medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Sex differences in the progression from NAFL to NASH: a functional meta-analysis of transcriptomic studies

José F. Català-Senent[1,2], Marta R. Hidalgo[1], Marina Berenguer[3,4,5,6], Gopanandan Parthasarathy[7], Harmeet Malhi[7], Pablo Malmierca-Merlo[1,8], María de la Iglesia-Vayá[9] and Francisco García-García[1,2]

Affiliations:

1. Bioinformatics and Biostatistics Unit, Principe Felipe Research Center, Valencia, Spain.
2. Spanish National Bioinformatics Institute, ELIXIR-Spain (INB, ELIXIR-ES).
3. Liver Transplantation and Hepatology Unit, Hospital Universitario y Politécnico La Fe, València, Spain.
4. Grupo de hepatología, Cirugía HBP y Trasplantes, Instituto de Investigación Sanitaria La Fe, València, Spain.
5. CIBERehd, Centro de Investigación Biomédica en Red en Enfermedades Hepáticas y Digestivas, Instituto de Salud Carlos III, Madrid, Spain.
6. Department of Medicine, Universitat de València, València, Spain.
7. Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA.
8. Atos Research Innovation (ARI), Spain.
9. Biomedical Imaging Unit FISABIO-CIPF. Fundación para el Fomento de la Investigación Sanitario y Biomédica de la Comunidad Valenciana. Valencia, Spain.

Keywords: Non-alcoholic Fatty Liver Disease; Sex Characteristics; Precision Medicine; Computational Biology; Transcriptome Profiling.

Corresponding authors:
mhidalgo@cipf.es (Marta R. Hidalgo), fgarcia@cipf.es (Francisco García-García)

Bioinformatics and Biostatistics Unit, Principe Felipe Research Center, Valencia, Spain.
C/Eduardo Primo Yúfera 3, 46012, Valencia, Spain
Tel. +34 963 289 680 (Ext. 2003)
Fax. +34 963 289 701

List of Abbreviations: BH, Benjamini & Hochberg p-value adjusting method; GEO, Gene Expression Omnibus; GO, Gene Ontology; GSA, Gene set analysis; KEGG, Kyoto Encyclopaedia of Genes and Genomes; LOR, Log-odds ratio; NAFLD, Non-alcoholic fatty liver disease; NAFL, Non-alcoholic steatosis; NASH, Non-alcoholic steatohepatitis; PCA, Principal component analysis; PPA, Pathway activity analysis; PRISMA, Preferred Reporting Items for Systematic Reviews and Meta-Analyses; RMA, Robust Multichip Average.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Financial support statement: This research was supported by the Principe Felipe Research Centre. This study was co-funded by European Regional Development Funds (FEDER) awarded within the Valencian Community 2014–2020, and ISCIII PT17/0009/0015 FEDER.

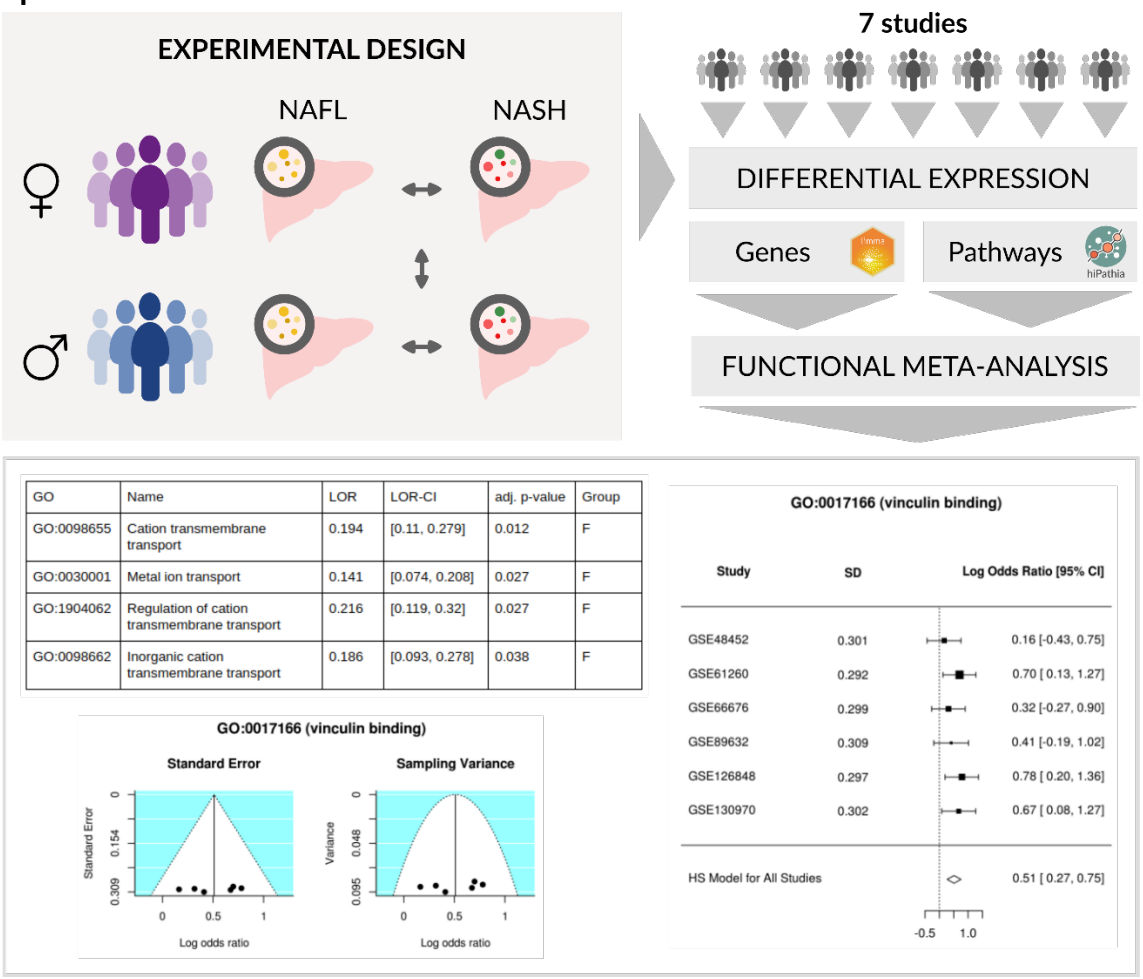
Conflict of interest statement: The authors who have taken part in this study declared that they do not have anything to disclose regarding funding or conflict of interest with respect to this manuscript.

Author contributions: JFC analysed the data; MRH and FGG designed and supervised the analysis; PMM designed and implemented the web tool; JFC, MRH, and FGG wrote the manuscript; MRH designed the graphical abstract; MB, GP, MIV and HM helped in the interpretation of the results; MRH and FGG conceived the work. All the authors revised the manuscript and added their suggestions.

Introductory Statement:

Progression from NAFL to NASH differently affects cellular functions in women and men. Here we systematically reviewed publicly available transcriptomic data and then performed a meta-analysis to find these affected functions. Thus, we identified 13 biological functions implicated in the progression of NAFLD that were differentially affected by sex.

