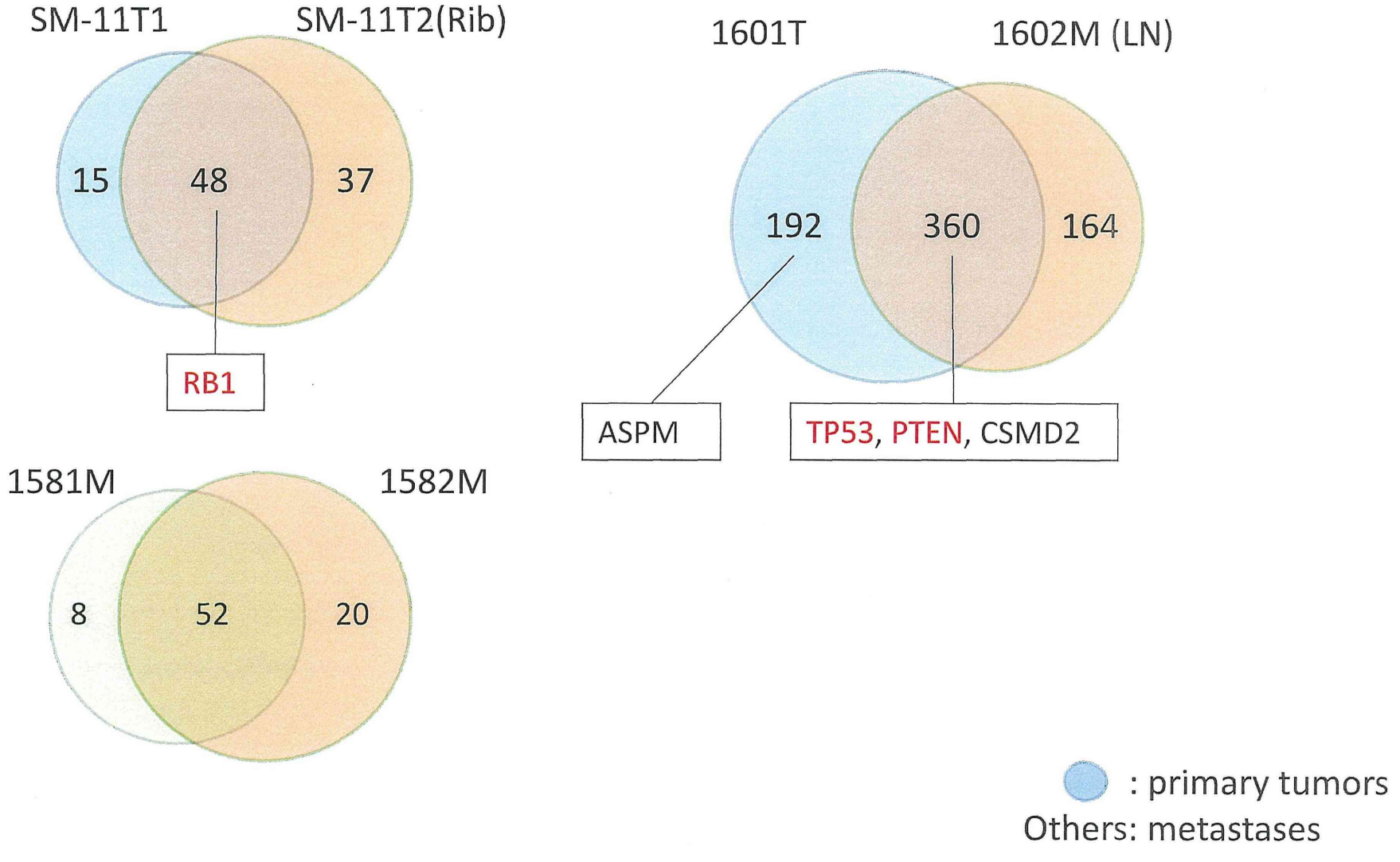
Supplementary Figure S7. Heterogeneity in the Accumulated Mutations among Multiple Tumors Obtained from Same Patients



Comprehensive Exploration of Novel Chimeric Transcripts in Clear Cell Renal Cell Carcinomas Using Whole Transcriptome Analysis

Masahiro Gotoh,[1†] Hitoshi Ichikawa,[2†] Eri Arai,[1] Suenori Chiku,[3] Hiromi Sakamoto,[2] Hiroyuki Fujimoto,[4] Masaki Hiramoto,[5] Takao Nammo,[5] Kazuki Yasuda,[5] 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]Science Solutions Division, Mizuho Information and Research Institute, Inc., Tokyo, Japan
[4]Department of Urology, National Cancer Center Hospital, Tokyo, Japan
[5]Department of Metabolic Disorder, Diabetes Research Center, National Center for Global Health and Medicine, Tokyo, Japan

The aim of this study was to clarify the participation of expression of chimeric transcripts in renal carcinogenesis. Whole transcriptome analysis (RNA sequencing) and exploration of candidate chimeric transcripts using the deFuse program were performed on 68 specimens of cancerous tissue (T) and 11 specimens of non-cancerous renal cortex tissue (N) obtained from 68 patients with clear cell renal cell carcinomas (RCCs) in an initial cohort. As positive controls, two RCCs associated with Xp11.2 translocation were analyzed. After verification by reverse transcription (RT)-PCR and Sanger sequencing, 26 novel chimeric transcripts were identified in 17 (25%) of the 68 clear cell RCCs. Genomic breakpoints were determined in five of the chimeric transcripts. Quantitative RT-PCR analysis revealed that the mRNA expression levels for the *MMACHC, PTER, EPC2, ATXN7, FHIT, KIFAP3, CPEB1, MINPP1, TEX264, FAM107A, UPF3A, CDC16, MCCC1, CPSF3,* and *ASAP2* genes, being partner genes involved in the chimeric transcripts in the initial cohort, were significantly reduced in 26 T samples relative to the corresponding 26 N samples in the second cohort. Moreover, the mRNA expression levels for the above partner genes in T samples were significantly correlated with tumor aggressiveness and poorer patient outcome, indicating that reduced expression of these genes may participate in malignant progression of RCCs. As is the case when their levels of expression are reduced, these partner genes also may not fully function when involved in chimeric transcripts. These data suggest that generation of chimeric transcripts may participate in renal carcinogenesis by inducing dysfunction of tumor-related genes.

INTRODUCTION

Clear cell renal cell carcinoma (RCC) is the most common histological subtype of adult kidney cancer (Ljungberg et al., 2011). In general, RCCs at an early stage are curable by nephrectomy. However, some RCCs relapse and metastasize to distant organs. Even though molecular targeting agents have been developed for treatment of RCCs, their effectiveness for relapsed or metastasized RCCs after nephrectomy is very limited. To improve prognostication and the effectiveness of targeting therapy in patients with RCCs, the molecular background of renal carcinogenesis should be further elucidated.

We and other groups have revealed both genetic and epigenetic events during renal carcinogenesis (Arai and Kanai, 2010). Especially, recent developments in high-throughput sequence capture methods and next-generation sequencing technologies have made exome sequencing technically feasible. Such whole exome analyses have revealed that renal carcinogenesis involves inactivation of histone-modifying genes such as *SETD2* (Dalgliesh et al., 2010), *KDM5C* (Dalgliesh et al., 2010), *UTX* (van Haaften et al., 2009), and *PBRM1*

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

Supported by: the Program for Promotion of Fundamental Studies in Health Sciences (10-42 and 10-43) of the National Institute of Biomedical Innovation (NiBio), Japan and the National Cancer Center Research and Development Fund (23-A-1), Japan.

†Masahiro Gotoh and Hitoshi Ichikawa contributed equally to this work.

*Correspondence to: Yae Kanai; Division of Molecular Pathology, National Cancer Center Research Institute, 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan. E-mail: ykanai@ncc.go.jp

Received 17 April 2014; Accepted 7 August 2014

DOI 10.1002/gcc.22211

Published online 18 September 2014 in Wiley Online Library (wileyonlinelibrary.com).

TABLE 1. The Clinicopathological Parameters of Clear Cell Renal Carcinomas Belonging to the Initial and Second Cohorts

Clinicopathological parameters	Initial cohort ($n = 68$)	Second cohort ($n = 26$)	P
Age (mean ± SD)	62.25 ± 11.00	57.12 ± 10.80	0.078[a]
Sex			
Male	49	17	0.616[b]
Female	19	9	
Tumor diameter (cm, mean ± SD)	5.55 ± 3.21	5.86 ± 2.84	0.407[a]
Macroscopic configuration[c]			
Type 1	25	13	0.562[b]
Type 2	17	5	
Type 3	26	8	
Predominant histological grades[d,e]			
G1	36	12	0.696[b]
G2	21	8	
G3	9	4	
G4	2	2	
Highest histological grades[f]			
G1	5	1	0.105[b]
G2	32	6	
G3	16	11	
G4	15	8	
Vascular involvement[g]			
Negative	38	13	0.649[b]
Positive	30	13	
Predominant growth pattern[e]			
Expansive	61	24	1.000[b]
Infiltrative	7	2	
Most aggressive growth pattern[f]			
Expansive	43	21	0.139[b]
Infiltrative	25	5	
Tumor necrosis			
Negative	51	16	0.212[b]
Positive	17	10	
Renal pelvic invasion			
Negative	61	23	1.000[b]
Positive	7	3	
Distant metastasis			
Negative	58	24	0.500[b]
Positive	10	2	
Pathological TNM stage[h]			
Stage I	33	13	0.531[b]
Stage II	4	3	
Stage III	19	8	
Stage IV	12	2	

[a]Mann–Whitney *U* test.
[b]Fisher's exact test. No significant differences of clinicopathologcial parameters were observed between two cohorts.
[c]Macroscopic configuration was evaluated on the basis of previously described criteria (Arai et al., 2006).
[d]All the tumors were graded on the basis of previously described criteria (Fuhrman et al., 1982).
[e]If the tumor showed heterogeneity, findings in the predominant area were described.
[f]If the tumor showed heterogeneity, the most aggressive features of the tumor were described.
[g]The presence or absence of vascular involvement was examined microscopically on slides stained with hematoxylin-eosin and elastica van Gieson.
[h]All the tumors were classified according to the pathological Tumor-Node-Metastasis classification (Sobin et al., 2009).

(Varela et al., 2011). 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 (Baldewijns et al., 2010). Frequent mutation of a further component of the ubiquitin-mediated proteolysis pathway gene, *BAP1* (Guo et al., 2012), and *VHL*-associated transcription elongation factor, *TCEB1* (Sato et al., 2013), has also been demonstrated on the basis of exome analyses. However, only a limited number of reports have described next-generation sequencing-based whole transcriptome analysis (RNA sequencing) of RCCs, and the molecular background of renal carcinogenesis has not been fully elucidated.

TABLE 2. The 33 Chimeric Transcripts[a] from the 61 Genes Verified by Reverse Transcription-PCR and Sanger Sequencing

