Figure 2. Genomic position and LD blocks. (A) Genomic position of non-synonymous (ns)SNPs and regulatory (r)SNPs in *NFKBIE* and *RTKN2*. *NFKBIE* (top) and *RTKN2* (bottom) correspond to transcripts NM_004556.2 and NM_145307.2, respectively. Exons are shown as boxes, where black boxes represent coding regions and open boxes represent untranslated regions. Intron sequences are drawn as lines. Open triangles represents nsSNPs and open diamond shapes indicate candidate rSNPs. dbSNP IDs of candidate causal variants were boxed in a solid line. (B) LD patterns for nsSNPs and candidate rSNPs in *NFKBIE* (left) and *RTKN2* (right) gene regions. LD blocks were constructed from genotype data of 3,290 control individuals of the GWAS. The diagrams show pairwise LD values as quantified using the D' and r^2 values.
doi:10.1371/journal.pgen.1002949.g002

independent effects of each variant would be the first step. For this purpose, a recent attempt to fine-map the known autoimmunity risk loci in Celiac disease (MIM 212750) using an "Immunochip" brought us several insights [34]. First, no stronger signals compared to the GWAS signals were detected in most of the known loci, while additional independent signals were found in several loci. Second, none of the genome-wide significant common SNP signals could be explained by any rare highly penetrant variants. Third, although the fine-mapping strategy could localize the association signals into finer scale regions, it could not identify the actual causal variants due to strong LD among the variants, indicating that an additional approach, such as functional evaluation of candidate variants, is needed.

In the present study, we focused on common variants to find causal variants. Instead of re-sequencing additional samples, we utilized the 1000 Genome Project dataset, where the theoretically estimated cover rate for common variants (frequency of >0.05) in our population is >0.99 [12,35]. To fine-map the association signals, we performed imputation-based association analysis, where we could not find any association signals that statistically exceeded the effect of landmark SNPs (rs2233434 for *NFKBIE* and rs3125734 for *RTKN2*) in both gene regions (Figures S3 and S4).

We also performed a conditional logistic regression analysis, and found no additional independent signals of association when conditioned on each landmark SNP (data not shown). Although the imputation-based association tests may yield some bias compared to direct genotyping of the variants, these results suggested that variants in strong LD with the landmark SNPs were strong candidates for causal variants.

Following the analysis of nsSNPs, we evaluated *cis*-regulatory effects of variants in the two regions by ASTQ analysis using both B-cell lines and primary cells (PBMC), the majority of which consisted of T and B lymphocytes. As the mechanism of gene-regulation is substantially different between cell types [26], ASTQ analysis in more specific cell types that are relevant to the disease etiology, such as Th1 and Th17 cells, would be ideal to evaluate the *cis*-regulatory effects of variants. In this context, a more comprehensive catalog of the eQTL database of multiple cell types should be established for genetic study of diseases. As our ASTQ analysis demonstrated *cis*-regulatory effects of variants in both regions, we then performed an integrated *in silico* and *in vitro* analysis to identify candidate regulatory variants. Accumulating evidence by recent ChIP-seq and DNase-seq studies suggested that *cis*-regulatory variants are
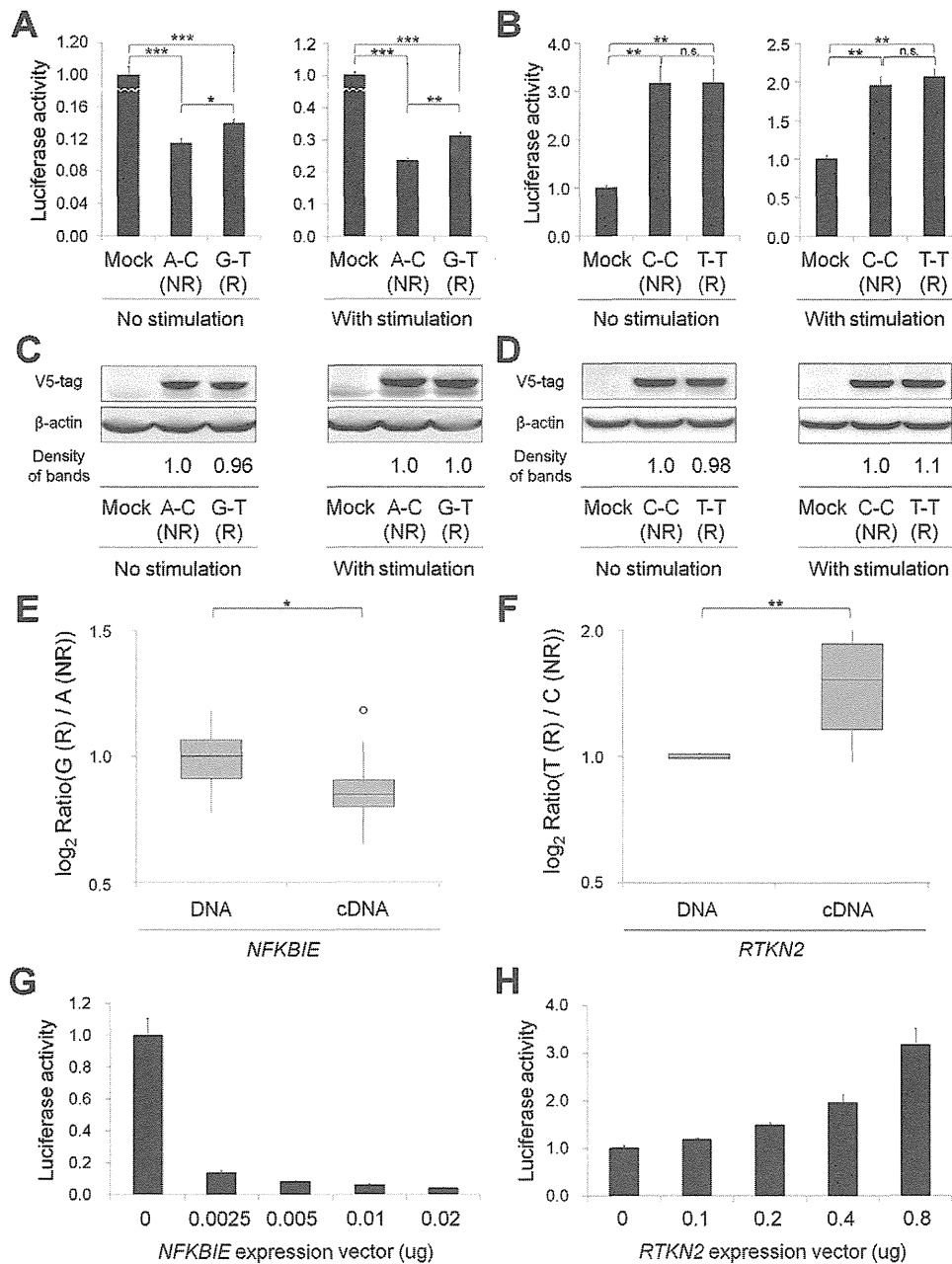
Figure 3. Functional evaluation of nsSNPs and allelic imbalance of expression in *NFKBIE* and *RTKN2*. (A, B) Effects of nsSNPs in *NFKBIE* (A) and *RTKN2* (B) on NF-κB activity by luciferase assays. Two haplotype constructs (A-C (rs2233434-rs2233433; non-risk (NR)) and G-T (risk (R)) for *NFKBIE* and C-C (rs3125734-rs61850830; NR) and T-T (R) for *RTKN2*) were used. The expression vector of each construct, pGL4.32[*luc2P*/NF-κB-RE] vector and pRL-TK vector were transfected into HEK293A cells. Data represent the mean ± s.d. Each experiment was performed in sextuplicate, and experiments were independently repeated three times. *$P<0.05$, **$P<1.0\times10^{-5}$, and ***$P<1.0\times10^{-10}$ by Student's *t*-test. n.s.: not significant. (C, D) Protein expression levels of each haplotype construct. Anti-V5 tag antibody was used in the Western blotting analysis to detect the expression of exogenous IκBε (C) and RTKN2 (D). Beta-actin expression was used as an internal control. The densities of the bands were quantified and normalized to that of the risk allele. (E, F) Allelic imbalance of expression in *NFKBIE* (E) and *RTKN2* (F). ASTQ was performed using samples from individuals heterozygous for rs2233434 (G/A) in *NFKBIE* and rs3125734 (T/C) in *RTKN2*. Genomic DNAs and cDNAs were extracted from PBMCs ($n=14$ for *NFKBIE* and $n=6$ for *RTKN2*). The y-axis shows the $\log_2$ ratio of the transcript amounts in target SNPs (risk allele/non-risk allele). The top bar of the box-plot represents the maximum value and the lower bar represents the minimum value. The top of box is the third quartile, the bottom of box is the first quartile, and the middle bar is the median value. The circle is an outlier. *$P=0.012$, **$P=0.016$, by Student's *t*-test. (G, H) Dose-dependent inhibition of *NFKBIE* (G) and activation of *RTKN2* (H) on NF-κB activity. Various doses of expression vectors carrying the non-risk allele of each gene were transfected into HEK293A cells with pGL4.32 and pRL-TK vectors.
doi:10.1371/journal.pgen.1002949.g003

located in the key regions of transcriptional regulation [26,36], warranting the prioritization of variants before evaluation by *in vitro* assays. This could also minimize false-positive results of the

in vitro assays. However, there may be additional causal variants, including rare variants, unsuccessfully selected at each step of our integrated screening. Therefore, the screening strategy
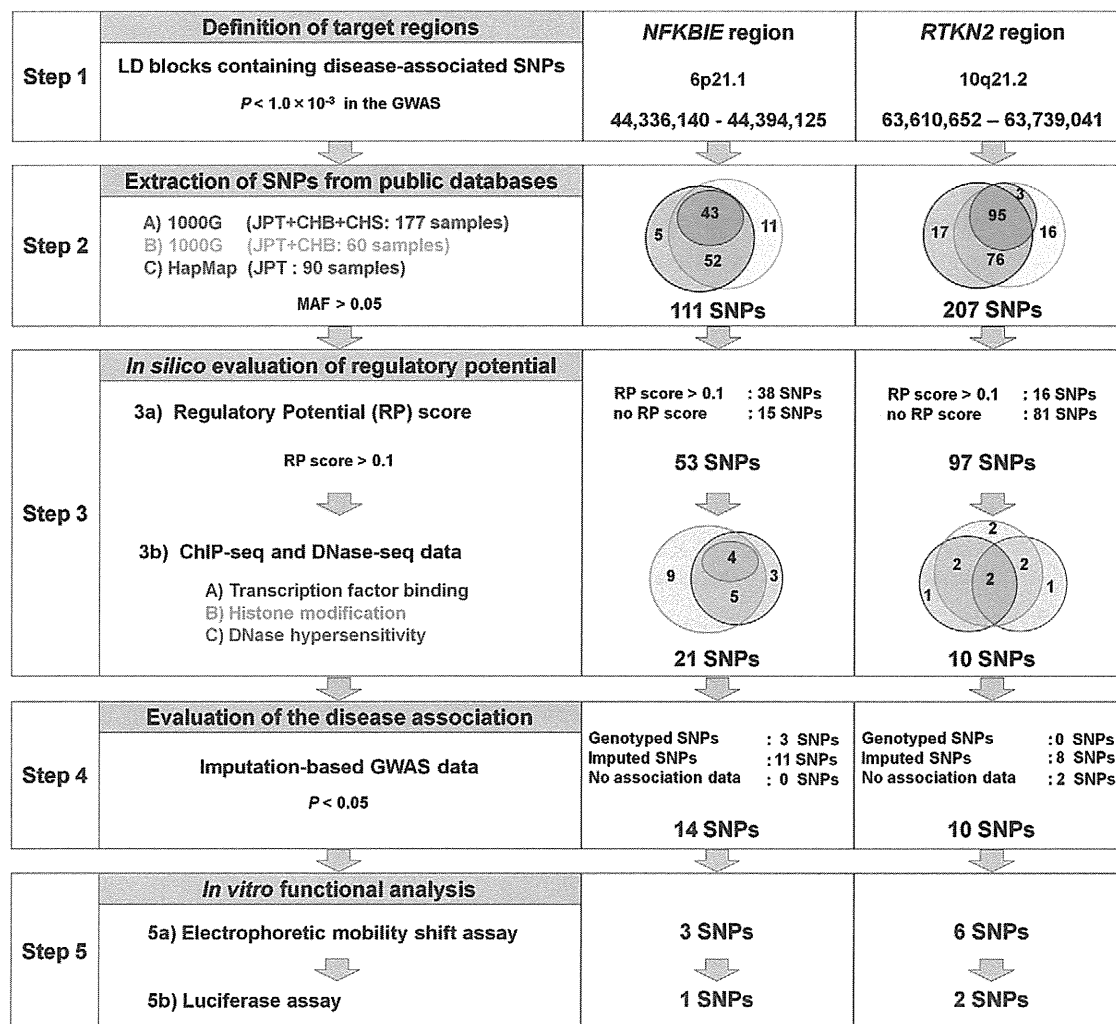
	Definition of target regions	*NFKBIE* region	*RTKN2* region
Step 1	LD blocks containing disease-associated SNPs *P* < 1.0 × 10⁻³ in the GWAS	6p21.1 44,336,140 - 44,394,125	10q21.2 63,610,652 - 63,739,041

