Integration of Machine Learning-Based Histopathology and Hepatic Transcriptomic Data Identifies Genes Associated With Portal Inflammation and Ductular Proliferation as Predictors of Disease Progression in Advanced Fibrosis Due to NASH

PathAI

GILEAD

Maryam Pouryahya,[1] Amaro Taylor-Weiner,[1] Harsha Pokkalla,[1] Kishalve Pethia,[1] Hunter Elliott,[1] Benjamin Glass,[1] Yevgeniy Gindin,[2] Ling Han,[2] Catherine Jia,[2] Marianne Camargo,[2] Chuhan Chung,[2] Robert P. Myers,[2] Michael C. Montalto,[1] Aditya Khosla,[1] Andrew H. Beck,[1] Murray Resnick,[1,3] Stephen A. Harrison,[4] Zobair M. Younossi,[5] Michael Trauner,[6] Quentin M. Anstee,[7] Ilan Wapinski,[1] Oscar M. Carrasco-Zevallos[1]

[1]PathAI, Inc., Boston, MA; [2]Gilead Sciences, Inc., Foster City, CA; [3]The Warren Alpert Medical School, Brown University, Providence, RI; [4]Pinnacle Clinical Research, San Antonio, TX; [5]Inova Fairfax Hospital, Falls Church, VA; [6]Medical University of Vienna, Austria; [7]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

Introduction

- Portal inflammation, ductular reaction, and fibrosis are associated with progressive nonalcoholic steatohepatitis (NASH)[1-3]
- Machine learning (ML) approaches to assess NASH histology are potentially more sensitive and reproducible than standard, manual evaluation[4]
- RNA-sequencing (RNA-Seq) efforts in NASH have identified genes associated with advanced fibrosis, but have not examined genomic changes in relation to histologic features of NASH such as portal inflammation and ductular reaction[5]

Objective

- To assess whether integration of transcriptomic data with histologic features, identified via an ML approach, would yield novel mechanistic insights into disease progression in study subjects with advanced fibrosis due to NASH

Methods

Study Population, Standard Histology, and Gene Expression Quantitation

- Formalin-fixed, paraffin-embedded tissue blocks of liver biopsies were obtained from 823 subjects with F3–F4 NASH in the STELLAR trials[6]
- Fibrosis was staged at baseline and Week 48 by a central pathologist according to the NASH Clinical Research Network (CRN) classification
- Extracted RNA was used to quantify expression of ~16,500 genes by RNA-Seq (TruSeq protocol, Illumina, Inc., San Diego, CA)

ML-Based Histologic Assessment[4]

