# Identification of *Fat4* and *Tsc22d1* as novel candidate genes for spontaneous pulmonary adenomas

**Annerose Berndt**[1,2], **Clinton L Cario**[2], **Kathleen A Silva**[1], **Victoria E Kennedy**[1], **David E Harrison**[1], **Beverly Paigen**[1], and **John P Sundberg**[1]

[1]The Jackson Laboratory, Bar Harbor, ME 04609, USA

[2]currently at the University of Pittsburgh, Pittsburgh, PA 15261, USA

## Abstract

Genetic influences that underlie spontaneous lung oncogenesis are poorly understood. The objective of this study was to determine the genetic influences on spontaneous pulmonary adenoma frequency and severity in 28 strains of mice as part of a large-scale aging study conducted at the Jackson Aging Center (http://agingmice.jax.org/). Genome-wide association studies were performed in these strains with both low-density (132,000) and high-density (4,000,000) panel of single nucleotide polymorphisms (SNPs). Our analysis revealed that adenomas were relatively less frequent and less severe in females than males, and that loci implicated in frequency and severity were often different between male and female mice. While some of the significant loci identified mapped to genomic locations known to be responsible for carcinogen-induced cancers (e.g., *Pas1*), others were unique to our study. In particular, *Fat4* was influential in males and *Tsc22d1* was influential in females. SNPs implicated were predicted to alter amino acid sequence and change protein function. In summary, our results suggested that genetic influences that underlie pulmonary adenoma frequency are dependent on gender, and that *Fat4* and *Tsc22d1* are likely candidate genes to influence formation of spontaneous pulmonary adenoma in aging male and female mice, respectively.

## Keywords

genetics; aging mice; *Fat4*; *Tsc22d1*; lung cancer; strain survey; tumor suppressor; Jackson Aging Center

## INTRODUCTION

Lung cancer is the second leading cause of deaths in western societies and is epidemic worldwide. The Center of Disease Control reports more than 203,000 individuals diagnosed with and more than 158,000 deaths related to lung cancer for the year 2007. The most common causes for lung cancers are cigarette smoke (1), radon gas exposure (2), and viruses (e.g., human papillomavirus, polyomaviruses, cytomegalovirus) (3, 4). Yet, a subset of lung cancers occurs with no apparent environmental stimulus (i.e., spontaneous). The genetic background of spontaneous lung oncogenesis is difficult to study in human populations due to the length of such studies, the uncertainty of the time of onset, and the various environments within study cohorts. Mice are valuable substitutes due to the variety of mouse

strains that develop spontaneous lung tumors, their relative short life span, possibilities of cross-sectional studies, and the availability of many publicly available databases that allow for genetic investigations.

The genetic basis to lung cancer development in mice was investigated intensely in carcinogen-induced models using both quantitative trait locus (QTL) or genome-wide association studies (GWAS) (5–7). Those studies lead to the repetitive identification of pulmonary adenoma susceptibility (*Pas*) and resistance (*Par*) loci and to the isolation of genes involved in the oncogenesis (e.g., *Kras*). In contrast, spontaneous lung cancer development was studied less frequently without the success of repetitive identification of the same or similar genomic loci (7, 8) (Table 1). Wang and You (8) reported the first GWAS analysis for spontaneous lung oncogenesis using 13 mouse strains and about 135,900 SNPs and identified five loci associated with spontaneous lung cancer development (i.e., spontaneous lung tumorigenesis (*Slt*) loci) on chromosomes (Chrs) 6 at about 84 Mb (*Slt1*), 7 at about 4 Mb (*Slt2*), 8 at about 123 Mb (*Slt3*), 19 at about 33 Mb (*Slt4*), and X at about 138 Mb (*Slt5*). Another GWAS on spontaneous lung cancer incidence was reported by Manenti et al. (7) using 28 mouse strains. The investigators found a locus on Chr 5 at about 15 Mb (*Slt6*) with about 13,000 SNPs but not when a larger SNP panel of about 140,000 SNPs was used. Therefore, none of the currently reported loci for spontaneous lung cancer were repetitive findings, thus, making the identification of true influential genetic determinants difficult.

GWAS has been proven successful for some disease-phenotypes to identify causative genomic loci. In comparison to QTL studies, GWAS analysis excels because of the identification of smaller genomic regions due to utilization of denser SNP panels, which leads to fewer candidate genes for subsequent verification experiments. Whereas QTL studies are based on the genomic variation between two mouse strains, GWAS studies are often most successful with strain numbers greater than thirty (9). Nevertheless, GWAS carry the risk for detecting false-positive associations depending on factors such as density of SNP panels, numbers of strains, and population structure (i.e., over representation of certain genetic backgrounds) (10). The risk can be decreased by increasing the SNP density and the numbers of strains as well as by controlling the population structure among the studied strains.

As part of the large-scale aging study at The Jackson Aging Center we aimed to identify genes associated with spontaneous pulmonary adenoma frequency and severity in a cohort of 20-month-old mice of 28 strains. We utilized both, a low-density (132,000 (132k)) and a high-density (4 million (4 Mio)) SNP panel, and an efficient mixed-model algorithm to control for population structure (11). We hypothesized that we would identify a subset of the previously reported *loci* for spontaneous oncogenesis and also novel candidate regions and possibly genes due to the use of a denser SNP data set.

## MATERIALS AND METHODS

### Mice

The following 28 strains of inbred mice were examined for pulmonary adenomas: 129S1/SvImJ (129S1), A/J (A), BALB/cByJ (BALB), BTBRT+*tf*/J (BTBR), BUB/BnJ (BUB), C3H/HeJ (C3H), C57BL/10J (B10), C57BL/6J (B6), C57BLKS/J (C57BLKS), C57BR/cdJ (C57BR), C57L/J (C57L), CBA/J (CBA), DBA/2J (D2), FVB/NJ (FVB), KK/HIJ (KK), LP/J (LP), MRL/MpJ (MRL), NOD.B10Sn-H2^b/J (NOD; a congenic strain with the NOD genetic background but with a histocompatibility locus from a diabetes-resistant strain), NON/ShiLtJ (NON), NZO/HlLtJ (NZO), NZW/LacJ (NZW), P/J (P), PL/J (PL), PWD/PhJ (PWD), RIIIS/J (RIIIS), SM/J (SM), SWR/J (SWR), and WSB/J (WSB). Originally two

more strains (AKR/J and SJL/J) were part of our study, but no mice of those strains survived until the time of euthanization. The strains in our study were representatives from each of the seven groups of inbred strains based on their derivation and haplotype (12). Thus, strains were selected for genetic diversity and not specifically for life span or cancer susceptibility. All investigated strains were part of a large-scale aging study by The Jackson Aging Center, which has been described in detail elsewhere (13). Briefly, mice were obtained, raised, and maintained at the breeding facilities of The Jackson Laboratory (Bar Harbor, ME). At an age of 6 to 8 weeks animals were transferred from the breeding facilities to a specific pathogen-free room, in which they were maintained until sacrifice. At 20 months ($\pm$ 28 days) of age mice were euthanized by $CO_2$ asphyxiation using methods approved by the American Veterinary Medical Association and subjected to necropsy. The breeding facilities and the mouse room were regulated on a 12 hr light/12 hr dark cycle and were maintained at an ambient temperature of 21–23 °C. Mice of the same sex (4 per pen) were housed in duplex polycarbonate cages ($31 \times 31 \times 214$ cm) on pressurized individually ventilated mouse racks (Thoran Caging System) with a high efficiency particulate air-filtered supply and exhaust. Mice were allowed *ad libitum* access to acidified water (pH 2.8 – 3.2) and pellets containing 6% fat (LabDiet 5K52, PMI Nutritional International, Bentwood, MO). The mouse colonies were regularly monitored (4-times/year) for and found to be free of 17 viruses (Ectromelia virus, GDVII (Theiler's) virus, Hantaan virus, K virus, Lactic dehydrogenase elevating virus, Lymphocytic choriomeningitis virus, Mouse adenovirus, Mouse cytomegalovirus, Mouse hepatitis virus, Mouse minute virus, Mouse norovirus, Mouse parvovirus, Mouse thymic virus, Pneumonia virus of mice, Polyoma virus, Reovirus 3, and Rotavirus), 17 bacterial species (including *Helicobacter* spp., *Pasteurella pneumotropica*, and two *Mycoplasma* spp.), external and internal parasites, and the microsporidium *Encephalitozoon cuniculi*. Animal protocols were reviewed and approved by The Jackson Laboratory Institutional Animal Care and Use Committee. Mouse handling and care were followed according to the Public Health Service animal welfare policies.