		Exon boundaries[b]								
		5′-partner gene			3′-partner gene					
Subtype	Sample	Symbol	RNA accession no.	Exon (chromosome, position)	Symbol	RNA accession no	Exon (chromosome, position)	Distance (bp)	Flanking sequences[c]	In-frame transcripts
(A) Initial cohort										
Clear	K1	*ANTXR1*	NM_032208	17 (ch 2, 69420547)	*GKN1*	NM_019617	2 (ch 2, 69204627)	215920	TCTGTGATGCGTCCACAGCCAGGAGACACG|ATCGTCTTTGCTGGACTTCTTGGAGTCTTT	+
		B4GALT2	NM_001005417	3 (ch 1, 44450537)	*ZSWIM5*	NR_024270	1 (ch 1, 45769599)	1319062	CTTGGCCCGGTTGAAGGTGTCCTCACCATG|GGTTTCGGAGTAAGGGGAGCGGCCCGGCGA	–
		C9orf3	NM_001193329	12 (ch 9, 97767898)	*PRUNE2*	NM_015225	19 (ch 9, 79229516)	18538382	CGCAGGGAGAAGGAAGAGGTGTTTGAAAAG|TATTGACTTGAAGCTGAAAGAAAAGCCTTA	–
		CTNNB1	NM_001904	1 (ch 3, 41241161)	*PLAG1*	NM_002655	3 (ch 8, 57083748)		CCCGCGGCGGGAGGAGCCTGTTCCCCTGAG|ATTGGCCAAAATGGGAAGGATTGGATTCCA	–
		ERBB2	NM_001005862	15 (ch 17, 37872192)	*LTBP4*	NM_001042545	12 (ch 19, 41116438)		ACACTGCCAACCGGCCAGAGGACGAGTGTG|ATGTGGACGAGTGCCACCGCGTGCCGCCGC	+
		MMACHC	NM_015506	1 (ch 1, 45966085)	*BX004987.7*		(ch 1, 143529235)	97563150	CCTTTTGGCTTCGAGGTTTACCCCTTCCAG|GTTATATGCAGTACTGAAGAGCAACTTCAG	–
		MMACHC	NM_015506	2 (ch 1, 45973222)	*BX004987.7*		(ch 1, 143529235)	97556013	GTGGCCTACCATCTGGGCCGTGTTAGAGAG|GTTATATGCAGTACTGAAGAGCAACTTCAG	–
		PHF3	NM_015153	1 (ch 6, 64356700)	*PTP4A1*	NM_003463	1 (ch 6, 64237724)	118976	ATCCCAATTTCCAGATGCCTTGTTCAACAG|AGACTTGCTTCAACACCAGGAAGAACATGA	–
		PTER	NM_001001484	5 (ch 10, 16547159)	*MSRB2*	NM_012228	4 (ch 10, 23399171)	6852012	CATGCCTGATGATAACAAAAGAATTAGAAG|CCTTTCAGTGGGATCTACCTGAATAACAAG	–
Clear	K2	*DNAH11*	NM_003777	79 (ch 7, 21934617)	*RAPGEF5*	NM_012294	7 (ch 7, 22306631)	372014	TTATCCTTCTACTTATGGCCTAGCCCAGTG|GTGCAGAGAGAGCTAGCAGCTGTTATTGCT	–
Clear	K3	*PLOD2*	NM_000935	1 (ch 3, 145878668)	*CCNYL1*	NM_001142300	4 (ch 2, 208602135)		CTCGGAGAAGCCCTCGAGCATCCCCACAG|TGTGACCTTAGCAATATATTACCACATAAAG	–
Clear	K4	*EPC2*	NM_015630	2 (ch 2, 149447942)	*RP11–715D1.1*		(ch 3, 108828611)		AACAGCCAAAACAGTTCATTCATATTCAGC|GAGAGAGCTGCCTGCAGAGAGCGTGAGTCC	–
Clear	K5	*ATXN7*	NM_000333	5 (ch 3, 63938159)	*FHIT*	NM_001166243	2 (ch 3, 61186339)	2751820	TCAAACCGCAGGCATTTCAATCACATTATG|CTTTGAAGCTCAGGAAAGAAGAGAAATCCA	–
		RP1–45C12.1		(ch 1, 171137243)	*KIFAP3*	NM_001204514	21 (ch 1, 169890922)	1246321	AGATGACAATTCATGTTGTTTTCAGGAAAG|CCTTGGAATGGATGGCTTTGGCCAACCAGT	–
Clear	K6	*POLR2G*	NM_002696	2 (ch 11, 62529376)	*CYP1A2*	NM_000761	7 (ch 15, 75047132)		CACCGAGGTGGAGGGGACCTGCACAGGGAA|AGAGCTGTGGGAGGACCCCTCTGAGTTCCG	+
Clear	K7	*AC010724.1*		(ch 15, 83207631)	*CPEB1*	NM_001079533	8 (ch 15, 83218408)	10777	AGAGAAAGGAGAGACAATTATGTTCCTGAG|GTGCAGGTGATCCCCTGGGTATTAGCCGAC	+
Clear	K8	*RIC8A*	NM_021932	1 (ch 11, 211349)	*RP11–34P13.7*		(ch 1, 92240)		CTCATCTTCCTAGAGAAGCGTTTGCACAAG|ACCAAACCAATGCAGACCAAACCAATGCAG	–
Clear	K9	*SEMA6A*	NM_020796	18 (ch 5, 115803279)	*CAMK4*	NM_001744	11 (ch 5, 110825288)	4977991	TGTCTTCCCATAATCACCAAGACAAGAAGG|AGGTCCACCAGGGCAAAATAGAAGCCTTGC	–
Clear	K10	*RP11–322M19.1*		(ch 10, 89005872)	*MINPP1*	NM_001178118	2 (ch 10, 89268093)	262221	CCTCAGCCTCCAAAGTAGCTGAGACTACAG|ATATGGAGTTTGGACCTCCAACAGTTAATG	–
Clear	K11	*TEX264*	NM_001129884	3 (ch 3, 51708578)	*FAM107A*	NM_001076778	1 (ch 3, 58594984)	6886406	ATCGCTGTCTACTATGACAACCCCCACATG|AAAAAGAAATGAAGGCCCAGACACGTTACG	–
Clear	K12	*UPF3A*	NM_023011	5 (ch 13, 115052104)	*CDC16*	NM_001078645	15 (ch 13, 115027362)	24742	AGATGGAGGCGAAGACAAGAGAGCTCATTG|GTAACAGTTGACAAATGGGAACCTTTGTTG	–
Clear	K13	*ASAP1*	NM_001247996	29 (ch 8, 131072825)	*ADCY8*	NM_001115	10 (ch 8, 131862049)	789224	CCCGTACCACTGCCCAGAAAAATCAATACG|AGTGATGCCAATGACCATCCAGTTCTCCAT	–

TABLE 2. (Continued)

		Exon boundaries[b]								
		5′-partner gene			3′-partner gene					
Subtype	Sample	Symbol	RNA accession no.	Exon (chromosome, position)	Symbol	RNA accession no	Exon (chromosome, position)	Distance (bp)	Flanking sequences[c]	In-frame transcripts
Clear	K14	*DCUN1D1*	NM_020640	(ch 3, 182679014)	*MCCC1*	NM_020166	18 (ch 3, 182735125)	56111	TTGCAAAGAATCCAGGACAAAAAGGATTAG| GTGTTTGTCAAAGCTGGAGACAAAGTGAAA	−
Clear	K15	*CPSF3*	NM_016207	7 (ch 2, 9576490)	*ASAP2*	NM_001135191	23 (ch 2, 9533611)	42879	GAAGGGCTCAGGAGCTGCTCTTGATTCTAG| ATCCCCTGACCCCCACGCCGCCCCCACCCG	+
Clear	K16	*ADAMTS2*	NM_014244	2 (ch 5, 178770768)	*RP11–798K23.4*		(ch 5, 178930741)	159973	TCTGTGGCGCTCAGCAACTGCGATGGGCTG| ACCTCTGAATAAGTCGTGGGAGCCCTCGGG	−
Clear	K17	*TPPP*	NM_007030	2 (ch 5, 677865)	*TERT*	NM_198253	3 (ch 5, 1282739)	604874	TGACGTGGACATCGTCTTCAGCAAGATCAA| GGGTTGGCTGTGTTCCGGCCGCAGAGCACC	−
(B) Positive controls										
Xp11.2	K69	*EEF2*	NM_001961	1 (ch 19, 3985376)	*ENHO*	NM_198573	2 (ch 9, 34521854)		TGGGAGAATCCACCGCCATCCGCCACCATG| GCTCAGGACTGCAGGTAGACATCTCCACTG	−
		NONO	NM_001145408	10 (ch X, 70517788)	*TFE3*	NM_006521	6 (ch X, 48891766)	21626022	GAAGGATTCAAGGGAACCTTCCCTGATGCG| CTGCCTGTGTCAGGGAATCTGCTTGATGTG	+
		PARG	NM_003631	8 (ch 10, 51093249)	*BMS1*	NM_014753	6 (ch 10, 43287075)	7806174	GCACTCTGTCTGCCAAATATTTGCACCCAG| GGTGTCAAGCTGTTCTACCTTTCTGGAATG	+
		RAGE	NM_014226	1 (ch 14, 102771299)	*EML1*	NM_001008707	2 (ch 14, 100317190)	2454109	TGTCACCGGCTTCCGCATCCAAGATGAAGA| ATGACAGCGCTTCTGCTGCAAGTAGCATGG	+
Xp11.2	K96	*DPP6*	NM_001039350	1 (ch 7, 153584819)	*ACTR3B*	NM_020445	4 (ch 7, 152498706)	1086113	AAGACCGCTAAGATGCAGGGGAACGTGATG| GTATACCAAGCTTGGCTACGCAGGCAACAC	+
		TFE3	NM_006521	4 (ch X, 48895722)	*RBM10*	NM_001204468	18 (ch X, 47044454)	1851268	CTCACCATCGGGTCCAGCTCAGAGAAGGAG| ATTGCCAAGGACATGGAACGCTGGGCCCG	+
		RBM10	NM_001204468	17 (ch X, 47041725)	*TFE3*	NM_006521	5 (ch X, 48895639)	1853914	GAGAAGCACAAGACCAAGACAGCTCAACAG| ATTGATGATGTCATTGATGAGATCATCAGC	+

[a]Chimeric transcripts (33) included the transcripts (*MMACHC-BX004987.7* and *TFE3-RBM10*) consisting of the same partner gene sets with a different exon boundary or different transcriptional direction, and the transcripts sharing a partner gene, *TFE3*.
[b]National Center for Biotechnology Information Database (Genome Build 37).
[c]The exon boundaries are indicated by a vertical bar. Clear, clear cell renal cell carcinoma; Xp11.2, renal cell carcinoma associated with Xp11.2 translocation; ch, chromosome; +, positive; −, negative.

TABLE 3. Genomic Breakpoints of Chimeric Transcripts

	Genomic breakpoints[a]						
	5′-partner gene			3′-partner gene			
Sample	Symbol	Chr	Genomic position	Symbol	Chr	Genomic position	Flanking sequences[b]
K6	*POLR2G*	11	62530558	*CYP1A2*	15	75045983	TAGTCTCTCGGAAGATCTGGGTTGGGTTCT|GAGAATTGCTTGAACTCTGGAGGTAGAGGC
K7	*AC010724.1*	15	83207075	*CPEB1*	15	83219352	GAGATTATTGAAGTAGATCCTGACACTAAG|GAAATTGGCTCCTCTCTTGTAACTTCTGCC
K9	*SEMA6A*	5	115796806	*CAMK4*	5	110823275	CGTAAGAAATTTGGTACATAAGCTGGTATT|TTAATCCAATTCATCCAAATTATTCTATCG
K13	*ASAP1*	8	131070249	*ADCY8*	8	131862252	GGCAGACAACGATGACGAGCTCACATTCAT|TGCAAAGTTTCTCAATAGAGAGAGTGCTCT
K15	*CPSF3*	2	9578689	*ASAP2*	2	9532071	ACCCTGTCACCCAGGCTGGAGTGTGGTGGC|ACAATCATGGCTCACTGCAGCCTCCAACTC

[a]National Center for Biotechnology Information Database (Genome Build 37).
[b]The genomic breakpoints are indicated by a vertical bar.

Human hematologic (Shima and Kitabayashi, 2011) and soft tissue malignancies (Cantile et al., 2013), prostatic adenocarcinoma (Tomlins et al., 2005), and distinct subtypes of lung adenocarcinoma (Soda et al., 2007; Kohno et al., 2012; Takeuchi et al., 2012) show "addiction" for gene fusion events. Although their incidence is low, fusion events involving the transcription factor *TFE3* gene have been reported in RCCs: RCC associated with Xp11.2 translocation, which harbors *TFE3* fusion, is considered to represent a distinct subtype according to the World Health Organization (WHO) classification (Eble et al., 2004). Moreover, fusion events including anaplastic lymphoma kinase (*ALK*), such as *TMP3-ALK*, *EML4-ALK*, and *VCL-ALK* fusion, have been reported in a distinct group of RCCs, including so-called "unclassified RCC" and papillary RCC in adults (Sugawara et al., 2012) and pediatric RCCs associated with the sickle cell trait (Debelenko et al., 2011; Mariño-Enríquez et al., 2011), based on fluorescence in situ hybridization (FISH) and immunohistochemistry. These findings have prompted us to perform comprehensive exploration of chimeric transcripts in the most common subtype, clear cell RCC, using next-generation sequencing technology. In the present study, to clarify the participation of expression of chimeric transcripts in renal carcinogenesis, whole transcriptome analysis was performed using tissue specimens of 68 clear cell RCCs in adults.

MATERIALS AND METHODS

Patients and Tissue Samples

The initial cohort subjected to whole transcriptome analysis comprised 68 samples of cancerous tissue (T) and 11 samples of non-cancerous renal cortex tissue (N) obtained from materials that had been surgically resected from 68 patients with primary clear cell RCCs. There were 49 men and 19 women with a mean (±standard deviation) age of 62.3 ± 11.0 years (range, 36 to 85 years). All patients underwent nephrectomy at the National Cancer Center Hospital, Tokyo, and had not received any preoperative treatment. Two expert pathologists specializing in genitourinary pathology, E.A. and Y.K., examined all histological slides and performed histological diagnosis in accordance with the WHO classification (Eble et al., 2004). All the tumors were graded on the basis of previously described criteria (Fuhrman et al., 1982) and classified according to the macroscopic configuration (Arai et al., 2006) and the pathological Tumor-Node-Metastasis (TNM) classification (Sobin et al., 2009). As a positive control for chimeric transcript detection, two T samples showing histological findings compatible with Xp11.2 translocation RCC based on the WHO criteria were also subjected to whole transcriptome analysis. For comparison, three T samples of papillary RCCs diagnosed in accordance with the WHO criteria were also subjected to whole transcriptome analysis.

