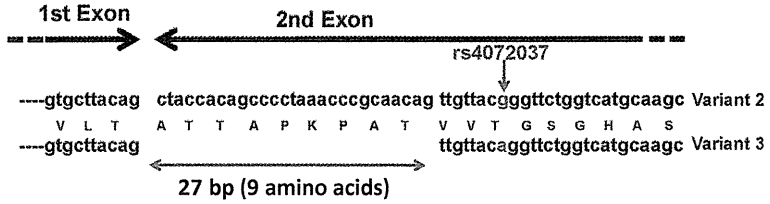
In general, each ethnic population has a distinct set of SNPs and haplotypes. In Japan, the SNPs were already catalogued in the early 2000s as a JSNP (Japanese SNP) database [14,15]. The database contributed to a number of GWASes on genetic factors for common diseases including, for example, lung cancer,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

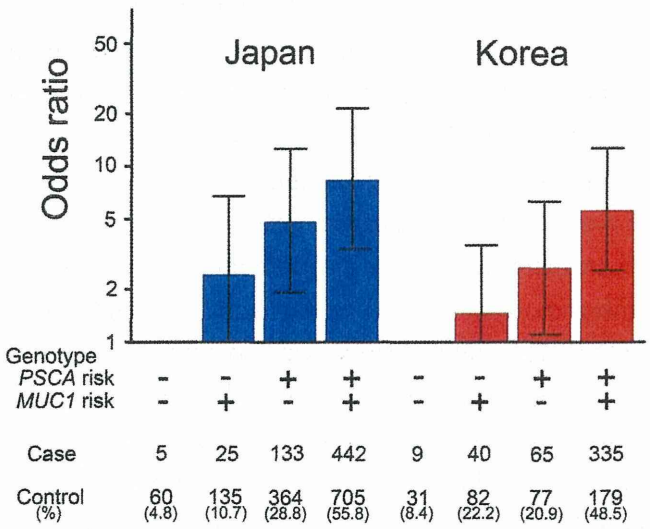
Moreover, MUC1 could have some role in GC stem cells, as it acts as a growth factor receptor on undifferentiated human embryonic stem cells and is expressed in acute myeloid leukemia stem cells [76,77]. Intriguingly, it is also known that MUC1 facilitates cancer cell survival under hypoxic and nutrient-deprived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

39. Utsunomiya, T.; Yonezawa, S.; Sakamoto, H.; Kitamura, H.; Hokita, S.; Aiko, T.; Tanaka, S.; Irimura, T.; Kim, Y.S.; Sato, E. Expression of MUC1 and MUC2 mucins in gastric carcinomas: its relationship with the prognosis of the patients. *Clin. Cancer Res.* **1998**, *4*, 2605–2614.
40. Reis, C.A.; David, L.; Seixas, M.; Burchell, J.; Sobrinho-Simões, M. Expression of fully and under-glycosylated forms of MUC1 mucin in gastric carcinoma. *Int. J. Cancer* **1998**, *79*, 402–410.
41. Lee, H.S.; Lee, H.K.; Kim, H.S.; Yang, H.K.; Kim, Y.I.; Kim, W.H. *MUC1*, *MUC2*, *MUC5AC*, and *MUC6* expressions in gastric carcinomas: Their roles as prognostic indicators. *Cancer* **2001**, *92*, 1427–1434.
42. Wang, R.Q.; Fang, D.C. Alterations of *MUC1* and *MUC3* expression in gastric carcinoma: Relevance to patient clinicopathological features. *J. Clin. Pathol.* **2003**, *56*, 378–384.
43. Wang, J.Y.; Chang, C.T.; Hsieh, J.S.; Lee, L.W.; Huang, T.J.; Chai, C.Y.; Lin, S.R. Role of *MUC1* and *MUC5AC* expressions as prognostic indicators in gastric carcinomas. *J. Surg. Oncol.* **2003**, *83*, 253–260.
44. Barresi, V.; Vitarelli, E.; Grosso, M.; Tuccari, G.; Barresi, G. Relationship between immunoexpression of mucin peptide cores MUC1 and MUC2 and Lauren's histologic subtypes of gastric carcinomas. *Eur. J. Histochem.* **2006**, *50*, 301–309.
45. Kocer, B.; Soran, A.; Kiyak, G.; Erdogan, S.; Eroglu, A.; Bozkurt, B.; Solak, C.; Cengiz, O. Prognostic significance of mucin expression in gastric carcinoma. *Dig. Dis. Sci.* **2004**, *49*, 954–964.
46. Terada, T. An immunohistochemical study of primary signet-ring cell carcinoma of the stomach and colorectum: II. Expression of *MUC1*, *MUC2*, *MUC5AC*, and *MUC6* in normal mucosa and in 42 cases. *Int. J. Clin. Exp. Pathol.* **2013**, *6*, 613–621.
47. Tanaka, M.; Kitajima, Y.; Sato, S.; Miyazaki, K. Combined evaluation of mucin antigen and E-cadherin expression may help select patients with gastric cancer suitable for minimally invasive therapy. *Br. J. Surg.* **2003**, *90*, 95–101.
48. Ohno, T.; Aihara, R.; Kamiyama, Y.; Mochiki, E.; Asao, T.; Kuwano, H. Prognostic significance of combined expression of MUC1 and adhesion molecules in advanced gastric cancer. *Eur. J. Cancer* **2006**, *42*, 256–263.
49. Owen, D.A. Stomach. In *Histology for Pathologists*; Mills, S.E., Ed.; Lippincott Williams and Wilkins: Philadelphia, PA, USA, 2007; p. 589.
50. Reis, C.A.; David, L.; Correa, P.; Carneiro, F.; de Bolós, C.; Garcia, E.; Mandel, U.; Clausen, H; Sobrinho-Simões, M. Intestinal metaplasia of human stomach displays distinct patterns of mucin (*MUC1*, *MUC2*, *MUC5AC*, and *MUC6*) expression. *Cancer Res.* **1999**, *59*, 1003–1007.
51. Vernygorodskyi, S. Immunohistochemical evaluation of mucin expression in precancerous tissue of stomach. *Exp. Oncol.* **2013**, *35*, 114–117.
52. Jonckheere, N; van Seuningen, I. The membrane-bound mucins: From cell signalling to transcriptional regulation and expression in epithelial cancers. *Biochimie* **2010**, *92*, 1–11.
53. Bafna, S.; Kaur, S.; Batra, S.K. Membrane-bound mucins: the mechanistic basis for alterations in the growth and survival of cancer cells. *Oncogene* **2010**, *29*, 2893–2904.
54. Kufe, D.W. Mucins in cancer: Function, prognosis and therapy. *Nat. Rev. Cancer* **2009**, *9*, 874–885.
55. Gendler, S.J. MUC1, the renaissance molecule. *Mammary Gland Biol. Neoplasia* **2001**, *6*, 339–353.

