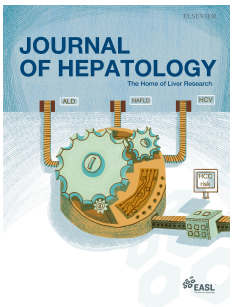


Rare *ATG7* genetic variants predispose patients to severe fatty liver disease

Guido A. Baselli, Oveis Jamialahmadi, Serena Pelusi, Ester Ciociola, Francesco Malvestiti, Marco Saracino, Luigi Santoro, Alessandro Cherubini, Paola Dongiovanni, Marco Maggioni, Cristiana Bianco, Federica Tavaglione, Annalisa Cespiati, Rosellina M. Mancina, Roberta D'Ambrosio, Valentina Vaira, Salvatore Petta, Luca Miele, Umberto Vespasiani-Gentilucci, Alessandro Federico, Jussi Pihlajamaki, Elisabetta Bugianesi, Anna L. Fracanzani, Helen L. Reeves, Giorgio Soardo, Daniele Prati, Stefano Romeo, Luca VC. Valenti, EPIDEMIC Study Investigators

PII: S0168-8278(22)00207-0

DOI: https://doi.org/10.1016/j.jhep.2022.03.031

Reference: JHEPAT 8666

To appear in: *Journal of Hepatology*

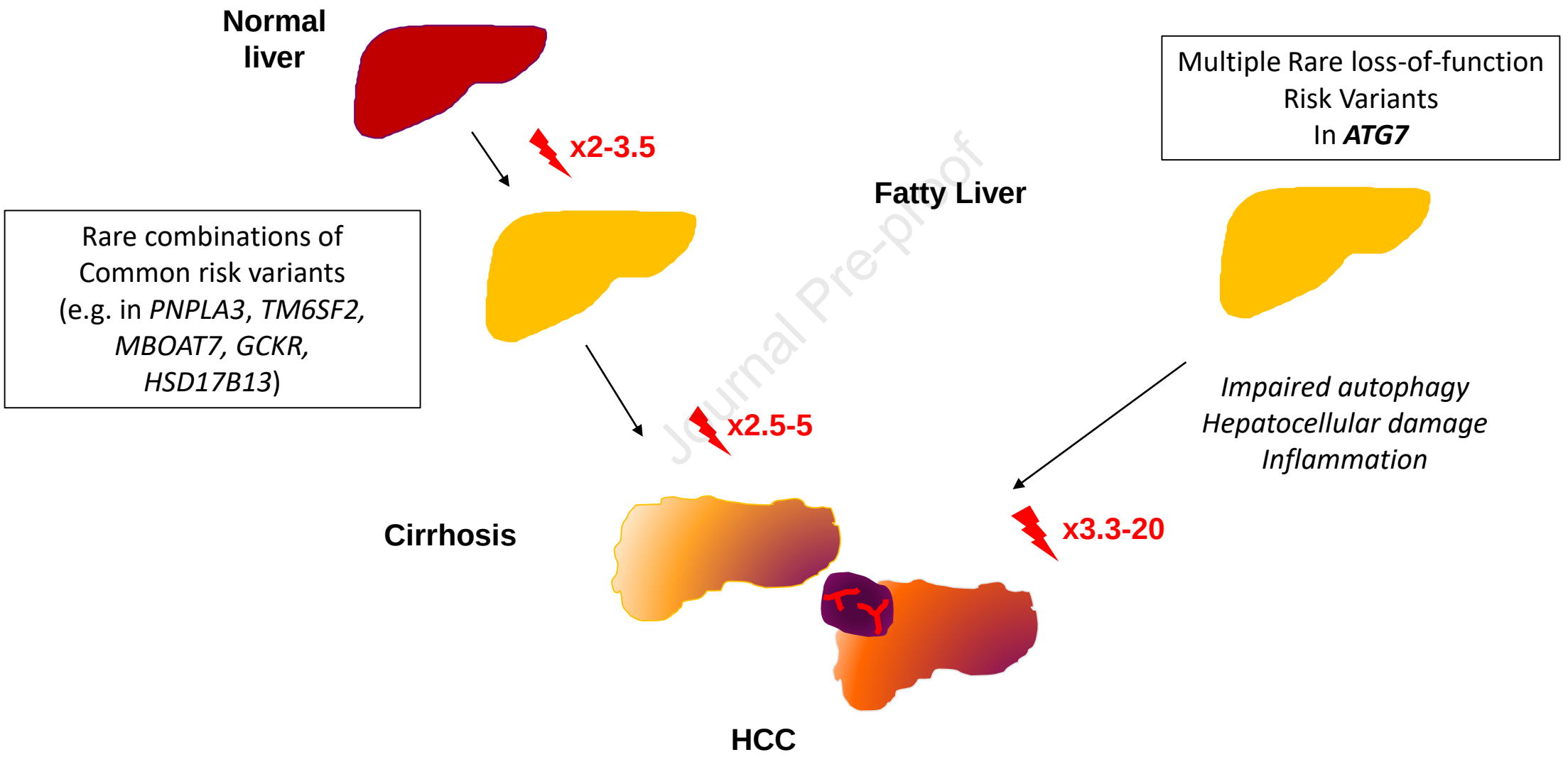
Received Date: 27 July 2021

Revised Date: 25 February 2022

Accepted Date: 15 March 2022

Please cite this article as: Baselli GA, Jamialahmadi O, Pelusi S, Ciociola E, Malvestiti F, Saracino M, Santoro L, Cherubini A, Dongiovanni P, Maggioni M, Bianco C, Tavaglione F, Cespiati A, Mancina RM, D'Ambrosio R, Vaira V, Petta S, Miele L, Vespasiani-Gentilucci U, Federico A, Pihlajamaki J, Bugianesi E, Fracanzani AL, Reeves HL, Soardo G, Prati D, Romeo S, Valenti LV, EPIDEMIC Study Investigators, Rare *ATG7* genetic variants predispose patients to severe fatty liver disease, *Journal of Hepatology* (2022), doi: https://doi.org/10.1016/j.jhep.2022.03.031.

INDIVIDUALS WITH DYSMETABOLISM:



Rare *ATG7* genetic variants predispose patients to severe fatty liver disease

Guido A Baselli[1,2]†, Oveis Jamialahmadi[3], Serena Pelusi[1,2], Ester Ciociola[3], Francesco Malvestiti[2], Marco Saracino[2], Luigi Santoro[1], Alessandro Cherubini[1], Paola Dongiovanni[4], Marco Maggioni[5], Cristiana Bianco[1], Federica Tavaglione[3], Annalisa Cespiati[4], Rosellina M Mancina[3], Roberta D'Ambrosio[6], Valentina Vaira[2,5], Salvatore Petta[7], Luca Miele[8], Umberto Vespasiani-Gentilucci[9], Alessandro Federico[10], Jussi Pihlajamaki[11,12], Elisabetta Bugianesi[13], Anna L Fracanzani[2,4], Helen L Reeves[14], Giorgio Soardo[15,16], Daniele Prati[2], EPIDEMIC Study Investigators[17], Stefano Romeo[3,18,19]‡, Luca VC Valenti[1,2]‡*.

[1]Precision Medicine – Department of Transfusion Medicine and Hematology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico; Milan, Italy,

[2]Department of Pathophysiology and Transplantation, Università degli Studi di Milano; Milan, Italy,

[3]Department of Molecular and Clinical Medicine, Institute of Medicine, Sahlgrenska Academy, Wallenberg Laboratory, University of Gothenburg; Gothenburg, Sweden,

[4]General Medicine and Metabolic Diseases, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico; Milan, Italy,

[5]Pathology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico; Milan, Italy,

[6]Division of Gastroenterology and Hepatology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico; Milan, Italy,

[7]Gastroenterology and Hepatology, PROMISE, Università di Palermo; Palermo, Italy,

[8]Department of Internal Medicine, Fondazione Policlinico A. Gemelli, Università Cattolica di Roma; Rome, Italy,

[9]Clinical Medicine and Hepatology Unit, Department of Internal Medicine and Geriatrics, Campus Bio-Medico University; Rome, Italy,

[10]Division of Hepatogastroenterology, Department of Precision Medicine, Università della Campania "Luigi Vanvitelli"; Naples, Italy,

[11]Institute of Public Health and Clinical Nutrition, University of Eastern Finland; Kuopio, Finland,

[12]Department of Medicine, Endocrinology and Clinical Nutrition, Kuopio University Hospital; Kuopio, Finland,

[13]Department of Medical Sciences, Division of Gastro-Hepatology, A.O. Città della Salute e della Scienza di Torino, Università di Torino; Turin, Italy,

[14]Translational and Clinical Research Institute, Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK; & Newcastle NIHR Biomedical Research Centre, Newcastle upon Tyne Hospitals NHS Foundation Trust; Newcastle upon Tyne, UK,

[15]Clinic of Internal Medicine – Liver Unit, Department of Medical Area (DAME), Università degli Studi di Udine; Udine, Italy,

[16]Italian Liver Foundation, Area Science Park, Basovizza Campus; Trieste, Italy,

[17]EPIDEMIC Study Investigators also include Vittorio Borroni, Antonio Liguori, Luisa Ronzoni, Alessandro Cherubini, Luigi Santoro, Melissa Tomasi, Angela Lombardi, Mahnoosh Ostadreza, Elia Casirati, Ilaria Marini, Silvia Maier, Chiara Rosso, Gianluca Svegliati Baroni, Carlo Santaniello, Marcello Dallio,

[18]Clinical Nutrition Unit, Department of Medical and Surgical Science, University Magna Graecia; Catanzaro, Italy,

[19]Department of Cardiology, Sahlgrenska University Hospital; Gothenburg, Sweden;

†Current address: SciLifeLab, Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Solna, Sweden

‡ Shared senior authors

*__Corresponding author:__
Luca Valenti, MD
Department of Pathophysiology and Transplantation, Università degli Studi di Milano; Milan, Italy,
Precision Medicine – Department of Transfusion Medicine and Hematology, Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico; Milan, Italy,
via Francesco Sforza 35, 20122, Milan, Italy
Email: luca.valenti@unimi.it

Word count: 6549 (including references and Tables, see Letter to the Editors); Figures: 5; Tables: 3; References: 38

Funding:

Italian Ministry of Health (Ministero della Salute), Ricerca Finalizzata RF-2016-02364358 ("Impact of whole exome sequencing on the clinical management of patients with advanced nonalcoholic fatty liver and cryptogenic liver disease"), (LV)

Fondazione IRCCS Ca' Granda Ospedale Maggiore Policlinico, Ricerca corrente LV, DP, ALF)

Fondazione IRCCS Ca' Granda core COVID-19 Biobank (RC100017A), "Liver BIBLE" (PR-0391) (LV)

Innovative Medicines Initiative 2 joint undertaking of European Union's Horizon 2020 research and innovation programme and EFPIA European Union (EU) Programme Horizon 2020 (under grant agreement No. 777377) for the project LITMUS (LV, EB, SP, LM)

The European Union, programme "Photonics" under grant agreement "101016726" (LV)

Gilead_IN-IT-989-5790 (LV, EB)

Swedish Research Council [Vetenskapsradet (VR), 2021-005208](SR)

Swedish state under the agreement between the Swedish government and the county councils (the ALF-agreement) [SU 2018-04276] (SR)

Novo Nordisk Foundation Grant for Excellence in Endocrinology [Excellence Project, 9321-430] (SR),

Swedish Diabetes Foundation [DIA2020-518] (SR)Swedish Heart Lung Foundation [20200191] (SR)

Wallenberg Academy Fellows from the Knut and Alice Wallenberg Foundation [KAW 2017.0203] (SR)

Astra Zeneca Agreement for Research (SR),

Swedish Foundation for Strategic Research (SSF) [ITM17-0384] (SR),

Novo Nordisk Project Grants in Endocrinology & Metabolism - Nordic Region 2020 [NNF20OC0063883] (SR) Cancer Research UK (CR UK) centre grant C9380/A18084, programme grant C18342/A23390 and Accelerator award C9380/A26813 (HR).

Author contributions:

Conceptualization: LV, SR, GB, SP
Methodology: LV, SR, GB, SP, OJ
Investigation: GB, OJ, SP, EC, FM, PD, MM, CB, FT, AC, RMM, RDA, SP, LM, UVG, AF, JP, EB, ALF, HR, GS, MS, LS, AC, VV
Visualization: GB, OJ, FM, MM, AC, LV
Funding acquisition: LV, SR, DP, HR, EB, SP, LM
Project administration: LV
Supervision: LV, SR
Writing – original draft: GB, LV
Writing – review & editing: LV, SR, OJ, FM, RMM

Competing interests: The authors declare that they have no conflict of interest relevant to the present study.

LV has received speaking fees from MSD, Gilead, AlfaSigma and AbbVie, served as a consultant for Gilead, Pfizer, AstraZeneca, Novo Nordisk, Intercept, Diatech Pharmacogenetics and Ionis Pharmaceuticals, and received research grants from Gilead.

SR has served as a consultant for AstraZeneca, Celgene, Sanofi, Amgen, Akcea Therapeutics, Camp4, AMbys, Medacorp and Pfizer in the past 5 years, and received research grants from AstraZeneca, Sanofi and Amgen.

Data and materials availability: The ethical approval of the study does not allow to publicly share individual patients' genetic data. All data, code, and materials used in the analysis are available upon reasonable request for collaborative studies regulated by materials/data transfer agreements (MTA/DTAs) to the corresponding author. The frequencies of genetic variants in the EPIDEMIC-2021 cohort are reported in the open access Mendeley database (doi:10.17632/gff2ftfhmr.1)

ABSTRACT

Background & Aims: Nonalcoholic fatty liver disease (NAFLD) is the leading cause of liver disorders and has a strong heritable component. The aim of this study was to identify new loci contributing to severe NAFLD by examining rare variants.

Methods: We performed whole-exome sequencing in individuals with NAFLD and advanced fibrosis or hepatocellular carcinoma (n=301) and examined the enrichment of likely pathogenic rare variants vs. the general population, followed by validation at gene level.

Results: In patients with severe NAFLD, we observed an enrichment of the p.P426L variant (rs143545741 C>T; OR 5.26, 2.1-12.6; p=0.003) of autophagy-related 7 (*ATG7)*, which we characterized as a loss-of-function, vs. the general population, and an enrichment in rare variants affecting the catalytic domain (OR 13.9, 1.9-612; p=0.002). In the UK Biobank cohort, loss-of-function *ATG7* variants increased the risk of cirrhosis and hepatocellular carcinoma (OR 3.30, 1.1-7.5 and OR 12.30, 2.6-36, respectively; p<0.001 for both). The low-frequency loss-of-function p.V471A variant (rs36117895 T>C) was also associated with severe NAFLD in the clinical cohort (OR=1.7, 1.2-2.5; p=0.003), predisposed to hepatocellular ballooning (p=0.007) evolving to fibrosis in a Liver biopsy cohort (n=2268), and was associated with liver injury in the UK Biobank (AST levels, p<0.001), with a larger effect in severely obese individuals where it was linked to hepatocellular carcinoma (p=0.009). ATG7 protein localized to periportal hepatocytes, more so in the presence of ballooning. In the Liver Transcriptomic cohort (n=125) *ATG7* expression correlated with suppression of the TNFα pathway, which was conversely upregulated in p.V471A carriers.

Conclusions: We identified rare and low-frequency *ATG7* loss-of-function variants as modifiers of NAFLD progression by impairing autophagy and facilitating ballooning and inflammation.

Keywords: NAFLD; NASH; genetics; liver fibrosis; autophagy

LAY SUMMARY

- We found that rare mutations in a gene called autophagy related (*ATG7*) increase the risk of developing severe liver disease in individuals with dysmetabolism.
- These mutations cause an alteration in protein function and impairment of self-renewal of cellular content, leading to liver damage and inflammation.

INTRODUCTION

Nonalcoholic fatty liver disease (NAFLD) is most frequently associated with metabolic dysfunction and is the leading cause of liver-related mortality worldwide [1]. Fatty liver disease is a spectrum of conditions ranging from a mild liver fat accumulation to steatohepatitis characterized by ballooning degeneration of hepatocytes, inflammation and fibrogenesis, to life-threatening conditions at the end of the disease spectrum, namely advanced liver fibrosis and hepatocellular carcinoma (HCC). Fatty liver disease has a strong heritable component, and common germline variants regulating the hepatic remodeling, synthesis, and secretion of lipids account for a fraction of the risk [2]. The identification of these variants has allowed improving individual risk stratification and identifying novel therapeutic targets [2, 3]. However, a substantial fraction of NAFLD heritability remains unexplained [2, 4].

Within this context, we aimed to identify new inherited risk factors contributing to NAFLD by using whole exome sequencing (WES) followed by using the following strategy: we first examined a well-defined trait at the extreme of the NAFLD spectrum, namely advanced fibrosis and HCC, and next prioritized variants based on multiple *in silico* predictions before conducting the enrichment analysis of rare variants. The rationale for this approach was that rare variants with a large effect on protein function may contribute to variation in the susceptibility to develop severe NAFLD.

Indeed, most rare genetic variants have been proposed to be deleterious in humans with an effect size that is inversely related to their population frequency [5, 6]. WES studies have identified a handful of genes [6] whose rare variants primarily associate with a complex trait. However, a) the large number of rare variants identified; b) the limited prioritization of these variants; c) the dilution of the effect size of these variants in poorly defined or mild phenotypes

have so far limited the statistical power of WES studies. As a proof-of-principle, by using a candidate gene approach, we identified an enrichment of rare variants in Apolipoprotein (*APOB*) in individuals with severe NAFLD [7].

Here, we first identified an enrichment of rare variants of autophagy-related 7 (*ATG7*) in patients with severe NAFLD. Strikingly, deletion of *ATG7* has previously been shown to facilitate liver injury due to defective autophagy and activation of inflammation [8-10]. We next validated the association of *ATG7* variants with advanced NAFLD in at-risk individuals and in a population-based cohort. Finally, we showed that ATG7 is expressed in hepatocytes and linked to hepatocellular damage, specifically ballooning degeneration.

PATIENTS AND METHODS

Study cohorts

The study cohorts and design are presented in Figure S1. The discovery EPIDEMIC cohort included 72 Italian patients with NAFLD-hepatocellular carcinoma, who were enrolled during 2010-2015. The EPIDEMIC validation cohort was composed by further 129 patients with advanced NAFLD including 59 Italian patients with severe fibrosis (stage F3-F4) and 70 patients with NAFLD-hepatocellular carcinoma from the UK [7]. The validation cohort included additional 100 Italian patients with severe fibrosis (PERSPECTIVE cohort). Secondary causes of FLD were excluded by history, including at-risk alcohol intake (≥30/20 g/day in M/F), use of drugs known to precipitate steatosis, other viral, autoimmune, and genetic liver disorders. Severe NAFLD was defined in the presence of advanced fibrosis (histological stage F3-F4 or clinically overt cirrhosis) or HCC [11]. We also considered a local ethnically matched control group of comparable sex distribution including 50 healthy Italian blood donors without NAFLD or metabolic abnormalities [7, 12]. Clinical features of these individuals (clinical cohort) are presented in Table S1. Single variants allele frequency in the discovery cohort as well as the mutational burden at the *ATG7* locus were compared to those of a sample representative of the European population (Figure S1A, S1B): we considered non-Finnish Europeans (NFE) studied by WES from the genome aggregation consortium (gnomAD, N=56,885) [13]. As a further control group for the burden test analysis, we included 404 non-Finnish European (NFE) healthy individuals from the 1000 genomes project with individual-level genotype data, whose genetic data were processed by the same pipeline of cases [14]. Evaluation of family members is reported in the Supplementary methods.

