International Journal of
Molecular Sciences

MDPI

Communication

GeneHarmony: A Knowledge-Based Tool for Biomarker Discovery in Disease: Sjögren's Disease vs. Rheumatoid Arthritis and Systemic Lupus Erythematosus

Micaela F. Beckman [1], Adam Alexander [1], Jean-Luc C. Mougeot [1,2,*] and Farah Bahrani Mougeot [1,*]

[1] Translational Research Laboratories, Department of Oral Medicine/Oral & Maxillofacial Surgery and Cannon Research Center, Atrium Health Carolinas Medical Center, Charlotte, NC 28203, USA; micaela.beckman@atriumhealth.org (M.F.B.); aalexa78@outlook.com (A.A.)
[2] Otolaryngology/Head and Neck Surgery, Wake Forest University School of Medicine, Winston-Salem, NC 27110, USA
* Correspondence: jean-luc.mougeot@atriumhealth.org (J.-L.C.M.); farah.mougeot@atriumhealth.org (F.B.M.)

Abstract

Sjögren's Disease (SjD), Rheumatoid Arthritis (RA), and Systemic Lupus Erythematosus (SLE) are autoimmune diseases with overlapping genetic features, yet the etiologies of these diseases are poorly understood. Using these rheumatic diseases as an example of proof of concept, our aim was to develop a tool that simplifies analysis of gene–disease associations applicable to any disease and to perform comparisons. This tool is meant to provide insights into associated gene symbols and gene expression data to identify candidate biomarkers in common among these diseases. The Diseasesv2.0 and GTExv8 databases were utilized for data collection, providing searchable disease names, affiliated gene symbols, confidence scores (ranging from 0 to 5, with 5 being the most confident), and gene expression across the panel of 54 tissue types present in GTExv8. Data infrastructure was established on a Postgres database using Plotlyv5.17.0 and Streamlitv1.27.2 Python packages. The resulting database was used to investigate the genetic associations among SjD, RA, and SLE, including confidence scores from 2.50 to 5.00. STRINGv12 analysis determined significant pathways (FDR < 0.05). Analysis using our tool revealed the following refined gene associations for each disease: SjD based on 'Sjogren' search term (n = 12 genes), RA (n = 231 genes), and SLE (n = 137 genes). We found seven genes in common, namely, *CD4*, *CD8A*, *IL6*, *IL17A*, *TNFS13B*, *TNF*, and *TRIM21*. With the exception of IL17A, these genes were expressed in tissue types known or suggested to be affected by SjD. STRINGv12 determined significant KEGG pathways involving interleukin signaling, cytokine signaling, and the immune system. We developed a tool that simplifies the data mining process, allowing users to search for diseases of interest and view common gene associations and gene expression. Some of the genes identified through our tool may be further explored to better understand SjD pathogenesis and systemic impact.

Keywords: computational biology; autoimmune disease; Sjögren's disease; Sjögren's syndrome; data mining

check for updates

Academic Editor: Yuriy L. Orlov

Received: 3 June 2025
Revised: 27 June 2025
Accepted: 30 June 2025
Published: 2 July 2025

Citation: Beckman, M.F.; Alexander, A.; Mougeot, J.-L.C.; Bahrani Mougeot, F. GeneHarmony: A Knowledge-Based Tool for Biomarker Discovery in Disease: Sjögren's Disease vs. Rheumatoid Arthritis and Systemic Lupus Erythematosus. *Int. J. Mol. Sci.* **2025**, *26*, 6379. https://doi.org/10.3390/ijms26136379

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Autoimmune diseases are complex disorders where the immune system mistakenly attacks the body's own tissues [1]. Sjögren's Syndrome (SS), recently renamed Sjögren's Disease (SjD), Rheumatoid Arthritis (RA), and Systemic Lupus Erythematosus (SLE) share

overlapping clinical features and complicated diagnoses [2]. Furthermore, our understanding of the underlying mechanisms of these diseases is lacking. Genetic and/or microbial biomarkers could provide insights into disease pathogenesis, aid in early diagnosis, and guide the development of targeted therapies. Recent advances in bioinformatics and genomics offer new ways to explore the genetic basis of diseases. Integrating gene–disease association data with gene expression profiles could reveal novel biomarkers and therapeutic targets. However, the vast amount of data and the specialized knowledge required for analysis still face significant challenges.

To address this, we developed GeneHarmony, a knowledge-based tool to simplify gene–disease associations and gene expression in various tissue types. GeneHarmony leverages a database infrastructure incorporating data from sources like the Diseasesv2.0 database and the Genotype-Tissue Expression (GTExv8) database [3–5]. This tool aims to identify candidate biomarkers and serve as a versatile platform for the study of multiple diseases with overlapping or similar pathogenesis. By streamlining data mining and making complex datasets accessible, further insights into the mechanisms of diseases and similar conditions may be discovered, thereby promoting diagnostic and therapeutic development. GeneHarmony provides information obtained from publicly available resources and is available for download at https://www.github.com/mbeckm01/GeneHarmony.

2. Results

Data Usage and Application Results

Data and accompanying scripts can be downloaded from the GeneHarmonyv1.0 GitHub page (http://www.github.com/mbeckm01/GeneHarmony). There are 40,288 unique diseases included in GeneHarmony for analysis from Diseasesv2.0. GTExv8 data contains gene expression values of 56,200 Ensembl gene ID's for 54 tissue types (excluding normal immune cells). Figure 1 shows the entity-relationship diagram of GeneHarmony.

