National Cancer Center (NCC) Biobank. These samples were from patients with LADC who received therapy at the NCC Hospital (Tokyo, Japan) between 1997 and 2012. All frozen samples were confirmed to be positive for *KIF5B-RET* fusion by reverse-transcriptase polymerase chain reaction (PCR) analysis, according to a previously described method.[3] *CCDC6-RET* fusion was detected by fusion fluorescence in situ hybridization (FISH) analysis of paraffin-embedded tissues using *RET*- and *CCDC6*-specific probes (Chromosome Science Labo Inc., Sapporo, Japan). This study was approved by the Institutional Review Board of the NCC.

Cloning and Sequencing of DNAs Containing Breakpoint Junctions

Genomic DNAs were extracted from cancer and noncancerous tissues using the QIAamp DNA Mini Kit or the QIAamp DNA Micro Kit (Qiagen, Hilden, Germany). Genomic DNA fragments containing breakpoint junctions were amplified by genomic PCR using primers that hybridized within the *KIF5B* and *RET* loci. PCR products specifically amplified in samples of interest were subjected to direct Sanger sequencing. The primers used are listed in Supplementary Table 1 (Supplementary Digital Content 1, http://links.lww.com/JTO/A541).

Genome-Capture Deep Sequencing Using a Next-Generation Speed Sequencer

Nucleotide sequences of *CCDC6-RET* fusion breakpoints were examined by targeted genome capture and massively parallel sequencing using an Ion Torrent Personal Genome Machine (Ion Torrent PGM) sequencing system and the Ion TargetSeq Custom Enrichment Kit (Life Technologies, Carlsbad, CA). One microgram of genomic DNA was subjected to enrichment using the probes listed in Supplementary Table 2 (Supplementary Digital Content 1, http://links.lww.com/JTO/A541). The mean depth of sequencing was approximately 1000.

Analysis of Sequence Reads Obtained by a Second-Generation Sequencer

Sequence reads were analyzed using a program developed by the authors. Briefly, reads were mapped to sequences of the *RET* and *CCDC6* genes using the Burrows-Wheeler Aligner, Smith-Waterman alignment (BWA-SW) software[18] to detect reads that mapped to both the *RET* and *CCDC6* genes. Breakpoints were extracted from the local alignment results of BWA-SW. The detailed procedure is described in Supplementary Notes (Supplementary Digital Content 2, http://links.lww.com/JTO/A542). Structures of breakpoint junctions were verified by Sanger sequencing of genomic PCR products.

Loss of Heterozygosity Analysis

Genomic DNAs obtained from cancerous and noncancerous tissues were subjected to single nucleotide polymorphism (SNP) genotyping using the Illumina HumanOmni1 2.5M Chip (Illumina, San Diego, CA). Based on the B-allele frequencies obtained using the Illumina GenomeStudio software, loss of heterozygosity (LOH) regions in *RET* and surrounding regions were deduced. Representative SNP loci were subjected to analysis of allelic imbalance using the Sequenom MassARRAY system (Sequenom, San Diego, CA).

Analysis of Nucleotide Sequences

Nucleotide sequence analysis, including search for sequence homology, was performed using the Genetyx-SV/RC Ver 8.0.1. software (Genetyx, Tokyo, Japan). Information about the distribution of repetitive elements, GC contents, conservation, DNA methylation, DNase sensitivity, and histone modification within the *RET* gene was obtained using the UCSC genome browser (http://genome.ucsc.edu/cgi-bin/hgGateway).

RESULTS

KIF5B-RET Fusion Variations in LADC

In our previous study, six of 319 LADC cases (1.9%) carried *KIF5B-RET* fusions.[3] In this study, we examined *KIF5B-RET* fusion by reverse-transcriptase PCR in a further 352 LADC cases and found eight additional *KIF5B-RET* fusion-positive cases. In total, 14 of 671 cases (2.1%) were positive for *KIF5B-RET* fusions (cases 1–4 and 7–16 in Table 1 and Supplementary Table 3, Supplementary Digital Content 1, http://links.lww.com/JTO/A541), and this frequency was consistent with those reported for other cohorts.[9,10,19]

Among those 14 cases, 10 (71%) contained a fusion of *KIF5B* exon 15 to *RET* exon 12 (K15;R12), whereas the remaining four each contained other variants. Thus, K15;R12 is the most frequent variant (Fig. 1*B*). The prevalence of the K15;R12 variant (45 of 60, 75%) was verified in a total of 60 cases, including 46 cases from eight other cohorts published to date[1–4,9,10,19,20] (Fig. 1*B*, Supplementary Table 4, Supplementary Digital Content 1, http://links.lww.com/JTO/A541). This preference was similar among cohorts from Japan, other Asian countries, and the United States ($p > 0.05$ by Fisher's exact test).

Distribution of Breakpoints in the *RET* and *KIF5B* Genes

To explore the molecular processes underlying *RET* fusion in LADC, we examined the location (clustering) of the breakpoints and the structure of the breakpoint junctions; information about the former enabled us to deduce the genomic or chromosomal features that make DNA susceptible to strand breaks, whereas information about the latter enabled us to deduce the mechanism underlying the illegitimate joining of DNA ends by DNA repair pathways.

The locations of the 28 breakpoints in the 14 *KIF5B-RET* fusion-positive cases mentioned above were identified by Sanger sequencing analysis of genomic PCR products and mapped (yellow arrowheads in Fig. 2*A* and *B*). The breakpoints in a single Korean case from another study were also identified and mapped (orange arrowheads in Fig. 2*A*; case 17 in Table 1). Consistent with the predominance of K15;R12 variants, most of the breakpoints were mapped to intron 11 of *RET* and intron 15 of *KIF5B* (Fig. 2, detailed information in Supplementary Table 5, Supplementary Digital Content 1, http://links.lww.com/JTO/A541).

Copyright © 2014 by the International Association for the Study of Lung Cancer

TABLE 1. Structure of Breakpoint Junctions of *RET* Fusions in Lung Adenocarcinoma

No.	Sample Name	Fusion Partner	Reciprocal/ Nonreciprocal	Deletion in the Joining		DNA Segment Duplication by Inversion		Nucleotide Overlap at Junction		Nucleotide Insertion at Junction		Mode of DNA End Joining	LOH Proximal to *RET*	Smoking
				RET	Partner	*RET*	Partner	Partner -*RET*	*RET*-Partner	Partner -*RET*	*RET*-Partner			
1	BR0020	*KIF5B*	Reciprocal	—	—	—	—	—	—	—	—	NHEJ	NT	No
2	L07K201T	*KIF5B*	Reciprocal	14 bp	19 bp	—	—	C	—	—	ATA	NHEJ	NT	Yes
3	349T	*KIF5B*	Reciprocal	1 bp	7 bp	—	—	—	—	A	A	NHEJ	NT	Yes
4	AD08-341T	*KIF5B*	Reciprocal	16 bp	26 bp	—	—	—	—	—	—	NHEJ	NT	No
5	RET-030	*CCDC6*	Reciprocal	52 bp	1021 bp		—	— —		—	—	NHEJ	NT	No
6	RET-024	*CCDC6*	Reciprocal	14 bp	2 bp	—	—	—	—	—	—	NHEJ	NT	Yes
7	AD12-106T	*KIF5B*	Reciprocal	—	573 bp	490 bp	—	—	—	—	—	BIR	NT	Yes
8	BR0030	*KIF5B*	Reciprocal	—	—	—	211 bp	—	—	—	—	BIR	NT	No
9	442T	*KIF5B*	Reciprocal	269 bp	—	—	232 bp	—	—	—	—	BIR	NT	No
10	AD08-144T	*KIF5B*	Reciprocal	5 bp	—	—	33 bp	—	—	—	—	BIR	NT	No
11	BR1001	*KIF5B*	Nonreciprocal					—		AGT		NHEJ	+	No
12	AD09-369T	*KIF5B*	Nonreciprocal					CTC		—		NHEJ (alternative end joining)	NT	No
13	BR1002	*KIF5B*	Nonreciprocal					A		—		NHEJ	NT	No
14	AD12-001T	*KIF5B*	Nonreciprocal					—		—		NHEJ	NT	Yes
15	BR1003	*KIF5B*	Nonreciprocal					—		CTTT		NHEJ	+	No
16	BR1004	*KIF5B*	Nonreciprocal					—		RET exon 7 to intron 7 (359 bp)		Complex rearrange	+	No
17	AK55[a]	*KIF5B*	Nonreciprocal					—		GT		NHEJ	NT	No
18	LC-2/ad[b]	*CCDC6*	Nonreciprocal					—		—		NHEJ	NT	Unknown

[a]Ju et al.[4]
[b]Suzuki et al.[21]
LOH, loss of heterozygosity; NHEJ, nonhomologous end joining; NT, not tested; BIR, break-induced replication; blank, not applicable.

