Linkage Disequilibrium and Physical Mapping of *Pas1* in Mice

Giacomo Manenti, Amanda Stafford, Laura De Gregorio, et al.

Genome Res. 1999 9: 639-646
Access the most recent version at doi:10.1101/gr.9.7.639

Letter

Linkage Disequilibrium and Physical Mapping of *Pas1* in Mice

Giacomo Manenti,[1] Amanda Stafford,[2] Laura De Gregorio,[1] Manuela Gariboldi,[1] F. Stefania Falvella,[1] Philip Avner,[2] and Tommaso A. Dragani[1,3]

[1]*Department of Experimental Oncology, Istituto Nazionale Tumori, Milan, Italy;* [2]*Unite de Genetique Moleculaire Murine, Unite de Recherche Associe Centre National de la Recherche Scientifique (URA CNRS) 1968, Institut Pasteur, Paris, France*

By using linkage disequilibrium (LD) analysis in 21 strains of known susceptibility to lung cancer and by assembling a YAC contig, we mapped to a ~1.5-Mb region on distal mouse chromosome 6 the *Pas1* locus, the major determinant of lung cancer predisposition in mice. Our results, on the basis of haplotype and phenetic analysis, suggest that the *Pas1*^{s} susceptibility allele is shared by several mouse-inbred strains of independent origin, which show either high or intermediate predisposition to lung tumorigenesis. Therefore, the *Pas1*^{s} allele is probably derived from an ancestral mouse rather than from independent mutations of the same gene. We showed the feasibility of LD in common inbred strains for the fine mapping of disease loci, and provided the biological basis and the reagents for the cloning of the *Pas1* gene.

The A/J mouse strain is highly susceptible to lung tumorigenesis and we have previously mapped the *Pulmonary adenoma susceptibility 1* (*Pas1*) locus affecting inherited predisposition to lung tumorigenesis in this strain, to the distal region of chromosome 6 (Gariboldi et al. 1993). Subsequently, independent studies with the A/J strain have confirmed the major role of *Pas1* in mouse lung tumorigenesis and have also supported the fact that the quantitative trait locus (QTL) peak for *Pas1* is localized around *Kras2* (Devereux et al. 1994; Festing et al. 1994; Manenti et al. 1995). As in many QTL studies, the linked region is too large (>10 cM) for undertaking positional cloning of the causative gene. Approaches other than genetic linkage analysis should be devised to narrow the region.

Several strains of independent origin show high or intermediate susceptibility to lung tumorigenesis. Among these strains, we have obtained evidence recently that, in addition to the A/J strain, the SWR/J and the BALB/c strains also carry the *Pas1*^{s} susceptibility allele, mapping close to the *Kras2* locus (Manenti et al. 1997b). Strain polymorphisms at the *Kras2* gene correlate with susceptibility to lung tumorigenesis (Malkinson and You 1994). On the basis of these observations, we hypothesize that the *Pas1*^{s} allele originated from a single founder and have designed a linkage disequilibrium (LD) study in mouse strains to test this hypothesis and to eventually narrow the candidate region for *Pas1*.

LD may be defined as the nonrandom association of marker alleles, usually mapping within a short chromosomal region, with a phenotype. Genetic linkage analysis is based only on recombination events that occur in a specific cross. In contrast, LD patterns rely on recombinations that have occurred over generations starting from the origin of the mutated allele in the founder mouse to the fixation of the mutation at homozygosity during inbreeding. LD analysis may therefore provide access to the equivalent of millions of meioses. LD has been successfully used in humans for the precise location of disease loci and has proved to be an important tool for the positional cloning of several disease genes (de la Chapelle 1993; Jorde 1995). The power of LD analysis for the precise mapping of a disease locus is especially clear in isolated populations, in which the disease-causing mutation originated from a single founder (Jorde 1995).

In experimental systems for the analysis of polygenic inheritance, disease loci have been mapped by genetic linkage studies typically carried out with a sample size of <500 meioses and haplotype analysis used to verify the candidacy of genes identified (Malo et al. 1994; MacPhee et al. 1995). To our knowledge, the potential of LD analysis for the fine mapping of disease genes has not, however, been tested in animals. If our hypothesis of a single founder of *Pas1*^{s} is correct, LD could be applied for the fine mapping of *Pas1*. Herein, we report that several markers located in the telomeric region of chromosome 6 exhibit significant LD with the genetic predisposition to lung cancer development. By combining LD and physical mapping, we shortened the *Pas1* region to a ~1.5-Mb interval, between *D6Mit57* and *D6Mit304* markers. These results should make feasible the positional cloning of *Pas1*.

RESULTS

The genetic markers used are listed in Table 1. During our analysis, we observed several discrepancies be-

[3]Corresponding author.
E-MAIL dragani@istitutotumori.mi.it; FAX 39-02-2390642.

