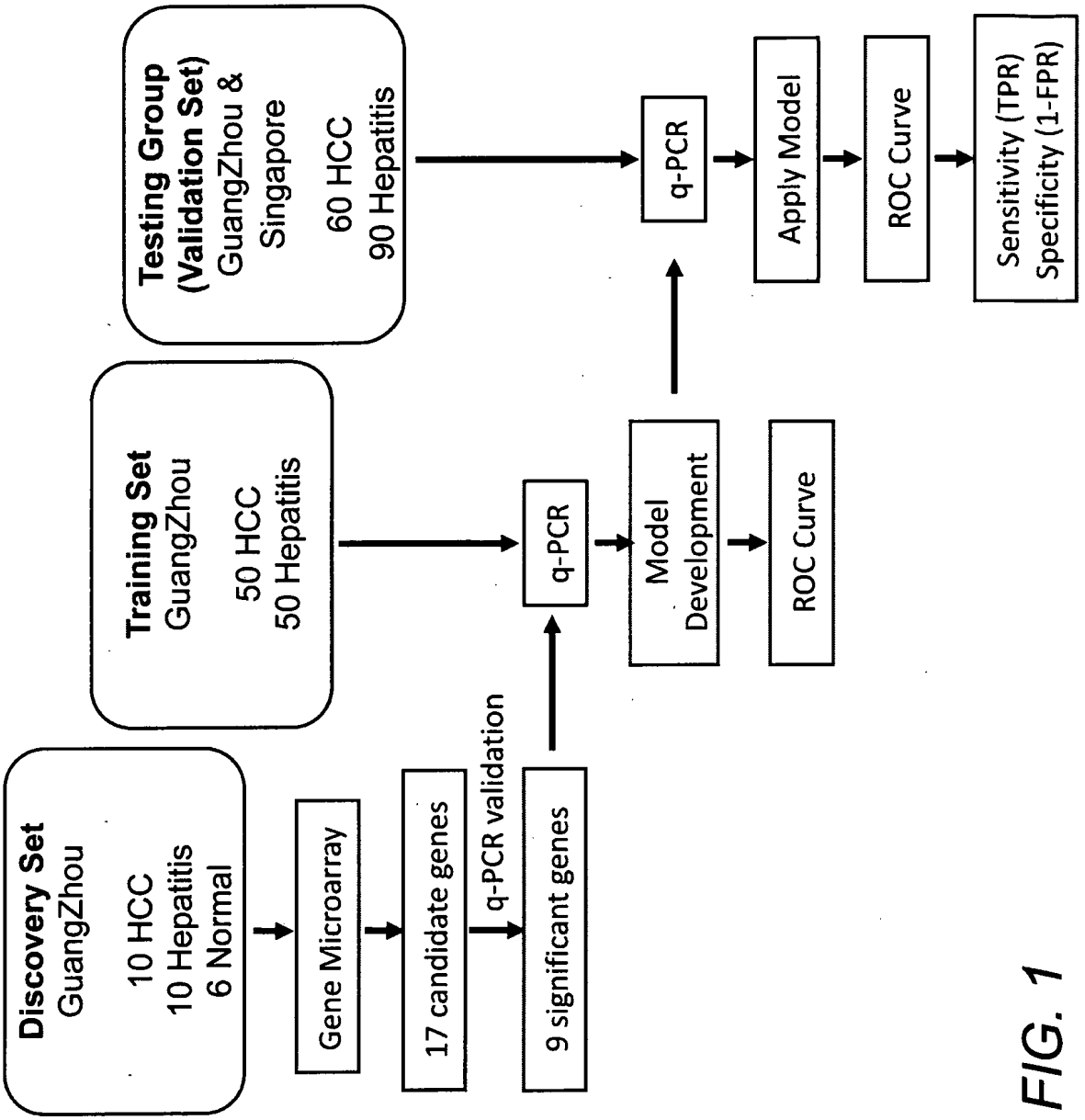

Discovery Set
GuangZhou
10 HCC
10 Hepatitis
6 Normal
Gene Microarray
17 candidate genes
q-PCR validation
9 significant genes
Training Set
GuangZhou
50 HCC
50 Hepatitis
q-PCR
Model Development
ROC Curve
Testing Group
(Validation Set)
GuangZhou &
Singapore
60 HCC
90 Hepatitis
q-PCR
Apply Model
ROC Curve
Sensitivity (TPR)
Specificity (1-FPR)

FIG. 1

Table 1. Clinical characteristics of study participants (Guangzhou sample)