Copyright © 2014 by the International Association for the Study of Lung Cancer

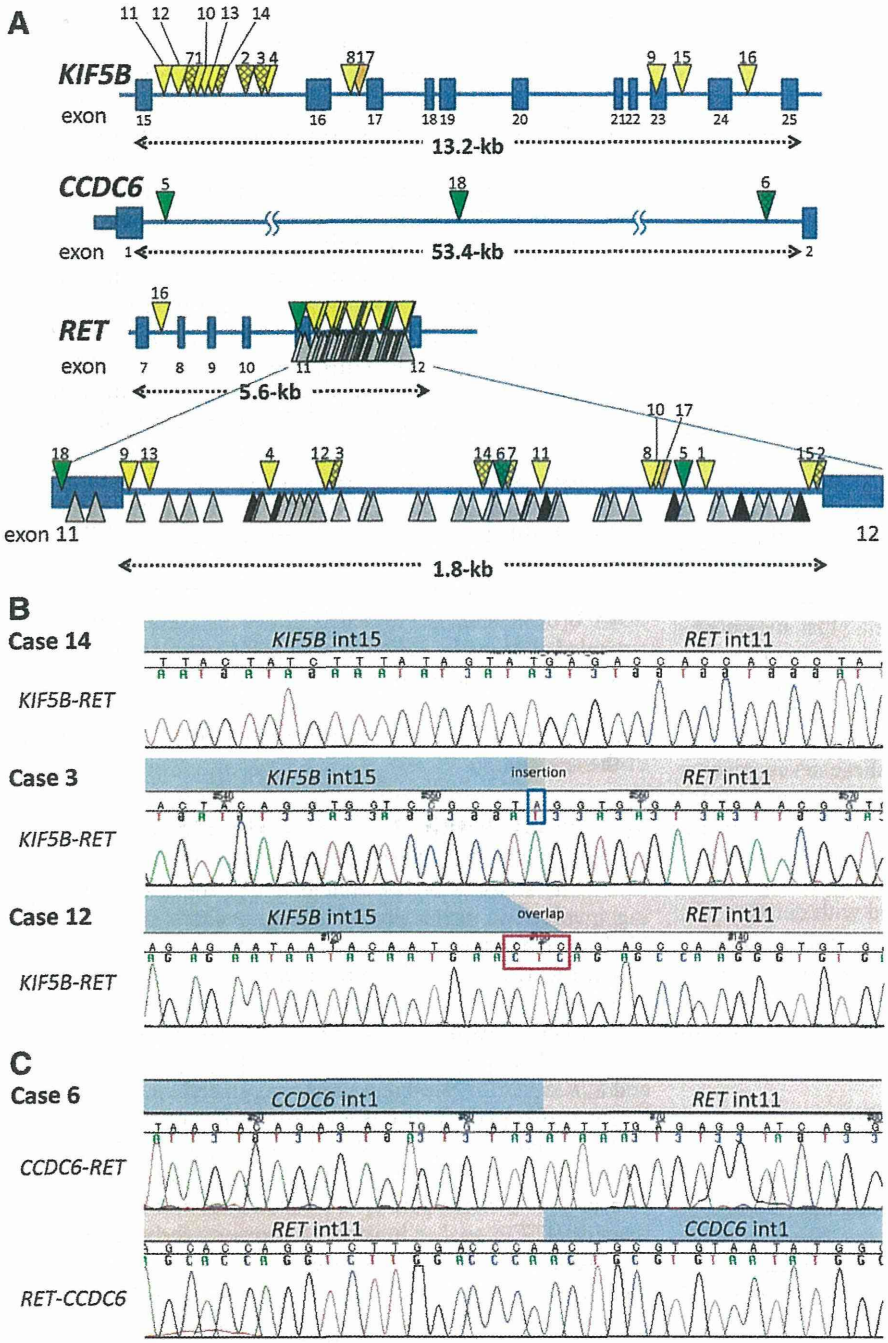


FIGURE 2. Breakpoint analysis. *A*, Distribution of breakpoints in the *CCDC6*, *KIF5B*, and *RET* genes. Yellow arrowheads indicate the locations of breakpoints for *KIF5B-RET* fusions in Japanese cases (cases 1–4 and 7–16 in Table 1), whereas the orange arrowhead indicates the breakpoints in a single Korean case (case 17). Green arrowheads indicate the locations of breakpoints of *CCDC6-RET* fusions in three Japanese cases (cases 5, 6, and 18). Arrowheads for ever-smoker LADC cases are hatched. Gray and black arrowheads indicate breakpoints of *RET-ELE1* fusion in 38 radiation-induced post-Chernobyl PTCs and six sporadic PTCs, respectively.[14–17] *B*, Electropherograms for Sanger sequencing of genomic fragments encompassing *KIF5B-RET* breakpoint junctions. PCR products were directly sequenced. Examples of three fusion patterns (joined without any nucleotide insertions or overlaps, joined with a nucleotide insertion, and joined with three nucleotide overlap) are shown. Inserted and overlapping nucleotides at breakpoint junctions are indicated, respectively, by the blue and red boxes. *C*, Electropherogram for Sanger sequencing of genomic fragments encompassing *CCDC6-RET* and *RET-CCDC6* breakpoint junctions. LDAC, lung adenocarcinoma; PCR, polymerase chain reaction; PTC, papillary thyroid carcinoma.

None of the *RET* and *KIF5B* breakpoints mapped at the same position, and no breakpoint was within 6 bp of another. To further investigate the breakpoint clustering, we mapped breakpoints in three cases of *CCDC6-RET* fusion, a minor fusion variant (cases 5, 6, and 18 in Table 1 and Supplementary Table 3, Supplementary Digital Content 1, http://links.lww.com/JTO/A541). Two of these cases were primary tumors, diagnosed by break apart and fusion *FISH*, and their breakpoints were determined by genome-capture deep sequencing of genomic DNAs using a second-generation sequencer. The remaining case was a LADC cell line from a Japanese patient, for which the breakpoints had previously been determined by the same method.[21] Two breakpoints and one breakpoint in the *RET* gene were mapped to intron 11 and exon 11, respectively (green arrowheads in Fig. 2), and no breakpoint was located within 5 bp of another. In total, a 2.0-kb region spanning exon 11 to intron 11 of *RET* and a 5.6-kb region spanning intron 15 of *KIF5B* (10 of 15, 75%) contained the majority of breakpoints (17 of 18 [94%] and 10 of 15 [75%], respectively), and these breakpoints

Copyright © 2014 by the International Association for the Study of Lung Cancer

were at least 5 bp from each other. Breakpoints within exon 11 to intron 11 of *RET* and intron 15 of *KIF5B* were not distributed in an evidently biased manner, nor did they exhibit any particular nucleotide sequence or composition (Supplementary Table 5, Supplementary Digital Content 1, http://links.lww.com/JTO/A541). Therefore, DNA strand breaks triggering oncogenic *RET* fusions in LADC occur preferentially in a few defined regions, but at nonspecific sites within those regions.

Reciprocal and Nonreciprocal Inversions Causing *RET* Fusions

To explore the modes of DNA end joining that give rise to *RET* fusion, we investigated the structures of *RET* fusion breakpoint junctions. To address whether chromosome inversion events were reciprocal, we cloned genomic segments containing reciprocal breakpoint junctions (i.e., *RET-KIF5B* and *RET-CCDC6*) from 17 Japanese cases (Table 1). Ten of the 17 cases, consisting of eight *KIF5B-RET* and two *CCDC6-RET* cases, allowed amplification of reciprocal genomic segments using PCR primers set 1 kb away from the identified *KIF5B-RET* or *CCDC6-RET* breakpoints. This indicated that these fusions were the results of simple reciprocal inversions (cases 1–10 in Table 1, Fig. 2*C*). On the other hand, the remaining seven cases did not allow amplification of genomic segments encompassing the reciprocal breakpoint junctions (cases 11–16 and 18 in Table 1). Three of these seven cases, for which corresponding noncancerous DNA was available, were subjected to LOH analysis at the *RET* locus. LOH was detected at a region proximal (N-terminal) to the breakpoints in all three cases (cases 11, 15, and 16 in Table 1, Fig. 1*A*), indicating nonreciprocal inversion associated with deletion of a copy of the region proximal to the breakpoints. In addition, the inversion in the aforementioned Korean case (case 17) is also nonreciprocal.[4] These data suggested that only a fraction of *RET* fusions (10 of 18, 56%) are caused by simple reciprocal inversions.

Modes of DNA End Joining That Give Rise to Reciprocal Inversions

Two major types of DNA repair pathways cause structural variations.[11,12] The first type is nonhomologous end joining (NHEJ) of DNA double strand breaks (DSBs). which requires very short (a few base pairs) or no homology, and often inserts a few nucleotides at breakpoint junctions.[8,22,23] NHEJ has canonical and noncanonical forms; in the latter, called alternative end joining, DNA ends are joined using microhomology of a few nucleotides at breakpoints.[24] The second type includes repair pathways that use long (>10 bp) homology at DNA ends, such as break-induced replication (BIR) and nonallelic homologous recombination.[12,25]

Sequence analysis of breakpoint-containing genomic segments in 10 reciprocal cases revealed that deletions frequently (8 of 10, 80%) occur in *RET* and/or its partner locus (i.e., *KIF5B* or *CCDC6*) upon DNA end joining (Table 1). This analysis also enabled us to deduce that both types of repair pathways described above are involved in these joining events. In six of the cases (cases 1–6 in Table 1), four DNA ends were joined, and in two cases, insertions were observed (representative cases in Supplementary Fig. 1, Supplementary Digital Content 3, http://links.lww.com/JTO/A543). The lack of significant homology between the sequences of the *RET* and *KIF5B/CCDC6* breakpoints led us to deduce that DNA end joining was mediated by NHEJ in these six cases: two DSBs formed, one each in *RET* and its partner locus, and the four resultant DNA ends were illegitimately joined by canonical or noncanonical NHEJ (Fig. 3*A*).

The remaining four cases (cases 7–10 in Table 1) had a distinctive feature. DNA segments of 33 to 490 bp from either the *RET* or *KIF5B* locus were retained at both the *KIF5B-RET* and *RET-KIF5B* breakpoints, resulting in duplication of these segments. Notably, two regions encompassing the breakpoint in a locus exhibited sequence homology to the duplicated segment of the other locus (representative cases in Supplementary Fig. 2, Supplementary Digital Content 3, http://links.lww.com/JTO/A543). This feature led us to deduce that these joining events were mediated by BIR, using both DNA ends generated by DNA single-strand breaks at the *RET* or fusion-partner locus (Fig. 3*B*). Specifically, two DNA broken ends generated at the *RET* (or partner locus) annealed with the DSB sites of the fusion-partner (or *RET*) locus through sequence homology and were then subjected to ectopic DNA replication. This process left the same DNA segment at both breakpoint junctions, resulting in duplication of the segment.