Table 1. Genetic Markers Used for the Composite Linkage Disequilibrium and Physical Mapping of *Pas1*, and YAC Contig Assembling

Marker	cM	PCR primers [a]	Polymorphism [b]	-log *P* [c]
D6Mit1	3	MIT	Length	0.58
D6Mit8	35	MIT	Length	0.61
D6Mit52	62.5	MIT	Length	0.00
D6Mit374	62.5	MIT	Length	0.73
D6Mit198	67	MIT	Length	1.71
D6Mit59	67	MIT	Length	0.26
D6Mit113	70	MIT	Length	0.33
D6Mit114	71	MIT	Length	0.82
D6Mit57	71.2	MIT	Length	0.48
Kras2_4B1067 [e]	71.2	TTGAGGAGACCAGAGTTG GGACTGTTACATGGTGAGA	A/G	3.29
Kras2_4B1155 [e]	71.2	TTGAGGAGACCAGAGTTG GGACTGTTACATGGTGAGA	G/C	3.29
Kras2_4B513 [e]	71.2	TGTCCATCTACTCATCCAA GAGGGAGTGACCAAGATTA	A/C	3.29
Kras2_4B561 [e]	71.2	TGTCCATCTACTCATCCAA GAGGGAGTGACCAAGATTA	A/G	3.29
Kras2_4B629 [e]	71.2	TGTCCATCTACTCATCCAA GAGGGAGTGACCAAGATTA	C/G	3.29
Kras2_4Bins [e]	71.2	TTGAGGAGACCAGAGTTG GGACTGTTACATGGTGAGA	Length	3.29
Kras2_Alu	71.2	GAAACCTGTCTCTTGGATA ACCATAGGCACATCTTCAG	RFLP	2.65
Kras2_Sca	71.2	GAAACCTGTCTCTTGGATA ACCATAGGCACATCTTCAG	RFLP	2.65
Kras2_296 [d]	71.2	GGGTGTTAGGGAACCATA CAGGGCTGCATAGTAAGA	A/C	2.12
Kras2_288 [d]	71.2	GGGTGTTAGGGAACCATA CAGGGCTGCATAGTAAGA	C/G	3.29
Kras2_37	71.2	GGGTGTTAGGGAACCATA CAGGGCTGCATAGTAAGA	Length	3.29
Kras2_494	71.2	GGGTGTTAGGGAACCATA CAGGGCTGCATAGTAAGA	C/T	4.16
D6Mit14	71.3	MIT	Length	1.67
D6Mit303	71.4	MIT	NI [f]	
Y95m1f5	71.5	GGATGTGTGGTAATCGGAAGA CAAAGACACTTGCCACTAAGC	T/C	2.56

YACs	Size(kb)
ICRF-62M8B7	630
MIT-88M6B4	800/750
MIT-76M8B4	460
ICRF-66M4G6	580
MIT-95M1F5	550/350
MIT-94M3C6	480
ICRF-70M6H6	440
MIT-83M2C2	680
MIT-82M6B7	680
MITIII-168E9	750
MITIII-151A11	>1500
MITIII-392E7	>1500
MITIII-312F6	900
MIT-95M5E3	800
MITIII-287B9	800
MIT-94M2H10	740
MITIII-170E4	>1500
MITIII-384B12	>1500
MITIII-167B4	>1500
MITIII-277D5	840
ICRF-60M7E10	700/840
ICRF-67M1B5	900
MIT-90M1B9	950/700
MIT-100M6C5	350
MIT-78M2E5	470
ICRF-71M1E4	600
MIT-97M3B7	780
MIT-78M6G9	500
MIT-100M8H10	1050
MIT-99M6F2	780
ICRF-55M5G2	550
ICRF-59M3G2	1400/440
ICRF-71M5F1	550
MIT-97M1A3	700
MIT-90M8E2	750/440
MIT-78M6A9	350
MIT-98M6G10	580/350
MIT-98M2D1	900/440
ICRF-65M6H1	1400
MITIII-273A8	1500/800/580
MITIII-384F2	1500/440/380
MITIII-382B3	1200
ICRF-67M1A7	550
MIT-95M4D4	650
MIT-91M8F11	580/440/350
MIT-88M6A6	500/350/230

Table 1. (*Continued*)

Y95m5e3	71.5	GGCTCCTGGTGGTTTGTCAA TTGCAACTCTCTGCTGGACT	C/T	1.26
Krag	71.6	AATAACCTTGAAAGCTGTCC GCTGAACTTGCACTGTACC	Length (3'-UTR)	0.69
D6Mit200	71.6	MIT	Length	2.20
D6Mit137	71.6	MIT	Length	2.34
Y94m2h10	71.7	AACTGGTGAGAACACTTCTCC CAGAAGACCCCAGGTCAATTC	NI[f]	
Y83m2c2	71.7	CATTGCTGGCATTGCTCACTT AATGGGCTGGTCTGAACAAAC	NI[f]	
D6Mit371	71.7	MIT	Length	1.38
D6Mit294	71.8	MIT	Length	1.15
D6Mit295	71.8	MIT	Length	2.00
D6Mit15	71.9	MIT	Length	0.98
D6Mit201	72	MIT	Length	1.85
Y78m6g9	72.1	CATCTACAACAACCAGTAAGG GGGATCAGACTCAGGTTATCA	A/T	1.16
D6Mit373	72.2	MIT	Length	2.91
D6Mit26	72.3	MIT	Length	3.29
Y97m1a3	72.4	TTTCCATGCATCTCACAGTTG TGTATGGGTGAAGAATCCGAT	NI[f]	
Y88m6a6	72.5	GATCAACATCTGTCGGGAAGA TCTGGAGGTTGTGGTACCCA	C/A	2.50
Y91m8f11	72.6	GCATTTCTGGATGACGGTTT AAGGTGAATGCTGATGTTTGG	NI[f]	
D6Mit304	72.6	MIT	Length	2.56
D6Mit372	73	MIT	Length	1.22
D6Int1	74	MIT	Length	0.66