- An ML platform (PathAI, Inc., Boston, MA) based on pathologist annotations and convolutional neural networks (CNNs; >20 layers and 8 million parameters) developed for NASH was used to assess NASH histology
  - For quantification of fibrosis, an "end-to-end" CNN was trained using slide-level pathologist scores to recognize unique patterns associated with each stage within fibrotic regions of images of trichrome (TC)–stained biopsies
  - For quantification of nonalcoholic fatty liver disease activity score and other features (eg, portal inflammation and bile ducts), a CNN model was trained based on annotations collected from 75 board-certified pathologists on images of hematoxylin and eosin (H&E)–stained slides to produce pixel-level predictions of each feature
- ML quantitative features were computed at the image level to summarize predictions on H&E and TC images
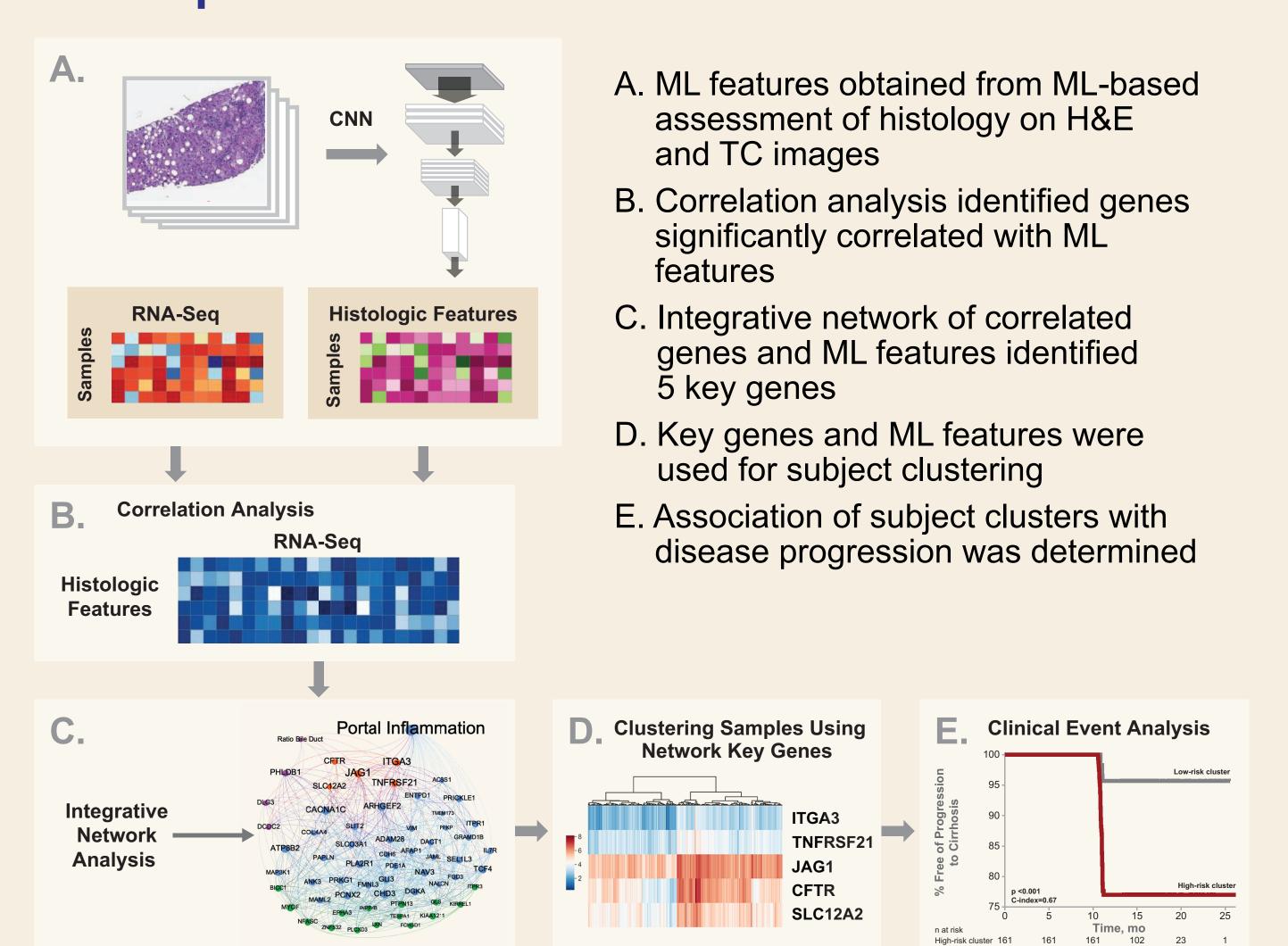
Statistical Analyses

