A univariate Cox proportional hazard model showed that expression levels of 97 probes (representing 88 nonredundant genes) were correlated with PFS time ($p<0.01$). In case of multiple-tagged 8 genes (represented by 17 probes), we selected 8 probes (one probe per gene) with the largest sum of the squares of individual expression values for the respective genes as representatives [24]. A total of 88 genes (represented by 88 unique probes) were thereby identified as PFS-related profile. Furthermore, we applied the ridge regression model to estimate optimal regression coefficients (β) for 88 genes of the PFS-related profile (Table 2), and calculated the prognostic index for each sample using equation (1) as reported previously [18]. The 88-gene prognostic indices obtained were in the range of -5.09 to 4.14 (median, 0.11), and the frequency distribution of the indices among 110 patients was unimodal.

To assess the prognostic index as a categorical variable, we attempted to divide this dataset into two groups based on median prognostic index of 0.11 [9]. Patients were assigned to the "high-risk" group if their prognostic index was greater than or equal to the median value, whereas "low-risk" group was composed of cases with the prognostic indices that were less than the median. As shown in Figure 1A, patients with high-risk prognostic indices had shorter median PFS times than those belonging to low-risk group (median PFS, 12 months vs. 51 months; log rank test, $p<0.0001$).

We then performed univariate and multivariate Cox proportional hazard analyses to prove that the 88-gene prognostic index was an independent prognostic factor (Table 3). A univariate Cox's proportional hazard analysis showed that the prognostic index, stage, optimal surgery, and histological grade were correlated with PFS ($p<0.0001$, $p=0.022$, $p<0.0001$ and $p=0.016$, respectively). Moreover, a multivariate analysis showed that the prognostic index was most significantly associated with PFS time [hazard ratio (HR), 3.80; 95% confidence interval (CI), 2.68–5.61; $p<0.0001$].

Validation by Quantitative Real-Time RT-PCR

To validate the microarray expression data, we performed quantitative real-time RT-PCR for a subset of the discovery dataset (53 samples). The four genes, *E2F2*, *FOXJ1*, *DNAH7*, and *FILIP1*, were randomly selected for this purpose. There were significant correlations between microarray expression data and real-time RT-PCR expression data (Figure 2). In spite of the smaller sample size, we confirmed a significant association between PFS time and each of the real-time RT-PCR data for the four genes in the univariate Cox hazard model (data not shown).

Appling PFS-Related Profile to the External Dataset

We translated the 88 prognostic genes with Agilent Probe IDs to Affymetrix 196 probes using a translation function in GeneSpring GX 10 and evaluated the present PFS-related profile in the external dataset (Figure S2). We calculated the prognostic index for each sample in the external dataset by the weighted sum of the expression values of 88 PFS-related genes according to the equation (1), in which the ridge regression coefficients (β) identified in the discovery set were used as weights for the respective genes (See Materials and Methods). We obtained prognostic indices ranging from -5.37 to 4.56 in the external dataset. The frequency distribution of the prognostic indices was not statistically different from that in the discovery set by Kolmogorov Smirnov test ($p>0.05$).

When we divided the external dataset into two subgroups by the median prognostic index (0.11) in the discovery set, a significant correlation was observed between risk classification and PFS (log rank test, $p=0.0004$) (Figure 1B). In univariate analysis of the external data, the estimated prognosis index and optimal surgery were correlated with PFS time ($p=0.0001$ and 0.049, respectively) (Table 3). Multivariate analysis showed that prognostic index was an independent prognostic factor for PFS time (HR, 1.64; 95% CI, 1.27–2.13, $p=0.0001$).

Assessment of Our Prognostic Index

To assess the sensitivity and specificity of our prognostic index, we used ROC curves for the index. An area under ROC curve of 0.5 (indicated by diagonal dotted lines in Figure S3) represents equality between true positive and false positive test results. The extent to which the ROC curve departs from the diagonal line to left and top axes is a measure of the effectiveness of the 88-gene prognostic index in the prediction of clinical outcome. The area under the ROC curves to distinguish early-relapse patients with less than 18 months of PFS times from late-relapse patients was 0.959 and 0.674 in the discovery set and the external dataset, respectively (Figure S3). When we used median value of prognostic index in the discovery set as the cut-off, the sensitivity and specificity were 88.9% and 85.7% in discovery dataset and 64.4% and 69.2% in the external dataset.

We performed survival analysis after the stratification of patients according to the status of debulking surgery which was an independent prognostic factor in multivariate analysis of the discovery dataset (Table 3). We divided patients into two groups ("optimal group" and "suboptimal group") in each of the discovery and external datasets, and assigned each patient to "high-risk" or "low-risk" based on the median value of the current prognostic index in each stratum according to the debulking status. Kaplan-Meier survival analysis showed that high-risk patients had significant shorter PFS time than low-risk patients in each of the four strata from the two datasets (Figure 3) as follows: optimal group ($p<0.0001$) and suboptimal group ($p<0.0001$) in our dataset; optimal group ($p=0.0034$) and suboptimal group ($p=0.015$) in the external dataset. This stratified analysis also indicated that the prognostic index was associated with PFS time independently of the debulking status.

Correlation between This Prognostic Index and Overall Survival

Overall survival is another important endpoint in patients with advance-stage ovarian cancers, and hence we further examined if the present 88-gene prognostic index could be extended to use for predicting the overall survival of patients. To evaluate correlation between this prognostic index and overall survival time, we performed Kaplan-Meier survival curve analysis. Patients with high-risk prognostic indices had shorter overall survival times than the low-risk patients in the two datasets (log rank test, $p<0.0001$ and $p=0.0010$, respectively) (Figure 1C, D). Furthermore, the prognostic index was significantly associated with overall survival time in both the discovery set and the external dataset in multivariate analysis (Table 4).

In addition, we examined the predictive ability of our prognostic index in publicly available Dressman's dataset [25], in which patients were longer followed-up (median overall survival, 31 months; range, 1–185 months). Dressman's dataset [25] was composed of 119 advanced-stage serous ovarian cancer patients treated with platinum-based chemotherapy (including non-taxane chemotherapy). Because their data were generated by a different platform with the foregoing two datasets, 75% of 88 PFS-related genes were translated for survival prediction in this dataset. When we divided Dressman's dataset [25] into two subgroups by the median prognostic index in discovery dataset, a significant association was observed between risk classification and overall survival (log rank test, $p=0.0008$) (Figure S4). Its prognostic index

Table 2. Eighty-eight genes composing the progression-free survival-related profile.

GenBank Acc.	GeneSymbol	Cytoband	β_{ridge}[a]	Description
NM_001123	*ADK*	10q22.2	0.006	adenosine kinase
NM_006408	*AGR2*	7p21.1	0.128	anterior gradient homolog 2 (Xenopus laevis)
NM_080429	*AQP10*	1q21.3	−0.162	aquaporin 10
NM_001040118	*ARAP1*	11q13.4	0.141	ArfGAP with RhoGAP domain, ankyrin repeat and PH domain 1
NM_006420	*ARFGEF2*	20q13.13	0.032	ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin A-inhibited)
NM_181575	*AUP1*	2p13.1	0.129	ancient ubiquitous protein 1
NM_004776	*B4GALT5*	20q13.13	0.215	UDP-Gal:betaGlcNAc beta 1,4- galactosyltransferase, polypeptide 5
NM_138639	*BCL2L12*	19q13.33	−0.189	BCL2-like 12 (proline rich)
NM_020643	*C11orf16*	11p15.4	0.221	chromosome 11 open reading frame 16
NM_145061	*C13orf3*	13q12.11	−0.107	chromosome 13 open reading frame 3
NM_024032	*C17orf53*	17q21.31	−0.184	chromosome 17 open reading frame 53
NM_001144956	*C1orf230*	1q21.3	0.012	chromosome 1 open reading frame 230
NM_022106	*C20orf177*	20q13.33	0.167	chromosome 20 open reading frame 177
NM_000715	*C4BPA*	1q32.2	−0.505	complement component 4 binding protein, alpha
NM_012337	*CCDC19*	1q23.2	−0.162	coiled-coil domain containing 19
NM_015603	*CCDC9*	19q13.32	0.263	coiled-coil domain containing 9
NM_005408	*CCL13*	17q12	−0.228	chemokine (C-C motif) ligand 13
NM_001252	*CD70*	19p13.3	−0.204	CD70 molecule
NM_078481	*CD97*	19p13.12	−0.137	CD97 molecule
NM_006383	*CIB2*	15q25.1	0.359	calcium and integrin binding family member 2
NM_182848	*CLDN10*	13q32.1	−0.292	claudin 10
NM_001316	*CSE1L*	20q13.13	−0.220	CSE1 chromosome segregation 1-like (yeast)
NM_024295	*DERL1*	8q24.13	0.007	Der1-like domain family, member 1
NM_001042517	*DIAPH3*	13q21.2	0.022	diaphanous homolog 3 (Drosophila)
NM_021120	*DLG3*	Xq13.1	−0.039	discs, large homolog 3 (Drosophila)
NM_020877	*DNAH2*	17p13.1	−0.378	dynein, axonemal, heavy chain 2
NM_018897	*DNAH7*	2q32.3	0.226	dynein, axonemal, heavy chain 7
NM_001394	*DUSP4*	8p21.1	0.007	dual specificity phosphatase 4
NM_004091	*E2F2*	1p36.12	0.220	E2F transcription factor 2
NM_006795	*EHD1*	11q13.1	0.248	EH-domain containing 1
NM_020819	*FAM135A*	6q13	0.142	family with sequence similarity 135, member A
NM_032181	*FAM176A*	2p12	−0.096	family with sequence similarity 176, member A
NM_015687	*FILIP1*	6q14.1	−0.188	filamin A interacting protein 1
NM_021784	*FOXA2*	20p11.21	0.184	forkhead box A2
NM_001454	*FOXJ1*	17q25.1	−0.344	forkhead box J1
NM_000819	*GART*	21q22.11	0.140	phosphoribosylglycinamide formyltransferase, phosphoribosylglycinamide synthetase, phosphoribosylaminoimidazole synthetase
NM_178172	*GPIHBP1*	8q24.3	0.147	glycosylphosphatidylinositol anchored high density lipoprotein binding protein 1
NM_000189	*HK2*	2p13.1	−0.087	hexokinase 2
NM_002118	*HLA-DMB*	6p21.32	−0.288	major histocompatibility complex, class II, DM beta
NM_022465	*IKZF4*	12q13.2	−0.092	IKAROS family zinc finger 4 (Eos)
NM_016584	*IL23A*	12q13.2	0.493	interleukin 23, alpha subunit p19
NM_006801	*KDELR1*	19q13.32	−0.001	KDEL (Lys-Asp-Glu-Leu) endoplasmic reticulum protein retention receptor 1
NM_014895	*KIAA1009*	6q14.3	−0.150	KIAA1009
NM_017527	*LY6K*	8q24.3	0.226	lymphocyte antigen 6 complex, locus K
NM_005906	*MAK*	6p24.2	0.271	male germ cell-associated kinase
NM_024871	*MAP6D1*	3q27.1	−0.038	MAP6 domain containing 1
NM_031417	*MARK4*	19q13.32	0.040	MAP/microtubule affinity-regulating kinase 4
NM_024298	*MBOAT7*	19q13.42	−0.058	membrane bound O-acyltransferase domain containing 7
NM_002421	*MMP1*	11q22.2	−0.336	matrix metallopeptidase 1 (interstitial collagenase)
NM_181526	*MYL9*	20q11.23	0.058	myosin, light chain 9, regulatory

Table 2. Cont.

