

Original article

RAvariome: a genetic risk variants database for rheumatoid arthritis based on assessment of reproducibility between or within human populations

Yoko Nagai¹ and Tadashi Imanishi^{1,2,*}

¹Department of Molecular Life Science, Division of Basic Medical Science and Molecular Medicine, Tokai University School of Medicine, 143 Shimokasuya, Isehara, Kanagawa 259-1193, Japan and ²Data Management and Integration Team, Molecular Profiling Research Center for Drug Discovery, National Institute of Advanced Industrial Science and Technology, Koto-ku, Tokyo 135-0064, Japan

*Correspondence author: Tel: +81 463 93 1121; Fax: +81 463 93 5418; Email: imanishi@tokai-u.jp

Submitted 14 August 2013; Revised 24 September 2013; Accepted 25 September 2013

Citation details: Nagai Y. and Imanishi T. RAvariome: a genetic risk variants database for rheumatoid arthritis based on assessment of reproducibility between or within human populations. *Database*, Vol. 2013, Article ID bat073, doi:10.1093/database/bat073.

Rheumatoid arthritis (RA) is a common autoimmune inflammatory disease of the joints and is caused by both genetic and environmental factors. In the past six years, genome-wide association studies (GWASs) have identified many risk variants associated with RA. However, not all associations reported from GWASs are reproduced when tested in follow-up studies. To establish a reliable set of RA risk variants, we systematically classified common variants identified in GWASs by the degree of reproducibility among independent studies. We collected comprehensive genetic associations from 90 papers of GWASs and meta-analysis. The genetic variants were assessed according to the statistical significance and reproducibility between or within nine geographical populations. As a result, 82 and 19 single nucleotide polymorphisms (SNPs) were confirmed as intra- and inter-population-reproduced variants, respectively. Interestingly, majority of the intra-population-reproduced variants from European and East Asian populations were not common in two populations, but their nearby genes appeared to be the components of common pathways. Furthermore, a tool to predict the individual's genetic risk of RA was developed to facilitate personalized medicine and preventive health care. For further clinical researches, the list of reliable genetic variants of RA and the genetic risk prediction tool are provided by open access database RAvariome.

Database URL: <http://hin.v.jp/hinv/rav/>

Introduction

Rheumatoid arthritis (RA; MIM180300) is a common autoimmune disease characterized by the chronic inflammation of the bones and joints. Several epidemiological studies reported that RA prevalence varies among different populations (1). In North America and Northern Europe, the estimated prevalence of RA is 0.5–1.1%, but in Southern Europe, a lower prevalence of 0.3–0.7% has been reported. In East Asia, RA prevalence in the urban areas of Japan and Taiwan is 1.04% and 0.93%, respectively, but that in

mainland China ranges from 0.2% to 0.37% (2, 3). Twin studies on RA have led to a heritability estimate for RA of 65% in the Finnish study and 53% in the UK study, and genetic factors account for an estimated 60% of the disease risk (4).

Large numbers of association studies have been conducted, and numerous single nucleotide polymorphisms (SNPs), genes and chromosomal regions have been reported to be associated with RA. Several SNP-trait association databases provide information about RA-associated loci, including the NHGRI Catalog of Published Genome-Wide

Association Studies (GWASs) (5), GWASdb (6), GWAS Central (7), HuGE Navigator (8) and PharmGKB (9). On the other hand, some databases have been developed to provide detailed genetic information of a specific disease, e.g. AlzGene (<http://www.alzgene.org/>) for Alzheimer's disease, CADgene (<http://www.bioguo.org/CADgene/>) for coronary artery disease and GOLD.db (<https://gold.tugraz.at/>) for lipid-associated disorders. There are a few databases focusing on autoimmune diseases; type 1 diabetes database T1Dbase (<http://www.t1dbase.org>) and Lymphoproliferative syndrome database ALPSbase (<http://research.nhgri.nih.gov/ALPS/>).

As the first human genetic variation database for RA, RAvariome aims to provide a reliable set of RA risk variants that was systematically assessed according to its reproducibility. Candidate gene association studies and GWASs are known to be vulnerable to a range of errors and biases, especially those arising from differences in the experimental and study designs (10). These problems are reflected in equivocal or inconsistent results and may lead researchers to design inappropriate follow-up studies or medical applications. Accordingly, we have collected association studies comprehensively, classified the studies by ethnicity of subjects, re-evaluated the associations by unified significance level and assessed the associations by population-based reproducibility. All data are publicly available in a regularly updated web database, RAvariome. Additionally, an online tool for predicting the genetic risk of RA for an individual was developed to support further analysis for preventive intervention of genetic RA risk carrier.

Collection and Extraction of Data

We collected 153 English-language literatures from the NHGRI catalog and HuGE Navigator and the automatic paper recommendation system PubMedScan (<http://medals.jp/pubmedscan/>) (Figure 1). Then, manual screening of their abstracts excluded articles about other autoimmune diseases, pharmacogenomics and gene environment studies. After filtering, 90 literatures about GWASs, fine mapping studies and meta-analyses were kept for further reading. Not only statistically significant but also statistically not significant association results, ethnicity of subjects, the country where subjects were recruited, the total number of cases and controls, the analysis platform and the study design were extracted manually from full text, tables or supplementary data. The association results of human genetic variants included SNPs, HLA alleles, copy number variations and variable number of tandem repeats markers. Finally, 7730 association results were stored in a database.

