A functional variant in \textit{NKX3.1} associated with prostate cancer susceptibility down-regulates \textit{NKX3.1} expression

Shusuke Akamatsu\textsuperscript{1,4}, Ryo Takata\textsuperscript{1,5}, Kyoto Ashikawa\textsuperscript{2}, Naoya Hosono\textsuperscript{2}, Naoyuki Kamatani\textsuperscript{3}, Tomoaki Fujikawa\textsuperscript{5}, Osamu Ogawa\textsuperscript{4}, Michiaki Kubo\textsuperscript{2}, Yusuke Nakamura\textsuperscript{6} and Hidewaki Nakagawa\textsuperscript{1,}\textsuperscript{*}

\textsuperscript{1}Laboratory for Biomarker Development, \textsuperscript{2}Laboratory for Genotyping Development and \textsuperscript{3}Laboratory for Statistical Analysis, Center of Genome Medicine, RIKEN, Yokohama, Japan, \textsuperscript{4}Department of Urology, Graduate School of Medicine, Kyoto University, Kyoto, Japan, \textsuperscript{5}Department of Urology, Iwate Medical University, Morioka, Japan and \textsuperscript{6}Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan

Received July 2, 2010; Revised and Accepted August 11, 2010

Genome-wide association studies (GWAS) identified multiple susceptible loci for prostate cancer (PC), and recent GWAS implicated that a common variant rs1512268 on chromosome \textit{8p21} is associated with PC susceptibility, which is located at 14 kb downstream of a prostate tumor suppressor gene \textit{NKX3.1}. To clarify a susceptibility gene and functional variants in this locus, we performed re-sequencing and fine mapping of this region and identified 12 candidates of functional single nucleotide polymorphisms that were absolutely linked with each other. Screening of these variants by RNA stability assay, electrophoretic mobility shift assay (EMSA) and reporter assay indicated that rs11781886 in the 5'-UTR of \textit{NKX3.1} displayed different binding affinity to nuclear proteins between the alleles, and that the transcriptional activity of the \textit{NKX3.1} promoter was significantly lower in the susceptible allele of this variant. Sp1 was determined to be the transcription factor that binds to the susceptible G allele, but not to the non-susceptible A allele. Allele-specific transcript quantification (ASTQ) and quantitative PCR analyses showed that the expression of \textit{NKX3.1} in the prostate was significantly lower in the subjects with the haplotype carrying the susceptible allele. These results suggest that the functional variant rs11781886 in the 5'-UTR of \textit{NKX3.1} can affect its transcription by altering the binding affinity of a transcriptional factor Sp1, and might result in PC susceptibility by lowering expression of \textit{NKX3.1} in the prostate.

INTRODUCTION

Prostate cancer (PC) is the most common malignancy among males in western countries (1,2). Asians have the lowest incidence of PC in the world; however, its incidence is now rapidly increasing in Asian countries including Japan probably due to the shift to a westernized lifestyle and the rapid increase in the aging population (3). Although the precise mechanisms of PC development are not fully elucidated, it is evident that genetic factors play some important roles in the etiology of PC (4,5). A positive family history of PC has been recognized as one of the most important risk factors for PC (5), and twin studies indicated that the contribution of genetic factors to the development of PC is larger than in other types of human common tumors (4). Genome-wide association studies (GWAS) have identified nearly 30 single nucleotide polymorphisms (SNPs) on various genes or chromosome loci which are related to PC susceptibility (6–11). However, most of these SNPs are merely the associated markers and only a few variants that functionally affect PC carcinogenesis have been clarified (12).

*To whom correspondence should be addressed at: Laboratory for Biomarker Development, Center for Genomic Medicine, RIKEN, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan. Tel: +81 354495786; Fax: +81 354495785; Email: hidewaki@ims.u-tokyo.ac.jp

© The Author 2010. Published by Oxford University Press. All rights reserved. For Permissions, please email: journals.permissions@oxfordjournals.org
Recently, GWAS and worldwide meta-analysis identified rs1512268 on chromosome 8p21 as a genetic marker for PC susceptibility, which is located at approximately 14 kb downstream of a well-known prostate tumor suppressor gene NKX3.1 (7). NKX3.1 is an androgen-regulated homeobox gene which is exclusively expressed in the prostate (13). In a normal prostate, it is proposed to play a key role in regulation of growth and differentiation of prostate epithelium (14,15). On the other hand, with regard to carcinogenesis, NKX3.1 maps to 8p21, a region which was reported to be most frequently subject to loss of heterozygosity (LOH) in human PCs (13,16) and it has been recognized to function as a tumor suppressor of prostate. Although the locus around rs1512268 on 8p21 is an intriguing region to find an association with human prostate carcinogenesis through NKX3.1, further confirmation is required to determine whether the locus around this SNP can actually affect NKX3.1 both in genetic and functional aspects.

