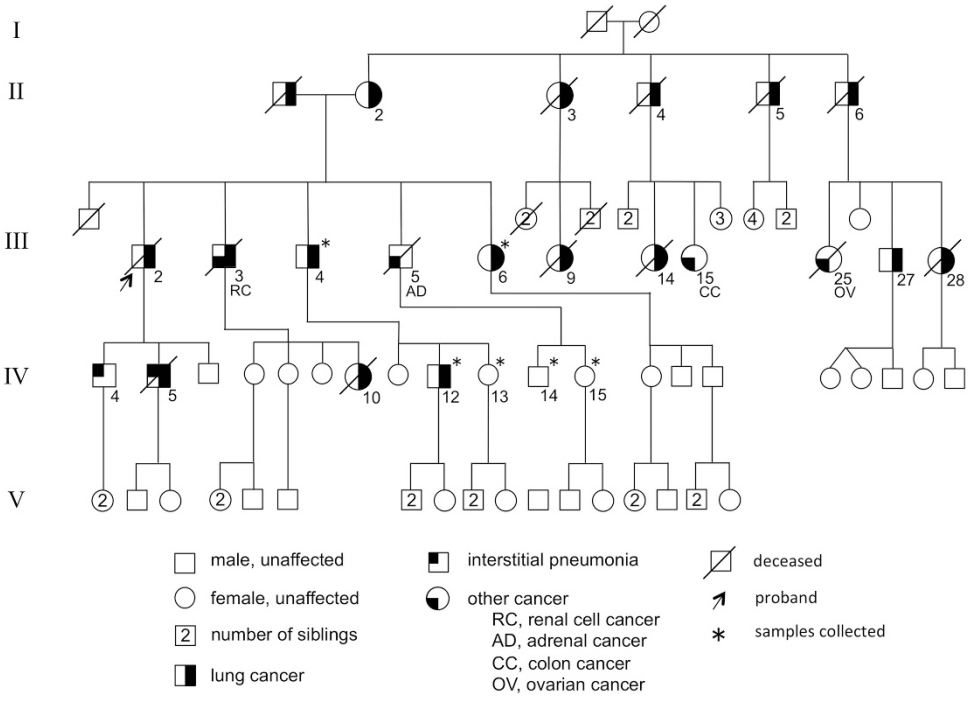

Figure 1 Familial lung cancer pedigree. Seventeen members, 16 of whom were cognate and one who was a spouse, were diagnosed as lung cancer. Three out of the 16 had multicentric lung cancer (III-13, IV-5, IV-12), and one (III-3) also had renal cell cancer. Three members also had other types of cancer (III-5, adrenal cancer; III-14, colon cancer; and III-25: ovarian cancer). Individuals IV-4 and IV-5 had interstitial pneumonia as a respiratory complication.

of adenocarcinoma and two of bronchiolo-alveolar carcinoma. Three of the 16 had multicentric lung cancer (III-13, IV-5, IV-12), and one (III-3) also had renal cell cancer. Three members also had other types of cancer (III-5: adrenal cancer, III-14: colon cancer, III-25: ovarian cancer). Individuals III-5 and III-25 died of adrenal cancer at 59 years, and ovarian cancer at 42 years, respectively. Individuals IV-4 and IV-5 had interstitial pneumonia as a respiratory complication. The age of unaffected control individuals (IV-13,14 and 15) was 58, 56 and 54 years, respectively. There were no consanguineous marriages and no history of exposure to asbestos. The family tree is shown in Figure 1 and clinical information is summarized in Table 1.

Genomic analysis

Peripheral blood was collected from affected individuals III-4, III-6 and IV-12, and from three unaffected control individuals, IV-13, IV-14 and IV-15. Samples from sporadic lung cancer patients were obtained from specimens resected in the Division of Surgical Oncology, Nagasaki University Hospital between 2004 and 2013; control samples were selected from healthy inhabitants of Nagasaki, Japan. DNA was extracted using a QIAamp DNA Mini kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. A Qubit fluorometer (Invitrogen, Carlsbad, CA, USA) was used to assess the concentration and purity of DNA. This study was conducted with the approval of the Genetic and Medical Ethics Commission at Nagasaki University and written consents were obtained from all participants.

Library preparation and whole-exome sequencing (WES)

We performed exon enrichment by hybridization capture using samples from individuals III-4, III-6, IV-12 and IV-14 (peripheral blood) with a SureSelect Human All Exon v4+UTR kit (Agilent, Santa Clara, CA, USA) following the manufacturer's protocol. Sequence data were obtained using a SOLiD5500 (Invitrogen) by 75-bp forward and 50-bp reverse paired-end sequencing. Emulsion PCR for the SOLiD5500 was carried out following the manufacturer's protocol but using KAPA HiFi Taq Polymerase (KAPA Biosystems, Wilmington, MA, USA). Read sequence data were aligned to the hg19 human reference genome using NovoalignCS (Novocraft Technologies Sdn Bhd, Petaling Jaya, Malaysia). NovoalignCS recalibrated the base-quality scores during the alignment. PCR and optical duplications were marked using Picard MarkDuplicates (http://picard.sourceforge.net/) and omitted from subsequent analyses. Reads near to insertions/deletions (INDELs) were locally realigned using the Genome Analysis Toolkit (GATK) IndelRealigner.[15] Single-nucleotide variants (SNVs) and INDELs were detected with GATK's UnifiedGenotyper according to the GATK Best Practice recommendations.[16,17] Detected SNVs and INDELs were annotated using ANNOVAR software (http://www.openbioinformatics.org/annovar/).[18] We selected SNVs and INDELs as candidate variants if they satisfied the following criteria: (1) marked as PASS after GATK VariantFiltration using GATK's recommended conditions; (2) alternative allele frequency $<0.5\%$ in these databases: (a) 69 Genomes Data from Complete Genomics (Mountain View, CA, USA); (b) National Heart, Lung and Blood Institute Grand Opportunity Exome Sequencing Project 6500 (https://esp.gs.washington.edu/drupal/); and (c) 1000 Genomes (http://www.1000genomes.org/); (3) no variation in our in-house data; and (4) not included within the table of segmental duplicated regions downloaded from the University of California Santa Cruz Genome Browser (2011-09-26 update). 'Potentially deleterious mutations' were defined as: (1) nonsynonymous change; or (2) change within 2-bp upstream of a splice acceptor site; or (3) change within 5-bp downstream of a splice donor site. Variants were annotated using following databases: RefSeq; ENSENBL gene; and GENCODE basic v12 and dbSNP135 downloaded from the University of California Santa Cruz Genome Browser at the beginning of this study. To report this study, we used the updated databases GENCODE_basic v19 and dbSNP138 to annotate variants. Sequence variants detected by WES were validated by capillary sequencing on a Genetic Analyzer 3130xl (Applied Biosystems, Foster City, CA, USA).

