Driver Gene and Novel Mutations in Asbestos-Exposed Lung Adenocarcinoma and Malignant Mesothelioma Detected by Exome Sequencing

Maki-Nevala, Satu

2016-02

Maki-Nevala , S , Sarhadi , V K , Knuuttila , A , Scheinin , I , Ellonen , P , Lagstrom , S , Ronty , M , Kettunen , E , Husgafvel-Pursiainen , K , Wolff , H & Knuutila , S 2016 , ' Driver Gene and Novel Mutations in Asbestos-Exposed Lung Adenocarcinoma and Malignant Mesothelioma Detected by Exome Sequencing ' , Lung , vol. 194 , no. 1 , pp. 125-135 . https://doi.org/10.1007/s0040

http://hdl.handle.net/10138/297141
https://doi.org/10.1007/s00408-015-9814-7

Lung (2016) 194:125–135
DOI 10.1007/s00408-015-9814-7

LUNG CANCER

Driver Gene and Novel Mutations in Asbestos-Exposed Lung Adenocarcinoma and Malignant Mesothelioma Detected by Exome Sequencing

Satu Mäki-Nevala[1] · Virinder Kaur Sarhadi[1] · Aija Knuuttila[2] · Ilari Scheinin[1,3] · Pekka Ellonen[4] · Sonja Lagström[4] · Mikko Rönty[5] · Eeva Kettunen[6] · Kirsti Husgafvel-Pursiainen[6] · Henrik Wolff[6] · Sakari Knuutila[1]

Received: 14 June 2015 / Accepted: 27 September 2015 / Published online: 13 October 2015

Abstract
Background Asbestos is a carcinogen linked to malignant mesothelioma (MM) and lung cancer. Some gene aberrations related to asbestos exposure are recognized, but many associated mutations remain obscure. We performed exome sequencing to determine the association of previously known mutations (driver gene mutations) with asbestos and to identify novel mutations related to asbestos exposure in lung adenocarcinoma (LAC) and MM.
Methods Exome sequencing was performed on DNA from 47 tumor tissues of MM (21) and LAC (26) patients, 27 of whom had been asbestos-exposed (18 MM, 9 LAC). In addition, 9 normal lung/blood samples of LAC were sequenced. Novel mutations identified from exome data were validated by amplicon-based deep sequencing. Driver gene mutations in *BRAF*, *EGFR*, *ERBB2*, *HRAS*, *KRAS*, *MET*, *NRAS, PIK3CA*, *STK11,* and ephrin receptor genes (*EPHA1-8*, *10* and *EPHB1-4*, *6*) were studied for both LAC and MM, and in *BAP1*, *CUL1*, *CDKN2A*, and *NF2* for MM.
Results In asbestos-exposed MM patients, previously non-described *NF2* frameshift mutation (one) and *BAP1* mutations (four) were detected. Exome data mining revealed some genes potentially associated with asbestos exposure, such as *MRPL1* and *SDK1*. *BAP1* and *COPG1* mutations were seen exclusively in MM. Pathogenic *KRAS* mutations were common in LAC patients (42 %), both in non-exposed ($n = 5$) and exposed patients ($n = 6$). Pathogenic *BRAF* mutations were found in two LACs.
Conclusion *BAP1* mutations occurred in asbestos-exposed MM. *MRPL1*, *SDK1*, *SEMA5B*, and *INPP4A* could possibly serve as candidate genes for alterations associated with asbestos exposure. *KRAS* mutations in LAC were not associated with asbestos exposure.

Keywords Asbestos · Mutation · Lung adenocarcinoma · Mesothelioma · Exome sequencing

Electronic supplementary material The online version of this article (doi:10.1007/s00408-015-9814-7) contains supplementary material, which is available to authorized users.

✉ Sakari Knuutila
sakari.knuutila@helsinki.fi

[1] Department of Pathology, Faculty of Medicine, University of Helsinki, P.O. Box 21, 00014 Helsinki, Finland

[2] Department of Pulmonary Medicine, Heart and Lung Center, University of Helsinki and Helsinki University Hospital, P.O. Box 340, 00029 Helsinki, Finland

[3] VU University Medical Center, De Boelelaan 1118, 1081 HZ Amsterdam, The Netherlands

[4] Sequencing Unit, Institute for Molecular Medicine Finland, BM2U, Tukholmankatu 8, 00029 Helsinki, Finland

[5] HUSLAB, Department of Pathology, Helsinki University Central Hospital, P.O. Box 400, 00029 Helsinki, Finland

[6] Finnish Institute of Occupational Health, P.O. Box 40, 00251 Helsinki, Finland

Introduction

Asbestos, which are naturally occurring mineral silicate fibers, are the most important work-related carcinogens being responsible for lung and mesothelial malignancies [1]. Asbestos fibers are inhaled into the deep parts of the lungs, where the fibers can penetrate the pleural space and encounter mesothelial cells [2]. MM has a long latency after the exposure. Thus, despite prohibitions on the use of asbestos in many industrialized countries, new MM cases still represent a major health problem.

Complex chromosomal abnormalities, molecular genetic and epigenetic (methylation, acetylation) alterations, as well as miRNA deregulations are typical features encountered in MM [3–5]. There are some other commonly seen alterations, e.g., either deletions or downregulation in *NF2*, *CDKN2A* and mutations in *BAP1*, and upregulation of *EGFR*, *VEGF*, *BCL2,* and *MET* [5]. Recent studies have indicated that patients with germline BAP1 mutations are more prone to develop asbestos-induced malignant pleural mesothelioma [6, 7]. At present, very little is known about the genomic changes that are associated with asbestos exposure. There is one early cytogenetic study, which did reveal that chromosomal deletions and translocations in the short arm of chromosome 1 and partial or total losses of chromosomes 1 and 4 were significantly associated with a high asbestos fiber count in MM [3].

Occupational asbestos exposure is an important risk factor for lung cancer and all fiber types increase the lung cancer risk [1]. Asbestos in combination with tobacco smoke acts as a co-carcinogen and has activities with the characteristics of both multiplicative and additive factors [1, 8, 9]. The genetic alterations occurring in asbestos-related lung cancer appear to be different from those encountered in tobacco smoke-related lung cancer [9–11]. Gene expression, miRNA, and copy number alteration (CNA) studies have provided evidence that there are differences in genomic alterations between asbestos-exposed and non-exposed lung tumors [12–14]. However, the specific mutations occurring in asbestos-related lung cancer still remain obscure.

