Cancer Letters ■■ (2016) ■■–■■

Contents lists available at ScienceDirect

Cancer Letters

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

Original Articles

Expression quantitative trait analysis reveals fine germline transcript regulation in mouse lung tumors

Chiara E. Cotroneo [a,1,2], Alice Dassano [a,1], Francesca Colombo [a], Angela Pettinicchio [a], Daniele Lecis [b], Matteo Dugo [b], Loris De Cecco [b], Tommaso A. Dragani [a,*], Giacomo Manenti [a]

[a] Department of Predictive and Preventive Medicine, Fondazione IRCCS, Istituto Nazionale dei Tumori, Milan, Italy
[b] Department of Experimental Oncology and Molecular Medicine, Fondazione IRCCS, Istituto Nazionale dei Tumori, Milan, Italy

ARTICLE INFO

Article history:
Received 2 December 2015
Received in revised form 25 February 2016
Accepted 26 February 2016

Keywords:
Animal models
Expression quantitative trait loci
Gene expression
Lung cancer
Single nucleotide polymorphisms
Transcriptome
Pas1

ABSTRACT

Gene expression modulates cellular functions in both physiologic and pathologic conditions. Herein, we carried out a genetic linkage study on the transcriptome of lung tumors induced by urethane in an (A/J x C57BL/6)F4 intercross population, whose individual lung tumor multiplicity (Nlung) is linked to the genotype at the Pulmonary adenoma susceptibility 1 (*Pas1*) locus. We found that expression levels of 1179 and 1579 genes are modulated by an expression quantitative trait locus (eQTL) in *cis* and in *trans*, respectively (LOD score > 5). Of note, the genomic area surrounding and including the *Pas1* locus regulated 14 genes in *cis* and 857 genes in *trans*. In lung tumors of the same (A/J x C57BL/6)F4 mice, we found 1124 genes whose transcript levels associated with Nlung (FDR < 0.001). The expression levels of about a third of these genes (n = 401) were regulated by the genotype at the *Pas1* locus. Pathway analysis of the sets of genes associated with Nlung and regulated by *Pas1* revealed a set of 14 recurrently represented genes that are components or targets of the Ras–Erk and Pi3k–Akt signaling pathways. Altogether our results illustrate the architecture of germline control of gene expression in mouse lung cancer: they highlight the importance of *Pas1* as a tumor-modifier locus, attribute to it a novel role as a major regulator of transcription in lung tumor nodules and strengthen the candidacy of the *Kras* gene as the effector of this locus.

Introduction

Expression quantitative trait loci (eQTLs) are genomic regions containing genetic variations that influence the transcription of a gene [1,2]. eQTLs can act either in *cis*, when the genetic variant specifically affects the expression of the allele of a gene located on the same chromosome, or in *trans*, when the variation alters the expression of both alleles [3]. *Cis*-acting eQTLs usually contain genetic variants that affect the molecular properties of DNA regulatory elements (e.g. promoters, UTRs, splicing sites, enhancers) located close to, or within, their target genes. *Trans*-acting loci, usually located in separate areas of the genome from their target genes, contain DNA variants which alter the functional properties or the abundance of diffusible regulatory elements such as transcription factors and microRNAs. However, in genome-wide eQTL studies, the definition of *cis*- (or local) and *trans*- (or distal) eQTLs is usually simplified by considering only the physical distance between the variant and the regulated gene(s).

eQTL studies can help uncover the functional mechanisms through which disease-associated variants, previously identified in genome-wide association or linkage studies, exert their role in pathogenesis. Indeed, most of these variants map in non-coding regions and therefore are likely to act as regulatory elements [4]. So far, just one genome-wide eQTL study of human lung cancer tissue has been reported, but it focused on genetic regulation acting in *cis* [5]. Indeed, in humans, the wide genetic heterogeneity among individuals [6] limits the statistical power for the genome-wide detection of eQTLs, hindering the detection of *trans*-eQTLs which typically have smaller phenotypic effects [3].

The difficulties in detecting eQTLs can be overcome through the use of simpler genetic models, such as inbred mouse strains that are homozygous at almost every genetic locus and that can be crossed to obtain progeny with lower genetic complexity than pedigrees of wild-type animals. In mice, numerous genetic loci are known to modify the development and progression of several cancer types, including lung cancer [7,8]. Except for a few cases where the identification of functional variations in candidate genes suggested a

* Corresponding author. Tel.: +39 0223902642; fax: +39 0223902764.
E-mail address: tommaso.dragani@istitutotumori.mi.it (T.A. Dragani).
[1] These authors contributed equally to the work.
[2] Present address: UCD School of Biomolecular and Biomedical Science, University College Dublin, Belfield, Dublin 4, Ireland.

