An Exploratory Genome-Wide Analysis of Genetic Risk for Alcoholic Hepatitis

James J Beaudoin[1], **Nanye Long**[1], **Suthat Liangpunsakul**[2], **Puneet Puri**[3], **Patrick S Kamath**[4], **Vijay Shah**[4], **Arun J Sanyal**[3], **David W Crabb**[2], **Naga P Chalasani**[2,*], **Thomas J Urban**[1], and **the TREAT Consortium**

[1]Division of Pharmacotherapy and Experimental Therapeutics, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC, USA

[2]Division of Gastroenterology/Hepatology, Indiana University School of Medicine, Indianapolis, IN, USA

[3]Division of Gastroenterology, Hepatology and Nutrition, Virginia Commonwealth University, Richmond, VA, USA

[4]Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA

Abstract

Objectives—To elucidate the genetic variability between heavy drinkers with and without alcoholic hepatitis.

Materials and methods—An exploratory genome-wide association study (GWAS; NCT02172898) was conducted comparing 90 alcoholic hepatitis cases with 93 heavy drinking matched controls without liver disease in order to identify variants or genes associated with risk for alcoholic hepatitis. Individuals were genotyped using the Multi-Ethnic Genotyping Array, after which the data underwent conventional quality control. Using bioinformatics tools, pathways associated with alcoholic hepatitis were explored on the basis of individual variants, and based on genes with a higher "burden" of functional variation.

Results—Although no single variant reached genome-wide significance, an association signal was observed for *PNPLA3* rs738409 (p = 0.01, OR 1.9, 95% CI 1.1 – 3.1), a common single nucleotide polymorphism that has been associated with a variety of liver-related pathologies including alcoholic cirrhosis. Using the Improved-Gene Set Enrichment Analysis for GWAS tool it

[*]Correspondence to: Naga P Chalasani, Professor, Division of Gastroenterology/Hepatology, Indiana University School of Medicine, 702 Rotary Circle, Suite 225, Indianapolis, IN 46202, USA. nchalasa@iu.edu; Telephone: +1 (317) 278-0414; Fax: +1 (317) 278-1949.

James J Beaudoin (jbeaudoin@unc.edu), Nanye Long (nanye.long@gmail.com), Thomas J Urban (holdem.callfield@gmail.com), Division of Pharmacotherapy and Experimental Therapeutics, Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, NC 27599, USA

Suthat Liangpunsakul (sliangpu@iu.edu), David W Crabb (dcrabb@iu.edu), Naga P Chalasani (nchalasa@iu.edu), Division of Gastroenterology/Hepatology, Indiana University School of Medicine, Indianapolis, IN 46202, USA

Puneet Puri (ppuri@mcvh-vcu.edu), Arun J Sanyal (asanyal@mcvh-vcu.edu), Division of Gastroenterology, Hepatology and Nutrition, Virginia Commonwealth University, Richmond, VA 23298, USA

Patrick S Kamath (kamath.patrick@mayo.edu), Vijay Shah (Shah.Vijay@mayo.edu), Division of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN 55905, USA

was shown that, based on the single variants' trait-association p-values, multiple pathways were associated with risk for alcoholic hepatitis with high confidence (FDR < 0.05), including several pathways involved in lymphocyte activation and chemokine signaling, which coincides with findings from other research groups. Several Tox Functions and Canonical Pathways were highlighted using Ingenuity Pathway Analysis, with an especially conspicuous role for pathways related to ethanol degradation, which is not surprising considering the phenotype of the genotyped individuals.

Conclusion—This preliminary analysis suggests a role for *PNPLA3* variation and several gene sets/pathways that may influence risk for alcoholic hepatitis among heavy drinkers.

Keywords

alcoholic hepatitis; genome-wide association study; genetic risk; single nucleotide polymorphism; pathway enrichment

Introduction

Alcoholic hepatitis (AH) is an illness marked by liver inflammation triggered by excessive alcohol consumption over a prolonged period of time. It may be rapidly fatal and is often associated with fibrosis and the development of cirrhosis [1], which is life-threatening and can require liver transplantation [2]. Interestingly, only a minority of heavy drinkers develop alcoholic hepatitis and this differential risk is poorly understood [3]. One approach to understanding this difference is to examine the genetic variability between heavy drinkers with and without alcoholic hepatitis. In the current exploratory study, we describe a genome-wide association study (GWAS) of alcoholic hepatitis. Although analysis of the AH-associated signal for individual variants was one of the objectives, it is clear that this preliminary study is underpowered to solely attempt to obtain a genome-wide significant signal for a disease like alcoholic hepatitis. With the currently available sample size, we were more interested in identifying gene sets and pathways associated with AH, than individual variants. Analysis of our GWAS data sheds some light on the genetic predisposition to, and biology of, alcoholic hepatitis.

Materials and Methods

Study Cohort

The study was conducted on a cohort consisting of 90 AH cases and 93 heavy drinking controls without liver diseases who were of European descent and were enrolled prospectively into the ongoing TREAT Alcoholic Hepatitis Prospective Study (NCT02172898; registered at ClinicalTrials.gov). Alcoholic hepatitis was defined based on clinical, biochemical and histological (wherever appropriate) criteria as described in an earlier publication from the TREAT consortium [4]. This definition is largely consistent with a recent consensus definition proposed by the NIAAA Alcoholic Hepatitis Consortium Investigators [5]. Based on the history, individuals with AH were deemed to not have drug induced liver injury as a reason for their liver disease presentation. Heavy drinking controls were those with average daily alcohol consumption similar to that of AH cases but had normal AST or ALT (< 50 U/L), total bilirubin (< 1 mg/dL), platelet count (> 140,000/mm^3)

and lacked physical examination evidence for significant alcoholic liver disease such as hepatosplenomegaly or ascites. Abdominal ultrasound or elastography was not performed on controls. All subjects signed the informed consent and a local institutional review board at each site approved the study protocol.

