Title: Whole exome sequencing in the diagnosis of liver diseases

Xiao-Fei Kong[1#], Kelsie Bogyo[2], Sheena Kapoor[2], Emily E. Groopman[2], Enrico Cocchi[2], Hila Milo Rasouly[2], Beishi Zheng[1], Siming Sun[1], Junying Zhang[2], Mercedes Martinez[4], Jennifer M Vittorio[4], Lorna M. Dove[1,4], Maddalena Marasa[2], Timothy C. Wang[1], Elizabeth C. Verna[1,4], Howard J. Worman[1,5], Julia Wattacheril[1,4*], David B. Goldstein[3*], Ali G. Gharavi[2,3*#]

1. Division of Digestive and Liver Diseases, Department of Medicine;

2. Division of Nephrology, Department of Medicine;

3. Institute for Genomic Medicine;

4. Center for Liver Disease and Transplantation;

5. Department of Pathology and Cell Biology.

Columbia University Irving Medical Center, New York, NY 10032, USA

[*]: Contributed equally

[#]: Correspondence

Keywords:

Chronic liver disease, whole-exome sequencing, pathogenic variants annotation, Human Gene Mutation Database, ClinVar

Footnote Page:

Contact information:

Xiao-Fei Kong, MD, Ph.D

Hammer Health Sciences Building Rm 402, 701 W 168th St, New York, NY 10032

Email: xk2137@cumc.columbia.edu T: (212) 305-1909; F: (212) 432-5759

Ali G. Gharavi, MD

1150 St. Nicholas Ave., Russ Berrie Pavilion 413, New York, NY 10032

Email: ag2239@cumc.columbia.edu

List of Abbreviations:

whole-exome sequencing; WES, chronic liver disease; CLD, chronic kidney disease; CKD, Human Gene Mutation Database; HGMD, protein truncating variants; PTV, minor allele frequency; MAF, next-generation sequencing: NGS, genome-wide association studies; GWAS, American College of Medical Genetics and Genomics and the Association for Molecular Pathology; ACMG-AMP, Online Mendelian Inheritance in Man database; OMIM, Human Phenotype Ontology; HPO, Genome Aggregation Database; gnomAD, Exome Aggregation Consortium; ExAC, autosomal dominant; AD, autosomal recessive; AR, X-linked dominant; XLD, X-linked recessive disorders; XLR, disease-causing mutation; DM, pathogenic/likely pathogenic; P/LP, Combined Annotation Dependent Depletion; CADD, quality control; QC.

Financial Support:

Research reported in this publication was supported by the National Institute Of Diabetes And Digestive And Kidney Diseases of the National Institutes of Health under Award Number K08DK128631. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Abstract

Background: We investigated the diagnostic yield of exome sequencing in patients with liver diseases**.**

Methods: We retrospectively analyzed WES data for 10,801 individuals: 758 patients with CLD, 7,856 self-declared healthy controls (HC), and 2,187 patients with chronic kidney disease (CKD). We searched for variants in a total of 502 genes causing Mendelian disorders with a primary or secondary liver disease phenotype. Candidate variants were previously reported as pathogenic (P) or likely pathogenic (LP) in the Human Gene Mutation Database (HGMD) or ClinVar databases, or novel protein-truncating variants (PTV) .

Results: Candidate pathogenic variants were initially identified in 19.9% of participants, most of which were attributable to previously reported pathogenic variants with implausibly high allele frequencies. After stringent filtering based on population minor allele frequency and variant annotation, P/LP variants were detected in 0.93% of the cohort, with a significant enrichment in the CLD cohort compared to the HC cohort (X^2 test OR: 5.00, 95% CI:3.06-8.18). After two levels of manual annotation, the diagnostic rate of monogenic disorders was 5.7% in the CLD cohort, attributable to P/LP variants in 25 genes. We also identified concordant liver phenotypes for 15/22 kidney disease patients with P/LP variants in liver genes. Clinically confirmed sequencing results had many implications for management: familial testing for early diagnosis and management, preventative screening for associated comorbidities, such as aneurysm or cancer. Some patients with previously undiagnosed congenital disorders of glycosylation, would benefit from selective nutritional management.

Conclusions: Gene list-based WES analysis provided a 5.7% diagnostic rate among patients with CLD, with implications regarding their clinical care and that of their families.

Introduction

Liver disease accounts for approximately 2 million deaths per year worldwide. In the United States, the mortality rate for chronic liver diseases (CLD) increased 31% from 2000 to 2015, making it the fifth leading cause of death in 2017 for persons aged 45-64 years [1]. With the development of effective direct-acting antivirals to treat viral hepatitis C, major research efforts have shifted to other liver diseases. Human genetics of liver diseases dates back to 1865-1890 when Triouseau and von Recklinghausen described hemochromatosis [2]. Subsequently, the cloning, mapping, and functional characterization of homeostatic iron regulator (*HFE*) gene in the 1990s paved a unique pathway for the diagnosis of hemochromatosis [3]. The advent of next-generation sequencing (NGS) approaches have led to the discovery of many novel genetic alterations causing liver phenotypes such as fibrolamellar hepatocellular carcinoma [4], recurrent acute liver failure[5–7], or idiopathic non-cirrhotic portal hypertension[8,9]. These findings demonstrate the power of NGS for identifying novel and poorly understood rare liver diseases.

The implications of NGS for common, complex liver diseases such as nonalcoholic fatty liver disease (NAFLD) are currently being investigated. Physicians now can incorporate genetic testing to diagnose monogenic forms of liver diseases [10,11]. A comprehensive genetic diagnostic pipeline for liver disease could benefit patients and clinicians, improving clinical diagnoses and the early recognition of rare genetic disorders that manifest as a common liver phenotype. For example, numerous monogenic disorders, can manifest as hepatic steatosis as a secondary phenotype and may not be recognized based on their clinical workup[12]. The American College of Medical Genetics and Genomics and the Association for Molecular Pathology (ACMG-AMP) have issued a joint consensus recommendation for the interpretation of sequence variants [13]. Genetic testing of single genes or small gene panels has been used for some suspected hereditary liver diseases [14–18], but NGS approaches

have not been widely adopted into the routine evaluation of liver disease. In this paper, we outline an analytic approach based on ACMG-AMP guidelines and conduct a clinical sequence interpretation for whole-exome sequencing (WES) data from 10,801 individuals. We present the diagnostic utility of WES for liver diseases, and also highlight the potential for misclassification and other special considerations in the genetic workup for liver diseases.