The European NAFLD Liver biopsy cohort included 1946 European individuals undergoing a liver biopsy for suspected nonalcoholic steatohepatitis [3, 15]. Clinical features of the patients stratified by *ATG7* p.V471A and p.P426L genotypes are reported in Table S2.

The Liver-Bible-2021 cohort is a Milan cohort, where healthy individuals with at least three features of metabolic syndrome were phenotyped non-invasively for the presence of liver damage by vibration controlled transient elastography and for metabolic traits [16]. Cohort composition and characterization is presented in the Supplementary Methods and Table S3. Fatty liver was defined for CAP>275 dB/m [17], severe NAFLD for altered values of Fibroscan-AST (FAST) score (>0.35) [18].

The UK Biobank (UKBB) is a large-scale study including more than 500,000 participants (age 40-69 years), recruited between 2006-2010 from 22 assessment centers across the UK; clinical features are presented in Table S4 [19].

The study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki, was approved by the Ethical committees of the involved Institutions and was performed according to the recommendations of the hospitals involved. Informed consent was obtained from each patient.

Clinical cohorts genotyping, transcriptomic analysis in the Transcriptomic cohort [20], liver histology and immunohistochemical analysis and cellular models are described in the Supplementary Methods.

Whole Exome Sequencing, variant calling, prioritization and statistical analysis

The bioinformatic pipeline for variant calling and annotation and quality control have previously been described and reported in detail in the Supplementary methods [7]. The variant

prioritization pipeline is summarized in Fig. S1 and described in the Supplementary Methods. Our purpose was to increase the study power by reducing the number of variants tested (and therefore the denominator when correcting for multiplicity of testing), by selecting variants with a higher likelihood of being pathogenic. To this end, we focused on rare missense/splice variants robustly predicted *in silico* to damage protein function and those in genes with a very low likelihood to carry damaging variants in the population. As a proof of principle, study power gain conferred by the lower testing multiplicity and thus a less stringent alpha level (applying Bonferroni correction) is shown in Fig. S2B.

Genotype to phenotype correlations were assessed by generalized linear models using multivariate binomial logistic, ordinal logistic or linear modelling when appropriate. Burden test analysis is described in the Supplementary Methods.

Statistical analysis was performed using R version 4.0.3.

RESULTS

Enrichment of *ATG7* rs143545741 p.P426L in severe NAFLD

To identify novel genetic determinants of severe NAFLD, we first developed a prioritization pipeline, based on frequency, predicted impact on protein function, and genic intolerance prediction, which is described in the Supplementary Methods and in Fig. S1. This approach increased the study power to detect rare variants with a large effect size by reducing the burden of multiple hypothesis testing (Fig. S2B). We detected 27 variants enriched in the Discovery cohort (whose features are reported in Table S1) compared to the general population (gnomAD NFE, Supplementary dataset A; adjusted $p<0.05$). Among those variants only two were also present in the Validation cohort (Table 1), and only *ATG7* rs143545741 C>T, encoding for the p.P426L aminoacidic substitution, was nominally associated with an increased risk of severe NAFLD also in this group ($p=0.034$, OR=4.33, Table 1). Considering both the discovery and validation sets the p.P426L variant was associated with a moderate-large increase in the risk of severe NAFLD (OR 5.26, 95% c.i. 2.06-12.61; $p=0.0031$). However, in family pedigrees two young women (daughters of probands with severe NAFLD) with no risk factors for liver disease carried the p.P426L variant and did not have evidence of liver damage (Supplementary Results).

Increased burden of *ATG7* C-terminal domain variants in severe NAFLD

To evaluate the overall impact of *ATG7* variants on the risk of severe NAFLD, we next examined the enrichment of missense functional variants, as compared to the general population (Non-Finnish European individuals included in the 1000 genomes cohort for whom individual data are available, 1000G-NFE, and local controls, which were processed by the same bioinformatic

pipeline; n=454). The gnomAD database could not be used for this specific analysis due to the lack of individual-level genotype data. The burden test for enrichment (SKAT test) of variants in severe NAFLD vs. controls is presented in Table 2, while the list of variants in Table S5 and Fig. 1. We detected a consistent enrichment in *ATG7* variants for all case cohorts vs. controls (Fig. 1 and Table 2; $p<0.05$). Mutations clustering at the "Apg7 homology domain" within the *ATG7* C-terminal catalytic domain were consistently enriched in all clinical cohorts (Table 2 and Fig. 2, $p<0.05$ at both SKAT and CAST tests). Overall, carriage of at least one rare variant in the *ATG7* C-terminal domain was associated with a 13.9 increase in the risk of advanced NAFLD (95% c.i. 1.9-611.5, $p=0.002$ at CAST test). The association remained significant after removing the rs143545741 p.P426L variant (OR=∞; 95% c.i.=1.0-∞; $p=0.025$ at CAST test; $p=0.012$ at SKAT test), indicating an independent contribution of the other missense variants located in this protein domain.

Enrichment in functional low-frequency and in rare variants of *ATG7* in patients with severe NAFLD (n=301) vs. considering the possible bias due to the use of external controls sequenced by a different platform (Table S6; $p<0.05$ irrespective of the threshold to filter rare variants). This integrated approach confirmed the absence of bias due to different sequencing platform used in the experiments and the resulting association from previous burden tests.

To highlight other clinically relevant *ATG7* variants, enrichment of specific variants occurring more than once in the clinical cohort of severe NAFLD patients (n=301) was assessed against the European population (gnomAD-NFE, n=56,885). Besides the p.P426L, we found an enrichment in the low-frequency variant rs36117895 T>C (MAF=0.060 vs. 0.035) encoding for p.V471A (Table S2; OR=1.7; 95% c.i. 1.2-2.5; $p=0.003$).

ATG7 p.P426L and p.V471A variants behave as loss-of-function in hepatocytes *in vitro*

We next investigated the functional impact of *ATG7* variants identified in the clinical cohort on lipid levels in human hepatoma cells. Downregulation of *ATG7* by siRNA technology (Fig. S3A) resulted in increased fat content in both HepG2 and HepaRG cells ($p<0.05$ for both, Fig. S3B and S3C). Importantly, *ATG7* silencing increased intracellular lipid accumulation, in primary human hepatocytes culture in 2D culture and in 3D, with a larger effect size compared to immortalized cells (Fig. 2A and 2B, respectively).

To corroborate previous results, ATG7 wild-type, p.P426L and p.V471A were transiently overexpressed in HepaRG cultured with oleic acid. Overexpression of the p.P426L ATG7 variant did not affect protein stability, whereas p.V471A showed non-significantly trend for lower protein levels compared to the wild-type (Fig. S3D). The impact of additional variants identified in the clinical cohort on ATG7 protein stability is reported in the supplementary results. Over-expression of wild-type ATG7 reduced the intracellular neutral lipid levels ($p<0.05$ Fig. 2C and 2D), whereas either the p.P426L or the p.V471A mutants reduced the ability of ATG7 to decrease the intracellular neutral lipid content ($p<0.05$; Fig. 2C and 2D).

To address the functional consequence of p.V471A variant on protein activity, we engineered HepG2 hepatocytes (which resulted heterozygous for the p.V471A, thereafter called *ATG7* V471A$^{+/-}$) to generate two syngenic independent cell clones, carrying the p.V471A in homozygosity (thereafter termed *ATG7* V471A$^{+/+}$) or *ATG7* gene deletion (thereafter termed *ATG7*$^{-/-}$), by using CRISPR-Cas9 (Fig S4). Despite the unaltered transcript abundance (Fig. S5A), *ATG7* V471A$^{+/+}$ hepatocytes displayed reduced ATG7 protein level, similar to the *ATG7*$^{-/-}$ clone (Fig. S5B). We next investigated the autophagic flux by inducing autophagy and blocking the lysosomal degradation by chloroquine treatment. ATG7 catalyzes the conversion of microtubule-

associated protein 1A/1B-light chain 3 by phosphatidylethanolamine conjugation (LC3B-I to LC3B-II) resulting in reduced levels of the autophagosome cargo protein p62. We observed a reduction of LC3B-II conversion in V471A$^{+/+}$ vs. V471A$^{+/-}$ cells ($p<0.05$; Fig. 2E and 2F), whereas expectedly in *ATG7*$^{-/-}$ cells pathway activation was almost completely suppressed (Fig. 2E and 2F). In keeping, p62 levels were higher in *ATG7* V471A$^{+/+}$ and *ATG7*$^{-/-}$ cells as compared to reference V471A$^{+/-}$ cells (Fig. 2E and 2F). Finally, *ATG7* V471A$^{+/+}$ cells showed also a higher susceptibility to accumulate intracellular lipid droplets compared to *ATG7* V471A$^{+/-}$ hepatocytes after exposure to fatty acids, similarly to *ATG7*$^{-/-}$ cells ($p<0.001$; Fig. S5C-D).

Taken together, these data suggest that the p.V471A is a hypomorphic variant, namely it results in an unstable protein, leading to an impairment of the autophagic flux that in turn facilitates lipid accumulation in hepatocytes.

Loss-of-function *ATG7* variants predispose to liver disease in the population

Missense variants in the active domain of *ATG7* are very likely to be causing a LoF of protein activity. To confirm the association of rare LoF variants with severe liver disease at a population level, we examined rare (MAF<0.005) LoF variants in unrelated Europeans from the UK Biobank population-based study (UKBB). LoF were defined as variation causing a stop gained, frameshift, splice acceptor, or splice donor variants (Table S5). The impact of LoF variants on liver disease in UKBB is reported in Table 3. We observed a progressive enrichment of these variants with severity of liver disease going from a 55% enrichment in chronic liver disease (SKAT-O p=0.025) to a more than 12-fold enrichment in HCC (SKAT-O p=0.001). To test if excess in body weight may interact with these rare variants in determining liver disease, we stratified the cohort based on body mass index (BMI). In obese individuals (BMI≥30) there was a

similar progressive enrichment of rare LoF variants throughout the spectrum of liver disease that was approximately twice larger for any given class of liver disease, with a more than 23-fold enrichment for HCC. In addition, carriage of *ATG7* variants was associated with increased risk of HCC ($p<0.05$) and of liver related phenotypes both in individuals irrespective of the reported alcohol intake (Table S7).

ATG7 p.V471A is associated with liver injury and disease trough hepatocellular ballooning

In the Liver-Bible-2021 cohort, carriage of either p.P426L or p.V471A variants was associated with higher AST, ALT, and risk of fatty liver and non-invasively assessed severe NAFLD (Table S8; $p<0.05$ for all). However, it was not associated with metabolic features.

In the overall UKBB the rs36117895 p.V471A variant was associated with AST levels ($p=7.7*10^{-4}$; Table S4). After stratification for BMI there was an association between HCC and the p.V471A variant in those with BMI$\geq$35 (OR: 2.31, 95% c.i. 1.2-4.4 p=0.009, Fig. S6), corresponding to the largest impact on AST levels (Fig. S6).

To understand the mechanism underlying the association with liver disease, we next examined the impact of the p.V471A variant on histological liver damage in 2268 patients from the NAFLD Liver biopsy cohort (LBC). The rs143545741 (p.P426L) variant was rare (MAF=0.003, n=14 carriers) among Italians and absent in Finns (in line with public database data, not shown). Clinical features of LBC patients stratified by rs36117895 p.V471A are shown in Table S2. We first assessed whether the impact of p.V471A on liver damage was accounted for by the association with a specific feature of liver damage. The p.V471A variant was not associated with steatosis or inflammation (Fig. 3A and 3B), but it was an independent predictor of ballooning (p=0.007, Fig. 3C, Table S9 left panel, Bonferroni adjusted p=0.028) and explained a relevant

proportion of the susceptibility to severe ballooning (Population Attributable Fraction, PAF=0.09). In UKBB, *ATG7* variants accounted for a meaningful fraction of HCC variability (PAF=0.02). The effect of p.V471A on ballooning was also independent of hepatic fat accumulation (adjusted p=0.009, Fig.3, Table S9 right panel). The p.V471A variant was nominally associated with the presence of advanced fibrosis stage F3-F4 (p=0.02, Table S2; adjusted p=0.1, Fig. 2E). However, the impact of the p.V471A on severe fibrosis was consistent in those with the most severe steatosis grade S3 (n=424, p=0.002; Fig. 3F). Sensitivity analyses in *PNPLA3* p.I148M carriers and patients with a high genetic risk to develop steatosis according to the polygenic risk score - hepatic fat content (PRS-HFC) [3, 4] are presented in Fig. S7 and Table S10. Similarly, in UKBB individuals, the association between the p.V471A *ATG7* variant and liver phenotypes was independent of PRS-HFC. However, it was significant only in those with high (≥0.53; adjusted OR 1.85, 95% c.i. 1.0-3.3; p=0.022).

Determinants of hepatic *ATG7* mRNA and protein expression

To further understand the role of ATG7 in NAFLD, we examined the liver protein levels in a cohort of 22 patients with NAFLD by immunohistochemistry (age 55±years, 64% males, 71% with severe fibrosis). We observed a strong ATG7 expression in non-parenchymal cells (Fig. 4A). However, ATG7 was also present in periportal hepatocytes (p<0.01 vs. centrilobular hepatocytes) and staining was particularly intense around lipid droplets (Fig. 4B-D). Staining intensity in hepatocytes was qualitatively higher in patients with hepatocellular ballooning (Fig. 4E, p=NS), as well as in those with severe (grade S3) steatosis (Fig. 4F, p=NS). On the other hand, ATG7 staining diffusion correlated with steatosis severity (estimate 0.51±0.21, adjusted p=0.046), with

a tendency to be more restricted to periportal hepatocytes in carriers of *PNPLA3* p.I148M (estimate 0.50±0.23; adjusted p=0.11).

To investigate the molecular mechanisms underpinning increased susceptibility to hepatocellular damage in carriers of *ATG7* p.V471A, we examined its impact on hepatic transcriptome (Transcriptomic cohort, Supplementary dataset B). First, in keeping with protein expression data isoform profiling of *ATG7* expression revealed that the main *ATG7* transcript was overexpressed in patients with hepatocellular ballooning independently of age, sex, BMI, and *PNPLA3* p.I148M (Fig. S8, β=0.55±0.19, p=0.004). The association between *ATG7* expression and ballooning was also independent of steatosis, lobular inflammation, and fibrosis (β=0.49±0.20, p=0.015).

To further explore *ATG7* role in liver biology, co-expression analysis was performed. *ATG7* expression level was positively correlated with that of genes involved in several metabolic pathways, mTORC1 signaling and coagulation. *ATG7* expression levels were inversely correlated with genes involved in cell cycle progression and TNF-α signaling (Fig. 5A, adjusted p<0.05). Conversely, gene set enrichment analysis (GSEA) of nominally associated transcripts revealed an overexpression in p.V471A carriers of the TNFα signaling, hypoxia, DNA damage, and p53 pathways (Supplementary Dataset C, Fig. 5B, FDR adjusted p<0.05). The impact of p.V471A on autophagy is shown in the Supplementary and Fig. S9A-B. The expression of genes belonging to this pathway and nominally associated with p.V471A variant carriage, including several genes involved in liver inflammation, is shown in Fig. S9C. Taken together, the dissociation of the impact of the p.V471A variant and *ATG7* gene expression with activation of the TNFα pathway and autophagy genes together suggest that the p.V471A variant behaves as a loss-of-function inducing TNFα dependent inflammation in carriers. By *in situ* hybridization (ISH) analysis of liver tissue

samples, we detected and association between the presence of the p.V471A variant with the number of TNFα mRNA positive (non-parenchymal) cells (5-fold increase as compared to lack of the *ATG7* variants; p=0.05; Figure S10).

DISCUSSION

Aiming to circumvent the main limitation of current WES studies for complex traits, namely the difficulty in achieving an adequate statistical power [21], as well as identifying new genetic determinants of severe NAFLD, we examined a cohort of individuals with severe NAFLD and developed an integrative data-mining approach to prioritize genetic variants contributing to this trait. To fulfil the selection criteria, genetic variants had to be: a) rare, b) causing protein sequence changes, as these are more likely to be deleterious [5], c) predicted to affect protein function with a high confidence, and d) having a low rate of spontaneous variation in the general population. To increase the power of our study we specifically examined individuals at the extreme of a well-defined clinical spectrum, namely advanced fibrosis and HCC.

We first identified an association of the rare *ATG7* rs143545741 p.P426L variant with severe NAFLD. This rare variant conferred a large, greater than seven-fold, increase in the risk of severe liver disease in European individuals. To validate this first association, we considered the mutational burden at the *ATG7* locus highlighting the enrichment of rare protein-coding variants of *ATG7* in patients with severe NAFLD compared to the general population. In the clinical cohort missense variants in the highly conserved catalytic domain were associated with a ~14-fold increased risk of severe liver disease, in the range of monogenic disorders. The phylogenetic conservation of the locus and high intolerance mutations of *ATG7* in the population, and the identification of one individual carrying a nonsense variant (R659*) are consistent with the notion that these missense variants resulted in a LoF.

Next, we confirmed a higher risk of liver disease in carriers of LoF *ATG7* variants at a population level in unrelated Europeans from the UKBB. Consistent with a gene-environment interaction, the association effect size of *ATG7* variants was amplified by the presence of obesity,

resulting in a ~23-fold higher risk of HCC in obese individuals. Taken altogether these results suggest that excess in body weight acts as a permissive factor exposing the deleterious effect of rare *ATG7* LoF variants on liver disease.

A fine-grained analysis of the role of specific *ATG7* variants highlighted a ~70% higher risk of severe NAFLD in carriers of the low-frequency rs36117895 p.V471A variant, which is comparable to that of the main NAFLD risk variants in *PNPLA3* and *TM6SF2* [2]. Of note, this variant was previously associated with a younger age of onset in Huntington's disease by impairing autophagy [22]. The association of p.V471A with increased risk of liver injury (AST levels) was detectable in the UKBB, where the variant was also associated with increased risk of HCC in severely obese individuals. In individuals with dysmetabolism, carriage of *ATG7* variants was associated with non-invasively assessed liver damage, including aminotransferases, hepatic fat accumulation and liver damage, but not with metabolic traits including insulin resistance.

To characterize the mechanism underpinning the epidemiological association with liver damage, we next examined the histological correlates of p.V471A carriage in the liver biopsy cohort. Although the p.V471A variant was not associated with steatosis, it was one of the main independent determinants of hepatocellular ballooning, particularly in those patients with more marked steatosis and in carriers of the *PNPLA3* p.I148M variant. These results are consistent with previous experimental data, supporting the combination of loss-of-function of *ATG7* with *PNPLA3* p.I148M - a main modifier of hepatic lipid droplets remodeling - having a synergic impact on liver disease [23], resulting in reduced lipo-autophagy in hepatocytes [24]. Data are also consistent with evidence from the family segregation analysis conducted in this study, showing that carriage of p.P426L did not confer a strong predisposition to NAFLD in the absence of metabolic triggers.

Taken together, these data suggest that genetic variants in *ATG7* lead to impaired autophagy in hepatocytes and predispose to fatty liver disease progression in individuals with metabolic risk factors. On the other hand, initiation of autophagy and intact ATG7 are required for hepatic stellate cells trans-differentiation and initiation of fibrogenesis [25], which may limit the detrimental impact of *ATG7* LoF mutations in individuals without impairment in lipid droplets remodeling.

The overexpression of the main *ATG7* transcript and protein in individuals with more severe lipid accumulation and progressing to ballooning degeneration also speaks in favor of a direct involvement of autophagy in the cell toxicity caused by lipids. Indeed, ATG7 staining intensity tended to correlate with steatosis and to localize around large lipid droplets, in keeping with a role in macro-lipo-autophagy [26]. This is also consistent with reduced ATG7 expression in perivenular hepatocytes, more susceptible to fat accumulation, especially in *PNPLA3* p.I148M carriers.