	Extraction of SNPs from public databases	*NFKBIE* region	*RTKN2* region
Step 2	A) 1000G (JPT+CHB+CHS: 177 samples) B) 1000G (JPT+CHB: 60 samples) C) HapMap (JPT : 90 samples) MAF > 0.05	 **111 SNPs**	 **207 SNPs**

	In silico evaluation of regulatory potential	*NFKBIE* region	*RTKN2* region
Step 3	3a) Regulatory Potential (RP) score RP score > 0.1	RP score > 0.1 : 38 SNPs no RP score : 15 SNPs **53 SNPs**	RP score > 0.1 : 16 SNPs no RP score : 81 SNPs **97 SNPs**
	3b) ChIP-seq and DNase-seq data A) Transcription factor binding B) Histone modification C) DNase hypersensitivity	 **21 SNPs**	 **10 SNPs**

	Evaluation of the disease association	*NFKBIE* region	*RTKN2* region
Step 4	Imputation-based GWAS data *P* < 0.05	Genotyped SNPs : 3 SNPs Imputed SNPs : 11 SNPs No association data : 0 SNPs **14 SNPs**	Genotyped SNPs : 0 SNPs Imputed SNPs : 8 SNPs No association data : 2 SNPs **10 SNPs**

	In vitro functional analysis	*NFKBIE* region	*RTKN2* region
Step 5	5a) Electrophoretic mobility shift assay	3 SNPs	6 SNPs
	5b) Luciferase assay	1 SNPs	2 SNPs

Figure 4. Overview of SNP selection using integrated *in silico* and *in vitro* approaches. The figure shows the SNP selection process (left) and the results of *NFKBIE* (middle) and *RTKN2* (right). (Step 1) LD blocks that contain disease-associated SNPs (P_{GWAS}<1.0×10⁻³) were selected. (Step 2) SNPs were extracted from three databases (A–C). 1000G, 1000 Genome Project; HapMap, International HapMap Project. A) JPT, CHB, and CHS samples (*n* = 177) from the 1000G (the August 2010 release). B) JPT and CHB samples (*n* = 60) from the pilot 1 low coverage study data of 1000G (the March 2010 release). C) JPT samples (*n* = 90) from HapMap phase II+III (release #27). SNPs with minor allele frequency >0.05 were selected. (Step 3) Prediction of regulatory potential *in silico*. 3a) Regulatory potential (RP) scores were used for SNP selection, where an RP score >0.1 indicated the presence of regulatory elements. SNPs without RP scores were also selected. 3b) Prediction of regulatory elements by ChIP-seq data and DNase-seq data. (A) Transcription factor binding sites, (B) histone modification sites (CTCF binding, H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K9ac), and (C) DNase I hypersensitivity sites were evaluated. ChIP-seq and DNase-seq data derived from GM12878 EBV-transformed B cells were used for *NFKBIE* and *RTKN2*. DNase-seq data of Th1, Th2, and Jurkat cells were also used for *RTKN2*. (Step 4) Association data of the imputation-based GWAS using 1000G reference genotypes were used. SNPs with a significance level of *P*<0.05 were selected. SNPs without association data were also selected. (Step 5) EMSAs and luciferase assays were performed for evaluation of regulatory potentials *in vitro*.
doi:10.1371/journal.pgen.1002949.g004

should be refined as the quality and quantity of genomic databases improves in the future.

We identified multiple candidate causal variants in *NFKBIE* (two nsSNPs and one rSNP) and *RTKN2* (two rSNPs). We could not statistically distinguish the primary effect of each candidate causal variant, because these variants are in strong LD and on the same common haplotype. However, multiple causal variants could be involved in a single locus, which is also seen in another well-known autoimmune locus in 6q23 (*TNFAIP3* gene locus), where both an nsSNP and a regulatory variant have been shown to be functionally related to the disease [8,37]. The risk haplotype of nsSNPs in *NFKBIE* (rs2233433 and rs2233434) showed an enhancement of NF-κB activity, which might reflect an impaired inhibitory effect of IκB-ε on nuclear translocation of NF-κB. On the other hand, down-regulated *NFKBIE* expression and up-regulated *RTKN2* expression were observed at the risk haplotypes, which may be regulated in *cis* by the rSNPs (rs2233424 in *NFKBIE*, rs12248974 and rs61852964 in *RTKN2*). As overexpression studies have also demonstrated dose-dependent attenuation of NF-κB activity by *NFKBIE*, and dose-dependent enhancement by *RTKN2*, the *cis*-regulatory effects of these rSNPs should enhance the NF-κB activity in the risk allele. Taken together with the effect of nsSNPs in *NFKBIE*, the enhancement of NF-κB activity may play a role in the pathogenesis of the disease. This is further supported by evidence that previous GWAS for RA have also identified genes related to the NF-κB pathway, such as *TNFAIP3* [13], v-rel
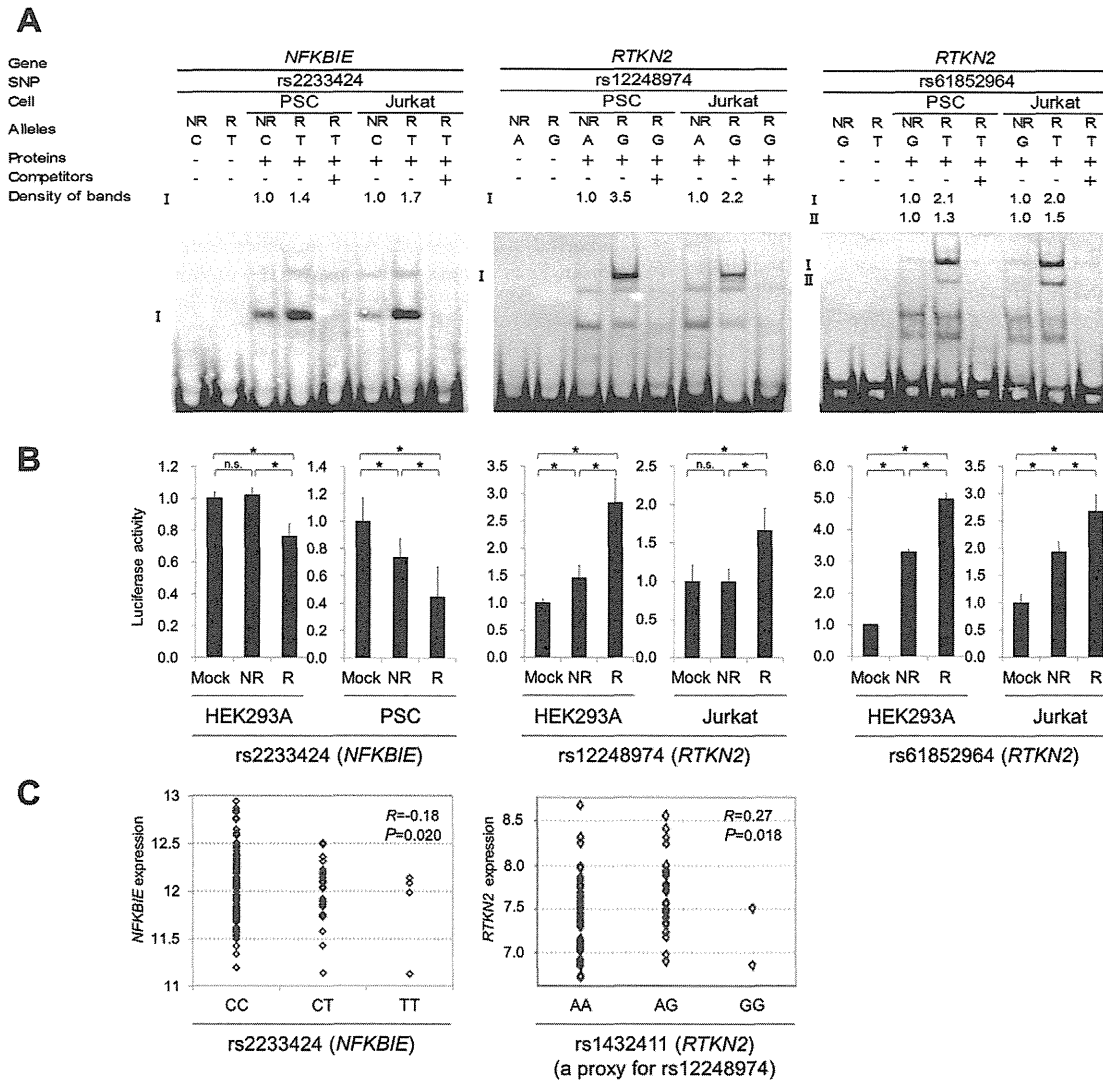
618

A

Gene		*NFKBIE*							*RTKN2*							*RTKN2*					
SNP		rs2233424							rs12248974							rs61852964					
Cell		PSC			Jurkat				PSC			Jurkat				PSC			Jurkat		
Alleles	NR R	NR	R	R	NR	R	R	NR R	NR	R	R	NR	R	R	NR R	NR	R	R	NR	R	R
	C T	C	T	T	C	T	T	A G	A	G	G	A	G	G	G T	G	T	T	G	T	T
Proteins	- -	+	+	+	+	+	+	- -	+	+	+	+	+	+	- -	+	+	+	+	+	+
Competitors	- -	-	-	+	-	-	+	- -	-	-	+	-	-	+	- -	-	-	+	-	-	+
Density of bands	I	1.0	1.4		1.0	1.7		I	1.0	3.5		1.0	2.2		I	1.0	2.1		1.0	2.0	
															II	1.0	1.3		1.0	1.5	

B

rs2233424 (*NFKBIE*) HEK293A PSC

rs12248974 (*RTKN2*) HEK293A Jurkat

rs61852964 (*RTKN2*) HEK293A Jurkat

C

rs2233424 (*NFKBIE*) $R=-0.18$ $P=0.020$ CC CT TT

rs1432411 (*RTKN2*)
(a proxy for rs12248974) $R=0.27$ $P=0.018$ AA AG GG

Figure 5. Evaluation of candidate regulatory SNPs *in vitro*. (A) Binding of nuclear factors from lymphoblastoid B-cells (PSC cells) and Jurkat cells to the 31-bp sequences around each SNP was evaluated by EMSA. Unlabeled probes in 200-fold excess as compared to the labeled probes were used for the competition experiment. The densities of the bands were quantified and normalized to that of the risk allele. rs2233424 in *NFKBIE* (C(NR)/T(R)) (left), rs12248974 (A(NR)/G(R)) (middle) and rs61852964 (G(NR)/T(R)) (right) in *RTKN2*. (B) Transcriptional activities were evaluated by luciferase assays. Each 31-bp oligonucleotide was inserted into the pGL4.24[*Luc2P*/minP] vector. Luc, luciferase; minP, minimal promoter. Transcfection was performed with HEK293A (for all the SNPs), PSC cells (for rs2233424), and Jurkat cells (for rs12248974 and rs61852964). rs2233424 (left), rs12248974 (middle), and rs61852964 (right). Data represent the mean ± s.d. Each experiment was performed in sextuplicate and independently repeated three times. *$P<0.05$ by Student's *t*-test. n.s.: not significant. (C) Liner regression analysis of the relationship between SNP genotype and gene expression level. *NFKBIE* expression data in lymphoblastoid B-cell lines of HapMap individuals (JPT+CHB, CEU and YRI; $n=151$), and *RTKN2* expression data in primary T cells from umbilical cords of Western European individuals ($n=85$) were used. The x-axis shows the SNP genotypes and the y-axis represents the $\log_2$-transformed gene expression level. R: correlation coefficient between SNP genotype and gene expression. Rs2233424 genotypes and *NFKBIE* expression level (left). The genotype classification by population: JPT+CHB, CC = 52, CT = 1; CEU, CC = 35, CT = 2; YRI, CC = 32, CT = 2, TT = 4. Rs1432411 genotypes and *RTKN2* expression level (right). Rs1432411 was used as a proxy SNP of rs12248974 ($r^2=0.97$).
doi:10.1371/journal.pgen.1002949.g005

reticuloendotheliosis viral oncogene homolog (*REL* [MIM 164910]) [5], TNF receptor-associated factor 1 (*TRAF1* [MIM 601711]) [3], and CD40 molecule TNF receptor superfamily member 5 (*CD40* [MIM 109535]) [38].

