Identification of clinically relevant genetic variation in immune-mediated inflammatory diseases using genome-wide approaches

Adrià Aterido Ballonga

TESI DOCTORAL UPF / 2018

DIRECTORS DE LA TESI

Dr. Antonio Julià Cano
(Grup de Recerca de Reumatologia, Vall d'Hebron Institut de Recerca)

Dra. Sara Marsal Barril
(Grup de Recerca de Reumatologia, Vall d'Hebron Institut de Recerca)

Departament de Ciències Experimentals i de la Salut

This is your life.

Do what you want and do it often.

If you don't like something, change it.

If you don't like your job, quit.

If you don't have enough time, stop watching TV.

If you are looking for the love of your life, stop; they will be waiting for you when you start doing things you love.

Stop over analyzing, life is simple.

All emotions are beautiful.

When you eat, appreciate every last bite.

Open your mind, arms and heart to new things and people, we are united in our differences.

Ask the next person you see what their passion is and share your inspiring dream with them.

Travel often; getting lost will help you find yourself.

Some opportunities only come once, seize them.

Life is about the people you meet, and the things you create with them so go out and start creating.

Life is short.

Live your dream and share your passion.

The Holstee Manifest, 2009

Acknowledgements - Agraïments

Un acord, un gir de guió inesperat i la plena convicció d'haver pres la decisió encertada. Aquest és el pròleg d'una tesi que no havia de ser escrita.

Quan l'any 2013 vaig arribar al GRR per realitzar el projecte final de màster, poc m'imaginava que la recerca translacional em despertaria una motivació tan gran com per dedicar quatre anys de la meva vida a la realització d'una tesi doctoral.

Aquesta tesi no hauria estat possible sense moltes de les persones que m'han acompanyat al llarg d'aquests anys, sobretot el Dr. Antonio Julià i la Dra. Sara Marsal. D'una manera molt especial m'agradaria agrair al Toni el seu rigor científic, que no només ha servit per dirigir-me la tesi sinó que també ha estat fonamental per formar-me com a investigador. No m'oblido que va ser ell qui va confiar i apostar per mi des del primer dia. A la Sara li estic enormement agraït per transmetre'm la seva serenitat, clarividència i fermesa a l'hora de prendre decisions, tant en recerca com en la vida mateixa. A tots dos vull donar-los les gràcies per haver-me fet sentir valorat i una peça important de l'engranatge del grup.

També voldria donar les gràcies al Dr. Robert Castelo per la tutoria de la tesi, sempre accessible i disposat a facilitar-ne la realització.

Pertànyer a un grup multidisciplinari com el GRR té un valor que cal posar de manifest. Creuar-me amb professionals de formació molt diversa m'ha permès eixamplar els coneixements en disciplines diferents. Des d'un punt de vista clínic, la Maria, el Pablo, la Gaby, l'Estefania, la Mireia, l'Andrea, l'Ernesto i el Santos m'han ajudat a visualitzar l'impacte que la recerca pot arribar a tenir a la pràctica mèdica. Al laboratori, el Raül i la Núria han estat imprescindibles per coordinar el biobanc i la Raquel, l'Ivan, la Natàlia i l'Alba han estat cabdals tant en l'obtenció de dades clíniques com en el processament de les mostres biològiques. En l'àrea de bioinformàtica, l'Arnald, el Toni, el Sergio, el David i el Jordi sempre han estat disposats a aportar una visió crítica i constructiva per tal d'analitzar rigorosament les dades genòmiques dels pacients. També voldria donar les gràcies a l'Elena per fer de la burocràcia un joc de nens i per estar sempre al peu del canó.

En el marc del consorci IMID, m'agradaria agrair a tots els membres la seva professionalitat i participació activa en els projectes de recerca liderats pel GRR. De la mateixa manera, voldria agrair als pacients la seva voluntat i disposició a participar en estudis d'investigació com els que es presenten en aquesta tesi. Indubtablement, la realització d'aquesta tesi no hauria estat possible sense ells.

Però no tot ha estat recerca. Quan tancava la porta de l'oficina, començava la doble vida: Puig-reig i Barcelona. Agafar oxigen al poble durant el cap de setmana, per consumir-lo a la ciutat al llarg de la setmana.

A Barcelona, l'atzar m'ha brindat l'oportunitat de fer bons amics, com el Bernardo. Amb ell he compartit llargues tardes de "running" per Collserola tot evadint-nos de la tesi. També el Gerard, amics i amigues de la carrera, del màster i de vòlei, així com companys i companyes de pis que venen i van. Tots ells han fet que la realització d'aquesta tesi fos molt més plaent.

Al poble, família i amics són casa meva. M'agradaria agrair als meus pares tot l'esforç i sacrifici que han fet per tal que jo pogués gaudir de l'educació que ells no van poder rebre. Sobretot, voldria donar-los les gràcies per respectar sempre les meves decisions. A la meva germana, Nina, tan diferents i tan iguals, per pintar els dies grisos de colors. Als meus avis, per encoratjar-me i transmetre'm sense saber-ho que el meu progrés és el seu orgull. I a la Ivet, per estar sempre al meu costat, no només compartint les meves il·lusions sinó fent-les encara més grans.

Als tiparrets, Arnau, Coll, Coli, Edu, Jaume, Migue i Trulls. No crec que sigui l'únic que pensa que l'heterogeneïtat de les nostres personalitats té un efecte sinèrgic en cadascun de nosaltres. Podria exemplificar-ho amb mil històries viscudes al llarg d'aquests quatre anys, però em considero incapaç d'expressar tot el valor que li dono en tan sols un parell de línies.

No podria concloure els agraïments sense tornar al punt de partida per veure com decisions d'incerta transcendència poden acabar marcant el nostre camí. Amb la publicació d'aquesta tesi queda palès que una part important de la meva vida ja no es pot explicar sense l'etapa viscuda al GRR.

Abstract

Rheumatoid arthritis, psoriasis, psoriatic arthritis, systemic lupus erythematosus, Crohn's disease and ulcerative colitis are six of the most prevalent immune-mediated inflammatory diseases (IMIDs) and are associated with a high socio-economic impact. There is compelling evidence that IMIDs are genetically complex diseases. To date, however, the genetic component of IMIDs has been only partially explained. Identifying new clinically relevant variation is therefore of major clinical interest. The objective of the present thesis was to identify new genetic variation underlying IMIDs. The research activity here presented is the result of analyzing high-throughput genomic data from a large cohort of IMID patients collected by the IMID Consortium. Using genome-wide approaches and functional analyses, we have identified new genetic variants associated to IMID susceptibility, IMID clinical phenotypes and specific treatment outcomes. Taken together, these findings contribute to better understanding the genetic basis of IMIDs and suggest more specific and preventive therapeutic strategies.

Resum

L'artritis reumatoide, la psoriasis, l'artritis psoriàsica, el lupus eritematós sistèmic, la malaltia de Crohn i la colitis ulcerosa són sis malalties inflamatòries mediades per immunitat (IMIDs) d'elevada prevalença i amb un fort impacte socioeconòmic. Totes elles comparteixen un component genètic important. No obstant, a dia d'avui, només s'ha caracteritzat una part dels factors genètics de les IMIDs. La identificació de factors genètics clínicament rellevants presenta doncs un gran interès clínic per tal d'incorporar la informació genètica a la pràctica mèdica. L'objectiu d'aquesta tesi és identificar noves variants genètiques associades a les IMIDs. La recerca que es presenta és el resultat d'analitzar dades genòmiques d'una gran cohort de pacients amb IMIDs, els quals es van obtenir a través del consorci IMID Consortium. Mitjançant estratègies d'anàlisi de genoma complet i estudis funcionals, en aquesta tesi s'han identificat noves variants genètiques associades al risc de desenvolupar IMIDs així com als seus fenotips clínics i tractament. Aquesta tesi contribueix significativament a la caracterització del component genètic de les IMIDs i, des d'un punt de vista clínic, suggereix noves estratègies terapèutiques.

Thesis outline

The objective of the present thesis is the identification of clinically relevant genetic variation in immune-mediated inflammatory diseases (IMIDs). IMIDs are a group of highly disabling chronic disorders characterized by the activation of multiple immune and inflammatory pathways against the self. From these, six of the most prevalent IMIDs that are associated with a high socio-economic impact are rheumatoid arthritis (RA), psoriasis (PS), psoriatic arthritis (PsA), systemic lupus erythematosus (SLE), Crohn's disease (CD) and ulcerative colitis (UC). These six IMIDs are genetically complex diseases with a strong genetic component. However, the genetic factors underlying IMIDs are still not completely understood. This lack of understanding and the high clinical interest to bridge genetic findings into the medical practice have motivated the development of the present PhD thesis.

For this objective, a large cohort of IMID patients has been used. All these patients were collected by the IMID Consortium during the execution of the singular and strategic *IMID-Kit* project between June 2007 and December 2010. This project led to the IMID Biobank creation, which stores biological samples and detailed clinical information from more than 13,000 IMID patients and 3,000 healthy controls from Spain. The analysis of high-throughput molecular data from these samples has set the starting point for the research activity described in this thesis.

The research line of the present thesis is focused on identifying new genetic variants associated to: (i) IMID susceptibility, (ii) IMID clinical phenotypes, and (iii) specific treatment outcomes in IMIDs. Accordingly, this thesis is divided into three sections. The first section shows the research activity that has been conducted to identify new genetic variation associated with IMID susceptibility. In this section, we present the first genome-wide association study (GWAS) for PS risk at the pathway level (A. Aterido *et al.* J Invest Dermatol '16) as well as the identification of new genetic variation that contributes to the risk of PsA but not purely cutaneous PS (A. Aterido *et al.* Ann Rheum Dis '18 -under review-). The second section shows the research studies that have been performed to

identify genetic factors associated with the development of clinically relevant phenotypes in IMIDs. In this section, we present the first two-stage GWAS on clinical phenotypes in SLE (A. Aterido *et al.* Arthritis Res Ther ’17) as well as the contribution that whole-genome variation has on the risk of developing a cardiovascular disease in IMIDs (PP. Perrotti *et al.* PLoS One ’17). Finally, the third section shows the research work that has been done to identify new genetic markers for anti-TNF treatment in IMIDs. Here, we present a novel multi-omic analysis aimed at identifying genetic markers for treatment response in RA (A. Aterido *et al.* Sci Rep ’18 -under review-), and the first GWAS for treatment immunogenicity in CD (A. Aterido *et al.* Pharmacogenomics J ’18 -under review-).

The research activity presented throughout the thesis has been funded by the Spanish Ministry of Economy and Competitiveness (grant numbers: PSE-010000-2006-6 and IPT-010000-2010-36) and by the “Agència de Gestió d’Ajuts Universitaris i de Recerca” (AGAUR, FI-DGR 2016, grant number: 00587), which is supported by the “Secretaria d’Universitats i Recerca” (Economy and Knowledge Department, Generalitat de Catalunya) and co-funded by the European Social Fund.

Contents

List of figures

List of tables

1 | INTRODUCTION

1.1 Immune-mediated inflammatory diseases

1.1.1 Epidemiology and pathogenesis

The overall prevalence of IMIDs is 5-7% in the general population[17]. Each particular disease has specific and unique epidemiological characteristics (Table 1)[17-32].

IMID	Prevalence (%)	Incidence (per 100,000 ind/year)	Gender (female:male)	Incidence peak (years)
RA	1.0	20-25	3:1 - 4:1	40-50
PsA	0.5	3-23	Balanced	15-20 and 55-60
SLE	0.24	80-230	9:1 - 10:1	16-55
PS	2.0	50-100	Balanced	16-22 and 57-60
CD	0.3	13-20	Up to 1.5:1	20-40 and 55-75
UC	0.5	6-25	Balanced	20-40 and 55-80

Table 1. Main epidemiological characteristics of IMIDs.

IMIDs are complex diseases characterized by a strong genetic component (Figure 1). In complex diseases, multiple genes and environmental factors contribute to disease susceptibility[33].

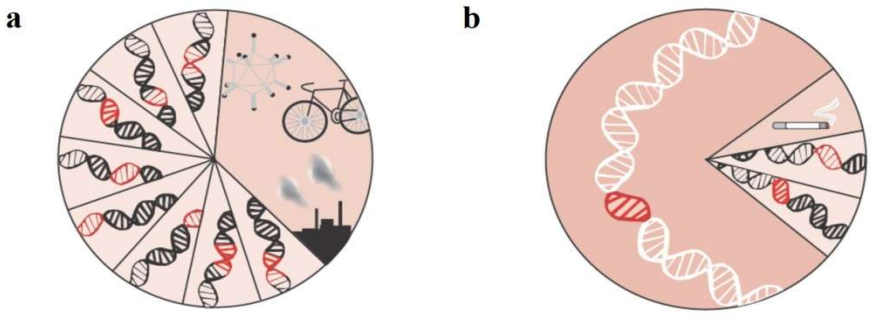

Figure 1. Genetic contribution on complex diseases compared to monogenic diseases. (a) In complex diseases like IMIDs, many genes and environmental factors contribute to disease risk, (b) whereas a single gene is responsible for monogenic diseases, which are poorly influenced by environmental factors. Source: adapted from *Manolio TA et al. J Clin Invest, 2008*[33].

In IMIDs, the mechanisms of immunological tolerance are dysregulated and, consequently, an immune cell reaction is raised against healthy tissues[34]. In normal conditions, T cells expressing both the CD4 and CD8 co-receptors are exposed to a wide variety of self-antigens in the thymus[35]. To avoid self-reactivity, only T cells that bind to self-antigens in the Major Histocompatibility Complex (MHC) with the appropriate affinity are selected. Depending on the MHC class recognized by the T cell receptor, T cells differentiate into CD4+ (MHC class II) or CD8+ T cells (MHC class I)[36]. Following T cell development in the thymus, CD4+ and CD8+ T cells continuously recirculate between the blood and lymphoid organs to find its specific antigen[37].

When CD4+ T cells recognize the antigen in the surface of antigen presenting cells (APCs), CD4+ T cells are activated[38]. The antigen recognition results from a complex interaction between the T cell receptor and the MHC class II expressed on the surface of APCs (Figure 2)[39]. This process requires also the interaction of the CD28 costimulation receptor on T cells with the CD80 and CD86 receptors expressed on APCs[34].

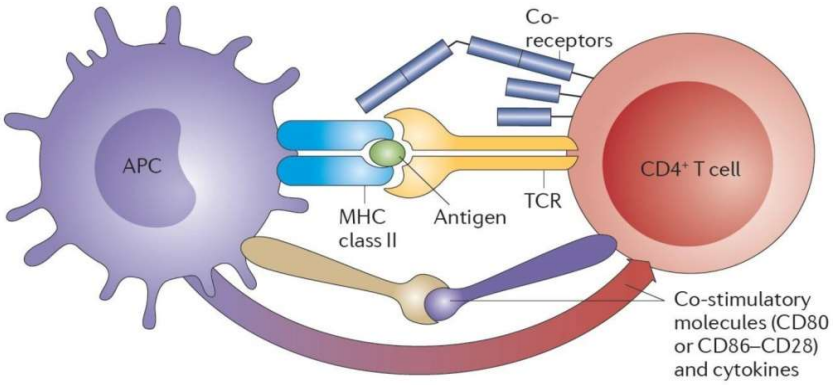

Figure 2. Antigen recognition is the process responsible of immune system activation. Source: *Gutierrez-Arcelius et al. Nat Rev Genet, 2016*[34].

Depending on the cytokine environment, the activated CD4+ T cells can differentiate into diverse helper T (Th) cells, including the Th1, Th2, Th9, Th17, Th22, follicular Th cells (Thf) and regulatory T (Treg) cells (Figure 3)[34].

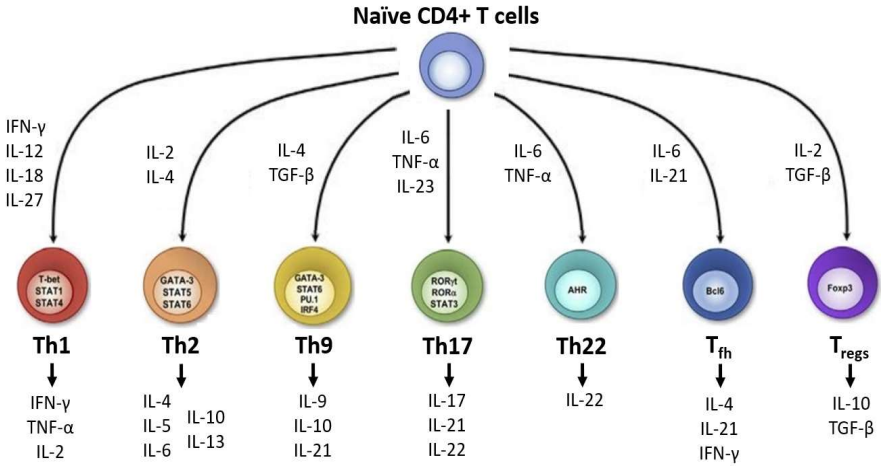

Figure 3. CD4+ T cell differentiation. CD4+ T cells can differentiate into diverse Th cells depending on the cytokine environment. The main subsets of Th cells are Th1, Th2, Th9, Th17, Th22 and Tregs. Source: adapted from *Russ BE et al. Front Genet, 2013*[40].

Th1 cells produce the proinflammatory cytokines TNF-α, IFN-γ and IL-2 as well as chemokines to recruit additional Th cells[41]. The activity of Th1 cells contribute to the activation of neighboring cells like macrophages. Of relevance, macrophages are the main producers of TNF-α[42], one of the key proinflammatory cytokines for IMID development[43,44]. Th2 cells contribute to the activation of eosinophils, mast cells and basophils[45]. Together with Th1 cells, the cytokines produced by the Th2 cell promote the activation of B cells from the bone marrow[46]. Once activated, B cells differentiate into plasma cells that produce specific antibodies against the recognized antigen, thereby triggering a targeted inflammatory response[47]. Like Th2 cells, Th9 cells produce cytokines to support the expansion of Th cells and the activation of mast cells[48]. Th17 cells are the main drivers of autoimmune tissue injury and are characterized by the production of IL-17, a relevant cytokine with potent inflammatory effects that facilitates the activation and recruitment of neutrophils[49]. Synergistically with the IL-17, the IL-22 produced by Th22 promotes pathological inflammation and tissue repair[50]. Thf cells secrete multiple cytokines and costimulatory molecules to assist B cells on the generation of high-affinity antibodies[51]. Tregs are fundamental to downregulate the immune response[52]. The

immunosuppressive activity of Tregs is crucial to reduce the proliferation of other Th cells as well as to inhibit the production of proinflammatory cytokines. The suppressor capacity of Tregs is reached by diverse biological mechanisms like the production of the anti-inflammatory cytokines IL-10 or TGF-β[53].

In addition to the crucial role that CD4+ T cells play on the immune response, CD8+ T cells have also an important contribution on this process[54]. After antigen recognition, CD8+ T cells can differentiate into cytotoxic T (Tc) cells or memory T cells[55]. The former cells release perforin, granulysin and granzymes, cytotoxins that are introduced into the cytoplasm of the target cells to induce apoptotic processes[56]. The latter cells provide long-term protective immunity.

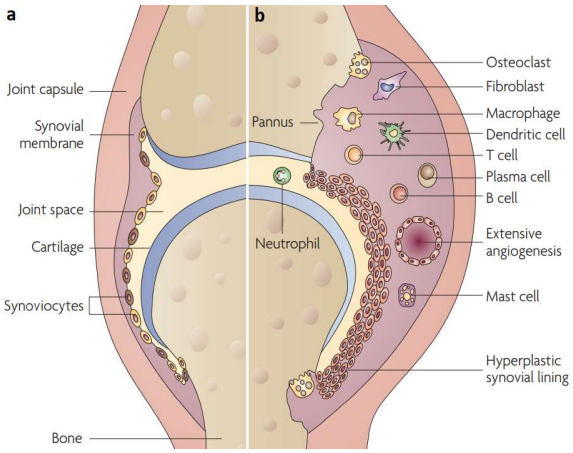

Figure 4. Cell type populations in the healthy joint and rheumatoid arthritis joints. Source: *Strand V et al. Nat Rev Drug Discov, 2007*[12].

Autoimmunity occurs when any of these immunity processes are dysregulated and a misdirected immune response is raised against human tissues leading to tissue destruction[57]. Although the exact etiological causes are unknown, there is compelling evidence that immune cells infiltrate the human tissues and promote a chronic inflammation that can lead to the tissue destruction if not properly treated[58]. In RA, immune cells are aberrantly activated leading to the chronic inflammation of the synovial membrane and the joint destruction[59]. In healthy individuals, the synovial membrane that lines the non-weight-bearing surface of joints contains macrophage-like and fibroblast-like synoviocytes (Figure 4a). In RA, instead, the synovial membrane is infiltrated by multiple immune cell

types including macrophages, dendritic cells, T cells, B cells, plasma cells and mast cells (Figure 4b)[12]. The interplay between these cell populations in the RA synovium drive the chronic phase in the pathogenesis of RA[60]. With disease progression, synoviocytes proliferate and the synovial membrane becomes hyperplastic destroying the articular cartilage and ultimately bone.

PS is characterized by an aberrant hyperproliferation and differentiation of keratinocytes that cause the thickening of the epidermis[61]. The infiltration of T lymphocytes and dendritic cells into the dermis and the presence of hyperplastic blood vessels are also two important hallmarks of PS[62]. In the epidermis, the chemokines produced by keratinocytes stimulate the activation of immune cells[63]. Keratinocytes also release cytokines and growth factors that induce the expression of adhesion molecules for T cells on keratinocytes[64]. In turn, the proinflammatory cytokines produced by the infiltrating immune cells induce the expression of inflammatory genes in keratinocytes and increase the proliferation of this cell type. The complex interaction between epidermal keratinocytes and infiltrating immune cells contributes to the formation of psoriatic plaque lesions in the skin.

In PsA, the dysregulated immune response causes the inflammation of the entheses, distal interphalangeal and sacroiliac joints[65]. Similar to PS and RA, patients with PsA are characterized by a prominent immune cell infiltration into the dermis in the skin and into the synovial membrane in the joint[66]. In both tissues, the immune cells activate mechanisms of chronic inflammation that can lead to the tissue destruction if not properly treated. Importantly, PsA can have dramatic functional implications for the patient and, generally, it requires an alternative therapeutic approach than purely cutaneous PS and RA.

In SLE patients, the aberrant immune system activation can affect diverse organs (e.g. skin, heart and kidney), which is translated into a high phenotypical diversity[24]. One of the key concepts in pathogenesis of SLE is the release of fragmented cellular material, including nuclear antigens,

by apoptotic cells[67]. The nuclear antigens can activate nucleic acid recognition receptors and, consequently, trigger an inflammatory response through the production of type I interferons[68]. The type I interferon promotes the differentiation of B cells and the formation of autoantibodies that lead to a loss of immune self-tolerance[69]. The apoptotic cells can also stimulate the release of other proinflammatory cytokines for the recruitment of immune cells. From these immune cells, neutrophils and T cells play a crucial role in autoreactivity processes. More specifically, T cells produce IL-17 that facilitate the tissue infiltration by neutrophils[70], immune cells that are responsible for the organ damage observed in SLE patients.

IBD is a chronic relapsing inflammatory disease of the intestine that is classified into two different disease entities: CD and UC. Patients with CD have segmental and transmural inflammation in any part of the gastrointestinal tract[71], whereas the inflammatory processes are limited to the colonic mucosa in UC patients[72]. In normal conditions, the innate and adaptative immunity in the gastrointestinal tract are balanced through complex interactions with the microbiota under homeostatic conditions[73]. In IBD, however, this homeostasis is disrupted. Together with the crucial role that Th1 and Th2 cells play for the pathogenesis of IBDs[74], the imbalance between Th17 cells and Tregs is also an important contributor to explain the chronic intestinal inflammation observed in patients with CD and UC[75].

Despite the increasing knowledge on the pathogenesis of IMIDs, to date, the biological mechanisms that contribute to the development of RA, PsA, PS, SLE, CD and UC are still not completely understood. The analysis of high-throughput genomic data from IMID patients offers a valuable opportunity to advance in the characterization of the pathophysiology underlying IMIDs.

1.1.2 Clinical phenotypes

IMIDs share not only inflammatory pathways, but also the presence of comorbidities and clinical phenotypes that can worsen the patients' quality of life[17]. Comorbidities are additional diseases that coexist with a particular disease[76]. Clinical phenotypes are disease manifestations or traits that describe differences among patients with the same disease. Both comorbidities and clinical phenotypes have important implications for disease diagnosis and treatment.

In IMIDs, clinical studies have demonstrated that SLE is a highly heterogeneous disease[24]. Patients with SLE can present a wide range of disease phenotypes that affect different organs, including the kidney, skin or heart (Figure 5a)[24]. Similarly, PsA can also present multiple clinical phenotypes, including skin and nail disease, dactylitis, uveitis, and osteitis[77]. Given the lack of robust molecular approaches for the diagnosis of IMIDs, the disease clinical phenotypes are currently used as diagnostic criteria in IMIDs[78].

Recent research using large cohorts of IMID patients have found that treatment efficacy is influenced by the presence of comorbidities like obesity, a prevalent disease that can affect up to 50% of IMID patients[79-83]. Obese patients have shown not only a more severe disease activity than non-obese patients[81], but also a significantly higher risk of anti-TNF treatment failure[84]. So far, however, little is known on the etiology of comorbidities in IMIDs. Understanding the biological mechanisms that underlie IMID comorbidities and also clinical phenotypes could guide treatment stratification as well as the development of more efficient and phenotype-specific treatments in IMIDs.

One of the most important comorbidities in IMIDs is the development of a cardiovascular disease (CVD). The common biological process underlying CVD is atherosclerosis, the artery narrowing resulting from complex cellular interactions in the intima layer. In this process, immune cells, cytokines (e.g. TNF-α) and antibodies accumulate in the arterial wall leading to the narrowing of the arterial lumen and the eventual thrombosis

(Figure 5b)[85]. The interplay between the adipose tissue and the immune system has been found to be crucial for the atherosclerotic plaque formation[86]. There is growing evidence that adipocytes not only increase the synthesis of proinflammatory cytokines but also radically diminish the secretion of adiponectin, an anti-inflammatory and cardioprotective protein[87,88]. Consistently, evidence from national registers suggests that patients treated with anti-TNF agents have a lower probability to develop CVD[89]. Taken together, these findings highlight the importance of the inflammatory state in patients with IMIDs for the development of a CVD.

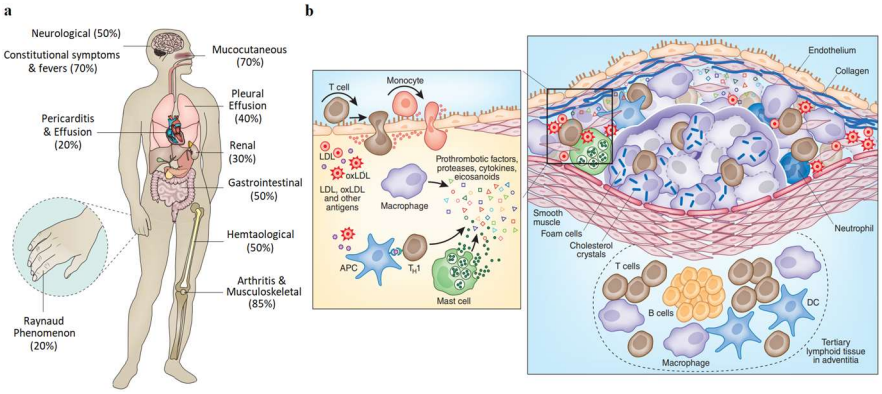

Figure 5. Clinical phenotypes in immune-mediated inflammatory diseases. (a) Clinical heterogeneity in SLE. The phenotypical variability is represented by the affectation of different organs together with the frequency of the most common manifestation. (b) Immune cells involved in the formation of the atherosclerotic plaque that leads to CVD. Source: adapted from *Hansson GK et al. Nat Immunol, 2011*[85]; *Kaul A et al. Nat Rev Dis Primers, 2016*[90]

To date, multiple epidemiological and clinical factors have been associated with CVD risk, including dyslipidemia, arterial hypertension and obesity. However, these classical risk factors only partially explain the elevated risk observed in IMIDs[91]. Growing evidence suggests that genetic factors could also contribute to the risk of developing a CVD[92]. Given that CVD is the leading cause of mortality worldwide[93], understanding the biological mechanisms of this relevant comorbidity is a priority for the public health system to reduce the socioeconomic costs of IMIDs.

