In *silico* search for modifier genes associated with pancreatic and liver disease in Cystic Fibrosis

Pascal Trouvé1*, Emmanuelle Génin1, Claude Férec1,2,3,4

1 Inserm, UMR1078, Brest, France, 2 Université de Bretagne Occidentale, Faculté de Médecine et des sciences de la santé, Brest, France, 3 C.H.U. Brest, Hôpital Morvan, Laboratoire de Génétique Moléculaire, Brest, France, 4 Etablissement Français du Sang - Bretagne, Brest, France

* pascal.trouve@univ-brest.fr

Abstract

Cystic Fibrosis is the most common lethal autosomal recessive disorder in the white population, affecting among other organs, the lung, the pancreas and the liver. Whereas Cystic Fibrosis is a monogenic disease, many studies reveal a very complex relationship between genotype and clinical phenotype. Indeed, the broad phenotypic spectrum observed in Cystic Fibrosis is far from being explained by obvious genotype-phenotype correlations and it is admitted that Cystic Fibrosis disease is the result of multiple factors, including effects of the environment as well as modifier genes. Our objective was to highlight new modifier genes with potential implications in the lung, pancreatic and liver outcomes of the disease. For this purpose we performed a system biology approach which combined, database mining, literature mining, gene expression study and network analysis as well as pathway enrichment analysis and protein-protein interactions. We found that IFI16, CCNE2 and IGFBP2 are potential modifiers in the altered lung function in Cystic Fibrosis. We also found that EPHX1, HLA-DQA1, HLA-DQB1, DSP and SLC33A1, GPNMB, NCF2, RASGRP1, LGALS3 and PTPN13, are potential modifiers in pancreas and liver, respectively. Associated pathways indicate that immune system is likely involved and that Ubiquitin C is probably a central node, linking Cystic Fibrosis to liver and pancreatic disease. We highlight here new modifier genes with potential implications in Cystic Fibrosis. Nevertheless, our *in silico* analysis requires functional analysis to give our results a physiological relevance.

Introduction

Cystic Fibrosis (CF) is the most common lethal autosomal recessive disorder in the white population. Its incidence is one in 2,500 with a carrier frequency of one in 25. The Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) gene, causing CF, was identified in 1989 [1] and located on chromosome 7q31.2, spanning a transcription unit of about 216,7 kb with 27 exons [2]. It encodes a transmembrane protein (CFTR, 1,480 amino acids) which is an ATP-binding cassette transporter functioning as a chloride (Cl-) channel [1–6]. The CFTR channel's
opening requires phosphorylation by cAMP-dependent protein kinases [6–8] and hydrolyzable MgATP [8]. The CFTR protein is located at the apical membrane of polarized epithelial cells in diverse tissues including the lungs, sweat ducts, pancreas, gastrointestinal tract, and vas deferens [9, 10].

There are currently 2,009 mutations listed in the CFTR mutation database (http://www.genet.sickkids.on.ca/cftr/app). The most common one, F508del-CFTR (a 3 bp deletion in exon 10 causing loss of the amino acid phenylalanine at position 508), encodes a cAMP-regulated Cl⁻ channel that is retained in the endoplasmic reticulum (ER) during translation and folding and is targeted to the early proteasomal degradation [11]. Besides an altered folding, the F508del-CFTR protein exhibits an altered function with an open time period of approximately 0.1 to 0.3 seconds [12, 13].

The phenotype due to a given mutation depends on its interaction with the second mutated CFTR allele and largely on disease modifiers. Whereas CF is a monogenic disease, there is a very complex relationship between genotype and phenotype [14]. A good genotype-phenotype correlation is observed in the pancreas. Severe and mild mutations are associated with pancreatic insufficiency (PI) and pancreatic sufficiency, respectively [15]. Although CFTR genotype is clearly predictive of PI [16], the genotype-phenotype link is unclear regarding lung and liver diseases. Environmental factors such as infection, nutritional status or socioeconomic status may influence in the pulmonary phenotype in CF. Nevertheless, they can’t explain the degree of variability observed in patients exhibiting a same CFTR genotypes. This is shown by studies in CF twins indicating that other genetic factors may explain the observed variability [17, 18]. Modifier genes may play a significant role in determining the severity of CF. Furthermore, for a single mutation such as F508del, the pulmonary status may range from insignificant to severe [19, 20]. It is therefore obvious that, whereas classic genotype-phenotype studies in CF are important, they are not sufficient and have to be complemented by the search for environmental effects on the phenotype of patients, to increase the basic knowledge of the disease and to develop new therapeutic approaches.

The influence of genetic modulators on CF [21] shows the importance to search for potential candidate gene acting upon CF phenotype. For a given CF genotype, modifier genes may influence the phenotype via the same or parallel pathways [22], leading to complex studies. Therefore, in many works, the candidate gene approach is used. Candidate genes are search in the interactome of the disease gene or in a pathway that is indirectly involved in the course of the disease. Some genome-wide association meta-analysis also identified modifier loci of lung disease severity [23]. In the present work, a different methodology was used.

To search for genetic factors that may play a role in CF, we used a system biology approach. This approach combined database and literature mining, gene expression, and network analysis. Pathway enrichment analysis and protein-protein interactions (PPI) were also searched, to give our results a physiological relevance. This approach allowed us to examine functional relationships between reported genes, and led us to identify novel genes and enriched pathways that may play a role in CF, regarding its lung, pancreatic and liver affections. We found that genes involved in immunity including Ubiquitin C are likely modifiers.

**Materials and methods**

**Data sources and gene selection from databases**

The gene selection was performed using a previously used methodology [24].

Genes were first identified utilizing the Online Mendelian Inheritance in Man (OMIM) database (http://www.omim.org/) because it is considered to be the best curated resource for genotype-phenotype relationships [25, 26]. In OMIM, we used the following key words: CF,
CF associated genes, PI and Liver Disease. CF and CF associated genes were used to retrieve genes directly involved in CF as well as genes potentially implicated in the disease.

The Comparative Toxicogenomics Database (CTD, http://ctdbase.org/, [27]) was also used. It curates relationships between genes and human diseases in a unique fashion which integrates gene/protein–disease relationships. In CTD, CF, Pancreatic diseases and Liver Disease key words were used. In CTD, disease-gene associations are reported as curated or inferred. We selected curated associations due to a higher confidence than inferred associations. Solely the genes that may be biomarkers of a disease or play a role in the etiology of a disease were selected.

Lastly, the Human Genome Epidemiology encyclopedia (HuGE Navigator, http://www.hugenavigator.net/HuGENavigator/startPagePubLit.do, [28]) which mines the scientific literature on human gene–disease associations was used. It is a database of population-based epidemiologic studies of human genes [29]. In HuGE Navigator we used the CF, PI and Liver Disease key words. The genes with a genetic association above 5 were selected (score formula is described in [30]).  

Venn diagrams were drawn to find common genes from these databases.

**Comparison with published gene lists from bibliography**

Differentially expressed genes in CF were retrieved from works published in PubMed (http://www.ncbi.nlm.nih.gov/pubmed). Up- and down-regulated genes were retrieved from a paper [31] in which 4 independent studies [32–35] and a re-analyzed list of genes [36] were compared. We used here a list of 75 up-regulated and 114 down-regulated genes which are shared by these studies and having the same direction of expression in at least two studies [31–37]. These genes are given in Table 1.

To search for up- and down-regulated genes in Liver Disease, we focused on gene expression in Human non alcoholic steato-hepatitis (NASH, [38–41]). The rationale that liver disease in NASH and liver of patients with CF share common pathways is based on the fact that steatosis is one of the most common hepatic affection in CF (prevalence of 23–75%). It is observed in 70% of children with liver disease, independently of their nutritional status. Together with a uniform hyperechogenicity of their liver, some pseudomasses may be seen by ultrasound and are due to lobulated fatty structures. In about 60% of the cases, the steatosis is associated with hepatic enzymes elevation (aminotransferases). Because this is observed in children, it is likely not due to alcohol consumption. Therefore, beside the many liver affections observed in livers from CF patients, they present many features of NASH (for review: [42]). Liver Disease key word was not used because it did not permit to retrieve studies with gene analysis. In a first study [38], 16 genes were differentially expressed in subjects with NASH when compared with healthy controls. 12 and 4 genes were significantly under- and 4 over-expressed in NASH, respectively. In a second study [39] comparing NASH samples versus controls, 14 genes were found with more than a two-fold difference. In the third study [40], comparison of non-obese controls patients with NASH exhibited 34 differentially expressed genes (> two fold). We noticed that the expression of 19 genes among 34, were not different in the obese and non-obese controls, showing that the second study using obese controls could be used here. In a last study [41], 9724 genes were differentially expressed in the normal liver versus NASH, (21,619 genes were unchanged).  

To search for disease-specific genes in Chronic Pancreatitis, a study showing 152 and 34 genes with increased and decreased expression, respectively, was used [43]. Full name and
OMIM identification for these 186 genes was searched. We used another study from which we extracted specific genes in Chronic Pancreatitis [44].

In our study, some genes were withdrawn because the corresponding record was deleted by NCBI or because the term was not found in Gene records. Pseudogenes and non human genes (artifacts) were also withdrawn.

### Expression analysis

Datasets restricted to humans, were collected from the Gene Expression Omnibus (GEO, [45]) and GEO2R was used for the analysis [46]. For this purpose, experimental groups were re-assigned. The following datasets were passed through the GEO2R online application:

- GSE40445 (Gene expression in CF-vs-non CF airway epithelial cells from nasal brushing), GSE48452 (Gene expression in control liver-vs-NASH) and GSE44314 (Gene expression in type 1 diabetes-vs-healthy controls). Type 1 diabetes was used because no PI dataset was found. Samples were assigned to the case or control groups. The GEO2R uses the limma (Linear Models for Microarray Data) package from Bioconductor and was designed to analyze complex experiments involving comparisons between many RNA targets simultaneously, with the idea to fit to a linear model to the expression data for each gene. Whereas limma provides

| Origin of the gene lists (References) | Regulated genes shared with [31] |
|--------------------------------------|---------------------------------|
| [33]                                 | UP: ACAA2, CDKN2B               |
|                                      | DOWN: IGFBP2                   |
| [34]                                 | UP: C9orf3, KRT14               |
|                                      | DOWN: CYP24A1, HLA-DQA1, SAA4  |
| [32]                                 | UP: CAV1, CCNE2                 |
|                                      | DOWN: CLGN, ENO2, EPB41L3, GPX3, TIMP4 |
| [36]                                 | UP: BCL2A1, G0S2, IL1B, MMP1, RGS2 |
|                                      | DOWN: CKB, CRIP1, CYP24A1, DNLAI1, FHL1, GSTT1, IGFBP2 |
| [37]                                 | UP: BCL2A1, G0S2, IL1B, IL1R2, LPC2, NDRG1, RGS2, RNF149, TCNT |
|                                      | DOWN: PROS1, SCGB1A1, SPAG8     |

All shared genes in the 6 studies [31–35, 37]

**UP (n = 75):**
- ACAA2, AGL, AKR1C1, ANXA8L2, BCHE, BLOC1S1, BTBD3, C9orf3, CAPG, CAV1, CCL20, CCNE2, CD24, CD83, CDKN2B, CLGN, CSF3, DD3X, ELK3, FOLR1, FOLR3, FOXL1, GCA, GFTPT2, HCLS1, HIST1H1C, HMCO1, HPCAL1, HSPB11, IFIT1, IFIT3, IFITM1, IL1R2, IL7R, ISG15, KCTD12, KRT14, KRT81, LCP2, LITAF, LYPD1, MLF1, MMP1, MRPL28, MX2, NCF1, NDRG1, NET1, PLAU, PLAU, PLTP, PRSS3, PSG9, PTTP13, RAC2, RAGE, RNF149, RPA3, SEMA3B, SERPINA3, SERPINF1, SLITRK5, SOD2, SULT1A3, TF51F5, TCIRG1, TCNT1, TXNIP, BCL2A1, G0S2, IL1B, NCF2, PLAT, PTGS2, RGS2.

**DOWN (n = 114):**
- ACTA2, ADAR, ALDH1A1, ANKR1D1, ASNS, ASS1, BEX4, BST2, BTG1, C9orf13, CALD1, CAP1, CCL20, CD164, CFB, CGRE1F, CKB, CLGN, COL8A1, COL9A3, CRIP1, CSTA, CXCR4, CYP51A1, DDB2, DDI4, DNLAI1, DSP, DSN3, DYNLT1, EDNRA, EFEMP1, EIF4A2, ENO2, EPO41L3, EPR8, F3, FBNL5, FCGBP, FHL1, GABRP, GCH1, GINS1, GPMB, GPR1, GPX3, GSTT1, GZMB, HES1, HLA-B, HLA-DOA1, HLA-DRB1, HLA-F, HLA-G, HMGCS1, HSPB1, HTRA1, ID1, IFIT1, IFITM1, IFRD1, IGFBP7, IL32, KCNN4, KIT, KLK1, KRT15, LCN2, LGALS3BP, LOX, MGAM, MUC, NID2, NPR3, NT5D2, PGD, PMAA2, PPIP1R3C, PRAS1, RASGRF1, RASD3, RUNX3, SAA4, SC5DL, SCGB1A1, SERPINB3, SERPINB4, SGK1, SLC2A3P1, SNAPC1, SPAG8, STAC, STAT4, TES, TFPI, THBD, TIMP4, TPBG, TRIB2, TWIST1, VDAC1, ZNF643, CTSC, CYP24A1, IFI27, IGFBP2, IGFBP3, NAMPT, PRSS23, SEL1L3, TMSB4X, TRIM22

https://doi.org/10.1371/journal.pone.0173822.t001

OMIM identification for these 186 genes was searched. We used another study from which we extracted specific genes in Chronic Pancreatitis [44].

In our study, some genes were withdrawn because the corresponding record was deleted by NCBI or because the term was not found in Gene records. Pseudogenes and non human genes (artifacts) were also withdrawn.
several p-value options, we applied the adjustment to the p-values also called multiple-testing corrections, to correct the occurrence of false positive results. The Benjamini & Hochberg false discovery rate method, commonly used to adjust microarray data, was selected. After GEO2R dataset analysis, genes of interest were selected by using the criteria of a Bonferroni-corrected p-value $\leq 0.05$. According to the annotation files, probes that were not corresponding to any genes and genes which were present several times within a single study were eliminated. Differentially expressed transcripts between normal and pathologic were selected and common genes between GSE40445, GSE48452 and GSE44314 were searched by drawing a Venn diagram (not shown).