Group	HCC (n=75)	Hepatitis (n=128)
Gender		
Male	70 (93%)	99 (77%)
Female	5 (7%)	29 (23%)
Age (years)	52 (29-76)	49 (18-69)
>50	38 (51%)	26 (20%)
AFP ng/ml [a]		
≤ 20	25 (34%)	66 (50%)
>20	48 (64%)	36 (28%)
ALB g/l [b]		
≥35	59 (79%)	74 (58%)
ALT U/L [c]		
≤ 40	34 (45%)	22 (17%)
Cirrhosis [d]	54 (72%)	29 (27%)
Tumor size [e]		
≤ 3cm	11 (15%)	NA
3-5 cm	22 (31%)	NA
>5 cm	39 (54%)	NA

a. 73 out of 75 HCC patients have AFP values; 100 out of 128 hepatitis patients have AFP values.

b. 118 out of 128 hepatitis patients have ALB values.

c. 123 out of 128 hepatitis patients have ALT values.

d. 29 hepatitis patients were confirmed with cirrhosis.

e. 72 out of 75 HCC patients have tumor size information

FIG. 2A

Table 2. Clinical characteristics of study participants (NCC/SGH sample)

Group	HCC (n=35)	Hepatitis (n=12)
Gender		
Male	17 (49%)	9 (75%)
Female	4 (11%)	2 (17%)
Age (years)		
>50	20 (57%)	0 (0%)
AFP ng/ml [a]		
≤ 20	19 (54%)	7 (58%)
>20	11 (31%)	0 (0%)
ALB g/l [b]		
≥35	17 (49%)	4 (33%)
ALT U/L [c]		
≤ 40	11 (31%)	5 (42%)
Cirrhosis [d]	13 (37%)	6 (0.5%)
Tumor size [e]		
≤ 3cm	5 (14%)	NA
3-5 cm	5 (14%)	NA
>5 cm	15 (43%)	NA

a. 31 out of 35 HCC patients have AFP values;
7 out of 12 hepatitis patients have AFP values.

b. 5 out of 12 hepatitis patients have ALB values.

c. 6 out of 12 hepatitis patients have ALT values.

d. 6 out of 12 hepatitis patients were confirmed with cirrhosis.

e. 26 out of 35 HCC patients have tumor size information

FIG. 2B

Table 3 Differential expression and diagnostic performance of the 9 significant genes in the training group (HCC n=50, CHB n=50).

Gene Name	Fold Change	AUC	Std. Error	P-value	95% Confidence Interval
GIMAP6	-1.94801	.841	.039	.000	.764 - .919
GIMAP4	-2.11316	.808	.043	.000	.723 - .892
GIMAP5	-2.70232	.906	.030	.000	.848 - .965
GIMAP8	-2.45173	.801	.043	.000	.716 - .885
TNFAIP	4.67122	.943	.024	.000	.896 - .989
AREG	8.28786	.926	.027	.000	.873 - .978
NFKBIA	4.64282	.869	.036	.000	.799 - .940
NKFBIZ	3.55833	.885	.032	.000	.823 - .947
CD83	2.82843	.776	.049	.000	.679 - .872

