Lung Cancer 85 (2014) 168–174

Contents lists available at ScienceDirect

Lung Cancer

journal homepage: www.elsevier.com/locate/lungcan

Oligogenic germline mutations identified in early non-smokers lung adenocarcinoma patients

Alessandra Renieri[a,b,c,*], Maria Antonietta Mencarelli[a], Francesco Cetta[d], Margherita Baldassarri[a,b], Francesca Mari[a,b], Simone Furini[e], Pietro Piu[f], Francesca Ariani[b], Tommaso A. Dragani[g], Elisa Frullanti[a,*]

[a] Genetica Medica, Azienda Ospedaliera Universitaria Senese, Siena, Italy
[b] Medical Genetics, University of Siena, Siena, Italy
[c] Istituto Toscano Tumori, Florence, Italy
[d] IRCCS MultiMedica, Milan, Italy
[e] Department of Medical Biotechnology, University of Siena, Siena, Italy
[f] Department of Medicine, Surgery & Neuroscience, University of Siena, Siena, Italy
[g] Fondazione IRCCS Istituto Nazionale dei Tumori, Milan, Italy

ARTICLE INFO

Article history:
Received 5 March 2014
Accepted 28 May 2014

Keywords:
Oligogenic model
Whole-exome sequencing
Lung cancer susceptibility
Germline variants
Next-generation sequencing
Lung adenocarcinoma

ABSTRACT

Objectives: A polygenic model is commonly assumed for the predisposition to common cancers. With respect to lung cancer, Genome Wide Association Studies (GWAS) have identified three loci at 15q25, 5p15.33, and 6p21. However, the relative risks associated with alleles at these loci are low; in addition, the data are limited to smokers, and have not been quite reproducible.
Materials and methods: In order to investigate genetic susceptibility we have adopted an entirely novel patient selection strategy. First, we have selected for adenocarcinoma (ADCA) histology only; second, we have selected non-smokers; third we have selected patients who developed ADCA of lung before the age of 60 and who had an older unaffected sib: we have identified 31 such sib-pairs. Among them, we selected two patients with very early age at disease onset (37- and 49-years old), and having a healthy sibling available for genome comparison older than at least 7 years.
Results: On germline DNA samples of four subjects of two such pairs we have carried out whole exome sequencing. Truncating mutations were detected in 8 'cancer genes' in one affected, and in 5 cancer genes in the other affected subject: but none in the two healthy sibs ($p = 0.0026$). Some of these mutant genes (such as *BAG6*, *SPEN* and WISP3) are recognized as major cancer players in lung tumors; others have been previously identified in other human cancers (*JAK2*, *TCEB3C*, *NELFE*, *TAF1B*, *EBLN2*), in mouse models (*GON4L*, *NOP58*, and *RBMX*) or in genome-wide association studies (*KIAA2018*, *ZNF311*).
Conclusions: This study identifies for the first time in non-smokers with lung adenocarcinoma specific sets of germline mutations that, together, may predispose to this tumor.
© 2014 Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Despite therapeutic advances in lung cancer, the overall 5-year survival has not improved in the last 20 years and remains only 15% [1]. Cigarette smoking is responsible for the vast majority of lung cancer cases. However, about less than 1 out of 5 smokers develop lung cancer [2] and 10–15% of lung cancer patients have never smoked [3], suggesting that genetic factors are also likely to contribute to the individual genetic susceptibility. This evidence was first suggested more than 50 years ago based on epidemiological evidence for familial aggregation of lung cancer [4].

In mice, there is a complex inheritance of genetic predisposition to lung tumorigenesis [5,6] suggesting the possibility of a polygenic inheritance model [7]. In inbred mouse strains about a dozen of genes within Pulmonary Adenoma susceptibility (*Pas*) and resistance (*Par*) loci regulate differential lung tumor susceptibility [8]. Among these, *Pas1* locus, mapping in the distal region of chromosome 6, contains six genes: *Bcat1*, *Lrmp*, *Casc1*, *Lyrm5*, *Kras*, and *Ifltd1*. *Kras* appears to play a major role in the genetic

* Corresponding authors at: Medical Genetics Unit, Azienda Ospedaliera Universitaria Senese, viale Bracci 2, 53100 Siena, Italy. Tel.: +39 0577 233303; fax: +39 0577 233325.
E-mail addresses: alessandra.renieri@unisi.it (A. Renieri), elisa.frullanti@dbm.unisi.it (E. Frullanti).

http://dx.doi.org/10.1016/j.lungcan.2014.05.020
0169-5002/© 2014 Elsevier Ireland Ltd. All rights reserved.

predisposition to lung cancer in mouse models, being linked to both tumor multiplicity and volume [5]. These data have never been replicated in humans, where similar studies are unreliable and familial aggregation of cases is very rare. Moreover, the search for the human equivalent regions of the *Pas1* locus on chromosome 12 and of the *Par1* locus on chromosome 17q21 has yielded no results.

Genome-wide association studies (GWAS) have been largely used to detect lung cancer susceptibility polymorphisms in humans, but have failed in the identification of a major locus: different GWAS found different associated loci for the same phenotype with relatively low risk (reviewed in [9,10]). The associations most frequently identified, at 15q25, 5p15.33, and 6p21 regions, explained a very low risk (odds ratios (OR) ~1.1–1.3)[11–14]. However, these regions were associated with lung cancer risk in smokers, but not in non-smokers [15–17]. Therefore, it is evident that much of the genetic risk remains unexplained, representing the so-called 'dark matter' of genetic risk or 'missing heritability' [9].

