ASSESSMENT OF SERUM LEVELS OF CHITINASE-3-LIKE PROTEIN 1 (CHI3L1) IMPROVES IDENTIFICATION OF THE NASH PATIENTS AT RISK WHO SHOULD BE TREATED

Arun J. Sanyal[2] , Stephen Harrison[6], Genevieve Cordonnier[1], John Brozek[1], Alice Roudot[1], Emilie Praca[1], Fouad Ben Sudrik[1], Sophie Megnien[1], Remy Hanf[1], Bart Staels[3], Pierre Bedossa[4] , Vlad Ratziu[5], Dean Hum[1], Raphael Darteil[1]

[1]GENFIT, Loos, France, [2]Virginia Commonwealth University, Richmond, United States, [3]Univ Lille INSERM U1011, Institut Pasteur de Lille, Lille, France, [4]Department of Pathology, Hôpital Beaujon, Paris, France, [5]Hôpital Pitié Salpêtrière, Paris, France, [6]Pinnacle Clinical Research, San Antonio, TX, United States

GENFIT
TOWARDS BETTER MEDICINE

BACKGROUND AND AIMS

New non-invasive diagnostic tools dedicated to the identification of NASH patients at risk of liver outcomes are needed. As defined by the regulatory agencies for the development of new NASH drugs (FDA and EMA), patients at risk of liver outcomes who should be pharmacologically treated are those with NAS≥4 (with score≥1 for steatosis, lobular inflammation and ballooning) and NASH-CRN fibrosis score (F) ≥2.

Based on these characteristics of the patients to be treated and using the GOLDEN-505 cohort, we recently reported a discriminating algorithm including five independent variables including two miRNAs (mir34a and mir200a), alpha2-macroglobulin, PIIINP and HbA1C (Sanyal *et al.*, J.Hepatol.2016, vol 64, S717).

Among the emerging biomarkers for chronic liver diseases is Chitinase-3-like protein 1 (CHI3L1 also known as YKL-40), a secreted glycoprotein. Elevated serum YKL-40 levels have been proposed as a marker of liver fibrosis (Johanssen, 2000; Tao, 2014; Kjaergaard, 2016).

The aims of this study were:

- To compare the diagnostic performance of CHI3L1 to that of other individual variables in the identification of NASH patients to be treated.
- To assess the diagnostic performance of a new CHI3L1-including algorithm in comparison with our previous algorithm and other scoring systems.

METHODS

PATIENTS AND DATA SET

- Data set and plasma samples from the 274 biopsy proven NASH patients included in the GOLDEN505 phase IIb trial with elafibranor was used.
- This trial included a wide spectrum of NASH disease activity (NAS=3-8) and fibrosis (F0-F3). Cirrhotic patients (F4) were excluded.
- Initial dataset at inclusion included more than 100 variables:
  - Sex, age, BMI, concommitent T2D…
  - Classical biochemistry/hematology:
    - Liver enzymes: ALT, AST, GGT, Alkaline Phosphatase (ALK)
    - Lipids and apolipoproteins: Tot-Chol, LDL-C, HDL-C, Triglycerides, ApoB, ApoA1…
    - Glucose metabolism: FPG, Insulin, HOMA-IR, Hb1AC, FFA, fructosamine
  - Oriented biochemistry:
    - Inflammatory markers: CRP, IL6, TNFα, IL1β, PAI1…
    - Apoptosis markers: CK18-M30, CK18-M65
    - Fibrosis markers: Alpha2-macroglobulin, hyaluronic acid, PIIINP, TIMP1
  - Biopsy scoring (centralized reading):
    - Steatosis, lobular inflammation, hepatocyte ballooning, NAS
    - Fibrosis score
- Additional variables:
  - Serum levels of 9 different miRNAs measured by RT-qPCR
  - Serum level of Chitinase-3-like protein 1 (CHI3L1) measured by using Elisa Test.

BIOSTATISTICAL APPROACHES AND ALGORITHMS

- Two independent biostatistical approaches (Median and Bootstrap) were used to generate diagnostic algorithms discriminating patients To-Be-Treated (TBT) from patients not-to-be-treated (NTBT).
- In both approaches, a stepwise logistic regression method was used to generate the most discriminating algorithm in the dataset and/or selected subsets.
- The Median approach:
  - 27000 equilibrated patient partitionings of the GOLDEN-505 cohorte (2/3 training; 1/3 confirmation) generated 27000 corresponding algorithms.
  - Variables were ranked by measuring their frequency of appearance in 27000 algorithms.
  - A median algorithm has been then obtained by calculating median coefficient of the most frequent variables.
  - The robustness of the median algorithm was then tested in the 27000 confirmation sets.
- The Bootstrap approach:
  - The most powerfull algorithm has been obtained from the complete dataset.
  - Robustness of this algorithm was then tested by a bootstrapping approach (1000 bootstrapping generated datasets).