### Tissue Preparation

All mice were subjected to complete necropsies according to methods described in detail elsewhere (14). The lung was inflated with Fekete's acid alcohol formalin and allowed to fix overnight, after which the lungs were stored in 70% ethanol. Lungs were then trimmed longitudinally and cross-sectionally to yield a minimum of 4 lobes. After trimming, the tissues were embedded routinely in paraffin, sectioned at 6 µm, and stained with hematoxylin and eosin (H&E).

### Definition of Pathological Lesions in the Lung

All lung tissue slides were reviewed by the same experienced, board-certified pathologist (JPS), who has been an advisor to the European Mouse Pathology Consortium (15). We applied some of the physiological phenotyping assays used in the Knockout Mouse Project (i.e., Tier) (16) that are available online via the Mouse Phenome Database (MPD) (17). However, the detailed histopathology presented here is unique to this investigation.

Aging mice develop a variety of proliferative lung lesions, but the most common spontaneous cancer type is pulmonary adenoma (18). Thus, in our study diagnoses were limited to adenomas and occasional primary epithelial hyperplasia. The later has been argued to potentially represent an early stage in development of pulmonary neoplasias (19) and, therefore, was grouped with pulmonary adenomas. Rarely, we also found adenocarcinomas, which were not included in this study. All lesions were characterized by severity, which is based on tumor size and numbers per mouse (i.e., multiplicity) (scores: 1 - mild; 2 - moderate; 3 - severe; 4 - extreme), and from which average scores for female or male mice per strain were calculated. Lesions were also defined by their frequency —

percentage of female or male mice per strain diagnosed with one or more of the lesions. Data were recorded and coded using the Mouse Disease Information System (MoDis) (20, 21).

## Genome-Wide Association Mapping

GWAS for pulmonary adenoma frequency and severity were performed using the efficient mixed-model association (EMMA) correction server, which is publically available through the University of California, Los Angeles (UCLA). Input phenotypes were per-animal data for adenoma frequency (mice with adenoma(s) were scored 1; mice without adenomas were scored 0) and adenoma severity (mice without adenomas obtained a severity score of 0; other severity scores were described above). EMMA uses a linear mixed-model algorithm to control for population structure and genetic relatedness (11). The association scans were performed using both available SNP panels (132k and 4 Mio) on the EMMA server. The 132k panel is a subset of the 4 Mio SNP panel. We used both SNP panels because previous GWAS on spontaneous lung oncogenesis were performed using similar low-density SNP panels (135k and 140k or 13k) and, therefore, the EMMA 132k panel in our study was in comparable size to previous SNP densities. The second, high density, SNP panel (4 Mio) was used to compare the results of the 132k panel and potentially to detect locations that cannot be found with the smaller SNP set. Each SNP was evaluated individually and *P*-values were recorded as the strength of the genotype-phenotype associations.

## Gene Identification

Due to the possibility that the identified significant SNPs were not the actual disease-causing variations but are inherited in co-segregation with true disease-causing variations, we applied an arbitrary confidence interval of ± 1 million base pairs (Mb) around each association peak. For each interval we isolated the list of unique genes using the *Genes and Markers Query Form* provided by the Mouse Genome Informatics (MGI) database. For each gene we identified all SNPs known to cause an amino acid change in the gene product (i.e., non-synonymous SNP, Cn SNP) using the *SNP wizard* at the MPD and their Center for Genome Dynamics 1 (CGD1) imputed SNP dataset (22, 23). To determine if Cn SNPs are predicted to cause functional changes of the protein we used the SIFT web tool by the J. Craig Venter Institute (24). Finally, we grouped strains with the same Cn SNP genotype and compared the phenotypes between groups.

## Statistical Analyses

Lung sections from 280 female and 275 male mice were examined. Genotype groups were compared using two-sided Student's t-test. Effects on tumor frequency and severity of sex, strain, and sex-by-strain were tested using one-way ANOVA. Statistical significance was assigned at a *P*-value < 0.05. All statistical analyses were performed using JMP 8 statistical analysis software (SAS institute).

# RESULTS

## Comparison of Strains Between Genetic Studies

We studied several strains (54%) that were common between ours and at least one of the two previous publications on genome-wide association studies of spontaneous pulmonary adenomas: A, C3H, B10, B6, C57BR, C57L, CBA, D2, FVB, LP, NZW, P, RIIIS, SM, and SWR. We also investigated 13 strains (46%) that were unique to our study — i.e., 129S1, BALB, BTBR, BUB, C57BLKS, KK, MRL, NOD, NON, NZO, PL, PWD, and WSB (Table 2). Seven strains were common between all three investigations.

### Frequency of Pulmonary Adenomas

Lungs of each mouse were examined histologically and were found either to be free of or to contain adenomas. Photomicrographs of representative lesions are available on the Mouse Tumor Biology Database (MTB) (25) and on Pathbase (26, 27).

Among the 28 mouse strains, we found that females and males of 14 strains (129S1, A, BALB, C57L, CBA, FVB, LP, NON, NZO, PWD, RIIIS, SM, SWR, WSB) and males of 5 additional strains (BTBR, B6, C57BR, NOD, PL) developed lung adenomas. Nine strains did not develop lung cancers until the age of 20 months in both females and males (BUB, C3H, B10, C57BLKS, D2, KK, MRL, NZW, P). In addition, female mice of five strains did not develop adenomas (BTBR, B6, C57BR, NOD, PL) (Table 2). The frequency of mice with tumors varied among strains from 7% (C57BR) to 40% (SWR) in females and from 7% (B6) to 67% (A) in males. Among all strains the average frequency of tumor development is about twice as high in males (16%) than in females (9%) (Table 3).

### Severity of Pulmonary Adenomas

For each mouse with pulmonary adenomas a severity score was determined using a scoring system ranging from 1 (mild) to 4 (extreme). Among all strains the percentage of mild, medium, severe, and extreme tumors was 6%, 2%, 0.4%, and 0.4%, respectively, for female mice and 11%, 5%, 0.7%, and 0%, respectively, for male mice (Table 3). While most of the strains had mild and medium severity scores we found that female mice of strain RIIIS/J and male mice of strain A/J had severe and extreme severity scores.

### Sex Differences in Pulmonary Adenoma Frequency and Severity

Because gender is an important determinant in lung cancer development and survival rates, we determined if sex was an influential factor for frequency and severity in our study. Among all strains, both frequency and severity of the adenomas were significantly lower in females than in males ($P = 0.0039$ and $P = 0.0041$, respectively) (Table 3). The effects of strain and strain-by-sex were not significant, most likely due to the high numbers of mice that did not develop pulmonary adenomas with age.

### Genome-Wide Association Mapping for Pulmonary Adenoma Frequency and Severity

GWAS for frequency and severity was performed using the UCLA EMMA correction server (http://mouse.cs.ucla.edu/emma), which uses a linear mixed model algorithm to correct for population structure and shared ancestry (11). Compared to previous methods such as single SNP, 3-SNP sliding window, and haplotype association mapping the EMMA method reduces false-positive associations and increases the power.

We performed GWAS using the available low and high-density SNP panels (132k and 4 Mio, respectively) because the low-density panel is comparable to SNP densities in previous GWAS publications and the high-density panel provides many more SNPs and, thus, denser genome coverage. We compared the scan results for unique SNPs within the 200 most significant associations (Table 4). For frequency we found 113 unique SNPs of the 132k panel and 113 unique SNPs of the 4 Mio panel for female mice as well as 163 unique SNPs of the 132k panel and 126 SNPs of the 4 Mio panel for male mice. Between the 132k and the 4 Mio SNP panels for frequency only 17 and 21 SNPs were common for females and males, respectively. For severity we found 113 unique SNPs of the 132k panel and 103 unique SNPs of the 4 Mio panel for female mice as well as 135 unique SNPs of the 132k panel and 180 SNPs of the 4 Mio panel for male mice. Between the 132k and the 4 Mio SNP panels for severity only 7 and 29 SNPs were common for females and males, respectively (Table 4).