Demographic data, past medical history and routine blood tests were collected prospectively on all consented participants. The quantity and pattern of alcohol consumption were determined using the Alcohol Use Disorders Identification Test (AUDIT), and the Time Line Follow-Back (TLFB).

Genotyping

The Multi-Ethnic Genotyping Array (MEGA) (Illumina, San Diego, CA) was used to genotype individuals according to the manufacturer's protocol. Genotyping was performed at the Mammalian Genotyping Core, University of North Carolina at Chapel Hill.

Data Quality Control

Our initial study population of 196 subjects of European ancestry was genotyped on the Illumina MEGA beadchip. Before quality control (QC), the raw number of single nucleotide polymorphisms (SNPs) was 1,705,969. SNPs were excluded based on the following criteria: 1) missing genotyping rate > 0.01 both SNP-wise and sample-wise; 2) SNP missing rate associated with case/control status (p < 0.0001); 3) monomorphic SNPs. Individual subjects were excluded if: 1) the inbreeding coefficient was outside of the normal range; 2) identity by descent (IBD) values between any two individuals was > 0.2; or 3) they were outliers on principal component analysis (PCA) (defined as subjects whose values on any of the top 10 PCs exceed 6 times the standard deviation). A detailed breakdown of the QC process is depicted in the Figure of Supplemental Digital Content 1. The final analysis-ready dataset contained 183 subjects (93 controls and 90 cases) and 819,401 SNPs.

Whole Genome Association Analysis

Single-variant tests for association were performed using an allelic Fisher's exact test as implemented in PLINK v1.07 [6] employing the --fisher option. A standard threshold for statistical significance was applied ($p < 5 \times 10^{-8}$).

Collapsing Test of Genes

In the collapsing test of each gene, we collapsed all functional (defined as those with putative impact being "high" or "moderate" in SNPEff's annotation system; see http://snpeff.sourceforge.net/VCFannotationformat_v1.0.pdf for details) and rare (defined as those with minor allele frequency (MAF)   1% in the 93 controls) variants by creating a binary indicator variable for each subject, which is equal to 1 if the individual carries any rare variant in that gene and 0 otherwise. Fisher's exact test was then used to test the association between case/control status and "burden" of rare variants carried by a gene. From a total of 22,222 unique Ensembl IDs, 10,429 had qualifying variants based on aforementioned inclusion criteria and were thus tested. Note that a gene is ineligible for the test if no functional variants in the gene are observed in our data, or if their MAF in controls > 1%.

Using Bonferroni correction for multiple testing, and $\alpha = 0.05$, the threshold for statistical significance in the gene-wise tests was $p < 4.8 \times 10^{-6}$.

Enrichment Test of a Gene Set

From the output of the gene collapsing test, we obtained a ranked list of p-values for all the genes tested. For a given gene set of interest (e.g., a particular pathway), we tested whether members of it are randomly distributed along the ranked list or tend to concentrate on the lower end (i.e., those showing stronger association with the case/control status). To this end, we applied the methodology of Gene Set Enrichment Analysis (GSEA) [7] by calculating an enrichment score (ES) for the given gene set, followed by estimating its statistical significance (nominal p-value) through a permutation test. We implemented GSEA in R following the procedure outlined in the Appendix of Subramanian et al., 2005, with 300 permutations.

Obtaining Lists of SNPs Associated with Alcohol-Related Diseases

We used the GWAS Catalog at https://www.ebi.ac.uk/gwas to find SNPs shown to be associated with other alcohol-related diseases. The SNP Annotation and Proxy Search tool at https://www.broadinstitute.org/mpg/snap/ldsearch.php was used to find single variants in high linkage disequilibrium (LD) with SNPs of interest. Lists of proxies were generated using the 1000 Genomes Pilot 1 SNP data set, the CEU (Northern Europeans from Utah) population panel, an r^2 threshold of 0.8, and a distance limit of 500kb. The Batch Coordinate Conversion (liftOver) tool at https://genome.ucsc.edu/cgi-bin/hgLiftOver was used to convert genome coordinates across different genome builds.

Pathway Analysis Tools

Improved-Gene Set Enrichment Analysis for Genome-Wide Association Study (i-GSEA4GWAS) was used to obtain pathway information on as many variants and genes in our dataset as possible [8]. The disease association p-value of each variant and gene is incorporated into i-GSEA4GWAS's analysis. The i-GSEA4GWAS tool implements GSEA through permutation of SNP labels, and uses an improved version of GSEA by focusing on gene sets with a large proportion of significant genes. In QIAGEN's Ingenuity® Pathway Analysis (IPA®, QIAGEN Redwood City, www.qiagen.com/ingenuity), we tested gene lists using a p-value threshold of 0.05 or 0.5. We considered experimentally observed, direct and indirect relationships in humans, from all of IPA's data sources. The p-value associated with a function or pathway in IPA, calculated using the right-tailed Fisher Exact Test, indicates the likelihood that the function or pathway is associated with a set of focus genes due to random chance. In g:Profiler [9] and Gene Ontology enRIchment anaLysis and visuaLizAtion (GOrilla) [10] we uploaded our dataset of ranked genes using a trait association p-value cutoff of 0.05, and no cutoff, respectively. Both GOrilla and g:Profiler produce effective graphical representations, but the former generates output images based on enrichment of GO terms, whereas the latter performs functional profiling, biomolecule integration tasks and genomic data mining using additional databases such as those from KEGG, Reactome, GEO and BioGrid. The Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.7 [11] was used to gather additional information on our list of genes with a p-value < 0.05.