GenBank Acc.	GeneSymbol	Cytoband	β_{ridge}[a]	Description
NM_032344	*NUDT22*	11q13.1	0.198	nudix (nucleoside diphosphate linked moiety X)-type motif 22
NM_007224	*NXPH4*	12q13.3	−0.310	neurexophilin 4
NM_015311	*OBSL1*	2q35	−0.045	obscurin-like 1
NM_014982	*PCNX*	14q24.2	−0.098	pecanex homolog (Drosophila)
NM_014317	*PDSS1*	10p12.1	0.001	prenyl (decaprenyl) diphosphate synthase, subunit 1
NM_024420	*PLA2G4A*	1q31.1	0.107	phospholipase A2, group IVA (cytosolic, calcium-dependent)
NM_016341	*PLCE1*	10q23.33	0.029	phospholipase C, epsilon 1
NM_001031745	*RIBC1*	Xp11.22	0.209	RIB43A domain with coiled-coils 1
NM_015653	*RIBC2*	22q13.31	0.053	RIB43A domain with coiled-coils 2
NM_006987	*RPH3AL*	17p13.3	−0.043	rabphilin 3A-like (without C2 domains)
NM_001025070	*RPS14*	5q33.1	0.013	ribosomal protein S14
NM_152732	*RSPH9*	6p21.1	−0.102	radial spoke head 9 homolog (Chlamydomonas)
NM_014433	*RTDR1*	22q11.22	0.034	rhabdoid tumor deletion region gene 1
NM_005500	*SAE1*	19q13.32	0.038	SUMO1 activating enzyme subunit 1
NM_020150	*SAR1A*	10q22.1	0.277	SAR1 homolog A (S. cerevisiae)
NM_031469	*SH3BGRL2*	6q14.1	−0.281	SH3 domain binding glutamic acid-rich protein like 2
NM_003951	*SLC25A14*	Xq25	−0.344	solute carrier family 25 (mitochondrial carrier, brain), member 14
NM_014585	*SLC40A1*	2q32.2	0.065	solute carrier family 40 (iron-regulated transporter), member 1
NM_052910	*SLITRK1*	13q31.1	−0.314	SLIT and NTRK-like family, member 1
NM_172312	*SPAG8*	9p13.3	−0.123	sperm associated antigen 8
NM_145263	*SPATA18*	4q12	0.041	spermatogenesis associated 18 homolog (rat)
NM_006100	*ST3GAL6*	3q12.1	−0.192	ST3 beta-galactoside alpha-2,3-sialyltransferase 6
NM_018414	*ST6GALNAC1*	17q25.1	−0.175	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3) -N-acetylgalactosaminide alpha-2,6-sialyltransferase 1
NM_032872	*SYTL1*	1p36.11	−0.084	synaptotagmin-like 1
NM_014466	*TEKT2*	1p34.3	−0.226	tektin 2 (testicular)
NM_005424	*TIE1*	1p34.2	0.250	tyrosine kinase with immunoglobulin-like and EGF-like domains 1
NM_198276	*TMEM17*	2p15	0.025	transmembrane protein 17
NM_199203	*TMEM189 -UBE2V1*	20q13.13	0.174	TMEM189-UBE2V1 readthrough transcript
NM_033550	*TP53RK*	20q13.12	0.054	TP53 regulating kinase
NM_139075	*TPCN2*	11q13.2	0.034	two pore segment channel 2
NM_018430	*TSNAXIP1*	16q22.1	0.170	translin-associated factor X interacting protein 1
NM_014023	*WDR37*	10p15.3	0.296	WD repeat domain 37
NM_018053	*XKR8*	1p35.3	0.106	XK, Kell blood group complex subunit-related family, member 8
NM_015896	*ZMYND10*	3p21.31	0.052	zinc finger, MYND-type containing 10
NM_005773	*ZNF256*	19q13.43	0.048	zinc finger protein 256
NM_024691	*ZNF419*	19q13.43	−0.042	zinc finger protein 419
NM_021089	*ZNF8*	19q13.43	0.093	zinc finger protein 8
NM_017975	*ZWILCH*	15q22.31	−0.074	Zwilch, kinetochore associated, homolog (Drosophila)

[a]A regression coefficient of each gene in ridge regression extension of multivariate Cox hazard model.

doi:10.1371/journal.pone.0009615.t002

was significantly correlated with overall survival time in multivariate analysis (HR, 1.51; 95% CI, 1.19–1.93, $p = 0.0008$).

Characterization of PFS-Related Profile

We conducted GO analysis to understand the biological characteristics of 88 PFS-related genes. To characterize the gene list based on GO classification on 'biological process', 'molecular function', and 'cellular component', we examined which categories were highly associated with the 88 genes. After multiple testing corrections using the FDR method [26], 8 categories were significantly overrepresented (FDR q-value<0.10) (Figure 4). In the 88 PFS-related genes, genes involved in GTPase binding (GO17016, GO31267 and GO51020), cellular localization (GO51649 and GO51641), intracellular transport (GO46907 and GO6886), and/or ciliary or flagellar motility (GO1539) were notably enriched. We investigated similarities in overrepresented GO categories between our 88 PFS-related genes and the previously reported gene expression profiles which were correlated to prognosis in ovarian cancer [11,13]. However, we could not identify common GO categories between our profile and the previously reported profiles (data not shown).

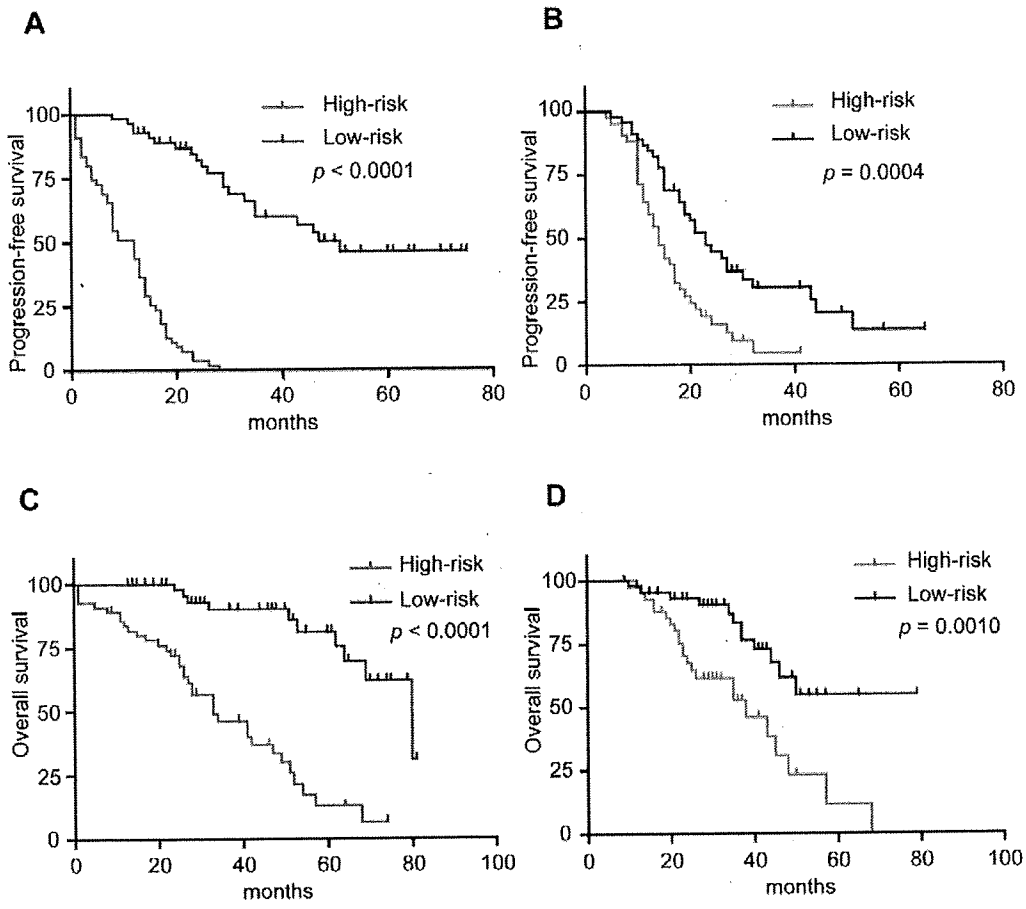

Figure 1. Prediction of prognosis in high-risk and low-risk patients based on the prognostic index. High-risk patients had significantly short progression-free survival times compared to low-risk patients (A) in the discovery set (log rank test, $p<0.0001$) and (B) in the external set (log rank test, $p=0.0004$). Similarly, high-risk patients had significantly shorter overall survival times compared to low-risk patients (C) in the discovery set (log rank test, $p<0.0001$) and (D) in the external set (log rank test, $p=0.0010$).
doi:10.1371/journal.pone.0009615.g001

We further used IPA software to analyze 88 PFS-related genes from the viewpoint of molecular interaction or pathway. Top three significant networks (score>25) are shown in Figures S5-7. The network 1 included 15 of the 88 prognostic genes, and was significantly associated with IPA-defined several networks: cell death, neurological disease, and cellular assembly and organization (Figure S5). Fourteen prognostic genes were included in the network 2, which was defined as networks related to cancer, cell morphology, and renal and urological disease (Figure S6). The network 3 displayed significant interactions and interrelations between genes involved in cell-to-cell signaling and interaction, hematological system development and function, and immune cell trafficking (Figure S7). In the 88 genes, we found several genes interacting with *SRC* or *MYC* (Figure S6), each of which was reported as a representative gene in oncogenic pathways of ovarian cancer [25,27].

Discussion

In this study, we identified the prognostic index for predicting PFS time in patients with advanced-stage serous ovarian cancer treated with platinum/taxane-based adjuvant chemotherapy across two types of microarray expression data from the present discovery set and publicly available external set by using the ridge regression Cox model. The significant correlation between our prognostic index and OS time was also indicated in the two independent datasets.

In expression microarray analysis, there is a so-called "curse of dimensionality" problem that the number of genes is much larger than the number of samples. To improve the reliability of a gene expression-based prognostic model, it is necessary to avoid overfitting to the dataset, and to confirm the reproducibility of the predictive ability in external independent datasets [28]. Until now, several bioinformatics approaches have been proposed to establish a model for survival prediction using microarray data [18,29]. Bøvelstad *et al.* [18] recently examined the prediction performance of the following seven methods: univariate selection, forward stepwise selection, principal components regression, supervised principal components regression, partial least squares regression, ridge regression and the lasso using three microarray datasets [Dutch breast cancer data (n = 295), diffuse large B-cell lymphoma data (n = 240), and Norway/Stanford breast cancer data (n = 115)] [7,30–32]. They concluded that the univariate Cox model alone was insufficient for predicting survival and that the ridge regression

Table 3. Univariate and multivariate Cox's proportional hazard model analysis of prognostic factors for progression-free survival.

	Univariate analysis		Multivariate analysis	
Prognostic factor	**Hazard ratio (95%CI*)**	***p*-value**	**Hazard ratio (95%CI)**	***p*-value**
A) Present study (n = 110)				
Age	0.99 (0.97–1.01)	0.41	1.00 (0.99–1.02)	0.68
Stage IV (vs Stage III)	1.40 (1.05–1.81)	0.022	0.93 (0.69–1.24)	0.65
Optimal Surgery (vs not optimal)	0.57 (0.45–0.72)	<0.0001	0.73 (0.56–0.94)	0.016
Grade				
Grade2 (vs Grade1)	1.21 (0.89–1.67)	0.23	1.08 (0.78–1.50)	0.66
Grade3 (vs Grade1)	1.44 (1.07–1.98)	0.016	1.34 (0.98–1.88)	0.065
Prognostic Index				
High (vs Low)	3.95 (2.85–5.74)	<0.0001	3.80 (2.68–5.61)	<0.0001
B) Tothill's dataset [20] (n = 87)				
Age	1.01 (0.98–1.03)	0.61	1.00 (0.98–1.03)	0.82
Stage IV (vs Stage III)	1.26 (0.51–2.28)	0.55	0.83 (0.33–1.55)	0.60
Optimal Surgery (vs not optimal)	0.78 (0.62–0.99)	0.049	0.76 (0.60–0.98)	0.035
Prognostic Index				
High (vs Low)	1.62 (1.26–2.09)	0.0001	1.64 (1.27–2.13)	0.0001

*CI denotes confidence interval.

doi:10.1371/journal.pone.0009615.t003

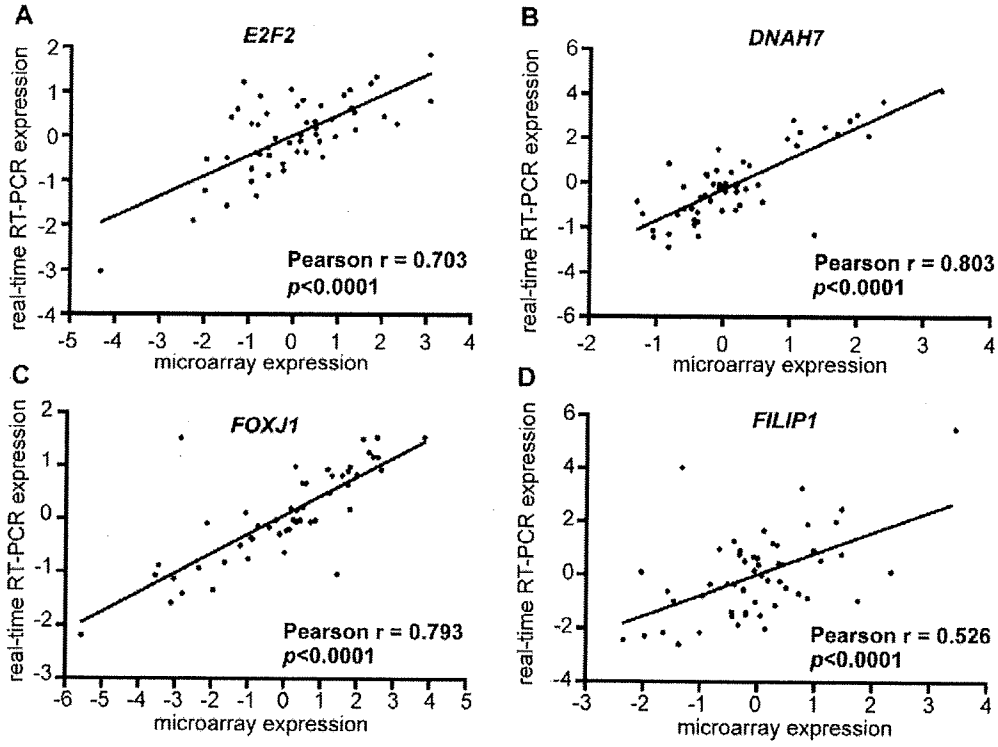

Figure 2. Validation of microarray expression data using quantitative real-time reverse transcript polymerase chain reaction (RT-PCR) analysis. There were significant correlations between microarray expression and real-time RT-PCR expression in (A) *E2F2*, (B) *DNAH7*, (C) *FOXJ1*, and (D) *FILIP1*.