ROC ANALYSIS

- Diagnostic performance of plasma level of CHI3L1 to discriminate To-Be-Treated (TBT=NAS≥4; F≥2) from Non-To-Be-Treated (NTBT) patients were compared to other variables using ROC analysis.
- Algorithms with and without CHI3L1 obtained by two different statistical approaches were also compared to discriminate TBT and NTBT through ROC analysis.

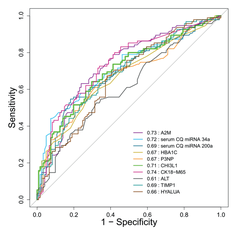
DISCRIMINATING INDIVIDUAL VARIABLES IN GOLDEN-505

	Histology and patient characteristics			
	NTBT mean (SD)	TBT mean (SD)	p value	ALL mean (SD)
Patient number	121	95		216
Steatosis (0-3)	2.12 (0.79)	2.44 (0.63)	0,0016	2.26 (0.74)
Lobular inflammation (0-3)	1.16 (0.37)	1.73 (0.63)	< 0,0001	1.41 (0.57)
Hepatocyte Ballooning (0-2)	1.23 (0.42)	1.57 (0.50)	< 0,0001	1.38 (0.49)
NAS (0-8)	4.52 (1.10)	5.74 (1.00)	< 0.0001	5.06 (1.22)
Fibrosis (0-4)	0.84 (0.67)	2.45 (0.50)	< 0.0001	1.55 (1.00)
Female/Male	53/68	43/52		96/120
Age (years)	50.36 (11.66)	54.31 (11.30)	0.0132	52.1 (11.6)
BMI (kg/m2)	30.9 (4.94)	31.6 (4.43)	0.2838	31.2 (4.72)
Diabetic (%)	27.3%	48.4%		36.6%

	Biomarkers			
	NTBT mean (SD)	TBT mean (SD)	p value	ALL mean (SD)
Number of Patients	121	95		216
ALT (UI/L)	56.07 (30.12)	70.87 (43.23)	0.0034	62.58 (37.11)
AST (UI/L)	35.12 (15.36)	51.45 (28.97)	< 0.0001	42.31 (23.76)
Alkaline Phosphatase (UI/L)	74.64 (19.16)	76.02 (22.75)	0,63	75.25 (20.78)
GGT (UI/L)	69.49 (72.94)	85.6 (99.87)	0.1722	76.57 (85.99)
Fasting Glucose (mM)	1.21 (0.43)	1.65 (0.67)	0.0006	1.40 (0.59)
HOMA-IR	6.07 (7.04)	9.10 (9.90)	0.0093	7.40 (8.53)
HbA1c (%)	5.77 (0.63)	6.37 (1.08)	< 0.0001	6.03 (0.90)
CK18-M30	321.25 (168.75)	488.99 (266.75)	< 0.0001	395.02 (232.31)
CK18-M65	384.39 (360.44)	718.13 (653.22)	< 0.0001	531.17 (535.39)
alpha2-macroglobulin	2.05 (0.80)	2.69 (0.87)	< 0.0001	2.33 (0.89)
PIIINP	8.08 (2.18)	10.51 (4.85)	< 0.0001	9.15 (4.02)
TIMP-1	239.2 (44.67)	274.9 (67.57)	< 0.0001	254.9 (58.53)
Hyaluronic acid	53.42 (75.74)	77.83 (71.93)	0.0171	64.15 (74.91)
CHI3L1 (ng/mL)	**51.8 (38.3)**	**94.1 (72.9)**	**< 0.0001**	**70.4 (59.9)**

- **Table 1: Characteristics in Not-To-Be-Treated Patients (NTBT) and To-Be-Treated Patients (TBT).**
  Variables were expressed as Means (Standard Deviation) and compared using the Student t-test for the statistical significance.
  1A: Histology and anthopometry; 1B: plasma markers
- **Compared to NTBT, TBT patients:**
  - Are older and more often diabetics
  - Have impaired glucose homeostasis, but comparable levels of plasma lipids and apoliproteins (data not shown).
  - Have higher levels of ALT and AST but comparable levels of GGT and alkaline phosphatase.
  - Have higher levels of apoptosis markers (CK18) and fibrosis associated markers (alpha2-macroglobulin, PIIINP, TIMP1)
  - **Have significantly higher levels of plasma CHI3L1**

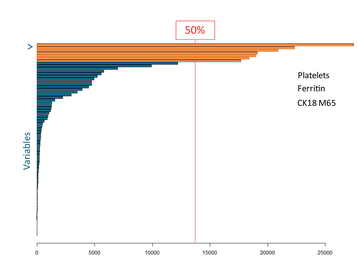
ROC curves of individuals markers



- **Figure 1: ROC curves of biomarkers differentially expressed in TBT and NTBT.**
  - Plasma CHI3L1 discriminates TBT and NTBT in GOLDEN-505 cohort (AUROC=0.71)
  - Plasma CHI3L1 has a diagnostic performance comparable to that of CK18 (apoptosis marker) , HbA1C (glucose homeostasis) or fibrosis associated biomarkers like alpha2-macroglobulin or PIIINP.
  - Plasma CHI3L1, mir34a, miR200a have comparable diagnostic performances.