1.1.3 Socioeconomic impact

IMIDs are chronic diseases that have functional implications for the patient and, therefore, it has a negative impact on the patient's quality of life[94]. These relapsing disorders are a major burden to society. For example, the annual economic burden of RA has been estimated on €45.3 billions in Europe and €41.6 billions in the USA[95]. In PsA, the annual cost per patient has been shown to range from €9,475 to 14,800 in European countries[96]. The annual cost of IBDs has been estimated to be approximately $12,000 per patient[97]. The economic costs of IMIDs are generally classified into healthcare and non-healthcare costs[98].

Healthcare costs include medications, hospitalizations and clinic visits, among others[99]. In SLE, the disease heterogeneity has a strong impact on the annual healthcare costs, reaching up to $70,000 per patient and increasing with renal involvement[100]. The medical costs of PS have been estimated on $12 billion in the North American population[101]. In Canada, the annual medical costs of IBDs are higher than $1.2 billion[97]. The healthcare costs of IMIDs like RA have been mostly dominated by the inpatient care[102]. With the advent of anti-TNF drugs, which are effective but much more costly than previous disease-modifying drugs, the contribution of medications on the overall healthcare costs has substantially increased[103]. In US, the average annual cost of anti-TNF therapy per treated patient ranges from $15,000 to $24,000[104]. In PsA, the introduction of this effective therapy has triplicated the healthcare costs[96].

Non-healthcare costs represent indirect expenses primarily associated with decreased work productivity, disability payments and early retirements that are also substantial[94]. In PS, the loss of productivity observed in patients from the North American population has an estimated cost of $114 million per year[105]. In IBDs, non-healthcare costs have been estimated to be $1.6 billions, which mainly result from a long-term work loss[97]. Therefore, identifying clinically relevant genetic variation for the development of new therapeutic approaches is key to increase the patients' quality of life and, in turn, to reduce the high socioeconomic cost of IMIDs.

1.1.4 IMID Consortium

Figure 6. Trademark of the IMID Biobank.

The IMID Consortium (IMIDC) is a Spanish network of biomedical researchers focused on the study of the molecular basis of IMIDs through the analysis of high-throughput data. The IMIDC was created during the execution of the *IMID-Kit* research project (Singular and Strategic Project funded by the Spanish Ministry of Economy and Competitiveness) and was launched in 2006. The creation of this network of clinical researchers is essential to collect large cohorts of patients with standardized phenotype measurements. Coordinated by Prof. Sara Marsal (GRR-VHIR), this national biomedical consortium is still ongoing and is currently composed by more than 90 rheumatology, dermatology and gastroenterology clinical departments from all over Spain.

The need to store biological samples from a very large cohort of patients led to the creation of the IMID Biobank (Figure 6). The IMID Biobank is a storage infrastructure that ensures the robust processing, storage and delivery of the biological samples from IMID patients. To date, the IMID Biobank stores biological samples from more than 13,000 IMID patients and 3,000 healthy controls from Spain. For each biological sample, more than 350 clinical and 150 epidemiological variables are available for analysis. The integrative analysis of high-throughput genomic data from this large cohort of IMID patients with their corresponding clinical and epidemiological information has been the starting point for the present thesis.

1.2 Genomics

1.2.1 The human genome

One of the major achievements of science has been the characterization of the human genome sequence. In February 2001, the first reference sequence of the human genome was released as a result of two independent public and private projects[106,107].

The DNA molecule

The human genome is a DNA sequence of 3×10^9 base pairs (bp) that is distributed in 23 chromosomes, including 22 autosomal chromosomes and two sexual chromosomes. Humans are diploid organisms and therefore each individual carry two sets of 23 chromosomes, one inherited from the father and the other one inherited from the mother.

In 1953, James Watson and Francis Crick discovered that the DNA molecule is composed by two strands of nucleotides that are coiled forming a three-dimensional double helix structure (Figure 7)[108]. Nucleotides are organic molecules that contain a five-carbon sugar deoxyribose, a phosphate group and a nitrogenous base. The nucleotide chain of each strand is formed by a covalent bound between the sugar of one molecule and the phosphate of the next. To make the double-stranded DNA, the nitrogenous bases of the two separate polynucleotide strands are bound together by hydrogen bonds according to base pairing DNA rules.

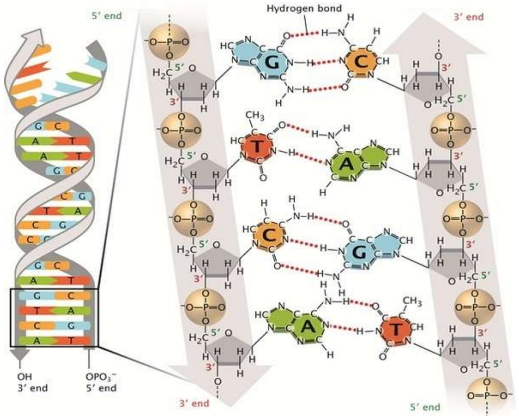

Figure 7. DNA molecular structure. Base pairing of thymine with adenine and guanine with cytosine. Source: *Pray L, NatEducat, 2008*[5].

The code of life

The human genome encodes the information that determines the biological functions of human cells. After the elucidation of the DNA structure, the mechanism by which genetic information flows within a cell and is used to synthetize proteins was discovered in 1970[109]. As stated by the central dogma of the molecular biology, this biological mechanism consists of two steps: the transcription and translation. In transcription (Figure 8a), once the DNA has been replicated, DNA sequences that encode for human genes are converted to messenger RNA molecules (mRNA). Unlike the DNA, the mRNA is a single-stranded molecule that contains the sugar ribose instead of deoxyribose and integrates the nitrogenous base uracil instead of thymine. The transcription of DNA to mRNA determines the gene expression profile of the human cells. In translation (Figure 8b), the mRNA sequence is interpreted by ribosomal enzymes that synthetize the encoded proteins. In this process, groups of three consecutive nucleotides (i. e. codons) are translated into specific amino acids using the genetic code.

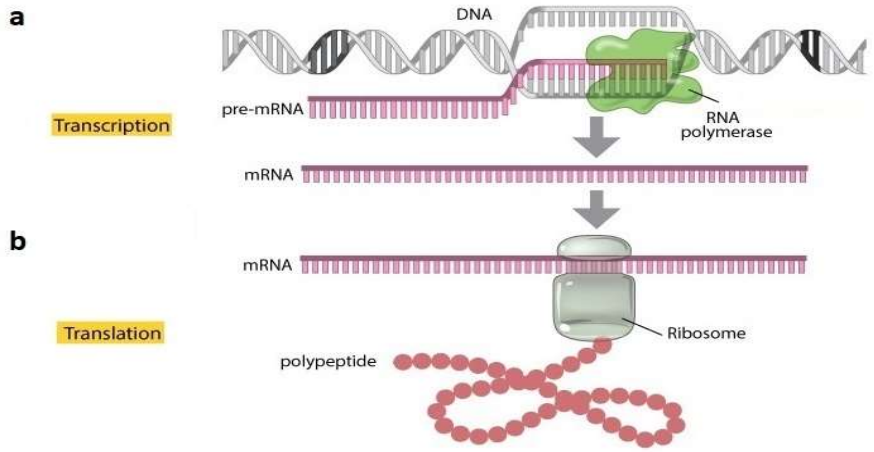

Figure 8. Central dogma of molecular biology. Non-coding regions are removed from the transcribed pre-mRNA and exons are subsequently assembled to form the mature mRNA during the splicing process. The mature mRNA is finally converted to an amino-acid sequence. Source: adapted from *Clancy S et al. Nat Educat, 2008*[6].

DNA sequence organization

The DNA sequence is a cryptic store of information that is commonly divided into coding and noncoding DNA[110]. Coding DNA are sequences that can be transcribed into mRNA and translated into proteins, whereas non-coding DNA includes all sequences that are not used to encode proteins.

Coding DNA represents the most widely studied component of the human genome. The complete protein-coding capacity of the genome is contained within the exome. Exons and introns (i.e. DNA segments located within protein-coding sequences that are copied into mRNA molecules but not translated into amino-acids) compose the structure of the human genes. The ~20,000 protein-coding human genes that have been characterized so far represent only a small fraction of the human genome (i.e. <2%)[111,112].

Non-coding DNA includes the sequences that are located outside protein-coding sequences and are not translated into proteins. The large proportion of non-coding DNA (i.e. >98% of the human genome) was initially described to have no biological function and, controversially, it has been called junk DNA since 1960[113]. Instead, recent evidence has shown that non-coding DNA has important functions on the regulation of genetic activity in human cells[114]. To date, the non-coding DNA forms that have been identified include: (i) non-coding RNAs; (ii) regulatory DNA elements; (iii) introns; (iv) telomeres and centromeres; (v) transposons and retrotransposons; (vi) untranslated regions; (vii) DNA methylated regions; and (viii) pseudogenes. The major advances in the comprehension of the non-coding DNA have been achieved by the Encyclopedia of DNA Elements project[115]. This worldwide research project has had the daunting task of identifying the functional elements encoded by the human genome as well as revealing the effect that genetic variation has on gene expression and disease development.

1.2.2 Human genetic variation

A striking observation from human genome sequencing is that the human population exhibits approximately 99% of genetic similarity[116]. The remaining genetic variation is an important contributor to the phenotypic variability observed in humans, which can predispose to disease risk, clinical phenotypes or treatment outcomes[117].

A single nucleotide polymorphism (SNP) is the most abundant form of genetic variation that occurs once every 100-300 bp[116]. Each SNP is a variation of two or more nucleotides at a single position in the DNA sequence. Given that most SNPs have two alleles, they are generally considered biallelic. According to the diploid nature of the human genome, each individual can therefore carry three possible genotype combinations. The allele composition, frequency, chromosome position and functional impact are the main properties of SNPs. The minor allele frequency (MAF) has been the predominant strategy for classifying the SNPs (Figure 9) into mutations (MAF<1%), low frequency variants (1%<MAF<5%) and common variation (MAF>5%).

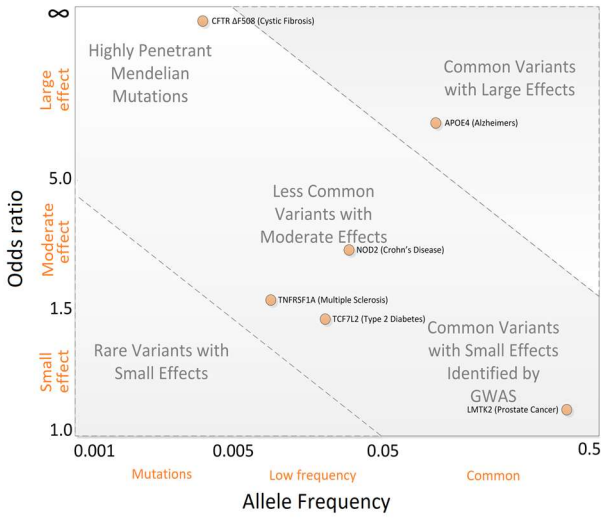

Figure 9. Spectrum of allele effects according to their frequency. Mutations have high penetrance and low frequency. Conversely, common variants have small effects on phenotypes. Source: adapted from *Bush WS et al. PLoS Comput Biol, 2012*[11].

At the functional level, for example, exonic SNPs can be classified into synonymous (i.e. SNPs that do not lead to a change in the amino-acid sequence), non-synonymous missense (i.e. SNPs that lead to an amino-

acid change in the resulting protein) and non-synonymous nonsense (i.e. SNPs that lead to a gain or a loss of a stop codon that produces a shortened or elongated version of the protein, respectively).

The SNP Database (dbSNP, www.ncbi.nlm.nih.gov/snp) is the reference public archive of genetic variation that is hosted by the National Center for Biotechnology Information[118]. The dbSNP lists more than 324 million variants found in sequenced human genomes, 15 million of which are present at frequencies of 1% or higher across different populations.

The vast majority of genome variation consists of SNPs. However, the human genome contains also an estimated 3,700 structural variants that affect DNA segments of ≥1 Kb[119]. In addition to the MAF, functional impact and chromosome position, structural variants are also classified into copy-number variants, inversions and translocations.

Origin of genetic variation

Genetic population studies have shown that genetic variation emerges as a consequence of two main formation mechanisms: DNA mutations and genetic recombination[120]. Importantly, only genetic variation that emerges in germ cells can be inherited[121].

A mutation is a permanent DNA alteration that can be originated by internal errors during DNA replication and repair (Figure 10a), transposable elements (Figure 10b) and environmental damaging agents like ultraviolet radiation[14]. Although single mutations can have large effects, the human evolution is mainly based on the accumulation of multiple small-effect mutations[122]. The occurrence of mutations in the somatic tissue of an organism is mostly harmless and can be accumulated in different cell types leading to local molecular changes in human tissues[123].

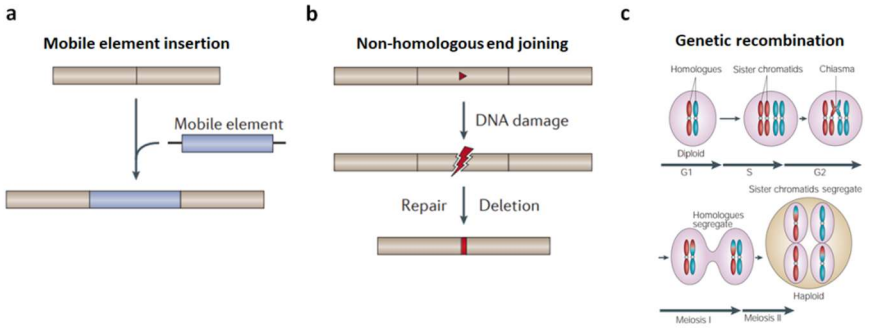

Figure 10. Principal mechanisms to generate genetic variation. (a) mobile element insertion by retrotransposition, (b) non-homologous end joining process during DNA repair, and (c) genetic recombination. Source: adapted from *Weischenfeldt J et al. Nat Rev Genet, 2013*[14]; *Marston AL et al. Nat Rev Mol Cell Biol*, 2004[16].

Genetic recombination occurs during the production of germ cells in cell meiosis (Figure 10c)[124]. In this process, the chromosome pairs of each parent are mixed to increase the genetic diversity of the offspring cells leading to new chromosome sequences. Nonetheless, the probability of recombination is not equal along the chromosome[125]. The physically closer that two alleles are located in a chromosome, the lower the probability that a genetic recombination event occurs between them. Consequently, alleles that lie close in the chromosome sequence are more likely to be inherited together compared to distant alleles or alleles from different chromosomes[11]. The allele correlation or linkage disequilibrium (LD) is commonly defined as a correlation coefficient between two loci (r^2, ranging from 0 to 1). The LD determines groups of highly correlated alleles referred as haplotypes (Figure 11)[126].

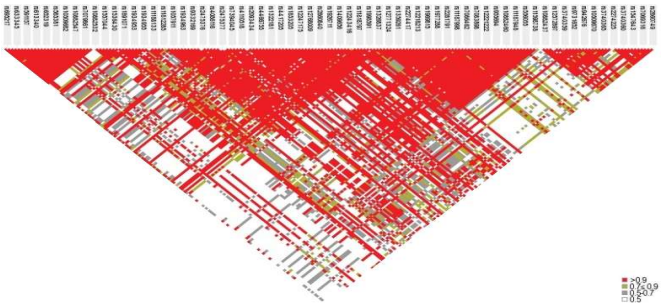

Figure 11. Definition of haplotype blocks based on LD. Correlation patterns between adjacent SNPs allow to define haplotype blocks (i.e. red triangles). Source: adapted from *Walton R et al. Nat Genet, 2005*[7].

1.2.3 High throughput genotyping: microarray technology

Microarray genotyping platforms allow the simultaneous genotyping of genome-wide genetic variants (*N*>500,000 SNPs) with >99% accuracy[127]. The principle underlying the microarray technology is the ability of nucleotides to bind to their complementary bases. If genetic variants were independent, genotyping technologies would have to determine each polymorphism along the genome. Instead, the high correlation existent among SNPs from a certain haplotype makes it unnecessary to genotype the complete genome variation[127]. According to the existence of highly correlated SNPs, genome-wide genotyping microarrays determine the genotype of a representative SNP for each LD block. These SNPs, known as tag SNPs, have set the basis for the development of GWAS to study the genetic basis of complex human diseases (Figure 12).

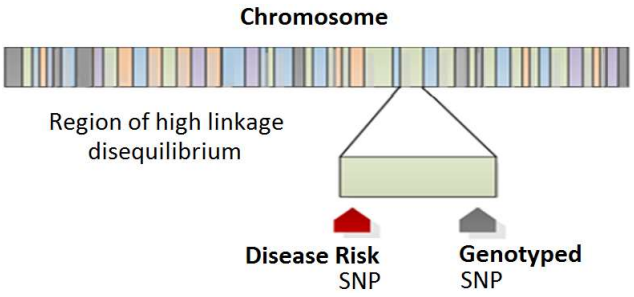

Figure 12. Underlying principle of genome-wide genotyping technologies. The associated SNPs are generally in linkage disequilibrium with the causal SNPs. Source: adapted from *Bush WS et al. PLoS Comput Biol, 2012*[11].

The desire for precise genomic mapping encouraged the development of two reference projects for genetic studies: the International HapMap project (www.hapmap.org)[128] and the 1000 Genomes project (1KG, www.1000genomes.org)[129]. The HapMap project was launched soon after the completion of the human genome sequence to catalog common SNPs and determine the LD relationships across the genome for guiding genetic studies. The 1KG project was launched with the goal of providing a more extensive catalog of variation to the scientific community (i.e. common variants and variants with a MAF<5%). The genomic data released by these reference projects have allowed the characterization of the human

LD patterns and, consequently, have paved the way for designing high throughput microarray genotyping platforms.

The first commercial genotyping platform targeted ~1,500 SNPs and was released by Affymetrix in 1996 (Santa Clara, CA, US)[130]. Since then, multiple manufacturers have commercialized high-density genotyping microarrays[131]. From these, Illumina has been the most widely-used platform to study the genetic component of human diseases due to its superior quality and coverage[11]. The genome-wide genotyping of the IMID patients analyzed in the present thesis has been performed using the Illumina platform.

In the last years, genotyping microarrays have gradually increased their capacity (Table 2). To date, for example, the *Illumina Infinium Omni5Exome-4 BeadChip Array* allows the genotyping of >4.3 million variants. The increasing genome coverage that expands to rare variants and the capacity to analyze large sample collections at an affordable cost has made the genotyping microarray a commonly used technology to study the genetics of human traits.

Illumina genotyping microarray	**Markers#**	**Samples#**
HumanCytoSNP-12 DNA Analysis BeadChip	299,140	12
Human660W-Quad v1 DNA Analysis BeadChip	657,366	4
HumanOmniExpress BeadChip	731,442	12
Human1M-Duo DNA Analysis BeadChip	1,199,187	2
HumanOmni1-Quad BeadChip	1,140,419	4
HumanOmni1S-8 BeadChip	1,200,000	8
HumanOmni2.5-Quad BeadChip	2,450,000	4

Table 2. Illumina genotyping microarrays. Number of genetic markers and samples that can be genotyped by each microarray. Source: adapted from *Lamy P et al. Hum Genomics, 2011*[131].

1.2.4 Genome-wide association studies

Genetic linkage studies scan genotype-phenotype associations by analyzing the allele segregation within family individuals[132]. This analytical approach has been the predominant methodology used to characterize the genetic component of monogenic disorders in the latter half of the twentieth century[133]. However, the results of genetic linkage studies proved hard to reproduce for complex traits[134]. Genome-wide genotyping has allowed to shift from genetic linkage studies in families to GWAS at the population level[135]. GWAS allow to analyze the association between genetic markers and the development of complex diseases. Since 2005, this statistical approach has successfully identified a large number of genetic variants that predispose to complex human diseases[135]. Recently, a statistical approach that is able to analyze genetic associations at the pathway level has been developed[136]. The pathway-based GWAS is starting to provide previously unsuspected biological mechanisms for human diseases that will contribute to identify new drug targets and develop more effective therapeutic strategies. To date, over 3,000 GWAS have identified more than 60,000 single markers associated with complex traits[137].

Single-marker genome-wide association studies

Study design

In GWAS, the association analysis between genetic variation and the phenotype of interest is performed following a population-based study design. In population-based studies, genetic associations are assessed by analyzing the phenotype and allelic frequency distribution in large cohorts of unrelated individuals[135]. Given that large cohorts are easier to be recruited, population-based studies can be more powerful to identify genetic risk variation than family-based studies[138,139]. The golden standard

for discarding false positive results is replication in an independent cohort from a different population (Figure 13)[140].

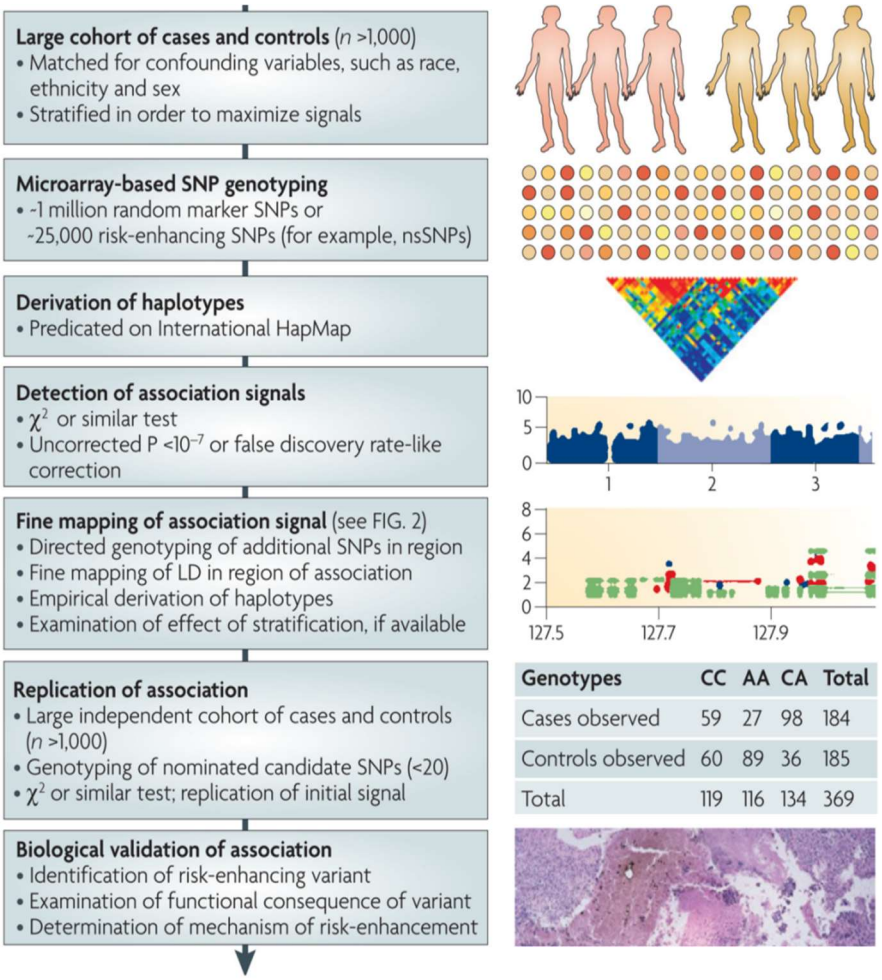

Figure 13. Overview of the general study design of GWAS. In the discovery phase, patients are genome-wide genotyped and the genotype data is tested for association with the phenotype of interest. The most significant associations are subsequently analyzed in a replication cohort. Finally, the validated variants are studied at the functional level to investigate how they are linked to the studied phenotype. Source: *Kingsmore SF et al. Nat Rev Drug Discov, 2010*[15].

There are several factors that can lead to irreproducibility, including variable phenotype definition, disease heterogeneity, population-specific LD, population stratification, model misspecification, insufficient sample size or heterogeneous bias in the estimated effects detected among the discovery and replication cohorts[15]. Therefore, considering these factors can substantially increase the statistical power of the genetic association analysis.

Statistical methodology

A major component of the success of GWAS is the use of rigorous criteria for clinical classification[141]. In addition, the statistical model should be adjusted for those factors that are known to influence the phenotype tested for association. This adjustment reduces spurious associations due to sampling artifacts or biases in study design at the cost of using additional degrees of freedom that can impact the statistical power of the analysis[142].

One of the most important confounding factors to be considered in GWAS is population stratification. This stratification results from the presence of systematic differences in allele frequencies that are due to ancestry[143]. GWAS using patients and control subjects from diverse geographic areas or ethnicities can therefore lead to false positive and negative associations[144]. To avoid these spurious associations, population stratification must be addressed. To date, different statistical approaches have been devised to prevent population stratification. Principal component analysis is the most commonly used approach to adjust for population stratification. With this strategy, genetic variation is used to compute the main axes of variation (i.e. principal components)[145]. In multiple studies and populations, the first principal components derived from this analysis have been found to mirror the geographical distribution of the samples (Figure 14). Accordingly, these variables can be used as covariates in the statistical model or to exclude samples showing an outlier genetic background.

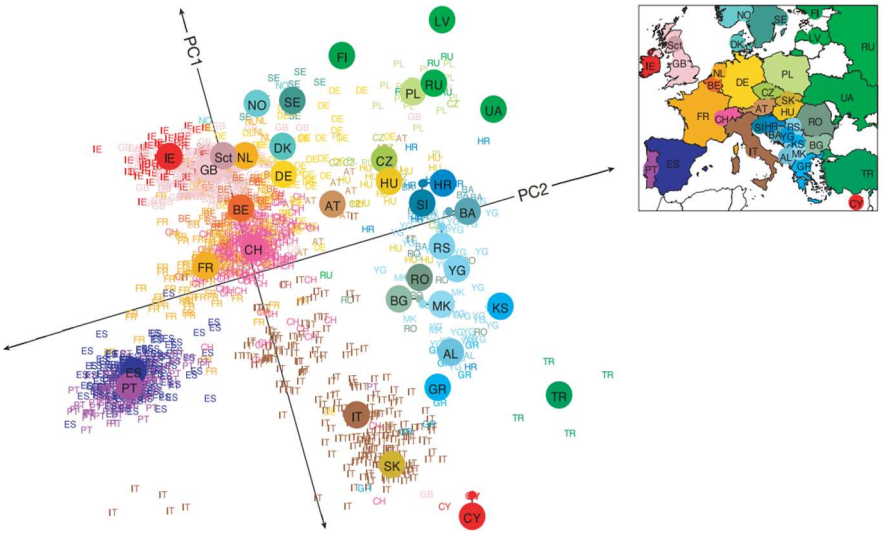

Figure 14. Genetic substructure of the European population. The geographical map of Europe arises as an efficient two-dimensional summary of genetic variation in the European population. This population stratification needs to be considered in GWAS. Source: adapted from *Novembre J et al. Nature, 2008*[1].

In GWAS, thousands of SNPs are simultaneously tested for association and, consequently, the association results must be adjusted by multiple testing. The probability of rejecting the null hypothesis when it is true (i.e. type-1 error rate, false positive detection) is generally controlled by setting a significance level α equal to 0.05 for a single test. In multiple testing, the probability of the type-1 error can be estimated as follows[146]:

$$P=1-(1-\alpha)^{n}$$

As shown, the probability of the type-1 error is a function of the number of statistical comparisons that are performed (n) as well as the significance threshold (α). Therefore, the probability of identifying a false positive is extremely high in GWAS. Based on a Bonferroni correction that assumes ~1 million independent SNPs along the human genome, the standard significance threshold to identify a genuine genotype-phenotype association has been established as $\alpha=5\times10^{-8}$.

Statistical power

The statistical power of GWAS depends on the sample size of the study cohort, effect size, allele frequency and statistical model[147]. GWAS can be underpowered to detect associations of modest effect sizes (i.e. odds ratio ~1.1-1.5) and large cohorts are required to detect variants of moderate effect (i.e. odds ratio ~1.5-2)[148]. In order to increase the statistical power of the analysis, GWAS meta-analysis and genotype imputation have emerged as efficient strategies.