- The top 150 genes significantly correlated with ML-derived proportionate area of portal inflammation or bile ducts were identified (all p <0.001)
- An integrative network was built using 61 genes that were significantly correlated with both proportionate area of portal inflammation and bile ducts
  - The network's connectivity was derived from graphical lasso, which approximately removed an edge between 2 nodes that were conditionally independent given the values of other nodes
- 5 key genes (hub nodes) were identified based on connectivity to both ML histologic features in the integrative network
- Subjects were stratified via hierarchical clustering using key genes or ML histologic features
- Associations of subject clusters with disease progression were determined using Cox proportional hazards regression analysis:
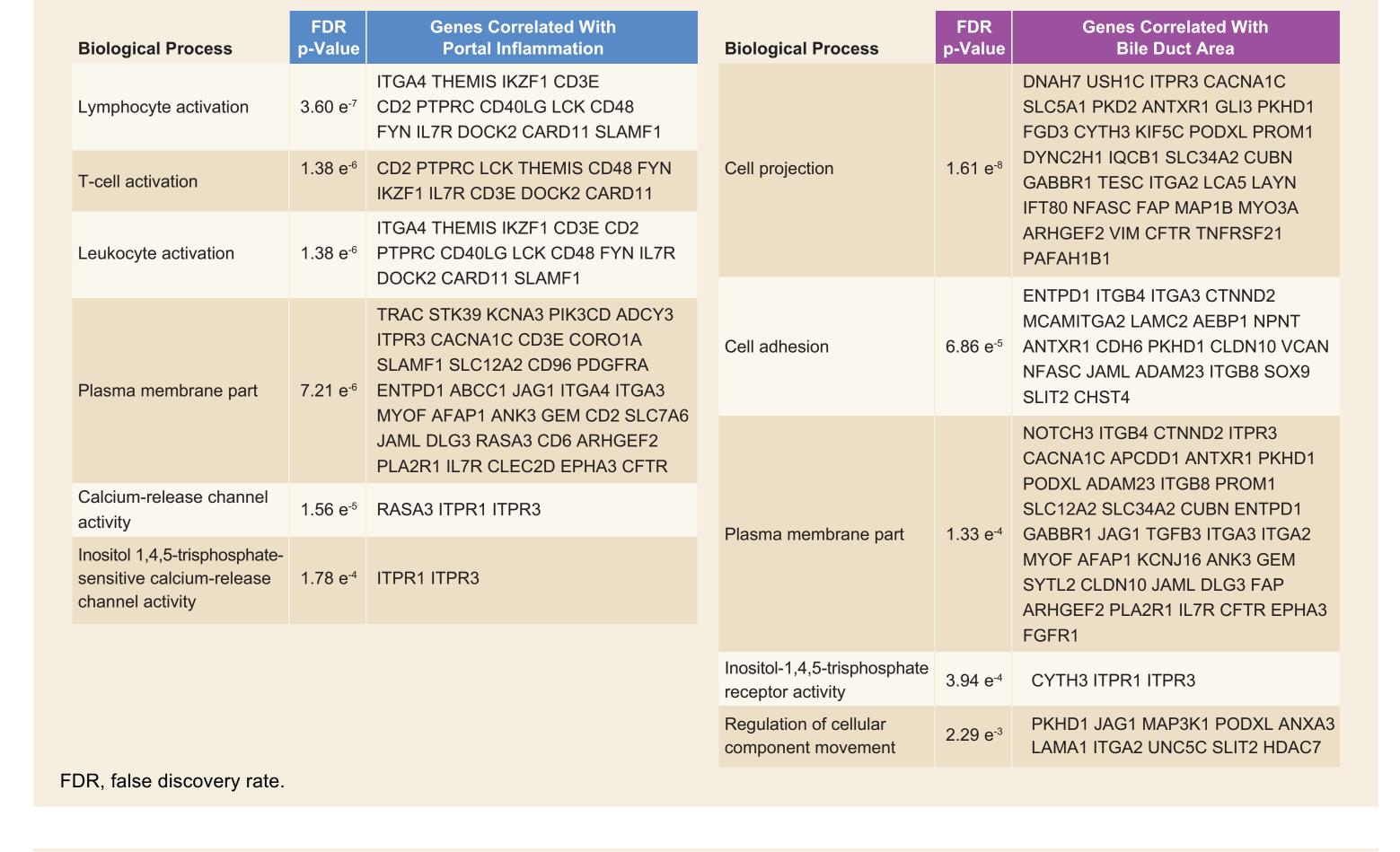 - Progression to cirrhosis for subjects with bridging (F3) fibrosis at baseline
  - Liver-related clinical events (eg, hepatic decompensation, transplantation, and death) for subjects with cirrhosis (F4) at baseline

Pipeline for Integrative Analysis of Histology and Gene Expression in NASH
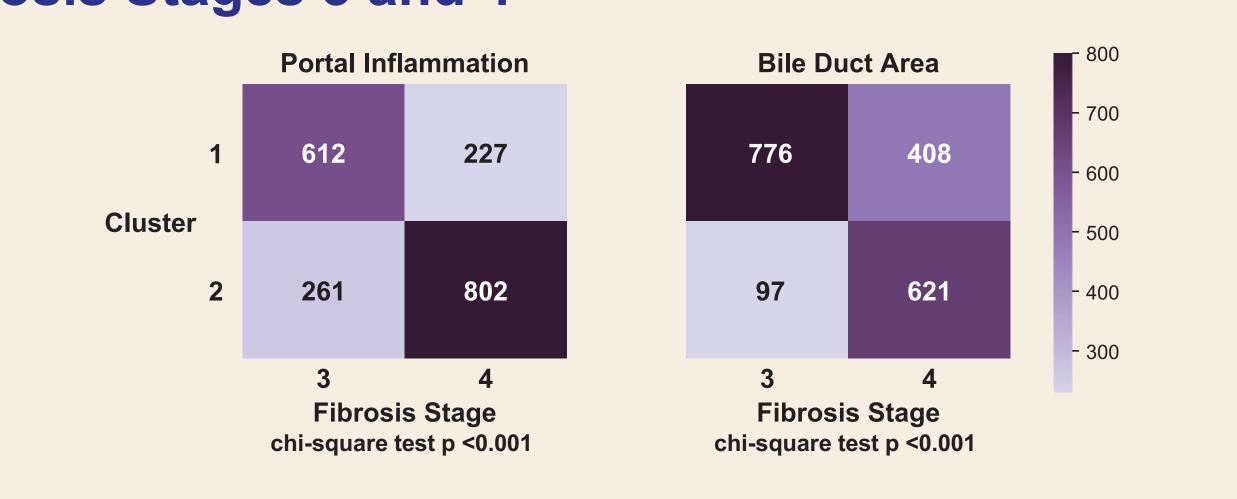

