incidence has been increasing in contrast to the decreasing trend of the intestinal type.[17–20] We have been especially interested in diffuse-type gastric cancer, because this type includes a distinct form called linitis plastica.[18] Linitis plastica, or type 4 advanced gastric cancer by the Japanese Classification,[13] accounts for approximately 10% of all gastric cancer in Japan, but its 5-year survival rate there is 10–20%[21] with no significant improvement in the past decade.

JSNP-based GWAS by the Millennium Genome Project. Various environmental factors, including diet and smoking, have been suggested in relation to a predisposition to gastric cancer.[17] In Japan, *H. pylori* infection appears almost as a prerequisite for the development of this cancer, because 99% of Japanese with gastric cancer and 90% of the Japanese adult (>40 years old) population were seropositive for *H. pylori*, and the infection was significantly associated with both the differentiated (≈ intestinal) and undifferentiated (≈ diffuse) types of gastric cancers.[22] Overall, however, development of the intestinal type is greatly influenced by environmental factors, while it is presumed that for the diffuse type a non-environmental factor, such as genetic predisposition, is also important. Several genetic and epigenetic alterations in gastric cancer were reported, including the *CDH1* (E-cadherin) gene mutation, which is causal to hereditary diffuse-type gastric cancer.[23] However, the pathogenic germline *CDH1* mutation is rare among Japanese pedigrees of familial gastric cancer,[24] and little is known about genetic factors involved in the polygenic, common type of diffuse-type gastric cancer. For instance, the role of the *CDH1* polymorphism is not clear in the Asian meta-analysis.[25] Therefore, we performed the first GWAS for this type of gastric cancer as a part of the five-disease joint GWAS (other diseases: Alzheimer's disease, type 2 diabetes, hypertension, and asthma) in the Millennium Genome Project.

The design and results of the GWAS were published previously,[11,26] and only their essential points are summarized here (Fig. 2 and Table 1). Briefly, the first stage of this two-stage GWAS analyzed 85576 JSNPs in 188 cases with linitis plastica and 752 references (a 752 case mix of four other common diseases and another 752 population control data from the JSNP database). We selected 2753 SNPs from the first-stage screening purely by statistical criteria, without relying on gene annotation, and genotyped another 749 cases with diffuse-type gastric cancer and 750 controls. The screening and a subsequent high-density typing identified a significant association with an SNP (rs2294008) in the first exon of the *PSCA* (prostate stem cell antigen) gene: gender- and age-adjusted odds ratio (OR) by dominant model = 4.18, 95% confidence interval (CI) = 2.88–6.21, $P = 1.5 \times 10^{-17}$ for a total of 925 cases with diffuse-type gastric cancer and 1396 controls. Because case-control association studies are so prone to error and bias, validating in independent populations is critical. The association of the *PSCA* SNP was replicated on diffuse-type gastric cancer in 454 cases and

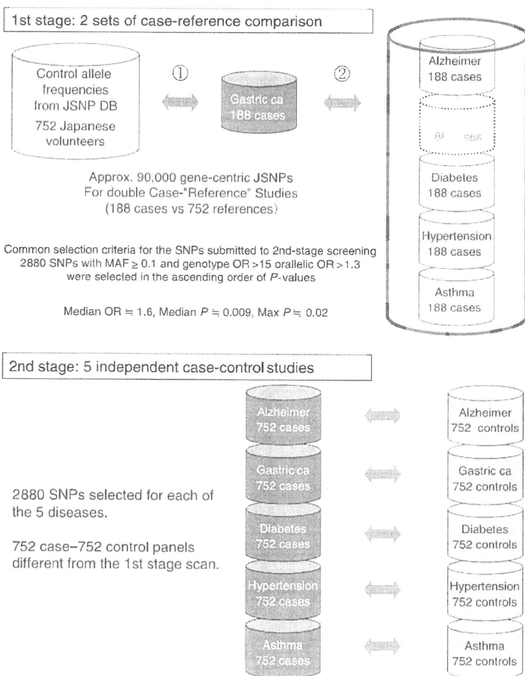

Fig. 2. Design of the two-stage genome-wide association study (GWAS) by Millennium Genome Project. As shown in Supplementary Table S1 online, GWAS is often performed in multiple stages. In the Millennium Genome Project, five diseases, including gastric cancer, were analyzed simultaneously in the same GWAS. Therefore, one of the reference (control) groups in the first screening was a case mix of the other four diseases, while the second stage was a more conventional case-control study. DB, database; MAF, minor allele frequency; OR, odds ratio.

doi: 10.1111/j.1349-7006.2010.01590.x
© 2010 Japanese Cancer Association

Table 1. Association of the *PSCA* SNP (rs2294008) and gastric cancers in Japan and Korea[11]

	Allele OR	95% CI	*P*-value (Fisher)	Dominant model OR	95% CI	*P*-value (logistic)
Risk allele frequency among 1396 Japanese control individuals = 0.617						
Diffuse type 925 cases	1.67	1.47–1.90	2.2×10^{-15}	4.18	2.88–6.21	1.5×10^{-17}
Intestinal type 599 cases	1.29	1.11–1.49	5.1×10^{-4}	1.59	1.15–2.21	0.0041
Risk allele frequency among 390 Korean control individuals = 0.462						
Diffuse type 454 cases	1.91	1.57–2.33	6.3×10^{-11}	3.61	2.41–5.51	3.2×10^{-11}
Intestinal type 417 cases	1.37	1.12–1.68	0.0017	1.85	1.27–2.71	0.0011

CI, confidence interval; OR, odds ratio; PSCA, prostate stem cell antigen.

390 controls in Korea (adjusted dominant model OR = 3.61, 95% CI = 2.41–5.51, $P = 3.2 \times 10^{-11}$). The SNP was far less significant in 599 and 417 cases with intestinal-type gastric cancer in Japan ($P = 0.0041$) and Korea ($P = 0.0011$), respectively. Moreover, Matsuo *et al.*[27] reported an independent replication of the association with the diffuse-type gastric cancer in Japan. They also detected a significant heterogeneity between the diffuse and intestinal types (*P*-heterogeneity = 0.007).

Functional studies on *PSCA* and its SNPs. The discovery of the *PSCA* SNP as a novel genetic susceptibility factor for gastric cancer was totally unexpected. The gene was originally identified as a prostate-specific antigen overexpressed in prostate cancers,[28,29] but it was also expressed strongly in the stomach and less intensely in the bladder, gallbladder, and tonsils.[30] It is noteworthy that the *PSCA* SNP identified for gastric cancer was recently found to be associated also with bladder cancer.[31] To explore a biological basis for this association, we performed a series of functional analyses of the gene and its SNPs as detailed previously.[11]

First, our anti-PSCA monoclonal antibody localized PSCA protein expression in differentiating epithelial cells in the isthmus and neck regions of normal gastric epithelium (Fig. 3). This was a very interesting observation, because these regions of the gastric gland are considered to harbor stem cells and precursors for the two-directional differentiation of gastric epithelial cells (Fig. 3) and because diffuse-type gastric cancer is suggested to arise from the stem cells and/or progenitors of the isthmus region.[16] Second, immunohistochemistry and quantitative RT-PCR revealed a frequent silencing of the gene in gastric cancer tissues. We analyzed 19 diffuse-type and 21 intestinal-type gastric cancers, and no PSCA staining was detectable in any of the diffuse-type cancers, while 20 of the 21 intestinal-type cancers appeared to retain some, but definitely reduced, PSCA expression, as compared to the surrounding normal gastric epithelium on the same section. Third, the transfection and overexpression of the *PSCA* cDNA induced inhibition of *in vitro* colony-formation and growth of the HSC57 gastric cancer cells, which do not detectably express the endogenous *PSCA* gene (Fig. 4). Fourth, a reporter assay on the 3.2-kb upstream fragment of the gene revealed that the fragment containing C allele at SNP (rs2294008) residing in the first exon has higher transcriptional activity than that containing T, the risk allele. In particular, a single substitution of the C allele with the risk allele T of rs2294008 reduces transcriptional activity of an upstream fragment of the *PSCA* gene, suggesting that the SNP is functionally responsible for the observed association with diffuse-type gastric cancer.

PSCA (prostate stem cell antigen) hypothesis generated by a hypothesis-free GWAS on diffuse-type gastric carcinogenesis. These functional studies suggest a simple but attractive hypothesis regarding the *PSCA* gene and diffuse-type gastric carcinogenesis as follows: "The PSCA protein is expressed predominantly at the stem cell/precursor-rich region of the gastric epithelium (which is also considered as the origin of diffuse type gastric cancer), has a tumor suppressor-like activity, and is involved in the regulation of gastric epithelial-cell proliferation. Individuals with a low promoter activity of the tumor-suppressive *PSCA* gene are susceptible to diffuse-type gastric cancer development. By contrast, the polymorphism does not significantly predispose to intestinal-type gastric cancer, congruous to the hypothesis of the two distinctive carcinogenesis pathways for the two major types of gastric cancer as depicted in Figure 1. PSCA (prostate stem cell antigen) appears to play a different role in a different tissue context, because the protein is

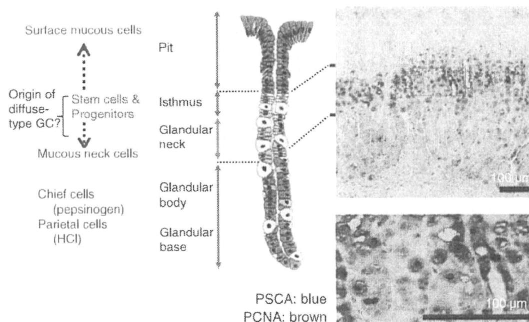

Fig. 3. PSCA (prostate stem cell antigen) expression at isthmus and neck region of gastric epithelium.[11] Center and left, a schematic architecture of the gastric gland. The isthmus is considered to contain stem cells and precursors of three major cell lineages of the gastric gland: pit, parietal, and zymogenic cells. Right, an immunohistochemistry by anti-PSCA monoclonal antibody (stained blue) and by anti-PCNA antibody (brown). Bar, 100 µm. GC, gastric adenocarcinomas; HCl, hydrochloric acid; PCNA, proliferating cell nuclear antigen. Modified from G.R. Van Den Brink *et al. Gastroenterol* 2001; 121:317.

© 2010 Japanese Cancer Association

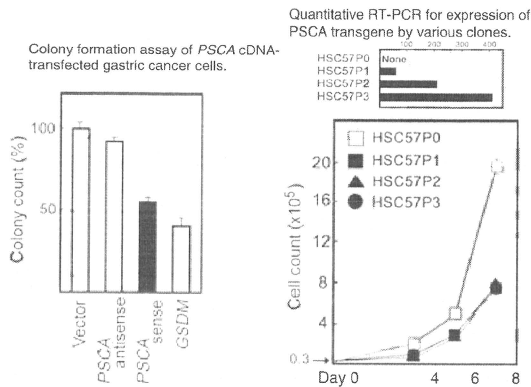

Fig. 4. *In vitro* tumor suppressor-like activities of *PSCA* (prostate stem cell antigen) cDNA.[11] Left, the *PSCA* cDNA was cloned into an expression vector driven by CMV promoter in the sense and antisense orientations and transfected to HSC57 gastric cancer cells. Colonies of stable transfectants selected by G418 selection were counted. The *GSDM* (gasdermin) cDNA was used as a positive control for the tumor suppressor-like activity.[47] Right, cell growth inhibition of the stable clones of the *PSCA*-transfected HSC57. The parental HSC57P0 cells do not detectably express endogenous *PSCA*.

overexpressed in prostate cancer cells.'' This study may become another example in which an agnostic GWAS has unveiled an unexpected molecular mechanism of a disease development.

Genome-Wide Association Study (GWAS) Databases and GeMDBJ

Genome-wide association study (GWAS) databases in the world. Collaboration, as we experienced at the validation stage on the Korean gastric cancer cases and controls in the *PSCA* research, has been not only crucial, as in much other human disease research, but also potentially robust for genetic association studies, because the genotype part of the relevant data is less prone to ascertainment error, unlike many other molecular phenotypes or clinico-epidemiological and lifestyle-related information. One step forward to such an ad hoc, gene- or marker-specific collaboration for a replication purpose would be the sharing of genome scan data through public accessible databases. In fact, there has been an increasing demand for an efficient use of research resources, such as invaluable human genome samples and phenotype information, especially those obtained by public funded research. The mission of such a database includes, at least (i) validation and evaluation of published research results and conclusions by allowing access to the authors' original data; (ii) offering reference data to allow comparison with the users' own experimental data; (iii) hypothesis generation through a biostatistical/bioinformatics exploration of the database data; and (iv) development and validation of tools and methods in biostatistics/bioinformatics. Therefore, a database should present data in two ways simultaneously: (i) through an easily navigatable, reference-dictionary type interface to respond to each user's specific data search; and (ii) as a downloadable flat file of as-raw-as-possible data.

