International Journal of
Molecular Sciences

MDPI

Article

Cross-Domain Text Mining of Pathophysiological Processes Associated with Diabetic Kidney Disease

Krutika Patidar [1], Jennifer H. Deng [2], Cassie S. Mitchell [2,3,*] and Ashlee N. Ford Versypt [1,4,5,*]

1 Department of Chemical and Biological Engineering, University at Buffalo, Buffalo, NY 14260, USA
2 Wallace H. Coulter Department of Biomedical Engineering, Georgia Institute of Technology and Emory University, Atlanta, GA 30332, USA
3 Center for Machine Learning at Georgia Tech, Georgia Institute of Technology, Atlanta, GA 30332, USA
4 Department of Biomedical Engineering, University at Buffalo, Buffalo, NY 14260, USA
5 Institute for Artificial Intelligence and Data Science, University at Buffalo, Buffalo, NY 14260, USA
* Correspondence: cassie.mitchell@bme.gatech.edu (C.S.M.); ashleefv@buffalo.edu (A.N.F.V.)

Abstract: Diabetic kidney disease (DKD) is the leading cause of end-stage renal disease worldwide. This study's goal was to identify the signaling drivers and pathways that modulate glomerular endothelial dysfunction in DKD via artificial intelligence-enabled literature-based discovery. Cross-domain text mining of 33+ million PubMed articles was performed with SemNet 2.0 to identify and rank multi-scalar and multi-factorial pathophysiological concepts related to DKD. A set of identified relevant genes and proteins that regulate different pathological events associated with DKD were analyzed and ranked using normalized mean HeteSim scores. High-ranking genes and proteins intersected three domains—DKD, the immune response, and glomerular endothelial cells. The top 10% of ranked concepts were mapped to the following biological functions: angiogenesis, apoptotic processes, cell adhesion, chemotaxis, growth factor signaling, vascular permeability, the nitric oxide response, oxidative stress, the cytokine response, macrophage signaling, NFκB factor activity, the TLR pathway, glucose metabolism, the inflammatory response, the ERK/MAPK signaling response, the JAK/STAT pathway, the T-cell-mediated response, the WNT/β-catenin pathway, the renin–angiotensin system, and NADPH oxidase activity. High-ranking genes and proteins were used to generate a protein–protein interaction network. The study results prioritized interactions or molecules involved in dysregulated signaling in DKD, which can be further assessed through biochemical network models or experiments.

Keywords: diabetic kidney disease; immune response; glomerular endothelial cells; biomedical text mining; disease relatedness; functional ontology; machine learning; artificial intelligence

check for updates

Citation: Patidar, K.; Deng, J.H.; Mitchell, C.S.; Ford Versypt, A.N. Cross-Domain Text Mining of Pathophysiological Processes Associated with Diabetic Kidney Disease. *Int. J. Mol. Sci.* **2024**, *25*, 4503. https://doi.org/10.3390/ijms25084503

Academic Editor: Dong-Jun Yu

Received: 13 March 2024
Revised: 16 April 2024
Accepted: 17 April 2024
Published: 19 April 2024

Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Diabetic kidney disease (DKD) is a major microvascular complication in the kidney that affects patients with type I diabetes and type II diabetes. Approximately 20–50% of patients with type II diabetes will develop DKD [1]. DKD can lead to a decline in kidney function and has the potential to develop into chronic kidney disease or end-stage renal disease (ESRD) [2,3]. A high clinical and socio-economic impact of DKD is burdensome because of the risk of progression to ESRD and other related comorbidities [4]. The progression of chronic and end-stage renal failure is estimated to affect over 10% of the general population, or more than 800 million people, worldwide [5]. The pathophysiology of DKD is multi-factorial and characterized by metabolic impairment, an uncontrolled inflammatory response, increased apoptosis, and tissue fibrosis [4,6]. In diabetes, aberrant glucose metabolism leads to dysregulation of the immune response and signaling [6–8]. Metabolic abnormalities activate the mononuclear phagocyte system, which releases pro-inflammatory cytokines and paracrine signals, leading to immune cell infiltration [6–9]. In the early stages of diabetes, glomerular and tubular cells have increased expression

of tumor necrosis factor (TNF) α, interleukin (IL)-6, IL-1, and adhesion molecules [10,11]. Previous experimental studies showed that hyperglycemia promotes the imbalance in macrophage phenotypes and chronic glucose exacerbates the regulation of inflammatory responses [12,13]. The dysregulated signaling response leads to progressive kidney damage through loss of glomerular endothelial fenestrations, thickening of the basement membrane, the detachment of podocyte foot processes, mesangial matrix expansion, and glomerular fibrosis [7,14–17]. For excellent illustrations of the pathological processes, readers are referred to [7,8,16,17].