doi:10.1371/journal.pone.0009615.g002

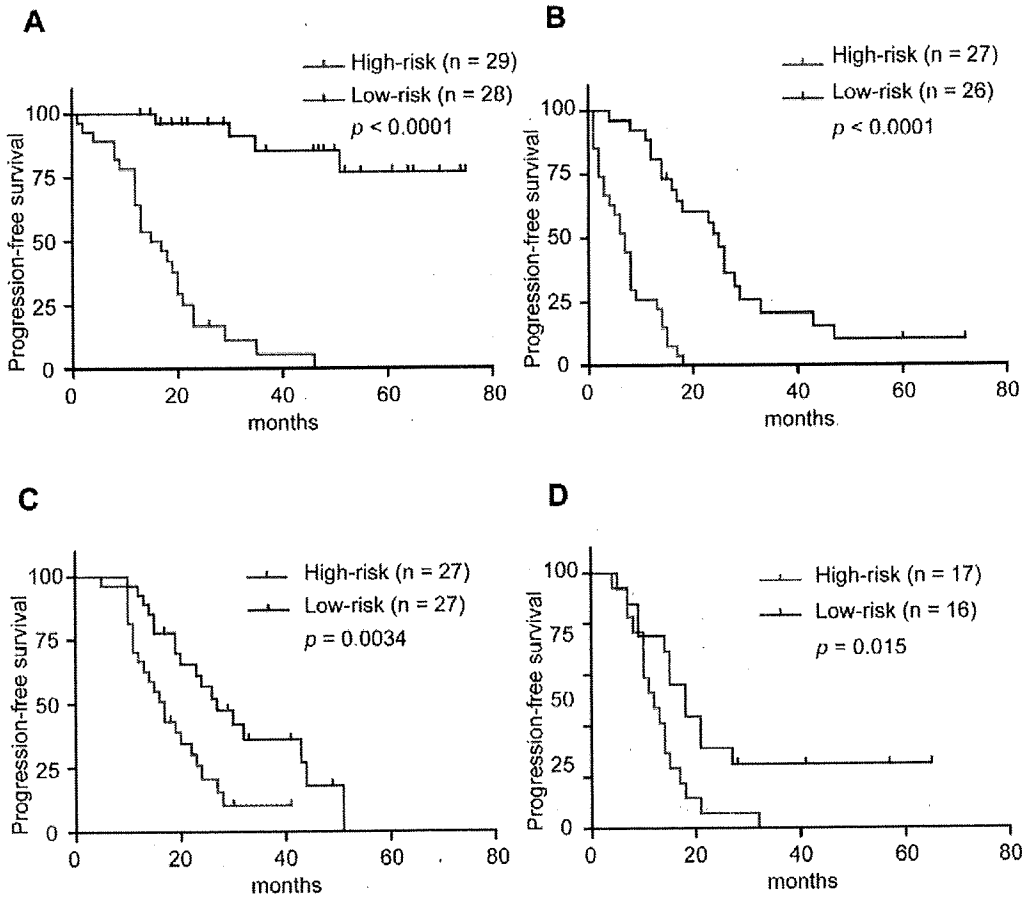

Figure 3. Prediction of prognosis in high-risk and low-risk patients based on the prognostic index after the stratification of patients according to the status of debulking surgery. High-risk patients had significantly short progression-free survival times compared to low-risk patients (A) in optimal (log rank test, $p<0.0001$) and (B) suboptimal group of discovery dataset (log rank test, $p<0.0001$). Similarly, high-risk patients had significantly shorter overall survival times compared to low-risk patients (C) in optimal (log rank test, $p=0.0034$) and (D) suboptimal group of the external dataset (log rank test, $p=0.015$).
doi:10.1371/journal.pone.0009615.g003

Cox model demonstrated the best performance in three datasets. Therefore, we used univariate Cox model only for selecting genes related to PFS time, and adjusted the regression coefficients by the ridge regression Cox model in order to increase the predictive performance of the prognostic index in our dataset.

The current study is intended to identify gene expression profile with a superior ability to predict prognosis than other clinicopathological factors. The stratification of patients with ovarian cancer according to clinicopathological prognostic factors is one of important analysis methods for the identification of highly accurate prognostic index [11]. After we stratified patients according to grade, FIGO stage, and status of debulking surgery, we investigated gene expression profile for predicting PFS time in stage III grade 2/3 serous ovarian cancer patients received optimal surgery or suboptimal surgery. However, we could find poorer predictive performance of the prognostic indices from the stratified analyses than that from the non-stratified analysis (Table S3). Besides the reduction of sample size in the discovery and external datasets after the stratification, a variety in clinical features and grading systems between the two datasets (Table S1) might influence the results from these stratified analyses. This is the main reason why we planned to identify prognostic index based on PFS-related genes in 110 advanced-stage serous ovarian cancers and then evaluate the significance of the prognostic index using multivariate analysis including grade, stage, and status of debulking surgery.

Although we enrolled ovarian cancer patients screened carefully by the following three categories: advanced-stage, histological serous-type, and platinum/taxane-based chemotherapy after primary surgery, we established no inclusion or exclusion criterion of histological grade for the enrollment as well as Crijns and colleagues did [12]. This is because a standard system for grading ovarian carcinomas is still under construction in the world, although several grading systems have been proposed for epithelial ovarian cancer [21–23,33,34]. According to the three criteria above, we recruited 110 Japanese ovarian cancer patients as a discovery set for the PFS analysis. The prognostic index for each patient was simply calculated by the ridge-regression-weighted sum of 88-gene expression values, and the prognostic power of our index was assessed using Tothill's dataset [20]. Further, subsequent stratified analysis according to debulking status, which was an independent prognostic factor in multivariate analysis of the discovery dataset, indicated that our prognostic index was associated with PFS time

Table 4. Univariate and multivariate Cox's proportional hazard model analysis of prognostic factors for overall survival.

	Univariate analysis		Multivariate analysis	
Prognostic factor	**Hazard ratio (95%CI*)**	***p*-value**	**Hazard ratio (95%CI)**	***p*-value**
A) Present study (n = 110)				
Age	1.01 (0.98–1.03)	0.56	-	-
Stage IV (vs Stage III)	1.14 (0.78–1.59)	0.49	0.75 (0.50–1.08)	0.12
Optimal Surgery (vs not optimal)	0.69 (0.50–0.92)	0.012	0.98 (0.70–1.35)	0.90
Grade				
Grade2 (vs Grade1)	1.30 (0.85–2.09)	0.23	1.23 (0.80–2.01)	0.35
Grade3 (vs Grade1)	1.68 (1.12–2.68)	0.012	1.83 (1.18–3.02)	0.0065
Prognostic Index				
High (vs Low)	2.72 (1.91–4.08)	<0.0001	2.99 (2.02–4.65)	<0.0001
B) Tothill's dataset [20] (n = 87)				
Age	1.01 (0.97–1.05)	0.73	1.00 (0.97–1.04)	0.88
Stage IV (vs Stage III)	2.13 (0.85–3.95)	0.093	1.60 (0.62–3.21)	0.28
Optimal Surgery (vs not optimal)	0.89 (0.62–1.23)	0.42	0.94 (0.66–1.37)	0.74
Prognostic Index				
High (vs Low)	1.76 (1.24–2.55)	0.0013	1.71 (1.20–2.49)	0.0029

*CI denotes confidence interval.

doi:10.1371/journal.pone.0009615.t004

independently of the debulking status. However, the sensitivity and specificity of the prognostic index for discriminating between early- and late-relapse patients were lower in Tothill's dataset than those in the discovery set. This might be caused by different backgrounds in respects of ethnicity or microarray platform. Although the differences in gene expression of cancer tissues among ethnicities have not been reported previously, several studies indicate that the proportions of clear cell and endometrioid histological types in

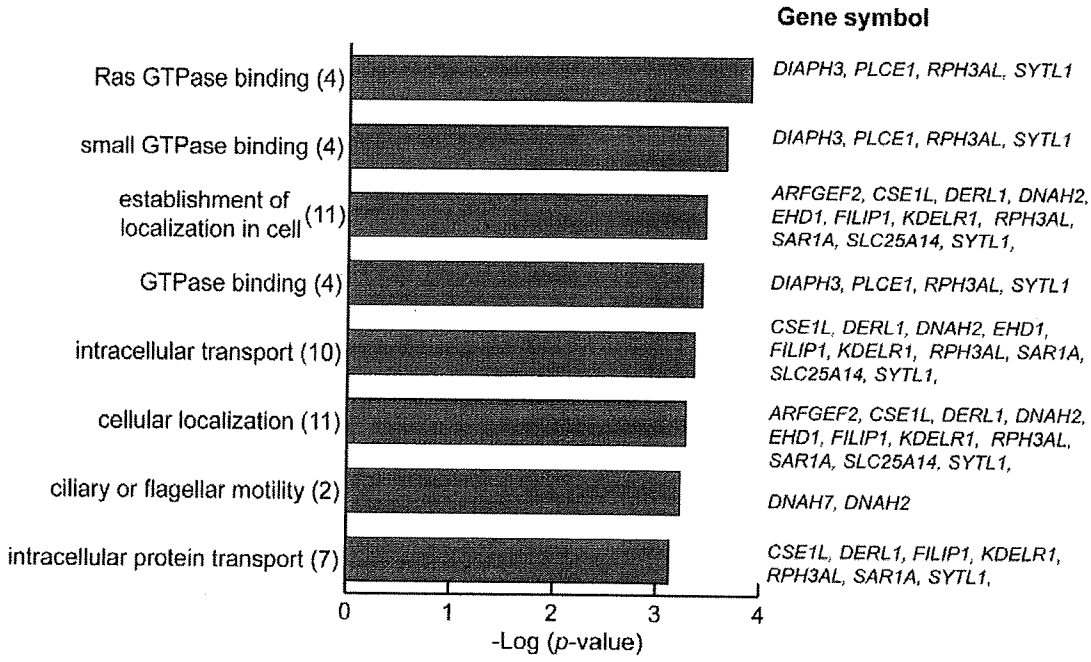

Figure 4. Biological characteristics of 88 progression-free survival-related genes. Significantly over-represented 8 gene ontology (GO) categories in GO-based profiling of 88 genes after multiple testing correction of the Benjamini–Hochberg false discovery rate method (FDR q-value≤0.10). Over-represented GO categories were identified using all genes on Agilent platform as a background set of genes for the determining p-values. The actual number of the PFS-related genes involved in each category is given in parentheses.

doi:10.1371/journal.pone.0009615.g004

epithelial ovarian cancer in Asian population are higher than those in non-Asian populations [35,36]. Recent genome-wide association study has identified a single nucleotide polymorphism at 9p22 associated with ovarian cancer risk in subjects with European ancestry but not in non-European descendants [37]. This type of differences between studies could be also attributed to genetic as well as environmental factors. In addition, we cannot rule out the possibility that the present PFS-associated classifiers with ridge-regression-based weights still have insufficient generalization properties on the external dataset due to the problem of overfitting. Therefore, we will reconsider these important issues such as between-study differences in ethnicities and microarray platforms and the overfitting problem using a larger number of microarray data from advanced-stage serous ovarian cancer patients in order to obtain better classifiers for the prediction of prognosis. And to improve the accuracy of prognostic index, development of prognostic index after the stratification of patients will be a research agenda for further study.

Interestingly, the present 88-gene prognostic index for prediction of PFS time was also significantly associated with overall survival time in both our dataset and Tothill's dataset [20]. Moreover, we examined the predictive ability of our prognostic index in Dressman's dataset [25] since patients in their dataset received longer-term follow-up than those in the above two datasets. Although Dressman's dataset (n = 119) [25] included 34 patients treated with platinum/cyclophospamide chemotherapy and 3 with single-agent platinum, the significance of this prognostic index for overall survival was still statistically supported in the longer followed-up dataset. As treatments for recurrent ovarian cancer patients remain an open area of investigation aiming to lead to survival benefit [38], our prognostic index for patient with advanced-stage serous ovarian cancer displays a potential to predict not only PFS time but also overall survival time. In the future, we may apply the prognostic indices to estimation of risk of recurrence for serous ovarian cancer patients and select a novel treatment such as dose-dense chemotherapy [39] or molecular-targeted agent for the purpose of improving prognosis of high-risk patients.

