Supplementary File 1.

Correlation of *KRT19* mRNA expression with clinicopathological characteristics

For microarray experiments, 139 HCC samples were included from individuals undergoing surgical treatment (resection or transplantation) from two ethnic groups (61 Chinese and 78 Caucasian) as previously published. [1,2] Overall survival time (n=113) was determined from the date of the histopathological diagnosis to the end of the study, when the status of the patient was ascertained from the medical records and by consulting municipality registers. Out of the 78 Caucasian patients treated at the University Hospitals Leuven Belgium a detailed periodic follow-up for local or remote recurrence of disease after resection was assessed for 66 patients until the patient's death or the end of the study by clinical assessment, chest X-rays and abdominal ultrasonography or computed tomography as previously reported.[3] The gene expression profile was generated at the Advanced Technology Centre at the National Cancer Institute, USA. Data transformation and normalization of gene expression was performed as previously reported.[1,2]

From a total of 21,329 the genes associated with *KRT19* expression were identified using Pearson's correlation. A cut-off of more than 0.5 correlation or less than -0.5 correlation was applied. The relationship between the *KRT19* expression, as a continuous variable, and various clinicopathological parameters was analyzed using the Mann-Whitney test. *KRT19* expression was also dichotomized, using a cut-off value of 1, 1.5 and 2 in log2-transformed scale of gene expression ratios, to allow application of the Mann-Whitney test for a continuous variable, i.e. age, and the use of Kaplan-Meier method. The prognostic relevance of *KRT19* expression was assessed using the log rank test and the univariate and multivariate Cox proportional hazards model. The tests were two-sided. The probabilities of overall survival and recurrence were also estimated according to the Kaplan-Meier method. The statistical analyzes were performed in StatView 5.0.1 software (SAS institute Inc., Cary, USA). Network analyzes were performed using the MetaCore™ software algorithm (GeneGo Inc., St. Joseph, MI, USA).

The relationship between *KRT19* expression and clinicopathological parameters was analysed using the Mann-Whitney test and prognostic relevance using the multivariate Cox proportional-hazards model. *KRT19* expression, as a continuous variable, is significantly associated with high serum AFP (> 300 ng/ml) (p=0,027), poor tumour differentiation (Edmondson Steiner differentiation grade III and IV versus I-II, p=0,0018), and with a larger tumour size (tumours ≥ 5cm) (p=0,041) (Mann-Whitney test). No significant association of *KRT19* expression with gender, proportion of patients with cirrhosis or former contact with Hepatitis B or C was observed (see Table and Fig below).

Table . Clinical and pathological features of HCC patients with correlation to *KRT19* expression.

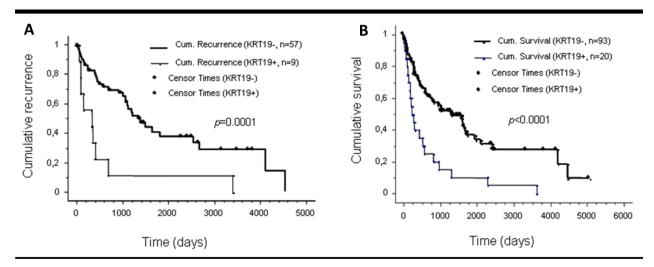|		*Amount (percentage)*	*Correlation p-value*
Age (years) (median 57, range 19-85)			
	>60	61 (43.9%)	NS
	<60	78 (56.1%)	NS
Gender			
	Male	102 (73.4%)	NS
	Female	37 (26.1%)	NS
Aetiology			
	HBV	58 (41.7%)	NS
	HCV	14 (10.1%)	NS
	HBV + HCV	4 (2.9%)	NS
	ALD	15 (10.8%)	NS
	ALD + viral infection	4 (2.9%)	NS
	Other	9 (6.5%)	NS
	Unknown	35 (25.2%)	NS
AFP(ng/ml)			
	>300	55 (39.6%)	p=0,027
	<300	73 (52.5%)	NS
	NA	11 (7.9%)	NS
Tumour size			
	≥ 5 cm	82 (59.0%)	p=0,041
	< 5 cm	57 (41.0%)	NS
Edmondson Steiner differentiation grade			
	I	2 (1.4%)	NS
	II	57 (41.0%)	NS
	III	74 (53.2%)	p=0,0018
	IV	6 (4.3%)	NS
Cirrhosis			
	NO	70 (50.4%)	NS
	YES	69 (49.6%)	NS

HBV, Hepatitis B Virus

HCV, Hepatitis C Virus

ALD, Alcoholic Liver Disease

NA, not available

NS, not significant

A

B

C

Multivariate Cox Proportional Hazards analysis for recurrence

Variable	p-value	Hazards ratio
KRT19 expression	**0.0007**	1.70
Edmondson Steiner differentiation grade III-IV versus I-II	0.45	0.77
Serum AFP (> 300 ng/ml)	0.19	1.64
Tumour size (≥ 5cm)	0.23	1.56

Fig. Correlation of *KRT19* mRNA expression with clinicopathological characteristics.

Panel A: *KRT19* mRNA expression and cumulative recurrence. Dichotomized according to *KRT19* expression using a cut-off value of 1.5 in log2 transformed scale of gene expression ratios (n=66). *KRT19* negative, <1.5: n=57, Events n=34, Censored n=23, *KRT19* positive, ≥ 1.5: n=9, Events n=9, Censored n=0.

Panel B: *KRT19* mRNA expression and cumulative survival. Dichotomized according to *KRT19* expression using a cut-off value of 1.5 in log2 transformed scale of gene expression ratios (n=113). *KRT19* negative HCCs, < 1.5: n=93, Events n=54, Censored n=39, *KRT19* positive ≥ 1.5 : n=20, Events n=20, Censored n=0).

Panel C: Multivariate Cox analysis shows that *KRT19* expression is an independent predictor of tumor recurrence (Log rank test, p=0,007; HR, 1.70) when corrected for serum AFP, tumor size and differentiation.

1. Lee, J.S.*, et al.* Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. *Hepatology* **40**, 667-676 (2004).
2. Lee, J.S.*, et al.* A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. *Nat Med* **12**, 410-416 (2006).
3. Durnez, A.*, et al.* The clinicopathological and prognostic relevance of cytokeratin 7 and 19 expression in hepatocellular carcinoma. A possible progenitor cell origin. *Histopathology* **49**, 138-151 (2006).