Persistent homology analysis of type 2 diabetes genome-wide association studies in protein–protein interaction networks

Euijun Song [iD] [1,2]*

[1]Yonsei University College of Medicine, Seoul, Republic of Korea, [2]Present: Independent Researcher, Gyeonggi, Republic of Korea

Genome-wide association studies (GWAS) involving increasing sample sizes have identified hundreds of genetic variants associated with complex diseases, such as type 2 diabetes (T2D); however, it is unclear how GWAS hits form unique topological structures in protein–protein interaction (PPI) networks. Using persistent homology, this study explores the evolution and persistence of the topological features of T2D GWAS hits in the PPI network with increasing p-value thresholds. We define an n-dimensional persistent disease module as a higher-order generalization of the largest connected component (LCC). The 0-dimensional persistent T2D disease module is the LCC of the T2D GWAS hits, which is significantly detected in the PPI network (196 nodes and 235 edges, $P < 0.05$). In the 1-dimensional homology group analysis, all 18 1-dimensional holes (loops) of the T2D GWAS hits persist over all p-value thresholds. The 1-dimensional persistent T2D disease module comprising these 18 persistent 1-dimensional holes is significantly larger than that expected by chance (59 nodes and 83 edges, $P < 0.001$), indicating a significant topological structure in the PPI network. Our computational topology framework potentially possesses broad applicability to other complex phenotypes in identifying topological features that play an important role in disease pathobiology.

KEYWORDS

persistent homology, topological data analysis, disease module, genome-wide association study, network medicine, systems biology, algebraic topology algorithms

1 Introduction

Understanding the genotype–phenotype relationships is challenging owing to their polygenicity and nonlinearity. Complex diseases result from interactions between diverse cellular processes and genes. Elucidating the genetic basis of complex diseases in the context of protein–protein interaction (PPI) networks is essential (Barabasi et al., 2011; Barrio-Hernandez et al., 2023). In the PPI network, genes (or gene products) that have similar biological functions are likely to interact closely with each other. Thus, genes associated with a specific phenotype tend to be clustered into a connected component called a *disease module* in the PPI network (Goh et al., 2007). Disease modules that significantly overlap with each other exhibit similar pathobiological pathways, co-expression patterns, and clinical manifestations (Menche et al., 2015). This disease module concept is useful in identifying novel disease–disease or disease–drug relationships (Menche et al., 2015;

Guney et al., 2016), enabling the implementation of network-based drug repurposing for complex traits/diseases (Song et al., 2020).

Genome-wide association studies (GWAS) have identified numerous genetic variants associated with various complex diseases and can be used to characterize disease-associated modules in the PPI network (Barrio-Hernandez et al., 2023). Genes associated with GWAS loci or GWAS hits tend to be mapped onto coherent network modules in the PPI network. As the p-value threshold increases from 0 to the standard genome-wide significance threshold of 5×10^{-8}, the GWAS hits mapped to the PPI network tend to gradually form a single, connected component (Menche et al., 2015; Leopold and Loscalzo, 2018). This largest connected component (LCC) of disease genes is occasionally called an *observable disease module* (Menche et al., 2015; Paci et al., 2021). However, owing to the limited sample size and coverage of current GWAS data as well as the interactome incompleteness, disease-associated seed genes are often scattered in the PPI network. To detect disease modules, various seed-expanding and/or heuristic-based algorithms have been developed to expand and merge the scattered seed genes in the PPI network (Ghiassian et al., 2015; Vlaic et al., 2018; Wang and Loscalzo, 2018; Choobdar et al., 2019). In addition, machine learning and graph embedding algorithms have been used to predict disease-associated genes (Yang et al., 2019; 2022; Hou et al., 2022) and disease treatment mechanisms (Zitnik et al., 2018; Ruiz et al., 2021) in the context of biological networks. Most studies have focused on identifying connected components by mapping disease-associated seed genes onto the PPI network and expanding these seed genes. However, the mechanism by which GWAS hits are mapped onto the LCC or other unique topological structures in the PPI network as the p-value threshold increases remains unclear. Therefore, the topological features of GWAS hits mapped onto the PPI network warrant investigation.

One mathematical method for analyzing the topological features of complex networks is simplicial homology (Hatcher, 2002; Salnikov et al., 2019). Simplicial homology is an algebraic topology tool used to analyze the topological features of a simplicial complex, which is a collection of higher-order interactions called simplices, including points (0-simplices), line segments (1-simplices), triangles (2-simplices), and higher-dimensional simplices. Simplicial homology can be used to examine the connectivity patterns within biological networks, such as gene-regulatory networks or brain connectivity networks. It can identify topological features, such as connected components (0-holes), loops (1-holes), voids (2-holes), and higher-dimensional holes in the data. For example, the LCC is the largest 0th homology class (connected component). Persistent homology is a method for capturing the persistence of simplicial homology features across multiple thresholds corresponding to a filtration of the simplicial complex (Otter et al., 2017; Salnikov et al., 2019). It can identify important topological features that are persistent across different levels of interaction, rather than artifacts of noise or parameter uncertainty. Persistent homology features of biological networks potentially correspond to biologically relevant components that play a crucial role in disease mechanisms (Sizemore et al., 2019; Masoomy et al., 2021).

The mathematical details of simplicial complex and homology concepts are described in the Method section.