The aim of the present study was to dissect this 'dark matter'. Specifically, we tried to prove the hypothesis that genetic susceptibility to cancer may rely on a restricted number of genes. We used the combination of an advanced technical tool, the exome sequencing, with a new selection strategy. We selected two patients with early onset lung cancer in absence of tobacco smoking and having a healthy sibling older by at least 7 years. The comparison of exome sequences of cases with healthy sib controls would allow identifying germline mutations affecting structure of encoded proteins and, thus, potentially associated with risk of disease.

2. Materials and methods

2.1. Study design

This study involved whole-exome sequencing in two patients with early onset lung adenocarcinoma (cases) and in their healthy older siblings (controls). Because siblings share 50% of the genome, the genetic comparison in discordant sibs can reveal variants and mutations associated with the incidence of cancer. The sib-pairs for this study were identified from a sample collection maintained at Fondazione IRCCS Istituto Nazionale Tumori in Milan (Italy) and containing DNAs and associated clinical data from Italian patients with lung cancer who were diagnosed in the institute. At the time of this study, the collection had samples from 964 patients with lung adenocarcinoma, including 188 non-smokers at the time of diagnosis. For 31 never-smokers, DNA samples and demographic data were also available for a healthy sibling; these samples had been collected with the help of the Marta Nurizzo Association in Brugherio, Milan (Italy).

For the present study, we chose two sib-pairs with the affected sib as young as possible at time of cancer diagnosis and the cancer-free sibling as old as possible. These selection criteria were chosen to maximize phenotypic difference between sibs. It is so likely that in these cases and controls genetic factors could have played a role in cancer susceptibility or resistance.

To test the hypothesis that cancer susceptibility depends on restricted number of genes, we analyzed the coding part of the entire genome of affected individuals and may a comparison of the affected with non-affected sib. The study protocol was approved by the Institutional Ethics Committee of the Istituto Nazionale Tumori in Milan (Italy). All patients and siblings had provided written informed consent for the use of their samples and clinical data for research purposes.

2.2. Whole-exome sequencing

After quality check using the NanodropND-1000 spectrophotometer (ThermoScientific, Wilmington, MA, USA) and gel-electrophotometry, genomic DNA (6 ug per sample) was used to prepare sample libraries. Briefly, DNA was randomly fragmented by Covaris ultrasonicator (Covaris™, Woburn, MA, USA) and adapters were ligated to both ends of the fragments. DNA fragments were amplified by ligation-mediated PCR using NimbleGen SeqCap EZExomePlus 64 M (Roche NimbleGen, Inc., Madison, WI, USA), and hybridized to the NimbleGen Human array for enrichment. Post-library quality controls were performed using the Agilent High-Sensitivity DNA assay on Agilent2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA, USA). Captured libraries were then loaded on Hiseq2000 sequencing platform (Illumina-Life Technologies, Grand Island, USA) for high-throughput sequencing. Finally, the adapter sequence in the resulting raw data was removed, and reads with low-quality (with >10% unspecified bases) and >50% low-quality bases (with base quality <5) were discarded using Illumina Data Analysis Pipeline software v1.8.2. The reads generated on the HiSeq2000 were provided under FASTQ format.

2.3. Read alignment and variants detection

FASTQ sequences were aligned to the human reference genome (NCBI-Build37/hg19) using the Burrows–Wheeler algorithm [18]. The resulting Sequence Alignment and Mapping (SAM) files were then converted to Binary Alignment and Mapping files (BAM) using SAMtools [19]. PCR duplicates were removed with Picard (http://picard.sourceforge.net/). The Genome Analysis-Toolkit [19] was used for base quality recalibration, local realignment around the potential insertion/deletion sites, and for calling variants. Variants were annotated against the NCBI-RefGene database, dbSNP137, and the 1000-Genomes Project (release April 2012). A quality filter was applied to minimize false positives removing variants with low-coverage (occurring in less than 5 total reads).

We carried out two different strategies by which we searched for interesting variants: (i) candidate gene approach, (ii) whole genome "unsupervised" analysis. We started analysis in searching for mutations in candidate genes. We firstly checked if there was involvement of the human homologous regions of *Pas1* and *Par1* loci. Then, we compared our results of whole-exome sequencing to lung cancer susceptibility previously reported loci, namely 15q24–25.1, 5p15.33, and 6p21. Then, we selected a panel of genes known to be affected by somatic alterations in lung cancer (Supplementary Table 1) and we checked for them in exome data. Non-synonymous mutations in candidate genes were evaluated for the effect on protein function using two predictive tools: SIFT (http://sift.jcvi.org/www/SIFT_BLink_submit.html) and PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/index.shtml). We lastly went through a whole genome "unsupervised" analysis. We filtering for exonic mutations, not present in dbSNP137, and present in affected but not in unaffected sibs and vice versa. We identified truncating variants, i.e. frameshift insertion and deletion and nonsense variants that could alter protein function. Genes affected by mutations were checked in literature in order to identify cancer-related genes, i.e. genes that are involved in tumors.

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.lungcan.2014.05.020.

To determine if the genes affected by mutations were involved in common biochemical or biological signatures and interaction networks, we investigated gene ontology (GO) enrichment using the Database for Annotation, Visualization and Integrated Discovery functional annotation tool (DAVID Bioinformatics Resources 6.7, NIAID/NIH, http://david.abcc.ncifcrf.gov/) and the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING; http://string-db.org).

