Choosing haplotype-tagging snps based on unphased genotype data using a preliminary sample of unrelated subjects with an example from the multiethnic cohort study. *Hum Hered.* 2003;55:27-36

22.Stram DO, Leigh Pearce C, Bretsky P, Freedman M, Hirschhorn JN, Altshuler D, Kolonel LN, Henderson BE, Thomas DC. Modeling and e-m estimation of haplotype-specific relative risks from genotype data for a case-control study of unrelated individuals. *Hum Hered.* 2003;55:179-190

23.Hixson JE, Vernier DT. Restriction isotyping of human apolipoprotein e by gene amplification and cleavage with hhai. *J Lipid Res.* 1990;31:545-548

24.Johnson GC, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RC, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SC, Clayton DG, Todd JA. Haplotype tagging for the identification of common disease genes. *Nat Genet.* 2001;29:233-237

25.Tregouet DA, Escolano S, Tiret L, Mallet A, Golmard JL. A new algorithm for haplotype-based association analysis: The stochastic-em algorithm. *Ann Hum Genet.* 2004;68:165-177

26.Greenwood TA, Rana BK, Schork NJ. Human haplotype block sizes are negatively correlated with recombination rates. *Genome Res.* 2004;14:1358-1361

Tables

TABLE 1. Characteristics of Cases vs Controls, and Controls in Osaka vs Controls in Kyoto

	Cases	Controls	*P* value	Osaka	Kyoto	*P* value
Number	362	332		136	66	
Female, %	67.4	54.5	0.0005	57.4	59.1	0.4408
Age at diagnosis, y						
Mean±SD	59.2 ± 10.8	62.2 ± 9.9	0.000172*	62.9± 9.3	60.6± 9.8	0.1134*
Range	26-90	40-88		40-81	43-86	
Hypertension, %	56.1	42.5	0.0003	47.8	40.9	0.0019
Current or ex smoker, %	39.5	37.7	0.617	39.7	27.3	<0.0001
Drinker, %	39.2	43.7	0.235	46.3	33.3	<0.0001
Family history of IA and/or SAH, %	24.0	0		0	0	
Ruptured IA, %	50.6	0		0	0	

P values were calculated by χ^2 test. * was calculated by Student's t test.

Osaka represents control subjects in Osaka; Kyoto, control subjects in Kyoto.

TABLE 2. Comparisons of Allele Frequencies between Cases with IA and Controls by Adjusting with Covariates

