Immunome Research
ISSN: 1745-7580

Castiblanco and Anaya, Immunome Res 2017, 13:3
DOI: 10.4172/1745-7580.1000136

Research Article Open Access

Homozygosity Analysis in Autoimmunity Affected Individuals and Multiplex Autoimmune Disease Families

John Castiblanco[1,2]* and Juan-Manuel Anaya[1]

[1]*Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24 #63-C-69, Bogota, Colombia*
[2]*Doctoral Program in Biomedical Sciences, Universidad del Rosario, Bogotá, Colombia*

Abstract

Autoimmune diseases (AD) are responsible for a substantial amount of disability and morbidity worldwide. Research generally focuses on a single disease, although autoimmune phenotypes could represent pleiotropic outcomes of non-specific disease genes underlying similar immunogenetic mechanisms. This report examined the effect and importance of the homozygosity status, using genome-wide interspersed markers, in individuals and multiplex families affected with AD. This study presented two approaches: (I) a case-control comparison and evaluation on the effect of homozygosity at the genome-wide level and per marker, including 453 unrelated individuals (121 late-, 79 early-onset AD, 40 polyautoimmunity (PolyA), 30 multiple autoimmune syndrome (MAS) and 183 healthy control individuals); and (II) a model-free affected pair linkage approach which included 35 MAS, 49 polyA, 104 late-, and 83 early-onset multiplex families. A total of 372 genome-wide markers were used in the analysis. The standardized observed homozygosity (S_{OH}) was calculated and the association of the homozygosity status and the autoimmune trait was evaluated. The multipoint model-free linkage analysis was applied by using RELPAL from S.A.G.E v6.3. Results for the S_{OH} showed significant differences between controls and early-onset individuals, where early-onset affected individuals showed lower homozygosity relative to controls. No differences were observed relative to controls for MAS, polyA and late-onset disease at the genome-wide level. The local marker homozygosity effect showed share and specific risk and/or protective effects for 24 markers. The model-free affected pair linkage approach lacked any suggestive linkage signals, but marginal signals displayed excess allele sharing for extreme phenotypes in autoimmunity. This study presumed autoimmunity as a trait rather than a clinical phenotype and tried to approach AD as a continuous trait presenting extreme phenotypes. Future approaches would be expected to dwell on the data presented here to corroborate and expand on sample size, marker coverage and their effects.

Keywords: Autoimmunity; Autoimmune disease; Familial autoimmunity; Homozygosity; Polyautoimmunity; Multiple autoimmune syndrome; Late-onset; Early-onset

Introduction

Human genetic markers reflect the differences in DNA sequence within the genome of individuals within populations. These markers can take many forms, including single nucleotide variants (i.e., SNP-Single nucleotide polymorphisms), short tandem repeats (STRs) (i.e., microsatellites and/or variable number of tandem repeats), small indels (i.e., insertions and deletions of a short DNA sequence) and duplications or deletions that change the copy number of a larger segment [1]. STRs have been the workhorse of human genetic analysis since the late 1980s. Their polymorphism is due to variations in the number of tandem repeats of short sequence units typically ranging from two to four nucleotides in size [2].

STRs are highly prone to mutations due to their susceptibility to slippage events during DNA replication, have been linked to at least 40 monogenic disorders [3], and are suggested to contribute to an array of complex traits [4]. Furthermore, STR variations convey high information content due to their rapid mutation and multi-allelic spectra, making this type of variants key for population genetics pilot and or proof of principle studies, when applied in a wide-range of methods to find signatures of selection, to elucidate mutation patterns in nearby SNPs, in DNA forensics and in genetic genealogy [5,6].

Heterozygosity is often used as proxy for homozygosity. Previous reports have studied the relationship between individual genetic diversity and fitness using heterozygosity–fitness correlations (HFCs). STRs have been the most commonly used markers to investigate HFCs. Heterozygosity and homozygosity estimation would help to shed light on underlying mechanisms, and provide tools for further population-based studies in humans [7]. Two primary mechanisms have been suggested to explain HFCs [8,9]. First, homozygous individuals may be more susceptible to disease because they are inbred and a second mechanism that may generate HFCs involves chance linkage between one or more of the markers and gene(s) experiencing balancing selection. Balancing selection has often been thought to be rather rare, particularly in humans where the classical example-sickle cell anemia - remains one of very few examples. Moreover, while some argue that polymorphisms at immune function genes are maintained by over dominant balancing selection, there is evidence that this is unlikely to be effective at maintaining more than two alleles. Regardless of theory, a number of recent HFC studies report-convincing associations between heterozygosity at particular loci. A correlation between inbreeding and blood pressure has been reported for isolate populations from Croatia [10]. There is also evidence suggesting homozygosity is an important risk factor in susceptibility to infectious diseases in humans, such as tuberculosis in Gambia, Leprosy in India and Hepatitis B infections [11].

***Corresponding author:** John Castiblanco, Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24#63-C-69, Bogota, Colombia, E-mail: jokaki@icloud.com

Received June 11, 2016; **Accepted** June 28, 2017; **Published** July 07, 2017

Citation: Castiblanco J, Anaya JM (2017) Homozygosity Analysis in Autoimmunity Affected Individuals and Multiplex Autoimmune Disease Families. Immunome Res 13: 136. doi: 10.4172/17457580.1000136

Similarly, the modified use of traditional linkage approaches remains a useful tool for the study of polygenic diseases. In some cases, genetic loci overlap or co-localize between related disorders. Becker et al., first reported based on previous autoimmune disease (AD) linkage studies, 18 common non-major histocompatibility complex (MHC) loci clusters and hypothesized a shared and common genetic basis for autoimmunity as a trait [12]. Other studies of linkage for specific-diseases (i.e., single disease approach) have found shared autoimmunity loci [13-15]. Limitations of genome-wide scans when applied to complex ADs, involve heterogeneity in disease phenotypes, population and ethnic differences and unavailable statistical and analytical models [13].

