Supplement Table 1.Baseline Characteristics of Patients by Outcome: Overall Training Cohort

Variable	Fibrosis Progression N= 274			Clinical Outcome N=533		
	No N=193 Mean or %	Yes N=81 Mean or %	P value	No N= 381 Mean or %	Yes N=152 Mean or %	P value
Age (yr)	49.5	49.4	0.93	49.3	50.3	0.12
% Female	29.5	33.3	0.53	28.8	26.3	0.55
Race (% White)	72.5	72.8	0.40	71.9	69.7	0.43
% HCV genotype 1	92.7	92.5	0.69	92.3	89.47	0.18
Duration of Infection (yr)	26.2	27.2	0.37	26.7	29.0	**0.004**
BMI	29.3	30.5	0.10	29.8	30.3	0.32
Diabetes (%)	11.9	19.7	0.09	15.2	20.39	0.14
Alcohol intake/day (gm)	28.0	25.8	0.69	28.2	29.0	0.85
Tobacco Use (pack yr)	14.7	16.4	0.46	15.7	13.6	0.22
Log HCV RNA ($\log_{10}$ IU/ml)	6.54	6.48	0.29	6.5	6.3	**0.0003**
Platelet count (1000/mm3)	199	169	**0.0001**	182	122	**<0.0001**
INR	0.99	1.02	**0.01**	1.01	1.09	**<0.0001**
AST ratio to ULN*	1.65	2.24	**0.001**	1.93	2.51	**<0.0001**
ALT ratio to ULN*	1.93	2.41	0.05	2.17	2.21	0.80
AST/ALT	0.81	0.85	0.21	0.80	1.02	**<0.0001**
Alkaline Phosphatase ratio to ULN*	0.74	0.89	**0.005**	0.77	1.02	**<0.0001**
Albumin (g/dL)	3.96	3.90	0.17	3.94	3.64	**<0.0001**
Total Bilirubin (mg/dL)	0.65	0.81	**0.0003**	0.71	0.95	**<0.0001**
AFP ratio to ULN*	1.01	2.13	**0.0001**	1.21	3.30	**<0.0001**
MELD	6.55	7.13	**0.0001**	6.84	7.72	**<0.0001**
APRI	0.94	1.63	**<0.0001**	1.30	2.44	**<0.0001**
Ishak	3.1	3.2	0.17	3.8	4.8	**<0.0001**
HAI	7.2	7.3	0.62	7.43	7.84	**0.03**
Steatosis (0-4)	1.11	1.62	**<0.0001**	1.33	1.34	0.96

AFP, alpha-fetoprotein; ALT, alanine aminotransferase; APRI, AST to platelet ratio index; AST, aspartate aminotransferase; BMI, body mass index; HAI, histologic activity index; HCV, hepatitis C virus; INR, international normalized ratio; MELD, model of end stage liver disease; RNA, ribonucleic acid; ULN, upper limit of normal

*Variable expressed relative to the ULN to account for differences in reference ranges for normal results among different clinical trial sites