Several mathematical models have demonstrated the pathophysiological processes affected at the cellular or tissue scale due to underlying kidney disease [9,15,18–20] (Table A1). Mathematical models integrate literature-based information into a quantitative framework, identifying knowledge gaps and generating new experimentally testable hypotheses. Studies have demonstrated that it is relevant to comprehend previously applied knowledge and relate it across disease domains to identify potential mechanisms and targeted treatment. A recent mathematical model of the disease etiology of glomerular fibrosis in DKD has successfully adapted and related the key attributes of glomerular fibrosis in lupus nephritis, a type of kidney disease [15]. Hyperglycemia-induced podocyte injury in DKD has been previously modeled using the local renin–angiotensin system in renal podocyte cells [18]. It is well known that dysregulation of the renin–angiotensin–aldosterone system is a phenomenon implicated in the pathophysiology of both renal disorders [6] and cardiovascular disorders, such as hypertension and congestive heart failure [21].

Mathematical models have also studied the pathophysiological interplay of the kidney with other organ disorders related to cardiavascular conditions [22], cancer [23], or bone metabolism [24,25] at a multi-scale and multi-organ level using existing individual cardiac, renal, or metabolic bone disorder models (Table A1). Examples of such models vary across a spectrum of first-principles-based mathematical models to complex network models. In each case, the parameters are typically estimated from the literature, and decisions must be made as to what concepts should be included. A few examples of such relevant models of kidney disease are highlighted below.

A relevant mathematical model was developed to examine the progression from tubulointerstitial inflammation to fibrosis in lupus nephritis using limited available knowledge of the pathogenesis of inflammation in lupus nephritis [20]. The model was built using the first principles of engineering and physics with parameters estimated using experimental literature. The advantage of such a first-principles model is the inherent interpretability. However, many pathophysiological concepts were simplified to make the mathematics tractable.

A different approach to mathematically modeling kidney disease used a signaling network topology and regulatory motifs in podocytes. The model contributed to the understanding of the establishment and maintenance of the morphologically distinct state of the podocytes in the kidney [19]. An extensive literature survey of cell culture studies and proteomics data was performed to build network models with varying levels of detail and provide insights into treatment strategies for kidney disease [19].

Another example of an approach to mathematically modeling kidney disease used a protein–protein interaction network model (Figure 1). The model was proposed using interaction databases and produced in vitro experimental data to understand the progression of early signs of endothelial dysfunction in DKD [9]. The multi-cell network was a manually curated, simplified network of pathway interactions and signaling molecules that affect glomerular endothelial fenestrations in the diabetic kidney. However, that biochemical network [9] only incorporated a subset of the relevant pathways, interactions, or molecules governing DKD progression that were derived from relevant publications.

A further way to understand DKD-related pathways is the use of bioinformatics analyses across different cell types in DKD. A single-cell renal endothelial transcriptomic atlas using differential gene expression revealed the alteration of signaling pathways including oxidative phosphorylation and growth factor signaling in renal endothelial cell subtypes [26]. The recent application of single-cell RNA and bulk RNA sequencing data to study the co-stimulatory interactions between renal endothelial cells and macrophages has been useful in identifying the immunological markers of DKD [27,28].

