# Supplementary Materials for Exploratory analysis of HLA variants and occult HBV infection: a multicenter case-control study in Chinese blood donors

Lin Zhou *et al.*

\*Corresponding author. Email: lunan99@163.com

# Supplementary Methods

## Sample processing and sequencing

Peripheral blood samples (5 mL per EDTA-coated tube, 5 tubes total) were processed for plasma separation and cellular component isolation. All aliquots were stored at -80°C until analysis. After centrifugation at 1,200 × g for 10 min at 4°C, genomic DNA was extracted using the QIAamp DNA Blood Mini Kit according to the manufacturer's protocol. DNA quality was assessed using an Agilent 5400 Fragment Analyzer, with the following inclusion criteria: uncontaminated genomic DNA with ≥0.3 μg input and predominant fragment sizes >3,000 bp. Sequencing libraries were prepared using the Agilent SureSelect Human All Exon V6 kit and sequenced on an Illumina NovaSeq 6000 platform (150 bp paired-end reads).

## HBV detection methods

All HBV serological testing was performed using ARCHITECT immunoassays (Abbott Diagnostics), including the HBsAg Qualitative II assay (limit of detection [LOD]: 0.05 IU/mL) with confirmatory testing for samples showing reactive results (signal-to-cutoff [S/CO] ratios 0.8-5.0). Additional serological markers (anti-HBs, HBeAg, anti-HBe, anti-HBc) were similarly analyzed using ARCHITECT platforms from Abbott Diagnostics (Ireland and Germany). Nucleic acid detection employed Roche Molecular Systems assays: qualitative screening used the Cobas TaqScreen MPX Test v2.0 (LOD: 2.3 IU/mL), while viral load quantification utilized the AmpliPrep/COBAS Taqman HBV Test v2.0 (LOD: 9 IU/mL). All procedures followed manufacturers' protocols with appropriate quality controls.

## Allele-specific analysis for HLA variants

The allele-specific analysis was conducted through a three-stage approach to ensure comprehensive evaluation of HLA variants. For alleles with population frequencies exceeding 5%, we performed: (1) direct comparison against all other alleles combined, (2) pairwise comparisons between specific allele combinations, and (3) dominant/recessive model assessments. This multi-faceted approach accounted for the complex genetic architecture of HLA polymorphisms while maintaining statistical power. The pipeline incorporated specific quality checks for multiallelic sites, including strand bias verification and Hardy-Weinberg equilibrium testing (P>0.01 threshold).

## Functional and structural characterization​ of HLA variants

We performed comprehensive functional annotation of HLA missense variants using ProtVar (v1.4) to obtain evolutionary conservation scores and predicted stability changes. Structural analyses were conducted using experimentally determined crystal structures of HLA-A\*02:01 (PDB: 6Q3K), HLA-C\*04:01 (PDB: 5W67) and HLA-DQA1\*05-DQB1\*02 (DQ2; PDB:1S9V) as reference templates. The HLA-A p.Ala114Asp variant structure was generated through in silico mutagenesis in PyMOL (v2.5), with particular focus on conformational changes in the antigen-binding groove. Comparative structural analysis included electrostatic potential mapping (APBS plugin, 0.15M ionic strength, 298K temperature), hydrophobicity analysis using PyMOL's hydrophobicity scale, and surface property visualization with standardized rendering parameters. For the HLA-DQA1 p.Gln57Glu variant, we retrieved and compared the 2D chemical structures of glutamine (Gln, PubChem CID: 5961) and glutamate (Glu, PubChem CID: 33032) to analyze the functional impact of this substitution. All paired figures were rendered with consistent lighting and perspective parameters to enable direct visual comparison.