Table 1
Clinical characteristics and sequencing data for two cases with lung adenocarcinoma and their healthy older siblings.

	Case 1	Control 1	Case 1	Control 1
Gender	F	F	M	F
Age at diagnosis, years	49	NA	37	NA
Age at sampling, years	NA	47	NA	42
Smoker status	Never	Never	Never	Never
Histologic type	ADCA	NA	ADCA	NA
Follow-up status	Dead	Alive	Dead	Alive
Age at follow-up	52	56[a]	39	50[a]
Exome sequence data, Gb	71,8	71,3	72,8	71,3
Depth of coverage, fold	34.49	35.11	34.18	34.25
Total sequence variants, *n*[b]	47,639	46,680	47,648	47,421

NA, not applicable.
[a] Cancer-free.
[b] Compared to human reference genome.

2.4. Statistical analysis

Fisher's exact test was used to evaluate the association between genomic profile (cancer-related vs. non-cancer-related germline mutations) in affected and unaffected sibs.

3. Results

We performed whole-exome sequencing in two cases with lung adenocarcinoma and their unaffected older sibling as controls (Table 1). The two patients had early onset lung cancer in the absence of a history of tobacco smoking and a healthy sibling who was older by 7 and 13 years, respectively. Exome sequencing generated a mean of 71.8 Gb of sequence data (SD = 0.72), at a mean depth of 34.5-fold exon coverage (SD = 0.42).

Overall, we identified a total of 189,388 sequence variants in the four samples. After removing variants with low coverage, there were 180,424 variants (Fig. 1), including 95,141 in the coding region. Among these, 171,329 variants were listed in dbSNP137.

We started examining regions of candidate genes for potential involvement in lung cancer. No truncating mutations were found in the genes located in the human homologous regions of the *Pas1* and *Par1* loci (Table 2). Concerning the *Pas1* region, we found a non-synonymous mutation in *CASC1* that was predicted to be benign by both SIFT and PolyPhen-2 predictive tools in the unaffected sib (control 1). Concerning *Par1* locus, we found a synonymous polymorphism in *PPP1R9B* gene in both controls but not in cases. Concerning the 15q24–25.1 region, no truncating mutations were found. We found the risk allele of rs1051730, and rs16969968, in both siblings of a pair (case 1 and control 1), concerning the 5p15.33 region, we found rs402710 in case 2 and control 1 of different sib-pairs, whereas concerning the 6p21 region, we did not find any truncating mutations/polymorphisms (Table 2). Regarding the 36 associations identified in our previous GWAS on discordant lung cancer sibs [15], we found non-synonymous mutations in SLC6A6 in case 1, in PLEKHG1 and APBA1 in case 2, in TOP3B in control 1, and in MST1R, CDH13, and SDK2 in control 2. All these mutations are polymorphisms and were predicted to be benign by both predictive tools. Among genes frequently reported as somatically mutated in lung cancers, no truncating mutations were found.

We then went through a whole genome "unsupervised" analysis, searching for mutation in affected but not in unaffected sibs and vice versa. Because of the known difficult interpretation of non-synonymous and splice site mutations as responsible for cancer occurrence, we focused on disrupting mutations in exome region that lead to truncation and then *bona fide* inactivation of the protein product: frameshifts insertions and deletions, and non-sense mutations. Among the 95,141 exonic variants, we found 133 truncating mutations not present in dbSNP137 in the four

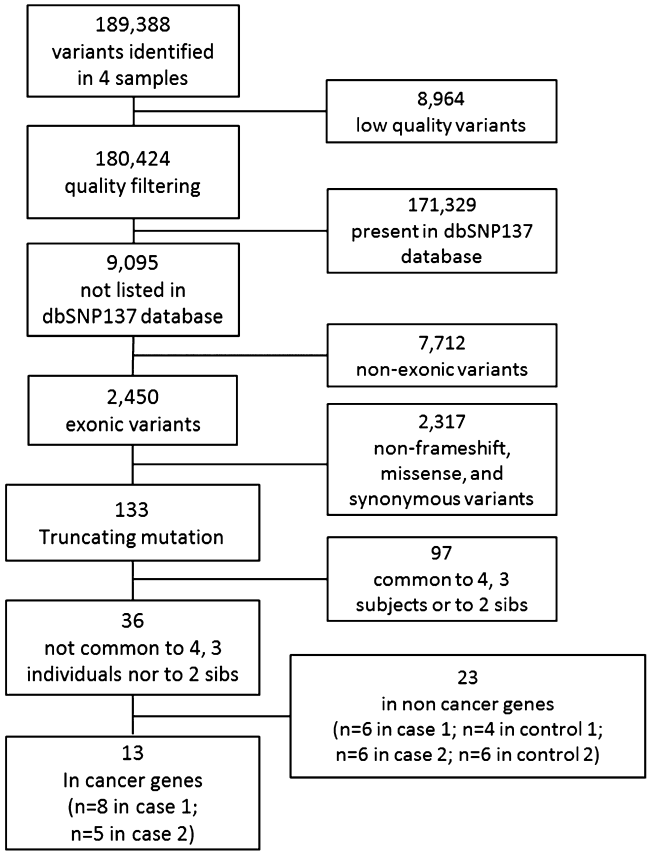

Fig. 1. Summary of exome sequence mutations.

analyzed samples. After exclusion of mutations that were present in all 4 subjects ($n = 32$), or 3 of the 4 ($n = 21$), and those present in the two sibs with discordant phenotype ($n = 44$), 36 variants in 35 genes were selected (Supplementary Table 2). All the 36 sequence variants identified were in heterozygosity. No common disrupting mutations were found between affected sibs. We found the same truncating mutation in healthy controls in a gene with unknown function (*C15orf40*).

