Genetic Epidemiology of Polyautoimmunity and Common Autoimmunity in Colombia — Proof of Principle

By

John Enrique Castiblanco Quinche

This dissertation is presented
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in Biomedical Sciences

Universidad del Rosario
Faculty of Sciences and Mathematics
Faculty of Medicine
Biomedical Doctoral School
Center for Autoimmune Disease Research (CREA)

Bogota, Colombia
October 2016

Genetic Epidemiology of Polyautoimmunity and Common Autoimmunity in Colombia — Proof of Principle

By

John Enrique Castiblanco Quinche

Mentor/Director:

Prof. Dr. Juan Manuel Anaya-Cabrera

Committee/Jurors:

Prof. Dr. Mauricio Arcos-Burgos

Prof. Dr. Carlos Martín Restrepo Fernandez

Prof. Gabriel Bedoya Berrio

This dissertation is presented
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy in Biomedical and
Biological Sciences

Bogota, Colombia
October 2016

To my Caludia (“Claudia”)
for her unconditional love and support

Acknowledgements

I would like to acknowledge all those who have supported me during my research.

My deepest gratitude to my mentor (“my Profe”) Dr. Juan-Manuel Anaya for initiating me in immunogenetic research and showing me the way with his advice, example and guidance. Dr. Anaya steered my project while giving me enough opportunities to develop my own ideas. All his teachings have become and will shape the rest for my life at the personal and scientific level.

I am also especially grateful to Dr. Mauricio Arcos-Burgos, Dr. Carlos Martín Restrepo and Prof. Gabriel Bedoya. Their insights and constructive criticism helped me a lot in the development of this project. All my appreciation to Dr. Luisa Marina Mateus Merino, her guidance and constant support would not had allowed me to complete my program from the distance.

I would like to extend my appreciation to Dr. Jim Weber from the Marshfield Genotyping service, and Dr. Sunil Ahuja from the Center of Personalized Medicine, for letting me be part of his lab when I came to the United States. The time spent in their labs marked me professionally and personally.

To all lab colleagues and friends whom helped me, shared with and lived research to the fullest at the Center for Autoimmune Disease Research (CREA) in Medellín and Bogotá, my best and complete friendship, love and wishes.

I am forever in debt with my parents, Jairo Castiblanco and Maria Teresa Quinche, and my sister, Diana Gatt, for their love, encouragement and support in the pursuit of my scientific career.

Last but not least, to my Caludia, also known as the wife, edition manager, graphic designer, and grammar advisor for Spanish and English among other tasks, my complete and limitless love and thankfulness for her continuous support and encouragement in the good and bad times.

Abstract

Background: 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. While it is apparent that multiple cases of a single disease cluster within families, more striking are the individuals in those families afflicted with multiple autoimmune diseases. This study explored the dynamics of familial aggregation and segregation in AD (i.e., having at least one AD), polyautoimmunity (polyA) (i.e., having at least two ADs) and multiple autoimmune syndrome (MAS) (i.e., having three or more ADs) patients. Moreover, this project examined the effect and importance of homozygosity and whether the ancestry component of Colombian affected individuals is associated with susceptibility/protection to develop an AD.

Methods: Familial aggregation was examined for first-degree relatives. Segregation analysis for a binary trait was implemented on 210 single ascertained multiplex AD families. Homozygosity was examined by two approaches: (I) a case – control comparison and evaluation on the effect of homozygosity at the genome-wide level, including 453 genotyped unrelated individuals (121 late-onset AD, 79 early-onset, 40 polyA, 30 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. The admixture effect was examined in all included Colombian affected and healthy individuals, as well as in individuals originated from reference populations, assuming three ancestral groups (k=3) (i.e., European, Amerindian and African). The ancestry component effect for the studied traits was compared and examined by logistic regression relative to controls. All individuals and families were treated and recruited at the Center for Autoimmune Diseases Research (CREA) from Medellin and Bogota, Colombia, South America.

Results: This project provided data supporting that polyA and MAS are not AD independent traits and that gender, age and age of onset represent factors that define and allow the study of the dynamics of the traits within the familial group. Also, segregation data provided evidence for the genetic component role in the etiology of AD in late-onset families, while for early-onset families and perhaps because of their the relatively familial young status, eluded a clear picture of autoimmunity segregation and aggregation. The data also 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. Moreover, ancestry and autoimmunity in Colombian samples showed how the autoimmune trait landscape is not a black and white scenario but rather a colorful mix of genetic and environmental factors. All markers analyzed were highly informative with a low null allele frequency making them optimal and reliable for genetic diversity studies.

Conclusions This study presumed autoimmunity as a trait rather than a clinical phenotype and tried to approach AD as a continuous trait presenting extreme phenotypes**.** Data suggested that AD are not independent traits and that gender, age and age of onset represent factors play a role and allow to study of the dynamic of the traits. Finally, a clinical defined individual AD, defined by symptoms and signs, might not be completely juxtaposed to the AD trait defined by environment and genetics, which makes even more difficult the task to define and untangle disease mechanisms.

Resumen

Introducción: Las enfermedades autoinmunes (EA) son responsables de una gran porción de discapacidad y morbilidad a nivel mundial. Generalmente, las investigaciones científicas se centran en una sola enfermedad, aunque los fenotipos autoinmunes podrían estar representados por efectos pleiotrópicos en genes no-específicos al presentar mecanismos inmunogenéticos similares. Múltiples casos de una sola enfermedad dentro de familia son evidentes, y aún más sorprendente son los individuos en aquellas familias que sufren de múltiples EAs. Este estudio exploró la dinámica de agregación familiar y la segregación en pacientes con EA, poliautoimmunidad (polyA) (presentar por lo menos dos AD) y el síndrome autoinmune múltiple (MAS) (presentar tres o más EA). Por otra parte, se examinó el efecto y la importancia de la homocigosis y la ancestría en individuos afectados colombianos con respecto a desarrollar una EA.

Métodos: La agregación y segregación familiar se examinó en familiares de primer grado para un rasgo binario en 210 familias afectadas por EA. La homocigosis se estudió en dos enfoques: (I) Comparación de casos y controles mediante la evaluación del efecto de la homocigosis al nivel de todo el genoma en 453 individuos no relacionados (121 EA tardía, 79 EA temprana, 40 polyA, 30 MAS y 183 individuos control); (II) por un estudio de ligamiento no-paramétrico en parientes afectados en 35 familias con MAS, 49 con polyA, 104 con EA tardía, y 83 con EA temprana. La ancestría se examinó en todos los individuos afectados y sanos colombianos, así como individuos originados a partir de poblaciones de referencia, suponiendo tres grupos ancestrales (k = 3) (i.e., europea, amerindia y africana). El efecto de ancestría para los rasgos estudiados se comparó y analizó mediante regresión logística con respecto a los controles. Todos los individuos y las familias fueron tratados e invitados a participar en el Centro de Investigación de Enfermedades Autoinmunes (CREA) en Medellín y Bogotá, Colombia.

Resultados: Este proyecto sugiere que las EA no son rasgos independientes y que el género, la edad y la edad de inicio representan factores que definen y permiten el estudio de la dinámica de los rasgos dentro del grupo familiar. Más allá, los datos de segregación proporcionaron soporte para el papel del componente genético en la etiología de las EAs en las familias de aparición tardía, mientras que para las familias de inicio temprano no se observó un papel claro, tal vez debido a la edad relativamente joven familiar.

Los datos también mostraron diferencias en la homocigosidad en relación con los controles para las personas de aparición temprana, mientras que en la inspección de varios marcadores locales se sugiere que la homocigosis se encuentra asociada con la protección/susceptibilidad para la EA temprana, de inicio tardío, polyA y MAS. Por otra parte, la ancestría y la autoinmunidad en muestras de colombianos mostraron cómo el paisaje rasgo autoinmune no es un escenario blanco y negro, sino más bien una colorida mezcla de factores genéticos y ambientales. Todos los marcadores analizados fueron muy informativos con una baja frecuencia del alelo nulo haciéndolos óptimos y fiables para estudios de diversidad genética.

Conclusiones: Este estudio asumió a la autoinmunidad como un rasgo más que un fenotipo clínico y como un rasgo continuo que presenta fenotipos extremos. Los datos sugieren que las EAs no son independientes. Por último, una EA individual, definida por síntomas y signos, podría no ser completamente yuxtapuesta con una EA definida por el medio ambiente y genética, lo que hace aún más difícil la tarea de definir y dilucidar los mecanismos de las enfermedades.

Abbreviations

AD - Autoimmune disease

ADA - American Diabetes Association

AITD - Autoimmune thyroid disease

AS - Ankylosing Spondylitis

CD - Celiac disease

CREA - Center for Autoimmune Diseases Research

FAI - familial autoimmunity

FAID - Familial autoimmune disease

FDR - First-degree relatives

GWAS – Genome-wide association study

HFC – Heterozygosity-fitness correlations

IBD - Identical by descent

MAS - Multiple autoimmune syndrome

MG - Myasthenia gravis

MHC - Major histocompatibility complex

MS - Multiple sclerosis

NIH - National Institute of Health

OR - Odds ratio

PASII - Polyglandular autoimmune syndrome II

PIC - Polymorphic information content

PolyA - Polyautoimmunity

PSO - Psoriasis

RA - Rheumatoid arthritis

SAGE - Statistical Analysis for Genetic Epidemiology

SLE - Systemic lupus erythematosus

SNP - Single nucleotide variants

SOH - Standardized observed homozygosity

SS - Sjögren's syndrome

SSc - Scleroderma

STRs - Short tandem repeats

T1D - Type 1 diabetes

VIT – Vitiligo

Scientific Production Linked with this Project

Published papers:

1. Anaya JM, Duarte-Rey C, Sarmiento-Monroy JC, Bardey D, **Castiblanco J**, Rojas-Villarraga A. Personalized medicine. Closing the gap between knowledge and clinical practice. Autoimmun Rev. 2016 Aug;15(8):833-42
2. **Castiblanco J** and Anaya JM. Genetics and Vaccines in the Era of Personalized Medicine. Current Genomics 2015; 16(1):47-59.
3. **Castiblanco J**, Arcos-Burgos M, Anaya JM. What is next after the genes for autoimmunity? BMC Med. 2013 Sep 4; 11: 197.
4. Anaya JM, **Castiblanco J**, Rojas-Villarraga A, Pineda-Tamayo R, Levy RA, Gómez-Puerta J, Dias C, Mantilla RD, Gallo JE, Cervera R, Shoenfeld Y, Arcos-Burgos M. The multiple autoimmune syndromes. A clue for the autoimmune tautology. Clin Rev Allergy Immunol. 2012 Dec; 43(3):256-64.
5. **Castiblanco J**, Anaya JM. The nature and nurture of common autoimmunity. Ann NY Acad Sci. 2007 Aug; 1109: 1-8.
6. Anaya JM, Corena R, **Castiblanco J**, Rojas-Villarraga A, Shoenfeld Y. The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity. Expert Rev Clin Immunol. 2007 Jul; 3 (4): 623-35.
7. Anaya JM, Gómez L, **Castiblanco J**. Is there a common genetic basis for autoimmune diseases? Clin Dev Immunol. 2006 Jun-Dec; 13 (2-4): 185-95.
8. Anaya JM, Delgado-Vega AM, **Castiblanco J**. Genetic basis of Sjögren's syndrome. How strong is the evidence? Clin Dev Immunol. 2006 Jun-Dec; 13 (2-4): 209-22.
9. Anaya JM, Tobon GJ, Vega P, **Castiblanco J**. Autoimmune disease aggregation in families with primary Sjögren's syndrome. J Rheumatol. 2006 Nov; 33 (11): 2227-34.
10. Anaya JM, **Castiblanco J**, Tobón GJ, García J, Abad V, Cuervo H, Velásquez A, Angel ID, Vega P, Arango A. Familial clustering of autoimmune diseases in patients with type 1 diabetes mellitus. J Autoimmun. 2006 May; 26 (3): 208-14.

Book Chapters:

- Cruz-Tapias P, **Castiblanco J**, Anaya JM. *Major histocompatibility complex: Antigen processing and presentation* in: Autoimmunity. From Bench to the Bedside. Edited by: Juan-Manuel Anaya Yehuda Shoenfeld Adriana Rojas-Villarraga Roger A. Levy Ricard Cervera. Bogota, Colombia. 2013. ISBN: 9789587383768.
- **Castiblanco J**, Arcos-Burgos M, Anaya JM. *Introduction to genetics of autoimmune diseases* in: Autoimmunity. From Bench to the Bedside. Edited by: Juan-Manuel Anaya Yehuda Shoenfeld Adriana Rojas-Villarraga Roger A. Levy Ricard Cervera. Bogota, Colombia. 2013. ISBN: 9789587383768.
- Cruz-Tapias P, **Castiblanco J**, Anaya JM. *HLA association with autoimmune diseases* in: Autoimmunity. From Bench to the Bedside. Edited by: Juan-Manuel Anaya Yehuda Shoenfeld Adriana Rojas-Villarraga Roger A. Levy Ricard Cervera. Bogota, Colombia. 2013. ISBN: 9789587383768.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Chapter 4

General Introduction

Autoimmune diseases (ADs) are responsible for a substantial amount of disability and morbidity worldwide. Although their epidemiology varies according to individual conditions, collectively, autoimmune prevalence is at least 5% in the general population and is one of the major causes of premature mortality in young and middle aged women (1).

ADs represent a diverse collection of diseases in terms of their demographic profile and primary clinical manifestations (2). The National Institutes of Health (NIH) estimates up to 23.5 million Americans suffer from AD with more than 80 known forms of disease and at least 40 more having an autoimmune basis making ADs the top ten leading causes of death in Americans. The commonality between ADs is the damage to tissues and organs arising from the loss of tolerance and in most cases a gender imbalance (2). Research generally focuses on a single disease, although autoimmune phenotypes could represent pleiotropic outcomes of non-specific disease genes underlying similar immunogenetic mechanisms (3). While it is apparent that multiple cases of a single disease cluster within families (4), more striking are the individuals in those families afflicted with multiple ADs (5).

Although ADs encompass a broad range of phenotypic manifestations and severity, their pathogenesis is considered to be multifactorial and several of their features suggest they share common etiologic factors (6). ADs can be categorized into two types of disorders: (1) systemic, such as systemic lupus erythematosus (SLE), in which the loss of immune tolerance is directed towards systemic antigens and disease manifestations can occur at a variety of different sites in the body; and (2) organ-specific in which predominantly or exclusively directed towards tissue-specific elements (e.g., type 1 diabetes (T1D) targets the pancreas, autoimmune thyroid disease (AITD) which attacks the thyroid gland). Most ADs are characterized by female predominance, and many are associated with the production of autoantibodies. These shared disease features, in conjunction with epidemiologic evidence that demonstrates the clustering of multiple ADs within individuals and families, strongly implicates shared etiologic factors that might include shared genetic loci.

Recent advances in genomics have led to increased understanding of the molecular underpinnings of disease. 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 (7). 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 (8, 9). However, due to their multifactorial and polygenic nature, accompanied by a differential penetrance influenced by environmental factors and genetic heterogeneity among populations (10, 11), untangling of the genetic determinants defining their outcome and onset has proven to be extremely challenging. Likewise, 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 (12).

Why Multiple Autoimmune Diseases Instead of One?

ADs represent a diverse collection of diseases in terms of their demographic profile and primary clinical manifestations. The phenotypic commonality between AD, however, is the damage to tissues and organs, which arises from the loss of tolerance, and that for most of them more women are affected than men (2). While it is apparent that multiple cases of a single disease cluster within families (4), even more striking are the numbers of individuals in those families afflicted with multiple different ADs (4, 13, 14). As heterogeneous diseases, ADs develop from the cumulative effect of diverse events on the immune system (15). Moreover, it is clear that ADs do not begin at the time of clinical appearance, but rather many years before.

The Commonality in Autoimmunity — A common origin for diverse ADs is sustained by three levels of evidence (5). 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., polyautoimmunity, polyA) (4, 16-19), or in the same family (i.e., familial autoimmunity) (Figure 1) (20). This level corresponds to the mosaic of clinical syndromes manifested in the form of co-occurrence of various ADs within an individual, or co-occurrence within members of a family. As support for the first level of clinical evidence, four entities have been described previously in the literature (Figure 1).

The **second level of evidence** refers to known shared pathophysiological mechanisms between ADs. Epidemiological studies have shown correlations among certain ADs, linking epidemiological observations to physiopathological evidence for autoimmune diseases might contribute to our knowledge for the shared etiological and immunogenetic mechanisms (15).

The **third level of evidence** corresponds to the evidence implying common genetic factors (18). The importance of this concept focuses on the probability of having multiple ADs simultaneously in one patient, which goes beyond epidemiologic inferences. Therefore, family history of ADs should be considered when performing genetic analysis as this new approach incorporates all accepted pathologies for which evidence suggests an autoimmune origin. Genetic approaches have postulated the possible scenarios in a defined population influencing the risk to develop an AD. Each population would have its own polymorphism repertoire, in which several would be common among populations and other would be characteristic of each one.

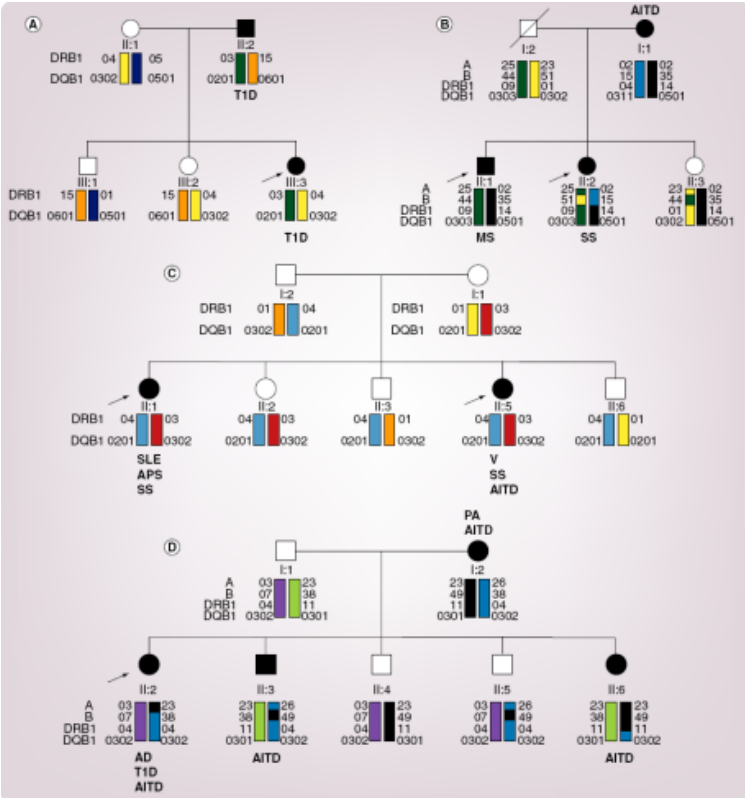


Figure 1. Types of familial autoimmunity. **A. Familial autoimmune disease (FAID)** is defined as the presence of one specific AD in various members of a nuclear family. The FAID importance stands behind the fact that the presence of the same AD in various generations of one family should require whether a stronger common genetic component compared to other syndromes or a common environmental risk factor or a combination of both, making these families more informative in terms of genetic epidemiology due to their homogeneity in the phenotype. Moreover, there is increased evidence for aggregation of AD in families of patients with a single AD (13, 21). Thus, the family history of AD should be considered when performing such genetic analysis. In this case a proband and a first degree relative (i.e. her father) have T1D. **B. Familial autoimmunity (FAI):** This phenomenon corresponds to the presence of different autoimmune diseases in a nuclear family. This definition uses the term "autoimmune disease" as a trait that encircles all accepted pathologies for which evidence suggests an autoimmune origin. **C. Multiple Autoimmune Syndrome (MAS)**: Corresponds to the presence of a least three ADs in a single individual. In this case two siblings have the same syndrome although different phenotypes. Also, this family represents a case of FAI. **MAS** was described by Humbert and Dupond in 1988 as a syndrome consisting of the presence of three or more ADs in one patient (22). The importance of this concept focuses on the probability of having three ADs simultaneously in one patient, goes beyond epidemiologic inferences or statistical chance. Thus, these previous notion arguments are in favor of common pathophysiological mechanisms giving origin to all three

diseases. **D. Polyglandular Autoimmune Syndrome, Type II (PASII).** Corresponds to the presence of Addison´s disease, autoimmune thyroid disease (AITD) and T1D. In this family, however, familial autoimmune disease and familial autoimmune coexist. PASII is also universally known as *Schmidt's Syndrome* (23). In 1964, Carpenter included in an extensive review of the literature the presence of T1D to the syndrome defining the classic triad for PASII (24). The diagnosis of PASII is defined by the presence of at least two ADs in one patient. There are three more types of PAS (24, 25). There some controversy around the definition of each type and some authors argument that PAS type II, III and IV are different manifestations of the same syndrome (25). This disagreement has strong foundations, for there are several reports showing association between T1D, AITD, CD, VIT, and AD. Herein, we will refer to the co-occurrence of distinct ADs within an individual by following the clinical evidence of which the autoimmune tautology highlights as **Polyautoimmunity** (19). Polyautoimmunity proposes the association of disorders, which encompasses the concept of a common origin for these diseases. Taken from Anaya et al (26).

ADs Genetic Epidemiology – Current status quo

Autoimmune disorder epidemiology varies according to individual conditions. Though, ADs encompass a broad range of phenotypic manifestations and severity, their pathogenesis is considered to be multifactorial and several of their features suggest they share common etiologic factors. These shared disease features, in conjunction with epidemiologic evidence that demonstrates the clustering of multiple ADs within individuals and families, strongly implicate shared etiologic factors including shared genetic loci. Reasons for the diverse manifestations exhibited by different ADs remain unclear, but recent progress in elucidating genetic susceptibility loci for this group of disorders promises to shed light on this important issue (27).

ADs are multifactorial in nature with susceptibility controlled by multiple genetic and environmental factors and they develop from the cumulative effect of diverse events on the immune system (15) (Figure 2). Diverse populations present different allelic structure, depending on the population natural and epidemiological history (28). In addition, the effects of genotype on phenotype in any given population may depend upon environment and length of exposure to an undefined etiological insult. Consequently, there is a need to explore genetic associations in diverse populations.

Genetic studies have revealed important conclusions on the genetic architecture of ADs. A plausible disease trait scenario has been projected as multiple environmental variants interacting with several genes to conferred susceptibility among individuals over a population (29). Accordingly with this theory, individuals would express the disease trait if they were located among the wrong side of the normal distribution. This concept has been widely

accepted, but additivity must be assumed for all genetic variants accompanied by an equal effect among the trait. Autoimmunity might involve a genetic distribution not as straightforward as a normal distribution but an unknown one in which many loci would not add but complement the individuals risk to develop the autoimmune phenotype.

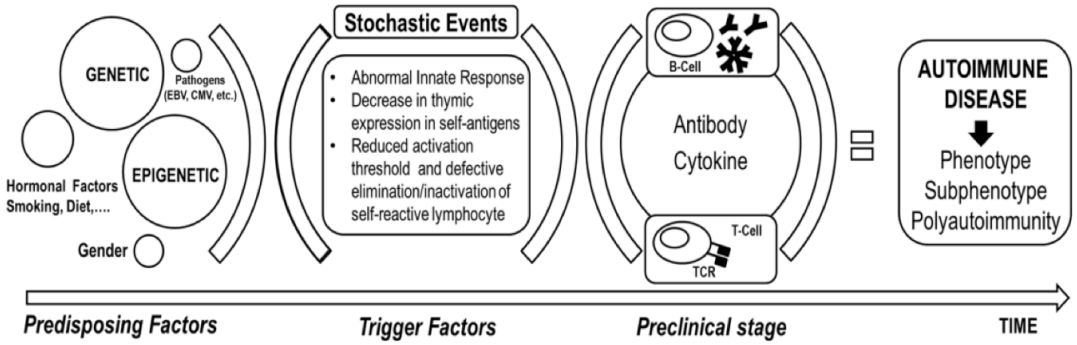


Figure 2. Outline showing the plausible stages for a multifactorial heterogeneous etiology to develop over time. Each stage shows the known phenomena that cumulative will be the causative scenario for the onset of autoimmune disease (s). As predisposing factors impact over the life of the individuals, they also converge and interact to create and increase or decrease the liability an individual would have to develop the phenotype. The former is accompanied by the physiological phenomena caused or derived from the functional and biological configuration of the subject which in turn will lead to disease markers, signs and symptoms needed to be taken into account to be prognosed, diagnosed and in the best case scenario prevented.

For many perhaps most traits, the interaction with other genes and environmental factors might be genetically programmed or may be purely stochastic. Most common diseases probably contain major subsets that fall into this sort of causation, even when the phenotypic manifestation is usually considered as being monogenic which could be the result of gene-environment interaction (30).

Main Factors Affecting the Genetic Epidemiology of ADs

Genetic evidence now suggests that genetic susceptibility to common disease is probably due to hundreds or even thousands of alleles that may or may not be rare, ancestral derived, and/or can vary in frequency among human populations. Likewise, evidence for **epistatic interaction** among risk alleles and their role in ADs has started to gain more focus (31).

It is important to distinguish between the clinical sense of familial clustering (extended families that happen to have multiple cases of a disease or syndrome of interest) and the epidemiological sense of familial aggregation (there is, on average, a greater frequency of disease in close relatives of individuals with the disease than in relatives of individuals without

the disease). Analyses of **familial aggregation** treat the family like any other unit of clustering. In addressing whether there is phenotypic aggregation within families, no attempt is made to determine the cause of any aggregation (32). Nevertheless, the observation and portrayal of familial autoimmunity and the outline of the MAS has put aside the environmental aggregation and given a greater value towards the common/rare genetic component for diverse autoimmune phenotypes with a generally common background (4).

