Journal of Pathology
J Pathol 2016
Published online in Wiley Online Library
(wileyonlinelibrary.com) DOI: 10.1002/path.4803

ORIGINAL PAPER

Mitochondrial genome architecture in non-alcoholic fatty liver disease

Silvia Sookoian,[1]* Diego Flichman,[2] Romina Scian,[1,3] Cristian Rohr,[4] Hernán Dopazo,[4] Tomas Fernández Gianotti,[3] Julio San Martino,[5] Gustavo O Castaño[6] and Carlos J Pirola[3]*

[1] *Department of Clinical and Molecular Hepatology, Institute of Medical Research A Lanari-IDIM, University of Buenos Aires – National Scientific and Technical Research Council (CONICET), Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina*
[2] *Department of Virology, School of Pharmacy and Biochemistry, University of Buenos Aires, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina*
[3] *Department of Molecular Genetics and Biology of Complex Diseases, Institute of Medical Research A Lanari-IDIM, University of Buenos Aires – National Scientific and Technical Research Council (CONICET), Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina*
[4] *Biomedical Genomics and Evolution Laboratory. Ecology, Genetics and Evolution Department, Faculty of Science, IEGEBA, University of Buenos Aires – National Scientific and Technical Research Council (CONICET), Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina*
[5] *Department of Pathology, Hospital Diego Thompson, San Martin, Buenos Aires, Argentina*
[6] *Liver Unit, Medicine and Surgery Department, Hospital Abel Zubizarreta, Ciudad Autónoma de Buenos Aires, Buenos Aires, Argentina*

**Correspondence to: S Sookoian, Instituto de Investigaciones Médicas A Lanari-IDIM-CONICET, Combatiente de Malvinas 3150, Buenos Aires (1427), Argentina. E-mail: sookoian.silvia@lanari.fmed.uba.ar*
Or CJ Pirola, Instituto de Investigaciones Médicas A Lanari-IDIM-CONICET, Combatiente de Malvinas 3150, Buenos Aires (1427), Argentina. E-mail: pirola.carlos@lanari.fmed.uba.ar

Abstract

Non-alcoholic fatty liver disease (NAFLD) is associated with mitochondrial dysfunction, a decreased liver mitochondrial DNA (mtDNA) content, and impaired energy metabolism. To understand the clinical implications of mtDNA diversity in the biology of NAFLD, we applied deep-coverage whole sequencing of the liver mitochondrial genomes. We used a multistage study design, including a discovery phase, a phenotype-oriented study to assess the mutational burden in patients with steatohepatitis at different stages of liver fibrosis, and a replication study to validate findings in loci of interest. We also assessed the potential protein-level impact of the observed mutations. To determine whether the observed changes are tissue-specific, we compared the liver and the corresponding peripheral blood entire mitochondrial genomes. The nuclear genes *POLG* and *POLG2* (mitochondrial DNA polymerase-γ) were also sequenced. We observed that the liver mtDNA of patients with NAFLD harbours complex genomes with a significantly higher mutational (1.28-fold) rate and degree of heteroplasmy than in controls. The analysis of liver mitochondrial genomes of patients with different degrees of fibrosis revealed that the disease severity is associated with an overall 1.4-fold increase in mutation rate, including mutations in genes of the oxidative phosphorylation (OXPHOS) chain. Significant differences in gene and protein expression patterns were observed in association with the cumulative number of OXPHOS polymorphic sites. We observed a high degree of homology (~98%) between the blood and liver mitochondrial genomes. A missense *POLG* p.Gln1236His variant was associated with liver mtDNA copy number. In conclusion, we have demonstrated that OXPHOS genes contain the highest number of hotspot positions associated with a more severe phenotype. The variability of the mitochondrial genomes probably originates from a common germline source; hence, it may explain a fraction of the 'missing heritability' of NAFLD.

Keywords: fatty liver; non-alcoholic fatty liver disease; NASH; mitochondrial dysfunction; gene expression; liver fibrosis; mitochondrial genome

Received 10 May 2016; Revised 29 July 2016; Accepted 18 August 2016

No conflicts of interest were declared.

Introduction

Non-alcoholic fatty liver (NAFL) disease (NAFLD) is a common chronic liver disease whose prevalence has reached global epidemic proportions [1]. NAFLD may progress from a relatively benign and mild histological stage (simple steatosis or NAFL) to a more severe form characterized by liver cell injury, a mixed inflammatory lobular infiltrate, and variable fibrosis, known as non-alcoholic steatohepatitis (NASH) [2]. The occurrence of NASH is of particular clinical importance, as it may progress to cirrhosis and eventually develop into hepatocellular carcinoma [3].

The disease pathogenesis and severity are both determined by multiple factors that include not only genetic and environmental influences, but also the interplay

between them, as mediated by epigenetic modifications [4]. The development and progression of NAFLD both entail a reprogramming of liver metabolism to fit the stressful metabolic environment caused by the intracellular lipid overload [5,6]. Recent evidence from experimental [7,8] and human studies [9–11] have consistently demonstrated that NAFLD is associated with impaired tricarboxylic acid cycle functioning, which takes place primarily in the mitochondrial matrix. The above-mentioned metabolic impairment is explained by liver mitochondrial dysfunction, which has been shown to be strongly involved in the disease's biology [5,12–15]. In fact, patients with NAFLD not only show decreased liver mitochondrial DNA (mtDNA) copy numbers [16], but also downregulation of liver expression of members of the oxidative phosphorylation (OXPHOS) system, as explained by novel epigenetic mechanisms that affect liver mtDNA [14]. These molecular changes are of particular relevance, as both the amount of mtDNA in a cell and the full functioning of mtDNA molecules are necessary to maintain a normal energy production level [17]. Nevertheless, it is still unknown whether the observed defects in the mitochondrial function of patients with NAFLD are explained by defects in the liver mitochondrial genomes. To understand the clinical implications of mtDNA diversity in the pathogenesis of NAFLD, we sequenced complete liver mitochondrial genomes of patients and controls under the hypothesis that NAFLD severity is associated with a mutational spectrum affecting the protein-encoding genes of the bioenergetics mitochondrial machinery. Furthermore, we explored the role of the variability of liver mitochondrial genomes in the severity of NAFLD. To analyse whether the observed changes in the liver mitochondrial genomes are tissue-specific mutations acquired in response to the associated metabolic stress, or, in contrast, if the observed mutational profile is part of the inherited variability of the mtDNA that predisposes patients to develop the disease, we further compared entire mitochondrial genomes from the liver and corresponding blood samples. The aforementioned explorations were all performed by deep-coverage whole sequencing of the mtDNA. Finally, we sequenced all of the nuclear genes that encode mitochondrial DNA polymerase-γ (*POLG* and *POLG2*), which is involved in mtDNA replication, to explore its contribution to the liver mtDNA copy number.

Patients and methods

Study design

Liver biopsies from all patients, and serum samples and DNA from white blood cells, were obtained with written, informed consent in accordance with the Institutional Review Board-approved protocols (protocol number: 104/HGAZ/09, 89/100 and 1204/2012, Comite de Bioetica Hospital Abel Zubizarreta, 2009/2010/2012). All of the investigations performed in this study were conducted in accordance with the guidelines of the 1975 Declaration of Helsinki, as revised in 1993.

To estimate the genome-wide liver mtDNA variability in NAFLD and its impact on disease severity, we used a multistage study design. In Study 1 (discovery study), we explored the diversity of whole liver mtDNA in a case–control study of patients with NAFLD ($n = 20$) and matched controls ($n = 8$). In Study 2, which was focused on disease severity, we included an independent exploration of the liver mtDNA mutational landscape in patients with NASH at different stages of liver fibrosis, including absent or mild fibrosis (scores 0–1, $n = 10$), versus patients with moderate or advanced fibrosis (scores 2–4, $n = 14$). Study 3 was a replication study performed in an independent sample of unrelated patients with NAFLD (NAFL, $n = 62$; NASH, $n = 76$) and healthy subjects ($n = 100$) that was oriented to validate findings of target mitochondrial variants in coding regions with significant associations with disease severity. In addition, the phenotypic impact of OXPHOS genomic variation on both gene and protein expression levels was explored.

In Study 4, to explore whether the observed variability in the liver mitochondrial genomes was somatically acquired (tissue-specific) or was part of a general mutational or variation pattern (germline), we compared paired complete mtDNA sequences ($n = 18$) from liver tissue and peripheral blood samples.

Finally, in Study 5, we explored the genetic variation of *POLG* and *POLG2* by targeted next-generation sequencing (NGS) in a larger sample of 96 individuals, including 64 NAFLD patients and 32 control subjects.

Details of physical, anthropometric and biochemical evaluations, liver biopsy and histopathological evaluation, including immunohistochemistry and electron microscopy, quantification of the liver mtDNA copy number and quantitative reverse transcription polymerase chain reaction (RT-qPCR) for quantitative assessment of mRNA levels are shown in Supplementary materials and methods. A graphical summary of the five studies is presented in supplementary material, Figure S1. Primer sequences are shown in supplementary material, Table S1.