Speculated Mode of DNA End Joining Giving Rise to Nonreciprocal Inversion

Our study also speculated about the modes of joining involved in the eight remaining cases, which were not likely to have been subjected to simple reciprocal inversion and are therefore defined here as nonreciprocal (cases 11–18 in Table 1). Due to the lack of sequence information from breakpoints in reciprocal counterparts, deletions could not be assessed. The lack of significant homology between the *RET* and *KIF5B/CCDC6* breakpoints suggested the involvement of NHEJ. Consistent with this idea, insertion of a few nucleotides, a common trace of NHEJ, was observed in three cases (cases 11, 15, and 17). A single case (case 16) had an insertion of 349 nucleotides, corresponding to the inverted segment of *RET* exon 7 to intron 7, suggesting the occurrence of an unspecified complex rearrangement mediated by a process other than NHEJ, such as fork stalling and template switching (Lee et al., 2007). These results suggest that the predominant molecular process is illegitimate NHEJ repair, in which two DSBs are formed both in the *RET* and partner loci, and one end of the partner locus (the N-terminal part of *KIF5B* or *CCDC6*) and one end of the *RET* locus (the C-terminal part) are joined by NHEJ. Nevertheless, the remaining two DNA ends were not joined in a simple manner. DNA segments within the DNA ends were either lost or joined with DNA ends other than those at the *RET*, *KIF5B*, and *CCDC6* loci, consistent with the observations of LOH at regions proximal to breakpoints in *RET* (Table 1). In fact, in case 17, the 3′ part of the *KIF5B* gene was fused to the *KIAA1462* gene, 2.0 Mb away from *KIF5B*.[4]

Copyright © 2014 by the International Association for the Study of Lung Cancer

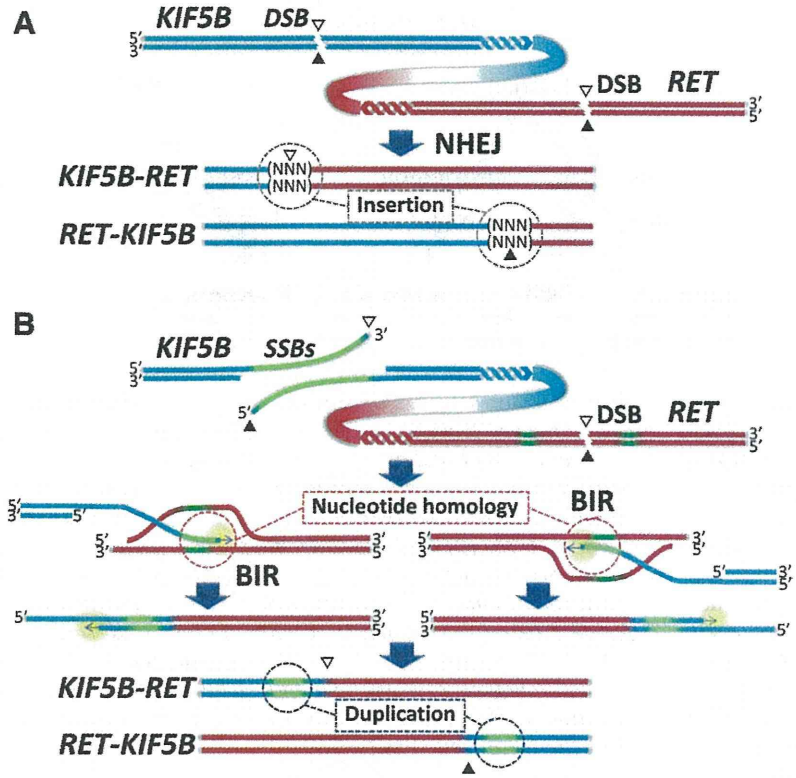


FIGURE 3. Deduced processes of reciprocal inversion by NHEJ and BIR. *A*, NHEJ. Four DNA ends generated by DSBs at *RET* and a partner locus were directly joined. Often, insertions of nucleotides (NNN) at breakpoint junctions are observed. *B*, BIR. Here, DNA single-strand breaks (SSBs) occur in the *KIF5B* locus and a DSB occurs in the *RET* locus. The two SSBs at the *KIF5B* locus trigger BIR by annealing at two homologous sites in the *RET* locus. BIR results in duplication of a *KIF5B* segment. As a result, the *RET* breakpoints in the *KIF5B-RET* and *RET-KIF5B* fusions are located at the same position (a DSB site), whereas the *KIF5B* breakpoints in these fusions are located at different positions (two SSB sites). ▽, breakpoints for partner-*RET* fusion; ▲, breakpoints for *RET*-partner fusion. NHEJ, nonhomologous end joining; BIR, break-induced replication.

DISCUSSION

In this study, we investigated the molecular mechanisms underlying oncogenic *RET* fusion in LADCs. Distribution of breakpoints made us consider a 2.0-kb segment spanning *RET* exon 11 to intron 11 (and also a 5.6-kb segment spanning *KIF5B* intron 15) as a breakpoint cluster region(s). The breakpoints in these regions were dispersed at intervals larger than 4 bp. The inferred breakpoints do not necessarily indicate the sites of actual DNA breaks because resection of nucleotides from DNA ends sometimes occurs during the DNA repair.[23] In fact, we observed nucleotide deletions in eight of 10 LADC cases with reciprocal *KIF5B/CCDC6-RET* inversions. Nevertheless, when the locations of putative breakpoints before DNA end resection were included, the breakpoint distribution remained scattered. These data strongly suggested that the majority of DNA breaks triggering *RET* fusions occur at nonspecific sites in defined regions of a few kb in size. Furthermore, this seems to hold true irrespective of etiology and tumor type: the distribution of breakpoints was not significantly different between ever- and never-smokers, and *RET* exon 11 to intron 11 was also defined as a breakpoint cluster region for *RET* fusions in PTCs, as previously reported.[14–17] The cases shown in Figure 2 (gray and black arrowheads) include PTCs induced by post-Chernobyl irradiation, in which DNA breaks were presumably caused exclusively by irradiation; the random breakpoint distributions in these PTCs were similar to those of the LADCs we analyzed.

We investigated the DNA end-joining pathways that gave rise to *RET* fusions by analyzing the structures of breakpoint junctions. NHEJ was found to be one of the major pathways of DNA end joining. We and others also showed that NHEJ is also prominently involved in interstitial deletions that inactivate tumor-suppressor genes, such as *CDKN2A/p16* and *STK11/LKB1*, in lung cancer.[13,26,27] Thus, NHEJ contributes to the occurrence of driver mutations in both tumor-suppressor genes and oncogenes during lung carcinogenesis. Our data also reveal a possible contribution of BIR in DNA end joining to generate reciprocal inversions. We deduced that BIR occurred from DNA ends, probably generated by DNA single-strand breaks, in the *RET* or partner locus, beginning with annealing with the other locus through nucleotide homologies of tens to hundreds of base pairs. This process resulted in duplication of breakpoint-flanking DNA segments of tens to hundreds of base pairs. BIR has recently been implicated in oncogenic *RAF* fusions in pediatric brain tumors.[28] In those cases, the sequence homology used for annealing of DNA ends was on the order of a few base pairs. Thus, BIR might generate oncogenic fusions frequently, although the detailed process may differ according to tumor type.

Irrespective of the similarities in breakpoint distribution, several processes involved in *RET* fusions differed between LADC and PTC (Fig. 4). Reciprocal inversion was unlikely to have occurred by BIR in PTC because none of the PTC cases exhibited the duplication of DNA segments that were observed in LADC; therefore, the joining of DNA ends in PTC was likely to have been mediated exclusively by NHEJ.[17] This is plausible because *RET* fusions preferentially occur in PTCs in patients suffering from high-dose radiation exposure, suggesting that DSBs generated at the *RET* or partner loci triggered the chromosome rearrangements that generated *RET* fusions.[29] Repetitive NHEJ repair of abundant

Copyright © 2014 by the International Association for the Study of Lung Cancer

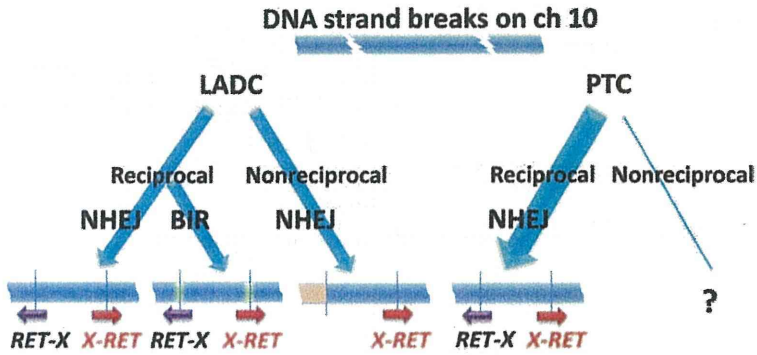


FIGURE 4. Molecular processes underlying *RET* gene fusions in LADC and PTC. Different processes are involved in *RET* fusion in different tumor types. Both reciprocal and nonreciprocal inversions occur in LADC. In LADC, BIR and NHEJ are responsible for DNA end joining in reciprocal inversion, whereas NHEJ is exclusively involved in nonreciprocal inversion. In PTC, reciprocal inversion by NHEJ is dominant. LADC, lung adenocarcinoma; PTC, papillary thyroid carcinoma; NHEJ, nonhomologous end joining; BIR, break-induced replication.

DSBs, which occurs in the context of irradiation, may increase the likelihood of illegitimate repair generating *RET* fusion. On the other hand, in LADC, both DSBs and single-strand breaks formed by multiple causes might trigger rearrangements by multiple DNA repair pathways. The high frequency of nonreciprocal inversion also distinguishes LADC from PTC, for previous study revealed that *RET* fusions result from reciprocal inversion in most cases (43 of 47, 91%).[14,15] Frequent nonreciprocal inversion is consistent with the observation that *KIF5B-RET* fusion-positive tumors contain deletions of the 5′ part of *RET*, as revealed by FISH staining patterns.[1] The present study provides a molecular basis for such a distinct FISH finding and will help to define the criteria used to diagnose *RET*-fusion-positive LADC. Interestingly, FISH analysis also revealed that another driver mutation, *EML4-ALK* fusion, in LADC, caused by a paracentric inversion of chromosome 2, also involves deletion of the 5′ region of the *ALK* oncogene locus.[30,31] Although the structures of breakpoint junctions of *ALK* fusions have not been characterized to the best of our knowledge, these results indicate that a significant fraction of chromosome inversions that cause oncogenic fusions in lung cancer are likely to be nonreciprocal.