The second cohort subjected to quantitative reverse transcription-polymerase chain reaction (RT-PCR) analysis comprised 26 paired T and N samples obtained from materials that had been surgically resected from 26 other patients with primary clear cell RCCs. These patients comprised 17 men and nine women with a mean (±standard

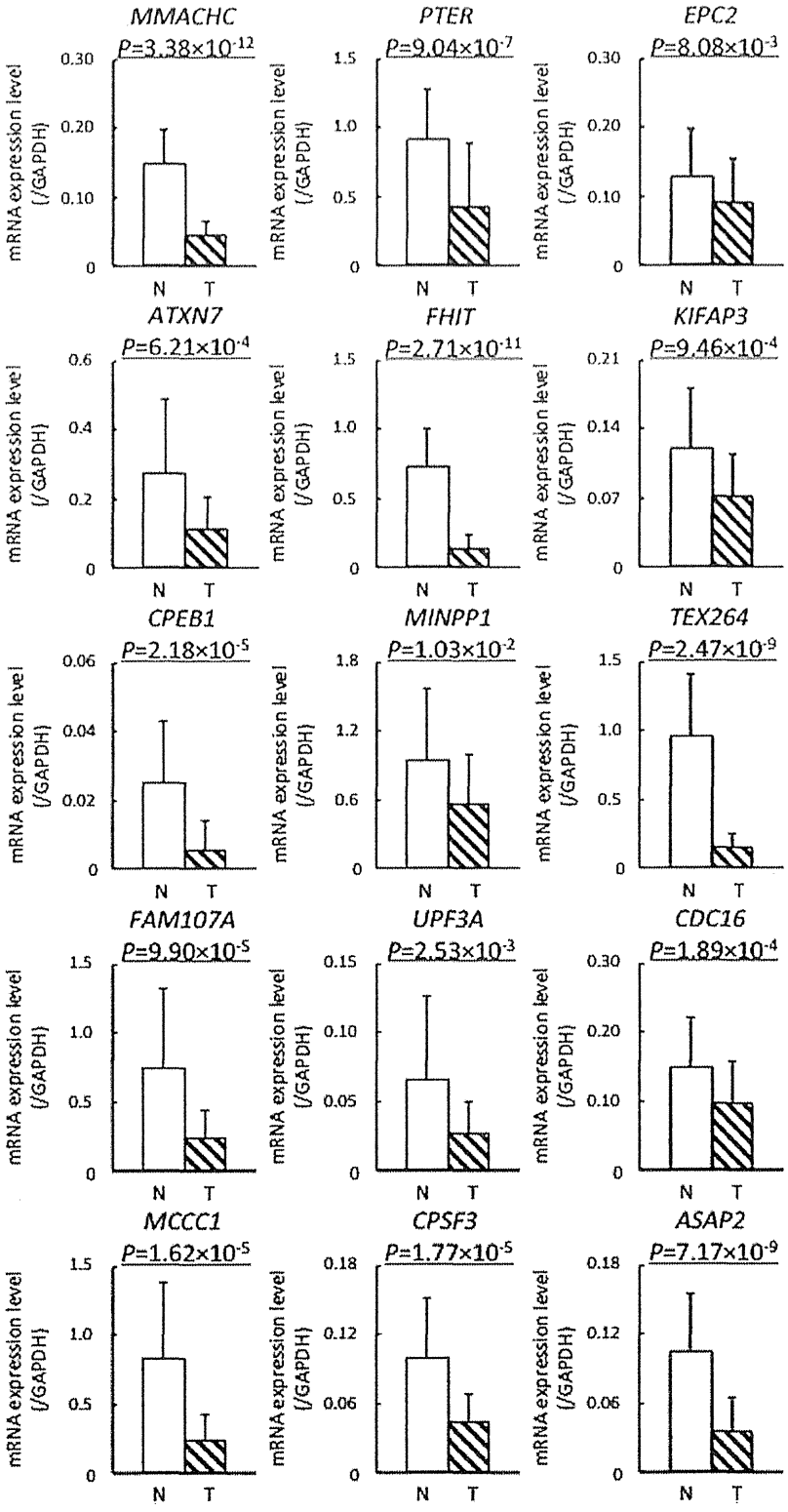


Figure 1. Levels of mRNA expression for the partner genes involved in chimeric transcripts in 26 paired samples of tumorous tissue (T) and non-cancerous renal cortex tissue (N) in the second cohort. mRNA expression was analyzed using custom TaqMan Gene Expression Assays on the 7500 Fast Real-Time PCR System (Life Technologies) employing the relative standard curve method. The probes and PCR primer sets used are summarized in Supporting Information Table S6. Experiments were performed in triplicate for each sample-primer set, and the mean value for the three experiments was used as the CT value. All CT values were normalized to that of *GAPDH* in the same sample. Levels of mRNA expression for the *MMACHC, PTER, EPC2, ATXN7, FHIT, KIFAP3, CPEB1, MINPP1, TEX264, FAM107A, UPF3A, CDC16, MCCC1, CPSF3*, and *ASAP2* genes were significantly reduced in T samples (shaded column) relative to N samples (white column). Bar, standard deviation.

TABLE 4. Correlations Between Levels of mRNA Expression for Each of the Partner Genes Involved in Chimeric Transcripts in Tumorous Tissue Samples and Clinicopathological Parameters Reflecting Tumor Aggressiveness in the Second Cohort

Clinicopathological parameters	Number of tumors	*MMACHC* Expression[a]	*MMACHC* *P*	*PTER* Expression[a]	*PTER* *P*	*EPC2* Expression[a]	*EPC2* *P*	*ATXN7* Expression[a]	*ATXN7* *P*
Macroscopic configuration[b]									
Type 1	13	0.0528 ± 0.0226	1.61×10^{-2c}	0.558 ± 0.557	1.88×10^{-1c}	0.114 ± 0.065	1.23×10^{-1c}	0.125 ± 0.099	2.12×10^{-1c}
Type 2	5	0.0250 ± 0.0116		0.324 ± 0.400		0.0546 ± 0.0480		0.0598 ± 0.0374	
Type 3	8	0.0384 ± 0.0202		0.277 ± 0.248		0.0784 ± 0.0596		0.128 ± 0.099	
Histological grades[d,e]									
G1	1	0.086	5.05×10^{-2c}	0.574	1.15×10^{-1c}	0.246	4.65×10^{-3c}	0.228	1.64×10^{-1c}
G2	6	0.0443 ± 0.0222		0.530 ± 0.608		0.112 ± 0.083		0.134 ± 0.144	
G3	11	0.0489 ± 0.0233		0.498 ± 0.479		0.106 ± 0.033		0.124 ± 0.082	
G4	8	0.0285 ± 0.0105		0.232 ± 0.309		0.0373 ± 0.0185		0.0696 ± 0.0311	
Vascular involvement[f]									
Negative	13	0.0533 ± 0.0254	2.56×10^{-2g}	0.571 ± 0.433	8.60×10^{-3g}	0.125 ± 0.067	7.24×10^{-3g}	0.139 ± 0.100	5.01×10^{-2g}
Positive	13	0.0327 ± 0.0132		0.282 ± 0.452		0.0586 ± 0.0383		0.0876 ± 0.078	
Growth pattern[e]									
Expansive	21	0.046 ± 0.0239	2.00×10^{-1g}	0.482 ± 0.492	1.57×10^{-1g}	0.0954 ± 0.0659	7.53×10^{-1g}	0.121 ± 0.100	7.05×10^{-1g}
Infiltrative	5	0.0306 ± 0.0077		0.194 ± 0.129		0.0762 ± 0.0541		0.0836 ± 0.035	
Tumor necrosis									
Negative	16	0.0503 ± 0.0242	1.69×10^{-2g}	0.548 ± 0.513	1.44×10^{-2g}	0.120 ± 0.064	7.13×10^{-4g}	0.142 ± 0.106	2.68×10^{-2g}
Positive	10	0.0314 ± 0.0134		0.232 ± 0.277		0.0472 ± 0.0274		0.0672 ± 0.0281	
Renal pelvic invasion									
Negative	23	0.0451 ± 0.0230	1.34×10^{-1g}	0.469 ± 0.471	8.46×10^{-3g}	0.0973 ± 0.0649	1.57×10^{-1g}	0.120 ± 0.095	2.11×10^{-1g}
Positive	3	0.0270 ± 0.0030		0.101 ± 0.018		0.0487 ± 0.0218		0.0620 ± 0.0265	
Distant metastasis									
Negative	24	0.0450 ± 0.0221	2.46×10^{-2g}	0.413 ± 0.460	4.98×10^{-1g}	0.0974 ± 0.0625	5.54×10^{-2g}	0.120 ± 0.093	3.69×10^{-2g}
Positive	2	0.0185 ± 0.0007		0.589 ± 0.572		0.0235 ± 0.0021		0.0365 ± 0.0106	
Pathological TNM stage[h]									
Stage I	13	0.0430 ± 0.0166	5.54×10^{-2c}	0.561 ± 0.573	4.69×10^{-1c}	0.104 ± 0.056	1.78×10^{-1c}	0.112 ± 0.095	1.87×10^{-1c}
Stage II	3	0.0277 ± 0.0155		0.204 ± 0.232		0.0607 ± 0.0503		0.112 ± 0.099	
Stage III	8	0.0549 ± 0.0287		0.251 ± 0.163		0.100 ± 0.078		0.135 ± 0.097	
Stage IV	2	0.0185 ± 0.0007		0.589 ± 0.572		0.0235 ± 0.0021		0.0365 ± 0.0106	

Clinicopathological parameters	Number of Tumors	*FHIT* Expression[a]	*FHIT* *P*	*KIFAP3* Expression[a]	*KIFAP3* *P*	*CPEB1* Expression[a]	*CPEB1* *P*	*TEX264* Expression[a]	*TEX264* *P*
Macroscopic configuration[b]									
Type 1	13	0.177 ± 0.125	4.73×10^{-2c}	0.0884 ± 0.0432	3.82×10^{-2c}	0.00369 ± 0.00572	3.31×10^{-1c}	0.155 ± 0.111	9.22×10^{-1c}
Type 2	5	0.0782 ± 0.0187		0.0362 ± 0.0222		0.0058 ± 0.0102		0.142 ± 0.104	
Type 3	8	0.100 ± 0.069		0.0675 ± 0.0345		0.0084 ± 0.0113		0.143 ± 0.086	
Histological grades[d,e]									
G1	1	0.280	6.62×10^{-3c}	0.118	1.18×10^{-1c}	0.004	4.23×10^{-2c}	0.148	9.75×10^{-2c}
G2	6	0.211 ± 0.172		0.0862 ± 0.0535		0.00383 ± 0.00722		0.233 ± 0.154	
G3	11	0.130 ± 0.044		0.0718 ± 0.0259		0.00173 ± 0.00168		0.142 ± 0.068	
G4	8	0.0631 ± 0.0308		0.0556 ± 0.0481		0.0123 ± 0.0119		0.0955 ± 0.0427	
Vascular involvement[f]									
Negative	13	0.178 ± 0.126	1.20×10^{-2g}	0.0839 ± 0.0384	5.68×10^{-2g}	0.00492 ± 0.00742	1.00[g]	0.155 ± 0.112	6.50×10^{-1g}
Positive	13	0.0898 ± 0.0503		0.0599 ± 0.0421		0.00615 ± 0.00978		0.143 ± 0.088	
Growth pattern[e]									
Expansive	21	0.150 ± 0.109	4.09×10^{-2g}	0.0752 ± 0.0446	5.69×10^{-1g}	0.00433 ± 0.00664	1.05×10^{-1g}	0.158 ± 0.104	3.40×10^{-1g}
Infiltrative	5	0.0660 ± 0.0410		0.0580 ± 0.0207		0.0106 ± 0.0139		0.112 ± 0.066	
Tumor necrosis									
Negative	16	0.174 ± 0.114	5.55×10^{-4g}	0.0820 ± 0.0381	3.09×10^{-2g}	0.00275 ± 0.00449	4.08×10^{-2g}	0.183 ± 0.110	6.05×10^{-3g}
Positive	10	0.0699 ± 0.0316		0.0558 ± 0.0431		0.0100 ± 0.0115		0.0943 ± 0.0395	
Renal pelvic invasion									
Negative	23	0.143 ± 0.107	1.34×10^{-1g}	0.0763 ± 0.0420	6.38×10^{-2g}	0.00461 ± 0.00639	5.94×10^{-1g}	0.156 ± 0.103	3.12×10^{-1g}
Positive	3	0.0643 ± 0.0397		0.0380 ± 0.0044		0.0127 ± 0.0193		0.0922 ± 0.0222	
Distant metastasis									
Negative	24	0.141 ± 0.106	5.54×10^{-2g}	0.0755 ± 0.0409	3.69×10^{-2g}	0.00458 ± 0.00790	5.54×10^{-2g}	0.156 ± 0.100	5.54×10^{-2g}
Positive	2	0.0520 ± 0.0014		0.0285 ± 0.0078		0.0170 ± 0.0099		0.0638 ± 0.0021	
Pathological TNM stage[h]									
Stage I	13	0.161 ± 0.124	1.40×10^{-1c}	0.0730 ± 0.0375	7.75×10^{-2c}	0.00300 ± 0.00478	1.15×10^{-1c}	0.153 ± 0.111	2.26×10^{-1c}
Stage II	3	0.0893 ± 0.0302		0.0400 ± 0.0271		0.00100 ± 0.00100		0.195 ± 0.116	
Stage III	8	0.127 ± 0.090		0.0930 ± 0.0446		0.0085 ± 0.0117		0.147 ± 0.083	
Stage IV	2	0.0520 ± 0.0014		0.0285 ± 0.0078		0.0170 ± 0.0099		0.0638 ± 0.0021	

Clinicopathological parameters	Number of Tumors	*FAM107A* Expression[a]	*FAM107A* *P*	*CDC16* Expression[a]	*CDC16* *P*	*CPSF3* Expression[a]	*CPSF3* *P*	*ASAP2* Expression[a]	*ASAP2* *P*
Macroscopic configuration[b]									
Type 1	13	0.312 ± 0.184	5.51×10^{-2c}	0.113 ± 0.054	1.35×10^{-1c}	0.0476 ± 0.0255	2.78×10^{-1c}	0.0455 ± 0.0346	1.85×10^{-1c}

TABLE 4. (Continued)

