博士論文

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

List of Contents

ABSTRACT

CHAPTER 1: Primary open angle glaucoma (POAG) is one of leading causes of adult blindness worldwide. To identify genetic variants associated with susceptibility to POAG, I conducted a genome-wide association study (GWAS) using a total of 3,196 cases and 13,811 controls. I found two associated loci: 9p21 and 14q23. For described SNPs, minor alleles are suspected to have a protective effect from the disease. *SIX6* gene in 14q23 was implicated to have a critical role in eye development, and *CDKN2A-CDKN2B* genes on chromosome 9p21 are known to be expressed in human ocular tissues, including the retina.

CHAPTER 2: Proteins are most abundant compounds in the serum and play several biological roles. I performed a GWAS on levels of serum total protein (TP), albumin (ALB), and non-albumin protein (NAP) using data from > 10,700 Japanese individuals. **Novel findings:** *TNFRSF13B* and *TNFSF13* encode a tumor necrosis factor (TNF) receptor and its ligand that is important for B-cell homeostasis and immunoglobulin production, and 4q21.2 near *ANXA3* indicated significant associations with TP, NAP, IgG, IgM, and IgA.

CHAPTER 3: CADM1 (Cell Adhesion Molecule 1) encodes an immunoglobulin-superfamily adhesion molecule involved in cell-cell adhesion in calcium-independent manner and well-known tumor suppressor. Microarray data in 41 adenocarcinoma cell lines indicated CADM1 is co-expressed with Rap1 guanine exchange factor RapGEF2. Further experiments indicate co-expression of both proteins in lung adenocarcinoma cell lines and in poorly-differentiated human adenocarcinoma tissues, co-localization at membrane areas with a direct evidence of interaction. The biological role of this interaction is still to be functioned.

List of figures

Figures for chapter 1:

Figures for chapter 2:

Figures of chapter 3:

List of tables

Tables for chapter 1:

Tables for chapter 2:

Tables of chapter 3:

ABBREVIATIONS

CCT	Central corneal thickness
CADM1	Cell adhesion molecule 1
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotides
EMT	Epithelial-mesenchymal transition
GWAS	Genome-wide association study
ICC	Immunocytochemistry
IgAN	IgA nephropathy
IHC	Immunohistochemistry
IOP	Intra-ocular pressure
IP	Immunoprecepitation
LD	Linkage disequilibrium
OR	Odds ratio
PCR	Polymerase chain reaction
QC	Quality control
POAG	Primary open-angle glaucoma
Q-Q plot	Quantile-Quantile plot
Rap1	Ras-proximate-1 or Ras-related protein 1
RapGEF2	Rap guanine nucleotide exchange factor 2
SE	Standard error
RGCs	Retinal ganglion cells

SLE	Systemic lupus erythematosus
SNP	Single nucleotide polymorphism
TMA	Tissue microarray
VCDR	Vertical cup-disc ratio
5’ and 3’ UTR	5’ and 3’ un translated region

Chapter 1: Genome-wide association study of primary open-angle glaucoma in the Japanese population

Section 1.1.1 Introduction

Glaucoma is one of leading causes of adult visual impairment and irreversible blindness worldwide **(1)**. It is a complex, chronic neurodegenerative optic neuropathy causing progressive degeneration of retinal ganglion cells (RGCs) and their axons along with the supporting glia cells and vasculature, finally resulting in a thinning of the neuro-retinal rim of the optic nerve and visual-fields defect **(2).**

Primary open angle glaucoma (POAG; MIM # 137760) is the most common form of the disease (**Fig. 1-1**). The term primary open-angle is due to the fact that the iridocorneal angle in the eye remains open in the disease (closed angle indicates closed-angle type). In addition, cupping of the optic-nerve head (or optic disc), with corresponding loss of visual field is another diagnostic criterion. The elevated intraocular pressure (IOP), as widely thought, is not a part of the clinical definition. That is because POAG can occur when intraocular pressure is normal (typically 10 to 21 mm Hg, and called normal tension glaucoma). However, elevated intraocular pressure is an important risk factor (when high, the disease called high tension glaucoma) **(3)**. In fact, currently all treatments of primary open-angle glaucoma are aimed at reducing intraocular pressure by medical or surgical methods. The IOP is produced by the effect of aqueous humor, fluid inside the eye which is essentially the same as blood plasma although with less protein. This fluid maintains the intraocular pressure and inflates the globe of the eye, provides nutrition, carries away waste products from metabolism, transport medium, and carries immunoglobulins.

The IOP is affected by physiological factors like: heart rate, respiration, exercise, fluid intake, systemic medication and topical drugs.

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 **(4)**. In summary, this was a cross-sectional epidemiological study carried out by the Japan Glaucoma Society in Tajimi city, which is located in center of Japan. Among 3,870 eligible people (age > 40 years) 3,021 (78.1%) participated in the study. Each individual was subjected to the standard tests for glaucoma diagnosis (**see Methods section**). Tajimi study is now used as the reference study for prevalence of glaucomas in Japan.

POAG is considered to have a multifactorial etiology and was reported that black race, older age, elevated intra ocular pressure, central corneal thickness (CCT), myopia, and systemic diseases such as diabetes and hypertension are factors increasing risk of the disease **(5)**. In addition, the family history of POAG is a well-known risk factor as first-degree relatives of patients have 3-9 folds higher risk than the general population **(6)**.

Considering the genetics, among more than 20 genetic loci that have been suggested by linkage analyses **(7)**, three genes encoding myocilin (*MYOC*) **(8)**, optineurin (*OPTN*) **(9)** and WD-repeat domain 36 (*WDR36*) **(10)** have been identified as causes of the disease although mutations in these genes were found to contribute to a small subset of the cases **(11)**. Recently, genome-wide association studies for POAG in the Japanese population as well as for those in individuals with European ancestry were reported **(12-15)**. Variants reported in two Japanese studies did not reach to the genome-wide significant level and have not been replicated in other populations. SNPs in caveolin 1 and 2 (*CAV1-CAV2)* on chromosome 7q31 found in Caucasians

are not polymorphic in the Japanese population. Furthermore, variants in *TMCO1* (transmembrane and coiled-coiled domains 1) on 1q24 and in *CDKN2B-AS* on 9p21 have not been studied in other populations. In addition, there are several studies which used the candidate gene approach to highlight the genetics of the disease **(16)**. These genes were selected based on the functional attraction to the disease; however, they remain unsuccessful because they couldn't be replicated in other populations and because the pathophysiology of the diseases remains not fully understood.

Section 1.1.2 Objective

The objective of this study is to identify the genetic variations contributing to the susceptibility of the POAG in the Japanese population, incorporating a total of 3,196 cases and 13,811 controls.

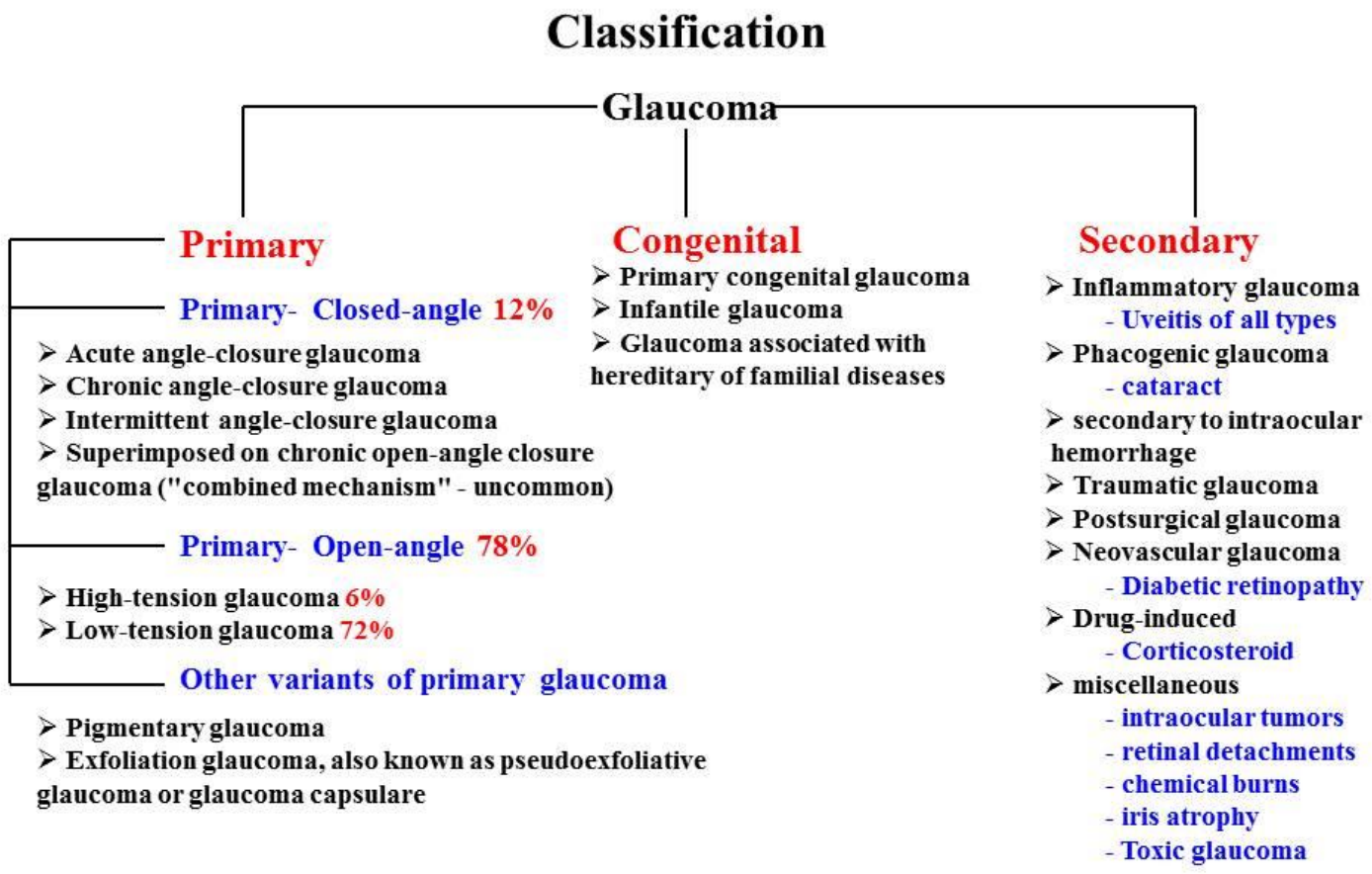


Figure 1-1: Classification of glaucoma.

Section 1.2 Subjects and Methods

1.2.1 Cases and Controls

Characteristics of study groups are shown in **Table 1-1**. All case and control subjects were collected in BioBank Japan which was supported as the Japan leading project for personalized medicine from the Ministry of Education, Culture, Science, Sports, and Technology **(17)**. For the GWAS, 1,394 subjects who were diagnosed as POAG were selected.

The diagnosis of POAG was judged by expert ophthalmologists in the 66 collaborating network hospitals according mainly to the glaucomatous optic neuropathy with corresponding glaucomatous visual field deficit caused by the disease assessed using different perimeters. The following examinations were also included for diagnosis: slit-lamp microscopy, ophthalmoscopic examination of the optic disc and retinal nerve fibers layer, measuring of the vertical cup-disc ratio (VCDR), tonometry for measuring the IOP, and gonioscopy for assessment of the iridocorneal angle which should necessarily to be open as inclusion criteria. All patients with normal tension or high tension glaucoma were included in this study as cases under the general term POAG. Individuals with secondary, close angle or congenital glaucoma were not included in this study. I obtained control data for GWAS from the genome scanning of other five cohorts in BioBank Japan (**Table 1-1**). All control subjects were not known to have glaucoma according to their clinical records. For a replication study, another cohort of 1,802 POAG cases was obtained also from BioBank Japan. Controls for the replication study were obtained from five different cohort sets as shown in **Table 1-1**.

1.2.2 Ethical approval

All participants provided a written informed consent for this study. The study was approved by the ethical committees at the Institute of Medical Science, the University of Tokyo, and the Center for Genomic Medicine, RIKEN, Yokohama Institute, Japan.

1.2.3 SNP genotyping

For the GWAS stage, DNA from POAG cases and disease- mix controls were genotyped using Illumina HumanOmniExpress BeadChip (Illumina, CA, USA). The BeadChip contains 733,202 SNPs, among which 712, 726 SNPs were located in autosomal chromosomes. Non-autosomal SNPs (20,476) and non-polymorphic loci in our samples (65,799 loci) were excluded from subsequent analysis. 602,216 SNPs in autosomal chromosomes were passed the quality control (Q.C) measures indicated as following: call rate ≥ 0.99 in both cases and controls, exact *P*-value of the test of the Hardy-Weinberg equilibrium ≥ 1.0×10^{-6} in controls, and minor allele frequency ≥ 0.01. Cluster plots of the top 100 SNPs showing the smallest *P*-values were checked by visual observation and 4 SNPs were excluded from further analysis because of ambiguous cluster plots. Genotype data was analyzed for screening of duplicate and closely related samples in the GWAS stage using identity-by-state (IBS) approach in PLINK 1.0.6. . To exclude individuals who do not belong to the Japanese ancestry, I applied EIGENSTRAT 3.0 to perform principle component analysis (PCA) using the case-control samples in this GWAS.

In replication analyses, I applied the multiplex PCR-based Invader Assay method **(18)** for genotyping the selected SNPs in the case samples. Typically, the primer pair for each SNP has a temperature between 55 to 60°C. I performed multiplex-PCR amplification using 20ng of genomic DNA with a total reaction volume is 50ul. The reaction mix was consisting of: 50pmol of each primer, 10 units of HotStart DNA polymerase, 10x PCR buffer, and dNTPs. The reaction

condition was one cycle at 94°C for 2 min for initial denaturation, then 35 cycles of: denaturation at 94°C for 15s, annealing at 60°C for 45s, and extension at 72°C for 2min.

I diluted the PCR product in x10 dilutions before distributed them into 384 well PCR plate. The PCR products were air-dried overnight before Invader assay. Briefly, the assay relies on the specific recognition and cleavage by flap-specific endonuclease of the three-dimensional structure formed when two overlapping oligonucleotides (labeled with FAM or VIC that corresponded to alleles), an invader oligonucleotide and signal oligonucleotide with a reporter arm hybridize to a target DNA sequence containing the polymorphic site. Genotypes were called by visual inspection, following application of Q.C measures of individuals call rates of > 98% of SNPs, and SNPs call rates > 99% of individuals. Each total reaction volume of 3µl per well with dried diluted PCR product, contained 0.5µl of signal buffer, 0.5µl of FRET probes, 0.5µl of structure-specific cleavage enzyme, 1µl of allele-specific probe mix. Samples were incubated at 95°C for 5min, and then at 63°C for 5min. Further incubation was required if the allele discrimination plot did not reveal distinct cluster of risk homozygote, heterozygote and non-risk homozygote after the reaction read by Applied Biosystem 7900 HT Fast Real-time PCR system. Control samples for replication study were genotyped using the Illumina HumanOmniExpress BeadChip, and the same Q.C filters were applied.

1.2.4 Statistical analysis

In the GWAS and replication analyses, I used the logistic regression test to assess the statistical significance of each SNP (age, gender, and the first two eigenvectors were incorporated as covariates) in addition to applying a genomic control method. I also applied the logistic regression model incorporating age and gender as covariates for the replication and

combined analyses. The significance levels used were 1.5 x 10^{-3} (0.05 / 33) for the replication stage and 8 x 10^{-8} (0.05 / 602,216) in the combined analysis following the Bonferroni correction for multiple testing. Meta-analysis of the two sets of data was performed using the inverse-variance method and heterogeneity between GWAS and replication cohorts were tested using Breslow- Day test. Odds ratios and 95% confidence intervals were calculated using the minor allele genotype as a reference.

Linkage disequilibrium (LD) analyses were performed using PLINK 1.0.6, Haploview software version 4.2 and SNAP database, meta-analyses were performed using Metal software, and regional association plots were generated using Locus Zoom (**see web resources**).

1.2.5 Imputation analyses

Imputation for prediction of un- genotyped SNPs in the case- control GWAS samples on chromosomes 2, 9 and 14 using MACH program version 1.0.16 (**see web resources**). The references for the analyses were the observed genotypes in the GWAS stage, and the data of JPT+CHB in the 1000 Genomes Project, release 2010 June. Dosage score for the imputation analysis were obtained using MACH2DAT **(19)**. Q.C. for imputation was done by exclusion of SNPs with estimated square correlation (r^2) between imputed genotypes and actual genotypes of < 0.5.

1.2.6 DNA re-sequencing

To identify possible causative variant(s) in *SIX6* gene, the coding exons, 5' UTR, 3' UTR and all splice sites of *SIX6* gene were sequenced in 191 cases together with their healthy controls. Primers are given in **Table 1-8**. Each fragment was amplified by PCR using specific primers set, and the PCR amplification products were purified on a Multiscreen PCR_{96} Plate (MSNU 03050,

Millipore). DNA products were then purified again using ethanol precipitation, re-suspended by 10μl Hi-Di Formamid and denatured at 95℃ for 5min to using GeneAmp 9708 (Applied Biosystems) before directly sequenced on ABI PRISM 3730x1 Genetic Analyzer (Applied Biosystems) with BigDye Terminators (Applied Biosystems) according to standard protocols. Sequence fragments were screened using 4.8 Sequencher software (Gene Codes Corporation, Ann Arbor, MI, USA).

1.2.7 eQTL analysis

The expression quantitative tract loci analysis (eQTL) was performed to test the association of identified SNPs and the expression of genes in the same genetic locus of the SNPs. I performed this analysis using Genevar database. Within this database, I focused on the data reported by Barbara E. Stranger et al. (**20**). Briefly, authors utilized lymphoblastoid cell lines from a total of 726 individuals from 8 global populations from the HapMap3 project and correlated gene expression levels with HapMap3 SNPs located in cis to the genes. I focused on Japanese data (N = 82), since this GWAS was performed using DNA samples from Japanese cohorts. The reported association was performed within 1 Mb of the SNP flanks. The database considered the significance association of (P < 0,001), using Spearman's rank correlation coefficient (rho).

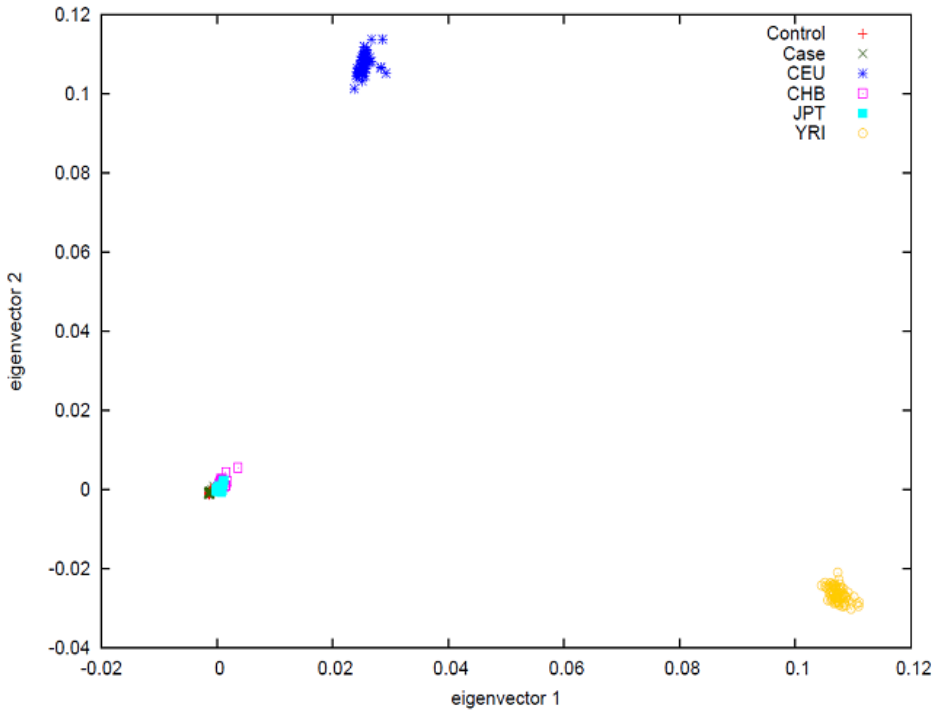
Section 1.3 Results

1.3.1 Genome-wide association study of POAG

To screen genetic variants contributing to the susceptibility of POAG in the Japanese population, I conducted a GWAS using BioBank Japan samples of 1,394 POAG cases and 6,599 control subjects, who genotyped with Illumina HumanOmniExpress BeadChip (**see Methods, Table 1-1**). Following applying stringent quality control (Q.C) filtering (**see Methods**), 602,216 SNPs were selected for further association analysis with POAG.

Principal component analysis (PCA) indicated that all cases and controls were of Asian ancestry (**Fig. 1-2 A**) and > 99% of the cases and controls were clustered within either of the two known major clusters of the Japanese population **(21)** (**Fig. 1-2 B**).

A



B

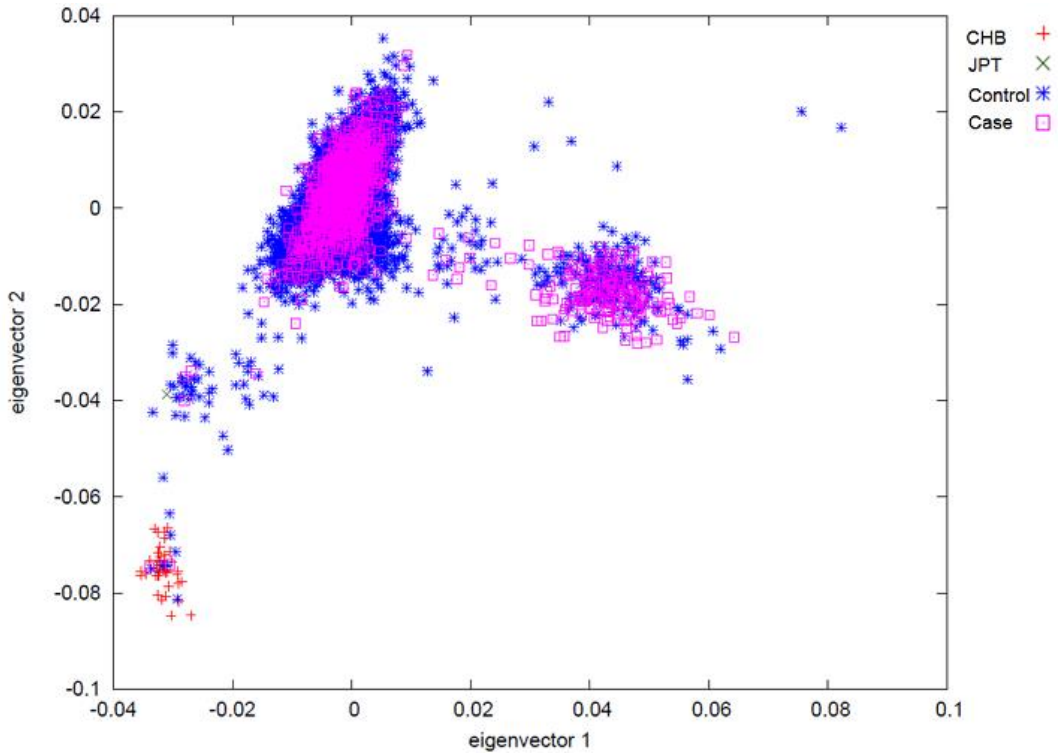


Figure 1-2: Principal component analysis plot of GWAS of POAG.

(A) All cases and controls who were selected for the GWAS together with the four population in the HapMap phase II database (Japanese: JPT; Chinese: CHB; Africans: YRI and European: CEU) were plotted based on the first two eigenvectors.

(B) All cases and controls who were selected for the GWAS together with the two populations of Asian ancestry in the HapMap phase II database (Japanese: JPT and Chinese: CHB) were plotted based on the first two eigenvectors. > 99% of cases and controls were clustered within either of the two major clusters of the Japanese population.

Since quantile-quantile (Q.Q) plot analysis using all the samples revealed the genomic inflation factor (λ_{GC}) of 1.30, I applied a logistic regression model incorporating age, gender, and the first two eigenvectors as covariates for the analysis, in addition to apply the genomic control method **(22)** (λ_{GC} =1.00; **Fig. 1-3**).

The association analysis indicated one SNP, rs7588567, on chromosome 2q21 was within the GWAS-significance level (P = 4.79 x 10^{-8}), in addition to another 57 SNPs to be possibly associated with POAG (P < 1 x 10^{-4}) as shown in **Fig. 1-4**.

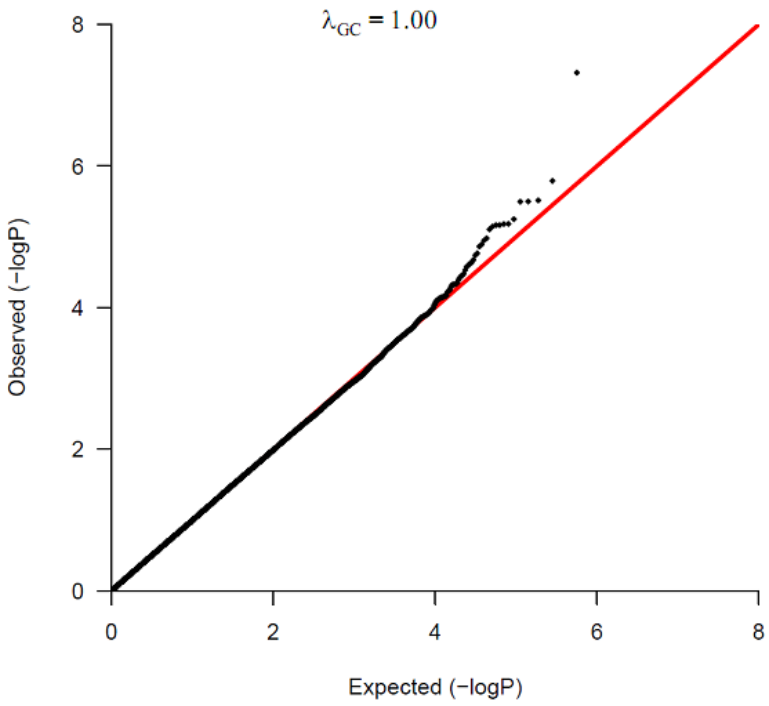


Figure 1-3: Q-Q plot of the GWAS of POAG.

Under the null hypothesis of no association at any locus, the points would be expected to follow the black line (y=x). The genomic inflation factor (λ_{GC}) for the analysis is shown above the graph. λ_{GC} = 1.00

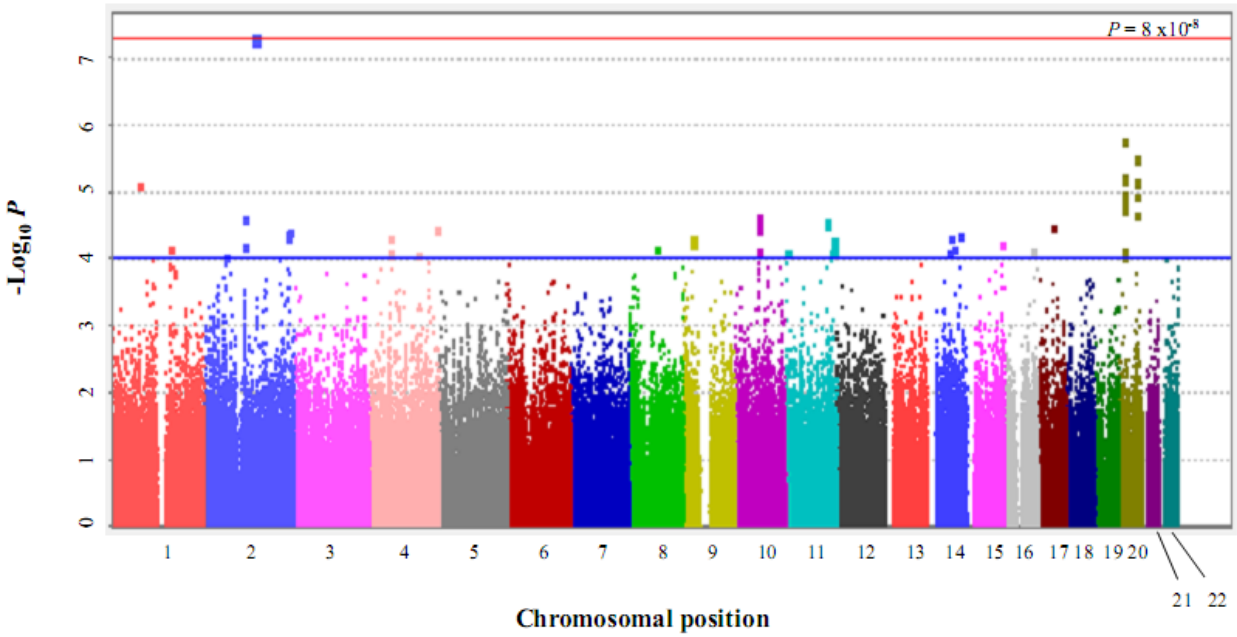


Figure 1-4: Manhattan plot for all the SNPs tested for association with POAG.

The horizontal axis shows the chromosomal positions. The vertical axis shows -log10 *P*-values from the test of association by logistic regression (age, gender, first two eigenvectors, and GC-adjusted). The red horizontal line indicates the GWAS-significance level ($P < 8 \times 10^{-8}$) and the blue line shows the threshold of SNPs selected for replication study, which corresponds to *P*-value of $1.0 \times 10^{-4.}$

1-3-2 Replication study

After examining the linkage-disequilibrium (LD) of these SNPs, I further selected 33 SNPs for replication studies (**Table 1-2**) using an additional set of independent 1,802 POAG cases and 7,212 controls collected from BioBank Japan (**Table 1-1**). Among 33 examined SNPs, 2 SNPs; rs1063192 in *CDKN2B* gene on chromosome 9p21, and rs10483727 on 14q23 were significantly replicated following the Bonferroni-correction of multiple testing (P-value < 0.0015), and both indicated stronger statistical association following combined analyses of the GWAS and the replication study (P_{meta} = 5.2 x 10^{-11}, OR = 0.75 and 9.49 x 10^{-8}, OR = 0.79, respectively) (**Fig. 1-5 and Table 1-3**).

No significant heterogeneity was observed following combining the GWAS and replication data for both SNPs (**Table 1-3**). For these two SNPs, the minor alleles are considered to be protective from the disease and risk alleles seem to have the additive effect in the development of the disease. The SNP rs7588567 on 2q21 revealed no association in the replication analysis (P = 0.07), and indicated a suggestive association following meta-analysis (P = 3.89 x 10^{-7}), although a significant evidence of heterogeneity was observed between the GWAS and replication sets for this SNP (**Table 1-3**).

Since the disease-mix controls in our GWAS included cases of intracranial aneurysm (IA), one of diseases being associated with 9p21 locus **(25)**, I performed the analysis of the SNP rs1063192 after exclusion of IA samples from the controls, and found stronger association of this SNP with the OPAG (P = 8.00 x 10^{-12}; **Table 1-4**), implying that my results are unlikely to be false-positive.

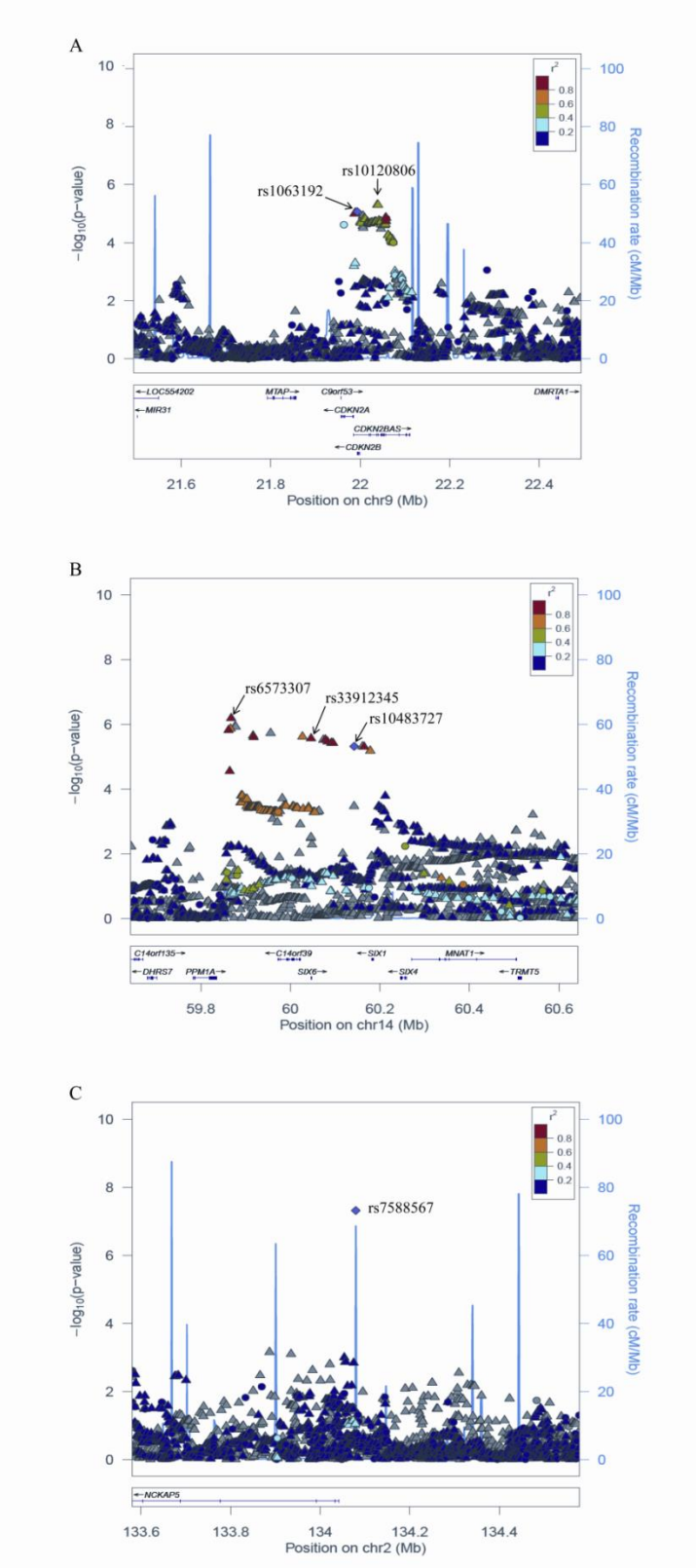


Figure 1-5: Regional plots for association of genotyped and imputed SNPs in the GWAS stage with POAG.