In line with this hypothesis and with a possible role of *ATG7* in lipid droplets catabolism [26], *ATG7* knockdown resulted in lipid overload in human hepatoma cells, while *ATG7* overexpression rescued the intracellular lipid accumulation induced by oleic acid exposure. Conversely, overexpression of p.P426L and p.V471A mutants did not reverse the phenotype. Furthermore, we directly confirmed that the p.V471A variant results in a LoF of ATG7 activity in hepatocytes, leading to a reduction in autophagy, the accumulation of p62 (SQSTM1), and of intracellular lipids.

The highly conserved C-terminal domain is involved in the binding to the Ubiquitin-like proteins (UBL) and contains the catalytic cysteine (Cys507) of ATG7 [27], an E1 enzyme involved in the regulation of autophagy. A growing body of evidence shows that autophagy regulates lipid

metabolism in the liver, and that *ATG7* plays a key role in initiating the signaling cascade eventually leading to lipo-autophagy [26]. Autophagy also regulates the activity of other hepatic and extra-hepatic cell types with a potential impact on liver disease progression. For example, adipose specific *Atg7* knock-down ameliorated NAFLD progression in mice [28]. However, the human genetic data suggest that, overall, the effect of *ATG7* LoF is detrimental for the liver. In keeping with this, adult *Atg7* mouse KO models displayed hepatic lipid accumulation due to reduced lipo-autophagy during fasting resulting in liver damage. In hepatocytes, steatosis was accompanied by accumulation of p62/LC-III aggregates [29], consistent with failure of autophagy [30], that are typically associated with hepatocellular ballooning [26, 31]. In both experimental models and in individuals with NAFLD, impairment in autophagy leads to hepatocellular ballooning characterized by accumulation of p62 ubiquitylated aggregates [32], a process also involved in hepatic carcinogenesis [33, 34], which was promoted by the p.V471A variant and by *ATG7* deletion in human primary hepatocytes *in vitro*. Furthermore, hepatocyte specific *Atg7* deletion initiates TNFα dependent liver injury [35].

Transcriptomic analysis provided an independent confirmation of the mechanism linking impaired ATG7 activity with liver disease. Indeed, the expression levels of the autophagy genes upstream of *ATG7* were consistent with a strong induction of autophagy initiation in p.V471A variant carriers. Conversely, most of the genes downstream of *ATG7* were downregulated suggesting an interruption of the pathway at this level. In keeping with this result, while ATG7 transcript levels correlated with fat oxidation [36] and suppression of TNFα signaling, while this latter pathway was upregulated in p.V471A variant carriers, concomitantly with an increased frequency of non-parenchymal TNFα positive cells. Overall, the data indicate that the p.V471A variant results in a LoF, which prevents ATG7 from suppressing inflammation in response to an

impaired lipo-autophagy and to lipotoxicity. The disease mechanism may encompass lysosomal permeabilization, caspase-8 dependent activation of the mitochondrial death pathway and TNFα release [35, 37].

The present study does not provide a comprehensive evaluation of the role of rare variants in NAFLD pathogenesis. Rather, due to current limitations in the size of cohort of patients with advanced disease, the study was designed to screen, by exploiting a heavy prioritization pipeline and an initial validation, single likely pathogenic variants associated with severe NAFLD, aiming to find robust candidates for subsequent evaluation. This however was followed by extensive validation at the gene level and in multiple clinical and population-based cohorts, with functional studies. We found several other candidates in the first analyses, which will require further validation. A limitation of this study is that we could not demonstrate if the impact of *ATG7* LoF variants on liver disease progression is entirely mediated by predisposition to ballooning, or whether the variants have a direct carcinogenic effect. While previous studies in one experimental model have reported that full *Atg7* knockdown by deletion during adult life may inhibit hepatic carcinogenesis [38], our data are consistent with the notion that heterozygous carriage of LoF variants favors liver disease development and is potentially carcinogenic. Finally, results may not be applicable to different ethnicities.

In conclusion, by several independent approaches we identified rare and low-frequency LoF genetic *ATG7* variants as modifiers of NAFLD progression in Europeans. The proposed mechanism involves the facilitation of hepatocellular ballooning degeneration in individuals with fatty liver and dysmetabolism. These findings highlight that examining well-characterized and extreme phenotypes together with prioritization of rare genetic variants can be used to identify novel candidate genetic determinants of human disease.

Abbreviations: NAFLD: nonalcoholic fatty liver disease; ATG7: autophagy related 7; TNF: tumor necrosis factor; HCC: hepatocellular carcinoma; WES: whole exome sequencing; GWAS: genomewide association study; ApoB: Apolipoprotein B; UKBB: United Kingdom Biobank cohort; BMI: body mass index; ORO: oil red O staining; NFE: non-Finnish Europeans; VEP: variant effect predictor; CAST: Cohort Allelic Sum Test; SKAT: sequence Kernel association test; GSEA: gene set enrichment analysis; MAF: minor allele frequency; LoF: loss-of-function.

Acknowledgments: We thank Rossana Carpani for administrative support and Alessandro Pietrelli for his expert opinion and support at the early stages of the project.

REFERENCES

[1] Younossi Z, Henry L. Contribution of Alcoholic and Nonalcoholic Fatty Liver Disease to the Burden of Liver-Related Morbidity and Mortality. Gastroenterology 2016;150:1778-1785.

[2] Trepo E, Valenti L. Update on NAFLD genetics: From new variants to the clinic. J Hepatol 2020;72:1196-1209.

[3] Bianco C, Jamialahmadi O, Pelusi S, Baselli G, Dongiovanni P, Zanoni I, et al. Non-invasive stratification of hepatocellular carcinoma risk in non-alcoholic fatty liver using polygenic risk scores. J Hepatol 2021;74:775-782.

[4] Dongiovanni P, Stender S, Pietrelli A, Mancina RM, Cespiati A, Petta S, et al. Causal relationship of hepatic fat with liver damage and insulin resistance in nonalcoholic fatty liver. J Intern Med 2018;283:356-370.

[5] Kryukov GV, Pennacchio LA, Sunyaev SR. Most rare missense alleles are deleterious in humans: implications for complex disease and association studies. Am J Hum Genet 2007;80:727-739.

[6] Timpson NJ, Greenwood CMT, Soranzo N, Lawson DJ, Richards JB. Genetic architecture: the shape of the genetic contribution to human traits and disease. Nat Rev Genet 2018;19:110-124.

[7] Pelusi S, Baselli G, Pietrelli A, Dongiovanni P, Donati B, McCain MV, et al. Rare Pathogenic Variants Predispose to Hepatocellular Carcinoma in Nonalcoholic Fatty Liver Disease. Sci Rep 2019;9:3682.

[8] Czaja MJ. Function of Autophagy in Nonalcoholic Fatty Liver Disease. Dig Dis Sci 2016;61:1304-1313.

[9] Yang L, Li P, Fu S, Calay ES, Hotamisligil GS. Defective hepatic autophagy in obesity promotes ER stress and causes insulin resistance. Cell Metab 2010;11:467-478.

[10] Xiao Y, Liu H, Yu J, Zhao Z, Xiao F, Xia T, et al. Activation of ERK1/2 Ameliorates Liver Steatosis in Leptin Receptor-Deficient (db/db) Mice via Stimulating ATG7-Dependent Autophagy. Diabetes 2016;65:393-405.

[11] Liver EAFTSOT, Cancer EOFRATO. EASL-EORTC clinical practice guidelines: management of hepatocellular carcinoma. J Hepatol 2012;56:908-943.

[12] Donati B, Pietrelli A, Pingitore P, Dongiovanni P, Caddeo A, Walker L, et al. Telomerase reverse transcriptase germline mutations and hepatocellular carcinoma in patients with nonalcoholic fatty liver disease. Cancer Med 2017;6:1930-1940.

[13] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. bioRxiv 2020:531210.

[14] Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, et al. A global reference for human genetic variation. Nature 2015;526:68-74.

[15] Pelusi S, Cespiati A, Rametta R, Pennisi G, Mannisto V, Rosso C, et al. Prevalence and Risk Factors of Significant Fibrosis in Patients With Nonalcoholic Fatty Liver Without Steatohepatitis. Clin Gastroenterol Hepatol 2019;17:2310-2319.e2316.

[16] Valenti L, Pelusi S, Bianco C, Ceriotti F, Berzuini A, Iogna Prat L, et al. Definition of Healthy Ranges for Alanine Aminotransferase Levels: A 2021 Update. Hepatol Commun 2021;5:1824-1832.

[17] European Association for the Study of the Liver. Electronic address eee, Clinical Practice Guideline P, Chair, representative EGB, Panel m. EASL Clinical Practice Guidelines on non-invasive tests for evaluation of liver disease severity and prognosis - 2021 update. J Hepatol 2021;75:659-689.

[18] Newsome PN, Sasso M, Deeks JJ, Paredes A, Boursier J, Chan WK, et al. FibroScan-AST (FAST) score for the non-invasive identification of patients with non-alcoholic steatohepatitis with significant activity and fibrosis: a prospective derivation and global validation study. Lancet Gastroenterol Hepatol 2020;5:362-373.

[19] Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. PLoS Med 2015;12:e1001779.

[20] Baselli GA, Dongiovanni P, Rametta R, Meroni M, Pelusi S, Maggioni M, et al. Liver transcriptomics highlights interleukin-32 as novel NAFLD-related cytokine and candidate biomarker. Gut 2020.

[21] Panoutsopoulou K, Tachmazidou I, Zeggini E. In search of low-frequency and rare variants affecting complex traits. Hum Mol Genet 2013;22:R16-21.

[22] Metzger S, Walter C, Riess O, Roos RA, Nielsen JE, Craufurd D, et al. The V471A polymorphism in autophagy-related gene ATG7 modifies age at onset specifically in Italian Huntington disease patients. PLoS One 2013;8:e68951.

[23] BasuRay S, Wang Y, Smagris E, Cohen JC, Hobbs HH. Accumulation of PNPLA3 on lipid droplets is the basis of associated hepatic steatosis. Proc Natl Acad Sci U S A 2019;116:9521-9526.

[24] Negoita F, Blomdahl J, Wasserstrom S, Winberg ME, Osmark P, Larsson S, et al. PNPLA3 variant M148 causes resistance to starvation-mediated lipid droplet autophagy in human hepatocytes. J Cell Biochem 2019;120:343-356.
[25] Hernandez-Gea V, Ghiassi-Nejad Z, Rozenfeld R, Gordon R, Fiel MI, Yue Z, et al. Autophagy releases lipid that promotes fibrogenesis by activated hepatic stellate cells in mice and in human tissues. Gastroenterology 2012;142:938-946.
[26] Singh R, Kaushik S, Wang Y, Xiang Y, Novak I, Komatsu M, et al. Autophagy regulates lipid metabolism. Nature 2009;458:1131-1135.
[27] Kaiser SE, Mao K, Taherbhoy AM, Yu S, Olszewski JL, Duda DM, et al. Noncanonical E2 recruitment by the autophagy E1 revealed by Atg7-Atg3 and Atg7-Atg10 structures. Nat Struct Mol Biol 2012;19:1242-1249.
[28] Sakane S, Hikita H, Shirai K, Myojin Y, Sasaki Y, Kudo S, et al. White Adipose Tissue Autophagy and Adipose-Liver Crosstalk Exacerbate Nonalcoholic Fatty Liver Disease in Mice. Cell Mol Gastroenterol Hepatol 2021;12:1683-1699.
[29] Karsli-Uzunbas G, Guo JY, Price S, Teng X, Laddha SV, Khor S, et al. Autophagy is required for glucose homeostasis and lung tumor maintenance. Cancer Discov 2014;4:914-927.
[30] Collier JJ, Guissart C, Olahova M, Sasorith S, Piron-Prunier F, Suomi F, et al. Developmental Consequences of Defective ATG7-Mediated Autophagy in Humans. N Engl J Med 2021;384:2406-2417.
[31] Komatsu M, Waguri S, Ueno T, Iwata J, Murata S, Tanida I, et al. Impairment of starvation-induced and constitutive autophagy in Atg7-deficient mice. J Cell Biol 2005;169:425-434.
[32] Carotti S, Aquilano K, Zalfa F, Ruggiero S, Valentini F, Zingariello M, et al. Lipophagy Impairment Is Associated With Disease Progression in NAFLD. Front Physiol 2020;11:850.
[33] Takamura A, Komatsu M, Hara T, Sakamoto A, Kishi C, Waguri S, et al. Autophagy-deficient mice develop multiple liver tumors. Genes Dev 2011;25:795-800.
[34] Komatsu M, Waguri S, Koike M, Sou YS, Ueno T, Hara T, et al. Homeostatic levels of p62 control cytoplasmic inclusion body formation in autophagy-deficient mice. Cell 2007;131:1149-1163.
[35] Amir M, Zhao E, Fontana L, Rosenberg H, Tanaka K, Gao G, et al. Inhibition of hepatocyte autophagy increases tumor necrosis factor-dependent liver injury by promoting caspase-8 activation. Cell Death Differ 2013;20:878-887.
[36] Saito T, Kuma A, Sugiura Y, Ichimura Y, Obata M, Kitamura H, et al. Autophagy regulates lipid metabolism through selective turnover of NCoR1. Nature communications 2019;10:1567.
[37] Shen Y, Malik SA, Amir M, Kumar P, Cingolani F, Wen J, et al. Decreased Hepatocyte Autophagy Leads to Synergistic IL-1beta and TNF Mouse Liver Injury and Inflammation. Hepatology 2020;72:595-608.
[38] Cho KJ, Shin SY, Moon H, Kim BK, Ro SW. Knockdown of Atg7 suppresses Tumorigenesis in a murine model of liver cancer. Transl Oncol 2021;14:101158.

TABLES

Table 1. Variants enriched in individuals with severe NAFLD in the Discovery cohort observed the Validation cohort. Frequencies were compared to the general population (n=56,885 gnomAD non-Finnish Europeans (NFE) plus 50 local ethnically matched healthy individuals from the same geographical area) for the Discovery cohort and to European individuals without liver disease included in UKBB for the Validation cohort (n=123,456).

Variants		Variant	2:g.32740353C>T	3:g.11389502C>T
		dbSNP	rs61757638	rs143545741
		Gene	*BIRC6*	*ATG7*
		Aminoacidic change	A3622V	P426L
Discovery	**Cases (n=72)**	**Alleles**	3/144	2/144
		MAF	0.0208	0.0139
	Controls (n=56,935)	**Alleles**	347/113340	196/112822
		MAF	0.0031	0.0017
		OR (95% c.i.)	6.93 (1.86–20.39)	8.09 (1.41–29.46)
		p	0.0103	0.0267
Validation	**Cases (n=229)**	**Alleles**	2/458	3/458
		MAF	0.0044	0.0066
	Controls (n=123,456)	**Alleles**	672/246912	375/246912
		MAF	0.0027	0.0015
		OR (95% c.i.)	1.61 (0.28–5.97)	4.33 (1.17–13.07)
		p	0.3548	0.0339
Overall	**Cases (n=301)**	**Alleles**	5/602	5/602
		MAF	0.0083	0.0083
	Controls (n=180,391)	**Alleles**	1019/359580	571/359359
		MAF	0.0028	0.0016
		OR (95% c.i.)	2.95 (1.16–7.08)	5.26 (2.06-12.70)
		p	0.0300	0.0031

AA: aminoacidic change; ATG7: autophagy related 7; BIRC6: baculoviral-inhibitor of apoptosis repeat containing 6; MAF: Minor Allele Frequency; OR: odds ratio; 95% c.i.: 95% confidence interval; FDR: false discovery rate. Only variants represented in both the Discovery and Validation cohort are reported. Of patients positive for the *ATG7* p.P426L variant, in the discovery cohort 2/2 had cirrhosis and HCC, while in the validation cohort 3/3 had advanced fibrosis without HCC.

Table 2. **Enrichment in *ATG7* variants in patients with severe NAFLD vs. the general population** (n=404 from 1000 genomes project NFE plus n=50 local controls).

***ATG7* gene**										
	All variants				**Rare variants (MAF<0.005)**					
Cohort	**n variants**	**Cases (%)**	**Controls (%)**	**SKAT p**	**n variants**	**Cases (%)**	**Controls (%)**	**SKAT p**	**CAST OR (95% c.i.)**	**CAST p**
EPIDEMIC discovery (N=72)	7	0.24	0.09	0.001	6	0.06	0.01	0.006	6.6 (1.2 - 36.2)	0.015
EPIDEMIC validation (N=129)	7	0.11	0.09	0.040	6	0.04	0.01	0.009	4.5 (1.0 - 23.1)	0.029
EPIDEMIC overall (N=201)	9	0.15	0.09	0.007	8	0.05	0.01	0.002	5.3 (1.4 - 23.7)	0.004
PERSPECTIVE (N=100)	8	0.14	0.09	0.142	7	0.03	0.01	0.069	3.5 (0.5 - 20.9)	0.115
ALL (N=301)	12	0.15	0.09	0.029	11	0.04	0.01	0.005	4.7 (1.4 - 20)	0.008
ATG7 C-terminal domain ("Apg7 homology domain" - CD01486)										
	All variants				**Rare variants (MAF<0.005)**					
Cohort	**n variants**	**Cases (%)**	**Controls (%)**	**SKAT p**	**n variants**	**Cases (%)**	**Controls (%)**	**SKAT p**	**CAST OR (95% c.i.)**	**CAST p**
Discovery (N=72)	3	0.22	0.09	0.002	2	0.04	0.002	0.006	19.5 (1.5 - 1029.2)	0.009
Validation (N=129)	2	0.10	0.09	0.039	1	0.02	0.002	0.019	10.7 (1.0 - 565.5)	0.036
EPIDEMIC (N=201)	3	0.14	0.09	0.006	2	0.03	0.002	0.003	13.9 (1.7 - 640.4)	0.004
PERSPECTIVE (N=100)	5	0.14	0.09	0.045	4	0.03	0.002	0.011	13.9 (1.1 - 733.7)	0.020
ALL (N=301)	6	0.14	0.09	0.027	5	0.03	0.002	0.004	13.9 (1.9 - 611.5)	0.002

SKAT: Sequence Kernel Association test; CAST: Cohort Allelic Sum test; OR: odds ratio; 95% c.i.: 95% confidence interval.

Table 3. Analysis of *ATG7* rare LoF variants in the UKBB (n=365,495). The association of 33 rare (MAF<0.005) loss-of-function variants of *ATG7* and liver related traits in UKBB was tested overall and within different BMI classes of non-obese (BMI<30) and obese (BMI≥30).