Graphical abstract:



GO	Name	LOR	LOR-CI	adj. p-value	Group
GO:0098655	Cation transmembrane transport	0.194	[0.11, 0.279]	0.012	F
GO:0030001	Metal ion transport	0.141	[0.074, 0.208]	0.027	F
GO:1904062	Regulation of cation transmembrane transport	0.216	[0.119, 0.32]	0.027	F
GO:0098662	Inorganic cation transmembrane transport	0.186	[0.093, 0.278]	0.038	F

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Abstract

Background & aims:

Sex differences in non-alcoholic fatty liver disease (NAFLD) are well known and yet, most of the studies available in the literature do not include this factor when analysing the data. Here we present a functional meta-analysis of NAFLD studies to detect the molecular mechanisms involved in its progression, distinguishing any differences relative to patient sex.

Methods:

We systematically reviewed the Gene Expression Omnibus (GEO) database functional Gene Ontology (GO) terms detailed in transcriptomic studies, following the PRISMA statement guidelines. For each study, we compared the progression from steatosis (NAFL) to steatohepatitis (NASH) in premenopausal women and men using a dual strategy: a gene-set analysis and a pathway activity analysis. The functional results of all the studies were integrated in a meta-analysis.

Results:

A total of 114 abstracts were reviewed and 7 studies, which included 323 eligible patients, were finally analysed. The meta-analyses highlighted significant functions in both sexes. In premenopausal women, the overrepresented functions referred to DNA regulation, vinculin binding, IL-2 responses, negative regulation of neuronal death, and the transport of ions and cations. In men, they referred to the negative regulation of IL-6 and the establishment of planar polarity involved in neural tube closure.

Conclusions:

Our meta-analysis of this transcriptomic data provides a powerful approach to identify sex differences in the progression from NAFL to NASH. We detected relevant biological functions and molecular terms that were affected differently between premenopausal women and men. Changes in the immune responsiveness between men and women with NAFLD suggested that women have a more immune tolerant milieu while men have an impaired liver regenerative response.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Introduction

Non-alcoholic fatty liver disease (NAFLD) encompasses a spectrum of liver disorders ranging from fat accumulation in hepatocytes (NAFL) to non-alcoholic steatohepatitis (NASH), which can all potentially lead to cirrhosis or liver cancer. Interest in this disease has increased in recent years because of its worldwide impact on health, given that it is now the most common liver disease in developed countries (1). It is estimated that the worldwide prevalence of this disease is around 25%, although it is more common in South America and the Middle East and less prevalent in Africa (2).

Sex also affects the distribution of this pathology, with current studies showing a higher prevalence in men and postmenopausal rather than premenopausal women (3). Indeed, the liver shows the second largest amount of sexual dimorphism in humans. Both physiological and pathological hepatic processes, such as the detoxifying metabolism of cholesterol, as well as the prevalence of hepatic diseases differ between men and women (4). Hence, the progression of NAFLD and the effect of treatments are also likely affected by patient sex (5). Nonetheless, most studies on this disease do not consider metabolic differences between the sexes (6).

Although the presence of NAFLD can be strongly suspected based on imaging results, clinical analyses, and the presence of comorbidities, histological analysis is the only definitive way to diagnose it (7). Because of its invasive nature, the fact that a liver biopsy is the only completely reliable technique to diagnose NAFLD is a limitation when conducting studies in human patients.

The progression of the disease is different in each patient, and can have a considerable impact on the clinical consequences to the individual. Patients in the initial stage of the disease (NAFL) have a low risk of adverse outcomes (8), but its progression to NASH increases the possibility of clinical complications, both liver-related and otherwise.

In this context, we carried out a systematic review of the literature related to this topic and selected transcriptomic studies which had deposited data in the Gene Expression Omnibus (GEO) datasets database (https://www.ncbi.nlm.nih.gov/geo/). We used these to clarify the molecular basis of the progression from NAFL to NASH in premenopausal women and in men. Sex differences related to this progression may have important clinical implications to patients, and could potentially favour a personalised approach to this disease in the future.

Materials and methods

Search studies

This review was conducted in June 2020, following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement guidelines (9). We searched the GEO datasets database using the keywords: “NAFLD”, “NAFL”, “steatosis”, “NASH”, and

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

"steatohepatitis" for transcriptomic studies published in English.

Study exclusion criteria

We applied the following exclusion criteria: (i) studies conducted in organisms other than humans; (ii) studies without information about the sex of the participants or that had not included both sexes; (iii) studies without individuals from the NAFL and NASH stages of the disease; (iv) studies in which the disease had not been diagnosed with a biopsy. In the latter case, we decided to require the use of a biopsy as the method for NAFLD and NASH diagnosis because other less invasive methods such as conventional imaging techniques have a low sensitivity to detect mild steatosis and/or to differentiate NAFL from NASH (7).

Bioinformatics analysis strategy

We applied the same three-step transcriptomic analysis strategy to each separate study: (i) data acquisition and preprocessing; (ii) differential gene expression and functional enrichment analysis; (iii) differential pathway activation and functional analysis. The functional results of all the studies were then integrated using meta-analysis techniques (Figure 1). This work was carried out using the programming language R 3.6.0 (10); information about the packages we used and their version numbers are provided in the supplementary material.

Data acquisition and preprocessing

During the data preprocessing we standardised the nomenclature of the different stages of the disease (by grouping the patients into Control, NAFL, and NASH groups) as well as the association of the probe identifiers with their corresponding genes. In those studies, with treated and untreated patients, treated patients were eliminated for further analysis. For studies that had used microarrays, the probe codes were transformed to their respective Entrez identifiers from the NCBI database.

For repeated probes, we calculated the median of their expression values. When available, data normalised by the original authors of the studies were used. Otherwise (studies GSE61260 and GSE66676), the raw data was normalised using the Robust Multichip Average (RMA) algorithm (11). The gene nomenclature in RNA-Seq studies was also standardised to the Entrez identifiers and the raw data matrix was processed using the TMM standardisation method (12) followed by a log-transformation of the data.

To meet the objective of this work, women were separated into premenopausal and postmenopausal groups. Since the selected studies did not include this information, we assumed that women aged under 50 years had been premenopausal, based on previous literature indicating that the average age of menopause is around 48–52 years (13,14).

After the normalisation of the data we attempted to detect any possible anomalous effects within the studies by completing an exploratory clustering and a principal component analysis (PCA).

