UAB
Universitat Autònoma de Barcelona

Bioinformatics and Systems Biology Approaches in the Study of Giant Cell Arteritis

Filippo Guerri

Bioinformatics and Systems Biology Approaches in the Study of Giant Cell Arteritis

Thesis for the degree of Doctor in Bioinformatics

Directors:
Prof. Dr. Xavier Daura
Dr. Judith Farrés Marisch

Candidate:
Filippo Guerri

Universitat Autònoma de Barcelona

2024
Programa de Doctorat en Bioinformàtica
Institut de Biotecnologia i de Biomedicina
Universitat Autònoma de Barcelona

Acknowledgement

A special thanks Goes to Xavier Daura, Judith Farrés and José Manuel Mas for the support and guidance trough my days in Anaxomics and my PhD Journey. I also want to thanks all my colleagues in Anaxomics, in particular Chiara, Pedro, JuanMa, Valentine, Guillem and Jose Luis.

This work was funded as part of the HELICAL project from the European Unions Horizon 2020 research and innovation program under the Marie Skłodowska-Curie grant agreement No. 813545. I am thankful for the funding that allowed me to conduct this research and to all the members of the HELICAL training network for being part of an exciting innovation program.

A special thanks goes to my family, all my dear friends and the nice people I have met in my years in Barcelona.

Summary

Giant Cell Arteritis (GCA) is a complex autoimmune vasculitis that affects large arteries mainly in the elderly population and can lead to severe complications. This work explores and describes the pathophysiological mechanisms of GCA, studies the side effects of prednisone and the mechanism of action of Tocilizumab using quantitative analysis and in-silico system biology method, and lastly, presents and evaluates MMPred, a Bioinformatics tool that incorporates alignment and HLA-II epitope prediction algorithm to predicted autoantigens that could be responsible to post-infection autoimmunity.

This work identified relevant associations between the use of prednisone and known and/or potential side effects in Real World Data and also investigated further those associations providing in-silico evidence of proteins that could be involved in the insurgence of these side effects. Data mining is applied to gene expression data of PBMCs (Peripheral Blood Mononuclear Cells) revealing a key role for STAT3 and MYC as predictors of response to Tocilizumab treatment.

At last, the bioinformatics tool MMPred (Molecular Mimicry Predictor) is described, tested, and applied to two relevant test cases in which a relation between viral infection and the insurgence of autoimmune diseases is known. While a statistical evaluation of the tool is limited by data availability, the analysis suggests the capability of the tool to identify biologically relevant autoantigens that could trigger the autoimmune response in the context of SARS-CoV-2 and Varicella Zoster Virus infections. In the first case, an interesting similarity between the human Helicase MOV10 and the NSP13 protein from SARS-CoV-2 is identified and described. In the second instead, the VZV protein 66 shows a high degree of similarity with many human proteins, among which KIT and SYK which are predicted to have strong functional relation with the early stage of inflammation and the dysfunction of immune checkpoints in Giant Cell Arteritis, and the proteins MAPK3 and MAPK1 which are known to related to the process of vascular occlusion in Giant Cell Arteritis.

Contents

Chapter 1

Introduction

1.1. Giant Cell Arteritis

Giant Cell Arteritis (GCA) is the most common granulomatous systemic vasculitis in people over the age of 50, reporting a peak between 70-80 years [Kuret et al., 2019, Terrades-Garcia and Cid, 2018]. This disorder appears predominantly in Caucasian individuals, showing more affectation in the female population. The risk of suffering from GCA for women is 1% whereas for men is 0.5% during their life time [Al-Mousawi et al., 2019]. An annual incidence range of GCA has been estimated between 1.1 to 32.8 cases per 100 000 individuals of both sex with 50 years or more [Terrades-Garcia and Cid, 2018, Al-Mousawi et al., 2019].

GCA is a systemic autoimmune disorder that typically affects medium and large vessels, with a special tropism for the aorta branches, carotid and vertebral arteries. This disorder can lead to occlusive granulomatous vasculitis with transmural infiltrate containing multinucleated giant cells, where the temporal artery is commonly involved. This scenario can lead to severe complications such as blindness or cerebrovascular events [Al-Mousawi et al., 2019].

Giant Cell Arteritis is characterized by chronic inflammation of large arteries, such as the aorta and the carotid arteries and its extra-cranial branches, which may lead to severe clinical sequelae if not treated promptly, including vision loss, scalp and tongue necrosis, aortic dissection/rupture, and cerebral infarction [Unizony and Kermani, 2018, Dinkin and Johnson, 2021]. Giant Cell Arteritis is best understood as an inflammatory vascular syndrome with features of cranial and/or large-vessel vasculitis, systemic inflammation, and polymyalgia rheumatica (PMR), which frequently overlap [Dejaco et al., 2017]. Histopathology and immunopathology studies reveal inflammation of the artery wall with predominance of CD4+ T lymphocytes and macrophages, which frequently undergo granulomatous organization with formation of giant cells [Cid et al., 1989]. Inflammation-induced vascular remodeling leads to intimal hyperplasia and lumen occlusion, the source of the ischemic complications of the

disease [Dejaco et al., 2017, Cid et al., 1989].

1.1.1. Symptoms

Symptoms include fever, fatigue, headache, visual impairment, pain in the jaw and tongue, and aggravation of pain by cold temperatures. Furthermore, symptoms include general signs of vascular occlusion, such as headache, mandibular pain, and stroke [Ness and Nölle, 2024].

1.1.2. Clinical Diagnosis

The presence of giant cells in a temporal artery biopsy specimen is strongly suggestive of GCA and giant cells are probably the most carefully sought feature when GCA is suspected. However, the frequency with which giant cells are identified is substantially lower than the name of the disease suggests [Parker et al., 1975, Lie, 1996, Zhou et al., 2009]. In some cases, multiple cuts of an artery are necessary to identify giant cells, and giant cells are not detected in around 2% of clinically diagnosed cases of GCA [Regan et al., 2002]. In the absence of giant cells, more subtle histopathological changes must suffice as clues to possible presence of GCA [Stacy et al., 2011, Banz and Stone, 2018].

People with symptoms are diagnosed by biopsy of cranial artery. In this case the markers of GCA are the destruction of the external muscular layer and the narrowing of the internal part due to hyperplasia/fibroplasia [Farina et al., 2023].

Other markers of GCA include Erythrocyte sedimentation rate $\geq 50 mm/hour$, anemia and inflammation markers [Maz et al., 2021]. Immune signals are more dramatic compared to other inflammatory diseases, with high levels of IL-6 detected in all patients [Cid, 2014], making IL-6 this the main molecular marker of the disease.

1.1.3. Etiology

The causes of GCA remain unknown. While antigen has been suspected to function as a trigger of intramural inflammation, more recent data point towards defects in the threshold setting of the immune system, leading to unopposed T-cells activation in the vessel wall [Maleszewski et al., 2017]. Vascular inflammation is tightly linked to dysfunctional immune checkpoints. Therefore, it is not surprising that inhibition of immune checkpoints in the context of cancer immunotherapy may lead to vasculitis (Immune Checkpoint Inhibitor-induced vasculitis) [Melissaropoulos et al., 2020].

1.2. Pharmacological treatment of Giant Cell Arteritis

1.2.1. Prednisone and its side effects

Prednisone is a synthetic glucocorticoid that plays a crucial role in the management of various inflammatory and autoimmune conditions because of its anti-inflammatory and immunosup-

pressive properties. As a corticosteroid, it mimics the effects of cortisol, a hormone produced by the adrenal glands, which is vital for regulating metabolism, immune response, and stress reactions [Puckett et al., 2023].

Upon administration, prednisone is metabolized in the liver to its active form, prednisolone, which exerts its effects by binding to glucocorticoid receptors within cells. This binding leads to alterations in gene expression that ultimately reduce inflammation and modulate immune responses [Puckett et al., 2023].

The mechanism of action involves several pathways: prednisone decreases the migration of leukocytes to sites of inflammation, reverses increased capillary permeability, and inhibits the production of pro-inflammatory cytokines. These actions contribute to the alleviation of symptoms associated with various conditions such as asthma, rheumatoid arthritis, lupus, inflammatory bowel disease, and certain cancers [Puckett et al., 2023]. The rapid onset of action—often within hours—makes prednisone a preferred choice for acute exacerbations of chronic conditions.

After its discovery in 1950 by Arthur Nobile [Nobile, 1994], prednisone is still widely used nowadays to treat allergic, dermatologic, gastrointestinal, hematologic, ophthalmologic, nervous, renal, respiratory, rheumatologic, infectious, endocrine, or neoplastic conditions as well as in organ transplant [U.S. Food and Drug Administration, 2013]. It is in fact the first treatment used when Giant Cell Arteritis is diagnosed [Fraser et al., 2008].

However, prednisone users have a high risk of showing various kinds of side effects. The side effects can range from mild to severe and include weight gain, hypertension, osteoporosis, increased susceptibility to infections, mood changes, alterations of blood levels of lipids and glucose to gastrointestinal, cardiovascular, cardiac, dermatological, ophthalmological side effects. Long-term use can lead to adrenal suppression due to feedback inhibition on the hypothalamic-pituitary-adrenal (HPA) axis [Oray et al., 2016].

1.2.2. Tocilizumab, a new frontier in the treatment of GCA

Despite advances in understanding its pathogenesis and the advent of glucocorticoids as a cornerstone of treatment, managing GCA remains a clinical conundrum marked by relapses, glucocorticoid-related adverse events, and refractory disease courses [Maleszewski et al., 2017, Opriș-Belinski et al., 2024, Ness and Nölle, 2024].

In recent years, the emergence of targeted immunomodulatory agents such as Tocilizumab represents a new era in the management of GCA (DrugBankID: DB06273) [Opriș-Belinski et al., 2024, Régent and Mouthon, 2022, Nepal et al., 2023].

Tocilizumab, a humanized monoclonal antibody, has emerged as a pivotal therapeutic agent

in the management of various autoimmune and inflammatory diseases. Tocilizumab was first produced in the late 1990s when researchers at Osaka University identified IL-6 as a critical mediator of inflammation. Following its discovery, Chugai Pharmaceuticals initiated clinical trials that culminated in the drug's approval by the FDA in January 2010 for RA treatment [Preuss CV, 2024]. Since then, ongoing research has expanded its indications significantly.

Initially developed for rheumatoid arthritis (RA), its applications have expanded significantly, particularly in response to the global health challenges posed by conditions such as cytokine release syndrome (CRS) and COVID-19 [Schett et al., 2020, Kishimoto, 2021]. In the last years, has demonstrated remarkable efficacy in refractory GCA, offering a corticosteroid-sparing alternative and achieving sustained disease remission [Calderón-Goercke et al., 2019]. Its ability to down-regulate the IL-6-driven inflammatory cascade not only mitigates acute symptoms but also attenuates vascular inflammation and reduces the risk of relapse, making it a promising treatment in GCA [Regola et al., 2020].

Tocilizumab functions as an interleukin-6 (IL-6) receptor antagonist. IL-6 is a pro-inflammatory cytokine that plays a crucial role in the immune response and is implicated in the pathogenesis of several autoimmune diseases. By binding to both soluble and membrane-bound IL-6 receptors, tocilizumab effectively inhibits IL-6 from exerting its inflammatory effects on target cells, including hepatocytes and synovial fibroblasts. This blockade results in decreased production of acute-phase reactants such as C-reactive protein (CRP) and fibrinogen, which are markers of inflammation and consequently help mitigate the inflammatory processes that characterize GCA [Sebba, 2008].

1.3. Molecular Mimicry and Autoimmune Diseases

Epidemiological, clinical, and experimental evidence support the association between infections and autoimmune diseases (AIDs) for a variety of conditions, e.g. Rheumatoid Arthritis, Lupus Erythematosus Systemic and Diabetes type I [Seyyed Mousavi et al., 2017, Knight et al., 2021, Ehrenfeld et al., 2020, Bergamin and Dib, 2015, Starshinova et al., 2022, Yeung et al., 2011, Gómez-Rial et al., 2020].

Based on this relation the concept of "Molecular Mimicry" has been proposed as one of the many factors that could induce autoimmunity [Damian, 1964, Kaplan and Meyeserian, 1962]. Molecular mimicry occurs when the similarity between host and pathogen could trigger an immune response against self-antigen, due to the incapability of the immune system to discriminate between self and non-self [Peterson and Fujinami, 2007, Tam et al., 2007].

1.3.1. GCA and the Relation with Varicella Zoster Virus

While the exact etiology of GCA remains uncertain, recent studies have investigated the potential role of Varicella Zoster Virus (VZV) as a trigger for the disease.

VZV is unique among human viruses for its ability to replicate in arteries and cause vascular disease. It is well-established that VZV can lead to vasculopathy in intracerebral arteries, characterized by granulomatous arteritis, a pathology also seen in GCA [Gilden and Nagel, 2016]. This similarity has prompted researchers to investigate whether VZV infection could also be linked to GCA.

Immunohistochemical analyses have consistently detected VZV antigens in a significant proportion of GCA-positive temporal arteries (TAs). For instance, VZV antigen was found in 70% of GCA-positive TAs and 58% of TAs negative for GCA pathology but with clinical features of GCA, compared to only 18% in normal TAs [Amlie-Lefond and Gilden, 2016]. The presence of VZV antigen primarily in the adventitia of affected arteries suggests that inflammation might follow VZV reactivation from ganglia, supporting the hypothesis that VZV could be a contributing factor to GCA.

Several epidemiological studies have examined the association between VZV events and the incidence of GCA. One study reported that complicated and uncomplicated VZV was associated with an increased risk of developing GCA, with hazard ratios of approximately 2.0 and 1.4, respectively, in large cohorts [England et al., 2017]. Another study found that prior infections, including HZ, were modestly associated with incident GCA, indicating that long-standing immune alterations might increase susceptibility to GCA [Rhee et al., 2017].

While there is considerable evidence suggesting that VZV may play a role in the pathogenesis of GCA, this relationship is not unequivocally established. The detection of VZV antigens in GCA-affected tissues supports the hypothesis that VZV reactivation could trigger the inflammatory cascade seen in GCA. However, contradictory findings and methodological challenges highlight the need for further research to definitively clarify the role of VZV in GCA.

1.3.2. The Varicella Zoster Virus

Varicella Zoster Virus (VZV) is a double-stranded DNA virus from the Herpesviridae family, it presents its symptoms in the form of chickenpox (varicella) and shingles (herpes zoster) when reactivated. After initial infection, which occurs typically in childhood, VZV establishes latency in the sensory ganglia. The reactivation, often occurring in older adults or those with weakened immune systems, leads to herpes zoster, a painful rash that can result in complications such as postherpetic neuralgia [Khalil et al., 2015].

VZV entry into host cells is facilitated by multiple glycoproteins on its envelope, including gB, gE, and gH, which mediate attachment and membrane fusion with host cells. Glycoprotein E (gE) is particularly significant, as it interacts with various cellular receptors to enable viral spread and persistence. Upon entry, VZV navigates to the nucleus, where its genome replicates and viral proteins are synthesized [Olson et al., 1997]. Once inside the cell, VZV

utilizes cell-to-cell spread mechanisms, which allow it to bypass extracellular space and evade certain immune defenses. Studies have shown that gE pairs with gI to form a complex that promotes efficient cell-to-cell spread, especially within neurons and epithelial cells [Oliver et al., 2020].

VZV has evolved mechanisms to avoid detection and elimination by the immune system, contributing to its ability to establish lifelong latency and episodic reactivation. A primary strategy involves glycoprotein E (gE) binding to the Fc region of host IgG, thereby inhibiting antibody-mediated neutralization [Abendroth et al., 2010]. Additionally, VZV downregulates major histocompatibility complex class I (MHC-I) molecules on infected cells, reducing recognition by cytotoxic CD8+ T cells. This allows VZV to persist without significant immune clearance. Furthermore, the virus modulates local immune responses through the upregulation of anti-inflammatory cytokines such as interleukin-10 (IL-10), which dampens local inflammation and immune activity [Abendroth and Arvin, 2001].

After primary infection, VZV establishes latency primarily in the cranial, dorsal root, and autonomic ganglia. During this period, the virus remains in a quiescent state with limited transcriptional activity, which helps it evade immune surveillance. Key latency-associated transcripts (LATs) are expressed at low levels, maintaining the viral genome in an episomal form without producing infectious virions [Zerboni et al., 2014]. Reactivation, often triggered by stress, aging, or immunosuppression, leads to viral replication and the spread of VZV along sensory nerves, resulting in the clinical manifestations of herpes zoster.

There is growing evidence to suggest that VZV may play a role in the pathogenesis of certain autoimmune diseases, possibly through molecular mimicry. Molecular mimicry occurs when viral antigens resemble host proteins closely enough to trigger an autoimmune response. For example, the structural similarity between VZV antigens and self-antigens may lead to cross-reactive T cells that attack host tissues. This mechanism has been implicated in VZV-associated vasculitis and may explain the virus's association with Giant Cell Arteritis and other inflammatory diseases of the blood vessels [Maguire et al., 2024, Ishihara et al., 2024, Abendroth and Arvin, 2001].

1.4. Predicting Molecular Mimicry

1.4.1. The variability of the Human Leukocyte Antigen (HLA) locus and its relation with autoimmunity

The Human Leukocyte Antigen (HLA) locus is part of the major histocompatibility complex (MHC). It spans approximately 4 megabases on chromosome 6 and includes several classes of genes, notably HLA class I (e.g., HLA-A, HLA-B, HLA-C) and class II (e.g., HLA-DRB1, HLA-DQB1, HLA-DQA1). It is known for its extensive genetic variability, which plays a critical role in immune response and susceptibility to autoimmune diseases [Carapito et al.,

2016, Dholakia et al., 2022].

These genes a show many different alleles within the population [Pierini and Lenz, 2018], and while their polymorphism allows for a wide range of antigen presentation capabilities, it also complicates the genetic landscape associated with autoimmune diseases [Dholakia et al., 2022].

Numerous studies have established strong associations between specific HLA alleles and various diseases. According to Dholakia et al. around 144,000 studies pairing 28,320 HLA alleles with different diseases, among which autoimmune diseases. For example Diabetes type I is linked to the variants DRB1*04:01 and DQA1*03:01 [Berryman et al., 2023, Krishna et al., 2024], Rheumatoid Arthritis is instead related to alleles of HLA-DRB1 [Matzaraki et al., 2017, Debebe et al., 2020]. Interestingly Giant Cell Arteritis is known to be related to HLA-II allele DRB1*04:01 [Prieto-Peña et al., 2021].

1.4.2. The mechanisms of action relating HLA to autoimmune diseases

The mechanisms by which HLA variability contributes to autoimmune diseases are different and can be summarized as follows:

- 1. **Altered Antigen Presentation**

  HLA molecules are responsible for presenting peptides (including self-peptides) to T cells. Variability in HLA genes can lead to differences in how these peptides are presented, which may trigger autoimmune responses. For example, as been demonstrated that HLA alleles associated with rheumatoid arthritis (RA) expose citrullinated peptides more effectively, potentially activating autoreactive T cells and initiating or worsening the disease [Miyadera and Tokunaga, 2015].

- 2. **Genetic Polymorphism and Linkage Disequilibrium**

  The high degree of polymorphism within the HLA locus results in a diverse array of haplotypes that can influence susceptibility to various autoimmune conditions. This genetic variability complicates the identification of specific causal variants due to linkage disequilibrium, where alleles at different loci are inherited together more often than expected by chance [Berryman et al., 2023, Krishna et al., 2024]. The combination of these factors contributes to the complex patterns of disease associations observed in autoimmune disorders.

- **3. Influence of Gut Microbiome**

  Recent research suggests that HLA variability may interact with gut microbiota, in-

fluencing immune responses and contributing to autoimmunity. An imbalance in gut microbiome has been linked to increased expression of HLA class II molecules, which can exacerbate systemic inflammation and further alter the gut microbiome composition [Berryman et al., 2023]. This relationship between HLA variability, gut microbiome, and autoimmune disease is still poorly explored but could unveil novel mechanism by which the autoimmune diseases arises.

- 4. **Systemic Inflammation and Immune Activation**

  Certain HLA haplotypes have been associated with systemic inflammation, which may precede the onset of autoimmune diseases. This systemic inflammation is thought to create an environment favoring the development of autoimmunity [Berryman et al., 2023].

- **5. Molecular Mechanisms and Pathophysiology**

  The molecular mechanisms underlying HLA-associated autoimmunity include specific amino acid variations within the peptide-binding grooves of HLA molecules. These variations can affect how well self-peptides are presented, potentially leading to inappropriate immune responses against the body's own tissues [Pavlos et al., 2017].

1.4.3. Epitopes

Epitopes are critical triggers of the immune response, as they are the specific parts of the antigen that bind to receptors of the Immune system. The epitope interacts both with HLA-I and HLA-II, B-cells, and T-cells [Wieczorek et al., 2017, Collesano et al., 2024, Sadegh-Nasseri, 2016].

HLA Class I molecules primarily present epitopes derived from intracellular proteins, typically 8 to 14 amino acids in length. These epitopes are recognized by CD8+ cytotoxic T cells. The binding of peptides to HLA Class I is highly selective, with specific anchor residues at defined positions (such as the second and last residues) being crucial for stable binding. This interaction is essential for the immune system to monitor and eliminate infected or malignant cells [Wieczorek et al., 2017].

In contrast, HLA Class II molecules present longer peptides (typically 12 to 25 amino acids) derived from extracellular proteins, which are processed by antigen-presenting cells (APCs). These epitopes are recognized by CD4+ helper T cells. The binding mechanism for HLA Class II is more flexible, allowing for multiple anchor points along the peptide chain, which facilitates the presentation of a broader range of epitopes [Wieczorek et al., 2017, Collesano et al., 2024].

The interaction between T cell receptors (TCRs) and peptide-HLA (pHLA) complexes is fundamental for T cell activation. Epitopes must be presented in a stable form by HLA molecules for effective recognition by T cells, which subsequently leads to an adaptive immune response [Sadegh-Nasseri, 2016].

1.4.4. Epitope Prediction Algorithms

Epitope prediction algorithms are essential tools in the fields of immuno-informatics, enabling researchers to identify potential epitopes that can elicit immune responses. These algorithms couple machine learning and bioinformatics techniques to analyze protein sequences and/or structures in order to predict which segments of a pathogen's or human protein could trigger an immune response [Yan et al., 2024].

Epitope prediction algorithms can be broadly categorized into two main types: sequence-based methods and structure-based methods.

- **Sequence-based methods**: These algorithms utilize the amino acid sequences of proteins to predict epitopes. Among which ***(i)*** *Motif Search-Based Approaches*: these methods identify specific patterns or motifs within protein sequences that are associated with epitope recognition, ***(ii)*** *Machine Learning Approaches*: apply mainly Artificial Neural Networks (ANN) trained on known epitope datasets to learn predictive features [Vita et al., 2019]. ***(iii)*** *Statistical Methods*: Techniques like propensity scales assess the likelihood of amino acids forming epitopes based on their physical and chemical properties, such as hydrophilicity and flexibility [Yurina and Adianingsih, 2022].

- **Structure-based methods**: These approaches focus on the three-dimensional structures of proteins to predict epitopes. ***(i)*** *Molecular Docking*: This technique simulates how peptides bind to Human Leukocyte Antigen (HLA) molecules, providing insights into potential T-cell epitopes [Ramana and Mehla, 2020]. ***(ii)*** *Conformational Epitope Prediction*: Algorithms that utilize spatial data from protein structures to predict discontinuous epitopes, which are formed by residues that are not sequentially adjacent but come together in the folded protein [Yurina and Adianingsih, 2022, Høie et al., 2024].

Overall, it must be highlighted that the integration of neural network-based machine learning techniques has enhanced significantly the predictive capabilities of these algorithms [Nilsson et al., 2023, Junet and Daura, 2021, Collatz et al., 2021, Ivanisenko et al., 2024, Yang et al., 2024].

Despite significant improvement in epitope prediction algorithms, limitations are still present. The accuracy of predictions is influenced by the quality of the datasets, their representation of the epitope universe as a whole, and the variability of the immune responses among indi-

viduals in relation to their genetic background [Barra et al., 2024, Yan et al., 2024].

1.4.5. Sequence Alignment & Sequence Database Search Algorithm

Sequence alignment is a fundamental technique in bioinformatics for analyzing the similarities and differences between protein sequences. Aligning sequences is useful for inferring functional and evolutionary relationships, identifying conserved regions, and predicting the structure and function of proteins [Barton, 1998]. Furthermore, sequence alignment is used within database search algorithms to identify similarities between distinct groups of proteins. In our case, sequence database search algorithms are used to compare human and microbial proteomes to identify potential similar peptides under the hypothesis of molecular mimicry [Altschul et al., 1990, Schäffer et al., 2001, Damian, 1964, Kaplan and Meyeserian, 1962].

Sequence alignment refers to the arrangement of two or more biological sequences (protein or nucleotide) to identify regions of similarity. These similarities may indicate functional, structural, or evolutionary relationships between the sequences. Alignments can be classified into two main types: *(i) global alignment*, which aligns entire sequences, and *(ii) local alignment*, which focuses on aligning the most similar sub-regions of sequences [Barton, 1998].

Pairwise alignment involves comparing two sequences to find the best-matching segments. This can be achieved through two primary algorithms:

- **Global Alignment**: The Needleman-Wunsch algorithm is commonly used for global alignments. It aligns every residue in both sequences from start to finish, maximizing overall similarity [Needleman and Wunsch, 1970]. This method is ideal for sequences of similar length.

- **Local Alignment**: The Smith-Waterman algorithm is utilized for local alignments, focusing on identifying the most similar segments within the sequences. This approach is particularly useful when comparing sequences that may differ significantly in length or composition [Smith and Waterman, 1981].

Multiple sequence alignment extends pairwise alignment techniques to align three or more sequences simultaneously. This method helps identify conserved motifs across a family of proteins and can be performed using algorithms such as Clustal [Sievers and Higgins, 2021], MUSCLE [Edgar, 2004], or FAMSA [Deorowicz et al., 2016]. MSAs are essential for phylogenetic analysis, functional prediction, and clustering of protein families.

Sequence alignment provides the basis for identifying similarities between sequences, but large-scale comparisons require similarity search algorithms. Tools like BLASTP and PSI-BLAST apply alignment methods to efficiently compare query sequences against protein databases, focusing on local regions of similarity. In this work, we use these algorithms to

compare human and microbial proteomes in order to identify similar peptides.

- **BLASTP (Basic Local Alignment Search Tool for proteins):** performs a straightforward comparison of a protein query against a protein database. It identifies regions of local similarity based on predefined scoring matrices, such as BLOSUM62 [Henikoff and Henikoff, 1992]. The algorithm operates by first identifying short matches (words) between the query and database sequences, then extending these matches to find longer alignments that meet a specified significance threshold [Altschul et al., 1990].

- **PSI-BLAST (Position-Specific Iterated BLAST):** builds upon the initial BLASTP search by incorporating iterative refinement. The first iteration of PSI-BLAST is identical to a standard BLASTP search; however, subsequent iterations utilize a position-specific scoring matrix (PSSM) derived from the sequences identified in the previous round. This PSSM allows PSI-BLAST to capture more subtle patterns of similarity that may not be detected in a single BLASTP run [Schäffer et al., 2001].

One of the primary advantages of PSI-BLAST over BLASTP is its increased sensitivity for detecting distant homologs. By iteratively refining the scoring matrix based on the most significant hits from previous searches, PSI-BLAST can uncover relationships between sequences that may be too weak to be identified by a single BLASTP search. This iterative process enables PSI-BLAST to find biologically relevant similarities that would otherwise remain hidden. For example, studies have shown that PSI-BLAST can identify more distant evolutionary relationships due to its ability to leverage information from multiple aligned sequences, making it particularly useful in cases where sequence conservation is low but functional similarities exist.

In BLASTP, the scoring is based on static substitution matrices that do not adapt throughout the search process. This can limit its ability to accurately score alignments involving divergent sequences. In contrast, PSI-BLAST dynamically generates a PSSM after each iteration, which reflects the specific amino acid composition and conservation patterns observed in the aligned sequences from previous iterations. This adaptability enhances the accuracy of scoring and alignment for subsequent searches. Moreover, PSI-BLAST allows users to set different expectation values for each iteration, enabling finer control over the sensitivity of the search process. This flexibility can be crucial when exploring databases with varying levels of sequence divergence.

While PSI-BLAST offers enhanced sensitivity, it may require more computational resources due to its iterative nature. Each iteration involves recalculating alignments and updating the PSSM, which can increase processing time compared to a single BLASTP search. However, this trade-off is often justified by the improved detection of relevant homologs in complex datasets.

1.4.6. Molecular Mimicry Predictor, MMPred

Epidemiological, clinical, and experimental evidence supports the association between infections and autoimmune diseases, with molecular mimicry proposed as one of the key mechanisms underlying this relationship [Oldstone, 1998, Rojas et al., 2018]. Identifying or predicting epitope mimicry events could therefore serve as a valuable clinical tool. From a theoretical standpoint, predicting peptide mimicry involves identifying similarities between self and exogenous proteins [Doxey and McConkey, 2013] and predicting their recognition by the immune system [Rojas et al., 2018]. In this context, bioinformatics can provide valuable insights through the application of sequence alignment and epitope-prediction algorithms, particularly for Human Leukocyte Antigen (HLA) epitopes [Caron et al., 2015, Wang et al., 2010]. To our knowledge, CRESSP [An et al., 2022] is the only tool currently integrating and streamlining both sequence alignment and epitope prediction. However, CRESSP primarily focuses on B-cell epitopes. Although the SARS-CoV-2 study by An et al.included HLA class II epitopes as well, the actual tool (`https://pypi.org/project/cressp/`) only incorporates HLA class I epitope predictions, as provided by MHCflurry [O'Donnell et al., 2020].

Here, we present the program MMPred (Molecular Mimicry Predictor), a tool that integrates sequence alignment and HLA class II epitope-prediction algorithms into a single pipeline. The tool is designed to be flexible, user-friendly, and amenable to non-expert use, requiring as sole inputs the fasta files for the exogenous (query) and endogenous (target) protein (or peptide) sets and a list of HLA alleles of interest to perform the predictions for. The application of sequence alignment is optional, and if used, the two sets of sequences are compared and epitope prediction is applied to those target sequences that show a significant alignment with query sequences. The tool is an extension of CNN-PepPred [Junet and Daura, 2021], and offers the possibility to include predictions from NetMHCIIpan4.1, both with the BA (trained on binding affinities) and EL (trained on eluted ligands) models [Reynisson et al., 2020]. The tool is programmed in such a way that additional predictors, including HLA class I epitope predictors, can be added with minimal effort. The alignments can be performed with BLASTP [Altschul et al., 1990] or by means of a position-specific scoring matrix (PSSM) with PSI-BLAST [Schäffer et al., 2001].

Chapter 2

Objectives

- **1. Provide a molecular description of the pathophysiology of Giant Cell Arteritis through a systematic review of published literature, to utilize this information within a systems biology framework.**
  Giant Cell Arteritis (GCA) is an autoimmune disease marked by inflammation in large arteries, predominantly affecting the elderly population. Despite significant advancements in the understanding of the disease, a detailed molecular description of the disease is not available. Through a systematic review of recent scientific literature, this work aims to gather data on causative and symptomatic motifs related to GCA. The knowledge gathered in this way will be applied within systems biology analyses in other parts of this thesis.

- **2. Evaluate the adverse effects associated with prednisone use by analyzing real-world data from the NHANES database, and propose molecular hypothesis to further understand the underlying mechanisms**.
  Prednisone, a synthetic glucocorticoid, is widely used in managing GCA due to its anti-inflammatory and immunosuppressive properties. However, its prolonged use is associated with a range of side effects, which can have severe consequences, especially for older adults. Given the absence of GCA-specific data in the NHANES database, this objective employs a proxy-based approach, analyzing prednisone's impact on populations with similar inflammatory conditions (e.g., asthma, rheumatoid arthritis, and chronic obstructive pulmonary disease). The study aims to quantify the association between prednisone and the insurgence of side effects within these populations using odds ratio analysis and logistic regression, controlling for confounding variables. When significant associations are identified, system biology analysis is applied to investigate the underlying molecular mechanisms.

- **3. Examine the effect of Tocilizumab mining gene expression data produced**

from ex-vivo models of PBMCs (Peripheral Blood Mononuclear Cells) from GCA patients, and use system biology method to explore its mechanism of action.

Tocilizumab, an IL-6 receptor antagonist, has emerged as a promising treatment option for GCA, offering a potential glucocorticoid-sparing alternative. Supervised and unsupervised analyses are used to identify molecular markers of response and to stratify the population under study. Furthermore, the GCA molecular description is used within system biology analysis to further understand the mechanism of action of Tocilizumab.

- **4. Develop and validate MMPred, a bioinformatics tool designed to predict autoantigens based on molecular mimicry, applying it to study the association between infectious agents and autoimmune conditions.**

  Molecular mimicry, where structural similarities between microbial and host proteins trigger an autoimmune response, is a hypothesized mechanism for GCA onset, potentially linked to infections like Varicella Zoster Virus (VZV). MMPred is a novel tool that explores molecular mimicry by integrating epitope prediction and sequence alignment algorithms. In this part of the work, MMpred is presented, tested, and applied to SARS-CoV-2-related autoimmune diseases, and to explore the relation between Varicella Zoster Virus infection and the insurgence of GCA.

Chapter 3

Methods

3.1. System Biology Analysis

In subsection 3.3.4 and subsubsection 3.5.4 of this work a system biology analysis is performed to quantify the functional relation between a certain protein and a pathophysiological motifs.

This approach, detailed in [Segú-Vergés et al., 2022], evaluates the likelihood of a functional relationship between a given protein and a set of proteins based on their connections within a model of the human protein network. The underlying algorithm employs a supervised machine-learning technique, specifically Artificial Neural Networks (ANNs), trained on a large dataset comprising pharmacological drug targets and molecular descriptors of clinical phenotypes —such as drug indications and adverse effects– from the Biological Effector Database (BED, [Jorba et al., 2020]) compiled by the authors.

In BED, each condition (e.g., giant-cell arteritis) is characterized by a list of motifs (e.g., dysfunction of immune checkpoints), and each motif is linked to a set of proteins that map to a specific subgraph within the network. The training objective of the ANNs is to correctly associate drug targets with their respective clinical conditions. The resulting score, S, represents the probability that a given relationship is a true positive, expressed as a percentage, to which is associated a $p - value$. Here, a "relationship" refers to a scenario where a perturbation of one element (e.g., a drug target) leads to an observable perturbation in another (e.g., a clinical condition).

Although the algorithm was originally trained on drug targets, it is designed to assign probabilities to the association of any node within a protein network with a specific subgraph. Since the propagation of perturbations or signals between proteins operates on the same principles regardless of whether the context involves drug targets or other protein types, the algorithm can calculate the probability of a relationship between any protein in the network, and any

clinical conditions annotated in the network, such as autoimmune diseases.

Two protein networks, or topologies, are available for the ANNs. In one of them, the functional associations are based on experimental evidence; in the other, they are based on inference and computed using a variety of resources (protein-protein interactions, gene expression, etc.). In these experiments, the analysis is run for both topologies and the largest score is taken, or a lower p-value is taken based on the purpose of the analysis. This analysis is performed both in subsection 3.3.4 and in subsubsection 3.5.4. The score S for a given protein-condition/motif pair may be zero when the protein is not present in the network or no possible association is found for the pair. In such a case, the pair is removed from the analysis.

3.2. GCA molecular description

GCA was molecularly characterized through manual curation of the current scientific literature regarding Giant Cell Arteritis pathophysiology. The information found has been used to perform system biology analysis. The characterization consists of the following protocol.

First, a search for reviews on the last 10 years on the molecular pathogenesis and pathophysiology of the condition was performed in the PubMed database (https://pubmed.ncbi.nlm.nih.gov/) on date October 7th, 2020. The specific search was as follows:

• (" Giant cell arteritis " [Title] OR "GCA" [Title]) AND ("molecular"[Title/Abstract] OR "pathogenesis"[Title/Abstract] OR "pathophysiology"[Title/Abstract])

• (" Giant cell arteritis " [Title] OR "GCA" [Title]) AND ("molecular"[Title/Abstract] OR "genetics"[Title/Abstract] OR "aneurism"[Title/Abstract] OR "inflamation"[Title/Abstract] OR "remodelling"[Title/Abstract] OR "angiogenesis"[Title/Abstract] OR "occlusion"[Title/Abstract])

The abstracts of the publications identified in this search were retrieved and assessed at the title and abstract level, and if molecular information describing the condition pathophysiology was found, the full texts were thoroughly reviewed seeking to identify the main pathophysiological processes known to be involved in this condition. These processes are referred to as 'motifs' and can be classified into two levels, depending on their involvement in the pathology:

- Causative (C): Motifs are directly related to the onset or pathophysiology of the condition characterized.

- Manifestative (M): Motifs that are a consequence of the pathology.

Subsequently, each pathophysiological process has been further characterized at protein level. The publications retrieved were reviewed in order to identify protein/gene candidates to

be effectors of the condition, i.e., proteins whose activity (or lack thereof) is functionally associated with the development of the condition. If the evidence of the implication of a candidate in the condition was judged not consistent enough to be assigned as an effector, an additional PubMed search was performed specifically for the candidate, including all the protein names according to UniProtKB; e.g.:

("Giant Cell Arteritis"[TITLE] OR " Granulomatous Vasculitis"[TITLE]) AND ("Interleukin-1 beta" [TITLE/ABSTRACT]) OR "IL1B" [TITLE/ABSTRACT] OR "IL-1 beta" [TITLE/ABSTRACT] OR "IL1F2" [TITLE/ABSTRACT] OR "Catabolin" [TITLE/ABSTRACT])

If novel candidates were identified in this phase, they were included as effectors following the same criteria and protocol.

3.3. The side-effects of Prednisone

3.3.1. NHANES Database Collection

Each individual in the continuous NHANES (1999-2018) [National Center for Health Statistics (NCHS), 2024] has been mapped to a new system of variables describing conditions/symptoms, drug prescription, demographic and anthropomorphic information (ethnicity, sex, age, body mass index (BMI)). Although there are specific variables in NHANES that map directly to these concepts, there are other variables that can also map to them, e.g. total cholesterol [mg/dl] > 240 to hypercholesterolemia. The 67 different NHANES variables used for the study are shown in Table 7.3. Different entries have been mapped to standardized terms, Drug Names for drugs, and BED (Biological Effectors Database) terms for clinical conditions. BED, is a hand-curated database from ANAXOMICS Biotech that relates clinical conditions to their molecular effectors [Jorba et al., 2020]. The terms for clinical conditions in BED are based on MESH vocabulary (Medical Subject Headings). BED version used for this study contained 302 clinical conditions, the full list of NHANES variables used is reported in Table 7.3, and the query for the translation from NHANES to BED is reported in Table 7.4.

The resulting extract from 1999 to 2018 contains 101316 individuals described by demographic and anthropometric information (AGE, SEX, BMI, and ethnicity), medical conditions, and drugs used. Apart from demographic information, the rest of the data is treated as binary (0/1). Due to the high presence of missing data on medical conditions variable in the first years of NHANES, we used as the basis for this study only the individuals from 2013 to 2018, resulting in 29400 individuals, 101 different medical conditions, and 1517 different drugs.

We have further labeled the clinical conditions in terms of the relation they have with prednisone treatment if any with four annotations.

- **IND:** when the term is a condition or symptom for which the drug could be indicated

- **SE:** when the term is a known side-effect of the drug
- **UNK:** when there is neither an indication nor a side-effect
- **BOTH:** when the term can be both a side effect and an indication.

The references for the annotation of the indication are extracted from DrugBank [Online, 2024] prednisone code DB00635 while for the side effect are extracted from SIDER 4.1 [Kuhn et al., 2016], the annotations are shown in Table 7.5.

3.3.2. Statistical Analysis

To evaluate the association between prednisone-use and medical conditions, we used the odds ratio (OR) [Szumilas, 2010].

To evaluate the association between a certain condition Y and the use of prednisone, the odds ratio OR is computed by means of multivariate logistic regression. Given a population, it is first split in two: Y^+ represents the population showing the condition Y, while Y^- represents the population not showing it. The analysis includes confounding variables in the model (age, sex, BMI, ethnicity) along with prednisone use.

Considering the logistic model:

$$p(Y^+) = \frac{\exp(\beta_0 + \beta_p x_p + \beta_1 x_1 + \cdots + \beta_m x_m)}{1 - \exp(\beta_0 + \beta_p x_p + \beta_1 x_1 + \cdots + \beta_m x_m)} \tag{3.1}$$

where $p(Y^+)$ is the probability of condition Y appearing, β_0 is the intercept, x_p is the variable for prednisone {0,1}, β_p is its regression coefficient, and $x_1, \ldots, x_m$ are confounding variables with their regression coefficients $\beta_1, \ldots, \beta_m$.

Define p_0 as the probability of having condition Y in the presence of prednisone, and p_1 the probability in its absence:

$$p_0 = p(Y^+|x_p = 0) = \frac{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)}{1 - \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)} \tag{3.2}$$

$$p_1 = p(Y^+|x_p = 1) = \frac{\exp(\beta_0 + \beta_p + \beta_1 x_1 + \cdots + \beta_m x_m)}{1 - \exp(\beta_0 + \beta_p + \beta_1 x_1 + \cdots + \beta_m x_m)} \tag{3.3}$$

Applying the logit transformation:

$$\log\left(\frac{p_1}{1 - p_1}\right) = \beta_0 + \beta_p + \beta_1 x_1 + \cdots + \beta_m x_m \tag{3.4}$$

$$\log\left(\frac{p_0}{1-p_0}\right) = \beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m \tag{3.5}$$

Exponentiating these, we get:

$$\frac{p_1}{1-p_1} = \exp(\beta_0 + \beta_p x_p + \beta_1 x_1 + \cdots + \beta_m x_m) \tag{3.6}$$

$$\frac{p_0}{1-p_0} = \exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m) \tag{3.7}$$

The odds ratio OR is defined as:

$$OR = \frac{\frac{p_1}{1-p_1}}{\frac{p_0}{1-p_0}} = \frac{\exp(\beta_0 + \beta_p + \beta_1 x_1 + \cdots + \beta_m x_m)}{\exp(\beta_0 + \beta_1 x_1 + \cdots + \beta_m x_m)} = \exp(\beta_p) \tag{3.8}$$

The lower bound $lb(OR)$ and upper bound $ub(OR)$ of the OR for prednisone are computed according to the Wald test [A., 2002]. An association is considered significant if $lb(OR) > 1$ at a confidence level $\alpha < 0.05$.

3.3.3. Homogenization of Population Datasets

To analyze the OR of different medical conditions between prednisone users and non-users, we addressed the large imbalance between the populations (232 vs. 29168). To correct for this, we sampled the non-prednisone population by selecting subsets with a size four times that of the prednisone group, while ensuring homogeneity in terms of age, gender, BMI, and ethnicity. Weighted random sampling [E., 1997] was used to select population subsets. The OR was calculated across different subsets, repeated N_{rep} times, and averaged. Since OR is lognormally distributed, the mean OR can be computed as:

$$OR(X) = \exp\left(\frac{1}{N_{rep}} \sum_{n=1}^{N_{rep}} \log(OR(X))\right) \tag{3.9}$$

The same method estimates the average lower and upper bounds $lb(OR)$ and $ub(OR)$. N_{rep} was set to 500. We calculated the OR between prednisone-use and medical conditions in four sets of individuals with the scope to have more homogeneus subpopulation in terms of conditions and comorbidities:

- **Overall:** all the 29400 Individuals extracted from NHANES.
- **RA:** set of 866 individuals with Rheumatoid Arthritis.

- **COPD:** set of 672 individuals with Chronic Obstructive Polmunary Disease.
- **AS:** set of 2585 individuals with Asthma.

Summary statistics describing the 4 populations with all the individuals are shown in Table 3.1 and in Figure 3.1.