http://dx.doi.org/10.1016/j.canlet.2016.02.054
0304-3835/

molecular mechanism of action [8–12], the effectors of most of these loci are still to be understood. Regarding the Pulmonary adenoma susceptibility 1 (*Pas1*) locus, the major locus affecting spontaneous and chemically induced lung tumorigenesis in mouse inbred strains [7], we recently suggested that it exerts its function through the modulation in *cis* of the expression of the 4A isoform of the cancer-related gene *Kras* (*Kras* 4A) in normal lung and in lung tumors [13]. This work was done in (A/J x C57BL/6)F4 intercross mice (hereafter called ABF4 mice) derived from crossing the A/J and the C57BL/6 inbred strains, which carry the susceptibility and resistance alleles of the *Pas1* locus, respectively.

Here, in the same ABF4 cross, we explored the genome-wide architecture of genetic regulation of gene expression in mouse lung tumor nodules, and identified the *cis*- and *trans*-eQTLs acting in this tissue. We observed that the *Pas1* locus exerts a pleiotropic effect on transcription in lung tumors, with a large number of *trans*-regulated genes not comparable with that of any other locus. By intersecting the set of genes whose levels in lung cancer tissue associated with lung tumor multiplicity with those whose levels were controlled by the genotype at the *Pas1* locus, we identified a set of 14 key genes of *Pas1* biological function in lung cancer susceptibility.

Materials and methods

Ethics statement

All animals received humane care according to the criteria outlined in a protocol approved on December 21, 2006, by the institutional ethical committee for animal use (CESA) at the Fondazione IRCCS Istituto Nazionale dei Tumori.

Mouse crosses, biological samples and genotype data

In a previous paper [13], we described the generation, treatment and sampling of intercross mice. Briefly, ABF4 mice were treated with a single intraperitoneal injection of urethane (1 g/kg body weight) according to a standard procedure for inducing the development of lung tumors [14]. When the animals were 40 weeks of age, they were killed, the lungs were removed for tumor counting (values of lung tumor multiplicity (Nlung) varied substantially among individuals, ranging from one to 33 in the animals that developed tumors), and a single lung tumor, 1–1.5 mm in diameter, was excised from each for RNA extraction. By histologic analysis of some excised tumors, they were classified as lung adenomas, i.e., the typical mouse lung tumor histotype [15]. Here, we took advantage of existing genotype data, for 142 male mice, regarding 548 informative (polymorphic) non-redundant SNPs dispersed over the whole genome excluding the Y chromosome (average density of coverage, ~ 6.0 Mb/SNP). We also used Nlung data and samples of total RNA from lung tumors of these 142 mice.

Exon array analysis

Total RNA was amplified using the Low Input Quick Amp WT Labeling kit, according to the manufacturer's instructions (Agilent Technologies, Santa Clara, USA). Fluorescent dye-labeled cRNA was hybridized to SurePrint G3 Mouse Exon 4 × 180K arrays (Agilent); hybridization and washing were performed on Agilent's microarray platform, according to standard protocols. Microarray images were acquired using an Agilent DNA microarray scanner; raw data were generated using Agilent Feature Extraction software. All microarray data were MIAME compliant and were deposited into the NCBI's GEO database (http://www.ncbi.nmlm.nih.gov/projects/geo/) with accession number GSE71232.

Microarray data were analyzed with a set of customized scripts in R (http://www.r-project.org/). First, raw median probe intensities were pre-processed using the *limma* package [16] of the R/Bioconductor project [17]; then, expression data were summarized at the gene level calculating the geometric mean of probes associated with the same gene. Only probes whose intensity was at least 10% brighter than the 95th percentile of negative control probes in at least 15% of samples were used for gene-level summarization. Finally, we removed genes not listed in the current mouse genome assembly (Mouse GRCm38, according to the Ensembl database [18]) as well as genes now listed as duplicates. After this filtering, performed using the *biomaRt* R package [19], we obtained a set of expression data for 18,020 genes for each individual of our population.

eQTL analysis

To identify QTLs controlling gene expression, we did simple interval mapping to assess genetic linkages for all the examined 18,020 genes, using a set of custom R scripts based on the R package *qtl* [20]. For each examined gene, we obtained a LOD score curve covering the whole genome; peaks in these curves corresponded to the position of putative eQTLs for that gene on the genetic map (expressed in cM). The genomic coordinates of each peak (in basepairs) were taken as those of the genotyped marker having the highest LOD score among all the markers mapping in proximity of the LOD peak. However, since in this mouse cross each marker identified a genomic region (locus) with an average length of 6 Mb, each of the 548 characterized markers was in linkage disequilibrium with several dozens or hundreds of genetic variants. This means that the observed modulation of gene expression can be due not only to the marker itself, but also to one or more variants in linkage disequilibrium with it. We defined a *cis*-acting eQTL as a locus that modulates the expression of a gene whose annotated start or end point is ≤ 3 Mb (i.e., ≤ 50% of the average distance between the genotyped markers) from the peak of the LOD score curve.