SNPs plotted with their -log10 P values of the logistic regression test (age, gender, first two eigenvectors, and GC-adjusted) against their physical chromosomal positions (NCBI Build 36). Genotyped SNPs are indicated by circles, while imputed SNPs are indicated as triangles. The color scheme indicated the linkage disequilibrium displayed as r^2 values between all SNPs and the top-ranked SNP in each plot. Top-ranked SNPs are shown as purple diamonds in the top of each chromosomal locus; (A) 9p21, (B) 14q23, and (C) 2q21.The blue lines represent the estimated recombination rates.

A recent report suggested imputation analysis using the 1000 Genomes Project resource is sufficient to capture most of SNPs with frequency >5%, and provides a relevant method for fine-mapping of the GWAS data **(26)**. Therefore; I performed imputation-based fine-mapping for the promising loci; 9p21, 14q23, and 2q21 using the data of the East-Asian populations in the 1000 Genomes Project as a reference (**Methods**). For 9p21 locus, a comparative evidence of association to rs1063192 was observed for rs10120806 in *CDKN2B-AS* (**Fig. 1-5 A**), confirming *CDKN2B-AS* is another possible candidate for the disease in 9p21 region as reported **(15)**. Several highly linked SNPs on 14q23 region revealed comparative associations with rs10483727, including the missense variant rs33912345 in *SIX6* gene and rs6573307 which is located within 32 kb distance from *PPM1A* (protein phosphatase, Mg2+/Mn2+ dependent, 1A) (**Fig. 1-5 B**).

1-3-3 Replication of previously-reported loci

I also examined the associations of SNPs in *TMCO1*, *MYOC*, *OPTN*, *WDR36*, and *CAV1-CAV2* genes that were previously reported as candidates for POAG. Although the reported SNPs in *CAV1-CAV2* in the Icelandic population (rs4236601 and rs1052990) **(14)** are not polymorphic in the Japanese population, other SNPs in this locus were replicated in a comparable OR (top SNP rs7795356; P = 0.0073, OR = 1.31) (**Table 1-6**). Rs4656461 (the reported variant in *TMCO1*) did not pass the Q.C criteria in this GWAS, and no proxy SNP was included. For other genes, no SNP revealed significant association with POAG (P > 0.05) as summarized in **Table 1-6**. In addition, using our GWAS set, the results of SNPs reported in the Japanese population by Nakano *et al*. **(12)** (rs7081455), and Meguro *et al*. (13) (rs735860), were not replicated (P = 0.16, 0.93, respectively) in association with POAG.

1-3-4 association with clinical parameters IOP and VCDR

I then tested the association of the SNPs rs1063192 and rs10483727 with the clinical parameters; VCDR and IOP (intraocular pressure), important predicting signs for POAG in the clinical practice. Data were available from 1,082 cases for VCDR analysis, and 2,434 cases for IOP (**see Methods**). For VCDR, I observed no association with neither a SNP (P = 0.057) as in **Table 1-7**. However, for IOP, I observed a strong association of a SNP rs10483727 ($P = 3.75 \times 10^{-6}$; **Table 1-7**).

1.3.5 Cumulative risk of associated loci

I estimated the cumulative risk effect of the two SNPs described above by counting the number of risk alleles of each SNP by assumption of the additive effect model and found that individuals with four risk alleles have 2.26 times higher risk for the disease than those with 0-1 risk alleles (**Fig. 1-6**).

1.3.6 Sequence analysis of *SIX6* gene

Nonetheless, as further in-depth functional analysis is needed to identify causative variant(s) in 14q23 region; I carried out sequence analysis for *SIX6* gene as being expressed in the eye and for its reported role in eye development (**37-39**). I selected 191 cases and their corresponding controls and then sequenced the two exons of *SIX6* gene, together with 3' and 5' UTR regions, and exon-intron boundaries to discover possible functional variants which could be a causative in this gene (**Table 1-8**).

The sequence analysis revealed seven variations within the gene and its regulatory sequences in the 3' or 5' UTR regions: three of them are reported and four are novel variations (**Table 1-9, Fig. 1-7**). Of the novel variations: one variation is a novel SNP at the 5' UTR and three are rare variations all in exon regions (**Fig. 1-7**). Considering the function, the SNP

rs33912345 is the most functionally attractive variation, not only because it is a missense variant, but also because it is located in codon 141 which is located inside the HOX domain (amino acids from 128 to 187). HOX domain is the functional domain of the protein and can binds to the DNA through helix-turn-helix structure, therefore, proteins harbor this domain are considered as transcription factors.

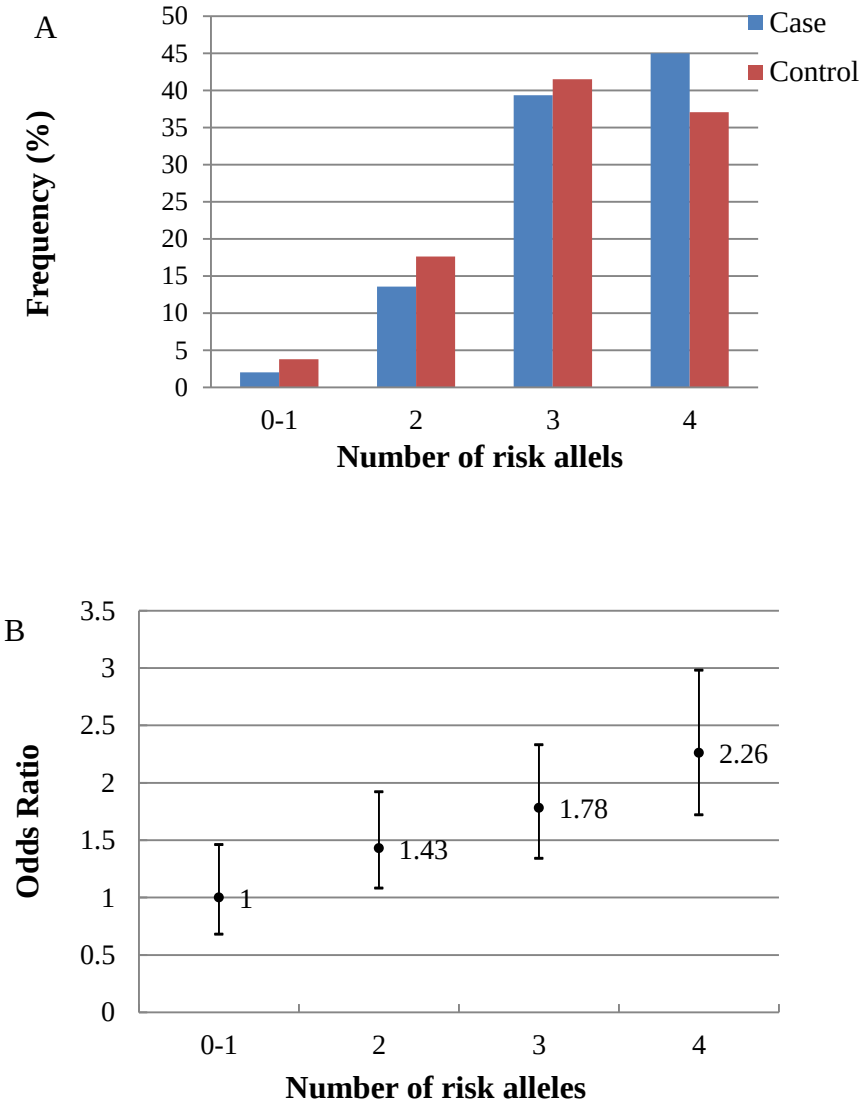


Figure 1-6: Cumulative effect of risk alleles of POAG of rs1063192 and rs10483727.

(A) The distribution of risk alleles among cases and controls. As the numbers of individuals having 0 or 1 risk alleles is very small, their combining numbers were used for the analysis. Cases are shown in blue bars, and controls in red bars. (B) Plot of the association between odds ratio (OR) and number of risk alleles. Every vertical bar represents 95% confidence intervals for the O.R.

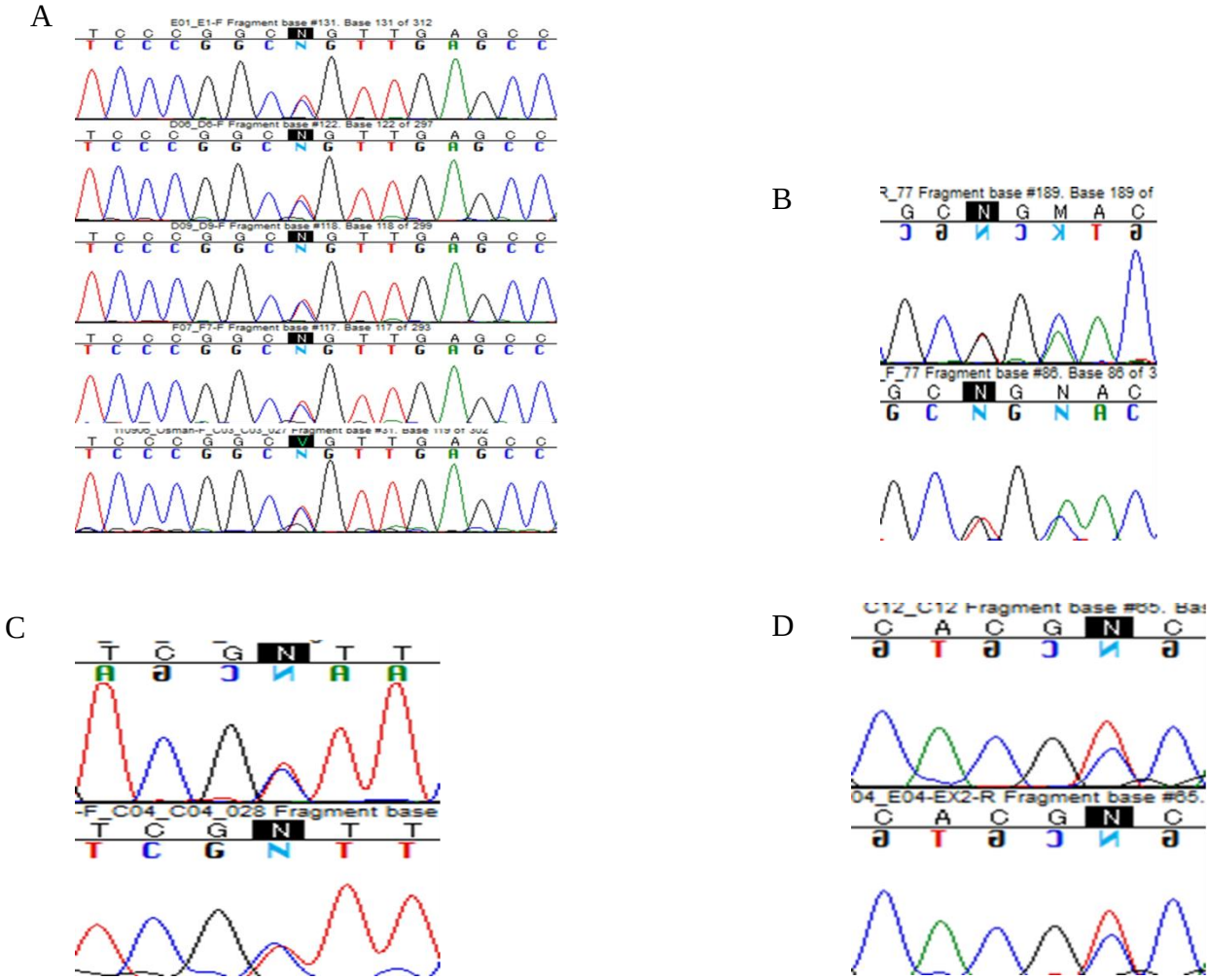


Figure 1-7: Novel variations identified by sequence analysis of *SIX6* gene.

(A) 5` UTR (t>c 31), frequency : 0.013
(B) Exon 1 (codon 140, CGG Arg →CTG Lue), frequency : 0.003
(C) Exon 1 (codon 32, CGC Arg →CGT Arg), frequency: 0.003
(D) Exon 2 (codon 213, CCA Pro → TCA Ser), frequency: 0.005

1.3.7 eQTL analysis of SNPs associated with POAG

To further investigate the functional effects of the loci associated with POAG, I used the data in Genevar eQTL (expression quantitative trait loci) associations of SNPs within each identified loci with genes in the locus

For the SNP rs1063192 in 9p21 locus, there are 25 probes of genes to be tested with this SNP; none of them indicted significant association (**Fig. 1-8 A**). The specific associations with rs1063192 with *CDKN2A* and *CDKN2B* are shown in figures **1-8 B and C** respectively. Similarly, rs10483727 in 14q23 locus did not indicate significance association with 17 probes of genes in this locus (**Fig. 1-9 A**), and with *SIX6* (**Fig. 1-9 B**) or *SIX1* (**Fig. 1-9 C**).

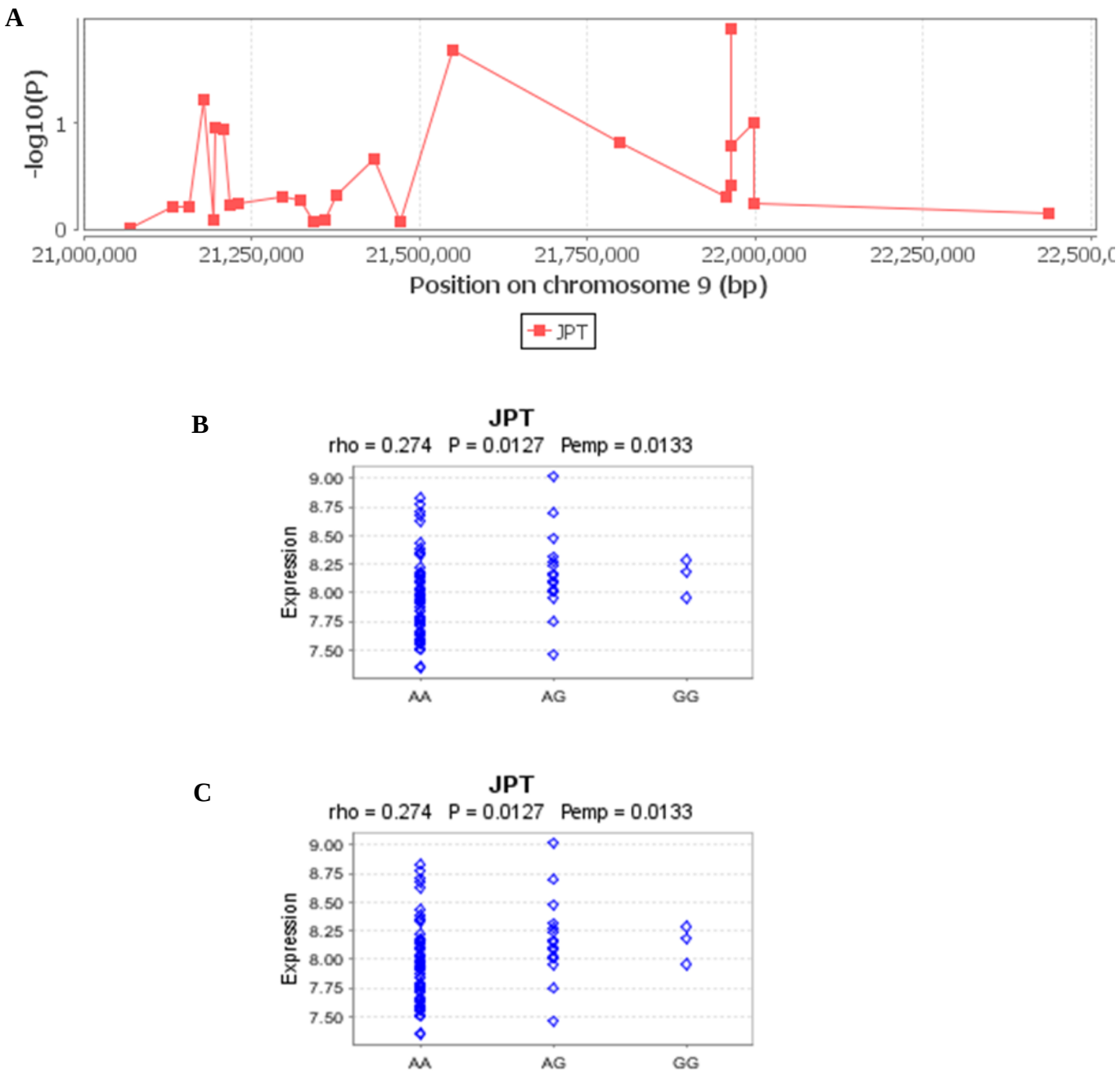


Figure 1-8: eQTL analysis of rs1063192 and genes in 9p21 locus.

(A) Association between the SNP rs1063192 and the expression of gene probes in 9p21 locus according to Genevar database. Each dot represents one probe. Significance level ($P < 0.001$).
(B) Association between the genotypes of rs1063192 and the expression of *CDKN2A*
(C) Association between the genotypes of rs1063192 and the expression of *CDKN2B*

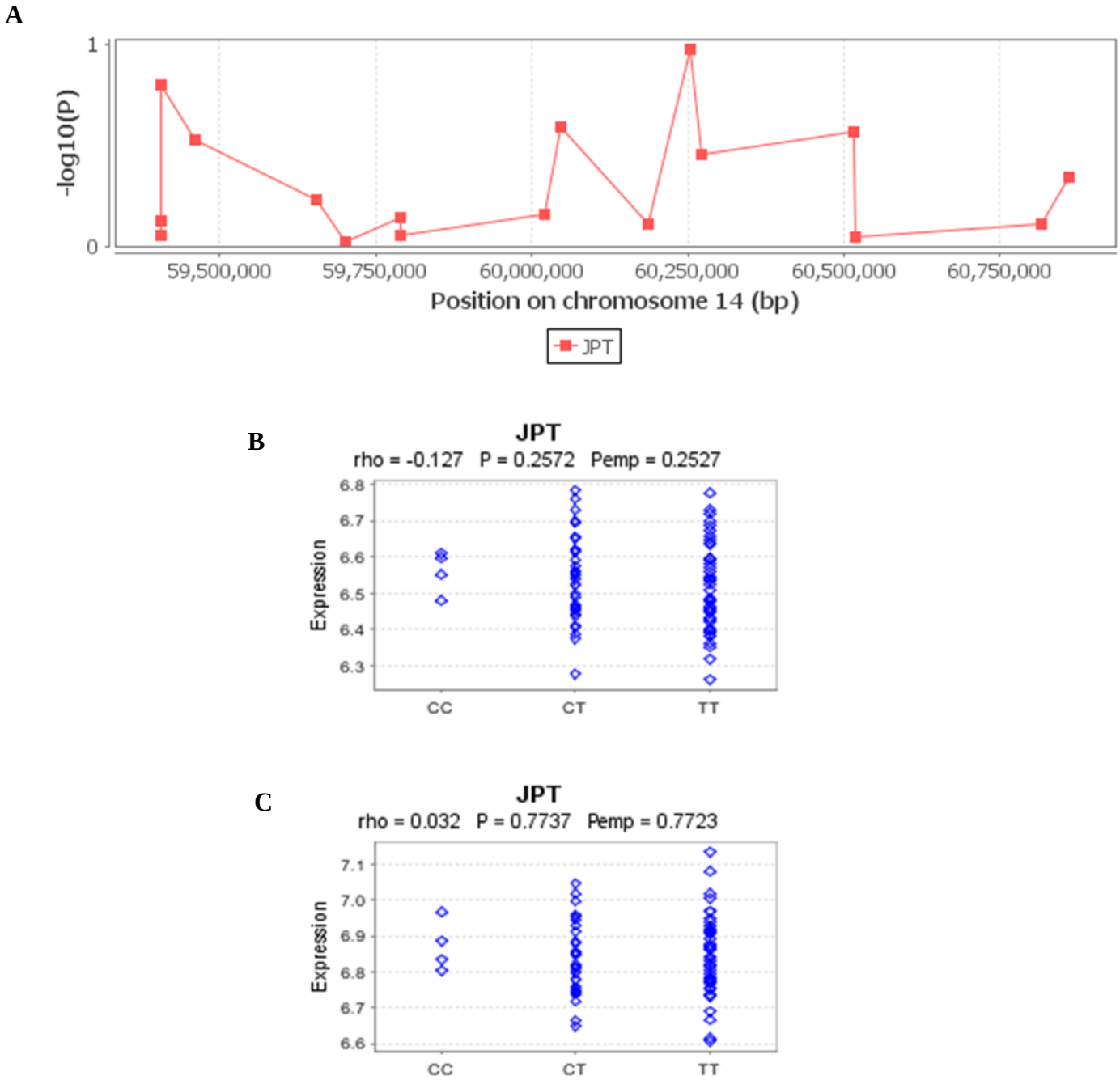


Figure 1-9: eQTL analysis of rs10483727 and genes in 14q23 locus.

(A) Association between the SNP rs10483727 and the expression of gene probes in 9p21 locus according to Genevar database. Each dot represents one probe. Significance level ($P < 0.001$).

(B) Association between the genotypes of rs10483727 and the expression of *SIX6*

(C) Association between the genotypes of rs10483727 and the expression of *SIX1*

Section 1.4 Discussion

I performed a GWAS and replication analyses for POAG in the Japanese population, and identified two susceptibility loci.

1.4.1 9p21 locus

For the most significantly associated locus on chromosome 9p21, a SNP rs1063192 is located within a gene encoding cyclin-dependent kinase inhibitor 2B (CDKN2B) which is also known as p15. Notably, an LD including this SNP contains three additional genes; *CDKN2A* (also known as *p16*), *p14/ARF* (encoded by alternate reading frame) as well as *CDKN2B-AS* (CDKN2B antisense RNA, and is also known as ANRIL) that is transcribed in the antisense direction of p14, p15 and p16, and emerged as a possible candidate for the disease using a 1000 Genomes imputation-based mapping in this study (**Fig. 1-5 A**). In confirmation of the association of this locus with POAG in Japanese, two studies by Takamoto M. et al. (**23**) and Nakano M. et al. (**24**) have reported this association. *CDKN2A* and *CDKN2B* are well-known tumor suppressor genes involved in the retinoblastoma (Rb) pathway, and ANRIL is a long non-coding RNA which is thought to regulate the transcription of p14/ p15/p16 genes. This region on chromosome 9p21 was implicated as a hotspot locus showing the association with various diseases including intracranial aneurysm (I.A) (**25, 27**), coronary artery disease (**28, 29**), type 2 diabetes (**30-32**), glioma (**33**), and endometriosis (**34**). The SNP rs1063192 was reported as being associated with glioma, but the C allele that was indicated to be a risk allele for glioma is considered to be protective from POAG in this study. The association of rs1063192 in 9p21 became stronger after exclusion of IA cases from the controls (**Table 1-4**); hence, our results represent most likely a true association of this region with POAG. A recent GWAS report for the optic disc parameters

indicated significant association of rs1063192 with the vertical cup-disc ratio (VCDR) and POAG in the Caucasian population (**35**), which was supported by further reports (**36, 37**). *CDKN2B* and *CDKN2A* genes are expressed in human ocular tissues including the RGCs, and were found to be upregulated in response to elevation of intraocular pressure (**15**). Although the SNP rs1063192 is located in a region that was predicted to be a binding site of a regulatory micro RNA (miRNA) in the 3`UTR of CDKN2B in the SNPinfo database (**38**), further functional analysis for this region is required to identify the causative variant(s). However, our data at least provide further evidence for the association of 9p21 region with POAG in Asians as well. There was no association with rs1063192 and the expression of genes in this locus including *CDKN2A* and *CDKN2B* (**Fig. 1-8**)

1.4.2 14q23 locus

For the second locus on 14q23 region, a SNP rs10483727 (combined P = 9.49 x 10-8, OR = 0.79) is located between two homologous genes belonging to the SIX/Sine oculis homeobox family; *SIX1* and *SIX6* (also known as Optx2), and is located about 94 kb downstream from SIX6 gene. The comparative mapping analysis between the marker SNP and imputed SNPs from the 1000 genomes project revealed several SNPs with comparable associations in or around SIX6 (**Fig. 1-5 B and Table 1-5**), including the SNP rs33912345 (r2 = 0.92 with the SNP rs10483727), that is considered to be a possible functional candidate. Rs33912345 causes an amino-acid substitution of codon 141 of the SIX6 protein which causes the loss of positive charge (a histidine residue with a positive charge is replaced with an asparagine residue with a negative charge) although the significance of this substitution on the protein function needs to be clarified. The SIX6 mRNA was reported to be expressed in retina, optic nerve and brain (**39, 40**). Interestingly, mutations in this gene have been found in patients with defects in eye development

(**41**). Furthermore, the SNP rs10483727 on 14q23 reported to be associated with VCDR (**35**) and POAG in the Caucasian population (**36, 37**). The mapping analysis indicated that the association in this locus is extended to rs6573307 (r2 = 0.92 with rs10483727) which is located within 32 kb distance from *PPM1A* (protein phosphatase, Mg2+/Mn2+ dependent, 1A) (**Fig. 1-5 B and Table 1-5**). The protein encoded by this gene is a member of the PP2C family of Ser/Thr protein phosphatases. PP2C family members are known to be negative regulators of cell stress response pathways, and PPM1A is known to terminate TGF-β signaling through Smad phosphatase (**42**). As most of the known genes involved in the control of glaucoma-related traits belong to developmental and TGF-β signaling pathways (**36, 43**), PPM1A constitutes another candidate in this region. My data confirmed 14q23 locus as a common susceptibility locus for POAG in different ethnic populations. There was no association with rs10483727 and the expression of genes in this locus including *SIX6* and *SIX1* (**Fig. 1-9**).

1.4.3 2q21.2 locus

For the third possibly-associated locus on 2q21.2, a SNP rs7588567 (combined $P = 3.89 \times 10^{-7}$, OR = 0.85) is located about 37 kb upstream to *NCKAP5* (Nck-associated protein 5) gene (**Figure 1-5 C**). The expression of NCKAP5 protein was detected in fetal and adult brain, leukocytes and fetal fibroblasts **(44)**. It was shown to be interacting with an SH3 (Src homology region 3) domain of the adaptor protein Nck **(44)**. Chromosome 2q21.2 locus was identified to be associated with optic disc area (DA) **(43, 45)**. The marker rs7588567 is located about 3.3 Mb form the suggestive linkage marker (rs113906) described by Axenovich T. et al. **(43)**. The association of 2q21.2 with POAG needs further validation, owing to the nature of the single-marker association presented here, which might be due to the high recombination rates in this locus (**Figure 1-5 C**).

1.4.4 previously-reported loci

I also replicated the association of *CAV1-CAV2* on chromosome 7q31 with POAG in the Japanese population. In my GWAS, rs7795356 (P = 0.007, OR = 1.31) (**Table 1-6**) indicated a comparable OR with the original report in the Icelandic population **(14)** and a USA cohort **(46)**, although this SNP exhibited no linkage with the marker rs42336601 described in the Icelandic population by Thorleifsson G. et al. **(14)**. Another SNP, rs959173 (P = 0.027, OR = 1.25; **Table 1-6**) was reported to be a part of a haplotype that was significantly associated with POAG **(46)**. Although I could not genotype rs7795356 in our replication cohort due to difficulty to make a suitable Invader assay, these results at least provide an evidence for *CAV1-CAV2* as a candidate for POAG in Asians.

1.4.5 Association with VCDR and IOP

For VCDR, I observed no association as being reported before (**Table 1-7)**. It is important to note that all individuals tested are cases of POAG; therefore, it is expected to show higher average values for VCDR. However, interestingly, for the SNP rs1063192 individuals homozygous for the protective alleles revealed lower values of VCDR although I had no data for the stages of patients. For IOP, I observed a strong association of a SNP rs10483727 with IOP ($P = 3.75 \times 10^{-6}$) (**Table 1-7**). Individuals homozygous for the protective allele for this SNP revealed higher values of IOP in an additive-effect model. Although the analysis was not adjusted by the effect of IOP-reducing therapy, I assume that those individuals who were homozygous for the minor allele might have a tendency to have less severe disease than those with two copies of the risk alleles. In general, this analysis is restricted by the fact that the

BioBank Japan project collects samples from several different medical institutions; each has its own settings and criteria.

1.4.6 Resequence of *SIX6* gene

Among the seven variations identified in *SIX6* gene, four are novel. (**Table 1-9, Fig. 1-7**). However, three of them are rare variations. Functional-wise; the SNP in the 5'UTR is predicted to be located in a transcription factor binding site, the variant in codons 140 and 213 are missense variations although they are very rare (0.1%) when compared to the prevalence of the disease. Further functional analysis might be needed to examine the effect of each of those variations in the protein structure or the function of SIX6 protein.

1.4.7 Functional insights of *SIX6*

Recent reports revealed functional insights for loci linked to POAG in this study. Adriana I. Iglesias et al. reported a knocked-down *SIX6* model in zebra fish (**47**). They found that knock-down of *SIX6* in zebra fish results in small eye phenotype, and the co-injection of *SIX6* mRNA partially rescued this small eye phenotype. In addition, *SIX6* knock-down resulted in underdeveloped eyes (optic nerve is less thick 21 μm in wild-type *SIX6* vs. 12 μm in knock-down) and also less developed lenses. And lastly, *SIX6* found to be involved in proliferation and differentiation of retinal cells. For 9p21 locus,

Section 1.5 Conclusions and limitations

I here 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 loci as common susceptibility loci for POAG among different ethnic populations. I showed that minor alleles of each of the representative SNPs are likely to be protective for the disease. Further in-depth functional analyses of these loci are needed to identify causative variants and clarify their contribution to the disease. My data should give more insight for better understanding of the pathogenesis of POAG.

There are number of limitations in this study, which are mainly related to the nature of the Bio-Bank Japan (BBJ) project:

1- The control group represents disease-mix groups in BBJ. All control subjects didn't undergo ophthalmic examination to rule out the diagnosis of POAG. Therefore, I can't rule out the fact that some controls may have the disease.

2- Hence the BBJ is collaborating project which gathers patients from several hospitals; the diagnosis of POAG and other diseases is performed by different doctors using different settings which may affect the "case definition".

3- Although BBJ requests filling of a standard questionnaire from each hospital, I noticed that an important data was missed from some patients (e.g. values of IOP and VCDR), a major limitations in this study.

Section 1.6 Acknowledgments

I want to acknowledge Dr. Atsushi Takahashi and the members of Laboratory of Laboratory for Statistical Analysis, Center for Genomic Medicine, Institute of Physical and Chemical Research (Center for Genomic Medicine, RIKEN), Kanagawa, Japan, for their assistance in the statistical analyses in this study.

Also, I want to thank Dr. Michiaki Kubo and the members of laboratory for Genotyping Development (Center for Genomic Medicine, RIKEN), Kanagawa, Japan, for their role in the genotyping work of the samples.

