

1 **Human genomics of acute liver failure due to hepatitis B virus**
2 **infection: an exome sequencing study in liver transplant recipients**

3

4

5 Samira Asgari^{1,2,3}, Nimisha Chaturvedi^{1,2}, Petar Scepanovic^{1,2}, Christian Hammer^{1,2},
6 Nasser Semmo⁴, Emiliano Giostra⁵, Beat Müllhaupt⁶, Peter Angus⁷, Alexander J
7 Thompson⁸, Darius Moradpour⁹, Jacques Fellay^{1,2,10}

8

9

10 ¹ School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne, Lausanne,
11 Switzerland

12 ² Swiss Institute of Bioinformatics, Lausanne, Switzerland

13 ³ Brigham and Women's Hospital, Harvard Medical School, Boston, USA

14 ⁴ Department for BioMedical Research, Hepatology, University of Bern, Switzerland

15 ⁵ Department of Gastroenterology and Hepatology, Geneva University Hospital,
16 Geneva, Switzerland

17 ⁶ Department of Gastroenterology and Hepatology, University Hospital Zürich, Zürich,
18 Switzerland

19 ⁷ Gastroenterology and Hepatology Department, Austin Health and the University of
20 Melbourne, Melbourne, Australia

21 ⁸ Department of Gastroenterology, St Vincent's Hospital and the University of
22 Melbourne, Melbourne, Australia

23 ⁹ Service of Gastroenterology and Hepatology, Lausanne University Hospital,
24 Lausanne, Switzerland

25 ¹⁰ Precision Medicine Unit, Lausanne University Hospital, Lausanne, Switzerland

26 **Abstract**

27 Acute liver failure (ALF) or fulminant hepatitis is a rare, yet severe outcome of
28 infection with hepatitis B virus (HBV) that carries a high mortality rate. The occurrence
29 of a life-threatening condition upon infection with a prevalent virus in individuals
30 without known risk factors is suggestive of pathogen-specific immune dysregulation.
31 In the absence of established differences in HBV virulence, we hypothesized that ALF
32 upon primary infection with HBV could be due to rare deleterious variants in the human
33 genome. To search for such variants, we performed exome sequencing in 21 previously
34 healthy adults who required liver transplantation upon fulminant HBV infection and
35 172 controls that were positive for anti-HBc and anti-HBs antibodies but had no clinical
36 history of jaundice or liver disease. After a series of hypothesis-driven filtering steps,
37 we searched for putatively pathogenic variants that were significantly associated with
38 case-control status. We did not find any causal variant or gene, a result that does not
39 support the hypothesis of a shared monogenic basis for human susceptibility to HBV-
40 related ALF in adults. This study represents a first attempt at deciphering the human
41 genetic contribution to the most severe clinical presentation of acute HBV infection in
42 previously healthy individuals.

43 **Author Summary**

44 Infection with hepatitis B virus (HBV) is very common and causes a variety of liver
45 diseases including acute and chronic hepatitis, cirrhosis and liver carcinoma. Acute
46 HBV infection is often asymptomatic, still about 1% of newly infected people develop
47 a rapid and severe disease known as acute liver failure or fulminant hepatitis. Acute
48 liver failure has a high mortality rate and is an indication for urgent liver
49 transplantation. It is not clear why some people, who are otherwise healthy, develop
50 such severe symptoms upon infection with a common pathogen. Here, we hypothesized
51 that rare DNA variants in the human genome could contribute to this unusual
52 susceptibility. We sequenced the exome (i.e. the regions of the genome that encode the
53 proteins) of 21 previously healthy adults who required liver transplantation upon
54 fulminant HBV infection and searched for rare genetic variants that could explain the
55 clinical presentation. We did not identify any variant that could be convincingly linked
56 to the extreme susceptibility to HBV observed in the study participants. This suggests
57 that HBV-induced acute liver failure is more likely to result from the combined
58 influence of multiple genetic and environmental factors.

59 **Introduction**

60 Hepatitis B virus (HBV) is a common human pathogen that attacks the liver and can
61 cause both acute and chronic disease. There is high inter-individual variability in the
62 clinical presentation of HBV infection, which ranges from self-limited to fulminant
63 acute disease, and from mild chronic hepatitis to liver cirrhosis and hepatocellular
64 carcinoma [1]. Differences in viral or environmental factors only explain a fraction of
65 this variability [2–5]. Previous studies have identified some human genetic factors that
66 play a modulating role in the clinical course of HBV infection [6,7]. However, our
67 understanding of host genetic influences on the disease is still very limited.

68

69 Fulminant hepatitis or acute liver failure (ALF) is defined as the rapid development of
70 liver injury leading to severe impairment of the synthetic capacity and to hepatic
71 encephalopathy in patients without previous liver disease [8,9]. ALF due to HBV
72 infection, or fulminant hepatitis B, is observed in less than 0.1% of infected individuals
73 but carries a high mortality and is an indication for urgent liver transplantation [10–15].