Populations	**Individuals (N)**	**Ethnicity (%)**					**SEX (%)**		**Prednisone (%)**
		Mexican-American	**Non-Hispanic Black**	**Non-Hispanic White**	**Other Hispanic**	**Other Race**	**M**	**F**	
Overall	29,400	17	22	34	10	17	49	51	0.8
PREDN	232	11	23	39	12	15	40	60	100
RA	866	15	28	35	12	10	42	58	6.3
COPD	672	4	18	65	5	9	53	47	6.5
AS	2,585	12	30	35	10	13	44	56	2.2

Table 3.1: Populations and Prednisone Use Statistics

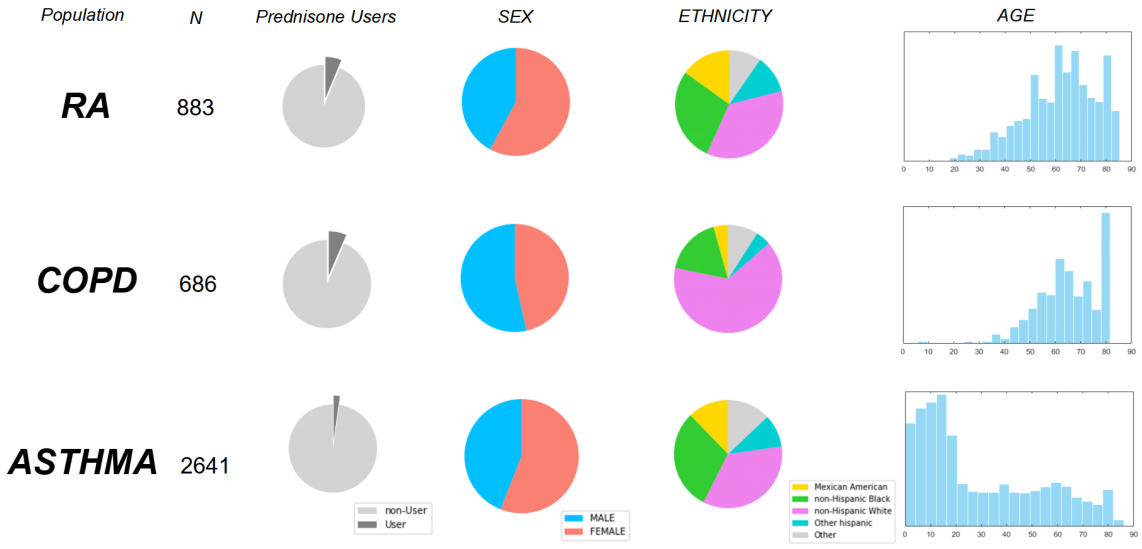

Figure 3.1: Graphical representation of the three population used in the study. The number of individuals N is shown, together with the distribution of prednisone users, sex, ethnicity and age within the population.

3.3.4. Evaluating the functional relation between the use of prednisone and the side-effects using System Biology Analysis

We used the systems biology-based approach described in section 3.1 to identify potential molecular relationships that could explain the associations identified between prednisone use and medical conditions.

In this case, the ANNs analysis is used in two steps. *(i)* At first the functional relation between the prednisone receptor and the motifs describing a certain side-effect (e.g. ADHD) is computed for those conditions that show a significant association with the use of prednisone in the Overall population. *(ii)* If the estimated functional relation is significant (p-value ¡ 0.2) and a significant association is identified in any of the three subpopulations (RA, AS, COPD) then the functional association between all the proteins that are known to be related to the condition describing the population and the motifs describing the side-effect is estimated. The results are discussed in case of significance. Results are reported in Table 4.1.

3.4. The effect of Tocilizumab in GCA-patients PBMCS (Peripheral Blood Mononuclear Cells)

3.4.1. GCA Patients, PBMCS treatment

PBMCs were sampled from 17 GCA female patients from the Hospital Clinical Barcelona.

3.4.2. PBMCS treatment

The PBMCs sampled from the participant's blood have been treated in three different ways.

- UNT: Untreated group.
- IL6: The PBMCs are treated with IL6 to exacerbate the pathophysiology of GCA.
- IL6+TCZ: The PBMCs are treated with IL6 and Tocilizumab.

3.4.3. Data generation

The expression level of 245 genes was measured by means of quantification by sequencing using NanoString techonology. The genes taken into account are shown in Table 7.6.

3.4.4. Differential Differential Analysis

Two differential expression analysis have been performed:

- UNT vs IL6
- IL6+TCZ vs IL6

Differential expression analysis is applied using Wilcoxon signed-rank for the paired experiments and the Wilcoxon rank-sum for the unpaired [Rey and Neuhäuser, 2011], multiple hypothesis correction is applied using the Benjamini-Hochberg method [Benjamini and Hochberg, 1995].

3.4.5. Data analysis: Clustering and Classification

Clustering of the 17 samples is applied to a subset of genes of interest. Z-score normalization is applied to each gene, the clustering method is hierarchical clustering with average aggregation function [Nielsen, 2016], spearman correlation coefficient [Corder and Foreman, 2014] has been used as distance. The clustering solution has been evaluated considering the mean silhouette index [Rousseeuw, 1987], Dunn-index [Dunn, 1973], and the Hopkins statistics [B. and G., 1954].

Classifiers are trained to reclassify the samples into the obtained clusters. The classifier used is linear regression with threshold optimization [Freedman, 2009], Leave One Out (LOO) is the validation method applied [Trevor Hastie, 2009]. The indexes to evaluate the validity of the classifier are True-Positive (TP), True-Negative (TN), False-Positive (FP), False-Negative (FN), Accuracy (ACC), Precision (PREC), Negative Predicted Value (NPV), True-Positive Rate (TPR), True-Negative Rate (TNR) and Matthew-Correlation-Coefficient (MCC) [Matthews, 1975].

$$\mathrm{ACC} = \frac{\mathrm{TN} + \mathrm{TP}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{3.10}$$

$$\mathrm{PREC} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FP}} \tag{3.11}$$

$$\mathrm{TPR} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{3.12}$$

$$\mathrm{NPV} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FN}} \tag{3.13}$$

$$\mathrm{TNR} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{3.14}$$

$$\mathrm{MCC} = \frac{\mathrm{TP} \times \mathrm{TN} - \mathrm{FP} \times \mathrm{FN}}{\sqrt{(\mathrm{TP} + \mathrm{FP})(\mathrm{TP} + \mathrm{FN})(\mathrm{TN} + \mathrm{FP})(\mathrm{TN} + \mathrm{FN})}} \tag{3.15}$$

The criterion for the selection of the best classifiesr is the Matthew Correlation Coefficient (MCC) [Matthews, 1975]. The clustering and classification tool used in this analysis have been described by Coto-Segura et al. and Gil et al., respectively.

3.4.6. System Biology Analysis. The TPMS

The Therapeutic Performance Mapping System (TPMS) is a systems biology tool designed to study the mechanisms of action (MoAs) of drugs by simulating patient-specific protein-protein interaction networks [SL, 2018]. It is built by integrating protein-protein interaction data into Human Protein Networks (HPNs) and can be used to simulate how drugs interact with the human biological system, focusing particularly on the molecular pathways that contribute to clinical outcomes or adverse effects [Jorba et al., 2020]. In this case, TPMS is used to study the effect of tocilizumab on GCA, i.e. to study how the inhibition of IL6 can reverse the activation state of the protein differentially expressed in GCA as reported in the molecular description of the disease in Table 7.2. Furthermore, TPMS allows users to specify the activation state of certain proteins and incorporate these specifications as constraints within the solution search space. MoAs have been computed with restriction, using as constraints a set of proteins/genes that have been identified to be differentially expressed PBMCS treated with tocilizumab.

3.5. MMPred, Molecular Mimicry Predictor

3.5.1. MMPred

This study introduces MMPred, a software tool integrating epitope prediction and sequence alignment algorithms to simplify the setup of computational analyses aimed at the generation, investigation or testing of hypotheses relative to molecular mimicry events in the context of autoimmune diseases. As it stands, the tool provides epitope predictions for HLA class II only, as alleles involved in autoimmunity belong often to this class [Fiorillo et al., 2017]. But the tool can be easily extended to HLA Class I prediction by incorporating the corresponding pre-trained models from CNN-PepPred and NetMHCpan [Reynisson et al., 2020].

3.5.2. Algorithm

The MMPred algorithm is illustrated in Figure 3.2 for uses combining sequence alignment and epitope prediction. In addition, the program may be also used without the alignment feature to streamline epitope prediction using NetMHCIIpan and CNN-PepPred.

Software specifications

The software has been developed using Python3.6.8 in a Linux environment. The Python3 libraries used are Pandas (`https://pandas.pydata.org/`), NumPy [Harris et al., 2020], Matplotlib [Hunter, 2007], pickle (`https://github.com/python/cpython/blob/3.6/Lib/pickle.py`), sklearn [Pedregosa et al., 2011] and tensorflow [Abadi et al., 2016]. Soft-

ware, installation instructions and program user-guide are available in GitHub (`https://github.com/ComputBiol-IBB/MMPRED`).

Input

The program takes a set of protein sequences in the form of a fasta file (QUERY) and a list of HLA alleles (ALLELES) as minimal input. If no additional input is provided, the program runs an epitope prediction for the sequences in QUERY and each of the HLA alleles specified.

The user can specify if QUERY contains entire protein sequences (*protein* mode) or peptides (*peptide* mode). When in *protein* mode, the program makes a prediction for each fragment of size W in the protein (with a 1-residue step), where W is a user-defined parameter. Instead, when using the *peptide* mode a single prediction is performed for the full length of the peptide.

The user can provide a second fasta file (TARGET). In this case, an alignment is performed between the peptides in or derived from QUERY and the TARGET sequences to produce the TARGET* set (see Figure 3.2), containing the sequences from TARGET that show significant (to a value defined by the user) alignment with QUERY peptides. Epitope prediction is then applied to the TARGET* sequences.

In this study, we are using microbial sequences as QUERY and human ones as TARGET. Yet, QUERY and TARGET may be whatever the user thinks appropriate for the analysis in question.

Alignment

Protein sequence alignment is performed to identify the potential similarity between QUERY and TARGET sequences, if TARGET is specified. The alignment can be done using either of two strategies: *i*) using BLASTP to perform an ungapped alignment, with automatic adjustment of parameters for short input sequences or *ii*) using PSI-BLAST to compute a PSSM by aligning the QUERY fragment against a user-defined set of epitopes provided in a fasta file and then using the PSSM to perform a search in TARGET.

If the alignment satisfies a certain significance threshold (either E-value or bit score; by default E-Value < 0.05) the aligned TARGET sequence is stored in TARGET* for epitope prediction. The parameter (E-value or bit score) and its value can be defined by the user.

To perform the epitope prediction, a sequence of length $\geq W$ that we shall call "prediction window" has to be extracted from the alignment. If the length of the alignment is $< W$, then a prediction window of size W centered on the aligned sequence is extracted. If the number of extra residues at left and right cannot be the same, the program automatically takes the extra residue to the right. If the window falls outside the ends of the sequence, the algorithm will take the first or last W residues accordingly. W is by default 15.

Epitope prediction

Whether the alignment is performed or not, epitope predictions are run using NetMHCIIpan 4.1, with both the model trained on Binding Affinity (BA) and that trained on mass spectrometry Eluted Ligands (EL) [Reynisson et al., 2020], and CNN-PepPred [Junet and Daura, 2021]. The three predictions are kept and no consensus score is generated. Note that NetMHCIIpan reports the prediction score and the %Rank for each peptide-HLA pair. The %Rank is a normalized prediction score that enables comparison between different HLA alleles and models (BA and EL). The %Rank of a query sequence is determined by comparing the prediction score to a score distribution for a random set of natural peptides, with %Rank = 1 meaning that the queried sequence obtained a prediction score in the highest 1% of the distribution. On the other hand, CNN-PepPred reports only prediction scores. To make results from CNN-PepPred comparable to NetMhcIIpan, a score distribution for natural peptides was generated for each HLA allele available in CNN-PepPred, using a random sample of 10,000 peptides extracted from UniRef50 [The UniProt Consortium, 2023].

Output

The output depends on the input parameters as follows:

- *No Alignment, Protein mode:* The program will return the predicted core (sequence of length 9) for every window of size W achieving %Rank ≤ 10 in every protein included in QUERY, together with the ID of the protein, start and end position of the W-residue peptide, start and end position of the predicted core, prediction method, score, %Rank and the allele for which the prediction has been made.

- *No Alignment, Peptide mode:* The program will return the predicted 9-mer core for every peptide included in QUERY achieving %Rank ≤ 10, together with the ID of the peptide, start and end position of the predicted core, prediction method, score, %Rank and the allele for which the prediction has been made.

- *Alignment*: For every sequence in TARGET* achieving %Rank ≤ 10, the program will return the predicted 9-mer core, the TARGET sequence ID, the start and end position of the alignment in the TARGET sequence, the start and end position of the window of size W extracted from the alignment in the TARGET sequence, the start and end position of the predicted core in the TARGET sequence, identity, E-value and bit score of the alignment, the aligned TARGET sequence, the QUERY sequence ID, the start and end position in the QUERY sequence of the alignment, the aligned QUERY sequence, prediction method, score, %Rank and the allele for which the prediction has been made.

3.5.3. Evaluation Datasets

Microbial Epitope Dataset (MEDS)

A dataset of known HLA class II epitopes from infectious agents was manually downloaded from IEDB [Vita et al., 2019] (`https://www.iedb.org`). The search terms "Epitope: Linear peptide", "Epitope source: Bacteria, Virus", "Host: Human", "Assay: MHC ligand", "Outcome: positive", "MHC Restriction: Class II", "Disease: any" were used (date 11/10/2023).

Epitopes lacking either the UniProtKB accession number (AC) [The UniProt Consortium, 2023] (`https://www.uniprot.org`) of the protein, start or end position in the sequence or the HLA allele for which they were tested, were excluded. Epitopes containing modified amino-acid residues were also discarded. Additionally, only epitopes from microorganisms known to be related to autoimmune diseases were considered. The relationship between an organism and the occurence of autoimmune diseases was determined by review of the literature when not annotated in IEDB. For each organism with epitopes fulfilling the previous selection criteria, a PubMed (`https://pubmed.ncbi.nlm.nih.gov/`) search for reviews from the last 10 years using the keywords "organism_name AND (autoimmunity OR autoimmune)" was performed. As last filter, epitopes associated to HLA alleles with no model available in CNN-PepPred and NetMHCIIPan were also discarded.

After these filters, MEDS contained 3,676 epitopes from 88 proteins of 13 microorganisms, associated to 50 HLA class II alleles, with a total of 9,229 epitope-allele pairs.

Human Autoepitope DataSet (HADS)

A dataset containing known human autoepitopes was downloaded from IEDB (11/10/2023). The search terms were: "Epitope: Linear peptide", "Epitope source: Human", "Host: Human", "Assay: MHC ligand", "Outcome: positive", "MHC Restriction: Class II", "Disease: autoimmune".

To avoid redundancy in the form of nested sets, epitopes from the same protein that showed overlap and were linked to the same allele were merged into a single epitope. After the merging, only sequences with length ≥ 15 were kept. Epitopes associated to HLA alleles with no model available in CNN-PepPred and NetMHCIIPan were also discarded.

HADS thus contained 807 epitopes from 608 different human proteins, associated to 5 different HLA class II alleles in the context of Rheumatoyd Arthritis and Multiple Sclerosis (Supplementary material file HADS_summary.xslx).

Human Proteome Dataset (HPDS)

HPDS contains the sequences of the 20,426 reviewed human proteins found in UniProt [The UniProt Consortium, 2023] on the date of the download (18/10/2023).

SARS-CoV-2 Proteome Dataset (SC2DS)

SC2DS was generated from the SARS-CoV-2 reference proteome [Wu et al., 2020] (UniProt identifier: UP000464024, downloaded 11/01/2024) and the Pangolin variants [Rambaut et al., 2020] (downloaded 01/03/2024) A.23.1-like, A.23.1-like+E484K, Alpha_(B.1.1.7-like), AV.1-like, B.1.1.318-like, B.1.1.7-like+E484K, B.1.617.1-like, B.1.617.3-like, Beta_(B.1.351-like), Delta_(AY.4.2-like), Delta_(AY.4-like), Delta_(B.1.617.2-like), Delta_(B.1.617.2-like)_+K417N, Epsilon_(B.1.427-like), Epsilon_(B.1.429-like), Eta_(B.1.525-like), Gamma_(P.1-like), Iota_(B.1.526-like), Lambda_(C.37-like), Mu_(B.1.621-like), Omicron_(BA.1-like), Omicron_(BA.2-like), Omicron_(BA.3-like), Omicron_(BA.4-like), Omicron_(BA.5-like), Omicron_(Unassigned), Omicron_(XBB.1.16-like), Omicron_(XBB.1.5-like), Omicron_(XBB.1-like), Omicron_(XBB-like), Theta_(P.3-like), XBB-parent1, XBB-parent2, XE-parent1, XE-parent2, Zeta_(P.2-like). The downloaded sequences were manually split into all possible overlapping fragments of size 15. Fragments of variant sequences were only kept if they had indels or mutations relative to the reference proteome. SC2DS thus contained 9,608 15-mers, of which 243 from Pangolin variants.

MHC class II epitopes Dataset (MHCII-EDS)

MHCII-EDS contains all HLA class II epitopes available in IEDB (date 18/03/2024), obtained with the search terms: "Epitope: Linear peptide", "Epitope source: Any", "Host: Any", "Assay: MHC ligand", "Outcome: positive", "MHC Restriction: Class II", "Disease: Any". A total of 485,020 epitopes were downloaded. Redundancy was eliminated by clustering the sequences at 95% identity with CD-HIT [Li and Godzik, 2006], then using the centroid as cluster representative to obtain a final set of 155,923 epitopes. This dataset was used for the computation of the PSSM in all the analyses where PSI-BLAST was used as alignment algorithm.

3.5.4. MMPred evaluation

Evaluation setup

To evaluate the algorithm, five sets of predictions were obtained:

1. Sequences from the human autoepitope dataset (HADS) (TARGET) that significantly align with sequences from the microbial epitope dataset (MEDS) (QUERY) using BLASTP and were positive for binding to HLA class II (with or without allele match, see below) with CNN-PepPred and/or NetMHCIIpan.

2. Same as prediction set 1 but using PSI-BLAST for the alignment.

3. Same as prediction set 1 but with the human proteome dataset (HPDS) as TARGET.

4. Same as prediction set 3 but using PSI-BLAST for the alignment.

5. Sequences from HPDS (TARGET) that significantly align with sequences from the SARS-CoV-2 proteome dataset (SC2DS) (QUERY) using BLASTP and where positive for binding to HLA class II with CNN-PepPred and/or NetMHCIIpan.

6. Same as prediction set 5 but using PSI-BLAST for the alignment.

Prediction sets 1 and 2 were based on a threshold E-value of 0.05 for the alignments and a %Rank ≤ 2 as condition for binding —Reynisson et al. [Reynisson et al., 2020] defined peptides with %Rank ≤ 2 as strong binders and peptides with $2 <$ %Rank ≤ 10 as weak binders. Prediction sets 3 and 4 were instead obtained in replicates by using threshold E-values from 0.1 to 0.001 and threshold %Rank values of 2 and 10. Owing to the results obtained for prediction sets 3 and 4, prediction sets 5 and 6 were based on a threshold E-value of 0.01 for the alignments and a %Rank ≤ 2.

Prediction sets 1 to 4 were evaluated according to two allele-selection criteria:

1. AllHLA: epitope prediction for HADS (or HPDS) sequences that had at least one significant alignment with MEDS sequences was performed for all alleles.

2. OneHLA: epitope prediction for HADS (or HPDS) sequences that had at least one significant alignment with MEDS sequences was only performed for the allele(s) corresponding to the microbial epitope-allele pair(s) indicated in MEDS.

Furthermore, they were also evaluated both ignoring and considering allele matches:

1. *Epitope* prediction: sequences in HADS (or HPDS) were considered to be predicted as mimicry-induced autoepitopes if there was at least one alignment with MEDS sequences that satisfied the threshold E-value and there was at least one prediction from CNN-PepPred or NetMHCIIpan (BA or EL models) that satisfied the threshold %Rank for any HLA allele.

2. *Epitope-allele* prediction: epitope-allele pairs in HADS (or HPDS) were considered to be predicted as autoepitope-allele pairs if there was at least one alignment with MEDS that satisfied the threshold E-value, and there was at least one prediction from CNN-PepPred or NetMHCIIpan (BA or EL models) that satisfied the threshold %Rank for the same HLA allele of the MEDS pair.

Prediction sets 1 and 2: Supervised evaluation

Prediction sets 1 and 2 may be viewed as a supervised evaluation, since the TARGET sequences are labeled, i.e., they are known to be human autoepitopes, albeit not necessarily relatable to infection events. The remaining prediction sets have the entire, unlabeled human

proteome (HPDS) as TARGET.

The allele set used for prediction contained all alleles for which a model exists in either CNN-PepPred or NetMHCIIpan and are present in either HADS or MEDS, totalling 58 alleles shown in Table 7.8.

Prediction sets 3 and 4: Functional evaluation

Prediction sets 3 and 4 were used to investigate potential functional relationships between predicted autoantigens and the pathophysiological pathways associated with specific autoimmune diseases. This investigation involved a post-analysis of the predicted autoantigens using a systems biology approach as described in section 3.1.

In this study, S was determined for the relationship between each of the human proteins from HPDS that we predicted to be autoantigens as a consequence of microbial peptide mimicry and a list of patophysiological motifs characteristic of each selected autoimmune disease. The list of autoimmune diseases was compiled from the same articles used for the generation of MEDS (see section 3.5.3), and each autoimmune disease was then mapped to corresponding pathophysiological motifs compiled in BED AS as shown in Table 7.7:

For a given BED condition, both the separate protein sets corresponding to the individual motifs and a single protein set corresponding to all motifs of the condition were used. The motifs to be tested for each predicted autoantigen were selected using the following logical sequence: predicted autoantigen $\rightarrow$ microbial protein with matching epitope sequence $\rightarrow$ infectious organism $\rightarrow$ organism's related autoimmune diseases $\rightarrow$ BED motifs.

The distribution of S for the predicted autoantigens was compared to the distribution for a random subset of 1000 samples from HPDS, as surrogate for a random distribution, using the one-sided Mann-Whitney U test [Di Bucchianico, 1999].

Prediction sets 5 and 6: SARS-CoV-2 peptide mimicry

Prediction sets 5 and 6 illustrate an actual application of the tool: the identification of potential human autoantigens resulting from SARS-CoV-2 peptide mimicry.

The analysis included those HLA class II alleles for which there is experimental evidence of their binding of human autoepitopes from HADS or SARS-CoV-2 epitopes from MEDS, plus a set of alleles from HLA-Spread [Dholakia et al., 2022] associated to autoimmune diseases that have been linked to SARS-CoV-2. As further filter, only those alleles for which there was a model in either CNN-PepPred or NetMHCIIpan were considered, leading to a total of 45 alleles (see Table 7.8.

We applied the same systems biology approach used on prediction sets 3 and 4 to explore

potential functional relationships between predicted autoantigens and pathopysiological pathways associated to specific autoimmune diseases. To that end, we choose the BED motifs corresponding to the following autoimmune diseases associated with SARS-CoV-2 infection as shown in Table 7.7: Anemia, Diabetes type 1, Guillain-Barre Syndrome, Myasthenia Gravis, Rheumatoyd Arthritis and, Lupus Erythematosus Systemic [Ehrenfeld et al., 2020, Knight et al., 2021]. For each condition motif, the background distribution of the score S was calculated using proteins that had $S > 0$, lacked a predicted autoepitope, and were not present in the motif's description. For the predicted autoantigens, the score S and its corresponding percentile, Perc(S) (indicating where it falls within the background distribution), were determined. Perc(S) served as an indicator of a potential functional relationship between the autoantigen and the pathophysiological motif. Thresholds were established as follows: Perc(S) > 95 indicated a weak functional relationship, Perc(S) > 99 indicated a functional relationship, and Perc(S) > 99.9 indicated a strong functional relationship.

3.6. MMPred, Varicella Zoster Virus and Giant Cell Arteritis

3.6.1. VZVDS, Varicella Zoster Virus Dataset

Sequences belonging to Varicella Zoster Virus have been extracted from UniprotKB [The UniProt Consortium, 2023] downloading all the Reviewed sequences related to the Varicell Zoster Virus (strain Dumas, taxon ID: 10388). A total of 69 proteins have been manually split into 34747 overlapping 15-mers.

3.6.2. MMPred analysis

VZVDS is aligned against HPDS. MMPred is applied both using BLASTP and PSI-BLAST, in the latter case the MHCII-EDS dataset is used for the computation of the PSSMs. Only the allele DRB1*04:01 is used in the prediction, being the only HLA-II allele known to be related to GCA according to HLA-Spread [Dholakia et al., 2022, Prieto-Peña et al., 2021]. $E - Value < 0.01$ and $\%Rank < 2$ are the thresholds applied.

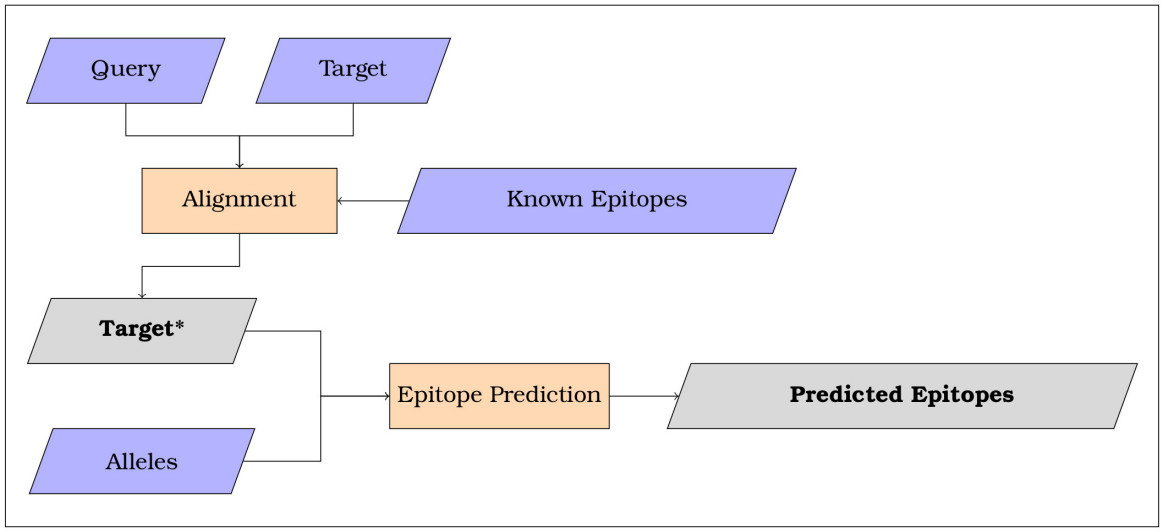

Figure 3.2: Scheme summarizing the overall workflow of MMPred when used with alignment. Input in purple, process in orange and output in grey.

Chapter 4

Results & Discussion

4.1. GCA Molecular Description

The systematic review of papers related to GCA provided 106 papers describing the behavior of 88 different proteins related to the disease. Those proteins have been categorized into 8 motifs, which in turn are organized into causative and manifestative motifs. These results are summarized in Table 7.1 and Table 7.2.

4.1.1. Causative motifs

GCA genetic associations

A genetic component in the pathogenesis of GCA is supported by observations of sporadic family clustering of affected members, along with the predominance of the disease in whites, particularly those from northern Europe or of northern European descent [Liozon et al., 2009, Carmona et al., 2014]. Indeed, an increased risk of GCA is associated with polymorphisms in a variety of genes that mediate immune, inflammatory, and vascular responses [Terrades-Garcia and Cid, 2018].

A genetic predisposition has been found, with strong relevance on some alleles of the Human Leukocyte Antigen (HLA) class I and II regions. For what concerns class II, the HLA-DRA locus and the allele HLA-DRB1*04:01 are known to be related to the disease. DQA1*03 and DQB1*03 are also reported to contribute to the risk development of the disorder [Terrades-Garcia and Cid, 2018, Carmona et al., 2015, Stamatis, 2020, Al-Mousawi et al., 2019, Prieto-Peña et al., 2021]). About class I, GCA is associated with HLA-B locus [Al-Mousawi et al., 2019]. HLA-II is a highly polymorphic protein complex present in antigen-presenting cells (APC). Genetic regions such as DR define the peptide-binding groove, which is formed by alpha and beta chains. The proper function of HLA-II is needed in order to develop acquired immunity, as its main role is displaying antigenic peptides to CD4+ T lymphocytes [Alvaro-

Benito and Freund, 2020, Wieczorek et al., 2017]. Alterations located in HLA-II chains that create peptide-binding grooves may lead to an antigen-presentation disease and thus a dysfunction in adaptative immunity [Terrades-Garcia and Cid, 2018]HLA class-I is a protein complex present in all nuclear cells, and its main role is to activate CD8+ T lymphocytes in pathological situations. A dysfunction of it could also contribute to an immune disorder [Wieczorek et al., 2017].

Despite the fact that HLA protein complex suppose the strongest genetic predisposition to suffer from GCA, there are also other genes involved. For example, TNF-alpha microsatellite or interleukin-6 (IL-6) are suggested to contribute to the development of this disease. However, further studies are needed to establish a more robust genetic association. It is not the case of IL-12 beta single nucleotide polymorphism (SNP) rs755374, which reported to have a strong genetic relationship with GCA in several meta-analysis and cohorts [Al-Mousawi et al., 2019, Carmona et al., 2017a, Stamatis, 2020]. Another gene with confirmed association with this pathology was the one that codes for tyrosine phosphatase non-receptor type 22 (PTPN22). The relevance of this receptor lies in its function, which is to participate in T and B cell receptor signaling pathways [Kuret et al., 2019]. Literature also describes as an important genetic factor to develop GCA Interleukin 10 (IL-10) and vascular endothelial growth factor (VEGF) [Kuret et al., 2019].

Regarding vascular remodeling and angiogenesis, PLG and P4HA2 variants showed GCA predisposition in a GWA study. PLG gene encodes a secreted blood zymogen that can be converted into two different proteins, plasmin and angiostatin, which are important in a wide spectrum of physiological processes such as angiogenesis and lymphocyte recruitment and inflammation via production of cytokine and ROS, all of them relevant processes in GCA. P4HA2 encodes an isoform of alpha subunit of the collagen prolyl 4-hydroxylase, essential for collagen biosynthesis. This gene has been considered important in hypoxia response and its expression is regulated by hypoxia-inducible factor-1 (HIF-1), which also targets other genes related to the processes involved in GCA, suggesting the involvement of P4HA2 in GCA [Carmona et al., 2017b]. Apart from the genetic variants or alterations mentioned above, there are other polymorphisms in a variety of genes encoding for molecules participating in immune, inflammatory, and vascular responses (i.e. NOS2) that are associated with increased GCA risk [Carmona et al., 2014, Kuret et al., 2019].

Dysfunction of immune checkpoints

Large and medium vessels have their walls protected from inflammation processes due to the expression of Programmed Cell Death 1 Ligand 1 (PD-L1) by the endothelial and dendritic cells. The display of this ligand allows it to maintain self-tolerance by binding Programmed Cell Death Protein 1 (PD-1) receptors from T lymphocytes. This interaction confers an immunoprivileged environment due to the function carried by these structures are vital and need to be protected. PD1/PDL1 checkpoint impairs T cell activation, effectors secretion

and survival, avoiding the destruction of the blood vessel tissue [Weyand et al., 2018].

PD1/PDL1 checkpoint is compromised in GCA patients and thus T cells and macrophages cause damage to wall arteries and fostering its pathological remodeling. This balance is altered because of a synergy created by low PD-L1 expression from Dendritic cells and high PD-1 expression from T cells. Therefore, in this scenario self-tolerance is lost and immunoprotection of the arterial wall is disrupted. As a consequence, T cells proliferation and activation is fostered [Zhang et al., 2017]. IFN-γ, IL-17, IL-9 and IL-21 are secreted by lymphocytes T which generates an inflammation environment that enhances hyperplasia of intimal artery layer, fragmentation of the elastic lamina and intramural angiogenesis [Weyand et al., 2018, Zhang et al., 2017, Watanabe et al., 2017]. Moreover, activated T cells, activated macrophages, multinucleated giant cells as well as dendritic cells will start to accumulate and form granulomatous infiltrates that will obstruct affected arteries [Watanabe et al., 2017].

The disruption PD1/PDL1 checkpoint and its chronification over time cause autoimmune inflammation that will severely damage aorta and its related branches. Specifically, those structures will experiment lumen occlusion, rupture and aneurysm formation [Watanabe et al., 2017].

Macrophages have a key role in vessel remodeling. According to the layer where they are located, a specific combination of substances will be secreted to produce the wall alteration. For example, macrophages infiltrated in the Media layer produce growth factors such as platelet-derived growth factor (PDGF) , Vascular endothelial growth factor (VEGF) and Fibroblast growth factor (FGF), whereas those infiltrated in the adventitia secrete transforming growth factor-β1, IL-6, and IL-1β. These cytokines foster adventitial fibrosis, contributing to structural rigidity and localized inflammation. Together, these processes orchestrate vascular remodeling that is central to the pathology of Giant Cell Arteritis [Watanabe et al., 2017].

4.1.2. Symptomatic Motifs of GCA

The current understanding on the pathophysiology of GCA invokes interactions occurring between the innate and adaptive immune system and the different compartments of the arterial wall, including endothelial cells and vascular smooth muscle cells (VSMCs), such that vascular inflammation, remodelling, and occlusion may occur. GCA is initiated by an unknown danger factor (may be bacterial or viral products) that triggers the early vascular inflammation (Motif: "$VascularInflammation - Early$"). This first response is amplified (Motif: "$VascularInflammation - Amplificationcascades, Persistent$") and may further result in a systemic inflammation (Motif: "$SystemicInflamation$"). During the filtration of immune cells into the site of lesion, the vascular wall results damaged (Motif 4) and a repairing response, vascular remodelling (Motif 5), is initiated. Although the prevalence of GCA is the highest in temporal artery, extracranial vessels are also found to be affected by GCA, being the aorta the main site of inflammation. Aortic aneurysm may be formed

secondary to GCA (Motif 6).

Vascular Inflammation - Early

The inflammation starts in the adventitia and progresses to the other layers of the arterial wall, culminating in transmural damage. Activated dendritic cells (DCs) play an important role in the immune response of GCA. The activation of DC breaks immune tolerance and renders the arteries susceptible to inflammatory injury [Terrades-Garcia and Cid, 2018].

During the initiation phase of GCA, dendritic cells (DCs) within the adventitia of arteries are activated through toll-like receptors (TLRs) triggered by an unknown antigen via pathogen-associated molecular patterns (PAMPs), microorganism-associated molecular patterns (MAMPs) and damage-associated molecular patterns (DAMPs). The DC activation results in the production of chemokines (CCL19 and CCL21) that attract and retain additional DCs and expression of co-stimulatory molecules CD83 and CD86, which interact with the T cell receptor complex and CD28 to activate CD4+T cells. In turn, the expression of CD83/86 is modulated by immune checkpoints. In GCA, it has been observed that the programmed death 1 (PD-1) receptor/programmed death ligand 1 (PD-L1) immune checkpoint is inefficient in GCA-affected temporal arteries; this is thought to contribute to the excessive infiltration of activated T cells into affected medium- and large-sized blood vessels [Terrades-Garcia and Cid, 2018]. Furthermore, studies demonstrated the efficacy of administrating a fusion protein of cytotoxic T-lymphocyte-associated protein 4 (CTLA4) and the fragment crystallizable region of a human IgG1 to treat GCA patients. CTLA4, a competitive protein binding to CD80/86, can inhibit T cell activation by impeding the interaction of CD28 with CD80/86 [Langford et al., 2017].

DCs also release cytokines, such as IL-1β, IL-6, IL-23, and IL-21 or IL-12 and IL-18, that are responsible for triggering the differentiation into two distinct T cell subgroups. The first set of cytokines induces the differentiation of activated T cells into Th17 cells; the second drives Th1 cell formation. Th17 cells secrete IL-17A, modulating the function of endothelial cells (ECs), vascular smooth muscle cells (VSMCs), fibroblasts, and bone marrow stromal cells, while Th1 cells release interferon γ (IFN-γ), involving in activation of macrophages, ECs, VSMCs, and cytotoxic cells [Dammacco et al., 2020].

Levels of IL-17A are found to be elevated in GCA lesions and are rapidly reduced following treatment with glucocorticoids, suggesting that IL-17A suppression may contribute to the improvement of symptoms in patients with GCA who receive high-dose glucocorticoid therapy. Interestingly, strong expression of IL-17A in the involved arteries of patients with GCA was associated with a better response to glucocorticoid therapy with few relapses [Terrades-Garcia and Cid, 2018, Espígol-Frigolé et al., 2013].

B cells are scarcely present in the arterial wall of all the patients with GCA, suggesting

that its effect is not crucial in the pathogenesis. Decreased levels of B cells are observed in patients with newly diagnosed GCA but return to normal levels during corticosteroid-induced remission. Interestingly, returned B cells in patients showed an enhanced potential to produce IL-6, suggesting the presence of a delicate interplay between B cells and T cells in GCA as IL-6 production by B cells is promoted by T cell help, and IL-6 is important for the development of Th17 cells [van der Geest et al., 2014].

DNA methylation is a component of the epigenetic and transcriptional regulatory system in the cell. Studies have shown that hypomethylation of several genes involved in T cell maturation and activation, such as IFNG, TNF, NLRP1, PTPN22, RUNX3, CCR7, PPP3CC, NFATC1 and NFATC2, are present in GCA patients, suggesting a central role of T cells in GCA [Coit et al., 2016].

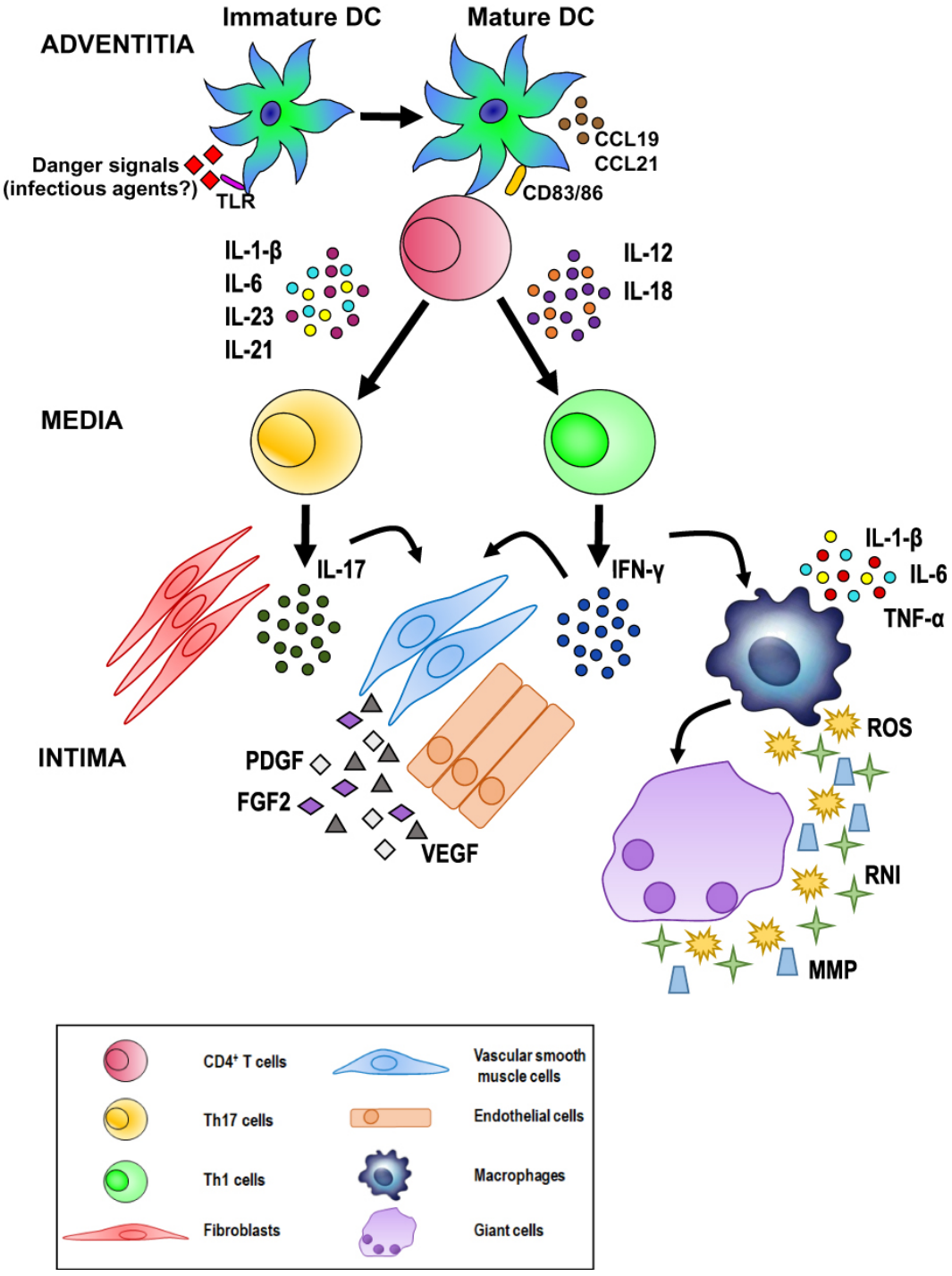

Figure 4.1: Possible pathogenetic algorithm of giant cell arteritis (GCA). After activation by danger signals, dendritic cells resident in the arterial adventitia mature, produce chemokines such as CCL19 and CCL21, and express the co-stimulatory molecules CD83 and CD86 required for their interaction with CD4+ T cells. Dendritic cells also release cytokines, such as IL-1 β, IL-6, IL-23, and IL-21 or IL-12 and IL-18, that trigger two distinct networks. The first network induces the differentiation of activated T cells into Th17 cells; the second drives Th1 cell formation. Both T cell lineages participate in the evolving granulomatous inflammationImage taken from Dammacco et al.. *Reprinted from Clinical Ophthalmology 2023:17 623-632'* Originally published by and used with permission from Dove Medical Press Ltd.

Vascular inflammation – Amplification cascades - Persistent

Following the initiating events of GCA, amplification cascades play an important role in the development and progression of inflammatory infiltrates, the development of full-blown transmural inflammation, vascular wall injury, and remodeling, the pathological substrate of clinical symptoms and complications of GCA [Terrades-Garcia and Cid, 2018].

Th1 cells release IFN-γ, and this induces the production of several chemokines (CCL2, CXCL9, CXCL10, and CXCL11) by VSMCs. CCL2 leads to the recruitment of monocytes which express its receptor (CCR2) and then differentiate into macrophages and multinucleated giant cells, which contributes to the development of granulomas at the intima-media junction. CXCL9, CXCL10, and CXCL11 trigger the recruitment of immune cells expressing their receptor (CXCR3), i.e. Th1 and CD8+ T cells, thus leading to an increase in the production of IFN-γ, which probably contributes to the initiation of a positive forward loop supporting the chronic Th1 inflammatory response observed in GCA [Samson et al., 2017]. CD8+ T cell may play a role in initiating vascular remodeling as its strong infiltration is associated with more severe disease. CD8+ T cells are infiltrated into the arterial wall and produce cytokines (IL-17 and IFN-γ) and cytotoxic molecules (granzymes and perforin) [Samson et al., 2017].

Once activated, macrophages release pro-inflammatory cytokines such as tumor necrosis factor *alpha* (TNF-α), IL-1β, and IL-6, thus amplifying the inflammatory response [Dammacco et al., 2020]. Upregulation of chemokines, endothelial adhesion molecules, and colony-stimulating factors in lesions may reinforce the inflammatory loops associated with GCA. Consequently, it results in the continuous recruitment and expansion of additional inflammatory cells. A study has shown that IL-23p19 might be involved in leukocytes attachment and transmigration by promoting the overexpression of adhesion molecules ICAM-1 and VCAM-1 (http://hdl.handle.net/10803/401863).