Sequencing of complete liver mitochondrial genomes by NGS and the characterization of genetic variation in *POLG* and *POLG2* by deep-coverage targeted sequencing

NGS technology (316 chip; Ion Torrent PGM system; Life Technologies, Carlsbad, CA, USA) was employed to search for the presence of genetic variation/mutations in the entire circular double-stranded (16 569 bp) mtDNA from liver and blood specimens. Complete details on the amplification of entire mtDNA by long-range polymerase chain reaction, NGS, variant calling, estimation of quality control, data analysis and prediction of variant/mutation effect and NGS for the *POLG* and *POLG2* genetic variations are provided in Supplementary materials and methods. Primer

Table 1. Study 1 (discovery case–control study): clinical and biochemical features of control subjects and patients with non-alcoholic fatty liver disease (NAFLD)

Variable	Controls	NAFL	NASH
Number of subjects	8	10	10
Female (*n*)	4	6	5
Age (years)	40 ± 10	46 ± 6	48 ± 10
Physical activity (h/week)	0 ± 0	0.66 ± 1.6	0.2 ± 0.6
BMI (kg/m^2)	24.5 ± 2.5	30.3 ± 3.8*	30.5 ± 5.0†
Fasting plasma glucose (mmol/l)	4.88 ± 0.72	5.38 ± 0.89	5.61 ± 0.44†
Fasting plasma insulin (pmol/l)	41.7 ± 3.5	111.1 ± 104.2*	69.5 ± 23.6†
HOMA-IR index	1.2 ± 0.3	3.7 ± 3.0*	3.0 ± 1.0†
SABP (mmHg)	110 ± 14	119 ± 18	122 ± 9
DABP (mmHg)	65 ± 2	75 ± 15	76 ± 7
Total cholesterol (mmol/l)	5.12 ± 0.91	5.53 ± 1.11	6.03 ± 1.91
HDL cholesterol (mmol/l)	1.60 ± 0.39	1.42 ± 0.39	1.45 ± 0.44
LDL cholesterol (mmol/l)	2.82 ± 1.16	2.69 ± 1.37	3.62 ± 1.63
Triglycerides (mmol/l)	1.46 ± 0.38	2.20 ± 0.77	2.54 ± 1.60
Uric acid (mmol/l)	262 ± 149	274 ± 119	357 ± 119
ALT (U/l)	69 ± 27	58 ± 18	66 ± 48
AST (U/l)	45 ± 12	36 ± 6	35 ± 18
GGT (U/l)	84 ± 27	85 ± 50	64 ± 31
AP (U/l)	217 ± 146	213 ± 138	180 ± 84
Degree of steatosis (%)	0	45 ± 32	71 ± 21
Lobular inflammation (0–3)	0	0.4 ± 0.5	1.1 ± 0.6‡
Portal inflammation (0–2)	0	0.0 ± 0.0	1.6 ± 0.7‡
Hepatocellular ballooning (0–2)	0	0.0 ± 0.0	1.1 ± 0.5‡
Fibrosis stage	0	0.0 ± 0.0	1.4 ± 1.1‡
NAS	NA	2.4 ± 1.4	5.3 ± 0.9‡

ALT, serum alanine aminotransferase; AP, alkaline phosphatase; AST, serum aspartate aminotransferase; BMI, body mass index; DABP, diastolic arterial blood pressure; GGT, γ-glutamyltransferase; HDL, high-density lipoprotein; HOMA-IR, homeostatic model assessment insulin resistance; LDL, low-density lipoprotein; NA, not applicable; NAFL, non-alcoholic fatty liver or simple steatosis; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; SABP, systolic arterial blood pressure.
Results are expressed as mean ± standard deviation.
*Statistically significant difference ($p < 0.01$) between NAFL and control.
†NAFL versus NASH.
‡NASH versus controls.
The *p*-value was obtained with the Mann–Whitney *U*-test, with the exception of the female/male ratio, where the *p*-value was obtained with the chi-squared test.

sequences used for the amplification of complete mtDNA prior to sequencing, to avoid nuclear mtDNA sequences, are shown in supplementary material, Table S2.

Statistical analysis

Quantitative data are expressed as mean ± standard deviation, unless otherwise indicated; complete details are provided in Supplementary materials and methods.

Results

NAFLD is associated with an increased mtDNA mutation rate

The clinical and biochemical features of patients and controls included in Study 1 are shown in Table 1. On the basis of alignments to the reference sequence (NC_012920.1), we achieved an average read depth (coverage) of >800 per individual (see supplementary material, Table S3). Liver mtDNA sequencing revealed a total of 689 variants, which represent changes in 227 mitochondrial positions (supplementary material, Table S4); a novel variant in *MT-CYB* was observed in two patients with NAFLD (m.15728 http://www.mitomap.org/foswiki/bin/view/MITOMAP, last accessed 16 June 2016) but not in controls. Remarkably, 525 of 689 variants (76%) were found in NAFLD patients, showing an overall 1.28-fold increase in mutation rate as compared with controls ($p = 0.0056$). The mutation rates of controls and patients with NAFL were not significantly different (Figure 1A). Conversely, patients with NASH had a significantly higher number of variants in *MT-ATP6*, *MT-CYB* and members of the NADH dehydrogenase complex (*ND1*, *ND2*, *ND3*, *ND4*, *ND4L*, *ND5*, and *ND6*) (Figure 1A) than controls. As a whole, patients with NASH had a 1.5-fold increased mutation rate as compared with controls ($p = 3.2E^{-5}$). A comparison of liver mtDNA diversity among patients with NAFL and NASH showed that disease severity was associated with an increased number of variants in members of the NADH dehydrogenase complex ($p = 7.4E^{-3}$) (Figure 1A).

Overall, we identified 134 non-synonymous mtDNA variants: 46 (33.6%) of them were predicted to be deleterious by the SIFT tool, and 26.86% were predicted to be possible/probably damaging by Polyphen (possible, 11.19%; probable, 14.92%); 40% were located in mitochondrially encoded cytochrome *b* (*MT-CYB*), 21% were located in mitochondrially encoded ATP synthase 6 (*MT-ATP6*), and 27.6% were located in other genes of the OXPHOS group.

Patients with NAFLD harboured a significantly higher proportion of heteroplasmic mtDNA variants

We also analysed the proportion of putative pathogenic mutations coexisting with wild-type mtDNA, which is a phenomenon known as heteroplasmy. The overall liver mtDNA heteroplasmy rate among subjects of the control group was 0.08 ± 0.03 (number of mtDNA sites, $n = 4$) versus 0.13 ± 0.08 in patients with NAFLD (number of mtDNA sites, $n = 15$) (χ^2 test, $p = 0.048$). The rate of liver mtDNA heteroplasmy was significantly higher across coding mtDNA regions (0.22 ± 0.08) than in the D-loop (0.08 ± 0.02, $p = 0.0001$); results are expressed as a fraction of 1.

It is important to note that 10 of 16 site positions containing heteroplasmic variants were reported to be highly polymorphic (>0.1), and all of them were observed in patients with NAFLD (supplementary material, Table S5). Interestingly, four of these 10 highly polymorphic variants were located in the control region, with the remaining six being located in OXPHOS-coding genes, including members of the mitochondrially encoded cytochrome *c* oxidase family and NADH dehydrogenase complex (supplementary material, Table S5). Furthermore, those variants located in mitochondrially encoded members of cytochrome

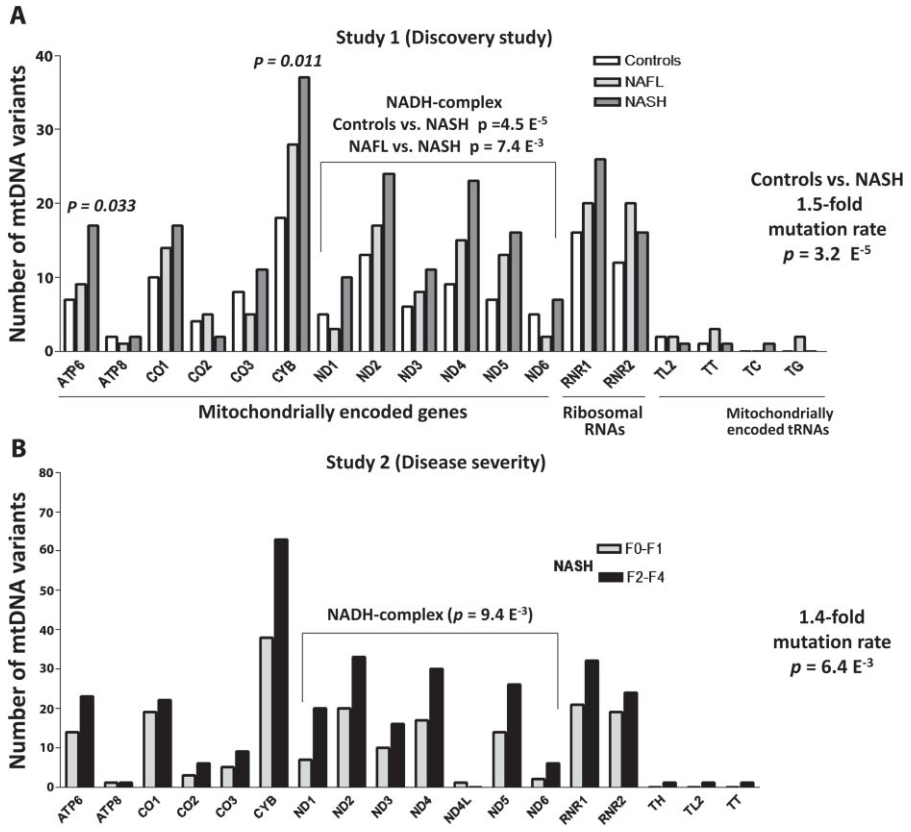

Figure 1. The role of mtDNA variability in NAFLD. (A) Summary of findings from Study 1: distribution of polymorphic sites per mitochondrial genomic region according to disease status. Significance stands for nominal *p*-values; empirical *p*-values are as follows: *ATP6*, $p = 0.10$; *CYB*, $p = 0.049$; and NADH dehydrogenase complex, $p = 0.0045$ (*p*-values are for the comparison between controls and NASH). Comparison between controls and NAFL: $p =$ NS. Empirical *p*-value for NADH dehydrogenase complex in the comparison between NAFL and NASH: $p = 0.029$. (B) Summary of findings from Study 2: the mtDNA variability in OXPHOS genes as the driving force of progressive NAFLD. The figure shows the mtDNA mutational landscape in livers of patients with NASH according to fibrosis status. Significance stands for nominal *p*-values; empirical *p*-value for NADH dehydrogenase complex: $p = 0.037$. Variants were defined as positions when the primary base differed from the reference rCRS, and were classified by location across the mitochondrial genome.

