Population-based Mapping of Pulmonary Adenoma Susceptibility 1 Locus

Tommaso A. Dragani, Setsuo Hirohashi, Takeo Juji, Kaname Kawajiri, Masahiro Kihara, Masako Ono-Kihara, Giacomo Manenti, Tomoko Nomoto, Haruhiko Sugimura, Keiichiro Genka, Jun Yokota, Takashi Takahashi, Tetsuya Mitsudomi, and Minako Nagao

National Cancer Center Research Institute, Tokyo, Japan [T. A. D., S. H., T. N., J. Y., M. N.]; Istituto Nazionale TumorI, 20133 Milan, Italy [T. A. D., G. M.]; Japanese Red Cross Central Blood Center, Tokyo, Japan [T. J.]; Saitama Cancer Center Research Institute, Saitama, Japan [K. K.]; Kanagawa Cancer Center Research Institute, Yokohama, Japan [M. K., M. O-K.]; Hamamatsu University of Medicine, Hamamatsu, Japan [H. S.]; National Okinawa Hospital, Okinawa, Japan [K. G.]; and Aichi Cancer Center Research Institute, Nagoya, Japan [T. T., T. M.]

Abstract

Pulmonary adenoma susceptibility 1 (Pas1), the major locus affecting inherited predisposition to lung tumor development in mice, maps near the Kras2 gene. We previously reported a significant association between a KRAS2/Run1 polymorphism and the risk and prognosis of lung adenocarcinoma (ADCA) in the Italian population. In the present case-control study, we examined 269 lung ADCA patients, 121 squamous cell lung carcinoma patients, and 632 healthy individuals (general population controls) in the Japanese population with genetic markers spanning ~1200 kb in the KRAS2 region. Allele-specific oligonucleotide hybridization revealed the same KRAS2/Run1 polymorphism associated with risk and prognosis as in Italian lung ADCA patients; the polymorphism was significantly associated with clinical stage (P < 0.001) and survival rate (log rank = 0.0014), confirming the mapping of PAS1 and pointing to the role of this locus in human lung cancer.

Introduction

Inbred mouse strains are available that show either high, intermediate, or low genetic susceptibility to lung cancer (1). Pulmonary adenoma susceptibility 1 (Pas1) plays a major role in inherited predisposition to lung tumor development in mice, as shown by genetic linkage studies carried out in different crosses and in different laboratories (reviewed in Ref. 2). Our recent genetic linkage and LD studies suggest that the same Pas1 susceptibility allele is carried by most mouse strains with high and intermediate genetic susceptibility to cancer, leading us to hypothesize an ancestral origin of this allele (3). We have fine-mapped the mouse Pas1 locus on a ~2-Mb region of chromosome 6 using a combination of physical mapping and LD analysis (3).

In an Italian population study, we demonstrated a significant association between genetic polymorphisms located in the human 12p12 chromosomal region, i.e., the region homologous to the Pas1 site in the mouse, and risk and prognosis of lung ADCA. However, population-based studies may be biased because of population admixture and stratification (4). Thus, we have designed a case-control study in the Japanese population to determine whether a similar association exists between genetic markers and risk and prognosis of lung cancer. In lung ADCA patients, we found a significant association between clinical stage, prognosis, and the same marker polymorphism (KRAS2/Run1) associated with lung cancer risk and prognosis in the Italian population. Pairwise LD between the biallelic markers used showed that LD can extend up to 200–300 kb.

Materials and Methods

Subjects. A total of 269 pathologically documented lung ADCA and 121 SCC patients were enrolled at several Japanese institutions. Clinical data included stage and survival time. A random sample of 632 unrelated control subjects was established for estimating population gene frequency. Controls were included in part using a case-control design and in part as a random sampling of the general population. They were sampled from the same areas and were of the same ethnic origin (Japanese) as cases. Genomic DNAs were isolated according to standard methodologies. Smoking information was obtained from clinical records; the ex-smoker category included subjects who stopped smoking more than 1 year before diagnosis/sampling.