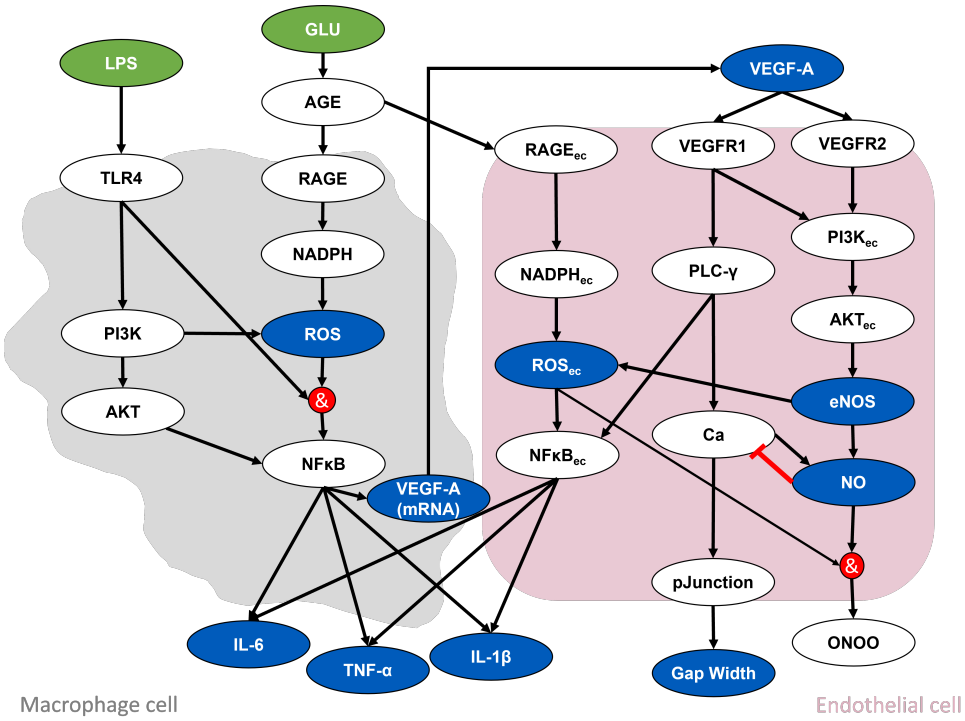

Figure 1. A multi-cellular protein–protein interaction network of crosstalk between macrophages (**left**, grey shape) and glomerular endothelial cells (**right**, pink shape) stimulated with glucose (GLU) and lipopolysaccharide (LPS), a pro-inflammatory stimulus, was created in our previous work through manual curation of the literature [9]. Green nodes (ovals) are input nodes, blue nodes are output nodes, and white nodes are regulatory nodes. Black arrows are activating interactions, a red line with a flat-head arrow is an inhibiting interaction, and red circles indicate logic AND gates. An OR logic rule connects two or more edges to a subsequent node throughout the network unless indicated otherwise by an AND logic gate. The subscript (ec) denotes an intracellular species expressed in glomerular endothelial cells. IL-6, TNF-α, IL-1β, and VEGF-A are protein levels expressed in extracellular space. ROS, ROSec, VEGF-A (mRNA), and NO are expressed within the cells. The Gap Width node denotes a fractional change in the glomerular endothelial cell fenestration size. The pJunction node represents the phosphorylated junction protein levels. TLR: toll-like receptor. AGE: advanced glycation end product. RAGE: receptor of advanced glycation end product. NADPH: nicotinamide adenine dinucleotide phosphate. NFκB: nuclear factor kappa B. IL: interleukin. TNF: tumor necrosis factor. PI3K: phosphoinositide 3-kinase. AKT: serine/threonine-specific protein kinases. ROS: reactive oxygen species. VEGF: vascular endothelial growth factor. VEGFR: vascular endothelial growth factor receptor. PLC: phospholipase C. NO: nitric oxide. ONOO: peroxynitrite. eNOS: endothelial nitric oxide synthase. Ca: calcium. Reprinted/adapted with permission from Ref. [9], 2023, K. Patidar and A. N. Ford Versypt.

In short, regardless of the mathematical method implemented, the relevant pathway interconnections to include are typically decided using domain knowledge and knowledge acquired from manual literature curation. However, a human cannot read all possible sources of important relationships that should be included in a multi-scalar, multi-factorial model of a complex pathophysiology like DKD. Moreover, signaling molecules in the disease pathway may be located further away from the target molecule or event of interest within the scientific literature [29]. Thus, a manual literature search may be insufficient for finding relationships not only within the DKD literature, but also in the cross-domain literature (cardiology, endocrinology, immunology, etc.) where relevant concepts and relationships may reside. In contrast, advanced high-throughput artificial intelligence approaches to literature-based discovery (LBD) can provide a holistic understanding of relevant dysregulated pathways and molecules. LBD approaches enabled by text mining infer disparate sources of information [30] at a scale not otherwise possible.