## ​Binding affinity prediction

Known HBV surface antigen (HBsAg) epitopes (HBV-B and -C included) were extracted from the Immune Epitope Database (IEDB, http://www.iedb.org) using the following parameters: (1) Organism: Hepatitis B virus (HBV), (2) Antigen: "Surface antigen" or "HBsAg," (3) Epitope type: Linear peptides (excluding discontinuous/conformational epitopes), (4) Host: Homo sapiens, and (5) Assay criteria: Positive binding (IC50 ≤ 500 nM) to HLA class I/II molecules. For class I epitopes, we prioritized 8-12mer peptides. For class II epitopes, we prioritized 15-20mer peptides. NetMHCpan (v4.1) and NetMHCstabpan (v1.0) were used for peptide-HLA I binding affinity and stability assessments, and NetMHCIIpan (v4.3) for peptide-HLA II binding affinity. MHCflurry (v2.0) was also used for HLA class I binding affinity predictions, complementing NetMHCpan results with an alternative neural network-based approach.

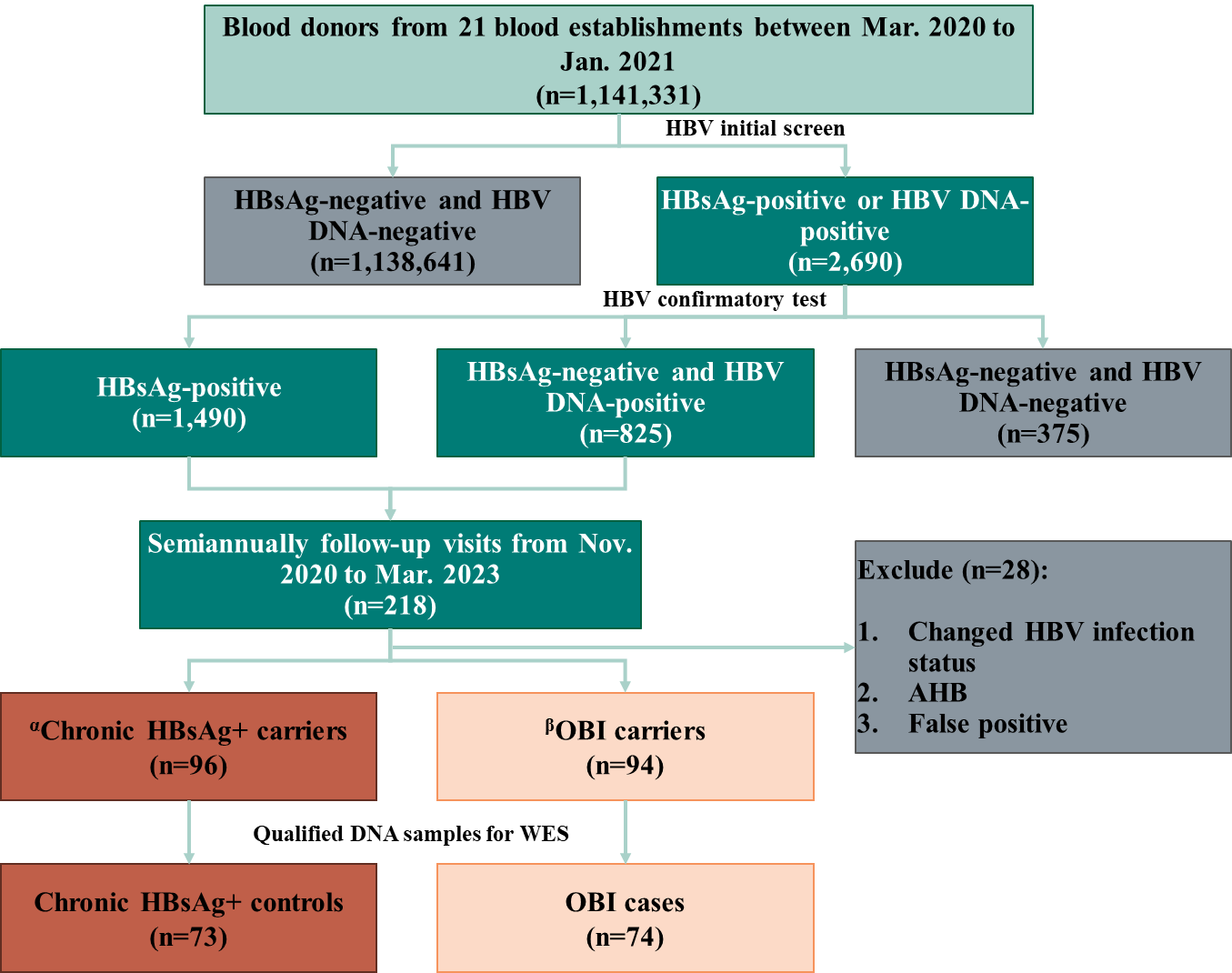
## HLA typing by HLA-HD

HLA-DRB4 typing was performed using HLA-HD software (v1.7.0) on the WES data with default parameters. HLA-HD aligns qualified sequencing reads to the IPD-IMGT/HLA database (release 3.52.0) and employs a Bayesian statistical framework to assign HLA alleles at 6-digit resolution.

## Statistical analysis of HLA variant effects​ on OBI susceptibility

The predictive performance of the HLA-enhanced models was assessed through a multi-step analytical approach. Model discrimination was evaluated using receiver operating characteristic (ROC) curve analysis, with 1000 bootstrap replicates performed to estimate the mean area under the curve (AUC), standard deviation, and 95% confidence intervals. The DeLong test for correlated ROC curves was applied to compare nested models (S-G vs. S and B-G vs. B), with statistical significance determined through 1000 permutations. Reclassification analysis was conducted to quantify improvements in risk stratification, calculating net reclassification improvement (NRI) and integrated discrimination improvement (IDI) metrics. Continuous NRI was derived by comparing predicted probabilities between models, while categorical NRI assessed movement across predefined risk thresholds (low, intermediate, high). Reclassification tables were generated to document shifts in individual risk classifications between baseline and HLA-enhanced models. All statistical analyses were performed in R version 4.3.1 using the pROC, PredictABEL, and rms packages, with a 7:3 random split for training and validation subsets maintained across all comparisons to ensure consistency.