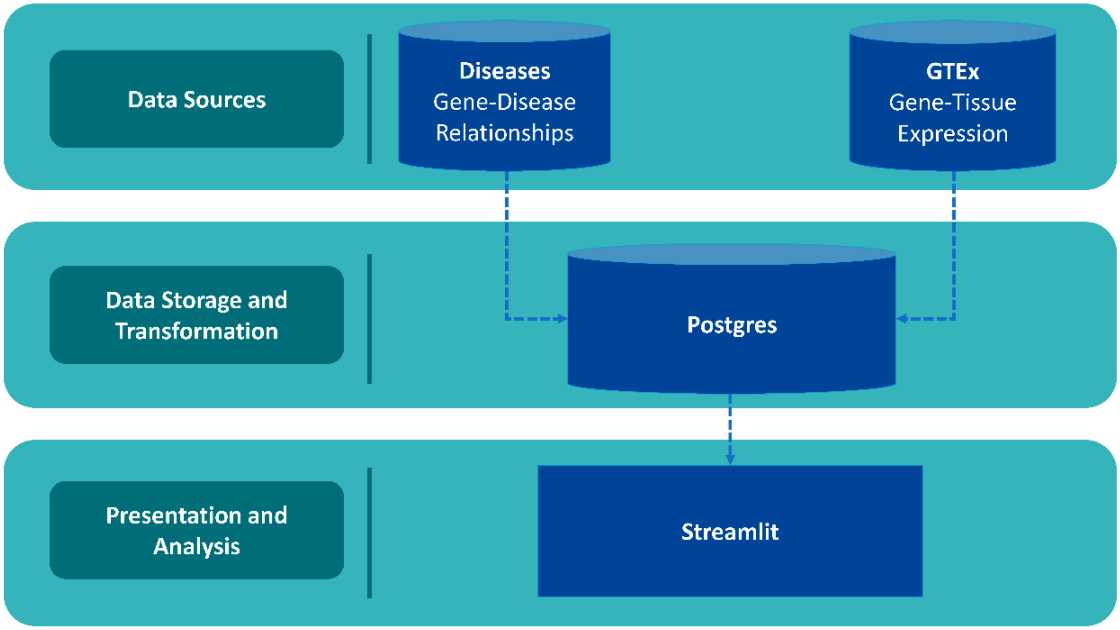

Figure 1. Entity-relationship diagram of GeneHarmony. Legend. Data was collected from the Diseasesv2.0 database applying the Amyloidoses Collection, MedlinePlus, Target Illumination GWAS Analytics (TIGA), UniProtKB-KW, and the Jensen Lab text mining sources for gene–disease associations and the Gene Tissue Expressionv8 (GTExv8) database for gene expression data. Data was utilized through PostgreSQLv16 with front-end development utilizing Streamlitv1.27.2 with Plotlyv5.17.0.

A screenshot of the GeneHarmony-selected disease groups is presented in Figure 2. Initial query results revealed a significant number of genes associated with each disease group: 2000 genes for the first group, 6837 for the second, and 5453 for the third. Among these, 1421 genes were identified as common across all selected disease groups. Seeking to refine these results, we adjusted the gene–disease confidence score filter to a range from 2.50 to 5.00. This adjustment yielded a smaller set of genes: 12 for the first disease group, 231 for the second, and 137 for the third. The associated genes, pre- and post-filtering, are available in Supplemental File S1. Further, sensitivity testing of multiple confidence score ranges with the number of genes in common are presented in Supplemental Table S1. Seven genes were found to be common across all three disease groups after applying the refined filter. A Venn diagram showing the overlapping output can be found in Supplemental Figure S1. The resulting genes in common were CD4 molecule (CD4), CD8 subunit alpha (CD8A), interleukin 17A (IL17A), interleukin 6 (IL6), tumor necrosis factor (TNF), TNF superfamily member 13b (TNFSF13B), and tripartite motif containing 21 (TRIM21).

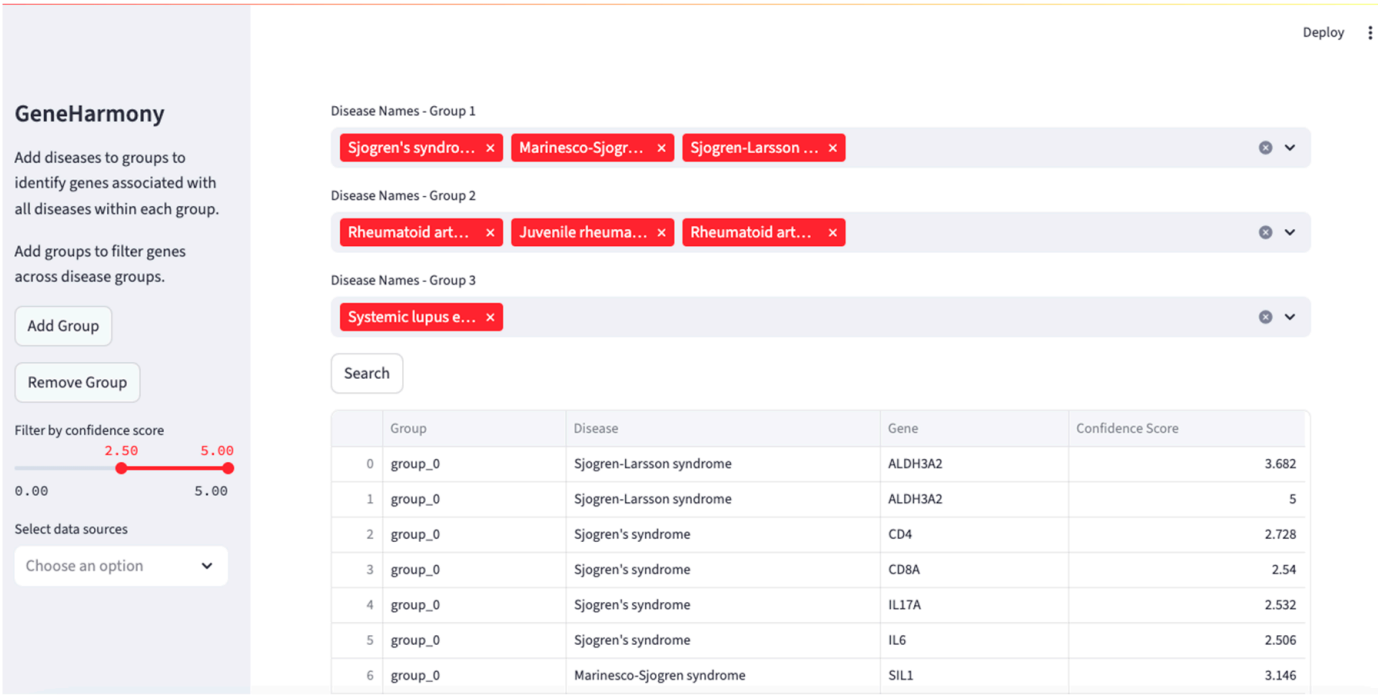

Figure 2. GeneHarmony case study investigating gene interactions between Sjögren's Syndrome (i.e., Sjögren's Disease), Rheumatoid Arthritis, and Systemic Lupus Erythematosus. Legend. Using GeneHarmony, Sjögren's Syndrome (i.e., Sjögren's Disease), Sjögren–Larsson Syndrome, and Marinesco–Sjögren Syndrome were selected for the first disease group. The second group focused on Rheumatoid Arthritis, including 'Rheumatoid Arthritis', 'Juvenile Rheumatoid Arthritis', and 'Rheumatoid Arthritis Interstitial Lung Disease'. The query for the third group included only 'Systemic Lupus Erythematosus'. All available gene–disease data sources were selected and the gene–disease confidence scores were set to span a range of 2.50 to 5.00.