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

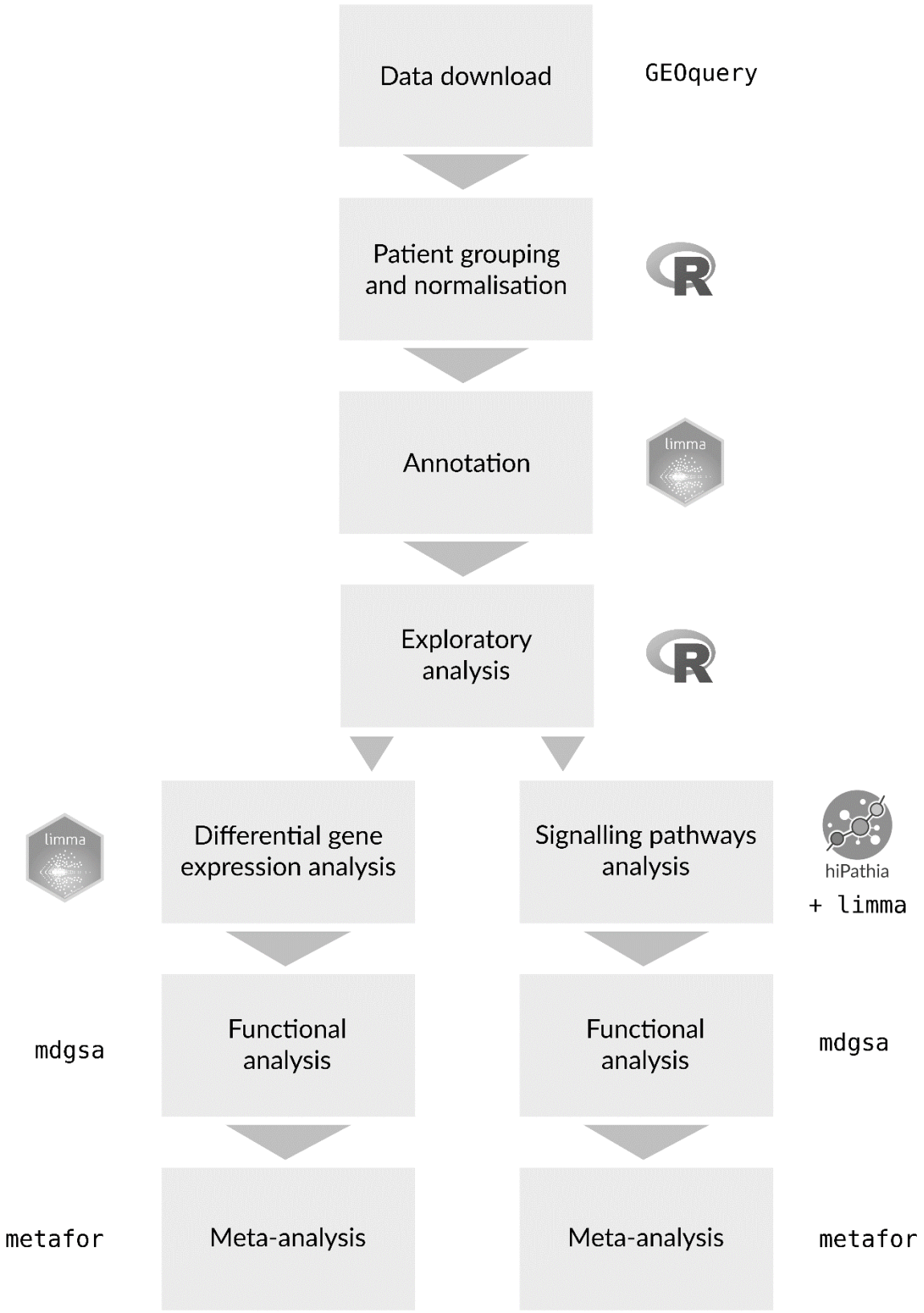


Fig. 1. Data-analysis workflow.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Differential gene expression and functional enrichment analysis

We carried out a differential expression analysis of these studies using the *limma* and *edgeR* packages (15,16) to detect genes that were differentially expressed during the progression of NAFL to NASH in premenopausal women versus men. For each gene, we adjusted a linear model, which included possible batch effects, contrasting (NASH.W – NAFL.W) and (NASH.M – NAFL.M), where NASH.W, NAFL.W, NASH.M, and NAFL.M corresponded to NASH-affected premenopausal women, NAFL-affected premenopausal women, NASH-affected men, and NAFL-affected men, respectively. The statistics of the differential expression were calculated and the *p*-values were adjusted using the Benjamini & Hochberg (BH) method (17).

Next, based on the differential gene expression results, we performed a functional enrichment analysis using gene set analysis (GSA) (18). First, the genes were ordered according to their *p*-value and the sign of the contrast statistic. Next, the GSA was performed using the logistic regression model implemented in the *mdgsa* R package (19) along with their corresponding functional annotations obtained from the Gene Ontology (GO) (20) and Kyoto Encyclopaedia of Genes and Genomes (KEGG) PATHWAY (21) databases.

Because of their hierarchical structure, the gene annotations with GO terms applied in the *mdgsa* package were propagated so that they inherited the annotations of the ancestor terms. Excessively specific or generic annotations (blocks smaller than 10 or larger than 500 words) were subsequently filtered out. Finally, functions with a BH-adjusted *p*-value under 0.05 were considered significant.

For the two functional elements (GO terms and KEGG paths), we analysed the number of over-represented elements shared by the studies. These results were graphically represented as UpSet plots (22) to show the number of elements in common between the different sets. First, we compared the overrepresented elements as separate graphs to detect the common functions in each group: GO terms in premenopausal women, GO terms in men, KEGG pathways in premenopausal women, and KEGG pathways in men. We then visualised any significant GO terms in each of the individual studies in the same plot to highlight significant terms with different signs from among the studies.

Differential pathway activation and functional analysis

The Hipathia algorithm (23) was used to perform Pathway Activity Analysis (PAA). This method transforms gene expression values for stimulus-response signalling sub-pathways into the activation levels which ultimately trigger cellular responses. In this current work we used the *hipathia* R package to analyse 1,654 GO terms and 142 UniProt functions (24), associated with 146 KEGG routes (21).

The activation signal of each subpathway was computed from gene expression values using this algorithm. These values were used to detect significant differential activations in the (NASH.W – NAFL.W) – (NASH.M – NAFL.M) contrast pair (as defined above). The functional annotations included in Hipathia and the differential activation results were

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

integrated using the GSA method to identify differences at the functional level (GO terms and UniProt functions) in relation to progression of the disease.

Meta-analysis

Once the gene functional analysis was applied to each individual study, we carried out a functional meta-analysis to summarise the results. Similarly, once the pathway functional analysis was applied to each individual study, the results were summarised via a homologous functional meta-analysis.

Both meta-analyses were performed following the methodology described in detail in García-García (25). Briefly, the *metafor* R package (26) was used to evaluate the combined effect with a random-effects model; this model more precisely detects overrepresented elements than performing the studies individually at a higher statistical power. Likewise, variability in the individual studies was incorporated into the global estimation of the measured effect so that less-variable results had a higher weight in the overall calculation of the logarithm of the odds ratio (LOR). Finally, we confirmed the suitability of the selected studies by evaluating the heterogeneity of these indicators.

For each meta-analysis and function, *p*-values (adjusted using the BH method), LORs, and 95% confidence intervals (CIs) were calculated. A term was considered significant when its adjusted *p*-value was lower than 0.05; a positive LOR in significant functions indicated a greater overrepresentation in premenopausal women than in men. In contrast, a negative LOR indicated higher representation in men than in premenopausal women. Finally, funnel plots and forest plots were used to assess the variability and to measure the contribution of each study to the meta-analysis.

A total of 8,223 elements were analysed in the gene functional meta-analysis (7,994 GO terms and 229 KEGG pathways); 1,654 GO terms and 142 UniProt functions (24) were assessed in the functional pathway meta-analysis in association with 146 KEGG pathways (21).