This study analyzes the persistent homology features of GWAS hits in the PPI network to identify important topological structures that potentially play a significant role in disease pathobiology. We analyze the simplicial homology features of GWAS hits in the PPI network as the p-value threshold increases from 0 to 5×10^{-8}. For example, the LCC of the mapped GWAS hits, which is occasionally called an observable disease module, can be considered a connected component (0th homology class) that lives forever. This study aims to expand the LCC concept using higher-order topological structure analysis. We use GWAS summary statistics data of type 2 diabetes (T2D) because T2D has undergone extensive genetic study across diverse ancestry populations with large sample sizes. GWAS with increasing sample sizes have recently identified more than 300 genetic loci associated with T2D (Vujkovic et al., 2020); however, many of these GWAS loci have small effect sizes of unclear pathobiological meaning. Therefore, this study systematically explores the evolution and persistence of the topological features of T2D GWAS hits in the PPI network as the p-value threshold increases from 0 to 5×10^{-8}. We also analyze biological pathways, transcription factors, and microRNAs associated with the persistent homology features.

2 Methods

2.1 Overview of the computational topology framework

This study analyzes the topological features of GWAS hits in the human PPI network. Using persistent homology, we systematically explore n-dimensional holes associated with a specific phenotype, as follows.

1. Map GWAS hits onto the human PPI network.
2. Using persistent homology, identify n-dimensional holes of GWAS hits in the PPI network, as the p-value threshold increases from 0 to 5×10^{-8}.
3. Detect *nth persistent disease modules*, which we define as unions of n-dimensional holes that live forever over all p-value thresholds.
4. Compute the statistical significance of nth persistent disease modules by comparing the result with the randomized distribution of a set of randomly selected nodes in the PPI network.

Since the LCC can be considered a connected component (0th homology class) that lives forever, the nth persistent disease module can be viewed as a higher-order generalization of the LCC concept. We test our computational framework using T2D GWAS summary statistics data, and perform functional enrichment analysis to validate the pathobiological significance of the persistent homology features.

2.2 Consolidated human protein–protein interactome

We used a consolidated human PPI network constructed previously by Wang and Loscalzo (Wang and Loscalzo, 2021; Wang and Loscalzo, 2023). Briefly, the protein–protein interactome was compiled from various sources, including high-throughput yeast-two-hybrid studies, the Center for Cancer Systems Biology (CCSB) human interactome, binary PPIs from other laboratories, protein–protein co-complex interactions, signaling interactions, kinase–substrate interactions, and the Human Reference Interactome (HuRI) binary PPIs. This network possesses a scale-free topology (Wang and Loscalzo, 2021). The LCC of the protein–protein interactome, comprising 16,422 proteins (nodes) and 233,940 interactions (links), was used for the downstream analyses.

2.3 T2D GWAS hits

We used a GWAS meta-analysis summary statistics dataset of 228,499 T2D cases and 1,178,783 controls encompassing multi-ancestral groups (Vujkovic et al., 2020) (downloaded from the GWAS catalog https://www.ebi.ac.uk/gwas/). The standard genome-wide significance threshold of 5×10^{-8} was applied. Each genetic variant was annotated with the closest gene(s) via GWAS catalog gene-mapping data. Only GWAS loci that had been annotated with at most two genes were included. Some genes were linked to multiple GWAS loci with multiple p-values. To extract GWAS hits, for each gene, we assigned the lowest p-value from the different GWAS loci mapped onto that gene (Ratnakumar et al., 2020). We only considered genes (or proteins) in the human PPI network.

2.4 Simplicial complex and homology theory

Here, we briefly describe the fundamentals of the simplicial complex and persistent homology theory (Hatcher, 2002; Otter et al., 2017; Salnikov et al., 2019). An n-dimensional simplex (n-simplex) is formed by $n + 1$ nodes

$$\sigma_n = (v_0, v_1, \ldots, v_n) \tag{1}$$

with an assigned orientation. For example, a 0-simplex is a vertex (node), a 1-simplex is an edge (link), and a 2-simplex is a triangle. An n'-face of an n-simplex ($n' < n$) is a proper subset of the nodes of the simplex with order $n' + 1$. A simplicial complex K is a set of simplices closed under the inclusion of the faces of each simplex. Given a set of n-simplices of a simplicial complex K, an n-dimensional chain (n-chain) is defined as a finite linear combination of n-simplices of K, as follows:

$$c_n = \sum_i b_i \sigma_n^{(i)} \tag{2}$$

where $b_i \in \mathbb{Z}/2\mathbb{Z}$. In this study, we restrict our analysis to homology with $\mathbb{Z}/2\mathbb{Z}$ coefficients. The set of n-chains forms an abelian group denoted by C_n (n-chain group). For any n-simplex

$\sigma_n = (v_0, v_1, \ldots, v_n)$, the boundary operator $\partial_n: C_n \to C_{n-1}$ is the homomorphism defined as follows:

$$\partial_n(\sigma_n) = \sum_{i=0}^n (-1)^i (v_0, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n). \tag{3}$$

An n-chain is said to be a n-cycle if its boundary is zero; that is, elements of the subgroup $Z_n := \ker \partial_n \subseteq C_n$ are n-cycles. Similarly, elements of the subgroup $B_n := \mathrm{im}\ \partial_{n+1} \subseteq C_n$ are said to be n-boundaries. Based on the definition of the boundary operator, it is obvious that any boundary has no boundary (i.e., $\partial_n \partial_{n+1} = 0$). Thus, $B_n \subseteq Z_n \subseteq C_n$. Hence, the nth simplicial homology group H_n of the simplicial complex K can be defined as the quotient abelian group:

$$H_n(K) := Z_n / B_n = \ker \partial_n / \mathrm{im}\ \partial_{n+1}. \tag{4}$$

The rank of the nth homology group H_n is called the nth Betti number β_n. The nth homology group H_n is isomorphic to $\mathbb{Z}^{\beta_n}$, with the basis of independent n-cycles on Z_n modulo boundaries. Intuitively, it represents n-dimensional holes in the simplicial complex K. For example, β_0, β_1, and β_2 represent the number of connected components, loops, and voids, respectively.