A number of omics databases, especially those offering genome and/or transcriptome datasets, have been established on the web. For germline genomic data for human diseases, examples include: dbGaP (database of Genotyping and Phenotypes), which is presented by NCBI (National Center for Biotechnology Information) and archives and distributes the results of genome-wide association studies, medical sequencing, and molecular diagnostic assays, as well as the association between genotype and non-clinical traits (http://www.ncbi.nlm.nih.gov/gap/); CGEMS (Cancer Genetic Markers of Susceptibility), which presents GWAS data on various cancers such as breast, prostate, pancreatic, and lung from the NCI (National Cancer Institute) Cohort Consortium as well as collaborative case-control epidemiologic studies with biospecimens (http://cgems.cancer.gov/); WTCCC (Wellcome Trust Case Control Consortium), which provides GWAS data on case-control and cohort studies on various diseases, mostly on UK samples (http://www.wtccc.org.uk/); and GWAS DB and Mutation Database (http://gwas.lifesciencedb.jp/), which is a part of the Integrated Database project by the Ministry of Education, Culture, Sports, Science and Technology of Japan (http://lifesciencedb.mext.go.jp/en/index.html) and offers GWAS data on several diseases. Availability of GWAS databases representing different ethnic groups is important, because a minor allele frequency (MAF) of disease susceptibility locus is one of the critical factors determining the statistical power of a GWAS, and a significant ethnic difference in MAF is often observed among Caucasians, Africans, and Asians; GWAS in each ethnic group has actually detected a different set of SNPs in the same disease.[32]

Genome Medicine Database of Japan (GeMDBJ). GeMDBJ (Genome Medicine Database of Japan, http://gemdbj.nibio.go.jp/) was originally developed to release GWAS data from the Millennium Genome Project in Japan. Since the Project was finished in 2005, the database has been maintained by the National Institute of Biomedical Innovation (NiBio), a funding agency supported by the Ministry of Health, Labour and Welfare of Japan, and the contents of GeMDBJ have been further developed by a collaboration of five national centers and a national research institute: National Center for Neurology and Psychiatry (NCNP), National Cancer Center (NCC), National Center for Global Health and Medicine (NCGM, formerly International Medical Center of Japan), National Cardiovascular Center (NCVC), National Center for Child Health and Development (NCCHD), and National Institute of Health Sciences (NIHS) (Fig. 5). The database offers genome-wide SNP typing data (allele and genotype frequency data) for five major diseases (Alzheimer's disease, gastric cancer, type 2 diabetes, hypertension, and asthma) targeted by the Millennium Genome Project. In addition to the SNP data, the database contains Affymetrix GeneChip transcriptome data on >1000 samples, most of which are various types of clinical cancer tissues. A recent addition to the database is GeMDBJ Proteomics,[33] which presents cancer proteome analysis data based on the standardized protocol[34]

doi: 10.1111/j.1349-7006.2010.01590.x
© 2010 Japanese Cancer Association

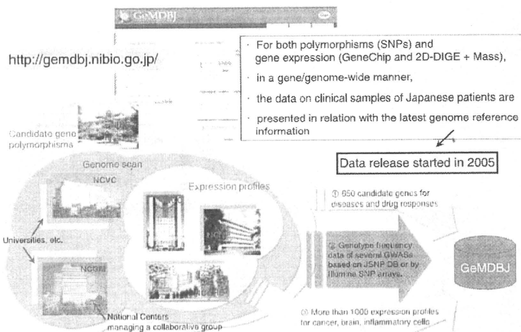

Fig. 5. Status of GeMDBJ (Genome Medicine Database of Japan). Since the completion of the Millennium Genome Project, GeMDBJ (http://gemdbj.nibio.go.jp/) has been developed by a joint effort of five national centers (NCNP, NCC, NCGM, NCVC, and NCCHD) and a national research institute (NIHS) (see text for full names). The left figure shows that expression profile data have been provided by three centers, genome scan data from two additional (total five) centers, and candidate gene SNP data from all six institutions, each specializing in different diseases (five national centers) or on drug responses (NIHS). GeMDBJ Proteomics was constructed by NCC researchers.[33]

with protein spot quantification by 2D-DIGE (two-dimensional difference gel electrophoresis) and protein identification by mass spectrometry.

Privacy protection issue with GWAS databases. In August 2008, a shockwave spread among GWAS database constructors. A paper by Homer *et al.*[35] demonstrated that it is feasible to infer whether an individual DNA sample was included in an aggregate genotype dataset, such as genotype and allele frequency data of a group of patients, which had been considered safe for posting in the publicly accessible database. Their simulation suggested that 50 000 SNP data can detect an individual within aggregate data of 1000 people. To make such a group assignment (case or control) possible for an individual, it is necessary to have high-density genomic data for that specific individual, as well as the allele frequencies of the control (or reference) population. While the latter population control genomic data are available for many major ethnic groups, the former individual genome-wide data are currently unlikely to be easily obtained outside the research context. However, such a situation will rapidly change by the introduction of commercial or medical massive genotyping or resequencing services. In response to the Homer paper, the major GWAS databases removed aggregate datasets from the open access tier and moved them into the controlled access category.[36] In Japan, the Integrated Database Project has recently established a policy to share the GWAS and resequencing data for disease-associated genes (http://gwas.lifesciencedb.jp/gwasdb/db_policy_en.html), which we have decided to adopt for GeMDBJ, because the policy was very well contemplated and also because we believe that policy standardization is important.

Towards an integrated database. Currently, GeMDBJ should be considered still in its infancy in the sense that each dataset is presented relatively separately, and we are exploring a way for a powerful synthetic analysis by correlating the multiple types of the omics and phenotype data. However, the potential power of such omics data integration was already hinted at in our first experience with gastric cancer GWAS. Because the data from the Millennium Genome Project was posted in the database as soon as possible, the first screening data from the two-stage JSNP genome scan had been already made available to the public when we were still in the midst of narrowing down our target to the *PSCA* gene through extensive second-stage screening, validation by the Korean researchers, and functional analyses. At that time, hundreds of Affymetrix GeneChip data were also accessible at GeMDBJ on various types of cells and tissues, including those of the stomach. When we combined the database search for a low *P*-value in the GWAS first-screening data and the increased expression in the gastric tissue, we were surprised to see that the *PSCA* gene was actually selected, without the then unpublished second-screening data. Of course, this could not be generalized for all of the susceptibility genes, because some of them may not necessarily be expressed in the tissues from which the cancer arises; for instance, some cancer susceptibility genes may not be expressed in the target tissues but in the liver or kidney involved in carcinogen metabolism and transport. Moreover, the *PSCA* case should be considered a rather fortunate case, because, as described in the Introduction, a hypothesis-free GWAS often results in a genome region in which a protein-coding gene was not known or cannot be singled out from many possible candidates. Integration of various data, knowledge, and information by an interactive database is expected to play a crucial role in such difficult GWAS cases and may accelerate a novel discovery in science and medicine.

Prospects for the Next Generation Germline Analyses

The genetic architecture of disease susceptibility has not yet been fully resolved for any common disease. Even for cases with identified causative gene mutations of classical Mendelian diseases, inter-pedigree differences are often observed regarding disease penetrance and expressivity, which may be explained by an involvement of modifier genes. For common diseases, one popular hypothesis for genetic architecture is a CD-CV, or common disease-common variant, hypothesis which predicts a relatively frequent (common) disease susceptibility allele at each of the major underlying disease loci.[37] Because disease susceptibility alleles have been able to achieve a high equilibrium frequency in a given population, they should at least have little or no effect on reproductive fitness,[38] and the overall effect size (such as risk ratio) may also be small. The classical example of the CD-CV model is the *APOE* e4 allele for Alzheimer's disease.[39] Although common diseases are believed to be polygenic, the number of common alleles involved in a disease is still a matter of speculation; for instance, as suggested by one simulation for breast cancer,[40] there may be 30–40 alleles with a relative risk of 1.5 and a minor allele frequency of 0.1. However, there are several cases against the CD-CV hypothesis, with the alternative model being called CD-MRV (multiple rare variants) or by other related names.[41] The simple but strong argument for CD-MRV is that a common disease is common because of highly prevalent non-genetic factors, not because of common "susceptibility alleles" in a given population.[39] It

© 2010 Japanese Cancer Association

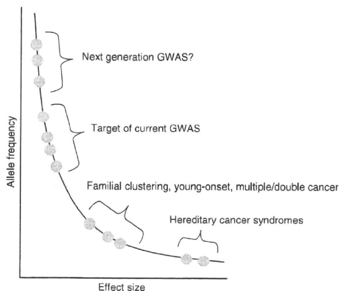

Fig. 6. Genetic architecture of disease susceptibility (an image). It is assumed that multiple genetic polymorphisms with a range of allele frequencies and effect sizes constitute an individual and population genetic architecture of disease risk. However, current and past linkage and association studies could access only limited segments of the curve. Personal genome sequencing by next or future generation sequencers is expected to fill, at least some, gaps. GWAS, genome-wide association study. Adapted from ref. 39.

has been predicted that "neutral" susceptibility alleles without evolutionary selection pressure become non-polymorphic, while susceptibility alleles under weak selection tend to remain polymorphic, especially at loci with high mutation rates. In this model, there will be extensive allelic heterogeneity in the genetic architecture of a common disease. Although some of the susceptibility alleles may be rare in the population, the collective frequency of these alleles may be quite high.[39,41,42]

Figure 6 illustrates an image of the genetic architecture of cancer (modified from ref. 39). The current GWAS covers a relatively limited range of allele frequency (i.e. common variants) and effect size (OR of approximately 1.2 or higher) for most diseases. To capture the entire landscape of genetic factors, the current GWAS can be extended along the upper left direction (higher frequency and lower effect size), if a sufficient research resource is available, but the clinical/public health significance is doubtful.[41] On the other hand, it is conceivable that a number of significant genetic factors are yet to be disclosed on the right lower area; in addition to some cases with clinically definitive, but mutation-undetectable, hereditary cancer syndromes, many cases with familial clustering, young age at onset, and/or multiple/double cancers may be due to underlying genetic causes.

For these cases, the recent advent of whole genome or even exome sequencing is beginning to show several early successes on recessive non-cancer hereditary diseases.[43–45] Unlike classical linkage analyses on large pedigrees, it is remarkable that these studies reached the causative genes on a small number of subjects, who were not necessarily related,[43] although a family-based sequencing would be more powerful.[44] The hurdles will be higher for dominant Mendelian traits,[43] which may account for most of the hereditary cancer syndromes. However, many of them are "dominant" because of the second-hit somatic inactivation of a causative tumor suppressor gene. The combination of the exomic sequencing of the tumor and blood (germline) DNAs identified a novel susceptibility gene for a familial pancreatic cancer.[46] In the coming era of personal genome sequencing, we will witness amazing progress in unveiling the missing portion of the genetic architecture of cancer (Fig. 6), which will be followed by intervention research for the development of personalized prevention based on genomic factors.

Acknowledgments

We thank all the co-authors of reference 11, which is the major original work for this review article. As regards GeMDBJ, we are deeply grateful to Drs Yuichi Goto, Kazuki Yasuda, Naoharu Iwai, Kenji Matsumoto, Yoshiro Saito, Tadashi Kondo, and all the other investigators who kindly contributed their data and ideas to the database. We also thank the engineers and scientists of Mitsui Knowledge Industry Co., Ltd., for their efforts in the database construction. This work was supported by the Program for Promotion of Fundamental Studies in Health Sciences of the National Institute of Biomedical Innovation (NiBio) and by a Grant-in-Aid for Cancer Research from the Ministry of Health, Labour and Welfare of Japan.

Disclosure Statement

The authors have no conflict of interest.