A. ML features obtained from ML-based assessment of histology on H&E and TC images
B. Correlation analysis identified genes significantly correlated with ML features
C. Integrative network of correlated genes and ML features identified 5 key genes
D. Key genes and ML features were used for subject clustering
E. Association of subject clusters with disease progression was determined

Results

Top Enriched Biological Processes for Genes Significantly Correlated With Portal Inflammation and Bile Duct Area
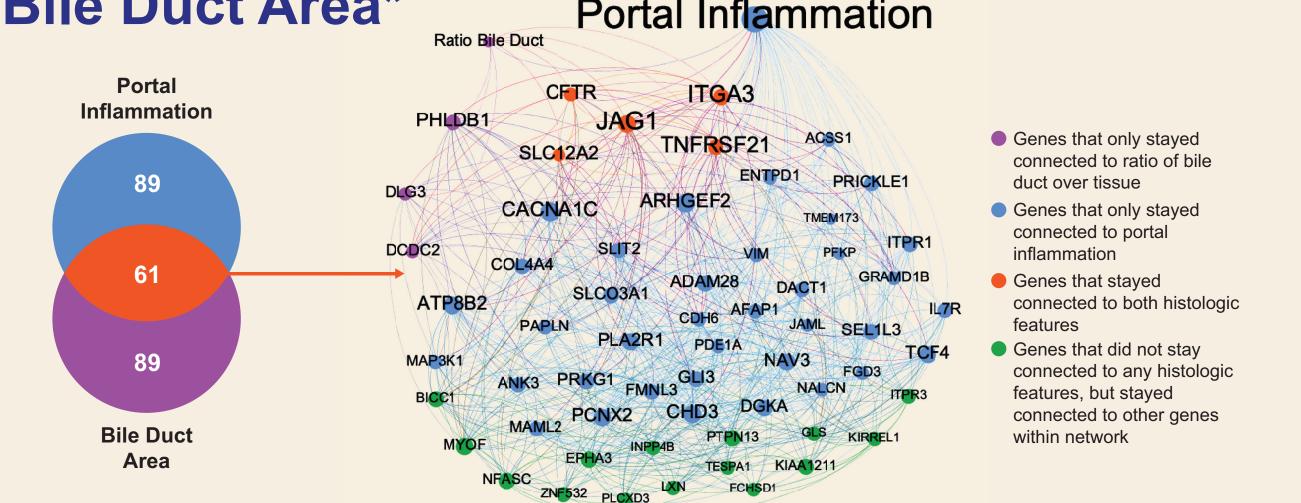

FDR, false discovery rate.