Families with multiple affected relatives appear to share common risk alleles with sporadic patients, but may have a higher genetic load. A consequence of the polygenic model for complex diseases is that patients are inevitably highly heterogeneous in terms of the particular set of risk alleles they carry. It has been suggested that this may translate in different genetically determined disease mechanisms in subgroups of patients or a common disease mechanism that is complemented by additional pathways that are more or less predominant in different subgroups (33). Familial approaches have documented the clustering of certain ADs among the relatives of individuals who have RA, MS, SLE, T1D and other diseases (13, 21, 34-38).

Age has been proposed as a responsible factor for the onset of ADs at midlife (age, 40-60). The problem with age in epidemiologic studies dwells in the fact that many ADs have different ages of onset. For children for example, the most common diseases are T1D, CD and VIT (39). For young adults, MS, myasthenia gravis (MG), VIT, and SLE are the most frequent (40). Mid-age patients are more likely to have Sjögren's syndrome (SS) (41), scleroderma (SSc) (42), and RA (43). Finally, people at older ages are more prone to have SS (41), AITD (44), and MG (45). This mosaic of ages constitutes one of the biggest problems in aggregation and co-occurrence studies (15).

The vast majority of patients affected with ADs are **females**; the reason for this prevalence is poorly understood. The proportion of females with AD varies depending on the disease; from 18:1 in AITD to 1:1 in PSO to 1:2 in ankylosing spondylitis (AS). Even for specific populations, reports have described differences in gender and clinical presentation pertaining sex and RA (46). ADs, more prevalent in men, are characterized by acute inflammation, appearance of autoantibodies and a pro-inflammatory Th1 immune response whereas ADs more prevalent in women have known antibody-mediated pathologies. Moreover, ADs more prevalent in women that appear clinically in women past age 50 years are associated with chronic Th2-mediated

pathology (47). The more frequent the AD and the later it appears, the more women are affected.

Sex hormones have been considered to be candidates responsible for susceptibility to AD through modulation of Th1/Th2 response. Impact of hormonal changes on the disease course in females is documented in pregnancy: severity of MS and RA has been reported to decrease during pregnancy whereas severity in SLE is either exacerbated or unaffected during pregnancy. High levels of hormones during pregnancy, which enhance Th2 response suppressing RA and MS driven by a Th1 response, may explain this. In contrast, SLE is Th2 driven and might not be suppressed by the hormones (48). Theoretically, X-chromosome inactivation and the resultant tissue chimerism might explain the female predisposition to systemic autoimmunity (49, 50).

A preferential **inheritance** of the autoimmunity trait from mothers has been observed in patients with primary SS (13, 51), SLE (52) and T1D (53), indicating a preferential transmission of susceptibility genes from mothers to their offspring. Maternal transmission of autoimmunity could be influenced by the high preponderance of ADs in females compared to the general population given their greater intrinsic susceptibility to develop these diseases that can potentially arise from sex related physiological factors (54). Nonetheless, other reports have postulated a preferential transmission of the autoimmune trait through the father, more specifically T1D (55) and MS (56). Given an inherently excess of susceptibility of women to develop an AD, men would require an augmented risk to overcome the resistance towards autoimmunity relatively to women. Thus, men would need a greater content of susceptibility variants that would trigger their phenotype and also would guarantee more often transmission of the autoimmunity trait to his offspring. This has been previously reported for MS by Kantarci *et al.* (56), under the Carter effect, defined by the observed higher incidence of the trait in relatives when the index case is the least commonly affected sex (57).

ADs are not **inherited** in a simple, classical Mendelian way, but have instead a complex or a yet unknown mode of inheritance (10, 58). Bias *et al.* were the first to consider a single major gene conferring susceptibility for autoimmunity, and suggested an autosomal dominant inheritance pattern with penetrance of approximately 92% in females and 9% in males (10). As well, Arcos-Burgos *et al.* showed the presence of a dominant major gene and strong environmental effects as the most parsimonious model of segregation for VIT (59). On the other hand, when analyzing

RA together with other ADs a mixed model fitted the data significantly better than the major gene or polygenic models (60).

Finally, it is well recognized that **geography and ethnicity** can affect the incidence and prevalence of autoimmune diseases. Both factors are thought to reflect etiological heterogeneity between environmental, genetic factors, and their interactions; compelling examples of this effect have been reported for ADs such as RA (61, 62), primary SS (61, 63), SSc (64) and SLE (65).

Research Plan

This project was portrait in order to serve as a proof of principle in a genetic epidemiology framework for AD affected individuals, their relatives and healthy controls recruited from the same background population. Each aim was proposed to support information for common autoimmunity and polyA, reported generally in independent studies using individuals not always from the same background population.

- **General Approach and State of the Art (Chapter 1)**

  Manuscript I - Castiblanco John, Arcos-Burgos Mauricio, Anaya Juan-Manuel. What is next after the genes for autoimmunity? BMC Med. 2013. Review. PMID: 24107170

- **Specific Aims**
  - **Aim 1 — Familial Aggregation and Segregation (Chapter 2)**

    Manuscript II – Castiblanco John, Sarmiento-Monroy Juan-Camilo, Mantilla Ruben-Dario, Rojas-Villarraga Adriana, Anaya Juan-Manuel. Familial Aggregation and Segregation Analysis in Families Presenting Autoimmunity, Polyautoimmunity, and Multiple Autoimmune Syndrome. J Immunol Res. 2015; 2015: 572353. Epub 2015 Nov 30. PMID: 26697508
  - **Aim 2 — Homozygosity mapping (Chapter 3)**

    Manuscript III - Castiblanco John, Mantilla Ruben-Dario , Rojas-Villarraga Adriana and Juan-Manuel Anaya. Homozygosity Genetic Analysis in Autoimmunity Affected Individuals and Multiplex Autoimmune Disease Families. Immunome Res 13: 136. doi: 10.4172/17457580.1000136.

  - **Aim 3 — Population structure and Admixture Mapping (Chapter 4)**

    Manuscript IV - Castiblanco John, Mantilla Ruben-Dario , Rojas-Villarraga Adriana and Juan-Manuel Anaya. Ancestry effect in Colombian Individuals presenting Autoimmunity, Polyautoimmunity and Multiple Autoimmune Syndrome. (*Manuscript in submission*)

	Aim 1	Aims 2 and 3
	Familial Aggregation and Segregation Analysis	**Homozygosity, Population Structure and Admixture Mapping**
FOCUS	Phenotype Autoimmunity, Polyautoimmunity and MAS. Epidemiology	Genome-wide Microsatellite Screening
INNOVATION	Autoimmune disease as a collection of syndromes throughout families in a Specific Population.	Effect of Homozygosity and Admixture at a population level
SYNERGY	Familial Clustering and Segregation modeling in Uni- and Poly-disease families	Uni- and Poly-disease individuals and families. Ancestry differential
STUDY SAMPLE	Uni-, Poly-disease and non-affected relatives of Autoimmune disease and Control Families	Familial and Polyautoimmune, Cases and Healthy Individuals
LEVEL OF KNOWLEDGE	Inheritance, Segregation and aggregation of familial and Polyautoimmune traits	Homozygosity, Admixture and Population Structure

Chapter 1 - What is next after the genes for autoimmunity?

Castiblanco John, Arcos-Burgos Mauricio, Anaya Juan-Manuel. What is next after the genes for autoimmunity? BMC Med. 2013. Review. PMID: 24107170

Castiblanco et al. BMC Medicine 2013, 11:197
http://www.biomedcentral.com/1741-7015/11/197

Cutting edge: issues in autoimmunity

BMC Medicine

MINIREVIEW

Open Access

What is next after the genes for autoimmunity?

John Castiblanco[1,2,3,4], Mauricio Arcos-Burgos[5] and Juan-Manuel Anaya[1*]

Abstract

Clinical pathologies draw us to envisage disease as either an independent entity or a diverse set of traits governed by common physiopathological mechanisms, prompted by environmental assaults throughout life. Autoimmune diseases are not an exception, given they represent a diverse collection of diseases in terms of their demographic profile and primary clinical manifestations. Although they are pleiotropic outcomes of non-specific disease genes underlying similar immunogenetic mechanisms, research generally focuses on a single disease. Drastic technologic advances are leading research to organize clinical genomic multidisciplinary approaches to decipher the nature of human biological systems. Once the currently costly omic-based technologies become universally accessible, the way will be paved for a cleaner picture to risk quantification, prevention, prognosis and diagnosis, allowing us to clearly define better phenotypes always ensuring the integrity of the individuals studied. However, making accurate predictions for most autoimmune diseases is an ambitious challenge, since the understanding of these pathologies is far from complete. Herein, some pitfalls and challenges of the genetics of autoimmune diseases are reviewed, and an approximation to the future of research in this field is presented.

Keywords: Autoimmunity, Common, Genetics, Genomics, Personalized, Predictive medicine, Polyautoimmunity, Translational medicine

Introduction

The everlasting vision of a predictive and preventive framework for disease assessment has pushed the medical sciences to search for new means to manage health care and translate basic research into clinical practice. However, as we dig deeper into the cell and disease mechanisms, the path is not always clear because each new achievement and tool leads to more intricate definitions and targets [1]. Likewise, the cost and configuration of health care plans do not take into consideration the move towards personalized medicine, due in part to the lack of interaction between basic and clinical research. Advances in technology are now prompting this interaction, preparing for more realistic bench to bedside implementation [1-3].

The lack of pathognomonic diagnostic tools and clear-cut diagnostic criteria for complex conditions exposes patients to a bureaucratic limbo, stuck in the system in search of an accurate and complete diagnosis to receive appropriate treatment. Clinical pathologies lead us to consider disease as either an independent entity or a diverse set of traits governed by common physiopathological mechanisms that are prompted by environmental assaults throughout life [4,5]. Autoimmune diseases (ADs) are not an exception. Though the damage to tissues and organs arising from the loss of tolerance is the common attractor to ADs, they represent a diverse collection of diseases defined by their demographic and epidemiological profile, genetic configuration of susceptibility, environmental spectrum and clinical manifestations [4]. Although research more often focuses on a single disease (phenotype), autoimmune phenotypes could represent heterogeneous outcomes of genes underlying similar immunogenic mechanisms, by either cross-phenotype association or by pleiotropy [4,6]. In this sense, clinical observations indicate the possible shift from one disease to another, or the fact that more than one AD may coexist in a single patient (that is, polyautoimmunity) or in the same family (that is, familial autoimmunity) [7].

This article provides a glimpse of the current and future directions for autoimmunity and ADs, discussing the many variables affecting the potential use and application of genetic, evolutionary, demographic, environmental and immunopathological information that could

* Correspondence: juan.anaya@urosario.edu.co
[1]Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24 #63-C-69, Bogota, Colombia
Full list of author information is available at the end of the article

be used for prediction, prevention and eventually treatment of ADs.

The genetic component of ADs

As multifactorial conditions, ADs develop from the cumulative effect of diverse events on the immune system. It is now clear they do not begin at the time of clinical appearance but rather many years before (Figure 1). This window of clinical silence offers the possibility of predicting ADs [8].

Familial aggregation is observed in ADs, but the prevalence in close relatives of affected individuals is usually lower than would be expected if these conditions were Mendelian-like [9]. Recurrent associations have been reported in the literature [10-12]. The diseases of this aggregated pattern share similar genetic risk factors, including the major histocompatibility complex and also non-major histocompatibility complex variants [13-15] (Figure 2). A higher concordance rate of ADs in monozygotic than in dizygotic twins supports a significant effect of genes additively contributing to autoimmunity [16]. Although there is higher concordance in monozygotic twins, environment, stochastic phenomena and exposure still result in discordance in disease thresholds among such twin pairs [17]. Reported heritability, based on available twin concordance rates and prevalence estimated for ADs as a group, ranges from 0.008 for systemic sclerosis to 1.0 for Crohn's disease, with a median value close to 0.6 [18]. ADs are not inherited in a classical Mendelian pattern, but instead have a complex, yet incompletely defined mode of inheritance [19-21]. Further study is needed on environmental and epigenetic factors to clarify their role and effect to allow a greater understanding of their influence, along with genetics, in defining the onset and progression of ADs. The National Institute of Environmental Health Sciences through expert panel workshops has started revisions of such factors to support this growing field of autoimmunity research [22]. For instance, exposure to organic solvents has been shown to affect the risk to develop ADs [23].

Age remains an important topic in autoimmunity, not only because of the biological implications of aging on the immune system but also because of the setback it constitutes for epidemiologic studies [27]. Further complications arise when two diseases are so far apart at their time of diagnosis that a rigorous follow-up becomes imperative to find co-occurrence in one patient [28].

The reason for a major prevalence of ADs among women is poorly understood. The more frequent the AD and the later it appears, the more women are affected [29]. The most convincing explanation of female-biased autoimmunity remains the hormonal theory. Hormones such as estrogens and prolactin have been studied for increasing susceptibility to ADs and can affect both innate and adaptive immune systems [29]. Generally, women have a stronger humoral and cellular immune response than men.

In complex traits, allelic architecture challenges the identification of common and rare genomic variants and their potential effect on risk or protection to develop ADs [15]. Several strategies have been considered to

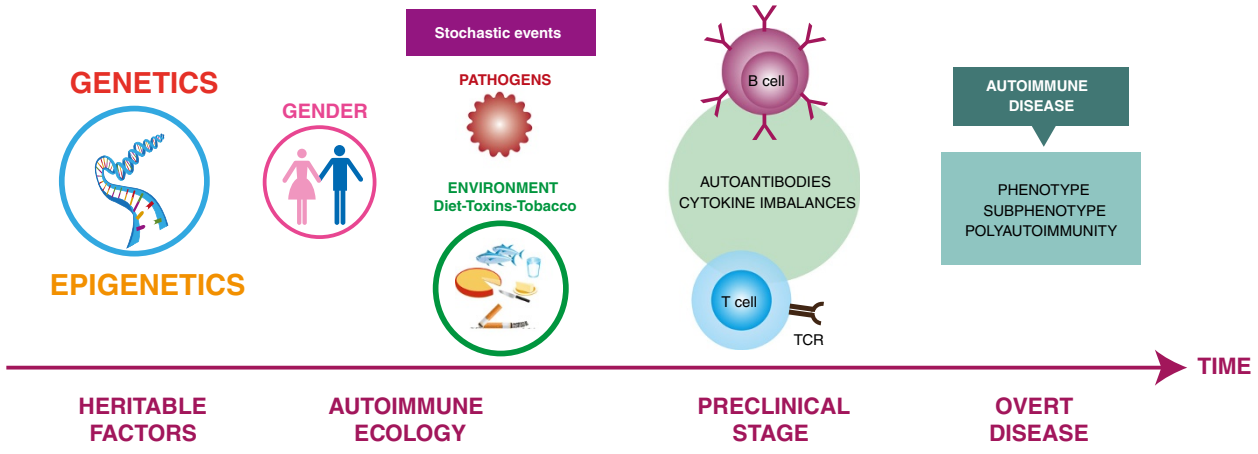


Figure 1 Etiopathogenesis of autoimmune diseases. Outline showing the plausible stages for a multifactorial etiology to develop over time. Each stage shows the known phenomena that cumulatively will be the causative scenario for the onset of disease(s). First, heritable factors (that is, genetics, including ancestry, and epigenetics) impact over the life of the individuals. They converge and interact to create and increase or decrease the liability an individual would have to develop the phenotype depending on risk and protective effects. Women are more affected than men. Second, the autoimmune ecology is characterized by the interactions between an individual and its environment, which acting stochastically will also influence the risk and course of disease. The additive effects of heritable and environmental risk factors favor the loss of autoimmune tolerance. Then, a preclinical stage characterized by B and T cell dysregulation arises. This third phase may take years before the phenotype becomes clinically evident. Adapted from Anaya [4] (with permission from Elsevier). This model may apply to all complex diseases. TCR, T cell receptor.

Figure 2 Weighted list created from the reported significant mapped genes in the current genome-wide association studies curated from the National Human Genome Research Institute and the database of genotypes and phenotypes. The word cloud shows the frequency of genes and its associated variants relative to their font size using a freely available java applet [24]. Both databases (accessed April 2013) [25,26] were queried taking into account *P-values* reported for the genetic variants associated with autoimmune disease. For the National Human Genome Research Institute, a total of 12,064 genetic variants were encountered, out of which 1,370 were variants significantly associated with autoimmune disease susceptibility. In the database of genotypes and phenotypes, out of 31,246 reported variants, 972 were mutually exclusive from the National Human Genome Research Institute, for a grand total of 2,342 genetic variants related to genes associated in a genome-wide association study of any population. The autoimmune diseases of interest were autoimmune thyroid disease, Behcet's disease, celiac disease, rheumatoid arthritis, inflammatory bowel disease, juvenile rheumatoid arthritis, Kawasaki disease, multiple sclerosis, primary biliary cirrhosis, primary sclerosing cholangitis, psoriasis, systemic sclerosis, systemic lupus erythematosus, type 1 diabetes and vitiligo.

dissect variants either associated or co-segregating with ADs (that is, association or linkage approaches such as family-based co-segregation analysis) [9,15]. For association studies, two approaches are available: genome-wide association studies (GWAS) and candidate gene studies. The genome-wide association approach is usually hypothesis-free whereas the candidate gene is hypothesis-driven.

A leap forward towards the recognition of more genes coincided with the advent of high-throughput genotyping technologies and genetic variation repositories, which allowed the use of large sample cohorts to screen for new variants. GWAS interrogate the vast majority of known common polymorphisms [30,31]. This strategy led to a broad array of studies of different AD cohorts (Figure 3), aiming to disclose either new genes or loci associated with ADs or to replicate previously reported associations (Figure 2). Guidelines for the design, quality control and interpretation of GWAS have been presented elsewhere [32-34], as well as novel approaches to study shared genetic factors (for example, cross-phenotype meta-analysis) [35,36].

The overreaching conclusion after the first round of GWAS reports is that genetic heterogeneity, epistasis and complex interactions, plus demographic and environmental factors, underpin the susceptibility to ADs [13-15]. It is unclear how many genetic variants are associated with ADs, and what the immunomolecular mechanisms underlying epistasis among them are. However, a full inventory of variants is not far away and new approaches to examine epistasis will tell us how genes interact to confer either susceptibility or protection against ADs [37]. On top of this genetic view, newly published and publicly available data (for example, exome sequencing project, HapMap and the 1000 genomes project) are at par with technological approaches probing other omic layers like gene expression (for example, RNA-seq, Ribo-seq), methylation (for example, Methyl-seq; BS-seq, Bisulfite Sequencing), other epigenetic marks (for example, ChIP-seq, Chromatin Immunoprecipitation sequencing; FAIRE-seq, formaldehyde-assisted isolation of regulatory elements–sequencing) and genome structure (for example, Immuno-seq; PhIT-Seq, phenotypic interrogation via tag sequencing) [38] are gaining further attention and application to be compared and matched between their omic counterparts. Current ongoing approaches mapping genetic variation contributing to transcriptional variation, referred to as expression quantitative trait locus analyses [39,40], are assessing the role of genetic variants on the expression of genes in their vicinity; empirically, these approaches have been demonstrated to be well-powered to detect regulatory

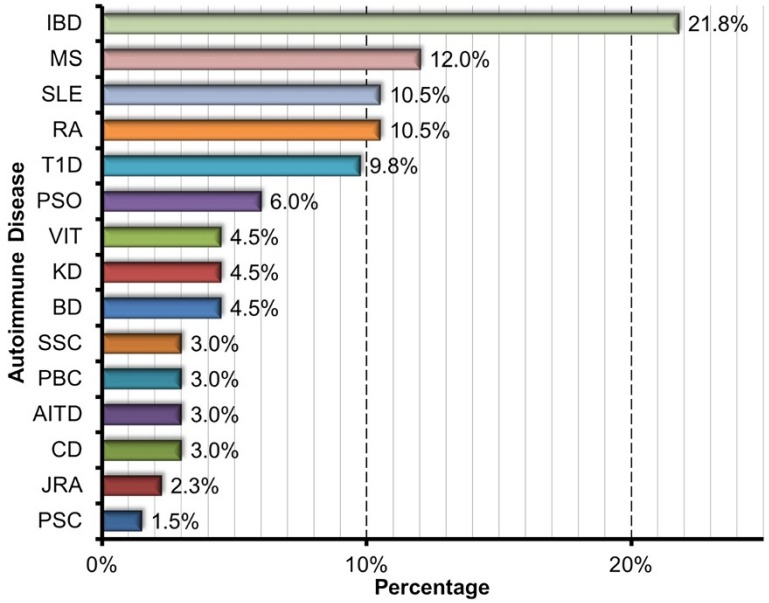


Figure 3 Histogram showing the percentage of autoimmune diseases with significant reported genetic variants in the current genome-wide association studies curated from the National Human Genome Research Institute and the database of genotypes and phenotypes. Both databases [25,26] were accessed in April 2013. AITD, autoimmune thyroid disease; BD, Behcet's disease; CD, celiac disease; IBD, inflammatory bowel disease; JRA, juvenile rheumatoid arthritis; KD, Kawasaki disease; MS, multiple sclerosis; PBC, primary biliary cirrhosis; PSC, primary sclerosing cholangitis; PSO, psoriasis; RA, rheumatoid arthritis; SCL, systemic sclerosis; SLE, systemic lupus erythematosus; SSC, systemic sclerosis; T1D, type 1 diabetes; VIT, vitiligo.

effects [41,42]. This type of post-omic information will add to current knowledge and provide new insights for mechanism and molecular processes for specific phenotyped cells and traits related to the autoimmunity phenomena.

Pitfalls and challenges of complex trait analysis

In recent years, a plethora of new susceptibility genetic variants for ADs has emerged. The advent and advance of microarray and next-generation sequencing technologies has resulted in commercially available tools to provide and obtain genotypes and sequencing information in a fast but costly manner. This exponential production of data is reflected in the number of manuscripts reporting associations of hundreds of loci to ADs. Thus far, the human leukocyte antigen locus has disclosed the strongest association with ADs [43]. In the case of systemic lupus erythematosus, a simple search in PubMed reported more than 5,000 papers on the genetics of the disease. These describe more than 40 loci, replicated by several independent studies, that modify the risk to acquire the disease. However, these systemic lupus erythematosus-associated loci explain a minimal portion of the additive heritability, challenging the idea that this new genetic knowledge might allow for a better predictive and preventive assessment of ADs (that is, missing heritability). Table 1 summarizes the main pitfalls and challenges of complex trait analyses, which we will comment upon next.

Two major challenges in studying ADs are the genetic heterogeneity, referring to how a set of genetic variants might define a trait onset either by their combination or differential effect, and pleiotropy [6], where a single gene leads to multiple phenotypic expressions or disorders. As mentioned by Lehner [44], the sharp statement by Sewal Wright in the 1930s that 'each character is affected by many characters…' is very much true today.

Diverse human populations present different allelic and genotype structures depending upon their evolutionary and epidemiological history [45]. In addition, the effects of genotype on phenotype for any given population may depend on the environment and length of exposure to an undefined etiological insult. Differences in allele and genotype frequencies among populations reflect the contribution of evolutionary forces such as selection, genetic drift, mutation and migration [46], which might explain why some risk alleles to autoimmunity may be protective factors to infectious diseases and vice versa [47]. Immune and infectious agents have been recognized as among the strongest selective pressures for natural populations [47]. Further research regarding exploration of the interplay between infection, type of exposure, additional environmental factors (for example, microbioma) and autoimmunity will result in the discovery of multiple factors underpinning perhaps newly identified physiopathology mechanisms of ADs.

The relatively short evolutionary time since the rise of modern humans after the clash of cultures in America (500 years) is a perfect scenario to dissect specific immunity associated with infectious diseases and its role in predisposition to ADs. Classical examples are Chagas'

Table 1 Pitfalls and challenges of complex trait analysis

Pitfall and challenge	Perspective
Complex epistatic interactions	- Better algorithms and control for phenotype and subphenotype studies. Data analysis is the next most expensive tool to develop.
Genetic heterogeneity	- Larger size cohorts.
Pleiotropy	- Familial studies to control for environmental and stochastic factors.
History of mutations and difference in allele frequencies.	- Description and study of population genetic structure in light of reported information from other reported and publicly available data.
Population stratification	- Usage of newly reported algorithms for admixture analysis and pan-meta-analysis approaches.
Genetics in admixed populations	
Statistical power and sample size	- Correspondence in the use of specific clinical criteria or diagnostic biomarkers to define phenotypes to enhance prediction and diagnosis.
Refining the phenotype - subphenotypes	Development and application of bioinformatical approaches to classify disease as quantitative and categorical entities.
Family based studies versus case–control studies	Application of classical genetic and epidemiological tools to characterize new information available for other 'omic' layers in the context of the genome from a familial and population viewpoint.
Gene-environment interaction	Further research in environmental factors that might influence onset of disease (for example, tobacco, coffee consumption, organic solvents)
Post-genomic era ('omics')	Use of the publicly available 'omic' information already reported (for example, ENCODE, GEO, HapMap, 1000 genomes project) to explore, replicate and hypothesize new experimental functional designs.
Personalized medicine	Genomic medicine-generated information to be applicable from the bench to bedside and also from the bedside to bench.
Pharmacogenomics	Disentangle markers capable of predicting and diagnosing risk of disease even before onset of symptoms and signs.

disease (originally found in America and absent in other continents) and typhoid fever (brought to America by the Spaniards conquers). Indeed, it is not only the knowledge that might be contributed by this type of population, but also the specific and direct epidemiological and health care approach that must be provided to them. Admixed populations such as Afro-American and Latin-American are often medically underserved and bear a disproportionately high burden of disease. Thus, given the diversity of their genomes, these populations have both advantages and disadvantages for genetic studies of complex phenotypes [48]. Advances in statistical methodologies that use genetic contributions from ancestral populations contributing to the current admixed population have proven to be a powerful method to leverage the confounder effect of ancestry, and this information is used to identify chromosomal segments linked to disease [46].