FIG. 3A

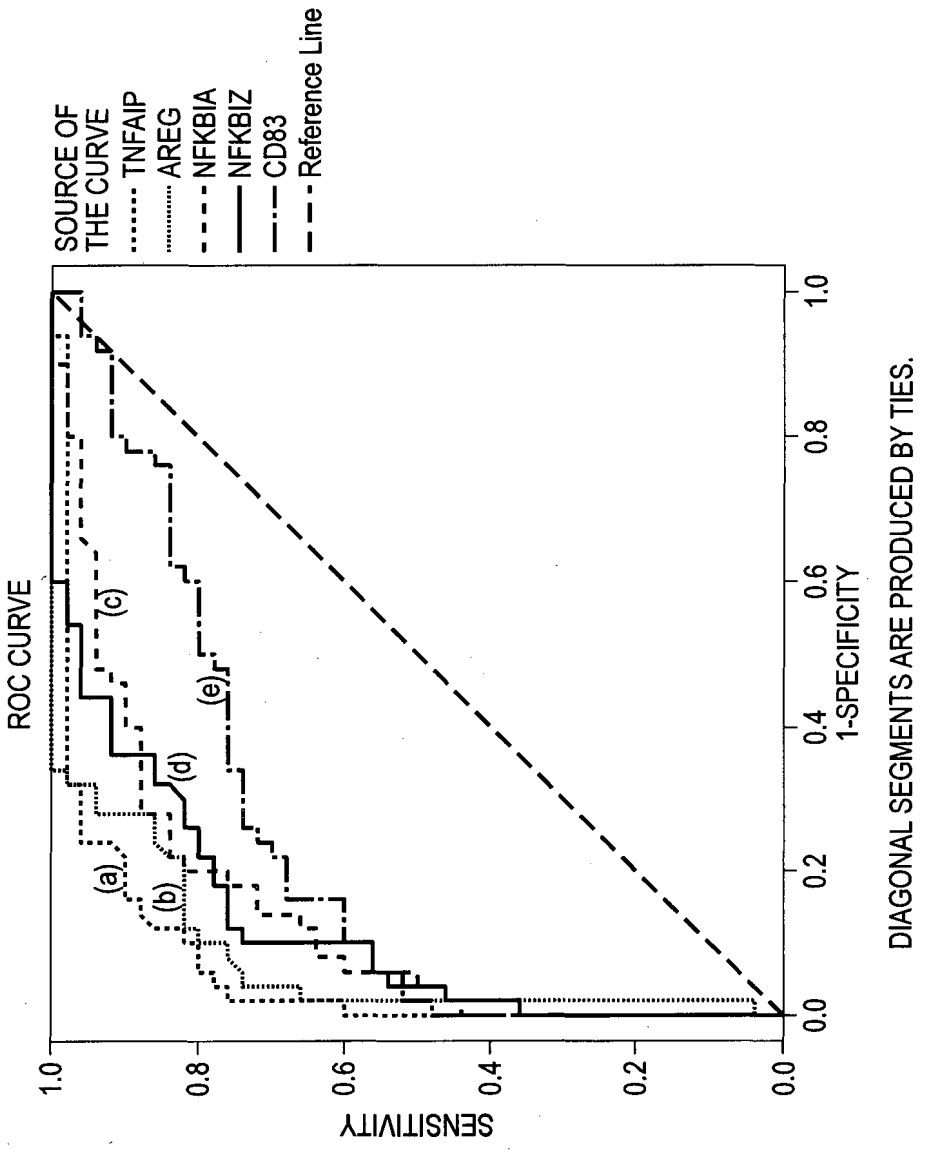

ROC CURVE
SOURCE OF THE CURVE
TNFAIP
AREG
NFKBIA
NFKBIZ
CD83
Reference Line
(a)
(b)
(c)
(d)
(e)
SENSITIVITY
1-SPECIFICITY
1.0
0.8
0.6
0.4
0.2
0.0
DIAGONAL SEGMENTS ARE PRODUCED BY TIES.

FIG. 3B

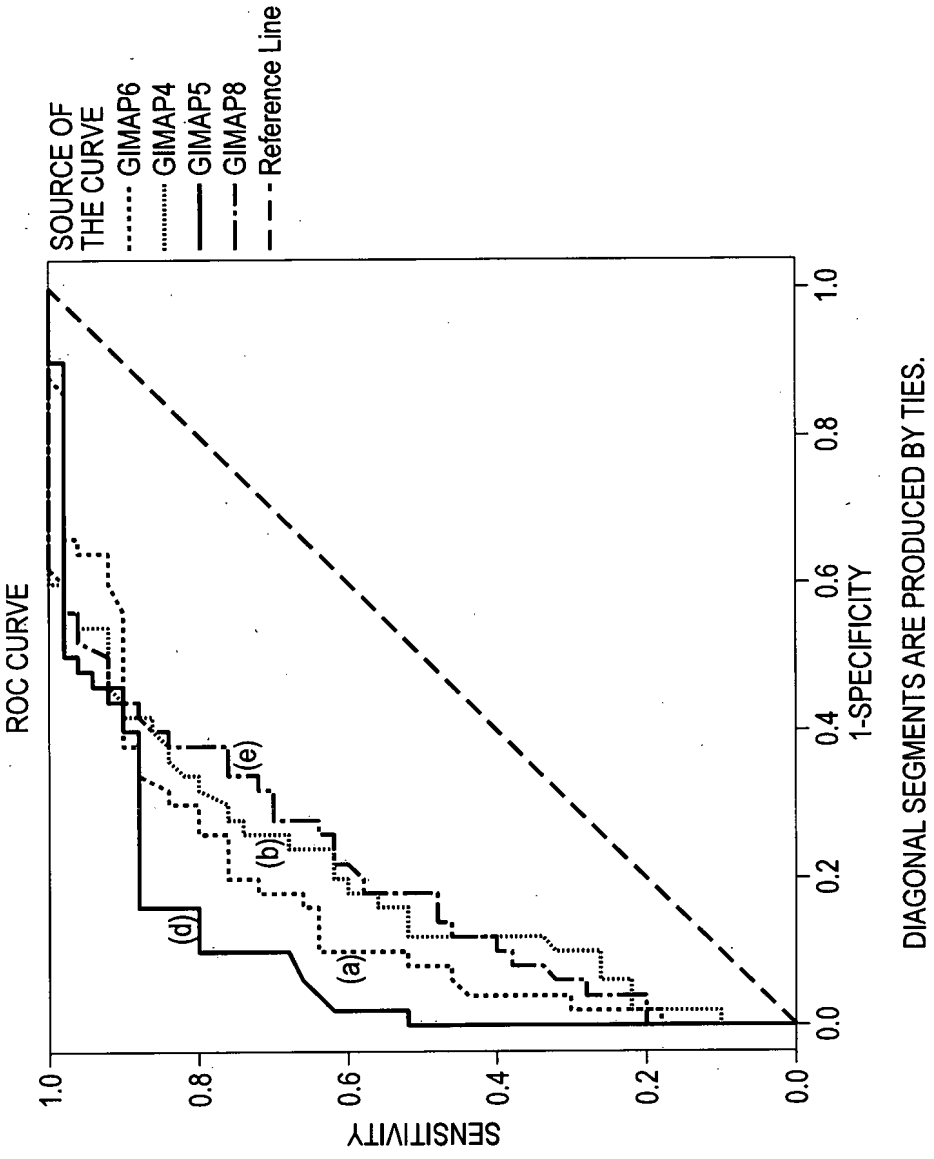
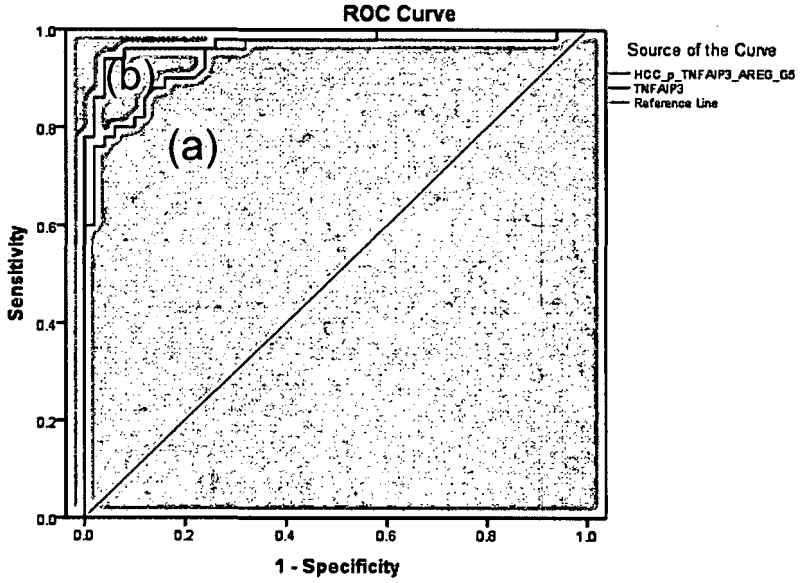