Persistent homology is a method for analyzing simplicial topological features at different resolutions of a given simplicial complex (Otter et al., 2017; Salnikov et al., 2019). Formally, a filtration of the simplicial complex K is a finite sequence of subcomplexes $\{K_i \mid 0 \leq i \leq m\}$ such that

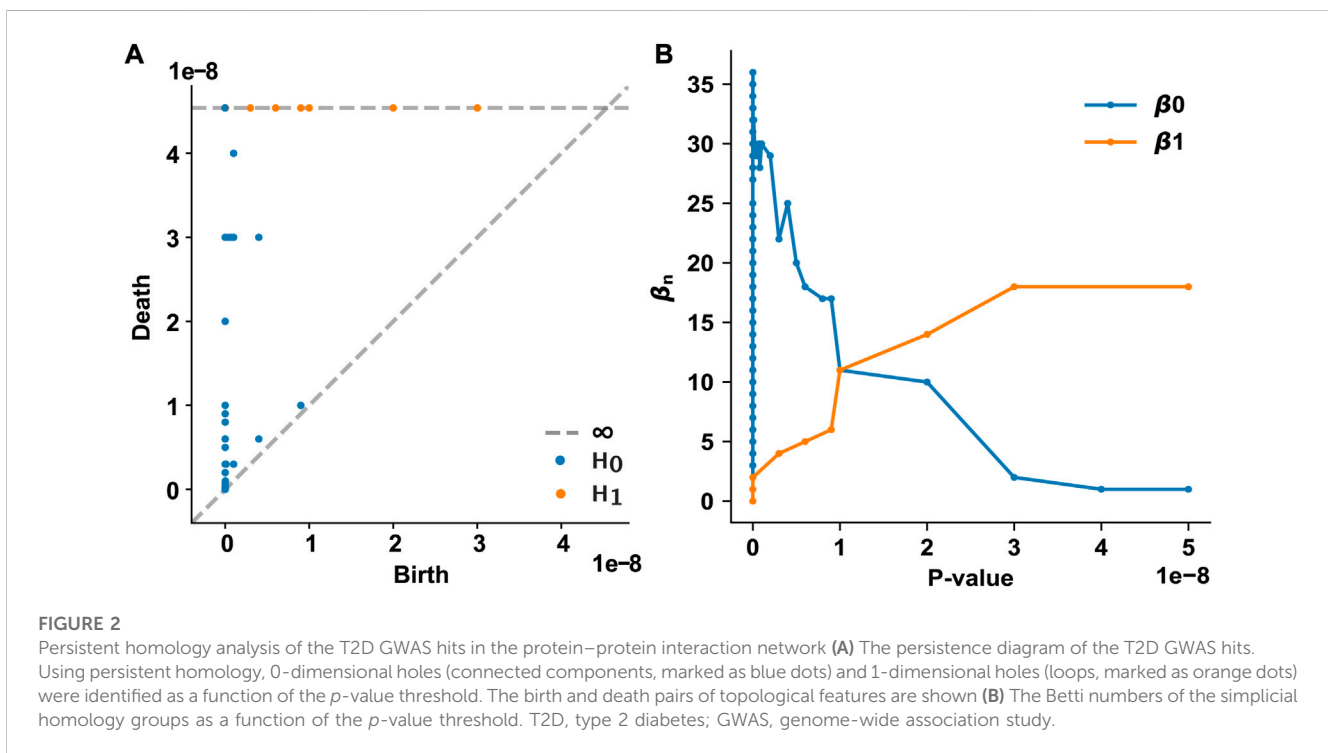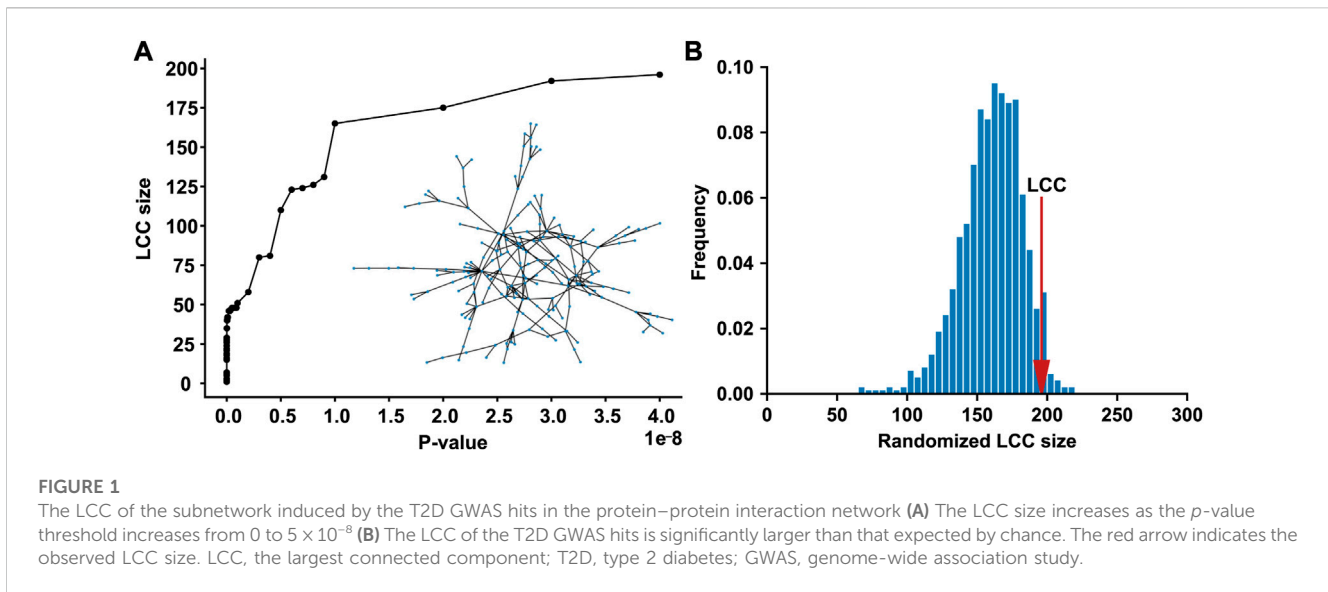$$\varnothing = K_0 \subseteq K_1 \subseteq \cdots \subseteq K_m = K. \tag{5}$$