Trait	p				MAC case (%)	MAC control (%)
	SKAT	burden	SKAT-O	β/OR (95% c.i.)		
CLD, yes	0.014	0.199	0.025	1.55 (0.72 - 2.9)	9 (0.25)	243 (0.17)
Cirrhosis, yes	0.001	0.008	0.002	3.30 (1.1 - 7.5)	5 (0.55)	247 (0.17)
SLD, yes	$2.4*10^{-4}$	0.003	$3.4*10^{-4}$	6.47 (2.1 - 15)	4 (0.88)	201 (0.16)
HCC, yes	$7.7*10^{-4}$	0.006	0.001	12.30 (2.6 - 36)	2 (1.54)	250 (0.17)
ALT, IU/l	0.436	0.620	0.631	0.03 (-0.08 - 14)	-	-
AST, IU/l	0.250	0.285	0.395	0.07 (-0.05, - 0.19)		
BMI<30						
CLD, yes	0.772	0.423	0.614	0.71 (0.15 - 2)	2 (0.10)	194 (0.17)
Cirrhosis, yes	0.366	0.859	0.543	1.73 (0.2 - 6.2)	1 (0.19)	195 (0.17)
SLD, yes	0.112	0.442	0.186	3.41 (0.39 - 12)	1 (0.37)	159 (0.17)
HCC, yes	>0.99	0.748	0.901	4.71 (0.04 - 33)	0 (0)	196 (0.17)
ALT, IU/l	0.731	0.842	0.892	-0.01 (-0.14 - 0.12)	-	-
AST, IU/l	0.671	0.965	0.851	0.003 (-0.13 - 0.14)		
BMI≥30						
CLD, yes	$2.6*10^{-4}$	0.006	$5.1*10^{-4}$	3.6 (1.5 - 7.3)	7 (0.48)	49 (0.15)
Cirrhosis, yes	$7.1*10^{-5}$	$8.2*10^{-4}$	$9.8*10^{-5}$	8.96 (2.9 - 22)	4 (1.06)	52 (0.15)
SLD, yes	$5.2*10^{-5}$	$7.3*10^{-4}$	$8.7*10^{-5}$	14.7 (3.9 - 39)	3 (1.62)	42 (0.15)
HCC, yes	$9.5*10^{-5}$	0.002	0.001	23.4 (4.5 - 77)	2 (3.08)	54 (0.15)
ALT, IU/l	0.476	0.128	0.211	0.20 (-0.06 - 0.45)	-	-
AST, IU/l	0.246	0.030	0.051	0.29 (0.03 - 0.54)		

Default beta weight was used for weighted kernel in SKAT. Odds ratios (OR) were calculated by Firth's penalized likelihood analysis in a gene-based collapsing model adjusted for age, gender, BMI and first ten genomic principal components. Abbreviations: LoF: loss-of-function; MAC, minor allele counts; OR: odds ratio; CLD: chronic liver disease; SLD: severe liver disease; HCC: hepatocellular carcinoma; ALT: alanine aminotransferase; 95% c.i.: 95% confidence interval. No association was detected with hepatic fat content was detected in individuals where information

was available, diabetes, hypertension, glucose, HbA1c and lipid levels, renal disease, and other gastroenteric tract cancers, including oesophagus, stomach, intestine, biliary tract and pancreatic cancer ($P>0.05$ at SKAT-O test for all).

FIGURE LEGENDS

Fig. 1. Nonsense and missense variants in the ATG7 main isoform identified by their aminoacidic change in cases and control groups according to their protein sequence localization. Scheme of the variants retrieved in the advanced NAFLD (n=301) and in the general population (n=404 from 1000 genomes project NFE plus n=50 local controls), which were evaluated at burden test analysis. Missense variants are indicated in yellow, while non-sense (truncating *) variants in red. The height of the variant symbols in the lollypop graph reflects their frequency. The Apg7 homology domain is indicated in blue. CD: conserved domain number (NCBI database), MAF: minor allele frequency.

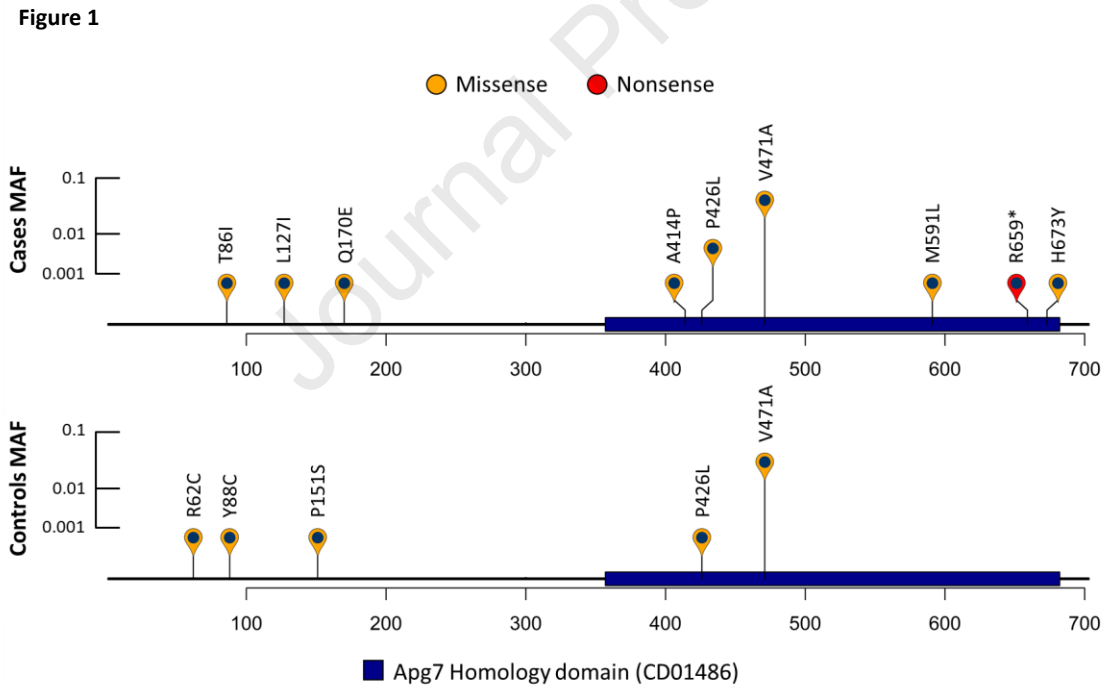


Fig. 2. ATG7 downregulation and overexpression influence intracellular lipid hepatocytes content and p.V471A and p.P426L result in a loss of function (LoF). A-B) Primary human hepatocytes were cultured in 2D (A) and as 3D (B) and incubated with negative control (SCR) siRNA or ATG7 siRNA for downregulation. Intracellular neutral lipid content was visualized by Oil Red-O staining (ORO). ORO area was quantified per DAPI stained nuclei by Image J. Two representative images of ORO-stained sections were presented. Data are shown as average and ± SD of 20 different sections. The p value was calculated by Mann-Whitney test. The ATG7 knockdown efficiency was evaluated by qRT-PCR. **C-D)** Overexpression of recombinant wild-type-V5 and mutant ATG7-V5 in HepaRG cells. Empty vector was used as negative control. Panel C shows representative images of the ORO staining. Panel D shows fluorescence intensity levels after ORO staining (n=4 for each group). Data are shown as mean and standard deviation; p values were calculated by using Mann-Whitney. RU, relative unit; RFU, relative Fluorescence Unit; W, wild type. **E-F)** Western blot and relative quantification of the conversion of LC3B-I to LC3B-II, a process catalyzed by ATG7, and accumulation of the p62 autophagy-related cargo in *ATG7* V471A$^{+/-}$, V471A$^{+/+}$ and ATG7$^{-/-}$ HepG2 cells treated with (+) or without (−) AZD8055 and chloroquine (CQ), to promote autophagy and block lysosomal degradation, respectively. *$p<0.05$, ** $p<0.005$.

Figure 2

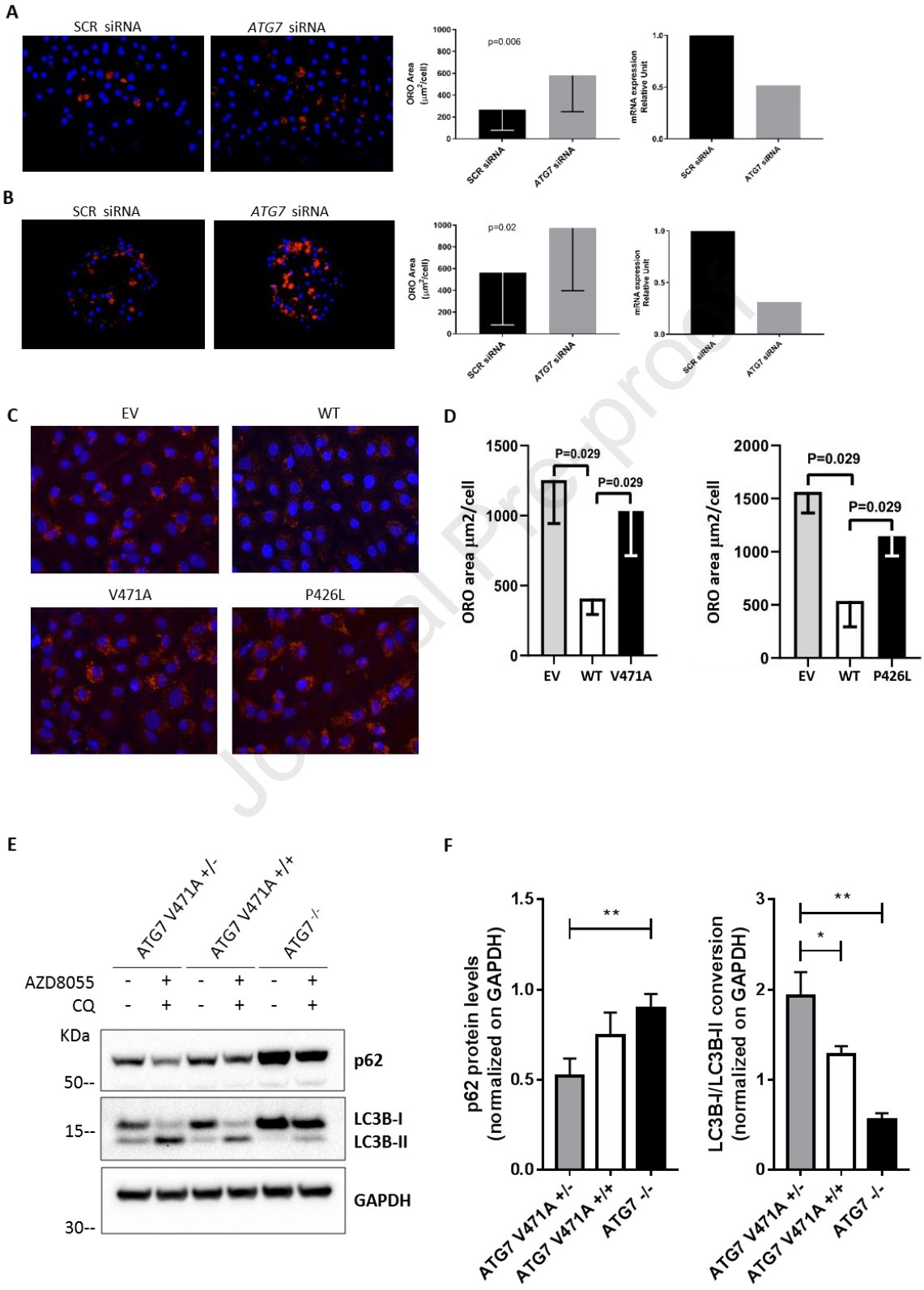


Fig. 3. Association of p.V471A *ATG7* variant with liver damage in the NAFLD Liver biopsy cohort (n=2268). Impact on steatosis (A), lobular inflammation (B), and hepatocellular ballooning (C) according to the rs36117895 p.V471A variant carriage; D) ballooning grade stratified by rs36117895 p.V471A variant carriage and by liver steatosis severity; impact on severe fibrosis (stage F3-F4) in the overall cohort (E) and in patients with S3 steatosis (F). All statistical analyses were performed under an additive model by multivariate ordinal logistic or binomial logistic regression accounting for sex, age, BMI, T2D, and *PNPLA3* rs738409 C>G p.I148M, *TM6SF2* rs58542926 C>T p.E167K, *MBOAT7* rs641738 C>T, *GCKR* rs1260326 T>C p.P446L variant genotypes. Statistics reported in D were further adjusted for steatosis grade. BMI: body mass index; T2D: type 2 diabetes; OR: Odds ratio; CI: confidence interval.

Figure 3

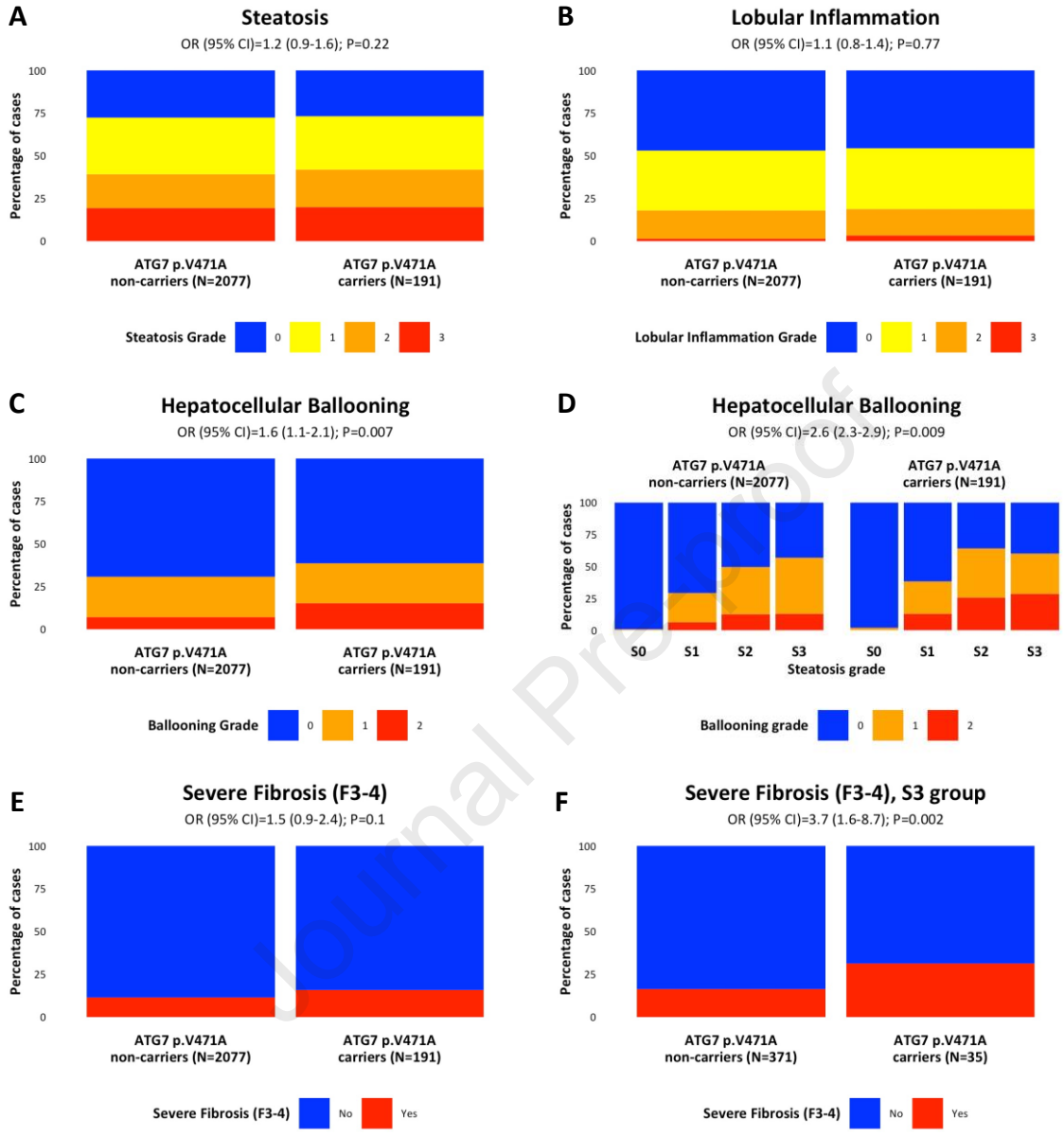


Fig. 4. ATG7 protein expression in liver biopsies. A-D) Representative images of ATG7 immunohistochemical staining in liver biopsies. Arrows highlight non-parenchymal cells (A) and lipid droplets (B-D); E-F) ATG7 staining intensity distribution according to presence of hepatocellular ballooning (E) or S3 steatosis (F).

Figure 4

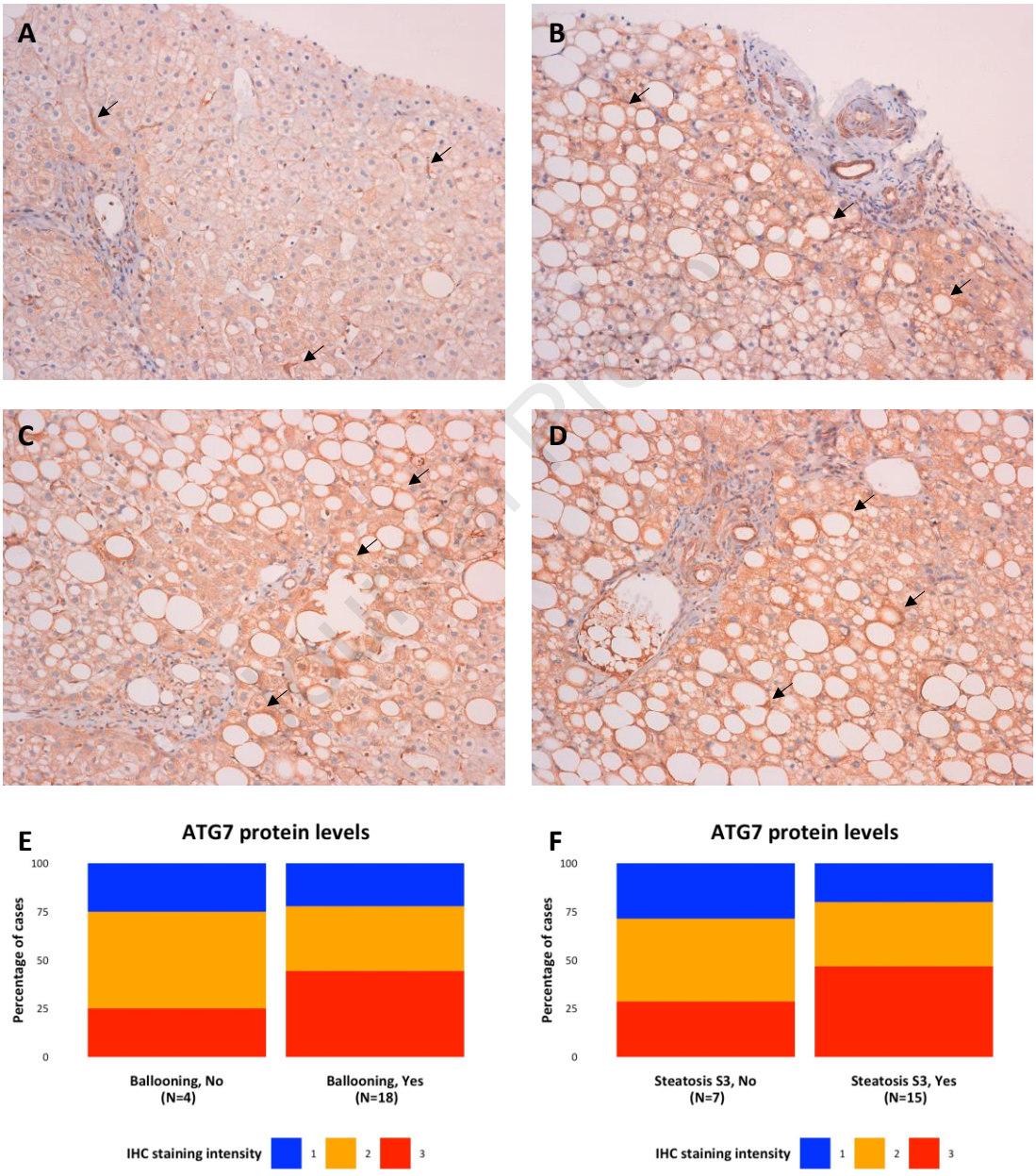


Fig. 5. Transcriptomic analysis. A) Gene set enrichment analysis (GSEA) was performed on pathways co-expressed with ATG7 in the Transcriptomic cohort B) GSEA was performed on genes nominally associated with ATG7 p.V471A genotype in the Transcriptomic cohort (n=125, at p<0.05 multivariate negative binomial regression for both).

