Comparative study for haplotype block partitioning methods – Evidence from chromosome 6 of the North American Rheumatoid Arthritis Consortium (NARAC) dataset

Mohamed N. Saad [1]⊙*, **Mai S. Mabrouk** [2]⊙, **Ayman M. Eldeib** [3]‡, **Olfat G. Shaker** [4]‡

1 Biomedical Engineering Department, Faculty of Engineering, Minia University, Minia, Egypt, **2** Biomedical Engineering Department, Faculty of Engineering, Misr University for Science and Technology (MUST), 6th of October City, Egypt, **3** Systems and Biomedical Engineering Department, Faculty of Engineering, Cairo University, Giza, Egypt, **4** Medical Biochemistry and Molecular Biology Department, Faculty of Medicine, Cairo University, Cairo, Egypt

⊙ These authors contributed equally to this work.
‡ These authors also contributed equally to this work.
* m.n.saad@ieee.org, m.n.saad@minia.edu.eg

Abstract

Haplotype-based methods compete with "one-SNP-at-a-time" approaches on being preferred for association studies. Chromosome 6 contains most of the known genetic biomarkers for rheumatoid arthritis (RA) disease. Therefore, chromosome 6 serves as a benchmark for the haplotype methods testing. The aim of this study is to test the North American Rheumatoid Arthritis Consortium (NARAC) dataset to find out if haplotype block methods or single-locus approaches alone can sufficiently provide the significant single nucleotide polymorphisms (SNPs) associated with RA. In addition, could we be satisfied with only one method of the haplotype block methods for partitioning chromosome 6 of the NARAC dataset? In the NARAC dataset, chromosome 6 comprises 35,574 SNPs for 2,062 individuals (868 cases, 1,194 controls). Individual SNP approach and three haplotype block methods were applied to the NARAC dataset to identify the RA biomarkers. We employed three haplotype partitioning methods which are confidence interval test (CIT), four gamete test (FGT), and solid spine of linkage disequilibrium (SSLD). P-values after stringent Bonferroni correction for multiple testing were measured to assess the strength of association between the genetic variants and RA susceptibility. Moreover, the block size (in base pairs (bp) and number of SNPs included), number of blocks, percentage of uncovered SNPs by the block method, percentage of significant blocks from the total number of blocks, number of significant haplotypes and SNPs were used to compare among the three haplotype block methods. Individual SNP, CIT, FGT, and SSLD methods detected 432, 1,086, 1,099, and 1,322 associated SNPs, respectively. Each method identified significant SNPs that were not detected by any other method (Individual SNP: 12, FGT: 37, CIT: 55, and SSLD: 189 SNPs). 916 SNPs were discovered by all the three haplotype block methods. 367 SNPs were discovered by the haplotype block methods and the individual SNP approach. The P-

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: BP, Base Pair; CIT, Confidence Interval Test; D´, Normalized Deviation; FGT, Four Gamete Test; HLA, Human Leukocyte Antigen; HWE, Hardy-Weinberg Equilibrium; IGV, Integrative Genomics Viewer; LD, Linkage Disequilibrium; MAF, Minor Allele Frequency; Mb, Mega Base Pair; MHC, Major Histocompatibility Complex; NARAC, North American Rheumatoid Arthritis Consortium; Q-Q, Quantile-Quantile; RA, Rheumatoid Arthritis; SNP, Single Nucleotide Polymorphism; SSLD, Solid Spine of Linkage Disequilibrium.

values of these 367 SNPs were lower than those of the SNPs uniquely detected by only one method. The 367 SNPs detected by all the methods represent promising candidates for RA susceptibility. They should be further investigated for the European population. A hybrid technique including the four methods should be applied to detect the significant SNPs associated with RA for chromosome 6 of the NARAC dataset. Moreover, SSLD method may be preferred for its favored benefits in case of selecting only one method.

Introduction

Many researchers associate RA disease with genetic biomarkers through individual SNP studies. Recently, the availability of high genomic density of SNPs allows the application of the haplotype block methods. These methods discover a group of SNPs within an associated block in only one test [1, 2].

The main advantages of haplotype block methods over individual SNP approaches are: (a) the reduction of the association testing dimension by using a single test for a block containing more than one SNP; (b) leading to power preservation and ensuring accepted false-positive rates [3]; (c) acquiring the synergy among SNPs. The main disadvantages of haplotype block methods are: (a) the more haplotypes within a block leads to a higher degree of freedom of the block ending in scaled down power; (b) each partitioning method ends up with haplotype blocks that differ from the others. Therefore, a comparison study should be performed to evaluate all partitioning methods' performance [1, 4].

Singular SNP approaches achieve inspiring results if there should be an occurrence of monogenetic disorders (for example, sickle cell anemia). Then again, they don't achieve a similar accomplishment in complex diseases. The power of association was studied for individual SNP approaches and haplotype block methods resulting in contradictory findings. This inconsistency may be arisen from the dependency of the method's performance on the nature of the experimented dataset itself [1, 5].

RA is a chronic autoimmune disease that is prevalent in women more than in men (with a ratio of about 3:1) [6–9]. The MHC (major histocompatibility complex) region extends on the short arm of chromosome 6 (6p21.3) from 26 to 34 Mb (mega base pair) [10]. The MHC region contains the HLA (human leukocyte antigen) region [11] which includes about 50% of the detected biomarkers for RA susceptibility. The association between HLA region and RA disease has been verified in multiethnic populations [12].

There are two objectives of this study. The first objective is comparing the association results of haplotype block methods and individual SNP approach on chromosome 6 of the NARAC dataset. The second objective is selecting the best haplotype block method suitable for the NARAC dataset (if applicable).

Materials and methods

Study population

The NARAC dataset consisted of 2,062 participants (1,493 females and 569 males), 868 RA patients and 1,194 healthy controls. All cases and controls were of European descent. All participants were genotyped on the HumanHap500 v1, Human Hap500 v3, HumanHap300, and HumanHap240 Illumina arrays [13]. The studied genetic variants were 35,574 SNPs included in chromosome 6. After removing 1,452 SNPs, 34,122 SNPs were retained for further analysis.

The reasons for excluding the 1,452 SNPs were the biomarker checks: (a) less than 75% geno-type percent, (b) less than 0.001 Hardy-Weinberg equilibrium (HWE) *P*-value or (c) less than 0.001 minor allele frequency (MAF) in the total sample.

Materials

The chromosome 6 data file was extracted from the NARAC data file using the programming language (Perl). The chromosome 6 data file was reformatted to be ready for processing by the program (PLINK) by the statistical package (R 3.1.0). Moreover, the R language was used to extract the chromosome 6 map file from the NARAC map file (SNP ID, physical position, and chromosome number). The chromosome 6 reformatted data and map files were processed by (PLINK 1.07) and (gPLINK 2.05) programs to be ready for processing by the program (Haplo-view) [14].

The computer program (Haploview 4.2) was used to partition chromosome 6 into successive blocks using CIT, FGT, and SSLD methods and to calculate the corresponding *P*-value for each haplotype in each block. The default parameters for the three methods were used. In addition, the program (Haploview 4.2) was used to apply the individual SNP approach and to provide the corresponding *P*-value for each SNP [15]. The distribution of *P*-values on chromosome 6 for the individual SNP approach was presented using the Integrative Genomics Viewer (IGV 2.3.72) [16, 17]. The significant blocks and the associated SNPs were selected using the pro-gramming language (Matlab Release 2010a). Fig 1 showed a block diagram for the whole chro-mosomic association analysis process.

Haplotype block methods

The haplotype blocks have been defined through more than one method. Several algorithms for haplotype partitioning have been proposed, among which the CIT, FGT, and SSLD have been implemented in HaploView 4.2. Each method differs greatly from the others in its scope of the definition of the haplotype blocks. Therefore, an objective has been arisen for comparing the results of these methods which was studied through this research paper [18].

Confidence interval test.   A haplotype block is defined by the CIT method as a region over which a very small fraction ($\leq$ 5%) of the measures among the informative SNP pairs are in weak linkage disequilibrium (LD). The informative SNP pairs are pairs showing strong LD or weak LD. Biological and artefactual forces in addition to recombination events are the rea-sons for allowing 5% of weak LD in the haplotype block. These forces could be recurrent muta-tion, gene conversion, or errors of the genome assembly or genotyping.

D´ (normalized deviation) is used to measure the LD between a pair of SNPs. CI is used to assess the reliability of the estimate of D´. Strong LD is defined as the upper limit of the 90% CI is (0.98), and the lower limit is (0.7). In contrast, weak LD is defined as the upper limit of the 90% CI is (0.9), as shown in S1 Fig. The thresholds were obtained by Gabriel et al. Short blocks (2–5 SNPs) were treated with different thresholds for different populations to select the used thresholds [19].

Four gamete test.   The FGT is a haplotype block partitioning method that assumes recom-bination events are not allowed within each block. When the four gametes are identified, a recombination event has been occurred. The rare gamete must be observed at a frequency greater than 0.01 to count a recombination event. The recombination events are only accepted between blocks. The FGT method differs from other haplotype block definitions that it does not require a threshold for D´.

The recombination events interrupt the continuity of the testing process. When a recombi-nation event is observed between the (k^{th}) locus and any preceding locus, the locus (k-1) is
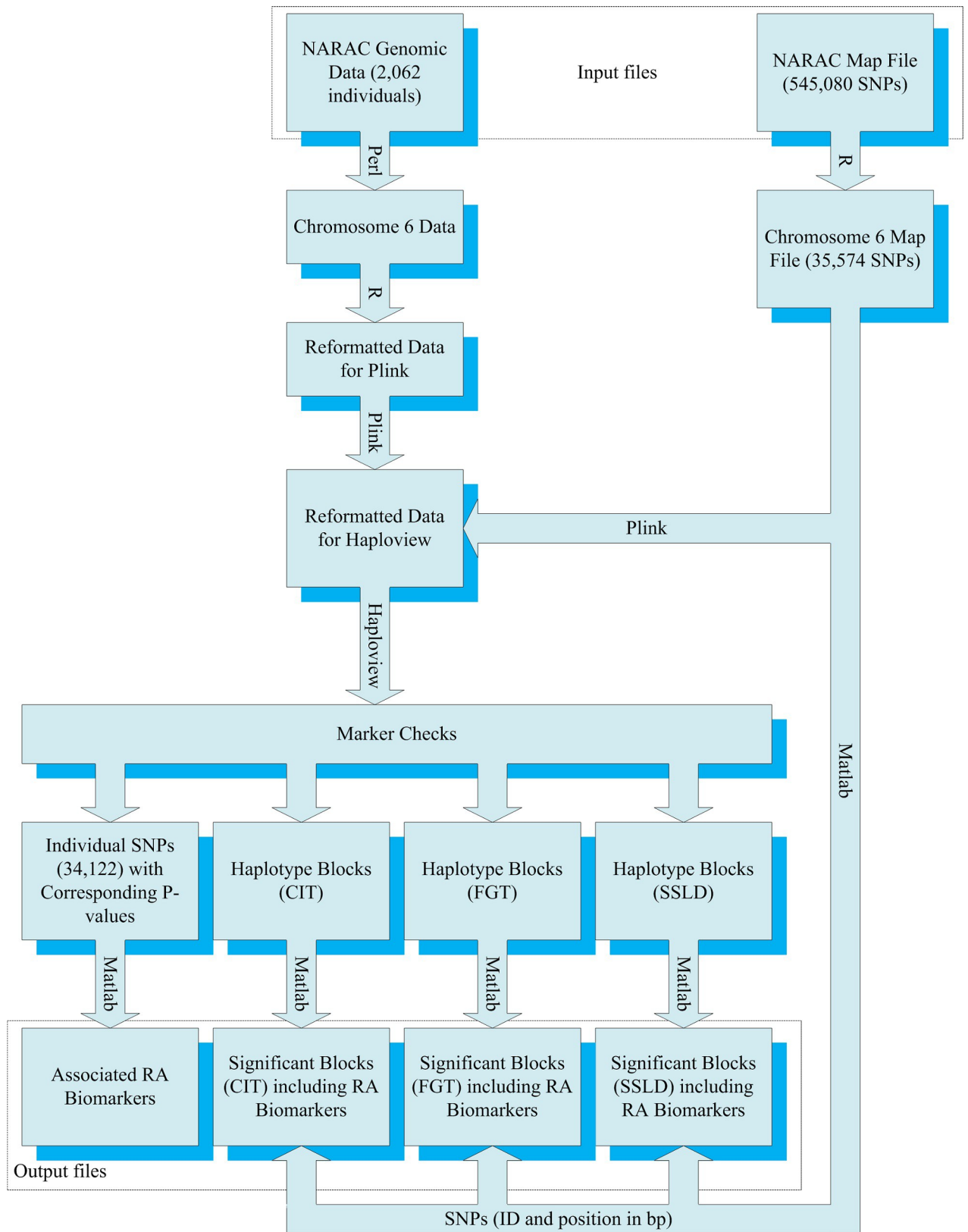
Fig 1. Experimental design of the proposed system.

