The rs738409 G Allele in PNPLA3 Is Associated With a Reduced Risk of COVID-19 Mortality and Hospitalization

Innes, Hamish; Buch, Stephan; Barnes, Eleanor; Hampe, Jochen; Marjot, Thomas; Stickel, Felix

Published in:
Gastroenterology

DOI:
10.1053/j.gastro.2021.02.059

Publication date:
2021

Document Version
Author accepted manuscript

The rs738409 G allele in *PNPLA3* is associated with a reduced risk of COVID-19 mortality and hospitalisation.

AIUTHORS: **Hamish Innes,** PhD[1,2,3*]**, Stephan Buch** PhD [4*]**,** Eleanor Barnes PhD [5] Jochen Hampe MD PhD [4], Thomas Marjot PhD[6], Felix Stickel MD PhD [7]

AFFILIATIONS:

[1] School of Health and Life Sciences; Glasgow Caledonian University. Glasgow UK.

[2] Division of Epidemiology and Public Health, School of Medicine, University of Nottingham, Nottingham, United Kingdom;

[3] Public Health Scotland, Glasgow, UK.

[4] Medical Department 1, University Hospital Dresden, TU Dresden, Germany;

[5] Peter Medawar Building for Pathogen Research, Nuffield Department of Medicine and the Oxford NIHR Biomedical Research Centre, Oxford University, UK.

[6] Oxford Liver Unit, Translational Gastroenterology Unit, Oxford University Hospitals NHS Foundation Trust, University of Oxford, Oxford, UK

[7] Department of Gastroenterology and Hepatology, University Hospital of Zurich, Switzerland

***Author names in bold designate shared co-first authorship**

CONFLICTS OF INTEREST: There are no conflicts of interest to disclose.

WORD COUNT: 986 (introduction, methods, results and discussion)

AUTHOR CONTRIBUTIONS: Specific author contributions are outlined in the table below.

Author	CONTRIBUTIONS								
	Conceptualisation	Data curation	Formal analysis	Funding acquisition	Study design & Methods	Resources	Supervision	Writing original draft	Writing, review and editing
HI	**X**	**X**	**X**	**X**	**X**	**X**		**X**	**X**
SB	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**	**X**
EB	**X**			**X**	**X**	**X**			**X**
TM	**X**			**X**	**X**	**X**			**X**
JH	**X**			**X**	**X**	**X**			**X**
FS	**X**	**X**		**X**	**X**	**X**	**X**	**X**	**X**

CORRESPONDENCE: Address correspondence to: Dr Hamish Innes; Glasgow Caledonian University; Cowcaddens Road; G40BA; Glasgow, UK. Tel: 0141-5332950.

Email: Hamish.innes@gcu.ac.uk

FUNDING: HI is a Medical Research Foundation Viral Hepatitis Fellow. TM has received a registry grant from European Association for Study of the Liver (EASL) (2020RG03) and is supported by the Wellcome Trust as a Clinical Research Fellow (MSDTC_868356). EB is an NIHR Senior Investigator. This work was also, in part, supported by the Swiss National Science Foundation (SNF no. 310030_169196).

INTRODUCTION

The natural history of SARS-CoV-2 infection is highly variable, ranging from asymptomatic infection on one hand, to pneumonia, septic shock, multiple organ failure and death on the other. [1] The physiological pathways that influence prognosis following infection are still incompletely understood; this represents a key barrier to optimal management.

The rs738409 G variant in *patatin-like phospholipase domain containing 3* (*PNPLA3)* is a prominent genetic risk factor for steatosis, cirrhosis and hepatocellular carcinoma. [2] For two reasons, we hypothesised that rs738409 G may also affect COVID-19 severity. Firstly, rs738409 G is associated with retinoid storage levels in liver mesenchymal cells [3], which may have bearing on the ability to mount an effective immune response following viral infection [4]. Secondly, previous data suggest the rs738409 G increases cellular abundance of omega-3 polyunsaturated fatty acids (PUFAs) such as alpha linolenic acid [5], which may modulate inflammatory processes levels during infections.