Figure 5

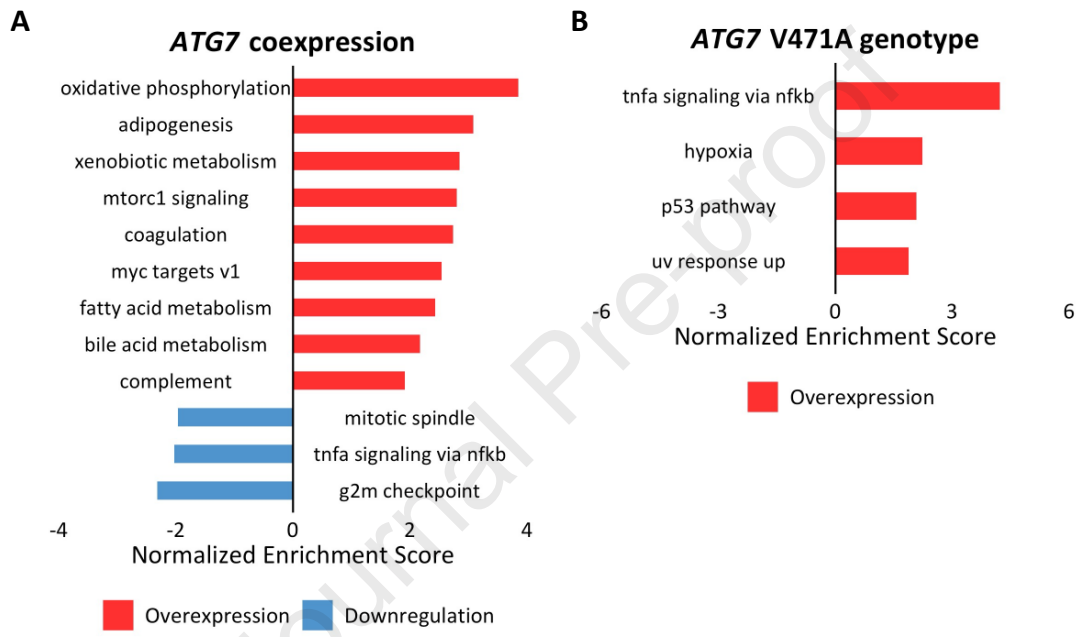


Figure 1

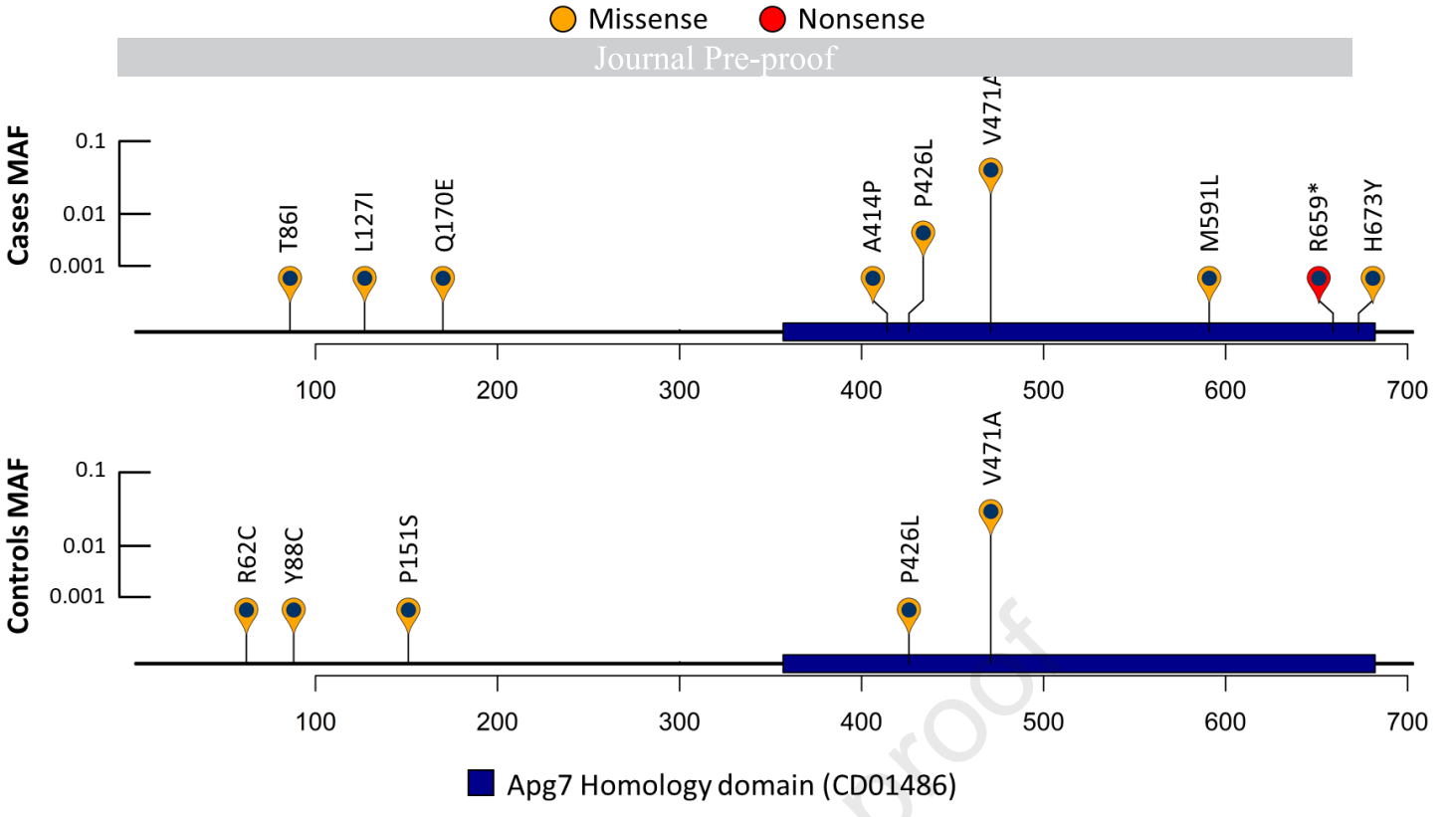


Figure 2

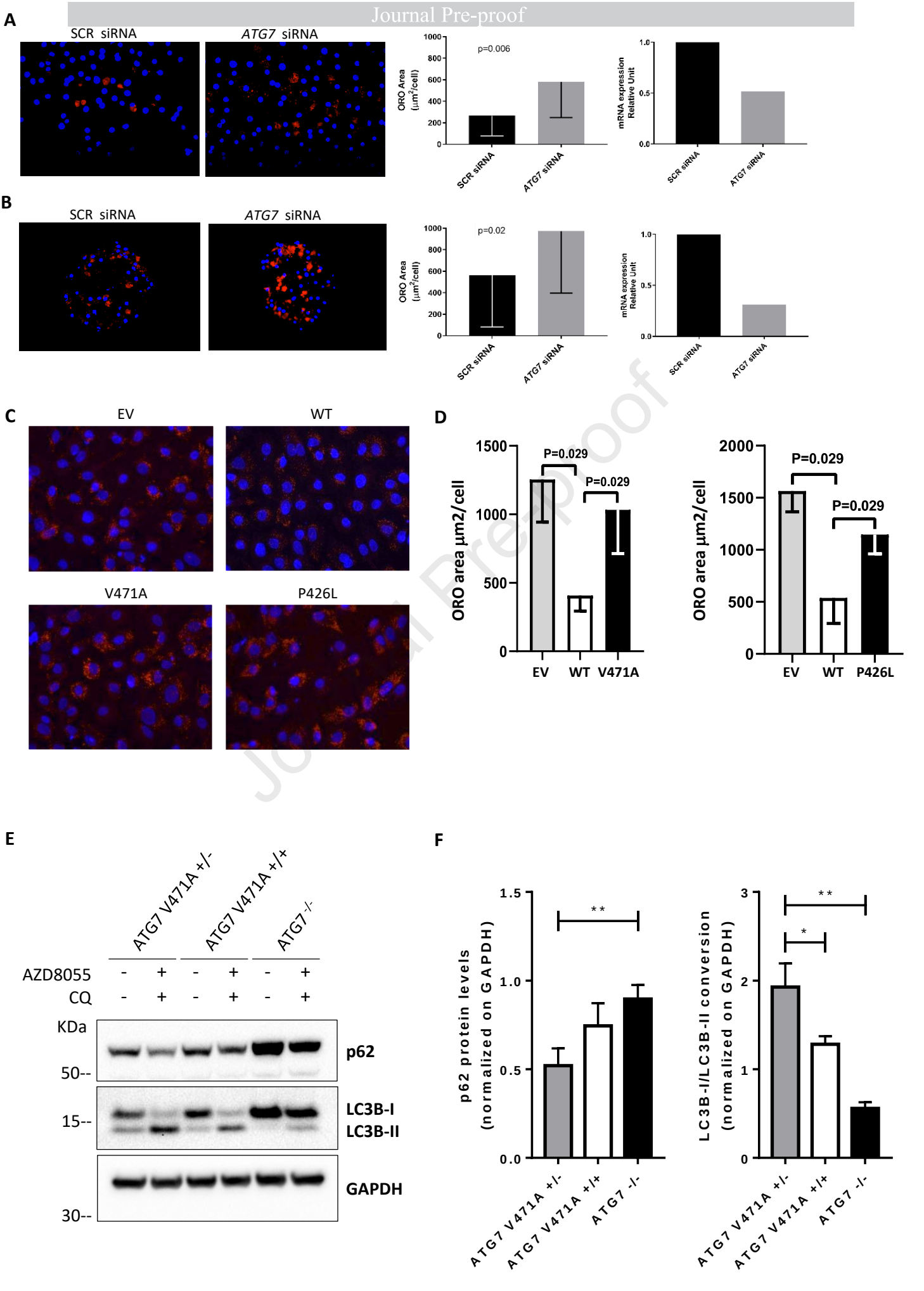


Figure 3

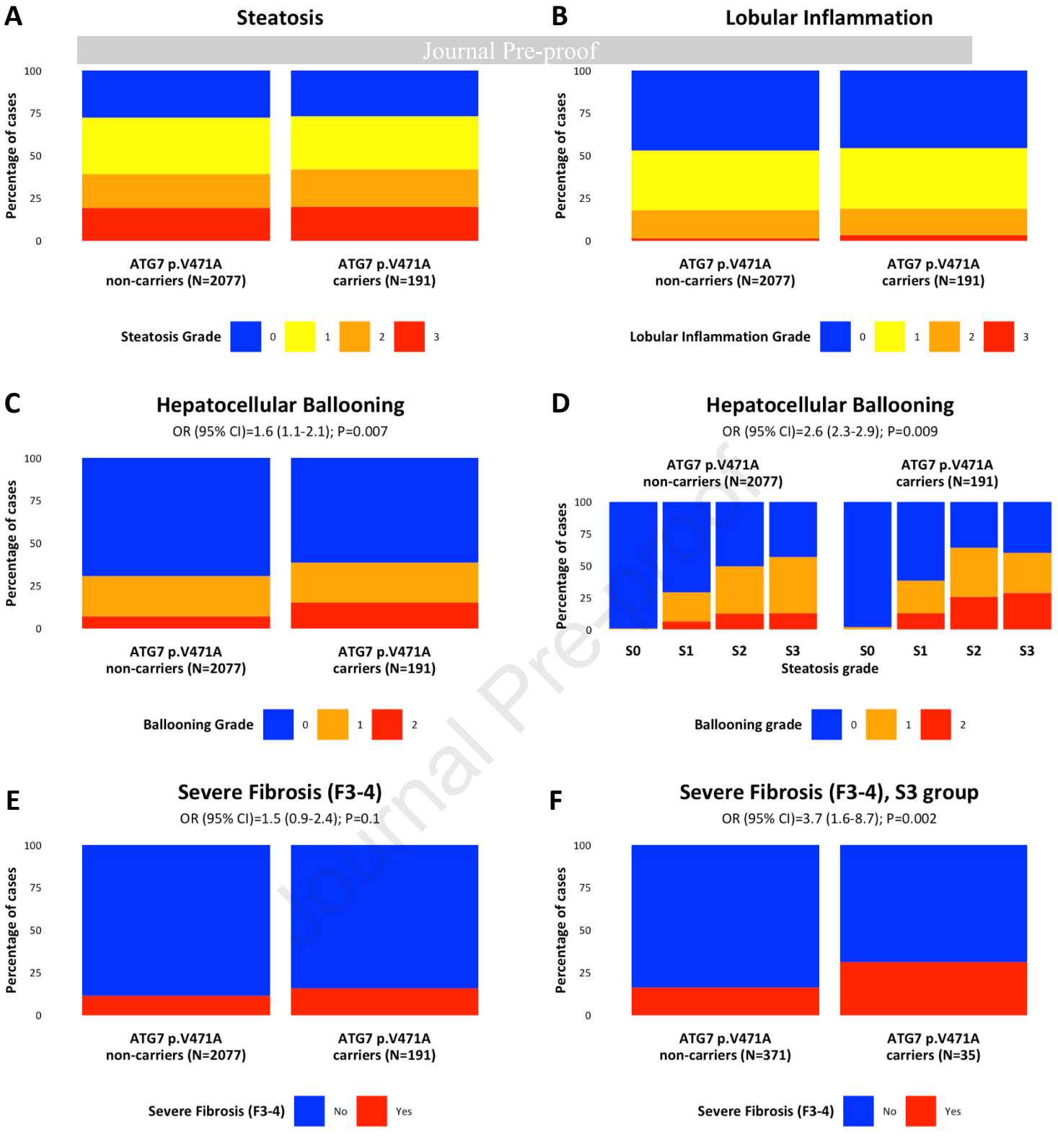


Figure 4

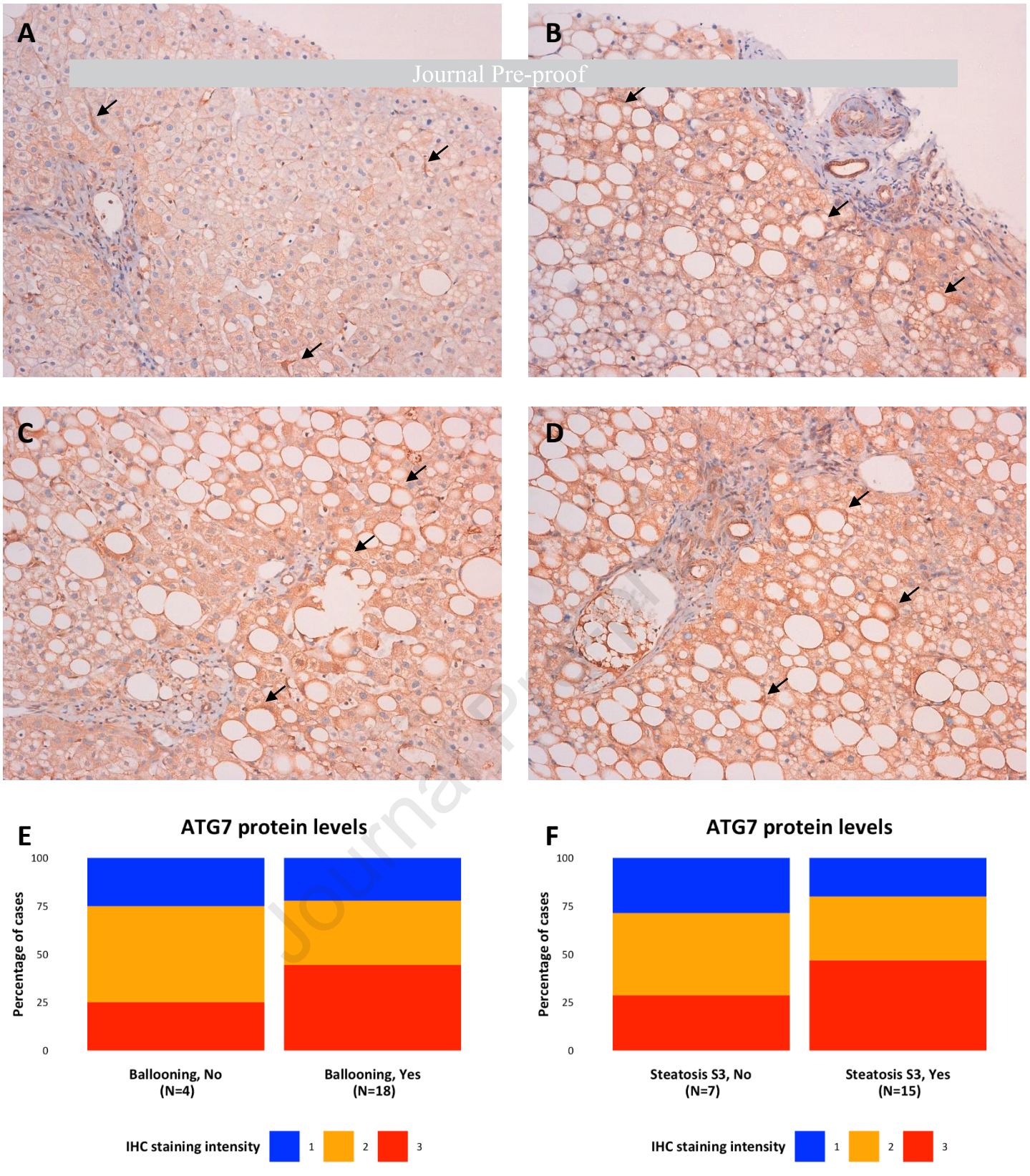


Figure 5

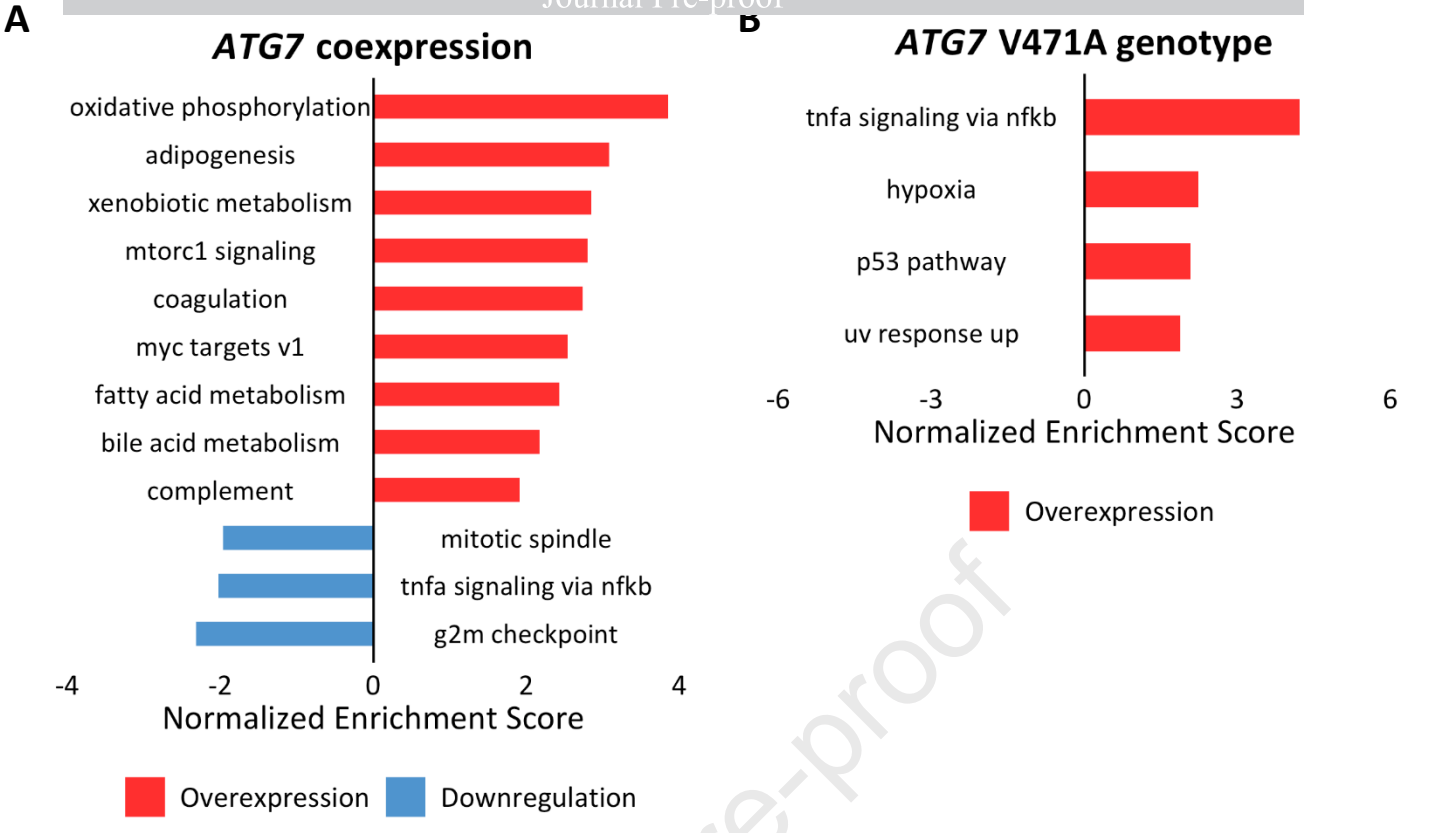


Figure S1