Angiogenic factors, such as VEGF, fibroblast growth factor-2, and PDGFs produced by macrophages may promote the formation of new vessels in vascular injuries of GCA. Acute phase proteins such as haptoglobin and serum amyloid A, typically increased in patients with GCA, may also be angiogenic [Cid et al., 1993, O'Neill et al., 2015]. The expression of endothelial adhesion molecules by neovessels facilitates the recruitment of additional leucocytes. While angiogenesis is an important process in the progression and maintenance of chronic inflammatory diseases, such as GCA, inflammation-induced angiogenic activity may also play a compensatory role for ischaemia at distal sites in patients with GCA, thus protecting against ischaemic complications [Terrades-Garcia and Cid, 2018].

Systemic Inflammation

Systemic inflammation is a syndrome associated with arteritis in the majority of the patients with GCA. This reaction takes place distant from the areas of inflammation and it involves the

participation of multiple organs and systems. As consequence, clinical manifestations such as fever, anorexia, weight loss, hematologic abnormalities (i.e., anemia and thrombocytosis), biochemical alterations (acute phase protein synthesis), and metabolic changes (i.e., increased lipolysis and muscle loss) are characteristically occurred [Weyand et al., 2012, Hernández-Rodríguez et al., 2002]. Macrophages at the site of injury produce pro-inflammatory cytokines, mostly IL-1β, TNF-α and IL-6, that have both local and systemic effects. In fact, the intensity of the systemic inflammatory response in GCA correlates with the expression of these pro-inflammatory cytokines. Moreover, circulating TNF-α and IL-6, along with tissue expression of TNF-*alpha*, have been shown to correlate with relapses and disease persistence [Terrades-Garcia and Cid, 2018, Hernández-Rodríguez et al., 2002].

Vascular Injury

In the media, macrophages activated by IFN-γ are involved in the production of toxic mediators to the arterial tissue like reactive oxygen species, causing lipid peroxidation of phospholipids.

Nitric oxide (NO), produced through induced NO-synthase (iNOS), triggers nitration of endothelial proteins and matrix metalloproteinase 2 (MMP2) and MMP9, also produced by vascular smooth muscle cells (VSMCs), causing the destruction of the media and digestion of the internal elastic lamina, degrading the extracellular matrix [Samson et al., 2017].

Indeed, an increased proteolytic activity is observed in GCA lesions, MMPs are found to be upregulated whereas its natural inhibitor, tissue inhibitor of metalloproteinases (TIMP-1 and TIMP-2), are downregulated [Terrades-Garcia and Cid, 2018]. This increment may contribute to the disruption of elastic fibers, favoring aortic dilatation, and abnormal vascular remodeling [Cid, 2014].

Arterial remodeling and vascular occlusion, Vascular injury and remodeling

Activated macrophages or injured VSMCs release growth factors such as platelet-derived growth factor (PDGF), fibroblast growth factor-2 (FGF2), and vascular endothelial growth factor (VEGF), that are responsible for triggering a vascular remodeling process leading to myofibroblast differentiation of VSMCs, migration towards the intimal layer and deposition of cellular extramatrix proteins [Planas Rigol and Corbera Bellalta, 2016, Hid Cadena et al., 2018]. This vascular remodeling program can be considered as dysfunction repair in response to the injured arterial cells that lead to intimal hyperplasia, luminal occlusion, ischemia, and eventually end organ damage [Weyand et al., 2012]. Several factors including TGF-β, endothelin-1, nerve growth factor (NGF), brain-derived neurotrophic factor (BDNF), and sortilin are also thought to contribute to myofibroblast activation and production of matrix, eventually leading to vascular occlusion. Furthermore, endothelin-1 expression in GCA injuries is not downregulated after glucocorticoid therapy, suggesting that mechanisms of vessel occlusion may require a specific approach in vessel vasculitis [Cid, 2014, Dejaco et al., 2017].

Vessel stenosis and luminal occlusion are generated as a consequence of excessive intimal hyperplasia. Of all the mechanisms leading to vascular injury, uncontrolled hyperproliferation of the intimal later produces the most significant consequences of GCA, particularly blindness and stroke [Weyand and Goronzy, 2002].

Chronic inflammation has long been associated with angiogenesis. In GCA, it has been found that the number of new vessels observed correlates with the extent of internal elastic lamina damage, suggesting the relationship between elastic tissue digestion and neovascularization. Moreover, a dual role of angiogenesis has also been suggested as it might play role both in vasculitic processes with a primary proinflammatory role and a secondary role compensating for ischemia [Mitchell and Cestari, 2009]. On these newly formed vessels in GCA, intense expression of constitutive endothelial adhesion molecules such as PECAM-1, P-selectin, and ICAM-1 are found, particularly at the medial-intimal junction. In addition, expression of the inducible endothelial adhesion molecules, E-selectin and VCAM-1, are also found in many patients. This suggests that neovascularization may be involved in providing new entries for infiltrating leukocytes, thus recruiting new cells into the inflammatory infiltrate [Samson et al., 2017].

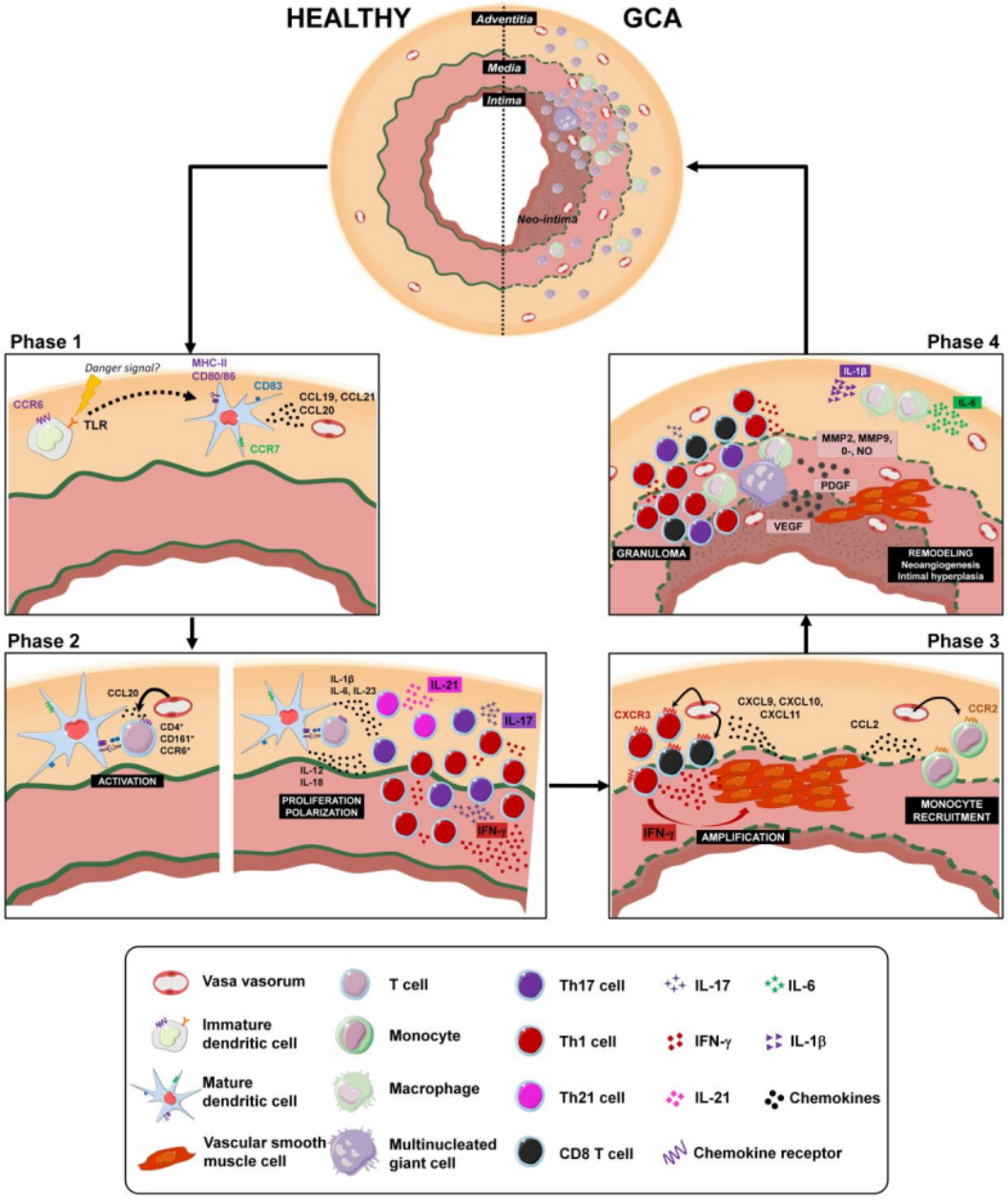

Figure 4.2: Immunopathological model of GCA. Phase 1: loss of tolerance and activation of resident dendritic cells in the adventitia. Phase 2: recruitment, activation, and polarization of CD4+ T cells. Phase 3: recruitment of CD8+ T cells and monocytes. Activation of macrophages, formation of giant cells, and injury of VSMCs. Phase 4: vascular remodeling. Reprinted from *Reprinted from Autoimm. Rev., 16(8), Samson et al., Recent advances in our understanding of giant cell arteritis pathogenesis, Pages 833-844, Copyright (2017), with permission from Elsevier.*

Aortic Aneurysm Formation

Although GCA predominantly involves branches of the external carotid artery, extratemporal large vessels such as arteries of the upper and lower extremities as well as the aorta can also be the site of GCA injury. GCA is rare but is the most frequent cause of non-atheromatous aortic aneurysm. Aortic involvement is generally asymptomatic and concurrently identified at the time of GCA diagnosis. In symptomatic aortitis, clinical manifestations include back pain, abdominal pain, and dyspnea [Kwon et al., 2015]. Aortic aneurysm can be classified either as thoracic aortic aneurysm or abdominal aortic aneurysm according to its spatial distribution, in the chest or below the diaphragm, respectively. Aneurysms are clinically important complications as spontaneous aortic rupture can be life-threatening [Salameh et al., 2018]. The formation of aortic aneurysm is found to be elevated in GCA patients five years after the diagnosis and remains elevated up to 20 years after diagnosis, suggesting that the long-term vascular inflammation in GCA patients may lead to a reduced ability to resist the high pressures milieu of the aorta that eventually causes aortic aneurysm or dissection [Carvajal Alegria et al., 2021].

Generally, formation and rupture of aortic aneurysms have been attributed to the infiltration of inflammatory cells, apoptosis and dysfunction of VSMCs, and breakdown of collagen and elastic fibers, as VSMCs and extracellular matrix (ECM) proteins are critical in a functional vessel, particularly in preserving its mechanical compliance with pulsative blood flow. The imbalance between levels of matrix metalloproteinases (MMP2 and MMP9) and their inhibitors (TIMP1) points towards the excessive degradation of the ECM and progressive aortic wall deterioration. MMP9 expression level can be regulated through the JNK1/2 and ERK1/2 signaling pathways, these kinases phosphorylate and activate the transcription factor AP-1, which drives the transcription of several MMP genes. In addition, inflammatory cytokines such as TNF-α can also activate AP-1, thus promoting MMP expression. Several MMP promotors contain NF-kB binding sites, suggesting its role in aneurysm formation [Barbour et al., 2007, Ince and Nienaber, 2007]. The pathology of aortic aneurysm is also associated with heritability. Several pathogenic genes have been found to cause aortic aneurysm and they encode proteins involved in ECM of the aortic wall, fibrillin-1 (FBN1) and microfibrillar-associated protein 5 (MFAP5); VSMC contraction or metabolism, ACTA2, MYH11, MYLK, PRKG1, and MAT2A; and canonical transforming growth factor-beta signaling, TGFBR1, TGFBR2, TGFB2, and SMAD3 [Wu, 2018].

4.2. The side-effects of Prednisone

Results of the analysis are shown in Table 4.1.

Population	***Condition***	***Annotation***	***OR***
Overall	ACNE VULGARIS ALL	SE	1.3 (0.1; 16.4)
RA	ACNE VULGARIS ALL	SE	-

COPD	ACNE VULGARIS ALL	SE	-
AS	ACNE VULGARIS ALL	SE	-
Overall	ALZHEIMER DISEASE	UNK	0.7 (0.0; 13.3)
RA	ALZHEIMER DISEASE	UNK	3.4 (0.2; 51.9)
COPD	ALZHEIMER DISEASE	UNK	1.1 (0.1; 14.3)
AS	ALZHEIMER DISEASE	UNK	20.9 (0.7; 661.4)
Overall	ANGINA PECTORIS	UNK	1.7 (0.8; 3.7)
RA	ANGINA PECTORIS	UNK	1.0 (0.3; 3.5)
COPD	ANGINA PECTORIS	UNK	0.8 (0.3; 2.4)
AS	ANGINA PECTORIS	UNK	0.4 (0.0; 2.8)
Overall	ANXIETY	SE	1.1 (0.6; 1.9)
RA	ANXIETY	SE	1.3 (0.6; 3.1)
COPD	ANXIETY	SE	1.0 (0.4; 2.5)
AS	ANXIETY	SE	0.9 (0.3; 2.4)
Overall	ARRHYTHMIA	SE	1.6 (0.9; 2.9)
RA	ARRHYTHMIA	SE	1.4 (0.5; 3.8)
COPD	ARRHYTHMIA	SE	1.6 (0.7; 3.7)
AS	ARRHYTHMIA	SE	1.6 (0.6; 4.4)
Overall	**ARTHRALGIA**	**SE**	**2.7 (1.6; 4.5)**
RA	ARTHRALGIA	SE	1.8 (0.9; 3.6)
COPD	ARTHRALGIA	SE	1.2 (0.5; 3.0)
AS	ARTHRALGIA	SE	0.6 (0.2; 2.0)
Overall	**ASTHMA**	**BOTH**	**4.0 (2.7; 6.0)**
RA	**ASTHMA**	**BOTH**	**1.9 (1.0; 3.7)**
COPD	ASTHMA	BOTH	1.1 (0.5; 2.2)
AS	ASTHMA	BOTH	-
Overall	**ASTHMA ALLERGIC**	**BOTH**	**9.9 (5.6; 17.4)**
RA	**ASTHMA ALLERGIC**	**BOTH**	**4.7 (2.1; 10.3)**
COPD	ASTHMA ALLERGIC	BOTH	1.5 (0.7; 3.5)
AS	**ASTHMA ALLERGIC**	**BOTH**	**5.2 (2.8; 9.5)**
Overall	ATHEROSCLEROSIS	UNK	1.1 (0.1; 14.0)

RA	ATHEROSCLEROSIS	UNK	-
COPD	ATHEROSCLEROSIS	UNK	-
AS	ATHEROSCLEROSIS	UNK	-
Overall	ATRIAL FIBRILLATION	UNK	0.6 (0.1; 3.2)
RA	ATRIAL FIBRILLATION	UNK	-
COPD	ATRIAL FIBRILLATION	UNK	-
AS	ATRIAL FIBRILLATION	UNK	-
Overall	ATTENTION DEFICIT DISORDER WITH HYPERACTIVITY	UNK	0.6 (0.1; 3.2)
RA	ATTENTION DEFICIT DISORDER WITH HYPERACTIVITY	UNK	8.4 (0.2; 327.2)
COPD	**ATTENTION DEFICIT DISORDER WITH HYPERACTIVITY**	**UNK**	**10.3 (1.1; 98.0)**
AS	ATTENTION DEFICIT DISORDER WITH HYPERACTIVITY	UNK	2.2 (0.5; 9.9)
Overall	BACTERIAL INFECTION	SE	1.0 (0.6; 1.5)
RA	BACTERIAL INFECTION	SE	0.9 (0.4; 2.0)
COPD	BACTERIAL INFECTION	SE	1.0 (0.4; 2.2)
AS	BACTERIAL INFECTION	SE	0.5 (0.2; 1.2)
Overall	BIPOLAR DISORDER	UNK	0.5 (0.0; 4.3)
RA	BIPOLAR DISORDER	UNK	-
COPD	BIPOLAR DISORDER	UNK	-
AS	BIPOLAR DISORDER	UNK	-
Overall	BLURRED VISION	SE	1.1 (0.3; 3.9)
RA	BLURRED VISION	SE	1.2 (0.3; 5.5)
COPD	BLURRED VISION	SE	-
AS	BLURRED VISION	SE	-
Overall	BREAST NEOPLASMS	UNK	2.1 (0.8; 5.1)
RA	BREAST NEOPLASMS	UNK	2.5 (0.8; 8.3)
COPD	BREAST NEOPLASMS	UNK	1.6 (0.4; 7.5)

AS	**BREAST NEOPLASMS**	**UNK**	**5.3 (1.3; 20.7)**
Overall	CARDIOMYOPATHY	SE	1.1 (0.5; 2.1)
RA	CARDIOMYOPATHY	SE	0.3 (0.1; 1.2)
COPD	CARDIOMYOPATHY	SE	0.9 (0.4; 2.4)
AS	CARDIOMYOPATHY	SE	0.6 (0.2; 1.9)
Overall	**CHRONIC OBSTRUCTIVE PULMONARY DISEASE**	**IND**	**4.3 (2.4; 7.6)**
RA	**CHRONIC OBSTRUCTIVE PULMONARY DISEASE**	**IND**	**4.5 (2.3; 8.8)**
COPD	CHRONIC OBSTRUCTIVE PULMONARY DISEASE	IND	-
AS	**CHRONIC OBSTRUCTIVE PULMONARY DISEASE**	**IND**	**3.3 (1.6; 6.8)**
Overall	COLORECTAL NEOPLASMS	UNK	0.8 (0.1; 4.9)
RA	COLORECTAL NEOPLASMS	UNK	-
COPD	COLORECTAL NEOPLASMS	UNK	-
AS	COLORECTAL NEOPLASMS	UNK	-
Overall	CONJUNCTIVITIS	IND	3.7 (0.3; 42.9)
RA	CONJUNCTIVITIS	IND	-
COPD	CONJUNCTIVITIS	IND	-
AS	CONJUNCTIVITIS	IND	-
Overall	CONSTIPATION	SE	3.9 (0.9; 17.7)
RA	CONSTIPATION	SE	1.7 (0.2; 14.6)
COPD	CONSTIPATION	SE	2.9 (0.6; 14.3)
AS	CONSTIPATION	SE	2.9 (0.6; 13.3)
Overall	**COUGH**	**UNK**	**15.5 (3.5; 69.1)**
RA	COUGH	UNK	1.8 (0.2; 15.5)
COPD	**COUGH**	**UNK**	**5.4 (1.0; 29.3)**
AS	**COUGH**	**UNK**	**6.4 (2.1; 19.2)**
Overall	CROHN DISEASE	BOTH	6.9 (0.7; 64.3)

RA	CROHN DISEASE	BOTH	13.0 (0.8; 207.1)
COPD	CROHN DISEASE	BOTH	-
AS	CROHN DISEASE	BOTH	-
Overall	**DEPRESSION**	**SE**	**1.8 (1.1; 2.9)**
RA	DEPRESSION	SE	1.5 (0.7; 3.2)
COPD	DEPRESSION	SE	1.8 (0.9; 3.9)
AS	**DEPRESSION**	**SE**	**2.6 (1.3; 5.3)**
Overall	DERMATITIS CONTACT SUSCEPTIBILITY	BOTH	7.4 (0.9; 59.2)
RA	DERMATITIS CONTACT SUSCEPTIBILITY	BOTH	-
COPD	DERMATITIS CONTACT SUSCEPTIBILITY	BOTH	-
AS	**DERMATITIS CONTACT SUSCEPTIBILITY**	**BOTH**	**8.5 (1.0; 72.7)**
Overall	DIABETES TYPE I	SE	0.8 (0.4; 1.8)
RA	DIABETES TYPE I	SE	1.1 (0.4; 3.0)
COPD	DIABETES TYPE I	SE	0.6 (0.2; 2.0)
AS	DIABETES TYPE I	SE	0.8 (0.2; 2.8)
Overall	DIABETES TYPE II	SE	0.7 (0.4; 1.2)
RA	DIABETES TYPE II	SE	1.5 (0.7; 2.9)
COPD	DIABETES TYPE II	SE	0.6 (0.2; 1.6)
AS	DIABETES TYPE II	SE	1.2 (0.5; 2.9)
Overall	DIABETIC NEPHROPATHY	UNK	0.2 (0.0; 4.9)
RA	DIABETIC NEPHROPATHY	UNK	-
COPD	DIABETIC NEPHROPATHY	UNK	1.8 (0.2; 16.6)
AS	DIABETIC NEPHROPATHY	UNK	-
Overall	DIABETIC NEUROPATHIES	UNK	2.1 (0.3; 14.9)
RA	DIABETIC NEUROPATHIES	UNK	1.3 (0.2; 10.2)
COPD	DIABETIC NEUROPATHIES	UNK	1.1 (0.1; 10.3)
AS	DIABETIC NEUROPATHIES	UNK	2.1 (0.3; 18.1)
Overall	DIABETIC RETINOPATHY	UNK	0.7 (0.3; 1.8)

RA	DIABETIC RETINOPATHY	UNK	0.8 (0.2; 2.7)
COPD	DIABETIC RETINOPATHY	UNK	0.7 (0.2; 3.4)
AS	DIABETIC RETINOPATHY	UNK	0.4 (0.1; 3.0)
Overall	**DIZZINESS**	**SE**	**5.9 (1.2; 28.8)**
RA	DIZZINESS	SE	3.1 (0.6; 15.9)
COPD	**DIZZINESS**	**SE**	**6.4 (1.5; 27.6)**
AS	**DIZZINESS**	**SE**	**5.4 (1.1; 26.1)**
Overall	DYSPEPSIA	SE	1.7 (0.9; 3.4)
RA	DYSPEPSIA	SE	1.5 (0.5; 4.1)
COPD	DYSPEPSIA	SE	1.2 (0.4; 3.5)
AS	**DYSPEPSIA**	**SE**	**3.2 (1.3; 7.7)**
Overall	**DYSPNEA**	**UNK**	**8.1 (2.5; 26.7)**
RA	**DYSPNEA**	**UNK**	**7.5 (2.3; 23.8)**
COPD	DYSPNEA	UNK	2.1 (0.9; 5.2)
AS	DYSPNEA	UNK	2.5 (0.8; 7.3)
Overall	**EDEMA**	**BOTH**	**2.2 (1.1; 4.3)**
RA	EDEMA	BOTH	1.4 (0.6; 3.6)
COPD	**EDEMA**	**BOTH**	**2.9 (1.3; 6.4)**
AS	EDEMA	BOTH	1.5 (0.5; 4.4)
Overall	FEVER	UNK	0.0 (0.0; 0.0)
RA	FEVER	UNK	-
COPD	FEVER	UNK	-
AS	FEVER	UNK	-
Overall	**FIBROMYALGIA**	**UNK**	**4.8 (1.4; 16.8)**
RA	FIBROMYALGIA	UNK	2.3 (0.6; 8.3)
COPD	FIBROMYALGIA	UNK	1.0 (0.1; 7.9)
AS	FIBROMYALGIA	UNK	2.2 (0.5; 10.0)
Overall	**GASTROESOPHAGEAL REFLUX DISEASE**	**SE**	**1.6 (1.0; 2.7)**
RA	GASTROESOPHAGEAL REFLUX DISEASE	SE	1.1 (0.5; 2.3)

COPD	GASTROESOPHAGEAL REFLUX DISEASE	SE	1.0 (0.4; 2.3)
AS	GASTROESOPHAGEAL REFLUX DISEASE	SE	1.5 (0.7; 3.3)
Overall	**GLAUCOMA**	**SE**	**3.8 (1.4; 10.4)**
RA	**GLAUCOMA**	**SE**	**3.6 (1.1; 11.8)**
COPD	**GLAUCOMA**	**SE**	**3.4 (1.0; 12.0)**
AS	GLAUCOMA	SE	1.5 (0.2; 12.4)
Overall	**GOUT**	**BOTH**	**1.9 (1.1; 3.4)**
RA	GOUT	BOTH	1.3 (0.5; 3.0)
COPD	GOUT	BOTH	1.1 (0.4; 2.9)
AS	GOUT	BOTH	1.7 (0.6; 4.6)
Overall	HEADACHE	SE	2.3 (0.9; 6.1)
RA	HEADACHE	SE	2.5 (0.8; 7.8)
COPD	HEADACHE	SE	1.6 (0.3; 8.4)
AS	HEADACHE	SE	2.0 (0.6; 6.6)
Overall	HEART FAILURE	SE	1.4 (0.5; 3.9)
RA	HEART FAILURE	SE	0.9 (0.2; 3.8)
COPD	HEART FAILURE	SE	0.9 (0.3; 3.3)
AS	HEART FAILURE	SE	0.9 (0.1; 7.2)
Overall	HEPATOCELLULAR CARCINOMA	UNK	0.5 (0.0; 15.9)
RA	HEPATOCELLULAR CARCINOMA	UNK	6.4 (0.3; 133.4)
COPD	HEPATOCELLULAR CARCINOMA	UNK	-
AS	HEPATOCELLULAR CARCINOMA	UNK	-
Overall	HYPERCALCEMIA	BOTH	1.4 (0.2; 10.9)
RA	HYPERCALCEMIA	BOTH	2.5 (0.2; 26.7)
COPD	HYPERCALCEMIA	BOTH	-
AS	HYPERCALCEMIA	BOTH	-
Overall	HYPERCHOLESTEROLEMIA	SE	1.1 (0.8; 1.5)

RA	HYPERCHOLESTEROLEMIA	SE	1.0 (0.6; 1.8)
COPD	HYPERCHOLESTEROLEMIA	SE	0.9 (0.5; 1.8)
AS	HYPERCHOLESTEROLEMIA	SE	1.6 (0.8; 3.3)
Overall	HYPERGLYCEMIA	UNK	1.0 (0.5; 1.8)
RA	HYPERGLYCEMIA	UNK	1.3 (0.5; 3.4)
COPD	HYPERGLYCEMIA	UNK	1.0 (0.3; 3.1)
AS	HYPERGLYCEMIA	UNK	0.9 (0.3; 3.0)
Overall	HYPERNANTREMIA	UNK	4.5 (0.5; 37.2)
RA	HYPERNANTREMIA	UNK	-
COPD	**HYPERNANTREMIA**	**UNK**	**22.7 (1.0; 505.8)**
AS	**HYPERNANTREMIA**	**UNK**	**18.0 (1.6; 206.3)**
Overall	**HYPERTENSION**	**SE**	**1.5 (1.0; 2.1)**
RA	HYPERTENSION	SE	0.8 (0.5; 1.5)
COPD	HYPERTENSION	SE	1.2 (0.6; 2.4)
AS	HYPERTENSION	SE	1.1 (0.5; 2.2)
Overall	HYPERTRIGLYCERIDEMIA	SE	1.0 (0.4; 2.0)
RA	HYPERTRIGLYCERIDEMIA	SE	2.8 (0.6; 12.4)
COPD	HYPERTRIGLYCERIDEMIA	SE	-
AS	HYPERTRIGLYCERIDEMIA	SE	0.8 (0.1; 4.8)
Overall	HYPERURICEMIA	UNK	1.2 (0.7; 2.0)
RA	HYPERURICEMIA	UNK	0.7 (0.3; 2.0)
COPD	HYPERURICEMIA	UNK	0.9 (0.3; 2.6)
AS	HYPERURICEMIA	UNK	1.2 (0.4; 3.1)
Overall	HYPOCALCEMIA	UNK	2.7 (0.9; 8.4)
RA	HYPOCALCEMIA	UNK	0.7 (0.1; 5.9)
COPD	HYPOCALCEMIA	UNK	3.1 (0.7; 13.8)
AS	HYPOCALCEMIA	UNK	2.3 (0.5; 10.6)
Overall	HYPOCALEMIA	UNK	1.0 (0.4; 2.4)
RA	HYPOCALEMIA	UNK	0.5 (0.1; 4.2)
COPD	HYPOCALEMIA	UNK	2.5 (0.5; 12.9)
AS	HYPOCALEMIA	UNK	1.4 (0.3; 6.2)

Overall	HYPONANTREMIA	UNK	0.2 (0.0; 2.0)
RA	HYPONANTREMIA	UNK	-
COPD	HYPONANTREMIA	UNK	-
AS	HYPONANTREMIA	UNK	-
Overall	HYPOPHOSPHATEMIA	UNK	2.7 (0.6; 13.0)
RA	HYPOPHOSPHATEMIA	UNK	3.9 (0.4; 40.0)
COPD	HYPOPHOSPHATEMIA	UNK	3.5 (0.4; 35.2)
AS	**HYPOPHOSPHATEMIA**	**UNK**	**5.5 (1.1; 27.1)**
Overall	INSOMNIA	SE	1.6 (0.8; 3.3)
RA	INSOMNIA	SE	1.6 (0.6; 4.2)
COPD	INSOMNIA	SE	0.6 (0.1; 2.4)
AS	INSOMNIA	SE	1.3 (0.4; 4.3)
Overall	LUNG NEOPLASMS	UNK	0.5 (0.0; 9.7)
RA	LUNG NEOPLASMS	UNK	-
COPD	LUNG NEOPLASMS	UNK	0.5 (0.1; 4.2)
AS	LUNG NEOPLASMS	UNK	-
Overall	**LUPUS ERYTHEMATOSUS SYSTEMIC**	**BOTH**	**18.7 (2.7; 131.5)**
RA	**LUPUS ERYTHEMATOSUS SYSTEMIC**	**BOTH**	**12.7 (3.5; 46.2)**
COPD	LUPUS ERYTHEMATOSUS SYSTEMIC	BOTH	2.9 (0.1; 58.1)
AS	**LUPUS ERYTHEMATOSUS SYSTEMIC**	**BOTH**	**42.2 (4.8; 370.6)**
Overall	MELANOMA	UNK	0.7 (0.1; 3.9)
RA	MELANOMA	UNK	-
COPD	MELANOMA	UNK	-
AS	MELANOMA	UNK	-
Overall	MENOPAUSE	UNK	0.8 (0.1; 9.0)
RA	MENOPAUSE	UNK	-
COPD	MENOPAUSE	UNK	-
AS	MENOPAUSE	UNK	-

Overall	**MUSCLE SPASMS**	**UNK**	**3.3 (1.3; 8.5)**
RA	MUSCLE SPASMS	UNK	1.5 (0.5; 4.5)
COPD	**MUSCLE SPASMS**	**UNK**	**3.2 (1.0; 10.6)**
AS	MUSCLE SPASMS	UNK	1.4 (0.3; 5.9)
Overall	**MYALGIA**	**SE**	**3.3 (1.4; 7.5)**
RA	MYALGIA	SE	2.2 (0.9; 5.5)
COPD	MYALGIA	SE	1.1 (0.2; 5.4)
AS	MYALGIA	SE	1.5 (0.4; 6.6)
Overall	**MYOCARDIAL INFARCTION**	**SE**	**2.5 (1.4; 4.4)**
RA	MYOCARDIAL INFARCTION	SE	1.2 (0.5; 2.7)
COPD	MYOCARDIAL INFARCTION	SE	1.5 (0.7; 3.2)
AS	MYOCARDIAL INFARCTION	SE	1.6 (0.6; 4.4)
Overall	NASOPHARYNGITIS	UNK	2.1 (0.1; 33.9)
RA	NASOPHARYNGITIS	UNK	-
COPD	NASOPHARYNGITIS	UNK	-
AS	NASOPHARYNGITIS	UNK	-
Overall	**NAUSEA**	**SE**	**4.2 (1.0; 18.1)**
RA	NAUSEA	SE	0.9 (0.1; 7.5)
COPD	NAUSEA	SE	-
AS	NAUSEA	SE	-
Overall	**NEUROPATHIC PAIN**	**SE**	**2.4 (1.2; 5.0)**
RA	NEUROPATHIC PAIN	SE	1.8 (0.7; 4.5)
COPD	NEUROPATHIC PAIN	SE	2.1 (0.9; 5.0)
AS	**NEUROPATHIC PAIN**	**SE**	**3.2 (1.3; 7.7)**
Overall	OSTEOARTHRITIS	BOTH	1.0 (0.7; 1.6)
RA	OSTEOARTHRITIS	BOTH	-
COPD	OSTEOARTHRITIS	BOTH	0.6 (0.2; 1.5)
AS	OSTEOARTHRITIS	BOTH	0.6 (0.2; 1.4)
Overall	OSTEOPOROSIS	SE	2.2 (0.2; 24.2)
RA	OSTEOPOROSIS	SE	-

COPD	OSTEOPOROSIS	SE	6.1 (0.7; 50.4)
AS	**OSTEOPOROSIS**	**SE**	**12.1 (1.2; 125.7)**
Overall	OSTEOSARCOMA	UNK	3.2 (0.3; 40.2)
RA	OSTEOSARCOMA	UNK	-
COPD	OSTEOSARCOMA	UNK	-
AS	OSTEOSARCOMA	UNK	-
Overall	**PAIN**	**SE**	**4.4 (2.1; 9.1)**
RA	**PAIN**	**SE**	**3.0 (1.2; 7.1)**
COPD	**PAIN**	**SE**	**3.9 (1.7; 8.7)**
AS	PAIN	SE	0.9 (0.2; 3.7)
Overall	PANCREATIC NEOPLASMS	UNK	1.9 (0.1; 33.1)
RA	PANCREATIC NEOPLASMS	UNK	-
COPD	PANCREATIC NEOPLASMS	UNK	-
AS	PANCREATIC NEOPLASMS	UNK	-
Overall	**PHARYNGITIS**	**UNK**	**6.8 (1.0; 48.0)**
RA	PHARYNGITIS	UNK	-
COPD	PHARYNGITIS	UNK	-
AS	PHARYNGITIS	UNK	-
Overall	PHOTOSENSITIVITY	UNK	1.0 (0.4; 2.2)
RA	PHOTOSENSITIVITY	UNK	-
COPD	PHOTOSENSITIVITY	UNK	0.3 (0.0; 2.5)
AS	PHOTOSENSITIVITY	UNK	1.3 (0.3; 4.8)
Overall	PROSTATIC NEOPLASMS	IND	1.8 (0.6; 5.0)
RA	PROSTATIC NEOPLASMS	IND	3.2 (0.6; 17.9)
COPD	PROSTATIC NEOPLASMS	IND	1.4 (0.4; 5.4)
AS	PROSTATIC NEOPLASMS	IND	4.6 (0.8; 27.9)
Overall	**PRURITUS**	**SE**	**7.6 (1.1; 55.2)**
RA	PRURITUS	SE	-
COPD	PRURITUS	SE	13.8 (0.8; 247.3)
AS	PRURITUS	SE	7.9 (0.8; 76.7)
Overall	PSORIASIS	BOTH	4.8 (0.7; 33.5)

RA	**PSORIASIS**	**BOTH**	**7.7 (1.3; 46.5)**
COPD	PSORIASIS	BOTH	-
AS	PSORIASIS	BOTH	-
Overall	RENAL CELL CARCINOMA	UNK	1.7 (0.2; 16.0)
RA	RENAL CELL CARCINOMA	UNK	-
COPD	RENAL CELL CARCINOMA	UNK	-
AS	RENAL CELL CARCINOMA	UNK	-
Overall	**RHEUMATOID ARTHRITIS**	**BOTH**	**5.0 (3.0; 8.4)**
RA	RHEUMATOID ARTHRITIS	BOTH	-
COPD	**RHEUMATOID ARTHRITIS**	**BOTH**	**5.2 (2.4; 11.1)**
AS	**RHEUMATOID ARTHRITIS**	**BOTH**	**5.0 (2.4; 10.6)**
Overall	**RHINITIS ALLERGIC**	**IND**	**3.8 (2.0; 7.2)**
RA	RHINITIS ALLERGIC	IND	1.8 (0.7; 4.9)
COPD	RHINITIS ALLERGIC	IND	1.2 (0.5; 2.9)
AS	RHINITIS ALLERGIC	IND	1.6 (0.9; 2.9)
Overall	SEIZURES	UNK	0.6 (0.1; 3.7)
RA	SEIZURES	UNK	2.5 (0.5; 13.0)
COPD	SEIZURES	UNK	2.7 (0.5; 14.3)
AS	SEIZURES	UNK	2.8 (0.6; 12.3)
Overall	**SINUSITIS**	**IND**	**14.5 (2.5; 84.2)**
RA	SINUSITIS	IND	-
COPD	SINUSITIS	IND	-
AS	SINUSITIS	IND	-
Overall	**SKIN ERUPTIONS**	**SE**	**9.8 (1.3; 73.9)**
RA	SKIN ERUPTIONS	SE	17.2 (0.9; 315.4)
COPD	SKIN ERUPTIONS	SE	-
AS	**SKIN ERUPTIONS**	**SE**	**32.5 (1.6; 642.8)**
Overall	STROKE	UNK	1.2 (0.6; 2.4)
RA	STROKE	UNK	1.4 (0.6; 3.3)

COPD	STROKE	UNK	1.4 (0.5; 3.5)
AS	STROKE	UNK	0.9 (0.3; 3.1)
Overall	THROMBOPHLEBITIS	SE	1.9 (0.3; 13.8)
RA	THROMBOPHLEBITIS	SE	3.4 (0.3; 34.4)
COPD	THROMBOPHLEBITIS	SE	-
AS	THROMBOPHLEBITIS	SE	-
Overall	THYROIDITIS	BOTH	1.2 (0.7; 1.9)
RA	THYROIDITIS	BOTH	0.9 (0.4; 2.0)
COPD	THYROIDITIS	BOTH	0.7 (0.3; 1.9)
AS	THYROIDITIS	BOTH	0.7 (0.2; 2.2)
Overall	URINARY INCONTINENCE	UNK	1.2 (0.2; 6.3)
RA	URINARY INCONTINENCE	UNK	1.8 (0.2; 14.8)
COPD	URINARY INCONTINENCE	UNK	-
AS	URINARY INCONTINENCE	UNK	-
Overall	URTICARIA	SE	1.9 (0.1; 31.7)
RA	URTICARIA	SE	-
COPD	URTICARIA	SE	-
AS	**URTICARIA**	**SE**	**33.0 (2.4; 453.3)**
Overall	UTERINE CERVICAL NEOPLASMS	UNK	3.0 (0.8; 10.7)
RA	UTERINE CERVICAL NEOPLASMS	UNK	0.7 (0.1; 5.4)
COPD	UTERINE CERVICAL NEOPLASMS	UNK	1.7 (0.2; 14.3)
AS	UTERINE CERVICAL NEOPLASMS	UNK	1.3 (0.2; 10.5)

Table 4.1: The table summarize the computed OR and the lower and upper bound computed in the four different population for each BED condition taken into account.

Of the 101 terms, only 85 have been used for the analysis. Of this 85, 5 were annotated as indication (IND), 31 as Side-effects (SE), 12 as Side-Effects or Indication (BOTH) and 37 as Unknown (UNK). Among the SE terms, the ones showing a positive and significant correlation with the use of prednisone in the Overall population are Pain, Arthralgia, Myalgia

and Neuropathic Pain for pain-related terms, followed by major conditions like Myocardial Infarction, Glaucoma, Depression and Hypertension. At last, other minor symptoms are present, Dizziness, Skin Eruptions, Pruritus and Nausea.

4.2.1. Known Side-effects

Pain, Arthralgia, neuropathic Pain

Pain shows a positive association in the Overall population with 4.4 (2.1; 9.1), the same correlation is also found in RA 3.0 (1.2; 7.1) and COPD 3.9 (1.7; 8.7), but not in AS 0.9 (0.2; 3.7). Arthralgia shows a similar pattern with 2.7 (1.6; 4.5) in the Overall population and 1.8 (0.9; 3.6) in RA and 1.2 (0.5; 3.0) in COPD while it is lower than 1 in AS. Myalgia is 3.3 (1.4; 7.5) in the Overall population and remains strong but not significant in RA 2.2 (0.9; 5.5) while is weaker and closer to 1 in COPD 1.1 (0.2; 5.4) and AS 1.5 (0.4; 6.6). Neuropathic pain is significant in the Overall population with 2.4 (1.2; 5.0), is still positive yet not significant in RA and COPD with 1.8 (0.7; 4.5) and 2.1 (0.9; 5.0), while is much stronger and significant in AS 3.2 (1.3; 7.7).

The term Pain, which includes individual taking any drugs for any kind of pain as described in Table 7.5 is particularly strong among RA and COPD patients, Arthralgia instead is stronger in RA while is closer to 1 or less in COPD and AS, similarly, Myalgia is stronger in RA and ASTHMA but weaker in COPD. Note that Myalgia and Arthralgia are particularly strong in RA with respect to the other conditions, these two terms are typical symptoms of RA [Walsh and McWilliams, 2012], and the prolonged use of prednisone could be a worsening factor in the context of rheumatoid arthritis. Prednisone results as a strong predictor of Neuropathic Pain in the context of asthma, some underlying mechanism explaining this observation could be hypothesized considering the risk of neuritis associated with the use of prednisone and the molecular evidence suggesting the presence of neurogenic inflammation in the context of asthma [Butler and Heaney, 2007, Joos et al., 2003].

Myocardial Infarction

Myocardial Infarction shows significant association in the Overall population with 2.5 (1.4; 4.4), while is not significant in the other populations with 1.2 (0.5; 2.7) for RA, 1.5 (0.7; 3.2) for COPD, and 1.6 (0.6; 4.4) in AS.

It must be highlighted that individuals described has having a Myocardial Infarction are those that reported at least one episode of infarction in their life, before or after they started use of prednisone.

Glaucoma, Depression, Hypertension

Glaucoma shows a significant association with prednisone in the RA population with 3.6 (1.1; 11.8), the same occurs in the COPD population 3.4 (1.0; 12.0) yet the lower bound is slightly

lower than one, the OR is 2 times smaller in the AS population respect to RA and COPD, with 1.5 (0.2; 12.4). Note that the OR for glaucoma in the Overall population is 3.8 (1.4; 10.4), similarly to what obtained in RA and COPD.

Depression shows Association in the Overall population with 1.8 (1.1; 2.9), while is not significant yet positive in RA 1.5 (0.7; 3.2) and COPD 1.8 (0.9; 3.9), though it shows significant association in AS patients with 2.6 (1.3; 5.3).

Hypertension show significant Association in the Overall population with 1.5 (1.0; 2.1) yet its lower bound is very close to one, correlation is not significant instead in none of the 3 population and always very close to one with 0.8 (0.5; 1.5) in RA, 1.2 (0.6; 2.4) in COPD and 1.1 (0.5; 2.2) in AS.

For what concern Glaucoma, Depression and Hypertension, significant association is observed with different entity across the population taken into account, the differences observed could be related to frequency, dose and path of administration which vary across the different population taken into account, on the other hand a quantification of the association between prednisone and the insurgence of this conditions is provided.

Nausea, Pruritus, Skin eruptions and Dizziness

Some minor conditions show significant association with the use prednisone, in particular Dizziness, which is significant in the Overall population with 5.9 (1.2; 28.8) and in COPD and AS with 6.4 (1.5; 27.6) and 5.4 (1.1; 26.1) while is two-fold smaller and not significant in RA 3.1 (0.6; 15.9). Skin Eruptions has an OR of 9.8 (1.3; 73.9) in the Overall population and appears also in AS with 32.5 (1.6; 642.8) and RA 17.2 (0.9; 315.4). Pruritus with 7.6 (1.1; 55.2) in the Overall population only appears in COPD and AS with 13.8 (0.8; 247.3) and 7.9 (0.8; 76.7). At last, Nausea with 4.2 (1.0; 18.1) in the Overall Population. The relation of prednisone use with the insurgence conditions of minor entities (Nausea, Pruritus, skin eruptions, and dizziness) is quantified, also in this case some variation is observed across the different populations.