Finally, a few issues remain to be elucidated regarding the molecular processes generating oncogenic *RET* fusions. First, although this and previous PTC studies imply that the 2.0-kb region spanning the *RET* exon 11 to intron 11 region is susceptible to DNA strand breaks, the underlying mechanisms remain unknown. For, this region does not exhibit distinctive features known to make DNA susceptible to breaks (Supplementary Fig. 3, Supplementary Digital Content 3, http://links.lww.com/JTO/A543; details in Supplementary Notes, Supplementary Digital Content 2, http://links.lww.com/JTO/A542). Second, the etiological factors that cause DNA strand breaks, and the factors that determine reciprocal or nonreciprocal inversion and selection of DNA repair pathways, also remain unknown. The mode of joining and breakpoint distribution was irrespective of smoking history, and therefore, DNA damage due to smoking is unlikely to be an important factor. The fact that *RET* fusions are more frequent in LADC of never-smokers than in that of ever-smokers indicates that undefined etiological factors play major roles in the occurrence of *RET* fusions.

ACKNOWLEDGMENTS

We thank Hiromi Nakamura, Isao Kurosaka, Sumiko Ohnami, and Sachiyo Mitani of National Cancer Center (NCC) Research Institute for data analysis and technical assistance. The NCC Biobank is supported by the NCC Research and Development Fund of Japan. SNP array analysis was performed by the genome core facility of the NCC. This study was supported in part by Grants-in-Aid for Scientific Research on Innovative Areas (22131006), from the Ministry of Education, Culture, Sports, Science, and Technology of Japan, for the Third-term Comprehensive 10-year Strategy for Cancer Control, from the Ministry of Health, Labor, and Welfare, and for the Program for Promotion of Fundamental Studies in Health Sciences, from the National Institute of Biomedical Innovation (NiBio), and by Management Expenses Grants from the Government to the NCC.

REFERENCES

1. Takeuchi K, Soda M, Togashi Y, et al. RET, ROS1 and ALK fusions in lung cancer. *Nat Med* 2012;18:378–381.
2. Lipson D, Capelletti M, Yelensky R, et al. Identification of new ALK and RET gene fusions from colorectal and lung cancer biopsies. *Nat Med* 2012;18:382–384.
3. Kohno T, Ichikawa H, Totoki Y, et al. KIF5B-RET fusions in lung adenocarcinoma. *Nat Med* 2012;18:375–377.
4. Ju YS, Lee WC, Shin JY, et al. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. *Genome Res* 2012;22:436–445.
5. Gautschi O, Zander T, Keller FA, et al. A patient with lung adenocarcinoma and RET fusion treated with vandetanib. *J Thorac Oncol* 2013;8:e43–e44.
6. Drilon A, Wang L, Hasanovic A, et al. Response to Cabozantinib in patients with RET fusion-positive lung adenocarcinomas. *Cancer Discov* 2013;3:630–635.
7. Kohno T, Tsuta K, Tsuchihara K, Nakaoku T, Yoh K, Goto K. RET fusion gene: translation to personalized lung cancer therapy. *Cancer Sci* 2013;104:1396–1400.
8. Shaw AT, Hsu PP, Awad MM, Engelman JA. Tyrosine kinase gene rearrangements in epithelial malignancies. *Nat Rev Cancer* 2013;13:772–787.
9. Wang R, Hu H, Pan Y, et al. RET fusions define a unique molecular and clinicopathologic subtype of non-small-cell lung cancer. *J Clin Oncol* 2012;30:4352–4359.
10. Suehara Y, Arcila M, Wang L, et al. Identification of KIF5B-RET and GOPC-ROS1 fusions in lung adenocarcinomas through a comprehensive mRNA-based screen for tyrosine kinase fusions. *Clin Cancer Res* 2012;18:6599–6608.
11. Yang L, Luquette LJ, Gehlenborg N, et al. Diverse mechanisms of somatic structural variations in human cancer genomes. *Cell* 2013;153:919–929.
12. Gu W, Zhang F, Lupski JR. Mechanisms for human genomic rearrangements. *Pathogenetics* 2008;1:4.
13. Kohno T, Yokota J. Molecular processes of chromosome 9p21 deletions causing inactivation of the p16 tumor suppressor gene in human cancer: deduction from structural analysis of breakpoints for deletions. *DNA Repair (Amst)* 2006;5:1273–1281.

Copyright © 2014 by the International Association for the Study of Lung Cancer

14. Nikiforov YE, Koshoffer A, Nikiforova M, Stringer J, Fagin JA. Chromosomal breakpoint positions suggest a direct role for radiation in inducing illegitimate recombination between the ELE1 and RET genes in radiation-induced thyroid carcinomas. *Oncogene* 1999;18:6330–6334.
15. Bongarzone I, Butti MG, Fugazzola L, et al. Comparison of the breakpoint regions of ELE1 and RET genes involved in the generation of RET/PTC3 oncogene in sporadic and in radiation-associated papillary thyroid carcinomas. *Genomics* 1997;42:252–259.
16. Minoletti F, Butti MG, Coronelli S, et al. The two genes generating RET/PTC3 are localized in chromosomal band 10q11.2. *Genes Chromosomes Cancer* 1994;11:51–57.
17. Klugbauer S, Pfeiffer P, Gassenhuber H, Beimfohr C, Rabes HM. RET rearrangements in radiation-induced papillary thyroid carcinomas: high prevalence of topoisomerase I sites at breakpoints and microhomology-mediated end joining in ELE1 and RET chimeric genes. *Genomics* 2001;73:149–160.
18. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–595.
19. Cai W, Su C, Li X, et al. KIF5B-RET fusions in Chinese patients with non-small cell lung cancer. *Cancer* 2013;119:1486–1494.
20. Yokota K, Sasaki H, Okuda K, et al. KIF5B/RET fusion gene in surgically-treated adenocarcinoma of the lung. *Oncol Rep* 2012;28:1187–1192.
21. Suzuki M, Makinoshima H, Matsumoto S, et al. Identification of a lung adenocarcinoma cell line with CCDC6-RET fusion gene and the effect of RET inhibitors in vitro and in vivo. *Cancer Sci* 2013;104:896–903.
22. Mahaney BL, Meek K, Lees-Miller SP. Repair of ionizing radiation-induced DNA double-strand breaks by non-homologous end-joining. *Biochem J* 2009;417:639–650.
23. Lieber MR. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu Rev Biochem* 2010;79:181–211.
24. Bennardo N, Cheng A, Huang N, Stark JM. Alternative-NHEJ is a mechanistically distinct pathway of mammalian chromosome break repair. *PLoS Genet* 2008;4:e1000110.
25. Lee JA, Carvalho CM, Lupski JR. A DNA replication mechanism for generating nonrecurrent rearrangements associated with genomic disorders. *Cell* 2007;131:1235–1247.
26. Sasaki S, Kitagawa Y, Sekido Y, et al. Molecular processes of chromosome 9p21 deletions in human cancers. *Oncogene* 2003;22:3792–3798.
27. Matsumoto S, Iwakawa R, Takahashi K, et al. Prevalence and specificity of LKB1 genetic alterations in lung cancers. *Oncogene* 2007;26:5911–5918.
28. Lawson AR, Hindley GF, Forshew T, et al. RAF gene fusion breakpoints in pediatric brain tumors are characterized by significant enrichment of sequence microhomology. *Genome Res* 2011;21:505–514.
29. Hamatani K, Eguchi H, Ito R, et al. RET/PTC rearrangements preferentially occurred in papillary thyroid cancer among atomic bomb survivors exposed to high radiation dose. *Cancer Res* 2008;68:7176–7182.
30. Dai Z, Kelly JC, Meloni-Ehrig A, et al. Incidence and patterns of ALK FISH abnormalities seen in a large unselected series of lung carcinomas. *Mol Cytogenet* 2012;5:44.
31. Yoshida A, Tsuta K, Nitta H, et al. Bright-field dual-color chromogenic in situ hybridization for diagnosing echinoderm microtubule-associated protein-like 4-anaplastic lymphoma kinase-positive lung adenocarcinomas. *J Thorac Oncol* 2011;6:1677–1686.

Copyright © 2014 by the International Association for the Study of Lung Cancer

Int. J. Mol. Sci. **2014**, *15*, 7958-7973; doi:10.3390/ijms15057958

OPEN ACCESS
International Journal of
Molecular Sciences
ISSN 1422-0067
www.mdpi.com/journal/ijms

Review

Mucin 1 Gene (*MUC1*) and Gastric-Cancer Susceptibility

Norihisa Saeki *, Hiromi Sakamoto and Teruhiko Yoshida

Division of Genetics, National Cancer Center Research Institute, Tsukiji 5-1-1, Chuo-ku, Tokyo 104-0045, Japan; E-Mails: hsakamot@ncc.go.jp (H.S.); tyoshida@ncc.go.jp (T.Y.)

* Author to whom correspondence should be addressed; E-Mail: nsaeki@ncc.go.jp; Tel.: +81-3-3542-2511 (ext. 3145); Fax: +81-3-3248-1631.

Received: 3 April 2014; in revised form: 11 April 2014 / Accepted: 21 April 2014 / Published: 7 May 2014

Abstract: Gastric cancer (GC) is one of the major malignant diseases worldwide, especially in Asia. It is classified into intestinal and diffuse types. While the intestinal-type GC (IGC) is almost certainly caused by *Helicobacter pylori* (HP) infection, its role in the diffuse-type GC (DGC) appears limited. Recently, genome-wide association studies (GWAS) on Japanese and Chinese populations identified chromosome 1q22 as a GC susceptibility locus which harbors mucin 1 gene (*MUC1*) encoding a cell membrane-bound mucin protein. *MUC1* has been known as an oncogene with an anti-apoptotic function in cancer cells; however, in normal gastric mucosa, it is anticipated that the mucin 1 protein has a role in protecting gastric epithelial cells from a variety of external insults which cause inflammation and carcinogenesis. HP infection is the most definite insult leading to GC, and a protective function of mucin 1 protein has been suggested by studies on *Muc1* knocked-out mice.