Abbreviations

CI	confidence interval
OR	odds ratio
PSCA	prostate stem cell antigen

References

1 Risch N, Merikangas K. The future of genetic studies of complex human diseases. *Science* 1996; **273**: 1516–17.
2 Yokoyama A, Omori T. Genetic polymorphisms of alcohol and aldehyde dehydrogenases and risk for esophageal and head and neck cancers. *Jpn J Clin Oncol* 2003; **33**: 111–21.
3 Li H, Hao X, Zhang W, Wei Q, Chen K. The hOGG1 Ser326Cys polymorphism and lung cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2008; **17**: 1739–45.
4 Horvath S, Baur MP. Future direction of research in statistical genetics. *Stat Med* 2000; **19**: 3337–43.
5 Pritchard JK, Przeworski M. Linkage disequilibrium in humans: models and data. *Am J Hum Genet* 2001, **69**: 1–14.
6 Hirakawa M, Tanaka T, Hashimoto Y, Kuroda M, Takagi T, Nakamura Y. JSNP: a database of common gene variations in the Japanese population. *Nucleic Acids Res* 2002; **30**: 158–62.
7 Haga H, Yamada R, Nakamura Y, Tanaka T. Gene-based SNP discovery as part of the Japanese Millennium Genome Project: identification of 190562 genetic variations in the human genome. *J Hum Genet* 2002; **47**: 605–10.
8 Ozaki K, Ohnishi Y, Iida A *et al.* Functional SNPs in the lymphotoxin-α gene that are associated with susceptibility to myocardial infarction. *Nat Genet* 2002; **32**: 650–4.
9 The International HapMap Consortium. The International HapMap project. *Nature* 2003; **426**: 789–96.
10 Yoshida T, Yoshimura K. Outline of disease gene hunting approaches in the Millennium Genome Project of Japan. *Proc Jpn Acad* 2003; **79**: 34–50.
11 Sakamoto H, Yoshimura K, Saeki N *et al.* Genetic variation in *PSCA* is associated with susceptibility to diffuse-type gastric cancer. *Nat Genet* 2008; **40**: 730–40.
12 http://www.fpcr.or.jp/publication/pdf/statistics2009/date04.pdf
13 Japanese Gastric Cancer Association. Japanese classification of gastric carcinoma. 2nd English edition. *Gastric Cancer* 1998; **1**: 10–24.
14 Lauren P. The two histological main types of gastric carcinoma: diffuse and so-called intestinal-type carcinoma. An attempt at a histo-clinical classification. *Acta Pathol Microbiol Scand* 1965; **64**: 31–49.
15 Hohenberger P, Gretschel S. Gastric cancer. *Lancet* 2003; **362**: 305–15.
16 Schier S, Wright NA. Stem cell relationships and the origin of gastrointestinal cancer. *Oncology* 2005; **1** (Suppl 1): 9–13.
17 Crew KD, Neugut AI. Epidemiology of gastric cancer. *World J Gastroenterol* 2006; **12**: 354–62.
18 Rosai J. Gastrointesitinal tract–stomach. In: *Rosai and Ackerman's Surgical Pathology*. Rosai J, ed. Edinburgh: Mosby, 2004; 648–711.
19 Yokota T, Teshima S, Saito T, Kikuchi S, Kunii Y, Yamauchi H. Borrmann's type IV gastric cancer: clinicopathologic analysis. *Can J Surg* 1999; **42**: 371–6.

doi: 10.1111/j.1349-7006.2010.01590.x
© 2010 Japanese Cancer Association

20 Henson DE, Dittus C, Younes M, Nguyen H, Albores-Saavedra J. Differential trend in the intestinal and diffuse types of gastric carcinoma in United States, 1973–2000–increase in the signet ring cell type. *Arch Pathol Lab Med* 2004; **128**: 765–70.
21 Japanese Gastric Cancer Association Registration Committee, Maruyama K, Kaminishi M *et al.* Gastric cancer treated in 1991 in Japan: data analysis of nationwide registry. *Gastric Cancer* 2006; **9**: 51–66.
22 Sasazuki S, Inoue M, Iwasaki M *et al.* Effect of *Helicobacter pylori* infection combined with CagA and pepsinogen status on gastric cancer development among Japanese men and women: a nested case-control study. *Cancer Epidemiol Biomarkers Prev* 2006; **15**: 1341–7.
23 Guilford P, Hopkins J, Harraway J *et al.* E-cadherin germline mutations in familial gastric cancer. *Nature* 1998; **392**: 402–5.
24 Shinmura K, Kohno T, Takahashi M *et al.* Familial gastric cancer: clinicopathological characteristics, RER phenotype and germline p53 and E-cadherin mutations. *Carcinogenesis* 1999; **20**: 1127–31.
25 Wang G-Y, Lu C-Q, Zhang R-M, Hu X-H, Luo ZW. The E-cadherin gene polymorphism -160C/A and Cancer Risk: a HuGE review and meta-analysis of 26 case-control studies. *Am J Epidemiol* 2008; **167**: 7–14.
26 Sato Y, Suganami H, Hamada C, Yoshimura I, Yoshida T, Yoshimura K. Designing a multistage, SNP-based, genome screen for common diseases. *J Hum Genet* 2004; **49**: 669–76.
27 Matsuo K, Tajima K, Suzuki T *et al.* Association of prostate stem cell antigen gene polymorphisms with the risk of stomach cancer in Japanese. *Int J Cancer* 2009; **125**: 1961–4.
28 Reiter RE, Gu Z, Watabe T *et al.* Prostate stem cell antigen: a cell surface marker overexpressed in prostate cancer. *Proc Natl Acad Sci USA* 1998; **95**: 1735–40.
29 Gu Z, Thomas G, Yamashiro J *et al.* Prostate stem cell antigen (PSCA) expression increases with high gleason score, advanced stage and bone metastasis in prostate cancer. *Oncogene* 2000; **19**: 1288–96.
30 Bahrenberg G, Brauers A, Joost H-G, Jakse G. Reduced expression of PSCA, a member of the LY-6 family of cell surface antigen, in bladder, esophagus, and stomach tumors. *Biochem Biophys Res Commun* 2000; **275**: 783–8.
31 Wu X, Ye Y, Kiemeney LA *et al.* Genetic variation in the prostate stem cell antigen gene PSCA confers susceptibility to urinary bladder cancer. *Nat Genet* 2009; **41**: 991–5.
32 Yasuda K, Miyake K, Horikawa Y *et al.* Variants in KCNQ1 are associated with susceptibility to type 2 diabetes mellitus. *Nat Genet* 2008; **40**: 1092–7.
33 Kondo T. Proteome expression database; Genome Medicine Database of Japan Proteomics. *Expert Rev Proteomics* 2010; **7**: 21–7.
34 Kondo T, Hirohashi S. Application of highly sensitive fluorescent dyes (CyDye DIGE Fluor saturation dyes) to laser microdissection and two-dimensional difference gel electrophoresis (2D-DIGE) for cancer proteomics. *Nat Protoc* 2007; **1**: 2940–56.
35 Homer N, Szelinger S, Redman M *et al.* Resolving individuals contributing trace amounts of DNA to highly complex mixtures using high-density SNP genotyping microarrays. *PLoS Genet* 2008; **4**: e1000167.
36 Couzin J. Genetic privacy. Whole-genome data not anonymous, challenging assumptions. *Science* 2008; **321**: 1278.
37 Reich DE, Lander ES. On the allelic spectrum of human disease. *Trends Genet* 2001: **17**: 502–10.
38 Kimura M, Ohta T. The age of a neutral mutant persisting in a finite population. *Genetics* 1973; **75**: 199–212.
39 Wright AF, Hastie ND. Complex genetic diseases: controversy over the Croesus code. *Genome Biol* 2001;**2**: commnet2007.1-2007.8. Epub 2001 Aug 1.
40 Ponder BA. Cancer genetics. *Nature* 2001; **411**: 336–41.
41 Bodmer W, Bonilla C. Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 2008; **40**: 695–701.
42 Pritchard JK. Are rare variants responsible for susceptibility to complex diseases? *Am J Hum Genet* 2001: **69**: 124–37.
43 Ng SB, Buckingham KJ, Lee C *et al.* Exome sequencing identifies the cause of a mendelian disorder. *Nat Genet* 2010; **42**: 30–5.
44 Roach JC, Glusman G, Smit AF *et al.* Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 2010; **328**: 636–9.
45 Lupski JR, Reid JG, Gonzaga-Jauregui C *et al.* Whole-genome sequencing in a patient with Charcot-Marie-Tooth neuropathy. *N Engl J Med* 2010; **362**: 1181–91.
46 Jones S, Hruban RH, Kamiyama M *et al.* Exomic sequencing identifies *PALB2* as a pancreatic cancer susceptibility gene. *Science* 2009; **324**: 217.
47 Saeki N, Kim DH, Usui T *et al.* GASDERMIN, suppressed frequently in gastric cancer, is a target of LMO1 in TGF-β-dependent apoptotic signalling. *Oncogene* 2007; **26**: 6488–98.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1. Genome-wide association study (GWAS) and related studies on cancers reported in *Nature Genetics* (as of December 2009).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

© 2010 Japanese Cancer Association

ELSEVIER

Cancer Genetics and Cytogenetics 197 (2010) 8–15

CANCER GENETICS AND CYTOGENETICS

Analysis of poly(ADP-ribose) polymerase-1 (*PARP1*) gene alteration in human germ cell tumor cell lines

Hideki Ogino[a,b], Robert Nakayama[c], Hiromi Sakamoto[c], Teruhiko Yoshida[c], Takashi Sugimura[a], Mitsuko Masutani[a,b,*]

[a]*Biochemistry Division, National Cancer Center Research Institute, 1-1 Tsukiji 5-chome, Chuo-ku, Tokyo 104-0045, Japan*
[b]*ADP-ribosylation in Oncology Project, National Cancer Center Research Institute, 1-1 Tsukiji 5-chome, Chuo-ku, Tokyo 104-0045, Japan*
[c]*Genetics Division, National Cancer Center Research Institute, 1-1 Tsukiji 5-chome, Chuo-ku, Tokyo 104-0045, Japan*

Received 28 June 2009; received in revised form 17 October 2009; accepted 17 October 2009

Abstract

The poly(ADP-ribose) polymerase-1 protein (PARP-1) functions in DNA repair, maintenance of genomic stability, induction of cell death, and transcriptional regulation. We previously analyzed alterations of the *PARP1* gene in 16 specimens of human germ cell tumors, and found a heterozygous sequence alteration that causes the amino acid substitution Met129Thr (M129T) in both tumor and normal tissues in a single patient. In this study, aberration of the *PARP1* gene and protein was further analyzed in human germ cell tumor cell lines. We found a nonheterozygous sequence alteration that causes the amino acid substitution Glu251Lys (E251K) located at a conserved peptide stretch of PARP-1 in cell line NEC8. Sequencing of 95 samples from Japanese healthy volunteers revealed that all the samples were homozygous for the wild-type alleles at M129T and E251K. The M129T allele is thus suggested to be a rare single-nucleotide polymorphism (SNP). We observed a decrease in auto-poly(ADP-ribosyl)ation activity of PARP-1 proteins harboring M129T or E251K amino acid substitution, but the difference was not statistically significant. The levels of PARP-1 and poly(ADP-ribosyl)ation were heterogeneous among germ cell tumor cell lines. The SNPs of the *PARP1* gene, as well as differences in the levels of PARP-1 and poly(ADP-ribosyl)ation of proteins, may influence germ cell tumor development and responses to chemotherapy and radiotherapy. © 2010 Elsevier Inc. All rights reserved.

1. Introduction

Poly(ADP-ribose) polymerase-1 (PARP-1) is activated by DNA damage and catalyzes poly(ADP-ribosyl)ation of various proteins, including PARP-1 itself, using nicotinamide adenine dinucleotide (NAD^+) as a substrate. PARP-1 is involved in DNA repair, maintenance of genomic stability, and cell death induction. We and others have previously reported that *Parp-1* knockout (*Parp-1*$^{-/-}$) mice showed higher susceptibility to carcinogenesis induced by alkylating agents in the colon [1] and lung [2], compared with wild-type (*Parp-1*$^{+/+}$) mice. The incidence of spontaneous tumors developed at an advanced age in the liver was also higher in *Parp-1*$^{-/-}$ mice [3,4].

Involvement of PARP-1 in the development of human cancer has not yet been fully clarified. In human cancers, increased expression of the *PARP1* gene has been reported in Ewing's sarcoma [5,6], in malignant lymphoma [7], and in the familial adenomatous polyposis (FAP) tumors [8]. Decreased expression of the *PARP1* gene has been observed in several gastric and colon cancer cell lines [6], grade II and III endometrial carcinomas [9], and in some breast cancers [10].

The A/A homozygotes of the V762A single-nucleotide polymorphism (SNP) in the *PARP1* gene have been reported to be associated with decreased activity of PARP-1. The A/A homozygotes are shown to be associated with an increased risk for prostate cancer in European-origin subjects [11] and in lung cancer and esophageal cancers in Chinese heavy smokers [12,13]. In the case of lung and esophageal cancers, a twofold increase in risk with the A/A homozygotes was observed in Chinese smokers [12,13]. The combination of the 762A allele of the *PARP1* gene and the 399 G allele of the *XRCC1* gene was associated with increased risk of lung, esophageal, and gastric cardia cancers [12–15].

* Corresponding author. Tel.: +81-3-3542-2511; fax: +81-3-2542-2530.

E-mail address: mmasutan@ncc.go.jp (M. Masutani).