https://doi.org/10.1371/journal.pone.0209603.g001

considered the ending point of the tested block. The block size could be measured as the distance between the start locus and the end locus. In this situation, the locus (k) is considered the starting point to search for a new block [20].

Solid spine of linkage disequilibrium. The SSLD method defines the haplotype block as a region at which a (spine) of strong LD (D´ > 0.8) moving from one allele to the next allele along the legs of the triangle in the LD chart, as shown in S2 Fig. In other words, for each block, there must be a strong LD between (the first SNP and the last SNP) and all the intermediate SNPs. However, the intermediate SNPs should not be in strong LD with each other [15].

Table 1 represents the concept of the SSLD method. Five SNPs are tested, having the same results as in S2 Fig. Here, (SNP2 and SNP3), (SNP3 and SNP4), and (SNP2 and SNP4) are not in strong LD. All other combinations of SNPs are in strong LD. Thus, there is a haplotype block extends from SNP1 to SNP5. Table 2 shows a comparison among the three haplotype block definitions.

Testing for associations with disease status

Both the individual SNP associations and the haplotype associations were measured with the aid of the *P*-values. The statistically significant SNPs were detected using their corresponding *P*-values after stringent Bonferroni correction for multiple testing ($< 0.05/\#$ of tests) [21]. The Bonferroni thresholds were 8.49×10^{-6} for the CIT method, 7.95×10^{-6} for the FGT method, 9.41×10^{-6} for the SSLD method, and 1.47×10^{-6} for the individual SNP approach where the total number of tests were 5,888 for the CIT method, 6,293 for the FGT method, 5,313 for the SSLD method, and 34,122 for the individual SNP approach.

Results

The testing algorithms were applied on Intel Core i7-4720HQ 2.6 GHz system with 16 GB of RAM. S1 Table provided the processing time for each used program. A total working time was 181 minutes.

The distribution of the observed *P*-values for all the used models showed evidence for population stratification (Fig 2) [22, 23]. The reason for this stratification might be the selection of chromosome 6 that contains the HLA region where many highly significant associations occur. The distribution of the *P*-values across chromosome 6 was presented in Fig 3 for the individual SNP approach. The top SNP (rs660895) in the HLA region (32,685,358 bp) had the lowest *P*-value (1.03×10^{-113}) as previously reported in Arya et al. [24].

The results related to the haplotype block methods are shown in Table 3. The associated SNPs properties are included in S1 Spreadsheet for the CIT method, S2 Spreadsheet for the FGT method, S3 Spreadsheet for the SSLD method, and S4 Spreadsheet for the individual SNP approach.

The SSLD significant blocks included more associated SNPs (1,322) than FGT (1,099) and CIT (1,086) blocks. Moreover, the number of the associated SNPs by the individual SNP

Table 1. Example on the solid spine of linkage disequilibrium (SSLD) method.

SNP#	SNP1	SNP2	SNP3	SNP4	SNP5
SNP1	-	0.97	0.99	0.93	0.96
SNP2		-	0.18	0.67	0.98
SNP3			-	0.03	0.94
SNP4				-	0.95
SNP5					-

https://doi.org/10.1371/journal.pone.0209603.t001

Table 2. A comparison among haplotype block definitions.

Items	FGT	CIT	SSLD
Recombination Event within Block	Not Allowed	≤ 5%	Allowed only between intermediate SNPs
Strong LD	LD is not used	D´ upper limit = 0.98 D´ lower limit = 0.7	D´ > 0.8
Weak LD	LD is not used	D´ upper limit = 0.9	D´ ≤ 0.8
Morphology in the LD Chart	No recombination event between all SNPs in the block	> 95% Strong LD between all SNPs in the block	Strong LD in the legs of the LD chart

https://doi.org/10.1371/journal.pone.0209603.t002

approach was (432). The SSLD method did the best job of representing the hits of the individual SNP approach with (415) SNPs. While, the CIT method represented (381) SNPs, and the FGT method represented (387) SNPs (S5 Spreadsheet). Interestingly, the SSLD method totally represented the top 50 associated SNPs by the individual SNP approach from (*P*-values) point of view. In addition, the CIT and FGT methods represented (47) and (48) SNPs respectively from the top 50 SNPs, as shown in Table 4.

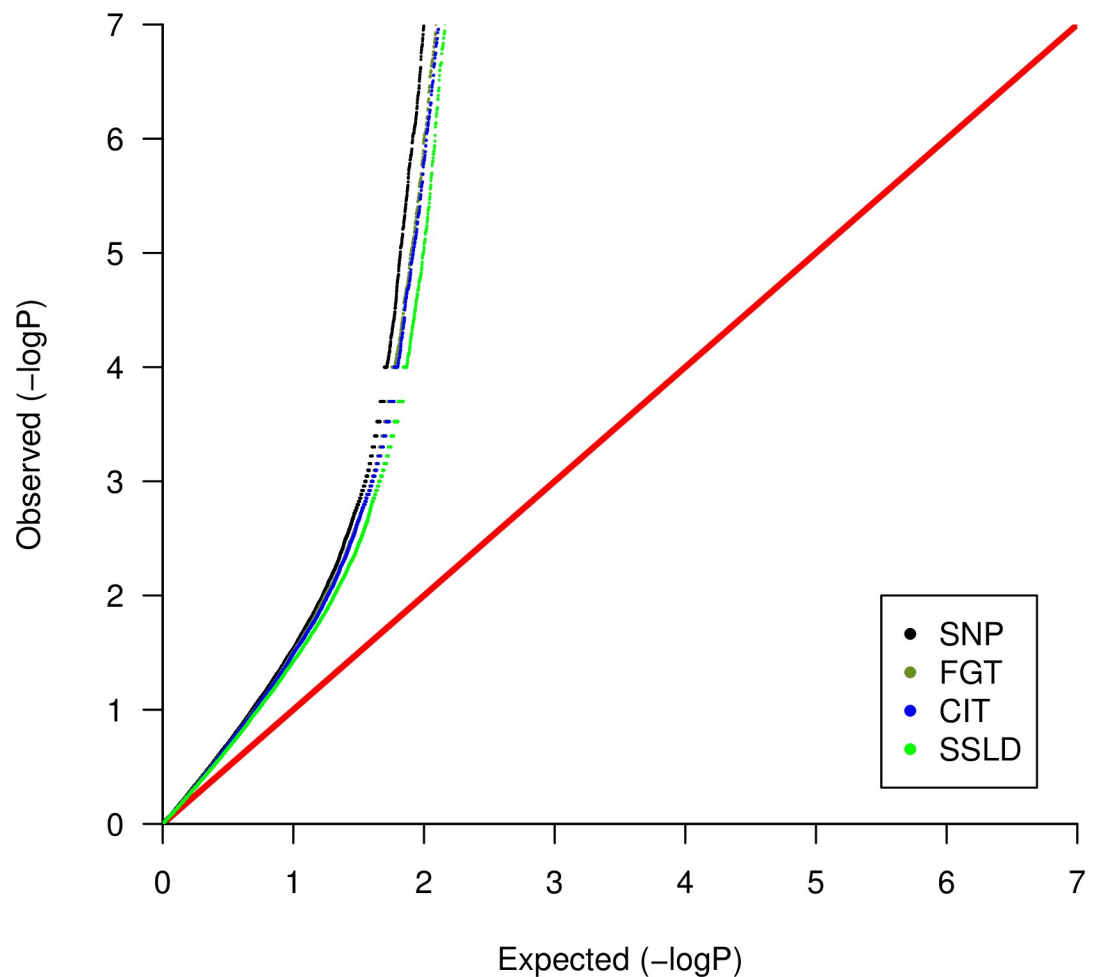**Fig 2. Quantile–quantile (Q–Q) plots of the chromosome 6 association results for all the used methods.** Q-Q plot shows the observed distribution of $-\log_{10}(P$-values) on the Y-axis compared to the expected distribution of $-\log_{10}(P$-values) on the X-axis. The red line (X = Y) represents the null distribution. SNP refers to the individual SNP approach.

https://doi.org/10.1371/journal.pone.0209603.g002
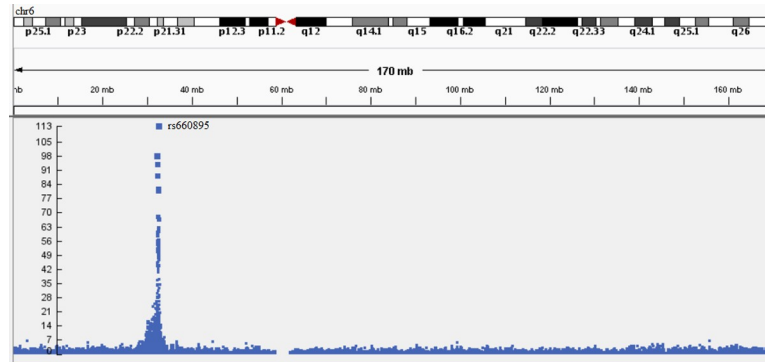
Fig 3. Manhattan plot for the associations between genotyped SNPs (chromosome 6 of the NARAC dataset) and RA susceptibility using individual SNP approach. The upper part represents chromosome 6 ideogram and its genomic coordinates. The Y-axis shows $-\log_{10}(P\text{-values})$, which represent the strength of association. The larger the point and the higher the point on the scale, the more significant the association with RA susceptibility.

https://doi.org/10.1371/journal.pone.0209603.g003

From Table 4, the intergenic region between HLA-DQB1 and HLA-DQA2 –at the p21.32 band of chromosome 6 (MHC class II)–contained blocks with the largest number of RA-associated biomarkers by the individual SNP approach. Within this region, the SSLD method had two blocks with thirteen (P-value = 1.49 X 10^{-53}, no. of SNPs = 20) and twelve (P-value = 1.55 X 10^{-52}, no. of SNPs = 17) biomarkers. The CIT method was represented by a block (P-value = 3.68 X 10^{-54}, no. of SNPs = 31) having twenty-two biomarkers. The FGT block (P-value = 3.26 X 10^{-52}, no. of SNPs = 28) was detected containing nineteen biomarkers.