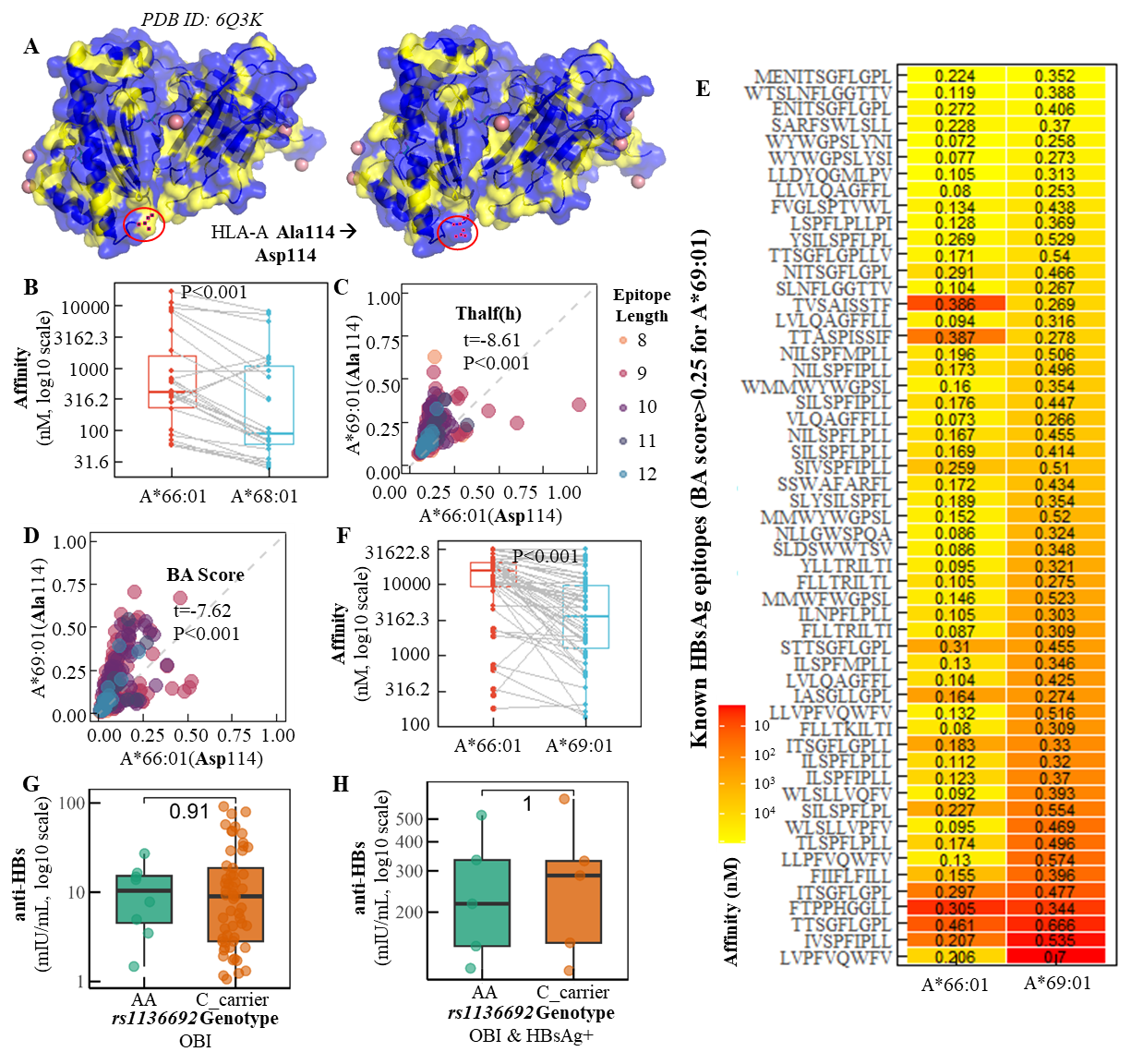
# Supplementary Figures

****

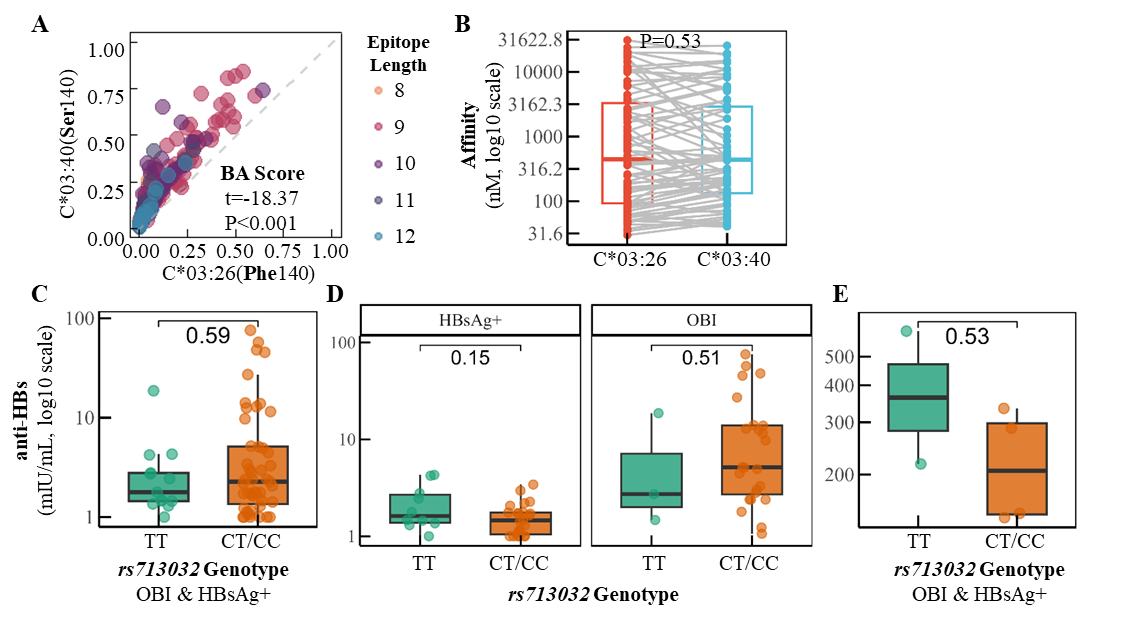
**Figure S1. Flow chart of multi-center cohort selection and longitudinal follow-up workflow for OBI identification in Chinese blood donors.** Initial HBV screen was conducted in 21 blood establishments in China during March 2020 to January 2021. 2690 blood donors were tested positive for HBsAg or HBV DNA. HBV confirmatory test was conducted by National Center for Clinical Laboratories (NCCL). 2315/2690 donors were confirmed. 218/2315 donors agreed and participated in a two-year follow-up visit. After two years’ follow up, 190/218 participants exhibited consistent CHB infection state. 147/190 participants provided qualified DNA samples for WES including 74 occult and 73 overt HBV infection.