74

75 Such an unusual clinical presentation fits the definition of an extreme phenotype.
76 Electing patients with extreme phenotype increases the power to detect causal gene as
77 variants as these patients are more likely to carry alleles with profound functional
78 consequences that are otherwise very rare in the population, due to purifying selection
79 [16–20]. In this study, we used exome sequencing and statistical analysis in a cohort of
80 21 cases and 172 controls to search for human genetic variants conferring extreme
81 susceptibility to HBV. Cases were previously healthy adults who required liver
82 transplantation for fulminant hepatitis B and controls were HBV-infected adults who
83 did not develop fulminant hepatitis (Figure 1).

84

85

86 **Results**

87 **Study participants**

88 Of the 21 cases, 13 (62%) were female. The median age at transplantation was 36.5
89 years (range 22-58). Of 21 cases, 16 (76%) were European, four (19%) were Asian and
90 one (5%) was African (Supplementary Figure 1).

91

92 **Exome sequencing, variant calling and variant filtering**

93 Exome sequencing data were generated from DNA extracted from whole blood for all
94 study participants. On average per sample, 96% of reads passing filtering criteria were
95 unique (not marked as duplicate). Ninety-seven percent of unique reads could be
96 aligned to the human reference genome GRCh37. The mean on-bait coverage was 73x,
97 with 99% of target bases reaching at least 2x coverage, 97% of target bases achieving
98 at least 10x coverage and 84% achieving at least 30x coverage. 205,642 variants were
99 detected after GATK quality control filtering including 520 novel variants. The average
100 transition to transversion ratio (Ti/Tv) was 2.66, and the average heterozygous to
101 homozygous ratio was 1.5. A total of 38,062 low-frequency variants ($MAF \leq 0.05$)
102 passed filtering criteria including 31,620 rare variants ($MAF \leq 0.01$, Table 1).

103

104 **Table 1:** Total number of rare ($MAF \leq 0.01$) and rare plus low-frequency ($MAF \leq 0.05$)
105 variants that passes quality control and filtering criteria.

Effect	$MAF \leq 0.01$	$MAF \leq 0.05$
Inframe indel	321	368
Frameshift indel	880	954

Missense SNV	29,549	35,756
Splice site acceptor SNV	152	177
Splice site donor SNV	177	200
Nonsense	541	607
Total variants	31,620	38,062
Total genes	11,595	12,295

106

107 **Single variant associate analysis**

108 All putatively pathogenic variants were tested for association with case-control status
109 using Fisher's exact test. We first restricted the analysis to rare variants ($MAF \leq 0.01$)
110 and European cases only. One variant passed the Bonferroni correction threshold (p-
111 value $< 1.6e-6$). The variant was a missense SNV in *IGSF3* (rs78806598, p-value=1.8e-
112 18). Visualizing the aligned reads for this variant convinced us that this variant is called
113 due to misalignment. This gene was excluded from our further analyses. Expanding the
114 analysis to include the five none-European cases and the low-frequency variants did
115 not lead to discovery of any significant associations (Supplementary Tables 1-3).

116

117 **Gene-based association analysis**

118 11,595 genes were included in the gene burden analysis. Two different burden scores
119 were calculated for each gene using the approached described in the methods section.
120 Using the weighted sum method, *SLC29A1* had the lowest p-value (p-value=1.7e-5).
121 CTSW had the lowest p-value in binary collapsing method (p-value=1.8e-5). However,
122 none of these genes passed the Bonferroni correction threshold (p-value $< 2.5e-6$ for
123 20,000 protein coding genes, Supplementary Tables 4-5). Including the low-frequency
124 variants (12,295 genes in total) did not change these results. The top associations

125 including low-frequency variants were ADAM32 (p-value=1.7e-5) and PREX2 (p-
126 value=8.7e-6) in for weighted sum method binary collapsing method respectively
127 (Supplementary Tables 6-7). Overall, the results from the two collapsing methods and
128 the results between rare variants ($MAF \leq 0.01$) and low-frequency variants ($MAF \leq$
129 0.05) analyses were highly concordant (Figure 2, Supplementary Figure 2). The highest
130 correlation ($r^2=0.927$, CI:0.924-0.929) was observed between the results of weighted
131 sum and binary collapsing methods for rare variants. The lowest correlation ($r^2=0.739$,
132 CI:0.731-0.747) was observed between the results of binary collapsing method for rare
133 and low-frequency variants (Figure 2).