Discussion and conclusions

In this study, 34,122 SNPs were used to examine the association with RA susceptibility in the NARAC dataset. The examined SNPs belonged to chromosome 6. The surveyed SNPs on chromosome 6 of the NARAC dataset were dense enough for the application of the haplotype block methods. Four methods were applied to assign the associations (three haplotype block

Table 3. A comparison among haplotype block methods results.

Items	CIT	FGT	SSLD
Total number of blocks	5,888	6,293	5,313
Maximum block size (bp)	480,781	351,442	430,369
Maximum significant block size (bp)	209,236	142,115	198,772
Median block size (bp)	8,456	9,582	13,943
Median significant block size (bp)	8,671	7,447	10,122
Minimum block size (bp)	6	6	6
Minimum significant block size (bp)	25	25	76
Median number of SNPs within each block	3	4	5
Median number of SNPs within each significant block	5	5	7
Percentage of uncovered SNPs by the block method	19.46%	10.25%	3.3%
Percentage of significant blocks from the total number of blocks	2.72%	2.84%	2.73%
Total number of significant haplotypes	307	343	302
Total number of significant SNPs	1,086	1,099	1,322
Intersection with SNPs detected by individual SNP approach (432 SNPs)	381	387	415
Intersection with top 50 SNPs (lowest *P*-values) detected by the individual SNP approach	47	48	50

https://doi.org/10.1371/journal.pone.0209603.t003

Table 4. The top 50 associated SNPs discovered by the individual SNP approach with the corresponding haplotype blocks.

SNP ID	Position (bp)	Assoc. Allele[a]	AAF[b] (Case, Control)	P-value	Gene / Nearest Genes	Haplotype Block (Method, P-value, No. of SNPs in Block)	Haplotype Block Position (bp) (Start, End, Size)	Previously Studied in
rs1033500	32415360	A	0.619, 0.393	2.1 E-45	C6orf10	CIT, 3.49 E-44, 15	32415360, 32445664, 30305	[25]
						FGT, 3.12 E-44, 8	32398932, 32425613, 26682	
						SSLD, 1.48 E-97, 26	32390832, 32445664, 54832	
rs1980495	32454772	C	0.511, 0.269	3.45 E-52	C6orf10 / BTNL2	CIT, 4.08 E-59, 9	32454772, 32471794, 17023	[26, 27]
						FGT, 4.08 E-59, 9	32454772, 32471794, 17023	
						SSLD, 1.46 E-81, 3	32449376, 32456123, 6748	
rs2076530	32471794	G	0.699, 0.444	2.97 E-59	BTNL2	CIT, 4.08 E-59, 9	32454772, 32471794, 17023	[28–32]
						FGT, 4.08 E-59, 9	32454772, 32471794, 17023	
						SSLD, 4.62 E-59, 7	32462406, 32471794, 9388	
rs2395157	32456123	G	0.528, 0.276	8.58 E-60	C6orf10 / BTNL2	CIT, 4.08 E-59, 9	32454772, 32471794, 17023	[33]
						FGT, 4.08 E-59, 9	32454772, 32471794, 17023	
						SSLD, 1.46 E-81, 3	32449376, 32456123, 6748	
rs2395163	32495787	G	0.509, 0.203	1.61 E-94	BTNL2 / HLA-DRA	CIT, 5.66 E-95, 11	32495758, 32512856, 17098	[34–39]
						FGT, 4.41 E-95, 5	32491086, 32495787, 4702	
						SSLD, 4.86 E-95, 14	32491086, 32512837, 21752	
rs2395185	32541145	A	0.589, 0.315	1.01 E-68	HLA-DRA / HLA-DRB5	CIT, 6.63 E-131, 11	32541145, 32713862, 172718	[32, 36, 37, 40, 41]
						FGT, 5.2 E-72, 4	32536263, 32678378, 142116	
						SSLD, 3.47 E-77, 9	32519501, 32678378, 158878	
rs2516049	32678378	G	0.576, 0.305	2.61 E-67	HLA-DRB1 / HLA-DQA1	CIT, 6.63 E-131, 11	32541145, 32713862, 172718	[32, 33, 36, 37, 42, 43]
						FGT, 5.2 E-72, 4	32536263, 32678378, 142116	
						SSLD, 3.47 E-77, 9	32519501, 32678378, 158878	
rs2647012	32772436	G	0.837, 0.624	1.76 E-50	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[32, 44, 45]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs2856717	32778286	G	0.838, 0.624	6.97 E-51	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs2856725	32774716	A	0.837, 0.623	1.05 E-50	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45, 46]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs2858305	32778442	A	0.838, 0.624	6.97 E-51	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[33, 45]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs3129871	32514320	C	0.835, 0.631	2.14 E-46	BTNL2 / HLA-DRA	SSLD, 5.38 E-41, 2	32512856, 32514320, 1465	[31, 46, 47]
rs3763309	32483951	A	0.507, 0.209	6.48 E-89	BTNL2 / HLA-DRA	CIT, 1.99 E-89, 8	32483951, 32491201, 7251	[35–38, 48]
						FGT, 1.46 E-87, 7	32483951, 32488240, 4290	
						SSLD, 1.46 E-87, 6	32481290, 32487361, 6072	
rs3763312	32484326	A	0.505, 0.207	1.05 E-88	BTNL2 / HLA-DRA	CIT, 1.99 E-89, 8	32483951, 32491201, 7251	[35, 36, 39, 46]
						FGT, 1.46 E-87, 7	32483951, 32488240, 4290	
						SSLD, 1.46 E-87, 6	32481290, 32487361, 6072	
rs3817963	32476065	G	0.534, 0.294	1.23 E-54	BTNL2	CIT, 1.07 E-54, 3	32474399, 32476065, 1667	[31, 37, 39, 49–52]
						FGT, 2.51 E-42, 4	32476065, 32481676, 5612	
						SSLD, 1.54 E-54, 4	32474399, 32477466, 3068	
rs3817973	32469089	A	0.698, 0.440	5.48 E-61	C6orf10 / BTNL2	CIT, 4.08 E-59, 9	32454772, 32471794, 17023	[31, 32, 39]
						FGT, 4.08 E-59, 9	32454772, 32471794, 17023	

(Continued)

Table 4. (Continued)

SNP ID	Position (bp)	Assoc. Allele[a]	AAF[b] (Case, Control)	P-value	Gene / Nearest Genes	Haplotype Block (Method, P-value, No. of SNPs in Block)	Haplotype Block Position (bp) (Start, End, Size)	Previously Studied in
						SSLD, 4.62 E-59, 7	32462406, 32471794, 9389	
rs3957148	32790115	G	0.277, 0.094	1.88 E-53	HLA-DQB1 / HLA-DQA2	CIT, 6.72 E-54, 7	32788906, 32790115, 1210	[53, 54]
						FGT, 2.32 E-53, 4	32789654, 32790286, 633	
						SSLD, 6.72 E-54, 7	32788906, 32790115, 1210	
rs4424066	32462406	G	0.698, 0.440	1.18 E-60	C6orf10 / BTNL2	CIT, 4.08 E-59, 9	32454772, 32471794, 17023	[32, 55]
						FGT, 4.08 E-59, 9	32454772, 32471794, 17023	
						SSLD, 4.62 E-59, 7	32462406, 32471794, 9389	
rs477515	32677669	A	0.575, 0.304	2.72 E-67	HLA-DRB1 / HLA-DQA1	CIT, 6.63 E-131, 11	32541145, 32713862, 172718	[32, 36, 42, 56]
						FGT, 5.2 E-72, 4	32536263, 32678378, 142116	
						SSLD, 3.47 E-77, 9	32519501, 32678378, 158878	
rs5000634	32771542	G	0.614, 0.382	1.09 E-48	HLA-DQB1 / HLA-DQA2	CIT, 2.62 E-81, 3	32767856, 32771829, 3974	[32, 33, 57]
						FGT, 1.03 E-48, 4	32766602, 32771829, 5228	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs547077	32397296	G	0.622, 0.397	3.98 E-46	C6orf10	CIT, 3.14 E-98, 18	32332366, 32406350, 73985	[39]
						FGT, 1.5 E-98, 15	32332366, 32397296, 64931	
						SSLD, 1.48 E-97, 26	32390832, 32445664, 54833	
rs6457617	32771829	A	0.803, 0.508	2.55 E-82	HLA-DQB1 / HLA-DQA2	CIT, 2.62 E-81, 3	32767856, 32771829, 3974	[32, 33, 36, 43, 48, 58–75]
						FGT, 1.03 E-48, 4	32766602, 32771829, 5228	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs660895	32685358	G	0.529, 0.192	1.03 E-113	HLA-DRB1 / HLA-DQA1	CIT, 6.63 E-131, 11	32541145, 32713862, 172718	[33, 35–43, 46, 47, 51, 60, 65, 68, 71, 76–92]
						FGT, 1.84 E-113, 8	32680229, 32713862, 33634	
						SSLD, 8.37 E-108, 11	32680229, 32760295, 80067	
rs6903608	32536263	A	0.882, 0.688	2.84 E-48	HLA-DRA / HLA-DRB5	FGT, 5.2 E-72, 4	32536263, 32678378, 142116	[32, 33, 40, 55]
						SSLD, 3.47 E-77, 9	32519501, 32678378, 158878	
rs6910071	32390832	G	0.506, 0.195	1.25 E-98	C6orf10	CIT, 3.14 E-98, 18	32332366, 32406350, 73985	[33, 35–40, 43, 46–48, 50, 65, 68, 71, 79, 86–89, 93, 94]
						FGT, 1.5 E-98, 15	32332366, 32397296, 64931	
						SSLD, 1.48 E-97, 26	32390832, 32445664, 54833	
rs6932542	32488240	G	0.769, 0.527	1.06 E-56	BTNL2 / HLA-DRA	CIT, 1.99 E-89, 8	32483951, 32491201, 7251	[33, 95]
						FGT, 1.46 E-87, 7	32483951, 32488240, 4290	
						SSLD, 4.87 E-53, 3	32487714, 32488240, 527	
rs7192	32519624	C	0.826, 0.608	1.68 E-51	HLA-DRA	CIT, 2.05 E-25, 3	32519501, 32521295, 1795	[31, 33, 39, 46, 96, 97]
						FGT, 2.05 E-25, 3	32519501, 32521295, 1795	
						SSLD, 3.47 E-77, 9	32519501, 32678378, 158878	
rs9268528	32491086	G	0.596, 0.365	8.28 E-49	BTNL2 / HLA-DRA	CIT, 1.99 E-89, 8	32483951, 32491201, 7251	[98–100]
						FGT, 4.41 E-95, 5	32491086, 32495787, 4702	
						SSLD, 4.86 E-95, 14	32491086, 32512837, 21752	
rs9268542	32492699	G	0.597, 0.371	1.56 E-46	BTNL2 / HLA-DRA	FGT, 4.41 E-95, 5	32491086, 32495787, 4702	[98, 99]
						SSLD, 4.86 E-95, 14	32491086, 32512837, 21752	
rs9268832	32535767	G	0.819, 0.597	1.82 E-52	HLA-DRA / HLA-DRB5	CIT, 8.33 E-54, 2	32522251, 32535767, 13517	[32, 39, 41, 96]
						FGT, 8.33 E-54, 2	32522251, 32535767, 13517	
						SSLD, 3.47 E-77, 9	32519501, 32678378, 158878	
rs9275224	32767856	G	0.804, 0.511	1.74 E-81	HLA-DQB1 / HLA-DQA2	CIT, 2.62 E-81, 3	32767856, 32771829, 3974	[32, 36, 39, 99, 101]
						FGT, 1.03 E-48, 4	32766602, 32771829, 5228	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	

(Continued)

PLOS ONE

Does the method of haplotype partitioning affect the association results of chromosome 6 of the NARAC dataset?