The large volume of data and results generated in this work is available in metafun-NAFLD web tool: https://bioinfo.cipf.es/metafun-NAFLD. Its access is free for any user and allows us to check the results shown in this manuscript, as well as to review other results that may be of interest to researchers. The front-end was developed using the Bootstrap library. All the graphics used in this tool have been implemented with Plot.ly except for the exploratory analysis cluster plot which was generated with ggplot2 package.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Results

Systematic review and exploratory analysis

As shown in the PRISMA flow diagram (Figure 2), our systematic review yielded 114 non-duplicated studies. After applying the exclusion criteria described above, a total of 7 studies were included in this work. Studies were excluded for the following reasons: studies not in humans or unrelated to NAFLD (*n* = 55), in vitro studies (*n* = 15), lack of information on sex or inclusion of only one sex (*n* = 24), no patients had the NAFL and NASH disease stages (*n* = 12), and diagnoses not based on histology (*n* = 1).

The 7 selected studies (27–33) included a total of 323 eligible patients (Table 1); 164 individuals were men and 159 women under 50 years (49.2% men and 50.8% respectively). By disease stage, 148 belonged to the NAFL group and 175 to NASH group (45.8 and 54.2%, respectively).

Of note, although the GSE83452 study contained patients at baseline and at one-year post-treatment, only baseline individuals were used in this current study. Table 1 shows the number of individuals from each study that were eligible in our work based on the criteria we established. Our exploratory analysis showed abnormal behaviour for the GSE83452 study; the PCA showed two different blocks: the first 64 samples were grouped into one cluster and the last 88 into another one. We considered this a batch effect and took it into account in our subsequent analyses.

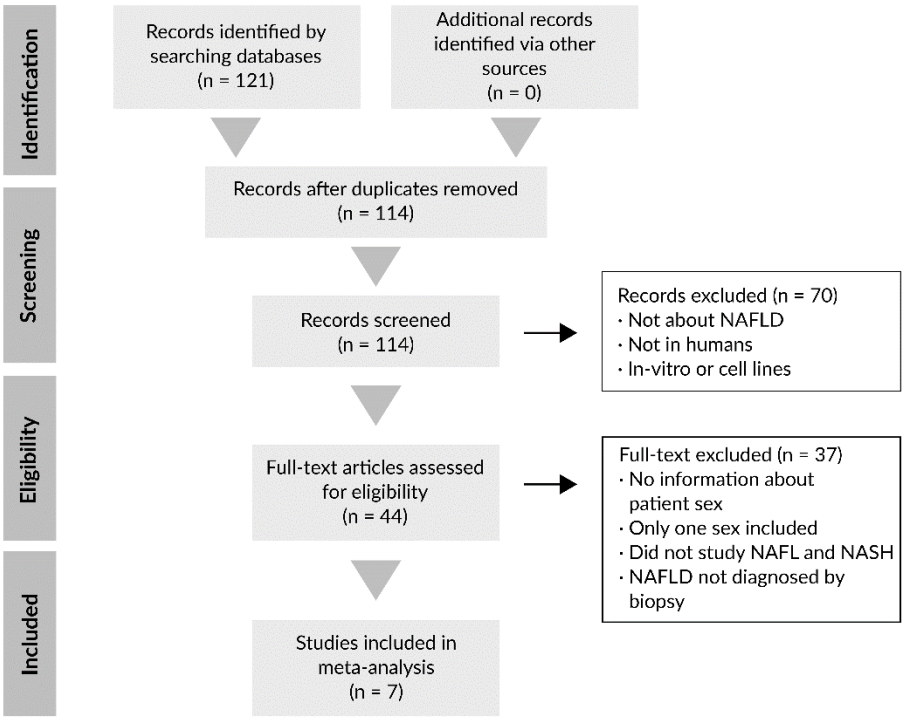


Fig. 2. Flow diagram of our systematic review of the literature and selection of studies for this meta-analysis, according to the PRISMA statement guidelines.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Table 1. Characteristics of the studies selected for inclusion in our analysis in terms of disease stage, sex, and patient age.

GEO accession, Authors, Year [Ref.]	**Patients by disease stage and sex**				**Age**	
	Premenopausal women		**Men**		**Mean**	***SD***
	NAFL	**NASH**	**NAFL**	**NASH**		
GSE48452 / Ahrens M. *et al.* 2013 [27]	10	9	4	4	41.62	9.35
GSE61260 / Horvath S. *et al.* 2014 [28]	10	7	12	12	41.51	9.09
GSE66676 / Xanthakos S.A. *et al.* 2015 [29]	20	4	6	3	17.14	1.66
GSE83452 / Lefebvre P. *et al.* 2017 [30]	32	35	6	46	39.61	11.51
GSE89632 / Arendt B.M. *et al*. 2015 [31]	2	4	14	9	39.55	8.19
GSE126848 / Suppli M.P. *et al.* 2019 [32]	6	4	9	12	40*	13*
GSE130970 / Hoang S.A. *et al.* 2019 [33]	4	12	13	14	44.72	9.80

*The information corresponding to the patients in this study was kindly provided by the author of the study, although individual patient ages were not available.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Individual analysis

The functional analysis of individual genes highlighted significant results overrepresented in both the groups of premenopausal women and of men, as summarised in Table 2. Analysis (using UpSet plots) of the terms or pathways that were common through the different studies showed that none of the elements were common to more than 4 studies (Supplementary Figure 1); in fact, the results for some elements were opposite in different studies (Figure 3A). The functional analysis of individual pathways only exhibited significant results for the GSE83452 study (20 sub-paths and 96 GO terms).

Table 2. Number of significant GO terms and KEGG pathways in each individual study after gene functional analysis. Positive and negative logarithm of the odds ratio (LORs) represent overrepresentation in premenopausal women and in men, respectively.

	GO terms		**KEGG pathways**	
	Positive LOR	**Negative LOR**	**Positive LOR**	**Negative LOR**
GSE48452	166	642	5	22
GSE61260	249	35	5	6
GSE66676	721	168	30	2
GSE83452	90	931	2	25
GSE89632	52	67	2	1
GSE126848	1168	150	55	5
GSE130970	1244	407	31	20

Meta-analysis

In addition to the *p*-value associated with each of the GO terms, the results of the meta-analysis measured the overall effect of that term throughout all the individual studies, i.e., the LOR. This effect indicated the magnitude of the overexpression of each term (the LOR value) and its direction (the LOR sign). Figure 3B shows, for one function, the LOR of each individual study and the joint LOR of the meta-analysis. The results of all significant elements are detailed in the Supplementary Figures 2 and 3. We confirmed the absence of bias and heterogeneity in all cases, as in Figure 3C (Supplementary Figures 4 and 5).

Gene functional analysis

The functional meta-analyses of the individual gene differential-expression pipeline results showed a total of 9 significant GO terms (Table 3 and Figure 4), although no KEGG

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

pathways were significant. Among the significant GO terms, 6 were overrepresented in premenopausal women; these elements encompassed a broad spectrum of biological processes and molecular functions, including *RNA polymerase II regulatory region DNA binding* (GO:0001012), *RNA polymerase II regulatory region sequence-specific DNA binding* (GO:0000977), *Vinculin binding* (GO:0017166), *Cellular response to interleukin-2* (GO:0071352), *Negative regulation of neuron death* (GO:1901215), and *Response to interleukin-2* (GO:0070669). The 3 remaining terms were overrepresented in men: *Negative regulation of interleukin-6 secretion* (GO:1900165), *Establishment of planar polarity involved in neural tube closure* (GO:0090177), and *Regulation of establishment of planar polarity involved in neural tube closure* (GO:0090178).