Results

We genotyped a sample of 90 alcoholic hepatitis cases and 93 heavy drinking controls (see Table 1 for demographics and clinical characteristics) of European ancestry for approximately 1.7 million SNPs using the MEGA array. Based on this sample of 183 subjects, no single genetic marker showed genome-wide significant association with alcoholic hepatitis (see Figure 1 for Manhattan plot; and Figure 2 for Quantile-Quantile plot). Although no single common or rare variant reached genome-wide significance, we proceeded to perform an exploratory analysis on the dataset of this GWAS of alcoholic hepatitis.

Immune-Mediated and Inflammatory Processes

I-GSEA4GWAS was used to assess whether gene sets are enriched on the basis of the p-values of individual variants (see Tables a & b, Supplemental Digital Content 2, for detailed results; and Table 2a & b for condensed versions). Many immune-mediated and inflammatory processes appear to be associated with risk for alcoholic hepatitis with high confidence when assessing all SNPs within a range of 500kb upstream and downstream of genes. These processes include gene sets related to chemokines such as the *CXCR4 Pathway*, and the *CCR3 Pathway* (each with $p < 0.001$; and False discovery rate (FDR) 0.012). It has previously been shown that, in contrast to circulating neutrophils, infiltrating neutrophils express receptors for these chemokines [12]. Alcoholic hepatitis is characterized by neutrophils infiltrating the liver, and the regulation of these chemokine pathways may play a role in that regard. In addition, we observed enrichment of lymphocyte-related gene sets including *Positive Regulation of T Cell Activation*, *Regulation of Lymphocyte Activation*, *Regulation of T Cell Activation*, *Positive Regulation of Lymphocyte Activation*, *Positive Regulation of T Cell Proliferation*, *T Cell Proliferation*, and *Regulation of T Cell Proliferation* (each with $p < 0.001$; and $FDR < 0.041$); and the autophagy-related gene set *HSA04140 Regulation of Autophagy* ($p < 0.001$; $FDR < 0.008$). The observations by other research groups that cytotoxic T lymphocytes have impaired cytotoxicity and reduced activation in alcoholic hepatitis [13], and that alcohol-induced liver diseases are associated with upregulation of autophagy in hepatocytes [14] may be related to the enrichment of these pathways. In addition, the *Nuclear Chromosome Part* gene set was found to be associated with risk for alcoholic hepatitis with high confidence when evaluating all SNPs within a range of 500kb upstream and downstream of genes ($p < 0.001$; $FDR < 0.001$), or assessing only functional variants (e.g., nonsynonymous SNPs; those inducing a frame shift; those affecting an essential splice site) ($p < 0.001$; $FDR = 0.049$). Although the *Nuclear Replication Fork* gene set is associated with the trait with statistical significance when only considering functional SNPs ($p < 0.001$; $FDR = 0.032$), the gene set loses its significance after multiple testing correction when SNPs within a range of 500kb on either side of genes are taken into account ($p = 0.007$; $FDR = 0.201$). A total of 21 additional gene sets are enriched with high confidence based on SNPs within a range of 500kb upstream and downstream of genes. Masking all the genes in the extended region of the major histocompatibility complex (MHC) had a negligible effect on these associations.

Comparison of Single Variant Data with Other Alcohol-Related GWASs

GWASs on other alcohol-related traits may hold clues about the genetic nature of alcoholic hepatitis, and/or of excessive alcohol consumption without developing alcoholic hepatitis (i.e., the control population in the current study). Using the NHGRI-EBI GWAS Catalog (at https://www.ebi.ac.uk/gwas) a list of all genome-wide significant SNPs associated with an alcohol-related phenotype was compiled, and subsequently divided into a list with pathological and a list with behavioral traits (see Tables a & b, Supplemental Digital Content 3). A total of 24 unique SNPs were associated with behavioral phenotypes (e.g., alcohol consumption, alcohol dependence, alcoholism), and two SNPs with pathological phenotypes (pancreatitis). The 105 genome-wide significant SNPs identified in a recently published alcoholic cirrhosis GWAS [15] were not yet listed in the GWAS Catalog at the time of writing this manuscript - these were manually added to the list of SNPs associated with alcohol-related pathological traits. To increase the number of SNPs with which we can make comparisons between our dataset and the findings from other studies, we included variants in high LD with an r^2 value of at least 0.8 (see Tables a & b, Supplemental Digital Content 4).

Pathology-Associated SNPs

We compared the strength of SNP association between our alcoholic hepatitis cases versus controls, against the results seen in other studies which are similarly comparing patients with organ-related illnesses with controls (although no SNP in our study reached genome-wide significance, and only SNPs which reached at least genome-wide significance were selected from other studies; see Table b, Supplemental Digital Content 3). From the list of alcohol pathology-associated SNPs we compared, the variants with the strongest signal in our study were all at the *PNPLA3* locus, of which many, such as the three with the highest signal (rs2294915, rs3747207, and rs738409) are in high LD (see Table b, Supplemental Digital Content 4). Previous studies have shown that one of these three top variants, rs738409 (I148M), is an important risk variant for alcoholic cirrhosis, and this has recently been confirmed by a large alcoholic cirrhosis GWAS [15]. We were able to find an association of this SNP with alcoholic hepatitis as well (p = 0.01, OR 1.9, 95% CI 1.1 – 3.1), although our sample size was too small to show genome-wide significance. The overall MAF of rs738409 was 0.339 in cases, and 0.215 in controls (see Table 3). Furthermore, there was no significant difference in rs738409 MAF (p-value > 0.05) between the 71 subjects with AH who survived compared to the 19 who died during the 12 month follow-up period (see Table 4).