Table 4. (Continued)

SNP ID	Position (bp)	Assoc. Allele[a]	AAF[b] (Case, Control)	P-value	Gene / Nearest Genes	Haplotype Block (Method, P-value, No. of SNPs in Block)	Haplotype Block Position (bp) (Start, End, Size)	Previously Studied in
rs9275312	32773706	G	0.297, 0.121	7.05 E-45	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[33, 41]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs9275371	32776274	G	0.488, 0.259	2.57 E-47	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[46, 99]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs9275374	32776504	A	0.482, 0.246	1.69 E-55	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45, 47, 102]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs9275383	32776824	A	0.261, 0.085	6.20 E-50	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[103]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs9275388	32777062	G	0.482, 0.241	7.21 E-57	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[33, 45]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs9275390	32777134	G	0.482, 0.246	1.69 E-55	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[47, 51, 57, 64, 75, 99, 104]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs9275393	32777417	A	0.482, 0.246	2.72 E-55	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45, 99]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs9275406	32777933	A	0.482, 0.246	1.38 E-55	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[41, 45, 51, 105]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.49 E-53, 20	32767856, 32777978, 10123	
rs9275407	32778015	A	0.483, 0.245	9.06 E-56	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45, 99]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275408	32778088	G	0.463, 0.231	6.75 E-52	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	-
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275418	32778222	G	0.482, 0.246	1.4 E-55	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45, 106–108]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275424	32778554	G	0.482, 0.246	1.4 E-55	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45, 99]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275425	32778852	A	0.477, 0.244	5.74 E-54	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45, 99]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275427	32778893	A	0.482, 0.246	1.86 E-55	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45]

(Continued)

Table 4. (Continued)

SNP ID	Position (bp)	Assoc. Allele[a]	AAF[b] (Case, Control)	P-value	Gene / Nearest Genes	Haplotype Block (Method, P-value, No. of SNPs in Block)	Haplotype Block Position (bp) (Start, End, Size)	Previously Studied in
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275428	32778956	G	0.482, 0.247	3.3 E-55	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45, 99, 109]
						FGT, 3.26 E-52, 28	32772436, 32778956, 6521	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275439	32779499	G	0.479, 0.245	9.76 E-55	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[41, 45, 99]
						FGT, 2.66 E-54, 2	32779081, 32779499, 419	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275555	32785066	A	0.475, 0.222	3.04 E-63	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[45, 46]
						FGT, 3.19 E-61, 3	32785066, 32786977, 1912	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275572	32786977	G	0.822, 0.599	7.07 E-53	HLA-DQB1 / HLA-DQA2	CIT, 3.68 E-54, 31	32772436, 32786977, 14542	[40, 45, 110]
						FGT, 3.19 E-61, 3	32785066, 32786977, 1912	
						SSLD, 1.55 E-52, 17	32778015, 32787668, 9654	
rs9275595	32789333	G	0.462, 0.218	2.59 E-61	HLA-DQB1/ HLA-DQA2	CIT, 6.72 E-54, 7	32788906, 32790115, 1210	[39, 40, 45, 104, 111]
						SSLD, 6.72 E-54, 7	32788906, 32790115, 1210	

[a]Assoc. Allele: Associated Allele.

[b]AAF: Associated Allele Frequency.

methods and individual SNP approach). The three used haplotype block methods were CIT, FGT, and SSLD. The individual SNP analysis was to point to chromosome regions that are genetically linked to the disease. The haplotype block methods were to further expand from the single SNPs with the strongest signal to the actual causal variants [112].

The aim of this study is to test the NARAC dataset to find out if haplotype block methods or single-locus approach alone can sufficiently provide the significant biomarkers associated with RA. Our research failed to select the best method as each method yielded significant results that were not detected using any of the other methods. S2 Table showed the SNP IDs that were uniquely identified by each method. The individual SNP, CIT, FGT, and SSLD methods exclusively detected 12, 55, 37, and 189 SNPs, respectively. Our findings were in line with Shim et al. conclusion (but they didn't test the SSLD method) that both the individual SNP approach and the haplotype block methods should be applied side by side to discover the valuable associations in the NARAC dataset [5].

In addition, the 367 SNPs- that were significantly associated with RA susceptibility by the individual SNP approach and the haplotype block methods- represent potent candidates for further investigations (S6 Spreadsheet). The three haplotype block methods were able to detect 916 associated SNPs in common. The SSLD method detected more significant SNPs (1,322) than CIT (1,086), FGT (1,099), and individual SNP (432) methods. This observation could be understood, as the SSLD does not take into account the LD between the intermediate SNPs. Therefore, the SSLD method is the lowest conservative method in including SNPs inside it's blocks.

The block similarity for the three applied methods were shown in Table 5. The similarity measure represented the intersected SNPs divided by the union SNPs for the two studied

Table 5. Block similarity among the used haplotype block methods.

Block Method	CIT	FGT	SSLD
CIT	1	0.76	0.71
FGT		1	0.74
SSLD			1

methods. The highest block similarity was between the CIT method and the FGT method. While, the lowest block similarity was between the CIT method and the SSLD method. The results showed that the FGT method had the most similarity with the other methods.

Fig 4 demonstrated the overlapping of the significant blocks resulted from the three haplotype block methods. Nearly, all chromosome 6 regions that were associated with RA were in
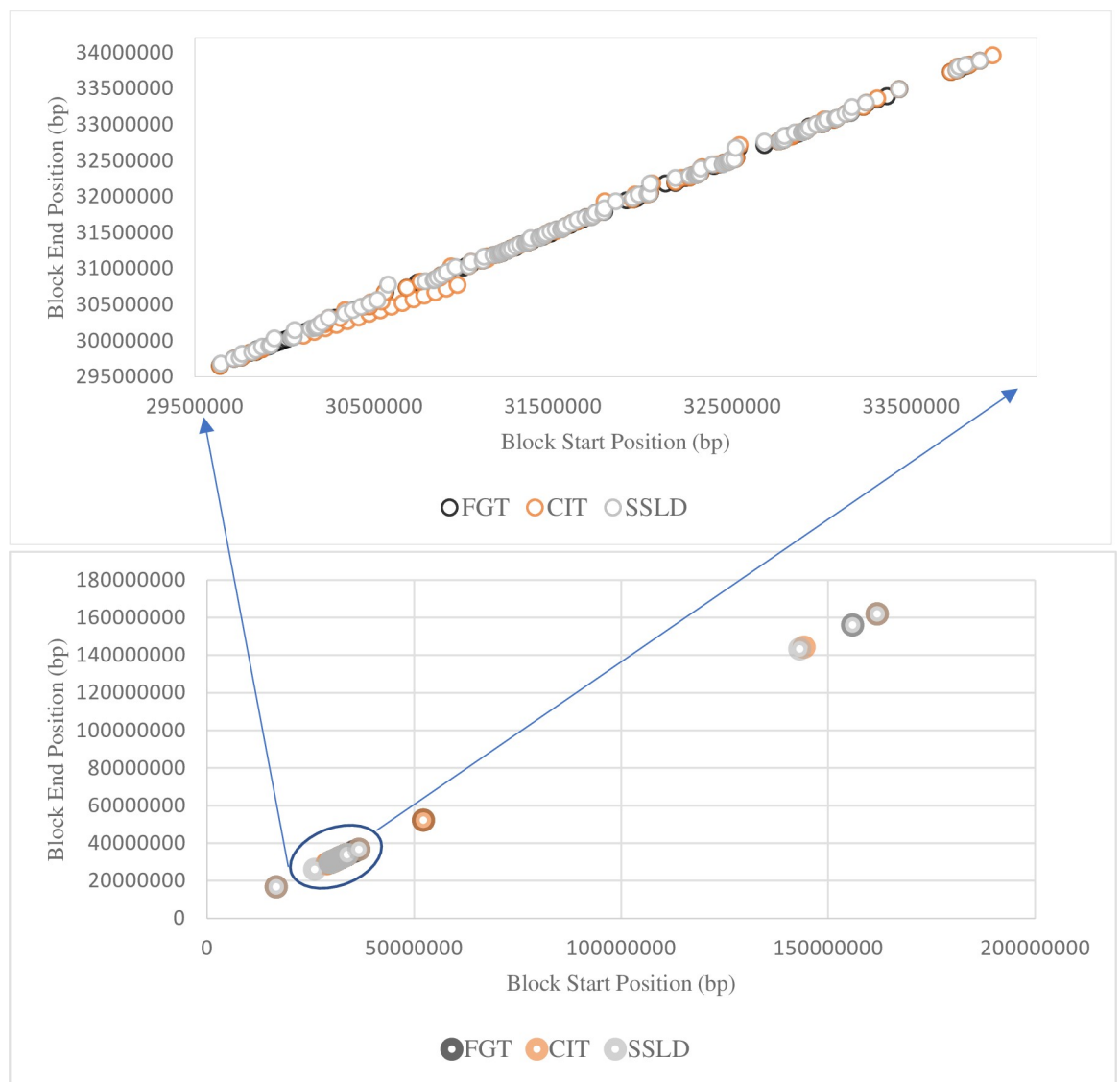
Fig 4. Comparison of the haplotype blocks obtained by the three methods. Each circle conistituted a significant haplotype block associated with RA susceptibility. The plot area represented chromosome 6, then zooming in on the MHC region.

PLOS ONE

Does the method of haplotype partitioning affect the association results of chromosome 6 of the NARAC dataset?

blocks that were detected by more than one method. Most of the haplotype blocks that showed significant associations with RA disease were in the MHC region or near it (+ 3 Mb). Most of the 916 SNPs that were detected by the three block methods were in the MHC region. These outcomes confirmed the strong association between the MHC region and RA susceptibility.

For further analysis of the results, a comparison between the properties of the SNPs that were detected by all the methods (intersection) and that were uniquely detected by one method (unique) was shown in Table 6. The properties of all the associated SNPs detected by the corresponding method (reference) were added to Table 6 for more clarification. For all the methods, the P-values of the (intersection) were lower than that of the (reference) and more lower than that of the (unique). This observation confirmed that the 367 SNPs (intersection) represents potent candidates for further investigations. For the CIT and FGT methods, the median block size (bp) of the (unique) was greater than that of the (intersection) and more greater than the (reference). However, this observation might be due to the small number of blocks representing the (unique) in comparison to the (intersection) and the (reference). While, when the number of blocks representing the (unique) (53) was sufficiently comparable to that of the (intersection) (104) and the (reference) (145) in the SSLD method, the median block size (bp) of the (intersection) was greater than that of the (unique) and more greater than that of the (reference).

Some associated SNPs were discovered using all the methods but others were observed by only one method. This finding might be explained by some reasons. For the associations that were observed using individual SNP approach only, it may be that only one SNP represent strong LD with the causal SNP. Therefore, studying haplotypes could decrease the power of association as they consist of several SNPs.

Table 6. A comparison between the associated SNPs detected with different categories.