Gene	SNP	Allele	Cases, n (%)	Controls, n (%)	Odds ratio (95%CI)	*P* value
ELN	INT4	G	515 (81.5%)	488 (79.7%)	0.99 (0.95-0.98)	0.94
		A	117 (18.5%)	124 (20.3%)		
	EX 5	C	682 (94.2%)	637 (95.9%)	1.48 (0.92-2.37)	0.10
		T	42 (5.8%)	27 (4.1%)		
	INT5	G	462 (63.8%)	446 (67.4%)	1.09 (0.87-1.37)	0.46
		A	262 (36.2%)	216 (32.6%)		
	EX20	G	591 (81.6%)	543 (81.8%)	1.05 (0.78 - 1.43)	0.74
		A	133 (18.4%)	121 (18.2%)		
	INT20	T	553 (76.4%)	510 (76.8%)	1.05 (0.81-1.36)	0.70
		C	171 (23.6%)	154 (23.2%)		
	INT21	G	704 (97.2%)	648 (97.8%)	1.29 (0.61-2.67)	0.50
		A	20 (2.8%)	14 (2.2%)		
	INT22	T	517 (71.4%)	454 (68.8%)	0.90 (0.71-1.14)	0.39
		C	207 (28.6%)	206 (31.2%)		
	EX25	G	706 (97.8%)	653 (98.3%)	1.28 (0.60-2.72)	0.52
		C	16 (2.2%)	11 (1.7%)		
NOS2A	INT7	A	364 (50.3%)	346 (52.1%)	1.08 (0.87-1.33)	0.49
		G	360 (49.7%)	318 (47.9%)		
	INT7'	I	665 (91.9%)	607 (91.4%)	0.90 (0.60-1.37)	0.63
		D	59 (8.1%)	57 (8.6%)		
	INT8	A	332 (45.9%)	331 (50.2%)	0.83 (0.67-1.05)	0.12
		G	392 (54.1%)	329 (49.8%)		
	INT12	C	535 (73.9%)	495 (74.8%)	1.07 (0.84-1.36)	0.59
		T	189 (26.1%)	167 (25.2%)		
	EX16	C	676 (93.6%)	622 (93.7%)	0.93 (0.59-1.48)	0.77
		T	46 (6.4%)	42 (6.3%)		
	INT16	G	611 (85.1%)	568 (86.1%)	1.06 (0.78-1.44)	0.71
		T	107 (14.9%)	92 (13.9%)		
	EX19	A	696 (97.2%)	644 (97.0%)	1.06 (0.56-2.01)	0.85
		G	20 (2.8%)	20 (3.0%)		
	EX22	G	537 (74.2%)	493 (74.5%)	1.04 (0.82-1.33)	0.73
		A	187 (25.8%)	169 (25.5%)		
APOE	EX4	ε2	13 (3.6%)	13 (4.1%)	1.35 (0.74-2.46)	0.33
		ε3	44 (12.2%)	38 (11.5%)		
		ε4	305 (84.2%)	280 (84.4%)	0.88 (0.63-1.22)	0.42
ACE2	INT3	T	65 (55.1%)	76 (50.3%)	0.60*	0.44*
	(male)	C	53 (44.9%)	75 (49.7%)		
	INT3	T	248 (50.8%)	193 (53.3%)	1.17 (0.88-1.54)	0.28
	(female)	C	240 (49.2%)	169 (46.7%)		

CI indicates confidential interval. * was calculated by χ^2 test. No difference in genotype frequencies was detected between cases and controls.

TABLE 3. Comparisons of Haplotype Frequencies in *ELN* and *NOS2A* between Cases with IA and Controls by Adjusting with Covariates

Gene	Variables	Haplotype	Cases, %	Controls, %	Odds ratio (95%CI)	*P* value
ELN	Haplotype	GCGGTGTG	13.4	14.2	0.92 (0.63-1.36)	0.69
		GCGGTGCG	6.6	6.8	0.90 (0.52-1.56)	0.72
		GCGATGTG	7.4	6.1	1.16 (0.68-1.98)	0.59
		GCGATGCG	8.3	9.9	0.90 (0.56-1.45)	0.66
		GCAGTGTG	21.6	21.6	Reference	
		ACGGCGTC	9.7	10.2	0.90 (0.58-1.40)	0.65
	INT20/INT22	TT	52.8	53.0	Reference	
		TC	23.6	23.7	1.01 (0.75-1.36)	0.95
		CT	18.6	15.6	1.21 (0.87-1.70)	0.26
		CC	5.0	7.6	0.71 (0.39-1.33)	0.29
	INT4/INT5/INT21	GGG	43.8	45.3	Reference	
		GAG	30.9	29.8	0.96 (0.73-1.26)	0.78
		AGG	19.4	20.8	0.90 (0.64-1.25)	0.53
NOS2A	Haplotype	AIACCGAG	18.4	21.8	0.78 (0.57-1.07)	0.12
		AIATCGAA	20.0	20.7	0.95 (0.71-1.27)	0.72
		GIGCCGAG	30.0	29.1	Reference	
		GIGCCTAG	6.6	5.8	1.05 (0.64-1.72)	0.83

In haplotypes of *NOS2A*, I indicates insertion.