134

135

136 **Discussion**

137 The role of human genetic factors in susceptibility to fulminant hepatitis B is poorly
138 understood. Monogenic defects in key immune genes and pathways have been shown
139 to cause extreme susceptibility to other common pathogens in apparently healthy
140 individuals [7,21,22]. A prime example is herpes simplex encephalitis (HSE), the most
141 common form of sporadic viral encephalitis in the western world, which is only
142 observed in an extremely low fraction of people infected with type 1 herpes simplex
143 virus (HSV-1). Children who develop HSE upon primary HSV-1 infection are not
144 particularly susceptible to other infections, and children with other primary
145 immunodeficiencies are not more susceptible to HSE [22]. Since 2006, multiple genetic
146 variants have been causally linked with HSE [23–29]. Similarly, ALF only occurs in <
147 1/1000 of individuals after primary infection with HBV. Because this is an extremely
148 rare clinical event, we hypothesized that it could be the first manifestation of a rare
149 monogenic defect, resulting in pathogen-specific immune dysregulation.

150

151 We used exome sequencing to systematically search for rare, putatively pathogenic
152 variants that could explain extreme susceptibility to HBV infection. We analyzed the
153 genetic variants present in the exomes of 21 liver transplant recipients and compared
154 them to 172 controls who were exposed to HBV but did not develop fulminant hepatitis
155 B. First, we performed a single variant association analysis using Fisher's exact test.
156 Fisher's exact test is a conservative test of association but guarantees type I error control
157 for small sample sizes [30]. We found one significant association (p-value < 1.4e-6) in
158 one gene: *IGSF3*. However, the manual inspection of the mapped reads in the region
159 demonstrated that this variant was wrongly called, due to a mapping error. False-
160 positive incidental findings are a major problem in small-scale exome sequencing
161 studies [31]. Previous studies have proposed guidelines to avoid misinterpretations and
162 erroneous reports of potential causality due to false-positive findings [32,33]. Our
163 results show that even after applying these guidelines, it is important to ensure the
164 quality of final findings by visualizing the mapped regions and manually verifying the
165 quality of each variant call.

166

167 Rare variant association studies are usually underpowered. To enrich association
168 signals and reduce the penalty of multiple testing correction, it is common to aggregate
169 information across multiple rare variants within a region (gene, exon, sliding window,
170 etc.) and test for the association of all variants in the region with the phenotype of
171 interest [34]. We performed gene-based association analysis using two different
172 aggregation methods: weighted sum collapsing and binary collapsing. Both methods
173 assume that all the variants included in the test have the same direction of effect
174 (increasing disease risk in our scenario) and thus are underpowered to detect disease-

175 gene associations if variants exert their effects in opposite directions. Binary collapsing
176 assumes that all putatively pathogenic variants have the same effect size. Weighted sum
177 collapsing assumes that rarer variants have larger effect sizes and that the risk of disease
178 is a function of the sum of the variant effect sizes. We did not find any genes to be
179 significantly associated with case-control status. The p-values and the top ranked genes
180 in both analyses were highly concordant (Figure 2, Supplementary Figure 2). The high
181 correlation between the p-values of weighted sum and binary collapsing methods
182 suggests that most individuals carry only one putatively pathogenic variant per gene.
183 This implies that larger sample sizes or linkage studies in families with multiple
184 affected individuals will be needed to increase statistical power for detecting potential
185 associations between rare variants and HBV-related ALF.

186

187 We did not identify any genetic variant conferring monogenic susceptibility to
188 fulminant hepatitis B in adults. Our results suggest that ALF upon primary infection
189 with HBV is likely to be multifactorial. This conclusion is in line with a previous exome
190 sequencing study of fulminant hepatitis A, which also failed to find any convincing
191 casual gene or genetic variant [35]. Our failure to detect a Mendelian cause for
192 fulminant hepatitis B, despite previous success for comparable phenotypes, could be
193 due to a number of factors and limitations of our study: 1- The severe liver injury
194 observed in patients with fulminant hepatitis B can be due to opposite pathogenic
195 mechanisms: an inefficient innate immune response, which is unable to prevent viral
196 replication, activate the adaptive immune system and clear the virus; and an over-
197 activation of innate immune signaling pathways leading to cytokine storm and
198 uncontrolled inflammation [36–38]. This implies that genetic variants with opposite
199 effects (e.g. gain-of-function and loss-of-function variants in the same gene or pathway)

200 could contribute synergistically to the disease. Such a genetic architecture would be
201 extremely difficult to identify. 2- Our study was performed in adults, while most
202 previous examples come from pediatric studies. A previous twin study has shown that
203 the estimated heritability of many immune parameters decreases with age, suggesting
204 that the cumulative influence of environmental exposures alters the role of human
205 genetics in susceptibility to infectious diseases in older patients [39]. 3- Due to our
206 recruitment criteria, we did not have access to information about the viral genome.
207 HBV genetic variation has previously been shown to be associated with disease severity
208 and infection outcome. For example, mutations in the pre-core and core protein genes
209 or X region of the HBV genome, which encode the HBx protein, were associated with
210 severity of chronic hepatitis [40,41], and mutations in HBV genotypes C and D
211 associated with liver cirrhosis and progression to hepatocellular carcinoma [42]. The
212 inclusion of viral genome information would allow for the stratification of patients
213 based on known HBV mutations, thus increasing the signal-to-noise ratio in the human
214 genetic analyses.