For $0 \leq i \leq j \leq m$, the inclusion $K_i \hookrightarrow K_j$ induces a homomorphism $h_n^{i,j}: H_n(K_i) \to H_n(K_j)$, and the nth persistent homology groups $PH_n^{i,j}$ are defined as the images of these homomorphisms:

$$PH_n^{i,j} := \mathrm{im}\ h_n^{i,j}. \tag{6}$$

Intuitively, the nth persistent homology groups represent n-dimensional holes that persist from K_i to K_j. We can track when n-dimensional holes appear (birth) and disappear (death) at different threshold values of the filtration. Persistence diagrams, representations of persistent homology, can be constructed by plotting the birth and death sites of topological features.

2.5 Persistent homology analysis of GWAS hits

In this study, the PPI network $G = (V, E)$ is considered a simplicial complex K: genes (or proteins) are regarded as 0-simplexes (nodes), PPIs as 1-simplexes (links), and higher-order connections (or cliques) as high-dimensional simplices. The T2D disease module was identified as the LCC of the PPI subnetwork induced by the T2D GWAS hits. The statistical significance of the LCC was calculated by comparing the observed LCC size with the randomized LCC distribution of a set of randomly selected nodes of the same size in a degree-preserving manner over 1,000 repetitions. The z-score was estimated as $z = \frac{LCC_{obs} - \langle LCC \rangle_{rnd}}{\sigma_{rnd}}$, where LCC_{obs} is the observed LCC size, and $\langle LCC \rangle_{rnd}$ and σ_{rnd} are the mean and SD of the randomized LCC distribution, respectively.
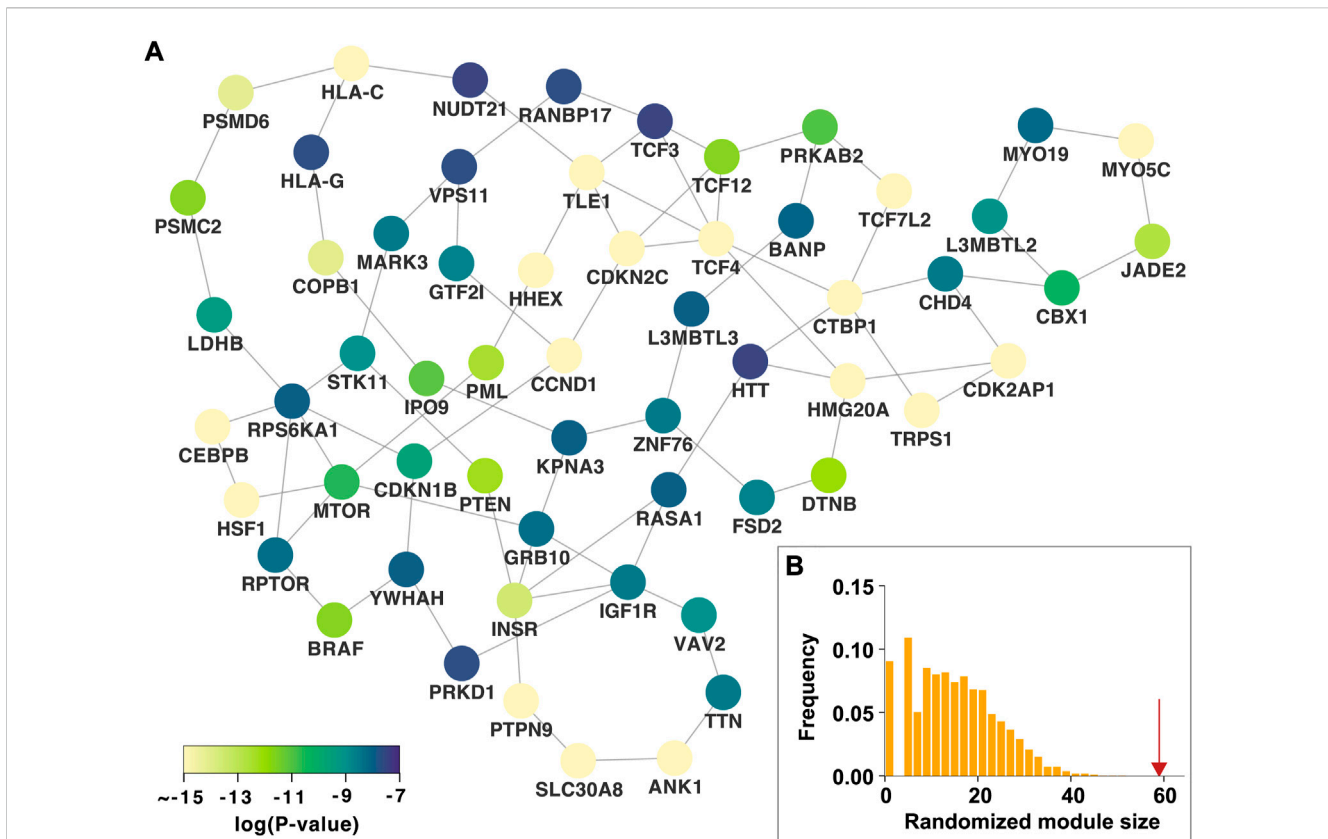
FIGURE 1
The LCC of the subnetwork induced by the T2D GWAS hits in the protein–protein interaction network **(A)** The LCC size increases as the p-value threshold increases from 0 to 5×10^{-8} **(B)** The LCC of the T2D GWAS hits is significantly larger than that expected by chance. The red arrow indicates the observed LCC size. LCC, the largest connected component; T2D, type 2 diabetes; GWAS, genome-wide association study.