[a]MIT primers were obtained from Research Genetics, nucleotide sequences are given as 5′ to 3′ directions.
[b]Type of polymorphism detected in the relative PCR fragment, including nucleotide variations for single nucleotide polymorphisms.
[c]Minus logarithms of *P* values obtained by Fisher's exact test (see Methods for further details).
[d]Direct oligonucleotide sequence of the NGP/N, STS/A, and *M. spretus* alleles (data from Chen et al. 1994 for the other alleles).
[e]Polymorphism described in Matzinger et al. (1997).
[f](NI) Not informative; polymorphisms among strains not detected.

tween the strain allele size of SSLPs reported by MIT and the PCR fragment length we obtained from our own analysis. Fifteen of 22 SSLPs (68%) showed small differences in the allele size of at least one of the strains analyzed (data not shown). In the case of *D6Mit26*, for example, we found that the BALB/c allele is 196 bp long instead of null, and the size of the *Mus spretus* allele is 202 bp instead of the reported size of 200 bp. When these errors for *D6Mit26* were corrected, all of the investigated strains fell into two classes that correlated well with susceptibility and resistance to lung tumor development, as indicated by the statistically significant association ($-\log P = 3.29$; $-\log P$ being defined as the negative value of the logarithm of the P value of the test statistic). Both the highly susceptible and the intermediate susceptible strains showed a common allele of 196 bp, whereas the resistant strains shared the same 202-bp-long allele (Table 2).

A YAC contig was constructed across the candidate region to provide reagents for the eventual cloning of the *Pas1* gene. The *D6Mit* markers that localized in this interval (Dietrich et al. 1996) were screened across the mouse YAC libraries. Continuity of coverage required the addition of new STS markers obtained by YAC end-clone isolation (Table 1). New markers lying within the contig were identified, including a 36-bp repeat polymorphism at *Krag* 3′-UTR, corresponding to nucleotide 4011–4046 in the GenBank sequence (accession no. MM02487). The strains carried either one or three repeats (data not shown). Strain polymorphisms detected in the YAC end-clone sequences yielded additional markers for LD analysis. A minimum set of five YACs of average size 840 kb is required to cover the *D6Mit57–D6Mit304* interval. The size of this contig can be estimated at 1.5–2 Mb if an average overlap of 50% between YAC clones is assumed. The order of markers across the YAC contig was in agreement with existing genetic maps. The one exception was *D6Mit15*, placed in the most telomeric group of markers on the MIT genetic map (Dietrich et al. 1996) but more proximally on the YAC contig as well as on EUCIB and MGD genetic maps (Mouse Genome Database 1998; Rhodes et al. 1998). A combined genetic and physical map of the region was assembled, with *D6Mit57* (71 cM) and *D6Mit304* (73 cM) markers as loci to anchor the genetic and physical maps. With the aim of defining genetic distance between markers in the contig, we attributed a 0.1-cM distance interval to markers separated by partially overlapping YAC clones (Table 1; Fig. 1).

A significant LD was found for markers extending across the whole YAC contig. Contingency tables with strains grouped into susceptible, intermediate, and resistance phenotypes resulted in the same LD pattern of two levels of tables (susceptible/intermediate and resis-

Table 2. Mouse Strains and Haplotypes at Genetic Markers Mapping in the *Pas1* Region

		Haplotype at selected markers[b]					
Strain	Susceptibility[a]	*Kras2_494*	*Kras2_37*	*Kras2_4B561* *Kras2_4B1067*	*Y95m1f5*	*D6Mit26*	Putative *Pas1* allele[c]
AKR/J	R	T	2	G	C	202	r
M. spretus	R	T	1	A	C	202	r
C3H/HeJ	R	T	2	G	C	202	r
C57L/J	R	T	2	G	T	202	r
C57BL/6J	R	T	2	G	T	202	r
SJL/J	R	T	2	G	C	202	r
DBA/2J	R	T	2	G	C	202	r
SM/J	R	C	1	A	T	196	s
CBA/J	I	C	1	A	T	196	s
LP/J	I	C	1	A	T	196	s
PL/J	I	C	1	A	T	202	?
RF/J	I	C	1	A	T	196	s
129/SvJ	I	C	1	A	T	196	s
ST/bJ	I	C	1	A	T	196	s
BALB/cJ	I	C	1	A	T	196	s
MA/MyJ	I	C	1	A	T	196	s
STS/A	S	C	1	A	T	196	s
O20/A	S	C	1	A	T	196	s
SWR/J	S	C	1	A	T	196	s
A/J	S	C	1	A	T	196	s
NGP/N	S	C	1	A	T	196	s

[a](R) Resistant; (I) Intermediate; (S) susceptible (see text for further details).
[b]Number of repeats at *Kras2_37* marker, allele size for *D6Mit26,* and single nucletide polymorphisms for the other markers (see Table 2 and text for further details).
[c]Based on haplotype sharing of marker loci in LD with phenotype.