Methods

Developing a list of monogenic disorders associated with liver phenotypes

We first developed a comprehensive gene list, or "liver gene list", to identify genes causing monogenic diseases with a wide range of liver manifestations. We used Online Mendelian Inheritance in Man database (OMIM), Orphanet, and the Human Phenotype Ontology (HPO) database [19] to search for potential genes with Mendelian inheritance that have been associated with or shown to be causative in liver disease before December 2018. For the search, we used a total of 30 keywords or phrases (**Supplementary Figure 1**), then manually reviewed OMIM and related literature [20–27]. We excluded: 1) genes not reported to be linked to any abnormal liver phenotypes; 2) genes within a locus reported from linkage analysis without any known pathogenic variants; 3) genes only discovered in GWAS but lacking any evidence for Mendelian inheritance; 4) genes with only somatic mutations reported in abnormal liver phenotypes; 5) genes within a locus associated with abnormal liver phenotypes due to chromosomal abnormalities. The selected genes were annotated for biological functions, clinical liver presentations, and gene constraint score [28]. We annotated the inheritance mode of liver genes based on OMIM and ClinGen then manually curated the list of genes by reviewing relevant literature.

Cohorts

We analyzed WES data obtained by sequencing of genomic DNA from 758 patients with CLD. A detailed clinical description is provided in **Supplementary Table 1** and **Supplementary Table 2**. We enrolled any patients from both pediatric and adult liver clinics at Columbia University Irving Medical Center (CUIMC) who were interested in and able to consent to participating in genetics research, without setting particular inclusion or exclusion criteria (**Table 1**). [sentence missing?]Additional controls were used to evaluate the gene-list based WES analysis, including 7,856 self-identified healthy individuals and 2,187 patients with chronic kidney disease (CKD) (**Supplementary Table 1**)[29,30]. Informed consent in writing was obtained from each patient and the study protocol conformed to the ethical guidelines of the 1975 Declaration of Helsinki as reflected in a priori approval by CUIMC Institutional Review Board.

Sequence analysis and variant annotation

Sample preparation, target-enrichment, sequencing process, read alignment, and variant calling were previously published[29,30]. We focused on variants that were predicted to have at least moderate to strong biological effects toward protein function and excluded those in intergenic and promoter regions. We used stringent quality filters and removed potential technical false-positive insertions and deletions (indels) using ATAV as previously described [29,31]. We excluded variants failing quality cutoffs in gnomAD or those identified as sequencing artifacts through a comparison of in-house control sequencing data. Current guidelines recommend considering all variants with a minor allele frequency (MAF) of less than 1% at the population level. Thus, we filtered the variants based on the overall MAF of less than 1% in the Genome Aggregation Database (gnomAD)[32]. Variants previously reported as pathogenic were identified using the HGMD and ClinVar. We included only those annotated as pathogenic/likely pathogenic (P/LP) in ClinVar or disease-causing mutation (DM) in HGMD without any conflicting evidence within each database. In addition,

we identified novel protein-truncating variants (PTVs) not previously reported in either HGMD or ClinVar. As the initial yield of individuals carrying candidate pathogenic variants was significantly higher than expected, we employed a stringent filter by inheritance mode and subpopulation MAF based on the data from gnomAD and Exome Aggregation Consortium (ExAC): MAF $\leq 10^{-4}$ for dominant disorders and MAF $\leq 10^{-3}$ for recessive disorders[29,33,34]. We used Loss-Of-Function Transcript Effect Estimator (LOFTEE) filter to exclude PTVs with a false prediction. A detailed description of genetic terminology in this study has been described previously[29].

Manual variant classification and clinical data review

Two independent genetic analysts performed a first-tier, stringent analysis of the CLD cohort to reach a consensus classification according to the ACMG-AMP guidelines[35]. We next performed a second-level manual curation of the CLD cohort using lower stringency filters, which identified several well-defined pathogenic variants that were excluded because they either have a MAF above 1% in some ethnic subpopulations or did not pass the stringent sequencing quality filters. This procedure had been successfully used to increase diagnostic yield in prior studies[13,36]. Subsequently, a multidisciplinary group of experts, including genetic counselors, geneticists, molecular pathologists, and clinicians, reviewed the available clinical information in individuals carrying P/LP variants to detect phenotypic concordance with the associated mode of inheritance of disease. If diagnostic evidence was insufficient based on chart review, a follow-up plan was recommended to clarify the significance of the genetic findings.

Statistical analysis

We compared the probability of being loss-of-function intolerant (pLI) and Z scores for genes using an analysis of variance (ANOVA) test to compare differences between the three groups. We analyzed the clinical variables between those with and those without a genetic diagnosis

using the Chi-squared test. All statistics and genetic analyses were done in R statistical software (Version 4.0.0). A p-value of <0.05 was considered significant after correction for multiple hypothesis testing.

Results

1. Characterization of 502 liver genes with Mendelian hepatobiliary disorders

In a comprehensive search for Mendelian genetic disorders with any abnormal liver clinical phenotypes, we manually curated a total of 959 genes. Of these, 502 had a confirmed abnormal liver phenotype, with 193 genetic disorders having primarily liver diseases, for instance, *ABCB11* or *ATP8B1* causing progressive familial intrahepatic cholestasis, other genes might lead to liver problems as a secondary cause (**Supplementary Figure 1** and **Supplementary Appendix 1**). We then annotated inheritance modes and detailed clinical phenotypes related to these 502 genes. In total, 75% of genes were associated with a recessive mode of inheritance (363 AR and 15 XLR). Sixty-two autosomal genes can result in both dominant and recessive disorders and 62 other genes associated with exclusively dominant disorders (61 AD and 1 XLD) (**Supplementary Table 3**). The most common clinical presentation of Mendelian hepatobiliary disorders was hepatomegaly, manifesting in 236/502 disorders (47%) Other common clinical manifestations of Mendelian liver disorders included metabolic disease (25%), liver fibrosis or cirrhosis (25%), elevated hepatic transaminase level (20%), and cholestasis (19%) (**Figure 1A**). Based on the clinical presentations and biological functions, we grouped these genes into five broad categories: hepatocellular, biliary, metabolic, immunologic, and developmental. Most of the genes (298/502, 59%) were associated with a developmental or congenital disorder with liver manifestations (**Supplementary Table 3**). The 62 genes exclusively associated with dominant inheritance showed significantly higher pLI (**Figure 1B**) and missense Z scores (**Figure 1C**) compared to the 378 genes associated with recessive diseases or the 62 genes

associated with both dominant and recessive inheritance. Based on these results, genes associated with both dominant and recessive disorders were analyzed under the recessive inheritance model.