In this study, we performed re-sequencing and fine mapping of this region, and identified a SNP in the 5′-UTR of NKX3.1 that could alter the binding affinity for a nuclear protein, and decrease the transcriptional activity of the NKX3.1 promoter in prostate cells. We also demonstrated that the haplotype carrying the susceptible allele at this locus could decrease NKX3.1 expression in the normal prostate, contributing to prostate carcinogenesis. These findings should provide new insights into understanding the pathogenesis and genetics of PC and facilitate developing new approaches to prevent or diagnose PC.

RESULTS

Fine mapping of the locus around rs1512268

A recent GWAS identified rs1512268 on chromosome 8p21 as a marker associated with PC susceptibility (7). We also confirmed this association in the GWAS of PC in a Japanese population (17). To clarify the susceptibility gene and functional variants in this locus, we first selected 66 tag-SNPs in the 120 kb region (Chr8: 23.51 – 63 Mb) from Phase II of HapMap JPT data, and genotyped these SNPs using 1583 PC cases and 3386 controls. We found that rs1512268 represented a large linkage disequilibrium (LD) block between 23.547 and 23.605 Mb (Supplementary Material, Fig. S1).

To further define the region of interest, we next genotyped all the 32 SNPs registered in Phase II of HapMap JPT data in this region with minor allele frequency (MAF) of 0.05 or higher. All the SNPs genotyped in this region were in one strong LD block spanning 57.6 kb (Fig. 1A). NKX3.1 was the sole gene included in this region, and there were no known microRNAs (miRNAs) or ESTs registered in this region. Therefore, we determined NKX3.1 to be the causative gene in this region. To search for candidates of functional SNPs in this region, we re-sequenced a 13 kb region from 5 kb upstream to the 3′-UTR of NKX3.1 using DNAs of 94 PC cases. We found 23 dbSNPs and 18 newly identified SNPs in this region. Of these SNPs, we further genotyped 16 SNPs with MAF of 0.05 or higher. Finally, we selected 12 strongest associated SNPs for the candidates for the functional analysis. These SNPs were absolutely linked with each other (r^2 > 0.95). Among them, six SNPs were located at the upstream of NKX3.1, one in the 5′-UTR of NKX3.1 (rs11781886), one in the intron (SNP02) and four in the 3′-UTR (rs4872176, rs34783892, rs4872177 and rs1567669) as shown in Figure 1B.

rs11781886 and rs11786738 are candidates of functional SNPs

Since our re-sequencing did not find non-synonymous SNPs or splice site variants of NKX3.1, we hypothesized that one or a combination of candidate SNPs could have functional significance to alter the expression of NKX3.1 transcript. Among the 12 candidate variants, we first tested the possibility that the four SNPs (rs4872176, rs34783892, rs4872177 and rs1567669) in the 3′-UTR might affect the stability of mRNA (18–21). We measured the amount of mRNA after mixing in vitro transcribed full-length NKX3.1 RNAs from the susceptible and non-susceptible alleles with extracts of LNCaP cells. However, there was no difference in the degradation ratio of mRNA between the alleles (Supplementary Material, Fig. S2). Binding of miRNA is another common mechanism by which the 3′-UTR regulates transcription. However, there was no database registered miRNA binding site that could be altered by the four SNPs in the 3′-UTR of NKX3.1. Therefore, we considered that the four SNPs in the 3′-UTR could not affect NKX3.1 expression, and excluded these SNPs from further analysis.

