論文の内容の要旨

Title: Genomic analyses of primary open angle glaucoma and major serum protein components, and functional analysis of CADM1 in lung adenocarcinoma.

(開放隅角緑内障、及び主要血清タンパク質分画のゲノム解析、並びに肺腺がんにおける CADM1 の機能解析.)

ワイル　モハメド　サイド　オスマン

Wael Mohammed Saeed Osman

Part I: Genome-wide association study of primary open-angle glaucoma in the Japanese population

Background

Glaucoma is one of leading causes of adult visual impairment and irreversible blindness worldwide. It is a complex, chronic neurodegenerative optic neuropathy resulting in a thinning of the neuro-retinal rim of the optic nerve and visual-fields defect. Primary open angle glaucoma POAG (open iridocorneal angle) is the most common form of the disease. A large cohort study in Tajimi, Japan, implicated the prevalence of glaucoma and POAG to be approximately 5.0% and 3.9%, respectively, in individuals of > 40 years old. Linkage analyses indicated three genes encoding myocilin (*MYOC*), optineurin (*OPTN*) and WD-repeat domain 36 (*WDR36*) as causes of the disease. Recently, GWAS for POAG and candidate genes were reported; however, the pathophysiology of the disease remains not fully understood.

Methodology

For the GWAS, 1,394 subjects who were diagnosed as OPAG were selected from BioBank Japan (BBJ) mainly due to the glaucomatous optic neuropathy with corresponding visual field deficit together with 6,599 controls that were not known to have glaucoma according to their clinical records. For a replication study, another cohort of 1,802 POAG cases and 7,212 controls was also obtained from BBJ.

GWAS genotyping: using Illumina HumanOmniExpress BeadChip (733,202 SNPs). Quality control (Q.C) measures indicated as following: call rate ≥ 0.99 in both cases and controls, *P*-value of the test of Hardy-Weinberg equilibrium ≥ 1.0×10^{-6} in controls, and minor allele frequency ≥ 0.01. Relatives were excluded using identity-by-state (IBS) approach in PLINK 1.0.6. I applied EIGENSTRAT 3.0 to perform principle component analysis (PCA) using the case-control samples to excluded non-Japanese individuals.

Replication-stage genotyping: I applied the multiplex PCR-based Invader Assay method.

Results and Discussion

PCA indicated that all cases and controls were of Asian ancestry. Genomic inflation factor (λ_{GC}) of 1.30, so I applied a logistic regression model incorporating age, gender, and the first two eigenvectors as covariates for the analysis, and I applied the genomic control method (λ_{GC} =1.00).

33 SNPs from the GWAS stage which possibly associated with POAG ($P < 1 \times 10^{-4}$) were selected for replication analyses. Among them, 2 SNPs; rs1063192 in *CDKN2B* gene on chromosome 9p21, and rs10483727 on

14q23 were significantly replicated (P_{meta} = 5.2 x 10^{-11}, OR = 0.75 and 9.49 x 10^{-8}, OR = 0.79, respectively). For these two SNPs, the minor alleles are protective from the disease and risk alleles have additive effects in the development of the disease. Individuals with four risk alleles have 2.26 times higher risk for the disease than those with 0-1 risk alleles.

Among genes that were previously reported as candidates for POAG, only SNPs in caveolin 1 and 2 (*CAV1-CAV2*) were replicated in a comparable OR (top SNP rs7795356; P = 0.0073, OR = 1.31).

No association with neither a SNP with the diseases clinical parameter vertical cup-disc ratio (VCDR) (number = 1,082 cases) (P = 0.057). However, for intra-ocular pressure IOP (number = 2,434), we observed a strong association of a SNP rs10483727 (P = 3.75 x 10^{-6}).

Chromosome 9p21 was implicated as a hotspot locus showing the association with various diseases and rs1063192 reported to be associated with VCDR and POAG in the Caucasians. *CDKN2B* and *CDKN2A* genes are expressed in human ocular tissues, and were found to be upregulated in response to elevation of intraocular pressure. The SNP rs1063192 is located in a binding site of a regulatory miRNA in the 3`UTR of *CDKN2B*.

The SNP rs10483727 in 14q23 is located between two homologous genes belonging to the SIX/Sine oculis homeobox family; *SIX1* and *SIX6* (also known as Optx2). The SIX6 mRNA was reported to be expressed in retina and optic nerve and mutations in the gene have been found in patients with defects in eye development. Furthermore, rs10483727 also reported to be associated with VCDR and POAG in the Caucasians.