Section 1.7 Web resources

BioBank Japan Project, http://biobankjp.org/

R statistical environment, http://www.r-project.org/

EIGENSTRAT software, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm

The International HapMap Project, http://www.hapmap.org/

MACH software, http://www.sph.umich.edu/csg/abecasis/MaCH/index.html

Metal software, http://www.sph.umich.edu/csg/abecasis/Metal/index.html

PLIKN software, http://pngu.mgh.harvard.edu/~purcell/plink/

Locus Zoom, http://csg.sph.umich.edu/locuszoom/

SNAP, http://www.broadinstitute.org/mpg/snap/ldsearch.php

Section 1.8 References

1- Resnikoff, S., Pascolini, D., Etya'ale, D., Kocur, I., Pararajasegaram, R., Pokharel, G.P. and Mariotti, S.P. Global data on visual impairment in the year 2002. ***Bull World Health Organ***, 82, 844-851. (2004)

2- Kwon, Y.H., Fingert, J.H., Kuehn, M.H. and Alward, W.L. Primary open-angle glaucoma. ***New England Journal of Medicine***, 360, 1113-1124. (2009)

3- Bahrami, H. Causal inference in primary open angle glaucoma: specific discussion on intraocular pressure. ***Ophthalmic Epidemiology***, 13, 283-289. (2006)

4- Iwase, A., Suzuki, Y., Araie, M., Yamamoto, T., Abe, H., Shirato, S., Kuwayama, Y., Mishima, H.K., Shimizu, H., Tomita, G. et al. The prevalence of primary open-angle glaucoma in Japanese: the Tajimi Study. ***Ophthalmology***, 111, 1641-1648. (2004)

5- Allingham, R., Liu, Y. and Rhee, D. The genetics of primary open-angle glaucoma: a review. ***Experimental Eye Research***, 88, 837-844. (2009)

6- Tielsch, J.M., Katz, J., Sommer, A., Quigley, H.A. and Javitt, J.C. Family history and risk of primary open angle glaucoma. The Baltimore Eye Survey. ***Archives of Ophthalmology***, 112, 69-73. (1994)

7- Fan, B.J., Wang, D.Y., Lam, D.S. and Pang, C.P. Gene mapping for primary open angle glaucoma. ***Clinical Biochemistry***, 39, 249-258. (2006)

8- Stone, E.M., Fingert, J.H., Alward, W.L., Nguyen, T.D., Polansky, J.R., Sunden, S.L., Nishimura, D., Clark, A.F., Nystuen, A., Nichols, B.E. et al. Identification of a gene that causes primary open angle glaucoma. ***Science***, 275, 668-670. (1997)

9- Rezaie, T., Child, A., Hitchings, R., Brice, G., Miller, L., Coca-Prados, M., Héon, E., Krupin, T., Ritch, R., Kreutzer, D. et al. Adult-onset primary open-angle glaucoma caused by mutations in optineurin. ***Science***, 295, 1077-1079. (2002)

10- Monemi, S., Spaeth, G., DaSilva, A., Popinchalk, S., Ilitchev, E., Liebmann, J., Ritch, R., Héon, E., Crick, R.P., Child, A. et al. Identification of a novel adult-onset primary open-angle glaucoma (POAG) gene on 5q22.1. ***Human Molecular Genetics***, 14, 725-733. (2005)

11- Hewitt, A.W., Craig, J.E. and Mackey, D.A. Complex genetics of complex traits: the case of primary open-angle glaucoma. ***Clinical and Experimental Ophthalmology***, 34, 472-484. (2006)

12- Nakano, M., Ikeda, Y., Taniguchi, T., Yagi, T., Fuwa, M., Omi, N., Tokuda, Y., Tanaka, M., Yoshii, K., Kageyama, M. et al. Three susceptible loci associated with primary open-angle glaucoma identified by genome-wide association study in a Japanese population. ***Proceeding of National Academy of Science USA***, 106, 12838-12842. (2009)

13- Meguro, A., Inoko, H., Ota, M., Mizuki, N., Bahram, S. and Society, W.C.f.t.N.T.G.G.S.G.o.J.G. Genome-wide association study of normal tension glaucoma: common variants in SRBD1 and ELOVL5 contribute to disease susceptibility. ***Ophthalmology***, 117, 1331-1338.e1335. (2010)

14- Thorleifsson, G., Walters, G.B., Hewitt, A.W., Masson, G., Helgason, A., DeWan, A., Sigurdsson, A., Jonasdottir, A., Gudjonsson, S.A., Magnusson, K.P. et al. Common variants near CAV1 and CAV2 are associated with primary open-angle glaucoma. ***Nature Genetics***, 42, 906-909. (2010)

15- Burdon, K.P., Macgregor, S., Hewitt, A.W., Sharma, S., Chidlow, G., Mills, R.A., Danoy, P., Casson, R., Viswanathan, A.C., Liu, J.Z. et al. Genome-wide association study identifies susceptibility loci for open angle glaucoma at TMCO1 and CDKN2B-AS1. ***Nature Genetics***. (2011)

16- Fuse, N. Genetic bases for glaucoma. ***Tohoku Journal of Experimental Medicine***, 221, 1-10. (2010)

17- Nakamura, Y. The BioBank Japan Project. ***Clinical Advances in Hematology and Oncology***, 5, 696-697. (2007)

18- Ohnishi, Y., Tanaka, T., Ozaki, K., Yamada, R., Suzuki, H. and Nakamura, Y. A high-throughput SNP typing system for genome-wide association studies. ***Journal of Human Genetics***, 46, 471-477. (2001)

19- Li, Y., Willer, C., Sanna, S. and Abecasis, G. Genotype imputation. ***Annual Reviews of Genomics and Human Genetics***, 10, 387-406. (2009)

20- Stranger BE, Montgomery SB, Dimas AS, Parts L, Stegle O, Ingle CE, Sekowska M, Smith GD, Evans D, Gutierrez-Arcelus M, Price A, Raj T, Nisbett J, Nica AC, Beazley C, Durbin R, Deloukas P, Dermitzakis ET. Patterns of cis regulatory variation in diverse human populations. ***PLos Genetics***. 8 (4):e1002639. (2012)

21- Yamaguchi-Kabata, Y., Nakazono, K., Takahashi, A., Saito, S., Hosono, N., Kubo, M., Nakamura, Y. and Kamatani, N. Japanese population structure, based on SNP genotypes from 7003 individuals compared to other ethnic groups: effects on population-based association studies. ***Americal Journal of Human Genetics***, 83, 445-456. (2008)

22- Devlin, B. and Roeder, K. Genomic control for association studies. ***Biometrics***, 55, 997-1004. (1999)

23- Bilguvar, K., Yasuno, K., Niemelä, M., Ruigrok, Y.M., von Und Zu Fraunberg, M., van Duijn, C.M., van den Berg, L.H., Mane, S., Mason, C.E., Choi, M. et al. Susceptibility loci for intracranial aneurysm in European and Japanese populations. ***Nature Genetics***, 40, 1472-1477. (2008)

Takamoto M, Kaburaki T, Mabuchi A, Araie M, Amano S, Aihara M, Tomidokoro A, Iwase A, Mabuchi F, Kashiwagi K, Shirato S, Yasuda N, Kawashima H, Nakajima F, Numaga J, Kawamura Y, Sasaki T, Tokunaga K. Common variants on chromosome 9p21 are associated with normal tension glaucoma. ***PLoS ONE***. 7(7):e40107.(2012)

24- Nakano M, Ikeda Y, Tokuda Y, Fuwa M, Omi N, Ueno M, Imai K, Adachi H, Kageyama M, Mori K, Kinoshita S, Tashiro K. Common variants in CDKN2B-AS1 associated with optic-nerve vulnerability of glaucoma identified by genome-wide association studies in Japanese. ***PLoS One***. 7(3):e33389. (2012)

25-

26- Shea, J., Agarwala, V., Philippakis, A.A., Maguire, J., Banks, E., Depristo, M., Thomson, B., Guiducci, C., Onofrio, R.C., Kathiresan, S. et al. Comparing strategies to fine-map the association of common SNPs at chromosome 9p21 with type 2 diabetes and myocardial infarction. ***Nature Genetics***, 43, 801-805. (2011)

27- Helgadottir, A., Thorleifsson, G., Magnusson, K.P., Grétarsdottir, S., Steinthorsdottir, V., Manolescu, A., Jones, G.T., Rinkel, G.J., Blankensteijn, J.D., Ronkainen, A. et al. The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. ***Nature Genetics***, 40, 217-224. (2008)

28- Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A., Masson, G. et al. A common variant on chromosome 9p21 affects the risk of myocardial infarction. ***Science***, 316, 1491-1493. (2007)

29- McPherson, R., Pertsemlidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D.R., Hinds, D.A., Pennacchio, L.A., Tybjaerg-Hansen, A., Folsom, A.R. et al. A common allele on chromosome 9 associated with coronary heart disease. ***Science***, 316, 1488-1491. (2007)

30- Zeggini, E., Weedon, M.N., Lindgren, C.M., Frayling, T.M., Elliott, K.S., Lango, H., Timpson, N.J., Perry, J.R., Rayner, N.W., Freathy, R.M. et al. Replication of genome-wide association signals in UK samples reveals risk loci for type 2 diabetes. ***Science***, 316, 1336-1341. (2007)

31- Saxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N., Daly, M.J. et al. Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. ***Science***, 316, 1331-1336. (2007)

32- Scott, L.J., Mohlke, K.L., Bonnycastle, L.L., Willer, C.J., Li, Y., Duren, W.L., Erdos, M.R., Stringham, H.M., Chines, P.S., Jackson, A.U. et al. A genome-wide association study of type 2 diabetes in Finns detects multiple susceptibility variants. ***Science***, 316, 1341-1345. (2007)

33- Shete, S., Hosking, F.J., Robertson, L.B., Dobbins, S.E., Sanson, M., Malmer, B., Simon, M., Marie, Y., Boisselier, B., Delattre, J.Y. et al. Genome-wide association study identifies five susceptibility loci for glioma. ***Nature Genetics***, 41, 899-904. (2009)

34- Uno, S., Zembutsu, H., Hirasawa, A., Takahashi, A., Kubo, M., Akahane, T., Aoki, D., Kamatani, N., Hirata, K. and Nakamura, Y. A genome-wide association study identifies genetic variants in the CDKN2BAS locus associated with endometriosis in Japanese. ***Nature Genetics***, 42, 707-710. (2010)

35- Ramdas, W.D., van Koolwijk, L.M., Ikram, M.K., Jansonius, N.M., de Jong, P.T., Bergen, A.A., Isaacs, A., Amin, N., Aulchenko, Y.S., Wolfs, R.C. et al. A genome-wide association study of optic disc parameters. ***PLoS Genetics***, 6, e1000978. (2010)

36- Ramdas, W.D., van Koolwijk, L.M., Lemij, H.G., Pasutto, F., Cree, A.J., Thorleifsson, G., Janssen, S.F., Jacoline, T.B., Amin, N., Rivadeneira, F. et al. Common genetic variants associated with open-angle glaucoma. ***Human Molecular Genetics***, 20, 2464-2471. (2011)

37- Fan, B.J., Wang, D.Y., Pasquale, L.R., Haines, J.L. and Wiggs, J.L. Genetic variants associated with optic nerve vertical cup-to-disc ratio are risk factors for primary open angle glaucoma in a US Caucasian population. ***Investigative Ophthalmology and Visusal Science***, 52, 1788-1792. (2011)

38- Xu, Z. and Taylor, J.A. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. ***Nucleic Acids Research***, 37, W600-605. (2009)

39- Jean, D., Bernier, G. and Gruss, P. Six6 (Optx2) is a novel murine Six3-related homeobox gene that demarcates the presumptive pituitary/hypothalamic axis and the ventral optic stalk. ***Mechanisms of Development***, 84, 31-40. (1999)

40- Toy, J., Yang, J.M., Leppert, G.S. and Sundin, O.H. The optx2 homeobox gene is expressed in early precursors of the eye and activates retina-specific genes. ***Proceedings of National Academy of Science USA***, 95, 10643-10648. (1998)

41- Gallardo, M.E., Lopez-Rios, J., Fernaud-Espinosa, I., Granadino, B., Sanz, R., Ramos, C., Ayuso, C., Seller, M.J., Brunner, H.G., Bovolenta, P. et al. Genomic cloning and characterization of the human homeobox gene SIX6 reveals a cluster of SIX genes in chromosome 14 and associates SIX6 hemizygosity with bilateral anophthalmia and pituitary anomalies. ***Genomics***, 61, 82-91. (1999)

42- Lin, X., Duan, X., Liang, Y.Y., Su, Y., Wrighton, K.H., Long, J., Hu, M., Davis, C.M., Wang, J., Brunicardi, F.C. et al. PPM1A functions as a Smad phosphatase to terminate TGFbeta signaling. ***Cell***, 125, 915-928. (2006)

43- Axenovich, T., Zorkoltseva, I., Belonogova, N., van Koolwijk, L., Borodin, P., Kirichenko, A., Babenko, V., Ramdas, W., Amin, N., Despriet, D. et al. Linkage and association analyses of glaucoma related traits in a large pedigree from a Dutch genetically isolated population. ***Journal of Medical Genetics***, 48, 802-809. (2011)

44- Matuoka, K., Miki, H., Takahashi, K. and Takenawa, T. A novel ligand for an SH3 domain of the adaptor protein Nck bears an SH2 domain and nuclear signaling motifs. ***Biochemical and Biophysical Research Communications***, 239, 488-492. (1997)

45- Macgregor, S., Hewitt, A.W., Hysi, P.G., Ruddle, J.B., Medland, S.E., Henders, A.K., Gordon, S.D., Andrew, T., McEvoy, B., Sanfilippo, P.G. et al. Genome-wide association identifies ATOH7 as a major gene determining human optic disc size. ***Human Molecular Genetics***, 19, 2716-2724. (2010)

46- Wiggs, J.L., Kang, J.H., Yaspan, B.L., Mirel, D.B., Laurie, C., Crenshaw, A., Brodeur, W., Gogarten, S., Olson, L.M., Abdrabou, W. et al. Common variants near CAV1 and CAV2 are associated with primary open-angle glaucoma in Caucasians from the USA. ***Human Molecular Genetics***, 20, 4707-4713. (2011)

47- Iglesias AI, Springelkamp H, Linde HV, Severijnen LA, Amin N, Oostra B, Kockx CE, van den Hout MC, van Ijcken WF, Hofman A, Uitterlinden AG, Verdijk RM, Klaver CC, Willemsen R, van Duijn CM. Exome sequencing and functional analyses suggest that SIX6 is a gene involved in an altered proliferation-differentiation balance early in life and optic nerve degeneration at old age. ***Human Molecular Genetics***, 20, 4707-4713. (2013)

Table 1-1: Characteristics of study cohorts of the GWAS of POAG

Stage	Genotyping method	Phenotype	Phenotype N (%) [a]	Age (Mean ± S.D)	Female (%)
GWAS					
Case	Illumina HumanOmniExpress	POAG	1,394	60.3 ± 9.20	50.9
Control	Illumina HumanOmniExpress		6,599	57.9 ± 14.0	40.1
		Healthy controls	1,929 (29.2)		
		Intracranial aneurysm	1,383 (21.0)		
		COPD [b]	1,260 (19.1)		
		Esophageal Cancer	1,116 (16.9)		
		Uterine Cancer	911 (13.8)		
Replication					
Case	Invader assay	POAG	1,802	72.0 ± 9.59	49.8
Control	Illumina HumanOmniExpress		7,212	44.9 ± 18.3	47.1
		Basedow's disease	1,960 (27.2)		
		Epilepsy	1,849 (25.6)		
		Atopic dermatitis	1,469 (20.4)		
		Urolithisasis	1,101 (15.3)		
		Nephrotic syndrome	831 (11.5)		

[a] The number and percentage of each phenotype in the disease-mix controls . [b] COPD: Chronic Obstructive Pulmonary Disease.

Table 1-2: Results of SNPs selected for replication study ($P < 1 \times 10^{-4}$).

	Allele		Chr.		Cases				Control							
CHR	1/2	SNP	Position *	Stage	11	12	22	MAF	11	12	22	MAF	***P*** [a]	**OR**	**95% C.I**	P_*het* [b]
2	A/G	rs7588567	134,079,502	GWAS	176	656	558	0.363	1,170	3,135	2,264	0.417	4.79×10^{-8}	0.78	0.72-0.85	
				Rep.	269	880	652	0.394	1,226	3,412	2,503	0.411	0.07	0.92	0.83-1.00	
				Combined	445	1,536	1,210	0.38	2,396	6,547	4,767	0.414	3.89×10^{-7}	0.85	0.80-0.91	0.0062
20	G/A	rs4813974	11,190,433	GWAS	399	681	312	0.531	1,559	3,231	1,803	0.482	1.62×10^{-6}	1.23	1.13-1.34	
				Rep.	412	912	477	0.482	1,683	3,590	1,935	0.483	0.48	1.03	0.94-1.14	
				Combined	811	1,593	789	0.503	3,242	6,821	3,738	0.482	1.41×10^{-4}	1.11	1.05-1.18	0.000432
20	A/G	rs17446462	43,940,067	GWAS	30	334	1,029	0.141	81	1,277	5,235	0.109	3.05×10^{-6}	1.34	1.19-1.53	
				Rep.	32	375	1,395	0.122	102	1,502	5,582	0.119	0.48	1.05	0.91-1.22	
				Combined	62	709	2,424	0.13	183	2,779	10,817	0.114	2.00×10^{-4}	1.21	1.09-1.33	0.0014
20	C/T	rs6066093	44,926,671	GWAS	38	414	942	0.176	277	2,129	4,189	0.203	3.19×10^{-6}	0.76	0.69-0.85	
				Rep.	81	582	1,137	0.207	331	2,281	4,595	0.204	0.71	0.98	0.87-1.10	
				Combined	119	996	2,079	0.193	608	4,410	8,784	0.204	5.33×10^{-4}	0.89	0.83-0.96	
20	A/G	rs6040463	11,172,778	GWAS	361	685	347	0.495	1,423	3,258	1,915	0.463	6.60×10^{-6}	1.22	1.12-1.34	
				Rep.	391	869	538	0.459	1,504	3,614	2,091	0.459	0.95	1	0.91-1.10	
				Combined	752	1,554	885	0.479	2,927	6,872	4,006	0.461	1.73×10^{-3}	1.09	1.02-1.16	0.0024
20	T/C	rs6514063	43,866,424	GWAS	29	329	1,036	0.139	71	1,254	5,273	0.106	6.85×10^{-6}	1.34	1.18-1.52	

				Rep.	28	355	1,418	0.114	95	1,440	5,674	0.113	0.47	1.06	0.91-1.22	
				Combined	57	684	2,454	0.125	166	2,694	10,947	0.11	3.1×10^{-4}	1.21	1.11-1.34	
20	T/C	rs1333301	44,927,833	GWAS	17	274	1,103	0.11	123	1,512	4,961	0.133	7.13×10^{-6}	0.73	0.64-0.84	
				Rep.	43	419	1,339	0.14	161	1,630	5,420	0.135	0.96	1	0.87-1.14	
				Combined	60	693	2,442	0.127	284	3,142	10,381	0.134	1.83×10^{-3}	0.9	0.82-0.98	0.0027
1	T/C	rs875727	71,146,325	GWAS	143	581	670	0.311	863	2,963	2,773	0.355	7.84×10^{-6}	0.81	0.74-0.89	
				Rep.	231	819	751	0.356	916	3,279	3,017	0.354	0.66	0.98	0.89-1.08	
				Combined	374	1,400	1,421	0.336	1,779	,6242	5,790	0.355	7.15×19^{-4}	0.9	0.84-0.96	0.00054
20	C/A	rs6032531	43,845,974	GWAS	234	709	451	0.422	985	3,101	2,512	0.384	2.12×10^{-5}	1.21	1.11-1.32	
				Rep.	293	857	651	0.401	1,180	3,321	2,708	0.394	0.7	1.02	0.93-1.12	
				Combined	527	1,566	1,102	0.41	2,165	6,422	5,220	0.389	1.38×10^{-3}	1.12	1.05-1.19	0.023
10	C/T	rs11813512	61,006,298	GWAS	22	240	1,130	0.102	45	947	5,597	0.079	2.32×10^{-5}	1.36	1.19-1.57	
				Rep.	22	270	1,509	0.087	38	1,108	6,062	0.082	0.08	1.16	0.98-1.37	
				Combined	44	510	2,639	0.094	83	2,055	11,659	0.08	2.96×10^{-5}	1.19	1.07-1.33	0.023
2	A/C	rs1478516	108,867,498	GWAS	92	524	778	0.254	268	2,099	4,230	0.2	2.39×10^{-5}	1.24	1.13-1.37	
				Rep.	84	579	1,138	0.207	317	2,410	4,482	0.211	0.27	0.93	0.83-1.05	
				Combined	176	1,103	1,916	0.228	585	4,509	8,712	0.206	2.11×10^{-4}	1.13	1.05-1.22	6.31 x 10-7
11	T/G	rs5005886	110,201,457	GWAS	24	256	1,114	0.109	46	1,024	5,528	0.085	2.54×10^{-5}	1.35	1.18-1.55	
				Rep.	14	300	1,484	0.091	77	1,219	5,916	0.095	0.4	0.93	0.80-1.09	
				Combined	38	556	2,598	0.099	123	2,243	11,444	0.09	4.58×10^{-4}	1.15	1.04-1.27	0.00047

17	A/G	rs3760379	40,349,381	GWAS	280	672	432	0.445	1,053	3,160	2,374	0.4	3.33×10^{-5}	1.2	1.10-1.31	
				Rep.	284	866	651	0.398	1,151	3,425	2,622	0.398	0.3	1.05	0.96-1.16	
				Combined	564	1,538	1,083	0.419	2,204	6,585	4,996	0.399	3.14×10^{-4}	1.11	1.04-1.18	0.0011
4	A/G	rs11132079	182,707,185	GWAS	128	543	722	0.287	707	2,871	3,021	0.325	3.59×10^{-5}	0.82	0.75-0.90	
				Rep.	183	768	845	0.316	732	3,115	3,364	0.318	0.5	0.97	0.88-1.07	
				Combined	311	1,311	1,567	0.303	1,439	5,986	6,385	0.321	8.81×10^{-4}	0.89	0.83-0.95	0.0052
2	C/T	rs9288645	229,923,176	GWAS	198	633	563	0.369	1,087	3,114	2,397	0.401	3.79×10^{-5}	0.83	0.76-0.91	
				Rep.	285	846	669	0.393	1,133	3,445	2,632	0.396	0.27	1.05	0.96-1.16	
				Combined	483	1,479	1,232	0.383	2,220	6,559	5,029	0.398	2.89×10^{-4}	0.94	0.89-1.00	0.0034
2	C/A	rs7558186	224,404,511	GWAS	279	704	409	0.453	1,698	3,242	1,653	0.503	4.08×10^{-5}	0.84	0.77-0.91	
				Rep.	388	915	496	0.47	1,801	3,476	1,930	0.491	0.41	0.96	0.88-1.06	
				Combined	667	1,619	905	0.463	3,499	6,718	3,583	0.497	6.5×10^{-4}	0.88	0.82-0.93	0.0038
14	G/A	rs2372922	85,074,724	GWAS	51	430	913	0.191	181	1,786	4,632	0.163	4.49×10^{-5}	1.26	1.29-1.4	
				Rep.	52	479	1,268	0.162	230	2,050	4,931	0.174	0.87	0.99	0.87-1.12	
				Combined	103	909	2,181	0.175	411	3,836	9,563	0.169	3.53×10^{-3}	1.09	1.00-1.18	0.00015
9	A/G	rs1350996	22,977,182	GWAS	10	208	1,175	0.082	26	748	5,824	0.061	4.60×10^{-5}	1.4	1.19-1.64	
				Rep.	6	215	1,577	0.063	37	805	6,370	0.061	0.34	1.1	0.91-1.33	
				Combined	16	423	2,752	0.071	63	1,553	12,194	0.061	4.78×10^{-4}	1.21	1.07-1.36	
14	C/T	rs10483727	60,142,628	GWAS	51	450	893	0.198	369	2,404	3,825	0.238	4.68×10^{-5}	0.8	0.72-0.89	
				Rep.	79	566	1,153	0.2	352	2,560	4,297	0.226	0.0005	0.82	0.73-0.92	

				Combined	130	1,016	2,046	0.2	721	4,964	8,122	0.232	9.49×10^{-8}	0.79	0.73-0.85	0.21
4	C/T	rs10517449	59,361,782	GWAS	55	477	862	0.211	228	1,931	4,429	0.181	4.73×10^{-5}	1.25	1.12-1.38	
				Rep.	60	537	1,194	0.183	240	2,237	4,721	0.189	0.19	0.92	0.82-1.04	
				Combined	115	1,014	2,056	0.195	468	4,168	9,150	0.185	1.82×10^{-4}	1.08	1.00-1.17	0.0017
11	C/T	rs12795702	127,661,524	GWAS	46	438	909	0.19	341	2,337	3,919	0.229	5.14×10^{-5}	0.8	0.72-0.89	
				Rep.	94	642	1,063	0.231	374	2,428	4,409	0.22	0.23	1.07	0.96-1.20	
				Combined	140	1,080	1,972	0.213	715	4,765	8,328	0.224	2.56×10^{-4}	0.92	0.85-0.98	1.97 x 10-5
9	C/T	rs1063192	21,993,366	GWAS	37	396	960	0.169	276	2,159	4,156	0.206	5.68×10^{-5}	0.79	0.71-0.86	
				Rep.	50	480	1,270	0.161	346	2,388	4,473	0.214	1.73×10^{-7}	0.73	0.64-0.82	
				Combined	87	876	2,230	0.164	622	4,547	8,629	0.21	5.2×10^{-11}	0.75	0.70-0.82	0.16
15	G/A	rs1378593	92,564,676	GWAS	6	192	1,196	0.073	19	686	5,894	0.055	5.69×10^{-5}	1.42	1.20-1.67	
				Rep.	12	203	1,584	0.063	18	798	6,394	0.058	0.29	1.11	0.91-1.35	
				Combined	18	395	2,780	0.067	37	1,484	12,288	0.056	4.1×10^{-4}	1.23	1.08-1.39	0.057
1	G/A	rs984553	157,866,038	GWAS	15	229	1,149	0.093	28	782	5,788	0.064	6.84×10^{-5}	1.37	1.18-1.60	
				Rep.	12	250	1,534	0.076	43	997	6,172	0.075	0.75	1.03	0.87-1.22	
				Combined	27	479	2,683	0.084	71	1,779	11,960	0.07	3.0×10^{-3}	1.25	1.12-1.40	0.00011
14	A/G	rs12589622	67,060,077	GWAS	6	192	1,194	0.073	15	690	5,891	0.055	6.86×10^{-5}	1.42	1.20-1.67	
				Rep.	4	198	1,599	0.057	27	790	6,394	0.059	0.88	0.99	0.81-1.20	
				Combined	10	390	2,793	0.064	42	1,480	12,285	0.057	4.55×10^{-3}	1.17	1.03-1.32	0.0031
8	C/T	rs6983293	70,400,478	GWAS	85	509	798	0.244	536	2,701	3,357	0.286	6.98×10^{-5}	0.82	0.74-0.90	

				Rep.	151	654	993	0.266	601	2,874	3,734	0.283	0.25	0.94	0.85-1.04	
				Combined	236	1,163	1,791	0.256	1,137	5,575	7,091	0.284	3.66×10^{-4}	0.86	0.81-0.92	0.038
16	G/A	rs7501148	71,875,058	GWAS	5	124	1,261	0.048	6	422	6,164	0.033	7.10×10^{-5}	1.52	1.24-1.87	
				Rep.	1	129	1,669	0.036	8	510	6,685	0.037	0.55	0.92	0.72-1.19	
				Combined	6	253	2,930	0.042	14	932	12,849	0.035	1.55×10^{-3}	1.2	1.03-1.40	0.0047
10	A/G	rs1930465	59,437,004	GWAS	3	112	1,277	0.042	6	376	6,207	0.029	7.12×10^{-5}	1.56	1.26-1.94	
				Rep.	5	115	1,675	0.035	9	427	6,773	0.031	0.71	1.05	0.81-1.36	
				Combined	8	227	2,952	0.038	15	803	12,980	0.03	2.71×10^{-3}	1.27	1.09-1.51	0.089
11	C/T	rs1706879	1,966,470	GWAS	227	660	506	0.4	1,275	3,324	1,998	0.445	7.53×10^{-5}	0.84	0.77-0.91	
				Rep.	351	891	557	0.443	1,385	3,561	2,261	0.439	0.93	1	0.91-1.09	
				Combined	578	1,551	1,063	0.424	2,660	6,885	4,259	0.442	5.47×10^{-3}	0.91	0.85-0.97	0.00041
11	T/C	rs745517	120,597,261	GWAS	286	669	439	0.445	1,652	3,256	1,691	0.497	7.80×10^{-5}	0.84	0.77-0.92	
				Rep.	425	918	455	0.492	1,662	3,600	1,948	0.48	0.3	1.05	0.96-1.15	
				Combined	711	1,587	894	0.471	3,314	6,856	3,639	0.488	5.37×10^{-4}	0.91	0.86-0.97	5.65 x 10-6
14	C/A	rs6573104	56,266,008	GWAS	47	435	912	0.19	183	1,759	4,655	0.161	7.96×10^{-5}	1.25	1.12-1.39	
				Rep.	51	530	1,217	0.176	196	1,985	5,031	0.165	0.13	1.1	0.97-1.25	
				Combined	98	965	2,129	0.182	379	3,744	9,686	0.163	1.37×10^{-4}	1.21	1.12-1.31	0.097
4	C/T	rs4975300	129,379,766	GWAS	188	629	577	0.36	1,076	3,175	2,347	0.404	8.17×10^{-5}	0.84	0.77-0.91	
				Rep.	284	861	653	0.397	1,180	3,435	2,594	0.402	0.27	0.95	0.86-1.04	
				Combined	472	14,90	1,230	0.381	2,256	6,610	4,941	0.403	4.48×10^{-4}	0.89	0.84-0.95	0.0044

2	A/G	rs843653	54,322,787	GWAS	126	601	667	0.306	802	3,043	2,753	0.352	8.94×10^{-5}	0.83	0.76-0.91	
				Rep.	213	814	771	0.345	865	3,254	3,093	0.346	0.8	0.99	0.90-1.09	
				Combined	339	1,415	1,438	0.328	1,667	6,297	5,846	0.349	4.05×10^{-3}	0.9	0.84-0.96	0.00054

[a] *P*-value of the logistic regression test . For the GWAS: age, gender, first two eigenvectors, and GC-adjusted; for the replication: age and gender adjusted; for meta-analysis: obtained using inverse-variance method.

[b] *P*-value of heterogeneity of Breslow-Day test.

[c] Relative location of the SNP from the nearest gene is shown. The minus sign means upstream to the gene.

Odds ratios and 95% confidence intervals (C.I) were calculated using the minor genotype as a reference.

* Chromosomal position as in NCBI Build 36.

Table 1-3: Summary results of SNPs associated with primary open angle glaucoma in GWAS and replication samples

SNP ID		Locus		MAF				
A1 / A2	Chr : position	Candidate	Stage (N_{case} / $N_{control}$) [a]	Case	Control	P [a]	OR (95% CI)	P_{het} [b]
rs1063192	9 : 21,993,367	9p21.3	GWAS (1,393 / 6,591)	0.169	0.206	5.68×10^{-5}	0.79 (0.71-0.86)	
(C/T)		*CDKN2B*	Replication (1,800 / 7,207)	0.161	0.214	1.73×10^{-7}	0.73 (0.64-0.82)	
			Meta-analysis	0.164	0.210	5.20×10^{-11}	0.75 (0.70-0.82)	0.16
rs10483727	14 : 60,142,628	14q23.1	GWAS (1,394 / 6,598)	0.198	0.238	4.68×10^{-5}	0.80 (0.72-0.89)	
(C/T)		*SIX1-SIX6*	Replication (1,798/ 7,209)	0.200	0.226	5.00×10^{-4}	0.82 (0.73-0.92)	
			Meta-analysis	0.200	0.232	9.49×10^{-8}	0.79 (0.73-0.85)	0.21
rs7588567	2 : 134,079,502	2q21.2	GWAS (1,390 / 6,569)	0.363	0.417	4.79×10^{-8}	0.78 (0.72-0.85)	
(A/G)		*NCKAP5*	Replication (1,801 / 7,141)	0.394	0.411	0.07	0.92 (0.83-1.00)	
			Meta-analysis	0.380	0.414	3.89×10^{-7}	0.85 (0.80-0.91)	0.0062

[a] *P*-values in the GWAS stage were age, gender, first two eigenvectors, and GC-adjusted; in the replication stage were age and gender adjusted, and the inverse-variance method was used for meta-analyses.

[b] *P*-value of heterogeneity of Breslow-Day test

Table 1-4: Summary results of the SNP rs1063192 association with POAG in GWAS and replication samples following exclusion of intracranial aneurysm (IA) samples from the GWAS controls.

	Chr.	Locus		Case			Control			MAF		Risk			
SNP ID	**Position**	**Gene**	**Stage**	**CC**	**CT**	**TT**	**CC**	**CT**	**TT**	**Case**	**Cntl**	**allele**	***P*** [a]	**OR (95% CI)**	***P*** $_{het}$
rs1063192	21,993,367	9p21	GWAS	37	396	960	219	1,735	3,256	0.169	0.209	T	1.03×10^{-5}	**0.77** (0.69-0.87)	
(C/T)		*CDKN2B*	Rep.	50	480	1,270	346	2,388	4,473	0.161	0.214		1.73×10^{-7}	**0.73** (0.64-0.82)	
			Meta-analysis	87	876	2,230	565	4,123	7,729	0.164	0.212		8.0×10^{-12}	**0.75** (0.70-0.82)	0.25

[a] *P*-value of the logistic regression test. For GWAS: age, gender, the first two eigenvectors, and GC-adjusted. For replication stage: adjusted by age and gender. Meta-analysis was performed using inverse-variance method.

P $_{het}$: *P*-value of heterogeneity of Breslow-Day test model.