Clinicopathological parameters	Number of Tumors	*FAM107A* Expression[a]	*FAM107A* *P*	*CDC16* Expression[a]	*CDC16* *P*	*CPSF3* Expression[a]	*CPSF3* *P*	*ASAP2* Expression[a]	*ASAP2* *P*
Type 2	5	0.0986 ± 0.0779		0.0584 ± 0.0377		0.0260 ± 0.0151		0.0218 ± 0.0191	
Type 3	8	0.203 ± 0.242		0.0926 ± 0.0753		0.0465 ± 0.0286		0.0250 ± 0.0227	
Histological grades[d,e]									
G1	1	0.685	2.14×10^{-2c}	0.172	1.28×10^{-2c}	0.0613	1.30×10^{-1c}	0.112	1.93×10^{-2c}
G2	6	0.209 ± 0.140		0.113 ± 0.083		0.0575 ± 0.0335		0.0399 ± 0.0320	
G3	11	0.313 ± 0.197		0.116 ± 0.049		0.0464 ± 0.0236		0.0411 ± 0.0252	
G4	8	0.100 ± 0.129		0.0475 ± 0.0206		0.0256 ± 0.0118		0.0122 ± 0.0068	
Vascular involvement[f]									
Negative	13	0.258 ± 0.182	3.11×10^{-1g}	0.123 ± 0.063	2.56×10^{-2g}	0.0518 ± 0.0271	1.13×10^{-1g}	0.0508 ± 0.030	7.95×10^{-4g}
Positive	13	0.217 ± 0.226		0.0696 ± 0.0458		0.0344 ± 0.0214		0.0185 ± 0.0203	
Growth pattern[e]									
Expansive	21	0.255 ± 0.211	4.47×10^{-1g}	0.0975 ± 0.0608	8.01×10^{-1g}	0.0437 ± 0.0263	9.00×10^{-1g}	0.0372 ± 0.0316	4.09×10^{-1g}
Infiltrative	5	0.166 ± 0.159		0.0914 ± 0.0662		0.0409 ± 0.0246		0.0240 ± 0.0211	
Tumor necrosis									
Negative	16	0.317 ± 0.205	2.24×10^{-3g}	0.121 ± 0.063	5.02×10^{-3g}	0.0535 ± 0.0266	1.44×10^{-2g}	0.0475 ± 0.0317	2.77×10^{-3g}
Positive	10	0.110 ± 0.117		0.0567 ± 0.0278		0.0266 ± 0.0115		0.0142 ± 0.0086	
Renal pelvic invasion									
Negative	23	0.259 ± 0.205	7.85×10^{-2g}	0.100 ± 0.063	3.52×10^{-1g}	0.0453 ± 0.0263	3.95×10^{-1g}	0.0368 ± 0.0313	4.42×10^{-1g}
Positive	3	0.0726 ± 0.0602		0.0650 ± 0.0128		0.0263 ± 0.0073		0.0179 ± 0.0058	
Distant metastasis									
Negative	24	0.256 ± 0.199	5.54×10^{-2g}	0.102 ± 0.059	2.46×10^{-2g}	0.0457 ± 0.0249	1.23×10^{-2g}	0.0363 ± 0.0307	3.94×10^{-1g}
Positive	2	0.0165 ± 0.0030		0.0245 ± 0.0007		0.0125 ± 0.0019		0.0151 ± 0.0005	
Pathological TNM stage[h]									
Stage I	13	0.246 ± 0.152	2.72×10^{-1c}	0.109 ± 0.053	4.59×10^{-2c}	0.0445 ± 0.0279	9.67×10^{-2c}	0.0404 ± 0.0279	1.86×10^{-1c}
Stage II	3	0.209 ± 0.167		0.0513 ± 0.0307		0.0324 ± 0.0170		0.0103 ± 0.0082	
Stage III	8	0.290 ± 0.284		0.110 ± 0.071		0.0526 ± 0.0220		0.0394 ± 0.0376	
Stage IV	2	0.0165 ± 0.0030		0.0245 ± 0.0007		0.0125 ± 0.0019		0.0151 ± 0.0005	

[a]Average mRNA levels/GAPDH ± standard deviation.
[b]Macroscopic configuration was evaluated on the basis of previously described criteria (Arai et al., 2006).
[c]Kruskal–Wallis test. *P* values of < 0.05 are underlined.
[d]All the tumors were graded on the basis of previously described criteria (Fuhrman et al., 1982).
[e]If the tumor showed heterogeneity, the most aggressive features of the tumor were described.
[f]The presence or absence of vascular involvement was examined microscopically on slides stained with hematoxylin-eosin and elastica van Gieson.
[g]Mann–Whitney *U* test. *P* values of <0.05 are underlined.
[h]All the tumors were classified according to the pathological Tumor-Node-Metastasis classification (Sobin et al., 2009). Although no significant correlation between expression of any of the 26 chimeric transcripts and clinicopathological parameters was observed in the initial cohort (Supporting Information, Table S4), downregulation of mRNA levels for each of the partner genes did show significant correlations with the above clinicopathological parameters in the second cohort.

deviation) age of 57.1 ± 10.8 years (range, 33–81 years). Copy number analysis using the HumanOmni1-Quad BeadChip (Illumina, San Diego, CA) and Global Parameter Hidden Markov Model (http://bioinformatics.ustc.edu.cn/gphmm/; Li et al., 2011) revealed copy number alterations in chromosome 3 in all 91 clear cell RCCs in the initial and second cohorts (with three exceptions, Supporting Information Tables S1 and S2 and Supporting Information Fig. S1). These findings were considered to be the hallmark of clear cell RCCs in the initial and second cohorts. The clinicopathological parameters of RCCs belonging to the initial and second cohorts are summarized in Table 1.

Tissue specimens were taken and frozen immediately after surgical removal, and thereafter stored in liquid nitrogen until use. These tissue specimens were provided by the National Cancer Center Biobank, Tokyo. This study was approved by the Ethics Committees of the National Cancer Center and National Center for Global Health and Medicine, Tokyo, and was performed in accordance with the Declaration of Helsinki. All the patients provided written informed consent prior to inclusion in the study.

Whole Transcriptome Analysis

Total RNA was isolated using TRIzol reagent (Life Technologies, Carlsbad, CA). A total of 84 (73 T and 11 N) samples in the initial cohort were subjected to whole transcriptome analysis. Sequencing libraries were prepared from 1.0 to 2.5 μg of total RNA using an mRNA-Seq Sample Prep Kit or a TruSeq RNA Sample Prep Kit (Illumina), according to the manufacturer's standard protocols. An mRNA-Seq Sample Prep Kit was used for libraries of 35 (30 T and 5 N) samples, and these libraries were prepared using a procedure including a gel purification step, in which a

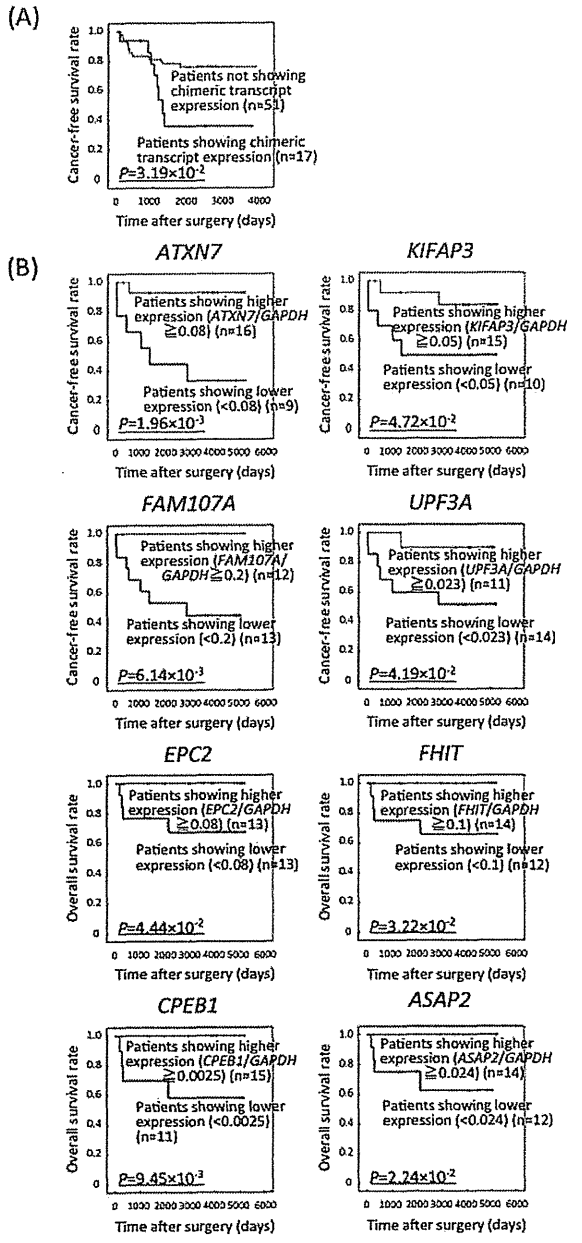


Figure 2. Kaplan–Meier survival curves of patients with clear cell RCCs in the initial (A) and second (B) cohorts. (A) Expression of any of 26 chimeric transcripts was inversely correlated with the cancer-free survival rate of patients in the initial cohort (the log-rank test, $P = 3.19 \times 10^{-2}$). (B) ROC curves were generated for levels of mRNA expression of each partner gene of chimeric transcripts, and the thresholds were set to the top left corner of the graph (data not shown). Using these thresholds, Kaplan–Meier curves were generated. mRNA levels for the *ATXN7* ($P = 1.96 \times 10^{-3}$), *KIFAP3* ($P = 4.72 \times 10^{-2}$), *FAM107A* ($P = 6.14 \times 10^{-3}$), and *UPF3A* ($P = 4.19 \times 10^{-2}$) genes in T samples were inversely correlated with the cancer-free survival rate of patients who underwent complete resection ($n = 25$), whereas those for the *EPC2* ($P = 4.44 \times 10^{-2}$), *FHIT* ($P = 3.22 \times 10^{-2}$), *CPEB1* ($P = 9.45 \times 10^{-3}$), and *ASAP2* ($P = 2.24 \times 10^{-2}$) genes were inversely correlated with the overall survival rate of all patients ($n = 26$) in the second cohort.

fraction of 250–300 bp (insert size: 150–200 bp) was collected. A TruSeq RNA Sample Prep Kit was used for libraries of the other 49 (43 T and 6 N) samples, and these libraries were prepared without gel purification. The resulting libraries were subjected to paired-end sequencing of 50-base reads on a GAIIx or HiSeq2000 sequencer (Illumina).

Detection of Chimeric Transcripts

To avoid multiple counting of each chimeric transcript, RNA sequencing data were used after removal of paired-end reads with the identical nucleotide sequence, which had probably been derived from PCR duplicates during library preparation. For prediction of chimeric transcripts, the deFuse program version 0.4.3 was used (McPherson et al., 2011). After applying default filtering of this program, potential alternative splicing and read-through products that the program predicted were eliminated, and candidates that had exon boundary junctions were selected. Finally, we discarded candidates that were also predicted from the data of 11 N samples.

RT-PCR and Sanger Sequencing

cDNA was reverse-transcribed from the total RNA (500 ng) of the initial cohort samples, in which candidate chimeric transcripts were detected by whole transcriptome analysis, using random primers and Superscript III RNase H⁻Reverse Transcriptase (Life Technologies). cDNA (corresponding to 10 ng total RNA) was subjected to PCR amplification using an optimal DNA polymerase among AmpliTaq Gold DNA Polymerase (Life Technologies), HotStar Taq DNA polymerase (Qiagen, Hilden, Germany) or KAPA Taq DNA Polymerase (KAPA Biosystems, Woburn, MA). The PCR products were separated electrophoretically on 2% agarose gel and stained with ethidium bromide to confirm that specific products of the size estimated on the basis of whole transcriptome analysis were obtained, and that no nonspecific products appeared on amplification. The PCR products were then directly sequenced in both directions using the same primers with the BigDye Terminator v3.1 Cycle Sequencing kit and an ABI 3130xl DNA Sequencer (Life Technologies).

Genomic PCR and Sanger Sequencing

High-molecular-weight genomic DNA was extracted from the initial cohort samples, in which candidate chimeric transcripts were verified by the above RT-PCR and Sanger sequencing, using phenol–chloroform followed by dialysis. Genomic

DNA (10 ng) was subjected to PCR amplification using an optimal DNA polymerase among AmpliTaq Gold DNA Polymerase, Platinum Taq DNA-polymerase high fidelity (Life Technologies), HotStar Taq DNA polymerase (Qiagen) or KAPA Taq DNA Polymerase (KAPA Biosystems). The PCR products were separated electrophoretically on 1% agarose gel and stained with ethidium bromide to confirm that no nonspecific products appeared on amplification. The PCR products were then directly sequenced in both directions using the same primers with the BigDye Terminator v3.1 Cycle Sequencing kit and an ABI 3130xl DNA Sequencer (Life Technologies).

Quantitative RT-PCR Analysis

cDNA was reverse transcribed from total RNA (500 ng) of the 26 paired T and N samples of the second cohort using random primers and Superscript III RNase H⁻Reverse Transcriptase (Life Technologies). mRNA expression was analyzed using custom TaqMan Gene Expression Assays and TaqMan Fast Advanced Master Mix (Life Technologies) on a 7500 Fast Real-Time PCR System (Life Technologies) employing the relative standard curve method. Experiments were performed in triplicate for each sample-primer set, and the mean value for the three experiments was used as the CT value. All CT values were normalized to that of *GAPDH* in the same sample.

Statistics

Differences in clinicopathological parameters between the initial and second cohorts were assayed by Mann–Whitney *U* test and Fisher's exact test. Differences in the levels of mRNA expression between N and T samples were examined by Mann–Whitney *U* test. Correlations between levels of mRNA expression and clinicopathological parameters were assayed by Kruskal–Wallis test and Mann–Whitney *U* test. Receiver operating characteristic (ROC) curves were generated for the levels of mRNA expression of each partner gene involved in the chimeric transcripts, and the thresholds were set at the top left corner of the graph. Subsequently, the impact of chimeric transcript expression and downregulation of mRNA levels for each partner gene on patient outcome was analyzed by the Kaplan–Meier method using the set thresholds and the log-rank test. Differences at $P < 0.05$ were considered to be significant.