***α.*** Chronic HBsAg+ carriers were defined as those who maintain persistent HBsAg+ from initial screening, confirmatary test to semi-anual follow-up visits (1-6 times).

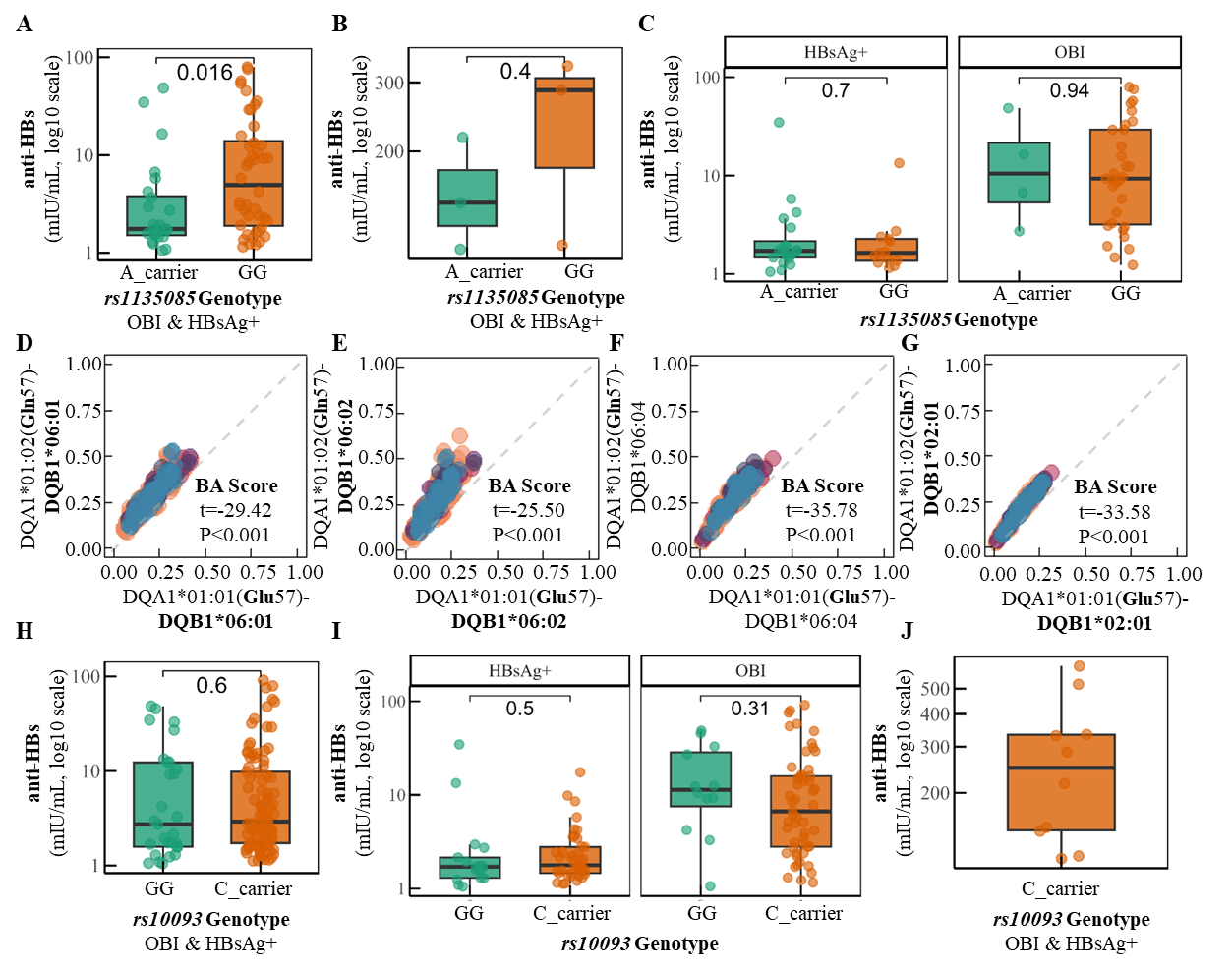
***β.***OBI were defined as those who maintain persistent HBsAg- with detectable HBV DNA<200 IU/mL from initial screening, confirmatary test to semi-anual follow-up visits (1-6 times).

