財團法人明日醫學基金會研究計畫申請書

計畫名稱	（中文）宿主遺傳因子與 B 型肝炎表面抗原消失：建立在臺灣人體生物資料庫的兩階段研究計畫	
	（英文）Host Genetic Determinants of Hepatitis B Surface Antigen Seroclearance: A Two-Stage Cross-Sectional and Longitudinal Study from the Taiwan BioBank	
計畫類別	☑個別型	☐整合型
計畫歸屬	☐基礎醫學☐生物醫學☑臨床醫學☐資訊系統☐醫院管理☐整合性醫學研究	
計畫主持人	姓名：許耀峻 Yao-Chun Hsu	機關：**義守大學、義大醫院**
	職稱：**教授**	單位：**醫學研究所**
通訊地址	**高雄市左營區重化街六號 13 樓**	
聯絡電話	**(02)6150011, 0988687726**	
計畫執行期限	自 115 年 1 月 1 日起至 115 年 12 月 31 日止	
共同主持人（一）	姓名： 吳嘉玲	機關：
	職稱：	單位：
共同主持人（二）	姓名：	機關：
	職稱：	單位：
共同主持人（三）	姓名：	機關：
	職稱：	單位：

研究計畫摘要

一、研究目的：

本研究將以「臺灣人體生物資料庫（Taiwan BioBank, TWB）」為基礎，利用橫斷面與縱貫性兩階段設計，系統性地鑑定與 B 型肝炎表面抗原（HBsAg）消失相關的非 HLA 宿主遺傳決定因子，並特別聚焦於非酒精性脂肪肝病（NAFLD）相關基因及多基因風險分數（PRS）的角色，以分離其對 HBsAg 清除的直接效應、透過脂肪肝與代謝途徑的中介效應，以及透過 ALT 升高與抗病毒治療的間接效應。

二、研究背景：

HBsAg 消失被視為慢性 B 型肝炎之「functional cure」，可顯著降低肝硬化及肝細胞癌風險。然而自發性 HBsAg 清除的年發生率僅約 1–2%，其宿主決定因子至今仍未完整闡明。目前已知的遺傳預測因子幾乎均集中於 HLA class II 區域，非 HLA 位點的證據有限。近年多項臨床觀察顯示脂肪肝及代謝異常與 HBsAg 清除呈正相關（Mak 2020；Hsu 2024；Mao 2023），但僅 Hsueh 等（2022）在 2,385 名男性追蹤 30 年的世代中直接檢驗 PNPLA3-148M 變異，發現其同時與 HBsAg 清除及肝癌風險增加有關，呈現「基因脂肪肝易感性」的雙面矛盾效應；該研究未校正 HLA 亦僅限男性，亦未檢驗其他 NAFLD 相關基因（TM6SF2、GCKR、SAMM50、HSD17B13）及多基因風險分數（PRS）。Taiwan BioBank 蓄積了大量含基因型及臨床生化資料之世代，正是回答此一問題的理想平台。

三、研究方法：

本研究採兩階段設計。階段一（橫斷面 discovery）將納入 TWB 中登錄時 HBcAb 陽性之參與者（約 80,000 人，HBsAg+ 約 28,500、HBsAg− 約 50,000），以 HBsAg 陽性（慢性帶原）vs. 陰性（已清除）為結果，進行（1）全基因組關聯分析（GWAS）；（2）候選基因分析（PNPLA3、TM6SF2、GCKR、SAMM50、HSD17B13）；（3）NAFLD PRS 及多種代謝 PRS-wide 分析；（4）HLA × NAFLD PRS 交互作用；（5）MAGMA/SKAT 基因與路徑分析。階段二（縱貫性 replication）將納入 TWB 基線 HBsAg 陽性且完成追蹤者（約 5,000 人，追蹤期間清除者約 200 人），以 cloglog 迴歸搭配 ln(ΔT) offset 處理不等追蹤間隔；核心分析採逐步校正模型（Model 1 基因型＋年齡＋性別＋PC1–10；Model 2 加 HLA；Model 3 加 HBeAg；敏感性 S1–S4 逐步加入抗病毒用藥、ALT、FLI/BMI 以分離路徑 A、B、C 之貢獻），並以因果中介分析分解 NAFLD PRS 總效應。最後以階段一資料訓練、階段二資料外部驗證，建構結合 HLA、NAFLD PRS 及臨床因子的 HBsAg 清除預測模型，並以 DeLong’s test 比較加入 PRS 前後之 ΔAUC。