Dyspepsia, Osteoporosis, Urticaria

Note that the only side-effects that have a significant association with prednisone in either the populations RA, COPD, or AS but are not significantly associated with prednisone in the Overall population only occurs in the AS population. AS shows a positive and significant association with Dyspepsia 3.2 (1.3; 7.7) Osteoporosis 12.1 (1.2; 125.7) and Urticaria 33.0 (2.4; 453.3). Note that the confidence intervals are extremely wide for both Osteoporosis and Urticaria, this is due to the very small number of individuals showing these conditions, also note that results for Urticaria are absent in COPD and RA while Osteoporosis is absent only in RA due to the absence of individuals showing these conditions in these populations. For what concerns Dyspepsia, note that the OR in the AS population is nearly 2 times the one

Overalll population with 3.2 (1.3; 7.7) against 1.7 (0.9; 3.4), a weaker yet positive relation is observed for this condition also in RA with 1.5 (0.5; 4.1) and COPD with 1.2 (0.4; 3.5). Dyspepsia, Osteoporosis, and Urticaria all show positive association with prednisone in the AS population. While for Osteoporosis and Urticaria the confidence intervals are extremely wide, these are not for dyspepsia. Dyspepsia is a frequent symptom in patients with asthma, [Tomyo et al., 2019] reported that 44% of patients with asthma show functional dyspepsia, and different authors describe a relationship of asthma with Gastroesophageal Reflux Disease [Boos et al., 2014, Liang et al., 2013, Tsai et al., 2010]. An association between the use of glucocorticoid and the insurgence of Dyspepsia has been reported by [Hallas and Bytzer, 1998] for a relative risk (RR 1.1, CI 1.0-1.3). Our results suggest a strong presence of Dyspepsia among AS patients using prednisone.

4.2.2. Potential Side-effects

Of the 38 conditions marked as UNK only Cough 15.5 (3.5; 69.1), Dyspnea 8.1 (2.5; 26.7), Fibromyalgia 4.8 (1.4; 16.8), and Muscle Spasms 3.3 (1.3; 8.5) show a significant association with the use of prednisone in the Overall population. Of these four conditions, only Dyspnea and Cough show associations also in the other populations, in particular Dyspnea shows OR of 7.5 (2.3; 23.8) in RA while Cough shows an OR of 6.4 (2.1; 19.2) in AS. Among the potential side effects, some don't show association in the Overall population while some show association in either COPD or AS, but not in RA. In AS we observe Breast Neoplasm 5.3 (1.3; 20.7), Hypophosphatemia 5.5 (1.1; 27.1) and Hypernatremia 18.0 (1.6; 206.3), this latter is also present in COPD with 22.7 (1.0; 505.8). In COPD we also observe Attention Deficit Disorder with Hyperactivity 10.3 (1.1; 98.0).

Attention Deficit Disorder with Hyperactivity (ADHD)

Attention Deficit Disorder with Hyperactivity shows a strong association with the use of prednisone in COPD patients. A common mechanism explaining the relation between ADHD and prednisone could be hypothesized considering that both have a relation with the HPA-axis dysregulation [Vogel et al., 2017, Alten and Wiebe, 2015]. On the other hand, [Vogel et al., 2017] suggests that the relations between ADHD and the HPA axis are mostly due to anxiety disorder and depression. Note that Anxiety and depression are comorbidities both for COPD [Panagioti et al., 2014] and for Attention Deficit Disorder with Hyperactivity [Kittel-Schneider and Reif, 2020], prednisone could be among the factors contributing to these mental conditions but further studies are needed.

Dyspnea and Cough

Dyspnea appears among the UNK terms because is not annotated as an indication in drug-bank, yet prednisone is widely used in the treatment of dyspnea when caused by Chronic Obstructive Pulmonary Disease [Woods et al., 2014], association with prednisone and dyspnea is particularly high in the RA population. Of the 5 patients showing both RA and

Dyspnea and using prednisone 1 is using it for unspecified allergies, 2 for dyspnea or breathing, and 2 for chronic obstructive pulmonary disease, so generally for respiratory problems. These results can be attributed to the lung involvement of rheumatoid arthritis [Chatzidiontsiou, 2016] and the use of prednisone for their management, rather than considering it a side-effect of prednisone. A similar situation is obtained for Cough, it is not annotated as an indication for prednisone in drugbank but prednisone is actually used for cough treatment when caused by Chronic Obstructive Pulmonary Disease [Thompson et al., 1996]. The use of prednisone for Cough treatment seems to be more frequent in the AS population, the strong association could be due to the occurrence of cough as asthmatic symptoms.

Fibromyalgia

Prednisone appears to be a predictor of Fibromyalgia in the Overall population, the association is positive but not significant in the RA and AS population while equal to 1 in the COPD population. Different papers discussed already the relationship between the use of prednisone and the occurrence of fibromyalgia. A study conducted in RA patients shows that individuals with strong fibromyalgia symptoms were more likely to use prednisone than those with weak symptoms (Odds Ratio 4.99 [95% Confidence Interval 1.20-20.73]) [Szumilas, 2010]. A prospective cohort study [Chakr et al., 2017] investigates the use of DMARD in RA patients and its relation with Fibromyalgia, concluding that RA patients with FM used more leflunomide and prednisone. Another study investigates the molecular aspect of Fibromyalgia, particularly focusing on Glucocorticoid (GR) and Mineral-corticoid receptor (MR) involvement, stating that lower levels of GR and MR are present in Fibromyalgia patients and that this under-expression could be related to a malfunctioning of the hypothalamic–pituitary–adrenal axis (HPA) and to the presence of low levels of anti-inflammatory mediator. Considering that the term fibromyalgia is defined by drug indications, our results suggest that people using prednisone are more likely to take drugs to treat fibromyalgia.

Muscle Spasms

Prednisone is associated with muscle spasms, a positive and almost significant association is observed among COPD patients. Myopathies are a known side effect of prednisone [Schweiger and Zdanowicz, 2010], and muscle spasms is a manifestation of it.

Hypernatremia and Hypophosphatemia

Hypernatremia shows positive association with prednisone in both AS and COPD patients. This result is in line with different other studies. A nested case-control study [Imaizumi et al., 2021] identify high dosage glucocorticoid has a predictor of IAH (Intensive-care-unit Acquired Hypernatremia) (odds ratio (OR), 4.15 [95% confidence interval (CI) 1.29–13.4]) while a meta-analysis of 37 reports a relative risk 1.57 (95% CI, 1.24-1.99) for hypernatremia due to use of corticosteroids. Our results are in line with these evidences yet OR estimated by our method shows a very wide confidence interval due to the small number of samples showing hypernatremia. Hypophosphatemia shows significant association with the use of prednisone in

AS patients. Hypophosphatemia has been identified as a complication in Asthmatic patients as a consequence of urinary excretion of phosphate due to the use of glucocorticoid [Braddy et al., 1989]. The relation between the use of corticosteroids and phosphate levels is poorly described in literature.

Breast Neoplasm

A significant association between Breast Neoplasm and prednisone is observed in AS patients. This association appears because prednisone is often used together with chemotherapies in the treatment breast neoplasm [Marini et al., 1996], so it can't be considered a side effect of prednisone.

4.2.3. System Biology Analysis

System biology analysis has been conducted for those associations that we found interesting and that show incomplete understanding of the molecular mechanisms underlying it. Analysis have been performed whether the condition in exam is a known side-effect or not. The results of this analysis are shown in Table 4.2 and Table 4.3, respectively for the identified side effects in the AS and COPD population. The analysis on the RA population didn't produce any significant results.

IN	**Dyspepsia** gallbladder alterations	**Hypophosphatemia** excessive secretion of parathyroid hormone (PTH)	**Hypophosphatemia** reduced bone resorption	**Neuropathic pain** pronociceptive facilitation at the spinal dorsal horn
Prednisone Receptor	*	*	**	*
P22301	-	-	*	***
P17931	-	-	*	***
P05231	-	-	*	***
Q16620	-	-	*	***
P35367	-	-	-	***
P23560	-	-	-	***
P13501	-	-	-	***
P13500	-	-	-	***
P09341	-	-	-	***
Q99731	-	-	-	***
P02778	-	-	-	***
O95998	-	-	-	***
O00175	-	-	-	***
Q14116	-	-	-	***
P80075	-	-	-	***
P51671	-	-	-	***
P42830	-	-	-	***
Q03135	*	-	*	*
P21731	*	-	-	*
P10721	-	-	-	*
P10145	-	-	-	*
P09429	-	-	-	*
P08637	-	-	-	*
P01584	-	-	-	*

P01583	-	-	-	*
P01133	-	-	-	*
O75015	-	-	-	*
Q99075	-	-	-	*
O00206	-	-	-	*
Q92769	*	-	-	*
Q16665	-	*	-	*
Q15109	-	-	-	*
Q13557	-	-	-	*
Q9NPH3	-	-	-	*
P42226	*	-	-	*
P09603	-	-	***	-
P35225	-	-	*	-
P29966	-	-	*	-
P29460	-	-	*	-
P29459	-	-	*	-
P27930	-	-	*	-
P21980	-	-	*	-
P20701	-	-	*	-
P15692	-	-	*	-
P08700	-	-	*	-
P05113	-	-	*	-
P05112	-	-	*	-
P04141	-	-	*	-
P01579	-	-	*	-
O60602	-	-	*	-
Q969D9	-	-	*	-
P47712	-	-	*	-
P43699	-	-	*	-
P14780	*	-	-	-
Q9BQ51	-	*	-	-
P05362	-	*	-	-
P01574	*	-	-	-
P01375	*	-	-	-
P00533	*	*	-	-
O95238	-	*	-	-
Q9Y261	-	*	-	-
Q15796	*	-	-	-
Q13478	*	-	-	-
P84022	*	*	-	-
P37231	*	-	-	-

Table 4.2: Results of the system biology analysis for relevant associations found in the AS population. ***: p-value $<$ 0.05, **: p-value $<$ 0.1, *:p-value $<$ 0.2 and -: p.value $>$ 0.2.

IN	**ADHD** Circadian System Imbalance	**ADHD** Neurotrasmitter imbalance
Prednisone	*	*
P02794	-	***
P02792	-	***
P48200	-	***
P21731	-	**
P13612	-	*
P05112	-	*
O94817	-	*
O15264	*	*
Q6NYC1	-	*
Q16539	*	*
Q15759	*	*
Q14790	-	*
P55211	-	*
P53778	*	*
P43115	-	*
P42574	-	*
P35225	-	*
Q9HBY0	-	*
P19838	*	-
P10144	*	-
Q96EB6	*	-
P05412	*	-
Q92769	*	-
Q14145	*	-
Q03405	*	-
P35968	*	-
P28482	*	-

Table 4.3: Results of the system biology analysis for relevant associations found in the COPD population. ***: p-value $<$ 0.05, **: p-value $<$ 0.1, *:p-value $<$ 0.2 and -: p.value $>$ 0.2.

Considering the functional association between the prednisone receptor, the motives that have been identified as functionally related are CIRCADIAN SYSTEM IMBALANCE and NEUROTRANSMITTER IMBALANCE for ADHD, with p-value $<$ 0.2 for both. In Dyspepsia the only significantly related module is GALL BLADDER ALTERATIONS with p-value $<$ 0.2. In hypophosphatemia, REDUCED BONE RESORPTION shows p-value $<$ 0.1 while EXCESSIVE SECRETION OF PARATHYROID HORMONE (PTH) shows p-value $<$0.2. For neuropathic pain, PRONOCICEPTIVE FACILITATION AT THE SPINAL DORSAL HORN is the most significant one with p-value $<$ 0.1.

We evaluated the molecular relation between the use of prednisone and (1) neuropathic pain in asthmatic individuals, (2) dyspepsia in asthmatic individuals, (3) hypophosphatemia in asthmatic individuals, and (4) ADHD in COPD.

For AS, COPD and RA populations, COPD proteins were identified to be functionally associated with the ADHD modules CIRCADIAN SYSTEM IMBALANCE (13 proteins with p < 0.2) and NEUROTRANSMITTER IMBALANCE (2 proteins with p-value < 0.05, 2 with p-value < 0.1 and 8 with p-value <0.2).

Functional relation between asthma protein and modules of Dyspepsia, Neuropathic Pain, and Hypophosphatemia has been identified. In particular 12 asthma proteins were identified to be functionally related to the dyspepsia motive GALLBLADDER ALTERATION (all with p-value <0.2), 7 proteins relate to the hypophosphatemia motive of EXCESSIVE SECRETION OF PARATHYROID HORMONE (PTH) (all with p-value < 0.2) and 24 proteins relates to REDUCED BONE RESORPTION (23 with p-value < 0.2 and 1 with p-value < 0.05). At last. some proteins relates with Neuropathic Pain modules of PRONOCICEPTIVE FACILITATION AT THE SPINAL DORSAL HORN (17 with p-value <0.05 and 15 p-value < 0.1).

ADHD-NEUROTRASMITTER IMBALANCE in COPD

Among the COPD proteins which show functional relation with the "NEUROTRASMITTER IMBALANCE" motif of ADHD are the Ferritin Heavy Chain (Uniprot: P02794) and the Iron-responsive element-binding protein 2 (Uniprot: P48200) with p-value < 0.05, while the Thromboxane A2 receptor (Uniprot: P21731), and the Ferritin Light Chain (Uniprot: P02792) shows p-value < 0.1. Ferritin is a well-known biomarker in COPD [Hoepers et al., 2015], while the relation of ferritin and/or Fe levels with ADHD is supported by much evidence, though with some conflicting results [Robberecht et al., 2020, Oner et al., 2012]. We could hypothesize (1) ferritin miss-regulation as a common underlying mechanism between COPD and ADHD and (2) prednisone as a potential contributor to this misregulation, somehow these findings confirm the results reported in [Fasmer et al., 2011], where a correlation is observed between ADHD and the use of drugs for asthma treatment, among which prednisone. It's interesting to notice that the Thromboxane A2 receptor is part of a variety of inflammatory processes [Yan et al., 2017].

HYPOPHOSPATEMIA-REDUCED BONE ABSORPTION in Asthma

The only asthma protein that has been identified to have a strong functional relation with the hypophosphatemia module REDUCED BONE RESORPTION is the Macrophage colony-stimulating factor 1 (Uniprot: P09603). Considering the involvement of the protein in inflammatory processes and bone-remodeling, the importance of phosphorus in bone mineralization, and the worsening effect of corticosteroids on bone general health [Mitra, 2011, Picado and Luengo, 1996], we can hypothesize prednisone to be a cause hypophosphatemia in asthma patients in relation to molecular process involving bone-remodeling.

NEUROPATHIC PAIN in ASTHMA

The number of Asthma proteins identified as strongly related to the Neuropathic Pain module PRONOCICEPTIVE FACILITATION AT THE SPINAL DORSAL HORN is 17. We found the brain-derived neurotrophic factor BDNF and the BDNF/NT-3 growth factors receptor (Uniprot: P23560, Q16620), the C-C motifs chemokines 2, 5, 8, 19, and 24 (Uniprot: P13500, P13501, P80075, Q99731, O00175) and the C-X-C motif chemokines 5 and 10 (Uniprot: P42830, P022778) together with interleukins 6, 10, 18 (Uniprot: P05231, P22301, Q14116) and the interleukin-18-binding protein (Unirprot O95998) and at last, Eoxitin (Uniprot: P51671), Galectin-3 (Uniprot: P17931), Growth Related Alpha Protein (Uniprot: P09341) and Histamine H1 Receptor (Uniprot: P35367). The first interesting aspect of the results obtained is that 15 out of the 17 proteins identified are related to immune mechanisms, while the two remaining regard the Brain-Derived Neurotrophic factor and its receptor. Various yet conflicting evidence relate BDNF and its signaling pathway to the glucocorticoid receptor and so to the use of Glucocorticoids [Zhang et al., 2020] while others report a relation with asthma [Barrios and Ai, 2018]. The presence of an underlying mechanism for which prednisone could be a predictor of neuropathic pain and neuritis could be hypothesized taking in consideration the proteins identified by this analysis.

4.3. The effect of Tocilizumab in GCA-patiens PBMCS (Peripheral Blood Mononuclear Cells)

4.3.1. Differential Analysis

Significant results for multiple hypothesis testing were obtained only in the p-value TCZ+IL6 vs IL6 experiment, with 81 differentially expressed transcripts, of which 7 with a negative fold-change (FC), and 71 with a positive one. In the IL6 vs UNT experiment, no statistically significant protein is found when adjusting for multiple hypotheses. Without adjustment, we found 49 significant proteins, of which 34 with $FC < 0$ and 15 with $FC > 0$.

Considering instead non-adjusted p-values both in the TCZ+IL6 vs IL6 and IL6 vs UNT experiments, 41 proteins are significant in both experiments, and all of them show opposite FC. A subgroup of 7 (BCL6, C3AR1, CCL2, CCR1, HIF1A, IL10, MYC, STAT3) is activated by IL6 and inhibited by TCZ (named "negative_set"), is the opposite for the other 31 ("positive_set").

Interestingly, if we consider a p-value threshold of 0.1, we have significant p-values also in the IL6 vs UNT experiment, and a total of 6 protein significant in both experiments. 3 are upregulated by IL6 and inhibited by TCZ (MYC, HIF1A, STAT3) while 3 have the opposite behaviour (STAT1, NLRP3, TREM2).

The results of the differential analysis are shown in Table 4.4.

	TCZ+IL6 vs IL6			IL6 vs UNT			
Protein	p	fdr(p)	FC	p	fdr(p)	FC	SET
CSF1	**0.0003**	**0.0113**	U	**0.0168**	0.1229	D	POS
MAP2K4	**0.0003**	**0.0113**	U	0.2659	0.5087	D	
TNF	**0.0004**	**0.0113**	U	0.1626	0.4039	D	
MYC	**0.0004**	**0.0113**	D	**0.0007**	0.0572	U	NEG
CCL24	**0.0005**	**0.0113**	U	**0.0168**	0.1229	D	POS
IFIT2	**0.0005**	**0.0113**	U	0.1626	0.4039	D	
RIPK2	**0.0005**	**0.0113**	U	**0.0277**	0.1629	D	POS
IFNG	**0.0005**	**0.0113**	U	0.4348	0.659	D	
CCL4	**0.0005**	**0.0113**	U	0.3318	0.5712	U	
TNFSF14	**0.0006**	**0.0113**	U	0.1773	0.4189	D	
NFE2L2	**0.0006**	**0.0113**	U	0.1024	0.3527	D	
CSF2	**0.0006**	**0.0113**	U	0.8313	0.9025	U	
IFI44	**0.0007**	**0.0113**	U	**0.0129**	0.1156	D	POS
CYSLTR1	**0.0007**	**0.0113**	U	**0.0056**	0.1128	D	POS
NOD1	**0.0007**	**0.0113**	U	0.1239	0.3982	D	
HIF1A	**0.0008**	**0.0125**	D	**0.0007**	0.0572	U	NEG
TREM2	**0.0012**	**0.0128**	U	**0.0019**	0.0776	D	POS
NLRP3	**0.0012**	**0.0128**	U	**0.0016**	0.0776	D	POS
IFIT3	**0.0012**	**0.0128**	U	0.2868	0.5237	D	
HMGB2	**0.0012**	**0.0128**	U	**0.0065**	0.1156	D	POS
MEF2A	**0.0012**	**0.0128**	U	**0.0129**	0.1156	D	POS
LIMK1	**0.0014**	**0.0144**	U	**0.0036**	0.1008	D	POS
CDC42	**0.0016**	**0.015**	U	**0.0352**	0.2017	D	POS
IL15	**0.0016**	**0.015**	U	0.554	0.7377	D	
BCL6	**0.0016**	**0.015**	D	**0.0217**	0.1375	U	NEG

IRF5	**0.0019**	**0.0153**	U	**0.0113**	0.1156	D	POS
TGFBR1	**0.0019**	**0.0153**	U	**0.0113**	0.1156	D	POS
HLA-DRA	**0.0019**	**0.0153**	U	**0.0168**	0.1229	D	POS
IFIT1	**0.0023**	**0.0153**	U	**0.0129**	0.1156	D	POS
C3	**0.0023**	**0.0153**	U	**0.0168**	0.1229	D	POS
ITGB2	**0.0023**	**0.0153**	U	**0.0113**	0.1156	D	POS
PTK2	**0.0023**	**0.0153**	U	**0.0191**	0.128	D	POS
MAPK14	**0.0023**	**0.0153**	U	0.2274	0.4684	D	
CCL3	**0.0023**	**0.0153**	U	0.6874	0.8121	U	
CD40LG	**0.0026**	**0.017**	U	**0.0113**	0.1156	D	POS
HLA-DRB1	**0.0026**	**0.017**	U	**0.0075**	0.1156	D	POS
TNFAIP3	**0.0031**	**0.0188**	U	0.0552	0.2609	D	
IL1RN	**0.0031**	**0.0188**	U	0.4631	0.6889	D	
CXCL10	**0.0036**	**0.0194**	U	**0.0086**	0.1156	D	POS
IL18	**0.0036**	**0.0194**	U	**0.0191**	0.128	D	POS
CXCL8	**0.0036**	**0.0194**	U	0.1626	0.4039	D	
NR3C1	**0.0036**	**0.0194**	U	0.2659	0.5087	D	
CCR7	**0.0036**	**0.0194**	D	0.4925	0.69	U	
STAT1	**0.0042**	**0.0207**	U	**0.0005**	0.0572	D	POS
HMGN1	**0.0042**	**0.0207**	U	0.0759	0.3099	D	
NFKB1	**0.0042**	**0.0207**	U	0.1239	0.3982	D	
CCL2	**0.0042**	**0.0207**	D	**0.0113**	0.1156	U	NEG
DDIT3	**0.0049**	**0.0235**	U	0.4074	0.6376	U	
RHOA	**0.0056**	**0.0251**	U	**0.0042**	0.1008	D	POS
MAX	**0.0056**	**0.0251**	U	**0.0099**	0.1156	D	POS
MAPK1	**0.0056**	**0.0251**	U	0.6529	0.7987	U	
C3AR1	**0.0056**	**0.0251**	D	**0.0217**	0.1375	U	NEG
ATF2	**0.0065**	**0.0285**	U	0.2461	0.4985	D	

PTGDR2	**0.0075**	**0.0311**	U	0.4925	0.69	D	
TLR3	**0.0075**	**0.0311**	U	0.4631	0.6889	U	
STAT3	**0.0075**	**0.0311**	D	**0.0019**	0.0776	U	NEG
TGFB1	**0.0086**	**0.0329**	U	**0.0245**	0.1516	D	POS
RAF1	**0.0086**	**0.0329**	U	0.6529	0.7987	D	
RAPGEF2	**0.0086**	**0.0329**	U	0.5228	0.7118	D	
TOLLIP	**0.0086**	**0.0329**	U	0.7583	0.854	U	
CXCR2	**0.0099**	**0.035**	U	0.1024	0.3527	D	
MMP9	**0.0099**	**0.035**	U	**0.0495**	0.2433	D	POS
PLCB1	**0.0099**	**0.035**	U	0.1488	0.4039	D	
MAP3K9	**0.0099**	**0.035**	U	0.0929	0.3391	D	
STAT2	**0.0099**	**0.035**	U	0.1626	0.4039	D	
IRF1	**0.0113**	**0.0368**	U	0.2274	0.4684	D	
RAC1	**0.0113**	**0.0368**	U	0.1488	0.4039	D	
CREB1	**0.0113**	**0.0368**	U	0.4348	0.659	D	
GNB1	**0.0113**	**0.0368**	U	0.4925	0.69	U	
IL6	**0.0113**	**0.0368**	U	0.5228	0.7118	U	
KEAP1	**0.0129**	**0.0398**	U	0.2659	0.5087	D	
IL18RAP	**0.0129**	**0.0398**	U	0.1626	0.4039	D	
RPS6KA5	**0.0129**	**0.0398**	U	0.2868	0.5237	D	
CCL20	**0.0129**	**0.0398**	U	0.9058	0.925	U	
PRKCB	**0.0148**	**0.0443**	U	**0.0129**	0.1156	D	POS
CD40	**0.0148**	**0.0443**	U	0.3088	0.5554	U	
ROCK2	**0.0168**	**0.0492**	U	0.0929	0.3391	D	
IL23A	**0.0168**	**0.0492**	U	0.2274	0.4684	D	
PTGS1	**0.0217**	0.0584	U	**0.0113**	0.1156	D	POS
OASL	**0.0217**	0.0584	U	**0.0036**	0.1008	D	POS
NFATC3	**0.0217**	0.0584	U	0.3812	0.6292	D	

MEF2D	**0.0217**	0.0584	U	0.6529	0.7987	D	
MAP3K7	**0.0217**	0.0584	U	0.3318	0.5712	D	
IL1A	**0.0217**	0.0584	U	0.4925	0.69	U	
FLT1	**0.0217**	0.0584	U	0.4348	0.659	U	
CCL22	**0.0277**	0.0714	U	**0.0395**	0.2162	D	POS
CXCL9	**0.0277**	0.0714	U	0.4074	0.6376	D	
CCL5	**0.0277**	0.0714	U	0.0552	0.2609	D	
TLR2	**0.0277**	0.0714	D	0.084	0.3266	U	
MX1	**0.0312**	0.078	U	0.193	0.4388	D	
CXCL2	**0.0312**	0.078	U	0.4925	0.69	D	
FASLG	**0.0312**	0.078	U	0.9058	0.925	D	
TRAF2	**0.0352**	0.0833	U	0.4925	0.69	D	
MAPKAPK5	**0.0352**	0.0833	U	0.1488	0.4039	D	
CXCR4	**0.0352**	0.0833	U	0.2274	0.4684	D	
C1R	**0.0352**	0.0833	U	0.9058	0.925	U	
CCL8	**0.0352**	0.0833	D	0.1359	0.3995	U	
NOD2	**0.0395**	0.09	U	0.0759	0.3099	D	
IL1RAP	**0.0395**	0.09	U	0.1626	0.4039	D	
LY96	**0.0395**	0.09	D	0.2097	0.4636	U	
CFB	**0.0395**	0.09	U	0.5228	0.7118	U	
C5	**0.0442**	0.0971	U	0.4074	0.6376	D	
LTB	**0.0442**	0.0971	U	0.1359	0.3995	D	
PDGFA	**0.0442**	0.0971	U	0.5862	0.7554	D	
SMAD7	**0.0442**	0.0971	U	0.6874	0.8121	U	
CD86	**0.0495**	0.1047	U	0.8684	0.9139	D	
OAS2	**0.0495**	0.1047	U	0.1488	0.4039	D	
SHC1	**0.0495**	0.1047	U	0.6529	0.7987	D	
CCR1	**0.0495**	0.1047	D	**0.0442**	0.2317	U	NEG

MAP2K1	0.0552	0.1148	U	0.4074	0.6376	D	
CCL21	0.0552	0.1148	U	0.3318	0.5712	U	
AGER	0.0615	0.1246	U	0.1359	0.3995	D	
RIPK1	0.0615	0.1246	U	0.9058	0.925	U	
IL23R	0.0615	0.1246	U	0.193	0.4388	U	
HSH2D	0.0684	0.1362	U	0.3318	0.5712	D	
TLR1	0.0684	0.1362	D	**0.0277**	0.1629	U	
PRKCA	0.0759	0.1474	U	0.084	0.3266	D	
GNAQ	0.0759	0.1474	U	0.0929	0.3391	D	
IL2	0.0759	0.1474	U	0.5228	0.7118	U	
MAPK8	0.084	0.1619	U	0.3812	0.6292	D	
FXYD2	0.0929	0.1733	U	**0.0442**	0.2317	D	
AREG	0.0929	0.1733	D	0.2274	0.4684	U	
NOX1	0.0929	0.1733	U	0.7564	0.854	U	
PTGS2	0.0929	0.1733	U	0.2659	0.5087	U	
LTB4R2	0.1024	0.1867	U	0.4348	0.659	D	
BORCS8-MEF2B	0.1024	0.1867	U	0.6874	0.8121	U	
BORCS8-MEF2B	0.1024	0.1867	U	0.6874	0.8121	U	
C9	0.1024	0.1867	U	0.5862	0.7554	U	
HMGB1	0.1128	0.1995	U	**0.0168**	0.1229	D	
CEBPB	0.1128	0.1995	U	0.1773	0.4189	U	
MAP3K5	0.1128	0.1995	D	0.3812	0.6292	U	
CCL11	0.1128	0.1995	U	0.193	0.4388	U	
MAFF	0.1239	0.2144	U	0.9811	0.9811	U	
C1S	0.1239	0.2144	U	0.0929	0.3391	U	
IL12B	0.1239	0.2144	U	0.1359	0.3995	U	
TLR5	0.1359	0.2285	U	**0.0191**	0.128	D	
MAP2K6	0.1359	0.2285	U	0.0759	0.3099	D	

CCR4	0.1359	0.2285	U	0.7583	0.854	D	
IL21	0.1359	0.2285	U	0.554	0.7377	U	
DAXX	0.1488	0.2415	U	0.2868	0.5237	D	
ALOX5	0.1488	0.2415	D	0.6529	0.7987	D	
JUN	0.1488	0.2415	U	0.6791	0.8121	U	
OXER1	0.1488	0.2415	U	0.7226	0.8412	U	
CCL13	0.1488	0.2415	U	0.5862	0.7554	U	
CFD	0.1626	0.2552	U	0.0759	0.3099	D	
IRF3	0.1626	0.2552	U	0.6874	0.8121	D	
MAFK	0.1626	0.2552	U	0.8684	0.9139	D	
TLR6	0.1626	0.2552	U	0.6874	0.8121	D	
MAP3K1	0.1626	0.2552	U	0.6192	0.7812	U	
CXCL3	0.1773	0.2676	U	0.8684	0.9139	D	
CYSLTR2	0.1773	0.2676	U	0.9811	0.9811	U	
CCR3	0.1773	0.2676	U	0.8684	0.9139	U	
CXCR1	0.1773	0.2676	U	0.1626	0.4039	U	
CSF3	0.1773	0.2676	U	0.2659	0.5087	U	
C8A	0.1773	0.2676	U	0.7583	0.854	U	
MRC1	0.193	0.2895	U	0.4074	0.6376	U	
IL4	0.2097	0.3067	U	0.5862	0.7554	D	
IL10RB	0.2097	0.3067	U	0.8684	0.9139	D	
RELB	0.2097	0.3067	U	0.0759	0.3099	U	
IL3	0.2097	0.3067	U	0.8313	0.9025	U	
LTA	0.2274	0.3246	U	0.5228	0.7118	D	
PTGIR	0.2274	0.3246	U	0.7946	0.8825	U	
IL11	0.2274	0.3246	U	0.1239	0.3982	U	
IFNB1	0.2274	0.3246	U	0.4925	0.69	U	
HSPB1	0.2659	0.3708	U	0.4925	0.69	D	

IRF7	0.2659	0.3708	D	0.6192	0.7812	U	
IL7	0.2659	0.3708	U	0.6192	0.7812	U	
CCL16	0.2659	0.3708	U	0.5862	0.7554	U	
CXCL1	0.2868	0.3975	D	0.1359	0.3995	U	
MAPKAPK2	0.3088	0.4158	U	0.8313	0.9025	U	
TLR7	0.3088	0.4158	D	0.4348	0.659	U	
IL22	0.3088	0.4158	U	0.8684	0.9139	U	
ALOX15	0.3088	0.4158	D	0.2274	0.4684	U	
CCL7	0.3088	0.4158	D	0.3812	0.6292	U	
DEFA1	0.3318	0.4321	U	0.2274	0.4684	D	
MYD88	0.3318	0.4321	D	0.2097	0.4636	D	
MKNK1	0.3318	0.4321	U	0.7226	0.8412	D	
HRAS	0.3318	0.4321	U	0.9058	0.925	U	
ELK1	0.3318	0.4321	U	0.8313	0.9025	U	
CCL17	0.3318	0.4321	U	0.1488	0.4039	U	
PPP1R12B	0.356	0.4585	U	0.9434	0.9513	U	
TBXA2R	0.356	0.4585	D	**0.0495**	0.2433	U	
IL13	0.3812	0.4831	U	0.7946	0.8825	U	
MBL2	0.3812	0.4831	U	0.4074	0.6376	U	
C6	0.3812	0.4831	U	0.9058	0.925	U	
IL12A	0.4074	0.5003	U	0.0759	0.3099	D	
TRADD	0.4074	0.5003	D	0.193	0.4388	U	
IL1R1	0.4074	0.5003	D	0.554	0.7377	U	
IL1B	0.4074	0.5003	U	0.6192	0.7812	U	
MASP1	0.4074	0.5003	U	0.1626	0.4039	U	
PTGFR	0.4074	0.5003	U	0.1773	0.4189	U	
CD4	0.4348	0.5257	U	0.1773	0.4189	D	
RELA	0.4348	0.5257	U	0.7226	0.8412	U	

PTGER3	0.4348	0.5257	U	0.7946	0.8825	U	
IL6R	0.4631	0.5515	U	**0.0395**	0.2162	D	
BCL2L1	0.4631	0.5515	U	0.7583	0.854	U	
NOS2	0.4631	0.5515	U	0.7583	0.854	U	
TLR4	0.4925	0.5807	U	0.2868	0.5237	D	
BIRC2	0.4925	0.5807	U	0.2868	0.5237	U	
GRB2	0.5228	0.6134	D	0.084	0.3266	U	
MX2	0.554	0.6343	D	0.4074	0.6376	D	
PLA2G4A	0.554	0.6343	U	0.0684	0.3099	U	
TLR9	0.554	0.6343	D	0.2659	0.5087	U	
CCL23	0.554	0.6343	U	0.1024	0.3527	U	
IL5	0.554	0.6343	U	0.4631	0.6889	U	
IFNA1	0.5695	0.6489	U	0.2274	0.4684	U	
HDAC4	0.5862	0.6615	D	0.9058	0.925	U	
CRP	0.5862	0.6615	U	0.4925	0.69	U	
CFL1	0.6192	0.6857	U	0.4925	0.69	U	
C4A	0.6192	0.6857	U	0.356	0.6041	U	
C1QA	0.6192	0.6857	U	0.3318	0.5712	U	
CXCL6	0.6192	0.6857	U	0.2461	0.4985	U	
GNAS	0.6529	0.7164	U	0.9434	0.9513	D	
CD55	0.6529	0.7164	U	0.3088	0.5554	U	
MEF2C	0.6874	0.7405	U	0.2097	0.4636	D	
CXCL5	0.6874	0.7405	U	0.8313	0.9025	U	
MMP3	0.6874	0.7405	U	0.2659	0.5087	U	
PIK3C2G	0.6874	0.7405	U	0.1359	0.3995	U	
LTB4R	0.7226	0.7749	U	0.5862	0.7554	D	
FOS	0.7583	0.7952	U	0.3318	0.5712	D	
PTGER4	0.7583	0.7952	U	0.8684	0.9139	U	

MAFG	0.7583	0.7952	U	0.1239	0.3982	U	
C1QB	0.7583	0.7952	U	0.1626	0.4039	U	
TGFB3	0.7583	0.7952	U	**0.0148**	0.1229	U	
TGFB2	0.7946	0.8296	D	0.356	0.6041	U	
GNGT1	0.8313	0.8641	U	0.2868	0.5237	U	
MAPK3	0.8684	0.891	D	0.9434	0.9513	U	
ARG1	0.8684	0.891	D	0.554	0.7377	U	
C7	0.8684	0.891	U	0.1488	0.4039	U	
TCF4	0.9058	0.9213	U	0.6874	0.8121	D	
C2	0.9058	0.9213	U	0.0684	0.3099	U	
CCR2	0.9811	0.9811	U	0.1128	0.3828	D	
TLR8	0.9811	0.9811	U	0.1773	0.4189	U	
CCL19	0.9811	0.9811	U	0.7583	0.854	U	
ALOX12	0.9811	0.9811	U	0.1359	0.3995	U	
MYL2	-	-	D	0.4074	0.6376	D	
HSPB2	-	-	D	0.6529	0.7987	D	
TYROBP	-	-	D	**0.0042**	0.1008	D	
IL9	-	-	D	0.1024	0.3527	U	
MASP2	-	-	D	-	-	D	
KNG1	-	-	D	-	-	D	
C8B	-	-	D	-	-	D	
PTGER1	-	-	D	-	-	D	
TWIST2	-	-	D	-	-	D	
TSLP	-	-	D	-	-	D	

Table 4.4: This table shows the results of the TCZ+IL6 vs IL6 and IL6 vs UNT differential expression analysis. "Protein" indicates the protein's gene name, "p" the p-value of the Wilcoxon-test, "fdr(p)" the adjusted p-value of the test, and "FC" the sign of the fold change with "U" for positive fold change and "D" for negative fold change, "SET" indicates whether the protein has been assigned to the negative (NEG) or positive (NEG).

4.3.2. Clustering and Classification

Clustering is adopted to identify the natural stratification of the GCA patients. Clustering has been applied using the "negative_set" of protein to identify two groups. The expression values of the seven proteins from the three experiments UNT, IL6, and IL6+TCZ are considered for a total of 7*3=21 features.

We define the responder group (those who respond better to Tocilizumab) as the cluster for which the average expression value in the TCZ+IL6 data is lower than in the IL6 data.

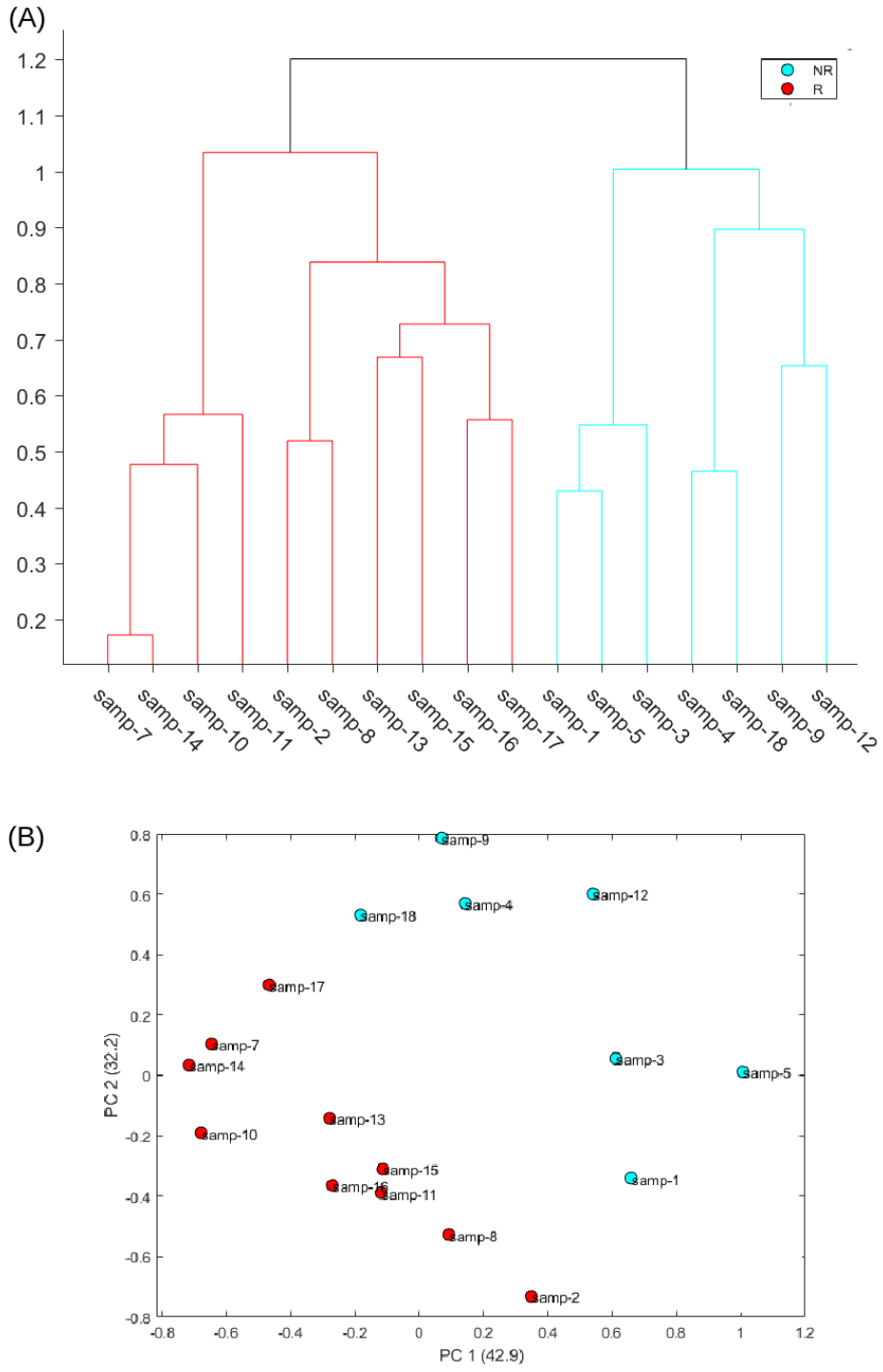

Figure 4.3: (A) shows the dendrogram obtained from the clustering analysis using average aggregation function and Spearman as distance. (B) shows the MDS generated using Spearman as distance function. The Non-Responder group (NR) is in blue, the Responder group (R) is in Red. The clustering solution reported $Mean - Silhouette - index = 1.5, Dunn - Index = 0.41$, and $Hopkins - statistics = 0.62$.

Furthermore, classification is performed considering all the combinations of 7 or fewer features using the only data from the UNT dataset to identify stratification in gene expression without treatment. Results are shown in Table 4.5.

Protein-Set	TP	TN	FP	FN	ACC	PREC	TNR	TPR	NPV	MCC
STAT3,MYC,C3AR1,BCL6	7	9	1	0	0.94	0.88	1.0	0.9	1.0	0.89
STAT3,MYC,HIF1A,CCL2	7	9	1	0	0.94	0.88	1.0	0.9	1.0	0.89
STAT3,MYC,CCR1,C3AR1,BCL6	7	9	1	0	0.94	0.88	1.0	0.9	1.0	0.89
STAT3,MYC,CCR1,CCL2,BCL6	7	9	1	0	0.94	0.88	1.0	0.9	1.0	0.89
STAT3,MYC,C3AR1	6	10	0	1	0.94	1.0	0.86	1.0	0.91	0.88
STAT3,MYC	7	8	2	0	0.88	0.78	1.0	0.8	1.0	0.79
STAT3	4	6	4	3	0.59	0.5	0.57	0.6	0.67	0.17
MYC	0	8	2	7	0.47	0.0	0.0	0.8	0.53	-0.31

Table 4.5: The table shows the results of the classification experiment (logistic regression, leave one out validation method). The "Protein-set" indicates the gene names of the proteins used to train the classifier, "TP" are the true positives, "TN" are the true negatives, "FP" are the false-positives, "FN" the false negatives, "ACC" is accuracy, "PREC" is precision, "TNR" is the true negative rate, "TPR" is the true positive rate, "NPV" is the negative predicted value and "MCC" is the Matthew correlation coefficient.

It is interesting to note that the classifiers maximizing the quality of the prediction ($MCC \approx 0.9$) all include STAT3 and MYC as features. The classifier with only STAT3 and MYC scores 0.79, whereas classifiers with the single genes STAT3 or MYC perform poorly, with MCC values of 0.17 and -0.31, respectively. These results suggest a cooperative role of STAT3 and MYC in the response to Tocilizumab. From this evidence, is clear that lower levels of STAT3 and MYC before treatment cause a stronger response to Tocilizumab, i.e. a stronger decrease of IL6 levels after the drug is administered. This information could be helpful to calibrate the dose of the drug and to better understand the stratification of the patients.

4.3.3. System biology analysis, TPMS

The mechanism of action of Tocilizumab in GCA has been computed using the TPMS, the MoAs have been computed using as restriction the genes of both the negative and positive set and their respective activation state after the treatment with Tocilizumab, With the genes belonging to the negative set being downregulated by Tocilizumab and the genes of the positive set being up-regulated by it. The results are shown in Figure 4.4.