Hence, we hypothesized that the remaining eight SNPs in the intron, the 5′-UTR or the upstream of NKX3.1 could have functional significance to alter NKX3.1 expression. To find out the functional SNPs among these SNPs, we performed electrophoretic mobility shift assays (EMSAs) for all the eight SNPs. We observed different binding affinity of nuclear proteins from LNCaP cells between the alleles for the probes of rs11781886 and rs11786738, but not for the others (Fig. 2A). The band for the probe of rs11781886 appeared only at the susceptible allele G, and that of rs11786738 appeared only at the non-susceptible allele A. The different binding affinity between the alleles at rs11781886 was confirmed by a competition assay (Fig. 2B). To determine the nuclear proteins that bound to the SNP, we searched TFSEARCH (22) database and found that the susceptible allele G of rs11781886 had a consensus binding sequence of Sp1, but the non-susceptible allele A did not. When we performed a supershift assay in this EMSA, a shifted band was observed after the incubation with anti-Sp1 antibody, confirming that Sp1 transcriptional factor could bind to the susceptible allele of rs11781886 (Fig. 2B). Chromatin immunoprecipitation (ChIP) assay also confirmed the binding of Sp1 to a genomic fragment including rs11781886 in vivo (Fig. 2C). Similarly, difference in the binding affinity between the alleles at rs11786738 was confirmed by a competition assay (Supplementary Material, Fig. S3A). Two candidate transcription factors, Oct-1 and CHOP-C/EBPα (a heterodimer of GADD153 and C/EBPα), were identified on the database as candidate proteins which could only bind to the non-susceptible A allele of rs11786378. However, no shifted band was observed when incubated with either anti-Oct-1 or C/EBPα antibodies (Supplementary Material,
Fig. S3A), and the nuclear protein that bound to rs11786738 could not be determined.

rs11781886 at the 5′-UTR regulates the transcriptional activity of NKX3.1 promoter

To examine whether rs11781886 and/or rs11786738 could affect gene expression, we performed reporter assays by inserting these regions into luciferase expressing vectors. rs11781886 is located at the 5′-UTR of NKX3.1, 15 bases upstream of the start codon. To examine the difference between the alleles at rs11781886 in the context of the NKX3.1 promoter, we cloned a 278 bp NKX3.1 promoter region including a TAAT box, two CAAT boxes and rs11781886 flanking region into pGL3 basic vector for each allele of rs11781886. There was no other common SNP in this region. The NKX3.1 promoter with the risk allele G at rs11781886 showed significantly lower luciferase activity compared with that with the non-risk allele A (Fig. 2D). For assessment of rs11786738, which is 4986 bp upstream of the transcription start site of NKX3.1, a sequence around rs11786738 that was examined in the EMSA was cloned into pGL3 vector with SV40 promoter to evaluate a difference in its enhancer or silencer activity between the alleles (Supplementary Material, Fig. S3B). However, there was no difference in the transcription activity in either allele compared with the mock vector. Combining the results of EMSA and reporter assays, we determined that rs11781886 at the 5′-UTR of NKX3.1 is a functional variant conferring PC susceptibility.

Correlation of NKX3.1 genotype with NKX3.1 expression in the prostate

To investigate the effect of genotypes on NKX3.1 expression in vivo, we performed ASTQ assay using a TaqMan probe for rs1567669, a SNP located in the 3′-UTR and absolutely linked with rs11781886, on 13 normal prostate samples with heterozygosity of this SNP. The allele T, which represents the susceptible genotype, showed 2.5-fold lower expression than the allele C, which represents the non-susceptible
genotype (\(P < 0.01\), Student’s \(t\)-test), consistent with the tumor suppressive nature of \(\text{NKX3.1}\) (Fig. 3A). Furthermore, quantitative PCR analysis of 33 normal prostate tissue samples also revealed significant association between the genotypes and \(\text{NKX3.1}\) expression levels (\(P < 0.05\) by one-way ANOVA), and addition of susceptible alleles significantly decreased expression of \(\text{NKX3.1}\) in the prostate tissues (Fig. 3B).