We performed exome sequencing with the aim of studying recurrent novel somatic mutations in asbestos-exposed lung adenocarcinoma (LAC) and MM, as they are the largest groups of tumor types related to asbestos exposure, and also to investigate known driver genes for probable pathogenic mutations in these patients.

Materials and Methods

Patients

We selected 26 LAC (9 asbestos-exposed) and 21 epithelioid MM (18 asbestos-exposed) tumor samples for exome sequencing based on asbestos fiber counts (Table 1). Additionally, normal tissue samples (leucocytes or normal lung tissue) from 9 of the LAC patients (3 asbestos-exposed) were also examined. All patients were of Finnish origin and diagnosed and operated in the Hospital District of Helsinki and Uusimaa (HUS), Finland. All samples were collected before any treatments. All MM samples were formalin-fixed, paraffin-embedded (FFPE) tumor tissues, and all tumorous LAC material was from fresh frozen (FF) samples with average tumor content of 60 % (range 10–97 %, 45/47 samples with more than 25 %). The asbestos fiber content of lung tissue in patients not considered as being exposed was set as follows: less than 0.2×10^6/g (of dry lung tissue) and 1.0×10^6/g for MM and LAC, respectively. In the asbestos-exposed group, lung samples contained fibers more than 1.0×10^6/g and 2.0×10^6/g in MM and LAC, respectively. The actual asbestos fiber ranges are listed in Table 1. Ethical permissions for this study were obtained.

Asbestos Fiber Measurement

The asbestos fiber count was performed on normal lung tissue samples, obtained during the operation from the surrounding normal lung tissue, by scanning electron microscopy (SEM) on LAC specimens [15] and by transmission electron microscopy (TEM) on MM samples. The assessment of asbestos fibers in lung tissue was conducted at the Finnish Institute of Occupational Health, Helsinki, according to the standardized protocol [16].

DNA Extraction

DNA was extracted from both FFPE and FF samples by the QIAamp DNA Mini Kit (Qiagen GmbH, Hilden, Germany) according to the manufacturer's instructions. The protocol for FFPE tissue samples included the modifications described in our previous study [17]. The Qubit® fluorometer (Life Technologies, Carlsbad, CA) was used to quantify the isolated DNA.

Exome Sequencing

Exome libraries were prepared from 1–3 μg of each DNA according to NimbleGen SeqCap EZ Exome 2.0 Library SR User's Guide. Sequencing was performed on Illumina's HiSeq sequencer (Illumina, Inc., San Diego, CA, USA). Detailed protocol is described in supplemental file 1.

Validation of Novel Mutations by Amplicon Sequencing

Novel variations seen in the exome sequencing were validated and checked for their somatic/germline origin by PCR amplification of the region of interest, performed on DNA from paired tumor and normal adjacent lung tissue. PCR amplicons were sequenced on Illumina MiSeq instrument (Illumina, Inc., San Diego, CA, USA). A detailed protocol is described in supplemental file 2.

Primary Data Analysis

Primary analysis for exome data was performed by the variant-calling pipeline (VCP) developed in the Finnish

Table 1 Features of the patients included in this study

	MM ($n = 21$) 19 pleural 2 peritoneal	LAC ($n = 26$)
Sample type	FFPE	FF
Thoracoscopic or core-needle biopsies	20	None
Surgical tumor samples	1	26
Gender		
Male, *n*	21	20
Smoking status		
Never-smoker, *n*	9	1
Ever-smoker, *n*	12	25
Smokers, pack-years[b], median (range)	NA	36 (14-105)
Asbestos exposure		
Exposed, *n* (fiber range[a])	18 (2.1–1300)	9 (2.1–72.9)
Non-exposed, *n* (fiber range[a])	3 (<0.2)	17 (0.0–0.3)
Normal paired samples		
Exome sequencing	None	9
Deep sequencing (validation)	6	5

FF fresh frozen, *FFPE* formalin fixed, paraffin embedded, *LAC* lung adenocarcinoma, *MM* malignant mesothelioma

[a] Million fibers per gram of dry lung tissue

[b] Number of years of smoking × average number of packs smoked per day

Institute of Molecular Medicine (FIMM) [18]. VCP uses commonly used sequencing data analysis software combined with their own in-house algorithms. Prior to alignment, the overlapping paired reads were merged into single longer reads using SeqPrep [19]. Exome sequencing data were processed further for quality.

Data obtained from amplicon sequencing were processed with an in-house amplicon pipeline that similarly to VCP utilizes common NGS software combined with in-house algorithms. Bowtie 2 [20] was used for the read alignment to the reference genome of GRCh37 with Ensemble release 70 annotation, SAMtools [21], and BCFtools [22] for variant calling and GATK IndelRealigner [23] for indel calling.

Secondary Data Analysis

Exome Sequencing

For novel somatic mutations associated with asbestos exposure, all single nucleotide variants (SNVs) and small insertion and deletion variants (indels) were combined. We selected novel mutations occurring in the protein coding regions of genes and removed all those which had been recorded in the 1000 Human Genomes project or the NCBI dbSNP database (build 137) or which were present in the exomes of paired normal samples. Two in silico analysis tools, PROVEAN/SIFT, were used for prediction of the effect of the missense variants on the produced protein. Of those, we selected mutations resulting in indel, nonsense, or deleterious/damaging missense mutations, as predicted in in silico by PROVEAN or SIFT analyses [24, 25]. Of those, we selected those mutations or genes mutated exclusively in asbestos-exposed patient samples. We analyzed the exome data according to the most frequently mutated chromosomal positions and the genes involved. Due to the small set of samples, no statistical significance was found, and thus, we set the threshold for recurrent variants/genes as only those occurring in three or more exposed patients. All results obtained by previously described workflow and thresholds were checked by the Integrative Genomics Viewer (IGV) for visualization [26] and NCBI dbSNP (build 142) to remove variants reported in a newly built database.