An example of a recent, comprehensive LBD approach to identifying and ranking relevant concepts is SemNet 2.0 [31,32]. SemNet is an open-source framework consisting of a knowledge graph that identifies and ranks the most important concepts to a user-defined target concept(s) (e.g., keyword). The graph consists of relationships extracted from 33+ million published articles in the PubMed database. The nodes are concepts defined by the Unified Medical Language System. The framework uses SemMedDB as a relationship extraction system for making the graph [33]. The unsupervised learning ranking algorithm within SemNet 2.0 examines relationship patterns in the literature to rank cross-domain concepts with respect to the user-defined concept(s) [31]. SemNet 2.0 has been used for drug repurposing for COVID-19 [34] and Parkinson's disease [35], identifying unknown disease mechanisms of resistant hypertension following COVID-19 infection [36], predicting adverse events from chronic tyrosine kinase inhibitor therapy in chronic myeloid leukemia [29], and identifying clinical features by which to better stratify chemotherapy-related infection risk in pediatric acute leukemia [37].

The objective of the present study was to identify and rank cross-domain concepts that comprise the complex multi-factorial and multi-scalar nature of DKD using advanced artificial intelligence-assisted LBD software. In this study, SemNet 2.0 was used to identify and rank critical signaling molecules associated with the glucose- and inflammation-mediated development and progression of DKD. The general workflow is shown in Figure 2. This comprehensive evaluation enabled the prioritization of highly ranked concepts, as well as the identification of concepts or phenomena missing from current mathematical network models of DKD. Our study results indicate that the utilized LBD approach provided a less biased and more comprehensive manner of integrating cross-domain knowledge into the mechanistic understanding of DKD.

The contributions of this work are as follows:

- The developed workflow with SemNet 2.0 and additional post hoc analysis with Gene Ontology (GO) networks and Cytoscape support the integrative prioritization of literature relationships that advance the study of DKD.
- The top-ranked source nodes identified by SemNet 2.0 enable the more comprehensive construction of protein–protein interactions, the efficient modeling of biochemical pathways, and the expedited testing of literature-based hypotheses through experiments.
- The application of artificial intelligence-enabled LBD techniques to DKD research enables faster integration and processing of new biomedical literature towards secondary analyses that improve drug discovery and target treatment design.

