Table 3. Genes for which DNA methylation levels were hallmarks for Clusters I, II and III in the learning cohort (Continued)

(C) Hallmark genes for Cluster III

Target ID [15]	Chrom[16]	Position[17]	Gene symbol	DNA methylation level in non-cancerous lung tissue (N) samples[18] (mean ± SD)			*p*-Value of Welch's *t*-test (III *vs.* I and II)[19]	Δβ (III-I and II) [20]	*p*-Value of Jonckheere–Terpstra trend test in III[21]
				Cluster I	Cluster II	Cluster III			
cg21063899	13	78,109,801	*SCEL*	0.033 ± 0.088	0.013 ± 0.054	−0.081 ± 0.086	3.06×10^{-12}	−0.103	1.47×10^{-9} (Hypo)
cg14074641	16	48,181,753	*ABCC12*	−0.002 ± 0.091	0.025 ± 0.054	−0.109 ± 0.105	1.40×10^{-12}	−0.121	2.44×10^{-1} (Hypo)

[1]Probe ID for the Infinium HumanMethylation27 Bead Array.
[2]Chromosome.
[3]National Center for Biotechnology Information (NCBI) Database (Genome Build 37).
[4]$\Delta\beta_{N\text{-}averageC}$.
[5]Average β_{N-C} in Cluster I *versus* average β_{N-C} in Clusters II and III. Such **p** values were calculated to reveal the hallmark genes of Cluster I that showed DNA methylation statuses significantly different in their N samples in comparison with N samples from other clusters (Clusters II and III).
[6]Average β_{N-C} in Cluster I minus average β_{N-C} in Clusters II and III. If Δβ (I-II and III) was more than 0.1, N samples in Cluster I were considered to show DNA hypermethylation relative to N samples in other clusters, and if Δβ (I-II and III) was less than −0.1, N samples in Cluster I were considered to show DNA hypomethylation relative to N samples in other clusters.
[7]Stepwise DNA hypermethylation (Hyper) and hypomethylation (Hypo) from normal lung tissue samples to N samples, and then to tumorous tissue samples in Cluster I.
[8]Probe ID for the Infinium HumanMethylation27 Bead Array.
[9]Chromosome.
[10]National Center for Biotechnology Information (NCBI) Database (Genome Build 37).
[11]$\Delta\beta_{N\text{-}averageC}$.
[12]Average β_{N-C} in Cluster II *versus* average β_{N-C} in Clusters I and III. Such *p* value was calculated to reveal the hallmark gene of Cluster II that showed DNA methylation status significantly different in their N samples in comparison with N samples from other clusters (Clusters I and III).
[13]Average β_{N-C} in Cluster II minus average β_{N-C} in Clusters I and III. If Δβ (II−I and III) was more than 0.1, N samples in Cluster II were considered to show DNA hypermethylation relative to N samples in other clusters.
[14]Stepwise DNA hypermethylation (Hyper) and hypomethylation (Hypo) from normal lung tissue samples to N samples, and then to tumorous tissue samples in Cluster II.
[15]Probe ID for the Infinium HumanMethylation27 Bead Array.
[16]Chromosome.
[17]National Center for Biotechnology Information (NCBI) Database (Genome Build 37).
[18]$\Delta\beta_{N\text{-}averageC}$.
[19]Average β_{N-C} in Cluster III *versus* average β_{N-C} in Clusters I and II. Such *p* values were calculated to reveal the hallmark gene of Cluster III that showed DNA methylation statuses significantly different in their N samples in comparison with N samples from other clusters (Clusters I and II).
[20]Average β_{N-C} in Cluster III minus average β_{N-C} in Clusters I and II. If Δβ (III-I and II) was more than 0.1, N samples in Cluster III were considered to show DNA hypermethylation relative to N samples in other clusters and if Δβ (III-I and II) was less than −0.1, N samples in Cluster III were considered to show DNA hypomethylation relative to N samples in other clusters.
[21]Stepwise DNA hypermethylation (Hyper) and hypomethylation (Hypo) from normal lung tissue samples to N samples, and then to tumorous tissue samples in Cluster III.

Table 4. Correlation between DNA methylation levels of hallmark genes for Clusters I, II and III and the clinicopathological parameters in the validation cohort