As expected, we found both, differences and commonalities, in the detected loci between the two SNP panels. Loci containing more than one SNP that were found with the 132k SNP panel are shown in Table 5 (Supplemental Table 1A for single SNP loci) and those found with the 4 Mio SNP panel are listed in Table 6 (Supplemental Table 1B for single SNP loci). The following comparisons were performed on only those loci that contained at least two SNPs. For adenoma frequency in females all loci detected with 4 Mio SNPs were also detected with 132k SNPs. Multiple loci were detected on Chrs 8, 13, 15, and 17 only with the 132k panel. For adenoma severity in females loci were detected on Chrs 7, 11, and 17 only with the 4 Mio panel and on Chrs 1, 2, 3, 4, 11, 12, 13, 17, 18, and 19 only with the 132k. Only four loci on Chrs 3, 14, and 15 were common between the panels.

For adenoma frequency in males loci were detected on Chrs 4, 7, and 19 only with the 4 Mio panel and on Chrs 4, 6, 10, 11, 13, 18, and 19 only with the 132k SNP set. Loci on Chrs 1, 3, 6, 10, 14, and 18 were similar between those two sets. Finally, for severity in males one locus was detected on Chr 1 only with the 4 Mio set and loci on Chrs 3, 6, 12, 13, 18, and 19 were only found with the 132k SNP set. Loci on Chr 1, 4, 6, 7, 10, 18, and 19 were common between both panels.

Associations for frequency and severity often differed in their genome locations for female mice. For example, when using the 4 Mio SNP panel adenoma frequency is influenced by SNPs on Chrs 1, 12, and 18 whereas severity is influenced by SNPs on Chrs 7, 15, and 17. In males, there were less significant differences than for females; however, some chromosomal locations differed, such as for Chrs 3, 7, and 16 (Table 6). Genome locations for influences of pulmonary frequency differed markedly between females and males. Whereas Chrs 2, 11, and 12 were specific for females, Chrs 4, 6, 7, 9, 10, and 19 were specific for males. The same was true for adenoma severity. Here, Chrs 2, 11, and 15 were specific for females and Chrs 1, 4, 6, 10, 18, and 19 were specific for males. Also, the locations on Chrs 7 and 17 differed between the traits for male mice. Only the locus on Chr 1 at around 117 Mb was common between female and male mice for frequency. For severity, none of the loci were common between the sexes (Table 6).

## Genetic Influences of Pulmonary Adenoma Frequency and Severity

Our GWAS analyses identified several previously reported loci for carcinogen-induced and spontaneous pulmonary cancer. The Chr 6 association at around 145.7 Mb — the *Pas1* locus (6, 28) — was found for frequency and severity only in male but not in female mice. In males, we also observed an association on Chr 18 at around 62 Mb, which was previously reported as locus (*Halt7* locus) (7), and another association on Chr 6 at around 88 Mb, which was only 4 Mb downstream of *Slt1* at 84 Mb (8). For females we identified a locus for severity on Chr 2 at around 153 Mb, which was found previously only 5 Mb upstream at about 148 Mb (*Halt2* locus) (7).

To identify novel genetic influences we searched for Cn SNPs in a ± 1Mb region around the most significant SNP association loci containing at least two SNPs from the 4 Mio SNP panel (Table 6). Alleles from Cn SNPs were used to assign strains into groups for phenotype comparisons and SIFT analysis was used to identify those Cn SNPs that are predicted to cause functional changes in the proteins. All Cn SNP with significant allele group differences are listed in Table 6. Of the 19 identified genes, four genes (i.e., *Fat4, Tsc22d1, Ccdc122*, and *Fbxo38*) contained Cn SNPs that lead to an amino acid change, which predicts a change in function of the protein. Two of the four genes (*Fat4* and *Tsc22d1*) are known tumor suppressor genes and, thus make them likely candidate genes for lung cancer development.

On Chr 3 we found the Cn SNP rs29949720 in exon 9 of the Fat tumor suppressor homolog 4 (*Fat4*) at 38.882537 Mb. The C/G allele change of rs29949720 causes a histidine (basic polar, hydrophilic) to glutamic acid (neutral polar, hydrophilic) change at position 3805. The locus was identified for adenoma frequency in male mice (Table 6) and we found male mice of strains with the C allele at rs29949720 developed significantly less pulmonary adenomas than males of strains with the G allele ($P = 0.039$) (Table 7).

On Chr 14 we identified the Cn SNP rs31392028 in exon 1 (or 3, transcript dependent) of the *Tgf*-beta-stimulated clone-22 (*Tsc22d1*) at 76.818029 Mb. The G/C allele change causes an alanine (neutral polar, hydrophobic) to proline (neutral polar, hydrophilic) change at position 714 of the protein. Because the locus was identified for adenoma frequency in females (Table 5) we compared the two allele groups using the female frequency data. We found that strains with the G allele had significantly fewer tumors than the C allele group ($P < 0.001$) (Table 7).

Finally, we also observed another two genes, which are predicted to lead to a protein function change. Coiled-coil domain containing 122 (*Ccdc122*) was significantly associated with adenoma frequency in females and F-box protein 38 (*Fbxo38*) with frequency in males. Beside the changes described in Table 6, those genes have not yet been described for roles in cancers or other lung diseases and, thus, would need further investigation to suggest them as candidate genes for spontaneous adenomas.

## DISCUSSION

Identification of genetic influences of spontaneous pulmonary adenomas has been difficult due to the small body of available data and the lack of repetitive findings between investigations. Here we present data on frequency and severity of spontaneous pulmonary adenomas in a survey of 28 inbred and wild-derived mouse strains at 20 months of age. Our study represents the first GWAS investigation of spontaneous pulmonary adenoma incidence for which all mouse strains were maintained under the same environmental conditions and analyzed by the same pathologist. Previous studies used a common set of 13 strains in which lung cancer was diagnosed by multiple investigators: Malkinson (29), Drinkwater and Bennett (30), the MTB of The Jackson Laboratory, and Wang and You (8). Additionally, we used a 4 Mio SNP set whereas previous GWAS were performed with less than 140k SNP panels. Our hypothesis was that we would find at least partially those *Slt* loci that were discovered previously but also that we would identify loci unique to our study design. In the original paper for spontaneous pulmonary adenomas by Wang and You (8) the investigators found associations on Chrs 6, 7, 8, 19, and X (*Slt1-5*) (Table 1). We also found an association on Chr 6 at about 88 Mb, which was only 4 Mb downstream of *Slt1* locus at about 84 Mb. This could potentially be the same locus considering that the SNP density between the studies differed greatly. Beside our repetition for this locus for spontaneous pulmonary adenomas, we also observed loci previously reported for carcinogen-induced lung cancers. For example, one of our significant associations was identified on Chr 6 at 145.7 Mb (i.e., *Pas1* locus) (6, 28), which is only 800k base pairs downstream from the cancer susceptibility candidate 1 (*Casc1*) and lymphoid-restricted membrane protein (*Lrmp*).The identified Cn SNPs in *Casc1* and *Lrmp*; however, did not cause significant phenotypic changes between the allele groups, thus, making them unlikely candidate genes for our criteria. Additionally, we also identified significant associations on the *Halt3* (Chr 2) and *Halt7* (Chr 18) loci, which were previously identified with a GWAS for urethane-induced lung cancer multiplicity (7). Conclusively, our investigations repeated multiple loci from previous studies, thus, supporting their importance for lung oncogenesis in both, spontaneous and carcinogen-induced, pulmonary cancers. Further studies, such as global

gene expression or traditional QTL studies need to be performed to identify and prove candidate genes within those regions.

