論文の内容の要旨

論文題目　Genome-wide association study and HLA fine-mapping for childhood steroid-sensitive nephrotic syndrome in the Japanese population
(日本人小児ステロイド感受性ネフローゼ症候群に関するゲノムワイド関連解析とHLA マッピング)

氏名　賈　暁媛

Nephrotic syndrome (NS) is a kidney disorder characterized by heavy proteinuria, hypoalbuminemia, edema and dyslipidemia. NS can be classified as idiopathic NS (INS) and secondary NS. INS is the most common glomerular disease in children, accounting for about 90% of children with NS. In Japan, the estimated incidence of INS is 6.49 cases/100,000 children annually. Between 80 and 90% of patients achieve remission with steroid therapy (steroid-sensitive NS [SSNS]), whereas 10-20% exhibit resistance to the therapy (steroid-resistant NS [SRNS]).

The underlying pathogenesis and genetic background of INS is not well understood. Previous genetic studies were mainly conducted using linkage analyses and candidate gene methods. More than 45 podocyte-related genes have been reported as disease-associated genes in patients with SRNS of hereditary NS. However, the knowledge of specific genetic mutation for SSNS is still limited. As a powerful approach investigate disease-associated genetic factors, genome-wide association studies (GWASs) have also been utilized in NS field. *HLA-DQA1* missenses were reported as susceptibility factors associated childhood SSNS in South Asian population and Europeans. Common variant in *GPC5* was identified as a genetic factor contributing to acquired NS in Japanese adults. However, there's no published GWAS working on childhood-onset SSNS in the Japanese population yet.

In the first part of the current study, I conducted a two-stage GWAS to identify the genetic factors in susceptibility to childhood SSNS in the Japanese population. In the discovery stage, 224 patients with childhood SSNS and 419 adult healthy controls were genotyped using Affymetrix 'Japonica Array'. After quality control (QC) procedure, 224 cases and 412 controls with 495,895 autosomal SNPs were retained for association analysis. The most significant association was detected in the *HLA-DR/DQ* region using Cochran-Armitage trend test (rs4642516, $P_{GC\text{-}corrected}=5.44\times10^{-10}$). Then, whole-genome imputation was performed using 2KJPN reference panel, which was generated using whole-genome sequence data of 2,049 healthy Japanese individuals by Tohoku Medical Megabank. 224 cases and 412 controls with 4,105,543 autosomal SNPs and short insertions and deletions (INDELs) passed QC and were included in the association analysis. 10 SNPs and short INDELs were selected from *HLA-DR/DQ* region and another two candidate loci ($P_{GC\text{-}corrected} < 1\times10^{-5}$) for further validation and replication. 5 SNPs were genotyped by DigiTag2 assay successfully in discovery cases as validation (mean concordance 97%).

To confirm the association, candidate variants were genotyped in an independent Japanese sample set including 216 cases and 719 healthy controls by DigiTag2 assay. Two SNPs from *HLA-DR/DQ* region were replicated (rs4642516, 21kb 5'of *HLA-DQB1*, $P=6.69\times10^{-13}$; rs3134996, 705bp 5' of *HLA-DQB1*, $P=9.44\times10^{-10}$) using Cochran-Armitage trend test. Combined analyses were performed using data in discovery stage and replication stage. The two SNPs showed significant associations under allelic and dominant model [rs4642516, combined $P=7.84\times10^{-23}$, odds ratio (OR)= 0.33 under allelic model, combined $P=4.87\times10^{-23}$, OR= 0.29 under dominant model; rs3134996, combined $P=1.72\times10^{-25}$, OR=0.29 under allelic model, combined $P=9.57\times10^{-21}$, OR= 0.28 under dominant model]. There was no secondary signal in *HLA* region when conditioning on the top SNP in *HLA-DR/DQ* region. 9.7% of variance (Nagelkerke's pseudo-R^2) in Japanese childhood SSNS can be explained by rs4642516 (the top SNP in *HLA* region).

