https://hhh.k8s.dev.maayanlab.cloud/

Under-studied Genes Likely Associated with Hepatitis D

Abinanda Prabhakaran*

Abstract

Hepatitis D (HDV) infection, a severe form of chronic viral hepatitis, lacks curative therapies and its host-genetic determinants remain incompletely characterized. To uncover understudied host genes potentially involved in HDV pathology, we integrated disease-associated gene sets from Enrichr, RummaGEO, Rummagene, Geneshot, MONDO, DO, the GWAS Catalog and ClinVar, and ranked genes by their prevalence in these sets versus their PubMed publication counts. This approach highlighted ten understudied candidates with high disease-set frequency yet few publications: *IFIH1*, *MYOM2*, *THPO*, *NR0B1*, *EIF2AK2*, *GPT2*, *ERVW-1*, *SLC10A1*, *PPIA* and *FBL*. A complementary analysis employed the Gene Set Foundation Model (GSFM) to augment known HDV genes and predict additional relevant genes; the top ten GSFM-ranked, low-publication genes were *IFNA13*, *IFNA4*, *IFNA5*, *IFNA7*, *B3GAT1*, *DPEP1*, *IFNA17*, *HAVCR1*, *CLEC4M* and *IFNA16*. Differential expression profiling of the representative GEO dataset GSE245916 (HDV patient versus healthy liver samples) identified the up-regulated genes *LGALS3BP* and *HLA-DRB4*, and confirmed dysregulation of several understudied candidates, thereby providing an initial experimental foothold. Together, this computational pipeline prioritizes a set of neglected host factors for functional validation—through CRISPR screens, interaction mapping, patient-cohort analyses and drug-repurposing—to advance mechanistic understanding and therapeutic development for hepatitis D.

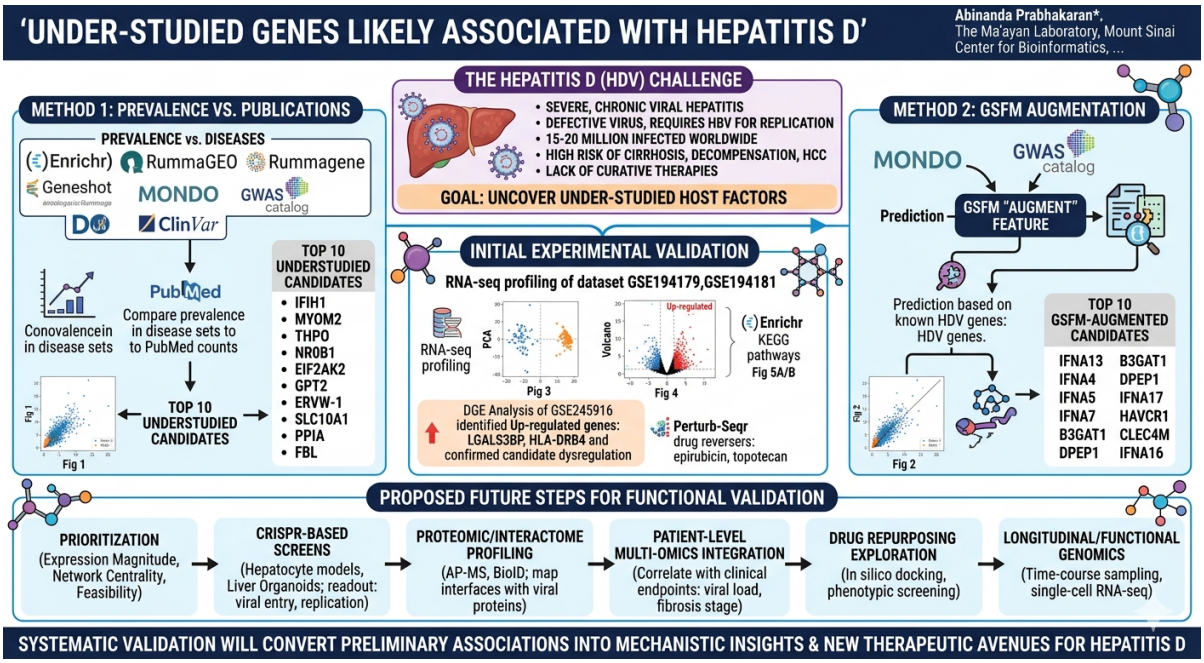
**The Ma'ayan Laboratory, Mount Sinai Center for Bioinformatics, Department of Pharmacological Sciences, Windreich Department of Artificial Intelligence and Human Health, Icahn School of Medicine at Mount Sinai, New York, NY 10029, USA.*

1. Introduction

Hepatitis D virus (HDV) is a small, defective, circular single-stranded RNA virus that can replicate only in the presence of hepatitis B virus (HBV), which supplies the envelope proteins required for virion assembly and transmission [1–4]. Worldwide, more than 15–20 million individuals are co-infected with HDV and HBV, making HDV the most severe form of chronic viral hepatitis [1, 2]. Chronic HDV infection is associated with rapid progression to cirrhosis, hepatic decompensation and hepatocellular carcinoma (HCC), and it carries a higher risk of liver-related mortality than HBV monoinfection [5–10].