There are small number of genes overlapped between our 88 PFS-related profile and previously reported expression-profiles that were related to prognosis or sensitivity of platinum/taxane-based chemotherapy [11–15,40,41]. Konstantinopoulos *et al.* [6] have discussed that these discrepancies might be related to the use of different microarray platforms with different normalization methods and different degree of contamination by noncancerous cells in a tumor sample, as well as differences in the patient populations under study. Nevertheless, several survival-associated genes such as *E2F2* and *HLA-DMB* [42,43] are included in 88 PFS-related genes. Reimer *et al.* [42] have reported that *E2F2* is associated with grade 3 ovarian tumors and residual disease (more than 2cm in diameter) after initial surgery, and that low *E2F2* expression is significantly associated with favorable disease-free and overall survival in epithelial ovarian cancer. Callahan *et al.* [43] have recently reported that the high expression of HLA-DMB in ovarian cancer cells is correlated with increased numbers of tumor-infiltrating CD8-positive T lymphocytes, and with good prognosis in advanced-stage high-grade serous ovarian cancer.

We performed GO analysis and IPA to assess biological characteristics of PFS-related genes. GO analysis revealed the significant associations of GTPase binding, intracellular transport, and ciliary or flagellar motility with PFS (Figure 4). *PLCE1* belongs to the GTPase binding category and activates MAP kinase or ERK as shown in IPA network 3 (Figure S7). In particular, previous report indicates that *PLCE1* activates the small G protein Ras/MAP kinase signaling [44], which is one of important pathways associated with cell growth and differentiation. Intriguingly, *CSE1L* included in the intracellular transport category is involved in the regulation of multiple cellular mechanisms, proliferation, and apoptosis [45]. Tanaka *et al.* [46] have reported that *CSE1L* is associated with regulated expression of p53 target genes, and that downregulation of *CSE1L* protects cancer cell from DNA damage-induced apoptosis. *DNAH2* and *DNAH7* are components of the inner dynein arm of cilialy axonemes, and axonemal dyneins are molecular motors that drive the beating of cilia and flagella. Plotnikova *et al.* [47] have reported that loss of cilia in cancer cells may contribute to the insensitivity of cancer cells to environmental repressive signals, partly owing to derangement of cell cycle checkpoints governed by cilia and centrosomes. On the other hand, IPA analysis showed several genes interacting with *SRC* or *MYC* (Figure S6), each of which was reported as a representative gene in oncogenic pathways of ovarian cancer [25,27]. Dressman *et al.* [25] have demonstrated that Src pathway activity is associated with chemotherapy response because of a significant correlation between the activation of Src pathway and poor prognosis in patients with platinum-resistant ovarian cancer. *MYC* is a multifunctional proto-oncogene and activated in about 30% of ovarian cancer by several mechanisms [48]. Iba *et al.* [49] report that MYC expression is associated with responsiveness to platinum-based chemotherapy and with prognosis in patients with epithelial ovarian cancer. Our PFS-related profile might have potentially functional relevance to altered activities of several oncogenic pathways. Although we identified several genes whose molecular function could be linked to prognosis in ovarian cancer patients, further functional study will be necessary to clarify the biological and pathological implications of the PFS-related profile.

These results suggest that the gene expression profile could be a useful tool to predict disease progression or recurrence of advanced-stage serous ovarian cancer. To apply the gene expression profile in clinical practice, we will need to improve the predictive ability of the profile and confirm the reliability of survival profile in a prospective multi-center study. Nevertheless, the survival-related profile could provide an optimization of the clinical management and development of new therapeutic strategies for the serous ovarian cancer patients.

Materials and Methods

Tissue Samples

One hundred ten Japanese patients who were diagnosed with advanced-stage serous ovarian cancer between July 1997 and June 2008 were included in this study. Fresh-frozen samples were obtained from primary tumor tissues during primary debulking surgery prior to chemotherapy. All patients with advanced-stage serous ovarian cancer were treated with platinum/taxane-based chemotherapy after surgery. In principle, patients were seen every 1 to 3 months for the first 2 years. Thereafter, follow-up visits had an interval of 3 to 6 months in the third to fifth year, and 6 to 12 months in the sixth to tenth year. At every follow-up visits, general physical and gynecologic examination were performed. CA125 serum levels were routinely determined. Staging of the disease was assessed according to the criteria of the International Federation of Gynecology and Obstetrics (FIGO) [19]. Optimal debulking surgery was defined as ≤1cm of gross residual disease. The histological characteristics of surgically resected specimens were assessed on formalin-fixed and paraffin-embedded hematoxylin and eosin sections by two or three gynecological pathologists belonging to the Japanese Society of Pathology at each institute,

and frozen tissues containing more than 80% of tumor cells upon histological evaluation were used for RNA extraction. In this study, the degree of histological differentiation is determined according to the increase in the proportion of solid growth within the adenocarcinoma as follows: grade 1, less than 5% solid growth; grade 2, 6-50% solid growth; grade 3, over 50% solid growth based on grading system proposed by Japan Society of Gynecologic Oncology.

PFS time was calculated as the interval from primary surgery to disease progression or recurrence. Based on standard Response Evaluation Criteria In Solid Tumors (RECIST) guidelines [50], disease progression was defined as at least 20% increase in the sum of the longest diameters of all target lesions or as the appearance of one or more new lesions and/or unequivocal progression existing non-target lesions. Overall survival time was calculated as the interval from primary surgery to the death due to ovarian cancer. This study was approved by the institutional ethics review board at Niigata University (No. 239, 282, 285, and 318), Niigata Cancer Center Hospital (No. 25), Jichi Medical University (G07-01), Kagoshima City Hospital (H19-21), Hiroshima University (Hi-11), Nagasaki University (080509), Kumamoto University (No. 309), and Tokai University (07I-29). All patients provided written informed consent for the collection of samples and subsequent analysis.

Microarray Experiments

Total RNA was extracted from tissue samples as previously described [17]. Five hundred nanograms of total RNA were converted into labeled cRNA with nucleotides coupled to a cyanine 3-CTP (Cy3) (PerkinElmer, Boston, MA, USA) using the Quick Amp Labeling Kit, one-color (Agilent Technologies). Cy3-labeled cRNA (1.65 μg) was hybridized for 17 hours at 65°C to an Agilent Whole Human Genome Oligo Microarray, which carries 60-mer probes to more than 40,000 human transcripts. The hybridized microarray was washed and then scanned in Cy3 channel with the Agilent DNA Microarray Scanner (model G2565AA). Signal intensity per spot was generated from the scanned image using Feature Extraction Software version 9.1 (Agilent Technologies) in the default settings. Spots that did not pass quality control procedures were flagged as "Absent". The MIAME-compliant microarray data were deposited into the Gene Expression Omnibus data repository (accession number GSE17260).

Microarray Data Analysis

We analyzed our dataset as a "discovery set" and the publicly available dataset as an "external dataset". Considering differences in microarray platforms, we selected common genes between the Agilent Whole Human Genome Oligo Microarray and Affymetrix Human Genome U133 Plus 2.0 Array, which was the platform in an external dataset (GSE9891) [20].

Data normalization was performed in GeneSpring GX 10 (Agilent Technologies) as follows: (i) Threshold raw signals were set to 1.0, (ii) 75th percentile normalization was chosen as normalized algorithm, (iii) Baseline was transformed to median of all samples. Furthermore, the expression level was normalized by Z-transformation (the mean expression was set to 0 and standard deviation to 1 for each gene in each dataset). In our dataset, 18,178 probes with expression levels marked as "Present" in all microarrays were used to remove missing and uncertain signals on gene expression.

The PFS-related genes from the 18,178 probes were identified by univariate Cox proportional hazard analysis, followed by a ridge regression, a penalized Cox regression analysis for survival prediction (Figure S2). We first identified 97 probes with expression levels correlating with the PFS time determined using the univariate Cox proportional hazard model ($p<0.01$). In case of multiple probes representing a given gene (so-called multiple tagged gene) in microarrays, only the probe with the largest magnitude (i.e., sum of the squares of per-individual expression values) was extracted as a representative probe for the gene [24]. To avoid the problem of overfitting, ridge regression extension of the multivariate Cox model was employed [18]. The ridge regression shrinks regression coefficients (β) of genes in multivariate Cox model by imposing a penalty on squared values of the coefficients, and is able to handle the problem of having larger number of expression values than individuals in an appropriate way [30]. We estimated regression coefficients of the prognostic genes by the ridge regression Cox model using M-files (available at http://www.med.uio.no/imb/stat/bmms/software/microsurv/) for MATLAB (Mathworks, Natick, MA, USA). Using 10-fold cross-validation, we obtained regression coefficients with optimal penalty parameter for the penalized Cox model, and calculated a prognostic index for each patient as defined by

$$\text{Prognostic index} = \sum_{i=1}^{88} \beta_i \times X_i \qquad (1)$$

where β_i is the estimated regression coefficient of each gene in discovery dataset under ridge regression multivariate Cox model and X_i is the Z-transformed expression value of each gene [18]. The estimated regression coefficient of each PFS-related gene given by ridge regression in the discovery set was also applied to calculate a prognostic index for each patient in external dataset using the equation above. We classified all patients into the two groups (high- and low-risk groups) by the median of the prognostic index in discovery set [9]. PFS between high- and low-risk groups was compared using Kaplan-Meier curves and the log rank test using GraphPad PRISM version 4.0 (GraphPad Software, San Diego, CA, USA). Furthermore, We then evaluated the prognostic index in the multivariate Cox proportional hazard model using JMP version 6 (SAS Institute, Cary, NC, USA). We also examined the discrimination performance of the prognostic index between early and late relapse in patients by plotting a receiver operating characteristic (ROC) curve for each dataset (JMP). Because 18 months is the median PFS time for advanced-stage ovarian cancer patients treated with cisplatin-paclitaxel [1], we used 18 months as the cut-off between early and late relapse. We performed ROC curve analysis for our prognostic index in only patients with follow-up for more than 18 months (Discovery set 103 samples; External dataset 84 samples).

To investigate the biological functions of PFS-related gene expression profiles, we used GO Ontology Browser, embedded in GeneSpring GX [17,51]. The GO Ontology Browser was used to analyze which categories of gene ontology were statistically overrepresented among the gene list obtained. Statistical significance was determined by Fisher's exact test, followed by multiple testing corrections by the Benjamini and Hochberg false discovery rate (FDR) method [26]. Furthermore, we tried to explore molecular interaction networks among the PFS-related genes using Ingenuity Pathway Analysis (IPA) [17].

Quantitative Real-Time Reverse Transcription Polymerase Chain Reaction (RT-PCR) Analysis

Real-time PCR was performed on *E2F2* (Hs00231667_m1, Applied Biosystems), *FOXJ1* (Hs00230964_m1, Applied Biosystems), *DNAH7* (Hs01022427_m1, Applied Biosystems), and *FILIP1* (Hs00325074_m1, Applied Biosystems) for a subset of serous

ovarian cancer (n = 53) as previously described [17]. The relative quantification method [52] was used to measure the amounts of the respective genes in serous ovarian cancer samples, normalized to *ACTB* (Hs99999903_m1, Applied Biosystems) and *TBP* (Hs99999910_m1, Applied Biosystems).

Evaluation of PFS-Related Genes in the External Dataset

To confirm whether our expression profile could predict prognosis of serous ovarian cancer patients in an independent data set, we selected to use publicly available microarray data (GSE9891) only because the data also disclosed individual clinical characteristics including PFS time. We examined clinical information of these dataset using supplementary data [20]. From this original dataset (n = 285), we selected 87 samples that were (i) diagnosed as advanced-stage serous adenocarcinoma, (ii) treated by platinum/taxane-based chemotherapy, (iii) obtained from primary lesion, and (iv) followed-up for more than 12 months (Table S1). Their samples are histologically graded by Silverberg classification [22] whose grading system is different from that in this study.

Supporting Information

Figure S1 Kaplan-Meier survival curves between 110 patients in this dataset and 87 in Tothill's dataset.
Found at: doi:10.1371/journal.pone.0009615.s001 (0.24 MB TIF)

Figure S2 Analytical process to develop a prognostic index for predicting survival.
Found at: doi:10.1371/journal.pone.0009615.s002 (0.48 MB TIF)

Figure S3 Assessment of the sensitivity and specificity of 88-gene prognostic index using receiver-operating characteristic (ROC) curves. When early relapse is positive in the analysis, the area under ROC curve to distinguish early-relapse patients with less than 18 months of progression-free survival times from late-relapse patients was 0.959 and 0.674 in (A) discovery set (early, n = 54; late, n = 49) and in (B) external set (early, n = 45; late, n = 39), respectively.
Found at: doi:10.1371/journal.pone.0009615.s003 (0.42 MB TIF)

Figure S4 Appling PFS-related gene expression profile to Dressman's dataset [25]. (A) Multivariate analysis showed a significant association of overall survival with the prognostic index estimated using the 88-gene linear combination model with the ridge regression coefficients from the present discovery set in Dresssman's dataset (HR, 1.51; 95% CI, 1.19–1.93, p = 0.0008) (B) Kaplan-Meier survival curves and the log rank test showed that high-risk patients had shorter overall survival compared to low-risk patients (median survival, 31 and 87 months for high- and low-risk patients, respectively; p = 0.0008).
Found at: doi:10.1371/journal.pone.0009615.s004 (0.23 MB TIF)

Figure S5 Molecular interaction networks of 88 progression-free survival-related genes using Ingenuity Pathway Analysis (IPA) software. The prognostic genes incorporated into the respective networks were marked as gray-colored.
Found at: doi:10.1371/journal.pone.0009615.s005 (2.42 MB TIF)

Figure S6 Molecular interaction networks of 88 progression-free survival-related genes using Ingenuity Pathway Analysis (IPA) software. The prognostic genes incorporated into the respective networks were marked as gray-colored.
Found at: doi:10.1371/journal.pone.0009615.s006 (1.68 MB TIF)

Figure S7 Molecular interaction networks of 88 progression-free survival-related genes using Ingenuity Pathway Analysis (IPA) software. The prognostic genes incorporated into the respective networks were marked as gray-colored.
Found at: doi:10.1371/journal.pone.0009615.s007 (1.82 MB TIF)

Table S1 Clinical characteristics of advanced-stage serous ovarian cancer patients in Tothill's dataset [20] (n = 87).
Found at: doi:10.1371/journal.pone.0009615.s008 (0.04 MB DOC)

Table S2 Univariate and multivariate Cox's proportional hazard model analysis of prognostic factors for progression-free survival.
Found at: doi:10.1371/journal.pone.0009615.s009 (0.04 MB DOC)

Table S3 Univariate Cox's proportional hazard model analysis of prognostic index for progression-free survival in the two datasets.
Found at: doi:10.1371/journal.pone.0009615.s010 (0.04 MB DOC)

Acknowledgments

We thank tissue donors and supporting medical staff for making this study possible. We are grateful to C. Seki and A. Yukawa for their technical assistance.