56. McAuley, J.L.; Linden, S.K.; Png, C.W.; King, R.M.; Pennington, H.L.; Gendler, S.J.; Florin, T.H.; Hill, G.R.; Korolik, V.; McGuckin, M.A. MUC1 cell surface mucin is a critical element of the mucosal barrier to infection. *J. Clin. Investig.* **2007**, *117*, 2313–2324.
57. Lindén, S.K.; Sheng, Y.H.; Every, A.L.; Miles, K.M.; Skoog, E.C.; Florin, T.H.; Sutton, P.; McGuckin, M.A. MUC1 limits Helicobacter pylori infection both by steric hindrance and by acting as a releasable decoy. *PLoS Pathog.* **2009**, *5*, e1000617.
58. McGuckin, M.A.; Every, A.L.; Skene, C.D.; Linden, S.K.; Chionh, Y.T.; Swierczak, A.; McAuley, J.; Harbour, S.; Kaparakis, M.; Ferrero, R.; *et al.* Muc1 mucin limits both Helicobacter pylori colonization of the murine gastric mucosa and associated gastritis. *Gastroenterology* **2007**, *133*, 1210–1218.
59. Vinall, L.E.; King, M.; Novelli, M.; Green, C.A.; Daniels, G.; Hilkens, J.; Sarner, M.; Swallow, D.M. Altered expression and allelic association of the hypervariable membrane mucin MUC1 in Helicobacter pylori gastritis. *Gastroenterology* **2002**, *123*, 41–49.
60. Li, M.; Huang, L.; Qiu, H.; Fu, Q.; Li, W.; Yu, Q.; Sun, L.; Zhang, L.; Hu, G.; Hu, J.; *et al.* Helicobacter pylori infection synergizes with three inflammation-related genetic variants in the GWASs to increase risk of gastric cancer in a Chinese population. *PLoS One* **2013**, *8*, e74976.
61. Carson, D.D. The cytoplasmic tail of MUC1: A very busy place. *Sci. Signal.* **2008**, *1*, doi:10.1126/scisignal.127pe35.
62. Behrens, M.E.; Grandgenett, P.M.; Bailey, J.M.; Singh, P.K.; Yi, C.H.; Yu, F.; Hollingsworth, M.A. The reactive tumor microenvironment: MUC1 signaling directly reprograms transcription of CTGF. *Oncogene* **2010**, *29*, 5667–5677.
63. Masckauchán, T.N.; Shawber, C.J.; Funahashi, Y.; Li, C.M.; Kitajewski, J. Wnt/beta-catenin signaling induces proliferation, survival and interleukin-8 in human endothelial cells. *Angiogenesis* **2005**, *8*, 43–51.
64. Murata-Kamiya, N.; Kurashima, Y.; Teishikata, Y.; Yamahashi, Y.; Saito, Y.; Higashi, H.; Aburatani, H.; Akiyama, T.; Peek, R.M., Jr.; Azuma, T.; *et al.* Helicobacter pylori CagA interacts with E-cadherin and deregulates the β-catenin signal that promotes intestinal transdifferentiation in gastric epithelial cells. *Oncogene* **2007**, *26*, 4617–4626.
65. Lévy, L.; Neuveut, C.; Renard, C.A.; Charneau, P.; Branchereau, S.; Gauthier, F.; van Nhieu, J.T.; Cherqui, D.; Petit-Bertron, A.F.; Mathieu, D.; *et al.* Transcriptional activation of interleukin-8 by β-catenin-Tcf4. *J. Biol. Chem.* **2002**, *277*, 42386–42393.
66. Guang, W.; Twaddell, W.S.; Lillehoj, E.P. Molecular Interactions between MUC1 Epithelial Mucin, β-Catenin, and CagA Proteins. *Front. Immunol.* **2012**, *3*, doi:10.3389/fimmu.2012.00105.
67. Park, Y.S.; Guang, W.; Blanchard, T.G.; Kim, K.C.; Lillehoj, E.P. Suppression of IL-8 production in gastric epithelial cells by MUC1 mucin and peroxisome proliferator-associated receptor-γ. *Am. J. Physiol. Gastrointest. Liver Physiol.* **2012**, *303*, G765–G774.
68. Guang, W.; Czinn, S.J.; Blanchard, T.G.; Kim, K.C.; Lillehoj, E.P. Genetic regulation of MUC1 expression by Helicobacter pylori in gastric cancer cells. *Biochem. Biophys. Res. Commun.* **2014**, *445*, 145–150.
69. Perrais, M.; Rousseaux, C.; Ducourouble, M.P.; Courcol, R.; Vincent, P.; Jonckheere, N.; van Seuningen, I. *Helicobacter pylori* urease and flagellin alter mucin gene expression in human gastric cancer cells. *Gastric Cancer* **2014**, *17*, 235–246.