c oxidase and NADH family showed the highest level of heteroplasmy (>0.2). Finally, the liver mtDNA heteroplasmy rate was unrelated to the age of the individuals, as subjects who did not show mtDNA point heteroplasmic sites had similar ages (43.3 ± 8.6 years), unlike those who had heteroplasmic mtDNA variants (46.5 ± 27.0) ($p = 0.37$).

The analysis of genetic variability in tRNAs showed the trend to be higher in NAFLD patients ($p = 0.059$) than in controls (supplementary material, Table S6).

Analysis of haplogroup markers

On the basis of the analysis of haplogroup markers, mitochondrial genomes from patients and controls were equally distributed into two major geographical populations (European and American), including haplogroups H1 (representing the most common subclade in western Europe), U3/U5 (primarily representing the Caucasian lineage), and C and D (representing the American linage). Overall, there were no significant differences in either the total number of variants or variant distribution in the coding and non-coding regions and major haplogroups; moreover, there were no differences in the proportions of missense/synonymous variants between haplogroups.

MtDNA variability as a driving force of NASH and advanced fibrosis

To explore the hypothesis that the genetic diversity of mtDNA has a role in disease severity, we performed a second study that included complete sequencing of liver mitochondrial genomes of an independent sample of patients with NASH at different stages of liver fibrosis (Table 2).

As compared with patients who had mild or absent fibrosis, subjects with advanced fibrosis had an overall 1.4-fold increase in the mutation rate ($p = 6.4E^{-3}$) and a significantly higher number of mutations in members of the NADH dehydrogenase complex ($p = 9.4E^{-3}$) (Figure 1B). Supplementary material, Table S7 presents the novel missense mtDNA variants (defined as non-registered in the MITOMAP database; last update 16 June 2016) according to the severity of fibrosis.

Detailed analysis of the whole liver mtDNA mutational profile of nine patients of Study 2 whose disease worsened (progressed from a mild to an advanced fibrosis stage) as demonstrated by paired liver biopsies

Table 2. Study 2 (disease severity): clinical and biochemical features of patients with non-alcoholic steatohepatitis (NASH) at different stages of liver fibrosis

Variable	Fibrosis stage 0–1	Fibrosis stage 2–4*
Number of subjects	10	14
Female (*n*)	6	8
Age (years)	55.4 ± 9.0	45 ± 10, †$p = 0.01$
Physical activity (h/week)	1.0 ± 1.5	4.4 ± 8.8
BMI (kg/m^2)	32.0 ± 7.5	35.0 ± 6.5
Fasting plasma glucose (mmol/l)	6.05 ± 0.50	6.61 ± 1.61
Fasting plasma insulin (pmol/l)	115.3 ± 48.6	104.2 ± 45.1
HOMA-IR index	4.4 ± 1.8	3.8 ± 1.5
SABP (mmHg)	122 ± 22	120 ± 17
DABP (mmHg)	74.0 ± 9.0	76.4 ± 13.0
Total cholesterol (mmol/l)	5.59 ± 0.72	4.91 ± 1.09
HDL cholesterol (mmol/l)	1.22 ± 0.18	1.32 ± 0.39
LDL cholesterol (mmol/l)	3.44 ± 0.83	2.77 ± 0.72
Triglycerides (mmol/l)	1.92 ± 1.06	2.26 ± 1.71
Uric acid (mmol/l)	291 ± 119	363 ± 131
ALT (U/l)	49 ± 19	84 ± 49, †$p = 0.04$
AST (U/l)	37 ± 11	60 ± 27, †$p = 0.04$
GGT (U/l)	68 ± 68	69 ± 37
AP (U/l)	204 ± 90	188 ± 96
Degree of steatosis (%)	70 ± 18	61 ± 23
Lobular inflammation (0–3)	1.12 ± 0.35	0.85 ± 0.66
Portal inflammation (0–2)	1.3 ± 0.7	1.2 ± 0.9
Hepatocellular ballooning (0–2)	1.0 ± 0.5	0.9 ± 0.7
Characteristics of patients with paired liver biopsies		
Variable	**Baseline**	**Follow-up**
Number of subjects	9	9
Age (years)	48.6 ± 6.5	55 ± 4.7, †$p = 0.034$
ALT (U/l)	54 ± 18	79 ± 36
AST (U/l)	44 ± 20	49 ± 25
GGT (U/l)	73 ± 65	95 ± 98
AP (U/l)	214 ± 106	198 ± 48
Degree of steatosis (%)	63 ± 23	59 ± 18
Lobular inflammation (0–3)	0.67 ± 0.5	1.67 ± 0.87, †$p = 0.008$
Portal inflammation (0–2)	1.1 ± 0.3	1.2 ± 0.3
Hepatocellular ballooning (0–2)	0.2 ± 0.4	1.0 ± 0.7, $p = 0.01$
NAS	3.6 ± 1.7	5.7 ± 1, †$p = 0.0049$
Fibrosis stage (F0–F4)	0.3 ± 0.7	2.3 ± 1.3, †$p = 0.0007$

ALT, serum alanine aminotransferase; AP, alkaline phosphatase; AST, serum aspartate aminotransferase; BMI, body mass index; DABP, diastolic arterial blood pressure; GGT, γ-glutamyltransferase; HDL, high-density lipoprotein; HOMA-IR, homeostatic model assessment insulin resistance; LDL, low-density lipoprotein; NAS, non-alcoholic fatty liver disease activity score; SABP, systolic and arterial blood pressure.

Results are expressed as mean ± standard deviation.

*The large majority of patients had F2–F3 fibrosis, and only one case had F4 fibrosis.

†Indicates statistically significant differences.

The *p*-value was obtained with the Mann–Whitney *U*-test.

(baseline and follow-up 5.22 ± 1.92 years apart) showed that there were sites in the mitochondrial genomes that had probable non-random point mutations. Specifically, three sites (m.4769 A > G *ND2*, m.7028 C > T *CO1*, and m.14766 C > T *CYB*) were shared across all of the explored genomes; there was 100% similarity between the two samples in each patient. Upon further comparison with the control group, these mitochondrial sites were found to be significantly associated with disease status (m.4769, $p = 8.1E^{-3}$; m.7028, $p = 3.3E^{-3}$; and m.14766, $p = 0.028$).

Of particular interest is the m.14766 *CYB* missense variant, for which deleterious effects are predicted. Electron microscopy analysis of liver tissue from a patient who carried the three mentioned variants showed striking changes in mitochondrial morphology, including giant and elongated mitochondria, mitochondria with condensed matrix and signs of injury, loss of mitochondrial membranes and cristae, and proliferation of peroxisomes (supplementary material, Figure S2); all of these morphological findings have previously been reported to be associated with mitochondrial dysfunction [12,14,15].

The presence of heteroplasmic variants was significantly associated with liver fibrosis as a dichotomous trait (fibrosis absent or present) ($p = 0.042$) (supplementary material, Table S8), but not with the classification of stage of fibrosis. Overall, patients with advanced disease had high levels of heteroplasmy (results are expressed as a fraction of 1) for members of the OXPHOS group, including m.3579-*ND1* (0.41), m.9101-*ATP6* (0.21), m.15237 (0.10) and m.15314 (0.13) located in *CYB*, and the D-loop region (m.16278: 0.23).

The accumulation of mtDNA polymorphic sites in OXPHOS subunits is paralleled by the emergence of an OXPHOS-deficient phenotype

We next explored the potential functional impact of the observed mutations; for this purpose, we reasoned that subjects carrying a high number of point mutations in OXPHOS subunits would have a different gene expression profile from those with a low OXPHOS mutation burden. The transcriptional analysis was focused on the OXPHOS genes that showed the highest number of point mutations in our study, including *MT-ND1*, *MT-ND4*, *MT-CYB*, *MT-CO1*, and *MT-ATP6*; this approach covered all of the structural mitochondrial OXPHOS subunits (complexes I, III, IV, and V). Significant changes in mRNA abundance in livers were observed in response to the accumulated number of polymorphic sites in *MT-CYB*, *MT-CO1*, and *MT-ATP6* (Figure 2A–C), suggesting that the more polymorphic the locus, the lower its transcriptional efficiency. Likewise, *MT-CYB*-mRNA levels were significantly lower in subjects carrying the m.15326 and m.14766 polymorphic sites, and *MT-ND4* mRNA levels were significantly lower in patients carrying the m.11719 mutation (Figure 2D).