**DISCUSSION**

The \(\text{NKX3.1}\) gene is localized to the region subjected to LOH in human PCs and further functional analyses have been indicating its property of a prostate tumor suppressor. Gene targeting studies in mouse models have suggested that the haploinsufficiency of \(\text{Nkx3.1}\) is semidominant (14). Thus, it is likely that functional germline variants at one of the alleles confer predisposition to PC in humans as well. However, no somatic or germline mutations have been discovered except for a very rare variant that cosegregates with hereditary PC (23,24). Although some epigenetic changes are reported to contribute to the expression of \(\text{NKX3.1}\) in PCs (25,26), it is still not clear how the expression of \(\text{NKX3.1}\) is regulated in normal prostate cells and in cancer cells. Therefore, our finding of a common susceptibility variant in the flanking region of \(\text{NKX3.1}\), which could functionally regulate its expression, would provide an important and novel clue to understanding the regulatory mechanism of \(\text{NKX3.1}\) and prostate carcinogenesis. In addition, despite the well-known property of \(\text{NKX3.1}\) as a tumor suppressor, its association with predisposition to PC has never been reported. Our finding of the functional association between the common germline variant and \(\text{NKX3.1}\) expression in the prostate underscores the critical role of \(\text{NKX3.1}\) in PC carcinogenesis.

SNPs in gene regulatory regions such as promoter and enhancer regions can affect gene expression by altering binding affinities to transcription factors (27) and we searched for functional SNPs in \(\text{NKX3.1}\) and the 5′-flanking region. Together with the results of our EMSA and reporter assays, \(\text{Sp1}\) binding to the susceptible allele G at rs11781886 region represses the transcriptional activity of \(\text{NKX3.1}\) promoter compared with the non-susceptible allele A. Regarding another candidate rs11786738 at the upstream of \(\text{NKX3.1}\) promoter compared with the non-susceptible allele A, we could not confirm it to be a functional variant due to the discordance between the EMSA and reporter assays, and the failure to discover the transcription factor that could bind to this region.
The 3′-UTR of NKX3.1 has recently been found to harbor functional androgen responsive elements (ARE), which are presumed to play important regulatory roles of NKX3.1 expression (28). However, we could not find any involvement of the four 3′-UTR SNPs we found in this study with these ARE sequences at the 3′-UTR. Of the SNPs in the same LD block, we did not examine the variants that were downstream of NKX3.1 in this study. However, it is possible that these SNPs exert additional influence on NKX3.1 expression.

In summary, we identified a functional variant rs11781886 that affects NKX3.1 expression in the prostate by altering the binding affinity for a transcription factor Sp1. The haplotype that carried the susceptible allele at this locus showed lower expression of NKX3.1. Our finding of a germline common variant that affects the expression of NKX3.1, a gene which plays a central role in PC development and prostate differentiation, should provide new insights into understanding the pathogenesis and genetics of PC.

**MATERIALS AND METHODS**

**Samples for genotyping**

All PC cases and controls were obtained from the BioBank Japan at the Institute of Medical Science, the University of Tokyo (29). This project was started in 2003 to collect a total of 300,000 cases that have at least one of 47 diseases by a collaborative network of 66 hospitals located at all areas of Japan. For fine mapping of the 8p21 PC susceptibility region, we utilized the same case and control sample set that we used in our GWAS (17). From the registered samples in the BioBank Japan, we selected 1583 individuals that were clinically diagnosed as having PC based on the pathological evaluation of prostatic biopsy. The controls consisted of 2480 individuals that were registered in the BioBank Japan as subjects with 13 diseases other than PC and 906 healthy volunteers collected at the Osaka-Midosuji Rotary Club. All participants provided written informed consent. This research project was approved by the ethical committees at the Institute of Medical Science, the University of Tokyo and the Center of Genomic Medicine, RIKEN.

**Genotyping and re-sequencing**

Genotyping was performed using the multiplex PCR-based Invader assay (30). We assessed case–control association analysis using a 1-degree-of-freedom Cochrane-Armitage trend test. LD and haplotype analysis were performed by the use of Haploview (Broad Institute, MA, USA). Direct sequencing of the candidate region was performed in 94 PC cases using ABI3730 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA).

**Cell line and clinical prostate samples**

PC cell line LNCaP was purchased from the American Type Culture Collection (ATCC, Rockville, MD, USA). LNCaP was grown in RPMI (Invitrogen, Carlsbad, CA, USA) supplemented with 10^{-8}M dihydrotestosterone (Teikoku Hormone MFG, Kawasaki, Japan), 10% fetal bovine serum.
Whole cell lysate and nuclear cell lysate preparation

Whole cell lysate and nuclear cell lysate were prepared from LNCaP cells using TransFactor Whole Cell Extraction Kit (Clontech, Mountain View, CA, USA) and TransFactor Extraction Kit (Clontech), respectively, according to the manufacturer’s instructions.

