J Appl Biotechnol Rep. 2025 March;12(1):1570-1585

Journal of
Applied Biotechnology Reports

doi 10.30491/jabr.2024.470731.1767

Original Article

Enrichment Analysis of Significant Variants from Multiple GWAS Datasets for Type 1 Diabetes Mellitus

Naga Vignesh Selvaraj [1], Nalinkanth Veerabadran Ghone [2*]

[1] Department of Biotechnology, Sri Venkateswara College of Engineering, Post Bag No.1, Sriperumbudur Taluk - 602117, Kancheepuram, Tamil Nadu, India
[2] Department of Chemical Engineering, Sri Sivasubramaniya Nadar College of Engineering, Rajiv Gandhi Salai (OMR), Kalavakkam - 603110, Tamil Nadu, India

Corresponding Author: Nalinkanth Veerabadran Ghone, PhD, Associate Professor, Department of Chemical Engineering, Sri Sivasubramaniya Nadar College of Engineering, Rajiv Gandhi Salai (OMR), Kalavakkam - 603110, Tamil Nadu, India. Tel: +91 8122603376, E-mail: nalinkanthv@ssn.edu.in

Received July 30, 2024; **Accepted** October 29, 2024; **Online Published** March 20, 2025

Abstract

Introduction: Genome-wide association studies (GWAS) have identified numerous genetic variants associated with Type 1 Diabetes Mellitus (T1DM). However, GWAS has limitations in detecting the cumulative effects of these variants. The current study aims to understand T1DM genetic architecture based on the cumulative effects of significant genetic variants from GWAS summary statistics, their corresponding genes and their association with regulatory pathways.

Materials and Methods: Understanding the cumulative effects of SNPs in pathways requires an integrated approach of combining results of multiple GWAS studies to perform Over-representation analysis and Gene Ontology enrichment analysis. Based on large sample sizes, 12 independent GWAS summary datasets were selected from the GWAS central database. Enrichment analyses were performed using the WEB-based Gene SeT AnaLysis Toolkit.

Results: Preliminary pathway enrichment analysis results containing 119 variants from the HLA region and 238 variants from other chromosomal regions show 49 gene ontology and four functional pathways as enriched categories. After assigning gene ID to noncoding variants, the pathway enrichment analysis identified 65 gene ontology and 15 functional pathways as enriched categories. Overall results emphasize the role of human leukocyte antigen class II and other significant variants from non-HLA regions correspondingly in immune response pathways and T cell differentiation, cytokine signalling, and immune system regulation pathways.

Conclusions: This study provides a foundation for future studies in understanding T1DM pathogenesis. Further investigation is needed to identify novel genetic signatures to reveal the T1DM pathogenesis mechanisms and detection at the genetic level.

Keywords: Type 1 Diabetes Mellitus, Genome Wide Association Studies, Pathway Enrichment Analysis, Over Representation Analysis, Gene Ontology

Citation: Selvaraj NV, Ghone NV. Enrichment Analysis of Significant Variants from Multiple GWAS Datasets for Type 1 Diabetes Mellitus. J Appl Biotechnol Rep. 2025;12(1):1570-1585. doi:10.30491/jabr.2024.470731.1767

Introduction

Autoimmune disorders are complex diseases with complications in clinical diagnosis due to detecting causing factors. Among them, Type 1 diabetes mellitus (T1DM) is a well-known chronic autoimmune disorder that arises from a complex interplay between genetic and environmental factors.[1] T1DM is a multifaceted disease that does not follow a simple inheritance pattern. T1DM is most often associated with several other multi-system autoimmune diseases, including but not limited to thyroid diseases, parathyroid diseases, celiac disease, vitiligo, gastritis, skin diseases, and rheumatic diseases.[2,3] In addition, T1DM causes excess morbidity and mortality by associating with macrovascular complications such as cardiovascular disease and microvascular diseases such as chronic kidney disease, neuropathy and retinopathy.[4] On a global scale, T1DM affected 1.2 million children and adolescents younger than 20 years as per the 10th edition of the International Diabetes Federation (IDF) Atlas. According to the IDF Atlas, India reported to have a much higher number of 19,194 incident cases of type 1 diabetes in children aged 0–14 years.[5] According to Gregory et al, the prediction model, the global population of individuals with T1DM will reach a staggering 13.5 to 17.4 million by 2040. The reasons for this increased incidence remain unknown. As per the model prediction, among worldwide T1DM affected individuals 64% of 8.4 million belong to the 20–59 age group. Although historically, T1DM has been a disease usually associated with childhood onset, according to the results, a more significant number of adults than children are diagnosed every year, with a mean diagnosis age of 32 years.[6] It further confirms that T1DM can develop at any age and in people of diverse genetic ancestry.[7,8]

Both candidate gene association and genome-wide association studies (GWAS) approaches have resulted in the identification of genes and genomic risk loci to understand the complex disease etiology for T1DM.[9-11] Each GWAS study summarizes the correlation between genetic variants with small to strong significant *p*-values from multiple genomic risk loci and the trait of interest.[12] The genetic determinants of T1DM have been extensively studied through genomic dissection to understand how specific genes contribute to the development of the disease.[13,14] Identifying the shared genetic component in autoimmunity has described several common susceptibility loci with a potential role in T1DM.[15] GWAS identified more prevalent genetic variants that alter T1DM, where genetic variants influence multiple associated autoimmune diseases, indicating shared common causal pathways among them.[16-20] Beyond the successful mapping of thousands of loci associated with complex traits, the traditional GWAS approach has several limitations. Through genomic approaches, researchers have identified more than 50 variants contributing to the genetic risk of T1DM, shedding light on previously unexpected factors involved in this disease.[21-23] However, the remaining genetic risk for T1DM, characterized as non-HLA, is spread across multiple loci throughout the genome.[24] Further research has confirmed at least 57 non-HLA T1DM susceptibility regions, with the most recent study identifying 78 genome-wide significant regions that influence immune cell and pancreatic β-cell functions.[24] It is impossible to decipher all the genomic variants to understand the complex disease genetic architecture. The genomic variants can explain only a tiny fraction, up to 25%, of the heritability of complex traits and up to 56% of disease-related traits.[25] Top-ranked SNPs are considered the tip of the iceberg and reveal only a limited number of associated functions.[26] The complexity of T1DM's genetic landscape is compounded by the presence of "dark variants"—variants that may not be captured across all GWAS databases but could play critical roles in disease etiology.[27]

Interestingly most variants identified through GWAS are disease-associated loci found in non-coding genome regions.[28,29] Understanding the role of non-coding variants such as promoters, long non-coding RNAs (lncRNAs), and untranslated regions particularly in enhancer regions can disrupt gene expression through cis or trans mechanisms. This linkage of genetic variation to T1DM pathogenesis identifies potential targets for therapeutic interventions and confirms that the regulatory pathway controlling gene expression underlies the genetic risk of T1DM.[30] Recent studies indicate that genetic variants from non-coding genomic regions are enriched in cell type-specific transcriptional enhancers.[31,32] However, their positional mapping, regulatory and functional mapping and the cell types or physiological contexts in which this regulation occurs are also unknown.[15]

Identifying genetic association signals driven by common low-impact risk variants provides valuable insights into the mechanisms underlying the disease's pathogenesis and the essential genes involved in T1DM development.[33] While many loci have been identified as being enriched in immune-related pathways related to T1DM, the functional implications of these loci and genetic variants are still largely unknown. Due to the increasing number of genetic variants associated with T1DM, it is crucial to develop additional screening strategies to reduce the high number of false positive results and obtain a significant number of authentic positive results.[34,35] A critical shortcoming of GWAS stems from testing each marker once for its association with the disease. As these studies assess the significance of variants individually, they are likely to miss SNPs that individually contribute little to the disease but may be essential in a collective interaction.[36] However, variant-level associations are hard to identify (because of small effects) and can be difficult to interpret biologically. “Enrichment analyses” help address both these problems by focusing on sets of biologically related variants.[37] Elucidation of this genetic overlap has been useful in identifying key molecular pathways with a pathogenic role in multiple disorders.[38]

Pathway analysis can increase the ability to identify genetic factors relevant to disease mechanisms.[39] This is because it is often easier to detect the accumulation of minor genetic effects that act in a common pathway than to map individual genes in a pathway that contributes to disease susceptibility. This proposition has received support from various investigations such as Yoon et al.,[26] and García-Campos et al.[39] Moreover, most significant GWAS variants reside in non-protein-coding regions of the genome, which presents challenges for functional interpretation.[35]

Over-representation analysis (ORA) is a statistical method used to identify if a particular set of genes or variants is over-represented in a specific biological pathway, suggesting a potential functional connection.[40] Recently, considering the functional importance, validation of non-coding variants and the role of non-coding variants in cancer, cardiomyopathy, and heritable diseases were explored.[41-44]

While the most renowned T1DM genetic variants express risk potentials, we must understand their association with other genes, regulatory mechanisms, and functional significance. Therefore, we need to better understand their clinical interventions.[45] A comprehensive understanding of how these variations affect the underlying pathophysiology is still unknown. Understanding the causes and their consequent immunological pathways involved in T1DM is crucial for developing effective etiopathogenetic treatments. Further research is needed to unravel the complex genetic architecture underlying T1DM and its heterogeneous clinical manifestations.[46] Unfortunately, our knowledge of the molecular mechanisms connecting genetic variation and

environmental triggers with T1DM is limited. Additional experimental data must be analyzed to comprehend the functional consequences of risk-associated variants identified in GWAS.[47]

Reanalyzing GWAS through bioinformatic approaches is a valuable strategy for expanding our understanding of complex diseases.[48] It allows the exploration of disease variants and gene loci crucial in processing and interpreting GWAS summary data and suggests new drug targets for treating rheumatoid arthritis.[49] The current approach is based on gene set enrichment analysis, which aims to investigates the cumulative effect of multiple functionally related genes (e.g., pathways) and provides an alternative yet effective approach for discovering genes associated with regulatory pathways from the original GWAS datasets.

Materials and Methods

GWAS Catalog

The GWAS catalog is the largest manually curated resource based on the latest reference genome version (GRCh38.p13) and variant database version (dbSNP Build 154). The information is a publicly available resource of Findable, Accessible, Interoperable, and Reusable (FAIR) GWAS data. Based on published GWAS, it contains approximately 400,000 curated SNP-trait associations from over 45,000 published GWAS across more than 5,000 human traits, and over 40,000 full *p*-value summary statistics datasets. It is developed by the NHGRI and EMBL-EBI. The catalog data provides the most significant variant-trait associations, structured metadata about the publication, study design, samples, and traits for all published human GWAS.