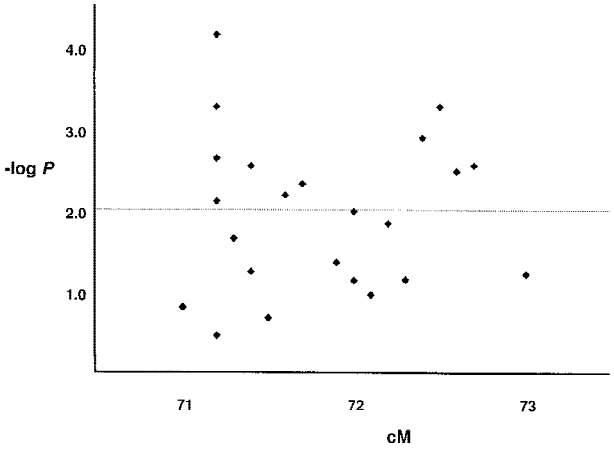

Figure 1 Composite genetic and physical map of the telomeric region of mouse chromosome 6 for LD mapping of the *Pas1* locus. Statistical significance of LD between marker polymorphisms and phenotype was estimated by Fisher's exact test and expressed as negative logarithm of *P* values ($-\log P$). The dotted line represents the $P = 0.01$ ($-\log P = 2$) cutoff significance value. Each point represents either a single genetic marker or multiple markers in the same position with the same $-\log P$ value (Table 1).

tant), although the statistical significance values were lower because of the increased degrees of freedom (data not shown). Analysis of variance by lung tumor multiplicity data gave the same LD pattern with significance values similar to those herein reported (data not shown). A total of 12 markers were typed in and around the *Kras2* locus, a short genomic region extending over <40 kb. In this region we found markers showing borderline significant LD (e.g. *Kras2_296*) as well as markers with a highly significant LD, such as the *Kras2_494* polymorphism that showed the highest LD ($-\log P = 4.16$) (Table 1). The *D6Mit26* marker, located distally in the contig, showed the second highest LD value ($-\log P = 3.29$) (Table 1). Several other markers, located between the proximal and distal ends of the contig, were also associated with a significant LD ($-\log P > 2$) (Table 1; Fig. 1). We tested 10 additional genetic markers in chromosomal regions other than that of the distal part of chromosome 6. None of these markers showed significant LD with lung tumor susceptibility (data not shown).

Haplotype analysis with markers showing the most significant LD indicated that strains 129/SvJ, A/J, BALB/cJ, CBA/J, LP/J, MA/MyJ, NGP/J, O20/A, RF/J, ST/bJ, STS/A, SM/J, and SWR/J carried the same haplotype (Table 2). Most of these strains show a high or intermediate predisposition to lung tumorigenesis (Malkinson 1989). Strain PL/J shared an identical haplotype to the above strains, except for the *D6Mit26* marker (Table 2). On the other hand, the AKR/J, C3H/HeJ, C57BL/6J, C57L/J, DBA/2J, *M. spretus*, and SJL/J strains, which are resistant to lung tumorigenesis, showed a variable haplotype. Parsimony analysis of discrete state data, as well as the results of distance matrix programs, essentially produced the same phenetic tree (Fig. 2; data not shown). Two main branches, in addition to a branch containing only the *M. spretus* strain, were clearly separated. One of the main branches contained the strains with the high or intermediate predisposition to lung cancer, whereas the other branch contained the resistant strains.

DISCUSSION

We performed LD analysis in mouse strains of known susceptibility and resistance to lung tumorigenesis, using marker alleles around the *Pas1* locus. The use of LD analysis is based on the assumption that recombination is the major force determining the presence or absence of LD. Because one would expect to find genomic regions identical by descent in inbred strains, LD analysis in strains with a known phenotype may allow testing of whether the same locus is responsible for that given phenotype. LD is maintained only along

