

**Additional file 2. Clinical characteristics of HCC patients (n = 839)**

<b>Characteristic</b>	<b>GSE14520</b>	<b>TCGA-LIHC</b>	<b>LIRI-JP</b>
Patient number	Total	242	226
	Male	211 (87%)	166 (73%)
	Female	31 (13%)	60 (27%)
Age (years)	Range	21 - 77	31 - 89
	Median	50	68
Viral etiology	Hepatitis B	242 (100%)	59 (26%)
	Hepatitis C		120 (53%)
	Hepatitis B & C		4 (2%)
	Negative		43 (19%)
	Not available		90 (23%)
Edmondson grade	I		28 (12%)
	II		156 (69%)
	III		21 (9%)
	IV		1 (0.4%)
	Not available	242 (100%)	371 (100%)
Tumour size	Small (< 5cm)	153 (63%)	174 (77%)
	Large (> 5cm)	89 (37%)	52 (23%)
	Not available		371 (100%)
Fibrosis	Negative	19 (8%)	21 (9%)
	Portal fibrosis (stages 1 -2)		73 (32%)
	Fibrous septa (stages 3 - 4)		132 (58%)
	Nodular formation (stage 5)		9 (2%)
	Cirrhosis (stage 6)	223 (92%)	72 (19%)
	Not available		153 (41%)
Alcohol consumption	Yes		86 (38%)
	No		128 (57%)
	Not available	242 (100%)	253 (68%)
Smoker	Yes		122 (54%)
	No		94 (42%)
	Not available	242 (100%)	358 (96%)
AFP	Low (< 300ng/mL)	128 (53%)	218 (59%)
	High (> 300ng/mL)	110 (45%)	66 (18%)
	Not available	4 (2%)	87 (23%)
BCLC stage	0	20 (8%)	226 (100%)
	A	152 (63%)	
	B	24 (10%)	
	C	29 (12%)	
	Not available	17 (7%)	371 (100%)
TNM stage	I	96 (40%)	35 (15%)
	II	78 (32%)	101 (45%)
	III	51 (21%)	68 (30%)
	IV		18 (8%)
	Not available	17 (7%)	89 (24%)