Further interaction with the gene expression heatmap provided insights into the expression levels of the above-mentioned genes in various tissues. Notably, high expression levels were observed in log-adjusted TPM for the following tissues and genes: EBV-transformed lymphocyte tissues for TNF, spleen tissues for *CD8A* and *CD4*, and lung tissues for *CD4*. The heatmap presented within GeneHarmony is presented in Figure 3. IL17A, known to be expressed by Th17 cells and at a lower level by $CD4^+$ regulatory rTh17 cells, was not expressed in any of the 54 tissue types presented in the heatmap [6].

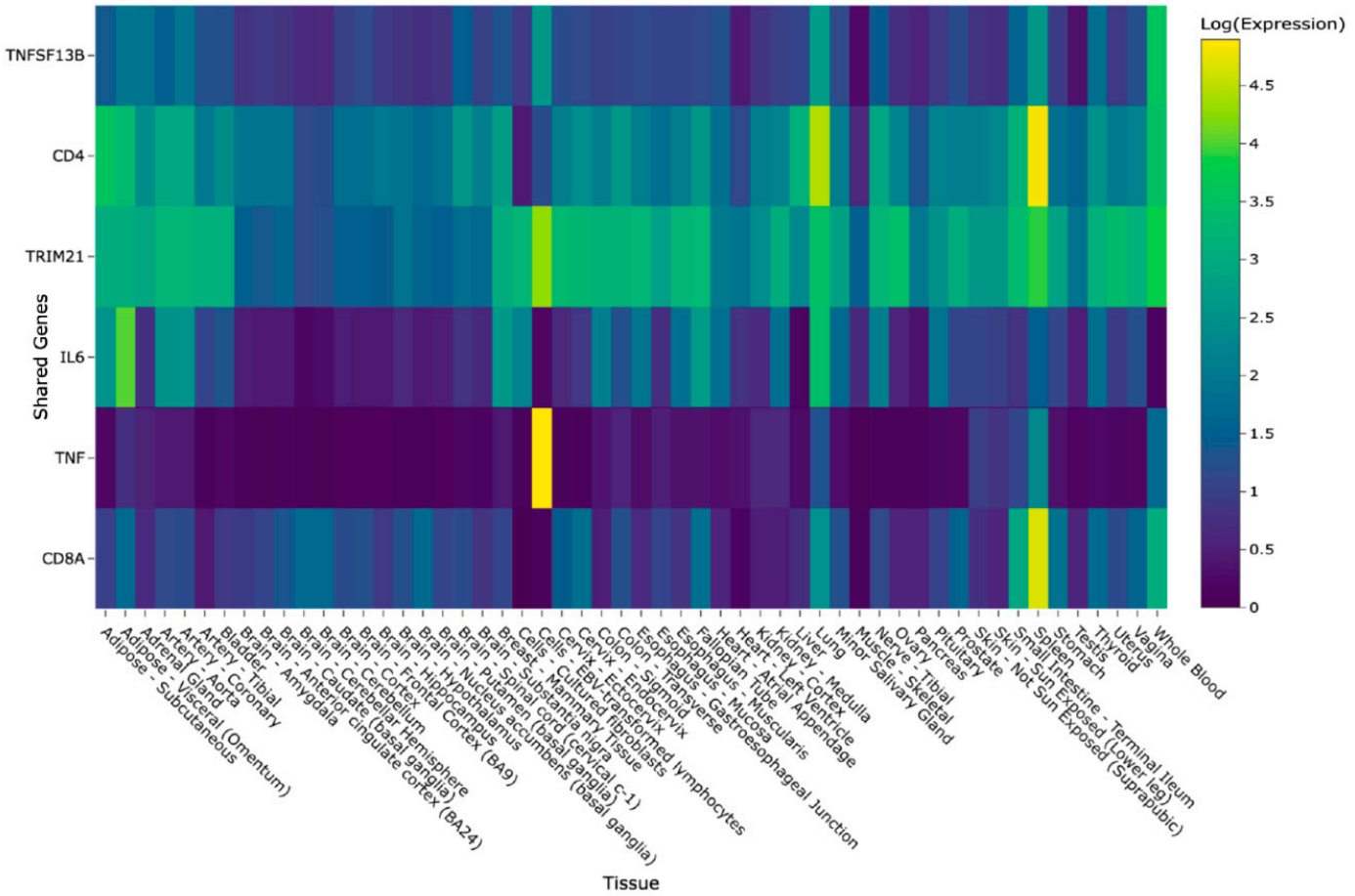

Figure 3. GeneHarmony case study heatmap showing intersecting genes within Sjögren's Syndrome (i.e., Sjögren's Disease), Rheumatoid Arthritis, and Systemic Lupus Erythematosus. Legend. Gene-tissue expression data was retrieved from the Gene Tissue Expression (GTExv8) database using normalized (N = log(N) + 1) transcripts per million (TPM) values. A screenshot from GeneHarmony showing a heatmap of the intersecting genes (n = 6) of three disease groups: (i) Sjögren's Syndrome (i.e., Sjögren's Disease), (ii) Rheumatoid Arthrtitis, and (iii) Systemic Lupus Erythematosus at a confidence score range from 2.00 to 5.00. IL17A, known to be expressed by Th17 cells and at a lower level by $CD4^+$ regulatory rTh17 cells, was not expressed in any of the 54 tissue types presented in heatmap and is not shown.

The Search Tool for Retrieval of Interacting Genes (STRINGv12) analysis at a medium confidence resulted in all genes in common having protein–protein interactions. Additionally, 41 Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways were determined as significant. Significant pathways included Antifolate resistance (hsa01523; FDR = 0.0017), Graft-versus-Host Disease (hsa05332; FDR= 0.0021), Rheumatoid Arthritis (hsa05323; FDR = 4.13×10^{-6}), Hematopoietic cell lineage (hsa04640; FDR = 4.13×10^{-6}), Yersinia infection (hsa05135; FDR = 4.92×10^{-6}), and Systemic Lupus Erythematosus (hsa05322; FDR = 0.0082). The pathways for Rheumatoid Arthritis, Hematopoietic cell lineage, and Yersinia infection resulted in 4 of the 12 genes being included in the gene set. The gene TNF was present in 17 of 20 (85%) of the top KEGG pathways, while IL6 was present in 16 of the 20 (80%). The top 20 pathways are presented in Table 1.