Furthermore, we observed that the accumulation of polymorphic sites in all of the above-mentioned members of the OXPHOS group was paralleled by the emergence of an 'OXPHOS-deficient phenotype', as the individual transcriptional activity of each locus was downregulated in response to the overall cumulative number of mutations (Figure 2E).

Finally, we observed a significant association between histological features associated with disease severity and the overall cumulative number of mutations in OXPHOS genes (steatosis score Spearman's rank correlation test R 0.40, $p = 0.02$; hepatocellular ballooning R 0.38, $p = 0.034$; lobular inflammation R 0.56, $p = 0.001$; and fibrosis R 0.38, $p = 0.031$).

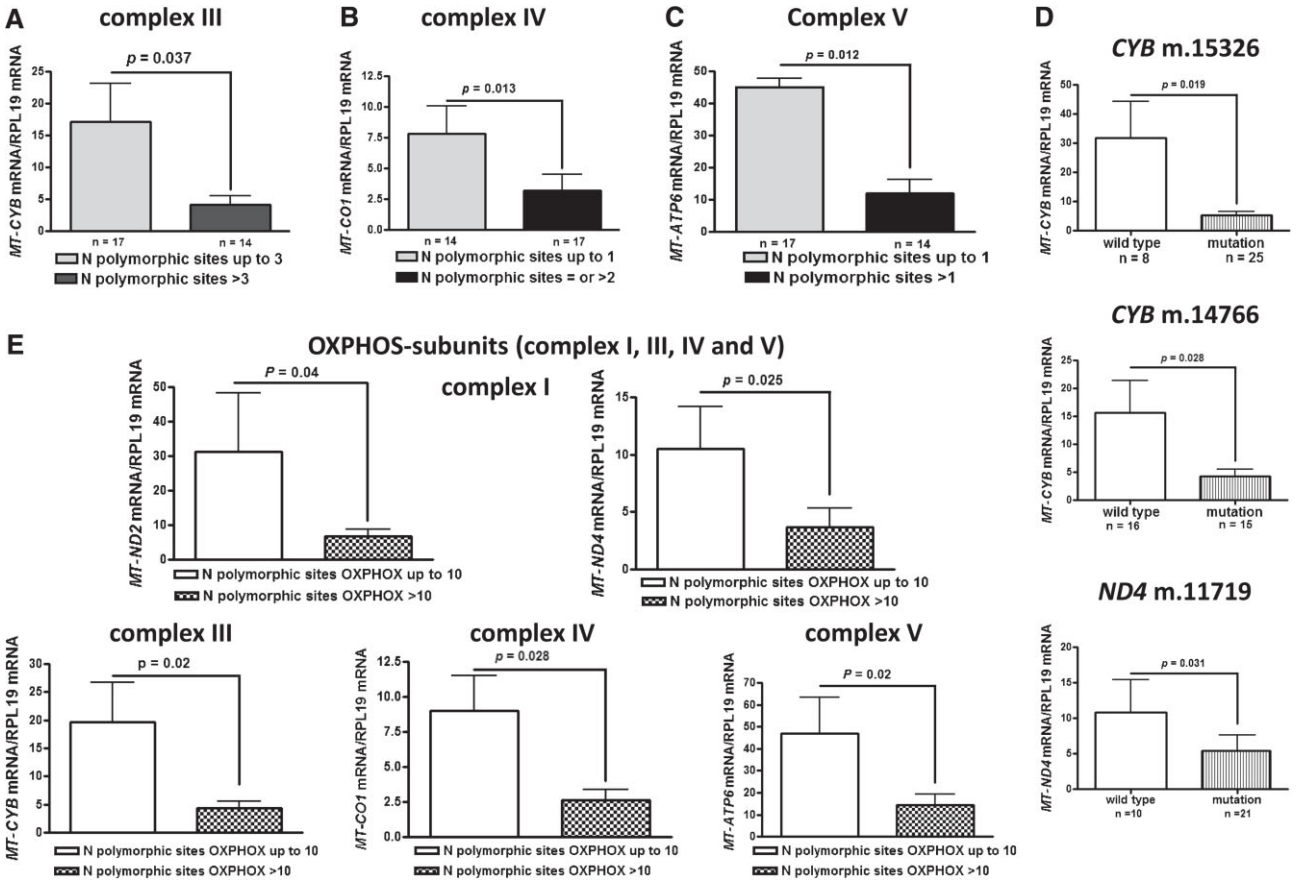

Figure 2. Phenotypic impact of OXPHOS variation: mtDNA variability explains variation in gene expression levels. We assessed changes in mRNA levels for those OXPHOS genes for which the number of polymorphic sites was highest in our population; mRNA abundance was assessed in the 31 patients for whom liver tissue mRNA was available by the use of RT-qPCR, and normalized to the amount of *RPL19* (used as a reference transcript). (A) Relative abundance of *MT-CYB* classified by the total number of polymorphic sites in the locus; the mean of *CYB* polymorphic sites was 3.26 ± 2.19, and the median was 3. A median value of 3 was set as the cutoff for classifying patients into two groups. (B) Relative abundance of *MT-CO1* classified by the total number of polymorphic sites in the locus; the mean of *CO1* polymorphic sites was 1.53 ± 0.99, and the median was 2. A median value of 2 was set as the cutoff for classifying patients into two groups. (C) Relative abundance of *MT-ATP6* classified by the total number of polymorphic sites in the locus; the mean of *ATP6* polymorphic sites was 1.26 ± 0.96, and the median was 1. A median value of 1 was set as the cutoff for classifying patients into two groups. (D) Relative abundances of *MT-CYB* and *MT-ND4* classified according to the presence of m.15326, m.14766 and m.11719 mutations. (E) Relative abundances of *MT-ND2*, *MT-ND4*, *MT-CYB*, *MT-CO1* and *MT-ATP6* mRNA classified according to the accumulated number of polymorphic sites; the mean of overall accumulated polymorphic sites was 9.61 ± 4.81, and the median was 10. A median value of 10 was set as the cutoff for classifying patients into two groups (up to 10, $n = 14$; and >10, $n = 17$). Statistical differences between groups were evaluated with the non-parametric Mann–Whitney test.

The replication study confirms the association of mtDNA genetic diversity, particularly in *MT-CYB* coding for cytochrome *b* in complex III, with NAFLD and disease severity

We performed a replication study aimed at validating the role of the missense p.Thr7Ile (rs527236041 C > T) variant located at position m.14766 of the *MT-CYB* locus that showed a significant association with disease severity. The features of subjects included in the replication study are shown in supplementary material, Table S9. The selected variant showed the strongest functional-predicted evidence and had significant phenotypic impact on liver *MT-CYB* gene expression (Figure 2). We observed that the m.14766 variant was significantly associated not only with NAFLD as a disease trait (healthy subjects versus NAFLD, $p = 0.0001$) even after adjustment for age, sex, body mass index, and homeostatic model assessment (odds ratio 3.04, 95% confidence interval 1.36–6.79, $p = 0.0063$), but also with disease severity (NAFL versus NASH, $p = 0.023$, χ^2 with Yates' correction).

We also aimed to validate the finding that the cumulative frequency of mtDNA variants in coding regions, including *MT-CYB*, is associated with disease phenotype. Sequencing of the entire *MT-CYB* gene was then performed with the Sanger method in a subsample of NAFLD patients ($n = 53$) and healthy subjects ($n = 23$) to explore the overall variability in this locus; patients and controls were randomly selected from the replication cohort, depending on the amount of DNA available. Sequence analysis revealed that patients with NAFLD harboured not only a significantly higher number of total accumulated variants (207 versus 70 polymorphic sites, respectively, $p = 0.0044$)