Target enrichment sequencing

DNA bait was generated for target resequencing using SureDesign (Agilent), then target enrichment of a DNA fragment of the 69 genes was performed using a SureSelect XT custom kit (Agilent) following the manufacturer's protocol. Using a Hiseq2500 (Illumina, San Diego, CA, USA), we generated two 100-bp paired-end sequences, which were aligned to hg19 using Novoalign (Novocraft Technologies Sdn Bhd). The data were processed in the same way as were the WES data, with a slight modification. In addition to the 'deleterious' criteria for WES data, a judgment of 'deleterious' was only given if the variants

Table 1 Summary of patient characteristics

ID	*Age*	*Gender*	*Location*	*Pathology*	*Complication*	*Others*	*Smoking history*
II-2	51	F	NA	NA			NA
II-3	72	F	NA	NA			NA
II-4	58	M	NA	NA			NA
II-5	66	M	NA	NA			NA
II-6	54	M	RUL	Adenocarcinoma, well differentiated			None
III-2	67	M	RUL	Adenocarcinoma, papillary type			10/Day × 45years
III-3	64	M	LUL	Adenocarcinoma, well differentiated		Synchronous double cancer	10/Day×45years
			Left kidney	Clear cell carcinoma			
III-4	74	M	LLL	Adenocarcinoma, well differentiated	Gout		
III-6	56	F	RUL	Adenocarcinoma, papillary type		Synchronous double cancer	None
			LUL	Adenocarcinoma, papillary type			
III-9	58	F	NA	NA			
III-14	55	F	RUL	Bronchioloalveolar carcinoma			None
III-27	57	M	RUL	Adenocarcinoma, poorly differentiated			30/Day×45years
III-28	49	F	RUL	Adenocarcinoma, well differentiated			30/Day×45years
IV-5	39	M	RUL	Adenocarcinoma, well differentiated	Interstitial pneumonia	Metachronous double cancer	20/Day×45years
	44		LUL	Adenocarcinoma, well differentiated			
IV-10	31	F	NA	NA			
IV-12	46	M	LUL	Bronchioloalveolar carcinoma		Metachronous double cancer	10/Day×45years
	58		LUL	Adenocarcinoma, well differentiated			

Abbreviations: F, female; LUL, left upper lobe; M, male; NA, not available; RUL, right upper lobe.

Table 2 WES data using SureSelect V4 UTRs

	Total (bp)	*Mean*	*Median*	*%⩾5x*	*%⩾10x*	*%⩾20x*
III-4(aff)	5 591 210 926	78.56	61	98.4	95.9	88.2
III-6(aff)	7 218 985 687	101.44	80	98.7	97.3	92.6
IV-12(aff)	7 207 615 334	101.28	82	98.7	97.3	93.1
IV-14(unaff)	9 956 305 555	139.9	110	99	98	95.2

Abbreviations: aff, affected; bp, base pair; unaff, unaffected; WES, whole-exome sequencing.

had an alternative allele frequency $<0.5\%$ in Human Genetic Variation Database (www.genome.med.kyoto-u.ac.jp/SnpDB/).

Capillary sequencing

To validate the next-generation sequencing results, we designed 140 pairs of primers for the somatic mutation candidate sites using PrimerZ[19] (http://genepipe.ngc.sinica.edu.tw/primerz/beginDesign.do) or Primer3Plus[20] (http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi) (Supplementary Table 1). For PCR, 10 ng of genomic DNA was amplified in a 10-μl volume using the following conditions in a T1 thermocycler (Biometra, Göttingen, Germany): 94 °C for 2 min, followed by 35 cycles of 94 °C for 10 s, 60 or 65 °C for 20 s and 68 or 72 °C for 30 s, followed by a final cycle of 68 or 72 °C for 5 min. The reactions were performed with ExTaq HS (Takara Bio, Shiga, Japan) or KOD FX (Toyobo, Osaka, Japan). Samples were sequenced using a BigDye Terminator v3.1 cycle sequencing kit (Applied Biosystems) and separated on a Genetic Analyzer 3130xl (Applied Biosystems). Sequence electropherograms were aligned using ATGC software (Genetyx Corporation, Tokyo, Japan).

RESULTS

Whole-exome sequencing with peripheral blood DNA and tumor DNA from the family

The family tree and clinical characteristics of the family (Family N) are shown in Figure 1 and Table 1, respectively.

The results of WES, targeting ~50 Mbp using a SureSelect Human All Exon v4+UTRs kit are shown in Table 2. The raw data filtering process is summarized in Tables 3A and 3B,. Seventy-one alterations were found in all three affected individuals but not in the unaffected individual IV-14 in DNA from blood. The candidate genes, variations and loci are summarized in Table 4. The 71 alterations comprised 69 SNVs and two short deletions in 69 genes; we annotated them as 'potentially deleterious' and considered them to be candidate mutations/genes for familial lung cancer development. All 71 alterations were confirmed by direct sequencing on a capillary sequencer and were heterozygous. After comparing the 71 variants with the updated GENCODE_basic v19 instead of v12, we selected 41 variants in 40 genes (40 SNVs and one INDEL), as our candidate variants (Table 4) and 30 variants were excluded because those are not on the list of defined genes in GENCODE_basic v19. Two putative deleterious variants were found in *NOTCH1*. One of these genes, *MET*, has been previously associated with several kinds of cancer[21,22]; 39 genes were newly identified as candidates for cancer susceptibility genes.