2. Assessment of the frequency of candidate pathogenic/likely pathogenic variants

To investigate the prevalence of candidate pathogenic variants in the liver genes, we analyzed WES data from 10,801 individuals, agnostic to the clinical phenotypes. We filtered out variants based on the quality matrix, and initially identified a total of 555 candidate pathogenic variants (**Supplement Figure 2A and 2B**). Altogether, 2,152 (19.9%) individuals carried these 555 candidate pathogenic variants: 1,567 (20.2%) in healthy controls, 416 (19.0%) in the CKD cohort, and 159 (21%) in the CLD cohort (**Supplementary Table 4**). Based on analysis of the maximal MAF from eight populations originating from the ExAC and gnomAD as well as CADD score (**Supplement Figure 2C**), we concluded that many of the previously reported pathogenic variants are not disease-causing and were erroneously reported as pathogenic (methods and supplemental data). We next used the maximal MAF to filter variants, and also performed manual review of 403 variants. (**Supplementary Table 5, Supplementary Figure 2A and 2B**)[29,33,34]. This manual annotation resulted in 112 variants being classified as either P/LP based on ACMG-AMP classification (including 78 PTVs, **Supplement Figure 2D**), detected across 45 genes in a total of 100 individuals (0.93% of three cohorts). The prevalence of these P/LP variants significantly differed between healthy controls (51/7856, 0.65%), patients with CKD (25/ 2187, 1.14%), and patients with CLD (24/758, 3.17%) (X^2 test OR: 5.00, 95%CI:3.06-8.18, p-value=4.55e - 12, Supplementary **Figure 2A)**. Variants for genes for four dominant genetic disorders were found in all three cohorts, namely *HNF1A* (maturity-onset diabetes of the young 3),

PKD1(polycystic kidney disease 1), *PRKCSH* (polycystic liver disease 1), *SEC63* (polycystic liver disease 2) (**Supplementary Figure 2E**).

3. Second-level annotation of the CLD cohort identifies additional pathogenic variants

To maximize the identification of diagnostic variants in the CLD cohort, we performed a second-level manual assessment, using more relaxed sequence quality thresholds that we had previously deployed for diagnostic evaluation of other cohorts [30,37]. This second-level analysis led to the identification of 16 additional diagnostic variants that explained the liver phenotypes in 14 additional patients (13 genes, **Figure 2A**). All 16 variants were missed because of the high stringency sequence quality thresholds and were all confirmed by Sanger sequencing. In addition, we evaluated five well-known pathogenic variants or risk alleles for liver disease that have a MAF above 1%: *HFE* C282Y and H63D, *SERPINA1*E264V (Pi*S) and E342K (Pi*Z). We found two patients with P/LP variants in *HFE* (one with a homozygous *HFE* C282Yvariant, and one with an H63D/c.340+1G>A genotype, **Table 2**). Both had high serum iron transferrin saturation and ferritin levels, and clinical presentations consistent with hereditary hemochromatosis. For *SERPINA1*, three patients in the CLD cohort had a homozygous Pi*Z genotype, and all of them had a clinical diagnosis of alpha-1 anti-trypsin deficiency (**Table 2**). Altogether, this second level analysis increased the diagnostic yield in the liver cohort to 43/758 cases (5.7%, **Figure 2A**).

4. Genetic diagnoses and their clinical implications

Overall, we identified a total of 25 genetic disorders in the liver disease cohort, with Alagille syndrome, alpha-1 anti-trypsin deficiency, cystic fibrosis, and progressive familial intrahepatic cholestasis-2 found in at least three patients each **(Figure 2B)**. Between the patients with or without a genetic diagnosis in the liver disease cohort, no differences were observed in sex, race, or ethnicity. From an univariate analysis, younger age and the clinical diagnosis of congenital liver disorders, abnormally elevated serum transaminase activities

due to unknown causes were associated with a higher rate of a genetic diagnosis (**Table 1**). We next performed a case-level review to assess concordance between genotype and phenotype. Among 43 liver disease patients with P/LP variants, we confirmed a previous clinical diagnosis for eleven, identified a genetic disease that partially explains phenotype for eleven, reclassified disease for seven, identified a molecular subtype of inherited liver diseases for six, and identified a cause for undiagnosed liver diseases for five. We also recommended a further workup in three patients to confirm or refute the liver diagnosis (**Figure 2C**). Finally, we examined the phenotypic concordance for the 25 CKD patients carrying P/LP variants in liver genes: 15/25 patients had a corresponding liver phenotype, which were mostly attributable to genes like *PKD1* and MODY or ciliopathy genes causing both kidney and liver disease (**Supplement Appendix 2).** Benefits of a genetic diagnosis included the ability to guide familial testing and obtain an early diagnosis of affected family members for 24 families, or to perform surveillance for known complications, such as brain aneurysms in individuals carrying a pathogenic variant in *PKD1*, and to conduct an appropriate cancer screening surveillance for four patients with *HFE* and *PFIC2*. Patients with *PGM1* and *PHKA2* mutations, diagnostic of congenital disorders of glycosylation, can benefit from selective nutritional management. Other implications for better treatment include targeted therapy, clinical trials, or surgical options. A review of clinicaltrials.gov identified 255 clinical trials are enrolling patients with monogenic forms of liver diseases identified in this study (**Figure 2D**).