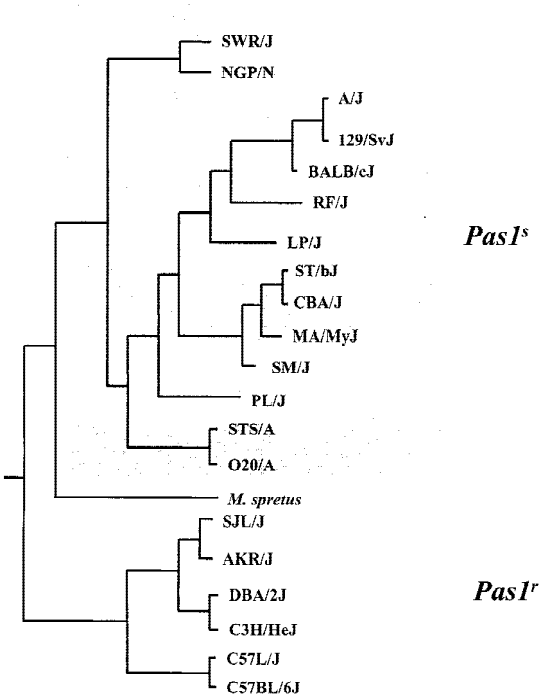

Figure 2 Phenetic tree of a 21-mouse inbred strain of known susceptibility to lung carcinogenesis, obtained by the Fitch–Margoliash distance matrix method, with marker polymorphisms mapped in the YAC contig containing the *Pas1* gene. The tree represents the estimated relationships among the strains for that short chromosomal region (~1.5Mb), and it therefore provides an estimation of the phylogenetic origin of the *Pas1*^{s} allele. The gray box contains the putative *Pas1*^{s} mice, which include the A/J, SWR/J, and BALB/c strains proved by genetic linkage to carry the *Pas1*^{s} allele. Also, the putative *Pas1*^{r} strains include the C57BL/6J, C3H/HeJ, and *M. spretus* mice, established by genetic linkage studies as *Pas1*^{r} strains.

short chromosomal regions, and it could therefore represent a new approach for the fine mapping of polygenic traits in mice.

However, several phenomena, including homoplasy and mutations, may have an important role in the maintenance or loss of LD. These confounding phenomena may explain why we found, within the same region, marker alleles showing a highly significant LD as well as marker alleles that were not in LD with the phenotypic trait (Fig. 1). As this is expected in LD studies, it is necessary to type many markers in LD studies to avoid false-positive and false-negative results. We suggest that the significance of LD analysis can only be assessed after comprehensive analysis of the whole region under investigation. We should also be aware that the prerequisite to limit LD mapping to those strains with a known phenotype (21 in our case) can weaken the confidence of any positive association, with some danger of type-1 error. However, combining LD mapping with conventional cross-linkage data (as we partially did, as 6 of 21 strains carry a known *Pas1* allele), we could restrict the mapping of disease genes to a 1- to 2-cM region, suitable for positional cloning.

Our results indicate that inbred strains resistant to lung tumorigenesis show variations in their marker alleles. In contrast, all inbred strains susceptible or partially susceptible to lung tumorigenesis belonged to a single haplotype group (Table 2). Phenetic analysis confirmed the haplotype analysis by separating two branches of putative *Pas1*^{s} and *Pas1*^{r} strains (Fig. 2). Phenetic analysis formed new phylogenetic branches that were unrelated to the known historical origin of these strains (Festing 1993) and different from phylogenetic trees published previously (Fitch and Atchley 1985; Atchley and Fitch 1991) (Fig. 2). For example, the C3H/He (resistant) and the CBA/J (intermediate) strains that derived from a single cross of a Bagg female and a DBA male (Festing 1993) are clearly separated in our analysis, whereas they are phylogenetically closely related (Atchley and Fitch 1991). The reason for this discrepancy is due to the selection for the genetic markers that lie within the short chromosomal region under examination. In this context, strain clustering suggests an origin of this small region from a common ancestor. Whereas homoplasy at microsatellite loci might also produce such strain clustering (Garza and Freimer 1996; Orti et al. 1997), the results we obtained with microsatellite-length polymorphisms were highly concordant with those derived from the analysis of single-nucleotide polymorphisms. Putative *Pas1*^{s} and *Pas1*^{r} groups of mice are separable (Fig. 2).

As was reported previously, strains A/J, SWR/J, and BALB/c carry the same *Pas1*^{s} allele (Manenti et al. 1997b); therefore, we can now infer that all strains sharing the same haplotype for these markers also carry the *Pas1*^{s} allele (Table 2). A discrepancy is represented by the SM/J strain that carries the same haplotype as the putative *Pas1*^{s} animals, even though it is resistant to lung tumorigenesis (Malkinson 1989). However, the SM/J strain carries the *Par1* and *Par3* loci (Abujiang et al. 1997), which may inhibit phenotypic expression of the putative *Pas1*^{s} allele.

Putative *Pas1*^{s} strains are of varied geographical derivation (e.g., A/J is United States derived and SWR/J strain is Swiss derived) and have originated at different times (e.g., A/J originated >70 years ago, whereas the NGP/N is of a much more recent derivation; Festing 1993). The finding of an identical haplotype in such strains would seem to indicate that the *Pas1*^{s} susceptibility allele probably derived from an ancestral mouse, rather than originating in different strains as independent mutations. The presence of the *Pas1*^{s} allele in several strains would indicate a relatively high frequency of this allele in the genus *Mus*. Genetic linkage experiments with crosses involving individual putative *Pas1*^{s} strains with a known *Pas1*^{r} strain (e.g., C3H/HeJ, C57BL/6J, *M. spretus*) (Gariboldi et al. 1993; Devereux et al. 1994; Festing et al. 1994; Manenti et al. 1995) may allow us to verify whether this prediction is correct. However, the phenotypic expression of *Pas1*^{s} allele is reduced in most of these strains by the lung cancer modifier loci (*Par* loci) as we reported for the BALB/c strain (Dragani and Manenti 1997; Manenti et al. 1997c). This can explain why some *Pas1*^{s} strains (e.g., A/J, SWR/J, NGP/N) are highly susceptible, whereas most of the other *Pas1*^{s} strains show an intermediate susceptibility to lung tumorigenesis. The *Par* loci might therefore have an essential role in phenotypic expression of the inherited predisposition to lung cancer. Their identification and cloning should be further influenced by our results. The *Par* genes may represent interesting new targets for prospective lung cancer chemoprevention and therapeutic strategies (Dragani and Manenti 1997).