Consequently, there is a need to explore genetic associations in diverse populations. Proper matching of cases and controls is a major consideration for GWAS, as well as in any case–control association study. The use of ancestry informative markers either to match or exclude cases and controls given specific patterns of genetic stratification allows us to overcome this limitation, diminishing the possibility of reaching spurious associations as a consequence of case–control ethnic microdifferentiation.

Determinants of statistical power such as sample size, disease heterogeneity, pedigree and genotyping errors, as well as the effect of the type and density of genetic markers, are a key factor in genetic studies. Studies should either have sufficient power to detect a small effect size of multiple genes or consider the use of extreme and well-defined phenotypes to detect the effect of major genes [30,31].

The term 'metagenomics' defines the set of mechanisms by which a community of microorganisms interacts, lives and infects animal tissues. New metagenomic approaches have disclosed crucial information about the shaping of resistance, susceptibility and loss of auto-tolerance for both infectious and ADs [49]. Indeed, new reports demonstrate that host-gene-microbial interactions are major determinants for the development of ADs. Commensal microbial communities may alter sex hormone levels and regulate AD fate in individuals with a high genetic risk load [50].

Although ADs are often diagnosed according to classification criteria, they share similar subphenotypes including signs and symptoms, non-specific autoantibodies and high levels of cytokines, which are prone to taxonomic problems [51]. ADs have a heterogeneous spectrum, the disease course differs from patient to patient and through different phases within the same patient [52]. Refining the phenotype will make the effect of certain genes in the sample more easily detectable [4]. Genetic effects may be stronger for extremes of the risk factor distribution (for example, people with onset at a very young or very old age) and for particular presentations. Therefore, restricting the sample to patients with specific characteristics, or minimizing the effect of known environmental confounders will increase the chances for genetic research to be successful.

Disease heterogeneity should be minimized by considering subphenotypes or otherwise by adjusting for known sources of heterogeneity as a covariate. Meta-analysis and data pooling between different research groups can provide a sizeable study, but both approaches require a high level of vigilance about locus and disease heterogeneity when data come from different populations. Spurious associations are often due to population stratification, cryptic relatedness and differential bias [53].

GWAS have a high power to detect common variants of high or moderate effect. For weaker effects (for example, relative risk <1.2), the power is greatly reduced, particularly for recessive loci if the frequency of the variant is common (that is, rare variants) [54]. Larger size cohorts can be used to study common diseases, but meta-analyses and data pooling are required to attain a study size of sufficient magnitude for many other diseases [53]. GWAS approaches are known to be poor in detecting effects from rare alleles (that is, frequency <5%), but novel methods and technology, such as exome and whole genome sequencing will fill this gap to further support the genetic commonality of autoimmune traits [55]. However, once a polymorphism has been found to be associated with a trait, it's functional relevance must be examined and its biological effect on such a trait understood (that is, functional genomics).

Recent advances in multiplexed assay technology are taking us closer toward the identification of 'actionable markers', capable of informing and providing biological metrics of use in clinical practice. Not only will they help gain insights into the onset, remission and exacerbation of a pathology, they will improve and enhance treatment, diagnosis and classification [56].

What is next?

Genomics normally implies the use of sequence and genome information to annotate, describe and curate functionality and structure, in order to decipher and disentangle functionality and organization. New 'omics' approaches are starting to take this further by correlating and matching layers of genome-wide information to explain and to explore mechanisms of interaction between genetic and environmental factors. Significant advances in human 'omics' are giving rise to new possibilities in medicine, such as clinical bioinformatics [57] and translational bioinformatics [58]. All these options lead to one common premise: ways of mining meaningful information from the vast amount of 'omics' data being generated. In this sense, application of comprehensive molecular information to clinical settings is been referred to as 'genomic medicine' [59] with the ultimate goal to nurture, improve and frame personalized medicine. A genomic medicine approach will always require participation at a multidisciplinary research expertise level.

Personalized medicine is committed to survey, monitor and diagnose risks to provide patients with a specific treatment, taking into account their particular genetic profile and molecular phenotype. Thus evaluation, comparison, correlation, cross-matching and interaction of the nascent 'omic' information would not only aid in the prediction, diagnosis and treatment at the individual level but also provide insights into the physiopathological mechanisms of disease onset and progression. For such purposes, an integrative personal 'omics' profile such as the one suggested by Chen *et al.* [60] will be useful to examine as many biological components as possible. Although these components might change during healthy and diseased states, this information combined with genomic information will be useful to estimate disease risk and gain new insights into diseased states [60]. Disease would be considered as a hierarchical biological system composed of molecular and functional cell, tissue and organ interactive networks. Any aberration in one or more networks will not only have local effects but also systemic effects because no cell, tissue or organ is isolated or independent.

Last but not least, safeguarding for all study participants, whether healthy or affected, and studied family members has to be warranted. Individuals are the 'why' behind this overhauling of 'omic' and genomics approaches and research, thus their legal rights and *status quo* have to be defined in order to eventually be successful in applying genomic-based medicine for the benefit of human kind. We shall not forget the understated idea '…we should not only be interested in the human genome but also in the human beings that carry it' [61].

Abbreviations

ADs: Autoimmune diseases; GWAS: Genome-wide association study.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

JMA designed the review. JC, MAB and JMA jointly wrote the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgements

This work was supported by Universidad del Rosario and Colciencias (122254531722), Bogota, Colombia.

Author details

[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. [3]Center for Personalized Medicine, South Texas Veterans Health Care System, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. [4]Department of Medicine, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA. [5]Genome Biology Department, John Curtin School of Medical Research, ANU College of Medicine, Biology and Environment, The Australian National University, Canberra, Australia.

Received: 8 April 2013 Accepted: 12 August 2013
Published: 4 September 2013

References

1. Green ED, Guyer MS: **Charting a course for genomic medicine from base pairs to bedside.** *Nature* 2011, **470:**204–213.
2. Hennekam RC, Biesecker LG: **Next-generation sequencing demands next-generation phenotyping.** *Hum Mutat* 2012, **33:**884–886.
3. Chan IS, Ginsburg GS: **Personalized medicine: progress and promise.** *Ann Rev Genomics Hum Genet* 2011, **12:**217–244.
4. Anaya JM: **Common mechanisms of autoimmune diseases (the autoimmune tautology).** *Autoimmun Rev* 2012, **11:**781–784.
5. Arcos-Burgos M, Velez JI, Solomon BD, Muenke M: **A common genetic network underlies substance use disorders and disruptive or externalizing disorders.** *Hum Genet* 2012, **131:**917–929.
6. Solovieff N, Cotsapas C, Lee PH, Purcell SM, Smoller JW: **Pleiotropy in complex traits: challenges and strategies.** *Nature Rev Genet* 2013, **14:**483–495.
7. Cardenas-Roldan J, Rojas-Villarraga A, Anaya JM: **How do autoimmune diseases cluster in families? A systematic review and meta-analysis.** *BMC Med* 2013, **11:**73.
8. Tobon GJ, Pers JO, Canas CA, Rojas-Villarraga A, Youinou P, Anaya JM: **Are autoimmune diseases predictable?** *Autoimmun Rev* 2012, **11:**259–266.
9. Anaya JM, Gomez L, Castiblanco J: **Is there a common genetic basis for autoimmune diseases?** *Clin Dev Immunol* 2006, **13:**185–195.
10. Somers EC, Thomas SL, Smeeth L, Hall AJ: **Autoimmune diseases co-occurring within individuals and within families: a systematic review.** *Epidemiology* 2006, **17:**202–217.
11. Anaya JM, Corena R, Castiblanco J, Rojas-Villarraga A, Shoenfeld Y: **The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity.** *Expert Rev Clin Immunol* 2007, **3:**623–635.
12. Rojas-Villarraga A, Amaya-Amaya J, Rodriguez-Rodriguez A, Mantilla RD, Anaya JM: **Introducing polyautoimmunity: secondary autoimmune diseases no longer exist.** *Autoimmune Dis* 2012, **2012:**254319.
13. Richard-Miceli C, Criswell LA: **Emerging patterns of genetic overlap across autoimmune disorders.** *Genome Med* 2012, **4:**6.
14. Becker KG, Simon RM, Bailey-Wilson JE, Freidlin B, Biddison WE, McFarland HF, Trent JM: **Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases.** *Proc Natl Acad Sci U S A* 1998, **95:**9979–9984.
15. Goris A, Liston A: **The immunogenetic architecture of autoimmune disease.** *Cold Spring Harb Perspect Biol* 2012, **4:**a007260.
16. Bogdanos DP, Smyk DS, Rigopoulou EI, Mytilinaiou MG, Heneghan MA, Selmi C, Gershwin ME: **Twin studies in autoimmune disease: genetics, gender and environment.** *J Autoimmun* 2012, **38:**J156–J169.
17. Czyz W, Morahan JM, Ebers GC, Ramagopalan SV: **Genetic, environmental and stochastic factors in monozygotic twin discordance with a focus on epigenetic differences.** *BMC Med* 2012, **10:**93.
18. Selmi C, Lu Q, Humble MC: **Heritability versus the role of the environment in autoimmunity.** *J Autoimmun* 2012, **39:**249–252.
19. Bias WB, Reveille JD, Beaty TH, Meyers DA, Arnett FC: **Evidence that autoimmunity in man is a Mendelian dominant trait.** *Am J Hum Genet* 1986, **39:**584–602.
20. Arcos-Burgos M, Parodi E, Salgar M, Bedoya E, Builes J, Jaramillo D, Ceballos G, Uribe A, Rivera N, Rivera D, Fonseca I, Camargo M, Palacio G: **Vitiligo: complex segregation and linkage disequilibrium analyses with respect to microsatellite loci spanning the HLA.** *Hum Genet* 2002, **110:**334–342.
21. Lin JP, Cash JM, Doyle SZ, Peden S, Kanik K, Amos CI, Bale SJ, Wilder RL: **Familial clustering of rheumatoid arthritis with other autoimmune diseases.** *Hum Genet* 1998, **103:**475–482.
22. Miller FW, Alfredsson L, Costenbader KH, Kamen DL, Nelson LM, Norris JM, De Roos AJ: **Epidemiology of environmental exposures and human autoimmune diseases: findings from a National Institute of Environmental Health Sciences Expert Panel Workshop.** *J Autoimmun* 2012, **39:**259–271.
23. Barragan-Martinez C, Speck-Hernandez CA, Montoya-Ortiz G, Mantilla RD, Anaya JM, Rojas-Villarraga A: **Organic solvents as risk factor for autoimmune diseases: a systematic review and meta-analysis.** *PLoS One* 2012, **7:**e51506.
24. Feinberg J: *Wordle.* http://www.wordle.net.
25. Hindorff LA, MacArthur J, (European Bioinformatics Institute), Morales J, European Bioinformatics Institute, Junkins HA, Hall PN, Klemm AK, Manolio TA: *A Catalog of Published Genome-Wide Association Studies.* http://www.genome.gov/gwastudies.
26. *The Database of Genotypes and Phenotypes.* http://www.ncbi.nlm.nih.gov/gap.
27. Hsu HC, Mountz JD: **Origin of late-onset autoimmune disease.** *Immunol Allergy Clin North Am* 2003, **23:**65–82. vi.
28. Sloka S: **Observations on recent studies showing increased co-occurrence of autoimmune diseases.** *J Autoimmun* 2002, **18:**251–257.
29. Quintero OL, Amador-Patarroyo MJ, Montoya-Ortiz G, Rojas-Villarraga A, Anaya JM: **Autoimmune disease and gender: plausible mechanisms for the female predominance of autoimmunity.** *J Autoimmun* 2012, **38:**J109–J119.
30. Hirschhorn JN, Gajdos ZK: **Genome-wide association studies: results from the first few years and potential implications for clinical medicine.** *Ann Rev Med* 2011, **62:**11–24.
31. Lessard CJ, Ice JA, Adrianto I, Wiley GB, Kelly JA, Gaffney PM, Montgomery CG, Moser KL: **The genomics of autoimmune disease in the era of genome-wide association studies and beyond.** *Autoimmun Rev* 2012, **11:**267–275.
32. Pearson TA, Manolio TA: **How to interpret a genome-wide association study.** *JAMA* 2008, **299:**1335–1344.
33. Barsh GS, Copenhaver GP, Gibson G, Williams SM: **Guidelines for genome-wide association studies.** *PLoS Genet* 2012, **8:**e1002812.
34. Austin MA, Hair MS, Fullerton SM: **Research guidelines in the era of large-scale collaborations: an analysis of Genome-wide Association Study Consortia.** *Am J Epidemiol* 2012, **175:**962–969.
35. Cotsapas C, Hafler DA: **Immune-mediated disease genetics: the shared basis of pathogenesis.** *Trends Immunol* 2013, **34:**22–26.
36. Cotsapas C, Voight BF, Rossin E, Lage K, Neale BM, Wallace C, Abecasis GR, Barrett JC, Behrens T, Cho J, De Jager PL, Elder JT, Graham RR, Gregersen P, Klareskog L, Siminovitch KA, van Heel DA, Wijmenga C, Worthington J, Todd JA, Hafler DA, Rich SS, Daly MJ, FOCiS Network of Consortia: **Pervasive sharing of genetic effects in autoimmune disease.** *PLoS Genet* 2011, **7:**e1002254.
37. Lee S, Kwon MS, Park T: **Network graph analysis of gene-gene interactions in genome-wide association study data.** *Genomics Inform* 2012, **10:**256–262.
38. Shendure J, Lieberman Aiden E: **The expanding scope of DNA sequencing.** *Nature Biotechnol* 2012, **30:**1084–1094.
39. Li J, Burmeister M: **Genetical genomics: combining genetics with gene expression analysis.** *Hum Mol Genet* 2005, **14 Spec No. 2:**R163–R169.
40. Costa V, Aprile M, Esposito R, Ciccodicola A: **RNA-Seq and human complex diseases: recent accomplishments and future perspectives.** *Eur J Hum Genet* 2013, **21:**134–142.
41. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK: **High-resolution mapping of expression-QTLs yields insight into human gene regulation.** *PLoS Genet* 2008, **4:**e1000214.
42. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, Ellis P, Langford C, Vannberg FO, Knight JC: **Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles.** *Nature Genet* 2012, **44:**502–510.
43. Cruz-Tapias P, Perez-Fernandez OM, Rojas-Villarraga A, Rodriguez-Rodriguez A, Arango MT, Anaya JM: **Shared HLA class II in six autoimmune diseases in Latin America: a meta-analysis.** *Autoimmune Dis* 2012, **2012:**569728.
44. Lehner B: **Genotype to phenotype: lessons from model organisms for human genetics.** *Nature Rev Genet* 2013, **14:**168–178.
45. Mori M, Yamada R, Kobayashi K, Kawaida R, Yamamoto K: **Ethnic differences in allele frequency of autoimmune-disease-associated SNPs.** *J Hum Genet* 2005, **50:**264–266.
46. Molineros JE, Maiti AK, Sun C, Looger LL, Han S, Kim-Howard X, Glenn S, Adler A, Kelly JA, Niewold TB, Gilkeson GS, Brown EE, Alarcón GS, Edberg JC, Petri M, Ramsey-Goldman R, Reveille JD, Vilá LM, Freedman BI, Tsao BP, Criswell LA, Jacob CO, Moore JH, Vyse TJ, Langefeld CL, Guthridge JM, Gaffney PM, Moser KL, Scofield RH, Alarcón-Riquelme ME, *et al*: **Admixture mapping in lupus identifies multiple functional variants within IFIH1 associated with apoptosis, inflammation, and autoantibody production.** *PLoS Genet* 2013, **9:**e1003222.
47. Boechat AL, Ogusku MM, Sadahiro A, Dos Santos MC: **Association between the PTPN22 1858C/T gene polymorphism and tuberculosis resistance.** *Infect Genet Evol* 2013, **16:**310–313.
48. Seldin MF, Pasaniuc B, Price AL: **New approaches to disease mapping in admixed populations.** *Nature Rev Genet* 2011, **12:**523–528.
49. Cho I, Blaser MJ: **The human microbiome: at the interface of health and disease.** *Nature Rev Genet* 2012, **13:**260–270.
50. Markle JG, Frank DN, Mortin-Toth S, Robertson CE, Feazel LM, Rolle-Kampczyk U, von Bergen M, McCoy KD, Macpherson AJ, Danska JS: **Sex differences in the gut microbiome drive hormone-dependent regulation of autoimmunity.** *Science* 2013, **339:**1084–1088.
51. Cho JH, Gregersen PK: **Genomics and the multifactorial nature of human autoimmune disease.** *N Engl J Med* 2011, **365:**1612–1623.

52. Castiblanco J, Anaya JM: **The nature and nurture of common autoimmunity.** *Ann N Y Acad Sci* 2007, **1109:**1–8.
53. Nsengimana J, Bishop DT: **Design considerations for genetic linkage and association studies.** *Methods Mol Biol* 2012, **850:**237–262.
54. Hunt KA, Mistry V, Bockett NA, Ahmad T, Ban M, Barker JN, Barrett JC, Blackburn H, Brand O, Burren O, Capon F, Compston A, Gough SC, Jostins L, Kong Y, Lee JC, Lek M, MacArthur DG, Mansfield JC, Mathew CG, Mein CA, Mirza M, Nutland S, Onengut-Gumuscu S, Papouli E, Parkes M, Rich SS, Sawcer S, Satsangi J, Simmonds MJ, *et al*: **Negligible impact of rare autoimmune-locus coding-region variants on missing heritability.** *Nature* 2013, **498:**232–235.
55. Voight BF, Cotsapas C: **Human genetics offers an emerging picture of common pathways and mechanisms in autoimmunity.** *Curr Opin Immunol* 2012, **24:**552–557.
56. Maecker HT, Lindstrom TM, Robinson WH, Utz PJ, Hale M, Boyd SD, Shen-Orr SS, Fathman CG: **New tools for classification and monitoring of autoimmune diseases.** *Nature Rev Rheumatol* 2012, **8:**317–328.
57. Knaup P, Ammenwerth E, Brandner R, Brigl B, Fischer G, Garde S, Lang E, Pilgram R, Ruderich F, Singer R, Wolff AC, Haux R, Kulikowski C: **Towards clinical bioinformatics: advancing genomic medicine with informatics methods and tools.** *Methods Inf Med* 2004, **43:**302–307.
58. Kann MG: **Advances in translational bioinformatics: computational approaches for the hunting of disease genes.** *Brief Bioinform* 2010, **11:**96–110.
59. Feero WG, Guttmacher AE, Collins FS: **Genomic medicine–an updated primer.** *N Engl J Med* 2010, **362:**2001–2011.
60. Chen R, Mias GI, Li-Pook-Than J, Jiang L, Lam HY, Miriami E, Karczewski KJ, Hariharan M, Dewey FE, Cheng Y, Clark MJ, Im H, Habegger L, Balasubramanian S, O'Huallachain M, Dudley JT, Hillenmeyer S, Haraksingh R, Sharon D, Euskirchen G, Lacroute P, Bettinger K, Boyle AP, Kasowski M, Grubert F, Seki S, Garcia M, Whirl-Carrillo M, Gallardo M, Blasco MA, *et al*: **Personal omics profiling reveals dynamic molecular and medical phenotypes.** *Cell* 2012, **148:**1293–1307.
61. Bernal JE: **Human diversity in Colombia.** *Science* 1995, **267:**774.

doi:10.1186/1741-7015-11-197
Cite this article as: Castiblanco *et al.*: What is next after the genes for autoimmunity? *BMC Medicine* 2013 **11**:197

Chapter 2 - Familial Aggregation and Segregation Analysis in Families Presenting Autoimmunity, Polyautoimmunity, and Multiple Autoimmune Syndrome

Castiblanco John, Sarmiento-Monroy Juan-Camilo, Mantilla Ruben-Dario, Rojas-Villarraga Adriana, Anaya Juan-Manuel. Familial Aggregation and Segregation Analysis in Families Presenting Autoimmunity, Polyautoimmunity, and Multiple Autoimmune Syndrome. J Immunol Res. 2015; 2015: 572 353. Epub 2015 Nov 30. PMID: 26697508

Hindawi Publishing Corporation
Journal of Immunology Research
Volume 2015, Article ID 572353, 10 pages
http://dx.doi.org/10.1155/2015/572353

Research Article

Familial Aggregation and Segregation Analysis in Families Presenting Autoimmunity, Polyautoimmunity, and Multiple Autoimmune Syndrome

John Castiblanco,[1,2] Juan Camilo Sarmiento-Monroy,[1] Ruben Dario Mantilla,[1] Adriana Rojas-Villarraga,[1] and Juan-Manuel Anaya[1]

[1]*Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24 No. 63-C-69, Bogotá, Colombia*
[2]*Doctoral Program in Biomedical Sciences, Universidad del Rosario, Bogotá, Colombia*

Correspondence should be addressed to Juan-Manuel Anaya; juan.anaya@urosario.edu.co

Received 17 August 2015; Accepted 29 October 2015

Academic Editor: Fulvia Ceccarelli

Studies documenting increased risk of developing autoimmune diseases (ADs) have shown that these conditions share several immunogenetic mechanisms (i.e., the autoimmune tautology). This report explored familial aggregation and segregation of AD, polyautoimmunity, and multiple autoimmune syndrome (MAS) in 210 families. Familial aggregation was examined for first-degree relatives. Segregation analysis was implemented as in S.A.G.E. release 6.3. Data showed differences between late- and early-onset families regarding their age, age of onset, and sex. Familial aggregation of AD in late- and early-onset families was observed. For polyautoimmunity as a trait, only aggregation was observed between sibling pairs in late-onset families. No aggregation was observed for MAS. Segregation analyses for AD suggested major gene(s) with no clear discernible classical known Mendelian transmission in late-onset families, while for polyautoimmunity and MAS no model was implied. Data suggest that polyautoimmunity and MAS are not independent traits and that gender, age, and age of onset are interrelated factors influencing autoimmunity.

1. Introduction

Autoimmune diseases (ADs) are responsible for a substantial amount of disability and morbidity worldwide. Although their epidemiology varies according to individual conditions, collectively, autoimmune prevalence is at least 5% in the general population and is one of the major causes of premature mortality in young and middle aged women [1].

As heterogeneous diseases, ADs develop from the cumulative effect of diverse events on the immune system [2]. It is clear that ADs do not begin at the time of clinical appearance but rather many years before. A common origin for diverse ADs is sustained by three levels of evidence [3]: 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., polyautoimmunity) [4–8] or in the same family (i.e., familial autoimmunity) [9]; a second level of evidence refers to known shared pathophysiological mechanisms between ADs [10, 11]. Epidemiological studies show correlations among certain ADs, linking epidemiological observations to physiopathological evidence for AD might contribute to our knowledge for the shared etiological and immunogenetic mechanisms [2]; and a third level of evidence corresponds to the evidence implying common genetic factors [7]. The importance of this concept focuses on the probability of having multiple ADs simultaneously in one patient, which goes beyond epidemiologic inferences.

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 (ADs) [12]. However, due to their multifactorial and polygenic nature, accompanied by differential penetrance

influenced by environmental factors and genetic heterogeneity among populations [13, 14], untangling of the genetic determinants defining their outcome and onset has proven to be extremely challenging. Likewise, 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 [15].

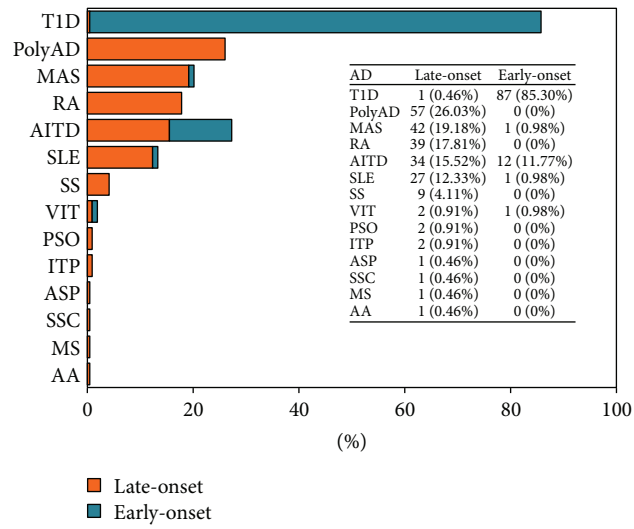
Therefore, family history of ADs should be considered when performing genetic analysis as this new approach incorporates all accepted pathologies for which evidence suggests an autoimmune origin. Families with multiple affected relatives appear to share common risk alleles with sporadic patients but may have a higher genetic load. A consequence of the polygenic model for complex diseases is that patients are inevitably highly heterogeneous in terms of the particular set of risk alleles they carry. It has been suggested that this may translate in different genetically determined disease mechanisms in subgroups of patients or a common disease mechanism that is complemented by additional pathways that are more or less predominant in different subgroups [16]. Familial approaches have documented the clustering of certain ADs among the relatives of individuals who have rheumatoid arthritis (RA), multiple sclerosis (MS), systemic lupus erythematosus (SLE), and type 1 diabetes mellitus (T1D) among other diseases [17–23].

ADs are not inherited in a simple, classical Mendelian way but have instead a complex or a yet uncharacterized mode of inheritance [13, 24]. Bias et al. were the first to consider a single major gene conferring susceptibility for autoimmunity and suggested an autosomal dominant inheritance pattern with penetrance of approximately 92% in females and 9% in males [13]. In addition, Arcos-Burgos et al. showed the presence of a dominant major gene and strong environmental effects as the most parsimonious model of segregation for VIT [25]. On the other hand, when analyzing RA together with other ADs, a mixed model fitted the data significantly better than the major gene or polygenic models [26].