Epidemiological surveys have highlighted a heterogeneous global distribution of HDV. Meta-analyses estimate a prevalence of HDV infection among HBsAg-positive individuals ranging from 5 % to 15 % and a total burden of approximately 15 million infections worldwide [11–13]. While prevalence is declining in some historically endemic regions, recent data show a resurgence in northern and central Europe, largely driven by immigration [14–16]. Genotypic diversity further complicates the epidemiology, with at least eight reported genotypes displaying distinct geographic patterns and variable pathogenicity [1, 17, 18].

At the molecular level, entry of both HBV and HDV into hepatocytes is mediated by the sodium taurocholate cotransporting polypeptide (NTCP), a liver-specific bile-acid transporter that binds the myristoylated pre-S1 domain of the HBV large envelope protein [19]. Structural studies have shown that this interaction overlaps with heparan sulfate proteoglycan binding sites on the viral envelope, and that mutations in NTCP can abrogate viral entry while preserving bile-acid transport [20–23]. These insights have spurred the development of entry-inhibitors such as Myrcludex B (bulevirtide), which blocks the pre-S1/NTCP interaction [24, 25].



Clinically, HDV coinfection markedly accelerates liver disease progression. Longitudinal cohorts demonstrate a higher incidence of cirrhosis, earlier hepatic decompensation and an increased risk of HCC compared with HBV monoinfection [5–8]. The presence of HDV also worsens outcomes in patients with other viral hepatitis or HIV coinfections, underscoring its role as a driver of severe liver pathology [8, 9]. Current therapeutic options are limited; pegylated interferon-α remains the only approved antiviral, achieving sustained virologic response in only about 25 % of treated patients [26–28]. Recent clinical trials have evaluated novel agents, including the entry inhibitor bulevirtide, the prenylation inhibitor lonafarnib, and nucleic-acid polymers such as REP2139, either alone or in combination with interferon, showing promising antiviral activity and improved biochemical outcomes [24, 29–33]. Nevertheless, durable viral clearance and functional cure remain elusive, highlighting the need for continued research into HDV biology, epidemiology, and therapeutic strategies.

2. Methods

2.1 Detailed introduction on the disease from DeepDive2.0

The introduction section of this Hepatitis Dreport was generated using the DeepDive2.0 pipeline. First, query NCBI PubMed search with the Hepatitis Dterm, we get top 50 highly cited publications and then have the LLM summarize these top 50 highly cited abstracts for input disease term. The introduction for the Hepatitis Dcontains valid citations to these top 50 articles used to write the introduction section of the report that describes the current knowledge about the disease.

2.2 Understudied genes by observed gene prevalence in disease gene sets

The method of ranking genes by their prevalence takes understudied genes from the collection of disease gene sets from resources - Enrichr [34], RummaGEO [35], Rummagene [36], Geneshot [37], MONDO [38], DO [39], GWAS Catalog [40] and ClinVar [41], and compares the gene occurrence in these sets with the number of publications per gene from PubMed [42]. From all the disease-associated genes extracted for the disease, we count their number of publications in PubMed using the NCBI E-utilities API [43]. To extract publication counts for each gene, returned publications from each search are filtered to only consider PubMed IDs where the gene appears in either the title or the abstract. A scatter plot is created for Hepatitis D, displaying publication counts vs. frequency of the genes considering all disease gene sets. The highlighted understudied genes are those with fewer publications than the median of all disease gene publication counts, and the top 10 genes ranked by their frequency in Hepatitis Dgene sets.

2.3 Understudied genes predicted using gene set foundation model

Another approach to rank understudied genes for Hepatitis Dis the utility of the "augment" feature of Gene Set Foundation Model (GSFM) [44]. Gene sets for Hepatitis Dextracted from MONDO [38] and GWAS catalog [40] were submitted to GSFM for augmentation. The genes from these resources contain the known disease-gene. GSFM [44] uses this information to predict additional genes ranked by the model's

scores. The predicted genes are filtered by the genes with fewer publications and ranked by the GSFM score to select the top 10 understudied genes for Hepatitis D.