We did not identify any other common *Slt* regions among previous studies and our investigations. Possible explanations for the lack of further overlap may be the differences in the size of the SNP panels. We showed that the results between the 132k and the 4 Mio SNP panels in our study differed. Often locations were identified with the 132k SNP panel that could not be detected with the 4 Mio panel. Although many of the locations identified with the 4 Mio panel were identified with the low-density SNP set, we also found associations that were unique to the 4 Mio panel. Because the *P*-values for the associations with the 4 Mio were higher than those with the 132k set panel it is likely that more false positive associations were identified with the smaller SNP panel. Additional causes for the lack of further overlap of identified loci could be the number and type of mouse strains used (we did not investigate 129X1/SvJ, CAST/EiJ, NZB/BlnJ, and SPRET/EiJ of the previous common 13 strain set and we did also not investigate AKR/J, C58/J, DBA/1J, MA/MyJ, NGP, O20, RF/J, SJL/J, ST/bJ and STS/A from the study by Manenti et al. (7)), and differences in phenotyping procedures (previous studies obtained diagnoses from different investigators and thus different pathologists vs. our study obtained diagnoses from one pathologist). Other factors, such as maintenance conditions (31), age of mice at termination of the study, diet (32), and the number of tissue samples for histopathological examination (33) could have also caused differences in study results. Finally, differences in the GWAS algorithms can cause different results. Whereas Wang and You (8) applied a Fisher exact test on a $2 \times 2$ contingency table (to test for resistance or susceptibility for each allele of a SNP), Manenti et al. (7) applied a t-test (continuous phenotype) for each SNP-*Slt* association and we used the EMMA mapping method by Kang et al. (11).

Our study identified differences between female and male mice for frequency and severity of pulmonary adenomas. Male mice developed adenomas more frequently and with greater severity than female mice. Differences between sexes translated into different genetic locations that determine adenoma frequency and severity in female and male mice. For example, the Chr 3 locus (i.e., *Fat4*) was only identified in males and the Chr 14 locus (i.e., *Tsc22d1*) was only found in females. Indeed, only the Chr 1 locus around 117 Mb seemed to be common between the sexes. Sex differences in mouse models of lung cancers were reported previously (34). For example, sex differences were observed in lung cancer development in mouse models, such as those with the *Kras* and *Trp53* missense mutations (35). In addition, ovariectomized female mice showed the same lung cancer characteristics as males, whereas estrogen application increased tumor rates more in ovariectomized females than intact males (36). In the same strain survey as presented here, gender-specific differences in other age-related phenotypes, such as electrocardiogram time intervals and heart rate (37), albuminuria (38), and IGF1 and lifespan (39), have been observed. However, only the study on albuminuria investigated their genetic basis. In the latter only one association was significant for male mice on Chr 3 at 157 Mb. As expected (due to the different phenotypes), we did not find associations in this area of the genome for pulmonary adenoma susceptibility. Reasons for the small number of significant loci identified for albuminuria might have been the use of a different association mapping algorithm (haplotype association mapping vs. EMMA) and a smaller SNP panel compared to our study (63k vs. 4 Mio).

Disease-causing polymorphisms can induce changes in gene expression (if located in the promoter region) or functional changes of the gene-encoded protein (if located in protein coding regions of the gene). Our focus for this study was to identify polymorphisms located in coding regions of genes that would lead to an amino acid and functional differences in the protein. Gene expression analysis of the lung tissues was not part of this aging project,

therefore, limiting our ability to search for global gene expression differences. Yet, we identified genes with Cn SNPs that are predicted to cause functional changes in the proteins. Two of the four genes, *Fat4* and *Tsc22d1*, have known functions in cancers in other organs. *Fat4* — originally described as a tumor suppressor in *Drosophila* — was later shown to be a candidate tumor suppressor gene in breast cancer cell lines and primary tumors (40). *Tsc22d1* is also a known tumor suppressor by promoting cell growth, survival, and proliferation and by preventing apoptosis, and has been associated with several cancers types (e.g., breast cancer, salivary gland cancer, prostate cancer) (41–43). It has been shown to be highly expressed in the lung (44). Both genes were important for frequency but not for severity. This may be due to the possibility that the genes that control severity differ from genes that are involved in the frequency of the lesions. From these results, we concluded that it is likely that both genes, *Fat4* and *Tsc22d1*, could be involved in molecular processes that lead to oncogenesis in the lung.

In comparison to previous studies we found different lung cancer frequencies (7, 8). For example, in previous publications strain A/J showed a lung tumor frequency of 82% whereas we only found a frequency of 29% in females and 67% in males. Another example, strain SM/J was reported with a frequency of 0% previously but our analysis found a frequency of 18% in females and 29% in males. The large frequency differences could be accounted for because previous studies focused on lung cancer identification whereas this study was part of routine, multi-organ evaluations. Many of these adenomas are very small and not seen on gross examination. Larger numbers of sections per mouse, more than the 4 used here, might increase the frequency. However, the same approaches were used for all mice within this study so the error rates were consistent. Environment varies between institutions, which can have an effect on frequency, so it is not surprising that actual numbers varied by study.

In summary, spontaneous pulmonary adenomas in mice are determined by multiple genetic loci that differ between female and male mice. The candidate gene *Fat4* is potentially involved in pulmonary adenoma frequency in male mice where as the candidate gene *Tsc22d1* seems to be involved in the oncogenesis in females mice. Replications between GWAS are important to identify candidate genes, but they are dependent on multiple factors, including SNP density, number of strains, and algorithms used.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

## REFERENCES

1. Hecht SS. Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nat Rev Cancer. 2003; 3:733–744. Available from: http://www.nature.com/nrc/journal/v3/n10/full/nrc1190.html. [PubMed: 14570033]

2. Catelinois, O.; Rogel, A.; Laurier, D., et al. Lung cancer attributable to indoor radon exposure in france: impact of the risk models and uncertainty analysis; Environ Health Perspect. 2006. p. 1361-1366.Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1570096/?tool=pubmed

3. Palmarini M, Fan H. Retrovirus-induced ovine pulmonary adenocarcinoma, an animal model for lung cancer. J Natl Cancer Inst. 2001; 93:1603–1614. Available from: http://jnci.oxfordjournals.org/content/93/21/1603.long. [PubMed: 11698564]

4. Cheng YW, Chiou HL, Sheu GT, et al. The association of human papillomavirus 16/18 infection with lung cancer among nonsmoking Taiwanese women. Cancer Res. 2001; 61:2799–2803. Available from: http://cancerres.aacrjournals.org/content/61/7/2799.long. [PubMed: 11306446]

5. Fenske TS, McMahon C, Edwin D, et al. Identification of candidate alkylator-induced cancer susceptibility genes by whole genome scanning in mice. Cancer Res. 2006; 66:5029–5038. Available from: http://cancerres.aacrjournals.org/content/66/10/5029.full. [PubMed: 16707424]

6. Liu P, Wang Y, Vikis H, et al. Candidate lung tumor susceptibility genes identified through whole-genome association analyses in inbred mice. Nat Genet. 2006; 38:888–895. Available from: http://www.nature.com/ng/journal/v38/n8/abs/ng1849.html. [PubMed: 16862160]

7. Manenti G, Galvan A, Pettinicchio A, et al. Mouse genome-wide association mapping needs linkage analysis to avoid false-positive Loci. PLoS Genet. 2009; 5 e1000331. Available from: http://www.plosgenetics.org/article/info%3Adoi%2F10.1371%2Fjournal.pgen.1000331.

8. Wang D, You M. Five loci, SLT1 to SLT5, controlling the susceptibility to spontaneously occurring lung cancer in mice. Cancer Res. 2005; 65:8158–8165. Available from: http://cancerres.aacrjournals.org/content/65/18/8158.abstract. [PubMed: 16166290]

9. Cervino AC, Darvasi A, Fallahi M, Mader CC, Tsinoremas NF. An integrated in silico gene mapping strategy in inbred mice. Genetics. 2007; 175:321–333. Available from: http://www.genetics.org/content/175/1/321.long. [PubMed: 17028314]

10. Devlin B, Roeder K. Genomic control for association studies. Biometrics. 1999; 55:997–1004. [PubMed: 11315092]

11. Kang HM, Zaitlen NA, Wade CM, et al. Efficient control of population structure in model organism association mapping. Genetics. 2008; 178:1709–1723. Available from: http://cancerres.aacrjournals.org/content/65/18/8158.abstract. [PubMed: 18385116]

12. Petkov PM, Ding Y, Cassell MA, et al. An efficient SNP system for mouse genome scanning and elucidating strain relationships. Genome Res. 2004; 14:1806–1811. Available from: http://genome.cshlp.org/content/14/9/1806.long. [PubMed: 15342563]

13. Sundberg JP, Berndt A, Sundberg BA, et al. The mouse as a model for understanding chronic diseases of aging: the histopathologic basis of aging in inbred mice. Pathobiology of Aging & Age-related Diseases. 2011 in press. Available from: http://www.pathobiologyofaging.net/index.php/pba/article/viewArticle/7179/html_161.