TABLE 4. Regional Difference of Allele Frequency of *ELN* in Controls

SNP	allele	Osaka, n	Kyoto, n	χ^2	*P* value	P_{corr}
INT4	G	176	109	0.35	0.5512	
	A	44	23			
EX5	C	267	122	8.18	**0.0042**	**0.034**
	T	5	10			
INT5	G	182	97	1.54	0.2143	
	A	88	35			
EX20	G	220	104	0.25	0.6203	
	A	52	28			
INT20	T	224	97	4.28	**0.0385**	0.308
	C	48	35			
INT21	G	267	125	6.42	**0.0113**	0.090
	A	3	7			
INT22	T	181	83	0.53	0.4649	
	C	87	47			
EX25	G	268	128	1.12	0.2912	
	C	4	4			

Significant valiables ($P<0.05$) were shown in bold. P_{corr} indicates *P* value after Bonferroni correction.

Search on Chromosome 17 Centromere Reveals *TNFRSF13B* as a Susceptibility Gene for Intracranial Aneurysm: A preliminary study

Kayoko Inoue, MD, MPH, PhD; Youhei Mineharu, MD; Sumiko Inoue, PhD; Shigeki Yamada, MD, PhD; Fumihiko Matsuda, PhD; Kazuhiko Nozaki, MD, PhD; Katsunobu Takenaka, MD, PhD; Nobuo Hashimoto, MD, PhD; Akio Koizumi, MD, PhD

From the Department of Health and Environmental Sciences (K.I., Y.M., S.I., A.K.), Department of Neurosurgery (Y.M., S.Y., K.N., N.H.), Center for Genomic Medicine (F.M.); Kyoto University Graduate School of Medicine, Kyoto, Japan; and the Department of Neurosurgery, Takayama Red Cross Hospital (K.T.), Gifu, Japan.

Correspondence to Akio Koizumi, MD, PhD, Professor, Department of Health and Environmental Sciences, Graduate School of Medicine Kyoto University,

Konoe-cho, Yoshida, Sakyo-ku, Kyoto, 606-8501, Japan.

Phone: +81-75-753-4456

Fax: +81-75-753-4458

E-mail: koizumi@pbh.med.kyoto-u.ac.jp

Total number of words: 5,269 words

Journal Subject Heads: Stroke [55] Genetics of Stroke

Abstract

Background-Our previous studies have shown a significant linkage of intracranial aneurysms (IAs) to chromosome 17.

Methods and Results- Nine genes (*TNFRSF13B, M-RIP, COPS3, RAI1, SREBF1, GRAP, MAPK7, MFAP4 and AKAP10*) were selected from 108 genes which are located between D17S1857 and D17S1871 by excluding 99 genes that were pseudogenes, hypothetical genes, or well characterized genes but not likely associated with IA. Direct sequencing of all coding and regulatory regions in 58 cases (29 pedigree probands and 29 unrelated non-pedigree cases) was performed. Deleterious changes were found only in *TNFRSF13B*, K154X and c.585-586insA in exon4. Association of IA with *TNFRSF13B* was further studied in 304 unrelated cases and 332 control subjects. Rare non-synonymous changes, a splicing acceptor site change and a frame shift were found in unrelated cases (2.3%: 14/608) more frequently than in control subjects (0.8%: 5/664) (p=0.035). The association study using single nucleotide polymorphisms (SNPs) in an unrelated case-control cohort revealed a protective haplotype (Odds=0.69, 95%CI: 0.52-0.92, p=0.012) as compared with the major haplotype by adjusting for covariates.

Conclusion-We propose that *TNFRSF13B* is one of the susceptibility genes for IA.

Key Words: aneurysm, cerebrovascular disorders, genes, immune system

Introduction

Intracranial aneurysms (IAs) are one of the major public health problems in Japan. Mortality rate of subarachnoid hemorrhage (SAH), of which more than 90% is attributable to IA rupture, is estimated at 70 deaths per 10,000 person years and accounts for 2% of annual total deaths.[1] The consequences of SAH are catastrophic, with approximately half of IA ruptures resulting in immediate death.