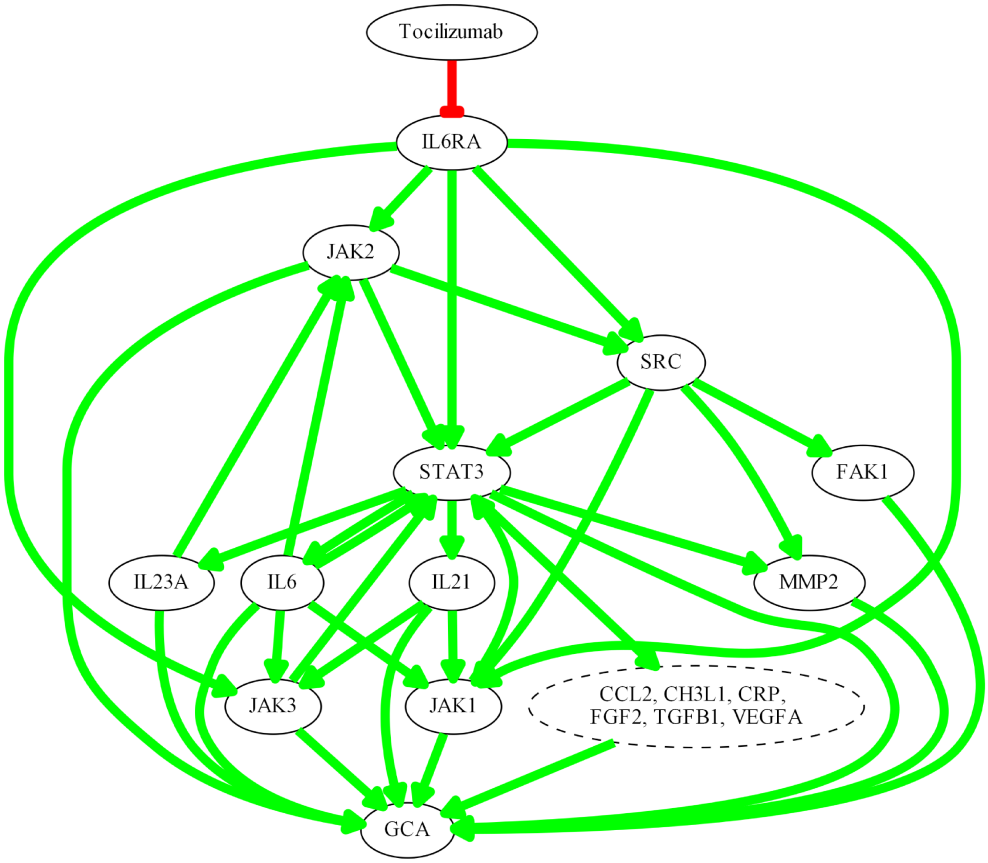

Figure 4.4: The image show the average MoA obtained using has restriction both the positive and negative sets identified in the differential expression analysis. The red lines indicate inhibition, the green line indicates activation.

The MoA highlights the effect of Tocilizumab on IL6RA, which is directly inhibited by the drug. This inhibition effectively blocks the IL6 signaling cascade and its relative inflammatory response. The downstream effects of IL6RA inhibition were observed across several key proteins and cytokines involved in immune signaling, including JAK1, JAK2, JAK3, STAT3, IL6, IL21, and IL23A.

STAT3, a central node in the network, displayed extensive connectivity, suggesting its pivotal role in mediating the downstream effects of IL6RA inhibition. The activation (green arrows) of STAT3 influenced a range of cellular processes by interacting with multiple proteins such as SRC, FAK1, and MMP2, further propagating the signal.

The network analysis also identified significant downstream targets influenced by the pathway's modulation, including molecules involved in inflammation and angiogenesis: CCL2, CH3L1, CRP, FGF2, TGFB1, and VEGFA. The simulation indicates that Tocilizumab's

inhibition of IL6RA disrupts the signaling pathways that lead to the production of these factors, potentially reducing inflammatory and proliferative responses.

In summary, the simulation of Tocilizumab's action within the protein network elucidates its broad impact on immune signaling pathways. By inhibiting IL6RA, Tocilizumab modulates key proteins and cytokines, which collectively contribute to its therapeutic effects in reducing inflammation and altering the pathological signaling of Giant Cell Arteritis.

4.4. MMPred

4.4.1. Supervised evaluation

Prediction sets 1 (using BLASTP) and 2 (using PSI-BLAST) evaluate the capacity of the tool to identify human peptides from a pool of known autoepitopes (HADS dataset) that significantly align with known epitopes from microbial species known to be associated to autoimmune diseases (MEDS dataset) and are recognized as HLA class II epitopes by one or both predictors used. In essence, starting from a pool of known human autoepitopes, the tool predicts which of them could induce autoimmunity as a consequence of a previous infection and microbial peptide mimicry. Out of the 807 known human autoepitops contained in HADS, 21 had at least one sequence fragment that significantly aligned with a microbial epitope from MEDS and was predicted as an autoepitope by CNN-PepPred and/or NetMHCIIpan (Table 7.9). The matching microbial epitopes are from SARS-CoV-2, *Mycobacterium tuberculosis* (MT) and Human Alphaherpesvirus (HHV) 1 and 3. PSI-BLAST and BLASTP produced significant alignments in all cases.

In four of the cases, the predicted autoepitopes were matched at the epitope-allele level and using OneHLA as allele selection criteria (see section 3.5.4), meaning that the predictions matched the allele that has been found to bind, experimentally, both the microbial and human epitopes. The corresponding autoantigens are H1-2, H1-4 and MPO, showing sequence similarity with HbhA and RplV from MT and the Spike Glicoprotein from SARS-CoV-2 (Table 7.9). Furtermore, the alleles DRB1*15:01 and DRB5*01:01 are know to be linked to the autoimmune disease —multiple sclerosis– that has been associated with these autoantigens [Karni et al., 1999, Finn et al., 2004, Prat et al., 2005, Živković et al., 2009, Shahbazi et al., 2010, Alcina et al., 2012, Quandt et al., 2012, Apperson et al., 2013, Kaushansky and Ben-Nun, 2014, Stürner et al., 2019].

4.4.2. Functional evaluation

Prediction sets 3 (using BLASTP) and 4 (using PSI-BLAST) evaluate the capacity of the tool to identify peptides from the full human proteome (HPDS dataset) that significantly align with known epitopes from microbial species known to be associated to autoimmune diseases (MEDS dataset) and are recognized as HLA class II epitopes by one or both predictors used. In short, it extends the analysis reported in section 4.4.1 to the full human proteome.

The identified autoantigens were then subjected to the functional evaluation described in section 3.5.4. The results are summarized in Figure 4.5 and Table 4.6, where the dependence of the distribution of scores S —the probability that there exists a relationship between the predicted autoantigen and a pathophysiological pathway associated to one of the autoimmune diseases considered– on the various parameters —E-value and %Rank thresholds, use of BLASTP or PSI-BLAST, use of the allHLA or oneHLA allele-selection criteria– is evaluated. In addition, the different score distributions are compared to a score distribution for a random subset of 1000 proteins from HPDS.

Using PSI-BLAST notably reduces the number of hits, but its strength lies in incorporating specific biological information into the alignments. By leveraging the pool of peptides known to bind HLA class II molecules to generate a Position-Specific Scoring Matrix, PSI-BLAST may enhance the alignment's relevance. Although this approach results in fewer hits, it achieves the highest mean score ratio and the most significant distinction between predicted autoantigen score distributions and the random peptide score distribution, particularly at E-value and %Rank thresholds of 0.01 and 2, respectively. These thresholds optimize the selection of predicted autoantigens, increasing the likelihood of identifying proteins associated with a specific autoimmune disease in the human protein network. However, using a more stringent E-value threshold of 0.001 drastically reduces the number of aligned sequences, which, when combined with a %Rank threshold of 2, can prevent the prediction of any autoantigens. The choice of allele-selection criteria for the binding prediction, allHLA or oneHLA, plays also a significant role in the results, particularly regarding the number of hits.

The results of this analysis require some biological context for proper interpretation. We are examining the potential relationship, within a graph representing the human protein network, between a protein identified as a potential autoantigen and a set of proteins that have been linked to the pathophysiology of a specific autoimmune disease. Clearly, the proteins in this set tend to be elements of the immune system. Therefore, a high score indicates that the predicted autoantigen is, in network terms, associated to these immune-system elements. While this may not be a universal characteristic of all autoantigens, Figure 4.5 and Table 4.6 show that as we apply more stringent alignment and epitope prediction thresholds, thereby increasing our confidence in the autoepitope, the mean score and significance of the relationship between autoantigen and autoimmune disease motif also increase. This focused approach not only enhances the reliability of our predictions but also allows us to propose potential links between the predicted autoantigens and specific autoimmune diseases, which would be difficult to establish otherwise.

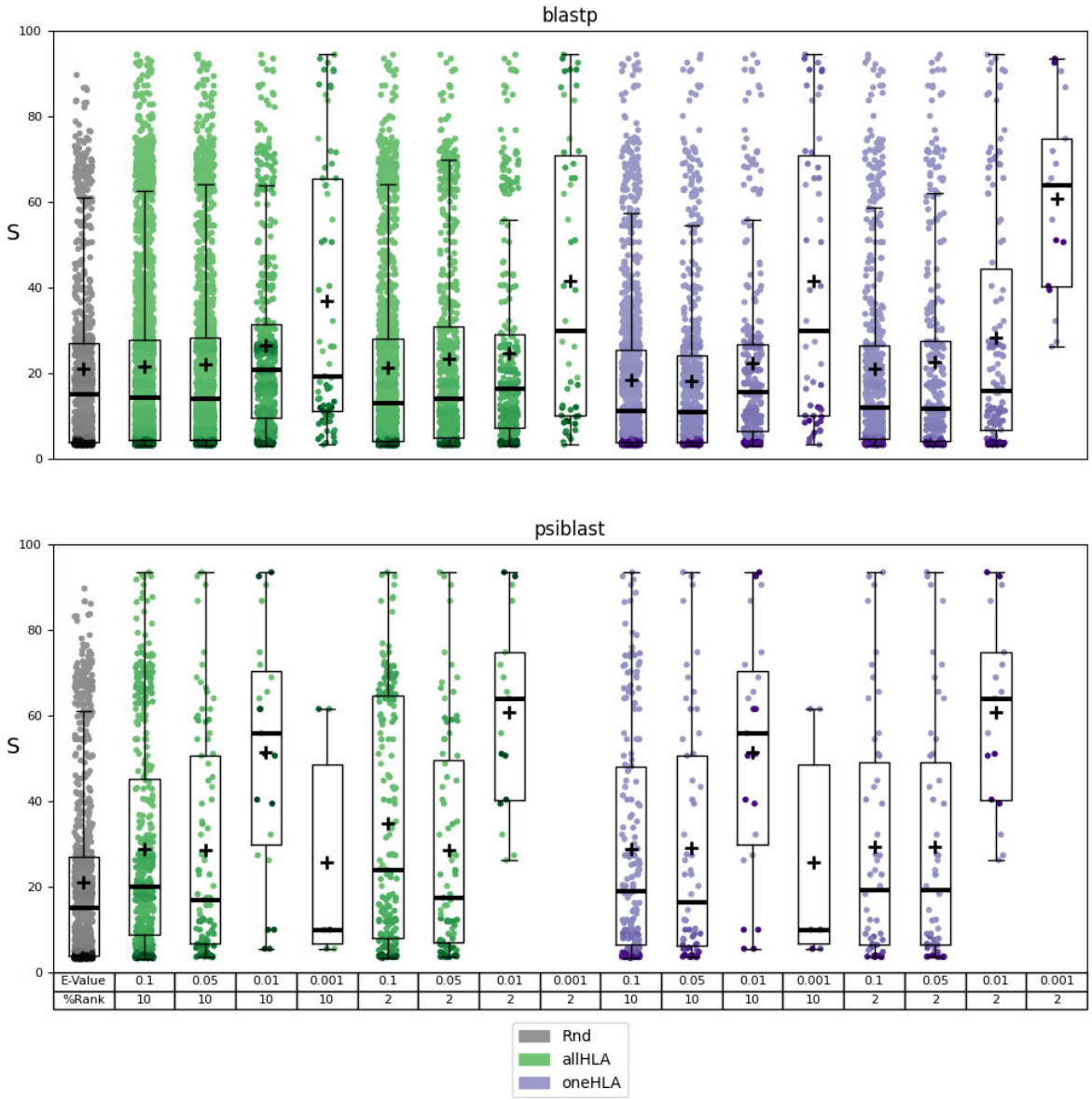

Figure 4.5: Results of the functional analysis for prediction sets 3 (using BLASTP) and 4 (using PSI-BLAST). Dependence of the distribution of scores S (box plots overlaid with scatterplots) on the following parameters: use of BLASTP or PSI-BLAST and threshold E-value for the alignment, %Rank threshold and allele-selection criterion (allHLA or one HLA, see section 3.5.4) for the epitope prediction. The random distribution (Rnd) is represented in grey. The mean of the distribution is indicated with a cross.

Alignment	Alleles	%Rank	E-Value	N	Score Ratio	p-Value	Significance
BLASTp	allHLA	10	0.1	3144	1.11	3.41e-02	-
BLASTp	allHLA	10	0.05	1770	1.15	2.33e-02	-
BLASTp	allHLA	10	0.01	516	1.36	1.10e-13	***
BLASTp	allHLA	10	0.001	67	1.91	2.05e-06	***
BLASTp	allHLA	2	0.1	1594	1.1	1.99e-01	-
BLASTp	allHLA	2	0.05	794	1.2	1.07e-02	-
BLASTp	allHLA	2	0.01	301	1.28	4.25e-04	**
BLASTp	allHLA	2	0.001	50	2.14	9.51e-07	***
BLASTp	oneHLA	10	0.1	1346	0.96	9.88e-01	-
BLASTp	oneHLA	10	0.05	711	0.95	9.82e-01	-
BLASTp	oneHLA	10	0.01	299	1.15	1.26e-02	-
BLASTp	oneHLA	10	0.001	50	2.14	9.51e-07	***
BLASTp	oneHLA	2	0.1	618	1.09	3.53e-01	-
BLASTp	oneHLA	2	0.05	330	1.16	2.95e-01	-
BLASTp	oneHLA	2	0.01	134	1.47	2.43e-03	*
BLASTp	oneHLA	2	0.001	17	3.14	9.69e-10	***
PSI-BLAST	allHLA	10	0.1	487	1.49	1.53e-13	***
PSI-BLAST	allHLA	10	0.05	89	1.47	6.60e-04	**
PSI-BLAST	allHLA	10	0.01	23	2.66	3.85e-08	***
PSI-BLAST	allHLA	10	0.001	6	1.33	3.26e-01	-
PSI-BLAST	allHLA	2	0.1	184	1.8	3.03e-11	***
PSI-BLAST	allHLA	2	0.05	83	1.48	6.45e-04	**
PSI-BLAST	allHLA	2	0.01	17	3.14	9.69e-10	***
PSI-BLAST	allHLA	2	0.001	0	-	-	-
PSI-BLAST	oneHLA	10	0.1	188	1.49	1.22e-05	***
PSI-BLAST	oneHLA	10	0.05	56	1.5	1.04e-02	-
PSI-BLAST	oneHLA	10	0.01	23	2.66	3.85e-08	***
PSI-BLAST	oneHLA	10	0.001	6	1.33	3.26e-01	-
PSI-BLAST	oneHLA	2	0.1	50	1.52	1.08e-02	-
PSI-BLAST	oneHLA	2	0.05	50	1.52	1.08e-02	-
PSI-BLAST	oneHLA	2	0.01	17	3.14	9.69e-10	***
PSI-BLAST	oneHLA	2	0.001	0	-	-	-

Table 4.6: Results of the functional analysis for prediction sets 3 (using BLASTp) and 4 (using PSI-BLAST). Alleles: allele-selection criterion (see section 3.5.4); Score ratio: mean-score ratio; N: size of the sample; p-Value: from the Mann-Whitney U test; Significance: - (not significant), * (p-value < 0.01), ** (p-value < 0.001), *** (p-value < 0.0001).

4.4.3. SARS-CoV-2 peptide mimicry

The predicted autoantigens and their relation to autoimmune diseases

Prediction sets 5 (using BLASTP) and 6 (using PSI-BLAST) evaluate the capacity of the tool to identify peptides from the full human proteome (HPDS dataset) that significantly align with known epitopes from SARS-CoV-2 (SC2DS dataset) and are recognized as HLA class II epitopes by one or both predictors used. It thus focuses the analysis on finding SARS-CoV-2 epitopes that could induce an autoimmune disease through peptide mimicry. As in the previous section, the identified autoantigens were evaluated for their potential relationship with autoimmune-disease motifs in the human protein network (section 3.5.4).

The results of this evaluation are reported in Figure 4.6. Using BLASTP, MMPred identified 14 potential autoantigens: MYT1L, BAZ1A, CHD5, MCM8, ATF7, MOV10, MOV10L1, DNA2, BRI3, PARVG, CALD1, MICAL3, SLC35E4, and UNC50, using a threshold E-value of 0.01. In contrast, PSI-BLAST did not detect any autoepitopes with an E-value below 0.01 and predicted only one autoepitope from HELZ2 with an E-value below 0.05. This outcome is consistent with the previous section's analysis, where BLASTP yielded a significantly higher number of positive predictions at the same thresholds. Most alignments involved SARS-CoV-2 Non-Structural Proteins (NSPs), particularly NSP3, NSP5, NSP13, NSP14, NSP15, and NSP16. Additionally, one alignment involved Nucleoprotein N, and two alignments were related to the Spike protein of the Omicron variants BA.1-like and BA.4-like [Wu et al., 2020]. These findings are summarized in Table 7.10.

In the functional analysis, the 15 predicted autoantigens were assessed for potential associations with 32 autoimmune-disease motifs annotated in the network. Eight proteins —BAZ1A, ATF7, MOV10, DNA2, PARVG, MICAL3, SLC35E4, and HELZ2– scored in the 95th percentile (Perc(S), see Section 3.5.4) or higher for 19 different motifs (see Figure 4.6). While no autoantigen met the minimum threshold for certain motifs, at least one motif from each selected autoimmune disease had a significant hit. Notably, BAZ1A, MOV10, and PARVG were the only proteins with a Perc(S) exceeding 99 for at least one motif. The highest Perc(S) (above 99.9) was achieved by MOV10 in association with the "Lupus Erythematosus Systemic" motif.

With a Perc(S) > 95, Rheumatoid Arthritis and Thyroiditis are associated with the same predicted autoantigens, MOV10 and ATF7, which are also linked to Lupus Erythematosus Systemic along with PARVG. Additionally, SLC35E4 is also connected to Rheumatoid Arthritis. Type I Diabetes is associated with the predicted autoantigens MOV10 and PARVG, with PARVG also linked to Guillain-Barré Syndrome, alongside ATF7, BAZ1A, and DNA2. Myasthenia Gravis shows predicted associations with HELZ2, ATF7 and PARVG. Lastly, Anemia is associated with MICAL3 and BAZ1A.

When the threshold is raised to Perc(S) > 99, MOV10 is associated with both Lupus Erythematosus Systemic and Rheumatoid Arthritis. MOV10 is also linked to Thyroiditis, along with ATF7. Anemia is connected to BAZ1A, and Myasthenia Gravis is linked to PARVG.

The human proteins MYT1L, CHD5, MCM8, MOV10L1, BRI3 and CALD1 were predicted as autoantigens but did not show a Perc(S) > 95 for any of the motifs tested.

The predicted human autoepitopes align with known SARS-CoV-2 epitopes

While there is no experimental evidence in IEDB linking our predicted autoepitopes to autoimmune diseases, many of the SARS-CoV-2 sequences that align significantly with these predicted autoepitopes are known to bind HLA class II molecules. Specifically, the SARS-

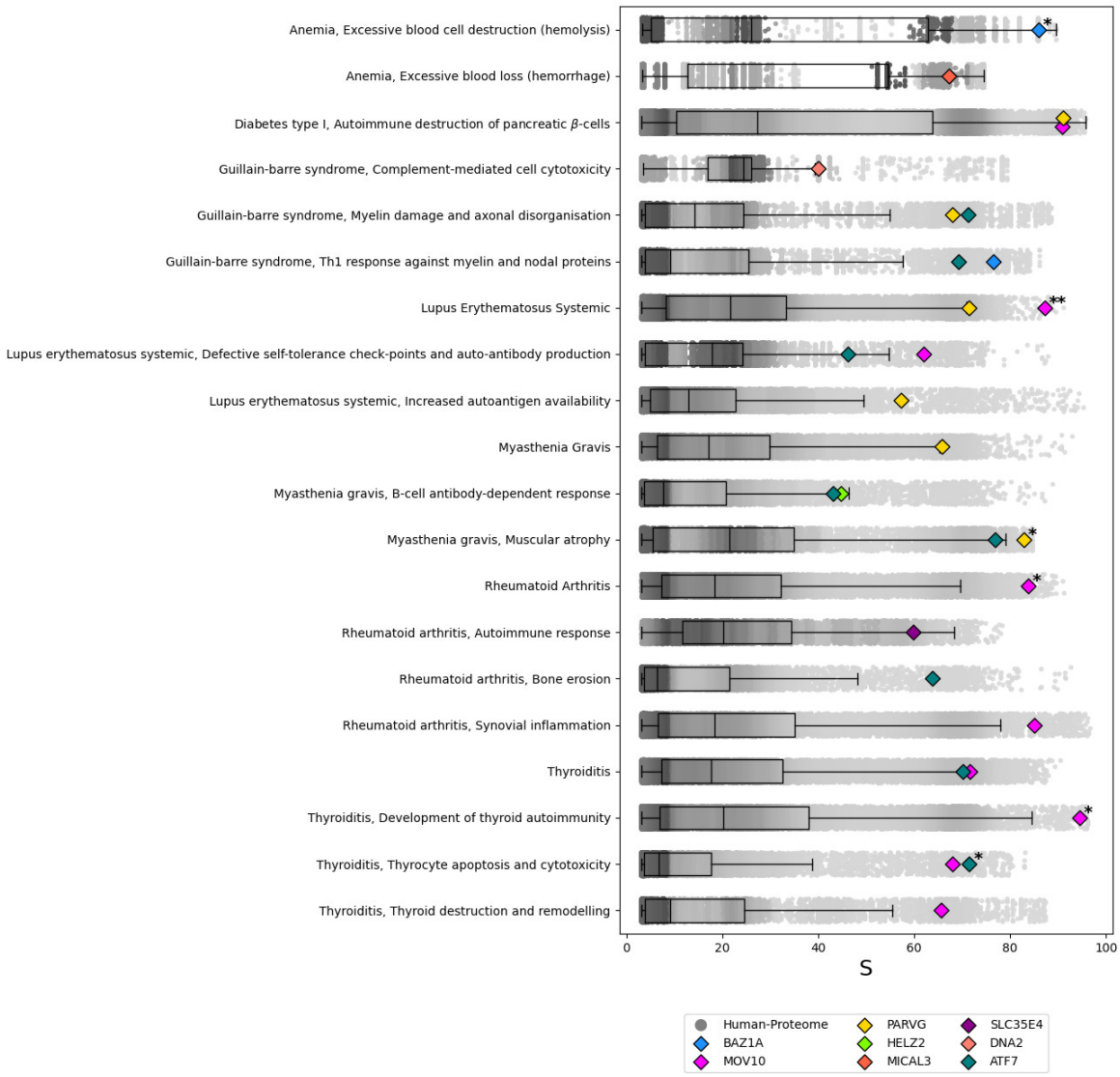

Figure 4.6: Results of the ANNs analysis for the predicted autoantigens of the SARS-CoV-2 vs. Human protome prediction sets. For each of the autoimmune-disease motifs tested, a boxplot with overlapped scatterplot represent the background distribution of the score S. The eight predicted autoantigens that satisfy Perc(S) $>$ 95 are shown. Those with Perc(S) $>$ 99 are marked with *, and those with Perc(S) $>$ 99.9 are marked with **.

CoV-2 sequences 369-378 of the Nucleoprotein, 47-55, 275-289, and 557-569 of NSP13, 389-398 of NSP14, 248-258 of NSP15, and 257-267 of NSP16 overlap with regions that have been experimentally validated [Obermair et al., 2022, Huisman et al., 2022] (see Table 7.10). Notably, the regions 275-289 and 557-569 of NSP13 show allele-specific hits, where the autoepitopes have been predicted for the same alleles experimentally observed to bind these SARS-CoV-2 protein regions. Thus, DRB1*07:01 is predicted to bind an autoepitope in MOV10, and DRB1*04:04 is predicted to bind an autoepitope in HELZ2.

The functional analysis reveals an interesting pattern among the predicted autoantigens that align with known SARS-CoV-2 epitopes. Specifically, nine human proteins —ATF7, HELZ2, MOV10, MOV10L1, DNA2, BRI3, PARVG, CALD1, MICAL3– exhibit significant alignment with these SARS-CoV-2 epitopes. Among these, ATF7, HELZ2, MOV10, DNA2, PARVG and MICAL3 rank in the 95th percentile or higher (Perc(S) > 95) across 18 different motifs, with ATF7, MOV10, and PARVG showing Perc(S) > 99 for six motifs. MOV10 alone achieves a Perc(S) > 99.9 for a single motif. In contrast, the associations are less significant when examining predicted autoantigens derived from alignments with SARS-CoV-2 sequences not known to be antigenic. Of the six proteins with peptides matching these regions (MYT1L, BAZ1A, CHD5, MCM8, SLC35E4 and UNC50), only BAZ1A and SLC35E4 yield significant results, both achieving Perc(S) > 95 for just three motifs, with only BAZ1A reaching Perc(S) > 99 for a single motif.

The antiviral activity of the predicted autoantigens MOV10 and HELZ2

We examined baseline expression patterns of the predicted autoantigens in lung cells using the Expression Atlas [George et al., 2024]. Proteomics data from Wang et al. [Wang et al., 2019] indicate that 11 out of the 15 predicted autoantigens (BAZ1A, MCM8, ATF7, HELZ2, MOV10, MOV10L1, DNA2, PARVG, CALD1, MICAL3, UNC50) are highly expressed, supporting the hypothesis that cross-reaction with HLA class II molecules in previously infected lung cells is plausible [Kawasaki et al., 2022, Hoffmann et al., 2020]. Additionally, a gene enrichment analysis conducted using the online tool g:Profiler [Kolberg et al., 2023] (threshold = 0.05, multiple hypothesis testing method g:SCS) revealed significant enrichment for the Gene Ontology (GO, `https://geneontology.org/`) Molecular Function terms "helicase activity" (GO:0004386) and "single-stranded DNA-helicase" (GO:0017116), as well as the Cellular Component terms "P granule" (GO:0043186) and "intracellular non-membrane-bounded organelle" (GO:0043232). Notably, the human proteins MOV10, HELZ2, and DNA2, which are reported to have helicase activity with GO evidence code "inferred by direct assay" (IDA), contain segments that align with the helicase NSP13 of SARS-CoV-2 (see Table 7.10). MOV10 is implicated in the modulation of viral infectivity [Goodier et al., 2012] and promotes type I interferon production [Cuevas et al., 2016, Yang et al., 2022, Balinsky et al., 2017], while HELZ2 is known to respond to interferon production during viral infection [Huntzinger et al., 2023, Du et al., 2024]. DNA2 does not appear to have any known antiviral activity. On the other hand, the helicase activity of NSP13 is crucial for viral replication [Yan et al.,

2021] and this protein interacts with the host to inhibit interferon-beta production, thereby evading the immune response [Xia et al., 2020]. Thus, it seems that MOV10 and HELZ2 play roles antagonistic to that of NSP13.

Overall, transcriptomic and proteomic data on our predicted autoantigens in the context of SARS-CoV-2 infection primarily focus on MOV10 and HELZ2. These studies utilize samples obtained from either the lungs of infected patients or lung cell lines.

Wang et al. employed Dermatan Sulfate (DS)-affinity proteomics to define the autoantigenome of lung fibroblasts, complemented by bioinformatics analyses to explore the relationship between autoantigenic proteins and COVID-19-induced alterations [Wang et al., 2022]. Notably, they discovered that 86% of their predicted autoantigens were either up- or down-regulated in COVID-19 patients or SARS-CoV-2-infected cells. Among the previously unknown autoantigens identified in this study, MOV10 (with very high DS affinity) and CALD1 (with medium to high DS affinity) align with our predictions. Both proteins exhibited altered expression in COVID-19 patients and/or SARS-CoV-2-infected cells.

In a study published by An et al., the authors developed a bioinformatics pipeline similar to ours (see comment in Introduction) and performed a differential expression analysis using the Calu-3 human lung adenocarcinoma cell line [An et al., 2022]. Their pipeline predicts cross-reactivity for MOV10 and HELZ2 and reports a positive correlation of these proteins' transcripts with SARS-CoV-2 viral load in infected cells.

In a study by An et al., the authors developed a bioinformatics pipeline similar to ours (as noted in the Introduction) and conducted a differential expression analysis using the Calu-3 human lung adenocarcinoma cell line [An et al., 2022]. Their pipeline predicts that MOV10 and HELZ2 may exhibit cross-reactivity, while expression data reveal a positive correlation between transcript abundance of these proteins, particularly HELZ2, and SARS-CoV-2 viral load in infected cells.

A study by Ariumi explored the epigenetic mechanisms triggered by SARS-CoV-2 infections [Ariumi, 2022]. The findings indicate that the knockdown of MOV10 leads to a significant increase in viral load/replication in infected cells, suggesting that MOV10 plays a role in the host's suppression of SARS-COV-2 replication.

A study examining gene expression in cell lines and patient samples in the context of epigenetic regulation during SARS-CoV-2, SARS-CoV, and MERS infections identified various differentially expressed genes involved in the epigenetic response during infection in pulmonary cell lines [Salgado-Albarrán et al., 2021]. Although the study does not report expression data for MOV10, it highlights MOV10's functional and physical relationships with the differentially expressed genes through protein-protein interaction (PPI) data [Kotlyar et al., 2019] and co-expression analysis [Langfelder and Horvath, 2008].

Aside from MOV10 and HELZ2, the literature also reports the overexpression of MICAL3 in convalescent COVID-19 patients who retested positive [Fang et al., 2022]. However, no relevant studies have been found that link the other predicted autoantigens to SARS-CoV-2.

Sequence and structural similarity between NSP13 and MOV10, an immune-excape mechanism?

Considering the reported evidence regarding MOV10, we extended our investigation to NSP13 due to its shared helicase activity and the presence of a similar HLA class II epitope. Given MOV10's suggested suppressor role in SARS-CoV-2 replication, its presence in lung epithelial cells, its upregulated expression in SARS-CoV-2-infected cells, and the APC-like properties of lung epithelial cells [Kawasaki et al., 2022, Hoffmann et al., 2020], we propose that the presence of epitopes in NSP13 that are cross-reactive with MOV10 epitopes might represent a mechanism that facilitates SARS-CoV-2 replication in the lungs.

We would therefore expect the NSP13 epitope NVNRFNVAITRAK (positions 557 to 569, see Table 7.10) to be conserved across NSP13 variants. However, it is important to note that this conservation might also arise from the sequence's involvement in NSP13's catalytic activity [Newman et al., 2021].

A total of 1,725,419 protein sequences of the ORF1ab polyproteins were downloaded from NCBI Virus (15/02/2024) [Hatcher et al., 2017], with the maximum number of ambiguous characters set to zero, and considering only sequences collected for baseline surveillance (random sampling = "Only"). To extract NSP13 from the ORF1ab sequences, a BLASTP alignment was performed using the reference sequence of NSP13 as the query and the ORF1ab sequences as the target. Due to the high similarity of the variants to the reference, all alignments yielded significant results, allowing us to heuristically extract the corresponding NSP13 variants. A multiple sequence alignment of all NSP13 variants, along with the NSP13 reference, was then performed using FAMSA [Deorowicz et al., 2016], which is optimized for datasets with high dimensionality and high pairwise identity. To compute the conservation of the NSP13 epitope, we averaged the Shannon Entropy [Shenkin et al., 1991] across its positions and compared it to the distribution of all windows of the same length across the alignment. The NSP13 epitope at positions 557 to 569 showed a higher conservation score than 95.7% of the windows of the same size, indicating the strong conservation of this region. A logo plot of the alignment is presented in Figure 4.7.

Furthermore, the superposition of the crystallographic structure of NSP13 (PDB entry 6ZSL, chain B) [Newman et al., 2021] with the predicted AlphaFold structure of MOV10 (UniProt entry Q9HCE1) [Jumper et al., 2021, Varadi et al., 2022] reveals significant structural similarity between two regions of these proteins of 214 residues in length (36.5% of the sequence of 6ZSL and 21.3% of that of the AlphaFold model of Q9HCE1, with a C_α RMSD of 1.98 Å, see Figure 4.7). Notably, the regions corresponding to the epitopes —NSP13 residues 557-

569 and MOV10 residues 901-913– are superimposed in the structural alignment, suggesting that epitope mimicry could arise from both sequence and structure. Although the MOV10 structure is based on an AlphaFold model [Terwilliger et al., 2024], the model confidence in this region is very high (pLDDT > 90).

Overall, these findings support the hypothesis of cross-reactivity between the epitopes of NSP13 and MOV10, which may influence SARS-CoV-2 replication in the lungs. While the potential for an autoimmune response against MOV10 and its connection to autoimmune diseases remains speculative at this stage (see section 4.4.3), further investigation is warranted.

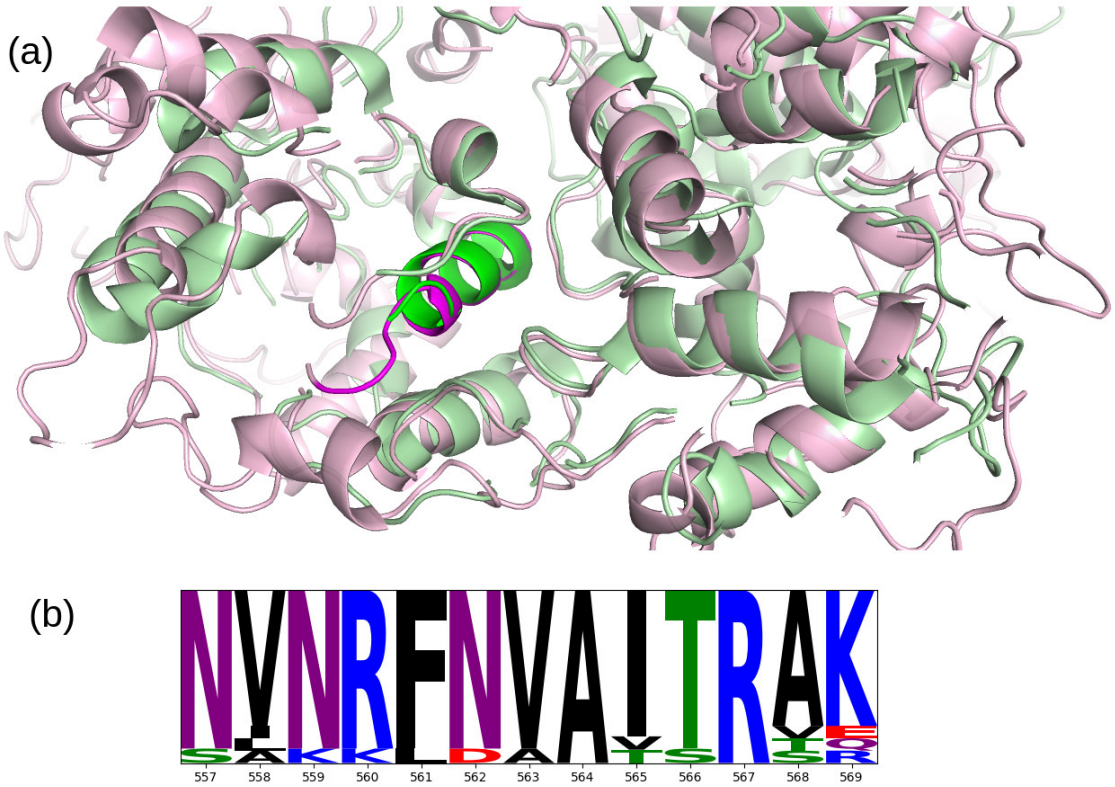

Figure 4.7: (a) Superposition of the crystallographic structure of NSP13 (PDB entry 6ZSL, chain B) (green) with the predicted AlphaFold structure of MOV10 (UniProt entry Q9HCE1) (pink), the highlighted α-helices in the center correspond to residues 557-569 of NSP13 and 901-913 of MOV10. Superposition and image were generated with PyMOL (`https://www.pymol.org/`). (b) Logo plot of the multiple sequence alignment for the NSP13 epitope. To facilitate visualization, a pseudocount of 0.1 is used and a min-max normalization of each position is applied. Image generated with the logomaker python package [Tareen and Kinney, 2020]

.

4.5. MMPred applied to VZV and GCA

The analysis produced 44 predicted autoantigens, 34 using Psi-BLAST, 8 using BLASTP, and 12 by PSI-BLAST or BLASTP. The results of the prediction are summarized in Table 7.11.

The system biology analysis produced significant results. Out of the 54 predicted autoantigens, 32 show a $Perc(S) > 95$ for at least one motif. All motifs have at least one protein scoring higher than this threshold. When considering a $Perc(S) > 99$, 19 predicted autoantigens satisfy the threshold for five different motifs. In particular, BRAF, LIMK1, and RAF1 relate to "Aortic Aneurysm Formation"; LCK, MAPK1, MAPK3, PDGFRA, and SYK to "Dysfunction of Immune Checkpoints", MAPKAPK3 to "Giant Cell Arteritis (GCA)"; KIT, MAPK12, MAPK13, MAPK7, and SYK to "Vascular Inflamation Early"; and DDR1, DDR2, EPHA2, EPHAB1, EPHB3, PDGFRB, and SYK to "Vascular Inflammation Amplification Cascade". Only the two genes KIT and SYK show a $Perc(S) > 99.9$, with KIT relating to "Vascular Inflammation – Early" and SYK to "Dysfunction of Immune check-points". The results are summarized in Figure 4.8.

Previous knowledge about the predicted autoantigens

None of the predicted autoepitopes and their similar microbial sequences have been previously identified to cross-link with DRB1*04:01 at experimental level according to IEDB (date: 25/06/2024).

Among the predicted autoantigens, the genes MAPK3(ERK1), MAPK1(ERK2) also appear in the molecular description of GCA, both for the motif "Aortic Aneurysm Formation" and identified by Barbour et al.. Is interesting to note that MAPK3 and MAPK1 are also known to be upregulated in Varicella Zoster Virus infected cells [Liu et al., 2012] increasing the likelihood of these two proteins being autoantigens within the context of GCA.

For all the other predicted autoantigens only PDGFRB appears in previous studies, which is identified among the hub genes that could relate to ischemic stroke in relation to VZV replication according to bioinformatic analysis [Wang et al., 2023].

The protein 66

All the alignments occur in the six VZV proteins gE (74 to 83), ORF13(149 to 164 and 186 to 201), ORF28 (686 to 701 and 888 to 895), ORF59 (242 to 254), ORF62 (346 to 351) and 66 (191 to 209). Is interesting to observe that most of the predicted autoantigens align with the VZV protein 66, with a total of 44 predicted autoantigens aligning with this VZV protein. Among these 44 proteins, 19 satisfy the treshold $Perc(S) > 99$, BRAF, DDR1, DDR2, EPHA2, EPHB1, EPHB3, KIT, LCK, LIMK1, MAPK1, MAPK12, MAPK13, MAPK3, MAPK7, MAPKAPK3, PDGFRA, PDGFRB, RAF1, and SYK. It also must be noted that SYK and KIT are the only two proteins to satisfy $Perc(S) > 99.9$. Furthermore the 9 predicted autoantigens not aligning with this protein score poorly in the ANNs analysis, with only TYMS, POLA1, and EMD satisfying $Perc(S) > 95$, making the case for the protein 66 to be among the VZV proteins inducing autoimmunity in Giant Cell Arteritis.

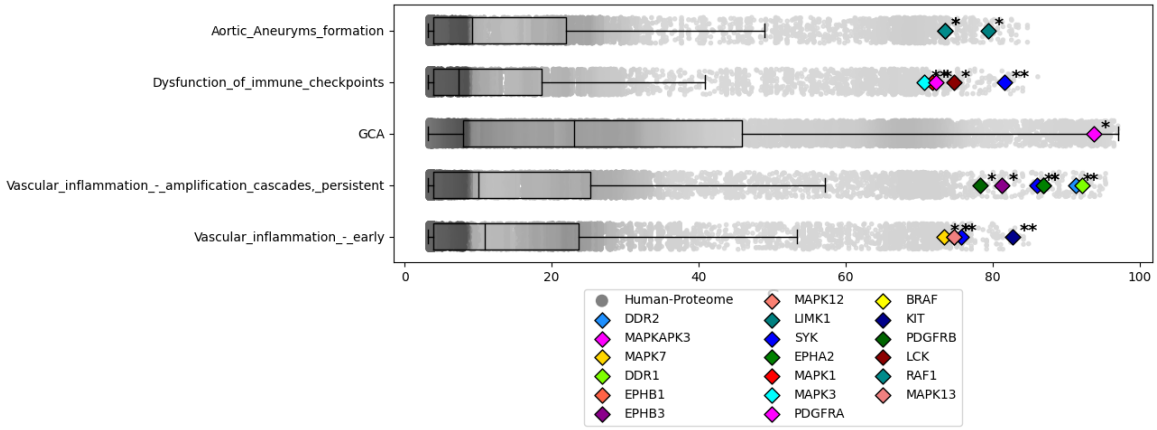

Figure 4.8: The image show the results of the ANNs analysis for the predicted autoantigens of the VZV vs Human-protome experiment. For each of the motifs tested a boxplot with overlapped scatterplot represent the background distribution of S, the predicted autoantigens that satisfies $Perc(S) > 99$ are shown. Those that shows $Perc(S) > 99$ are marked with *, and those that show $Perc(S) > 99.9$ are marked with **.

Chapter 5

General Discussion & Outlook

This thesis has explored statistics, bioinformatics, and systems biology approaches to deepen our understanding of Giant Cell Arteritis. The results contribute to the comprehension of the disease providing (i) a comprehensive molecular description of the pathophysiological pathways related to the condition, (ii) a further understanding of its pharmacological treatment regarding the side effects of prednisone and the predictor of response for tocilizumab, and *(iii)* the development of the molecular mimicry predictor MMPred, which identified different autoantigens that could be responsible for the insurgence of the disease in relation to VZV infections and also identified interesting autoantigens in the context of autoimmune disease related to SARS-CoV-2 infections.

The analysis of prednisone's adverse effects highlighted several significant associations, including fibromyalgia, muscle spasms, hypernatremia, hypophosphatemia, and breast neoplasms, with different entities across the different populations taken into account.

Using systems biology analysis, the study further explored prednisone's mechanisms, identifying molecular links between prednisone receptors and specific motifs associated with adverse effects. While in some cases the association regards known side effects, for others the association is uncertain according to previous knowledge. Among the novel and interesting relations identified, reduced bone resorption (linked to hypophosphatemia) and neurotransmitter imbalances (related to neuropathic pain and circadian rhythm disruptions) are the most interesting ones. For each of these associations, a set of proteins whose misregulation could be linked to the insurgence of the adverse effect is provided. This part of the work provides a starting point for further understanding of the molecular mechanism underlying the occurrences of these side effects. On the other side, limitations of this part of the work regard a lack of information on doses and path of administration of the drug and, in some cases, the sample size studied.

The study examined Tocilizumab's impact on Peripheral Blood Mononuclear Cells (PBMCs) from GCA patients by comparing gene expression profiles under three conditions: untreated, IL-6-stimulated, and IL-6 combined with Tocilizumab. Differential expression analysis identified 81 significantly modulated genes in the IL-6 + Tocilizumab group compared to IL-6 alone. Key findings included the downregulation of STAT3, MYC, and HIF1A in the presence of Tocilizumab, suggesting an inhibitory effect on IL-6-driven pathways, which are crucial in GCA inflammation. Furthermore, an important relationship between baseline levels of STAT3 and MYC in the modulation of the response to Tocilizumab has been detected by means of clustering and classification analysis, suggesting that these molecular markers could play a predictive role in determining individual patient responses to treatment. This study is limited by the relatively small number of genes analyzed, focusing only on key inflammatory pathways. Future research could expand the gene panel to include a broader range of genes involved in immune regulation and inflammatory signaling pathways, as well as explore the potential interactions between STAT3, MYC, and other relevant biomarkers. A systems biology approach using the Therapeutic Performance Mapping System (TPMS) further elucidated Tocilizumab's mechanism of action by simulating the effects of IL-6 receptor inhibition on downstream immune signaling pathways. Inhibiting IL-6RA led to reduced activation of pro-inflammatory cytokines and chemokines such as CCL2, IL21, and IL23A, potentially contributing to decreased inflammation and vascular remodeling in GCA.