Genes Correlated With Portal Inflammation and Bile Duct Area Were Associated With NASH CRN Fibrosis Stages 3 and 4
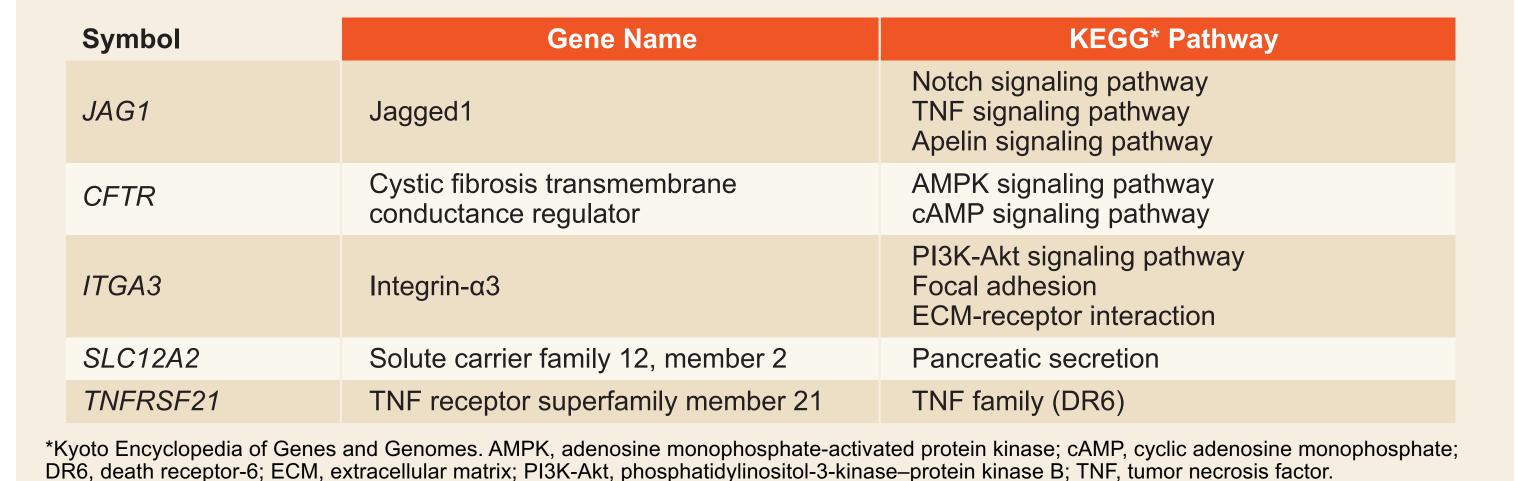

- Using the top correlated genes, subjects were stratified into 2 clusters that were significantly associated with CRN fibrosis stages 3 and 4

A 5-Gene Signature Derived From an Integrative Network of Transcriptomic Data, Portal Inflammation, and Bile Duct Area*
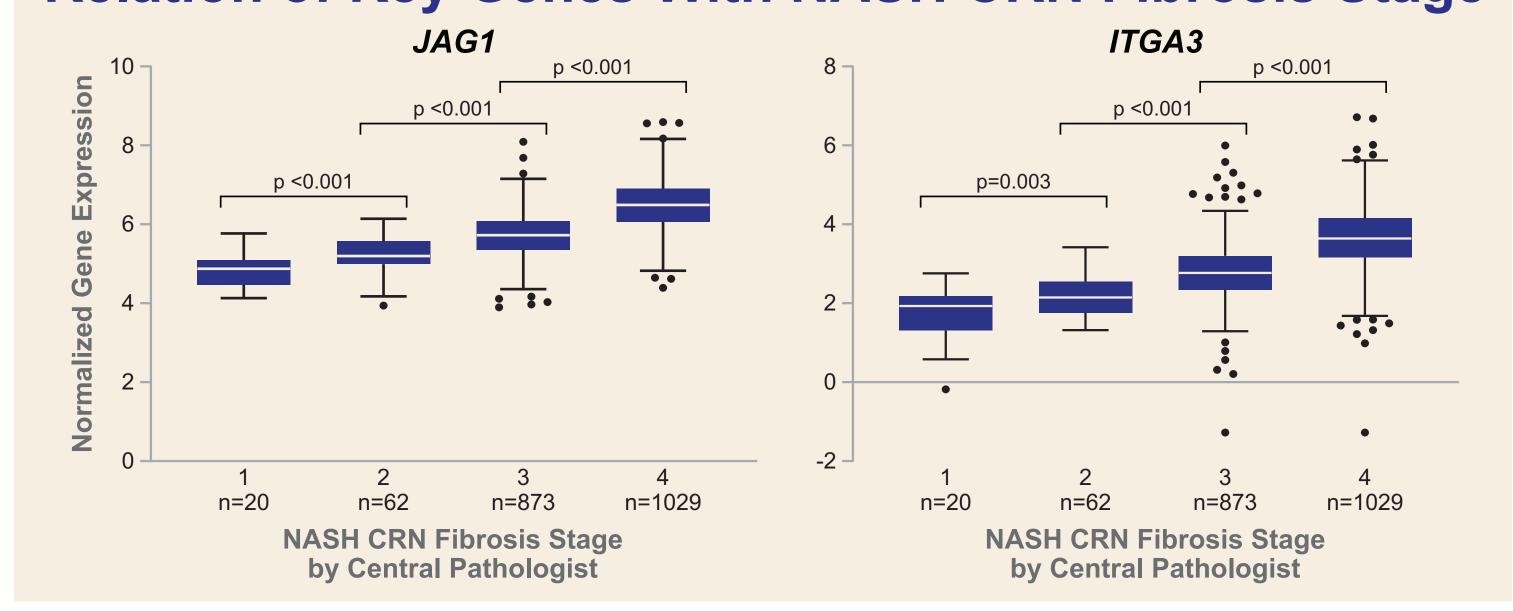

*Sizes of nodes represent degrees; colors of nodes represent connectivity of genes within integrative network.