In vitro RNA stability assays

We synthesized cDNA from RNA of normal prostate tissue samples that were homozygous for each of the inferred haplotypes. A gene-specific primer 5′-AAATCTAATAAGACAT GGAGTACTTTATTTC-3′ was used for reverse transcription with SuperScript III First-Strand Synthesis System (Invitrogen). We then amplified two types of the full-length NKX3.1 (risk allele and non-risk allele) from the cDNA by PCR. The PCR primers were 5′-GGGGATCCGGGCGGGGCGGGTGCA TT-3′ and 5′-AAATCTAATAAGACATGGGTACTTTAT TTTCAAAAACAC-3′. The PCR products were cloned into pCR2.1 vector (Invitrogen), which had a T7 promoter. Vectors were then digested with BamHI, and NKX3.1 was expressed in vitro using Ribomax Large Scale RNA Production System-T7 (Promega) and purified according to the manufacturer’s instructions. We mixed each 1 µg of the synthesized RNAs and diluted whole cell extract (1:50–1:1000) and incubated at room temperature. The reaction was stopped with the addition of formaldehyde dye at different time points, and the samples were then heated at 68°C. After the reaction, RNA was separated on a denaturing gel. We scanned results on Molecular Imager FX (Bio-Rad Laboratories, Hercules, CA, USA) and measured signal intensities of full-length RNAs. The test was repeated four times with duplicate samples.

Electrophoretic mobility shift assays (EMSAs)

EMSA assays were carried out using DIG Gel Shift Kit, 2nd Generation (Roche) according to the manufacturer’s instructions. Thirty-one basepair sense and anti-sense strand for each SNP allele were prepared, and double-stranded oligonucleotides were generated. Their sequences are shown at Supplementary Material, Table S1. After pre-incubation with poly(dIc) and poly-l-lysine, DIG-labeled probes were incubated with 10 µg of nuclear extracts prepared from LNCaP cells for 20 min at 25°C. For competition studies, we pre-incubated a 5–125-fold excess of unlabeled oligonucleotide probe with the nuclear extract before adding the DIG-labeled probe. In supershift assays, the nuclear extract was incubated with DIG-labeled probe at 25°C for 30 min, and then incubated for another 60 min on ice after addition of 2 µg of antibody to Sp1 (PEP2, Santa Cruz Biotechnology, Santa Cruz, CA, USA). Oct-1 (c-21, Santa Cruz Biotechnology) or C/EBPα (sc-61, Santa Cruz Biotechnology) antibodies were used for supershift assays. The protein–oligonucleotide complexes were visualized by autoradiography. All EMSAs were repeated twice to check for reproducibility.

Chromatin immunoprecipitation (ChIP)

ChIP assays were performed using EZChIP kit (Millipore, Billerica, MA, USA) according to the manufacturer’s instructions with minor modifications. LNCaP cells were cross-linked with 1% formaldehyde for 10 min and then sonicated into 200–1000 kb fragments by Microson XL-2000 (Misonix, Farmingdale, NY, USA). DNA–protein complex was immunoprecipitated with 4 µg of antibody to Sp1 (Santa Cruz Biotechnology). Antibody to rabbit IgG was prepared as a negative control. Dynabeads Protein G (Invitrogen) was used instead of Protein G Agarose slurry for precipitation. After washing once with the high-salt buffer (0.1% SDS, 1% Triton X, 2 mM EDTA, 20 mM Tris–HCl, 500 mM NaCl), three times with the LiCl buffer (250 mM LiCl, 1% NP-40, 1% sodium deoxycolate, 1 mM EDTA, 10 mM Tris–HCl), and once with the TE buffer, immunoprecipitated complex was reverse cross-linked and purified according to the manufacturer’s instructions. PCRs were performed using primer sequences 5′-CGCTCCCTTGGGTATAAG-3′ and 5′-GGA TGTCTTGGATGAGGAAG-3′ for the rs11781886 region, and 5′-AGTGGGCGGAGAATTTGG-3′ and 5′-GAGACA TGCCAGGGCAAG-3′ for the known Sp1 binding region on DHFR promoter as a positive control.