Table 1. Significant KEGG pathways.

KEGG Pathway [a]	Pathway ID [b]	Total # Genes [c]	# Genes Included in Network [d]	Genes Included from Analysis in Network [e]	FDR [f]
Antifolate resistance	hsa01523	31	2	IL6; TNF	0.0017
Graft-versus-host disease	hsa05332	36	2	IL6; TNF	0.0021
African trypanosomiasis	hsa05143	36	2	IL6; TNF	0.0021
Primary immunodeficiency	hsa05340	37	2	CD4; CD8A	0.0021
Inflammatory bowel disease	hsa05321	59	3	IL6; IL17A; TNF	6.91×10^{-5}
Rheumatoid arthritis	hsa05323	83	4	IL6; IL17A; TNF; TNFSF13B	4.13×10^{-6}
Antigen processing and presentation	hsa04612	64	3	CD4; CD8A; TNF	7.29×10^{-5}
Intestinal immune network for IgA production	hsa04672	43	2	IL6; TNFSF13B	0.0024
Hematopoietic cell lineage	hsa04640	90	4	CD4; CD8A; IL6; TNF	4.13×10^{-6}
Malaria	hsa05144	46	2	IL6; TNF	0.0025
Legionellosis	hsa05134	55	2	IL6; TNF	0.0034
IL-17 signaling pathway	hsa04657	91	3	IL6; IL17A; TNF	0.00017
Yersinia infection	hsa05135	124	4	CD4; CD8A; IL6; TNF	4.92×10^{-6}
T cell receptor signaling pathway	hsa04660	100	3	CD4; CD8A; TNF	0.00019
Th17 cell differentiation	hsa04659	99	3	CD4; IL6; IL17A	0.00019
Pertussis	hsa05133	73	2	IL6; TNF	0.0055
Hypertrophic cardiomyopathy	hsa05410	88	2	IL6; TNF	0.0075
Systemic lupus erythematosus	hsa05322	94	2	TNF; TRIM21	0.0082
AGE-RAGE signaling pathway in diabetic complications	hsa04933	96	2	IL6; TNF	0.0082
Viral protein interaction with cytokine and cytokine receptor	hsa04061	96	2	IL6; TNF	0.0082

[a] Top 20 Kyoto Encyclopedia of Genes and Genomes (KEGG) Pathways from the Search Tool for the Retrieval of Interacting Genes (STRINGv12) analysis of 12 genes in common between Sjögren's Syndrome (i.e., Sjögren's Disease), Rheumatoid Arthritis, and Systemic Lupus Erythematosus. [b] KEGG pathway identification. [c] Total number of genes included in the KEGG pathway. [d] Total number of genes included out of the 12 genes in common. [e] Entrez gene symbols of genes included in KEGG pathway. [f] False discovery rate.

3. Discussion

This case study demonstrates the power of GeneHarmony to facilitate complex analyses of gene–disease associations, allowing for a deeper understanding of the genetic overlaps in chronic diseases. Here, we were able to identify important genes in common between Sjögren's Disease, Rheumatoid Arthritis, and Systemic Lupus Erythematosus, as well as their gene expression values in multiple tissue types. Furthermore, we showed how GeneHarmony can be used in downstream analysis to identify significant pathways related to genes in common. Through the application of specific filters and the analysis of gene expression data, we were able to highlight genes and pathways of interest that may play crucial roles in the pathology of these autoimmune diseases.

We were able to determine seven genes in common between the three disease groups, leading to downstream analysis confirming that the genes involved are related to the KEGG pathways of 'Rheumatoid Arthritis' and 'Systemic Lupus Erythematosus'. Using GeneHarmony, we were able to identify TNF (*alias* TNF-α) and IL6 as important genes associated with all three disease groups, which might provide further insight into SjD de-

velopment of autoimmunity systemically. The genes TNF and IL6 were the most important among all significant pathways being present in 85% and 80% of the top 20 significant KEGG pathways, respectively.

TNF may have conflicting roles in the immune system, having the ability to induce inflammatory processes as well as fight infections and provide immunity against pathogens [7]. TNF has been shown to play a role in the pathogenesis of SjD using mice models but its role in the immune system is complex and not fully understood [8]. In SjD, targeting TNF-α has been shown to be ineffective, resulting in continued dysregulated cytokine levels and even increased TNF-α levels after treatment [9]. In RA, TNF activates synovial fibroblasts causing cathepsin and MMP overproduction, resulting in the breakdown of collagen, cartilage, and bones [7]. There are several available TNF inhibitors that have proved to be effective in the reduction in disease progression in patients with RA [10]. A study investigating the role of TNF-α in SLE suggested that anti-autoimmune effects could lead to autoantibodies production and the stimulation of pro-inflammatory mechanisms in SLE development [11]. It has also been suggested that TNF-a inhibitors block access to regulatory TNFR2 and the pro-apoptotic TNFR1 receptors, resulting in deregulated signaling function of TNFR2 [12]. Due to this and the possibility of increased risk of infection, drug-induced Lupus Erythematosus, and possible increased risk of cancer, many healthcare professionals avoid using anti-TNF-α therapies as treatment for SLE [13–15].

A study by Grisius et al. found elevated concentrations of IL6 in serum of patients with SjD compared to healthy controls [16]. IL6 functions in the production of Th17 cells (expressing IL17A) and can activate local B cells. In SjD, this mechanism can lead to the release of more IL6 [17]. Like TNF-α, IL6 has diverse functions in the immune system and has recently become a prospective target for treatment of the disease. Targeting IL6 may disrupt its protective effects while failing to address the multi-inflammatory pathways involved in the development of SjD and SLE [18]. Additionally, IL6 is not seen as the primary driver of inflammation in patients with SjD or SLE. Instead, SjD and SLE are characterized by infiltration of T cells into tissues, creating a positive feedback loop through the production of pro-inflammatory cytokines through the induction of B cell activation [19,20]. Moreover, the ability of Th17 cells to change their characteristics and functions based on their microenvironment proves their status as challenging drug targets to alleviate symptoms and slow down the progression of autoimmune diseases, warranting a broader treatment strategy [21]. Drugs such as rituximab, tocilizumab, and bortezomib have conflicting results in the treatment of SjD, demonstrating inconsistencies, clinical improvements, or lack of systemic involvement improvement in IL6 and Th17/ IL17 [17,22–25].