Discussion:

Our primary goal was to evaluate the utility of exome sequencing for diagnosis of liver disease. Currently available clinical genetic testing for heritable liver diseases exists and is mostly utilized in the pediatric diagnostic framework . For instance, one lab provides a panel of 72 genes for well-defined monogenic liver diseases, especially cholestasis and biliary

atresia[38–40]. To facilitate the WES analysis, we developed a list of 502 genes associated with a Mendelian disease with potential liver phenotypes (**Figure 3**). Several new genetic disorders, such as *TULP3*[41], *KIF12*, *USP54*[42], *KCNN3*[43,44], *GIMAP5*[9] and so on, were discovered and recently reported to cause liver phenotypes. These discoveries coupled with their detailed annotated phenotypes provide an important resource to look for a genetic diagnosis for those patients with a related clinical phenotype and available WES result. Although we did not find any patients with the newly discovered genetic disorders mentioned above, nonetheless, our work joined the global efforts to develop a list of genes with sufficient evidence causing clinical phenotypes. In the future, the creation of a liver disease workgroup, for instance, under the ClinGen platform, will accelerate the development of a better version of gene list for CLD.

The current challenge of genetic analysis is to determine the pathogenicity of variants. Consistent with prior studies of other genetic disorders, our variant level analyses indicated that many previously reported P/LP variants for liver diseases are too common to be disease-causing and are erroneously annotated in reference databases. We report liver disease genes with the most frequently encountered false-positive P/LP variants to help with the reannotation of reference databases (Supplementary Appendix 2). We also performed a manual annotation at the data, which confirmed that the application of hard filters for allele frequency and sequence quality may lead to the omission of true pathogenic variants. For example, in addition to *HFE* and *SERPINA1* pathogenic variants, two patients with progressive familial intrahepatic cholestasis type 3 carried an *ABCB4* Ala934Thr missense variant with a MAF of 1.2% in the African American population, which should be interpreted as a pathogenic variant. Thus, a balanced approach was necessary for maximizing the diagnostic rate. A case-level review indicated that the genetic results were consistent with the clinical findings in the majority of liver and kidney disease cases, validating our approach.

For the patients who did not have a concordant phenotype, the mutations may be non-penetrant or may develop in the future, or the variant may be downgraded based on future evidence of non-pathogenicity. We note that our study is limited by the lack of clinical information for most self-reported healthy controls, which hampers our ability to determine the causality of P/LP variants in this cohort.

Altogether, our single-center study indicates a significant diagnostic utility for WES in the evaluation of patients with CLD. Future studies will have to evaluate the diagnostic utility across multiple healthcare settings and prospectively demonstrate the impact of genetic testing on clinical decision-making.

Figure and table legends

Figure 1: A summary of liver phenotypes in Mendelian genetic disorders.

A) Inheritance mode, annotated clinical liver phenotypes, and biological effects of 502 genes related to Mendelian disorders. The liver phenotypes and inheritance were curated based on OMIM, ClinGen, and a literature search. AD: autosomal dominant disorder; AR: autosomal recessive disorder; XLD: X-linked disorder; XLR: X-linked recessive disorder; AD and AR: Genes with both autosomal dominant and autosomal recessive inheritance were reported. The right lower box showed the numbers of genes with corresponding biological effects and inheritance mode. B) Box plot of pLI scores of 502 genes in three groups based on inheritance mode. The dark line inside the box represents the median of pLI score. The top of box is 75% and bottom of box is 25%. The endpoints of the lines are at a distance of 1.5*IQR, where inter quartile range is the distance between 25th and 75th percentiles. The points outside the whiskers are marked as dots and are considered as extreme points. C) Violin plot of missense Z scores of 502 genes in three different groups based on inheritance mode. P-values in B and C for differences between dominant and recessive genes were determined using ANOVA.

Figure 2: Genetic diagnosis and clinical implications of WES findings in the liver disease cohort. A) A total of 43 CLD patients with P/LP variants from three searching approaches; B. A total of 25 genetic disorders were found in the CLD cohort. Red star indicated the genetic disorders causing primarily liver diseases. A genetic characterization (C.) and clinical implication (D.) with the findings.

Figure 3: A summary of the genetic analytic strategy and outcomes for liver diseases.

Table 1: A comparison of clinical characteristics for those with or without a genetic diagnosis in the liver disease cohort.

Table 2: Hemochromatosis, alpha1 antitrypsin deficiency, and genetic diagnosis in the liver disease cohort. High MAF candidate pathogenic variants in *HFE* and *SERPINA1*. # Individuals with a homozygous pathogenic variant or two heterozygous pathogenic variants reported in HGMD or ClinVar. [*]Unable to determine the phase of two variants. [$]*:* cases with sufficient clinical evidence of liver phenotypes consistent with the genetic diagnosis. One case with *H63D* has liver phenotypes consistent with hemochromatosis.

Reference:

[1] *QuickStats:* Death Rates* for Chronic Liver Disease and Cirrhosis,† by Sex and Age Group — National Vital Statistics System, United States, 2000 and 2015. MMWR Morb Mortal Wkly Rep 2017;66:1031. https://doi.org/10.15585/mmwr.mm6638a9.

[2] Barton JC, Edwards CQ, Acton RT. HFE gene: Structure, function, mutations, and associated iron abnormalities. Gene 2015;574:179–92. https://doi.org/10.1016/j.gene.2015.10.009.

[3] Feder JN, Gnirke A, Thomas W, Tsuchihashi Z, Ruddy DA, Basava A, et al. A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. Nat Genet 1996;13:399–408. https://doi.org/10.1038/ng0896-399.

[4] Honeyman JN, Simon EP, Robine N, Chiaroni-Clarke R, Darcy DG, Lim IIP, et al. Detection of a recurrent DNAJB1-PRKACA chimeric transcript in fibrolamellar hepatocellular carcinoma. Science (80-) 2014;343:1010–4. https://doi.org/10.1126/science.1249484.

[5] Belkaya S, Michailidis E, Korol CB, Kabbani M, Cobat A, Bastard P, et al. Inherited IL-18BP deficiency in human fulminant viral hepatitis. J Exp Med 2019;216:1777–90. https://doi.org/10.1084/jem.20190669.