70. Navabi, N.; Johansson, M.E.; Raghavan, S.; Lindén, S.K. *Helicobacter pylori* infection impairs the mucin production rate and turnover in the murine gastric mucosa. *Infect. Immun.* **2013**, *81*, 829–837.
71. Wei, X.; Xu, H.; Kufe, D. Human mucin 1 oncoprotein represses transcription of the p53 tumor suppressor gene. *Cancer Res.* **2007**, *67*, 1853–1858.
72. Weinberg, R.A. *The Biology of Cancer*; Garland Science: New York, NY, USA, 2007.
73. Raina, D.; Kharbanda, S.; Kufe, D. The MUC1 oncoprotein activates the anti-apoptotic phosphoinositide 3-kinase/Akt and Bcl-xL pathways in rat 3Y1 fibroblasts. *J. Biol. Chem.* **2004**, *279*, 20607–20612.
74. Raina, D.; Ahmad, R.; Kumar, S.; Ren, J.; Yoshida, K.; Kharbanda, S.; Kufe, D. MUC1 oncoprotein blocks nuclear targeting of c-Abl in the apoptotic response to DNA damage. *EMBO J.* **2006**, *25*, 3774–3783.
75. Rahn, J.J.; Chow, J.W.; Horne, G.J.; Mah, B.K.; Emerman, J.T.; Hoffman, P.; Hugh, J.C. MUC1 mediates transendothelial migration in vitro by ligating endothelial cell ICAM-1. *Clin. Exp. Metastasis* **2005**, *22*, 475–483.
76. Hikita, S.T.; Kosik, K.S.; Clegg, D.O.; Bamdad, C. MUC1* mediates the growth of human pluripotent stem cells. *PLoS One* **2008**, *3*, e3312.
77. Stroopinsky, D.; Rosenblatt, J.; Ito, K.; Mills, H.; Yin, L.; Rajabi, H.; Vasir, B.; Kufe, T.; Luptakova, K.; Arnason, J.; *et al.* MUC1 is a potential target for the treatment of acute myeloid leukemia stem cells. *Cancer Res.* **2013**, *73*, 5569–5579.
78. Mehla, K.; Singh, P.K. MUC1: A novel metabolic master regulator. *Biochim. Biophys. Acta* **2014**, *1845*, 126–135.
79. Saeki, N.; Ono, H.; Sakamoto, H.; Yoshida, T. Genetic factors related to gastric cancer susceptibility identified using a genome-wide association study. *Cancer Sci.* **2013**, *104*, 1–8.

IJC
International Journal of Cancer

Multilayer-omics analysis of renal cell carcinoma, including the whole exome, methylome and transcriptome

Eri Arai[1]*, Hiromi Sakamoto[2]*, Hitoshi Ichikawa[2], Hirohiko Totsuka[3], Suenori Chiku[4], Masahiro Gotoh[1], Taisuke Mori[1], Tamao Nakatani[1], Sumiko Ohnami[2], Tohru Nakagawa[5], Hiroyuki Fujimoto[5], Linghua Wang[6], Hiroyuki Aburatani[6], Teruhiko Yoshida[2] and Yae Kanai[1]

[1] Division of Molecular Pathology, National Cancer Center Research Institute, Tokyo, Japan
[2] Division of Genetics, National Cancer Center Research Institute, Tokyo, Japan
[3] Bioinfomatics Group, Research and Development Center, Solution Division 4, Hitachi Government and Public Corporation System Engineering Ltd, Tokyo, Japan
[4] Science Solutions Division, Mizuho Information and Research Institute Inc., Tokyo, Japan
[5] Department of Urology, National Cancer Center Hospital, Tokyo, Japan
[6] Genome Science Division, Research Center for Advanced Science and Technology (RCAST), The University of Tokyo, Japan

The aim of this study was to identify pathways that have a significant impact during renal carcinogenesis. Sixty-seven paired samples of both noncancerous renal cortex tissue and cancerous tissue from patients with clear cell renal cell carcinomas (RCCs) were subjected to whole-exome, methylome and transcriptome analyses using Agilent SureSelect All Exon capture followed by sequencing on an Illumina HiSeq 2000 platform, Illumina Infinium HumanMethylation27 BeadArray and Agilent SurePrint Human Gene Expression microarray, respectively. Sanger sequencing and quantitative reverse transcription-PCR were performed for technical verification. MetaCore software was used for pathway analysis. Somatic nonsynonymous single-nucleotide mutations, insertions/deletions and intragenic breaks of 2,153, 359 and 8 genes were detected, respectively. Mutations of *GCN1L1*, *MED12* and *CCNC*, which are members of *CDK8* mediator complex directly regulating β-catenin-driven transcription, were identified in 16% of the RCCs. Mutations of *MACF1*, which functions in the Wnt/β-catenin signaling pathway, were identified in 4% of the RCCs. A combination of methylome and transcriptome analyses further highlighted the significant role of the Wnt/β-catenin signaling pathway in renal carcinogenesis. Genetic aberrations and reduced expression of *ERC2* and *ABCA13* were frequent in RCCs, and *MTOR* mutations were identified as one of the major disrupters of cell signaling during renal carcinogenesis. Our results confirm that multilayer-omics analysis can be a powerful tool for revealing pathways that play a significant role in carcinogenesis.

