Identification of *Las2*, a Major Modifier Gene Affecting the Pas1 Mouse Lung Tumor Susceptibility Locus

Peng-Yuan Liu, Haris Vikis, Michael James, Yan Lu, Dao-Long Wang, Hong-Bo Liu, Wei-Dong Wen, Yian Wang and Ming You

Supplementary Data

Supplementary Methods

Lung tumor susceptibility and SNP data. The lung tumor data were collected from several previous studies (**Supplementary Table 1**). 593 mice from 21 inbred strains were measured for lung tumor multiplicity (tumors per mouse) at 14-16 weeks after a single injection of urethane (1 mg/g). The analyzed strains were derived from different genealogies, including 11 Castle's mice (129S1/SvJmJ, A/J, AKR/J, BALB/c, C3H/HeJ, CBA/J, DBA/2J, LP/J, NZB/BlNJ, O20 and SM/J), 5 C57-related strains (C57BL/10J, C57BL/6J, C57BR/cdJ, C57L/J and MA/MyJ), 2 Swiss mice (SJL/J and SWR/J), and 3 other inbred strains (PL/J, RIIIS/J and ST/bJ). The mean tumor multiplicity ranged from 0 to 27.5 with a mean of 4.4, suggesting a wide range of variation in lung tumorigenesis in the sample. The genotyped SNP data were obtained from the Mouse Phenome Database (MPD) (http://phenome.jax.org/) and the imputed SNP data were obtained from the Center for Genome Dynamics (CGD) at The Jackson Laboratory (Bar Harbor, ME) (http://cgd.jax.org/index.php). The genotyped data contained 190,903 SNPs on commonly used mouse inbred strains. The SNP genotyping accuracy reported by the Wellcome Trust Centre for Human Genetics and the Broad Institute is over 99.8%. We removed SNPs with less than 16 strains typed or without genetic mapping information. Each SNP allele must be present in two or more inbred strains. The resulting data consisted of 115,904 genotyped SNPs, spanning the mouse genome at an average density of approximately 20 kb per SNP. The imputed data contained 7,870,134 imputed SNPs with confidence levels greater than 0.9 (1). Similar filtering criteria were applied to these imputed SNPs. As a result, 1,952,918 imputed SNPs were retained in our final analysis, which span the mouse genome at an average density of approximately 1.1 kb per SNP.

Statistical analysis. Prior to statistical analysis, the raw tumor multiplicity data were converted to approach normality by Box-Cox transformation, $[(y+1)^{\lambda}-1]/\lambda$, where $\lambda = -0.6017$. To correct for population structure and genetic relatedness among inbred strains, we used a recently developed method, efficient mixed-model association (EMMA), to assess association of lung tumor susceptibility with SNPs (2). Specifically, the mixed model in the EMMA method can be represented by:

$$\mathbf{y} = \mathbf{X\beta} + \mathbf{Zu} + \mathbf{e}$$

where y is an n × 1 vector of observed phenotypes (i.e., lung tumor multiplicity), and X is an n × q matrix of fixed effects including mean, SNPs, and other covariate variables. $\boldsymbol{\beta}$ is a q × 1 vector representing coefficients of the fixed effects. Epistatic interaction can be also modeled into fixed effects. *Z* is an n × t incidence matrix mapping each observed phenotype to one of t inbred strains. *u* is the random effect (i.e., strain effects) of the mixed model with Var(u) = $2\mathbf{K}\sigma_g^2$ where K is the t × t kinship matrix inferred from genotypes, and *e* is an n × n matrix of residual effect such that Var(e) = $\mathbf{I}\sigma_e^2$. The overall phenotypic variance–covariance matrix can be represented as $\mathbf{V} = 2\mathbf{ZKZ}'\sigma_g^2 + \mathbf{I}\sigma_e^2$. An R package implementation of the EMMA method is publicly available (http://mouse.cs.ucla.edu/emma/). In the EMMA approach, a kinship matrix based on genetic similarity was inferred from SNP genotype data. However, most MPD SNPs were initially discovered by comparison of the genomes of several classical inbred laboratory mouse strains (such as C57BL/6J and 129S1/SvImJ). Although this SNP ascertainment will not likely introduce false positives in association mapping, it certainly affects any inference about the population based on these SNPs (3). Recently, Perlegen Science reported 8.27 million high-quality SNPs by resequencing the genomes of four wild-derived and eleven classical strains (4), which is a more representative set of SNPs. Therefore, to correctly infer a kinship among inbred strains, we randomly chose 624 SNPs from the Perlegen SNP database (http://mouse.perlegen.com/mouse/download.html) and typed these SNPs in 56 commonly used inbred strains to infer kinship matrix. SNP genotyping was performed by the Sequenom mass array spectrometry system at the Human Genetics Division Genotyping Core of Washington University (St Louis, MO). In addition to single SNP analyses, we also performed haplotype-based association analyses using a sliding window of three SNPs. Similarly, the haplotype-based association analyses were implemented in the EMMA R package. The results from haplotype analyses were generally consistent with the above single SNP analyses (data not shown). Particularly, in the refined region at the Pas1 and Par2 loci where SNPs show complete linkage disequilibria, the results from haplotype analyses are identical to those from single SNP analyses.