The clinical evidence of the autoimmune tautology highlights the cooccurrence of distinct ADs within an individual [27]. ADs coexistence in a single individual has led researchers to consider different terms like autoimmune diathesis [28] or kaleidoscope of autoimmunity [29] both of which point to a common genetic background of ADs [6]. In an effort to understand and further support the commonality of autoimmunity as a trait among ADs, the present study examined the dynamics of familial aggregation and segregation in AD, polyautoimmunity, and multiple autoimmune syndrome (MAS) in well-defined and characterized patients and their relatives from Colombia, South America.

2. Materials and Methods

2.1. Study Population and Family Collection. This study sample consisted of multiplex families of varying size ascertained through patients treated at the Center for Autoimmune Diseases Research (CREA) in Medellin and Bogotá at the University of Rosario, Colombia (Table 1). (i) Each recruited family presented a proband with at least one AD according to validated international classification criteria; (ii) each recruited family presented at least one family member with polyautoimmunity (i.e., cooccurrence of distinct ADs within an individual); (iii) each recruited family presented evidence of familial autoimmunity (i.e., different ADs within members of a nuclear family); and (iv) each other affected individual presented a well-defined autoimmune phenotype (i.e., fulfillment of international classification criteria in first-degree relatives (FDRs)). Moreover, families in which the proband presented with T1D were included and used as early-onset AD families (Figure 1). FDRs were defined as parents and siblings.



AD	Late-onset	Early-onset
T1D	1 (0.46%)	87 (85.30%)
PolyAD	57 (26.03%)	0 (0%)
MAS	42 (19.18%)	1 (0.98%)
RA	39 (17.81%)	0 (0%)
AITD	34 (15.52%)	12 (11.77%)
SLE	27 (12.33%)	1 (0.98%)
SS	9 (4.11%)	0 (0%)
VIT	2 (0.91%)	1 (0.98%)
PSO	2 (0.91%)	0 (0%)
ITP	2 (0.91%)	0 (0%)
ASP	1 (0.46%)	0 (0%)
SSC	1 (0.46%)	0 (0%)
MS	1 (0.46%)	0 (0%)
AA	1 (0.46%)	0 (0%)

Figure 1: Frequency and distribution of autoimmune disease (AD) in late-onset and early-onset families included in this report. For analytical purposes, families were divided into two types: late-onset (i.e., families where a proband presents a late-onset AD) and early-onset (i.e., T1D families) (Figure 1).

Patients with AD, polyautoimmunity, and MAS fulfilled validated classification criteria and were part of a multicenter cohort followed at the CREA. Their information on demographics and cumulative clinical manifestations over the course of disease were obtained by both chart review and discussion with the patient and were collected in a standard data collection form. Only relatives of Colombian patients were included and interviewed, following the methodology described by Priori et al. [30], using a standardized questionnaire that incorporates demographics and medical information including a check-point list of 18 ADs [21]. 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 [6, 21].

In T1D families, recruited cases were children all of whom fulfilled the diagnostic classification criteria proposed by the American Diabetes Association (ADA) [31] and had been previously described [32] (Table 1). Their information on demographics and cumulative clinical manifestations over

Table 1: Characteristics of probands and families classified as late-onset and early-onset.

Characteristic	Late-onset				Early-onset		
	All	AD	PolyAD	MAS	All	AD	PolyAD/MAS
Age (yrs)	49.43	45.99	45.49	44.81	32.32**	19.54**	16
[Min, Max]	[11, 91]	[13, 83]	[16, 78]	[20, 64]	[3, 94]	[4, 70]	
Age at onset (yrs)	—	32.80	33.42	33.97	—	7.77**	11
[Min, Max]		[5, 62]	[5, 62]	[5, 62]		[1, 24]	
Male							
Aff (Unaff)	265	24 (156)	8 (172)	2 (178)	227	50 (141)	0 (191)
Female							
Aff (Unaff)	451	195 (216)	91 (320)	41 (370)	216	52 (152)	1 (203)
Number of Peds	127				83		
Mean size ± SD	5.64 ± 2.76				5.34 ± 2.94		
[Min, Max]	[3, 16]				[3, 20]		

AD: autoimmune disease; PolyAD: polyautoimmunity; MAS: multiple autoimmune syndrome. Data correspond to FDRs affected or unaffected and taking into account the analysis. Aff: affected; Unaff: unaffected.
** p value < 0.001 t-test when comparing late-onset versus early-onset variables.

the course of disease were obtained by both chart review and discussion with the patient and were collected in a standard data collection form. A total of 87 patients with T1D were analyzed and their relatives were included (Table 1).

For individuals (i.e., probands and FDR) 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 hematological diseases and hepatitis B virus, hepatitis C virus, or human immunodeficiency virus infections. As for the family characteristics in our population, most of them are nuclear and at least 30% are multigenerational [33, 34]. The great majority of our country households still contain related persons. In addition, all family members participating in this study were living in the same city and approved informed consent in order to participate in the present study. This research is being carried out in accordance with Resolution number 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 Universidad del Rosario approved the present project.

2.2. Statistical and Genetic Data Analysis. Data was managed and stored using the R software version 3.1.1 [35] and Excel spreadsheets. Results are presented as means ± standard deviation (SD) and minimum/maximum and/or in percentages. Comparison between means was performed by Student's t-test and those between percentages by the χ^2 test and two-sided Fisher's exact test, where appropriate. 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), and (iii) family/pedigree relationships. Estimation of the distributions of relationship types and affection status among relatives pairs were performed using the Statistical Analysis for Genetic Epidemiology (S.A.G.E.) program PEDINFO, release 6.3 [36]. 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.

Familial Aggregation Analysis. Recurrent risk ratios (λ_R) were calculated for first-degree relatedness (parent/offspring and sibling/sibling pairs) using the formula $\lambda_R = K_{\text{Relative}}/K$, where K_{Relative} (K_R) is the prevalence for a specific degree of relatedness in the sample and K is the mean prevalence in the population [37] and/or the previously reported K in specific pairs of relatives in the same population [21]. Information about the prevalence of ADs in our population is not clear and available; for this matter prevalence values in the range of 0.1%–0.5% were chosen as reported in the literature [1, 38–45]. Therefore, 0.5% (5/1000 individuals) for AD and 2.5% (25/1000 individuals) for all ADs taken together were selected as putative population prevalence as previously reported [1, 21, 38–45]. These methods were extended to ascertain whether or not clustering of two or more autoimmune disorders in relatives increased the probability or the risk for the presence of the disorder in the affected proband.

Familial Segregation Analysis. Analyses on 210 single ascertained pedigrees (Table 1) to identify the most plausible model explaining the segregation of AD, polyautoimmunity, and MAS in late-onset (non-T1D families) and early-onset families (T1D families) were performed for a binary trait as implemented in SEGREG S.A.G.E. release 6.3 (Table 2). SEGREG uses maximum-likelihood methods to estimate the parameters of mathematical models of disease occurrence in families. Each model assumes that the presence (or absence) of a putative disease allele influences susceptibility to the trait and applies the regressive multivariate logistic model allowing us to include available covariates into the fitted models.

Table 2: Parameter estimates from segregation analysis of autoimmune disease proband-ascertained pedigrees.

Model/parameter	Type susceptibilities			Transmission probabilities			Freq	Multifactorial/polygenic effect[b]
	β_{AA}	β_{AB}	β_{BB}	τ_{AA}	τ_{AB}	τ_{BB}	q_A	$\rho_{FM} = 0$[a]; $\rho_{F0} = \rho_{M0}$
(1) Random environmental	—	—	—	q_A	q_A	q_A	*	0
(2) Dominant	*	β_{AA}	*	1	0.5	0	*	0
(3) Dominant multifactorial	*	β_{AA}	*	1	0.5	0	*	*
(4) Recessive	*	β_{BB}	*	1	0.5	0	*	0
(5) Recessive multifactorial	*	β_{BB}	*	1	0.5	0	*	*
(6) Codominant	*	*	*	1	0.5	0	*	0
(7) Additive	*	$(1/2)(\beta_{AA} + \beta_{BB})$	*	1	0.5	0	*	0
(8) Mayor gene	*	*	*	*	*	*	*	0
(9) General transmission[c]	*	*	*	*	*	*	*	*

*Parameters freely estimated within an appropriate range; q_A: allele frequency; when $\tau_{AA} = 1.0$, $\tau_{AB} = 0.5$, and $\tau_{BB} = 0.0$, Mendelian transmission is assumed; when q_A is estimated under Mendelian transmission, Hardy-Weinberg proportions ($\psi_{AA} = q_A^2$; $\psi_{AB} = 2q_A^2(1 - 2q_A^2)$; $\psi_{BB} = q_B^2$) are assumed.

[a]Father-mother correlations, set to 0 assuming absence of assortative mating or consanguineous mating.

[b]Polygenic transmission effect inclusion assumes that the phenotype is determined by polygenic inheritance, so the phenotype has one distribution, and familial correlations can explain the familial aggregation of the trait.

[c]All parameters are estimated in Model 9. As a result, all other models are nested, and thus the general model is used as the baseline to compare all other models in this study.

Models Description. Random environmental model (Model 1) assumes that the trait segregation is caused purely by a random environmental factor and there is no transmission from generation to generation ($\tau_{AA} = \tau_{AB} = \tau_{BB} = q_A$). Pure major locus transmission models (Models 2, 4, 6, and 8) assume major locus transmission in a Mendelian mode, without multifactorial/polygenic inheritance. Major gene plus multifactorial/polygenic models (Models 3 and 5) assumes that both a major locus (transmitted in a Mendelian mode) and a multifactorial/polygenic effect influence the trait. The general model (Model 9) is the unrestricted full model, which subsumes all of the other models.

The fitted models assumed that the likelihood for any two individuals presenting with the phenotype and having the major type over nuclear families is independent. Consequently, the susceptibility (marginal probability) that any pedigree member has a particular phenotype is the same for all members who have the same values of any covariates in the model. This susceptibility is given the cumulative logistic function $\lambda = e^{\theta y}/(1 + e^{\theta y})$, where y is the affection status phenotype of ith individual and θ is the logit of the susceptibility for ith individual defined as $\theta(i) = \log[p(Y = 1)/1-p(Y = 1)] = \beta g+\varphi X$, where β is the baseline parameter, g is the susceptibility type and X is the covariate vector.

Analyses were performed by estimating the following parameters: type frequencies Ψ_u (u = AA, AB, BB): if the type frequencies were in Hardy-Weinberg equilibrium proportions, they were defined in terms of q_A (frequency of allele A); transmission probabilities τ_u (the probability that a parent of type u transmits allele A to an offspring: under Mendelian transmission, $\tau_{AA} = 1$, $\tau_{AB} = 0.5$, and $\tau_{BB} = 0$); and baseline parameter β, which can be sex dependent and/or type dependent. Sporadic/environmental and genetic models that were considered in assessing type of familial association and possible evidence of transmission of major effect are shown in Table 2.

Every model was tested against the likelihood of the general (unrestricted) model, in which all parameters were unrestricted and allowed to fit the empirical data. The estimated model hypotheses of transmission were as follows: major gene type, Mendelian dominant, Mendelian recessive, Mendelian additive, random environmental effect, codominant, and no transmission (Table 2). A likelihood ratio test (LRT) was used to test the significance of the departure from a specified null hypothesis model using the asymptotic properties of the LRT distributed as chi-square distribution with degrees of freedom equal to the difference in the number of parameters estimated in both models. Using this test, a significant chi-square test indicates that the submodel tested can be rejected at the given alpha level, which means the hypothesized model does not fit the data. Models were also compared using Akaike's information criterion (AIC), which is defined as AIC = $-2 \ln L + 2x$ (number of parameters estimated). A lower value of AIC represents a better fitting model.

3. Results

In this study, 127 late-onset diseases and 83 early-onset families were examined. The general statistics of the pedigrees are disclosed in Table 1. The mean pedigree size and standard deviation as well as the total number of relative pairs were obtained in order to calculate the prevalence for AD, polyautoimmunity, and MAS as main traits. Analyses were restricted to FDR. 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.

In total 716 and 443 individuals were included for the analyses, for late-onset and early-onset families, respectively (Table 1). Late-onset families included 37% males and 63% females while early-onset presented 51% males and 49% females. Moreover, females represented the most affected ones in late-onset families while in early-onset the ratio of the affected was close to 1 : 1 (male : female). In early-onset

Table 3: Familial aggregation (λ_R) of autoimmune disease (AD), polyautoimmunity, and multiple autoimmune syndrome (MAS) in late-onset and early-onset families.

Type of family	Pairs of relatives	Total pairs	Pairs	K (%)	$\lambda_R = K_R/K_{HI}$	$\lambda_R = K_R/K_{pop}$
Late-onset	AD			K_{AD}	λ_{HI}	λ_{pop}
	Parent/offspring	876	55/190/208	6.28	4.76	2.51
	Sibling/sibling	706	86/267/353	12.1	13.39	4.87
	Sister/sister	336	67/92/177	19.9	21.91	7.98
	Brother/brother	64	0/44/20	0.00	0.00	0.00
	Brother/sister	306	19/131/156	6.21	6.82	2.48
Late-onset	Polyautoimmunity			K_{PolyAD}	λ_{HI}	λ_{pop}
	Parent/offspring	876	8/333/112	0.91	0.69	0.37
	Sibling/sibling	706	23/450/233	3.26	3.58	1.30
	Sister/sister	336	20/181/135	5.95	6.54	2.38
	Brother/brother	64	0/59/5	0.00	0.00	0.00
	Brother/sister	306	3/210/93	0.98	1.08	0.39
Late-onset	MAS			K_{MAS}	λ_{HI}	λ_{pop}
	Parent/offspring	876	1/403/49	0.11	0.09	0.05
	Sibling/sibling	706	4/581/121	0.57	0.62	0.23
	Sister/sister	336	3/260/73	0.89	0.98	0.36
	Brother/brother	64	0/60/4	0.00	0.00	0.00
	Brother/sister	306	1/261/44	0.33	0.36	0.13
Early-onset	AD			K_{AD}	λ_{HI}	λ_{pop}
	Parent/offspring	498	9/199/155	1.81	1.37	0.72
	Sibling/sibling	245	9/130/106	3.67	4.04	1.47
	Sister/sister	61	3/30/28	4.92	5.40	1.97
	Brother/brother	60	2/33/25	3.33	3.66	1.33
	Brother/sister	120	4/67/53	3.33	3.66	1.33
Early-onset	Polyautoimmunity/MAS			K_{MAS}	λ_{HI}	λ_{pop}
	Parent/offspring	498	0/361/2	0.00	0.00	0.00
	Sibling/sibling	245	0/244/1	0.00	0.00	0.00
	Sister/sister	61	0/61/0	0.00	0.00	0.00
	Brother/brother	60	0/60/0	0.00	0.00	0.00
	Brother/sister	120	0/123/1	0.00	0.00	0.00

[a] Affected/unaffected/discordant pairs.

[*] K_{AD}, K_{PolyAD}, and K_{MAS} = prevalence for AD, polyautoimmunity, and MAS, respectively. K_{HI} = prevalence for AD in healthy individual's pedigrees as previously reported (K_{PO} = 1.32%; $K_{S/S}$ = 0.91%) [21]. K_{pop} = chosen prevalence for the general population. Recurrent risk ratio ($\lambda_R = K_R/(K_{HI}$ or $K_{pop})$), where R is the specific relative pair used (P/O = parent/offspring; SIB = sibling/sibling). The chosen population prevalence (K) for AD was considered as 25/1000 individuals [21]. Prevalence is given in percentages.

families, there was only one individual presenting with MAS among the 102 affected individuals.

3.1. Familial Aggregation (λ_R). The distribution of relationship types and total number of study subjects included in this study is presented in Table 3. No two probands belonged to the same family. Pairs of relatives discordant or concordant for AD, polyautoimmunity, and MAS were calculated in order to examine the family aggregation. Overall, the data is composed of 876 parent-offspring pairs and 706 different sib-pairs broken down to sister-sister (n = 336), sister-brother (n = 64), and brother-brother (n = 306) pairs (Table 3).

The prevalence of AD, polyautoimmunity, and MAS for each pair of relatives (parent/offspring [P/O], sibling/sibling [S/S]) is disclosed in Table 3. Previously reported prevalence values for familial pairs for AD in healthy individuals were taken into account for the examination of aggregation (K_{PO} = 1.32%; $K_{S/S}$ = 0.91%) [21]. Also, using a putative chosen prevalence for all AD taken together as trait (K_{pop} = 2.5%), λ_R were calculated (Table 3). Values supporting familial aggregation ($\lambda_R > 1.0$) were observed for AD in late-onset families in P/O (λ_{HI} = 4.76, λ_{pop} = 2.51) and S/S (λ_{HI} = 13.39, λ_{pop} = 4.87) pairs, with the highest familial aggregation within sister-pairs (λ_{HI} = 21.91, λ_{pop} = 7.98). For

Table 4: Parameter estimates from segregation analyses of early-onset families. For details in each model check Table 2. AD: autoimmune disease; PolyAD: polyautoimmunity; MAS: multiple autoimmune syndrome; ND: model not able to maximize.

Model/parameter	β_{AA}	β_{AB}	β_{BB}	q_A	ρ_{SS}	τ_{AA}	τ_{AA}	τ_{AA}	Sex	$-2\ln(L)$	d.f.	p value	AIC
AD													
Random environmental	—	—	—	0.19	1.43	q_A	q_A	q_A	2.36	698.079	3	<0.05	708.079
Dominant	1.21	β_{AA}	−109	0.05	0.00	1.00	0.50	0.00	2.16	707.994	5	<0.05	**715.994**
Dominant multifactorial	1.00	β_{AA}	−1.16	0.07	−0.06	1.00	0.50	0.00	2.19	706.492	4	<0.05	716.492
Recessive	−1.09	β_{BB}	1.21	0.95	0.00	1.00	0.50	0.00	2.16	707.994	5	<0.05	**715.994**
Recessive multifactorial	−1.15	β_{BB}	1.33	0.94	−0.05	1.00	0.50	0.00	2.24	706.653	4	<0.05	716.653
Codominant	−33.00	1.41	−1.21	0.06	0.00	1.00	0.50	0.00	2.35	707.529	5	<0.05	717.529
Additive	1.94	0.38	−1.18	0.10	−0.09	1.00	0.50	0.00	2.08	706.956	5	<0.05	716.956
Mayor locus	−0.73	1.54	−2.07	0.01	0.00	0.56	0.00	1.00	2.14	667.079	1	0.52	**679.079**
General transmission	−0.75	1.49	−2.07	0.02	−0.05	0.56	0.00	1.00	2.13	666.871		Ref.	680.871
PolyAD													
Random environmental	—	—	—	0.35	2.27	q_A	q_A	q_A	2.30	491.607	3	<0.05	501.607
Dominant	1.50	β_{AA}	−2.09	0.01	0	1	0.5	0	2.10	499.629	5	<0.05	**507.629**
Dominant multifactorial	−0.89	β_{AA}	−2.21	0.08	−0.22	1	0.5	0	1.89	499.127	4	<0.05	509.127
Recessive	−2.09	β_{BB}	1.51	0.99	0	1	0.5	0	2.10	499.629	5	<0.05	**507.629**
Recessive multifactorial	−2.29	β_{BB}	−0.98	0.90	−0.24	1	0.5	0	1.90	499.135	4	<0.05	509.135
Codominant	−48.15	1.44	−2.09	0.01	0	1	0.5	0	2.11	499.614	5	<0.05	509.614
Additive	−0.65	−1.56	−2.47	0.24	−0.27	1	0.5	0	1.85	499.233	5	<0.05	509.233
Mayor gene	−2.02	−0.44	−17.60	0.04	0	0.86	0.00	1.00	2.09	472.191	1	<0.05	**484.191**
General transmission	−66.10	−1.05	−3.47	0.00	1.86	1.00	0.00	0.39	1.86	459.356		Ref.	471.356
MAS													
Random environmental	—	—	—	0.51	2.65	q_A	q_A	q_A	2.73	286.846	3	<0.05	**296.846**
Dominant	ND	β_{AA}	ND	ND	0	1.00	0.50	0.00	ND	ND	5		
Dominant multifactorial	−2.27	β_{AA}	−4.32	0.25	−0.06	1.00	0.50	0.00	2.25	286.875	4	<0.05	296.875
Recessive	ND	β_{BB}	ND	ND	0	1.00	0.50	0.00	ND	ND	5		
Recessive multifactorial	−4.84	β_{BB}	−2.28	0.72	−0.05	1.00	0.50	0.00	0.72	286.856	4	<0.05	296.856
Codominant	−2.27	−2.27	−4.66	0.27	0	1.00	0.50	0.00	−0.98	286.838	5	<0.05	298.838
Additive	−2.27	−4.14	−6.07	0.66	−0.97	1.00	0.50	0.00	2.25	287.122	5	<0.05	297.122
Mayor gene	24.70	−14.29	−18.36	0.00	0	0.00	0.00	0.15	37.20	271.525	1	<0.05	**281.525**
General transmission	−152.53	−3.36	−1.91	0.96	4.42	0.42	0.12	0.99	1.95	260.304		Ref.	276.304

polyautoimmunity, familial aggregation was not observed for P/O pairs but for S/S pairs ($\lambda_{HI} = 3.58$, $\lambda_{pop} = 1.30$). In early-onset families, familial aggregation was observed for AD in P/O ($\lambda_{HI} = 1.37$) and in S/S ($\lambda_{HI} = 4.04$, $\lambda_{pop} = 1.47$). No aggregation for MAS was observed in any pair of relatives.

3.2. Segregation Analysis. The parameter estimates and test statistics from the segregation analyses for late- and early-onset families for AD, polyautoimmunity, and MAS are presented in Tables 4 and 5, respectively.

To determine support for familial or residual association in the data, initially we compared four no-transmission models, each having different type of familial association, to inspect whether the sibling (S) correlation equals the parent-offspring correlation (FO and/or MO, F: father, M: mother, and O: offspring). Four no major models were fitted and compared; each, respectively, assumed (1) ρFO; ρMO; ρSS-free; (2) ρFO = ρMO, ρSS-free; (3) ρFO = ρMO = ρSS; and (4) ρFO = ρMO = ρSS = 0 (the no multifactorial component model). ρFM was assumed to be 0 for all models. The model where both parent-offspring and sibling residual associations are equal (i.e., ρFO = ρMO = ρSS) fitted the data better than any of the other three models for AD, polyautoimmunity, and MAS for both late- and early-onset families (results not shown), thereby providing support for the existence of familial association in the data and inclusion and estimation of familial association parameters in the subsequent models. To determine whether sex should be included in the segregation models, two nontransmission models were initially fitted, one including the covariate and the other not, and then compared by AIC. Results showed that including sex as a covariate in the models allowed better model fitting (data not shown).

The hypothesis of no major gene was tested by comparing the random environmental (Model 1) and general transmission model (Model 9) (Table 2). The random transmission model was rejected in late-onset disease families, supporting the existence of a major gene in AD ($p < 0.05$, AIC = 708.08), polyautoimmunity ($p < 0.05$, AIC = 501.61), and MAS ($p < 0.05$, AIC = 296.46) (Table 4), while in early-onset

Table 5: Parameter estimates from segregation analyses of early-onset families. AD: autoimmune disease. For details in each model check Table 2.

Model/parameter	β_{AA}	β_{AB}	β_{BB}	q_A	ρ_{SS}	τ_{AA}	τ_{AA}	τ_{AA}	Sex	$-2\ln(L)$	d.f.	p value	AIC
AD													
Random environmental	—	—	—	0.01	−0.83	q_A	q_A	q_A	−0.02	426.292	3	0.55	438.292
Dominant	−1.05	β_{AA}	−1.05	0.02	0	1	0.5	0	−0.03	451.220	5	<0.05	459.22
Dominant multifactorial	−1.99	β_{AA}	−1.05	0.08	0.01	1	0.5	0	0.01	441.228	4	<0.05	451.228
Recessive	−1.07	β_{BB}	−1.05	0.00	0	1	0.5	0	−0.03	451.220	5	<0.05	459.22
Recessive multifactorial	−2.80	β_{BB}	−1.04	0.32	−0.53	1	0.5	0	0.01	440.46	4	<0.05	450.46
Codominant	−2.78	−1.05	−1.08	0.29	0	1	0.5	0	0.01	440.408	5	<0.05	452.408
Additive	−1.17	−1.17	−1.17	0.10	−0.48	1	0.5	0	0.01	441.265	5	<0.05	451.265
Mayor gene	115.3	21.2	−2.68	0.00	0	0.3	0.0	0.1	0.54	400.587	1	<0.05	412.587
General transmission	−9.57	−0.71	−0.91	0.32	−0.84	0.20	0.33	0.34	−0.005	427.342	0		443.342

families the model could not be rejected ($p = 0.55$, AIC = 438.29) (Table 5). Subsequently, the major gene hypothesis was further tested by comparing the major gene only model (Model 8) and the general transmission model (Model 9) (Table 2). For this comparison, the hypothesis for the major gene was rejected only for AD in late-onset families ($p < 0.05$, AIC = 679.08) (Table 4), while it was not rejected for late-onset families when taking polyautoimmunity and MAS as main traits, as well as in early-onset families for AD (Table 5). Of note, for early-onset families due to low frequency of polyautoimmunity and MAS, only models for AD as a main trait were estimated.