MMPred was presented and its capability to identify biologically relevant results was demonstrated, using both BLASTP and PSI-BLAST as alignment methods. Interestingly, an interesting relation between the likelihood of a protein to be an autoantigen and the functional relation of the same protein to the motif related to a wide range of autoimmune diseases was identified. While the statistical evaluation of the method was not possible because of the lack of resources reporting experimental evidence for molecular mimicry about HLA-II, biologically relevant results were produced in the context of SARS-CoV-2 and VZV infections. For SARS-CoV-2, MMPred identified 15 potential autoantigens, with MOV10 being the most interesting one for its strong functional relation with many autoimmune conditions but also for its anti-viral role and its similarity with the SARS-CoV-2 protein NSP13, suggesting a potential immune-escape mechanism.

For VZV, MMPred detected multiple autoantigens, with protein 66 from VZV aligning with 44 predicted human autoantigens related to GCA. Among these, SYK and KIT emerged because of the strong functional relationship with the motifs of "vascular inflammation - early" and "Dysfunction of immune check-points". Furthermore, the protein 66 also aligns with MAPK3 and MAPK1, two proteins known to be highly expressed in GCA aortas and involved in the motif "Aortic aneurysm formation".

This analysis strongly supports VZV's possible role as a trigger in GCA, driven by molecular mimicry between viral and host proteins. Further studies could unveil the role of the identified autoantigens both with in-silico and in-vivo experiments. The relation between any

other infective microorganism and autoimmune disease can be explored using MMPred in a systematic way. At last, MMPred can be further extended to include any other linear epitope predictor for HLA-I and HLA-II molecules and T and B cell receptors.

Chapter 6

Conclusions

Through a systematic review of recent literature, the current knowledge of the pathophysiology of GCA has been extracted from a molecular perspective. Pathways related to vascular inflammation, immune checkpoint dysfunction, and arterial remodeling have been described, helping future researchers in a deeper understanding of the pathology.

The side effects of Prednisone have been explored by mining real-world-data and producing real-world-evidence. In addition, an in-silico system biology method has been applied to identify human proteins that can be related to the insurgence of certain side effects and provide the background to formulate hypotheses at a molecular level.

By mining gene-expression data from ex-vivo models of PBMCs, STAT3 and MYC have been identified as crucial genes in modulating response to Tocilizumab, offering two key biomarkers for patient stratification in further studies.

The bioinformatics tool MMPred was presented, and evaluated and its capability to produce biologically relevant results has been demonstrated. In particular, the autoantigen MOV10 has been identified as a potential trigger of autoimmunity in the context of SARS-CoV 2 infections, and the interplay with NSP13 could be at the base of an immune-escape mechanism from the virus. In the case of Varicella Zoster Virus, the viral protein 66 showed high similarity with many human proteins among which SYK and KYT show a strong functional relation with the pathophysiology of GCA, and MAPK3 and MAPk1 which are known to be upregulated in VZV-infected cells and be involved in aortic aneurysm formation in GCA patients.

Bibliography

Alan A. *Categorical Data Analysis*. John Wiley & Sons, Hoboken, NJ, 2nd edition, 2002. ISBN 978-0-471-36093-3. See Section 4.3 for Wald Test.

M. Abadi, A. Agarwal, P. Barham, E. Brevdo, Z. Chen, C. Citro, G. S. Corrado, A. Davis, J. Dean, M. Devin, S. Ghemawat, I. Goodfellow, A. Harp, G. Irving, M. Isard, Y. Jia, R. Jozefowicz, L. Kaiser, M. Kudlur, J. Levenberg, D. Mané, R. Monga, S. Moore, D. Murray, C. Olah, M. Schuster, J. Shlens, B. Steiner, I. Sutskever, K. Talwar, P. Tucker, V. Vanhoucke, V. Vasudevan, F. Viégas, O. Vinyals, P. Warden, M. Wattenberg, M. Wicke, Yuan Yu, and X. Zheng. TensorFlow: Large-scale Machine Learning on heterogeneous distributed systems. *arXiv*, page 1603.04467v2, 2016. doi: 10.48550/arXiv.1603.04467.

A. Abendroth and A. Arvin. Immune evasion mechanisms of varicella-zoster virus. *Arch. Virol. Suppl.*, 1(17):99–107, 2001.

A. Abendroth, P. R. Kinchington, and B. Slobedman. Varicella zoster virus immune evasion strategies. *Curr. Top. Microbiol. Immunol.*, 342:155–171, 2010.

A. Z. Al-Mousawi, S. P. Gurney, A. R. Lorenzi, U. Pohl, M. Dayan, and S. P. Mollan. Reviewing the Pathophysiology Behind the Advances in the Management of Giant Cell Arteritis. *Ophthalmol Ther*, 8(2):177–193, Jun 2019.

A. Alcina, M. M. Abad-Grau, M. Fedetz, G. Izquierdo, M. Lucas, O. Fernández, D. Ndagire, A. Catalá-Rabasa, A. Ruiz, J. Gayán, C. Delgado, C. Arnal, and F. Matesanz. Multiple sclerosis risk variant HLA-DRB1*1501 associates with high expression of DRB1 gene in different human populations. *PLoS One*, 7(1):e29819, 2012. doi: 10.1371/journal.pone.0029819.

R. Alten and E. Wiebe. Hypothalamic-pituitary-adrenal axis function in patients with rheumatoid arthritis treated with different glucocorticoid approaches. *Neuroimmunomodulation*, 22(1-2):83–88, 2015.

S. F. Altschul, W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. Basic local alignment search tool. *J. Mol. Biol.*, 215(3):403–410, 1990. doi: 10.1016/S0022-2836(05)80360-2.

M. Alvaro-Benito and C. Freund. Revisiting nonclassical HLA II functions in antigen presentation: Peptide editing and its modulation. *HLA*, 96(4):415–429, Oct 2020.

C. Amlie-Lefond and D. Gilden. Varicella zoster virus: A common cause of stroke in children and adults. *J. Stroke Cerebrovasc. Dis.*, 25(7):1561–1569, July 2016.

H. An, M. Eun, J. Yi, and J. Park. CRESSP: A comprehensive pipeline for prediction of immunopathogenic SARS-CoV-2 epitopes using structural properties of proteins. *Brief. Bioinform.*, 23(2):bbac056, 2022. doi: 10.1093/bib/bbac056.

M. L. Apperson, Y. Tian, B. Stamova, B. P. Ander, G. C. Jickling, M. A. Agius, and F. R. Sharp. Genome wide differences of gene expression associated with HLA-DRB1 genotype in multiple sclerosis: a pilot study. *J. Neuroimmunol.*, 257(1-2):90–96, 2013. doi: 10.1016/j.jneuroim.2013.02.004.

Y. Ariumi. Host cellular RNA helicases regulate SARS-CoV-2 infection. *J. Virol.*, 96(6):e0000222, 2022. doi: 10.1128/jvi.00002-22.

Hopkins B. and Skellam J. G. A new method for determining the type of distribution of plant individuals. *Annals of Botany*, 18:213–227, 1954. URL `https://api.semanticscholar.org/CorpusID:89564582`.

C. A. Balinsky, H. Schmeisser, A. I. Wells, S. Ganesan, T. Jin, K. Singh, and K. C. Zoon. IRAV (FLJ11286), an interferon-stimulated gene with antiviral activity against dengue virus, interacts with MOV10. *J. Virol.*, 91(5):e01606–16, 2017. doi: 10.1128/JVI.01606-16.

Y. Banz and J. H. Stone. Why do temporal arteries go wrong? Principles and pearls from a clinician and a pathologist. *Rheumatology (Oxford)*, 57(suppl_2):ii3–ii10, Feb 2018.

J. R. Barbour, F. G. Spinale, and J. S. Ikonomidis. Proteinase systems and thoracic aortic aneurysm progression. *J Surg Res*, 139(2):292–307, May 2007.

C. Barra, J. B. Nilsson, A. Saksager, I. Carri, S. Deleuran, H. Garcia Alvarez, Ma. H. Høie, Y. Li, J. N. Clifford, Y. R. Wan, L. S. Moreta, and M. Nielsen. In silico tools for predicting novel epitopes. *Methods Mol. Biol.*, 2813:245–280, 2024.

J. Barrios and X. Ai. Neurotrophins in asthma. *Curr. Allergy Asthma Rep.*, 18(2):10, February 2018.

G. J. Barton. Protein sequence alignment techniques. *Acta Crystallogr. D Biol. Crystallogr.*, 54(Pt 6 Pt 1):1139–1146, November 1998.

Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. Roy. Statist. Soc. Ser*, 57:289–300, 1995.

C. S. Bergamin and S. A. Dib. Enterovirus and type 1 diabetes: What is the matter? *World J. Diabetes*, 6(6):828–839, June 2015.

M. A. Berryman, J. Ilonen, E. W. Triplett, and Johnny Ludvigsson. Important denominator between autoimmune comorbidities: a review of class II HLA, autoimmune disease, and the gut. *Front. Immunol.*, 14:1270488, September 2023.

C. J. Boos, M. Nam, and A. J. Camm. Novel oral anticoagulants and stroke prevention in atrial fibrillation and chronic heart failure. *Heart Fail. Rev.*, 19(3):391–401, May 2014.

H. R. Braddy, F. Ryan, J. Cunningham, W. Tormey, M. P. Ryan, and R. O'Neill. Hypophosphatemia complicating bronchodilator therapy for acute severe asthma. *Arch Intern Med"*, 149(10):2367–2368, October 1989.

B. Burja, T. Kuret, S. Sodin-Semrl, K. Lakota, Ž. Rotar, R. Ješe, K. Mrak-Poljšak, P. Žigon, G. G. Thallinger, J. Feichtinger, S. Čučnik, M. Tomšič, S. Praprotnik, and A. Hočevar. A concise review of significantly modified serological biomarkers in giant cell arteritis, as detected by different methods. *Autoimmun Rev*, 17(2):188–194, Feb 2018.

C. A. Butler and L. G. Heaney. Neurogenic inflammation and asthma. *Inflamm. Allergy Drug Targets*, 6(2):127–132, June 2007.

P. Cacoub and B. Terrier. Hepatitis b-related autoimmune manifestations. *Rheum. Dis. Clin. North Am.*, 35(1):125–137, February 2009.

M. Calderón-Goercke, J. Loricera, V. Aldasoro, S. Castañeda, I. Villa, A. Humbría, C. Moriano, S. Romero-Yuste, J. Narváez, C. Gómez-Arango, E. Pérez-Pampín, R. Melero, E. Becerra-Fernández, M. Revenga, N. Álvarez-Rivas, C. Galisteo, F. Sivera, A. Olivé-Marqués, M. Álvarez Del Buergo, L. Marena-Rojas, C. Fernández-López, F. Navarro, E. Raya, E. Galindez-Agirregoikoa, B. Arca, R. Solans-Laqué, A. Conesa, C. Hidalgo, C. Vázquez, J. A. Román-Ivorra, P. Lluch, S. Manrique-Arija, P. Vela, E. De Miguel, C. Torres-Martín, J. C. Nieto, C. Ordas-Calvo, E. Salgado-Pérez, C. Luna-Gomez, F. J. Toyos-Sáenz de Miera, N. Fernández-Llanio, A. García, C. Larena, N. Palmou-Fontana, V. Calvo-Río, D. Prieto-Peña, C. González-Vela, A. Corrales, M. Varela-García, E. Aurrecoechea, R. Dos Santos, Á. García-Manzanares, N. Ortego, S. Fernández, F. Ortiz-Sanjuán, Montserrat Corteguera, J. L. Hernández, M. Á. González-Gay, and R. Blanco.

Tocilizumab in giant cell arteritis. observational, open-label multicenter study of 134 patients in clinical practice. *Semin. Arthritis Rheum.*, 49(1):126–135, August 2019.

Raphael Carapito, Mirjana Radosavljevic, and Seiamak Bahram. Next-Generation sequencing of the HLA locus: Methods and impacts on HLA typing, population genetics and disease association studies. *Hum. Immunol.*, 77(11):1016–1023, November 2016.

F. D. Carmona, M. A. Miguel A González-Gay, and J. Martín. Genetic component of giant cell arteritis. *Rheumatology (Oxford)*, 53(1):6–18, Jan 2014.

F. D. Carmona, S. L. Mackie, J. E. Martín, J. C. Taylor, A. Vaglio, S. Eyre, L. Bossini-Castillo, S. Castañeda, M. C. Cid, J. Hernández-Rodríguez, S. Prieto-González, R. Solans, M. Ramentol-Sintas, M. F. González-Escribano, L. Ortiz-Fernández, I. C. Morado, J. Narváez, J. A. Miranda-Filloy, Spanish GCA Group, L. Beretta, C. Lunardi, M. A Cimmino, D. Gianfreda, D. Santilli, G. A. Ramirez, A. Soriano, F. Muratore, G. Pazzola, O. Addimanda, C. Wijmenga, T. Witte, J. H. Schirmer, F. Moosig, V. Schönau, A. Franke, Ø. Palm, Ø. Molberg, A. P. Diamantopoulos, S. Carette, D. Cuthbertson, L. J. Forbess, G. S. Hoffman, N. A. Khalidi, C. L. Koening, C. A. Langford, C. A. McAlear, L. Moreland, P. A. Monach, C. Pagnoux, P. Seo, R. Spiera, A. G. Sreih, K. J. Warrington, S. R. Ytterberg, P. K. Gregersen, C. T. Pease, A. Gough, M. Green, L. Hordon, S. Jarrett, R. Watts, S. Levy, Y. Patel, S. Kamath, B. Dasgupta, J. Worthington, B. P. C. Koeleman, P. I. W. de Bakker, J. H. Barrett, C. Salvarani, P. A. Merkel, M. A. González-Gay, A. W. Morgan, and J. Martín. A large-scale genetic analysis reveals a strong contribution of the HLA class II region to giant cell arteritis susceptibility. *Am J Hum Genet*, 96(4):565–580, Apr 2015.

F. D. Carmona, P. Coit, G. Saruhan-Direskeneli, J. Hernández-Rodrguez, M. C. Cid, R. Solans, S. Castañeda, A. Vaglio, H. Direskeneli, P. A. Merkel, L. Boiardi, C. Salvarani, M. A. González-Gay, J. Martín, A. H. Sawalha, A. Martínez-Berriochoa, A. Unzurrunzaga, A. Hidalgo-Conde, A. B. Vuelta, A. Fernández-Nebro, M. C. Ordóñez-Cañizares, B. Fernández-Gutiérrrez, L. Rodríguez-Rodríguez, B. Escalante, B. Marí-Alfonso, B. Sopeña, C. Carmen Gómez-Vaquero, E. Raya, E. Grau, J. A. Román, E. F. Vicente, E. de Miguel, F. J. López-Longo, L. Martínez, I. C. Morado, J. B. Díaz-López, L. Caminal-Montero, A. Martínez-Zapico, J. Narváez, J. Monfort, L. Tío, J. A. Miranda-Filloy, J. Sánchez-Martín, J. J. Alegre-Sancho, L. Sáez-Comet, M. Pérez-Conesa, M. Corbera-Bellalta, M. Ramentol-Sintas, M. J. García-Villanueva, M. Guijarro Rojas, N. Ortego-Centeno, R. Raquel Ríos Fernández, J. L. Callejas, O. Sanchez Pernaute, P. Fanlo Mateo, R. Blanco, S. Sergio Prieto-González, V. M. Martínez-Taboada, A. Soriano, C. Lunardi, D. Gianfreda, D. Santilli, F. Bonatti, F. Muratore, G. Pazzola, O. Addimanda, G. Emmi, G. A. Ramirez, L. Beretta, M. Govoni, M. A. Cimmino, A. Mesut Onat, A. Cefle, A. Yazici, B. Kısacık, E. Dalkilic, E. Seyahi, I. Fresko, E. Tunc, E. Erken, H. Te Ozer, K. Aksu, G. Keser, M. A. Ozturk, M. Bıcakcıgil, N. Duzgun, O. Karadag, S. Kiraz, N. Ö. Pamuk, S. Akar, F. Onen, N. Akkoc, S. Kamali, M. Inanc, S. P. Yentür, S. Z.

Aydin, F. Alibaz-Oner, T. Kaşifoğlu, V. Cobankara, Z. Ozbalkan, A. Ates, Y. Karaaslan, S. Carette, S. A. Chung, D. Cuthbertson, L. J. Forbess, G. S. Hoffman, N. A. Khalidi, C. L. Koening, C. A. Langford, C. A. McAlear, K. McKinnon-Maksimowicz, P. A. Monach, L. Moreland, C. Pagnoux, P. Seo, R. Spiera, A. G. Sreih, K. J. Warrington, and S. R. Ytterberg. Analysis of the common genetic component of large-vessel vasculitides through a meta-Immunochip strategy. *Sci Rep*, 7:43953, Mar 2017a.

F. D. Carmona, A. Vaglio, S. L. Mackie, J. Hernández-Rodríguez, P. A. Monach, S. Castañeda, R. Solans, I. C. Morado, J. Narváez, M. Ramentol-Sintas, C. T. Pease, B. Dasgupta, R. Watts, N. Khalidi, C. A. Langford, S. Ytterberg, L. Boiardi, L. Beretta, M. Govoni, G. Emmi, F. Bonatti, M. A. Cimmino, T. Witte, T. Neumann, J. Holle, V. Schönau, L. Sailler, T. Papo, J. Haroche, A. Mahr, L. Mouthon, Ø. Molberg, A. P. Diamantopoulos, A. Voskuyl, E. Brouwer, T. Daikeler, C. T. Berger, E. S. Molloy, L. O'Neill, D. Blockmans, B. A. Lie, P. Mclaren, T. J. Vyse, C. Wijmenga, Y. Allanore, B. P. C. Koeleman, J. H. Barrett, M. C. Cid, C. Salvarani, P. A. Merkel, A. W. Morgan, M. A. González-Gay, J. Martín, J. L. Callejas, L. Caminal-Montero, M. Corbera-Bellalta, E. de Miguel, J. B. D. López, M. J. García-Villanueva, C. Gómez-Vaquero, M. Guijarro-Rojas, A. Hidalgo-Conde, B. Marí-Alfonso, A. M. Berriochoa, A. M. Zapico, V. M. Martínez-Taboada, J. A. Miranda-Filloy, J. Monfort, N. Ortego-Centeno, M. Pérez-Conesa, S. Prieto-González, E. Raya, R. R. Fernández, J. Sánchez-Martín, B. Sopeña, L. Tío, A. Unzurrunzaga, A. Gough, J. D. Isaacs, M. Green, N. McHugh, L. Hordon, S. Kamath, M. Nisar, Y. Patel, C. S. Yee, R. Stevens, P. Nandi, A. Nandagudi, S. Jarrett, C. Li, S. Levy, S. Mollan, A. Salih, O. Wordsworth, E. Sanders, E. Roads, A. Gill, L. Carr, C. Routledge, K. Culfear, A. Nugaliyadde, L. James, J. Spimpolo, A. Kempa, F. Mackenzie, R. Fong, G. Peters, B. Rowbotham, Z. Masqood, J. Hollywood, P. Gondo, R. Wood, S. Martin, L. H. Rashid, J. I. Robinson, M. Morgan, L. Sorensen, J. Taylor, S. Carette, S. Chung, D. Cuthbertson, L. J. Forbess, O. Gewurz-Singer, G. S. Hoffman, C. L. Koening, K. M. Maksimowicz-McKinnon, C. A. McAlear, L. W. Moreland, C. Pagnoux, P. Seo, U. Specks, R. F. Spiera, A. Sreih, K. J. Warrington, and M. Weisman. A Genome-wide Association Study Identifies Risk Alleles in Plasminogen and P4HA2 Associated with Giant Cell Arteritis. *Am J Hum Genet*, 100(1):64–74, Jan 2017b.

E. Caron, D. J. Kowalewski, C. Chiek Koh, T. Sturm, H. Schuster, and R. Aebersold. Analysis of Major Histocompatibility Complex (MHC) immunopeptidomes using mass spectrometry. *Mol. Cell. Proteomics*, 14(12):3105–3117, 2015. doi: 10.1074/mcp.O115.052431.

G. Carvajal Alegria, Y. van Sleen, J. C. Graver, M. Sandovici, V. Devauchelle-Pensec, E. Brouwer, and D. Cornec. Aortic involvement in giant cell arteritis. *Joint Bone Spine*, 88(2):105045, Mar 2021.

R. M. S. Chakr, C. Brenol, A. Ranzolin, A. Bernardes, A. P. Dalosto, G. Ferrari, S. Scalco, V. Olszewski, C. Kohem, O. Monticielo, J. C. T. Brenol, and R. M. Xavier. Rheumatoid

arthritis seems to have DMARD treatment decision influenced by fibromyalgia. *Rev. Bras. Reumatol. Engl. Ed.*, 57(5):403–411, September 2017.

A. Chatzidiontsiou. The lung in rheumatoid arthritis, cause or consequence? erratum. *Curr. Opin. Rheumatol.*, 28(2):1, March 2016.

M. C. Cid. 3. pathogenesis of giant cell arteritis. *Rheumatology*, 53(suppl 2):i2–i3, July 2014. ISSN 1462-0332. doi: 10.1093/rheumatology/keu184. URL `http://dx.doi.org/10.1093/rheumatology/keu184`.

M. C. Cid, E. Campo, G. Ercilla, A. Palacin, J. Vilaseca, J. Villalta, and M. Ingelmo. Immunohistochemical analysis of lymphoid and macrophage cell subsets and their immunologic activation markers in temporal arteritis. Influence of corticosteroid treatment. *Arthritis Rheum*, 32(7):884–893, Jul 1989.

M. C. Cid, D. S. Grant, G. S. Hoffman, R. Auerbach, A. S. Fauci, and H. K. Kleinman. Identification of haptoglobin as an angiogenic factor in sera from patients with systemic vasculitis. *J Clin Invest*, 91(3):977–985, Mar 1993.

P. Coit, L. B. De Lott, B. Nan, V. M. Elner, and A. H. Sawalha. DNA methylation analysis of the temporal artery microenvironment in giant cell arteritis. *Ann Rheum Dis*, 75(6): 1196–1202, Jun 2016.

M. Collatz, F. Mock, E. Barth, M. Hölzer, K. Sachse, and M. Marz. EpiDope: a deep neural network for linear b-cell epitope prediction. *Bioinformatics*, 37(4):448–455, May 2021.

L. Collesano, M. Łuksza, and M. Lässig. Energy landscapes of peptide-MHC binding. *PLoS Comput. Biol.*, 20(9):e1012380, September 2024.

G. W. Corder and D. I. Foreman. *Nonparametric statistics*. John Wiley & Sons, Nashville, TN, 2 edition, May 2014.

P. Coto-Segura, C. Segú-Vergés, A. Martorell, D. Moreno-Ramírez, G. Jorba, V. Junet, F. Guerri, X. Daura, B. Oliva, C. Cara, O. Suárez-Magdalena, S. Abraham, and J. M. Mas. A quantitative systems pharmacology model for certolizumab pegol treatment in moderate-to-severe psoriasis. *Front Immunol*, 14:1212981, 2023.

R. A. Cuevas, A. Ghosh, C. Wallerath, V. Hornung, C. B. Coyne, and S. N. Sarkar. MOV10 provides antiviral activity against RNA viruses by enhancing RIG-I - MAVS-independent IFN induction. *J. Immunol.*, 196(9):3877–3886, 2016. doi: 10.4049/jimmunol.1501359.

R. T. Damian. Molecular mimicry: antigen sharing by parasite and host and its consequences.

The American Naturalist, 98(900):129–149, 1964.

R. Dammacco, G. Alessio, E. Giancipoli, P. Leone, A. Cirulli, L. Resta, A. Vacca, and F. Dammacco. Giant Cell Arteritis: The Experience of Two Collaborative Referral Centers and an Overview of Disease Pathogenesis and Therapeutic Advancements. *Clin Ophthalmol*, 14:775–793, 2020.

B. J. Debebe, L. Boelen, J. C. Lee, IAVI Protocol C Investigators, C. L. Thio, J. Astemborski, G. Kirk, S. I. Khakoo, S. M. Donfield, J. J. Goedert, and B. Asquith. Identifying the immune interactions underlying HLA class I disease associations. *Elife*, 9, April 2020.

C. Dejaco, E. Brouwer, J. C. Mason, F. Buttgereit, E. L. Matteson, and B. Dasgupta. Giant cell arteritis and polymyalgia rheumatica: current challenges and opportunities. *Nat Rev Rheumatol*, 13(10):578–592, Oct 2017.

S. Deorowicz, A. Debudaj-Grabysz, and A. Gudyś. FAMSA: Fast and accurate multiple sequence alignment of huge protein families. *Sci. Rep.*, 6:33964, 2016. doi: 10.1038/srep33964.

S. Deshayes, H. de Boysson, A. Dumont, D. Vivien, A. Manrique, and A. Aouba. An overview of the perspectives on experimental models and new therapeutic targets in giant cell arteritis. *Autoimmun Rev*, 19(10):102636, Oct 2020.

D. Dholakia, A. Kalra, B. R. Misir, U. Kanga, and M. Mukerji. HLA-SPREAD: a natural language processing based resource for curating HLA association from PubMed abstracts. *BMC Genomics*, 23(1):10, Jan 2022. [PubMed Central:PMC6602571] [DOI:10.1016/j.berh.2020.101499] [PubMed:32279929].

A. Di Bucchianico. Combinatorics, computer algebra and the Wilcoxon-Mann-Whitney test. *J. Stat. Plan. Inference*, 79(2):349–364, 1999. doi: 10.1016/S0378-3758(98)00261-4.

M. Dinkin and E. Johnson. One giant step for giant cell arteritis: Updates in diagnosis and treatment. *Curr. Treat. Options Neurol.*, 23(2):6, January 2021.

A. C. Doxey and B. J. McConkey. Prediction of molecular mimicry candidates in human pathogenic bacteria. *Virulence*, 4(6):453–466, 2013. doi: 10.4161/viru.25180.

L. Du, F. Deiter, M. S. Bouzidi, J. Billaud, G. Simmons, P. Dabral, S. Selvarajah, A. F. Lingappa, M. Michon, S. F. Yu, K Paulvannan, B Manicassamy, V. R. Lingappa, H. Boushey, J. R. Greenland, and S. K. Pillai. A viral assembly inhibitor blocks SARS-CoV-2 replication in airway epithelial cells. *Commun. Biol.*, 7:486, 2024. doi: 10.1038/s42003-024-06130-8.

J. C. Dunn. A fuzzy relative of the isodata process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3):32–57, January 1973. ISSN 0022-0280. doi: 10.1080/01969727308546046. URL `http://dx.doi.org/10.1080/01969727308546046`.

Knuth D. E. *The Art of Computer Programming, Volume 2: Seminumerical Algorithms*. Addison-Wesley, Reading, MA, 3rd edition, 1997. ISBN 978-0-201-89684-8. See Section 3.4.1 for Random Sampling and Weighted Sampling.

R. C. Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, 32(5):1792–1797, March 2004.

M. Ehrenfeld, A. Tincani, L. Andreoli, M. Cattalini, A. Greenbaum, D. Kanduc, J. Alijotas-Reig, V. Zinserling, N. Semenova, H. Amital, and Y. Shoenfeld. Covid-19 and autoimmunity. *Autoimmun. Rev.*, 19(8):102597, 2020. doi: 10.1016/j.autrev.2020.102597.

B. R. England, T. R. Mikuls, F. Xie, S. Yang, L. Chen, and J. R. Curtis. Herpes zoster as a risk factor for incident giant cell arteritis. *Arthritis Rheumatol.*, 69(12):2351–2358, December 2017.

G. Espígol-Frigolé, M. Corbera-Bellalta, E. Planas-Rigol, E. Lozano, M. Segarra, A. García-Martínez, S. Prieto-González, J. Hernández-Rodríguez, J. M. Grau, M. U. Rahman, and M. C. Cid. Increased IL-17A expression in temporal artery lesions is a predictor of sustained response to glucocorticoid treatment in patients with giant-cell arteritis. *Ann Rheum Dis*, 72(9):1481–1487, Sep 2013.

K. Y. Fang, G. N. Liang, Z. Q. Zhuang, Y. X. Fang, Y. Q. Dong, C. J. Liang, X. Y. Chen, and X. G. Guo. Screening the hub genes and analyzing the mechanisms in discharged COVID-19 patients retesting positive through bioinformatics analysis. *J. Clin. Lab. Anal.*, 36(7):e24495, 2022. doi: 10.1002/jcla.24495.

N. Farina, A. Tomelleri, C. Campochiaro, and L. Dagna. Giant cell arteritis: Update on clinical manifestations, diagnosis, and management. *Eur. J. Intern. Med.*, 107:17–26, January 2023.

O. B. Fasmer, A. Halmøy, T. M. Eagan, K. J. Oedegaard, and J. Haavik. Adult attention deficit hyperactivity disorder is associated with asthma. *BMC Psychiatry*, 11(1):128, August 2011.

T. P. Finn, R. E. Jones, C. Rich, R. Dahan, J. Link, C. S. David, Y. K. Chou, H. Offner, and A. A. Vandenbark. HLA-DRB1*1501 risk association in multiple sclerosis may not be related to presentation of myelin epitopes. *J. Neurosci. Res.*, 78(1):100–114, 2004. doi: 10.1002/jnr.20227.

M. T. Fiorillo, F. Paladini, V. Tedeschi, and R. Sorrentino. HLA Class I or Class II and disease association: Catch the difference if you can. *Front. Immunol.*, 8:1475, 2017. doi: 10.3389/fimmu.2017.01475.

J. A. Fraser, C. M. Weyand, N. J. Newman, and V. Biousse. The treatment of giant cell arteritis. *Rev. Neurol. Dis.*, 5(3):140–152, 2008.

D. A. Freedman. *Statistical models*. Cambridge University Press, Cambridge, England, 2 edition, April 2009.

N. George, S. Fexova, A. M. Fuentes, P. Madrigal, Y. Bi, H. Iqbal, U. Kumbham, N. F. Nolte, L. Zhao, A. S. Thanki, I. D. Yu, J. C. Marugan Calles, K. Erdos, L. Vilmovsky, S. R. Kurri, A. Vathrakokoili-Pournara, D. Osumi-Sutherland, A. Prakash, S. Wang, M. K. Tello-Ruiz, S. Kumari, D. Ware, D. Goutte-Gattat, Y. Hu, N. Brown, N. Perrimon, J. A. no, T. Burdett, S. Teichmann, A. Brazma, and I. Papatheodorou. Expression Atlas update: insights from sequencing data at both bulk and single cell level. *Nucleic Acids Res.*, 52 (D1):D107–D114, 2024. doi: 10.1093/nar/gkad1021.

J. Gil, M. Marques-Pamies, M. Sampedro, S. M. Webb, G. Serra, I. Salinas, A. Blanco, E. Valassi, C. Carrato, A. Picó, A. García-Martínez, L. Martel-Duguech, T. Sardon, A. Simó-Servat, B. Biagetti, C. Villabona, R. Cámara, C. Fajardo-Montañana, C. Álvarez Escolá, C. Lamas, C. V. Alvarez, I. Bernabéu, M. Marazuela, M. Jordà, and M. Puig-Domingo. Data mining analyses for precision medicine in acromegaly: a proof of concept. *Sci Rep*, 12(1):8979, May 2022.

D. Gilden and M. A. Nagel. Varicella zoster virus triggers the immunopathology of giant cell arteritis. *Curr. Opin. Rheumatol.*, 28(4):376–382, July 2016.

J. Gómez-Rial, I. Rivero-Calle, A. Salas, and F Martinón-Torres. Rotavirus and autoimmunity. *J. Infect.*, 81(2):183–189, August 2020.

J. L. Goodier, L. E. Cheung, and H. H. Kazazian. MOV10 RNA helicase is a potent inhibitor of retrotransposition in cells. *PLoS Genet.*, 8(10):e1002941, 2012. doi: 10.1371/journal.pgen.1002941.

J. Hallas and P. Bytzer. Screening for drug related dyspepsia: an analysis of prescription symmetry. *Eur. J. Gastroenterol. Hepatol.*, 10(1):27–32, January 1998.

R. Harrington, S. A. Al Nokhatha, and R. Conway. Biologic therapies for giant cell arteritis. *Biologics*, 15:17–29, January 2021.

C. R. Harris, K. J. Millman, S. J. van der Walt, R. Gommers, P. Virtanen, D. Courna-

peau, E. Wieser, J. Taylor, S. Berg, N. J. Smith, R. Kern, M. Picus, S. Hoyer, M. H. van Kerkwijk, M. Brett, A. Haldane, J. Fernández del Río, M. Wiebe, P. Peterson, P. Gérard-Marchant, K. Sheppard, T. Reddy, W. Weckesser, H. Abbasi, C. Gohlke, and T. E. Oliphant. Array programming with NumPy. *Nature*, 585(7825):357–362, 2020. doi: 10.1038/s41586-020-2649-2.

E. L. Hatcher, S. A. Zhdanov, Y. Bao, O. Blinkova, E. P. Nawrocki, Y. Ostapchuck, A. A. Schäffer, and J. R. Brister. Virus Variation Resource - improved response to emergent viral outbreaks. *Nucleic Acids Res.*, 45(D1):D482–D490, 2017. doi: 10.1093/nar/gkw1065.

S. Henikoff and J. G. Henikoff. Amino acid substitution matrices from protein blocks. *Proc Natl Acad Sci U S A*, 89(22):10915–10919, November 1992.

J. Hernández-Rodríguez, A. García-Martínez, J. Casademont, X. Filella, M. J. Esteban, A. López-Soto, J. Fernández-Solà, A. Urbano-Márquez, J. M. Grau, and M. C. Cid. A strong initial systemic inflammatory response is associated with higher corticosteroid requirements and longer duration of therapy in patients with giant-cell arteritis. *Arthritis Rheum*, 47(1):29–35, Feb 2002.

R. Hid Cadena, W. H. Abdulahad, G. A. P. Hospers, T. T. Wind, A. M. H. Boots, P. Heeringa, and E. Brouwer. Checks and Balances in Autoimmune Vasculitis. *Front Immunol*, 9:315, 2018.

A. T. de Carvalho Hoepers, M. M. Menezes, and T. S. Fröde. Systematic review of anaemia and inflammatory markers in chronic obstructive pulmonary disease. *Clin. Exp. Pharmacol. Physiol.*, 42(3):231–239, March 2015.

M. Hoffmann, H. Kleine-Weber, S. Schroeder, N. Krüger, T. Herrler, S. Erichsen, T. S. Schiergens, G. Herrler, N. H. Wu, A. Nitsche, M. A. Müller, C. Drosten, and S. Pöhlmann. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell*, 181(2):271–280, 2020. doi: 10.1016/j.cell.2020.02.052.

M. H. Høie, F. S. Gade, J. M. Johansen, C. Würtzen, O. Winther, M. Nielsen, and P. Marcatili. DiscoTope-3.0: improved b-cell epitope prediction using inverse folding latent representations. *Front. Immunol.*, 15:1322712, February 2024.

B. D. Huisman, Z. Dai, D. K. Gifford, and M. E. Birnbaum. A high-throughput yeast display approach to profile pathogen proteomes for MHC-II binding. *eLife*, 11:e78589, 2022. doi: 10.7554/eLife.78589.

J. D. Hunter. Matplotlib: A 2D graphics environment. *Comput. Sci. Eng.*, 9(3):90–95, 2007. doi: 10.1109/MCSE.2007.55.

E. Huntzinger, J. Sinteff, B. Morlet, and B. Séraphin. HELZ2: a new, interferon-regulated, human 3'-5' exoribonuclease of the RNB family is expressed from a non-canonical initiation codon. *Nucleic Acids Res.*, 51(17):9279–9293, 2023. doi: 10.1093/nar/gkad673.

T. Imaizumi, M. Nakatochi, Y. Fujita, R. Yamamoto, K. Watanabe, M. Maekawa, T. Yamawaka, T. Katsuno, and S. Maruyama. Glucocorticoid treatment is associated with ICU-acquired hypernatremia: a nested case-control study. *Clin. Exp. Nephrol.*, 25(2):131–139, February 2021.

H. Ince and C. A. Nienaber. Etiology, pathogenesis and management of thoracic aortic aneurysm. *Nat Clin Pract Cardiovasc Med*, 4(8):418–427, Aug 2007.

R. Ishihara, R. Watanabe, M. Shiomi, M. Katsushima, K. Fukumoto, S. Yamada, T. Okano, and M. Hashimoto. Exploring the link between varicella-zoster virus, autoimmune diseases, and the role of recombinant zoster vaccine. *Biomolecules*, 14(7):739, June 2024.

N. V. Ivanisenko, T. I. Shashkova, A. Shevtsov, M. Sindeeva, D. Umerenkov, and O. Kardymon. SEMA 2.0: web-platform for b-cell conformational epitopes prediction using artificial intelligence. *Nucleic Acids Res.*, 52(W1):W533–W539, July 2024.

G. F. Joos, K. O. De Swert, V. Schelfhout, and R. A. Pauwels. The role of neural inflammation in asthma and chronic obstructive pulmonary disease. *Ann. N. Y. Acad. Sci.*, 992(1):218–230, May 2003.

G. Jorba, J. Aguirre-Plans, V. Junet, C. Segú-Vergés, J. L. Ruiz, A. Pujol, N. Fernández-Fuentes, J. M. Mas, and B. Oliva. In-silico simulated prototype-patients using TPMS technology to study a potential adverse effect of sacubitril and valsartan. *PLoS One*, 15 (2):e0228926, 2020. doi: 10.1371/journal.pone.0228926.

J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S. A. A. Kohl, A. J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A. W. Senior, K. Kavukcuoglu, P. Kohli, and D. Hassabis. Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873):583–589, 2021. doi: 10.1038/s41586-021-03819-2.

V. Junet and X. Daura. CNN-PepPred: an open-source tool to create convolutional NN models for the discovery of patterns in peptide sets –application to peptide-MHC class II binding prediction. *Bioinformatics*, 37(23):4567–4568, 2021. doi: 10.1093/bioinformatics/btab687.

M. H. Kaplan and M. Meyeserian. An immunological cross-reaction between group-a streptococcal cells and human heart tissue. *Lancet*, 1:706–710, 1962.

A. Karni, Y. Kohn, C. Safirman, O. Abramsky, L. Barcellos, J. R. Oksenberg, E. Kahana, D. Karussis, J. Chapman, and C. Brautbar. Evidence for the genetic role of human leukocyte antigens in low frequency DRB1*1501 multiple sclerosis patients in Israel. *Mult. Scler. J.*, 5(6):410–415, 1999. doi: 10.1177/135245859900500i607.

N. Kaushansky and A. Ben-Nun. DQB1*06:02-associated pathogenic anti-myelin autoimmunity in multiple sclerosis-like disease: Potential function of DQB1*06:02 as a disease-predisposing allele. *Front. Oncol.*, 4:280, 2014. doi: 10.3389/fonc.2014.00280.

T. Kawasaki, M. Ikegawa, and T. Kawai. Antigen presentation in the lung. *Front. Immunol.*, 13:860915, 2022. doi: 10.3389/fimmu.2022.860915.

M. I. Khalil, M. H. Sommer, J. Hay, W. T. Ruyechan, and A. M. Arvin. Varicella-zoster virus (VZV) origin of DNA replication oris influences origin-dependent DNA replication and flanking gene transcription. *Virology*, 481:179–186, July 2015.

T. Kishimoto. IL-6: from arthritis to CAR-T-cell therapy and COVID-19. *Int. Immunol.*, 33 (10):515–519, September 2021.

S. Kittel-Schneider and A. Reif. Adult attention deficit hyperactivity disorder and comorbidity: new findings on epidemiological and genetic factors. *Nervenarzt*, 91(7):575–582, July 2020.

J. S. Knight, R. Caricchio, J.L. Casanova, A. J. Combes, B. Diamond, S. E. Fox, D. A. Hanauer, J. A. James, Y. Kanthi, V. Ladd, P. Mehta, A. M. Ring, I. Sanz, C. Selmi, R. P. Tracy, P. J. Utz, C. A. Wagner, J. Y. Wang, and W. J. McCune. The intersection of COVID-19 and autoimmunity. *J. Clin. Invest.*, 131(24), December 2021.

L. Kolberg, U. Raudvere, I. Kuzmin, P. Adler, J. Vilo, and H. Peterson. g:Profiler-interoperable web service for functional enrichment analysis and gene identifier mapping (2023 update). *Nucleic Acids Res.*, 51(W1):W207–W212, 2023. doi: 10.1093/nar/gkad347.

M. Kotlyar, C. Pastrello, Z. Malik, and I. Jurisica. IID 2018 update: context-specific physical protein-protein interactions in human, model organisms and domesticated species. *Nucleic Acids Res.*, 47(D1):D581–D589, 2019. doi: 10.1093/nar/gky1037.

C. Krishna, J. Chiou, S. Sakaue, J. B. Kang, S. M. Christensen, I. Lee, M. A. Aksit, H. I. Kim, D. von Schack, S. Raychaudhuri, D. Ziemek, and X. Hu. The influence of HLA genetic variation on plasma protein expression. *Nat. Commun.*, 15(1):6469, July 2024.

M. Kuhn, I. Letunic, L. J. Jensen, and P. Bork. The SIDER database of drugs and side effects. *Nucleic Acids Res.*, 44(D1):D1075–9, January 2016.

T. Kuret, B. Burja, J. Feichtinger, G. G. Thallinger, M. Frank-Bertoncelj, K. Lakota, P. Žigon, S. Sodin-Semrl, S. Čučnik, M. Tomšič, and A. Hočevar. Gene and miRNA expression in giant cell arteritis-a concise systematic review of significantly modified studies. *Clin Rheumatol*, 38(2):307–316, Feb 2019.

H. Kwon, Y. Han, D. H. Son, Y. P. Cho, and T. W. Kwon. Abdominal aortic aneurysm in giant cell arteritis. *Ann Surg Treat Res*, 89(4):224–227, Oct 2015.

P. Langfelder and S. Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9:559, 2008. doi: 10.1186/1471-2105-9-559.

C. A. Langford, D. Cuthbertson, S. R. Ytterberg, N. Khalidi, P. A. Monach, S. Carette, P. Seo, L. W. Moreland, M. Weisman, C. L. Koening, A. G. Sreih, R. Spiera, C. A. McAlear, K. J. Warrington, C. Pagnoux, K. McKinnon, L. J. Forbess, G. S. Hoffman, R. Borchin, J. P. Krischer, P. A. Merkel, R. Hajj-Ali, K. Tuthill, K. Gartner, L. Madden, E. L. Matteson, T. Kermani, J. Jaquith, N. Amudala, M. Clark-Cotton, S. Messier, J. Farquharson, S. Jagadeesh, D. McBride, S. Venuturupalli, D. Wallace, R. Phan, N. Verde, D. Salinas, J. Godina, M. Davids, U. Udeh, L. Sejismundo, and J. Harris. A Randomized, Double-Blind Trial of Abatacept (CTLA-4Ig) for the Treatment of Giant Cell Arteritis. *Arthritis Rheumatol*, 69(4):837–845, Apr 2017.

W. Li and A. Godzik. Cd-hit: A fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659, 2006. doi: 10.1093/bioinformatics/btl158.

B. Liang, Q. Yi, and Y. Feng. Association of gastroesophageal reflux disease with asthma control. *Dis. Esophagus*, 26(8):794–798, November 2013.

J. T. Lie. Temporal artery biopsy diagnosis of giant cell arteritis: lessons from 1109 biopsies. *Anat Pathol*, 1:69–97, 1996.