[6] Haack TB, Staufner C, Köpke MG, Straub BK, Kölker S, Thiel C, et al. Biallelic Mutations in NBAS Cause Recurrent Acute Liver Failure with Onset in Infancy. Am J Hum Genet 2015;97:163–9. https://doi.org/10.1016/j.ajhg.2015.05.009.

[7] Cousin MA, Conboy E, Wang J-S, Lenz D, Schwab TL, Williams M, et al. RINT1 Bi-allelic Variations Cause Infantile-Onset Recurrent Acute Liver Failure and Skeletal Abnormalities. Am J Hum Genet 2019. https://doi.org/10.1016/J.AJHG.2019.05.011.

[8] Vilarinho S, Sari S, Yilmaz G, Stiegler AL, Boggon TJ, Jain D, et al. Recurrent recessive mutation in deoxyguanosine kinase causes idiopathic noncirrhotic portal hypertension. Hepatology 2016;63:1977–86. https://doi.org/10.1002/hep.28499.

[9] Drzewiecki K, Choi J, Brancale J, Leney-Greene MA, Sari S, Dalgiç B, et al. GIMAP5 maintains liver endothelial cell homeostasis and prevents portal hypertension. J Exp Med 2021;218. https://doi.org/10.1084/JEM.20201745.

[10] Denny JC, Rutter JL, Goldstein DB, Philippakis A, Smoller JW, Jenkins G, et al. The "All of Us" Research Program. N Engl J Med 2019;381:668–76. https://doi.org/10.1056/NEJMsr1809937.

[11] Dewey FE, Murray MF, Overton JD, Habegger L, Leader JB, Fetterolf SN, et al. Distribution and clinical impact of functional variants in 50,726 whole-exome sequences from the DiscovEHR study. Science (80-) 2016;354. https://doi.org/10.1126/science.aaf6814.

[12] Liebe R, Esposito I, Bock HH, vom Dahl S, Stindt J, Baumann U, et al. Diagnosis and management of secondary causes of steatohepatitis. J Hepatol 2021;74:1455–71. https://doi.org/10.1016/j.jhep.2021.01.045.

[13] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–24. https://doi.org/10.1038/gim.2015.30.

[14] Kong XF, Ye J, Gao DY, Gong QM, Zhang DH, Lu ZM, et al. Identification of a ferrochelatase mutation in a Chinese family with erythropoietic protoporphyria. J Hepatol 2008;48:375–9. https://doi.org/S0168-8278(07)00589-2 [pii]10.1016/j.jhep.2007.09.013.

[15] Li X-H, Lu Y, Ling Y, Fu Q-C, Xu J, Zang G-Q, et al. Clinical and molecular characterization of Wilson's disease in China: identification of 14 novel mutations. BMC Med Genet 2011;12:6. https://doi.org/10.1186/1471-2350-12-6.

[16] Kong X-F, Han Y, Li X-H, Gao D-Y, Zhang X-X, Gong Q-M. Recurrent porphyria attacks in a Chinese patient with a heterozygous PBGD mutation. Gene 2013;524:401–2. https://doi.org/10.1016/j.gene.2013.03.130.

[17] Karlsen TH, Lammert F, Thompson RJ. Genetics of liver disease: From pathophysiology to clinical practice. J Hepatol 2015;62:S6–14. https://doi.org/10.1016/j.jhep.2015.02.025.

[18] Pelusi S, Ronzoni L, Malvestiti F, Bianco C, Marini I, D'Ambrosio R, et al. Clinical exome sequencing for diagnosing severe cryptogenic liver disease in adults: A case series. Liver Int 2022;42:864–70. https://doi.org/10.1111/LIV.15185.

[19] Köhler S, Vasilevsky NA, Engelstad M, Foster E, McMurry J, Aymé S, et al. The human phenotype ontology in 2017. Nucleic Acids Res 2017;45:D865–76. https://doi.org/10.1093/nar/gkw1039.

[20] Marques-da-Silva D, dos Reis Ferreira V, Monticelli M, Janeiro P, Videira PA, Witters P, et al. Liver involvement in congenital disorders of glycosylation (CDG). A systematic review of the literature. J Inherit Metab Dis 2017;40:195–207. https://doi.org/10.1007/s10545-016-0012-4.

[21] Kulecka M, Habior A, Paziewska A, Goryca K, Dąbrowska M, Ambrozkiewicz F, et al. Clinical Applicability of Whole-Exome Sequencing Exemplified by a Study in Young Adults with the Advanced Cryptogenic Cholestatic Liver Diseases. Gastroenterol Res Pract 2017;2017:1–8. https://doi.org/10.1155/2017/4761962.

[22] Schonfeld EA, Brown RS. Genetic Testing in Liver Disease: What to Order, in Whom, and When. Clin Liver Dis 2017;21:673–86. https://doi.org/10.1016/j.cld.2017.06.001.

[23] Stephens MC, Boardman LA, Lazaridis KN. Individualized Medicine in Gastroenterology and Hepatology. Mayo Clin Proc 2017;92:810–25. https://doi.org/10.1016/j.mayocp.2017.03.002.

[24] Nicastro E, D'Antiga L. Next generation sequencing in pediatric hepatology and liver transplantation. Liver Transplant 2018;24:282–93. https://doi.org/10.1002/lt.24964.

[25] Pericleous M, Kelly C, Ala A, Schilsky ML. The epidemiology of rare hereditary

metabolic liver diseases. Clin. Epidemiol. Chronic Liver Dis., Springer International Publishing; 2018, p. 307–30. https://doi.org/10.1007/978-3-319-94355-8_17.

[26] Nicastro E, D'Antiga L. Next generation sequencing in pediatric hepatology and liver transplantation. Liver Transplant 2018;24:282–93. https://doi.org/10.1002/lt.24964.

[27] Date E, Name E, Address S, Code Z, Tax C, Program R, et al. Commuter Tax Relief (CTR) Program Enrollment/Certification Tax-Reduction (Public Transportation) n.d.