****

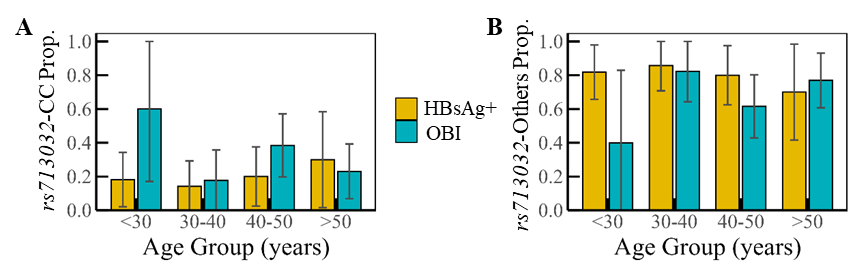
**Figure S2. Comparative analysis of HLA-A Ala114 vs. Asp114 on HBsAg binding stability, affinity and anti-HBs antibody response.** **(A)** Hydrophobicity shift analysis comparing Ala114 variant (hydrophobic, yellow; associated with high OBI susceptibility) and Asp114 variant (hydrophilic, blue) in the HLA-A P4 pocket. **(B-C)** Impaired HBsAg binding capacity of HLA-A Asp114 variant (A\*66:01) vs. Ala114 (A\*69:01). **(B)** Scatter plot showing reduced complex half-life (Thalf) for A\*66:01 compared to A\*69:01 (t=-8.61, P<0.001, NetMHCstabpan prediction). Each dot represents an IEDB-curated HBsAg epitope (8-12 aa) from HBV genotype B/C, colored by epitope length. **(C)** A\*66:01 showing decreased binding affinity (t=-7.62, P<0.001, NetMHCpan). **(D)**​​ Significant lower peptide-MHC affinity (MHCflurry) for Asp114 variant (A\*66:01) vs. Ala114 variant (A\*68:01) for HBsAg epitopes with A\*68:01 BA score>0.25 (paired Wilcoxon signed-rank test, V=349, P<0.001). Boxplots with paired points (log10 nM). **(E-F)** MHCflurry evaluation for HBsAg epitopes (BA score>0.25 for A\*69:01). **(E)** Heatmap of HBsAg epitopes (BA score>0.25 for A\*69:01) with color intensity representing MHCflurry affinity value (nM) and numerical values showing NetMHCpan BA scores. **(F)** Significant affinity reduction for A\*66:01 (paired Wilcoxon signed-rank test, V=1,464, P<0.001, MHCflurry). **(G)**​​ For anti-HBs<100 mIU/mL, no significant difference in anti-HBs levels was observed between C allele carriers (n=57, median=7.99 mIU/mL; associated with high OBI susceptibility) and AA homozygotes (n=8, median=9.83 mIU/mL) among OBI cases (U=234, P=0.91). **(H)**​​ For anti-HBs ≥100 mIU/mL, neither AA homozygote (n=5, median=216.0 mIU/mL) nor C allele carriers (n=5, median=285.7 mIU/mL) showed a significant difference (U=12, P=1).