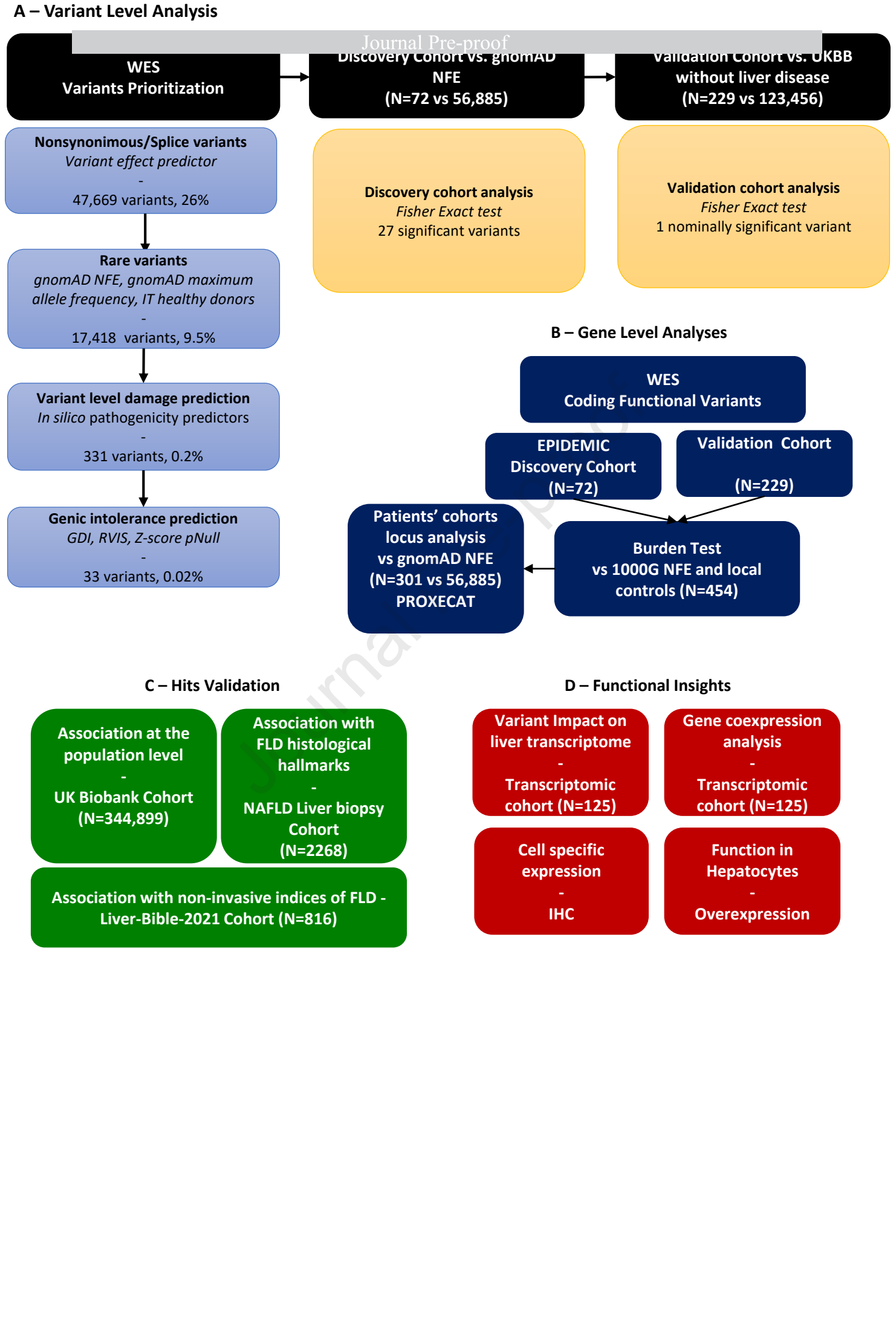


Figure S2

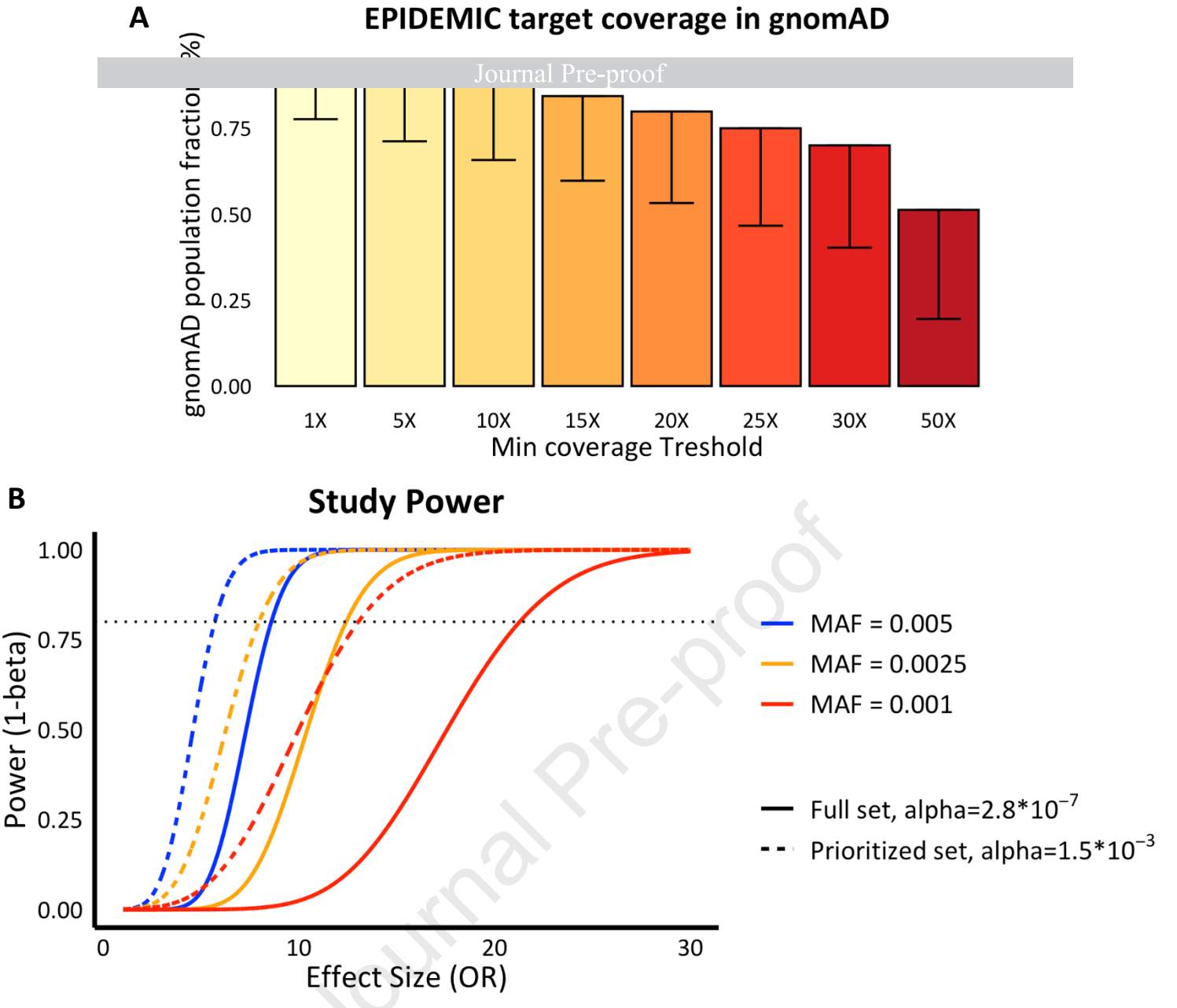
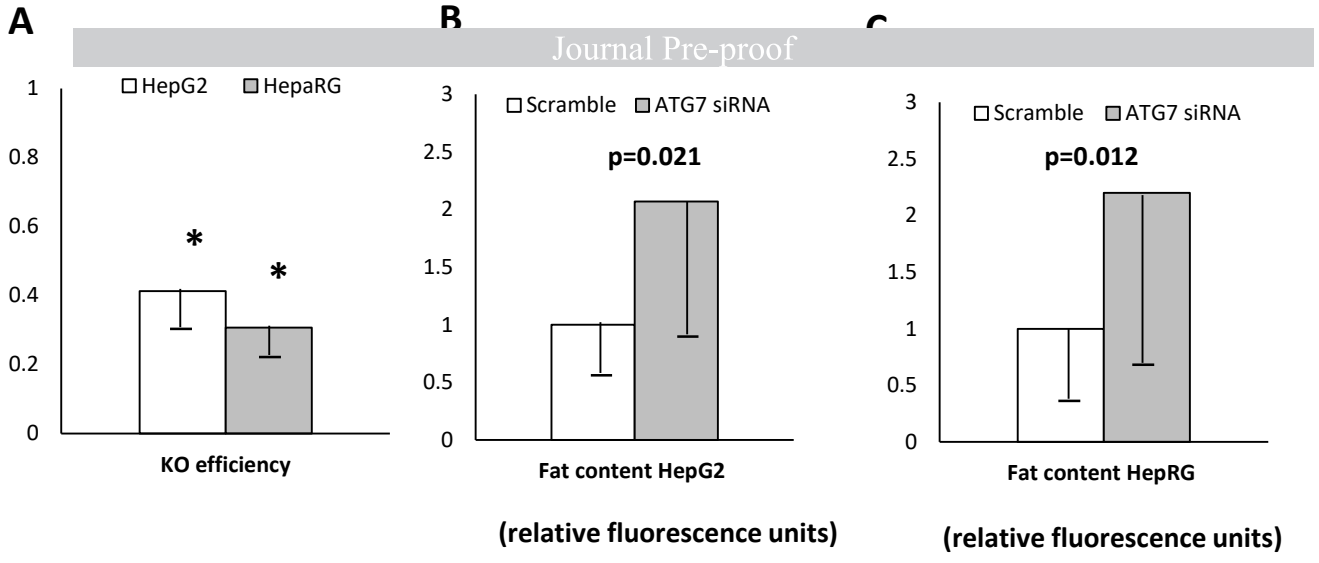
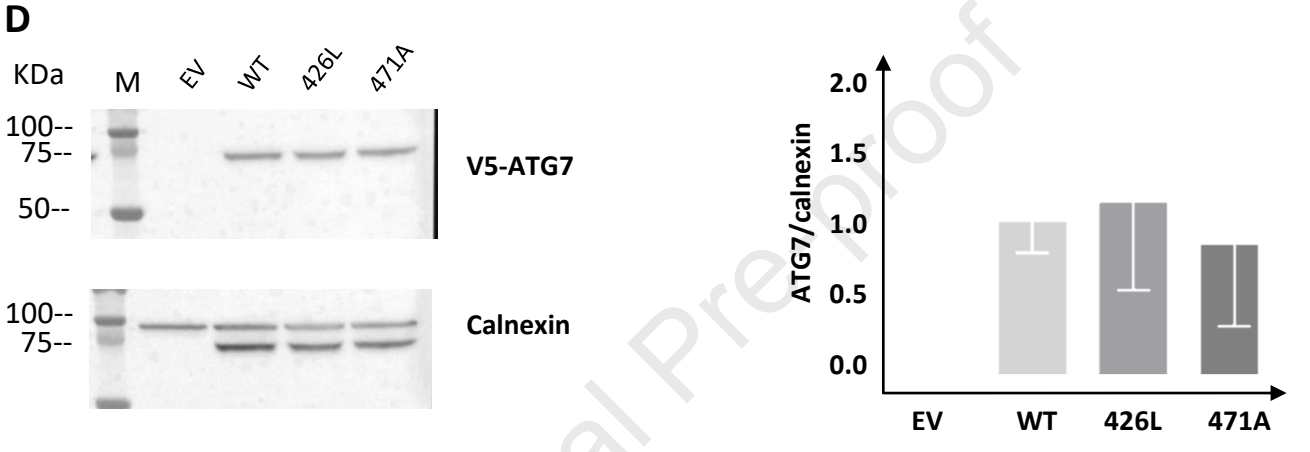


Figure S3





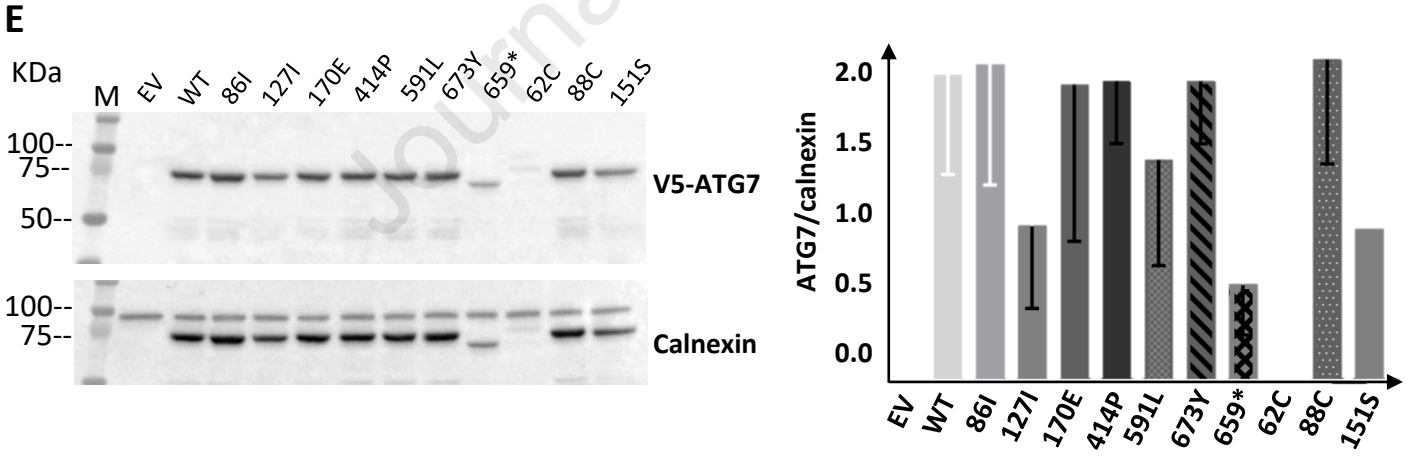
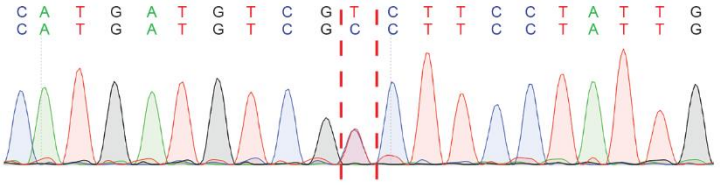


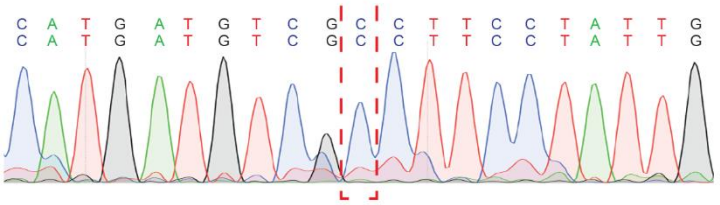
Figure S4

A

ATG7 p.