Two genome search strategies. In the conventional genome search for QTLs, one SNP was tested for association at a time. The EMMA approach was used for the association analysis (2). A two-sided p value for each SNP was obtained for testing hypothesis of no association between the SNP and lung tumor multiplicity. In the genetic background-controlled search, the Pas1

genetic background was taken into account for association tests (5). Specifically, the tumor multiplicity was first adjusted by the *Pas1* allelic status using a simple linear regression model. The residual from this regression model was then used as a phenotypic variable for the association analysis. Once a SNP is identified with strong association, the epistatic interaction between that SNP and the *Pas1* was further assessed by the EMMA approach (2).

LOH analysis. The LOH data were downloaded from the Tumor Sequencing Project (TSP) public datasets (http://caintegrator.nci.nih.gov/csp). Specifically, 384 lung adenocarcinomas and matched normal DNA were used in the LOH analysis, which were genotyped with the StyI chip of the 500K Human Mapping Array set (Affymetrix Inc). SNP arrays were processed as a plate of 96 samples using the GenePattern software package with modules based on dChipSNP algorithms. LOH calls were produced using the LOHPaired GenePattern module. The genotype of each SNP in a tumor was compared directly to the corresponding genotype in its paired normal sample to give a heterozygosity call of Loss, Retention, Noninformative or Conflict. We defines a deletion region as a minimum of 15 consecutive SNPs having the following characteristics (6): (1) there be a minimum of 10 informative (heterozygous) normal samples at each tested SNPs; (2) the SNPs be mapped to the NCBI human genome build 36.2; and (3) LOH frequencies be greater than 50%.

Microsatellite markers were amplified using [α^{32}P]dCTP and using primers specified in the NCBI UniSTS database. Products were separated on 8% polyacrylamide gels and exposed to film. Loss of heterozygosity at microsatellite markers was measured by band intensity measured using ImageJ software (NIH). Loss of heterozygosity was defined by an alelle ratio (T1/T2 : N1/N2) of $\leq$ .67 or $\geq$ 1.5.

References

1. Szatkiewicz JP, Beane GL, Ding Y, et al. An imputed genotype resource for the laboratory mouse. Mamm Genome 2008; 19: 199-208.
2. Kang HM, Zaitlen NA, Wade CM, et al. Efficient control of population structure in model organism association mapping. Genetics 2008; 178: 1709-23.
3. Clark AG, Hubisz MJ, Bustamante CD, Williamson SH, Nielsen R. Ascertainment bias in studies of human genome-wide polymorphism. Genome Res 2005; 15: 1496-502.
4. Frazer KA, Eskin E, Kang HM, et al. A sequence-based variation map of 8.27 million SNPs in inbred mouse strains. Nature 2007; 448: 1050-3.
5. Zeng ZB. Precision mapping of quantitative trait loci. Genetics 1994; 136: 1457-68.
6. Hu N, Wang C, Hu Y, et al. Genome-wide loss of heterozygosity and copy number alteration in esophageal squamous cell carcinoma using the Affymetrix GeneChip Mapping 10 K array. BMC Genomics 2006; 7: 299.
7. Malkinson AM. The genetic basis of susceptibility to lung tumors in mice. Toxicology 1989; 54: 241-71.
8. Manenti G, Dragani TA. Pas1 haplotype-dependent genetic predisposition to lung tumorigenesis in rodents: a meta-analysis. Carcinogenesis 2005; 26: 875-82.
9. Malkinson AM, Beer DS. Major effect on susceptibility to urethan-induced pulmonary adenoma by a single gene in BALB/cBy mice. J Natl Cancer Inst 1983; 70: 931-6.
10. Manenti G, Acevedo A, Galbiati F, et al. Cancer modifier alleles inhibiting lung tumorigenesis are common in inbred mouse strains. Int J Cancer 2002; 99: 555-9.
11. Manenti G, Gariboldi M, Fiorino A, et al. Pas1 is a common lung cancer susceptibility locus in three mouse strains. Mamm Genome 1997; 8: 801-4.
12. Falconer DS, Bloom JL. Changes In Susceptibility To Urethane-Induced Lung Tumours Produced By Selective Breeding In Mice. Br J Cancer 1964; 18: 322-32.
13. Trainin N, Precerutti A, Law LW. Trends In Carcinogenesis By Urethan Administration To New-Born Mice Of Different Strains. Nature 1964; 202: 305-6.
14. Thaete LG, Nesbitt MN, Malkinson AM. Lung adenoma structure among inbred strains of mice: the pulmonary adenoma histologic type (Pah) genes. Cancer Lett 1991; 61: 15-20.
15. Festing MF, Lin L, Devereux TR, et al. At least four loci and gender are associated with susceptibility to the chemical induction of lung adenomas in A/J x BALB/c mice. Genomics 1998; 53: 129-36.
16. Obata M, Nishimori H, Ogawa K, Lee GH. Identification of the Par2 (Pulmonary adenoma resistance) locus on mouse chromosome 18, a major genetic determinant for lung carcinogen resistance in BALB/cByJ mice. Oncogene 1996; 13: 1599-604.
17. Manenti G, Gariboldi M, Fiorino A, et al. Genetic mapping of lung cancer modifier loci specifically affecting tumor initiation and progression. Cancer Res 1997; 57: 4164-6.
18. Lee GH, Matsushita H, Kitagawa T. Fine chromosomal localization of the mouse Par2 gene that confers resistance against urethane-induction of pulmonary adenomas. Oncogene 2001; 20: 3979-85.
19. Zhang Z, Lin L, Liu G, et al. Fine mapping and characterization of candidate lung tumor resistance genes for the Par2 locus on mouse chromosome 18. Exp Lung Res 2000; 26: 627-39.