Method	Items	Significant SNPs detected by all methods (intersection)	All significant SNPs detected by a method (reference)	Significant SNPs detected by only one method (unique)
CIT	# SNPs	367	1,086	55
	# blocks	127	160	8
	Average P-values	1.63 E-07	7.20 E-07	3.33 E-06
	Median Block Size (#SNPs)	5	5	7
	Median Block Size (BP)	9,755	8,671	17,021
FGT	# SNPs	367	1,099	37
	# blocks	134	179	12
	Average P-values	1.37 E-07	5.6 E-07	7.84 E-07
	Median Block Size (#SNPs)	6	5	7
	Median Block Size (BP)	9,227	7,447	11,394
SSLD	# SNPs	367	1,322	189
	# blocks	104	145	53
	Average P-values	1.86 E-07	6.57 E-07	9.34 E-07
	Median Block Size (#SNPs)	9	7	8
	Median Block Size (BP)	17,054	10,122	11,604
Individual SNP	# SNPs	367	432	12
	Average P-values	1.25 E-07	1.35 E-07	3.26 E-07

For the associations that were observed using the haplotype block methods only, Individual SNP approach required approximately 83% more tests than the haplotype block methods. Consequently, the Bonferroni correction was more severe for the individual SNP approach. Moreover, the haplotype block methods were able to capture the interactions among many causal SNPs. In addition, haplotypes could capture rare alleles that individual SNPs may not detect. This reason could be clarified as the power to observe associations is maximized when the frequencies of the studied biomarker and the causal SNP are similar.

For the associations that were observed using a haplotype block method but not by the other haplotype block methods, each method differs greatly from the others in its scope of the definition of the haplotype blocks. At last, we conclude that the application of the individual SNP approach and the three haplotype block methods altogether on chromosome 6 of the NARAC dataset will in turn maximize the system's ability for discovering crucial associations. In case of selecting one method, the SSLD would be the most appropriate method for the NARAC dataset. The SSLD method has valuable advantages such as the highest genomic coverage, the largest minimum, median, and maximum significant block sizes, the biggest number of significant SNPs included in blocks, and the biggest number of associated SNPs discovered exclusively by a method.

The limitations of this study are as follows: a) the effects of population stratification were not accounted for; b) a replication study in other datasets was not performed. In addition, further investigations of other haplotype block methods, such as hidden Markov model [113, 114], dynamic programming-based algorithm [115–119], wavelet decomposition [120], greedy algorithm [121], minimum description length [122, 123], spatial correlation of SNPs [124], sequence kernel association tests [125], and block entropy [126] should be applied and compared to show the effect of the changes in block partitions on the resulting associated biomarkers.

Supporting information

S1 Fig. The confidence interval test (CIT) method showing the definition of strong LD and weak LD.
(TIF)

S2 Fig. LD block structure as found by the solid spine of linkage disequilibrium (SSLD) method.
(TIF)

S1 Table. The processing time for each performed step.
(DOCX)

S2 Table. SNP IDs that were uniquely identified by each method.
(DOCX)

S1 Spreadsheet. Properties for the associated SNPs using CIT method. Sheet1 "ID" represents SNPs IDs, and each row represents a block. Sheet2 "Bp" represents SNPs physical positions in base pairs, and each row represents a block. Sheet3 "No. of SNPs in Block" represents the number of SNPs in each block. Sheet4 "Start-Stop"–first column represents blocks start physical positions in base pairs. Sheet4 "Start-Stop"–second column represents blocks end physical positions in base pairs. Sheet4 "Start-Stop"–third column represents blocks sizes in base pairs. Sheet5 "Block no." represents the block numbers (positions) from all blocks partitioned by CIT method. Sheet6 "P-values" represents the P-values of the blocks.
(XLS)

PLOS ONE

Does the method of haplotype partitioning affect the association results of chromosome 6 of the NARAC dataset?

S2 Spreadsheet. Properties for the associated SNPs using FGT method. Sheet1 "ID" represents SNPs IDs, and each row represents a block. Sheet2 "Bp" represents SNPs physical positions in base pairs, and each row represents a block. Sheet3 "No. of SNPs in Block" represents the number of SNPs in each block. Sheet4 "Start-Stop"–first column represents blocks start physical positions in base pairs. Sheet4 "Start-Stop"–second column represents blocks end physical positions in base pairs. Sheet4 "Start-Stop"–third column represents blocks sizes in base pairs. Sheet5 "Block no." represents the block numbers (positions) from all blocks partitioned by FGT method. Sheet6 "*P*-values" represents the *P*-values of the blocks.
(XLS)

S3 Spreadsheet. Properties for the associated SNPs using SSLD method. Sheet1 "ID" represents SNPs IDs, and each row represents a block. Sheet2 "Bp" represents SNPs physical positions in base pairs, and each row represents a block. Sheet3 "No. of SNPs in Block" represents the number of SNPs in each block. Sheet4 "Start-Stop"–first column represents blocks start physical positions in base pairs. Sheet4 "Start-Stop"–second column represents blocks end physical positions in base pairs. Sheet4 "Start-Stop"–third column represents blocks sizes in base pairs. Sheet5 "Block no." represents the block numbers (positions) from all blocks partitioned by SSLD method. Sheet6 "*P*-values" represents the *P*-values of the blocks.
(XLS)

S4 Spreadsheet. Properties for the associated SNPs using individual SNP approach. Sheet1 "ID" represents SNPs IDs. Sheet2 "Bp" represents SNPs physical positions in base pairs. Sheet6 "*P*-values" represents the *P*-values of the SNPs.
(XLS)

S5 Spreadsheet. Intersection between haplotype methods and individual SNP approach. Sheet1 "CIT" represents SNPs IDs detected by both CIT method and Individual SNP Approach. Sheet2 "FGT" represents SNPs IDs detected by both FGT method and Individual SNP Approach. Sheet3 "SSLD" represents SNPs IDs detected by both SSLD method and Individual SNP Approach.
(XLS)

S6 Spreadsheet. The SNPs IDs identified by all the used methods.
(XLS)

Author Contributions

Conceptualization: Mohamed N. Saad, Mai S. Mabrouk, Ayman M. Eldeib, Olfat G. Shaker.

Data curation: Mohamed N. Saad.

Formal analysis: Mohamed N. Saad.

Investigation: Mohamed N. Saad.

Methodology: Mohamed N. Saad, Mai S. Mabrouk, Ayman M. Eldeib, Olfat G. Shaker.

Resources: Mai S. Mabrouk.

Software: Mohamed N. Saad.

Supervision: Mai S. Mabrouk, Ayman M. Eldeib, Olfat G. Shaker.

Validation: Mohamed N. Saad.

Visualization: Mohamed N. Saad.

Writing – original draft: Mohamed N. Saad.

Writing – review & editing: Mohamed N. Saad, Mai S. Mabrouk, Ayman M. Eldeib, Olfat G. Shaker.

References

1. Saad M. N., Mabrouk M. S., Eldeib A. M. and Shaker O. G., " Identification of rheumatoid arthritis bio-markers based on single nucleotide polymorphisms and haplotype blocks: a systematic review and meta-analysis," J Adv Res, vol. 7, pp. 1–16, 2016. https://doi.org/10.1016/j.jare.2015.01.008 PMID: 26843965

2. Peise E., Fabregat-Traver D. and Bientinesi P., "High performance solutions for big-data GWAS," Parallel Computing, vol. 42, pp. 75–87, 2015.

3. Clark A. G., "The role of haplotypes in candidate gene studies," Genet Epidemiol, vol. 27, pp. 321–33, Dec 2004. https://doi.org/10.1002/gepi.20025 PMID: 15368617

4. Su S. C., Kuo C. C. and Chen T., "Single nucleotide polymorphism data analysis—state-of-the-art review on this emerging field from a signal processing viewpoint," IEEE Signal Process Mag, vol. 24, pp. 75–82, 2007.

5. Shim H., Chun H., Engelman C. D. and Payseur B. A., "Genome-wide association studies using single-nucleotide polymorphisms versus haplotypes: an empirical comparison with data from the North American Rheumatoid Arthritis Consortium," BMC Proc, vol. 3 Suppl 7, p. S35, 2009.

6. Saad M. N., Mabrouk M. S., Eldeib A. M. and Shaker O. G., "Genetic case-control study for eight polymorphisms associated with rheumatoid arthritis," PLoS One, vol. 10, e0131960, 2015. https://doi.org/10.1371/journal.pone.0131960 PMID: 26147289

7. Caliz R., Canet L. M., Lupianez C. B., Canhao H., Escudero A., Filipescu I., et al., "Gender-specific effects of genetic variants within Th1 and Th17 cell-mediated immune response genes on the risk of developing rheumatoid arthritis," PLoS One, vol. 8, p. e72732, 2013. https://doi.org/10.1371/journal.pone.0072732 PMID: 24023637

8. M. N. Saad, M. S. Mabrouk, A. M. Eldeib and O. G. Shaker, "Vitamin D receptor gene polymorphisms in rheumatoid arthritis patients associating osteoporosis," presented at the 7th Cairo International Biomedical Engineering Conference, Cairo, Egypt, 2014.

9. Saad M. N., Mabrouk M. S., Eldeib A. M. and Shaker O. G., "Effect of MTHFR, TGFβ1 and TNFB polymorphisms on osteoporosis in rheumatoid arthritis patients," Gene, vol. 568, pp. 124–128, 2015. https://doi.org/10.1016/j.gene.2015.05.037 PMID: 25981594

10. Kitajima H., Sonoda M. and Yamamoto K., "HLA and SNP haplotype mapping in the Japanese population," Genes Immun, vol. 13, pp. 543–8, Oct 2012. https://doi.org/10.1038/gene.2012.35 PMID: 22914434

11. Sekar A., Bialas A. R., de Rivera H., Davis A., Hammond T. R., Kamitaki N., et al., "Schizophrenia risk from complex variation of complement component 4," Nature, vol. 530, pp. 177–183, 2016. https://doi.org/10.1038/nature16549 PMID: 26814963

12. Perricone C., Ceccarelli F. and Valesini G., "An overview on the genetic of rheumatoid arthritis: a never-ending story," Autoimmun Rev, vol. 10, pp. 599–608, Aug 2011. https://doi.org/10.1016/j.autrev.2011.04.021 PMID: 21545847

13. Amos C. I., Chen W. V., Seldin M. F., Remmers E. F., Taylor K. E., Criswell L. A., et al., "Data for Genetic Analysis Workshop 16 Problem 1, association analysis of rheumatoid arthritis data," BMC Proceedings, vol. 3, pp. 1–4, 2009.