ROC CURVE
SOURCE OF THE CURVE
GIMAP6
GIMAP4
GIMAP5
GIMAP8
Reference Line
(a)
(b)
(d)
(e)
SENSITIVITY
1-SPECIFICITY
0.0
0.2
0.4
0.6
0.8
1.0
DIAGONAL SEGMENTS ARE PRODUCED BY TIES.

FIG. 3C

FIG. 4A

Training Group

50 HCC and 50 Hepatitis



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s)	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
HCC_p_TNFAIP3_AREG_G5	.977	.014	.000	.940	1.000
TNFAIP3	.943	.024	.000	.896	.989

Area Under the Curve

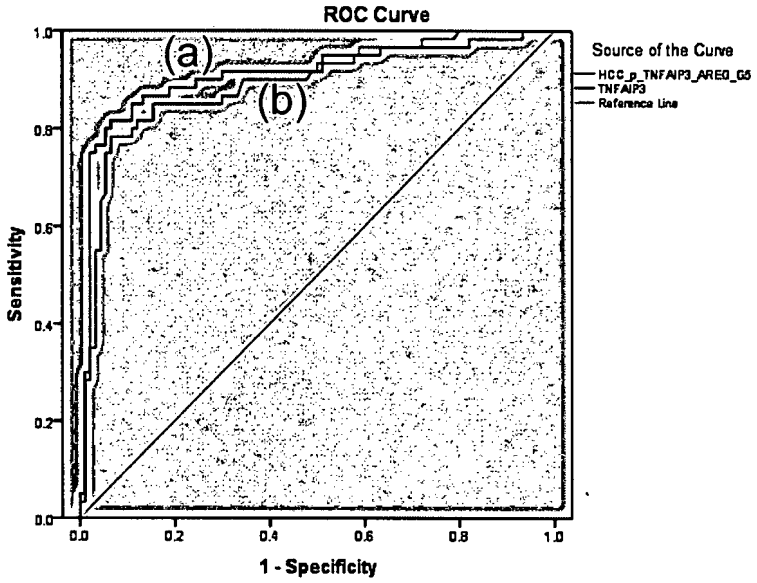
Test Result Variable(s)	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
HCC_p_TNFAIP3_AREG_G5	.909	.028	.000	.853	.964
TNFAIP3	.891	.029	.000	.834	.947

ROC curves analysis of different marker models in the training group (A) and testing group (B). Area Under Curve (AUC) is shown with 95% confidence Interval. Models were developed by stepwise logistic regression (forward method). The probability of being HCC was calculated from the odds ratio and was given as a score ranging from 0 to 100.