Genetic Markers. A pilot study to identify genetic polymorphisms was conducted on selected Japanese cases and controls, using PCR primers allowing amplification of 1–2-kb DNA fragments containing the genetic markers of interest. Nucleotide sequences of the amplified fragments were obtained using an ABI PRISM 377 automatic sequencer (Perkin-Elmer). Nucleotide sequences were aligned and compared using the Genetics Computer Group software package to identify putative polymorphisms. PCR primers were designed to amplify DNA fragments of 100–200 bp (Table 1), and the SNPs identified were tested by ASO hybridization in the entire sample set, using the oligonucleotide probes reported in Table 1.

Genetic markers were amplified by PCR under the following conditions: 100–200 ng of genomic DNA; 1.5 mm MgCl2; 200 μm deoxynucleotide triphosphates; 0.25 unit of Taq polymerase (Perkin-Elmer/Roche); and 20 pmol of specific primers in a final volume of 20 μl. Thirty PCR cycles were performed. Aliquots of the PCR products were analyzed on an agarose gel and stained with ethidium bromide to check the quality of the PCR reactions. Physical mapping of genetic markers was performed using the GeneBridge4 RH panel (Research Genetics; Ref. 5).

ASO Hybridization. SNPs were detected by ASO hybridization. The PCR mix was denatured in 0.4 M NaOH/25 mm EDTA at room temperature and spotted onto a nylon membrane. Fifteen-mer ASOs, including the SNP at the central position, were 5’-end-labeled with [γ-32P]ATP (3000 Ci/mmol; Amerham, Branchburg, NY) and T4 polynucleotide kinase (New England Biolabs, Beverly, MA). ASO hybridizations were performed in tetramethylammonium chloride as described previously (6).

Statistical Analysis. Molecular genetic analysis was performed without knowledge of the clinical data, which were available after completion of the marker analysis. Fisher’s exact test was used to evaluate LD in the case of two alleles at both loci because it detects significance with a low probability of false positives, and P < 0.001 was considered significant (7, 8). Negative logarithms of the statistical P values were used to present the pairwise LD evidence more clearly. Allele and genotype associations with lung cancer risk were also tested by the Fisher’s exact test. The Kaplan-Meier product-limit method (9) was adopted to estimate survival functions. The null hypothesis concerning the differential effects of genotypes on univariate (unadjusted)

Received 1/3/00; accepted 8/2/00.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

4 The abbreviations used are: LD, linkage disequilibrium; ADCA, adenocarcinoma; SCC, squamous cell lung carcinoma; SNP, single-nucleotide polymorphism; ASO, allele-specific oligonucleotide; RH, radiation hybrid; YAC, yeast artificial chromosome.
analysis or after adjustment for gender and smoking habits (adjusted analysis) was tested by means of the log-rank test (10), and all \( P \) values were related to a two-sided significance test. In addition, Cox’s multiple regression analysis (11) was performed. The regression coefficients have been estimated by maximum likelihood criteria, and their significance was tested by Wald’s test (12). The relative risks reported in the text were estimated as hazard rate ratios.

### Results

Epidemiological data of the cases and controls are reported in Table 2. Our cases constituted representative samples of Japanese lung ADCA and SCC patients collected consecutively in different institutions. Mean age was slightly lower in ADCA cases than in SCC cases (Table 2), and nonsmokers had ADCA more frequently than SCC, as expected. Smoking habits were available for more than 90% of cases and for \( \sim 60\% \) of controls, which showed a distribution similar to that of ADCA cases (Table 2). Males were more frequent in SCC cases than in ADCA cases (Table 2).

#### LD and Physical Distance between Genetic Markers.

The genetic markers used span from the KRAS2 to ITPR2 genes (Table 1). A physical map of the genetic markers was assembled based on reported data (13, 14), on our YAC contig data, and on RH mapping (data not shown). In this RH panel, a distance of 234 kb is attributed to 1 cR for chromosome 12 (5). Although the algorithm used in RH mapping may lead to estimated physical distances \( < 1 \) cR, computer simulation indicated that the RH panel provides a mapping resolution of 1.3–2.1 cR (304–490 kb) in the 12p12 region. The mapping resolution represents a range of minimal estimated distances between two non-cosegregating markers, due to single variations in typing. Considering these uncertainties in the physical distances, the following marker order with distances in cR was obtained: KRAS2-(4.2 cR)-KRAG-(0.8 cR)-M4-(0.0 cR)-ITPR2. The marker order and estimated distances were in agreement with the YAC contig map (data not shown). The physical distance between the KRAS2 and ITPR2 markers was found to be 5.0 cR (\( \sim 1200 \) kb; Table 3).