E. Liozon, B. Ouattara, K. Rhaiem, K. Ly, H. Bezanahary, V. Loustaud, P. Letellier, M. Drouet, and E. Vidal. Familial aggregation in giant cell arteritis and polymyalgia rheumatica: a comprehensive literature review including 4 new families. *Clin Exp Rheumatol*, 27(1 Suppl 52):89–94, 2009.

X. Liu, Q. Li, K. Dowdell, E. R. Fischer, and J. I. Cohen. Varicella-Zoster virus ORF12 protein triggers phosphorylation of ERK1/2 and inhibits apoptosis. *J. Virol.*, 86(6):3143–3151, March 2012.

K. H. Ly, A. Régent, E. Molina, S. Saada, P. Sindou, C. Le-Jeunne, A. Brézin, V. Witko-Sarsat, F. Labrousse, P. Y. Robert, P. Bertin, J. L. Bourges, A. L. Fauchais, E. Vidal, L. Mouthon, and M. O. Jauberteau. Neurotrophins are expressed in giant cell arteritis lesions and may contribute to vascular remodeling. *Arthritis Res Ther*, 16(6):487, Nov 2014.

S. L. Mackie, A. W. Morgan, and P. F. Jones. *Angiogenesis and Giant Cell Arteritis*, page 383–402. Springer Netherlands, 2010. ISBN 9789048194957. doi: 10.1007/978-90-481-9495-7_17. URL http://dx.doi.org/10.1007/978-90-481-9495-7_17.

C. Maguire, C. Wang, A. Ramasamy, C. Fonken, B. Morse, N. Lopez, D. Wylie, and E. Melamed. Molecular mimicry as a mechanism of viral immune evasion and autoimmunity. *Nat. Commun.*, 15(1):9403, October 2024.

J. J. Maleszewski, B. R. Younge, J. T. Fritzlen, G. G. Hunder, J. J. Goronzy, K. J. Warrington, and C. M. Weyand. Clinical and pathological evolution of giant cell arteritis: a prospective study of follow-up temporal artery biopsies in 40 treated patients. *Mod Pathol*, 30(6):788–796, Jun 2017.

G. Marini, S. Murray, A. Goldhirsch, R. D. Gelber, M. Castiglione-Gertsch, K. N. Price, M. H. Tattersall, C. M. Rudenstam, J. Collins, J. Lindtner, F. Cavalli, H. Cortés-Funes, A. Gudgeon, J. F. Forbes, E. Galligioni, A. S. Coates, and H. J. Senn. The effect of adjuvant prednisone combined with CMF on patterns of relapse and occurrence of second malignancies in patients with breast cancer. international (ludwig) breast cancer study group. *Ann. Oncol.*, 7(3):245–250, March 1996.

B.W. Matthews. Comparison of the predicted and observed secondary structure of t4 phage lysozyme. *Biochimica et Biophysica Acta (BBA) - Protein Structure*, 405(2): 442–451, October 1975. ISSN 0005-2795. doi: 10.1016/0005-2795(75)90109-9. URL http://dx.doi.org/10.1016/0005-2795(75)90109-9.

V. Matzaraki, V. Kumar, C. Wijmenga, and A. Zhernakova. The MHC locus and genetic susceptibility to autoimmune and infectious diseases. *Genome Biol.*, 18(1):76, April 2017.

M. Maz, S. A. Chung, A. Abril, C. A. Langford, M. Gorelik, G. Guyatt, A. M. Archer, D. L. Conn, K. A. Full, P. C. Grayson, M. F. Ibarra, L. F. Imundo, S. Kim, P. A. Merkel, R. L. Rhee, P. Seo, J. H. Stone, S. Sule, R. P. Sundel, O. I. Vitobaldi, A. Warner, K. Byram, A. B. Dua, N. Husainat, K. E. James, M. A. Kalot, Y. C. Lin, J. M. Springer, M. Turgunbaev, A. Villa-Forte, A. S. Turner, and R. A. Mustafa. 2021 american college of rheumatology/vasculitis foundation guideline for the management of giant cell arteritis and takayasu arteritis. *Arthritis Rheumatol.*, 73(8):1349–1365, August 2021.

K. Melissaropoulos, K. Klavdianou, A. Filippopoulou, F. Kalofonou, H. Kalofonos, and D. Daoussis. Rheumatic Manifestations in Patients Treated with Immune Checkpoint Inhibitors. *Int J Mol Sci*, 21(9), May 2020.

E. B. Mitchell and D. M. Cestari. Giant cell arteritis and angiogenesis: a review. *Semin Ophthalmol*, 24(3):190–193, 2009.

R. Mitra. Adverse effects of corticosteroids on bone metabolism: a review. *PM R*, 3(5): 466–71; quiz 471, May 2011.

H. Miyadera and K. Tokunaga. Associations of human leukocyte antigens with autoimmune diseases: challenges in identifying the mechanism. *J. Hum. Genet.*, 60(11):697–702, November 2015.

National Center for Health Statistics (NCHS). National Health and Nutrition Examination Survey (NHANES). `https://www.cdc.gov/nchs/nhanes/index.htm`, 2024. Accessed: 2022.

S. B. Needleman and C. D. Wunsch. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.*, 48(3):443–453, March 1970.

D. Nepal, M. Putman, and S. Unizony. Giant Cell Arteritis and Polymyalgia Rheumatica: Treatment Approaches and New Targets. *Rheum Dis Clin North Am*, 49(3):505–521, Aug 2023.

T. Ness and B. Nölle. Giant Cell Arteritis. *Klin Monbl Augenheilkd*, 241(5):644–652, May 2024.

J. A. Newman, A. Douangamath, S. Yadzani, Y. Yosaatmadja, A. Aimon, J. Brandão Neto, L. Dunnett, T. Gorrie-Stone, R. Skyner, D. Fearon, M. Schapira, F. von Delft, and O. Gileadi. Structure, mechanism and crystallographic fragment screening of the SARS-CoV-2 NSP13 helicase. *Nat. Commun.*, 12(1):4848, 2021. doi: 10.1038/s41467-021-25166-6.

F. Nielsen. *Hierarchical Clustering*, page 195–211. Springer International Publishing, 2016. ISBN 9783319219035. doi: 10.1007/978-3-319-21903-5_8. URL `http://dx.doi.org/10.1007/978-3-319-21903-5_8`.

J. B. Nilsson, S. Kaabinejadian, H. Yari, M. G. D. Kester, P. van Balen, W. H. Hildebrand, and M. Nielsen. Accurate prediction of HLA class II antigen presentation across all loci using tailored data acquisition and refined machine learning. *Sci. Adv.*, 9(47):eadj6367, November 2023.

A. Nobile. The discovery of the delta 1,4-steroids, prednisone, and prednisolone at the schering corporation (USA). *Steroids*, 59(3):227–230, March 1994.

F. J. Obermair, F. Renoux, S. Heer, C. H. Lee, N. Cereghetti, M. Loi, G. Maestri, Y. Haldner, R. Wuigk, O. Iosefson, P. Patel, K. Triebel, M. Kopf, J. Swain, and J. Kisielow. High-resolution profiling of MHC II peptide presentation capacity reveals SARS-CoV-2 CD4 T cell targets and mechanisms of immune escape. *Sci. Adv.*, 8(17):eabl5394, 2022. doi: 10.1126/sciadv.abl5394.

T. J. O'Donnell, A. Rubinsteyn, and U. Laserson. MHCflurry 2.0: Improved pan-allele prediction of MHC class I-presented peptides by incorporating antigen processing. *Cell Syst.*, 11(1):42–48, 2020. doi: 10.1016/j.cels.2020.06.010.

M. B. Oldstone. Molecular mimicry and immune-mediated diseases. *FASEB J.*, 12(13): 1255–1265, 1998. doi: 10.1096/fasebj.12.13.1255.

S. L. Oliver, M. Zhou, and A. M. Arvin. Varicella-zoster virus: molecular controls of cell fusion-dependent pathogenesis. *Biochem. Soc. Trans.*, 48(6):2415–2435, December 2020.

J. K. Olson, G. A. Bishop, and C. Grose. Varicella-zoster virus fc receptor ge glycoprotein: serine/threonine and tyrosine phosphorylation of monomeric and dimeric forms. *J. Virol.*, 71(1):110–119, January 1997.

L. O'Neill, P. Rooney, D. Molloy, M. Connolly, J. McCormick, G. McCarthy, D. J. Veale, C. C. Murphy, U. Fearon, and E. Molloy. Regulation of Inflammation and Angiogenesis in Giant Cell Arteritis by Acute-Phase Serum Amyloid A. *Arthritis Rheumatol*, 67(9): 2447–2456, Sep 2015.

P. Oner, O. Oner, F. M. Azik, E. Cop, and K. M. Munir. Ferritin and hyperactivity ratings in attention deficit hyperactivity disorder. *Pediatr. Int.*, 54(5):688–692, October 2012.

DrugBank Online. Drugbank: Open data drug & drug target database, 2024. URL `https://go.drugbank.com`.

D. Opriș-Belinski, C. O. Cobilinschi, and I. Săulescu. Current Perspectives in Giant Cell Arteritis: Can We Better Connect Pathogenesis and Treatment? *Medicina (Kaunas)*, 60 (3), Feb 2024.

M. Oray, K. Abu Samra, N. Ebrahimiadib, H. Meese, and C. S. Foster. Long-term side effects of glucocorticoids. *Expert Opin. Drug Saf.*, 15(4):457–465, February 2016.

M. Panagioti, C. Scott, A. Blakemore, and P. A. Coventry. Overview of the prevalence, im-

pact, and management of depression and anxiety in chronic obstructive pulmonary disease. *Int. J. Chron. Obstruct. Pulmon. Dis.*, 9:1289–1306, November 2014.

F. Parker, L. A. Healey, K. R. Wilske, and G. F. Odland. Light and electron microscopic studies on human temporal arteries with special reference to alterations related to senescence, atherosclerosis and giant cell arteritis. *Am J Pathol*, 79(1):57–80, Apr 1975.

R. Pavlos, E. J. McKinnon, D.A. Ostrov, B. Peters, S. Buus, D. Koelle, A. Chopra, R. Schutte, C. Rive, A. Redwood, S. Restrepo, A. Bracey, T. Kaever, P. Myers, E. Speers, S.A. Malaker, J. Shabanowitz, Y. Jing, S. Gaudieri, Hunt D.F., M. Carrington, D.W. Haas, S. Mallal, and Phillips E. Shared peptide binding of HLA class I and II alleles associate with cutaneous nevirapine hypersensitivity and identify novel risk alleles. *Sci. Rep.*, 7(1):8653, August 2017.

F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res.*, 12(85):2825–2830, 2011.

L. K. Peterson and R. S. Fujinami. Molecular mimicry. In *Autoantibodies*, pages 13–19. Elsevier, 2007.

C. Picado and M. Luengo. Corticosteroid-induced bone loss. prevention and management. *Drug Saf.*, 15(5):347–359, November 1996.

Federica Pierini and Tobias L Lenz. Divergent allele advantage at human MHC genes: Signatures of past and ongoing selection. *Mol. Biol. Evol.*, 35(9):2145–2158, September 2018.

E. Planas Rigol and M. Corbera Bellalta. Giant-cell arteritis: Immunopathogenic mechanisms involved in vascular inflammation and remodeling. *Journal of Vasculitis*, 2(1), 2016.

E. Planas-Rigol, N. Terrades-Garcia, M. Corbera-Bellalta, E. Lozano, M. A. Alba, M. Segarra, G. Espígol-Frigolé, S. Prieto-Gonzálezlez, J. Hernández-Rodríguez, S. Preciado, R. Lavilla, and M. C. Cid. Endothelin-1 promotes vascular smooth muscle cell migration across the artery wall: a mechanism contributing to vascular remodelling and intimal hyperplasia in giant-cell arteritis. *Ann Rheum Dis*, 76(9):1624–1634, Sep 2017.

E. Prat, U. Tomaru, L. Sabater, D. M. Park, R. Granger, N. Kruse, J. M. Ohayon, M. P. Bettinotti, and R. Martin. HLA-DRB5*0101 and -DRB1*1501 expression in the multiple sclerosis-associated HLA-DR15 haplotype. *J. Neuroimmunol.*, 167(1-2):108–119, 2005. doi: 10.1016/j.jneuroim.2005.04.027.

Anjum F Preuss CV. *Tocilizumab*. StatPearls, Treasure Island (FL), 2024. URL `https://www.ncbi.nlm.nih.gov/books/NBK570644/`.

S. Prieto-González, N. Terrades-García, M. Corbera-Bellalta, E. Planas-Rigol, C. Miyabe, M. A. Alba, A. Ponce, I. Tavera-Bahillo, G. Murgia, G. Espígol-Frigolé, J. Marco-Hernández, J. Hernández-Rodríguez, A. García-Martínez, S. H. Unizony, and M. C. Cid. Serum osteopontin: a biomarker of disease activity and predictor of relapsing course in patients with giant cell arteritis. Potential clinical usefulness in tocilizumab-treated patients. *RMD Open*, 3(2):e000570, 2017.

D. Prieto-Peña, S. Remuzgo-Martínez, J. G. Ocejo-Vinyals, B. Atienza-Mateo, F. Genre, A. Muñoz-Jimenez, F. Ortiz-Sanjuán, S. Romero-Yuste, C. Moriano, E. Galindez-Agirregoikoa, I. Calvo, N. Ortego-Centeno, N. Álvarez-Rivas, J. A. Miranda-Filloy, I. Llorente, J. García-García, R. Blanco, O. Gualillo, J. Martin, S. Castañeda, R. Lopez-Mejías, and M. A. González-Gay. The presence of both HLA-DRB1*04:01 and HLA-B*15:01 increases the susceptibility to cranial and extracranial giant cell arteritis. *Clin. Exp. Rheumatol.*, 39 Suppl 129(2):21–26, March 2021.

Y. Puckett, A. Gabbar, and Bokhari A.A. *Prednisone*. StatPearls, Treasure Island (FL), 2023. URL `https://www.ncbi.nlm.nih.gov/books/NBK534809/`.

J. A. Quandt, J. Huh, M. Baig, K. Yao, N. Ito, M. Bryant, K. Kawamura, C. Pinilla, H. F. McFarland, R. Martin, and K. Ito. Myelin basic protein-specific TCR/HLA-DRB5*01:01 transgenic mice support the etiologic role of DRB5*01:01 in multiple sclerosis. *J. Immunol.*, 189(6):2897–2908, 2012. doi: 10.4049/jimmunol.1103087.

J. Ramana and K. Mehla. Immunoinformatics and epitope prediction. *Methods Mol. Biol.*, 2131:155–171, 2020.

A. Rambaut, E. C. Holmes, A. O'Toole, V. Hill, J. T. McCrone, C. Ruis, L. du Plessis, and O. G. Pybus. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.*, 5(11):1403–1407, 2020. doi: 10.1038/s41564-020-0770-5.

M. J. Regan, B. J. Wood, Y. H. Hsieh, M. L. Theodore, T. C. Quinn, D. B. Hellmann, W. R. Green, C. A. Gaydos, and J. H. Stone. Temporal arteritis and Chlamydia pneumoniae: failure to detect the organism by polymerase chain reaction in ninety cases and ninety controls. *Arthritis Rheum*, 46(4):1056–1060, Apr 2002.

F. Regola, E. Cerudelli, G. Bosio, L. Andreoli, A. Tincani, F. Franceschini, and P. Toniati. Long-term treatment with tocilizumab in giant cell arteritis: efficacy and safety in a monocentric cohort of patients. *Rheumatol. Adv. Pract.*, 4(2):rkaa017, May 2020.

D. Rey and M. Neuhäuser. *Wilcoxon-Signed-Rank Test*, pages 1658–1659. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011. ISBN 978-3-642-04898-2. doi: 10.1007/978-3-642-04898-2_616. URL https://doi.org/10.1007/978-3-642-04898-2_616.

B. Reynisson, B. Alvarez, S. Paul, B. Peters, and M. Nielsen. NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic Acids Res.*, 48(W1): W449–W454, 2020. doi: 10.1093/nar/gkaa379.

R. L. Rhee, P. C. Grayson, P. A. Merkel, and G. Tomasson. Infections and the risk of incident giant cell arteritis: a population-based, case-control study. *Ann. Rheum. Dis.*, 76 (6):1031–1035, June 2017.

H. Robberecht, A. A. J. Verlaet, A. Breynaert, T. De Bruyne, and N. Hermans. Magnesium, iron, zinc, copper and selenium status in attention-deficit/hyperactivity disorder (ADHD). *Molecules*, 25(19):4440, September 2020.

M. Rojas, P. Restrepo-Jiménez, D. M. Monsalve, Y. Pacheco, Y. Acosta-Ampudia, C. Ramírez-Santana, P. S. C. Leung, A. A. Ansari, M. E. Gershwin, and J. M. Anaya. Molecular mimicry and autoimmunity. *J. Autoimmun.*, 95:100–123, 2018. doi: 10.1016/j.jaut.2018.10.012.

Peter J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, November 1987. ISSN 0377-0427. doi: 10.1016/0377-0427(87)90125-7. URL http://dx.doi.org/10.1016/0377-0427(87)90125-7.

A. Régent and L. Mouthon. Treatment of Giant Cell Arteritis (GCA). *J Clin Med*, 11(7), Mar 2022.

A. Régent, K. H. Ly, M. Groh, C. Khifer, S. Lofek, G. Clary, P. Chafey, V. Baud, C. Broussard, C. Federici, F. Labrousse, L. Mesturoux, C. Le Jeunne, E. Vidal, A. Brezin, V. Witko-Sarsat, L. Guillevin, and L. Mouthon. Molecular analysis of vascular smooth muscle cells from patients with giant cell arteritis: Targeting endothelin-1 receptor to control proliferation. *Autoimmun Rev*, 16(4):398–406, Apr 2017.

S. Sadegh-Nasseri. A step-by-step overview of the dynamic process of epitope selection by major histocompatibility complex class II for presentation to helper T cells. *F1000Res.*, 5: 1305, June 2016.

M. J. Salameh, J. H. Black, and E. V. Ratchford. Thoracic aortic aneurysm. *Vasc Med*, 23 (6):573–578, Dec 2018.

M. Salgado-Albarrán, E. I. Navarro-Delgado, A. Del Moral-Morales, N. Alcaraz, J. Baumbach, R. González-Barrios, and E. Soto-Reyes. Comparative transcriptome analysis reveals key epigenetic targets in SARS-CoV-2 infection. *npj Syst. Biol. Appl.*, 7:21, 2021. doi: 10.1038/s41540-021-00181-x.

M. Samson, M. Corbera-Bellalta, S. Audia, E. Planas-Rigol, L. Martin, M. C. Cid, and B. Bonnotte. Recent advances in our understanding of giant cell arteritis pathogenesis. *Autoimmun Rev*, 16(8):833–844, Aug 2017.

A. A. Schäffer, L. Aravind, T. L. Madden, S. Shavirin, J. L. Spouge, Y. I. Wolf, E. V. Koonin, and S. F. Altschul. Improving the accuracy of PSI-BLAST protein database searches with composition-based statistics and other refinements. *Nucleic Acids Res.*, 29(14):2994–3005, 2001. doi: 10.1093/nar/29.14.2994.

G. Schett, B. Manger, D. Simon, and R. Caporali. COVID-19 revisiting inflammatory pathways of arthritis. *Nat. Rev. Rheumatol.*, 16(8):465–470, August 2020.

T. A. Schweiger and M. Zdanowicz. Systemic corticosteroids in the treatment of acute exacerbations of chronic obstructive pulmonary disease. *Am. J. Health. Syst. Pharm.*, 67(13): 1061–1069, July 2010.

A. Sebba. Tocilizumab: the first interleukin-6-receptor inhibitor. *Am. J. Health. Syst. Pharm.*, 65(15):1413–1418, August 2008.

C. Segú-Vergés, S. Caño, E. Calderón-Gómez, H. Bartra, T. Sardón, S. Kaveri, and J. Terencio. Systems biology and artificial intelligence analysis highlights the pleiotropic effect of IVIg therapy in autoimmune diseases with a predominant role on B cells and complement system. *Front. Immunol.*, 13:901872, 2022. doi: 10.3389/fimmu.2022.901872.

M. N. Seyyed Mousavi, B. Mehramuz, J. Sadeghi, N. Alizadeh, M. A. Oskouee, and H. S. Kafil. The pathogenesis of staphylococcus aureus in autoimmune diseases. *Microb. Pathog.*, 111:503–507, October 2017.

M. Shahbazi, D. Roshandel, H. Ebadi, D. Fathi, M. Zamani, M. Boghaee, M. Mohammadhoseeeni, A. Rshaidbaghan, A. Bakhshandeh, and S. Shahbazi. High frequency of the IL-2 - 330T/HLA-DRB1*1501 haplotype in patients with multiple sclerosis. *Clin. Immunol.*, 137(1):134–138, 2010. doi: 10.1016/j.clim.2010.05.010.

P. S. Shenkin, B. Erman, and L. D. Mastrandrea. Information-theoretical entropy as a measure of sequence variability. *Proteins*, 11(4):297–313, 1991. doi: 10.1002/prot.340110408.

F. Sievers and D. G. Higgins. The clustal omega multiple alignment package. *Methods Mol.*

Biol., 2231:3–16, 2021.

Anaxomics Biotech SL. Tpms technology [internet], 2018. URL `http://www.anaxomics.com/tpms.php`.

M. K. Smatti, F. S. Cyprian, G. K. Nasrallah, A. A. Al Thani, R. O. Almishal, and H. M. Yassine. Viruses and autoimmunity: A review on the potential interaction and molecular mechanisms. *Viruses*, 11(8):762, August 2019.

T. F. Smith and M. S. Waterman. Identification of common molecular subsequences. *J. Mol. Biol.*, 147(1):195–197, March 1981.

R. C. Stacy, J. F. Rizzo, and D. M. Cestari. Subtleties in the histopathology of giant cell arteritis. *Semin Ophthalmol*, 26(4-5):342–348, Jul-Sep 2011.

P. Stamatis. Giant Cell Arteritis versus Takayasu Arteritis: An Update. *Mediterr J Rheumatol*, 31(2):174–182, Jun 2020.

A. Starshinova, A. Malkova, I. Kudryavtsev, D. Kudlay, Y. Zinchenko, and P. Yablonskiy. Tuberculosis and autoimmunity: Common features. *Tuberculosis (Edinb.)*, 134(102202): 102202, May 2022.

K. H. Stürner, I. Siembab, G. Schön, J. P. Stellmann, N. Heidari, B. Fehse, C. Heesen, T. H. Eiermann, R. Martin, and T. M. Binder. Is multiple sclerosis progression associated with the HLA-DR15 haplotype? *Mult. Scler. J. Exp. Transl. Clin.*, 5(4), 2019. doi: 10.1177/2055217319894615.

M. Szumilas. Explaining odds ratios. *J. Can. Acad. Child Adolesc. Psychiatry*, 19(3):227–229, August 2010.

C. C. Tam, S. J. O'Brien, I. Petersen, A. Islam, A. Hayward, and L. C. Rodrigues. Guillain-barré syndrome and preceding infection with campylobacter, influenza and epstein-barr virus in the general practice research database. *PloS one*, 2(4):e344, 2007.

A. Tareen and J. B. Kinney. Logomaker: Beautiful sequence logos in Python. *Bioinformatics*, 36(7):2272–2274, 2020. doi: 10.1093/bioinformatics/btz921.

J. S. Temme, D. L. Butler, and J. C. Gildersleeve. Anti-glycan antibodies: roles in human disease. *Biochem. J.*, 478(8):1485–1509, April 2021.

N. Terrades-Garcia and M. C. Cid. Pathogenesis of giant-cell arteritis: how targeted therapies are influencing our understanding of the mechanisms involved. *Rheumatology (Oxford)*, 57

(suppl_2):ii51–ii62, Feb 2018.

T. C. Terwilliger, D. Liebschner, T. I. Croll, C. J. Williams, A. J. McCoy, B. K. Poon, P. V. Afonine, R. D. Oeffner, J. S. Richardson, R. J. Read, and P. D. Adams. AlphaFold predictions are valuable hypotheses and accelerate but do not replace experimental structure determination. *Nat. Methods*, 21(1):110–116, 2024. doi: 10.1038/s41592-023-02087-4.

The UniProt Consortium. UniProt: the Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.*, 51(1):523–531, 2023. doi: 10.1093/nar/gkac1052.

W. H. Thompson, C. P. Nielson, P. Carvalho, N. B. Charan, and J. J. Crowley. Controlled trial of oral prednisone in outpatients with acute COPD exacerbation. *Am. J. Respir. Crit. Care Med.*, 154(2 Pt 1):407–412, August 1996.

F. Tomyo, N. Sugimoto, H. Toyota, A. Ito, M. Ujino, H. Sakasegawa, K. Kobayashi, Y. Koizumi, S. Miyoshi, M. Kuramochi, M. Yamaguchi, T. Yamamoto, and H. Nagase. Prevalence of functional dyspepsia symptoms in patients with asthma. *Arerugi*, 68(9): 1132–1140, 2019.

Jerome Friedman Trevor Hastie, Robert Tibshirani. Model assessment and selection. In Trevor Hastie, Robert Tibshirani, and Jerome Friedman, editors, *The Elements Of Statistical Learning - Data Mining, Inference, And Prediction, Second Edition*, pages 219–259. Springer Science & Business Media, 2009. ISBN 978-0-387-84858-7.

M. Tsai, H. Lin, C. Lin, H. Lin, Y. Chen, S. Pfeiffer, and H. Lin. Increased risk of concurrent asthma among patients with gastroesophageal reflux disease: a nationwide population-based study. *Eur. J. Gastroenterol. Hepatol.*, 22(10):1169–1173, October 2010.

S. Unizony and T. A. Kermani. IL-6 blockade and its therapeutic success in giant cell arteritis. *J. Neuroophthalmol.*, 38(4):551–558, December 2018.

U.S. Food and Drug Administration. Fda approved label for drug 202020, 2013. `https://www.accessdata.fda.gov/drugsatfda_docs/label/2013/202020s003lbl.pdf`.

M. Živković, A. Stanković, E. Dinčić, M. Popović, S. Popović, R. Raičević, and D. Alavantić. The tag SNP for HLA-DRB1*1501, rs3135388, is significantly associated with multiple sclerosis susceptibility: cost-effective high-throughput detection by real-time PCR. *Clin. Chim. Acta*, 406(1-2):27–30, 2009. doi: 10.1016/j.cca.2009.05.004.

K. S. van der Geest, W. H. Abdulahad, P. Chalan, A. Rutgers, G. Horst, M. G. Huitema, M. P. Roffel, C. Roozendaal, P. M. Kluin, N. A. Bos, A. M. Boots, and E. Brouwer. Disturbed B cell homeostasis in newly diagnosed giant cell arteritis and polymyalgia rheumatica.

Arthritis Rheumatol, 66(7):1927–1938, Jul 2014.

M. Varadi, S. Anyango, M. Deshpande, S. Nair, C. Natassia, G. Yordanova, D. Yuan, O. Stroe, G. Wood, A. Laydon, A. Žídek, T. Green, K. Tunyasuvunakool, S. Petersen, J. Jumper, E. Clancy, R. Green, A. Vora, M. Lutfi, M. Figurnov, A. Cowie, N. Hobbs, P. Kohli, G. Kleywegt, E. Birney, D. Hassabis, and S. Velankar. AlphaFold Protein Structure Database: massively expanding the structural coverage of protein-sequence space with high-accuracy models. *Nucleic Acids Res.*, 50(D1):D439–D444, 2022. doi: 10.1093/nar/gkab1061.

R. Vita, S. Mahajan, J. A. Overton, S. K. Dhanda, S. Martini, J. R. Cantrell, D. K. Wheeler, A. Sette, and B. Peters. The Immune Epitope Database (IEDB): 2018 update. *Nucleic Acids Res.*, 47(D1):D339–D343, 2019. doi: 10.1093/nar/gky1006.

S. W. N. Vogel, D. Bijlenga, J. Verduijn, T. I. Bron, A. T. F. Beekman, J. J. S. Kooij, and B. W. J. H. Penninx. Attention-deficit/hyperactivity disorder symptoms and stress-related biomarkers. *Psychoneuroendocrinology*, 79:31–39, May 2017.

D. A. Walsh and D. F. McWilliams. Pain in rheumatoid arthritis. *Curr. Pain Headache Rep.*, 16(6):509–517, December 2012.

D. Wang, B. Eraslan, T. Wieland, B. m, T. Hopf, D. P. Zolg, J. Zecha, A. Asplund, L. H. Li, C. Meng, M. Frejno, T. Schmidt, K. Schnatbaum, M. Wilhelm, F. Ponten, M. Uhlen, J. Gagneur, H. Hahne, and B. Kuster. A deep proteome and transcriptome abundance atlas of 29 healthy human tissues. *Mol. Syst. Biol.*, 15(2):e8503, 2019. doi: 10.15252/msb.20188503.

J. Y. Wang, W. Zhang, V. B. Roehrl, M. W. Roehrl, and M. H. Roehrl. An autoantigen atlas from human lung HFL1 cells offers clues to neurological and diverse autoimmune manifestations of COVID-19. *Front. Immunol.*, 13:831849, 2022. doi: 10.3389/fimmu.2022.831849.

P. Wang, J. Sidney, Y. Kim, A. Sette, O. Lund, M. Nielsen, and B. Peters. Peptide binding predictions for HLA DR, DP and DQ molecules. *BMC Bioinformatics*, 11:568, 2010. doi: 10.1186/1471-2105-11-568.

X. Wang, H. Chen, F. Song, K. Zuo, X. Chen, X. Zhang, L. Liang, Q. Ta, L. Zhang, and J. Li. Resveratrol: a potential medication for the prevention and treatment of varicella zoster virus-induced ischemic stroke. *Eur. J. Med. Res.*, 28(1):400, October 2023.

R. Watanabe, H. Zhang, G. Berry, J. J. Goronzy, and C. M. Weyand. Immune checkpoint dysfunction in large and medium vessel vasculitis. *Am J Physiol Heart Circ Physiol*, 312

(5):H1052–H1059, May 2017.

C. M. Weyand and J. J. Goronzy. Pathogenic mechanisms in giant cell arteritis. *Cleve Clin J Med*, 69 Suppl 2:28–32, 2002.

C. M. Weyand and J. J. Goronzy. Immune mechanisms in medium and large-vessel vasculitis. *Nat Rev Rheumatol*, 9(12):731–740, Dec 2013.

C. M. Weyand, Y. J. Liao, and J. J. Goronzy. The immunopathology of giant cell arteritis: diagnostic and therapeutic implications. *J Neuroophthalmol*, 32(3):259–265, Sep 2012.

C. M. Weyand, G. J. Berry, and J. J. Goronzy. The immunoinhibitory PD-1/PD-L1 pathway in inflammatory blood vessel disease. *J Leukoc Biol*, 103(3):565–575, Mar 2018.

M. Wieczorek, E. T. Abualrous, J. Sticht, M. Álvaro Benito, S. Stolzenberg, F. Noé, and C. Freund. Major Histocompatibility Complex (MHC) Class I and MHC Class II Proteins: Conformational Plasticity in Antigen Presentation. *Front Immunol*, 8:292, 2017.

J. A. Woods, J. S. Wheeler, C. K. Finch, and N. A. Pinner. Corticosteroids in the treatment of acute exacerbations of chronic obstructive pulmonary disease. *Int. J. Chron. Obstruct. Pulmon. Dis.*, 9:421–430, May 2014.

F. Wu, S. Zhao, B. Yu, Y. M. Chen, W. Wang, Z. G. Song, Y. Hu, Z. W. Tao, J. H. Tian, Y. Y. Pei, M. L. Yuan, Y. L. Zhang, F. H. Dai, Y. Liu, Q. M. Wang, J. J. Zheng, L. Xu, E. C. Holmes, and Y. Z. Zhang. A new coronavirus associated with human respiratory disease in China. *Nature*, 579(7798):265–269, 2020. doi: 10.1038/s41586-020-2008-3.

L. Wu. The pathogenesis of thoracic aortic aneurysm from hereditary perspective. *Gene*, 677:77–82, Nov 2018.

H. Xia, Z. Cao, X. Xie, X. Zhang, J. Y. Chen, H. Wang, V. D. Menachery, R. Rajsbaum, and P. Y. Shi. Evasion of Type I Interferon by SARS-CoV-2. *Cell Rep.*, 33(1):108234, 2020. doi: 10.1016/j.celrep.2020.108234.

A. Yan, G. Cai, W. Xia, and Y Fu. Thromboxane A2 receptor antagonist SQ29548 suppresses the LPS-induced release of inflammatory cytokines in BV2 microglia cells via suppressing MAPK and NF-κB signaling pathways. *Mol. Med. Rep.*, 16(3):2491–2496, September 2017.

L. Yan, J. Ge, L. Zheng, Y. Zhang, Y. Gao, T. Wang, Y. Huang, Y. Yang, S. Gao, M. Li, Z. Liu, H. Wang, Y. Li, Y. Chen, L. W. Guddat, Q. Wang, Z. Rao, and Z. Lou. Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis. *Cell*, 184(1):184–193, 2021. doi: 10.1016/j.cell.2020.

11.016.

Z. Yan, K. Kim, H. Kim, B. Ha, A. Gambiez, J. Bennett, M. F. de Almeida Mendes, R. Trevizani, J. Mahita, E. Richardson, D. Marrama, N. Blazeska, Z. Koşaloğlu-Yalçın, M. Nielsen, A. Sette, B. Peters, and J. A. Greenbaum. Next-generation IEDB tools: a platform for epitope prediction and analysis. *Nucleic Acids Res.*, 52(W1):W526–W532, July 2024.

X. Yang, Z. Xiang, Z. Sun, F. Ji, K. Ren, and D. Pan. Host MOV10 is induced to restrict herpes simplex virus 1 lytic infection by promoting type I interferon response. *PLoS Pathog.*, 18(2):e1010301, 2022. doi: 10.1371/journal.ppat.1010301.

Y. Yang, Z. Wei, G. Cia, X. Song, F. Pucci, M. Rooman, F. Xue, and Q. Hou. MHCII-peptide presentation: an assessment of the state-of-the-art prediction methods. *Front. Immunol.*, 15:1293706, March 2024.

W. G. Yeung, W. D. Rawlinson, and M. E. Craig. Enterovirus infection and type 1 diabetes mellitus: systematic review and meta-analysis of observational molecular studies. *BMJ*, 342(feb03 1):d35, February 2011.

V. Yurina and O. R. Adianingsih. Predicting epitopes for vaccine development using bioinformatics tools. *Ther. Adv. Vaccines Immunother.*, 10:25151355221100218, May 2022.

A. Zerbini, F. Muratore, L. Boiardi, F. Ciccia, M. Bonacini, L. Belloni, A. Cavazza, L. Cimino, A. Moramarco, R. Alessandro, A. Rizzo, M. Parmeggiani, C. Salvarani, and S. Croci. Increased expression of interleukin-22 in patients with giant cell arteritis. *Rheumatology (Oxford)*, 57(1):64–72, Jan 2018.

L. Zerboni, N. Sen, S. L. Oliver, and A. M. Arvin. Molecular mechanisms of varicella zoster virus pathogenesis. *Nat. Rev. Microbiol.*, 12(3):197–210, March 2014.

B. Zhang, M. Bai, X. Xu, M. Yang, F. Niu, F. Gao, and B. Liu. Corticosteroid receptor rebalancing alleviates critical illness-related corticosteroid insufficiency after traumatic brain injury by promoting paraventricular nuclear cell survival via Akt/CREB/BDNF signaling. *J. Neuroinflammation*, 17(1):318, October 2020.

H. Zhang, R. Watanabe, G. J. Berry, A. Vaglio, Y. J. Liao, K. J. Warrington, J. J. Goronzy, and C. M. Weyand. Immunoinhibitory checkpoint deficiency in medium and large vessel vasculitis. *Proc Natl Acad Sci U S A*, 114(6):E970–E979, Feb 2017.

L. Zhou, K. Luneau, C. M. Weyand, V. Biousse, N. J. Newman, and H. E. Grossniklaus. Clinicopathologic correlations in giant cell arteritis: a retrospective study of 107 cases.

Ophthalmology, 116(8):1574–1580, Aug 2009.

Chapter 7

Appendix

7.1. Giant Cell Arteritis Molecular Description

Motif name	N Proteins
Vascular inflammation-early	16
Vascular inflammation-amplification cascade	32
Systemic inflammation	7
Vascular injury	5
Arterial remodelling and vascular occlusion	20
Aortic Aneurysm formation	15
Dysfunction of immune checkpoints	12
GCA genetic associations	6
total	88

Table 7.1: Summary of the motifs identified as involved in GCA, and the number of protein composing these.

Motif	Protein	State	Reference
GCA genetic associations	HLA-DRA		[Al-Mousawi et al., 2019]
GCA genetic associations	HLA-B		[Al-Mousawi et al., 2019]
GCA genetic associations	HLA-DQA1		[Al-Mousawi et al., 2019]
GCA genetic associations	HLA-DQB1		[Al-Mousawi et al., 2019]
GCA genetic associations	HLA-DRB1		[Al-Mousawi et al., 2019, Terrades-Garcia and Cid, 2018, Stamatis, 2020]
GCA genetic associations	IL12B	U	[Al-Mousawi et al., 2019]
Dysfunction of immune checkpoints	CD274	D	[Zhang et al., 2017, Weyand et al., 2018]
Dysfunction of immune checkpoints	IFNG	U	[Weyand et al., 2018, Watanabe et al., 2017]
Dysfunction of immune checkpoints	IL1B	U	[Watanabe et al., 2017]
Dysfunction of immune checkpoints	IL6	U	[Watanabe et al., 2017]
Dysfunction of immune checkpoints	IL9	U	[Watanabe et al., 2017]
Dysfunction of immune checkpoints	PDCD1	U	[Zhang et al., 2017]
Dysfunction of immune checkpoints	TGFB1	U	[Watanabe et al., 2017]
Dysfunction of immune checkpoints	FGF1	U	[Watanabe et al., 2017]
Dysfunction of immune checkpoints	IL17A	U	[Weyand et al., 2018]
Dysfunction of immune checkpoints	IL21	U	[Weyand et al., 2018]
Dysfunction of immune checkpoints	VEGFA	U	[Watanabe et al., 2017]
Dysfunction of immune checkpoints	PDGFA	U	[Watanabe et al., 2017]
Vascular inflammation – early	IL12	U	[Dammacco et al., 2020]
Vascular inflammation – early	IL18	U	[Dammacco et al., 2020]

Vascular inflammation – early	IL23	U	[Dammacco et al., 2020]
Vascular inflammation – early	CCL19	U	[Dammacco et al., 2020]
Vascular inflammation – early	CCL21	U	[Dammacco et al., 2020]
Vascular inflammation – early	ICAM1	U	[Samson et al., 2017]
Vascular inflammation – early	IFNG	U	[Cid, 2014]
Vascular inflammation – early	IL17A	U	[Planas Rigol and Corbera Bellalta, 2016]
Vascular inflammation – early	IL1B	U	[Dammacco et al., 2020]
Vascular inflammation – early	IL21	U	[Planas Rigol and Corbera Bellalta, 2016]
Vascular inflammation – early	IL6	U	[Planas Rigol and Corbera Bellalta, 2016]
Vascular inflammation – early	TGFA	U	[Planas Rigol and Corbera Bellalta, 2016]
Vascular inflammation – early	TLR2	U	[Samson et al., 2017]
Vascular inflammation – early	TLR3	U	[Weyand and Goronzy, 2013]
Vascular inflammation – early	TLR4	U	[Samson et al., 2017]
Vascular inflammation – early	VCAM1	U	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	CXCL9	U	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	CCL2	U	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	TNFA	U	[Dammacco et al., 2020]
Vascular inflammation – amplification cascades, persistent	CXCL10	U	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	CXCL11	U	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	CSF	U	[Terrades-Garcia and Cid, 2018]
Vascular inflammation – amplification cascades, persistent	FGF2	U	[Terrades-Garcia and Cid, 2018]

Vascular inflammation – amplification cascades, persistent	GMCSF	U	[Harrington et al., 2021]
Vascular inflammation – amplification cascades, persistent	GRA	U	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	IL17	U	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	ICAM1	U	[O'Neill et al., 2015]
Vascular inflammation – amplification cascades, persistent	IFNG	U	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	JAK1	U	[Harrington et al., 2021]
Vascular inflammation – amplification cascades, persistent	JAK2	U	[Harrington et al., 2021]
Vascular inflammation – amplification cascades, persistent	JAK3	U	[Harrington et al., 2021]
Vascular inflammation – amplification cascades, persistent	IL1B	U	[Terrades-Garcia and Cid, 2018]
Vascular inflammation – amplification cascades, persistent	IL22	U	[Zerbini et al., 2018]
Vascular inflammation – amplification cascades, persistent	NOX2	D	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	IL33	U	[Terrades-Garcia and Cid, 2018]
Vascular inflammation – amplification cascades, persistent	IL6	U	[Terrades-Garcia and Cid, 2018]
Vascular inflammation – amplification cascades, persistent	PDGFA	U	[Terrades-Garcia and Cid, 2018]
Vascular inflammation – amplification cascades, persistent	JAG1	U	[Deshayes et al., 2020]
Vascular inflammation – amplification cascades, persistent	PRF1	U	[Samson et al., 2017]
Vascular inflammation – amplification cascades, persistent	ASAA	U	[O'Neill et al., 2015]
Vascular inflammation – amplification cascades, persistent	GP130	U	(http://hdl.handle.net/10803/401863)

Vascular inflammation – amplification cascades, persistent	HPT	U	[Cid et al., 1993]
Vascular inflammation – amplification cascades, persistent	IL23P19	U	[Deshayes et al., 2020], http://hdl.handle.net/10803/401863
Vascular inflammation – amplification cascades, persistent	NOTCH	U	[Deshayes et al., 2020]
Vascular inflammation – amplification cascades, persistent	PECAM1	U	[O'Neill et al., 2015]
Vascular inflammation – amplification cascades, persistent	STAT3	U	http://hdl.handle.net/10803/401863
Vascular inflammation – amplification cascades, persistent	VCAM1	U	[O'Neill et al., 2015]
Vascular inflammation – amplification cascades, persistent	VEGFA	U	[Terrades-Garcia and Cid, 2018]
Systemic inflammation	CRP	U	[Terrades-Garcia and Cid, 2018]
Systemic inflammation	TNFA	U	[Cid, 2014]
Systemic inflammation	IL1B	U	[Planas Rigol and Corbera Bellalta, 2016]
Systemic inflammation	IL33	U	[Planas Rigol and Corbera Bellalta, 2016]
Systemic inflammation	IL6	U	[Cid, 2014]
Systemic inflammation	ASAA	U	[O'Neill et al., 2015]
Systemic inflammation	OPN	U	[Prieto-González et al., 2017]
Vascular injury	TIMP1	D	[Terrades-Garcia and Cid, 2018]
Vascular injury	TIMP2	D	[Terrades-Garcia and Cid, 2018]
Vascular injury	iNOS	U	[Samson et al., 2017]
Vascular injury	MMP2	U	[Cid, 2014]
Vascular injury	MMP9	U	[Cid, 2014]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	EDN1	U	[Samson et al., 2017, Deshayes et al., 2020, Cid, 2014]

Arterial remodeling and vascular occlusion / Vascular injury and remodelling	CHI3L1	U	[Burja et al., 2018]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	FGF2	U	[Dammacco et al., 2020]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	FAK	U	[Planas-Rigol et al., 2017]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	ICAM1	U	[Mackie et al., 2010]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	PDGFA	U	[Dammacco et al., 2020, Cid, 2014]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	PECAM1	U	[Mackie et al., 2010]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	SELPLG	U	[Mackie et al., 2010]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	TFGB1	U	[Cid, 2014]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	VWF	U	[Burja et al., 2018]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	BDNF	U	[Dejaco et al., 2017, Ly et al., 2014]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	NGF	U	[Dejaco et al., 2017]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	PTX3	U	[Burja et al., 2018]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	SELE	U	[Mackie et al., 2010]

Arterial remodeling and vascular occlusion / Vascular injury and remodelling	PAXI	U	[Régent et al., 2017]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	SORT	U	[Dejaco et al., 2017]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	TNC	U	[Régent et al., 2017]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	VCAM1	U	[Mackie et al., 2010]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	VEGFA	U	[Dammacco et al., 2020]
Arterial remodeling and vascular occlusion / Vascular injury and remodelling	VEGFA	U	[Samson et al., 2017]
Aortic Aneuryms formation	ERK1	U	[Barbour et al., 2007]
Aortic Aneuryms formation	ERK2	U	[Barbour et al., 2007]
Aortic Aneuryms formation	FOS	U	[Barbour et al., 2007]
Aortic Aneuryms formation	JNK1	U	[Barbour et al., 2007]
Aortic Aneuryms formation	JNK2	U	[Barbour et al., 2007]
Aortic Aneuryms formation	JUN	U	[Barbour et al., 2007]
Aortic Aneuryms formation	TIMP1	D	[Barbour et al., 2007]
Aortic Aneuryms formation	MMP12	U	[Barbour et al., 2007]
Aortic Aneuryms formation	MMP2	U	[Barbour et al., 2007]
Aortic Aneuryms formation	MMP2	U	[Barbour et al., 2007]
Aortic Aneuryms formation	ADAM17	U	[Barbour et al., 2007]
Aortic Aneuryms formation	CST3	U	[Barbour et al., 2007]
Aortic Aneuryms formation	MMP1	U	[Barbour et al., 2007]
Aortic Aneuryms formation	MMP13	U	[Barbour et al., 2007]
Aortic Aneuryms formation	MMP8	U	[Barbour et al., 2007]

Table 7.2: The table reports what is obtained in the molecular description of GCA. Each protein is described by the pathopysiological motif "Motif", the gene name of the protein "Protein", the state of the protein "U" or "D" respectively up or down regulated, and the "Reference" where the information has been found.