GWAS meta-analyses can increase the power to detect new risk variants by increasing the sample size of the study cohort[149]. In meta-analysis, the statistical significance and effect size of a particular marker is combined across distinct studies[150]. Genotype imputation can increase the power of the statistical analysis to identify the causal variant by examining additional non-directly genotyped markers at the genome-wide scale. Genotype imputation exploits the human LD patterns from densely genotyped reference panels (e.g. 1KG project) to estimate the genotype of non-directly genotyped SNPs (Figure 15)[9]. Generally, most imputation methodologies start by computing the haplotypes within the study samples. Using these estimated haplotypes, the genotypes of the non-directly genotyped markers are then estimated by extrapolating LD patterns from the reference panel of the sample ancestry. Optionally, the estimated genotypes can be weighted according to a probability score that is based on the haplotype overlap[151,152].

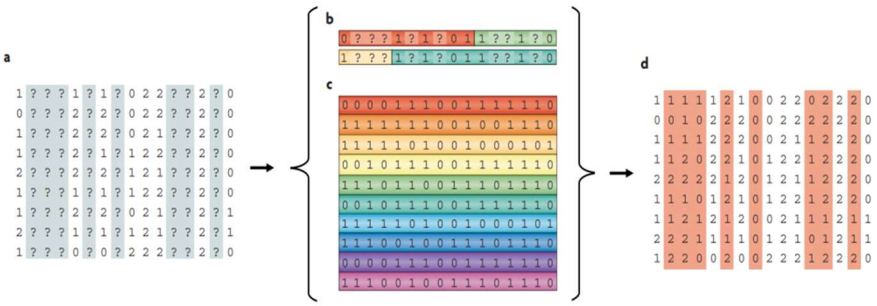

Figure 15. Genotype imputation algorithm. (a) Genotyped data with eight non-directly genotyped SNPs represented as question marks (b, one individual shown) are phased and (c) compared to the dense haplotypes in the reference panel (d) to estimate the non-directly genotyped SNPs. Source: adapted from *Marchini J et al. Nat Rev Genet, 2010*[9].

Genetic insights from genome-wide association studies

After a decade of GWAS in humans, these studies have led to the identification of many thousands of variants reproducibly associated with complex human diseases (Figure 16).

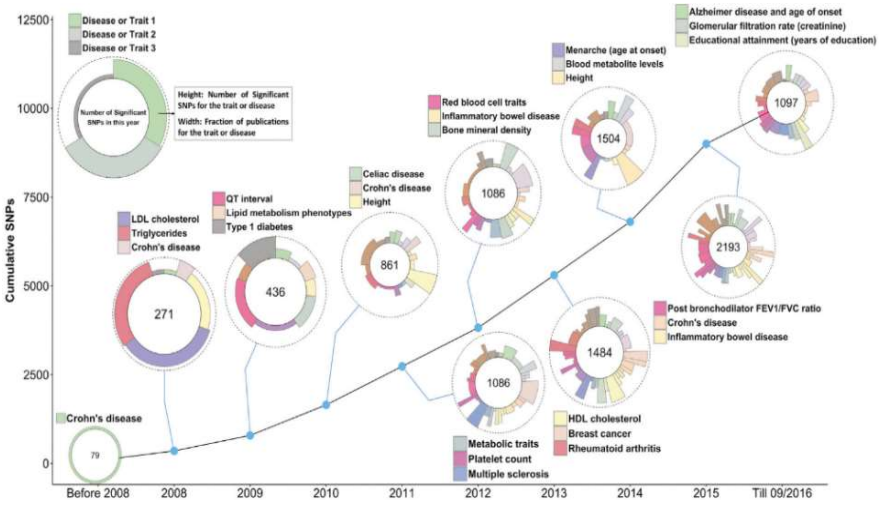

Figure 16. Timeline of disease risk genes discovery using GWAS. For each year of discovery, only the three traits with the largest number of SNPs are shown in the circle ($P<5\times10^{-8}$ and $r^2<0.5$). Source: *Visscher PM et al. Am J Hum Genet, 2017*[153].

While still ongoing, GWAS have already led to remarkable findings[153]:

- Complex diseases are frequently associated with non-coding regions. Only ~10% of disease associations lie in protein-coding regions[154].
- Pleiotropy is pervasive. Hundreds of loci are significantly associated with multiple traits[155].
- Disease risk loci explain less than 30% of the phenotypic variance in complex diseases. These findings indicate the existence of disease missing heritability[156].

- Complex diseases are highly polygenic. Multiple loci contribute to the susceptibility to human diseases[157]. Generally, each SNP explains a small proportion of the phenotypic variance[158].
- Susceptibility genes can guide drug discovery. Genetically supported candidate targets can lead to double the success rate in drug development (Figure 17)[159].

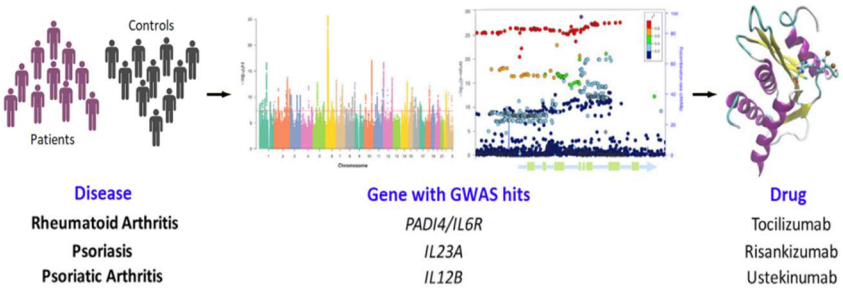

Figure 17. Genetic evidence supporting drug indications in IMIDs. Source: adapted from *Visscher PM et al. Am J Hum Genet, 2017*[153].

- Sharing of genetic data enables new research discoveries. The public availability of genetic data is fundamental for the scientific community to replicate and meta-analyze GWAS results. The main public resources of GWAS data are: (i) GWAS catalog, a manually curated collection of all known SNP-trait associations[158]; (ii) dbGAP, a resource of genotyping and phenotypic data from several published GWAS[160]; and (iii) UK Biobank, a repository of GWAS and phenotypic data from 500,000 individuals (www.ebi.ac.uk/ega/studies).

The missing heritability of complex diseases

One of the major challenges in the genetics of complex diseases like IMIDs is to characterize the remaining missing heritability. There is evidence that single-marker GWAS are underpowered to detect genetic variants with small effect sizes unless extremely large cohorts are used[148]. In order to characterize these small-effect genetic factors, alternative analytical approaches need to be devised.

In addition to common genetic variants with small effect sizes, other factors could also contribute to explain part of the estimated missing heritability[161]. Genetic factors like rare variants (MAF<0.01) and copy number variants represent potential factors to contribute on the heritable basis of complex traits[162]. Epigenetic factors, epistasis and genotype-environment interactions represent additional contributors on the missing heritability of complex traits[163-165].

Pathway-based genome-wide association studies

Genes do not work isolatedly in human cells. Instead, they act within complex molecular networks and cellular pathways[166]. Functionally related genes have been shown to predispose to disease susceptibility, including loci that do not reach individually the genome-wide significant threshold[167]. Statistical approaches that are able to analyze genetic associations at the pathway level have been developed[168]. These pathway-based approaches are analytical strategies that integrate genetic and biological information to test if sets of genes are jointly associated with complex traits[169].

Unit of analysis

In pathway-based GWAS, genetic variation is integrated with biological information. For this objective, SNPs are mapped to genes and these are subsequently aggregated into biological pathways. In order to perform this preprocessing step, three methodological aspects must be considered:

- Pathway annotation data. Multiple biological annotation databases currently offer the possibility to select curated pathways to be tested for association (www.pathguide.org). The Molecular Signature Database is a public repository that provides genetic pathways representing the universe of common biological processes for meaningful interpretation of large-scale genomic data[170]. To date, a total of 217, 674 and 186 pathways that have been manually curated by

expert biologists are collected in the BioCarta, Reactome and Kyoto Encyclopedia of Genes and Genomes, respectively. These genetic pathways can be used in pathway-based GWAS to identify new biological processes associated with human diseases.

- SNP-Gene mapping. The annotation of genetic variants to genes and pathways is a complex issue. To date, proximity-based criteria are clearly the dominant approach[171]. The definition of gene boundaries includes the coding region as well as the upstream and downstream boundary regions (i.e. SNP-gene distance windows range from 10 to 500Kb)[172,173].

- Pathway size. Genetic pathways including an extremely large or small number of SNPs can give false positive results.[174,175] To avoid these spurious associations, the predominant strategy is to focus the analysis on pathways with a gene content typically ranging from 10 to 300 genes.

Statistical methodology

The statistical methodology used for pathway-based GWAS is designed to aggregate the association evidence from multiple genetic variants into a single association statistic. This methodology can be classified according to the input genomic data analyzed and the null hypothesis tested for association:

- Input genomic data. The required input data set for pathway-based strategies can be a collection of SNP p-values or SNP genotype data (Figure 18a)[176].

- Null hypothesis. The null hypothesis tested for association in pathway-based GWAS can be: (i) self-contained, indicating that pathway genes are not expected to be associated with the phenotype of interest, or (ii) competitive, indicating that genes from the analyzed pathway and the rest of the genome are expected to show the same magnitude of association with the phenotype of interest (Figure 18b)[176].

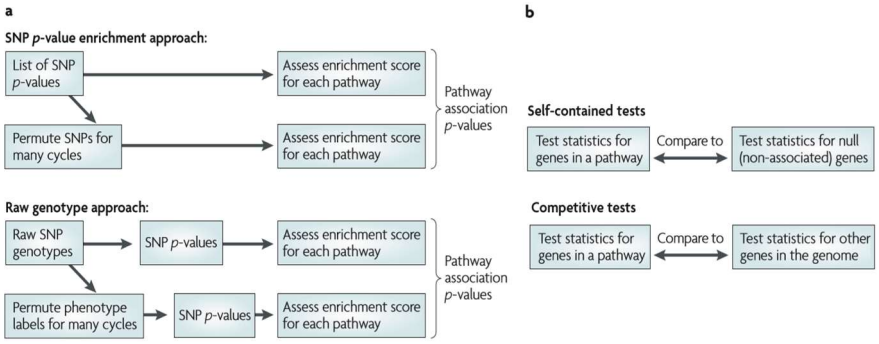

Figure 18. Pathway-based GWAS methodology. (a) Pathway-based approaches analyze SNP p-values or raw genotype data and are classified as (b) self-contained or competitive according to the null hypothesis. Source: adapted from *Wang K et al. Nat Rev Genet, 2010*[136].

To date, multiple analytical methods have been developed (Table 3)[177-187]. Although there is no consensus about the most appropriate approach, the set-based test implemented in PLINK software is currently one of the predominant strategies for pathway-based GWAS[184].

Name	Input data	Hypothesis tested
ALIGATOR	P-values	Competitive
i-GSEA4GWAS	P-values	Competitive
GenGen	Genotype data	Competitive
GESBAP	P-values	Competitive
GRASS	Genotype data	Self-contained
GSA-SNP	P-values	Competitive
GSEA-SNP	Genotype data	Competitive
PLINK set-test	Genotype data	Self-contained
SNP ratio test	Genotype data	Competitive
MAGENTA	P-values	Competitive
PARIS	P-values	Self-contained

Table 3. Pathway-based GWAS methods. List of the main publicaly available methodologies for pathway-based analysis on GWAS datasets. Source: *Wang K et al. Nat Rev Genet, 2010*[136]*; Ramanan VK et al. Trends Genet. 2012*[173].

The statistical test implemented in PLINK software tests a self-contained null hypothesis using SNP genotype data. In this method, independent SNPs are first identified using the LD structure. Unlike other methods, the PLINK method calculates the LD structure using the genotype data from the cohort of study. Accounting for LD is crucial to avoid an increase in

false positive results due to the presence of highly correlated SNPs. The observed pathway statistic is then computed as the average of the statistics of the associated pathway SNPs and it is contrasted to the null distribution generated using a permutation-based approach. The resulting statistical association must be interpreted according to:

- Multiple testing correction. Genetic pathways from publicly available databases can share genes[188]. According to this gene overlap among pathways, the multiple testing correction should not be over-conservative (i.e. stringent Bonferroni correction). The False Discovery Rate is the most commonly used method to control the expected proportion false positives[136].
- Genetic architecture of the disease. Pathway-based GWAS can be powerful to identify new clinically relevant variation when a particular disease results from the interplay of multiple small-effect genes[136]. Instead, the presence of large-effect loci within common biological pathways might lead to false positive pathway associations[189]. In such a case, this approach might not be informative.
- Replication of significant genetic pathways. Like single-marker GWAS, the associated pathways need to be replicated in an independent cohort. In pathway-based GWAS, multiple genes from genuinely associated pathways are expected to be consistently replicated[190].

To date, pathway-based GWAS has been successfully used to characterize the genetic component of complex human traits like psychiatric disorders, including bipolar disorder and schizophrenia[191,192], obesity[193], height[194], and diverse cancer subtypes[195-197]. This has led to novel hypotheses for therapeutic intervention on the basis of the involvement of new biological pathways[198]. However, the pathway-based GWAS has not been exploited to study the genetic basis of IMIDs.

From pathway-based analysis to network medicine

As a polygenic analytical approach, the pathway-based GWAS requires a systems biology interpretation of the resulting associated pathways. To investigate the downstream mechanisms by which genetic variation at the pathway level cause phenotypic changes, pathway genes can be connected according to their functional association and create a biological network[199,200]. A biological network is a graph model that represents biological pathways as nodes and edges, corresponding to molecular entities and their pairwise relationships, respectively[8]. Given that complex biological networks are likely to underlie most genotype-phenotype associations[10], the analysis of the functional connectivity among genes from disease-associated pathways in the form of complex biological networks can have relevant implications for the medical practice (Figure 19)[201]. In translational research, this analytical approach can also provide valuable insights for drug discovery[202].

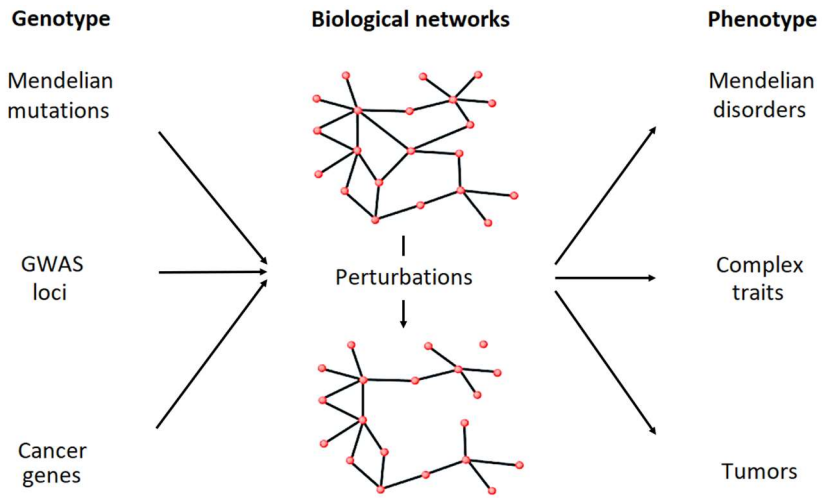

Figure 19. Genetic perturbations in biological networks underlie phenotypic alterations. GWAS loci can perturb the normal pathway topology and functionality thus determining the phenotype developed. Source: adapted from *Vidal M et al. Cell, 2011*[10].

The nature of the nodes (genes, proteins or metabolites) and edges (protein-protein interactions, gene-gene interactions, drug-target interactions or gene-gene functional associations) depends on the type of

data used for the network building[203]. The release of experimental data from high-throughput technologies and computational strategies has led to the creation of multiple network-based datasets. Most of these datasets are accessible through diverse public databases (Table 4).

Database	Data type	#Species	URL
HPRD	PPI, PHI	1	http://hprd.org
BioGRID	PPI, GI, CA	64	http://thebiogrid.org
DIP	PHI	10	http://dip.doe-mbi.ucla.edu/dip/
STRING	PPI, PHI, CP	2,031	https://string-db.org/
MINT	PPI	90	https://mint.bio.uniroma2.it/
IntAct	PPI	>10	https://www.ebi.ac.uk/intact/
TRED	TFG	3	http://rulai.cshl.edu/TRED

Table 4. Public biological network databases. Main publicly available databases of human networks. Networks can be built using protein-protein interactions (PPI), genetic interactions (GI), chemical associations (CA), physical interactions (PHI), transcriptional factor-gene interaction (TFG) or computational predictions (CP). Source: adapted from *Yu D et al. Genomics Inform, 2013*[8].

The non-random organization and structure of biological networks has shown that network topology encodes information about how molecular connections contribute to complex phenotypes[204]. The main topological measures that are used to characterize the structure of biological networks are shown in Figure 20[203]. In network analysis, nodes with the highest

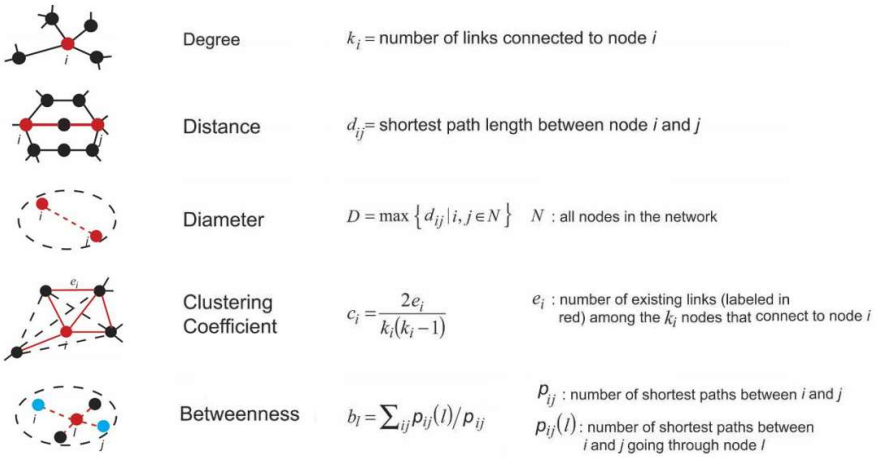

Figure 20. Main topological measures from biological networks. Graph and formulae illustration of five topological parameters. Source: adapted from *Vidal M et al. Cell, 2011*[10].

centrality properties have also the highest probability to be essential for the network functionality[205-207]. There is compelling evidence that degree centrality (DC) and betweenness centrality (BC) are two of the most important centrality properties of biological networks. The DC measures the number of edges that connect a node[208], while the BC measures the extent to which a node connect subnetworks[209]. Based on the DC distribution, most of the human biological networks has been found to be scale-free[210]. This statistical property refers to networks whose DC distribution follows a power law instead of a Poisson distribution[211]. In scale-free networks, the majority of nodes have only a few interactions (i.e. non-hubs) and coexists with a few highly connected nodes (i.e. hubs). *In silico* network analysis has demonstrated that the random removal of non-hubs does not lead to observable changes in the biological network structure, while the removal of hubs significantly changes the structure and function of the network[212]. Therefore, the study of the network centrality properties is of a major interest to convert genetic findings into translational opportunities.

The emerging network-based medicine is starting to provide new insights into the pathogenesis of complex diseases[213]. For example, the enrichment analysis of subnetworks overlapping GWAS risk loci in differentially expressed genes has provided new insights into biological mechanisms underlying Type 1 Diabetes pathogenesis[214]. Also, the characterization of the network distribution of *de novo* mutations in autism spectrum disorders has led to the identification of a protein subnetwork that participates in β-catenin/chromatin remodeling[215]. At the pharmacological level, deliberate network perturbation has been suggested as a potential therapeutic strategy in cancer[216]. Taken together, these findings support the analysis of the biological networks identified by the pathway-based GWAS as a powerful strategy to discover new genes and biological mechanisms underlying complex diseases.

Network-based approaches are also a promising strategy to advance in three of the main research lines of drug discovery, including multi-targeting therapy, drug repositioning and side effect prediction. For

example, the analysis of the drug action at the network level has enabled to identify new synergistic drug combinations for the treatment of breast cancer (Figure 21)[3].

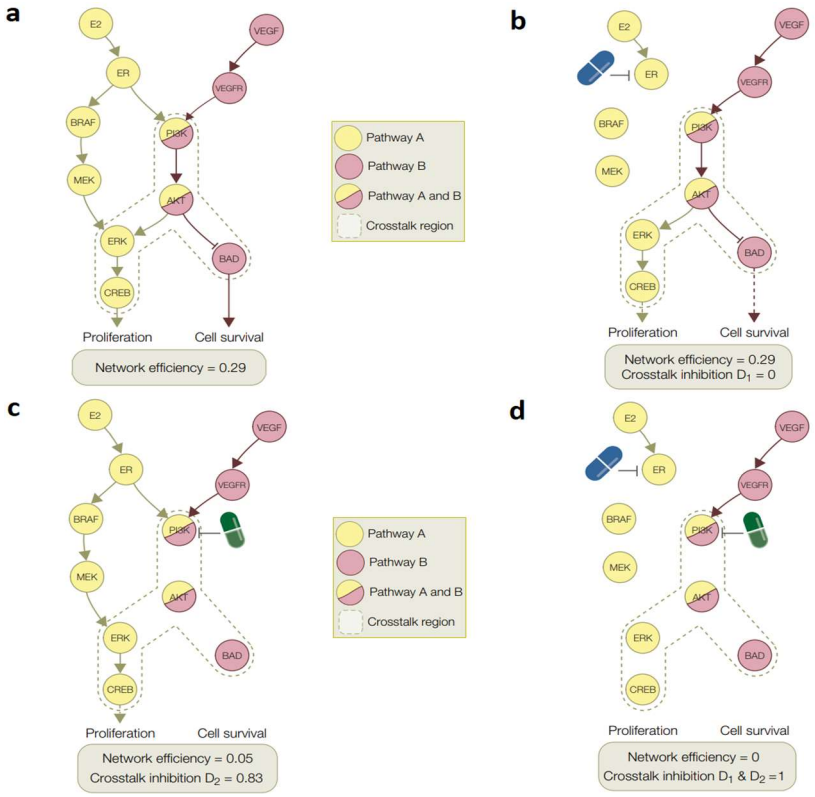

Figure 21. Novel network-based approach to discover anti-cancer drug combinations. The use of multiple drugs that inhibit shared or downstream protein interactions within two pathways (i.e. crosstalk inhibition) aimed at reducing the flow information within the pathways (i.e. network efficiency) has proven successful. Source: adapted from *Jaeger S et al. Cancer Res, 2017*[3].

Also, the analysis of drug-target networks has shown that most of the GWAS risk genes are not directly targeted by common drugs, but the distance from these genes to drug targets is shorter than expected by chance[217,218]. Therefore, mapping GWAS risk genes on drug-target networks can help to identify new indications for conventional drugs. Finally, network modeling has proven successful for side effect prediction, which exemplifies the potential of network-based approaches to shorten the overall time of drug development[219].

Integration of genotype and multi-omic information

There is an increasing effort to integrate GWAS datasets with additional layers of molecular information[220]. High-throughput technologies have expanded the breadth of available omic information from whole-genome genotype data to epigenomic, transcriptomic, metabolomic and proteomic data[221-225]. Although the analysis of GWAS data provides a comprehensive view of disease associations at the genetic level, the cross-talk between multiple layers of molecular data can provide deeper insights into the relationship between genetic variation and the biological mechanisms underlying complex traits (Figure 22)[2,226].

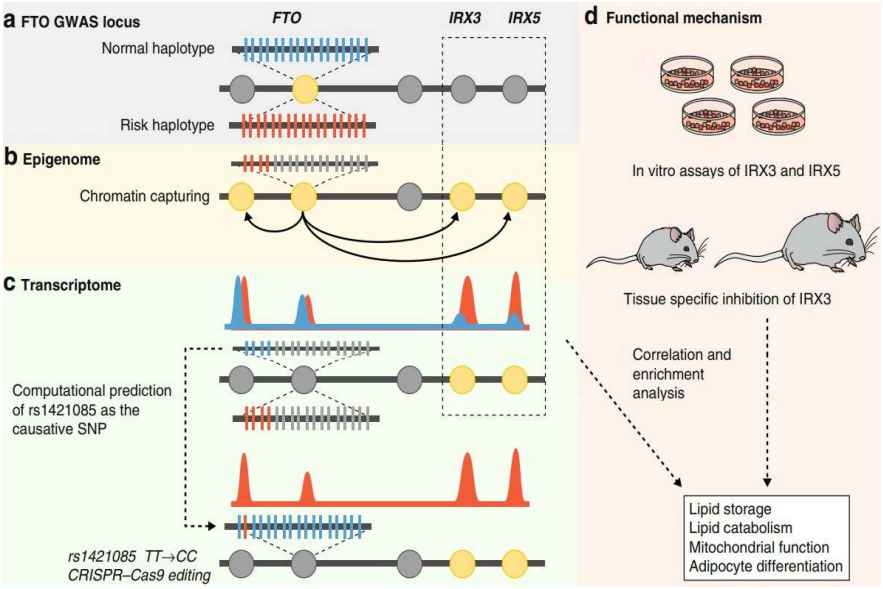

Figure 22. Multi-omic integrative analysis to identify regulatory mechanisms in obesity. The combination of genomic, epigenomic and transcriptomic data has enabled to identify functional mechanisms mediating the genetic effect at the *FTO* locus in obesity. Source: *Claussnitzer M et al. N Engl J Med, 2015[2]; Hasin Y et al. Genome Biol, 2017[4].*

In 2005, transcriptomic and GWAS data were integrated for the first time using the expression quantitative trait loci (eQTL) approach. This analytical approach led to the identification of relevant genes for complex traits like obesity[227]. Since then, eQTL mapping studies have provided functional hypotheses to explain the linkage between GWAS loci and the

disease phenotypes via gene expression regulation[228,229]. In addition, GWAS loci mapping to non-coding DNA regions have been shown to co-localize with eQTL regulatory elements[227], thereby showing the importance of non-coding variation for disease susceptibility.

To date, multiple studies have combined genetic and epigenetic data to investigate the genetic regulatory basis of the human epigenome[230-235]. Certain epigenetic marks have shown a significant phenotypic cell-type specificity (e.g. H3K4me3 and H3K27Ac)[236,237]. Analyzing the overlap of GWAS associated variants in these cell-type specific epigenetic marks has emerged as a powerful approach to discover relevant cell types for the development of complex traits[238]. With the increasing epigenomic data for primary cells and tissues derived from the reference human epigenome collection[223], the combination of these information with genetic variation is starting to provide functional and causal insights into autoimmunity[237].

Similarly, other omic data types like metabolite and protein abundance have been integrated with GWAS datasets. At the metabolomic level, genetic variation has been strongly associated with the metabolite abundance. The associated regions have shown a significant overlap with disease risk loci[239,240]. At the protein level, recent genome-wide quantitative trait loci mapping studies have shown that alterations in the proteome driven by genetic polymorphisms can influence the development of particular human traits[241].

1.2.5 Genetic component of IMIDs

Before the advent of GWAS, only a few large-effect genes had been associated to IMID susceptibility. The use of GWAS has enabled to identify IMID risk loci with modest effect sizes[34]. More recently, GWAS are starting to be used to identify genetic factors in IMIDs that underlie the risk to develop phenotypes of clinical relevance[242-244].

Genetic susceptibility to disease risk

Familial aggregation and twin studies have demonstrated a marked sibling recurrence rate (λ_s) in IMIDs (e.g. PsA λ_s~37, SLE λ_s 8-29 and PS λ_s~7)[245-249]. These findings indicate the existence of a strong genetic basis for IMID risk. In addition, there is compelling evidence that more than 50% of the IMID risk can be attributed to genetic factors[34].