215

216 This study represents the first attempt at identifying human genetic variants involved
217 in the pathogenesis of fulminant hepatitis B in previously healthy individuals. The
218 absence of any conclusive finding indicates that ALF due to primary HBV infection is
219 unlikely to be the result of a single monogenic disorder, and that a more complex
220 genetic architecture is probably involved, intermixed with viral and environmental
221 factors. Going forward, studies that aim at identifying the genetic causes of fulminant
222 hepatitis B will need to include more patients and to better characterize them at the
223 molecular level (e.g. to stratify them based on specific immune activation markers
224 measured during acute disease). To obtain a more complete description of human

225 genetic variation, full genome sequencing would be preferable, which will allow the
226 exploration of non-coding variants, large structural variants and exonic variants that are
227 not well-covered by current exome capture methods. Finally, a parallel evaluation of
228 the viral genome and of any potentially interfering factor will be necessary, as
229 individual susceptibility to HBV is the result of a complex interplay between host,
230 pathogen and environment.

231

232

233 **Methods**

234 **Ethics statement**

235 The study was approved by the responsible institutional Human Research Ethics
236 Committees in Switzerland and Australia. Each study participant provided written
237 informed consent for genetic testing.

238

239 **Study participants**

240 Twenty-one liver transplant recipients who developed ALF due to fulminant HBV
241 infection were recruited in the transplantation units of the University Hospitals of
242 Lausanne, Zurich, Bern, Geneva, and Melbourne. Patients with fulminant hepatitis B
243 due to reactivation after withdrawal of anti-HBV drugs and patients with pre-existing
244 liver diseases, known immune deficiency or other chronic conditions were excluded.
245 The following demographic and clinical information were collected: age at
246 transplantation date, gender, and ethnicity. For each study participant, we obtained 3ml
247 of blood in EDTA vacutainer tubes and 2,5ml blood in PAXgene blood RNA tubes.
248 Samples were immediately frozen at -70°C, and then shipped and analyzed in batch.

249

250 **Control population**

251 One hundred seventy-two controls were selected from our in-house database of exome-
252 sequenced individuals. They were adults of European ancestry, who were positive for
253 anti-HBc and anti-HBs antibodies, but had no clinical history of jaundice or liver
254 disease. The controls were HBV-eliminated at the time of blood collection for exome-
255 sequencing.

256

257 **DNA sequencing and alignment**

258 Genomic DNA was extracted from whole blood using QIAgen DNeasy Blood and
259 Tissue kit. Cluster generation was performed using Illumina TruSeq PE Cluster Kit v5
260 reagents. Libraries were sequenced as 100-basepair long, paired-end reads on Illumina
261 HiSeq 2500 using TruSeq SBS Kit v5 reagents. Sequencing reads were processed using
262 CASAVA v1.82, and aligned to the human reference genome hg19 using BWA [43,44]
263 version 0.6.2. PCR duplicates were removed using Picard 1.27-1
264 (<http://picard.sourceforge.net/>). We used Samtools [45] Visualization of aligned reads.

265

266 **Variant calling**

267 We used Genome Analysis Toolkit (GATK) [46,47] version 3.1-1 to call single
268 nucleotide variants (SNVs) and small insertion and deletions (indels) from duplicate-
269 marked bam files. We used HaplotypeCaller for multi-sample variant calling on all
270 samples following GATK best practice.

271

272 **Variant effect prediction, frequency estimation and filtering**

273 We used SnpEff [48] version 4.3T to predict the functional impact of variants. As a
274 single variant can have several predicted effects, we only considered the most severe

275 effect for each variant according to SnpEff order of impact severity. We used genome
276 aggregation database (gnomAD) to assign minor allele frequency (MAF) to variants
277 (gnomAD, includes 123,136 exome sequences and 15,496 whole-genome sequences)
278 [49]. For variants that were not present in gnomAD were assigned $MAF=1-e8$ to avoid
279 having $-\log(0)$ in the following burden analysis. Only biallelic variants that were
280 flagged as PASS by GATK, and were called in all cases and all controls were included
281 in the analysis. Known polymorphic genes and genes in noisy alignment regions were
282 excluded from the analysis [31–33]. We restricted all the downstream analyses to
283 protein modifying variants (missense, inframe indels, frame-shift indels, splice-site
284 disrupting, nonsense). All analyses were done on both rare ($MAF \leq 0.01$) and low-
285 frequency ($MAF \leq 0.05$) variants. We refer to variants that passed above filtering
286 criteria as putatively pathogenic variants.