(A) Hallmark genes for Cluster I

		DNA methylation level in non-cancerous lung tissue (N) samples[2] (mean ± SD)											
		Pleural anthracosis			Emphysematic change			Atypical adenomatous hyperplasia			Tumor anthracosis		
Target ID [1]	Gene symbol	G1	G2-3	*p*-Value[3]	Negative	Positive	*p*-Value[3]	Absence	Presence	*p*-Value[3]	Negative	Positive	*p*-Value[3]
cg20249919	*PCSK6*	−0.126 ± 0.049	−0.049 ± 0.102	1.83×10^{-2}	−0.069 ± 0.067	−0.049 ± 0.131	4.99×10^{-1}	−0.056 ± 0.101	−0.096 ± 0.082	3.57×10^{-1}	−0.077 ± 0.085	−0.047 ± 0.105	3.18×10^{-1}
cg23349790	*IGSF21*	−0.044 ± 0.101	−0.002 ± 0.100	4.13×10^{-1}	−0.035 ± 0.072	0.028 ± 0.118	3.50×10^{-2}	−0.005 ± 0.101	−0.029 ± 0.086	5.79×10^{-1}	−0.054 ± 0.079	0.017 ± 0.098	1.68×10^{-2}
cg22285621	*SSH3*	−0.073 ± 0.043	0.002 ± 0.101	1.25×10^{-2}	−0.001 ± 0.077	−0.014 ± 0.121	6.52×10^{-1}	−0.004 ± 0.101	−0.035 ± 0.059	3.41×10^{-1}	−0.015 ± 0.077	0.001 ± 0.106	5.69×10^{-1}
cg15433631	*IRX2*	−0.041 ± 0.061	0.034 ± 0.074	4.73×10^{-2}	0.025 ± 0.068	0.029 ± 0.083	8.52×10^{-1}	0.026 ± 0.077	0.028 ± 0.056	9.46×10^{-1}	0.010 ± 0.074	0.037 ± 0.072	2.77×10^{-1}
cg21949305	ADORA2A, CYTSA	0.025 ± 0.091	0.026 ± 0.060	9.73×10^{-1}	0.015 ± 0.054	0.039 ± 0.069	1.91×10^{-1}	0.029 ± 0.063	−0.003 ± 0.036	1.28×10^{-1}	−0.004 ± 0.058	0.039 ± 0.061	3.48×10^{-2}
cg10942056	*DISP1*	0.014 ± 0.088	0.015 ± 0.068	9.71×10^{-1}	0.009 ± 0.062	0.023 ± 0.077	5.00×10^{-1}	0.019 ± 0.071	−0.022 ± 0.027	2.48×10^{-2}	−0.007 ± 0.056	0.026 ± 0.072	1.08×10^{-1}
cg15149645	NUPR1	−0.007 ± 0.124	0.013 ± 0.073	7.37×10^{-1}	0.006 ± 0.070	0.015 ± 0.085	7.06×10^{-1}	0.015 ± 0.079	−0.036 ± 0.022	2.81×10^{-3}	−0.015 ± 0.081	0.021 ± 0.077	1.77×10^{-1}
cg06954481	*GBX2*	−0.044 ± 0.031	0.013 ± 0.075	9.57×10^{-3}	0.008 ± 0.062	0.003 ± 0.085	7.95×10^{-1}	0.012 ± 0.073	−0.047 ± 0.040	2.27×10^{-2}	−0.012 ± 0.058	0.016 ± 0.078	1.90×10^{-1}
cg21250978	*PRKAR2B*	−0.013 ± 0.092	0.002 ± 0.058	7.44×10^{-1}	−0.010 ± 0.050	0.014 ± 0.070	1.95×10^{-1}	0.005 ± 0.061	−0.037 ± 0.033	4.81×10^{-2}	−0.032 ± 0.058	0.013 ± 0.059	2.63×10^{-2}
cg22418909	SFRP1	−0.043 ± 0.076	0.002 ± 0.058	2.55×10^{-1}	0.000 ± 0.065	−0.003 ± 0.053	8.32×10^{-1}	0.003 ± 0.061	−0.041 ± 0.022	4.86×10^{-3}	−0.020 ± 0.057	0.007 ± 0.060	1.64×10^{-1}
cg26200585	*PRX*	0.020 ± 0.079	0.015 ± 0.066	8.81×10^{-1}	0.013 ± 0.069	0.016 ± 0.063	8.89×10^{-1}	0.019 ± 0.067	−0.029 ± 0.035	3.48×10^{-2}	−0.006 ± 0.054	0.025 ± 0.070	1.19×10^{-1}
cg24396745	*HCN4*	−0.056 ± 0.035	0.020 ± 0.072	3.53×10^{-3}	0.017 ± 0.066	0.003 ± 0.079	4.87×10^{-1}	0.015 ± 0.073	−0.025 ± 0.054	1.86×10^{-1}	−0.005 ± 0.074	0.022 ± 0.071	2.60×10^{-1}
cg04330449	*NEUROG1*	−0.040 ± 0.033	0.010 ± 0.073	2.33×10^{-2}	0.010 ± 0.065	−0.005 ± 0.078	4.52×10^{-1}	0.006 ± 0.072	−0.020 ± 0.056	3.82×10^{-1}	−0.006 ± 0.044	0.012 ± 0.078	3.40×10^{-1}
cg19589427	*TNFSF18*	−0.008 ± 0.078	−0.010 ± 0.070	9.65×10^{-1}	−0.010 ± 0.078	−0.012 ± 0.057	9.26×10^{-1}	−0.007 ± 0.071	−0.042 ± 0.036	1.10×10^{-1}	−0.040 ± 0.057	0.005 ± 0.069	3.13×10^{-2}
cg16731240	*ZNF577*	−0.042 ± 0.037	0.014 ± 0.087	2.50×10^{-2}	0.012 ± 0.1000	0.003 ± 0.060	7.02×10^{-1}	0.007 ± 0.086	0.014 ± 0.078	8.48×10^{-1}	−0.015 ± 0.074	0.020 ± 0.087	1.79×10^{-1}
cg03544320	*CRMP1*	−0.097 ± 0.083	0.018 ± 0.094	3.14×10^{-2}	0.010 ± 0.109	−0.005 ± 0.084	5.67×10^{-1}	0.008 ± 0.102	−0.040 ± 0.036	4.56×10^{-2}	−0.014 ± 0.095	0.019 ± 0.097	3.04×10^{-1}
cg12864235	*CDH9*	0.032 ± 0.054	0.027 ± 0.057	8.51×10^{-1}	0.025 ± 0.045	0.030 ± 0.066	7.86×10^{-1}	0.031 ± 0.057	−0.004 ± 0.019	1.29×10^{-2}	0.012 ± 0.043	0.034 ± 0.060	1.77×10^{-1}
cg15898840	IGFBP3	−0.043 ± 0.030	0.001 ± 0.057	2.53×10^{-2}	−0.004 ± 0.052	−0.007 ± 0.061	8.16×10^{-1}	−0.002 ± 0.056	−0.036 ± 0.037	1.14×10^{-1}	−0.003 ± 0.044	−0.002 ± 0.060	9.36×10^{-1}
cg08044694	*BRD4*	−0.067 ± 0.036	−0.020 ± 0.049	3.84×10^{-2}	−0.021 ± 0.046	−0.028 ± 0.053	6.30×10^{-1}	−0.022 ± 0.050	−0.039 ± 0.035	3.55×10^{-1}	−0.041 ± 0.041	−0.017 ± 0.052	1.12×10^{-1}
cg03734874	*TMEM179*	−0.032 ± 0.034	0.023 ± 0.074	1.62×10^{-2}	0.021 ± 0.076	0.015 ± 0.068	7.56×10^{-1}	0.024 ± 0.072	−0.037 ± 0.039	1.89×10^{-2}	−0.008 ± 0.052	0.030 ± 0.075	5.63×10^{-2}
cg10599444	*MMP14*	−0.063 ± 0.026	−0.010 ± 0.060	4.85×10^{-3}	−0.008 ± 0.057	−0.023 ± 0.060	3.87×10^{-1}	−0.013 ± 0.060	−0.027 ± 0.041	5.25×10^{-1}	−0.026 ± 0.047	−0.007 ± 0.059	2.66×10^{-1}
cg24133115	*PDE10A*	−0.020 ± 0.025	0.022 ± 0.060	1.58×10^{-2}	0.014 ± 0.059	0.020 ± 0.057	7.02×10^{-1}	0.018 ± 0.060	0.004 ± 0.027	3.88×10^{-1}	0.009 ± 0.037	0.024 ± 0.063	3.25×10^{-1}
cg12594641	LYPD6	−0.024 ± 0.041	0.036 ± 0.068	2.45×10^{-2}	0.021 ± 0.062	0.036 ± 0.075	4.61×10^{-1}	0.029 ± 0.071	0.013 ± 0.038	4.43×10^{-1}	0.025 ± 0.045	0.035 ± 0.075	5.94×10^{-1}
cg05724065	PHKG1	0.016 ± 0.100	0.011 ± 0.057	9.10×10^{-1}	0.014 ± 0.055	0.008 ± 0.067	7.39×10^{-1}	0.016 ± 0.061	−0.032 ± 0.027	1.05×10^{-2}	−0.002 ± 0.066	0.019 ± 0.059	3.33×10^{-1}
cg19466563	SPARCL1	0.018 ± 0.081	0.012 ± 0.055	8.86×10^{-1}	0.007 ± 0.052	0.021 ± 0.060	3.98×10^{-1}	0.019 ± 0.056	−0.035 ± 0.015	4.46×10^{-5}	−0.015 ± 0.047	0.025 ± 0.057	2.24×10^{-2}
cg24433189	SSTR5	0.024 ± 0.075	0.034 ± 0.060	7.96×10^{-1}	0.031 ± 0.052	0.032 ± 0.071	9.63×10^{-1}	0.035 ± 0.062	−0.003 ± 0.027	2.95×10^{-2}	0.026 ± 0.051	0.036 ± 0.065	5.79×10^{-1}
cg24453664	CD59	−0.057 ± 0.039	0.000 ± 0.053	2.44×10^{-2}	−0.002 ± 0.054	−0.009 ± 0.055	6.40×10^{-1}	−0.004 ± 0.054	−0.021 ± 0.050	5.00×10^{-1}	−0.018 ± 0.050	0.001 ± 0.055	2.53×10^{-1}
cg26609631	GSX1	−0.051 ± 0.058	0.006 ± 0.067	9.21×10^{-2}	0.002 ± 0.066	−0.005 ± 0.069	7.29×10^{-1}	0.004 ± 0.068	−0.050 ± 0.025	3.65×10^{-3}	−0.020 ± 0.054	0.010 ± 0.071	1.35×10^{-1}
cg10604646	RGS5	0.038 ± 0.039	0.033 ± 0.058	8.08×10^{-1}	0.013 ± 0.062	0.056 ± 0.041	5.70×10^{-3}	0.037 ± 0.056	−0.014 ± 0.057	1.16×10^{-1}	−0.006 ± 0.063	0.049 ± 0.047	1.07×10^{-2}
cg03355526	ZNF454	−0.061 ± 0.044	0.001 ± 0.075	2.98×10^{-2}	0.003 ± 0.077	−0.012 ± 0.070	4.81×10^{-1}	0.002 ± 0.074	−0.052 ± 0.053	8.77×10^{-2}	−0.007 ± 0.058	0.000 ± 0.078	7.53×10^{-1}
cg27096144	MSX2	−0.066 ± 0.049	0.001 ± 0.063	3.29×10^{-2}	−0.006 ± 0.057	−0.006 ± 0.072	9.95×10^{-1}	−0.003 ± 0.065	−0.026 ± 0.040	2.94×10^{-1}	−0.023 ± 0.055	0.002 ± 0.068	2.05×10^{-1}
cg15520279	HOXD8	−0.021 ± 0.040	0.016 ± 0.070	1.15×10^{-1}	0.015 ± 0.075	0.009 ± 0.056	7.16×10^{-1}	0.014 ± 0.070	−0.006 ± 0.025	2.05×10^{-1}	−0.010 ± 0.038	0.024 ± 0.075	4.76×10^{-2}

Table 4. Correlation between DNA methylation levels of hallmark genes for Clusters I, II and III and the clinicopathological parameters in the validation cohort (Continued)

(A) Hallmark genes for Cluster I

		DNA methylation level in non-cancerous lung tissue (N) samples[2] (mean ± SD)											
		Pleural anthracosis			Emphysematic change			Atypical adenomatous hyperplasia			Tumor anthracosis		
Target ID [1]	Gene symbol	G1	G2-3	*p*-Value[3]	Negative	Positive	*p*-Value[3]	Absence	Presence	*p*-Value[3]	Negative	Positive	*p*-Value[3]
cg11733245	IL2RA	−0.037 ± 0.042	−0.035 ± 0.051	9.29×10^{-1}	−0.023 ± 0.038	−0.047 ± 0.060	1.10×10^{-1}	−0.035 ± 0.051	−0.013 ± 0.023	1.09×10^{-1}	−0.009 ± 0.035	−0.047 ± 0.050	$\underline{7.07\times10^{-3}}$
cg22325572	CD53	−0.024 ± 0.083	−0.028 ± 0.058	9.23×10^{-1}	−0.017 ± 0.051	−0.040 ± 0.067	1.90×10^{-1}	−0.029 ± 0.060	−0.008 ± 0.057	4.68×10^{-1}	0.003 ± 0.057	−0.041 ± 0.057	$\underline{2.96\times10^{-2}}$
cg15691199	CEBPE	−0.029 ± 0.080	−0.018 ± 0.064	7.94×10^{-1}	−0.016 ± 0.050	−0.022 ± 0.079	7.79×10^{-1}	−0.022 ± 0.066	0.015 ± 0.021	$\underline{1.52\times10^{-2}}$	−0.002 ± 0.061	−0.028 ± 0.067	2.28×10^{-1}
cg16927606	U2AF1L4	−0.012 ± 0.079	−0.004 ± 0.054	8.25×10^{-1}	−0.005 ± 0.048	−0.006 ± 0.065	9.54×10^{-1}	−0.010 ± 0.057	0.032 ± 0.025	$\underline{1.48\times10^{-2}}$	0.009 ± 0.051	−0.012 ± 0.057	2.31×10^{-1}
cg16240480	EDARADD	−0.040 ± 0.103	−0.047 ± 0.072	8.91×10^{-1}	−0.032 ± 0.071	−0.062 ± 0.075	1.61×10^{-1}	−0.048 ± 0.074	−0.017 ± 0.070	3.86×10^{-1}	0.001 ± 0.049	−0.066 ± 0.076	$\underline{1.24\times10^{-3}}$
cg05596756	FAM113B	−0.010 ± 0.085	−0.010 ± 0.062	9.92×10^{-1}	0.000 ± 0.052	−0.022 ± 0.073	2.22×10^{-1}	−0.014 ± 0.064	0.026 ± 0.025	$\underline{2.14\times10^{-2}}$	0.011 ± 0.057	−0.020 ± 0.065	1.30×10^{-1}
cg08040471	C17orf62	0.000 ± 0.073	−0.014 ± 0.066	7.11×10^{-1}	−0.002 ± 0.051	−0.024 ± 0.080	2.80×10^{-1}	−0.015 ± 0.068	0.013 ± 0.029	1.21×10^{-1}	0.019 ± 0.053	−0.027 ± 0.066	$\underline{2.03\times10^{-2}}$
cg20622019	ADA	−0.021 ± 0.073	−0.033 ± 0.068	7.39×10^{-1}	−0.028 ± 0.069	−0.035 ± 0.066	7.31×10^{-1}	−0.036 ± 0.069	0.008 ± 0.021	$\underline{6.34\times10^{-3}}$	0.002 ± 0.051	−0.047 ± 0.068	$\underline{1.14\times10^{-2}}$
cg05109049	EVI2B	−0.057 ± 0.106	−0.041 ± 0.090	7.56×10^{-1}	−0.023 ± 0.093	−0.062 ± 0.084	1.32×10^{-1}	−0.044 ± 0.090	−0.009 ± 0.102	5.07×10^{-1}	−0.002 ± 0.054	−0.058 ± 0.099	$\underline{1.72\times10^{-2}}$
cg07973967	CD79B	−0.034 ± 0.105	−0.028 ± 0.075	9.16×10^{-1}	−0.017 ± 0.060	−0.042 ± 0.094	2.83×10^{-1}	−0.031 ± 0.080	0.007 ± 0.021	$\underline{2.06\times10^{-2}}$	−0.007 ± 0.067	−0.039 ± 0.081	1.69×10^{-1}