Table 1-5: List of full SNPs showed suggestive association with POAG in this study

Locus	SNP ID	Chr. position	A1	A2	*P*	OR [a]	SE
9p21	rs10120806	22,047,945	C	T	5.00E-06	1.32	0.061
	rs10811645	22,049,656	A	G	5.02E-06	1.32	0.061
	rs1063192	22,003,367	A	G	8.59E-06	1.28	0.055
	rs2811713	21,999,328	G	A	9.27E-06	1.28	0.056
	rs61271866	21,997,015	T	A	1.01E-05	1.28	0.056
	rs1101329	22,015,997	C	T	1.20E-05	1.32	0.064
	rs10965223	22,067,004	A	G	1.39E-05	1.28	0.057
	rs10965224	22,067,276	A	T	1.40E-05	1.28	0.057
	rs10811648	22,067,542	T	C	1.41E-05	1.28	0.057
	rs518394	22,019,673	G	C	1.41E-05	1.32	0.064
	rs10811651	22,067,830	A	G	1.42E-05	1.28	0.057
	rs523096	22,019,129	A	G	1.43E-05	1.32	0.064
	rs4977756	22,068,652	A	G	1.44E-05	1.28	0.057
	rs1333039	22,065,657	C	G	1.44E-05	1.28	0.056
	rs597816	22,021,172	T	C	1.56E-05	1.32	0.064
	rs944801	22,051,670	C	G	1.73E-05	1.31	0.063

9p21	rs10811649	22,067,554	T	C	1.82E-05	1.27	0.057
	rs6475604	22,052,734	C	T	1.88E-05	1.31	0.063
	rs7866783	22,056,359	G	A	1.88E-05	1.31	0.063
	rs7030641	22,054,040	T	C	1.88E-05	1.31	0.063
	rs944800	22,050,898	G	A	1.89E-05	1.31	0.063
	rs1360589	22,045,317	T	C	1.90E-05	1.31	0.063
	rs1412829	22,043,926	A	G	1.91E-05	1.31	0.063
	rs615552	22,026,077	T	C	1.91E-05	1.31	0.063
	rs1412830	22,043,612	C	T	1.92E-05	1.31	0.063
	rs2069418	22,009,698	C	G	1.96E-05	1.31	0.064
	rs1333037	22,040,765	T	C	1.96E-05	1.31	0.063
	rs613312	22,026,594	G	A	1.98E-05	1.31	0.063
	rs543830	22,026,639	A	T	2.00E-05	1.31	0.063
	rs1008878	22,036,112	T	G	2.04E-05	1.31	0.063
	rs1556515	22,036,367	T	C	2.04E-05	1.31	0.063
	rs599452	22,027,402	G	A	2.06E-05	1.31	0.063
	rs2157719	22,033,366	T	C	2.09E-05	1.31	0.063
	rs573687	22,011,642	G	A	2.11E-05	1.31	0.064

9p21	rs634537	22,032,152	T	G	2.18E-05	1.31	0.063
	rs7865618	22,031,005	A	G	2.20E-05	1.31	0.063
	rs62560774	22,028,406	C	A	2.20E-05	1.31	0.063
	rs564398	22,029,547	T	C	2.22E-05	1.31	0.063
	rs679038	22,029,080	G	A	2.23E-05	1.31	0.063
	rs4977755	22,066,363	A	T	2.32E-05	1.31	0.064
	rs7853090	22,056,295	C	T	2.37E-05	1.31	0.064
	rs10757268	22,059,905	C	T	2.44E-05	1.31	0.063
	rs3731239	21,974,218	A	G	2.44E-05	1.33	0.068
	rs2095144	22,060,136	G	A	2.45E-05	1.31	0.063
	rs2383205	22,060,935	G	A	2.45E-05	1.31	0.064
	rs1537378	22,061,614	G	A	2.49E-05	1.31	0.064
	rs8181047	22,064,465	G	A	2.51E-05	1.31	0.064
	rs8181050	22,064,391	A	G	2.51E-05	1.31	0.064
	rs1101330	22,015,465	C	A	2.68E-05	1.31	0.065
	chr9:20714495	20,714,495	G	A	2.91E-05	0.20	0.391
	chr9:20712384	20,712,384	C	A	2.92E-05	0.20	0.391
	rs62556546	22,016,510	C	G	3.11E-05	1.30	0.064

9p21	rs7039467	22,056,213	G	A	3.32E-05	1.30	0.064
	rs17696405	22,990,235	G	A	3.95E-05	0.70	0.085
	rs1350996	22,987,182	G	A	4.04E-05	0.73	0.078
	rs12349792	22,985,934	A	T	4.46E-05	0.74	0.075
14q23	rs6573307	60,798,009	G	T	6.49E-07	1.31	0.055
	rs12897140	61,961,410	G	C	1.10E-06	1.23	0.043
	rs6573308	60,806,976	T	C	1.15E-06	1.30	0.053
	rs4902071	61,963,561	C	T	1.16E-06	1.23	0.043
	rs4436712	60,808,002	T	G	1.19E-06	1.30	0.053
	rs2180833	60,796,648	C	T	1.37E-06	1.31	0.056
	rs2350892	60,792,103	G	A	1.52E-06	1.32	0.058
	rs1272131	60,886,150	C	T	1.88E-06	1.29	0.053
	rs1254276	60,847,001	T	C	2.25E-06	1.28	0.053
	rs2093210	60,957,279	C	T	2.41E-06	1.29	0.054
	rs1313237	60,848,224	C	A	2.51E-06	1.29	0.053
	rs33912345	60,976,537	C	A	2.74E-06	1.29	0.054
	rs7159392	61,008,596	T	C	2.98E-06	1.29	0.054

14q23	rs4899012	61,003,889	G	C	3.04E-06	1.29	0.054
	rs2351174	61,013,237	G	T	3.43E-06	1.28	0.054
	rs1955695	61,021,891	A	G	3.77E-06	1.28	0.053
	rs4442732	61,025,617	A	G	3.79E-06	1.28	0.053
	rs10483727	61,072,875	T	C	4.82E-06	1.27	0.052
	rs34935520	61,091,401	G	A	4.88E-06	1.27	0.053
	rs35155027	61,095,174	G	C	4.97E-06	1.27	0.053
	rs4902067	61,956,958	G	C	6.07E-06	0.83	0.042
	rs35320790	61,108,825	C	A	6.59E-06	1.27	0.054
	rs11622805	61,954,594	A	G	7.40E-06	0.83	0.042
	rs35014123	61,954,761	A	C	7.61E-06	0.83	0.042
	rs4902063	61,953,839	T	C	8.00E-06	0.83	0.042
	rs12886896	61,954,272	T	C	8.20E-06	0.83	0.042
	rs1957898	61,953,406	C	T	8.56E-06	0.83	0.043
	rs2253026	61,990,633	G	A	8.74E-06	1.21	0.043
	rs3783774	61,951,101	G	C	1.06E-05	0.83	0.042
	rs912619	61,949,933	C	T	1.06E-05	0.83	0.042
	rs1886464	61,949,051	A	G	1.08E-05	0.83	0.042

14q23	rs2252267	61,984,684	A	G	1.54E-05	1.20	0.043
	rs1088683	61,978,571	C	T	1.56E-05	1.20	0.043
	rs1091331	61,998,747	C	T	1.85E-05	1.20	0.042
	rs1091680	61,984,087	A	G	1.94E-05	1.20	0.043
	rs1088667	61,991,812	G	C	2.55E-05	0.84	0.042
	rs12883769	61,990,861	C	T	2.79E-05	0.84	0.042
	rs1885095	60,794,782	C	T	2.82E-05	1.26	0.054

A1, A2: alleles 1 and 2, OR: odds ratio, SE: standard error

Table 1-6: Replication results of previously-reported associated loci using our sample sets

Locus	Gene	SNP ID	Chr. position	A1	A2	MAF		P^a	OR	95% C.I	
						Case	Control			L	U
1q23-q24	***MYOC***	rs2236875	169,881,306	A	C	0.145	0.155	0.61	0.97	0.86	1.09
		rs12035960	169,881,829	T	C	0.145	0.154	0.69	0.98	0.87	1.10
		rs235917	169,887,202	A	G	0.109	0.102	0.14	1.11	0.97	1.27
		rs2075648	169,888,457	A	G	0.036	0.037	0.92	0.99	0.79	1.24
		rs235913	169,885,279	T	G	0.278	0.285	0.92	1.01	0.92	1.10
		rs604864	169,875,088	A	G	0.183	0.177	0.43	1.05	0.94	1.16
		rs235920	169,888,598	A	G	0.356	0.359	0.51	1.03	0.94	1.13
		rs235875	169,880,379	A	G	0.489	0.484	0.73	0.98	0.91	1.07
		rs235881	169,870,169	A	G	0.199	0.196	0.62	1.03	0.93	1.14
		rs2032555	169,874,324	C	T	0.182	0.179	0.80	1.02	0.91	1.13
		rs12076134	169,872,677	G	T	0.179	0.182	0.91	0.99	0.89	1.11
1q24	***TMCO1***	rs1913845	163,962,203	T	C	0.132	0.128	0.94	1.01	0.89	1.14
		rs12723198	163,968,970	T	C	0.320	0.320	0.92	1.01	0.92	1.10

1q24	***TMCO1***	rs10918271	163,971,356	T	C	0.188	0.193	0.97	1.00	0.90	1.11
		rs7411708	163,992,901	A	G	0.321	0.323	0.93	1.00	0.91	1.09
5q22.1	***WDR36***	rs1457113	110,494,973	C	T	0.180	0.163	0.058	1.12	1.00	1.25
		rs6870356	110,495,136	T	C	0.088	0.088	0.51	0.95	0.82	1.10
		rs10045255	110,466,256	G	A	0.317	0.334	0.15	0.94	0.86	1.02
		rs10038177	110,464,349	A	G	0.316	0.332	0.15	0.94	0.86	1.02
		rs4530809	110,490,129	G	A	0.292	0.307	0.089	0.92	0.84	1.01
		rs10051830	110,480,744	T	C	0.315	0.330	0.17	0.94	0.86	1.03
		rs1043828	110,491,907	C	T	0.046	0.043	0.25	1.13	0.92	1.38
		rs11241095	110,467,408	G	A	0.228	0.243	0.24	0.94	0.85	1.04
		rs12522383	110,483,165	T	C	0.226	0.241	0.24	0.94	0.85	1.04
		rs2416257	110,463,388	A	G	0.027	0.027	0.40	1.12	0.86	1.45
7q31.1	***CAV1 - CAV2***	**rs7795356**	**116,004,265**	**T**	**C**	**0.053**	**0.044**	**0.0073**	**1.31**	**1.08**	**1.58**
		rs926198	**115,954,443**	**G**	**A**	**0.054**	**0.046**	**0.024**	**1.25**	**1.03**	**1.51**
		rs959173	**115,969,289**	**G**	**A**	**0.052**	**0.045**	**0.027**	**1.25**	**1.03**	**1.51**
		rs6975771	**115,910,080**	**A**	**G**	**0.052**	**0.046**	**0.037**	**1.23**	**1.02**	**1.49**

7q31.1	***CAV1 - CAV2***	rs6466579	115,938,390	A	G	0.276	0.271	0.24	1.06	0.96	1.16
		rs2191502	116,005,322	C	T	0.436	0.427	0.19	1.06	0.97	1.15
		rs17138749	115,920,333	C	A	0.225	0.223	0.60	1.03	0.93	1.13
		rs3807989	115,973,477	T	C	0.348	0.339	0.33	1.05	0.96	1.14
		rs1052990	115,935,606	C	A	0.224	0.225	0.82	1.01	0.92	1.12
		rs1049334	115,987,615	T	C	0.292	0.292	0.90	0.99	0.91	1.09
		rs3815412	115,977,928	G	A	0.294	0.293	0.96	1.00	0.91	1.09
10p13	***OPTN***	rs1953314	13,175,837	T	C	0.393	0.365	0.075	1.08	0.99	1.18
		rs7095146	13,201,415	A	C	0.453	0.452	0.54	1.03	0.95	1.12
		rs1561570	13,195,732	G	A	0.453	0.454	0.63	1.02	0.94	1.11
		rs825411	13,209,380	T	C	0.475	0.479	0.86	1.01	0.93	1.10
		rs660592	13,188,272	A	G	0.433	0.416	0.45	1.03	0.95	1.13
		rs10906308	13,210,473	A	G	0.392	0.385	0.86	0.99	0.91	1.08
		rs17512962	13,209,072	T	C	0.091	0.094	0.88	1.01	0.87	1.17
		rs2234968	13,191,230	T	C	0.169	0.172	1.00	1.00	0.89	1.12
10p13	***OPTN***	rs583277	13,206,881	G	A	0.393	0.385	0.89	0.99	0.91	1.08

rs2244380	13,198,268	G	A	0.278	0.276	0.81	1.01	0.92	1.11
rs10796028	13,209,518	C	A	0.433	0.425	0.85	0.99	0.91	1.08

[a] P-value of the logistic regression test (age, gender, the first two eigenvectors, and GC-adjusted).

The results of SNPs shown are from the GWAS stage

Table 1-7: correlation of top SNPs in combined analyses of POAG with IOP and VCDR

Trait / SNP	No.	Effect	S.E.	*P* [a]
Intraocular pressure	2,434			
rs1063192		-0.32	0.21	0.13
rs10483727		-0.91	0.19	3.75×10^{-6}
Vertical cup-disk ratio	1,082			
rs1063192		-0.01	0.01	0.2
rs10483727		-0.01	0.01	0.3

[a] *P*-value obtained by linear regression model following adjustment for age and gender.

MAF: minor allele frequency, S.E: standard error,

Table 1-8: Primers for *SIX6* re-sequencing analysis

Primer name	Sequence	Length	Tm (^{o}C)	Product size (bp)
5' UTR_ F	ttcaagctttgattggcaga	20	59.5	400
5' UTR_ R	tcttccagggtctcacatacc	21	59.0	
Exon 1-1_F	cagctgcccatcttgaattt	20	60.2	299
Exon1-1_R	agcttctcagcctcctggta	20	59.2	
Exon1-2_F	cacaagttcaccaaggagtcg	21	60.7	448
Exon1-2_R	ctcctgctgaattggagtcg	20	60.9	
Exon2-3' UTR 1_F	gacccgtgttccctttcttc	20	60.9	481
Exon2-3' UTR 1_R	ttccgaaggagactttgcag	20	60.5	
3' UTR 2_ F	gcaccacgttcttcttgctt	20	60.4	391
3' UTR 2_ R	gcctggtgaagggaatgata	20	59.9	

Table 1-9: Results of re-sequence analysis of *SIX6* gene

Location	Varian	Position	Type	Alleles 1/2	Case (191)				Control (93)				Function *
					11	12	22	MAF	11	12	22	MAF	
5` UTR	rs1956558	60045709	Reported SNP	C/T	150	37	3	0.113	72	20	1	0.118	TFBS
5` UTR	t>c 31	60045721	Novel: SNP	C/T	186	5	0	0.013	90	3	0	0.016	
Exon 1	t>c 275	60045965	Novel: rare variant	C/T	188	1	0	0.003	90	0	0		Codon 32 (CGC → CGT) synonymous (Arg → Arg)
Exon 1	t>g 598	60046288	Novel: rare variant	G/T	188	1	0	0.003	90	0	0		Codon 140 (CGG → CTG) missense (Arg [+] → Lue) Effect (Benign)
Exon 1	rs33912345	60046290	Reported SNP	C/A	72	63	54	0.452	46	37	7	0.283	Codon 141 (CAC → AAC) missense (Arg [+] → Asn [-]) Effect (Benign)
Exon 2	t>c 816	60047618	Novel: rare variant	C/T	189	2	0	0.005	88	3	0	0.016	Codon 213 (CCA → TCA) missense (Pro → Ser) Effect (Benign)
3` UTR	rs1061108	60047824	SNP	G/C	144	40	5	0.132	64	26	1	0.154	microRNA-BS

The effect of amino acid change on the function was detected using SNPinfo database.

Abbreviations: TFBS: transcription factor binging site

Chapter 2: Genome-wide association study of major components of serum proteins in the Japanese population

Section 2.1.1 Introduction

Serum proteins possess various biological functions such as hormones, enzymes, antibodies, and clotting agents, and some serve as valuable biomarkers that reflect several disease conditions. For example, many fractions of major serum proteins (like albumin) are synthesized in liver, therefore, decreased total protein level and albumin are observed in patients with liver dysfunction such as liver cirrhosis. Nephropathy patients also indicate low total protein level due to increased secretion of protein to urine. In contrast, patients with multiple myeloma or chronic infection indicated elevated production of immunoglobulin which caused high total serum protein level.

Major components of serum proteins are albumin (ALB) (approximately 60%), globulins (mainly as γ-globulins, approximately 30%), and fibrinogens (**Fig. 2-1**). Total serum protein levels range from 6.5 to 8.5 g/dl and show significant inter-individual variation. These variations are found to be influenced by environmental factors stress, diet, and exposure to toxins, pathogens, radiation and chemicals, although sex, alcohol use, age and BMI also contribute significantly to inter-individual variation (**1**). However, genetic factors are also known to affect their levels although the range of genetic effects varies by the reports from 20% to 77% **(1)**. Genome-wide association studies (GWAS) recently demonstrated that serum levels of several proteins can be strongly influenced by common genetic variants through either *cis* or *trans* effects **(2-4)**.

Our group previously reported the GWAS results for hematological and biochemical traits, including TP and ALB, in the Japanese population **(5)**. An associated SNP for

TP, rs4273077 (*P*-value = 4.5×10^{-10}), is located in an intron of *TNFRSF13B* (Tumor Necrosis Factor Receptor Superfamily member 13B), which encodes TACI (transmembrane activator and calcium-modulator and cytophilin interactor), one of three TNF-receptor family members (BAFF-R, TACI, and BCMA) **(6)**. However, since rs4273077 showed no significant association with the serum ALB level (P = 0.089), I suspected that this SNP would have genetic effects primarily on the levels of the non-albumin fraction. TACI is expressed mainly in activated B cells and binds with a high affinity to two TNF ligands; APRIL (a Proliferation-Inducing Ligand, encoded by *TNFSF13*), and BAFF (B Cell-Activating Factor, encoded by *TNFSF13B*) **(7)**. TACI is implicated in B- cell homeostasis (including B- cell survival, activation, and differentiation), immunoglobulin production, and antibody class switching **(8-10)**. Hence, the association of variants in *TNFRSF13B* with TP is likely to reflect the immunoglobulin serum levels.

2.1.2 Objective

To identify the genetic variations associated with serum levels of non-albumin proteins (NAP), particularly those of immunoglobulins by GWAS of Japanese subjects.

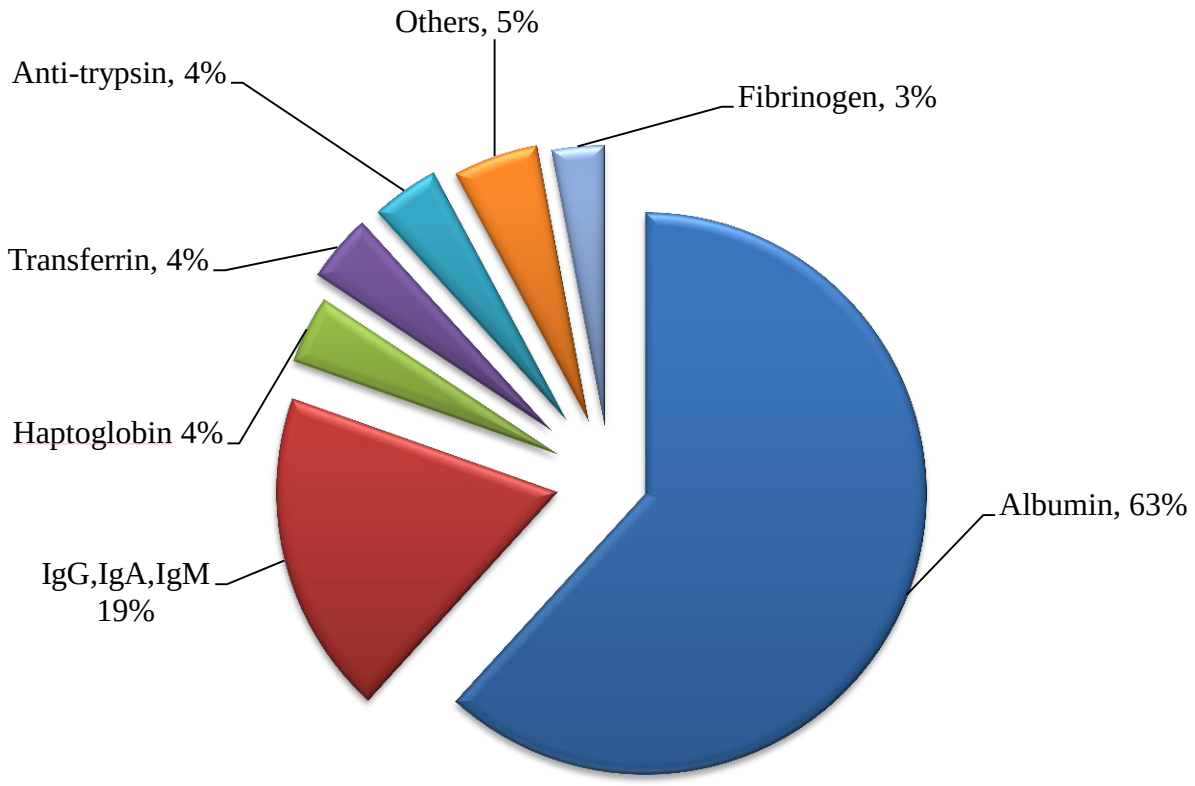


Figure 2-1: Serum total protein fractions.

Section 2.2 Subjects and Methods

2.2.1 Study cohorts

For the GWAS, 9,103 subjects derived from 10 disease cohorts (colorectal cancer, breast cancer, prostate cancer, lung cancer, gastric cancer, diabetes mellitus, peripheral artery disease, atrial fibrillation, ischemic stroke, and myocardial infarction) were selected, and for the replication study, I used data from >1600 independent individuals selected from the BioBank Japan Project **(11)** (**Table 2-1**). For immunoglobulin isotypes analyses, the data from ~1,600 additional individuals in BioBank Japan was used (**Table 2-1**). The clinical information for the samples is updated annually using a standard questionnaire in the 66 hospitals participating in the project. DNA of individuals with SLE and IgAN were obtained from BioBank Japan also, as shown in our previous genome-wide association studies of SLE and IgAN (**12, 13**).

2.2.2 Ethical approval

Written informed consent was obtained from all subjects. The research project was approved by the ethical committees in the Institute of Medical Science, the University of Tokyo, and the Center of Genomic Medicine, RIKEN, Yokohama, Japan.

2.2.3 Genotyping and quality control (Q.C) filters

In the GWAS, SNPs were genotyped using the Illumina HumanHap610-Quad BeadChip (Illumina, CA, USA). After the exclusion of samples with call rates of <0.98, I excluded closely related individuals (in 1st or 2nd degree kinships) using identity-by-descent (IBD) evaluated by PLINK version 1.0.6 **(14)**. I also excluded individuals who were outliers in the cluster analysis using the principle component analysis performed by EIGENSTRAT 3.0 along with HapMap Phase Ⅱ populations (**Fig. 2-2 A and B**). In addition, SNPs with call rates of <0.99, MAF of < 0.01 and Hardy Weinberg equilibrium of P <1.0 x 10^{-7} were excluded.

Genotyping data of the SNPs selected for replication analyses and for testing with immunoglobulin levels were generated using multiplex PCR- based Invader Assay (Third Wave Technologies, Madison, WI, USA) **(15)**. Genotypes were judged by visual inspection, following the application of QC measures of individuals' call rates of >98% and SNPs call rates of >99% of individuals. I could not obtain the genotype data of rs3817588 in *GCKR* using the Invader assay.

2.2.4 Whole-genome imputation of genotypes

Yukinori Okada (Laboratory of Statistical analysis, RIKEN) performed whole-genome imputation of the GWAS subjects in a two-step procedure, as described by Okada et al. **(16)**. HapMap phase Ⅱ Japanese (JPT) and Han Chinese (CHB) individuals (release 24) were adopted as reference panels. I excluded the imputed SNPs with MAF of <0.01 or *Rsq* of <0.7. As a result, a total of 2,178,644 SNPs on autosomal chromosomes were used for the GWAS.

2.2.5 Statistical analysis

I obtained the non-transformed values of TP, ALB and NAP (mg/dl) for the subjects from the clinical information stored in BioBank Japan, and adjusted them 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 and subjects with *Z* scores of <-4 or >4 were removed from each trait analysis. The associations of the SNPs with *Z* scores were evaluated in linear regression models assuming additive effects of allele dosages, using mach2qtl software. The same methods of data normalization and statistical models were applied for the replication analyses and for testing the association with common log-transformed values of immunoglobulin isotypes (IgG IgA, IgM, and IgE). Meta-analyses of the GWAS and the replication study were performed using the inverse-variance method assuming a fixed-effects

model. To test if any heterogeneity exists among the 10 cohorts used for the analysis, I performed Pearson's Chi-squared test for homogeneity for the identified SNPs. For rs4273077, *P*-value = 0.53, for rs11552708, *P*-value = 0.44, and for rs3803800, *P*-value = 0.46. These *P*-values indicate that there is no significant heterogeneity among those 10 cohorts (they are homogenous), and hence the results of the association analysis are not affected by being from different cohorts.

The significance level used was $5x10^{-8}$ in the GWAS stage. For the replication stage, I considered 0.05 as significant for the association of rs4985726 with TP and rs1260326 with ALB. For the association of SNPs rs4985726 in TNFRSF13B, rs3803800 and rs11552708 in TNFSF13 with NAP, 0.017 (0.05/3) was considered to be significant. These significance levels represent the Bonferroni correction for multiple statistical tests. In addition, I set a level of 0.05 to consider the association of the selected SNPs with immunoglobulin isotypes 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.

The haplotype analyses were performed using the Haplo Stats package (version 1.4.0) implemented in *R* statistical software. Epistatic effects of the SNPs in *TNFRSF13B* and *TNFSF13* were evaluated using a linear regression model incorporating the product of the allele dosages of the SNPs in the loci as an independent variable. All statistical analyses including haplotype analyses were performed using the *R* statistical software version 2.9.1 except for genome-wide linear regression analyses. LD analyses were performed using Haploview 4.2 software, PLINK, and the SNAP database.

2.2.6 ELISA for serum APRIL

Serum APRIL levels were assayed using set of 193 serum samples stored in BioBank Japan Project**.** Serum samples were assayed for Human APRIL by antigen-capture enzyme-linked immunosorbent assay (ELISA) (Bender MedSystems, Vienna, Austria) .The detection level of the test was set as (0.8ng/ml) according to the manufacture's structures. Samples were analyzed in duplicated.

2.2.7 eQTL analysis

The expression quantitative tract loci analysis (eQTL) was performed as mentioned in chapter 1.

Section 2.3 Results

2.3.1 GWAS of total protein (TP), albumin (ALB), and non-albumin protein (NAP)

I conducted a GWAS using genotyping data and clinical information on 9,103 individuals who had been collected in the BioBank Japan Project (**Table 2-1**).

A

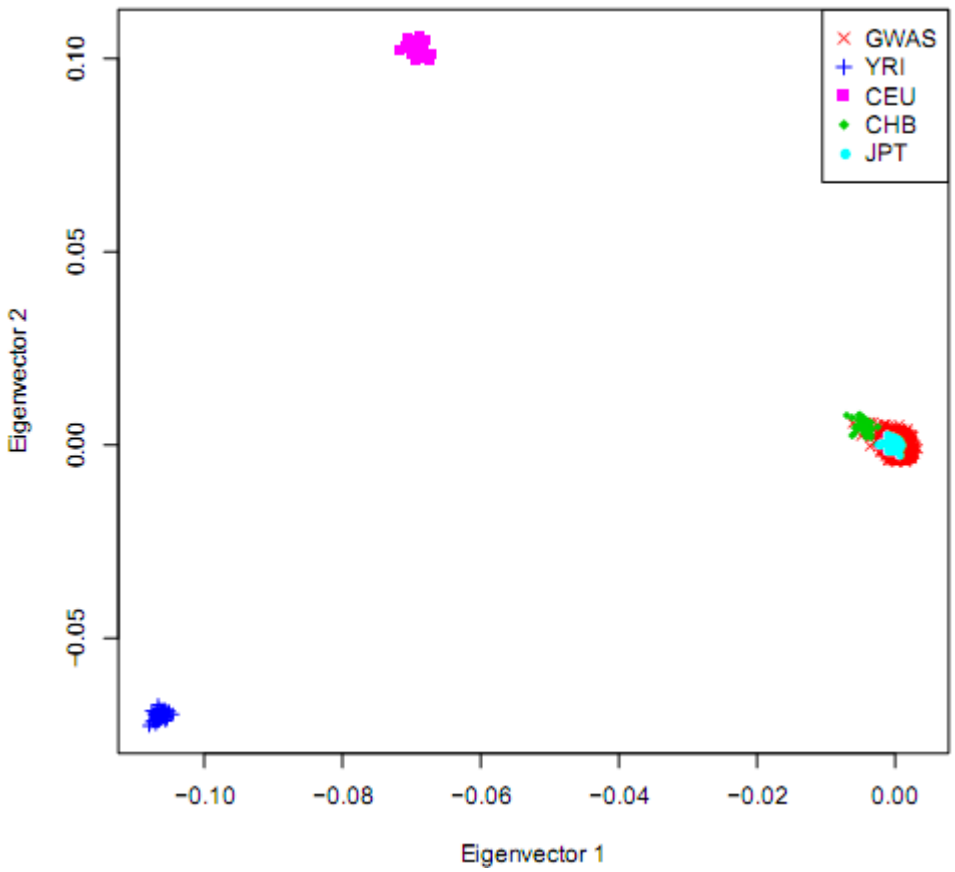


Figure 2-2 (A): Principal component analysis Plot of cohorts included in the GWA major components of serum proteins.
All individuals who were finally incorporated in the GWAS together with the four populations in the HapMap Phase II database (Japanese: JPT; Han Chinese: CHB; Africans: YRI, and European: CEU) were plotted based on the first two eigenvectors.

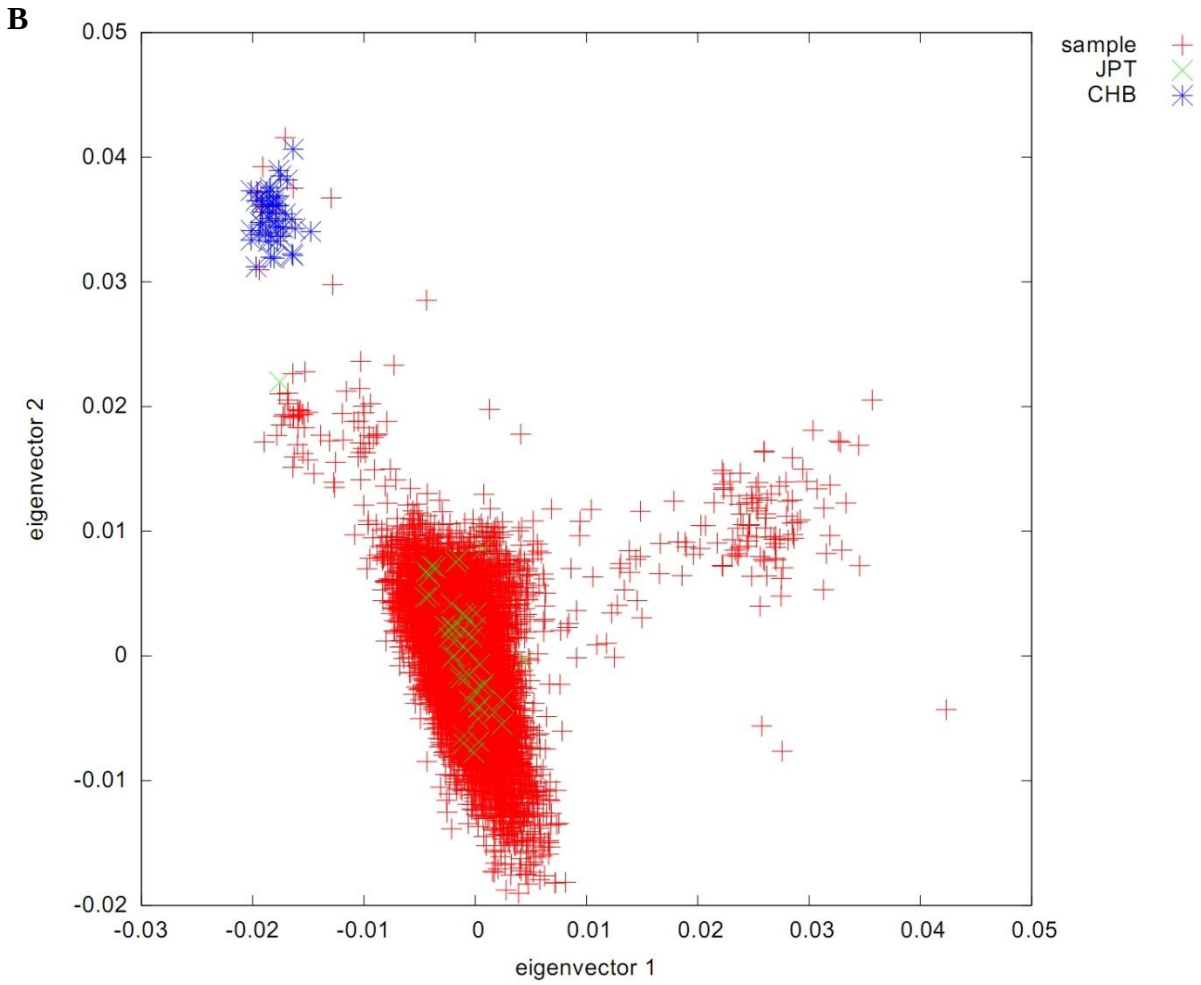


Figure 2-2 (B): Principal component analysis Plot of cohorts included in the GWA major components of serum proteins.

All cases and controls who were selected for the GWAS together with the two populations of Asian ancestry in the HapMap phase II database (Japanese: JPT and Chinese: CHB) were plotted based on the first two eigenvectors. > 99% of cases and controls were clustered within either of the two major clusters of the Japanese population.

Genotyping was performed using Illumina Human610-Quad BeadChip (Illumina, CA, USA). After applying stringent quality control (QC) filters for selection of individuals and SNPs (**Methods**), whole-genome imputation analysis was performed using the data of HapMap Phase II East Asian populations, and I obtained the information of 2,178,644 SNPs on autosomal chromosomes with minor allele frequencies (MAF) of ≥ 0.01 and *Rsq* of ≥ 0.7. I then evaluated the association of the SNPs with the adjusted *Z* scores of serum levels of total protein (TP), albumin (ALB), and non-albumin protein (NAP). A Quantile-quantile (Q-Q) plot for each trait indicated low possibility of population stratification (inflation factors (λ_{GC}) for TP, ALB and NAP were 1.04, 1.02 and 1.02, respectively) (**Fig. 2-3**).

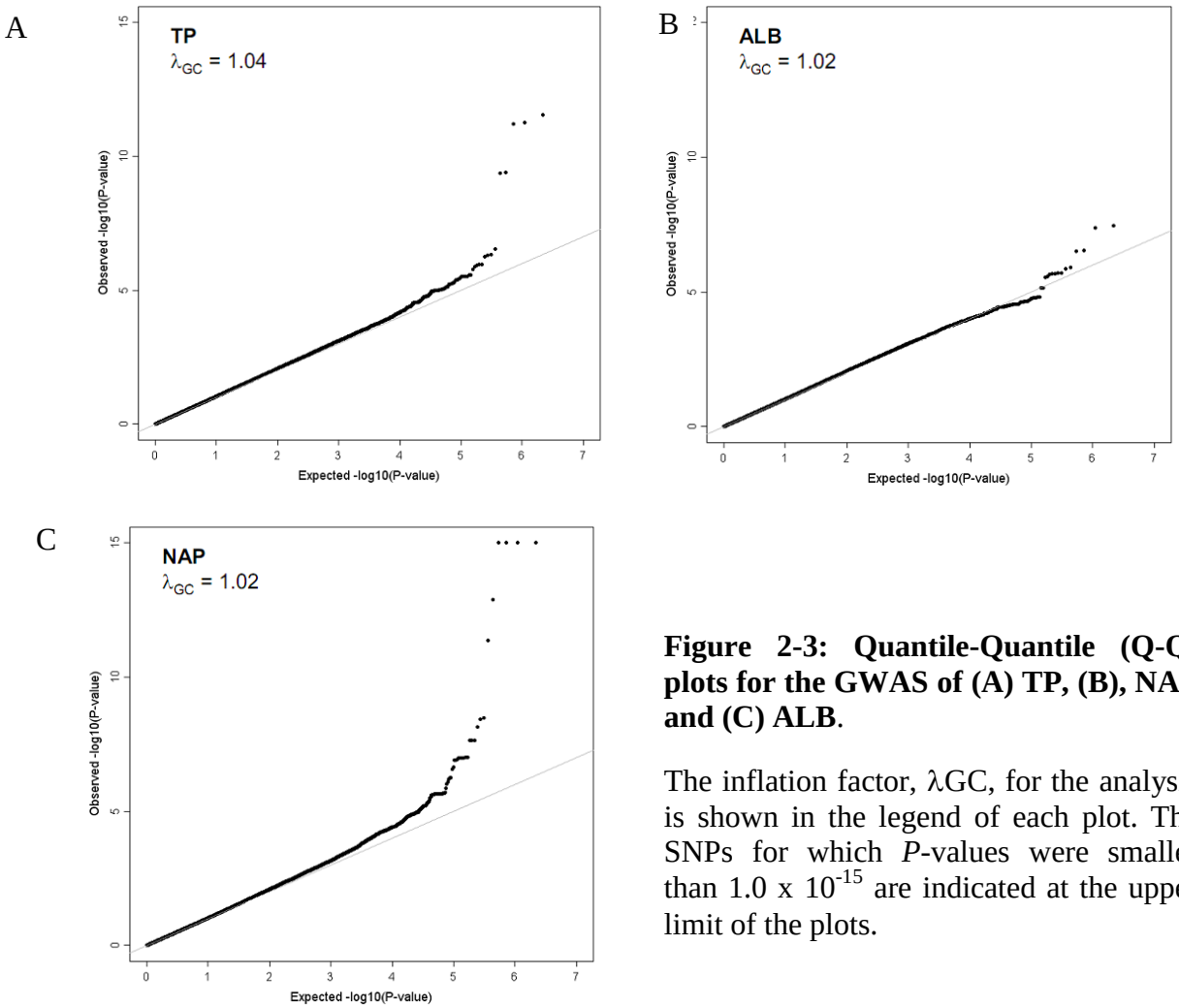


Figure 2-3: Quantile-Quantile (Q-Q) plots for the GWAS of (A) TP, (B), NAP and (C) ALB.