Dual luciferase reporter assays

For assessment of the rs11781886 region, a 278 bp fragment of the 5′-flanking region of NKX3.1 was obtained by PCR amplification using genomic DNA of samples homozygous for each allele. Primer sequences were forward 5′-CGCACG CGTGGGGAAGAGAGGAAATTTGG-3′ and reverse 5′-CCG CTCCAGCTCCGGAAACCTGAGCAT-3′. The amplified products were digested with Xhol and MluI (TAKARA Bio Inc., Shiga, Japan), and inserted into the upstream of luciferase reporter gene in the pGL3 basic (Promega). For assessment of the rs11786738 region, KpnI and MluI restriction enzyme sites were added to the 31 bp fragment used as a probe in the EMSA. The annealed double-stranded oligonucleotide was digested with KpnI and MluI, and inserted into the upstream of luciferase reporter gene in the pGL3 promoter (Promega). The sequences of these constructs were verified using ABI3730 Genetic Analyzer (Applied Biosystems). Twenty-four hours after plating 3×10^4 cells on six-well plates, LNCaP cells were transfected with 0.5 µg of the luciferase reporter plasmid and 0.3 µg of pRL-TK (renilla luciferase) using Lipofectamine 2000 (Promega). At 48 h after transfection, the cells were solubilized, and the luciferase activity was measured using Pikkagene Dual Luciferase Assay System (Toyo Ink, Tokyo, Japan). The luciferase activities were normalized by renilla luciferase activity. Each
experiment was carried out more than two times with hexaplicate replicates.

Allele-specific transcript quantification and quantitative real-time PCR
DNA and RNA samples were extracted from 33 normal prostate tissue samples by using QIAamp DNA Mini Kit (QIAGEN, Chatsworth, CA, USA) and RNaseasy Mini Kit (QIAGEN), respectively. cDNA was prepared using SuperScript III First-Strand Synthesis System (Invitrogen). For the 13 normal prostate tissue samples showing heterozygosity of the haplotypes we inferred, we performed allelic expression analysis by TaqMan assays using the probe for rs1567669 (Applied Biosystems). Ct values of FAM and VIC, each representing non-risk and risk alleles, were obtained for genomic DNA and cDNA samples at 40 cycles of real-time PCR. We also prepared genomic DNA and DNA samples of homogeneous for each allele and mixed them at different ratios (2:8, 3:7, 4:6, 5:5, 6:4, 7:3, 8:2) to create a standard curve by plotting Ct values of VIC/FAM against the allelic ratio of VIC/FAM for each mixture. Using the standard curve, we calculated allelic ratio for each genomic DNA and cDNA samples. We measured each sample in quadruplicate in one assay; tests were independently repeated twice. We carried out real-time PCR on Light Cycler 480 using LightCycler Fast Start DNA Master SYBR Green I (Roche, Basel, Switzerland). Primer sequences for NKX3.1 were 5′-CAGTCCCTA.CTGAATCAC TCTTTCCTC-3′ and 5′-CAGTGAAATGTGTAATCC TGC-3′. The results were normalized to ACTB and analyzed by the standard curve method. The statistical analysis was performed by one-way ANOVA, and differences between groups were analyzed by Tukey’s test.

SUPPLEMENTARY MATERIAL
Supplementary Material is available at HMG online.

ACKNOWLEDGEMENTS
We thank the staffs of Laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN for their contribution to SNP genotyping and Ms Kumiko Watanabe in Laboratory for Biomarker Development for her technical assistance. We are also grateful to the members of the BioBank Japan project and the Rotary Club of Osaka-Midosuji District 2660 Rotary International in Japan for supporting our study.

Conflict of Interest statement. None declared.

FUNDING
This work was conducted as a part of the BioBank Japan Project that was supported by the Ministry of Education, Culture, Sports, Sciences and Technology of the Japanese government, and was supported in part by Research grant no. 22390306 (H.N.) from the Japan Society for the Promotion of Science. S.A. is a Japan Society for the Promotion of Science Research Fellow.

REFERENCES