Testing Group

69 HCC and 93 Hepatitis

FIG. 4B



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s)	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
HCC_p_TNFAIP3_AREG_G5	.977	.014	.000	.940	1.000
TNFAIP3	.943	.024	.000	.896	.989

Area Under the Curve

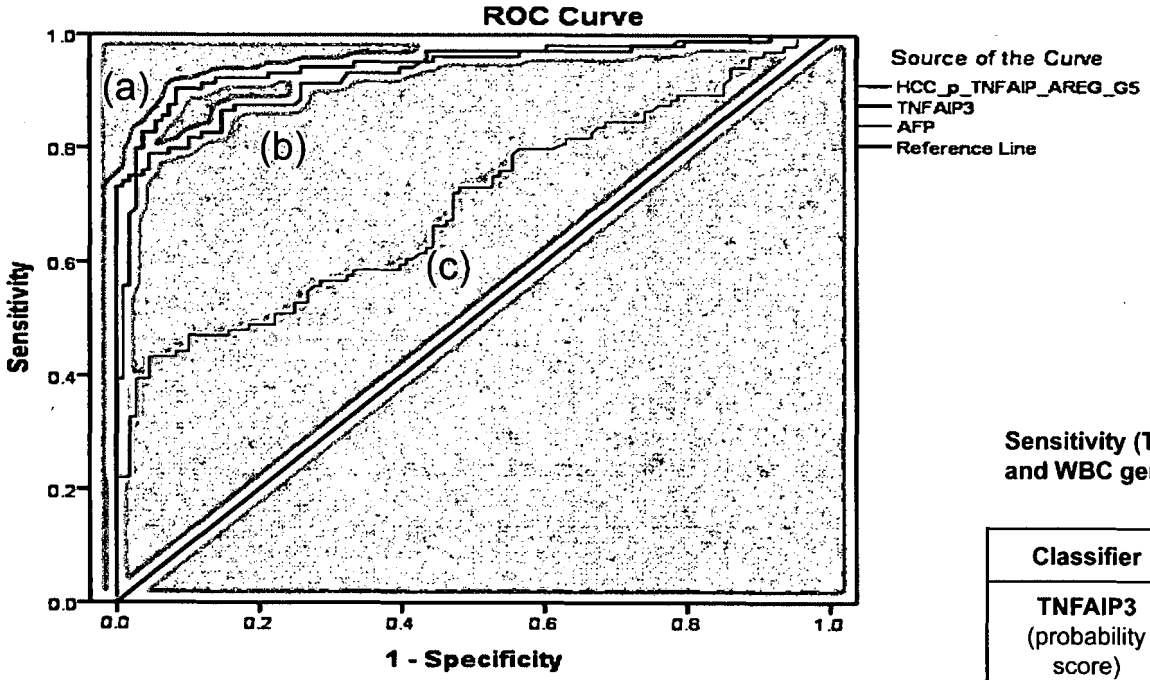
Test Result Variable(s)	Area	Std. Error	Asymptotic Sig.	Asymptotic 95% Confidence Interval Lower Bound	Asymptotic 95% Confidence Interval Upper Bound
HCC_p_TNFAIP3_AREG_G5	.909	.028	.000	.853	.964
TNFAIP3	.891	.029	.000	.834	.947

ROC curves analysis of different marker models in the training group (A) and testing group (B). Area Under Curve (AUC) is shown with 95% confidence Interval. Models were developed by stepwise logistic regression (forward method). The probability of being HCC was calculated from the odds ratio and was given as a score ranging from 0 to 100.

Table 3. Sensitivity (TPR) and specificity (1-FPR) from ROC analysis of training and testing groups at different cutoff points.

Model	Cut-off	Training		Testing	
		TPR	1-FPR	TPR	1-FPR
TNFAIP3 alone	> 0.58	80%	92%	80%	88%
	> 0.73	76%	98%	75%	94%
	> 0.90	58%	100%	55%	96%
TNFAIP3 Set (TNFAIP, AREG, GIMAP5)	> 0.58	92%	96%	85%	87%
	> 0.73	88%	96%	82%	90%
	> 0.90	72%	100%	77%	97%

FIG. 5



Diagonal segments are produced by ties.

