Genome-wide Association Study in the Taiwan Biobank Identify Genetic Determinants of *Helicobacter pylori* Infection

Yi-Jiun Pan[1], Wei-Yi Kao[2], Chi-Fung Cheng[2], Shi-Heng Wang[2]
[1]Department of Microbiology and Immunology, School of Medicine, College of Medicine, China Medical University, Taichung, Taiwan
[2]National Center for Geriatrics and Welfare Research, National Health Research Institutes, Zhunan, Taiwan

Background: *Helicobacter pylori* (*H. pylori*) infection contributes significantly to the disease burden, especially in East Asia. The lack of genome-wide scans in East Asians underscores the need for large-scale studies.

Methods: We performed a genome-wide association study in the Taiwan Biobank of over 100,000 individuals of East Asian populations. With linkage to nationwide medical records, we defined symptomatic *H. pylori* infection by ICD code (ICD-9 code: 041.86; ICD-10 code: B96.81). We utilized the HIBAG package to establish a Taiwanese population-based HLA imputation reference and applied this model to impute HLA regions, including locus *A*, *B*, *C*, *DPA1*, *DPB1*, *DQA1*, *DQB1*, and *DRB1*, with a resolution reaching field2 (specific HLA protein).

Results: A total of 6,072 subjects with *H. pylori* infection and 100,509 *H. pylori*-negative controls were included in the genome-wide association analysis. We identified 85 genome-wide significant variants, all located in the *HLA* class II region. Three *HLA* class II alleles—*DQA1*01:03*, *DRB1*08:03*, and *DQB1*06:01*—were associated with increased infection risk, while *DQB1*05:02* conferred a protective effect.

Conclusions: Through a genome-wide scan of *H. pylori* genetic susceptibility in a large Taiwanese population-based cohort, followed by *HLA* allele association analyses, we identified genetic predisposition in the *HLA* class II region associated with *H. pylori* infection.