The results of statistical tests in original literatures were re-evaluated according to the following criteria. We determined two kinds of significance levels for GWAS that assayed genome-wide SNPs and for follow-up study, such

as meta-analysis, that assayed only few SNPs of interest. For a GWAS, if the corresponding P -value was $<5.0 \times 10^{-8}$, the result of the statistical analysis was judged as significant evidence of strong association, and P -value between 1.0×10^{-5} and 5.0×10^{-8} was judged as significant evidence of moderate association. Associations with $P > 1.0 \times 10^{-5}$ were judged as not significant. For follow-up study, associations were judged as 'strong' for $P < 0.01$, 'moderate' for P -value between 0.01 and 0.05 and 'not significant' for $P > 0.05$. Exceptionally, a meta-analysis of GWAS result and combined analysis of GWAS and replication results were judged by genome-wide significance threshold. According to our significance level, 5970 associations were classified as statistically not significant, and 1760 associations were classified as either strong or moderate associations.

Confirmation of Reproducibility of RA Risk Variants

To confirm reproducibility of genetic associations based on a geographical information, all associations including non-significant results were grouped by the following nine populations; European (including European American, European Australian, European New Zealander, West European, North European, South European and East European), East Asian (including South-East Asian), West Asian, South Asian, South American, Central American, North African, South African and African American. Because many follow-up studies of previous GWAS researches were conducted, the representative association for the variant in each geographical population was selected from the result of a study of the largest number of case subjects. Out of 6740 representative associations, 1296 were statistically significant and remained for further reproducibility assessment. Notably, 34 variants reached significant level in small-scale studies, but were not confirmed by larger studies.

The genetic associations were classified into two classes based on reproducibility between or within population. To assess reproducibility of an association between different populations, the variants that was reproduced by independent representative associations were identified. A genetic variant that showed opposite direction of association between different populations was excluded. Accordingly, 40 representative associations of 19 variants located in 12 loci were confirmed as 'inter-population reproduced' based on independent studies of different populations such as African American, Central American, East Asian, European, South American, South Asian and West Asian (Table 1).

To assess reproducibility of an association within the particular population, we identified the statistically significant representative association that was reproduced by an