In familial IAs, there is a three to five fold increase in risk for first-degree relatives of affected individuals compared with the general population.[2,3] A positive family history is a risk factor as strong as smoking, hypertension and heavily drinking alcohol.[4,5]

In an attempt to isolate susceptibility gene(s) for IA, four genome-wide linkage analyses have been reported.[6-9] In a series of studies, we have failed to identify a positive association with reported candidate genes.[9,10]

Since disease and genetic heterogeneity are postulated for IA,[11] extensive efforts are required to find the susceptibility gene(s) for IA, if the approach is limited to traditional positional cloning. On the other hand, the candidate gene approach relies serendipity.

In the present study, we hypothesize that the many rare variants contribute to a common phenotype.[12] We further assume that while deleterious changes are likely to be rare in the unaffected cohort, they may be more common in aggregate in the affected cohort. Consequently

we have assumed that variants associated with functional changes such as nonsense or non-synonymous variants should be more abundant in candidate genes that determine susceptibility for IA. On this rationale, candidate genes were searched in a primary gene set in the 17 centromere region between D17S1857 and D17S1871, where we found the maximum non-parametric LOD score (MNS) peak [MNS 3.00] at D17S2196.[9]

Methods

Study Design

Subjects from three groups participated. The first group was probands of 29 pedigrees with IA clustering.[9] The second group was consisted of 333 unrelated non-pedigree cases with IA and the third group had 332 control subjects.[10] Members of the first group and 29 unrelated cases which were selected randomly from the second group, constituted the first cohort. Remaining 304 unrelated non-pedigree cases of the second group constituted the second cohort and the third group constituted the third cohort. The response rates to our request of participation in our study were 95.1% in the second group and 94.8% in the third group, respectively.

The target region was 4.3 Mb encompassing D17S1857 and D17S1871 where we found significant linkage in families with an IA cluster.[9] 108 genes are now assigned to this region

(Supplement Table 1). We set an exclusion principle to choose the primary candidate gene set: we excluded 99 genes; 26 pseudogenes; 29 hypothetical genes; 22 enzymes and transporters; 8 developmentally regulated genes, 4 genes associated with Smith-Magenis Syndrome; 3 genes associated with neoplastic syndrome; 3 zinc finger proteins genes; 3 similar to keratin genes and one open reading frame. Finally, nine genes remained. These were *TNFRSF13B*, *M-RIP*, *COPS3*, *RAI1*, *SREBF1*, *GRAP*, *MAPK7*, *MFAP4* and *AKAP10* (Table 1). These nine genes were directly sequenced in all subjects of the first cohort. Whether or not sequence variants were functional was predicted by PolyPhen (http://tux.embl-heidelberg.de/ramensky).[13] Except for *TNFRSF13B*, none of the sequence variants in other genes were predicted to be deleterious. Further analysis was thus limited to *TNFRSF13B*. Using observed polymorphisms, an association study was conducted in the second cohort and the third cohort.

Study population

The probands of pedigrees and unrelated non-pedigree cases were diagnosed by digital subtraction angiography (DSA) or in operations throughout collaborating hospitals in western Japan. We have excluded cases with IA affected with known heritable diseases or autoimmune diseases. Control subjects were screened at the brain check-up in the same hospitals as cases and met the following criteria: (a) confirmation of absence of IA by DSA, 3-dimensional computerized tomography (3-D CT), or magnetic resonance angiography (MRA), (b) an age at

screening of ≥ 40 years old, (c) no medical history of any stroke including IA or SAH, and (d) no family history of IA or SAH in first-degree relatives.[10]

Individual and family history and life style was obtained by interviews. Past history and co-morbidity were also examined by clinical charts at the hospitals or interview charts at the brain check-ups.

The study was approved by the Ethics Committee of Kyoto University Institutional Review Board and appropriate informed consent was obtained from all subjects.