After having procured evidence for the segregation of major gene(s) in late-onset families with AD as the main trait and not for polyautoimmunity and MAS for late-onset and for AD in early-onset families, the hypothesis of Mendelian transmission was tested by comparing the Mendelian proposed models (Models 2, 4, 6, and 8) with the general transmission model (Model 9) (Table 2). Dominant, recessive, codominant, and additive Mendelian transmission models were rejected for late-onset families when taking AD as a trait. All the same, when a multifactorial/polygenic parameter was added to the dominant and recessive Mendelian models (Models 3 and 5, resp.) and compared with the Mendelian counterpart without the multifactorial component, no change in the rejection of the models was observed (Table 4).

4. Discussion

The commonality between ADs is the damage to tissues and organs arising from the loss of tolerance and in most cases a gender imbalance [46]. Research generally focuses on a single disease, although autoimmune phenotypes could represent pleiotropic outcomes of nonspecific disease genes underlying similar immunogenetic mechanisms [47]. While it is apparent that multiple cases of a single disease cluster within families [4], more striking are the individuals in those families afflicted with multiple ADs [3].

This report presents the familial aggregation and segregation analyses of AD, polyautoimmunity, and MAS in Colombian families. We have analyzed 210 families (i.e., 127 late-onset diseases and 83 early-onset ones) in Table 1, for which a total of 716 and 443 individuals were analyzed (Table 1). Each pedigree was ascertained through an affected proband fulfilling the inclusion criteria presented in Section 2. This study is restricted and takes into account AD, polyautoimmunity, and MAS as main traits presented in the recruited families (Figure 1). The recruited families were divided into two types of family given by the pathology presented in the proband (i.e., early-onset families are constituted mainly by T1D probands and late-onset families by AD known to develop later in life). Results show differences between late- and early-onset families regarding their age, age of onset, and sex distribution, which is expected given the particular and specific autoimmune disorder prevalence (Table 1, Figure 1).

Analyses of familial aggregation treat the family like any other unit of clustering. In addressing whether there is phenotypic aggregation within families, no attempt is made to determine the cause of any aggregation [48]. The observation and portrayal of familial autoimmunity and the outline of MAS have put aside the environmental aggregation and given a greater value towards the common/rare genetic component for diverse autoimmune phenotypes with a generally common background [4]. When considering the familial aggregation of AD, polyautoimmunity, and MAS for both types of families, values supporting the aggregation of AD in late- and early-onset families for P/O and S/S pairs, with the highest aggregation observed between sister-pairs of late-onset families, were observed (Table 3). For polyautoimmunity as a trait only aggregation was observed between S/S pairs in late-onset families. No familial aggregation for MAS was observed for any type of family. This suggests and confirms that polyautoimmunity and MAS are not AD independent traits and that gender, age, and age of onset represent factors that define and allow the study of the dynamics of the traits within the familial group.

Segregation analyses help to assess the possible genetic mode of segregation of a trait by consideration of relevant hypothesis-based mathematical models. Findings from segregation analyses are often used to formulate tailored research hypotheses for the trait under investigation and/or to decide the type of investigative effort to be put forward. This study was carried out to assess types of familial dependence in AD, polyautoimmunity, and MAS to investigate possible evidence

of transmission of major gene(s) and to determine the best mode of transmission for such major gene(s). The presented analyses indicate evidence for the familial transmission of major gene(s) with no clear discernible classical known Mendelian transmission in late-onset families when AD is taken as the main trait, while for polyautoimmunity and MAS familial transmission fails to be demonstrated. In early-onset families analyses did not demonstrate a major gene effect but a random environmental model explaining the presence of the phenotypes in the families. These results thus provide evidence for the genetic role in the etiology of AD in late-onset families by showing support for major gene(s) mode of segregation of susceptibility to AD, while for the early-onset families and perhaps by their relatively young status eludes a clear picture of autoimmunity segregation and aggregation in these families.

Previous segregation analyses have proposed models in families with more than one member affected by autoimmune hemolytic anemia and chronic thrombocytopenic purpura compatible with a Mendelian dominant trait [49]. In African Americans [50, 51] and EA [52] SLE families, presenting FAD, a dominant inheritance is reported, while in Chinese families segregation analyses describe a polygenetic model and major gene model, suggesting a polygenetic multifactorial disease [53]. Other analyses in VIT for Chinese families suggest a dominant inheritance model [54], while other reports suggest a non-Mendelian pattern supporting a multifactorial, polygenic inheritance [38]; even so other models describe a major dominant gene and the existence of strong environmental effects acting on a recessive genotype [25]. More generally, a Mendelian dominant genetic inheritance is proposed in many ADs, like SS [55] and T1D [56], while segregation is better explained by either dominant or codominant or polygenic models in APS [57], RA [26], and idiopathic inflammatory myopathies [58]. Others suggest that several major ADs result from pleiotropic effects of a single major gene on a polygenic background [26]. Finally, in traits such as MS segregation results are indeterminate and cannot be explained by a genetic model [59].

5. Conclusions

Overall, aggregation and segregation analyses in Colombian families enriched by autoimmunity as a trait show how ADs, polyautoimmunity, and MAS are not independent entities. Familial aggregation for ADs was observed between parents and offspring as well as in sibling pairs in late-onset families, while aggregation for polyautoimmunity and MAS was lesser given by the fact that both traits represent a more complex etiology with lower prevalence but still a common autoimmunity background. Segregation analyses were not able to discern a Mendelian transmission model but still suggested major gene(s) transmission for AD in late-onset families, while for early-onset families a stochastic model was suggested. Thus, a clinical defined individual AD, defined by symptoms and signs, might not be completely juxtaposed to the AD trait defined by environment and genetics, which makes the task to define and untangle disease mechanisms even more difficult. Last but not least, to further study and describe the familial dynamics of two or more cluster ADs, approaches such as familial coaggregation might find their place towards the exploration of common familial factors on top of studies taking into account AD, polyautoimmunity, and MAS as a trait in order to disentangle the common/rare genetic landscape of autoimmunity.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgments

This work was supported by Universidad del Rosario and Colciencias (122254531722), Bogotá, Colombia. Some of the results of this paper were obtained by using the software package S.A.G.E., which was supported by a US Public Health Service Resource Grant (RR03655) from the National Center for Research Resources.

References

[1] D. L. Jacobson, S. J. Gange, N. R. Rose, and N. M. H. Graham, "Epidemiology and estimated population burden of selected autoimmune diseases in the United States," *Clinical Immunology and Immunopathology*, vol. 84, no. 3, pp. 223–243, 1997.

[2] S. Sloka, "Observations on recent studies showing increased co-occurrence of autoimmune diseases," *Journal of Autoimmunity*, vol. 18, no. 3, pp. 251–257, 2002.

[3] J.-M. Anaya, L. Gómez, and J. Castiblanco, "Is there a common genetic basis for autoimmune diseases?" *Clinical and Developmental Immunology*, vol. 13, no. 2–4, pp. 185–195, 2006.

[4] J.-M. Anaya, R. Corena, J. Castiblanco, A. Rojas-Villarraga, and Y. Shoenfeld, "The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity," *Expert Review of Clinical Immunology*, vol. 3, no. 4, pp. 623–635, 2007.

[5] M. Blank and M. E. Gershwin, "Autoimmunity: from the mosaic to the kaleidoscope," *Journal of Autoimmunity*, vol. 30, no. 1-2, pp. 1–4, 2008.

[6] A. Rojas-Villarraga, J. Amaya-Amaya, A. Rodriguez-Rodriguez, R. D. Mantilla, and J.-M. Anaya, "Introducing polyautoimmunity: secondary autoimmune diseases no longer exist," *Autoimmune Diseases*, vol. 2012, Article ID 254319, 9 pages, 2012.

[7] J.-M. Anaya, J. Castiblanco, A. Rojas-Villarraga et al., "The multiple autoimmune syndromes. A clue for the autoimmune tautology," *Clinical Reviews in Allergy and Immunology*, vol. 43, no. 3, pp. 256–264, 2012.

[8] J.-M. Anaya, A. Rojas-Villarraga, and M. García-Carrasco, "The autoimmune tautology: from polyautoimmunity and familial autoimmunity to the autoimmune genes," *Autoimmune Diseases*, vol. 2012, Article ID 297193, 2 pages, 2012.

[9] J. Cárdenas-Roldán, A. Rojas-Villarraga, and J.-M. Anaya, "How do autoimmune diseases cluster in families? A systematic review and meta-analysis," *BMC Medicine*, vol. 11, no. 1, article 73, 2013.

[10] S. E. Baranzini, "The genetics of autoimmune diseases: a networked perspective," *Current Opinion in Immunology*, vol. 21, no. 6, pp. 596–605, 2009.

[11] A. Zhernakova, C. C. van Diemen, and C. Wijmenga, "Detecting shared pathogenesis from the shared genetics of immune-related diseases," *Nature Reviews Genetics*, vol. 10, no. 1, pp. 43–55, 2009.

[12] J. M. Anaya, "Common mechanisms of autoimmune diseases (the autoimmune tautology)," *Autoimmunity Reviews*, vol. 11, no. 11, pp. 781–784, 2012.

[13] W. B. Bias, J. D. Reveille, T. H. Beaty, D. A. Meyers, and F. C. Arnett, "Evidence that autoimmunity in man is a Mendelian dominant trait," *American Journal of Human Genetics*, vol. 39, no. 5, pp. 584–602, 1986.

[14] A. Wandstrat and E. Wakeland, "The genetics of complex autoimmune diseases: non-MHC susceptibility genes," *Nature Immunology*, vol. 2, no. 9, pp. 802–809, 2001.

[15] P. Invernizzi, "Future directions in genetic for autoimmune diseases," *Journal of Autoimmunity*, vol. 33, no. 1, pp. 1–2, 2009.

[16] A. Goris and A. Liston, "The immunogenetic architecture of autoimmune disease," *Cold Spring Harbor Perspectives in Biology*, vol. 4, no. 3, 2012.

[17] L. F. Barcellos, B. B. Kamdar, P. P. Ramsay et al., "Clustering of autoimmune diseases in families with a high-risk for multiple sclerosis: a descriptive study," *The Lancet Neurology*, vol. 5, no. 11, pp. 924–931, 2006.

[18] E. C. Somers, S. L. Thomas, L. Smeeth, and A. J. Hall, "Autoimmune diseases co-occurring within individuals and within families: a systematic review," *Epidemiology*, vol. 17, no. 2, pp. 202–217, 2006.

[19] L. Michou, A.-C. Rat, S. Lasbleiz, T. Bardin, and F. Cornélis, "Prevalence and distribution of autoimmune diseases in 368 rheumatoid arthritis families," *Journal of Rheumatology*, vol. 35, no. 5, pp. 790–796, 2008.

[20] D. Alarcón-Segovia, M. E. Alarcón-Riquelme, M. H. Cardiel et al., "Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort," *Arthritis & Rheumatism*, vol. 52, no. 4, pp. 1138–1147, 2005.

[21] J.-M. Anaya, G. J. Tobon, P. Vega, and J. Castiblanco, "Autoimmune disease aggregation in families with primary Sjögren's syndrome," *Journal of Rheumatology*, vol. 33, no. 11, pp. 2227–2234, 2006.

[22] J.-M. Anaya, J. Castiblanco, G. J. Tobón et al., "Familial clustering of autoimmune diseases in patients with type 1 diabetes mellitus," *Journal of Autoimmunity*, vol. 26, no. 3, pp. 208–214, 2006.

[23] S. Prahalad, E. S. Shear, S. D. Thompson, E. H. Giannini, and D. N. Glass, "Increased prevalence of familial autoimmunity in simplex and multiplex families with juvenile rheumatoid arthritis," *Arthritis and Rheumatism*, vol. 46, no. 7, pp. 1851–1856, 2002.

[24] J. M. Anaya, Y. Shoenfeld, P. A. Correa, M. Garcia-Carrasco, and M. R. Cervera, *Autoimmunity and Autoimmune Disease*, Edited by J. M. Anaya, CIB, Medellin, Colombia, 1st edition, 2005.

[25] M. Arcos-Burgos, E. Parodi, M. Salgar et al., "Vitiligo: complex segregation and linkage dissequilibrium analyses with respect to microsatellite loci spanning the HLA," *Human Genetics*, vol. 110, no. 4, pp. 334–342, 2002.

[26] J.-P. Lin, J. M. Cash, S. Z. Doyle et al., "Familial clustering of rheumatoid arthritis with other autoimmune diseases," *Human Genetics*, vol. 103, no. 4, pp. 475–482, 1998.

[27] J.-M. Anaya, "The autoimmune tautology," *Arthritis Research & Therapy*, vol. 12, no. 6, article 147, 2010.

[28] N. R. Rose, "Autoimmune diseases: tracing the shared threads," *Hospital Practice*, vol. 32, no. 4, pp. 147–154, 1997.

[29] M. Lorber, M. E. Gershwin, and Y. Shoenfeld, "The coexistence of systemic lupus erythematosus with other autoimmune diseases: the kaleidoscope of autoimmunity," *Seminars in Arthritis and Rheumatism*, vol. 24, no. 2, pp. 105–113, 1994.

[30] R. Priori, E. Medda, F. Conti et al., "Familial autoimmunity as a risk factor for systemic lupus erythematosus and vice versa: a case-control study," *Lupus*, vol. 12, no. 10, pp. 735–740, 2003.

[31] "Report of the expert committee on the diagnosis and classification of diabetes mellitus," *Diabetes Care*, vol. 26, no. 11, pp. 3160–3167, 2003.

[32] G. J. Tobon, A. Arango, V. Abad et al., "Clinical and immunological characteristics of type 1 diabetes mellitus in a northwestern Colombian population," *Diabetes Research and Clinical Practice*, vol. 72, no. 2, pp. 170–175, 2006.

[33] M. Ordonez, "The family at the end of the century in Colombia," *Profamilia*, vol. 16, no. 31, pp. 11–18, 1998.

[34] V. Gutierrez de Pineda, "Familia extensa," in *Estructura, función y cambio de la familia colombiana*, pp. 338–359, Universidad de Antioquía, Medellin, Colombia, 2nd edition, 1999.

[35] R. C. Team, *R: A Language and Environment for Statistical Computing*, 2014.

[36] S.A.G.E., Statistical Analysis for Genetic Epidemiology, Release 6.3, 2012, http://darwin.cwru.edu.

[37] N. Risch, "Linkage strategies for genetically complex traits. I. Multilocus models," *The American Journal of Human Genetics*, vol. 46, no. 2, pp. 222–228, 1990.

[38] A. Alkhateeb, P. R. Fain, A. Thody, D. C. Bennett, and R. A. Spritz, "Epidemiology of vitiligo and associated autoimmune diseases in Caucasian probands and their families," *Pigment Cell Research*, vol. 16, no. 3, pp. 208–214, 2003.

[39] S. J. Bowman, G. H. Ibrahim, G. Holmes, J. Hamburger, and J. R. Ainsworth, "Estimating the prevalence among Caucasian women of primary Sjögren's syndrome in two general practices in Birmingham, UK," *Scandinavian Journal of Rheumatology*, vol. 33, no. 1, pp. 39–43, 2004.

[40] A. Firooz, A. Mazhar, and A. R. Ahmed, "Prevalence of autoimmune diseases in the family members of patients with pemphigus vulgaris," *Journal of the American Academy of Dermatology*, vol. 31, no. 3, part 1, pp. 434–437, 1994.

[41] W. Gatling, S. Budd, D. Walters, M. A. Mullee, J. R. Goddard, and R. D. Hill, "Evidence of an increasing prevalence of diagnosed diabetes mellitus in the Poole area from 1983 to 1996," *Diabetic Medicine*, vol. 15, no. 12, pp. 1015–1021, 1998.

[42] M. C. Hochberg, "The application of genetic epidemiology to systemic lupus erythematosus," *Journal of Rheumatology*, vol. 14, no. 5, pp. 867–869, 1987.

[43] E. R. Senna, A. L. P. De Barros, E. O. Silva et al., "Prevalence of rheumatic diseases in Brazil: a study using COPCORD approach," *Journal of Rheumatology*, vol. 31, no. 3, pp. 594–597, 2004.

[44] J. C. Shank, "A study of thyroid disease in family practice," *Journal of Family Practice*, vol. 3, no. 3, pp. 247–253, 1976.

[45] W. M. G. Tunbridge, D. C. Evered, R. Hall et al., "The spectrum of thyroid disease in a community: the Whickham survey," *Clinical Endocrinology*, vol. 7, no. 6, pp. 481–493, 1977.

[46] S. M. Mariani, "Genes and autoimmune diseases—a complex inheritance," *Medscape General Medicine*, vol. 6, no. 4, article 18, 2004.

[47] T. J. Vyse and J. A. Todd, "Genetic analysis of autoimmune disease," *Cell*, vol. 85, no. 3, pp. 311–318, 1996.

[48] P. R. Burton, M. D. Tobin, and J. L. Hopper, "Key concepts in genetic epidemiology," *The Lancet*, vol. 366, no. 9489, pp. 941–951, 2005.

[49] S. M. Lippman, F. C. Arnett, C. L. Conley, P. M. Ness, D. A. Meyers, and W. B. Bias, "Genetic factors predisposing to autoimmune diseases. Autoimmune hemolytic anemia, chronic thrombocytopenic purpura, and systemic lupus erythematosus," *The American Journal of Medicine*, vol. 73, no. 6, pp. 827–840, 1982.

[50] S. K. Nath, B. Namjou, J. Kilpatrick et al., "A candidate region on 11p13 for systemic lupus erythematosus: a linkage identified in African-American families," *Journal of Investigative Dermatology Symposium Proceedings*, vol. 9, no. 1, pp. 64–67, 2004.

[51] J. A. Kelly, K. Thompson, J. Kilpatrick et al., "Evidence for a susceptibility gene (SLEH1) on chromosome 11q14 for systemic lupus erythematosus (SLE) families with hemolytic anemia," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 18, pp. 11766–11771, 2002.

[52] S. K. Nath, J. A. Kelly, J. Reid et al., "SLEB3 in systemic lupus erythematosus (SLE) is strongly related to SLE families ascertained through neuropsychiatric manifestations," *Human Genetics*, vol. 111, no. 1, pp. 54–58, 2002.

[53] J. Wang, S. Yang, J. J. Chen et al., "Systemic lupus erythematosus: a genetic epidemiology study of 695 patients from China," *Archives of Dermatological Research*, vol. 298, no. 10, pp. 485–491, 2007.

[54] X. Sun, A. Xu, X. Wei et al., "Genetic epidemiology of vitiligo: a study of 815 probands and their families from south China," *International Journal of Dermatology*, vol. 45, no. 10, pp. 1176–1181, 2006.

[55] J. D. Reveille, R. W. Wilson, and T. T. Provost, "Primary Sjogren's syndrome and other autoimmune diseases in families. Prevalence and immunogenetic studies in six kindreds," *Annals of Internal Medicine*, vol. 101, no. 6, pp. 748–756, 1984.

[56] N. Risch, "Segregation analysis incorporating linkage markers. I. Single-locus models with an application to type I diabetes," *The American Journal of Human Genetics*, vol. 36, no. 2, pp. 363–386, 1984.

[57] N. Goel, T. L. Ortel, D. Bali et al., "Familial antiphospholipid antibody syndrome: criteria for disease and evidence for autosomal dominant inheritance," *Arthritis and Rheumatism*, vol. 42, no. 2, pp. 318–327, 1999.

[58] L. R. Ginn, J.-P. Lin, P. H. Plotz et al., "Familial autoimmunity in pedigrees of idiopathic inflammatory myopathy patients suggests common genetic risk factors for many autoimmune diseases," *Arthritis & Rheumatism*, vol. 41, no. 3, pp. 400–405, 1998.

[59] R. W. Haile, L. Iselius, S. E. Hodge, N. E. Morton, and R. Detels, "Segregation and linkage analysis of 40 multiplex multiple sclerosis families," *Human Heredity*, vol. 31, no. 4, pp. 252–258, 1981.

Chapter 3 - Homozygosity Genetic Analysis in Autoimmunity Affected Individuals and Multiplex Autoimmune Disease Families

Castiblanco John, Mantilla Ruben-Dario, Rojas-Villarraga Adriana and Anaya Juan-Manuel. Homozygosity Genetic Analysis in Autoimmunity Affected Individuals and Multiplex Autoimmune Disease Families. Immunome Res 13: 136. doi: 10.4172/17457580.1000136.

Homozygosity Genetic Analysis in Autoimmunity Affected Individuals and Multiplex Autoimmune Disease Families

John Castiblanco[1,2], Ruben Dario Mantilla[1], Adriana Rojas-Villarraga[1] 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.

Corresponding author: Juan-Manuel Anaya Cabrera, MD, PhD

Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24 #63-C-69, Bogota, Colombia.

e-mail: juan.anaya@urosario.edu.co

Key Words: Autoimmunity; Autoimmune disease; Familial autoimmunity; Homozygosity; polyautoimmunity; multiple autoimmune syndrome; late-onset; early-onset.

Abstract

Background: Autoimmune diseases (AD) are responsible for a substantial amount of disability and morbidity worldwide. The commonality between ADs is the damage to tissues and an organ arising from the loss of tolerance and in most cases a gender imbalance. 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 homozygosity in individuals and multiplex families affected with AD using a panel of microsatellite markers.

Methods: This study presented two approaches: (I) a case – control comparison and evaluation on the effect of homozygosity at the genome-wide level, including 453 genotyped unrelated individuals (121 late-onset AD, 79 early-onset, 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. All individuals were treated and recruited at the Center for Autoimmune Diseases Research (CREA) from Medellin and Bogota, Colombia, South America. A total of 372 markers were used in the analysis. The standardized observed homozygosity (S_{OH}) was calculated and the association of homozygosity and autoimmune trait was evaluated. The multipoint model-free linkage analysis was applied by using RELPAL from S.A.G.E v6.3.

Results: The S_{OH} showed significant differences between controls and early-onset individuals, where 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 homozygosity effect showed a 1:1 relation on elevated risk and/or protective effects for 24 markers with marginal significance. The model-free affected pair linkage approach did not provide any suggestive linkage signals for early-, late-onset, polyA or MAS. Marginal signals displayed excess allele sharing for early-onset and MAS, both extreme phenotypes in autoimmunity.

Conclusions: This study presumed autoimmunity as a trait rather than a clinical phenotype and tried to approach AD as a continuous trait presenting extreme phenotypes**.** This report is an 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 corroborate and expand on sample size, marker coverage and their effects.

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 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 overdominant 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).

Likewise, 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 homozygosity in individuals and multiplex families affected with AD using a panel of microsatellite markers 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 (Table 1). Individuals included in this study presented with: (i) at least one AD according to specififc 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).

Table 1. Characteristics of AD affected and healthy control study individuals.

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

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 lupue 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.

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 (26) (Table 1). 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 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.

Table 2. Characteristics of probands and families included in the analysis.

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	4.05 ± 2.21	4.41 ± 2.72	4.14 ± 2.92	4.60 ± 2.08
[Min, Max]	[3,17]	[3,17]	[3,17]	[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

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

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 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).

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 $1x10^{-3}$ to $5x10^{-4}$ (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).

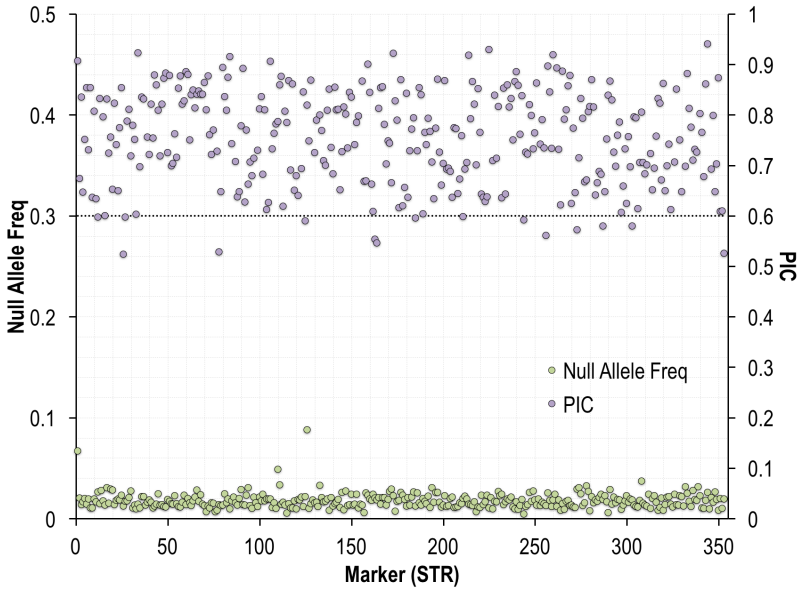
Results

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 16^{th}-17^{th} centuries and flourished in relative isolation until the late 19^{th} 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.

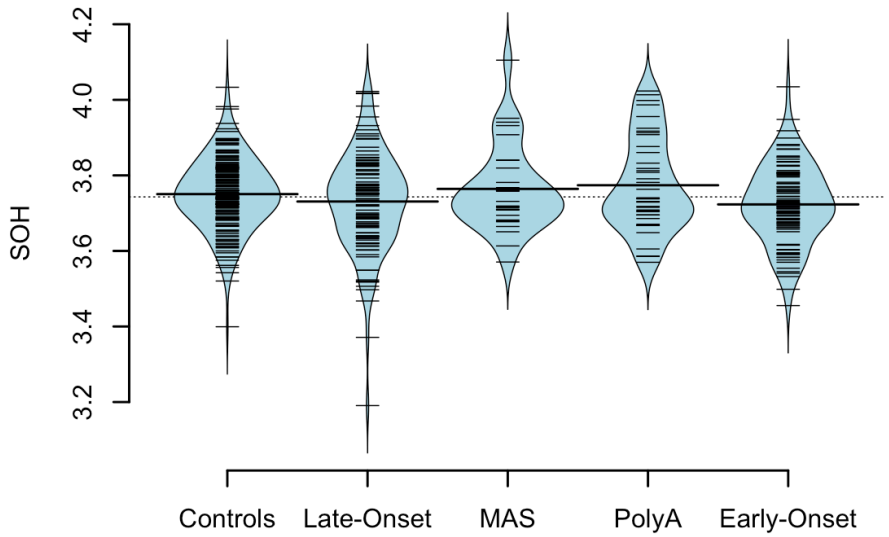
Figure 1. Marker reliability, gene diversity and population relationship indicators. **A.** Calculated null allele frequency (left axis) and polymorphic information content (PIC) (Right axis).