Pathway functional analysis

The functional meta-analysis of the individual results in the pathway differential-activation pipeline returned a total of 4 significant GO terms, all of them overrepresented in the progression of NAFLD in premenopausal women and with a LOR value lower than 0.25 (Table 3 and Figure 4). All these terms were related to the transport of ions and cations: *Cation transmembrane transport* (GO:0098655), *Metal ion transport* (GO:0030001), *Regulation of cation transmembrane transport* (GO:1904062), and *Inorganic cation transmembrane transport* (GO:0098662). The meta-analysis of the UniProt functions did not provide any significant results.

Metafun-NAFLD web tool

Metafun-NAFLD web tool (https://bioinfo.cipf.es/metafun-NAFLD) contains information related to 7 studies and 323 samples used in this work. For each study, the portal included fold-changes of genes and log odds ratios of functions and pathways, which can be explored by users to identify profiles of interest to them.

A total of 8,223 meta-analysis were carried out. For each of the 13 significant functions, metafun-NAFLD showed the global activation level by sex, for all studies and the specific contribution of each one of them, using statistical indicators (log odds ratio, confidence interval and p-value) and graphical representations by function, as forest and funnel plots. This open resource contributes to share data and results between researchers to improve other further works.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

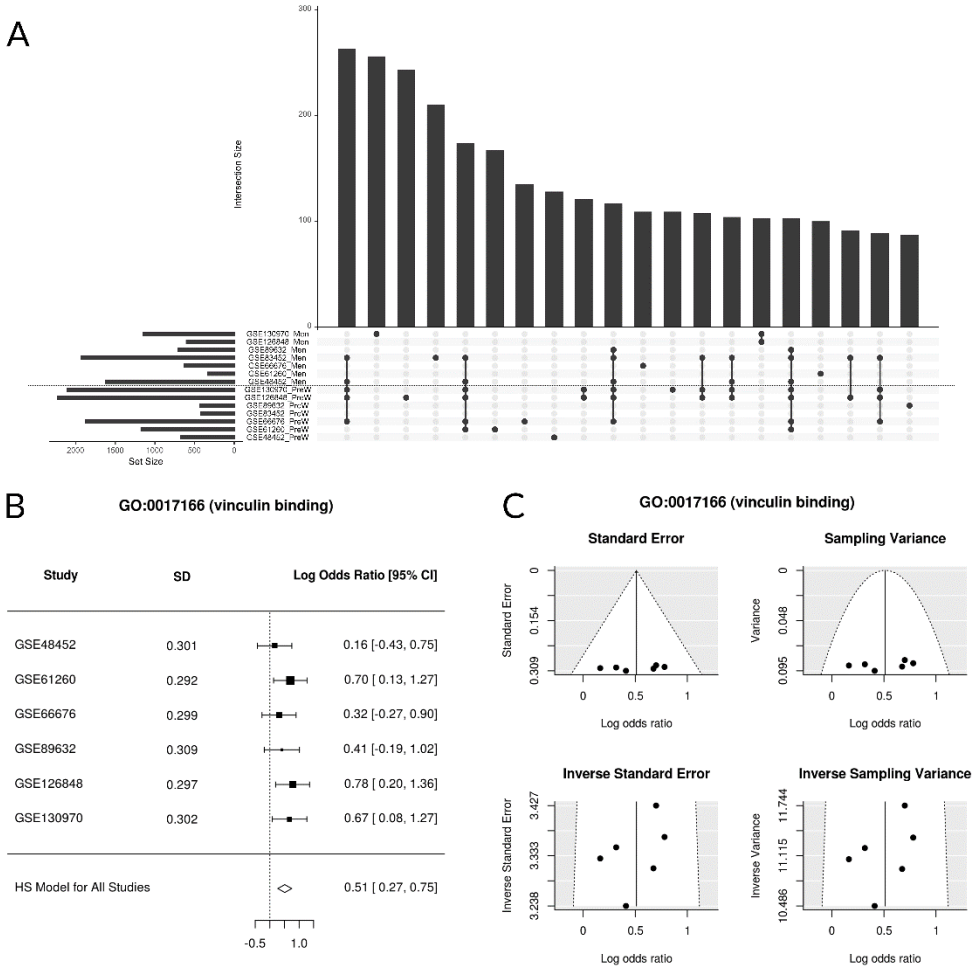


Fig. 3. **(A) UpSet plot showing the number of common elements among the significant GO terms in our functional gene analysis**. Only the 20 most abundant interactions are shown. Horizontal bars indicate the number of significant elements in each study; the two groups (premenopausal women and men) are separated by a dotted line. The vertical bars indicate the common elements in the sets, indicated with dots under each bar. The single points represent the number of unique elements in each group. Cases which cross the red line are significant GO terms, both in premenopausal women and men. **(B) A forest plot of the GO:0017166 term, showing the LOR of each study and the global result. (C) Funnel plot of the GO:0017166 term; dots in the white area indicate the absence of bias and heterogeneity.**