Direct sequencing and prediction of functional analysis for detected variants

All exons, intron-exon boundaries, putative promoter sequences and the 3'UTR were analyzed by direct sequencing of 9 genes for 58 cases (the first cohort). For sequencing, we referred to *TNFRSF13B*, *M-RIP*, *COPS3*, *RAI1*, *SREBF1*, *GRAP*, *MAPK7*, *MFAP4* and *AKAP10*, NCBI Map Viewer, http://www.ncbi.nih.gov/mapview/maps.cgi?. Primers for coding exons were designed from an intronic sequence >50 base pairs (bp) away from the intron-exon boundary and commercially synthesized by PROLIGO (PROLIGO Primers & Probes, Kyoto, Japan; http://www.proligo.com). For regulatory regions, about 500 bp upstream to the first exon was sequenced. However, if the database suggested the existence of a regulatory region further upstream, sequencing was done to cover the entire regulatory region. After PCR amplification and purification, sequencing was done on an ABI Prism 3100 Avant DNA sequencer (Applied

Biosystems, USA). We checked the SNP database (dbSNP) as a reference (http://www.ncbi.nlm.nih.gov/SNP/index.html). Primers and PCR conditions for each gene are available from the Supplement Table 2.

Among all the sequence changes identified by direct sequencing, we selected nonsense mutations and non-synonymous variants as primary candidate variants. Then, we conducted functional analysis for each non-synonymous variant by PolyPhen.

The number of subjects who had rare non-synonymous or deleterious changes was compared between the second cohort and the third cohort by Fisher's exact test using SAS software (Version 8.2. SAS Institute Inc).

Testing segregation in pedigrees

Three variants (K154X, c.585-586insA, G76C) of *TNFRSF13B* found in three probands were investigated for concordance of segregation in these families (Pedigree10, Pedigree26 and Pedigree15).[9]

Association study

SNPs of *TNFRSF13B* with allele frequency ≥1% in 58 cases (the first cohort: 29 probands of the pedigrees and 29 unrelated cases) were all genotyped by direct sequencing (P251L and S277S) or by polymerase chain reaction-restriction fragment length polymorphism (PCR-RFLP) using AlwI for c-247G>T and BfuAI for IVS3+25C>A in 304 unrelated cases (the second

cohort) and 332 controls (the third cohort).

Haplotypes were constructed using sequence variants with allele frequency ≥1% in the third cohort by THESIAS[14] (http://genecanvas.ecgene.net/). We used the following criteria to choose a set of haplotypes for association study: a set of the minimum number of haplotypes of which cumulative haplotype frequency was ≥80%[15] or a set of all halotypes of which frequencies were ≥5%.[16] Associations were analyzed by adjusting for covariates including sex, hypertension, smoking, and drinking habit. Bonferroni correction was done for comparison of multiple haplotypes, not for experiment-wide multiple testing.

Linkage Disequilibrium (LD) was analyzed and visualized with the Genotype2LDBlock (http://cgi.uc.edu/cgi-bin/kzhang/genotype2LDBlock.cgi).

Population attributable risk

The population attributable risk for a given haplotype was calculated as follows:

Population attributable risk= (Odds ratio-1)*IE*P/IT

Where IE is incidence of IA in the control cohort, IT is the incidence of IA in the general population and P is the reference haplotype frequency in the general population. We assumed that IE was equal to IT, and P in the control cohort was equal to that in the general population. Thus the population attributable risk will be obtained as follows:

Population attributable risk= (Odds ratio-1)* P

The authors had full access to the data and take full responsibility for its integrity. All authors have read and agree to the manuscript as written.

Results

Demographic features of the three cohorts

As shown in Table 2, among unrelated subjects the proportion of females or that of hypertension was higher in the second cohort than in the third cohort. Age at diagnosis was lower in the second cohort. No significant difference was found for either smoking or drinking habits.

Candidate genes

The primary candidate gene set after excluding genes based on a defined criteria was found to be related with immunity (*TNFRSF13B)*, regulatory component (*M-RIP*), protein kinase (*COPS3*, *MAPK7*), transcriptional factor (*RAI1*, *SREBF1*), signaling protein (*GRAP*), cell adhesion (*MFAP4*) and signal transduction (*AKAP10*).

Detected sequence changes in 58 cases (the first cohort) and their predicted effects on function are shown in Supplement Table 3. We identified seven sequence changes in *TNFRSF13B*, 20 sequence changes in *M-RIP*, six sequence changes in *COPS3*, 23 sequence changes in *RAI1*, nine sequence changes in *SREBF1*, nine sequence changes in *GRAP*, five sequence changes in *MAPK7*, two sequence changes in *MFAP4* and ten sequence changes in

AKAP10.