The inflation factor, λGC, for the analysis is shown in the legend of each plot. The SNPs for which *P*-values were smaller than 1.0 x 10^{-15} are indicated at the upper limit of the plots.

Several SNPs with strong linkage disequilibrium (LD) ($r^2 > 0.8$) in intronic regions of *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) (**Table 2-3, Table 2-4, Fig. 2-4 A**). 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) (**Table 2-3, Table 2-4, and Fig. 2-4 B**).

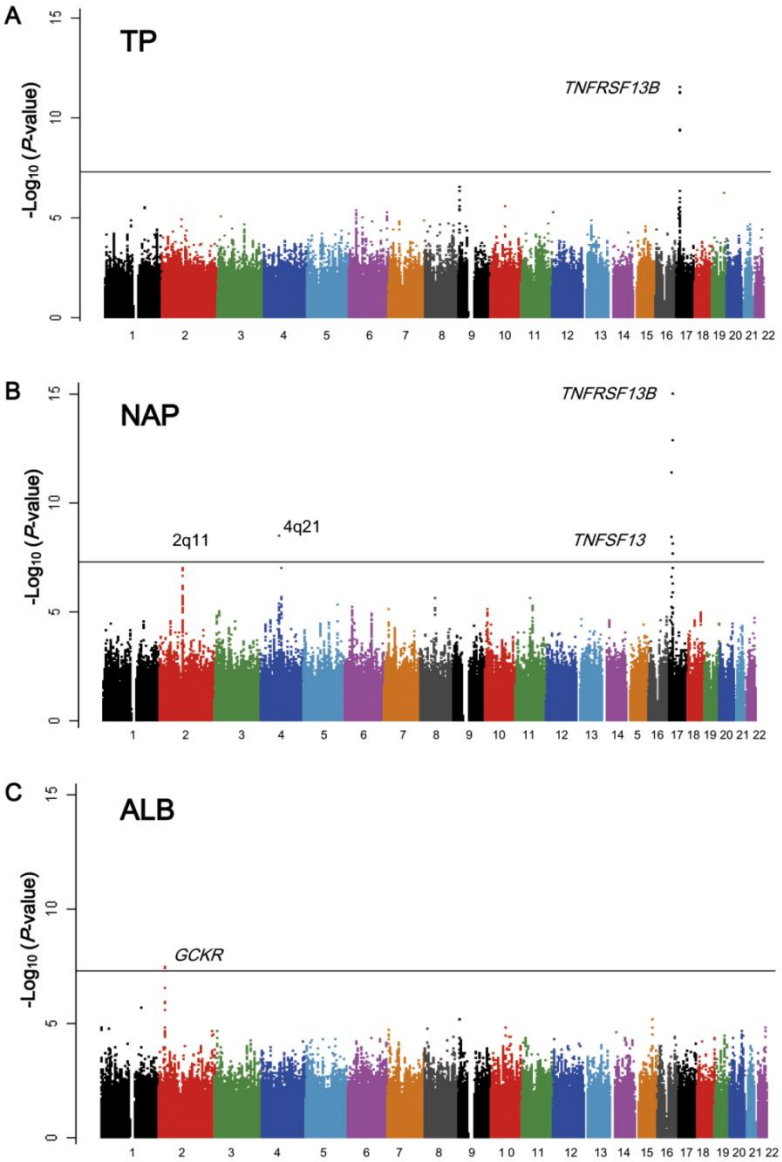


Figure 2-4: Manhattan plots for the GWAS of (A) TP, (B) NAP, and (C) ALB.

SNPs were plotted based on their physical chromosomal positions (horizontal axis) together with their –log10 (*P*-values) in the GWAS (vertical axis). The black horizontal line shows the genome-wide significance threshold of $P = 5.0 \times 10^{-8}$. The SNPs for which P-values were smaller than 1.0×10^{-15} are indicated at the upper limit of the

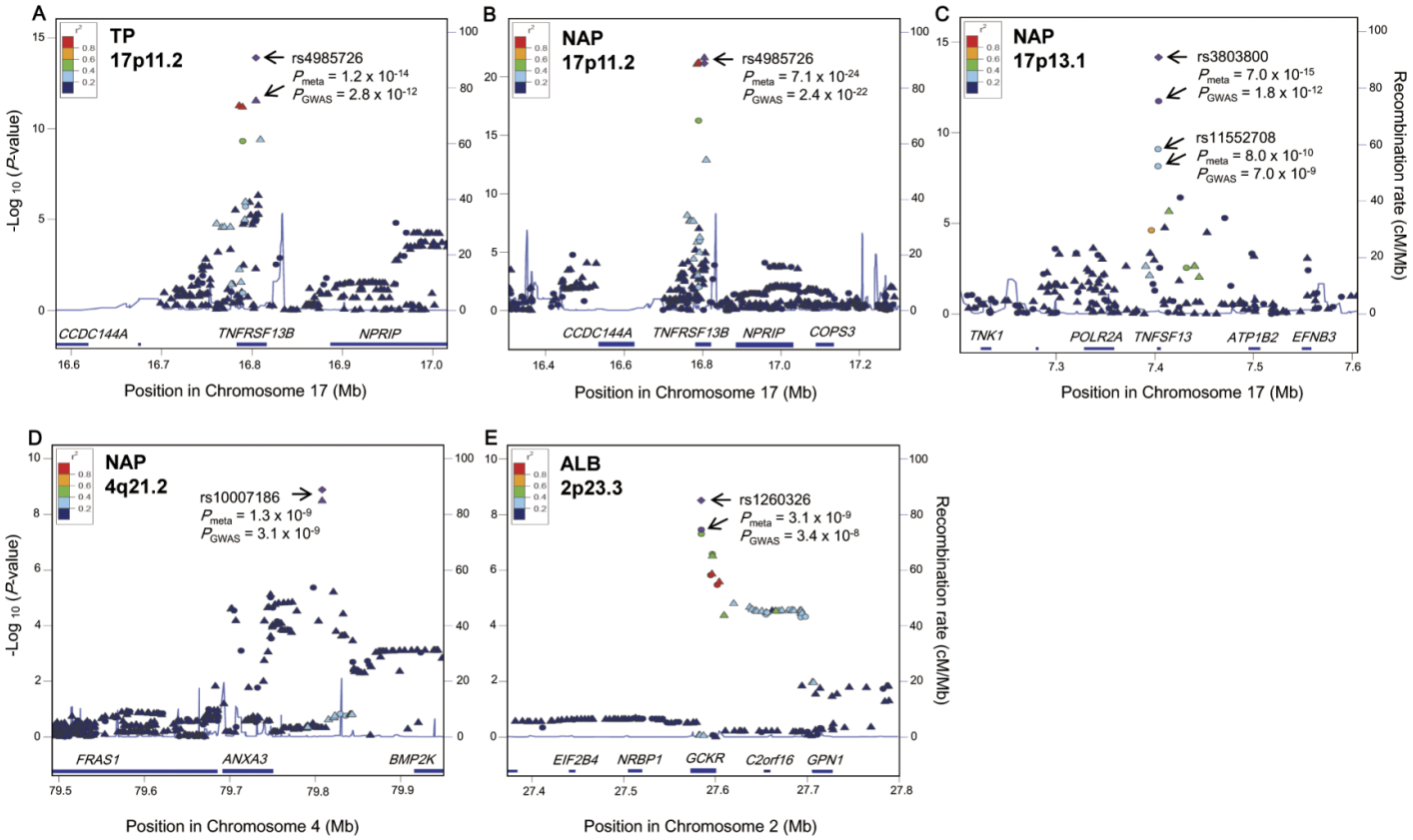


Figure 2-5: Regional plots for the associations of the SNPs in the GWAS stage of TP, ALB and NAP.

SNPs plotted with their –log10 (*P*-values) in the GWAS based on their physical chromosomal positions. Genotyped SNPs are indicated as circles, while imputed SNPs are indicated as triangles. The color scheme indicated the linkage disequilibrium displayed as r^2 values between all SNPs and the top-ranked SNP in each plot. The tested trait, chromosomal locus, and the top-ranked SNPs (in purple color) in the GWAS and combined analyses together with their P-values are shown in each plot. The blue lines represent the recombination rates estimated based on HapMap Phase II database. The plots were drawn using Locus Zoom software.

Since *TNFSF13* encodes APRIL, a ligand of TACI encoded by *TNFRSF13B*, this ligand-receptor interaction is likely to play a critical role in regulation of the serum NAP levels. However, I did not find any synergistic effects between SNPs in the receptor and ligand on NAP levels.

Rs10007186 located near *ANXA3* (annexin A3) on chromosome 4q21.2 also revealed significant association with NAP (P = 3.3x10^{-9}; **Table 2-3, Table 2-4, Fig. 2-4 B, and Fig. 2-5 D**), and a cluster of highly linked SNPs near the 5' flanking region of *AFF3* (AF4/FMR2 family, member 3) on 2q11.2 indicated suggestive associations with NAP (rs4851274, P = 9.95x10^{-8}) (**Table 2-3**). For serum ALB, SNPs rs1260326 (in exon) and rs3817588 (in intron) in *GCKR* (glucokinase regulator) on 2p23.3 revealed significant associations (P = 3.4x10^{-8} and 4.1x10^{-8}, respectively) (**Table 2-3, Table 2-4, and Fig. 2-5 E**).

Conditional logistic regression analysis for the SNPs on 17p13.1 indicated that both rs3803800 and rs11552708 conferred independent associations with NAP levels when adjusted for each other (P <0.023). These two SNPs were in strong LD (D`= 0.99, r^2= 0.30) and the haplotype analysis of these two SNPs identified that a haplotype (rs3803800 [A] – rs11552708 [G]) revealed stronger association with NAP than individual SNP (P = 2.59x10^{-13}) (**Table 2-5**). Similarly, rs1260326 and rs3817588 in *GCKR* exhibited independent associations with ALB levels (P < 0.022), and were in LD (D` = 0.95, r^2= 0.50). Moreover, the haplotype (rs1260326 [C] – rs3817588 [C]) indicated stronger association with serum ALB (P = 2.83 x 10^{-9}) (**Table 2-6**). For the 17p11.2 and 4q21.2 loci, no SNP remained significant after accounting for the effect of marker the SNPs rs4985726, and rs10007186, respectively.

When I examined the genetic contribution of these variations for the traits, the combinations of the SNPs indicated above could explain nearly 0.5%, 2.3%, and 0.3% of variations in serum TP, NAP, and ALB, respectively.

2.3.2 Replication study

To validate the GWAS results, I performed a replication study using an independent set of ~1,600 subjects from BioBank Japan (**Table 2-1**). For each trait, I selected marker SNPs for the replication analysis at each locus that indicated the genome-wide significant level of 5.0×10^{-8} (rs4985726 in *TNFRSF13B*, rs3803800 in *TNFSF13*, rs1260326 in *GCKR*, and rs10007186 on 4q21.2). In addition, the two SNPs that remained significant after accounting for the effect of each marker SNP at two loci (rs11552708 in *TNFSF13* and rs3817588 in *GCKR*) were also further investigated.

SNPs rs4985726 in the *TNFRSF13B* locus as well as rs3803800 and rs11552708 in the *TNFSF13* locus revealed significant associations with both TP and NAP (**Table 2-4**). The association of rs1260326 in *GCKR* with serum ALB was also replicated ($P = 0.029$; **Table 2-4**). Meta-analyses combining the GWAS and the replication study yielded stronger associations of these SNPs than the GWAS alone (**Table 2-4 and Fig. 2-5 A-E**). 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 ($P = 1.3 \times 10^{-9}$) (**Table 2-4 and Fig. 2-5 D**).

2.3.3 Association of the SNPs identified in the GWAS of NAP with serum immunoglobulin isotypes

Immunoglobulin isotypes constitute the major components of NAP. Hence, I further examined the NAP-associated SNPs in the GWAS (*TNFRSF13B*, *TNFSF13*, and *ANXA3*) for the

association with various serum immunoglobulins using the samples in BioBank Japan (IgG: n = 1,794, IgA: n = 1,675, IgM: n = 1,649, and IgE: n = 549; **Table 2-2**).

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 ($P < 0.018$) (**Table 2-7**). For IgA, rs3803800 and rs11552708 in *TNFSF13* also revealed the significant association ($P < 0.013$), while rs4985726 in *TNFRSF13B* revealed no significant association ($P = 0.099$) (**Table 2-7**). Rs10007186 near *ANXA3* indicated significant association with IgA ($P = 0.036$), IgM ($P = 0.019$), and IgE ($P = 4.9 \times 10^{-4}$). However, these associated SNPs explained only 1.4%, 0.9%, 1.3%, and 2.0% of the variances of log-transformed values of serum IgG, IgA, IgM, and IgE, respectively.

2.3.4 Association of SNPs in Non-albumin Protein GWAS with SLE and IgA nephropathy

Both SNPs rs3803800 and rs11552708 identified in my non-albumin analysis are reported to be associated with SLE in the Japanese population **(17, 18)**. To test the association of identified polymorphisms in my non-albumin analysis with the process of autoimmunity, SNPs were examined for association with SLE. Cohorts of 193 cases of SLE and 934 healthy controls were analyzed (**Table 2-8**). I selected also to genotype the SNP rs4273077 in *TNFRSF13B* using DNA from these two diseases. This step was done very early before discovering the association of rs4985726, yet rs4273077 is a proxy for rs4985726 because they are highly linked (D' = 1, r^2 = 1) in Europeans but no available data in Japanese. 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). Besides, the other SNP rs3803800 in *TNFSF13* showed a trend of association with SLE. Furthermore, SNPs were also examined for association with IgAN using a cohort of 370 cases (**Table 2-9**). All 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). Moreover, I detected a cumulative effect of risk alleles of SNPs rs4273077 in TACI and rs11552708 in APRIL with the increasing in risk of both SLE and IgAN (**Fig. 2-6 and Fig. 2-7**). 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. These results suggest that TACI-APRIL interaction may play a significant role in both SLE and IgAN susceptibility.

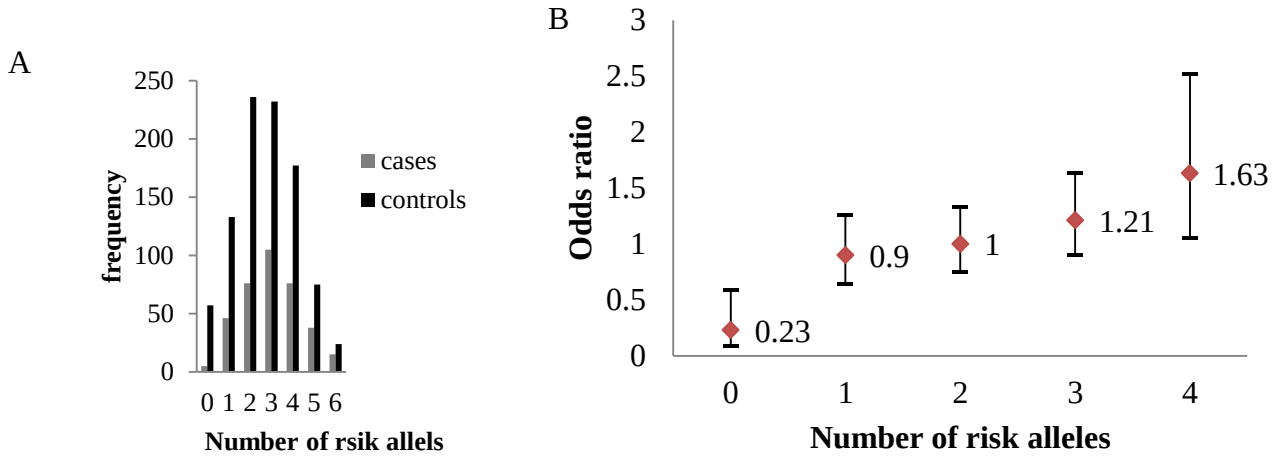


Fig. 2-6: Cumulative risk effect of SLE.

(A) Allele distribution of rs4273077 and rs11552708 in cases and controls (B) Association of number of risk alleles and odds ratio.

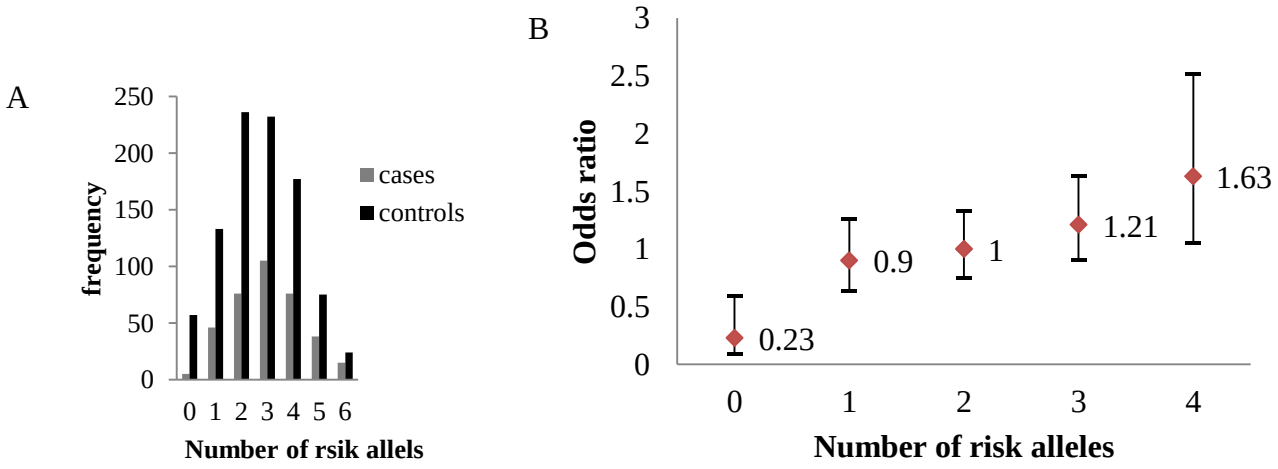


Fig. 2-7: Cumulative risk effect of IgAN.

(A) Allele distribution of rs4273077 and rs11552708 in cases and controls (B) Association of number of risk alleles and odds ratio.

2.3.5 Association of rs3803800 and rs11552708 in *TNFSF13* with serum APRIL levels

To test the association between the two SNPs, rs3803800 and rs11552708, in *TNFSF13* and serum APRIL levels, I measured the serum APRIL levels by ELISA in 193 sera from the samples stored in the BioBank Japan. Although I did not detect a statistically significant difference between serum APRIL levels in regard to genotypes of both SNPs (**Fig. 2-8 A**) as well as by haplotype analysis (*P*-value of haplotype score test = 0.58, 0.41, and 0.23 of haplotype effects as additive, dominant and recessive, respectively), individuals homozygous for the NAP-increasing alleles of both SNPs (rs3803800 [A], rs11552708 [G]) revealed higher percentage of samples above the detection level of the protein in the serum of 0.8ng/ml (**see Methods**), as well as higher serum APRIL levels (**Fig. 2-8 A**). Moreover, individuals with four copies of NAP-increasing alleles demonstrated two times higher serum APRIL levels (13.3ng/ml) than those with zero alleles (6.2ng/ml; **Fig. 2-8 B**).

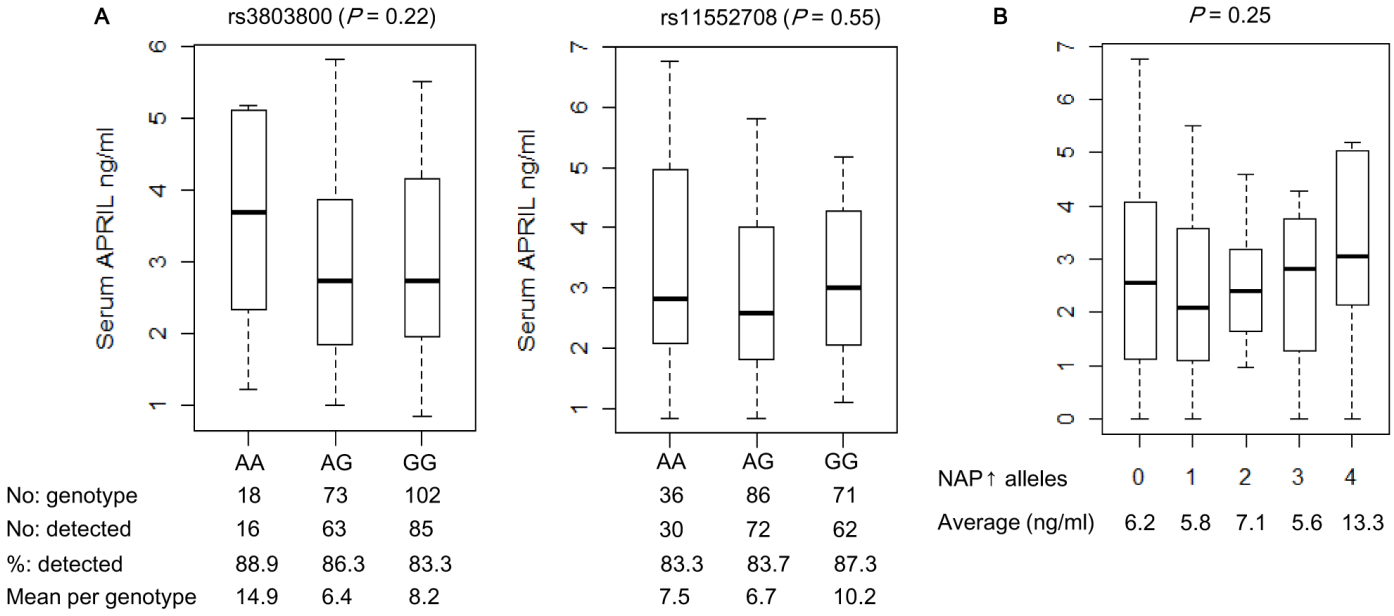


Figure 2-8: Comparison of serum APRIL with regard to the genotypes of (A) rs11552708 and (B) rs3803800.

Serum APRIL levels were measured by ELISA in 193 individuals with brain infarction. The detection level of this test is 0.8ng/ml. In below each genotype; the number of samples above the detection level in relation to the total genotype number is shown. All *P* values were obtained by Kruskal-Wallis rank sum test

2.3.6 eQTL analysis

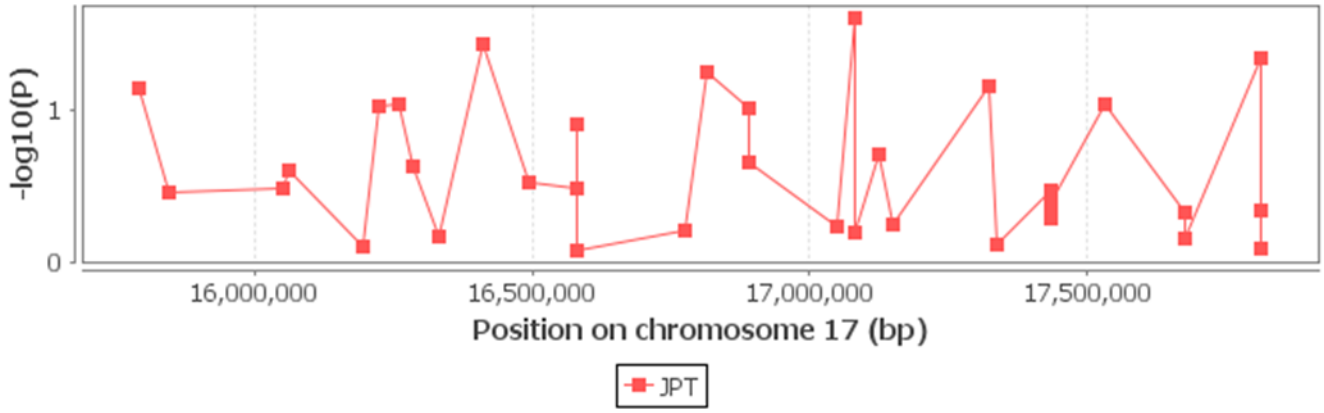
The eQTL analysis was performed using Genevar similarly to that in chapter 1. The SNP rs4985726 can't found in the database, so I used the proxy SNP rs4273077 (in complete LD). Rs4273077does not affect the expression of 33 gene probes in 17p11.2 locus or the gene *TNFRSF13B* (**Fig. 2-9**).

Similarly, rs380300 does not affect the expression of 86 gene probes in 17p13.1 locus or the gene *TNFSF13* (**Fig. 2-10**). However, rs11552708 can significantly affect the expression of the gene *CLEC10A* (C-type lectin domain family 10, member A) in the same locus (**Fig. 2-11 A**) but not with *TNFSF13* (**Fig. 2-11 B**). CLEC10A gene encodes a member of the C-type lectin/C-type lectin-like domain (CTL/CTLD) superfamily. Members of this family share a common protein fold and have diverse functions, such as cell adhesion, cell-cell signaling, glycoprotein turnover, and roles in inflammation and immune response, but it link to NAP or immunoglobulins is not clear.

There is no data for the SNP rs10007186 in 4q21.

In 2p23 locus, no association detected between rs1260326 and 48 gene probes in this locus or with the genes *GCKR* (**Fig. 2-12**).

A



B

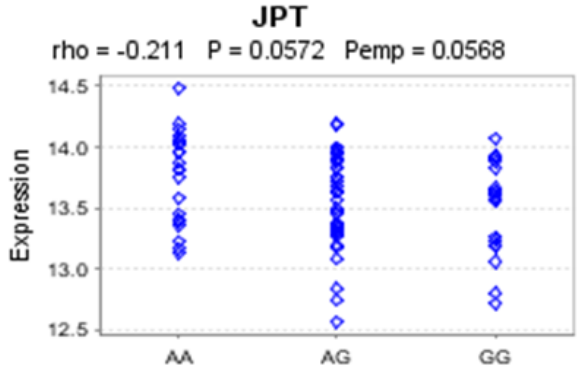
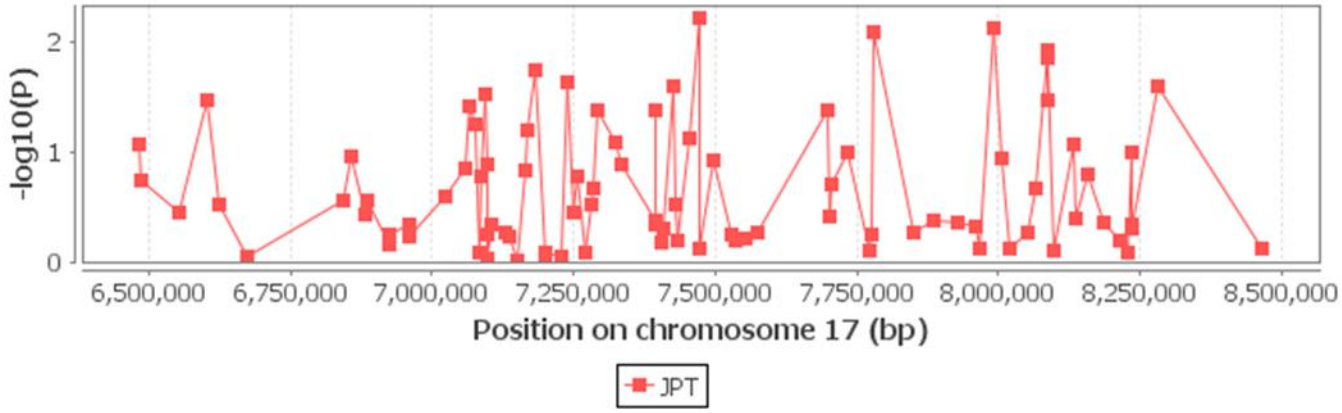


Figure 2-9: eQTL analysis of rs4273077 and genes in 17p11.2 locus.

(A) Association between the SNP rs4273077 and the expression of gene probes in 17p11.2 locus according to Genevar database. Each dot represents one probe. Significance level ($P < 0.001$).

(B) Association between the genotypes of rs4273077 and the expression of *TNFRSF13B*

A



B

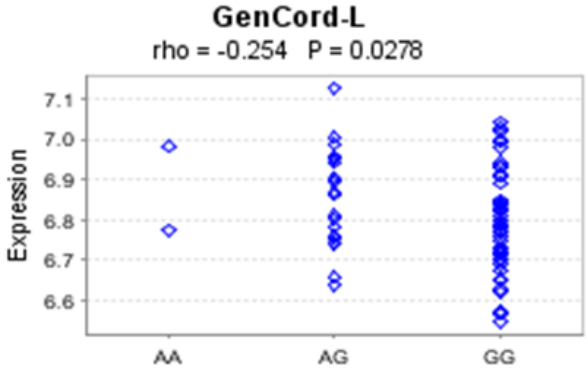
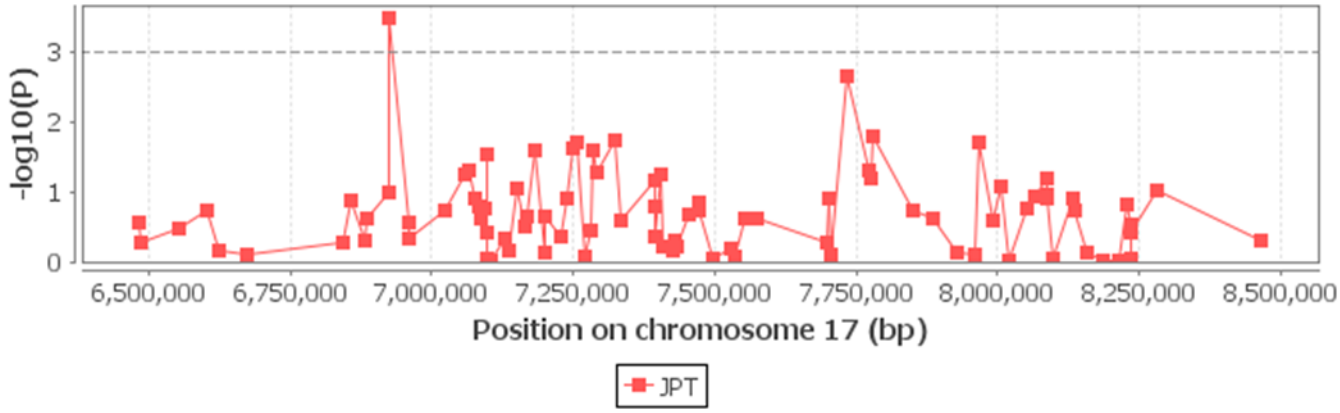


Figure 2-10: eQTL analysis of rs3803800 and genes in 17p13.1 locus.

(A) Association between the SNP rs3803800 and the expression of gene probes in 17p13.1 locus according to Genevar database. Each dot represents one probe. Significance level ($P < 0.001$).

(B) Association between the genotypes of rs3803800 and the expression of *TNFSF13*

A



B

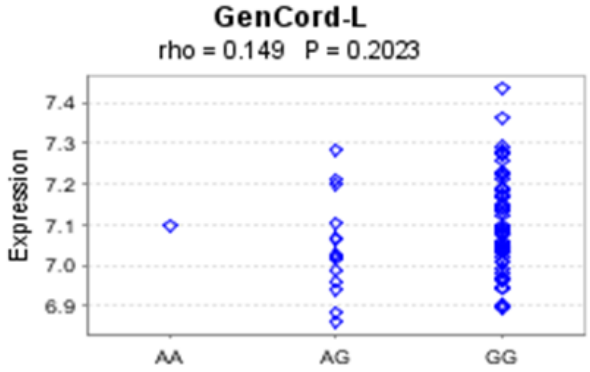
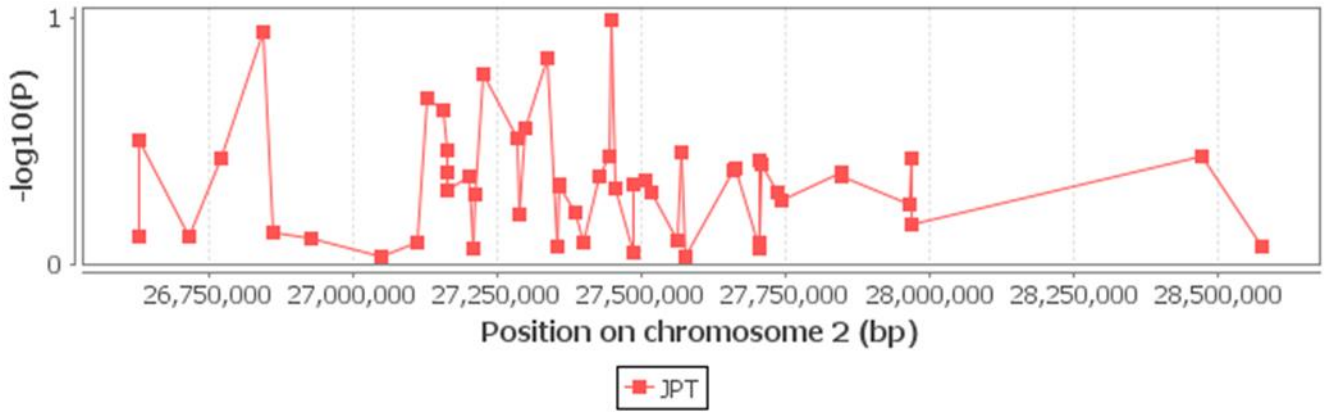


Figure 2-11: eQTL analysis of rs11552708 and genes in 17p13.1 locus.

(A) Association between the SNP rs11552708 and the expression of gene probes in 17p13.1 locus according to Genevar database. Each dot represents one probe. Significance level ($P < 0.001$).

(B) Association between the genotypes of rs11552708 and the expression of *TNFSF13*

A



B

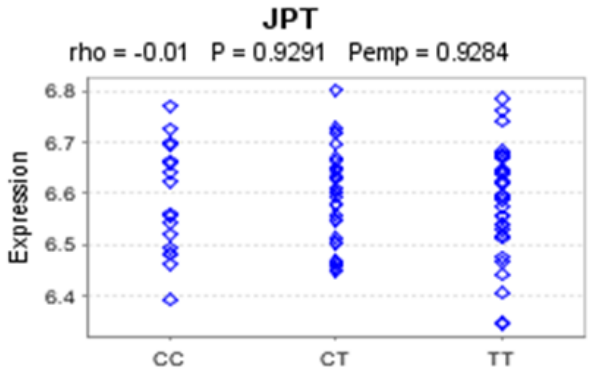


Figure 2-12: eQTL analysis of rs1260326 and genes in 2p23 locus.

(A) Association between the SNP rs1260326 and the expression of gene probes in 2p23 locus according to Genevar database. Each dot represents one probe. Significance level ($P < 0.001$).