2.4 Differential gene expression analysis of a GEO study

To better understand the potential role of the understudied genes in the context of the disease, we perform differential gene expression (DGE) analysis by selecting a representative GEO study related to the disease. From all GEO studies with up and down signatures obtained by querying the RummaGEO resource [35], we picked one GEO study for downstream analysis and inclusion in the reports (Table 1). We then performed DGE analysis [45] on the gene expression data for the study. Statistically significant up and down regulated genes were identified by comparing a group of healthy controls to disease samples using limma-voom [46, 47]. Significantly expressed genes are determined by a p-value of <0.05 and the direction of regulation or increase/decrease in expression is determined by the log2FC to separate the up and down gene sets. These up and down gene sets are then given as separate inputs to Enrichr [34] for enrichment analysis with the KEGG pathways [48] library. In addition, the up and down genes are submitted to Perturb-Seqr [49] to identify drugs that may reverse the disease condition towards the normal state of gene expression. The top ranked understudied genes from each method that are also differentially expressed in the GEO study for the disease are identified. As part of each GEO study analysis and disease report, PCA plots [50] and volcano plots are added to the enrichment bar plots from the Enrichr analysis [34], and the drug predictions from Perturb-Seqr [49].

2.5 Constructing disease reports using the gpt-oss:120b LLM model

Reports are constructed for each disease with abstract, introduction, results, methods, discussion, acknowledgements and references sections. The abstract, introduction and discussion sections are generated by prompting the gpt-oss:120b model, while the other sections are created with templates and custom code. The report starts with an introduction section that provides a summary with verified citations about the disease using the DeepDive2.0 pipeline. DeepDive2.0 queries the disease term in PubMed, and then uses an LLM to summarize the top 50 most cited articles that mentioned the disease. Followed by the results section that highlights the understudied genes for each disease found by both the methods discussed above, and the findings from the DGE analysis. A graphical abstract for each report is generated using the Gemini-3.1 flash image model by uploading the entire report to Gemini manually. So, all of the figures along with the text are then made into a LaTex bundle to produce the final disease report.

2.6 Generating videos for disease reports using Paper2Video pipeline

To generate videos with generated slides and narration, we used Paper2Video pipeline to automatically convert research publication-like report (given as LaTeX bundle) into a narrated video presentation. For each disease report LaTeX bundle, along with the given reference speaker image, and a short reference audio sample, the system uses LLMs to generate slides, synthesizes per-slide speech via voice cloning (F5-TTS) to produce a final MP4 with subtitles.

3. Results

After extracting gene sets for Hepatitis Dfrom various resources including Enrichr [34], RummaGEO [35], Rummagene [36], Geneshot [37], MONDO [38], DO [39], GWAS Catalog [40] and ClinVar [41], we try to identify those genes that are understudied for Hepatitis Dwith fewer publications on PubMed [42].

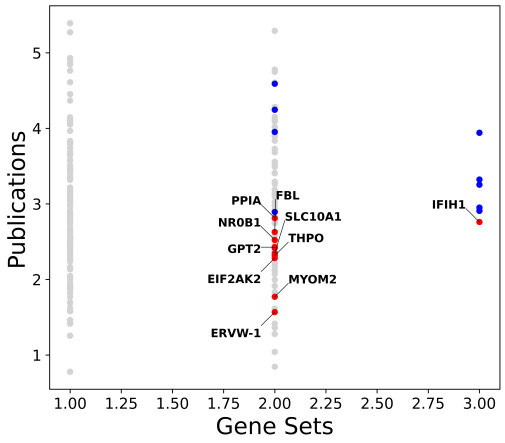


Figure 1. Scatterplot of publication counts vs gene set counts across only Hepatitis Dgene sets for each of the Hepatitis Dgenes. Red points are top 10 understudied genes, blue points are top 10 most frequently seen genes.

We plot publication counts and gene set counts for each Hepatitis Dgene using only the Hepatitis Ddisease gene sets (Figure 1). The points in red signify top 10 understudied genes with fewer publications and high frequency in Hepatitis Dgene sets, while the blue points are top 10 frequently appearing genes in the Hepatitis Dgene sets. The top 10 understudied genes for Hepatitis Dare *IFIH1, MYOM2, THPO, NR0B1,*

EIF2AK2, GPT2, ERVW-1, SLC10A1, PPIA, FBL.

Another approach to get understudied genes for disease could be to use GSFM model [44] to augment the disease genes for Hepatitis Dfrom MONDO [38] and Gwas Catalog [40] resources and get unknown highly related genes for Hepatitis D.

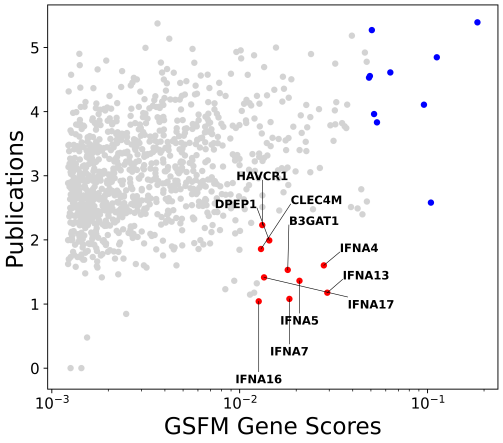


Figure 2. Scatterplot of publication counts vs GSFM gene scores for each of the predicted Hepatitis Dgenes. Red points are top 10 understudied genes with high GSFM scores but fewer publications, blue points are top 10 genes with high GSFM scores.