Direct sequencing in unaffected individuals

We next checked for the presence of these variants by direct sequencing of DNA from the peripheral blood of unaffected individuals IV-13, IV-14 and IV-15. We found that 22 out of 41 variants were not present in individuals IV-13 or IV-15 and were therefore completely linked to lung cancer development in Family N (Table 4). Assuming complete penetrance among the six individuals in this study, many of these 22 alterations were located on 17p13, 19p13 and 19q13, so these loci could be defined as regions linked to lung cancer in Family N. Consider base sequence and map information together from exome analyses in the family, one of the genes including *CLUH*, *TRPV3* and *P2RX5* on 17p13; *MAST1* and *CD97* on 19p13; and *PPP5C* and *EMC10* on 19q13 is most likely the causative candidate gene for this family. However, other genes cannot be excluded from candidates by exome sequence and variant map information.

Target enrichment sequencing of 192 sporadic lung cancers and 192 control samples

We considered that one of the alterations in the 40 genes would act as a driver mutation in the development of lung cancer. Because we

Table 3A Process of SNV filtering from raw data

Total SNVs	*III-4* 143 286	*III-6* 154 479	*IV-12* 170 522	*IV-14* 171 955
Commonly Shared SNPs				
(1) GATK-Unified Genotyper		7694		
(2) Not reported/AAF < 0.5% in complete genomics, ESP6500, or 1000GP		541		
(3) No in-house data base		430		
(4) Nonsynonymous alteration, stop gain, stop loss, splice site		75		
(5) Not in segmental duplication		71		

Abbreviations: GATK, Genome Analysis Toolkit; SNV, single-nucleotide variant; WES, whole-exome sequencing.

Table 3B Process of INDEL filtering from raw data

Total indels	*III-4* 6897	*III-6* 7434	*IV-12* 8400	*IV-14* 8085
Commonly Shared SNPs				
(1) GATK-Unified Genotyper		366		
(2) Not reported/AAF <0.5% in complete genomics, ESP6500 or 1000GP		70		
(3) No in-house data base		54		
(4) Nonsynonymous alteration, stop gain, stop loss, splice site		2		
(5) Not in segmental duplication		2		

Abbreviation: GATK, Genome Analysis Toolkit.

expected that somatic mutations would accumulate in one of the 40 genes, we performed exon target enrichment sequencing in 192 sporadic lung cancer patients and 192 healthy individuals. The breakdown of the pathological diagnosis in the 192 lung cancers was as follows: 117 (60.9%) adenocarcinoma; 48 (25%) squamous cell carcinoma; 10 (5.2%) large cell carcinoma; 8 (4.2%) small cell carcinoma; and 9 (4.6%) other carcinoma.

We used a custom SureSelect target enrichment system to extract mutations found only in the sporadic cancer patients. We use the word 'inherited' variants to mean those that came from the zygote; thus, 'inherited variants' means germline variants or nonsomatic mutations. We identified 69 alterations in the 40 candidate genes in 192 sporadic lung cancers and considered them to be 'deleterious mutations', as detailed in the Materials and Methods.

Twenty-eight changes out of the 69 were confirmed to be somatic mutations by comparing them with sequenced DNA from corresponding normal tissue. All somatic mutations were heterozygous and were not observed recurrently (Table 5). Among the 28 somatic mutations, one was nonsense mutation (*CACNB2*; c.C1380A:p.Y460X) and 27 were nonsynonymous (Table 5). Mutations found in nontumorous regions (considered to be inherited variants) are listed in Table 5. Genes in which two or more somatic mutations were found were: five somatic mutations in *CENPE*; three in *LCT*, *ATG2A* and *MAST1*; and two in *PCDH10*, *MET*, *CACNB2* and *SYMPK*. During validation by capillary sequencing, we noted that the wild-type allele for *GRN* in Sample 127 and that for *ATG2A* in sample 151 were detected in a mosaic state due to loss of heterozygosity, because the peak height of the wild-type allele was very low.

DISCUSSION

In this study, 71 variants were annotated as 'deleterious' by the first screening. After filtering against GENCODE_basic v19, we ultimately selected 41 'inherited' variants as candidates causing familial lung cancer in Family N. All 41 variants were changes occurring heterozygously in all three patients; thus, it is conceivable that loss or gain of function due to any one of the alterations induces lung cancer. Some of the variants have an 'rs number' in DbSNP138 and/or are present in the Japanese population according to the Human Genetic Variation Database. It is less likely that these variants cause lung cancer in this pedigree.

If our 41 variants include a causative mutation for lung cancer, we expected that somatic or inherited mutations would be identified in one of these candidate genes in sporadic lung cancer. To this end, we performed exon target enrichment sequencing. For inherited mutations, we could not conclude which gene is responsible for lung cancer, because we identified many germline alterations but no particular gene showed many mutations. In contrast, for somatic mutations, we identified 28 in 40 candidate genes. More than two somatic mutations were detected in eight genes: *LCT*, *CENPE*, *PCDH10*, *MET*, *CACNB2*, *ATG2A*, *MAST1* and *SYMPK*; these genes may be generally related to cancer development. In particular, *CACNB2* is a very good candidate because an A-to-G variant found in affected family members (chr10: 18690944) was not present in any variant database, and because this gene harbored two somatic mutations including a stop-gain mutation. Similarly, five somatic mutations were found in *CENPE*. However, one *CENPE* variant found in a member of Family N, chr4:104059558 C>T, is also present, albeit rarely, in the Japanese population (alternative allele frequency = 0.00271 in Human Genetic Variation Database) (Table 4). *LCT*, *ATG2A* and *MAST1* each harbored one somatic mutation that was not present in databases of normal variation. In addition, none of our variants was present in the Sanger COSMIC lung cancer database (http://www.sanger.ac.uk/genetics/CGP/cosmic/). The somatic mutational frequency of *CACNB2*, *CENPE*, *LCT*, *ATG2A* and *MAST1* in sporadic lung cancer patients was 1.0% (2/192 samples), 2.6% (5/192), 1.6% (3/192), 1.6% (3/192) and 1.6% (3/192), respectively. In a previous study, mutations in the well-known driver genes, *HER2*, *BRAF*, *PIK3CA*, *AKT1*, *MAP2K1* and *MET*, accounted for <5% of mutations,[23] so it is not surprising that the somatic mutation rate is low among just these five genes.