(B) Association between the genotypes of rs1260326 and the expression of *GCKR*

Section 2.4 Discussion

On the basis of the information of 10,716 Japanese individuals, I identified one genetic locus (*TNFRSF13B*) on chromosome 17p11.2 associated with both TP and NAP, two loci (*TNFSF13* on 17p13.1 and a region near *ANXA3* on 4q21.2) associated with NAP, and one locus (*GCKR*) on 2p23.3 associated with ALB at the level of genome-wide significance.

The marker SNP rs4985726 shows association with TP and NAP is located in an intron of *TNFRSF13B* on chromosome 17p11.2. A possible mechanism for its association with these traits could be explained by its strong LD with rs34562254 ($D`$ = 1, r^2 = 0.97), which exhibits a missense variation (C>T, Pro251Leu) located in the intracellular domain of the receptor molecule. The *in silico* prediction of the amino acid substitution by rs34562254 in the PolyPhen-2 and SNPinfo database **(19, 20)** suggested a "probably damaging" effect on the protein structure.

The SNPs in *TNFSF13* (encoding APRIL) that identified as being associated with NAP are missense variants; rs3803800 (A>G, Asn96Ser), and rs11552708 (G>A, Gly67Arg). APRIL was first described as having a promoter function for tumor-cell proliferation and survival **(21)**. APRIL is cleaved in the Golgi apparatus by furin at its 104Arg/105Ala site **(22)**, and interestingly, rs3803800 is closely located to this cleavage site. Hence, this SNP might affect the cleavage affinity. Another possibility is the effect on splicing, because both SNPs are predicted to be located within binding sites of splicing regulatory elements **(20)**. However, further investigation should be required to address these possibilities

The SNP rs4985726 in *TNFRSF13B* as well as rs3803800 and rs11552708, in *TNFSF13* also revealed significant associations with serum levels of IgG, IgA, and IgM. It is notable that the two genes encode a TNF-receptor and ligand axis that plays important roles for mediating antibody class switching and regulating immunoglobulin production **(8, 9)**.

Furthermore, knockout mice of either TNFRSF13B or TNFSF13 presented a common phenotype of reduced IgA levels, with impaired antibody response to T cell-independent type I (TNFSF13$^{-/-}$) or type II (TNFRSF13B$^{-/-)}$ antigens **(23, 24)**. In addition, germ-line mutations in *TNFRSF13B* were reported in cases of common variable immunodeficiency (CVID; MIM # 607594) and selective IgA deficiency (IGAD; MIM # 137100) **(25)**. The combination of these significant statistical and biological evidences would suggest that the association of these two loci with NAP reflect at least their associations with regulation of serum immunoglobulin levels. It is also known that immunoglobulins are the major components of NAP, which provides compelling evidence for our results. The facts that both SNPs rs3803800 [A] and rs11552708 [G] in *TNFSF13* were reported to be associated with the susceptibility to the Systemic Lupus Erythematosus (SLE) in the Japanese population and that high serum APRIL was detected in the sera of individuals with the rs3803800 [A]–rs11552708 [G] haplotype **(17)** further support the significance of these SNPs in the regulation of immunoglobulin production. In this study, I observed that possession of two copies of SLE-risk alleles was associated with higher serum levels of NAP, IgG, IgA, and IgM (**Fig. 2-13**), providing a good example of genetic loci that influence both quantitative traits and susceptibility to complex diseases.

Rs10007186, which was associated with NAP (P_{meta} = 1.3 x 10^{-9}) is located about 57.4 kb downstream of *ANXA3* encoding annexin A3, a member of annexin family of calcium-dependent phospholipid-binding proteins **(26)**. Annexin A3 was found to be translocated into phagosomes in dendritic cells **(27)**, which are antigen-presenting cells that serve as messengers between the innate and adaptive immune response, and play a key role in allergic, inflammatory, and autoimmune conditions. In addition, annexin A3 was also found to be associated with neutrophil granule membranes **(28)**, where it can play a regulatory role in calcium-dependent

granule secretions that contribute to acute inflammation and chronic tissue destruction. The association of rs10007186 with IgA, IgM, and IgE, would suggest additional biological roles of annexin A3 in the immune response.

My results showed that SNPs in both *TNFRSF13B* and its ligand *TNFSF13* are significantly associated with SLE and IgAN. Previous study found association between polymorphisms in the tissue necrosis family receptor *TNFRSF6B* (but not *TNFRSF13B* or *TNFSF13*) and IgAN in Caucasians **(29)**. Some reports have also demonstrated modestly increased serum levels of IgA in IGAN **(30, 31)**. The association of this receptor-ligand pathway with IGAN gives more evidence for the role of this pathway in the regulation of IgA serum levels. I replicated the association of the SNPs rs3803800 and rs11552708 with SLE in Japanese population. Furthermore, my data indicate that the associated alleles in my GWAS analyses that lead to positive effect on the total protein, non-albumin protein and globulin values are the risk alleles for autoimmune disorders (**Fig. 2-13**).

I also confirmed the association of SNPs in *GCKR* with serum ALB levels (rs1260326, $P_{meta} = 3.1 \times 10^{-9}$). Rs1260326 is a missense variant (T>C, Leu446Pro) and predicted to cause a damaging effect on the protein structure. *GCKR* is a locus frequently associated with several metabolic traits **(4, 32-34)** and rs1260326 has been reported to be associated with serum triglycerides **(4)**.

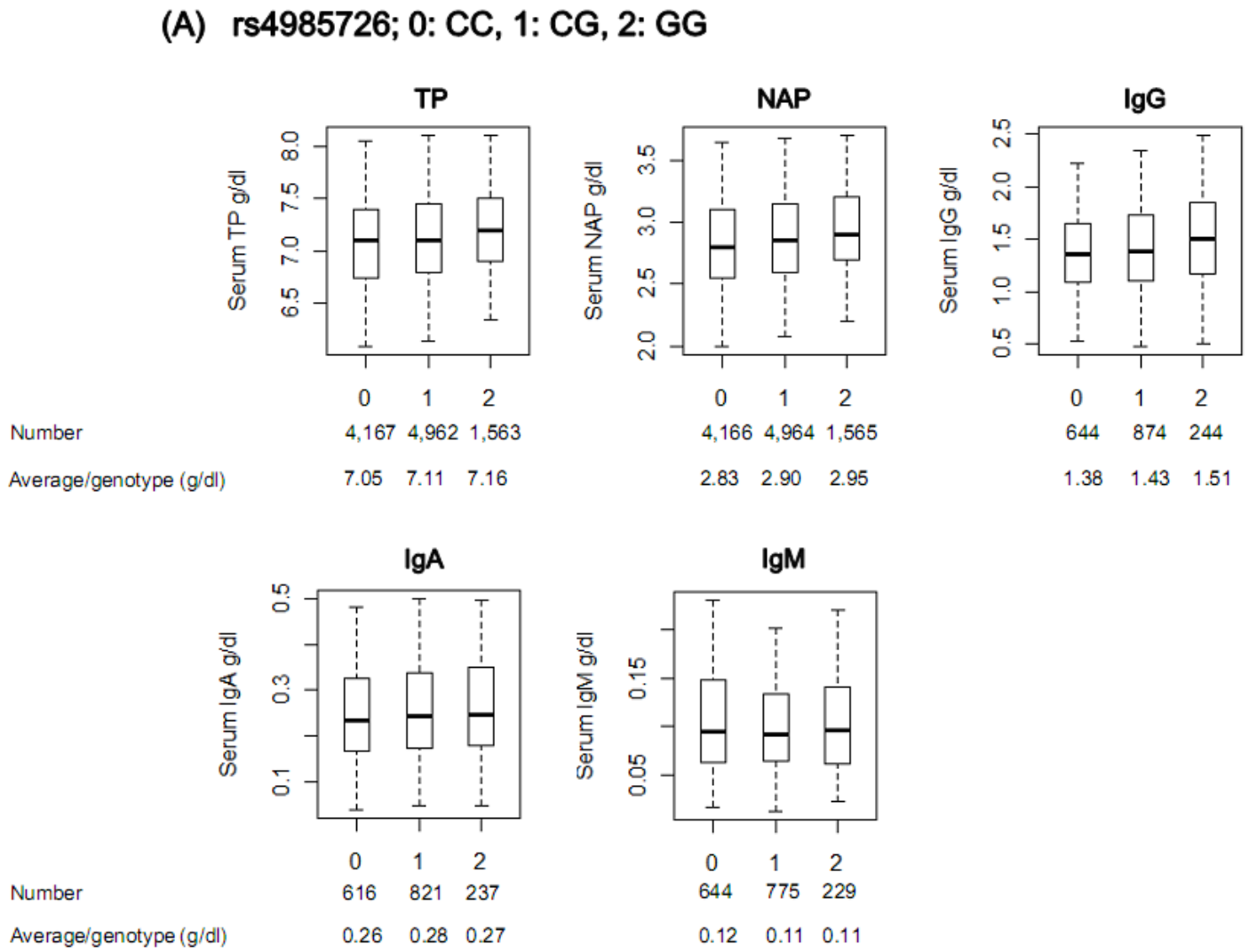


Figure 2-13 A: Relationship between the genotypes of the SNP rs4985726 and the levels of tested proteins.
For each box plot, the bold line indicates the median value which is the 50th quartile. The limits of each box are the 25th and 75th quartiles.

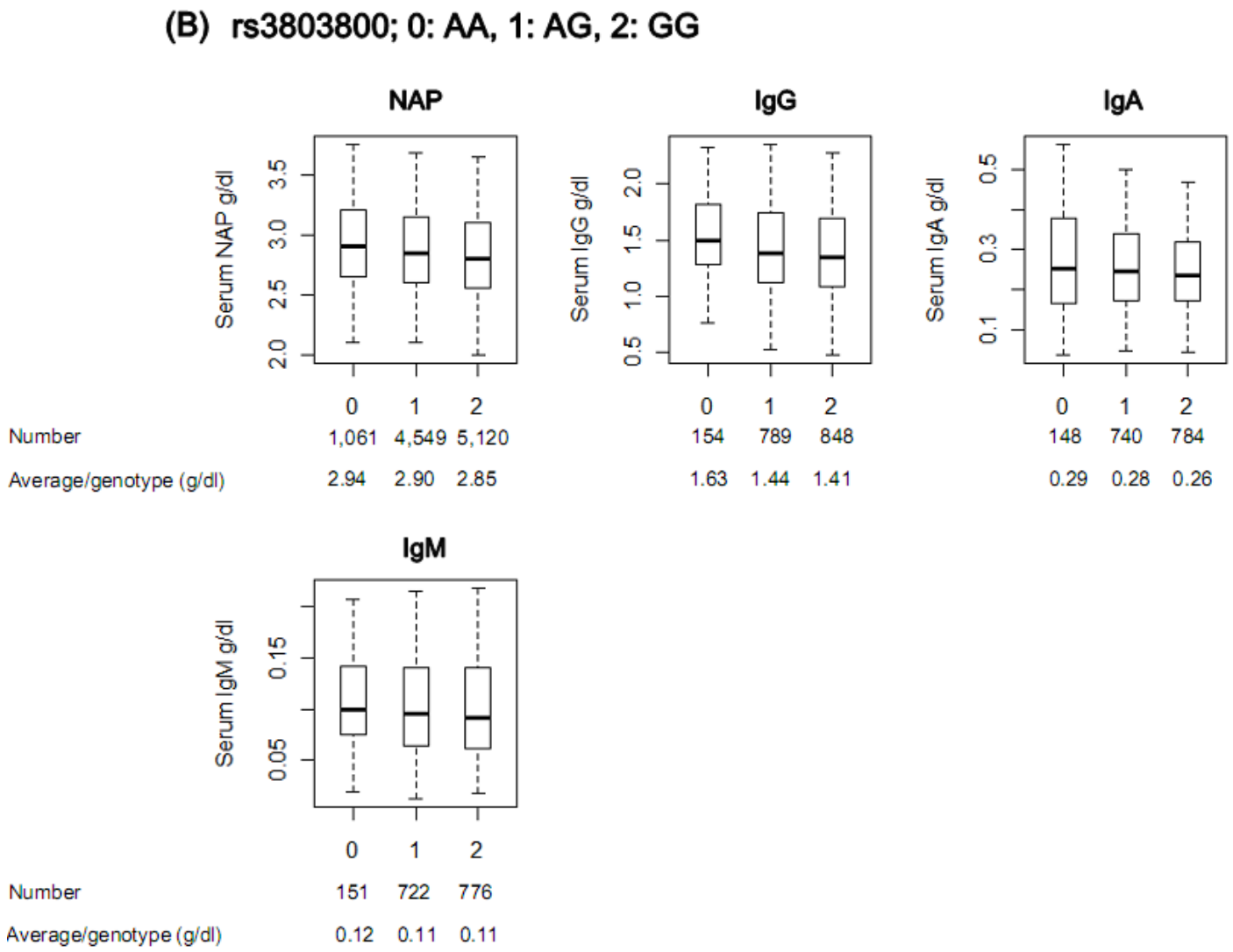


Figure 2-13 B: Relationship between the genotypes of the SNP rs3803800 and the levels of tested proteins.
For each box plot, the bold line indicates the median value which is the 50th quartile. The limits of each box are the 25th and 75th quartiles.

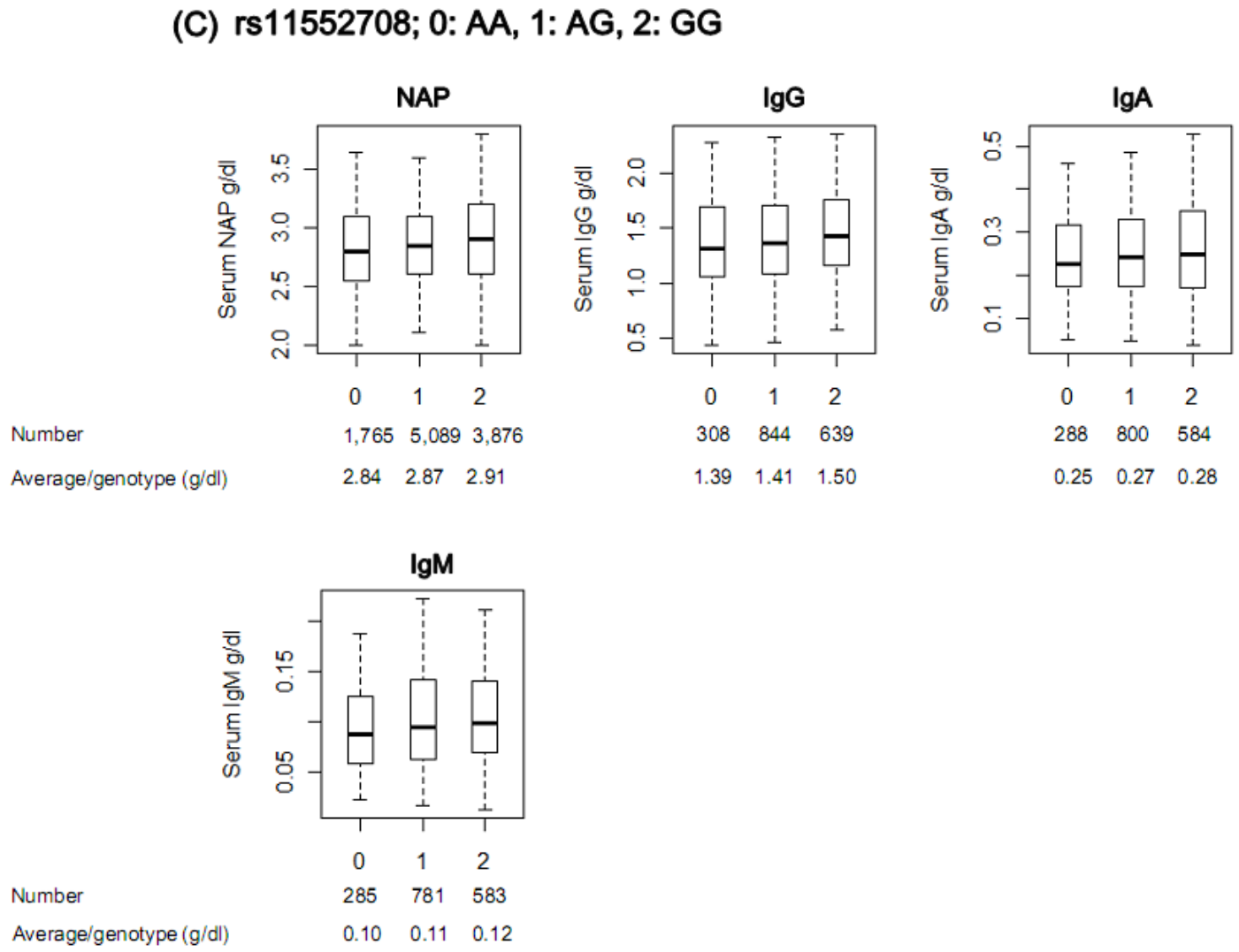


Figure 2-13 C: Relationship between the genotypes of the SNP rs11552708 and the levels of tested proteins.
For each box plot, the bold line indicates the median value which is the 50th quartile. The limits of each box are the 25th and 75th quartiles.

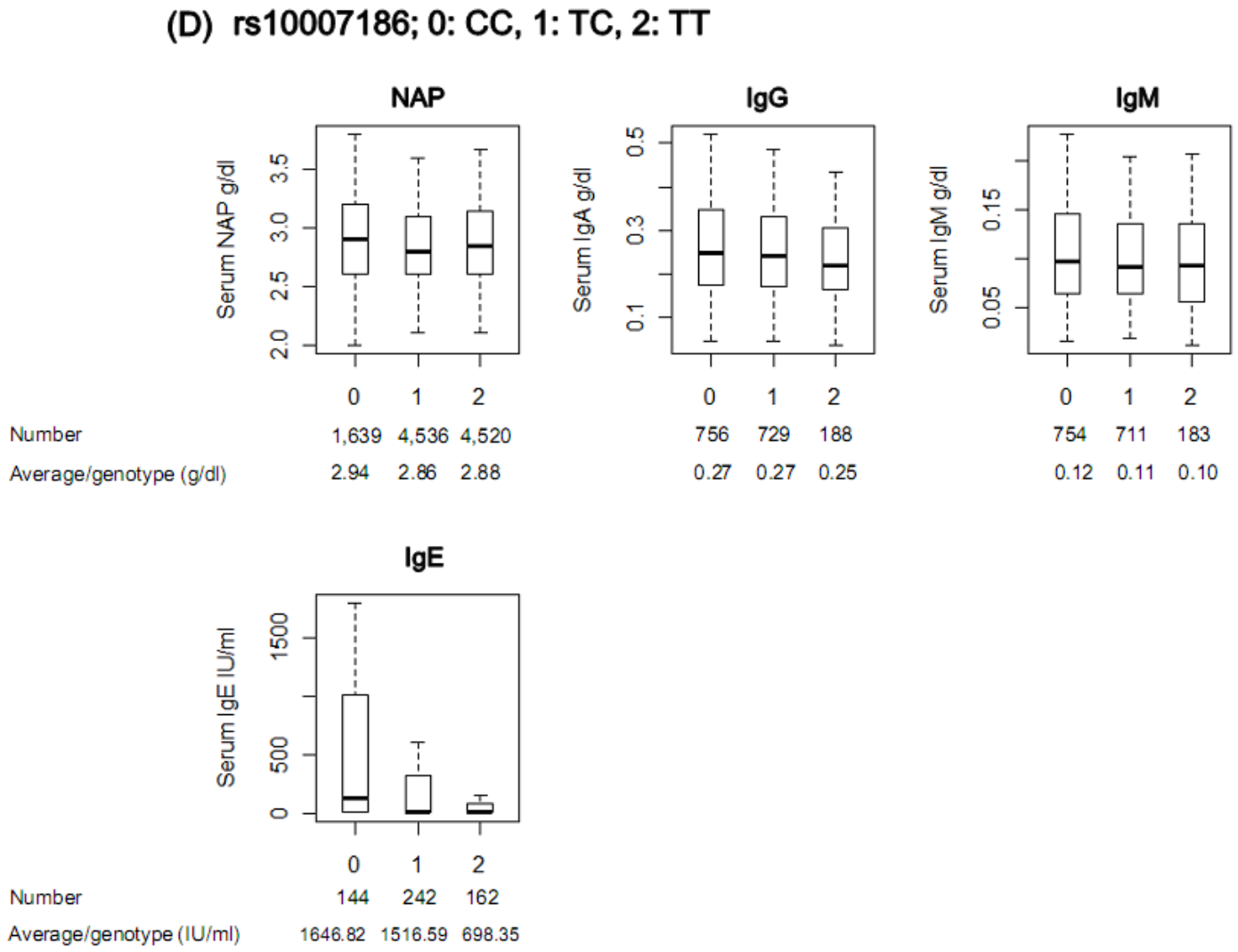


Figure 2-13 D: Relationship between the genotypes of the SNP rs10007186 and the levels of tested proteins.
For each box plot, the bold line indicates the median value which is the 50th quartile. The limits of each box are the 25th and 75th quartiles.

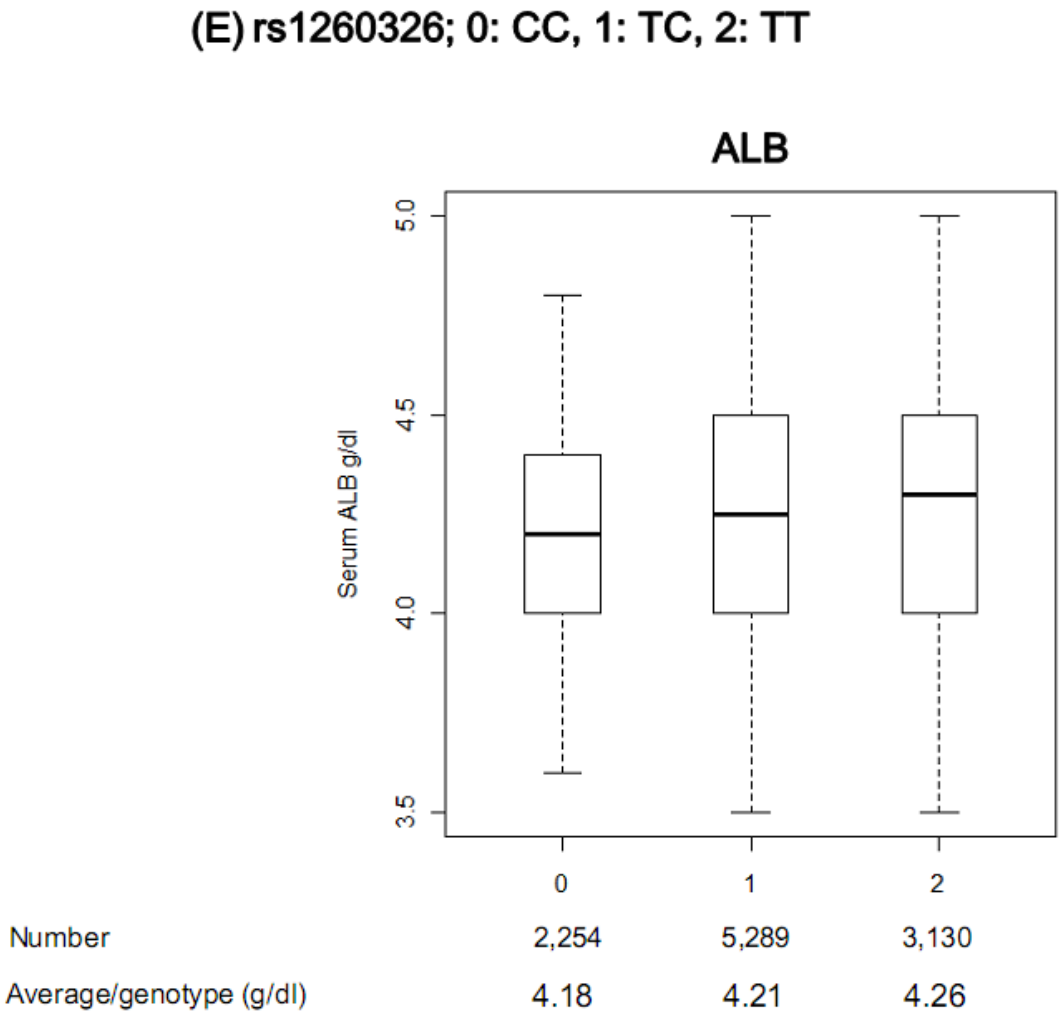


Figure 2-13 E: Relationship between the genotypes of the SNP rs1260326 and the levels of albumin.
For each box plot, the bold line indicates the median value which is the 50th quartile. The limits of each box are the 25th and 75th quartiles.

2.5 Conclusions

As a conclusion, this 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. My results should add novel insight toward understanding the genetic background contributing to the regulation of the serum levels of NAP and its major components.

Section 2.6 Acknowledgments

I want to acknowledge Dr. Yukinori Okada and the members of Laboratory of Laboratory for Statistical Analysis, Center for Genomic Medicine, Institute of Physical and Chemical Research (Center for Genomic Medicine, RIKEN), Kanagawa, Japan, for their assistance in the statistical analyses in this study.

Also, I want to thank Dr. Michiaki Kubo and the members of laboratory for Genotyping Development, Center for Genomic Medicine, RIKEN, Kanagawa, Japan, for their role in the genotyping work of the samples.

Section 2.7 Web resources

The URLs for the data presented in this paper are as follows:

The BioBank Japan Project, http://biobankjp.org/

PLIKN software, http://pngu.mgh.harvard.edu/~purcell/plink/

EIGENSTRAT software, http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm

The International HapMap Project, http://www.hapmap.org/

MACH and mach2qtl software, http://www.sph.umich.edu/csg/abecasis/MaCH/index.html

R statistical environment, http://www.r-project.org/

Haploview software, www.broad.mit.edu/mpg/haploview/

SNAP, http://www.broadinstitute.org/mpg/snap/ldsearch.php

Locus Zoom, http://csg.sph.umich.edu/locuszoom/

Section 2.8 References

1- Rahmioglu N, Andrew T, Cherkas L, Surdulescu G, Swaminathan R, et al. Epidemiology and genetic epidemiology of the liver function test proteins. ***PLoS One***, 4, e4435. (2009)

2- Melzer D, Perry J, Hernandez D, Corsi A, Stevens K, et al. A genome-wide association study identifies protein quantitative trait loci (pQTLs). ***PLoS Genetics***, 4, e1000072. (2008)

3- Weidinger S, Gieger C, Rodriguez E, Baurecht H, Mempel M, et al. Genome-wide scan on total serum IgE levels identifies FCER1A as novel susceptibility locus. ***PLoS Genetics***, 4, e1000166. (2008)

4- Sabatti C, Service SK, Hartikainen AL, Pouta A, Ripatti S, et al. Genome-wide association analysis of metabolic traits in a birth cohort from a founder population. ***Nature Genetics***, 41, 35-46. (2009)

5- Kamatani Y, Matsuda K, Okada Y, Kubo M, Hosono N, et al. Genome-wide association study of hematological and biochemical traits in a Japanese population. ***Nature Genetics***, 42, 210-215. (2010)

6- Bossen C, Schneider P. BAFF, APRIL and their receptors: structure, function and signaling. ***Seminars in Immunology***, 18, 263-275. (2006)

7- Mackay F, Kalled S. TNF ligands and receptors in autoimmunity: an update. ***Current Opinion in Immunology***, 14, 783-790. (2002)

8- Castigli E, Wilson S, Scott S, Dedeoglu F, Xu S, et al. TACI and BAFF-R mediate isotype switching in B cells. ***Journal of Experimental Medicine***, 201, 35-39. (2005)

9- Sakurai D, Hase H, Kanno Y, Kojima H, Okumura K, et al. TACI regulates IgA production by APRIL in collaboration with HSPG. ***Blood***, 109, 2961-2967. (2007)

10- Mackay F, Schneider P. TACI, an enigmatic BAFF/APRIL receptor, with new unappreciated biochemical and biological properties. ***Cytokine and Growth Factor Reviews***, 19, 263-276. (2008)

11- Nakamura Y. The BioBank Japan Project. ***Clinical Advances in Hematology and Oncology***, 5, 696-697. (2007)

12- Kamatani Y, Matsuda K, Ohishi T, et al. Identification of a significant association of a single nucleotide polymorphism in TNXB with systemic lupus erythematosus in a Japanese population. ***Journal of Human Genetics***, 53(1):64-73. (2008)

13- Obara W, Iida A, Suzuki Y, et al. Association of single-nucleotide polymorphisms in the polymeric immunoglobulin receptor gene with immunoglobulin A nephropathy (IgAN) in Japanese patients. ***Journal of Human Genetics***, 48(6):293-9. (2003)

14- Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. ***American Journal of Human Genetics***, 81, 559-575. (2007)

15- Olivier M. The Invader assay for SNP genotyping. ***Mutation Research***, 573, 103-110. (2005)

16- Okada Y, Hirota T, Kamatani Y, Takahashi A, Ohmiya H, et al. Identification of nine novel loci associated with white blood cell subtypes in a Japanese population. ***PLoS Genetics***, 7, e1002067. (2011)

17- Koyama T, Tsukamoto H, Masumoto K, Himeji D, Hayashi K, Harada M, Horiuchi T. A novel polymorphism of the human APRIL gene is associated with systemic lupus erythematosus. ***Rheumatology*** (Oxford), 42(8):980-5. (2003)

18- Kawasaki A, Tsuchiya N, Ohashi J, et al. Role of APRIL (TNFSF13) polymorphisms in the susceptibility to systemic lupus erythematosus in Japanese. ***Rheumatology*** (Oxford), 46(5):776-82. (2007)

19- Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, et al. A method and server for predicting damaging missense mutations. ***Nature Methods***, 7, 248-249. (2010)

20- Xu Z, Taylor JA. SNPinfo: integrating GWAS and candidate gene information into functional SNP selection for genetic association studies. ***Nucleic Acids Research***, 37, W600-605. (2009)

21- Hahne M, Kataoka T, Schröter M, Hofmann K, Irmler M, et al. APRIL, a new ligand of the tumor necrosis factor family, stimulates tumor cell growth. ***Journal of Experimental Medicine***, 188, 1185-1190. (1998)

22- López-Fraga M, Fernández R, Albar JP, Hahne M.Biologically active APRIL is secreted following intracellular processing in the Golgi apparatus by furin convertase. ***EMBO Reports***, 2, 945-951. (2001)

23- Castigli E, Scott S, Dedeoglu F, Bryce P, Jabara H, et al. Impaired IgA class switching in APRIL-deficient mice. ***Proceedings of National Academy of Science USA***, 101, 3903-3908. (2004)

24- von Bülow GU, van Deursen JM, Bram RJ.Regulation of the T-independent humoral response by TACI. ***Immunity***, 14, 573-582. (2001)

25- Castigli E, Wilson S, Garibyan L, Rachid R, Bonilla F, et al. TACI is mutant in common variable immunodeficiency and IgA deficiency. ***Nature Genetics***, 37, 829-834. (2005)

26- Gerke V, Moss SE. Annexins: from structure to function. ***Physiological Reviews***, 82, 331-371. (2002)

27- Larsson M, Majeed M, Ernst JD, Magnusson KE, Stendahl O, et al. Role of annexins in endocytosis of antigens in immature human dendritic cells. ***Immunology***, 92, 501-511. (1997)

28- Rosales JL, Ernst JD. Calcium-dependent neutrophil secretion: characterization and regulation by annexins. ***Journal of Immunology***, 159, 6195-6202. (1997)

29- Liu X, Paterson A, He N, et al. IL5RA and TNFRSF6B gene variants are associated with sporadic IgA nephropathy. Search Results. ***Journal of the American Society of Nephrology***, 19(5):1025-33. (2008)

30- Zhang J, Xu L, Liu G, Zhao M, Wang H. The level of serum secretory IgA of patients with IgA nephropathy is elevated and associated with pathological phenotypes. ***Nephrology Dialysis Transplantation***, 23(1):207-12. (2008)

31- Montenegro V, Monteiro R. Elevation of serum IgA in spondyloarthropathies and IgA nephropathy and its pathogenic role. ***Current Opinion in Rheumatology***, 11(4):265-72. (1999)

32- Ridker PM, Pare G, Parker A, Zee RY, Danik JS, et al. Loci related to metabolic-syndrome pathways including LEPR,HNF1A, IL6R, and GCKR associate with plasma C-reactive protein:

the Women's Genome Health Study. ***American Journal of Human Genetics***, 82, 1185-1192. (2008)

33- Kolz M, Johnson T, Sanna S, Teumer A, Vitart V, et al. Meta-analysis of 28,141 individuals identifies common variants within five new loci that influence uric acid concentrations. ***PLoS Genetics***, 5, e1000504. (2009)

34- Orho-Melander M, Melander O, Guiducci C, Perez-Martinez P, Corella D, et al. Common missense variant in the glucokinase regulatory protein gene is associated with increased plasma triglyceride and C-reactive protein but lower fasting glucose concentrations. ***Diabetes***, 57, 3112-3121. (2008)

Table 2-1: Characteristics of the cohorts of the GWAS of major serum protein components.

Disease	No.	Age [a]	Female %	BMI [a]	Smokers (%)	Drinkers (%)
Colorectal cancer	1,349	60.9 ± 11.0	40.2	22.5 ± 3.4	42.7	32.3
Breast cancer	1,346	54.8 ± 9.10	100	22.8 ± 3.6	17.0	14.7
Prostate cancer	1,096	69.0 ± 7.40	0.00	23.5 ± 2.8	39.2	42.7
Lung cancer	1,183	65.9 ± 9.30	36.6	22.3 ± 3.3	48.1	31.8
Gastric cancer	1,307	65.1 ± 8.60	19.7	21.3 ± 3.1	54.3	36.0
Diabetes Mellitus	758	65.3 ± 10.4	38.8	24.1 ± 3.7	40.6	23.4
Peripheral Artery Disease	197	70.7 ± 9.20	12.7	22.8 ± 3.2	72.1	28.9
Atrial Fibrillation	250	67.9 ± 10.1	27.6	24.1 ± 3.6	44.8	32.8
Ischemic stroke	997	67.6 ± 8.50	34.1	23.8 ± 3.3	39.7	25.1
Myocardial infarction	574	59.6 ± 10.4	12.7	24.4 ± 3.3	59.6	26.3

[a] Age and body mass index (BMI) are shown as mean ± standard deviation.

All cohorts were genotyped using Illumina Human610-Quad BeadChip.