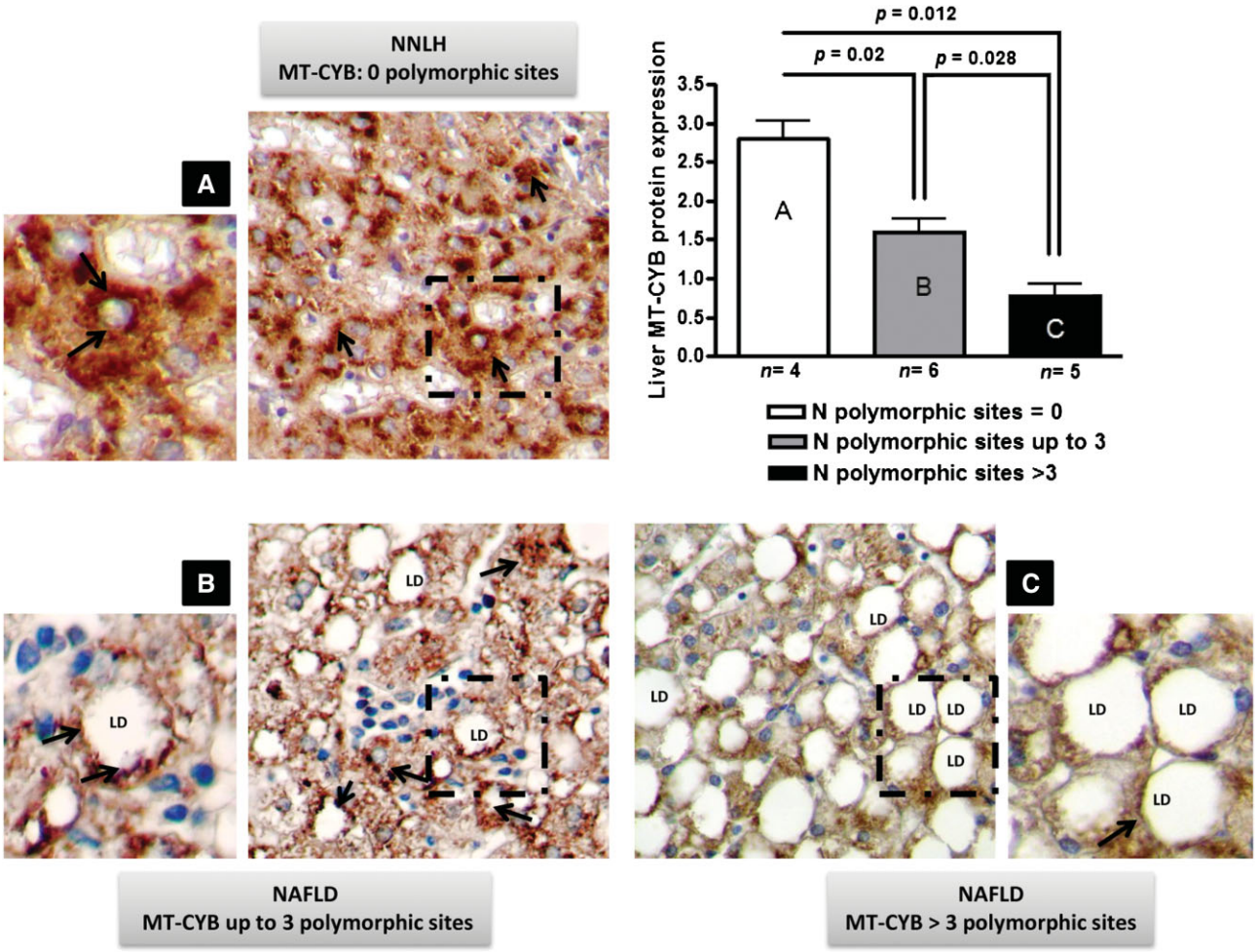

Figure 3. Phenotypic expression at the protein level of the accumulated number of polymorphic sites in mitochondrially encoded cytochrome *b* (*MT-CYB*). The figure shows functional assessment of the impact of variants/mutations in *MT-CYB* on NAFLD pathobiology by the use of immunohistochemistry with haematoxylin counterstaining. (A) Representative expression pattern of MT-CYB in control liver [near-normal liver histology (NNLH)] showing strong staining. (B, C) Representative expression pattern of MT-CYB in livers of patients with NAFLD according to the accumulated number of polymorphic sites. LD, lipid droplet. The arrow indicates MT-CYB immunostaining. Original magnification: ×400. The bar chart shows MT-CYB protein expression in liver specimens of subjects with NNLH having no polymorphic sites (bar A), and patients with NAFLD classified according to the total number of polymorphic sites in the locus (bar B and bar C). Statistical differences between groups were evaluated with the Mann–Whitney *U*-test.

but also a significantly higher proportion of heteroplasmic sites (25 versus 2, respectively, $p = 0.0013$) than controls; for details, see supplementary material, Figure S3.

To assess the functional impact of mtDNA mutations, we explored whether mutations in the *MT-CYB* locus had an impact on protein expression levels. The abundance of liver MT-CYB protein was significantly associated with the overall cumulative number of mutations in the locus, meaning that the higher the number of polymorphic sites, the lower the expression level of the protein (Figure 3). In addition, we observed that MT-CYB immunoreactivity was significantly decreased in the liver tissue of patients with NAFLD ($p = 0.0024$) (Figure 3); MT-CYB abundance was significantly and inversely correlated with the score of hepatocyte ballooning (Spearman's rank correlation test R −0.78, $p = 0.001$). Accordingly, we observed that liver *MT-CYB* gene expression levels were significantly associated with the histological features (supplementary material, Figure S4).

NAFLD is associated with a germline-derived mtDNA mutational profile

To explain whether NAFLD is associated with a tissue-specific mutational burden, we also sequenced the entire mtDNA from the liver tissue specimens of NAFLD patients ($n = 9$) and their corresponding peripheral blood samples ($n = 9$). The comparison of the overall number of variants per base position in coding regions between intra-patient liver and peripheral blood mitochondrial genomes showed a high degree of correlation, with an average of 98%. In fact, seven of nine paired samples shared 100% similarity, whereas the remaining two were 87% and 95% similar (Figure 4). Although the overall mutational burden of liver mitochondrial genomes did not differ from that of the corresponding blood samples, we did observe the emergence of

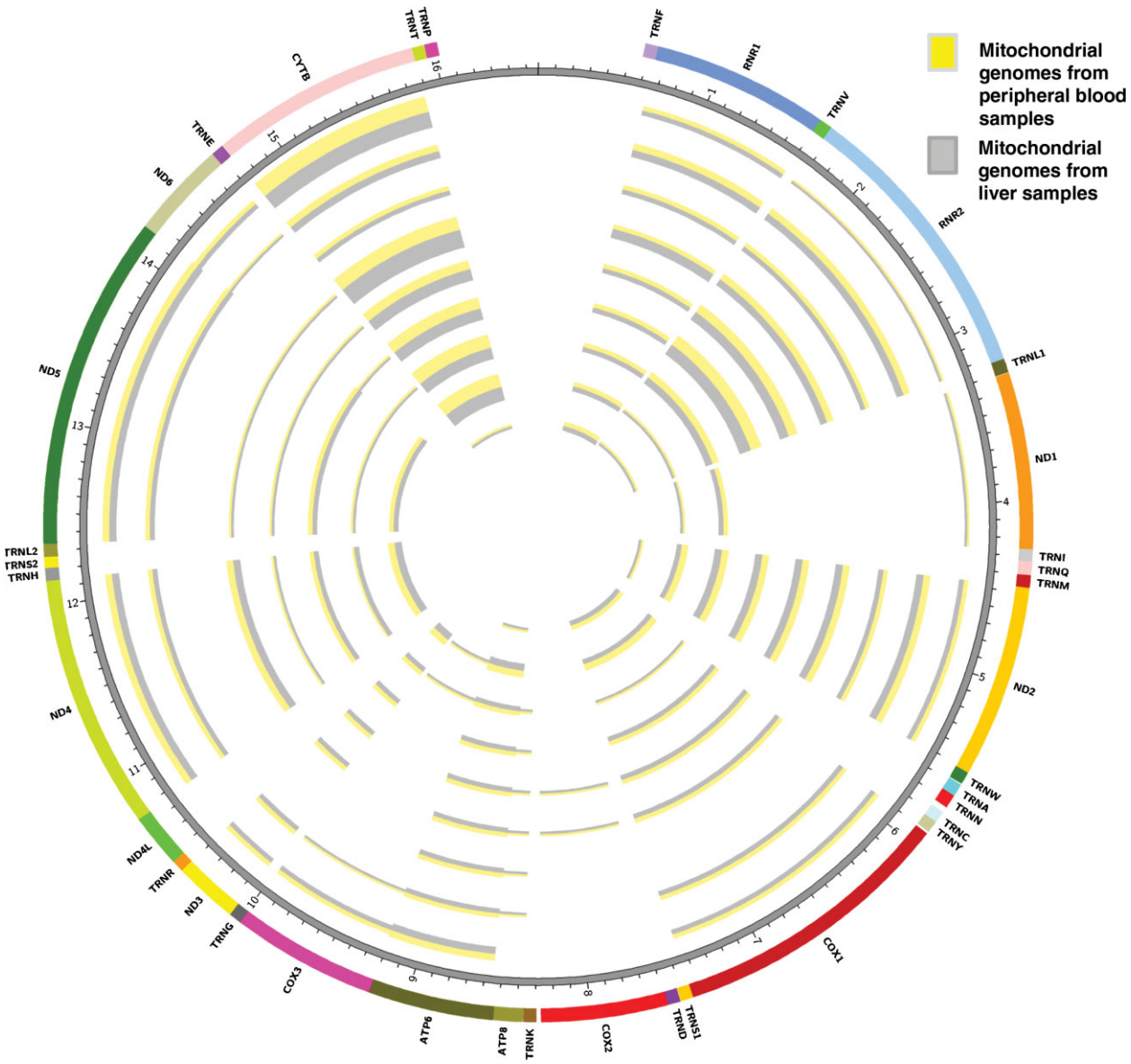

Figure 4. NAFLD is associated with a germline-derived mtDNA mutational profile. A concentric Circos plot (generated with Circos version 0.67) of the comparative analysis between liver and peripheral blood mitochondrial genomes in patients with NAFLD; on average, 98% of mtDNA variants observed in the liver of a given individual were shared with their corresponding blood sample. From the outside to the inner centre, each concentric circle represents a patient, and a total of nine pairs are depicted; variants in liver mitochondrial genomes are in grey, and variants in the peripheral blood samples are in yellow. The size of the segments corresponds to the overall accumulative number of variants per mitochondrial region; the segments do not depict specific variants per coding region. Numbers in the outside grey concentric circle indicate kb of the mtDNA. *MT-ATP6*, mitochondrially encoded ATP synthase 6; *MT-ATP8*, mitochondrially encoded ATP synthase 8; *MT-CO1-3*, mitochondrially encoded cytochrome *c* oxidase I-3; *MT-CYB*, mitochondrially encoded cytochrome *b*, *MT-ND1-6*, mitochondrially encoded NADH dehydrogenase 1–6; *RNR1*, mitochondrially encoded 12S ribosomal RNA; *RNR2*, mitochondrially encoded 16S ribosomal RNA. TRN stands for mitochondrially encoded tRNA transfers, with the final letter referring to a universal single-letter amino acid code. Variants were defined as positions where the primary base differed from the reference rCRS, and were classified by location across the mitochondrial genome.