This tool allowed us to see genes in common and determine possible future diagnostic and/or drug targets that have not previously been investigated for their effectiveness in the treatment of these three diseases. For example, TRIM21 was implicated in all three diseases but we were unable to identify any studies using drugs targeting TRIM21, such as vilazodone, in the treatment of SjD, RA, or SLE despite research suggesting it may lead to reduced pro-inflammatory gene expression and immune modulation [26]. Additionally, CD4 and CD8A are crucial components of the immune system, playing a vital role in T cell activation and function. The drug ibalizumab works to target CD4, binding to the CD4 receptor that prevents HIV from binding to CXCR4 or CCR5. CXCR4 has been shown to have increased expression in SjD patients, causing increased migration of immune cells into lacrimal glands [27]. Moreover, CXCR4 has been shown to be highly expressed in the serum and joint synovial fluid of patients with RA and upregulated in circulating B cells in patients with active SLE, promoting infiltration into renal tissues [28,29]. However, no studies have investigated ibalizumab for its efficacy in these diseases.

Currently, this tool does not consider sex-specific gene expression patterns and their contributions to diseases. In rheumatic diseases such as SjD, RA, and SLE, there is a known disparity in female-to-male incidence. Recently, global studies have placed the female-to-male ratio at a 14:1 ratio for SjD, a 3:1 ratio for RA, and a 4–14:1 ratio for SLE, depending on ethnicity [30–32]. More recent studies have highlighted the relationship between X chromosome presence and risk of developing autoimmune diseases [32,33]. These studies revealed women with Klinefelter (XXY) or Triple X (XXX) syndromes and men with Klinefelter syndrome (XXY) have similar risks of autoimmune diseases. However, it has been determined that women with Turner's syndrome (XO) experience a decreased risk [32,34–36]. A gene found on the X chromosome, *CXorf21*, was found to be upregulated at the mRNA and protein level in women's vs. men's monocytes and overexpressed when under the effects of lipopolysaccharide (LPS) immunological challenge [37,38]. Further, the X-linked gene CXCR3, belonging to the same gene family as the previously mentioned CXCR4, has been implicated for its role in autoimmunity through its ability to direct the migration of T cells to sites of inflammation or injury [32,39]. Additionally, as women age, these genes have been shown evade X-chromosome inactivation, causing a dosage-dependent increase in gene expression and possibly contributing to the development of autoimmune diseases [40].

Imminent updates to GeneHarmony include incorporation of sex-specific gene expression data via the sex-associated gene database (SAGD) [41] to account for sex-based genetic factors and their contributions to diseases such as SjD, RA, and SLE. We also hope to incorporate GTExv8 filtering based on tissue type and expression quantitative trait loci (eQTL) data to identify common genetic variations across diseases of interest and to give insight into mutual tissue specificity and possible regulatory mechanisms. Furthermore, we hope to incorporate microbe set enrichment analysis (MSEA) data to explore the potential impact of LPS-producing bacteria on human gene expression in complex diseases, as well as data from the Human Microbiome Project to examine bacteria–tissue interactions, among other pertinent datasets [42,43]. This analysis has also highlighted the future need for experimental studies to confirm findings and standardized scoring methods to determine the relationships between genes and diseases/phenotypes from experimental, knowledge-based, and/or text mining databases.

In conclusion, while necessitating further development in integrating omics data to investigate rheumatic autoimmune diseases, GeneHarmony provides a valuable and user-friendly initial proof of concept data mining tool.

4. Methods

4.1. Database Architecture and Data Collection

Our selection criteria for integrating data sources emphasize databases with a consistent history of updates, ensuring that our tool remains relevant and adapts to evolving data sources. Diseasesv2.0 was chosen for its gene–disease association data, weekly update schedule, comprehensive analysis from multiple sources, and its methodology for ranking the strength of gene–disease relationships [3]. Full-text ranking was taken into consideration using the scoring method calculated as confidence scores from 0 to 5 [3,5]. There is a reward mechanism built in for genes and diseases mentioned together, and there are penalties for genes and diseases being mentioned in conjunction with other diseases or genes [3,5]. Namely, confidence scores are used to determine how strongly a gene is linked to a disease using text mining. First, the number of times a gene and disease are mentioned together in an abstract or a sentence is determined. Next, a score is calculated in which the score increases if a gene and disease are mentioned frequently together but decreases if the gene is mentioned with many diseases or other genes. Scores are then normalized by being

converted to a z-score to reduce background noise and refined into a confidence score from 0.0 to 5.0 to account for automatic text mining never being as reliable as a human manually determining annotations and associations [3,5].

The GTExv8, offering gene–tissue expression data from RNA-seq experiments, is similarly updated regularly, contributing to the continued improvement in our tool [4]. For visualization, gene–tissue expression data is normalized by placing it in a matrix and applying an adjustment [$N = \log(n + 1)$] to the transcripts per million (TPM) values, creating a heatmap.

The data storage infrastructure, database, and schema were established using PostgreSQLv16 [44]. PostgreSQL is a free and open-source database, also known for its capability to integrate full-text indexing, essential for potential future machine learning applications [44].

For front-end development and user interaction, we used Streamlitv1.27.2 [45]. Despite the availability of many robust web application frameworks, we chose Streamlitv1.27.2 for its ability to quickly integrate data visualizations from various libraries commonly used in bioinformatics. This tool is easy to use by bioinformaticians without the need for extensive web application development experience. We implemented Streamlitv1.27.2 data frames for displaying query results, while matplotlibv3.10 and Plotlyv5.17.0 visualizations were used for the graphical representation of common gene counts and gene expression data, respectively [45–47].

4.2. Download and Usage

GeneHarmony requires the installation of miniconda, pip, psycopg2v2.9.10, and Streamlitv1.27.2, as well as matplotlibv3.10.3, pandasv2.3.0, numpyv2.3.1, kaleidov1.0.0, and plotlyv5.17.0 pythonv4.3.0 packages [45–52]. Furthermore, PostgreSQL is a prerequisite for GeneHarmony. All database files can be downloaded from OneDrive. The 'create_disease_names.sql' script within the 'Streamlit ETL directory' is used to create a streamlined table for the Streamlitv1.27.2 application while 'modify_gtex_tpm.py' will set up all gene expression data for use. Streamlitv1.27.2 can be started using the following command: 'Streamlit run streamlit.py'. From here, a browser should appear, and the local IP address should appear in the terminal after running the command. The initial setup does not require any user data.

User queries are initiated by selecting a disease name from a textbox. Multiple disease names can be selected for related diseases, allowing the identification of genes in common between similarly named diseases. Additional disease groups can be added to the query to identify genes intersecting between the disease groups. When the query is executed, genes associated with each disease group are returned, along with their intersection, which is then visualized via a Venn diagram and a heatmap. Users can further refine the results using filters such as the strength of the gene–disease relationship and the source of the gene–disease association.