(B) Hallmark genes for Cluster II and III

		DNA methylation level in non-cancerous lung tissue (N) samples[5] (mean ± SD)										
		Lymphatic invasion			Nodal status				Pathological Tumor-Node-Metastasis stage			
Target ID [4]	Gene symbol	Negative	Positive	*p*-Value[6]	N0	N1	N2-3	*p*-Value[6]	IA-IB	IIA-IIB	IIIA-IV	*p*-Value[6]
cg26606064	*EI24*	0.018 ± 0.098	0.098 ± 0.113	$\underline{3.01\times10^{-2}}$	0.004 ± 0.098	0.010 ± 0.094	0.118 ± 0.091	$\underline{1.27\times10^{-3}}$	0.009 ± 0.102	−0.008 ± 0.082	0.118 ± 0.091	$\underline{1.14\times10^{-3}}$
cg17872476	*VTI1A*	−0.035 ± 0.134	−0.126 ± 0.137	$\underline{4.32\times10^{-2}}$	−0.023 ± 0.132	0.022 ± 0.127	−0.149 ± 0.121	$\underline{5.69\times10^{-3}}$	−0.020 ± 0.116	−0.016 ± 0.166	−0.149 ± 0.121	$\underline{6.70\times10^{-3}}$
cg21063899	*SCEL*	−0.043 ± 0.093	−0.109 ± 0.110	6.08×10^{-2}	−0.024 ± 0.082	−0.024 ± 0.090	−0.141 ± 0.096	$\underline{2.81\times10^{-4}}$	−0.021 ± 0.089	−0.033 ± 0.062	−0.141 ± 0.096	$\underline{2.60\times10^{-4}}$
cg14074641	*ABCC12*	−0.017 ± 0.118	−0.083 ± 0.117	8.60×10^{-2}	0.005 ± 0.104	−0.006 ± 0.112	−0.118 ± 0.114	$\underline{2.09\times10^{-3}}$	0.000 ± 0.111	0.014 ± 0.084	−0.118 ± 0.114	$\underline{2.00\times10^{-3}}$

[1]Probe ID for the Infinium HumanMethylation27 Bead Array.
[2]$\Delta\beta_{N\text{-}averageC}$.
[3]*p* values (Welch's *t*-test) of <0.05 are underlined.
[4]Probe ID for the Infinium HumanMethylation27 Bead Array.
[5]$\Delta\beta_{N\text{-}averageC}$.
[6]*p* values (Welch's *t*-test) of <0.05 are underlined.

strengthened in, T samples (Table 1). These findings are compatible with the "field cancerization" concept in the lung.[29] In our previous study using the Infinium assay, we proved that DNA methylation alterations in N samples resulted in silencing of tumor-related genes in tumorous tissue.[15] However, the correlation between the results of the Infinium assay in N samples and carcinogenetic factors was not examined in detail.

In this epigenetic clustering of patients with LADCs based on DNA methylation profiles in N samples, many of the patients belonging to Cluster I were heavy smokers. In fact, pleural anthracosis, which mainly reflects the long-term cumulative effects of cigarette smoking, was marked in the lungs of patients belonging to Cluster I. Smoking is known to be a cause of COPD. In fact, many patients in Cluster I actually suffered from obstructive ventilation impairment, and histological findings compatible with emphysema and lung fibrosis were observed in their N samples. Moreover, recurrent inflammation is generally associated with COPD,[30] and histological findings compatible with respiratory bronchiolitis[20,21] were actually observed in the lungs of patients belonging to Cluster I. Inflammation is known to be one of the major causes of DNA methylation alterations in precancerous conditions in various organs, such as chronic hepatitis[16,17] and chronic pancreatitis.[31,32] Taken together, the data suggest that the DNA methylation profiles characterizing Cluster I may be established in lung tissue through the long-term cumulative effects of cigarette smoking *via* chronic inflammation under the conditions of COPD. Unlike the previous study, which revealed aberrant DNA methylation of several tumor-related genes in lung cancers themselves of patients with COPD,[33] this study demonstrated for the first time the presence of distinct DNA methylation profiles related to COPD in N samples, based on genome-wide analysis.

The majority of patients belonging to Cluster II were non-smokers, especially young females. DNA methylation profiles characterizing Cluster II may reflect the carcinogenetic pathway that is unrelated to cigarette smoking. Mutation of the *EGFR* gene is well known to be a driver of LADCs in young female non-smokers, especially in Asia.[34] However, Cluster II included LADCs without *EGFR* gene mutations in non-smokers (data not shown), indicating that DNA methylation profiles in Cluster II were not entirely induced by *EGFR* mutation.

Although many of the patients belonging to Cluster III were smokers, the average number of cigarettes smoked per day × year index was lower in Cluster III than in Cluster I. In fact severe pleural anthracosis was not so frequently evident in the lungs of patients belonging to Cluster III. In addition, the incidence of emphysematous change and fibrosis was lower in the adjacent lung tissue of patients in Cluster III than in that of patients in Cluster I. DNA methylation profiles in Cluster III did not develop from a background of chronic inflammation in COPD, but may have developed rapidly before the long-term effects of cigarette smoking had accumulated in the adjacent lung tissue (possibly through more direct effects of carcinogens related or unrelated to cigarette smoking). However, to evaluate more precisely the effects of smoking on DNA methylation profiles, detailed DNA methylation analysis should be performed using purified epithelial cells, such as those from the airway epithelium.

Distinct DNA methylation profiles seem to be established in the non-cancerous lung during the carcinogenetic pathway *via* inflammation in COPD in heavy smokers (Fig. 2*a*), the carcinogenetic pathway unrelated to cigarette smoking (Fig. 2*b*), and the carcinogenetic pathway that occurs not *via* COPD but possibly *via* more direct effects of carcinogens (Fig. 2*c*). Each pathway may have distinct target genes as hallmarks for Clusters I, II and III (Table 3 and Supporting Information Table S3). Among 120 hallmark genes for Clusters I, II and III, 119 (one exception, *ABCC12*, being shared between Clusters II and III) showed ordered differences of DNA methylation from C to N, and then to T samples of the relevant cluster ($p < 0.05$, Jonckheere–Terpstra trend test, Table 3 and Supporting Information Table S3), indicating that a distinct DNA methylation profile in N samples of each cluster is inherited during progression to Ts.

A proportion of genes described in Table 3 and Supporting Information Table S3 may be simple hallmarks of each cluster (simple target genes of each carcinogenetic pathway). However, at least a proportion of DNA methylation alterations occurring during each carcinogenetic pathway actually result in altered expression of target genes, and may participate in establishment of the clinicopathological characteristics of LADCs in each cluster. The DNA methylation profiles in Cluster I may participate in the generation of locally invasive LADCs, which have a large diameter, a progressed T stage, a high histological grade, frequent pleural invasion and tumor anthracosis. DNA methylation profiles in Cluster II may participate in the generation of clinicopathologically less aggressive LADCs with a favorable outcome. DNA methylation profiles in Cluster III may participate in the generation of the most aggressive LADCs showing frequent lymphatic vessel invasion, blood vessel invasion, a high N stage, a high TNM stage and a poor outcome.

Table 3 includes homeobox genes, such as *IRX2* and *HOXD8*, a gene that has been implicated in cell migration, *SPARCL1*, and genes that have been implicated in apoptosis, such as *RGS5* and *EI24*. *IRX2* is known to participate in early lung development in mouse embryos.[35] *HOXD8* is known to be methylated and/or down-regulated in human malignancies, especially in metastatic, rather than in primary lesions.[36,37] *SPARCL1* is an extracellular matrix glycoprotein known to be correlated with cancer invasion.[38,39] *RGS5* is a member of the family of molecules regulating G protein signaling, and stimulates hypoxia-inducible apoptosis.[40] Positive correlations between *RGS5* expression and both tumor differentiation and a favorable outcome have been reported.[41,42] *EI24* is induced by *p53*, suppresses cell growth and induces apoptosis.[43] Reduced expression associated with DNA methylation of *IRX2, HOXD8, SPARCL1, RGS5* and *EI24* in our cohort of LADCs has been confirmed using expression microarray (data not shown). It is

feasible that these target genes of each carcinogenetic pathway participate in determining the clinicopathological characteristics of LADCs in each cluster.