probable *de novo* mutations in the liver of genes of the OXPHOS group, most specifically in *MT-CYB* (m.15110 G > A, read coverage = 1962; m.15237 T > C, read coverage = 1934; and m.15314 G > A, read coverage = 587), which were observed in two patients with NASH and advanced fibrosis. The observed changes correspond to missense mutations, two of which (m.15237 and m.15314) showed highly predicted scores for a deleterious functional impact (Polyphen: 0.47 and 0.6, respectively). Furthermore, two mutations *de novo* in liver 12S RNA (m.1339 G > A and m.1373 T > C) were observed in one patient with advanced fibrosis.

Whereas the total number of variants per base position in coding regions was almost the same in the liver and in the corresponding blood samples, variants at higher degrees of heteroplasmy (>0.1) were present only in the liver (three of nine NAFLD patients), except for the m.9101 T > C located in *MT-ATP6*. Interestingly, the highly polymorphic variants were located in genes of the OXPHOS group, including *MT-ATP6* (m.9101 T > C),

Table 3. Clinical and biochemical features of patients with non-alcoholic fatty liver disease (NAFLD) and controls included in the exploration of *POLG* and *POLG2* genetic variability

Variable	Healthy subjects	NAFL	NASH	*p*-value*	*p*-value†	*p*-value‡
Number of subjects	32	32	32	–	–	–
Female/male (%)	50/50	64/36	60/40	NS	NS	NS
Age (years)	48 ± 13	52 ± 9.7	51 ± 11	NS	NS	NS
Physical activity (h/week)	1.9 ± 2.5	3.3 ± 0.9	0.9 ± 1.9	NS	NS	NS
BMI (kg/m^2)	23.0 ± 2.5	32.0 ± 5	36.0 ± 5.6	<0.000001	<0.002	<0.000001
Fasting plasma glucose (mmol/l)	4.50 ± 0.39	5.33 ± 1.05	7.05 ± 2.72	<0.0002	<0.001	<0.000001
Fasting plasma insulin (pmol/l)	38.2 ± 17.4	83.3 ± 45.1	145.8 ± 97.2	<0.000001	<0.0005	<0.000001
HOMA-IR index	1.1 ± 0.5	3.0 ± 2.0	6.2 ± 4.6	<0.000001	<0.00007	<0.000001
SABP (mmHg)	117 ± 10	125 ± 12	133 ± 16	<0.03	NS	<0.001
DABP (mmHg)	73 ± 9	79 ± 8	79 ± 15	<0.04	NS	<0.01
Total cholesterol (mmol/l)	5.66 ± 1.22	5.43 ± 1.53	5.38 ± 1.19	NS	NS	NS
HDL cholesterol (mmol/l)	1.47 ± 0.39	1.32 ± 0.75	1.22 ± 0.34	NS	NS	<0.09
LDL cholesterol (mmol/l)	3.75 ± 0.93	3.28 ± 1.50	3.13 ± 1.09	NS	NS	NS
Triglycerides (mmol/l)	1.03 ± 0.33	1.71 ± 3.95	2.51 ± 1.60	<0.01	<0.07	<0.007
Uric acid (mmol/l)	196 ± 59	280 ± 167	309 ± 149	<0.09	NS	<0.006
ALT (U/l)	21 ± 5	68 ± 99	58 ± 35	<0.008	NS	<0.0002
AST (U/l)	22 ± 4	38 ± 19	49 ± 33	<0.001	<0.04	<0.0002
GGT (U/l)	34 ± 24	51 ± 31	87 ± 82	NS	<0.05	<0.01
AP (U/l)	159 ± 62	222 ± 89	269 ± 92	<0.01	<0.05	<0.008
CK18 fragment	–	235 ± 200	355 ± 284	–	<0.06	–
Histological features						
Degree of steatosis (%)	–	48 ± 23	60 ± 20	–	<0.04	–
Lobular inflammation (0–3)	–	0.5 ± 0.5	1.4 ± 0.5	–	<0.0006	–
Portal inflammation (0–2)	–	0.0 ± 0.0	1.7 ± 0.7	–	<0.000001	–
Hepatocellular ballooning (0–2)	–	0.0 ± 0.0	0.7 ± 0.6	–	<0.005	
Fibrosis stage (F0–F4)	–	0.1 ± 0.5	2.0 ± 1.2	–	<0.000001	
NAS	–	2.4 ± 1	6.4 ± 1.6	–	<0.00002	

ALT, serum alanine aminotransferase; AP, alkaline phosphatase; AST, serum aspartate aminotransferase; BMI, body mass index; CK, cytokeratin; DABP, diastolic arterial blood pressure; GGT, γ-glutamyltransferase; HDL, high-density lipoprotein; HOMA-IR, homeostatic model assessment insulin resistance; LDL, low-density lipoprotein; NAFL, non-alcoholic fatty liver or simple steatosis; NAS, NAFLD activity score; NASH, non-alcoholic steatohepatitis; NS, not significant; SABP, systolic arterial blood pressure.

Results are expressed as mean ± standard deviation.

*Statistically significant difference ($p < 0.01$) for comparisons between NAFL and controls.

†NAFL versus NASH.

‡NASH versus controls.

The *p*-value was obtained with the Mann–Whitney *U*-test, with the exception of the female/male ratio, where the *p*-value was obtained with the chi-squared test.

MT-CYB (m.15314 G > A and m.15237 T > C), and *MT-ND4* (m.11040 T > C).

The observed spectrum of mtDNA variants is not related to oxidative damage

It was previously suggested that reactive oxygen species might be a possible driving force of mtDNA mutagenesis; however, this concept has been questioned [18]. To confirm our findings that NAFLD could be associated with a common germline-derived variability of the mitochondrial genomes, rather than somatic mutations acquired in response to oxidative stress, we measured circulating levels of 8-hydroxy-2′-deoxyguanosine (8-OHdG; one of the predominant forms of free radical-induced mtDNA oxidative damage) in NAFLD patients at different stages of disease severity (NAFL, $n = 10$; NASH, $n = 10$). Notably, serum 8-OHdG levels were not associated with either disease severity ($p = 0.12$) or fibrosis score ($p = 0.20$), and neither did they correlate with the mtDNA mutation rate ($p = 0.9$), suggesting that the observed spectrum of mtDNA variants is not explained by oxidative damage. We analysed the ratio of transition (purine-to-purine or pyrimidine-to-pyrimidine) to transversion (purine-to-pyrimidine or pyrimidine-to-purine) (ts/tv) mutations in the entire mtDNA sequences of patients (ts/tv ratio: NAFL, 9.33; NASH, 9.1) and controls (ts/tv ratio: 6.9), and we did not observe any significant difference. This result is of particular relevance, as oxidative damage yields transversions more frequently [18].

Variants in *POLG* and *POLG2* are associated with liver mtDNA copy number

POLG is essential for mtDNA replication and repair, and *POLG2* encodes the catalytic subunit of POLG. We applied an NGS approach to all of the exons and regulatory regions of the two polymerase loci. The study cohort comprised 96 subjects, including 64 patients with NAFLD and 32 healthy subjects (Table 3). Quantification of liver mtDNA copy number and NGS of *POLG*/*POLG2* was performed on all patients and controls, and we identified 11 single-nucleotide polymorphisms (SNPs) in the entire population, including eight variants in *POLG* and three in *POLG2* (supplementary material, Table S10); two of them were novel, and the others are already registered in the dsSNP database. The missense p.Gln1236His variant in *POLG*, which encodes a change from glutamine (Q) to histidine (H)

at position 1236 (p.Q1236H), was significantly associated with the liver mtDNA copy number. The variant p.Q1236H occurs at a low frequency [minor allele frequency (MAF) of 0.027 in our study; MAF of 0.03 in 1000 Genomes], and was observed in four patients and one control subject. Carriers of the minor A allele had a significantly higher liver mtDNA content than homozygous carriers for the ancestral C allele (127 ± 9 and 65 ± 27, respectively, $p = 0.01$); the MAF in our study was 0.027.