14. Purcell S., Neale B., Todd-Brown K., Thomas L., Ferreira M. A., Bender D., et al., "PLINK: a tool set for whole-genome association and population-based linkage analyses," Am J Hum Genet, vol. 81, pp. 559–75, Sep 2007. https://doi.org/10.1086/519795 PMID: 17701901

15. Barrett J. C., Fry B., Maller J. and Daly M. J., "Haploview: analysis and visualization of LD and haplotype maps," Bioinformatics, vol. 21, pp. 263–5, Jan 2005. https://doi.org/10.1093/bioinformatics/bth457 PMID: 15297300

16. Robinson J. T., Thorvaldsdottir H., Winckler W., Guttman M., Lander E. S., Getz G., et al., "Integrative genomics viewer," Nat Biotechnol, vol. 29, pp. 24–6, Jan 2011. https://doi.org/10.1038/nbt.1754 PMID: 21221095

17. Thorvaldsdottir H., Robinson J. T. and Mesirov J. P., "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration," Brief Bioinform, vol. 14, pp. 178–92, Mar 2013. https://doi.org/10.1093/bib/bbs017 PMID: 22517427

18. Wall J. D. and Pritchard J. K., "Haplotype blocks and linkage disequilibrium in the human genome," Nat Rev Genet, vol. 4, pp. 587–97, Aug 2003. https://doi.org/10.1038/nrg1123 PMID: 12897771

19. Gabriel S. B., Schaffner S. F., Nguyen H., Moore J. M., Roy J., Blumenstiel B., et al., "The structure of haplotype blocks in the human genome," Science, vol. 296, pp. 2225–9, Jun 21 2002. https://doi.org/10.1126/science.1069424 PMID: 12029063

20. Wang N., Akey J. M., Zhang K., Chakraborty R. and Jin L., "Distribution of recombination crossovers and the origin of haplotype blocks: the interplay of population history, recombination and mutation," Am J Hum Genet, vol. 71, pp. 1227–34, Nov 2002. https://doi.org/10.1086/344398 PMID: 12384857

21. Balding D. J., "A tutorial on statistical methods for population association studies," Nat Rev Genet, vol. 7, pp. 781–91, Oct 2006. https://doi.org/10.1038/nrg1916 PMID: 16983374

22. Saxena R., Voight B. F., Lyssenko V., Burtt N. P., de Bakker P. I. W., Chen H., et al., "Genome-Wide Association Analysis Identifies Loci for Type 2 Diabetes and Triglyceride Levels," Science, vol. 316, pp. 1331–1336, 2007. https://doi.org/10.1126/science.1142358 PMID: 17463246

23. Wu Y., Fan H., Wang Y., Zhang L., Gao X., Chen Y., et al., "Genome-wide association studies using haplotypes and individual SNPs in Simmental cattle," PloS one, vol. 9, e109330, 2014. https://doi.org/10.1371/journal.pone.0109330 PMID: 25330174

24. Arya R., Hare E., del Rincon I., Jenkinson C. P., Duggirala R., Almasy L., et al., "Effects of covariates and interactions on a genome-wide association analysis of rheumatoid arthritis," BMC Proceedings, vol. 3, p. S84, 2009. PMID: 20018080

25. Mo X.-B., Zhang Y.-H. and Lei S.-F., "Genome-Wide Identification of N(6)-Methyladenosine (m(6)A) SNPs Associated With Rheumatoid Arthritis," Frontiers in genetics, vol. 9, pp. 299–299, 2018. https://doi.org/10.3389/fgene.2018.00299 PMID: 30123242

26. Fattal I., Rimer J., Shental N., Molad Y., Gabrielli A., Livneh A., et al., "Pemphigus vulgaris is characterized by low IgG reactivities to specific self-antigens along with high IgG reactivity to desmoglein 3," Immunology, vol. 143, pp. 374–380, 2014. https://doi.org/10.1111/imm.12316 PMID: 24820664

27. Kim K., Kim H. and Yoon J., "The expression and role of KLF4 in psoriasis," in JOURNAL OF INVESTIGATIVE DERMATOLOGY, 2013, pp. S17–S17.

28. Orozco G., Eerligh P., Sánchez E., Zhernakova S., Roep B. O., González-Gay M. A., et al., "Analysis of a Functional BTNL2 Polymorphism in Type 1 Diabetes, Rheumatoid Arthritis and Systemic Lupus Erythematosus," Human Immunology, vol. 66, pp. 1235–1241, 2005/12/01/ 2005. https://doi.org/10.1016/j.humimm.2006.02.003 PMID: 16690410

29. Steer S., Abkevich V., Gutin A., Cordell H. J., Gendall K. L., Merriman M. E., et al., "Genomic DNA pooling for whole-genome association scans in complex disease: empirical demonstration of efficacy in rheumatoid arthritis," Genes And Immunity, vol. 8, p. 57, 12/07/online 2006. https://doi.org/10.1038/sj.gene.6364359 PMID: 17159887

30. Mitsunaga S., Hosomichi K., Okudaira Y., Nakaoka H., Kunii N., Suzuki Y., et al., "Exome sequencing identifies novel rheumatoid arthritis-susceptible variants in the BTNL2," Journal Of Human Genetics, vol. 58, p. 210, 01/31/online 2013. https://doi.org/10.1038/jhg.2013.2 PMID: 23364395

31. Wang M., Chen X., Zhang M., Zhu W., Cho K. and Zhang H., "Detecting significant single-nucleotide polymorphisms in a rheumatoid arthritis study using random forests," BMC Proceedings, vol. 3, p. S69, December 15 2009. PMID: 20018063

32. Jiang R., Dong J. and Dai Y., "Genome-wide association study of rheumatoid arthritis by a score test based on wavelet transformation," BMC Proceedings, vol. 3, p. S8, 2009/12/15 2009.

33. Chiu Y.-F., Kao H.-Y., Chen Y.-S., Hsu F.-C. and Yang H.-C., "Assessment of gene-covariate interactions by incorporating covariates into association mapping," BMC Proceedings, vol. 3, p. S85, December 15 2009. PMID: 20018081

34. Liu Y., Aryee M. J., Padyukov L., Fallin M. D., Hesselberg E., Runarsson A., et al., "Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis," Nature Biotechnology, vol. 31, p. 142, 01/20/online 2013. https://doi.org/10.1038/nbt.2487 PMID: 23334450

35. Chen L., Zhong M., Chen W. V., Amos C. I. and Fan R., "A genome-wide association scan for rheumatoid arthritis data by Hotelling's T2tests," BMC Proceedings, vol. 3, p. S6, 2009/12/15 2009.

36. Park J., Namkung J., Jhun M. and Park T., "Genome-wide analysis of haplotype interaction for the data from the North American Rheumatoid Arthritis Consortium," BMC Proc, vol. 3(Suppl 7):S34, 2009.

37. Black M. H. and Watanabe R. M., "A principal-components-based clustering method to identify multiple variants associated with rheumatoid arthritis and arthritis-related autoantibodies," BMC Proceedings, vol. 3, p. S129, December 15 2009. PMID: 20017995

38. Üstünkar G. and Son Yeşim A., "METU-SNP: An Integrated Software System for SNPComplex Disease Association Analysis," in Journal of Integrative Bioinformatics vol. 8, ed, 2011, p. 204.

PLOS | ONE

Does the method of haplotype partitioning affect the association results of chromosome 6 of the NARAC dataset?

39. Taliun D., Gamper J. and Pattaro C., "Efficient haplotype block recognition of very long and dense genetic sequences," BMC Bioinformatics, vol. 15, p. 10, January 14 2014. https://doi.org/10.1186/1471-2105-15-10 PMID: 24423111

40. Cho S., Kim H., Oh S., Kim K. and Park T., "Elastic-net regularization approaches for genome-wide association studies of rheumatoid arthritis," BMC Proceedings, vol. 3, p. S25, December 15 2009. PMID: 20018015

41. Lemire M., "On the association between rheumatoid arthritis and classical HLA class I and class II alleles predicted from single-nucleotide polymorphism data," BMC Proceedings, vol. 3, p. S33, December 15 2009. PMID: 20018024

42. Jung J., Song J. J. and Kwon D., "Allelic based gene-gene interactions in rheumatoid arthritis," BMC Proceedings, vol. 3, p. S76, 2009/12/15 2009. https://doi.org/10.1186/1753-6561-3-S7-S76 PMID: 20018071

43. 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, pp. bat073–bat073, 2013. https://doi.org/10.1093/database/bat073 PMID: 24158836

44. Tamiya G., Shinya M., Imanishi T., Ikuta T., Makino S., Okamoto K., et al., "Whole genome association study of rheumatoid arthritis using 27 039 microsatellites," Human Molecular Genetics, vol. 14, pp. 2305–2321, 2005. https://doi.org/10.1093/hmg/ddi234 PMID: 16000323

45. Sengupta Chattopadhyay A., Hsiao C.-L., Chang C. C., Lian I.-B. and Fann C. S. J., "Summarizing techniques that combine three non-parametric scores to detect disease-associated 2-way SNP-SNP interactions," Gene, vol. 533, pp. 304–312, 2014/01/01 2014. https://doi.org/10.1016/j.gene.2013.09.041 PMID: 24076437

46. Chanda P., Zhang A., Sucheston L. and Ramanathan M., "A two-stage search strategy for detecting multiple loci associated with rheumatoid arthritis," BMC Proceedings, vol. 3, p. S72, December 15 2009. PMID: 20018067

47. Alekseyenko A. V., Lytkin N. I., Ai J., Ding B., Padyukov L., Aliferis C. F., et al., "Causal graph-based analysis of genome-wide association data in rheumatoid arthritis," Biology Direct, vol. 6, p. 25, May 18 2011. https://doi.org/10.1186/1745-6150-6-25 PMID: 21592391

48. Orozco G., Viatte S., Bowes J., Martin P., Wilson A. G., Morgan A. W., et al., "Novel Rheumatoid Arthritis Susceptibility Locus at 22q12 Identified in an Extended UK Genome-Wide Association Study," Arthritis & Rheumatology, vol. 66, pp. 24–30, 2014.

49. Roederer M., Quaye L., Mangino M., Beddall Margaret H., Mahnke Y., Chattopadhyay P., et al., "The Genetic Architecture of the Human Immune System: A Bioresource for Autoimmunity and Disease Pathogenesis," Cell, vol. 161, pp. 387–403, 2015/04/09/ 2015. https://doi.org/10.1016/j.cell.2015.02.046 PMID: 25772697

50. Zheng W. and Rao S., "Knowledge-based analysis of genetic associations of rheumatoid arthritis to inform studies searching for pleiotropic genes: a literature review and network analysis," Arthritis Research & Therapy, vol. 17, p. 202, August 08 2015.

51. Wallace A. D., Wendt G. A., Barcellos L. F., de Smith A. J., Walsh K. M., Metayer C., et al., "To ERV Is Human: A Phenotype-Wide Scan Linking Polymorphic Human Endogenous Retrovirus-K Insertions to Complex Phenotypes," Frontiers in Genetics, vol. 9, 2018-August-14 2018.

52. Ju J. H., Shenoy S. A., Crystal R. G. and Mezey J. G., "An independent component analysis confounding factor correction framework for identifying broad impact expression quantitative trait loci," PLOS Computational Biology, vol. 13, p. e1005537, 2017. https://doi.org/10.1371/journal.pcbi.1005537 PMID: 28505156

53. Fang H., Knezevic B., Burnham K. L. and Knight J. C., "XGR software for enhanced interpretation of genomic summary data, illustrated by application to immunological traits," Genome Medicine, vol. 8, p. 129, December 13 2016. https://doi.org/10.1186/s13073-016-0384-y PMID: 27964755

54. Goris A., Pauwels I., Gustavsen M. W., van Son B., Hilven K., Bos S. D., et al., "Genetic variants are major determinants of CSF antibody levels in multiple sclerosis," Brain, vol. 138, pp. 632–643, 2015. https://doi.org/10.1093/brain/awu405 PMID: 25616667

55. Rivera N. V., Ronninger M., Shchetynsky K., Franke A., Nöthen M. M., Müller-Quernheim J., et al., "High-Density Genetic Mapping Identifies New Susceptibility Variants in Sarcoidosis Phenotypes and Shows Genomic-driven Phenotypic Differences," American Journal of Respiratory and Critical Care Medicine, vol. 193, pp. 1008–1022, 2016. https://doi.org/10.1164/rccm.201507-1372OC PMID: 26651848

56. Rai E. and Wakeland E. K., "Genetic predisposition to autoimmunity–What have we learned?," Seminars in Immunology, vol. 23, pp. 67–83, 2011/04/01/ 2011. https://doi.org/10.1016/j.smim.2011.01.015 PMID: 21288738

57. Wei W.-H., Loh C.-Y., Worthington J. and Eyre S., "Immunochip Analyses of Epistasis in Rheumatoid Arthritis Confirm Multiple Interactions within MHC and Suggest Novel Non-MHC Epistatic Signals," The Journal of Rheumatology, 2016.