A *trans*-acting eQTL was initially defined as any statistically significant LOD score peak located further than 3 Mb from the target gene on the same chromosome, or located on another chromosome. For some genes, however, we found multiple peaks, with decreasing LOD scores, on the same chromosome. In most cases, these peaks were partially overlapping and centered about very close positions on the genetic map, suggesting that their presence was caused by only one regulatory locus. However, the low number of meiotic recombinations in the F4 intercross did not allow us to determine if this behavior of the LOD score curves was caused by a single locus or by multiple, very close, eQTLs. Therefore, for every chromosome displaying multiple, statistically significant LOD score peaks, we only considered the one with the highest LOD score, i.e. the peak with the strongest linkage with the examined phenotype.

To determine which peaks of a LOD score curve indicated significant linkage, we calculated threshold LOD values by permutation analysis on 62 genes randomly selected among all the genes that had LOD score peaks higher than the genome-wide LOD threshold of 3.3 proposed in the literature for mouse crosses [21]. In detail, 31 out of the 62 genes had a LOD peak corresponding to a putative *cis*-acting eQTL and 31 had a LOD peak corresponding to a putative *trans*-acting eQTL. After permutation analysis, these 62 genes showed threshold LOD scores peaks ($\alpha = 0.05$) ranging from 3.8 to 5.7, with a median of 4.0. Therefore, we decided to use a threshold LOD score of 5.0 for all eQTLs; such a threshold is conservative over the proposed genome-wide LOD threshold of 3.3.

A circular genome map to visualize inter-chromosomal *trans*-eQTLs was generated with R scripts for data formatting and Circos software [22].

Statistical analysis of the transcriptome

Statistical association of transcript levels with lung tumor multiplicity (square-root transformed Nlung values) was tested with custom R scripts by simple linear regression modeling, examining the association of one gene at a time, with Nlung and the expression level of the examined gene treated as continuous variables. The obtained *P*-values were adjusted for multiple testing following the Benjamini–Hochberg procedure [23]; the threshold for genome-wide significance was placed at an FDR < 0.001.

Pathway analysis

For genes whose expression level was linked to the genotype in the genomic area surrounding the *Pas1* locus, we searched for their participation in annotated biological pathways, looking for cases of functional enrichment, using the Ingenuity Pathway Analysis (IPA, Qiagen; http://www.ingenuity.com/products/ipa) online tool and the Ingenuity Canonical Pathways database. The significance threshold was placed at an FDR < 0.05. The *Pas1*-regulated genes involved in all the pathways identified by IPA were visually represented as a network using R scripts for data formatting and Cytoscape software [24]. The network was generated by connecting genes (nodes) with a line (edge) when they were involved in at least one common pathway. The size of nodes was scaled proportionally to the total number of IPA-identified pathways in which each gene was involved (Npathways). Nodes were colored in accordance with the direction of association of their expression levels with the number of lung tumor susceptibility (A/J-derived) alleles at the *Pas1* locus (red, genes positively associated with the A/J-derived allele; green, genes inversely associated with the A/J-derived allele).

Western blot analysis

Tissues from normal lungs and nodules, collected from mice and stored at –80 °C, were finely minced and lysed by boiling 10 minutes in lysis buffer (125 mM Tris HCl pH 6.8, 5% sodium dodecyl sulfate/SDS). After the addition of protease and phosphatase inhibitors, samples were sonicated and centrifuged at 11,000 rpm for 15 minutes at RT. Cleared supernatants were separated by SDS-PAGE using precast 4–12% Bis-Tris NuPAGE gels (Thermo Fisher Scientific) and blotted onto PVDF membranes (Merck Millipore) using the XCell II blot module (Thermo Fisher Scientific). Membranes were saturated in blocking buffer containing Tris-buffered saline (TBS) with 4% BSA for 30 minutes and then incubated overnight with the primary antibodies purchased from Cell Signaling Technology (Phospho-Akt Ser473 #9271; Akt #2920; Phospho-MEK1/2 #9121; MEK1/2 #4694) and Sigma-Aldrich (Phospho-ERK1/2 #M8159; ERK1/2 #M5670; Vinculin #V9131). After 1 h incubation with the appro-

Please cite this article in press as: Chiara E. Cotroneo, et al., Expression quantitative trait analysis reveals fine germline transcript regulation in mouse lung tumors, Cancer Letters (2016), doi: 10.1016/j.canlet.2016.02.054