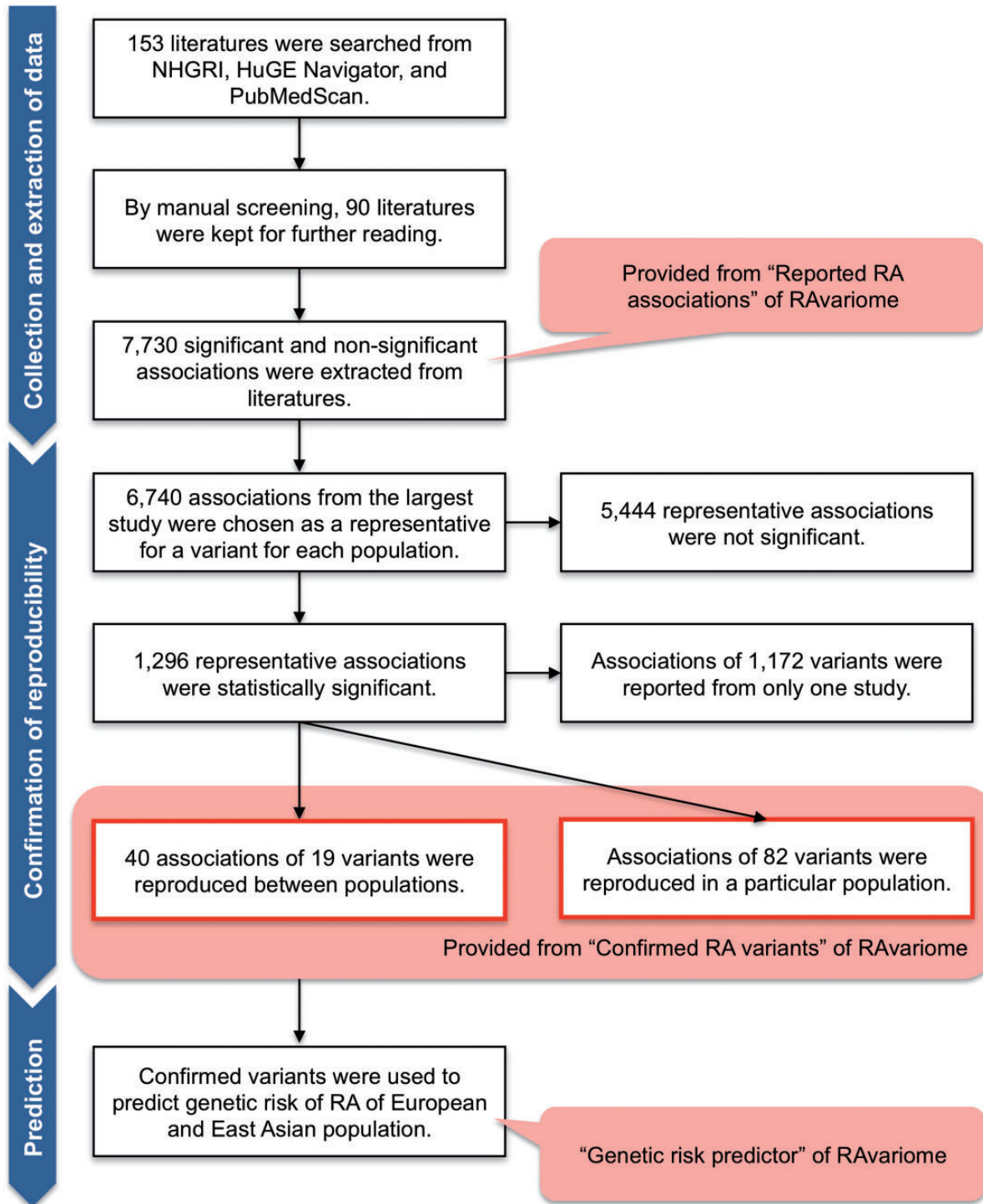


Figure 1. The flowchart for the assessment of association results from articles. From comprehensive search by NHGRI Genome-Wide Association catalog, HuGE Navigator and PubMedScan, 153 literatures were collected. In all, 7730 associations were collected from 90 literatures by manual screening and reading. The associations were grouped into the nine geographical populations based on sample information. A total of 6740 associations (6614 unique genetic variants) were chosen as representative associations of the variants in either population. In all, 5444 representative associations (5405 unique variants) did not reach significance level, and 1296 representative associations (1274 unique variants) were statistically significant and remained for further analysis. Forty representative associations of 19 variants were reproduced between different populations, and associations of 82 variants were reproduced in the particular population. The confirmed variants were used to predict genetic risk of RA based on individual's genotype. The data can be view and downloaded from the web database RAvariome.

Table 1. List of RA risk variants confirmed between different populations

Related genes	SNP-allele	P-value	Genotyping coverage	OR	Case/control numbers	Population	References
APOE	rs805297-A	6.65e-005	Selected	1.40	578/711	East Asian	Hu HJ, <i>Exp Mol Med</i> , 2011 (11)
	rs805297-A	1.49e-016	GW		1147/1079	European	Padyukov L, <i>Ann Rheum Dis</i> , 2011 (12)
CCR6	rs3093023-A	2.1e-017	GW	1.27	4074/16891	East Asian	Okada Y, <i>Nat Genet</i> , 2012 (13)
	rs3093023-A	1.5e-011	GW	1.12	12 307/28975	European	Stahl EA, <i>Nat Genet</i> , 2010 (14)
CTLA4, ICOS	rs231775-G	0.01	Selected	1.45	199/199	Central American	Li X, <i>J Clin Immunol</i> , 2012 (15)
	rs231775-G	0.002	Selected (Meta)	1.18	2734/1441	East Asian	Li X, <i>J Clin Immunol</i> , 2012 (15)
HLA-DRB1	rs231775-G	0.009	Selected (Meta)	1.07	6160/6684	European	Plant D, <i>Ann Rheum Dis</i> , 2010 (16)
	rs13192471-G	1.9e-058	GW	1.97	2303/3380	East Asian	Kochi Y, <i>Nat Genet</i> , 2010 (17)
IKZF3 ^a	rs13192471-G	6.7e-016	Selected	2.16	983/1007	South Asian	Prasad P, <i>PLoS One</i> , 2012 (18)
	rs2157337-C	2.6e-118	GW	1.99	4074/16891	East Asian	Okada Y, <i>Nat Genet</i> , 2012 (13)
	rs2157337-C	1e-299	GW	2.50	5539/20169	European	Stahl EA, <i>Nat Genet</i> , 2010 (14)
	rs2516049-C	3.6e-031	GW	2.18	874/855	East Asian	Terao C, <i>PLoS One</i> , 2011 (19)
	rs2516049-C	1.64e-042	GW		1147/1079	European	Padyukov L, <i>Ann Rheum Dis</i> , 2011 (12)
	rs6457617-T	3.44e-076	GW		1861/2938	European	WTCCC, <i>Nature</i> , 2007 (20)
	rs6457617-A	1.6e-009	Selected	1.48	983/1007	South Asian	Prasad P, <i>PLoS One</i> , 2012 (18)
	rs660895-G	5.83e-053	GW		1147/1079	European	Padyukov L, <i>Ann Rheum Dis</i> , 2011 (12)
	rs660895-G	2.56e-005	Selected	1.52	983/1007	South Asian	Prasad P, <i>PLoS One</i> , 2012 (18)
	rs6910071-G	1e-299	GW	2.88	5539/20169	European	Stahl EA, <i>Nat Genet</i> , 2010 (14)
	rs6910071	0.04	Selected	1.27	983/1007	South Asian	Prasad P, <i>PLoS One</i> , 2012 (18)
	IL2, IL21	rs2872507-G	0.006	Selected	0.79	440/795	African American
rs2872507-G		0.003	Selected	0.91	2414/14245	East Asian	Kurreaman FA, <i>Am J Hum Genet</i> , 2012 (21)
IL6	rs2872507-A	9.4e-007	GW	1.09	12 307/28975	European	Stahl EA, <i>Nat Genet</i> , 2010 (14)
	rs13119723-G	6.8e-007	GW	0.88	12 307/28975	European	Stahl EA, <i>Nat Genet</i> , 2010 (14)
IL2RA	rs13119723-G	0.008	Selected	0.75	983/1007	South Asian	Prasad P, <i>PLoS One</i> , 2012 (18)
	rs13119723-G	5.89e-008	Selected (Meta)	0.88	12953/13370	European	Plant D, <i>Ann Rheum Dis</i> , 2010 (16)
IL6	rs6822844	0.019	Selected	0.61	354/368	South American	Maiti AK, <i>Arthritis Rheum</i> , 2010 (22)
	rs2104286	2.48e-05	Selected (Meta)	0.91	10 112/10450	European	Plant D, <i>Ann Rheum Dis</i> , 2010 (16)
IL6	rs2104286-G	0.00019	Selected	0.73	983/1007	South Asian	Prasad P, <i>PLoS One</i> , 2012 (18)
	rs1800795-G	1e-008	Selected	0.074	120/168	East Asian	Lee YH, <i>Inflamm Res</i> , 2012 (23)
OLIG3, TNFAIP3	rs1800795-G	7.4e-005	Selected	0.46	425/247	West Asian	Lee YH, <i>Inflamm Res</i> , 2012 (23)
	rs2230926	0.032	Selected	1.847	148/1513	African American	Lee YH, <i>Inflamm Res</i> , 2012 (23)
SPRED2	rs2230926-C	1.6e-006	GW	1.31	7069/15876	East Asian	Kochi Y, <i>Nat Genet</i> , 2010 (17)
	rs5029937-T	3.9e-009	GW	1.33	4074/16891	East Asian	Okada Y, <i>Nat Genet</i> , 2012 (13)
STAT4	rs5029937	4.62e-010	Selected (Meta)	1.42	7731/9403	European	Plant D, <i>Ann Rheum Dis</i> , 2010 (16)
	rs934734-G	3.2e-007	GW	1.19	4074/16891	East Asian	Okada Y, <i>Nat Genet</i> , 2012 (13)
ZEB1 ^a	rs934734-G	5.3e-010	GW	1.13	12 307/28975	European	Stahl EA, <i>Nat Genet</i> , 2010 (14)
	rs7574865-T	1.8e-006	GW	1.17	7069/15876	East Asian	Kochi Y, <i>Nat Genet</i> , 2010 (17)
ZEB1 ^a	rs7574865	2.99e-015	Selected (Meta)	1.18	14 394/16 131	European	Plant D, <i>Ann Rheum Dis</i> , 2010 (16)
	rs2793108-T	0.001	Selected	1.11	2414/14245	East Asian	Kurreaman FA, <i>Am J Hum Genet</i> , 2012 (21)
ZEB1 ^a	rs2793108-C	0.006	Selected	0.84	983/1007	South Asian	Prasad P, <i>PLoS One</i> , 2012 (18)