Keywords: gastric cancer; mucin 1; *Helicobacter pylori*; genome-wide association study; single nucleotide polymorphism; cancer susceptibility gene

1. Introduction

Gastric cancer (GC) is one of the major cancers and the second most deadly form of cancer worldwide [1]. Gastric adenocarinoma, a major type of GC, can be histologically classified into two types: intestinal and diffuse, a classification that is thought to reflect its pathogenesis [2]. In the carcinogenesis of the intestinal-type GC (IGC), *Helicobacter pylori* (HP) infection has an important

role. The HP infection results in a sequence of inflammatory change of the gastric epithelium leading to neoplasma: chronic inflammation-intestinal metaplasia-dysplasia-adenocarcinoma [3]. On the other hand, diffuse-type GC (DGC) is thought to develop as a consequence of some genetic change that occurred in gastric stem cells and/or epithelial precursor cells.

In the carcinogenic contribution by HP infection, two bacterial proteins are important: CagA, a product of the cytotoxin-associated gene A, and vacuolating cytotoxin (VacA). There are many excellent review articles on the function of the two proteins. In brief, the proteins induce signaling related to the pro-inflammatory (e.g., interleukin-17, -21 and Nod1), proliferative (e.g., epidermal growth factor-related peptides, EGF receptor, Ras-MAPK pathway, cyclooxygenase 2 and nuclear translocation of β-catenin) and anti-apoptotic (e.g., nuclear factor kappa-B) pathways in the gastric epithelial cells [4].

Consequently, the likelihood is that almost all IGC can be prevented by the eradication of HP infection, and the International Agency of Research on Cancer, sponsored by the World Health Organization, has categorized HP as a class I carcinogen and a definite cause of human gastric cancer, contributing to about 75% of the cases [4,5]. Although the contribution of HP infection is suggested [6], DGC has no established environmental risk factor but does have a tendency to develop in younger people than does IGC, suggesting a genetic factor as a major contributor in its carcinogenesis. Moreover, some countries have a higher prevalence of HP infection but a much lower GC incidence than other countries. Japan, for example, is a country with a high incidence of GC (age-standardized incidence rate 62.7/100,000) but a lower HP seroprevalence (39.3%) than other Asian countries such as Bangladesh (92%) and India (79%), which have a much lower GC incidence, 1.6/100,000 and 5.7/100,000, respectively [7]. The geographical enigma suggests that genetic factors may also contribute to IGC development. With this as a background, three genome-wide association studies (GWASes) were recently performed for detecting the genetic factors related to GC susceptibility, and two of them identified chromosome 1q22 harboring the mucin 1 (*MUC1*) gene as a GC susceptibility locus [8–11].

2. Association between GC (Gastric Cancer) and *MUC1*

A common disease-common variant hypothesis proposes the idea that common and multifactorial diseases are attributed by multiple common genetic variants with a weak to moderate pathogenic effect [12]. Single nucleotide polymorphisms (SNPs) are genetic variants and observed on average once in every 300 nucleotides, which means there are roughly 10 million SNPs in the human genome. Although the Mendelian inheritance law states that separate genetic loci are passed independently of one another from parents to offspring, the SNPs actually descend to offspring as multiple clusters, *i.e.*, many SNPs linked to each other in each chromosome, because there are recombination hotspots in each chromosome when crossing-over events occur during mitosis. This condition, *i.e.*, the SNPs existing as heritable clusters rather than conforming to the Mendelian inheritance law in the genome is called linkage disequilibrium (LD), and current GWASes using SNPs have been exploring genetic susceptibility loci using LD in the genome [13]. In GWASes, an association of SNPs with a disease suggests that the genetic factors or genes exist in the clusters (called LD block or haplotype block) to which the SNPs belong.

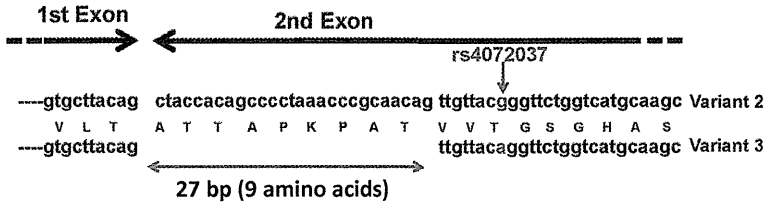
In general, each ethnic population has a distinct set of SNPs and haplotypes. In Japan, the SNPs were already catalogued in the early 2000s as a JSNP (Japanese SNP) database [14,15]. The database contributed to a number of GWASes on genetic factors for common diseases including, for example, lung cancer, myocardial infarction, asthma, intracranial aneurism and Kawasaki disease [16–20].

Japan is a country with one of the highest GC incidences, *i.e.*, it is a common disease in the population. Recently, we performed a GWAS on DGC, which consisted of two steps of the association study [8]. The first step was performed on 85,576 SNPs using 188 DGC cases and 752 references, and the second step on 2753 selected SNPs with 749 DGC cases and 750 controls. Finally, we identified ten SNPs related to DGC with statistical significance, which included four SNPs located in chromosome 8q24.3 and 2 SNPs in 1q22. Haplotype block analyses for detecting the susceptibility genes revealed two candidates at 8q24.3 and 5 at 1q22.

In the 8q24.3 haplotype block, prostate stem cell antigen gene (*PSCA*) was identified as a DGC susceptibility gene, with a significant association between DGC and two SNPs in the gene (rs2976392: 926 cases, 1397 controls, allele-specific odds ratio = 1.71, 95% confidence interval = 1.50–1.94, $p = 1.5 \times 10^{-16}$; rs2294008: 925 cases, 1396 controls, allele-specific odds ratio = 1.67, 95% confidence interval = 1.47–1.90, $p = 2.2 \times 10^{-15}$) [8]. The association was replicated in the Korean population, which has a GC incidence as high as the Japanese (rs2976392: 449 cases, 390 controls, allele-specific odds ratio = 1.90, 95% confidence interval = 1.56–2.33, $p = 8.0 \times 10^{-11}$; rs2294008: 454 cases, 390 controls, allele-specific odds ratio = 1.91, 95% confidence interval = 1.57–2.33, $p = 6.3 \times 10^{-11}$) [8]. *PSCA* also showed a weak correlation to IGC in populations from both Japan (rs2976392: 599 cases, 1397 controls, allele-specific odds ratio = 1.29, 95% confidence interval = 1.12–1.49, $p = 5.0 \times 10^{-4}$) and Korea (rs2976392: 416 cases, 390 controls, allele-specific odds ratio = 1.37, 95% confidence interval = 1.12–1.68, $p = 0.0017$) [8]. Later, the association of rs2976392 or rs2294008 with GC was validated in other Japanese and Korean panels and also in Chinese and Caucasian populations [21–28].

In the other DGC susceptibility locus 1q22, the haplotype block contained five genes, in which we identified the mucin 1 gene (*MUC1*) as the susceptibility gene [9]. A representative SNP in *MUC1*, rs2070803, showed the association with DGC ($p = 2.20 \times 10^{-6}$, adjusted per allele OR = 1.63, 606 cases and 1264 controls), which was replicated in additional Japanese ($p = 3.93 \times 10^{-5}$, OR = 1.81, 304 cases and 1465 controls) and Korean ($p = 2.19 \times 10^{-4}$, OR = 1.82, 452 cases and 372 controls) case-control panels. Moreover, we identified a functional SNP rs4072037 (A/G) in the *MUC1* gene, and the A allele was associated with DGC patients [9]. The SNP influences the splicing of the primary transcripts. We revealed that there are two major *MUC1* transcripts in the gastric epithelium: variants 2 and 3. The rs4072037 located in the 5' side of the second exon determines the splicing acceptor site in the second exon, which in turn determines the type of variants; the G and A alleles result in the expression of variants 2 and 3, respectively (Figure 1) [9,29]. The structural difference between the two variants is nine amino acids in the second exon that are involved in the *N*-terminal signal peptide. This difference in the signal peptide may lead to a difference in the function of the encoded protein between the two splicing variants.

Figure 1. SNP (single nucleotide polymorphism) rs4072037 (G/A, red arrow) in the *MUC1* gene determines the major splicing variants expressed in the gastric mucosa. In the gastric mucosa, major splicing forms were variants 2 and 3, and the allele of SNP rs4072037 is related to the splicing acceptor site selection in the second exon (1st and 2nd exons are indicated by black arrows) and consequently determines the variant type. The variant 2 but not the variant 3 transcript contains the first 27 bp (double-headed red arrow) of the 2nd exon.



In addition to the GWAS conducted in Japan [8,9], GWASes on other ethnic populations also listed 1q22 as a candidate for a GC-related locus (Table 1). The GWAS on the Chinese population revealed the association between the rs4072037 in *MUC1* and GC (rs4072037; OR = 0.75, $p = 4.22 \times 10^{-7}$) [11].

Table 1. Association between GC (gastric cancer) and *MUC1* SNPs (single nucleotide polymorphisms). Ref. = Reference.

SNPs (Major/Minor)	Risk Allele	Odds Ratio (95% CI) and Genotype	*p* Value	Ethnic	Cancer Type	Ref.
rs4072037 (A/G)	A	1.62 (1.32–1.99) [#] A to G, allelic	4.04×10^{-6}	Japanese	DGC	[9]
rs4072037 (A/G)	A	1.74 (1.26–2.39) [#] A to G, allelic	7.82×10^{-4}	Korean	DGC	[9]
rs4072037 (A/G)	A	0.78 (0.67–0.91) AG to AA	0.031	Korean	All	[30]
rs4072037 (A/G)	A	0.72 (0.62–0.85) G to A, allelic	5.74×10^{-5}	Chinese	Non-cardia	[11]
rs4072037 (A/G)	A	0.75 (0.65–0.87) G to A, allelic	9.45×10^{-5}	Chinese	Cardia	[11]
rs4072037 (A/G)	A	0.73 G to A, allelic	1.0×10^{-4}	Chinese	Non-cardia	[10]
rs4072037 (A/G)	A	1.81 AA to AG + GG	0.031	Chinese	All	[31]
rs2070803 (G/A)	G	0.46 (0.32–0.67) AA + AG to GG	<0.001	Chinese	All	[32]
rs4072037 (G/A)	A	2.20 (1.41–3.44) AA to GG	<0.01	Caucasian	All	[33]
rs4072037 (G/A)	A	0.5 (0.3–0.9) AG to AA	-	Caucasian	Cardia	[34]
rs4072037 (G/A)	A	0.4 (0.2–0.9) AG to AA	-	Caucasian	Non-cardia, Intestinal	[34]

[#] additive model.