Further, we selected the genes that are known to be altered in MM and/or LAC according to reports in the literature. Driver gene mutations in *BRAF*, *EGFR*, *ERBB2*, *HRAS*, *KRAS*, *MET*, *NRAS, PIK3CA*, and *STK11* were studied for both LAC and MM and in *BAP1*, *CUL1*, *CDKN2A*, and *NF2* for MM. Moreover, for MM and LAC, we selected ephrin receptor genes *EPHA1-8*, *10* and *EPHB1-4*, *6* based on our previous study of frequently mutated receptor tyrosine kinases (RTKs) in lung cancer [27]. From those, we selected the variants occurring in coding regions and causing nonsense, missense, and indel mutation and occurring less than 2 % in the 1000 Human

Genomes project. NCBI dbSNP build 142 was used for studying SNPs. We performed PROVEAN/SIFT in silico analyses for rare variants and selected those missense variants with deleterious effects predicted by either algorithm [24, 25].

Validation by Deep Sequencing

A bioinformatics pipeline was used for analyzing the data. When a frequency of variant base was 0.5 % of all reads covering a given position, a variant was called. The base frequency was compared to the quality value of the corresponding base. All variants with a frequency ratio of minimum of 0.7 were considered to be true sequence variants. The depth of those variant sequences varied between 212 and 48217, and the frequency ratio was at least 0.83.

Results

Exome sequencing analysis of 21 MM and 26 LAC (9 with paired normal sample) cases resulted in 1504431 variants occurring in the coding region. After removing all variants found in the 1000 Human Genomes projects and/or described in NCBI dbSNP (build 137), and then removing all variants found in normal samples and those predicted as neutral by PROVEAN, a number of variants left were 9448. All variants found in non-exposed group of samples were removed, leaving 3048 variants that were found to occur exclusively in asbestos-exposed samples. In order to detect recurrent mutations associated with asbestos exposures, we selected only those mutations that occurred in three or more cases. For exome sequencing, mean average target coverage was 38.1 (range 12.8–54.1). Mean target coverage was on an average of 36.7 (range 12.8–54.0) in FFPE samples and 39.0 (range 20.3–54.1) in FF samples.

Asbestos-Associated Novel Mutations

We found a recurrent novel mutation in *MRPL1* (Tyr87Cys), which was present in three asbestos-exposed patients. Mutations were predicted as deleterious/damaging by SIFT/PROVEAN analysis and they were not seen in non-exposed LAC or MM samples or paired normal LAC samples. The other genes most commonly (with predicted deleterious protein product) and exclusively mutated in asbestos-exposed patients were *BAP1*, *COPG1*, *INPP4A*, *MBD1*, *SDK1*, *SEMA5B*, *TTLL6*, and *XAB2* (Table 2); of those, mutations in *BAP1* and *COPG1* occurred only in MM patients.

Validation of Novel Mutations by Amplicon Sequencing

Deep sequencing revealed mutations in *BAP1* as somatic, i.e., those were seen in tumor material but not in normal paired material from the same patient (Table 2; Fig. 1). From one patient, normal material was not available, but this mutation was reproducible in the tumor sample.

In addition, the *SDK1* mutation (Gln963Ter) was validated as being somatic (Fig. 1). Moreover, mutations in the following genes were validated in the tumor material which was the only sample material available from those patients: *COPG1* (Cys230Arg), *SEMA5B* (Thr1040Pro), *INPP4A* (Lys954Arg), and *TTLL6* (Glu56 fs). The *MRPL1* (Tyr87Cys) mutation was not seen in one paired normal sample, which supports the somatic nature of the recurrent MRPL1 mutation.

Association of Driver Gene Mutations with Asbestos Exposure

In LAC, a total of 42 % (11/26) harbored the *KRAS* mutation (codons 12, 13 and 61). *KRAS* mutations occurred both in asbestos-exposed ($n = 6$) and non-exposed ($n = 5$) individuals. *BRAF* mutations (codon 469 and 601) were found in two non-exposed patients. We did not detect any of the known activating *EGFR* mutations. One of the *EGFR* mutations (His870Arg) detected has been reported previously (COSM33725). All these *KRAS*, *EGFR*, and *BRAF* mutations were mutually exclusive. No possible deleterious missense, nonsense, or indel alterations in coding regions were detected in *NRAS*, *HRAS*, and *PIK3CA*.

In MM, a *BAP1* mutation was found in four patients, all asbestos-exposed. A single nucleotide deletion in *NF2* was detected in one asbestos-exposed patient. One novel *EGFR* mutation (Pro243Ala) was seen in one asbestos-exposed patient. No likely deleterious missense, nonsense, or indel alterations in coding regions were observed in *BRAF*, *CUL1*, *CDKN2A*, *ERBB2*, *HRAS*, *KRAS*, *MET*, *NRAS*, and *PIK3CA*. The results are presented in Table 3.

Ephrin Receptor Mutations

The ephrin receptor mutations found in this study are shown in Table 4. These were present in both asbestos-exposed and non-exposed patients. Some rare SNPs of *EPHA2* (rs11543934) and *EPHA3* (rs34437982) were detected in our previous study [27]. No normal paired material was sequenced from those patients, so that the somatic nature of those SNPs remains obscure.

Table 2 The most recurrently mutated genes in asbestos-exposed LAC and MM patients detected by exome sequencing