RESULTS

Identification of Novel Chimeric Transcripts in RCCs of the Initial Cohort

We performed RNA sequencing of 68 T samples (K1 to K68) and 11 N samples in the initial cohort, and a T sample (K69) showing histological findings compatible with Xp11.2 translocation RCC. At least 30,000,000 reads (average read count 50,000,000) were obtained for each sample. The deFuse program version 0.4.3 (McPherson et al., 2011) provided 3,746 fusion gene candidates from the data obtained using the 69 T samples by applying default filtering. From those candidates, 95 were extracted by eliminating potential alternative splicing and read-through products that the program predicted, and by selecting candidates that had exon boundary junctions. Next, candidates that were predicted even from the data obtained using the 11 N samples were discarded, and finally 35 candidates were obtained. Three candidates were abandoned because of difficulty with the primer design and shortage of samples, and then RT-PCR analysis was performed for the 32 candidates in the same T sample. The PCR and sequencing primers used are shown in Supporting Information Table S3.

After a T sample (K96) of Xp11.2 translocation RCC and three T samples of papillary RCCs (K97 to K99) had been additionally analyzed for comparison, expression of 33 fusion transcripts (including two transcripts [*MMACHC-BX004987.7* and *TFE3-RBM10*] consisting of the same partner gene sets with a different exon boundary or a different transcriptional direction and three transcripts sharing a partner gene, *TFE3*) from the 61 genes was finally verified by RT-PCR, and the exon boundaries and flanking sequences were determined by Sanger sequencing analysis (Table 2 and Supporting Information Fig. S2).

Previously reported in-frame fusion transcripts including *TFE3* (Table 2B; Clark et al., 1997), which are attributable to translocation of the X chromosome, were detected in samples K69 and K96 showing histological findings compatible with Xp11.2 translocation RCC, indicating the reliability of our study. Other than *TFE3* fusion transcripts, three additional transcripts (*EEF2-ENHO*, *PARG-BMS1*, and *RAGE-EML1*, Table 2B) and one additional transcript (*DPP6-ACTR3B*, Table 2B), which have never been reported in RCCs, were also detected in the K69 and K96, respectively. *NONO-TFE3*, *PARG-BMS1*, *RAGE-EML1*, *RBM10-TFE3*, and *DPP6-ACTR3B* transcripts

were predicted to generate in-frame chimeric proteins. These observations of additional chimeric transcripts in K69 and K96 were different from the previously reported characteristics of RCCs associated with Xp11.2 translocation [Pflueger et al. (2013) reported that expression of the *TMED6-COG8* chimeric transcript and higher expression levels of the *EEF1A2* and *CNTN3* genes characterize RCCs associated with Xp11.2 translocation]. K69 showed grade 3 histology, vascular involvement, and tumor necrosis in surgically resected materials, and the patient developed lymph node metastasis 6 months after surgery, whereas K96 showed grade 3 histology. Such phenotypes, especially those of K69, which are more aggressive than those generally described for RCCs with Xp11.2 translocation (Eble et al., 2004), may be attributable to expression of multiple additional chimeric transcripts. Conversely, in three papillary RCCs (K97 to K99) analyzed for comparison, no chimeric transcript was detected.

All 26 chimeric transcripts detected in the initial cohort of clear cell RCCs (Table 2A) have never been reported previously. Even though chimeric transcripts involving the *FHIT* and *TERT* genes have recently been sequenced by The Cancer Genome Atlas (TCGA, The Cancer Genome Atlas Research Network, 2013), the partner gene of *FHIT*, *FAM172A*, and that of *TERT*, *PDCD6*, listed in TCGA each differed from those (*ATXN7* and *TPPP*) in the present study. Each of the detected chimeric transcripts was expressed in a single clear cell RCC. *ANTXR1-GKN1*, *ERBB2-LTBP4*, *POLR2G-CYP1A2*, *AC010724.1-CPEB1*, and *CPSF3-ASAP2* chimeric transcripts were predicted to generate in-frame chimeric proteins, whereas other chimeric transcripts resulted in a premature stop codon in the 3′-partner gene or were generated in the untranslated regions.

The chimeric transcripts were expressed in 17 clear cell RCCs in the initial cohort [17/68 (Table 2), 25%]. Samples K1 and K5 had multiple chimeric transcripts (Table 2). No significant correlation between expression of any of 26 chimeric transcripts and clinicopathological parameters was observed in the initial cohort (Supporting Information Table S4). However, when examined individually, each clear cell RCC with chimeric transcripts showed tumor aggressiveness: e.g., K11 carrying a *TEX264-FAM107A* chimeric transcript showed a type 3 macroscopic configuration and K15 carrying a *CPSF3-ASAP2* chimeric transcript showed a type 3 macroscopic configuration, grade 4 histology, vascular involvement, an invasive growth pattern, and tumor necrosis. Moreover, expression of any of the 26 chimeric transcripts was inversely correlated with the cancer-free survival rate of patients in the initial cohort (the period covered ranged from 42 to 4,783 days [mean, 2,015 days]; log-rank test, $P = 3.19 \times 10^{-2}$; Fig. 2).

Identification of Genomic Breakpoints in RCCs of the Initial Cohort

Long-range genomic PCR and Sanger sequencing were performed for 17 clear cell RCCs (K1 to K17) harboring chimeric transcripts using the primers shown in Supporting Information Table S5. Genomic breakpoints for five chimeric transcripts, *POLR2G-CYP1A2*, *AC010724.1-CPEB1*, *SEMA6A-CAMK4*, *ASAP1-ADCY8*, and *CPSF3-ASAP2*, were successfully revealed, but the genomic PCR failed for the other transcripts. The genomic breakpoints for these five chimeric transcripts are summarized in Table 3.

Levels of mRNA Expression for the Genes Involved in Chimeric Transcripts

The levels of mRNA expression for 20 representative partner genes involved in chimeric transcripts in the initial cohort were quantitatively examined in 26 paired T and N samples in the second cohort. The probes and PCR primer sets used are shown in Supporting Information Table S6.

The levels of mRNA expression for the *MMACHC*, *PTER*, *EPC2*, *ATXN7*, *FHIT*, *KIFAP3*, *CPEB1*, *MINPP1*, *TEX264*, *FAM107A*, *UPF3A*, *CDC16*, *MCCC1*, *CPSF3*, and *ASAP2* genes were significantly reduced in T samples relative to the corresponding N samples (Fig. 1, Mann–Whitney *U* test, $P = 3.38 \times 10^{-12}$, $P = 9.04 \times 10^{-7}$, $P = 8.08 \times 10^{-3}$, $P = 6.21 \times 10^{-4}$, $P = 2.71 \times 10^{-11}$, $P = 9.46 \times 10^{-4}$, $P = 2.18 \times 10^{-5}$, $P = 1.03 \times 10^{-2}$, $P = 2.47 \times 10^{-9}$, $P = 9.90 \times 10^{-5}$, $P = 2.53 \times 10^{-3}$, $P = 1.89 \times 10^{-4}$, $P = 1.62 \times 10^{-5}$, $P = 1.77 \times 10^{-5}$, and $P = 7.17 \times 10^{-9}$, respectively). The levels of mRNA expression for the *MMACHC*, *PTER*, *EPC2*, *ATXN7*, *FHIT*, *KIFAP3*, *CPEB1*, *TEX264*, *FAM107A*, *CDC16*, *CPSF3*, and *ASAP2* genes in T samples in the second cohort were significantly correlated with clinicopathological parameters reflecting tumor aggressiveness, such as invasive macroscopic configuration, higher histological grades, vascular involvement, invasive growth pattern, tumor necrosis, renal pelvic invasion, distant metastasis,

and higher TNM stages (Table 4). Moreover, mRNA levels for the *ATXN7* ($P = 1.96 \times 10^{-3}$), *KIFAP3* ($P = 4.72 \times 10^{-2}$), *FAM107A* ($P = 6.14 \times 10^{-3}$), and *UPF3A* ($P = 4.19 \times 10^{-2}$) genes in T samples were inversely correlated with the cancer-free survival rate, whereas those for the *EPC2* ($P = 4.44 \times 10^{-2}$), *FHIT* ($P = 3.22 \times 10^{-2}$), *CPEB1* ($P = 9.45 \times 10^{-3}$), and *ASAP2* ($P = 2.24 \times 10^{-2}$) genes were inversely correlated with the overall survival rate in the second cohort (the period covered ranged from 88 to 5,207 days [mean, 3,038 days], the log-rank test; Fig. 2).

DISCUSSION

To comprehensively explore chimeric transcripts in clear cell RCCs, whole transcriptome analysis was performed using tissue specimens. The significance of generation of chimeric transcripts has not been fully elucidated in adult solid tumors other than well-studied exceptions, such as sarcomas and adenocarcinomas of the prostate and the lung. Although previous reports of fusion events involving the *ALK* gene based on FISH and immunohistochemistry have been restricted to nonclear cell RCCs (Sugawara et al., 2012), when comprehensively explored using next-generation sequencing technology, chimeric transcripts were detected in 25% (17/68) of the clear cell RCCs. In some RCCs (K1 and K5), multiple chimeric transcripts were observed. Moreover, the genomic breakpoints revealed for five chimeric transcripts in clear cell RCCs indicate that such transcripts have actually arisen through genomic rearrangement. Gene fusion events may thus play a greater role in renal carcinogenesis than previously anticipated. Conversely, mechanisms other than genomic rearrangements (Yuan et al., 2013), e.g., trans-splicing (Li et al., 2008), may generate chimeric transcripts in which genomic breakpoints are not revealed.

The WHO classification defines RCC associated with Xp11.2 translocation, which involves *TFE3* fusion, as a distinct subtype (Eble et al., 2004). Diagnosis of RCC associated with Xp11.2 translocation depends on detection of *TFE3* protein overexpression using immunohistochemistry or detection of gene fusion using FISH and/or RT-PCR analysis (Green et al., 2013; Rao et al., 2013). The procedure for final diagnosis of RCC associated with Xp11.2 translocation differs from that for other RCC subtypes, such as clear cell RCC, papillary RCC, and chromophobe RCC, which generally can be diagnosed on the basis of histological observation. As RCC associated with Xp11.2 translocation and other RCCs were lumped into the same category as the RCC subtypes, the final diagnosis of RCC subtypes could not be made based solely on conventional histological examination of surgically resected or biopsy specimens. As the present comprehensive study demonstrated multiple chimeric transcripts in various RCCs, it seems that the use of Xp11.2 translocation as the only criterion for defining a distinct subtype of RCC may not be a rational approach. The classification of RCC subtypes should therefore be revised after a more comprehensive appraisal of correlations between histological features and genetic background.

All 26 chimeric transcripts detected in the initial cohort were novel chimeric transcripts that have never been reported previously in RCCs. However, only five of them were predicted to generate in-frame chimeric proteins in clear cell RCCs. Expression microarray analysis did not necessarily suggest prominent overexpression of in-frame chimeric transcripts in the initial cohort (data not shown). Moreover, in-frame chimeric transcripts observed in clear cell RCCs do not necessarily result in constitutive activation of protein kinases, which frequently cause addiction for gene fusion events. Conversely, many genes for which reduced expression and/or tumor suppressive function have been reported in human cancers were included in chimeric transcripts observed in the initial cohort. Therefore, we examined the levels of mRNA expression for 20 representative genes involved in chimeric transcripts in the second cohort (Supporting Information Table S6) and revealed significantly reduced mRNA expression of the *MMACHC, PTER, EPC2, ATXN7, FHIT, KIFAP3, CPEB1, MINPP1, TEX264, FAM107A, UPF3A, CDC16, MCCC1, CPSF3*, and *ASAP2* genes in T samples in the second cohort (Fig. 1).

It has been reported that reduced expression of the *MMACHC* gene, which participates in intracellular trafficking of cobalamin, can result in increased tumorigenicity and methionine dependence of cancer cells (Loewy et al., 2009). Although its implication in human cancers has been unclear, the *PTER* gene was first cloned as a rat homolog of bacterial phosphotriesterase, and its expression in the normal proximal tubules of the kidney has been reported (Davies et al., 1997). Single nucleotide polymorphism (SNP) of the *EPC2* gene has been reported to be associated with response to gemcitabine in human cancer cell lines (Jarjanazi et al., 2008). SNP of the *ATXN7* gene, which

encodes a subunit of the GCN5 histone acetyltransferase-containing coactivator complex (Helmlinger et al., 2006), is reportedly associated with susceptibility to breast cancer (Milne et al., in press). The fragile *FHIT* gene, encompassing the chromosomal fragile site *FRA3B*, is a target of DNA damage-induced cancer initiation and progression through modulation of genomic stability (Karras et al., 2014). KIFAP3 is colocalized with KIF3, which participates in subcellular transport of several cancer-related proteins including beta-catenin and cadherins (Tanuma et al., 2009). Down regulation of *CPEB1*, which participates in the regulation of mRNA translation and processing of the 3′ untranslated region (Bava et al., 2013), has been reported in human cancers (Caldeira et al., 2012). As has been reported for the *PTEN* gene, somatic mutation and germline variants of the *MINPP1* gene, located in proximity to *PTEN* in 10q23.3, have been reported in patients with follicular thyroid tumors (Gimm et al., 2001). *FAM107A* was first identified in a commonly deleted region in 3p21 in RCCs (Wang et al., 2000), and transfection of this gene induces growth suppression and apoptosis of FAM107A-negative cancer cell lines (Wang et al., 2000; Liu et al., 2009). UPF3A is a crucial factor of nonsense-mediated decay, an RNA decay pathway that downregulates aberrant mRNAs (Chan et al., 2009). CDC16 is a component of the Anaphase Promoting Complex/Cyclosome (APC/C), which governs cell cycle progression and has crucial functions in maintaining genomic integrity and tumorigenesis (Zhang et al., 2014). Genetic imbalance at the *MCCC1* gene locus has recently been reported in clinical specimens of oral squamous cell carcinoma (Ribeiro et al., 2014). *CPSF3* is required for site-specific endonucleolytic cleavage and poly (A) addition (Keller and Minvielle-Sebastia, 1997) and directly interacts with (Zhu et al., 2009) tumor suppressor gene product CSR1 (Yu et al., 2006). The src homology 3 domain of the paxillin-binding protein (Kondo et al., 2000; Coutinho-Camillo et al., 2006), *ASAP2*, directly interacts with the SAMP repeat region of the *APC* tumor suppressor gene (Matsui et al., 2008).