0165-4608/10/$ – see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cancergencyto.2009.10.012

PARP-1 also participates in the transcriptional regulation of some genes [16–18] and in cellular differentiation [18–20]. *Parp-1*$^{-/-}$ mouse embryonic stem cells show preferential induction of the trophoblast lineage [20], including trophoblast giant cells (TGCs), during teratocarcinoma formation in vivo or during cell culture in vitro [19]. The biochemical properties of TGCs resemble those of syncytiotrophoblastic giant cells (STGCs) of human germ cell tumors [21,22]. It is thus suggested that *PARP1* deficiency may possibly trigger differentiation to STGCs during germ cell tumor formation. The appearance of STGCs in trophoblastic or choriocarcinomatous human germ cell tumors has been reported to be associated with poor prognosis [21]. Teratocarcinoma cells undergo differentiation into epithelial cells in vitro, at least in part, in the presence of the PARP inhibitor 3-aminobenzamide [23].

The aberrations of the *PARP1* gene in 16 human germ cell tumors were previously analyzed, and a heterozygous sequence alteration (ATG to ACG) that causes amino acid substitution, Met129Thr (M129T) [24] was found in one patient in both cancer and normal tissues. In the present study, we further analyzed aberration of the *PARP1* gene and poly(ADP-ribosyl)ation level in human germ cell tumor cell lines and evaluated effects of any amino acid alterations found on PARP-1 function.

2. Materials and methods

2.1. Cell culture

Cell lines JEG-3, NCCIT, PA-1, Tera-1, and Tera-2 were purchased from the American Type Culture Collection (ATCC, Manassas, VA). Cell lines ITOII and NEC8 were purchased from the Japanese Collection of Research Bioresources (http://cellbank.nibio.go.jp/cellbank_e.html). Cell lines NEC14 and NEC15 were purchased from the Riken Bioresource Center (http://www.brc.riken.go.jp/inf/en/index.shtml). Each cell line was cultured under the conditions recommended by the providers. The growing cells at mid-late log phase were used as the materials for preparation of genomic DNA, total RNA, and total protein.

For transfection experiments, an immortalized *Parp-1*$^{-/-}$ mouse embryonic fibroblast (MEF) clone, PH13b, established from spontaneously immortalized *Parp-1*$^{-/-}$ MEFs [25] was used. The cells were cultured at 37°C under 5% CO_2 and 95% humidity.

2.2. Direct sequencing of the human PARP1 gene

Polymerase chain reaction (PCR)-based direct sequencing of all 23 exons of the *PARP1* gene in germ cell tumor cell lines was performed as previously described [24]. Oligonucleotide primer sets for the 23 exons were designed from intron sequences of each exon as previously described [24]. Amplified PCR products were subjected to sequence analysis (ABI PRISM 310 genetic analyzer, Applied Biosystems, Carlsbad, CA; model CEQ8000, Beckman Coulter, Fullerton, CA). Sequence comparison was performed against the sequence of the human *PARP1* gene (NCBI accession numbers NT_004559 and NT_167186) and its cDNA (NCBI accession numbers M18112, M32721, M17081, J03473, BC037545, and BC014206).

2.3. Pyrosequencing

Pyrosequencing for codon 129 and codon 251 of the *PARP1* gene was performed as previously described [26]. Briefly, genomic DNA samples were extracted from blood of 95 healthy Japanese volunteers, and samples were subjected to genotyping by pyrosequencing using the PSQ96 System (Pyrosequencing, Uppsala, Sweden). For sequencing of codons 129 and 251, we used 5′-CGTGCAAGGGGTGTA-3′ and 5′-TGAACACACTTTCTTTAGC-3′ as sequencing primers, respectively. In the case of codon 251, the sequencing result of one sample was not informative.

All subjects provided informed consent, and the study was approved by the Ethical Committee of the National Cancer Center of Japan.

2.4. Construction of PARP1 mutants

PARP1 mutant cDNA harboring either M129T, E251K, or K940R amino acid substitution was prepared using primers harboring respective mutation and Phusion polymerase (Fynnzymes, Espoo, Finland) according to the protocol of the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA). The human *PARP1* cDNA [27] was modified using PCR to harbor restriction enzyme recognition sites of *Sal*I, *Sma*I, *Age*I, and *Xba*I at the 5′ and 3′ terminus, respectively, and modified Kozack's sequences derived from pEGFP-C1 (Clontech Laboratories, Mountain View, CA) in the 5′-UTR (untranslated region) of the *PARP1* cDNA. The *PARP1* cDNA harboring either M129T, E251K, and K940R amino acid substitution was inserted into pcDNA3.1(+)hygro (Invitrogen, Carlsbad, CA) for measurement of PARP-1 enzymatic activity. pEGFP-C1 was used for construction of the GFP protein fused to the N-terminus of PARP-1 protein for analysis of subcellular localization.

2.5. Transfection of the PARP1 mutant constructs

Four micrograms of each construct was transfected into the *Parp-1*$^{-/-}$ MEFs in six-well plates using Lipofectamine 2000 (Invitrogen). At 24 hours after transfection, whole-cell extracts were prepared by suspending cells in a lysis solution containing 50 mmol/L Tris–HCl (pH 6.8), 10% glycerol, 2% sodium dodecyl sulfate (SDS), complete protease inhibitor cocktail tablets (Roche Applied Science, Mannheim, Germany), and 10% β-mercaptoethanol. The samples were frozen in liquid nitrogen and stored at −80°C until use.

2.6. Activity gel analysis

PARP-1 enzymatic activity was measured as auto-poly(ADP-ribosyl)ation activity of PARP-1, as described elsewhere [28]. Briefly, crude extracts were separated by 6% SDS–polyacrylamide gel electrophoresis (SDS-PAGE) containing 100 μg/mL of sonicated salmon sperm DNA as an activated DNA. After renaturation of proteins in the gel, the gel was incubated in a reaction mixture containing 50 mmol/L Tris–HCl (pH 8.0), 1 mmol/L dithiothreitol, 50 μmol/L [^{32}P]adenylate-labeled nicotinamide adenine dinucleotide (1 μCi/mL, NEN–PerkinElmer, Waltham, MA), and 25 mmol/L $MgCl_2$ at 37°C for 1 hour. The gel was fixed with 10% methanol–30% acetic acid (v/v) solution, and washed with 5% trichloroacetic acid–0.2% sodium pyrophosphate. The radioactivities of the dried gels were analyzed using a BAS-2500 bio-imaging analyzer (Fujifilm, Tokyo, Japan).

2.7. Western blot analysis

Whole-cell extracts were prepared by suspending cells in a lysis solution containing 50 mmol/L Tris-HCl (pH 6.8), 10% glycerol, 2% SDS, complete protease inhibitor cocktail tablets (Roche Applied Science), and 10% β-mercaptoethanol, followed by sonication. Equivalent protein amounts of lysate (8 μg) were separated by 4–20% gradient SDS-PAGE. After transfer of proteins to Immobilon-P polyvinylidene difluoride membranes (Millipore, Billerica, MA), the membrane was incubated with anti-poly(ADP-ribose) monoclonal antibody 10H (Alexis Biochemicals–Enzo Life Sciences, Lausanne, Switzerland) [29], anti-PARP-1 monoclonal antibody C2-10 (Oncogene Research Products, Merck Chemicals, Darmstadt, Germany), F1-23 (Alexis Biochemicals) or anti-α-tubulin monoclonal antibody DM1A (MP Biomedicals, Irvine, CA). Immune complexes were visualized using a horseradish peroxidase-linked secondary antibody and an enhanced chemiluminescence reaction ECL kit (Amersham Biosciences, Piscataway, NJ). The PARP-1 protein level was quantified using a LAS-3000 bio-imaging analyzer (Fujifilm).

3. Results

3.1. Sequence alterations in germ cell tumor cell lines

The human germ cell tumor cell lines used in this study are listed in Table 1. All 23 exons and their flanking regions of the *PARP1* gene in nine germ cell tumor cell lines were sequenced. Sequence alterations and SNPs found in the *PARP1* gene are listed in Table 2. We found that the two missense SNPs, Phe54Leu (F54L) and V762A, were described in the NCBI (National Center for Biotechnology Information) database of SNPs. These SNPs of F54L and V762A in NEC14 and NEC15 were observed as nonheterozygous sequence alterations. A nonheterozygous sequence alteration (GAG to AAG) that causes amino acid substitution Glu251Lys (E251K) in NEC8 was also found (Table 2 and Fig. 1). This sequence alteration has not been listed in the NCBI database of SNPs. Further information for NEC8 was not available, and it is not known whether this is a somatic or germ line mutation or sequence alteration caused during establishment of this cell line.

Table 1
Human germ cell tumor cell lines used in this study

Cell line	Type	Reference
NEC8	Embryonal carcinoma: testis	Motoyama et al., 1987 [39]
NEC14	Embryonal carcinoma, choriocarcinoma: testis	Motoyama et al., 1987 [39]
NEC15	Embryonal carcinoma, yolk sac tumor: testis	Motoyama et al., 1987 [39]
ITOII	Embryonal carcinoma: testis	Motoyama et al., 1987 [39]
Tera-1	Embryonal carcinoma: lung	Fogh et al., 1978 [40]
Tera-2	Embryonal carcinoma: lung	Fogh et al., 1978 [40]
NCCIT	Embryonal carcinoma	Teshima et al., 1988 [41]
PA-1	Teratoma: ovary	Zeuthen et al., 1980 [42]
JEG-3	Choriocarcinoma: placenta	Kohler et al., 1971 [43]

To determine whether E251K is a common SNP in the Japanese population, we sequenced 94 samples from Japanese healthy volunteers using a pyrosequencing method. Similarly, sequencing data were obtained from 95 volunteers for M129T, which we previously found in a germinoma and its normal tissue from a patient [24]. In both cases, all the sequenced samples were homozygous: GAG (251E) and ATG (129M). None were heterozygous, which suggests that the two amino acid substitutions, E251K and M129T, are not common SNPs in the Japanese population.

Three synonymous SNPs were also found at Asp81, Ala284, and Lys352 (Table 2). In the noncoding region, a SNP of G to C in 5′-UTR, 17 bases upstream of the translation initiation site, downstream of a putative ETS-1-binding site (base −26 to −22) [30]), was found (rs907187). NEC15 had a nonheterozygous C allele, and ITOII and Tera-1 had heterozygous G/C alleles. We also noted that the nonheterozygous allele at a SNP in intron 2 (rs1805405) was observed at higher frequency (3/9) than a heterozygous allele (1/9) in germ cell tumor cell lines, but the difference was not statistically significant. NEC8, NEC15, and ITOII had the nonheterozygous A allele, whereas NCCIT had heterozygous C/A alleles.

3.2. Effect of amino acid alteration on the activity of PARP-1

The effect of amino acid substitution of methionine to threonine at codon 129 (Met129Thr) and glutamic acid to lysine at codon 251 (Glu251Lys) on PARP-1 enzymatic activity was examined. We transiently expressed the mutated PARP-1 harboring either M129T or E251K substitution in the *Parp-1*$^{-/-}$ MEFs, and measured the enzymatic activity of PARP-1 by the activity gel method. Both *Parp-1* mutants showed a decrease in PARP-1 activity relative to the wild type, although the difference was not

Table 2
Sequence alterations and SNPs found in the *PARP1* gene in human germ cell tumor cell lines

	Exon	Nucleotide[a]	No.	Germ cell tumor cell line[b]	SNP ID[c]	Heterozygosity[c]
				Amino acid substitution		
Phe54Leu	2	TTC	8	NEC8, NEC14, ITOII, Tera-1, Tera-2, NCCIT, PA-1, JEG-3	rs3738708	0.023
		TTC/TTG	0			
		TTG	1	NEC15		
Glu251Lys	6	GAG	8	NEC14, NEC15, ITOII, Tera-1, Tera-2, NCCIT, PA-1, JEG-3	not listed	not listed
		GAG/AAG	0			
		AAG	1	NEC8		
Val762Ala	17	GTG	7	NEC8, ITOII, Tera-1, Tera-2, NCCIT, PA-1, JEG-3	rs1136410	0.351
		GTG/GCG	0			
		GCG	2	NEC14, NEC15		
				SNPs without amino acid substitution		
Asp81Asp	2	GAC	5	NEC8, Tera-2, NCCIT, PA-1, JEG-3	rs1805404	0.372
		GAC/GAT	2	ITOII, Tera-1		
		GAT	2	NEC14, NEC15		
Ala284Ala	7	GCT	5	NEC15, ITOII, Tera-2, PA-1, JEG-3	rs1805414	0.498
		GCT/GCC	2	NCCIT, Tera-1		
		GCC	2	NEC8, NEC14		
Lys352Lys	8	AAA	8	NEC14, NEC15, ITOII, Tera-1, Tera-2, NCCIT, PA-1, JEG-3	rs1805415	0.379
		AAA/AAG	0			
		AAG	1	NEC8		
				SNPs in noncoding region		
5′-UTR (−17 bp)		G	6	NEC8, NEC14, Tera-2, NCCIT, PA-1, JEG-3	rs907187	0.357
		G/C	2	ITOII, Tera-1		
		C	1	NEC15		
Intron 2 (5592 bp)		C	5	NEC14, Tera-1, Tera-2, PA-1, JEG-3	rs1805405	0.362
		C/A	1	NCCIT		
		A	3	NEC8, NEC15, ITOII		

Abbreviations: SNP, single nucleotide polymorphism; UTR, untranslated region.
[a] The altered nucleotide is indicated by underscoring.
[b] See Table 1 for cell lines.
[c] SNP identifiers and average estimated heterozygosity data are from the NCBI database of SNPs, available at http://www.ncbi.nlm.nih.gov/SNP/index.html.

statistically significant ($P = 0.1266$ for 129T, $P = 0.2752$ for 251K) (Fig. 2A).