Analysis of enrichment of KEGG biochemical pathways in CTD

Our target genes were mapped to biological process pathways using CTD (http://ctd.mdibl.org). Biocurators at CTD manually curated gene-disease relationships from the literature as well as pathways. The core data are integrated to construct pathway networks. The Venn-Viewer function, which permits to compare associated datasets, was used for the genes that were found to be potential modifiers in CF, Liver Disease and Chronic Pancreatitis. The Pathway associations function was selected to compare annotated enriched pathways, as well as pathways for disease for these genes. Associated data for interacting genes and diseases, inferred KEGG and REACTOME pathways, or enriched GO terms were downloaded. Annotated pathways for genes are associations established by KEGG and REACTOME curation. The significance of a given pathway is reported as a p-value.

Gene-disease association search

Gene-disease associations were search in CTD in which associations are extracted from the published literature or are derived from the OMIM database using the mim2gene file from the NCBI Gene database. Curated associations were retrieved according to their Bonferroni-corrected p-value $< 0.01$.

PPI networks design

PPI networks were obtained by STRING 10.0 (Search Tool for the Retrieval of Interacting Genes and Proteins, http://string-db.org/). STRING is a global resource for searching connections (edges) between genes or proteins (nodes) currently covering 9,643,763 proteins. Each edge has a confidence value between 0 and 1, lowest and highest confidence, respectively. The gene’s names were used as an initial input and experimental data sets were selected to display interactive networks. The parameters were defined as follows: all PPI prediction methods were enabled, except for High-throughput Lab Experiments; maximum of 5 interactions by node; cut-off criterion of combined score $\geq 0.7$ (interactions at high confidence or better); Homo sapiens.

Results and discussion

Study design

Fig 1 depicts the workflow of our study. We retrieved genes from OMIM, CTD and HuGE. Common genes were searched in CF, pancreatic disease and Liver Disease. Differentially expressed genes in CF were also retrieved from published results in PubMed. Finally, Datasets were collected from the GEO and re-analyzed. Because the CF, Chronic Pancreatitis and Liver Disease keywords did not give any result, type 1 diabetes-vs-healthy control and control liver-vs-NASH were used in GEO. Candidate genes from these databases were compared and
analyzed to search for potential common genes and pathways. PPI were further searched to give our results a physiological relevance.

**Gene selection from databases**

OMIM is a comprehensive, daily-updated human phenotype database, containing more than 12,000 genes of all human genetic diseases. The searches in OMIM using CF, CF associated genes, PI and Liver Disease key words led to 100, 109, 152, 1192 genes, respectively (Fig 2). The intersection of CF and CF associated genes showed 42 different genes and 46 genes were shared by PI and Liver Disease. The 6 genes shared by CF, CF associated genes and PI key words were S100A8/S100A9 (OMIM Accession ID: 123885), CFTR (OMIM Accession ID: 602421), SCNN1G (OMIM Accession ID: 600761), TGFB1 (OMIM Accession ID: 190180),
SERPINA1 (OMIM Accession ID: 107400) and PKD1 (OMIM Accession ID: 601313). In the intersection of CF plus CF associated genes with Liver Disease key words, we found 24 genes (Fig 3). Finally, the number of genes in the intersection of the pooled CF and CF associated genes search and the pooled PI and Liver Disease search was only 4. These genes were CFTR (OMIM Accession ID: 602421), TGFB1 (OMIM Accession ID: 190180), SERPINA1 (OMIM Accession ID: 107400) and PKD1 (OMIM Accession ID: 601313).

Using the CTD database, 11,142 and 813 genes were retrieved using the CF, Pancreatic diseases and Liver Disease key words, respectively. Fig 4 presents a Venn diagram showing the shared genes between these key words. Only 3 genes were found to be shared among the 3 used key words: TGFB1 (NCBI Accession ID: 7040), TNFRSF1A (NCBI Accession ID: 7132), CFM1 (NCBI Accession ID: 10167). Surprisingly, among some differences with the OMIM search, the CFTR gene was not found to be present in the list of genes retrieved using the Liver Disease key word in CTD.

HuGE Navigator database is an integrated, searchable, Web-based knowledge base which mines the scientific literature on human genetic associations and human genome epidemiology [29, 30]. In HuGE Navigator, common genes (score > 5) in PI, CF and Liver Disease were CFTR, SPINK1, GSTP1, PRSS1, ADRB2, GSTM1, GSTT1, HRAS, MIF, OGG1 and TNF. The corresponding OMIM Accession IDs are given in Fig 5.

Differential gene expressions retrieved from bibliographic analysis
Differentially expressed genes in CF were retrieved from PubMed. First, a small-scale microarray study performed with human native nasal epithelial cells (F508del homozygous
patients vs. controls) was used [31]. We also retrieved genes from other papers aimed to compare gene expression in CF vs normal cells [32]. Another work using non-CF and F508del-CFTR homozygote samples, showed significant changes in the expression in 24 genes (two-sample t-test, \( p < 0.00001 \)). A three-filter comparative analysis showed that 18 genes were significantly increased and 6 genes were decreased in CF relative to non-CF samples, respectively [33]. We also retrieved genes from a microarray study in which results from 12 CF and 11 non-CF participants were used [34], and from data in which functional CFTR was absent of the plasma membrane [35]. Finally, we used a paper in which 4 studies evaluating the effect of the F508del-CFTR mutation on airway epithelial cells gene expression were analyzed [36]. A profiled gene expression in CF and non-CF nasal and bronchial epithelium samples, using Illumina HumanRef-8 Expression BeadChips was used. It showed that 863 genes were differentially expressed between CF and non-CF bronchial epithelium and that only 15 were differentially expressed between CF and non-CF nasal epithelium [37]. This indicated that within airways, gene expression varies depending on the region that is studied. Up- and down-regulated genes were compared within these studies [32–37] and common genes were searched (Table 1). We found that 75 genes were up-regulated
and 114 were down-regulated in human airway cells expressing F508del-CFTR (Table 1, lower part).

Nonalcoholic fatty liver disease has a large spectrum ranging from simple steatosis to NASH, which may lead to progressive fibrosis. In the first study that we used [38], the abundance of intra-hepatic messenger RNA for a broad array of genes was measured. From this study we retrieved differentially expressed genes in NASH-vs-normal liver. Among 6,412 genes, only 16 were differentially expressed in subjects with NASH when compared with controls (Table 2). We used another study using microarrays [39]. 14 genes for NASH-vs-obese controls were found to be up-regulated (Table 2). In a third study, genes in NASH patients and controls were selected [40]. Whereas, 34 genes were differentially expressed in NASH vs non-obese controls, 19 of these genes had no significant differences in obese vs non-obese, suggesting a stronger association of these genes to NASH. Therefore, we used this list of 15 genes (Table 2). Finally, a study with normal, steatosis, NASH with fatty liver, and NASH without fatty liver samples was analyzed. Among 11,633 genes with altered expression out of 33,252 genes, 39 genes were changed in expression, between normal and NASH [21]. Thus, a list of 84 differentially expressed genes from 4 different studies was used in the present study (Table 2).

Fig 4. Venn diagrams of the CTD database search. The 3 shared genes in CF, Pancreatic diseases and Liver Disease were TGFB1, TNFRSF1A and CFM1.

https://doi.org/10.1371/journal.pone.0173822.g004
For pancreatic disease, a study with normal and Chronic Pancreatitis specimens was used [43]. Comparison of the expression of 5,600 genes between the normal and Chronic Pancreatitis was performed. GenBank accession numbers of 152 genes with increased expression ([43], Footnote 1) and 34 with decreased genes levels in Chronic Pancreatitis ([43], Footnote 2) were retrieved. We analyzed those genes one by one and removed non human genes and pseudo genes. This led us to list 129 increased genes and 23 decreased genes in Chronic Pancreatitis (Table 3). Because 152 genes were simultaneously increased in pancreatic cancer, only 5 of 5,600 genes were significantly over expressed in Chronic Pancreatitis compared with normal pancreas: Mucin-6 (GenBank Accession Number: L07517), COMP (GenBank Accession Number: L32137), TPSB1 (GenBank Accession Number: M33493), Rearranged Ig-lambda light chain (GenBank Accession Number: X57809) and CRISP-3 (GenBank Accession Number: X95240). An analysis of 6,800 different genes expressed in samples of normal pancreas and Chronic Pancreatitis was studied [44]. 107 genes were predicted to be expressed within cells with Chronic Pancreatitis ([44], Table 1). Genes from both studies [43 and 44] were retrieved and non human, pseudo genes and common genes were withdrawn. Finally, we found 23 decreased genes and 229 increased genes in Chronic Pancreatitis, when compared to normal pancreas (Table 4).

Differential gene expressions retrieved from microarray studies

For pancreatic disease, a study with normal and Chronic Pancreatitis specimens was used [43]. Comparison of the expression of 5,600 genes between the normal and Chronic Pancreatitis was performed. GenBank accession numbers of 152 genes with increased expression ([43], Footnote 1) and 34 with decreased genes levels in Chronic Pancreatitis ([43], Footnote 2) were retrieved. We analyzed those genes one by one and removed non human genes and pseudo genes. This led us to list 129 increased genes and 23 decreased genes in Chronic Pancreatitis (Table 3). Because 152 genes were simultaneously increased in pancreatic cancer, only 5 of 5,600 genes were significantly over expressed in Chronic Pancreatitis compared with normal pancreas: Mucin-6 (GenBank Accession Number: L07517), COMP (GenBank Accession Number: L32137), TPSB1 (GenBank Accession Number: M33493), Rearranged Ig-lambda light chain (GenBank Accession Number: X57809) and CRISP-3 (GenBank Accession Number: X95240). An analysis of 6,800 different genes expressed in samples of normal pancreas and Chronic Pancreatitis was studied [44]. 107 genes were predicted to be expressed within cells with Chronic Pancreatitis ([44], Table 1). Genes from both studies [43 and 44] were retrieved and non human, pseudo genes and common genes were withdrawn. Finally, we found 23 decreased genes and 229 increased genes in Chronic Pancreatitis, when compared to normal pancreas (Table 4).

Differential gene expressions retrieved from microarray studies

We also acquired gene array data from the GEO database, a public available archive of individual microarrays studies [46, 47]. GSE40445 [31], is a study of gene expression in human native
Table 2. Retrieved differentially expressed genes in Liver Disease (from [38–41]). When possible, names or abbreviations as well as OMIM accession number were added (in italic). The number sign # is used because α1-antitrypsin deficiency is caused by mutation in the SERPINA1 gene (OMIM: 107400). “?” is used when no result was obtained in our OMIM search for the corresponding gene name or when several results could be retrieved. The lower list is retrieved from [21] shows the Solute Carrier Family (SLC, sodium/potassium/chloride transporter family) with differential gene expression in NASH.

| Gene name                          | NASH vs. controls | Reference |
|------------------------------------|-------------------|-----------|
| Aldehyde oxidase (AOX1, OMIM: 602841) | -2.4              |           |
| Cu/Zn SOD (SOD1, OMIM: 147450)      | -3.7              |           |
| Catalase (CAT, OMIM: 115500)       | -2.4              |           |
| Glucose-6 phosphatase (G6PC, OMIM: 613742) | -4.9              |           |
| Alcohol dehydrogenase (?)          | -2.6              |           |
| Ubiquitin (?)                      | +2.6              |           |
| methylmalonate semialdehyde dehydrogenase (?) | -3.3              |           |
| Elongation factor TU (TUFM, OMIM: 602389) | -2.9              |           |
| HM CoA-synthase (COASY, OMIM: 609855) | -2.2 [38]         |           |
| LCA CoA-synthetase (?)              | -2.6              |           |
| MO CoA-thiolase (?)                 | -2.3              |           |
| a1-antitrypsin#                    | +3.7              |           |
| Complement C3 (C3, OMIM: 120700)   | +3.3              |           |
| HFREP-1 (?)                        | +2.7              |           |
| Cytochrome P-450 4A (CYP4A11 OMIM: 601310) | -3.9              |           |
| MC-specific LRP (LRP1, OMIM: 107770) | -3.7              |           |
| POU2AF1 (POU Domain, Class 2, Associating Factor 1, OMIM: 601206) | +2.8              |           |
| MMP2 (Matrix Metalloproteinase 2, OMIM: 120360) | +2.4              |           |
| C-fos-induced growth factor/vascular endothelial growth factor D (FIGF, OMIM: 300091) | +2.2              |           |
| c12orf14 (?)                       | +2.2              |           |
| CSPG2 (?)                          | +2.1              |           |
| Hypothetical protein dj462o23.2    | +2.1              |           |
| ROD1 (Regulator of Differentiation 1, OMIM: 607527) | +2.1 [39]         |           |
| STXB1 (Syntaxin-Binding Protein 1, OMIM: 602926) | +2.1              |           |
| Serine dehydratase (SDS, OMIM: 182128) | +0.5              |           |
| SULT1A2 (Sulfotransferase Family 1A, OMIM: 601292) | +0.5              |           |
| CUTL2 (Cut-Like 2, OMIM: 610648)   | +0.5              |           |
| ARL6IP (ADP-Ribosylation-Like Factor 6-Interacting Protein 1, OMIM: 607669) | +0.5              |           |
| Acyl-coenzyme A (CoA) dehydrogenase | +0.5              |           |
| Ornithine aminotransferase (OAT, OMIM: 613349) | +0.4              |           |
| NNMT (Nicotinamide N-methyltransferase, OMIM: 600008) | +0.3              |           |
| ODC1 (Ornithine Decarboxylase 1, OMIM: 165640) | +0.4              |           |
| IGFBP1 (Insulin-Like Growth Factor-Binding Protein 1, OMIM: 146730) | +0.1              |           |
| ATF3 (Activating Transcription Factor 3, OMIM: 603148) | +0.1              |           |
| CHI3L1 (Chitinases 3-Like 1, OMIM: 601525) | +0.4              |           |
| NFIL-3 (Nuclear Factor, Interleukin 3-Regulated, OMIM: 605327) | +0.4              |           |
| ZFP36 (Zinc Finger Protein 36, OMIM: 190700) | +0.3              |           |
| IL-15RA (Interleukin 15 Receptor, Alpha, OMIM: 601070) | +0.4 [40]         |           |
| TGFβ1 (Transforming Growth Factor, Beta-1, OMIM: 190180) | +0.4              |           |
| CAT (Catalase, OMIM: 115500)       | +2.0              |           |
| GSTA4 (Glutathione S-Transferase, Alpha-4, OMIM: 605450) | +2.1              |           |
| LTF (Lactotransferrin, OMIM: 150210) | +2.0              |           |
| TF (transferrin, OMIM: 190000)     | +2.0              |           |
| ABR (Active BCR-Related Gene, OMIM: 600365) | +0.3              |           |
| PNRC1 (Proline-Rich Nuclear Receptor Coactivator 1, OMIM: 606714) | +0.3              |           |
| SULT1C4 (Sulfotransferase Family 1C, Member 4?) | +1              |           |
| MTR (5'-Methyltetrahydrofolate-Homocysteine S Methyltransferase, OMIM: 156570) | +1              |           |