Author Contributions

Conceived and designed the experiments: KY AT TY H KT. Performed the experiments: KY AT. Analyzed the data: KY AT. Contributed reagents/materials/analysis tools: KY TY SK HF MS YO MH KS HF YK KK HM HT HK H KT. Wrote the paper: KY AT TY H KT.

References

1. McGuire WP, Hoskins WJ, Brady MF, Kucera PR, Partridge EE, et al. (1996) Cyclophosphamide and cisplatin compared with paclitaxel and cisplatin in patients with stage III and stage IV ovarian cancer. N Engl J Med 334: 1–6.
2. Piccart MJ, Bertelsen K, James K, Cassidy J, Mangioni C, et al. (2000) Randomized intergroup trial of cisplatin-paclitaxel versus cisplatin-cyclophosphamide in women with advanced epithelial ovarian cancer: three-year results. J Natl Cancer Inst 92: 699–708.
3. Cannistra SA (2004) Cancer of the ovary. N Engl J Med 351: 2519–29.
4. du Bois A, Reuss A, Pujade-Lauraine E, Harter P, Ray-Coquard I, et al. (2009) Role of surgical outcome as prognostic factor in advanced epithelial ovarian cancer: a combined exploratory analysis of 3 prospectively randomized phase 3 multicenter trials: by the Arbeitsgemeinschaft Gynaekologische Onkologie Studiengruppe Ovarialkarzinom (AGO-OVAR) and the Groupe d'Investigateurs Nationaux Pour les Etudes des Cancers de l'Ovaire (GINECO). Cancer 115: 1234–44.
5. Winter WE, 3rd, Maxwell GL, Tian C, Carlson JW, Ozols RF, et al. (2007) Prognostic factors for stage III epithelial ovarian cancer: a Gynecologic Oncology Group Study. J Clin Oncol 25: 3621–7.
6. Konstantinopoulos PA, Spentzos D, Cannistra SA (2008) Gene-expression profiling in epithelial ovarian cancer. Nat Clin Pract Oncol 5: 577–87.
7. van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. Nature 415: 530–6.
8. Motoori M, Takemasa I, Yano M, Saito S, Miyata H, et al. (2005) Prediction of recurrence in advanced gastric cancer patients after curative resection by gene expression profiling. Int J Cancer 114: 963–8.
9. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, et al. (2007) A five-gene signature and clinical outcome in non-small-cell lung cancer. N Engl J Med 356: 11–20.
10. Schramm A, Schulte JH, Klein-Hitpass L, Havers W, Sieverts H, et al. (2005) Prediction of clinical outcome and biological characterization of neuroblastoma by expression profiling. Oncogene 24: 7902–12.
11. Bonome T, Levine DA, Shih J, Randonovich M, Pise-Masison CA, et al. (2008) A gene signature predicting for survival in suboptimally debulked patients with ovarian cancer. Cancer Res 68: 5478–86.
12. Crijns AP, Fehrmann RS, de Jong S, Gerbens F, Meersma GJ, et al. (2009) Survival-related profile, pathways, and transcription factors in ovarian cancer. PLoS Med 6: e24.
13. Denkert C, Budczies J, Darb-Esfahani S, Györffy B, Sehouli J, et al. (2009) A prognostic gene expression index in ovarian cancer - validation across different independent data sets. J Pathol 218: 273–80.

14. Hartmann LC, Lu KH, Linette GP, Cliby WA, Kalli KR, et al. (2005) Gene expression profiles predict early relapse in ovarian cancer after platinum-paclitaxel chemotherapy. Clin Cancer Res 11: 2149–55.
15. Spentzos D, Levine DA, Ramoni MF, Joseph M, Gu X, et al. (2004) Gene expression signature with independent prognostic significance in epithelial ovarian cancer. J Clin Oncol 22: 4700–10.
16. Agarwal R, Kaye SB (2006) Expression profiling and individualization of treatment for ovarian cancer. Curr Opin Pharmacol 6: 345–9.
17. Yoshihara K, Tajima A, Komata D, Yamamoto T, Kodama S, et al. (2009) Gene expression profiling of advanced-stage serous ovarian cancers distinguishes novel subclasses and implicates ZEB2 in tumor progression and prognosis. Cancer Sci 100: 1421–8.
18. Bøvelstad HM, Nygård S, Størvold HL, Aldrin M, Borgan O, et al. (2007) Predicting survival from microarray data–a comparative study. Bioinformatics 23: 2080–7.
19. FIGO Cancer Committee. (1986) Staging Announcement: FIGO Cancer Committee. Gynecol Oncol 25: 383–5.
20. Tothill RW, Tinker AV, George J, Brown R, Fox SB, et al. (2008) Novel molecular subtypes of serous and endometrioid ovarian cancer linked to clinical outcome. Clin Cancer Res 14: 5198–208.
21. International Federation of Gynecology and Obstetrics (1971) Classification and staging of malignant tumours in the female pelvis. Acta Obstet Gynecol Scand 50: 1–7.
22. Silverberg SG (2000) Histopathologic grading of ovarian carcinoma: a review and proposal. Int J Gynecol Pathol 19: 7–15.
23. Kommoss S, Schmidt D, Kommoss F, Hedderich J, Harter P, et al. (2009) Histological grading in a large series of advanced stage ovarian carcinomas by three widely used grading systems: consistent lack of prognostic significance. A translational research subprotocol of a prospective randomized phase III study (AGO-OVAR 3 protocol). Virchows Arch 454: 249–56.
24. Woo HG, Park ES, Cheon JH, Kim JH, Lee JS, et al. (2008) Gene expression-based recurrence prediction of hepatitis B virus-related human hepatocellular carcinoma. Clin Cancer Res 14: 2056–64.
25. Dressman HK, Berchuck A, Chan G, Zhai J, Bild A, et al. (2007) An integrated genomic-based approach to individualized treatment of patients with advanced-stage ovarian cancer. J Clin Oncol 25: 517–25.
26. Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. J R Statist Soc B 57: 289–300.
27. Bild AH, Yao G, Chang JT, Wang Q, Potti A, et al. (2006) Oncogenic pathway signatures in human cancers as a guide to targeted therapies. Nature 439: 353–7.
28. Dupuy A, Simon RM (2007) Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 99: 147–57.
29. Bair E, Tibshirani R (2004) Semi-supervised methods to predict patient survival from gene expression data. PLoS Biol 2: e108.
30. van Houwelingen HC, Bruinsma T, Hart AA, Van't Veer LJ, Wessels LF (2006) Cross-validated Cox regression on microarray gene expression data. Stat Med 25: 3201–16.
31. Rosenwald A, Wright G, Chan WC, Connors JM, Campo E, et al. (2002) The use of molecular profiling to predict survival after chemotherapy for diffuse large-B-cell lymphoma. N Engl J Med 346: 1937–47.
32. Sorlie T, Tibshirani R, Parker J, Hastie T, Marron JS, et al. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100: 8418–23.
33. Tawassoli FA, Devilee P (2003) Pathology and Genetics. Tumours of the Breast and Female Genital Organs. IARC Press, Lyon.
34. Malpica A, Deavers MT, Lu K, Bodurka DC, Atkinson EN, et al. (2004) Grading ovarian serous carcinoma using a two-tier system. Am J Surg Pathol 28: 496–504.
35. Goodman MT, Howe HL, Tung KH, Hotes J, Miller BA, et al. (2003) Incidence of ovarian cancer by race and ethnicity in the United States, 1992–1997. Cancer 97(10 Suppl): 2676–85.
36. McGuire V, Jesser CA, Whittemore AS (2002) Survival among U.S. women with invasive epithelial ovarian cancer. Gynecol Oncol 84: 399–403.
37. Song H, Ramus SJ, Tyrer J, Bolton KL, Gentry-Maharaj A, et al. (2009) A genome-wide association study identifies a new ovarian cancer susceptibility locus on 9p22.2. Nat Genet 41: 996–1000.
38. Ozols RF (2005) Treatment goals in ovarian cancer. Int J Gynecol Cancer 15 Suppl 1: 3–11.
39. Katsumata N, Yasuda M, Takahashi F, Isonishi S, Jobo T, et al. (2009) Dose-dense paclitaxel once a week in combination with carboplatin every 3 weeks for advanced ovarian cancer: a phase 3, open-label, randomised controlled trial. Lancet 374: 1331–8.
40. Berchuck A, Iversen ES, Luo J, Clarke JP, Horne H, et al. (2009) Microarray analysis of early stage serous ovarian cancers shows profiles predictive of favorable outcome. Clin Cancer Res 15: 2448–55.
41. Helleman J, Jansen MP, Span PN, van Staveren IL, Massuger LF, et al. (2006) Molecular profiling of platinum resistant ovarian cancer. Int J Cancer 118: 1963–71.
42. Reimer D, Sadr S, Wiedemair A, Stadlmann S, Concin N, et al. (2007) Clinical relevance of E2F family members in ovarian cancer–an evaluation in a training set of 77 patients. Clin Cancer Res 13: 144–51.
43. Callahan MJ, Nagymanyoki Z, Bonome T, Johnson ME, Litkouhi B, et al. (2008) Increased HLA-DMB expression in the tumor epithelium is associated with increased CTL infiltration and improved prognosis in advanced-stage serous ovarian cancer. Clin Cancer Res 14: 7667–73.
44. Lopez I, Mak EC, Ding J, Hamm HE, Lomasney JW (2001) A novel bifunctional phospholipase c that is regulated by Galpha 12 and stimulates the Ras/mitogen-activated protein kinase pathway. J Biol Chem 276: 2758–65.
45. Behrens P, Brinkmann U, Wellmann A (2003) CSE1L/CAS: its role in proliferation and apoptosis. Apoptosis 8: 39–44.
46. Tanaka T, Ohkubo S, Tatsuno I, Prives C (2007) hCAS/CSE1L associates with chromatin and regulates expression of select p53 target genes. Cell 130: 638–50.
47. Plotnikova OV, Golemis EA, Pugacheva EN (2008) Cell cycle-dependent ciliogenesis and cancer. Cancer Res 68: 2058–61.
48. Darcy KM, Brady WE, Blancato JK, Dickson RB, Hoskins WJ, et al. (2009) Prognostic relevance of c-MYC gene amplification and polysomy for chromosome 8 in suboptimally-resected, advanced stage epithelial ovarian cancers: a Gynecologic Oncology Group study. Gynecol Oncol 114: 472–9.
49. Iba T, Kigawa J, Kanamori Y, Itamochi H, Oishi T, et al. (2004) Expression of the c-myc gene as a predictor of chemotherapy response and a prognostic factor in patients with ovarian cancer. Cancer Sci 95: 418–23.
50. Therasse P, Arbuck SG, Eisenhauer EA, Wanders J, Kaplan RS, et al. (2000) New guidelines to evaluate the response to treatment in solid tumors. European Organization for Research and Treatment of Cancer, National Cancer Institute of the United States, National Cancer Institute of Canada. J Natl Cancer Inst 92: 205–16.
51. Okada H, Tajima A, Shichiri K, Tanaka A, Tanaka K, et al. (2008) Genome-wide expression of azoospermia testes demonstrates a specific profile and implicates ART3 in genetic susceptibility. PLoS Genet 4: e26.
52. Livak KJ, Schmittgen TD (2001) Analysis of Relative Gene Expression Data Using Real-Time Quantitative PCR and the $2^{-\Delta\Delta CT}$ Method. Methods 25: 402–8.