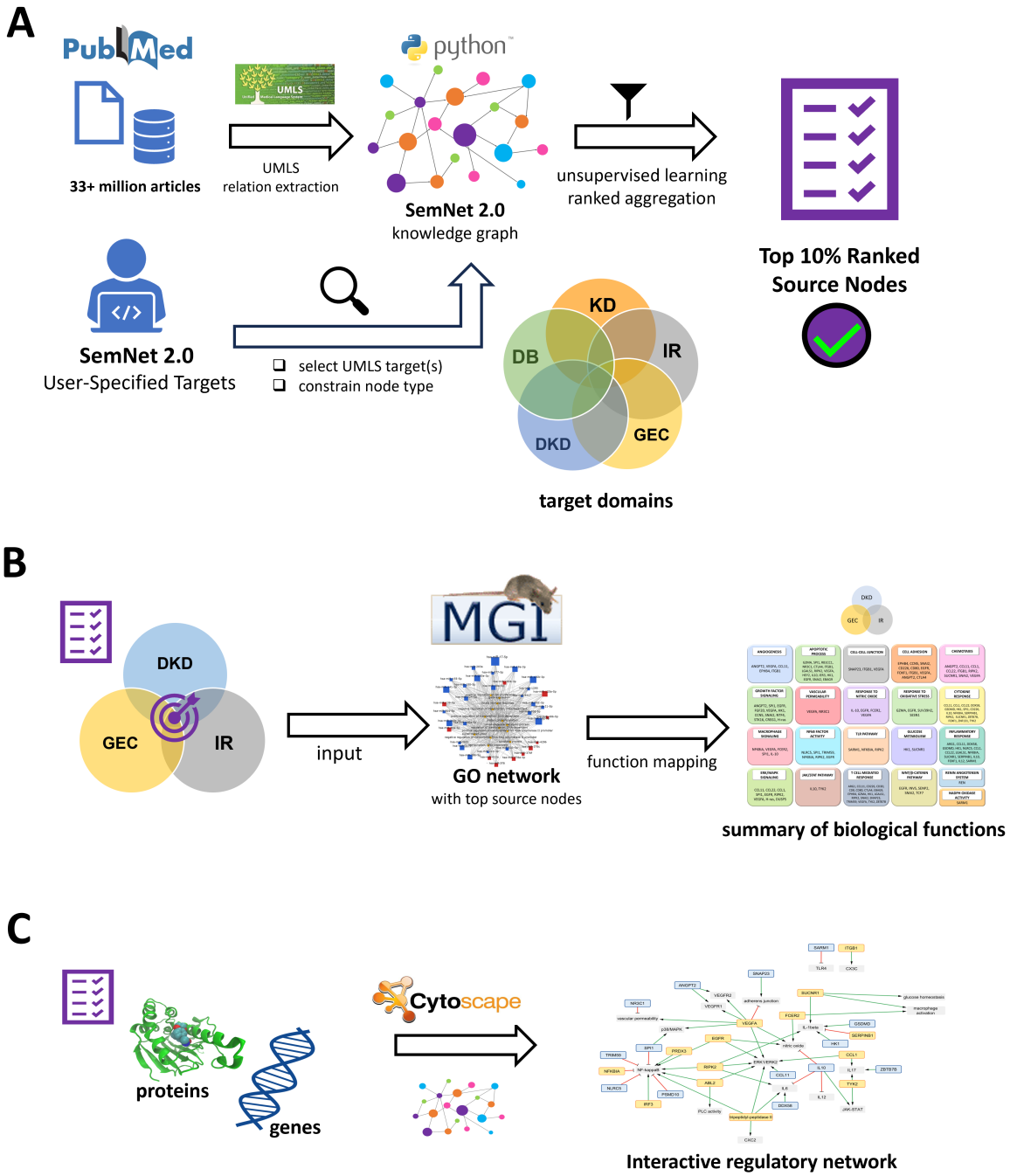

Figure 2. The baseline data used in this project included a knowledge graph with semantic text relationships extracted from 33+ million PubMed articles. The workflow consisted of (**A**) performing a SemNet 2.0 analysis to identify the top-ranked source nodes across domains, (**B**) performing a Gene Ontology (GO) network analysis to summarize the biological functions of the top-ranked source nodes, and (**C**) visualizing regulatory relationships from the top-ranked genes and proteins using Cytoscape [38,39]. Note that the target domains are labeled as diabetes (DB), kidney disease (KD), immune response (IR), diabetic kidney disease (DKD), and glomerular endothelial cells (GEC).

2. Results

The LBD results from SemNet 2.0 identified and ranked the shared concepts among five Unified Medical Language System (UMLS)-defined pathophysiological concepts relevant to the study of DKD, including DKD itself, as well as diabetes (DB), kidney disease (KD), the immune response (IR), and glomerular endothelial cells (GEC). The intersecting top 10% of source genes and proteins at the intersection of the DKD, IR, and GEC domains were used to map the biological functions with GO network analysis. Finally, the mapped biological functions and potential regulatory relationships of the identified source nodes were visualized in a Cytoscape [38,39] network to expedite subsequent biochemical models. Collectively, these results provide a comprehensive, prioritized summary of key relationships important to the study and treatment of DKD. The specific results are discussed below with deeper supporting context.