Gene	Mutation	No. of patients	Tumor	Validation status	Comment
MRPL1	NP_064621.3:p.(Tyr87Cys)	3	2 MM 1 LAC	Probable somatic	No mutation in normal tissue (result available for only one normal sample)
COPG1	NP_057212.1:p.(Cys230Arg)	1	MM	Validated in tumor	Normal sample not available
	NP_057212.1:p.(Pro287Leu)	1	MM	Uncertain	Low coverage
	NP_057212.1:p.(Ala344Val)	1	MM	Not validated[a]	
	NP_057212.1:p.(Glu643Ter)	1	MM	Germline	
BAP1	NP_004647.1:p.(Phe170Cys)	1	MM	Somatic	
	NP_004647.1:p.(Asp184Tyr)	1	MM	Somatic	
	NP_004647.1:p.(Glu242Ter)	1	MM	Somatic	
	NP_004647.1:p.(Arg238 fs)	1	MM	Validated in tumor	Normal paired sample not available
SEMA5B	NP_001026872.2:p.(Arg273Leu)	1	LAC	Probable germline	Low variant frequency in both tumor and normal
	NP_001026872.2:p.(Gly700Ser)	1	LAC	Probable somatic	Variant in 22 % of reads in tumor and in 3 % in normal tissue
	NP_001026872.2:p.(Pro1015His)	1	LAC	Uncertain	Low coverage
	NP_001026872.2:p.(Thr1040Pro)	1	MM	Validated in tumor	Normal sample not available
INPP4A	NP_004018.1:p.(Ser712Tyr)	1	LAC	Uncertain	Low coverage in tumor sample; variant not detected in normal
	NP_004018.1:p.(Arg846Cys)	1	MM	Not validated[a]	
	NP_001557.1:p.(Lys954Arg)	1	MM	Validated in tumor	Normal sample not available
MBD1	NP_001191071.1:p.(Arg369Leu)	1	LAC	Not validated[a]	
	NP_001191071.1:p.(Arg369Cys)	1	MM	Not validated[a]	
	NP_001191071.1:p.(Ala80Glu)	1	LAC	Not validated[a]	
SDK1	NP_689957.3:p.(Gln963Ter)	1	LAC	Somatic	
	NP_689957.3:p.(Val965Ala)	1	MM	Not validated[a]	
	NP_689957.3:p.(Thr1772Ile)	1	MM	Not validated[a]	
TTLL6	NP_001124390.1:p.(Glu56 fs)	2	2 MM	Validated in tumor	Normal sample not available
	NP_001124390.1:p.(Asp422Asn)	1	LAC	Not validated[a]	
XAB2	NP_064581.2:p.(Arg138Trp)	1	MM	Germline	
	NP_064581.2:p.(Arg725His)	1	LAC	Not tested	No material available
	NP_064581.2:p.(Gln813Ter)	1	MM	Not validated[a]	

Validation performed by amplicon-based deep sequencing on both normal on tumorous material of the patient, if available

LAC lung adenocarcinoma, *MM* malignant mesothelioma

[a] Variant not detected by amplicon sequencing

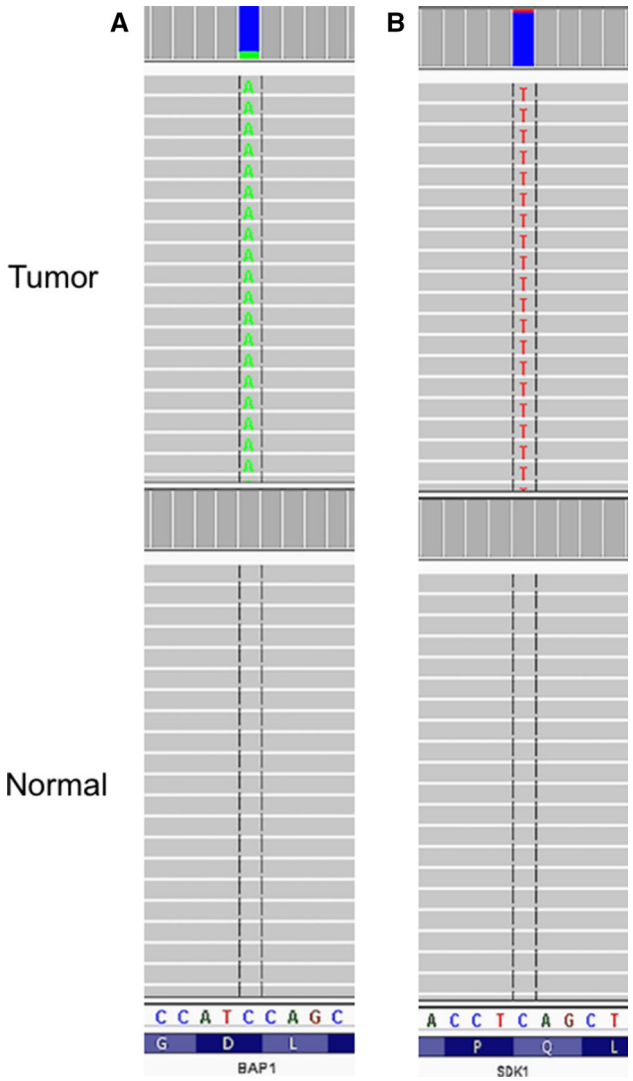

Fig. 1 IGV visualization showing the somatic nature of mutation. **a** BAP1 Asp184Tyr present in asbestos-exposed malignant mesothelioma patient. **b** SDK1 Gln963Ter present in asbestos-exposed lung adenocarcinoma patient

Discussion

Novel Asbestos-Associated Mutations

The exome data mining identified genes *BAP1*, *COPG1*, *INPP4A*, *MBD1*, *SDK1*, *SEMA5B*, *TTLL6*, and *XAB2* as being frequently mutated (at least in three patients) and exclusively in asbestos-exposed patients. After validation with amplicon-based deep sequencing, mutations in *BAP1* and one mutation in *SDK1* (Gln963Ter) could be validated reliably as being somatic. Unfortunately due to the lack of normal tissue and deep-sequencing challenges, somatic status of other candidate mutations remains elusive.

BAP1 and *COPG1* were the most frequently mutated genes seen exclusively in MM; a fact is in line with previous studies reporting *BAP1* mutations in MM. All of the detected *BAP1* mutations occurred in the region coding ubiquitin carboxyl hydrolase (UCH) site of the protein, which is known to be frequently mutated in MM or immediately after that region (five amino acids upstream) [28]. Nonetheless, none of these mutations have been reported previously in MM, although Phe170Cys has been found in kidney (COSM480289). Sporadic, somatic mutations have been found in 20 % of MM [28, 29], and in COSMIC database, the mutation frequency of *BAP1* in MM is 32 %, which are in accordance with our finding. A recent study showed *BAP1* mutations in malignant pleural mesothelioma to be more common in smokers [29]. In the present study, three out of four *BAP1* mutations were found in never-smokers, and one former smoker harbored this mutation.