Although the *TEX264* gene has been simply identified as one of the protein-encoding open reading frames deposited in a database (Lamesch et al., 2007), the above characteristics of each of the partner genes suggest that down-regulation of the *MMACHC, PTER, EPC2, ATXN7, FHIT, KIFAP3, CPEB1, MINPP1, FAM107A, UPF3A, CDC16, MCCC1, CPSF3*, and *ASAP2* genes may participate in renal carcinogenesis. Moreover, the levels of mRNA expression for many of the partner genes in T samples were significantly correlated with the clinicopathological aggressiveness of RCCs (Table 4) and were inversely correlated with the cancer-free and/or overall survival rates of patients with clear cell RCCs (Fig. 2), indicating that such reduced expression may continue to play a role in multistage malignant progression during renal carcinogenesis.

Even if the same chimeric transcripts detected in the initial cohorts had been expressed in the second cohort, quantitative RT-PCR analysis for each partner gene would not have evaluated chimeric transcripts lacking target exons (Supporting Information Table S6). Therefore, to reveal the presence or absence of the same chimeric transcripts detected in the initial cohort, RT-PCR analysis using total RNA samples and the primer sets indicated in Supporting Information Table S3 and long-range PCR analysis using genomic DNA samples and the primer sets described in Supporting Information Table S5 were performed in the second cohort. These analyses did not detect the same chimeric transcripts in the second cohort (data not shown). As all detected chimeric transcripts were expressed only in a single clear cell RCC in the initial cohort, it is possible that the same chimeric transcripts may have been absent in the second cohort. Downregulation of mRNA levels for each of the genes described in Figure 1 in the second cohort would have been attributable to mechanisms other than expression of chimeric transcripts, such as gene deletion, DNA methylation status around the promoter regions and/or alterations in the expression levels, and accessibility of transcription factors. In fact, silencing of the *MMACHC* (Loewy et al., 2009) and *CPEB1* (Caldeira et al., 2012) genes due to DNA methylation, and gene deletion and DNA methylation of *FHIT* (Karras et al., 2014), have been reported in human cancers. However, further studies are needed to reveal the mechanisms responsible for downregulation of each of the partner genes in the second cohort.

Conversely, it is feasible that dysfunction of each partner gene is induced by generation of chimeric transcripts in the initial cohort of clear cell RCCs, as such mechanisms of tumor suppressor gene functional impairment have been reported in adult malignancies such as acute myeloid leukemia (McNerney et al., 2013). Even though prominent overexpression and/or constitutive activation of growth factors and/or protein kinases due to

gene fusion events is rare, generation of chimeric transcripts may participate in renal carcinogenesis through dysfunction of tumor-related genes.

REFERENCES

Arai E, Kanai Y. 2010. Genetic and epigenetic alterations during renal carcinogenesis. Int J Clin Exp Pathol 4:58–73.

Arai E, Kanai Y, Ushijima S, Fujimoto H, Mukai K, Hirohashi S. 2006. Regional DNA hypermethylation and DNA methyltransferase (DNMT) 1 protein overexpression in both renal tumors and corresponding nontumorous renal tissues. Int J Cancer 119: 288–296.

Baldewijns MM, van Vlodrop IJ, Vermeulen PB, Soetekouw PM, van Engeland M, de Bruïne AP. 2010. VHL and HIF signalling in renal cell carcinogenesis. J Pathol 221:125–138.

Bava FA, Eliscovich C, Ferreira PG, Miñana B, Ben-Dov C, Guigó R, Valcárcel J, Méndez R. 2013. CPEB1 coordinates alternative 3′-UTR formation with translational regulation. Nature 495:121–125.

Caldeira J, Simões-Correia J, Paredes J, Pinto MT, Sousa S, Corso G, Marrelli D, Roviello F, Pereira PS, Weil D, Oliveira C, Casares F, Seruca R. 2012. CPEB1, a novel gene silenced in gastric cancer: A Drosophila approach. Gut 61:1115–1123.

Cantile M, Marra L, Franco R, Ascierto P, Liguori G, De Chiara A, Botti G. 2013. Molecular detection and targeting of EWSR1 fusion transcripts in soft tissue tumors. Med Oncol 30:412.

Chan WK, Bhalla AD, Le Hir H, Nguyen LS, Huang L, Gécz J, Wilkinson MF. 2009. A UPF3-mediated regulatory switch that maintains RNA surveillance. Nat Struct Mol Biol 16:747–753.

Clark J, Lu YJ, Sidhar SK, Parker C, Gill S, Smedley D, Hamoudi R, Linehan WM, Shipley J, Cooper CS. 1997. Fusion of splicing factor genes PSF and NonO (p54nrb) to the TFE3 gene in papillary renal cell carcinoma. Oncogene 15:2233–2239.

Coutinho-Camillo CM, Salaorni S, Sarkis AS, Nagai MA. 2006. Differentially expressed genes in the prostate cancer cell line LNCaP after exposure to androgen and anti-androgen. Cancer Genet Cytogenet 166:130–138.

Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, Davies H, Edkins S, Hardy C, Latimer C, Teague J, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Forbes S, Jia M, Jones D, Knott H, Kok CY, Lau KW, Leroy C, Lin ML, McBride DJ, Maddison M, Maguire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, O'Meara S, Pleasance E, Rajasingham A, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turrell K, Dykema KJ, Khoo SK, Petillo D, Wondergem B, Anema J, Kahnoski RJ, Teh BT, Stratton MR, Futreal PA. 2010. Systematic sequencing of renal carcinoma reveals inactivation of histone-modifying genes. Nature 463:360–363.

Davies JA, Buchman VL, Krylova O, Ninkina NN. 1997. Molecular cloning and expression pattern of rpr-1, a resiniferatoxin-binding, phosphotriesterase-related protein, expressed in rat kidney tubules. FEBS Lett 410:378–382.

Debelenko LV, Raimondi SC, Daw N, Shivakumar BR, Huang D, Nelson M, Bridge JA. 2011. Renal cell carcinoma with novel VCL-ALK fusion: New representative of ALK-associated tumor spectrum. Mod Pathol 24:430–442.

Eble JN, Togashi K, Pisani P. 2004. Renal cell carcinoma. In: Eble JN, Sauter G, Epstein JI, Sesterhenn IA, editors. World Health Organization classification of tumours. Pathology and genetics. Tumours of the urinary system and male genital organs. Lyon: IARC, pp. 10–43.

Fuhrman SA, Lasky LC, Limas, C. 1982. Prognostic significance of morphologic parameters in renal cell carcinoma. Am J Surg Pathol 6:655–663.

Gimm O, Chi H, Dahia PL, Perren A, Hinze R, Komminoth P, Dralle H, Reynolds PR, Eng C. 2001. Somatic mutation and germline variants of MINPP1, a phosphatase gene located in proximity to PTEN on 10q23.3, in follicular thyroid carcinomas. J Clin Endocrinol Metab 86:1801–1805.

Green WM, Yonescu R, Morsberger L, Morris K, Netto GJ, Epstein JI, Illei PB, Allaf M, Ladanyi M, Griffin CA, Argani P. 2013. Utilization of a TFE3 break-apart FISH assay in a renal tumor consultation service. Am J Surg Pathol 37:1150–1163.

Guo G, Gui Y, Gao S, Tang A, Hu X, Huang Y, Jia W, Li Z, He M, Sun L, Song P, Sun X, Zhao X, Yang S, Liang C, Wan S, Zhou F, Chen C, Zhu J, Li X, Jian M, Zhou L, Ye R, Huang P, Chen J, Jiang T, Liu X, Wang Y, Zou J, Jiang Z, Wu R, Wu S, Fan F, Zhang Z, Liu L, Yang R, Liu X, Wu H, Yin W, Zhao X, Liu Y, Peng H, Jiang B, Feng Q, Li C, Xie J, Lu J, Kristiansen K, Li Y, Zhang X, Li S, Wang J, Yang H, Cai Z, Wang J. 2012. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. Nat Genet 44:17–19.

Helmlinger D, Hardy S, Abou-Sleymane G, Eberlin A, Bowman AB, Gansmüller A, Picaud S, Zoghbi HY, Trottier Y, Tora L, Devys D. 2006. Glutamine-expanded ataxin-7 alters TFTC/STAGA recruitment and chromatin structure leading to photoreceptor dysfunction. PLoS Biol 4:e67.

Jarjanazi H, Kiefer J, Savas S, Briollais L, Tuzmen S, Pabalan N, Ibrahim-Zada I, Mousses S, Ozcelik H. 2008. Discovery of genetic profiles impacting response to chemotherapy: Application to gemcitabine. Hum Mutat 29:461–467.

Karras JR, Paisie CA, Huebner K. 2014. Replicative Stress and the FHIT Gene: Roles in tumor suppression, genome stability and prevention of carcinogenesis. Cancers (Basel) 6:1208–1219.

Keller W, Minvielle-Sebastia L. 1997. A comparison of mammalian and yeast pre-mRNA 3′-end processing. Curr Opin Cell Biol 9:329–336.

Kohno T, Ichikawa H, Totoki Y, Yasuda K, Hiramoto M, Nammo T, Sakamoto H, Tsuta K, Furuta K, Shimada Y, Iwakawa R, Ogiwara H, Oike T, Enari M, Schetter AJ, Okayama H, Haugen A, Skaug V, Chiku S, Yamanaka I, Arai Y, Watanabe S, Sekine I, Ogawa S, Harris CC, Tsuda H, Yoshida T, Yokota J, Shibata T. 2012. KIF5B-RET fusions in lung adenocarcinoma. Nat Med 18:375–377.

Kondo A, Hashimoto S, Yano H, Nagayama K, Mazaki Y, Sabe H. 2000. A new paxillin-binding protein, PAG3/Papalpha/KIAA0400, bearing an ADP-ribosylation factor GTPase-activating protein activity, is involved in paxillin recruitment to focal adhesions and cell migration. Mol Biol Cell 11:1315–1327.

Lamesch P, Li N, Milstein S, Fan C, Hao T, Szabo G, Hu Z, Venkatesan K, Bethel G, Martin P, Rogers J, Lawlor S, McLaren S, Dricot A, Borick H, Cusick ME, Vandenhaute J, Dunham I, Hill DE, Vidal M. 2007. hORFeome v3.1: A resource of human open reading frames representing over 10,000 human genes. Genomics 89:307–315.

Li A, Liu Z, Lezon-Geyda K, Sarkar S, Lannin D, Schulz V, Krop I, Winer E, Harris L, Tuck D. 2011. GPHMM: An integrated hidden Markov model for identification of copy number alteration and loss of heterozygosity in complex tumor samples using whole genome SNP arrays. Nucleic Acids Res 39:4928–4941.

Li H, Wang J, Mor G, Sklar J. 2008. A neoplastic gene fusion mimics trans-splicing of RNAs in normal human cells. Science 321:1357–1361.

Liu Q, Zhao XY, Bai RZ, Liang SF, Nie CL, Yuan Z, Wang CT, Wu Y, Chen LJ, Wei YQ. 2009. Induction of tumor inhibition and apoptosis by a candidate tumor suppressor gene DRR1 on 3p21.1. Oncol Rep 22:1069–1075.

Ljungberg B, Campbell SC, Choi HY, Jacqmin D, Lee JE, Weikert S, Kiemeney LA. 2011. The epidemiology of renal cell carcinoma. Eur Urol 60:615–621.

Loewy AD, Niles KM, Anastasio N, Watkins D, Lavoie J, Lerner-Ellis JP, Pastinen T, Trasler JM, Rosenblatt DS. 2009. Epigenetic modification of the gene for the vitamin B(12) chaperone MMACHC can result in increased tumorigenicity and methionine dependence. Mol Genet Metab 96:261–267.

Mariño-Enríquez A, Ou WB, Weldon CB, Fletcher JA, Pérez-Atayde AR. 2011. ALK rearrangement in sickle cell trait-associated renal medullary carcinoma. Genes Chromosomes Cancer 50:146–153.

Matsui C, Kaieda S, Ikegami T, Mimori-Kiyosue Y. 2008. Identification of a link between the SAMP repeats of adenomatous polyposis coli tumor suppressor and the Src homology 3 domain of DDEF. J Biol Chem 283:33006–33020.

McNerney ME, Brown CD, Wang X, Bartom ET, Karmakar S, Bandlamudi C, Yu S, Ko J, Sandall BP, Stricker T, Anastasi J, Grossman RL, Cunningham JM, Le Beau MM, White KP. 2013. CUX1 is a haploinsufficient tumor suppressor gene on chromosome 7 frequently inactivated in acute myeloid leukemia. Blood 121:975–983.

McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG, Griffith M, Heravi Moussavi A, Senz J, Melnyk N, Pacheco M, Marra MA, Hirst M, Nielsen TO, Sahinalp SC, Huntsman D, Shah SP. 2011. deFuse: An algorithm for gene

fusion discovery in tumor RNA-Seq data. PLoS Comput Biol 7: e1001138.