We plot publication counts and GSFM gene scores for each of the predicted Hepatitis Dgenes from GSFM by augmenting its disease genes (Figure 2). The red points are top 10 genes with fewer publications and high GSFM scores that are not in the input Hepatitis Ddisease genes, while the blue points are top 10 genes that have high GSFM scores. The top 10 understudied genes with high GSFM scores are *IFNA13, IFNA4, IFNA5, IFNA7, B3GAT1, DPEP1, IFNA17, HAVCR1, CLEC4M, IFNA16*.

These understudied genes identified might play a unexplored critical role in the pathology of Hepatitis Dthat should be analyzed further through valid scientific RNA -seq experiments that knockout the genes in the healthy vs Hepatitis Ddisease samples.

To understand the role these understudied genes play in Hepatitis Dpathology, we can find GEO studies where some of these genes are significantly up or down regulated. Using RummaGEO [35], we can get these differentially expressed gene signatures related to Hepatitis D. Out of all the published GEO stuies for Hepatitis Dqueried using RummaGEO [35], we perform differential expression analysis on only one selected representative GEO study for Hepatitis D(Table 1).

Differential Gene Expression (DGE) [45] analysis for the GEO study reveals the up and down regulated differentially expressed genes between two conditions control vs disease samples.

For Hepatitis DGEO study GSE194179,GSE194181, raw counts data can be downloaded from NCBI FTP server or from ARCHS4 [51] platform that contains uniformly processed counts data available for all human and mouse GEO studies. To explore the similarity of biological samples in RNA-seq dataset, we apply Principal Component Analysis (PCA) [50] and the scatterplot of the first two Principal Components (PCs) of the transformed gene expression data available for the samples considered for the analysis (Figure 3). To perform DGE analysis, limma-voom [46, 47] technique is appiled to this raw counts data after clear case and control samples are identified for the study. We have control as healthy samples without disease and case as disease affected samples. Identify differentially expressed genes (DEGs) by P-value <0.05 and direction of regulation with log2FC >1 as up regulated and log2FC <-1 as down regulated differentially expressed genes for control vs disease samples. A volcano plot shows the DEGs identified for GSE194179,GSE194181 study (Figure 4).

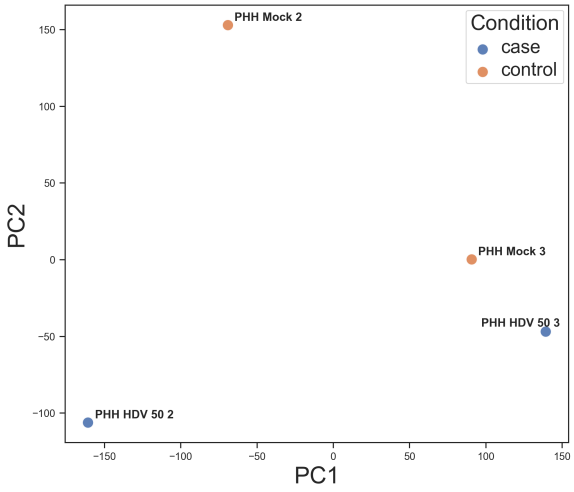


Figure 3. PCA plot of control and disease samples from the GEO study GSE194179,GSE194181. Blue points are control samples and orange points are disease samples. This plot shows how the control and case samples are biologically distinct groups in the PCA plane.

Genes are significantly down regulated in Hepatitis Dsamples compared to healthy ones. While are up regulated.

For the list of up and down regulated genes we can then perform enrichment analysis using Enrichr API [34] to get enriched terms with these DEGs as input queries (Figure 5A, Figure 5B).

Using both the up and down genes, we obtain drugs reversers from Perturb-Seqr [49] that reerses the disease

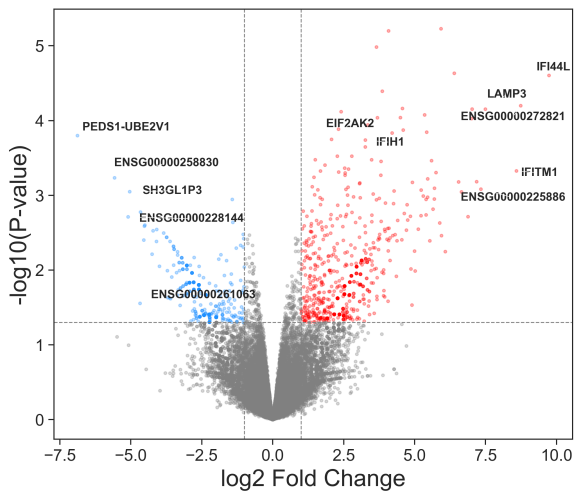


Figure 4. Volcano plot of P-value and Log2FC on the limma-voom results for the GEO study for differentially expressed Hepatitis Dgenes.