Pairwise LD analysis between genetic markers revealed statistically significant LD (\( -\log P > 3 \)) between close markers. The highest LD (\( -\log P = 137 \)) was seen for the two KRAS2 polymorphisms, which were separated by \( \sim 20 \) kb. As expected, intragenic polymorphisms maintained a high degree of disequilibrium (15). The two M4 polymorphisms, located within a 1.3-kb DNA fragment that coamplifies with KRAS2 in some tumors (16), also showed a highly significant LD (\( -\log P = 43 \)). Pairwise LD results were in agreement with the marker order predicted by the physical map and showed that in our population sample, a statistically significant LD can be detected up to 1 cR (200–300 kb), as in the case of KRAG-M4 markers (\( -\log P = 14.7 \); Table 3).

#### Case-Control and Clinical Associations.

Allele frequencies at all loci were in Hardy-Weinberg equilibrium for both patients and controls (data not shown). Analysis of association of genetic markers with risk of lung cancer showed no statistically significant association at any markers, even after adjustment for gender and smoking history. However, genotype A2/A2 at marker KRAS2/RsaI, i.e., homozygous for the presence of the restriction site, was \( \sim 2\)-fold more frequent in ADCA patients than in controls (Table 4). In SCC, only three individuals carried the A2/A2 genotype, and no significant difference from controls was observed.

### Table 2 Epidemiological data of lung cancer patients and general population controls

<table>
<thead>
<tr>
<th>Group</th>
<th>M</th>
<th>F</th>
<th>Age</th>
<th>N</th>
<th>EXS</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>443</td>
<td>168</td>
<td>46.0 ± 0.6</td>
<td>160</td>
<td>36</td>
<td>169</td>
</tr>
<tr>
<td>ADCA</td>
<td>167</td>
<td>102</td>
<td>61.9 ± 0.6</td>
<td>85</td>
<td>38</td>
<td>126</td>
</tr>
<tr>
<td>SCC</td>
<td>101</td>
<td>20</td>
<td>61.6 ± 0.8</td>
<td>16</td>
<td>36</td>
<td>66</td>
</tr>
</tbody>
</table>

* Years (mean ± SE).

* N, nonsmoker; EXS, ex-smoker; S, smoker; number of missing subjects for smoking data, 246 controls, 20 ADCAs, and 3 SCCs.

### Table 3 Pairwise LD between genetic markers mapping in the human PAS1 homologous region (chromosome 12p12)*

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>~20</td>
<td>KRAS2/TaqI</td>
<td>137</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>~1000</td>
<td>KRAG/DdeI</td>
<td>0.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20–300</td>
<td>M4/StuI</td>
<td>1.7</td>
<td>1.4</td>
<td>14.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>~1</td>
<td>M4/BstXI</td>
<td>2.0</td>
<td>3.1</td>
<td>1.3</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>200–300</td>
<td>ITPR2/DdeI</td>
<td>0.7</td>
<td>0.1</td>
<td>2.7</td>
<td>5.7</td>
<td>10.7</td>
</tr>
</tbody>
</table>

* LD analysis in the Japanese population (lung ADCA and SCC cases, general population controls, \( n = 1022 \)) was performed using Fisher’s exact test; results were considered significant at \( -\log P > 3.0 \) (\( P < 0.001 \)).

* Estimated physical distance between flanking markers is based on known intragenic distance, YAC contig data, and screening results obtained using the Genebridge4 Radiation Hybrid panel, considering a distance of 234 kb as 1 cR (see "Results").
Clinical stage according to KRAS2/RsaI genotype showed a significant difference in the distribution of the stages among the genotypes ($P < 0.001$). An excess of stage 3 and stage 4 patients was observed in the A2/A2 and A1/A2 genotypes, respectively, as compared with A1/A1 genotypes (Table 4).