關鍵詞：B 型肝炎；HBsAg 清除；宿主遺傳；NAFLD；多基因風險分數；Taiwan BioBank；兩階段研究

研究計畫目的及背景說明

Part I. Background and Literature Review

1.1 Clinical significance of HBsAg seroclearance

Loss of hepatitis B surface antigen (HBsAg) is regarded as the functional cure of chronic hepatitis B (CHB) and is associated with markedly reduced risks of cirrhosis and hepatocellular carcinoma (HCC). However, the annual incidence of spontaneous HBsAg seroclearance is only about 1–2%, and its host determinants remain incompletely defined.

1.2 Known predictors of HBsAg seroclearance

Low baseline HBV DNA, low quantitative HBsAg (qHBsAg < 250 IU/mL), HBeAg-negative status, and HBV genotypes A and B (vs. C and D) are viral predictors of seroclearance. Older age increases cumulative clearance, and females clear HBsAg more often than males (pooled OR ≈ 0.54 for male sex). Metabolic factors have emerged as important modulators: Mak et al. (2020, J Hepatol) showed that concomitant hepatic steatosis tripled the probability of HBsAg clearance (HR 3.246); Hsu et al. (2024, CGH) reported 17.6% vs. 9.8% clearance with vs. without MASLD in 4,084 CHB patients; Mao et al. (2023, Hepatology) meta-analyzed 34 studies (n = 68,268) confirming this association (OR 2.22; 95% CI 1.58–3.10). Known host genetic determinants, however, are almost exclusively in the HLA class II region (HLA-DPB1 rs9277535, HLA-DQB1). The TWB own GWAS (Ou 2020) confirmed HLA class II as the dominant determinant of chronic carriage; Kim (2018) reported non-HLA loci (MPEG1, DTX4) using a very small extreme-phenotype design (100 vs. 100) that has not been replicated.

1.3 NAFLD-related genes and HBsAg seroclearance: the only direct evidence

Hsueh et al. (2022, J Hepatocell Carcinoma) followed 2,385 HBsAg-positive male Taiwanese civil servants for up to 30 years and reported that both PNPLA3-

148M and ultrasonographic fatty liver were associated with higher HBsAg seroclearance; homozygosity for PNPLA3-148M simultaneously conferred higher HCC risk (sHR 1.83), illustrating a paradoxical dual effect. Critically, the study was restricted to men, examined only PNPLA3, and did not adjust for HLA.

1.4 Knowledge gaps

(i) Only one study has directly tested an NAFLD gene (PNPLA3) against HBsAg seroclearance, in men only, without HLA adjustment.

(ii) Whether TM6SF2, GCKR, SAMM50, HSD17B13 predict HBsAg clearance is unknown.

(iii) Whether NAFLD polygenic risk scores (PRS) predict HBsAg seroclearance has not been tested.

(iv) Whether NAFLD-gene effects are independent of HLA is the single most critical methodological gap.

(v) Causal pathways (direct vs. steatosis-mediated vs. ALT/treatment-mediated) have not been dissected.

(vi) Whether metabolic factors retain independent effects after adjustment for host genetics is unknown.

(vii) Effects in female carriers are unknown.

(viii) Systematic PRS-wide interrogation against HBsAg clearance has never been performed.

(ix) No prior GWAS has searched for non-HLA loci for HBsAg clearance after HLA adjustment.

Part II. Causal Pathway Framework

We propose four plausible pathways from NAFLD-related genotype to HBsAg clearance. Pathway A (direct effect): NAFLD genotypes alter hepatic lipid metabolism and thereby directly modulate HBV persistence without going through

clinically measurable steatosis, ALT, or antiviral treatment; this is the core hypothesis. Pathway B (steatosis/metabolic mediation): NAFLD genotype → hepatic steatosis/metabolic abnormality → HBsAg clearance. Pathway C (ALT/treatment mediation): NAFLD genotype → elevated ALT → antiviral treatment → treatment-associated clearance. Pathway D (HLA/immune): HLA genotype → antigen presentation → HBsAg clearance, independent of NAFLD genotype. Because genotype is fixed at conception, it is immune to reverse causation and to virtually all conventional confounders except population structure; therefore, main analyses estimate total effects by not adjusting for mediators, and sensitivity analyses progressively add mediators to infer each pathway from effect attenuation.