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.lungcan.2014.05.020.

At gene level, we then checked if mutated genes have previously been reported as involved in lung or other cancer. Among these 36 truncating mutations, 13 fall in genes that have previously been involved in cancer occurrence (*BAG6*, *SPEN*, WISP3, *JAK2*, *TCEB3C*, *NELFE*, *TAF1B*, *EBLN2*, *GON4L*, *NOP58*, *RBMX*, *KIAA2018*, *ZNF311*). Interestingly, all 13 were found in the two patients with lung adenocarcinoma, but not in the healthy sibs (Fig. 2). Comparing each sib-pair we observed a statistically significant enrichment ($p = 0.0026$, OR = 16.88, 95%CI 1.99–142.48) of disrupting mutations in cancer-related genes in the two cases. Among the 13 cancer-related genes, there were 3 genes known to be involved in lung adenocarcinoma and other non-small cell lung cancers (NSCLC). In particular, we found truncating mutations in *BAG6*, and *SPEN* genes in case 1, and in *WISP3* gene in case2 (Fig. 2). *BAG6* was found associated with lung adenocarcinoma and other NSCLC susceptibility in GWAS [13]; *SPEN* recurrently showed truncating somatic mutations in patients with splenic marginal zone lymphoma and adenoid cystic carcinoma and more recently in squamous cell lung cancer [20–22]; *WISP3* encodes a protein with tumor-suppressor activity, that is relevant to malignant transformation in NSCLC [23,24].

Table 2
Sequence variants in lung cancer candidate genes identified by whole-exome sequencing of two cases with lung adenocarcinoma and their healthy siblings.

Gene symbol	Gene name	Chr	Mutation type	SNP/position (bp)	Case 1	Control 1	Case 2	Control 2
CASC1	Cancer susceptibility candidate 1	12	Non-synonymous	25,311,457		√		
PPP1R9B	Protein phosphatase 1, regulatory subunit 9B	17	Synonymous	rs1569116		√		√
SLC6A6	Solute carrier family 6, member 6	3	Non-synonymous	rs62233560	√			
PLEKHG1	Pleckstrin homology domain containing, family G member 1	6	Non-synonymous	rs79596384			√	
APBA1	Amyloid beta precursor protein-binding, family A, member 1	9	Non-synonymous	rs34788368			√	
TOP3B	Topoisomerase (DNA) III beta	22	Non-synonymous	rs75602167		√		
MST1R	Macrophage stimulating 1 receptor	3	Non-synonymous	rs2230593				√
CDH13	Cadherin 13	16	Non-synonymous	rs34106627				√
SDK2	Sidekick cell adhesion molecule 2	17	Non-synonymous	rs2270716				√
CLPTM1L	Cleft lip and palate transmembrane protein 1-like	5	Intronic	rs402710		√		√
VEGFA	Vascular endothelial growth factor A	6	5′UTR	rs2010963			√	√
CHRNA3	Cholinergic receptor, nicotinic, alpha 3	15	Synonymous	rs1051730	√	√		
CHRNA5	Cholinergic receptor, nicotinic, alpha 5	15	Non-synonymous	rs16969968	√	√		

5′UTR, five prime untranslated region.

Among genes previously found somatically mutated in other cancers, we found a truncating mutation in *JAK2, TCEB3C, NELFE, TAF1B*, and *EBLN2* (Table 3). Specifically, *TCEB3C* was recently identified as tumor suppressor gene in small intestinal neuroendocrine tumors [25]; *JAK2* is a frequent target for point mutations, deletion, or gene fusion in leukemia and myeloproliferative neoplasms [26–29]; *NELFE* depletion has been reported to lead to developmental defects and tumorigenesis [30]; *TAF1B* is a target for frameshift mutations in colorectal carcinomas [31]; and *EBLN2* is down-regulated in cervical cancer [32]. Truncating mutations were identified in 3 genes previously involved in tumorigenesis in mouse models or in vitro studies: *GON4L, RBMX*, and *NOP58*. Mice with recessive mutation in *Gon4l* gene spontaneously developed salivary gland tumors [33]; *RBMX* depletion sensitizes cells to DNA damaging agents [34]; and *NOP58* expression confers a greater ability to cell growth in response to stress [35]. Finally, we identified truncating mutations in 2 genes previously associated with cancer risk in GWAS: *KIAA2018* is as one of 23 prostate cancer susceptibility loci [36], and *ZNF311* is associated with hormone-related cancer risk [37] (Table 3).

3.1. Pathways related to identified genes

Gene Ontology analysis showed enrichment in genes with nucleic acid binding function (10/13 genes). Five of them were known regulators of nuclear transcription. The remaining 5 genes were classified as ribosome or receptor binding (Supplementary Table 3). Overall, each patient had disrupting mutations in genes belonging to the chromatin binding pathway (*GON4L*, *KIAA2018*, *SPEN*, *TCEB3C* in case1, and *TAF1B*, *EBLN2*, *ZNF311* in case2), and to the receptor binding category (*JAK2* in case1 and *ISP3* in case2).

Supplementary material related to this article can be found, in the online version, at http://dx.doi.org/10.1016/j.lungcan.2014.05.020.