To further fine-map the *HLA* association detected by SNP-based analyses, HLA-imputation and direct HLA-typing were conducted at 4-digit level. First, 6 classical *HLA* genes in both class I and class II regions (*HLA-A*, *-C*, *-B*, *-DRB1*, *-DQB1*, and *-DPB1*) were imputed in the discovery sample set (224 cases and 412 controls) using Japanese-specific reference by 'HIBAG' R package. A total of 197 cases and 411 controls passed post-imputation QC (call-threshold>0.4) and were included in subsequent analyses. Significant *HLA* alleles and haplotypes were identified using imputed data. *HLA-A*02:06* (Pc=8.84×10^{-3}, OR=1.91), *HLA-DRB1*08:02* (Pc=2.56×10^{-4}, OR=2.79), and *HLA-DQB1*03:02* (Pc=4.14×10^{-4}, OR=2.08) were highly associated with the susceptibility to Japanese childhood SSNS. *HLA-A*33:03* (Pc=2.43×10^{-3}, OR=0.30), *HLA-B*44:03* (Pc=1.25×10^{-4}, OR=0.13), *HLA-C*14:03* (Pc=5.25×10^{-5}, OR=0.13), *HLA-DRB1*13:02* (Pc=7.47×10^{-4}, OR=0.18), and *HLA-DQB1*06:04* (Pc=3.40×10^{-5}, OR=0.07) showed significantly protective effects. *HLA-DRB1*08:02-DQB1*03:02* (*Pc*=4.34×10^{-6}, OR=4.33) and *HLA-DRB1*13:02-DQB1*06:04* (*Pc*=6.17×10^{-5}, OR=0.07) demonstrated the most significant positive and negative association, respectively. As validation, HLA-DRB1/-DQB1 genotyping was conducted in the discovery sample set using PCR-SSO (sequence-specific oligonucleotide probing) method on Luminex platform. The concordance rate between the imputed (passed post-imputation QC) and genotyped *HLA-DRB1/-DQB1* genotypes was 99.2%.

As replication, *HLA-DRB1/-DQB1* genotypes of 213 patients and 710 controls were obtained successfully by direct HLA-typing or high-accuracy HLA-imputation. Almost all the significantly associated *HLA-DRB1/-DQB1* alleles and *HLA-DRB1-DQB1* haplotypes were replicated except for *HLA-DQB1*03:02*. Combined analyses were conducted using validation dataset and replication dataset. *HLA-DRB1*08:02* (*Pc*=1.82×10^{-9}, OR=2.62), and *HLA-DQB1*06:04* (*Pc*=2.09×10^{-12}, OR=0.10) were considered as primary *HLA* alleles with Japanese childhood SSNS. *HLA-DRB1*08:02-DQB1*03:02* (*Pc*=7.01×10^{-11}, OR=3.60) was identified as the most significant susceptibility haplotype and showed more significant and stronger association than single *HLA* allele. *HLA-DRB1*13:02-DQB1*06:04* (*Pc*=4.18×10^{-12}, OR=0.10) was identified as the most significant protective haplotype.

Interestingly, consistent with previous researches in the Japanese population, single-tag SNP rs3129888 works with high sensitivity and specificity for capturing *HLA-DRB1*08:02* and *HLA-DRB1*08:02-DQB1*03:02* in this study. Single-tag SNP rs3129888 could be used as an economical biomarker to capture the susceptibility genetic factors in clinical practice.

This study is the first GWAS for childhood SSNS in the Japanese population. Using hypothesis-free method, the predominant role of *HLA-DR/DQ* region in the pathogenesis of childhood SSNS was detected and confirmed in the Japanese population. 9.7% of variance (Nagelkerke's pseudo-R^2) in Japanese childhood SSNS can be explained by rs4642516 (the top SNP in *HLA* region). *HLA* association was dissected by high-resolution HLA fine-mapping. Primary *HLA* alleles and disease-associated *HLA-DRB1-DQB1* haplotypes were identified, providing better understanding of the disease mechanism from molecular and functional aspects. In the present study, the significantly associated *HLA* alleles were consistent with previous Japanese studies, but different from the reports in other populations, suggesting the disease-associated *HLA* alleles may vary depending on geographic or ethnic origins. The findings indicate that HLA fine-mapping after SNP-based analysis is essential for a better understanding of disease mechanism. The current sample size put limits on the ability of detecting genetic variants with modest associations. GWAS with larger sample size in the Japanese population is still needed in the future, with the purpose to identify more disease-associated loci (especially in non-*HLA* regions).