There are no previous reports of *COPG1* mutations in mesothelioma. COPG1 is a subunit of a coatomer protein complex that is involved in the COPI coat of vesicles during protein transport in the secretory pathway [30]. Little is known about the role of COPI coat vesicles in tumorigenesis or carcinogenesis, and very few somatic mutations in COPG have been described in COSMIC. An elevated expression of COPA, the alpha subunit of coatomer, has been reported in mesothelioma cell lines and COPA knockdown has been associated with a suppression of tumor growth and with the induction of apoptosis [31]. Since COPG1 and COPA are both part of the coatomer protein complex, our finding suggests that the coatomer protein complex might play an important role in MM.

In mesothelioma, it is very difficult to obtain asbestos-non-exposed cases and it is challenging to find sufficient numbers of these rare cases for mutation analyses with adequate statistical power. So, although all of the *BAP1* and *COPG1* mutations occurred in asbestos-exposed MM patients, it is not possible to conclude their exclusive association with asbestos exposure.

Our exome sequencing revealed a novel recurrent mutation in *MRPL1* seen only in asbestos-exposed MM and LAC. MRPL1 is involved in protein synthesis within mitochondria. *MRPL1* is a nuclear gene encoding the 39S subunit of the mitochondrial ribosome. The mutation found in the present study has not been described previously, but another somatic missense mutation in *MRPL1* has been described in two small-cell lung cancers (COSM325848, COSM317641). Furthermore, some mutations have been reported in other cancers, such as in colorectal carcinoma tumors (COSMIC). The possible role of *MRPL1* mutations in tumor biology is still not well understood; we can only speculate that it might be related to aberrant translation of mt-mRNAs derived from all 13 mitochondrial genes, which could affect cell metabolism. In particular, any interference with the production of ROS species is

Table 3 Mutations of the selected genes (for LAC: *BRAF*, *EGFR*, *ERBB2*, *KRAS*, *NRAS*, *HRAS*, *MET*, *PIK3CA*, and *STK11*, for MM: *BAP1*, *BRAF*, *CUL1*, *CDKN2A*, *EGFR*, *ERBB2*, *HRAS*, *KRAS*, *MET*, *NF2*, *NRAS*, *PIK3CA*, and *STK11*) predicted as deleterious by either in silico tool (PROVEAN/SIFT) with frequency >0.02 in the 1000 Human Genomes project, missense, nonsense, and indel mutations included, in LAC and MM patients

Gene	Mutation	No. of Patients (asbestos±)	Somatic [a]	Concurrent Mutations in Gene (AA)	dbSNP 142 rs#	Notes	Variant Reads/ Total Reads (QS)[b]	in silico Prediction PROVEAN/SIFT[c]
KRAS	NP_004976.2:p.Gly12Cys	2 LAC (+,−)	Yes (1 case)	*EPHA2* (R876)	rs121913530	Pathogenic	71/128 (2672); 24/71 (918)	Del/dam
	NP_004976.2:p.Gly12Ser	2 LAC (2+)		*EPHB1* (R222) *ERRB2* (G704) *EPHB4* (G221)	rs121913530	Pathogenic	16/80 (620); 23/52 (910)	Del/dam
	NP_004976.2:p.Gly12Ala	2 LAC (2−)	Yes (1 case)	*EPHB6* (D653)[a] *MET* (T1010) *EPHA2* (E157)	rs121913529	Pathogenic	247/293 (9170); 10/46 (400)	Del/dam
	NP_004976.2:p.Gly12Asp	2 LAC (2+)	Yes (1 case)	*EPHA10* (G260)[a] *STK11* (E293)	rs121913529	Pathogenic	68/133 (2432); 35/51 (1369)	Del/dam
	NP_004976.2:p.Gly13Asp	1 LAC (−)		*EPHA1* (G398) *EPHA2* (R762) *EPHA10* (P70)	rs112445441	Pathogenic	25/70 (955)	Del/dam
	NP_004976.2:p.Gln61His	1 LAC (+)	Yes		rs17851045	Pathogenic	60/89 (2171)	Del/dam
	NP_004976.2:p.Gln61Leu	1 LAC (−)		*EPHA8* (R441)	rs121913240	Pathogenic	55/91 (2013)	
STK11	NP_000446.1:p.(Val34Phe)	1 MM (+)		none			2/8 (75)	Del/dam
	NP_000446.1:p.(Arg74Gly)	1 LAC (-)		*BRAF* (G469) *EPHA2* (P350)			17/24 (594)	Del/dam
	NP_000446.1:p.(Glu293Ter)	1 LAC (+)		*KRAS* (G12) *EPHB3* (D785)	rs398123405	Pathogenic	4/8 (157)	Del/dam
BAP1	NP_004647.1:p.(Phe170Cys)	1 MM (+)		None		COSM480289 (kidney)	4/20 (162)	Del/dam
	NP_004647.1:p.(Asp184Tyr)	1 MM (+)		None			3/13 (111)	Del/dam
	NP_004647.1:p.(Glu242Ter)	1 MM (+)		None			10/30 (370)	Del/dam
	NP_004647.1:p.(Arg238 fs)	1 MM (+)		None			8/19 (128)	NA
BRAF	NP_004324.2:p.Gly469Val	1 LAC (-)		*STK11* (R74) *EPHA2* (P350)	rs121913355	Pathogenic	14/49 (531)	Del/dam
	NP_004324.2:p.Lys601Glu	1 LAC (-)	Yes	*STK11*(V34)	rs121913364	Pathogenic	23/108 (867)	Del/dam
MET	NP_001120972.1:p.(Thr1010Ile)	1 LAC (−)		*KRAS* (G12) *EPHA2* (E157)	rs56391007	Somatic; COSM707 (lung and others)	11/23 (439)	Del/dam
	NP_001120972.1:p.(Tyr1021His)	1 LAC (−)		*EPHA3* (A777) *EPHA8* (A611)		Other mutations in the same codon, e.g. COSM598583 (lung)	11/43 (407)	Del/dam