ATG7 p.V471A$^{+/+}$



B

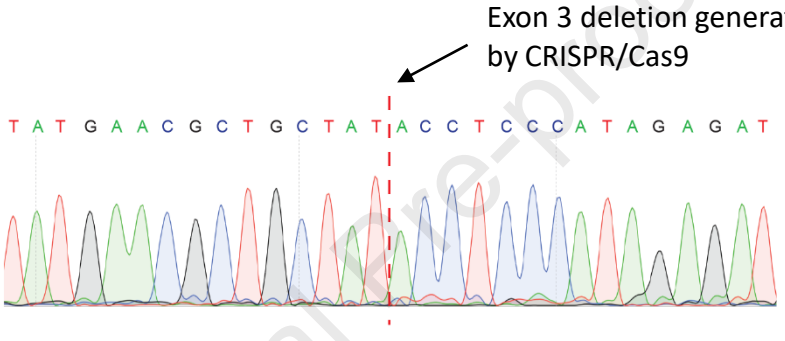
ATG7 p.V471A$^{+/-}$

```
                                        Exon 13
                  L   I   E   S   H   D   V   V   F   L   L   M   D   T   R
Allele 1   CTC ATC GAA AGC CAT GAT GTC GTC TTC CTA TTG ATG GAC ACC AGG
Allele 2   CTC ATC GAA AGC CAT GAT GTC GCC TTC CTA TTG ATG GAC ACC AGG
                  L   I   E   S   H   D   V   A   F   L   L   M   D   T   R
```

ATG7 p.V471A$^{+/+}$

```
                                        Exon 13
                  L   I   E   S   H   D   V   A   F   L   L   M   D   T   R
Allele 1   CTC ATC GAA AGC CAT GAT GTC GCC TTC CTA TTG ATG GAC ACC AGG
Allele 2   CTC ATC GAA AGC CAT GAT GTC GCC TTC CTA TTG ATG GAC ACC AGG
                  L   I   E   S   H   D   V   A   F   L   L   M   D   T   R
```

C



D

```
                                        Intron 2
ATG7 wt      ttgttatgaa cgctgctatt tctgacttgt cattgaaaat gtgataatgc ttctgtccag
ATG7 KO-/-   ttgttatgaa cgctgctat- ---------- ---------- ---------- ----------

                                        Exon 3
ATG7 wt      GTGACTCTGC TGGGCTGCCA GCTCGCTTAA CATTGGAGTT CAGTGCTTTT GACATgtgag
ATG7 KO-/-   ---------- ---------- ---------- ---------- ---------- ----------

                                        Intron 3
ATG7 wt      tatttatttg ttcaaaatct gaagtaaaga atactacttt tggcaaggaa taagcatgct
ATG7 KO-/-   ---------- ---------- ---------- ---------- ---------- ----------

ATG7 wt      tgcctccctc ataggtggac cacaggagga ctagggaagt tcccggtggg cagagtcagc
ATG7 KO-/-   ---------- ---------- ---------- ---------- ---------- ----------

ATG7 wt      ccccctcatg attctcttat acctcccata gagatgagtg ctggatactt aggtctgcct
ATG7 KO-/-   ---------- ---------- acctcccata gagatgagtg ctggatactt aggtctgcct
```

E

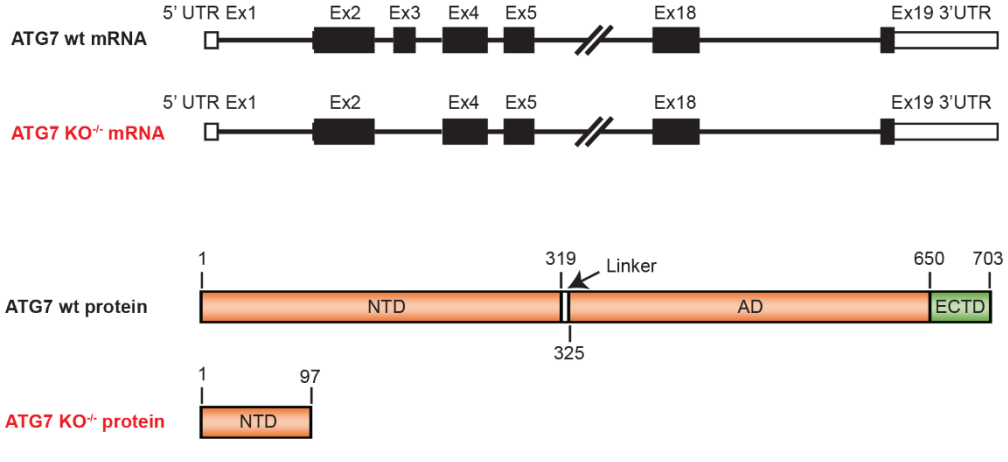


Figure S5

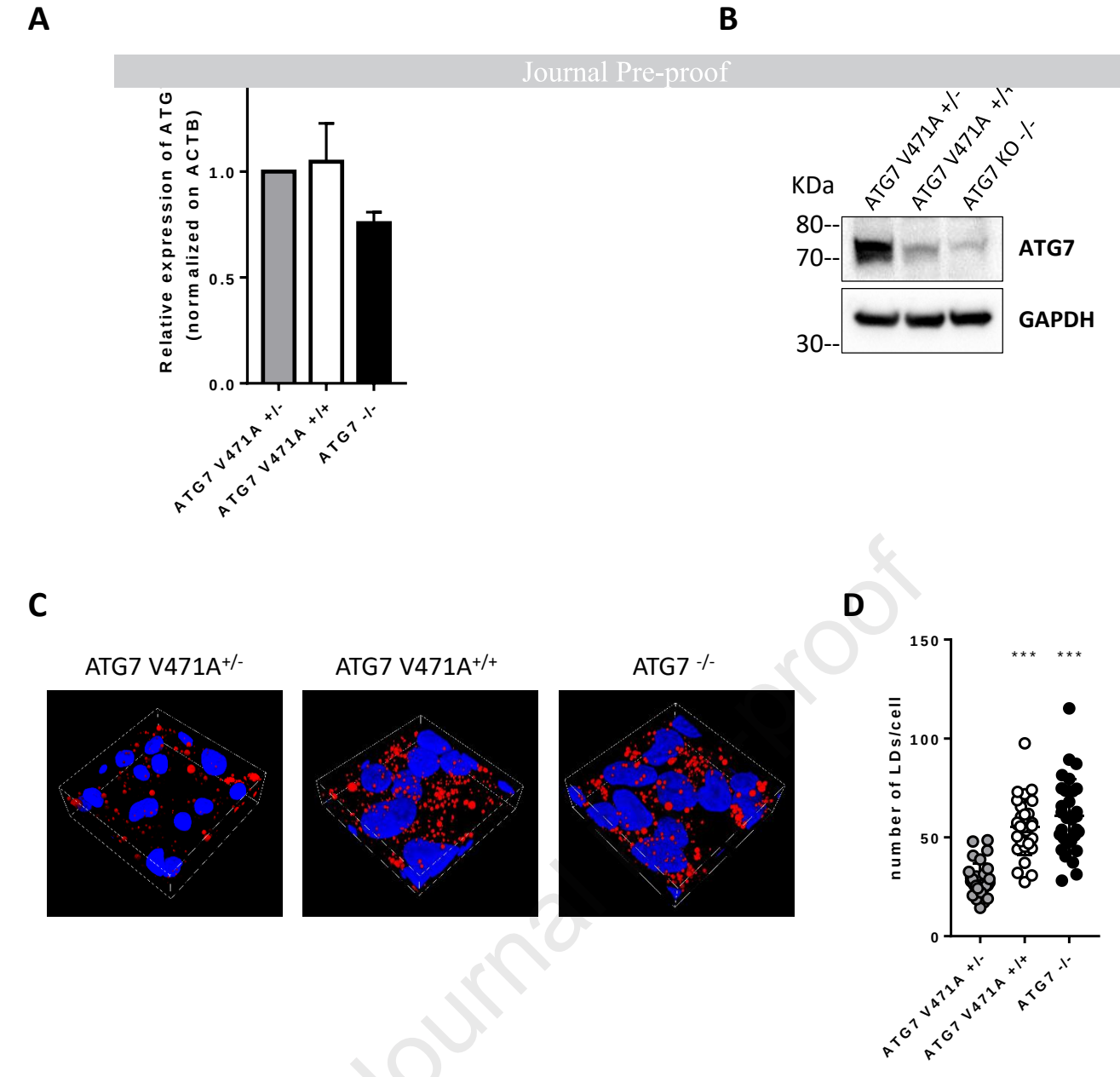
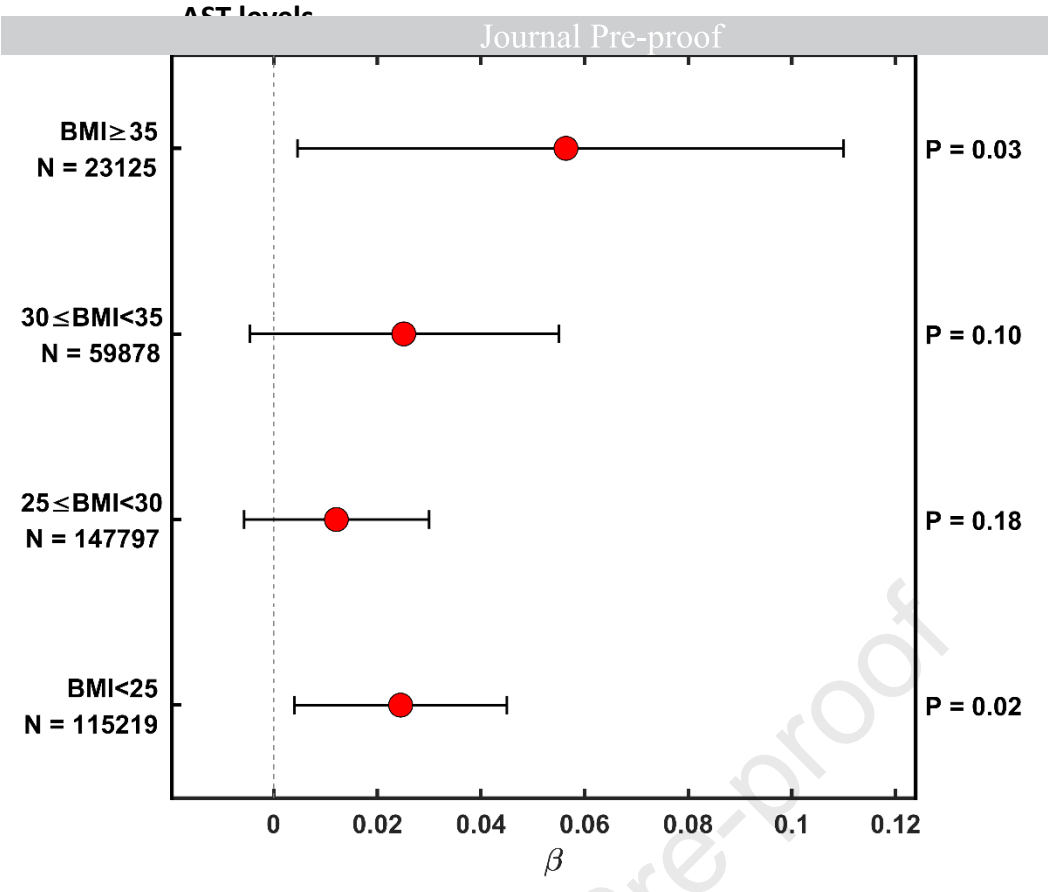


Figure S6

A



B

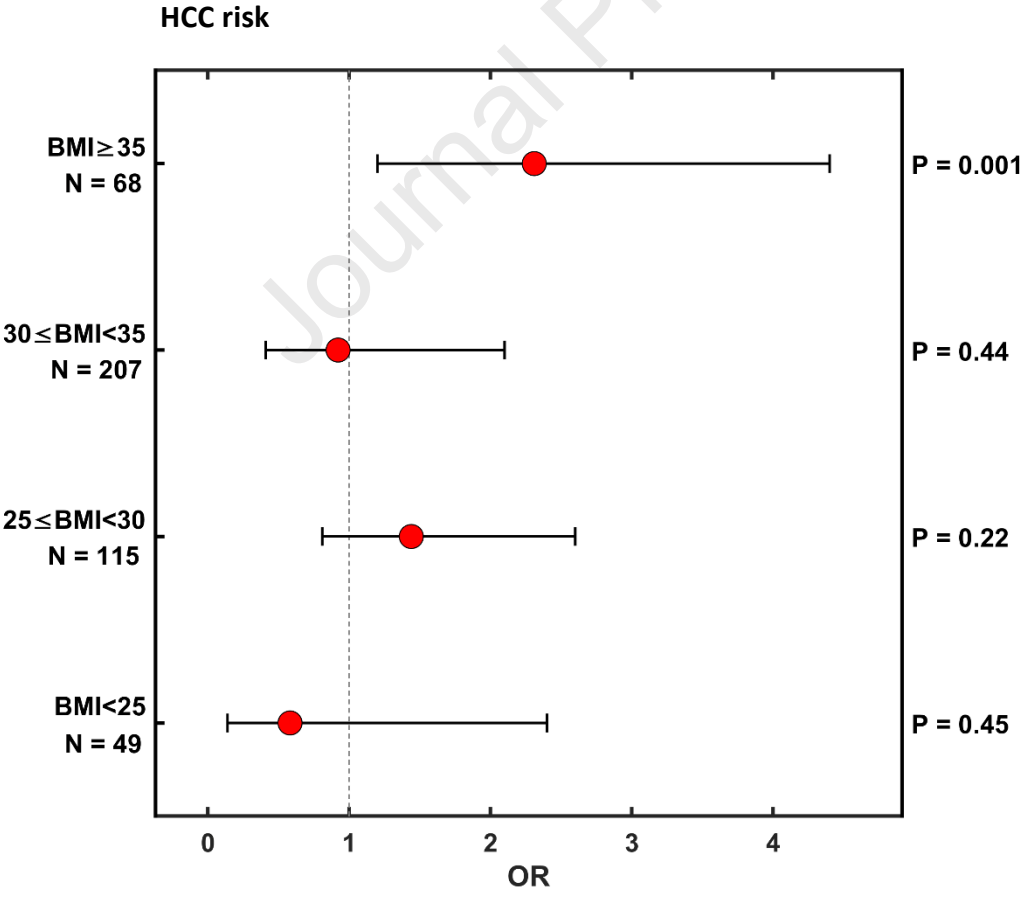


Figure S7

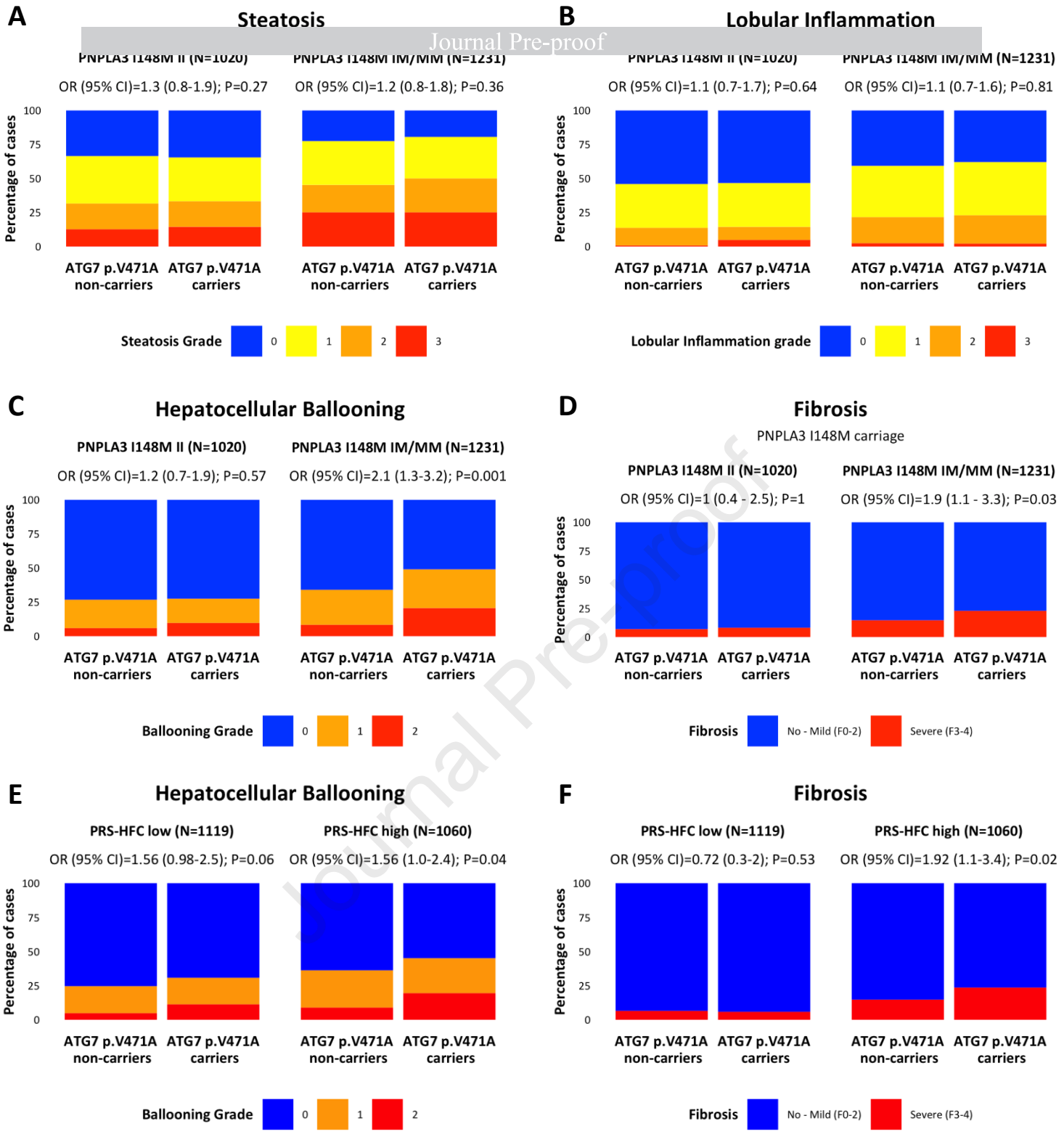


Figure S8

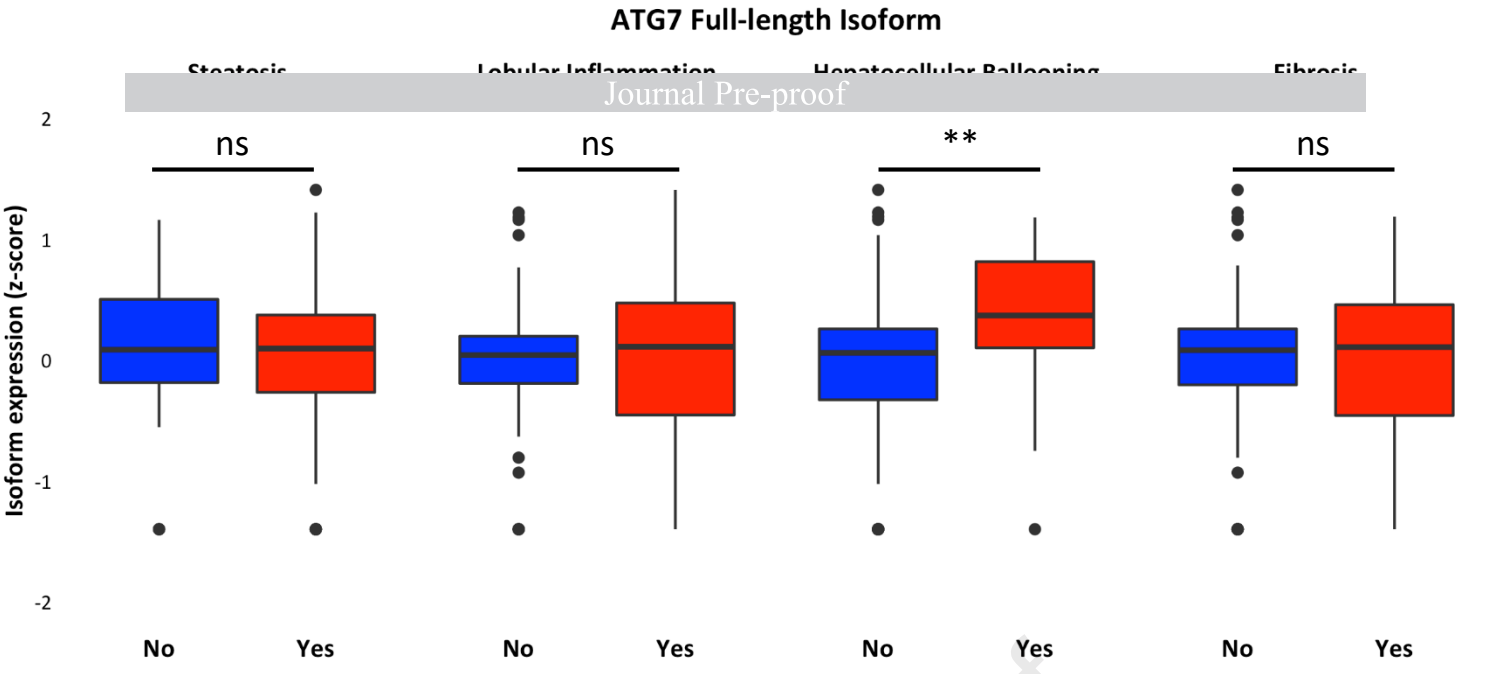
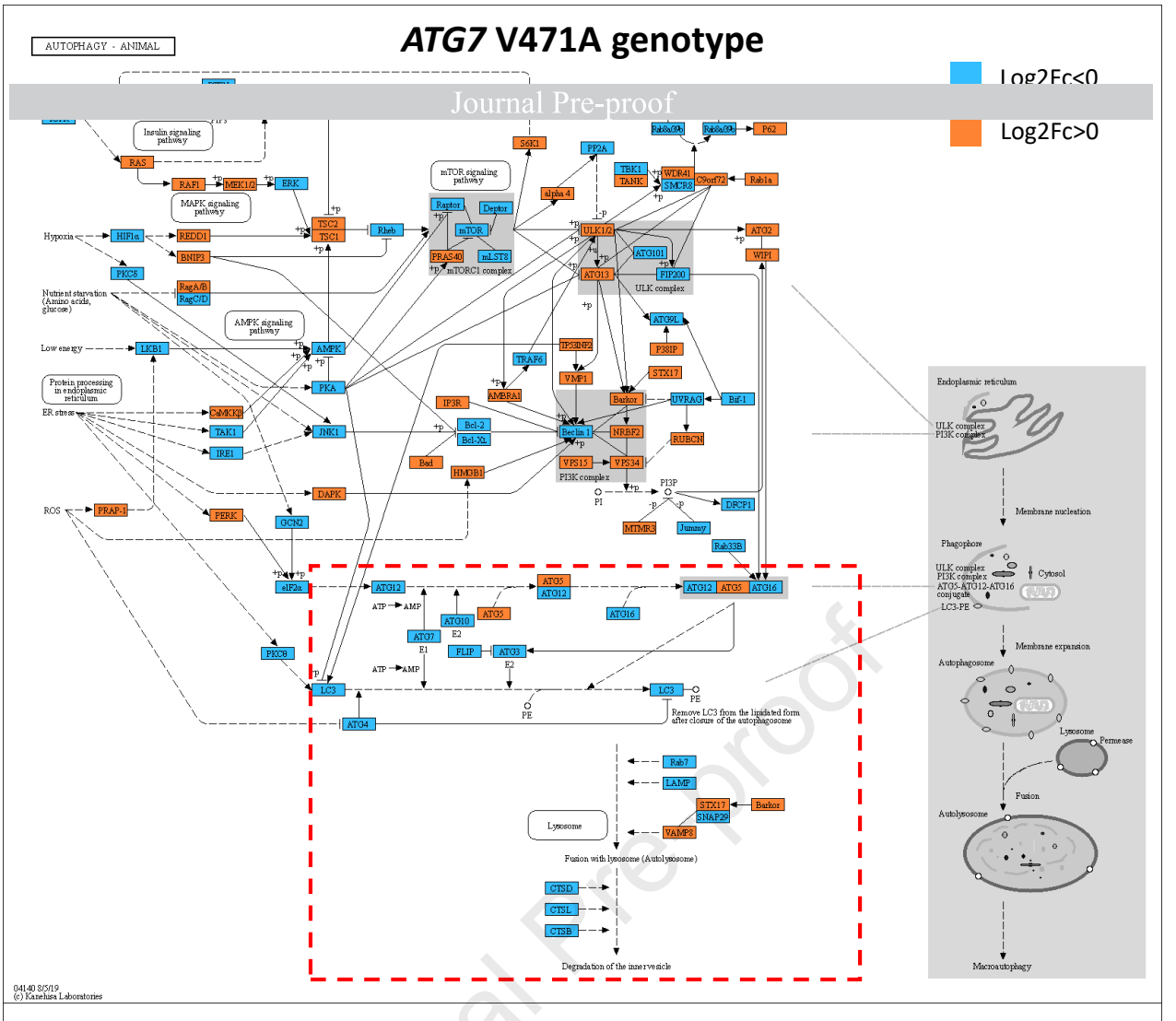