The median follow-up time for all survivors was 1490 days (range, 190-7027 days). All markers were tested for possible allelic association with survival rate in ADCA patients and SCC patients separately. In ADCA patients, a statistically significant association was found only for marker KRAS2/RsaI; the A2 allele, which was present at an excess frequency in ADCA cases as compared with the general population controls, was associated with poorer survival ($P = 0.0014$; Fig. 1). The probabilities of survival at 1200 days were 0.58 ± 0.05 ($n = 119$), 0.30 ± 0.07 ($n = 47$), and 0.20 ± 0.18 ($n = 5$) for those with the A1/A1, A1/A2, and A2/A2 genotypes, respectively ($P = 0.0030$), indicating a significant trend toward poor survival by copy number of the A2 allele. Cox’s regression analysis, including genotypes, sex, and smoking habits, was carried out by resorting to a backward procedure. Genotypes (relative risk, 1.8; 95% confidence interval, 1.14–2.85; $P = 0.01$) and sex (relative risk, 2.17; 95% confidence interval, 1.24–3.79; $P = 0.006$) remained statistically significant, whereas smoking habits failed to reach statistical significance.

No significant association was found between KRAS2/RsaI genotypes or any other marker genotypes and survival in SCC patients.

## Discussion

In our previous analysis of the KRAS2/RsaI marker and risk/prognosis of lung cancer in the Italian population, individuals homozygous for the KRAS2/RsaI A2 allele were ~2-fold more frequent in the general population controls than in lung ADCA patients, with a difference that was of borderline statistical significance (13). The same A2 allele associated with a protective effect on lung cancer risk was significantly associated with better survival (13). In the Japanese population, there was a ~2-fold excess frequency of the homozygous A2 genotype in lung ADCA patients as compared with controls (“risk allele”), although this was not statistically significant, due at least in part to the lower frequency of the KRAS2/RsaI A2 allele in the Japanese general population as compared with the Italian population.

The A2 allele was seen more frequently in the general population than in cancer patients in the Italian population, but the opposite distribution was seen in the Japanese population; the reason for this difference may rest in the different creation of the LD between the KRAS2/RsaI marker and the PAS1 locus in the two populations. The A2 allele, in excess frequency in lung ADCA patients, was significantly associated with poor prognosis in the Japanese cases. Therefore, the same KRAS2/RsaI A2 allele showed a significant and consistent pattern of association with disease prognosis in lung ADCA patients from two unrelated populations. The maintenance of LD between a disease allele (PAS1) and the same marker allele (KRAS2/RsaI) in two unrelated populations provides strong support for the implication of the PAS1 gene in human lung tumorigenesis and suggests a close relationship between the loci of the disease and the marker.

In a study of LD in human populations, Laan and Paabo (17) showed that recently expanded populations, such as the Finns, are well suited to map rare single-disease genes affected by recent mutations, whereas populations that have been of constant size may be much better suited to map genes involved in complex traits caused by older mutations. Slatkin (8) reached the same conclusions, finding that there is a substantial probability of obtaining significant nonrandom associations between closely linked polymorphic loci in a population of constant size at equilibrium, whereas in a rapidly growing population, the probability of detecting significant LD is low, even between completely linked loci. Our results indicate that both the Italian and Japanese populations are well suited for mapping lung cancer predisposition loci, in agreement with the findings of other studies (18).

Several studies have demonstrated a good correlation between physical distance and LD (15, 19, 20). However, the correlation is necessarily imperfect due to the type, age, frequency, mutation rate, and so forth of the polymorphisms. In a survey of LD between microsatellite loci spread over an anonymous genome region in the Finnish population, LD was detected between loci separated by up to 1 Mb (19). However, the detection limit for LD is ~50–300 kb when only biallelic markers are used (15, 20). In our total sample set of 1022 Japanese individuals, the biallelic markers showed a statistically significant LD with their flanking markers located in close vicinity, i.e., significant LD was detected up to a physical distance of 200–300 kb (Table 3). Thus, in light of the significant association between KRAS2 polymorphism and lung tumor prognosis, the PAS1 gene is most probably located within 200–300 kb of this polymorphism.