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

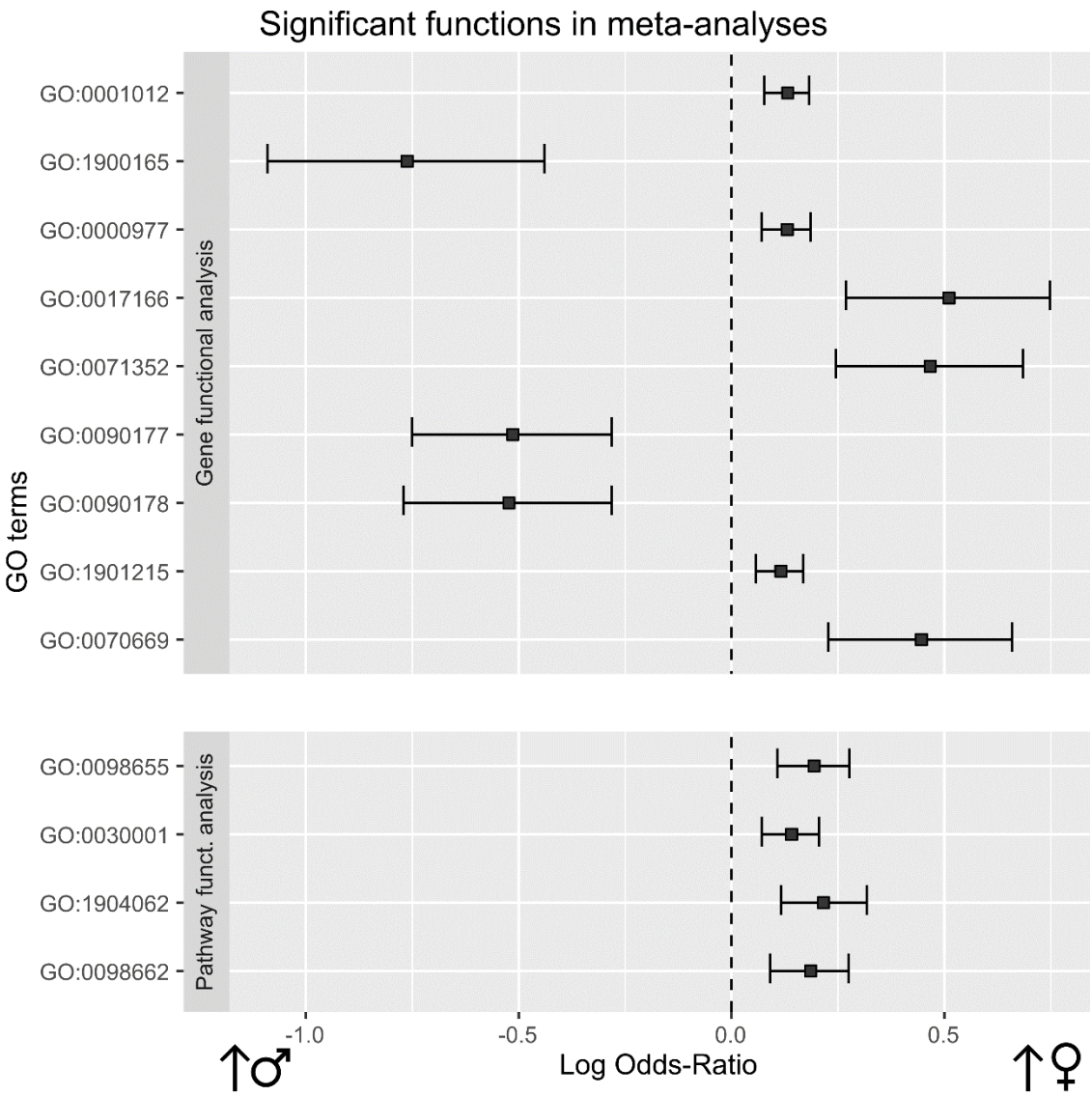


Fig. 4. **Differential functional profiling by sex**. This plot shows significant functional terms of each meta-analysis. On the right: biological functions more overrepresented in women. On the left: biological functions more overrepresented in men. For each function, LOR (red square) and its confidence interval have been represented.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Table 3. **Significant functions in the functional meta-analyses.** The logarithm of the odds ratio (LOR), its confidence interval (LOR-CI), the adjusted *p*-value, and the group in which the function was overexpressed are shown.
(W: women; M: men).

	GO ID	Name	LOR	LOR-CI	adj. *p*-value	Group
Gene functional meta-analysis	**GO:0001012**	RNA polymerase II regulatory region DNA binding	0.132	[0.079, 0.185]	0.008	W
	GO:1900165	Negative regulation of interleukin-6 secretion	-0.762	[-1.088, -0.437]	0.018	M
	GO:0000977	RNA polymerase II regulatory region sequence-specific DNA binding	0.131	[0.073, 0.188]	0.02	W
	GO:0017166	Vinculin binding	0.511	[0.271, 0.75]	0.035	W
	GO:0071352	Cellular response to interleukin-2	0.467	[0.248, 0.687]	0.035	W
	GO:0090177	Establishment of planar polarity involved in neural tube closure	-0.514	[-0.748, -0.279]	0.035	M
	GO:0090178	Regulation of establishment of planar polarity involved in neural tube closure	-0.523	[-0.768, -0.279]	0.035	M
	GO:1901215	Negative regulation of neuron death	0.116	[0.06, 0.171]	0.043	W
	GO:0070669	Response to interleukin-2	0.446	[0.23, 0.662]	0.046	W
Pathway functional meta-analysis	**GO:0098655**	Cation transmembrane transport	0.194	[0.11, 0.279]	0.012	W
	GO:0030001	Metal ion transport	0.141	[0.074, 0.208]	0.027	W
	GO:1904062	Regulation of cation transmembrane transport	0.216	[0.119, 0.32]	0.027	W
	GO:0098662	Inorganic cation transmembrane transport	0.186	[0.093, 0.278]	0.038	W

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

Discussion

Although previous studies have analysed pathobiological mechanisms in NAFLD, and separately sexual differences in metabolic regulation, the impact of gender (sociocultural factors) and sex (biological factors) on NAFLD pathobiology remains incompletely defined. Epidemiological studies have shown a higher prevalence of NAFLD in men and postmenopausal women compared to premenopausal women (3). To better understand the molecular basis of these sex differences, here we present a systematic review and functional meta-analysis of publicly available genomic studies, to analyse differences in NAFLD disease progression between premenopausal women and men.

The individual results of the 7 selected studies revealed that no GO or KEGG terms were consistently different between groups across all studies, perhaps related to smaller sample sizes, underlying genetic and environmental variance in the source population, and methodological heterogeneity. Also, as shown in Figure 3A, surprisingly, some functions were overexpressed in both groups, depending on the study. Therefore, a meta-analysis, such as this one, can help improve identification of relevant and consistent differences across studies, subject to the caveat that patients in each study may have other differences in clinical characteristics. Also, a meta-analysis can increase the overall sample size and hence the relevance of findings in a disease known to be heterogeneous.

Our functional meta-analysis identified a total of 13 GO terms that were different between men and premenopausal women using two different approaches– differentially expressed genes (9 terms) and differentially activated pathways (4 terms). These terms were related to processes such as inflammation, cell binding, and the establishment of polarity during neural tube closure.

Functional meta-analysis of differentially expressed genes identified differences in 9 significant GO terms. Among these, our results showed the overrepresentation of *Vinculin binding term* (GO:0017166) in premenopausal women compared to men. This cytoplasmic actin-binding protein is enriched in focal adhesions and adherens junctions, contributes to junction stability (34), and plays an important role as a regulator of apoptosis (35). Of note, although cellular adhesion is a broad topic and not a specific biologic pathway, it was reported to be a GO term consistently enriched in male NASH (36) and Arendt et al. also identified genes related to cell-cell adhesion (*ANXA2* and *MAG*) that were differentially expressed between patients with NAFL and NASH (31).

Our analysis also identified differences between the sexes related to the inflammatory response, specifically the expression of cytokine-related terms. *Negative regulation of interleukin-6 secretion* (GO:1900165) was overrepresented in men, while *Cellular response to interleukin-2* (GO:0071352) and *Response to interleukin-2* (GO:0070669) were overrepresented in premenopausal women. These terms related to interleukin 2 (IL-2) and interleukin 6 (IL-6), had LOR values close to or above 0.5, indicating a strong difference between the groups. IL-2 induced T cell responses are regulated by the level of IL-2, such that low level IL-2 regulates T-cell central tolerance via the formation of regulatory T cells (37). Higher levels of IL-2 induce proliferation of T cells and aid differentiation into memory

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

T-cells. IL-2 receptors are also expressed by B-cells and cells of the innate immune system (38). Interestingly, IL-2 signalling is deficient in many chronic liver diseases (39) and levels of soluble IL-2 receptor alpha were higher in children with NASH and advanced fibrosis (40). These data suggest that low-level IL-2 induced immune tolerance may mitigate the risk of developing NASH in premenopausal women and raise the intriguing possibility of restoring low-level IL-2 signalling in men as a potential treatment of NAFLD.