Our goal was to test a possible association of rs738409 G in *PNPLA3* with outcomes of COVID-19 using data from the United Kingdom Biobank (UKB) study.

METHODS

Our hypothesis was tested in UKB, a population-based cohort of ~500,000 middle-aged individuals in the UK. Latterly, data on participants testing positive for SARS-CoV-2 in England after 16 March 2020 have been released.

All UKB unrelated participants with: a) a positive PCR test result for SARs-CoV-2 between 16th March 2020 and 15th August 2020, and b) data available for the *PNPLA3*:rs738409 genotype were analyzed.

Primary endpoints in this study were: a) hospital admission for COVID-19; and b) death from COVID-19.

Logistic regression was used to assess the association between rs738409 G allele and each endpoint under additive and recessive genetic models. All models were adjusted for potential confounding factors.

We performed sensitivity analyses to check assumptions. First, we looked to see if the association between rs738409 G and each outcome changed when individuals with liver disease were excluded. We also assessed if the association varied according to ethnicity, coronary artery disease (CAD) and obesity, age and liver disease. Third, we determined the rs738409 G association with two specific types of COVID-19 hospitalisation; namely: i) hospital admission for COVID-19 with pneumonia recorded; and ii) hospital admission for COVID-19 entailing advanced respiratory support. Fourth, we calculated the association between rs738409 G and prevalent liver disease to ensure concordance with previous research [2].

Further information, including detailed covariate and outcome definitions, is provided in the supplementary methods.

VALIDATION ANALYSIS:

We retrieved and meta-analysed the association between rs738409 G and COVID-19 morbidity from three studies/sources comparing patients with COVID-19 morbidity to SARS-CoV-2 positive patients with no/minimal morbidity. These data sources were: a) The FinnGen biobank; b) Geisinger Health Study; 3) Data on COVID-19 patients hospitalised in Palermo, Italy.[6]

RESULTS

The primary analysis included 1585 UKB participants. Mean age was 67.9 years (sd:9.1), 52.7% were male, and three-quarter were White British (78.1%). The rs738409 G allele frequency was 20.9% (Table S1). The number of patients with a COVID-19 hospital admission, a COVID-19 admission with pneumonia, and a COVID-19 admission requiring advanced respiratory support was 759 (47.9%), 450(28.4%) and 76(4.8%), respectively. Almost one-fifth of the sample died from COVID-19 (16.9%; n=267).

The rs738409 G allele was independently associated with a reduced risk of COVID-19 hospitalisation and mortality (Figure 1). On average, each additional G allele carried was associated with a 21% lower odds of COVID-19 hospitalisation (aOR:0.79; 95%CI 0.64-0.97. *P*=0.027); and a 25% lower odds of COVID-19 death (aOR:0.75; 95%CI:0.57-0.98. *P*=0.037). The adjusted odds ratio for the recessive model was 0.47 (95%CI:0.25-0.88. *P*=0.018) for COVID-19 hospitalisation and 0.19 (95%CI: 0.06-0.64;*P*=0.007) for COVID-19 mortality.

In sensitivity analyses, the association between rs738409 G allele and COVID-19 mortality/hospitalisation did not attenuate after excluding individuals with liver disease (e.g. adjusted additive OR for hospitalisation: 0.78; 95%CI: 0.63-0.97. *P*=0.027). There was a suggestion that the rs738409 G effect size was greater for younger individuals. For example, the additive OR for COVID-19 hospitalisation was 0.59 (95%CI:0.41-0.84) for ages<65yrs versus 0.92 (95%CI:0.71-1.19) for individuals aged ≥65yrs (*Interaction P*=0.040). Finally, as expected, rs738409 G allele was associated with an increased risk of liver disease (aOR:1.44; 95%CI: 1.01-2.05; *P*=0.045); see Figure 1.