Supplementary Table S1 Urethane-induced lung tumor multiplicity among inbred strains

Strains	Mean multiplicity*	No. of mice
C57BR/cdJ	0.00	10
AKR/J	0.10	11
C57L/J	0.10	10
C3H/HeJ	0.20	55
SJL/J	0.29	7
NZN/B1NJ	0.30	12
C57BL/10J	0.31	19
DBA/2J	0.40	20
C57BL/6J	0.50	27
SM/J	0.50	10
LP/J	1.10	8
CBA/J	1.70	40
PL/J	2.00	9
129S1/SvJmJ	2.10	9
ST/bJ	3.20	7
BALB/c	3.30	95
RIIIS/J	6.70	48
MA/MyJ	8.90	9
O20	12.10	20
SWR/J	20.10	41
A/J	27.50	126

*The mean tumor multiplicity (tumors per mouse) was recorded at 14-16 weeks after a single injection of urethane (1 mg/g) (from ref. (7-14)).

Supplementary Table S2 Significant SNPs identified in the *Pas1* locus using the conventional genome search.

dbSNP	Chr	Pos*	Minor allele	Maf†	P value
rs30843330	6	144,898,833	A	0.40	1.24E-05
rs30118733	6	144,908,118	G	0.37	3.78E-05
rs3661605	6	144,917,954	C	0.40	1.24E-05
rs3669753	6	144,922,189	G	0.40	1.24E-05
rs3671011	6	144,922,401	A	0.40	1.24E-05
rs30766031	6	144,937,701	C	0.42	2.46E-06
rs30219361	6	144,937,898	G	0.44	3.85E-05
rs30858405	6	144,939,800	A	0.40	1.24E-05
rs30937125	6	144,940,700	C	0.40	1.24E-05
rs30074927	6	144,964,523	T	0.40	1.24E-05
rs32239397	6	144,989,793	G	0.50	5.34E-05
rs30117907	6	144,990,941	C	0.40	1.24E-05
rs13459098	6	145,123,190	T	0.43	1.07E-05
rs30514198	6	145,127,703	T	0.40	1.24E-05
rs30495428	6	145,145,489	G	0.40	1.24E-05
rs29921541	6	145,145,851	T	0.37	3.65E-05
rs47252111	6	145,204,498	T	0.40	1.24E-05
rs13479086	6	145,244,336	A	0.45	1.68E-05
rs30414389	6	145,246,380	C	0.42	1.98E-05
rs30112226	6	145,257,630	C	0.40	1.24E-05
rs30655674	6	145,265,429	G	0.40	1.24E-05
rs30161519	6	145,307,624	A	0.35	9.11E-05
rs29967373	6	145,325,876	T	0.40	1.24E-05
rs30851797	6	145,351,119	C	0.50	4.54E-06
rs13479046	6	145,351,205	T	0.42	1.98E-05
rs30271782	6	145,364,721	T	0.40	1.24E-05
rs30709422	6	145,371,339	C	0.40	1.24E-05
rs30029517	6	145,372,552	G	0.40	1.24E-05

*The SNP positions (bp) were based on the NCBI mouse genome build 37.1.

†Minor allele frequency in the sample of inbred mice

Supplementary Table S3 Other loci associated with lung tumor susceptibility.

dbSNP*	Chr	Pos	Minor allele	Maf	P value†
rs28272180	2	110,709,431	A	0.45	1.00E-05
rs28255336	2	110,736,542	A	0.45	1.00E-05
rs13477626	4	30,326,006	A	0.47	4.65E-06
rs30644386	19	12,846,315	A	0.38	2.28E-06
rs30641699	19	12,877,218	C	0.38	2.28E-06
rs30756662	19	12,891,228	C	0.38	9.78E-06
rs30756662	19	12,891,228	C	0.37	2.28E-06
rs31044785	19	12,891,663	T	0.35	6.45E-06
rs40024169	19	12,896,063	T	0.35	6.45E-06
rs30990597	19	12,898,631	T	0.35	6.45E-06
rs31197265	19	12,899,887	C	0.35	6.45E-06
rs36938738	19	12,900,284	T	0.35	6.45E-06
rs36450565	19	12,900,304	C	0.35	6.45E-06

*SNPs in *italic* are imputed SNPs.

†SNPs with P values ≤ 1.00E-05 were presented.

Supplementary Table S4 Significant SNPs/polymorphisms in the *Par2* locus using the genetic background-controlled genome search.

dbSNP/polymorphism	Chr	Pos	Minor allele	Maf	P value*
rs29622203	18	70,602,803	G	0.40	7.09E-05
rs29557643	18	70,613,785	A	0.42	1.08E-04
rs30301901	18	70,622,787	T	0.42	5.59E-05
rs29686328	18	70,622,848	T	0.42	5.59E-05
rs30273275	18	70,623,572	T	0.43	9.33E-05
rs30120015	18	70,626,673	T	0.43	9.33E-05
Las2-3-bp-deletion (234 codon)	18	70,626,717-70,626,719	del	0.40	1.68E-04
rs29675801	18	70,628,766	A	0.43	9.33E-05
rs30259292	18	70,628,787	G	0.43	9.33E-05
rs30245983/*Las2*-T94I	18	70,629,114	T	0.43	9.33E-05
Las2-S37A	18	70,629,286	G	0.40	1.68E-04
Las2-9-bp-insertion (32-34 codons)	18	70,629,292-70,629,301	ins	0.40	1.68E-04
Las2-S30R	18	70,629,307	C	0.40	1.68E-04
rs30066196/*Las2*-R13Q	18	70,629,357	A	0.43	9.33E-05
rs51575278	18	70,630,714	G	0.43	9.33E-05
rs46078787	18	70,630,994	T	0.43	9.33E-05
rs49840286	18	70,631,087	A	0.43	9.33E-05
rs50790044	18	70,631,338	A	0.43	9.33E-05
rs45700290	18	70,637,806	T	0.43	9.33E-05
rs48238521	18	70,638,098	A	0.43	9.33E-05

*Sequence variants R13Q and T94I are identical to SNPs rs30066196 and rs30245983, respectively. These two SNPs have available genotype data in strain O20 in the SNP panel from the MPD. Therefore, R13Q and T94I yielded smaller p values from statistical tests than the other coding sequence variants in the *Las2*.