2.1. Source Nodes Revealed by Cross-Domain Analysis

The distribution of source nodes in each semantic type for each pairwise domain analysis is shown in Figure 3. For instance, the distribution of source nodes at the intersection of both the DB and DKD domains was found in the cross-domain analysis between the DB and DKD (DB-DKD) domains. Higher distributions of intersecting nodes of the semantic type gene (gngm) or type protein (aapp) were observed in each pairwise domain (Figure 3). Nodes belonging to the gngm or aapp semantic type from the top 10% of source nodes were selected for further evaluation of their biological themes and functional relevance in the progression of DKD. The distributions of these top 10% of source nodes in each paired domain are shown and characterized by their mean HeteSim scores and counts (Figure 4). The source nodes are represented by a unique bubble with a color indicative of the pairwise domain and a size indicative of the frequency of node occurrence (Figure 4). The top 10% of source nodes obtained from the post hoc analysis belonged mainly to the GEC, DKD, and IR domains. As seen in Figure 4, more source nodes were present at the intersection of the GEC-IR, DKD-IR, and DKD-GEC domains relative to the other pairwise domains. Most of the identified source nodes with a high mean HeteSim score were present at the intersection of the DKD-IR and GEC-IR domains (Figure 4).

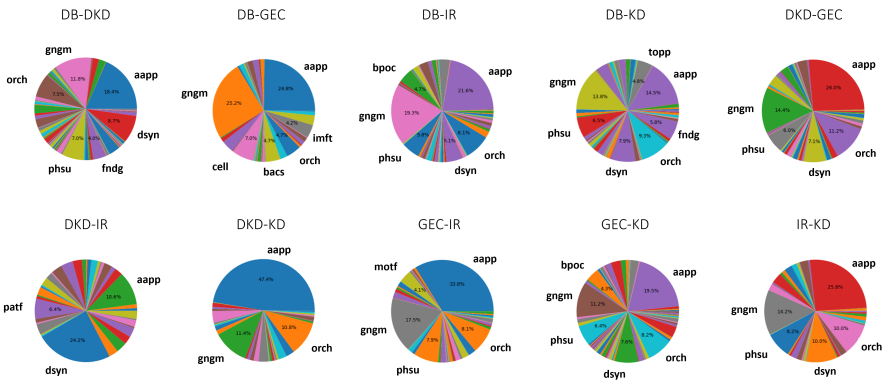

Figure 3. Semantic type distribution of source nodes identified at the intersection of each pairwise domain. Arbitrary colors are assigned to each pie chart segment. Labels are provided in the diagram for types that reach at least 4% share of a distribution. gngm: gene or genome. aapp: amino acid, peptide, or protein. dsyn: disease or syndrome. phsu: pharmacological substance. fndg: finding. imft: immunologic factor. orch: organic chemical. cell: cell. bacs: biologically active substance. patf: pathologic function. bpoc: body part, organ, or organ component. topp: therapeutic or preventive procedures. DB: diabetes domain. DKD: diabetic kidney disease domain. GEC: glomerular endothelial cells domain. IR: immune response domain. KD: kidney disease domain.