IL-6 has been previously implicated in NASH with an increased hepatic expression correlating with disease severity as well as insulin resistance (41). It is also a regulator of hepatic regeneration (42) and a major driver of hepatic carcinogenesis (43). In NAFLD, IL-6 may enhance hepatic repair and regeneration, however, it might also be involved in inducing insulin resistance and stimulating hepatocyte apoptosis, thus contributing to NASH progression (42,44). Indeed IL-6 gene polymorphisms have been associated with an increased susceptibility to chronic liver disease (45).

Thus, differences in immune mechanisms between men and women may be one possible factor in for the sex-differences observed in NAFLD. For example, in mice, the macrophage phenotype in males is pro-inflammatory and pro-fibrotic, compared to pro-resolution and anti-fibrotic in women (46). To speculate, women may have a relatively decreased immune response to lipotoxicity, whereas, men may have an impaired liver-healing response to chronic injury perhaps related to deficient IL-6 signalling. These ideas remain unsubstantiated and will require further examination and validation in larger studies.

A recent study by Herrera-Marcos et al. detected sex differences in the control of the *Cidec/Fsp27β* gene at the level of the liver (47). In fact, expression of the mRNA of this gene was associated with the presence and density of liver lipid droplets in mice fed a high fat diet. Among other functions, *Cidec/Fsp27β* is suggested to be activator for the neuronal pathway in the steatoic liver (48). Therefore, the fact that the result for the function *Negative regulation of neuron death* (GO:1901215) was significant in our study could be related to the differences detected by Herrera-Marcos et al.

Interestingly, we found other significant GO terms that had not been associated with NAFLD in previous studies. For instance, *Establishment of planar polarity involved in neural tube closure* (GO:0090177) and *Regulation of establishment of planar polarity involved in neural tube closure* (GO:0090178) were overrepresented in men with a LOR exceeding 0.5 (−0,514 and −0,523). However, we did not find a relationship between these terms and NAFLD.

Our functional pathway meta-analysis identified 4 significant GO terms involved in cation transport: *Cation transmembrane transport* (GO:0098655), *Metal ion transport* (GO:0030001), *Regulation of cation transmembrane transport* (GO:1904062), and *Inorganic cation transmembrane transport* (GO:0098662). The alteration of pathways and functions related to cation transport have been described to be associated with NAFLD (49), but information about their specific relevance and significance remains scarce.

Therefore, our results support the notion that there are differences between men and premenopausal women in some biological processes and molecular functions that may be

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

relevant in the NAFLD. Given the well-known heterogeneity in NAFLD progression between males and females this transcriptomic meta-analysis also suggests that pathophysiologically-informed disease biomarkers for males and females may need to be different. Can we further extrapolate these findings to disease definition and clinical trials? For example, biological sex-driven differences may be utilized to better classify any individual with NAFLD and their inclusion into clinical trials. Increased understanding of the differences in the risk of progression and hepatocellular carcinoma may be included in future practice guidelines. Thus, our meta-analysis underscores the importance of recognizing and accounting for sex-related differences in the study of NAFLD pathobiology.

Conclusions

Differences in clinical characteristics and pathogenic mechanisms of NAFLD in men versus women have been explored previously but have been not been well characterized. In this regard, starting from a total of 114 GEO studies we selected 7 studies suitable for a functional meta-analysis. We identified a total of 13 significant GO terms from these data, some of which have been previously identified to be relevant to the disease. This supports the notion that sex is an important variable in understanding NAFLD pathobiology and heterogeneity. However, other functional terms we identified have not been previously linked to NAFLD, opening new avenues for further study. In conclusion, meta-analysis of transcriptomic data is a helpful approach to identify sex-related differences in the progression from NAFL to NASH.

Acknowledgements

The authors would like to thank the Principe Felipe Research Centre (CIPF) for providing access to its computational cluster in order to carry out this work. We would also like to thank Dr. Kristoffer T.G. Rigbolt for providing information about ages of the sample used in study GSE126848.

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

References

1. Friedman SL, Neuschwander-Tetri BA, Rinella M, Sanyal AJ. Mechanisms of NAFLD development and therapeutic strategies. Nat Med 2018;24:908–922.
2. Younossi ZM, Koenig AB, Abdelatif D, Fazel Y, Henry L, Wymer M. Global epidemiology of nonalcoholic fatty liver disease-Meta-analytic assessment of prevalence, incidence, and outcomes. Hepatology 2016;64:73–84.
3. Ballestri S, Nascimbeni F, Baldelli E, Marrazzo A, Romagnoli D, Lonardo A. NAFLD as a Sexual Dimorphic Disease: Role of Gender and Reproductive Status in the Development and Progression of Nonalcoholic Fatty Liver Disease and Inherent Cardiovascular Risk. Adv Ther 2017;34:1291–1326.
4. Zheng D, Wang X, Antonson P, Gustafsson J-Å, Li Z. Genomics of sex hormone receptor signaling in hepatic sexual dimorphism. Mol Cell Endocrinol 2018;471:33–41.
5. **Ling G, Sugathan A,** Mazor T, Fraenkel E, Waxman DJ. Unbiased, Genome-Wide In Vivo Mapping of Transcriptional Regulatory Elements Reveals Sex Differences in Chromatin Structure Associated with Sex-Specific Liver Gene Expression. Mol Cell Biol 2010;30:5531–5544.
6. Lonardo A, Suzuki A. Nonalcoholic fatty liver disease: does sex matter? HepatoBiliary Surg Nutr 2019;8:164–166.
7. Aller R, Fernández-Rodríguez C, lo Iacono O, Bañares R, Abad J, Carrión JA, et al. Consensus document. Management of non-alcoholic fatty liver disease (NAFLD). Clinical practice guideline. Gastroenterol y Hepatol 2018;41:328–349
8. Singh S, Allen AM, Wang Z, Prokop LJ, Murad MH, Loomba R. Fibrosis Progression in Nonalcoholic Fatty Liver vs Nonalcoholic Steatohepatitis: A Systematic Review and Meta-analysis of Paired-Biopsy Studies. Clin Gastroenterol Hepatol 2015;13:643-654.e9.
9. Moher D, Liberati A, Tetzlaff J, Altman DG. Preferred Reporting Items for Systematic Reviews and Meta-Analyses: The PRISMA Statement. PLoS Med 2009;6:e1000097.
10. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. 2019;Available from: http://www.r-project.org/
11. Bolstad BM, Irizarry RA, Astrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinformatics 2003;19:185–193.
12. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol 2010;11:R25.
13. Davis SR, Lambrinoudaki I, Lumsden M, Mishra GD, Pal L, Rees M, et al. Menopause. Nat Rev Dis Prim 2015;1:15004.
14. Freeman EW, Sammel MD, Sanders RJ. Risk of long-term hot flashes after natural menopause. Menopause 2014;21:924–932.
15. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 2015;43:e47–e47.
16. **Robinson MD, McCarthy DJ,** Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139–140.
17. Benjamini Y, Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. J R Sta. Soc Ser B 1995;57:289–300.
18. **Subramanian A, Tamayo P,** Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci 2005;102:15545–15550.
19. Montaner D, Dopazo J. Multidimensional Gene Set Analysis of Genomic Data. PLoS