Discussion

In this study, we took advantage of high-throughput sequencing technologies to explore the role of the whole liver mitochondrial mutational landscape as a driver and sustainer of an aggressive NAFLD phenotype. We observed that the liver mtDNA of patients with NAFLD harbours complex mitochondrial genomes with a significantly higher mutational rate than found in controls. The amount of mutant liver mtDNA, in either a heteroplasmic or a homoplasmic state, was significantly higher in NAFLD patients than in controls. The increased pathogenic potential of mitochondrial heteroplasmy was previously demonstrated in experimental and clinical settings [19,20], and the importance of both homoplasmic and heteroplasmic variants is highly recognized, not only in canonical mitochondrial diseases, but also in common complex diseases of the adult [21].

The comparison of liver mtDNA genetic diversity among patients with NAFL and NASH showed that disease severity was associated with an increased number of variants in the NADH dehydrogenase complex. The analysis of entire liver mitochondrial genomes of patients with advanced fibrosis not only confirmed our initial findings, but also revealed that disease severity is paralleled by a higher overall mutation rate (1.4-fold change), in addition to the presence of liver mtDNA variants in members of the OXPHOS group, some of them with a predicted functional impact. Although it is known that deleterious mtDNA mutations are frequent events during cancer development [21], the presence of these variants in the livers of patients with advanced NASH is a novel finding of our present study. In addition, missense mutations may compromise the function of the organ's respiratory chain. In fact, owing to the particular nature of the mtDNA in showing little redundancy throughout its sequence, a single variant or a random combination of them in essential proteins of the mitochondrial respiratory chain certainly impacts on the mitochondrial phenotypic traits. In the case of NASH, the combination of the observed variants in members of the OXPHOS group may modulate the overall mitochondrial performance towards a mitochondrial-deficient phenotype. Accordingly, Koliaki *et al* recently showed that NAFL progresses to NASH because of a deficit in the compensatory upregulation of the hepatic mitochondrial respiration required to satisfy bioenergetic demands [22].

By integrating mutation data with gene expression analysis, we observed that accumulation of mutated/polymorphic sites in members of the OXPHOS subunits (complexes I, III, IV, and V) modulates the overall efficiency of the gene expression program associated with mitochondrial bioenergetics, leading to the emergence of an OXPHOS-deficient phenotype. This finding, although novel, is biologically plausible, and can be explained by the conspicuous nature of the mitochondrial genome, the transcription of which is initiated from a single and major promoter that produces a polycistronic precursor RNA [23].

In addition to the profiling of liver OXPHOS gene expression, we have also provided evidence that mtDNA variants impact on protein function. Specifically, we observed that protein levels of MT-CYB are less abundant in the livers of patients who show the highest numbers of accumulated polymorphic sites. This finding is of particular relevance, because protein expression is what ultimately dictates the phenotypic manifestation of any given mitochondrial pathogenic mutation [24].

By comparing the genetic variability of mitochondrial genomes of peripheral blood cells with those from the liver, we found that most of the changes observed in the liver seemed to be part of a 'whole body' common mitochondrial mutational spectrum, rather than somatic mutations, with the majority being homoplasmic. Although there was a near-perfect match of correlation (on average, 98%) between variants in the mitochondrial genomes of the blood and the liver, we observed the emergence of predicted damaging somatic mutations in *MT-CYB*; interestingly, these mutations were seen only in patients with advanced fibrosis. The identification of a high degree of similarity between liver and blood mitochondrial genomes suggests that most of the observed mutations could have been inherited from the mother, or have occurred during embryonic or the germ-cell development in previous tissue differentiation. Therefore, it is plausible to speculate that mtDNA diversity may explain part of the 'missing heritability' of NAFLD. Likewise, the presence of a probably 'inherited' mtDNA mutational profile may also explain the development of NAFLD at an early age, which, depending on the ability of the variants to pass the transmission bottleneck, will allow the expansion of the mutant alleles, thereby resulting in the 'NASH mitochondrial-deficient' phenotype at the lowest ages. Our observations, if extended to paediatric populations, would help to explain why NASH develops in children with a prevalence that parallels that in the adult population [25]. Furthermore, it would be interesting to explore the hypothesis that an inherited mtDNA mutator profile explains the distinctive histological picture of paediatric NASH, which is characterized by changes in liver zone 1, including portal inflammation and portal fibrosis [25]. It is of note that zone 1 hepatocytes are adapted to high oxidative activities, which include β-oxidation of fatty acids and gluconeogenesis; in fact, hepatocytes in zone 1 have more abundant mitochondria. On the other hand, ultrastructural analysis of liver biopsies of children with

NASH showed substantial mitochondrial abnormalities, including changes in the mitochondrial matrix [26].

In addition, we showed that oxidative damage may not play a determining role in the increased mtDNA variability observed in patients with NASH. This observation also supports the notion that NASH may be associated with an 'inherited' mtDNA mutational profile that is independent of the 'acquired' oxidative damage. Another observation from our work is that the common *POLG* missense p.Gln1236His variant can result in liver mtDNA depletion. This variant, which is located in the conserved C-terminal polymerase domain (exon 23), seems to affect the performance of the POLG catalytic unit, and Glu1236 is conserved among species [27].

We would like to highlight the strengths and limitations of our study. These include the use of NGS technology, which, as compared with Sanger sequencing, for example, ensures the identification of mtDNA variants with a minimal error rate in a putative diverse mitochondrial population. Nevertheless, although a coverage of approximately × 800 is suitable for detecting homoplastic variants, a potential limitation should be highlighted, because the number of reads may potentially limit the sensitivity for detecting heteroplasmic variants with low frequency, so an average coverage of approximately × 800 could be underpowered to detect somatic mutations.

A potential caveat could be the combination of very stringent quality filters that could have underestimated the real number of variants per mitochondrial region; nevertheless, the use of appropriate computational filters ensures differentiation between true mtDNA variants from artefacts in the data. In fact, reported variants were confirmed by a thorough one-by-one visual inspection. In addition, it might be argued that the sample size is modest; nevertheless, it is worth highlighting that the molecular explorations were carried out in liver tissue samples obtained by percutaneous liver biopsy. Moreover, because mitochondrial genetics are different from nuclear genetics (the polyploid mtDNA genome consists of multiple circular dsDNA molecules with a high mutation rate, 1–5 copies of mtDNA/mitochondrion, >1000 mitochondria/cell), conventional criteria for power calculations cannot be applied, and neither may any prior assumption about the distribution of mutations be made (in part explained by the fact that variants/mutations are related to mtDNA haplogroups). Nevertheless, the replication study validated the main results, and supported the concept that the cumulative number of variants in coding regions of the mitochondrial genome is a key factor influencing disease severity.

A drawback of our study is the absence of functional data to support the impact of the observed missense mutations. In contrast, we did explore the phenotypic impact of OXPHOS variation in the liver, which provided relevant information on gene and protein expression. Considering that the large majority of mutations are found in some but not all mitochondrial genomes, any *in vitro* exploration supporting a mechanistic role for the variant(s) in the disease would not necessarily recapitulate the impact *in vivo*, as the resultant *in vivo* phenotype is indeed the consequence of a mixture of mitochondrial genomes, including mitochondrial genes that are functionally disrupted by mutations, together with normally functioning ones.

In conclusion, to the best of our knowledge, this is the first human study with a comprehensive characterization of the complete liver mitochondrial genome architecture in NAFLD. We have provided new insights into the pathogenesis of NAFLD by showing that the accumulation of mutations in the mitochondrial genomes of patients with advanced disease may lead to the development of an OXPHOS-deficient phenotype; altogether, these observations support the hypothesis that NASH may be regarded as a 'mitochondrial-deficient' disease [28].

Author contributions statement

The authors contributed in the following way: SS: study concept and design, data acquisition, liver biopsies and collection of biological material, data analysis and interpretation, general study supervision, drafting of the manuscript, and securing funding; DF: DNA sequencing and sequence data analysis; RS: gene expression analysis, genotyping, and NGS; HD, CR: sequence data analysis; TFG: NGS; JSM: immunohistochemistry; GOC: liver biopsies and collection of biological samples; CJP: study concept and design, data acquisition, data analysis and interpretation, statistical analysis, drafting of the manuscript, general study and NGS supervision, and securing funding; SS, DF, RS, HD, CR and CJP belong to the National Scientific and Technical Research Council (CONICET).

Acknowledgements

This study was partially supported by grants PICT 2010-0441/PICT 2014–1816 (CJP) and PICT 2012-0159/PICT 2014–0432 (SS) (Agencia Nacional de Promoción Científica y Tecnológica).

References

1. Rinella ME. Nonalcoholic fatty liver disease: a systematic review. *JAMA* 2015; **313:** 2263–2273.
2. Brunt EM. Histopathology of non-alcoholic fatty liver disease. *Clin Liver Dis* 2009; **13:** 533–544.
3. Bhala N, Angulo P, van der Poorten D, *et al.* The natural history of nonalcoholic fatty liver disease with advanced fibrosis or cirrhosis: an international collaborative study. *Hepatology* 2011; **54:** 1208–1216.
4. Sookoian S, Pirola CJ. The genetic epidemiology of nonalcoholic fatty liver disease: toward a personalized medicine. *Clin Liver Dis* 2012; **16:** 467–485.
5. Begriche K, Massart J, Robin MA, *et al.* Mitochondrial adaptations and dysfunctions in nonalcoholic fatty liver disease. *Hepatology* 2013; **58:** 1497–1507.