VALIDATION ANALYSIS:

In meta-analysis of three independent data sources, rs738409 G was associated with a reduced risk of COVID-19 hospitalisation/severe disease versus mild disease (additive OR: 0.83: 95%CI: 0.66-1.05; *P*=0.12; *P* heterogeneity=0.68); see Figure 1.

DISCUSSION:

These data show a lower risk of COVID-19 hospitalisation and death in carriers of the rs738409 G allele in *PNPLA3*, which is athwart to its risk-increasing effect on liver disease (Figure 1). However, although this association was robust in the UKB study against a broad set of COVID-19 related endpoints and with adjustment for a comprehensive set of covariates, it was only moderately supported by independent validation data. Further data from larger well-defined cohorts including data on mortality are therefore required to verify the association of rs738409 G with COVID-19 sequelae. Functionally, this association could reflect the influence of lipid metabolism on the immune

response to COVID-19. E.g., retinoids are stored as retinyl esters in hepatic mesenchymal cells and also in adipose tissue where *PNPLA3* is expressed. When required, retinoids are mobilised to extrahepatic tissues where they can stimulate interferon type 1 production as a potent antiviral cytokine response to viral infections.[4] PNPLA3 has retinyl-palmitate lipase activity which stimulates release of retinol into the systemic circulation.[7] Accordingly, individuals with the I148M loss-of-function variant (encoded by rs738409 G) exhibit lower circulating retinoid levels [7] and increased retinoid storage levels in the liver [3]. In contrast, some risk factors for severe COVID-19 (e.g. obesity and liver disease [8]) are associated with diminished retinoid levels and/or impaired retinoid signalling, which may limit retinoid availability during infection. This finding may also reflect a lower omega-6 to omega-3-PUFA ratio in individuals with the rs738409 G allele, [5] which may temper inflammation, and safeguard against cytokine-storm-syndrome.

In summary, we report a putative association between PNPLA3:rs738409 and COVID-19 severity. In doing so, we provide an example of how pleiotropic effects of certain genetic risk loci affect disease endpoints differently.

ACKNKOWLEDGEMENTS:

This research has been conducted using the United Kingdom (UK) Biobank Resource (application number: 8764). We also want to acknowledge the participants and investigators of the FinnGen study There are no conflicts of interest to declare.

REFERENCES:

[1] Huang C, et al. Lancet 2020; 395:497-506.

[2] Trepo E, et al. J Hepatol 2016;399-412.

[3] Kovarova M, et al. J Clin Endocrinol Metab 2015;100: E1568-74.

[4] Trottier C, et al. Antiviral Res 2008;80:45-53.

[5] Peter A, et al. Diabetologia 2014;57:2103-7.

[6] Grimaudo S, et al. Preprint at https://www.researchsquare.com/article/rs-40510/v1 (2020)

[7] Mondul A, et al. J Nutr 2015;145:1687-91.

[8] Trasino SE, et al. Sci Rep 2015;5:15893.

FIGURE LEGEND:

Figure 1. Association of rs738409 G allele with COVID-19 severity and prevalent liver disease.

Table S1: Characteristics of UKB participants with a positive SARS-CoV-2 test, according to COVID-19 hospital admission and death status

