

Identification of Clinically Shared Genetic Risk Variants Among Juvenile Type 1 Diabetes and Other Autoimmune Diseases for Drug Discovery Research Using Multi-Omics Analysis

Suhani Pahuja

Received January 26, 2025

Accepted August 11, 2025

Electronic access September 30, 2025

Approximately 8 million people worldwide including 2 million people in the United States are affected by Type 1 Diabetes (T1D) which primarily develops in children under the age of 18. This project aimed to identify shared clinically genetic risk variants and molecular pathways among T1D and other autoimmune diseases such as Rheumatoid Arthritis (RA), Hashimotos Hypothyroidism Disease, and Celiac Disease. The systematic in-depth analysis of GWAS (Genome wide association studies) and dbSNP (single nucleotide polymorphism database) led to the identification of potentially pathogenic coding variants in genes such as PTPN22, FUT2, TYK2 and NCF2. In addition, several non-coding variants were found in regulatory regions of CTLA4, SH2B3, and LRRK2. The structural and functional effect of coding variant of TYK2 was further investigated using the SWISS-MODEL protein modelling platform. These findings contribute to a comprehensive understanding of the genotype-phenotype relationship underlying these common clinically risk variants in pathogenesis of autoimmune diseases and provide a basis for future in vitro and in vivo functional studies using CRISPR (clustered regularly interspaced short palindromic repeats) genome editing strategies.

Introduction

Type 1 Diabetes (T1D) or juvenile diabetes is a chronic disease, characterized by the destruction of pancreatic beta cells, resulting in lifelong loss of insulin secretion. This is a life-threatening condition if left untreated and typically managed through continuous monitoring and external insulin administration. T1D often begins with genetic predisposition, resulting in individuals with certain genetic risk variants at higher likelihood of developing the disease. Diagnosis usually occurs at Stage 1 or Stage 2 when two or more T1D associated auto-antibodies are detected in the blood. However, these stages are often late, after the immune system has already started attacking pancreatic beta cells, leading to abnormal blood glucose levels.

There are ongoing efforts to screen for genetic risk factors early that lead to T1D, allowing for preventative measures to be implemented sooner. Previous studies have¹⁻⁴ indicated that T1D frequently have co-occurring autoimmune disorders (AIDs) such as Rheumatoid Arthritis (RA), Hashimotos thyroiditis, and celiac disease. The autoimmune process underlying T1D can also affect other organs, leading to the development of additional autoimmune diseases and further complicating diabetes management. T1D is most commonly associated with autoimmune thyroid diseases (Hashimotos disease and Graves disease) in 30% of the cases, celiac disease in 8%, and Rheumatoid Arthritis in 1.2%⁴.

The reason co-occurring autoimmune diseases are so preva-

lent remains unclear, though genetics plays a significant role. This study aims to explore the genetic relationships among T1D and other AIDs to support the development of effective prevention strategies, early screening, intervention and treatment.

Previous studies using immunochip analysis have identified shared genetic risk variants between T1D and RA that include CTLA4, STAT4, IL-2RA⁵ and primarily focused on HLA associations. However, none of the previous studies have systematically investigated cross-disease and molecular pathways analysis of shared genetic risk variants among T1D, RA, celiac and Hashimotos disease. This is the first study to examine non-HLA risk association among four different autoimmune diseases using comprehensive genomic and functional pathway analysis. The goal is to identify shared clinically risk variants or cellular pathways linking Type 1 Diabetes (T1D) with these commonly associated autoimmune diseases.

Genome Wide Association Studies (GWAS) detects association between single nucleotide polymorphism (SNPs) and phenotypic traits using genotypic data collected from microarray or next generation sequencing of whole genome or whole exome sequencing^{6,7}. GWAS has successfully identified risk variants or risk alleles for complex diseases such as Type 2 diabetes, Parkinsons disease, Crohns disease, heart diseases and various cancers. It identifies significant association with p value less than 5×10^{-8} , but does not identify causality leaving the biological significance of many variants unknown. The dbSNP database catalogues all human variants including synonymous

and non-synonymous variants, insertions and deletions^{8,9}. It includes both pathogenic (clinically relevant) mutations and common naturally occurring benign mutations. dbSNP database is instrumental in annotating GWAS variants database and filtering out benign variations. All the tools used in this study are open sourced and readily accessible.

This study will help to calculate polygenic risk score which will help in early screening of Type 1 diabetes patients at risk of developing additional autoimmune diseases. The goal of this investigation is eventually to pave the way to potential therapeutic targets for drug discovery research.