Besides the GWAS, the association of SNPs in *MUC1* with GC has been demonstrated in other ethnic populations, especially in Chinese (Table 1). An association study with imputation analysis on Chinese case-control samples demonstrated the association of the SNP (OR = 0.73, $p = 1.0 \times 10^{-4}$) [10]. In a study on 300 cases and 300 controls, the association was also successfully replicated (rs2070803 AA/AG to GG, OR = 0.46, the permutation $p < 0.001$) [32]. Still another study on the Chinese (138 cases and 241 controls) showed the association (rs4072037 AA against AG + GG, OR = 1.81, 95% CI = 1.06–3.12) [31]. In addition to the Chinese populations, the association was also replicated in the Korean population (3245 cases and 1700 controls, rs4072037 AG to AA, OR = 0.78, 95% CI = 0.67–0.91) [30]. Moreover, it was replicated in a study on a Caucasian population (290 cases and 376 controls) in which an association between rs4072037 and non-cardia intestinal GC was demonstrated (OR = 0.4, 95% CI = 0.2–0.9) [34]. Another study on 273 cases and 377 controls also revealed the association (rs4072037 AA against GG, OR = 2.20, the permutation $p < 0.01$) [33]. Finally, a meta-analysis on the data obtained in the association studies with Asian or European ethnicities showed an association of rs4072037 with both IGC (G allele, OR = 0.74, 95% CI = 0.66–0.83, p value of Z-test = 1.79×10^{-7}) and DGC (G allele, OR = 0.66, 95% CI = 0.58–0.74, p value of Z-test = 1.29×10^{-7}) [35]. It is noteworthy that the A allele was associated with GC and is a major allele in the Japanese, Chinese and Korean populations, which have a high GC incidence, but a minor one in a European population with a low GC incidence.

Surprisingly, an association between *MUC1* gene polymorphisms other than SNP and GC has also been demonstrated in other studies previous to the GWASes. The *MUC1* gene has a variable tandem repeat region, which results in large (L) and small (S) alleles shown in Southern blot analyses when DNA samples are digested with restriction enzymes. It was demonstrated in a Caucasian population (159 GC cases and 324 controls) that SS genotypes of *MUC1* had an increased risk of developing GC (SS to LL, OR = 4.3, 95% CI = 1.8–10.5, $p < 0.0001$) [36], and the two alleles, the S and the A of rs4072037, as well as the L and the G of the SNP are in LD, respectively, in Japanese and European populations [9,29]. The association in different ethnic populations strongly supports the suggestion that *MUC1* is a GC susceptibility gene.

3. MUC1 Expression in Gastric Carcinogenesis

Several immunohistochemical studies identified the mucin 1 protein in normal and malignant gastric epithelial cells. However, the pattern of the staining for the protein was a little different depending on the antibodies used in the studies, which is likely to have originated from variableness in the glycosylation state of the antigen used in raising the antibodies. In summary, the MUC1 protein was observed in the surface foveolar cells in the entire stomach, in mucous neck cells and chief cells of the gastric fundus and antrum, and also in the pyloric gland, typically in the manner of staining at the apical side of the cell membrane and also diffusely in cytoplasm [37–39]. An immunohistochemical study using two anti-mucin 1 antibodies, HMFG1 reacting with the fully glycosyrated mucin 1 protein and SM3 reacting with the under-glycosylated protein, revealed a zonal pattern of the glycosylation state of the protein [40]. The HMFG1 stained the protein in the foveolar cells of the antrum but not of the corpus. On the other hand, staining of SM3 was limited to the perinuclear area of the foveolar cells of the antrum.

There are many immunohistochemical studies on MUC1 expression in GC, and most of them reported MUC1 staining in roughly more than 50% of both IGC and DGC except for signet-ring cell carcinoma, a poorly differentiated GC in which the MUC1 expression was observed only 10% (Table 2) [38–46]. Although MUC1 staining seems to be related to a better differentiation state of the tumor cells since it can be considered as a differentiation marker, most of the reports suggested an association of MUC1 expression with a worse prognosis. It was also reported that abnormal E-cadherin expression in tumor cells was correlated to MUC1 expression, which was observed in the cases of poor prognosis or advanced stage [47,48]. Downregulation of MUC1 was observed in pre-cancerous lesion. There are two types of intestinal metaplasia, complete and incomplete: the former has fully developed intestinal goblet cells and enterocytes with a brush border and the latter has no absorptive cells [49]. Several studies revealed none, or a marked reduction of MUC1 expression in the tissues of complete intestinal metaplasia, a pre-neoplastic condition, although it was expressed in the incomplete type [37,39,40,50,51]. The suppression in the pre-neoplastic lesion and the frequent reactivation in GC of MUC1 expression, especially in the cases with a poor prognosis, implicated its distinct function in normal gastric epithelial cells and in GC cells.

The structure and function of the promoter region of *MUC1* gene have been elucidated. It contains responsive elements for several signalings executed by external molecules, such as transforming growth factor-β and interferon-γ. Moreover, hypomethylation of the tandem-repeat region is required for *MUC1* gene expression in epithelial tissues [52].

4. MUC1 Function in Normal Gastric Epithelial Cells

MUC1 belongs to the mucin family (MUC1 to MUC21), which consists of secretory and membrane-bound types, and MUC1 is the latter [53]. In normal epithelial cells, MUC1 is located at the apical surface of the cells and acts as a barrier against exogenous insults to the cells [54]. The MUC1 protein on the cell surface consists of *N*- and *C*-terminal subunits, designated as MUC1-N and MUC1-C, respectively. After being translated, a single MUC1 protein is cleaved to the two subunits by autoproteolysis, but both the subunits remain associated by non-covalent binding and are localized to the cell membrane. MUC1-N, present on the cell surface, has multiple glycosylation sites and has a protective role for cells against many types of insults [55].

HP infection is a definite carcinogen for gastric epithelial cells, leading to carcinogenesis, and there is experimental and epidemiological evidence for the role of MUC1 in protecting the gastrointestinal tract from bacterial infection. *Muc1* knocked-out (KO) mice with oral infection of *Campylobactor jejuni*, showed damage in the small intestine as well as systemic infection more frequently than did the wild type [56]. A study on *Muc1*-deficient cultured cells and mice demonstrated that mucin 1 protected the gastric epithelium from both non-MUC1 binding bacteria (by inhibiting adhesion to the cell surface with its steric hindrance effect) and MUC1-binding bacteria (by acting as a releasable decoy) [57]. In one study, mice lacking *Muc1* were colonized by five-fold more HP within one day of infection, and developed an atrophic gastritis marked by loss of parietal cells, although wild-type mice developed only a mild gastritis, when infected for two months with HP [58].

Table 2. MUC1 expression in gastric cancer observed by immunohistochemistry.

Study	MUC1 Staining						Note	Correlation to Clinical Information
	Intestinal		Diffuse		Intestinal + Diffuse			
	Case No.	(%)	Case No.	(%)	Case No.	(%)		
Ho, *et al.* [38]	-	-	-	-	25/33	(75.8)	-	-
Reis, *et al.* [40]	31/31	(100)	24/24	(100)	-	-	fully glycosylated MUC1	lymphatic invasion *, nodal metastasis *, advanced stage
-	73/90	(81.1)	30/49	(61.2)	-	-	under-glycosylated MUC1	wall penetration, lymphatic invasion *, nodal metastasis, advanced stage
Utsunomiya, *et al.* [39]	(60/68)	(88)	(45/68)	(66)	-	-	fully glycosylated MUC1	worse prognosis *
Lee, *et al.* [41]	37/113	(32.7)	28/159	(17.6)	-	-	-	worse prognosis *
Wang, *et al.* [42]	13/21	(61.9)	11/17	(64.7)	-	-	-	better prognosis
Wang, *et al.* [43]	14/26	(53.8)	30/44	(68.2)	-	-	-	worse prognosis *
Kocer, *et al.* [45]	10/16	(62.5)	13/19	(68.4)	-	-	-	worse prognosis *
Barresi, *et al.* [44]	23/27	(85.2)	3/10	(30)	-	-	-	-
Terada, *et al.* [46]	-	-	3/30	(10)	-	-	signet-ring cell carcinoma	-

* Statistically significant correlation was demonstrated.

As mentioned previously, our study demonstrated that rs4072037 determines a major variant expressed in the stomach by influencing the splicing acceptor site of the second exon (Figure 1) [9]. It is likely that rs4072037 affects the barrier function in the stomach of individuals through this determination of a major variant. In addition, our study revealed that rs4072037 also influences the transcriptional activity of the *MUC1* gene promoter; the A allele associated with GC reduced the transcriptional activity, which may result in decreased MUC1 expression [9]. These findings suggest that rs4072037 influences the quantity and/or the quality of the MUC1 protein, which causes a difference in its barrier function in the stomach and subsequently the difference in GC susceptibility between individuals. Indeed, it was reported on Caucasians that those having the S allele of *MUC1*, which is linked to the A allele of rs4072037, were more susceptible to HP gastritis than the people with the L allele [59]. A study on a Chinese population revealed that HP seropositivity and AA genotypes for rs4072037 synergistically enhance the risk of GC [60]. In the study, compared to the subjects with HP seronegativity and the AG or GG genotype, those with HP seropositivity and the AG or GG genotype had more risk (OR = 2.30, 95% CI = 1.23–4.31, $p = 0.017$), and those with HP seropositivity and the AA genotype has significant risk (OR = 3.95, 95% CI = 2.29–6.79, $p = 6.5 \times 10^{-6}$). However, as the risk of those with HP seronegativity and the AA genotype was also increased (OR = 2.46, 95% CI = 1.42–4.27, $p = 0.003$), it is certain that the genotype would also contribute to GC development in an HP-independent manner. The effect of HP seropositivity and rs4072037 state is summarized in Table 3 [9,59,60].

Table 3. Effect of HP (*Helicobacter pylori)* infection and MUC1 polymorphism on GC risk [60].

Factors		GC Risk	
MUC1 polymorphism	rs4072037	GG, AG	AA
	tandem-repeat [59]	LL, LS	SS
	splicing variant [9]	2/2, 2/3	3/3
HP infection	seronegative	1.00 (reference)	2.46 (1.42–4.27) [#]
	seropositive	2.30 (1.23–4.31) [#]	3.95 (2.29–6.79) [#]

[#] Odds ratio (95% CI).