In the validation cohort, the DNA methylation status of hallmark genes identified in N samples of Cluster I was significantly correlated with pleural anthracosis, which reflects the long-term cumulative effects of smoking, and COPD (pulmonary emphysema) in the adjacent lung and tumor anthracosis, which reflect active cancer–stromal interaction in LADCs. The DNA methylation status of the hallmark gene identified in N samples of Cluster II was significantly correlated with lower aggressiveness (low N stage and low TNM stage) of LADCs in the validation cohort. The DNA methylation status of hallmark genes identified in N samples of Cluster III was significantly correlated with aggressiveness of LADCs, such as lymph vessel invasion, a high N stage and a high TNM stage, in the validation cohort. Thus, correlations between distinct DNA methylation profiles in N samples and both carcinogenetic background factors in the adjacent lung tissue and clinicopathological characteristics of LADCs were confirmed in the validation cohort (Table 4).

Receiver operating characteristic curve (ROC) analysis was performed for N samples in the learning cohort, and the thresholds of the representative hallmark genes described in Table 4 were set so that they were nearest to the top left corner of the ROC. Using these thresholds, the sensitivity, specificity and accuracy for prediction of lymphatic vessel involvement, lymph node metastasis, TNM stage and patient outcome (recurrence and death) were calculated in both the learning and validation cohorts (Supporting Information Table S4). Even though Supporting Information Table S4 suggests that the aggressiveness of tumors developing in the same individual patients and patient outcome may be predictable on the basis of DNA methylation status in N samples, further examinations will be needed to set strict criteria for maximal sensitivity, specificity and accuracy.

Although bulk tissue comprising several cell lineages, for a large number of C, N and T samples, was examined in this study, it would be preferable to examine the DNA methylation status of purified cells. Therefore, the DNA methylation status of the representative gene *CASP8* (Infinium probe ID: cg26799474), included in Table 1B, was compared between cancer cells and normal peripheral airway epithelial cells obtained by microdissection from formalin-fixed, paraffin-embedded tissues of representative patients with LADCs and patients without primary lung cancers, respectively, using pyrosequencing. The DNA methylation levels in T samples (0.279 ± 0.184) were significantly lower than those in C samples (0.689 ± 0.042) by Infinium assay ($p = 3.64 \times 10^{-4}$). Such a significant difference was reproduced upon comparison with microdissected normal airway epithelium: pyrosequencing showed that the DNA methylation levels in microdissected cancer cells (0.273 ± 0.313) were significantly lower than those in microdissected normal airway epithelial cells (0.765 ± 0.104) ($p = 2.74 \times 10^{-3}$).

Differences in DNA methylation levels among different cell lineages, such as epithelial and stromal components, are also an important issue. Cancer cells and their stromal cells, such as cancer-associated fibroblasts, were again collected separately by microdissection from formalin-fixed, paraffin-embedded tissues from representative patients with LADCs. The DNA methylation levels of representative genes described in Table 1B were evaluated quantitatively by pyrosequencing. In one of the examined genes (*CASP8* [Infinium probe ID: cg26799474]), the DNA methylation statuses of cancer cells (0.273 ± 0.313) and stromal cells (0.219 ± 0.094) were almost equal, indicating that both may be affected by carcinogenetic factors. For the other examined gene (*LHX1* [Infinium probe ID: cg22660578]), the DNA methylation statuses of cancer cells (0.096 ± 0.141) and stromal cells (0.538 ± 0.486) differed from each other, probably reflecting differences in susceptibility to the effects of carcinogens, or differences in cell lineage.

In summary, DNA methylation profiles reflecting carcinogenetic background factors, such as smoking, inflammation and COPD, appear to be established in adjacent lung tissue in patients with LADCs. Such DNA methylation profiles in adjacent lung tissue may play a role in determining the aggressiveness of tumors developing in the same individual patients, and thus patient outcome.

Acknowledgement

T. Sato is an awardee of a research resident fellowship from the Foundation for Promotion of Cancer Research in Japan.

References

1. Siegel R, Naishadham D, Jemal A. Cancer statistics, 2013. *CA Cancer J Clin* 2013;63:11–30.
2. Sun S, Schiller JH, Gazdar AF. Lung cancer in never smokers—a different disease. *Nat Rev Cancer* 2007;7:778–90.
3. Baylin SB, Jones PA. A decade of exploring the cancer epigenome—biological and translational implications. *Nat Rev Cancer* 2011;11:726–34.
4. Heller G, Zielinski CC, Zochbauer-Muller S. Lung cancer: from single-gene methylation to methylome profiling. *Cancer Metastasis Rev* 2010;29: 95–107.
5. Arai E, Kanai Y. DNA methylation profiles in precancerous tissue and cancers: carcinogenetic risk estimation and prognostication based on DNA methylation status. *Epigenomics* 2010;2: 467–81.
6. Kanai Y. Genome-wide DNA methylation profiles in precancerous conditions and cancers. *Cancer Sci* 2010;101:36–45.
7. Arai E, Chiku S, Mori T, et al. Single-CpG-resolution methylome analysis identifies clinicopathologically aggressive CpG island methylator phenotype clear cell renal cell carcinomas. *Carcinogenesis* 2012;33:1487–93.
8. Zöchbauer-Müller S, Lam S, Toyooka S, et al. Aberrant methylation of multiple genes in the upper aerodigestive tract epithelium of heavy smokers. *Int J Cancer* 2003;107: 612–6.
9. Eguchi K, Kanai Y, Kobayashi K, et al. DNA hypermethylation at the D17S5 locus in non-small cell lung cancers: its association with smoking history. *Cancer Res* 1997;57: 4913–5.
10. Lamy A, Sesboüé R, Bourguignon J, et al. Aberrant methylation of the CDKN2a/p16INK4a gene promoter region in preinvasive bronchial lesions: a prospective study in high-risk patients without invasive cancer. *Int J Cancer* 2002;100: 189–93.
11. Belinsky SA, Palmisano WA, Gilliland FD, et al. Aberrant promoter methylation in bronchial epithelium and sputum from current and former smokers. *Cancer Res* 2002;62:2370–7.