[28] Lek M, Karczewski KJ, Minikel E V., Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285–91. https://doi.org/10.1038/nature19057.

[29] Rasouly HM, Groopman EE, Heyman-Kantor R, Fasel DA, Mitrotti A, Westland R, et al. The burden of Candidate pathogenic variants for kidney and genitourinary disorders emerging from exome sequencing. Ann Intern Med 2019;170:11–21. https://doi.org/10.7326/M18-1241.

[30] Groopman EE, Marasa M, Cameron-Christie S, Petrovski S, Aggarwal VS, Milo-Rasouly H, et al. Diagnostic Utility of Exome Sequencing for Kidney Disease. N Engl J Med 2019;380:142–51. https://doi.org/10.1056/nejmoa1806891.

[31] Ren Z, Povysil G, Goldstein DB. ATAV - a comprehensive platform for population-scale genomic analyses. BioRxiv 2020:2020.06.08.136507. https://doi.org/10.1101/2020.06.08.136507.

[32] Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434–43. https://doi.org/10.1038/s41586-020-2308-7.

[33] Whiffin N, Minikel E, Walsh R, O'Donnell-Luria AH, Karczewski K, Ing AY, et al. Using high-resolution variant frequencies to empower clinical genome interpretation. Genet Med 2017;19:1151–8. https://doi.org/10.1038/GIM.2017.26.

[34] Whiffin N, Roberts AM, Minikel E, Zappala Z, Walsh R, O'Donnell-Luria AH, et al. Using High-Resolution Variant Frequencies Empowers Clinical Genome Interpretation and Enables Investigation of Genetic Architecture. Am J Hum Genet 2019;104:187–90. https://doi.org/10.1016/J.AJHG.2018.11.012.

[35] Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: A joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med 2015;17:405–24. https://doi.org/10.1038/gim.2015.30.

[36] Kalia SS, Adelman K, Bale SJ, Chung WK, Eng C, Evans JP, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): A policy statement of the American College of Medical Genetics and Genomics. Genet Med 2017;19:249–55. https://doi.org/10.1038/gim.2016.190.

[37] Shashi V, Schoch K, Spillmann R, Cope H, Tan QKG, Walley N, et al. A comprehensive iterative approach is highly effective in diagnosing individuals who are exome negative. Genet Med 2019;21:161–72. https://doi.org/10.1038/S41436-018-0044-2.

[38] CINCINATTI hospital liver genetics testing panel n.d.

[39] Shagrani M, Burkholder J, Broering D, Abouelhoda M, Faquih T, El-Kalioby M, et al. Genetic profiling of children with advanced cholestatic liver disease. Clin Genet 2017;92:52–61. https://doi.org/10.1111/CGE.12959.

[40] S A, A H. A Comprehensive Review of Progressive Familial Intrahepatic Cholestasis (PFIC): Genetic Disorders of Hepatocanalicular Transporters. Gastroenterol Res 2014;7. https://doi.org/10.14740/GR609E.

[41] Devane J, Ott E, Olinger EG, Epting D, Decker E, Friedrich A, et al. Progressive liver, kidney, and heart degeneration in children and adults affected by TULP3 mutations. Am J Hum Genet 2022;109:928–43. https://doi.org/10.1016/J.AJHG.2022.03.015.

[42] Maddirevula S, Alhebbi H, Alqahtani A, Algoufi T, Alsaif HS, Ibrahim N, et al. Identification of novel loci for pediatric cholestatic liver disease defined by KIF12, PPM1F, USP53, LSR, and WDR83OS pathogenic variants. Genet Med 2019;21:1164–72. https://doi.org/10.1038/S41436-018-0288-X.

[43] Koot BGP, Alders M, Verheij J, Beuers U, Cobben JM. A de novo mutation in KCNN3 associated with autosomal dominant idiopathic non-cirrhotic portal hypertension. J Hepatol 2016;64:974–7. https://doi.org/10.1016/J.JHEP.2015.11.027.

[44] Bauer CK, Schneeberger PE, Kortüm F, Altmüller J, Santos-Simarro F, Baker L, et al. Gain-of-Function Mutations in KCNN3 Encoding the Small-Conductance Ca 2þ-Activated K þ Channel SK3 Cause Zimmermann-Laband Syndrome. Am J Hum Genet 2019;104:1139–57. https://doi.org/10.1016/j.ajhg.2019.04.012.