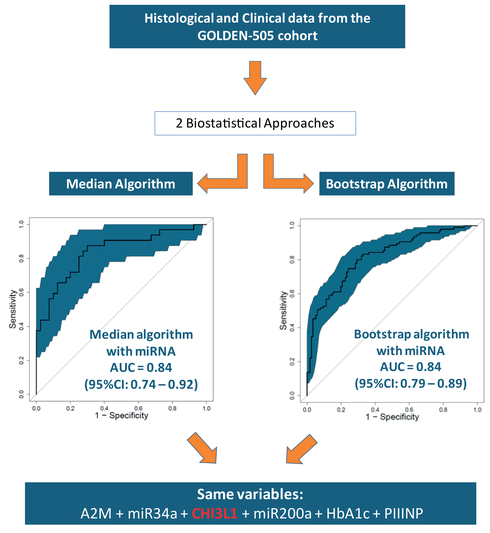
MEDIAN AND BOOTSTRAP ALGORITHMS

Ranking of variables (median approach)



- **Figure 2: Ranking of individual variables (median approach).**
  After removing co-linear variables from the dataset, the frequency of each variable in 27,000 algorithms generated from 27,000 training sets was calculated.
  - CHI3L1 is the second most represented variable after miR200a.
  - All other variables displaying a frequency >50% were those included in our previously reported algorithm: miR200a, PIIINP, alpha2-macroglobulin, HbA1C and miR34a.
  - All variables displaying a frequency >50% were used to establish a "median algorithm".

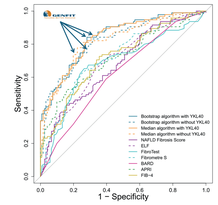
Median vs bootstrap algorithms



- **Figure 3: ROC analysis and robustness of median and bootstrap algorithms.**
  - The bootstrap approach confirmed CHI3L1 as a component of the most discriminating algorithm obtained using the whole dataset.
  - The median and bootstrap approaches identified the same explicative variables.
  - The robustness of the CHI3L1 including median algorithm was confirmed when tested on the 27,000 confirmatory sets
  - Performance and robustness of the CHI3L1 including bootstrap algorithm were comparable to median algorithm

COMPARISON OF ALGORITHMS WITH EXISTING SCORES

ROC curves of Median and Bootstrap algorithms with and without CHI3L1 vs other existing scores



Algorithms	AUC	Cut-off	Accuracy (%)	Sensitivity (%)	Specificity (%)	PPV (%)	NPV (%)
Median model with CHI3L1	**0.84**	**0.2302**	**76**	**80**	**74**	**70**	**82**
Bootstrap model with CHI3L1	**0.84**	**0.3502**	**77**	**84**	**72**	**70**	**85**
Median model without CHIDL1	0.82	0.227	79	81	77	74	83
Bootstrap model without CHIL1	0.82	0.4343	75	75	76	72	79
Angulo Index	0.68	0,676	35	67	10	37	28
ELF	0.70	7,7	45	98	4	45	71
FibroTest	0.68	0,48	66	42	85	69	65
Fibrometre S	0.72	0,62	64	33	88	69	63
BARD	0.62	2	60	59	60	54	65
APRI	0.72	1	67	28	98	90	63
FIB-4	0.72	3,25	59	7	100	100	58

For all algorithms, cut-offs are the most efficient in GOLDEN-505 cohorte.

- **Table 2: Comparison of performances of Median and Bootstrap algorithms with and without CHI3L1 vs other existing scores.**
  - CH3L1 including algorithm has slightly better diagnostic performances (accuracy, sensitivity, specificity, PPV and NPV) than a comparator algorithm without CHI3L1.
  - Compared to other previously reported algorithms, our median and bootstrap algorithms with and without CHI3L1 have higher diagnostic performances for discriminating TBT and NTBT in GOLDEN-505 cohort.

CONCLUSION

- This study demonstrates that individually plasma CHI3L1 as a diagnostic value to identify patients who should be treated in the GOLDEN-505 cohort.
- The inclusion of circulating levels of CHI3L1 in both the Median and the Boostrap algorithms slightly, but not significantly enhances the diagnostic performances of our previously described algorithms.
- Median and bootstrap algorithms with or without CHI3L1 showed significantly better performances than existing scoring systems such as the NAFLD Fibrosis score, the ELF test, the Fibrotest, the Fibrometer and FIB-4.

REFERENCES

Johansen JS, Christoffersen P, Møller S, Price PA, Henriksen JH, Garbarsch C, Bendtsen F. (2000)Serum YKL-40 is increased in patients with hepatic fibrosis. J Hepatol. 32(6):911-20.
Kjaergaard AD, Johansen J. S. , Bojesen S. E. , Nordestgaard BG (2016) Role of inflammatory marker YKL-40 in thediagnosis, prognosis, and cause of cardiovascularand liver diseases Crit Rev Clin Lab Sci
Tao H, Yang J-J, Shi K-H, Zhang C, Lv X-W, Li J(2014) The significance of YKL-40 protein in liver fibrosis. Inflamm. Res. 63:249–254