Our results provide information for the identification of the *Pas1* gene. Both the fine physical mapping with YAC clones of the region containing the *Pas1* gene and the identification of putative *Pas1*^{s} strains with haplotype analysis are relevant elements for subsequent strain comparison of germ-line variations of candidate *Pas1* genes. The *Pas1* gene may be located in the central region of the contig that shows significant LD. However, because the markers showing the best LD (*Kras2_494* and *D6Mit26*) are positioned at the ends of the contig, we cannot rule out that two *Pas1* genes might exist, located close to each of these two markers. Our approach cannot distinguish whether strains have both trait loci rather than one. Because the *Pas1*^{s} allele is common in mouse strains, it may be predicted that it is also common in other species, including humans. Results that we have obtained in humans are in agreement with our present ones, because we have found a

significant association of polymorphisms at KRAS2 and PTHLH loci (ends of the homologous human contig) with risk and prognosis for lung adenocarcinoma (Manenti et al. 1997a).

METHODS

Mouse DNAs and Genetic Markers

Genomic DNA samples were obtained from Jackson Laboratories (Bar Harbor, ME) (129/SvJ, A/J, AKR/J, BALB/cJ, C3H/HeJ, C57BL/6J, C57L/J, CBA/J, DBA/2J, LP/J, MA/MyJ, *M. spretus*, PL/J, RF/J, SJL/J, SM/J, ST/bJ, SWR/J) or were kindly provided by Dr. I. Nakashima (Nagoya University, Japan; O20/A), Dr. M. Nishimura (Hamamatsu University, Japan; STS/A), and Dr. M. Mandel (National Cancer Institute, Bethesda, MD; NGP/N).

PCR primers for single nucleotide polymorphisms (SNPs) and for simple sequence length polymorphisms (SSLPs) are as reported in Table 1. For SNP markers, aliquots of PCR reactions were loaded on an agarose gel to check the size and amount of amplified fragments. The remaining PCR mix was denatured in 0.4 M NaOH/25 mM EDTA at room temperature and spotted onto nylon membrane. Allele-specific 15-mer oligonucleotides empassing the SNP were 5′- end-labeled with [γ-^{32}P]dATP (3000 Ci/mM) (Amersham, Branchburg, NY) and T4 polynucleotide kinase (New England Biolabs, Beverly, MA). Allele-specific oligonucleotide (ASO) hybridizations were performed in tetramethylammonium chloride (TMAC) as reported (Manenti et al. 1994). SSLP markers were typed with PCR primers obtained from Research Genetics (Huntsville, AL) and 25 radioactive PCR cycles (55°C annealing temperature); results were scored on 6% denaturing polyacrylamide gels. As markers to measure SSLPs length, we loaded in a single well the four ddNTPs-stopped samples of a radioactive sequencing reaction. The resulting 1-bp ladder was used to assign a precise length to the SSLP allele of each strain.

YAC Library Screening

The ICRF and MITII mouse YAC libraries (Larin et al. 1991; Kusumi et al. 1993) were screened by PCR. This involved screening a total of 49 super pools followed by a further 24-reaction per positive super pools to identify the clone address. PCR reactions were carried out on a custom built Waffle Iron PCR machine with 35 cycles of 30 sec at 94°C/30 sec at 55°C/30 sec at 72°C. PCR products were visualized on 2% agarose gels. YAC clones from the MITIII library (Haldi et al. 1996) were obtained from Research Genetics on the basis of the coordinates provided on MIT's public access database. The individual YAC clones identified by these two strategies were checked for STS content by PCR screening with all STSs in the region to exclude false-positive and false-negative results that may occur in high-throughput YAC library screening. YAC clones were grown from frozen library stocks on AHC plates as unpurified well aliquots in AHC media. Agarose-embedded yeast DNA was loaded onto a 1% agarose gel and electrophoresed in 0.5× TBE at 14°C on a Bio-Rad CHEF DRII-pulsed field gel apparatus with the following program: ramped switch time 50–110 sec over 24 hr at 200 V. The DNA was transferred to a nylon membrane (Hybond N$^+$, Amersham) by Southern blotting and hybridized with a radioactive probe prepared from 100 ng of mouse *cot1* DNA (GIBCO-BRL) labeled with Amersham's Megaprime DNA Labeling System with [α- ^{32}P]dCTP (ICN). The membrane was washed to a stringency of 0.1× SSC at 65°C for 20 min and exposed for autoradiography.

YAC End Clone Isolation and Polymorphism Detection

YAC end clones were isolated by Vectorette, following the protocol of Riley et al. (1990) with YAC DNA embedded in blocks and digested with the restriction enzymes *Alu*I and *Rsa*I. The Vectorette products were gel purified and sequenced by cycle sequencing with fluorescence-labeled dideoxynucleotides with ABI's dideoxy *Taq* F.S. kit and electrophoresed on an ABI 377 automated sequencer. This sequence was used to derive PCR primers that were used to amplify YAC end clone sequences in four mouse strains, A/J, SWR/J, C57BL6/J and C3H/HeJ. The amplified fragments were cloned into the pCR2.1 vector (Invitrogen, San Diego, CA) and at least two clones for each fragment were sequenced in both orientations on an ABI 377 sequencer. Sequences obtained from different strains were compared. When base differences were found among the four strains, ASOs were synthesized and hybridized to all mouse strains under investigation (Table 1).