Cellular localization of E251K and M129T mutants was also analyzed as GFP-fusion protein expressed in the *Parp-1*$^{-/-}$ MEF. The cellular localization of another *PARP1* mutant harboring K940R amino acid substitution was analyzed as well [24]. Wild-type, 129T, 251K, and 940R protein localized exclusively in the nuclei (Fig. 2B). The localization is observed in a punctuated manner in the nuclei, and the localization pattern did not differ between wild-type and the mutants.

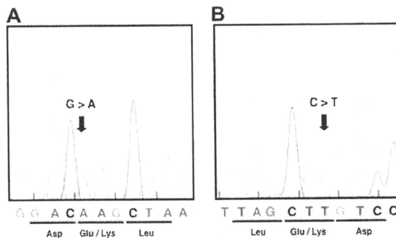

Fig. 1. Electropherograms of sequences surrounding codon 251 of the *PARP1* gene in NEC8 cells. Sequence alterations of both strands were confirmed by sequencing using the sense (A) and anti-sense (B) primers.

3.3. Levels of PARP-1 protein, activity, and poly(ADP-ribosyl)ation

The levels of PARP-1 proteins were measured by Western blot analysis (Fig. 3). PA-1 and NEC8 cells both showed lower levels of PARP-1 proteins. We also examined the enzymatic activity of PARP-1 as PARP-1 auto-poly(ADP-ribosyl)ation activity using whole-cell extracts (Fig. 3, activity gel analysis). PA-1 and NEC8 cells both showed lower levels of auto-poly(ADP-ribosyl)ation activity, compared with the other cell lines analyzed. PA-1, NEC8, and also Tera-2 exhibited lower levels of overall poly(ADP-ribosyl)ation of proteins, compared with the other cell lines.

4. Discussion

Among the germ cell tumor cell lines analyzed in this study, five cell lines were established in Japan, and these

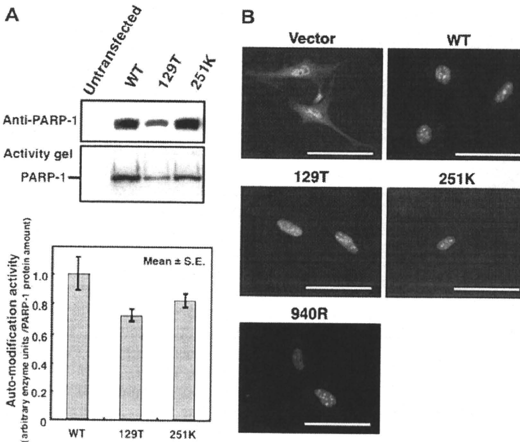

Fig. 2. Effects of amino acid alteration in PARP-1 on enzymatic activity and subcellular localization. (A) Semiquantitative analysis of auto-poly(ADP-ribosyl)ation activity of wild-type and mutant PARP-1 proteins in the whole cell extracts from the *Parp-1*$^{-/-}$ mouse embryonic fibroblasts transfected with either the wild-type (WT) or mutant PARP-1 expression plasmid. The representative result of the activity gel analysis (middle panel) and Western blot analysis of PARP-1 (top panel) is shown. The auto-poly(ADP-ribosyl)ation activity normalized to the expressed PARP-1 level measured by Western blot analysis. A linear relationship between the amount of PARP-1 and auto-poly(ADP-ribosyl)ation activity was confirmed (data not shown). (B) Subcellular localization of GFP-fused PARP-1. At 24 hours after transfection, localization of wild-type PARP-1 and mutant proteins harboring either 129T, 251K, or 940R amino acid substitution was observed exclusively in the nuclei. Scale bars: 10 μm.

may reflect a spectrum of polymorphisms in the Japanese population. We found a nonheterozygous sequence alteration (GAG to AAG) that causes amino acid substitution of E251K of PARP-1 in NEC8. Because the corresponding normal tissue samples were not available for the NEC8 cell line, we could not examine whether E251K is a SNP. It is not known whether the nonheterozygosity represents a homozygosity or an allelic loss, nor could we exclude the possibility that this sequence alteration was introduced during or after establishment of NEC8. We previously reported that a heterozygous sequence alteration that causes amino acid alteration of M129T was observed in a human germ cell tumor specimen [24]. Our analysis in the present study suggests that the sequence alterations of E251K and M129T are not common SNPs in the Japanese population.

E251K is located at a peptide stretch conserved among species in the C-terminus of a DNA binding domain close to the nuclear localization signal. The third zinc-binding motif (codon 295–321) [28,31] is present close to codon 251. The third zinc-binding domain of codon 216–366 is required for dimerization of PARP-1. K249E substitution is reported to decrease PARP-1 enzymatic activity [32]. A decrease in auto-poly(ADP-ribosyl)ation activity but no alteration in nuclear localization of PARP-1 harboring M129T or E251K amino acid substitution was observed; however, we noted that NEC8, which has a nonheterozyous E251K allele, had a lower level of PARP-1 protein in the extract (Fig. 3). The effects on DNA binding or DNA repair regulation, as well as stability, should be further analyzed.

Within nine cell lines, both nonheterozygous minor alleles of V762A and intron 2 (5,592 bp) (SNP ID, rs1805405) showed a tendency of higher frequencies than expected, although it was not statistically significant. These tendencies are similar to the result obtained with our previous study using 16 germ cell tumor specimens [24]. The V762A SNP was found to be associated with the risk of prostate cancers in European-origin subjects, in whom the A/A genotype showed a twofold increase in susceptibility [11]. Recently the *PARP1* V762A polymorphism has been reported to reduce the enzymatic activity of PARP-1 and the ability of interaction with *XRCC1* [15,33]; however, a decrease in PARP-1 auto-poly(ADP-ribosyl)ation activity and overall poly(ADP-ribosyl)ation levels was not observed in NEC14, which harbors a nonheterozygous V762A allele (Table 2 and Fig. 3).

The nonheterozygous allele of SNP rs1805405, located within the polypyrimidine tract close to the 3′ splice acceptor site in the intron 2, was observed at a higher frequency (3/9) than expected, as in the case with human germ cell tumor specimens (3/16) [24]. A stretch of (C/T)$_6$NCAGG(C/T) at a splicing acceptor site is relatively conserved in the introns [34]. The polypyrimidine tract is

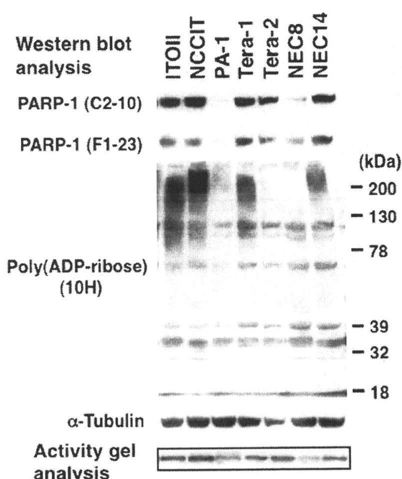

Fig. 3. PARP-1 protein levels and activity, with overall poly(ADP-ribosyl)ation protein levels. Western blot analysis was performed for detection of PARP-1 protein and poly(ADP-ribosyl)ated proteins (upper and middle panels). For detection of PARP-1 enzymatic activity, activity gel analysis was performed (bottom panel) to analyze PARP-1 auto-poly(ADP-ribosyl)ation activity; 30 μg of cell extracts was subjected to 6% SDS-PAGE in the presence of activated DNA in the gel.

highly conserved, and the frequency of adenine at this position is low (~7%) [35]. In the case of all introns of *PARP1* reference sequences (NCBI numbers NT_004559 and NT_167186), the frequency of adenine at this position is ~9%. In the case of hereditary nonpolyposis colorectal cancer proband, a single base-pair T-to-A transversion at position −11 of the *MLH1* gene intron 1 splice acceptor site caused exon 2 skipping [36]. We detected only a full-length transcript of *PARP1* in the cell lines harboring minor alleles of the SNP at intron 2 by Northern blot analysis (data not shown). Whether the SNP at intron 2 of *PARP1* affects the splicing efficiency of exon 3 needs further investigation.

It is noteworthy that the poly(ADP-ribosyl)ation level is lower in NEC8, PA-1, and Tera-2, compared with other germ cell tumor cell lines. The lower poly(ADP-ribosyl)ation level in NEC8 and PA-1 could be explained by the lower PARP-1 level, but that is not the case for Tera-2. Activites of other PARP family proteins or of PARG, a major poly(ADP-ribose) degradation enzyme, may also affect the poly(ADP-ribosyl)ation level in Tera-2. We did not find any SNPs or other base alterations in the 3′UTR of the *PARP1* gene in nine cell lines. Therefore, the mechanism for lower levels of PARP-1 in NEC8 and PA-1 may not be due to altered posttranscriptional regulation of PARP-1.

In this study, we identified sequence alterations, including SNPs, in the *PARP1* gene of human germ cell tumor cell lines. The nonheterozygous minor alleles of SNPs at V762A and intron 2 showed a slightly higher frequency. Differences in the levels of PARP-1 and poly(ADP-ribosyl)ation were observed. Because poly(ADP-ribosyl)ation reaction is involved in several physiological processes in cancer cells, including DNA repair and differentiation, the alteration of PARP-1 activity may affect the development of cancers through multiple processes. It is also suggested that PARP family proteins, including PARP-2, may complement PARP-1 functions [37]. It may therefore be necessary to examine aberrations of PARP-1 and other PARP family proteins in cancers not only at gene expression levels but also at protein or enzymatic activity levels. Clinical trials are ongoing with PARP inhibitors in combination with chemotherapeutic agents [38]. Understanding the mechanisms of functional regulation of PARP-1 in cancer cells is important. Because the activity of PARP-1 and other PARP family members is important in DNA repair and cell death induction, the levels of PARP-1 and poly(ADP-ribosyl)ation activity may also substantially affect the outcome of cancer therapies that target DNA.

Acknowledgments

We thank T. Otsubo, T. Okada, and S. Mimaki for technical help and suggestions. We are grateful for the kind support and suggestions of H. Nakagama. This work was supported in part by a Grant-in-Aid for the Second and Third Term Comprehensive 10-year Strategy for Cancer Control and a Grant-in-Aid for Cancer Research from the Ministry of Health, Labor, and Welfare of Japan.

References

[1] Nozaki T, Fujihara H, Watanabe M, Tsutsumi M, Nakamoto K, Kusuoka O, Kamada N, Suzuki H, Nakagama H, Sugimura T, Masutani M. Parp-1 deficiency implicated in colon and liver tumorigenesis induced by azoxymethane. Cancer Sci 2003;94:497–500.
[2] Tsutsumi M, Masutani M, Nozaki T, Kusuoka O, Tsujiuchi T, Nakagama H, Suzuki H, Konishi Y, Sugimura T. Increased susceptibility of poly(ADP-ribose) polymerase-1 knockout mice to nitrosamine carcinogenicity. Carcinogenesis 2001;22:1–3.
[3] Tong WM, Cortes U, Hande MP, Ohgaki H, Cavalli LR, Lansdorp PM, Haddad BR, Wang ZQ. Synergistic role of Ku80 and poly(ADP-ribose) polymerase in suppressing chromosomal aberrations and liver cancer formation. Cancer Res 2002;62:6990–6.
[4] Masutani M, Nakagama H, Sugimura T. Poly(ADP-ribosyl)ation in relation to cancer and autoimmune disease. Cell Mol Life Sci 2005;62:769–83.
[5] Prasad SC, Thraves PJ, Bhatia KG, Smulson ME, Dritschilo A. Enhanced poly(adenosine diphosphate ribose) polymerase activity and gene expression in Ewing's sarcoma cells. Cancer Res 1990; 50:38–43.
[6] Masutani M, Nozaki T, Sasaki H, Yamada T, Kohno T, Shimizu K, Gotoh M, Shiraishi M, Yokota J, Hirohashi S, Nakagama H, Sugimura T. Poly(ADP-ribose) polymerase-1 gene in human tumor

cell lines: its expression and structural alteration. Proc Jpn Acad Ser B Phys Biol Sci 2004;80B:114–8.

[7] Menegazzi M, Scarpa A, Carcereri de Prati A, Menestrina F, Suzuki H. Correlation of poly(ADP-ribose)polymerase and p53 expression levels in high-grade lymphomas. Mol Carcinog 1999; 25:256–61.