287

288 **Single variant association tests**

289 We used Fisher's exact test to look for association of single variants with case-control
290 status. Each variant was given an allele count based on the number of alternate alleles
291 $G_{ij} \in \{0,1,2\}$, where G_{ij} is the genotype of variant j in individual i . We summarized the
292 reference and alternate allele counts for cases and controls, into 2x2 contingency tables.
293 These tables were analyzed using one-tailed Fisher's exact test. We used Bonferroni
294 correction to correct for multiple testing.

295

296 **Gene burden association tests**

297 Gene burden test was performed using GMMAT [50] version 0.7-1, a generalized linear
298 mixed model framework as follows:

299

300
$$H = W\alpha + C\beta + M; \quad M \sim N\left(0, \sum_{k=1}^K \tau_k V_k\right)$$

301

302 Where H is an n -vector of case-control status for n individuals, W is an n -vector of
303 gender covariate, and C is an n -vector of the gene burden scores. M is an n -vector of
304 random effects. τ_k is the variance component parameter and V_k are known $n \times n$
305 matrices. We ran this model using an $n \times n$ kinship coefficients matrix calculated using
306 PC-Relate [51]. We used Bonferroni correction to correct for multiple testing. To
307 calculate the gene burden scores, we used two different methods:

308

309 **i) Binary collapsing method**

310 Each gene was given a burden score of zero if no putatively pathogenic variant was
311 present in the gene and a gene burden score of one otherwise:

312

313
$$C_i = \begin{cases} 1 & \text{if } \sum_{j=1}^m G_{ij} > 0 \\ 0 & \text{if } \sum_{j=1}^m G_{ij} = 0 \end{cases}$$

314

315 Where $G_{ij} \in \{0,1,2\}$ is the genotype of variant j in individual i , and C_i is the gene burden
316 score for individual i . This approach is based on the Cohort Allelic Sum Test (CAST)
317 method [52].

318

319 **ii) Weighted sum collapsing method**

320 First, each gene was given a burden score as follows:

321

322
$$C_i = \sum_{j=1}^m (G_{ij} * -\log_{10}(AF_i\text{-gnomAD}))$$

323

324 Where $j \in \{0,1,2,\dots,m\}$ is the m th variant per gene, and $G_{ij} \in \{0,1,2\}$ is the genotype
325 of variant j in individual i , $AF_{i-gnomAD}$ is the minor allele frequency of j in gnomAD,
326 and C_i is the gene burden score for individual i . This approach is based on the Madsen
327 and Browning weighted sum method [53].

328

329

330

331 **Acknowledgements**

332 Sequencing was performed at the Lausanne Genomic Technologies Facility of the
333 University of Lausanne. The computations were performed at the Vital-IT
334 (<http://www.vital-it.ch>) Center for high-performance computing of the SIB Swiss
335 Institute of Bioinformatics. We would like to thank the study participants, as well as
336 the study nurses, physicians and laboratories who participated in the recruitment.