Milne RL, Burwinkel B, Michailidou K, Arias-Perez JI, Zamora MP, Menéndez-Rodríguez P, Hardisson D, Mendiola M, González-Neira A, Pita G, Alonso MR, Dennis J, Wang Q, Bolla MK, Swerdlow A, Ashworth A, Orr N, Schoemaker M, Ko YD, Brauch H, Hamann U; The GENICA Network, Andrulis IL, Knight JA, Glendon G, Tchatchou S, Investigators K; Australian Ovarian Cancer Study Group, Matsuo K, Ito H, Iwata H, Tajima K, Li J, Brand JS, Brenner H, Dieffenbach AK, Arndt V, Stegmaier C, Lambrechts D, Peuteman G, Christiaens MR, Smeets A, Jakubowska A, Lubinski J, Jaworska-Bieniek K, Durda K, Hartman M, Hui M, Lim WY, Chan CW, Marme F, Yang R, Bugert P, Lindblom A, Margolin S, García-Closas M, Chanock SJ, Lissowska J, Figueroa JD, Bojesen SE, Nordestgaard BG, Flyger H, Hooning MJ, Kriege M, van den Ouweland AM, Koppert LB, Fletcher O, Johnson N, Dos-Santos-Silva I, Peto J, Zheng W, Deming-Halverson S, Shrubsole MJ, Long J, Chang-Claude J, Rudolph A, Seibold P, Flesch-Janys D, Winqvist R, Pylkäs K, Jukkola-Vuorinen A, Grip M, Cox A, Cross SS, Reed MW, Schmidt MK, Broeks A, Cornelissen S, Braaf L, Kang D, Choi JY, Park SK, Noh DY, Simard J, Dumont M, Goldberg MS, Labrèche F, Fasching PA, Hein A, Ekici AB, Beckmann MW, Radice P, Peterlongo P, Azzollini J, Barile M, Sawyer E, Tomlinson I, Kerin M, Miller N, Hopper JL, Schmidt DF, Makalic E, Southey MC, Teo SH, Yip CH, Sivanandan K, Tay WT, Shen CY, Hsiung CN, Yu JC, Hou MF, Guénel P, Truong T, Sanchez M, Mulot C, Blot W, Cai Q, Nevanlinna H, Muranen TA, Aittomäki K, Blomqvist C, Wu AH, Tseng CC, Van Den Berg D, Stram DO, Bogdanova N, Dörk T, Muir K, Lophatananon A, Stewart-Brown S, Siriwanarangsan P, Mannermaa A, Kataja V, Kosma VM, Hartikainen JM, Shu XO, Lu W, Gao YT, Zhang B, Couch FJ, Toland AE; TNBCC, Yannoukakos D, Sangrajrang S, McKay J, Wang X, Olson JE, Vachon C, Purrington K, Severi G, Baglietto L, Haiman CA, Henderson BE, Schumacher F, Le Marchand L, Devilee P, Tollenaar RA, Seynaeve C, Czene K, Eriksson M, Humphreys K, Darabi H, Ahmed S, Shah M, Pharoah PD, Hall P, Giles GG, Benítez J, Dunning AM, Chenevix-Trench G, Easton DF. Common non-synonymous SNPs associated with breast cancer susceptibility: Findings from the Breast Cancer Association Consortium. Hum Mol Genet (in press).

Pflueger D, Sboner A, Storz M, Roth J, Compérat E, Bruder E, Rubin MA, Schraml P, Moch H. 2013. Identification of molecular tumor markers in renal cell carcinomas with TFE3 protein expression by RNA sequencing. Neoplasia 15:1231–1240.

Rao Q, Williamson SR, Zhang S, Eble JN, Grignon DJ, Wang M, Zhou XJ, Huang W, Tan PH, Maclennan GT, Cheng L. 2013. TFE3 break-apart FISH has a higher sensitivity for Xp11.2 translocation-associated renal cell carcinoma compared with TFE3 or cathepsin K immunohistochemical staining alone: Expanding the morphologic spectrum. Am J Surg Pathol 37:804–815.

Ribeiro IP, Marques F, Caramelo F, Ferrão J, Prazeres H, Julião MJ, Rifi W, Savola S, de Melo JB, Baptista IP, Carreira IM. 2014. Genetic imbalances detected by multiplex ligation-dependent probe amplification in a cohort of patients with oral squamous cell carcinoma-the first step toward clinical personalized medicine. Tumour Biol 35:4687–4695.

Sato Y, Yoshizato T, Shiraishi Y, Maekawa S, Okuno Y, Kamura T, Shimamura T, Sato-Otsubo A, Nagae G, Suzuki H, Nagata Y, Yoshida K, Kon A, Suzuki Y, Chiba K, Tanaka H, Niida A, Fujimoto A, Tsunoda T, Morikawa T, Maeda D, Kume H, Sugano S, Fukayama M, Aburatani H, Sanada M, Miyano S, Homma Y, Ogawa S. 2013. Integrated molecular analysis of clear-cell renal cell carcinoma. Nat Genet 45:860–867.

Shima Y, Kitabayashi I. 2011. Deregulated transcription factors in leukemia. Int J Hematol 94:134–141.

Sobin LH, Gospodarowicz M, Wittekind C. 2009. International Union Against Cancer (UICC). TNM classification of malignant tumors, 7th ed. New York: Wiley-Liss.

Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, Fujiwara S, Watanabe H, Kurashina K, Hatanaka H, Bando M, Ohno S, Ishikawa Y, Aburatani H, Niki T, Sohara Y, Sugiyama Y, Mano H. 2007. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature 448:561–566.

Sugawara E, Togashi Y, Kuroda N, Sakata S, Hatano S, Asaka R, Yuasa T, Yonese J, Kitagawa M, Mano H, Ishikawa Y, Takeuchi K. 2012. Identification of anaplastic lymphoma kinase fusions in renal cancer: Large-scale immunohistochemical screening by the intercalated antibody-enhanced polymer method. Cancer 118:4427–4436.

Takeuchi K, Soda M, Togashi Y, Suzuki R, Sakata S, Hatano S, Asaka R, Hamanaka W, Ninomiya H, Uehara H, Lim Choi Y, Satoh Y, Okumura S, Nakagawa K, Mano H, Ishikawa Y. 2012. RET, ROS1 and ALK fusions in lung cancer. Nat Med 18:378–381.

Tanuma N, Nomura M, Ikeda M, Kasugai I, Tsubaki Y, Takagaki K, Kawamura T, Yamashita Y, Sato I, Sato M, Katakura R, Kikuchi K, Shima H. 2009. Protein phosphatase Dusp26 associates with KIF3 motor and promotes N-cadherin-mediated cell-cell adhesion. Oncogene 28:752–761.

The Cancer Genome Atlas Research Network. 2013. Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature 499:43–49.

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. 2005. Recurrent fusion of TMPRSS2 and ETS transcription factor genes in prostate cancer. Science 310:644–648.

van Haaften G, Dalgliesh GL, Davies H, Chen L, Bignell G, Greenman C, Edkins S, Hardy C, O'Meara S, Teague J, Butler A, Hinton J, Latimer C, Andrews J, Barthorpe S, Beare D, Buck G, Campbell PJ, Cole J, Forbes S, Jia M, Jones D, Kok CY, Leroy C, Lin ML, McBride DJ, Maddison M, Maquire S, McLay K, Menzies A, Mironenko T, Mulderrig L, Mudie L, Pleasance E, Shepherd R, Smith R, Stebbings L, Stephens P, Tang G, Tarpey PS, Turner R, Turrell K, Varian J, West S, Widaa S, Wray P, Collins VP, Ichimura K, Law S, Wong J, Yuen ST, Leung SY, Tonon G, DePinho RA, Tai YT, Anderson KC, Kahnoski RJ, Massie A, Khoo SK, Teh BT, Stratton MR, Futreal PA. 2009. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. Nat Genet 41:521–523.

Varela I, Tarpey P, Raine K, Huang D, Ong CK, Stephens P, Davies H, Jones D, Lin ML, Teague J, Bignell G, Butler A, Cho J, Dalgliesh GL, Galappaththige D, Greenman C, Hardy C, Jia M, Latimer C, Lau KW, Marshall J, McLaren S, Menzies A, Mudie L, Stebbings L, Largaespada DA, Wessels LF, Richard S, Kahnoski RJ, Anema J, Tuveson DA, Perez-Mancera PA, Mustonen V, Fischer A, Adams DJ, Rust A, Chan-on W, Subimerb C, Dykema K, Furge K, Campbell PJ, Teh BT, Stratton MR, Futreal PA. 2011. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. Nature 469:539–542.

Wang L, Darling J, Zhang JS, Liu W, Qian J, Bostwick D, Hartmann L, Jenkins R, Bardenhauer W, Schutte J, Opalka B, Smith DI. 2000. Loss of expression of the DRR 1 gene at chromosomal segment 3p21.1 in renal cell carcinoma. Genes Chromosomes Cancer 27:1–10.

Yuan H, Qin F, Movassagh M, Park H, Golden W, Xie Z, Zhang P, Sklar J, Li H. 2013. A chimeric RNA characteristic of rhabdomyosarcoma in normal myogenesis process. Cancer Discov 3: 1394–1403.

Yu G, Tseng GC, Yu YP, Gavel T, Nelson J, Wells A, Michalopoulos G, Kokkinakis D, Luo JH. 2006. CSR1 suppresses tumor growth and metastasis of prostate cancer. Am J Pathol 168:597–607.

Zhang J, Wan L, Dai X, Sun Y, Wei W. 2014. Functional characterization of Anaphase Promoting Complex/Cyclosome (APC/C) E3 ubiquitin ligases in tumorigenesis. Biochim Biophys Acta 1845:277–293.

Zhu ZH, Yu YP, Shi YK, Nelson JB, Luo JH. 2009. CSR1 induces cell death through inactivation of CPSF3. Oncogene 28:41–51.

Dig Dis Sci
DOI 10.1007/s10620-014-3476-9

ORIGINAL ARTICLE

Targeted Next-Generation Sequencing Effectively Analyzed the Cystic Fibrosis Transmembrane Conductance Regulator Gene in Pancreatitis

Eriko Nakano · Atsushi Masamune · Tetsuya Niihori · Kiyoshi Kume · Shin Hamada · Yoko Aoki · Yoichi Matsubara · Tooru Shimosegawa

Received: 5 September 2014 / Accepted: 28 November 2014

Abstract
Background The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene, responsible for the development of cystic fibrosis, is known as a pancreatitis susceptibility gene. Direct DNA sequencing of PCR-amplified *CFTR* gene segments is a first-line method to detect unknown mutations, but it is a tedious and labor-intensive endeavor given the large size of the gene (27 exons, 1,480 amino acids). Next-generation sequencing (NGS) is becoming standardized, reducing the cost of DNA sequencing, and enabling the generation of millions of reads per run. We here report a comprehensive analysis of *CFTR* variants in Japanese patients with chronic pancreatitis using NGS coupling with target capture.
Methods Exon sequences of the *CFTR* gene from 193 patients with chronic pancreatitis (121 idiopathic, 46 alcoholic, 17 hereditary, and nine familial) were captured by HaloPlex target enrichment technology, followed by NGS.
Results The sequencing data covered 91.6 % of the coding regions of the *CFTR* gene by ≥20 reads with a mean read depth of 449. We could identify 12 non-synonymous variants including three novel ones [c.A1231G (p.K411E), c.1753G>T (p.E585X) and c.2869delC (p.L957fs)] and seven synonymous variants including three novel ones in the exonic regions. The frequencies of the c.4056G>C (p.Q1352H) and the c.3468G>T (p.L1156F) variants were higher in patients with chronic pancreatitis than those in controls.
Conclusions Target sequence capture combined with NGS is an effective method for the analysis of pancreatitis susceptibility genes.

Keywords Chloride channel · HaloPlex · In silico analysis · MiSeq · Target enrichment

Abbreviations

bp	Base pair
CF	Cystic fibrosis
CFTR	Cystic fibrosis transmembrane conductance regulator
CP	Chronic pancreatitis
ERCP	Endoscopic retrograde cholangiopancreatography
NGS	Next-generation sequencing
PCR	Polymerase chain reaction
RD	Related disorder
SIFT	Sorting Intolerant From Tolerant

Introduction

Chronic pancreatitis (CP) is a progressive inflammatory disease that eventually leads to impairment of the exocrine and endocrine functions of the organ [1, 2]. Since the identification of mutations in the cationic trypsinogen

E. Nakano · A. Masamune (✉) · K. Kume · S. Hamada · T. Shimosegawa
Division of Gastroenterology, Tohoku University Graduate School of Medicine, 1-1 Seiryo-machi, Aoba-ku, Sendai 980-8574, Japan
e-mail: amasamune@med.tohoku.ac.jp

T. Niihori · Y. Aoki · Y. Matsubara
Department of Medical Genetics, Tohoku University Graduate School of Medicine, Sendai 980-8574, Japan

Y. Matsubara
National Research Institute for Child Health and Development, Tokyo 157-8535, Japan

Published online: 10 December 2014

(protease, serine, 1; *PRSS1*) gene as a cause of hereditary pancreatitis in 1996 [3], several pancreatitis susceptibility genes have been identified [3–6]. Gain-of-function mutations in the *PRSS1* gene as well as loss-of-function variants in the serine protease inhibitor Kazal type 1 (*SPINK1*) gene and the trypsin-degrading enzyme chymotrypsin C (*CTRC*) increase the risk of CP [3–5]. In 2013, carboxypeptidase A1 (*CPA1*) gene was identified as a novel pancreatitis susceptibility gene [6]. These studies have been replicated in the Japanese population [7–10].