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 Standardized Observed Homozygosity (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 were compared by using Wilcoxon-rank sum test to examine if there were significant differences between AD groups relative to healthy control individuals (Figure 2). 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).

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.



Subsequently, the local homozygosity effect for the genotyped markers 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 showed p-values less than 0.05; however, when correction for multiple comparisons was examined, 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 Benjamin-Hochberg procedure at the 0.05 level.

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.

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	21q22.13	D21S1440	< 0.05	< 0.05	< 0.05	0.08 (0 - 0.54) 0.002

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 4). Moreover, two markers were shared between late-onset, MAS and polyA showed the same directional effect (i.e., D2S2944, D10S1208) and another one was shared between polyA and MAS (i.e., D1S1677) (Table 4). Although the spacing between markers 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.

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

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

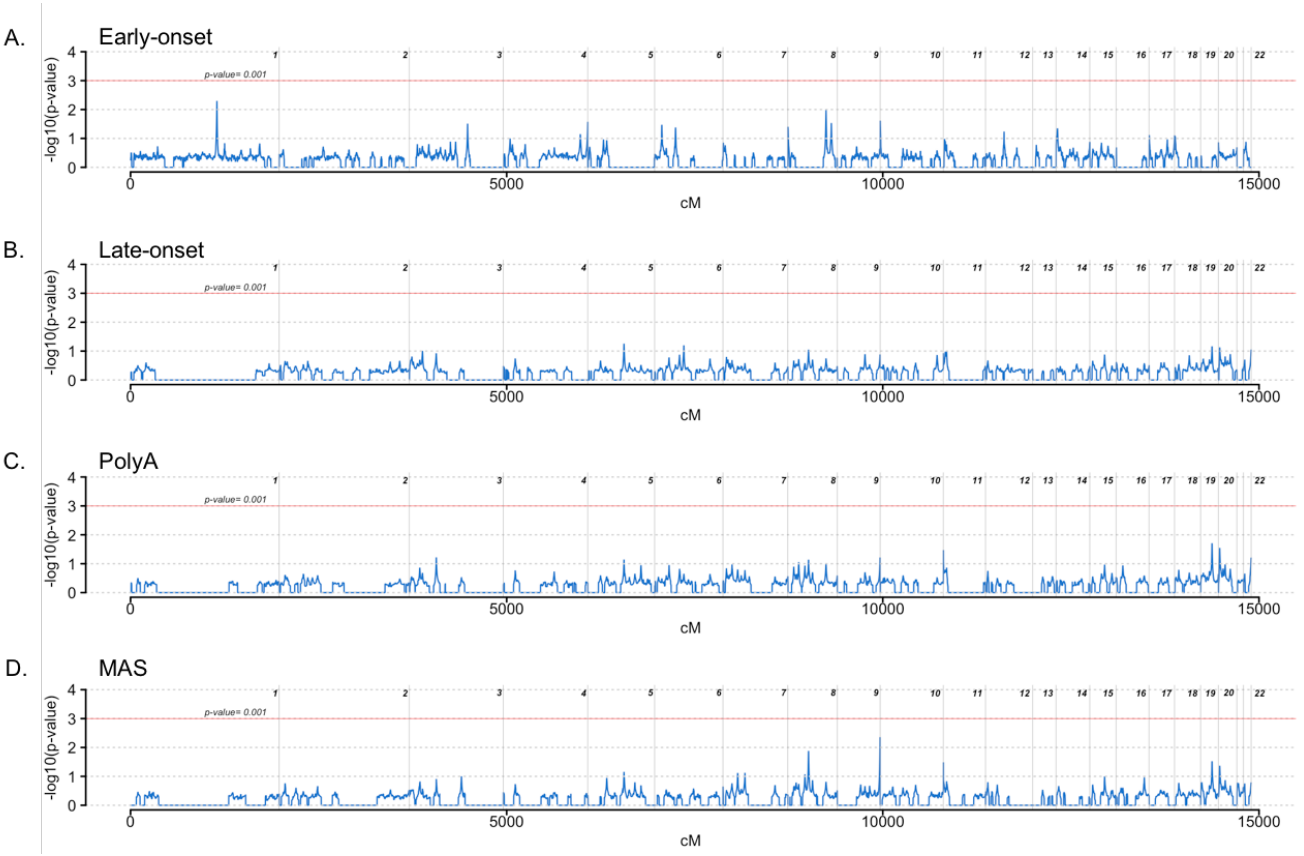
Familial Data and Multipoint Model-free Linkage Analysis

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.

Affected pair results for the non-parametric multipoint linkage analyses implemented in RELPAL for early-, late-onset, polyA and MAS are shown on figure 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 $1x10^{-3}$ to $5x10^{-4}$ [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; these were the highest linkage signals obtained, but present marginally evidence for linkage.

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 of suggestive linkage significance (p-value < 0.001).



Discussion

The commonality between ADs 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 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 homozygosity at the genome-wide level and a model-free affected pair linkage approach to identify IBD loci. The first is a systematic analysis where the status of the individuals per locus is taken into account, standardized and evaluated for the association between homozygosity and AD. The results for S_{OH} showed significant differences between controls and early-onset individuals, where 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 (Figure 3).

Detailed analysis for the local homozygosity effect showed about a 1:1 relation on elevated risk and/or protective effects (Table 3) conferred by the homozygosity status at specific loci

depending on the compared autoimmune trait. On top of this, some of the markers presented a shared component between trait and more interestingly between late-onset, polyA and MAS but not with early-onset. Since correction for multiple comparisons only provided suggestive and marginal significance values, no candidate regions or genes were put forward; no less, this data provides a baseline for future approaches with better coverage and larger sample sizes given by the fact that extreme phenotypes are not as prevalent to study. In general, “Paisa” population for affected and unaffected individuals included in this report showed to be highly diverse.

Homozygosity has been previously examined on a single disease basis for rheumatoid arthritis (RA) (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 countercurrent 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, did not provide any suggestive linkage signals for early-, late-onset, polyA or MAS. Putative/Marginal signals displayed excess allele sharing for early-onset and MAS, both extreme phenotypes in autoimmunity, but their signals warrants caution, although the marker regions have been previously linked at a single disease level (13).

The present study is 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 corroborate and expand on sample size, marker coverage and their effects. Closer inspection of clinical and phenotypic quantitative variants is warranted, as well as inclusion of environmental and clinical available variants. The affected relative pair 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 ADs, involve heterogeneity in disease phenotypes, population and ethnic differences and unavailable statistical and analytical models

Conclusions and Perspectives

Overall, this study presumed autoimmunity as a trait rather than a clinical phenotype and tried to approach AD as a continuous trait presenting 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 exome sequencing in individuals affected with polyA and MAS and their relatives (44, 45).

Acknowledgments

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.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this manuscript.

References

1. Frazer KA, Murray SS, Schork NJ, Topol EJ. Human genetic variation and its contribution to complex traits. Nature reviews Genetics. 2009;10(4):241-51.
2. Elahi E, Kumm J, Ronaghi M. Global genetic analysis. Journal of biochemistry and molecular biology. 2004;37(1):11-27.
3. Mirkin SM. Expandable DNA repeats and human disease. Nature. 2007;447(7147):932-40.
4. Gemayel R, Vinces MD, Legendre M, Verstrepen KJ. Variable tandem repeats accelerate evolution of coding and regulatory sequences. Annual review of genetics. 2010;44:445-77.
5. Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, et al. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. Science. 2001;293(5529):455-62.
6. Khan R, Mittelman D. Rumors of the death of consumer genomics are greatly exaggerated. Genome biology. 2013;14(11):139.
7. Miller JM, Malenfant RM, David P, Davis CS, Poissant J, Hogg JT, et al. Estimating genome-wide heterozygosity: effects of demographic history and marker type. Heredity. 2014;112(3):240-7.
8. Pemberton J. Measuring inbreeding depression in the wild: the old ways are the best. Trends in ecology & evolution. 2004;19(12):613-5.
9. Hansson B, Westerberg L. On the correlation between heterozygosity and fitness in natural populations. Molecular ecology. 2002;11(12):2467-74.
10. Rudan I, Smolej-Narancic N, Campbell H, Carothers A, Wright A, Janicijevic B, et al. Inbreeding and the genetic complexity of human hypertension. Genetics. 2003;163(3):1011-21.
11. Lyons EJ, Amos W, Berkley JA, Mwangi I, Shafi M, Williams TN, et al. Homozygosity and risk of childhood death due to invasive bacterial disease. BMC Med Genet. 2009;10:55.
12. Becker KG, Simon RM, Bailey-Wilson JE, Freidlin B, Biddison WE, McFarland HF, et al. Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. Proc Natl Acad Sci U S A. 1998;95(17):9979-84.
13. Anaya JM, Gomez L, Castiblanco J. Is there a common genetic basis for autoimmune diseases? Clinical & developmental immunology. 2006;13(2-4):185-95.
14. Becker KG. The common variants/multiple disease hypothesis of common complex genetic disorders. Medical hypotheses. 2004;62(2):309-17.
15. Tomer Y, Ban Y, Concepcion E, Barbesino G, Villanueva R, Greenberg DA, et al. 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. 2003;73(4):736-47.
16. Cardenas-Roldan J, Rojas-Villarraga A, Anaya JM. How do autoimmune diseases cluster in families? A systematic review and meta-analysis. BMC medicine. 2013;11:73.
17. Baranzini SE. The genetics of autoimmune diseases: a networked perspective. Current opinion in immunology. 2009;21(6):596-605.
18. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nature reviews Genetics. 2009;10(1):43-55.
19. Bias WB, Reveille JD, Beaty TH, Meyers DA, Arnett FC. Evidence that autoimmunity in man is a Mendelian dominant trait. American journal of human genetics. 1986;39(5):584-602.
20. Wandstrat A, Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol. 2001;2(9):802-9.
21. Invernizzi P. Future directions in genetic for autoimmune diseases. J Autoimmun. 2009;33(1):1-2.

22. Priori R, Medda E, Conti F, Cassara EA, Danieli MG, Gerli R, et al. Familial autoimmunity as a risk factor for systemic lupus erythematosus and vice versa: a case-control study. Lupus. 2003;12(10):735-40.
23. Anaya JM, Tobon GJ, Vega P, Castiblanco J. Autoimmune disease aggregation in families with primary Sjogren's syndrome. J Rheumatol. 2006;33(11):2227-34.
24. Rojas-Villarraga A, Amaya-Amaya J, Rodriguez-Rodriguez A, Mantilla RD, Anaya JM. Introducing polyautoimmunity: secondary autoimmune diseases no longer exist. Autoimmune diseases. 2012;2012:254319.
25. Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care. 2003;26 Suppl 1:S5-20.
26. Tobon GJ, Arango A, Abad V, Garcia J, Cuervo H, Velasquez A, et al. Clinical and immunological characteristics of type 1 diabetes mellitus in a northwestern Colombian population. Diabetes Res Clin Pract. 2006;72(2):170-5.
27. Team RC. R: A Language and Environment for Statistical Computing. R: A Language and Environment for Statistical Computing. v3.1.1 ed2014.
28. S.A.G.E. Statistical Analysis for Genetic Epidemiology, Release 6.3: http://darwin.cwru.edu. 2012.
29. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28(2):298-9.
30. Chapuis MP, Estoup A. Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol. 2007;24(3):621-31.
31. Lander E, Kruglyak L. Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. Nat Genet. 1995;11(3):241-7.
32. Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, Ospina-Duque J, et al. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. American journal of human genetics. 2000;67(5):1287-95.
33. Bravo ML, Valenzuela CY, Arcos-Burgos OM. Polymorphisms and phyletic relationships of the Paisa community from Antioquia (Colombia). Gene Geogr. 1996;10(1):11-7.
34. Mariani SM. Genes and autoimmune diseases - a complex inheritance. MedGenMed. 2004;6(4):18.
35. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. Cell. 1996;85(3):311-8.
36. Anaya JM, Corena R, Castiblanco J, Rojas-Villarraga A, Shoenfeld Y. The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity. Expert Review of Clinical Immunology. 2007;3(4):623-35.
37. Blank M, Gershwin ME. Autoimmunity: from the mosaic to the kaleidoscope. J Autoimmun. 2008;30(1-2):1-4.
38. Anaya JM, Castiblanco J, Rojas-Villarraga A, Pineda-Tamayo R, Levy RA, Gomez-Puerta J, et al. The Multiple Autoimmune Syndromes. A Clue for the Autoimmune Tautology. Clinical reviews in allergy & immunology. 2012.
39. Perez-Fernandez OM, Mantilla RD, Cruz-Tapias P, Rodriguez-Rodriguez A, Rojas-Villarraga A, Anaya JM. Spondyloarthropathies in autoimmune diseases and vice versa. Autoimmune diseases. 2012;2012:736384.
40. Cardenas-Roldan J, Rojas-Villarraga A, Anaya JM. How do autoimmune diseases cluster in families? A systematic review and meta-analysis. BMC medicine. 2013;11(1):73.
41. Sloka S. Observations on recent studies showing increased co-occurrence of autoimmune diseases. J Autoimmun. 2002;18(3):251-7.
42. Yang HC, Chang LC, Liang YJ, Lin CH, Wang PL. A genome-wide homozygosity association study identifies runs of homozygosity associated with rheumatoid arthritis in the human major histocompatibility complex. PloS one. 2012;7(4):e34840.

43. Anaya JM, Ramirez-Santana C, Alzate MA, Molano-Gonzalez N, Rojas-Villarraga A. The Autoimmune Ecology. Front Immunol. 2016;7:139.
44. Johar A, Sarmiento-Monroy JC, Rojas-Villarraga A, Silva-Lara MF, Patel HR, Mantilla RD, et al. Definition of mutations in polyautoimmunity. J Autoimmun. 2016;72:65-72.
45. Johar AS, Mastronardi C, Rojas-Villarraga A, Patel HR, Chuah A, Peng K, et al. Novel and rare functional genomic variants in multiple autoimmune syndrome and Sjogren's syndrome. Journal of translational medicine. 2015;13:173.

Chapter 4 - Ancestry Effect in Colombian Individuals Presenting Autoimmunity, Polyautoimmunity and Multiple Autoimmune Syndrome

Castiblanco John, Mantilla Ruben-Dario , Rojas-Villarraga Adriana and Juan-Manuel Anaya. Ancestry effect in Colombian Individuals presenting Autoimmunity, Polyautoimmunity and Multiple Autoimmune Syndrome. (*Manuscript in submission*)

Ancestry Effect in Colombian Individuals Presenting Autoimmunity, Polyautoimmunity and Multiple Autoimmune Syndrome

John Castiblanco[1,2], Ruben Dario Mantilla[1], Adriana Rojas-Villarraga[1] 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.

Corresponding author: Juan-Manuel Anaya Cabrera, MD, PhD

Center for Autoimmune Diseases Research (CREA), School of Medicine and Health Sciences, Universidad del Rosario, Carrera 24 #63-C-69, Bogota, Colombia.

e-mail: juan.anaya@urosario.edu.co

Key Words: Autoimmunity; autoimmune disease; admixture; ancestry; polyautoimmunity; multiple autoimmune syndrome; late-onset; early-onset.

Abstract

Background: 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. While it is apparent that multiple cases of a single disease cluster within families, more striking are the individuals in those families afflicted with multiple autoimmune diseases.

Most Latin American populations are a diverse genetic collection of Native American, European, African and/or Asian admixture, resulting from varying geographic origins and individual lineages. This study examined whether the ancestry component of Colombian affected individuals is associated with susceptibility/protection to develop an autoimmune disease (i.e., having at least one AD), polyautoimmunity (polyA) (i.e., having at least two ADs) and multiple autoimmune syndrome (MAS) (i.e., having three or more ADs) in reference with publicly available worldwide populations.

Methods: This study included 453 genotyped unrelated individuals (121 late-onset AD, 79 early-onset, 40 PolyA, 30 MAS and 183 healthy control individuals) treated at the Center for Autoimmune Diseases Research (CREA) from Medellin, Colombia, South America. A total of 334 markers were used in the analysis. Population genetic structure was examined in all included Colombian affected and healthy individuals, as well as in the individuals originated from reference populations, assuming three ancestral groups (k=3) (i.e., European, Amerindian and African). Estimation of individual ancestry proportions were obtained with the model-based MCMC Bayesian algorithm implemented in STRUCTURE v2.3.3. The ancestry component effect for the studied traits was compared and examined by logistic regression relative to controls.

Results: All markers analyzed were highly informative with a low null allele frequency making them optimal and reliable for genetic diversity studies. Amerindian and African ancestry significantly differed in late-onset samples (p-value=0.01) and European ancestry significantly differed in MAS samples (p-value=0.02). For late-onset individuals, Amerindian ancestry showed a protective effect, while African ancestry pertained a susceptibility effect. European ancestry showed a protective effect for MAS individuals.

Conclusion: Although the observed ancestry effects warrant further exploration due to the relatively small sample size, this data provides a starting point for approaches using markers that provide higher genome coverage. To our knowledge this would be the first proof of principle approach to examine the ancestry effect for individuals affected with AD (e.g., early- and late-onset) and with more than one AD (e.g., polyA and MAS).

Introduction

Most Latin American populations are a diverse genetic collection of Native American, European, African and/or Asian admixture, resulting from varying geographic origins and individual lineages (1-3). Such context is of particularly interest for genetic epidemiology and the study of variation and distribution of alleles involved in susceptibility to complex diseases (2). The use of individual genetic ancestry estimates for understanding complex disease risk and/or susceptibility is of particular interest in genetic association studies (1, 4-7). The estimation of genomic ancestry at the individual or group level and the use of this information in genotype-disease association studies in place of self-reported ethnicity to measure stratification can be considered by assigning individuals to subpopulations using information from a set of unlinked loci under an admixture model (i.e. Structured association methods) (8).

For Latin America, admixture occurred between the Spanish and Amerindians as a result of the Spanish conquest and colonization of the New World over three centuries ago. In surveys of ancestry estimates of Latino populations from California, Mexico, Brazil, and Colombia, the Latinos from Mexico and California had approximately 50% European ancestry and 40% North American Amerindian, whereas Latinos from Brazil and Colombia had approximately 71% European ancestry. The average African ancestry ranges from approximately 4% in Mexicans to 10–20% in South American Latinos (9). Other surveys of Latino ancestry indicate considerable heterogeneity among regions, with a range of 33–95% European ancestry, 0–58% Native American, and 0–29% West African; this proportions show regional and geographic variation (10), as well as differences in ancestry associated with socioeconomic status (11) .

Autoimmune diseases (ADs) are responsible for a substantial amount of disability and morbidity worldwide. Although their epidemiology varies according to individual conditions, collectively, autoimmune prevalence is at least 5% in the general population and is one of the major causes of premature mortality in young and middle aged women (12). The commonality between ADs is the damage to tissues and organs arising from the loss of tolerance and in most cases a gender imbalance (13). Research generally focuses on a single disease, although autoimmune phenotypes could represent pleiotropic outcomes of non-specific disease genes underlying similar immunogenetic mechanisms (14). While it is apparent that multiple cases of a single disease cluster within families (15), more striking are the individuals in those families afflicted with multiple ADs (16).

A common origin for diverse ADs is sustained by three levels of evidence (16). 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., polyautoimmunity [PolyA] and or Multiple Autoimmune syndrome [MAS]) (15, 17-20), or in the same family (i.e., familial autoimmunity) (21). The second level of evidence refers to known shared pathophysiological mechanisms between ADs. Epidemiological studies have shown correlations among certain ADs, linking epidemiological observations to physiopathological evidence for autoimmune diseases might contribute to our knowledge for the shared etiological and immunogenetic mechanisms (22). The third level of evidence corresponds to the evidence implying common genetic factors (19). The importance of this concept focuses on the probability of having multiple ADs simultaneously in one patient, which goes beyond epidemiologic inferences.

Recent advances in genomics have led to increased understanding of the molecular underpinnings of disease. 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 (23). 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 (24, 25). However, due to their multifactorial and polygenic nature, accompanied by a differential penetrance influenced by environmental factors and genetic heterogeneity among populations (26, 27), untangling of the genetic determinants defining their outcome and onset has proven to be extremely challenging. Likewise, 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 (28). This study is aimed to characterize the population structure of Colombian individuals affected by AD in reference with publicly available worldwide populations and also to examine whether the ancestry component of the affected individuals is associated with susceptibility/protection to develop AD, polyA and/or MAS.

Materials and Methods

Population Samples: All patients were treated at the Center for Autoimmune Diseases Research (CREA) at the University of Rosario in Bogotá and Medellin, Colombia (Table 1). Individuals included in this study presented: (i) with at least one AD according to specififc

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) with polyA and/or MAS. 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 ethnicity criteria for eligibility as the cases with no evidence of AD (Table 1).

Table 1. Characteristics of study individuals.

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

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 lupue 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.

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.* (29), using a standardized questionnaire that incorporates demographics and medical information including a check-point list of 18 ADs (30). 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 (18, 30).

All T1D affected cases were children all of whom fulfilled the diagnostic classification criteria proposed by the American Diabetes Association (ADA) (31), as has been previously described (32) (Table 1).

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 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.

Genetic markers: Genetic markers included in this study were autosomal microsatellites genotyped using screening set 16 at the NHLBI sponsored Mammalian Genotyping Service, Marshfield, Wisconsin. This study included out of study reference populations the marker and population compilation reported by Pemberton *et al.* (33), which includes 645 microsatellite loci with genotypes for 5795 individuals from 267 worldwide populations, this compilation defines subsets of unrelated individuals to be used in studies in which relatedness needs to be clearly characterized.

Statistical and Genetic Data Analysis: Data was managed and stored using the R software v3.1.2 (34) 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. 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 or unaffected 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). Logistic regression was used to assess the effect of genetic ancestry on autoimmune affection while taking into consideration the following covariates: age (as a quantitative variable), sex (male = 0, female = 1) and calculated local ancestry assuming three ancestral populations.

Genetic Markers Characterization and Population Structure Analysis: Individuals with less that 10% of missing genotypes were excluded from analysis for the Colombian and reference populations. 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 (35). 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 (36). Pairwise F_{st} values for populations were calculated by applying 1000 random bootstrap replicates.

In order to visualize the relationship between populations, Nei's D pairwise genetic distances were calculated after making ENA corrections on the allele frequencies by using GENETIX Software v. 4.05 (37). Nei's D genetic distance was used to construct a neighbor-joining tree using the ape package from R.

Structure Analysis: In order to estimate the ancestral estimated proportions, a total of 334 informative markers (AIMS) were chosen by comparing their F_{st} pairwise indexes. The selected AIMS included 195 markers with European/Amerindian $F_{st} \geq 0.25$, 48 markers with African/Amerindian $F_{st} \geq 0.25$ and 91 markers with European/African $F_{st} \geq 025$.

The allotment of genetic ancestral contributions was estimated using the software STRUCTURE v2.3.4 (38, 39). To estimate the ancestral membership proportions, a supervised analysis was performed using prior information on the geographic origin of the reference samples from Africa, Europe and Amerindians (Table 2). The STRUCTURE v2.3.4 runs comprised three replicates of 50,000 burning steps followed by 50,000 Markov Chain Monte Carlo (MCMC) iterations. A tri-hybrid contribution from Amerindians, Europeans and Africans was assumed (i.e., K = 3). The "Use population Information to test for migrants" option was used with the Admixture model. Allele frequencies were correlated and updated using only individuals with POPFLAG = 1.

Table 2. Out of study populations used in this study as reference for ancestry component calculation, geographic coordinates are provided and sample size per each population is provided.