^aConfirmed RA risk loci that were not mentioned in RA reviews (10, 24–27). Related genes, refSNP ID and allele, association P-value, study type to set significance level, OR, sample sizes of case/control, geographical population of samples and references are shown. Genotyping coverage indicated whether the study assayed genome-wide (GW) SNPs or selected SNPs.

independent study of the same population. As a result, associations of 82 variants were confirmed as 'intra-population reproduced'. Intra-population reproduced variants were only reported in European and East Asian populations (53 SNPs and 29 SNPs, respectively).

Prediction of Genetic RA Risk

To facilitate personalized medicine and preventive health care, a tool to calculate genetic risk of RA for an individual was developed based on confirmed RA risk variants (intra- and inter-population-reproduced variants). To avoid over-estimation, if confirmed RA risk variants are closely located to each other, i.e. in the same linkage disequilibrium block, only the genetic variant with a smaller *P*-value was used as a risk marker of that locus.

The genetic risk score (GRS) and the relative genetic risk (RGR) of RA for an individual were calculated by the combination of RA risk markers. As described in previous studies, an unweighted GRS, simply counting the number of risk alleles carried by an individual, was not applicable to RA because HLA-DRB1 alleles had substantially higher odds ratio (OR) than alleles at non-HLA loci (28–30). Accordingly, we used a weighted GRS that increases additively by log-OR of the risk allele (31). GRS was calculated as follows:

$$GRS = \sum_{i=1}^n X_i \log OR_i,$$

where *n* is the number of markers available and X_i represents the copy number of risk allele at a marker *i*. If an individual carries 1 risk allele at *i*th marker, X_i is 1, and if an individual carries 2 risk alleles, then X_i is 2.