Cancer Genetics

Key words: *CDK8* mediator complex, clear cell renal cell carcinoma (RCC), multilayer-omics analysis, whole exome analysis, Wnt/β-catenin signaling pathway

Abbreviations: ASCAT: allele-specific copy number analysis of tumors; GeMDBJ: genome medicine database of Japan; GPHMM: global parameter hidden Markov model; indel: insertion/deletion; mTOR: mammalian target of rapamycin; N: non-cancerous renal cortex tissue; PolyPhen: polymorphism phenotyping; RCC: renal cell carcinoma; SIFT: sorting intolerant from tolerant; SNP: single nucleotide polymorphism; T: cancerous tissue

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

*E.A. and H.S. contributed equally to this work

Grant sponsor: Program for Promotion of Fundamental Studies in Health Sciences (10-41, 10-42 and 10-43), National Institute of Biomedical Innovation (NiBio), Japan; **Grant sponsor:** National Cancer Center Biobank, National Cancer Center Research and Development Fund (23A-1), Japan

DOI: 10.1002/ijc.28768

History: Received 24 July 2013; Accepted 16 Jan 2014; Online 6 Feb 2014

Correspondence to: Yae Kanai, MD, PhD, Division of Molecular Pathology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan, Tel.: +81335422511, Fax: +81-3-3248-2463, E-mail: ykanai@ncc.go.jp or Teruhiko Yoshida, MD, PhD, Division of Genetics, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan, Tel.: +81335422511, Fax: +81-3-3541-2685, E-mail: tyoshida@ncc.go.jp

What's new?

Large-scale systems biology approaches are currently reshaping biomedical research identifying new pathways or reinforcing significance of previously discovered pathways in cancer biology. Here the authors performed multilayer -omics analyses in clear renal carcinoma or healthy control samples. They found frequent tumor-associated genetic aberrations of *GCN1L1*, *MED12*, and *CCNC*, all members of the *CDK8* Mediator complex involved in regulating β-catenin-driven transcription, as well as alterations in *MACF1*, also a member of the Wnt/β-catenin signaling pathway. These findings underscore the significance of the Wnt/β-catenin signaling pathway during renal carcinogenesis and confirm the power of large-scale sequencing efforts in revealing pathways that may become therapeutic targets in specific cancers.

Clear cell renal cell carcinoma (RCC) is the most common histological subtype of adult kidney cancer and frequently affects working-age adults in midlife.[1] Recently, large-scale PCR-based exon resequencing and whole-exome analysis by exon capturing have revealed that renal carcinogenesis involves inactivation of histone-modifying genes such as *SETD2*,[2] a histone H3 lysine 36 methyltransferase, *KDM5C*,[2] a histone H3 lysine 4 demethylase and *UTX*,[3] a histone H3 lysine 27 demethylase, as well as the SWI/SNF chromatin remodeling complex gene, *PBRM1*.[4] Moreover, it is well known that clear cell RCCs are characterized by inactivation of the *VHL* tumor-suppressor gene encoding a component of the protein complex that possesses ubiquitin ligase E3 activity.[5] Another exome analysis study has revealed frequent mutation of a further component of the ubiquitin-mediated proteolysis pathway, *BAP1*.[6] Non-synonymous mutations of the *NF2* gene and truncating mutations of the *MLL2* gene have also been reported.[2]

Not only genetic, but also epigenetic events appear to accumulate during carcinogenesis, and DNA methylation alterations are one of the most consistent epigenetic changes in human cancers.[7,8] In fact, we have shown that noncancerous renal tissue obtained from patients with RCCs is already at the precancerous stage associated with DNA methylation alterations, even though no remarkable histological changes are evident and there is no association with chronic inflammation or persistent infection with viruses or other pathogens.[9,10] Furthermore, using single-CpG resolution methylome analysis with the Infinium array, we have demonstrated that DNA methylation alterations at precancerous stages may determine tumor aggressiveness and patient outcome.[11]

It is well known that DNA methylation alterations around promoter regions affect the expression levels of tumor-related genes.[7] Once the DNA methylation status has been altered, such alterations are stably preserved on the DNA double strands by covalent bonds through maintenance-methylation mechanisms by *DNMT1* during carcinogenesis.[7] Therefore, tumor-related genes showing alterations of both expression level and DNA methylation may have a larger impact on carcinogenesis than those showing only alterations of expression. Therefore, subjecting tissue specimens to a combination of both methylome and transcriptome analyses may be a powerful approach for revealing genes that are involved in carcinogenetic pathways.

Although one article reporting the use of an integrated multilayer-omics approach including exome analysis to examine human clear cell RCCs was published while this manuscript was in preparation,[12] the entire pathway of carcinogenesis in the kidney may not yet be fully explained. In this study, to identify pathways having a significant impact during renal carcinogenesis, we subjected paired samples of both noncancerous renal cortex tissue (N) and cancerous tissue (T) from patients with clear cell RCCs to whole-exome, methylome and transcriptome analyses.