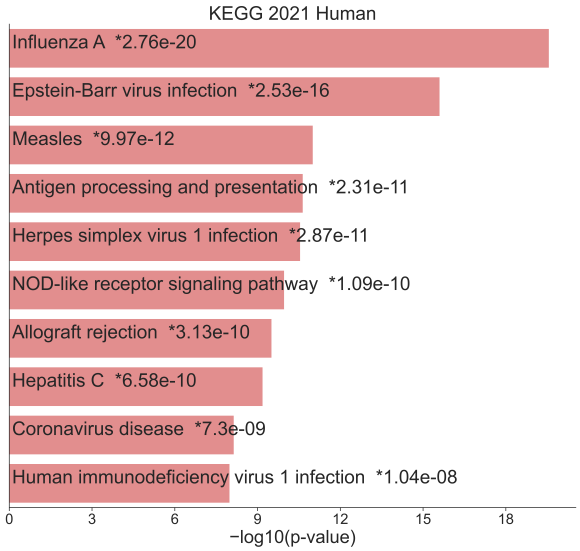


Figure 5. Bar chart of top enriched terms from the KEGG_2021_Human gene set library. The top 10 enriched terms for the input up gene set are displayed based on the -log10(p-value), with the actual p-value shown next to each term. The term at the top has the most significant overlap with the input up gene set.

gene signatures queried, with details of the predicted drugs or chemicals (Table 2).

4. Discussion

The present study leveraged large-scale disease-gene resources and a gene-set foundation model to pinpoint a set of understudied genes that recurrently appear in hepatitis-D–related gene collections yet have received limited attention in the literature. By intersecting publication frequency with gene-set prevalence (Method 1) and with model-derived relevance scores (Method 2), we identified two complementary panels of candidate genes (e.g., *IFIH1*, *MYOM2*, *THPO*, *NR0B1*, *EIF2AK2*, *GPT2*, *ERVW-1*, *SLC10A1*, *PPIA*, *FBL* and *IFNA13*,

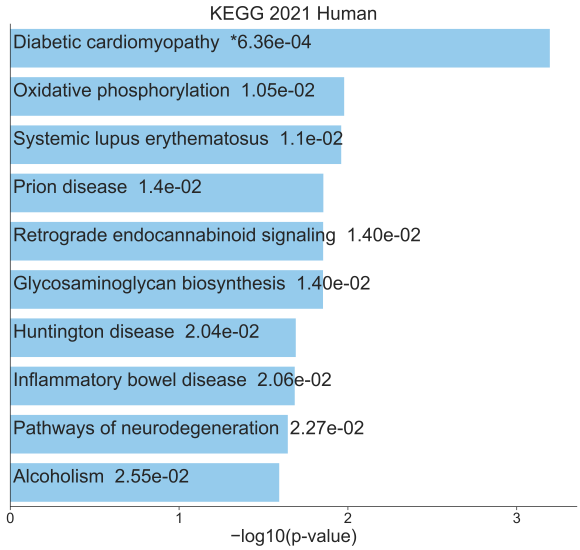


Figure 6. Bar chart of top enriched terms from the KEGG_2021_Human gene set library. The top 10 enriched terms for the input down gene set are displayed based on the -log10(p-value), with the actual p-value shown next to each term. The term at the top has the most significant overlap with the input down gene set.

IFNA4, *IFNA5*, *IFNA7*, *B3GAT1*, *DPEP1*, *IFNA17*, *HAVCR1*, *CLEC4M*, *IFNA16*). While the bulk of the analysis was computational, the downstream differential expression profiling in a representative GEO dataset demonstrated that several of these genes are indeed dysregulated in hepatitis-D patient samples, providing an initial experimental foothold.

Prioritisation for functional validation

The immediate next step is to prioritize the candidate list for experimental interrogation. A pragmatic approach combines three criteria: (i) magnitude of differential expression in patient-derived RNA-seq data, (ii) network centrality within hepatitis-D-pertinent pathways (e.g., innate immune signaling, viral entry, and interferon response), and (iii) feasibility of genetic manipulation in hepatocyte models. Genes such as *IFIH1* and *EIF2AK2*, which encode key pattern-recognition receptors and kinases in the antiviral response, rank highly on all three axes and thus merit early focus.