JOURNAL OF VIROLOGY, July 2009, p. 6922–6928 Vol. 83, No. 13
0022-538X/09/$08.00+0 doi:10.1128/JVI.02674-08

Genetic Analysis of Hepatitis C Virus with Defective Genome and Its Infectivity in Vitro▽

Kazuo Sugiyama,[1]* Kenji Suzuki,[2] Takahide Nakazawa,[3] Kenji Funami,[1] Takayuki Hishiki,[4] Kazuya Ogawa,[4] Satoru Saito,[5] Kumiko W. Shimotohno,[2] Takeshi Suzuki,[2] Yuko Shimizu,[1] Reiri Tobita,[6] Makoto Hijikata,[7] Hiroshi Takaku,[6] and Kunitada Shimotohno[1,4]

Center for Integrated Medical Research, Keio University, Shinjuku-ku, Shinanomachi 35, Tokyo 160-8582, Japan[1]; Division of Basic Biological Sciences, Faculty of Pharmacy, Keio University, Tokyo 105-8512, Japan[2]; Department of Gastroenterology, Internal Medicine, Kitasato University East Hospital, Kanagawa 228-8520, Japan[3]; Research Institute, Chiba Institute of Technology, Chiba 275-0016, Japan[4]; Yokohama City University Hospital, Kanagawa 236-0004, Japan[5]; Department of Life and Environmental Sciences, Chiba Institute of Technology, Chiba 275-0016, Japan[6]; and Institute for Virus Research, Kyoto University, Kyoto 606-8507, Japan[7]

Received 29 December 2008/Accepted 6 April 2009

Replication and infectivity of hepatitis C virus (HCV) with a defective genome is ambiguous. We molecularly cloned 38 HCV isolates with defective genomes from 18 patient sera. The structural regions were widely deleted, with the 5′ untranslated, core, and NS3-NS5B regions preserved. All of the deletions were in frame, indicating that they are translatable to the authentic terminus. Phylogenetic analyses showed self-replication of the defective genomes independent of full genomes. We generated a defective genome of chimeric HCV to mimic the defective isolate in the serum. By using this, we demonstrated for the first time that the defective genome, as it is circulating in the blood, can be encapsidated as an infectious particle by *trans* complementation of the structural proteins.

Viruses with a deletion mutation in their genome have been identified as defective interfering (DI) particles for many virus species (1, 3, 9, 16). Part of the DI virus genome is deleted, but regions indispensable for replication and packaging are preserved. Most DI viruses occur spontaneously in the course of cell culture infected with a high titer of wild-type viruses. Hepatitis C virus (HCV) with a defective genome has been found in liver and serum specimens of some HCV patients (4, 8, 15). HCV has a plus-strand RNA genome that encodes the viral core, E1, E2, and p7 structural proteins and NS2, NS3, NS4A, NS4B, NS5A, and NS5B nonstructural proteins (10). According to the reports, the deletions have been found mainly in the structural region and most of the deletions are in frame, but some deletions are out of frame (4), raising questions about whether the defective HCV genome is merely a by-product of a full genome or a self-replicating genome and whether it can be encapsidated into an infectious virus particle.

In the present study, we molecularly cloned 38 HCV isolates with defective genomes from HCV patient sera to address these questions by genetic analyses and infection experiments. As long as we explored, all of the deletions were in frame, indicating the potential to support translation from the authentic initiation codon to the termination codon, although the structural region was widely deleted, as reported previously. Phylogenetic analyses evidenced self-replication of the defective genomes independent of full genomes. We demonstrated for the first time, by *trans* complementation experiments, that the defective genome, as it is circulating in the blood, can be encapsidated as an infectious particle, designated HCV_{CCD}.

First, to amplify HCV cDNAs in 21 serum specimens from 18 HCV patients (genotype 1b), we performed three sets of long-distance reverse transcription (RT)-PCRs flanking (i) the 5′ untranslated region (UTR) to the 5′ part of the NS3 region, (ii) the remaining part of the NS3 region to the end of NS5B, and (iii) the 5′ UTR to the end of the NS5B region (Fig. 1A). The specimens were collected with informed consent. cDNA was synthesized with RNase H-deficient reverse transcriptase Superscript III (Invitrogen, Carlsbad, CA) at a higher temperature (55°C) to reduce template switching and mispriming. PCRs were performed in a (hemi)nested manner with high-fidelity polymerase KOD plus or KOD FX (Toyobo, Osaka, Japan) as described previously (5). For some target nucleotide positions, a mixture of two or three primers was used to reduce mismatches due to sequence heterogeneity (Table 1). Of the 21 specimens examined, representative results are shown in Fig. 1. An amplicon of the 5′ UTR-NS3 region of the predicted size (ca. 3.7 kb) was detected in all specimens (18/18), and representative results are shown in Fig. 1B. In addition, a shorter amplicon suggestive of a defective HCV genome was simultaneously present in four specimens from 1 (R4) of 12 cases of clinically mild hepatitis and from 3 (T5, K3, and K4-pre) of 6 cases of active hepatitis (clinical data not shown). Defective genomes were found in the patients with relatively higher copy numbers of HCV RNA (>8.1×10^5 copies/ml in the 5′ UTR, Table 2), suggesting that the coexistence of a defective genome is related to hepatitis severity. The authentic-size amplicon was poorly detected when coexisting with a defective HCV genome shorter than 2 kb (T5 and K3), presumably because of preferential amplification of the shorter amplicon. A shorter amplicon was not detected for the NS3-

* Corresponding author. Mailing address: Center for Integrated Medical Research, Keio University, Shinjuku-ku, Shinanomachi 35, Tokyo 160-8582, Japan. Phone: 81-3-3353-1211. Fax: 81-47-478-0527. E-mail: sygiyamkz@a8.keio.jp.

▽ Published ahead of print on 15 April 2009.

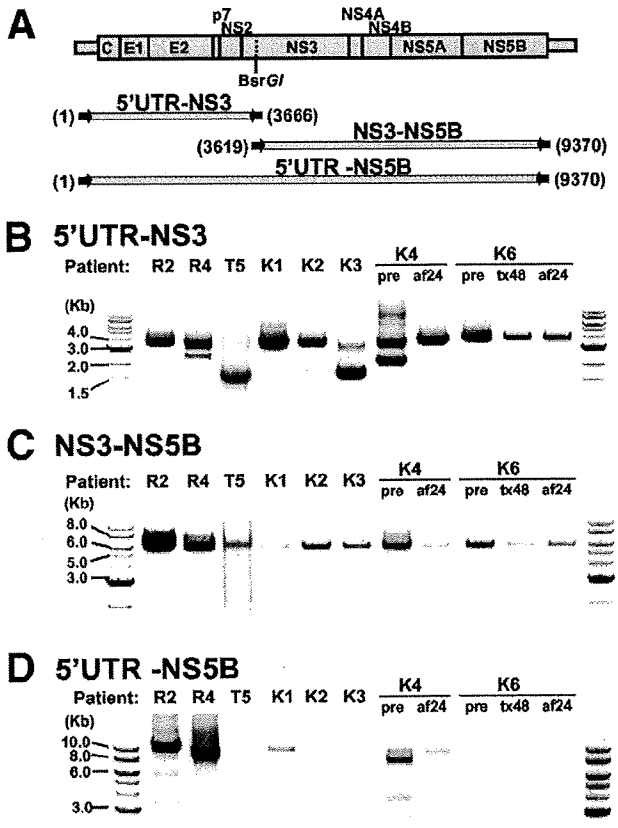

FIG. 1. Representative results of long-distance RT-PCRs for serum HCV. (A) The three sets of long-distance RT-PCR used: 5′ UTR-NS3 (5′ UTR to the 5′ part of the NS3 region), NS3-NS5B (the remaining part of the NS3 region to the end of NS5B), and 5′ UTR-NS5B (5′ UTR to the end of the NS5B region). The nucleotide positions of the 5′ and 3′ ends of each amplicon are indicated in parentheses. PCR products were electrophoresed and stained with ethidium bromide. Results of the representative 11 specimens (eight patients) are shown for 5′ UTR-NS3 (B), NS3-NS5B (C), and 5′ UTR-NS5B (D). Serum specimens were collected from patients K4 and K6 before interferon treatment (pre), at the end of the full 48-week treatment period (tx48), and 24 weeks after the full treatment period (af24). DNA molecular size markers are at both sides of panels B to D.

NS5B region, while the amplicon of the predicted size (ca. 5.6 kb) was detected in all of the specimens, albeit with various efficiencies (Fig. 1C). The 5′ UTR-NS5B region, which covers almost the whole genome, was then amplified, and an amplicon of the predicted size (9.4 kb) was detected in 10 specimens from nine patients (5 specimens in Fig. 1D). Of the successfully amplified specimens, two (R4 and K4-pre) also contained a shorter amplicon, in accordance with the results of the 5′ UTR-NS3 PCR. The NS3-NS5B region is essential for autonomous replication of HCV as an RNA replicon in vitro (5–7). It has been shown that NS5A is the only nonstructural protein that can *trans* complement HCV replication (13). They used a nonadaptive mutation of NS5A as a replication-incompetent NS5A protein instead of a deletion mutant protein. Thus, we speculate that deletion of the NS3-NS5B region cannot be complemented in *trans*. Intriguingly, the shorter amplicon was not detected after full-term interferon treatment in patient K4 (K4-af24), although it was detected prior to treatment (K4-pre) (Fig. 1B and D). The possible reasons for this are that (i) the defective genome disappeared naturally, (ii) packaging of the defective genome by the helper virus was impeded by an unknown mechanism of interferon, or (iii) replication of the defective genome is preferentially inhibited by the interferon pathway. Further studies are needed to reveal the effect of a defective HCV genome on the pathogenesis and treatment of HCV.

A total of 38 isolates with defective HCV genomes were molecularly cloned into plasmid vector pASGT (unpublished data) from the shorter amplicons of the 5′ UTR-NS3 PCR from four serum specimens (R4, T5, K3, and K4-pre in Fig. 1B) at the AscI and BsrGI restriction sites. The nucleotide sequences were determined with an autosequencer (3730 DNA analyzer; Applied Biosystems, Foster City, CA). Sequence analyses revealed that the structural region was widely deleted in all of the defective isolates and that the deletion ranges were quite diverse among the isolates (extending up to the NS2 region) (Fig. 2A). In contrast, the 5′ UTR and core regions were constantly preserved, suggesting that these regions, as well as the NS3-NS5B region, are indispensable for the production of HCV with a defective genome. Intriguingly, defective genomes with different deletion patterns coexisted in single specimens from two patients (three patterns in patient K3 and four patterns in patient K4-pre). Moreover, two deletions in a single genome were observed in five isolates from patient R4 (isolate R4S-5). As many as three deletions in a single genome were observed in the isolate from patient K3 (e.g., isolate K3S-15), in which two small deletions resulted in two tiny residual fragments. Such diversity in deletion ranges indicates flexibility of the remaining structural region for the replication of defective HCV genomes. Nevertheless, all of the deletions identified in the 38 isolates were in frame (Fig. 2B), implying that these defective HCV genomes have the potential for translation from the core to the authentic end of NS5B without a frameshift.

To determine the ratio of defective to full genomes, we performed quantitative PCRs targeting a relatively conserved E2 sequence, which is commonly deleted in the defective genomes, with primers listed in Table 1. Calculation of the 5′ UTR/E2 ratio, which must theoretically be 1 without the existence of the defective genome, showed higher values (1.7 to 2.45) in specimens containing the defective genomes (R4, K3, and K4pre in Table 2), indicating that the defective genome level in serum is 0.7 to 1.45 times the full genome level. However, to clarify the impact of defective genomes on pathogenesis and their effect on the treatment of HCV, accumulation of more data is needed.