GWAS Central

GWAS Central is a compilation of genetic variant trait association data and metadata from various GWAS based on HGVbaseG2P.[50] The database contains data from NHGRI-EBI GWAS Catalog, OpenGWAS, Japanese GWASdb, dbGaP, WTCCC, and several other sources. It provides access to summary-level findings from over 3800 studies with 70 million *p*-values, covering 1400 unique phenotypes to study the link between genetic variants and diseases.[51,52] The database provides information based on the query keywords such as specific traits, diseases, or genetic variants of interest, and filters to refine results. It allows users to compare and analyze up to 16 GWAS summary-level datasets simultaneously in a single view using the integrated genome browser.[53] This enables users to correlate diseases, visualize connections between various datasets and gain valuable insights from different perspectives.[54] The meticulous curation process involves verifying the presence of valid dbSNP unique reference numbers for all genetic markers, evaluating the accuracy of alleles and strand representation associated with these markers, removing any duplicate markers, amalgamating numerous data sets for individual studies, and enriching the data with extensive metadata.[48]

Selection of GWAS Summary Statistics Data of Type 1 Diabetes Mellitus

The GWAS summary statistics provide information on the association between SNPs and complex traits. Individual GWAS summary statistics summarize the size of effects, variances, minor allele frequency (MAF), odds ratio, *p*-values for individual SNPs, and correlations of the SNPs.[55] The GWAS studies were selected based on their volume of sample size. This study utilized a range of T1DM GWAS summary statistics obtained from the GWAS Catalog and GWAS Central databases. The T1DM summary dataset incorporated dominant cohorts, such as European ancestry, British Cohort, and Canadian Cohort. Further details are described in supplementary information SI1.

Applying prediction tools such as Variant Effect Predictor,[56] g:Profiler,[57] VannoPortal,[58] and VarNote has significantly advanced the exploration and understanding of genetic variations. Each tool offers a unique approach to annotating, analyzing, and interpreting genetic variations, particularly in non-coding genome regions that require more attention and regulatory pathway exploration.[59] VannoPortal is a diverse database for genetic variation annotations. Integrating information from various biological fields with detailed, genome-wide variant data. This unique combination allows VannoPortal to predict functional scores and provide a comprehensive view encompassing allele frequency, evolutionary signature, linkage disequilibrium and disease association. VarNote uses a random-sweep searching algorithm to extract annotations and filter linked variants in linkage disequilibrium. It is a novel index system with a high-speed parallel intersection algorithm that manages variant annotations and genomic features at scale. It is efficient in non-coding variant functional impact analysis integrated from GenoSkyline-Plus and FUN-LDA for 127 tissue/cell types.[60] The FUN-LDA method utilizes a Latent Dirichlet Allocation (LDA) model to incorporate a variety of epigenetic annotations from prominent genomics initiatives such as ENCODE and Roadmap Epigenomics.[61] Its purpose is to identify genomic regions with probable functional importance in distinct tissues, aiding in the evaluation of the functional impacts of non-coding genetic variants in a tissue- and cell-specific manner. This method is closely linked to GenoSkylinePlus, a tool to predict tissue-specific functional elements by analyzing functional genomics data. Statistically significant variants retrieved from the GWAS catalog and GWAS Central database were analyzed using the Variant Effect Predictor. The nearby gene symbol was assigned either upstream or downstream sequence positions within a 20 kb distance using VarNote.

Different regions of the genome can have different LD patterns. These disease-associated SNPs are typically found in non-coding genome regions, evenly distributed between intergenic and intronic regions. SNPs in high linkage disequilibrium are often functionally related, and their collective association with a disease within a pathway can provide insight into specific regulatory elements or interactions and strengthen the importance of pathway enrichment analysis in identifying high enrichment ratio pathways for SNPs associated with the disease. Considering linkage disequilibrium in over-representation analysis allows for exploration of the associated pathways and a better understanding of the underlying regulatory mechanisms.[62] The close association of alleles due to LD presents a significant obstacle to identifying the specific variant responsible for a particular trait, as its surrounding correlated neighbors are also inherited together.[63] Lower LD thresholds (≥ 0.2) can increase power to detect associations between variants that are in reasonably strong LD without being too stringent but may include more false positives. Higher thresholds (≥ 0.8) are more precise, conservative and can help pinpoint variants that are very likely to be inherited together but might miss some true associations. Multiple LD thresholds can provide a more comprehensive view of the LD structure and its impact on genetic associations. The five categories of genetic variants with definite gene symbols, variants including assigned gene symbols with linkage disequilibrium ranging from 0 (independent inheritance), 0.2 (Weak to moderate LD), 0.6 (Moderate to strong LD) and 0.8 (Strong LD). Table 1 provides the classification of genetic variants with gene symbols and linkage disequilibrium variation level.

Table 1. Classification of Genetic Variants based on Linkage Disequilibrium Variation Level

S.No	Dataset	Category
1	Genetic variants with defined gene symbol	1
2	All the genetic variants (including defined gene symbol and assigned gene symbol) with no LD value limit	2
3	All the genetic variants (including defined gene symbol and assigned gene symbol) with LD value 0.2	3
4	All the genetic variants (including defined gene symbol and assigned gene symbol) with LD value 0.6	4
5	All the genetic variants (including defined gene symbol and assigned gene symbol) with LD value 0.8	5

Enrichment Analysis Using WEB-based GEne SeT AnaLysis Toolkit

Web-based Gene Set Analysis Toolkit, a pathway-based analysis tool for prevalent disease pathways, is a frequently utilized gene set enrichment analysis tool that helps users extract biological insights from genes of interest.[64] To identify the enrichment of disease-related genes among all the genes in each pathway, we used Over-Representation Analysis (ORA), which is a hypergeometric test.[65] We calculated the *p*-value for observing at least "m" disease-related genes in each pathway using the following formula:

$$P = 1 - \sum_{i=0}^{m} \frac{\binom{n}{i}\binom{N-n}{M-i}}{\binom{N}{M}}$$

p-value = 1 - hypergeometric cumulative distribution function (m-1, M, N, n). Here, M represents the total number of genes of interest associated with a given disease, N is the number of reference genes, and n is the number of genes in the pathway. The smaller the *p*-value, the more likely the list of disease-related genes of interest will be overrepresented in this pathway. To avoid testing overly narrow or broad pathways, which could result in some false positives, we only analyzed pathways comprising at least 20 and at most 300 genes.[66] We considered a pathway associated with T1DM significant if its *p*-value was less than 0.05 in disease.

The WebGestaltR function can perform Over-Representation Analysis enrichment analysis. Based on the input gene symbols, the WebGestaltR function will first map the gene list to the Entrez gene IDs and then summarize the gene list based on the GO (Gene Ontology) Slim. After performing the enrichment analysis, the WebGestaltR function also returns a user-friendly HTML report containing the GO Slim summary and the enrichment analysis results. GO enrichment analysis was carried out in the WEB-based GEne SeT AnaLysis Toolkit using default settings, with FDR < 0.05 selected as the cutoff value.

Results

Significant Genetic Variants from GWAS Summary Statistics

From the reported GWAS studies, all 385 statistically significant variants were retrieved from the GWAS catalog and GWAS Central database. The effect of all the genetic variants with significance threshold *p*-values was analyzed using the Variant Effect Predictor and shown in Figure 1. Figure 1(a) shows that among the 385 genetic variants, 357 are unique variants. Synonymous variants are RNA variants

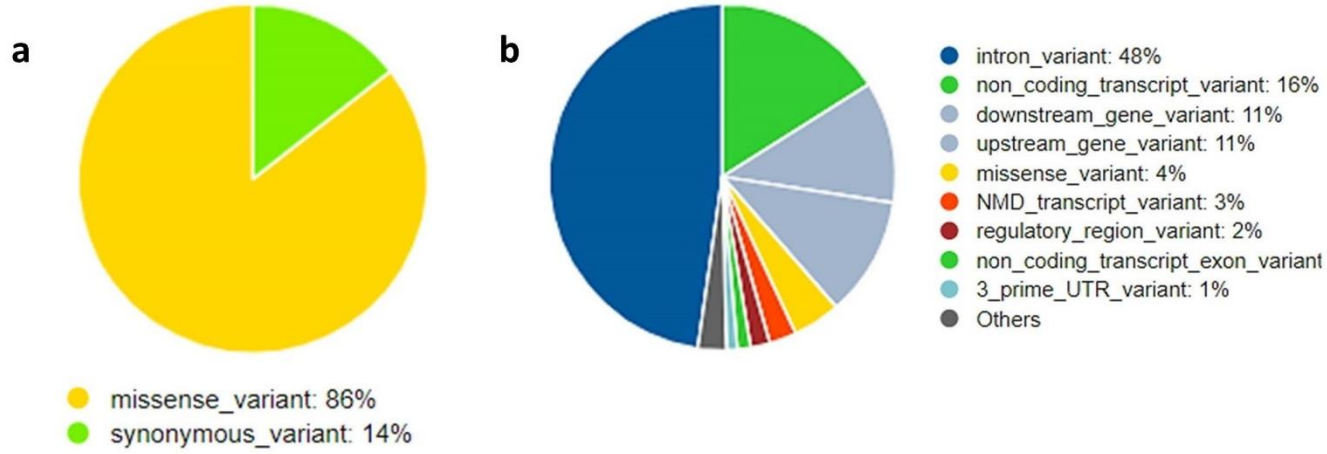

Figure 1. Genetic Variants Predicted Using Variant Effect Predictor Tool **(a)** Type of Genetic variants **(b)** Characteristics of Genetic Variants.

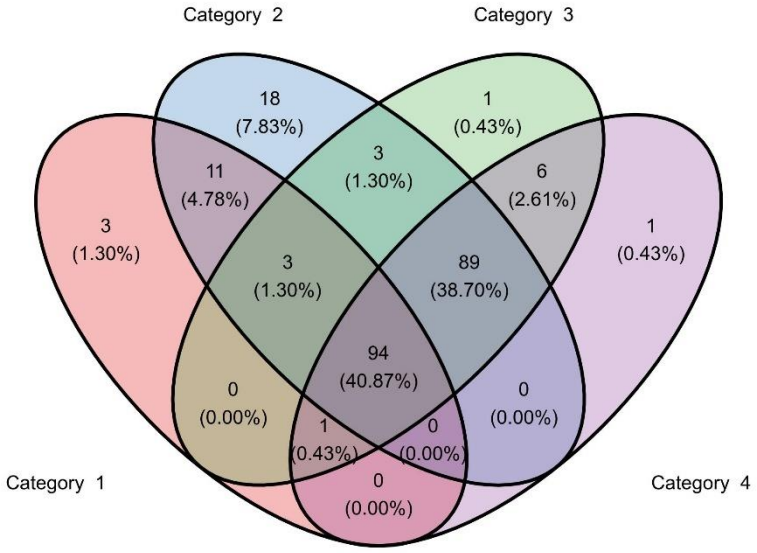

Figure 2. Number of Unique and Shared Genetic Variants with Gene Symbol Across Categories 1-4.

that do not change amino acids in the coded protein. The result shows 86% missense variants and 14% synonymous variants. Figure 1(b) shows further subclasses: 48% intron variants, 22% upstream and downstream variants, 16% non-coding transcript variants, 4% missense variants, 3% nonsense-mediated mRNA decay, and 2% regulatory region variants. In total, out of 206 variants have gene symbols and 151 variants without gene symbols. Hence, it is necessary to assign the gene symbol for the remaining variants and analyze the functional impacts due to the cumulative effect of associated SNPs.