Besides the protective function as a mucosal barrier, MUC1 may have an anti-carcinogenic role in another manner. As previously mentioned, the MUC1 protein consists of *N*- and *C*-terminal subunits, MUC1-N and MUC1-C. MUC1-C has a transmembrane domain and a cytoplasmic tail (CT), which contains several phosphorylation sites and a β-catenin binding site. Phosphorylation of threonine contained in the CT promotes interactions between MUC1 and β-catenin, and leads to a nuclear localization of the complex, resulting in regulation of genes including *p53* [61,62]. Namely, the CT is involved in subcellular signal transduction. Recently it has been suggested that the HP virulence factor CagA destabilizes the E-cadherin/β-catenin complex located in the cytoplasm of epithelial cells and enhances an accumulation of β-catenin in the nucleus [63]. The nuclear accumulation of β-catenin activates beta-catenin-dependent genes, such as *CDX1*, which encodes an intestinal specific transcription factor, and induces aberrant expression of molecules in gastric epithelial cells, including an intestinal-differentiation marker, goblet-cell mucin MUC2, which contributes to the development of

intestinal metaplasia, a pre-neoplastic lesion [64]. In addition, the nuclear accumulation of β-catenin also activates interlukin-8 expression, a chemotactic and inflammatory cytokine [65]. It is hypothesized that MUC1 binds to β-catenin and attenuates its nuclear accumulation [66,67]. Intriguingly, it was demonstrated that HP upregulates MUC1 expression in gastric cancer cells through STAT3 and CpG hypomethylation [68]. This cascade may exist in the normal gastric epithelium as an anti-carcinogenic mechanism against HP infection. It was reported that HP infection upregulates MUC2, MUC5AC and MUC6 genes in KATO-III, a cultured gastric cancer cell line [69]; however, it was demonstrated that HP infection reduced the rate of mucin turnover and decreased the levels of Muc1 in the gastric mucosa of mice [70].

5. MUC1 Function in Gastric Carcinogenesis

Contrary to its protective function in normal gastric epithelial cells, the two findings mentioned above suggest a different function of MUC1 in GC cells: the gene is silenced in intestinal metaplasia, a pre-neoplastic lesion, but frequently reactivated in GC, and its expression is correlated to poor prognosis. Indeed, MUC1 has been considered as an oncoprotein, because there is accumulating evidence which suggests its cancer-promoting function.

It was reported that, interacted with Kruppel-like factor 4 (KLF4), a MUC1 *C*-terminal subunit (MUC1-C) occupies the PE21 element of the *p53* gene promoter, which recruits histone deacetylases, and suppresses the transcription of the *p53* gene [71]. *p53* is one of the representative tumor suppressor genes functioning in apoptosis, genomic stability and the inhibition of angiogenesis. It is a master guardian and executioner that surveys genetic damage and responds to it by arresting the cell cycle and facilitating DNA damage repair, or by induction of cell death when the genetic damage is severe [72]. MUC1 activates anti-apoptotic protein Bcl-xL and attenuates the loss of mitochondrial transmembrane potential, mitochondrial cytochrome c release and caspase-9 activation, leading to the failure of apoptosis induction [73]. In response to DNA damage, the non-receptor c-Abl tyrosine kinase is translocated to the nucleus and induces apoptosis of the cells, but MUC1 protein attenuates this nuclear translocation [74]. As stated above, MUC1 is a tremendous oncoprotein that destroys apoptosis execution pathways, one of the most important anti-cancer machines contained in the cells. The anti-apoptotic function of molecules confers cancer cells with resistance to genotoxic anticancer drugs.

MUC1 may contribute to metastasis, as it was demonstrated *in vitro* that the MUC1 protein can bind to intercellular adhesion molecule-1 (ICAM-1), which facilitates adhesion of breast cancer cells to endothelial cells, leading to adhesion and subsequent migration through the vessel wall [75].

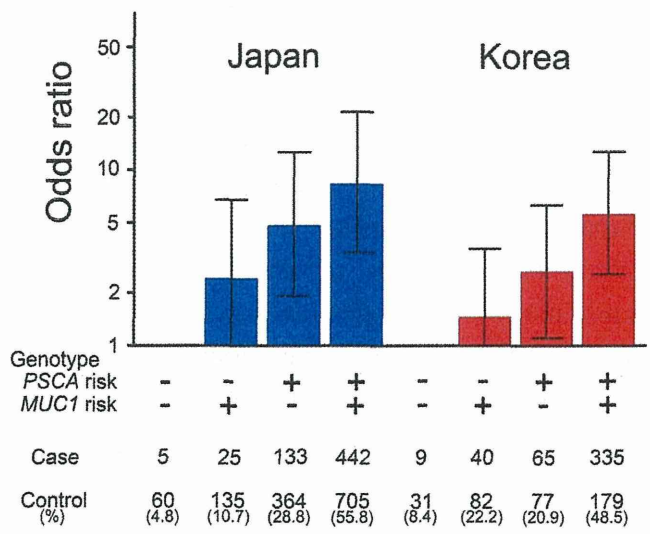
Moreover, MUC1 could have some role in GC stem cells, as it acts as a growth factor receptor on undifferentiated human embryonic stem cells and is expressed in acute myeloid leukemia stem cells [76,77]. Intriguingly, it is also known that MUC1 facilitates cancer cell survival under hypoxic and nutrient-deprived conditions by regulating glucose and lipid metabolism and the cellular energy state [78].

As previously mentioned, in normal epithelial cells, it is likely that the G allele of rs4072037 contributes to increasing MUC1 expression and maybe also to enhancing the quality of MUC1 protein. It would be interesting to know whether the G allele is correlated to a poor prognosis in GC, but no study on the relation of the SNP and a GC prognosis has yet been conducted.

6. Perspective and Conclusions

Needless to say, prevention is the best way for us to cope with diseases. GC susceptibility genes, which have been and will continue to be identified, may contribute to GC prevention because a population can be stratified based on the GC susceptibility defined by the genes. The stratification enables us to intervene in the subpopulations by, for example, modulating the intensity of health check procedures according to their GC-development risk: health check with an endoscopic examination every two years starting at age 40 or with an examination every six months starting at age 20. Interestingly, the Japanese and Korean populations can be stratified using the two GC susceptibility genes (Figure 2). The combined genotype association data of rs2294008 in *PSCA* and rs4072037 in *MUC1* suggested that 4.8% of the Japanese population has the risk genotype of rs4072037, 28.8% the risk genotype of rs2294008 and 55.8% with both, and that the people of the double risk genotype has the highest risk for GC development (OR = 8.4) [9,79]. Unfortunately, the DNA samples used in this study were not linked with the information on HP infection, but it is likely that the highest risk group can be further stratified based on HP infection state and other environmental factors. The combination of HP detection and the identification of the rs4072037 and rs2294008 genotypes may contribute to individual risk evaluation and GC prevention.

Figure 2. Japanese and Korean populations can be stratified based on *PSCA* and *MUC1* genotypes associated with risk for DGC. The stratification and risk estimation were performed using genotype data of rs2294008 in *PSCA* and rs4072037 in *MUC1* [9]. The risk allele's effect is assumed to be dominant for rs2294008 in *PSCA* (risk genotype: TT and TC; protective genotype: CC) and recessive for rs4072037 in *MUC1* (risk genotype: AA; protective genotype: GG and GA). Bar, upper and lower bounds of 95% confidence interval.



In this context, finding environmental factors is also important, as not all of the members of the high risk group with OR = 8.4—corresponding to about half of the Japanese population—contract GC. A stratification based on the genotype of GC susceptibility genes may contribute to investigating the environmental factors, as it enables us to concentrate on exploring those that have a critical effect on the high risk group for GC development.

In conclusion, identification of GC susceptibility genes can contribute to GC prevention. To date, aside from 1q22 and 8q24.3, 3 GC-associated loci, chromosome 3q13.31, 5p13.1 and 10q23, have been found [79]. To realize preventive intervention based on genetic risk, additional GC susceptibility genes should be identified in and outside of the loci in order to stratify the population in a more detailed manner.

Acknowledgments

This review is based on research grants from the Ministry of Education, Culture, Sports, Science and Technology, Japan (JST grant for the personalized medicine project) and Grants-in-Aid for Scientific Research (KAKENHI) by the Japan Society for the Promotion of Science (No. 23501327).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Brenner, H.; Rothenbacher, D.; Arndt, V. Epidemiology of gastric cancer. In *Methods of Molecular Biology, Cancer Epidemiology*; Verma, M., Ed.; Humana Press: Totowa, NJ, USA, 2009; Volume 472, p. 467.
2. Yasui, W.; Sentani, K.; Motoshita, J.; Nakayama, H. Molecular pathobiology of gastric cancer. *Scand. J. Surg.* **2006**, *95*, 225–231.
3. Peek, R.M., Jr.; Blaser, M.J. Helicobacter pylori and gastrointestinal tract adenocarcinomas. *Nat. Rev. Cancer* **2002**, *2*, 28–37.
4. Ricci, V.; Romano, M.; Boquet, P. Molecular cross-talk between *Helicobacter pylori* and human gastric mucosa. *World J. Gastroenterol.* **2011**, *17*, 1383–1399.
5. Forman, D. *Helicobacter pylori* and gastric cancer. *Scand. J. Gastroenterol. Suppl.* **1996**, *220*, 23–26.
6. Pilpilidis, I.; Kountouras, J.; Zavos, C.; Katsinelos, P. Upper gastrointestinal carcinogenesis: H. pylori and stem cell cross-talk. *J. Surg. Res.* **2011**, *166*, 255–264.
7. Fock, K.M.; Ang, T.L. Epidemiology of *Helicobacter pylori* infection and gastric cancer in Asia. *J. Gastroenterol. Hepatol.* **2010**, *25*, 479–486.
8. Sakamoto, H.; Yoshimura, K.; Saeki, N.; Katai, H.; Shimoda, T.; Matsuno, Y.; Saito, D.; Sugimura, H.; Tanioka, F.; Kato S.; *et al.* Genetic variation in *PSCA* is associated with susceptibility to diffuse-type gastric cancer. *Nat. Genet.* **2008**, *40*, 730–740.
9. Saeki, N.; Saito, A.; Choi, I.J.; Matsuo, K.; Ohnami, S.; Totsuka, H.; Chiku, S.; Kuchiba, A.; Lee, Y.S.; Yoon, K.A.; *et al.* A functional single nucleotide polymorphism in *mucin 1*, at chromosome 1q22, determines susceptibility to diffuse-type gastric cancer. *Gastroenterology* **2011**, *140*, 892–902.
10. Shi, Y.; Hu, Z.; Wu, C.; Dai, J.; Li, H.; Dong, J.; Wang, M.; Miao, X.; Zhou, Y.; Lu, F.; *et al.* A genome-wide association study identifies new susceptibility loci for non-cardia gastric cancer at 3q13.31 and 5p13.1. *Nat. Genet.* **2011**, *43*, 1215–1218.