TNFRSF13B had two nonsense mutations and two non-synonymous variants predicted as "probably damaging" by PolyPhen: K154X and frame shift (c.585-586insA) in exon4, G76C in exon3 and P251L in exon5. Apparent deleterious variants, including nonsense mutations or non-synonymous variants, which were predicted to be "probably damaging", were identified only in *TNFRSF13B* (Table 3).

Segregation of the *TNFRSF13B* variants with the IA phenotype in pedigrees

Two nonsense mutations and one non-synonymous variant ("probably damaging" by PolyPhen) were found in probands in three pedigrees (Figure 1) out of 29 families. In one family (Pedigree10);[9] K154X was found in two affected siblings and one daughter while it was not detected in an unaffected younger brother. Insertion A (c.585-586insA) was found in one family (Pedigree26)[9]; an affected mother and her son had this mutation. This mutation was also found in an unaffected sibling of the mother, that later developed stroke but was not investigated for pathogenesis. G76C was found in another family (Pedigree15)[9]; two affected sisters had this variant but an unaffected sister did not.

Direct sequencing exons 3 to 5 in 304 unrelated cases (the second cohort) and 332 controls (the third cohort) in *TNFRSF13B*

An extensive search was done in exons 3 to 5 because there are two deleterious variants and

two probably damaging variants in these regions in *TNFRSF13B*. We further found additional sequence variants in the second and third cohort (Table 3). The number of subjects having rare non-synonymous changes, a splicing acceptor site change and a frame shift in *TNFRSF13B* were significantly larger in the 304 unrelated cases than in the 332 controls (Fisher's exact test, p=0.035, Table 4).

Locations of these variants were summarized in Figure 2. These non-synonymous variants were located on the region critical for function.[17-19] The 70th peptide S was conserved in *Xenopus laevis* but not in mice, dogs or rats. On the other hand, 74th E, 76th G, and 177th C were conserved in mice, dogs and rats. It should also be pointed out that G76S and C177R were found in more than one unrelated subjects. But, no one had more than one variant.

Association study

Allele frequencies of four SNPs were found to be ≥1% in the third cohort (Table 3). We thus used four SNPs (c-247G>C, IVS3+25C>A, P251L, S277S) of *TNFRSF13B* to construct haplotypes. LD structure was shown in Figure 3. Application of the selection criteria chose four hypotypes, covering 87% of all haplotypes (Table 5). Haplotype H1 (GACC) was found to be protective (odds=0.69, 95%CI: 0.52-0.92, p=0.012) as compared with the major haplotype H4 (TCTC). After Bonferroni correction for multiple comparison, the p value of H1 was still statistically significant (p_{corr}=0.048).

Population attributable risks

The population attributable risk was calculated to be about –8% for H1 haplotype against H4 compared with 24% for smoking against nonsmoking. Therefore the attributable risk for the *TNFRSF13B* variants was approximately one-third that of smoking.

Discussion

Extensive efforts have been made to search for susceptibility gene(s) for IA. So far, three genome-wide linkage analyses have been done for the general population. Except *ELN*,[6] *LOX*[20] and *COL1A2*,[21] no gene has been claimed as a candidate gene. There have been contradictions, however, in terms of involvement of *ELN* in IA.[22,23]

In the present study, we have conducted a systematic approach targeting a linked region on chromosome 17. We selected nine candidates from 108 genes and sequenced entire coding exons and regulatory regions in 58 cases (the first cohort). Since we found several variants including obvious deleterious mutations in *TNFRSF13B*, we searched variants in 304 unrelated cases (the second cohort) and 332 control subjects (the third cohort), although searches were limited to those in exons 3 to 5 which covered the critical areas CRD2 (cysteine-rich domain 2), TM (trans-membrane) and ICD (intracellular regions). The rare variants were significantly more frequent in IA unrelated cases than in control subjects. In addition, deleterious variants [K154X,

frame shift (c.585-586insA) and G76C] were clearly segregated in the families except in a family sibling who had c.585-586insA but did not have IA, implying that the penetrance of IA is not complete as was expected.[24,25] Finally, case-control studies using sequence variants revealed a protective haplotype (GACC) against the most common haplotype (TCTC). With these lines of evidence, *TNFRSF13B* emerges as a candidate for susceptibility for IA.