The MHC locus was the first genetic risk factor identified for most IMIDs. This locus harbors human leukocyte antigen (*HLA*) genes encoding for proteins that are crucial to trigger the adaptive immune response[250]. In most of the IMIDs, the MHC locus contributes to disease risk more significantly than any other known locus, with relative risks generally greater than 3[251]. Considerable work prior to the GWAS era was focused on identifying genetic associations outside the MHC, particularly in immune-related pathways. Through linkage studies and candidate-gene analyses, several risk loci outside the MHC region were identified. From these, *PTPN22*, *CARD14*, *IRF5* and *NOD2* genes represent some of the most strongly associated loci with the risk of developing RA, PS, SLE and IBD, respectively[252-255].

GWAS have radically improved our knowledge of the genetic variability associated with the risk of developing IMIDs. To date, GWAS have identified at the genome-wide level of statistical significance more than 100 risk loci for IBDs and RA, more than 50 genes for PS and SLE risk as well as 15 PsA risk loci (Figure 23)[256-260]. For each IMID, these disease risk loci (i.e. genome-wide significant SNPs with $P<5\times10^{-8}$ and MAF≥0.01) explain less than 30% of the disease heritability[156]. Similar to

other complex diseases, most of the IMID risk loci have shown also modest effect sizes, with relative risks ranging from 1.1 to 1.5[250]. Taken together, these findings highlight that small-effect genes could be important contributors to explain the remaining missing heritability of IMIDs.

One of the major challenges in the study of the genetics of IMIDs is to identify genetic variation that contributes to the risk of PsA but not purely cutaneous PS. GWAS have shown that most of the established PsA risk loci are also shared with PS[261-268], indicating that the biological processes that cause autoimmunity to the skin are also central for PsA. To date, only *PTPN22*, *CSF2-P4HA2* and *ADAMTS9-MAGI1* loci have been found to be PsA-specific. Considering the effects of all known PsA and PS risk loci[269-271], less than 25% of the PsA heritability is currently explained. The identification of additional PsA-specific genes could raise new pathogenic mechanisms underlying PsA and, in turn, it could lead to the development of more effective therapies.

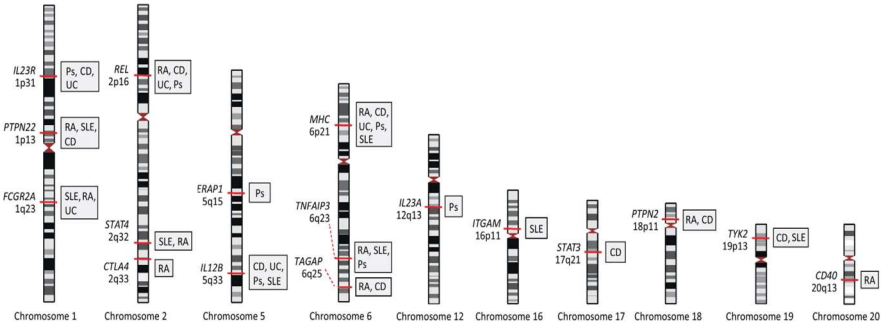

Figure 23. GWAS loci associated IMIDs risk. This illustration shows some examples of GWAS loci associated with IMID risk as well as the overlap between many of them. Source: adapted from *Knight JC, Cell, 2013*[13].

GWAS not only have contributed to discover hundreds of common variants associated with IMID risk, but also to refine disease risk associations with the *HLA* region[272]. The *HLA* is one of the most complex loci in the human genome that is characterized by the presence of highly

correlated SNPs (i.e. high LD), high gene density and large sequence diversity[273,274]. Stemming from the high LD existent in this locus, multiple SNPs could yield equivalent statistical evidence of association thereby preventing the identification of the causal variant[275]. The application of imputation methodologies and conditional association analyses on GWAS data offers a valuable opportunity for in-depth investigation of the *HLA* locus to identify genetic variants with independent effects on disease risk[276].

GWAS results have also revealed that a substantial part of disease risk factors is shared across IMIDs[277]. These complex cross-phenotype associations are not necessarily in the same direction (i.e. a risk allele for a particular IMID can act as a protective allele for other IMID) and underscore the importance of pleiotropy in IMIDs. Pleiotropy occurs when a genetic variant affects more than one phenotypic trait[278]. Accordingly, the use of genome-wide approaches embracing this prevalent property should increase the power to discover clinically relevant variation in IMIDs. Furthermore, detecting pleiotropic variants could have important implications for drug discovery based on repurposing strategies.

The simultaneous analysis of genetic and transcriptomic data has provided additional insights into the genetics of IMIDs[279]. Recent studies have found that some eQTLs are only present under active inflammatory states and disappear after disease treatment[280]. These findings demonstrate the existence of genetic variation with functional effects that are context dependent[281]. Accordingly, we have previously identified genetic regulatory variants that are key for the activity of CD4+ T cells in RA *(Annex* section*)*[282]. To date, the genetic regulatory mechanisms by which causal variants alter the cell functionality to induce IMID risk are still under intense study[283].

Genetic susceptibility to clinical phenotypes

IMIDs are characterized by a high phenotypical diversity and the development of important comorbidities. GWAS have been mostly performed using cohorts of patients without considering the disease phenotypical heterogeneity. However, the clinical phenotypes in IMIDs have also been shown to aggregate in families, thereby suggesting the existence of a genetic component for disease heterogeneity[284,285]. In order to characterize the genetic factors underlying IMID clinical phenotypes, the study design shifts from a case-control design towards a case-case design (Figure 24). To date, only candidate-gene studies and a few GWAS have been conducted to uncover the genetic basis of the IMID clinical heterogeneity.

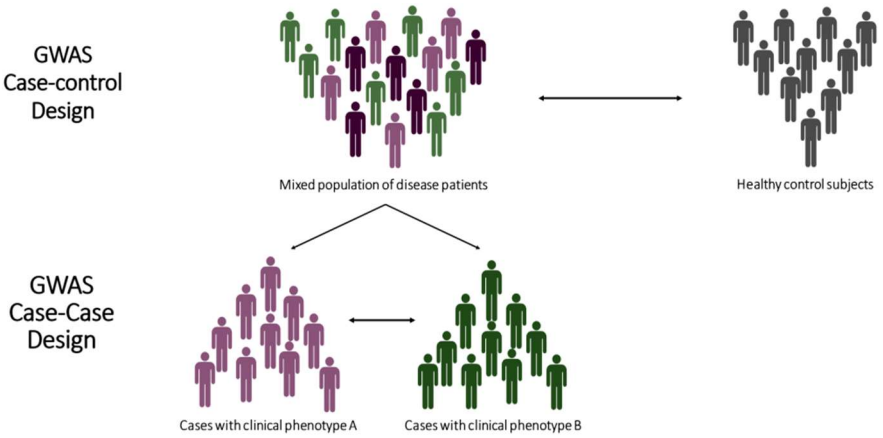

Figure 24. Case-control and case-case study designs used in GWAS. Different analytical strategies employed to identify genetic variation associated with either disease risk or clinical phenotypes. Source: adapted from *Verstockt B, Clin Transl Immunology, 2018*[286].

At the single gene level, different clinical phenotypes in CD have been consistently associated with variation in *NOD2* gene[287,288]. In SLE, the most strong genetic associations with clinical phenotypes have been detected between renal disorder and *ITGAM*, *STAT4* and *TNFSF4*, malar rash and *FCGR2A*, and between hematological disorder and *IL21*[289,290]. In PS, genetic variation at *LCE3D*, *IL1RN* and *GJB2* genes has been

associated with severity, nail disease and plaque development, respectively[291,292]. Very recently, a cross-sectional analysis conducted in PsA has found that the *HLA-B*27* allele is associated with enthesitis, especially in patients with longer disease duration[293]. Similarly, the *HLA-B*0801* allele has been found to be associated with radiographic sacroiliitis in patients with PsA[294]. In RA, a functional variant of *TLR10* gene has been associated with erosive disease in patients seropositive for anti-citrullinated protein antibodies (ACPA)[295]. Of relevance, this genetic variant is not associated with susceptibility to RA, thereby supporting family aggregation studies on the existence of a genetic basis for clinical heterogeneity that is independent from the overall disease risk.

At the genome-wide scale, seven loci have been associated with CD phenotypes including *MAGI1*, *CLCA2*, *2q24.1*, *LY75*, *NOD2*, *MST1* and the *HLA* region[242,296]. The unique GWAS that was conducted for PS phenotypic variability identified variants at *HLA-C*, *IL23R*, *LCE3A*, *TNFRSF9* and *TNFAIP3* showing a significantly different effect in PsA than in PS[244]. In RA, the *SLC8A3* gene has been identified as a new risk locus for ACPA-positivity using a GWAS strategy[243]. Before the development of the present thesis, no GWAS for SLE phenotypes had been performed.

CVD has an increased prevalence in patients with IMIDs. Despite this evidence, the genetic basis of this comorbidity has been poorly investigated. Using case-control cohorts of healthy individuals and CVD patients, the large number of GWAS performed so far have been focused on characterizing the genetic risk basis of CVD in the general population[92]. These studies have led to the identification of more than 300 genetic variants for CVD[297]. To date, however, little is known on the impact that these variants have on the risk of developing a cardiovascular event in IMID patients. Likewise, no comprehensive study has investigated the impact that IMID susceptibility variants have on CVD risk.

1.3 Pharmacogenomics

Pharmacogenomics studies the association between genetic variation and the clinical response or adverse reactions to a particular therapy. Anti-TNF therapies have revolutionized the treatment of IMIDs. However, there is a large fraction of patients that do not respond to anti-TNF therapy. Also, anti-TNF treated patients can produce antidrug antibodies. Given the high costs of the anti-TNF therapy and the availability of alternative drugs, it is of a major clinical interest to understand the genetic and biological mechanisms underlying treatment failure.

1.3.1 Anti-TNF therapy in IMIDs

Until the end of the 20th century, patients with IMIDs were treated with non-steroidal anti-inflammatory drugs, disease-modifying antirheumatic drugs and glucocorticoids[298]. In the last two decades, active research in the biotechnology sector has led to the development of anti-TNF agents. The TNF-α is a proinflammatory cytokine that is mainly produced by macrophages in acute and chronic inflammation. Multiple studies have demonstrated the crucial role that TNF plays to hyperactivate proinflammatory signaling pathways in IMIDs.

The three main anti-TNF drugs that have been predominantly used for the treatment of IMID patients include (Figure 25): infliximab, adalimumab and etanercept. Infliximab is a chimeric monoclonal antibody with a murine variable region (25%) bound to the Fc region of a human IgG1 (75%) that effectively blocks the TNF binding to its soluble and membrane receptors[299]. The Food and Drug Administration (FDA) has approved the clinical use of infliximab for the treatment of PS, PsA, RA, CD and UC. Adalimumab is a fully humanized IgG1 monoclonal antibody that blocks the TNF binding to both its soluble and membrane receptors[300]. Adalimumab has been approved by the FDA for the treatment of CD, PS, PsA and RA. Etanercept is a recombinant fusion protein containing the TNF receptor 2 that inhibits the TNF binding to its

cell surface receptors rendering TNF biologically inactive[301]. Etanercept is currently approved and indicated for the treatment of RA, PsA and PS.

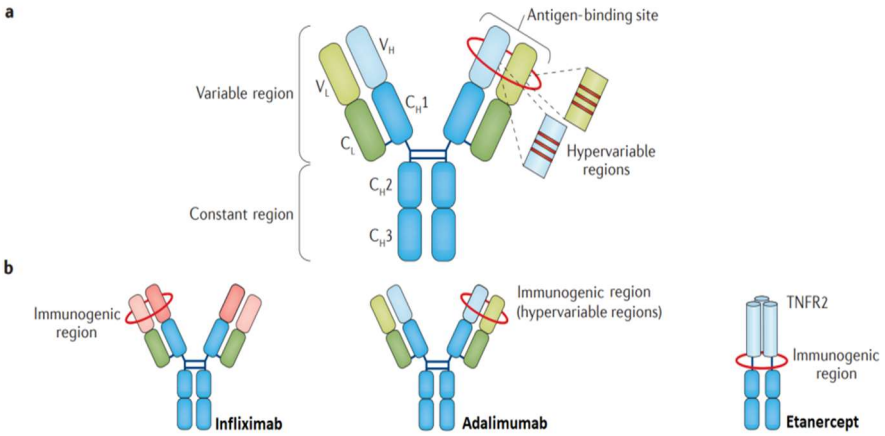

Figure 25. Structure of the main anti-TNF drugs. (a) Structure of an antibody molecule. The variable region forms the antigen-binding site, the constant region is responsible for the effector functions and the hypervariable regions determine the antigen specificity. (b) Main anti-TNF drugs used in IMIDs. Source: adapted from *Kalden JR et al. Nat Rev Rheumatol, 2017*[302].

The introduction of anti-TNF drugs has revolutionized the treatment of IMIDs (Table 5)[44], demonstrating a marked improvement in the patients' quality of life as well as a significant reduction in the number of surgeries and hospitalizations[303,304]. There is, however, a substantial fraction of anti-TNF treated patients that do not show a significant clinical improvement. In RA, ~30% of anti-TNF treated patients do not respond to the therapy[305]. In PsA, a non-significant clinical response to anti-TNF treatment has been observed in up to 40% of the treated patients[306,307]. In PS, approximately one-third of the anti-TNF treated patients do not respond to the treatment[308]. In IBDs, up to 50% of CD patients experience a loss of clinical response over time that will require either dose escalation or treatment discontinuation[309], and 30-40% of UC patients fail to achieve a significant clinical improvement after anti-TNF therapy[310]. Identifying the genetic and biological mechanisms that influence the treatment efficacy is therefore of high interest to guide the search for response biomarkers.

Clinical Feature	Infliximab	Adalimumab	Etanercept
Efficacy in RA	+++	+++	+++
Efficacy in PsA	+++	+++	+++
Efficacy in CD	+++	+++	+
Efficacy in UC	+++	+++	na
Efficacy in PS	+++	+++	++
Administration	iv	sc	sc
Dosage	3-10 mg/kg; q4-8w	40 mg eow; 40 mg qw	25 mg biw; 50 mg qw
Half-life (t ½)	8-10 days	10-20 days	4 days
V_d (L)	4.3 ± 2.5	4.7 - 6.0	8.0
Clearance (mL/h)	11.0	12.0	72.0 ± 5.0
C_{max} (µg/mL)	118.0	4.7 ± 1.6	1.1 ± 0.6

Table 5. Clinical profile of the anti-TNF drugs used in IMDs. Abbreviations: biw, twice a week; Cmax, maximum serum concentration; eow, every other week; iv, intravenous; na, not applicable; qw, every week; sc, subcutaneous; vd, volume of distribution; w, week; +++, strong; ++, moderate; +, weak. Source: adapted from *Tracey D et al. Pharmacol Ther, 2008*[44].

One of the main causes of anti-TNF treatment failure is the production of antidrug antibodies (ADAs). Anti-TNF agents are extraneous proteins and, therefore, this therapeutic strategy is inherently immunogenic. The treatment of IMIDs with anti-TNFs often requires lifelong administrations and, consequently, it can raise an immune system response against these drugs. This results in the formation of ADAs, which can make anti-TNF therapy ineffective.

Immunogenicity studies in anti-TNF treated IMID patients have shown that ADAs can reduce the drug efficacy by competing with the endogenous ligand and/or by forming immune complexes, which accelerates the drug clearance and the subsequent reduction of the drug bioavailability[311,312]. In IMIDs, the presence of ADAs has been associated with a reduced clinical response[313-315].

To date, several studies have detected ADAs in patients with IMIDs, including IBD and RA[315,316]. Although the quantification of ADAs partially depends on the type of assay, infliximab has been established as the most immunogenic anti-TNF therapy (~25% of infliximab-treated patients). Anti-adalimumab antibodies have been detected in approximately 14% of patients. The lowest immunogenicity rate has been detected in etanercept-treated patients (1.2%)[315]. This low

immunogenicity is likely to be explained by the molecular structure of etanercept, which is characterized by the lack of an Fc region[44].

The formation of ADAs might also lead to adverse effects like acute infusion reactions, delayed infusion reactions and disseminated skin reactions[317,318]. Although these adverse events and the reduction of the anti-TNF efficacy associated with the development of ADAs are of major clinical relevance, the biological processes that are responsible for the formation of ADAs are still unclear. In IMIDs, the study of patient-related factors like genetic variation that influences the development of ADAs could provide novel insights into this specific pharmacogenomics field[319].

1.3.2 Genetic basis of clinical response to anti-TNF therapy

There is increasing evidence that genetic variation is an important factor to explain the heterogeneity on the clinical response to anti-TNF therapy in IMIDs[320,321]. To date, most of the pharmacogenomic studies of anti-TNF response in IMIDs have been focused on candidate genes from inflammatory pathways[320]. In RA, genes participating in TNF and NFKB signaling pathways and genes associated with T cell function like *IRAK3*, *PTPRC* or *NFKBIB* have been investigated in relation to anti-TNF response by more than 40 candidate-gene studies[322]. Although these studies are limited to the current biological knowledge, they have identified new loci associated to anti-TNF response.

One of these candidate genes is *FCGR2A*, which encodes for a Fc receptor that is mainly expressed in dendritic cells and macrophages[323]. Given that alterations in the Fc binding affinity to anti-TNF drugs could directly influence the response[323-325], the study of genetic variants in Fc receptors is of major clinical interest[323-325]. Accordingly, we have previously analyzed the association between variation at this gene and anti-TNF response *(Annex* section*)*[326].

In IBDs and PS, multiple studies have identified genes associated with anti-TNF response. However, most of these loci have shown modest associations with the clinical response and lack of reproducibility[327-331].

The genetic component underlying anti-TNF response in PsA has been explored by only a few studies focusing on variation in *TNF*, *TNFR1A*, *TRAIL-R1*, *FCGR2A* and *PDE3A-SLCO1C1* loci[332,333]. Altogether, these studies highlight the need for replicating the genetic findings in larger cohorts of patients.

In the last decade, genome-wide genotyping has allowed to investigate the contribution of whole genome variation on the heterogeneity of anti-TNF response in RA and PS. To date, a total of six GWAS on anti-TNF response have been performed in RA[334-339]. These have led to the identification of four susceptibility loci that are significant at the genome-wide scale (i.e. *MED15, GFRA1, PDE3A-SLCO1C1* and *CD84* genes)[340], and that have also been replicated in at least an independent cohort of patients. The unique GWAS performed in PS has identified genetic variants at *ADRA2A*, *CDH23*, *GUCY1B3*, *JAG2*, *KCNIP1*, *LOC728724*, *MACC1*, *PDE6A*, *SHOC2* and *SPEN* genes showing non genome-wide significant associations with anti-TNF response[341]. The association between genetic variation and the heterogeneity of anti-TNF response has been yet investigated in PsA and IBDs.

Taken together, the pharmacogenomic studies on anti-TNF response in IMIDs have led to two general findings. First, the genetic variation of anti-TNF response can be drug-specific. In RA, for example, the association of *MED15* and *CD84* genes was detected in etanercept-treated patients and was not observed in adalimumab- or infliximab-treated patients. Likewise, we found that variation at *FCGR2A* gene is associated with the clinical response to adalimumab and infliximab, but not to etanercept therapy[326]. Second, genetic variation associated to anti-TNF response might be shared across IMIDs[342]. The *PDE3A-SLCO1C1* locus was initially associated with the response to anti-TNF therapy in RA[340]. Soon after this discovery, the same genetic variant was found to be also associated with anti-TNF response in PS and PsA[333,343]. Therefore, the biological variability among patients and the molecular diversity of anti-TNF agents should be considered in the search for response biomarkers in IMIDs.

1.3.3 Genetic basis of immunogenicity to anti-TNF therapy

Clinical and experimental studies have shown that not only treatment-related factors (e.g. drug dosage and structure) promote ADA formation[344,345], but also patient-related factors like genetic variation[319]. In IMIDs, however, the contribution of genetic variation on ADA formation has been only analyzed by a few candidate studies.

In RA, the production of anti-adalimumab antibodies has been associated with genetic variation in *IL10*, a key participant in B cell differentiation[346]. More recently, five alleles in the *HLA* region (i.e. *HLA-DRB1*01*, *HLA-DRB1*03*, *HLA-DQB1*05*, *HLA-DRB1*07* and *HLA-DRB1*011*) have been associated to adalimumab immunogenicity[347].

In IBDs, carrying the *HLA-DRB1*03* or *HLA-DRB1*13* alleles has been found to predispose to anti-infliximab antibodies[348]. Also, variation in *FCGR3A* has been recently associated with the production of ADAs in adalimumab or infliximab treated patients[349]. The *IGHG1* gene has been tested for association with antibodies to infliximab, but no significant association has been detected[350]. In other autoimmune diseases, variation in the HLA class II region (*HLA-DRB1*0401* and *HLA-DRB1*0408* alleles) has been strongly associated with the development of ADAs[351].

The influence of genetic variation on anti-TNF immunogenicity has not been yet investigated in PS and PsA. In these and other IMIDs, analyses at a genome-wide scale could enable the identification of genetic markers that help physicians to detect those patients at high risk of anti-TNF intolerance.

1.3.4 Translating pharmacogenomics to therapeutics

The past few years have witnessed exciting pharmacogenomic discoveries. Some of the most promising genetic associations in terms of clinical applicability have been found in pharmacogenomic studies[137]. The FDA has labeled more than 150 drugs with pharmacogenomic information and pharmacogenomic testing is starting to be integrated in the clinical practice[352]. One of the most successful examples is the screening for the *HLA-B*5701* allele to identify Abacavir treated patients at high risk of developing hypersensitivity reactions[353]. Clinical trials have also investigated the power of variation in *ADRB2* to predict which patients will develop venodilatation or desensitization after isoproterenol administration as well as acute airway response to albuterol (Figure 26)[354,355]. Variation in these two genes exemplifies how pharmacogenomic information can guide drug administration.

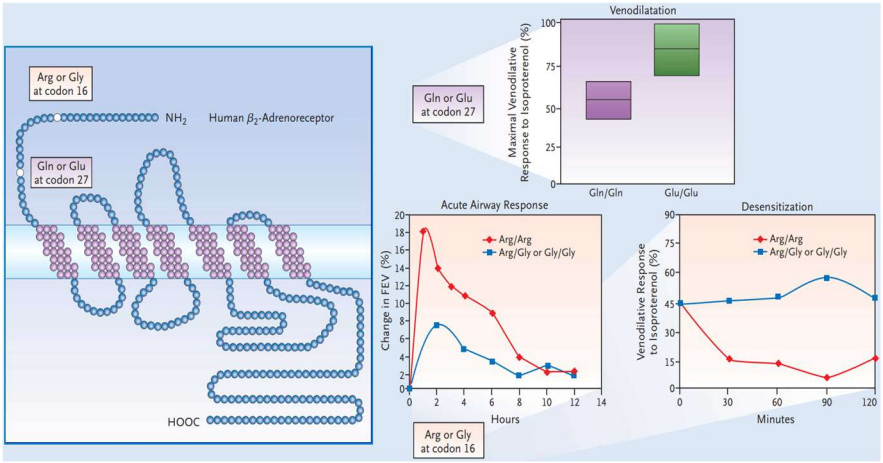

Figure 26. Impact of *ARDB2* variation on treatment outcomes. Homozygous Glu genotype at codon 27 is associated to greater venodilatation after isoproterenol administration. Homozygous Arg genotype codon 16 is associated to greater desensitization to isoproterenol and greater airway response to albuterol. Abbreviations: FEV, forced expiratory volume. Source: adapted from *Evans ME et al. N Engl J Med, 2003*[355].

In IMIDs, rigorous and systematic measurement of drug response can be as difficult as the genome-wide genotyping[356]. Therefore, the creation of biomedical consortia like the IMIDC will be essential to advance in pharmacogenomics research[342].

2 | OBJECTIVES

The three main objectives of the present thesis were:

1) **To identify new genetic variation associated with IMID susceptibility.** We sought to identify: (i) new genetic factors associated with the risk of developing PS, and (ii) new genetic factors that contribute to the risk of PsA but not purely cutaneous PS.

2) **To identify new genetic variation associated with clinical phenotypes in IMIDs.** We aimed to discover new genetic variation associated with: (i) the most clinically relevant phenotypes in SLE, and (ii) the risk of developing CVD in IMIDs.

3) **To identify new genetic markers for anti-TNF therapy in IMDs.** We sought to find new genetic markers associated with: (i) the clinical response to anti-TNF therapy in RA, and (ii) the production of anti-drug antibodies in CD.

3 | Identification of new genetic variation associated with IMID susceptibility

3.1 Genome-wide pathway analysis identifies genetic pathways associated with PS

Aterido A, Julià A, Ferrándiz C, Puig L, Fonseca E, Fernández-López E, Dauden E, Sánchez-Carazo JL, López-Estebaranz JL, Moreno-Ramírez D, Vanaclocha F, Herrera E, de la Cueva P, Dand N, Palau N, Alonso A, López-Lasanta M, Tortosa R, García-Montero A, Codó L, Gelpí JL, Bertranpetit J, Absher D, Capon F, Myers RM, Barker JN, Marsal S. *Genome-Wide Pathway Analysis Identifies Genetic Pathways Associated with Psoriasis.* Journal of Investigative Dermatology (2016)

http://www.jidonline.org/article/S0022-202X(15)00242-0/fulltext

Aterido A, Julià A, Ferrándiz C, Puig L, Fonseca E, Fernández-López E, et al. Genome-Wide Pathway Analysis Identifies Genetic Pathways Associated with Psoriasis. J Invest Dermatol. 2016;136(3):593–602. DOI: 10.1016/j.jid.2015.11.026

3.2 Genetic variation at the GAG metabolism pathway contributes to the risk of PsA but not psoriasis

Aterido A, Cañete JD, Tornero J, Ferrándiz C, Pinto JA, Gratacós J, Queiró R, Montilla C, Torre-Alonso JC, Pérez-Venegas JJ, Fernández Nebro A, Muñoz-Fernández S, González C, Roig D, Zarco P, Erra A, Rodríguez J, Castañeda S, Rubio E, Salvador G, Díaz-Torné C, Blanco R, Willisch Domínguez A, Mosquera JA, Vela P, Sánchez-Fernández S, Corominas H, Ramírez J, De la Cueva P, Fonseca E, Fernández E, Puig L, Dauden E, Sánchez-Carazo JL, López-Estebaranz JL, Moreno D, Vanaclocha F, Herrera E, Blanco F, Fernández-Gutiérrez B, González A, Pérez-García C, Alperi-López M, Olivé A, Martínez-Taboada V, González-Álvaro I, Sanmartí R, Tomás C, García-Montero AC, Bonàs-Guarch S, Mercader JM46, Torrents D46,47, Codó L48, Gelpí JL48, López-Corbeto M, Pluma A, López-Lasanta M, Tortosa R, Palau N, Absher D, Myers R, Marsal S, Julià A. *Genetic variation at the glycosaminoglycan metabolism pathway contributes to the risk of psoriatic arthritis but no psoriasis.* Annals of the Rheumatic Diseases (2018)

Under review

Aterido A, Cañete JD, Tornero J, Ferrándiz C, Pinto JA, Gratacós J, et al. Genetic variation at the glycosaminoglycan metabolism pathway contributes to the risk of psoriatic arthritis but not psoriasis. Ann Rheum Dis. 2019 Mar;78(3):355–64. DOI: 10.1136/annrheumdis-2018-214158

4 | Identification of new genetic variation associated with IMID clinical phenotypes

4.1 Genome-wide pathway analysis identifies VEGF pathway association with oral ulceration in SLE

Aterido A, Julià A, Carreira P, Blanco R, López-Longo JJ, Venegas JJP, Olivé À, Andreu JL, Aguirre-Zamorano MÁ, Vela P, Nolla JM, Marenco-de la Fuente JL, Zea A, Pego JM, Freire M, Díez E, López-Lasanta M, López-Corbeto M, Palau N, Tortosa R, Gelpí JL, Absher D, Myers RM, Fernández-Nebro A, Marsal S. *Genome-wide pathway analysis identifies VEGF pathway association with oral ulceration in systemic lupus erythematosus.* Arthritis Research & Therapy (2017)

https://arthritis-research.biomedcentral.com/articles/10.1186/s13075-017-1345-6

Aterido A, Julià A, Carreira P, Blanco R, López-Longo JJ, Venegas JJP, et al. Genome-wide pathway analysis identifies VEGF pathway association with oral ulceration in systemic lupus erythematosus. Arthritis Res Ther. 2017 Jun 15;19(1). DOI: 10.1186/s13075-017-1345-6

4.2 Genetic variation associated with cardiovascular risk in IMIDs

Perrotti PP, **Aterido A**, Fernández-Nebro A, Cañete JD, Ferrándiz C, Tornero J, Gisbert JP, Domènech E, Fernández-Gutiérrez B, Gomollón F, García-Planella E, Fernández E, Sanmartí R, Gratacós J, Martínez-Taboada VM, Rodríguez-Rodríguez L, Palau N, Tortosa R, Corbeto ML, Lasanta ML, Marsal S, Julià A; IMID Consortium. *Genetic variation associated with cardiovascular risk in autoimmune diseases.* PLoS One (2017)

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0185889

Perrotti PP, Aterido A, Fernández-Nebro A, Cañete JD, Ferrándiz C, Tornero J, et al. Genetic variation associated with cardiovascular risk in autoimmune diseases. PLoS One. 2017 Oct 1;12(10). DOI: 10.1371/journal.pone.0185889

5 | Identification of new genetic markers for anti-TNF therapy in IMIDs

5.1 A combined transcriptomic and genomic analysis identifies a gene signature associated with the response to anti-TNF therapy in RA

Aterido A, Cañete JD, Tornero J, Blanco F, Fernández-Gutiérrez B, Pérez C, Alperi-López M, Olivé A, Corominas H, Martínez-Taboada V, González I, Fernández-Nebro A, Erra A, López-Lasanta M, López-Corbeto M, Palau N, Marsal S, Julià A. *A combined transcriptomic and genomic analysis identifies a gene signature associated with the response to anti-TNF therapy in rheumatoid arthritis.* Scientific Reports (2018)

Under review

Aterido A, Cañete JD, Tornero J, Blanco F, Fernández-Gutierrez B, Pérez C, et al. A combined transcriptomic and genomic analysis identifies a gene signature associated with the response to anti-TNF therapy in rheumatoid arthritis. Front Immunol. 2019;10(JUL). DOI: 10.3389/fimmu.2019.01459

5.2 Genetic association between *CD96* locus and immunogenicity to anti-TNF therapy in CD

Aterido A[*], Palau N[*], Domènech E, Nos-Mateu P, Gutiérrez A, Gomollón F, Mendoza JL, Garcia-Planella E, Barreiro-de Acosta M, Muñoz F, Vera M, Saro C, Esteve M, Andreu M, Chaparro M, Panés J, García-Sánchez V, López-Lasanta M, Pluma A, Codó L, García-Montero A, Manyé J, Gisbert JP, Marsal S, Julià A. *Genetic association between CD96 locus and immunogenicity to anti-TNF therapy in CD.* The Pharmacogenomics Journal (2018)

[*] Equally contributed as first authors.