The nucleotide sequence comparison of 38 defective HCV isolates showed sequence diversity. Such diversity was observed even among isolates obtained from the same specimen. Perhaps such diversity is a result of self-replication and the subsequent evolution of the defective HCV genome. To explore this possibility, phylogenetic analyses were performed on the nucleotide sequence data from patient K4. Sequences at the 5′ and 3′ maximum overlapping regions located outside the deletions were separately compared (Fig. 2A), and phylogenetic trees were created by the neighbor-joining method with GENETYX software (Genetyx Inc., Tokyo, Japan). As a re-

TABLE 1. Primers used for long-distance and quantitative RT-PCRs in this study

Test and region(s)	Direction	Primer(s)[a]	Sequence[b]	Position[c]
Long-distance PCR				
5′ UTR-NS3	RT	606R/712R	GTTTCCATAGACTC(A/G)ACGGG	3930–3949
5′ UTR-NS3, 5′ UTR-NS5B	1st forward	420	GGCGACACTCCACCATAGATCACTC	1–42
5′ UTR-NS3	1st reverse	605R/713R	ACCGGAATGACATCAGCATG(T/C)CTCGT	3741–3766
5′ UTR-NS3, 5′ UTR-NS5B	2nd forward	AscT7-420	ATCGTAGGCGCGCCTC*TAATACGACTCACTATA*GC CAGCCCCCGATTGGGGGCGACACTCCACCATAGATCACTC	1–42
5′ UTR-NS3	2nd reverse	604R/714R	CGAGGTCCTGGTCTACATT(G/A)GTGTACAT	3639–3666
NS3-NS5B, 5′ UTR-NS5B	RT	386R	AATGGCCTATTGGCCTGGAG	9390–9392
NS3-NS5B	1st forward	602/723	CCACCGCAACACAATCTTTCCT(G/A)GCGAC	3529–3556
NS3-NS5B, 5′ UTR-NS5B	1st reverse	719R/720R/721R	GAGTGTTTAGCTCCCCGTTCA(T/C/G)CGGTTGGG	9363–9392
NS3-NS5B	2nd forward	603/724	CAAAGGGTCCAATCACCCA(A/G)ATGTACAC	3619–3646
NS3-NS5B, 5′ UTR-NS5B	2nd reverse	607R/654R/722R	CGGTTGGGGAGCAGGTA(G/A/G)A(T/T/C)GCCTAC	9345–9370
Quantitative PCR				
5′ UTR	RT	738RH	ACTCGCAAGCACCCTATCAGGC	291–312
5′ UTR	Forward	736	AAGCGTCTAGCCATGGCGTTAGTA	73–96
5′ UTR	Reverse	737R	GGCAGTACCACAAGGCCTTTCG	272–293
5′ UTR	Probe	733FB	FAM-TCTGCGGAACCGGTGAGTACAC-BHQ1	147–168
E2	RT	743RH/744RH/753RH/753RH	CAACGCTCTCCTCG(A/A/G/G)GTCCA(A/G/A/G)TTGCA	2271–2296
E2	Forward[d]	751/752	GGCCTCCACATGGCAA(C/T)TGGTTCGG	1972–1993
E2	Forward[d]	739/740	CCGCCGCAAGGCAACTGGTT(C/T)GG	1974–1993
E2	Reverse	741R/742R	GCCTCGGGGTGCTTCCGGAAGCA(G/A)TCCGT	2088–2116
E2	Probe	734FB/735FB	FAM-TGGATGAA(T/C)AGCACTGGGTTCACCAAGAC-BHQ1	2001–2029

[a] Primers separated by slashes harbor a nucleotide substitution(s) (in parentheses) in the sequence in the same order.
[b] An underline and a double underline indicate recognition sequences for AscI and BsrGI, respectively, with which the PCR products were subcloned into plasmid vector pASGT5. Italics denote the T7 promoter, which was used to synthesize RNA in vitro from the T5S2 isolate (Fig. 4A).
[c] Nucleotide positions correspond to the HCV-JS sequence (12).
[d] Forward primers for E2 were mixed in the reaction mixture.

sult, isolates with the same deletion pattern formed genetic clusters that were distinct from each other, as well as from those of nondefective HCV isolates (Fig. 3A and B). Similar results were obtained for the other patients with defective HCV genomes (data not shown). These results suggest that a defective HCV genome is capable of replication to accumulate mutations and to evolve independently of the nondefective HCV genome.

TABLE 2. Quantitative PCRs for the 5′ UTR and E2 regions of HCV[a]

Region for quantification	No. of copies/ml		5′ UTR/E2 ratio
	5′ UTR	E2	
R2	2.0×10^6	1.7×10^6	1.17
R4	5.3×10^6	2.2×10^6	2.44
T5	ND[b]	ND[b]	
K1	8.3×10^5	8.4×10^5	0.99
K2	3.6×10^5	3.5×10^5	1.01
K3	8.6×10^5	5.1×10^5	1.7
K4pre	8.1×10^5	3.3×10^5	2.45
K4af24	4.5×10^5	4.4×10^5	1.01

[a] For quantification of the 5′ UTR and E2 regions, the TaqMan Fast PCR Universal mixture and the 7500 Fast Real-Time PCR system (Applied Biosystems) were used in a two-step method with the primers and probes shown in Table 1 according to the manufacturer's protocol. The copy number of HCV was determined by the standard-curve method with serial dilutions of the synthesized full-length HCV RNA.
[b] ND, not determined due to sample shortage.

Next, the ability of the defective HCV genome to be encapsidated and released from cells as HCV_{CCD} was examined. A genotype 1b replicon RNA lacking the structural region was synthesized by using defective isolate T5S-2 from patient T5 (Fig. 2 and 4A) as the template in an in vitro transcription system (MEGAscript T7 kit; Ambion, Inc., Austin, TX) under the control of the T7 promoter. Also, capped mRNA encoding the genotype 1b structural proteins from the same patient (designated C-NS2 in Fig. 4A) was synthesized in vitro with the mMessage mMachine T7 kit (Ambion). Both synthesized RNAs were cotransfected into Huh7.5 hepatoma cells. However, HCV_{CCD} was not obtained, presumably because of low replication or virus productivity of genotype 1b HCV per se. In fact, we transfected the defective RNA alone and observed the replication and protein expression of HCV, but with low efficiency (data not shown). Thus, to augment virus productivity, a JFH1-based chimeric HCV genome (genotype 1b/2a) and its deletion mutant were generated to mimic isolate T5S-2 (designated TNS2J1 and TNS2J1ΔS, respectively, Fig. 4A). JFH1 is genotype 2a HCV isolate that can produce high levels of infectious virus (14). To verify the virus productivity of TNS2J1, Huh7.5 cells (10-cm plate) were transfected with 10 μg of in vitro-synthesized RNA from TNS2J1 or JFH1 by lipofection with TransMessenger transfection reagent (Qiagen, Valencia, CA) according to the manufacturer's protocol. Two days later, the culture medium was concentrated 10-fold and inoculated into naïve Huh7.5 cells (four-well chamber

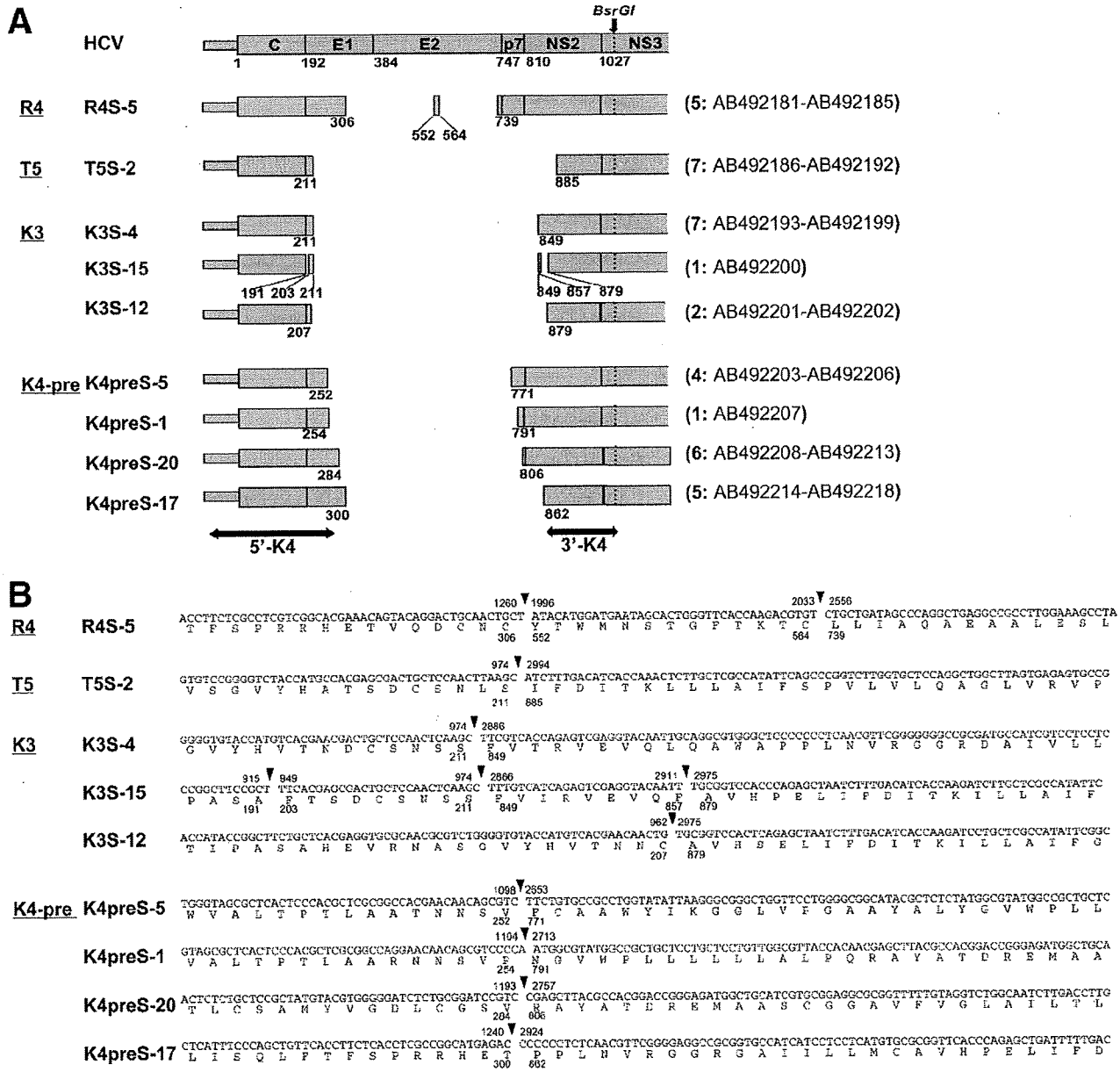

FIG. 2. Sequence analysis of the defective HCV genomes. A total of 38 isolates were molecularly cloned into a plasmid vector and sequenced. Data from representative isolates are presented. Nucleotide positions and deduced amino acid positions correspond to those of genotype 1b strain HCV-JS (12). (A) Defects located in the structural region were compared. The remaining regions are illustrated as shaded boxes. Below the boxes are numbers indicating amino acid positions at the end of each remaining region. At the top of the panel is the HCV genome with the amino acid position at the N terminus of each HCV protein below. The BsrGI restriction site that was used to clone the PCR products is shown as a dotted line. Each value in parentheses at the right is the number of isolates showing the same deletion pattern, followed by the GenBank accession number(s). The two-headed arrows indicate the 5′ and 3′ maximum overlapping regions among the defective HCV isolates in the K4-pre specimen that are compared in the following phylogenetic analyses (5′-K4 and 3′-K4; see Fig. 3). (B) Deletion breakpoints and their adjacent nucleotides and deduced amino acid sequences are indicated. Solid triangles denote breakpoints, and numbers indicate the nucleotide positions (above) and amino acid positions (below) at the junctions.

slide). Cells inoculated with the culture medium from TNS2J1 RNA-transfected cells markedly expressed HCV protein, as shown by immunofluorescent staining (Fig. 4B). The percentage of HCV-positive cells in chimera-infected cells, 40% (565/1,240), was greater than that of JFH1, 3% (37/1,210), demonstrating that the chimeric genome TNS2J1 can produce infectious HCV more robustly than JFH1 can ($P < 0.0001$).

Taking advantage of this chimeric genome, we conducted *trans* complementation experiments. To mimic the T5S-2 iso-

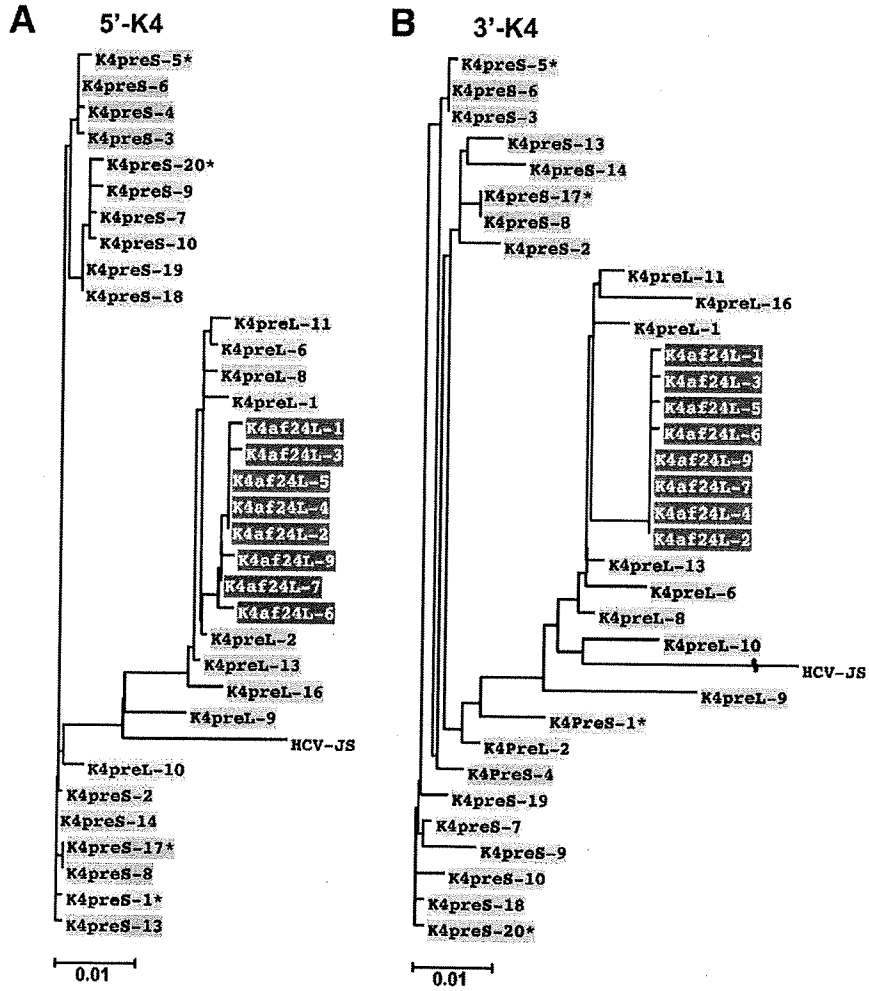

FIG. 3. Phylogenetic analyses of defective HCV genomes. Nucleotide sequence data from 33 isolates from patient K4 were used for phylogenetic analyses. The defective HCV genome (16 isolates) and the nondefective HCV genome coexisting before interferon treatment (9 isolates; GenBank accession no. AB492219 to AB492227) and those after treatment (8 isolates; GenBank accession no. AB492228 to AB492235) were compared in the 5′ and 3′ maximum overlapping regions separately (5′-K4 and 3′-K4 in Fig. 2A). Phylogenetic trees were created for the respective regions (A and B). In the isolate designations, pre and af24 stand for before and after interferon treatment and S and L stand for defective and nondefective HCV genomes, respectively. Isolates with the same deletion pattern (according to K4-pre in Fig. 2) are shaded in the same color. Asterisks denote the representative isolates illustrated in Fig. 2.