Linkage Disequilibrium Analysis and Phylogenetic Analysis

As susceptibility to lung tumorigenesis, mouse strains were placed into two phenotype groups: The first group contained resistant strains and the second group contained susceptible and intermediate strains, according to published classification (Table 2) (van der Valk 1981; Malkinson 1989; Manenti et al. 1995). Linkage disequilibrium between strain segregation of marker alleles and lung tumor susceptibility was evaluated by Fisher's exact test. *P* values were transformed in their negative logarithms, and a significant LD was considered if the $-\log P > 2$ ($P < 0.01$).

Phenetic analysis of genetic elements surrounding the *Pas1* locus was used to estimate phylogeny for this locus. In the region spanning the *D6Mit113* to *D6Int1* markers, we characterized 33 genetic markers (16 SSLPs, 12 SNPs, 2 RFLPs, and 3 small deletion/insertion) in 21 inbred strains with known susceptibility to lung tumor development. For the whole data set, a triangular matrix with the percentages of genotype differences was calculated by pairwise comparison of the strains (Canzian 1997). Among the 33 markers, 19 showed a binary state in all the inbred strains analyzed. The data set of discrete characters was first boot strapped and then used to estimate phylogeny according to Wagner and Dollo parsimony methods with MIX and DOLLOP programs, respectively (Fitch and Margoliash 1967; Canzian 1997; Farris 1997), with the PHYLYP package.

ACKNOWLEDGMENTS

We thank Dr. Federico Canzian for his suggestions, comments, and criticism. This work was partially funded by grants from Finalized Project Consiglio Nazionale Ricerche "Applicazioni Cliniche Ricerca Oncologica" and from the Associazione and Fondazione Italiana Ricerca Cancro (AIRC and FIRC) of Italy. Amanda Stafford was in receipt of a Wellcome Prize Travelling Research Fellowship. P.A. and A.S. were supported by grants from the GREG and the Juvenile Diabetes Foundation International.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be

hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES

Abujiang, P., M. Nishimura, T. Kamoto, K. Ichioka, M. Stato, and H. Hiai. 1997. Genetic resistance to urethane induced pulmonary adenomas in SMXA RI strains. *Cancer Res.* **57:** 2904–2908.

Atchley, W.R. and W.M. Fitch, 1991. Gene trees and the origins of inbred strains of mice. *Science* **254:** 554–558.

Canzian, F. 1997. Phylogenetics of the laboratory rat Rattus norvegicus. *Genome Res.* **7:** 262–267.

Chen, B., L. Johanson, J.S. Wiest, M.W. Anderson, and M. You. 1994. The second intron of the K-*ras* gene contains regulatory elements associated with mouse lung tumor susceptibility. *Proc. Natl. Acad. Sci.* **91:** 1589– 1593.

de la Chapelle, A. 1993. Disease gene mapping in isolated human populations: The example of Finland. *J. Med. Genet.* **30:** 857–865.

Devereux, T.R., R.W. Wiseman, N. Kaplan, S. Garren, J.F. Foley, C.M. White, C. Anna, M.A. Watson, A. Patel, S. Jarchow, R.R. Maronpot, and M.W. Anderson. 1994. Assignment of a locus for mouse lung tumor susceptibility to proximal chromosome 19. *Mamm. Genome* **5:** 749–755.

Dietrich, W.F., J. Miller, R. Steen, M.A. Merchant, D. Damron-Boles, Z. Husain, R. Dredge, M.J. Daly, K.A. Ingalls, T.J. O'Connor, et al. 1996. A comprehensive genetic map of the mouse genome. *Nature* **380:** 149–152.

Dragani, T.A. and G. Manenti. 1997. *Mom1* leads the pack. *Nat. Genet.* **17:** 7–8.

Farris, J.S. 1977. Phylogenetic analysis under Dollo's Law. *Syst. Zool.* **26:** 77–88.

Festing, M.F.W. 1993. Origins and characteristics of inbred strains of mice. *Mouse Genome* **91:** 393–509.

Festing, M.F.W., A. Yang, and A.M. Malkinson. 1994. At least four genes and sex are associated with susceptibility to urethane-induced pulmonary adenomas in mice. *Genet. Res.* **64:** 99–106.

Fitch, W.M. and E. Margoliash. 1967. Construction of phylogenetic trees. *Science* **155:** 279–284.

Fitch, W.M. and W.R. Atchley. 1985. Evolution in inbred strains of mice appears rapid. *Science* **228:** 1169–1175.

Gariboldi, M., G. Manenti, F. Canzian, F.S. Falvella, M.T. Radice, M.A. Pierotti, G. Della Porta, G. Binelli, and T.A. Dragani. 1993. A major susceptibility locus to murine lung carcinogenesis maps on chromosome 6. *Nat. Genet.* **3:** 132–136.