Methods

GWAS (<https://www.ebi.ac.uk/gwas/>) genomic variants datasets for (T1D, RA, Hashimotos disease, Celiac disease) were downloaded. Overlap analysis between rsIDs and genes was performed to identify common genetic variants and affected genes amongst all 4 diseases using Venny software tool <https://bioinfogp.cnb.csic.es/tools/venny/>. The rsID number is a unique label ("rs" followed by a number) are used by researchers and databases to identify a specific SNP (Single Nucleotide Polymorphism). Gene enrichment analysis was performed for SNPs using Shiny Go (<https://bioinformatics.sdstate.edu/go/>) to identify pathway enrichment for each of the disease. Overlap genes were further analyzed for protein-protein interactions using String database analysis (<https://string-db.org/>). String database contains information about protein-protein interactions reported through co-occurrence of genes, co-expressions or experimental data. Python programming using Jupyter notebook was used to shortlist clinically relevant variants from a huge dbSNP database (<https://www.ncbi.nlm.nih.gov/snp/>) of 1.5 million variants. rsIDs were input into this script and clinically relevant SNPs were derived with genomic location as output. VaradB <http://www.licpathway.net/VARAdB/> and G-profiler (<https://biit.cs.ut.ee/gprofiler/gost>) analysis was performed to further annotate non-coding and coding variants common amongst all 4 diseases. Structure modeling for WT and mutant TYK2 was performed using SWISS Model and superimposed using Pymol (<https://www.pymol.org/>). Methodology is explained in Figure 1.

Results & Discussion

Genomic analysis identified genetic risk alleles and associated genes common among all 4 autoimmune diseases

Genome-wide association studies (GWAS) datasets for all 4 diseases- Type 1 Diabetes, Rheumatoid Arthritis (RA), Hashimotos Hypothyroidism Disease, and Celiac Disease were obtained

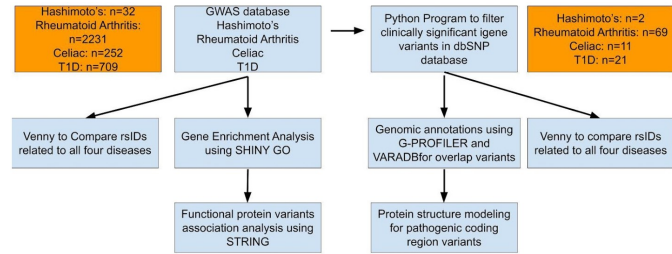


Fig. 1 Flow-chart diagram used for this study. Original database sources used were GWAS and dbSNP.

(Supplementary Data 1). The number of statistically significant risk allele reported identified by rsIDs and mapped genes were about 710 for T1D, 253 for Celiac, 2231 for Rheumatoid Arthritis and 33 for Hashimotos disease. To explore potential shared genetic susceptibility, an overlap analysis of these risk alleles (rsIDs) was conducted using the venny bioinformatics tool. Results are shown as Figure 2A. Two common risk alleles rs72928038-A (BACH2) and rs3184504-C (SH2B3) were found overlapping among all 4 diseases and 48 common alleles were found for T1D, Rheumatoid and celiac diseases. As Hashimotos disease dataset was limited and there were only 2 common alleles identified across all 4 diseases, gene mapped with these risk alleles were further extracted from GWAS database tables (Supplementary Data 1) and Venny overlap analysis was performed to analyze common genes across all 4 diseases (Figure 2B and Table 1). Overlapping genes can suggest potential novel biomarkers and drug discovery targets for T1D, Celiac, Hashimoto's disease and RA, however need to be further validated experimentally. Even though these are statistically significant genes and risk alleles, there is a possibility that these risk alleles are present in other autoimmune diseases and other unrelated diseases which cannot be ruled out and out of scope for this work.

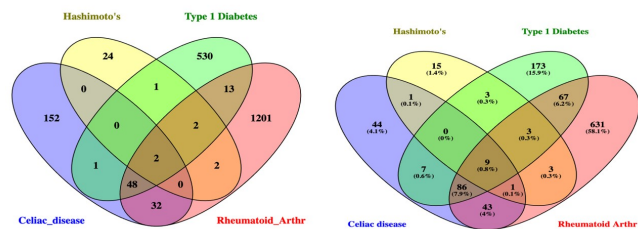


Fig. 2 A. Venny overlap analysis for genetic risk alleles among all 4 diseases. **B.** Venny overlap analysis for common genes among all 4 diseases. *Hashimotos disease has a limited dataset in GWAS.