Alcohol-Related Behavior-Associated SNPs

Next, we took advantage of the fact that both our alcoholic hepatitis cases and the matched controls are heavy alcohol users, and therefore share a behavioral trait. This allowed us to genetically compare the 183 heavy alcohol users as a whole with subjects in the control dataset from the Wellcome Trust Case Control Consortium (WTCCC) (http://www.wtccc.org.uk) GWAS, with the assumption that the WTCCC has a negligible proportion of heavy alcohol users. Only those SNPs which were studied in both our dataset and the one from the WTCCC were compared. None of the variants reached statistical

significance at $\alpha = 0.05$ or 0.10 (see Figure, Supplemental Digital Content 5, for a Quantile-Quantile plot; and Table a, Supplemental Digital Content 3). When including variants in high LD, we find three proxies of rs2168784 with a p-value between 0.05 and 0.10 (see Figure, Supplemental Digital Content 6, for a Quantile-Quantile plot; and Table a, Supplemental Digital Content 4). We were not able to compare rs2168784 directly between our sample and the WTCCC, because it was not genotyped on the MEGA array.

Gene-Based Tests for Burden of Functional Variation

Apart from focusing on individual variants, we performed gene-based tests to elucidate the "burden" of functional variation in patients with alcoholic hepatitis compared with matched controls. For these tests we only included rare variants (i.e., those with a MAF 1% in the controls) with obvious effects on protein structure or function (i.e., missense, nonsense and splice site variants). The rare variants in the current study map to a total of 10,429 unique Ensembl IDs (corresponding to 10,407 HUGO gene names) with qualifying variants (see Methods), 6,268 of which (60.1%) had a trait association p-value < 1 (see Table, Supplemental Digital Content 7). No single gene showed significant association after multiple testing correction, however, interestingly, among the top associated genes was SLC22A1, which is known to play a role in lipid metabolism in the liver [16].

Using various bioinformatics tools we explored which pathways may help characterize the differences between alcoholic hepatitis cases and matched controls from our dataset. Similar to the gene set enrichment assessment on the basis of the p-values of individual variants, we used i-GSEA4GWAS based on p-values of genes, where lower p-values imply a larger difference in being affected by rare variants between the two groups of subjects. As shown in the Table of Supplemental Digital Content 8, although some individual gene sets have an enrichment p-value 0.002, after multiple testing correction these gene sets are regarded as possibly associated with traits (i.e., FDR < 0.25), but none with statistical significance (i.e., FDR < 0.05). Masking an extended region of MHC again had negligible effect on the results.

Tox Functions

Using IPA we were able to identify several Tox Functions and Canonical Pathways, as well as various other signaling pathways – each of which might explain some of the differences between alcoholic hepatitis subjects and matched controls. In addition to using the IPA database as a reference set when running our analyses, we evaluated our dataset using itself as a reference. Since about half of all our original list of genes (22,222 Ensembl IDs, corresponding to 20,844 HUGO gene names) did not contain qualifying variants, the IPA database likely overestimates the degree of enrichment of pathways and gene sets. Using a relatively stringent association test p-value threshold of 0.05 and using the user dataset, or IPA database as reference, Liver Hyperplasia/Hyperproliferation was among highly enriched Tox Functions (see Figures, Supplemental Digital Content 9 & 10, respectively). Gene sets associated with Cardiac Arrhythmia, Kidney Failure, Liver Cirrhosis, Liver Fibrosis, Pulmonary Hypertension, Liver Inflammation/Hepatitis, and Liver Steatosis also reached statistical significance in our dataset.

Ethanol and Other Degradation Pathways

Canonical Pathway enrichment became apparent only after loosening the inclusion criterion for the individual genes. By specifying the threshold of the alcoholic hepatitis association test p-value to be 0.5, nearly 5,000 unique genes are included in the core analysis, the maximum number of genes IPA recommends users to analyze for statistical reasons. Using the user dataset as reference, three out of the top four Canonical Pathways were involved in Ethanol Degradation, which might be an indication that alcohol metabolism plays an important role in describing differences between the cases and controls in our subject sample (see Figure, Supplemental Digital Content 11). A total of 84.2%, 90.0% and 83.3% of genes comprising the Ethanol Degradation II, Oxidative Ethanol Degradation III, and Ethanol Degradation IV Canonical Pathways, respectively, were on the list of genes below the p-value cut-off (see Table, Supplemental Digital Content 12). Degradation of Noradrenaline and Adrenaline, Dopamine, Histamine, Putrescine, Tryptophan, Methionine, and Serotonin are among enriched Canonical Pathways as well. Using the IPA database as reference, the results look similar, albeit not identical (see Figure, Supplemental Digital Content 13).

To test the validity of these findings, we performed a disease/control relabeling permutation test of ten runs and assessed to what extent the aforementioned enrichments reappeared. Although the abovementioned pathways reached statistical significance in our true dataset, many permutations of the dataset reached statistical significance for the Tox Functions as well (see Figures, Supplemental Digital Content 9 & 10). None of the permutations reached statistical significance for any of the Ethanol Degradation Canonical Pathways, but some other Canonical Pathways were also enriched in one or more of our permutations (see Figures, Supplemental Digital Content 11 & 13). Although many permuted result sets did not reach the same level of significance as the true disease/control populations for the aforementioned pathways, the fact that some permutations did reach statistical significance suggests that no strong claims about the relevance of those pathways should be made in the context of alcoholic hepatitis.