CRISPR-based loss- and gain-of-function screens

To directly assess causal involvement, CRISPR-Cas9 knockout and CRISPRa activation screens can be deployed in hepatocyte-derived cell lines (e.g., HepG2-NTCP) infected with HDV pseudovirions. Parallel screens in primary human hepatocytes or liver organoids would capture context-specific effects. Readouts should include viral entry efficiency, replication competence (HDV RNA levels), and downstream host transcrip-

GSE Series	Title	Direction	Species	Samples	Genes
GSE194179,GSE194181	Study of HBV and HDV transcriptomic modulation in HepaRG and PHHs	↑	human	21	1322
GSE194179,GSE194181	Study of HBV and HDV transcriptomic modulation in HepaRG and PHHs	↓	human	21	1180

Table 1. RummaGEO differential expression signatures for Hepatitis D

perturbation	adjPvalue	oddsRatio	approved
epirubicin	4.300942384021754e-16	389.825576	True
topotecan	6.671402232713488e-06	153.475421	True
docetaxel	1.0	0.000000	True
nicotinamide	1.0	0.000000	True
dasatinib	1.0	0.000000	True
ibrutinib	1.0	0.000000	True
pimavanserin	1.0	0.000000	True
lapatinib	1.0	0.000000	True
olaparib	1.0	0.000000	True
propofol	1.0	0.000000	True

Table 2. Drug reversers from Perturb-Seqr using up and down gene set search

tional changes. Hits that modulate these phenotypes will provide mechanistic links between the understudied genes and HDV biology.

Proteomic and interactome profiling

Given that many candidates encode proteins with limited functional annotation, affinity-purification mass spectrometry (AP-MS) or proximity-labeling (BioID/TurboID) experiments can map their interaction partners in the hepatic environment. Overlaying these interaction networks with known HDV proteins (e.g., HDAg, L-HBs) may uncover direct or indirect viral–host interfaces, suggesting novel therapeutic targets.

Integration with patient-level multi-omics

Beyond in-vitro assays, validation in clinical cohorts is essential. Publicly available hepatitis-D transcriptomic and proteomic datasets (e.g., from liver biopsies or serum proteomics) should be mined to confirm consistent dysregulation of the candidate genes across independent cohorts. Correlating gene expression with clinical endpoints—viral load, liver fibrosis stage, and treatment response—will help stratify which genes have prognostic or predictive value.

Exploration of drug repurposing opportunities

The Perturb-Seqr analysis highlighted several FDA-approved compounds that reverse the disease signature. For understudied genes that encode druggable proteins (e.g., kinases, transporters), in silico docking and chemoinformatics pipelines can be applied to identify existing molecules that modulate their activity. Subsequent phenotypic screening of these compounds in HDV infection models will test whether targeting the newly implicated genes yields antiviral effects.

Longitudinal and functional genomics studies

Future work should incorporate longitudinal sampling from patients undergoing antiviral therapy (e.g., bulevirtide) to monitor dynamic changes in the expression of these genes. Coupling such time-course data with single-cell RNA-seq will resolve cell-type specific regulation and uncover potential compensatory pathways that emerge upon treatment.

Conclusion

Collectively, the computational pipeline presented here has surfaced a cadre of understudied genes that are plausibly linked to hepatitis-D pathogenesis. Systematic functional validation—through CRISPR screens, interaction mapping, patient-cohort analyses, and drug repurposing studies—will be pivotal in converting these preliminary associations into mechanistic insights and, ultimately, novel therapeutic avenues for a disease that currently lacks curative options.

Acknowledgements

This manuscript used assistance from the Ollama gpt-oss:120b large language model and DeepDive2.0 resource.

References

[1] Hughes S A, Wedemeyer H, and Harrison P M. Hepatitis delta virus. *The Lancet*, 378(9785), 2011. doi:10.1016/S0140-6736(10)61931-9.

[2] Sureau C and Negro F. The hepatitis delta virus: Replication and pathogenesis. *Journal of Hepatology*, 64(1), 2016. doi:10.1016/j.jhep.2016.02.013.

[3] Rizzetto M. Hepatitis d: Thirty years after. *Journal of Hepatology*, 50(5), 2009. doi:10.1016/j.jhep.2009.01.004.

[4] Farci P and Niro G. Clinical features of hepatitis d. *Seminars in Liver Disease*, 32(03), 2012. doi:10.1055/s-0032-1323628.

[5] Fattovich G, Giustina G, Christensen E, et al. Influence of hepatitis delta virus infection on morbidity and mortality in compensated cirrhosis type b. *Gut*, 46(3), 2000. doi:10.1136/gut.46.3.420.

[6] Romeo R, Del Ninno E, Rumi M, et al. A 28-year study of the course of hepatitis infection: A risk factor for cirrhosis and hepatocellular carcinoma. *Gastroenterology*, 136(5), 2009. doi:10.1053/j.gastro.2009.01.052.

[7] Su C, Huang Y, Huo T, et al. Genotypes and viremia of hepatitis b and d viruses are associated with outcomes of chronic hepatitis d patients. *Gastroenterology*, 130(6), 2006. doi:10.1053/j.gastro.2006.01.035.

[8] Sagnelli E, Coppola N, Scolastico C, et al. Virologic and clinical expressions of reciprocal inhibitory effect of hepatitis b, c, and delta viruses in patients with chronic hepatitis. *Hepatology*, 32 (5), 2000. doi:10.1053/jhep.2000.19288.