Association of Key Genes With Portal Inflammation and Bile Duct Area

Symbol	Gene Name	KEGG* Pathway
JAG1	Jagged1	Notch signaling pathway; TNF signaling pathway; Apelin signaling pathway
CFTR	Cystic fibrosis transmembrane conductance regulator	AMPK signaling pathway; cAMP signaling pathway
ITGA3	Integrin-α3	PI3K-Akt signaling pathway; Focal adhesion; ECM-receptor interaction
SLC12A2	Solute carrier family 12, member 2	Pancreatic secretion
TNFRSF21	TNF receptor superfamily member 21	TNF family (DR6)

*Kyoto Encyclopedia of Genes and Genomes. AMPK, adenosine monophosphate-activated protein kinase; cAMP, cyclic adenosine monophosphate; DR6, death receptor-6; ECM, extracellular matrix; PI3K-Akt, phosphatidylinositol-3-kinase–protein kinase B; TNF, tumor necrosis factor.
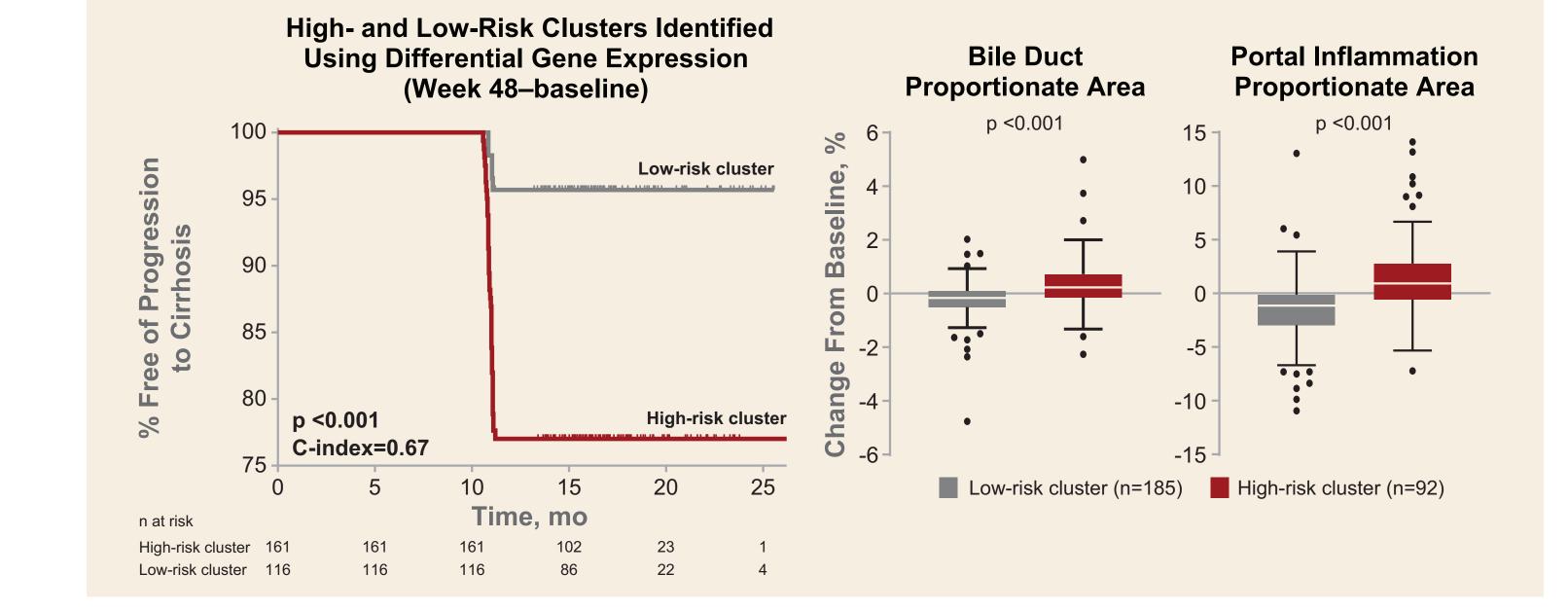
- The expression of 5 key genes was strongly associated with both portal inflammation and bile duct area (*JAG1:* Spearman ρ=0.71, 0.62; *CFTR:* ρ=0.68, 0.63; *ITGA3:* ρ=0.68, 0.60; *SLC12A2:* ρ=0.68, 0.58; and *TNFRSF21:* ρ=0.67, 0.56 [all FDR-corrected p <0.001])