Under review

Aterido A, Palau N, Domènech E, Nos Mateu P, Gutiérrez A, Gomollón F, et al. Genetic association between CD96 locus and immunogenicity to anti-TNF therapy in Crohn's disease. Pharmacogenomics J. 2019 Dec 1;19(6):547–55. DOI: 10.1038/s41397-019-0090-4

6 | DISCUSSION

6.1 Identification of new genetic variation associated with IMID susceptibility

Identification of genetic variation associated with psoriasis risk

PS is a chronic inflammatory disease of the skin that has a complex genetic architecture. GWAS have only partially explained the disease heritability[357,358]. There is increasing evidence that part of the missing heritability could be explained by multiple small-effect genes from common pathways[167]. Most of the GWAS are performed at the single-marker level and, consequently, the statistical power to detect new risk genes soon becomes insufficient[359]. One of the most successful strategies to overcome this problem is the genome-wide pathway analysis (GWPA)[136]. Before this work, however, the GWPA had not been applied to study the genetics of PS.

In this work, we have performed a two-stage GWPA in order to identify new genetic pathways associated with PS risk. Using this approach and two large case-control cohorts, we have identified three new genetic pathways that had not been previously associated with PS risk. These new pathways include retinol metabolism, transport of inorganic ions and amino acids and post-translational modification. From these pathways, we have found that the latter includes the gene with the strongest network centrality properties. This gene that is key for the pathway functionality is the *MGAT5* gene. Additionally, we have further confirmed the association of *MGAT5* with PS both at the functional and genetic levels.

In dermatological diseases, retinol has been shown to inhibit inflammatory processes by regulating the NFKB activity[360,361]. As a transcriptional factor, NFKB regulates proinflammatory genes that are key for the pathogenesis of PS, such as the *TNF* and *IL-17* genes[362]. The NFKB

signaling pathway has been also associated with the regulation of keratinocyte proliferation[363]. These evidences are consistent with the elevated levels of NFKB that have been found in psoriatic skin compared to non-psoriatic skin[364]. From a pharmacological perspective, the use of retinoids (i.e. retinol and its metabolites) for the treatment of PS is also in line with our findings[365]. Therefore, genetic variation in the retinol metabolism pathway could reduce the retinol production, weaken the NFKB signaling and promote both inflammatory and proliferative hallmarks of PS.

The transport of inorganic ions in CD4+ T cells has been shown to be implicated in the development of IMIDs[366]. In particular, the intracellular transport of calcium is crucial to control the expression of proinflammatory genes in T cells[367,368]. Importantly, the expression of amino acid transporters has been found to be differentially regulated in PS[369]. These evidences suggest that genetic variation in the transport of amino acids and inorganic ions pathway could increase the PS risk by modulating the activity of T cells.

A relevant function of the post-translational modification (PTM) pathway is the regulation of the immune tolerance to self-antigens through the N-linked glycosylation of key HLA molecules for antigen recognition[370,371]. Furthermore, specific post-translationally modified autoantigens have been associated with the development of PS[372]. Also, the glycosylation activity has been found to be markedly increased in PS compared to controls[373]. PTMs on glycoproteins from the T cell surface have been shown to target these cells towards the inflamed skin in PS[374]. Therefore, genetic variation in the PTM pathway could perturb glycosylation processes that are essential to maintain the immune system tolerance.

Using network analyses, we have found that *MGAT5* gene plays a central role for the functionality of the PTM pathway. *MGAT5* is a glucosaminyltransferase whose activity has been associated with autoimmunity[375,376]. In our study, we have found that genetic variation at *MGAT5* is associated with the glycosylation levels of *in vitro* activated T cells. Consistent with previous findings[377], our results indicate that genetic

variation could play a functional role in the development of PS by modulating the glycosylation-dependent activation of T cells. Importantly, after the publication of our work, the glycosylation levels of T cells were found to be increased in the target tissues of other IMIDs like IBDs and, consequently, the blockade of this biological process has been proposed as new therapeutic strategy[378].

Identification of genetic variation associated with psoriatic arthritis risk

PsA is a chronic inflammatory arthritis affecting up to 30% of PS patients[23]. PsA can have dramatic functional implications for the patient and, generally, it requires an alternative therapeutic approach than purely cutaneous PS. Familiar aggregation studies have shown that PsA has a higher heritability than PS[246-249], indicating the existence of PsA-specific genetic risk. However, identifying disease-specific loci has proven very challenging and, to date, only *PTPN22*, *CSF2-P4HA2* and *ADAMTS9-MAGI1* have been shown specific for PsA risk[379,380]. The effect of these loci explains <50% of the PsA heritability and, therefore, new genetic factors remain to be identified[269-271].

In this study, we have performed a GWAS at the single-marker and pathway levels on two independent PsA case-control cohorts. In these analyses, we have identified a new PsA risk SNP at *B3GNT2* gene and 14 genetic pathways associated with PsA. From these, the glycosaminoglycan (GAG) metabolism pathway was confirmed to be disease-specific when comparing the PsA cohort of patients with a cohort of purely cutaneous psoriasis patients and a cohort of RA patients. These findings are of a major interest for both the rheumatology and dermatology communities since the identification genetic variation that differentiates PsA from PS can guide the development of disease specific drugs. Using network and drug repurposing analyses, we have further identified candidate drug targets in the GAG metabolism pathway as well as new PsA indications for approved drugs.

B3GNT2 encodes for an acetylglucosaminyltransferase that synthesizes the carbohydrate structure of polylactosamine onto glycoproteins[381]. The *B3GNT2* locus has been previously associated with other arthritic diseases[382,383]. *B3gnt2* knockout mice show hyperactivation of T and B lymphocytes as well as enhanced macrophage activation[381,384], thereby providing a functional link between the *B3GNT2* locus and autoimmunity. Our study shows, for the first time, that *B3GNT2* is associated with PsA at the genome-wide scale and that the frequency of the risk allele is significantly higher in PsA than in purely cutaneous psoriasis. With these findings, we demonstrate that while *B3GNT2* is a common risk locus for PsC and PsA, it has a bigger contribution to PsA etiology.

GAG metabolism has been shown to be altered in autoimmune diseases[385,386]. GAGs are crucial components of proteoglycans and the major component of cartilage, which the target tissue of PsA[387,388]. In *in vitro* models, the proteolysis of a cartilage-specific proteoglycan (i.e. aggrecan) in response to proinflammatory cytokines promotes articular damage[389]. After aggrecan destruction, GAGs are released from the ECM to the synovial fluid[390]. Accordingly, the synovial levels of GAGs have been found increased in PsA compared to controls[391]. Consistent with these evidences, genetic variation at GAG metabolism pathway could diminish the biosynthesis of GAGs and, consequently, reduce their availability for aggrecan and cartilage formation in patients with PsA.

Current drug discovery research is shifting from targeting single genes towards the modulation of specific biological pathways[392]. Here, we have investigated the potential of the PsA-specific associations for drug discovery using network and drug repurposing analyses. The results of these analyses show that the GAG metabolism could be a druggable pathway for PsA treatment. Our findings also suggest that the FDA-approved drugs hyaluronic acid and tromethamine are good candidates for repurposing in PsA, since they target key genes for the GAG metabolism. With this study, we have demonstrated the power of genetics to identify new drug targets and opportunities for drug repurposing in PsA[393].

6.2 Identification of new genetic variation associated with IMID clinical phenotypes

Identification of genetic variation associated with clinical phenotypes of systemic lupus erythematosus

SLE is a rheumatic disease characterized by heterogeneous clinical manifestations of unknown etiology. One of the major challenges in the pathogenesis of SLE is understanding the biological basis of the disease clinical heterogeneity. In the last years, GWAS have enabled the identification of >40 loci associated with SLE risk[394]. There is also evidence that the main SLE clinical phenotypes aggregate in families, which suggests the existence of a genetic basis for disease heterogeneity[395]. Before this thesis, however, only a few candidate analyses were performed to study the genetics of clinical heterogeneity in SLE[396].

In order to identify new genetic variants associated with SLE phenotypes, we have performed the first GWPA on SLE clinical phenotypes using two independent cohorts of patients. The main clinical phenotypes represented by the 11 American College of Rheumatology criteria for SLE diagnosis were tested for association with 798 biological pathways. In this study, we have identified and validated the association between VEGF genetic pathway and the presence of oral ulceration in SLE. Therapies commonly used to treat mucocutaneous SLE phenotypes were found to strongly influence the expression of the VEGF pathway in phenotype-relevant cell types.

The VEGF pathway is a network of genes that act coordinately to modulate inflammatory and angiogenic processes[397,398]. There is growing evidence that angiogenesis is involved in the development of skin manifestations in SLE[399,400]. Here, we have shown for the first time that genetic variation in the VEGF pathway is associated with oral ulceration in SLE. Oral ulcers are lesions characterized by a high angiogenic activity in the oral mucosa[401,402]. In SLE, these mucocutaneous lesions are highly prevalent and frequently chronic[403,404]. Oral ulcers have been associated

with an increased disease activity and worse prognosis in SLE[405,406]. Our findings are consistent with previous evidences from other ulcer-related diseases like Behçet's disease (BD) or gastroduodenal ulcers (GDU). In BD, clinical evidence supports that VEGF could be implicated in the formation of oral ulcers[407]. In GDU, genetic variation at *VEGF* has been associated with the disease risk[408]. Taken together, these evidences support the implication of VEGF pathway in oral ulceration both at the genetic and functional level.

Finally, we have evaluated the utility of targeting the VEGF to treat oral ulcers. We have demonstrated that topical immunotherapies perturb the expression of the VEGF pathway using an *in silico* analysis. Our results indicate that the VEGF pathway could mediate the benefits of topical immunotherapies to reduce oral ulceration. Importantly, a drug repurposing analysis for SLE treatment based on GWAS findings was published one year after the release of our study[409]. This work demonstrates the power of genetic information to guide drug repurposing in SLE. Accordingly, we suggest that FDA-approved drugs targeting the VEGF pathway could be repurposed for the treatment of oral ulceration in SLE.

Identification of genetic variation associated with cardiovascular disease risk in IMIDs

The development of CVD has life-threatening consequences and therefore it is one of the most important comorbidities in IMIDs[410]. CVD has been associated with the reduction of the life expectancy observed in IMID patients[410]. There is growing evidence that the development of a CVD in IMIDs is only partially explained by the classical CVD risk factors[91]. In the general population, the identification of >100 disease risk loci by GWAS has demonstrated the existence of a genetic risk background for CVD[297,411,412]. Before our study, however, little was known on the impact of genetics on CVD risk in IMIDs. The identification of these genetic

factors could be fundamental to characterize the specific biological mechanisms that most contribute to CVD development.

In this work, we have investigated the genetic basis of CVD in IMIDs using GWAS data from 6,485 patients with RA, PsA, PS, SLE, CD or UC. We have found that 17 established CVD risk loci are associated with CVD in IMIDs. From these, four loci showed different genetic effects across the six IMIDs. In addition, we have detected that six IMID risk loci are associated with CVD. With a GWAS cross-phenotype meta-analysis, we have identified 10 genetic clusters associated with CVD risk in IMIDs.

The established CVD risk loci replicated in IMIDs include genes with functional roles that could link both diseases. For example, *ADAMTS7* encodes for a metalloproteinase that is implicated on arthritis and also on the thickening of the neointima[413,414]. *SMARCA4* has been involved in CD4+ T cell differentiation and it has been associated to lipoprotein levels, which directly contribute to CVD[415,416]. Also, the differential effects detected in four CVD risk genes suggest that the disease-specific proinflammatory state influences the risk conferred by the known genetic factors. This is the first time that CVD risk loci have shown to be modulated by the presence of an IMID. Supporting this, after the publication of our study, a genome-wide study found that the *RARB* gene is associated with carotid intima-media thickness in RA patients[417]. Taken together, these findings indicate the existence of a genetic basis for autoimmunity that is also associated with CVD risk.

Using a genome-wide cross phenotype meta-analysis on GWAS data from IMID patients, we have identified 10 genetic clusters associated with CVD risk. Importantly, two of these genetic clusters showed a significant enrichment in genes from immune response pathways, including TNF signaling. TNF is a key proinflammatory cytokine for IMIDs[418], but also an important regulator of the cardiac function[419]. The systematic inhibition of this cytokine has proven effective to treat IMID patients as well as to reduce the prevalence of CVD[89]. The marked statistical significance of the genetic clusters identified in this study strongly supports that CVD is associated with different genes across IMIDs.

6.3 Identification of new genetic markers for anti-TNF therapy in IMIDs

Identification of genetic variation associated with the clinical response to anti-TNF therapy in rheumatoid arthritis

The introduction of anti-TNF agents has significantly improved the management of many RA patients. However, approximately 30% of the anti-TNF treated patients do not show a significant clinical improvement. There is growing evidence that clinical response to anti-TNF therapy in RA has a genetic component[334]. However, only variation at four genes has been previously replicated[340], including two drug-specific associations[337]. Consistent with clinical observations[420], these findings support the existence of drug specific variation underlying anti-TNF response. At the transcriptomic level, the few gene expression signatures identified so far have shown a modest association with anti-TNF response as well as a low overlap of genes between studies[421]. Therefore, it is noteworthy that before this thesis genetic and transcriptomic data had been only separately analyzed to characterize the biological causes of anti-TNF failure in RA.

In this work, we have conducted an integrative genomic analysis. Using synovial biopsies from RA patients, we have identified 149 gene coexpression modules (GCMs) that characterize the inflamed RA synovium. From these, 13 GCMs were found to be associated with anti-TNF response. At the genetic level, one of the 13 GCMs showed a significant association with adalimumab response in two independent cohorts. Using pathway and cell type epigenetic enrichment analysis, we have further detected that the adalimumab-associated GCM is enriched in genes that participate in the nucleotide metabolism as well as in epigenetic marks from critical immune cells like Tregs. Our analysis demonstrates that integrating different layers of molecular data is a powerful strategy to discover new genetic and biological mechanisms of anti-TNF response in RA.

Previous expression analyses in the RA synovium have shown that adalimumab reduces the expression of genes implicated on cell

proliferation[422]. In addition to the essential role that the nucleotide metabolism plays in DNA replication, this biological process is responsible for the synthesis of adenosine, a purine nucleoside that exhibits a potent anti-inflammatory activity when bound to its cognate receptors[423]. Adenosine receptors, however, display a weaker affinity for adenosine in RA compared to controls, thereby dampening their anti-inflammatory effect. Importantly, adalimumab has been found to normalize the binding affinity of adenosine receptors in RA patients [424,425]. Our results are in line with these evidences and provide a functional link between the effectivity of adalimumab and the local production of adenosine in the synovial joint.

Tregs produce high levels of anti-inflammatory cytokines to modulate the action of cytotoxic CD8+ T cells and, therefore, are essential for self-tolerance[52,426,427]. In RA, however, Tregs are functionally defective resulting in a sustained immune response to self-antigens[428]. Anti-TNF therapy has been shown to restore the suppressor function of Tregs in RA[429]. There is evidence that this modulation of Tregs could be drug-specific[430]. In particular, adalimumab has been shown to induce a Treg-specific phenotype that restrains the progression of IL-17-related inflammation by regulating the expression of IL-6 by monocytes[431].

Identification of genetic variation associated with anti-TNF immunogenicity in Crohn's disease

Understanding the biological mechanisms of anti-TNF response in CD is of major interest to prevent the treatment failure. The production of antidrug antibodies is one of the main causes of the treatment failure[432]. However, only a few candidate studies in RA have previously investigated the genetics basis of anti-TNF immunogenicity[346,350]. In CD, this analysis has proven challenging due to the lack of well-characterized cohorts of patients and, therefore, the contribution of genetics to anti-TNF immunogenicity remains unknown before this study.

As a part of the present thesis, the first GWAS for immunogenicity to anti-TNF therapy in CD has been performed. Following a two-stage design, we have identified and validated a significant association between the *CD96* locus and the production of antibodies to anti-TNF therapy in CD. As expected, we have found that the risk allele for anti-TNF immunogenicity is also associated with a lack of clinical response to this therapy.

CD96 is a member of the immunoglobulin superfamily that is mainly expressed in the cell membrane of NK, CD8+ T and CD4+ T cells and some subsets of B cells[433]. The expression of *CD96* gene has been also found to be higher in the terminal ileum than any other human tissue[434]. Consistent with our results, these evidences suggest a functional role for the *CD96* gene in the target tissue of CD. CD96 bind to the CD155 ligand, which also binds to the membrane receptors CD226 and TIGIT[435]. *Cd96* knockout mice have demonstrated that CD96 regulates the cytokine response of NK cells competing with CD226 to bind to the ligand CD155 expressed in APCs[436]. Importantly, variation at *CD155* has been also associated to immunogenicity, in this case, against vaccination[437,438].

In *Cd155* knockout mice, this CD96 ligand has been implicated in humoral response development. *Cd155-/-* mice show a less efficient response to orally administered antigens due to a decreased production of IgG and IgA compared to wild-type mice[14]. Also, significantly higher titers of Th1-associated IgG isotypes are detected after immunization in *Cd155-/-* mice compared to wild-type littermates[15]. These findings suggest that CD155 participates in the polarization of naïve CD4+ T cells to the Th2 phenotype. Together, these experimental studies are consistent with our results. Genetic variation at CD96-CD155 signaling pathway might predispose to produce antidrug antibodies by promoting CD155 upregulation and the subsequent polarization to the Th2 phenotype that leads to B cell activation.

7 | CONCLUSIONS

The main conclusions of the present thesis are:

1) We have identified new genetic variation associated with IMID susceptibility:

   Retinol metabolism, transport of inorganic ions and amino acids and post-translational protein modification are biological pathways associated with PS risk at the genetic level.

   MGAT5 gene is a key factor for the post-translation protein modification pathway and variation at this gene is not only associated with PS risk, but also to the levels of glycosylation in T cells from patients with PS.

   Genetic variation in the glycosaminoglycan metabolism pathway contributes to the risk of PsA but not purely cutaneous PS.

   The FDA-approved drugs hyaluronic acid and tromethamine that target key genes for the glycosaminoglycan metabolism pathway could be repurposed for the treatment of PsA.

2) We have identified new genetic variation associated with clinically relevant phenotypes in IMIDs:

   The genetic basis underlying SLE clinical heterogeneity can be independent from the genetic component associated to disease risk.

   Genetic variation at the VEGF pathway is associated with the risk of developing oral ulceration in SLE.

   The VEGF pathway represents a potentially new target to develop phenotype-specific drugs in SLE.

IMID susceptibility loci can also predispose to CVD risk and the presence of an IMID can change the penetrance of established CVD risk loci.

Immune-related pathways are associated with CVD risk across IMIDs.

3) We have identified new genetic markers for anti-TNF treatment in IMIDs:

A gene coexpression module is associated with the clinical response to adalimumab at the genetic level, supporting the existence of drug-specific genetic factors for anti-TNF response in RA.

The nucleotide metabolism and immune cells like Tregs could mediate the response to adalimumab in RA.

CD96 locus is associated with immunogenicity to anti-TNF therapy and also to anti-TNF efficacy in CD.

8 | PUBLICATIONS

8.1 Original Research Articles

- **Aterido A**, Julià A, Ferrándiz C *et al.* (2016) *Genome-wide pathway analysis identifies genetic pathways associated with psoriasis.* Journal of Investigative Dermatology.
- Perroti PP, **Aterido A**, Fernández-Nebro A *et al.* (2017) *Genetic variation associated with cardiovascular risk in autoimmune diseases*. Plos One.
- **Aterido A**, Julià A, Carreira P *et al.* (2017) *Genome-wide pathway analysis identifies an association of VEGF pathway with oral ulcers in systemic lupus erythematosus*. Arthritis Res and Therapy.
- **Aterido A**, Cañete JD, Tornero J *et al.* (2018) *Genetic variation at the glycosaminoglycan metabolism pathway contributes to the risk of psoriatic arthritis but not psoriasis.* Annals of the Rheumatic Diseases -under review-
- **Aterido A**, Cañete JD, Tornero J *et al.* (2018) *A combined transcriptomic and genetic association analysis identifies a gene signature associated with the clinical response to anti-TNF therapy in rheumatoid arthritis.* Scientific Reports -under review-
- **Aterido A**, Palau N, Domènech E *et al.* (2018) *Genetic association between CD96 locus and immunogenicity to anti-TNF therapy in Crohn's Disease.* The Pharmacogenomics Journal -under review-

8.2 Additional Publications Authored By Adrià Aterido

- **Aterido A**, Palacio C, Marsal S *et al.* (2014) *Novel insights into the regulatory architecture of CD4+ T cells in rheumatoid arthritis.* Plos One
- Avila-Pedretti G, Tornero J, Fernández-Nebro A, Blanco F, González-Álvaro I, Cañete JD, Maymó J, Alperiz M, Fernández-Gutiérrez B, Olivé A, Corominas H, Erra A, **Aterido A** *et al.* (2015) *Variation at FCGR2A and functionally related genes is associated with the response to anti-TNF therapy in rheumatoid arthritis.* Plos One
- Chaparro M*, **Aterido A***, Guerra I *et al.* (2018) *New damaging variants identified by whole genome sequencing involved in the clinical response to anti-TNF therapy in Crohn's disease.* Alimentary Pharmacology & Therapeutics -under review-

  * Equally contributed as first authors.

8.3 Oral Communications

- **Multi-omics analysis identifies a gene signature associated with the clinical response to anti-TNF therapy in rheumatoid arthritis.** Annual American Congress of Rheumatology (ACR), Chicago, United States of America, 2018.
- **Identificación de un módulo genético asociado a la respuesta al tratamiento anti-TNF en artritis reumatoide mediante el análisis multiómico.** Annual Congress of the Spanish Society of Rheumatology (SER), A Coruña, Spain, 2018.
- **Identificación de variación genética específica de la artritis psoriásica mediante análisis de asociación de genoma completo.** Annual Congress of the Spanish Society of Rheumatology (SER), Bilbao, Spain, 2017.
- **Genetic variation associated with cardiovascular risk in autoimmune diseases.** Annual European Congress of Rheumatology (EULAR), Madrid, Spain, 2017.
- **Genome-wide pathway analysis of clinical phenotypes in systemic lupus erythematosus.** Annual Human Genome Meeting (HGM), Barcelona, Spain, 2017.
- **Genome-wide pathway analysis reveals that VEGF genetic pathway is associated with oral ulcers in systemic lupus erythematosus.** Annual European Congress of Rheumatology (EULAR), London, United Kingdom, 2016.
- **Genome-wide pathway analysis identifies new genetic pathways associated with psoriasis.** III Bioinformatics & Computational Biology Symposium, Bioinformatics Barcelona, Barcelona, Spain, 2015.
- **Novel insights into the regulatory architecture of CD4+ T cells in rheumatoid arthritis.** 8th Scientific Conference Vall d'Hebron Research Institute, Barcelona, Spain, 2014.

8.4 Poster Presentations

- **Multi-omics analysis identifies a gene signature associated with the clinical response to anti-TNF therapy in rheumatoid arthritis.** Annual European Congress of Rheumatology (EULAR), Amsterdam, Netherlands, 2018.
- **Identification of genetic variation specifically associated with psoriatic arthritis using genome-wide association studies.** Annual European Congress of Rheumatology (EULAR), Madrid, Spain, 2017.
- **Variación genética asociada al riesgo cardiovascular en las enfermedades autoinmunes.** Annual Congress of the Spanish Society of Rheumatology (SER), Bilbao, Spain, 2017.
- **Genome-wide pathway analysis reveals that VEGF genetic pathway is associated with oral ulcers in systemic lupus erythematosus.** Annual American Congress of Rheumatology (ACR), Washington DC, United States of America, 2016.

- **Asociación de la vía génica VEGF con la presencia de úlceras orales en lupus sistémico eritematoso mediante un análisis de asociación de genoma completo.** Annual Congress of the Spanish Society of Rheumatology (SER), Barcelona, Spain, 2016.
- **Genome-wide pathway analysis of clinical phenotypes in systemic lupus erythematosus.** IV Bioinformatics & Genomics Symposium, Bioinformatics Barcelona, Barcelona, Spain, 2016.
- **Genome-wide pathway analysis identifies new genetic pathways associated with psoriasis.** 9th Scientific Conference Vall d'Hebron Research Institute, Barcelona, Spain, 2015.
- **Genome-wide pathway analysis identifies new genetic pathways associated with psoriasis.** 9th Scientific Conference Vall d'Hebron Research Institute, Barcelona, Spain, 2015.
- **Genome-wide pathway analysis identifies new genetic pathways associated with psoriasis.** 4th DCEXS Symposium: Innovative *in silico* strategies in biomedical research, Barcelona, Spain, 2015.