Gene name	Gene description	Pathway association(s)
BACH2	BTB domain and CNC homolog 2	B-cell and T-cell differentiation
STAT4	Signal transducer and activator of transcription 4	Cytokine signaling
RNU6-474P	RNA, U6 small nuclear 474, pseudogene	Cytokine signaling
CTLA4	Cytotoxic T-lymphocyte associated protein 4	Allograft rejection
HLA-DQA1	HLA-DQA1 antisense RNA 1 (MHC Class II molecule)	Antigen presentation to T cells
ICOS	Inducible T cell costimulator	T cell activation
ATXN2	Ataxin 2	RNA processing and translation
SH2B3	SH2B adaptor protein 3	Cytokine signaling
IL2RA	Interleukin 2 receptor subunit alpha	Allograft rejection, cytokine signaling

Table 1. Common genes and associated molecular pathways identified among Type 1 Diabetes, Rheumatoid Arthritis (RA), Hashimotos Hypothyroidism Disease, and Celiac Disease

Pathway enrichment and gene network interactions analysis indicate JAK-STAT and immune signaling pathway as commonly affected

In order to understand the functional significance of these nine overlapping genes, STRING analysis was performed. String analysis extracts functional proteomic analysis for each of these genes and thread these in the protein interaction network (Figure 3). Most of these genes were involved in T cell regulation for inflammation response and immunosuppression. CTLA4 is majorly expressed in T cells and works as a critical regulator in inflammatory response. Patients with RA and T1D are reported to have lower levels of CTLA4 mRNA and protein levels¹⁰. IL2RA is expressed high in regulatory T cells and involved in T cell immunosuppression for auto, allo antigens¹¹. These common genes can be novel biomarkers and used to develop polygenic risk score for early detection and prevention of these autoimmune diseases.

Pathway enrichment analysis for risk allele associated genes for each of the 4 autoimmune diseases revealed several commonly affected pathways. These shared pathways include allograft rejection, T cell differentiation and the JAK-STAT pathway

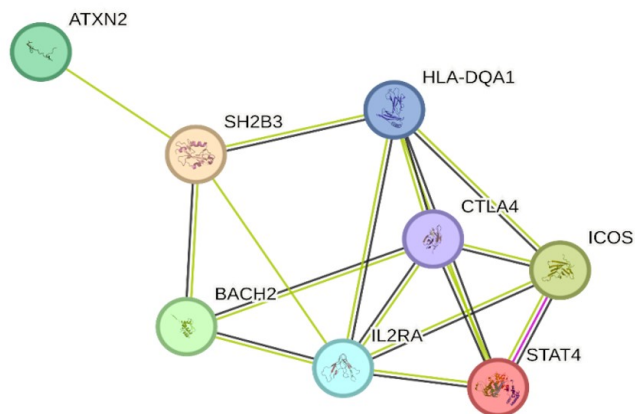


Fig. 3 String analysis of 9 common genes amongst T1D, Hashimotos disease, Celiac and Rheumatoid disease.

indicating a convergence of immune-related molecular mechanisms across diseases. Corresponding KEGG pathway analysis are shown for T1D (Figure 4A), Rheumatoid (Figure 4B), Celiac

(Figure 4C) and Hashimoto's (Figure 4D). The size of the dots represents the number of genes, and the color represents the significance of the FDR value. FDR is False discovery rate and the higher the value of $-\log(\text{FDR})$, smaller is FDR.

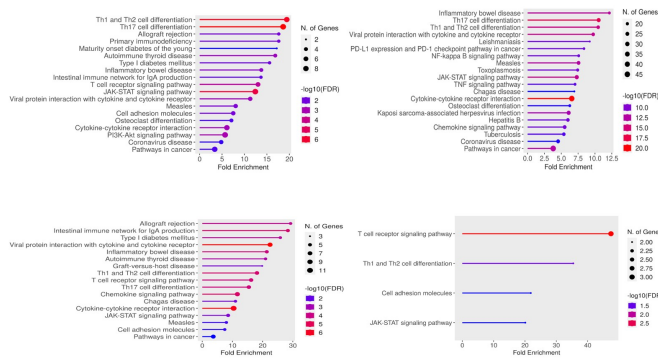


Fig. 4 KEGG pathway analysis for risk alleles associated genes for 4A. Type 1 Diabetes Figure 4B. Rheumatoid Arthritis Figure 4C. Celiac and Figure 4D. Hashimotos disease