To calculate RGR of the average population for each marker, an individual genotype-specific risk s_i is estimated as:

$$s_i = \begin{cases} 1 & \text{if } X_i = 0 \\ r_i & \text{if } X_i = 1 \\ r_i^2 & \text{if } X_i = 2 \end{cases},$$

where r_i is a risk factor assumed to equal the allelic OR of *i*th marker. RGR of individual marker is estimated as:

$$RGR_i = \frac{s_i}{(1 - p_i)^2 + 2p_i(1 - p_i)r_i + p_i^2 r_i^2}.$$

Here p_i is the risk allele frequency of *i*th marker in the control group. Thus, the overall RGR of an individual estimated by multi-markers is calculated as:

$$RGR_{overall} = \prod_{i=1}^n RGR_i.$$

RAvariome Web Interface and Usage

To provide our results and a genetic risk prediction tool for researchers and clinicians, a web database RAvariome was developed. By comparing number of literatures and association data with other existing literature-based GWAS databases, RAvariome provides the most comprehensive dataset of associations of RA (Table 2). RAvariome consists of four sections: overview of RA, the list of confirmed RA variants, the collection of reported RA associations and genetic risk prediction tool. In the overview section, the heritability, environmental factors and protective factors of RA are described.

The page entitled 'Confirmed RA variants' provides users a list of genetic variants with the reproducibility class (Figure 2A). The list provides detailed study information of the representative association of the specific variant and the particular geographical population. In this section, users can search the confirmed RA risk variants by the reproducibility class, the geographical population of the subjects, a gene name or refSNP ID.

The page entitled 'Reported RA associations' leads users to a list of all statistically significant and statistically not significant associations described in literatures. Users can search association data by study design, geographical population of the subjects, nationality of the subjects,

Table 2. Comparing the number of literatures and associations of RAvariome with other literature-based databases

Database	Condition of literature collection	Number of literatures	Condition of data collection	Number of associations
RAvariome	RA GWAS and follow-up studies	90	All associations described in literature	7730
NHGRI GWAS catalog (5)	GWAS literature	14 ^a	Associations reached P -values $< 1.0 \times 10^{-5}$	96 ^a
HuGE Navigator (8)	GWAS literature meta-analysis	40 ^a 98 ^a		84 ^a
GWASdb (6)	GWAS literature and dataset from GAD		Associations reached P -values $< 1.0 \times 10^{-3}$	1748 ^b

^aThe number for HuGE Navigator and NHGRI GWAS catalog were obtained by searching the term 'rheumatoid arthritis' at 12 March 2013. ^bFor GWASdb, number of data was the result as of July 2012.