Figure 2 provides the number of shared genes and unique genes across categories 1-4 (Category 5, where the linkage disequilibrium value is above 0.8, is not included due to no significance found under FDR>0.05). In Figure 2, 94 genes are commonly present in all four categories, such as PTPN2, IF1HI, MICB, CFB, BTNL2, BACH2, ANKRD55, CTLA4, IL2RA, CLEC16A, HSPA1B, IL6R, CD226, HLA-DQA1, HLA-DQA2, HLA-DRA, CD69, IL7R, and GSDMB, likely representing fundamental components in the biological pathways relevant to autoimmune diseases.[67] Around 89 genes are shared between categories 2, 3, and 4, as shown in Figure 2. Well-known genes such as CTLA4, IL2RA, and PTPN2 are involved in signalling pathways that regulate immune responses, which are critical in the autoimmune aspect of T1DM.[68,69] Similarly, the remaining genes in the study, such as ABR, ADAM30, ADCY3, CLDN10, CTRB1, CTSD, FADS2, IRF4, RAD51B, FLI1, FARP2, LPAR3, CAMK4, KLRG1, ITGB8, DUSP16, IRF2, EMSY, NCOA1, MAPT, HORMAD2, ADAMTS14, CD6, PRDX5, CCR3, JAZF1, have associations yet to be analyzed in relation to T1DM understanding. To analyze the cumulative impact of genetic variants and their associated genes in pathways, all the background genes were tested for significant enrichment of genes connected with T1DM.

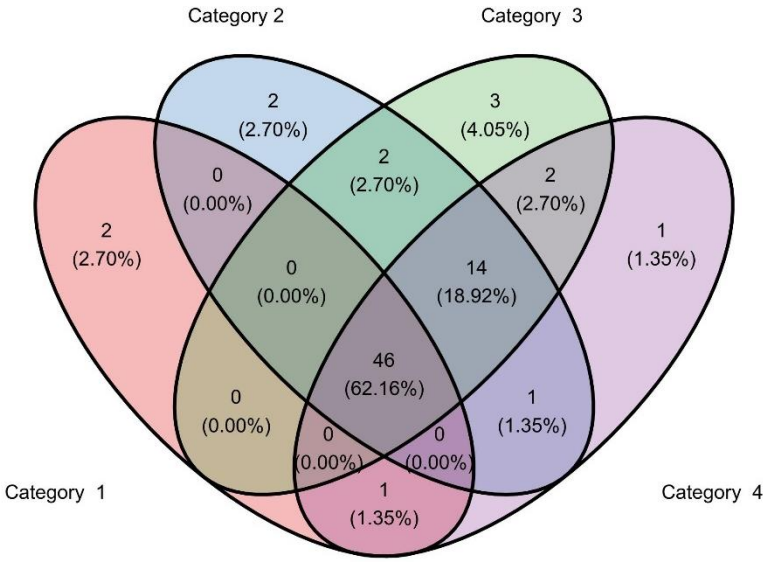

Figure 3. Number of Unique and Shared Gene Ontology Terms Across Categories 1-4.

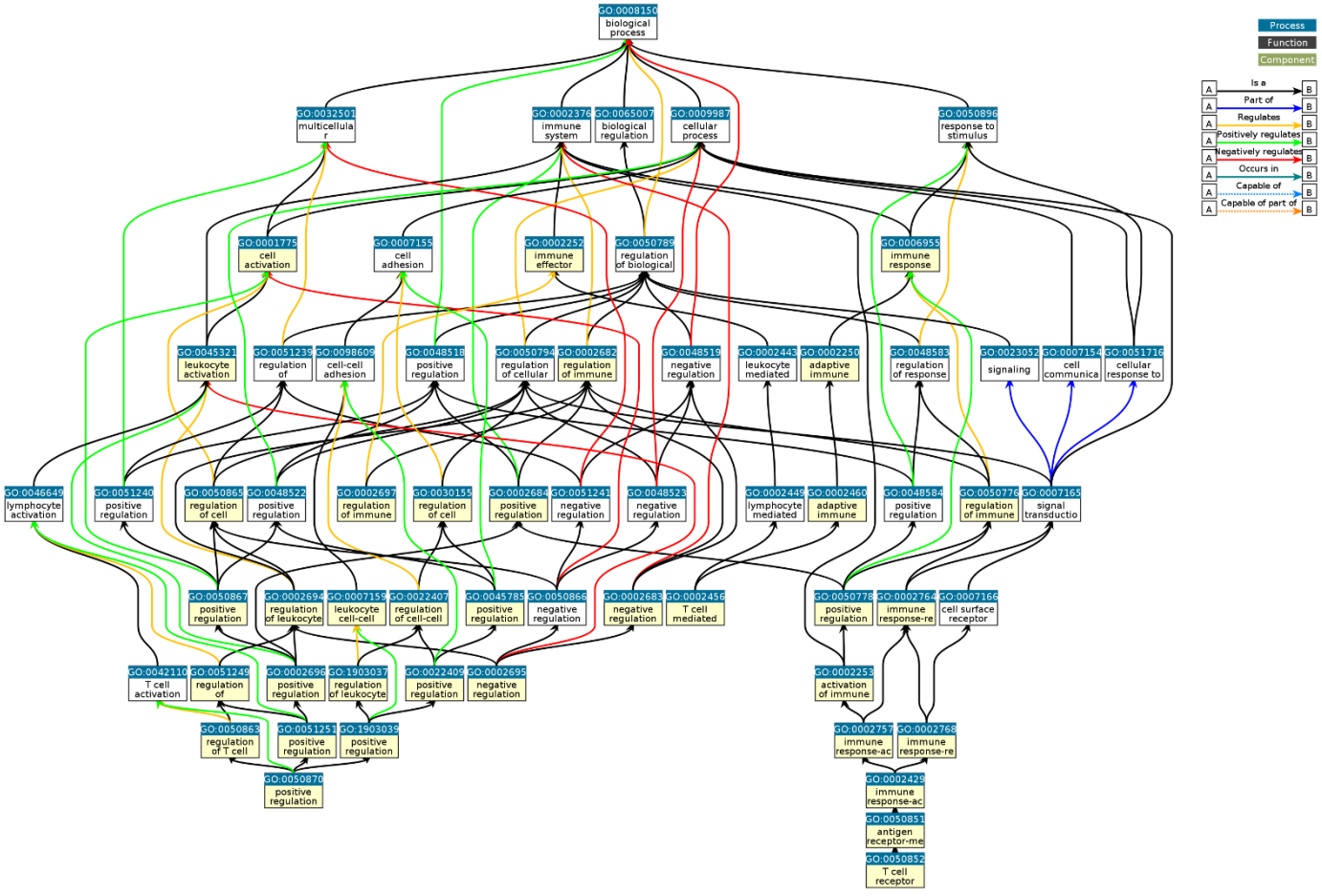

Figure 4. Inter-regulatory Network Connection among Category 2,3 and 4: Around 36 GO Biological Process Terms Shared among Categories 2, 3 and 4 in the Inter Regulatory Network.

GO Terms Enrichment Analysis

Gene ontology enriched biological processes (BP), cellular components (CC), molecular functions (MF), and KEGG Pathway analysis for all five categories are shown in

Supplementary Table ST1. The inter-regulatory network connection for biological processes of categories 1, 2, 3, and 4 is shown in Supplementary Figure SF1. Based on the supplementary information, the following GO term enrichment analysis was performed.

Figure 3 provides the gene ontology terms across categories 1-4. Category 5, where the linkage disequilibrium value is above 0.8, is not included due to no significance found under FDR>0.05. Based on Figure 3, the enrichment analysis provides 46 GO terms belonging to biological processes, two from cellular components, and one from molecular functions, providing a detailed view of the molecular mechanisms and genetic components involved. Overall, 17 GO terms are highly associated with T-cell receptor signalling pathways and reported in category one gene enrichment analysis.

Around 36 GO biological process terms are shared among categories 2, 3, and 4, and their inter-regulatory network connection is shown in Figure 4. The regulation of immune responses, including T-cell activation, lymphocyte activation, and leukocyte activation, is crucial for maintaining homeostasis and defending against pathogens. The GO terms highlight positive and negative regulatory mechanisms, indicating the complex balance the immune system must maintain. Gene Ontology-enriched biological process (BP), cellular component (CC), and molecular function (MF) analysis for categories 2, 3, and 4 are shown in Table 2.

Table 2. GO Enriched BP, CC and MF Analysis for the Categories 2, 3 and 4

Gene Set	Gene Ontology	Ratio			*p*-Value			FDR			Gene		
		2	3	4	2	3	4	2	3	4	2	3	4
GO:0034097	Response to cytokine	2.84	3.01	3.12	1.24 E-07	1.79 E-08	7.49 E-09	0.0015	2.18 E-04	9.09 E-05	31	32	32
GO:0071345	Cellular response to cytokine stimulus	2.88	3.06	3.17	2.52 E-07	3.79 E-08	1.67 E-08	0.003056	4.60 E-04	2.03 E-04	29	30	30
GO:0006952	Defense response	2.52	2.52	2.54	8.20 E-08	1.17 E-07	1.42 E-07	9.96E-04	0.001417	0.001727	38	37	36
GO:0009986	Cell surface	3.35	3.58	3.56	6.16 E-08	8.24 E-09	1.66 E-08	7.48E-04	1.00 E-04	2.01 E-04	26	27	26
GO:0002695	Negative regulation of leukocyte activation	6.97	7.17	6.75	5.46 E-07	4.12 E-07	2.44 E-06	0.006625	0.005008	0.02962	11	11	10
GO:0019221	Cytokine-mediated signaling pathway	3.57	3.67	3.80	3.29 E-08	1.86 E-08	9.10 E-09	3.99E-04	2.26 E-04	1.10 E-04	25	25	25
GO:2000026	Regulation of multicellular organismal development	#N/A	2.17	2.19	#N/A	1.66 E-06	1.74 E-06	#N/A	0.020174	0.021172	#N/A	40	39
GO:0002683	Negative regulation of immune system process	4.36	4.23	4.12	1.81 E-07	6.01 E-07	1.82 E-06	0.002196	0.007299	0.022073	18	17	16
GO:0001816	Cytokine production	#N/A	3.20	#N/A	#N/A	1.46 E-06	#N/A	#N/A	0.017766	#N/A	#N/A	22	#N/A

#N/A – Corresponding GO not available because of no significant under FDR<0.05.