58. Klareskog L., Stolt P., Lundberg K., Källberg H., Bengtsson C., Grunewald J., et al., "A new model for an etiology of rheumatoid arthritis: Smoking may trigger HLA–DR (shared epitope)–restricted immune reactions to autoantigens modified by citrullination," Arthritis & Rheumatism, vol. 54, pp. 38–46, 2006.

59. MacGregor A. J., Snieder H., Rigby A. S., Koskenvuo M., Kaprio J., Aho K., et al., "Characterizing the quantitative genetic contribution to rheumatoid arthritis using data from twins," Arthritis & Rheumatism, vol. 43, pp. 30–37, 2000.

60. Ding B., Padyukov L., Lundstrom E., Seielstad M., Plenge R. M., Oksenberg J. R., et al., "Different patterns of associations with anti-citrullinated protein antibody-positive and anti-citrullinated protein antibody-negative rheumatoid arthritis in the extended major histocompatibility complex region," Arthritis Rheum, vol. 60, pp. 30–8, Jan 2009. https://doi.org/10.1002/art.24135 PMID: 19116921

61. Cui J., Saevarsdottir S., Thomson B., Padyukov L., van der Helm-Van Mil A. H. M., Nititham J., et al., "Rheumatoid arthritis risk allele PTPRC is also associated with response to anti–tumor necrosis factor α therapy," Arthritis & Rheumatism, vol. 62, pp. 1849–1861, 2010.

62. Han T.-U., Bang S.-Y., Kang C. and Bae S.-C., "TRAF1 polymorphisms associated with rheumatoid arthritis susceptibility in Asians and in Caucasians," Arthritis & Rheumatism, vol. 60, pp. 2577–2584, 2009.

63. Freudenberg J., Lee H.-S., Han B.-G., Shin H. D., Kang Y. M., Sung Y.-K., et al., "Genome-wide association study of rheumatoid arthritis in Koreans: Population-specific loci as well as overlap with European susceptibility loci," Arthritis & Rheumatism, vol. 63, pp. 884–893, 2011.

64. Julia A., Ballina J., Canete J. D., Balsa A., Tornero-Molina J., Naranjo A., et al., "Genome-wide association study of rheumatoid arthritis in the Spanish population: KLF12 as a risk locus for rheumatoid arthritis susceptibility," Arthritis Rheum, vol. 58, pp. 2275–86, Aug 2008. https://doi.org/10.1002/art.23623 PMID: 18668548

65. Prasad P., Kumar A., Gupta R., Juyal R. C. and T. B. K, "Caucasian and Asian Specific Rheumatoid Arthritis Risk Loci Reveal Limited Replication and Apparent Allelic Heterogeneity in North Indians," PLOS ONE, vol. 7, p. e31584, 2012. https://doi.org/10.1371/journal.pone.0031584 PMID: 22355377

66. Snir O., Gomez-Cabrero D., Montes A., Perez-Pampin E., Gómez-Reino J. J., Seddighzadeh M., et al., "Non-HLA genes PTPN22, CDK6 and PADI4 are associated with specific autoantibodies in HLA-defined subgroups of rheumatoid arthritis," Arthritis Research & Therapy, vol. 16, p. 414, August 20 2014.

67. Rioux J. D., Goyette P., Vyse T. J., Hammarström L., Fernando M. M. A., Green T., et al., "Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases," Proceedings of the National Academy of Sciences, vol. 106, pp. 18680–18685, 2009.

68. Delgado-Vega A., Sánchez E., Löfgren S., Castillejo-López C. and Alarcón-Riquelme M. E., "Recent findings on genetics of systemic autoimmune diseases," Current Opinion in Immunology, vol. 22, pp. 698–705, 2010/12/01/ 2010. https://doi.org/10.1016/j.coi.2010.09.002 PMID: 20933377

69. Ritchie M. D., Denny J. C., Crawford D. C., Ramirez A. H., Weiner J. B., Pulley J. M., et al., "Robust Replication of Genotype-Phenotype Associations across Multiple Diseases in an Electronic Medical Record," The American Journal of Human Genetics, vol. 86, pp. 560–572, 2010/04/09/ 2010. https://doi.org/10.1016/j.ajhg.2010.03.003 PMID: 20362271

70. Li H., Hu Y., Zhang T., Liu Y., Wang Y., Yang T., et al., "Replication of British Rheumatoid Arthritis Susceptibility Loci in Two Unrelated Chinese Population Groups," Clinical and Developmental Immunology, vol. 2013, p. 6, 2013.

71. Lu Q., Cui Y., Ye C., Wei C. and Elston R. C., "Bagging Optimal ROC Curve Method for Predictive Genetic Tests, with an Application for Rheumatoid Arthritis," Journal of Biopharmaceutical Statistics, vol. 20, pp. 401–414, 2010/03/19 2010. https://doi.org/10.1080/10543400903572811 PMID: 20309765

72. Gall B. J., Wilson A., Schroer A. B., Gross J. D., Stoilov P., Setola V., et al., "Genetic variations in GPSM3 associated with protection from rheumatoid arthritis affect its transcript abundance," Genes And Immunity, vol. 17, p. 139, 01/28/online 2016. https://doi.org/10.1038/gene.2016.3 PMID: 26821282

73. Bermejo J. L., Fischer C., Schulz A., Cremer N., Hein R., Beckmann L., et al., "Representation of genetic association via attributable familial relative risks in order to identify polymorphisms functionally relevant to rheumatoid arthritis," BMC Proceedings, vol. 3, p. S10, December 15 2009.

74. Jeffries N. and Zheng G., "Evaluation of an optimal receiver operating characteristic procedure," BMC Proceedings, vol. 3, p. S56, December 15 2009. PMID: 20018049

75. Achour Y., Ben Hamad M., Chaabane S., Rebai A., Marzouk S., Mahfoudh N., et al., "Analysis of two susceptibility SNPs in HLA region and evidence of interaction between rs6457617 in HLA-DQB1 and HLA-DRB1*04 locus on Tunisian rheumatoid arthritis," Journal of Genetics, vol. 96, pp. 911–918, December 01 2017. PMID: 29321349

76. Mahdi H., Fisher B. A., Källberg H., Plant D., Malmström V., Rönnelid J., et al., "Specific interaction between genotype, smoking and autoimmunity to citrullinated α-enolase in the etiology of rheumatoid arthritis," Nature Genetics, vol. 41, p. 1319, 11/08/online 2009. https://doi.org/10.1038/ng.480 PMID: 19898480

77. Vignal C., Bansal A. T., Balding D. J., Binks M. H., Dickson M. C., Montgomery D. S., et al., "Genetic association of the major histocompatibility complex with rheumatoid arthritis implicates two non-DRB1 loci," Arthritis & Rheumatism, vol. 60, pp. 53–62, 2009.

78. Yarwood A., Han B., Raychaudhuri S., Bowes J., Lunt M., Pappas D. A., et al., "A weighted genetic risk score using all known susceptibility variants to estimate rheumatoid arthritis risk," Annals of the Rheumatic Diseases, vol. 74, pp. 170–176, 2015. https://doi.org/10.1136/annrheumdis-2013-204133 PMID: 24092415

79. Denny J. C., Bastarache L., Ritchie M. D., Carroll R. J., Zink R., Mosley J. D., et al., "Systematic comparison of phenome-wide association study of electronic medical record data and genome-wide association study data," Nature Biotechnology, vol. 31, p. 1102, 11/24/online 2013. https://doi.org/10.1038/nbt.2749 PMID: 24270849

80. Solus J. F., Chung C. P., Oeser A., Li C., Rho Y. H., Bradley K. M., et al., "Genetics of serum concentration of IL-6 and TNFα in systemic lupus erythematosus and rheumatoid arthritis: a candidate gene analysis," Clinical Rheumatology, vol. 34, pp. 1375–1382, August 01 2015. https://doi.org/10.1007/s10067-015-2881-6 PMID: 25652333

81. González-Recio O., de Maturana E. L., Vega A. T., Engelman C. D. and Broman K. W., "Detecting single-nucleotide polymorphism by single-nucleotide polymorphism interactions in rheumatoid arthritis using a two-step approach with machine learning and a Bayesian threshold least absolute shrinkage and selection operator (LASSO) model," BMC Proceedings, vol. 3, p. S63, December 15 2009. PMID: 20018057

82. Arya R., Hare E., del Rincon I., Jenkinson C. P., Duggirala R., Almasy L., et al., "Effects of covariates and interactions on a genome-wide association analysis of rheumatoid arthritis," BMC Proc, vol. 3 (Suppl 7):S84, 2009.

83. Scott I. C., Rijsdijk F., Walker J., Quist J., Spain S. L., Tan R., et al., "Do Genetic Susceptibility Variants Associate with Disease Severity in Early Active Rheumatoid Arthritis?," The Journal of Rheumatology, 2015.

84. Jiang X., Frisell T., Askling J., Karlson E. W., Klareskog L., Alfredsson L., et al., "To What Extent Is the Familial Risk of Rheumatoid Arthritis Explained by Established Rheumatoid Arthritis Risk Factors?," Arthritis & Rheumatology, vol. 67, pp. 352–362, 2015.

85. Deng F.-Y., Lei S.-F., Zhu H., Zhang Y.-H. and Zhang Z.-L., "Integrative Analyses for Functional Mechanisms Underlying Associations for Rheumatoid Arthritis," The Journal of Rheumatology, 2013.

86. Yarwood A., Viatte S., Okada Y., Plenge R., Yamamoto K., Barton A., et al., "Loci associated with N-glycosylation of human IgG are not associated with rheumatoid arthritis: a Mendelian randomisation study," Annals of the Rheumatic Diseases, vol. 75, pp. 317–320, 2016. https://doi.org/10.1136/annrheumdis-2014-207210 PMID: 26386125

87. Simmons C. R., Zou F., Younkin S. G. and Estus S., "Rheumatoid arthritis-associated polymorphisms are not protective against Alzheimer's disease," Molecular Neurodegeneration, vol. 6, p. 33, May 19 2011. https://doi.org/10.1186/1750-1326-6-33 PMID: 21595938

88. Xie G., Lu Y., Sun Y., Zhang S. S., Keystone E. C., Gregersen P. K., et al., "Identification of the NF-kappaB activating protein-like locus as a risk locus for rheumatoid arthritis," Ann Rheum Dis, vol. 72, pp. 1249–54, Jul 2013. https://doi.org/10.1136/annrheumdis-2012-202076 PMID: 23223422

89. Jalil S. F., Bhatti A., Demirci F. Y., Wang X., Ahmed I., Ahmed M., et al., "Replication of European Rheumatoid Arthritis Loci in a Pakistani Population," The Journal of Rheumatology, 2013.