Part III. Objectives and Hypotheses

Using TWB's cross-sectional registry (~80,000 HBcAb+ participants) and longitudinal follow-up (~5,000 baseline HBsAg-positive participants, ~200 with subsequent seroclearance), we will systematically identify non-HLA host genetic determinants of HBsAg seroclearance, focusing on NAFLD-related loci.

Specific Aim 1 (cross-sectional discovery). H1a: In HBcAb+ individuals, non-HLA host genetic variants are independently associated with HBsAg clearance/persistence; selected NAFLD loci (PNPLA3/SAMM50, TM6SF2, GCKR) may reach genome-wide significance after HLA adjustment. H1b: NAFLD PRS independently predicts HBsAg clearance after adjustment for HLA. H1c: At least one metabolic PRS (NAFLD, BMI, TG, T2DM) independently predicts HBsAg clearance.

Specific Aim 2 (longitudinal replication, causal-pathway dissection). H2a: Top GWAS hits and NAFLD PRS predict incident HBsAg clearance during prospective follow-up. H2b: The total effect of NAFLD PRS can be decomposed into direct, steatosis-mediated, and ALT/treatment-mediated components, with a substantial direct contribution. H2c: Genetic effects differ between treated and untreated participants;

gene × treatment interaction is clinically meaningful.

Specific Aim 3 (integration). H3: A prediction model combining HLA genotype, NAFLD PRS, and clinical factors (age, sex, HBeAg) yields significantly higher AUC than an HLA-only model.

Part IV. Study Design

Two-stage design. Stage 1 (cross-sectional discovery): TWB participants testing HBcAb-positive at enrollment; exclude anti-HCV positive, heavy alcohol use (≥150 mL/week for ≥6 months), QC failure (call rate < 97%, HWE $P < 1\times10^{-6}$), and first- or second-degree relatives. Outcome: HBsAg+ (chronic carrier = case) vs. HBsAg− (cleared = control). Stage 2 (longitudinal replication): TWB participants with baseline HBsAg-positive status and completed follow-up with HBsAg testing and complete genotype data; same exclusions. Outcome: conversion from baseline HBsAg+ to follow-up HBsAg− (binary); secondary endpoints are longitudinal changes in ALT, FLI, FIB-4. Follow-up interval varies (~2–4 years); handled with complementary log-log regression plus ln(ΔT) offset. All covariates use baseline values T_0 to guarantee temporal ordering.

Part V. Exposures

Genetic exposures (both stages, time-invariant): PNPLA3 rs738409 G (additive/dominant/recessive), TM6SF2 rs58542926 T, GCKR rs1260326 T, SAMM50 rs3761472 A, HSD17B13 rs72613567 T, NAFLD PRS-EA (4-SNP East-Asian model, standardized), NAFLD PRS-TWB (16 FLI-GWAS variants from TWB, standardized), and ~8 million imputed SNPs (Stage 1 GWAS only). Metabolic exposures (FLI, BMI, waist circumference, TG, fasting glucose, HbA1c, metabolic syndrome) are supporting variables in Stage 1 (reverse-causation risk because cleared individuals may have already altered metabolism) and use baseline values T_0 in Stage 2.

Part VI. Covariate Strategy Based on Causal Structure

Because genotype is fixed at conception, no post-natal factor can confound a genotype–outcome association; the only true confounder is population structure. Covariates other than ancestry PCs are included for precision, for answering specific scientific questions (e.g., independence from HLA), or for mediation dissection. Crucially, main analyses do not adjust for ALT, FLI, or antiviral treatment, because they lie on the causal pathway; adjusting for them blocks causal paths (Pathways B and C) and may introduce collider bias (ALT is jointly affected by genotype and by viral activity). Roles: PC1–PC10 (confounder, always adjusted); age (precision, always adjusted); sex (precision + effect modifier, adjusted and stratified/interaction); HLA genotype (independent effect, added in Model 2); HBeAg (precision in Stage 2, added in Model 3); ALT, FLI/BMI/MetS, antiviral treatment (mediators, not in main, stepwise in sensitivity); FIB-4 (ambiguous, sensitivity only).