Some support to the claim that there might be a true difference in enriched pathways between alcoholic hepatitis cases and matched controls, comes from the observation that each of the ten permutations had fewer genes which had an association test p-value < 0.5, implying that relabeling permutations may simply tend to dampen true biological differences between these populations.

Potential Enrichment of Additional Pathways

Examining pathway enrichment of gene classes that may contribute to risk for alcoholic hepatitis using the bioinformatics tools g:Profiler (see Figure, Supplemental Digital Content 14), GOrilla (see Figures, Supplemental Digital Content 15, 16 & 17), and DAVID (see Tables a & b, Supplemental Digital Content 18) revealed the potential importance of a multitude of additional pathways. Gene sets related to the nuclear pore complex assembly, and alpha-actinin binding were shown to be enriched using g:Profiler and GOrilla. Apart from that, gene sets related to signaling, the cilium axoneme, and microtubule-associated components, among others, reappeared in the results of both GOrilla and DAVID.

Discussion

This pilot report describes a GWAS of alcoholic hepatitis, but due to its limited sample size serves only as a proof of concept study of this understudied disease. This research was designed to find any kind of (preliminary) signal for individual variants, and to find potentially involved genes and pathways, which we were able to find. Although no individual SNP or gene was found to have a genome-wide significant association with the traits under study, various interesting observations were made which might prove important to the understanding of the underlying genetics of alcoholic hepatitis.

Since alcoholic hepatitis can result in cirrhosis, it would be prudent to evaluate whether specific variants are associated with risk for both phenotypes. Interestingly, the pivotal common SNP rs738409 in a recent alcoholic cirrhosis GWAS [15] was among the most strongly associated SNPs in our study as well, indicating that this may be a common risk factor for both alcoholic hepatitis and cirrhosis; in fact, it may suggest that *PNPLA3* is associated with alcoholic cirrhosis via conferring increased sensitivity of the patient to alcoholic hepatitis. The current exploratory study provides optimism for expanded genome-wide studies to be performed as additional patients are enrolled in and complete the TREAT trial. The fact that we see some evidence of association for rs738409, despite the currently small sample size, suggests that this common SNP may be among the strongest genetic risk factors for alcoholic hepatitis, as it is for alcoholic cirrhosis. However, rs738409 was far from being the SNP with the strongest signal in the current study, and it might turn out that other SNPs play a more important role in alcoholic hepatitis. It is furthermore important to note that we suspect a majority of AH cases to have significant underlying fibrotic disease, but without a biopsy it is difficult to be precise about the prevalence of underlying cirrhosis among AH cases. In the United States, routine liver biopsy is not standard of care in patients with AH, except when there is an uncertainty about the diagnosis, and thus a systematic histological characterization of our cases is lacking. This could have influenced the results in the current study.

By analyzing our dataset on a gene-based level, by only including variants which had a MAF 1% in our control population, we were able to elucidate which genes have the highest "burden" of functional variation in patients with alcoholic hepatitis compared with matched controls. Depending on which trait-association p-value cut-off we chose in pathway analysis tools, we obtained several interesting findings. For instance, we found several ethanol degradation pathways to be enriched in our dataset, which seems to agree with the traits under study. Similarly, the chemokine and lymphocyte activation pathways highlighted by iGSEA4GWAS are consistent with our current understanding of the role of infiltrating neutrophils and lymphocyte activity in hepatic inflammation.

The tools IPA, g:Profiler, GOrilla and DAVID showed enrichment of additional gene sets, some of which overlapped. An abundance of gene sets are highlighted by the various enrichment analysis methods used in the current study. Only after showing that enrichment of these pathways can be replicated in a new and/or larger cohort of alcoholic hepatitis patients and heavy drinking controls can we be confident that any of these are truly biologically important to alcoholic hepatitis.

The ultimate goal of this pilot study was to find preliminary risk factors for alcoholic hepatitis, but apart from detecting a possible association of AH with *PNPLA3*, none of the reported GWAS results are statistically significant. The main limitation of the current study is clearly the small sample size, and thus future work with a larger number of subjects and the use of imputation approaches to increase the power to detect associations, as well as the inclusion of a validation cohort and/or the performance of appropriate functional assays, will be required in order to establish definite risk factors. While subjects are continuously being enrolled in the TREAT trial, it will take many more years before a study can be performed that is sufficiently powered to obtain results that can be clinically significant. However, the preliminary results found here might be reproduced with more power when such a study is possible, which will give those results increasing support. Finally, making these results available can benefit future research of this understudied disease, such as meta-analysis efforts.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Appendix

Additional information: Members of the TREAT Consortium

Indiana University, Indianapolis, IN, USA:

David Crabb, MD, Naga Chalasani, MD, Suthat Liangpunsakul, MD, Barry Katz, PhD, Spencer Lourens, PhD, Andy Borst, BS, Ryan Cook, MPH, Andy Qigui Yu, PhD, David Nelson, PhD, Romil Saxena, MD, Sherrie Cummings, RN, Megan Comerford, BS, Lakye Edwards, BS

Mayo Clinic, Rochester, MN, USA:

Vijay Shah, MD, Gregory Gores, MD, Patrick Kamath, MD, Vikas Verma, PhD, Sarah Wilder, RN, BSN, Amy Olofson, RN, Amanda Schimek

Virginia Commonwealth University, Richmond, VA, USA:

Arun Sanyal, MD, Puneet Puri, MD, Susan Walker, RN, MSN

NIAAA:

Project Scientist – Svetlana Radeava, PhD

Program Official – Andras Orosz, PhD

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

References

1. Gao B, Bataller R. Alcoholic liver disease: pathogenesis and new therapeutic targets. Gastroenterology. 2011; 141:1572–85. [PubMed: 21920463]

2. Louvet A, Labreuche J, Artru F, Boursier J, Kim DJ, O'Grady J, Trepo E, Nahon P, Ganne-Carrie N, Naveau S, Diaz E, Gustot T, Lassailly G, Cannesson-Leroy A, Canva-Delcambre V, Dharancy S, Park SH, Moreno C, Morgan TR, Duhamel A, Mathurin P. Combining Data From Liver Disease Scoring Systems Better Predicts Outcomes of Patients With Alcoholic Hepatitis. Gastroenterology. 2015; 149:398–406 e8. quiz e16–7. [PubMed: 25935634]

3. Sozio MS, Liangpunsakul S, Crabb D. The role of lipid metabolism in the pathogenesis of alcoholic and nonalcoholic hepatic steatosis. Semin Liver Dis. 2010; 30:378–90. [PubMed: 20960377]

4. Liangpunsakul S, Puri P, Shah VH, Kamath P, Sanyal A, Urban T, Ren X, Katz B, Radaeva S, Chalasani N, Crabb DW, Translational R. Evolving Alcoholic Hepatitis Treatment C. Effects of Age, Sex, Body Weight, and Quantity of Alcohol Consumption on Occurrence and Severity of Alcoholic Hepatitis. Clin Gastroenterol Hepatol. 2016

5. Crabb DW, Bataller R, Chalasani NP, Kamath PS, Lucey M, Mathurin P, McClain C, McCullough A, Mitchell MC, Morgan TR, Nagy L, Radaeva S, Sanyal A, Shah V, Szabo G, Consortia NAH. Standard Definitions and Common Data Elements for Clinical Trials in Patients With Alcoholic Hepatitis: Recommendation From the NIAAA Alcoholic Hepatitis Consortia. Gastroenterology. 2016; 150:785–90. [PubMed: 26921783]

6. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, de Bakker PI, Daly MJ, Sham PC. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet. 2007; 81:559–75. [PubMed: 17701901]

7. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005; 102:15545–50. [PubMed: 16199517]

8. Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. Nucleic Acids Res. 2010; 38:W90–5. [PubMed: 20435672]

9. Reimand J, Arak T, Vilo J. g:Profiler--a web server for functional interpretation of gene lists (2011 update). Nucleic Acids Res. 2011; 39:W307–15. [PubMed: 21646343]

10. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: a tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009; 10:48. [PubMed: 19192299]

11. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. [PubMed: 19131956]

12. Hartl D, Krauss-Etschmann S, Koller B, Hordijk PL, Kuijpers TW, Hoffmann F, Hector A, Eber E, Marcos V, Bittmann I, Eickelberg O, Griese M, Roos D. Infiltrated neutrophils acquire novel chemokine receptor expression and chemokine responsiveness in chronic inflammatory lung diseases. J Immunol. 2008; 181:8053–67. [PubMed: 19017998]

13. Stoy S, Dige A, Sandahl TD, Laursen TL, Buus C, Hokland M, Vilstrup H. Cytotoxic T lymphocytes and natural killer cells display impaired cytotoxic functions and reduced activation in patients with alcoholic hepatitis. Am J Physiol Gastrointest Liver Physiol. 2015; 308:G269–76. [PubMed: 25501547]

14. Dolganiuc A, Thomes PG, Ding WX, Lemasters JJ, Donohue TM Jr. Autophagy in alcohol-induced liver diseases. Alcohol Clin Exp Res. 2012; 36:1301–8. [PubMed: 22551004]

15. Buch S, Stickel F, Trepo E, Way M, Herrmann A, Nischalke HD, Brosch M, Rosendahl J, Berg T, Ridinger M, Rietschel M, McQuillin A, Frank J, Kiefer F, Schreiber S, Lieb W, Soyka M, Semmo N, Aigner E, Datz C, Schmelz R, Bruckner S, Zeissig S, Stephan AM, Wodarz N, Deviere J, Clumeck N, Sarrazin C, Lammert F, Gustot T, Deltenre P, Volzke H, Lerch MM, Mayerle J, Eyer F, Schafmayer C, Cichon S, Nothen MM, Nothnagel M, Ellinghaus D, Huse K, Franke A, Zopf S, Hellerbrand C, Moreno C, Franchimont D, Morgan MY, Hampe J. A genome-wide association study confirms PNPLA3 and identifies TM6SF2 and MBOAT7 as risk loci for alcohol-related cirrhosis. Nat Genet. 2015; 47:1443–8. [PubMed: 26482880]

16. Chen L, Shu Y, Liang X, Chen EC, Yee SW, Zur AA, Li S, Xu L, Keshari KR, Lin MJ, Chien HC, Zhang Y, Morrissey KM, Liu J, Ostrem J, Younger NS, Kurhanewicz J, Shokat KM, Ashrafi K, Giacomini KM. OCT1 is a high-capacity thiamine transporter that regulates hepatic steatosis and is a target of metformin. Proc Natl Acad Sci U S A. 2014; 111:9983–8. [PubMed: 24961373]