(Continued)
| Gene name                                      | NASH vs. controls | Reference |
|-----------------------------------------------|-------------------|-----------|
| PAPSS2 (3′-Phosphoadenosine 5′-Phosphosulfate Synthase2, OMIM: 603005) | -1                |           |
| CYP4F11 (Cytochrome P450, Family 4, Subfamily F, Polypeptide 11, OMIM: 611517) | -1                |           |
| ALDH4A1 (Aldehyde Dehydrogenase, Family 4, Subfamily A, Member 1, OMIM: 606811) | -1                |           |
| AOX1 (Aldehyde Oxidase 1, OMIM: 602841)       | -1                |           |
| UGCG (UDP-Glucose Ceramide Glucosyltransferase, OMIM: 602874)              | 1                 |           |
| CYP2E1 (Cytochrome P450, Subfamily IIE, OMIM: 124040)              | -1                |           |
| GPX2 (Glutathione Peroxidase 2, OMIM: 138319)              | -1                |           |
| HGSNAT (Heparan-Alpha-Glucosaminide N-Acytethyltransferase, OMIM: 610453) | 1                 |           |
| FMQ2 (Flavin-Containing Monooxygenase 2, OMIM: 603955)              | 1                 |           |
| NNMT (Nicotinamide N-Methyltransferase, OMIM: 600008)              | -1                |           |
| SULT1C2 (Sulfotransferase Family 1C, Member2, OMIM: 608357)          | 1                 |           |
| HSD17B10 (17-β-HydroxySteroid Dehydrogenase X, OMIM: 300256)       | -1                |           |
| LOX (Lysyl Oxidase, OMIM: 135455)              | 1                 |           |
| GGT5 (Gamma-GlutamylTransferase 5, OMIM: 137168)              | 1                 |           |
| GSTP1 (Glutathione S-Transferase, PI, OMIM: 134660)              | 1                 |           |
| PAPSS1 (3′-Phosphoadenosine 5′-Phosphosulfate Synthase 1, OMIM: 603262) | 1                 |           |
| MTRR (Methionine Synthase Reductase, OMIM: 602568)              | -1                |           |
| CYP3A7 (Cytochrome P450, Subfamily IIIa, Polypeptide 7, OMIM: 605340) | 1                 |           |
| GCKR (Glucokinase Regulatory Protein, OMIM: 600842)              | -1                | [41]*     |
| CYP7B1 (Cytochrome P450, Family 7, Subfamily B, Polypeptide 1, OMIM: 603711) | 1                 |           |
| PTGS2 (Prostaglandine-Endoperoxide Synthase 2, OMIM: 600262)       | 1                 |           |
| CYP21A2 (Cytochrome P450, Family 21, Subfamily A, Polypeptide 2, OMIM: 613815) | -1                |           |
| HSD17B4 (17-β-Hydroxysteroid Dehydrogenase IV, OMIM: 601860)         | -1                |           |
| ALDHHA1 (Aldehyde Dehydrogenase 8 Family, Member A1, OMIM: 606467)   | -1                |           |
| GCLM (Glutamate-Cysteine Ligase, Modifier Subunit, OMIM: 601176)     | 1                 |           |
| NAT1 (N-Acetyltransferase 1, OMIM: 108345)                          | -1                |           |
| HNMT (Histamine N-Methyltransferase, OMIM: 605238)                  | 1                 |           |
| CYP51A1 (Cytochrome P450, Family 51, Subfamily A, Polypeptide 1, OMIM: 601637) | -1                |           |
| UCHL3 (Ubiquitin Carboxyl-Terminal Esterase L3, OMIM: 603090)        | -1                |           |
| UGT2B4 (Uridine Diphosphate Glycosyltransferase 2 Family, Member B4, OMIM: 600067) | 1                 |           |
| ALDH1A3 (Aldehyde Dehydrogenase 1 Family, Member A3, OMIM: 600463)  | 1                 |           |
| CYB5R3 (Cytochrome b5 Reductase 3, OMIM: 613213)                    | -1                |           |
| HSD3B2 (3-β-Hydroxysteroid Dehydrogenase 2, OMIM: 613890)            | -1                |           |
| BLVRA (Biliverdin Reductase A, OMIM: 109750)                        | 1                 |           |
| GZMA (Granzyme A, OMIM: 140050)                                    | 1                 |           |
| UGT2B (Uridine Diphosphate Glycosyltransferase)                      | 1                 |           |

*: 0 = No change, +1 = Up regulation, -1 = Down regulation.

https://doi.org/10.1371/journal.pone.0173822.1002
nasal epithelial cells from F508del-CFTR homozygous patients and non-CF controls. GSE48452 [48] is an expression profiling performed with samples of control human liver, healthy obese, steatosis and NASH). Only controls and NASH groups were re-analyzed here. For PI, we failed to find any array analysis. Therefore, we used a gene expression of type 1 diabetes-vs-classical type 1A diabetes-vs-healthy controls array (GSE44314) [49]. We solely re-analyzed classical type 1A diabetes-vs-healthy controls. 246, 211 and 145 genes were retrieved from our GSE40445, GSE48452 and GSE44314 re-analysis, respectively (Table 5). A Venn diagram was drawn (not shown) but no common genes among the genes selected from the 3 studies were found. Therefore, if there are some common genes in the 3 studies, they do not appear among the most highly regulated genes of each study.

Common genes in CF, retrieved from bibliography and microarray studies

We then search for common genes between the genes which were obtained by the bibliographic search for CF genes (Table 1), Liver Disease (Table 2) and Chronic Pancreatitis (Tables 3 and 4) and the genes which were retrieved by the re-analyzing of GSE40445, GSE48452 and GSE44314. The result is summarized in Table 6.

Only 4 genes were found to be common to both groups of genes retrieved from the bibliographic analysis and the GEO database analysis, regarding CF. These genes were IFRD1 (Interferon-related developmental regulator 1, OMIM Accession ID: 603502), IFI16 (Interferon gamma inducible protein 16, OMIM Accession ID: 147586), CCNE2 (Cyclin E2, OMIM Accession ID: 603775) and IGFBP2 (Insulin-like growth factor-binding protein 2, OMIM Accession ID: 146731). This low number of common genes is likely due to the fact that in microarray experiments, the recorded intensities depend on the conditions under which the measurements are made [50].

IFRD1, a histone deacetylase (HDAC)-dependent transcriptional co-regulator expressed during terminal neutrophil differentiation, was already identified of as a modifier gene in CF [51]. It was also found to be common for our CF, CF associated genes and Liver disease keywords search in OMIM.

---

Table 3. Retrieved differentially expressed genes in Chronic Pancreatitis (from [43]). Some genes were withdrawn according to the reasons explained in Methods.

| Genes                          | 129 Genes expressed at increased level | 23 Genes expressed at decreased level |
|-------------------------------|----------------------------------------|--------------------------------------|
| LGALS4, ARPC1B, MMP22, POSTN, CDH11, IQGAP1, HLA-A, LGMN, AP2M1, SERPINH1, FCGBP, Dyrk1A, PXDN, COL5A1, SPARC, CYCA, PLAG2A2, CTSA, COL4A2, HBB, F3, CEACAM5, MGMT, XPC, HLADR, PSMC3, GPB2, HLA-DPB1, TNFRSF1A, IGFBP4, NR2F2, Cfrh1, Tgfb1, Tagln, Hspg2, Cald1, Col1a1, Ppica, Sept7, Lcn2, Fscn1, Cetn1, Mdkg, Cagp, Col1a2, Ighg3, Aldh1a3, Atpp2a3, Tpbg, Gucy1b3, Acvr1, Ifitm2, Col3a1, Fn1, Tff1, Cd16, Hla-f, Gsn, Ass1, Glrx, Hif1a, Lum, Junb, Aoc3, Vcan, Lmb3, Pnmb, Polr2a, Empl, Trim22, Prmt1, Itga1, Mall, Atpp6v0d1, Snrpd3, Nnmnt, Pygb, Icam3, Gpx2, Col6a3, Agps, Tgb2, Sgs6, Loxl1, Rarres1, C3ar1, Hla-dra, Tax1bp3, Akap12, Lita, Ccl19, Abcc8, Filip1l, Gnr, Krt19, Abcg1, Srgn, Flna, Crisp3, Thbs1, Vasp, Ptprc, S100a13, Acta2, Cd14, Ighm, Col4a2, Pkm, Mpmp1, Nfe2l1, Hla-e, Lcn2, Fgll, Fut1, Pcdhb11, Pck1, Tkt, Aox1, Tep1, Bnip3, Art1, Vhl, Spin1, Ephx1, Selemnb1, Slc25a1, Slc14a1, Tym, Pex7, Btn3a1, Fasn, Gst2a, Anpep, Elane, Hoxb2 |

https://doi.org/10.1371/journal.pone.0173822.t003
Table 4. Combined differentially expressed genes in Chronic Pancreatitis [43, 44]. Some genes were withdrawn according to the reasons explained in Methods. OMIM ID and full names were searched for each gene. 23 genes were found to be decreased and 229 genes were found to be increased in Chronic Pancreatitis (p<0.05).

| Genes expressed at increased levels in Chronic Pancreatitis | OMIM ID | Reference |
|-------------------------------------------------------------|---------|-----------|
| LGALS4 (Lectin, Galactoside-Binding, Soluble, 4)            | 602518  |           |
| ARPC1B (actin related protein 2/3 complex subunit 1B)      | 604223  |           |
| PMP22 (peripheral myelin protein 22)                       | 601097/608777 |       |
| POSTN (periostin, osteoblast specific factor)              | 600023  |           |
| CDH11 (cadherin 11)                                        | 603379  |           |
| IGGAP1 (IQ motif containing GTPase activating protein 1)   | 142800  |           |
| HLA-A (major histocompatibility complex, class I, A)       | 602620  |           |
| LGMN (legumain)                                            | 601024  |           |
| AP2M1 (adapter related protein complex 2 mu 1 subunit)     | 600943  |           |
| SERPINH1 (serpin peptidase inhibitor, clade, member 1)     | 601437  |           |
| POSTN (periostin, osteoblast specific factor)              | 600023  |           |
| CDH11 (cadherin 11)                                        | 603379  |           |
| IGGAP1 (IQ motif containing GTPase activating protein 1)   | 142800  |           |
| HLA-A (major histocompatibility complex, class I, A)       | 602620  |           |
| LGMN (legumain)                                            | 601024  |           |
| AP2M1 (adapter related protein complex 2 mu 1 subunit)     | 600943  |           |
| SERPINH1 (serpin peptidase inhibitor, clade, member 1)     | 601437  |           |
| POSTN (periostin, osteoblast specific factor)              | 600023  |           |
| CDH11 (cadherin 11)                                        | 603379  |           |
| IGGAP1 (IQ motif containing GTPase activating protein 1)   | 142800  |           |
| HLA-A (major histocompatibility complex, class I, A)       | 602620  |           |
| LGMN (legumain)                                            | 601024  |           |
| AP2M1 (adapter related protein complex 2 mu 1 subunit)     | 600943  |           |
| SERPINH1 (serpin peptidase inhibitor, clade, member 1)     | 601437  |           |

(Continued)
| Table 4. (Continued) |
|----------------------|
| IGFBP4 (insulin like growth factor binding protein 4) | 146733 |
| NR2F2 (nuclear receptor subfamily 2 group F member 2) | 107773 |
| CFHR1 (complement factor H related) | 134371 |
| TGFBI (transforming growth factor beta induced) | 601692 |
| TAGLN (transgelin) | 600818 |
| MDK (midkine, neurite growth-promoting factor 2) | 162096 |
| CAPG (capping actin protein, gelsolin like) | 153615 |
| COL1A2 (collagen type I alpha 2) | 120160 |
| IGHG3 (immunoglobulin heavy constant gamma 3) | 147120 |
| HSPG2 (heparan sulfate proteoglycan 2) | 142461 |
| CALD1 (caldesmon 1) | 114213 |
| COL1A1 (collagen type I alpha 1) | 120150 |
| PPIC (peptidylprolyl isomerase C) | 123842 |
| SEPT7 (septin 7) | 603151 |
| LCN2 (lipocalin 2) | 600181 |
| FSCN1 (fascin actin-bundling protein 1) | 602689 |
| CETN1 (centrin 1) | 603187 |
| ALDH1A3 (aldehyde dehydrogenase 1 family member A3) | 600463 |
| ATP2A3 (ATPase sarcoplasmic/endoplasmic reticulum Ca2+ transporting 3) | 601929 |
| TPBG (trophoblast glycoprotein) | 190920 |
| GUCY1B3 (guanylate cyclase 1, soluble, beta 3) | 139397 |
| ACVR1 (activin A receptor type 1) | 102576 |
| GLRX (glutaredoxin) | 600443 |
| HIF1A (hypoxia inducible factor 1 alpha subunit) | 603348 |
| LUM (lumican) | 600616 |
| JUNB (jun B proto-oncogene) | 165161 |
| AOC3 (amine oxidase, copper containing 3) | 603735 |
| VCAN (versican) | 118661 |
| LAMB3 (laminin subunit beta 3) | 150310 |
| GPNMB (glycoprotein nmb) | 604368 |
| POLR2A (polymerase (RNA) II subunit A) | 180660 |
| EMP1 (epithelial membrane protein 1) | 602333 |
| TRIM22 (tripartite motif containing 22) | 606559 |
| PRMT1 (arginine methyltransferase 1) | 602950 |
| ITGA1 (integrin subunit alpha 1) | 192968 |
| MALL (mal, T-cell differentiation protein-like) | 602022 |
| ATP6V0D1 (ATPase H+ transporting V0 subunit d1) | 607028 |
| SNRPD3 (small nuclear ribonucleoprotein D3 polypeptide) | 601062 |
| NNMT (nicotinamide N-methyltransferase) | 600008 |
| PYGB (phosphorylase, glycogen; brain) | 138550 |
| ICAM3 (intercellular adhesion molecule 3) | 146631 |
| GPX2 (gluthathione peroxidase 2) | 138319 |
| COL6A3 (collagen type VI alpha 3) | 120250 |
| AGPS (alkylglycerone phosphate synthase) | 603051 |
| ITGB2 (integrin subunit beta 2) | 600065 |
| SGSH (N-sulfoglycosamine sulfohydrolase) | 605270 |
| LOXL1 (lysyl oxidase like 1) | 153456 |
| RARRES1 (retinoic acid receptor responder 1) | 605090 |