Figure 1: A summary of liver phenotypes in Mendelian genetic disorders

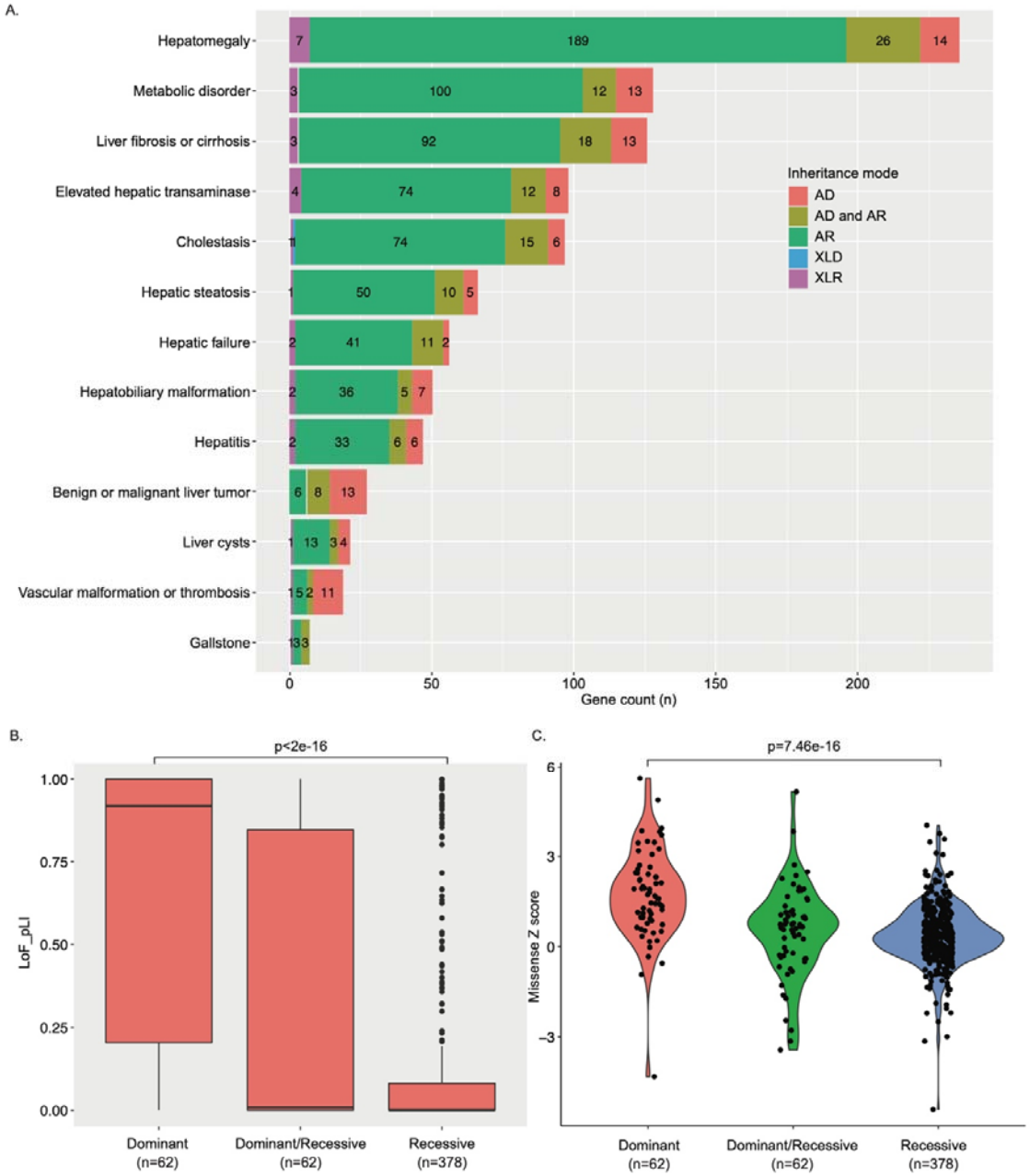

Figure 2. Genetic diagnosis and clinical implication of WES findings in the CLD cohort

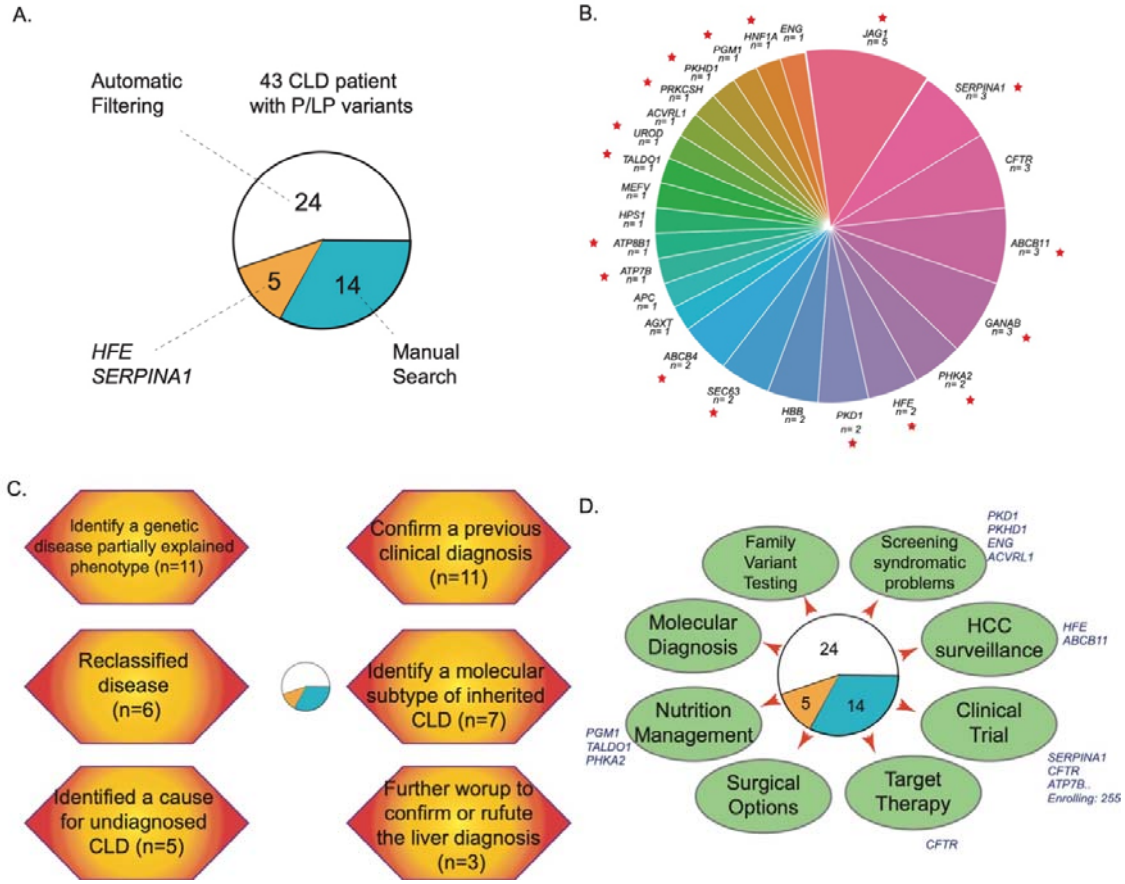

Figure 3. A summary of the genetic analytic strategy and outcomes for chronic liver diseases

