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]		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)].

Probes for the *ERC2* gene were designed for the Infinium array, and DNA hypermethylation around the 5′-region of the *ERC2* gene was detected in only 6% of RCCs, indicating that reduced expression of the *ERC2* gene may not be attributable to DNA methylation alterations during renal carcinogenesis. Since the probes for the *ABCA13* gene were not designed for the Infinium array, we examined DNA methylation levels in the 5′-region of the *ABCA13* gene by pyrosequencing. No significant differences in the DNA methylation levels of the *ABCA13* gene between T samples (0.528 ± 0.060, $n = 67$) and N samples (0.510 ± 0.149, $n = 67$) were observed (Supporting Information Fig. S2a). Our data for RCCs were consistent with the data in the public database Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/): no significant differences in DNA methylation levels of the *ABCA13* gene were evident between bile duct cancer and normal bile duct tissue (Accession number: GSE49656) and between breast cancer and normal breast tissue (GSE37754), indicating that reduced expression of the *ABCA13* gene may not be attributable to DNA methylation alterations during renal carcinogenesis.

Alterations of expression associated with DNA hypermethylation or hypomethylation

All genes showing DNA methylation alterations [0.2 or more $\Delta\beta$ ($\beta_T - \beta_N$) or -0.2 or less $\Delta\beta$ ($\beta_T - \beta_N$)] or mRNA expression alterations [4 or more ΔE ($E_T - E_N$) or -4 or less ΔE ($E_T - E_N$)] in each RCC are summarized in Supporting Information Table S6 along with genes showing genetic aberration scores of 1 or more. The DNA methylation status of the 5′-region can regulate the mRNA expression level of each gene. DNA methylation status is stably preserved on DNA double strands by covalent bonds and inherited through cell division by maintenance-methylation mechanisms by *DNMT1*. Therefore, altered mRNA expression due to DNA methylation alterations may be more stably fixed during multistage human carcinogenesis in comparison to mRNA expression alterations without DNA methylation alterations. Therefore, we have calculated upregulation and downregulation scores based on both DNA methylation status and expression levels described in the Material and Methods section: 86 genes showed reduced expression [-4 or less ΔE ($E_T - E_N$)] associated with DNA hypermethylation [0.2 or more $\Delta\beta$ ($\beta_T - \beta_N$)] in 5 or more patients (downregulation scores of 5 or more; Table 2) and 28 genes showed overexpression [4 or more ΔE ($E_T - E_N$)] associated with DNA hypomethylation [-0.2 or less $\Delta\beta$ ($\beta_T - \beta_N$)] in 5 or more patients (upregulation scores of 5 or more; Table 2).

Expression alterations of genes included in Table 2 were validated using the clear cell RCC database in the Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/; Supporting Information Table S7): reduced or increased mRNA expression of 97 (89%) of the109 genes, which are included in Table 2 and for which probes were designed in the expression microarrays described in the database, were found, indicating the reliability of our expression analysis. Since genome-wide DNA methylation data for RCCs obtained using array-based analysis with appropriate resolution were not available in the public database, Infinium assay data for other human malignant tumors deposited in the Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo/) were used instead for validation (Supporting Information Table S8). In addition, DNA methylation levels of the representative genes, *RAB25*, *GGT6*, *C3* and *CHI3L2*, included in Table 2 based on the Infinium assay were successfully verified using pyrosequencing (Supporting Information Figs. S2*b*–S2*e*), indicating the reliability of our Infinium assay.

Pathway analysis

MetaCore pathway analysis by GeneGo was performed for 61 genes assigned genetic aberration scores of 3 or more, 86 genes assigned downregulation scores of 5 or more (frequent reduction of expression associated with DNA hypermethylation) and 28 genes assigned upregulation scores of 5 or more (frequent overexpression associated with DNA hypomethylation; total 174 genes). Twenty potentially significant GeneGo pathways ($p < 0.05$) and the affected genes are listed in Table 3. Mutations of 5 (100%) of the 5 genes included in Table 3 were found in the clear cell RCC database of The Cancer Genome Atlas (Supporting Information Table S5). Reduced or increased mRNA expression of 11 (92%) of the 12 genes, which are included in Table 3 and for which probes had been designed in expression microarrays described in the clear cell RCC database of the Gene Expression Omnibus, were found (Supporting Information Table S7), supporting the participation of these genes in renal carcinogenesis.

Genes for which correlation with Wnt/β-catenin signaling was indicated by MetaCore pathway analysis, together with their genetic aberration, DNA methylation alterations and mRNA expression alterations, are illustrated schematically in Figure 1. Mutations, mRNA expression alterations or DNA methylation alterations of 32 (89%) of the 36 genes included in Figure 1 were found in Supporting Information Tables S5, S7 or S8, supporting the participation of the Wnt/β-catenin signaling pathway in renal carcinogenesis. In addition, MetaCore pathway analysis was separately performed for RCCs with and without genetic aberrations and/or DNA hypermethylation [$\Delta\beta$ ($\beta_T - \beta_N$) >0.2] of the *VHL* gene (Supporting Information Table S9 and Fig. S3).

Discussion

High frequencies of genetic aberrations of the *VHL* (53%), *PBRM1* (33%), *KDM5C* (12%) and *SETD2* (9%) genes, which have been highlighted in previous resequencing[2] and exome analyses,[4,6] supported the reliability of our approach. In addition to *PBRM1*, somatic mutation of another member of the SWI/SNF complex, *SMARCA4*, was detected. In addition to *SETD2* and *KDM5C*, somatic mutation of another histone modification protein, *JARID2*, was also detected. The significance of aberrations of chromatin remodeling and histone modification-related proteins in RCCs was confirmed.