(Continued)
Table 4. (Continued)

| Gene Symbol | Description                                      | Gene ID |
|-------------|--------------------------------------------------|---------|
| C3AR1       | complement component 3a receptor 1              | 605246  |
| HLA-DRA     | major histocompatibility complex, class II, DR alpha | 142860  |
| TAX1BP3     | Tax1 binding protein 3                           | 616484  |
| AKAP12      | (A-kinase anchoring protein 12)                  | 604698  |
| LITAF       | lipopolysaccharide induced TNF factor            | 603795  |
| CCL19       | (C-C motif chemokine ligand 19)                  | 602227  |
| ABCC8       | (ATP binding cassette subfamily C member 8)      | 600509  |
| FILIP1L     | (filamin A interacting protein 1-like)           | 612993  |
| GRN         | (granulin)                                       | 138945  |
| KRT19       | keratin 19                                       | 148020  |
| ABCG1       | (ATP binding cassette subfamily G member 1)      | 60376  |
| SRGN        | (serglycin)                                      | 177040  |
| FLNA        | (filamin A)                                      | 300017  |
| CRISP3      | (cysteine rich secretory protein 3)             | 600824  |
| THBS1       | (thrombospondin 1)                               | 188060  |
| VASP        | (vasodilator-stimulated phosphoprotein)          | 601703  |
| PTPRC       | (protein tyrosine phosphatase, receptor type C)  | 151460  |
| S100A13     | (S100 calcium binding protein A13)               | 601989  |
| ACTA2       | (actin, alpha 2, smooth muscle, aorta)           | 102620  |
| CD14        | (CD14 molecule)                                  | 158120  |
| IGHM        | (immunoglobulin heavy constant mu)               | 147020  |
| COL4A2      | (collagen type IV alpha 2)                       | 120090  |
| PKM         | (pyruvate kinase, muscle)                        | 179050  |
| MMP1        | (matrix metalloproteinase 1)                     | 120353  |
| IFITM2      | (interferon induced transmembrane protein 2)     | 605578  |
| COL3A1      | (collagen type III alpha 1)                      | 120180  |
| FN1         | (fibronectin 1)                                  | 135600  |
| TFF1        | (trefoil factor 1)                               | 113710  |
| CDH5        | (cadherin 5)                                     | 60120  |
| HLA-F       | (major histocompatibility complex, class I, F)   | 143110  |
| GSN         | (gelsolin)                                       | 137350  |
| ASS1        | (argininosuccinate synthase 1)                   | 603470  |
| NFE2L1      | (nuclear factor, erythroid 2 like 1)             | 163260  |
| HLA-E       | (major histocompatibility complex, class I, E)   | 143010  |
| LCN2        | (lipocalin 2)                                    | 600181  |
| ACTA2       | (actin, alpha 2, smooth muscle)                  | 102620  |
| AEBP1       | (AE-binding protein)                             | 602981  |
| ALDH2       | (aldehyde dehydrogenase 2, mitochondrial)        | 100650  |
| AOC3        | (amine oxidase, copper containing 3)             | 603735  |
| APOC1       | (apolipoprotein C-I)                             | 107710  |
| APOE        | (apolipoprotein E)                               | 107741  |
| BTN3A1      | (butyrophilin, subfamily 3, member A1)           | 613593  |
| BTN3A2      | (butyrophilin, subfamily 3, member A2)           | 613594  |
| CDH11       | (cadherin 11)                                    | 600023  |
| CALD1       | (caldesmon 1)                                    | 114213  |
| CTSC        | (cathepsin C)                                    | 602365  |
| CTSK        | (cathepsin K)                                    | 601105  |
| CTSS        | (cathepsin S)                                    | 116845  |

(Continued)
| Gene Name | Description | Accession Number |
|-----------|-------------|-----------------|
| CD37 (CD37 molecule) | | 151523 |
| CD74 (CD74 molecule) | | 142790 |
| FOXN3 (forkhead box N3) | | 602628 |
| CCR1 (chemokine (C-C motif) receptor 1) | | 601159 |
| CXCX4 (chemokine (C-X-C motif), receptor) | | 162643 |
| CHN1 (chimerin 1) | | 118423 |
| COL1A1 (collagen type I, alpha 1) | | 120150 |
| COL1A2 (collagen type I, alpha 2) | | 120160 |
| COL3A1 (collagen type III, alpha 1) | | 120180 |
| COL4A2 (collagen type IV, alpha 2) | | 120090 |
| COL5A2 (collagen type V, alpha 2) | | 120190 |
| COL6A3 (collagen type VI, alpha 3) | | 120250 |
| COL10A1 (collagen type X, alpha 1) | | 120110 |
| C1R (complement component 1, r subcomponent) | | 613785 |
| C1S (complement component 1, s subcomponent) | | 120580 |
| C2 (complement component 2) | | 613927 |
| CBFB (core-binding factor, beta subunit) | | 121360 |
| CDKN1C (cyclin-dependent kinase inhibitor 1C) | | 600856 |
| DPYS12 (dihydropyrimidinase-like 2) | | 602463 |
| DPYS13 (dihydropyrimidinase-like 3) | | 601168 |
| ENTPD1 (ectonucleoside triphosphate diphosphohydrolase 1) | | 601752 |
| ENPP2 (ectonucleotide pyrophosphatase/phosphodiesterase) | | 601060 |
| EDNRA (endothelin receptor type A) | | 131243 |
| FCER1G (Fc fragment of IgE, high-affinity I, receptor) | | 147139 |
| FBN1 (fibrillin 1) | | 13479 |
| FAP (fibroblast activation protein, alpha) | | 600403 |
| FN1(fibronectin 1) | | 135600 |
| FOLR1 (folate receptor) | | 136430 |
| GSN (gelsolin) | | 137350 |
| GPNMB (glycoprotein (transmembrane) nmb) | | 604368 600 784 |
| G2MK (granzyme K) | | 142995 |
| HLX (H2.0-like homeobox) | | 602179 |
| HSPB2 (heat-shock 27-KD protein 2) | | 603348 |
| HIF1A (hypoxia-inducible factor 1, alpha subunit) | | 188061 |
| THBS2 (thrombospondin 2) | | 188230 |
| THY1 (thy-1 cell surface antigen) | | 602272 |
| TCF4 (transcription factor 4) | | 600818 |
| TAGLN (transgelin) | | 600769 |
| TSPAN8 (tetraspanin 8) | | 602529 |
| TUBA1A (tubulin alpha 1a) | | 601693 |
| UCP2 (uncoupling protein 2 (mitochondrial, proton carrier)) | | 604078 |
| ZNF136 (zinc finger protein 136) | | 147120 |
| IGHG3 (immunoglobulin heavy constant gamma 3 (G3m marker)) | | 147200 |
| IGKC (immunoglobulin kappa constant) | | 192968 |
| ITGA1 (integrin subunit alpha 1) | | 135620 [44] |
| ITGA5 (integrin subunit alpha 5) | | 120980 |
| ITGAM (integrin subunit alpha M) | | 146933 |
| IL10RA (interleukin-10 receptor, alpha) | | 150390 |

(Continued)
Table 4. (Continued)

| Gene Symbol | Gene Name | OMIM ID |
|-------------|-----------|---------|
| LTBP1 | latent transforming growth factor beta binding protein 1 | 600616 |
| LUM | lumican | 109170 |
| LST1 | leukocyte specific transcript 1 | 600978 |
| LTB | lymphotixin beta | 601746 |
| LAPTM5 | lysosomal protein transmembrane 5 | 153455 |
| LOX | lysis oxidase | 153456 |
| LOXL1 | lysis oxidase-like 1 | 143010 |
| HLA-E | major histocompatibility complex, class I, E | 142855 |
| HLA-DMA | major histocompatibility complex, class II, DM alpha | 142856 |
| HLA-DMB | major histocompatibility complex, class II, DM beta | 142857 |
| HLA-DRB1 | major histocompatibility complex, class II, DR beta | 120360 |
| MMP2 | matrix metalloproteinase 2 | 602281 |
| MFGE8 | milk fat globule-EGF factor 8 protein | 600922 |
| MYLK | myosin light chain kinase | 608243 |
| NDN | necdin, MAGE family member | 600613 |
| NBL1 | neuroblastoma, suppression of tumorigenicity | 601488 |
| NCF4 | neutrophil cytosolic factor 4 | 600008 |
| NNMT | nicotinamide N-methyltransferase | 605399 |
| NID2 | nidogen 2 | 608777 |
| POSTN | peristin, osteoblast specific factor | 607332 |
| NREP | neuronal regeneration related protein | 601097 |
| PMP22 | peripheral myelin protein 22 | 153430 |
| LCP1 | lymphocyte cytosolic protein 1 | 173445 |
| PECAM1 | platelet/endothelial cell adhesion molecule | 173490 |
| PDGFR | platelet derived growth factor receptor alpha | 173410 |
| PDGFRA | platelet-derived growth factor receptor, beta polypeptide | 600270 |
| PCOLCE | procollagen C-endopeptidase enhancer | 602194 |
| HTRA1 | HtrA serine peptidase 1 | 176970 |
| PRKCB | protein kinase C beta | 151460 |
| PTPRC | protein tyrosine phosphatase, receptor type, C | 601890 |
| PTK7 | protein tyrosine kinase 7 (inactive) | 609492 |
| RASSF2 | Ras association domain family member 2 | 602189 |
| RGS3 | regulator of G-protein signalling 3 | 182120 |
| SPARC | secreted protein, acidic, cysteine rich | 107400 |
| SERPINA1 | serpin peptidase inhibitor, clade A1 | 106150 |
| AGT | angiotensinogen | 173360 |
| SERPINE1 | Serine (or cysteine) peptidase inhibitor, clade E, member 1 | 313020 |
| SAT1 | spermidine/spermine N1-acetyltransferase | 601737 |
| SMARC3 (SIMP/SNF related, matrix associated, actin dependent regulator of chromatin, subfamily d, member 3) | OMIM ID |
| FGL1 | fibrinogen like 1 | 605776 |
| FUT1 | fucosyltransferase 1 (H blood group) | 211100 |
| PCDHB11 | protocadherin beta 11 | 606337 |
| PCK1 | phosphoenolpyruvate carboxykinase 1 | 614168 |
| TKT | transketolase | 606781 |
| AOX1 | aldehyde oxidase 1 | 602841 |
| TEPI | telomerase associated protein 1 | 601686 |
| BNIP3 | BCL2/adenovirus E1B 19KDa interacting protein 3 | 603293 |

(Continued)
We failed to find published works from PubMed, showing that IFI16, CCNE2 and IGFBP2 are also possibly involved in CF. Therefore, we searched in previously published works some lines of evidence to link these genes to CF.

IFI16 has a possible role in chronic inflammatory autoimmune disorders. The IFI16 protein is normally expressed in nuclei but may be mislocalized in the cytoplasm and secreted. Indeed, significant levels of extracellular IFI16 protein have been identified in the sera of patients with autoimmune diseases [52]. These data provide evidence for a function of IFI16 upon inflammation and tissue damage, as it is also observed in CF.

The human CCNE2 gene encodes a 404-amino-acid protein, related to cyclin E. It is associated with Cdk2 in a functional kinase complex [53] and is involved in the G1 progression [54]. Activated neutrophils elastase is found in high concentrations in the airways of CF patients [55] and has injurious effects on airway epithelial cells. Because the treatment of normal human bronchial epithelial by elastase results in G arrest, due to cyclin E complex inhibition ([56]), we propose that CCNE2 is possibly involved in the pathophysiology of CF.

The insulin-like growth factors (IGF) stimulates growth of multiple cell types. IGF-II access to the cell surface receptor is mediated by IGFBP-2 [57]. It was previously described that serum IGF-II levels were significantly lower in CF patients than in controls, whereas serum IGFBP-2 was significantly higher [58]. IGF-II/IGFBP-2 molar ratios were also found to be significantly lower in CF. Because chronic inflammation is an important modulator of the IGF/IGFBP system in CF and because IGF2 promotes development of pancreatic beta cells, related to some forms of diabetes mellitus, we also propose that IGFBP-2 is a modifier gene in CF.