Characteristic		COVID-19 hospital admission			COVID-19 death			Total (N=1585)
		No (N=826)	Yes (n=759)	P-value*	No (N=1318)	Yes (N=267)	P-value*	
rs738409:g minor allele frequency		22.9%	18.7%	0.007	21.7%	16.9%	0.009	20.9%
mean age, years (sd)		64.9 (9.1)	71.1 (8.0)	<0.001	66.5 (9.1)	74.5 (5.9)	<0.001	67.9 (9.1)
% male gender		43.0%	63.4%	<0.001	49.9%	67.0%	<0.001	52.7%
Ethnic group	White British	78.7%	77.5%	0.543	77.4%	81.7%	0.162	78.1%
	Other	2.7%	2.4%		2.8%	1.1%		2.5%
	Asian	5.1%	4.6%		5.3%	2.6%		4.9%
	Black	3.9%	5.7%		4.6%	5.2%		4.7%
	missing	9.7%	9.9%		9.9%	9.4%		9.8%
mean BMI (sd)		28.0 (5.1)	29.6 (5.5)	<0.001	28.6 (5.3)	29.7 (5.8)	0.002	28.8 (5.4)
% Current smoker		11.4%	14.2%	0.089	12.5%	13.9%	0.550	12.7%
% Type 2 diabetes		5.3%	10.3%	<0.001	6.8%	12.4%	0.002	7.7%
rs11385942:ga minor allele frequency		7.0%	10.9%	<0.001	8.6%	10.3%	0.089	8.9%
rs657152:a minor allele frequency		35.7%	38.2%	0.331	36.8%	37.3%	0.539	36.9%
Commorbidities	% Coronary artery disease	11.6%	27.9%	<0.001	17.2%	30.7%	<0.001	19.7%
	% any liver hospitalisation	3.0%	8.4%	<0.001	5.0%	8.6%	0.020	5.6%
	% cirrhosis hospitalisation	0.5%	3.3%	<0.001	1.4%	3.8%	0.010	1.8%

*P-values calculated using the Chi-squared test for categorical variables, and the Student's T-test for numeric variables.

Table S2: Association between rs738409 G allele and COVID-19 related outcomes in the UK Biobank and validation cohorts.

Cohort	Case (N)	Controls (N)	Genetic model	Adjustment level	OR (rs738409 G)	95% CI	P-value
#1.UK Biobank	COVID-19 hospitalisation (N=759)	SARS-CoV-2 positive without Covid19 hospitalisation (N=826)	Additive	univariate	0.78	0.65-0.92	0.004
				age and sex	0.76	0.63-0.92	0.004
				full adjustment*	0.79	0.64-0.97	0.027
			Recessive**	univariate	0.50	0.30-0.82	0.006
				age and sex	0.48	0.28-0.83	0.008
				full adjustment*	0.47	0.25-0.88	0.018
	COVID-19 hospitalisation with pneumonia (N=450)	SARS-CoV-2 positive without Covid19 hospitalisation with pneumonia (N=1135)	Additive	univariate	0.75	0.61-0.91	0.004
				age and sex	0.74	0.60-0.91	0.004
				full adjustment*	0.74	0.60-0.93	0.009
			Recessive**	univariate	0.54	0.29-1.00	0.050
				age and sex	0.55	0.29-1.02	0.059
				full adjustment*	0.53	0.27-1.05	0.067
	COVID-19 hospitalisation requiring advanced ventilation support (N=76)	SARS-CoV-2 positive without Covid19 hospitalisation requiring advanced ventilation support (N=1509)	Additive	univariate	0.63	0.40-1.00	0.049
				age and sex	0.63	0.40-1.00	0.048
				full adjustment*	0.65	0.40-1.05	0.078
			Recessive**	univariate	insufficient events for recessive model§		
				age and sex			
				full adjustment*			
	COVID-19 mortality (N=267)	SARS-CoV-2 positive without Covid19 death (N=1318)	Additive	univariate	0.73	0.57-0.94	0.013
				age and sex	0.71	0.54-0.92	0.009
				full adjustment*	0.75	0.57-0.98	0.037
			Recessive**	univariate	0.21	0.06-0.66	0.008
				age and sex	0.20	0.06-0.65	0.007
				full adjustment*	0.19	0.06-0.64	0.007
	SARS-CoV-2 positive with prevalent liver disease (N=91)	SARS-CoV-2 positive without prevalent liver disease (N=1494)	Additive	univariate	1.34	0.95-1.89	0.094
				age and sex	1.36	0.97-1.93	0.079
				full adjustment*	1.44	1.01-2.05	0.045
			Recessive**	univariate	1.53	0.64-3.62	0.337
				age and sex	1.62	0.68-3.87	0.277
				full adjustment*	insufficient events for fully-adjusted estimate		
#2.FinnGen	COVID-19 hospitalisation (N=83)	SARS-CoV-2 positive without Covid19 hospitalisation (N=274)	Additive	univariate	0.86	0.54-1.34	0.50
			Recessive**	univariate	0.75	0.21-2.71	0.66
#3.Geisinger Health system (European Ancestry)	COVID-19 hospitalisation (n=165)	SARS-CoV-2 positive without Covid19 hospitalisation (N=689)	Additive	age, sex and principal components	0.89	0.64-1.23	0.47
#4.Grimaudo et al	Severe COVID-19 disease† (N=52)	SARS-CoV-2 positive and mild disease‡ (N=314)	Additive	univariate	0.68	0.41-1.13	0.14
			Recessive**	univariate	0.53	0.12-2.33	0.40
Meta analysis: #2+#3+#4	COVID-19 hospitalisation/severe disease (N=300)	SARS-Cov-2 positive without Covid19 hospitalisation or mild disease (N=1277)	Additive	mixed	0.83	0.66-1.05	0.12
Meta-analysis: #2+#4	COVID-19 hospitalisation/severe disease (N=135)	SARS-Cov-2 positive without Covid19 hospitalisation or mild disease (N=488)	Recessive**	univariate	0.65	0.25-1.71	0.38