Table 3 continued

Gene	Mutation	No. of Patients (asbestos±)	Somatic [a]	Concurrent Mutations in Gene (AA)	dbSNP 142 rs#	Notes	Variant Reads/ Total Reads (QS)[b]	in silico Prediction PROVEAN/SIFT[c]
EGFR	NP_005219.2:p.(Pro243Ala)	1 MM (+)		None			6/10 (199)	Del/dam
	NP_005219.2:p.(His870Arg)	1 LAC (−)		*EPHA2* (R876)		COSM33725 (lung)	7/22 (272)	Del/dam
ERBB2	NP_004439.2:p.(Gly704Arg)	1 LAC (+)		*KRAS* (G12)		COSM3378168 (pancreas)	2/8 (71)	Del/dam
				EPHB4 (G221)				
NF2	NP_000259.1:p.(Leu140fs)	1 MM (+)		None			9/39 (190)	NA

[a] Confirmed, if normal paired sample exome sequenced

[b] *QS* a phred quality score for the variant

[c] *Del/dam* deleterious/damaging by PROVEAN/SIFT analysis [24, 25]

AA amino acid residue, *LAC* lung adenocarcinoma, *MM* malignant mesothelioma

intriguing in asbestos-related cancer. Mutations in mt-rRNA genes are probably the most important group for pathogenic variations in mitochondria, but confirmation of pathogenicity remains difficult [32].

Our data showed *INPP4A*, *SDK1*, and *SEMA5B* as frequently mutated genes in asbestos-related LAC and MM. *INPP4A* and *SDK1* are related to oxidative stress. *INPP4A* dephosphorylates molecules, which function as second messengers and are important regulators in many signaling pathways. For example, INPP4A is a negative regulator of PI-3/Akt signaling, the dysfunction of which has been reported in many cancerous tissues [33], and its activation can induce oxidative stress [34]. INPP4A has been identified as an asthma candidate gene, and its downregulation has been described in mice with allergic inflamed lungs [35].

SDK1 is an adhesion molecule, which is activated by cellular stress especially in conditions with the reactive oxygen species. In starved cancer cells, *SDK1* is expressed at high levels [36]. Intriguingly, a recent GWAS study showed the *SDK1* gene and the region around the gene to be associated with the risk of malignant mesothelioma in Italian and Australian asbestos-exposed patients [37]. In our study, one somatic *SDK1* mutation was found in an asbestos-exposed LAC patient, which may suggest that *SDK1* may be associated with asbestos exposure, not only in MM but also in other asbestos-related lung malignancies.

SEMA5B belongs to the family of semaphorins. Somatic mutations in *SEMA5B* have been reported previously, but only three of them in lung tumors (COSM3944760, COSM326437, COSM3944757). In the GWAS study of esophageal cancer patients, *SEMA5B* was implicated as being a candidate gene at one susceptibility locus [38].

Association of Driver Gene Mutations to Asbestos Exposure

We found pathogenic *KRAS* mutations (codons 12, 13 and 61) in 42 % of both asbestos-exposed and non-exposed LAC patients, suggesting that these mutations are not linked to exposure to asbestos. The mutation frequency is higher than reported in smokers (34 %) [39], which might be due to the fact that a majority of our patients had heavy smoking history (median pack years 36), and also due to relatively smaller number of cases. One *KRAS* (Gly12Asp) positive patient harbored a concomitant *STK11* (Glu293-Ter) mutation. Similar concomitant *KRAS*/*STK11* mutations were recently reported in an adrenal metastasis from an LAC patient [40]. The *BRAF* mutations were found in two LAC patients and one of these patients harbored also *STK11* mutation.

Table 4 Indel, nonsense, and missense mutations with frequency less than 0.02 in the 1000 Human Genomes project predicted as deleterious by either in silico tool (PROVEAN/SIFT) of ephrin receptor genes, *EPHA1-8*, *10*, and *EPHB1-4*, *6,* in LAC and MM

Gene	Mutation	No. of Patients (asbestos±)	Somatic[a]	dbSNP 142 rs#	Notes	Variant Reads/ Total Reads (QS)[b]	in silico Prediction (PROVEAN/ SIFT)[c]
EPHA1	NP_005223.4:p.(Gly398Trp)	1 LAC (−)				2/10 (81)	Del/dam
EPHA2	NP_004422.2:p.(Glu157Lys)	1 LAC (−)				2/8 (71)	Neut/dam
	NP_004422.2:p.(Pro350Thr)	1 LAC (−)		rs11543934	Previous study 2 cases[d]	4/7 (121)	Del/dam
	NP_004422.2:p.(Arg762Ser)	1 LAC (−)			Other mutation in the same codon COSM3782397 (prostate)	2/10 (76)	Del/dam
	NP_004422.2:p.(Arg876His)	2 LAC (+, −)		rs35903225		13/28 (462); 11/27 (420)	Del/dam
EPHA3	NP_005224.2:p.(Tyr278Asn)	1 LAC (+)			Other mutation in the same codon COSM1538635 (lung)	26/65 (973)	Del/dam
	NP_005224.2:p.(Ala777Gly)	1 LAC (−)		rs34437982	Previous study 3 cases[c]	37/87 (1429)	Neut/dam
EPHA6	NP_001265229.1:p.(Trp18Arg)	1 LAC (+)				29/86 (1077)	Neut/dam
EPHA8	NP_065387.1:p.(Arg441Gln)	1 LAC (−)		rs146978261		3/7 (103)	Neut/dam
	NP_001006944.1:p.(Ala611Ser)	1 LAC (−)				2/9 (76)	Neut/dam
EPHA10	NP_001092909.1:p.(Pro70His)	1 LAC (−)			COSM341849 (lung)	2/8 (66)	Del/dam
	NP_001092909.1:p.(Gly260Val)	1 LAC (+)	Yes			2/9 (73)	Del/dam
	NP_001092909.1:p.(Leu472Met)	1 LAC (−)	Yes			2/8 (61)	Neut/dam
EPHB1	NP_004432.1:p.(Arg222Trp)	1 LAC (+)			Other mutation in the same codon COSM260704 (large intestine, skin)	21/68 (787)	Del/dam
	NP_004432.1:p.(Glu335Lys)	1 MM (+)				2/8 (74)	Del/dam
	NP_004432.1:p.(Arg470Trp)	1 MM (+)		rs202048188		5/15 (200)	Del/dam
	NP_004432.1:p.(Leu843Met)	1 MM (+)				2/7 (192)	Neut/dam
	NP_004432.1:p.(Thr981Met)	1 LAC (+)	Yes	rs56186270		40/91 (1468)	Neut/dam
EPHB2	NP_059145.2:p.(Ala783Val)	1 MM (+)				2/7 (197)	Neut/dam
EPHB3	NP_004434.2:p.(Asp785Asn)	1 LAC (+)				2/7 (71)	Del/dam
EPHB4	NP_004435.3:p.(Gly221Ser)	1 LAC (+)				2/9 (63)	Del/dam
EPHB6	NP_004436.4:p.(Asp653fs)	1 LAC (−)	Yes			6/34 (102)	NA