4.3. Application: Utilizing GeneHarmony for Cross-Disease Gene Association Analysis

In this case study, we employed GeneHarmony to investigate the genetic associations among Sjögren's Syndrome, Rheumatoid Arthritis, and Systemic Lupus Erythematosus. The exploration began by entering the term "Sjogren" into the search field, which autopopulated with three related diseases: Sjögren's Syndrome, Sjögren–Larsson Syndrome, and Marinesco–Sjögren Syndrome, all of which were selected for the first disease group. The second group focused on Rheumatoid Arthritis, including 'Rheumatoid Arthritis', 'Juvenile Rheumatoid Arthritis', and 'Rheumatoid Arthritis Interstitial Lung Disease'. The query for the third group included only 'Systemic Lupus Erythematosus'.

The analysis proceeded without the application of any filters. All available gene–disease data sources were selected, encompassing AmyCo, MedlinePlus, Target Illumination GWAS analytics (TIGA), Jensen lab texting mining sources, and UniProtKB-KW [53–56]. The gene–disease confidence score was set to span the full range from 0.00 to 5.00. Results were then refined to only include genes scoring within the top 50% of possible confidence scores between 2.50 and 5.00 to limit false positives. Genes in common were then used for STRINGv12 analysis to determine any possible protein–protein interactions using the STRINGv12 database (www.string-db.org) at a medium confidence (0.400) [57]. A Venn diagram was used to visualize the counts of overlapping genes within each disease group. A heatmap was created within GeneHarmony to visualize gene expression of the overlapping genes in 54 tissue types.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms26136379/s1.

Author Contributions: Conceptualization, J.-L.C.M., F.B.M., M.F.B. and A.A.; methodology, F.B.M., J.-L.C.M., M.F.B. and A.A.; formal analysis, A.A. and M.F.B.; interpretation, J.-L.C.M., F.B.M., A.A. and M.F.B.; writing—review and editing, F.B.M., J.-L.C.M., M.F.B. and A.A. All authors have read and agreed to the published version of the manuscript.

Funding: This study was supported by the Atrium Health Foundation Research fund.

Data Availability Statement: All supporting files and documentation for the download and use of GeneHarmony can be found on our lab's Github page (www.github.com/mbeckm01/GeneHarmony).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Wang, L.; Wang, F.S.; Gershwin, M.E. Human autoimmune diseases: A comprehensive update. *J. Intern. Med.* **2015**, *278*, 369–395. [CrossRef] [PubMed]
2. Ortíz-Fernández, L.; Martín, J.; Alarcón-Riquelme, M.E. A Summary on the Genetics of Systemic Lupus Erythematosus, Rheumatoid Arthritis, Systemic Sclerosis, and Sjögren's Syndrome. *Clin. Rev. Allergy Immunol.* **2023**, *64*, 392–411. [CrossRef] [PubMed]
3. Grissa, D.; Junge, A.; Oprea, T.I.; Jensen, L.J. Diseases 2.0: A weekly updated database of disease-gene associations from text mining and data integration. *Database J. Biol. Databases Curation* **2022**, *2022*, baac019. [CrossRef] [PubMed]
4. Lonsdale, J.; Thomas, J.; Salvatore, M.; Phillips, R.; Lo, E.; Shad, S.; Hasz, R.; Walters, G.; Garcia, F.; Young, N.; et al. GTEx Consortium the Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **2013**, *45*, 580–585. [CrossRef]
5. Pletscher-Frankild, S.; Pallejà, A.; Tsafou, K.; Binder, J.X.; Jensen, L.J. DISEASES: Text mining and data integration of disease-gene associations. *Methods* **2015**, *74*, 83–89. [CrossRef]
6. Mills, K.H.G. IL-17 and IL-17-producing cells in protection versus pathology. *Nat. Rev. Immunol.* **2023**, *23*, 38–54. [CrossRef]
7. Jang, D.I.; Lee, A.H.; Shin, H.Y.; Song, H.R.; Park, J.H.; Kang, T.B.; Lee, S.R.; Yang, S.H. The Role of Tumor Necrosis Factor Alpha (TNF-α) in Autoimmune Disease and Current TNF-α Inhibitors in Therapeutics. *Int. J. Mol. Sci.* **2021**, *22*, 2719. [CrossRef]
8. Zhou, J.; Kawai, T.; Yu, Q. Pathogenic role of endogenous TNF-α in the development of Sjögren's-like sialadenitis and secretory dysfunction in non-obese diabetic mice. *Lab. Investig. A J. Tech. Methods Pathol.* **2017**, *97*, 458–467. [CrossRef]
9. Moutsopoulos, N.M.; Katsifis, G.E.; Angelov, N.; Leakan, R.A.; Sankar, V.; Pillemer, S.; Wahl, S.M. Lack of efficacy of etanercept in Sjögren syndrome correlates with failed suppression of tumour necrosis factor alpha and systemic immune activation. *Ann. Rheum. Dis.* **2008**, *67*, 1437–1443. [CrossRef]
10. Santos-Moreno, P.; Rodríguez-Vargas, G.S.; Rodríguez-Linares, P.; Ibatá, L.; Martínez, S.; Rodríguez-Florido, F.; Rojas-Villarraga, A. Effectiveness and Persistence of Anti-TNFα Treatment in Patients with Rheumatoid Arthritis—A 7 Years Real-World Cohort Study. *Biol. Targets Ther.* **2024**, *18*, 339–347. [CrossRef]
11. Aringer, M.; Smolen, J.S. The role of tumor necrosis factor-alpha in systemic lupus erythematosus. *Arthritis Res. Ther.* **2008**, *10*, 202. [CrossRef] [PubMed]
12. Richter, P.; Macovei, L.A.; Mihai, I.R.; Cardoneanu, A.; Burlui, M.A.; Rezus, E. Cytokines in Systemic Lupus Erythematosus-Focus on TNF-α and IL-17. *Int. J. Mol. Sci.* **2023**, *24*, 14413. [CrossRef]
13. Fischer, R.; Kontermann, R.E.; Pfizenmaier, K. Selective Targeting of TNF Receptors as a Novel Therapeutic Approach. *Front. Cell Dev. Biol.* **2020**, *8*, 401. [CrossRef]