Numerous genetic factors are established to be important contributors to susceptibility in developing ADs based on several findings including the examination of the concordance rates between relatives for many autoimmune diseases [16]. A variety of pathogenic mechanisms are ultimately triggered during the progression of ADs and dysregulation involving major cell signaling pathways and inflammatory responses are consistent features in most ADs [17,18]. However, due to their multifactorial and polygenic nature, accompanied by a differential penetrance influenced by environmental factors and genetic heterogeneity among populations [19,20], untangling of the genetic determinants defining their outcome and onset has proven to be extremely challenging. Data showing the existence of different ADs within a single family or within the same individual, suggest a combination of genetic defects that may predispose individuals to different ADs sharing common pathogenic pathways [21]. This report examined the effect and importance of the homozygosity status, using genome-wide interspersed markers, in individuals and multiplex families affected with AD using a panel of microsatellites by applying a case-control approach and a model-free multipoint linkage affected relative pair approach.

Materials and Methods

Study population and family collection

All patients were treated and invited to participate at the Center for Autoimmune Diseases Research (CREA) at the University of Rosario in Bogotá and Medellín, Colombia. Individuals included in this study presented with: (i) at least one AD according to specific validated criteria. For analysis purposes, type 1 diabetes (T1D) cases were categorized as individuals with early-onset AD while any other affected AD individual was categorized as late-onset AD; and (ii) polyautoimmunity (polyA) (i.e., co-occurrence of distinct ADs within an individual) and/or multiple autoimmune syndrome (MAS) (i.e., co-occurrence of three or more distinct ADs within an individual). Healthy controls, matched by age, sex, ethnicity and socioeconomic status, were selected from women attending the same clinic, who met a similar age (± 5 years) and criteria for eligibility as the cases with no evidence of AD (Table 1). The distribution of AD among these four categories is as follows - Early-onset AD included 79 T1D individuals, Late-onset AD included systemic lupus erythematosus (SLE) (n=21), rheumatoid arthritis (RA) (n=23), Sjögren's syndrome (SS) (n=45), autoimmune thyroid disease (AITD) (n=27) and other AD (n=5) individuals. There were 70 individuals affected with 2 AD (i.e., PolyA) and 30 with more than 3 AD (i.e., MAS). Moreover, multiplex families consisted of varying size were ascertained through patients treated at the CREA in Medellín and Bogotá at the University of Rosario in Bogotá and Medellin, Colombia (Table 2). In each recruited family: (i) the proband presented with at least one AD according to validated criteria; (ii) presented evidence of familial autoimmunity (i.e., different ADs within members of a nuclear family), (iii) and each affected presented well-defined autoimmune phenotype (i.e., fulfillment of international classification criteria in probands and first-degree relatives [FDR]). Families in which the proband was affected with T1D were included and used as early-onset AD families (Table 2). FDR were defined as parents and siblings.

Patients with AD, polyA and MAS fulfilled validated classification criteria and were part of a multicenter cohort followed at the CREA. Information on demographics and cumulative clinical manifestations over the course of disease were obtained by both chart review and discussion with the patient and was collected in a standard data collection form, following the methodology described by Priori et al. [22], using a standardized questionnaire that incorporates demographics and medical information including a check-point list of 18 ADs [23]. In order to avoid ascertainment bias, the diagnosis of any AD was only considered reliable and consequently registered if made by a certified physician (i.e., internist, endocrinologist, or rheumatologist) and confirmed by chart review or verification during discussion with the relative.

All Patients fulfilled the diagnostic classification criteria proposed per disease as previously applied [23,24]. All T1D affected cases were children all of whom fulfilled the diagnostic classification criteria proposed by the American Diabetes Association (ADA) [25], as has been previously described (Table 1) [26]. For affected individuals with thyroid disorders, anti-thyroglobulin and anti-thyroperoxidase antibodies were measured by enzyme-linked immunosorbent assay (QUANTA Lite, INOVA Diagnostics, San Diego, CA, USA). Only patients with positive antibody profile for autoimmune thyroid disease (AITD) were included for analysis. Exclusion criteria were preexisting

Autoimmune Trait	Age ± Std Dev (Min, Max)	Age of Onset (Min, Max)	Female (%): Male (%)	Total (n=453)
Early-onset AD	15.61 ± 8.19 [4, 41]**	7.94 ± 5.22 [1, 24]**	34 (43): 45 (57)**	79
Late-onset AD	50.51 ± 15.73 [13, 85]	37.79 ± 14.54 [10, 74]	114 (94): 7 (6)	121
MAS	42.27 ± 14.1 [16, 71]	32.25 ± 13.14 [5, 59]	29 (97): 1 (3)	30
PolyA	47.70 ± 15.81 [16, 78]	35.63 ± 13.77 [5, 67]	68 (97):2 (3)	70
Controls	47.92 ± 16.42 [22, 85]	-	180 (98): 3 (2)	183

Table 1: Characteristics of AD affected and healthy control study individuals;AD: Autoimmune disease; PolyA: polyautoimmunity; MAS: Multiple autoimmune syndrome. Data corresponds to Colombian unrelated affected or unaffected and taken into account for the analysis. Number of PolyA individuals included in analysis includes MAS individuals. Late-onset AD included systemic lupus erythematosus (SLE) (n=21), rheumatoid arthritis (RA) (n=23), Sjögren's syndrome (SS) (n=45), autoimmune thyroid disease (AITD) (n=27) and other AD (n=5) individuals. Early-onset AD included 79 type 1 diabetes affected individuals;**p-value<0.001 two-tailed t-test when comparing Late-onset vs. Early-onset variables.