337 **References**

- 338 1. MacLachlan JH, Locarnini S, Cowie BC (2015) Estimating the global
339 prevalence of hepatitis B. *Lancet* 386: 1515–7. doi:10.1016/S0140-6736(15)61116-3.
- 340 2. Karvellas C, Cardoso F, Gottfried M, Reddy, Hanje, et al. (2017) HBV-
341 Associated Acute Liver Failure After Immunosuppression and Risk of Death. *Clinical*
342 *gastroenterology and hepatology : the official clinical practice journal of the*
343 *American Gastroenterological Association* 15: 113–122.
344 doi:10.1016/j.cgh.2016.06.008.
- 345 3. Aragri M, Alteri C, Battisti A, Carlo D, Minichini C, et al. (2016) Multiple
346 Hepatitis B Virus (HBV) Quasispecies and Immune-Escape Mutations Are Present in
347 HBV Surface Antigen and Reverse Transcriptase of Patients With Acute Hepatitis B.
348 *The Journal of infectious diseases* 213: 1897–905. doi:10.1093/infdis/jiw049.
- 349 4. Zhang A, Wan Z, You S, Liu H, Zhu B, et al. (2013) Association of Hepatitis
350 B Virus Mutations of A1846T and C1913A/G With Acute-on-Chronic Liver Failure
351 Development From Different Underlying Chronic Liver Diseases. *Hepatitis monthly*
352 13: e12445. doi:10.5812/hepatmon.12445.
- 353 5. Goldstein S, Alter M, Williams I, Moyer L, Judson F, et al. (2002) Incidence
354 and risk factors for acute hepatitis B in the United States, 1982-1998: implications for
355 vaccination programs. *The Journal of infectious diseases* 185: 713–9.
356 doi:10.1086/339192.
- 357 6. Matsuura K, Isogawa M, Tanaka Y (2016) Host genetic variants influencing
358 the clinical course of hepatitis B virus infection. *J Med Virol* 88: 371–9.
359 doi:10.1002/jmv.24350.
- 360 7. Kenney A, Dowdle J, Bozzacco L, McMichael T, Gelais C, et al. (2017)
361 Human Genetic Determinants of Viral Diseases. *Annual review of genetics* 51: 241–
362 263. doi:10.1146/annurev-genet-120116-023425.
- 363 8. Trey C, Davidson CS (1970) The management of fulminant hepatic failure.
364 *Prog Liver Dis* 3: 282–98.
- 365 9. Ritt DJ, Whelan G, Werner DJ, Eigenbrodt EH, Schenker S, et al. (1969)
366 Acute hepatic necrosis with stupor or coma. An analysis of thirty-one patients.
367 *Medicine (Baltimore)* 48: 151–72.
- 368 10. Papaevangelou G, Tassopoulos N, Roumeliotou-Karayannis A, Richardson C
369 (1984) Etiology of fulminant viral hepatitis in Greece. *Hepatology* 4: 369–72.
- 370 11. Schiødt F, Davern T, Shakil O, McGuire B, Samuel G, et al. (2003) Viral
371 Hepatitis-Related Acute Liver Failure. *The American Journal of Gastroenterology* 98:
372 448–453. doi:10.1111/j.1572-0241.2003.t01-1-07223.x.
- 373 12. Oketani M, Uto H, Ido A, Tsubouchi H (2014) Management of hepatitis B
374 virus-related acute liver failure. *Clin J Gastroenterol* 7: 19–26. doi:10.1007/s12328-
375 013-0447-1.
- 376 13. Fujiwara K, Mochida S, Matsui A, Nakayama N, Nagoshi S, et al. (2008)
377 Fulminant hepatitis and late onset hepatic failure in Japan. *Hepatology* 47: 646–57.
378 doi:10.1111/j.1872-034X.2008.00322.x.
- 379 14. Lee W, Squires R, Nyberg S, Doo E, Hoofnagle J (2008) Acute liver failure:
380 Summary of a workshop. *Hepatology* 47: 1401–1415. doi:10.1002/hep.22177.
- 381 15. Ostapowicz G, Fontana RJ, Schiødt FV, Larson A, Davern TJ, et al. (2002)
382 Results of a prospective study of acute liver failure at 17 tertiary care centers in the
383 United States. *Ann Intern Med* 137: 947–54. doi:10.7326/0003-4819-137-12-

- 384 200212170-00007.
- 385 16. Guey L, Kravic J, Melander O, Burt N, Laramie J, et al. (2011) Power in the
386 phenotypic extremes: a simulation study of power in discovery and replication of rare
387 variants. *Genetic epidemiology* 35: 236–46. doi:10.1002/gepi.20572.
- 388 17. Lander, Botstein (1989) Mapping mendelian factors underlying quantitative
389 traits using RFLP linkage maps. *Genetics* 121: 185–99.
- 390 18. Barnett I, Lee S, Lin X (2013) Detecting rare variant effects using extreme
391 phenotype sampling in sequencing association studies. *Genetic epidemiology* 37:
392 142–51. doi:10.1002/gepi.21699.
- 393 19. Li D, Lewinger J, Gauderman W, Murcray C, Conti D (2011) Using extreme
394 phenotype sampling to identify the rare causal variants of quantitative traits in
395 association studies. *Genetic epidemiology* 35: 790–9. doi:10.1002/gepi.20628.
- 396 20. Gibson G (2012) Rare and common variants: twenty arguments. *Nature*
397 reviews Genetics 13: 135–45. doi:10.1038/nrg3118.
- 398 21. Casanova J-LL (2015) Human genetic basis of interindividual variability in
399 the course of infection. *Proc Natl Acad Sci USA* 112: E7118–27.
400 doi:10.1073/pnas.1521644112.
- 401 22. Alcaïs A, Abel L, Casanova J-LL (2009) Human genetics of infectious
402 diseases: between proof of principle and paradigm. *J Clin Invest* 119: 2506–14.
403 doi:10.1172/JCI38111.
- 404 23. Zhang S-YY, Jouanguy E, Ugolini S, Smahi A, Elain G, et al. (2007) TLR3
405 deficiency in patients with herpes simplex encephalitis. *Science* 317: 1522–7.
406 doi:10.1126/science.1139522.
- 407 24. Andersen LL, Mørk N, Reinert LS, Kofod-Olsen E, Narita R, et al. (2015)
408 Functional IRF3 deficiency in a patient with herpes simplex encephalitis. *J Exp Med*
409 212: 1371–9. doi:10.1084/jem.20142274.
- 410 25. Herman M, Ciancanelli M, Ou Y-HH, Lorenzo L, Klaudel-Dreszler M, et al.
411 (2012) Heterozygous TBK1 mutations impair TLR3 immunity and underlie herpes
412 simplex encephalitis of childhood. *J Exp Med* 209: 1567–82.
413 doi:10.1084/jem.20111316.
- 414 26. Guo Y, Audry M, Ciancanelli M, Alsina L, Azevedo J, et al. (2011) Herpes
415 simplex virus encephalitis in a patient with complete TLR3 deficiency: TLR3 is
416 otherwise redundant in protective immunity. *J Exp Med* 208: 2083–98.
417 doi:10.1084/jem.20101568.
- 418 27. Pérez de Diego R, Sancho-Shimizu V, Lorenzo L, Puel A, Plancoulaine S, et
419 al. (2010) Human TRAF3 adaptor molecule deficiency leads to impaired Toll-like
420 receptor 3 response and susceptibility to herpes simplex encephalitis. *Immunity* 33:
421 400–11. doi:10.1016/j.immuni.2010.08.014.
- 422 28. Sancho-Shimizu V, Pérez de Diego R, Lorenzo L, Halwani R, Alangari A, et
423 al. (2011) Herpes simplex encephalitis in children with autosomal recessive and
424 dominant TRIF deficiency. *J Clin Invest* 121: 4889–902. doi:10.1172/JCI59259.
- 425 29. Casrouge A, Zhang S-YY, Eidenschenk C, Jouanguy E, Puel A, et al. (2006)
426 Herpes simplex virus encephalitis in human UNC-93B deficiency. *Science* 314: 308–
427 12. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16973841>.
- 428 30. Li B, Leal S (2008) Methods for detecting associations with rare variants for
429 common diseases: application to analysis of sequence data. *American journal of*
430 *human genetics* 83: 311–21. doi:10.1016/j.ajhg.2008.06.024.