In order to identify protein interaction and to examine possible relationships among genes, we performed bioinformatics analysis using STRING tool. In particular: (i) interaction was found between EP300 (E1A binding protein p300) and both BAG6 and SPEN; (ii) both BAG6 and GON4L interacted with UBC (ubiquitin C).

4. Discussion

In the present study we used a combination of the most advanced genetic tool, and a selection strategy based on the comparison of the genome of unaffected and affected siblings with very early age of cancer onset in the absence of cigarette smoke. This approach, using for the first time the whole-exome sequencing to investigate germline mutations associated with lung cancer risk, revealed that the affected sibs had an oligogenic combination of disrupted cancer predisposing genes.

The present study has a number of strengths and limitations. One of the strengths is that the study used the most updated genetic tool to detect truncating germline mutations in cancer genes. In addition, an original selection model (discordant sibs at

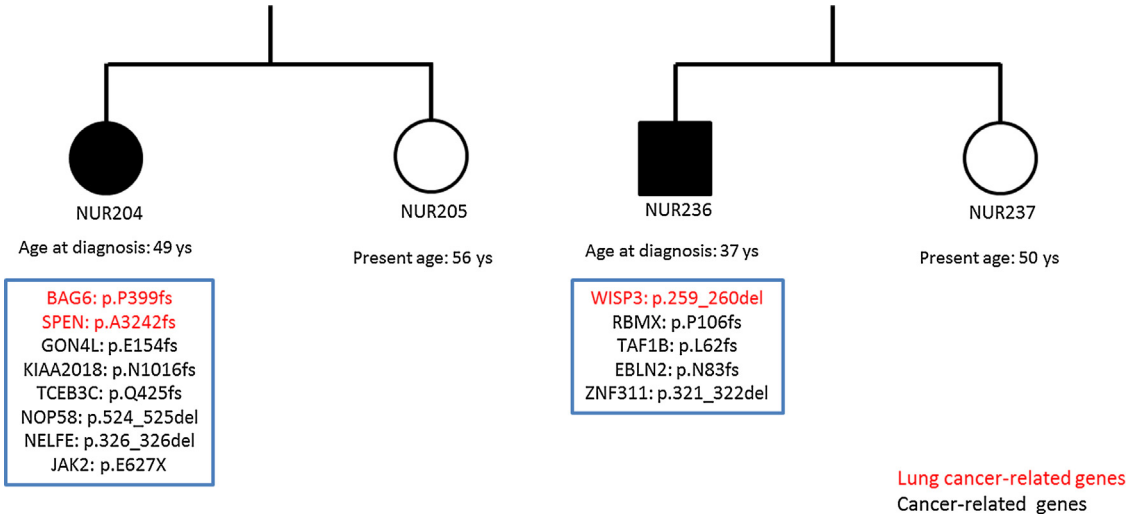

Fig. 2. Pedigree of the two sib-pairs and representation of disruptive mutations in cancer-related genes found in the two cases and in their unaffected sib.

Table 3
Truncating mutations identified by whole-exome sequencing, in two cases of lung adenocarcinoma and their two healthy siblings, and relevance of the affected gene to cancer development.

Gene symbol	Gene name	Chr	Mutation type	Protein function related to cancer development	Case 1	Contr 1	Case 2	Contr 2
ACOT4	Acyl-CoA thioesterase 4	chr14	Frameshift ins	NA	√			
ACOT4	Acyl-CoA thioesterase 4	chr14	Frameshift del	NA	√			
ANKRD36	Ankyrin repeat domain 20 family, member A4	chr2	Frameshift ins	NA				√
BAG6	BCL2-associated athanogene 6	chr6	Frameshift del	Nuclear protein implicated in apoptosis, cell-cycle regulation, and lung cancer risk [13,39]	√			
BEST3	Bestrophin 3	chr12	Frameshift del	NA			√	
C15orf40	Chromosome 15 open reading frame 40	chr15	Stop gain	NA		√		√
CCDC150	Coiled-coil domain containing 150	chr2	Frameshift ins	NA	√			
CLHC1	Clathrin heavy chain linker domain containing 1	chr2	Stop gain	NA				√
EBLN2	Endogenous Bornavirus-like nucleoprotein 2	chr3	Frameshift ins	RNA-binding protein down-regulated in cervical cancer [32]			√	
FAM200A	Family with sequence similarity 200, member A	chr7	Frameshift ins	NA		√		
FSIP2	Fibrous sheath interacting protein 2	chr2	Stop gain	NA	√			
FUT10	Fucosyltransferase 10	chr8	Stop gain	NA	√			
GABRG1	Gamma-aminobutyric acid (GABA) A receptor, gamma 1	chr4	Frameshift del	NA		√		
GOLGA6L6	Golgin A6 family-like 6	chr15	Stop gain	NA			√	
GON4L	Gon-4-Like	chr1	Frameshift del	Nuclear factor and transcriptional regulator. Mice with recessive mutation spontaneously develop salivary gland tumors [33,40]	√			
GPR135	G protein-coupled receptor 135	chr14	Frameshift ins	NA			√	
JAK2	Janus kinase 2	chr9	Stop gain	Tyrosine kinase involved in cytokine receptor signaling pathways and in the control of cellular growth and proliferation; frequent target of point mutations, deletions and gene fusions in cancer [26–29]	√			
KIAA2018	KIAA2018	chr3	Frameshift ins	Recently identified as a new prostate cancer susceptibility locus [36]	√			
LTN1	Listerin E3 ubiquitin protein ligase 1	chr21	Frameshift del	NA	√			
NELFE	Negative elongation factor E	chr6	Frameshift del	One of four subunits of the NELF complex that is required for maintaining the expression of key cell cycle regulators, ensuring proper rates of cell proliferation and cell cycle progression. NELFE depletion leads to tumorigenesis [30,41]	√			
NOP58	NOP58 ribonucleoprotein	chr2	Frameshift del	Nucleolar protein with small nucleolar RNA-binding function whose expression stimulates growth in response to stress [35]	√			
NUDT11	Nudix (nucleoside diphosphate linked moiety X)-type motif 11	chrX	Frameshift del	NA				√
OR1J2	Olfactory receptor, family 1, subfamily J, member 2	chr9	Frameshift ins	NA				√
RBMX	RNA binding motif protein, X-linked	chrX	Frameshift ins	Required for resistance to DNA damage; its depletion was sensitizes cells to DNA damage [34]			√	
REXO4	REX4, RNA exonuclease 4 homolog	chr9	Frameshift del	NA				√