Population Name	Location	Geographic Region	Latitude [Degrees North]	Longitude [Degrees East]	Sample size
Orcadian	Orkney Islands	Europe	59.0	-3.0	16
Adygei	Russia-Caucasus	Europe	44.0	39.0	17
Russian	Russia	Europe	61.0	40.0	25
Basque	France	Europe	43.0	0.0	24
French	France	Europe	46.0	2.0	29
Italian	Italy-Bergamo	Europe	46.0	10.0	13
Sardinian	Italy	Europe	40.0	9.0	28
Tuscan	Italy	Europe	43.0	11.0	8
Piapoco	Colombia	Amerindian	3.0	-68.0	13
Karitiana	Brazil	Amerindian	-10.0	-63.0	24
Surui	Brazil	Amerindian	-11.0	-62.0	21
Maya	Mexico	Amerindian	19.0	-91.0	25
Pima	Mexico	Amerindian	29.0	-108.0	25
Bantu South Africa	South Africa	Africa	-25.6	24.3	8
Bantu Kenya	Kenya	Africa	-3.0	37.0	12
Mandenka	Senegal	Africa	12.0	-12.0	24
Yoruba	Nigeria	Africa	8.0	5.0	25
Biaka Pygmy	Central African Republic	Africa	4.0	17.0	32
Mbuti Pygmy	Congo	Africa	1.0	29.0	15
San	Namibia	Africa	-21.0	20.0	7
Quichean	Guatemala	Amerindian	15.0	-91.0	12
Mixtec	Mexico	Amerindian	17.0	-97.0	40
Zapotec	Mexico	Amerindian	16.0	-97.0	19
Guaymi	Panama	Amerindian	8.5	-82.0	18
Cabecar	Costa Rica	Amerindian	9.5	-84.0	20
Aymara	Chile	Amerindian	-22.0	-70.0	94
Kogi	Colombia	Amerindian	11.0	-74.0	35
Ingano	Colombia	Amerindian	1.0	-77.0	17
Wayuu	Colombia	Amerindian	11.0	-73.0	17
Ticuna	Colombia	Amerindian	-4.0	-70.0	35
Embera	Colombia	Amerindian	7.0	-76.0	11
Waunana	Colombia	Amerindian	5.0	-77.0	20
Arhuaco	Colombia	Amerindian	11.0	-73.8	17

Results

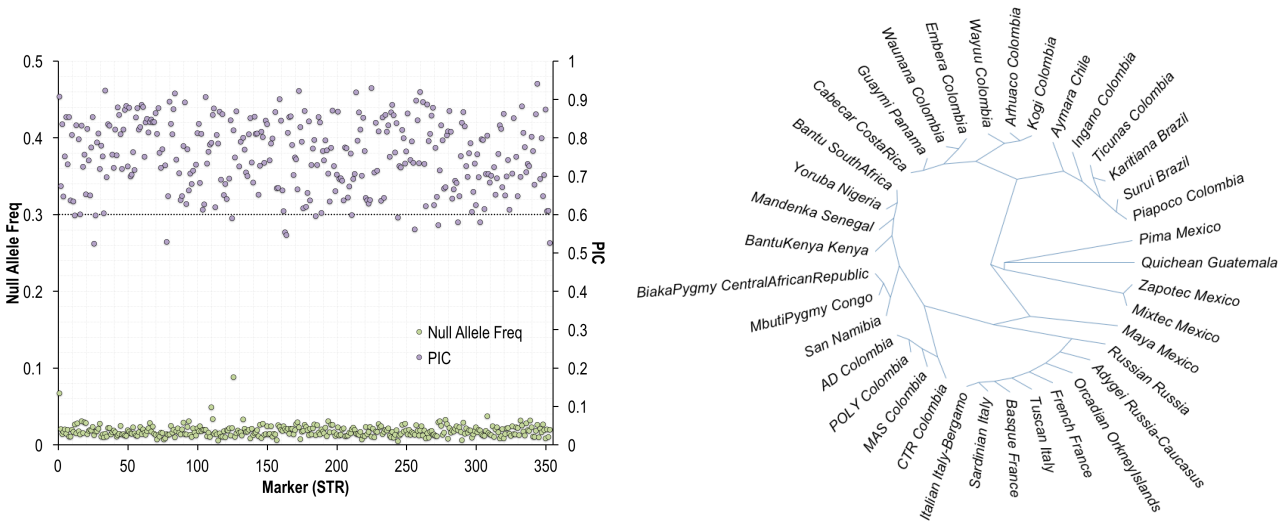
This study included 453 genotyped unrelated individuals (121 late-onset AD, 79 early-onset, 40 PolyA, 30 MAS and 183 healthy control individuals) from Medellin, Colombia South America (Table 1). In addition, 746 individuals were used as out of study controls, as they were compiled and assembled in a single dataset with available genotypes for the Marshfield screening set 16 (40).

A general description of the Colombian samples included in this study is disclosed on Table 1. When early-onset and late-onset individuals age and age of onset were compared, 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 43% males and 57% 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).

To examine the admixture characteristics in Colombian patients and controls, a total of 334 markers were used. These markers had been previously reported for individuals of 37 worldwide populations (40) (Table 2). These 334 markers were selected as AIMS out of a set of 353 initial markers based on a greater than 0.25 fixation index value (F_{st}) obtained by pairwise comparisons calculated between the out of study populations.

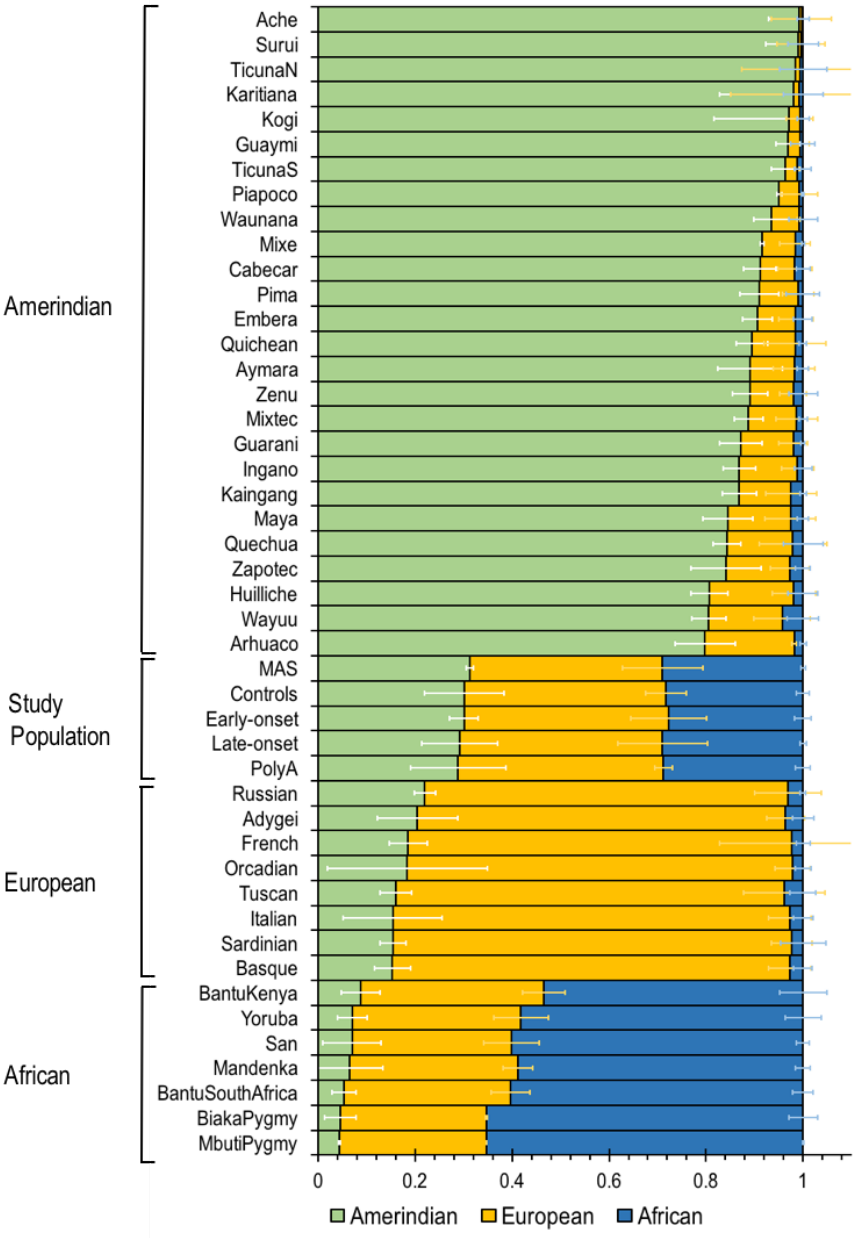
All markers were highly informative (PIC ≥ 0.50) with a low null allele frequency making them optimal and reliable for genetic diversity studies (Figure 1A). Moreover, average allelic richness observed per locus in the included markers was 4.30 ± 1.22. The observed heterozygosity and gene diversity varied across loci. The average observed (H_o) and expected (H_e) heterozygosity were 0.69 ± 0.16 and 0.68 ± 0.13, respectively. The calculated pairwise genetic distances between populations excluding null alleles were used to construct NJ trees to visualize the relationship between the 33 reference included populations and the Colombian samples (Figure 1B).

Figure 1. Marker reliability, gene diversity and population relationship indicators. **A.** Calculated null allele frequency (left axis) and polymorphic information content (PIC) (Right axis). **B.** Circular neighbor-joining tree using the pairwise calculated genetic distances.



Population genetic structure was examined in all included Colombian AD affected and healthy individuals, as well as in the individuals originated from reference populations, assuming three ancestral groups (k=3) (i.e., European, Amerindian and African). Estimation of individual ancestry proportions were obtained with the model-based MCMC Bayesian algorithm implemented in STRUCTURE, by using allelic frequencies for estimating a posterior distribution of the probability of membership to the predefined clusters (K), assuming that multiple loci are independent and are in Hardy-Weinberg equilibrium. The ancestry proportions (i.e., membership [Q]) obtained are depicted on figure 2. Reference populations from the three assumed ancestral populations presented contributions as follows: Amerindian (0.90 ± 0.06), European 0.79 ± 0.03 and African 0.60 ± 0.04 (Figure 2); while Colombian samples in general showed Amerindian (0.30 ± 0.0), European 0.42 ± 0.01 and African 0.29 ± 0.01 contributions.

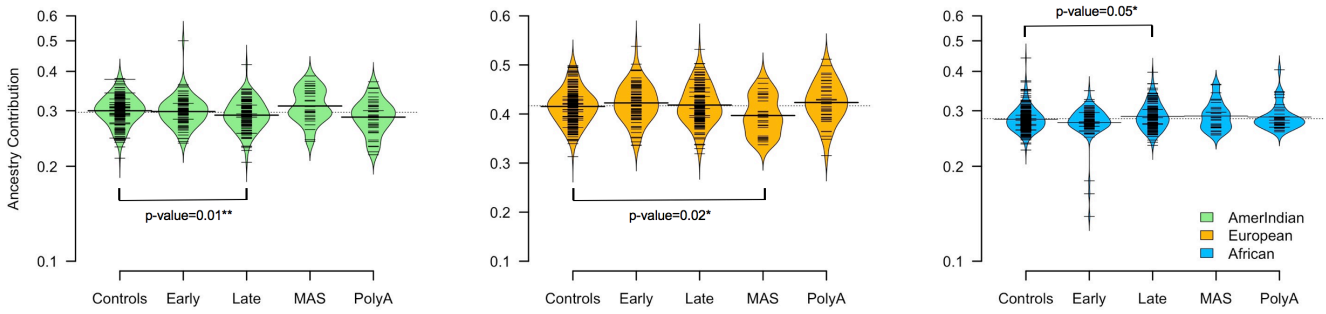
Figure 2. Population genetic structure analysis of Colombian individuals affected and unaffected by autoimmune disease. Each population shows the mean observed ancestry component predicted.



The calculated ancestry components for the Colombian studied population showed individual heterogeneity and similar results to previous reports (1-3). When the mean calculated proportions were compared with that of controls: the mean proportion of Amerindian ancestry significantly differed with late-onset samples (p-value=0.01), but not with early-onset, polyA or

MAS (Figure 3A); European ancestry mean proportions significantly differed with MAS samples (p-value=0.02) but not with early- Late-onset or polyA (Figure 3B) and African ancestry significantly differed in late-onset samples (p-value=0.05) but not for early-polyA or MAS (Figure 3C).

Figure 3. Beanplots of the calculated ancestry component in Colombian individuals affected and unaffected of 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 t-test.



The ancestry component effect for the studied traits was examined by logistic regression. For late-onset individuals, Amerindian ancestry showed a protective effect, for late-onset risk (odds ratio (OR) 0.00001 95% confidence interval (CI) 0.00 - 0.15), while African ancestry pertained a susceptibility effect (OR 3998 95% CI 1.05 - $1.90x10^{+07}$) (Table 3). European ancestry showed a protective effect for MAS individuals (OR $3.2x10^{+04}$ 95% CI 0.00 - 0.03). No other significant effect was observed with the evaluated ancestral components. It is possible that the very low or high effect of the OR observed is due to the sample size of the study. Of note, for quantitative covariates the OR is the ratio of two odds where the value of the covariate increases by 1.

Table 3. Logistic regression of autoimmune traits on genetic ancestry calculated components. AD: Autoimmune disease; PolyA: polyautoimmunity; MAS: Multiple autoimmune syndrome.

Trait	Covariate	OR (95% CI)	p-value
Late-Onset			
	Age	10.01 (0.99 - 1.04)	0.22
	Sex	3.68 (1.00 - 17.36)	0.06
	Amerindian	0.0001 (0.00 - 0.15)	0.01
	European	7.8 (0.02 - 3760)	0.51
	African	3998 (1.05 - $1.90x10^{+07}$)	0.05
Early-onset			
	Age	0.79 (0.73 - 0.85)	$1.04x10^{-09}$
	Sex	79.41 (27.06 - 340)	$2.55x10^{-12}$
	Amerindian	0.25 (0.00 - 707)	0.73
	European	226.21 (0.17 - $3.53x10^{+05}$)	0.14
	African	0.004 (0.00 - 7.35)	0.13
PolyA			
	Age	0.99 (0.98 - 1.02)	0.92
	Sex	1.76 (0.23 - 10.87)	0.54
	Amerindian	0.10 (0.00 - 348)	0.58
	European	0.10 (0.00 - 159)	0.54
	African	1244 (0.08 - $1.88x10^{+07}$)	0.14
MAS			
	Age	0.98 (0.95 - 1.00)	0.08
	Sex	2.07 (0.10 - 16.79)	0.535
	Amerindian	$3.2x10^{+04}$ (0.25 - $5.63x10^{+09}$)	0.08
	European	$5.64x10^{-07}$ (0.00 - 0.03)	0.01
	African	4264 (0.08 - $1.88x10^{+07}$)	0.19

Discussion

The great diversity of Latin American populations which is predominantly admixed represent a strong resource to decipher the genetic basis of complex traits (41). Population analysis and genetic association studies nurture and open up opportunities to examine gene-gene and gene-environment interactions (42).

Hispanic/Mestizo populations' present relative recent admixture derived from Amerindian inhabitants, European settlers (primarily from Spain), and West Africans brought to the Americas as a consequence of the slave trade. The contribution of each parental population and

the degree of admixture vary across regions in the Americas depending upon the local pattern of interaction among the different ethnic groups (43).

The entire group of Colombian individuals in this study 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 (44, 45). For all individuals, the ancestral component for Paisa individuals were estimated in order to examine specific contributions of expected and/or assumed founding populations (i.e., Amerindian, European and African) in AD affected (i.e., Early- and Late-onset) and in individuals presented with polyA and MAS, as well as, in healthy controls.

Data on this report showed appreciable differences in the admixture component of individuals affected with late-onset AD and MAS, while none were observed for early-onset and polyA individuals. Ancestry component estimates were different and have a protective effect on late-onset AD samples for Amerindian but a susceptibility one for African ancestry. Likewise, MAS individuals showed a protective effect for European ancestry when compared to control individuals. Of note, the calculated effects showed to be very low which warrants that the number of compared individuals confounds the effect and might indeed need further exploration, however this study support and provides a proof of principle of the possible role of ancestry in the AD as a trait.

Geography and ethnicity can affect the incidence and prevalence of AD. Both factors are thought to reflect etiological heterogeneity between environmental, genetic factors, and their interactions. Compelling examples of this effect have been reported for ADs such as RA (46, 47), primary SS (46, 48), scleroderma (SSc) (49) and SLE (50).

For indigenous populations AD prevalence information is limited. However, available data suggest SLE is more common in Native American Indians of Canada, Maori Pacific People in New Zealand, and Aborigines in Australia compared to their respective European populations (43). An increased proportion of the Amerindian genome correlates with the presence of an increased number of risk alleles (43).

Many Amerindian groups present high prevalence rates of RA, connective tissue diseases, and spondyloarthropathies (51, 52) including Pima, Chippewa and Yakima tribes (53). Previous reports have demonstrated that T1D in Mexican-Americans is associated with their European HLA contribution (54). Although other studies suggest that Amerindian ancestry may also

contribute increased risk for SLE (55, 56), others correlate it with lower socio-demographic status and increase the risk for developing renal involvement and SLE at an earlier age of onset (57).

Conclusions and perspectives

Autoimmune research generally focuses on a single disease, although autoimmune phenotypes could represent pleiotropic outcomes of non-specific disease genes underlying similar immunogenetic mechanisms (14). The current approach presumed autoimmunity as a trait rather than a clinical phenotype and also tries to approach the AD as a continuous trait presenting extreme phenotypes (i.e., early-onset and MAS traits, respectively). This reports shows appreciable differences in the admixture component of affected individuals relative to healthy controls for late-onset AD and MAS, while no significant differences were observed for early-onset and polyA.

Moreover, when the ancestry effect was examined a protective effect was observed for Amerindian ancestry and African ancestry for late-onset AD; while a protective effect was observed for European ancestry in MAS individuals. Although the observed effects warrant further exploration due to the relatively small sample size, this data provides a starting point for approaches using markers that provide higher genome coverage and better population discrimination. To our knowledge this would be the first proof of principle approach to examine the ancestry effect for individuals affected with AD (e.g., early- and late-onset) and with more than one AD (e.g., polyA and MAS). Overall, ancestry and autoimmunity in Colombian samples show how the autoimmune trait landscape is not a black and white scenario but rather a colorful mix of genetic and environmental factors (i.e., autoimmune ecology) (58).

Acknowledgments

This work was supported by Universidad del Rosario and Colciencias (122254531722), Bogota, Colombia.

Conflicts of Interest

The author(s) declare(s) that there is no conflict of interest regarding the publication of this manuscript.

References

1. Bedoya G, Montoya P, Garcia J, Soto I, Bourgeois S, Carvajal L, et al. Admixture dynamics in Hispanics: a shift in the nuclear genetic ancestry of a South American population isolate. Proc Natl Acad Sci U S A. 2006;103(19):7234-9.
2. Rojas W, Parra MV, Campo O, Caro MA, Lopera JG, Arias W, et al. Genetic make up and structure of Colombian populations by means of uniparental and biparental DNA markers. Am J Phys Anthropol. 2010;143(1):13-20.
3. Chacon-Duque JC, Adhikari K, Avendano E, Campo O, Ramirez R, Rojas W, et al. African genetic ancestry is associated with a protective effect on Dengue severity in colombian populations. Infect Genet Evol. 2014;27:89-95.
4. Choudhry S, Coyle NE, Tang H, Salari K, Lind D, Clark SL, et al. Population stratification confounds genetic association studies among Latinos. Hum Genet. 2006;118(5):652-64.
5. Gower BA, Fernandez JR, Beasley TM, Shriver MD, Goran MI. Using genetic admixture to explain racial differences in insulin-related phenotypes. Diabetes. 2003;52(4):1047-51.
6. Reiner AP, Ziv E, Lind DL, Nievergelt CM, Schork NJ, Cummings SR, et al. Population structure, admixture, and aging-related phenotypes in African American adults: the Cardiovascular Health Study. American journal of human genetics. 2005;76(3):463-77.
7. Salari K, Choudhry S, Tang H, Naqvi M, Lind D, Avila PC, et al. Genetic admixture and asthma-related phenotypes in Mexican American and Puerto Rican asthmatics. Genet Epidemiol. 2005;29(1):76-86.
8. Barnholtz-Sloan JS, McEvoy B, Shriver MD, Rebbeck TR. Ancestry estimation and correction for population stratification in molecular epidemiologic association studies. Cancer Epidemiol Biomarkers Prev. 2008;17(3):471-7.
9. Price AL, Patterson N, Yu F, Cox DR, Waliszewska A, McDonald GJ, et al. A genomewide admixture map for Latino populations. American journal of human genetics. 2007;80(6):1024-36.
10. Winkler CA, Nelson GW, Smith MW. Admixture mapping comes of age. Annual review of genomics and human genetics. 2010;11:65-89.
11. Florez JC, Price AL, Campbell D, Riba L, Parra MV, Yu F, et al. Strong association of socioeconomic status with genetic ancestry in Latinos: implications for admixture studies of type 2 diabetes. Diabetologia. 2009;52(8):1528-36.
12. Jacobson DL, Gange SJ, Rose NR, Graham NM. Epidemiology and estimated population burden of selected autoimmune diseases in the United States. Clinical immunology and immunopathology. 1997;84(3):223-43.
13. Mariani SM. Genes and autoimmune diseases - a complex inheritance. MedGenMed. 2004;6(4):18.
14. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. Cell. 1996;85(3):311-8.
15. Anaya JM, Corena R, Castiblanco J, Rojas-Villarraga A, Shoenfeld Y. The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity. Expert Review of Clinical Immunology. 2007;3(4):623-35.
16. Anaya JM, Gomez L, Castiblanco J. Is there a common genetic basis for autoimmune diseases? Clinical & developmental immunology. 2006;13(2-4):185-95.
17. Blank M, Gershwin ME. Autoimmunity: from the mosaic to the kaleidoscope. J Autoimmun. 2008;30(1-2):1-4.
18. Rojas-Villarraga A, Amaya-Amaya J, Rodriguez-Rodriguez A, Mantilla RD, Anaya JM. Introducing polyautoimmunity: secondary autoimmune diseases no longer exist. Autoimmune diseases. 2012;2012:254319.

19. Anaya JM, Castiblanco J, Rojas-Villarraga A, Pineda-Tamayo R, Levy RA, Gomez-Puerta J, et al. The Multiple Autoimmune Syndromes. A Clue for the Autoimmune Tautology. Clinical reviews in allergy & immunology. 2012.
20. Perez-Fernandez OM, Mantilla RD, Cruz-Tapias P, Rodriguez-Rodriguez A, Rojas-Villarraga A, Anaya JM. Spondyloarthropathies in autoimmune diseases and vice versa. Autoimmune diseases. 2012;2012:736384.
21. Cardenas-Roldan J, Rojas-Villarraga A, Anaya JM. How do autoimmune diseases cluster in families? A systematic review and meta-analysis. BMC medicine. 2013;11(1):73.
22. Sloka S. Observations on recent studies showing increased co-occurrence of autoimmune diseases. J Autoimmun. 2002;18(3):251-7.
23. Cardenas-Roldan J, Rojas-Villarraga A, Anaya JM. How do autoimmune diseases cluster in families? A systematic review and meta-analysis. BMC medicine. 2013;11:73.
24. Baranzini SE. The genetics of autoimmune diseases: a networked perspective. Current opinion in immunology. 2009;21(6):596-605.
25. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nature reviews Genetics. 2009;10(1):43-55.
26. Bias WB, Reveille JD, Beaty TH, Meyers DA, Arnett FC. Evidence that autoimmunity in man is a Mendelian dominant trait. American journal of human genetics. 1986;39(5):584-602.
27. Wandstrat A, Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol. 2001;2(9):802-9.
28. Invernizzi P. Future directions in genetic for autoimmune diseases. J Autoimmun. 2009;33(1):1-2.
29. Priori R, Medda E, Conti F, Cassara EA, Danieli MG, Gerli R, et al. Familial autoimmunity as a risk factor for systemic lupus erythematosus and vice versa: a case-control study. Lupus. 2003;12(10):735-40.
30. Anaya JM, Tobon GJ, Vega P, Castiblanco J. Autoimmune disease aggregation in families with primary Sjogren's syndrome. J Rheumatol. 2006;33(11):2227-34.
31. Report of the expert committee on the diagnosis and classification of diabetes mellitus. Diabetes Care. 2003;26 Suppl 1:S5-20.
32. Tobon GJ, Arango A, Abad V, Garcia J, Cuervo H, Velasquez A, et al. Clinical and immunological characteristics of type 1 diabetes mellitus in a northwestern Colombian population. Diabetes Res Clin Pract. 2006;72(2):170-5.
33. Pemberton TJ, DeGiorgio M, Rosenberg NA. Population structure in a comprehensive genomic data set on human microsatellite variation. G3 (Bethesda). 2013;3(5):891-907.
34. Team RC. R: A Language and Environment for Statistical Computing. R: A Language and Environment for Statistical Computing. v3.1.1 ed2014.
35. Lischer HE, Excoffier L. PGDSpider: an automated data conversion tool for connecting population genetics and genomics programs. Bioinformatics. 2012;28(2):298-9.
36. Chapuis MP, Estoup A. Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol. 2007;24(3):621-31.
37. Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F. GENETIX 4.05, logiciel sous Windows pour la genetique des populations. Universite de Montpellier, Montpellier, France. 1996-2004.
38. Falush D, Stephens M, Pritchard JK. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. Genetics. 2003;164(4):1567-87.
39. Pritchard JK, Stephens M, Donnelly P. Inference of population structure using multilocus genotype data. Genetics. 2000;155(2):945-59.
40. Pemberton TJ, Sandefur CI, Jakobsson M, Rosenberg NA. Sequence determinants of human microsatellite variability. BMC Genomics. 2009;10:612.
41. Amorim CE, Wang S, Marrero AR, Salzano FM, Ruiz-Linares A, Bortolini MC. X-chromosomal genetic diversity and linkage disequilibrium patterns in Amerindians and non-

Amerindian populations. American journal of human biology : the official journal of the Human Biology Council. 2011;23(3):299-304.
42. Gonzalez Burchard E, Borrell LN, Choudhry S, Naqvi M, Tsai HJ, Rodriguez-Santana JR, et al. Latino populations: a unique opportunity for the study of race, genetics, and social environment in epidemiological research. American journal of public health. 2005;95(12):2161-8.
43. Makeev AF, Dzhurko GN, Zotsenko AN. [Acute anaerobic paraproctitis complicated by phlegmon of the perineum and scrotum in a patient with diabetes mellitus]. Klinicheskaia khirurgiia. 1991(2):65.
44. Carvajal-Carmona LG, Soto ID, Pineda N, Ortiz-Barrientos D, Duque C, Ospina-Duque J, et al. Strong Amerind/white sex bias and a possible Sephardic contribution among the founders of a population in northwest Colombia. American journal of human genetics. 2000;67(5):1287-95.
45. Bravo ML, Valenzuela CY, Arcos-Burgos OM. Polymorphisms and phyletic relationships of the Paisa community from Antioquia (Colombia). Gene Geogr. 1996;10(1):11-7.
46. Alamanos Y, Tsifetaki N, Voulgari PV, Venetsanopoulou AI, Siozos C, Drosos AA. Epidemiology of primary Sjogren's syndrome in north-west Greece, 1982-2003. Rheumatology. 2006;45(2):187-91.
47. Jacobsson LT, Hanson RL, Knowler WC, Pillemer S, Pettitt DJ, McCance DR, et al. Decreasing incidence and prevalence of rheumatoid arthritis in Pima Indians over a twenty-five-year period. Arthritis and rheumatism. 1994;37(8):1158-65.
48. Goransson LG, Haldorsen K, Brun JG, Harboe E, Jonsson MV, Skarstein K, et al. The point prevalence of clinically relevant primary Sjogren's syndrome in two Norwegian counties. Scandinavian journal of rheumatology. 2011;40(3):221-4.
49. Zurauskas J, Beroukas D, Walker JG, Smith MD, Ahern MJ, Roberts-Thomson PJ. Scleroderma in Australian aborigines. Internal medicine journal. 2005;35(1):60-2.
50. Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis and rheumatism. 2008;58(1):15-25.
51. Hirsch R, Lin JP, Scott WW, Jr., Ma LD, Pillemer SR, Kastner DL, et al. Rheumatoid arthritis in the Pima Indians: the intersection of epidemiologic, demographic, and genealogic data. Arthritis and rheumatism. 1998;41(8):1464-9.
52. Harvey J, Lotze M, Stevens MB, Lambert G, Jacobson D. Rheumatoid arthritis in a Chippewa Band. I. Pilot screening study of disease prevalence. Arthritis and rheumatism. 1981;24(5):717-21.
53. Molokhia M, McKeigue P. Risk for rheumatic disease in relation to ethnicity and admixture. Arthritis research. 2000;2(2):115-25.
54. Shaw SJ, Vadheim CM, Rotter JI, Costin G, Beshkov Y, Zeidler A. Ancestral origin of insulin-dependent diabetes in Mexican-Americans. International journal of epidemiology. 1992;21(4):725-9.
55. Fernandez M, Alarcon GS, Apte M, Andrade RM, Vila LM, Reveille JD, et al. Systemic lupus erythematosus in a multiethnic US cohort: XLIII. The significance of thrombocytopenia as a prognostic factor. Arthritis and rheumatism. 2007;56(2):614-21.
56. Pons-Estel BA, Catoggio LJ, Cardiel MH, Soriano ER, Gentiletti S, Villa AR, et al. The GLADEL multinational Latin American prospective inception cohort of 1,214 patients with systemic lupus erythematosus: ethnic and disease heterogeneity among "Hispanics". Medicine. 2004;83(1):1-17.
57. Sanchez E, Rasmussen A, Riba L, Acevedo-Vasquez E, Kelly JA, Langefeld CD, et al. Impact of genetic ancestry and sociodemographic status on the clinical expression of systemic lupus erythematosus in American Indian-European populations. Arthritis and rheumatism. 2012;64(11):3687-94.