TACI (Transmembrane activator and calcium-modulator and cyclophilin ligand interactor) encoded by *TNFRSF13B* mediates isotype switching in B cells. The mutations in *TNFRSF13B* have recently been reported to be associated with common variable immunodeficiency (CVID) and IgA deficiency in humans.[26,27] In one of these studies, 11 mutations (4.1%) were found in 270 chromosomes from 135 sporadic CVID cases.[26] It is of particular interest that most sporadic cases with CVID had only one mutant allele, suggesting a mechanism of gain of function or haploinsufficiency.

Given that mutations of *TNFRSF13B* are associated with CVID or IgA deficiency, an unanswered question is why variants in *TNFRSF13B* are associated with IA. It is interesting that in our study 12 out of 17 rare variants in IA cases, and 3 out of 5 rare variants in control subjects were found in the CRD2 domain, while the majority of mutations in cases with CVID or IgA deficiency were found at the C terminal side to the CRD2 region, which transfers signals from cell surface to intracellular domains. We postulate that variants at the ligand binding site may

cause quantitative changes while mutations in signal transduction results in qualitative changes. Different modes of functional impairments might be associated with different phenotypes. Studies are needed to investigate this further.

In the present study, we found three nonsense mutations (one stop codon, one splicing acceptor site change, and one frame shift) and five rare non-synonymous changes in 17 cases. Each case had a single variant. It is interesting that these variants are novel and none was found in Caucasians.[26,27] The most common mutation among Japanese with IA is G76S (8/17) while that in Caucasians with CVID is A181E, suggesting founder mutations specific to ethnic groups. If so, genetic preposition to IA or CVID or IgA deficiency may be predicted by these founder mutations in the future.

The present study has several limitations. First, population attributable risks of IA are calculated to be 7% to 10% while that of smoking observed is about 24%, suggesting that the risk attributable to *TNFRSF13B* is about one-third of smoking in the present cohort. However, further studies are needed because only a small fraction of the risk is explained by *THNFRS13B*. Second, we have selected only nine genes as the primary gene set from 108 genes. We excluded genes of which functions are not well characterized or those of which well characterized functions are not considered to be involved in IA. Although this is primary screening, this study cannot be free from selection bias. In the next study, we are expanding the gene set so that it

includes some genes of which functions are unknown. Third, we tested with PolyPhen whether non-synonymous variants were functional or not. Bioinformatics approaches may sometime be misleading.[28] We should explore other genes which had "Possibly damaging" or "unknown" variants in future. Effects of variants on function of TACI should also be confirmed experimentally in future. Fourth, there may be an argument for the rare variants contribute to a common diseases hypothesis. It should be addressed, however, that the hypothesis can provide criteria for positive selection of a susceptibility gene, which would have been overlooked by a haplotype-based association study. Fifth, in the present study, we did not determine CVID related parameters such as B-cell expression of TACI and serum levels of immunoglobulin. Finally, we did not explore genes in LD with *TNFRSF13B*. The International HapMap Project (http://www.hapmap.org) suggests that there is LD between *LOC96597* and *TNFRSF13B*. Further exploration will be needed in future.

With positive findings, above rationale and reasonable background, we proposed that *TNFRSF13B* is one of the candidate genes for susceptibility for IA notwithstanding several limitations. This in turn proposes that immunological mechanisms may play a role in IA development to a discernible extent. Our hypothesis is in accordance with clinical experiences where IA is often found in subjects with autoimmune diseases.[29,30] Further studies are needed to strengthen our hypothesis. In addition, the present results might pave the way for investigating a