12. Bibikova M, Le J, Barnes B, et al. Genome-wide DNA methylation profiling using Infinium(R) assay. *Epigenomics* 2009;1:177–200.
13. Selamat SA, Chung BS, Girard L, et al. Genome-scale analysis of DNA methylation in lung adenocarcinoma and integration with mRNA expression. *Genome Res* 2012;22:1197–211.
14. Lockwood WW, Wilson IM, Coe BP, et al. Divergent genomic and epigenomic landscapes of lung cancer subtypes underscore the selection of different oncogenic pathways during tumor development. *PLoS One* 2012;7:e37775.
15. Sato T, Arai E, Kohno T, et al. DNA methylation profiles at precancerous stages associated with recurrence of lung adenocarcinoma. *PLoS One* 2013;8:e59444.
16. Nagashio R, Arai E, Ojima H, et al. Carcinogenetic risk estimation based on quantification of DNA methylation levels in liver tissue at the precancerous stage. *Int J Cancer* 2011;129:1170–9.
17. Arai E, Ushijima S, Gotoh M, et al. Genome-wide DNA methylation profiles in liver tissue at the precancerous stage and in hepatocellular carcinoma. *Int J Cancer* 2009;125:2854–62.
18. Ushijima T, Hattori N. Molecular pathways: involvement of Helicobacter pylori-triggered inflammation in the formation of an epigenetic field defect, and its usefulness as cancer risk and exposure markers. *Clin Cancer Res* 2012;18:923–9.
19. Kanai Y, Hirohashi S. Alterations of DNA methylation associated with abnormalities of DNA methyltransferases in human cancers during transition from a precancerous to a malignant state. *Carcinogenesis* 2007;28:2434–42.
20. Garibaldi BT, Illei P, Danoff SK. Bronchiolitis. *Immunol Allergy Clin North Am* 2012;32:601–19.
21. Travis WD, Colby TV, Koss MN, et al. Obstructive pulmonary disease. In: King DW, ed. Atlas of nontumor pathology. Non-neoplastic disorders of the lower respiratory tract. Washington, DC: American Registry of Pathology, 2002:435–71.
22. Kerr KM. Clinical relevance of the new IASLC/ERS/ATS adenocarcinoma classification. *J Clin Pathol* 2013;66:832–8.
23. Noguchi M, Shimosato Y. The development and progression of adenocarcinoma of the lung. *Cancer Treat Res* 1995;72:131–42.
24. Colby TV, Noguchi M, Henschke C, et al. Adenocarcinoma. In: Travis WD, Brambilla E, Muller-Hermelink HK, Harris CC, eds. World health classification of tumours pathology and genetics of tumours of the lung, pleura, thymus and heart. Lyon: IARC Press, 2004:35–44.
25. Wang D, Minami Y, Shu Y, et al. The implication of background anthracosis in the development and progression of pulmonary adenocarcinoma. *Cancer Sci* 2003;94:707–11.
26. Noguchi M, Morikawa A, Kawasaki M, et al. Small adenocarcinoma of the lung. Histologic characteristics and prognosis. *Cancer* 1995;75:2844–52.
27. UICC International Union Against Cancer. Lung and pleural tumours. In: Sobin LH, Gospodarowicz MK, Wittekind C, eds. TNM classification of malignant tumours, 7th edn. Oxford: Wiley-Blackwell, 2009:138–46.
28. Hu L, Sekine M, Gaina A, et al. Association of smoking behavior and socio-demographic factors, work, lifestyle and mental health of Japanese civil servants. *J Occup Health* 2007;49:443–52.
29. Kadara H, Kabbout M, Wistuba II. Pulmonary adenocarcinoma: a renewed entity in 2011. *Respirology* 2012;17:50–65.
30. Decramer M, Janssens W, Miravitlles M. Chronic obstructive pulmonary disease. *Lancet* 2012;379:1341–51.
31. Peng DF, Kanai Y, Sawada M, et al. DNA methylation of multiple tumor-related genes in association with overexpression of DNA methyltransferase 1 (DNMT1) during multistage carcinogenesis of the pancreas. *Carcinogenesis* 2006;27:1160–8.
32. Peng DF, Kanai Y, Sawada M, et al. Increased DNA methyltransferase 1 (DNMT1) protein expression in precancerous conditions and ductal carcinomas of the pancreas. *Cancer Sci* 2005;96:403–8.
33. Suzuki M, Wada H, Yoshino M, et al. Molecular characterization of chronic obstructive pulmonary disease-related non-small cell lung cancer through aberrant methylation and alterations of EGFR signaling. *Ann Surg Oncol* 2010;17:878–88.
34. Sharma SV, Bell DW, Settleman J, et al. Epidermal growth factor receptor mutations in lung cancer. *Nat Rev Cancer* 2007;7:169–81.
35. Mummenhoff J, Houweling AC, Peters T, et al. Expression of Irx6 during mouse morphogenesis. *Mech Dev* 2001;103:193–5.
36. Leshchenko VV, Kuo PY, Shaknovich R, et al. Genomewide DNA methylation analysis reveals novel targets for drug development in mantle cell lymphoma. *Blood* 2010;116:1025–34.
37. Kanai M, Hamada J, Takada M, et al. Aberrant expressions of HOX genes in colorectal and hepatocellular carcinomas. *Oncol Rep* 2010;23:843–51.
38. Hurley PJ, Marchionni L, Simons BW, et al. Secreted protein, acidic and rich in cysteine-like 1 (SPARCL1) is down regulated in aggressive prostate cancers and is prognostic for poor clinical outcome. *Proc Natl Acad Sci USA* 2012;109:14977–82.
39. Turtoi A, Musmeci D, Naccarato AG, et al. Sparc-like protein 1 is a new marker of human glioma progression. *J Proteome Res* 2012;11:5011–21.
40. Jin Y, An X, Ye Z, et al. RGS5, a hypoxia-inducible apoptotic stimulator in endothelial cells. *J Biol Chem* 2009;284:23436–43.
41. Huang G, Song H, Wang R, et al. The relationship between RGS5 expression and cancer differentiation and metastasis in non-small cell lung cancer. *J Surg Oncol* 2012;105:420–4.
42. Wang JH, Huang WS, Hu CR, et al. Relationship between RGS5 expression and differentiation and angiogenesis of gastric carcinoma. *World J Gastroenterol* 2010;16:5642–6.
43. Zhao X, Ayer RE, Davis SL, et al. Apoptosis factor EI24/PIG8 is a novel endoplasmic reticulum-localized Bcl-2-binding protein which is associated with suppression of breast cancer invasiveness. *Cancer Res* 2005;65:2125–9.

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

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

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

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

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

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

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

Sanger sequencing

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

SNP microarray analysis

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

Infinium analysis

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

Pyrosequencing

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

Expression microarray analysis

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

Quantitative RT-PCR analysis

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

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

Multilayer-omics scoring

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

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

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

Pathway analysis

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

Results

Genetic aberrations

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

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

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

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

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.

Table 2. Genes showing downregulation or upregulation scores of 5 or more in clear cell RCCs

Gene	Chromosome	Entrez GeneID	Downregulation score[1]
(a) Genes showing reduced mRNA expression associated with DNA hypemethylation in their 5′-regions			
CLCNKB	1	1,188	24
SCNN1A	12	6,337	24
RAB25	1	57,111	22
TMEM213	7	155,006	22
ATP6V0A4	7	50,617	22
NR0B2	1	8,431	21
KCNJ1	11	3,758	21
GGT6	17	124,975	21
CLDN8	21	9,073	20
CLDN19	1	149,461	19
MUC15	11	143,662	16
RANBP3L	5	202,151	15
HRG	3	3,273	14
TSPAN8	12	7,103	14
RGS7	1	6,000	11
PTH1R	3	5,745	11
CWH43	4	80,157	11
F11	4	2,160	11
IRX2	5	153,572	11
EHF	11	26,298	11
CBLC	19	23,624	11
ATP6V1B1	2	525	10
LRRC2	3	79,442	10
CLDN16	3	10,686	10
EGF	4	1,950	10
WISP3	6	8,838	10
PHYHD1	9	254,295	10
FLJ45983	10	399,717	10
WIT-AS	11	51,352	10
ACSF2	17	80,221	10
ALDOB	9	229	9
ANKRD2	10	26,287	9
WT1	11	7,490	9
SOST	17	50,964	9
CYP4F3	19	4,051	9
COL18A1-AS1	21	378,832	9
BSND	1	7,809	8
TACSTD2	1	4,070	8
SLC44A4	6	80,736	8
KHDRBS2	6	202,559	8
VWC2	7	375,567	8
CHRM1	11	1,128	8
COL4A6	X	1,288	8
XPNPEP2	X	7,512	8
PROM2	2	150,696	7
ACPP	3	55	7
CKMT2	5	1,160	7
NEFM	8	4,741	7
KCNA4	11	3,739	7
FLRT1	11	23,769	7
OLFM4	13	10,562	7
SERPINA4	14	5,267	7
STRA6	15	64,220	7
CRABP1	15	1,381	7
SLC7A10	19	56,301	7
CSDC2	22	27,254	7
VWA5B1	1	127,731	6
LAD1	1	3,898	6
SYN2	3	6,854	6
SLC22A13	3	9,390	6
ABHD14A	3	25,864	6
UPK1B	3	7,348	6
KCTD8	4	386,617	6
SFRP1	8	6,422	6
GATA3	10	2,625	6
DAO	12	1,610	6
TMPRSS3	21	64,699	6
CHD5	1	26,038	5
PRELP	1	5,549	5
PLD5	1	200,150	5
MAL	2	4,118	5
ENTPD3	3	956	5
TNNC1	3	7,134	5
ANK2	4	287	5
PART1	5	25,859	5
SVOPL	7	136,306	5
DMRT2	9	10,655	5
AMBP	9	259	5
RBP4	10	5,950	5
SLC22A12	11	116,085	5
PDZRN4	12	29,951	5
PROZ	13	8,858	5
RHCG	15	51,458	5
KLK6	19	5,653	5

Table 2. Genes showing downregulation or upregulation scores of 5 or more in clear cell RCCs (Continued)

Gene	Chromosome	Entrez GeneID	Downregulation score[1]
BEX1	X	55,859	5
ZCCHC16	X	340,595	5
Gene	**Chromosome**	**Entrez GeneID**	**Up-regulation score[2]**
(b) Genes showing increased mRNA expression associated with DNA hypomethylation in their 5′-regions.			
CA9	9	768	25
C3	19	718	23
CP	3	1,356	22
NNMT	11	4,837	21
FABP7	6	2,173	11
REG1A	2	5,967	10
UBD	6	10,537	8
ENPP3	6	5,169	8
MCHR1	22	2,847	7
FCGR3A	1	2,214	6
FGG	4	2,266	6
PMCHL1	5	5,369	6
CPA6	8	57,094	6
SAA2	11	6,289	6
SAA1	11	6,288	6
DNAJB13	11	374,407	6
VWF	12	7,450	6
FGF11	17	2,256	6
SPAG4	20	6,676	6
CHI3L2	1	1,117	5
FCRL3	1	115,352	5
TIGIT	3	201,633	5
APOLD1	12	81,575	5
CCL18	17	6,362	5
CARD14	17	79,092	5
LILRA2	19	11,027	5
CXorf36	X	79,742	5
SH2D1A	X	4,068	5

[1]If the probe of the Infinium array was designed in the 5′-region of the gene, if $\Delta\beta$ ($\beta_T - \beta_N$) 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 (T and N), then a gene downregulation score of 1 was assigned.
[2]If the probe of the Infinium array was designed in the 5′-region of the gene, if $\Delta\beta$ ($\beta_T - \beta_N$) 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 (T and N), then a gene upregulation score of 1 was assigned.

Among genes showing frequent genetic aberrations (genetic aberration score of 4 or more in Table 1), *GCN1L1* has recently been reported to be associated with the *CDK8* mediator complex, which includes *CDK8*, cyclin C (also known as *CCNC*), *MED12* and *MED13*.[25] *CDK8* directly regulates β-catenin-driven transcription[25] and human *CDK8* is known to be an oncogene that is amplified in a subset of colon cancers.[26] In addition, our quantitative RT-PCR analysis revealed a tendency for down regulation of β-catenin after knockdown of CDK8 by siRNA in RCC cell lines A-498 and ACHN (Supporting Information Fig. S4). These results are consistent with those of previous studies showing that knockdown of CDK8 in the human colon cancer cell line HCT116[27] and the human gastric cancer cell line SNU-638[28] resulted in significant reduction of β-catenin, indicating correlations between *CDK8* and the Wnt/β-catenin pathway.