Supplementary Figure Legends

Supplementary Figure S1 Cumulative distribution of P values from the EMMA method in the GWAS. CDF refers to cumulative distribution function. (**a**) Conventional genome search. (**b**) Genetic background-controlled search.

Supplementary Figure S2 Fine mapping of the *Pas1* locus on chromosome 6. (**a**) Association mapping of lung tumor multiplicity in chromosome 6. (**b)** Enhanced view of association mapping between 143.0-147.0 Mb. (**c**) Physical map of the complex multigenic *Pas1* locus. The annotated genes were based on the NCBI mouse genome build 37.1. Genes shown on the upper side of the chromosome (turquoise lines) are transcribed in the – orientation (from right to left), and those on the lower side (pink lines) in the + orientation (from left to right).

Supplementary Figure S3 Association mapping in chromosome 18 using the conventional genome search. The region (green) between D18Mit152 and D18Mit162 is one-LOD support interval from previous linkage analyses (15-17), and the region (brown) between D18Mit103 and D18Mit188 is the refined region by congenic mice (18, 19). The strongest association was located at 60.81 Mb on distal chromosome 18, which is about 10 Mb apart from the refined region from fine mapping. The *Par2* locus was not identified by the GWAS using the conventional genome search.

Supplementary Figure S4 Lung tumor multiplicities (mean ± SE) of inbred mice with different genotypes at *Pas1* and *Par2* loci. SS and RR indicate susceptible and resistant alleles at *Pas1* or *Par2* locus, respectively. One-way ANOVA was used to test difference in lung tumor multiplicities of inbred mice among different genotypes at *Pas1* and *Par2* loci (P = 6.68E-05).

Supplementary Figure S5 Expression of *Las2*. (**a**) *Las2* expression in mouse tissues was determined by quantitative PCR (PCR primers spanning exons 6 to 9) using cDNA panel (Clontech, Mountain View, CA). (**b**) *Las2* mRNA levels (mean ± SE) of the top five resistant strains (C57BR/cdJ, AKR/J, C57L/J, C3H/HeJ and SJL/J) and the top five susceptible strains (BALB/cByJ, RIIIS/J, MA/MyJ, SWR/J and A/J). No significant correlation between *Las2* mRNA levels and tumor multiplicity was observed (P > 0.05).

Supplementary Figure S6 Genomic organization of 4930503L19RIK/*Las2* and location of non-synonymous variants. *Las2* gene has nine exons encoding for a transcript of 2,291-bp and

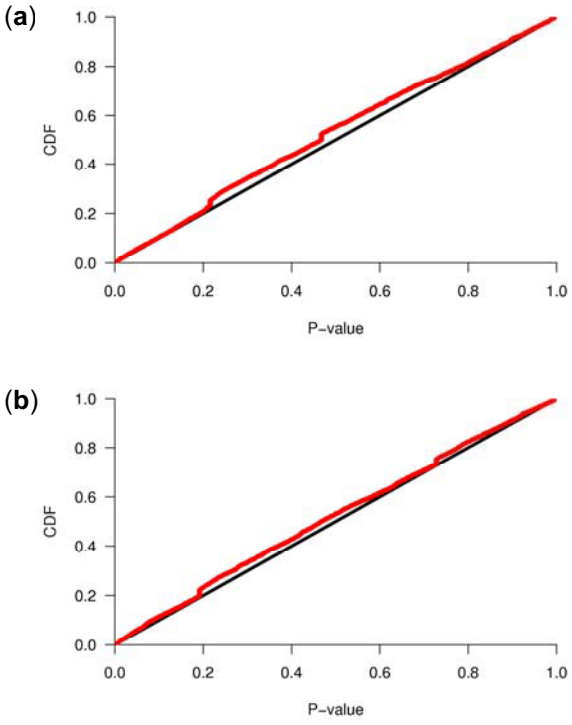
protein of 526 residues. Exons are indicated with relative positions of the sequence variants detected by sequencing. T94I is created by non-synonymous SNP rs30245983. Coding exons are indicated by filled in boxes.

Supplementary Figure S7 Alignment of *Las2* protein sequence of multiple species. (**a**) *Las2* protein identity and similarity. (**b**) Alignment of *Las2* protein sequence*s. Las2* protein sequence from C57BL/6J that contains susceptible alleles at the *Par2* locus was used as a query sequence to search putative orthologues using the latest NCBI non-redundant protein sequences database. The best hit to each organism in the BLAST results was present. Residues identical in at least five species are shown with black backgrounds.