Considering the genotype of unaffected family members, candidate mutations were located on chromosomes 17p13, 19p13 and 19q13. Although these regions have not been previously implicated in lung cancer by genome-wide association studies,[3,24–26] they could be linked to lung cancer in Family N. Within these three regions, *MAST1* is the most obvious candidate gene. Chromosome 12q24 has been previously linked to lung cancer by genome-wide association studies;[10] however, we did not identify any variants in this region in Family N.

Regarding inherited variants found in sporadic lung cancers and healthy controls, five candidate genes —*CACNB2*, *CENPE*, *LCT*, *ATG2A* and *MAST1*—merit consideration. This is because the count

Table 4 Variants found only in patients III-4, III-6 and IV-12

Gene	*Locus of mutation*	*CytoBand*	*RefGene/EnsGene/GENCODE_basicV12*	*GENCODE_basicV19*	*dbSNP138*	*HGVD*
RNF207	chr1:6270957–6270957; G>A	1p36.31	c.G970A:p.V324I		rs200628076	0.007 783
OSBPL9	chr1:52231523–52231523; C>G	1p32.3	c.C769G:p.L257V	c.C769G:p.L257V	None	0.002 743
DMRTB1	chr1:53927173–53927173; G>A	1p32.3	c.G605A:p.R202H		rs182273822	0.023 55
LCE6A	chr1:152816061–152816061; G>A	1q21.3	c.G65A:p.R22K	c.G65A:p.R22K	None	0.006 352
ATP1A2	chr1:160093019–160093019; G>T	1q23.2	c.G194T:p.R65L		rs187733403	0.01
LCT	chr2:136594615–136594615; T>C	2q21.3	c.A125G:p.N42S	c.A125G:p.N42S	None	None
NEB	chr2:152420386–152420386; C>T	2q23.3	c.G13427A:p.R4476H		rs147159176	0.018 843
TNS1	chr2:218770086–218770086; G>A	2q35	c.C83T:p.P28L		None	None
PNKD	chr2:219209238–219209238; C>A	2q35	c.C857A:p.A286D	c.C857A:p.A286D	None	None
CCDC108	chr2:219892643–219892643; A>G	2q35	c.T1745C:p.F582S		rs201618042	0.019 108
METTL6	chr3:15452789–15452789; C>T	3p25.1	c.G829A:p.V277I	c.G829A:p.V277I	None	None
LRRC3B	chr3:26751192–26751192; G>C	3p24.1	c.G29C:p.R10P	c.G29C:p.R10P	None	None
*CENPE**	chr4:104059558–104059558; C>T	4q24	c.G6253A:p.G2085R	c.G6253A:p.G2085R	None	0.002 71
*PCDH10**	chr4:134072488–134072488; T>A	4q28.3	c.T1193A:p.L398Q	c.T1193A:p.L398Q	None	None
*TRAPPC11**	chr4:184618711–184618711; T>C	4q35.1	c.T2684C:p.V895A	c.T2684C:p.V895A	rs200931036	0.004 08
FAT1	chr4:187538271–187538271; T>A	4q35.2	c.A8969T:p.K2990I		rs74986565	0.005 996
RHOBTB3	chr5:95116072–95116072; T>C	5q15	c.T1399C:p.S467P	c.T1399C:p.S467P	None	None
ERAP1	chr5:96127778–96127780; CAT>-	5q15	c.1304_1306del:p.435_436del	c.1304_1306del:p.435_436del	None	None
FGD2	chr6:36982699–36982699; A>T	6p21.2	c.A914T:p.Q305L		rs79274660	0.019 213
FBXO9	chr6:52945785–52945785; G>C	6p12.1	c.G457C:p.D153H	c.G457C:p.D153H	None	None
DST	chr6:56716270–56716270; G>A	6p12.1	c.C550T:p.H184Y	c.C550T:p.H184Y	None	None
LMBRD1	chr6:70409081–70409081; A>G	6q13	c.T1192C:p.Y398H		rs185334169	0.005 469
KHDC3L	chr6:74072893–74072893; A>T	6q13	c.A245T:p.N82I		rs199912037	0.013 043
*COL10A1**	chr6:116441239–116441239; C>G	6q22.1	c.G2040C:p.M680I	c.G2040C:p.M680I	rs200235459	None
*PON1**	chr7:94940879–94940879; C>A	7q21.3	c.G381T:p.M127I	c.G381T:p.M127I	rs202062288	0.004 468
GS1–259H13.2	chr7:99205373–99205373; T>A	7q22.1	c.T433A:p.F145I		None	0.001 258
STAG3	chr7:99798392–99798392; C>T	7q22.1	c.C1861T:p.H621Y	c.C1861T:p.H621Y	None	0.002 469
MET	chr7:116339990–116339990; A>G	7q31.2	c.A852G:p.I284M	c.A852G:p.I284M	None	None
C5	chr9:123716083–123716083; G>A	9q33.2	c.C4826T:p.A1609V	c.C4826T:p.A1609V	None	None
DBH	chr9:136505086–136505086; C>A	9q34.2	c.C458A:p.T153N	c.C458A:p.T153N	None	None
SDCCAG3	chr9:139304341–139304344; CTGT>-	9q34.3	c.244_247del:p.82_83del		None	None