- 431 31. Kohane I, Hsing M, Kong S (2012) Taxonomizing, sizing, and overcoming the
432 incidentalome. *Genetics in medicine : official journal of the American College of*
433 *Medical Genetics* 14: 399–404. doi:10.1038/gim.2011.68.
- 434 32. MacArthur, Manolio, Dimmock, Rehm, Shendure, et al. (2014) Guidelines for
435 investigating causality of sequence variants in human disease. *Nature* 508: 469–76.
436 doi:10.1038/nature13127.
- 437 33. Fajardo K, Adams D, Program N, Mason C, Sincan M, et al. (2012) Detecting
438 false-positive signals in exome sequencing. *Human mutation* 33: 609–13.
439 doi:10.1002/humu.22033.
- 440 34. Lee S, Abecasis G, Boehnke M, Lin X (2014) Rare-variant association
441 analysis: study designs and statistical tests. *American journal of human genetics* 95:
442 5–23. doi:10.1016/j.ajhg.2014.06.009.
- 443 35. Long D, Fix OK, Deng X, Seielstad M, Luring AS (2014) Whole genome
444 sequencing to identify host genetic risk factors for severe outcomes of hepatitis a
445 virus infection. *J Med Virol* 86: 1661–8. doi:10.1002/jmv.24007.
- 446 36. Han Q, Zhang C, Zhang J, Tian Z (2013) The role of innate immunity in HBV
447 infection. *Semin Immunopathol* 35: 23–38. doi:10.1007/s00281-012-0331-y.
- 448 37. Maini MK, Gehring AJ (2016) The role of innate immunity in the
449 immunopathology and treatment of HBV infection. *J Hepatol* 64: S60–70.
450 doi:10.1016/j.jhep.2016.01.028.
- 451 38. Jung M-CC, Pape GR (2002) Immunology of hepatitis B infection. *Lancet*
452 *Infect Dis* 2: 43–50.
- 453 39. Brodin P, Jojic V, Gao T, Bhattacharya S, Angel C (2015) Variation in the
454 human immune system is largely driven by non-heritable influences. *Cell*.
- 455 40. Kim H, Lee S-AA, Kim B-JJ (2016) X region mutations of hepatitis B virus
456 related to clinical severity. *World J Gastroenterol* 22: 5467–78.
457 doi:10.3748/wjg.v22.i24.5467.
- 458 41. Kim H, Lee S-AA, Do SY, Kim B-JJ (2016) Precore/core region mutations of
459 hepatitis B virus related to clinical severity. *World J Gastroenterol* 22: 4287–96.
460 doi:10.3748/wjg.v22.i17.4287.
- 461 42. Sunbul M (2014) Hepatitis B virus genotypes: global distribution and clinical
462 importance. *World J Gastroenterol* 20: 5427–34. doi:10.3748/wjg.v20.i18.5427.
- 463 43. Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-
464 Wheeler transform. *Bioinformatics* 25: 1754–60. doi:10.1093/bioinformatics/btp324.
- 465 44. Li H, Durbin R (2010) Fast and accurate long-read alignment with Burrows-
466 Wheeler transform. *Bioinformatics* 26: 589–95. doi:10.1093/bioinformatics/btp698.
- 467 45. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al. (2009) The Sequence
468 Alignment/Map format and SAMtools. *Bioinformatics* 25: 2078–9.
469 doi:10.1093/bioinformatics/btp352.
- 470 46. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, et al. (2010) The
471 Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation
472 DNA sequencing data. *Genome research* 20: 1297–303. doi:10.1101/gr.107524.110.
- 473 47. Auwera GA Van der, Carneiro MO, Hartl C, Poplin R, Angel G Del, et al.
474 (2013) From FastQ data to high confidence variant calls: the Genome Analysis
475 Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 43: 11.10.1–33.
476 doi:10.1002/0471250953.bi1110s43.