Figures 5 to 9 depict the pathway enrichment analysis results using the dot plot, lollipop plot, bubble plot, upset plot, and network tree, respectively, highlighting the significance of each pathway and gene in T1DM. These visualizations can help identify patterns, such as the enrichment of specific pathways and the overlap of genes across different pathways, suggesting a complex network of interactions contributing to the disease's pathogenesis. The enriched size, input size, gene ratio, and other metrics offer insights into each pathway and gene's relative importance and contribution to T1DM. The dot plots help display the magnitude of gene enrichment across different pathways. Each dot represents a pathway, and its size can indicate the number of genes involved or the fold enrichment. To identify the most relevant pathways, the colour intensity represents the significance level (*p*-value). A high gene ratio implies that many genes significantly enrich a pathway.

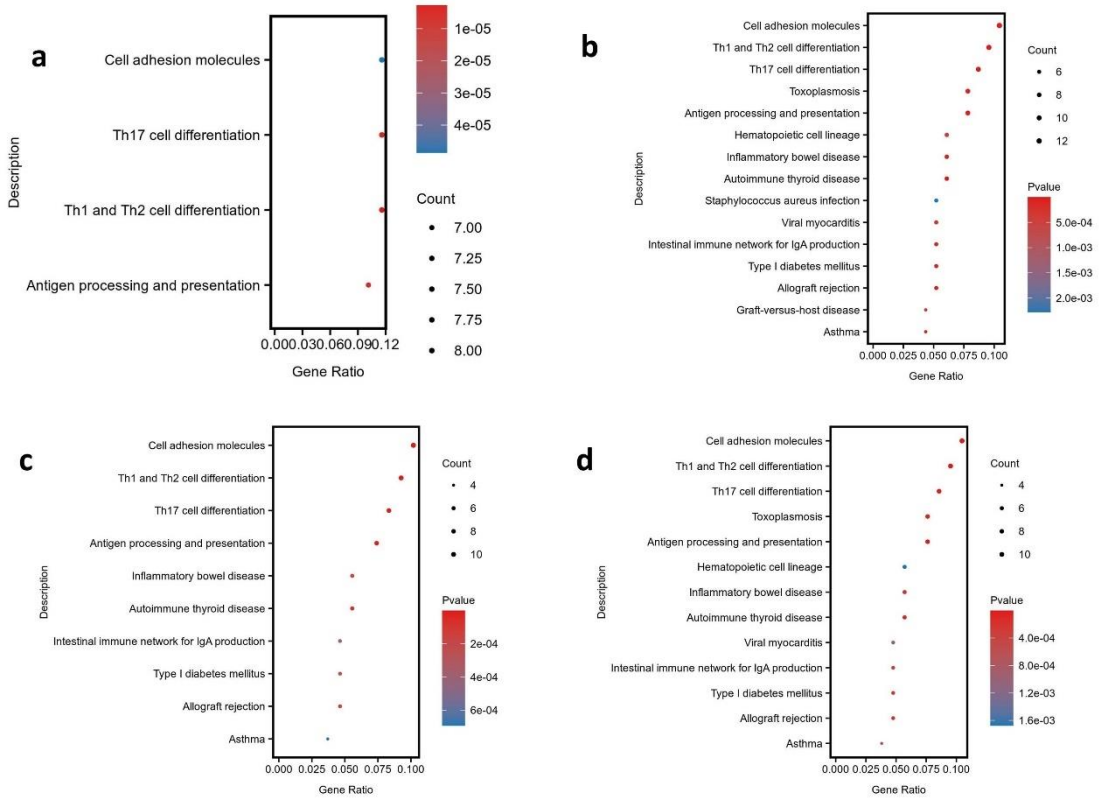

Figure 5. Dot Plots for **(a)** Category 1 **(b)** Category 2 **(c)** Category 3, and **(d)** Category 4: In Dot plots each dot represents a pathway, and its size can indicate the number of genes involved or the fold enrichment. the colour intensity represents the significance level (p-value) to identify the most relevant pathways.

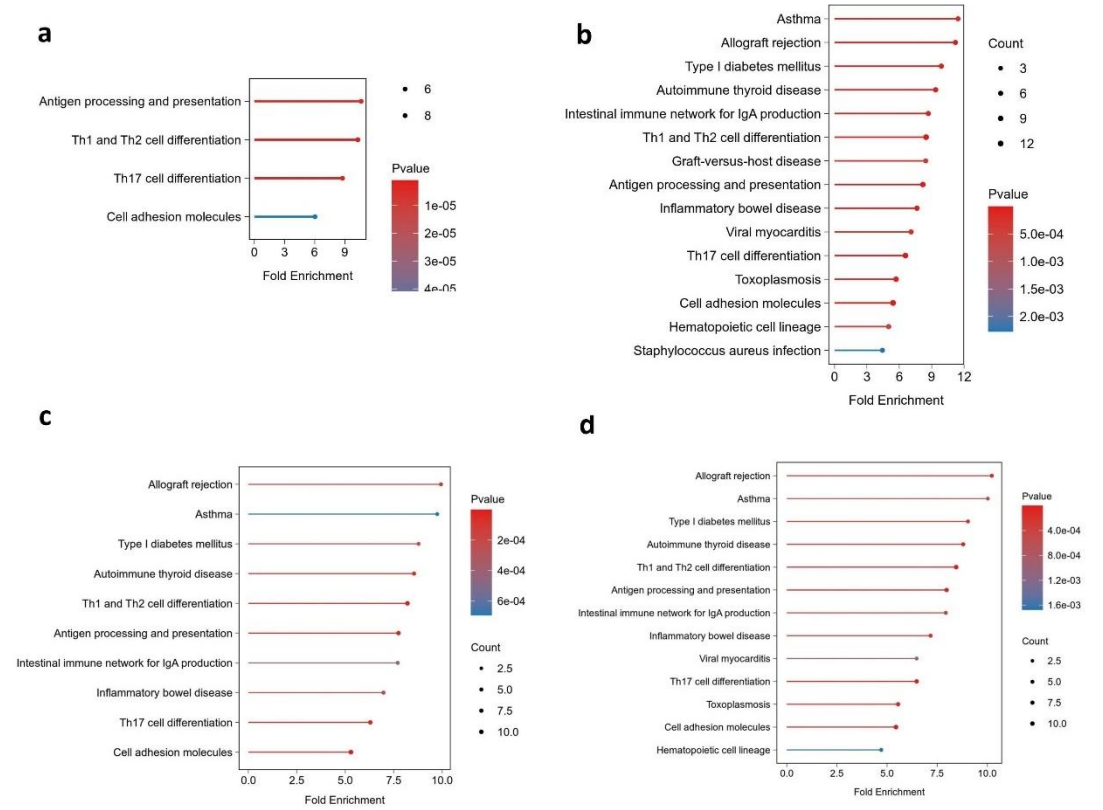

Figure 6. Lollipop plots for **(a)** Category 1 **(b)** Category 2 **(c)** Category 3, and **(d)** Category 4: Lollipop plots provide fold enrichment comparison across pathways using a "lollipop" marker at the end of a line. It shows each pathway's fold enrichment ratio, making comparing the relative importance or impact of different T1DM pathways.

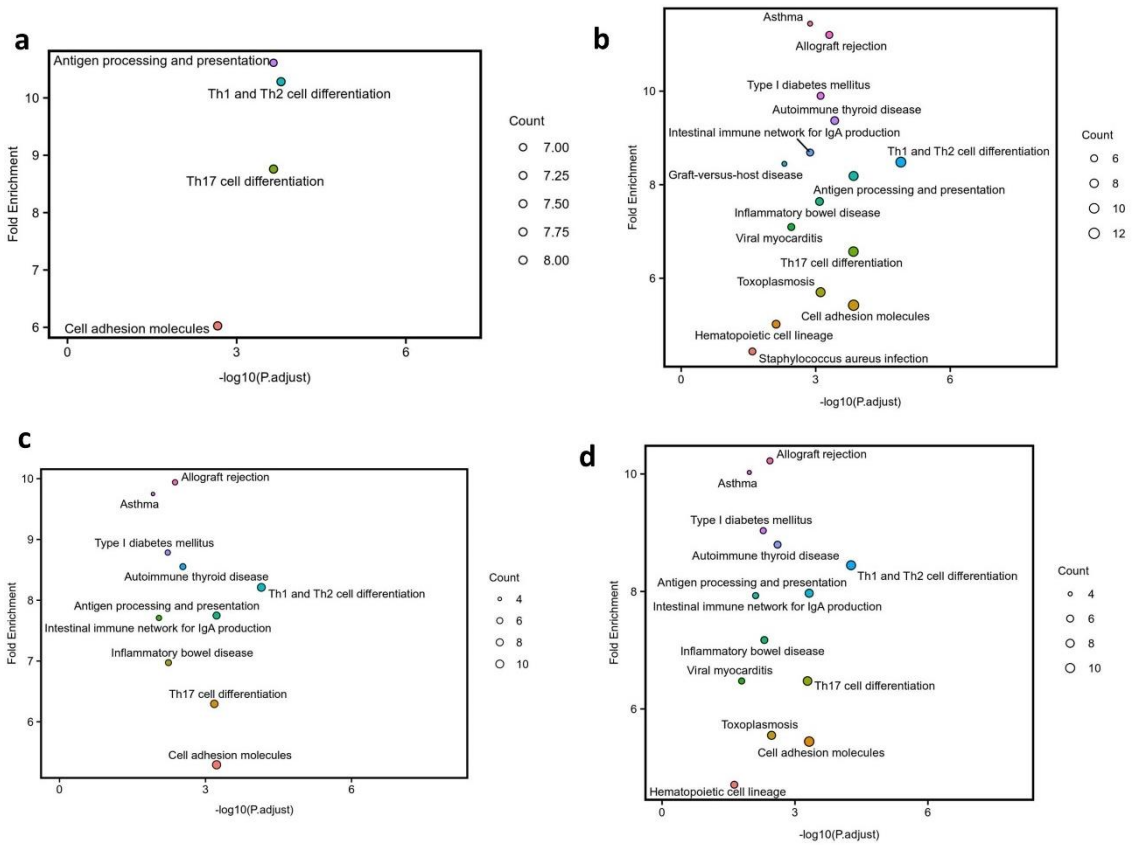

Figure 7. Bubble plots for **(a)** Category 1 **(b)** Category 2 **(c)** Category 3, and **(d)** Category 4: Bubble plot compares the enriched size, input size, gene ratio, and other variables across different pathways to show the complex interplay of various pathways.