In conclusion, we identified *NFKBIE* and *RTKN2* as genetic risk factors for RA. Considering the allelic effect of both genes, enhanced NF-κB activity may play a role in the pathogenesis of the disease. Because NF-κB regulates the expression of numerous genes, including inflammatory and immune response mediators, NF-κB and its regulators identified by GWAS are promising targets for the treatment of RA.

Materials and Methods

Ethics statement

All subjects were of Japanese origin and provided written informed consent for participation in the study, which was approved by the ethical committees of the institutional review boards.

Subjects

A total of 7,907 RA cases, 657 SLE cases, 1,783 GD cases, and 35,362 control subjects were enrolled in the study through medical

institutes in Japan under the support of the BioBank Japan Project, Center for Genomic Medicine at RIKEN, the University of Tokyo, Tokyo Women's Medical University, and Kyoto University. The same case and control samples were used in the previous meta-analysis of GWASs in the Japanese population (Table S1) [15] . RA and SLE subjects met the revised American College of Rheumatology (ACR) criteria for RA [39]. Diagnosis of individuals with GD was established on the basis of clinical findings and results of the routine examinations for circulating thyroid hormone and thyroid-stimulating hormone concentrations, thyroid-stimulating hormone receptors, ultrasonography, $^{[99m]}TCO_4^-$ (or $[^{123}I]$) uptake, and thyroid scintigraphy. DNAs were extracted from peripheral blood cells using a standard protocol. Total RNAs were also extracted from PBMCs of healthy individuals ($n = 20$) using an RNeasy kit (QIAGEN, Valencia, CA, USA). Details of the samples are summarized in Table S1.

Genotyping and quality control

In the GWAS, RA cases and controls were genotyped using Illumina Human610-Quad and Illumina Human 550v3 Genotyping BeadsChips (Illumina, San Diego, CA, USA), respectively, and quality control of genotyping was performed as described previously [6]. For replication study of candidate loci, a landmark SNP was selected from each locus that satisfied $5 \times 10^{-8} < P_{GWAS} < 5 \times 10^{-5}$ in the GWAS. If multiple candidate SNPs existed within ±100 kb, the SNP with the lowest P-value was selected. All case subjects in the replication study and both case and control subjects in the validation study of candidate causal variants were genotyped using TaqMan SNP genotyping assays (Table S12) (Applied Biosystems, Foster City, CA, USA) with an ABI Prism 7900HT Sequence Detection System (Applied Biosystems). Because of the availability of DNA samples, only a part of the control subjects were genotyped for the validation study ($n = 3,290$, 97.3%). To enlarge the number of subjects and enhance statistical power for replication studies, we used genotype data obtained from other GWAS projects genotyped using the Illumina platforms for the replication control panels (Table S1). All SNPs were successfully genotyped with call rates >0.98 and were in Hardy-Weinberg equilibrium (HWE) in control subjects ($P>0.05$ as examined by χ^2 test), except for rs2233434, which displayed a deviation from HWE ($P = 0.00091$). To evaluate possible genotyping biases between the platforms, we also genotyped rs2233434 and rs3125734 by TaqMan assays for randomly selected subjects genotyped using other genotyping platforms ($n = 376$), yielding high concordance rates of ≥0.99.

Association analysis

The associations of the SNPs were tested with the Cochran-Armitage trend test. Combined analysis was performed with the Mantel-Haenszel method. Haplotype association analysis and haplotype-based conditional association analysis were performed using Haploview v4.2 and the PLINK v1.07 program (see URLs) [40], respectively. The SNPs that were not genotyped in the GWAS were imputed using MACH 1.0.16 (see URLs), with genotype data from the 1000 Genome Project (JPT, CHB, and Han Chinese South (CHS): 177 individuals) as references (August 2010 release) [41]. All the imputed SNPs demonstrated Rsq values more than 0.60.

DNA re-sequencing

Unknown variants in the coding sequences of *NFKBIE* and *RTKN2* were revealed by directly sequencing the DNA of 48 individuals affected with RA. DNA fragments were amplified with the appropriate primers (Table S13). Purification of PCR products

was performed with Exonuclease I (New England Biolabs, Ipswich, MA, USA) and shrimp alkaline phosphatase (Promega, Madison, WI, USA). The amplified DNAs were sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems), and signals were detected using an ABI 3700 DNA Analyzer (Applied Biosystems).

Construction of haplotype-specific expression vectors

The full coding regions were amplified using cDNAs prepared from an Epstein-Barr virus-transfected lymphoblastoid B-cell line (Pharma SNP Consortium (PSC), Osaka, Japan) for *NFKBIE* (NM_004556.2) and from Jurkat cells (American Type Culture Collection (ATCC), Rockville, MD, USA) for *RTKN2* (NM_145307.2) with appropriate primers (Table S14) and DNA polymerases. PCR products were inserted into the pcDNA3.1D/V5-His-TOPO vector (Invitrogen, Camarillo, CA, USA) using the TaKaRa Ligation kit ver. 2.1 (Takara Bio Inc, Shiga, Japan), and mutagenized using the AMAP Multi Site-Directed Mutagenesis Kit (MBL, Nagoya, Japan). Each construct was then transformed into Jet Competent *Escherichia coli* cells (DH5α) (BioDynamics Laboratory Inc., Tokyo, Japan). These plasmids were purified using an Endofree Plasmid Maxi Kit (QIAGEN) after confirmation of the sequence.

NF-κB reporter assay

Human embryonic kidney (HEK) 293A cells (Invitrogen) were cultured in Dulbecco's modified Eagle's medium (Sigma-Aldrich, St. Louis, MO, USA) supplemented with 10% fetal bovine serum (BioWest, Nuaillé, France), 1% penicillin/streptomycin (Invitrogen), and 0.1 mM MEM Non-Essential Amino Acids (Invitrogen). Various doses of the haplotype-specific expression vector (0.0025–0.02 µg for *NFKBIE* and 0.1–0.8 µg for *RTKN2*), pGL4.32[*luc2P*/NF-κB-RE/Hygro] vector (Promega) (0.05 µg and 0.0125 µg, respectively), and pRL-TK vector (an internal control for transfection efficiency) (0.45 µg and 0.15 µg, respectively) were transfected into the HEK293A cells using the Lipofectamine LTX transfection reagent (Invitrogen) according to the manufacturer's protocol. The total amounts of DNAs were adjusted with empty pcDNA3.1 vector. After 22 h, cells were incubated with 1 ng/ml TNF-α (Sigma) for 2 h or with medium alone. Cells were collected, and luciferase activity was measured using a Dual-Luciferase Reporter Assay system (Promega) and a GloMax-Multi+ Detection System (Promega). Each experiment was independently repeated three times, and sextuplicate samples were assayed each time.

Western blotting

After 24 h of transfection as described for the NF-κB reporter assay, cells were lysed in NP-40 lysis buffer (150 mM NaCl, 1% NP-40, 50 mM Tris-HCl at pH 8.0, and a protease inhibitor cocktail), and incubated on ice for 30 min. After centrifugation, the supernatant fraction was collected and 4×Sodium dodecyl sulfate (SDS) sample buffer was added. After denaturation at 95°C for 5 min, proteins were analyzed by SDS-polyachrylamide gel electrophoresis (PAGE) on a 5% to 20% gradient gel (Wako, Osaka, Japan) and were transferred to polyvinylidene difluouride (PVDF) membranes (Millipore, Billerica, MA, USA). Target proteins on the membrane were probed with antibodies (mouse anti-V5 tag (Invitrogen), anti-β-actin-HRP (an internal control), and goat anti-mouse IgG2a-HRP (Santa Cruz Biotechnology, Santa Cruz, CA, USA)), visualized using enhanced chemiluminescence (ECL) detection reagent (GE Healthcare, Pollards Wood, UK), and detected using a LAS-3000 mini lumino-image analyzer

(Fujifilm, Tokyo, Japan). Band intensities were measured using MultiGauge software (Fujifilm).

Allele-specific transcript quantification (ASTQ) analysis

ASTQ analysis was performed as previously described [42]. Total RNAs and genomic DNAs were extracted from PBMCs and lymphoblastoid B-cell lines. cDNAs were synthesized using TaqMan reverse transcription reagents (Applied Biosystems). We selected SNPs (rs2233434 (A/G) for *NFKBIE* and rs3125734 (C/T) for *RTKN2*) as target SNPs. Allele-specific gene expression was measured by TaqMan SNP genotyping probes for these SNPs (Applied Biosystems). To make a standard curve, we selected two individuals that had homozygous genotypes of each target SNP. We mixed these DNAs at nine different ratios and detected the intensities. The $\log_2$ of (risk allele/non-risk allele intensity) for each SNP was plotted against the $\log_2$ of mixing homozygous DNAs. We generated a standard curve (linear regression line; y = ax+b), where y is the $\log_2$ of (risk allele/non-risk allele intensity) at a given mixing ratio, x is the $\log_2$ of the mixing ratio, a is the slope, and b is the intercept. We then measured the allelic ratio for each cDNA and genomic DNA from each individual by real-time TaqMan PCR. Based on a standard curve, we calculated the allelic ratio of cDNAs and genomic DNAs. Intensities were detected using an ABI Prism 7900HT Sequence Detection System (Applied Biosystems).

Electrophoretic mobility shift assays (EMSA)

EMSA and preparation of nuclear extract from lymphoblastoid B-cell lines and Jurkat cells were performed as previously described [43]. Cells were cultured in RPMI-1640 medium (Sigma-Aldrich) supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. Following stimulation with 50 ng/ml phorbol myristate acetate (Sigma-Aldrich) for 2 h, cells were collected and suspended in buffer A (20 mM HEPES at pH 7.6, 20% glycerol, 10 mM NaCl, 1.5 mM $MgCl_2$, 0.2 mM EDTA at pH 8.0, 1 mM DTT, 0.1% NP-40, and a protease inhibitor cocktail) for 10 min on ice. After centrifugation, the pellets were resuspended in buffer B (which contains buffer A with 500 mM NaCl). Following incubation on ice for 30 min and centrifugation to remove cellular debris, the supernatant fraction containing nuclear proteins was collected. Oligonucleotides (31-bp) were designed that corresponded to genomic sequences surrounding the SNPs (Table S15). Single-stranded oligonucleotide probes were labeled using a Biotin 3′ End DNA Labeling Kit (Pierce Biotechnology, Rockford, IL, USA), and sense and antisense oligonucleotides were then annealed. DNA-protein interactions were detected using a LightShift Chemiluminescent EMSA kit (Pierce Biotechnology). The DNA-protein complexes were separated on a non-denaturing 5% polyachrylamide gel in 1×TBE (Tris-borate-EDTA) running buffer for 60 min at 150 V. The DNA-protein complexes were then transferred from the gel onto a nitrocellulose membrane (Ambion, Carlsbad, CA, USA), and were cross-linked to the membrane by exposure to UV light. Signals were detected using a LAS-3000 mini lumino-image analyzer (Fujifilm). Allelic differences were analyzed using MultiGauge software (Fujifilm) by measuring the intensity of the bands.