7.2. The Side-Effects of Prednisone

1ST CANCER - WHAT KIND WAS IT?, 2ND CANCER - WHAT KIND WAS IT?; 3RD CANCER - WHAT KIND WAS IT?; AGE AT SCREENING ADJUDICATED - RECODE; AGE IN YEARS AT SCREENING; BLURRED VISION; BODY MASS INDEX (KG/M**2); CALCIUM, TOTAL (MMOL/L); CHOLESTEROL (MG/DL); CHOLESTEROL (MMOL/L); CHOLESTEROL, REFRIGERATED SERUM (MG/DL); CHOLESTEROL, REFRIGERATED SERUM (MMOL/L); CHOLESTEROL, TOTAL (MG/DL); CHOLESTEROL, TOTAL (MMOL/L); DIABETES AFFECTED EYES/HAD RETINOPATHY; DID SP HAVE EPISODE OF HAY FEVER/PAST YR; DO YOU STILL HAVE A THYROID PROBLEM; DO YOU STILL HAVE THYROID PROBLEM; DOCTOR EVER TOLD YOU THAT YOU HAD GOUT?; DOCTOR TOLD HAVE HAY FEVER; DOCTOR TOLD YOU - HIGH CHOLESTEROL LEVEL; DOCTOR TOLD YOU HAVE DIABETES; EPISODE OF HAY FEVER IN PAST 12 MONTHS; EPISODE OF HAY FEVER IN PAST 12 MONTHS?; EVER BEEN TOLD YOU HAVE JAUNDICE?; EVER HAD 3 OR MORE EAR INFECTIONS?; EVER TOLD BY DOCTOR HAVE SLEEP DISORDER?; EVER TOLD YOU HAD A STROKE; EVER TOLD YOU HAD ANGINA/ANGINA PECTORIS; EVER TOLD YOU HAD COPD?; EVER TOLD YOU HAD HEART ATTACK; GENDER; GENERAL CONDITION OF HEARING; GENERIC DRUG NAME; GLUCOSE (MMOL/L); GLUCOSE, REFRIGERATED SERUM (MMOL/L); GLUCOSE, SERUM (MMOL/L); HAD A HYSTERECTOMY?; ICD-10-CM CODE 1 DESCRIPTION; ICD-10-CM CODE 2 DESCRIPTION; ICD-10-CM CODE 3 DESCRIPTION; LDL-CHOLESTEROL (MG/DL); LDL-CHOLESTEROL (MMOL/L); MORE THAN 3 KINDS OF CANCER; NUMBER OF DAYS TAKEN MEDICINE; PHOSPHORUS (MMOL/L); POTASSIUM (MMOL/L); POTASSIUM: SI (MMOL/L); PULSE REGULAR OR IRREGULAR?; RACE/ETHNICITY - RECODE; RACE/HISPANIC ORIGIN; SEVERE PAIN IN CHEST MORE THAN HALF HOUR; SKIN REACTION TO SUN AFTER NON-EXPOSURE; SODIUM (MMOL/L); SODIUM: SI (MMOL/L); STANDARD GENERIC INGREDIENT NAME; STILL HAVE ASTHMA; TAKEN PRESCRIPTION MEDICINE, PAST MONTH; TAKEN PRESCRIPTION MEDICINE/PAST MONTH; TAKING INSULIN NOW; TAKING PRESCRIPTION FOR HYPERTENSION; TAKING TREATMENT FOR ANEMIA/PAST 3 MOS; TOLD HAD HIGH BLOOD PRESSURE - 2+ TIMES; TOTAL BONE MINERAL DENSITY (G/CM$\hat{2}$); TOTAL CALCIUM (MMOL/L); TOTAL CHOLESTEROL (MG/DL); TOTAL CHOLESTEROL (MMOL/L); TOTAL CHOLESTEROL(MG/DL); TOTAL CHOLESTEROL(MMOL/L); TRIGLYCERIDE (MMOL/L); TRIGLYCERIDES (MG/DL); TRIGLYCERIDES, REFRIG SERUM (MG/DL); TRIGLYCERIDES, REFRIGERATED (MG/DL); URIC ACID (MG/DL); WHAT KIND OF CANCER; WHICH TYPE OF ARTHRITIS WAS IT?, WHICH TYPE OF ARTHRITIS

Table 7.3: List of NHANES variables used.

BED	***QUERY***
ACNE VULGARIS ALL	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Acne
ALZHEIMER DISEASE	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Alzheimer s disease, unspecified
ANGINA PECTORIS	STATUS:ANGINA PECTORIS == yes
ANXIETY	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Anxiety disorder, unspecified
ARRHYTHMIA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Abnormalities of heart beat; Cardiac arrhythmia, unspecified; Tachycardia, unspecified)
ARRHYTHMIA	STATUS:PULSE REGULAR == irregular
ARTHRALGIA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Pain in hip; Pain in shoulder; Cervicalgia; Pain in knee; Dorsalgia, unspecified; Pain in joint; Dorsalgia)
ASTHMA	STATUS:ASTHMA TODAY == yes
ASTHMA ALLERGIC	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Prevent asthma; Asthma)
ATHEROSCLEROSIS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Atherosclerosis of native arteries of extremities; Atherosclerosis; Prevent atherosclerosis)
ATRIAL FIBRILLATION	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Unspecified atrial fibrillation; Unspecified atrial fibrillation and atrial flutter)

ATTENTION DEFICIT DISORDER WITH HYPERACTIVITY	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Attention-deficit hyperactivity disorders
BACTERIAL INFECTION	STATUS:BACTERIAL INFECTION ANY == yes
BIPOLAR DISORDER	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Bipolar disorder, unspecified
BLURRED VISION	STATUS:BLURRED VISION == blurred vision
BREAST NEOPLASMS	STATUS:CANCER TYPE == breast
CARDIOMYOPATHY	STATUS:CARDIOPATHY PAIN IN CHEST == yes
CHRONIC OBSTRUCTIVE PULMONARY DISEASE	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Chronic obstructive pulmonary disease, unspecified
CHRONIC OBSTRUCTIVE PULMONARY DISEASE	STATUS:COPD == yes
COLORECTAL NEOPLASMS	STATUS:CANCER TYPE == OR(colon; rectum (rectal))
CONJUNCTIVITIS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Conjunctivitis
CONSTIPATION	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Constipation
COUGH	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Cough
CROHN DISEASE	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Crohn s disease [regional enteritis]
DEPRESSION	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Manic episode; Major depressive disorder, recurrent, unspecified; Unspecified mood [affective] disorder; Major depressive disorder, single episode, unspecified)
DERMATITIS CONTACT SUSCEPTIBILITY	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Dermatitis, unspecified; Atopic dermatitis, unspecified)
DIABETES TYPE I	STATUS:DIABETES == yes -AND- STATUS:DIABETES INSULIN NOW == yes
DIABETES TYPE II	STATUS:DIABETES == yes -AND- STATUS:DIABETES INSULIN NOW == no
DIABETIC NEPHROPATHY	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Prevent diabetic kidney disease; Type 2 diabetes mellitus with kidney complications)
DIABETIC NEUROPATHIES	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Type 2 diabetes mellitus with neurological complications
DIABETIC RETINOPATHY	STATUS:DIABETES RETINOPATHY == yes
DIARRHEA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Infectious gastroenteritis and colitis, unspecified
DIZZINESS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Dizziness and giddiness
DYSPEPSIA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Functional dyspepsia
DYSPNEA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Shortness of breath; Prevent breathing difficulty; Unspecified abnormalities of breathing)
EDEMA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Edema, unspecified
ERECTILE DYSFUNCTION	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Male erectile dysfunction, unspecified
FEVER	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Fever, unspecified
FIBROMYALGIA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Fibromyalgia
GASTROESOPHAGEAL REFLUX DISEASE	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Gastro-esophageal reflux disease
GLAUCOMA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Glaucoma
GOUT	STATUS:GOUT == yes
HEADACHE	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Prevent migraine; Migraine; Headache)

HEART FAILURE	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Heart failure, unspecified
HEPATOCELLULAR CARCINOMA	STATUS:CANCER TYPE == liver
HODGKINS LYMPHOMA	STATUS:CANCER TYPE == lymphoma/hodgkins disease
HYPERCALCEMIA	LAB:CALCIUM TOTAL(mmol/L) > 2.6 -AND- DEMO:AGE > 17
HYPERCHOLESTEROLEMIA	LAB:CHOLESTEROL TOTAL(mg/dL) > 240 -AND- DEMO:AGE > 17
HYPERCHOLESTEROLEMIA	LAB:CHOLESTEROL TOTAL(mmol/dL) > 6.2 -AND- DEMO:AGE > 17
HYPERCHOLESTEROLEMIA	LAB:LDL CHOLESTEROL(mg/dL) > 160 -AND- DEMO:AGE > 17
HYPERCHOLESTEROLEMIA	LAB:LDL CHOLESTEROL(mmol/L) > 4.1 -AND- DEMO:AGE > 17
HYPERCHOLESTEROLEMIA	STATUS:HYPERCHOLESTEROLEMIA == yes
HYPERGLYCEMIA	LAB:GLUCOSE SERUM(mmol/L) > 7 -AND- DEMO:AGE > 17
HYPERNANTREMIA	LAB:SODIUM(mmol/L) > 145 -AND- DEMO:AGE > 17
HYPERTENSION	STATUS:HYPERTENSION2 == yes
HYPERTRIGLYCERIDEMIA	LAB:TRIGLYCERIDE(mmol/L) > 1.7 -AND- DEMO:AGE > 17
HYPERURICEMIA	LAB:URIC ACID(mg/dL) > 6 -AND- DEMO:SEX == female -AND- DEMO:AGE > 17
HYPERURICEMIA	LAB:URIC ACID(mg/dL) > 7 -AND- DEMO:SEX == male -AND- DEMO:AGE > 17
HYPOCALCEMIA	LAB:CALCIUM TOTAL(mmol/L) < 2.2 -AND- DEMO:AGE > 17
HYPOCALEMIA	LAB:POTASSIUM(mmol/L) < 3.5 -AND- DEMO:AGE > 17
HYPOGLYCEMIA	LAB:GLUCOSE SERUM(mmol/L) < 2.8 -AND- DEMO:AGE > 17
HYPONANTREMIA	LAB:SODIUM(mmol/L) < 135 -AND- DEMO:AGE > 17
HYPOPHOSPHATEMIA	LAB:PHOSPHORUS(mmol/L) < 0.81 -AND- DEMO:AGE > 17
INSOMNIA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Insomnia
LIPOSARCOMA	STATUS:CANCER TYPE == soft tissue (muscle or fat)
LUNG NEOPLASMS	STATUS:CANCER TYPE == lung
LUPUS ERYTHEMATOSUS SYSTEMIC	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Lupus erythematosus
MELANOMA	STATUS:CANCER TYPE == melanoma
MENOPAUSE	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Menopausal and other perimenopausal disorders
MUSCLE CRAMPS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Cramp and spasm; Prevent cramp and spasm)
MUSCLE SPASMS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Muscle spasm; Cramp and spasm; Prevent cramp and spasm)
MYALGIA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Myalgia
MYOCARDIAL INFARCTION	STATUS:HEART ATTACK == yes
NASOPHARYNGITIS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Prevent common cold; Acute nasopharyngitis [common cold])
NAUSEA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Nausea and vomiting; Vomiting; Nausea; Prevent nausea and vomiting)
NEUROPATHIC PAIN	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Neuralgia and neuritis, unspecified
OBESITY	DEMO:BODY MASS INDEX-kg/m**2 >= 30
OSTEOARTHRITIS	STATUS:ARTHRITIS TYPE == OR(osteoarthritis; osteoarthritis or degenerative arthritis)
OSTEOPOROSIS	LAB:BMD TOTAL(g/cm^2) < 0.86 -AND- DEMO:AGE >= 50
OSTEOSARCOMA	STATUS:CANCER TYPE == bone
OVARIAN NEOPLASMS	STATUS:CANCER TYPE == ovary (ovarian)

PAIN	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Other chest pain; Other acute postprocedural pain; Unspecified abdominal pain; Pain, not elsewhere classified; Abdominal and pelvic pain; Pain in limb, hand, foot, fingers and toes; Pain, unspecified; Chest pain, unspecified)
PANCREATIC NEOPLASMS	STATUS:CANCER TYPE == pancreas (pancreatic)
PARKINSON DISEASE	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Parkinson s disease
PEPTIC ULCER	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Peptic ulcer, site unspecified
PHARYNGITIS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Streptococcal pharyngitis; Acute pharyngitis)
PHOTOSENSITIVITY	STATUS:FOTOSENSIBILITY LEVEL == OR(get a severe sunburn with blisters; a severe sunburn for a few days with peeling)
PROSTATIC NEOPLASMS	STATUS:CANCER TYPE == prostate
PRURITUS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Pruritus, unspecified
PSORIASIS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Psoriasis
RENAL CELL CARCINOMA	STATUS:CANCER TYPE == kidney
RHEUMATOID ARTHRITIS	STATUS:ARTHRITIS TYPE == rheumatoid arthritis
RHINITIS ALLERGIC	STATUS:HAY FEVER LASTYEAR == yes
SEIZURES	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Epilepsy and recurrent seizures
SINUSITIS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Acute sinusitis, unspecified
SKIN ERUPTIONS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Rash and other nonspecific skin eruption
SKIN PIGMENTATION	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Disorder of pigmentation, unspecified
STOMACH NEOPLASMS	STATUS:CANCER TYPE == stomach
STROKE	STATUS:STROKE == yes
TACHYCARDIA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Tachycardia, unspecified
THROMBOPHLEBITIS	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Phlebitis and thrombophlebitis of lower extremities, unspecified
THYROID NEOPLASMS	STATUS:CANCER TYPE == thyroid
THYROIDITIS	STATUS:THYROID PROBLEM TODAY == yes
URINARY BLADDER NEOPLASMS	STATUS:CANCER TYPE == bladder
URINARY INCONTINENCE	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == OR(Urgency of urination; Other specified urinary incontinence; Unspecified urinary incontinence)
URTICARIA	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Urticaria
UTERINE CERVICAL NEOPLASMS	STATUS:CANCER TYPE == OR(cervix (cervical); uterus (uterine))
VAGINAL BLEEDING	DRUG:PRESCRIPTION CONDITION ICD10 DESC ALL == Excessive, frequent and irregular menstruation

Table 7.4: BED indicates the BED term used in the final datasets, QUERY is the query used to trasform the NHANES database into BED.

BED	*DrugBank Indication*	*Sider Side-effect*	*ANNOTATION*
ACNE VULGARIS		Acne; Acneiform eruption; Steroid acne	SE
ALZHEIMER DISEASE			UNK
ANGINA PECTORIS			UNK
ANXIETY		Anxiety	SE
ARRHYTHMIA		Arrhythmia	SE
ARTHRALGIA		Arthralgia	SE
ASTHMA	Asthma	Asthma	BOTH
ASTHMA ALLERGIC	Asthma	Asthma	BOTH
ATHEROSCLEROSIS			UNK
ATRIAL FIBRILLATION			UNK
ATTENTION DEFICIT DISORDER WITH HYPERACTIVITY			UNK
BACTERIAL INFECTION		Infection	SE
BIPOLAR DISORDER			UNK
BLURRED VISION		Vision blurred	SE
BREAST NEOPLASMS			UNK
CARDIOMYOPATHY		Hypertrophic cardiomyopathy	SE
CHRONIC OBSTRUCTIVE PULMONARY DISEASE	Acute Exacerbation of Chronic Obstructive Pulm...		IND
COLORECTAL NEOPLASMS			UNK
CONJUNCTIVITIS	Allergic Conjunctivitis (AC)		IND
CONSTIPATION		Constipation	SE
COUGH			UNK
CROHN DISEASE	Acute Crohn's Disease (CD)	Crohn's disease	BOTH
DEPRESSION		Depression	SE
DERMATITIS CONTACT SUSCEPTIBILITY	Dermatitis, Contact;	Dermatitis; Dermatitis contact	BOTH
DIABETES TYPE I		Diabetes mellitus	SE
DIABETES TYPE II		Diabetes mellitus	SE
DIABETIC NEPHROPATHY			UNK
DIABETIC NEUROPATHIES			UNK
DIABETIC RETINOPATHY			UNK
DIARRHEA		Diarrhoea	SE
DIZZINESS		Dizziness	SE
DYSPEPSIA		Dyspepsia	SE
DYSPNEA			UNK
EDEMA	Edema of the cerebrum	Oedema	BOTH
ERECTILE DYSFUNCTION			UNK
FEVER			UNK
FIBROMYALGIA			UNK
GASTROESOPHAGEAL REFLUX DISEASE		Gastrointestinal disorder; Gastrointestinal ir...	SE
GLAUCOMA		Glaucoma	SE
GOUT	Acute Gouty Arthritis	Gouty arthritis	BOTH
HEADACHE		Headache	SE
HEART FAILURE		Rheumatic heart disease	SE
HEPATOCELLULAR CARCINOMA			UNK
HODGKINS LYMPHOMA	Aggressive Lymphoma		IND
HYPERCALCEMIA	Hypercalcemia of Malignancy	Hypercalcaemia	BOTH
HYPERCHOLESTEROLEMIA		Hypercholesterolaemia	SE

HYPERGLYCEMIA			UNK
HYPERNANTREMIA			UNK
HYPERTENSION		Hypertension	SE
HYPERTRIGLYCERIDEMIA		Hypertriglyceridaemia	SE
HYPERURICEMIA			UNK
HYPOCALCEMIA			UNK
HYPOCALEMIA			UNK
HYPOGLYCEMIA			UNK
HYPONANTREMIA			UNK
HYPOPHOSPHATEMIA			UNK
INSOMNIA		Insomnia	SE
LIPOSARCOMA			UNK
LUNG NEOPLASMS			UNK
LUPUS ERYTHEMATOSUS SYSTEMIC	Systemic Lupus Erythematosus (SLE)	Systemic lupus erythematosus	BOTH
MELANOMA			UNK
MENOPAUSE			UNK
MUSCLE CRAMPS			UNK
MUSCLE SPASMS			UNK
MYALGIA		Myalgia	SE
MYOCARDIAL INFARCTION		Myocardial infarction	SE
NASOPHARYNGITIS			UNK
NAUSEA		Nausea	SE
NEUROPATHIC PAIN		Neuritis	SE
OBESITY		Obesity	SE
OSTEOARTHRITIS	Synovitis of osteoarthritis	Osteoarthritis	BOTH
OSTEOPOROSIS		Osteoporosis	SE
OSTEOSARCOMA			UNK
OVARIAN NEOPLASMS			UNK
PAIN		Arthralgia; Myalgia; Abdominal pain	SE
PANCREATIC NEOPLASMS			UNK
PARKINSON DISEASE			UNK
PEPTIC ULCER		Peptic ulcer	SE
PHARYNGITIS			UNK
PHOTOSENSITIVITY			UNK
PROSTATIC NEOPLASMS	Metastatic Castration Resistant Prostate Cance...		IND
PRURITUS		Pruritus	SE
PSORIASIS	Psoriatic Arthritis; Severe Psoriasis	Psoriasis	BOTH
RENAL CELL CARCINOMA			UNK
RHEUMATOID ARTHRITIS	Rheumatoid Arthritis	Rheumatoid arthritis	BOTH
RHINITIS ALLERGIC	Allergic Rhinitis (Disorder)		IND
SEIZURES			UNK
SINUSITIS	Allergic Rhinitis (Disorder)		IND
SKIN ERUPTIONS		Unspecified disorder of skin and subcutaneous ...	SE
SKIN PIGMENTATION		Unspecified disorder of skin and subcutaneous ...	SE
STOMACH NEOPLASMS			UNK
STROKE			UNK
TACHYCARDIA		Tachycardia	SE

THROMBOPHLEBITIS		Thrombophlebitis	SE
THYROIDITIS	Thyroiditis	Thyroiditis	BOTH
THYROID NEOPLASMS			UNK
URINARY BLADDER NEOPLASMS			UNK
URINARY INCONTINENCE			UNK
URTICARIA		Urticaria	SE
UTERINE CERVICAL NEOPLASMS			UNK
VAGINAL BLEEDING			UNK

Table 7.5: ***BED*** is the name of the BED condition, ***DrugBank Indication*** is the name of the prednisone indication as reported in DrugBank, ***Sider Side-effect*** is the side-effect has indicated in the Sider database, and ***ANNOTATION*** is the label assigned to each BED condition.

7.3. The effect of Tocilizumab in GCA-patiens PBMCS (Peripheral Blood Mononuclear Cells)

AGER, ALOX12, ALOX15, ALOX5, AREG, ARG1, ATF2, BCL2L1, BCL6, BIRC2, BORCS8-MEF2B, C1QA, C1QB, C1R, C1S, C2, C3, C3AR1, C4A, C5, C6, C7, C8A, C8B, C9, CCL11, CCL13, CCL16, CCL17, CCL19, CCL2, CCL20, CCL21, CCL22, CCL23, CCL24, CCL3, CCL4, CCL5, CCL7, CCL8, CCR1, CCR2, CCR3, CCR4, CCR7, CD4, CD40, CD40LG, CD55, CD86, CDC42, CEBPB, CFB, CFD, CFL1, CREB1, CRP, CSF1, CSF2, CSF3, CXCL1, CXCL10, CXCL2, CXCL3, CXCL5, CXCL6, CXCL8, CXCL9, CXCR1, CXCR2, CXCR4, CYSLTR1, CYSLTR2, DAXX, DDIT3, DEFA1, ELK1, FASLG, FLT1, FOS, FXYD2, GNAQ, GNAS, GNB1, GNGT1, GRB2, HDAC4, HIF1A, HLA-DRA, HLA-DRB1, HMGB1, HMGB2, HMGN1, HRAS, HSH2D, HSPB1, HSPB2, IFI44, IFIT1, IFIT2, IFIT3, IFNA1, IFNB1, IFNG, IL10RB, IL11, IL12A, IL12B, IL13, IL15, IL18, IL18RAP, IL1A, IL1B, IL1R1, IL1RAP, IL1RN, IL2, IL21, IL22, IL23A, IL23R, IL3, IL4, IL5, IL6, IL6R, IL7, IL9, IRF1, IRF3, IRF5, IRF7, ITGB2, JUN, KEAP1, KNG1, LIMK1, LTA, LTB, LTB4R, LTB4R2, LY96, MAFF, MAFG, MAFK, MAP2K1, MAP2K4, MAP2K6, MAP3K1, MAP3K5, MAP3K7, MAP3K9, MAPK1, MAPK14, MAPK3, MAPK8, MAPKAPK2, MAPKAPK5, MASP1, MASP2, MAX, MBL2, MEF2A, MEF2C, MEF2D, MKNK1, MMP3, MMP9, MRC1, MX1, MX2, MYC, MYD88, MYL2, NFATC3, NFE2L2, NFKB1, NLRP3, NOD1, NOD2, NOS2, NOX1, NR3C1, OAS2, OASL, OXER1, PDGFA, PIK3C2G, PLA2G4A, PLCB1, PPP1R12B, PRKCA, PRKCB, PTGDR2, PTGER1, PTGER3, PTGER4, PTGFR, PTGIR, PTGS1, PTGS2, PTK2, RAC1, RAF1, RAPGEF2, RELA, RELB, RHOA, RIPK1, RIPK2, ROCK2, RPS6KA5, SHC1, SMAD7, STAT1, STAT2, STAT3, TBXA2R, TCF4, TGFB1, TGFB2, TGFB3, TGFBR1, TLR1, TLR2, TLR3, TLR4, TLR5, TLR6, TLR7, TLR8, TLR9, TNF, TNFAIP3, TNFSF14, TOLLIP, TRADD, TRAF2, TREM2, TSLP, TWIST2, TYROBP

Table 7.6: List of gene used in the differential expression analysis.

7.4. MMPred

Evaluation			Sars-CoV-2 peptide mimicry		
DPA1*01:03/DPB1*02:01					
DPA1*01:03/DPB1*03:01					
DPA1*01:03/DPB1*04:01					
DPA1*01:03/DPB1*04:02					
DPA1*02:01/DPB1*01:01					
DPA1*02:02/DPB1*05:01			DPA1*01:03/DPB1*04:01		DPB1*01:01
DQA1*01:01/DQB1*05:01			DPB1*03:01	DPB1*04:01	DPB1*04:02
DQA1*01:02/DQB1*06:02			DPB1*05:01		DQA1*01:02
DQA1*02:01/DQB1*02:02			DQA1*01:02/DQB1*06:02		
DQA1*03:01/DQB1*03:02			DQA1*01:03		DQA1*03:01
DQA1*03:02/DQB1*04:01			DQA1*03:01/DQB1*03:02		DQA1*03:03
DQA1*05:01/DQB1*02:01			DQA1*05:01	DQA1*05:01/DQB1*02	
DQA1*05:01/DQB1*03:01			DQA1*05:01/DQB1*02:01		
DQA1*05:05/DQB1*03:01		DRB1*01:01	DQA1*05:01/DQB1*03:01		DQB1*02:01
DRB1*01:02	DRB1*01:03	DRB1*03:01	DQB1*02:02	DQB1*03:02	DQB1*03:03
DRB1*03:02	DRB1*03:03	DRB1*03:04	DQB1*04:01	DQB1*05:01	DQB1*06:01
DRB1*04:01	DRB1*04:02	DRB1*04:03	DQB1*06:02	DRB1*01:01	DRB1*03:01
DRB1*04:04	DRB1*04:05	DRB1*04:06	DRB1*04:01	DRB1*04:04	DRB1*04:05
DRB1*04:07	DRB1*07:01	DRB1*08:01	DRB1*04/DQB1*03:02		DRB1*07:01
DRB1*08:02	DRB1*08:03	DRB1*09:01	DRB1*08:02	DRB1*08:03	DRB1*09:01
DRB1*10:01	DRB1*11:01	DRB1*11:02	DRB1*11:01	DRB1*12:01	DRB1*13:01
DRB1*11:03	DRB1*11:04	DRB1*12:01	DRB1*13:02	DRB1*15:01	DRB1*15:02
DRB1*12:02	DRB1*13:01	DRB1*13:02	DRB1*15:03	DRB3*01:01	DRB3*02:02
DRB1*13:03	DRB1*13:04	DRB1*13:05	DRB4*01:01 DRB5*01:01		
DRB1*14:01	DRB1*14:02	DRB1*14:05			
DRB1*15:01	DRB1*15:02	DRB1*16:01			
DRB1*16:02	DRB3*01:01	DRB3*02:02			
DRB3*03:01	DRB4*01:01	DRB4*01:03			
DRB5*01:01					

Table 7.8: List of alleles used for prediction.

Microorganism	BED term	Reference
Enterovirus B	DIABETES TYPE I	[Bergamin and Dib, 2015, Yeung et al., 2011]
Hepatitis B virus	ARTHRALGIA	[Cacoub and Terrier, 2009]
Hepatitis B virus	PRURITUS	[Cacoub and Terrier, 2009]
Hepatitis B virus	PSORIASIS	[Cacoub and Terrier, 2009]
Influenza A virus	DIABETES TYPE I	[Smatti et al., 2019]
Mycobacterium tuberculosis	DERMATOMYOSITIS	[Starshinova et al., 2022]
Mycobacterium tuberculosis	LUPUS ERYTHEMATOSUS SYSTEMIC	[Starshinova et al., 2022]
Mycobacterium tuberculosis	RHEUMATOID ARTHRITIS	[Starshinova et al., 2022]
Neisseria meningitidis	GUILLAIN-BARRE SYNDROME	[Temme et al., 2021]
Rotavirus A	DIABETES TYPE I	[Gómez-Rial et al., 2020]
Rotavirus A	MYASTHENIA GRAVIS	[Gómez-Rial et al., 2020]
Sars-Cov 2	ANEMIA	[Knight et al., 2021, Ehrenfeld et al., 2020]
Sars-Cov 2	DIABETES TYPE I	[Knight et al., 2021, Ehrenfeld et al., 2020]
Sars-Cov 2	GUILLAIN-BARRE SYNDROME	[Knight et al., 2021, Ehrenfeld et al., 2020]
Sars-Cov 2	LUPUS ERYTHEMATOSUS SYSTEMIC	[Knight et al., 2021, Ehrenfeld et al., 2020]
Sars-Cov 2	MYASTHENIA GRAVIS	[Knight et al., 2021, Ehrenfeld et al., 2020]
Sars-Cov 2	RHEUMATOID ARTHRITIS	[Knight et al., 2021, Ehrenfeld et al., 2020]
Sars-Cov 2	THYROIDITIS	[Knight et al., 2021, Ehrenfeld et al., 2020]
Staphylococcus aureus	GUILLAIN-BARRE SYNDROME	[Seyyed Mousavi et al., 2017]
Staphylococcus aureus	HEPATITIS	[Seyyed Mousavi et al., 2017]
Staphylococcus aureus	LUPUS ERYTHEMATOSUS SYSTEMIC	[Seyyed Mousavi et al., 2017]
Staphylococcus aureus	MULTIPLE SCLEROSIS	[Seyyed Mousavi et al., 2017]
Staphylococcus aureus	RHEUMATOID ARTHRITIS	[Seyyed Mousavi et al., 2017]
Staphylococcus aureus	SCHOENLEIN-HENOCH PURPURA	[Seyyed Mousavi et al., 2017]

Table 7.7: This table report the autoimmune disease/condition associated to each microorganism present in BED. Only those microorganism that are related to at least one autoimmune condition/disease that is associated to a BED term are reported. ***Microorganism*** reports the name of the Microorganism, ***BED term*** the name of the BED term for the corresponding autoimmune condition/disease, and ***Reference*** reports the reference from which the information was extracted.

Human autoepitope	AID	Microbial epitope	Organism	Same allele
H1-1 (62-70)	MS	PPE68 (63-68)	MT	-
H1-2 (59-67)	MS	PPE69 (63-68)	MT	-
H1-2 (203-211)	MS	HbhA (167-185)	MT	DRB5*01:01
H1-4 (210-218)	MS	HbhA (168-185)	MT	DRB5*01:01
H1-4 (210-218)	MS	RplV (152-157)	MT	DRB5*01:01
H2BC3 (66-74)	MS	EsxB (18-24)	MT	-
MPO (234-242)	MS	Spike (326-330)	SARS-CoV-2	DRB1*15:01
IFT57 (60-68)	MS	Spike (457-462)	SARS-CoV-2	-
PTPRJ (693-701)	MS	Spike (778-785)	SARS-CoV-2	-
RPL31 (103-116)	MS	Spike (1065-1069)	SARS-CoV-2	-
RPL7A (60-68)	MS	Spike (1210-1218)	SARS-CoV-2	-
ARID4B (540-552)	MS	Replicase polyprotein 1a (972-977)	SARS-CoV-2	-
CCDC97 (248-256)	MS	Replicase polyprotein 1a (972-977)	SARS-CoV-2	-
GLT8D1 (48-56)	MS	Replicase polyprotein 1a (1910-1917)	SARS-CoV-2	-
TGFBI (235-244)	MS	Replicase polyprotein 1a (2147-2155)	SARS-CoV-2	-
VIM (52-64)	MS	Replicase polyprotein 1a (3381-3389)	SARS-CoV-2	-
HLA-A (57-65)	MS	Replicase polyprotein 1a (3991-3999)	SARS-CoV-2	-
HLA-A (57-66)	RA	Replicase polyprotein 1a (3991-3999)	SARS-CoV-2	-
MRPS15 (196-204)	MS	Nucleoprotein (86-93)	SARS-CoV-2	-
PLXDC2 (38-49)	MS	Nucleoprotein (267-274)	SARS-CoV-2	-
ACTA2 (153-161)	MS	ORF3a protein (164-176)	SARS-CoV-2	-
H3-4 (21-29)	MS	ORF8 protein (51-56)	SARS-CoV-2	-
RPL5 (18-26)	MS	Tegument protein UL46 (486-490)	HHV-1	-
JAK2 (102-110)	MS	gE (38-42)	HHV-3	-

Table 7.9: Results of the supervised evaluation. Human autoepitope: human protein name and position in the sequence of the predicted autoepitope; AID: autoimmune disease associated to the microorganism (MS: multiple sclerosis, RA: rheumatoid arthritis); Microbial epitope: microbial protein name and position in the sequence of the known microbial epitope; Organism: corresponding microrganism (MT: *Mycobacterium tuberculosis*, HHV: Human Alphaherpesvirus); Same allele: predicted allele, only shown when the same allele is known to recognise both human and microbial epitopes at the experimental level.

Autoantigen	Epitope	SARS-CoV-2 protein	Aligned Sequence	N positive
MYT1L	EEGDREEEE (125 to 133)	NSP3	DEDEEEGDCEEEE (110 to 122)	1
BAZ1A	VDGDEEEGQSEEEE (1229 to 1242)	NSP3	DEDEEEGDCEEEEFE (110 to 124)	2
CHD5	DDDDEEEEGGCEEEED (398 to 413)	NSP3	DEDEEEGDCEEEE (110 to 122)	2
MCM8	YNYEPLTQL (199 to 207)	NSP5	YNYEPLTQ (237 to 244)	1
ATF7	FVCNAPGCG (7 to 15)	NSP13	PYVCNAPGC (47 to 55)	1
HELZ2	FTVIQGPPG (2169 to 2177)	NSP13	QKYSTLQGPPGTGKS (275 to 289)	11
MOV10	KRFNVAVTRAKAL (903 to 915)	NSP13	NVNRFNVAITRAK (557 to 569)	5
MOV10L1	RFNVAITRPKAL (1131 to 1142)	NSP13	NVNRFNVAITRAK (557 to 569)	5
DNA2	LNVAITRAKH (1000 to 1009)	NSP13	RFNVAITRAK (560 to 569)	4
BRI3	VTRYPANSI (64 to 72)	NSP14	VDRYPANSIV (389 to 398)	7
PARVG	LHLLVALAKRFQ (140 to 151)	NSP15	LHLLIGLAKRF (248 to 258)	8
CALD1	VMSLKNGQI (225 to 233)	NSP16	TAVMSLKEGQI (257 to 267)	2
MICAL3	YKKDKKKKA (1747 to 1755)	N	KKDKKKKADE (369 to 378)	3
SLC35E4	SVLYNLASF (265 to 273)	spike(BA.1-like)	SVLYNLASFS (366 to 375)	1
UNC50	YKYLRRLFR (32 to 40)	spike(BA.4-like)	YNYLRRLFR (447 to 455)	7

Table 7.10: MMPred-predicted autoantigens from similarity to SARS-CoV-2 sequences. N positive: number of positive predictions for the epitope (pairs of prediction method and allele).

Predicted AA	autoepitope	VZV protein	aligned sequence	alignment method
AFF3 (664 to 672)	FIETESSSS	ORF62 (346 to 360)	PKSREFVSSSSSSSS	blastp
UNG (251 to 259)	FLLWGSYAQ	ORF59 (242 to 254)	GLVFMLWGAHAQK	blastp
POLD1 (609 to 618)	IMMAHNLCYY	ORF28 (686 to 701)	LYPSIIQAHNLCFTTL	blastp
POLA1 (867 to 875)	SIIQEFNIC	ORF28 (688 to 700)	PSIIQAHNLCFTT	blastp
LILRA2 (196 to 204)	YAYDSNSPY	gE (74 to 83)	AYDHNSPYIW	blastp
HEG1 (654 to 663)	FFVSDSSSSS	ORF62 (350 to 364)	EFVSSSSSSSSWGSS	blastp
EMD (190 to 198)	FMSSSSSSS	ORF62 (351 to 361)	FVSSSSSSSSW	blastp
ZNF594 (542 to 550)	LECEKTFSQ	ORF28 (888 to 895)	KLECEKTF	blastp
TAOK1 (140 to 149)	YLHSHTMIHH	66 (192 to 208)	RALQYLHNNSIIHRDIK	blastp, psiblast
PHKG2 (142 to 151)	FLHANNIVHH	66 (195 to 209)	QYLHNNSIIHRDIKS	blastp, psiblast
CDKL3 (114 to 122)	YLHSNNIIH	66 (191 to 207)	LRALQYLHNNSIIHRDI	blastp, psiblast
BRAF (565 to 574)	YYLHAKSIIH	66 (192 to 206)	RALQYLHNNSIIHRD	blastp, psiblast
MAPKAPK3 (155 to 164)	FLHSHNIAHH	66 (193 to 208)	ALQYLHNNSIIHRDIK	blastp, psiblast
TAOK3 (136 to 145)	YYLHSHALIH	66 (193 to 207)	ALQYLHNNSIIHRDI	blastp, psiblast
TSSK4 (137 to 146)	YYLHSKSIVH	66 (193 to 206)	ALQYLHNNSIIHRD	blastp, psiblast
MAPKAPK2 (175 to 184)	YLHSINIAHH	66 (193 to 208)	ALQYLHNNSIIHRDIK	blastp, psiblast
TYMS (200 to 208)	FYVVNSELS	ORF13 (186 to 201)	LCQFYVANGELSCQVY	blastp, psiblast
DCLK2 (504 to 513)	YLHGLSIVHH	66 (193 to 207)	ALQYLHNNSIIHRDI	blastp, psiblast
DCLK1 (500 to 509)	YYLHSLNIVH	66 (193 to 207)	ALQYLHNNSIIHRDI	blastp, psiblast
TYMS (164 to 172)	IDTIKTNPD	ORF13 (149 to 164)	LQTVIDTIKTNPESRR	blastp, psiblast
TNNI3K (575 to 583)	YLHNLTQPI	66 (193 to 208)	ALQYLHN–NSIIHRD	psiblast
TAOK2 (140 to 148)	YLHSHNMIH	66 (192 to 207)	RALQYLHNNSIIHRDI	psiblast
SYK (484 to 492)	YLEESNFVH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
RAF1 (457 to 466)	YYLHAKNIIH	66 (192 to 206)	RALQYLHNNSIIHRD	psiblast
TYRO3 (644 to 653)	YLSSRNFIHH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
ARAF (418 to 427)	YYLHAKNIIH	66 (192 to 206)	RALQYLHNNSIIHRD	psiblast
PDGFRB (815 to 824)	FFLASKNCVH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
PDGFRA (807 to 816)	FFLASKNCVH	66 (192 to 206)	RALQYLHNNSIIHRD	psiblast
NLK (253 to 261)	YLHSAGILH	66 (192 to 206)	RALQYLHNNSIIHRD	psiblast
MERTK (712 to 721)	YYLSNRNFLH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
CSNK1E (117 to 126)	YIHSKNFIHH	66 (194 to 207)	LQYLHNNSIIHRDI	psiblast
MAPK6 (141 to 150)	YYIHSANVLH	66 (192 to 206)	RALQYLHNNSIIHRD	psiblast
CSNK1D (117 to 126)	YIHSKNFIHH	66 (194 to 207)	LQYLHNNSIIHRDI	psiblast
CSNK1G1 (154 to 162)	YVHSKNLIY	66 (194 to 207)	LQYLHNNSIIHRDI	psiblast
CSNK1G3 (152 to 160)	YVHSKNLIY	66 (194 to 207)	LQYLHNNSIIHRDI	psiblast
DDR1 (755 to 764)	YLATLNFVHH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
DDR2 (699 to 708)	YLSSLNFVHH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
DSTYK (765 to 774)	IRFLHSQGLH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
CSNK1A1L (126 to 134)	YVHTKNFLH	66 (194 to 207)	LQYLHNNSIIHRDI	psiblast
EPHA2 (728 to 737)	YLANMNYVHH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
EPHB1 (733 to 742)	YYLAEMNYVH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
EPHB3 (747 to 756)	YYLSEMNYVH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
CSNK1A1 (126 to 134)	YVHTKNFIH	66 (194 to 207)	LQYLHNNSIIHRDI	psiblast
KIT (781 to 790)	FLASKNCIHH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
LCK (354 to 362)	FIEERNYIH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
CSF1R (767 to 775)	FLASKNCIH	66 (192 to 207)	RALQYLHNNSIIHRDI	psiblast
LIMK1 (449 to 458)	YYLHSMNIIH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
MAPK1 (138 to 146)	YIHSANVLH	66 (192 to 206)	RALQYLHNNSIIHRD	psiblast
ZAP70 (451 to 459)	YLEEKNFVH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
MAPK13 (138 to 146)	LKYIHSAGV	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast
MAPK3 (155 to 164)	YYIHSANVLH	66 (192 to 206)	RALQYLHNNSIIHRD	psiblast
MAPK4 (138 to 147)	YYIHSANVLH	66 (192 to 206)	RALQYLHNNSIIHRD	psiblast

MAPK7 (169 to 178)	LKYMHSAQVV	66 (191 to 206)	LRALQYLHNNSIIHRD	psiblast
MAPK12 (142 to 151)	YIHAAGIIHH	66 (193 to 206)	ALQYLHNNSIIHRD	psiblast

Table 7.11: The table summarizes the prediction and alignments obtained in the MMPred comparison of VZV and Human protein sequences. The column "Predicted AA" reports the predicted human autoantigens and the position of the overallped predicted autoepitopes, "autoepitope" reports the sequences of the overlapped 9-mers, "VZV protein" reports the gene name of the Varicella Zoster Virus proteins and the position of the overlapped aligned sequence, "aligned sequence" reports the overlapped sequences showing significant alignment, and "alignment method" report the alignment method used.

Chapter 8

Publications related to this thesis

P. Coto-Segura, C. Segú-Vergés, A. Martorell, D. Moreno-Ramírez, G. Jorba, V. Junet, F. Guerri, X. Daura, B. Oliva, C. Cara, O. Suárez-Magdalena, S. Abraham, and J. M. Mas. *A quantitative systems pharmacology model for certolizumab pegol treatment in moderate-to-severe psoriasis. Frontier in Immunology*, 14:1212981, 2023.

A different application of the Therapeutic Performance Mapping System and Clustering tool is utilized in this thesis in subsection 3.4.6 and in subsection 4.3.2. This study is not part of the thesis.

F. Guerri, V. Junet, J. Farrés and X. Daura. *MMPred: a tool to predict peptide mimicry events in MHC class II recognition. Frontiers in Genetics.* Accepted 25 November 2024, Volume 15 - 2024, doi: 10.3389/fgene.2024.1500684.

This paper includes the development and testing of MMPred, described in subsection 1.4.6, section 3.5, and section 4.4 of this thesis.