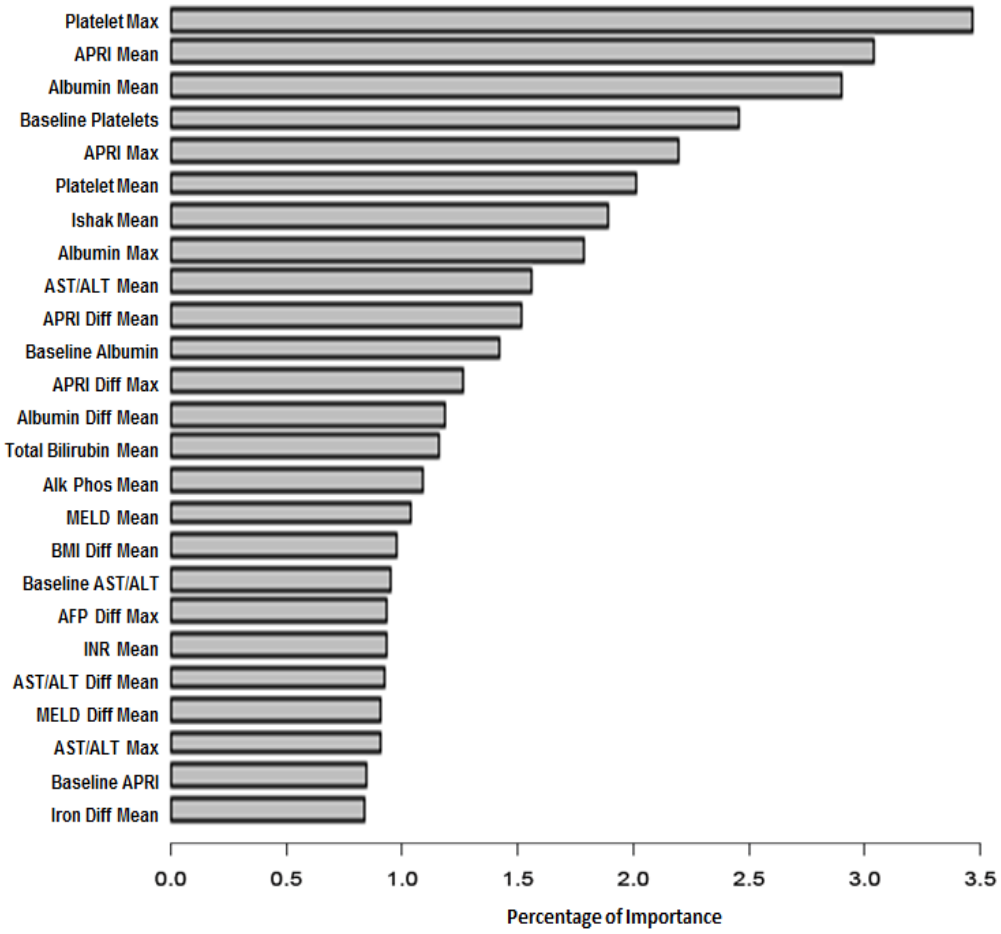
Supplement Table 2. Baseline Characteristics of Patients by Outcome: Internal Validation Cohort

Variable	Fibrosis Progression N= 183			Clinical Outcome N=183		
	No N= 137 Mean or %	Yes N=46 Mean or %	P value	No N=152 Mean or %	Yes N=31 Mean or %	P value
Age (yr)	51.6	49.2	**0.05**	50.8	52.1	0.36
% Female	26.3	21.7	0.54	25.6	22.6	0.72
Race (% White)	67.9	78.3	0.22	71.1	64.5	0.79
% HCV genotype 1	98.5	91.3	0.06	97.4	93.6	0.82
Duration of Infection (yr)	29.6	28.4	0.36	29.1	30.2	0.50
BMI	29.3	29.9	0.44	29.3	29.9	0.55
Diabetes (%)	15.3	21.7	0.31	17.1	16.1	0.89
Alcohol intake/day (gm)	20.2	19.7	0.89	19.8	21.3	0.77
Tobacco Use (pack yr)	14.4	12.7	0.51	14.9	9.7	0.09
Log HCV RNA ($\log_{10}$ IU/ml)	6.6	6.5	0.18	6.5	6.6	0.76
Platelet count (1000/mm3)	198	159	**0.0001**	196	153	**0.0001**
INR	1.00	1.04	**0.05**	1.00	1.04	0.08
AST ratio to ULN*	1.64	2.29	**0.002**	1.68	2.41	**0.002**
ALT ratio to ULN*	1.86	2.47	**0.02**	1.86	2.78	**0.002**
AST/ALT	0.80	0.91	**0.01**	0.82	0.85	0.54
Alkaline Phosphatase ratio to ULN*	0.76	0.77	0.82	0.77	0.73	0.49
Albumin (g/dL)	3.96	3.89	0.22	3.97	3.82	**0.03**
Total Bilirubin (mg/dL)	0.70	0.87	**0.004**	0.72	0.87	**0.02**
AFP ratio to ULN*	1.03	2.46	**0.004**	1.32	1.72	0.49
MELD	6.61	7.09	**0.03**	6.68	7.0	0.19
APRI	0.94	1.68	**<0.0001**	0.98	1.79	**<0.0001**
Ishak	3.1	3.3	0.26	3.1	3.4	**0.007**
HAI	7.23	7.5	0.42	7.17	7.9	**0.05**
Steatosis (0-4)	1.17	1.41	0.13	1.23	1.22	0.98

AFP, alpha-fetoprotein; ALT, alanine aminotransferase; APRI, AST to platelet ratio index; AST, aspartate aminotransferase; BMI, body mass index; HAI, histologic activity index; HCV, hepatitis C virus; INR, international normalized ratio; MELD, model of end stage liver disease; RNA, ribonucleic acid; ULN, upper limit of normal