Table 2-2: Characteristics of the proteins analyzed in the GWAS

	TP		ALB		NAP					
	GWAS	Replication	GWAS	Replication	GWAS	Replication	IgG *	IgA *	IgM *	IgE *
No.	9,090	1,626	9,103	1,607	9,077	1,629	1,794	1,675	1,649	549
M ± S.D [a]	7.10 ± 0.50	7.06 ± 0.73	4.25 ± 0.35	4.00 ± 0.51	2.85 ± 0.42	3.07 ± 0.57	1.44 ± 0.61	0.27 ± 0.15	0.11 ± 0.07	1306.54 ± 5598.06
Age [b]	69.52 ± 10.44	59.52 ± 15.43	69.52 ± 10.44	59.54 ± 15.39	69.51 ± 10.44	59.48 ± 15.52	59.70 ± 15.46	59.38 ± 15.73	59.42 ± 15.57	62.54 ± 18.61
Female %	37.45	45.08	37.41	45.12	37.46	45.12	55.30	54.57	54.88	63.93
BMI [b]	22.91 ± 3.45	23.31 ± 5.67	22.91 ± 3.45	23.34 ± 5.69	22.91 ± 3.45	23.29 ± 5.67	23.17 ± 5.00	23.20 ± 5.09	23.19 ± 5.07	22.73 ± 4.22
Smokers %	42.11	51.91	42.11	52.15	42.05	51.81	51.90	51.82	52.27	48.63
Drinkers %	29.37	51.97	29.37	52.08	29.40	51.81	51.00	50.81	50.82	41.35

[a] M ± S.D: mean value ± standard deviation of each protein is indicated in g/dl except for IgE, which is indicated as IU/ml.

[b] Age and body mass index (BMI) are shown as mean values ± standard deviation.

[*] Log-transformed values were applied in the analysis.

Abbreviations: GWAS: genome-wide association study, TP: total protein, ALB: albumin, NAP: non-albumin protein.

Table 2-3: SNPs showed significant or suggestive associations with each trait ($P < 1 \times 10^{-6}$)

Trait	SNP	Chr.	Position	A1	A2	A2 Freq	Gene	Effect [a]	S.E	*P*
TP	**rs4985726**	17	16804363	C	G	0.375	*TNFRSF13B*	-0.108	0.015	2.83×10^{-12}
	rs4792800	17	16785892	A	G	0.37	*TNFRSF13B*	-0.106	0.015	5.38×10^{-12}
	rs4561508	17	16789475	C	T	0.371	*TNFRSF13B*	-0.104	0.015	6.28×10^{-12}
	rs7226097	17	16809286	C	T	0.563	*TNFRSF13B*	-0.103	0.016	3.98×10^{-10}
	rs4273077	17	16789864	A	G	0.455	*TNFRSF13B*	-0.092	0.015	4.34×10^{-10}
	rs1052335	9	6320380	A	C	0.129	*TPD52L3*	-0.112	0.022	2.86×10^{-7}
	rs4985700	17	16806800	A	C	0.261	*TNFRSF13B*	0.086	0.017	4.73×10^{-7}
	rs560059	9	6437196	A	G	0.138	*UHRF2*	-0.112	0.022	4.89×10^{-7}
	rs2280410	19	54691821	A	G	0.845	*RPS11*	0.101	0.021	5.74×10^{-7}
Alb	**rs1260326**	2	27584444	C	T	0.555	*GCKR*	-0.082	0.015	3.42×10^{-8}
	rs3817588	2	27584716	C	T	0.698	*GCKR*	-0.089	0.016	4.10×10^{-8}
	rs780092	2	27596658	A	G	0.299	*GCKR*	0.0835	0.016	2.84×10^{-7}
	rs814295	2	27596719	A	G	0.299	*GCKR*	0.0832	0.016	3.11×10^{-7}
N-Alb P	**rs4985726**	17	16804363	C	G	0.375	*TNFRSF13B*	-0.148	0.015	2.38×10^{-22}
	rs4561508	17	16789475	C	T	0.371	*TNFRSF13B*	-0.143	0.015	6.36×10^{-22}
	rs4792800	17	16785892	A	G	0.37	*TNFRSF13B*	-0.145	0.015	8.48×10^{-22}
	rs4273077	17	16789864	A	G	0.455	*TNFRSF13B*	-0.12	0.014	9.87×10^{-17}
	rs7226097	17	16809286	C	T	0.563	*TNFRSF13B*	-0.12	0.016	1.34×10^{-13}
	rs3803800	17	7403693	A	G	0.688	*TNFSF13*	0.107	0.015	4.42×10^{-12}
	rs10007186	4	79808069	C	T	0.693	*ANXA3*	-0.095	0.016	3.27×10^{-9}
	rs11552708	17	7403279	A	G	0.597	*TNFSF13*	-0.086	0.015	3.81×10^{-9}
	rs16961828	17	16760949	C	G	0.618	*TNFRSF13B*	0.094	0.016	7.41×10^{-9}
	rs4792795	17	16766741	A	G	0.38	*TNFRSF13B*	-0.09	0.016	2.24×10^{-8}
	rs3751991	17	16776011	A	C	0.616	*TNFRSF13B*	0.088	0.016	2.32×10^{-8}
	rs3752005	17	16769438	A	T	0.384	*TNFRSF13B*	-0.089	0.016	2.32×10^{-8}
	rs11654088	17	16790538	C	G	0.79	*TNFRSF13B*	-0.1	0.019	9.93×10^{-8}
	rs6739695	2	100226054	A	G	0.587	*AFF3*	0.078	0.015	$1.0^{1} \times 10^{-7}$
	rs10865036	2	100227027	C	T	0.413	*AFF3*	-0.078	0.015	1.01×10^{-7}
	rs2309811	2	100228791	A	G	0.587	*AFF3*	0.078	0.015	1.01×10^{-7}

N-Alb P	rs17017652	4	91956523	A	C	0.606	*FAM190A*	0.082	0.016	1.07×10^{-7}
	rs4851274	2	100237294	C	T	0.414	*AFF3*	-0.078	0.015	1.07×10^{-7}
	rs7597861	2	100219619	A	G	0.413	*AFF3*	-0.078	0.015	1.17×10^{-7}
	rs6737502	2	100229335	C	T	0.587	*AFF3*	0.078	0.015	1.22×10^{-7}
	rs6542921	2	100230930	A	C	0.413	*AFF3*	-0.078	0.015	1.26×10^{-7}
	rs12474386	2	100223449	A	G	0.417	*AFF3*	-0.076	0.015	2.28×10^{-7}
	rs9901675	17	7425536	A	G	0.938	*CD68*	0.152	0.03	2.69×10^{-7}
	rs2274892	17	16792752	G	T	0.372	*TNFRSF13B*	0.074	0.015	5.61×10^{-7}
	rs3818716	17	16792559	C	T	0.628	*TNFRSF13B*	-0.074	0.015	5.63×10^{-7}
	rs4851273	2	100214753	A	G	0.544	*AFF3*	0.072	0.014	6.46×10^{-7}
	rs6542920	2	100211520	A	G	0.544	*AFF3*	0.071	0.014	7.94×10^{-7}

[a] The effect of allele 1 on the normalized values of the indicated trait.

SNPs in bold were selected for replication.

Abbreviations: A1/A2: alleles1 and 2, A2: allele 2, A2 Freq: the frequency of allele 2 estimated by the study, S.E: standard error.

Table 2-4: Summary results of the GWAS and the replication study of TP, ALB, and NAP.

Trait	SNP	Chr: Position	Nearest gene	A1/ A2[a]	MAF	GWAS Effect[b] (s.e)	GWAS *P*[c]	Replication Effect[b] (s.e)	Replication *P*[c]	Meta-analysis Effect[b] (s.e)	Meta-analysis *P*[c]	% variance explained
TP	rs4985726*	17:16804363	*TNFRSF13B*	C/G	0.375	0.108 (0.015)	2.8×10^{-12}	0.100 (0.030)	0.0010	0.107 (0.0138)	1.2×10^{-14}	0.53
ALB	rs1260326	2: 27584444	*GCKR*	T/C	0.445	-0.082 (0.015)	3.4×10^{-8}	-0.070 (0.032)	0.029	-0.080 (0.014)	3.1×10^{-9}	0.32
NAP	rs4985726*	17:16804363	*TNFRSF13B*	C/G	0.375	0.148 (0.015)	2.4×10^{-22}	0.090 (0.028)	0.0013	0.135 (0.013)	7.1×10^{-24}	1.03
	rs3803800	17:7403693	*TNFSF13*	G/A	0.311	0.108 (0.015)	1.8×10^{-12}	0.090 (0.029)	0.0022	0.104 (0.013)	7.2×10^{-15}	0.53
	rs11552708	17:7403279	*TNFSF13*	G/A	0.401	-0.084 (0.015)	7.0×10^{-9}	-0.070 (0.027)	0.0091	-0.081 (0.013)	7.5×10^{-10}	0.36
	rs10007186*	4:79808069	*ANXA3*	T/C	0.307	0.095 (0.016)	3.3×10^{-9}	0.053 (0.029)	0.065	0.085 (0.014)	1.3×10^{-9}	0.38

[a] A1/A2: major/minor alleles.

[b] The effect of the minor allele on the normalized values.

[c] For the GWAS and replication analysis, *P*-values were obtained by linear regression test model, for the Meta analysis by inverse-variance method.

*SNPs obtained by whole-imputation analysis.

Abbreviations: GWAS: genome-wide association study, MAF: minor allele frequency, TP: total protein, ALB: albumin, NAP: non-albumin protein, s.e: standard error.

Table 2-5: Haplotype analysis of rs3803800 and rs11552708 in association with NAP

	Haplotype				
	rs3803800	rs11552708	Frequency	Effect	S.E
1	A	G	0.310	0.124	0.017
2	G	G	0.289	0.042	0.018
3 [a]	A	A	0.001	0.404	0.234
4 [b]	G	A	0.399		

[a] Rare haplotype.

[b] The reference haplotype for the analysis.

S.E: standard error.

Table 2-6: Haplotype analysis of rs1260326 and rs3817588 in association with ALB

	Haplotype				
	rs1260326	rs3817588	Frequency	Effect	S.E
1	C	C	0.293	-0.101	0.017
2	C	T	0.150	-0.048	0.022
3 [a]	T	C	0.007	-0.033	0.093
4 [b]	T	T	0.550		

[a] Rare haplotype.

[b] The reference haplotype for the analysis.

S.E: standard error.

Table 2-7: Association of the SNPs in the GWAS of the NAP with immunoglobulin isotypes

		IgG			IgA			IgM			IgE		
SNP	Gene	Effect [a] (s.e)	*P* [b]	%EV	Effect [a] (s.e)	SNP	Gene	Effect [a] (s.e)	*P* [b]	%EV	Effect [a] (s.e)	*P* [b]	%EV
rs4985726	*TNFRSF13B*	0.071 (0.022)	1.4×10^{-3}	0.51	0.049 (0.030)	0.099	–	-0.090 (0.032)	5.9×10^{-3}	0.40	0.039 (0.064)	0.54	–
rs3803800	*TNFSF13*	-0.074 (0.024)	2.2×10^{-3}	0.47	-0.086 (0.031)	6.2×10^{-3}	0.39	-0.082 (0.034)	0.018	0.29	-0.117 (0.067)	0.080	–
rs11552708	*TNFSF13*	0.067 (0.022)	2.3×10^{-3}	0.46	0.072 (0.029)	0.013	0.31	0.078 (0.032)	0.014	0.31	0.059 (0.060)	0.33	–
rs10007186	*ANXA3*	-0.018 (0.022)	0.42	–	-0.063 (0.030)	0.036	0.20	-0.078 (0.033)	0.019	0.27	0.200 (0.057)	4.9×10^{-4}	2.02

[a] The effect of the minor alleles on the standardized values.

[b] *P*-values for the associations of SNPs with each normalized immunoglobulin isotype obtained by using a linear regression model.

Abbreviations: s.e: standard error, %EV: percentage of the explanatory variance.

Table 2-8: Association of top SNPs in Non-albumin Protein GWAS with SLE (N = 193)

		Candidate	*Case*			*Control*			*MAF*			
SNP	*Chr.*	*Gene*	11	12	22	11	12	22	*Case*	*control*	*P-value*	*O.R (95% C.I)*
rs4273077	17	*TNFRSF13B*	46	102	45	293	454	185	0.497	0.442	0.03613	1.25 (1.00 - 1.56)
rs3803800	17	*TNFSF13B*	84	79	30	431	399	104	0.360	0.325	0.08492	1.17 (0.93 - 1.47)
rs11552708	17	*TNFSF13B*	28	78	87	160	449	324	0.347	0.412	0.00655	1.32 (1.05 - 1.66)

C.I: confidence intervals, MAF: minor allele frequency, OR: odds ratio.

Table 2-9: Association of top SNPs in Non-albumin Protein GWAS with IgA nephropathy (N = 370)

		Candidate	*Case*			*Control*			*MAF*			
SNP	*Chr.*	*Gene*	11	12	22	11	12	22	*Case*	*control*	*P-value*	*O.R (95% C.I)*
rs4273077	17	*TNFRSF13B*	91	187	88	293	454	185	0.496	0.442	0.0132	1.24 (1.05 - 1.47)
rs3803800	17	*TNFSF13B*	46	177	144	104	399	431	0.367	0.325	0.0239	1.20 (1.01 - 1.42)
rs11552708	17	*TNFSF13B*	41	172	157	160	449	324	0.343	0.412	0.0012	1.34 (1.12 - 1.60)

C.I: confidence intervals, MAF: minor allele frequency, OR: odds ratio.

Chapter 3: Functional analysis of CADM1 in lung adenocarcinoma

Section 3.1.1 Introduction

CADM1 (Cell Adhesion Molecule 1) encodes an immunoglobulin-superfamily adhesion molecule with three immunoglobulin-like loops. It was first described as a tumor suppressor gene by functional complementation of a human lung cancer cell line, A549, through suppression of tumorigenicity in nude mice **(1)** . Later, CADM1 has been also found to be involved in cell-cell adhesion in calcium-independent manner as being recruited in in the lateral membrane of cells, and then being involved in hemophilic or heterophilic t*rans*-interaction between its molecules **(2)**. Mika Sakurai-Yageta et al. described in 2009 that CADM1 is involved in epithelial cell structure as suppression of CADM1 in HEK293 by siRNA results in loosing of epithelial-like structure and changing to flat morphology of cells with immature cell adhesion patterns **(3)**.

The intracellular (cytoplasmic) domain of CADM1 harbors protein 4.1 and class II PDZ domain binding motifs, both been involved in protein-interaction modules **(1)**. Protein 4.1 binding motif, as the name implies, can bind to members of protein 4.1 family and links CADM1 to the actin cytoskeleton **(4)**. The other motif is class II PDZ-binding motif, through which CADM1 interacts with membrane-associated guanylate kinase homologs (MAGUKs) through their PDZ domains **(5, 6)**.

Interestingly, despite being a tumor suppressor in various cancers **(7)**, CADM1 was reported to be upregulated in primary adult T-cell leukemia **(8)** and small cell lung cancer **(9)**, indicating CADM1 might work also as oncogene through unknown mechanism.

Lung cancer is a leading cause of death in many developed countries including Japan **(10)**, and adenocarcinoma is the most histological subtype of the disease **(11)**. In 47 primary lung adenocarcinoma, CADM1 was found to be expressed in normal pulmonary epithelia, and lost or decreased in primary pulmonary adenocarcinoma **(12)**. In other study in 48 non small-cell lung tumors, 21 showed CADM1 promoter hypermethylation that correlated to the advanced tumor **(13)**. Furthermore, loss of CADM1 expression was found to affect survival and clinco-pathological parameters in lung adenocarcinoma (advanced stage, lymph node involvement, lymphatic penetration, and vascular invasion) **(14)**. More examination of 93 primary lung adenocarcinoma tumors indicated loss of CADM1 expression in 60 tumors, more frequently in male, and also affected the survival and invasiveness in male **(15)**. In addition, CADM1 loss was reported to be correlated to heavy smoking and significantly associated with shorter survival rates when researchers examined another 103 tumor **(16)**. However, the mechanism by which CADM1 prevents cancer progression in lung adenocarcinoma remains a debated question. In this study, I tried to investigate the possible interacting proteins with CADM1 in lung adenocarcinoma utilizing microarray data in lung adenocarcinoma cell lines.

Section 3.1.2 Objective

The objective of this study is to perform functional analyses of CADM1 in lung adenocarcinoma through searching for interacting protein(s) utilizing microarray data in 41 adenocarcinoma cell lines.

Section 3.2 Methods

3.2.1 Reagents, plasmids, and antibodies

All lung adenocarcinoma cell lines were cultured in RPMI 1640 medium, supplied with 10% fetal bovine serum and 1% antibiotics. HEK293 cells were cultured in DMEM medium with the same supplies. All cell lines were kept in 5% CO_2 atmosphere.

RapGEF2 expression vector (FXC00500) was obtained from Kazusa DNA Research Institute. It was constructed using pF1KM T7 Flexi vector, and it is a tag-free **(17)**. After receiving the vector, I was proliferated it by transformation into DH5α competent bacterial cells, then plated in petri-dish supplied with ampicillin. Positive colonies were sub-cultured in 100 ml LB medium supplied with ampicillin O/N, and the plasmid was then extracted using Genopure Plasmid Maxi kit (Roche). The pCMV2FLAG-RapGEF2 vector was a generous gift from Prof. Lawrence A. Quilliam, Indiana University School of Medicine, USA. Un-tagged CADM1 and HA-CADM1 are pcDNA3.1/Hygro (+) full length plasmids and was prepared previously in our laboratory.

The antibodies used in this study and their concentrations are summarized in **Table 3-1**.

3.2.2 Transfection

1 x 10^6 HEK293 cells were seeded in 100 mm dishes to be 50-70 % confluent. Transfection was the performed using Lipofectamine LTX reagent (Invitrogen).

3.2.3 Western blot analysis

After growing to be confluent (usually within 48 hours), cells were collected on ice with chilled PBS and centrifuged for 5000 rpm for 5min. I then lysed the cells using a lysis buffer (50 mM Tris P^H 7.8, 150 mM NaCl, 1mM EDTA, and 1% Triton X-100). The cells were incubated on ice for 10min, and centrifuged in 15,000 rpm for 15min. The supernatants were collected into new tubes as the whole cell lysates. I determined the protein concentration of each samples using Bio-Rad Protein assay according to manufacture protocol. Samples were supplied with x3 sample buffer, and 7.5 μg of samples were applied to SDS-PAGE gel (7.5 %), and transferred into PVDF membrane using the semi-dry method. The membrane were blocked in 5% skimmed milk solution for 1 hour at RT, then incubated with the specific antibodies .After 3 washes (each 10 minutes), the membrane was incubated with second antibodies (anti-mouse or anti-rabbit) in 1:20,000 concentrations. Proteins were visualized using enhanced chemiluminescence with a protein-blotting detection (GE Healthcare).

3.2.4 Immunoprecepitation

For Immunoprecepitation, cell lysates were prepared as in western blot experiments. Lysates were first pre-cleared by incubation with protein A-Sepharose or protein G-Sepharose (GE Healthcare) for 3h at 4°C. The pre-cleared lysates were then incubated with a rabbit anti-CADM1 (C18) pAb, or a mouse anti-RapGEF2 mAb for 6 hours at 4°C. A rabbit IgG was used as a negative control. The protein-antibody conjugates were precipitated with 30 μl of protein A-Sepharose (for anti-rabbit antibodies) or protein G-

Sepharose (for anti-mouse antibodies) another 6h at 4°C. Immunoprecipitates were rinsed three times with the lysis buffer, and X2 sample buffer was added before lysates incubated at 98°C for 5min. 20 μl of the processed lysates (together with their unprocessed whole cell lysates) was fractionated in a 7.5 % SDS-PAGE gel and the process continued as in the western blot experiments.

3.2.5 Immunocytochemistry

HEK293 T Cells were cultured in coverslips and transfected with pCMV2FLAG-RapGEF2 and HA-CADM1 vectors. After 48h, the medium was removed and cells washed 2 times with PBS (−). The cells were then fixed with PBS (−) containing 4% paraformaldehyde for 30min at 4°C, and rendered permeable with PBS (−) containing 0.1% Triton X-100 at room temperature for 2min. Subsequently, the cells were covered with PBS (−) containing 3% bovine serum albumin for 1h at room temperature to block nonspecific hybridization and were then incubated with anti-FLAG, or anti-HA antibodies at 4°C O/N. After washing 3 times with PBS (−), cells were stained by an Alexa Fluor 488– conjugated anti-rat and Alexa Fluor 568– conjugated anti-mouse secondary antibodies (Life Technologies) at a 1:1000 dilution ratio. Nuclei were counterstained with 4′, 6′-diamidine-2′-phenylindole dihydrochloride (DAPI). Fluorescent images were obtained under Axio Observer. D1 microscope (ZEISS). Pictures were seen by Axio vision 4.7.1 program.

3.2.6 Patients

I used 7 different TMAs which were made to accommodate primary lung adenocarcinoma tissue core sections (n=168), which were collected from patients (n=166)

whom had undergone surgical resection at the University of Tokyo Hospital between June 2005 and September 2008. Informed consent was obtained from all the patients, and the study was approved by the Institutional Ethics Review Committee. Of 168 core sections, 2 were from the same patients and 12 were missing from TMA sections, so I performed immunohistochemical analysis of 154 cases. The patients (n=154) included 84 males and 70 females, ranging in age from 34 to 86 years (average 66.8 years). Each case was reassigned for tumor, node, metastasis (TNM) classification and pathological stage on the basis of the new IASLC staging system; 108 were Stage I (69 Stage IA, 39 Stage IB), 18 were Stage II (14 Stage IIA, 4 Stage IIB), 23 were Stage III (18 Stage IIIA, 5 Stage IIIB), and 1 was stage IV. Stages of 4 cases were unknown (3 were more than Stage I). Each case was classified by pathologist (Daisuke MATSUBARA) as follows, depending on predominant histopathological subtype on the TMA sections; lepidic growth (n=41), papillary adenocarcinoma (n=57), acinar adenocarcinoma (n=26), solid adenocarcinoma with mucin (n=28), and invasive mucinous adenocarcinoma (n=2). These cases were also graded into three grades; low-grade containing lepidic growth (n=41), intermediate-grade containing acinar and papillary adenocarcinomas (n=83), and high-grade containing solid adenocarcinoma with mucin and invasive mucinous adenocarcinoma (n=30), referring to the histopathological grading by Yoshizawa et al (Yoshizawa, 2011 modern pathol) with a slight modification.

3.2.7 Immunohistochemistry

This experiment was done with a kind help of Ms. Ando Tomoko at Aqua lab for core pathology in the IMSUT. The sections were first treated with 0.3% hydrogen peroxide in methanol for 30 min to block endogenous peroxidase activity. Then sections were either autoclaved in 10 mM citrate buffer (pH 6.0) for 10 min at 120°C (for RapGEF2 and CADM1). Sections were then preincubated with 10% normal horse serum in phosphate-buffered saline, incubated with a rabbit polyclonal antibody against RAPGEF2 (Abcam) at a dilution of 1:250, and a rabbit polyclonal antibody against CADM1 (C18) at a dilution of 1:500. The linked primary antibody was detected with DAKO Envision system according to manufacturer's instructions. 3, 3'-diaminobenzidine tetrahydrochloride (DAB) was used as a chromogen, whereas hematoxylin was used as a light counterstain. No significant staining was observed in the negative controls, which were prepared using the same class of mouse immunoglobulin at the same concentration.

3.2.8 Evaluation of IHC

The evaluation of the immunohistochemical staining was performed by a pathologist (Dr. Daisuke MATSUBARA at the Molecular Pathology lab in IMSUT), through light microscopic observation and without knowledge of the clinical data from each patient.

CADM1 and RapGEF2 were evaluated as follows: Cytoplasmic staining was assessed for RAPGEF2, while membranous staining was assessed for CADM1. Intensity was quantified as follows: 1+, weak staining (detection required high magnification); 2+, moderate staining (detected readily at medium magnification); 3+, strong staining (detected

readily at low magnification). The percentages of positive cells were scored into five categories: 0, 0%; 1, 1–25%; 2, 26–50%; 3, 51–75%; 4, 76–100%. The product of the intensity and percentage scores was used as the final staining score. The final score for the staining of CADM1 and RAPGEF2 was defined as low-level expression (final staining score <6) and high-level expression (final staining score =>6).

3.2.9 Statistics

The chi-squared test was used to evaluate the clinicopathologic correlations except histopathological grades. Mann-Whitney U-test was used to evaluate the histopathological grades, with each grade scored (low-grade 1. intermediate-grade 2. and high-grade 3.) Disease Free-Survival curves were generated using the Kaplan–Meyer method and differences in survival were analyzed by the Wilcoxon method. The results were considered significant if the p value was less than 0.05. All statistical calculations were performed using the StatView computer program (Abacus Concepts, Berkeley, USA).

3.2.10 Rap1 activation assay

Active Rap1 pull-down detection kit was used for this assay (Thermo Scientific) HEK293 cells were transfected with CADM1 and RapGEF2 vectors. Lysates were then prepared using a lysis buffer provided in the kit. 500 μg of the lysates were treated in-vitro by either GTPγS (positive control) or GDP (negative control). 100 μl of 50% glutathione resin were added to each sample, followed by 20 μg GST-RalGDS-RBD and immediately 500 μg of lysates including positive and negative controls. Samples were then incubated at

4° C O/N with gentle rocking, before washed 3 times. 50 μl of 2X SDS sample buffer were then added and collected by centrifugation at 6,000 g for 30s. The experiment was then preceded as described in the western blot with the following exceptions: blocking step was performed using 3% BSA, 15 % SDS-PAGE gel was used to fractionate the samples, and 5% BSA solution was used to prepared 1st and 2nd antibodies.

Section 3.3 Results

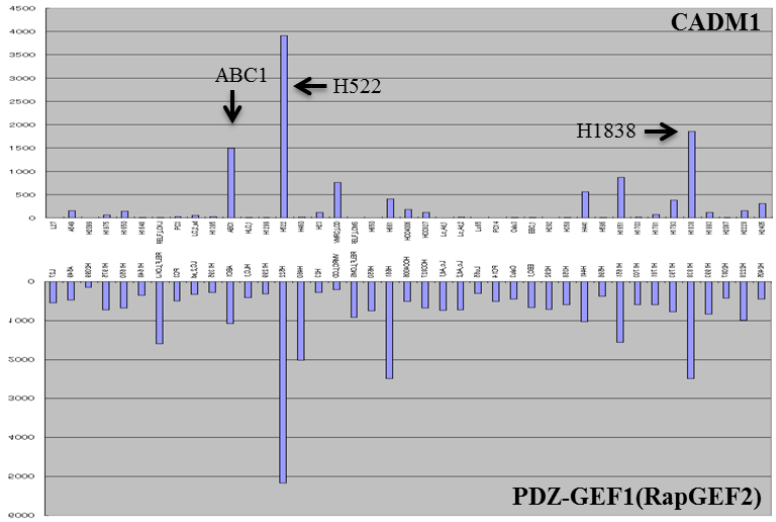
3.3.1 RapGEF2 indicates expression correlation with CADM1

I utilized the microarray data of 41 lung adenocarcinoma cell lines to predict possible CADM1 interacting proteins depending on the expression patterns. I used a cut-off of correlation coefficient of 0.6 to predict genes showing similar expression patterns with CADM1. Among the proteins detected, RapGEF2 (PDZ-GEF1) is found to have the strongest correlation with CADM1 expression (**Fig. 3-1 A**). RapGEF2 is a guanine exchange factor for the small GTPase Rap1. Interestingly, RapGEF2 possess a PDZ domain which by similarity to TIAM-1 (**Fig. 3-1 B**), which reported recently as CADM1 interacting protein **(18)**, is a type II PDZ domain, therefore it can potentially interacts with CADM1.

3.3.2 RapGEF2 and CADM1 show similar expression patterns in lung adenocarcinoma cell lines

To confirm the microarray data, I performed western analysis experiment using 12 lung adenocarcinoma cell lines to detect the expression levels of RapGEF2 and CADM1. GAPDH was used as a control. In concordance with the microarray data, both proteins (RapGEF2 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: H522 and H1838 cells indicate highest expression for both proteins. Taken together, this data confirms the possibility of interaction of these proteins in lung adenocarcinoma.

A



B

```
RapGEF2 411 IFVDSVDSGSKATEAGLKRGDQILEVNGQNFENIQLSKAMEILRNNTHLSITVKT 465
Tiam-1  873 LYVNSVKETGLASKKGLKAGDEILEINNRAADALNSSMLKDFLSQPS-LGLLVRT 926
```

Fig 3-1: Genome-wide microarray data in 41 lung adenocarcinomas.

(A) This microarray data was utilized to compare the expression between CADM1 and other genes using a cut-off point of correlation coefficient > 0.5. Among candidates, PDZ- GEF1 (RapGEF2) indicated a strong co-expression with CADM1.

(B) Comparison between the PDZ domains of RapGEF2 and Tiam-1. Red color shows the exact amino acids. Green color indicates amino acid from the same group and function. The two PDZ domains are 33% identical and 58% similar, which indicates PDZ domain of RapGEF2 is probably type 2.

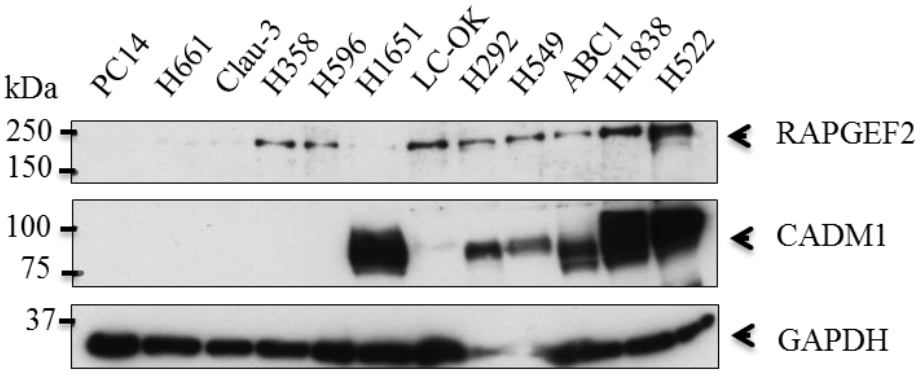


Fig 3-2: Western blot analysis of 12 adenocarcinoma cell lines.
7.5 μg of the total cell lysates were applied to 7.5% SDS-PAGE.

3.3.3 Interaction and localization of CADM1 and RapGEF2

To test the interaction between CADM1 and RapGEF2, I performed immunoprecepitation assays. First, I tested the interaction between the endogenous proteins in H522 cells as this cell line showed the highest expression of both proteins in microarray data (**Fig. 3-1 A**) and western blot results (**Fig. 3-2**). As a result, I couldn't get an evidence for interaction (**Fig. 3-3 A left**) of endogenous proteins. I speculated that these proteins may be temporally interacting under certain conditions or through collecting another interacting protein. I then performed same experiment after overexpressing RapGEF2 alone (**Fig. 3-3 A right**) or overexpressing both proteins (**Fig. 3-3 B**) in HEK293 cells. When both proteins were overexpressed, the interaction between these two proteins could be detected as antibodies against both proteins can notably precipitate both of them as in **Fig. 4-3 B**.

I also tested the localization of both proteins using immunocytochemistry. When I tried to detect the localization of endogenous proteins, the anti-RapGEF2 antibody gave a diffused signal; the major portion of the signal was located in the nucleus (data no shown). As RapGEF2 is suspected to be located in the cell membrane, I suspected this antibody is not suitable for ICC experiment. I then used the Flag-tagged RapGEF2 together with HA-tagged CADM1 to detect the localization of both proteins. Both vectors were transfected into HEK293 cells, and after 48 hours, I detected the localization of both proteins using anti-Flag and anti-HA antibodies. There was partial co-localization between both proteins mainly in the cell membrane region indicated by yellow color in **Fig. 3-3 C.**

Taken together, these results suggested that CADM1 and RapGEF2 are interacting proteins.

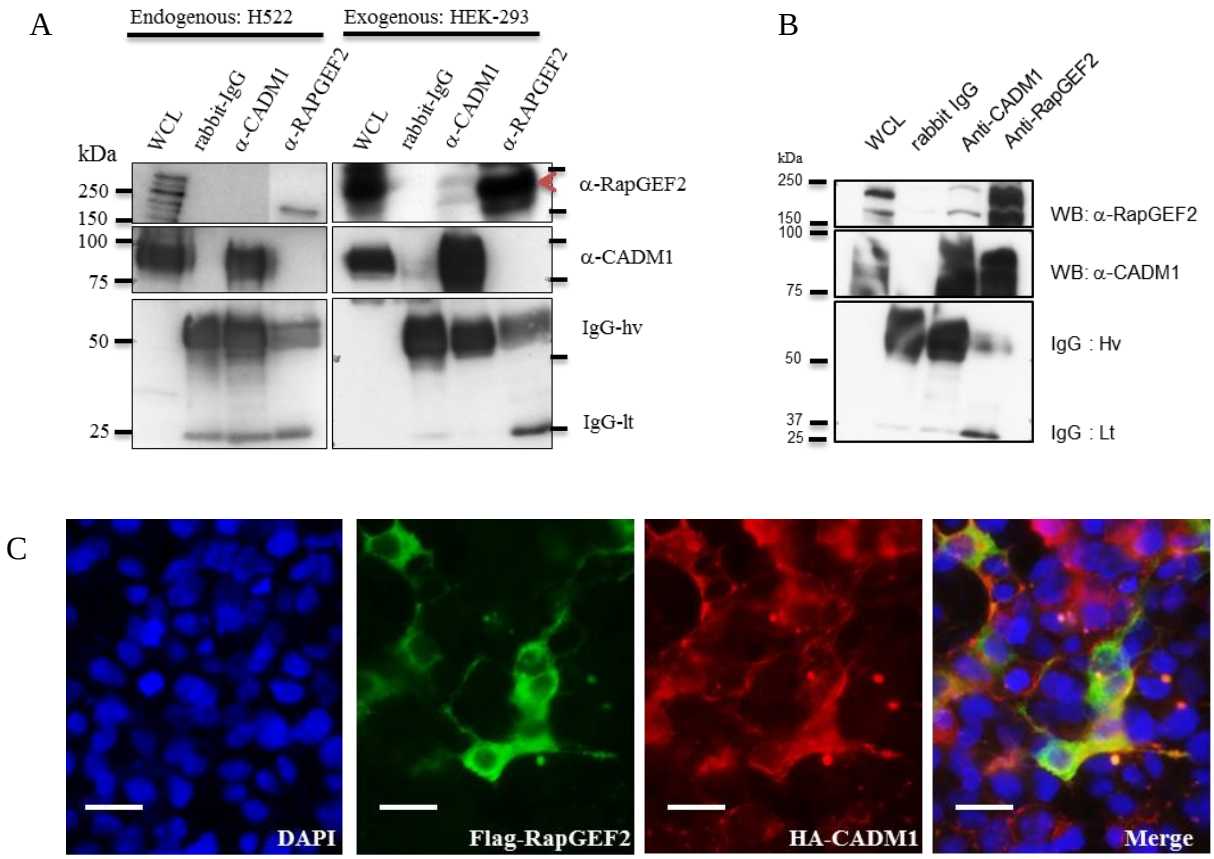


Fig 3-3: Interaction and localization of CADM1 and RapGEF2.
(A) and (B) For exogenous expression, RapGEF2 expression vector (pF1KT7 Flexi vector) and CADM1 expressing vectors were co-transfected into HEK-293 cells. 1 µg of antibodies were used to precipitate CADM1 and IgG and 5 µg to precipitate RapGEF2. Rabbit IgG used as an internal control. (C) Immunocytochemistry analysis for localization of CADM1 and RapGEF2 was done using tagged vectors; Flag-RapGEF2 and HA-CADM1 expression vectors. The localization was detected after 48 hours of transfection. Bar = 20 µm.