Candidate genes in CF were selected due to their involvement in CF-related diseases or in similar diseases. Published studies have mostly examined genes involved in immune or inflammatory response (for review: [59]). These genes are: ACE, ADRβ2, ATB, CAPN10, CICN2, DEFβ4, ENaC, FcβRII, GCLC, GSTM1, GSTM3, GSTP1, GSTT1, HFE, HLA-I, HLA-II, HLA-III, HSP70, IFN-γ, IL1-β, IL6, IL10, IL18, KCNJ11, MBL2, MIF, NOS1, NOS3, MASp-2, PPARβ, SERPINA1, SERPINA3, SFTPA2, TLR4, TGFBI, TNFα, TNFα-receptor, and TNFβ [59]. Among the published modifier genes [59], we found using databases that TGFBI, SERPINA1, TNFα-receptor, GSTP1, ADRβ2, GSTM1, GSTT1 and TNFα are also involved in pancreatic and liver diseases (Figs 3 and 4). Among candidate gene modifiers of clinical phenotypes in CF, we also retrieved MBL2, EDNRA, TGFBI, IFRD1, IL8, MSRA,
Table 5. List of the genes retrieved from SE40445, GSE48452 and GSE44314. Genes that were found in at least two groups are underlined.

| Genes obtained from SE40445 |
|----------------------------|
| BANF1, EPHX1, CAB39, C2orf17, PLD3, NUCB2, SOC, PDXP, FRMD4B, ETS2, C2orf98, SLCE3A1, TP53AP1, TJP1, TCEA2, AES, TETRAN, LOC348262, IFRD1, DSCR1L2, C1orf32, GCMS306, GCG20806, C1orf33, LNNEP, PHACTR2, CLDN1, RHBDL7, INSIG1, K1AA1847, ATP6V0A4, PTDS25, DKFPZ761D0211, TF11, MRPL40, TNRC5, MAOB, K1AA1434, COBR1A1, APBA2BP, GNA11, DKFPZp4341610, ARL8, SDK1, GSDMDC1, IFNGR2, TCTA, EP300, ARFGAP3, TOMM34, TTC13, HINT2, RPS6KB1, FEM1A, ERAL1, NUDT1, CCND3, RAB2A, MYCPBP, SEC14L3, CDR2, KIAA1033, RBM21, ARID1A, TTC14, ZADH1, DNCL2, VARS2, HSO2C6B, TM3C, ENPP5, PBP, DREV1, C14orf31, TCEG1, CPNE2, LOC203069, B3GAT3, GSTM1, DUSP10, MB, RNF152, HSCARG, AP1G1, FLJ32096, C12orf8, LOC124446, DCTN1, AGP16L, DGCR6, LOC144438, LOC90378, CUDCE2, ARRDC3, ZNF562, WDR13, EFA1, C2orf24, MECT1, FLNC, G6PC3, GRCC9, CHFR, KLF3, DKFPZ5640243, ZBTB34, POLK, ENTPD4, TM4C, IFI16, PYCR2, PEX14, C6orf18, SH3GL1B |
| Genes obtained from GSE48452 |
| LYZ, CDH23, SFMBT2, LOC654342/LOC654166, PECAM1, HLA-DRA, ZFPI, TL1, KRR1/QPR1, ARHGAP31, GPSC5, DPYS2L, GRT1P, GMFG, IFI16, Pap4K2A, ACSS1, MYOS1, SXBP1, PLA2G4A, H2AFY2, SYK, UROG1, SEMA3D, AMICA1, MSN, LAMA4, ZEB2, DFHC1, TMEM144, INS-IGF2, EPH3A, ENG, OGRFL1, SYP, F2R, AB, GRPNMB, GCDH, PIEZ20, SLC38A1, CSF2RA, TFAP2B, ABAT, CCND2, GCAT, NIN, FRZB, COL14A1, CA14, GUCY1A3, MRAS, NCF2, SULF2, ITGA9, HLA-DMA, PLXDC2, MYOF, TPM4, LOC100130691, FOX12, SAMD9L, SCPEP1, SEL1L3, IQGAP1, APOBEC3C, TMSB10, ALOX5, ALKBH7, ENTPD1, PODXL, CD74, IGK/IGK3-20/IGHV3-23, LCP1, SLC6A6, SUN2, FABP4, TACO1, PPM1K, PTGS1, ZNF532, CLEC10A, CPLX1, CNTNAP3B, CNTNAP2, AADAT, H2AFY2, FCGRC2/FCGRC8, DOCK11, MAPRE1, DOCK10, CACHD1, YBX3, IGKC, FABP5, SPARC, RASGRP1, GLS2, GUCY1B3, KLF15, FLRT2, FCGRC2, TME1M120A, LSP1, HIST2H2F/B, HIST2H2A, SLC31A4, PIK3R3, AHCY, EBP1, ADAMTS2L, ACVR1, MECOM, ELMO2, FANC, HEG1, RNASE1, ARL3, ACKR2, SARDH, SLIT2, PTPRCAP, MTHFD1L, GBQ, ACER3, PDGFRB, PMP2, SPP1, ASAP1, CMTM7, FMNL3, CD37, LGALS3, IGHG1/IGHV3-48, FAM20C, COL3A1, PNPT1, HLA-DPB1, SSBP2, MTHFD2, IGH1/IGHD, UTRN, CENPV, HDC7, HSD17B11, RAB31, SCFV/IGHV4-31/IGHV4-30, MAMLL2, PKC2, SNX29, RASAL2, RHOC, CD48, NAV1, MFSD6, MAFAP4, GYS2, WDR87, HLA-DRB4/HLA-DRB1, CD44, HIST1H1B, UG0898H09, GPR124, DZIP1/CLDN10, DPT, PHKX, SLC39A10, COL12A1, SCR1N, ARMCX6, GPR174, ITGA5, LAIR1, JAM2, IMP3, DOCK8, SPAN3, GSK, PKM, IGHG4/IHG32, IGF1, GLT8D2, HLA-DPA1, FOCAD, SUSD1, HIST3H1B, PIGV, STK39, MGMT, PREX1, PTPRM, TCF4, KRTAP4-7, KRTAP4-9/KRTAP4-8, KRTAP4-6, FAM65B, AMDHD1, FAM8A1, CRYBG3, MIMA, DOCK2, NOD1, MMP22, LOC100134868, TMEM165, ANKRD36/ANKRD36B, SORT1, FAP1, INPP5D, ACOT12, PTPLAD2, VIM, PTPN13, IKBKBE |

https://doi.org/10.1371/journal.pone.0173822.t005
ADIPOR2, TCF7L2 and SERPINA1 (for review: [60]). When compared to our database search, we found that TGFβ1 was present in OMIM and CTD and that IFRD1 and SERPINA1 were only present in OMIM, indicating that they are also possibly involved in pancreatic and liver diseases. IFRD1 was also found in our CF-vs-non CF re-analysis of SE40445 (Table 6).

In conclusion, we propose here that IFI16, CCNE2 and IGFBP2 are new candidates of modifier genes in CF.

Table 6. Genes involved in CF, Chronic Pancreatitis and Liver Disease. Common genes present in our re-analyzed data of the GEO database and in our bibliographic analysis. Common genes in CF and in pancreatic and in liver affections are underlined.

| GEO Analysis | Differentially expressed genes in CF | Differentially expressed genes in Chronic Pancreatitis | Differentially expressed genes in Liver Disease |
|--------------|--------------------------------------|-------------------------------------------------------|-----------------------------------------------|
| CF-vs-non CF re-analysis of SE40445 | IFRD1* [32–35, 37] IFI16 [32–35, 37] CCNE2 [32] IGFBP2 [33, 35] | EPHX1 [38] | SLC33A1 [41] |
| Diabetes-vs-healthy re-analysis of GSE44314 | HLA-DRB1** [32–35, 37] DSP [32–35, 37] HLA-DQA1 [34] | HLA-DRB1** [43, 44] CTSA [43] PKM [43] | SLC25A36 [41] TGFβ1*** [40] |
| NASH-vs-healthy re-analysis of GSE48452 | HLA-DRA [32–35, 37] GPNMB [32–35, 37] NCF2 [32–35, 37] RASGRP1 [32–35, 37] LGALS3 [32–35, 37] PTPN13 [32–35, 37] | LYZ [43] PECAM1 [43, 44] HLA-DRA [43] DPYSL2 [44] GPNMB [43, 44] HLA-DMA [44] IQGAP1 [43] ENTPD1 [44] CD74 [43, 44] LCP1 [44] IGKC [44] SPARC [43, 44] GUCY1B3 [43] PDGFRB [44] CD37 [44] COL3A1 [43, 44] HLA-DPB1 [43] GSN*** [43, 44] MGMT [43] TOF4 [44] PMP22 [43, 44] | STXB1 [39] ABR [40] SLC6A6 [41] SLC39A10 [41] |

Underlined: Genes that were found in the bibliographic analysis for CF and Chronic Panreatitis genes.
*: also found in our OMIM search and common within CF, CF associated genes and Liver disease keywords
**: also found in our CTD search and common within Pancreatic Diseases and Liver Diseases keywords
***: also found in our OMIM search and is common within CF, CF associated genes, Liver disease and PI keywords and in CTD search for CF, Pancreatic Diseases and Liver Disease keywords
****: also found in common our OMIM search and common within CF, CF associated genes keywords

https://doi.org/10.1371/journal.pone.0173822.t006
carcinoma [62]. Regarding CF, a low activity of EPHX1 was found to be a risk factor for chronic obstructive pulmonary disease (COPD) in Caucasian [63]. Genetic polymorphisms of GSTP1 and EPHX1 were shown to correlate with oxidative stress markers and lung function in COPD [64]. It has to be noticed that GSTP1 belongs to our list of common genes for PI, CF and Liver Disease search In HuGE Navigator. Mutations in EPHX1 were also suggested to modify the severity of respiratory disorders in CF [65].

SLC33A1 (AT1) is a transporter of the ER membrane permitting the entry of acetyl-CoA into the ER lumen. Its down-regulation induces autophagy and cell death [66]. CFTR needs N-linked glycosylation. Its altered glycosylation leads to a misfolded protein such as F508del-CFTR which is degraded [67]. The accumulation of this incorrectly folded protein in the ER, triggers the Unfolded Protein Response (UPR; [68, 69]). In CF, excessive UPR may be alleviated by the inhibition of ATF6 which possesses an endogenous inhibitor: XBP1 [69, 45] which controls ERAD by activating the expression of SLC33A1 [70]. Therefore, we propose here that SLC33A1 is likely a strong modulator in CF.

Differentially expressed genes in the bibliographic search for CF and in the Diabetes-vs-healthy re-analysis of GSE44314 were the Major Histocompatibility Complex Class II genes DRB1 and DQA1 (OMIM Accession ID: 142857 and 146880, respectively) and the Desmoplakin gene (DSP, OMIM Accession ID: 125647) (Table 6). HLA-DQA1 and HLA-DQB1 are likely genetic markers of the Celiac Disease [71] which was suggested to be a risk factor in CF [72, 73]. Therefore, together with our present results, we propose HLA-DQA1 and HLA-DQB1 genes, as new modifiers in CF. Desmoplakin is a cytoskeletal linker protein conferring the structural integrity of tissues by linking intermediate filament cytoskeleton from cell to cell (for review: [74]). In remodeled airways, as in CF, abnormalities of the cytokeratins and desmoplakins are observed [75]. According to the importance of intermediate filaments and cell-to-cell communications in CF [76, 77], DSP gene expression is likely involved.

Differentially expressed genes in the bibliographic search for CF and in NASH-vs-healthy re-analysis of GSE48452 were: HLA-DRA (OMIM Accession ID: 142860), GPNMB (Glycoprotein NMB, OMIM Accession ID: 604368), NCF2 (Neutrophil Cytosolic Factor 2, OMIM Accession ID: 608515), RASGRP1 (RAS Guanyl Nucleotide-Releasing Protein 1, OMIM Accession ID: 603962), LGALS3 (Lectin, Galactosidase-Binding, Soluble 3, OMIM Accession ID: 153619), PTPN13 (Protein-Tyrosine Phosphatase, Nonreceptor-Type 13; OMIM Accession ID: 600267) (Table 6).

Involvement of Class II major histocompatibility complex (MHC) genes in CF were discussed above. Because HLA Class II Polymorphism is already known to be a modifier of the pulmonary phenotype in CF [78, 79] our finding is not new.

GPNMB is a transmembrane glycoprotein expressed in the heart, lung, and small intestine [80]. The GPNMB gene is a p53- and androgen-dysregulated gene with antiproliferative and antitumorigenic effects in prostate [81]. To our knowledge, its role in CF has never been studied.

Chronic granulomatous disease (CGD) is the most common phagocyte immunodeficiency due to mutations in a gene encoding NADPH oxidase in phagocytes. As in CF, CGD is characterized by recurrent infections and inflammation in the lungs. Because patients with mutated NCF2 gene present severe infections and CGD [82, 83] NCF2 gene is a good candidate for modifying CF.

RASGRP1 plays a critical role in T cell receptor (TCR) stimulation and is essential in the Natural Killer (NK) cell activation and in the immune response [84, 85]. The NK cells are of particular importance in CF as they play a central role in clearing P. aeruginosa from the lung [86, 87].

In silico search for modifier genes in Cystic Fibrosis
LGALS3 gene encodes a β-galactoside-binding lectin which plays a role in apoptosis and innate immunity. Because LGALS3 corrects F508del-CFTR trafficking [88], its gene is a good candidate as a modifier in CF.

The protein encoded by PTPN13 is a tyrosine phosphatase with five PDZ (postsynaptic density protein 95/discs large/zonula occludens-1) domains that bind to plasma membrane and cytoskeleton. The C terminus of CFTR binds various PDZ domain-containing proteins including EBP50, CAL and members of the GRASP family [89]. Despite it was suggested a possible role of reduced PDZ interactions in the accelerated internalization of F508del-CFTR [90], the role of CFTR-PDZ interactions remains incompletely resolved [88]. The PTPN13 gene is therefore of interest and the encoded protein may bind and regulate CFTR.

New potential modifier genes

In conclusion, we found three groups of potential modifier genes. From our bibliographic analysis and microarray studies, we propose here that IFI16 (NCBI ID: 3428), CCNE2 (NCBI ID: 9134) and IGFBP2 (NCBI ID: 3485) are new modifiers candidates in CF. Common genes in CF and Chronic Pancreatitis are EPHX1 (NCBI ID: 2052), HLA-DQA1 (NCBI ID: 3117) and -DQB1 (NCBI ID: 3119) and DSP (NCBI ID: 1832). Modifier genes in CF, in link to Liver Disease, are SLC33A1 (NCBI ID: 9197), GPNMB (NCBI ID: 10457), NCF2 (NCBI ID: 4688), RASGRP1 (NCBI ID: 10125), LGALS3 (NCBI ID: 3958) and PTPN13 (NCBI ID: 5783).

Gene-pathway association of the new potential modifier genes

To better understand the biological function of our new potential modifier genes, we search for their annotated pathways, which are associations established by KEGG [91] and REACTOME [92]. Biological pathways describe biological processes. Therefore, they can be used to integrate and visualize gene products in different conditions (healthy vs disease). KEGG and REACTOME pathway data show known molecular interactions and reaction networks. These data, integrating genes and diseases provide insights into molecular networks involving our depicted candidate genes. Indeed, pathway enrichment analysis permits to decipher biological functions associated with these genes [93]. First, the three common genes in CF (IFI16, CCNE2 and IGFBP2), which were retrieved from Table 6, were submitted to the pathway search, using VennViewer in CTD. 1, 1 and 9 pathways were retrieved for IFI16, IGFBP2 and CCNE2, respectively (Fig 6). No pathway was common between those genes. The 4 common genes in CF and Chronic Pancreatitis and the 6 common genes in CF and Liver Disease were further submitted to the pathway search (Table 7). The only common pathway between CF, Chronic Pancreatitis and Liver Disease was “Immune System” (REACTOME: 6900). The implicated genes were IFI16 for CF, HLA-DQB1 for CF plus Chronic Pancreatitis and LGALS3, NCF2 and RASGRP1 for CF plus Liver Disease (Table 7). Therefore, genes of the Immune System are likely involved in CF and in its association with Chronic Pancreatitis and Liver Disease.