Part VII. Statistical Analyses

Stage 1 (~80,000 HBcAb+). Logistic-regression GWAS for HBsAg+ vs. HBsAg− with age, sex, and PC1–PC10; threshold $P < 5\times10^{-8}$; HLA top hits (rs3077, rs9277535) then added as covariates and GWAS repeated to search for non-HLA loci. Candidate NAFLD SNPs analyzed in a stepwise framework (Model 1: PCs + age + sex; Model 2: + HLA). NAFLD PRS-EA and PRS-TWB analyzed per SD and by quartiles. PRS-wide analysis for NAFLD, BMI, TG, HDL, T2DM, and CRP PRSs with Bonferroni correction. Gene- and pathway-based analyses with MAGMA/SKAT, focusing on lipid metabolism (KEGG hsa00071), lipid-droplet regulation (GO:0005811), cholesterol metabolism, and innate immunity. HLA × NAFLD-PRS interaction tested. All analyses stratified by birth cohort (pre-/post-1986). Power: with 28,500 cases and 50,000 controls, GWAS can detect OR ≥ 1.08 at MAF 30% (80% power).

Stage 2 (~5,000 HBsAg+, ~200 events). Primary model is complementary log-log regression with ln(ΔT) offset; exp(β) interpreted as hazard ratio. Stepwise adjustment: Model 1 (SNP/PRS + age_0 + sex + PC1–10 + offset) → Model 2 (+HLA) → Model 3 (+$HBeAg_0$); sensitivity analyses S1 (+antiviral treatment), S2 (+ALT_0), S3 (+FLI_0 + BMI_0), and S4 (strictest direct effect with mediators all included). Causal mediation analysis decomposes total PRS effect into Pathway B (PRS → FLI_0 → clearance), Pathway C (PRS → ALT_0 → treatment → clearance), and Pathway A (residual direct effect), with bootstrap 95% CIs. Interaction analyses include gene × treatment, gene × steatosis, gene × sex. Secondary analysis of ΔALT, ΔFLI, ΔFIB-4 uses all ~5,000 participants with change-score regression and linear mixed-effects models (SNP × time interaction) to substantially boost power. Additional sensitivity: exclude ALT_0 > 2×ULN; exclude FIB-4 > 2.67; stratify by sex and birth cohort; logistic with ΔT; restrict to never-treated.

Cross-stage integration. Stage-1 top hits ($P < 1\times10^{-5}$ outside HLA) formally replicated in Stage 2 (one-sided $P < 0.05$, concordant direction). Stage-1 and Stage-2 genotype effects combined in fixed-effects meta-analysis. Prediction models built on Stage 1 (base: HLA + age + sex; extended: + NAFLD PRS + metabolic PRSs) with 10-fold cross-validation; ΔAUC evaluated with DeLong's test and externally validated in Stage 2.

Part VIII. Statistical Power

Stage 1 GWAS: 28,500 / 50,000; MAF 30%, detectable OR ≥ 1.08 at $P < 5\times10^{-8}$; MAF 10%, OR ≥ 1.13. Stage 1 candidate PNPLA3: MAF 35%, detectable OR ≥ 1.04 at $P < 0.007$. Stage 1 PRS per SD: detectable OR ≥ 1.04 at $P < 0.007$. Stage 2 candidate/PRS (~200 events): detectable HR ≥ 1.40 at $P < 0.007$. Stage 2 continuous ΔALT (all ~5,000): β ≥ 0.08 SD at $P < 5\times10^{-8}$.

Part IX. Genotype Quality Control

SNP-level: call rate > 97%; HWE P > 1×10^{-6}; MAF > 1% (candidate) / > 0.5% (GWAS). Individual-level: call rate > 97%; sex concordance; PI_HAT > 0.2 exclusion. TWBv1/v2 harmonized over ~100,000 overlapping markers with batch adjusted as covariate. HLA alleles (HLA-DPB1, HLA-DQB1) imputed with HIBAG/CookHLA using East-Asian reference panel with posterior probability > 0.5. Ancestry PCs: PC1–PC10 from PCA; >6-SD outliers excluded.