Characteristic	Late-onset	PolyAD	MAS	Early-onset
Age (yrs) [Min, Max]	45.99 [13,83]	45.49 [16,78]	44.81 [20,64]	19.54 [4,70]**
Male [Aff (Unaff)]	8 (109)	6 (53)	5 (33)	49 (132)
Female [Aff (Unaff)]	130 (174)	71 (86)	48 (59)	47 (149)
No. of Peds	104	49	35	82
Mean Size ± SD [Min, Max]	4.05 ± 2.21 [3,17]	4.41 ± 2.72 [3,17]	4.14 ± 2.92 [3,17]	4.60 ± 2.08 [3,13]
Pairs of relatives [a]				
Parent/Offspring	27/162/121	20/88/52	11/51/31	8/191/151
Sibling/Sibling	25/194/201	20/129/118	12/86/81	5/117/88
Sister/Sister	17/73/121	14/56/74	9/44/51	1/29/23
Brother/Brother	0/27/6	0/15/5	0/9/5	2/27/22
Brother/Sister	8/94/74	6/58/39	3/30/25	2/61/43
Grandparent	1/6/5	1/3/4	0/3/3	2/30/22
Avuncular	8/21/23	8/12/20	5/17/12	0/48/34
Cousin	1/1/2	1/1/2	1/1/2	0/0/0

Table 2: Characteristics of probands and families included in the analysis; AD: Autoimmune disease; PolyA: poly autoimmunity; MAS: Multiple autoimmune syndrome; Data correspond to relatives affected or unaffected taken into account for the analysis; Aff: Affected; Unaff: Unaffected; **p-value<0.001 t-test when comparing Late-onset *vs*. Early-onset variables. [a] Affected/Unaffected/Discordant pairs.

hematological diseases and hepatitis B virus, hepatitis C virus, or human immunodeficiency virus infections. This research is being carried out in accordance with Resolution No 008430 of 1993 issued by the Ministry of Health of the Republic of Colombia and was classified as a minimal risk research. The Ethics Committee of the Universidad del Rosario approved the present project.

Statistical and genetic data analysis

Data was managed and stored using the R software v3.1.2 [27] and Excel spreadsheets. Results are presented as means ± standard deviation (SD), minimum/maximum and/or in percentages. Comparison between means was performed by the Student´s t-test and those between percentages by the χ^2 test and two–sided Fisher's exact test, where appropriate or unless stated otherwise. A p–value of less than 0.05 was considered as statistically significant.

The present study included information on (i) sex, (ii) autoimmunity affection status defined as affected, unaffected or unknown for AD (i.e., having at least one AD), polyautoimmunity (i.e., having at least two ADs) and MAS (i.e., having three or more ADs); (iii) family/pedigree relationships. Estimation of the distributions of relationship types and affection status among relatives pairs were examined using the Statistical Analysis for Genetic Epidemiology (S.A.G.E.) program PEDINFO, release v6.3 [28]. Where necessary, dummy individuals were added to families for the purpose of connecting relatives within pedigrees, and the affection status for such dummy individuals was set to missing and thus they were not used in the analyses.

Genetic marker characterization and homozygosity analysis

Genomic DNA from affected patients and relatives was extracted from 10 mL of an EDTA-anticoagulated blood sample using the classical salting out protocol. Genetic markers included in this study were autosomal microsatellites genotyped using Screening Set 16 diversity panel at the NHLBI sponsored Mammalian Genotyping Service, Marshfield, Wisconsin.

Individuals with less that 10% of missing genotypes were excluded from analysis. Descriptive gene diversity parameters, allelic richness, observed (H_o), and expected (H_e) heterozygosity and the polymorphic information content (PIC) were calculated at each locus and over all loci using PopGene and PopGeneKit R packages. When necessary, file conversions were performed using PGDSpider v2.0.7.4 [29]. Incidence of genotyping errors was examined to screen the data for null allele frequency estimators using the F_{st} Refined Estimation by Excluding Null Alleles (ENA) - FreeNA software [30]. The Standardized Observed Homozygosity (S_{OH}) for an individual genotyped for *i* loci was calculated as the ratio of the number of homozygote genotypes (N_{Hom}) observed in *i-th* individual and the sum of the frequency for the observed homozygotes in the *i* locus (H_{oi}) scored per individual across the full sample set (i.e., $S_{OH}=N_{Hom}/\Sigma H_{oi}$) [11]. The S_{OH} measures to which extend and individual presents a greater or lesser homozygosity relative to the homozygosity level expected per maker if all genotypes were randomly ascertained. This allowed to compare the global homozygosity per AD trait category and per marker.

Familial data cleaning and multipoint model-free linkage analysis

Affected relative pair methods were used to identify genetic linkage. Familial data was checked and corrected for Mendelian inconsistent genotypes and relationship errors by using the RELTEST and MARKERINFO programs in S.A.G.E program, v6.3 [28]. Allele frequency estimates were obtained by using the program FREQ in S.A.G.E by maximum likelihood estimates of the allele frequencies among the founders of the families using all genotyped family members.

Genotypes from all pairs of relatives were used to estimate the proportion of alleles shared identical by descent (IBD) using GENIBD from S.A.G.E v6.3, by calculating the likelihood of each inheritance vector at multiple vectors to generate IBD distributions at spacing of 2 cm. Multipoint model-free linkage was performed using the regression-based model-free two-level Haseman–Elston linkage analyses using RELPAL from S.A.G.E v6.3, which models trait data from relative pairs as a function of marker allele sharing IBD. All individuals were used at the first level and all pairs of relatives used at the second level linkage analysis. Empirical p-values were estimated with up to 1,000,000 permutations. Empirical P-values in the range of 1×10^{-3} to 5×10^{-4} (i.e., $-\log_{10}$ P-values ≤ 3.00 to 3.30), were presumed as suggestive linkage, as suggested by the Lander and Kruglyak criterion for studies involving a mixture of relative pairs [31].