NOTCH1	chr9:139390896–139390896; C>T	9q34.3	c.G7295A:p.S2432N	c.G7295A:p.S2432N	None	None
NOTCH1	chr9:139400036–139400036; G>A	9q34.3	c.C4312T:p.R1438C	c.C4312T:p.R1438C	None	0.001 179
PROSER2	chr10:11912353–11912353; A>C	10p14	c.A668C:p.E223A	c.A668C:p.E223A	None	0.003 243
RPP38	chr10:15146068–15146068; C>T	10p13	c.C755T:p.A252V	c.C755T:p.A252V	None	0.003 333
CACNB2	chr10:18690944–18690944; A>G	10p12.32	c.A305G:p.Q102R	c.A305G:p.Q102R	None	None
DLG5	chr10:79556239–79556239;C>G	10q22.3	c.G4258C:p.A1420P	c.G4258C:p.A1420P	None	None
SFTPA2	chr10:81316987–81316987;C>T	10q22.3	c.G725A:p.R242Q	c.G725A:p.R242Q	None	None
ANKRD1	chr10:92679979–92679979; G>T	10q23.31	c.C154A:p.P52T	c.C154A:p.P52T	None	None
SFRP5	chr10:99531301–99531301; G>T	10q24.2	c.C290A:p.A97D	c.C290A:p.A97D	rs200715195	None
HIF1AN	chr10:102305787–102305787; C>G	10q24.31	c.C910G:p.L304V	c.C910G:p.L304V	None	None
OR52A5	chr11:5153821–5153821; C>T	11p15.4	c.G52A:p.G18R		rs145258281	0.014 372
ATG2A	chr11:64665803–64665803; G>C	11q13.1	c.C4703G:p.T1568S	c.C4703G:p.T1568S	None	None
PACS1	chr11:65868344–65868344; C>T	11q13.1	c.C61T:p.R21W		rs148162349	None
RPS6KB2	chr11:67200226–67200226; C>A	11q13.2	c.C386A:p.S129X		rs200660106	0.017 889
RPS6KB2	chr11:67200279–67200279; A>T	11q13.2	c.A439T:p.S147C		rs2286622	0.018 336
DDX54	chr12:113617090–113617090; C>T	12q24.13	c.G422A:p.R141Q		None	0.008 097
ATP6V0A2	chr12:124236846–124236846; T>C	12q24.31	c.T2072C:p.I691T		rs144946016	0.007 226
*RNF31**	chr14:24618653–24618653; C>T	14q12	c.C217T:p.P73S	c.C217T:p.P73S	None	0.001 166
EFCAB11	chr14:90302987–90302987; G>C	14q32.11	c.C370G:p.L124V	c.C370G:p.L124V	None	None
*RP11–96O20.4-SQRDL**	chr15:45968331–45968331; C>G	15q21.1	c.C687G:p.F229L		rs148840707	0.025 195
*TSR1**	chr17:2235610–2235610; A>G	17p13.3	c.T1349C:p.I450T		rs147119549	0.008 174
*CLUH**	chr17:2593929–2593929; G>A	17p13.3	c.C3889T:p.P1297S	c.C3889T:p.P1297S	None	None
*TRPV3**	chr17:3433382–3433382;T>G	17p13.2	c.A1133C:p.N378T	c.A1133C:p.N378T	None	0.003 196
*P2RX5**	chr17:3582883–3582883;C>G	17p13.2	splicing variant	splicing variant	None	0.004 579
SREBF1	chr17:17722852–17722852; C>G	17p11.2	c.G801C:p.P267P		rs13306742	0.005 623
EVI2A	chr17:29645838–29645838; G>A	17q11.2	c.C263T:p.T88I		rs190013343	0.014
MYO1D	chr17:31094761–31094761; T>G	17q11.2	c.A724C:p.N242H		rs186742471	0.014 209
GRN	chr17:42430090–42430090; G>A	17q21.31	c.G1166A:p.G389D	c.G1166A:p.G389D	None	0.002 75
*ITGB4**	chr17:73733520–73733520; A>G	17q25.1	c.A1856G:p.K619R		rs56119997	0.017 621
*FBXO15**	chr18:71793304–71793304; T>C	18q22.3	c.A590G:p.N197S		rs79499419	0.008 152
*QTRT1**	chr19:10818296–10818296; A>G	19p13.2	splicing variant		rs200161085	0.005 921
*MAST1**	chr19:12981947–12981947; G>T	19p13.2	c.G3224T:p.R1075L	c.G3224T:p.R1075L	None	None
*CD97**	chr19:14518796–14518796; C>G	19p13.12	c.C2453G:p.S818C	c.C2453G:p.S818C	None	0.001 667
SYMPK	chr19:46345537–46345537; G>A	19q13.32	c.C1058T:p.S353L	c.C1058T:p.S353L	None	None
*PPP5C**	chr19:46857233–46857233; G>A	19q13.32	c.G311A:p.R104Q	c.G311A:p.R104Q	None	None
*PRR12**	chr19:50100273–50100273; C>A	19q13.33	c.C2681A:p.P894H		None	0.01
*ADM5**	chr19:50193431–50193431; G>A	19q13.33	c.G143A:p.R48H		rs2288919	0.015 041
*EMC10**	chr19:50983910–50983910; G>T	19q13.33	c.G475T:p.V159L	c.G475T:p.V159L	None	0.003 256
*SIGLEC9**	chr19:51633253–51633253; A>G	19q13.41	c.A1309G:p.S437G		rs199797698	0.023 451
EFCAB8*	chr20:31494437–31494437; T>C	20q11.21	splicing variant		None	None