Re-sequence analysis for *SIX6* gene in 191 cases and controls identified seven variations identified, four are novel and three of them are rare variations. Functional-wise; variants in codons 140 and 213 are missense variations although they are very rare (0.1%).

<u>Conclusion</u>

I reported results of the GWAS and replication study for POAG using a total of 3,196 cases and 13,811 controls in the Japanese population. This report confirmed the 9p21 and 14q23 as common susceptibility loci for POAG among different ethnic populations. Minor alleles of each of the representative SNPs are likely to be protective for the disease.

<u>Part II: Genome-wide association study of major components of serum proteins in the Japanese population</u>

<u>Background</u>

Serum proteins possess various biological functions such as hormones, enzymes, antibodies, and some serve as valuable disease biomarkers. Major components of serum proteins are albumin (ALB) (approximately 60%), globulins (mainly as γ-globulins, approximately 30%), and fibrinogens. Total serum protein levels range from 6.5 to 8.5 g/dl and show significant inter-individual variation influenced by environmental and genetic factors (20% to 77%). GWAS reports recently demonstrated that serum levels of several proteins can be strongly influenced by common genetic variants.

<u>Methodology</u>

For the GWAS, 9,103 subjects derived from BBJ, and for the replication study, I used data from >1600 independent individuals. For immunoglobulin isotypes analyses, the data from ~1,600 additional individuals in BBJ was used. DNA of individuals with SLE (193 case) and IgA nephropathy (370 cases) were also obtained from BBJ.

<u>GWAS genotyping:</u> the Illumina HumanHap610-Quad BeadChip (Illumina, CA, USA). Quality control measures are same to the GWAS of POAG.

Replication-stage genotyping: generated using multiplex PCR- based Invader Assay.

Values of TP, ALB and NAP (mg/dl) were adjusted in linear regression models with age, gender, body mass index (BMI), smoking, drinking status, and affection status of the disease as covariates. The residuals were then normalized as Z scores. The associations of the SNPs with Z scores were evaluated in linear regression models assuming additive effects of allele dosages, using mach2qtl software. Immunoglobulin isotypes (IgG IgA, IgM, and IgE) used as common log-transformed values. Meta-analyses of the GWAS and the replication study were performed using the inverse-variance method assuming a fixed-effects model.

The significance level used was $5x10^{-8}$ in the GWAS stage. For the replication stage, I considered 0.05 as significant. Pearson's chi squire test was used under 3 genetic models (allelic, dominant and recessive) to compare genotype and allele frequencies between SLE and IgA nephropathy cases, and the healthy control cohort. P-values < 0.05 were considered significant. Serum APRIL levels were assayed using set of 193 serum samples stored in BBJ by ELISA. The detection level of the test was set as (0.8ng/ml). Samples were analyzed in duplicated.

Results and Discussion

Q-Q analysis for each trait is like following: λ_{GC} for TP, ALB and NAP were 1.04, 1.02 and 1.02, respectively. *TNFRSF13B* on chromosome 17p11.2 showed significant associations with both TP and NAP (rs4985726, $P = 2.8 \times 10^{-12}$ and 2.4×10^{-22}, respectively). In addition, rs3803800 and rs11552708 in coding regions of *TNFSF13* on chromosome 17p13.1 demonstrated significant associations with NAP ($P = 1.8 \times 10^{-12}$ and 7.0×10^{-9}, respectively). In addition to rs10007186 located near *ANXA3* (annexin A3) on chromosome 4q21.2 ($P = 3.3x10^{-9}$). For serum ALB, SNPs rs1260326 (in exon) and rs3817588 (in intron) in *GCKR* (glucokinase regulator) on 2p23.3 revealed significant associations.

Replication and mate analyses: Meta-analyses combining the GWAS and the replication study yielded stronger associations of SNPs rs4985726 in *TNFRSF13B*, rs3803800 and rs11552708 in the *TNFSF13* and rs1260326 in *GCKR* with serum ALB. Rs10007186 near *ANXA3* revealed a suggestive association in the replication study ($P = 0.065$), and meta-analyses indicated that the association was unlikely to be false positive ($P = 1.3 \times 10^{-9}$).