LAC lung adenocarcinoma, *MM* malignant mesothelioma

[a] Confirmed, if normal paired sample exome sequenced

[b] *QS* a phred quality score for the variant

[c] *Del/dam* deleterious/damaging, *neut* neutral by PROVEAN/SIFT analysis [24, 25]

[d] Found in our previous study [27]

None of the activating *EGFR* mutations were detected, which we believe might be due to the fact that nearly all our LAC patients had a history of smoking [39, 41]. *EGFR* mutation (His870Arg) was found in a case without smoking history. Two *MET* mutations were detected, both occurring in non-exposed patients with smoking history. Both mutations have been reported in lung tumor in COSMIC. There is clinical interest for MET mutations, but

no clear clinical relevance has been defined as yet, as the right biomarkers for anti-MET therapy remain obscure [42].

Mutations in ephrin receptor genes were seen in both asbestos-exposed and non-exposed patients. We also detected two rare variants that had been observed also in our previous study [27]. We found that ephrin receptors were not only recurrently mutated in LAC but also in MM, especially *EPHB1* (with three mutations). However, their somatic status still remains elusive.

By conducting a detailed study of exomes from asbestos-exposed and non-exposed LAC and MM patients, we were able to identify mutations that were seen only in the exposed group. While mutations in *BAP1* have been reported previously, the identification of novel recurrent mutations/mutated genes is important discoveries and can aid in future studies of asbestos-associated biomarkers. Mutations in known driver genes, such as *KRAS* and *BRAF* mutations, are not associated with asbestos exposure and were detected in lung cancer, as may be expected. Mutations in both of these driver genes showed a putative association with smoking but not with asbestos.

Acknowledgments We thank Päivi Tuominen, Jaana Kierikki, Helinä Hämäläinen, Sauli Savukoski, Tuula Suitiala, Finnish Institute of Occupational Health, and Milja Tikkanen and Tiina Wirtanen, University of Helsinki, for excellent technical assistance. We are also grateful to Ewen MacDonald for the correction of grammar and style. This work was funded by the Finnish Work Environment Fund (no. 112268 to SK; 112269 to HW; 111100 to KHP), Sigrid Jusélius Foundation, Cancer Society of Finland (11/13/2013 to SK; 11/12/2014 to KHP).

Compliance with Ethical Standards

Conflicts of interest Aija Knuuttila received payment for consultancy from Pfizer, Boehringer-Ingelheim, Roche, BMS and for lectures, including service on speakers bureaus, from Pfizer, Lilly, BMS. All other authors declare that they do not have any conflict of interest.

References

1. IARC (2012) IARC Monographs on the Evaluation of Carcinogenic Risks to Humans. Arsenic, metals, fibres and dusts, A review of human carcinogens, p 100C
2. Kane A, Jean D, Knuutila S et al (2014) Malignant mesothelioma: mechanism of carcinogenesis. In: Anttila S, Boffetta P (eds) Occupational cancers. Springer, London, p 299
3. Tiainen M, Tammilehto L, Rautonen J et al (1989) Chromosomal abnormalities and their correlations with asbestos exposure and survival in patients with mesothelioma. Br J Cancer 60:618–626
4. Guled M, Lahti L, Lindholm PM et al (2009) CDKN2A, NF2, and JUN are dysregulated among other genes by miRNAs in malignant mesothelioma -A miRNA microarray analysis. Genes Chromosomes Cancer 48:615–623. doi:10.1002/gcc.20669
5. Kettunen E, Knuutila S (2014) Malignant mesothelioma: molecular markers. In: Anttila S, Boffetta P (eds) Occupational cancers. Springer, London, p 325
6. Xu J, Kadariya Y, Cheung M et al (2014) Germline mutation of Bap1 accelerates development of asbestos-induced malignant mesothelioma. Cancer Res 74:4388–4397. doi:10.1158/0008-5472.CAN-14-1328
7. Betti M, Casalone E, Ferrante D et al (2015) Inference on germline BAP1 mutations and asbestos exposure from the analysis of familial and sporadic mesothelioma in a high-risk area. Genes Chromosom Cancer 54:51–62. doi:10.1002/gcc.22218
8. Nielsen LS, Baelum J, Rasmussen J et al (2014) Occupational asbestos exposure and lung cancer–a systematic review of the literature. Arch Environ Occup Health 69:191–206. doi:10.1080/19338244.2013.863752
9. Inamura K, Ninomiya H, Nomura K et al (2014) Combined effects of asbestos and cigarette smoke on the development of lung adenocarcinoma: different carcinogens may cause different genomic changes. Oncol Rep 32:475–482. doi:10.3892/or.2014.3263
10. Husgafvel-Pursiainen K, Hackman P, Ridanpaa M et al (1993) K-ras mutations in human adenocarcinoma of the lung: association with smoking and occupational exposure to asbestos. Int J Cancer 53:250–256
11. Andujar P, Wang J, Descatha A et al (2010) p16INK4A inactivation mechanisms in non-small-cell lung cancer patients occupationally exposed to asbestos. Lung Cancer 67:23–30. doi:10.1016/j.lungcan.2009.03.018
12. Wikman H, Ruosaari S, Nymark P et al (2007) Gene expression and copy number profiling suggests the importance of allelic imbalance in 19p in asbestos-associated lung cancer. Oncogene 26:4730–4737 **1210270**
13. Nymark P, Guled M, Borze I et al (2011) Integrative analysis of microRNA, mRNA and aCGH data reveals asbestos- and histology-related changes in lung cancer. Genes Chromosom Cancer 50:585–597. doi:10.1002/gcc.20880
14. Nymark P, Aavikko M, Makila J et al (2013) Accumulation of genomic alterations in 2p16, 9q33.1 and 19p13 in lung tumours of asbestos-exposed patients. Mol Oncol 7:29–40. doi:10.1016/j.molonc.2012.07.006
15. Karjalainen A, Anttila S, Heikkilä L et al (1993) Asbestos exposure among Finnish lung cancer patients: occupational history and fiber concentration in lung tissue. Am J Ind Med 23:461–471
16. Tuomi T (1992) Fibrous minerals in the lungs of mesothelioma patients: comparison between data on SEM, TEM, and personal interview information. Am J Ind Med 21:155–162
17. Tuononen K, Maki-Nevala S, Sarhadi VK et al (2013) Comparison of targeted next-generation sequencing (NGS) and real-time PCR in the detection of EGFR, KRAS, and BRAF mutations on formalin-fixed, paraffin-embedded tumor material of non-small cell lung carcinoma-superiority of NGS. Genes Chromosom Cancer 52:503–511. doi:10.1002/gcc.22047
18. Sulonen AM, Ellonen P, Almusa H et al (2011) Comparison of solution-based exome capture methods for next generation sequencing. Genome Biol 12:R94-2011-12-9-r94, DOI: 10.1186/gb-2011-12-9-r94
19. St John J SeqPrep. https://github.com/jstjohn/SeqPrep
20. Langmead B, Salzberg SL (2012) Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. doi:10.1038/nmeth.1923
21. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and samtools. Bioinformatics 25:2078–2079. doi:10.1093/bioinformatics/btp352
22. BCFTools. http://samtools.github.io/bcftools/