Results

Homozygosity and susceptibility to autoimmunity as a trait

Clinical characteristics and demographics of the case- control autoimmunity samples: This study included 453 genotyped unrelated individuals (121 late-onset AD, 79 early-onset, 40 PolyA, 30 MAS and 183 healthy control individuals) from Medellin, Colombia South America (Table 1). Control individuals were comprised of 183 matched by age, sex, ethnicity, and socioeconomic status. A general description of the Colombian samples included is disclosed on Table 1. When age and age of onset were compared between early-onset and late-onset individuals, the difference was statistically significant (P-value<0.001), as expected given their autoimmune disorder characteristics (Table 1). Late-onset individuals presented 6% males and 94% females while early-onset presented 57% males and 43% females. Females represented the most affected in late-onset families while in early-onset the ratio of affected was close to 1:1 (Male: Female). The entire group of Colombian individuals belonged to a population from the Northwestern part of Colombia, South America (i.e., Paisa community). This population was established in the 16th-17th centuries and flourished in relative isolation until the late 19th century [32,33].

A total of 453 samples and controls, and 372 polymorphic markers were analyzed, giving a total of about 168,516 genotypes. All markers were highly informative (PIC ≥ 0.50) with a low null allele frequency, making them optimal and reliable for genetic diversity studies (Figure 1). Moreover, average allelic richness observed per locus for the markers was 4.30 ± 1.22. The average observed (H_o) and expected (H_e) heterozygosity were 0.69 ± 0.16 and 0.68 ± 0.13, respectively.

Assessment of homozygosity as a surrogate of heterozygosity could generate more disadvantages than advantages (i.e., null alleles, allele dropout, shutter bands, miscalling). In order to avoid these mishaps, the S_{OH} was used. S_{OH} measures the expected homozygosity from the allele frequencies and the observed homozygosity per locus. Thus, S_{OH} measures to which extent an individual presents a greater or lesser homozygosity relative to the homozygosity level expected if all genotypes were randomly ascertained.

After calculation of the S_{OH} per individual and for the whole set of 453 samples, the distribution of the S_{OH} values (i.e., the genome-wide Standardized observed homozygosity) was compared by using Wilcoxon-rank sum test to examine if there were significant differences between AD groups relative to healthy control individuals (Figure 2).

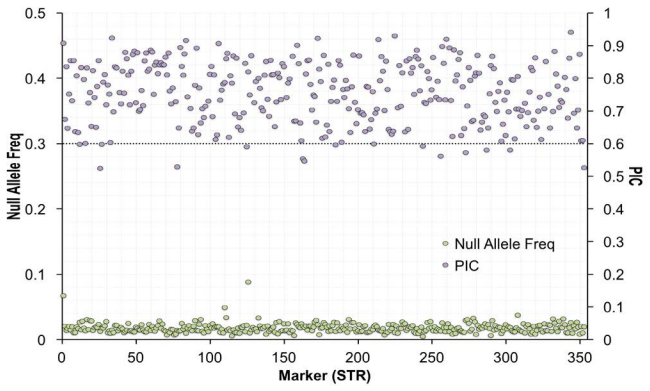

Figure 1: Marker reliability, gene diversity and population relationship indicators; A: Calculated null allele frequency (left axis) and polymorphic information content (PIC) (Right axis).

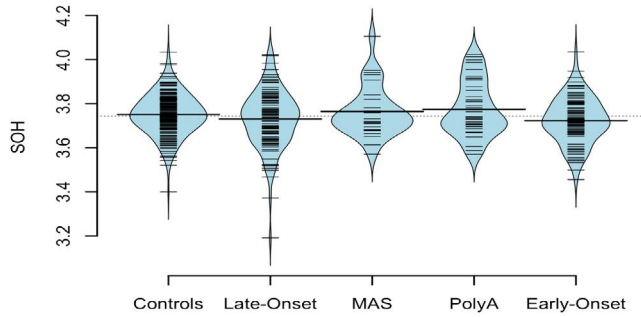

Figure 2: Beanplots for the calculated standardized observed homozygosity in Colombian individuals affected and unaffected with autoimmune diseases; Each bean consists of a mirrored density curve containing one- dimensional scatterplot of the data; Individual data points are represented as short lines, a solid line shows the average per each group and the dashed line represents the overall average; This plot was generated using the beanplot package from R; Comparisons relative to controls performed by a two-tailed Wilcoxon rank sum test; When Early-onset was compared with controls, p-value=0.02. No other comparison was significant.

Comparisons did not reveal statistical significant differences (p-value ≤ 0.05) when late-onset AD, MAS and polyA affected individuals were compared with controls, while early-onset AD individuals showed a statistically significant deviation towards reduced homozygosity in cases relative to controls (p-value=0.02) (Figure 2).

Subsequently, the local homozygosity effect for the genotyped markers (i.e., the observed homozygosity per marker) was examined. The odds ratio (OR) for each marker was calculated at each locus per each AD trait (Table 3). Most of the markers showed a non-significant association between homozygosity and susceptibility/protection to either present or not early-, late-onset, MAS and/or polyA. Instead, 24 markers presented p-values less than 0.05; however, when correction for multiple comparisons was applied, these significant values became suggestive or marginally significant (Table 3). Threshold of statistical significance was established at 0.00013 (0.05/372) conservatively applying the standard Bonferroni correction procedure at the 0.05 level.

The ORs observed for the suggestive or marginal effects were diverse. Twelve out of 24 markers showed a higher risk/susceptibility to acquire/develop AD traits (i.e., D1S1677, D3S2432, D3S2418, D4S2632, GATA104, D8S1477, D8S1136, D10S1208, ATA5A09N, AGAT113Z, D15S1507, ATA41E04), while the other 12 showed a protective effect (i.e., D2S2944, AAAT072, D7S821, D8S1110, D8S2324, D8S1132, D11S4459, D12S1045, D14S617, D15S643, AAT269, D21S1440) (Table 3). More important, two markers were shared between late-onset, MAS and polyA showing the same directional effect (i.e., D2S2944, D10S1208). Other marker was shared between polyA and MAS (i.e., D1S1677) (Table 3). Although the spacing between markers, approximately on average 10cM, is sufficient to ensure they behave as if unlinked, it is possible that multiple markers contribute to the same risk through linkage to related genes.