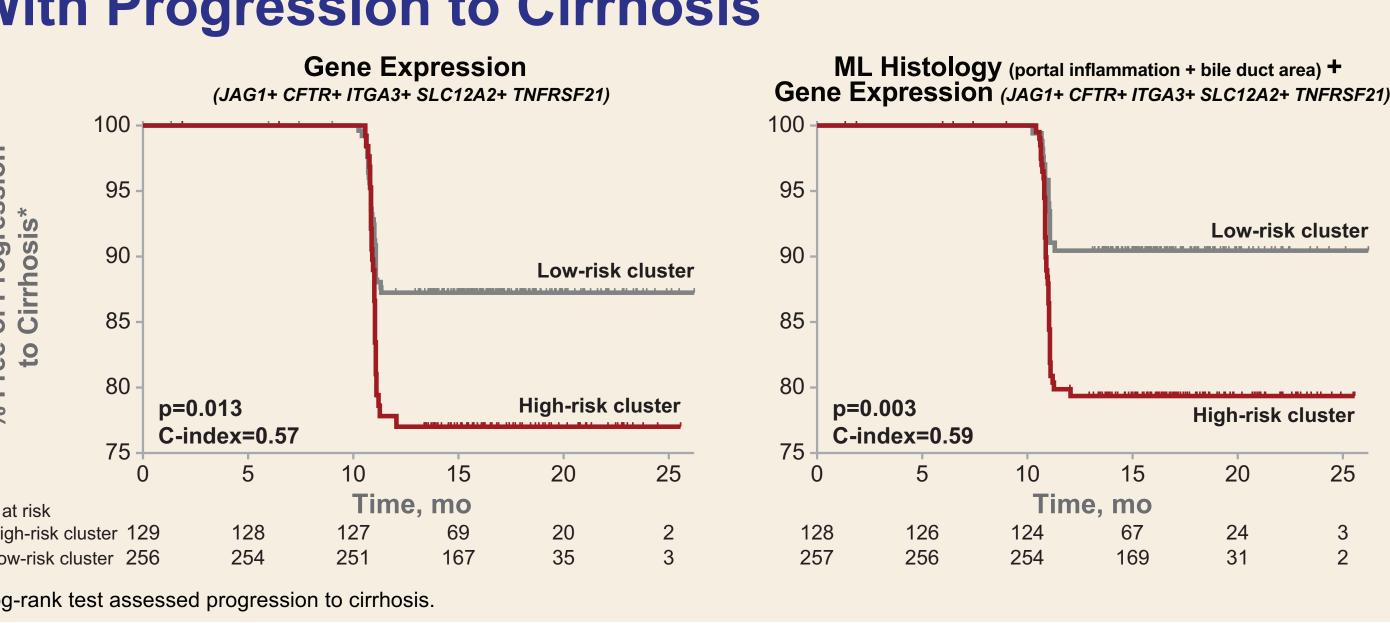### Relation of Key Genes With NASH CRN Fibrosis Stage

- Expression of each key gene significantly correlated with NASH CRN fibrosis stage
- A signature incorporating expression of these 5 genes (1st principal component) was positively associated with NASH CRN fibrosis stage (ρ=0.59; p <0.01)