11. Abnet, C.C.; Freedman, N.D.; Hu, N.; Wang, Z.; Yu, K.; Shu, X.O.; Yuan, J.M.; Zheng, W.; Dawsey, S.M.; Dong, L.M.; *et al.* A shared susceptibility locus in *PLCE1* at 10q23 for gastric adenocarcinoma and esophageal squamous cell carcinoma. *Nat. Genet.* **2010**, *42*, 764–767.
12. Gibson, G. Rare and common variants: Twenty arguments. *Nat. Rev. Genet.* **2012**, *13*, 135–145.
13. Palmer, L.J.; Cardon, L.R. Shaking the tree: Mapping complex disease genes with linkage disequilibrium. *Lancet* **2005**, *366*, 1223–1234.
14. 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*, 158–162.
15. Yoshida, T.; Ono, H.; Kuchiba, A.; Saeki, N.; Sakamoto, H. Genome-wide germline analyses on cancer susceptibility and GeMDBJ database: Gastric cancer as an example. *Cancer Sci.* **2010**, *101*, 1582–1589.
16. Miki, D.; Kubo, M.; Takahashi, A.; Yoon, K.A.; Kim, J.; Lee, G.K.; Zo, J.I.; Lee, J.S.; Hosono, N.; Morizono, T.; *et al.* Variation in TP63 is associated with lung adenocarcinoma susceptibility in Japanese and Korean populations. *Nat. Genet.* **2010**, *42*, 893–896.
17. Aoki, A.; Ozaki, K.; Sato, H.; Takahashi, A.; Kubo, M.; Sakata, Y.; Onouchi, Y.; Kawaguchi, T.; Lin, T.H.; Takano, H.; *et al.* SNPs on chromosome 5p15.3 associated with myocardial infarction in Japanese population. *J. Hum. Genet.* **2011**, *56*, 47–51.
18. Hirota, T.; Takahashi, A.; Kubo, M.; Tsunoda, T.; Tomita, K.; Doi, S.; Fujita, K.; Miyatake, A.; Enomoto, T.; Miyagawa, T.; *et al.* Genome-wide association study identifies three new susceptibility loci for adult asthma in the Japanese population. *Nat. Genet.* **2011**, *43*, 893–896.
19. Low, S.K.; Takahashi, A.; Cha, P.C.; Zembutsu, H.; Kamatani, N.; Kubo, M.; Nakamura, Y. Genome-wide association study for intracranial aneurysm in the Japanese population identifies three candidate susceptible loci and a functional genetic variant at EDNRA. *Hum. Mol. Genet.* **2012**, *21*, 2102–2110.
20. Onouchi, Y.; Ozaki, K.; Burns, J.C.; Shimizu, C.; Terai, M.; Hamada, H.; Honda, T.; Suzuki, H.; Suenaga, T.; Takeuchi, T.; *et al.* A genome-wide association study identifies three new risk loci for Kawasaki disease. *Nat. Genet.* **2012**, *44*, 517–521.
21. Wu, C.; Wang, G.; Yang, M.; Huang, L.; Yu, D.; Tan, W.; Lin, D. Two genetic variants in prostate stem cell antigen and gastric cancer susceptibility in a Chinese population. *Mol. Carcinog.* **2009**, *48*, 1131–1138.
22. Matsuo, K.; Tajima, K.; Suzuki, T.; Kawase, T.; Watanabe, M.; Shitara, K.; Misawa, K.; Ito, S.; Sawaki, A.; Muro, K.; *et al.* Association of prostate stem cell antigen gene polymorphisms with the risk of stomach cancer in Japanese. *Int. J. Cancer* **2009**, *125*, 1961–1964.
23. Lu, Y.; Chen, J.; Ding, Y.; Jin, G.; Wu, J.; Huang, H.; Deng, B.; Hua, Z.; Zhou, Y.; Shu, Y.; *et al.* Genetic variation of PSCA gene is associated with the risk of both diffuse- and intestinal-type gastric cancer in a Chinese population. *Int. J. Cancer* **2010**, *127*, 2183–2189.
24. Ou, J.; Li, K.; Ren, H.; Bai, H.; Zeng, D.; Zhang, C. Association and haplotype analysis of prostate stem cell antigen with gastric cancer in Tibetans. *DNA Cell Biol.* **2010**, *29*, 319–323.
25. Lochhead, P.; Frank, B.; Hold, G.L.; Rabkin, C.S.; Ng, M.T.; Vaughan, T.L.; Risch, H.A.; Gammon, M.D.; Lissowska, J.; Weck, M.N.; *et al.* Genetic variation in the prostate stem cell antigen gene and upper gastrointestinal cancer in white individuals. *Gastroenterology* **2011**, *140*, 435–441.

26. Zeng, Z.; Wu, X.; Chen, F.; Yu, J.; Xue, L.; Hao, Y.; Wang, Y.; Chen, M.; Sung, J.J.; Hu, P. Polymorphisms in prostate stem cell antigen gene rs2294008 increase gastric cancer risk in Chinese. *Mol. Carcinog.* **2011**, *50*, 353–358.
27. Song, H.R.; Kim, H.N.; Piao, J.M.; Kweon, S.S.; Choi, J.S.; Bae, W.K.; Chung, I.J.; Park, Y.K.; Kim, S.H.; Choi, Y.D.; *et al.* Association of a common genetic variant in prostate stem-cell antigen with gastric cancer susceptibility in a Korean population. *Mol. Carcinog.* **2011**, *50*, 871–875.
28. Sala, N.; Muñoz, X.; Travier, N.; Agudo, A.; Duell, E.J.; Moreno, V.; Overvad, K.; Tjonneland, A.; Boutron-Ruault, M.C. Clavel-Chapelon, F.; *et al.* Prostate stem-cell antigen gene is associated with diffuse and intestinal gastric cancer in Caucasians: Results from the EPIC-EURGAST study. *Int. J. Cancer* **2011**, *130*, 2417–2427.
29. Ng, W.; Loh, A.X.; Teixeira, A.S.; Pereira, S.P.; Swallow, D.M. Genetic regulation of *MUC1* alternative splicing in human tissues. *Br. J. Cancer* **2008**, *99*, 978–985.
30. Song, H.R.; Kim, H.N.; Kweon, S.S.; Choi, J.S.; Shim, H.J.; Cho, S.H.; Chung, I.J.; Park, Y.K.; Kim, S.H.; Choi, Y.D.; *et al.* Common genetic variants at 1q22 and 10q23 and gastric cancer susceptibility in a Korean population. *Tumour Biol.* **2014**, *35*, 3133–3137.
31. Xu, Q.; Yuan, Y.; Sun, L.P.; Gong, Y.H.; Xu, Y.; Yu, X.W.; Dong, N.N.; Lin, G.D.; Smith, P.N.; Li, R.W. Risk of gastric cancer is associated with the MUC1 568 A/G polymorphism. *Int. J. Oncol.* **2009**, *35*, 1313–1320.
32. Li, F.; Zhong, M.Z.; Li, J.H.; Liu, W.; Li, B. Case-control study of single nucleotide polymorphisms of *PSCA* and *MUC1* genes with gastric cancer in a Chinese. *Asian Pac. J. Cancer Prev.* **2012**, *13*, 2593–2596.
33. Jia, Y.; Persson, C.; Hou, L.; Zheng, Z.; Yeager, M.; Lissowska, J.; Chanock, S.J.; Chow, W.H.; Ye, W. A comprehensive analysis of common genetic variation in *MUC1*, *MUC5AC*, *MUC6* genes and risk of stomach cancer. *Cancer Causes Control* **2010**, *21*, 313–321.
34. Palmer, A.J.; Lochhead, P.; Hold, G.L.; Rabkin, C.S.; Chow, W.H.; Lissowska, J.; Vaughan, T.L.; Berry, S.; Gammon, M.; Risch, H.; *et al.* Genetic variation in *C20orf54*, *PLCE1* and *MUC1* and the risk of upper gastrointestinal cancers in Caucasian populations. *Eur. J. Cancer Prev.* **2013**, *21*, 541–544.
35. Zheng, L.; Zhu, C.; Gu, J.; Xi, P.; Du, J.; Jin, G. Functional polymorphism rs4072037 in MUC1 gene contributes to the susceptibility to gastric cancer: evidence from pooled 6580 cases and 10,324 controls. *Mol. Biol. Rep.* **2013**, *40*, 5791–5796.
36. Carvalho, F.; Seruca, R.; David, L.; Amorim, A.; Seixas, M.; Bennett, E.; Clausen, H.; Sobrinho-Simões, M. *MUC1* gene polymorphism and gastric cancer—An epidemiological study. *Glycoconj. J.* **1997**, *14*, 107–111.
37. Ho, S.B.; Niehans, G.A.; Lyftogt, C.; Yan, P.S.; Cherwitz, D.L.; Gum, E.T.; Dahiya, R.; Kim, Y.S. Heterogeneity of mucin gene expression in normal and neoplastic tissues. *Cancer Res.* **1993**, *53*, 641–651.
38. Ho, S.B.; Shekels, L.L.; Toribara, N.W.; Kim, Y.S.; Lyftogt, C.; Cherwitz, D.L.; Niehans, G.A. Mucin gene expression in normal, preneoplastic, and neoplastic human gastric epithelium. *Cancer Res.* **1995**, *55*, 2681–2690.