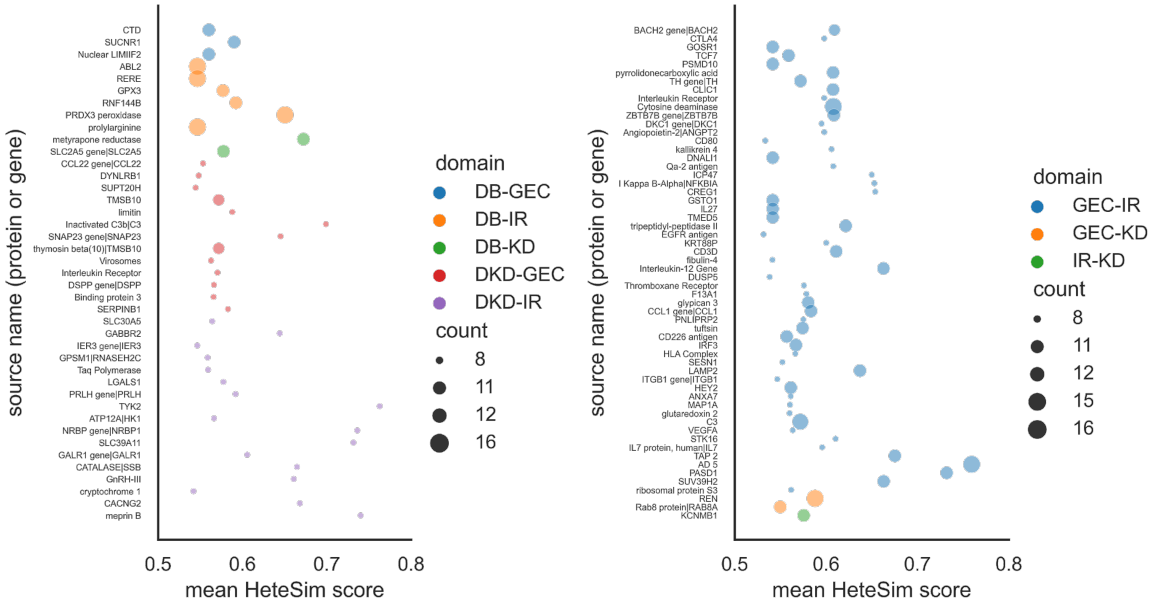

Figure 4. Bubble plot of source nodes identified by SemNet 2.0 at the intersection of each pairwise domain. Source nodes (genes or proteins) are shown on the vertical axis, each pairwise domain is denoted by a different bubble color, and the frequency (count) of each source node is distinguished by the bubble size. Source nodes are either genes or proteins. For clarity, the source nodes are presented in two plots (**left** and **right**). DB: diabetes domain. DKD: diabetic kidney disease domain. GEC: glomerular endothelial cells domain. IR: immune response domain. KD: kidney disease domain. Definitions of the source node abbreviations are provided in the Supplementary File S2.

2.2. Biological Themes Associated with Top 10% of Source Nodes

A unique functional ontology or biological process was mapped to each of the top 10% of source genes and proteins using the mouse genome informatics (MGI) term mapper [40,41]. A total of 99 genes and 117 proteins comprised the intersection of the DKD, IR, and GEC domains, which were mapped to 32 unique ontology IDs that described their biological functions. Figure 5 represents the identified genes and proteins and their respective mapped unique ontology IDs. These 21 unique ontologies (Figure 5) are also among the most common biological themes associated with DKD progression, a dysregulated immune response and inflammation, and the functional and structural integrity of endothelial cells. About 25% of the intersecting source nodes were found to be related to apoptotic function, 25% of the source nodes were related to the T-cell-mediated response, 22% of the intersecting source nodes were related to the cytokine response and inflammatory response, and 15% of the source nodes were related to growth factor signaling. Moreover, cell adhesion, chemotaxis, and the ERK/MAPK signaling pathway were also among the well-represented functional ontologies.

Further, the biological functions associated with these identified source genes (Figure 6) and source proteins (Figure 7) are summarized. The top 10% of source nodes were associated with biological processes, including the ERK/MAPK signaling cascade, the JAK/STAT pathway, NF-κB factor activity, growth factor signaling, the Wnt/β-catenin pathway, and the TLR pathway. Moreover, these genes and proteins were found to have multiple functional roles. The frequency of a biological process or function associated with a source node qualitatively describes the relative importance of the biological function. Figures 6 and 7 show heatmaps of the top 10% of identified source nodes and their respective biological processes specified as a unique ontology ID. The frequencies of these processes are represented by color bars on the right. These source nodes in the GEC, IR, and DKD domains play a crucial role in the immune response, the T-cell-mediated response, the cytokine response, apoptosis, and cell adhesion, among other critical biological func-

tions (Figures 6 and 7). The biological processes, including the TLR pathway, the immune response, apoptotic processes, calcium channel activity, and the response to cytokines, were relatively more prevalent processes associated with the identified source genes (Figure 6). The biological processes of the T-cell-mediated response, apoptotic processes, cell adhesion, the response to cytokines, chemotaxis, and growth factor signaling were relatively more prevalent processes associated with the identified source proteins (Figure 7).