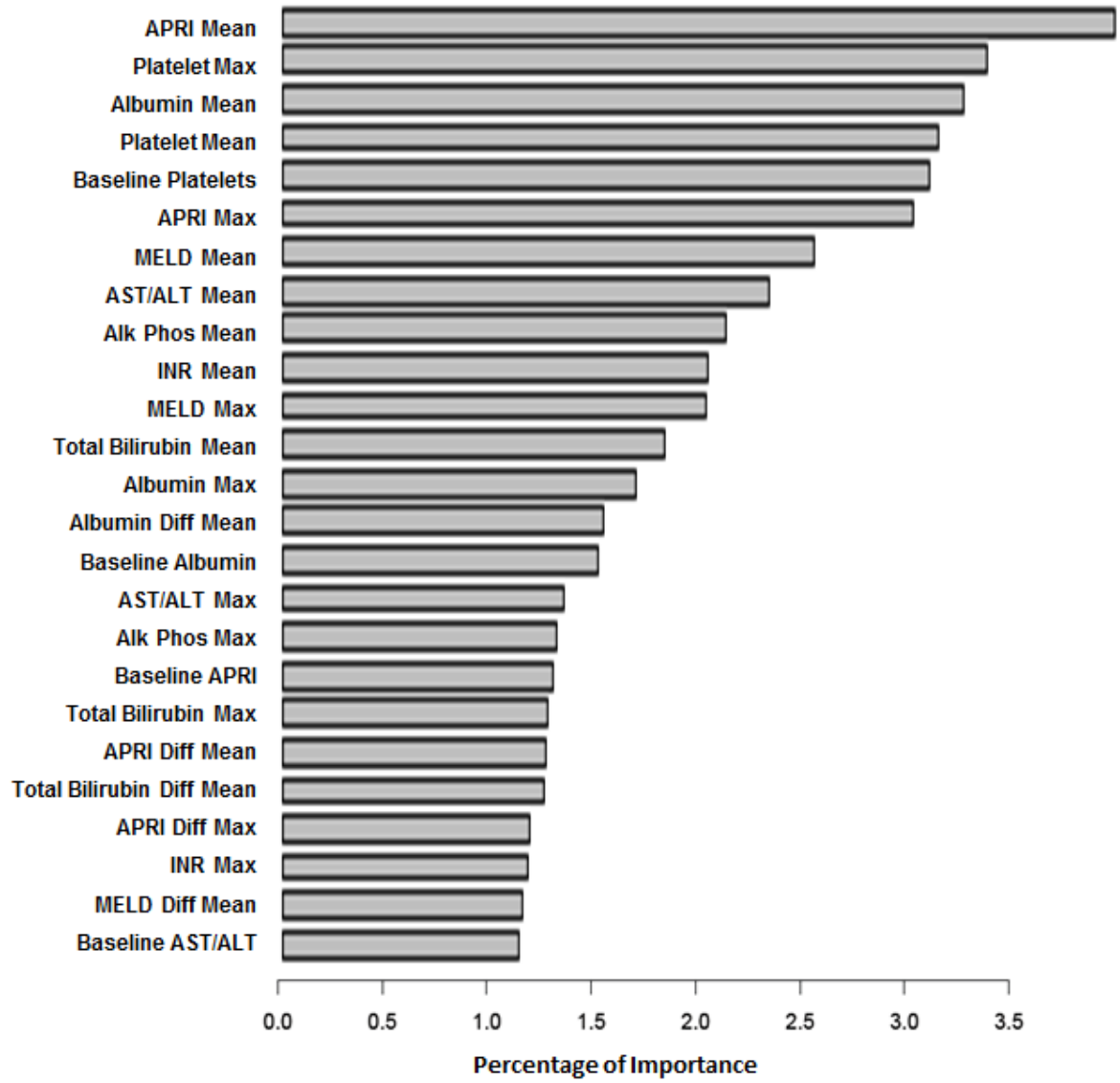
*Variable expressed relative to the ULN to account for differences in reference ranges for normal results among different clinical trial site

Supplement Figure 1. Longitudinal Random Forest Variable Importance Clinical Outcomes without HCC: Training Cohort



AFP, alfa-fetoprotein; ALT, alanine aminotransferase; Alk Phos, alkaline phosphatase; APRI, AST to platelet ratio index; AST, aspartate aminotransferase; BMI, body mass index; Diff, differential; HCC, hepatocellular carcinoma; INR, international normalized ratio; MELD, model of end stage liver disease;