Alcoholic hepatitis GWAS

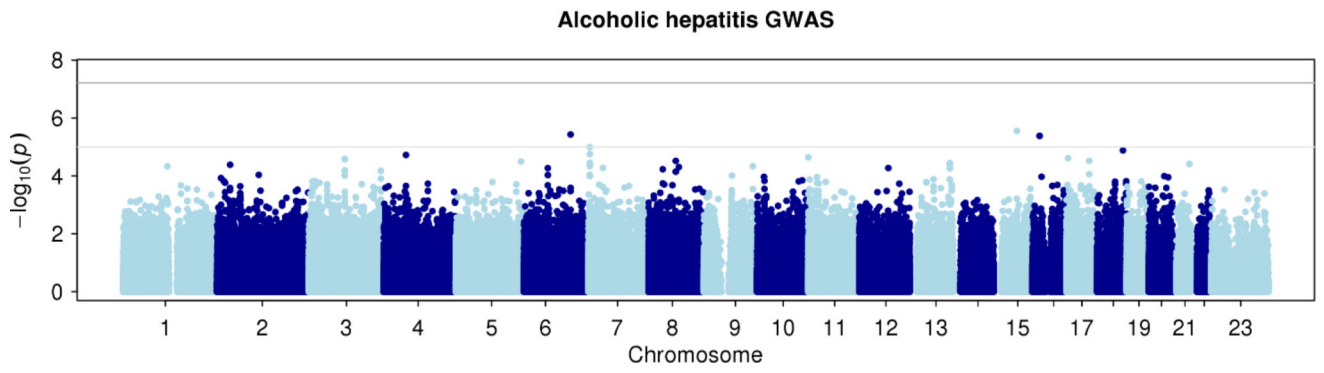**Figure 1. Manhattan plot for the current alcoholic hepatitis genome-wide association study**
None of the single nucleotide polymorphisms reached genome-wide significance after Bonferroni multiple testing correction (dark grey line). The light grey line represents a p-value of 1×10^{-5}.
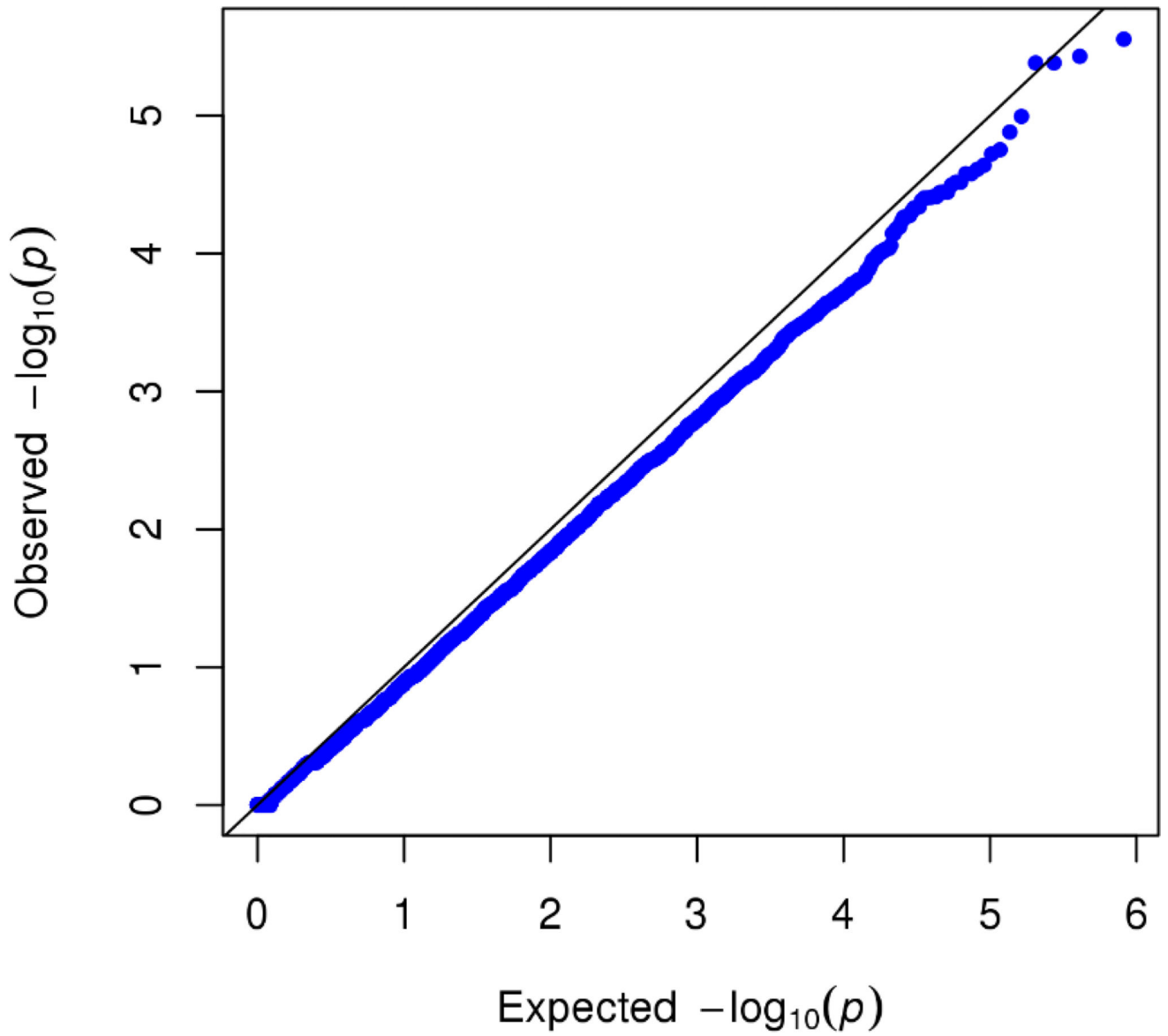
Alcoholic hepatitis GWAS

Figure 2. Quantile-quantile plot for the alcoholic hepatitis genome-wide association study

Table 1

Demographics of alcoholic hepatitis (AH) patients and heavy drinking matched controls without hepatitis.