Identification and annotation of clinically relevant genetic risk alleles

To address the challenge of analyzing a large dbSNP data containing over 1.5 million SNPs (single nucleotide polymorphism), a custom python program was developed (Supplementary method) to systematically identify clinically significant genetic risk variants associated with 4 autoimmune diseases. Statistically significant risk alleles (rsIDs) derived from GWAS dataset (Supplementary Data 1) were cross-referenced with dbSNP database to retain only those annotated with clinical significance (Supplementary Data 2). All other rsIDs lacking clinical annotations in dbSNP were filtered out. This stringent filtering resulted in a reduced and clinically significant SNP dataset consisting of 21 variants for T1D, 69 variants for RA, 11 variants for Celiac and 2 variants for Hashimotos. Venny analysis was performed again to analyze common variants among all 4 autoimmune diseases and are all listed in Figure 5.

Chromosomal genomic location and annotation were also extracted for each clinically significant common risk alleles. The gene names, genomic coordinates along with annotations for each of these genetic risk variants in all 4 diseases are shown in Figure 6. Further, VaradB analysis was used to validate these common clinically significant risk variants. VARAdB is a useful resource for selecting potential functional variations and interpreting their effects on human diseases and biological processes.

This helped in identifying coding variant (gene: TYK2) and non-coding clinically relevant variants (3UTR in CTLA4 and enhancer in SH2B3) common amongst all 4 diseases.

id	chr	start	end	strand	gene ids	gene names	T1D	RA	Celiac	Hashimoto's
rs1746626	12	40366829	40366829	+	ENSG00000188906	LRRK2	1	1	1	1
rs3184504	12	111446804	111446804	+	ENSG00000111252, ...	SH2B3, ATXN2	1	1	1	1
rs3087243										
rs602662	19	48703728	48703728	+	ENSG00000178920	FUT2	1	1	1	1
rs2476001	1	113834946	113834946	+	ENSG00000134242, ...	PTPRN2, ENSG0000...	1	1	1	1
rs178485502	1	183563445	183563445	+	ENSG00000116686, ...	SMGT1, NCF2	1	1	1	1
rs1893592	21	42434957	42434957	+	ENSG00000160185	UBASH3A	1	1	1	1
rs4625	3	49534707	49534707	+	ENSG00000173402	DAG1	1	1	1	1
rs2304256	19	10364976	10364976	+	ENSG00000105397	TYK2	1	1	1	1
rs12720356	19	10359299	10359299	+	ENSG00000105397	TYK2	1	1	1	1
rs35667974	2	162268127	162268127	+	ENSG00000115267	IFH1	1	1	1	1
rs34536443	19	10352442	10352442	+	ENSG00000105397	TYK2	1	1	1	1

Fig. 5 Genomic locations and Genetic annotations of clinically relevant genetic risk alleles common amongst T1D, celiac disease, Hashimoto's and RA. This information was extracted from dbSNP and varadb.

Clinically relevant mutation identified for TYK2 potentially alters protein function. Tyrosine Kinase 2 (TYK2) variant rs12720356 was identified as the most frequently associated clinically significant risk genetic allele amongst all 4 autoimmune diseases. Tyrosine kinase 2 (TYK2) is an intracellular enzyme that mediates immune signaling and inflammatory signaling pathways (18). This coding variant leads to ILE684 to SER684 mutation. Wild type TYK2 and mutant TYK2 ILE684 to SER684 were modeled using SWISS model and aligned using pymol (cyan ribbon diagram: wt TYK2 (4PO6 with 1.99 resolution), pink ribbon diagram: mutant TYK2. RMSD of two overlapped structures was very low and structures aligned pretty well indicating this mutation is not likely to perturb protein structure. This mutation probably affects function and is earlier reported to be catalytically inactive and has impaired signaling using cell based in vitro kinase activity assay¹². TYK2 activates a series of transcription factors called signal transducer and activator of transcription (STAT)¹³. Activated STATs promote expression of cytokines and cellular processes such as cellular division, differentiation and death¹³. By binding to specific receptors, cytokines signal through TYK2 to regulate the immune system. These cytokines include IL-12, IL-23 and Type I IFNs, which are critical in driving the function of Th1 cells, Th17 cells and the innate immune response.¹⁴ Activation by type I IFNs signaling through TYK2 in the B cell impacts pathways important in autoimmunity such as B cell differentiation, antibody production, and immunoglobulin isotype class switching.