Material and Methods

Patients and tissue samples

Sixty-seven paired T and N samples were obtained from materials that had been surgically resected from 67 patients with primary clear cell RCCs. N mainly consists of proximal tubules, which are the origin of clear cell RCCs. These patients had not received any preoperative treatment and had undergone nephrectomy at the National Cancer Center Hospital, Tokyo. Tissue specimens were provided by the National Cancer Center Biobank, Tokyo. Histological diagnosis was made in accordance with the World Health Organization classification.[13] All the tumors were graded on the basis of previously described criteria[14] and classified according to the pathological Tumor-Node-Metastasis classification.[15] The clinicopathological parameters of these RCCs are summarized in Supporting Information Table S1.

All patients included in this study provided written informed consent. This study was approved by the Ethics Committee of the National Cancer Center, Tokyo and was performed in accordance with the Declaration of Helsinki.

Exome analysis

High-molecular-weight DNA was extracted using phenol-chloroform, followed by dialysis. Three-microgram aliquots of genomic DNA from the 67 paired samples were fragmented by a Covaris-S2 instrument (Covaris, Woburn, MA) to provide DNA fragments with a base pair peak at 150–200 bp. The DNA fragments were end-repaired and ligated with paired-end adaptors (NEBNext DNA sample prep, New England Biolabs, Ipswich, MA). The resulting DNA library was purified using Agencourt AMPure XP Reagent (Beckman Coulter Genomics, Danvers, MA) and amplified by PCR (4 cycles). Five-hundred-nanogram aliquots of the adaptor-ligated libraries were

Table 1. Genes showing 3 or more genetic aberration scores in clear cell RCCs

			Genetic aberration score				Predicted protein function				
							Nonsynonymous single-nucleotide mutation[2]		Indel	Copy number aberration (%)[3]	
Genes	Chr[1]	Entrez Gene ID	Non-synonymous single-nucleotide mutation	Indel	Intragenic break	Total	SIFT	PolyPhen-2	SIFT	Loss	Gain
VHL	3	7,428	22	14	0	36	0	1	Damaging	77.61	11.94
PBRM1	3	55,193	11	10	1	22	0	1	Damaging	73.13	10.45
TTN	2	7,273	9	3	0	12	0.75	0.387878	Neutral	0.00	38.81
KDM5C	X	8,242	4	4	0	8	0	0.998	Damaging	53.73	26.87
MUC16	19	94,025	6	0	0	6	0	NA	–	2.99	29.85
CUBN	10	8,029	5	1	0	6	0.32	0.987	Damaging	0.00	26.87
SETD2	3	29,072	3	3	0	6	0	0.99	Damaging	76.12	7.46
ABCA13	7	154,664	5	0	0	5	0	NA	–	0.00	44.78
BIRC6	2	57,448	4	1	0	5	0.02	NA	Damaging	4.48	35.82
GCN1L1	12	10,985	3	2	0	5	0	0.735079	Damaging	0.00	37.31
HERC2	15	8,924	5	0	0	5	0.01	0.902	–	1.49	25.37
BAP1	3	8,314	4	0	0	4	0	1	–	74.63	10.45
KIAA0100	17	9,703	4	0	0	4	0.05	0.999	–	0.00	29.85
MTOR	1	2,475	4	0	0	4	0	0.999	–	7.46	25.37
SPTBN1	2	6,711	3	1	0	4	0	0.993	NA	0.00	35.82
SPTA1	1	6,708	2	2	0	4	0.09	0.513	Damaging	0.00	34.33
CADM2	3	253,559	1	0	3	4	0.09	0.012	–	29.85	25.37
ERC2	3	26,059	1	0	3	4	0.01	NA	–	71.64	10.45
ADAM23	2	8,745	3	0	0	3	0	0.998	–	2.99	37.31
AKAP9	7	10,142	3	0	0	3	0	0.986	–	0.00	46.27
ANKRD26	10	22,852	3	0	0	3	0	0.995	–	2.99	28.36
ARHGEF33	2	100,271,715	3	0	0	3	0	NA	–	2.99	35.82
BRD4	19	23,476	3	0	0	3	0	0.997	–	0.00	29.85
C1orf112	1	55,732	3	0	0	3	0	0.952	–	0.00	34.33
CCNC	6	892	3	0	0	3	0	0.876	–	2.99	22.39
CPAMD8	19	27,151	3	0	0	3	0	0.439286	–	0.00	29.85
CSMD3	8	114,788	3	0	0	3	0	0.999	–	1.49	31.34
DNAH5	5	1,767	3	0	0	3	0.1	0.169	–	0.00	46.27
FAT1	4	2,195	3	0	0	3	0	NA	–	1.49	22.39
FAT2	5	2,196	3	0	0	3	0	0.999	–	0.00	71.64
FMN2	1	56,776	3	0	0	3	0	0.957	–	0.00	34.33
FNIP1	5	96,459	3	0	0	3	0.1	0.45171	–	0.00	65.67
KIF26B	1	55,083	3	0	0	3	0	NA	–	0.00	34.33
LIMCH1	4	22,998	3	0	0	3	0	0.992	–	2.99	20.90
LRBA	4	987	3	0	0	3	0.01	0.939	–	1.49	23.88
MACF1	1	23,499	3	0	0	3	0	0.791225	–	4.48	25.37
MADD	11	8,567	3	0	0	3	0	0.999	–	0.00	29.85
MED12	X	9,968	3	0	0	3	0.01	0.576	–	55.22	25.37
MGAM	7	8,972	3	0	0	3	0	NA	–	0.00	46.27
OBSCN	1	84,033	2	1	0	3	0	NA	Neutral	1.49	34.33