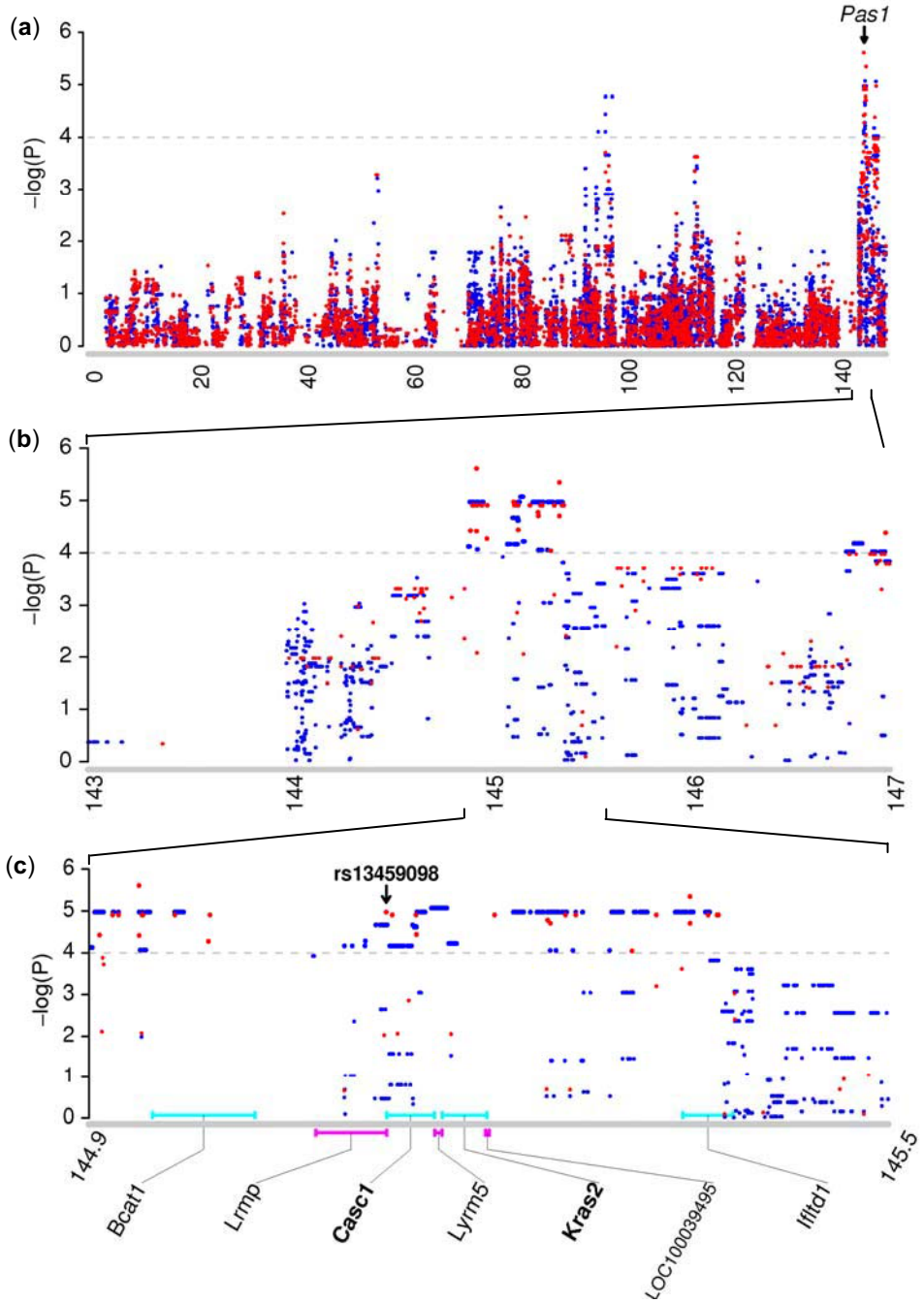
Supplementary Figure S8 Light microscopy of LM2 cells. Colonies of LM2 cells overexpressing *Las2-R* were somewhat dispersed compared to *Las2-S* overexpressing and vector alone LM2 colonies. These *Las2-R* overexpressing cells were also slightly larger than *Las2-S* or vector alone LM2.

Supplementary Figure S9 *LAS2* loss in human lung cancer from the analysis of SNP chips. (**a**) Deletions in human chromosome 18q21.2 (49-51 Mb). Red points indicate the average LOH frequency at tested SNPs and black bars indicate deletion regions. Two deletions were identified around the *LAS2* (C18orf54) region in lung tumors. One is located at 49.766-49.911 Mb; another is located at 49.990-50.216 Mb which covers *LAS2*. LOH was inferred from patient-matched normal DNA samples. (**b**) Physical map of *LAS2* region. The annotated genes were based on the NCBI human genome build 36.2. (**c**) LOH at the deletion region (49.990-50.216 Mb) among 384 lung adenocarcinomas. Each column indicates a SNP and each row indicates a case sample. Red represents LOH; blue represents retention; gray represents conflict; and blank represents non-informative at tested SNPs.