'None': rs number or allele frequency not reported in dbSNP138 or GVHD.
'*': Variants not found in three unaffected family members.

Table 5 Variants found in DNA from 192 sporadic lung cancer samples

Gene	*Sample no.*	*Locus of mutation*	*Somatic mutation*	*CytoBand*	*RefGene/EnsGene/GENCODE_basicV19*	*dbSNP138*	*HGVD*
LCE6A	106	chr1:152816061–152816061;G>A	NA	1q21.3	c.G65A:p.R22K	None	0.003 176
LCT	163	chr2:136562377–136562377;C>A	●	2q21.3	c.G2720T:p.R907M	None	None
LCT	103	chr2:136566294–136566294;G>T	●	2q21.3	c.C1919A:p.T640K	None	None
LCT	147	chr2:136567135–136567135;C>A	●	2q21.3	c.G2782T:p.G928C	None	None
LCT	128	chr2:136575285–136575285;C>A		2q21.3	c.G1333T:p.G445C	None	None
PNKD	110	chr2:219206291–219206291;C>T		2q35	c.G101A:p.R34Q	rs375550686	None
METTL6	180	chr3:15455644–15455644;A>G		3p25.1	c.T557C:p.L186S	None	0.003 193
LRRC3B	171	chr3:26751200–26751200;T>C		3p24.1	c.T37C:p.S13P	None	None
LRRC3B	114	chr3:26751456–26751456;A>G	●	3p24.1	c.A293G:p.N98S	None	None
CENPE	180	chr4:104027406–104027406;C>G		4q24	c.G7716C:p.K2572N	None	None
CENPE	149	chr4:104059558–104059558;C>T		4q24	c.G6253A:p.G2085R	None	0.002 71
CENPE	34	chr4:104061178–104061178;A>G	●	4q24	c.T5972C:p.M1991T	None	None
CENPE	140	chr4:104062033–104062033;C>T	●	4q24	c.G5617A:p.E1873K	None	None
CENPE	173	chr4:104065666–104065666;T>C	●	4q24	c.A4892G:p.Q1631R	None	None
CENPE	185	chr4:104067000–104067000;C>G	●	4q24	c.G4324C:p.E1442Q	None	None
CENPE	143	chr4:104074307–104074307;A>G	●	4q24	c.T3059C:p.I1020T	None	None
PCDH10	130	chr4:134071752–134071752;A>C		4q28.3	c.A457C:p.T153P	None	None
PCDH10	160	chr4:134071773–134071773;G>A		4q28.3	c.G478A:p.E160K	None	0.001 372
PCDH10	99	chr4:134072488–134072488;T>A		4q28.3	c.T1193A:p.L398Q	None	None
PCDH10	121	chr4:134073505–134073505;C>T	●	4q28.3	c.C2210T:p.A737V	rs200093699	None
PCDH10	152	chr4:134073562–134073562;G>C	●	4q28.3	c.G2267C:p.C756S	None	None
ERAP1	145	chr5:96124300–96124300;C>G		5q15	c.G1613C:p.R538T	None	0.001 166
ERAP1	7	chr5:96124318–96124318;A>-		5q15	c.1595delT:p.L532fs	None	None
ERAP1	157	chr5:96129625–96129625;T>C		5q15	c.A955G:p.M319V	None	None
ERAP1	158	chr5:96139107–96139107;T>C	●	5q15	c.A523G:p.R175G	None	None
FBXO9	193	chr6:52962562–52962562;G>T	●	6p12.1	c.G1278T:p.K426N	None	None
STAG3	11	chr7:99802295–99802295;G>A		7q22.1	c.G2848A:p.E950K	rs2293483	None
STAG3	128	chr7:99809455–99809455;G>C		7q22.1	c.G3553C:p.E1185Q	None	None
MET	24	chr7:116340263–116340263;C>G	●	7q31.2	c.C1125G:p.N375K	None	0.007 032
MET	100	chr7:116397565–116397565;C>A		7q31.2	c.C1937A:p.T646K	None	0.001 166
MET	68	chr7:116412045–116412045;T>C		7q31.2	splicing variant	None	None
MET	181	chr7:116415119–116415119;T>G		7q31.2	c.T3213G:p.I1071M	None	None
MET	197	chr7:116422152–116422155;GTAA>-		7q31.2	Splicing variant	None	None
MET	133	chr7:116436137–116436137;G>C	●	7q31.2	c.G4132C:p.E1378Q	None	None
C5	5	chr9:123751983–123751983;C>G		9q33.2	c.G3017C:p.G1006A	None	0.001
DBH	164	chr9:136518087–136518087;C>T		9q34.2	c.C1400T:p.T467M	rs13306303	None
NOTCH1	4	chr9:139390545–139390545;C>T		9q34.3	c.G7646A:p.R2549H	None	None
NOTCH1	119	chr9:139391634–139391634;C>A	●	9q34.3	c.G6557T:p.G2186V	None	0.001 497
NOTCH1	8	chr9:139401847–139401847;C>T		9q34.3	c.G3553A:p.D1185N	None	0.063 969
NOTCH1	155	chr9:139402570–139402570;C>T		9q34.3	c.G3347A:p.C1116Y	None	None
PROSER2	120	chr10:11912353–11912353;A>C		10p14	c.A1256C:p.E419A	None	0.003 244
RPP38	100	chr10:15145453–15145453;C>T		10p13	c.C140T:p.T47M	rs41284459	None
RPP38	197	chr10:15145453–15145453;C>T		10p13	c.C140T:p.T47M	rs41284459	None
CACNB2	115	chr10:18827184–18827184;T>G	●	10p12.31	c.T1378G:p.Y460D	None	None
CACNB2	115	chr10:18827186–18827186;C>A	●	10p12.31	c.C1380A:p.Y460X	None	None
DLG5	107	chr10:79567593–79567593;G>A		10q22.3	c.C3737T:p.T1246M	None	None
DLG5	165	chr10:79588722–79588722;T>C		10q22.3	c.A2207G:p.N736S	None	None
SFRP5	1	chr10:99527309–99527309;A>G		10q24.2	c.T916C:p.Y306H	None	None
ATG2A	114	chr11:64662363–64662363;C>A	●	11q13.1	c.G1225T:p.G409C	None	None
ATG2A	161	chr11:64668560–64668560;C>T		11q13.1	c.G4124A:p.R1375Q	None	None
ATG2A	151	chr11:64673625–64673625;C>T	●	11q13.1	c.G3224A:p.R1075H	None	None
ATG2A	3	chr11:64677224–64677224;C>A	●	11q13.1	c.G2036T:p.R679L	rs201771874	None
CLUH	57	chr17:2595902–2595902;A>G		17p13.3	c.T3284C:p.V1095A	None	None
CLUH	120	chr17:2597500–2597500;A>G		17p13.3	c.T2891C:p.I964T	None	None
TRPV3	163	chr17:3431384–3431384;C>G	●	17p13.2	c.G1438C:p.G480R	None	None
TRPV3	161	chr17:3445920–3445920;T>G		17p13.2	c.A539C:p.N180T	None	None
GRN	168	chr17:42429793–42429793;G>A		17q21.31	c.G1027A:p.V343I	None	None
GRN	127	chr17:42429793–42429793;G>A		17q21.31	c.G1027A:p.V343I	None	None
MAST1	3	chr19:12951874–12951874;C>A	●	19p13.2	c.C242A:p.S81Y	None	None
MAST1	147	chr19:12962845–12962845;G>T	●	19p13.2	c.G872T:p.C291F	None	None
MAST1	175	chr19:12962999–12962999;T>A	●	19p13.2	c.T947A:p.V316E	None	None
MAST1	142	chr19:12985451–12985451;C>G		19p13.2	c.C4480G:p.P1494A	None	None
CD97	13	chr19:14507970–14507970;G>A		19p13.12	c.G560A:p.R187H	None	None
CD97	163	chr19:14518739–14518739;A>G	●	19p13.12	c.A2117G:p.K706R	None	0.001 318
SYMPK	193	chr19:46333368–46333368;C>A	●	19q13.32	c.G1693T:p.A565S	None	None
SYMPK	99	chr19:46334651–46334651;G>T	●	19q13.32	c.C1589A:p.T530N	None	None
PPP5C	135	chr19:46850381–46850381;G>C		19q13.32	c.G28C:p.E10Q	None	None
EMC10	192	chr19:50982322–50982322;G>A		19q13.33	c.G296A:p.R99Q	None	None
EMC10	169	chr19:50984167–50984167;G>A		19q13.33	c.G611A:p.R204H	None	None