Table 1. Genes showing 3 or more genetic aberration scores in clear cell RCCs (Continued)

			Genetic aberration score				Predicted protein function: Nonsynonymous single-nucleotide mutation[2]		Predicted protein function: Indel	Copy number aberration (%)[3]	
Genes	Chr[1]	Entrez Gene ID	Non-synonymous single-nucleotide mutation	Indel	Intragenic break	Total	SIFT	PolyPhen-2	SIFT	Loss	Gain
PLCE1	10	51,196	3	0	0	3	0	0.999	–	10.45	25.37
PREX2	8	80,243	3	0	0	3	0	1	–	5.97	31.34
PTPN4	2	5,775	3	0	0	3	0	0.999	–	0.00	35.82
ROR2	9	4,920	3	0	0	3	0	1	–	7.46	20.90
RP1	8	6,101	3	0	0	3	0.01	0.992	–	5.97	31.34
RYR2	1	6,262	3	0	0	3	0	NA	–	0.00	34.33
SYNE1	6	23,345	3	0	0	3	0.04	0.918	–	4.48	20.90
TTI1	20	9,675	3	0	0	3	0	0.999	–	0.00	29.85
VWDE	7	221,806	3	0	0	3	0.04	NA	–	0.00	44.78
ATM	11	472	2	1	0	3	0	1	NA	7.46	28.36
DNAH2	17	146,754	2	1	0	3	0.14	0.048	Damaging	0.00	29.85
FOXN2	2	3,344	2	1	0	3	0.08	0.255	Neutral	1.49	35.82
PTEN	10	5,728	2	1	0	3	0.01	0.988	Damaging	8.96	25.37
SAMD9L	7	219,285	2	1	0	3	0	0.968	Damaging	0.00	46.27
SI	3	6,476	2	1	0	3	0.01	0.992	Damaging	10.45	32.84
TCHH	1	7,062	2	1	0	3	NA	0.998	Damaging	0.00	34.33
TUBGCP6	22	610,053	2	1	0	3	0	0.993	NA	1.49	29.85
UGGT2	13	55,757	2	1	0	3	0.01	0.726	Neutral	0.00	25.37
CCDC178	18	374,864	1	2	0	3	0	0.235	Damaging	2.99	22.39
HGSNAT	8	138,050	1	2	0	3	0	NA	Damaging	16.42	25.37
NIPBL	5	25,836	1	2	0	3	0.05	0.98	Damaging	0.00	46.27

[1]Chromosome.
[2]Minimum SIFT score and maximum PolyPhen-2 score among all detected mutations of each gene (A SIFT score of <0.05 means "damaging."[19] PolyPhen-2 scores of >0.85 and 0.15–0.85 mean "probably damaging" and "possibly damaging," respectively).[20] NA: not available using SIFT or PolyPhen-2; –: indels of the gene were not detected.
[3]The incidence of loss (1 or less copy number) or gain (3 or more copy number) detected using ASCAT or GPHMM in all 67 tumors. SIFT and PolyPhen-2 scores and copy numbers of each gene in each RCC were described in Supporting Information Table S3.

hybridized for 24 hr at 65°C with biotinylated oligo RNA bait, SureSelect Human All Exon 50 Mb (Agilent Technologies, Santa Clara, CA). The hybridized genomic DNA was subjected to 10 cycles of PCR reamplification. Following the manufacturer's standard protocols, the whole-exome DNA library was sequenced on an Illumina HiSeq 2000 (Illumina, San Diego, CA) using 75-bp paired-end reads.

After completion of the entire run, image analyses, error estimation and base calling were performed using the Illumina Pipeline (version 1.3.4) to generate primary data. First, the reads were aligned against the reference human genome from UCSC human genome 19 (Hg19) using the Burrows Wheeler Aligner Multi-Vision software package.[16] Because duplicated reads had been generated during the PCR amplification process, paired-end reads that were aligned to the same genomic positions were removed using SAMtools. Second, the following loci were removed: (*i*) read depth <6 and (*ii*) base quality score <3 in the T sample. Third, we used the following Bayesian data analysis pipeline developed in our laboratory: (*i*) single nucleotide polymorphism (SNP) array analysis was performed on each paired cancerous and noncancerous tissue samples using Illumina HumanOmni1-Quad BeadChip (see "SNP microarray analysis") and the genomic region, which is considered to be 1 copy in the pure cancerous genome was identified by the visual inspection of the log R ratio and B allele frequency plots on the Illumina Genome Viewer in the GenomeStudio software. (*ii*) Heterozygous SNP loci were selected from the above 1-copy region using GATK UnifiedGenotyper (Broad Institute, MA). (*iii*) At the SNP loci, which were 1 copy in the pure cancerous genome but heterozygous in the noncancerous genome, the ratio of the contaminating non-cancerous cells in the