We further search for the diseases that are statistically enriched among our genes. Modifiers candidates in CF (IFI16, CCNE2 and IGFBP2), the common genes in CF and Chronic Pancreatitis (EPHX1, HLA-DQA1 -DQB1, DSP) and in CF in link to Liver Disease (SLC33A1, GPNMB, NCF2, RASGRP1, LGALS3 and PTPN13) were entered as a single data set in CTD. According to CTD, a disease was considered enriched if the proportion of genes annotated to it in our test set was significantly larger than the proportion of all genes annotated to it in the genome. The results are presented Table 8.
Biological function and PPI networks association of the new potential modifier genes

For a full description of the proteins which are encoded by our candidate modifiers genes, we search for their potential interactions using the recent 10.0 version of STRING [94]. We used the basic interaction unit in STRING for functional association between proteins, derived from known experimental data. Given the importance of interactions in protein function, we search for groups of interacting proteins into functional sets, in physical complexes.

We first search for high confidence (≥ 0.7) interactions involving the proteins encoded by IFI16, CCNE2 and IGFBP2 (CF candidates). We added the CFTR protein to the query. As shown in Fig 7A, a network of 9 nodes and 10 edges was obtained (clustering coefficient: 0.744). Whereas IGFBP2 did not belong to the network, we observed that CFTR and IFI16 have a common partner: Ubiquitin C (UBC, OMIM: 191340). IFI16 and UBC are involved in the innate immune signaling pathways [95]. Furthermore, Ubiquitin modification is required for the proteasomal targeting of CFTR, which is ubiquitinated in the ER during assembly and during recycling from the cell surface [96]. Therefore, modifications in the IFI16 and/or UBC gene expression could lead to modifications of the maturation of CFTR, showing the importance of IFI16 as a modifier gene in CF, in link to immunity. CCNE2 binds to CDK2 in a catalytically active complex which modulates the cell cycle progression [97] and CDK inhibition corrects F508del-CFTR proteins [98]. Together with the role of UBC upon the proteasomal targeting of CFTR, the physical and functional complexes involving CCNE2, UBC and CFTR (Fig 7A) indicates that CCNE2 is likely a modifier gene in CF.

Fig 6. Gene-pathway association of the new potential modifier genes in KEGG and REACTOME. The 3 common genes in CF, which were retrieved from Table 6, were submitted to the pathway search, using VennViewer in CTD. 1, 1 and 9 pathways were retrieved for IFI16, IGFBP2 and CCNE2, respectively. No pathway was common between those genes.

https://doi.org/10.1371/journal.pone.0173822.g006
Table 7. Associated pathways with the retrieved potential modifiers in CF, in CF and Chronic Pancreatitis and in CF and Liver Disease. Common pathways are underlined.

| Genes name | Pathway | Pathway ID |
|------------|---------|------------|
| **Genes in CF (n = 3)** | | |
| CCNE2 | Cell cycle | KEGG:04110 |
| | Cell cycle | REACT:115566 |
| | Cellular responses to stress | REACT:120956 |
| | Measles | KEGG:05162 |
| | Oocyte meiosis | KEGG:04114 |
| | p53 signaling pathway | KEGG:04115 |
| | Pathways in cancer | KEGG:05200 |
| | Prostate cancer | KEGG:05215 |
| | Small cell lung cancer | KEGG:05222 |
| IFI16 | Immune System | REACT:6900 |
| GFBP2 | Metabolism of proteins | REACT:17015 |
| **Genes in CF and in Chronic Pancreatitis** | | |
| DSP | Apoptosis | REACT:578 |
| EPHX1 | Arrhythmogenic right ventricular cardiomyopathy | KEGG:05412 |
| HLA-DQA1 | Bile secretion | KEGG:04976 |
| | Metabolism of xenobiotics by cytochrome P450 | KEGG:00980 |
| | Allograft rejection | KEGG:05330 |
| | Antigen processing and presentation | KEGG:04612 |
| | Asthma | KEGG:05310 |
| | Autoimmune thyroid disease | KEGG:05320 |
| | Cell adhesion molecules (CAMs) | KEGG:04514 |
| | Graft-versus-host disease | KEGG:05332 |
| HLA-DQB1 | Immune System | REACT:6900 |
| | Intestinal immune network for IgA | KEGG:04672 |
| | Leishmaniasis | KEGG:05140 |
| | Phagosome | KEGG:04145 |
| | Rheumatoid arthritis | KEGG:05323 |
| | Staphylococcus aureus infection | KEGG:05150 |
| | Systemic lupus erythematosus | KEGG:05322 |
| | Toxoplasmosis | KEGG:05145 |
| | Tuberculosis | KEGG:05152 |
| | Type I diabetes mellitus | KEGG:04940 |
| | Viral myocarditis | KEGG:05416 |
| **Genes in CF and in Liver Disease** | | |
| GPNMB | None | |
| LGALS3 | Immune System | REACT:6900 |
| NCF2 | Immune System | REACT:6900 |
| | Leishmaniasis | KEGG:05140 |
| | Leukocyte transendothelial migration | KEGG:04670 |
| | Osteoclast differentiation | KEGG:04380 |
| | Phagosome | KEGG:04145 |
| | Signal Transduction | REACT:111102 |
| PTPN13 | None | |
| RASGRP1 | Immune System | REACT:6900 |
| | MAPK signaling pathway | KEGG:04010 |
| | Signal Transduction | REACT:111102 |
| | T cell receptor signaling pathway | KEGG:04660 |
| | Glycosphingolipid biosynthesis | KEGG:00604 |
| | Metabolic pathways | KEGG:01100 |
| SLC33A1 | Transmembrane transport of small molecules | REACT:15518 |

https://doi.org/10.1371/journal.pone.0173822.t007
Interactions involving the proteins encoded by common genes in CF and Chronic Pancreatitis (EPHX1, HLA-DQA1·DQB1, DSP and CFTR) were searched (Fig 7B). Whereas, HLA-DQA1 and HLA-DQB1 did not belong to any complex, EPHX1, DSP and CFTR were linked by UBC. The Ubiquitin-proteasome system (UPS) is an important regulator for the intracellular

### Table 8. Associated diseases with the retrieved potential modifiers in CF, in CF and Chronic Pancreatitis and in CF and Liver Disease.

The list of genes was submitted to CTD search and associated diseases were retrieved. In bold: diseases which may present relevant involvement in CF physiopathology.

| Disease Name                                      | Disease ID | P-value   | Corrected P-value | Annotated Genes (input)                                      | Genome Frequency |
|---------------------------------------------------|------------|-----------|--------------------|-------------------------------------------------------------|------------------|
| Skin and Connective Tissue Diseases               | D017437    | 3.24E-09  | 8.29E-07           | DSP|EPHX1|GPNM|HLA-DQA1|HLA-DQB1|IGFBP2|NCF2|RASGRP1 | 1499/41292 (3.63%) |
| Female Urogenital Diseases and Pregnancy Complications | D005261    | 5.68E-08  | 1.45E-05           | CCNE2|EPHX1|GPNM|HLA-DQA1|HLA-DQB1|LGALS3|RASGRP1 | 1350/41292 (3.27%) |
| Esophageal Achalasia                              | D004931    | 2.74E-07  | 7.03E-05           | HLA-DQA1|HLA-DQB1 | 3/41292 (0.01%) |
| Skin Diseases(a)                                  | D012871    | 7.73E-07  | 1.98E-04           | DSP|EPHX1|GPNM|HLA-DQA1|HLA-DQB1|IGFBP2 | 1179/41292 (2.86%) |
| Female Urogenital Diseases                        | D052776    | 9.53E-07  | 2.44E-04           | CCNE2|GPNM|HLA-DQA1|HLA-DQB1|LGALS3|RASGRP1 | 1222/41292 (2.96%) |
| Esophageal Motility Disorders                      | D015154    | 3.29E-06  | 8.42E-04           | HLA-DQA1|HLA-DQB1 | 9/41292 (0.02%) |
| Lung Diseases, Obstructive(b)                     | D008173    | 5.74E-06  | 0.00147            | EPHX1|HLA-DQA1|HLA-DQB1 | 114/41292 (0.28%) |
| Deglutition Disorders                              | D003680    | 7.12E-06  | 0.00182            | HLA-DQA1|HLA-DQB1 | 13/41292 (0.03%) |
| Nephritis                                         | D009393    | 7.39E-06  | 0.00189            | HLA-DQA1|HLA-DQB1|LGALS3 | 124/41292 (0.30%) |
| Connective Tissue Diseases                        | D003240    | 1.35E-05  | 0.00345            | HLA-DQA1|HLA-DQB1|NCF2|RASGRP1 | 496/41292 (1.20%) |
| Immune System Diseases                            | D007154    | 1.95E-05  | 0.005              | EPHX1|HLA-DQA1|HLA-DQB1|NCF2|RASGRP1 | 1173/41292 (2.84%) |
| Autoimmune Diseases                               | D001327    | 2.08E-05  | 0.00633            | HLA-DQA1|HLA-DQB1|NCF2|RASGRP1 | 554/41292 (1.34%) |
| Esophageal Diseases                                | D004935    | 2.15E-05  | 0.00649            | HLA-DQA1|HLA-DQB1|LGALS3 | 177/41292 (0.43%) |
| Male Urogenital Diseases(c)                       | D052801    | 2.61E-05  | 0.00668            | EPHX1|GPNM|HLA-DQA1|HLA-DQB1|LGALS3 | 1246/41292 (3.02%) |
| Kidney Diseases                                   | D007674    | 2.82E-05  | 0.00723            | GPNM|HLA-DQA1|HLA-DQB1|LGALS3 | 599/41292 (1.45%) |
| Digestive System Diseases(d)                      | D004066    | 3.18E-05  | 0.00813            | EPHX1|GPNM|HLA-DQA1|HLA-DQB1|IGFBP2|LGALS3 | 2247/41292 (5.44%) |
| Celiac Disease(e)                                  | D002446    | 3.44E-05  | 0.00881            | HLA-DQA1|HLA-DQB1 | 28/41292 (0.07%) |
| Pathological Conditions                           | D013568    | 3.45E-05  | 0.00883            | DSP|EPHX1|HLA-DQA1|IGFBP2|LGALS3|SLC33A1 | 2280/41292 (5.52%) |
| Drug-Induced Liver Injury                         | D056486    | 3.62E-05  | 0.00927            | EPHX1|HLA-DQB1|LGALS3 | 211/41292 (0.51%) |

(a): including Sweat Gland Diseases;  
(b): including Chronic Bronchitis, COPD, Mucus Insipissation of Respiratory Tract;  
(c): including Congenital bilateral aplasia of vas deferens;  
(d): including Biliary Tract Diseases, Digestive System Abnormalities, Liver Diseases and Pancreatic Diseases;  
(e): including Malabsorption Syndromes  
(1): Bonferroni-corrected p-value;  
(2): Genotype frequency in a population is the number of individuals with a given genotype divided by the total number of individuals in the population.

https://doi.org/10.1371/journal.pone.0173822.t008

Interactions involving the proteins encoded by common genes in CF and Chronic Pancreatitis (EPHX1, HLA-DQA1·DQB1, DSP and CFTR) were searched (Fig 7B). Whereas, HLA-DQA1 and HLA-DQB1 did not belong to any complex, EPHX1, DSP and CFTR were linked by UBC. The Ubiquitin-proteasome system (UPS) is an important regulator for the intracellular...
trafficking of proteins and a UPS-dependent stabilization of cell-cell contacts involving DSP was shown [99]. Because a low activity of EPHX1 is a risk factor for COPD in Caucasian [63] and because an increased UPS activity is observed in COPD [100], a functional link between EPHX1 and UBC is likely involved in CF and Chronic Pancreatitis.

Fig 7C shows the networks formed by the proteins encoded by the genes found in CF in link to Liver Disease (SLC33A1, GPNMB, NCF2, RASGRP1, LGALS3 and PTPN13). SLC33A1 forms a protein complex together with CFTR, with UBC as an intermediate. As mentioned above, IRE1/XBP1 controls ERAD by activating the expression of...
SLC33A1 [70]. UBC is also involved in ERAD which is a key element of the CFTR degradation. This reinforces our proposition that SLC33A1 is a strong modulator in CF.

Finally, a network was drawn using the proteins encoded by the genes involved in CF, in CF plus Chronic Pancreatitis and in CF plus Liver Disease, which were found in a network linked to the CFTR protein (Fig 7D). UBC was observed as a central node. It is well known that F508del-CFTR protein degradation involves Ubiquitin modification. It was previously shown that the de-ubiquitinating enzyme, Ubiquitin C-terminal hydrolase-L1 (UCH-L1), is highly expressed in CF airway epithelial cells and that there is a positive correlation between UCH-L1 expression and steady state levels of Wt-CFTR protein and F508del-CFTR [101]. Although it is not sufficient to rescue F508del-CFTR, the effect of UCH-L1 upon CFTR processing shows its potential roles in CF. UBC is thus a key component in CF.

Conclusion

CF is characterized by progressive lung disease with inflammation, pancreatic exocrine insufficiency and fatty liver disease. Liver disease is found in two-thirds of sick childrens [102–104]. Inflammation is indeed a hallmark of CF and is characterized by bacterial infection, neutrophils infiltrations in the lung and high levels of cytokines. Whereas, inflammatory response is deregulated and excessive, it is not known whether abnormal CFTR is responsible or a consequence. The most common hypothesis is that a defective CFTR leads to a decreased airway surface liquid which fails to clear infected secretions from the lung, triggering the excessive inflammatory response. It is therefore still controversial whether this hyper-inflammation is solely the result of the chronic infection or is a primary event due to CFTR defects [105].

Regarding liver, three types of Liver Disease are observed in CF patients: steatosis, cirrhosis and biliary fibrosis. Fatty infiltration of the liver being found in 30% at biopsy and in 60% at the autopsy of the CF patients, is the most common hepatic alteration [106]. In CF, PI phenotype is a hallmark in patients with two severe alleles such as F508del [16] and CF Related Diabetes (CFRD) is observed in more than 50% of patients. Whereas CFRD is a unique entity, it exhibits common features of both type 1 and 2 diabetes. However, the hallmark of CFRD is insulin deficiency, as in type 1 diabetes in which pancreatic beta cells are destroyed [107].