[8] Idogawa M, Yamada T, Honda K, Sato S, Imai K, Hirohashi S. Poly(ADP-ribose) polymerase-1 is a component of the oncogenic T-cell factor-4/β-catenin complex. Gastroenterology 2005;128: 1919–36.

[9] Ghabreau L, Roux JP, Frappart PO, Mathevet P, Patricot LM, Mokni M, Korbi S, Wang ZQ, Tong WM, Frappart L. Poly(ADP-ribose) polymerase-1, a novel partner of progesterone receptors in endometrial cancer and its precursors. Int J Cancer 2004;109:317–21.

[10] Bièche I, de Murcia G, Lidereau R. Poly(ADP-ribose) polymerase gene expression status and genomic instability in human breast cancer. Clin Cancer Res 1996;2:1163–7.

[11] Lockett KL, Hall MC, Xu J, Zheng SL, Berwick M, Chuang SC, Clark PE, Cramer SD, Lohman K, Hu JJ. The *ADPRT* V762A genetic variant contributes to prostate cancer susceptibility and deficient enzyme function. Cancer Res 2004;64:6344–8.

[12] Zhang X, Miao X, Liang G, Hao B, Wang Y, Tan W, Li Y, Guo Y, He F, Wei Q, Lin D. Polymorphisms in DNA base excision repair genes *ADPRT* and *XRCC1* and risk of lung cancer. Cancer Res 2005;65:722–6.

[13] Hao B, Wang H, Zhou K, Li Y, Chen X, Zhou G, Zhu Y, Miao X, Tan W, Wei Q, Lin D, He F. Identification of genetic variants in base excision repair pathway and their associations with risk of esophageal squamous cell carcinoma. Cancer Res 2004;64:4378–84.

[14] Zhang Z, Miao XP, Tan W, Guo YL, Zhang XM, Lin DX. Correlation of genetic polymorphisms in DNA repair genes *ADPRT* and *XRCC1* to risk of gastric cancer [In Chinese]. Ai Zheng 2006;25:7–10.

[15] Miao X, Zhang X, Zhang L, Guo Y, Hao B, Tan W, He F, Lin D. Adenosine diphosphate ribosyl transferase and x-ray repair cross-complementing 1 polymorphisms in gastric cardia cancer. Gastroenterology 2006;131:420–7.

[16] Ogino H, Nozaki T, Gunji A, Maeda M, Suzuki H, Ohta T, Murakami Y, Nakagama H, Sugimura T, Masutani M. Loss of *Parp-1* affects gene expression profile in a genome-wide manner in ES cells and liver cells. BMC Genomics 2007;8:41.

[17] Krishnakumar R, Gamble MJ, Frizzell KM, Berrocal JG, Kininis M, Kraus WL. Reciprocal binding of PARP-1 and histone H1 at promoters specifies transcriptional outcomes. Science 2008;319: 819–21.

[18] Quenet D, Gasser V, Fouillen L, Cammas F, Sanglier-Cianferani S, Losson R, Dantzer F. The histone subcode: poly(ADP-ribose) polymerase-1 (Parp-1) and Parp-2 control cell differentiation by regulating the transcriptional intermediary factor TIF1β and the heterochromatin protein HP1α. FASEB J 2008;22:3853–65.

[19] Nozaki T, Masutani M, Watanabe M, Ochiya T, Hasegawa F, Nakagama H, Suzuki H, Sugimura T. Syncytiotrophoblastic giant cells in teratocarcinoma-like tumors derived from *Parp*-disrupted mouse embryonic stem cells. Proc Natl Acad Sci USA 1999;96: 13345–50.

[20] Hemberger M, Nozaki T, Winterhager E, Yamamoto H, Nakagama H, Kamada N, Suzuki H, Ohta T, Ohki M, Masutani M, Cross JC. Parp1-deficiency induces differentiation of ES cells into trophoblast derivatives. Dev Biol 2003;257:371–81.

[21] von Hochstetter AR, Sigg C, Saremaslani P, Hedinger C. The significance of giant cells in human testicular seminomas: a clinico-pathological study. Virchows Arch A Pathol Anat Histopathol 1985;407:309–22.

[22] Masutani M, Nozaki T, Watanabe M, Ochiya T, Hasegawa F, Nakagama H, Suzuki H, Sugimura T. Involvement of poly(ADP-ribose) polymerase in trophoblastic cell differentiation during tumorigenesis. Mutat Res 2001;477:111–7.

[23] Ohashi Y, Ueda K, Hayaishi O, Ikai K, Niwa O. Induction of murine teratocarcinoma cell differentiation by suppression of poly(ADP-ribose) synthesis. Proc Natl Acad Sci USA 1984;81: 7132–6.

[24] Shiokawa M, Masutani M, Fujihara H, Ueki K, Nishikawa R, Sugimura T, Kubo H, Nakagama H. Genetic alteration of poly(ADP-ribose) polymerase-1 in human germ cell tumors. Jpn J Clin Oncol 2005;35:97–102.

[25] Nozaki T, Fujihara H, Kamada N, Ueda O, Takato T, Nakagama H, Sugimura T, Suzuki H, Masutani M. Hyperploidy of embryonic fibroblasts derived from *Parp-1* knockout mouse. Proc Jpn Acad Ser B Phys Biol Sci 2001;77:121–4.

[26] Nakayama R, Sato Y, Masutani M, Ogino H, Nakatani F, Chuman H, Beppu Y, Morioka H, Yabe H, Hirose H, Sugimura H, Sakamoto H, Ohta T, Toyama Y, Yoshida T, Kawai A. Association of a missense single nucleotide polymorphism, Cys1367Arg of the *WRN* gene, with the risk of bone and soft tissue sarcomas in Japan. Cancer Sci 2008; 99:333–9.

[27] Uchida K, Morita T, Sato T, Ogura T, Yamashita R, Noguchi S, Suzuki H, Nyunoya H, Miwa M, Sugimura T. Nucleotide sequence of a full-length cDNA for human fibroblast poly(ADP-ribose) polymerase. Biochem Biophys Res Commun 1987;148:617–22.

[28] Masutani M, Nozaki T, Hitomi Y, Ikejima M, Nagasaki K, de Prati AC, Kurata S, Natori S, Sugimura T, Esumi H. Cloning and functional expression of poly(ADP-ribose) polymerase cDNA from *Sarcophaga peregrina*. Eur J Biochem 1994;220:607–14.

[29] Kawamitsu H, Hoshino H, Okada H, Miwa M, Momoi H, Sugimura T. Monoclonal antibodies to poly(adenosine diphosphate ribose) recognize different structures. Biochemistry 1984;23: 3771–7.

[30] Soldatenkov VA, Albor A, Patel BK, Dreszer R, Dritschilo A, Notario V. Regulation of the human poly(ADP-ribose) polymerase promoter by the ETS transcription factor. Oncogene 1999;18: 3954–62.

[31] Langelier MF, Servent KM, Rogers EE, Pascal JM. A third zinc-binding domain of human poly(ADP-ribose) polymerase-1 coordinates DNA-dependent enzyme activation [Erratum in: J Biol Chem 2008;283:22884]. J Biol Chem 2008;283:4105–14.

[32] Trucco C, Flatter E, Fribourg S, de Murcia G, Ménissier-de Murcia J. Mutations in the amino-terminal domain of the human poly(ADP-ribose) polymerase that affect its catalytic activity but not its DNA binding capacity. FEBS Lett 1996;399:313–6.

[33] Wang XG, Wang ZQ, Tong WM, Shen Y. *PARP1* Val762Ala polymorphism reduces enzymatic activity. Biochem Biophys Res Commun 2007;354:122–6.

[34] Nakata K, Kanehisa M, DeLisi C. Prediction of splice junctions in mRNA sequences. Nucleic Acids Res 1985;13:5327–40.

[35] Stephens RM, Schneider TD. Features of spliceosome evolution and function inferred from an analysis of the information at human splice sites. J Mol Biol 1992;228:1124–36.

[36] Clarke LA, Veiga I, Isidro G, Jordan P, Ramos JS, Castedo S, Boavida MG. Pathological exon skipping in an HNPCC proband with *MLH1* splice acceptor site mutation. Genes Chromosomes Cancer 2000;29:367–70.

[37] Miwa M, Masutani M. PolyADP-ribosylation and cancer. Cancer Sci 2007;98:1528–35.

[38] Ratnam K, Low JA. Current development of clinical inhibitors of poly(ADP-ribose) polymerase in oncology. Clin Cancer Res 2007; 13:1383–8.

[39] Motoyama T, Watanabe H, Yamamoto T, Sekiguchi M. Human testicular germ cell tumors in vitro and in athymic nude mice. Acta Pathol Jpn 1987;37:431–48.

[40] Fogh J. Cultivation, characterization, and identification of human tumor cells with emphasis on kidney, testis, and bladder tumors. Natl Cancer Inst Monogr 1978;49:5–9.

[41] Teshima S, Shimosato Y, Hirohashi S, Tome Y, Hayashi I, Kanazawa H, Kakizoe T. Four new human germ cell tumor cell lines. Lab Invest 1988;59:328–36.

[42] Zeuthen J, Nørgaard JO, Avner P, Fellous M, Wartiovaara J, Vaheri A, Rosén A, Giovanella BC. Characterization of a human ovarian teratocarcinoma-derived cell line. Int J Cancer 1980;25: 19–32.

[43] Kohler PO, Bridson WE. Isolation of hormone-producing clonal lines of human choriocarcinoma. J Clin Endocrinol Metab 1971;32: 683–7.

Contribution of the *TP53*, *OGG1*, *CHRNA3*, and *HLA-DQA1* Genes to the Risk for Lung Squamous Cell Carcinoma

Takashi Kohno, PhD, Hideo Kunitoh, MD,† Sachiyo Mimaki, MSc,‡ Kouya Shiraishi, PhD,* Aya Kuchiba, PhD,‡ Seiichiro Yamamoto, PhD,§ and Jun Yokota, MD**

Introduction: Recent genome-wide association studies (GWASs) have identified polymorphisms in several genes associated with lung cancer risk. Nevertheless, functional polymorphisms in DNA repair and metabolic genes that had been reported as being associated with risk for lung cancer, particularly for lung squamous cell carcinoma (SQC), were not examined in those studies. Therefore, significance of these functional polymorphisms was evaluated in a population, in which polymorphisms in the GWAS genes showed associations with lung SQC risk.

Methods: Polymorphisms in three DNA repair genes, *TP53*, *MDM2*, and *OGG1*, and two metabolic genes, *CYP1A1* and *GSTM1*, were examined for associations with lung SQC risk in a hospital-based case-control study consisting of 377 cases and 325 controls, which had been previously subjected to association studies on GWAS genes, *CHRNA3*, *TERT*, and *HLA-DQA1*.

Results: Genotypes for two DNA repair genes, *TP53* and *OGG1*, showed significant associations with SQC risk ($p < 0.05$), and those for two GWAS genes, *CHRNA3* and *HLA-DQA1*, showed significant associations with SQC risk ($P < 0.05$) with odds ratios between 1.65 (95% confidence interval = 1.06–2.57 for *OGG1*) and 2.57 (95% confidence interval = 1.03–6.87 for *CHRNA3*). Marginally significant associations were also observed for *MDM2* and *CYP1A1* genes. Interactions among these polymorphisms on SQC risk were not observed.

Conclusions: Association of functional polymorphisms in DNA repair and metabolic genes with lung SQC risk was appreciated. This result indicates the necessity of reevaluation for the significance of functional polymorphisms in DNA repair and metabolic genes on lung cancer risk in other populations subjected to GWASs.

Key Words: Genome-wide association study, Single-nucleotide polymorphism, Lung squamous cell carcinoma, DNA repair gene, Metabolic gene.