**Figure S3. Functional impact of HLA-C Phe140 variant (C\*03:26) on HBsAg binding and anti-HBs antibody response.​** **(A-B)** Functional evaluation of HLA-C Phe140 variant (C\*03:26; associated with low OBI susceptibility) vs. Ser140 (C\*03:40). **(A)** Scatter plot showing decreased binding affinity for C\*03:26 compared to C\*03:40 (t=-18.37, P<0.001, NetMHCpan prediction). Each dot represents an IEDB-curated HBsAg epitope (8-12 aa) from HBV genotype B/C, colored by epitope length. **(B)** For HBsAg epitopes with BA score>0.25 for C\*04:40, no significant difference in MHCflurry affinity values was observed between C\*03:26 and C\*03:40 (paired Wilcoxon signed-rank test, V=1,134, P=0.53). **(C-E)** Comparison of anti-HBs level between HLA-C Phe140 homozygotes (TT genotype; associated with low OBI susceptibility) and CT/CC genotypes. **(C)** For anti-HBs<100 mIU/mL, no difference between CT/CC (n=47, median=1.27 mIU/mL) and TT (n=13, median=0.78 mIU/mL) (Mann-Whitney U=275, P=0.59). **(D)** Consistent null associations in both chornic HBsAg+ (P=0.33) and OBI groups (P=0.51). **(E)** For anti-HBs>100 mIU/mL, no statistical difference between CT/CC (n=4, median=216.25 mIU/mL) and TT (n=2, median=412.36 mIU/mL) (U=6, P=0.53).

****

**Figure S4. HLA-DRB4 and HLA-DQ polymorphisms differentially regulate anti-HBs antibody levels and HBsAg binding affinity.** **(A-C)** Comparison of anti-HBs level between HLA-DRB4 A allele carrier (GA/AA; associated with lower OBI susceptibility) and GG homozygotes at *rs1135085*. **(A)** For anti-HBs<100 mIU/mL, GG homozygotes (n=47, median=4.01 mIU/mL) showing higher anti-HBs levels than A allele carriers (n=29, median=0.76 mIU/mL) (Mann-Whitney U=340, P=0.016). **(B)** For anti-HBs>100 mIU/mL, no statistical difference between GG homozygotes (n=3, median=285.69 mIU/mL) and GA/AA (n=3, median=107.96 mIU/mL) (U=2, P=0.4). **(C)** For anti-HBs<100 mIU/mL, comparable levels between GG and GA/AA in both chornic HBsAg+ (P=0.7) and OBI groups (P=0.94). **(D-G)** Comparative binding analysis of DQA1 \* 01:02 (Gln57) versus DQA1 \* 01:01 (Glu57) paired with different DQB1 genotypes (DQB1\*06:01, DQB1\*06:02, DQB1\*06:04 and DQB1\*02:01, respectively, paired t-test, all P<0.001). **(H-J)** Comparison of anti-HBs level between HLA-DQA1 Glu57 homozygotes (GG genotype) and Gln57 carrier (CG/CC genotypes). **(H)** For anti-HBs<100 mIU/mL, no difference between CG/CC (n=47, median=1.92 mIU/mL) and GG (n=13, median=1.73 mIU/mL) (Mann-Whitney U=1,274, P=0.6). **(I)** Consistent null associations in both chornic HBsAg+ (P=0.5) and OBI groups (P=0.31). **(J)** Donors with anti-HBs>100 mIU/mL exclusively carried the Gln57 variants (n=10, 4 CG amd 6 CC genotype at *rs10093*).



**Figure S5. Age-stratified distribution of *rs713032* (HLA-C p.Ser140Phe) genotypes in chronic HBsAg+ (yellow) and OBI (blue) groups. (A)** CC genotype (associated with high OBI susceptibility) proprotion, **(B)** Other genotypes proprotion. Error bars indicate 95% CIs. In the <30y group, OBI showing higher CC genotype prevalence (60.0% [17.1-100], 3/5 cases) than HBsAg+ (18.2% [2.1-34.3%], 4/22 controls), though Fisher’s exact test showed marginal significance (P=0.091, OR=6.15 [0.53-97.90]). HBsAg+ CC proportions were consistently lower than OBI in <30y, 30-40 (14.3% vs. 17.6%) and 40-50y (20.0% vs. 38.5%), but reversed in >50y (30.0% vs. 23.1%). Note the exceptionallly wide CI for OBI <30y group reflects limited sample size (n=5).