Familial data and multipoint model-free linkage analysis: The affected relative approach examined 35 MAS, 49 polyA, 104 late-onset, and 83 early-onset multiplex families (Table 2). The mean pedigree size, standard deviation as well as the total number of relative pairs included in the analysis are depicted on table 2. When early-onset and late-onset families age and age of onset were compared, the difference was statistically significant (P-value<0.001) as expected given their autoimmune disorder characteristics.

Chr	Band GRCh37	Locus	Early-onset	Late-onset	PolyA	MAS
1	1q23.3	D1S1677	>0.05	>0.05	2.23 (1.15 -4.33) 0.013	3.12 (1.23 - 7.93) 0.009
2	2q34	D2S2944	>0.05	0.38 (0.17 - 0.79) 0.006	0.29 (0.1 - 0.73) 0.004	0.21 (0.02 - 0.91) 0.031
3	3p22.3	D3S2432	>0.05	2.06 (1.17 - 3.63) 0.009	>0.05	>0.05
3	3q28	D3S2418	>0.05	2.08 (1.17 - 3.73) 0.009	>0.05	>0.05
4	4p15-p14	D4S2632	>0.05	2.38 (1.18 - 4.79) 0.01	>0.05	>0.05
5	5q35.3	AAAT072	0.38 (0.18 - 0.73) 0.002	>0.05	>0.05	>0.05
7	7q21.3	D7S821	>0.05	>0.05	0.29 (0.07 -0.85) 0.016	>0.05
7	7q21.3	GATA104	>0.05	2.28 (1.28 - 4.05) 0.003	>0.05	>0.05
8	8p12	D8S1477	>0.05	>0.05	>0.05	3.24 (1.18 -8.61) 0.013
8	8q11.23	D8S1110	0.4 (0.16 - 0.89) 0.018	>0.05	>0.05	>0.05
8	8q13.1	D8S1136	>0.05	>0.05	2.49 (1.23-5.03) 0.008	>0.05
8	8q21.11	D8S2324	0.36 (0.16 - 0.73) 0.002	>0.05	>0.05	>0.05
8	8q23	D8S1132	>0.05	>0.05	0.12 (0 - 0.8) 0.019	>0.05
10	10p11.21	D10S1208	>0.05	1.91 (1.05 - 3.49) 0.027	2.02 (1.02 -3.93) 0.03	2.61 (0.99 - 6.75) 0.047
11	11q12.1	D11S4459	0.45 (0.23 - 0.84) 0.008	>0.05	>0.05	>0.05
12	8q24	D12S1045	0.42 (0.18 - 0.88) 0.017	>0.05	>0.05	>0.05
13	13q12	ATA5A09N	2.06 (1.11 - 3.84) 0.015	>0.05	>0.05	>0.05
13	13q33.3	AGAT113Z	>0.05	2 (1.1 - 3.64) 0.018	>0.05	>0.05
14	14q32.12	D14S617	>0.05	0.37 (0.15 - 0.83) 0.012	>0.05	>0.05
15	15q22.2	D15S643	>0.05	0.35 (0.12 - 0.84) 0.013	>0.05	>0.05
15	15q22.31	D15S1507	2.2 (1.25 - 3.89) 0.004	>0.05	>0.05	>0.05
16	16p13.3	ATA41E04	>0.05	2.16 (1.15 - 4.04) 0.013	>0.05	>0.05
20	20q13.13	AAT269	>0.05	>0.05	0.27 (0.08 -0.72) 0.004	>0.05
21	q22.13	D21S1440	>0.05	>0.05	>0.05	0.08 (0 -0.54) 0.002

Table 3: Short tandem repeats showing the strongest association between homozygosity and early-, late-onset, polyA and MAS. Odd ratios are presented with their correspondent 95% CI and p-value.

Affected discordant and concordant pair results for the non-parametric multipoint linkage analyses implemented in RELPAL for early-, late-onset, polyA and MAS are shown on Figures 3A-3D, respectively. Linkage was modeled without including any covariates. Markers were within 10 cM approximately of each other. The Lander and Kruglyak criterion for suggestive linkage for studies involving a mixture of relative pairs [31] was used to verify the linkage signals obtained (i.e., P-values in the range of 1×10^{-3} to 5×10^{-4} [i.e., $-\log_{10}$ P-value=3.00 to 3.30]. Results did not show any suggestive linkage for early-, late-onset, polyA and/or MAS. However, putative linkage signals were observed in early-onset and MAS families (Table 4). D1S518 and D8S1128 for early-onset and TTTA002 for MAS families are markers that displayed excess allele sharing in concordantly affected and unaffected relative pairs.

Discussion

The commonality between AD is the damage to tissues and organs arising from the loss of tolerance and in most cases a gender imbalance [34]. Research generally focuses on a single disease, although autoimmune phenotypes could represent pleiotropic outcomes of non-specific disease genes underlying similar immunogenetic mechanisms [35]. While it is apparent that multiple cases of a single disease cluster within families [36], more striking are the individuals in those families afflicted with multiple ADs [13].

A common origin for diverse ADs is sustained by three levels of evidence [13]. The first comes from clinical observations indicating the possible shift from one disease to another or to the fact that more than one AD may coexist in a single patient (i.e., polyA, MAS) [24,36-39], or