No LD is detectable between genetic markers and a disease locus resulting from multiple mutations, such as the NF1 locus (15). Our positive results with respect to the association between KRAS2 polymorphisms and risk/prognosis of human lung cancer suggest that the PAS1 susceptibility allele in humans results from a single or a predominant mutation/polymorphism. This prediction is consistent with an ancestral origin of the PAS1 susceptibility allele, which has been hypothesized based on LD analysis of mouse inbred strains (3).

The putative PAS1 locus showed a more statistically significant LD with prognosis than with risk of lung ADCA in both the Italian and Japanese populations. Prognosis of lung cancer patients may be affected by several factors, including gender and smoking habits (21). In

### Table 4 Genotype frequency of a KRAS2/RsaI polymorphism in Japanese lung ADCA patients by clinical stage and in the general population

<table>
<thead>
<tr>
<th>Genotype</th>
<th>Population controls</th>
<th>Lung ADCA</th>
<th>Stage *</th>
</tr>
</thead>
<tbody>
<tr>
<td>A1/A1</td>
<td>441</td>
<td>183</td>
<td>1</td>
</tr>
<tr>
<td>A1/A2</td>
<td>154</td>
<td>71</td>
<td>27</td>
</tr>
<tr>
<td>A2/A2</td>
<td>8 (1.3%)</td>
<td>6 (2.3%)</td>
<td>1</td>
</tr>
</tbody>
</table>

* Number of individuals by genotype; percentages of the A2/A2 genotypes are in parentheses.

* Clinical stage of lung ADCA patients by genotype ($P < 0.001$, $\chi^2$ test); percentages of stage 3 and 4 patients by genotype are in parentheses.
our ADCA patients, PAS1 and gender, but not smoking habits, were significantly associated with survival. Based on these data, it is possible that the low statistical level of LD with lung cancer risk is due to a relatively high frequency of the susceptibility allele in the general population. The PAS1 susceptibility allele may determine the risk of lung cancer in association with environmental risk factors (e.g., smoking). Indeed, Sellers et al. (22) suggested the involvement of a major predisposition gene in the risk of human lung cancer, i.e., the susceptibility allele of the predisposition gene causing lung cancer by interaction with tobacco smoking. Also, interaction of the PAS1 susceptibility allele with lung cancer resistance (Par) loci may mask the relevance of the PAS1 locus in association studies, underestimating the role of PAS1 in lung cancer risk (23). In any case, the putative human PAS1 locus showed a consistent association with lung ADCA prognosis, indicating a role for the PAS1 gene in lung tumor progression.

The use of several genetic markers over a short region in our population-based association study allowed us to estimate the physical distance between biallelic markers to which LD extends in the Japanese population and to confirm of the role and fine mapping of PAS1 in humans. The delineation of a short mapping region (200–300 kb) for the PAS1 locus by LD provides the basis for the positional cloning of the PAS1 gene. Availability of the cloned gene will enable analysis of the type of mutation/polyorphism associated with the risk and prognosis of lung ADCA and study of the frequency of the disease allele(s) in different populations. Identification of individuals at genetic risk for lung ADCA as well as epidemiological studies to characterize the interaction of genetic and environmental factors will also be facilitated.

Acknowledgments

We thank P. Valagussa for statistical assistance.

References

Population-based Mapping of Pulmonary Adenoma Susceptibility 1 Locus

Tommaso A. Dragani, Setsuo Hirohashi, Takeo Juji, et al.

Cancer Res 2000;60:5017-5020.

Updated version  Access the most recent version of this article at:  
http://cancerres.aacrjournals.org/content/60/18/5017

Cited articles  This article cites 22 articles, 10 of which you can access for free at:  
http://cancerres.aacrjournals.org/content/60/18/5017.full.html#ref-list-1

Citing articles  This article has been cited by 14 HighWire-hosted articles. Access the articles at:  
/content/60/18/5017.full.html#related-urls

E-mail alerts  Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions  To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions  To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.