9 | REFERENCES

1. Novembre, J. et al. Genes mirror geography within Europe. Nature 456, 98-101 (2008).
2. Claussnitzer, M. et al. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. N Engl J Med 373, 895-907 (2015).
3. Jaeger, S. et al. Quantification of Pathway Cross-talk Reveals Novel Synergistic Drug Combinations for Breast Cancer. Cancer Res 77, 459-469 (2017).
4. Hasin, Y., Seldin, M. & Lusis, A. Multi-omics approaches to disease. Genome Biol 18, 83 (2017).
5. Pray, L. Discovery of DNA structure and function: Watson and Crick. Nature Education 1(2008).
6. Clancy, S. & Brown, W. Translation: DNA to mRNA to Protein. Nature Education 1(2008).
7. Walton, R. et al. Haplotype block structure of the cytochrome P450 CYP2C gene cluster on chromosome 10. Nat Genet 37, 915-6; author reply 916 (2005).
8. Yu, D., Kim, M., Xiao, G. & Hwang, T.H. Review of biological network data and its applications. Genomics Inform 11, 200-10 (2013).
9. Marchini, J. & Howie, B. Genotype imputation for genome-wide association studies. Nat Rev Genet 11, 499-511 (2010).
10. Vidal, M., Cusick, M.E. & Barabasi, A.L. Interactome networks and human disease. Cell 144, 986-98 (2011).
11. Bush, W.S. & Moore, J.H. Chapter 11: Genome-wide association studies. PLoS Comput Biol 8, e1002822 (2012).
12. Strand, V., Kimberly, R. & Isaacs, J.D. Biologic therapies in rheumatology: lessons learned, future directions. Nat Rev Drug Discov 6, 75-92 (2007).
13. Knight, J.C. Genomic modulators of the immune response. Trends Genet 29, 74-83 (2013).
14. Weischenfeldt, J., Symmons, O., Spitz, F. & Korbel, J.O. Phenotypic impact of genomic structural variation: insights from and for human disease. Nat Rev Genet 14, 125-38 (2013).
15. Kingsmore, S.F., Lindquist, I.E., Mudge, J., Gessler, D.D. & Beavis, W.D. Genome-wide association studies: progress and potential for drug discovery and development. Nat Rev Drug Discov 7, 221-30 (2008).
16. Marston, A.L. & Amon, A. Meiosis: cell-cycle controls shuffle and deal. Nat Rev Mol Cell Biol 5, 983-97 (2004).

17. El-Gabalawy, H., Guenther, L.C. & Bernstein, C.N. Epidemiology of immune-mediated inflammatory diseases: incidence, prevalence, natural history, and comorbidities. J Rheumatol Suppl 85, 2-10 (2010).

18. Chandran, V. & Raychaudhuri, S.P. Geoepidemiology and environmental factors of psoriasis and psoriatic arthritis. J Autoimmun 34, J314-21 (2010).

19. Christophers, E. Psoriasis--epidemiology and clinical spectrum. Clin Exp Dermatol 26, 314-20 (2001).

20. Gladman, D.D., Antoni, C., Mease, P., Clegg, D.O. & Nash, P. Psoriatic arthritis: epidemiology, clinical features, course, and outcome. Ann Rheum Dis 64 Suppl 2, ii14-7 (2005).

21. Henseler, T. & Christophers, E. Psoriasis of early and late onset: characterization of two types of psoriasis vulgaris. J Am Acad Dermatol 13, 450-6 (1985).

22. Hou, J.K., Kramer, J.R., Richardson, P., Mei, M. & El-Serag, H.B. The incidence and prevalence of inflammatory bowel disease among U.S. veterans: a national cohort study. Inflamm Bowel Dis 19, 1059-64 (2013).

23. Ibrahim, G., Waxman, R. & Helliwell, P.S. The prevalence of psoriatic arthritis in people with psoriasis. Arthritis Rheum 61, 1373-8 (2009).

24. Lisnevskaia, L., Murphy, G. & Isenberg, D. Systemic lupus erythematosus. Lancet 384, 1878-1888 (2014).

25. Molodecky, N.A. et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. Gastroenterology 142, 46-54 e42; quiz e30 (2012).

26. Nestle, F.O., Kaplan, D.H. & Barker, J. Psoriasis. N Engl J Med 361, 496-509 (2009).

27. Norgard, B.M. et al. The incidence of ulcerative colitis (1995-2011) and Crohn's disease (1995-2012) - based on nationwide Danish registry data. J Crohns Colitis 8, 1274-80 (2014).

28. Parisi, R., Symmons, D.P., Griffiths, C.E. & Ashcroft, D.M. Global epidemiology of psoriasis: a systematic review of incidence and prevalence. J Invest Dermatol 133, 377-85 (2013).

29. Rees, F., Doherty, M., Grainge, M.J., Lanyon, P. & Zhang, W. The worldwide incidence and prevalence of systemic lupus erythematosus: a systematic review of epidemiological studies. Rheumatology (Oxford) 56, 1945-1961 (2017).

30. Shapira, Y., Agmon-Levin, N. & Shoenfeld, Y. Geoepidemiology of autoimmune rheumatic diseases. Nat Rev Rheumatol 6, 468-76 (2010).

31. Symmons, D. et al. The prevalence of rheumatoid arthritis in the United Kingdom: new estimates for a new century. Rheumatology (Oxford) 41, 793-800 (2002).

32. Wagtmans, M.J., Verspaget, H.W., Lamers, C.B. & van Hogezand, R.A. Gender-related differences in the clinical course of Crohn's disease. Am J Gastroenterol 96, 1541-6 (2001).

33. Manolio, T.A., Brooks, L.D. & Collins, F.S. A HapMap harvest of insights into the genetics of common disease. J Clin Invest 118, 1590-605 (2008).

34. Gutierrez-Arcelus, M., Rich, S.S. & Raychaudhuri, S. Autoimmune diseases - connecting risk alleles with molecular traits of the immune system. Nat Rev Genet 17, 160-74 (2016).

35. Zuniga-Pflucker, J.C. T-cell development made simple. Nat Rev Immunol 4, 67-72 (2004).

36. Germain, R.N. T-cell development and the CD4-CD8 lineage decision. Nat Rev Immunol 2, 309-22 (2002).

37. Masopust, D. & Schenkel, J.M. The integration of T cell migration, differentiation and function. Nat Rev Immunol 13, 309-20 (2013).

38. Lever, M., Maini, P.K., van der Merwe, P.A. & Dushek, O. Phenotypic models of T cell activation. Nat Rev Immunol 14, 619-29 (2014).

39. Davis, S.J. et al. The nature of molecular recognition by T cells. Nat Immunol 4, 217-24 (2003).

40. Russ, B.E., Prier, J.E., Rao, S. & Turner, S.J. T cell immunity as a tool for studying epigenetic regulation of cellular differentiation. Front Genet 4, 218 (2013).

41. Majchrzak, K. et al. Exploiting IL-17-producing CD4+ and CD8+ T cells to improve cancer immunotherapy in the clinic. Cancer Immunol Immunother 65, 247-59 (2016).

42. Arango Duque, G. & Descoteaux, A. Macrophage cytokines: involvement in immunity and infectious diseases. Front Immunol 5, 491 (2014).

43. Croft, M. The role of TNF superfamily members in T-cell function and diseases. Nat Rev Immunol 9, 271-85 (2009).

44. Tracey, D., Klareskog, L., Sasso, E.H., Salfeld, J.G. & Tak, P.P. Tumor necrosis factor antagonist mechanisms of action: a comprehensive review. Pharmacol Ther 117, 244-79 (2008).

45. Spencer, L.A. & Weller, P.F. Eosinophils and Th2 immunity: contemporary insights. Immunol Cell Biol 88, 250-6 (2010).

46. Smith, K.M. et al. Th1 and Th2 CD4+ T cells provide help for B cell clonal expansion and antibody synthesis in a similar manner in vivo. J Immunol 165, 3136-44 (2000).

47. Borghesi, L. & Milcarek, C. From B cell to plasma cell: regulation of V(D)J recombination and antibody secretion. Immunol Res 36, 27-32 (2006).

48. Kaplan, M.H. Th9 cells: differentiation and disease. Immunol Rev 252, 104-15 (2013).

49. Lee, Y. et al. Induction and molecular signature of pathogenic TH17 cells. Nat Immunol 13, 991-9 (2012).

50. Sonnenberg, G.F., Fouser, L.A. & Artis, D. Border patrol: regulation of immunity, inflammation and tissue homeostasis at barrier surfaces by IL-22. Nat Immunol 12, 383-90 (2011).

51. Ueno, H., Banchereau, J. & Vinuesa, C.G. Pathophysiology of T follicular helper cells in humans and mice. Nat Immunol 16, 142-52 (2015).

52. Wing, K. & Sakaguchi, S. Regulatory T cells exert checks and balances on self tolerance and autoimmunity. Nat Immunol 11, 7-13 (2010).

53. Bettini, M. & Vignali, D.A. Regulatory T cells and inhibitory cytokines in autoimmunity. Curr Opin Immunol 21, 612-8 (2009).

54. Milstein, O. et al. CTLs respond with activation and granule secretion when serving as targets for T-cell recognition. Blood 117, 1042-52 (2011).

55. Gray, S.M., Kaech, S.M. & Staron, M.M. The interface between transcriptional and epigenetic control of effector and memory CD8(+) T-cell differentiation. Immunol Rev 261, 157-68 (2014).

56. Dotiwala, F. et al. Killer lymphocytes use granulysin, perforin and granzymes to kill intracellular parasites. Nat Med 22, 210-6 (2016).

57. Perl, A. Pathogenesis and spectrum of autoimmunity. Methods Mol Biol 900, 1-9 (2012).

58. Kuek, A., Hazleman, B.L. & Ostor, A.J. Immune-mediated inflammatory diseases (IMIDs) and biologic therapy: a medical revolution. Postgrad Med J 83, 251-60 (2007).

59. Choy, E.H. & Panayi, G.S. Cytokine pathways and joint inflammation in rheumatoid arthritis. N Engl J Med 344, 907-16 (2001).

60. McInnes, I.B. & Schett, G. The pathogenesis of rheumatoid arthritis. N Engl J Med 365, 2205-19 (2011).

61. Lowes, M.A., Bowcock, A.M. & Krueger, J.G. Pathogenesis and therapy of psoriasis. Nature 445, 866-73 (2007).

62. Ferenczi, K., Burack, L., Pope, M., Krueger, J.G. & Austin, L.M. CD69, HLA-DR and the IL-2R identify persistently activated T cells in psoriasis vulgaris lesional skin: blood and skin comparisons by flow cytometry. J Autoimmun 14, 63-78 (2000).

63. Lebre, M.C. et al. Human keratinocytes express functional Toll-like receptor 3, 4, 5, and 9. J Invest Dermatol 127, 331-41 (2007).

64. Baliwag, J., Barnes, D.H. & Johnston, A. Cytokines in psoriasis. Cytokine 73, 342-50 (2015).

65. Veale, D.J. & Fearon, U. What makes psoriatic and rheumatoid arthritis so different? RMD Open 1, e000025 (2015).

66. Veale, D.J., Ritchlin, C. & FitzGerald, O. Immunopathology of psoriasis and psoriatic arthritis. Ann Rheum Dis 64 Suppl 2, ii26-9 (2005).

67. Tsokos, G.C., Lo, M.S., Costa Reis, P. & Sullivan, K.E. New insights into the immunopathogenesis of systemic lupus erythematosus. Nat Rev Rheumatol 12, 716-730 (2016).

68. Theofilopoulos, A.N., Kono, D.H., Beutler, B. & Baccala, R. Intracellular nucleic acid sensors and autoimmunity. J Interferon Cytokine Res 31, 867-86 (2011).

69. Crow, M.K. Type I interferon in the pathogenesis of lupus. J Immunol 192, 5459-68 (2014).

70. Crispin, J.C. & Tsokos, G.C. IL-17 in systemic lupus erythematosus. J Biomed Biotechnol 2010, 943254 (2010).

71. Peyrin-Biroulet, L., Loftus, E.V., Jr., Colombel, J.F. & Sandborn, W.J. The natural history of adult Crohn's disease in population-based cohorts. Am J Gastroenterol 105, 289-97 (2010).

72. Ungaro, R., Mehandru, S., Allen, P.B., Peyrin-Biroulet, L. & Colombel, J.F. Ulcerative colitis. Lancet 389, 1756-1770 (2017).

73. Khor, B., Gardet, A. & Xavier, R.J. Genetics and pathogenesis of inflammatory bowel disease. Nature 474, 307-17 (2011).

74. Abraham, C. & Cho, J.H. Inflammatory bowel disease. N Engl J Med 361, 2066-78 (2009).

75. Cheon, J.H. Genetics of inflammatory bowel diseases: a comparison between Western and Eastern perspectives. J Gastroenterol Hepatol 28, 220-6 (2013).

76. Valderas, J.M., Starfield, B., Sibbald, B., Salisbury, C. & Roland, M. Defining comorbidity: implications for understanding health and health services. Ann Fam Med 7, 357-63 (2009).

77. Veale, D.J. & Fearon, U. The pathogenesis of psoriatic arthritis. Lancet 391, 2273-2284 (2018).

78. Tan, E.M. et al. The 1982 revised criteria for the classification of systemic lupus erythematosus. Arthritis Rheum 25, 1271-7 (1982).

79. Qin, B. et al. Body mass index and the risk of rheumatoid arthritis: a systematic review and dose-response meta-analysis. Arthritis Res Ther 17, 86 (2015).

80. Bremmer, S. et al. Obesity and psoriasis: from the Medical Board of the National Psoriasis Foundation. J Am Acad Dermatol 63, 1058-69 (2010).

81. Singh, S., Dulai, P.S., Zarrinpar, A., Ramamoorthy, S. & Sandborn, W.J. Obesity in IBD: epidemiology, pathogenesis, disease course and treatment outcomes. Nat Rev Gastroenterol Hepatol 14, 110-121 (2017).

82. Khalili, H. et al. Measures of obesity and risk of Crohn's disease and ulcerative colitis. Inflamm Bowel Dis 21, 361-8 (2015).

83. Albrecht, K. et al. Body mass index distribution in rheumatoid arthritis: a collaborative analysis from three large German rheumatoid arthritis databases. Arthritis Res Ther 18, 149 (2016).

84. Singh, S. et al. Obesity and response to anti-tumor necrosis factor-alpha agents in patients with select immune-mediated inflammatory diseases: A systematic review and meta-analysis. PLoS One 13, e0195123 (2018).

85. Hansson, G.K. & Hermansson, A. The immune system in atherosclerosis. Nat Immunol 12, 204-12 (2011).

86. Berg, A.H. & Scherer, P.E. Adipose tissue, inflammation, and cardiovascular disease. Circ Res 96, 939-49 (2005).

87. Wang, B., Jenkins, J.R. & Trayhurn, P. Expression and secretion of inflammation-related adipokines by human adipocytes differentiated in culture: integrated response to TNF-alpha. Am J Physiol Endocrinol Metab 288, E731-40 (2005).

88. Fantuzzi, G. Adipose tissue, adipokines, and inflammation. J Allergy Clin Immunol 115, 911-9; quiz 920 (2005).

89. Jacobsson, L.T. et al. Treatment with tumor necrosis factor blockers is associated with a lower incidence of first cardiovascular events in patients with rheumatoid arthritis. J Rheumatol 32, 1213-8 (2005).

90. Kaul, A. et al. Systemic lupus erythematosus. Nat Rev Dis Primers 2, 16039 (2016).

91. Mankad, R. Atherosclerotic vascular disease in the autoimmune rheumatologic patient. Curr Atheroscler Rep 17, 497 (2015).

92. Deloukas, P. et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 45, 25-33 (2013).

93. Moran, A.E. et al. The global burden of ischemic heart disease in 1990 and 2010: the Global Burden of Disease 2010 study. Circulation 129, 1493-501 (2014).

94. Jacobs, P., Bissonnette, R. & Guenther, L.C. Socioeconomic burden of immune-mediated inflammatory diseases--focusing on work productivity and disability. J Rheumatol Suppl 88, 55-61 (2011).

95. Lundkvist, J., Kastang, F. & Kobelt, G. The burden of rheumatoid arthritis and access to treatment: health burden and costs. Eur J Health Econ 8 Suppl 2, S49-60 (2008).

96. Burgos-Pol, R. et al. The Cost of Psoriasis and Psoriatic Arthritis in 5 European Countries: A Systematic Review. Actas Dermosifiliogr 107, 577-90 (2016).

97. Rocchi, A. et al. Inflammatory bowel disease: a Canadian burden of illness review. Can J Gastroenterol 26, 811-7 (2012).

98. McCain, J. The Disease Burden of the Most Common Autoimmune Diseases. Manag Care 25, 28-32 (2016).

99. Taylor, P.C., Moore, A., Vasilescu, R., Alvir, J. & Tarallo, M. A structured literature review of the burden of illness and unmet needs in patients with rheumatoid arthritis: a current perspective. Rheumatol Int 36, 685-95 (2016).

100. Carter, E.E., Barr, S.G. & Clarke, A.E. The global burden of SLE: prevalence, health disparities and socioeconomic impact. Nat Rev Rheumatol 12, 605-20 (2016).

101. Vanderpuye-Orgle, J. et al. Evaluating the economic burden of psoriasis in the United States. J Am Acad Dermatol 72, 961-7 e5 (2015).

102. Kvien, T.K. Epidemiology and burden of illness of rheumatoid arthritis. Pharmacoeconomics 22, 1-12 (2004).

103. Kvamme, M.K. et al. Cost-effectiveness of TNF inhibitors vs synthetic disease-modifying antirheumatic drugs in patients with rheumatoid arthritis: a Markov model study based on two longitudinal observational studies. Rheumatology (Oxford) 54, 1226-35 (2015).

104. Bonafede, M.M., Gandra, S.R., Watson, C., Princic, N. & Fox, K.M. Cost per treated patient for etanercept, adalimumab, and infliximab across adult indications: a claims analysis. Adv Ther 29, 234-48 (2012).

105. Bickers, D.R. et al. The burden of skin diseases: 2004 a joint project of the American Academy of Dermatology Association and the Society for Investigative Dermatology. J Am Acad Dermatol 55, 490-500 (2006).

106. Lander, E.S. et al. Initial sequencing and analysis of the human genome. Nature 409, 860-921 (2001).

107. Venter, J.C. et al. The sequence of the human genome. Science 291, 1304-51 (2001).

108. Watson, J.D. & Crick, F.H.C. A structure for deoxyribose nucleic acid. Nature 171(1953).

109. Crick, F. Central dogma of molecular biology. Nature 227, 561-3 (1970).

110. Birney, E. et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature 447, 799-816 (2007).

111. Ezkurdia, I. et al. Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum Mol Genet 23, 5866-78 (2014).

112. Alexander, R.P., Fang, G., Rozowsky, J., Snyder, M. & Gerstein, M.B. Annotating non-coding regions of the genome. Nat Rev Genet 11, 559-71 (2010).

113. Palazzo, A.F. & Gregory, T.R. The case for junk DNA. PLoS Genet 10, e1004351 (2014).

114. Quinn, J.J. & Chang, H.Y. Unique features of long non-coding RNA biogenesis and function. Nat Rev Genet 17, 47-62 (2016).

115. ENCODE. An integrated encyclopedia of DNA elements in the human genome. Nature 489, 57-74 (2012).

116. Sachidanandam, R. et al. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409, 928-33 (2001).

117. Meyerson, M. Human genetic variation and disease. Lancet 362, 259-60 (2003).

118. Sherry, S.T. et al. dbSNP: the NCBI database of genetic variation. Nucleic Acids Res 29, 308-11 (2001).

119. Brandler, W.M. et al. Paternally inherited cis-regulatory structural variants are associated with autism. Science 360, 327-331 (2018).

120. Reich, D.E. et al. Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet 32, 135-42 (2002).

121. Konnick, E.Q. & Pritchard, C.C. Germline, hematopoietic, mosaic, and somatic variation: interplay between inherited and acquired genetic alterations in disease assessment. Genome Med 8, 100 (2016).

122. Eyre-Walker, A. & Keightley, P.D. The distribution of fitness effects of new mutations. Nat Rev Genet 8, 610-8 (2007).

123. Martincorena, I. & Campbell, P.J. Somatic mutation in cancer and normal cells. Science 349, 1483-9 (2015).

124. Matos, J., Blanco, M.G., Maslen, S., Skehel, J.M. & West, S.C. Regulatory control of the resolution of DNA recombination intermediates during meiosis and mitosis. Cell 147, 158-72 (2011).

125. Myers, S., Bottolo, L., Freeman, C., McVean, G. & Donnelly, P. A fine-scale map of recombination rates and hotspots across the human genome. Science 310, 321-4 (2005).

126. Wall, J.D. & Pritchard, J.K. Haplotype blocks and linkage disequilibrium in the human genome. Nat Rev Genet 4, 587-97 (2003).

127. LaFramboise, T. Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. Nucleic Acids Res 37, 4181-93 (2009).

128. The International HapMap Project. Nature 426, 789-96 (2003).

129. Auton, A. et al. A global reference for human genetic variation. Nature 526, 68-74 (2015).

130. Wang, D.G. et al. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280, 1077-82 (1998).

131. Lamy, P., Grove, J. & Wiuf, C. A review of software for microarray genotyping. Hum Genomics 5, 304-9 (2011).

132. Benyamin, B., Visscher, P.M. & McRae, A.F. Family-based genome-wide association studies. Pharmacogenomics 10, 181-90 (2009).

133. Ott, J., Wang, J. & Leal, S.M. Genetic linkage analysis in the age of whole-genome sequencing. Nat Rev Genet 16, 275-84 (2015).

134. Hirschhorn, J.N. & Daly, M.J. Genome-wide association studies for common diseases and complex traits. Nat Rev Genet 6, 95-108 (2005).

135. Manolio, T.A. Genomewide association studies and assessment of the risk of disease. N Engl J Med 363, 166-76 (2010).

136. Wang, K., Li, M. & Hakonarson, H. Analysing biological pathways in genome-wide association studies. Nat Rev Genet 11, 843-54 (2010).

137. MacArthur, J. et al. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). Nucleic Acids Res 45, D896-D901 (2017).

138. Teng, J. & Risch, N. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases. II. Individual genotyping. Genome Res 9, 234-41 (1999).

139. Risch, N. & Teng, J. The relative power of family-based and case-control designs for linkage disequilibrium studies of complex human diseases I. DNA pooling. Genome Res 8, 1273-88 (1998).

140. Chanock, S.J. et al. Replicating genotype-phenotype associations. Nature 447, 655-60 (2007).

141. MacRae, C.A. & Vasan, R.S. Next-generation genome-wide association studies: time to focus on phenotype? Circ Cardiovasc Genet 4, 334-6 (2011).

142. Aschard, H., Vilhjalmsson, B.J., Joshi, A.D., Price, A.L. & Kraft, P. Adjusting for heritable covariates can bias effect estimates in genome-wide association studies. Am J Hum Genet 96, 329-39 (2015).

143. Tian, C., Gregersen, P.K. & Seldin, M.F. Accounting for ancestry: population substructure and genome-wide association studies. Hum Mol Genet 17, R143-50 (2008).

144. Price, A.L., Zaitlen, N.A., Reich, D. & Patterson, N. New approaches to population stratification in genome-wide association studies. Nat Rev Genet 11, 459-63 (2010).

145. Price, A.L. et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 38, 904-9 (2006).

146. Johnson, R.C. et al. Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics 11, 724 (2010).

147. Spencer, C.C., Su, Z., Donnelly, P. & Marchini, J. Designing genome-wide association studies: sample size, power, imputation, and the choice of genotyping chip. PLoS Genet 5, e1000477 (2009).

148. Stranger, B.E., Stahl, E.A. & Raj, T. Progress and promise of genome-wide association studies for human complex trait genetics. Genetics 187, 367-83 (2011).

149. Begum, F., Ghosh, D., Tseng, G.C. & Feingold, E. Comprehensive literature review and statistical considerations for GWAS meta-analysis. Nucleic Acids Res 40, 3777-84 (2012).

150. Evangelou, E. & Ioannidis, J.P. Meta-analysis methods for genome-wide association studies and beyond. Nat Rev Genet 14, 379-89 (2013).

151. Li, Y., Willer, C., Sanna, S. & Abecasis, G. Genotype imputation. Annu Rev Genomics Hum Genet 10, 387-406 (2009).

152. Howie, B.N., Donnelly, P. & Marchini, J. A flexible and accurate genotype imputation method for the next generation of genome-wide association studies. PLoS Genet 5, e1000529 (2009).

153. Visscher, P.M. et al. 10 Years of GWAS Discovery: Biology, Function, and Translation. Am J Hum Genet 101, 5-22 (2017).

154. Hindorff, L.A. et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A 106, 9362-7 (2009).

155. Pickrell, J.K. et al. Detection and interpretation of shared genetic influences on 42 human traits. Nat Genet 48, 709-17 (2016).

156. Speed, D., Cai, N., Johnson, M.R., Nejentsev, S. & Balding, D.J. Reevaluation of SNP heritability in complex human traits. Nat Genet 49, 986-992 (2017).

157. Boyle, E.A., Li, Y.I. & Pritchard, J.K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. Cell 169, 1177-1186 (2017).

158. Welter, D. et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res 42, D1001-6 (2014).

159. Nelson, M.R. et al. The support of human genetic evidence for approved drug indications. Nat Genet 47, 856-60 (2015).

160. Mailman, M.D. et al. The NCBI dbGaP database of genotypes and phenotypes. Nat Genet 39, 1181-6 (2007).

161. Manolio, T.A. et al. Finding the missing heritability of complex diseases. Nature 461, 747-53 (2009).

162. Bomba, L., Walter, K. & Soranzo, N. The impact of rare and low-frequency genetic variants in common disease. Genome Biol 18, 77 (2017).

163. Cortijo, S. et al. Mapping the epigenetic basis of complex traits. Science 343, 1145-8 (2014).

164. Zuk, O., Hechter, E., Sunyaev, S.R. & Lander, E.S. The mystery of missing heritability: Genetic interactions create phantom heritability. Proc Natl Acad Sci U S A 109, 1193-8 (2012).

165. Robinson, M.R. et al. Genotype-covariate interaction effects and the heritability of adult body mass index. Nat Genet 49, 1174-1181 (2017).

166. Schadt, E.E. Molecular networks as sensors and drivers of common human diseases. Nature 461, 218-23 (2009).

167. Zhong, H., Yang, X., Kaplan, L.M., Molony, C. & Schadt, E.E. Integrating pathway analysis and genetics of gene expression for genome-wide association studies. Am J Hum Genet. 86, 581-91. doi: 10.1016/j.ajhg.2010.02.020. Epub 2010 Mar 25. (2010).

168. Gui, H., Li, M., Sham, P.C. & Cherny, S.S. Comparisons of seven algorithms for pathway analysis using the WTCCC Crohn's Disease dataset. BMC Res Notes. 4:386., 10.1186/1756-0500-4-386. (2011).

169. Ramanan, V.K., Shen, L., Moore, J.H. & Saykin, A.J. Pathway analysis of genomic data: concepts, methods, and prospects for future development. Trends Genet. 28, 323-32. doi: 10.1016/j.tig.2012.03.004. Epub 2012 Apr 3. (2012).

170. Liberzon, A. et al. Molecular signatures database (MSigDB) 3.0. Bioinformatics 27, 1739-40 (2011).

171. Sawcer, S. et al. Genetic risk and a primary role for cell-mediated immune mechanisms in multiple sclerosis. Nature 476, 214-9 (2011).

172. Wang, L., Jia, P., Wolfinger, R.D., Chen, X. & Zhao, Z. Gene set analysis of genome-wide association studies: methodological issues and perspectives. Genomics 98, 1-8 (2011).

173. Ramanan, V.K., Shen, L., Moore, J.H. & Saykin, A.J. Pathway analysis of genomic data: concepts, methods, and prospects for future development. Trends Genet 28, 323-32 (2012).

174. Elbers, C.C. et al. Using genome-wide pathway analysis to unravel the etiology of complex diseases. Genet Epidemiol. 33, 419-31. doi: 10.1002/gepi.20395. (2009).

175. Holmans, P. Statistical methods for pathway analysis of genome-wide data for association with complex genetic traits. Adv Genet. 72:141-79., 10.1016/B978-0-12-380862-2.00007-2. (2010).

176. de Leeuw, C.A., Neale, B.M., Heskes, T. & Posthuma, D. The statistical properties of gene-set analysis. Nat Rev Genet 17, 353-64 (2016).

177. Holmans, P. et al. Gene ontology analysis of GWA study data sets provides insights into the biology of bipolar disorder. Am J Hum Genet 85, 13-24 (2009).

178. Zhang, K., Cui, S., Chang, S., Zhang, L. & Wang, J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. Nucleic Acids Res 38, W90-5 (2010).

179. Wang, K., Li, M. & Bucan, M. Pathway-based approaches for analysis of genomewide association studies. Am J Hum Genet 81, 1278-83 (2007).