14. Ghorbaninezhad, F.; Leone, P.; Alemohammad, H.; Najafzadeh, B.; Nourbakhsh, N.S.; Prete, M.; Malerba, E.; Saeedi, H.; Tabrizi, N.J.; Racanelli, V.; et al. Tumor necrosis factorα in systemic lupus erythematosus: Structure, function and therapeutic implications (Review). *Int. J. Mol. Med.* **2022**, *49*, 43. [CrossRef]
15. Zhu, L.J.; Yang, X.; Yu, X.Q. Anti-TNF-alpha therapies in systemic lupus erythematosus. *J. Biomed. Biotechnol.* **2010**, *2010*, 465898. [CrossRef] [PubMed]
16. Grisius, M.M.; Bermudez, D.K.; Fox, P.C. Salivary and serum interleukin 6 in primary Sjögren's syndrome. *J. Rheumatol.* **1997**, *24*, 1089–1091.
17. Youinou, P.; Pers, J.O. Disturbance of cytokine networks in Sjögren's syndrome. *Arthritis Res. Ther.* **2011**, *13*, 227. [CrossRef] [PubMed]
18. Kaneko, Y.; Takeuchi, T. An update on the pathogenic role of IL-6 in rheumatic diseases. *Cytokine* **2021**, *146*, 155645. [CrossRef]
19. Verstappen, G.M.; Kroese, F.G.M.; Bootsma, H. T cells in primary Sjögren's syndrome: Targets for early intervention. *Rheumatology* **2021**, *60*, 3088–3098. [CrossRef]
20. Li, H.; Boulougoura, A.; Endo, Y.; Tsokos, G.C. Abnormalities of T cells in systemic lupus erythematosus: New insights in pathogenesis and therapeutic strategies. *J. Autoimmun.* **2022**, *132*, 102870. [CrossRef]
21. Yang, P.; Qian, F.Y.; Zhang, M.F.; Xu, A.L.; Wang, X.; Jiang, B.P.; Zhou, L.L. Th17 cell pathogenicity and plasticity in rheumatoid arthritis. *J. Leukoc. Biol.* **2019**, *106*, 1233–1240. [CrossRef]
22. St Clair, E.W.; Levesque, M.C.; Prak, E.T.; Vivino, F.B.; Alappatt, C.J.; Spychala, M.E.; Wedgwood, J.; McNamara, J.; Moser Sivils, K.L.; Fisher, L.; et al. Autoimmunity Centers of Excellence Rituximab therapy for primary Sjögren's syndrome: An open-label clinical trial and mechanistic analysis. *Arthritis Rheum.* **2013**, *65*, 1097–1106. [CrossRef] [PubMed]
23. Jakez-Ocampo, J.; Atisha-Fregoso, Y.; Llorente, L. Refractory primary Sjögren syndrome successfully treated with bortezomib. *J. Clin. Rheumatol. Pract. Rep. Rheum. Musculoskelet. Dis.* **2015**, *21*, 31–32. [CrossRef] [PubMed]
24. Shin, Y.W.; Lee, S.T.; Kim, T.J.; Jun, J.S.; Chu, K. Bortezomib treatment for severe refractory anti-NMDA receptor encephalitis. *Ann. Clin. Transl. Neurol.* **2018**, *5*, 598–605. [CrossRef] [PubMed]
25. Felten, R.; Devauchelle-Pensec, V.; Seror, R.; Duffau, P.; Saadoun, D.; Hachulla, E.; Pierre Yves, H.; Salliot, C.; Perdriger, A.; Morel, J.; et al. Interleukin 6 receptor inhibition in primary Sjögren syndrome: A multicentre double-blind randomised placebo-controlled trial. *Ann. Rheum. Dis.* **2021**, *80*, 329–338. [CrossRef]
26. Eyre, H.; Siddarth, P.; Cyr, N.; Yang, H.; Cole, S.; Forbes, M.; Lavretsky, H. Comparing the Immune-Genomic Effects of Vilazodone and Paroxetine in Late-Life Depression: A Pilot Study. *Pharmacopsychiatry* **2017**, *50*, 256–263. [CrossRef]
27. Hansen, A.; Reiter, K.; Ziprian, T.; Jacobi, A.; Hoffmann, A.; Gosemann, M.; Scholze, J.; Lipsky, P.E.; Dörner, T. Dysregulation of chemokine receptor expression and function by B cells of patients with primary Sjögren's syndrome. *Arthritis Rheum.* **2005**, *52*, 2109–2119. [CrossRef]
28. Peng, L.; Zhu, N.; Mao, J.; Huang, L.; Yang, Y.; Zhou, Z.; Wang, L.; Wu, B. Expression levels of CXCR4 and CXCL12 in patients with rheumatoid arthritis and its correlation with disease activity. *Exp. Ther. Med.* **2020**, *20*, 1925–1934. [CrossRef]
29. Hanaoka, H.; Okazaki, Y.; Hashiguchi, A.; Yasuoka, H.; Takeuchi, T.; Kuwana, M. Overexpression of CXCR4 on circulating B cells in patients with active systemic lupus erythematosus. *Clin. Exp. Rheumatol.* **2015**, *33*, 863–870.
30. Bruno, K.A.; Morales-Lara, A.C.; Bittencourt, E.B.; Siddiqui, H.; Bommarito, G.; Patel, J.; Sousou, J.M.; Salomon, G.R.; Paloka, R.; Watford, S.T.; et al. Sex differences in comorbidities associated with Sjögren's disease. *Front. Med.* **2022**, *9*, 958670. [CrossRef]
31. Maranini, B.; Bortoluzzi, A.; Silvagni, E.; Govoni, M. Focus on Sex and Gender: What We Need to Know in the Management of Rheumatoid Arthritis. *J. Pers. Med.* **2022**, *12*, 499. [CrossRef]
32. Vieira, A.A.; Almada-Correia, I.; Inácio, J.; Costa-Reis, P.; da Rocha, S.T. Female-bias in systemic lupus erythematosus: How much is the X chromosome to blame? *Biol. Sex Differ.* **2024**, *15*, 76. [CrossRef]
33. Panimolle, F.; Tiberti, C.; Spaziani, M.; Riitano, G.; Lucania, G.; Anzuini, A.; Lenzi, A.; Gianfrilli, D.; Sorice, M.; Radicioni, A.F. Non-organ-specific autoimmunity in adult 47,XXY Klinefelter patients and higher-grade X-chromosome aneuploidies. *Clin. Exp. Immunol.* **2021**, *205*, 316–325. [CrossRef]
34. Liu, K.; Kurien, B.T.; Zimmerman, S.L.; Kaufman, K.M.; Taft, D.H.; Kottyan, L.C.; Lazaro, S.; Weaver, C.A.; Ice, J.A.; Adler, A.J.; et al. X Chromosome Dose and Sex Bias in Autoimmune Diseases: Increased Prevalence of 47,XXX in Systemic Lupus Erythematosus and Sjögren's Syndrome. *Arthritis Rheumatol.* **2016**, *68*, 1290–1300. [CrossRef]
35. Cooney, C.M.; Bruner, G.R.; Aberle, T.; Namjou-Khales, B.; Myers, L.K.; Feo, L.; Li, S.; D'Souza, A.; Ramirez, A.; Harley, J.B.; et al. 46,X,del(X)(q13) Turner's syndrome women with systemic lupus erythematosus in a pedigree multiplex for SLE. *Genes Immun.* **2009**, *10*, 478–481. [CrossRef]
36. Scofield, R.H.; Bruner, G.R.; Namjou, B.; Kimberly, R.P.; Ramsey-Goldman, R.; Petri, M.; Reveille, J.D.; Alarcón, G.S.; Vilá, L.M.; Reid, J.; et al. Klinefelter's syndrome (47,XXY) in male systemic lupus erythematosus patients: Support for the notion of a gene-dose effect from the X chromosome. *Arthritis Rheum.* **2008**, *58*, 2511–2517. [CrossRef]