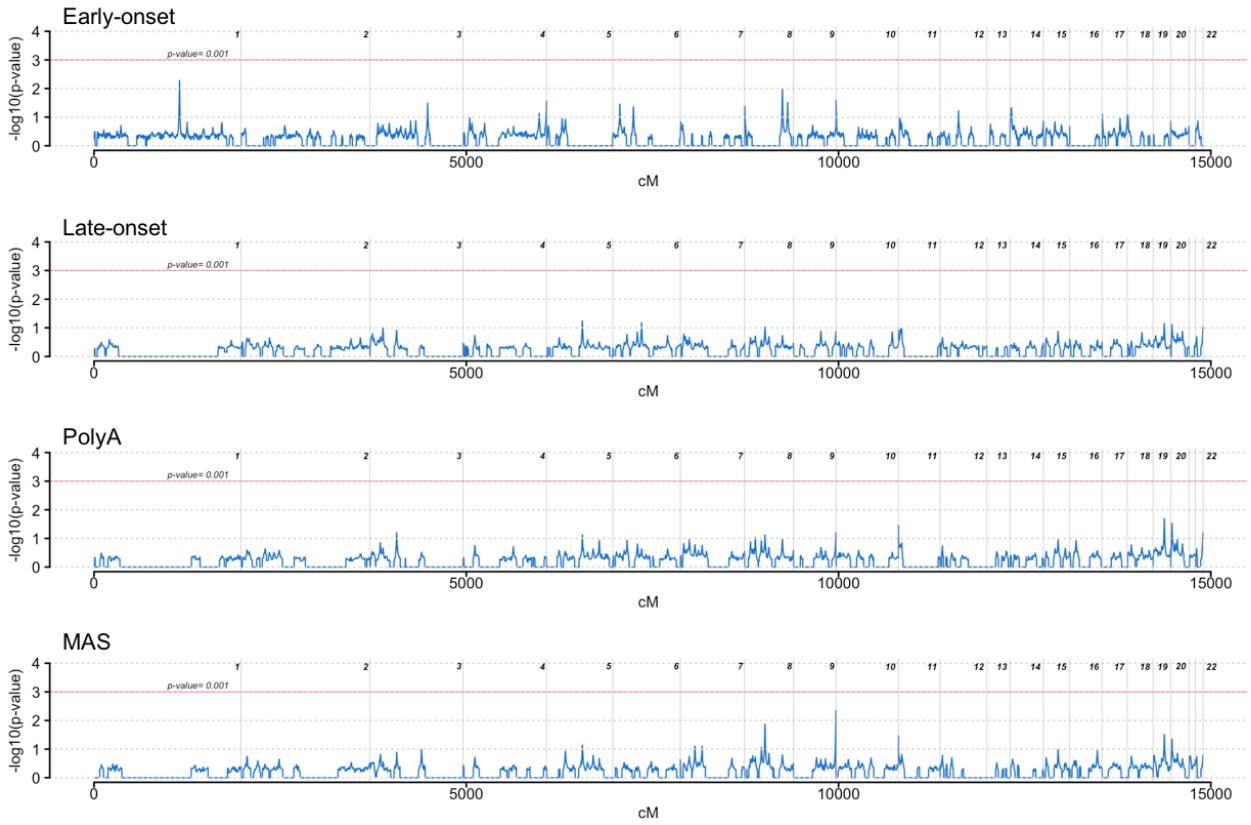

Figure 3: Model-free multipoint linkage by RELPAL using all affected relative pairs and the IBD variance; P-values evaluated on the basis of up to 1,000,000 permutations; Red line represents the lower threshold for suggestive linkage significance (p-value<0.001).

Marker	Trait	Chr.	p-value		
			Nominal	Empirical	$-\log_{10}$
D1S518	Early-Onset	1q31.1	0.000670	0.005	2.3
D8S1128	Early-Onset	8q22.1	0.000001	0.010	2.0
TTTTA002	MAS	9q34.3	0.000100	0.004	2.4

Table 4: Chromosome regions with the highest RELPAL $-\log_{10}$ p-value estimates.

in the same family (i.e., familial autoimmunity) [40]. The second level of evidence refers to known shared pathophysiological mechanisms between ADs [41]; and the third level of evidence corresponds to the evidence implying common genetic factors [38]. The importance of this concept focuses on the probability of having multiple ADs simultaneously in one patient, which goes beyond epidemiologic inferences. This study sought to consider autoimmune clinical phenotypes as traits to laid out the commonality and their complexity by including extreme phenotypes (i.e., Early-onset and MAS affected individuals and their families) and traits that might reside within their interim as a phenotype (i.e., Late-onset and PolyA affected and their families).

Over the last decade, association studies examining the genetic basis of human disease have switched overwhelmingly from STR markers to SNPs. SNPs are much less polymorphic than microsatellites, a deficiency that is usually compensated for by the vastly greater number of markers being genotyped. However, while there are many advantages to using SNPs for the assessment of local heterozygosity, microsatellites offer an arguably more direct approach that circumvents the need to reconstruct complex haplotypes [11]. Several authors contend that SNPs may be more suitable than microsatellites for HFCs. Extensive simulation studies have examined the effect of different mutational patterns (corresponding to SNPs and microsatellites) and demographic history on the expected correlation between heterozygosity and fitness. Their results point to a complex interplay between these two factors. The high mutation rate of microsatellites should make them more suitable to detect HFCs that result from recent inbreeding due to crosses between relatives or to a small population size [7].

This report presented two approaches: a case-control comparison and evaluation on the effect of the homozygosity status at the genome-wide level and per marker, and a model-free affected concordant/discordant pair linkage approach to identify IBD loci. The first was a systematic analysis where the status of the individuals per locus and at the genome level was taken into account, standardized, and evaluated for the association between homozygosity and AD. The results for S_{OH} at the genome-wide level showed significant differences between controls and early-onset individuals, where early-onset affected individuals showed lower homozygosity relative to controls. No differences were observed relative to controls for MAS, polyA and late-onset disease, which would support the clinical late-onset nature of these entities (Figure 2).

Detailed analysis for the local homozygosity status effect showed about a 1:1 relation on elevated risk and/or protective effects conferred by the homozygosity status at specific loci depending on the compared autoimmune trait (Table 3). On top of this, some of the markers

suggested a shared component between traits and more interestingly between late-onset, polyA and MAS but not with early-onset as observed with the S_{OH} genome-wide level results. Since correction for multiple comparisons only provided suggestive and marginal significance values, no candidate regions or genes were put forward; no less, data provides evidence of shared effect regions between late-onset, MAS and polyA traits and a plausible baseline for future approaches with broader marker coverage, and larger sample sizes.