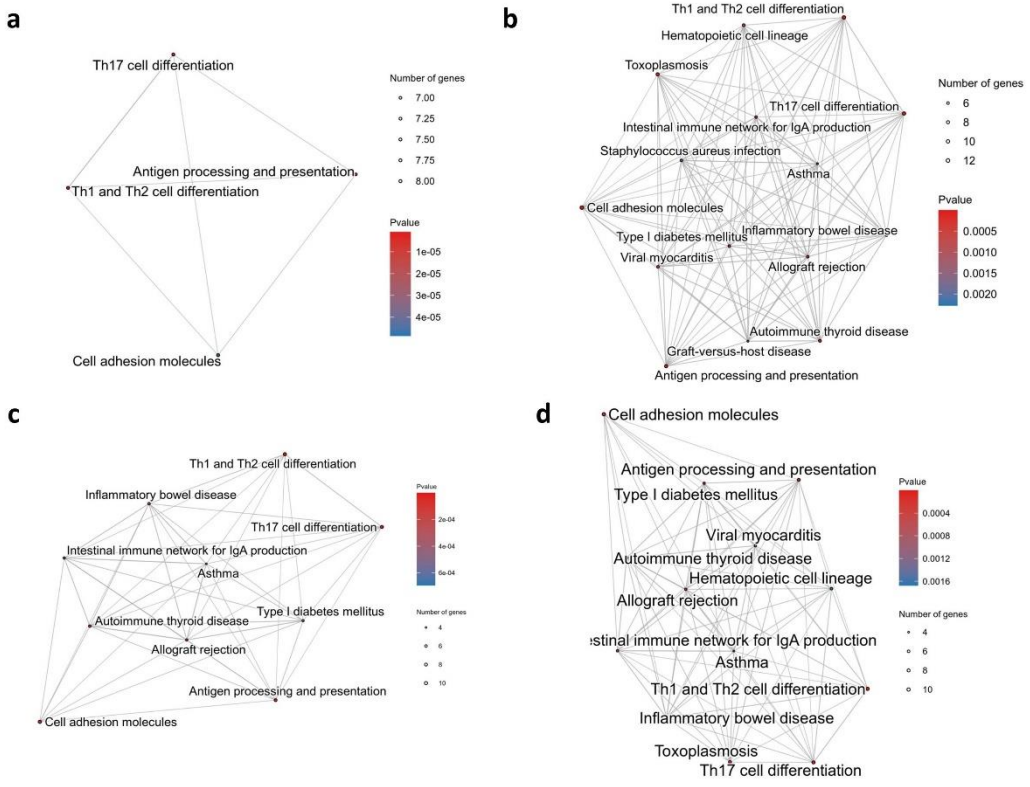

Figure 8. Network tree plots for **(a)** Category 1 **(b)** Category 2 **(c)** Category 3, and **(d)** Category 4: Network tree plots shows enriched terms serve as nodes, thickness of the edges indicates the extent of overlap between the gene sets and the overlapping gene sets are represented as edges, thicker edges suggesting a higher degree of shared genes.

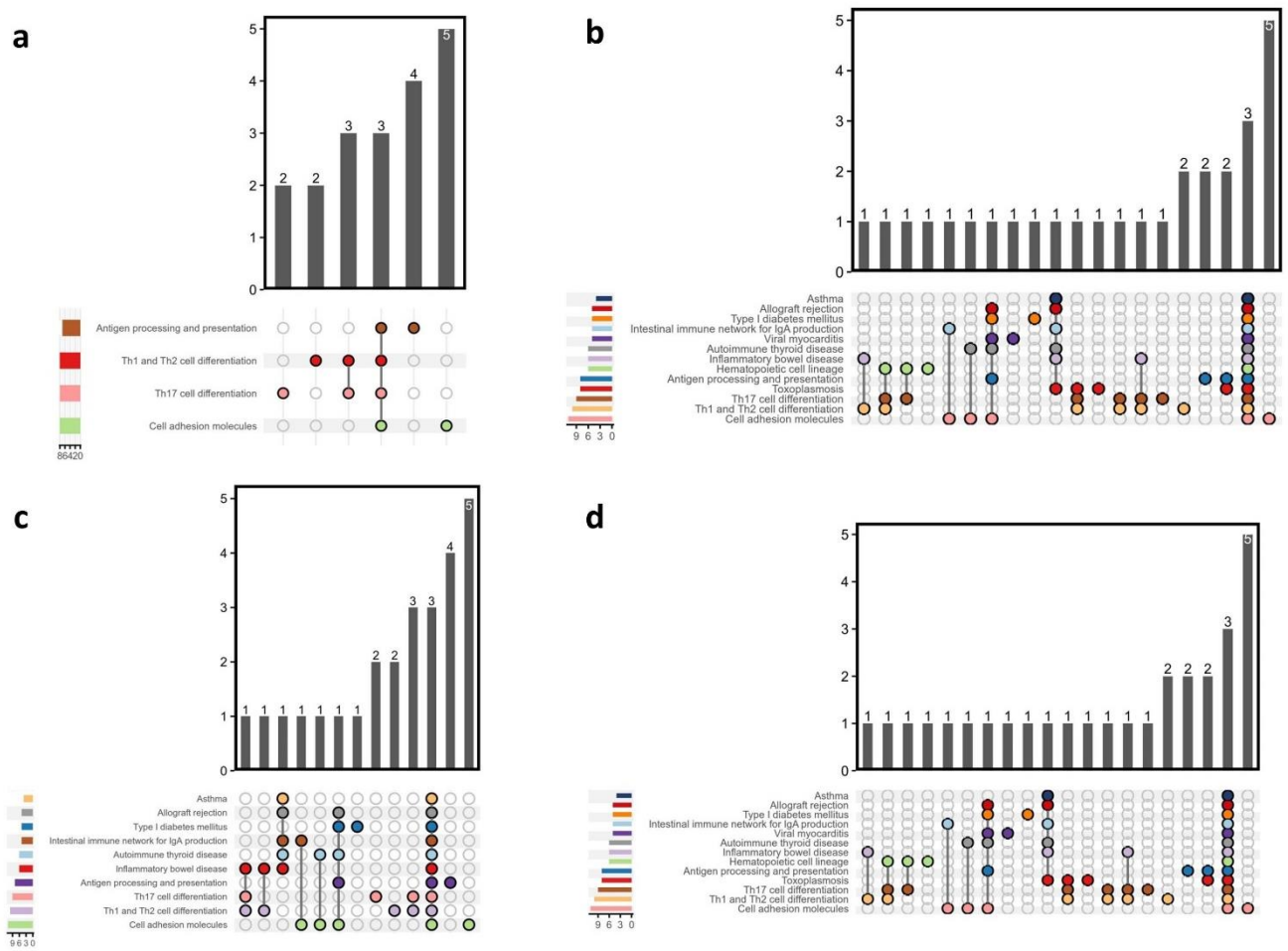

Figure 9. Upset plots for **(a)** Category 1 **(b)** Category 2 **(c)** Category 3, and **(d)** Category 4: Upset plots shows commonalities and unique elements across category. The intersection size along the Y-axis reflects the degree of overlap between different category, while the aggregate representation showcases the shared elements among these sets.

Discussion

A single gene can contain multiple variants, each with different LD values relative to other variants because LD is measured between pairs of genetic variants, not directly on genes themselves. Therefore, a gene might appear in multiple LD categories if its variants have varying LD relationships with other variants across the genome. Overall, 231 significant genes were retrieved from GWAS summary statistics and classified into four groups with multiple LD categories. Figure 2 shows 94 genes found across all four categories suggesting a core set of genes that might be crucial in the underlying mechanisms of the condition or related immune responses. The presence of common variants across different LD thresholds suggests they have strong and direct associations that are likely functionally essential. Further comprehensive analysis is required to confirm their potential involvement regardless of their linkage patterns with other variants in multiple biological processes or diseases, including T1DM. These genes might act as "hub genes" in genetic networks, interacting with many other genes and thus appearing in multiple LD-based categories.

Enrichment Analysis of GO Terms Commonly Found in the Category 1, 2, 3 and 4

As provided in the supplementary table ST1 Several pathways exhibit high enrichment ratios and extremely low *p*-values, indicating a strong and statistically significant association with the listed genetic variants. Majorly, the pathway for "adaptive immune response based on somatic recombination of immune receptors" (GO:0002460) shows the highest enrichment ratio (12.87477954) with a *p*-value of 1.52E-10, suggesting a potent involvement in T1DM pathogenesis. Cell Activation (GO:0001775) is represented by 34 variants, showing an enrichment ratio of 4.221 and a highly significant *p*-value of 1.69E-13. Similarly, Adaptive Immune Response (GO:0002250) involves 27 genes, with an enrichment ratio of 4.793 and a *p*-value of 5.52E-12. Immune Effector Process (GO:0002252) consists of 26 genes, with an enrichment ratio of 4.971 and a *p*-value of 6.60E-12. Finally, Activation of Immune Response (GO:0002253) exhibits the highest gene overlap with 36 genes, having an enrichment ratio of 3.261 and a *p*-value of 5.18E-11.

Various Gene Ontology (GO) terms related to immune response pathways provide a detailed view of their molecular mechanisms and genetic components, as shown in Supplementary Figure SF1. The immune response is the signalling pathways initiated by cross-linking antigen receptors on T cells (GO:0050852) and B or T cells (GO:0050851). These pathways involve 11 and 12 variants, respectively, with significant *p*-values indicating their critical roles in immune activation. These pathways are essential for initiating molecular signals that activate or perpetuate an immune response, highlighting the importance of antigen recognition in immune defence mechanisms. The immune response-activating cell surface receptor signalling pathway (GO:0002429) involves a broader set of 36 variants, demonstrating the complexity of signals initiated by extracellular ligands binding to cell surface receptors. Specifically, a few genetic variants are associated with multiple immune system pathways. rs2182419 appears in numerous pathways, including "cell activation" (GO:0001775), "adaptive immune response" (GO:0002250), and "T cell-mediated immunity" (GO:0002456). This overlap suggests that some genetic variants may have pleiotropic effects, influencing multiple aspects of the immune response relevant to T1DM. Pathways involved in the activation and regulation of the immune response, such as "activation of immune response" (GO:0002253) and "regulation of immune system process" (GO:0002682), are highlighted with significant associations to genetic variants.