FIGURE 2
Persistent homology analysis of the T2D GWAS hits in the protein–protein interaction network **(A)** The persistence diagram of the T2D GWAS hits. Using persistent homology, 0-dimensional holes (connected components, marked as blue dots) and 1-dimensional holes (loops, marked as orange dots) were identified as a function of the p-value threshold. The birth and death pairs of topological features are shown **(B)** The Betti numbers of the simplicial homology groups as a function of the p-value threshold. T2D, type 2 diabetes; GWAS, genome-wide association study.

Each T2D GWAS hit's p-value was used as a varying threshold to obtain a filtration of the PPI subnetwork induced by the T2D GWAS hits as a function of the p-value. As the threshold value increases from 0 to 5×10^{-8}, each node appears at the p-value assigned to that gene. Formally, we define the δ-simplicial complex for the p-value threshold $\delta \geq 0$ as follows:

$$\mathcal{W}_\delta := \{\sigma = (v_0, v_1, \ldots, v_k) \in K \mid \forall i \in \{0, 1, \ldots, k\}, \ p(v_i) \leq \delta\} \quad (7)$$

where $p(v) \in [0, 1]$ is the GWAS hit p-value assigned to the node $v \in V$. Using this δ-simplicial complex, we define the filtration as $\{\mathcal{W}_\delta \hookrightarrow \mathcal{W}_{\delta'}\}_{0 \leq \delta \leq \delta'}$. We subsequently examined the persistent homology features (n-dimensional holes) of this filtration for

each dimension as a function of the p-value threshold. Persistence diagrams were used to visualize the birth and death times of topological features. For each dimension, we also computed the Betti numbers (ranks) of the simplicial homology groups as a function of the p-value threshold.

We define an *nth persistent disease module* as a union of n-dimensional holes that live forever over all p-value thresholds. This definition is concordant with the conventional disease module concept–the 0th persistent disease module is the LCC, which is the persistent 0-dimensional hole (connected component) that lives forever. The statistical significance of the nth persistent disease module was calculated by comparing the observed persistent

FIGURE 3
The 1st persistent T2D disease module in the protein–protein interaction network **(A)** The 1st persistent T2D disease module comprising persistent 1-dimensional holes (loops) that live forever over all *p*-value thresholds. The *p*-values of the T2D GWAS hits are shown **(B)** The 1st persistent T2D disease module is significantly larger than that expected by chance. The red arrow indicates the observed module size. T2D, type 2 diabetes; GWAS, genome-wide association study.

disease module size with the randomized persistent disease module distribution of a set of randomly selected nodes of the same size in a degree-preserving manner. The persistent homology features of randomly selected nodes were analyzed over 1,000 repetitions. The network and homology analyses were performed using the NetworkX and Ripser packages of Python 3.8 (https://www.python.org/), and networks were visualized using Cytoscape 3.9.1 (https://cytoscape.org/). The core code for analyzing persistent homology is publicly available in our GitHub repository (https://github.com/esong0/PHGWAS).

2.6 Functional enrichment analysis

To infer the biological significance of the persistent disease module, a pathway enrichment analysis was performed based on the Kyoto Encyclopedia of Genes and Genomes (KEGG) 2021 database using the GSEApy Python package (Fang et al., 2022) with the Enrichr web server (Kuleshov et al., 2016). In addition, the transcription factor target enrichment analysis was conducted based on the ENCODE and ChEA Consensus databases. The microRNA target enrichment analysis was also performed based on the miRTarBase database. Adjusted *p*-values were computed using the Benjamini–Hochberg method, and statistical significance was set at P < 0.05.
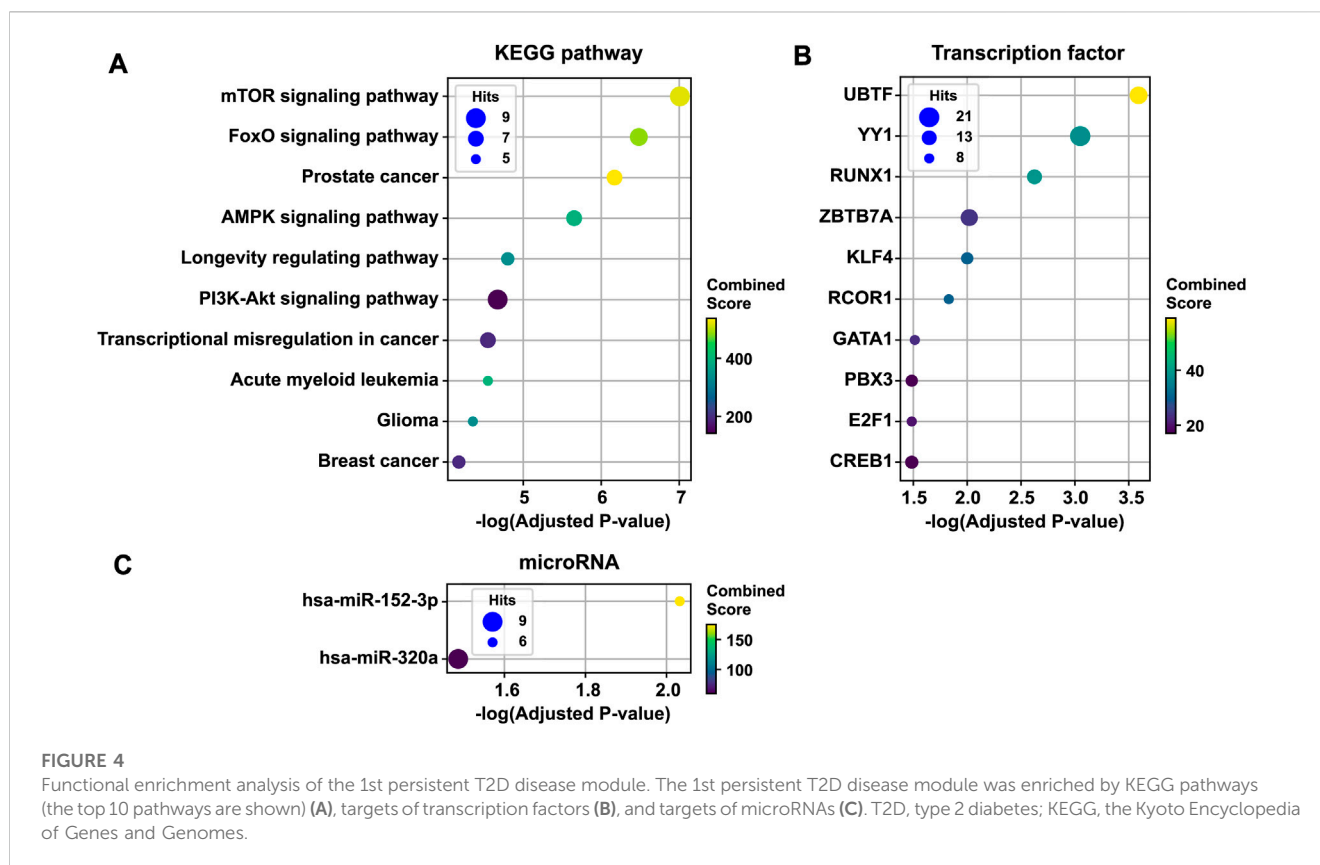
3 Results