(*J Thorac Oncol.* 2011;6: 813–817)

Several genome-wide association studies (GWASs) on single-nucleotide polymorphisms (SNPs) have led to the identification of three loci, chromosomes 15q25, 5p13, and 6p21, containing SNPs associated with lung cancer risk in Europeans and Americans.[1–3] Associations with lung cancer risk of a SNP in the *CHRNA3* gene at 15q25 encoding a nicotinic acetylcholine receptor subunit and a SNP in the *TERT* gene at 5p15 encoding a telomerase reverse transcription were also validated in Asians.[4,5] Significance of SNPs at 6p21 on lung cancer risk of Asians have not been investigated fully; however, our GWAS on Japanese indicated *HLA-DQA1*, encoding a human leukocyte antigen class II protein, as a responsible locus at 6p21.[5] A recent combined study on multiple GWASs in Europeans and Americans indicated that major common genes conferring lung cancer risk have been identified already.[6]

Studies on DNA adducts/damages, including those produced by tobacco carcinogens and their repair processes led to identification of various metabolic and DNA repair genes carrying functional polymorphisms, which possibly cause interindividual differences in the rate of somatic mutation and lung cancer susceptibility.[7] *CYP1A1* and *GSTM1* are representative, because their polymorphisms have been reported to be associated with risk for lung cancer of Asians, particularly for squamous cell carcinoma (SQC), a major histological type of lung cancer mostly developed in cigarette smokers.[8] The risk (462Val) allele for the Ile462Val SNP in the *CYP1A1* gene encodes a protein with a higher activity to bioactivate polycyclic aromatic hydrocarbons, major tobacco procarcinogens, than the 462Ile allele.[7] The risk (absence) allele for the presence or absence polymorphism in the *GSTM1* gene does not encode GSTM1 protein to detoxify activated polycyclic aromatic hydrocarbon-intermediates.[7] *TP53*, *MDM2*, and *OGG1* are representative DNA repair genes associated with lung cancer risk in Asians.[9–11] The risk (72Pro) allele for the TP53-Arg72Pro SNP in the *TP53* gene encodes a protein with a weaker apoptosis activity allowing survival of DNA-damaged cells than the 72Arg allele.[7] The risk (G) allele for a T/G SNP in the promoter region of the *MDM2* gene (called

*Biology Division, National Cancer Center Research Institute; †Division of Thoracic Oncology, National Cancer Center Hospital; ‡Genetics Division, National Cancer Center Research Institute; and §Statistics and Cancer Control Division, Research Center for Cancer Prevention and Screening, National Cancer Center, Tokyo, Japan.

Disclosure: The authors declare no conflicts of interest.

Address for correspondence: Jun Yokota, MD, Biology Division, National Cancer Center Research Institute, 1-1 Tsukiji 5-chome, Chuo-ku, Tokyo 104-0045, Japan. E-mail: jyokota@ncc.go.jp

Hideo Kunitoh is currently at Department of Respiratory Medicine, Mitsui Memorial Hospital, Tokyo, Japan.

K.S. is an awardee of a Research Resident Fellowship from the Foundation for Promotion of Cancer Research in Japan.

Copyright © 2011 by the International Association for the Study of Lung Cancer

ISSN: 1556-0864/11/0604-0813

Copyright © 2011 by the International Association for the Study of Lung Cancer.

SNP309) allows a lower level of expression of MDM2 protein to suppress *TP53* function than the T allele.[10] The risk (326Cys) allele for the Ser326Cys SNP in the *OGG1* gene encodes a DNA glycosylase with a weaker activity to repair 8-hydroxyguanine, an oxidatively damaged promutagenic base, in part, produced by tobacco carcinogens than the 326Ser allele.[7]

Nevertheless, associations of these functional polymorphisms were not investigated in the GWASs above[1–3] because of the lack of probes to discriminate these polymorphisms in the platforms used for GWASs (http://www.ncbi.nlm.nih.gov/snp). Therefore, significance of functional polymorphisms on lung SQC risk and their interactions with GWAS genes in the risk is still unknown. In this study, polymorphisms identified by GWASs and those functional polymorphisms were investigated together for their significance in the risk for lung SQC in the same population by conducting a hospital-based case-control study.

PATIENTS AND METHODS

Case-Control Study

All cases and controls were Japanese and were admissions to National Cancer Center Hospital from 1999 to 2007. Cases were individuals diagnosed with lung SQC by cytological and histological examinations according to World Health Organization classification,[8] whereas controls were those without lung and other cancers and with no history of cancers. All cases and controls, from whom informed consent and blood samples were obtained, were consecutively included in this study without any exclusion criteria. The participation rate was approximately 80%. From each individual, a 20 ml whole-blood sample was obtained. Information on these subjects has been described elsewhere.[5,12] This study was approved by the institutional review boards of the National Cancer Center. Smoking habit was expressed by pack-years, which was defined as the number of cigarette packs smoked daily multiplied by years of smoking. Nonsmokers were defined as individuals of the pack-years = 0. Smokers were defined as individuals of pack-years more than 0 and included both former and current smokers.

Genotyping for Polymorphisms

Genomic DNA from whole blood cells was subjected to genotyping. All 377 cases and 325 controls were previously genotyped for the *CHRNA3*, *TERT*, and *HLA-DQA1* genes.[5] One hundred eighty-eight cases and 203 controls had been genotyped previously for two SNPs, TP53-Arg72Pro and OGG1-Ser326Cys[12]; therefore, the remaining 189 cases and 122 controls were newly genotyped in this study by pyrosequencing as described.[12] All cases and controls were genotyped for the *CYP1A1*, *GSTM1*, and *MDM2* genes. Genotypes for the *CYP1A1* gene were determined by pyrosequencing using primers 5′-GCCTGTCCTCTATCCTTTGGG-3′ and 5′-GCTCCATCAGCATCTATGTGGC-3′ for polymerase chain reaction (PCR) amplification and 5′-GGAAGTGTATCGGTGAGACC-3′ for sequencing, whereas those for the *MDM2* gene were determined by pyrosequencing using primers 5′-GGGAGTTCAGGGTAAAGGTC-3′ and 5′-CACACTAGTGACCCGACAGG-3′ for PCR amplification and 5′-CTGCGGGGCCGCT-3′ for sequencing. The GSTM1-presence or absence polymorphism was analyzed by multiplex PCR.[13]

Statistical Analysis

The strength of association of genotypes with risk was measured as odds ratios (ORs) adjusted for gender, age (≥49, 50–59, 60–69, and ≤70 years), and smoking dosage (0, 1–50, and ≤50 pack-years) with 95% confidence intervals by unconditional logistic regression analysis.[14] Adjustment of smoking was performed with lifetime pack-years for former smokers. ORs for each genotype and those in dominant (i.e., AA versus Aa + aa; A: major allele, a: minor allele) or recessive (i.e., AA + Aa versus aa) modes were calculated. To find pairs of SNPs that interact with each other, associations between genotypes for two polymorphisms were examined in the case subjects. The statistical analyses were performed using the JMP version 6.0 software (SAS Institute Inc., Cary, NC). A level of p less than 0.05 was considered as statistically significant, whereas a level of $0.05 \leq p \leq 0.10$ was considered as marginally significant.

RESULTS AND DISCUSSION

We conducted a case-control study consisting of 377 SQC cases and 325 controls enrolled in the National Cancer Center Hospital, Tokyo, Japan (Table 1). Smoking rates of cases and controls were similar to the ones in a previous cohort study of a Japanese population,[15] and therefore, the cases showed a higher fraction of smokers than the controls. Polymorphisms for two genes identified by GWASs, *CHRNA3* and *HLA-DQA1*, showed significant associations with SQC risk, but a SNP in the *TERT* gene did not show significant associations in our previous study[5] (Table 2). Association of the *TERT* SNP with risk not for SQC but for adenocarcinoma, another major histological type of lung cancer, was also shown by other studies.[3,4]

All the cases and controls were examined for associations with SQC risk of polymorphisms in three DNA repair genes, *TP53*, *MDM*, and *OGG1*; and two metabolic genes, *CYP1A1* and *GSTM1* (Table 2). Significant increases in ORs were observed in one or more genotypes or modes for three

TABLE 1. Subjects for Case-Control Study

		Gender (%)			Smoking (%)	
Subject	No.	Male	Female	Age (Mean ± SD)	Nonsmoker	Smoker
Case	377	340 (90)	37 (10)	62.7 ± 7.6	13 (3)	364 (97)
Control	325	185 (57)	140 (43)	62.5 ± 11.3	179 (55)	146 (45)

Copyright © 2011 by the International Association for the Study of Lung Cancer

Copyright © 2011 by the International Association for the Study of Lung Cancer.

TABLE 3. Combined Genotypes for the *TP53* and *MDM2* Genes and Lung SQC Risk

TP53 Genotype	*MDM2* Genotype	Control (%)	Case (%)	OR[a]	(95% CI, *p* Value)	*p* for Interaction[a]
Arg/Arg	T/T	41 (13)	27 (7)	Reference		1.00
Arg/Arg	T/G + G/G	111 (34)	100 (27)	1.48	(0.72–3.10, 0.29)	
Arg/Pro + Pro/Pro	T/T	37 (11)	41 (11)	1.68	(0.68–4.22, 0.26)	
Arg/Pro + Pro/Pro	T/G + G/G	136 (42)	209 (55)	2.53	(1.27–5.11, 0.0081)	

[a] Adjusted for sex, age (≥49, 50–59, 60–69, ≤70 yr), and smoking dosage (0, 1–50, ≤50 pack-years).
SQC, squamous cell carcinoma; OR, odds ratio; CI, confidence interval.

We next examined associations between genotypes for every six pairs of four genes, *CHRNA3*, *HLA-DQA1*, *TP53*, and *OGG1*, whose significant associations with lung SQC risk were observed, in the case subjects to find pairs of polymorphisms that interact with each other. Interaction between these four genes and two genes with marginally significant associations was also examined. In this analysis, four polymorphisms, CHRNA3-C/T, TP53-Arg72Pro, MDM2-T/G, and CYP1A1-Ile462Val, were examined in the dominant mode, whereas the remaining two, HLA-DQA1*03 and OGG1-Ser326Cys, were examined in the recessive mode, based on the results of association with SQC risk (Table 2). Significant or marginally significant association was not observed for any pairs of polymorphisms ($p > 0.1$; Supplementary Table 1, http://links.lww.com/JTO/A33). Previously, polymorphisms of the *MDM2* and *TP53* genes were suggested to interact with each other, because they cooperate in apoptosis.[10] Nevertheless, the ORs of individuals with risk genotypes both for *MDM2* and *TP53* (2.53) were similar to the OR deduced from ORs of individuals with a single-risk genotype for *MDM2* and *TP53* (1.48 [for *MDM2*] × 1.68 [for *TP53*] = 2.49; Table 3). Therefore, *TP53* and *MDM2* would not interactively contribute to SQC risk as reported.[16] Therefore, these genes were suggested to independently contribute to SQC risk, although it remains possible that interaction between polymorphisms were overlooked because of the small number of study subjects.

In this study, association of a few functional polymorphisms in DNA repair and metabolic genes with lung SQC risk was suggested in a population, in which polymorphisms in the GWAS genes showed associations. Therefore, significance of those functional polymorphisms as genetic factors for lung SQC risk was reindicated. Nevertheless, our study has limitations. The present results were obtained from a hospital-based case-control study with a small number of case and control subjects, and frequencies of critical factors involved in lung SQC risk, that is, gender and smoking, are largely different between the cases and controls. Therefore, the statistical power for detection of association in this study was lower than that in a study using a larger number of case and control subjects and was also lower than that using control subjects that were matched to case subjects by gender and smoking. Therefore, associations of some polymorphisms might have been overlooked because of the lack of statistical power. In addition, associations of polymorphisms with lung SQC risk could be confounded by gender or smoking because of insufficient adjustments in statistical tests. Nevertheless, ORs for risk genotypes only in males or smokers were similar to those in all subjects (Supplementary Table 2, http://links.lww.com/JTO/A33). Therefore, it is unlikely that the associations observed in this study were confounded by these factors. On the basis of these results, we would like to emphasize the importance of analyzing various functional polymorphisms together with millions of GWAS marker SNPs. It will be a powerful way for this purpose to analyze those polymorphisms in the populations that were subjected to recent GWASs.

ACKNOWLEDGMENTS

Supported by Grants-in-Aid from the Ministry of Health, Labor, and Welfare for the third-term Comprehensive 10-year Strategy for Cancer Control and for Cancer Research, 19S-1 and 19-9, and the Ministry of Education, Culture, Sports, Science and Technology for Scientific Research on Priority Areas, KAKENHI 20014031.

The authors thank Dr. Daizo Saito for collection of control subjects.

REFERENCES

1. McKay JD, Hung RJ, Gaborieau V, et al. Lung cancer susceptibility locus at 5p15.33. *Nat Genet* 2008;40:1404–1406.
2. Wang Y, Broderick P, Webb E, et al. Common 5p15.33 and 6p21.33 variants influence lung cancer risk. *Nat Genet* 2008;40:1407–1409.
3. Landi MT, Chatterjee N, Yu K, et al. A genome-wide association study of lung cancer identifies a region of chromosome 5p15 associated with risk for adenocarcinoma. *Am J Hum Genet* 2009;85:679–691.
4. Jin G, Xu L, Shu Y, et al. Common genetic variants on 5p15.33 contribute to risk of lung adenocarcinoma in a Chinese population. *Carcinogenesis* 2009;30:987–990.
5. Kohno T, Kunitoh H, Shimada Y, et al. Individuals susceptible to lung adenocarcinoma defined by combined HLA-DQA1 and TERT genotypes. *Carcinogeneis* 2010;31:834–841.
6. Broderick P, Wang Y, Vijayakrishnan J, et al. Deciphering the impact of common genetic variation on lung cancer risk: a genome-wide association study. *Cancer Res* 2009;69:6633–6641.
7. Bartsch H, Haugen A, Risch A, et al. Genetic susceptibility. In WD Travis, E Brambilla, HK Muller-Hermelink, et al. (Eds.), World Health Organization Classification of Tumors: Pathology and Genetics, Tumours of Lung, Pleura, Thymus and Heart. Lyon: IARC Press, 2004. Pp. 24–25.
8. Hammar SP, Bramilla C, Pugatch B, et al. Squamous cell carcinoma. In WD Travis, E Brambilla, HK Muller-Hermelink, et al. (Eds.), World Health Organization Classification of Tumors: Pathology and Genetics, Tumours of Lung, Pleura, Thymus and Heart. Lyon, IARC Press, 2004. Pp. 26–30.
9. Kiyohara C, Yoshimasu K, Shirakawa T, et al. Genetic polymorphisms and environmental risk of lung cancer: a review. *Rev Environ Health* 2004;19:15–38.