The fly *MED12* and *MED 13* homologs, kohtalo and skuld, respectively activate Wnt/β-catenin target genes through direct interaction with the Wnt pathway component Pygopus.[29] However, *let-19* and *doy-22*, homologs of human *MED12* and *MED13*, respectively, in *Caenorhabditis elegans*, suppress the transcription of Wnt/β-catenin target genes.[30] Frequent mutation of human *MED12* has been reported in human uterine leiomyomas.[31] Deletion of the *CCNC* gene is frequently detected in human lymphoid malignancies[32] and sarcomas.[33] Wnt/β-catenin signaling is constitutively active in RCCs and activates their cell growth and metastasis.[34] However, unlike other human carcinomas, the incidence of mutation of exon 3 of the β-catenin gene is not so high in RCCs.[34] Analogously with other members of the *CDK8* mediator complex, mutations of *GCN1L1* may participate in renal carcinogenesis *via* Wnt/β-catenin signaling.

All 5 amino acid substitutions of the *GCN1L1* occurred within or near to Huntingtin protein, eEF3, protein phosphatase 2A and TOR (HEAT) repeats, which are crucial for protein-protein interaction[35] (Supporting Information Fig. S5). In addition, SIFT and PolyPhen-2 software predicted that amino acid substitutions due to mutations of the *GCN1L1* gene result in dysfunction of GCN1L1 protein (Table 1). The present study demonstrated not only a genetic aberration score of 5 for *GCN1L1*, but also a genetic aberration score of 3 for *MED12* and *CCNC* (Table 1). SIFT and PolyPhen-2 analyses have predicted that amino acid substitutions due to mutations of the *MED12* and *CCNC* genes also result in dysfunction of the proteins (Table 1). Taken together, the present data indicate that the function of the *CDK8* mediator complex may have been disturbed in 16% of the examined 67 RCCs. Genetic aberrations in members of the *CDK8* mediator complex may thus participate in the Wnt/β-catenin-related carcinogenetic pathway in clear cell RCCs.

MACF 1, a member of the plakin family of cytoskeletal linker proteins, regulates dynamic interactions between actin and microtubules to sustain directional cell movement.[36] *MACF1* is known to function in the Wnt signaling pathway through association with a complex containing axin, β-catenin, *GSK3β* and *APC* during mouse embryogenesis.[36] Somatic mutation of *MACF1* (Table 1) may also participate in the Wnt/β-catenin-related carcinogenetic pathway in clear cell RCCs. With respect

Table 3. Statistically significant GeneGo pathway maps revealed by MetaCore pathway analysis

		Involved genes		
Pathway	*P*-value	Genes	Entrez Gene ID	Multilayer-omics scoring (exome, methylome and transcriptome)
Cell adhesion_tight junctions	9.98×10^{-4}	*CLDN8*	9073	Downregulation score 20
		CLDN16	10686	Downregulation score 10
		CLDN19	149461	Downregulation score 19
Blood coagulation	1.26×10^{-3}	*VWF*	7450	Upregulation score 6
		F11	2160	Downregulation score 11
		FGG	2266	Upregulation score 6
Translation_non-genomic (rapid) action of androgen receptor	1.36×10^{-3}	*MTOR*	2475	Genetic score 4
		PTEN	5728	Genetic score 3
		EGF	1950	Downregulation score 10
Signal transduction_PTEN pathway	2.04×10^{-3}	*MTOR*	2475	Genetic score 4
		PTEN	5728	Genetic score 3
		EGF	1950	Downregulation score 10
Development_EGFR signaling *via* PIP3	7.04×10^{-3}	*PTEN*	5728	Genetic score 3
		EGF	1950	Downregulation score 10
Protein folding and maturation_-Bradykinin/ Kallidin maturation	1.34×10^{-2}	*KLK6*	5653	Downregulation score 5
		XPNPEP2	7512	Downregulation score 8
Transcription_receptor-mediated HIF regulation	1.95×10^{-2}	*MTOR*	2475	Genetic score 4
		PTEN	5728	Genetic score 3
Serotonin modulation of dopamine release in nicotine addiction	2.24×10^{-2}	*PTEN*	5728	Genetic score 3
		CHRM1	1128	Downregulation score 8
Signal transduction_AKT signaling	2.34×10^{-2}	*MTOR*	2475	Genetic score 4
		PTEN	5728	Genetic score 3
cAMP/ Ca(2+)-dependent Insulin secretion	2.34×10^{-2}	*PLCE1*	51196	Genetic score 3
		RYR2	6262	Genetic score 3
Immune response_interleukin-4 signaling pathway	2.45×10^{-2}	*MTOR*	2475	Genetic score 4
		GATA3	2625	Downregulation score 6
Role of alpha-6/beta-4 integrins in carcinoma progression	2.55×10^{-2}	*MTOR*	2475	Genetic score 4
		EGF	1950	Downregulation score 10
G-protein signaling_regulation of cAMP levels by muscarinic acetylcholine receptor	2.55×10^{-2}	*PLCE1*	51196	Genetic score 3
		CHRM1	1128	Downregulation score 8
Development_PIP3 signaling in cardiac myocytes	2.77×10^{-2}	*MTOR*	2475	Genetic score 4
		PTEN	5728	Genetic score 3

Table 3. Statistically significant GeneGo pathway maps revealed by MetaCore pathway analysis (Continued)

Pathway	*P*-value	Involved genes		
		Genes	Entrez Gene ID	Multilayer-omics scoring (exome, methylome and transcriptome)
Some pathways of EMT in cancer cells	3.22×10^{-2}	*MTOR*	2475	Genetic score 4
		EGF	1950	Downregulation score 10
Development_beta-adrenergic receptors signaling *via* cAMP	3.34×10^{-2}	*RYR2*	6262	Genetic score 3
		TNNC1	7134	Downregulation score 5
Development_IGF-1 receptor signaling	3.34×10^{-2}	*MTOR*	2475	Genetic score 4
		PTEN	5728	Genetic score 3
Translation _regulation of EIF4F activity	3.45×10^{-2}	*MTOR*	2475	Genetic score 4
		EGF	1950	Downregulation score 10
G-protein signaling_RAP2B regulation pathway	3.81×10^{-2}	*PLCE1*	51196	Genetic score 3
DNA damage_DNA-damage-induced responses	4.87×10^{-2}	*ATM*	472	Genetic score 3

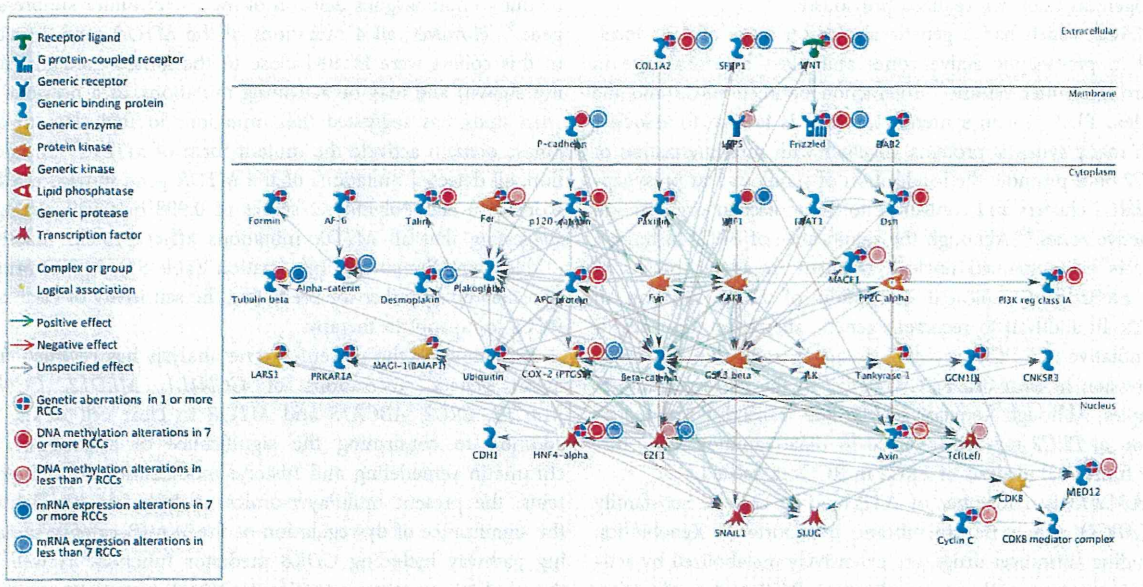

Figure 1. Genes for which a correlation with Wnt/β-catenin signaling was indicated by MetaCore pathway analysis. The numbers of genetic aberrations, DNA hyper- or hypo-methylation and/or increased or reduced mRNA expression (shown in Supporting Information Table S6) detected among the 67 examined RCCs are indicated schematically: legends are shown at the left of the panel. The 36 marked genes that showed genetic aberration, DNA methylation alterations and/or mRNA expression alterations in one or more RCCs were correlated with Wnt/β-catenin signaling.

to 29 RCCs for which transcriptome analysis was performed, mRNA expression levels of the targets genes of the Wnt/β-catenin signaling, such as *MYC*,[37] *MYCN*,[37] *IGF2*,[38] *POU5F1*,[39] *SOX9*,[40] *CYR61*,[41] *ENPP2*[42] and *MITF*,[43] tended to be higher in the 8 RCCs with mutations of any of the *GCN1L1*, *MED12*, *CCNC* and *MACF1* genes than in 21 RCCs without them (Supporting Information Table S10), indicating that such mutations may result in activation of Wnt/β-catenin signaling.