Subjects at Risk for Progression to Cirrhosis—Identified Using Differential Expression of 5 Key Genes—Had Increased Bile Duct Proliferation and Portal Inflammation From Baseline to Week 48
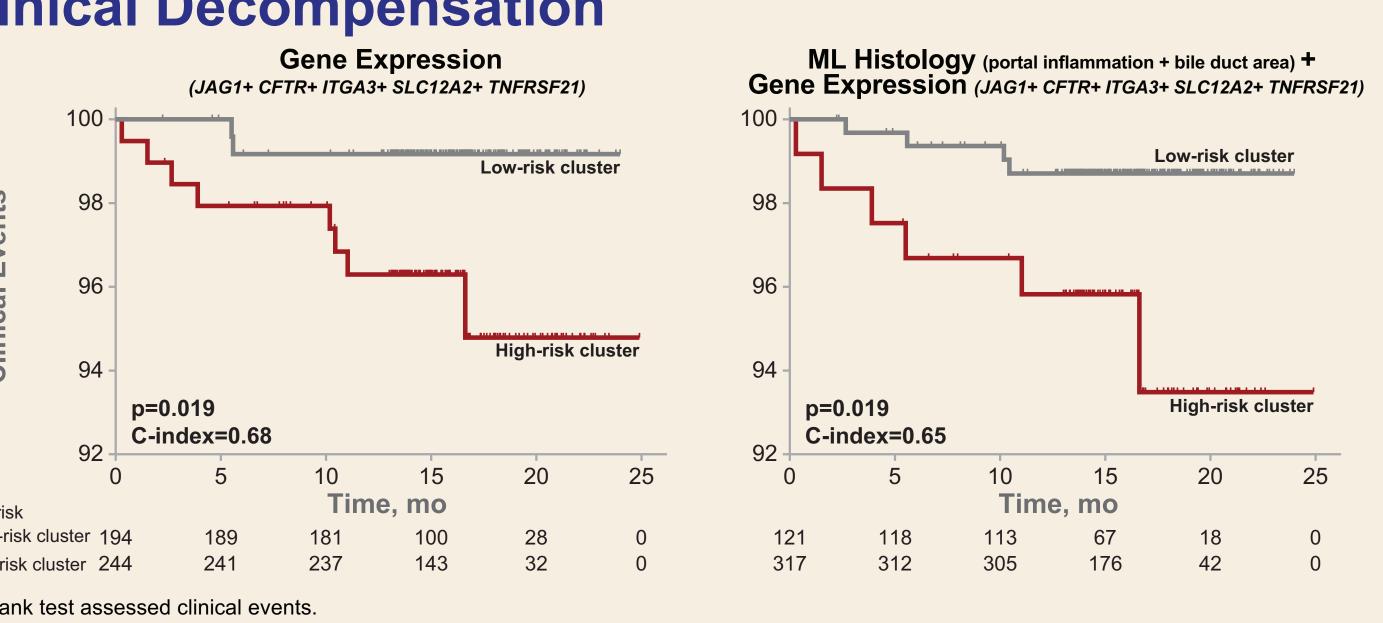

ML Histology and a 5-Gene Signature Were Associated With Progression to Cirrhosis

*Log-rank test assessed progression to cirrhosis.

- 16% (61/385) of subjects with bridging (F3) fibrosis at baseline progressed to cirrhosis
- F3 subjects in the high-risk cluster had increased risk of progression to cirrhosis (hazard ratio [HR] 1.87 [95% confidence interval (CI) 1.13, 3.09] for gene expression and HR 2.11 [95% CI 1.25, 3.49] for ML histology + gene expression)

ML Histology and a 5-Gene Signature Differentiated NASH Cirrhosis Subjects at High and Low Risk of Clinical Decompensation

*Log-rank test assessed clinical events.

- During a median follow-up of 15.6 mo, 2% (10/438) of subjects with cirrhosis at baseline had liver-related events
- F4 subjects in the high-risk cluster had an increased risk of clinical events (HR 4.05 [95% CI 1.40, 14.36] for gene expression, and HR 1.66 [95% CI 1.12, 24.85] for ML-derived portal inflammation and bile duct proliferation + gene expression)
- Performance of models for predicting progression to cirrhosis and liver-related clinical events using either 5 key genes alone or combination of 5 key genes with histologic features was similar

Conclusions

- ML-based quantification of NASH histology and integration with transcriptomic data has identified genomic predictors of disease progression in subjects with advanced fibrosis due to NASH participating in the STELLAR trials
- This approach identifies the importance of baseline portal inflammation and ductular reaction, and their increase as hallmark features of disease progression
- A 5-gene signature based on genomic changes associated with portal inflammation and ductular reaction differentiated subjects at low and high risk of clinical events
- These findings have implications for risk stratification based on histology and may help identify potential new targets to reduce NASH progression

References: 1. Brunt EM, et al. Hepatology 2009;49:809-82. 2. Neuschwander-Tetri BA, et al. Hepatology 2010;52:913-24. 3. Richard MM, et al. Gastroenterology 2007;133:80-90. 4. Pokkalla H, et al. AASLD 2019, abstr 187. 5. Gerhard GS, et al. J Endocr Soc 2018;2:710-26. 6. Harrison SA, et al. J Hepatology 2020;73:26-39.