Copyright © 2011 by the International Association for the Study of Lung Cancer

Copyright © 2011 by the International Association for the Study of Lung Cancer.

TABLE 2. Association of Polymorphisms in Eight Lung Cancer Susceptibility Genes with Lung SQC Risk

Gene	Polymorphism/ Allele	rs Number	Genotype/ Mode[a]	Control (%)	Case (%)	OR[b]	(95% CI, p Value)
GWAS genes							
CHRNA3	C/T	rs1051730	C/C	314 (97)	349 (93)	Reference	
			C/T	10 (3)	25 (7)	2.57	(1.03–6.87, 0.04996)
			T/T	0 (0)	0 (0)	—	(—)
			Dominant			2.57	(1.03–6.87, 0.04996)
			Recessive			—	(—)
TERT	T/G	rs2736100	T/T	116 (36)	142 (38)	Reference	
			T/G	165 (52)	175 (47)	0.74	(0.49–1.11, 0.15)
			G/G	39 (12)	53 (14)	1.22	(0.66–2.29, 0.53)
			Dominant			0.83	(0.56–1.23, 0.36)
			Recessive			1.54	(0.88–2.72, 0.13)
HLA-DQA1	DQA1*03	—	Others/others	121 (37)	120 (32)	Reference	
			Others/*03	169 (52)	194 (51)	0.97	(0.64–1.46, 0.88)
			*03/*03	35 (11)	63 (17)	1.74	(0.94–3.27, 0.080)
			Dominant			1.10	(0.74–1.63, 0.63)
			Recessive			1.78	(1.03–3.13, 0.043)
DNA repair genes							
TP53	Arg72Pro	rs1042522	Arg/Arg	152 (47)	127 (34)	Reference	
			Arg/Pro	139 (43)	191 (51)	1.73	(1.15–2.63, 0.0093)
			Pro/Pro	34 (10)	59 (16)	1.85	(1.03–3.38, 0.042)
			Dominant			1.73	(1.17–2.56, 0.0058)
			Recessive			1.36	(0.79–2.39, 0.27)
MDM2	T/G in intron 1 (SNP309)	rs2279744	T/T	79 (24)	68 (18)	Reference	
			T/G	151 (46)	183 (49)	1.59	(0.96–2.63, 0.071)
			G/G	95 (29)	126 (33)	1.54	(0.90–2.65, 0.12)
			Dominant			1.55	(0.98–2.47, 0.064)
			Recessive			1.09	(0.72–1.63, 0.69)
OGG1	Ser326Cys	rs1052133	Ser/Ser	98 (30)	115 (31)	Reference	
			Ser/Cys	164 (50)	162 (43)	0.86	(0.55–1.33, 0.49)
			Cys/Cys	63 (19)	100 (27)	1.50	(0.90–2.50, 0.12)
			Dominant			1.04	(0.69–1.57, 0.84)
			Recessive			1.65	(1.06–2.57, 0.027)
Metabolic genes							
CYP1A1	Ile462Val	rs1048943	Ile/Ile	188 (58)	196 (52)	Reference	
			Ile/Val	118 (36)	160 (42)	1.51	(1.02–2.24, 0.042)
			Val/Val	19 (6)	21 (6)	0.86	(0.40–1.89, 0.71)
			Dominant			1.40	(0.96–2.04, 0.081)
			Recessive			0.73	(0.34–1.61, 0.43)
GSTM1	Presence/absence	—	Presence	159 (50)	174 (47)	Reference	
			Absence	158 (50)	200 (53)	0.93	(0.64–1.36, 0.72)

[a] ORs in the dominant (ie AA vs Aa + aa, where A, major allele; a, minor allele) and recessive (ie AA + Aa vs aa) modes are shown.
[b] Adjusted for sex, age (≥49, 50–59, 60–69, ≤70 yr), and smoking dosage (0, 1–50, ≤50 pack-years).
SQC, squamous cell carcinoma; OR, odds ratio; CI, confidence interval.

genes, *TP53*, *OGG1*, and *CYP1A1* ($p < 0.05$), consistent with previous reports that their minor alleles are associated with increased risk.[7,9,11] The OR of heterozygotes for the *CYP1A1* SNP was significantly increased, but that of minor homozygotes did not increase. Accordingly, increase in the OR in the dominant mode was only marginal. Genotypes containing the MDM2-G allele showed increased ORs as reported[10]; however, their increases were only marginally significant. Therefore, it was concluded that these two SNPs are marginally significantly associated with lung SQC risk in this study population. The GSTM1-absence genotype did not show an increase in OR. Thus, polymorphisms in two DNA repair genes, *TP53* and *OGG1*, were significantly and those in a DNA repair gene, *MDM2*, and a metabolic gene, *CYP1A1*, were marginally significantly associated with lung SQC risk in a population, in which associations of polymorphisms in the *CHRNA3* and *HLA-DQA1* genes identified by GWASs had been observed.

Copyright © 2011 by the International Association for the Study of Lung Cancer

Copyright © 2011 by the International Association for the Study of Lung Cancer.

10. Wilkening S, Bermejo JL, Hemminki K. MDM2 SNP309 and cancer risk: a combined analysis. *Carcinogenesis* 2007;28:2262–2267.
11. Li H, Hao X, Zhang W, et al. The hOGG1 Ser326Cys polymorphism and lung cancer risk: a meta-analysis. *Cancer Epidemiol Biomarkers Prev* 2008;17:1739–1745.
12. Sakiyama T, Kohno T, Mimaki S, et al. Association of amino acid substitution polymorphisms in DNA repair genes TP53, POLI, REV1 and LIG4 with lung cancer risk. *Int J Cancer* 2005;114:730–737.
13. Arand M, Mühlbauer R, Hengstler J, et al. A multiplex polymerase chain reaction protocol for the simultaneous analysis of the glutathione S-transferase GSTM1 and GSTT1 polymorphisms. *Anal Biochem* 1996;236:184–186.
14. Breslow NE, Day NE. Statistical methods in cancer research. Volume I—the analysis of case-control studies. *IARC Sci Publ* 1980:5–338.
15. Sobue T, Yamamoto S, Hara M, et al. Cigarette smoking and subsequent risk of lung cancer by histologic type in middle-aged Japanese men and women: the JPHC study. *Int J Cancer* 2002;99:245–251.
16. Zhang X, Miao X, Guo Y, et al. Genetic polymorphisms in cell cycle regulatory genes MDM2 and TP53 are associated with susceptibility to lung cancer. *Hum Mutat* 2006;27:110–117.

Copyright © 2011 by the International Association for the Study of Lung Cancer

Copyright © 2011 by the International Association for the Study of Lung Cancer.

Published Ahead of Print on October 12, 2010 as 10.1200/JCO.2010.30.5334
The latest version is at http://jco.ascopubs.org/cgi/doi/10.1200/JCO.2010.30.5334

JOURNAL OF CLINICAL ONCOLOGY — ORIGINAL REPORT

Association of DNA Repair Gene Polymorphisms With Response to Platinum-Based Doublet Chemotherapy in Patients With Non–Small-Cell Lung Cancer

Kouya Shiraishi, Takashi Kohno, Chiharu Tanai, Yasushi Goto, Aya Kuchiba, Seiichiro Yamamoto, Koji Tsuta, Hiroshi Nokihara, Noboru Yamamoto, Ikuo Sekine, Yuichiro Ohe, Tomohide Tamura, Jun Yokota, and Hideo Kunitoh

From the National Cancer Center Research Institute; National Cancer Center Hospital; and the Center for Cancer Control and Information Services, National Cancer Center, Tokyo, Japan.

Submitted May 17, 2010; accepted July 30, 2010; published online ahead of print at www.jco.org on October 12, 2010.

Supported in part by Grant-in-Aid No. KAKENHI 20014031 from the Ministry of Education, Culture, Sports, Science and Technology for Scientific Research on Priority Areas, and Nos. 19S-1 and 19-9 from the Ministry of Health, Labor and Welfare for Cancer Research and by the 3rd-term Comprehensive 10-Year Strategy for Cancer Control. K.S. received a Research Resident Fellowship from the Foundation for Promotion of Cancer Research in Japan.

Authors' disclosures of potential conflicts of interests and author contributions are found at the end of this article.

Corresponding author: Takashi Kohno, PhD, Biology Division, National Cancer Center Research Institute, 5-1-1, Tsukiji, Chuo-ku, Tokyo 104-0045, Japan; e-mail: tkkohno@ncc.go.jp.

© 2010 by American Society of Clinical Oncology

0732-183X/10/2899-1/$20.00

DOI: 10.1200/JCO.2010.30.5334

ABSTRACT

Purpose
To identify polymorphisms in DNA repair genes that affect responses to platinum-based doublet chemotherapy in patients with non–small-cell lung cancer (NSCLC).

Patients and Methods
In total, 640 patients with NSCLC who received platinum-based doublet chemotherapy in the National Cancer Center Hospital in Japan from 2000 to 2008 and whose responses were evaluated by Response Evaluation Criteria in Solid Tumors (RECIST) participated in a study of the association between response and genotypes for 30 single nucleotide polymorphisms (SNPs) in 27 DNA repair genes. Candidate SNPs were selected in a discovery set of 201 patients, and their associations were validated in an independent set of 439 patients by prespecified *P* value criteria.

Results
Homozygotes for the minor allele TP53-72Pro of the Arg72Pro SNP in the *TP53* gene showed a better response rate (54.3%) than those for the major allele TP53-72Arg (29.1%; $P = 4.4 \times 10^{-5}$) irrespective of therapeutic regimens, and minor allele homozygotes had significantly longer progression-free and overall survivals than major allele homozygotes (hazard ratio [HR], 0.85; 95% CI, 0.74 to 0.98; $P = .020$; and HR, 0.86; 95% CI, 0.74 to 0.99; $P = .039$). Minor allele carriers for SNP Lys940Arg in the poly (ADP-ribose) polymerase 1 (*PARP1*) gene showed a better response rate to the paclitaxel regimen (45.8%) than to the gemcitabine regimen (10.5%; *P* for interaction = .019).

Conclusion
Polymorphisms in the *TP53* and *PARP1* genes are involved in inter-individual differences in the response to platinum-based doublet chemotherapy in patients with NSCLC.

J Clin Oncol 28. © *2010 by American Society of Clinical Oncology*

INTRODUCTION

Non–small-cell lung cancer (NSCLC) is a major cause of cancer-related death with 5-year survival rates of < 20%.[1] Cytotoxic chemotherapy is the standard care for patients with advanced NSCLC. The standards of therapeutic regimens are platinum-based doublets (platinum plus another agent).[2] The drugs paired with platinum include microtubule-targeted agents (paclitaxel, docetaxel, or vinorelbine) and DNA-damaging agents (gemcitabine or irinotecan). The efficacy of each combination has been demonstrated to be similar by a series of trials in unselected patients with response rates of 30% to 40%.[3-5] Therefore, predictive factors for the efficacy of these chemotherapy regimens are being investigated for the development of customized therapies.

Considering that agents that damage DNA or disturb chromosomal integrity are used for chemotherapy, activities that repair DNA or chromosome damage possibly influence the outcome of patients with NSCLC after chemotherapy. In fact, expression of *ERCC1*, which is involved in the repair of DNA adducts generated by platinum, has been shown to be a possible predictive factor for the efficacy of the postoperative cisplatin-based adjuvant chemotherapy in resected tumors.[6,7] More recently, a single nucleotide polymorphism (SNP) in the *ERCC1* gene, rs11615, which affects *ERCC1* mRNA levels, was suggested to be associated with response (ie, tumor regression) of patients with advanced NSCLC to platinum-based chemotherapy.[8] Since SNPs can be examined by using blood cells, they will be promising biomarkers in the clinical

Information downloaded from jco.ascopubs.org and provided by at Kokuritsu Gan Center on October 14, 2010 from
Copyright © 2010 American Society of Clinical Oncology. All rights reserved.
Copyright 2010 by American Society of Clinical Oncology