*refers to adjustment for: age, sex, month of positive SARS-CoV-2 test, ethnicity, tobacco smoking, diabetes, BMI, rs11385942 (*LZTFL1*), rs657152 (*ABO*), liver disease, cardiovascular disease, and top 5 principal components of genetic ancestry. However, where the outcome of interest is prevalent liver disease, then liver disease is omitted as a covariate.

**recessive model compares individuals with rs738409 GG genotype to those with rs738409 CC/GC genotype

† severe disease defined as admission to intensive care or death

‡ mild disease defined as patients hospitalised for COVID-19 but without complications

§ i.e. zero patients with rs738409 GG phenotype received advanced respiratory support

SUPPLEMENTARY METHODS:

INCLUSION CRITERIA:

The present analysis included all UKB participants with a positive test result for SARs-CoV-2 in the three-month period between 16th March 2020 and 15th August 2020. NHS England Hospital Episode Statistics (HES) and mortality data were complete until 30th September 2020 at the time of analysis. Thus, the 15th August cut-off date ensured there was a minimum of 6 weeks follow-up data between the first positive test for SARs-CoV-2 and any subsequent COVID-19 hospitalisation or death.

Participants who were first or second degree relatives with another participant in the sample were excluded. We identified first or second degree relations via a kinship coefficient ≥ 0.10, as recommended by UKB. We excluded participants with unreliable genetic data, using the UKB Field ID: 22010.

DEFINITION OF OUTCOME EVENTS.

A COVID-19 death was defined through the presence of one or more of the following three ICD-10 codes in any cause of death position. These were: U071 (COVID-19 virus identified); and/or U072 (COVID-19 virus not identified); and/or B342 (Coronavirus infection, unspecified site). Of note, the U072 ICD-10 code is used to denote clinical or epidemiological cases of COVID-19 where laboratory data are inconclusive or not available.

COVID-19 hospital admissions were identified using NHS England HES data. The same three ICD-10 codes (U071; and/or U072; and/or B342) in any diagnostic position were used to identify a COVID-19 related admission. Instances of COVID-19 related pneumonia were identified using the J12.8 ICD10 code (other viral pneumonia, not classified elsewhere) in combination with U071 and/or U072 and/or B342.