- 477 48. Cingolani P, Platts A, Wang L e, Coon M, Nguyen T, et al. (2012) A program
478 for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff:
479 SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6:
480 80–92. doi:10.4161/fly.19695.
- 481 49. Exome Aggregation Consortium (ExAC) (n.d.) Cambridge, MA (URL:
482 <http://exac.broadinstitute.org>) [August 2016].
- 483 50. Chen H, Wang C, Conomos M, Stilp A, Li Z, et al. (2016) Control for
484 Population Structure and Relatedness for Binary Traits in Genetic Association Studies
485 via Logistic Mixed Models. *American journal of human genetics* 98: 653–66.
486 doi:10.1016/j.ajhg.2016.02.012.
- 487 51. Conomos M, Reiner A, Weir B, Thornton T (2016) Model-free Estimation of
488 Recent Genetic Relatedness. *American journal of human genetics* 98: 127–48.
489 doi:10.1016/j.ajhg.2015.11.022.
- 490 52. Morgenthaler S, Thilly W (2007) A strategy to discover genes that carry
491 multi-allelic or mono-allelic risk for common diseases: a cohort allelic sums test
492 (CAST). *Mutation research* 615: 28–56. doi:10.1016/j.mrfmmm.2006.09.003.
- 493 53. Madsen B, Browning S (2009) A groupwise association test for rare mutations
494 using a weighted sum statistic. *PLoS genetics* 5: e1000384.
495 doi:10.1371/journal.pgen.1000384.
- 496
497

498 **Figure Legends**

499 **Figure 1: Overview of data production and data analysis pipeline.** MAF: minor
500 allele frequency, GATK: Genome Analysis Tool Kit

501

502 **Figure 2: Comparison between different gene burden analysis methods and**
503 **different MAF thresholds.** The circles below each plot show the top ten associated
504 genes in the two compared analyses and the number of shared genes between the two
505 sets: A) correlation between p-value for rare variants ($MAF \leq 0.01$) using weighted
506 sum method (light green circle) and binary collapsing method (light red circle) B)
507 correlation between p-value for binary collapsing method using $MAF \leq 0.01$ (light red
508 circle) and $MAF \leq 0.05$ (dark red circle).

509

510 **Supplementary Tables legends**

511 **Supplementary Tables 1-3: Fisher's exact test results for single variant association**
512 **analysis** for: 1- Rare variants ($MAF \leq 0.01$), 16 European cases and 172 controls, 2-
513 Rare variants ($MAF \leq 0.01$), all 21 cases and 172 controls, 3- Low-frequency variants
514 ($MAF \leq 0.05$), all 21 cases and 172 controls. Column names: chromosome, position,
515 reference allele, alternate allele, variant ID, gene, number of putatively pathogenic
516 alleles in cases, number of putatively pathogenic alleles in controls, number of non-
517 putatively pathogenic alleles in cases, number of non-putatively pathogenic alleles in
518 controls

519

520 **Supplementary Tables 4-7: Gene burden association results** for: 4- Rare variants
521 ($MAF \leq 0.01$) and weighted sum method, 5- Rare variants ($MAF \leq 0.01$) and binary
522 collapsing method, 6- Rare variants ($MAF \leq 0.05$) and weighted sum method, 7- Rare

523 variants ($MAF \leq 0.05$) and binary collapsing method. Column names: gene, score,
524 variance, p-value

525

526 **Supplementary Figures legends**

527 Supplementary Figure 1: **Principal component analysis (PCA)** A-B) PCA analysis
528 for 21 cases, 172 controls and continental populations from 1000 genomes project, C-
529 D) PCA analysis for 21 cases and 172 controls.

530

531 Supplementary Figure 2: **Comparison between different gene burden analysis**
532 **methods and different MAF thresholds**. The circles below each plot show the top ten
533 associated genes in the two compared analyses and the number of shared genes between
534 the two sets: A) correlation between p-value for low-frequency variants ($MAF \leq 0.05$)
535 using weighted sum method (dark green circle) and binary collapsing method (dark red
536 circle). B) correlation between p-value for weighted sum method using $MAF \leq 0.01$
537 (light green circle) and $MAF \leq 0.05$ (dark green circle).

bioRxiv preprint doi: <https://doi.org/10.1101/320374>; this version posted May 11, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY 4.0 International license.