14. Seymour, R.; Ichiki, T.; Mikaelian, I.; Boggess, D.; Silva, KA. Laboratory mouse: Necropsy methods. London: Academic Press; 2004.

15. Schofield PN, Dubus P, Klein L, et al. Pathology of the laboratory mouse: an International Workshop on Challenges for High Throughput Phenotyping. Toxicol Pathol. 2011; 39:559–562. Available from: http://tpx.sagepub.com/content/39/3/559.long. [PubMed: 21444746]

16. Austin CP, Battey JF, Bradley A, et al. The knockout mouse project. Nat Genet. 2004; 36:921–924. Available from: http://www.nature.com/ng/journal/v36/n9/full/ng0904-921.html. [PubMed: 15340423]

17. Grubb SC, Maddatu TP, Bult CJ, Bogue MA. Mouse phenome database. Nucleic Acids Res. 2009; 37:D720–D730. Available from: http://nar.oxfordjournals.org/content/37/suppl_1/D720.long. [PubMed: 18987003]

18. Nikitin AY, Alcaraz A, Anver MR, et al. Classification of proliferative pulmonary lesions of the mouse: recommendations of the mouse models of human cancers consortium. Cancer Res. 2004; 64:2307–2316. [PubMed: 15059877]

19. Hahn, FF.; Boorman, GA. Neoplasia and Preneoplasia of the Lung. In: Bannasch, P.; Goessner, W., editors. Pathology of Neoplasia and Preneoplasia in Rodents. Stuttgart, New York: Schattauer; 1997. p. 29-42.

20. Sundberg BA, Schofield PN, Gruenberger M, Sundberg JP. A data-capture tool for mouse pathology phenotyping. Vet Pathol. 2009; 46:1230–1240. Available from: http://vet.sagepub.com/content/46/6/1230.short?rss=1&ssource=mfc. [PubMed: 19605915]

21. Sundberg, JP.; Sundberg, BA.; Schofield, P. Integrating mouse anatomy and pathology ontologies into a phenotyping database: tools for data capture and training; Mamm Genome. 2008. p. 413-419.Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2844541/

22. Hutchins LN, Ding Y, Szatkiewicz JP, et al. CGDSNPdb: a database resource for error-checked and imputed mouse SNPs. Database (Oxford). 2010 baq008. Available from: http://database.oxfordjournals.org/content/2010/baq008.long.

23. Szatkiewicz, JP.; Beane, GL.; Ding, Y.; Hutchins, L.; Pardo-Manuel de Villena, F.; Churchill, GA. An imputed genotype resource for the laboratory mouse; Mamm Genome. 2008. p. 199-208.Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2725522/

24. Ng PC, Henikoff S. SIFT: Predicting amino acid changes that affect protein function. Nucleic Acids Res. 2003; 31:3812–3814. Available from: http://nar.oxfordjournals.org/content/31/13/3812.long. [PubMed: 12824425]

25. Krupke, DM.; Begley, DA.; Sundberg, JP.; Bult, CJ.; Eppig, JT. The Mouse Tumor Biology database; Nat Rev Cancer. 2008. p. 459-465.Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2574871/

26. Schofield PN, Bard JB, Boniver J, et al. Pathbase: a new reference resource and database for laboratory mouse pathology. Radiat Prot Dosimetry. 2004; 112:525–528. Available from: http://rpd.oxfordjournals.org/content/112/4/525.full. [PubMed: 15623888]

27. Schofield PN, Bard JB, Booth C, et al. Pathbase: a database of mutant mouse pathology. Nucleic Acids Res. 2004; 32:D512–D515. Available from: http://nar.oxfordjournals.org/content/32/suppl_1/D512.full. [PubMed: 14681470]

28. Devereux TR, Wiseman RW, Kaplan N, et al. Assignment of a locus for mouse lung tumor susceptibility to proximal chromosome 19. Mamm Genome. 1994; 5:749–755. [PubMed: 7894154]

29. Malkinson AM. The genetic basis of susceptibility to lung tumors in mice. Toxicology. 1989; 54:241–271. [PubMed: 2650016]

30. Drinkwater NR, Bennett LM. Genetic control of carcinogenesis in experimental animals. Prog Exp Tumor Res. 1991; 33:1–20. [PubMed: 2028021]

31. Engelhardt JA, Gries CL, Long GG. Incidence of spontaneous neoplastic and nonneoplastic lesions in Charles River CD-1 mice varies with breeding origin. Toxicol Pathol. 1993; 21:538–541. [PubMed: 8052799]

32. Rehm S, Rapp KG, Deerberg F. Influence of food restriction and body fat on life span and tumour incidence in female outbred Han:NMRI mice and two sublines. Z Versuchstierk. 1985; 27:240–283.

33. Rittinghausen, S.; Dungworth, DL.; Ernst, H.; Mohr, U. Pathobiology of the Aging Mouse. Washington, D.C.: 1996.

34. Festing MFW, Yang A, Malkinson AM. At least four genes and sex are associated with susceptibility to urethan-induced pulmonary adenomas in mice. Genet Res. 1994; 64:99–106. [PubMed: 7813906]

35. Zheng S, El-Naggar AK, Kim ES, Kurie JM, Lozano G. A genetic mouse model for metastatic lung cancer with gender differences in survival. Oncogene. 2007; 26:6896–6904. Available from: http://www.nature.com/onc/journal/v26/n48/full/1210493a.html. [PubMed: 17486075]

36. Hammoud Z, Tan B, Badve S, Bigsby RM. Estrogen promotes tumor progression in a genetically defined mouse model of lung adenocarcinoma. Endocr Relat Cancer. 2008; 15:475–483. Available from: http://erc.endocrinology-journals.org/content/15/2/475.abstract. [PubMed: 18509000]

37. Xing S, Tsaih SW, Yuan R, et al. Genetic influence on electrocardiogram time intervals and heart rate in aging mice. Am J Physiol Heart Circ Physiol. 2009; 296:H1907–H1913. Available from: http://ajpheart.physiology.org/content/296/6/H1907.abstract. [PubMed: 19395551]

38. Tsaih SW, Pezzolesi MG, Yuan R, Warram JH, Krolewski AS, Korstanje R. Genetic analysis of albuminuria in aging mice and concordance with loci for human diabetic nephropathy found in a

genome-wide association scan. Kidney Int. 2009; 77:201–210. Available from: http://www.nature.com/ki/journal/v77/n3/full/ki2009434a.html. [PubMed: 19924099]

39. Yuan R, Tsaih SW, Petkova SB, et al. Aging in inbred strains of mice: study design and interim report on median lifespans and circulating IGF1 levels. Aging Cell. 2009; 8:277–287. Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.1474-9726.2009.00478.x/abstract. [PubMed: 19627267]

40. Qi, C.; Zhu, YT.; Hu, L.; Zhu, YJ. Identification of Fat4 as a candidate tumor suppressor gene in breast cancers; Int J Cancer. 2009. p. 793-798.Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2667156/?tool=pubmed

41. Doi Y, Kawamata H, Ono Y, Fujimori T, Imai Y. Expression and cellular localization of TSC-22 in normal salivary glands and salivary gland tumors: implications for tumor cell differentiation. Oncol Rep. 2008; 19:609–616. Available from: http://www.spandidos-publications.com/or/19/3/609. [PubMed: 18288391]

42. Meijer D, Jansen MP, Look MP, et al. TSC22D1 and PSAP predict clinical outcome of tamoxifen treatment in patients with recurrent breast cancer. Breast Cancer Res Treat. 2009; 113:253–260. Available from: http://www.springerlink.com/content/p5249g74850061hl/. [PubMed: 18299979]