58. Anaya JM, Ramirez-Santana C, Alzate MA, Molano-Gonzalez N, Rojas-Villarraga A. The Autoimmune Ecology. Front Immunol. 2016;7:139.

General Discussion

In this time and era, lack of clear diagnostic tools and defined disease criteria leaves patients in a bureaucratic limbo soaring through a healthcare system in search of a complete an accurate diagnosis to receive appropriate treatment. Clinical pathologies draw us to envisage disease as either an independent entity or a diverse set of traits governed by common physiopathological mechanisms prompted by environmental assaults throughout life (66). ADs are not the exception for this premise, given they represent a diverse collection of diseases in terms of their demographic profile and primary clinical manifestations (2).

As multifactorial etiologies, ADs develop from the cumulative effect of diverse events on the immune system (15). It is now clear they do not begin at the time of clinical appearance but rather many years before. Emerging research has identified both common and rare genetic loci shared across the spectrum of AD and the biologic pathways whose involvement is implicated by these shared loci (9, 67-69). The extent to which immune-related signaling and/or other pathways are implicated for each of these disorders varies and suggests that most of these pathways contribute or has a role to a variable degree in most of these disorders.

Currently, available genetic information points out to nearly half of loci identified in GWAS studies of an individual disease influence risk to at least two diseases, arguing for a genetic basis to co-morbidity. Moreover, there are examples of several variants with opposing risk profiles in different diseases. Support for the idea of common patterns of association and shared biological processes is obtained by loci clustered over a pattern of diseases as they affect and harbor genes encoding for interacting proteins at a much higher rate than by chance. These results suggest that multi-phenotype mapping will identify the molecular mechanisms underlying co-morbid immune-mediated inflammatory and ADs.

This expected commonality has motivated several meta-analyses across pairs of diseases to establish their shared genetic basis. Approaches taking into account CD and RA (70), T1D (71) and IBD (72) have revealed loci overlap adding further support to a shared underlying pathogenesis. In addition, there is evidence that loci predisposing to one disease can have effects on risk of a second disease (71), although the risk allele for one disease may not be the same as for the second (73).

Modeling and simulation analysis estimate that in several diseases where GWAS have been successful, further low effect associations remain to be discovered (74). This problem gets worse when considering independent discoveries across single ADs, since power would be multiplicative across studies. Therefore, estimates of the true extent of genetic sharing are probably underestimated by either simple overlaps or pairwise meta-analyses.

This project extends support to an autoimmune commonality principle in a single well-characterized population of AD affected individuals and their families by exploring the familial aggregation and segregation, admixture and homozygosity effect. The current approach presumed autoimmunity as a trait rather than a clinical phenotype, approaching AD as a continuous trait presenting extreme phenotypes (i.e., early-onset and MAS traits, respectively).

Key Points of Each Chapter

Chapter 1 – What is next after the genes for autoimmunity?

- Presents the state of the art and perspective of the current and future directions of autoimmunity and ADs, by discussing the many components affecting the potential use and application of genetic, evolutionary, demographic, environmental and immunopathological information that could be used for prediction, prevention and eventually treatment of ADs.

Chapter 2 – Familial Aggregation and Segregation Analysis in Families Presenting Autoimmunity, Polyautoimmunity, and Multiple Autoimmune Syndrome

- Describes aggregation and segregation analyses in Colombian families enriched by autoimmunity as a trait and show how ADs, polyautoimmunity and MAS are not independent entities.

- Familial aggregation for ADs was observed between parents and offspring as well as in sibling pairs in late-onset families, while aggregation for polyautoimmunity and MAS was

lower given by the fact that both traits represent a more complex etiology with lower prevalence but still a common autoimmunity background.

- Segregation analyses were not able to discern a Mendelian transmission model but still suggested a major gene(s) transmission for AD in late-onset families, while for early-onset families a stochastic model was suggested.

Chapter 3 – Homozygosity Genetic Analysis in Autoimmunity Affected Individuals and Multiplex Autoimmune Disease Families

- This report presented two approaches: a case – control comparison and evaluation on the effect of homozygosity at the genome-wide level and a model-free affected pair linkage approach to identify IBD loci.

- 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.

- On the other hand, the model-free affected pair linkage approach, did not provide any suggestive linkage signals for early-, late-onset, polyA or MAS. Putative/Marginal signals displayed excess allele sharing for early-onset and MAS, both extreme phenotypes in autoimmunity, but their signals warrants caution.

Chapter 4 – Ancestry effect in Colombian Individuals presenting Autoimmunity, Polyautoimmunity and Multiple Autoimmune Syndrome

- This report showed appreciable differences in the admixture component of affected individuals relative to healthy controls for late-onset AD and MAS, while no significant differences were observed for early-onset and polyA. Moreover, when the ancestry effect was examined a protective effect was observed for Amerindian ancestry and African ancestry for late-onset AD; while a protective effect was observed for European ancestry in MAS individuals.

Limitations and Challenges

The present study is 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 corroborate and expand on sample size, marker coverage and their effects. Closer inspection of clinical and phenotypic quantitative variants is warranted, as well as inclusion of environmental and clinical available variants. The affected relative pair approach was only possible, instead of a sibling pair due to the sample size and available concordant and discordant pairs.

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 exome sequencing in individuals affected with polyA and MAS and their relatives (75, 76).

Moreover, with the extant amount and accumulation of information, the issue of interpretation of the genetic variants effect and their role comes together. Several correlated alleles appear to impart risk to some diseases but are protective for others (73, 77), plus it remains unclear how this evidence should be incorporated into a pathway view of disease. Each new genetic finding can suggest multiple hypotheses that need to be fit into an overall scheme of pathogenesis. Ongoing research points to some expected shared biology (78), however a complete and convincing share genetic background still needs to be further supported

Going forward for Genetic Research in Autoimmunity

This project suggested that polyA and MAS are not AD independent traits and that gender, age and age of onset represent factors that define and allow the study of the dynamics of the traits within the familial group. Also segregation data provided evidence for the genetic component role in the etiology of AD in late-onset families, while for early-onset families and perhaps because of their the relatively familial young status, eluded a clear picture of autoimmunity segregation and aggregation.

The data also 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. Moreover, ancestry and autoimmunity in Colombian samples showed how the autoimmune trait landscape is not a black and white scenario but rather a colorful mix of genetic and environmental factors (i.e., autoimmune ecology) (79). Thus, a clinical defined individual AD, defined by symptoms and signs, might not be completely juxtaposed to the AD trait defined by environment and genetics, which makes even more difficult the task to define and untangle disease mechanisms.

We are living an era of new technologies and by far the biggest impacts on the long run are the newly adopted technologies including next-generation sequencing (NGS) (80). Current studies in many phenotypes are presently using resequencing in regions found through GWAS to ensure that the majority of variation has been identified before embarking on detailed omic functional layers (81, 82). Studies must also carefully evaluate the impact of environmental influences in combination with genetic predisposition to disease to better understand the pathophysiological mechanisms underpinning autoimmune phenotypes.

Last but not least, association rather than causality result from the combined effect of many variants. Many different combinations of risk alleles are able to independently generate a high level of disease risk, without individual loci being necessary or sufficient for the development of disease. Thus, a long way to disentangle a complete AD genetic architecture understanding lies ahead, however technology and new multidisciplinary approaches will nurture this landscape.

References

1. Jacobson DL, Gange SJ, Rose NR, Graham NM. Epidemiology and estimated population burden of selected autoimmune diseases in the United States. Clinical immunology and immunopathology. 1997;84(3):223-43.
2. Mariani SM. Genes and autoimmune diseases - a complex inheritance. MedGenMed. 2004;6(4):18.
3. Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. Cell. 1996;85(3):311-8.
4. Anaya JM, Corena R, Castiblanco J, Rojas-Villarraga A, Shoenfeld Y. The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity. Expert Review of Clinical Immunology. 2007;3(4):623-35.
5. Anaya JM, Gomez L, Castiblanco J. Is there a common genetic basis for autoimmune diseases? Clinical & developmental immunology. 2006;13(2-4):185-95.
6. Cotsapas C, Hafler DA. Immune-mediated disease genetics: the shared basis of pathogenesis. Trends in immunology. 2013;34(1):22-6.
7. Cardenas-Roldan J, Rojas-Villarraga A, Anaya JM. How do autoimmune diseases cluster in families? A systematic review and meta-analysis. BMC medicine. 2013;11:73.
8. Baranzini SE. The genetics of autoimmune diseases: a networked perspective. Current opinion in immunology. 2009;21(6):596-605.
9. Zhernakova A, van Diemen CC, Wijmenga C. Detecting shared pathogenesis from the shared genetics of immune-related diseases. Nature reviews Genetics. 2009;10(1):43-55.
10. Bias WB, Reveille JD, Beaty TH, Meyers DA, Arnett FC. Evidence that autoimmunity in man is a Mendelian dominant trait. American journal of human genetics. 1986;39(5):584-602.
11. Wandstrat A, Wakeland E. The genetics of complex autoimmune diseases: non-MHC susceptibility genes. Nat Immunol. 2001;2(9):802-9.
12. Invernizzi P. Future directions in genetic for autoimmune diseases. J Autoimmun. 2009;33(1):1-2.
13. Anaya JM, Tobon GJ, Vega P, Castiblanco J. Autoimmune disease aggregation in families with primary Sjogren's syndrome. J Rheumatol. 2006;33(11):2227-34.
14. Tobon GJ, Arango A, Abad V, Garcia J, Cuervo H, Velasquez A, et al. Clinical and immunological characteristics of type 1 diabetes mellitus in a northwestern Colombian population. Diabetes Res Clin Pract. 2006;72(2):170-5.
15. Sloka S. Observations on recent studies showing increased co-occurrence of autoimmune diseases. J Autoimmun. 2002;18(3):251-7.
16. Blank M, Gershwin ME. Autoimmunity: from the mosaic to the kaleidoscope. J Autoimmun. 2008;30(1-2):1-4.
17. Rojas-Villarraga A, Amaya-Amaya J, Rodriguez-Rodriguez A, Mantilla RD, Anaya JM. Introducing polyautoimmunity: secondary autoimmune diseases no longer exist. Autoimmune diseases. 2012;2012:254319.
18. Anaya JM, Castiblanco J, Rojas-Villarraga A, Pineda-Tamayo R, Levy RA, Gomez-Puerta J, et al. The Multiple Autoimmune Syndromes. A Clue for the Autoimmune Tautology. Clinical reviews in allergy & immunology. 2012.
19. Anaya JM, Rojas-Villarraga A, Garcia-Carrasco M. The autoimmune tautology: from polyautoimmunity and familial autoimmunity to the autoimmune genes. Autoimmune diseases. 2012;2012:297193.
20. Cardenas-Roldan J, Rojas-Villarraga A, Anaya JM. How do autoimmune diseases cluster in families? A systematic review and meta-analysis. BMC medicine. 2013;11(1):73.

21. Anaya JM, Castiblanco J, Tobon GJ, Garcia J, Abad V, Cuervo H, et al. Familial clustering of autoimmune diseases in patients with type 1 diabetes mellitus. Journal of autoimmunity. 2006;26(3):208-14.
22. Humbert P, Dupond JL. [Multiple autoimmune syndromes]. Ann Med Interne (Paris). 1988;139(3):159-68.
23. Schmidt MB. Eine biglandulare Erkrankung (Nebennieren und Schilddrüse) bei Morbus Addisonii. Verh Dtsch Ges Pathol. 1926;21(212-21).
24. Betterle C, Lazzarotto F, Presotto F. Autoimmune polyglandular syndrome Type 2: the tip of an iceberg? Clinical and experimental immunology. 2004;137(2):225-33.
25. Eisenbarth GS, Gottlieb PA. Autoimmune polyendocrine syndromes. The New England journal of medicine. 2004;350(20):2068-79.
26. Anaya JM, Corena R, Castiblanco J, Rojas-Villarraga A, Shoenfeld Y. The kaleidoscope of autoimmunity: multiple autoimmune syndromes and familial autoimmunity. Expert review of clinical immunology. 2007;3(4):623-35.
27. Voight BF, Cotsapas C. Human genetics offers an emerging picture of common pathways and mechanisms in autoimmunity. Current opinion in immunology. 2012;24(5):552-7.
28. Mori M, Yamada R, Kobayashi K, Kawaida R, Yamamoto K. Ethnic differences in allele frequency of autoimmune-disease-associated SNPs. Journal of human genetics. 2005;50(5):264-6.
29. Gibson G. Decanalization and the origin of complex disease. Nature reviews Genetics. 2009;10(2):134-40.
30. Aylsworth AS. Defining disease phenotypes. In: Haines JL, Pericak-Vance MA, editors. Approaches to Gene Mapping in Complex Human Diseases. New York, NY: Wiley-Liss; 1998. p. 53-76.
31. Rose AM, Bell LC. Epistasis and immunity: the role of genetic interactions in autoimmune diseases. Immunology. 2012;137(2):131-8.
32. Burton PR, Tobin MD, Hopper JL. Key concepts in genetic epidemiology. Lancet. 2005;366(9489):941-51.
33. Goris A, Liston A. The immunogenetic architecture of autoimmune disease. Cold Spring Harbor perspectives in biology. 2012;4(3).
34. Barcellos LF, Kamdar BB, Ramsay PP, DeLoa C, Lincoln RR, Caillier S, et al. Clustering of autoimmune diseases in families with a high-risk for multiple sclerosis: a descriptive study. The Lancet Neurology. 2006;5(11):924-31.
35. Somers EC, Thomas SL, Smeeth L, Hall AJ. Autoimmune diseases co-occurring within individuals and within families: a systematic review. Epidemiology. 2006;17(2):202-17.
36. Michou L, Rat AC, Lasbleiz S, Bardin T, Cornelis F. Prevalence and distribution of autoimmune diseases in 368 rheumatoid arthritis families. The Journal of rheumatology. 2008;35(5):790-6.
37. Alarcon-Segovia D, Alarcon-Riquelme ME, Cardiel MH, Caeiro F, Massardo L, Villa AR, et al. Familial aggregation of systemic lupus erythematosus, rheumatoid arthritis, and other autoimmune diseases in 1,177 lupus patients from the GLADEL cohort. Arthritis and rheumatism. 2005;52(4):1138-47.
38. Prahalad S, Shear ES, Thompson SD, Giannini EH, Glass DN. Increased prevalence of familial autoimmunity in simplex and multiplex families with juvenile rheumatoid arthritis. Arthritis and rheumatism. 2002;46(7):1851-6.
39. Kumar V, Rajadhyaksha M, Wortsman J. Celiac disease-associated autoimmune endocrinopathies. Clin Diagn Lab Immunol. 2001;8(4):678-85.
40. Hafler DA. Multiple sclerosis. J Clin Invest. 2004;113(6):788-94.
41. Anaya JM, Delgado-Vega AM, Castiblanco J. Genetic basis of Sjogren syndrome. How strong is the evidence? Clinical & developmental immunology. 2006;13(2):209-22.

42. Herrick AL, Worthington J. Genetic epidemiology: systemic sclerosis. Arthritis Res. 2002;4(3):165-8.
43. Jawaheer D, Lum RF, Amos CI, Gregersen PK, Criswell LA. Clustering of disease features within 512 multicase rheumatoid arthritis families. Arthritis and rheumatism. 2004;50(3):736-41.
44. Chistiakov DA. Immunogenetics of Hashimoto's thyroiditis. J Autoimmune Dis. 2005;2(1):1.
45. Thanvi BR, Lo TC. Update on myasthenia gravis. Postgrad Med J. 2004;80(950):690-700.
46. Barragan-Martinez C, Amaya-Amaya J, Pineda-Tamayo R, Mantilla RD, Castellanos-de la Hoz J, Bernal-Macias S, et al. Gender differences in Latin-American patients with rheumatoid arthritis. Gender medicine. 2012;9(6):490-510 e5.
47. Fairweather D, Frisancho-Kiss S, Rose NR. Sex differences in autoimmune disease from a pathological perspective. The American journal of pathology. 2008;173(3):600-9.
48. Rubtsov AV, Rubtsova K, Kappler JW, Marrack P. Genetic and hormonal factors in female-biased autoimmunity. Autoimmunity reviews. 2010;9(7):494-8.
49. Chitnis S, Monteiro J, Glass D, Apatoff B, Salmon J, Concannon P, et al. The role of X-chromosome inactivation in female predisposition to autoimmunity. Arthritis Res. 2000;2(5):399-406.
50. Stewart JJ. Theory and treatment of the X-inactivation chimera in female-prevalent autoimmune disease. Arch Immunol Ther Exp (Warsz). 1999;47(6):355-9.
51. Reveille JD, Wilson RW, Provost TT, Bias WB, Arnett FC. Primary Sjogren's syndrome and other autoimmune diseases in families. Prevalence and immunogenetic studies in six kindreds. Ann Intern Med. 1984;101(6):748-56.
52. Priori R, Medda E, Conti F, Cassara EA, Danieli MG, Gerli R, et al. Familial autoimmunity as a risk factor for systemic lupus erythematosus and vice versa: a case-control study. Lupus. 2003;12(10):735-40.
53. Tait KF, Marshall T, Berman J, Carr-Smith J, Rowe B, Todd JA, et al. Clustering of autoimmune disease in parents of siblings from the Type 1 diabetes Warren repository. Diabet Med. 2004;21(4):358-62.
54. Coyle PK. Gender issues. Neurol Clin. 2005;23(1):39-60, v-vi.
55. Guo SW, Tuomilehto J. Preferential transmission of type 1 diabetes from parents to offspring: fact or artifact? Genet Epidemiol. 2002;23(4):323-34.
56. Kantarci OH, Barcellos LF, Atkinson EJ, Ramsay PP, Lincoln R, Achenbach SJ, et al. Men transmit MS more often to their children vs women: the Carter effect. Neurology. 2006;67(2):305-10.
57. Carter CO. The inheritance of congenital pyloric stenosis. Br Med Bull. 1961;17:251-4.
58. Anaya JM, Shoenfeld Y, Correa PA, Garcia-Carrasco M, Cervera R. Autoimmunity and Autoimmune Disease. 1 ed. Anaya JM, editor. Medellin: CIB; 2005.
59. Arcos-Burgos M, Parodi E, Salgar M, Bedoya E, Builes J, Jaramillo D, et al. Vitiligo: complex segregation and linkage disequilibrium analyses with respect to microsatellite loci spanning the HLA. Hum Genet. 2002;110(4):334-42.
60. Lin JP, Cash JM, Doyle SZ, Peden S, Kanik K, Amos CI, et al. Familial clustering of rheumatoid arthritis with other autoimmune diseases. Human genetics. 1998;103(4):475-82.
61. Alamanos Y, Tsifetaki N, Voulgari PV, Venetsanopoulou AI, Siozos C, Drosos AA. Epidemiology of primary Sjogren's syndrome in north-west Greece, 1982-2003. Rheumatology. 2006;45(2):187-91.
62. Jacobsson LT, Hanson RL, Knowler WC, Pillemer S, Pettitt DJ, McCance DR, et al. Decreasing incidence and prevalence of rheumatoid arthritis in Pima Indians over a twenty-five-year period. Arthritis and rheumatism. 1994;37(8):1158-65.
63. Goransson LG, Haldorsen K, Brun JG, Harboe E, Jonsson MV, Skarstein K, et al. The point prevalence of clinically relevant primary Sjogren's syndrome in two Norwegian counties. Scandinavian journal of rheumatology. 2011;40(3):221-4.

64. Zurauskas J, Beroukas D, Walker JG, Smith MD, Ahern MJ, Roberts-Thomson PJ. Scleroderma in Australian aborigines. Internal medicine journal. 2005;35(1):60-2.
65. Helmick CG, Felson DT, Lawrence RC, Gabriel S, Hirsch R, Kwoh CK, et al. Estimates of the prevalence of arthritis and other rheumatic conditions in the United States. Part I. Arthritis and rheumatism. 2008;58(1):15-25.
66. Roden DM. Personalized medicine and the genotype-phenotype dilemma. Journal of interventional cardiac electrophysiology : an international journal of arrhythmias and pacing. 2011;31(1):17-23.
67. McKinney C, Merriman ME, Chapman PT, Gow PJ, Harrison AA, Highton J, et al. Evidence for an influence of chemokine ligand 3-like 1 (CCL3L1) gene copy number on susceptibility to rheumatoid arthritis. Annals of the rheumatic diseases. 2008;67(3):409-13.
68. Rai E, Wakeland EK. Genetic predisposition to autoimmunity--what have we learned? Seminars in immunology. 2011;23(2):67-83.
69. Lettre G, Rioux JD. Autoimmune diseases: insights from genome-wide association studies. Human molecular genetics. 2008;17(R2):R116-21.
70. Zhernakova A, Stahl EA, Trynka G, Raychaudhuri S, Festen EA, Franke L, et al. Meta-analysis of genome-wide association studies in celiac disease and rheumatoid arthritis identifies fourteen non-HLA shared loci. PLoS genetics. 2011;7(2):e1002004.
71. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. The New England journal of medicine. 2008;359(26):2767-77.
72. Festen EA, Goyette P, Green T, Boucher G, Beauchamp C, Trynka G, et al. A meta-analysis of genome-wide association scans identifies IL18RAP, PTPN2, TAGAP, and PUS10 as shared risk loci for Crohn's disease and celiac disease. PLoS genetics. 2011;7(1):e1001283.
73. Sirota M, Schaub MA, Batzoglou S, Robinson WH, Butte AJ. Autoimmune disease classification by inverse association with SNP alleles. PLoS genetics. 2009;5(12):e1000792.
74. Stahl EA, Wegmann D, Trynka G, Gutierrez-Achury J, Do R, Voight BF, et al. Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis. Nature genetics. 2012;44(5):483-9.
75. Johar A, Sarmiento-Monroy JC, Rojas-Villarraga A, Silva-Lara MF, Patel HR, Mantilla RD, et al. Definition of mutations in polyautoimmunity. J Autoimmun. 2016;72:65-72.
76. Johar AS, Mastronardi C, Rojas-Villarraga A, Patel HR, Chuah A, Peng K, et al. Novel and rare functional genomic variants in multiple autoimmune syndrome and Sjogren's syndrome. Journal of translational medicine. 2015;13:173.
77. Wang K, Baldassano R, Zhang H, Qu HQ, 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. 2010;19(10):2059-67.
78. Eaton WW, Rose NR, Kalaydjian A, Pedersen MG, Mortensen PB. Epidemiology of autoimmune diseases in Denmark. Journal of autoimmunity. 2007;29(1):1-9.
79. Anaya JM, Ramirez-Santana C, Alzate MA, Molano-Gonzalez N, Rojas-Villarraga A. The Autoimmune Ecology. Front Immunol. 2016;7:139.
80. Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, et al. A map of human genome variation from population-scale sequencing. Nature. 2010;467(7319):1061-73.
81. Shendure J, Lieberman Aiden E. The expanding scope of DNA sequencing. Nature biotechnology. 2012;30(11):1084-94.
82. A user's guide to the encyclopedia of DNA elements (ENCODE). PLoS biology. 2011;9(4):e1001046.