6. Sookoian S, Pirola CJ. NAFLD. Metabolic make-up of NASH: from fat and sugar to amino acids. *Nat Rev Gastroenterol Hepatol* 2014; **11:** 205–207.
7. Cotter DG, Ercal B, Huang X, *et al.* Ketogenesis prevents diet-induced fatty liver injury and hyperglycemia. *J Clin Invest* 2014; **124:** 5175–5190.
8. Satapati S, Sunny NE, Kucejova B, *et al.* Elevated TCA cycle function in the pathology of diet-induced hepatic insulin resistance and fatty liver. *J Lipid Res* 2012; **53:** 1080–1092.
9. Befroy DE, Perry RJ, Jain N, *et al.* Direct assessment of hepatic mitochondrial oxidative and anaplerotic fluxes in humans using dynamic 13C magnetic resonance spectroscopy. *Nat Med* 2014; **20:** 98–102.
10. Sunny NE, Parks EJ, Browning JD, *et al.* Excessive hepatic mitochondrial TCA cycle and gluconeogenesis in humans with nonalcoholic fatty liver disease. *Cell Metab* 2011; **14:** 804–810.
11. Sookoian S, Castano GO, Scian R, *et al.* Serum aminotransferases in nonalcoholic fatty liver disease are a signature of liver metabolic perturbations at the amino acid and Krebs cycle level. *Am J Clin Nutr* 2016; **103:** 422–434.
12. Caldwell SH, de Freitas LA, Park SH, *et al.* Intramitochondrial crystalline inclusions in nonalcoholic steatohepatitis. *Hepatology* 2009; **49:** 1888–1895.
13. Garcia-Ruiz C, Baulies A, Mari M, *et al.* Mitochondrial dysfunction in non-alcoholic fatty liver disease and insulin resistance: cause or consequence? *Free Radic Res* 2013; **47:** 854–868.
14. Pirola CJ, Gianotti TF, Burgueno AL, *et al.* Epigenetic modification of liver mitochondrial DNA is associated with histological severity of nonalcoholic fatty liver disease. *Gut* 2013; **62:** 1356–1363.
15. Sanyal AJ, Campbell-Sargent C, Mirshahi F, *et al.* Nonalcoholic steatohepatitis: association of insulin resistance and mitochondrial abnormalities. *Gastroenterology* 2001; **120:** 1183–1192.
16. Sookoian S, Rosselli MS, Gemma C, *et al.* Epigenetic regulation of insulin resistance in nonalcoholic fatty liver disease: impact of liver methylation of the peroxisome proliferator-activated receptor gamma coactivator 1alpha promoter. *Hepatology* 2010; **52:** 1992–2000.
17. Rocher C, Taanman JW, Pierron D, *et al.* Influence of mitochondrial DNA level on cellular energy metabolism: implications for mitochondrial diseases. *J Bioenerg Biomembr* 2008; **40:** 59–67.
18. Kauppila JH, Stewart JB. Mitochondrial DNA: radically free of free-radical driven mutations. *Biochim Biophys Acta* **2015**; 1847: 1354–1361.
19. Wallace DC, Chalkia D. Mitochondrial DNA genetics and the heteroplasmy conundrum in evolution and disease. *Cold Spring Harb Perspect Med* 2013; **3:** a021220.
20. Ye K, Lu J, Ma F, *et al.* Extensive pathogenicity of mitochondrial heteroplasmy in healthy human individuals. *Proc Natl Acad Sci USA* 2014; **111:** 10654–10659.
21. Stewart JB, Chinnery PF. The dynamics of mitochondrial DNA heteroplasmy: implications for human health and disease. *Nat Rev Genet* 2015; **16:** 530–542.
22. Koliaki C, Szendroedi J, Kaul K, *et al.* Adaptation of hepatic mitochondrial function in humans with non-alcoholic fatty liver is lost in steatohepatitis. *Cell Metab* 2015; **21:** 739–746.
23. Taylor RW, Turnbull DM. Mitochondrial DNA mutations in human disease. *Nat Rev Genet* 2005; **6:** 389–402.
24. Tuppen HA, Blakely EL, Turnbull DM, *et al.* Mitochondrial DNA mutations and human disease. *Biochim Biophys Acta* **2010**; 1797: 113–128.
25. Ovchinsky N, Lavine JE. A critical appraisal of advances in pediatric nonalcoholic fatty liver disease. *Semin Liver Dis* 2012; **32:** 317–324.
26. Lotowska JM, Sobaniec-Lotowska ME, Bockowska SB, *et al.* Pediatric non-alcoholic steatohepatitis: the first report on the ultrastructure of hepatocyte mitochondria. *World J Gastroenterol* 2014; **20:** 4335–4340.
27. Luoma PT, Luo N, Loscher WN, *et al.* Functional defects due to spacer-region mutations of human mitochondrial DNA polymerase in a family with an ataxia-myopathy syndrome. *Hum Mol Genet* 2005; **14:** 1907–1920.
28. Sookoian S, Pirola CJ. DNA methylation and hepatic insulin resistance and steatosis. *Curr Opin Clin Nutr Metab Care* 2012; **15:** 350–356.
29. *Kleiner DE, Brunt EM, Van Natta M, *et al.* Design and validation of a histological scoring system for nonalcoholic fatty liver disease. *Hepatology* 2005; **41:** 1313–1321.
30. *Brunt EM, Kleiner DE, Wilson LA, *et al.* Nonalcoholic fatty liver disease (NAFLD) activity score and the histopathologic diagnosis in NAFLD: distinct clinicopathologic meanings. *Hepatology* 2011; **53:** 810–820.
31. *Pirola CJ, Scian R, Gianotti TF, *et al.* Epigenetic modifications in the biology of nonalcoholic fatty liver disease: the role of DNA hydroxymethylation and TET proteins. *Medicine (Baltimore)* 2015; **94:** e1480.
32. *Rutledge RG. Sigmoidal curve-fitting redefines quantitative real-time PCR with the prospective of developing automated high-throughput applications. *Nucleic Acids Res* 2004; **32**: e178.
33. *Gianotti TF, Sookoian S, Dieuzeide G, *et al.* A decreased mitochondrial DNA content is related to insulin resistance in adolescents. *Obesity (Silver Spring)* 2008; **16:** 1591–1595.
34. *Schlotter YM, Veenhof EZ, Brinkhof B, *et al.* A GeNorm algorithm-based selection of reference genes for quantitative real-time PCR in skin biopsies of healthy dogs and dogs with atopic dermatitis. *Vet Immunol Immunopathol* 2009; **129:** 115–118.
35. *Ruijter JM, Ramakers C, Hoogaars WM, *et al.* Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic Acids Res* 2009; **37:** e45.
36. *Bender A, Krishnan KJ, Morris CM, *et al.* High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease. *Nat Genet* 2006; **38:** 515–517.
37. *Seneca S, Vancampenhout K, Van Coster R, *et al.* Analysis of the whole mitochondrial genome: translation of the Ion Torrent Personal Genome Machine system to the diagnostic bench? *Eur J Hum Genet* 2015; **23:** 41–48.
38. *Sosa MX, Sivakumar IK, Maragh S, *et al.* Next-generation sequencing of human mitochondrial reference genomes uncovers high heteroplasmy frequency. *PLoS Comput Biol* 2012; **8:** e1002737.
39. *Sookoian S, Castano GO, Burgueno AL, *et al.* A nonsynonymous gene variant in the adiponutrin gene is associated with nonalcoholic fatty liver disease severity. *J Lipid Res* 2009; **50:** 2111–2116.
40. *Bleda M, Tarraga J, de Maria A, *et al.* CellBase, a comprehensive collection of RESTful web services for retrieving relevant biological information from heterogeneous sources. *Nucleic Acids Res* 2012; **40**: W609–W614.

*Cited only in supplementary material.

SUPPLEMENTARY MATERIAL ONLINE

Supplementary materials and methods

Supplementary figure legends

Figure S1. Flow chart of work undertaken.

Figure S2. Electron microscopy of the liver tissue of a patient who carries recurring probable non-random point mutations (m.4769 A>G *ND2*, m.7028 C>T *CO1* and m.14766 C>T *CYB*).

Figure S3. Replication study: the cumulative frequency of mtDNA variants in MT-CYB (cytochrome B in complex III) is associated with the disease phenotype.

Figure S4. Association between liver *MT-CYB* gene expression levels and NAFLD histological features.

Table S1. Primer sets for mtDNA copy number and mRNA expression.

Table S2. Primer sets for whole mtDNA amplification.

Table S3. Deep-coverage analysis.

Table S4. List of mtDNA positions in which variants of liver mtDNA were found.

Table S5. Heteroplasmic variants.

Table S6. mtDNA variants in tRNAs.

Table S7. Novel missense-mtDNA variants according to fibrosis status.

Table S8. Heteroplasmic variants according to fibrosis status.

Table S9. Study 3 (replication case-control study): Clinical and biochemical features of healthy subjects and patients with NAFLD.

Table S10. SNPs in *POLG* and *POLG2*.