cancerous tissue was estimated from the allele frequencies of the cancerous genome by fitting to a binominal mixture model. (*iv*) Considering the estimated ratio of the contaminating noncancerous cells, the posterior probability of the genotypes of the cancer cells was calculated. Mutation was called if the posterior probability of being homozygous for the allele recorded in the reference human genome sequence was 0.001 or lower, and the ratio of the nonreference allele was 0.02 or lower in the noncancerous tissue sample, which had a read depth of at least 15. Fourth, Annovar extracted candidates that were nonsynonymous and did not correspond to the refSNP number. Fifth, candidates were discarded if the frequency of the nonreference allele was >2% in the N sample. Somatic mutations were also removed from the candidates if the root mean square mapping quality score of the reads covering the somatic mutation was <20. Finally, if the Blast search did not detect homologous regions for which the edit distance was 7 or <7 within the neighboring 151-bp stretch (75 bp both up- and downstream), the candidate was considered as a somatic mutation. Somatic insertions/deletions (indels) were called using both SAMtools and Pindel[17] as described previously.[18] Effects of amino acid substitutions on protein function due to single nucleotide nonsynonymous mutations have been estimated using the Sorting Intolerant from Tolerant (SIFT) (http://sift.jcvi.org)[19] and polymorphism phenotyping (PolyPhen)-2 (http://genetics.bwh.harvard.edu/pph2/),[20] and those due to indels have been estimated using SIFT.[21] All data from exome analysis will be submitted to the Genome Medicine Database of Japan (GeMDBJ, https://gemdbj.nibio.go.jp/dgdb/).

Sanger sequencing

To verify the nonsynonymous single-nucleotide mutations and indels detected by the exome analysis and described in Table 1, the target sites and the flanking sequences of each patient's DNA template were amplified individually with specific primers designed using Primer6.0. The PCR products were then sequenced with an ABI 3730 DNA Analyzer using the BigDye Terminator v1.1 Cycle Sequencing kit (Life Technologies, Carlsbad, CA).

SNP microarray analysis

Two-hundred-nanogram aliquots of DNA from the 67 paired samples were genotyped with the HumanOmni1-Quad BeadChip (Illumina) in accordance with the manufacturer's protocols. The data were assembled using GenomeStudio software (Illumina). For the single-nucleotide mutation detection, we developed the Bayesian data analysis pipeline using SNP microarray data (see "Exome analysis"). Localization of intragenic breakpoints, in which the end point of a deletion or duplication lies within a gene, in each of the T samples was clearly identified by the visual inspection of the B allele frequency plots on the Illumina Genome Viewer in the GenomeStudio software (Supporting Information Fig. S1). Copy number data has been obtained using Allele-Specific Copy Number Analysis of Tumors (ASCAT; http://heim.ifi.uio.no/bioinf/Projects/ASCAT/)[22] and Global Parameter Hidden Markov Model (GPHMM; http://bioinformatics.ustc.edu.cn/gphmm/)[23] software.

Infinium analysis

Five-hundred-nanogram aliquots of DNA from the 67 paired samples were subjected to bisulfite conversion using an EZ DNA Methylation-Gold™ Kit (Zymo Research, Irvine, CA). Subsequently the DNA methylation status at 27,578 CpG loci was examined at single-CpG resolution using the Infinium HumanMethylation27 Bead Array (Illumina). The data were assembled using GenomeStudio methylation software (Illumina). At each CpG site, the ratio of the fluorescent signal was measured using a methylated probe relative to the sum of the methylated and unmethylated probes, that is, the so-called β-value, which ranges from 0.00 to 1.00, reflecting the methylation level of an individual CpG site. All data of Infinium analysis will be submitted to GeMDBJ.

Pyrosequencing

DNA methylation levels of Infinium probe sites of the *RAB25, GGT6, C3* and *CHI3L2* genes and the 5′-region of the *ABCA13* gene were measured by pyrosequencing. The PCR and sequencing primers were designed using Pyrosequencing Assay Design Software ver.1.0 (QIAGEN, Hilden, Germany). To overcome any PCR bias, we optimized the annealing temperature as described previously.[24] Each of the primer sequences and PCR conditions are given in Supporting Information Figure S2. The PCR product was generated from bisulfite-treated DNA and subsequently captured on streptavidin-coated beads. Quantitative sequencing was performed on a PyroMark Q24 (QIAGEN) using the Pyro Gold Reagents (QIAGEN) in accordance with the manufacturer's protocol.

Expression microarray analysis

Total RNA was isolated using TRIzol reagent (Life Technologies). From the 67 paired samples, 29 pairs, from which a sufficient amount of total RNA for both N and T samples was available, were subjected to expression microarray analysis. Two-hundred-nanogram aliquots of total RNA from the 29 paired samples were used for the production of fluorescent complementary RNA, and all samples were hybridized to the SurePrint G3 Human Gene Expression 8 × 60 K microarray (Agilent Technologies). The signal values were extracted using the Feature Extraction software (Agilent Technologies). All data of Expression microarray analysis will be submitted to GeMDBJ.

Quantitative RT-PCR analysis

cDNA was reverse-transcribed from total RNA using random primers and Superscript III RNase H⁻ Reverse Transcriptase (Life Technologies). From the 67 paired samples, 66 pairs, from which a sufficient amount of cDNA for both N and T

samples was available, were subjected to quantitative RT-PCR analysis. mRNA expression was analyzed using custom TaqMan Expression Assays (probe and PCR primer sets, Supporting Information Table S2) on the 7500 Fast Real-Time PCR System employing the relative standard curve method. All CT values were normalized to that of *GAPDH* in the same sample.