Table 3 (*Continued*)

Gene symbol	Gene name	Chr	Mutation type	Protein function related to cancer development	Case 1	Contr 1	Case 2	Contr 2
SPEN	Spen family transcriptional repressor	chr1	Frameshift ins	Involved in NOTCH1 signaling; has recurrent truncating somatic mutations in cancer (including lung cancer)[20–22,42]	√			
SPNS3	Spinster homolog 3	chr17	Stop gain	NA				√
TAF1B	TATA box binding protein (TBP)-associated factor, RNA polymerase I, B	chr2	Frameshift ins	Subunit of the TBP-containing promoter selectivity factor TIF-1B/SL1, related to ribosomal transcription; frequently mutated in colorectal carcinoma [31,43]			√	
TCEB3C, TCEB3CL, TCEB3CL2	Transcription elongation factor B polypeptide-3C, -3C-like, -3C-like2	chr18	Frameshift ins	Tumor suppressor gene in small intestinal neuroendocrine tumors [25]	√			
TMEM254	Transmembrane protein 178°	chr10	Frameshift ins	NA		√		
TPSAB1	Tryptase alpha/beta 1	chr16	Stop gain	NA			√	
TPSB2	Tryptase beta 2	chr16	Stop gain	NA			√	
UBTFL1	Upstream binding transcription factor, RNA polymerase I-like 1	chr11	Frameshift ins	NA			√	
WISP3	Wnt1-inducible signaling protein 3	chr6	Frameshift del	Tumor suppressor that may act downstream in the WNT1 signaling pathway; relevant to malignant transformation in non-small cell lung carcinoma [23,24]			√	
ZNF311	Zinc finger protein 311	chr6	Frameshift del	DNA binding protein that may be a transcription regulator; associated with hormone-related cancer risk [37]			√	
ZNF551	Zinc finger protein 551	chr19	Stop gain	NA		√		

Chr, chromosome; Contr, control; NA, not available info about relation to cancer; ins, insertion; del, deletion.

the extreme ends of the phenotype) has been used. Obviously, a larger number of informative pairs of siblings will improve the study strength, but it was very difficult to find "ideal" discordant sibs, i.e. the affected sib with early age of tumor onset (<50 years) and at least one or more older unaffected siblings with an age difference of at least 7 years.

Anyway, the final outcome was not to demonstrate a specific common locus, or a small number of genes determining cancer occurrence, but, on the contrary, to give the proof of a concept, i.e. that in common sporadic cancer each individual finds his/her own way to cancer, that also depends on congenital predisposing factors, which vary from an individual to another. This concept could have important implications for the "personalized medicine", and it could be used to reduce the burden of expectations from GWAS. As there is not a major common susceptibility locus, but many and different among individuals, it was no possible to identify them by population-based GWAS.

Among limitations, the first is the small number of "ideal" analyzed kindreds. In this model, using a greater number of kindreds, including "non ideal" discordant sibs could be not of benefit, but a confounding factor, because it increases the number of observations, but reduces the power of the model. Another limitation is that we focused only on frameshifts and non-sense mutations, because only in them there is the consensus that the truncated protein suffers a loss of function. But we cannot exclude that non-synonymous or in-frame mutations could also be of interest. It is also known that a lot of non-cancer genes, in particular those controlling/modulating the immune response, or repair and host defense mechanisms, or drug metabolizing enzymes, could also be important for cancer development [38]. Therefore, it is likely that, in addition to germline mutations in cancer genes, a similar heritage concerning immune response or defense mechanisms could play an important role and be part of the individual predisposition/resistance to cancer. In addition, our analysis focusing on truncating mutations leads to identification of tumor suppressors, and thus the number of predisposing mutations can be underestimated. Indeed, all identified genes are tumor suppressors, except *JAK2*, that is a known oncogene. However, a recent in vivo study reported that loss of function in *JAK2* was associated with cancer [29]. This is in line with the observation that some genes may act both as tumor suppressors and as oncogenes.

Our study represents the first report on whole exome sequencing in lung ADCA patients using a pedigree model comparing discordant sibs. Observed germline mutations were not polymorphisms, but truncating mutations, likely responsible for loss of protein function. In case 1 we can envisage the combined disruption of the p53 pathway (*BAG6*) and of the EGFR pathway (JAK2 interacting with STAT5), whereas in case 2 the disruption of the Wnt pathway (WISP1) can be suggested. The identification of common pathways underlying genetic heterogeneity could lead to the design of new tests for early diagnosis based on biochemical targets of the causal loci modulating the disease. In the future, exome sequencing may be offered to distinguish individuals with low/high risk to lung cancer. Future research will detect "which is which", i.e. which germline mutation facilitates which step and how this combination of mutations may be causative or facilitate tumor occurrence.