Area Under the Curve

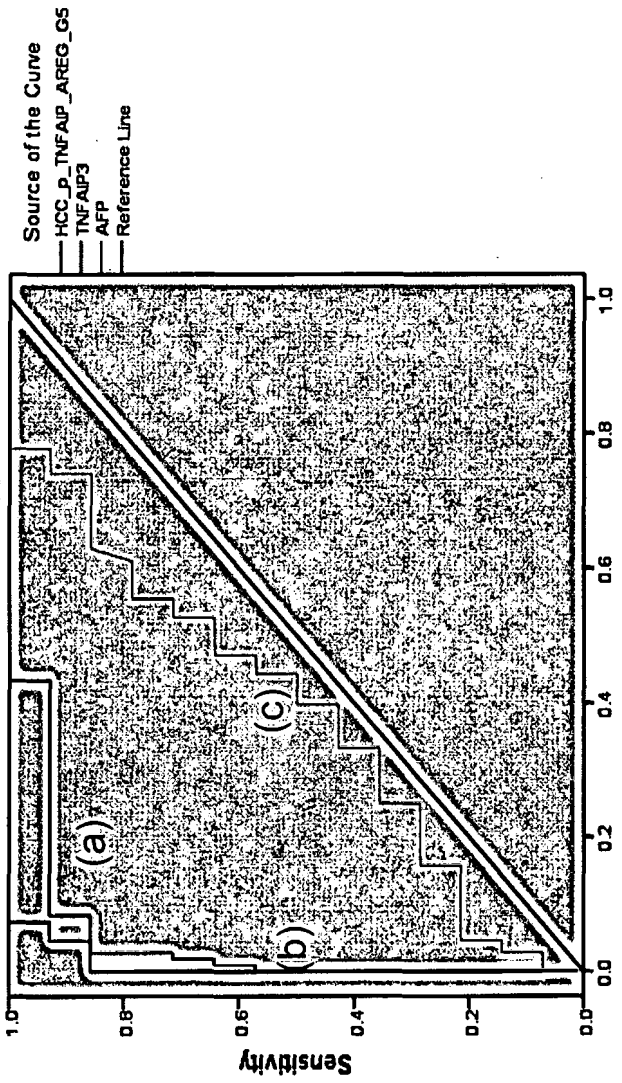
Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
HCC_p_TNFAIP_AREG_G5	.951	.015	.000	.921	.981
TNFAIP	.925	.019	.000	.889	.961
AFP	.697	.036	.000	.626	.769

Sensitivity (TPR) and specificity (1-FPR) of AFP and WBC gene markers (HCC n=104; CHB n=108).

Classifier	Cut-off	TPR	1-FPR
TNFAIP3 (probability score)	> 55	81%	90%
	> 69	79%	95%
	> 90	57%	98%
TNFAIP3 Set (probability score)	> 57	90%	92%
	> 70	87%	93%
	> 90	74%	99%
AFP (ng/ml)	> 20	58%	67%
	> 107	47%	89%
	> 209	43%	95%

FIG. 6

ROC curves analysis of WBC gene markers and serum AFP in 104 HCC and 108 CHB patients.



Diagonal segments are produced by ties.

Area Under the Curve

Test Result Variable(s)	Area	Std. Error[a]	Asymptotic Sig.[b]	Asymptotic 95% Confidence Interval	
				Lower Bound	Upper Bound
HCC_p_TNFAIP_AREG_G5	.963	.031	.000	.000	1.000
TNFAIP3	.985	.009	.000	.959	1.000
AFP	.618	.073	.153	.475	.761

FIG. 7

ROC curves analysis of WBC gene markers and serum AFP in 14 BCLC stage A HCC patients (single nodule 3 cm, no vascular invasion and 140 CHB patients.