Homozygosity has been previously examined on a single disease basis for rheumatoid arthritis [42]. This type of approach provides an alternative to allelic association mapping for the identification of recessive variants responsible related to ADs. It is suggested that the immune system genes would benefit from high diversity; a richer allele structure would indulge protection towards a pathogen/environment repertoire but could go counter current towards autoimmune phenomena. Thus, a "less is more" hypothesis could result in a limited repertoire response system towards external exposures but could be advantageous to promote stable and balanced autoimmune phenomena. The idea of the environment/exposure defining and driving a disease outcome is well accepted in autoimmunity (i.e., Autoimmune ecology) and it is starting to get the needed attention [43].

The second approach, which is a model-free affected pair linkage approach, lacked any suggestive linkage signals for early-, late-onset, polyA or MAS. However, putative/marginal signals displayed excess allele sharing for early-onset and MAS between affected relatives, both extreme phenotypes in autoimmunity. Although their signals warrant caution, these support a possible shared genetic predisposition; likewise, marker regions have been previously linked at a single disease level amid different reports [13].

Recently, 11 novel and rare functional variants were identified in AD and MAS cases for the *MACF1, KIAA0754, DUSP12, ICA1, CELA1, LRP1/STAT6, GRIN3B, ANKLE1, TMEM161A*, and *FKRP* genes [44]. On top of this, linkage analysis in multiplexed families affected by PolyA and familial autoimmunity showed significant linkage signals (LOD>3.0) in *SRA1, MLL4, ABCB8, DHX34* and *PLAUR* [45]. Network analysis and functional relatedness for previous autoimmunity associated genes affiliated *SRA1*, *PLAUR* and *ABCB8* as key players into regulation of apoptotic processes [45]. Moreover, LRP1, was functionally related to the *HLA-B* and *IL10* genes suggested a substantial impact within immunological pathways and/or reaction to bacterial and other foreign proteins. LRP1/STAT6 play key roles in maintaining the homeostasis of the immune system and are involved in extracellular and intracellular anti-inflammatory pathways [44]. However, due to their multifactorial, polygenic and oligogenic nature, accompanied by a differential penetrance influenced by environmental factors and genetic heterogeneity among populations [19,20], untangling of the genetic determinants defining their outcome and onset has proven to be extremely challenging. Likewise, data showing the existence of different AD within a single family or within the same individual, suggest a combination of genetic defects that may predispose individuals to different AD sharing common pathogenic pathways [13,21,44-46].

The present study was a pilot/exploratory approach, expected to serve as an initial proof of principle for the commonality of autoimmunity as a trait. Future approaches would be expected to dwell on the data presented here to broaden and expand on sample size, genetic marker type and coverage. Closer inspection of clinical and phenotypic quantitative variants is warranted, as well as inclusion of environmental and clinical available variants. The affected relative pairs approach was only possible, instead of a sibling pair due to the sample size and available concordant and discordant pairs. Limitations of genome-wide scans when applied to complex AD should involve heterogeneity in disease phenotypes, population and ethnic differences and unavailable statistical and analytical models.

Conclusions and Perspectives

Overall, this study assumed autoimmunity as a trait rather than a clinical phenotype and tried to approach AD as a continuous phenotype presented with extreme phenotypes (i.e., early-onset and MAS traits, respectively). On genome-wide homozygosity examination, results showed homozygosity differences relative to controls for early-onset individuals, while on local inspection several markers suggested homozygosity associated with protection/susceptibility to early-, late-onset, polyA and/or MAS.

Numerous genetic factors are established to be important contributors to susceptibility in developing ADs; on top of this genetic layer, environment/exposure would refine and tune towards either disease onset or tolerance. Usually association methods approach heterogeneity as the main cause of disease onset for ADs. This focus in part reflects the multifactorial and polygenic nature, accompanied by a differential penetrance influenced by environmental factors but does not reflect the recessive component of the puzzle. A common and rare component within the genetic landscape of the autoimmune trait should be expected, thus extreme phenotypes should bring to the table new clues and information that might serve and correlate towards the more homogenous component of the trait. This rare component has started to surface with approaches such as whole exome sequencing in individuals affected with polyA, MAS and multiplex autoimmune families [44,45].

Acknowledgments

We would like to thank all patients and their families who participated in this study. Also, we would like to thank Dr. Ruben Dario Mantilla and Dr. Adriana Rojas-Villarraga for their support and clinical expertise input. This work was supported by Universidad del Rosario and Colciencias (122254531722), Bogota, Colombia. Some of the results of this paper were obtained by using the software package S.A.G.E., which was supported by a U.S. Public Health Service Resource Grant (RR03655) from the National Center for Research Resources.

References

1. Frazer KA, Murray SS, Schork NJ, Topol EJ (2009) Human genetic variation and its contribution to complex traits. Nat Rev Genet 10: 241-251.
2. Elahi E, Kumm J, Ronaghi M (2004) Global genetic analysis. J Biochem Mol Biol 37: 11-27.
3. Mirkin SM (2007) Expandable DNA repeats and human disease. Nature 447: 932-940.
4. Gemayel R, Vinces MD, Legendre M, Verstrepen KJ (2010) Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annual review of genetics 44: 445-477.
5. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G,et al. (2001) Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science 293: 455-462.
6. Khan R, Mittelman D (2013) Rumors of the death of consumer genomics are greatly exaggerated. Genome Biology 14: 139.
7. Miller JM, Malenfant RM, David P, Davis CS, Poissant J, et al. (2014) Estimating genome-wide heterozygosity: effects of demographic history and marker type. Heredity 112: 240-247.
8. Pemberton J (2004) Measuring inbreeding depression in the wild: the old ways are the best. Trends Ecol Evol 19: 613-615.
9. Hansson B, Westerberg L (2002) On the correlation between heterozygosity and fitness in natural populations. Molecular ecology 11: 2467-2474.
10. Rudan I, Smolej-Narancic N, Campbell H, Carothers A, Wright A, et al. (2003)

Inbreeding and the genetic complexity of human hypertension. Genetics 163: 1011-1021.