3.1 The LCC of GWAS hits

We compiled the T2D GWAS hits using the large-scale T2D GWAS summary statistics data, 565 of which are present in the human PPI network. As the *p*-value threshold increased from 0 to 5×10^{-8}, the LCC of the subnetwork induced by the T2D GWAS hits increased (Figure 1A). When the standard genome-wide significance threshold of 5×10^{-8} was applied, we identified the LCC comprising 196 nodes and 235 edges, which is significantly larger than that expected by chance ($p = 0.0487$, Figure 1B). We defined a T2D observable disease module as this LCC of the T2D GWAS hits (Supplementary Table S1). Other connected components of the subnetwork induced by the T2D GWAS hits comprised ≤5 nodes, which were excluded from the downstream analyses.

3.2 Persistent homology analysis

We examined how the topological features of the T2D disease module evolve and persist in the PPI network as the *p*-value threshold increases from 0 to 5×10^{-8}. In our framework, each node appears at the *p*-value assigned to that gene. We used the *p*-value of each T2D GWAS hit as a varying threshold and determined the timing of the appearance

FIGURE 4
Functional enrichment analysis of the 1st persistent T2D disease module. The 1st persistent T2D disease module was enriched by KEGG pathways (the top 10 pathways are shown) **(A)**, targets of transcription factors **(B)**, and targets of microRNAs **(C)**. T2D, type 2 diabetes; KEGG, the Kyoto Encyclopedia of Genes and Genomes.

(birth) and disappearance (death) of n-dimensional holes at different threshold values. In the 0th homology group (H_0) analysis, 61 0-dimensional holes (connected components) were identified, of which only one persisted over all p-value thresholds (Figure 2A). This persistent 0-dimensional hole is the LCC, that is, the T2D observable disease module. In the 1st homology group (H_1) analysis, 18 1-dimensional holes (loops or 1-cycles) were identified, all of which persisted over all p-value thresholds (Figure 2A). No higher-dimensional hole ($n \geq 2$) existed in the T2D GWAS hit data. The Betti numbers (ranks) of the simplicial homology groups are shown in Figure 2B. As the p-value threshold increases, the number of 0-dimensional holes converges to 1 (i.e., the LCC), while the number of 1-dimensional holes increases and converges to 18.

We identified the nth persistent disease modules, which were defined as unions of persistent n-dimensional holes that live forever over all p-value thresholds. The 0th persistent disease module is the LCC, which is the persistent 0-dimensional hole that lives forever. As shown in Figure 1B, the LCC is significantly larger than that expected by chance. Since the LCC concept has been extensively investigated in various complex diseases, this study focused on the 1st persistent disease module. In our T2D GWAS data analysis, we identified 18 persistent 1-dimensional holes (loops or 1-cycles), which constitute the 1st persistent T2D disease module comprising 59 nodes and 83 edges (Figure 3A). This 1st persistent T2D disease module is significantly larger than that expected by chance ($P < 0.001$, Figure 3B), indicating a significant topological feature of the T2D GWAS hits in the PPI network. Since the lowest p-value in the T2D GWAS data is as extremely small as 3e-695 (rs7903146 in *TCF7L2*), we repeated our analysis using the log p-value scale. The

same 61 0-dimensional holes and 18 1-dimensional holes were also identified (Supplementary Figure S1).

3.3 Biological pathways, transcription factors, and microRNAs

To infer the pathobiological significance of the 1st persistent T2D disease module, we identified over-represented KEGG pathways. The top 10 enriched KEGG pathways included mTOR signaling, FoxO signaling, AMPK signaling, the longevity regulating pathway, PI3K-Akt signaling, the transcriptional misregulation pathway in cancer, and several cancer pathways (Figure 4A). In addition, the 1st persistent T2D disease module was enriched with targets of transcription factors, including UBTF, YY1, RUNX1, ZBTB7A, KLF4, RCOR1, GATA1, PBX3, E2F1, and CREB1 (Figure 4B). The 1st persistent T2D disease module was also enriched with targets of microRNAs, including hsa-miR-152–3p and hsa-miR-320a (Figure 4C).