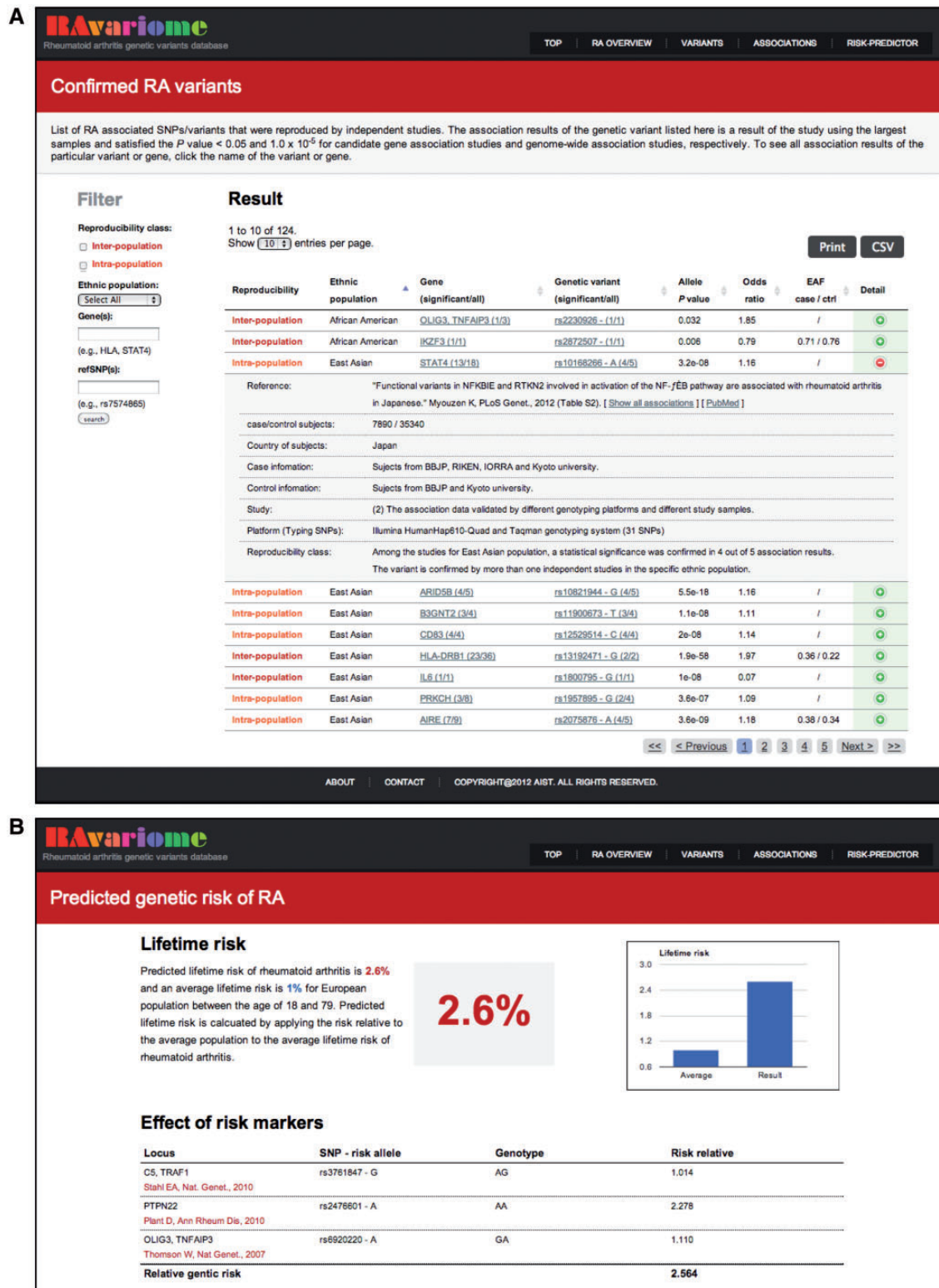


Figure 2. Screenshots of web database RAvariome. (A) View of the confirmed RA variants of RAvariome. In the left sidebar, users can search an association data by reproducibility class of genetic variants, population of subjects, genes and refSNP ID. The query results will be shown in a tabular format. Users can get detailed study information by click on the detail icon on rightmost column. The results can be easily downloaded as print format and comma-separated values (CSV) format. (B) The result of the genetic RA risk predictor. Lifetime risk and effects of each risk markers based on selected genotypes are shown.

Table 3. Result of gene set enrichment analysis using population unique RA risk genes that was confirmed only in European and East Asian populations, respectively

Description of MSigDB gene set (number of genes in gene set)	European unique 39 genes			East Asian unique 30 genes		
	Genes in overlap	P-value	FDR q-value ^a	Genes in overlap	P-value	FDR q-value ^a
Genes involved in Immune System (933)	CD40, HLA-DPB1, IRF5, CD28, IL2, IL2RA, CD226, IL2RB, TNFRSF14, IL6ST, C5, PRKCQ, REL, KIF5A	3.89E-15	2.34E-11	NFKBIE, IL6, IRF8, IRF4, CSF2, PTPN2	8.13E-06	2.26E-03
Genes upregulated by CD40 (GeneID = 958) signaling in Ramos cells (EBV negative Burkitt lymphoma) (101).	CD40, HLA-DPB1, IRF5, CD58, BATF, TRAF1, SH2B3	1.45E-12	4.35E-09	CD83, NFKBIE, ARID5B, IRF4	2.54E-07	2.18E-04
Genes upregulated in TK6, WTK1 and NH32 cell lines (lymphoblast) in response to ionizing radiation (149).	IL2RB, BATF, TRAF1, CCL21, CDK6, PTPN22	1.74E-09	8.73E-07	CD83, NFKBIE, IRF4	7.01E-05	1.11E-02
Genes involved in cytokine signaling in immune system (270).	HLA-DPB1, IRF5, IL2, IL2RA, IL2RB, IL6ST	6.04E-08	1.91E-05	IL6, IRF8, IRF4, CSF2, PTPN2	3.05E-07	2.29E-04
Genes whose promoters are bound by FOXP3 (GeneID = 50943) based a ChIP-chip analysis (491).	CD28, IL2RA, CD226, IL6ST, PRKCQ, PTPN22, ANKRD55	8.89E-08	2.67E-05	ETS1, ARID5B, B3GNT2, IRF4, PRKCH, FLI1, EXOC2	5.66E-09	1.13E-05
Genes upregulated in the normal-like subtype of breast cancer (476).	CD40, PRKCQ, HLA-DRA, CD2, CCL21, PTPN22	1.66E-06	2.68E-04	IL6, IRF8, FLI1, PDE2A	1.15E-04	1.47E-02
Genes upregulated in prostate cancer samples from African-American patients compared with those from the European-American patients.	HLA-DPB1, CD28, HLA-DRA, CD2, PTPN22	3.38E-06	4.83E-04	CD83, IRF8, FLI1	5.43E-04	3.82E-02
Genes from common genomic gains observed in a meta-analysis of copy number alterations across a panel of different cancer cell lines and tumor samples (323).	IL2, REL, CDK6, DDX6, AFF3	4.91E-06	6.29E-04	IRF4, FLI1, MNI	6.79E-04	4.39E-02
Calcineurin-regulated NFAT-dependent transcription in lymphocytes (47).	IL2, IL2RA, PRKCQ	6.72E-06	8.24E-04	IRF4, CSF2	2.95E-04	2.70E-02
Genes downregulated in freshly isolated CD31- [GeneID = 5175] (stromal stem cells from adipose tissue) versus the CD31+ (non-stem) counterparts (216).	HLA-DPB1, SH2B3, HLA-DRA, PODXL	2.36E-05	2.49E-03	PRKCH, FLI1, ANXA3, PDE2A	5.27E-06	1.90E-03
Genes modulated in HeLa cells (cervical carcinoma) by TNF [GeneID = 7124] via NFKB pathway (28).	REL, TRAF1	2.17E-04	1.55E-02	CD83, NFKBIE, GCH1, IL6	1.31E-09	7.87E-06
Cytokines and inflammatory response (29).	IL2, HLA-DRA	2.33E-04	1.57E-02	IL6, CSF2	1.11E-04	1.47E-02

^aThe gene set enrichment analysis was performed by computing overlaps between the population unique genes and gene sets of positional (c1), curated (c2) and motif (c3) of MSigDB (32). The MSigDB gene sets listed in left column were significantly enriched in both populations at 5% false discovery rate (FDR) q-value.

number of subjects and sub-phenotype of cases such as cases with 100% seropositive for rheumatoid factor and/or anti-CCP.