Variables	Controls (n = 93)	AH cases (n = 90)	p-value
Age (years)	44.6 ± 12.2	46.8 ± 11.0	0.21
Men, *n* (%)	56 (60%)	56 (62%)	0.78
BMI (kg/m^2)	29.1 ± 7.3	29.6 ± 8.3	0.71
WBC (10^3 cells/mm^3)	7.2 ± 2.8	11.1 ± 8.1	< 0.0001
Hemoglobin (g/dl)	13.3 ± 1.7	10.1 ± 1.9	< 0.0001
Platelet count (10^3 cells/mm^3)	245.2 ± 71.9	139.3 ± 82.0	< 0.0001
Total bilirubin (mg/dL)	0.5 ± 0.3	14.2 ± 12.5	< 0.0001
INR	1.0 ± 0.1	1.8 ± 0.5	< 0.0001
AST (U/L)	27.0 ± 8.4	134.2 ± 78.2	< 0.0001
ALT (U/L)	26.0 ± 9.9	64.4 ± 78.2	< 0.0001
Albumin (g/dL)	3.9 ± 0.6	2.9 ± 0.7	< 0.0001
Creatinine (mg/dL)	0.8 ± 0.2	1.1 ± 1.0	0.0143
MELD score	6.8 ± 1.5	22.4 ± 7.6	< 0.0001
Total drinks in the past 30 days (TLFB)	383.3 ± 294.2	232.3 ± 263.0	0.0004
AUDIT score	29.5 ± 6.7	22.8 ± 9.2	0.0001

BMI, body mass index; WBC, white blood cell; INR, international normalized ratio; AST, aspartate transaminase; ALT, alanine transaminase; MELD, Model For End-Stage Liver Disease; TLFB, Time Line Follow-Back; AUDIT, Alcohol Use Disorders Identification Test.

Table 2

a. Enrichment of gene sets based on trait-association p-values of functional single nucleotide polymorphisms.

Gene set name	Gene set p-value	Gene set FDR	Significant gene number	Selected gene number	All gene number
NUCLEAR_REPLICATION_FORK	<0.001	0.032	3	5	10
NUCLEAR_CHROMOSOME_PART	<0.001	0.049	7	15	34

b. Enrichment of gene sets based on trait-association p-values of single nucleotide polymorphisms within a range of 500kb upstream and downstream of genes.

Gene set name	Gene set p-value	Gene set FDR	Significant gene number	Selected gene number	All gene number
NUCLEAR_CHROMOSOME_PART	<0.001	0	19	29	34
POSITIVE_REGULATION_OF_T_CELL_ACTIVATION	<0.001	0.006	15	20	21
HSA04140_REGULATION_OF_AUTOPHAGY	<0.001	0.007667	14	26	30
CXCR4PATHWAY	<0.001	0.0086	17	23	24
REGULATION_OF_T_CELL_ACTIVATION	<0.001	0.01025	18	27	28
REGULATION_OF_LYMPHOCYTE_ACTIVATION	<0.001	0.010286	20	33	35
POSITIVE_REGULATION_OF_LYMPHOCYTE_ACTIVATION	<0.001	0.011667	16	23	24
NUCLEAR_CHROMOSOME	<0.001	0.01175	26	44	54
CCR3PATHWAY	<0.001	0.012	16	20	23
POSITIVE_REGULATION_OF_T_CELL_PROLIFERATION	<0.001	0.0143	9	12	13
T_CELL_PROLIFERATION	<0.001	0.019455	12	17	19
ANATOMICAL_STRUCTURE_MORPHOGENESIS	<0.001	0.02525	198	347	379
SPRYPATHWAY	<0.001	0.028077	12	17	18
ACETYLCHOLINE_SYNTHESIS	<0.001	0.0365	6	6	8
IGF1RPATHWAY	0.001	0.036533	9	15	15
CELLULAR_MONOVALENT_INORGANIC_CATION_HOMEOSTASIS	0.001	0.039625	7	11	11
REGULATION_OF_T_CELL_PROLIFERATION	<0.001	0.040706	10	15	16
SODDPATHWAY	0.003	0.041056	6	8	10
MAINTENANCE_OF_LOCALIZATION	0.002	0.046579	11	19	22
LATE_ENDOSOME	0.002	0.047591	6	8	12
RIBONUCLEOPROTEIN_COMPLEX	0.002	0.04815	42	118	143

b, Enrichment of gene sets based on trait-association p-values of single nucleotide polymorphisms within a range of 500kb upstream and downstream of genes.

Gene set name	Gene set p-value	Gene set FDR	Significant gene number	Selected gene number	All gene number
CORTICAL_ACTIN_CYTOSKELETON	< 0.001	0.048333	9	12	13

Only those gene sets with an FDR < 0.05 are shown here. See Table a, Supplemental Digital Content 2, for detailed results.
Only those gene sets with an FDR < 0.05 are shown here. See Table b, Supplemental Digital Content 2, for detailed results.

Table 3

Number of alcoholic hepatitis cases (n = 90) and controls (n = 93) for various rs738409 genotypes.

	C/C	G/C	G/G	MAF
Controls	57	32	4	0.215
Cases	39	41	10	0.339

NB: The SNP is oriented on the minus strand; i.e., it can be argued that rs738409(G) is actually the common allele. MAF, minor allele frequency.

Table 4

Mortality during the first 12 months after diagnosis in alcoholic hepatitis cases (n = 90) for various rs738409 genotypes.

	C/C	G/C	G/G	MAF
Survived	31	32	8	0.338
Died	8	9	2	0.342

MAF, minor allele frequency.