23. DePristo MA, Banks E, Poplin R et al (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet 43:491–498. doi:10.1038/ng.806
24. Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081. doi:10.1038/nprot.2009.86
25. Choi Y, Sims GE, Murphy S et al (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS One 7:e46688. doi:10.1371/journal.pone.0046688
26. Thorvaldsdottir H, Robinson JT, Mesirov JP (2013) Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Brief Bioinform 14:178–192. doi:10.1093/bib/bbs017
27. Maki-Nevala S, Kaur Sarhadi V, Tuononen K et al (2013) Mutated ephrin receptor genes in non-small cell lung carcinoma and their occurrence with driver mutations-targeted resequencing study on formalin-fixed, paraffin-embedded tumor material of 81 patients. Genes Chromosom Cancer 52:1141–1149. doi:10.1002/gcc.22109
28. Bott M, Brevet M, Taylor BS et al (2011) The nuclear deubiquitinase BAP1 is commonly inactivated by somatic mutations and 3p21.1 losses in malignant pleural mesothelioma. Nat Genet 43:668–672. doi:10.1038/ng.855
29. Zauderer MG, Bott M, McMillan R et al (2013) Clinical characteristics of patients with malignant pleural mesothelioma harboring somatic BAP1 mutations. J Thorac Oncol 8:1430–1433. doi:10.1097/JTO.0b013e31829e7ef9
30. Hahn Y, Lee YJ, Yun JH et al (2000) Duplication of genes encoding non-clathrin coat protein gamma-COP in vertebrate, insect and plant evolution. FEBS Lett 482:31–36
31. Sudo H, Tsuji AB, Sugyo A et al (2010) Knockdown of COPA, identified by loss-of-function screen, induces apoptosis and suppresses tumor growth in mesothelioma mouse model. Genomics 95:210–216. doi:10.1016/j.ygeno.2010.02.002
32. Smith PM, Elson JL, Greaves LC et al (2014) The role of the mitochondrial ribosome in human disease: searching for mutations in 12S mitochondrial rRNA with high disruptive potential. Hum Mol Genet 23:949–967. doi:10.1093/hmg/ddt490
33. Bauer TM, Patel MR, Infante JR (2015) Targeting PI3 kinase in cancer. Pharmacol Ther 146:53–60. doi:10.1016/j.pharmthera.2014.09.006
34. Kim JH, Chu SC, Gramlich JL et al (2005) Activation of the PI3K/mTOR pathway by BCR-ABL contributes to increased production of reactive oxygen species. Blood 105:1717–1723
35. Hakim S, Bertucci MC, Conduit SE et al (2012) Inositol polyphosphate phosphatases in human disease. Curr Top Microbiol Immunol 362:247–314. doi:10.1007/978-94-007-5025-8_12
36. Yoon S, Woo SU, Kang JH et al (2012) NF-kappaB and STAT3 cooperatively induce IL6 in starved cancer cells. Oncogene 31:3467–3481. doi:10.1038/onc.2011.517
37. Cadby G, Mukherjee S, Musk AW et al (2013) A genome-wide association study for malignant mesothelioma risk. Lung Cancer 82:1–8. doi:10.1016/j.lungcan.2013.04.018
38. Wu C, Hu Z, He Z et al (2011) Genome-wide association study identifies three new susceptibility loci for esophageal squamous-cell carcinoma in Chinese populations. Nat Genet 43:679–684. doi:10.1038/ng.849
39. Dogan S, Shen R, Ang DC et al (2012) Molecular epidemiology of EGFR and KRAS mutations in 3,026 lung adenocarcinomas: higher susceptibility of women to smoking-related KRAS-mutant cancers. Clin Cancer Res 18:6169–6177. doi:10.1158/1078-0432.CCR-11-3265
40. Gleeson FC, Kipp BR, Levy MJ et al (2015) Somatic STK11 and concomitant STK11/KRAS mutational frequency in stage IV lung adenocarcinoma adrenal metastases. J Thorac Oncol 10:531–534. doi:10.1097/JTO.0000000000000391
41. Varghese AM, Sima CS, Chaft JE et al (2013) Lungs don't forget: comparison of the KRAS and EGFR mutation profile and survival of collegiate smokers and never smokers with advanced lung cancers. J Thorac Oncol 8:123–125. doi:10.1097/JTO.0b013e31827914ea
42. Rolfo C, Van Der Steen N, Pauwels P et al (2015) Onartuzumab in lung cancer: the fall of Icarus? Expert Rev Anticancer Ther 15:487–489. doi:10.1586/14737140.2015.1031219