The downregulation score for the *SFRP1* gene was 6: reduced expression associated with DNA hypermethylation of *SFRP1* was frequent in clear cell RCCs. Members of the secreted frizzled-related protein (SFRP) family contain an N-terminal domain homologous to the cysteine-rich domain of the Wnt receptor Frizzled and lack a transmembrane region and the cytoplasmic domain required for transduction of signals into the cells.[44] This enables SFRPs to downregulate Wnt/β-catenin signaling by competing with Frizzled for Wnt binding *via* their cysteine-rich domain. Silencing of *SFRP1* due to DNA hypermethylation is known to result in activation of Wnt/β-catenin signaling.[44]

Since this study indicated possible alternative activation mechanisms (mutations of the *GCN1L1, MED12, CCNC* and *MACF1* genes and reduced expression of *SFRP1* due to DNA hypermethylation), we extensively examined Wnt/β-catenin signaling. MetaCore pathway analysis revealed that the 36 genes (marked in Fig. 1 and included in Supporting Information Table S6), which showed genetic aberration, DNA hypermethylation or hypomethylation and/or increased or reduced mRNA expression in one or more RCCs, are included in the Wnt/β-catenin signaling pathway. The present multilayer-omics analysis revealed that the Wnt/β-catenin signaling pathway may be of greater significance in renal carcinogenesis than was realized previously.

ERC2, which had a genetic aberration score of 4, is localized in presynaptic active zones and plays a critical role in neurotransmitter release.[45] Interaction between *ERC2* and the tandem PDZ protein syntenin-1, which is known to associate with many synaptic proteins, together with multimerization of *ERC2* both promote the localization of syntenin-1 at presynaptic *ERC2* clusters and contribute to the molecular organization of active zones.[45] Although the significance of *ERC2* in human cancers has remained unclear, frequent intragenic breaks in the *ERC2* gene indicated disruption of *ERC2* function in RCCs. In addition to recurrent genetic aberration, the present quantitative RT-PCR revealed frequent reduction of *ERC2* expression in clear cell RCCs relative to the corresponding N samples. Although frequent genetic and transcriptional inactivation of *ERC2* may be involved in renal carcinogenesis, further functional analysis of *ERC2* in RCCs is needed.

ABCA13 is a member of ATP-binding cassette sub-family A (*ABC1*) and a transmembrane transporter.[46] Xenobiotics, including anticancer drugs, are extensively metabolized by activation enzymes such as cytochromes *P450* and conjugation enzymes such as glutathione S-transferases or glucuronide transferases. Biotransformation represented by ABC transporters represents another important component of xenobiotic metabolism. In addition, ABC transporters play a crucial role in the development of resistance through efflux of anticancer agents from cancer cells.[46] The disease-free interval of patients with colorectal cancers treated by adjuvant chemotherapy is significantly shorter in patients with low *ABCA13* transcript levels.[47] In addition to recurrent genetic aberration (Table 1), the present quantitative RT-PCR revealed frequently reduced expression of *ABCA13* in RCCs relative to the corresponding N samples. Our findings suggest that it may be necessary to pay more attention to aberrations of *ABCA13* at both the genetic and expressional levels when deciding the indications for chemotherapy in patients with clear cell RCCs.

In Table 3 based on MetaCore pathway analysis, it is feasible that expression of *CLDNs* required for generating cation-selective paracellular channels[48] was reduced in clear cell RCCs, which lack the original absorptive function of the renal tubule. Moreover, *MTOR* mutations were highlighted as one of the major disrupters of multiple cell signaling during renal carcinogenesis: the *MTOR* gene participated in 10 (50%) of the 20 significant pathways in Table 3. The mammalian target of rapamycin (*mTOR*) encoded by the *MTOR* gene is a serine/threonine kinase that regulates cell growth, proliferation and autophagy.[49] *mTOR* inhibitors, such as rapamycin and its derivatives, are being introduced for targeted therapy of clear cell RCCs. Overactivation of *mTOR* is generally considered to be due to homozygous deletion of the *PTEN* tumor suppressor gene.[50] However, all 4 mutations of the *MTOR* gene detected in this cohort were located close to the kinase domain (data not shown) and may be activating mutations, as a previous *in vitro* study has suggested that mutations located close to the kinase domain activate the mutant form of *mTOR*.[50] In addition, all detected mutations of the *MTOR* gene showed a SIFT score of 0 and PolyPhen-2 scores of 0.998 or 0.999, strongly suggesting that all *MTOR* mutations affect protein function (Table 1 and Supporting Information Table S3). *MTOR* mutation may be a marker for predicting the sensitivity of clear cell RCCs to rapamycin therapy.

In summary, the present exome analysis has revealed frequent genetic aberrations of *GCN1L1, MED12, CCNC, MACF1, ERC2, ABCA13* and *MTOR* in clear cell RCCs. In addition to confirming the significance of aberrations of chromatin remodeling and histone modification-related proteins, the present multilayer-omics analysis has highlighted the significance of dysregulation of the Wnt/β-catenin signaling pathway including *CDK8* mediator function, as well as the need to pay closer attention to *MTOR* mutations, causing major disruption of cell signaling during renal carcinogenesis, in relation to chemosensitivity. Multilayer-omics analysis can be considered a powerful tool for revealing significant carcinogenetic pathways in human cancers.

References

1. Ljungberg B, Campbell SC, Choi HY, et al. The epidemiology of renal cell carcinoma. *Eur Urol* 2011;60:615–21.
2. Dalgliesh GL, Furge K, Greenman C, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. *Nature* 2010;463:360–3.
3. van Haaften G, Dalgliesh GL, Davies H, et al. Somatic mutations of the histone H3K27 demethylase gene UTX in human cancer. *Nat Genet* 2009;41:521–3.