Immunoglobulins association: I found significant associations of rs4985726 in *TNFRSF13B* as well as rs3803800 and rs11552708 in *TNFSF13* with serum levels of IgG ($P < 0.0023$) and IgM ($P < 0.018$). For IgA, rs3803800 and rs11552708 in *TNFSF13* also revealed the significant association ($P < 0.013$), while rs4985726 in *TNFRSF13B* revealed no significant association ($P = 0.099$). Rs10007186 near *ANXA3* indicated significant association with IgA ($P = 0.036$), IgM ($P = 0.019$), and IgE ($P = 4.9 \times 10^{-4}$).

SLE and IgA nephropathy association: I detected association of allele G of rs11552708 in *TNFSF13* with SLE (P= 0.007, O.R=1.32, 95% C.I: 1.05-1.66), and rs4273077 in *TNFRSF13B* (P= 0.036, O.R=1.25, 95% C.I: 1.00-1.56). SNPs showed significant association with IgAN (P< 0.05).The SNP rs11552708 showed strongest association (P= 0.0012, O.R=1.34, 95 % C.I:1.12- 1.60). Individuals carrying both risk alleles of both SNPs showed 2.16 higher risks for SLE than those carrying non-risk alleles, and 7.09 higher risks for IgAN.

Conclusion

This present study identified genetic loci that influence the inter-individual variation in serum levels of TP, ALB, and NAP. The loci associated with NAP encompass genes encoding a TNF-receptor and its ligand, which are implicated in biological roles in the immune system, and their associations with immunoglobulin isotypes were demonstrated here.

Part III: Functional analysis of CADM1 in lung adenocarcinoma

Background

CADM1 (Cell Adhesion Molecule 1) encodes an immunoglobulin-superfamily adhesion molecule with three immunoglobulin-like loops. It is a tumor suppressor protein and found to be involved in cell-cell adhesion in calcium-independent manner as being recruited in the lateral membrane of cells, and then being involved in hemophilic or heterophilic t*rans*-interaction between its molecules. Lung cancer is a leading cause of death in many developed countries including Japan, and adenocarcinoma is the most histological subtype of the disease. In primary lung adenocarcinoma, CADM1 was found to be expressed in normal pulmonary epithelia, and lost or decreased in primary pulmonary adenocarcinoma. However, the mechanism by which CADM1 prevents cancer progression in lung adenocarcinoma remains a debated question.

Results and discussion:

I utilized the microarray data of lung adenocarcinoma cell lines to predict possible CADM1 interacting proteins. Depending on the expression patterns, I found gene X to have the strongest correlation with CADM1 expression (correlation coefficient > 0.6). Gene X involves in a cell-signaling pathway. Western blot analysis using lung adenocarcinoma cell lines supported the microarray data as both proteins (gene X and CADM1) indicate similar patterns of expression; first: 8 cell lines (about 67%) tend to show co-expression or loss of expression of both proteins, second: mesenchyme-like phenotypes cell lines indicate highest expression for both proteins. ICC indicated co-localization of both proteins mainly in the cell membrane region. IHC analysis using primary lung adenocarcinoma tumors revealed that CADM1 and gene X protein product are co-expressed in poorly differentiated tumors (solid and pleomorphic phenotypes). Gene X is significantly correlated with poorer prognosis and tumor progression.

Using expression microarray expression data in lung adenocarcinoma cell lines, gene X product was identified as a potential CADM1-interacting protein. Subsequent experiments indicated similar expression patterns of both proteins both in cell lines and human lung adenocarcinoma tissues, evidence of co-immunoprecepitation and hence interaction and co-localization in cell membrane region. Gene X pathway members involve adherence junction (AJ, mediated by E-cadherin), and control endothelial junctions and tight junctions (TJ). Gene X product functions by activating the effector proteins which perform this function. Several observations indicate CADM1 may be involved in gene X product signaling: both molecules are involved in cell-cell adhesion; both may be involved in recruitment of E-cadherin to cell-cell contact sites,

I hypothesize that in advanced lung adenocarcinoma, CADM1 will work as strong tumor inducer, possible by recruiting gene X product. This interaction will then promote invasion and cancer progression

Conclusion:

I have here demonstrated in this ongoing study that the adhesion molecule CADM1 is co-expressed and co-localize with gene X. This indicates CADM1 may have a role in molecule X signaling pathway. Since both molecules (CADM1 and gene X product) are involved in cell-cell adhesion, and both of them seem to be promote cancer progression in poorly-differentiated tumors.