Enrichment Analysis of GO Terms Commonly Found in the Category 2, 3 and 4

Table 2 shows the enrichment ratios and *p*-values for various Gene Ontology (GO) terms related to immune system processes. Positive Regulation of T Cell Activation (GO:0050870) has an enrichment ratio of 6.539 and a *p*-value of 3.30E-07, while Lymphocyte Activation (GO:0051251) has an enrichment ratio of 5.38 and a *p*-value of 9.65E-07. These processes enhance T cell activation, which is crucial for initiating immune responses against pathogens. Genes like CTLA4, PRKCQ, and GATA3 are involved in signaling pathways that promote T cell proliferation and differentiation. Negative Regulation of Leukocyte Activation (GO:0002695) has an enrichment ratio of 7.17 and a *p*-value of 4.12E-07, while Immune System Process (GO:0002683) has an enrichment ratio of 3.87 and a *p*-value of 4.12E-07. These processes inhibit leukocyte activation, including T cells and B cells, to prevent autoimmunity and resolve inflammation. Genes like CTLA4 and IL10 play inhibitory roles in immune responses, maintaining immune balance. Cell adhesion is essential for various biological processes, including tissue formation, wound healing, and immune responses. GO terms related to the regulation of cell adhesion (e.g., GO:0022409, GO:0045785, and GO:0030155) highlight the importance of cell-cell interactions in physiological and pathological conditions. Positive Regulation of Cell-Cell Adhesion (GO: 0022409) and Leukocyte Cell-Cell Adhesion (GO:0007159) enhance cell attachment, crucial for immune synapse formation between T cells and antigen-presenting cells.[70] Genes like CTLA4 and ICAM1 mediate these adhesion processes. Regulation of Cell Adhesion (GO:0030155) includes both positive and negative modulation of cell attachment. The adaptive immune response (GO:0002250) is a key component of the immune system, characterized by its ability to remember previous pathogen encounters and mount stronger responses upon re-exposure.

This process involves the activation and differentiation of T cells, which is significant for the adaptive immune response. Genes such as CTLA4, IL10, PRKCQ, and GATA3 play significant roles in modulating T cell activation. CTLA4 is known for its inhibitory role in T cell activation, serving as a checkpoint in immune regulation. Its involvement in T1DM has been extensively studied, with CTLA4 polymorphisms associated with an increased risk of developing the disease. Similarly, IL10, a cytokine with anti-inflammatory properties, plays a major role in immune response regulation and maintaining immune tolerance.[71] The negative regulation of immune responses (GO:0002683 and GO:0002695) is essential for preventing autoimmunity and maintaining immune homeostasis. Genes such as IL10, CTLA4, and PTPN2 are involved in these processes. PTPN2 has been identified as a novel gene associated with T1DM. It encodes a tyrosine phosphatase that negatively regulates inflammation and immune responses. Variants of PTPN2 have been linked to an increased risk of T1DM, highlighting its potential role in the pathogenesis of the disease.[72]

Pathway Enrichment Analysis

The figures 5a to 5d represent dot plots for categories 1-4, respectively. Cell Adhesion Molecules (hsa04514), Th1 and Th2 Cell Differentiation (hsa04658), and Th17 Cell Differentiation (hsa04659) show enrichment in all four category datasets. Antigen processing and presentation (hsa04612), Inflammatory bowel disease (hsa05321), and autoimmune thyroid disease (hsa05320) are common pathways among categories 2, 3, and 4.

The figures 6a to 6d represent lollipop plots for categories 1-4, respectively. Lollipop plots use a "lollipop" marker at the end of a line to represent the value of each category. In gene enrichment analysis, lollipop plots show each pathway's fold enrichment ratio, making it easier to compare the relative importance or impact of different T1DM pathways. The visualization of pathways through dot plots and lollipop plots helps in the interpretation of complex data and highlights the interconnectedness of these pathways. The size and color intensity of dots in a dot plot can quickly reveal which pathways are most significantly

enriched and potentially impactful in T1DM. In contrast, lollipop plots provide a fold enrichment comparison across pathways. Cell Adhesion Molecules (hsa04514), Th1 and Th2 Cell Differentiation (hsa04658), and Th17 Cell Differentiation (hsa04659) are consistently enriched in all four category datasets. Antigen processing and presentation (hsa04612), inflammatory bowel disease (hsa05321), and autoimmune thyroid disease (hsa05320) are shared among figures 6b, 6c, and 6d of categories 2, 3, and 4. However, significant fold enrichment is shown in Figure 6b of Category 2, which has a higher fold enrichment than the others.

The figures 7a to 7d represent bubble plots for categories 1-4, respectively. A bubble plot is a powerful tool for visualizing the complex interplay of various pathways involved in T1DM. By comparing the enriched size, input size, gene ratio, and other variables across different pathways, the bubble plot can show which pathways are most significantly associated with T1DM. Pathways with low *p*-values and high fold enrichment, such as Th1 and Th2 cell differentiation and antigen processing and presentation, will likely be crucial in the disease's pathogenesis.[40] Based on the comparison level across the four categories, Type I Diabetes Mellitus (hsa04940) has a higher fold enrichment level in figures 7b, 7c, and 7d of categories 2, 3, and 4 under a low *p*-value. The Type I Diabetes Mellitus (hsa04940) pathway is directly relevant to T1DM and includes HLA-B, HLA-DRA, HLA-DRB1, and INS. Including the insulin gene (INS) is particularly noteworthy, as autoantibodies against insulin are a hallmark of T1DM. The HLA genes are also crucial for the genetic predisposition to T1DM, with specific alleles strongly associated with an increased disease risk.[40]

The figures 8a to figure 8d represents network plots for category 1-4 respectively. Network tree plots provide detailed insights into the intricate correlations and interpretations related to the interactions between genes and pathways implicated in T1DM. These figures uniquely showcase the enriched network levels, particularly evident in Figure 8, highlighting the complex interconnections among various components. Within the network, enriched terms serve as nodes, while the overlapping gene sets are represented as edges. The thickness of the edges indicates the extent of overlap between the gene sets, with thicker edges suggesting a higher degree of shared genes. Furthermore, the Upset plots depicted in Figures 9 dissect the complexity of biological datasets, illustrating both commonalities and unique elements across multiple sets. The intersection size along the Y-axis reflects the degree of overlap between different sets, while the aggregate representation showcases the shared elements among these sets. Specifically, Figure 9a of category 1 highlights specific pathways such as cell adhesion molecules, Th1 and Th2 cell differentiation, Th17 cell differentiation, and antigen processing and presentation, each containing a varying number of genes within the category 1 dataset. Moreover, the visualization of shared genes across different categories further emphasizes the interconnectedness of pathways and genes relevant to T1DM.[73] The association of genes in the T1DM pathway with other GO-enriched pathways, as depicted in Figures 9b, 9c, and 9d of Categories 2-4 (e) plots, underscores the intricate relationships between different biological processes. Notably, the balance between Th1 and Th2 responses plays a crucial role in immune regulation, with an imbalance potentially leading to autoimmune diseases like T1DM. Genes such as GATA3, HLA-DOB, HLA-DQA1, HLA-DRA, HLA-DRB1, IL2RA, NOTCH2, PRKCQ, RUNX3, STAT4, and TYK2 are significant in immune cell differentiation and function, influencing T1DM susceptibility and progression.[74] Overall, these findings suggest a complex interplay between genetic predisposition and immune regulation in the development and progression of T1DM.

Variations in these genes could influence the immune response balance, potentially affecting T1DM susceptibility and progression. Their involvement suggests a complex interplay between genetic predisposition and immune regulation in T1DM.[75] The autoimmune thyroid disease involving Cell Adhesion Molecules (hsa04514) plays a crucial role in the interaction between T cells and antigen-presenting cells, which ultimately leads to the autoimmune response against pancreatic beta cells. This destructive process is facilitated by various genes such as PTPRC, CTLA4, CD6, HLA-DQA1, HLA-DOB, HLA-DRA, ITGB8, CLDN10, and CD226. The significance of these genes in T1DM underscores the essential role of cell-cell interactions in the destruction of pancreatic beta cells.[76] In particular, CTLA4 stands out as a regulatory molecule capable of modulating T cell activation and tolerance in T1DM.[55] The importance of this pathway in T1DM pathogenesis is highlighted by the enriched size and input size depicted in the bubble plot. Another critical pathway in autoimmune diseases, especially in T1DM, is the Th17 Cell Differentiation pathway (hsa04659). Th17 cells, a subset of pro-inflammatory T helper cells, are involved in immune responses and autoimmune diseases. Genes like IL6R, IRF4, HLA-DQA1, HLA-DOB, HLA-DRA, GATA3, IL2RA, and TYK2 demonstrate the importance of Th17 cells in the inflammatory environment that contributes to beta-cell destruction in T1DM.[47] The autoimmune thyroid disease (hsa05320) pathway involves genes such as CTLA4, HLA-B, HLA-DRA, HLA-DRB1, and IL10, which are implicated in autoimmune thyroid disease but also play significant roles in T1DM with significant enrichment in categories 2, 3, and 4. Enriching these genes suggests a shared genetic basis between autoimmune thyroid disease and T1DM,

emphasising the immune regulation and the presentation of autoantigens. CTLA4, for instance, is a critical immune checkpoint that modulates T cell activation, while HLA genes are essential for antigen presentation, a process dysregulated in autoimmune diseases.[77] The Allograft Rejection (hsa05330) pathway, which is linked to the autoimmune thyroid disease pathway, highlights the crucial role of immune recognition and response in the development of T1DM. The presence of HLA genes and IL10 in this pathway serves to emphasize the delicate balance between immune activation and regulation that is essential for maintaining tolerance to self-antigens. Any disruption in this balance can trigger autoimmune reactions against the pancreatic beta cells, leading to the onset of T1DM.[78] The Antigen Processing and Presentation (hsa04612) pathway, which includes genes such as CTSB, HLA-B, HLA-DRA, HLA-DRB1, and HSPA1B, plays a central role in the pathogenesis of T1DM. Effective presentation of beta-cell autoantigens by HLA molecules to T cells is crucial for initiating the autoimmune attack in T1DM.[13,47] Variations in these genes could potentially impact the risk of developing T1DM by influencing the immune system's ability to differentiate between self and non-self.[79] The Cell Adhesion Molecules (hsa04514) pathway, which involves CD6, CTLA4, HLA-B, HLA-DRA, HLA-DRB1, and ITGB8, underscores the significance of cell-cell interactions in the immune response and T1DM.[80] These molecules play a crucial role in facilitating the interaction between T cells and antigen-presenting cells, which is a critical step in the autoimmune response against pancreatic beta cells. Disruption in these pathways may impact the gastrointestinal environment and potentially influence autoimmunity and the development of T1DM. The Cell Adhesion Molecules (hsa04514) pathway, involving CD6, CTLA4, HLA-B, HLA-DRA, HLA-DRB1, and ITGB8, underscores the importance of cell-cell interactions in the immune response and T1DM. Toxoplasmosis (hsa05145), Asthma (hsa05310), Graft-versus-host Disease (hsa05332), Intestinal Immune Network for IgA Production (hsa04672), and Viral Myocarditis (hsa05416) pathways illustrate the broad impact of immune regulation and response on various diseases and conditions.[46] The involvement of HLA genes and IL10 across these pathways emphasises the critical role of these genes in immune system modulation, which is significant to understanding T1DM pathogenesis [40]

Conclusion

With over 70 genomic risk loci regions with more than a hundred significant genetic variants, it is crucial to determine the underlying genetic mechanism for comprehending the pathophysiology of T1DM. Multiple T1DM GWAS summary statistics data from GWAS Central were analyzed to retrieve significant genetic variants associated with T1DM. Previous GWAS results established that T1DM is a complex trait where the genetic variants at the HLA genomic region contribute highly. In contrast, genetic variants from non-HLA regions regulate the regulatory pathways with minor effects.