43. Rentsch CA, Cecchini MG, Schwaninger R, et al. Differential expression of TGFbeta-stimulated clone 22 in normal prostate and prostate cancer. Int J Cancer. 2006; 118:899–906. Available from: http://onlinelibrary.wiley.com/doi/10.1002/ijc.21449/abstract. [PubMed: 16106424]

44. Ohta S, Shimekake Y, Nagata K. Molecular cloning and characterization of a transcription factor for the C-type natriuretic peptide gene promoter. Eur J Biochem. 1996; 242:460–466. [PubMed: 9022669]

**Table 1**

Spontaneous pulmonary adenoma loci identified by previous GWAS studies.

| Author | Year | N of strains | N* of SNPs | Chr | Mb | Locus |
|---|---|---|---|---|---|---|
| Wang and You | 2005 | 13 | 135,900 | 6 | 84 | *Slt1* |
| | | | | 7 | 4 | *Slt2* |
| | | | | 8 | 123 | *Slt3* |
| | | | | 19 | 33 | *Slt4* |
| | | | | X | 138 | *Slt5* |
| Manenti et al. | 2009 | 28 | 13,000 | 5 | 150 | *Slt6* |

* = approximate number; N = number; Chr = chromosome; Mb = Million base pairs; *Slt* = spontaneous lung tumorigenesis

**Table 2**

Inbred strains used in previous and current genome-wide association studies for spontaneous pulmonary adenomas.

| Wang and You | Manenti et al. | Present Study |
|---|---|---|
|  |  | 129S1/SvlmJ |
| 129X1/SvJ | 129X1/SvJ |  |
| A/J | A/J | A/J |
| AKR/J | AKR/J |  |
| BALB/cJ | BALB/cJ |  |
|  |  | BALB/cByJ |
|  |  | BTBRT+*tf*/J |
|  |  | BUB/BnJ |
| C3H/HeJ | C3H/HeJ | C3H/HeJ |
|  | C57BL/10J | C57BL/10J |
| C57BL/6J | C57BL/6J | C57BL/6J |
|  |  | C57BLKS/J |
|  | C57BR/cdJ | C57BR/cdJ |
|  | C57L/J | C57L/J |
| CAST/EiJ | CAST/EiJ |  |
|  | CBA/J | CBA/J |
|  | CE/J |  |
|  | DBA/1J |  |
| DBA/2J | DBA/2J | DBA/2J |
|  | FVB/NJ | FVB/NJ |
|  |  | KK/HlJ |
| LP/J | LP/J | LP/J |
|  | MA/MyJ |  |
|  |  | MRL/MpJ |
|  |  | NOD.B10Sn-H2$^{b}$/J |
|  |  | NON/ShiLtJ |
| NZB/BlnJ | NZB/BlnJ |  |
|  |  | NZO/HlLtJ |
| NZW/LacCJ | NZW/LacJ | NZW/LacJ |
|  | O20 |  |
|  | P/J | P/J |
|  |  | PL/J |
|  |  | PWD/PhJ |
|  | RF/J |  |
|  | RIIIS/J | RIIIS/J |
|  | SJL/J |  |
| SM/J | SM/J | SM/J |
| SPRET/EiJ | SPRET/EiJ |  |

| Wang and You | Manenti et al. | Present Study |
| --- | --- | --- |
| | STS/A | |
| | SWR/J | SWR/J |
| | | WSB/J |

Gray shaded strains were used in multiple studies.

**Table 3**

Frequency and severity of pulmonary adenomas in females and males of 20-month-old mice of 28 strains.

| Sex | Strain | N of mice without adenomas | N of mice with adenomas | | | | | Grand Total |
|---|---|---|---|---|---|---|---|---|
| | | | S1 | S2 | S3 | S4 | % | |
| female | 129S1/SvImJ | 11 | 2 | | | | 15 | 13 |
| | A/J | 5 | 2 | | | | 29 | 7 |
| | BALB/cByJ | 13 | 1 | 1 | | | 13 | 15 |
| | BTBRT+tf/J | 7 | | | | | 0 | 7 |
| | BUB/BnJ | 9 | | | | | 0 | 9 |
| | C3H/HeJ | 6 | | | | | 0 | 6 |
| | C57BL/10J | 12 | | | | | 0 | 12 |
| | C57BL/6J | 15 | | | | | 0 | 15 |
| | C57BLKS/J | 13 | | | | | 0 | 13 |
| | C57BR/cdJ | 11 | | | | | 0 | 11 |
| | C57L/J | 14 | 1 | | | | 7 | 15 |
| | CBA/J | 12 | 1 | | | | 8 | 13 |
| | DBA/2J | 7 | | | | | 0 | 7 |
| | FVB/NJ | 5 | 1 | | | | 17 | 6 |
| | KK/HlJ | 7 | | | | | 0 | 7 |
| | LP/J | 10 | 2 | 1 | | | 23 | 13 |
| | MRL/MpJ | 7 | | | | | 0 | 7 |
| | NOD/LtJ | 8 | | | | | 0 | 8 |
| | NON/LtJ | 12 | | 1 | | | 8 | 13 |
| | NZO/HlLtJ | 4 | 2 | | | | 33 | 6 |
| | NZW/LacJ | 8 | | | | | 0 | 8 |
| | P/J | 10 | | | | | 0 | 10 |
| | PL/J | 5 | | | | | 0 | 5 |
| | PWD/PhJ | 12 | 1 | | | | 8 | 13 |
| | RIIIS/J | 10 | 2 | | 1 | 1 | 29 | 14 |
| | SM/J | 9 | 1 | 1 | | | 18 | 11 |
| | SWR/J | 3 | 1 | 1 | | | 40 | 5 |

| Sex | Strain | N of mice without adenomas | N of mice with adenomas | | | | | Grand Total |
|---|---|---|---|---|---|---|---|---|
| | | | S1 | S2 | S3 | S4 | % | |
| | WSB/EiJ | 10 | | 1 | | | 9 | 11 |
| female Total | | 255 | 17 | 6 | 1 | 1 | 9 | 280 |
| male | 129S1/SvImJ | 10 | 2 | 3 | | | 33 | 15 |
| | A/J | 3 | 4 | | 2 | | 67 | 9 |
| | BALB/cByJ | 6 | 3 | | | | 33 | 9 |
| | BTBRT+tf/J | 3 | | 3 | | | 50 | 6 |
| | BUB/BnJ | 2 | | | | | 0 | 2 |
| | C3H/HeJ | 10 | | | | | 0 | 10 |
| | C57BL/10J | 14 | | | | | 0 | 14 |
| | C57BL/6J | 13 | 1 | | | | 7 | 14 |
| | C57BLKS/J | 15 | | | | | 0 | 15 |
| | C57BR/cdJ | 13 | 2 | | | | 13 | 15 |
| | C57L/J | 12 | 2 | | | | 14 | 14 |
| | CBA/J | 10 | | | 1 | | 9 | 11 |
| | DBA/2J | 6 | | | | | 0 | 6 |
| | FVB/NJ | 5 | 2 | | | | 29 | 7 |
| | KK/HlJ | 7 | | | | | 0 | 7 |
| | LP/J | 10 | 3 | 1 | | | 29 | 14 |
| | MRL/MpJ | 10 | | | | | 0 | 10 |
| | NOD/LtJ | 8 | 1 | | | | 11 | 9 |
| | NON/LtJ | 10 | 2 | | | | 17 | 12 |
| | NZO/HlLtJ | 3 | 2 | | | | 40 | 5 |
| | NZW/LacJ | 10 | | | | | 0 | 10 |
| | P/J | 6 | | | | | 0 | 6 |
| | PL/J | 1 | | 1 | | | 50 | 2 |
| | PWD/PhJ | 9 | | 1 | | | 10 | 10 |
| | RIIIS/J | 9 | 1 | 2 | | | 25 | 12 |
| | SM/J | 10 | 2 | 1 | | | 23 | 13 |
| | SWR/J | 5 | 1 | 1 | | | 29 | 7 |

| Sex | Strain | N of mice without adenomas | N of mice with adenomas | | | | % | Grand Total |
|---|---|---|---|---|---|---|---|---|
| | | | S1 | S2 | S3 | S4 | | |
| | WSB/EiJ | 10 | 1 | | | | 9 | 11 |
| male Total | | 230 | 29 | 14 | 2 | | 16 | 275 |