Abbreviations: HGVD, Human Genetic Variation Database; NA, non-tumor tissue was not available; NC, not confirmed.
Sixty-nine variants were confirmed by capillary sequencing; 28 out of 69 were somatic mutations.
'None': rs number or allele frequency not reported in dbSNP138 or HGVD.
No indication in Somatic mutation column means 'inherited variant'.

of rare variants indicates specificity for lung cancer. We counted the inherited variants (that were not present in databases of normal variation) in sporadic lung cancer cases and healthy controls (Table 5 and Supplementary Table 2). We found 2 and 4 variants in *CENPE*; 0 and 2 in *CACNB2*; 1 and 0 in *LCT*; 1 and 4 in *ATG2A*; and 1 and 0 in *MAST1*, respectively. Variants in *MAST1* and *LCT* are probably very rare in healthy control populations, so somatic mutation in lung cancer patients and variants in Family N might be significant for lung cancer development.

To sum up, it is most likely that the *MAST1*, c.G3224T: p.R1075L mutation is causative for the familial lung cancer in this study, with *CENPE*, *CACNB2* and *LCT* as second-place candidates. All have reported functions concerning tumor development. MAST1, microtubule-associated serine-threonine kinase 1, and in particular its PDZ domain, stabilizes and modulates phosphorylation of the C-terminal phospholipid-binding C2 domain of PTEN.[27] *PTEN*, a tumor suppressor gene, is connected with cancer development by regulating cell growth and apoptosis.[27] In addition, fusion genes

involving *MAST1—ZNF700–MAST1*, *NFIX–MAST1* and *TADA2A–MAST1*—were identified in breast cancer cell lines and tumor samples by transcriptome sequencing, and overexpression of these *MAST1* fusion genes had a proliferative effect both *in vitro* and *in vivo*.[28] It is possible that the *MAST1* mutation in Family N influences signal transduction involving PTEN regulation or increases MAST1 activity, leading to cancer development. We examined immunohistochemical data to know the expression level of MAST1 using formalin-fixed paraffin-embedded samples of normal lung in an affected and unaffected person. There were no significant differences in the expression after immunohistochemical examination.

CENPE, centromere-associated protein-E, is a member of the kinesin family that is a key receptor at the mitotic checkpoint. Inhibition of CENPE has a tumor-suppressive effect, such as tumor cell apoptosis or regression.[29] The mutation site in Family N is within a long flexible alpha-helical coiled-coil region (residues D336–A2471), with the mutation at a site that links two domains, an ATP-binding/microtubule-interacting region and a kinetochore-binding domain.[30] The mutation in Family N, c.G6253A:p.G2085R, may influence the domain and linker structure of CENPE.

LCT, lactase, has also been linked with cancer. A polymorphism in *LCT* influences calcium metabolism in colorectal cancer and is correlated with progression and/or incidence of colorectal cancer.[31] There have not yet been any reports linking CACNB2, calcium channel voltage-dependent beta-2 subunit, with cancer.

We conclude that *MAST1* is possibly a causative gene for familial lung cancer. Further genomic studies of sporadic cases and/or familial cases, and functional assays for mutations in *MAST1*, *CENPE* and *LCT*, are necessary to confirm our findings and reveal a novel gene related to lung cancer development.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

ACKNOWLEDGEMENTS

This work was supported by the Ministry of Health, Labor and Welfare of Japan (KY); the Japan Society for the Promotion of Science Grant-in-Aid for Scientific Research (B) grant number 25293084 (KY), Challenging Exploratory Research grant number 25550033 (KY), Grant-in-Aid for Scientific Research (C) grant number 25462180 (TM) from the Ministry of Education, Culture, Sports, Science and Technology of Japan.