Luciferase assay

Oligonucleotides (31-bp) were designed as described for the EMSAs (Table S15), and complementary sense and antisense oligonucleotides were annealed. To construct luciferase reporter plasmids, pGL4.24[*luc2P*/minP] vector (Promega) was digested with restriction enzymes (XhoI and BglII) (Takara Bio Inc), and annealed oligonucleotide was ligated into a pGL4.24 vector

upstream of the minimal promoter. HEK293A ($n = 2.5 \times 10^5$), lymphoblastoid B-cell lines ($n = 2.0 \times 10^6$) and Jurkat ($n = 5.0 \times 10^5$) cells were transfected with the allele-specific constructs (0.4 µg, 1.8 µg and 2.5 µg, respectively) and the pRL-TK vector (0.1 µg, 0.2 µg and 0.25 µg, respectively) using the Lipofectamine LTX transfection reagent (for HEK293A and Jurkat cells) and Amaxa nucleofector kit (Lonza, Basel, Switzerland) (for lymphoblastoid B-cell lines). Cells were collected, and luciferase activity was measured as described for the NF-κB reporter assay. Each experiment was independently repeated three times and sextuplicate samples were assayed each time.

Correlation analysis between gene expression and genotypes

The expression data in lymphoblastoid B-cell lines derived from HapMap individuals ($n = 210$; JPT, CHB, CEU, and YRI) and in primary T cells from umbilical cords of Western European individuals ($n = 85$) from the database of the Gene Expression Variation (Genevar) project were used. SNP genotypes were obtained from HapMap and 1000 Genome Project databases. The expression levels were regressed with the genotype in a liner model. The statistical significance of regression coefficients was tested using Student's *t*-test.

Statistical analysis

We used χ^2 contingency table tests to evaluate the significance of differences in allele frequency in the case-control subjects. We defined haplotype blocks using the solid spine of LD definition of Haploview v4.2, and estimated haplotype frequency and calculated pairwise LD indices (r^2) between pairs of polymorphisms using the Haploview program. Luciferase assay data and ASTQ analysis data were analyzed by Student's *t*-test.

Web resources

The URLs for data presented herein are as follows:

PLINK, http://pngu.mgh.harvard.edu/~purcekk/plink
MACH, http://www.sph.umich.edu/csg/abecasis/mach/
UCSC Genome Browser, http://genome.ucsc.edu/;
Genevar, http://www.sanger.ac.uk/resources/software/genevar/
HapMap Project, http://www.HapMap.org/
1000 Genome Project, http://www.1000genomes.org
Online Mendelian Inheritance in Man (OMIM), http://www.omim.org/

Supporting Information

Figure S1 NF-kB activity was influenced by nsSNPs in *NFKBIE*. NF-κB activities were evaluated by luciferase assays. Allele specific construct, pGL4.32[*luc2P*/NF-κB-RE] luciferase vector, and pRL-TK vector were transfected into HEK293A cells. Four haplotypes (rs2233434-rs2233433; A-C, G-C, A-T, and G-T) were examined. (rs2233434: A = non-risk (NR), G = risk (R); rs2233433: C = NR, T = R). Twenty-two hours after transfection, cells were stimulated with medium alone (A) or TNF-α (B) for 2 h. Data represent the mean ± s.d. Each experiment was performed in sextuplicate, and experiments were independently repeated three times. *$P<0.05$ and **$P<1.0\times10^{-5}$ by Student's *t*-test. n.s.: not significant. (TIF)

Figure S2 Allelic imbalance of expression in *NFKBIE*. ASTQ was performed using samples from individuals heterozygous for rs2233434 (G/A) in *NFKBIE*. Genomic DNAs and cDNAs were extracted from lymphoblastoid B cells ($n = 9$). The y-axis shows the $\log_2$ ratio of the transcript amounts in target SNPs (risk allele/non-risk

allele). The top bar of the box-plot represents the maximum value and the lower bar represents the minimum value. The top of box is the third quartile, the bottom of box is the first quartile, and the middle bar is the median value. The circle is an outlier. $*P = 5.3 \times 10^{-4}$ by Student's t-test.
(TIF)

Figure S3 SNP selection using *in silico* analysis in the *NFKBIE* region. Step 1: Definition of the target region. P-values of the SNPs in the GWAS (top) and genomic structure (middle), and the D'-based LD map (bottom). The green diamond shapes represent the $-\log_{10}$ of the Cochran-Armitage trend P-values. The dashed line indicates the significance threshold ($P < 1 \times 10^{-3}$). The LD map was drawn based on genotype data of the 1000 Genome Project (JPT, CHB and CHS: 177 samples) using Haploview software v4.2. LD blocks were defined by the solid spine method. The red box (top) represents the target region of the *in silico* analysis (Chr6: 44,336,140-44,394,125). Step 2: Target SNPs were extracted from public databases (HapMap and 1000 Genome Project). SNPs with MAF >0.05 were selected. Step 3: Evaluation of regulatory potential. Step 3a: The regulatory potential (RP) score was calculated for sequences surrounding the SNPs by ESPERR (evolutionary and sequence pattern extraction through reduced representations) method. SNPs with RP score >0.1 were selected. Step 3b: Subsequently, SNPs within the predicted, regulatory genomic elements were selected by using ChIP-seq data of transcription factor binding sites (Txn factor), histone modification sites (CTCF binding, H3K4me1, H3K4me2, H3K4me3, H3K27ac, H3K9ac) or DNase-seq data of DNase I hypersensitivity sites (DNase HS). ChIP-seq data and DNase-seq data used the signals derived from GM12878 EBV-transformed B cells. All these analyses of Steps 2 to 3 were performed by using the UCSC genome browser. Step 4: Evaluation of disease association. Association data of both genotyped (green diamonds) and imputed (black diamonds) SNPs in the GWAS samples were used. Red triangles represent 14 extracted SNPs *in silico*. The dashed line indicates the significance threshold ($P < 0.05$).
(TIF)

Figure S4 SNP selection using *in silico* analysis in the *RTKN2* region. SNP selection in the *RTKN2* region was performed the same as in the case of the *NFKBIE* region as described in Figure S3, except that we used DNase-seq data derived from Th1, Th2, and Jurkat cells in addition to GM12878 EBV-transformed B cells.
(TIF)

Figure S5 Results of EMSAs for candidate regulatory SNPs. Binding affinities of nuclear factors from lymphoblastoid B-cells (PSC cells) and Jurkat cells to the 31-bp sequences around each allele of the candidate regulatory SNPs were evaluated by EMSA. Nuclear factors from PSC cells were used for *NFKBIE*, and Jurkat cells were used for *RTKN2*. 14 SNPs in *NFKBIE* (A) and 10 SNPs in *RTKN2* (B) were tested. NR: non-risk allele; R: risk allele. Arrows indicate bands showing allelic differences in each SNP.
(TIF)

Figure S6 Luciferase assays for regulatory SNPs. Transcriptional activities of the 31-bp genomic sequences around the SNPs were evaluated by luciferase assays. Each oligonucleotide was inserted into the pGL4.24[*luc2P*/minP] vector upstream of the minimal promoter (minP), and allele-specific constructs were transfected into HEK293A cells. Relative luciferase activity is expressed as the ratio of luciferase activity of each allele-specific construct to the luciferase activity of the mock construct. Data represent the mean ± s.d. Each experiment was independently repeated three times, and each sample was measured in sextuplicate. $*P < 1 \times 10^{-3}$ by

Student's t-test. n.s.: not significant. (A) rs2233434 and rs77986492 in the *NFKBIE* region. (B) rs3864793, rs1864836, rs4979765, and rs4979766 in the *RTKN2* region. NR: non-risk allele; R: risk allele. (TIF)

Figure S7 The correlation between *NFKBIE* expression and rs2233434 and rs77986492 genotypes. Linear regression analysis of the relationship between SNP genotypes and *NFKBIE* expression. Gene expression data from EBV-transformed lymphoblastoid B cell lines of HapMap individuals (JPT+CHB, CEU, and YRI). (A) rs2233434 ($n = 204$) and (B) rs77986492 ($n = 152$). The genotype classification by population: rs2233434 (JPT+CHB, AA = 61, AG = 28, GG = 1; CEU, AA = 52, AG = 2; YRI, AA = 53, AG = 72) and rs77986492 (JPT+CHB, CC = 52, CT = 24; CEU, CC = 35, CT = 2; YRI, CC = 38, CT = 1). The x-axis shows SNP genotypes and the y-axis represents the $\log_2$-transformed *NFKBIE* expression level. R: the correlation coefficient between *NFKBIE* expression and SNP genotype.
(TIF)

Figure S8 The correlation between *RTKN2* expression and rs3852694 genotypes. Linear regression analysis of the relationship between the rs3852694 genotype and *RTKN2* expression. Rs3852694 was used as a proxy SNP of rs1864836 ($r^2 = 1.0$). Gene expression data in primary T cells from umbilical cords of Western European individuals ($n = 85$) were presented by using Genevar software. The x-axis shows the rs3852694 genotypes (AA, AG, GG) and the y-axis represents the $\log_2$-transformed *RTKN2* expression level. R: the correlation coefficient between *RTKN2* expression and rs3852694 genotype.
(TIF)

Table S1 Summary of samples.
(DOC)

Table S2 Association results of the GWAS and 1st replication study.
(DOC)

Table S3 Association analysis of *NFKBIE* and *RTKN2* with autoimmune diseases.
(DOC)

Table S4 Association analysis of nsSNPs with RA.
(DOC)

Table S5 Haplotype association study of nsSNPs in *NFKBIE*.
(DOC)

Table S6 Haplotype association study of nsSNPs in *RTKN2*.
(DOC)

Table S7 Predicting the effects of nsSNPs on protein function.
(DOC)

Table S8 Association analysis of candidate rSNPs with RA.
(DOC)

Table S9 Haplotype association study of candidate causal SNPs in *NFKBIE*.
(DOC)

Table S10 Haplotype association study of candidate causal SNPs in *RTKN2*.
(DOC)

Table S11 The conditional haplotype-based association analysis of candidate causal SNPs in *RTKN2*.
(DOC)

Table S12 Probes and Primers used for TaqMan assays.
(DOC)

Table S13 Primers used for DNA re-sequencing.
(DOC)

Table S14 Primers used for construction of expression vectors.
(DOC)

Table S15 Oligonucleotides used for EMSAs and Luciferase assays.
(DOC)

Author Contributions

Conceived and designed the experiments: K Myouzen, Y Kochi, Y Okada, C Terao, K Ikari, K Ohmura, R Yamada, K Yamamoto. Performed the experiments: K Myouzen, Y Kochi, C Terao, A Suzuki, K Ikari, K Ohmura. Analyzed the data: K Myouzen, Y Kochi, Y Okada, C Terao, T Tsunoda, A Takahashi, R Yamada. Contributed reagents/materials/analysis tools: M Kubo, A Taniguchi, F Matsuda, K Ohmura, S Momohara, T Mimori, H Yamanaka, N Kamatani, Y Nakamura. Wrote the paper: K Myouzen, Y Kochi, Y Okada, C Terao, K Yamamoto.