4 Discussion

Using persistent homology, this study explored the evolution and persistence of the topological features of T2D GWAS hits in the PPI network as the p-value threshold increased from 0 to 5×10^{-8}. The nth persistent disease module was defined as a union of persistent n-dimensional holes that live forever over all p-value thresholds. This is a higher-order generalization of the conventional disease module concept. The 0th persistent T2D disease module is the LCC of the T2D

GWAS hits, which is significantly larger than that expected by chance. In the 1st homology group analysis, all 18 1-dimensional holes (loops) of the T2D GWAS hits persist over all p-value thresholds. The 1st persistent T2D disease module comprising these 18 persistent 1-dimensional holes is significantly larger than that expected by chance, indicating a significant topological structure in the PPI network. The 1st persistent T2D disease module is enriched with the mTOR, FoxO, AMPK, and PI3K-Akt signaling pathways; longevity regulating pathway; and cancer pathways. It has been known that the mechanisms of T2D, aging, and cancer are closely related to each other (Wei et al., 2017). The pathobiological significance of this persistent disease module is subject to subsequent experimental validation.

Our computational topology framework potentially has broad applicability to other complex phenotypes. By analyzing the persistent homology features, the higher-order topological features that may be closely associated with a specific phenotype can be identified. We plan to expand this preliminary study to systematically analyze the topological features of the large-scale disease–gene networks (Menche et al., 2015; Guney et al., 2016). We expect that there are several mathematical ways to expand our persistent homology approach in PPI networks. The weighted topology (Baccini et al., 2022) of weighted PPI networks reflecting proteome-wide binding affinity and concentration information should provide more biologically plausible and reliable information. In addition, relational persistent homology (Stolz et al., 2023) may be a useful tool for dissecting multispecies data, such as multiomics data or multilayer biological networks.

Notwithstanding, this study has several potential limitations. While the conventional disease module concept typically relies on connected components (0th homology class) of disease seeds, the proposed persistent disease module concept is a higher-order generalization of the LCC. Hence, it is hard to directly compare our homology approach to most other disease module identification algorithms. Therefore, it is essential to develop higher-order versions of seed-expanding algorithms to detect robust and reliable persistent disease modules. The role of seed connectors (Wang and Loscalzo, 2018) in homology features also warrants elucidation. As the uncertainty and incompleteness of GWAS and PPI network data are inevitable (Menche et al., 2015), how these errors and uncertainty affect the robustness of persistent disease modules remains unclear. Although no higher-dimensional hole ($n \geq 2$) was present in our T2D GWAS hit data, higher-order interactions may play a significant role in disease pathobiology. Dynamic topological data analysis approaches based on sequential data would provide more rigorous and robust results (Ciocanel et al., 2021). Tissue- or cell-type-specific networks (Greene et al., 2015) should provide more biological information regarding persistent disease modules. Determining whether oncogenic mutations perturb PPI or higher-order interactions in the PPI network is a worthwhile endeavor (Cheng et al., 2021).

Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material, further inquiries can be directed to the corresponding author.

Author contributions

ES: Conceptualization, Formal Analysis, Investigation, Methodology, Visualization, Writing–original draft.

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2023.1270185/full#supplementary-material

References

Baccini, F., Geraci, F., and Bianconi, G. (2022). Weighted simplicial complexes and their representation power of higher-order network data and topology. *Phys. Rev. E* 106, 034319. doi:10.1103/PhysRevE.106.034319

Barabasi, A. L., Gulbahce, N., and Loscalzo, J. (2011). Network medicine: a network-based approach to human disease. *Nat. Rev. Genet.* 12, 56–68. doi:10.1038/nrg2918

Barrio-Hernandez, I., Schwartzentruber, J., Shrivastava, A., del Toro, N., Gonzalez, A., Zhang, Q., et al. (2023). Network expansion of genetic associations defines a pleiotropy map of human cell biology. *Nat. Genet.* 55, 389–398. doi:10.1038/s41588-023-01327-9

Cheng, F., Zhao, J., Wang, Y., Lu, W., Liu, Z., Zhou, Y., et al. (2021). Comprehensive characterization of protein-protein interactions perturbed by disease mutations. *Nat. Genet.* 53, 342–353. doi:10.1038/s41588-020-00774-y

Choobdar, S., Ahsen, M. E., Crawford, J., Tomasoni, M., Fang, T., Lamparter, D., et al. (2019). Assessment of network module identification across complex diseases. *Nat. Methods* 16, 843–852. doi:10.1038/s41592-019-0509-5

Ciocanel, M.-V., Juenemann, R., Dawes, A. T., and McKinley, S. A. (2021). Topological data analysis approaches to uncovering the timing of ring structure onset in filamentous networks. *Bull. Math. Biol.* 83, 21. doi:10.1007/s11538-020-00847-3

Fang, Z., Liu, X., and Peltz, G. (2022). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* 39, btac757. doi:10.1093/bioinformatics/btac757

Ghiassian, S. D., Menche, J., and Barabasi, A. L. (2015). A disease module detection (diamond) algorithm derived from a systematic analysis of connectivity patterns of disease proteins in the human interactome. *PLoS Comput. Biol.* 11, e1004120. doi:10.1371/journal.pcbi.1004120

Goh, K. I., Cusick, M. E., Valle, D., Childs, B., Vidal, M., and Barabasi, A. L. (2007). The human disease network. *Proc. Natl. Acad. Sci. U S A.* 104, 8685–8690. doi:10.1073/pnas.0701361104

Greene, C. S., Krishnan, A., Wong, A. K., Ricciotti, E., Zelaya, R. A., Himmelstein, D. S., et al. (2015). Understanding multicellular function and disease with human tissue-specific networks. *Nat. Genet.* 47, 569–576. doi:10.1038/ng.3259

Guney, E., Menche, J., Vidal, M., and Barabasi, A. L. (2016). Network-based *in silico* drug efficacy screening. *Nat. Commun.* 7, 10331. doi:10.1038/ncomms10331

Hatcher, A. (2002). *Algebraic topology.* Cambridge University Press.