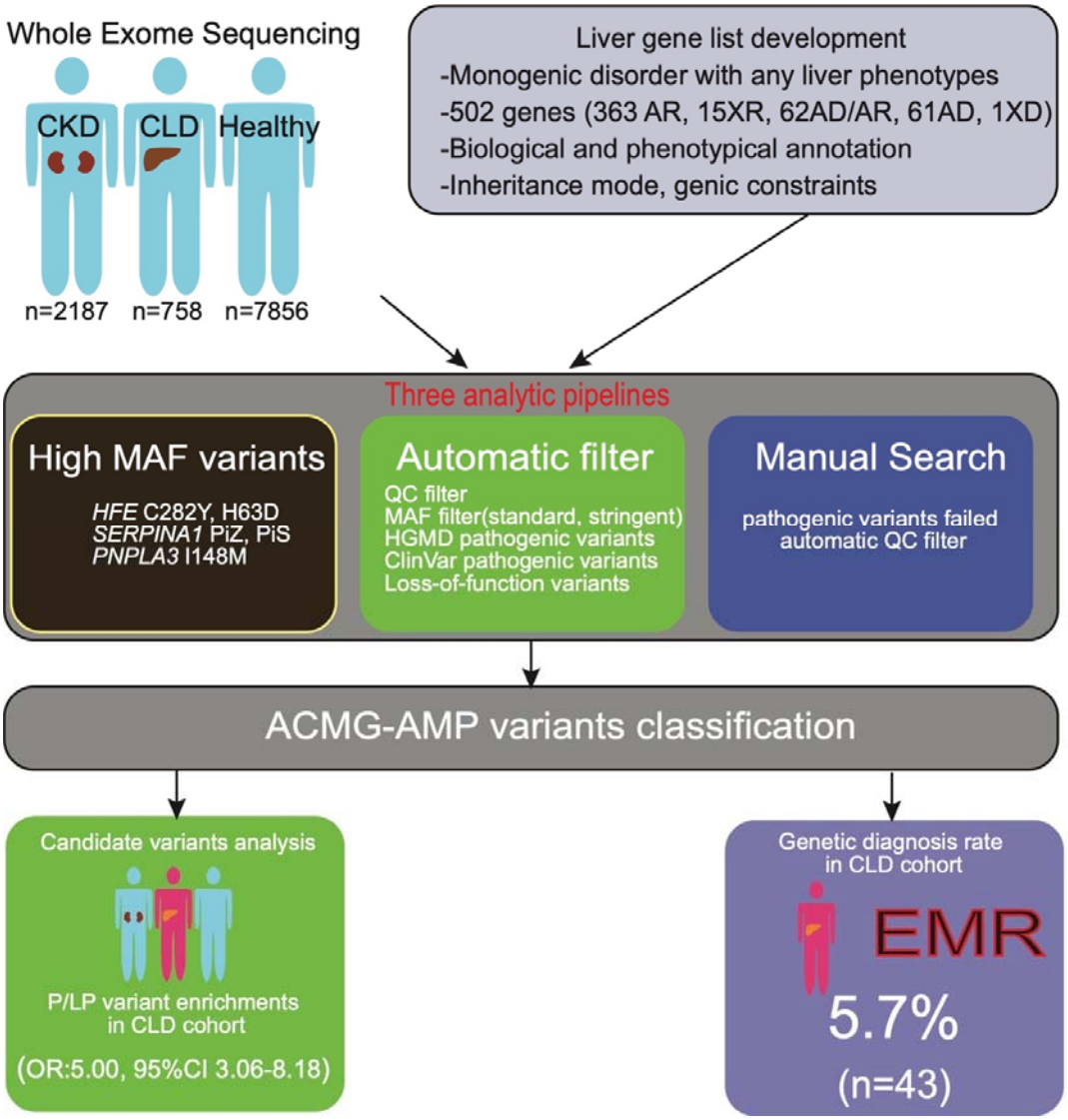

Table 1: Clinical characteristics for monogenic diagnosis in the liver cohort from WES analysis

Characteristics	Total	Nongenetic Dx(n, %)	Monogenic Dx(n, %)	P value
Total number of individuals	758	715(94.3%)	43(5.7%)	
Gender				
Male	351	328(93.4%)	23(6.6%)	0.415
Female	407	387(95.1%)	20(4.9%)	
Age group				
0-21 yr	255	232(91.0%)	23(9.0%)	0.019
22-44 yr	150	141(94.0%)	9(6.0%)	
45-64 yr	226	218(96.5%)	8(3.5%)	
≥ 65 yr	127	124(96.9%)	3(2.4%)	
Self-declared race/ethnicity				
White	367	349(95.1%)	18(4.9%)	0.903
Hispanic	136	128(94.1%)	8(5.9%)	
Black	95	88(92.6%)	7(7.4%)	
Asia	61	57(93.4%)	4(6.6%)	
Other or unspecified	99	93(93.9%)	6(6.1%)	
Primary liver diagnosis				
Metabolic/Congenital	17	7(41.2%)	10(58.8%)	6.48e-19
NAFLD/NASH	182	174(95.6%)	8(4.1%)	
AIH/PBC/PSC	128	122(85.3%)	6(4.7%)	
Abnormal LFT	52	47(90.4%)	5(9.6%)	
Biliary atresia	76	73(96.1%)	3(3.9%)	
Other	303	292(96.4%)	11(3.6%)	

*chi-square test was used to compare the difference between two groups, either negative or positive genetic diagnosis through WES.

Table 2: *HFE* and *SERPINA1* variants in three cohorts

Genotypes	HC (n=7856)	CKD (n=2187)	CLD (n=758)
***HFE* variants**			
Homozygous C282Y	2	1[$]	1[$]
Homozygous H63D	9	3[&]	0
Compound C282Y/H63D*	60	11	2
C282Yor H63D/PTV	22	5	1[$]
Total(n=117)	93 (1.20%)	20 (0.91%)	4(0.53%)
***SERPINA1* variants:**			
Pi ZZ	0	0	3[$]
Pi MZ (n=222)	138 (1.76%)	49 (2.24%)	15 (1.98%)
Pi SS	1	0	0
Pi MS(n=606)	482 (6.14%)	92 (4.21%)	32 (4.22%)
Pi Z/ Pi S* (n=12)	7	5	0

Individuals with a homozygous pathogenic variant, or two heterozygous pathogenic variants reported in HGMD or ClinVar; * phase was not determined to evaluate if these two variants are *in-trans*, or *cis*. *$:* cases with a sufficient clinical evidence of liver phenotypes consistent with the genetic diagnosis. One case with H63D has liver phenotypes consistent with hemochromatosis.