References

1. Gabriel SE (2001) The epidemiology of rheumatoid arthritis. Rheum Dis Clin North Am 27: 269–281
2. Suzuki A, Yamada R, Chang X, Tokuhiro S, Sawada T, et al. (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. Nat Genet 34: 395–402
3. Plenge RM, Seielstad M, Padyukov L, Lee AT, Remmers EF, et al. (2007) TRAF1-C5 as a risk locus for rheumatoid arthritis–a genomewide study. N Engl J Med 357: 1199–1209
4. Wellcome Trust Case Control Consortium (2007) Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 447: 661–678
5. Gregersen PK, Amos CI, Lee AT, Lu Y, Remmers EF, et al. (2009) REL, encoding a member of the NF-kappaB family of transcription factors, is a newly defined risk locus for rheumatoid arthritis. Nat Genet 41: 820–823
6. Kochi Y, Okada Y, Suzuki A, Ikari K, Terao C, et al. (2010) A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. Nat Genet 42: 515–519
7. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, et al. (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 75: 330–337
8. Adrianto I, Wen F, Templeton A, Wiley G, King JB, et al. (2011) Association of a functional variant downstream of TNFAIP3 with systemic lupus erythematosus. Nat Genet 43: 253–258
9. Thomas PD, Kejariwal A (2004) Coding single-nucleotide polymorphisms associated with complex vs. Mendelian disease: evolutionary evidence for differences in molecular effects. Proc Natl Acad Sci U S A 101: 15398–15403
10. Okada Y, Shimane K, Kochi Y, Tahira T, Suzuki A, et al. (2012) A Genome-Wide Association Study Identified AFF1 as a Susceptibility Locus for Systemic Lupus Eyrthematosus in Japanese. PLoS Genet 8: e1002455. doi:10.1371/journal.pgen.1002455
11. Dubois PC, Trynka G, Franke L, Hunt KA, Romanos J, et al. (2010) Multiple common variants for celiac disease influencing immune gene expression. Nat Genet 42: 295–302
12. 1000 Genomes Project Consortium (2010) A map of human genome variation from population-scale sequencing. Nature 467: 1061–1073
13. Plenge RM, Cotsapas C, Davies L, Price AL, de Bakker PI, et al. (2007) Two independent alleles at 6q23 associated with risk of rheumatoid arthritis. Nat Genet 39: 1477–1482
14. Remmers EF, Plenge RM, Lee AT, Graham RR, Hom G, et al. (2007) STAT4 and the risk of rheumatoid arthritis and systemic lupus erythematosus. N Engl J Med 357: 977–986
15. Okada Y, Terao C, Ikari K, Kochi Y, Ohmura K, et al. (2012) Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet 45: 511–516
16. Li Z, Nabel GJ (1997) A new member of the I kappaB protein family, I kappaB epsilon, inhibits RelA (p65)-mediated NF-kappaB transcription. Mol Cell Biol 17: 6184–6190
17. Whiteside ST, Epinat JC, Rice NR, Israel A (1997) I kappa B epsilon, a novel member of the I kappa B family, controls RelA and cRel NF-kappa B activity. Embo J 16: 1413–1426
18. Collier FM, Gregorio-King CC, Gough TJ, Talbot CD, Walder K, et al. (2004) Identification and characterization of a lymphocytic Rho-GTPase effector: rhotekin-2. Biochem Biophys Res Commun 324: 1360–1369
19. Collier FM, Loving A, Baker A, J., McLeod J, Walder K, et al. (2009) RTKN2 Induces NF-kappaB Dependent Resistance to Intrinsic Apoptosis in HEK cells and Regulates BCL-2 Gene in Human CD4+ Lymphocytes. J Cell Death 2: 9–23
20. Makarov SS (2001) NF-kappa B in rheumatoid arthritis: a pivotal regulator of inflammation, hyperplasia, and tissue destruction. Arthritis Res 3: 200–206
21. Kolbe D, Taylor J, Elnitski L, Eswara P, Li J, et al. (2004) Regulatory potential scores from genome-wide three-way alignments of human, mouse, and rat. Genome Res 14: 700–707
22. Taylor J, Tyekucheva S, King DC, Hardison RC, Miller W, et al. (2006) ESPERR: learning strong and weak signals in genomic sequence alignments to identify functional elements. Genome Res 16: 1596–1604
23. Johnson DS, Mortazavi A, Myers RM, Wold B (2007) Genome-wide mapping of in vivo protein-DNA interactions. Science 316: 1497–1502
24. Valouev A, Johnson DS, Sundquist A, Medina C, Anton E, et al. (2008) Genome-wide analysis of transcription factor binding sites based on ChIP-Seq data. Nat Methods 5: 829–834
25. Mikkelsen TS, Ku M, Jaffe DB, Issac B, Lieberman E, et al. (2007) Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. Nature 448: 553–560
26. Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, et al. (2011) Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 473: 43–49
27. Sabo PJ, Kuehn MS, Thurman R, Johnson BE, Johnson EM, et al. (2006) Genome-scale mapping of DNase I sensitivity in vivo using tiling DNA microarrays. Nat Methods 3: 511–518
28. Dimas AS, Deutsch S, Stranger BE, Montgomery SB, Borel C, et al. (2009) Common regulatory variation impacts gene expression in a cell type-dependent manner. Science 325: 1246–1250
29. Yang TP, Beazley C, Montgomery SB, Dimas AS, Gutierrez-Arcelus M, et al. (2010) Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies. Bioinformatics 26: 2474–2476
30. Stranger BE, Forrest MS, Dunning M, Ingle CE, Beazley C, et al. (2007) Relative impact of nucleotide and copy number variation on gene expression phenotypes. Science 315: 848–853
31. Stranger BE, Nica AC, Forrest MS, Dimas A, Bird CP, et al. (2007) Population genomics of human gene expression. Nat Genet 39: 1217–1224
32. Stahl EA, Raychaudhuri S, Remmers EF, Xie G, Eyre S, et al. (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. Nat Genet 42: 508–514
33. Chu X, Pan CM, Zhao SX, Liang J, Gao GQ, et al. (2011) A genome-wide association study identifies two new risk loci for Graves' disease. Nat Genet 43: 897–901
34. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, et al. (2011) Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet 43: 1193–1201
35. Li Y, Sidore C, Kang HM, Boehnke M, Abecasis GR (2011) Low-coverage sequencing: implications for design of complex trait association studies. Genome Res 21: 940–951
36. Degner JF, Pai AA, Pique-Regi R, Veyrieras JB, Gaffney DJ, et al. (2012) DNase I sensitivity QTLs are a major determinant of human expression variation. Nature 482: 390–394
37. Musone SL, Taylor KE, Lu TT, Nititham J, Ferreira RC, et al. (2008) Multiple polymorphisms in the TNFAIP3 region are independently associated with systemic lupus erythematosus. Nat Genet 40: 1062–1064
38. Raychaudhuri S, Remmers EF, Lee AT, Hackett R, Guiducci C, et al. (2008) Common variants at CD40 and other loci confer risk of rheumatoid arthritis. Nat Genet 40: 1216–1223
39. Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, et al. (1988) The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. Arthritis Rheum 31: 315–324
40. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81: 559–575
41. Li Y, Willer C, Sanna S, Abecasis G (2009) Genotype imputation. Annu Rev Genomics Hum Genet 10: 387–406

623

42. Akamatsu S, Takata R, Ashikawa K, Hosono N, Kamatani N, et al. (2010) A functional variant in NKX3.1 associated with prostate cancer susceptibility down-regulates NKX3.1 expression. Hum Mol Genet 19: 4265–4272

43. Andrews NC, Faller DV (1991) A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. Nucleic Acids Res 19: 2499

13

624

東大病院ゲノム医学センターにおける取り組み
パーキンソン病のパーソナルゲノム解析を中心に

三井　純　石浦浩之　辻　省次

特 集 次世代シーケンサーによる神経変性疾患の解析と展望

東大病院ゲノム医学センターにおける取り組み
パーキンソン病のパーソナルゲノム解析を中心に

Present Efforts in the Medical Genome Center at the University of Tokyo Hospital

三 井 　 純[1]　石 浦 浩 之[1]　辻 　 省 次[1,2]

Jun Mitsui[1], Hiroyuki Ishiura[1], Shoji Tsuji[1,2]

Abstract

Technologies associated with massively parallel sequencing have evolved rapidly over the last several years, making it possible to cost-effectively sequence the whole human genome and exome in a short period of time. These technologies are expected to bring about a better understanding of genetic components underlying monogenic diseases, as well as diseases inherited in a non-Mendelian fashion. They will eventually cause a paradigm shift in clinical practice, where the diagnosis and decision-making for appropriate therapeutic procedures is based on the "personal genome". In this review, we outline some of our recent efforts in the Medical Genome Center at the University of Tokyo Hospital, including an identification of the causative gene for a Mendelian disease (posterior column ataxia with retinitis pigmentosa), an approach to uncover susceptible genes for a non-Mendelian disease (Parkinson disease), and an application of exome sequencing for the molecular diagnosis of a disease with vast genetic heterogeneity (hereditary diffuse leukoencephalopathy with spheroids). We also discuss the advantages and limitations of these emerging technologies.

Key words : massively parallel sequencing, Parkinson disease, common disease- multiple rare variants, personal genome

はじめに

次世代シーケンサーと総称される大規模並列 DNA シーケンシング技術は，最近数年の間に加速的に進化しており，処理速度の向上，コストの低下が進んでいる。次世代シーケンサーが臨床遺伝学にもたらすインパクトとして以下の3つの事柄が挙げられる。

第1は，メンデル遺伝性疾患の原因遺伝子解明が進むことである。連鎖解析による絞り込みを十分に行うことが難しい小さな家系サイズの遺伝性疾患，*de novo* 変異などで生じる重篤で生殖適応度が低い遺伝性疾患など，従来の技術ではアプローチが困難だった遺伝性疾患の解明が期待される。実際，このようなメンデル遺伝性疾患の原因遺伝子の報告がここ数年で急速に増加している。問題点としては，現在普及している次世代シーケンサーでは，ひとつながりで配列決定できる塩基長（リード長）が高々100塩基程度であり，トリプレットリピート病などに代表される繰り返し配列の延長や挿入変異の検出がしばしば困難なことである。特に遺伝性神経変性疾患ではこの種類の変異が多く知られており，現在の次世代シーケンサーの技術的課題の1つである。

第2は，孤発性疾患の遺伝因子の解明が期待されることである。従来は一塩基多型（single nucleotide polymorphism：SNP）をマイクロアレイ上で大規模にタイピングする技術を利用して，患者群と対照群で多型の頻

度を比較することで疾患と関連する感受性遺伝子探索が行われてきた。候補となる遺伝子・領域だけではなく，全ゲノム上の多型を広範囲に探索できることから，このアプローチは全ゲノム関連解析（genome-wide association study：GWAS）と呼ばれ，多くの疾患で検討が行われた。新たな発見も多かったが，孤発性疾患の遺伝因子の大部分が解明できるのではないかという期待には届かず，まだ解明されていない遺伝因子（missing heritability）が残されている[1]。

多型マーカーと連鎖不平衡にある疾患感受性アレルを関連解析で検出する手法は，比較的少数の創始者に由来する疾患感受性アレルが，患者群に広く分布するという構造を持つ集団（common disease-common variants仮説）に対しては強い検出力を示すが，多数の独立した疾患感受性アレルが個々には稀に患者群に分布するという集団の遺伝的構造（common disease-multiple rare variants仮説）に対しては検出が困難になる。また，多型タイピングでは検出できないコピー数変異などの構造変異が寄与している可能性もある。今後，孤発性疾患における遺伝因子の解明を進めていくためには，パーソナルゲノム解析に基づく網羅的な変異の同定が大きな手掛かりになるであろう。

第3に，臨床における遺伝子診断の汎用化が挙げられる。神経内科領域の臨床では遺伝性疾患の占める割合が相対的に高く，需要も高いことから普及が期待される。特に原因遺伝子が多様な表現型・疾患群の遺伝子診断において高い効果を発揮するであろう。問題点としては，上述のように遺伝性神経筋疾患にみられる繰り返し配列の延長（優性遺伝性脊髄小脳変性症の多く，歯状核赤核淡蒼球ルイ体萎縮症，ハンチントン病，球脊髄性筋萎縮症，筋強直性ジストロフィー，フリードライヒ運動失調症，9p21に連鎖する筋萎縮性側索硬化症・前頭側頭型認知症，眼咽頭筋ジストロフィーなど）の検出は短いリード長では困難であり，フラグメント解析やサザン・ブロッティング解析を併用する必要がある。また，現状ではコスト・パフォーマンスの点からエクソーム解析が選択されることが多いと考えられるが，コピー数変異（遺伝性神経疾患では*APP*，*SNCA*，*PMP22*，*MPZ*などのコピー数変異による遺伝性疾患が報告されている）や大きな欠失・重複変異（デュシェンヌ・ベッカー型筋ジストロフィーにおける*DMD*や常染色体劣性遺伝若年性パーキンソニズムにおける*PARK2*の欠失・重複変異など）において，エクソーム解析では検出力が十分でない可能性があり適応に注意が必要である。

本稿では，以上3点について概説し，いくつかの具体例を挙げる。最後に2011年度に東京大学医学部附属病院の新たな組織として発足したゲノム医学センターの紹介と今後の展望を述べる。