Part X. Expected Contributions

Scientifically: (i) first genome-wide search for non-HLA determinants of HBsAg clearance after explicit HLA adjustment in a large East-Asian cohort; (ii) systematic test of multiple NAFLD-related variants and PRSs, extending host genetics of HBV beyond HLA; (iii) first formal dissection of direct, steatosis-mediated, and ALT/treatment-mediated pathways; (iv) clarification of the paradoxical dual effect of PNPLA3 reported by Hsueh et al. Translationally, if NAFLD PRS independently predicts clearance, combining with HLA enables precision risk stratification; if Pathway B is substantial, metabolic intervention is plausible as functional-cure adjunct; if Pathway A exists, hepatic lipid-metabolism pathways may be novel HBV therapeutic targets.

Part XI. Limitations and Mitigation

No quantitative HBV DNA (HBeAg proxy in Stage 2; genotype is free from reverse causation). No quantitative HBsAg (acknowledged). No HBV genotype (Taiwanese B/C roughly balanced). Reverse causation for metabolic factors in Stage 1 (Stage 1 focuses on genotype; pathway dissection deferred to Stage 2). Short follow-up (2–4 years) in Stage 2 (~200 events): candidate/PRS power adequate, GWAS discovery via Stage 1. Imprecise treatment information (main analyses estimate total effect; interaction in sensitivity). Cross-sectional design lacks event timing (addressed by Stage 2 longitudinal). Survivor bias (bias toward null). Over-adjustment for

mediators (main analyses exclude them; stepwise sensitivity only).

References:

1. Mak LY, Hui RWH, Fung J, et al. Effect of coexisting non-alcoholic fatty liver disease on the clinical outcomes of chronic hepatitis B: a prospective study. J Hepatol. 2020.

2. Hsu YC, Huang DQ, Nguyen MH. Global burden of hepatitis B virus: current status, missed opportunities and a call for action. Nat Rev Gastroenterol Hepatol. 2023;20(8):524-537.

3. Hsu YC, Lin YH, Lin JT, et al. Risk of hepatocellular carcinoma and HBsAg seroclearance in chronic hepatitis B patients with and without metabolic dysfunction-associated steatotic liver disease. Clin Gastroenterol Hepatol. 2024.

4. Mao X, Cheung KS, Peng C, et al. Steatotic liver disease and hepatitis B surface antigen seroclearance: a systematic review and meta-analysis. Hepatology. 2023.

5. Ou CY, Chen TH, Hsu YC, et al. Genome-wide association study of chronic hepatitis B virus infection in a Taiwan Biobank cohort. J Hepatol. 2020.

6. Kim YJ, Kim HY, Lee JH, et al. A genome-wide association study identifies non-HLA loci for hepatitis B virus persistence. Liver Int. 2018.

7. Hsueh CY, Hsu CS, Huang YT, et al. PNPLA3 variant and fatty liver are associated with HBsAg seroclearance and hepatocellular carcinoma in chronic hepatitis B. J Hepatocell Carcinoma. 2022.

8. Romeo S, Kozlitina J, Xing C, et al. Genetic variation in PNPLA3 confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2008;40(12):1461-1465.

9. Kozlitina J, Smagris E, Stender S, et al. Exome-wide association study identifies a TM6SF2 variant that confers susceptibility to nonalcoholic fatty liver disease. Nat Genet. 2014.

10. Abul-Husn NS, Cheng X, Li AH, et al. A protein-truncating HSD17B13 variant and protection from chronic liver disease. N Engl J Med. 2018;378(12):1096-1106.

11. Anstee QM, Darlay R, Cockell S, et al. Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort. J Hepatol. 2020.

12. Terrault NA, Lok ASF, McMahon BJ, et al. Update on prevention, diagnosis, and treatment of chronic hepatitis B: AASLD 2018 hepatitis B guidance. Hepatology. 2018;67(4):1560-1599.

13. Chen CL, Yang JY, Lin SF, et al. Taiwan Biobank: A rich biomedical research database of the Taiwanese population. Cell Genomics. 2023.

14. VanderWeele TJ. Mediation analysis: a practitioner's guide. Annu Rev Public Health. 2016;37:17-32.

15. Hsu YC, Yip TCF, Ho HJ, et al. Development of a scoring system to predict hepatocellular carcinoma in Asians on antivirals for chronic hepatitis B. J Hepatol. 2018;69(2):278-285.