1 Nagy, R., Sweet, K. & Eng, C. Highly penetrant hereditary cancer syndromes. *Oncogene* **23**, 6445–6470 (2004).
2 Amos, C. I., Pinney, S. M., Li, Y., Kupert, E., Lee, J., de Andrade, M. A. *et al.* A susceptibility locus on chromosome 6q greatly increases lung cancer risk among light and never smokers. *Cancer Res.* **70**, 2359–2367 (2010).
3 Hung, R. J., McKay, J. D., Gaborieau, V., Boffetta, P., Hashibe, M., Zaridze, D. *et al.* A susceptibility locus for lung cancer maps to nicotinic acetylcholine receptor subunit genes on 15q25. *Nature* **452**, 633–637 (2008).
4 Lynch, T. J., Bell, D. W., Sordella, R., Gurubhagavatula, S., Okimoto, R. A., Brannigan, B. W. *et al.* Activating mutations in the epidermal growth factor receptor underlying responsiveness of non-small-cell lung cancer to gefitinib. *N. Engl. J. Med.* **350**, 2129–2139 (2004).
5 Riely, G. J., Marks, J. & Pao, W. KRAS mutations in non-small cell lung cancer. *Proc. Am. Thorac. Soc.* **6**, 201–205 (2009).
6 Yamamoto, H., Higasa, K., Sakaguchi, M., Shien, K., Soh, J., Ichimura, K. *et al.* Novel germline mutation in the transmembrane domain of HER2 in familial lung adenocarcinomas. *J. Natl Cancer Inst.* **106**, djt338 (2014).
7 Soda, M., Choi, Y. L., Enomoto, M., Takada, S., Yamashita, Y., Ishikawa, S. *et al.* Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. *Nature* **448**, 561–566 (2007).
8 Liu, Y., Liu, P., Wen, W., James, M. A., Wang, Y., Bailey-Wilson, J. E. *et al.* Haplotype and cell proliferation analyses of candidate lung cancer susceptibility genes on chromosome 15q24-25.1. *Cancer Res.* **69**, 7844–7850 (2009).
9 Liu, P., Vikis, H. G., Wang, D., Lu, Y., Wang, Y., Schwartz, A. G. *et al.* Familial aggregation of common sequence variants on 15q24-25.1 in lung cancer. *J. Natl Cancer Inst.* **100**, 1326–1330 (2008).
10 Fang, S., Pinney, S. M., Bailey-Wilson, J. E., de Andrade, M. A., Li, Y., Kupert, E. *et al.* Ordered subset analysis identifies loci influencing lung cancer risk on chromosomes 6q and 12q. *Cancer Epidemiol. Biomarkers Prev.* **19**, 3157–3166 (2010).
11 Bailey-Wilson, J. E., Amos, C. I., Pinney, S. M., Petersen, G. M., de Andrade, M., Wiest, J. S. *et al.* A major lung cancer susceptibility locus maps to chromosome 6q23-25. *Am. J. Hum. Genet.* **75**, 460–474 (2004).
12 You, M., Wang, D., Liu, P., Vikis, H., James, M., Lu, Y. *et al.* Fine mapping of chromosome 6q23-25 region in familial lung cancer families reveals RGS17 as a likely candidate gene. *Clin. Cancer Res.* **15**, 2666–2674 (2009).
13 Liu, P., Vikis, H. G., Lu, Y., Wang, Y., Schwartz, A. G., Pinney, S. M. *et al.* Cumulative effect of multiple loci on genetic susceptibility to familial lung cancer. *Cancer Epidemiol. Biomarkers Prev.* **19**, 517–524 (2010).
14 Wang, Y., Kuan, P. J., Xing, C., Cronkhite, J. T., Torres, F., Rosenblatt, R. L. *et al.* Genetic defects in surfactant protein A2 are associated with pulmonary fibrosis and lung cancer. *Am. J. Hum. Genet.* **84**, 52–59 (2009).
15 McKenna, A., Hanna, M. & Banks, E. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
16 DePristo, M. A., Banks, E., Poplin, R., Garimella, K. V., Maguire, J. R., Hartl, C. *et al.* A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* **43**, 491–498 (2011).
17 Van der Auwera, G. A., Carneiro, M. O., Hartl, C., Poplin, R., del Angel, G., Levy-Moonshine, A. *et al.* From FastQ data to high-confidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr. Protoc. Bioinformatics* **11**, 1011.10.1–11.10.33 (2013).
18 Wang, K., Li, M. & Hakonarson, H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* **38**, e164 (2010).
19 Tsai, M.-F., Lin, Y.-J., Cheng, Y.-C., Lee, K.-H., Huang, C.-C., Chen, Y.-T. *et al.* PrimerZ: streamlined primer design for promoters, exons and human SNPs. *Nucleic Acids Res.* **35**, W63–W65 (2007).
20 Untergasser, A., Nijveen, H., Rao, X., Bisseling, T., Geurts, R. & Leunissen, J. A. Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Res.* **35**, W71–W74 (2007).
21 Christensen, J. G., Burrows, J. & Salgia, R. c-Met as a target for human cancer and characterization of inhibitors for therapeutic intervention. *Cancer Lett.* **225**, 1–26 (2005).
22 Davis, I. J., McFadden, A. W., Zhang, Y., Coxon, A., Burgess, T. L., Wagner, A. J. *et al.* Identification of the receptor tyrosine kinase c-Met and its ligand, hepatocyte growth factor, as therapeutic targets in clear cell sarcoma. *Cancer Res.* **70**, 639–645 (2010).
23 Pao, W. & Girard, N. New driver mutations in non-small-cell lung cancer. *Lancet Oncol.* **12**, 175–180 (2011).
24 Amos, C. I., Wu, X., Broderick, P., Gorlov, I. P., Gu, J., Eisen, T. *et al.* Genome-wide association scan of tag SNPs identifies a susceptibility locus for lung cancer at 15q25.1. *Nat. Genet.* **40**, 616–622 (2008).
25 McKay, J. D., Hung, R. J., Gaborieau, V., Boffetta, P., Chabrier, A., Byrnes, G. *et al.* Lung cancer susceptibility locus at 5p15.33. *Nat Genet.* **40**, 1404–1406 (2008).
26 Wang, Y., Broderick, P., Webb, E., Wu, X., Vijayakrishnan, J., Matakidou, A. *et al.* Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat. Genet.* **40**, 1407–1409 (2008).
27 Valiente, M., Andrés-Pons, A., Gomar, B., Torres, J., Gil, A., Tapparel, C. *et al.* Binding of PTEN to specific PDZ domains contributes to PTEN protein stability and phosphorylation by microtubule-associated serine/threonine kinases. *J. Biol. Chem.* **280**, 28936–28943 (2005).
28 Robinson, D. R., Kalyana-Sundaram, S., Wu, Y.-M., Shankar, S., Cao, X., Ateeq, B. *et al.* Functionally recurrent rearrangements of the MAST kinase and Notch gene families in breast cancer. *Nat. Med.* **17**, 1646–1651 (2011).
29 Wood, K. W., Lad, L., Luo, L., Qian, X., Knight, S. D., Nevins, N. *et al.* Antitumor activity of an allosteric inhibitor of centromere-associated protein-E. *Proc. Natl Acad. Sci. USA* **107**, 5839–5844 (2010).
30 Mirzaa, G. M., Vitre, B., Carpenter, G., Abramowicz, I., Gleeson, J. G., Paciorkowski, A. R. *et al.* Mutations in CENPE define a novel kinetochore-centromeric mechanism for microcephalic primordial dwarfism. *Hum. Genet.* **133**, 1023–1039 (2014).
31 Bácsi, K., Hitre, E., Kósa, J. P., Horváth, H., Lazáry, A., Lakatos, P. L. *et al.* Effects of the lactase 13910 C/T and calcium-sensor receptor A986S G/T gene polymorphisms on the incidence and recurrence of colorectal cancer in Hungarian population. *BMC Cancer* **8**, 317 (2008).

Supplementary Information accompanies the paper on Journal of Human Genetics website (http://www.nature.com/jhg).