Ⅰ．メンデル遺伝性疾患へのアプローチ

連鎖解析による候補領域の絞り込みが十分行えず，従来の技術ではアプローチが困難だった小さな家系でも，次世代シーケンサーによってアプローチできる可能性がある。網膜色素変性症を伴う後索運動失調（posterior column ataxia with retinitis pigmentosa：PCARP）は，常染色体劣性遺伝性・小児期発症で，その名のとおり後索性の運動失調症と網膜色素変性症を特徴とする極めて稀な神経変性疾患である。北米とスペインの家系の連鎖解析から1q31-q32に連鎖することが知られており，原因遺伝子は長らく不明であったが[2,3]，2010年末にエクソーム解析によって*FLVCR1*が原因遺伝子であることが報告された[4]。筆者らは，この原因遺伝子が報告される以前にPCARPの1家系について検討する機会があったので，その解析の経緯を紹介する[5]。

家系はFig.1に示すように両親いとこ婚で，発症者2例，非発症者1例である。Affymetrix社の50K Xba/Hindアレイを用いてSNPジェノタイピングを行い，SNPによる連鎖解析のために開発したパイプラインSNP-HiTLinkを用いて[6]，常染色体劣性遺伝性・完全浸透モデルでパラメトリック多点解析を行った。その結果，染色体1番（1q32.1-q41，約13Mbの範囲）と染色体20番（20p12.1-p11.23，約16Mbの範囲）に高いLODスコア（染色体上のマーカーと病原性変異が連鎖しているかどうかを統計学的に推定するための値。最もスコアの高い領域に病原性変異が存在する可能性が高い）を示す領域が得られ，特に染色体1番の領域は，これまでに報告されていたPCARP連鎖領域とオーバーラップしていた。

次に，この候補領域の配列をキャプチャーするためロシュ・ニンブルジェン社のプラットフォームでカスタムアレイを作製し，候補領域をエンリッチしたDNAサンプルをイルミナ社のGenome Analyzer Ⅱxを用いて100塩基・シングルリードでシーケンスした。得られたリードはbwa（Burrows-Wheeler Aligner）のデフォルト条件で参照配列（NCBI36/hg18）に対してアライメントを行い[7]，SAMtoolsを用いて変異の検出を行った[8]。染色体1番の候補領域から，13,616個の変異が検出され，そのうち翻訳領域には60個の変異が認められた。PCARPが極めて稀な疾患であることから，原因と
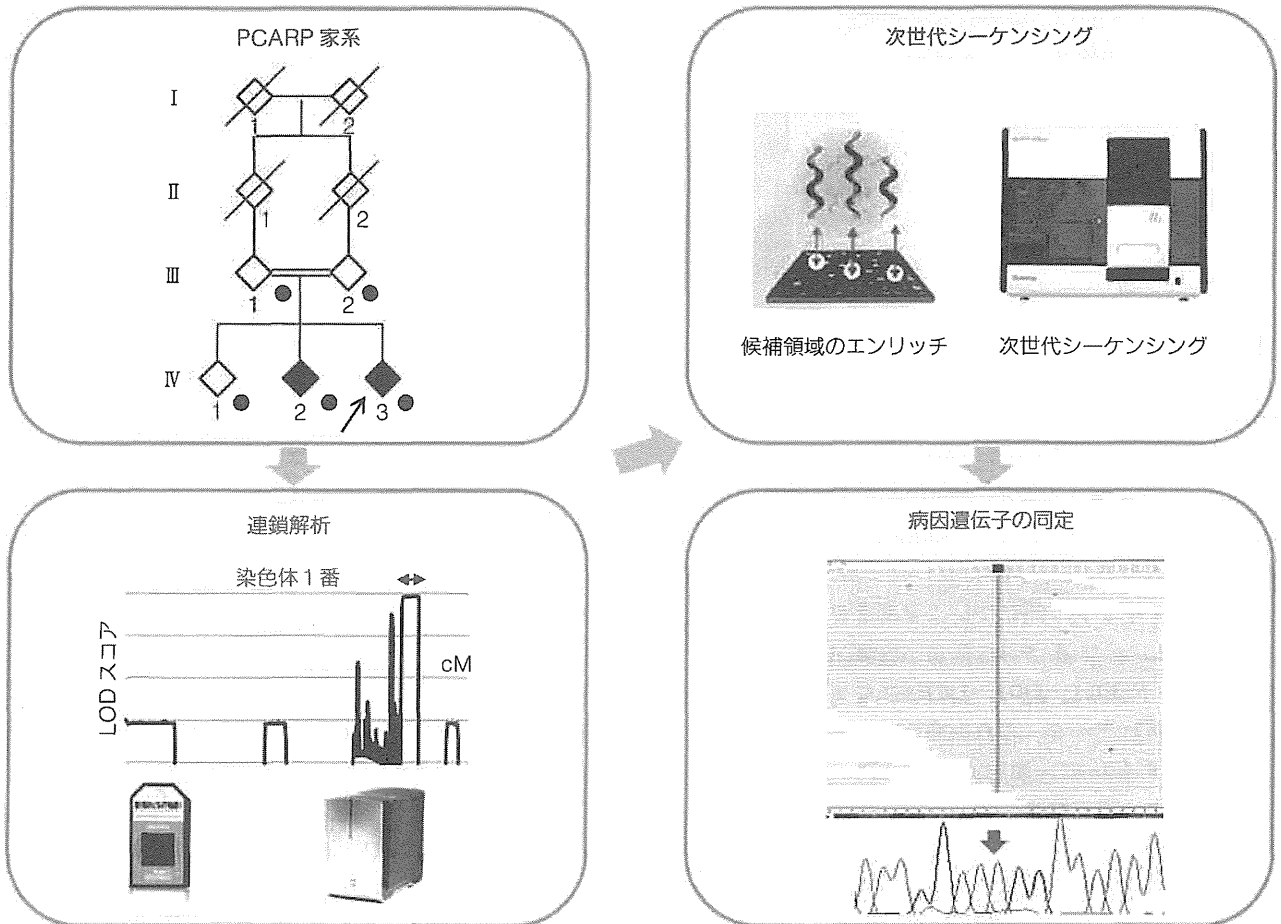
Fig. 1 メンデル遺伝性神経疾患の解明
〔略語〕PCARP：網膜色素変性症を伴う後索運動失調

なる変異は変異データベースに登録されていない可能性が高いと考え，翻訳領域の変異のうち dbSNP131 に登録されていない変異を検索したところ 5 個，そのうちアミノ酸置換を伴う変異は 4 個に絞られた。さらに，これまでに報告されていた PCARP 連鎖領域とオーバーラップする領域に絞り込むと，*FLVCR1* にある G493R ホモ接合性変異のみが残された。

　ここまで検討を進め，*FLVCR1* が PCARP の原因遺伝子である可能性が高いと考えられたが，単独の家系ではこれ以上の証明が困難であり，原因遺伝子と断定するためには別の複数家系による報告を待たねばならなかった[4]。次世代シーケンサーによって，候補領域における変異探索の労力は大幅に軽減されたが，その一方で正確な臨床診断に基づいてリソースを収集する臨床家の役割はこれまでと変わらず極めて重要なものである。

　シーケンサー側の課題としては，現在主流の次世代シーケンサーが高々100 塩基のショートリードによる配列解析であるため，繰り返し配列の延長や挿入変異の検出が難しい点が挙げられる。最近注目を集めている 9p21 に連鎖する筋萎縮性側索硬化症・前頭側頭型認知症における *C9orf71* イントロン部分の 6 塩基繰り返し配列延長変異[9,10]など，頻度の高い重要な疾患遺伝子がまだ見つけられていない可能性もある。数塩基繰り返し配列にターゲットを絞った配列アセンブラーの開発，PacBio RS システムなどによるロングリード配列解析の併用など検討を重ねている。

II．孤発性神経変性疾患の遺伝因子解明に向けたアプローチ

1．これまでの GWAS の成果と残された課題

　孤発性神経疾患の 1 例として，パーキンソン病を挙げる。パーキンソン病は振戦，筋強剛，寡動，姿勢反射障害などの運動障害を主な症候学的特徴とし，認知障害や自律神経障害などさまざまな随伴症状を呈する進行性の神経変性疾患である。神経変性疾患としてはアルツハイ
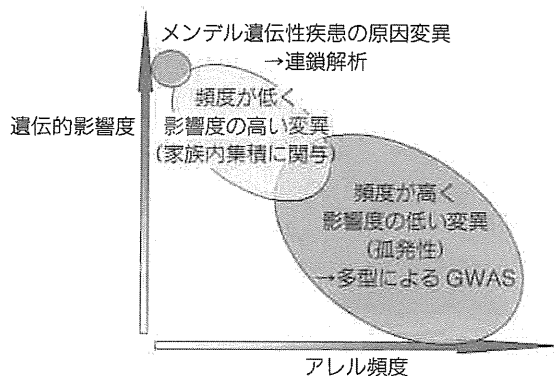
メンデル遺伝性疾患の原因変異
→連鎖解析

遺伝的影響度

頻度が低く
影響度の高い変異
(家族内集積に関与)

頻度が高く
影響度の低い変異
(孤発性)
→多型によるGWAS

アレル頻度

Fig. 2　孤発性神経疾患の遺伝因子解明に向けて

マー病に次いで頻度が高く，60歳以上の高齢者においては1％程度の頻度であり[11]，高齢化とともに患者数の急増が予想される。

家族性パーキンソン病については家系に対する連鎖解析によって，αシヌクレイン（SNCA），パーキン（PARK2），PINK1，DJ-1，LRRK2などが原因遺伝子として報告されている。これら原因遺伝子の機能解析により，αシヌクレイン蛋白の過剰発現・病的蓄積，ミトコンドリアの機能障害，ユビキチン・プロテアソーム系の障害，酸化ストレスなどが孤発性パーキンソン病とも共通する病態機序として推測され，病態解明・治療法開発に向けた研究が行われている[12]。しかし現時点での治療法の中心は，中脳黒質・線条体系のドパミン低下に対する補充療法にとどまり，長期的な病態の進行を抑制する方法がない。大部分を占める孤発性パーキンソン病の遺伝因子が解明されれば，病態自体の進行を抑制する新たな治療法の開発が可能になるのではないかと期待されている。

孤発性パーキンソン病の遺伝因子の大きさについては議論があるが，これまでの遺伝疫学研究からは遺伝因子の大きさを表す遺伝率を60％前後と推定している[13,14]。これまでの多型をマーカーにしたGWASはヨーロッパ集団からの報告が大部分であるが，SNCA，LRRK2，MAPT，PARK16，BST1，GAK，HLA-DRB5などを含む遺伝子座が報告されている[15-20]。日本人を対象にした大規模GWASでは，ヨーロッパ集団で報告されているMAPTが検出されなかった一方，新たな遺伝子座PARK16，BST1が検出された事実は，集団によって疾患の遺伝因子の構造が異なっていることを反映していて興味深い[17]。一方，SNCAなどは日本人集団でもヨーロッパ集団でも共通して再現性よく関連が確認されている。

昨年，ヨーロッパ集団におけるGWASをメタ解析した最大規模の研究が報告された[21]。このメタ解析では，これまで報告されていたSNCA，MAPT，LRRK2，PARK16，BST1，HLA，GAKに加えて，新たにSYT11，ACMSD，STK39，MCCC1/LAMP3，CCDC62/HIP1Rがゲノムワイド水準の有意差をもって確認され，これら11遺伝子座に基づく人口寄与リスク割合（population attribution risk：PAR）を合計で60.3％と推定している。PARはリスクアレルで説明できる疾患患者中の割合を指し，これまで報告されているパーキンソン病の遺伝率に近い数字に達していることは注目される。

もちろん，有意となった遺伝子座のみを集めてくるための確認バイアスや，検出されたマーカーと真の機能性変異の連鎖不平衡の程度があるため，この推定値は過剰に評価されている可能性がある。今後とも，遺伝因子構造の異なるさまざまな人種集団における再現性の確認，遺伝子間相互作用や環境因子に関する評価などさまざまな点を考慮する必要があるが，頻度の高い疾患感受性変異に限れば，GWASによって網羅されつつある状況であると考えられる。