Information on advanced respiratory support is collected on the HES critical care dataset. It has been defined as: “i) Invasive medical ventilatory support applied via a trans-laryngeal tracheal tube or applied via a tracheostomy; ii) Bi-level positive airway pressure applied via a trans-laryngeal tracheal

tube or applied via a tracheostomy; iii) Continuous positive airway pressure via a trans-laryngeal tracheal tube; iv) Extracorporeal respiratory support". We used the NHS England HES critical care dataset available for UKB participants to identify individuals receiving this intervention.

Prevalent liver disease was defined through one or more liver-related hospital admission prior to the first SARs-CoV-2 positive test. Liver-related hospital admissions were identified using the ICD10:K70-K77 "diseases of the liver" chapter, or the equivalent ICD9 codes (571-573).

ADJUSTMENT FOR CONFOUNDING FACTORS:

All regression models were adjusted for a broad range of potential confounding factors. These were: age; month of positive SARS-CoV-2 test; BMI; gender; coronary artery disease (CAD); chronic liver disease/liver cirrhosis; current smoker; Type-2 diabetes; rs657152; rs11385942; ethnic group and the top five principal components of genetic ancestry. Individuals with missing data for one or more of these variables were excluded from the final sample.

Ethnic group was defined through a combination of genetic and self-reported information (field IDs: 22006, and 21000). Type 2 diabetes was defined using field ID: 2443: "has a doctor ever told you that you have diabetes?". From here, we excluded participants with a history of gestational diabetes only (field ID:4041) and individuals with evidence of type1 diabetes from hospital admission records (ICD10: E10), or from detailed information collected on diagnosed medical conditions at UKB enrolment during the nurse interview (Field ID: 20002). The presence of Coronary Artery disease (CAD) was inferred through hospital admissions occurring prior to the first positive test for SARs-CoV-2; using ICD10:120-125, and/or ICD9:412[0-9], 413[0-9], 4140, 4141, 4149. As a covariate, liver disease was divided into two categories: those with a hospital admission for cirrhosis, and those with a non-cirrhosis liver related admission. As before, a liver-related admission was defined using the ICD10:K70-K77 liver disease chapter (or equivalent ICD9 codes). Cirrhosis was defined using ICD codes outlined previously by Ratib S, et al (Am J Gastroenterol. 2014:109:190-8). Age was based on age at the time of the first positive test for SARs-CoV-2. Current smoking was inferred only through information provided at UKB enrolment (Field ID:20116). BMI was determined from each

participant's height and weight at the time of their initial UKB assessment visit. Standing height was measured via the Seca202 height measure, whilst body weight was measured from the Tanita BC-418 MA body composition analyser.

ASSESSING VARIABILITY IN rs738409 EFFECT SIZE

We assessed if the rs738409 effect size varied according to three covariates: ethnic group, CAD, obesity, age<65 years and history of prior liver disease. This was done by adding interactions terms into the model (i.e. between rs738409 and the covariate of interest), and performing a likelihood ratio test to gauge whether the interaction model significantly improved model fit.

VALIDATION ANALYSIS:

We retrieved the association between rs738409 G and COVID-19 morbidity (versus SAS-CoV-2 infection but no/minimal morbidity) from three independent studies. These were:

1) The FinnGen population-based biobank study, comprising of 83 patients with a COVID-19 hospital admission versus 274 SARS-CoV-2 positive patients without a hospital admission. (Table S2)
2) Individuals in the Geisinger Health Study dataset of European descent. Comprising of 165 individuals COVID-19 hospitalisation versus 689 SARS-CoV-2 positive patients without a hospital admission. (Table S2)
3) Data from a study by Grimaudo et al, comprising 52 patients admitted to hospital for COVID-19 who either died or required intensive care therapy, versus 314 patients admitted to hospital for COVID-19 but without any complications.

The rs738409:G association from each study was then aggregated through fixed-effect meta-analysis, weighted according to study sample size.