Figure S9

A


AUTOPHAGY - ANIMAL
ATG7 V471A genotype
Log2Fc<0
Log2Fc>0
Journal Pre-proof
Insulin signaling pathway
mTOR signaling pathway
MAPK signaling pathway
AMPK signaling pathway
Protein processing in endoplasmic reticulum
Lysosome
Fusion with lysosome (Autolysosome)
Degradation of the inner vesicle
Remove LC3 from the lipidated form after closure of the autophagosome
Endoplasmic reticulum
ULK complex PI3K complex
Membrane nucleation
Phagophore
Cytosol
LC3-PE
Membrane expansion
Autophagosome
Lysosome
Permease
Fusion
Autolysosome
Macroautophagy
04140 8/5/19
(c) Kanehisa Laboratories

B

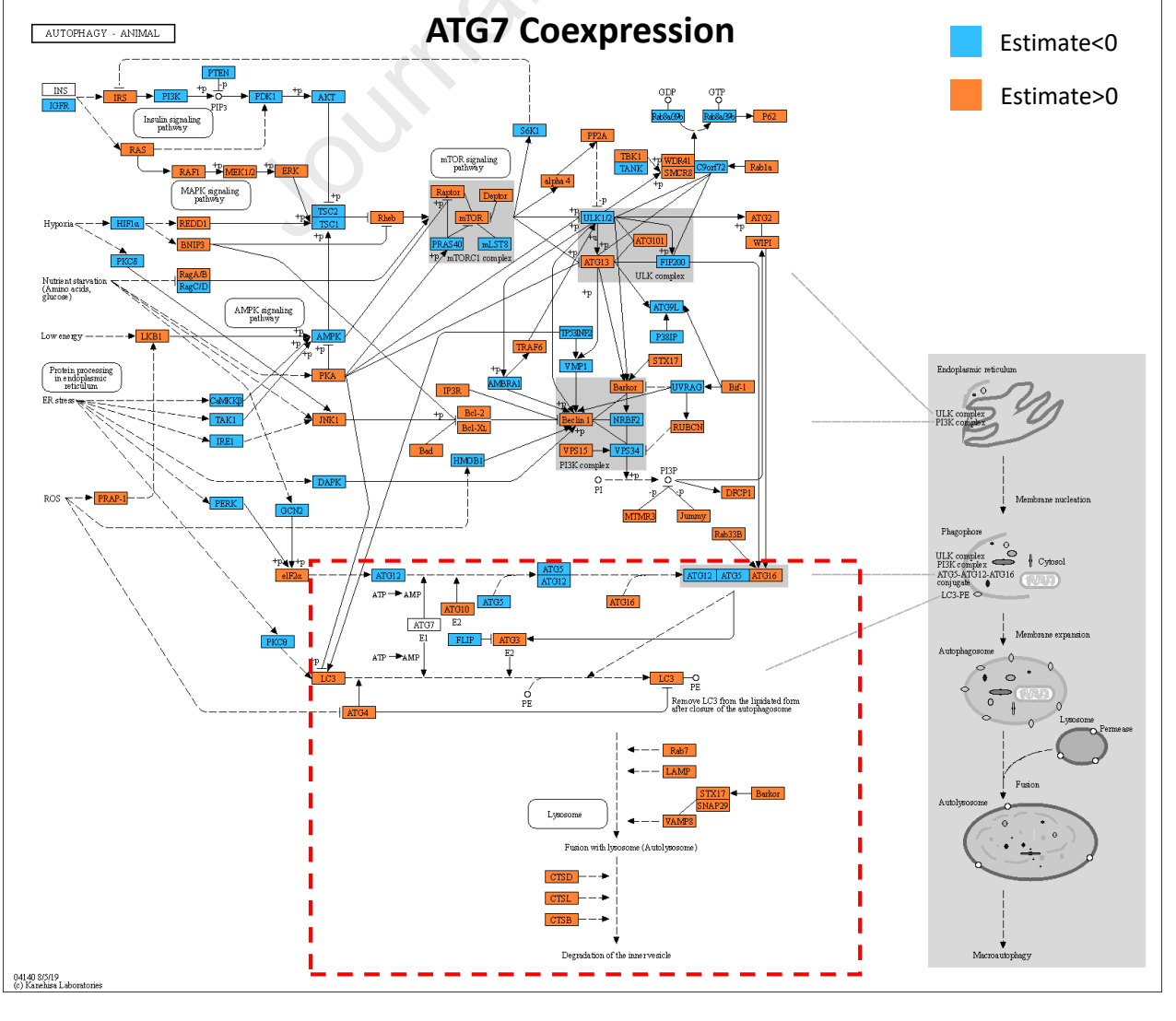

AUTOPHAGY - ANIMAL
ATG7 Coexpression
Estimate<0
Estimate>0
Journal Pre-proof
Insulin signaling pathway
mTOR signaling pathway
MAPK signaling pathway
AMPK signaling pathway
Protein processing in endoplasmic reticulum
Lysosome
Fusion with lysosome (Autolysosome)
Degradation of the inner vesicle
Remove LC3 from the lipidated form after closure of the autophagosome
Endoplasmic reticulum
ULK complex PI3K complex
Membrane nucleation
Phagophore
Cytosol
LC3-PE
Membrane expansion
Autophagosome
Lysosome
Permease
Fusion
Autolysosome
Macroautophagy
04140 8/5/19
(c) Kanehisa Laboratories

Figure S9C

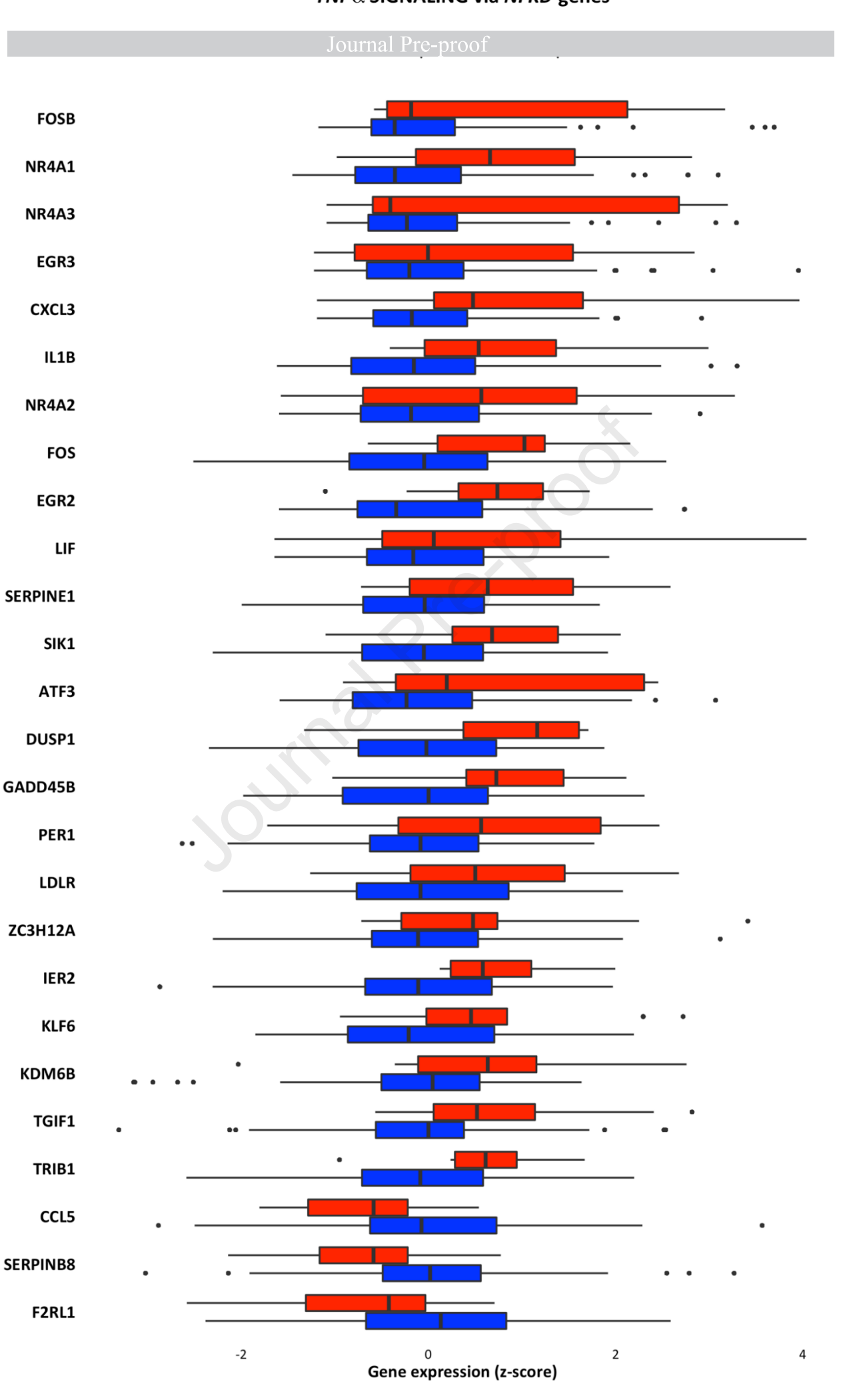
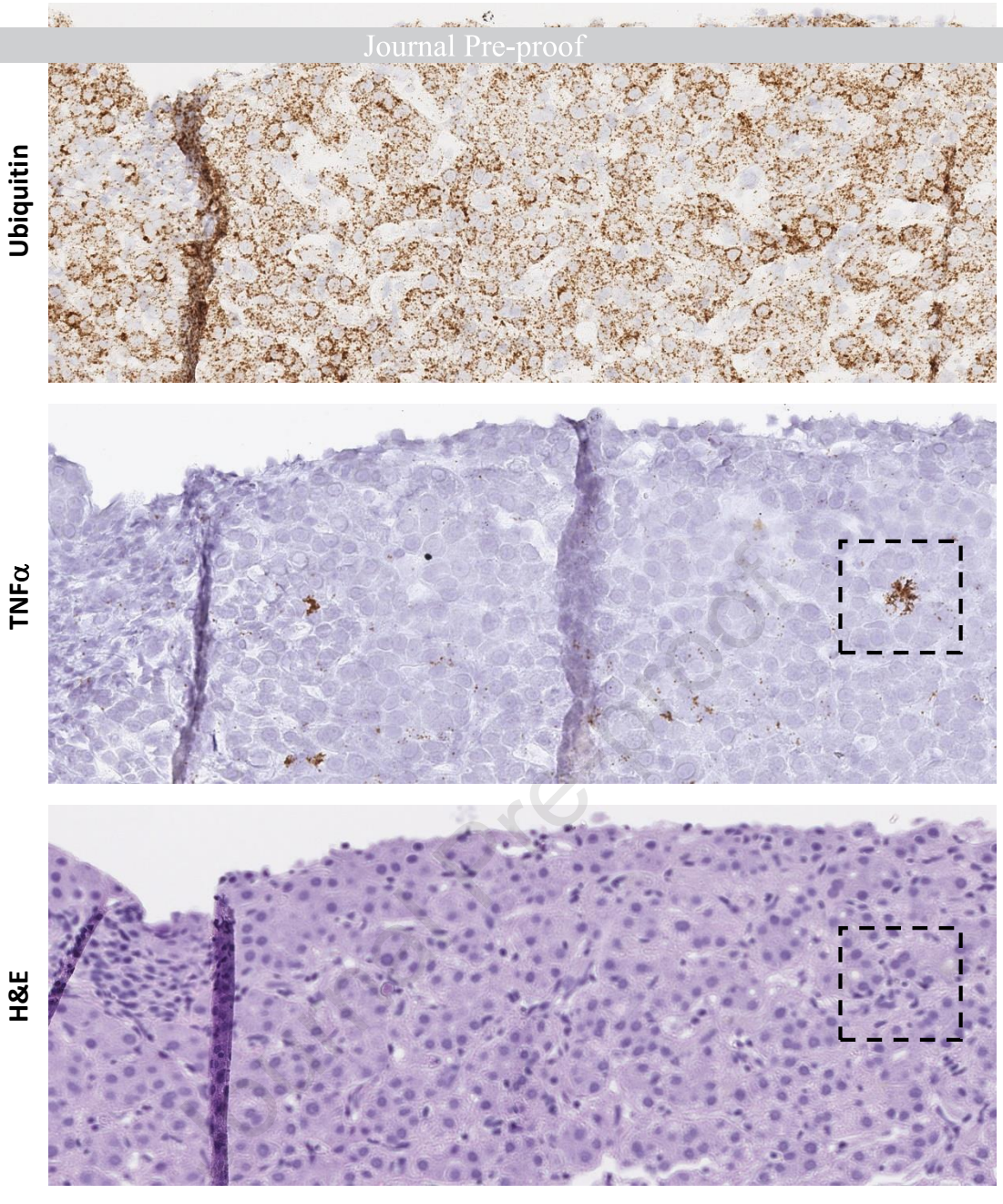
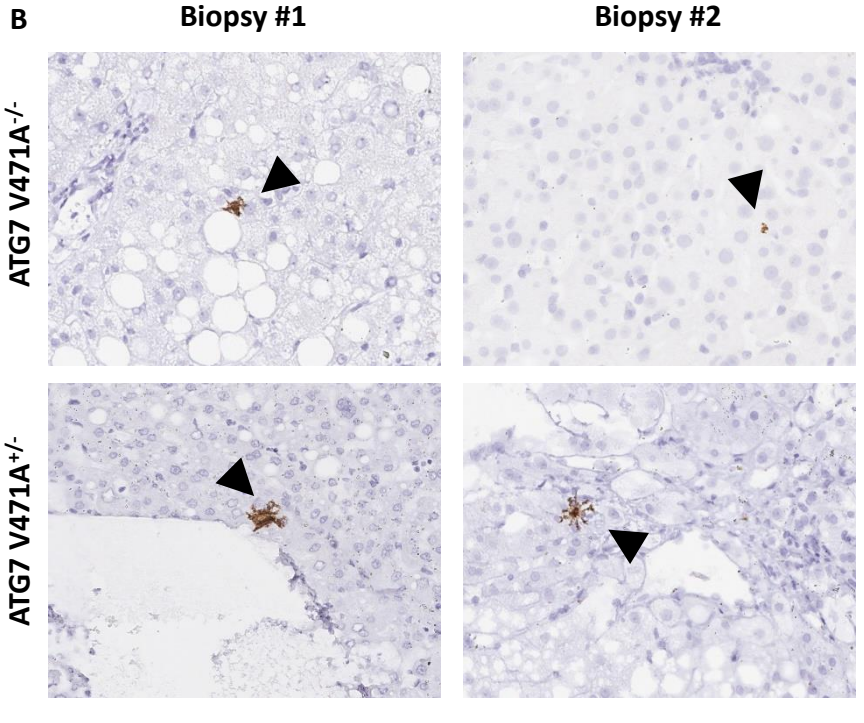
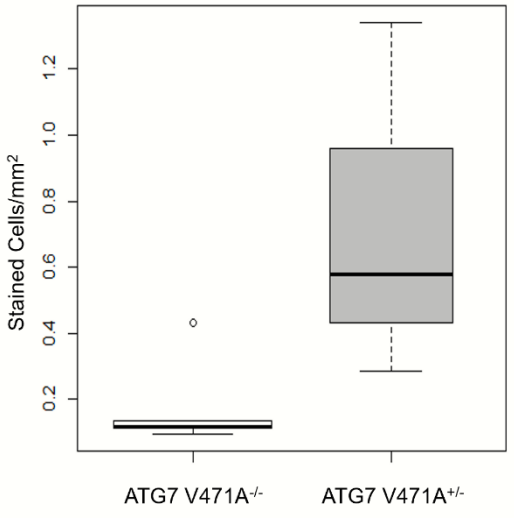

TNFα SIGNALING via NFKB genes
Journal Pre-proof
FOSB
NR4A1
NR4A3
EGR3
CXCL3
IL1B
NR4A2
FOS
EGR2
LIF
SERPINE1
SIK1
ATF3
DUSP1
GADD45B
PER1
LDLR
ZC3H12A
IER2
KLF6
KDM6B
TGIF1
TRIB1
CCL5
SERPINB8
F2RL1
-2
0
2
4
Gene expression (z-score)

Figure S10







Highlights:

- NAFLD is the leading cause of liver disorders and has a strong heritable component
- We found that rare loss-of-function mutations in the *ATG7* gene increase the risk of developing severe liver disease in individuals with dysmetabolism and patients with NAFLD
- ATG7 mutations cause an alteration in protein function and impairment of autophagy, leading to hepatocellular ballooning and inflammation
- The most frequent variant, namely rs36117895 T>C encoding for p.V471A, is responsible for a meaningful fraction of predisposition to ballooning and hepatocellular carcinoma