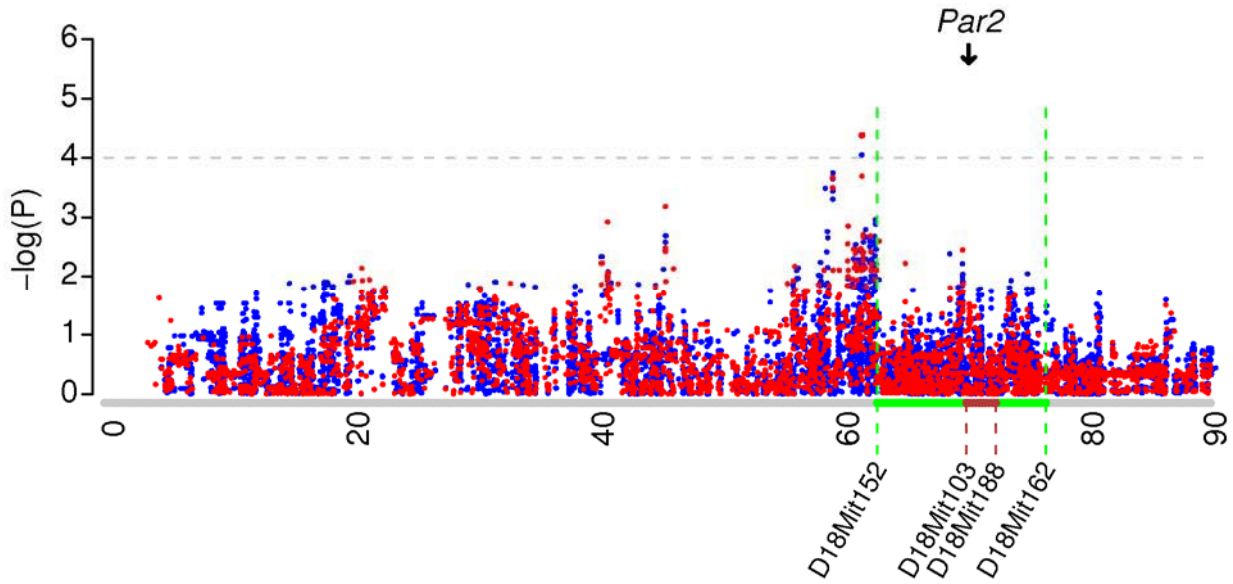
Supplementary Figure S1 Cumulative distribution of P values from the EMMA approach in the genome-wide association study



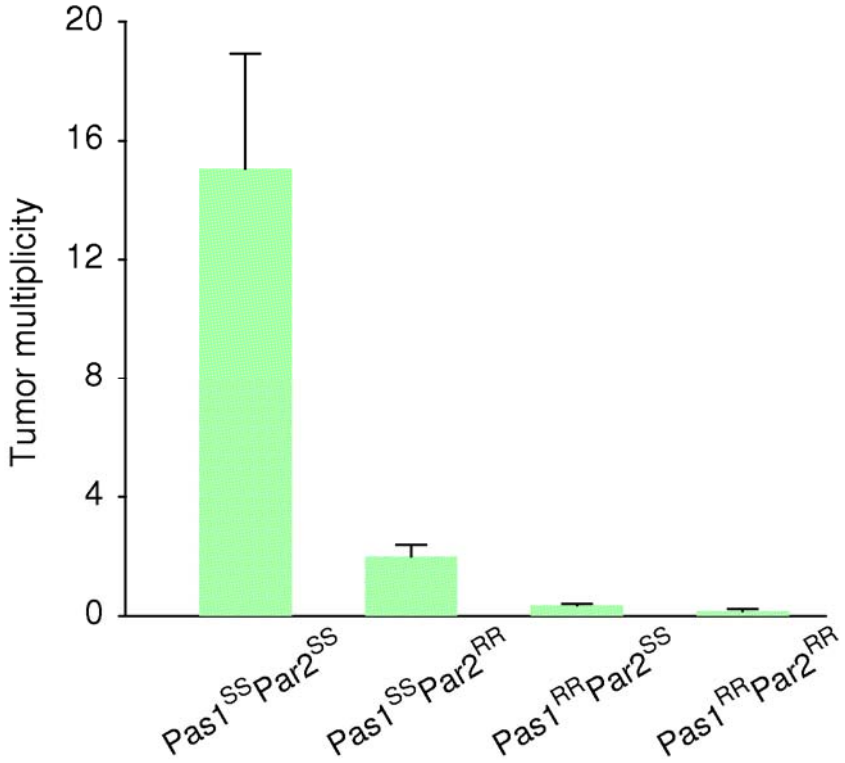
Supplementary Figure S2 Fine mapping of the *Pas1* locus on chromosome 6