Hou, S., Zhang, P., Yang, K., Wang, L., Ma, C., Li, Y., et al. (2022). Decoding multilevel relationships with the human tissue-cell-molecule network. *Briefings Bioinforma.* 23, bbac170. doi:10.1093/bib/bbac170

Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44, W90–W97. doi:10.1093/nar/gkw377

Leopold, J. A., and Loscalzo, J. (2018). Emerging role of precision medicine in cardiovascular disease. *Circ. Res.* 122, 1302–1315. doi:10.1161/CIRCRESAHA.117.310782

Masoomy, H., Askari, B., Tajik, S., Rizi, A. K., and Jafari, G. R. (2021). Topological analysis of interaction patterns in cancer-specific gene regulatory network: persistent homology approach. *Sci. Rep.* 11, 16414. doi:10.1038/s41598-021-94847-5

Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., et al. (2015). Disease networks. uncovering disease-disease relationships through the incomplete interactome. *Science* 347, 1257601. doi:10.1126/science.1257601

Otter, N., Porter, M. A., Tillmann, U., Grindrod, P., and Harrington, H. A. (2017). A roadmap for the computation of persistent homology. *EPJ Data Sci.* 6, 17. doi:10.1140/epjds/s13688-017-0109-5

Paci, P., Fiscon, G., Conte, F., Wang, R. S., Farina, L., and Loscalzo, J. (2021). Gene co-expression in the interactome: moving from correlation toward causation via an integrated approach to disease module discovery. *NPJ Syst. Biol. Appl.* 7, 3. doi:10.1038/s41540-020-00168-0

Ratnakumar, A., Weinhold, N., Mar, J. C., and Riaz, N. (2020). Protein-protein interactions uncover candidate 'core genes' within omnigenic disease networks. *PLoS Genet.* 16, e1008903. doi:10.1371/journal.pgen.1008903

Ruiz, C., Zitnik, M., and Leskovec, J. (2021). Identification of disease treatment mechanisms through the multiscale interactome. *Nat. Commun.* 12, 1796. doi:10.1038/s41467-021-21770-8

Salnikov, V., Cassese, D., and Lambiotte, R. (2019). Simplicial complexes and complex systems. *Eur. J. Phys.* 40, 014001. doi:10.1088/1361-6404/aae790

Sizemore, A. E., Phillips-Cremins, J. E., Ghrist, R., and Bassett, D. S. (2019). The importance of the whole: topological data analysis for the network neuroscientist. *Netw. Neurosci.* 3, 656–673. doi:10.1162/netn_a_00073

Song, E., Wang, R. S., Leopold, J. A., and Loscalzo, J. (2020). Network determinants of cardiovascular and repositioned drug treatments. *FASEB J.* 34, 11087–11100. doi:10.1096/fj.202001062R

Stolz, B. J., Dhesi, J., Bull, J. A., Harrington, H. A., Byrne, H. M., and Yoon, I. H. R. (2023). Relational persistent homology for multispecies data with application to the tumor microenvironment. arXiv 2308.06205. doi:10.48550/arXiv.2308.06205

Vlaic, S., Conrad, T., Tokarski-Schnelle, C., Gustafsson, M., Dahmen, U., Guthke, R., et al. (2018). Modulediscoverer: identification of regulatory modules in protein-protein interaction networks. *Sci. Rep.* 8, 433. doi:10.1038/s41598-017-18370-2

Vujkovic, M., Keaton, J. M., Lynch, J. A., Miller, D. R., Zhou, J., Tcheandjieu, C., et al. (2020). Discovery of 318 new risk loci for type 2 diabetes and related vascular outcomes among 1.4 million participants in a multi-ancestry meta-analysis. *Nat. Genet.* 52, 680–691. doi:10.1038/s41588-020-0637-y

Wang, R. S., and Loscalzo, J. (2021). Network module-based drug repositioning for pulmonary arterial hypertension. *CPT Pharmacometrics Syst. Pharmacol.* 10, 994–1005. doi:10.1002/psp4.12670

Wang, R. S., and Loscalzo, J. (2018). Network-based disease module discovery by a novel seed connector algorithm with pathobiological implications. *J. Mol. Biol.* 430, 2939–2950. doi:10.1016/j.jmb.2018.05.016

Wang, R. S., and Loscalzo, J. (2023). Uncovering common pathobiological processes between covid-19 and pulmonary arterial hypertension by integrating omics data. *Pulm. Circ.* 13, e12191. doi:10.1002/pul2.12191

Wei, M., Brandhorst, S., Shelehchi, M., Mirzaei, H., Cheng, C. W., Budniak, J., et al. (2017). Fasting-mimicking diet and markers/risk factors for aging, diabetes, cancer, and cardiovascular disease. *Sci. Transl. Med.* 9, eaai8700. doi:10.1126/scitranslmed.aai8700

Yang, K., Wang, R., Liu, G., Shu, Z., Wang, N., Zhang, R., et al. (2019). Hergepred: heterogeneous network embedding representation for disease gene prediction. *IEEE J. Biomed. Health Inf.* 23, 1805–1815. doi:10.1109/JBHI.2018.2870728

Yang, K., Zheng, Y., Lu, K., Chang, K., Wang, N., Shu, Z., et al. (2022). Pdgnet: predicting disease genes using a deep neural network with multi-view features. *IEEE/ACM Trans. Comput. Biol. Bioinforma.* 19, 575–584. doi:10.1109/TCBB.2020.3002771

Zitnik, M., Agrawal, M., and Leskovec, J. (2018). Modeling polypharmacy side effects with graph convolutional networks. *Bioinformatics* 34, i457–i466. doi:10.1093/bioinformatics/bty294