Multilayer-omics scoring

If any of the somatic nonsynonymous single-nucleotide mutations, indels or intragenic breaks was observed in one of the T samples, a genetic aberration score of one was assigned for the gene. If the $\Delta\beta$ ($\beta_T - \beta_N$) was 0.2 or more, the gene was considered to be hypermethylated in the T sample relative to the corresponding N sample. If the $\Delta\beta$ ($\beta_T - \beta_N$) was -0.2 or less, the gene was considered to be hypomethylated in the T sample relative to the corresponding N sample.

The expression level (E value) of each gene was expressed as the log2-signal intensity normalized by the median for all probes in the sample. If the ΔE ($E_T - E_N$) was 4 or more, the expression of the gene was considered to be elevated in the T sample relative to the corresponding N sample. If the ΔE ($E_T - E_N$) was -4 or less, the expression of the gene was considered to be reduced in the T sample relative to the corresponding N sample.

All probes of the Infinium HumanMethylation27 Bead Array and SurePrint G3 Human Gene Expression 8 × 60 K microarray were aligned against the reference human genome from Hg19. Infinium array probe and expression microarray probe pairs were annotated to each individual gene. If the probe of the Infinium array was designed for the upstream region including the promoter region, exon 1 or intron 1 of the gene, if $\Delta\beta$ ($\beta_T - \beta_N$) of the gene was 0.2 or more (DNA hypermethylation), and if ΔE ($E_T - E_N$) based on the expression microarray was -4 or less (reduced expression) in one paired sample of T and N, then a gene downregulation score of one was assigned. If the probe of the Infinium array was designed for the upstream region including the promoter region, exon 1 or intron 1 of the gene, if $\Delta\beta$ of the gene was -0.2 or less (DNA hypomethylation), and if ΔE ($E_T - E_N$) based on the expression microarray was 4 or more (overexpression) in one paired sample of T and N, then a gene upregulation score of one was assigned.

Pathway analysis

MetaCore software (http://www.genego.com) is a pathway analysis tool based on a proprietary manually curated database of human protein–protein, protein–DNA and protein compound interactions. The MetaCore pathway analysis by GeneGo was performed among genes showing genetic scores of 3 or more or showing downregulation or upregulation scores of 5 or more. Pathways for which the p value was <0.05 were considered to play a significant role in renal carcinogenesis.

Results

Genetic aberrations

Exome analysis detected somatic non-synonymous single-nucleotide mutations and indels of 2,153 and 359 genes among the 67 clear cell RCCs, respectively. SNP array analysis revealed intragenic breaks in 8 genes among the 67 RCC samples. In total, 2,440 genes showed non-synonymous single-nucleotide mutations, indels and/or intragenic breaks in RCCs and were assigned genetic aberration scores (described in "Multilayer-omics scoring" in the Material and Methods section) of 1 or more. Genetic alterations in each RCC are summarized in Supporting Information Table S3. The 2,131 and 248 genes that were assigned a genetic aberration score of 1 and 2 are listed in Supporting Information Table S4, and the 61 genes that were assigned genetic aberration scores of 3 or more are listed in Table 1. All 256 mutations (209 somatic nonsynonymous single-nucleotide mutations and 57 indels with 10 exceptions, for which Sanger sequencing failed due to difficulties with PCR primer design) listed in Table 1 were verified by Sanger sequencing. In addition, mutations of 54 (89%) of the 61 genes included in Table 1 were also found in the clear cell RCC database in The Cancer Genome Atlas (http://cancergenome.nih.gov/; Supporting Information Table S5), indicating the reliability of our whole-exome analysis results.

Effects of amino acid substitutions due to genetic aberrations on protein function estimated using SIFT[19,21] and PolyPhen-2[20] software are shown in Table 1 and Supporting Information Table S3. In 60 of 61 genes listed in Table 1, SIFT and PolyPhen-2 analyses (less than 0.05 SIFT score[19] or more than 0.15 PolyPhen-2 score[20] for nonsynonymous single-nucleotide mutations and "damaging" SIFT score[21] for indels) indicated that amino acid substitutions due to genetic aberrations impair the functions of proteins.

The incidence of copy number loss (1 or less) and gain (3 or more), detected using ASCAT[22] and GPHMM[23] software, of the genes that were assigned genetic aberration scores of 3 or more is described in Table 1. The copy number of each gene showing genetic aberrations in each RCC is described in Supporting Information Table S3. Nonsynonymous single-nucleotide mutations and indels were frequently concordant with copy number alterations (Table 1), suggesting that such genetic aberrations may actually result in dysfunction of proteins in RCCs.

In addition to recurrent genetic aberrations, expression microarray analysis revealed reduced mRNA expression [ΔE ($E_T - E_N$) was -4 or less as described in "Expression microarray analysis" in the Material and Methods section] of the *ERC2* and *ABCA13* genes in 21 and 31% of RCCs, respectively. These mRNA expression alterations were verified quantitatively by real-time RT-PCR analysis [mean *ERC2* expression levels in T samples ($n = 66$): 8.91 ± 29.72; those in N samples ($n = 66$): 110.02 ± 75.31 ($p < 1.00 \times 10^{-12}$, Mann-Whitney U-test) and mean *ABCA13* expression levels in T samples ($n = 66$): 8.43 ± 45.12; those in N samples ($n = 66$): 47.82 ± 89.51 ($p < 1.00 \times 10^{-12}$, Mann-Whitney U-test)].