N = number of mice; S1 = severity score of 1 (i.e., mild), S2 = severity score of 2 (i.e., moderate), S3 = severity score of 3 (i.e., severe), S4 = severity score of 4 (i.e., extreme)

**Table 4**

Number of unique and common SNPs among the 200 most significant associations for frequency and severity using 132k and 4 Mio SNP panels in GWAS.

| | 132k SNP panel | 4 Mio SNP panel | Common SNPs[*] |
|---|---|---|---|
| **Female mice** | | | |
| Frequency | 113 | 113 | 17 |
| Severity | 113 | 103 | 7 |
| Total | 226 | 216 | 24 |
| **Male mice** | | | |
| Frequency | 163 | 126 | 21 |
| Severity | 135 | 180 | 29 |
| Total | 298 | 306 | 50 |

[*] Common SNPs between 132k and 4 Mio SNP panels

**Table 5**

Location of the most significant SNP associations for pulmonary adenoma frequency and severity in male and female mice using a 132k SNP panel.

| Sex | Trait | Chr | N of SNPs[a] | Most significant SNP[b] | | |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | Position (Mb) | *P*-value | SNP ID |
| female | Frequency | 1 | 8 | 116.9 | 3.04E-06 | mm37-1-116948842 |
| | | 2 | 12 | 179.7 | 5.14E-06 | mm37-2-179749740 |
| | | 3 | 18 | 39.2 | 9.67E-06 | mm37-3-39229160 |
| | | 8 | 2 | 80.2 | 7.28E-05 | mm37-8-80162031 |
| | | 11 | 3 | 131 | 1.81E-05 | mm37-11-13101129 |
| | | 12 | 13 | 41.1 | 8.82E-06 | mm37-12-41147516 |
| | | 13 | 2 | 104.2 | 8.57E-06 | mm37-13-104204778 |
| | | 14 | 15 | 77.2 | 3.44E-07 | mm37-14-77179131 |
| | | 15 | 6 | 64.7 | 3.18E-05 | mm37-15-64685909 |
| | | 17 | 3 | 71 | 4.10E-05 | mm37-17-70991543 |
| | | 18 | 27 | 70.1 | 3.82E-06 | mm37-18-70084711 |
| | Severity | 1 | 3 | 116.9 | 3.62E-05 | mm37-1-116948842 |
| | | 2 | 3 | 29.6 | 1.87E-05 | mm37-2-29562500 |
| | | | 5 | 179.7 | 5.14E-05 | mm37-2-179749740 |
| | | 3 | 15 | 39.0 | 3.22E-06 | mm37-3-39033565 |
| | | | 4 | 71.1 | 4.84E-05 | mm37-3-71074670 |
| | | 4 | 2 | 100.4 | 1.24E-05 | mm37-4-100422365 |
| | | | 5 | 150.0 | 1.05E-06 | mm37-4-149964558 |
| | | 11 | 3 | 13.1 | 3.32E-05 | mm37-11-13076555 |
| | | 12 | 8 | 41.2 | 1.13E-05 | mm37-12-41217945 |
| | | 13 | 3 | 107.2 | 2.39E-05 | mm37-13-107195805 |
| | | 14 | 6 | 77.7 | 1.64E-06 | mm37-14-77693238 |
| | | 15 | 16 | 64.1 | 1.92E-06 | mm37-15-64134731 |
| | | 17 | 2 | 70.3 | 1.74E-05 | mm37-17-70295560 |
| | | 18 | 6 | 46–49 | 1.50E-05 | mm37-18-46455869 |
| | | | 2 | 60.7 | 3.07E-05 | mm37-18-60685321 |
| | | 19 | 2 | 21.5 | 4.84E-05 | mm37-19-21488761 |

| Sex | Trait | Chr | N of SNPs[a] | Most significant SNP[b] | | |
|---|---|---|---|---|---|---|
| | | | | Position (Mb) | P-value | SNP ID |
| | | | 7 | 27.5 | 3.70E-05 | mm37-19-27524472 |
| male | Frequency | 1 | 15 | 117.2 | 1.30E-09 | mm37-1-117172346 |
| | | 3 | 5 | 39.0 | 2.06E-08 | mm37-3-39013184 |
| | | 4 | 5 | 10.2 | 4.68E-07 | mm37-4-10243412 |
| | | 6 | 8 | 77.1 | 1.08E-07 | mm37-6-77167784 |
| | | | 4 | 147 | 8.89E-07 | mm37-6-146978458 |
| | | 10 | 64 | 3.1 | 2.30E-08 | mm37-10-3052424 |
| | | | 7 | 20.9 | 1.22E-07 | mm37-10-20873101 |
| | | | 3 | 122.1 | 7.82E-08 | mm37-10-122066015 |
| | | 11 | 2 | 67.1 | 5.42E-07 | mm37-11-67120998 |
| | | 13 | 22 | 4.2 | 1.61E-08 | mm37-13-4228684 |
| | | | 2 | 59.1 | 8.88E-08 | mm37-13-59092765 |
| | | 14 | 5 | 71.5 | 4.66E-08 | mm37-14-71474237 |
| | | 18 | 2 | 9.5 | 6.89E-07 | mm37-18-9463919 |
| | | | 9 | 62.4 | 1.01E-08 | mm37-18-62374081 |
| | | 19 | 2 | 47.8 | 3.90E-07 | mm37-19-47847846 |
| | Severity | 1 | 3 | 117.2 | 4.00E-08 | mm37-1-117172346 |
| | | 3 | 2 | 39 | 3.48E-08 | mm37-3-39013184 |
| | | 4 | 3 | 122.9 | 8.95E-09 | mm37-4-122905307 |
| | | 6 | 3 | 77.2 | 2.23E-07 | mm37-6-77167784 |
| | | | 22 | 85.5–90.3 | 8.44E-09 | mm37-6-90303862 |
| | | 7 | 15 | 127.9–134.1 | 7.48E-12 | mm37-7-133125150 |
| | | 10 | 25 | 3.1–8.5 | 6.57E-09 | mm37-10-3750455 |
| | | 12 | 4 | 114.6 | 2.00E-07 | mm37-12-11458210 |
| | | 13 | 5 | 4.9 | 1.25E-07 | mm37-13-4919866 |
| | | 17 | 6 | 74.4 | 1.74E-09 | mm37-17-74433060 |
| | | 18 | 10 | 9.5 | 2.08E-11 | mm37-18-9463919 |
| | | | 15 | 14.9 | 2.19E-07 | mm37-18-14935874 |
| | | | 7 | 62.4 | 1.41E-09 | mm37-18-62374081 |
| | | 19 | 3 | 27.7 | 1.30E-07 | mm37-19-27702998 |

| Sex | Trait | Chr | N of SNPs[a] | Most significant SNP[b] | | |
|---|---|---|---|---|---|---|
| | | | | Position (Mb) | P-value | SNP ID |
| | | | 3 | 60.6 | 2.63E-08 | mm37-19-60621343 |

Only peak associations with more than 1 SNP are listed. Single SNP associations are reported in Supplemental Table 1.

[a] total number of SNPs at peak association.

[b] most significant SNP at peak association. If the location of the SNPs were less than 2 Mb apart only the location of the Significant SNP was reported. If SNPs were in a continuous region the range of the region was reported.

N = Number; Mb = Million base pairs; SNP ID = SNP identification number from UCLA EMMA correction server SNP panels.