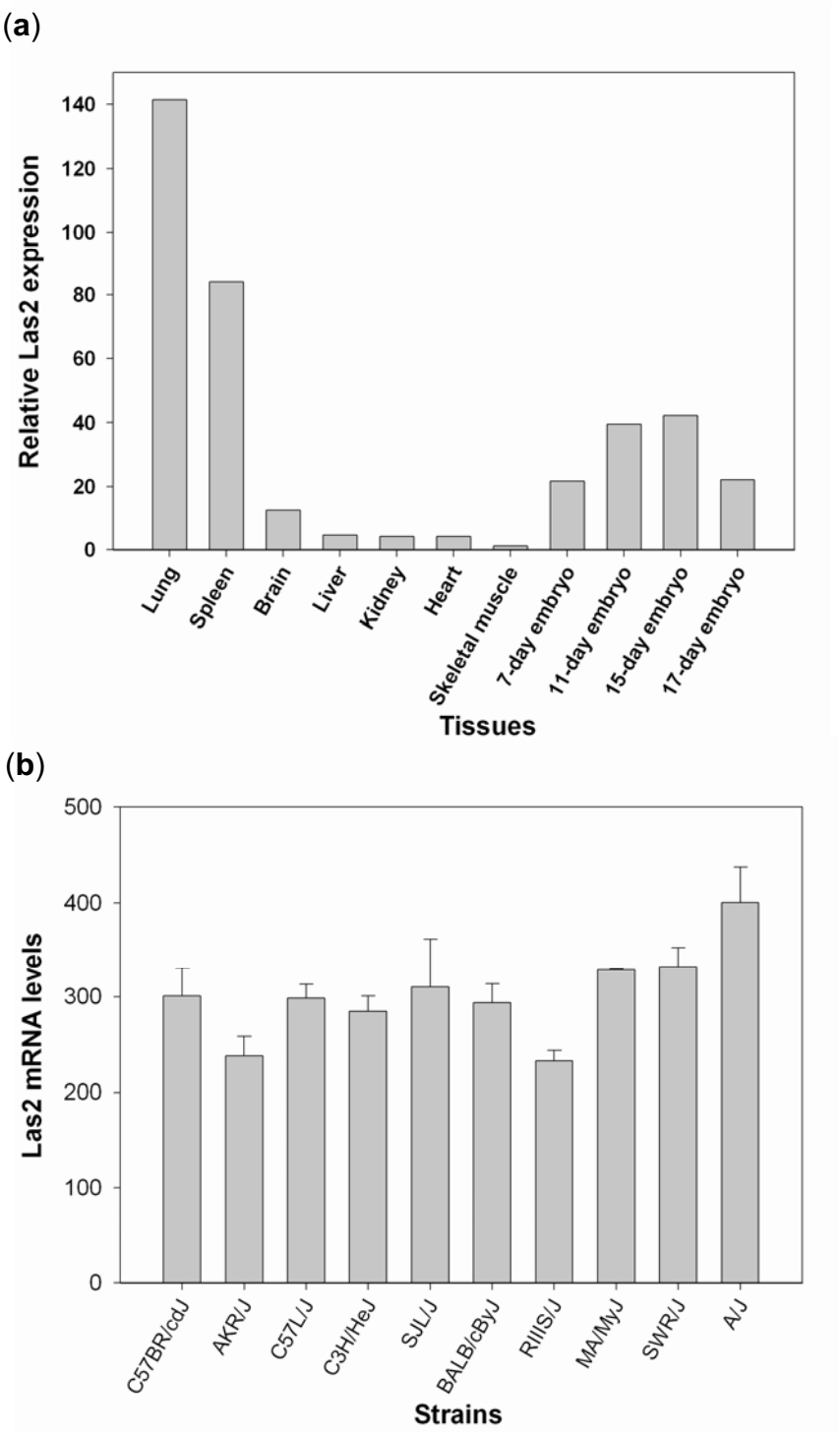
Supplementary Figure S3 Association mapping in chromosome 18 using the conventional genome search



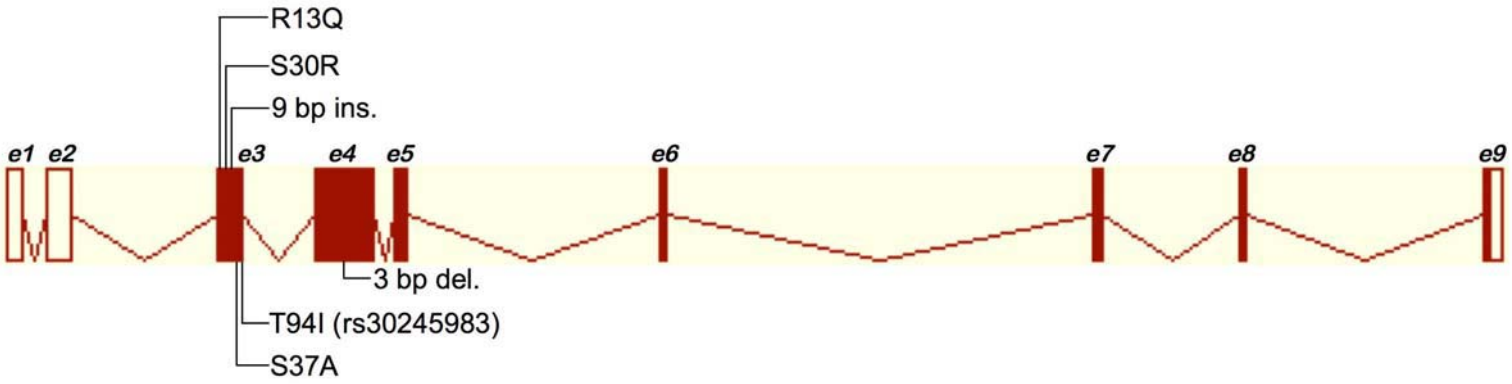
Supplementary Figure S4 Lung tumor multiplicities of inbred mice with different genotypes at *Pas1* and *Par2* loci