Garza, J.C. and N.B. Freimer. 1996. Homoplasy for size at microsatellite loci in humans and chimpanzees. *Genome Res.* **6:** 211–217.

Haldi, M.L., C. Strickland, P. Lim, V. VanBerkel, X. Chen, D. Noya, J.R. Korenberg, Z. Husain, J. Miller, and E.S. Lander. 1996. A comprehensive large- insert yeast artificial chromosome library for physical mapping of the mouse genome. *Mamm. Genome* **7:** 767–769.

Jorde, L.B. 1995. Linkage disequilibrium as a gene-mapping tool. *Am. J. Hum. Genet.* **56:** 11–14.

Kusumi, K., J.S. Smith, J.A. Segre, D.S. Koos, and E.S. Lander. 1993. Construction of a large-insert yeast artificial chromosome library of the mouse genome. *Mamm. Genome* **4:** 391–392.

Larin, Z., A.P. Monaco, and H. Lehrach. 1991. Yeast artificial chromosome libraries containing large inserts from mouse and human DNA. *Proc. Natl. Acad. Sci.* **88:** 4123–4127.

MacPhee, M., K.P. Chepenik, R.A. Liddell, K.K. Nelson, L.D. Siracusa, and A.M. Buchberg. 1995. The secretory phospholipase A2 gene is a candidate for the *Mom1* locus, a major modifier of Apc^{Min}-induced intestinal neoplasia. *Cell* **81:** 957–966.

Malkinson, A.M. 1989. The genetic basis of susceptibility to lung tumors in mice. *Toxicology* **54:** 241–271.

Malkinson, A.M. and M. You. 1994. The intronic structure of cancer-related genes regulates susceptibility to cancer. *Mol. Carcinog.* **10:** 61–65.

Malo, D., K. Vogan, S. Vidal, J. Hu, M. Cellier, E. Schurr, A. Fuks, N. Bumstead, K. Morgan, and P. Gros. 1994. Haplotype mapping and sequence analysis of the mouse Nramp gene predict susceptibility to infection with intracellular parasites. *Genomics* **23:** 51–61.

Manenti, G., S. Pilotti, F.C. Re, G. Della Porta, and M.A. Pierotti. 1994. Selective activation of *ras* oncogenes in follicular and undifferentiated thyroid carcinomas. *Eur. J. Cancer* **30A:** 987–993.

Manenti, G., F.S. Falvella, M. Gariboldi, T.A. Dragani, and M.A. Pierotti. 1995. Different susceptibility to lung tumorigenesis in mice with an identical *Kras2* intron 2. *Genomics* **29:** 438–444.

Manenti, G., L. De Gregorio, S. Pilotti, F.S. Falvella, M. Incarbone, F. Ravagnani, M.A. Pierotti, and T.A. Dragani, 1997a. Association of chromosome 12p genetic polymorphisms with lung adenocarcinoma risk and prognosis. *Carcinogenesis* **18:** 1917–1920.

Manenti, G., M. Gariboldi, A. Fiorino, A.I. Zedda, M.A. Pierotti, and T.A. Dragani. 1997b. *Pas1* is a common lung cancer susceptibility locus in three mouse strains. *Mamm. Genome* **8:** 801–809.

Manenti, G., M. Gariboldi, A. Fiorino, N. Zanesi, M.A. Pierotti, and T.A. Dragani. 1997c. Genetic mapping of lung cancer modifier loci specifically affecting tumor initiation and progression. *Cancer Res.* **57:** 4164– 4166.

Matzinger, S.A., B. Chen, Y. Wang, K.A. Crist, G.D. Stoner, G.J. Kelloff, RA, Lubet, and M. You. 1997. Tissue-specific expression of the K-ras allele from the A/J parent in (A/J x TSG-p53) F1 mice. *Gene* **188:** 261–269.

Mouse Genome Database (MGD) 1998. *Mouse genome informatics*, The Jackson Laboratory, Bar Harbor, ME. (URL:http://www.informatics.jax.org/).

Orti, G., D.E. Pearse, and J.C. Avise. 1997. Phylogenetic assessment of length variation at a microsatellite locus. *Proc. Natl. Acad. Sci.* **94:** 10745–10749.

Rhodes, M., R. Straw, S. Fernando, A. Evans, T. Lacey, A. Dearlove, J. Greystrong, J. Walker, P. Watson, P. Weston et al. 1998. A high-resolution microsatellite map of the mouse genome. *Genome Res.* **8:** 531–542.

Riley, J., R. Butler, D. Ogilvie, R. Finniear, D. Jenner, S. Powell, R, Anand, J.C. Smith, and A.F. Markham. 1990. A novel, rapid method for the isolation of terminal sequences from yeast artificial chromosome (YAC) clones. *Nucleic Acids Res.* **18:** 2887–2890.

van der Valk, M.A. 1981. Survival, tumor incidence and gross pathology in 33 mouse strains. In *Mammary tumors in the mouse*. (ed. J. Hilgers and M. Sluyser), pp. 46–115. Elsevier, Amsterdam, The Netherlands.

Received July 29, 1998; accepted in revised form May 26, 1999.