The cystic fibrosis transmembrane conductance regulator (*CFTR*) gene is another pancreatitis susceptibility gene [11, 12]. Acute recurrent pancreatitis and CP have been accepted as CFTR-related disorders (CFTR-RDs), a clinical entity associated with CFTR dysfunction that does not fulfill the diagnostic criteria for cystic fibrosis (CF; MIM# 219700) [13]. The *CFTR* gene, responsible for the development of CF, encodes for a cyclic adenosine monophosphate-dependent chloride channel that is located in the apical membrane of secretory and absorptive epithelial cells of the pancreas, intestine, liver, airways, vas deferens, and sweat glands [14]. In general, the clinical manifestations of CF arise from ductal and glandular obstruction because of an inability to hydrate macromolecules within the ductal lumens [15]. Until now, more than 1,900 variants have been identified in the Cystic Fibrosis Mutation Database (http://www.genet.sickkids.on.ca/cftr). The human *CFTR* gene spans 250 kb and contains 27 exons that encode for a protein with a total length of 1,480 amino acids [14]. Direct DNA sequencing of polymerase chain reaction (PCR)-amplified *CFTR* gene segments is a first-line method to detect unknown *CFTR* mutations [16], but this is a tedious and labor-intensive endeavor given the large size of the gene.

A new approach that uses massive parallel sequencing called next-generation sequencing (NGS) is becoming standardized, and the cost is rapidly dropping [17]. Using the ultimate platforms, such systems are able to perform billions of sequencing reactions with a read length of 150–250 nucleotides. For most research groups, whole-genome sequencing of many samples remains a costly endeavor, and targeted capture of selected regions of interest followed by sequencing provides a more efficient and cost-effective option. This strategy has allowed identification of causal variants in several Mendelian disorders, variants associated with complex diseases, and recurrently mutated cancer genes [18–20]. The HaloPlex target enrichment technology is a selective circularization-based method that is a further development of the principle of selector probes [21]. In the HaloPlex technology, genomic DNA is fragmented by restriction enzyme digestion and circularized by hybridization to probes whose ends are complementary to the target fragments. Compared to hybrid capture methods, the HaloPlex system requires smaller amounts of starting DNA, has higher specificity (fraction of sequence reads in the region of interest), and provides more uniform genome coverage [22]. Using the bench-top Illumina MiSeq platform, comprehensive analysis of many samples can be easily done. We here report the comprehensive analysis of *CFTR* variants in Japanese patients with CP.

Materials and Methods

Subjects

One hundred and ninety-three patients with CP (121 idiopathic, 46 alcoholic, 17 hereditary, and nine familial) were enrolled in this study. Because we initially aimed to identify novel pancreatitis-associated genes using the HaloPlex system, majority of the patients recruited were nonalcoholic. To extend our findings, we additionally recruited patients with alcoholic CP who had developed CP at relatively younger ages (mean: 44.1 years old). The diagnosis of CP was based on at least two separate episodes of abdominal pain and radiological findings of pancreatic calcifications by computed tomography, endoscopic ultrasonography, and/or morphological findings such as pancreatic ductal irregularities and dilatations revealed by endoscopic retrograde cholangiopancreatography (ERCP) or by magnetic resonance imaging [23]. Hereditary pancreatitis was diagnosed when one first-degree relative or two or more second-degree relatives had recurrent acute pancreatitis or CP without any apparent predisposing factor [24]. Patients with CP in whom the criteria for hereditary pancreatitis were not met but who had at least two affected family members were classified as having familial pancreatitis. Idiopathic CP was diagnosed in the absence of a positive family history or possible predisposing factors such as alcohol abuse, trauma, medication, and anatomical abnormalities. Patients who consumed alcohol over 80 g/day (for men) or 60 g/day (for women) for more than 2 years were classified as alcoholic CP. All subjects were Japanese. This study was performed with the informed consent of the patients in accordance with the principles of the declaration of Helsinki. This study was approved by the Ethics Committee of Tohoku University School of Medicine (article#: 2013-1-498).

Peripheral Blood Collection and DNA Preparation

After written informed consent was obtained, 5–10 mL of peripheral blood was collected in disposable vacuum tubes for genetic testing. Genomic DNA was extracted from peripheral blood leukocytes using the Wizard genomic DNA purification kit (Promega, Madison, WI, USA).

Targeted Next-Generation Sequencing

We used the online design tool SureDesign to generate a customized HaloPlex target enrichment system (Agilent Technologies, Santa Clara, CA, USA) targeting the regions including *CFTR* exons and 50 base pairs (bp) of flanking introns. The expected coverage of the *CFTR* coding region based on the amplicon design was 99.6 %. The HaloPlex target enrichment system relies on a tailored cocktail of restriction enzymes and customized probes to capture genomic regions of interest, which are subsequently amplified by PCR. Sequencing libraries were prepared according to the manufacturer's instructions. Briefly, genomic DNA was digested with restriction enzymes, followed by hybridization to the biotinylated HaloPlex probe library in the presence of the indexing primer cassette. Hybridization results in the circularization of genomic DNA fragments and incorporation of indices and Illumina sequencing motifs. Hybridized probes were captured with streptavidin-coated magnetic beads. Subsequently, libraries were amplified by PCR to produce a sequencing-ready, target-enriched sample.

All libraries of target-enriched DNA were analyzed on a 2200 TapeStation (Agilent Technologies) to verify successful enrichment. All samples were sequenced on the Illumina MiSeq platform (Illumina Japan, Tokyo, Japan) with paired-end 151-bp reads according to the manufacturer's instruction.

Bioinformatic Analysis of Sequencing Data

The reads were trimmed with the utility program Trim Galore (http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/) to remove possible adapter sequences, based on the Illumina TruSeq adapter index sequences. If either read from a pair was shorter than 20 bp after trimming, that pair was removed from the analysis. The remaining quality reads were mapped to the GRCh37 primary assembly of the human genome (http://ensembl.org/) using the Burrows–Wheeler Alignment tool 0.6.1 (http://bio-bwa.sourceforge.net/). Further sequence data processing, assessment of coverage rates, variant calling, and filtration were performed with the Genome Analysis Toolkit, version 1.6 software (Broad Institute, Cambridge, MA, USA; http://www.broadinstitute.org/gatk/). SNPs and insertions/deletions (indels) were annotated using the ANNOVAR (http://www.openbioinformatics.org/annovar/; BIOBASE, Wolfenbüttel, Germany).

Sanger Sequencing

Sanger sequencing was performed to analyze the DNA sequences which included any nucleotide variant identified by NGS. Exons and adjacent intronic regions of the *CFTR* gene containing the nucleotide variants were amplified by PCR using the primer sets (Table 1). The cycle conditions were as follows: preheating at 95 °C for 5 min, followed by 40 cycles of 95 °C for 30 s, 60 °C for 30 s, and 72 °C for 30 s; and then a final extension at 72 °C for 5 min. PCR products were cleaned up using the Illustra ExoProStar S (GE Healthcare Life Sciences; Little Chalfont, UK). The PCR products were sequenced using an ABI Prism BigDye Terminator Cycle Sequencing Kit, version 3.1 on ABI3730xl DNA Analyzer (Applied Biosystems, Foster City, CA, USA) according to the manufacturer's instructions. The results were compared with the reference sequence derived from the GenBank (http://www.ncbi.nlm.nih.gov/genbank, reference sequence NM_000492) to identify nucleotide substitutions. The A of the ATG start codon was used as nucleotide +1. The mutations are described according to the nomenclature recommended by

Table 1 Primers used for direct sequencing

Exon	Forward	Reverse	Size of PCR product (bp)
2	CCAGAAAAGTTGAATAGTATCA	AAGCAATCCTCTCATCTTGG	369
4	AATTCTCAGGGTATTTTATGAG	CCAGCTCACTACCTAATTTATG	363
10	AGCATCTATTGAAAATATCTGACAAAC	AAAGAGACATGGACACCAAATTAAG	315
11	GGAGGCAAGTGAATCCTGAG	AACCGATTGAATATGGAGCC	343
12	CAGATTGAGCATACTAAAAGTG	CATTTACAGCAAATGCTTGCTAG	224
13	TAGATGACCAGGAAATAGAGA	ATGAGGCGGTGAGAAAAGGT	351
15	GGTGGCATGAAACTGTACTG	TGTATACATCCCCAAACTATCT	251
17	TCAGTAAGTACTTTGGCTGC	CCTATTGATGGTGGATCAGC	390
21	TGTGCCCTAGGAGAAGTGTG	TGACAGATACACAGTGACCCTC	335
23	TATGTCACAGAAGTGATCCC	TGAGTACAAGTATCAAATAGC	252
25	GCTTGAGTGTTTTTAACTCTGTGG	AGACCCCCACACGCAGAC	335
27	CTCTGGTCTGACCTGCCTTC	AGCTCCAATTCCATGAGGTG	334

bp base pairs

the Human Genome Variation Society (http://www.hgvs.org/mutnomen).

In addition, all exons and adjacent intronic regions of the *PRSS1*, *SPINK1*, *CTRC*, and *CPA1* genes were amplified by PCR and directly sequenced as previously reported [5, 6, 8, 25].

In Silico Prediction

SIFT (Sorting Intolerant From Tolerant; http://sift.jcvi.org/) and PolyPhen-2 (http://genetics.bwh.harvard.edu/pph2/) were used to predict whether an amino acid substitution would affect the structure and function of a protein. SIFT uses sequence homology, whereas PolyPhen-2 offers predictions based on conservation, protein folding, and crystal structure [26, 27]. The SIFT scores range from zero to one, with zero predicted to be the most deleterious mutation and one the least deleterious. The PolyPhen-2 scoring predicts three outcomes for mutations: "benign" (most likely lacking any phenotypic effect), "possibly damaging" (may affect protein structure or function), and "probably damaging" (high degree of confidence that protein structure function will be affected).

Statistical Analysis

The variant frequencies in the Japanese population were obtained from the Human Genetic Variation Database (www.genome.med.kyoto-u.ac.jp/SnpDB/). The significance of the differences in variant frequencies between patients and controls was tested by two-tailed Fisher's exact test. A P value <0.05 was considered significant. All statistical analyses were performed using the SPSS version 17.0 statistical analysis software (SPSS Inc., Chicago, IL, USA).

Results

Approximately 10 kb of the coding regions and the adjacent noncoding regions of the *CFTR* gene were analyzed in this study. On average, 98.8, 97.0, and 95.1 % of the coding regions of the *CFTR* gene were covered by at least one, five, and 10 sequence reads, respectively. The sequencing data covered 91.6 % of the coding regions of the *CFTR* gene by ≥20 reads with a mean read depth of 449 and a median depth of 412 (Fig. 1). These results indicate a high-resolution capability for the identification of variants, such as mutations.

In our cohort of 193 CP patients, we identified 12 non-synonymous and seven synonymous variants in the exons of the *CFTR* gene by targeted NGS (Tables 2, 3, 4, 5). The presence of these variants was confirmed by Sanger sequencing. Based on the presence in dbSNP137, Exome Variant Server (NHLBI GO Exome Sequencing Project, Seattle, WA, USA; URL: http://evs.gs.washington.edu/EVS/), and the Human Genetic Variation Database, three non-synonymous variants [c.1231A>G (p.K411E), c.1753G>T (p.E585X) and c.2869delC (p.L957fs)] and three synonymous variants (c.372C>T, c.3975A>G and c.4254G>A) were novel. The frameshift variant c.2869delC (p.L957fs) leads to a stop codon afterward at amino acid 967, to premature termination of translation and a heavily truncated protein missing more than one-third of its amino acids. This variant was found in a 22-year-old female with idiopathic CP. She was admitted due to a pancreatitis attack and diagnosed as having CP. She had suffered from back and abdominal pain since 20 years old. The value of the *n*-benzoyl-*l*-tyrosyl-*p*-aminobenzoic acid test was 52 % (normal: >70 %), suggesting pancreatic exocrine dysfunction. The patient also had the c.4056G>C (p.Q1352H) variant in a

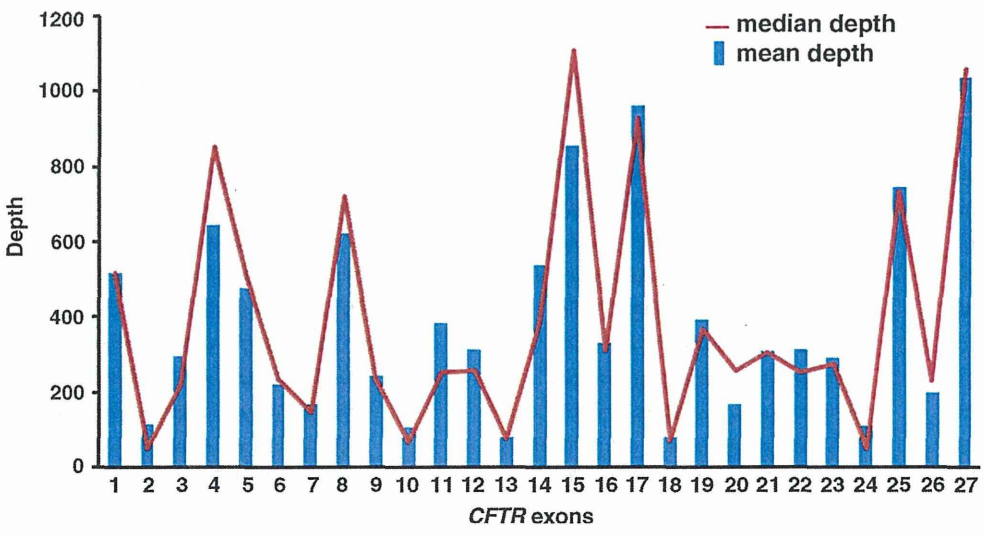


Fig. 1 Graph of the mean depth, median depth, and sequencing coverage for all the 27 exons in the *CFTR* gene. By MiSeq NGS, a high-quality sequence was obtained for 27 exons and the flanking sequences from the *CFTR* gene, including a mean depth of ×516 and a median depth of ×442. On average, 90.3 % of the coding region was successfully covered by ≥20 reads