11. Lyons EJ, Amos W, Berkley JA, Mwangi I, Shafi M, et al. (2009) Homozygosity and risk of childhood death due to invasive bacterial disease. BMC Med Genet 10: 55.
12. Becker KG, Simon RM, Bailey-Wilson JE, Freidlin B, Biddison WE, et al. (1998) Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. Proc Natl Acad Sci U S A 95: 9979-9984.
13. Anaya JM, Gomez L, Castiblanco J (2006) Is there a common genetic basis for autoimmune diseases? Clin Dev Immunol 13: 185-195.
14. Becker KG (2004) The common variants/multiple disease hypothesis of common complex genetic disorders. Med Hypotheses 62: 309-317.
15. Tomer Y, Ban Y, Concepcion E, Barbesino G, Villanueva R, et al. (2003) Common and unique susceptibility loci in Graves and Hashimoto diseases: results of whole-genome screening in a data set of 102 multiplex families. Am J Hum Genet 73: 736-747.
16. Cardenas-Roldan J, Rojas-Villarraga A, Anaya JM (2013) How do autoimmune diseases cluster in families? A systematic review and meta-analysis. BMC medicine 11: 73.
17. Baranzini SE (2009) The genetics of autoimmune diseases: a networked perspective. Current opinion in immunology 21: 596-605.
18. Zhernakova A, van Diemen CC, Wijmenga C (2009) Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nature reviews Genetics 10: 43-55.
19. Bias WB, Reveille JD, Beaty TH, Meyers DA, Arnett FC (1986) Evidence that autoimmunity in man is a Mendelian dominant trait. Am J Human Genet 39: 584-602.
20. Wandstrat A, Wakeland E (2001) The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol 2: 802-809.
21. Invernizzi P (2009) Future directions in genetic for autoimmune diseases. J Autoimmun 33: 1-2.
22. Priori R, Medda E, Conti F, Cassara EA, Danieli MG, et al. (2003) Familial autoimmunity as a risk factor for systemic lupus erythematosus and vice versa: a case-control study. Lupus 12: 735-740.
23. Anaya JM, Tobon GJ, Vega P, Castiblanco J (2006) Autoimmune disease aggregation in families with primary Sjogren's syndrome. J Rheumatol 33: 2227-2234.
24. Rojas Villarraga A, Amaya Amaya J, Rodriguez Rodriguez A, Mantilla RD, Anaya JM (2012) Introducing polyautoimmunity: secondary autoimmune diseases no longer exist. Autoimmune Dis 2012: 254319.
25. The Expert Committee on the Diagnosis and Classification of Diabetes Mellitus (2003) Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care 26: S5-20.
26. Tobon GJ, Arango A, Abad V, Garcia J, Cuervo H, et al. (2006) Clinical and immunological characteristics of type 1 diabetes mellitus in a northwestern Colombian population. Diabetes Res Clin Pract 72: 170-175.
27. Team RC (2014) R: A Language and Environment for Statistical Computing. In R: A Language and Environment for Statistical Computing.
28. S.A.G.E. (2012) Statistical Analysis for Genetic Epidemiology, Release 6: 3.
29. Lischer HE, Excoffier L (2012) PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics 28: 298-299.
30. Chapuis MP, Estoup A (2007) Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol 24: 621-631.
31. Lander E, Kruglyak L (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet 11: 241-247.
32. Carvajal Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, et al. (2000) Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. Am J Hum Genet 67: 1287-1295.
33. Bravo ML, Valenzuela CY, Arcos Burgos OM (1996) Polymorphisms and phyletic relationships of the Paisa community from Antioquia (Colombia). Gene Geogr 10: 11-17.
34. Mariani SM (2004). Genes and autoimmune diseases - a complex inheritance. MedGenMed 6: 18.
35. Vyse TJ, Todd JA (1996) Genetic analysis of autoimmune disease. Cell 85: 311-318.
36. Anaya JM, Corena R, Castiblanco J, Rojas Villarraga A, Shoenfeld Y (2007) The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity. Exp Rev Clin Immunol 3: 623-635.
37. Blank M,Gershwin ME (2008) Autoimmunity: from the mosaic to the kaleidoscope. J Autoimmun 30: 1-4.
38. Anaya JM, Castiblanco J, RojasVillarraga A, Pineda Tamayo R, Levy RA, et al. (2012) The Multiple Autoimmune Syndromes. A Clue for the Autoimmune Tautology. Clin Rev Allergy Immunol 43: 256-264.
39. Perez Fernandez OM, Mantilla RD, Cruz Tapias P, Rodriguez Rodriguez A, Rojas Villarraga A, et al. (2012) Spondyloarthropathies in autoimmune diseases and vice versa. Autoimmune Dis 2012: 736384.
40. Cardenas Roldan J, Rojas Villarraga A, Anaya JM (2013) How do autoimmune diseases cluster in families? A systematic review and meta-analysis. BMC Med 11: 73.
41. Sloka S (2002) Observations on recent studies showing increased co-occurrence of autoimmune diseases. J Autoimmun 18: 251-257.
42. Yang HC, Chang LC, Liang YJ, Lin CH, Wang PL (2012) A genome-wide homozygosity association study identifies runs of homozygosity associated with rheumatoid arthritis in the human major histocompatibility complex. PloS One 7: e34840.
43. Anaya JM, Ramirez Santana C, Alzate MA, Molano Gonzalez N, Rojas Villarraga A (2016) The Autoimmune Ecology. Front Immunol 7: 139.
44. Johar AS, Mastronardi C, Rojas Villarraga A, Patel HR, Chuah A, et al. (2015) Novel and rare functional genomic variants in multiple autoimmune syndrome and Sjogren's syndrome. J Translat Med 13: 173.
45. Johar A, Sarmiento Monroy JC, Rojas Villarraga A, Silva Lara MF, Patel HR, et al. (2016) Definition of mutations in polyautoimmunity. J Autoimmun 72: 65-72.
46. Johar AS, Anaya JM, Andrews D, Patel HR, Field M, et al. (2015) Candidate gene discovery in autoimmunity by using extreme phenotypes, next generation sequencing and whole exome capture. Autoimmu Rev 14: 204-209.