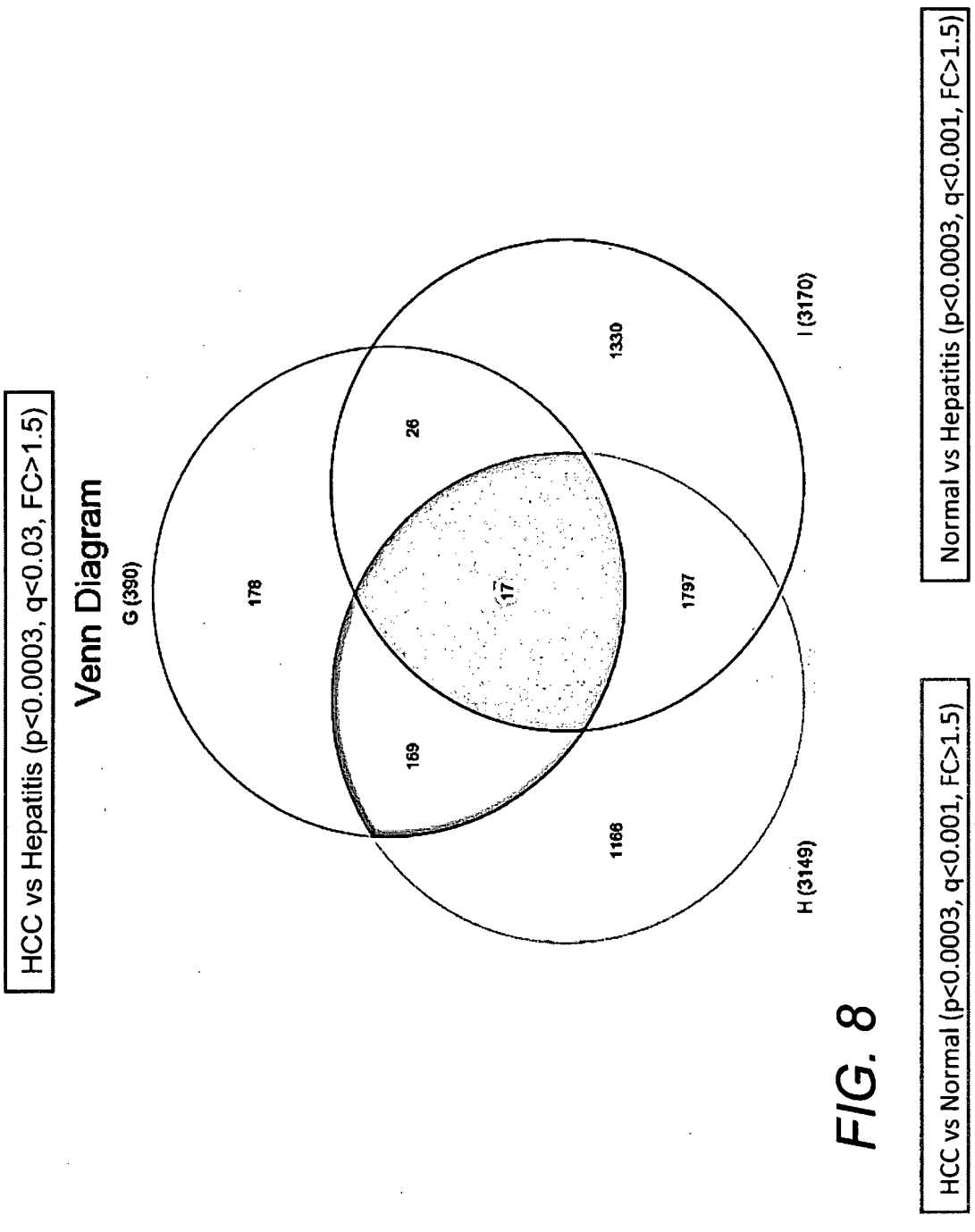


FIG. 8

Venn Diagram for three pairwise comparisons. Candidate gene markers were selected from those genes that are differentially expressed in HCC compared with CHB and healthy subjects.

Table 4 Differential gene expression of the 9 significant genes in HCC and CHB from gene microarray analysis.

Gene Symbol	Gene Title	p-value	q-value	HCC/CHB
TNFAIP3	Tumor necrosis factor, alpha-induced protein 3	3.68E-005	0.015	3.72
NFKBIA	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, al	0.00013	0.027	3.06
NFKBIZ	Nuclear factor of kappa light polypeptide gene enhancer in B-cells inhibitor, ze	5.05E-005	0.017	3.8
AREG	Amphiregulin /// amphiregulin B	5.04E-007	0.0004	8.6
CD83	CD83 molecule	0.0014	0.080	4.83
GIMAP4	GTPase, IMAP family member 4	2.05E-005	0.013	-2.04
GIMAP5	GTPase, IMAP family member 5	1.4E-05	0.008	-2.05
GIMAP6	GTPase, IMAP family member 6	5.11E-005	0.017	-1.70
GIMAP8	GTPase, IMAP family member 8	0.0001	0.024	-2.34