Despite residual involvement of CFTR’s sequence variation upon lung function, modifiers are of main importance. Indeed, it is clear that differences in the genotype in CF are not solely responsible for the disease variability [108] and large-scale genome-wide association studies (GWAS) were undertaken to search for genetic determinants of phenotypic variation (for review, [23]). Nevertheless, further studies are still needed. Therefore, our aim was to depict new candidate genes with modifying activity in CF, in association with liver and pancreatic diseases. Using databases, bibliographic mining and gene expression analysis we found some candidate genes specific to lung, liver and pancreas. Nevertheless, two groups of different genes could be built, depending on the way they were retrieved. Unsurprisingly, databases search led to known genes involved in CF. Therefore, we focused on genes obtained by microarray results re-analysis, compared to those from our bibliographic analysis. We highlighted 3 genes specific to CF, 4 genes specific to CF and Chronic Pancreatitis and 6 genes for CF and Liver Disease. Nevertheless, we failed to find common genes in CF, Chronic Pancreatitis and Liver Disease. Pathways associated with these retrieved potential modifiers indicated that immune system is likely involved and that Ubiquitin C is the central node linking CF to liver and pancreatic disease. A schematic representation of the results is proposed in Fig 8. Despite we propose here some new genes involved in CF, we are fully aware that some other modifier genes may be missing and that our in silico analysis requires functional analysis to give our results a physiological relevance. Indeed, our analysis may present some limitations due to the
statistical combination of results from several studies which may present variable quality and heterogeneity of the used tools. This could lead to misleading results, inherent to any systematic meta-analyses which may be non-exhaustive or over interpreted. Database and literature analysis can also be limiting because of incomplete knowledge base leading to byproduct results. Because in the manual curation of the literature we have not performed an exhaustive

Fig 8. Schematic representation of the results. Candidate genes are shown in each organ. The corresponding proteins presenting PPI are linked by a red line.

https://doi.org/10.1371/journal.pone.0173822.g008
review of all the genes and because the used studies have been performed using different methodologies and have different power, no statistical assessment was possible here. Finally, gene expression profiles may be affected by the variability between the used cell models, their intrinsic properties and their type of statistical analysis.

In conclusion, we propose here a methodology that may be used for some other genetic disease with great variability and identify new candidate modifier genes in CF, related to lung, pancreas and liver, such as Ubiquitin C.

Author Contributions

Conceptualization: PT.

Data curation: PT.

Formal analysis: PT.

Investigation: PT.

Methodology: PT.

Project administration: PT.

Resources: PT.

Software: PT.

Supervision: PT EG CF.

Validation: PT EG CF.

Visualization: PT.

Writing – original draft: PT.

Writing – review & editing: PT EG CF.

References

1. Riordan JR, Rommens JM, Kerem BS, Alon N, Rozmahel R, Grzelczak Z, et al. Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA. Science 1989; 245: 1066–1073. PMID: 2475911

2. Tsui LC, Dorfman R. The cystic fibrosis gene: a molecular genetic perspective. Cold Spring Harb Perspect Med. 2013; 3(2):a009472. https://doi.org/10.1101/cshperspect.a009472 PMID: 23378595

3. Welsh MJ, Tsui LC, Boat TF, Beaudet AL in: Scriver C.R., Beaudet A.L., Sly W.S. and Valle D. In The Metabolic and Molecular Bases of Inherited Disease ( 7th edition., McGraw-Hill, New York), 1995.pp. 3799–3876.

4. Drumm ML, Pope HA, Cliff WH, Rommens JM, Marvin SA, Tsui L, et al. Correction of the cystic fibrosis defect in vitro by retrovirus-mediator gene transfer. Cell 1990; 62:1227–1233. PMID: 1698126

5. Rich DP, Anderson MP, Gregory RJ, Cheng SH, Paul S, Jefferson DM, et al. Expression of the cystic fibrosis transmembrane conductance regulator corrects defective chloride channel regulation in cystic fibrosis airway epithelial cells. Nature 1990; 347:358–363. https://doi.org/10.1038/347358a0 PMID: 1699126

6. Szellas T, Nagel G. Apparent affinity of CFTR for ATP is increased by continous kinase activity. FEBS Letters 2003; 538:141–146. PMID: 12580093

7. Treharne KJ, Xu Z, Chen JH, Best OG, Cassidy DM, Gruenert DC et al. Inhibition of protein kinase CK2 closes the CFTR Cl– channel, but has no effect on the cystic fibrosis mutant DeltaF508-CFTR. Cell Physiol Biochem. 2009; 24:347–360. https://doi.org/10.1159/000257427 PMID: 19910675

8. Howell LD, Borchardt R, Kole J, Kaz AM, Randak C Cohn JA. Protein kinase A regulates ATP hydrolysis and dimerization by a CFTR domain. Biochem. J. 2004; 378:151–159. https://doi.org/10.1042/BJ20021428 PMID: 14602407
9. Bradbury NA. Intracellular CFTR: localization and function. Physiol Rev. 1999; 79(Suppl 1):S175–S191.

10. Bertrand CA, Frizzell RA. The role of regulated CFTR trafficking in epithelial secretion. Am J Physiol Cell Physiol. 2003; 285(1):C1–C18. https://doi.org/10.1152/ajpcell.00554.2002 PMID: 12777252

11. Ward CL, Omura S, Kopito RR. Degradation of CFTR by the ubiquitin-proteasome pathway. Cell. 1995; 83(1):121–7. PMID: 7553863

12. Haws CM, Nepomuceno IB, Krouse ME, Wakelee H, Law T, Xia Y, et al. Delta F508-CFTR channels: kinetics, activation by forskolin, and potentiation by xanthines. Am J Physiol. 1996; 270(5 Pt 1):C1544–55. PMID: 8967457

13. Schultz BD, Frizzell RA, Bridges RJ. Rescue of dysfunctional deltaF508-CFTR chloride channel activity by IBMX. J Membr Biol. 1999; 170(1):51–66. PMID: 10398760

14. Mekus F, Ballmann M, Bronsveld I, Bijman J, Veeze H, Tummler B. Categories of deltaF508 homozygous cystic fibrosis twin and sibling pairs with distinct phenotypic characteristics. Twin Res. 2000; 3(4):277–293. PMID: 11463149

15. Ferrari M, Cremonesi L. Genotype-phenotype correlation in cystic fibrosis patients. Ann Biol Clin (Paris) 1996; 54(6):235–241.

16. Kristidis P, Bozon D, Corey M, Markiewicz D, Wakelee H, Law T, Xia Y, et al. Delta F508-CFTR channels: kinetics, activation by forskolin, and potentiation by xanthines. Am J Physiol. 1996; 270(5 Pt 1):C1544–55. PMID: 8967457

17. Kerem E, Corey M, Kerem BS, et al. The relation between genotype and phenotype in cystic fibrosis—analysis of the most common mutation (delta F508) N Engl J Med. 1990; 323(22):1517–1522. https://doi.org/10.1056/NEJM199011293232203 PMID: 2233932

18. Hubert D, Bienvenu T, Desmazes-Dufeu N, et al. Genotype-phenotype relationships in a cohort of adult cystic fibrosis patients. Eur Respir J. 1996; 9(11):2207–2214. PMID: 8947061

19. Lipner EM, Garcia BJ, Strong M. Network Analysis of Human Genes Influencing Susceptibility to Mycobacterial Infections. PLoS One. 2016; 11(1):e0146585. https://doi.org/10.1371/journal.pone.0146585 PMID: 26751573

20. Hamosh A. Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders. Nucleic Acids Research. 2004; 33:D514–D517.

21. Amberger JS, Bocchini CA, Schiettecatte F, Scott AF, Hamosh A. OMIM.org: Online Mendelian Inheritance in Man (OMIM), an online catalog of human genes and genetic disorders. Nucleic Acids Research. 2015; 43:D789–D798. https://doi.org/10.1093/nar/gku1205 PMID: 25428349

22. Davis AP, Murphy CG, Johnson R, Lay JM, Lennon-Hopkins K, Saraceni-Richards C, et al. The Comparative Toxicogenomics Database: update 2013. Nucleic Acids Research. 2013; 41:D1104–1114. https://doi.org/10.1093/nar/gks994 PMID: 23093600

23. Yu W, Gwinn M, Clyne M, Yesupriya A, Khoury MJ. A navigator for human genome epidemiology. Nat Genet. 2008; 40(2):124–5. https://doi.org/10.1038/ng0208-124 PMID: 18227866

24. Yesupriya A, Evangelou E, Kavvoura FK, Patspououlos NA, Clyne M, Walsh MC, et al. Reporting of human genome epidemiology (HuGE) association studies: an empirical assessment. BMC Med Res Methodol. 2008; 8:31. https://doi.org/10.1186/1471-2288-8-31 PMID: 18492284

25. Yu W, Yesupriya A, Wulf A, Qu J, Khoury MJ, Gwinn M. An open source infrastructure for managing knowledge and finding potential collaborators in a domain-specific subset of PubMed, with an example from human genome epidemiology. BMC Bioinformatics 2007; 8: 436. https://doi.org/10.1186/1471-2105-8-436 PMID: 17996092
31. Clarke LA, Sousa L, Barreto C, Amaral MD. Changes in transcriptome of native nasal epithelium expressing F508del-CFTR and intersecting data from comparable studies. Respiratory Research 2013; 14:38. https://doi.org/10.1186/1465-9921-14-38 PMID: 23537407

32. Virella-Lowell I, Herlihy J, Liu B, Lopez C, Cruz P, Muller C, et al. Effects of CFTR, interleukin-10, and Pseudomonas aeruginosa on gene expression profiles in a CF bronchial epithelial cell line. Mol Ther. 2004; (10)(3):562–573. https://doi.org/10.1016/j.mther.2004.06.215 PMID: 15336656

33. Zabner J, Scheetz TE, Almabrazzi HG, Casavant TL, Huang J, Keshavjee S, et al. CFTR DeltaF508 mutation has minimal effect on the gene expression profile of differentiated human airway epithelia. Am J Physiol Lung Cell Mol Physiol. 2005; 289(4):L545–L553. https://doi.org/10.1152/ajplung.00065.2005 PMID: 15937068

34. Wright JM, Merlo CA, Reynolds JB, Zeitlin PL, Garcia JG, Guggino WB, et al. Respiratory epithelial gene expression in patients with mild and severe cystic fibrosis lung disease. Am J Respir Cell Mol Biol. 2006; 35(3):327–336. https://doi.org/10.1165/rcmb.2005-0359OC PMID: 16614352

35. Verhaeghe C, Remouchamps C, Hennuy B, Vanderplasschen A, Chariot A, Tabruyn S, et al. Role of IKK and ERK pathways in intrinsic inflammation of cystic fibrosis airways. Biochem Pharmacol. 2007; 73(12):1982–1994. https://doi.org/10.1016/j.bcp.2007.03.019 PMID: 17466952

36. Hampton T, Stanton B. A novel approach to analyze gene expression data demonstrates that the DeltaF508 mutation in CFTR downregulates the antigen presentation pathway. Am J Physiol Lung Cell Mol Physiol. 2010; 298(4):L473–L482. https://doi.org/10.1152/ajplung.00379.2009 PMID: 20044347

37. Ogilvie V, Passmore M, Hyndman L, Jones L, Stevenson B, Wilson A, et al. Differential global gene expression in cystic fibrosis nasal and bronchial epithelia. Genomics 2011; 98(5):327–36. https:// doi.org/10.1016/j.ygeno.2011.06.008 PMID: 21756994

38. Sreekumar R, Rosado B, Rasmussen D, Charlton M. Hepatic gene expression in histologically progressive nonalcoholic steatohepatitis. Hepatology 2003; 38:244–251. https://doi.org/10.1053/heap.2003.50290 PMID: 12830008

39. Younossi ZM, Baranova A, Ziegler K, Del Giacco L, Schlauch K, Born TL, et al. Hepatic gene expression in patients with obesity-related non-alcoholic steatohepatitis. Liver Int. 2005; 25:760–771. https://doi.org/10.1002/hep.20838 PMID: 16116632

40. Younossi ZM, Gorreta F, Ong JP, Schlauch K, Giacco LD, Elariny H, et al. Hepatic gene expression in patients with obesity-related non-alcoholic steatohepatitis. Liver Int. 2005; 25:760–771. https://doi.org/10.1002/hep.20838 PMID: 16116632

41. Lake AD, Novak P, Fisher CD, Jackson JP, Hardwick RN, Dean Billheimer D et al. Analysis of Global Absorption, Distribution, Metabolism, and Elimination Gene Expression in the Progressive Stages of Human Nonalcoholic Fatty Liver Disease. Drug Metab Dispos 2011; 39(10):1954–1960. https://doi.org/10.1124/dmd.111.040592 PMID: 21756994

42. Friess H, Ding J, Kleeff J, Liao Q, Berberat PO, Hammer J, et al. Identification of Disease-specific Genes in Chronic Pancreatitis Using DNA Array Technology. Annals of Surgery 2001; 234(6):769–779. PMID: 11729383

43. Binkley CE, Zhang L, Greenson JK, Giordano TJ, Kuick R, Misiek D, et al. The Molecular Basis of Pancreatic—FibrosisCommon Stromal Gene Expression in Chronic Pancreatitis and Pancreatic Adenocarcinoma. Pancreas 2004; 29(4):254–263. PMID: 15502640

44. Edgar R, Domrachev M, Lash AE. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic Acids Research 2002; 30(1):207–210. PMID: 11752295

45. Barrett T, Willhite SE, Ledoux P, et al. NCBIGEO: archive for functional genomics data sets-update. Nucleic Acids Research 2013; 41(1):D991–D995.