Our findings: (1) confirm the hypothesis of an oligogenic combination for cancer susceptibility [9]; (2) further support a model of "private genetic epidemiology" for a better understanding of the genetic effects in lung cancer families. Although these data require confirmation, the private nature of the oligogenic combination reduces the potential for significant contribution from the exploration of additional patients, who will have their personal way to cancer. The present study could open the way to preventive identification of a constitutive genomic background for lung

cancer susceptibility, and have important implications for personalized medicine.

Conflict of interest statement

The authors certify that there is no conflict of interest.

Acknowledgements

The authors thank Associazione Marta Nurizzo, Brugherio, Italy (http://www.martalive.org/foreign.htm) for recruiting pairs of Italian lung cancer patients and their healthy sibs. We also thank Fondazione Umberto Veronesi for its support to EF through the Young Investigator Research Program fellowship. This work was supported in part by grants from the Italian Association for Cancer Research (AIRC, Grants Nos. 10323 and 12162). This work was supported in part by funds obtained through an Italian law that allows taxpayers to allocate 0.5 percent share of their income tax contribution to a research institution of their choice. The funders had no role in the design or conduct of the study, in the collection, analysis or interpretation of the data, or in the preparation, review or approval of the manuscript. The authors also thank Prof. Lucio Luzzatto for scientific editing.