**Table 6**

Location of the most significant SNP associations for pulmonary adenoma frequency and severity in male and female mice using a 4 Mio SNP panel.

| Sex | Trait | Chr | N of SNPs[a] | Most significant SNP[b] Position (Mb) | P-value | SNP ID |
|---|---|---|---|---|---|---|
| female | Frequency | 1 | 16 | 116.9 | 3.04E-06 | mm37-1-116948842 |
| | | 2 | 2 | 179.7 | 5.14E-06 | mm37-2-179749740 |
| | | 11 | 9 | 131 | 1.08E-05 | mm37-11-13076555 |
| | | 12 | 63 | 41 | 2.49E-06 | NES15384728 |
| | | 14 | 17 | 77.2 | 3.44E-07 | mm37-14-77179131 |
| | | 18 | 4 | 70.1 | 1.08E-05 | mm37-18-60685321 |
| | Severity | 2 | 5 | 153.7 | 2.86E-06 | NES08911781 |
| | | 7 | 61 | 80.1 | 2.86E-06 | NES13954651 |
| | | 11 | 5 | 116.5 | 4.19E-06 | NES08527802 |
| | | 15 | 24 | 64.1 | 1.92E-06 | mm37-15-64134731 |
| | | 17 | 6 | 44.9 | 5.00E-06 | NES15205932 |
| male | Frequency | 1 | 41 | 117.1 | 3.38E-10 | NES16426410 |
| | | 3 | 26 | 38.9 | 7.32E-09 | NES12846366 |
| | | 4 | 8 | 108 | 6.11E-08 | NES08658635 |
| | | 6 | 7 | 145.7 | 7.69E-09 | NES13987853 |
| | | 7 | 4 | 41 | 5.03E-08 | NES11811971 |
| | | 10 | 16 | 3.1 | 2.30E-08 | mm37-10-3052424 |
| | | | 4 | 21.9 | 5.10E-10 | NES15226274 |
| | | 18 | 13 | 62.4 | 3.84E-09 | NES13455555 |
| | | 19 | 3 | 60.9 | 7.15E-08 | NES15820319 |
| | Severity | 1 | 7 | 115.8 | 3.96E-10 | NES12794203 |
| | | | 3 | 171.7 | 6.13E-09 | NES13039867 |
| | | | 26 | 123.9 | 8.95E-09 | NES09185926 |
| | | 6 | 78 | 88.4 | 2.51E-09 | mm37-6-88424482 |
| | | | 5 | 145.7 | 6.04E-10 | NES13987853 |
| | | 7 | 16 | 128.7 | 7.54E-10 | NES11585821 |

NIH-PA Author Manuscript  NIH-PA Author Manuscript  NIH-PA Author Manuscript

| Sex | Trait | Chr | N of SNPs[a] | Most significant SNP[b] | | |
| --- | --- | --- | --- | --- | --- | --- |
| | | | | Position (Mb) | P-value | SNP ID |
| | | 10 | 2 | 3.8 | 1.46E-09 | NES15060986 |
| | | 18 | 2 | 9.6 | 2.08E-11 | mm37-18-9616220 |
| | | | 11 | 62.4 | 1.04E-09 | NES13455555 |
| | | 19 | 21 | 60.9 | 5.98E-10 | NES15820137 |

Only peak associations with more than 1 SNP are listed. Single SNP associations are reported in Supplemental Table 1

[a] total number of SNPs at peak association.

[b] most significant SNP at peak association. If the location of the SNPs were less than 2 Mb apart only the location of the Significant SNP was reported. If SNPs were in a continuous region the range of the region was reported.

N = Number; Mb = Million base pairs; SNP ID = SNP identification number from UCLA EMMA correction server SNP panels.

**Table 7**

Candidate genes and their Cn SNPs for spontaneous pulmonary adenoma frequency and severity in female and male mice.

| Chr | Region (Mb) | Phenotype/Sex | Gene | Exon[a] | Rs number | Alleles[b] | P-value[d] | AA change[c] | SIFT[d] |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 114.8 – 118.9 | Freq/F & Freq/M | Cntnap5a | 19 | rs30647215 | T/C | 0.006 & 0.03 | L1034S[c] | tolerated |
| | 122.9 – 124.9 | Sever/ M | Ccdc93 | 17 | rs30776698 | A/G | 0.024 | I432V[c] | tolerated |
| | 170.7 – 172.7 | Sever/ M | Sh2d1b1 | 3 | rs31466875 | A/T | 0.029 | T71S | tolerated |
| 2 | 152.7 – 154.7 | Sever/F | Psp | 4 | rs13476849 | T/G | 0.046 | D111E | tolerated |
| | | | | 5 | rs27347468 | C/T | 0.046 | T130M | tolerated |
| | | | | 6 | rs27347448 | G/C | 0.046 | V198L | tolerated |
| | | | | 7 | rs27347441 | A/G | 0.046 | S212G | tolerated |
| | | | | 7 | rs27347440 | G/A | 0.046 | S217N | tolerated |
| | 178.7 – 180.7 | Freq/F | Lsm14b | 7 | rs27664746 | A/G | < 0.001 | A213T[c] | tolerated |
| | | | Lama5 | 75 | rs27679307 | C/T | 0.008 | R3495Q[c] | tolerated |
| | | | | 50 | rs27679266 | C/T | 0.021 | A2231T[c] | tolerated |
| | | | | 48 | rs27679257 | T/C | 0.045 | H2132R[c] | tolerated |
| | | | Dido1 | 12/6 | rs27664132 | T/C | 0.040 | M960V[c] | tolerated |
| | | | | 12/6 | rs27664130 | T/C | 0.030 | N890D[c] | tolerated |
| 3 | 37.9 – 39.9 | Freq/M | Fat4 | 9 | rs29949720 | C/G | 0.039 | H3805Q | yes |
| | | | | 9 | rs29949718 | G/A | 0.018 | I3916V | tolerated |
| 6 | 144.7 – 147.7 | Freq/M & Sever/M | Lrmp | 19 | rs13459098 | T/C | Sever: 0.045 | L537P[c] | tolerated |
| | | | Casc1 | 4/2 | rs32247843 | C/T | 0.007 & 0.009 | S60N[c] | tolerated |
| | | | Ifltd1 | 6/5/4 | rs36862482 | T/C | 0.008 & 0.008 | H218R[c] | tolerated |
| | | | Itpr2 | 13/12 | rs37701645 | C/G | 0.005 & 0.009 | V451L[c] | tolerated |
| 11 | 115.5 – 117.5 | Sever/F | Itgb4 | 29/30/31 | rs27001410 | C/A | 0.041 | R1263S[c] | tolerated |
| | | | | 33/34/35/36 | rs27001391 | T/C | 0.044 | M1536T[c] | tolerated |
| 14 | 76 – 78 | Freq/F | Tsc22d1 | 1/3 | rs30156847 | A/G | 0.012 | I591V[c] | tolerated |
| | | | | 1/3 | rs31392028 | G/C | < 0.001 | A714P[c] | yes |
| | | | Ccdc122 | 4/6 | rs31509286 | C/T | 0.003 | P285L[c] | yes |

| Chr | Region (Mb) | Phenotype/Sex | Gene | Exon[a] | Rs number | Alleles[b] | P-value[d] | AA change[c] | SIFT[d] |
|-----|-------------|---------------|------|---------|-----------|------------|------------|--------------|---------|
|     | 122 – 124   |               | *Nalcn* | 44   | rs30499120 | T/A | 0.029 | T1699S | tolerated |
| 17  | 43.9 – 45.9 | Sever/F       | *Enpp5* | 4    | rs33653328 | C/A | 0.042 | H461N | tolerated |
| 18  | 61.4 – 63.4 | Freq/M        | *Fbxo38* | 15  | rs36884937 | G/A | 0.027 | T659M | yes |
| 19  | 59.9 – 61.9 | Freq/M & Sever/M | *Sfxn4* | 9 | rs47497127 | G/A | 0.003 & 0.005 | S148F | tolerated |
|     |             |               |      | 8       | rs46394777 | C/T | 0.003 & 0.005 | G115S[c] | tolerated |

Chr = chromosome; Mb = million base pairs; AA = amino acid; Freq = pulmonary adenoma frequency; Sever = pulmonary adenoma severity; F = female mice; M = male mice;

[a] multiple if multiple transcripts exist for which exon number is different;

[b] the C57BL/6J allele is used as the reference allele;

[c] if multiple positions due to multiple transcripts exist (http://useast.ensembl.org/);

[d] SIFT, a tool by the Craig Venter Institute that predicts if Cn SNPs may change protein function, AA changes that cause a protein function change are marked with 'yes' and AA changes that do not cause protein function changes are marked with 'tolerated';

[d] strains were grouped according to their allele at the SNP and groups were compared using t-test analysis