37. Odhams, C.A.; Roberts, A.L.; Vester, S.K.; Duarte, C.S.T.; Beales, C.T.; Clarke, A.J.; Lindinger, S.; Daffern, S.J.; Zito, A.; Chen, L.; et al. Interferon inducible X-linked gene *CXorf21* may contribute to sexual dimorphism in Systemic Lupus Erythematosus. *Nat. Commun.* **2019**, *10*, 2164. [CrossRef]
38. Harris, V.M.; Harley, I.T.W.; Kurien, B.T.; Koelsch, K.A.; Scofield, R.H. Lysosomal pH Is Regulated in a Sex Dependent Manner in Immune Cells Expressing *CXorf21*. *Front. Immunol.* **2019**, *10*, 578. [CrossRef]
39. Groom, J.R.; Luster, A.D. CXCR3 ligands: Redundant, collaborative and antagonistic functions. *Immunol. Cell Biol.* **2011**, *89*, 207–215. [CrossRef]
40. Roberts, A.L.; Morea, A.; Amar, A.; Zito, A.; El-Sayed Moustafa, J.S.; Tomlinson, M.; Bowyer, R.C.E.; Zhang, X.; Christiansen, C.; Costeira, R.; et al. Age acquired skewed X chromosome inactivation is associated with adverse health outcomes in humans. *eLife* **2022**, *11*, e78263. [CrossRef]
41. Mayer, M.; Fey, K.; Heinze, E.; Wick, C.R.; Abboud, M.I.; Yeh, T.L.; Tumber, A.; Orth, N.; Schley, G.; Buchholz, B.; et al. A Fluorescent Benzo[g]isoquinoline-Based HIF Prolyl Hydroxylase Inhibitor for Cellular Imaging. *ChemMedChem* **2019**, *14*, 94–99. [CrossRef]
42. Kou, Y.; Xu, X.; Zhu, Z.; Dai, L.; Tan, Y. Microbe-set enrichment analysis facilitates functional interpretation of microbiome profiling data. *Sci. Rep.* **2020**, *10*, 21466. [CrossRef]
43. Integrative HMP (iHMP) Research Network Consortium. The Integrative Human Microbiome Project. *Nature* **2019**, *569*, 641–648. [CrossRef]
44. PostgreSQLGlobal Development Group. *PostgreSQL*, version 16; PostgreSQLGlobal Development Group: Santa Barbara, CA, USA; Available online: http://www.postgresql.org/ (accessed on 3 October 2023).
45. Snowflake Inc. *Streamlit*, version 1.27.2; Snowflake Inc.: Bozeman, MT, USA; Available online: https://streamlit.io/ (accessed on 12 November 2023).
46. Hunter, J.D. Matplotlib: A 2D Graphics Environment. *Comput. Sci. Eng.* **2007**, *9*, 90–95. [CrossRef]
47. Plotly Technologies Inc. *Collaborative Data Science*; Plotly Technologies Inc.: Montréal, QC, Canada, 2015; Available online: https://plot.ly (accessed on 3 October 2023).
48. *Miniconda*, Version 2-2.4.0; Anaconda Software Distribution. 2016. Available online: https://anaconda.com (accessed on 3 October 2023).
49. The Pip Developers. *Pip*, Version 25.0.1. Available online: https://pip.pypa.io (accessed on 3 October 2023).
50. The Psycopg Team. *Psycopg*, Version 2.9.10. Available online: https://www.psycopg.org (accessed on 3 October 2023).
51. McKinney, W. Pandas: A Foundational Python Library for Data Analysis and Statistics. 2011. Available online: https://www.pandas.pydata.org (accessed on 3 October 2023).
52. Harris, C.R.; Millman, K.J.; van der Walt, S.J.; Gommers, R.; Virtanen, P.; Cournapeau, D.; Wieser, E.; Taylor, J.; Berg, S.; Smith, N.J.; et al. Array programming with NumPy. *Nature* **2020**, *585*, 357–362. [CrossRef]
53. Nastou, K.C.; Nasi, G.I.; Tsiolaki, P.L.; Litou, Z.I.; Iconomidou, V.A. AmyCo: The amyloidoses collection. *Amyloid Int. J. Exp. Clin. Investig. Off. J. Int. Soc. Amyloidosis* **2019**, *26*, 112–117. [CrossRef]
54. *MedlinePlus [Internet]*; National Library of Medicine (US): Bethesda, MD, USA, 2024. Available online: https://medlineplus.gov/ (accessed on 1 July 2020).
55. Yang, J.J.; Grissa, D.; Lambert, C.G.; Bologa, C.G.; Mathias, S.L.; Waller, A.; Wild, D.J.; Jensen, L.J.; Oprea, T.I. TIGA: Target illumination GWAS analytics. *Bioinformatics* **2021**, *37*, 3865–3873. [CrossRef]
56. UniProt Consortium. UniProt: The Universal Protein Knowledgebase in 2023. *Nucleic Acids Res.* **2023**, *51*, D523–D531. [CrossRef]
57. Szklarczyk, D.; Gable, A.L.; Lyon, D.; Junge, A.; Wyder, S.; Huerta-Cepas, J.; Simonovic, M.; Doncheva, N.T.; Morris, J.H.; Bork, P.; et al. STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Res.* **2019**, *47*, D607–D613. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.