180. Medina, I. et al. Gene set-based analysis of polymorphisms: finding pathways or biological processes associated to traits in genome-wide association studies. Nucleic Acids Res 37, W340-4 (2009).

181. Chen, L.S. et al. Insights into colon cancer etiology via a regularized approach to gene set analysis of GWAS data. Am J Hum Genet 86, 860-71 (2010).

182. Nam, D., Kim, J., Kim, S.Y. & Kim, S. GSA-SNP: a general approach for gene set analysis of polymorphisms. Nucleic Acids Res 38, W749-54 (2010).

183. Holden, M., Deng, S., Wojnowski, L. & Kulle, B. GSEA-SNP: applying gene set enrichment analysis to SNP data from genome-wide association studies. Bioinformatics 24, 2784-5 (2008).

184. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 81, 559-75 (2007).

185. O'Dushlaine, C. et al. The SNP ratio test: pathway analysis of genome-wide association datasets. Bioinformatics 25, 2762-3 (2009).

186. Segre, A.V., Groop, L., Mootha, V.K., Daly, M.J. & Altshuler, D. Common inherited variation in mitochondrial genes is not enriched for associations with type 2 diabetes or related glycemic traits. PLoS Genet 6(2010).

187. Yaspan, B.L. et al. Genetic analysis of biological pathway data through genomic randomization. Hum Genet 129, 563-71 (2011).

188. Vivar, J.C., Pemu, P., McPherson, R. & Ghosh, S. Redundancy control in pathway databases (ReCiPa): an application for improving gene-set enrichment analysis in Omics studies and "Big data" biology. OMICS 17, 414-22 (2013).

189. Chen, D., Enroth, S., Ivansson, E. & Gyllensten, U. Pathway analysis of cervical cancer genome-wide association study highlights the MHC region and pathways involved in response to infection. Hum Mol Genet 23, 6047-60 (2014).

190. Wang, K. et al. Diverse genome-wide association studies associate the IL12/IL23 pathway with Crohn Disease. Am J Hum Genet 84, 399-405 (2009).

191. Askland, K., Read, C. & Moore, J. Pathways-based analyses of whole-genome association study data in bipolar disorder reveal genes mediating ion channel activity and synaptic neurotransmission. Hum Genet 125, 63-79 (2009).

192. Biological insights from 108 schizophrenia-associated genetic loci. Nature 511, 421-7 (2014).

193. Locke, A.E. et al. Genetic studies of body mass index yield new insights for obesity biology. Nature 518, 197-206 (2015).

194. Lango Allen, H. et al. Hundreds of variants clustered in genomic loci and biological pathways affect human height. Nature 467, 832-8 (2010).

195. Chen, D., Enroth, S., Ivansson, E. & Gyllensten, U. Pathway analysis of cervical cancer genome-wide association study highlights the MHC region and pathways involved in response to infection. Hum Mol Genet. 23, 6047-60. doi: 10.1093/hmg/ddu304. Epub 2014 Jun 16. (2014).

196. Koster, R. et al. Pathway-based analysis of GWAs data identifies association of sex determination genes with susceptibility to testicular germ cell tumors. Hum Mol Genet. 23, 6061-8. doi: 10.1093/hmg/ddu305. Epub 2014 Jun 18. (2014).

197. Menashe, I. et al. Pathway analysis of breast cancer genome-wide association study highlights three pathways and one canonical signaling cascade. Cancer Res. 70, 4453-9. doi: 10.1158/0008-5472.CAN-09-4502. Epub 2010 May 11. (2010).

198. Sullivan, P.F., Daly, M.J. & O'Donovan, M. Genetic architectures of psychiatric disorders: the emerging picture and its implications. Nat Rev Genet 13, 537-51 (2012).

199. Aterido, A. et al. Genome-wide pathway analysis identifies VEGF pathway association with oral ulceration in systemic lupus erythematosus. Arthritis Res Ther 19, 138 (2017).

200. Aterido, A. et al. Genome-Wide Pathway Analysis Identifies Genetic Pathways Associated with Psoriasis. J Invest Dermatol 136, 593-602 (2016).

201. Barabasi, A.L. & Oltvai, Z.N. Network biology: understanding the cell's functional organization. Nat Rev Genet 5, 101-13 (2004).

202. Barabasi, A.L., Gulbahce, N. & Loscalzo, J. Network medicine: a network-based approach to human disease. Nat Rev Genet 12, 56-68 (2011).

203. Zhu, X., Gerstein, M. & Snyder, M. Getting connected: analysis and principles of biological networks. Genes Dev 21, 1010-24 (2007).

204. Taylor, I.W. et al. Dynamic modularity in protein interaction networks predicts breast cancer outcome. Nat Biotechnol 27, 199-204 (2009).

205. Hahn, M.W. & Kern, A.D. Comparative genomics of centrality and essentiality in three eukaryotic protein-interaction networks. Mol Biol Evol 22, 803-6 (2005).

206. Joy, M.P., Brock, A., Ingber, D.E. & Huang, S. High-betweenness proteins in the yeast protein interaction network. J Biomed Biotechnol 2005, 96-103 (2005).

207. Vallabhajosyula, R.R., Chakravarti, D., Lutfeali, S., Ray, A. & Raval, A. Identifying hubs in protein interaction networks. PLoS One 4, e5344 (2009).

208. Jeong, H., Mason, S.P., Barabasi, A.L. & Oltvai, Z.N. Lethality and centrality in protein networks. Nature 411, 41-2 (2001).

209. Yu, H., Kim, P.M., Sprecher, E., Trifonov, V. & Gerstein, M. The importance of bottlenecks in protein networks: correlation with gene essentiality and expression dynamics. PLoS Comput Biol 3, e59 (2007).

210. Jeong, H., Tombor, B., Albert, R., Oltvai, Z.N. & Barabasi, A.L. The large-scale organization of metabolic networks. Nature 407, 651-4 (2000).

211. Seebacher, J. & Gavin, A.C. SnapShot: Protein-protein interaction networks. Cell 144, 1000, 1000 e1 (2011).

212. Albert, R., Jeong, H. & Barabasi, A.L. Error and attack tolerance of complex networks. Nature 406, 378-82 (2000).

213. Ideker, T. & Nussinov, R. Network approaches and applications in biology. PLoS Comput Biol 13, e1005771 (2017).

214. Bergholdt, R. et al. Identification of novel type 1 diabetes candidate genes by integrating genome-wide association data, protein-protein interactions, and human pancreatic islet gene expression. Diabetes 61, 954-62 (2012).

215. O'Roak, B.J. et al. Sporadic autism exomes reveal a highly interconnected protein network of de novo mutations. Nature 485, 246-50 (2012).

216. Lee, M.J. et al. Sequential application of anticancer drugs enhances cell death by rewiring apoptotic signaling networks. Cell 149, 780-94 (2012).

217. Yildirim, M.A., Goh, K.I., Cusick, M.E., Barabasi, A.L. & Vidal, M. Drug-target network. Nat Biotechnol 25, 1119-26 (2007).

218. Keiser, M.J. et al. Predicting new molecular targets for known drugs. Nature 462, 175-81 (2009).

219. Cami, A., Arnold, A., Manzi, S. & Reis, B. Predicting adverse drug events using pharmacological network models. Sci Transl Med 3, 114ra127 (2011).

220. Ala-Korpela, M., Kangas, A.J. & Inouye, M. Genome-wide association studies and systems biology: together at last. Trends Genet 27, 493-8 (2011).

221. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648-60 (2015).

222. Kim, M.S. et al. A draft map of the human proteome. Nature 509, 575-81 (2014).

223. Kundaje, A. et al. Integrative analysis of 111 reference human epigenomes. Nature 518, 317-30 (2015).

224. Shin, S.Y. et al. An atlas of genetic influences on human blood metabolites. Nat Genet 46, 543-550 (2014).

225. Ziller, M.J. et al. Charting a dynamic DNA methylation landscape of the human genome. Nature 500, 477-81 (2013).

226. Ritchie, M.D., Holzinger, E.R., Li, R., Pendergrass, S.A. & Kim, D. Methods of integrating data to uncover genotype-phenotype interactions. Nat Rev Genet 16, 85-97 (2015).

227. Schadt, E.E. et al. An integrative genomics approach to infer causal associations between gene expression and disease. Nat Genet 37, 710-7 (2005).

228. Fairfax, B.P. et al. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet 44, 502-10 (2012).

229. Pickrell, J.K. et al. Understanding mechanisms underlying human gene expression variation with RNA sequencing. Nature 464, 768-72 (2010).

230. Shi, J. et al. Characterizing the genetic basis of methylome diversity in histologically normal human lung tissue. Nat Commun 5, 3365 (2014).

231. Heyn, H. et al. DNA methylation contributes to natural human variation. Genome Res 23, 1363-72 (2013).

232. Teh, A.L. et al. The effect of genotype and in utero environment on interindividual variation in neonate DNA methylomes. Genome Res 24, 1064-74 (2014).

233. Gibbs, J.R. et al. Abundant quantitative trait loci exist for DNA methylation and gene expression in human brain. PLoS Genet 6, e1000952 (2010).

234. Grundberg, E. et al. Global analysis of DNA methylation variation in adipose tissue from twins reveals links to disease-associated variants in distal regulatory elements. Am J Hum Genet 93, 876-90 (2013).

235. Heyn, H. et al. Linkage of DNA methylation quantitative trait loci to human cancer risk. Cell Rep 7, 331-8 (2014).

236. Trynka, G. et al. Chromatin marks identify critical cell types for fine mapping complex trait variants. Nat Genet 45, 124-30 (2013).

237. Farh, K.K. et al. Genetic and epigenetic fine mapping of causal autoimmune disease variants. Nature 518, 337-43 (2015).

238. Akiyama, M. et al. Genome-wide association study identifies 112 new loci for body mass index in the Japanese population. Nat Genet 49, 1458-1467 (2017).

239. Kettunen, J. et al. Genome-wide association study identifies multiple loci influencing human serum metabolite levels. Nat Genet 44, 269-76 (2012).

240. Suhre, K. et al. Human metabolic individuality in biomedical and pharmaceutical research. Nature 477, 54-60 (2011).

241. Di Narzo, A.F. et al. High-Throughput Characterization of Blood Serum Proteomics of IBD Patients with Respect to Aging and Genetic Factors. PLoS Genet 13, e1006565 (2017).

242. Alonso, A. et al. Identification of risk loci for Crohn's disease phenotypes using a genome-wide association study. Gastroenterology 148, 794-805 (2015).

243. Julia, A. et al. A genome-wide association study identifies SLC8A3 as a susceptibility locus for ACPA-positive rheumatoid arthritis. Rheumatology (Oxford) 55, 1106-11 (2016).

244. Stuart, P.E. et al. Genome-wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture. Am J Hum Genet 97, 816-36 (2015).

245. Harley, J.B. et al. Genome-wide association scan in women with systemic lupus erythematosus identifies susceptibility variants in ITGAM, PXK, KIAA1542 and other loci. Nat Genet 40, 204-10 (2008).

246. Bhalerao, J. & Bowcock, A.M. The genetics of psoriasis: a complex disorder of the skin and immune system. Hum Mol Genet. 7, 1537-45. (1998).

247. Chandran, V. et al. Familial aggregation of psoriatic arthritis. Ann Rheum Dis. 68, 664-7. doi: 10.1136/ard.2008.089367. Epub 2008 Jun 4. (2009).

248. Elder, J.T. et al. The genetics of psoriasis. Arch Dermatol. 130, 216-24. (1994).

249. Gladman, D.D., Farewell, V.T., Pellett, F., Schentag, C. & Rahman, P. HLA is a candidate region for psoriatic arthritis. evidence for excessive HLA sharing in sibling pairs. Hum Immunol. 64, 887-9. (2003).

250. Lessard, C.J. et al. The genomics of autoimmune disease in the era of genome-wide association studies and beyond. Autoimmun Rev 11, 267-75 (2012).

251. Fernando, M.M. et al. Defining the role of the MHC in autoimmunity: a review and pooled analysis. PLoS Genet 4, e1000024 (2008).

252. Begovich, A.B. et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 75, 330-7 (2004).

253. Graham, R.R. et al. A common haplotype of interferon regulatory factor 5 (IRF5) regulates splicing and expression and is associated with increased risk of systemic lupus erythematosus. Nat Genet 38, 550-5 (2006).

254. Hugot, J.P. et al. Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. Nature 411, 599-603 (2001).

255. Tomfohrde, J. et al. Gene for familial psoriasis susceptibility mapped to the distal end of human chromosome 17q. Science 264, 1141-5 (1994).

256. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature 506, 376-81 (2014).

257. Bowes, J. et al. Dense genotyping of immune-related susceptibility loci reveals new insights into the genetics of psoriatic arthritis. Nat Commun 6, 6046 (2015).

258. Huffmeier, U. et al. Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis. Nat Genet 42, 996-9 (2010).

259. Bentham, J. et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet 47, 1457-64 (2015).

260. Jostins, L. et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 491, 119-24 (2012).

261. Stuart, P.E. et al. Genome-wide Association Analysis of Psoriatic Arthritis and Cutaneous Psoriasis Reveals Differences in Their Genetic Architecture. Am J Hum Genet. 97, 816-36. doi: 10.1016/j.ajhg.2015.10.019. Epub 2015 Nov 28. (2015).

262. Bowes, J. et al. Dense genotyping of immune-related susceptibility loci reveals new insights into the genetics of psoriatic arthritis. Nat Commun. 6:6046., 10.1038/ncomms7046. (2015).

263. Liu, Y. et al. A genome-wide association study of psoriasis and psoriatic arthritis identifies new disease loci. PLoS Genet. 4, e1000041. doi: 10.1371/journal.pgen.1000041. (2008).

264. Huffmeier, U. et al. Common variants at TRAF3IP2 are associated with susceptibility to psoriatic arthritis and psoriasis. Nat Genet. 42, 996-9. doi: 10.1038/ng.688. Epub 2010 Oct 17. (2010).

265. Ellinghaus, E. et al. Genome-wide meta-analysis of psoriatic arthritis identifies susceptibility locus at REL. J Invest Dermatol. 132, 1133-40. doi: 10.1038/jid.2011.415. Epub 2011 Dec 15. (2012).

266. Julia, A. et al. A deletion at ADAMTS9-MAGI1 locus is associated with psoriatic arthritis risk. Ann Rheum Dis. 74, 1875-81. doi: 10.1136/annrheumdis-2014-207190. Epub 2015 May 19. (2015).

267. Apel, M. et al. Variants in RUNX3 contribute to susceptibility to psoriatic arthritis, exhibiting further common ground with ankylosing spondylitis. Arthritis Rheum. 65, 1224-31. doi: 10.1002/art.37885. (2013).

268. Bowes, J. et al. PTPN22 is associated with susceptibility to psoriatic arthritis but not psoriasis: evidence for a further PsA-specific risk locus. Ann Rheum Dis. 74, 1882-5. doi: 10.1136/annrheumdis-2014-207187. Epub 2015 Apr 28. (2015).

269. Chen, H. et al. A genetic risk score combining ten psoriasis risk loci improves disease prediction. PLoS One 6, e19454 (2011).

270. Tsoi, L.C. et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat Genet 44, 1341-8 (2012).

271. Yin, X. et al. Common variants explain a large fraction of the variability in the liability to psoriasis in a Han Chinese population. BMC Genomics 15, 87 (2014).

272. Raychaudhuri, S. et al. Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. Nat Genet 44, 291-6 (2012).

273. de Bakker, P.I. et al. A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. Nat Genet 38, 1166-72 (2006).

274. Horton, R. et al. Gene map of the extended human MHC. Nat Rev Genet 5, 889-99 (2004).

275. Ioannidis, J.P., Thomas, G. & Daly, M.J. Validating, augmenting and refining genome-wide association signals. Nat Rev Genet 10, 318-29 (2009).

276. de Bakker, P.I. & Raychaudhuri, S. Interrogating the major histocompatibility complex with high-throughput genomics. Hum Mol Genet 21, R29-36 (2012).

277. Cotsapas, C. et al. Pervasive sharing of genetic effects in autoimmune disease. PLoS Genet 7, e1002254 (2011).

278. Solovieff, N., Cotsapas, C., Lee, P.H., Purcell, S.M. & Smoller, J.W. Pleiotropy in complex traits: challenges and strategies. Nat Rev Genet 14, 483-95 (2013).

279. Gabrielsen, I.S. et al. Genetic risk variants for autoimmune diseases that influence gene expression in thymus. Hum Mol Genet 25, 3117-3124 (2016).

280. Peters, J.E. et al. Insight into Genotype-Phenotype Associations through eQTL Mapping in Multiple Cell Types in Health and Immune-Mediated Disease. PLoS Genet 12, e1005908 (2016).

281. Kim-Hellmuth, S. et al. Genetic regulatory effects modified by immune activation contribute to autoimmune disease associations. Nat Commun 8, 266 (2017).

282. Aterido, A., Palacio, C., Marsal, S., Avila, G. & Julia, A. Novel insights into the regulatory architecture of CD4+ T cells in rheumatoid arthritis. PLoS One 9, e100690 (2014).

283. Yap, C.X. et al. Trans-eQTLs identified in whole blood have limited influence on complex disease biology. Eur J Hum Genet 26, 1361-1368 (2018).

284. Bayless, T.M. et al. Crohn's disease: concordance for site and clinical type in affected family members--potential hereditary influences. Gastroenterology 111, 573-9 (1996).

285. Olson, J.M. et al. A genome screen of systemic lupus erythematosus using affected-relative-pair linkage analysis with covariates demonstrates genetic heterogeneity. Genes Immun 3 Suppl 1, S5-S12 (2002).

286. Verstockt, B., Smith, K.G. & Lee, J.C. Genome-wide association studies in Crohn's disease: Past, present and future. Clin Transl Immunology 7, e1001 (2018).

287. Economou, M., Trikalinos, T.A., Loizou, K.T., Tsianos, E.V. & Ioannidis, J.P. Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. Am J Gastroenterol 99, 2393-404 (2004).

288. Adler, J., Rangwalla, S.C., Dwamena, B.A. & Higgins, P.D. The prognostic power of the NOD2 genotype for complicated Crohn's disease: a meta-analysis. Am J Gastroenterol 106, 699-712 (2011).

289. Taylor, K.E. et al. Specificity of the STAT4 genetic association for severe disease manifestations of systemic lupus erythematosus. PLoS Genet 4, e1000084 (2008).

290. Sanchez, E. et al. Phenotypic associations of genetic susceptibility loci in systemic lupus erythematosus. Ann Rheum Dis 70, 1752-7 (2011).

291. Julia, A. et al. Risk variants for psoriasis vulgaris in a large case-control collection and association with clinical subphenotypes. Hum Mol Genet 21, 4549-57 (2012).

292. Yao, F. et al. A genetic coding variant rs72474224 in GJB2 is associated with clinical features of psoriasis vulgaris in a Chinese Han population. Tissue Antigens 86, 134-8 (2015).

293. Polachek, A. et al. The Association Between HLA Genetic Susceptibility Markers and Sonographic Enthesitis in Psoriatic Arthritis. Arthritis Rheumatol 70, 756-762 (2018).

294. Haroon, M., Winchester, R., Giles, J.T., Heffernan, E. & FitzGerald, O. Clinical and genetic associations of radiographic sacroiliitis and its different patterns in psoriatic arthritis. Clin Exp Rheumatol 35, 270-276 (2017).

295. Torices, S. et al. A functional variant of TLR10 modifies the activity of NFkB and may help predict a worse prognosis in patients with rheumatoid arthritis. Arthritis Res Ther 18, 221 (2016).

296. Cleynen, I. et al. Inherited determinants of Crohn's disease and ulcerative colitis phenotypes: a genetic association study. Lancet 387, 156-67 (2016).

297. Traylor, M. et al. Genetic risk factors for ischaemic stroke and its subtypes (the METASTROKE collaboration): a meta-analysis of genome-wide association studies. Lancet Neurol 11, 951-62 (2012).

298. Tabas, I. & Glass, C.K. Anti-inflammatory therapy in chronic disease: challenges and opportunities. Science 339, 166-72 (2013).

299. Elliott, M.J. et al. Randomised double-blind comparison of chimeric monoclonal antibody to tumour necrosis factor alpha (cA2) versus placebo in rheumatoid arthritis. Lancet 344, 1105-10 (1994).

300. Kempeni, J. Preliminary results of early clinical trials with the fully human anti-TNFalpha monoclonal antibody D2E7. Ann Rheum Dis 58 Suppl 1, I70-2 (1999).

301. Moreland, L.W. et al. Treatment of rheumatoid arthritis with a recombinant human tumor necrosis factor receptor (p75)-Fc fusion protein. N Engl J Med 337, 141-7 (1997).

302. Kalden, J.R. & Schulze-Koops, H. Immunogenicity and loss of response to TNF inhibitors: implications for rheumatoid arthritis treatment. Nat Rev Rheumatol 13, 707-718 (2017).

303. Rutgeerts, P., Van Assche, G. & Vermeire, S. Optimizing anti-TNF treatment in inflammatory bowel disease. Gastroenterology 126, 1593-610 (2004).

304. Leonardi, C.L. et al. Etanercept as monotherapy in patients with psoriasis. N Engl J Med 349, 2014-22 (2003).

305. Kievit, W. et al. The effectiveness and medication costs of three anti-tumour necrosis factor alpha agents in the treatment of rheumatoid arthritis from prospective clinical practice data. Ann Rheum Dis 67, 1229-34 (2008).

306. Saad, A.A. et al. Efficacy and safety of anti-TNF therapies in psoriatic arthritis: an observational study from the British Society for Rheumatology Biologics Register. Rheumatology (Oxford) 49, 697-705 (2010).

307. Kristensen, L.E., Gulfe, A., Saxne, T. & Geborek, P. Efficacy and tolerability of anti-tumour necrosis factor therapy in psoriatic arthritis patients: results from the South Swedish Arthritis Treatment Group register. Ann Rheum Dis 67, 364-9 (2008).

308. Nast, A., Jacobs, A., Rosumeck, S. & Werner, R.N. Efficacy and Safety of Systemic Long-Term Treatments for Moderate-to-Severe Psoriasis: A Systematic Review and Meta-Analysis. J Invest Dermatol 135, 2641-2648 (2015).

309. Gisbert, J.P. & Panes, J. Loss of response and requirement of infliximab dose intensification in Crohn's disease: a review. Am J Gastroenterol 104, 760-7 (2009).

310. Rutgeerts, P. et al. Infliximab for induction and maintenance therapy for ulcerative colitis. N Engl J Med 353, 2462-76 (2005).

311. Chaparro, M., Guerra, I., Munoz-Linares, P. & Gisbert, J.P. Systematic review: antibodies and anti-TNF-alpha levels in inflammatory bowel disease. Aliment Pharmacol Ther 35, 971-86 (2012).

312. Garces, S., Demengeot, J. & Benito-Garcia, E. The immunogenicity of anti-TNF therapy in immune-mediated inflammatory diseases: a systematic review of the literature with a meta-analysis. Ann Rheum Dis 72, 1947-55 (2013).

313. Paul, S. et al. Pharmacokinetics of adalimumab in inflammatory bowel diseases: a systematic review and meta-analysis. Inflamm Bowel Dis 20, 1288-95 (2014).

314. Pascual-Salcedo, D. et al. Influence of immunogenicity on the efficacy of long-term treatment with infliximab in rheumatoid arthritis. Rheumatology (Oxford) 50, 1445-52 (2011).

315. Thomas, S.S. et al. Comparative Immunogenicity of TNF Inhibitors: Impact on Clinical Efficacy and Tolerability in the Management of Autoimmune Diseases. A Systematic Review and Meta-Analysis. BioDrugs 29, 241-58 (2015).

316. Vincent, F.B. et al. Antidrug antibodies (ADAb) to tumour necrosis factor (TNF)-specific neutralising agents in chronic inflammatory diseases: a real issue, a clinical perspective. Ann Rheum Dis 72, 165-78 (2013).

317. Atzeni, F. et al. Investigating the potential side effects of anti-TNF therapy for rheumatoid arthritis: cause for concern? Immunotherapy 7, 353-61 (2015).

318. Bremmer, M., Deng, A. & Gaspari, A.A. A mechanism-based classification of dermatologic reactions to biologic agents used in the treatment of cutaneous disease: Part 1. Dermatitis 20, 182-92 (2009).

319. van Schouwenburg, P.A., Rispens, T. & Wolbink, G.J. Immunogenicity of anti-TNF biologic therapies for rheumatoid arthritis. Nat Rev Rheumatol 9, 164-72 (2013).

320. Prieto-Perez, R., Cabaleiro, T., Dauden, E. & Abad-Santos, F. Gene polymorphisms that can predict response to anti-TNF therapy in patients with psoriasis and related autoimmune diseases. Pharmacogenomics J 13, 297-305 (2013).

321. Madian, A.G., Wheeler, H.E., Jones, R.B. & Dolan, M.E. Relating human genetic variation to variation in drug responses. Trends Genet 28, 487-95 (2012).

322. Bek, S. et al. Systematic review and meta-analysis: pharmacogenetics of anti-TNF treatment response in rheumatoid arthritis. Pharmacogenomics J 17, 403-411 (2017).

323. Guilliams, M., Bruhns, P., Saeys, Y., Hammad, H. & Lambrecht, B.N. The function of Fcgamma receptors in dendritic cells and macrophages. Nat Rev Immunol 14, 94-108 (2014).

324. Arora, T. et al. Differences in binding and effector functions between classes of TNF antagonists. Cytokine 45, 124-31 (2009).

325. Mimoto, F. et al. Crystal structure of a novel asymmetrically engineered Fc variant with improved affinity for FcgammaRs. Mol Immunol 58, 132-8 (2014).

326. Avila-Pedretti, G. et al. Variation at FCGR2A and functionally related genes is associated with the response to anti-TNF therapy in rheumatoid arthritis. PLoS One 10, e0122088 (2015).

327. Rufini, S., Ciccacci, C., Novelli, G. & Borgiani, P. Pharmacogenetics of inflammatory bowel disease: a focus on Crohn's disease. Pharmacogenomics 18, 1095-1114 (2017).

328. Ovejero-Benito, M.C. et al. Pharmacogenetics and Pharmacogenomics in Moderate-to-Severe Psoriasis. Am J Clin Dermatol 19, 209-222 (2018).

329. Ovejero-Benito, M.C. et al. Polymorphisms associated with etanercept response in moderate-to-severe plaque psoriasis. Pharmacogenomics 18, 631-638 (2017).

330. Gallo, E. et al. The relationship between tumour necrosis factor (TNF)-alpha promoter and IL12B/IL-23R genes polymorphisms and the efficacy of anti-TNF-alpha therapy in psoriasis: a case-control study. Br J Dermatol 169, 819-29 (2013).

331. Murdaca, G. et al. Pharmacogenetics of etanercept: role of TNF-alpha gene polymorphisms in improving its efficacy. Expert Opin Drug Metab Toxicol 10, 1703-10 (2014).

332. Jani, M., Barton, A. & Ho, P. Pharmacogenetics of treatment response in psoriatic arthritis. Curr Rheumatol Rep 17, 44 (2015).

333. Julia, A. et al. PDE3A-SLCO1C1 locus is associated with response to anti-tumor necrosis factor therapy in psoriatic arthritis. Pharmacogenomics 15, 1763-1769 (2014).

334. Cui, J. et al. Genome-wide association study and gene expression analysis identifies CD84 as a predictor of response to etanercept therapy in rheumatoid arthritis. PLoS Genet 9, e1003394 (2013).

335. Umicevic Mirkov, M. et al. Genome-wide association analysis of anti-TNF drug response in patients with rheumatoid arthritis. Ann Rheum Dis 72, 1375-81 (2013).

336. Honne, K. et al. A longitudinal genome-wide association study of anti-tumor necrosis factor response among Japanese patients with rheumatoid arthritis. Arthritis Res Ther 18, 12 (2016).

337. Julia, A. et al. A genome-wide association study identifies a new locus associated with the response to anti-TNF therapy in rheumatoid arthritis. Pharmacogenomics J 16, 147-50 (2016).

338. Krintel, S.B. et al. Investigation of single nucleotide polymorphisms and biological pathways associated with response to TNFalpha inhibitors in patients with rheumatoid arthritis. Pharmacogenet Genomics 22, 577-89 (2012).