late, the region corresponding to the defect found in T5S-2 was identically deleted from the TNS2J1 genome (designated TN2J1ΔS, Fig. 4A). Ten micrograms of synthesized RNA of TN2J1ΔS was cotransfected into Huh7.5 cells (10-cm plate) together with 10 μg of synthesized capped mRNA encoding the structural region, including part of the nonstructural region of TNS2J1, designated C-NS2 or C-NS3P (Fig. 4A). Two days later, the culture medium was concentrated and inoculated into naïve Huh7.5 cells as previously described. HCV protein was expressed when cells were inoculated with the medium of cells cotransfected with TN2J1ΔS RNA and C-NS2 or C-NS3P mRNA, whereas no expression was observed in the case of TN2J1ΔS RNA alone (Fig. 4C). To stably provide the structural proteins in *trans*, packaging cell lines were established by retroviral transduction (2) of Huh7.5 cells with genes encoding the C-NS2 or C-NS3P region (Fig. 4A). These packaging cell lines were transfected with TN2J1ΔS RNA, and HCV protein was expressed in cells inoculated with the culture medium from the RNA-transfected packaging cells (Fig. 4D). Notably, the construct C-NS2 helped to produce HCV_{CCD} more efficiently than C-NS3P did (Fig. 4C). We observed less expression of the structural proteins with the C-NS3 construct than with the C-NS2 construct in a transient expression experiment (data not shown). One possible reason for this is that the C-NS3 construct needs one additional process, i.e., cleavage between NS2 and NS3, to produce NS2 and may affect the other proteins. Otherwise, it is simply because of the difference in the lengths of the constructs. These results indicate that a defective HCV genome lacking the structural region can be encapsidated by *trans* complementation of the structural proteins, thus conferring infectivity in vivo. Recently, a *trans*-packaging system consisting of an HCV subgenomic replicon and a reporter gene

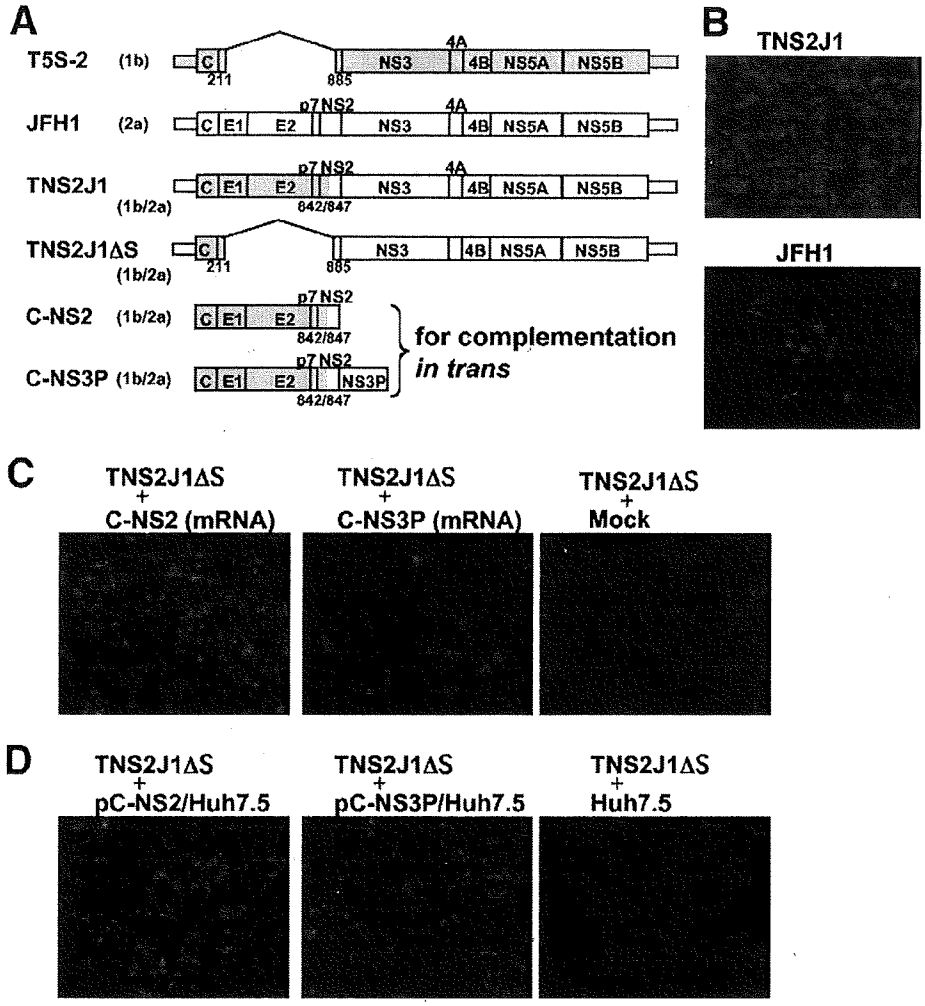

FIG. 4. In vitro infectivity of deletion mutant of chimeric HCV conferred by *trans* complementation of structural proteins. (A) Schematics of the following HCV genomic constructs: the defective HCV isolate (T5S-2), JFH1, chimeric virus of genotypes 1b and 2a (TNS2J1), and its deletion mutant (TNS2J1ΔS). C-NS2 and C-NS3 are fragments encoding the region from the core to the C terminus of the NS2 region and to the C terminus of the serine protease moiety in NS3, respectively. For the *trans* complementation experiments, the latter two constructs were inserted into pcDNA3.1 (Invitrogen) to synthesize capped mRNAs or into retroviral vector pCX4bsr (GenBank accession no. AB086384) to establish packaging cell lines stably expressing the proteins. Shaded and open boxes represent genotypes 1b (isolate from patient T5) and 2a (JFH1), respectively. The numbers below the boxes are amino acid positions at deletion breakpoints or at PCR-based recombination junctions. Naïve Huh7.5 cells were inoculated with the culture medium from cells transfected with JFH1 or TNS2J1 RNA (B), from cells cotransfected with TNS2J1ΔS RNA together with the structural region mRNA (C-NS2 or C-NS3P) or TNS2J1ΔS RNA alone (C), and from the packaging cell line (C-NS2/Huh7.5 or C-NS3P/Huh7.5) transfected with TNS2J1ΔS RNA and parental Huh7.5 cells transfected with TNS2J1ΔS RNA (D). HCV protein was detected by human HCV serum (1:500) by the indirect immunofluorescent method with Alexa Fluor 568 goat anti-human immunoglobulin G (1:200; red; Invitrogen). Nuclei were counterstained with 4′,6′-diamidino-2-phenylindole (DAPI; blue).

was also reported in which an intragenotypic chimera (2a/2a) was used as the most efficient packaging construct (11). Our packaging system used an efficient intergenotypic chimera (1b/2a) to encapsidate a genome mimicking a naturally occurring deletion (1b). Thus, although its efficiency may be different, our system could be a useful tool for the study of HCV_{CCD} of chimeric genome 1b/2a or genotype 1b.

Taken together, genetic analyses of the defective HCV genome showed the potential of its translation and self-replication. These defective genomes can be encapsidated into infectious virus-like particles by *trans* complementation of the structural proteins in vitro. The 5′ UTR and core regions, which are preserved in defective HCV genomes, are targets for the clinical quantification of HCV. Therefore, measured values may represent additive values for defective and nondefective HCVs and the method used for HCV quantification should be reevaluated.

We thank H. Kato, R. Shiina, and H. Yamamoto for technical assistance; T. Wakita for the gift of JFH1; C. Rice for the gift of Huh7.5 cells; and T. Akagi for the gift of retroviral vector pCX4bsr.

This work was supported by a grant-in-aid for scientific research (C) from the Japan Society for the Promotion of Science (KAKENHI18590454) and a grant-in-aid for research on hepatitis from the Ministry of Health, Labor, and Welfare.

REFERENCES

1. **Brinton, M. A.** 1983. Analysis of extracellular West Nile virus particles produced by cell cultures from genetically resistant and susceptible mice indicates enhanced amplification of defective interfering particles by resistant cultures. J. Virol. **46:**860–870.
2. **Chen, C. J., K. Sugiyama, H. Kubo, C. Huang, and S. Makino.** 2004. Murine coronavirus nonstructural protein p28 arrests cell cycle in G_0/G_1 phase. J. Virol. **78:**10410–10419.
3. **Huang, A. S., and D. Baltimore.** 1970. Defective viral particles and viral disease processes. Nature **226:**325–327.
4. **Iwai, A., H. Marusawa, Y. Takada, H. Egawa, K. Ikeda, M. Nabeshima, S. Uemoto, and T. Chiba.** 2006. Identification of novel defective HCV clones in liver transplant recipients with recurrent HCV infection. J. Viral Hepat. **13:**523–531.
5. **Kato, N., K. Sugiyama, K. Namba, H. Dansako, T. Nakamura, M. Takami, K. Naka, A. Nozaki, and K. Shimotohno.** 2003. Establishment of a hepatitis C virus subgenomic replicon derived from human hepatocytes infected in vitro. Biochem. Biophys. Res. Commun. **306:**756–766.
6. **Kishine, H., K. Sugiyama, M. Hijikata, N. Kato, H. Takahashi, T. Noshi, Y. Nio, M. Hosaka, Y. Miyanari, and K. Shimotohno.** 2002. Subgenomic replicon derived from a cell line infected with the hepatitis C virus. Biochem. Biophys. Res. Commun. **293:**993–999.
7. **Lohmann, V., F. Körner, J. Koch, U. Herian, L. Theilmann, and R. Bartenschlager.** 1999. Replication of subgenomic hepatitis C virus RNAs in a hepatoma cell line. Science **285:**110–113.
8. **Nopporpanth, S., S. L. Smits, T. X. Lien, Y. Poovorawan, A. D. Osterhaus, and B. L. Haagmans.** 2007. Characterization of hepatitis C virus deletion mutants circulating in chronically infected patients. J. Virol. **81:**12496–12503.
9. **Poidinger, M., R. J. Coelen, and J. S. Mackenzie.** 1991. Persistent infection of Vero cells by the flavivirus Murray Valley encephalitis virus. J. Gen. Virol. **72**(Pt. 3)**:**573–578.
10. **Shimotohno, K.** 1995. Hepatitis C virus as a causative agent of hepatocellular carcinoma. Intervirology **38:**162–169.
11. **Steinmann, E., C. Brohm, S. Kallis, R. Bartenschlager, and T. Pietschmann.** 2008. Efficient *trans*-encapsidation of hepatitis C virus RNAs into infectious virus-like particles. J. Virol. **82:**7034–7046.
12. **Sugiyama, K., N. Kato, T. Mizutani, M. Ikeda, T. Tanaka, and K. Shimotohno.** 1997. Genetic analysis of the hepatitis C virus (HCV) genome from HCV-infected human T cells. J. Gen. Virol. **78**(Pt. 2)**:**329–336.
13. **Tong, X., and B. A. Malcolm.** 2006. Trans-complementation of HCV replication by non-structural protein 5A. Virus Res. **115:**122–130.
14. **Wakita, T., T. Pietschmann, T. Kato, T. Date, M. Miyamoto, Z. Zhao, K. Murthy, A. Habermann, H. G. Kraussslich, M. Mizokami, R. Bartenschlager, and T. J. Liang.** 2005. Production of infectious hepatitis C virus in tissue culture from a cloned viral genome. Nat. Med. **11:**791–796.
15. **Yagi, S., K. Mori, E. Tanaka, A. Matsumoto, F. Sunaga, K. Kiyosawa, and K. Yamaguchi.** 2005. Identification of novel HCV subgenome replicating persistently in chronic active hepatitis C patients. J. Med. Virol. **77:**399–413.
16. **Yoon, S. W., S. Y. Lee, S. Y. Won, S. H. Park, S. Y. Park, and Y. S. Jeong.** 2006. Characterization of homologous defective interfering RNA during persistent infection of Vero cells with Japanese encephalitis virus. Mol. Cells **21:**112–120.