Gene set enrichment analysis identifies significantly enriched gene ontologies associated with T1DM. The significantly enriched GO terms spanned biological processes, cellular components, and molecular functions, primarily involved in immune response, activation/regulation, and T-cell development. The above GO results further strengthen the well-established crucial role of immune response in T1DM development. Significant genetic variants related to pathways such as T cell differentiation and its regulation, cytokine-mediated signalling pathways, and negative regulation of immune system processes are less-investigated immune mechanisms. The processes need further exploration to identify novel pathogenesis mechanisms of T1DM. This study provides a foundation for future studies in understanding the complex pathogenesis and identification of genetic signatures to reveal the disease prognosis and detection at the genetic level.

Authors' Contributions

Study conceptualization, design, material preparation, data collection and initial analysis were performed by NVS. The first draft of the manuscript was written by NVS and NVG. Study conceptualization, design, detailed analysis, supervision and correspondence was done by NVG. All authors read and approved the final manuscript.

Conflict of Interest Disclosures

The authors declare that they have no conflicts of interest.

Availability of Data and Materials

This study utilized a range of T1DM GWAS summary statistics obtained from the GWAS Central and GWAS Catalog databases. The sources are mentioned in detail in the materials and methods section.

References

1. Todd JA. Etiology of type 1 diabetes. Immunity. 2010;32(4):457-67. doi:10.1016/j.immuni.2010.04.001
2. Popoviciu MS, Kaka N, Sethi Y, Patel N, Chopra H, Cavalu S. Type 1 diabetes mellitus and autoimmune diseases: a critical review of the association and the application of personalized medicine. J Pers Med. 2023;13(3):422. doi:10.3390/jpm13030422
3. Redondo MJ, Steck AK, Pugliese A. Genetics of type 1 diabetes. Pediatr Diabetes. 2018;19(3):346-53. doi:10.1111/pedi.12597
4. Cade WT. Diabetes-related microvascular and macrovascular diseases in the physical therapy setting. Phys Ther. 2008;88(11):1322-35. doi:10.2522/ptj.20080008
5. Ogle GD, James S, Dabelea D, Pihoker C, Svennson J,

Maniam J, et al. Global estimates of incidence of type 1 diabetes in children and adolescents: Results from the International Diabetes Federation Atlas. Diabetes Res Clin Pract. 2022;183:109083. doi:10.1016/j.diabres.2021.109083
6. Gregory GA, Robinson TI, Linklater SE, Wang F, Colagiuri S, de Beaufort C, et al. Global incidence, prevalence, and mortality of type 1 diabetes in 2021 with projection to 2040: a modelling study. Lancet Diabetes Endocrinol. 2022;10(10):741-60. doi:10.1016/S2213-8587(22)00218-2
7. Vanderniet JA, Jenkins AJ, Donaghue KC. Epidemiology of type 1 diabetes. Curr Cardiol Rep. 2022;24(10):1455-65. doi:10.1007/s11886-022-01762-w
8. Michalek DA, Onengut-Gumuscu S, Repaske DR, Rich SS. Precision medicine in type 1 diabetes. J Indian Inst Sci. 2023;103(1):335-51. doi:10.1007/s41745-023-00356-x
9. Hirschhorn JN. Genetic epidemiology of type 1 diabetes. Pediatr Diabetes. 2003;4(2):87-100. doi:10.1034/j.1399-5448.2001.00013.x
10. Torres JM, Cox NJ, Philipson LH. Genome wide association studies for diabetes: perspective on results and challenges. Pediatr Diabetes. 2013;14(2):90-6. doi:10.1111/pedi.12015
11. Xia Y, Xie Z, Huang G, Zhou Z. Incidence and trend of type 1 diabetes and the underlying environmental determinants. Diabetes Metab Res Rev. 2019;35(1):e3075. doi:10.1002/dmrr.3075
12. Schaid DJ, Chen W, Larson NB. From genome-wide associations to candidate causal variants by statistical fine-mapping. Nat Rev Genet. 2018;19(8):491-504. doi:10.1038/s41576-018-0016-z
13. Roizen JD, Bradfield JP, Hakonarson H. Progress in understanding type 1 diabetes through its genetic overlap with other autoimmune diseases. Curr Diab Rep. 2015;15:102. doi:10.1007/s11892-015-0668-4
14. Polychronakos C, Li Q. Understanding type 1 diabetes through genetics: advances and prospects. Nat Rev Genet. 2011;12(11):781-92. doi:10.1038/NRG3069
15. Cano-Gamez E, Trynka G. From GWAS to function: using functional genomics to identify the mechanisms underlying complex diseases. Front Genet. 2020;11:424. doi:10.3389/fgene.2020.00424
16. Chouakria AD, Nagabhushan PN. Adaptive dissimilarity index for measuring time series proximity. Adv Data Anal Classif. 2007;1:5-21. doi:10.1007/s11634-006-0004-6
17. Parkes M, Cortes A, Van Heel DA, Brown MA. Genetic insights into common pathways and complex relationships among immune-mediated diseases. Nat Rev Genet. 2013;14(9):661-73. doi:10.1038/nrg3502
18. Ellinghaus D, Jostins L, Spain SL, Cortes A, Bethune J, Han B, et al. Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. Nat Genet. 2016;48(5):510-8. doi:10.1038/ng.3528
19. Gutierrez-Sacristan A, Bravo A, Giannoula A, Mayer MA, Sanz F, Furlong LI. comoRbidity: an R package for the systematic analysis of disease comorbidities. Bioinformatics. 2018;34(18):3228-30. doi:10.1093/bioinformatics/bty315
20. Sánchez-Valle J, Tejero H, Fernández JM, Juan D, Capella-Gutiérrez S, Al-Shahrour F, et al. Unveiling the molecular basis of disease co-occurrence: towards personalized comorbidity profiles. bioRxiv. 2018:431312. doi:10.1101/431312
21. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nat Rev Genet. 2010;11(12):843-54. doi:10.1038/nrg2884
22. Trynka G, Hunt KA, Bockett NA, Romanos J, Mistry V, Szperl A, et al. Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. Nat Genet. 2011;43(12):1193-201. doi:10.1038/ng.998
23. Fortune MD, Guo H, Burren O, Schofield E, Walker NM, Ban M, et al. Statistical colocalization of genetic risk variants for related autoimmune diseases in the context of common controls. Nat Genet. 2015 Jul;47(7):839-46. doi:10.1038/ng.3330
24. Robertson CC, Inshaw JR, Onengut-Gumuscu S, Chen WM, Santa Cruz DF, Yang H, et al. Fine-mapping, trans-ancestral and genomic analyses identify causal variants, cells, genes and drug targets for type 1 diabetes. Nat Genet. 2021;53(7):962-71. doi:10.1038/s41588-021-00880-5
25. Shi H, Kichaev G, Pasaniuc B. Contrasting the genetic architecture of 30 complex traits from summary association data. Am J Hum Genet. 2016 ;99(1):139-53. doi:10.1016/j.ajhg.2016.05.013
26. Yoon S, Nguyen HC, Yoo YJ, Kim J, Baik B, Kim S, et al. Efficient pathway enrichment and network analysis of GWAS summary data using GSA-SNP2. Nucleic Acids Res. 2018;46(10):e60. doi:10.1093/nar/gky175
27. Welter D, MacArthur J, Morales J, Burdett T, Hall P, Junkins H, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014;42(D1):D1001-6. doi:10.1093/nar/gkt1229
28. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. Science. 2012;337(6099):1190-5. doi:10.1126/science.1222794
29. Balliu B, Carcamo-Orive I, Gloudemans MJ, Nachun DC, Durrant MG, Gazal S, et al. An integrated approach to identify environmental modulators of genetic risk factors for complex traits. Am J Hum Genet. 2021;108(10):1866-79. doi:10.1016/j.ajhg.2021.08.014
30. Zhong W, Liu W, Chen J, Sun Q, Hu M, Li Y. Understanding the function of regulatory DNA interactions in the interpretation of non-coding GWAS variants. Front Cell Dev Biol. 2022;10:957292. doi:10.3389/fcell.2022.957292
31. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, et al. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. 2008;4(10):e1000214. doi:10.1371/journal.pgen.1000214
32. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. PLoS Genet. 2010;6(4):e1000888. doi:10.1371/journal.pgen.1000888
33. Lee HS, Hwang JS. Genetic aspects of type 1 diabetes. Ann Pediatr Endocrinol Metab. 2019;24(3):143-8. doi:10.6065/apem.2019.24.3.143
34. Boyle EA, Li YI, Pritchard JK. An expanded view of complex traits: from polygenic to omnigenic. Cell. 2017;169(7):1177-86. doi:10.1016/j.cell.2017.05.038
35. Goldstein DB. Common genetic variation and human traits. N Engl J Med. 2009;360(17):1696-8. doi:10.1056/NEJMp0806284
36. Chia R, Saez-Atienzar S, Murphy N, Chiт A, Blauwendraat C, International Myasthenia Gravis Genomics Consortium, et al. Identification of genetic risk loci and prioritization of genes and pathways for myasthenia gravis: a genome-wide association study. Proc Natl Acad Sci U S A. 2022;119(5):e2108672119. doi:10.1073/pnas.2108672119
37. Mooney MA, Wilmot B. Gene set analysis: a step-by-step