一方，残されている課題としてパーキンソン病の家族内集積性の問題が挙げられる。一般にパーキンソン病患者の5〜10％程度に家族歴があることが知られており，家族内集積性に関するメタ解析によると，患者の一親等の発症危険率は3倍程度と推定されている[22,23]。ところが，GWASで検出されている疾患感受性変異のオッズ比は高々1.5倍程度であり，家族内集積性のほとんどが説明できていない[21]。筆者らは，この missing heritability について，頻度が低く疾患への影響度が強い変異が未解明のまま残されていると推測している（Fig. 2）。すなわち，頻度が低すぎてGWASでは検出が難しい一方，浸透率が低すぎて連鎖解析の適応とならないような変異が家族内集積性に寄与していると考えている。このような特徴を持つ遺伝因子の具体例について，以下にグルコセレブロシダーゼ遺伝子を例に述べていく。

2. 孤発性パーキンソン病とグルコセレブロシダーゼ遺伝子

グルコセレブロシダーゼ（glucocerebrosidase）遺伝子（GBA）はライソゾームに局在する加水分解酵素グルコセレブロシダーゼをコードし，常染色体劣性遺伝性の先天代謝疾患ゴーシェ病の原因遺伝子である。ゴーシェ病患者は日本国内で100万人に3人程度，病原性変異のキャリア頻度は300人に1人程度と推測される稀な疾患である[24]。2004年，Goker-Alpanら[25]により，

Table 同定された GBA 病原性変異の種類と頻度

変異	パーキンソン病患者群の頻度	対照者群の頻度	フィッシャーの正確確率検定	オッズ比（95%信頼区間）
R120W	15/534	0/544	$P < 0.0001$	
L444P-A456P-V460V（RecNciI）	14/534	2/544	$P = 0.0020$	7.3（1.7〜66.4）
L444P	8/534	0/544	$P = 0.0035$	
N188S	4/534	0/544	有意差なし	
R329C	2/534	0/544	有意差なし	
R496C	2/534	0/544	有意差なし	
R120W-N188R-V191G-S196P-F213I	1/534	0/544	有意差なし	
R131C	1/534	0/544	有意差なし	
G193W	1/534	0/544	有意差なし	
F213I	1/534	0/544	有意差なし	
A456P-V460V	1/534	0/544	有意差なし	
合計	50/534（9.4%）	2/544（0.37%）	$P = 6.9 \times 10^{-14}$	28.0（7.3〜238.3）

ゴーシェ病患者の血縁者に何世代かにわたりパーキンソン病を発症する家系がいくつか報告された。これは GBA 変異のキャリアが，パーキンソン病発症の危険因子になる可能性を示唆しており，家系内で複数の発症者が観察されるということは影響度が強い可能性を示唆している。同年，Aharon-Peretz ら[26]により，アシュケナージ系ユダヤ人を対象に GBA の特定の変異をスクリーニングした関連解析が行われ，キャリア頻度のオッズ比 7.0 倍と高い関連を示す報告がなされた。しかし，特定の変異のみをスクリーニングしている点，サンプル数が比較的少ない点，GBA 変異の分布・頻度が特異なユダヤ人種を対象にしている点で，一般化できるかどうか解釈を困難にしていた。

これらの問題を克服するため，筆者らと神戸大学 戸田達史教授らの研究グループ（Japanese Parkinson Disease Susceptibility Gene Consortium）との共同研究で，日本人大規模サンプルに対して GBA の全エクソン配列解析にて網羅的に変異をスクリーニングする患者-対照関連解析を行った[27]。サンプル規模はパーキンソン病患者 534 例，対照者 544 例で，合計 27 種類の変異を同定した。うち 11 種類はゴーシェ病の原因変異として報告のある変異（病原性変異），16 種類はゴーシェ病の病原性変異として報告のない変異（非病原性変異）であった。個々の変異別に検討すると，R120W，L444P，RecNciI の 3 種類はキャリア頻度がパーキンソン病群で有意に高かった（それぞれ $P < 0.0001$，0.0035，0.0020）。これら 3 種類の変異はすべて病原性変異であるため，病原性変異に注目したところ，R120W，L444P，RecNciI 以外の 8 種類の病原性変異は，全例パーキンソン病患者群でのみ認められ，対照群には認めなかった（Table）。合計するとパーキンソン

病患者 534 例中 50 例（9.4%），対照者 544 例中 2 例（0.37%）で病原性変異をヘテロ接合性に認めた。全病原性変異のパーキンソン病患者に対するオッズ比は 28.0 倍（95%信頼区間 7.3〜238.3）となり，P 値は 6.9×10^{-14} と有意であった。一方，非病原性変異では個別でも合計してもパーキンソン病群と対照群の間に有意な関連は認めなかった。

次に，GBA 病原性変異キャリアであるパーキンソン病患者 50 例の臨床的特徴について検討した。発症年齢を検討したところ，GBA 病原性変異キャリア群の平均発症年齢は 52.5 歳で，非キャリア群の 58.8 歳と比べて有意に若年化していた。GBA 病原性変異キャリア 50 例のうち 49 例については臨床情報を入手可能だった。49 例中 41 例（83.7%）では抗パーキンソン病薬（レボドパ，ドパミン受容体アゴニスト）が有効であった。33 例でパーキンソン病の補助診断に用いられる [123I]-MIBG（metaiodobenzylguanidine）心筋シンチグラフィー検査が施行されていた。検査が施行されていた 33 例中 29 例（87.9%）では心筋への取り込みが低下しており，一般的なパーキンソン病の特徴に矛盾ないと考えられた。49 例中 13 例（26.5%）で認知障害（clinical dementia scale 1 以上[28]），17 例（34.7%）で幻視をきたしていた。非キャリア群の臨床的特徴が十分得られなかったため，有意かどうか結論できなかったが，キャリア群では認知障害や幻視の頻度が高い可能性があると考えられた。

さらに同年，筆者らも参加した国際多施設共同解析（北米 4 施設，南米 1 施設，アジア 3 施設，イスラエル 2 施設，ヨーロッパ 6 施設）の結果が発表され，GBA 変異がパーキンソン病発症に関与する遺伝因子であることが多くの人種で確認された[29]。この解析でも，GBA

変異キャリアの発症年齢は非キャリアに比べて若年化していることが確認され，さらに認知症を発症する頻度が高いことなどが示された。

3．*GBA* 変異の家族内集積性への関与

このように *GBA* 病原性変異は，パーキンソン病発症に対して非常に強い影響度を持つ遺伝因子であることがわかり，パーキンソン病の家族内集積性の一部を説明できるのではないかと考えられている。実際，筆者らの検討でも，*GBA* 病原性変異キャリア群の家族歴に注目したところ，50例中11例（22.0%）で両親または同胞の血縁者にパーキンソン病患者が1人ずついることがわかった。そのうち3例については，発症者のDNA解析が可能であったため解析を行ったところ，家系内の発症者全例で発端者の持つ変異と同じ変異を共有していることがわかった[27]。さらに，両親または同胞内にパーキンソン病発症者がいる多発家系34家系について *GBA* の全エクソン配列解析を行ったところ，34家系中5家系（14.7%）で *GBA* 病原性変異を認め，家系内でDNA解析が可能な全サンプルにおいて，発症者は *GBA* 病原性変異を共有し，非発症者では変異を認めないことが示された。

筆者らの少数例の検討では完全な共分離が確認されたが，発端者が *GBA* 変異を持つパーキンソン病多発家系21家系の共分離を調べた最近の報告では，非発症者の46%が発端者と同じ *GBA* 変異を共有する一方，家系内の発症者の17%が *GBA* 変異を共有していないことが報告されている[30]。このことは，*GBA* 変異を持っていても100%発症するわけではないこと，同じ家系内でも遺伝的異質性が一定程度存在することを示している。

それでは，*GBA* 変異を持っている人のどの程度がパーキンソン病を発症するのであろうか。*GBA* 変異キャリアにおけるパーキンソン病の年齢別累積発症頻度を調べた報告によると，50歳で7.6%，60歳で13.7%，70歳で21.4%，80歳で29.7%と推定されている[31]。このことから，*GBA* 変異は，メンデル遺伝性疾患ほどの強い浸透率ではないが，GWASで検出されている疾患感受性変異と比べるとはるかに強い影響度を持っているといえる。

4．パーソナルゲノム解析に基づくパーキンソン病における疾患遺伝子研究

GBA 変異のような特徴を持つ，稀で影響度の強い遺伝因子は，パーキンソン病を含め多くの疾患で未解明の状態にあると考えられる。頻度の高い多型をマーカーにしているGWASでは検出されていないこの領域へのアプローチは，パーソナルゲノム解析に基づく網羅的な変異の同定が前提となるであろう。しかし，頻度の低い変異について変異ごとに関連を検定する従来のGWAS的手法だと，検出される変異の種類が膨大なものとなることから，個別の変異における関連の検出力がかなり低くなってしまう。このことから，稀な変異の関連解析に際しては，変異をグループ化して検定する統計的な手法がいくつか提案されている[32-38]。しかしこれまでのところ，具体的な成功例や標準的といえるアプローチがなく，筆者らも試行錯誤を続けている状況である。

一般論として，グループ化した変異において中立的な変異や，影響する方向の異なる変異（疾患感受性変異と保護的な変異）が混在している可能性があり，それらによって検出力が低下してしまう懸念があることから，機能的な変異を選択してグループ化するようなアルゴリズムを設定する方向性が考えられる。また，既知の病態パスウェイから候補遺伝子を絞り込んだり，重みづけをしたりする方法も考えられる。あるいは，未解明の遺伝因子の多くは，比較的影響度が強いことが予想されることから，孤発性サンプルではなく，同朋例や親子例など家族内集積性のあるサンプルに集中して解析することも有効かもしれない。筆者らはこのようなアプローチのもと，パーキンソン病，多系統萎縮症（multiple system atrophy：MSA）などの神経変性疾患について解析を進めている。

III．遺伝子診断への適用とその役割

従来の遺伝子診断と本質的な違いはないが，エクソーム解析・全ゲノム解析では，まず膨大な変異のリストが得られ，その中から既知の原因遺伝子に変異がないか検索するという流れになる。従来は，臨床像などから可能性の高い候補遺伝子に絞って解析を行っていたが，たとえ可能性が低い候補遺伝子でも網羅的に解析できるという利点がある。当科では，遺伝形式・臨床像などからカテゴリーを設けて，検索すべき遺伝子のリストを作成し，エクソーム解析による遺伝子診断の標準化を目指している。神経筋疾患に限定しても，新たな原因遺伝子の発見は続いており，候補遺伝子リストの質の維持が重要である。

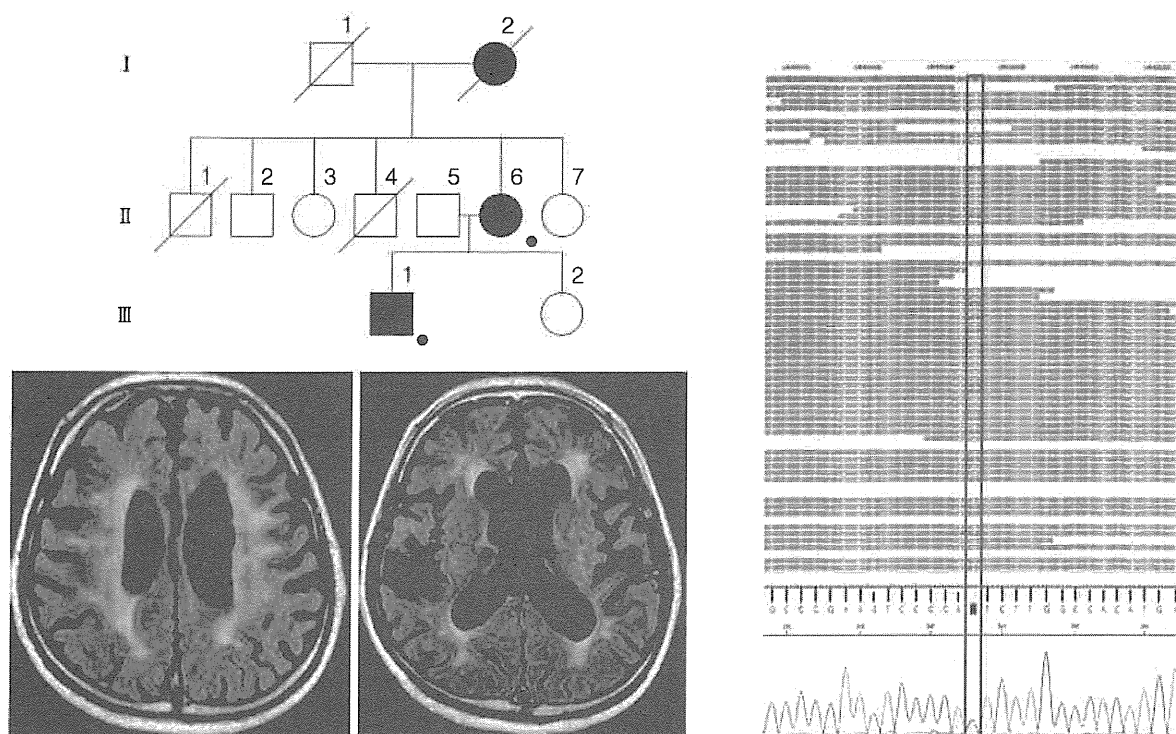本稿では，優性遺伝性の白質脳症・認知症のカテゴリーの中から遺伝子診断がついた例について紹介する。このようなカテゴリーでは，遺伝性脳小血管病（*NOTCH3，COL4A1，TREX1*），ラミンB1遺伝子の