The page entitled 'Genetic risk predictor' provides a tool for predicting the genetic risk of RA calculated by the ethnicity and genotypes of an individual (Figure 2B). By choosing a population from either European or East Asian, the confirmed RA risk markers and their common three genotypes are displayed in the list. Users can select one genotype for each risk marker, and selected genotypes will be used for estimation of RGR and GRS. RGR is the index that assumed the risk of average population as baseline. GRS is the index that assumed non-risk allele carrier as baseline and used in basic researches. For example, if genotypes AG for rs3761847, AA for rs2476601 and GA for rs6920220 are selected at the European population, a predicted lifetime risk will be 2.6%, 2.6-fold higher than the prevalence of RA in general European population (1%).

Discussion

RAvariome was developed to provide a list of RA-associated genetic variants with the degree of reproducibility, based on comprehensive re-evaluated results of the genetic association studies. By comparing our results with reviews (10, 24–27) of RA, we found that two loci from inter-population-reproduced variants were not reported in reviews and have been validated in recent meta-analyses (Table 1). Thus, we conclude that our database provides the newest and most reliable genetic risk variants based on the comparative verification.

As the intra-population-reproduced variants were observed only in European and East Asian populations, we used gene set enrichment analysis (32) to discover common functional mechanisms underlying different gene sets confirmed specifically in European and East Asian populations. From inter- and intra-reproduced-variants, 10 genes were confirmed in both European and East Asian populations, although 39 and 30 genes were confirmed only in the European and East Asian populations, respectively. Interestingly, the European and East Asian unique gene sets had the following gene sets in common: genes involved in immune system, gene upregulated by CD40 signaling in Ramos cells, genes upregulated in TK6, WTK1 and NH32 cell lines in response to ionizing radiation, genes modulated in HeLa cells by TNF via NFKB pathway, genes whose promoters are bounded by FOXP3 based on a chromatin immunoprecipitation (ChIP)-chip analysis, genes involved in cytokine signaling in immune system and genes downregulated in freshly isolated CD31– versus the CD31+ counterparts (Table 3). This result suggested that even if the same genes were not reproduced between European and East Asian populations, several pathways were common in both populations.

The purpose of open access database RAvariome is to be a standard resource not only for RA researchers but also for RA clinicians and the general public. RAvariome is therefore designed as simple as possible to get confirmed RA genetic risk variants for each geographical population. With ongoing progress in sequencing technology, the number of genetic studies of RA will continue to grow. RAvariome will be periodically updated in concert with progress in RA genetic research and will incorporate a new genetic risk prediction method (33). An open access resource would be valuable to raise the precision of the clinical genetic tests and to develop effective prevention programs of RA based on genetic and population differences among individuals.

Acknowledgements

We thank C. Yamasaki, M. Sera, K. Murakami, Y. Hara and T. Habara for their valuable suggestions and technical support for the database development. We also thank Dr Naoyuki Kamatani and Dr Katsushi Tokunaga for their thoughtful guidance.

Funding

AIST Innovation School to Y.N.; Grant-in-Aid for Scientific Research on Innovative Areas (22133002). Funding for open access charge: Grant-in-Aid for Scientific Research on Innovative Areas (22133002).

Conflict of interest. None declared.

References

1. Tobón,G.J., Youinou,P. and Saraux,A. (2010) The environment, geoeidemiology, and autoimmune disease: rheumatoid arthritis. *Autoimmun. Rev.*, **9**, A288–A292.
2. Hakoda,M. (2004) Research on epidemiology of rheumatoid arthritis and clinical dynamics (in Japanese). *Meneki Arerugi Shikkan Yobo, Chiryō Kenkyū Jigyo Kenkyū Hokokusho Heisei 15 Nendo*, **3**, 45–47.
3. Zeng,Q.Y., Chen,R., Darmawan,J. et al. (2008) Rheumatic diseases in China. *Arthritis Res. Ther.*, **10**, R17.
4. Bogdanos,D.P., Smyk,D.S., Rigopoulou,E.I. et al. (2012) Twin studies in autoimmune disease: genetics, gender and environment. *J. Autoimmun.*, **38**, J156–J169.
5. Hindorff,L., Sethupathy,P., Junkins,H. et al. (2009) Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc. Natl Acad. Sci. USA*, **106**, 9362–9367.
6. Li,M.J., Wang,P., Liu,X. et al. (2012) GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.*, **40**, D1047–D1054.
7. Thorisson,G., Lancaster,O., Free,R.C. et al. (2009) HGVbaseG2P: a central genetic association database. *Nucleic Acids Res.*, **37**, D797–D802.