46. Barrett T, Troup DB, Willhite SE, Ledoux P, Rudnev D, Evangelista C, et al. NCBI GEO: archive for high-throughput functional genomic data. Nucleic Acids Research 2009; 37:D885–D890. https://doi.org/10.1093/nar/gkn764 PMID: 19308857

47. Ahrens M, Ammerpohl O, von Schönfels W, Kolarova J et al. DNA methylation analysis in nonalcoholic fatty liver disease suggests distinct disease-specific and remodeling signatures after bariatric surgery. Cell Metab. 2013; 18(2):296–302. https://doi.org/10.1016/j.cmet.2013.07.004 PMID: 23931760

48. Nakata S, Imagawa A, Miyata Y, Yoshikawa A et al. Low gene expression levels of activating receptors of natural killer cells (NKG2E and CD94) in patients with fulminant type 1 diabetes. Immunol Lett. 2013; 156(1–2):149–55. https://doi.org/10.1016/j.imlet.2013.10.004 PMID: 24177169

49. Chen JJ, Hsueh HM, Delongchamp RR, Lin CJ, Tsai CA. Reproducibility of microarray data: a further analysis of microarray quality control (MAQC) data. BMC Bioinformatics 2007; 8:412–426. https://doi.org/10.1186/1471-2105-8-412 PMID: 17961233
51. Gu Y, Harley IT, Henderson LB, Aronow BJ, Vietor I, Huber LA, et al. Identification of IFRD1 as a modifier gene for cystic fibrosis lung disease. Nature 2009; 458(7241):1039–42. https://doi.org/10.1038/nature07811 PMID: 19242412

52. Mondini M, Costa S, Sponza S, Gugliesi F, Gariglio M, Landolfo S. The interferon-inducible HIN-200 gene family in apoptosis and inflammation: implication for autoimmunity. Autoimmunity 2010; 43:226–31. https://doi.org/10.3109/08916930903510922 PMID: 20187706

53. Payton M, Coats S. Cyclin E2, the cycle continues. Int J Biochem Cell Biol. 2002; 34(4):315–20. PMID: 11854029

54. Lauper N, Beck ARP, Cariou S, Richman L, Hofmann K, Reith W, et al. Cyclin E2: a novel CDK2 partner in the late G1 and S phases of the mammalian cell cycle. Oncogene 1998; 17(20):2637–2643. https://doi.org/10.1038/sj.onc.1202477 PMID: 9840927

55. Nakamura H, Yoshimura K, McElvaney NG, Crystal RG. Neutrophil elastase in respiratory epithelial lining fluid of individuals with cystic fibrosis induces interleukin-8 gene expression in a human bronchial epithelial cell line. J Clin Invest. 1992; 89:1478–1484. https://doi.org/10.1172/JCI115738 PMID: 1569186

56. Fischer BM, Zheng S, Fan R, Voynow JA. Neutrophil elastase inhibition of cell cycle progression in airway epithelial cells in vitro is mediated by p27kip1. Am J Physiol Lung Cell Mol Physiol. 2007; 293(3):L762–8. https://doi.org/10.1152/ajplung.00067.2007 PMID: 17586698

57. Kelley Kevin M, Oh Y, Gargosky SE, Gucev Z, Matsumoto T, Hwa V, et al. Insulin-like growth factor-binding proteins (IGFBPs) and their regulatory dynamics. The International Journal of Biochemistry & Cell Biology 1996; 28:619–637.

58. Street ME, Ziveri MA, Spaggiari C, Viani I, Volta C, Grzincich GL, et al. Inflammation is a modulator of the insulin-like growth factor (IGF)/IGF-binding protein system inducing reduced bioactivity of IGFs in cystic fibrosis. Eur J Endocrinol. 2006; 154(1):47–52. https://doi.org/10.1530/eje.1.02064 PMID: 16381990

59. Collaco JM, Cutting GR. Update on gene modifiers in cystic fibrosis. Curr Opin Pulm Med. 2008; 14(6):559–66. https://doi.org/10.1097/MCP.0b013e3283321c5c PMID: 18812833

60. Michael RK, Drumm M. The Influence of Genetics on Cystic Fibrosis Phenotypes. Cold Spring Harb Perspect Med. 2012; 2:a009548. https://doi.org/10.1101/cshperspect.a009548 PMID: 23209180

61. Václavíková R, Hughes DJ, Souček P. Microsomal epoxide hydrolase 1 (EPHX1): Gene, structure, function, and role in human disease. Gene 2015; 571(1):1–8. https://doi.org/10.1016/j.gene.2015.07.071 PMID: 26216302

62. Peluso ME, Munnia A, Srivatanakul P, Jedpiyawongse A, Sangrajrang S, Ceppi M, et al. DNA adducts and combinations of multiple lung cancer at-risk alleles in environmentally exposed and smoking subjects. Environ. Mol. Mutagen. 2013; 54:375–383. https://doi.org/10.1002/em.21788 PMID: 23797975

63. Li H, Fu WP, Hong ZH. Microsomal epoxide hydrolase gene polymorphisms and risk of chronic obstructive pulmonary disease: a comprehensive meta-analysis. Oncol. Lett. 2013; 5:1022–1030. https://doi.org/10.3892/ol.2012.1099 PMID: 23426996

64. Vibhuti A, Arif E, Deepak D, Singh B, Qadar Pasha MA. Genetic polymorphisms of GSTP1 and mEPHX correlate with oxidative stress markers and lung function in COPD. Biochem Biophys Res Commun. 2007; 359(1):136–42. https://doi.org/10.1016/j.bbrc.2007.05.076 PMID: 17532303

65. Korytina GF, Ianbaeva DG, Viktorova TV. Role of polymorphic variants of cytochrome P450 genes (CYP1A1, CYP2E1) and microsomal epoxide hydrolase (mEPHX) in pathogenesis of cystic fibrosis and chronic respiratory tract diseases. Mol Biol (Mosk). 2003; 37(5):784–92.

66. Jonas MC, Pehar M, Puglielli L. AT-1 is the ER membrane acetyl-CoA transporter and is essential for cell viability. J. Cell Sci. 2010; 123:3378–3388. https://doi.org/10.1242/jcs.068841 PMID: 20826464

67. Farinha CM, Amaral MD. Most F508del-CFTR is targeted to degradation at an early folding checkpoint and independently of calnexin. Mol Cell Biol. 2005; 25(12):5242–52. https://doi.org/10.1128/MCB.25.12.5242-5252.2005 PMID: 15926368

68. Chakrabarti A, Chen AW, Varner JD. A review of the mammalian unfolded protein response. Biotechnol Bioeng. 2011; 108(12):2777–93. https://doi.org/10.1002/bit.23282 PMID: 21809331

69. Kerbiriou M, Le Drévo MA, Férec C, Trouvé P. Coupling cystic fibrosis to endoplasmic reticulum stress: differential role of Grp78 and ATF6. BBA Molecular Basis of Diseases 2007; 1772(11–12):1236–49.

70. Pehar M, Cabell Jonas M, Hare TM, Puglielli L. SLC33A1/AT-1 Protein Regulates the Induction of Autophagy Downstream of IRE1/XBP1 Pathway. The Journal of biological chemistry 2012; 287(35):29921–29930. https://doi.org/10.1074/jbc.M112.363911 PMID: 22787145
71. Zubillaga P, Vidales MC, Zubillaga I, Ormaechea V, García-Urkia N, Vitoria JC. HLA-DQA1 and HLA-DQB1 Genetic Markers and Clinical Presentation in Celiac Disease. Journal of Pediatric Gastroenterology and Nutrition 2002; 34:548–554. PMID: 12050583

72. Walkowiak J, Blask-Osip A, Lisowska A, Oralewska B, Pogorzelski A, Cicy W, et al. Cystic fibrosis is a risk factor for celiac disease. Acta Biochim Pol. 2010; 57(1):115–8. PMID: 20300660

73. Fluge G, Olesen HV, Gilljam M, Meyer P, Pressler T, Storrösten OT, et al. Co-morbidity of cystic fibrosis and celiac disease in Scandinavian cystic fibrosis patients. J Cyst Fibros. 2009; 8(3):196–202. https://doi.org/10.1016/j.jcf.2009.02.002 PMID: 19303374

74. Schmidt A, Heid HW, Schafer S, Nuber UA, Zimbelmann R, Franke WW. Desmosomes and cytoskeletal architecture in epithelial differentiation: cell type-specific plaque components and intermediate filament anchorage. Eur J Cell Biol. 1994; 65(2):229–45. PMID: 7720719

75. Brezillon S, Dupuit F, Hinrnasky J, Marchand V, Kahn N, Tummler B, et al. Decreased expression of CFTR protein in remodeled human nasal epithelium from non-cystic fibrosis patients. Lab. Invest. 1995; 72:191–200. PMID: 7531792

76. Edelman A. Cytoskeleton and CFTR. The International Journal of Biochemistry & Cell Biology 2014; 52:68–72.

77. Losa D, Chanson M. The lung communication network. Cell Mol Life Sci. 2015; 72(15):2793–808. https://doi.org/10.1007/s00018-015-1960-9 PMID: 26100513

78. Aron Y, Polla BS, Bienvenu T, Dall’ava J, Dusser D, Hubert D. HLA class II polymorphism in cystic fibrosis. A possible modifier of pulmonary phenotype. Am J Respir Crit Care Med. 1999; 159(5 Pt1):1464–8.

79. Corvol H, Blackman SM, Boelle PY, Gallins PJ, Pace RG, Stonebraker JR, et al. Genome-wide association meta-analysis identifies five modifier loci of lung disease severity in cystic fibrosis. Nat Commun. 2015; 6:8382. https://doi.org/10.1038/ncomms9382 PMID: 26417704

80. Bénarsat K, Ben-Moustapha I, Ben-Ali M, Rouault K, Hamami S, Mekki N, et al. A Founder Effect of c.257 + 2T>C Mutation in NCF2 Gene Underlies Severe Chronic Granulomatous Disease in Eleven Patients. J Clin Immunol. 2016; [Epub ahead of print].

81. Ebinu JO, Bottorff DA, Chan EY, Stang SL, Dunn RJ, Stone JC. RasGRP, a Ras guanyl nucleotide-releasing protein with calcium- and diacylglycerol-binding motifs. Science 1998; 280:1082–1086. PMID: 9582122

82. Dower NA, Stang SL, Bottorff DA, Ebinu JO, Dickie P, Ostergaard HL, et al. RasGRP is essential for mouse thymocyte differentiation and TCR signaling. Nat Immunol. 2000; 1(4):317–21. https://doi.org/10.1038/79766 PMID: 11017103

83. Lee SH, Yun S, Lee J, Kim MJ, Piao ZH, Jeong M, et al. RasGRP1 is required for human NK cell function. J Immunol. 2009; 183(12):7931–8. https://doi.org/10.4049/jimmunol.0902012 PMID: 19933860

84. Valente CD, Lukacs GS, Verkman AS, Haggie PM. Reduced PDZ interactions of rescued ΔF508CFTR increases its cell surface mobility. J Biol Chem. 2012; 287(52):43630–8. https://doi.org/10.1074/jbc.M112.421172 PMID: 23115232

85. Kanehisa M, Akira M, Goto S, Hattori M, Hiraoka M, Itoh M, et al. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 2008; 36:D480–D484. https://doi.org/10.1093/nar/gkn882 PMID: 18077471

86. Croft D. Building models using Reactome pathways as templates In Silico Systems Biology: Springer 2013:p. 273–83.
93. Khatri P, Sirot M, Butte AJ. Ten years of pathway analysis: Current approaches and outstanding
challenges. PLoS Comput. Biol. 2012; 8:373.

94. Szkarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRING v10:
protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 2015; 43:
D447–52. https://doi.org/10.1093/nar/gku1003 PMID: 2532553

95. Cui J, Chen Y, Wang HY, Wang RF. Mechanisms and pathways of innate immune activation and
regulation in health and cancer. Hum Vaccin Immunother. 2014; 10(11):3270–85. https://doi.org/10.4161/
21654515.2014.979640 PMID: 25625930

96. Sato S, Ward CL, Kopito RR. Cotranslational ubiquitination of cystic fibrosis transmembrane conduct-
tance regulator in vitro. J Biol Chem. 1998; 273(13):7189–92. PMID: 9516408

97. Gudas JM, Payton M, Thukral S, Chen E, Bass M, Robinson MO, et al. Coats S. Cyclin E2, a Novel
G1 Cyclin That Binds Cdk2 and Is Aberrantly Expressed in Human Cancers. Mol. Cell. Biol. 1999;
1:612–622.

98. Norez C, Vandenbrouck C, Bertrand J, Noel S, Durieu E, Oumata N, et al. Roscovitine is a proteostasis
regulator that corrects the trafficking defect of F508del-CFTR by a CDK-independent mechanism. Br J
Pharmacol. 2014; 171(21):4831–4849. https://doi.org/10.1111/bph.12859 PMID: 25065395

99. Löffek S, Bruckner-Tuderman L, Magin TM. Involvement of the ubiquitin-proteasome system in the
stabilization of cell-cell contacts in human keratinocytes. Exp Dermatol. 2012; 21(10):791–3. https://
doi.org/10.1111/j.1600-0625.2012.01564.x PMID: 22882483

100. Ottenheijm CA, Heunks LM, Li YP, Jin B, Minnaard R, van Hees HW, et al. Activation of the ubiquitin-
proteasome pathway in the diaphragm in chronic obstructive pulmonary disease. Am J Respir Crit
Care Med. 2006; 174(9):997–1002. https://doi.org/10.1164/rcrm.200605-721OC PMID: 16917114

101. Henderson MJ, Vij N, Zeitlin PL. Ubiquitin C-terminal hydrolase-L1 protects cystic fibrosis transmem-
brane conductance regulator from early stages of proteasomal degradation. J Biol Chem. 2010; 285
(15):11314–25. https://doi.org/10.1074/jbc.M109.044057 PMID: 20147297

102. Kelly T, Buxbaum J. Gastrointestinal Manifestations of Cystic Fibrosis Dig Dis Sci. 2015; 60:1903–
1913. https://doi.org/10.1007/s10620-015-3546-7 PMID: 25648641

103. Karlas T, Neuschulz M, Oltmanns A, et al. Non-invasive evaluation of cystic fibrosis related liver dis-
ease in adults with ARFI, transient elastography and different fibrosis scores. Plos One. 2012; 7:8.

104. Kobelska-Dubiel N, Klineciczew B, Cichy W. Liver disease in cystic fibrosis. Prz Gastroenterol. 2014; 9
(3):136–141. https://doi.org/10.5114/pg.2014.43574 PMID: 25097709

105. Machen TE. Innate immune response in CF airway epithelia: hyperinflammatory? Am J Physiol Cell
Physiol. 2006; 291:C218–C230. https://doi.org/10.1152/ajpcell.00605.2005 PMID: 16825601

106. Akata D, Akhan O. Liver manifestations of cystic fibrosis. European Journal of Radiology. 2007; 61
(1):11–17. https://doi.org/10.1016/j.ejrad.2006.11.008 PMID: 17174503

107. Laguna TA, Nathan BM, Moran A. Managing diabetes in cystic fibrosis. Diabetes Obes Metab. 2010;
12(10):858–864. https://doi.org/10.1111/j.1463-1326.2010.01250.x PMID: 20920037

108. Kerem B, Kerem E. The molecular basis for disease variability in cystic fibrosis. Eur J Hum Genet.
1996; 4:65–73. PMID: 8744024