References

[1] Jemal A, Siegel R, Xu J, Ward E. Cancer statistics, 2010. CA Cancer J Clin 2010;60(5):277–300.
[2] Peto R, Darby S, Deo H, Silcocks P, Whitley E, Doll R. Smoking, smoking cessation, and lung cancer in the UK since 1950: combination of national statistics with two case-control studies. Br Med J 2000;321:323–9.
[3] Peto R, Lopez AD, Boreham J, Thun M, Heath Jr C. Mortality from smoking in developed countries 1950–2010. New York, NY: Oxford University Press; 2012.
[4] Tokuhata GK, Lillienfeld AM. Familial aggregation of lung cancer in humans. J Natl Cancer Inst 1963;30:289–312.
[5] Gariboldi M, Manenti G, Canzian F, Falvella FS, Radice MT, Pierotti MA, et al. A major susceptibility locus to murine lung carcinogenesis maps on chromosome 6. Nat Genet 1993;3:132–6.
[6] Manenti G, Galbiati F, Giannì Barrera R, Pettinicchio A, Acevedo A, Dragani TA. Haplotype sharing suggests that a genomic segment containing six genes accounts for the pulmonary adenoma susceptibility 1 (Pas1) locus activity in mice. Oncogene 2004;23:4495–504.
[7] Dragani TA, Canzian F, Pierotti MA. A polygenic model of inherited predisposition to cancer. FASEB J 1996;10:865–70.
[8] Dragani TA. 10 Years of mouse cancer modifier loci: human relevance. Cancer Res 2003;63:3011–8.
[9] Galvan A, Ioannidis JP, Dragani TA. Beyond genome-wide association studies: genetic heterogeneity and individual predisposition to cancer. Trends Genet 2010;26(3):132–41.
[10] Ioannidis JP. Invited commentary-genetic prediction for common diseases. Arch Intern Med 2012;172(9):744–6.
[11] Amos CI, Wu X, Broderick P, Gorlov IP, Gu J, Eisen T, et al. Genome-wide association scan of tag SNPs identifies a susceptibility locus for LC at 15q25.1. Nat Genet 2008;40:616–22.
[12] 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 2008;40:1407–9.
[13] Wang Y, Broderick P, Matakidou A, Eisen T, Houlston RS. Role of 5p15.33 (TERT-CLPTM1L), 6p21.33 and 15q25.1 (CHRNA5-CHRNA3) variation and lung cancer risk in never-smokers. Carcinogenesis 2010;31:234–8.
[14] Falvella FS, Galvan A, Frullanti E, Spinola M, Calabro E, Carbone A, et al. Transcription deregulation at the 15q25 locus in association with lung adenocarcinoma risk. Clin Cancer Res 2009;15:1837–42.
[15] Li Y, Sheu CC, Ye Y, de Andrade M, Wang L, Chang SC, et al. Genetic variants and risk of LC in never smokers: a genome-wide association study. Lancet Oncol 2010;11:321–30.
[16] Galvan A, Falvella FS, Frullanti E, Spinola M, Incarbone M, Nosotti M, et al. Genome-wide association study in discordant sibships identifies multiple inherited susceptibility alleles linked to LC. Carcinogenesis 2010;31:462–5.
[17] Landi MT, Chatterjee N, Caporaso NE, Rotunno M, Albanes D, Thun M, et al. GPC5 rs2352028 variant and risk of LC in never smokers. Lancet Oncol 2010;11:714–6.
[18] Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25(14):1754–60.
[19] McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20(9):1297–303.
[20] Parry M, Rose-Zerilli MJ, Gibson J, Ennis S, Walewska R, Foster J, et al. Whole exome sequencing identifies novel recurrently mutated genes in patients with splenic marginal zone lymphoma. PLoS ONE 2013;8(12):e83244.
[21] Stephens PJ, Davies HR, Mitani Y, Van Loo P, Shlien A, Tarpey PS, et al. Whole exome sequencing of adenoid cystic carcinoma. J Clin Invest 2013;123(7):2965–8.
[22] Zheng CX, Gu ZH, Han B, Zhang RX, Pan CM, Xiang Y, et al. Whole-exome sequencing to identify novel somatic mutations in squamous cell lung cancers. Int J Oncol 2013;43(3):755–64.
[23] Kleer CG, Zhang Y, Pan Q, van Golen KL, Wu ZF, Livant D, et al. WISP3 is a novel tumor suppressor gene of inflammatory breast cancer. Oncogene 2002;21:3172–80.
[24] Chen PP, Li WJ, Wang Y, Zhao S, Li DY, Feng LY, et al. Expression of Cyr61, CTGF, and WISP-1 correlates with clinical features of lung cancer. PLoS ONE 2007;2(6):e534.
[25] Edfeldt K, Ahmad T, Akerström G, Janson ET, Hellman P, Stalberg P, et al. TCEB3C a putative tumor suppressor gene of small intestine neuroendocrine tumors. Endocr Relat Cancer 2013 [Epub ahead of print].
[26] Xu W, Chen B, Tong X. Chronic myeloid leukemia patient with co-occurrence of BCR-ABL junction and JAK2 V617F mutation. Int J Hematol 2014;99(1):87–90.
[27] Wei J, Wang C, Qin YW, Zhu J, Gao YR, Cai Q, et al. JAK2 V617F positive essential thrombocythemia developing in a patient with CD5$^-$ chronic lymphocytic leukemia. Chin Med J 2012;125(11):2076–9.
[28] Alshemmari SH, Rajaan R, Ameen R, Al-Drees MA, Almosailleakh MR. JAK2V617F allele burden in patients with myeloproliferative neoplasms. Ann Hematol 2013 [Epub ahead of print].
[29] Akada H, Akada S, Hutchison RE, Mohi G. Loss of wild-type Jak2 allele enhances myeloid cell expansion and accelerates myelofibrosis in Jak2V617F knock-in mice. Leukemia 2014 [Epub ahead of print].
[30] Jennings BH. Pausing for thought: truncating the early transcription elongation checkpoint leads to developmental defects and tumourigenesis. Bioessays 2013;35(6):553–60.
[31] Kim NG, Rhee H, Li LS, Kim H, Lee JS, Kim JH, et al. Identification of MARCKS, FLJ11383 and TAF1B as putative novel target genes in colorectal carcinomas with microsatellite instability. Oncogene 2002;21(33):5081–7.
[32] Lando M, Wilting SM, Snipstad K, Clancy T, Bierkens M, Aarnes EK, et al. Identification of eight candidate target genes of the recurrent 3p12-p14 loss in cervical cancer by integrative genomic profiling. J Pathol 2013;230(1):59–69.
[33] Simons AL, Lu P, Gibson-Corley KN, Robinson RA, Meyerholz DK, Colgan JD, et al. The Justy mutant mouse strain produces a spontaneous murine model of salivary gland cancer with myoepithelial and basal cell differentiation. Lab Invest 2013;93(6):711–9.
[34] Adamson B, Smogorzewska A, Sigoillot FD, King RW, Elledge SJ. A genome-wide homologous recombination screen identifies the RNA-binding protein RBMX as a component of the DNA-damage response. Nat Cell Biol 2012;14(3):318–28.
[35] Ito A, Watabe K, Koma Y, Kitamura Y. An attempt to isolate genes responsible for spontaneous and experimental metastasis in the mouse model. Histol Histopathol 2002;17(3):951–9.
[36] Eeles RA, Olama AA, Benlloch S, Saunders EJ, Leongamornlert DA, Tymrakiewicz M, et al. Identification of 23 new prostate cancer susceptibility loci using the iCOGS custom genotyping array. Nat Genet 2013;45(4):385–91.
[37] Pooley KA, Bojesen SE, Weischer M, Nielsen SF, Thompson D, Amin Al Olama A, et al. A genome-wide association scan (GWAS) for mean telomere length within the COGS project: identified loci show little association with hormone-related cancer risk. Hum Mol Genet 2013;22(24):5056–64.
[38] Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. Cell 2011;144(5):646–74.
[39] Yong ST, Wang XF. A novel, non-apoptotic role for Scythe/BAT3: a functional switch between the pro- and anti-proliferative roles of p21 during the cell cycle. PLoS ONE 2012;7(6):e38085.
[40] Lu P, Hankel IL, Knisz J, Marquardt A, Chiang MY, Grosse J, et al. The Justy mutation identifies Gon4-like as a gene that is essential for B lymphopoiesis. J Exp Med 2010;207(7):1359–67.
[41] Sun J, Li R. Human negative elongation factor activates transcription and regulates alternative transcription initiation. J Biol Chem 2010;285(9):6443–52.
[42] Rossi D, Trifonov V, Fangazio M, Bruscaggin A, Rasi S, Spina V, et al. The coding genome of splenic marginal zone lymphoma: activation of NOTCH2 and other pathways regulating marginal zone development. J Exp Med 2012;209:1537–51.
[43] Comai L, Zomerdijk JCBM, Beckmann H, Zhou S, Admon A, Tjian R. Reconstitution of transcription factor SL1: exclusive binding of TBP by SL1 or TFIID subunits. Science 1994;266:1966–72.