8. Yu,W., Gwinn,M., Clyne,M. et al. (2008) A navigator for human genome epidemiology. *Nat. Genet.*, **40**, 124–125.
9. McDonagh,E.M., Whirl-Carrillo,M., Garten,Y. et al. (2011) From pharmacogenomic knowledge acquisition to clinical applications: the PharmGKB as a clinical pharmacogenomic biomarker resource. *Biomark. Med.*, **5**, 795–806.
10. Eyre,S. and Thomson,W. (2012) Case study on rheumatoid arthritis. In: Zeggini,E. and Morris,A. (eds), *Analysis of Complex Disease Association Study—A Practical Guide*. Academic Press, London, pp. 307–323.
11. Hu,H.J., Jin,E.H., Yim,S.H. et al. (2011) Common variants at the promoter region of the APOM confer a risk of rheumatoid arthritis. *Exp. Mol. Med.*, **43**, 613–621.
12. Padyukov,L., Seielstad,M., Ong,R.T.H. et al. (2011) A genome-wide association study suggests contrasting associations in ACPA-positive versus ACPA-negative rheumatoid arthritis. *Ann. Rheum. Dis.*, **70**, 259–265.
13. Okada,Y., Terao,C., Ikari,K. et al. (2012) Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. *Nat. Genet.*, **44**, 511–516.
14. Stahl,E.A., Raychaudhuri,S., Remmers,E.F. et al. (2010) Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci. *Nat. Genet.*, **42**, 508–514.
15. Li,X., Zhang,C., Zhang,J. et al. (2012) Polymorphisms in the CTLA-4 Gene and Rheumatoid Arthritis Susceptibility: A Meta-analysis. *J. Clin. Immunol.*, 10.1007/s10875-012-9650-y.
16. Plant,D., Flynn,E., Mbarek,H. et al. (2010) Investigation of potential non-HLA rheumatoid arthritis susceptibility loci in a European cohort increases the evidence for nine markers. *Ann. Rheum. Dis.*, **69**, 1548–1553.
17. Kochi,Y., Okada,Y., Suzuki,A. et al. (2010) A regulatory variant in CCR6 is associated with rheumatoid arthritis susceptibility. *Nat. Genet.*, **42**, 515–519.
18. Prasad,P., Kumar,A., Gupta,R. et al. (2012) Caucasian and Asian specific rheumatoid arthritis risk loci reveal limited replication and apparent allelic heterogeneity in north Indians. *PLoS One*, **7**, e31584.
19. Terao,C., Ohmura,K., Katayama,M. et al. (2011) Myelin basic protein as a novel genetic risk factor in rheumatoid arthritis—a genome-wide study combined with immunological analyses. *PLoS One*, **6**, 10.
20. 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.
21. Kurreeman,F.A.S., Stahl,E.A., Okada,Y. et al. (2012) Use of a multi-ethnic approach to identify rheumatoid- arthritis-susceptibility loci, 1p36 and 17q12. *Am. J. Hum. Genet.*, **90**, 1–9.
22. Maiti,A.K., Kim-Howard,X., Viswanathan,P. et al. (2010) Confirmation of an association between rs6822844 at the IL2-IL21 region and multiple autoimmune diseases: evidence of a general susceptibility locus. *Arthritis Rheum.*, **62**, 323–329.
23. Lee,Y.H., Bae,S.C., Choi,S.J. et al. (2012) The association between interleukin-6 polymorphisms and rheumatoid arthritis: a meta-analysis. *Inflamm. Res.*, **61**, 665–671.
24. Suzuki,A., Kochi,Y., Okada,Y. et al. (2011) Insight from genome-wide association studies in rheumatoid arthritis and multiple sclerosis. *FEBS Lett.*, **585**, 3627–3632.
25. Scott,I.C., Steer,S., Lewis,C.M. et al. (2011) Precipitating and perpetuating factors of rheumatoid arthritis immunopathology: linking the triad of genetic predisposition, environmental risk factors and autoimmunity to disease pathogenesis. *Best Pract. Res. Clin. Rheumatol.*, **25**, 447–468.
26. Raychaudhuri,S. (2010) Recent advances in the genetics of rheumatoid arthritis. *Curr. Opin. Rheumatol.*, **22**, 109–118.
27. Chatzikyriakidou,A., Voulgari,P.V., Lambropoulos,A. et al. (2013) Genetics in rheumatoid arthritis beyond HLA genes: what meta-analyses have shown? *Semin. Arthritis Rheum.*, **43**, 29–38.
28. Karlson,E.W., Chibnik,L.B., Kraft,P. et al. (2010) Cumulative association of twenty-two genetic variants with seropositive rheumatoid arthritis risk. *Ann. Rheum. Dis.*, **69**, 1077–1085.
29. Nakaoka,H., Cui,T., Tajima,A. et al. (2011) A systems genetics approach provides a bridge from discovered genetic variants to biological pathways in rheumatoid arthritis. *PLoS One*, **6**, e25389.
30. Chibnik,L.B., Keenan,B.T., Cui,J. et al. (2011) Genetic risk score predicting risk of rheumatoid arthritis phenotypes and age of symptom onset. *PLoS One*, **6**, e24380.
31. Sebastiani,P., Solovieff,N. and Sun,J.X. (2012) Naive bayesian classifier and genetic risk score for genetic risk prediction of a categorical trait: not so different after all! *Front. Genet.*, **3**, 26.
32. Subramanian,A., Tamayo,P., Mootha,V.K. et al. (2005) Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA*, **102**, 15545–15550.
33. Stahl,E., Wegmann,D., Trynka,G. et al. (2012) Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. *Nat. Genet.*, **44**, 483–489.