Supplementary Figure S5 Expression of *Las2*

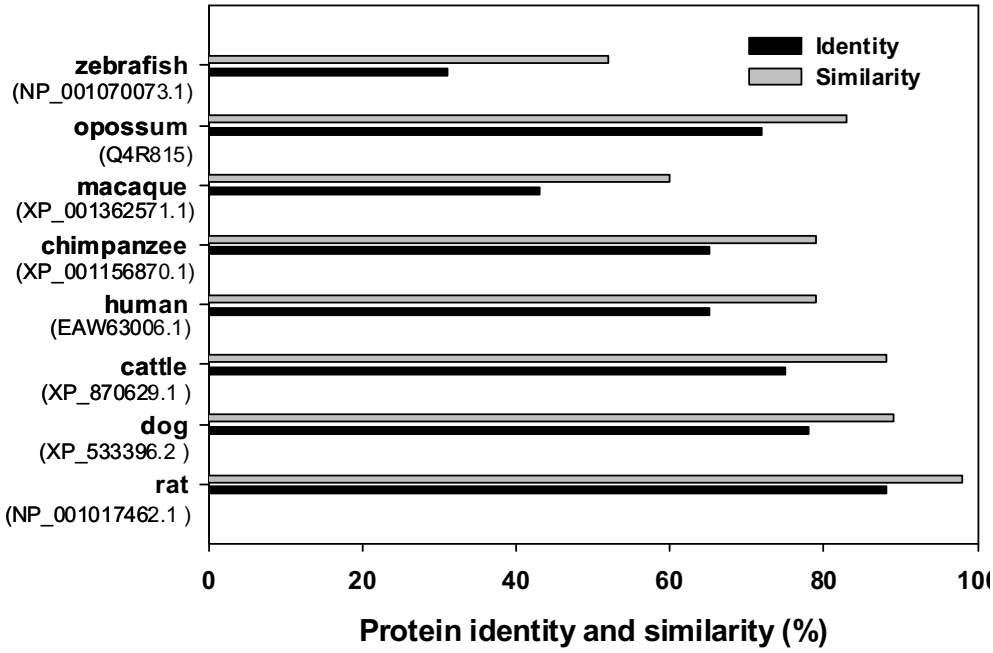


Supplementary Figure S6 Genomic organization of 4930503L19RIK/*Las2* and location of non-synonymous variants



Supplementary Figure S7 Alignment of *Las2* protein sequence of multiple species

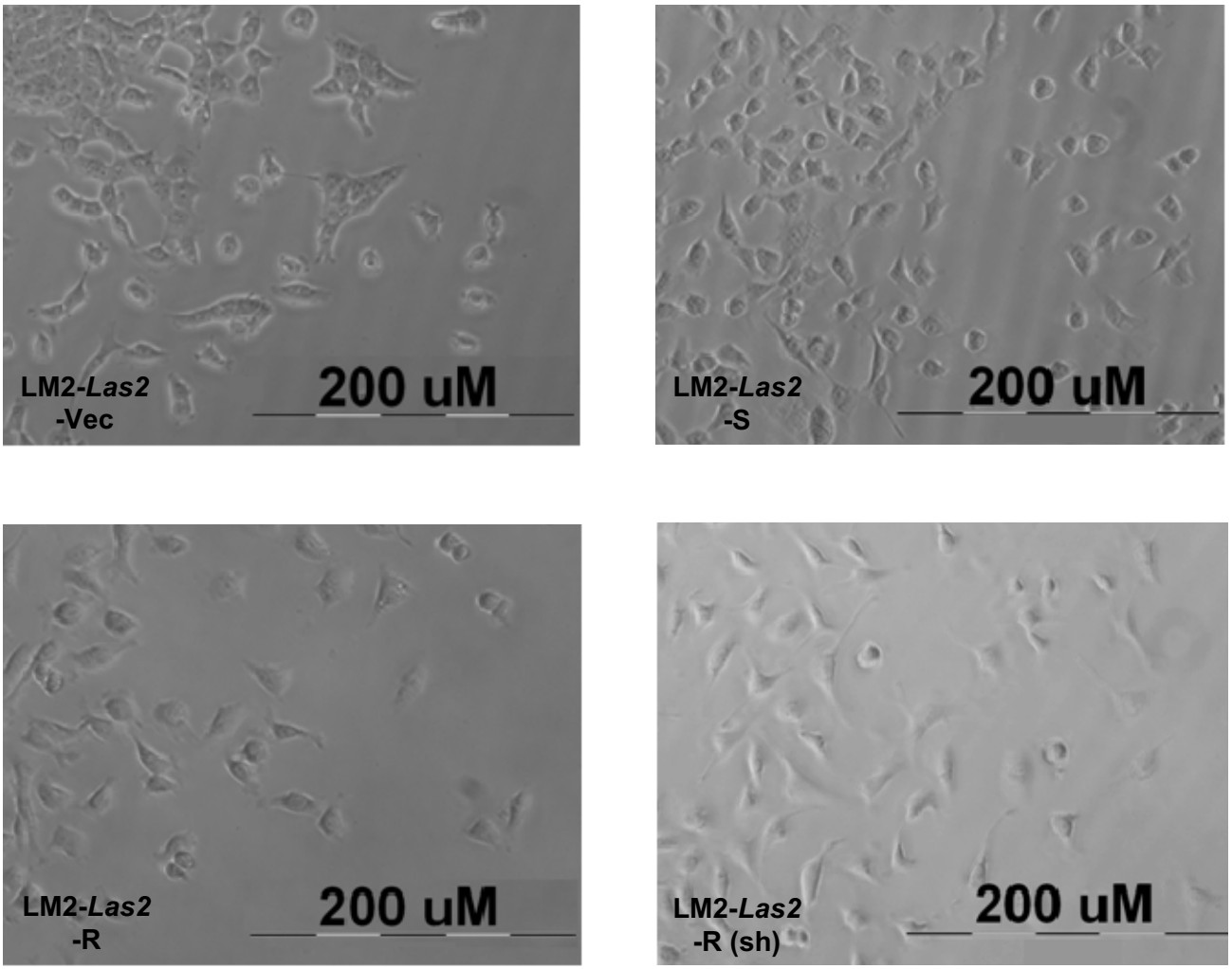
(a)



(b)


9-bp insertion
S37A
Mouse
Rat
Dog
Cattle
Human
Chimpanzee
Macaque
Opossum
Zebrafish

Supplementary Figure S8 Light microscopy of LM2 cells



Supplementary Figure S9 ***LAS2*** **loss in human lung cancer**