4. Varela I, Tarpey P, Raine K, et al. Exome sequencing identifies frequent mutation of the SWI/SNF complex gene PBRM1 in renal carcinoma. *Nature* 2011;469:539–42.
5. Baldewijns MM, van Vlodrop IJ, Vermeulen PB, et al. VHL and HIF signaling in renal cell carcinogenesis. *J Pathol* 2010;221:125–38.
6. Guo G, Gui Y, Gao S, et al. Frequent mutations of genes encoding ubiquitin-mediated proteolysis pathway components in clear cell renal cell carcinoma. *Nat Genet* 2012;44:17–9.
7. Baylin SB, Jones PA. A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* 2011;11:726–34.
8. Kanai, Y. Genome-wide DNA methylation profiles in precancerous conditions and cancers. *Cancer Sci* 2010;101:36–45.
9. Arai E, Ushijima S, Fujimoto H, et al. Genome-wide DNA methylation profiles in both precancerous conditions and clear cell renal cell carcinomas are correlated with malignant potential and patient outcome. *Carcinogenesis* 2009;30:214–21.
10. Arai E, Kanai Y, Ushijima S, et al. Regional DNA hypermethylation and DNA methyltransferase (DNMT) 1 protein overexpression in both renal tumors and corresponding nontumorous renal tissues. *Int J Cancer* 2006;119:288–96.
11. Arai E, Chiku S, Mori T, et al. Single-CpG-resolution methylome analysis identifies clinicopathologically aggressive CpG island methylator phenotype clear cell renal cell carcinomas. *Carcinogenesis* 2012;33:1487–93.
12. Sato Y, Yoshizato T, Shiraishi Y, et al. Integrated molecular analysis of clear-cell renal cell carcinoma. *Nat Genet* 2013;45:860–7.
13. Eble JN, Togashi K, Pisani P. Renal cell carcinoma. World Health Organization classification of tumours. Pathology and genetics. Tumours of the urinary system and male genital organs. Lyon: IARC Press, 2004. 10–43.
14. Fuhrman SA, Lasky LC, Limas, C. Prognostic significance of morphologic parameters in renal cell carcinoma. *Am J Surg Pathol* 1982;6:655–63.
15. Sobin LH, Gospodarowicz MK, Wittekind C, eds. International Union Against Cancer (UICC). TNM classification of malignant tumors, 7th edn. New York: Wiley, 2009.
16. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics* 2010;26:589–95.
17. Ye K, Schulz MH, Long Q, et al. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. *Bioinformatics* 2009;25: 2865–71.
18. Wang L, Tsutsumi S, Kawaguchi T, et al. Whole-exome sequencing of human pancreatic cancers and characterization of genomic instability caused by MLH1 haploinsufficiency and complete deficiency. *Genome Res* 2012;22:208–19.
19. Ng PC, Henikoff S. Accounting for human polymorphisms predicted to affect protein function. *Genome Res* 2002;12:436–46.
20. Hicks S, Wheeler DA, Plon SE, et al. Prediction of missense mutation functionality depends on both the algorithm and sequence alignment employed. *Hum Mutat* 2011;32:661–8.
21. Sim NL, Kumar P, Hu J, et al. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Res* 2012;40:W452–7.
22. Van Loo P, Nordgard SH, Lingjærde OC, et al. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci USA* 2010;107:16910–5.
23. Li A, Liu Z, Lezon-Geyda K, et al. 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* 2011;39: 4928–41.
24. Nagashio R, Arai E, Ojima H, et al. Carcinogenetic risk estimation based on quantification of DNA methylation levels in liver tissue at the precancerous stage. *Int J Cancer* 2011;129: 1170–9.
25. Firestein R, Hahn WC. Revving the Throttle on an oncogene: CDK8 takes the driver seat. *Cancer Res* 2009;69:7899–901.
26. Firestein R, Bass AJ, Kim SY, et al. CDK8 is a colorectal cancer oncogene that regulates beta-catenin activity. *Nature* 2008;455:547–51.
27. He SB, Yuan Y, Wang L, et al. Effects of cyclin-dependent kinase 8 specific siRNA on the proliferation and apoptosis of colon cancer cells. *J Exp Clin Cancer Res* 2011;30:109.
28. Seo JO, Han SI, Lim SC. Role of CDK8 and beta-catenin in colorectal adenocarcinoma. *Oncol Rep* 2010;24:285–91.
29. Carrera I, Janody F, Leeds N, et al. Pygopus activates Wingless target gene transcription through the mediator complex subunits Med12 and Med13. *Proc Natl Acad Sci USA* 2008;105:6644–9.
30. Yoda A, Kouike H, Okano H, et al. Components of the transcriptional mediator complex are required for asymmetric cell division in C. elegans. *Development* 2005;132:1885–93.
31. Mäkinen N, Mehine M, Tolvanen J, et al. MED12, the mediator complex subunit 12 gene, is mutated at high frequency in uterine leiomyomas. *Science* 2011;334:252–5.
32. Jackson A, Carrara P, Duke V, et al. Deletion of 6q16-q21 in human lymphoid malignancies: a mapping and deletion analysis. *Cancer Res* 2000; 60:2775–9.
33. Ohata N, Ito S, Yoshida A, et al. Highly frequent allelic loss of chromosome 6q16-23 in osteosarcoma: involvement of cyclin C in osteosarcoma. *Int J Mol Med* 2006;18:1153–8.
34. Banumathy G, Cairns P. Signaling pathways in renal cell carcinoma. *Cancer Biol Ther* 2010;10: 658–64.
35. Andrade MA, Petosa C, O'Donoghue SI, et al. Comparison of ARM and HEAT protein repeats. *J Mol Biol* 2001;309:1–18.
36. Chen HJ, Lin CM, Lin CS, et al. The role of microtubule actin cross-linking factor 1 (MACF1) in the Wnt signaling pathway. *Genes Dev* 2006;20:1933–45.
37. Karim R, Tse G, Putti T, et al. The significance of the Wnt pathway in the pathology of human cancers. *Pathology* 2004;36:120–8.
38. Heaton JH, Wood MA, Kim AC, et al. Progression to adrenocortical tumorigenesis in mice and humans through insulin-like growth factor 2 and β-catenin. *Am J Pathol* 2012;181:1017–33.
39. Li J, Li J, Chen B. Oct4 was a novel target of Wnt signaling pathway. *Mol Cell Biochem* 2012; 362:233–40.
40. Blache P, van de Wetering M, Duluc I, et al. SOX9 is an intestine crypt transcription factor, is regulated by the Wnt pathway, and represses the CDX2 and MUC2 genes. *J Cell Biol* 2004;166:37–47.
41. Li ZQ, Ding W, Sun SJ, et al. Cyr61/CCN1 is regulated by Wnt/β-catenin signaling and plays an important role in the progression of hepatocellular carcinoma. *PLoS One* 2012;7:e35754.
42. Zirn B, Samans B, Wittmann S, et al. Target genes of the WNT/beta-catenin pathway in Wilms tumors. *Genes Chromosomes Cancer* 2006; 45:565–74.
43. Syed DN, Afaq F, Maddodi N, et al. Inhibition of human melanoma cell growth by the dietary flavonoid fisetin is associated with disruption of Wnt/β-catenin signaling and decreased Mitf levels. *J Invest Dermatol* 2011;131:1291–9.
44. Suzuki H, Watkins DN, Jair KW, et al. Epigenetic inactivation of SFRP genes allows constitutive WNT signaling in colorectal cancer. *Nat Genet* 2004;36:417–22.
45. Ko J, Yoon C, Piccoli G, et al. Organization of the presynaptic active zone by ERC2/CAST1-dependent clustering of the tandem PDZ protein syntenin-1. *J Neurosci* 2006;26:963–70.
46. Cole SP, Bhardwaj G, Gerlach JH, et al. Overexpression of a transporter gene in a multidrug-resistant human lung cancer cell line. *Science* 1992;258:1650–4.
47. Hlavata I, Mohelnikova-Duchonova B, Vaclavikova R, et al. The role of ABC transporters in progression and clinical outcome of colorectal cancer. *Mutagenesis* 2012;27:187–96.
48. Angelow S, El-Husseini R, Kanzawa SA, et al. Renal localization and function of the tight junction protein, claudin-19. *Am J Physiol Renal Physiol* 2007;293:166–77.
49. Populo H, Lopes JM, Soares P. The mTOR signalling pathway in human cancer. *Int J Mol Sci* 2012;13:1886–918.
50. Sato T, Nakashima A, Guo L, et al. Single amino-acid changes that confer constitutive activation of mTOR are discovered in human cancer. *Oncogene* 2010;29:2746–52.

OPEN ACCESS Freely available online

PLOS | ONE

Associated Factors of Atrophic Gastritis Diagnosed by Double-Contrast Upper Gastrointestinal Barium X-Ray Radiography: A Cross-Sectional Study Analyzing 6,901 Healthy Subjects in Japan

Nobutake Yamamichi[1]*, Chigaya Hirano[2], Takeshi Shimamoto[1,2], Chihiro Minatsuki[1], Yu Takahashi[1], Chiemi Nakayama[1], Rie Matsuda[1], Mitsuhiro Fujishiro[1], Maki Konno-Shimizu[1], Jun Kato[3], Shinya Kodashima[1], Satoshi Ono[1], Keiko Niimi[1], Satoshi Mochizuki[1], Yosuke Tsuji[1], Yoshiki Sakaguchi[1], Itsuko Asada-Hirayama[1], Chihiro Takeuchi[1], Seiichi Yakabi[1], Hikaru Kakimoto[1], Ryoichi Wada[2], Toru Mitsushima[2], Masao Ichinose[3], Kazuhiko Koike[1]

1 Department of Gastroenterology, Graduate School of Medicine, The University of Tokyo, Tokyo, Japan, **2** Department of Gastroenterology, Kameda Medical Center Makuhari, Chiba, Japan, **3** Second Department of Internal Medicine, Wakayama Medical University, Wakayama, Japan

Abstract

Background: Double-contrast upper gastrointestinal barium X-ray radiography (UGI-XR) is one of the most widely conducted gastric cancer screening methods. It has been executed to find gastric cancer, but has not been usually executed to detect premalignant atrophic mucosa of stomach. To understand the meaning of UGI-XR-based atrophic gastritis, we analyzed its association with several causative factors including *Helicobacter pylori* (HP) infection.

Methods: We evaluated 6,901 healthy adults in Japan. UGI-XR-based atrophic gastritis was diagnosed based on the irregular shape of areae gastricae and its expansion in the stomach.

Results: Of the 6,433 subjects with no history of HP eradication and free from gastric acid suppressants, 1,936 were diagnosed as UGI-XR-based atrophic gastritis (mild: 234, moderate: 822, severe: 880). These were univariately associated with serum HP IgG and serum pepsinogen I/II ratio with statistical significance. The multiple logistic analysis calculating standardized coefficients (β) and odds ratio (OR) demonstrated that serum HP IgG (β = 1.499, OR = 4.48), current smoking (β = 0.526, OR = 1.69), age (β = 0.401, OR = 1.49), low serum pepsinogen I/II ratio (β = 0.339, OR = 1.40), and male gender (β = 0.306, OR = 1.36) showed significant positive association with UGI-XR-based atrophic gastritis whereas drinking and body mass index did not. Among the age/sex/smoking/drinking-matched 227 pairs derived from chronically HP-infected and successfully HP-eradicated subjects, UGI-XR-based atrophic gastritis was detected in 99.1% of the former but in only 59.5% of the latter subjects ($p<0.0001$). Contrastively, UGI-XR-based atrophic gastritis was detected in 13 of 14 HP-positive proton pump inhibitor users (92.9%) and 33 of 34 HP-positive histamine H_2-receptor antagonist users (97.1%), which are not significantly different from gastric acid suppressant-free subjects.

Conclusions: The presence of UGI-XR-based atrophic gastritis is positively associated with *Helicobacter pylori* infection, current smoking, age, decreased serum pepsinogen I/II ratio, and male gender. Eradication of *Helicobacter pylori* seems to superficially improve UGI-XR-based atrophic gastritis whereas intake of gastric acid suppressants does not.

Citation: Yamamichi N, Hirano C, Shimamoto T, Minatsuki C, Takahashi Y, et al. (2014) Associated Factors of Atrophic Gastritis Diagnosed by Double-Contrast Upper Gastrointestinal Barium X-Ray Radiography: A Cross-Sectional Study Analyzing 6,901 Healthy Subjects in Japan. PLoS ONE 9(10): e111359. doi:10.1371/journal.pone.0111359

Editor: Mitsunobu R. Kano, Okayama University, Japan

Received March 11, 2014; **Accepted** October 1, 2014; **Published** October 24, 2014

Data Availability: The authors confirm that all data underlying the findings are fully available without restriction. All relevant data are within the paper.

Funding: This work was supported in part by a research grant from Takeda Science Foundation, in part by a research grant from the Daiwa Security Health Foundation, and also in part by Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science. All the funders had no role in study design, data collection and analysis, decision to publish, or preparation of manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* Email: nyamamic-tky@umin.ac.jp

Introduction

The incidence and mortality of gastric cancer has gradually fallen in the recent several decades, but it is still the second leading cause of cancer death worldwide [1,2]. Many gastric cancer screening methods have been developed and executed especially in East Asia, where a high incidence of gastric cancer is observed [3]. Among them, the double-contrast upper gastrointestinal barium X-ray radiography (UGI-XR) is one of the most widely used screening methods for gastric cancer. It has been conducted in