[9] Pineda J A, García-García J A, Aguilar-Guisado M, et al. Clinical progression of hepatitis c virus–related chronic liver disease in human immunodeficiency virus–infected patients undergoing highly active antiretroviral therapy. *Hepatology*, 46(3), 2007. doi:10.1002/hep.21757.

[10] Brunetto M R, Ricco G, Negro F, et al. Easl clinical practice guidelines on hepatitis delta virus. *Journal of Hepatology*, 79(2), 2023. doi:10.1016/j.jhep.2023.05.001.

[11] Stockdale A J, Kreuels B, Henrion M Y R, et al. The global prevalence of hepatitis d virus infection: Systematic review and meta-analysis. *Journal of Hepatology*, 73(3), 2020. doi:10.1016/j.jhep.2020.04.008.

[12] Chen H, Shen D, Ji D, et al. Prevalence and burden of hepatitis d virus infection in the global population: a systematic review and meta-analysis. *Gut*, 68(3), 2018. doi:10.1136/gutjnl-2018-316601.

[13] Miao Z, Zhang S, Ou X, et al. Estimating the global prevalence, disease progression, and clinical outcome of hepatitis delta virus infection. *The Journal of Infectious Diseases*, 221(10), 2019. doi:10.1093/infdis/jiz633.

[14] Wedemeyer H and Manns M P. Epidemiology, pathogenesis and management of hepatitis d: update and challenges ahead. *Nature Reviews Gastroenterology & Hepatology*, 7(1), 2010. doi:10.1038/nrgastro.2009.205.

[15] Gaeta G B, Stroffolini T, Chiaramonte M, et al. Chronic hepatitis d: A vanishing disease? an italian multicenter study. *Hepatology*, 32(4), 2000. doi:10.1053/jhep.2000.17711.

[16] Lavanchy D. Worldwide epidemiology of hbv infection, disease burden, and vaccine prevention. *Journal of Clinical Virology*, 34, 2005. doi:10.1016/s1386-6532(05)00384-7.

[17] Radjef N, Gordien E, Ivaniushina V, et al. Molecular phylogenetic analyses indicate a wide and ancient radiation of african hepatitis delta virus, suggesting a deltavirus genus of at least seven major clades. *Journal of Virology*, 78(5), 2004. doi:10.1128/jvi.78.5.2537-2544.2004.

[18] Le Gal F, Gault E, Ripault M, et al. Eighth major clade for hepatitis delta virus. *Emerging Infectious Diseases*, 12(7), 2006. doi:10.3201/eid1209.060112.

[19] Yan H, Zhong G, Xu G, et al. Sodium taurocholate cotransporting polypeptide is a functional receptor for human hepatitis b and d virus. *eLife*, 1, 2012. doi:10.7554/eLife.00049.

[20] Urban S, Bartenschlager R, Kubitz R, et al. Strategies to inhibit entry of hbv and hdv into hepatocytes. *Gastroenterology*, 147(1), 2014. doi:10.1053/j.gastro.2014.04.030.

[21] Yan H, Peng B, Liu Y, et al. Viral entry of hepatitis b and d viruses and bile salts transportation share common molecular determinants on sodium taurocholate cotransporting polypeptide. *Journal of Virology*, 88(6), 2014. doi:10.1128/JVI.03478-13.

[22] Sureau C and Salisse J. A conformational heparan sulfate binding site essential to infectivity overlaps with the conserved hepatitis b virus a-determinant. *Hepatology*, 57(3), 2013. doi:10.1002/hep.26125.

[23] Yan H, Peng B, He W, et al. Molecular determinants of hepatitis b and d virus entry restriction in mouse sodium taurocholate cotransporting polypeptide. *Journal of Virology*, 87(14), 2013. doi:10.1128/JVI.03540-12.

[24] Bogomolov P, Alexandrov A, Voronkova N, et al. Treatment of chronic hepatitis d with the entry inhibitor myrcludex b: First results of a phase ib/iia study. *Journal of Hepatology*, 65(3), 2016. doi:10.1016/j.jhep.2016.04.016.

[25] Blank A, Markert C, Hohmann N, et al. First-in-human application of the novel hepatitis b and hepatitis d virus entry inhibitor myrcludex b. *Journal of Hepatology*, 65(3), 2016. doi:10.1016/j.jhep.2016.04.013.

[26] Wedemeyer H, Yurdaydìn C, Dalekos G N, et al. Peginterferon plus adefovir versus either drug alone for hepatitis delta. *New England Journal of Medicine*, 364(4), 2011. doi:10.1056/NEJMoa0912696.

[27] Castelnau C, Le Gal F, Ripault M, et al. Efficacy of peginterferon alpha-2b in chronic hepatitis delta: Relevance of quantitative rt-pcr for follow-up. *Hepatology*, 44(3), 2006. doi:10.1002/hep.21325.

[28] Niro G A, Ciancio A, Gaeta G B, et al. Pegylated interferon alpha-2b as monotherapy or in combination with ribavirin in chronic hepatitis delta. *Hepatology*, 44(3), 2006. doi:10.1002/hep.21296.