Conclusion

This study analyzed GWAS data across four autoimmune diseases Type 1 Diabetes, Rheumatoid Arthritis (RA), Hashimotos thyroiditis, and celiac disease to identify overlapping genetic variants and molecular pathways involved in disease susceptibility. A total of 3,227 significant risk variants were examined and comparative analysis revealed two shared variants across

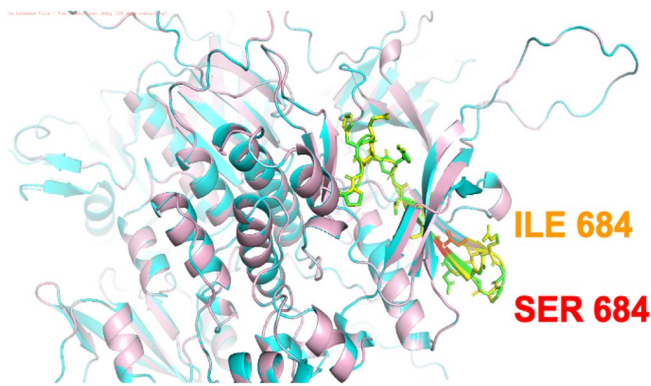


Fig. 6 Swiss modeling was performed using the primary amino acid sequence of wt TYK2 and I684S TYK2. Both structure model ribbon diagrams were overlaid using pymol. cyan ribbon diagram: wt TYK2 (4PO6 with 1.99 resolution), pink ribbon diagram: mutant TYK2. Mutation probably affects catalytic pseudo kinase domain of TYK2¹⁰

all 4 diseases (rs72928038-A in BACH2 and rs3184504-C in SH2B3) with 48 additional alleles common among T1D, Celiac and RA.

Further, gene overlap analysis identified nine common genes, many of which are involved in T-cell regulation, inflammatory response, and immunosuppression. Pathway enrichment and network analysis highlighted the JAK-STAT signalling and immune response pathway as commonly dysregulated, emphasizing their critical role in autoimmunity.

To prioritize clinically relevant risk variants, a filtering pipeline using dbSNP and varadB was implemented. This yielded variants in coding regions for genes such as PTPN22, FUT2, TYK2 and NCF2 as well as non-coding variants in regulatory regions of CTLA4, SH2B3, and LRRK2. TYK2 ILE684 to SER684 mutation was identified as common among all 4 diseases and can be a potential target for drug discovery. Structural modeling suggested this mutation does not alter protein structure significantly but likely impairs TYK2 function, consistent with its previously reported catalytic inactivity.

Future directions

The potential limitation of using GWAS dataset is biased associations as 79% of the population in this study are of European origin. However, it will be good to explore other databases such as OMIM and HGMD in future as these are not very comprehensive yet for autoimmune diseases. It is imperative to validate these findings using functional experimental studies in cell lines or animal models to understand the mechanism of these variants in regulating immune signaling and disease pathogenesis. This clinically relevant genetic risk alleles gene set can be used to develop polygenic risk scores for early diagnosis of celiac,

rheumatoid arthritis and Hashimotos disease in T1D patients and help in development of personalized treatment plans.

References

- 1 *Genetic overlap between type 1 diabetes and other autoimmune diseases*, Review.
- 2 *Type 1 Diabetes Mellitus and Autoimmune Diseases: A Critical Review of the Association and the Application of Personalized Medicine*.
- 3 *Type 1 diabetes and associated autoimmune diseases*.
- 4 *Screening of monogenic autoimmune diabetes among children with type 1 diabetes and multiple autoimmune diseases: is it worth doing?*
- 5 *Meta-analysis of Immunochip data of four autoimmune diseases reveals novel single-disease and cross-phenotype associations*.
- 6 <https://www.ebi.ac.uk/gwas/>, GWAS -.
- 7 *Genome-wide association studies*.
- 8 *Clin VAR database*, <https://www.ncbi.nlm.nih.gov/clinvar/>.
- 9 *The Database of Short Genetic Variation (dbSNP)*. 2013 Jun 30.
- 10 *Current understanding of CTLA-4: from mechanism to autoimmune diseases*.
- 11 *IL-2 immunotherapy for targeting regulatory T cells in autoimmunity*.
- 12 *Two Rare Disease-Associated Tyk2 Variants Are Catalytically Impaired but Signaling Competent*.
- 13 <https://cytoscape.org/>, Cytoscape -.
- 14 *Current understanding of the role of tyrosine kinase 2 signaling in immune responses*.