FIG. 9

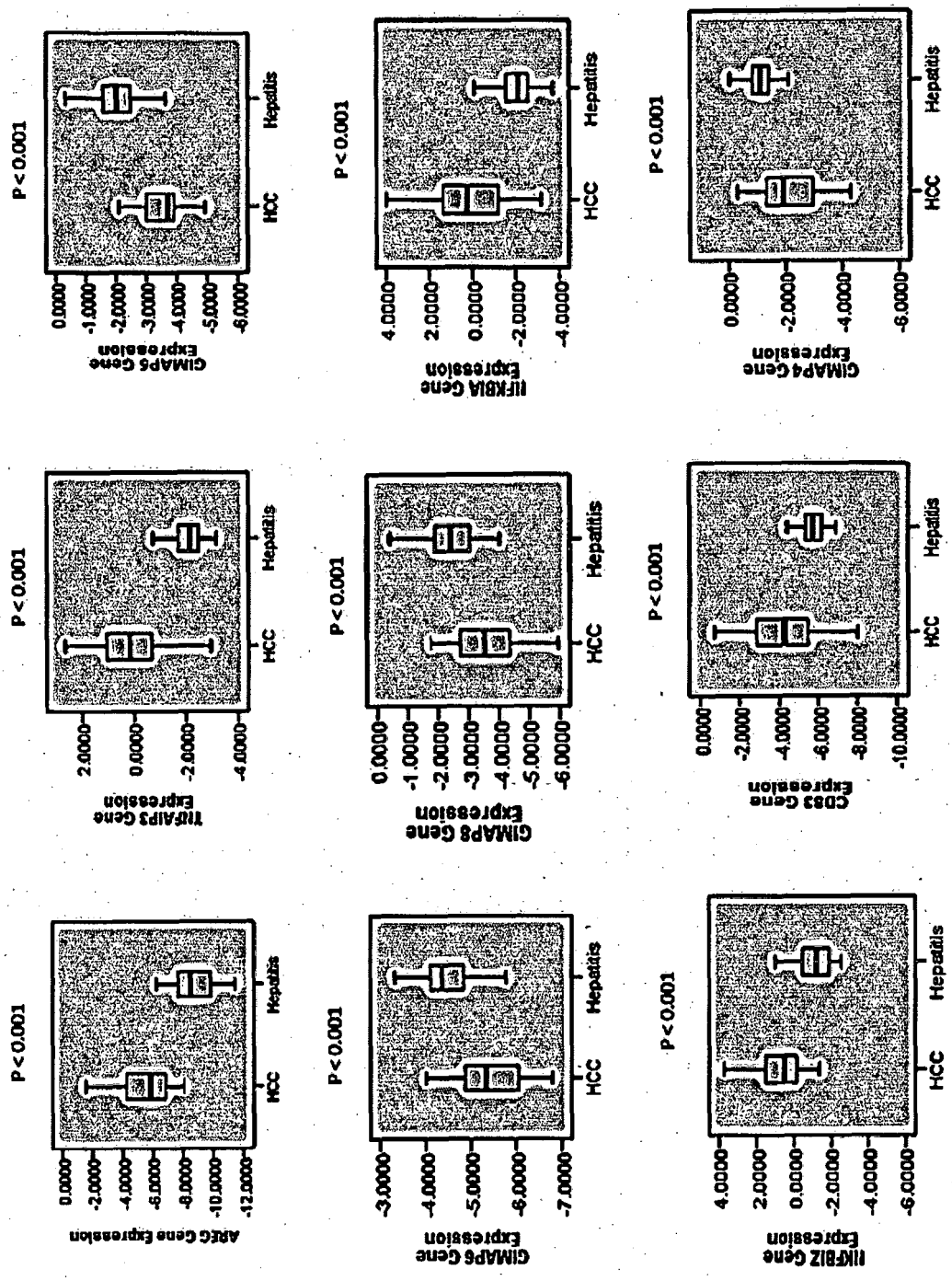


FIG. 10

Differential expression of WBC gene markers in the training group validated by q-PCR (HCC n=50; CHB n=50; Healthy n=6). Gene expression levels were normalized to that of CD45 and presented as percentage of CD45 expression level. Box refers to the 25th and 75th percentile, with the line indicating the median. Whiskers represent the minimum and maximum values. Mann-Whitney test was performed to determine the significance.

		Forward 5'-3'	Reverse 5'-3'
1	NFKBIZ	TCCTGTTGCACATCCGAAGTC	TCCATCAGACAACGAATCGGG
2	GIMAP8	GGGTCGCTCTCCGGCCATTC	CAGGCTCCCGCTTGTTCTGGG
3	GIMAP5	GTGCAGCTGAGTCATGGAGCTT	TTCTCTCCAGAAACGGTTGTTGTGC
4	AREG	GTGGTGCTGTCGCTCTTGATACTC	TCAAATCCATCAGCACTGTGGTC
5	CD83	TGCACTCTGCAGGACCCGGA	TGTAGCCGTGCAAACTTACAAGTGA
6	NFKBIA	CTCCGAGACTTTCGAGGAAATAC	GCCATTGTAGTTGGTAGCCTTCA
7	GIMAP6	GTCTTCGAGTCTAAACTCAGCAC	TGGGTGTGTCAATCACCTCAA
8	TNFAIP3	TTGTCCTCAGTTTCGGGAGAT	TTCTCGACACCAGTTGAGTTTC
9	GIMAP4	GCCCAATACGGCAGTATGAG	CCTGCTCCGGTTTTACCCAC

FIG. 11