Supplement Figure 2. Longitudinal Random Forest Variable Importance for Condensed Clinical Outcomes Model: Training Cohort



AFP, alpha-fetoprotein; ALT, alanine aminotransferase; Alk Phos, alkaline phosphatase; APRI, AST to platelet ratio index; AST, aspartate aminotransferase; Diff, differential; INR, international normalized ratio; MELD, model of end stage liver disease

Supplement Table 3. Misclassification Table for Condensed Longitudinal Predictive Model of Clinical Outcomes: Internal Validation Cohort

Clinical Outcomes								
		Clinical Progressors (N=46)		Clinical Non-Progressors (N=227)				
	Cutoff	Predicted Clinical Progression	Predicted No Clinical Progression	Predicted Clinical Progression	Predicted No Clinical Progression	Brier score	NPV	PPV
Random Forest	0.211	35 (76.1%)	11 (23.9%)	67 (29.5%)	160 (70.5%)	0.289	**93.6%**	34.3%
Boosting	-16.17	36 (78.3%)	10 (21.7%)	76 (33.5%)	151 (66.5%)	0.315	**93.8%**	32.1%
Logistic Regression	-1.40	32 (69.6%)	14 (30.4%)	60 (26.4%)	167 (73.6%)	0.271	92.3%	34.8%

NPV, negative predictive value; PPV, positive predictive value

SUPPLEMENTAL METHODS SECTION

Development of the Logistic Regression Models

Classic logistic regression models are subject to overfitting if too many predictor variables are included compared to the number of subjects and outcome rates in the study population. In these instances, this will result in inaccurate regression coefficients, large standard errors and confidence intervals. Depending on the ratio of predictor variables to subjects and outcomes, certain models will not converge at all (i.e. the calculation for the prediction cannot be performed and the equation cannot be solved). Due to these restrictions, incorporating the results of longitudinal data has proved difficult using the standard statistical regression approaches.

One approach to address this limitation of classic regression is to apply a lasso technique.(9) This approach is a shrinkage and selection method that minimizes the usual sum of squares errors subject to the sum of the absolute value of the coefficients being less than a constant. It facilitates variable selection in regression to incorporate those variables most strongly correlated with the outcome of interest. In creating a more condensed clinical prediction model that only included variables that were highly associated with the outcome, the condensed model can outperform a more comprehensive model .

Development of the Machine Learning Models

We used random forest analysis, a type of ML algorithm that can build classification prediction models, to identify baseline and longitudinal predictors associated with the development of our outcome.(11) The RF approach divides the initial cohort into two groups—x1 and x2 samples. The x1 sample is created using random sampling from the initial cohort. The x2 sample is composed of the unsampled data from the initial cohort, and typically includes about one-tenth of the initial cohort (10-fold cross validation). This process is repeated 50 times to get

a precise point estimate. For each pairing, a decision tree is constructed, using a random set of potential candidate variables for each split, and then validated using the x2 sample. As each tree is built, only a random subset of the predictor variables is considered as possible splitters for each binary partitioning. The predictions from each tree are used as “votes”, and the outcome with the most votes is considered the dichotomous outcome prediction for that sample. Using this method, multiple decision trees are constructed to create the final classification prediction model and determine overall variable importance. Accuracies and error rates are computed for each observation using the x2 samples, and then averaged over all observations. Because the x2 observations were not used in the fitting of the trees, they serve as cross-validated accuracy estimates. Variable importance identifies the most important variables based on their contribution to the predictive accuracy of the model. The most important variables are identified as those that most frequently result in early splitting of the decision trees. The final algorithms, consisting of 500 trees each, are not presented here for the sake of brevity.

Boosting is another decision tree-based ML algorithm. (10,12) Boosting is based on the observation that identifying a single accurate prediction rule is difficult. Instead, boosting focuses on finding a combination of many rough prediction rules that together can yield an accurate prediction for outcome of interest. This method begins with “weak” learning algorithms that repeatedly assess different subsets and weights of predictors. A “weak” rule is defined as a classifier that only slightly correlates with the true classifier. In this technique, the most weight is placed on the examples that are most often misclassified by the prior “weak” prediction rules. After multiple rounds, the algorithm then generates a combination of these “weak” prediction rules in order to create a single prediction rule. When these rules are combined, a weighted majority vote is effective at accurate predictions.