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

One 2010;5:e10348.
20. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene Ontology: tool for the unification of biology. Nat Genet 2000;25:25–29.
21. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res 2000;28:27–30.
22. Lex A, Gehlenborg N. Sets and intersections. Nat Methods 2014;11:779–779.
23. Hidalgo MR, Cubuk C, Amadoz A, Salavert F, Carbonell-Caballero J, Dopazo J. High throughput estimation of functional cell activities reveals disease mechanisms and predicts relevant clinical outcomes. Oncotarget 2017;8:5160–5178.
24. The UniProt Consortium. UniProt: the universal protein knowledgebase. Nucleic Acids Res 2016;45:D158–D169.
25. García-García F. Métodos de análisis de enriquecimiento funcional en estudios genómicos [Internet]. 2016. Available from: https://www.educacion.gob.es/teseo/mostrarRef.do?ref=1307283
26. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. J Stat Softw 2010;36.
27. **Ahrens M, Ammerpohl O, von Schönfels W,** Kolarova J, Bens S, Itzel T, et al. DNA Methylation Analysis in Nonalcoholic Fatty Liver Disease Suggests Distinct Disease-Specific and Remodeling Signatures after Bariatric Surgery. Cell Metab 2013;18:296–302.
28. Horvath S, Erhart W, Brosch M, Ammerpohl O, von Schonfels W, Ahrens M, et al. Obesity accelerates epigenetic aging of human liver. Proc Natl Acad Sci 2014;111:15538–15543.
29. Xanthakos SA, Jenkins TM, Kleiner DE, Boyce TW, Mourya R, Karns R, et al. High Prevalence of Nonalcoholic Fatty Liver Disease in Adolescents Undergoing Bariatric Surgery. Gastroenterology 2015;149:623-634.e8.
30. Lefebvre P, Lalloyer F, Baugé E, Pawlak M, Gheeraert C, Dehondt H, et al. Interspecies NASH disease activity whole-genome profiling identifies a fibrogenic role of PPARα-regulated dermatopontin. JCI Insight 2017;2:e92264.
31. Arendt BM, Comelli EM, Ma DWL, Lou W, Teterina A, Kim T, et al. Altered hepatic gene expression in nonalcoholic fatty liver disease is associated with lower hepatic n-3 and n-6 polyunsaturated fatty acids. Hepatology 2015;61:1565–1578.
32. **Suppli MP, Rigbolt KTG,** Veidal SS, Heebøll S, Eriksen PL, Demant M, et al. Hepatic transcriptome signatures in patients with varying degrees of nonalcoholic fatty liver disease compared with healthy normal-weight individuals. Am J Physiol Gastrointest Liver Physiol 2019;316:G462–G472.
33. Hoang SA, Oseini A, Feaver RE, Cole BK, Asgharpour A, Vincent R, et al. Gene Expression Predicts Histological Severity and Reveals Distinct Molecular Profiles of Nonalcoholic Fatty Liver Disease. Sci Rep 2019;9:12541.
34. **Konishi S, Yano T,** Tanaka H, Mizuno T, Kanoh H, Tsukita K, et al. Vinculin is critical for the robustness of the epithelial cell sheet paracellular barrier for ions. Life Sci Alliance 2019;2:e201900414.
35. Peng X, Nelson ES, Maiers JL, DeMali KA. Chapter five - New Insights into Vinculin Function and Regulation. In: Jeon KW, eds. International Review of Cell and Molecular Biology. Volume 287. Academic Press, 2011:191–231.
36. Vandel J, Dubois−Chevalier J, Gheeraert C, Derudas B, Raverdy V, Thuillier D, et al. Hepatic molecular signatures highlight the sexual dimorphism of Non - Alcoholic SteatoHepatitis (NASH). Hepatology 2020;hep.31312.
37. **Ward NC, Yu A,** Moro A, Ban Y, Chen X, Hsiung S, et al. IL-2/CD25: A Long-Acting Fusion Protein That Promotes Immune Tolerance by Selectively Targeting the IL-2 Receptor on Regulatory T Cells. J Immunol 2018;201:2579–2592.
38. Waters RS, Perry JSA, Han S, Bielekova B, Gedeon T. The effects of interleukin-2 on

medRxiv preprint doi: https://doi.org/10.1101/2020.06.03.20118570; this version posted August 4, 2020. The copyright holder for this preprint **(which was not certified by peer review)** is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY 4.0 International license .

immune response regulation. Math Med Biol 2018;35:79–119.
39. Saxena S, Nouri-Aria KT, Anderson MG, Eddleston AL, Williams R. Interleukin 2 activity in chronic liver disease and the effect of in vitro alpha-interferon. Clin Exp Immunol 1986;63:541–548.
40. Perito ER, Ajmera V, Bass NM, Rosenthal P, Lavine JE, Schwimmer JB, et al. Association Between Cytokines and Liver Histology in Children with Nonalcoholic Fatty Liver Disease. Hepatol Commun 2017;1:609–622.
41. Wieckowska A, Papouchado BG, Li Z, Lopez R, Zein NN, Feldstein AE. Increased hepatic and circulating interleukin-6 levels in human nonalcoholic steatohepatitis. Am. J Gastroenterol 2008;103:1372–1379
42. Braunersreuther V, Viviani GL, Mach F, Montecucco F. Role of cytokines and chemokines in non-alcoholic fatty liver disease. World J Gastroenterol 2012;18:727–735.
43. Schmidt-Arras D, Rose-John S. IL-6 pathway in the liver: From physiopathology to therapy. J. Hepatol 2016;64:1403–1415.
44. Yamaguchi K, Itoh Y, Yokomizo C, Nishimura T, Niimi T, Fujii H, et al. Blockade of interleukin-6 signaling enhances hepatic steatosis but improves liver injury in methionine choline-deficient diet-fed mice. Lab Investig 2010;90:1169–1178.
45. Wang X, Yan Z, Ye Q. Interleukin-6 gene polymorphisms and susceptibility to liver diseases: A meta-analysis. Medicine 2019;98:e18408.
46. Lonardo A, Nascimbeni F, Ballestri S, Fairweather D, Win S, Than TA, et al. Sex Differences in Nonalcoholic Fatty Liver Disease: State of the Art and Identification of Research Gaps. Hepatology 2019;70:1457–1469.
47. Herrera-Marcos L V., Sancho-Knapik S, Gabás-Rivera C, Barranquero C, Gascón S, Romanos E, et al. Pgc1a is responsible for the sex differences in hepatic Cidec/Fsp27β mRNA expression in hepatic steatosis of mice fed a Western diet. Am J Physiol Metab 2020;318:E249–E261.
48. Matsusue K. A Physiological Role for Fat Specific Protein 27/Cell Death-Inducing DFF45-Like Effector C in Adipose and Liver. Biol Pharm Bull 2010;33:346–350.
49. **Pirola CJ, Sookoian S.** Multiomics biomarkers for the prediction of nonalcoholic fatty liver disease severity. World J Gastroenterol. 2018;24:1601–1615.