[29] Koh C, Canini L, Dahari H, et al. Oral prenylation inhibition with lonafarnib in chronic hepatitis d infection: a proof-of-concept randomised, double-blind, placebo-controlled phase 2a trial. *The Lancet Infectious Diseases*, 15(10), 2015. doi:10.1016/S1473-3099(15)00074-2.

[30] Bazinet M, Pântea V, Cebotarescu V, et al. Safety and efficacy of rep 2139 and pegylated interferon alfa-2a for treatment-naive patients with chronic hepatitis b virus and hepatitis d virus co-infection (rep 301 and rep 301-ltf): a non-randomised, open-label, phase 2 trial. *The Lancet Gastroenterology & Hepatology*, 2(12), 2017. doi:10.1016/S2468-1253(17)30288-1.

[31] Wedemeyer H, Aleman S, Brunetto M R, et al. A phase 3, randomized trial of bulevirtide in chronic hepatitis d. *New England Journal of Medicine*, 389(1), 2023. doi:10.1056/NEJMoa2213429.

[32] Bordier B B, Ohkanda J, Liu P, et al. In vivo antiviral efficacy of prenylation inhibitors against hepatitis delta virus. *Journal of Clinical Investigation*, 112(3), 2003. doi:10.1172/JCI17704.

[33] Urban S, Neumann-Haefelin C, and Lampertico P. Hepatitis d virus in 2021: virology, immunology and new treatment approaches for a difficult-to-treat disease. *Gut*, 70(9), 2021. doi:10.1136/gutjnl-2020-323888.

[34] Xie Z et al. Gene set knowledge discovery with Enrichr. *Current Protocols*, 1(3), 2021. doi:10.1002/cpz1.90. URL https://maayanlab.cloud/Enrichr/.

[35] Marino GB et al. RummaGEO: Automatic mining of human and mouse gene sets from GEO. *Patterns*, 5(10), 2024. doi:10.1016/j.patter.2024.101072. URL https://rummageo.com/.

[36] Clarke DJB et al. Rummagene: massive mining of gene sets from supporting materials of biomedical research publications. *Communications Biology*, 7 (1), 2024. doi:10.1038/s42003-024-06177-7. URL https://rummagene.com/.

[37] Lachmann A et al. Geneshot: search engine for ranking genes from arbitrary text queries. *Nucleic Acids Research*, 47(W1), 2019. doi:10.1093/nar/gkz393. URL https://maayanlab.cloud/geneshot/.

[38] Vasilevsky NA et al. Mondo: Integrating disease terminology across communities. *Genetics*, 232 (4), 2025. doi:10.1093/genetics/iyaf215. URL https://mondo.monarchinitiative.org/.

[39] Schriml LM et al. The human disease ontology 2022 update. *Nucleic Acids Research*, 50(D1), 2021. doi:10.1093/nar/gkab1063. URL https://www.disease-ontology.org/.

[40] Sollis E et al. The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucleic Acids Research*, 51(D1), 2022. doi:10.1093/nar/gkac1010. URL https://www.ebi.ac.uk/gwas/.

[41] Landrum MJ et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, 42 (D1), 2013. doi:10.1093/nar/gkt1113. URL https://www.ncbi.nlm.nih.gov/clinvar/.

[42] Canese K and Weis S. PubMed: the bibliographic database. https://pubmed.ncbi.nlm.nih.gov/, 2013. Bethesda (MD): National Center for Biotechnology Information (US).

[43] Sayers EW et al. Database resources of the national center for biotechnology information in 2025. *Nucleic Acids Research*, 53(D1):D20–D29, 2024. doi:10.1093/nar/gkae979.

[44] Clarke DJB et al. A gene set foundation model pre-trained on a massive collection of diverse gene sets. *Preprint on bioRxiv*, 2025. doi:10.1101/2025.05.30.657124. URL https://gsfm.maayanlab.cloud/.

[45] Robinson MD et al. edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–140, 2010. doi:10.1093/bioinformatics/btp616.

[46] Ritchie ME et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Research*, 43(7):e47, 2015. doi:10.1093/nar/gkv007.

[47] Law CW et al. voom: precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biology*, 15(2), 2014. doi:10.1186/gb-2014-15-2-r29.

[48] Kanehisa M and other. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1):D353–D361, 2016. doi:10.1093/nar/gkw1092.

[49] Gardner JK et al. Perturb-Seqr [internet]. https://perturbseqr.maayanlab.cloud/, 2026.

[50] Abdi H, Williams L J, et al. Principal component analysis. *WIREs Computational Statistics*, 2(4): 433–459, 2010. doi:10.1002/wics.101.

[51] Lachmann A et al. Massive mining of publicly available RNA-seq data from human and mouse. *Nature Communications*, 9(1), 2018. doi:10.1038/s41467-018-03751-6. URL `https://archs4.org/`.