339. Plant, D. et al. Genome-wide association study of genetic predictors of anti-tumor necrosis factor treatment efficacy in rheumatoid arthritis identifies associations with polymorphisms at seven loci. Arthritis Rheum 63, 645-53 (2011).

340. Acosta-Colman, I. et al. GWAS replication study confirms the association of PDE3A-SLCO1C1 with anti-TNF therapy response in rheumatoid arthritis. Pharmacogenomics 14, 727-34 (2013).

341. Nishikawa, R. et al. Genetic prediction of the effectiveness of biologics for psoriasis treatment. J Dermatol 43, 1273-1277 (2016).

342. Julia, A. & Marsal, S. Pharmacogenomics of anti-TNF response in psoriasis, where are we? Pharmacogenomics 17, 323-6 (2016).

343. Julia, A. et al. Association of the PDE3A-SLCO1C1 locus with the response to anti-TNF agents in psoriasis. Pharmacogenomics J 15, 322-5 (2015).

344. Maini, R.N. et al. Therapeutic efficacy of multiple intravenous infusions of anti-tumor necrosis factor alpha monoclonal antibody combined with low-dose weekly methotrexate in rheumatoid arthritis. Arthritis Rheum 41, 1552-63 (1998).

345. Chirino, A.J. & Mire-Sluis, A. Characterizing biological products and assessing comparability following manufacturing changes. Nat Biotechnol 22, 1383-91 (2004).

346. Bartelds, G.M. et al. Anti-adalimumab antibodies in rheumatoid arthritis patients are associated with interleukin-10 gene polymorphisms. Arthritis Rheum 60, 2541-2 (2009).

347. Liu, M. et al. Identification of HLA-DRB1 association to adalimumab immunogenicity. PLoS One 13, e0195325 (2018).

348. Billiet, T. et al. Immunogenicity to infliximab is associated with HLA-DRB1. Gut 64, 1344-5 (2015).

349. Romero-Cara, P. et al. A FCGR3A Polymorphism Predicts Anti-drug Antibodies in Chronic Inflammatory Bowel Disease Patients Treated With Anti-TNF. Int J Med Sci 15, 10-15 (2018).

350. Magdelaine-Beuzelin, C. et al. IgG1 heavy chain-coding gene polymorphism (G1m allotypes) and development of antibodies-to-infliximab. Pharmacogenet Genomics 19, 383-7 (2009).

351. Hoffmann, S. et al. HLA-DRB1*0401 and HLA-DRB1*0408 are strongly associated with the development of antibodies against interferon-beta therapy in multiple sclerosis. Am J Hum Genet 83, 219-27 (2008).

352. Scharfe, C.P.I., Tremmel, R., Schwab, M., Kohlbacher, O. & Marks, D.S. Genetic variation in human drug-related genes. Genome Med 9, 117 (2017).

353. Mallal, S. et al. HLA-B*5701 screening for hypersensitivity to abacavir. N Engl J Med 358, 568-79 (2008).

354. Dishy, V. et al. The effect of common polymorphisms of the beta2-adrenergic receptor on agonist-mediated vascular desensitization. N Engl J Med 345, 1030-5 (2001).

355. Evans, W.E. & McLeod, H.L. Pharmacogenomics--drug disposition, drug targets, and side effects. N Engl J Med 348, 538-49 (2003).

356. Crews, K.R., Hicks, J.K., Pui, C.H., Relling, M.V. & Evans, W.E. Pharmacogenomics and individualized medicine: translating science into practice. Clin Pharmacol Ther 92, 467-75 (2012).

357. Tsoi, L.C. et al. Identification of 15 new psoriasis susceptibility loci highlights the role of innate immunity. Nat Genet. 44, 1341-8. doi: 10.1038/ng.2467. Epub 2012 Nov 11. (2012).

358. Yin, X. et al. Common variants explain a large fraction of the variability in the liability to psoriasis in a Han Chinese population. BMC Genomics. 15:87., 10.1186/1471-2164-15-87. (2014).

359. Johnson, R.C. et al. Accounting for multiple comparisons in a genome-wide association study (GWAS). BMC Genomics. 11:724., 10.1186/1471-2164-11-724. (2010).

360. Balato, A. et al. Interleukin-1 family members are enhanced in psoriasis and suppressed by vitamin D and retinoic acid. Arch Dermatol Res. 305, 255-62. doi: 10.1007/s00403-013-1327-8. Epub 2013 Feb 24. (2013).

361. Austenaa, L.M. et al. Vitamin A status significantly alters nuclear factor-kappaB activity assessed by in vivo imaging. Faseb J. 18, 1255-7. Epub 2004 Jun 4. (2004).

362. Goldminz, A.M., Au, S.C., Kim, N., Gottlieb, A.B. & Lizzul, P.F. NF-kappaB: an essential transcription factor in psoriasis. J Dermatol Sci. 69, 89-94. doi: 10.1016/j.jdermsci.2012.11.002. Epub 2012 Nov 14. (2013).

363. Tsuruta, D. NF-kappaB links keratinocytes and lymphocytes in the pathogenesis of psoriasis. Recent Pat Inflamm Allergy Drug Discov. 3, 40-8. (2009).

364. Lizzul, P.F. et al. Differential expression of phosphorylated NF-kappaB/RelA in normal and psoriatic epidermis and downregulation of NF-kappaB in response to treatment with etanercept. J Invest Dermatol. 124, 1275-83. (2005).

365. Beckenbach, L., Baron, J.M., Merk, H.F., Loffler, H. & Amann, P.M. Retinoid treatment of skin diseases. Eur J Dermatol 25, 384-91 (2015).

366. Lang, F., Stournaras, C. & Alesutan, I. Regulation of transport across cell membranes by the serum- and glucocorticoid-inducible kinase SGK1. Mol Membr Biol. 31, 29-36. doi: 10.3109/09687688.2013.874598. Epub 2014 Jan 14. (2014).

367. Vig, M. & Kinet, J.P. Calcium signaling in immune cells. Nat Immunol. 10, 21-7. doi: 10.1038/ni.f.220. (2009).

368. Khananshvili, D. The SLC8 gene family of sodium-calcium exchangers (NCX) - structure, function, and regulation in health and disease. Mol Aspects Med. 34, 220-35. doi: 10.1016/j.mam.2012.07.003. (2013).

369. Jaeger, K., Paulsen, F. & Wohlrab, J. Characterization of cationic amino acid transporters (hCATs) 1 and 2 in human skin. Histochem Cell Biol. 129, 321-9. doi: 10.1007/s00418-007-0367-6. Epub 2008 Jan 3. (2008).

370. Rudd, P.M., Elliott, T., Cresswell, P., Wilson, I.A. & Dwek, R.A. Glycosylation and the immune system. Science. 291, 2370-6. (2001).

371. Ryan, S.O. & Cobb, B.A. Roles for major histocompatibility complex glycosylation in immune function. Semin Immunopathol. 34, 425-41. doi: 10.1007/s00281-012-0309-9. Epub 2012 Mar 30. (2012).

372. Iversen, O.J., Lysvand, H. & Hagen, L. The autoantigen Pso p27: a post-translational modification of SCCA molecules. Autoimmunity. 44, 229-34. doi: 10.3109/08916934.2010.530628. Epub 2011 Jan 14. (2011).

373. Damasiewicz-Bodzek, A. & Wielkoszynski, T. Advanced protein glycation in psoriasis. J Eur Acad Dermatol Venereol. 26, 172-9. doi: 10.1111/j.1468-3083.2011.04024.x. Epub 2011 Mar 14. (2012).

374. Fuhlbrigge, R.C., Kieffer, J.D., Armerding, D. & Kupper, T.S. Cutaneous lymphocyte antigen is a specialized form of PSGL-1 expressed on skin-homing T cells. Nature. 389, 978-81. (1997).

375. Mkhikian, H. et al. Genetics and the environment converge to dysregulate N-glycosylation in multiple sclerosis. Nat Commun. 2:334., 10.1038/ncomms1333. (2011).

376. Grigorian, A. & Demetriou, M. Mgat5 deficiency in T cells and experimental autoimmune encephalomyelitis. ISRN Neurol. 2011:374314., 10.5402/2011/374314. Epub 2011 Aug 17. (2011).

377. Demetriou, M., Granovsky, M., Quaggin, S. & Dennis, J.W. Negative regulation of T-cell activation and autoimmunity by Mgat5 N-glycosylation. Nature. 409, 733-9. (2001).

378. Fujii, H. et al. Core Fucosylation on T Cells, Required for Activation of T-Cell Receptor Signaling and Induction of Colitis in Mice, Is Increased in Patients With Inflammatory Bowel Disease. Gastroenterology 150, 1620-1632 (2016).

379. Julia, A. et al. A deletion at ADAMTS9-MAGI1 locus is associated with psoriatic arthritis risk. Ann Rheum Dis 74, 1875-81 (2015).

380. Bowes, J. et al. PTPN22 is associated with susceptibility to psoriatic arthritis but not psoriasis: evidence for a further PsA-specific risk locus. Ann Rheum Dis 74, 1882-5 (2015).

381. Togayachi, A. et al. Polylactosamine on glycoproteins influences basal levels of lymphocyte and macrophage activation. Proc Natl Acad Sci U S A 104, 15829-34 (2007).

382. Okada, Y. et al. Meta-analysis identifies nine new loci associated with rheumatoid arthritis in the Japanese population. Nat Genet. 44, 511-6. doi: 10.1038/ng.2231. (2012).

383. Reveille, J.D. et al. Genome-wide association study of ankylosing spondylitis identifies non-MHC susceptibility loci. Nat Genet. 42, 123-7. doi: 10.1038/ng.513. Epub 2010 Jan 10. (2010).

384. Togayachi, A. et al. Beta3GnT2 (B3GNT2), a major polylactosamine synthase: analysis of B3GNT2-deficient mice. Methods Enzymol 479, 185-204 (2010).

385. Afratis, N. et al. Glycosaminoglycans: key players in cancer cell biology and treatment. Febs J. 279, 1177-97. doi: 10.1111/j.1742-4658.2012.08529.x. Epub 2012 Mar 12. (2012).

386. Hansen, C., Otto, E., Kuhlemann, K., Forster, G. & Kahaly, G.J. Glycosaminoglycans in autoimmunity. Clin Exp Rheumatol. 14, S59-67. (1996).

387. Schett, G., Coates, L.C., Ash, Z.R., Finzel, S. & Conaghan, P.G. Structural damage in rheumatoid arthritis, psoriatic arthritis, and ankylosing spondylitis: traditional views, novel insights gained from TNF blockade, and concepts for the future. Arthritis Res Ther. 13, S4. doi: 10.1186/1478-6354-13-S1-S4. (2011).

388. Caterson, B., Flannery, C.R., Hughes, C.E. & Little, C.B. Mechanisms involved in cartilage proteoglycan catabolism. Matrix Biol. 19, 333-44. (2000).

389. Arner, E.C., Hughes, C.E., Decicco, C.P., Caterson, B. & Tortorella, M.D. Cytokine-induced cartilage proteoglycan degradation is mediated by aggrecanase. Osteoarthritis Cartilage. 6, 214-28. (1998).

390. van den Boom, R. et al. Relationship between synovial fluid levels of glycosaminoglycans, hydroxyproline and general MMP activity and the presence and severity of articular cartilage change on the proximal articular surface of P1. Equine Vet J. 37, 19-25. (2005).

391. Patterson, A.M. et al. Differential expression of syndecans and glypicans in chronically inflamed synovium. Ann Rheum Dis 67, 592-601 (2008).

392. Gibbs, J.B. Mechanism-based target identification and drug discovery in cancer research. Science. 287, 1969-73. (2000).

393. Okada, Y. et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 506, 376-81. doi: 10.1038/nature12873. Epub 2013 Dec 25. (2014).

394. Bentham, J. et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat Genet. 47, 1457-64. doi: 10.1038/ng.3434. Epub 2015 Oct 26. (2015).

395. Olson, J.M. et al. A genome screen of systemic lupus erythematosus using affected-relative-pair linkage analysis with covariates demonstrates genetic heterogeneity. Genes Immun. 3, S5-S12. (2002).

396. AS, J.d., C, A., P, S.G. & S, C. Systemic Lupus Erythematosus: Old and New Susceptibility Genes versus Clinical Manifestations. Curr Genomics. 15, 52-65. doi: 10.2174/1389202915011403061137l5. (2014).

397. Shibuya, M. & Claesson-Welsh, L. Signal transduction by VEGF receptors in regulation of angiogenesis and lymphangiogenesis. Exp Cell Res. 312, 549-60. Epub 2005 Dec 5. (2006).

398. Ferrara, N. Vascular endothelial growth factor: basic science and clinical progress. Endocr Rev. 25, 581-611. (2004).

399. Walchner, M., Meurer, M., Plewig, G. & Messer, G. Clinical and immunologic parameters during thalidomide treatment of lupus erythematosus. Int J Dermatol. 39, 383-8. (2000).

400. Cuadrado, M.J. et al. Thalidomide for the treatment of resistant cutaneous lupus: efficacy and safety of different therapeutic regimens. Am J Med. 118, 246-50. (2005).

401. Munoz-Corcuera, M., Esparza-Gomez, G., Gonzalez-Moles, M.A. & Bascones-Martinez, A. Oral ulcers: clinical aspects. A tool for dermatologists. Part II. Chronic ulcers. Clin Exp Dermatol. 34, 456-61. doi: 10.1111/j.1365-2230.2009.03219.x. Epub 2009 Apr 14. (2009).

402. Arbiser, J.L., Johnson, D., Cohen, C. & Brown, L.F. High-level expression of vascular endothelial growth factor and its receptors in an aphthous ulcer. J Cutan Med Surg. 7, 225-8. (2003).

403. Mustafa, M.B., Porter, S.R., Smoller, B.R. & Sitaru, C. Oral mucosal manifestations of autoimmune skin diseases. Autoimmun Rev. 14, 930-51. doi: 10.1016/j.autrev.2015.06.005. Epub 2015 Jun 24. (2015).

404. Khatibi, M., Shakoorpour, A.H., Jahromi, Z.M. & Ahmadzadeh, A. The prevalence of oral mucosal lesions and related factors in 188 patients with systemic lupus erythematosus. Lupus. 21, 1312-5. Epub 2012 Jul 25. (2012).

405. Urman, J.D., Lowenstein, M.B., Abeles, M. & Weinstein, A. Oral mucosal ulceration in systemic lupus erythematosus. Arthritis Rheum. 21, 58-61. (1978).

406. Parodi, A. et al. Measuring the activity of the disease in patients with cutaneous lupus erythematosus. Br J Dermatol. 142, 457-60. (2000).

407. Yalcin, B., Arda, N., Tezel, G.G., Erman, M. & Alli, N. Expressions of vascular endothelial growth factor and CD34 in oral aphthous lesions of Behcet's disease. Anal Quant Cytol Histol. 28, 303-6. (2006).

408. Kim, Y.S. et al. Novel single nucleotide polymorphism of the VEGF gene as a risk predictor for gastroduodenal ulcers. J Gastroenterol Hepatol. 23, S131-9. doi: 10.1111/j.1440-1746.2008.05404.x. (2008).

409. Wang, Y.F. et al. Identification of ST3AGL4, MFHAS1, CSNK2A2 and CD226 as loci associated with systemic lupus erythematosus (SLE) and evaluation of SLE genetics in drug repositioning. Ann Rheum Dis 77, 1078-1084 (2018).

410. Mason, J.C. & Libby, P. Cardiovascular disease in patients with chronic inflammation: mechanisms underlying premature cardiovascular events in rheumatologic conditions. Eur Heart J 36, 482-9c (2015).

411. Deloukas, P. et al. Large-scale association analysis identifies new risk loci for coronary artery disease. Nat Genet 45, 25-33 (2013).

412. Matsukura, M. et al. Genome-Wide Association Study of Peripheral Arterial Disease in a Japanese Population. PLoS One 10, e0139262 (2015).

413. Lin, E.A. & Liu, C.J. The role of ADAMTSs in arthritis. Protein Cell 1, 33-47 (2010).

414. Reilly, M.P. et al. Identification of ADAMTS7 as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. Lancet 377, 383-92 (2011).

415. Kanno, Y., Vahedi, G., Hirahara, K., Singleton, K. & O'Shea, J.J. Transcriptional and epigenetic control of T helper cell specification: molecular mechanisms underlying commitment and plasticity. Annu Rev Immunol 30, 707-31 (2012).

416. Talmud, P.J. et al. Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip. Am J Hum Genet 85, 628-42 (2009).

417. Lopez-Mejias, R. et al. A genome-wide association study identifies a 3'UTR genetic variant of RARB associated with carotid intima-media thickness in rheumatoid arthritis. Arthritis Rheumatol (2018).

418. Blandizzi, C. et al. The role of tumour necrosis factor in the pathogenesis of immune-mediated diseases. Int J Immunopathol Pharmacol 27, 1-10 (2014).

419. Ferrari, R. The role of TNF in cardiovascular disease. Pharmacol Res 40, 97-105 (1999).

420. Kekow, J., Mueller-Ladner, U. & Schulze-Koops, H. Rituximab is more effective than second anti-TNF therapy in rheumatoid arthritis patients and previous TNFalpha blocker failure. Biologics 6, 191-9 (2012).

421. van der Pouw Kraan, T.C. et al. Responsiveness to anti-tumour necrosis factor alpha therapy is related to pre-treatment tissue inflammation levels in rheumatoid arthritis patients. Ann Rheum Dis 67, 563-6 (2008).

422. Badot, V. et al. Gene expression profiling in the synovium identifies a predictive signature of absence of response to adalimumab therapy in rheumatoid arthritis. Arthritis Res Ther 11, R57 (2009).

423. Cronstein, B.N. & Sitkovsky, M. Adenosine and adenosine receptors in the pathogenesis and treatment of rheumatic diseases. Nat Rev Rheumatol 13, 41-51 (2017).

424. Varani, K. et al. Normalization of A2A and A3 adenosine receptor up-regulation in rheumatoid arthritis patients by treatment with anti-tumor necrosis factor alpha but not methotrexate. Arthritis Rheum 60, 2880-91 (2009).

425. Vincenzi, F. et al. A(2A) adenosine receptors are differentially modulated by pharmacological treatments in rheumatoid arthritis patients and their stimulation ameliorates adjuvant-induced arthritis in rats. PLoS One 8, e54195 (2013).

426. Lundy, S.K., Sarkar, S., Tesmer, L.A. & Fox, D.A. Cells of the synovium in rheumatoid arthritis. T lymphocytes. Arthritis Res Ther 9, 202 (2007).

427. Bluestone, J.A. & Tang, Q. How do CD4+CD25+ regulatory T cells control autoimmunity? Curr Opin Immunol 17, 638-42 (2005).

428. Huang, Z. et al. Anti-TNF-alpha therapy improves Treg and suppresses Teff in patients with rheumatoid arthritis. Cell Immunol 279, 25-9 (2012).

429. Ceeraz, S., Hall, C., Choy, E.H., Spencer, J. & Corrigall, V.M. Defective CD8+CD28+ regulatory T cell suppressor function in rheumatoid arthritis is restored by tumour necrosis factor inhibitor therapy. Clin Exp Immunol 174, 18-26 (2013).

430. Byng-Maddick, R. & Ehrenstein, M.R. The impact of biological therapy on regulatory T cells in rheumatoid arthritis. Rheumatology (Oxford) 54, 768-75 (2015).

431. McGovern, J.L. et al. Th17 cells are restrained by Treg cells via the inhibition of interleukin-6 in patients with rheumatoid arthritis responding to anti-tumor necrosis factor antibody therapy. Arthritis Rheum 64, 3129-38 (2012).

432. Atzeni, F. et al. Immunogenicity and autoimmunity during anti-TNF therapy. Autoimmun Rev. 12, 703-8. doi: 10.1016/j.autrev.2012.10.021. Epub 2012 Nov 30. (2013).

433. Wang, P.L., O'Farrell, S., Clayberger, C. & Krensky, A.M. Identification and molecular cloning of tactile. A novel human T cell activation antigen that is a member of the Ig gene superfamily. J Immunol 148, 2600-8 (1992).

434. The Genotype-Tissue Expression (GTEx) project. Nat Genet 45, 580-5 (2013).

435. Martinet, L. & Smyth, M.J. Balancing natural killer cell activation through paired receptors. Nat Rev Immunol 15, 243-54 (2015).

436. Chan, C.J. et al. The receptors CD96 and CD226 oppose each other in the regulation of natural killer cell functions. Nat Immunol 15, 431-8 (2014).

437. Haralambieva, I.H. et al. Associations between single nucleotide polymorphisms in cellular viral receptors and attachment factor-related genes and humoral immunity to rubella vaccination. PLoS One 9, e99997 (2014).

438. Ovsyannikova, I.G., Salk, H.M., Larrabee, B.R., Pankratz, V.S. & Poland, G.A. Single nucleotide polymorphisms/haplotypes associated with multiple rubella-specific immune response outcomes post-MMR immunization in healthy children. Immunogenetics 67, 547-61 (2015).

10 | ANNEX

10.1 Abbreviations

The abbreviations included in the following list have been used two or more times throughout the body of the thesis.

ACPA Anti-citrullinated protein antibodies. 56

ADAs Antidrug antibodies. 59-60, 62

APCs Antigen presenting cells. 18, 170

BC Betweenness centrality. 48

BD Behçet's disease. 166

BP Base pairs. 27, 30

CD Crohn's disease. 17, 22, 55-59, 65, 149, 167, 169-170, 172

CVD Cardiovascular disease. 23-24, 56, 65, 166-167, 171-172

dbSNP The SNP database. 31

DC Degree centrality. 48

DNA Deoxyribonucleic acid. 27-32, 34, 51, 169

eQTL Expression quantitative trait loci. 50, 54

FDA Food and Drug Administration. 57, 171

GAG Glycosaminoglycan. 163-164

GCM Gene coexpression module. 168

GDU Gastroduodenal ulcers. 166

GWAS Genome-wide association study. 33-56, 61, 161-170

GWPA Genome-wide pathway analysis. 161, 165

HLA Human leukocyte antigen. 52-56, 62-63, 162

IBD	Inflammatory bowel disease. 22, 25, 52, 58-62, 163
IMIDC	IMID Consortium. 26, 63
IMID	Immune-mediated inflammatory disease. 17-26, 34, 41, 45, 52-63, 65, 67, 97, 111, 131, 161-168, 171-172
LD	Linkage disequilibrium. 32-34, 44, 53-54
MAF	Minor allele frequency. 30-31, 33, 42, 52
MHC	Major histocompatibility complex. 18, 52
mRNA	Messenger ribonucleic acid. 28-29
PsA	Psoriatic arthritis. 17, 21-23, 52-53, 56-59, 61-62, 65, 79, 163-164, 167, 171
PTM	Post-translational modification. 162
RA	Rheumatoid arthritis. 17, 20-22, 25, 44, 52, 54, 56-62, 65, 131, 163, 167-169, 172
SLE	Systemic lupus erythematosus. 17, 21-25, 52, 55-56, 65, 97, 165, 167, 171
SNP	Single nucleotide polymorphism. 30-34, 38-45, 52-54, 163
Tregs	Regulatory T cells. 18-20, 22, 168-169, 172
Th	Helper T cells. 18-20, 22, 170
Thf	Follicular helper T cells. 18, 19
UC	Ulcerative colitis. 17, 22, 57-59, 167
1KG	1000 Genomes project. 33, 39

10.2 Summary of the additional publications authored by Adrià Aterido

The development of this thesis has led also to the elaboration of three additional research articles.

In order to investigate the genetic regulatory mechanisms of CD4+ T cells associated with RA etiology, we have performed a genome-wide analysis of eQTLs in CD4+ T cells from RA patients. Whole genome expression profiling of CD4+ T cells and genome-wide genotyping (598,258 SNPs) of 29 RA patients with an active disease were performed. We also developed a novel systems genetics approach to avoid the excessive burden of multiple testing associated with genome-wide *trans*-eQTL analysis. The genomic regulation pattern of CD4+ T cells in RA was compared to the genomic regulation observed in reference lymphoblastoid cell lines (LCLs). In the genome-wide eQTL analysis, we detected a significant *cis*-eQTL associated with the expression of the *FAM66C* gene (P=6.51×10^{-9}). Using the new systems genetics approach, we identified significant *trans*-eQTLs associated with the expression of key genes for RA pathogenesis like *BIRC5* (P=5.35×10^{-8}). Comparing the genomic regulation profiles between RA CD4+ T cells and control LCLs, we found 20 genes showing differential regulatory patterns between both cell types. The new genetic regulatory elements that are key for the activity of CD4+ T cells in RA were published in the PLoS One journal:

Aterido A, Palacio C, Marsal S, Avila G, Julià A. *Novel insights into the regulatory architecture of CD4+ T cells in rheumatoid arthritis.* PLoS One (2014)

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0100690

A pharmacogenetics study aimed at validating the genetic association between the *FCGR2A* locus and the clinical response to anti-TNF therapy in RA has been also performed. For this objective, a total of 348 RA patients treated with an anti-TNF therapy from the Spanish population were included in the study. All these patients were genotyped for the *FCGR2A* polymorphism rs1081274. The clinical response to each anti-TNF drug (i.e. infliximab, etanercept and adalimumab) was determined at week 12 and was globally and independently tested for association with genetic variation at *FCGR2A* locus. In this analysis, we detected a significant association between the *FCGR2A* locus and ADL response (P=0.022). Analyzing the subset of anti-CCP positive RA patients (78%) a significant association between the *FCGR2A* and the response to infliximab was also identified (P=0.035). This work was published in the PLoS One journal:

Avila-Pedretti G, Tornero J, Fernández-Nebro A, Blanco F, González-Alvaro I, Cañete JD, Maymó J, Alperiz M, Fernández-Gutiérrez B, Olivé A, Corominas H, Erra A, **Aterido A**, López Lasanta M, Tortosa R, Julià A, Marsal S. *Variation at FCGR2A and functionally related genes is associated with the response to anti-TNF therapy in rheumatoid arthritis.* PLoS One (2015)

http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0122088

Finally, a pharmacogenomics study aimed at characterizing the impact that functional rare variants have on anti-TNF response in CD has been conducted. For this objective, a total of 41 CD patients starting anti-TNF therapy were analyzed. Whole genome sequencing was performed using the Illumina Hiseq4000 platform. Low-frequency variants were annotated and classified according to their damaging potential. The clinical response was determined at week 14 of treatment. Screening the whole genome to identify homozygous loss-of-function (LoF) variants, we identified a total of 3,250 functional rare variants, including 2,682 damaging and 568 LoF variants. Two homozygous LoF mutations at *HLA-B* and *HLA-DRB1* genes were found to be associated with anti-TNF response. Genome-wide LoF variants were found to be enriched in specific epigenetic marks for the gastrointestinal tissue ($P<0.05$). We also tested the TNF signaling pathway for overabundance of damaging variants using the SKAT-O test. We found that the burden of damaging variation in this pathway is associated with anti-TNF response ($P=0.018$). Moreover, we found that damaging variation in the TNF signaling pathway is enriched in epigenetic marks from CD8+ ($P=6.01\times10^{-4}$) and CD4+ ($P=0.032$) T cells. At the time of thesis deposit, these findings were under review for publication in the Alimentary Pharmacology & Therapeutics journal:

Chaparro M[*], **Aterido A**[*], Guerra I, Iborra M, Cabriada JL, Bujanda L, Taxonera C, García-Sánchez V, Marín-Jiménez I, Barreiro-de Acosta M, Vera I, Martín-Arranz MD, Hernández-Breijo B, Mesonero F, Sempere L, Gomollón F, Hinojosa J, Bermejo F, Beltrán B, Rodríguez Pescador A, Banales JM, Olivares D, Aguilar-Melero P, Menchén L, Ferreiro-Iglesias R, Blázquez Gómez I, Benítez García B, Guijarro LG, Marín AC, Bernardo D, Marsal S, Julià A, Gisbert JP. *Functional rare variants influence the clinical response to anti-TNF therapy in Crohn's disease.* Alimentary Pharmacology & Therapeutics (2018)

[*] **Equally contributed as first authors.**

Under review