guide. Am J Med Genet B Neuropsychiatr Genet. 2015;168(7):517-27. doi:10.1002/ajmg.b.32328
38. Pujar M, Vastrad B, Kavatagimath S, Vastrad C, Kottursheti S. Identification of candidate biomarkers and pathways associated with type 1 diabetes mellitus using bioinformatics analysis. Sci Rep. 2022;12(1):9157. doi:10.1038/s41598-022-13291-1
39. García-Campos MA, Espinal-Enríquez J, Hernбndez-Lemus E. Pathway analysis: state of the art. Front Physiol. 2015;6:383. doi:10.3389/fphys.2015.00383
40. Chu X, Janssen AW, Koenen H, Chang L, He X, Joosten I, et al. A genome-wide functional genomics approach uncovers genetic determinants of immune phenotypes in type 1 diabetes. Elife. 2022;11:e73709. doi:10.7554/ELIFE.73709
41. Khurana E, Fu Y, Chakravarty D, Demichelis F, Rubin MA, Gerstein M. Role of non-coding sequence variants in cancer. Nat Rev Genet. 2016;17(2):93-108. doi:10.1038/nrg.2015.17
42. French JD, Edwards SL. The role of noncoding variants in heritable disease. Trends Genet. 2020;36(11):880-91. doi:10.1016/J.TIG.2020.07.004
43. Alsheikh AJ, Wollenhaupt S, King EA, Reeb J, Ghosh S, Stolzenburg LR, et al. The landscape of GWAS validation; systematic review identifying 309 validated non-coding variants across 130 human diseases. BMC Med Genomics. 2022;15(1):74. doi:10.1186/s12920-022-01216-w
44. Htet M, Lei S, Bajpayi S, Zoitou A, Chamakioti M, Tampakakis E. The role of noncoding genetic variants in cardiomyopathy. Front Cardiovasc Med. 2023;10:1116925. doi:10.3389/fcvm.2023.1116925
45. Manolio TA. Bringing genome-wide association findings into clinical use. Nat Rev Genet. 2013;14(8):549-58. doi:10.1038/nrg3523
46. Johnson MB, Cerosaletti K, Flanagan SE, Buckner JH. Genetic mechanisms highlight shared pathways for the pathogenesis of polygenic type 1 diabetes and monogenic autoimmune diabetes. Curr Diab Rep. 2019;19:20. doi:10.1007/S11892-019-1141-6
47. Ramos-Rodríguez M, Pérez-González B, Pasquali L. The β-cell genomic landscape in T1D: implications for disease pathogenesis. Curr Diab Rep. 2021;21:1. doi:10.1007/S11892-020-01370-4
48. Svishcheva GR, Belonogova NM, Zorkoltseva IV, Kirichenko AV, Axenovich TI. Gene-based association tests using GWAS summary statistics. Bioinformatics. 2019;35(19):3701-8. doi:10.1093/bioinformatics/btz172
49. Lypez de Maturana E, Rodrнguez JA, Alonso L, Lao O, Molina-Montes E, Martнn-Antoniano IA, et al. A multilayered post-GWAS assessment on genetic susceptibility to pancreatic cancer. Genome medicine. 2021;13:15. doi:10.1186/s13073-020-00816-4
50. Thorisson GA, Lancaster O, Free RC, Hastings RK, Sarmah P, Dash D, et al. HGVbaseG2P: a central genetic association database. Nucleic Acids Res. 2009;37(suppl_1):D797-802. doi:10.1093/nar/gkn748
51. Beck T, Rowlands T, Shorter T, Brookes AJ. GWAS Central: an expanding resource for finding and visualising genotype and phenotype data from genome-wide association studies. Nucleic Acids Res. 2023;51(D1):D986-93. doi:10.1093/nar/gkac1017
52. Beck T, Shorter T, Brookes AJ. GWAS Central: a comprehensive resource for the discovery and comparison of genotype and phenotype data from genome-wide association studies. Nucleic Acids Res. 2020;48(D1):D933-40. doi:10.1093/nar/gkz895
53. Beck T, Hastings RK, Gollapudi S, Free RC, Brookes AJ. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. Eur J Hum Genet. 2014;22(7):949-52. doi:10.1038/ejhg.2013.274
54. Ray D, Boehnke M. Methods for meta-analysis of multiple traits using GWAS summary statistics. Genet Epidemiol. 2018;42(2):134-45. doi:10.1002/gepi.22105
55. Laine AP, Valta M, Toppari J, Knip M, Veijola R, Ilonen J, et al. Non-HLA gene polymorphisms in the pathogenesis of type 1 diabetes: phase and endotype specific effects. Front Immunol. 2022;13:909020. doi:10.3389/fimmu.2022.909020
56. McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The ensembl variant effect predictor. Genome Biol. 2016;17:122. doi:10.1186/s13059-016-0974-4
57. Reimand J, Arak T, Adler P, Kolberg L, Reisberg S, Peterson H, et al. g: Profiler—a web server for functional interpretation of gene lists (2016 update). Nucleic Acids Res. 2016;44(W1):W83-9. doi:10.1093/nar/gkw199
58. Huang D, Zhou Y, Yi X, Fan X, Wang J, Yao H, et al. VannoPortal: multiscale functional annotation of human genetic variants for interrogating molecular mechanism of traits and diseases. Nucleic Acids Res. 2022;50(D1):D1408-16. doi:10.1093/nar/gkab853
59. Huang D, Yi X, Zhou Y, Yao H, Xu H, Wang J, et al. Ultrafast and scalable variant annotation and prioritization with big functional genomics data. Genome Res. 2020;30(12):1789-801. doi:10.1101/gr.267997.120
60. Lu Q, Powles RL, Abdallah S, Ou D, Wang Q, Hu Y, et al. Systematic tissue-specific functional annotation of the human genome highlights immune-related DNA elements for late-onset Alzheimer's disease. PLoS Genet. 2017;13(7):e1006933. doi:10.1371/journal.pgen.1006933
61. Backenroth D, He Z, Kiryluk K, Boeva V, Pethukova L, Khurana E, et al. FUN-LDA: A latent Dirichlet allocation model for predicting tissue-specific functional effects of noncoding variation. bioRxiv. 2016:069229. doi:10.1101/069229
62. Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Fet al. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet. 2015;47(4):381-6. doi:10.1038/ng.3245
63. Cirillo E, Parnell LD, Evelo CT. A review of pathway-based analysis tools that visualize genetic variants. Front Genet. 2017;8:174. doi:10.3389/fgene.2017.00174
64. Pomyen Y, Segura M, Ebbels TM, Keun HC. Over-representation of correlation analysis (ORCA): a method for identifying associations between variable sets. Bioinformatics. 2015;31(1):102-8. doi:10.1093/bioinformatics/btu589
65. Karp PD, Midford PE, Caspi R, Khodursky A. Pathway size matters: the influence of pathway granularity on over-representation (enrichment analysis) statistics. BMC Genomics. 2021;22:191. doi:10.1186/s12864-021-07502-8
66. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. Nucleic Acids Res. 2017;45(W1):W130-7. doi:10.1093/nar/gkx356
67. Zajec A, Trebušak Podkrajšek K, Tesovnik T, Šket R, Čugalj Kern B, Jenko Bizjan B, et al. Pathogenesis of type 1 diabetes: established facts and new insights. Genes. 2022;13(4):706. doi:10.3390/genes13040706
68. Steck AK, Rewers MJ. Genetics of type 1 diabetes. Clin Chem. 2011;57(2):176-85. doi:10.1373/clinchem.2010.148221

69. Morran MP, Vonberg A, Khadra A, Pietropaolo M. Immunogenetics of type 1 diabetes mellitus. Mol Aspects Med. 2015;42:42-60. doi:10.1016/j.mam.2014.12.004
70. Mallone R, Eizirik DL. Presumption of innocence for beta cells: why are they vulnerable autoimmune targets in type 1 diabetes?. Diabetologia. 2020;63:1999-2006. doi:10.1007/S00125-020-05176-7/FIGURES/2
71. Ramos-Rodríguez M, Raurell-Vila H, Colli ML, Alvelos MI, Subirana-Granés M, Juan-Mateu J, et al. The impact of proinflammatory cytokines on the β-cell regulatory landscape provides insights into the genetics of type 1 diabetes. Nat Genet. 2019;51(11):1588-95. doi:10.1038/S41588-019-0524-6
72. Gootjes C, Zwaginga JJ, Roep BO, Nikolic T. Functional impact of risk gene variants on the autoimmune responses in type 1 diabetes. Front Immunol. 2022;13:886736. doi:10.3389/fimmu.2022.886736
73. Johnson MB, Cerosaletti K, Flanagan SE, Buckner JH. Genetic mechanisms highlight shared pathways for the pathogenesis of polygenic type 1 diabetes and monogenic autoimmune diabetes. Curr Diab Rep. 2019;19:20. doi:10.1007/s11892-019-1141-6
74. Cheng M, Mei B, Zhou Q, Zhang M, Huang H, Han L, et al. Computational analyses of obesity associated loci generated by genome-wide association studies. PloS One. 2018;13(7):e0199987. doi:10.1371/journal.pone.0199987
75. Bhushan R, Rani A, Ali A, Singh VK, Dubey PK. Bioinformatics enrichment analysis of genes and pathways related to maternal type 1 diabetes associated with adverse fetal outcomes. J Diabetes Complications. 2020;34(5):107556. doi:10.1016/j.jdiacomp.2020.107556
76. Kaur S, Mirza AH, Overgaard AJ, Pociot F, Stwrling J. A dual systems genetics approach identifies common genes, networks, and pathways for type 1 and 2 diabetes in human islets. Front Genet. 2021;12:630109. doi:10.3389/fgene.2021.630109
77. Sharma A, Liu X, Hadley D, Hagopian W, Chen WM, Onengut-Gumuscu S, et al. Identification of non-HLA genes associated with development of islet autoimmunity and type 1 diabetes in the prospective TEDDY cohort. J Autoimmun. 2018;89:90-100. doi:10.1016/j.jaut.2017.12.008
78. Márquez A, Martín J. Genetic overlap between type 1 diabetes and other autoimmune diseases. Semin Immunopathol. 2022;44(1):81-97. doi:10.1007/s00281-021-00885-6
79. Nyaga DM, Vickers MH, Jefferies C, Perry JK, O'Sullivan JM. Type 1 diabetes mellitus-associated genetic variants contribute to overlapping immune regulatory networks. Front Genet. 2018;9:535. doi:10.3389/fgene.2018.00535
80. Kim SS, Hudgins AD, Yang J, Zhu Y, Tu Z, Rosenfeld MG, et al. A comprehensive integrated post-GWAS analysis of Type 1 diabetes reveals enhancer-based immune dysregulation. PLoS One. 2021;16(9):e0257265. doi:10.1371/journal.pone.0257265