3.4.5 Expression of CADM1 and RapGEF2 in lung adenocarcinoma tissues

To investigate the expression of CADM1 and RapGEF2 in human lung adenocarcinoma tissues, I used tissue microarray (TMA) sections which include a total of 165 cases. Generally, immunostaining of these sections indicated a significant inverse correlation between CADM1 and RapGEF2. RapGEF2 showed granular staining pattern in cytoplasm. High-grade subtype frequently showed high expression of RapGEF2. On TMA sections, there are 7 pleomorphic carcinoma cases (high grade adenocarcinoma), and among them 6 cases showed high level expressions of RapGEF2 and 3 cases were double positive for CADM1 and RapGEF2 (**Fig. 3-4**). Pathological analyses indicated that high expression of RapGEF2 is significantly correlated with advanced T-factor, pleural invasion, histological high grade, and heavy smoking (**Table 3-2**).

I performed survival analysis, and high expression of RapGEF2 is significantly correlated with poorer prognosis (**Fig. 3-5 A**). Cases could be divided into two groups; CADM1 positive cases and CADM1 negative cases. Among CADM1 positive cases, RapGEF2 positive cases significantly showed poorer prognosis, but among CADM1 negative cases, there is no significant difference between RapGEF2 positive cases and RapGEF2 negative cases (**Fig. 3-5 B**).

This data indicates that CADM1 is acting mainly as tumor suppressor, but rarely, it can interact with RapGEF2, and strongly induce malignant progression. The number of RapGEF2 positive and CADM1 adenocarcinoma positive cases were only 2, but all of them were solid predominant adenocarcinoma, a high-grade subtype.

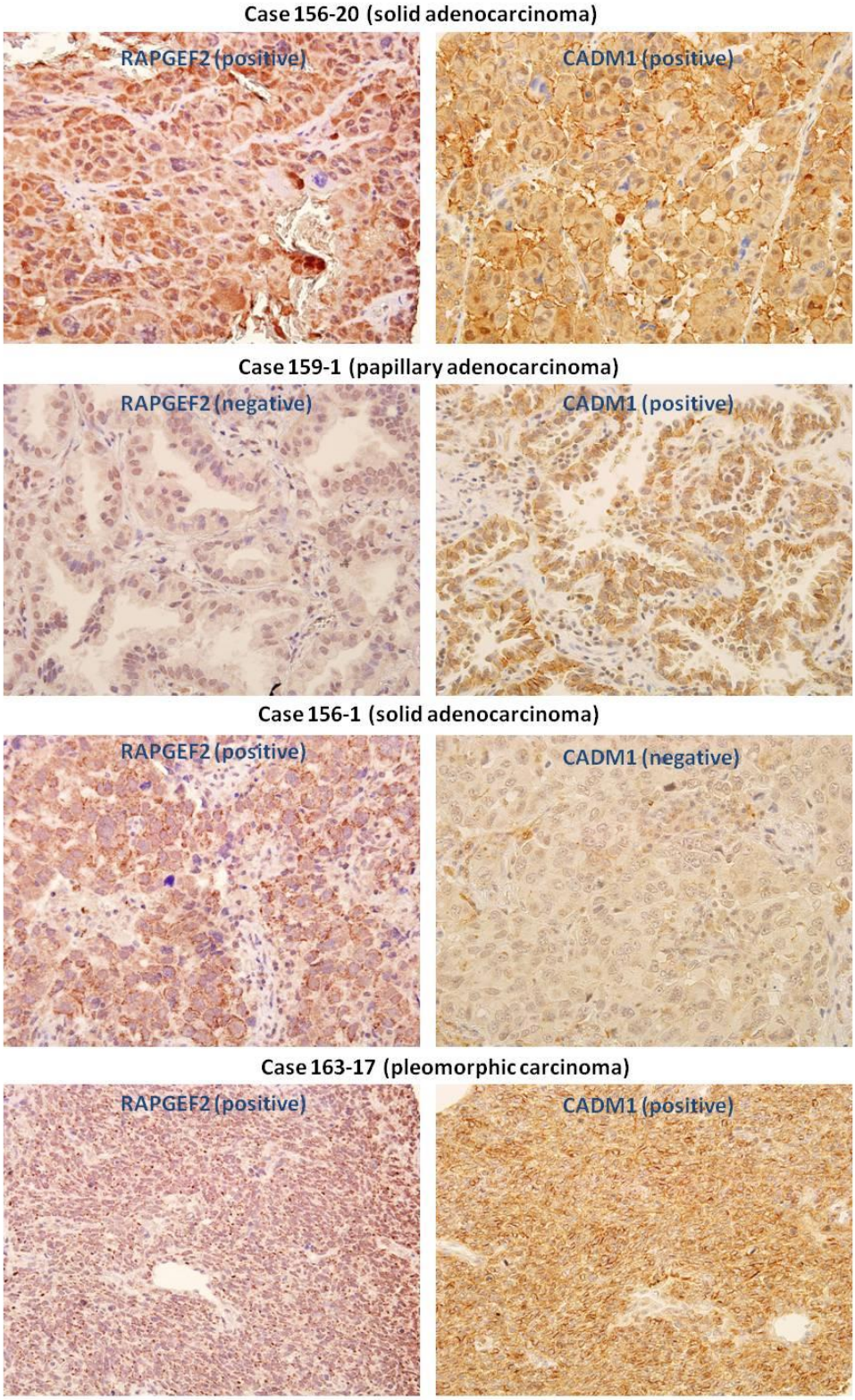


Figure 3-4: Immunohistochemical analysis of human adenocarcinoma tissues.

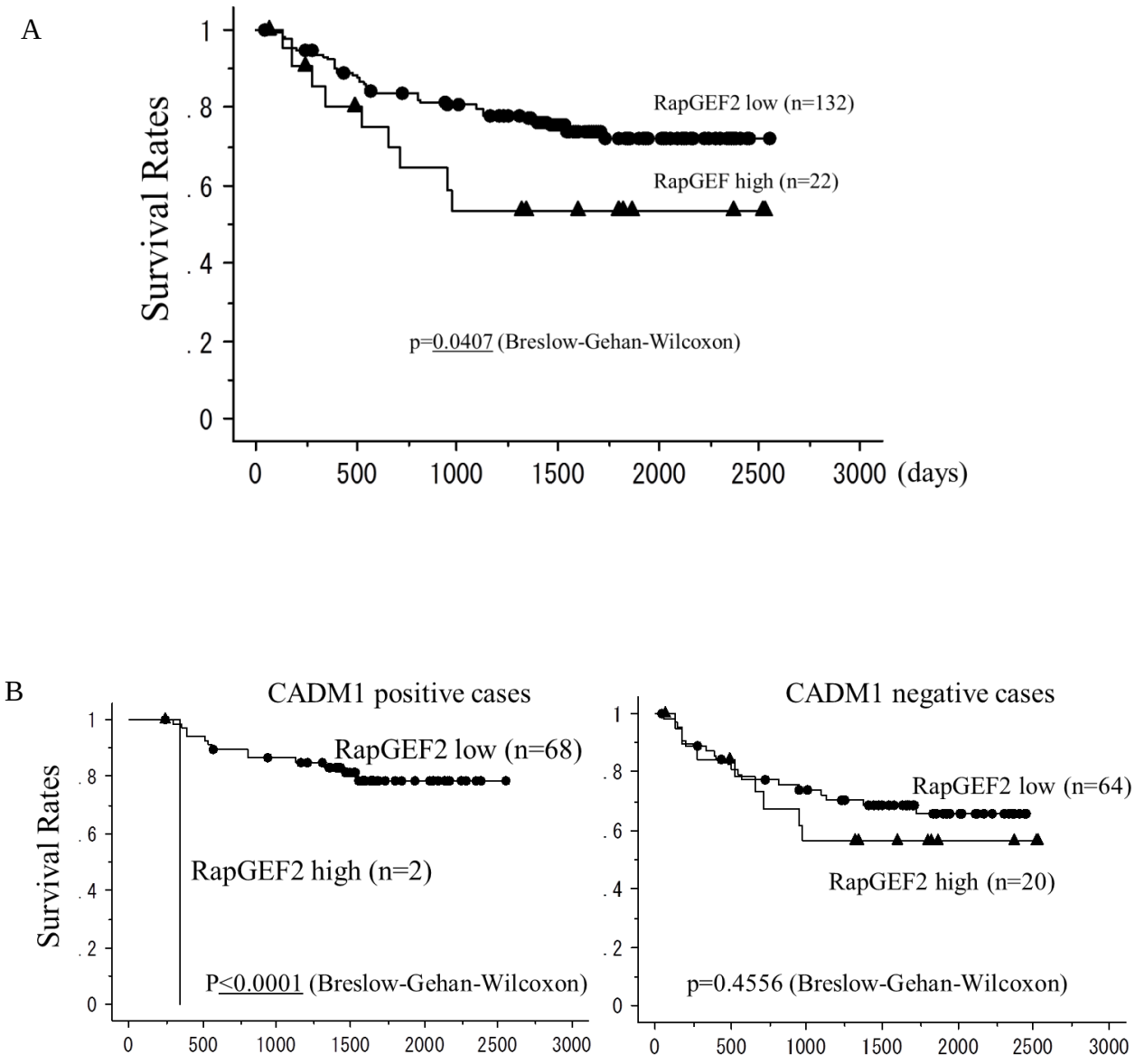


Figure 3-5: Survival analysis of RapGEF2 and CADM1 in lung adenocarcinoma tissues.

(A) Disease-free survival rates of RapGEF2 alone. (B) Disease-free survival rates of RapGEF2 in CADM1 positive cases (left) and CADM1 negative cases (right)

3.4.4 Simple over-expression of CADM1 and RapGEF2 is not sufficient to activate Rap1

To test the biological role of interaction between CADM1 and RapGEF2, I performed Rap1 activation assay. I first transfected CADM1 and RapGEF2 to HEK293 cells (both tagged and untagged vectors). This test is a pull-down assay depends on the idea that Rap1 is activated when binding to GTP molecules (positive control) and inactivated when binds to GDP molecules (negative control). Control and tested lysates were then incubated with glutathione beads before adding GST-RalGDS-RBD, which is a GST-tagged fusion protein, corresponding to amino acids 788-884 of human Ral GDS-Rap binding domain (RBD). The expected band of Rap1 is around 25 kDa.

The test showed a strong Rap1 band for positive control and no band for negative control, indicates it was working well (**Fig. 3-6**). However, no band was detected when transfecting both tagged and untagged expression vectors for CADM1 and RapGEF2 to HEK293 cells, implying simple over-expression of these proteins is not sufficient to activate Rap1.

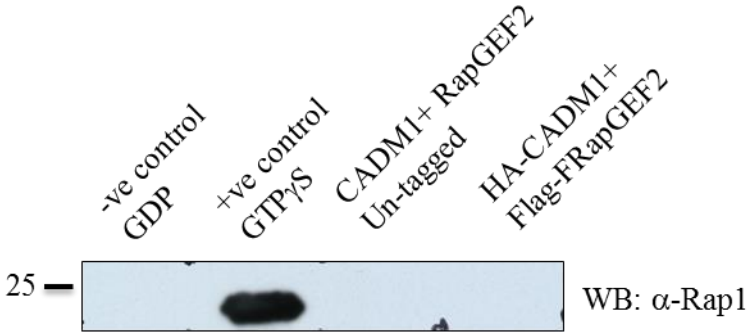


Fig 3-6: Rap1 activation assay. GDP and GTP molecules were used as negative and positive controls respectively.

Section 3-4 Discussion

Using microarray expression data in 41 lung adenocarcinoma cell lines, RapGEF2 was identified as a potential CADM1-interacting protein. Subsequent experiments indicated similar expression patterns of both proteins in cell lines, evidence of co-immunoprecepitation both proteins, and co-localization in cell membrane region.

Rap1 is a small GTPase that regulates adherence junction (AJ, mediated by E-cadherin) **(19)**, and controls endothelial junctions and tight junctions (TJ) **(20)**. It has been found that Rap1 is activated during early AJ formation and that Rap1 is activated by disruption of AJ. For being active, Rap1, like other GTPases, should be binding to GTP molecules provided by the effect of guanine exchange factors (RapGEFs) in the presence of effector proteins known as GTPase activating proteins (GAPs). RapGEF2 (PDZ-GEF1) has been found to mediate Rap1 activation upon AJ disruption and mice with RapGEF2 knockdown died at 9.5 days post coitum due to failure of vasculature in yolk sac and defective B-raf/ERK signaling **(21)**.

Several observations indicate CADM1 may be involved in Rap1 signaling:

i. Both molecules Rap1 and CADM1 are involved in cell-cell adhesion.
ii. CADM1 may be involved in recruitment of E-cadherin to cell-cell contact sites and E-cadherin is lost from cell boundaries in cells lacking CADM1 **(3)**.
iii. CADM1 reported to interact with Tiam1 to regulate actin cytoskeleton through Rac activation. Rap1 has also been shown to be interact directly

with Tiam1 (a Rac-GEF) leading to activation of cdc42 and Par polarity complex that ultimately activate PKCζ and Rac1 in T cells **(18)**.

iv. CADM1 reported to interact with MAGuKs (MMP1, 2, 3) **(3)** and RapGEF2 was also reported to interact with MAGuK (MAG1) to interact with E-cadherin through β-catenin **(22)**.

Overall, these observations suggested CADM1 might be involved in Rap1-mediated AJ formation and repair.

Rap1 is activated by several GEFs, that response to different second messengers (**Fig. 3-6**). For example, Epac1 and 2 (RapGEFs 3 and 4) respond to cAMP **(23, 24)**, C3G (RapGEF1) respond to receptor tyrosine kinase activation **(25)**, and CalDAG1 is responsive to Calcium and diacylglycerol **(26)**. Although RapGEF2 harbors a cAMP binding domain in the N terminus and that RapGEF2 is closely related to Epacs which are responsive to cAMP, RapGEF2 doesn't interact with cAMP or cGMP **(27)**, so the exact stimulus of RapGEF2 is not known.

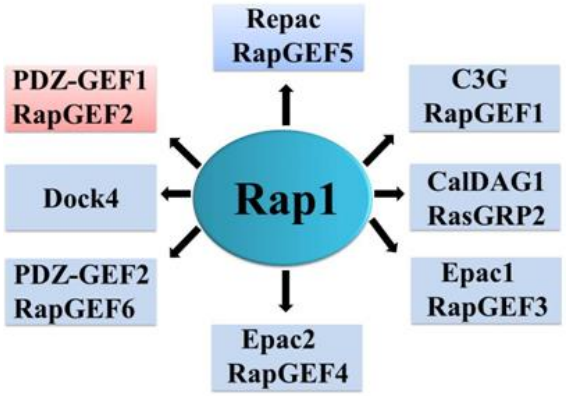


Fig. 3-7: Rap1 GEF2
Several RapGEFs have been described. They respond to different stimuli; Epac1 and 2 (RapGEFs 3 and 4) respond to cAMP, C3G (RapGEF1) respond to receptor tyrosine kinase activation, and CalDAG1 is responsive to Calcium and diacylglycerol.

My results showed that CADM1 interacts with RapGEF2 (**Fig. 3-3 B**) and both are co-localized in the cell-cell boundaries (**Fig. 3- 3 C**). However, simple over-expression of both proteins is insufficient to activate Rap1, the basic step needed for Rap1 signaling. This indicates this interaction might be facilitated by other molecules, or a specific physiological stimulus is needed which I am planning to investigate in the future.

In cancer, the role of Rap1 has reported in quite contradictory reports. For example, Rap1 has reported roles in prevention of cancer:

i. Rap1 was reported to antagonize the Ras-induced transformation by binding to Raf-1 **(28)**. Although Rap1, like Ras, can bind to Raf-1 with high affinity, but it doesn't activate it because Raf-1 will not be phosphorylated at this case **(29)**.

ii. Rap1 was also reported to antagonize MAP kinases (ERKs). The inactivation of Rap1 by disrupting Crk-C3G leads to activation of ERKs **(30)**.

However, contrary reports suggested that Rap1 is involved in tumorigenesis independently of Ras by enhancing the proliferation and morphological transformation of cancer cells **(31-33)**.

Overall, this contradiction suggests that Rap1 may function in a tissue-specific fashion or functions in dependence of the candidate GEF molecules.

CADM1 on other hand has the same contradiction as although it is a well-known tumor suppressor protein, but it was found to have the potentiality to act as oncogene in adult T-cell leukemia **(34)**.

Many of cell lines loose CADM1 as tumor suppressor, but some cell lines can keep CADM1. In primary tumor, CADM1 is frequently seen in well differentiated adenocarcinoma, such as BAC. But, interestingly, among cell lines, CADM1 is negative in bronchial epithelial phenotype, and positive in mainly mesenchyme-like phenotype. In addition, positive high expression of both molecules (CADM1 and RapGEF2) was detected in solid and also in pleomorphic phenotypes (poorly-differentiated phenotypes); I speculate that in such phenotypes, CADM1 will work as strong tumor inducer, possible by recruiting RapGEF2. This interaction will then promote invasion and cancer progression.

My hypothesis to explain the discrepancies between results of cell lines and TMA is as follows: CADM1 will work as tumor suppressor in bronchial epithelial phenotype lung adenocarcinoma with (i) high expression of bronchial epithelial markers, and (ii) frequent mutations of EGFR. So cell lines derived from poorly differentiated bronchial epithelial phenotype primary tumors showed loss of CADM1, because CADM1 is an obstacle for their survive. However, in the cases of EMT phenotype, though many of them showed loss of CADM1, some cases can use CADM1 as oncogene, with interaction between RapGEF2 and CADM1 where RapGEF2 can stabilize CADM1 expression. Those tumors are high-grade, so they can be easily established as cell lines. So there is positive correlation between CADM1 and RAPGEF2 in lung adenocarcinoma cell lines.

Section 3-5 Conclusions and future plan

I have here demonstrated in this ongoing study that Adhesion molecule CADM1 is co-expressed and co-localize with Rap1 exchange molecule RapGEF2. This indicates CADM1 may have a role in Rap1 signaling pathway.

Since both molecules (CADM1 and Rap1) are involved in cell-cell adhesion, and both of them seem to be promote cancer progression in poorly-differentiated tumors (**Fig. 3-7**). The future plan is to study the role of this interaction.

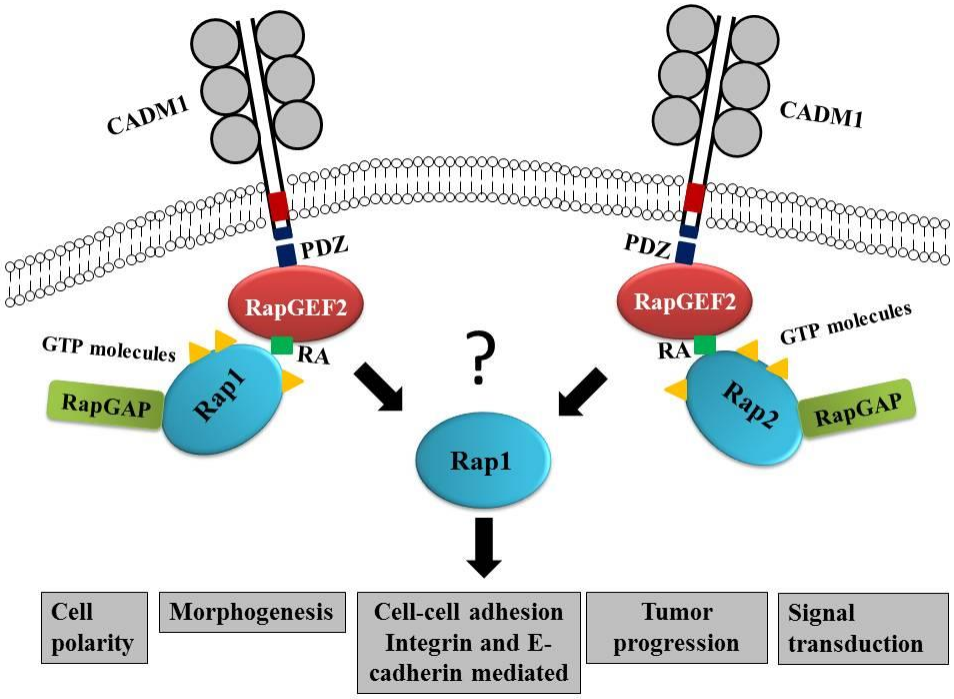


Fig. 3-8: Proposed scheme for CADM1-RapGEF2 interaction.

Interaction between CADM1 and RapGEF2 will affect one of Rap1 functions shown in this scheme in below. As both molecules have a main function of being involved in cell-cell adhesion, this function was shown in red and will be investigated as the first step.

Section 3-6 References

1- Kuramochi M, Fukuhara H, Nobukuni T, Kanbe T, Maruyama T, Ghosh HP, Pletcher M, Isomura M, Onizuka M, Kitamura T, Sekiya T, Reeves RH, Murakami Y. TSLC1 is a tumor-suppressor gene in human non-small-cell lung cancer. ***Nature Genetics***, 27:427-430. (2001)

2- Masuda M, Yageta M, Fukuhara H, Kuramochi M, Maruyama T, Nomoto A, Murakami Y. The tumor suppressor protein TSLC1 is involved in cell-cell adhesion. ***Journal of Biological Chemistry***, 277:31014-31019. (2002)

3- Sakurai-Yageta M, Masuda M, Tsuboi Y, Ito A, Murakami Y. Tumor suppressor CADM1 is involved in epithelial cell structure. ***Biochemical and Biophysical Research Communications***, 390:977-982. (2009)

4- Yageta M, Kuramochi M, Masuda M, Fukami T, Fukuhara H, Maruyama T, Shibuya M, Murakami Y. Direct association of TSLC1 and DAL-1, two distinct tumor suppressor proteins in lung cancer. ***Cancer Research***, 62:5129-5133. (2002)

5- Shingai T, Ikeda W, Kakunaga S, Morimoto K, Takekuni K, Itoh S, Satoh K, Takeuchi M, Imai T, Monden M, Takai Y. Implications of nectin-like molecule-2/IGSF4/RA175/SgIGSF/TSLC1/SynCAM1 in cell-cell adhesion and transmembrane protein localization in epithelial cells. ***Journal of Biological Chemistry***, 278:35421-35427. (2003)

6- Fukuhara H, Masuda M, Yageta M, Fukami T, Kuramochi M, Maruyama T, Kitamura T, Murakami Y, Masvuda M. Association of a lung tumor suppressor TSLC1 with MPP3, a human homologue of Drosophila tumor suppressor Dlg. ***Oncogene***, 22:6160-6165. (2003)

7- Liang QL, Chen GQ, Li ZY, Wang BR. Function and histopathology of a cell adhesion molecule TSLC1 in cancer. ***Cancer Investigation***, 29:107-112. (2011)

8- Sasaki H, Nishikata I, Shiraga T, Akamatsu E, Fukami T, Hidaka T, Kubuki Y, Okayama A, Hamada K, Okabe H, Murakami Y, Tsubouchi H, Morishita K. Overexpression of a cell adhesion molecule, TSLC1, as a possible molecular marker for acute-type adult T-cell leukemia. ***Blood***, 105:1204-1213. (2005)

9- Kikuchi S, Iwai M, Sakurai-Yageta M, Tsuboi Y, Ito T, Maruyama T, Tsuda H, Kanai Y, Onizuka M, Sato Y, Murakami Y. Expression of a splicing variant of the CADM1 specific to small cell lung cancer. ***Cancer Science***, 103:1051-1057. (2012)

10- Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, Thun MJ. Cancer statistics, 2006. ***CA: A Cancer Journal for Clinicians***, 56:106-130. (2006)

11- Russell PA, Wainer Z, Wright GM, Daniels M, Conron M, Williams RA. Does lung adenocarcinoma subtype predict patient survival?: A clinicopathologic study based on the new International Association for the Study of Lung Cancer/American Thoracic Society/European Respiratory Society international multidisciplinary lung adenocarcinoma classification. ***Journal of Thoracic Oncology***, 6:1496-1504. (2011)

12- Ito A, Okada M, Uchino K, Wakayama T, Koma Y, Iseki S, Tsubota N, Okita Y, Kitamura Y. Expression of the TSLC1 adhesion molecule in pulmonary epithelium and its down-regulation in pulmonary adenocarcinoma other than bronchioloalveolar carcinoma. ***Laboratory Investigation***, 83:1175-1183. (2003)

13- Fukami T, Fukuhara H, Kuramochi M, Maruyama T, Isogai K, Sakamoto M, Takamoto S, Murakami Y. Promoter methylation of the TSLC1 gene in advanced lung tumors and various cancer cell lines. ***International Journal of Cancer***, 107:53-59. (2003)

14- Uchino K, Ito A, Wakayama T, Koma Y, Okada T, Ohbayashi C, Iseki S, Kitamura Y, Tsubota N, Okita Y, Okada M. Clinical implication and prognostic significance of the tumor suppressor TSLC1 gene detected in adenocarcinoma of the lung. ***Cancer***, 98:1002-1007. (2003)

15- Goto A, Niki T, Chi-Pin L, Matsubara D, Murakami Y, Funata N, Fukayama M. Loss of TSLC1 expression in lung adenocarcinoma: relationships with histological subtypes, sex and prognostic significance. ***Cancer Science***, 96:480-486. (2005)

16- Kikuchi S, Yamada D, Fukami T, Maruyama T, Ito A, Asamura H, Matsuno Y, Onizuka M, Murakami Y. Hypermethylation of the TSLC1/IGSF4 promoter is associated with tobacco smoking and a poor prognosis in primary nonsmall cell lung carcinoma. ***Cancer***, 106:1751-1758. (2006)

17- Nagase T, Yamakawa H, Tadokoro S, Nakajima D, Inoue S, Yamaguchi K, Itokawa Y, Kikuno RF, Koga H, Ohara O. Exploration of human ORFeome: high-throughput preparation of ORF clones and efficient characterization of their protein products. ***DNA Research***, 15:137-149. (2008)

18- Masuda M, Maruyama T, Ohta T, Ito A, Hayashi T, Tsukasaki K, Kamihira S, Yamaoka S, Hoshino H, Yoshida T, Watanabe T, Stanbridge EJ, Murakami Y. CADM1 interacts with Tiam1 and promotes invasive phenotype of human T-cell leukemia virus type

I-transformed cells and adult T-cell leukemia cells. ***Journal of Biological Chemistry***, 285:15511-15522. (2010)

19- Dubé N, Kooistra MR, Pannekoek WJ, Vliem MJ, Oorschot V, Klumperman J, Rehmann H, Bos JL. The RapGEF PDZ-GEF2 is required for maturation of cell-cell junctions. ***Cell Signaling***, 20:1608-1615. (2008)

20- Pannekoek WJ, van Dijk JJ, Chan OY, Huveneers S, Linnemann JR, Spanjaard E, Brouwer PM, van der Meer AJ, Zwartkruis FJ, Rehmann H, de Rooij J, Bos JL. Epac1 and PDZ-GEF cooperate in Rap1 mediated endothelial junction control. ***Cell Signaling***, 23:2056-2064. (2011)

21- Satyanarayana A, Gudmundsson KO, Chen X, Coppola V, Tessarollo L, Keller JR, Hou SX. RapGEF2 is essential for embryonic hematopoiesis but dispensable for adult hematopoiesis. ***Blood***, 116:2921-2931. (2010)

22- Sakurai A, Fukuhara S, Yamagishi A, Sako K, Kamioka Y, Masuda M, Nakaoka Y, Mochizuki N. MAGI-1 is required for Rap1 activation upon cell-cell contact and for enhancement of vascular endothelial cadherin-mediated cell adhesion. ***Molecular Biology of the Cell***, 17:966-976. (2006)

23- de Rooij J, Zwartkruis FJ, Verheijen MH, Cool RH, Nijman SM, Wittinghofer A, Bos JL. Epac is a Rap1 guanine-nucleotide-exchange factor directly activated by cyclic AMP. ***Nature***, 396:474-477. (1998)

24- Kawasaki H, Springett GM, Mochizuki N, Toki S, Nakaya M, Matsuda M, Housman DE, Graybiel AM. A family of cAMP-binding proteins that directly activate Rap1. ***Science***, 282:2275-2279. (1998)

25- Gotoh T, Hattori S, Nakamura S, Kitayama H, Noda M, Takai Y, Kaibuchi K, Matsui H, Hatase O, Takahashi H. Identification of Rap1 as a target for the Crk SH3 domain-binding guanine nucleotide-releasing factor C3G. ***Molecular and Cellular Biology***, 15:6746-6753. (1995)

26- Kawasaki H, Springett GM, Toki S, Canales JJ, Harlan P, Blumenstiel JP, Chen EJ, Bany IA, Mochizuki N, Ashbacher A, Matsuda M, Housman DE, Graybiel AM. A Rap guanine nucleotide exchange factor enriched highly in the basal ganglia. ***Proceedings of the National Academy of Sciences USA***, 95:13278-13283. (1998)

27- de Rooij J, Boenink NM, van Triest M, Cool RH, Wittinghofer A, Bos JL. PDZ-GEF1, a guanine nucleotide exchange factor specific for Rap1 and Rap2. ***Journal of Biological Chemistry***, 274:38125-38130. (1999)

28- Bos JL, de Rooij J, Reedquist KA. Rap1 signalling: adhering to new models. ***Nature Reviews Molecular Cell Biology***, 2:369-377. (2001)

29- Carey KD, Watson RT, Pessin JE, Stork PJ. The requirement of specific membrane domains for Raf-1 phosphorylation and activation. ***Journal of Biological Chemistry***, 278:3185-3196. (2003)

30- Buensuceso CS, O'Toole TE. The association of CRKII with C3G can be regulated by integrins and defines a novel means to regulate the mitogen-activated protein kinases. ***Journal of Biological Chemistry***, 275:13118-13125. (2000)

31- Dupuy AJ, Morgan K, von Lintig FC, Shen H, Acar H, Hasz DE, Jenkins NA, Copeland NG, Boss GR, Largaespada DA. Activation of the Rap1 guanine nucleotide

exchange gene, CalDAG-GEF I, in BXH-2 murine myeloid leukemia. ***Journal of Biological Chemistry***, 276:11804-11811. (2001)

32- Gao Q, Srinivasan S, Boyer SN, Wazer DE, Band V. The E6 oncoproteins of high-risk papillomaviruses bind to a novel putative GAP protein, E6TP1, and target it for degradation. ***Molecular and Cellular Biology***, 19:733-744. (1999)

33- Singh L, Gao Q, Kumar A, Gotoh T, Wazer DE, Band H, Feig LA, Band V. The high-risk human papillomavirus type 16 E6 counters the GAP function of E6TP1 toward small Rap G proteins. ***Journal of Virolology***, 77:1614-1620. (2003)

34- Murakami Y. Involvement of a cell adhesion molecule, TSLC1/IGSF4, in human oncogenesis. ***Cancer Science***, 96:543-552. (2005)

Table 3-1: list of antibodies used in this study

Antibody	Name	Company	Species	Purpose (Concentration)
RapGEF2	1E8	Abnova	Mouse mAb	WB (1:750), IP (1:50)
CADM1	C18	MBL	Rabbit pAb	WB (1:5000), IP (1: 1000), (IHC) 1:500
GAPDH	6C5	Millipore	Mouse mAb	WB (1:10000)
Flag M2	F3165	Sigma	Mouse mAb	ICC (1: 1000)
Anti-HA	3F10	Roche	Rat mAb	ICC (1: 1000)
RapGEF2	ab105110	Abcam	Rabbit pAb	IHC (1: 250)

WB: western blot, IP: Immunoprecepitation, ICC: immunocytochemistry, IHC: immunohistochemistry. mAb: monoclonal antibody, pAb: polyclonal antibody.

Table 3-2: Correlations between expression levels of RapGEF2 and clinico-pathological factors

	RapGEF2 expression		
	high	low	*p*-value
Pathological stage			
- I	12	96	0.0743
- II+III+IV	10	35	
T-stage			
- T1	5	71	0.0070
- T2,T3,T4	17	61	
Nodal involvement			
- positive	7	28	0.2346
- negative	14	102	
Distant metastasis			
- positive	0	2	0.5611
- negative	22	130	
Tumor size			
- large (3cm=<)	9	34	0.1425
- small (3cm>)	13	98	
Lymphatic invasion			
- positive	5	28	0.8726
- negative	17	104	
Vessel invasion			
- positive	10	45	0.3031
- negative	12	87	
Pleural invasion			
- positive	14	49	0.0207
- negative	8	82	
Dissemination			
- positive	0	2	0.5611
- negative	22	130	
Pulmonary metastasis			
- positive	2	8	0.5933
- negative	20	124	
Histological grades			
- High	15	15	
- Intermediate	5	78	<0.0001
- Low	2	39	
CADM1			
- high	2	68	0.0002
- low	20	64	
Smoking Index			
- 600=<	14	47	0.0019
- 600>	5	83	

Acknowledgments

This work couldn't be accomplished without those people whom I want to acknowledge:

My mother: for giving me life.

My father: in his resting life.

My lovely family: my wife Reem and my lovely daughters Aya and Alaa: for giving means to my life.

My teachers: Yoshinori Murakami, Yusuke Nakamura, and Daisuke Matsubara.

My colleagues: in Murakami lab and Nakamura lab.

Onishi Akiko sensei: for her kind support for me and my family in my stay in Japan

To Japan: for the precious opportunity to study in this prestigious university.

Thank you very much!