Fig. 3　遺伝子診断への具体的応用
優性遺伝性の家系で，40〜50歳代発症の認知症。頭部 MRI にて白質病変を認めた。鑑別診断による候補遺伝子の中から，*CSF1R* 変異を検出した。

重複変異，アレクサンダー病（*GFAP*），遺伝性脳アミロイド血管症（*APP*，*PSEN1*，*PSEN2*，*CST3*，*ITM2B*），さらに2011年末にスフェロイドを伴う遺伝性びまん性白質脳症（hereditary diffuse leukoencephalopathy with axonal spheroids：HDLS）の原因として報告された *CSF1R* などが挙げられる。筆者らは，白質病変・脳萎縮を伴い，進行性の認知障害をきたす原因遺伝子未同定家系についてエクソーム解析を行い，これらの候補遺伝子変異の有無について検索した。その結果，2つの家系で *CSF1R* 遺伝子のチロシンキナーゼ領域にミスセンス変異が同定され，HDLS の診断に至った（**Fig. 3**）。なお，他の候補遺伝子には，病原性を疑う変異は認められなかった。

　HDLS は常染色体優性遺伝性で白質病変を特徴とする成人発症の非常に稀な神経変性疾患であり，中枢神経病理でスフェロイドが多数みられることが特徴である。その臨床症状は多様であり，性格や行動の変化，認知障害，抑うつ，パーキンソニズムなど運動障害，痙攣などが報告されている。また画像所見も初期には虚血性変化でみられるような非特異的所見である。HDLS の臨床所見が多様で非特異的であること，また優性遺伝性の白質脳症・認知症のカテゴリーに属する遺伝性疾患が多様

であることから，従来はこのような状況の遺伝子診断は難しく，エクソーム解析による遺伝子診断が有用であったと考えられる。

　エクソーム解析による遺伝子診断はコストの点からも現実化しつつあり，近い将来，一般的な臨床検査に近づくのではないかと考えられる。もちろんトリプレットリピート病をはじめとして，ショートリードのシーケンサーでは検出が難しい変異もあるため，解析方法の特徴と限界を知ったうえで，臨床カテゴリーに応じて適応を判断する必要がある。また，検査が汎用化する過程で，膨大な変異データの解釈をどのように標準化して，診療に役立てていくのかは非常に大きな課題である。例えば，同定された変異が病原性を持つかどうかという判断に限っても，さまざまな状況がありうる。当該変異で病原性を持つという報告が既にされていたり，ナンセンス変異やフレームシフト変異など翻訳される蛋白質の機能に大きな影響を与える変異であったりすれば，病原性の判断は比較的容易かもしれない。しかし，例えばミスセンス変異では，変異の起こった蛋白質ドメインの機能，置換されたアミノ酸の性質，種間の保存性，一般集団における変異の頻度など，病原性の判断についてさまざまな条件が考えられる。

おわりに

　数年前から起こっている大規模シーケンシング技術の飛躍的な進歩は目覚ましく，ヒトのエクソーム解析や全ゲノム配列解析はすでに低コストかつ簡便に行えるようになっている。シーケンシングの低コスト化・大規模並列化のスピードに対して，情報解析のハード面・ソフト面のインフラが追いついておらず，ボトルネックになりつつある。今後，爆発的に増大するパーソナルゲノム情報を医学研究・診療に役立てていくためには，情報解析の基盤づくりは必須である。計算機リソースのハード面の拡充に加えて，日本人における参照配列整備，変異データベースの構築・共有など，ソフト面でも早急な整備が望まれる。さらに，これまで述べたように疾患の遺伝的研究のみならず，診療に果たす役割も大きくなることが予想される。臨床にフィードバックする際には，検査としての信頼度管理，変異解釈の標準化，さらに意図しない偶発的な所見の取り扱いなど，大きな課題が山積している。

　東京大学医学部附属病院では，このような背景のもと，病院内の共同利用施設としてゲノム医学センター・ゲノム解析部門を 2011 年度に発足させた。複数台の次世代シーケンサー（Hiseq2000 2 台，5500 XL 1 台，PacBio RS 1 台，GAIIx 1 台，Miseq 1 台）およびサンプル調製自動化ロボットを含む周辺機器，さらに大規模計算機と安全性に配慮したネットワークシステムを擁する病院のコア施設の 1 つである。パーソナルゲノム情報を遺伝的研究および診療に有効に役立てていくためには，情報を安全に管理するしくみと集約化・共有化が必要になっていくであろう。

文　献

1) Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al: Finding the missing heritability of complex diseases. Nature 461: 747-753, 2009

2) Higgins JJ, Morton DH, Loveless JM: Posterior column ataxia with retinitis pigmentosa (AXPC1) maps to chromosome 1q31-q32. Neurology 52: 146-150, 1999

3) Higgins JJ, Kluetzman K, Berciano J, Combarros O, Loveless JM: Posterior column ataxia and retinitis pigmentosa: a distinct clinical and genetic disorder. Mov Disord 15: 575-578, 2000

4) Rajadhyaksha AM, Elemento O, Puffenberger EG, Schierberl KC, Xiang JZ, et al: Mutations in FLVCR1 cause posterior column ataxia and retinitis pigmentosa. Am J Hum Genet 87: 643-654, 2010

5) Ishiura H, Fukuda Y, Mitsui J, Nakahara Y, Ahsan B, et al: Posterior column ataxia with retinitis pigmentosa in a Japanese family with a novel mutation in FLVCR1. Neurogenetics 12: 117-121, 2011

6) Fukuda Y, Nakahara Y, Date H, Takahashi Y, Goto J, et al: SNP HiTLink: a high-throughput linkage analysis system employing dense SNP data. BMC Bioinformatics 10: 121, 2009

7) Li H, Durbin R: Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25: 1754-1760, 2009

8) Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, et al; 1000 Genome Project Data Processing Subgroup: The Sequence Alignment/Map format and SAMtools. Bioinformatics 25: 2078-2079, 2009

9) DeJesus-Hernandez M, Mackenzie IR, Boeve BF, Boxer AL, Baker M, et al: Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. Neuron 72: 245-256, 2011

10) Renton AE, Majounie E, Waite A, Simón-Sánchez J, Rollinson S, et al: A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. Neuron 72: 257-268, 2011

11) de Lau LM, Breteler MM: Epidemiology of Parkinson's disease. Lancet Neurol 5: 525-535, 2006

12) Saiki S, Sato S, Hattori N: Molecular pathogenesis of Parkinson's disease: update. J Neurol Neurosurg Psychiatry 83: 430-436, 2012

13) Hamza TH, Payami H: The heritability of risk and age at onset of Parkinson's disease after accounting for known genetic risk factors. J Hum Genet 55: 241-243, 2010

14) Moilanen JS, Autere JM, Myllylä VV, Majamaa K: Complex segregation analysis of Parkinson's disease in the Finnish population. Hum Genet 108: 184-189, 2001

15) Edwards TL, Scott WK, Almonte C, Burt A, Powell EH, et al: Genome-wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for Parkinson disease. Ann Hum Genet 74: 97-109, 2010

16) Hamza TH, Zabetian CP, Tenesa A, Laederach A, Montimurro J, et al: Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease. Nat Genet 42: 781-785, 2010

17) Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, et al: Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet 41: 1303-1307, 2009

18) Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, et al: Genome-wide association study

reveals genetic risk underlying Parkinson's disease. Nat Genet 41: 1308-1312, 2009

19) Saad M, Lesage S, Saint-Pierre A, Corvol JC, Zelenika D, et al; French Parkinson's Disease Genetics Study Group: Genome-wide association study confirms *BST1* and suggests a locus on 12q24 as the risk loci for Parkinson's disease in the European population. Hum Mol Genet 20: 615-627, 2011

20) UK Parkinson's Disease Consortium; Wellcome Trust Case Control Consortium 2, Spencer CC, Plagnol V, Strange A, Gardner M, Paisan-Ruiz C, et al: Dissection of the genetics of Parkinson's disease identifies an additional association 5' of *SNCA* and multiple associated haplotypes at 17q21. Hum Mol Genet 20: 345-353, 2011

21) International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin UM, et al: Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet 377: 641-649, 2011

22) Thacker EL, Ascherio A: Familial aggregation of Parkinson's disease: a meta-analysis. Mov Disord 23: 1174-1183, 2008

23) Shino MY, McGuire V, Van Den Eeden SK, Tanner CM, Popat R, et al: Familial aggregation of Parkinson's disease in a multiethnic community-based case-control study. Mov Disord 25: 2587-2594, 2010

24) Ida H, Rennert OM, Kawame H, Maekawa K, Eto Y: Mutation prevalence among 47 unrelated Japanese patients with Gaucher disease: dentification of four novel mutations. J Inherit Metab Dis 20: 67-73, 1997

25) Goker-Alpan O, Schiffmann R, LaMarca ME, Nussbaum RL, McInerney-Leo A, et al: Parkinsonism among Gaucher disease carriers. J Med Genet 41: 937-940, 2004

26) Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R: Mutations in the glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med 351: 1972-1977, 2004

27) Mitsui J, Mizuta I, Toyoda A, Ashida R, Takahashi Y, et al: Mutations for Gaucher disease confer high susceptibility to Parkinson disease. Arch Neurol 66: 571-576, 2009

28) Morris JC: The Clinical Dementia Rating (CDR): current version and scoring rules. Neurology 43: 2412-2414, 1993

29) Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, et al: Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med 361: 1651-1661, 2009

30) Lesage S, Anheim M, Condroyer C, Pollak P, Durif F, et al; French Parkinson's Disease Genetics Study Group: Large-scale screening of the Gaucher's disease-related glucocerebrosidase gene in Europeans with Parkinson's disease. Hum Mol Genet 20: 202-210, 2011

31) Anheim M, Elbaz A, Lesage S, Durr A, Condroyer C, et al; French Parkinson Disease Genetic Group: Penetrance of Parkinson disease in glucocerebrosidase gene mutation carriers. Neurology 78: 417-420, 2012

32) Li B, Leal SM: Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data. Am J Hum Genet 83: 311-321, 2008

33) Price AL, Kryukov GV, de Bakker PI, Purcell SM, Staples J, et al: Pooled association tests for rare variants in exon-resequencing studies. Am J Hum Genet 86: 832-838, 2010

34) Madsen BE, Browning SR: A groupwise association test for rare mutations using a weighted sum statistic. PLoS Genet 5: e1000384, 2009

35) Liu DJ, Leal SM: A novel adaptive method for the analysis of next-generation sequencing data to detect complex trait associations with rare variants due to gene main effects and interactions. PLoS Genet 6: e1001156, 2010

36) Wu MC, Lee S, Cai T, Li Y, Boehnke M, et al: Rare-variant association testing for sequencing data with the sequence kernel association test. Am J Hum Genet 89: 82-93, 2011

37) Basu S, Pan W: Comparison of statistical tests for disease association with rare variants. Genet Epidemiol 35: 606-619, 2011

38) Wang Y, Chen YH, Yang Q: Joint rare variant association test of the average and individual effects for sequencing studies. PLoS One 7: e32485, 2012