90. Hinks A., Marion M. C., Cobb J., Comeau M. E., Sudman M., Ainsworth H. C., et al., "Brief Report: The Genetic Profile of Rheumatoid Factor–Positive Polyarticular Juvenile Idiopathic Arthritis Resembles That of Adult Rheumatoid Arthritis," Arthritis & Rheumatology, vol. 70, pp. 957–962, 2018.

91. Manning A. K., Ngwa J. S., Hendricks A. E., Liu C.-T., Johnson A. D., Dupuis J., et al., "Incorporating biological knowledge in the search for gene × gene interaction in genome-wide association studies," BMC Proceedings, vol. 3, p. S81, 2009/12/15 2009. PMID: 20018077

92. Das S., Baruah C., Saikia A. K. and Bose S., "Associative role of HLA-DRB1 SNP genotypes as risk factors for susceptibility and severity of rheumatoid arthritis: A North-east Indian population-based study," International Journal of Immunogenetics, vol. 45, pp. 1–7, 2018. https://doi.org/10.1111/iji.12347 PMID: 29168332

93. Stahl E. A., Raychaudhuri S., Remmers E. F., Xie G., Eyre S., Thomson B. P., et al., "Genome-wide association study meta-analysis identifies seven new rheumatoid arthritis risk loci," Nat Genet, vol. 42, pp. 508–14, Jun 2010. https://doi.org/10.1038/ng.582 PMID: 20453842

94. Viatte S., Flynn E., Lunt M., Barnes J., Singwe-Ngandeu M., Bas S., et al., "Investigation of Caucasian rheumatoid arthritis susceptibility loci in African patients with the same disease," Arthritis Res Ther, vol. 14, p. R239, 2012. https://doi.org/10.1186/ar4082 PMID: 23121884

95. Govind N., Choudhury A., Hodkinson B., Ickinger C., Frost J., Lee A., et al., "Immunochip identifies novel and replicates known, genetic risk loci for rheumatoid arthritis in black South Africans," Molecular medicine (Cambridge, Mass.), vol. 20, pp. 341–349, 2014.

96. Lee Y. H., Bae S.-C., Choi S. J., Ji J. D. and Song G. G., "Genome-wide pathway analysis of genome-wide association studies on systemic lupus erythematosus and rheumatoid arthritis," Molecular Biology Reports, vol. 39, pp. 10627–10635, December 01 2012. https://doi.org/10.1007/s11033-012-1952-x PMID: 23053960

97. Kochi Y., Yamada R., Kobayashi K., Takahashi A., Suzuki A., Sekine A., et al., "Analysis of single-nucleotide polymorphisms in Japanese rheumatoid arthritis patients shows additional susceptibility markers besides the classic shared epitope susceptibility sequences," Arthritis & Rheumatism, vol. 50, pp. 63–71, 2004.

98. Kirsten H., Al-Hasani H., Holdt L., Gross A., Beutner F., Krohn K., et al., "Dissecting the genetics of the human transcriptome identifies novel trait-related trans-eQTLs and corroborates the regulatory relevance of non-protein coding loci†," Human Molecular Genetics, vol. 24, pp. 4746–4763, 2015. https://doi.org/10.1093/hmg/ddv194 PMID: 26019233

99. Brorsson C. A., Onengut S., Chen W.-M., Wenzlau J., Yu L., Baker P., et al., "Novel association between immune-mediated susceptibility loci and persistent autoantibody positivity in type 1 diabetes," Diabetes, 2015.

100. Wolin A., Lahtela E. L., Anttila V., Petrek M., Grunewald J., van Moorsel C. H. M., et al., "SNP Variants in Major Histocompatibility Complex Are Associated with Sarcoidosis Susceptibility—A Joint Analysis in Four European Populations," Frontiers in Immunology, vol. 8, 2017-April-19 2017.

101. Herráez D. L., Martínez-Bueno M., Riba L., Torre I. G., Sacnún M., Goñi M., et al., "Rheumatoid Arthritis in Latin Americans Enriched for Amerindian Ancestry Is Associated With Loci in Chromosomes 1, 12 and 13 and the HLA Class II Region," Arthritis & Rheumatism, vol. 65, pp. 1457–1467, 2013.

102. Okada Y., Yamada R., Suzuki A., Kochi Y., Shimane K., Myouzen K., et al., "Contribution of a haplotype in the HLA region to anti–cyclic citrullinated peptide antibody positivity in rheumatoid arthritis, independently of HLA–DRB1," Arthritis & Rheumatism, vol. 60, pp. 3582–3590, 2009.

103. Wang K., Baldassano R., Zhang H., Qu H.-Q., Imielinski M., Kugathasan S., et al., "Comparative genetic analysis of inflammatory bowel disease and type 1 diabetes implicates multiple loci with opposite effects," Human Molecular Genetics, vol. 19, pp. 2059–2067, 2010. https://doi.org/10.1093/hmg/ddq078 PMID: 20176734

104. Liu C., Ackerman H. H. and Carulli J. P., "A genome-wide screen of gene–gene interactions for rheumatoid arthritis susceptibility," Human Genetics, vol. 129, pp. 473–485, May 01 2011. https://doi.org/10.1007/s00439-010-0943-z PMID: 21210282

105. Zhang L., Pan Q., Wang Y., Wu X. and Shi X., "Bayesian Network Construction and Genotype-Phenotype Inference Using GWAS Statistics," IEEE/ACM Transactions on Computational Biology and Bioinformatics, pp. 1–1, 2018. https://doi.org/10.1109/TCBB.2016.2599867

106. Eleftherohorinou H., Wright V., Hoggart C., Hartikainen A.-L., Jarvelin M.-R., Balding D., et al., "Pathway Analysis of GWAS Provides New Insights into Genetic Susceptibility to 3 Inflammatory Diseases," PLOS ONE, vol. 4, p. e8068, 2009. https://doi.org/10.1371/journal.pone.0008068 PMID: 19956648

107. Feng T. and Zhu X., "Genome-wide searching of rare genetic variants in WTCCC data," Human Genetics, vol. 128, pp. 269–280, September 01 2010. https://doi.org/10.1007/s00439-010-0849-9 PMID: 20549515

108. Schaub M. A., Kaplow I. M., Sirota M., Do C. B., Butte A. J. and Batzoglou S., "A Classifier-based approach to identify genetic similarities between diseases," Bioinformatics, vol. 25, pp. i21–i29, 2009. https://doi.org/10.1093/bioinformatics/btp226 PMID: 19477990

109. Wei W.-H., Viatte S., Merriman T. R., Barton A. and Worthington J., "Genotypic variability based association identifies novel non-additive loci DHCR7 and IRF4 in sero-negative rheumatoid arthritis," Scientific Reports, vol. 7, p. 5261, 2017/07/13 2017. https://doi.org/10.1038/s41598-017-05447-1 PMID: 28706201

110. Matzaraki V., Kumar V., Wijmenga C. and Zhernakova A., "The MHC locus and genetic susceptibility to autoimmune and infectious diseases," Genome Biology, vol. 18, p. 76, April 27 2017. https://doi.org/10.1186/s13059-017-1207-1 PMID: 28449694

111. Andreasi R. B., Khan M., Galuppi E., Govoni M. and Rubini M., "THU0022 Replication analysis of gene-gene interaction between HLA-DQA2 and HLA-DQB2 variants in italian rheumatoid arthritis patients," Annals of the Rheumatic Diseases, vol. 76, pp. 207–207, 2017.

112. Kim S. A. and Yoo Y. J., "Effects of single nucleotide polymorphism marker density on haplotype block partition," Genomics & Informatics, vol. 14, pp. 196–204, 2016.

113. Daly M. J., Rioux J. D., Schaffner S. F., Hudson T. J. and Lander E. S., "High-resolution haplotype structure in the human genome," Nat Genet, vol. 29, pp. 229–32, Oct 2001. https://doi.org/10.1038/ng1001-229 PMID: 11586305

114. Kimmel G. and Shamir R., "A block-free hidden Markov model for genotypes and its application to disease association," J Comput Biol, vol. 12, pp. 1243–60, Dec 2005. https://doi.org/10.1089/cmb.2005.12.1243 PMID: 16379532

115. Zhang K., Deng M., Chen T., Waterman M. S. and Sun F., "A dynamic programming algorithm for haplotype block partitioning," Proc Natl Acad Sci U S A, vol. 99, pp. 7335–9, May 28 2002. https://doi.org/10.1073/pnas.102186799 PMID: 12032283

116. Zhang K., Qin Z. S., Liu J. S., Chen T., Waterman M. S. and Sun F., "Haplotype block partitioning and tag SNP selection using genotype data and their applications to association studies," Genome Res, vol. 14, pp. 908–16, May 2004. https://doi.org/10.1101/gr.1837404 PMID: 15078859

117. Katanforoush A., Sadeghi M., Pezeshk H. and Elahi E., "Global haplotype partitioning for maximal associated SNP pairs," BMC Bioinformatics, vol. 10, p. 269, 2009. https://doi.org/10.1186/1471-2105-10-269 PMID: 19712447

118. Zahiri J., Mahdevar G., Nowzari-Dalini A., Ahrabian H. and Sadeghi M., "A novel efficient dynamic programming algorithm for haplotype block partitioning," J Theor Biol, vol. 267, pp. 164–70, Nov 21 2010. https://doi.org/10.1016/j.jtbi.2010.08.019 PMID: 20728452

119. Chen W.-P., Hung C.-L. and Lin Y.-L., "Efficient haplotype block partitioning and tag SNP selection algorithms under various constraints," BioMed Research International, vol. 2013, 2013.

120. Pugach I., Matveyev R., Wollstein A., Kayser M. and Stoneking M., "Dating the age of admixture via wavelet transform analysis of genome-wide data," Genome Biol, vol. 12, p. R19, 2011. https://doi.org/10.1186/gb-2011-12-2-r19 PMID: 21352535

121. Patil N., Berno A. J., Hinds D. A., Barrett W. A., Doshi J. M., Hacker C. R., et al., "Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21," Science, vol. 294, pp. 1719–23, Nov 23 2001. https://doi.org/10.1126/science.1065573 PMID: 11721056

122. Anderson E. C. and Novembre J., "Finding haplotype block boundaries by using the minimum-description-length principle," Am J Hum Genet, vol. 73, pp. 336–54, Aug 2003. https://doi.org/10.1086/377106 PMID: 12858289

123. Koivisto M., Perola M., Varilo T., Hennah W., Ekelund J., Lukk M., et al., "An MDL method for finding haplotype blocks and for estimating the strength of haplotype block boundaries," Pac Symp Biocomput, vol. 8, pp. 502–13, 2003.

124. Pattaro C., Ruczinski I., Fallin D. M. and Parmigiani G., "Haplotype block partitioning as a tool for dimensionality reduction in SNP association studies," BMC Genomics, vol. 9, p. 405, 2008. https://doi.org/10.1186/1471-2164-9-405 PMID: 18759977

125. Ionita-Laza I., Lee S., Makarov V., Buxbaum J. D. and Lin X., "Sequence kernel association tests for the combined effect of rare and common variants," Am J Hum Genet, vol. 92, pp. 841–53, Jun 6 2013. https://doi.org/10.1016/j.ajhg.2013.04.015 PMID: 23684009

126. Su S. C., Kuo C. C. and Chen T., "Inference of missing SNPs and information quantity measurements for haplotype blocks," Bioinformatics, vol. 21, pp. 2001–7, May 1 2005. https://doi.org/10.1093/bioinformatics/bti261 PMID: 15699029