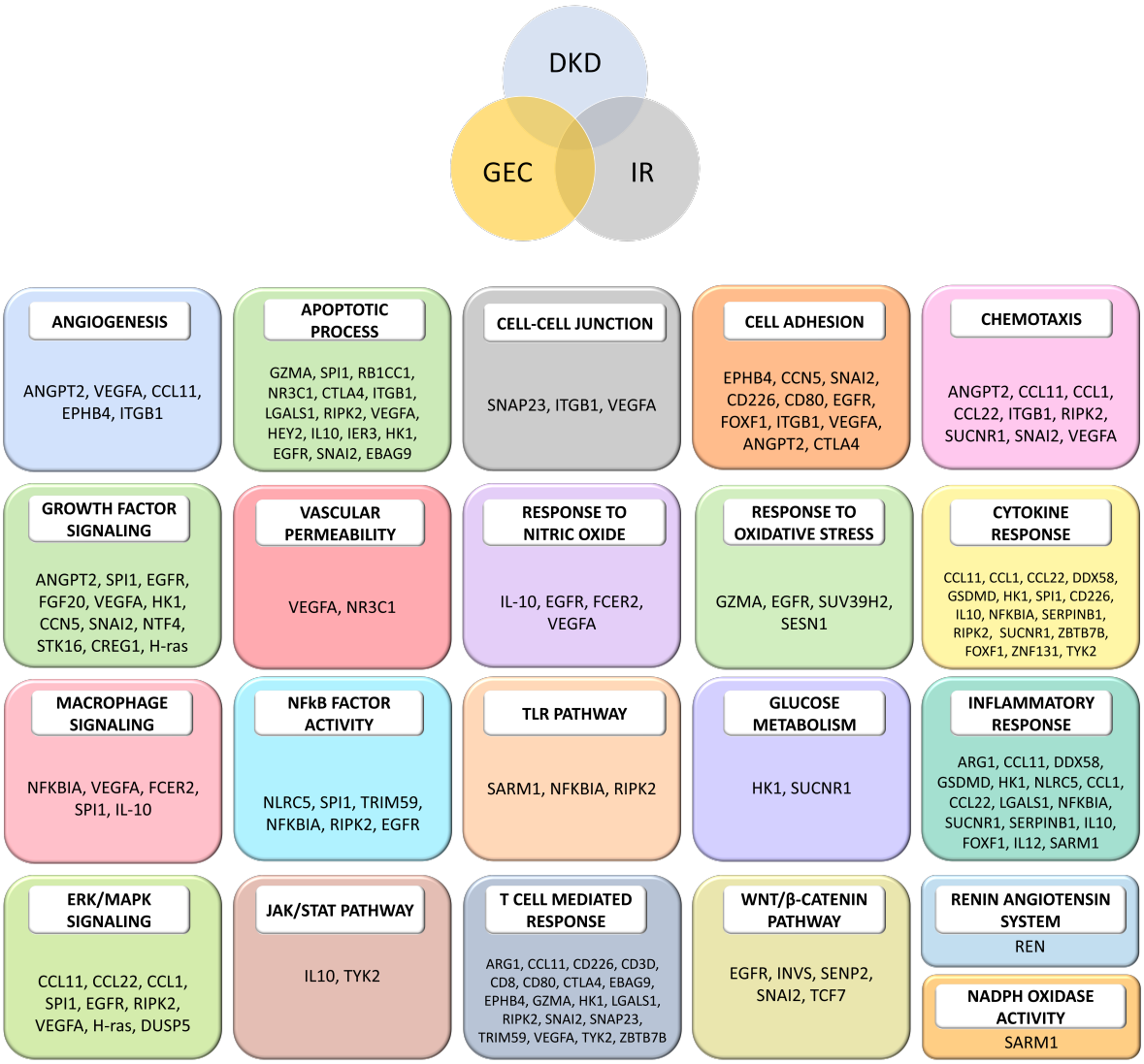

Figure 5. The top 10% of genes and proteins that intersect in the three domains—diabetic kidney disease (DKD), immune response (IR), and glomerular endothelial cells (GEC)—are grouped into their biological processes (block headers). NFkB: nuclear factor κ B. TLR: toll-like receptor. ERK: extracellular signal-regulated kinase. MAPK: mitogen-activated protein kinase. JAK: Janus kinase. STAT: signal transducer and activator of transcription. WNT: wingless/integrated. NADPH: nicotinamide adenine dinucleotide phosphate. The colors for the domains correspond to those used in Figure 9. The colors for the biological processes are arbitrary and are purely for aesthetic purposes. Definitions of the source node (gene and protein) abbreviations are provided in the Supplementary File S2.

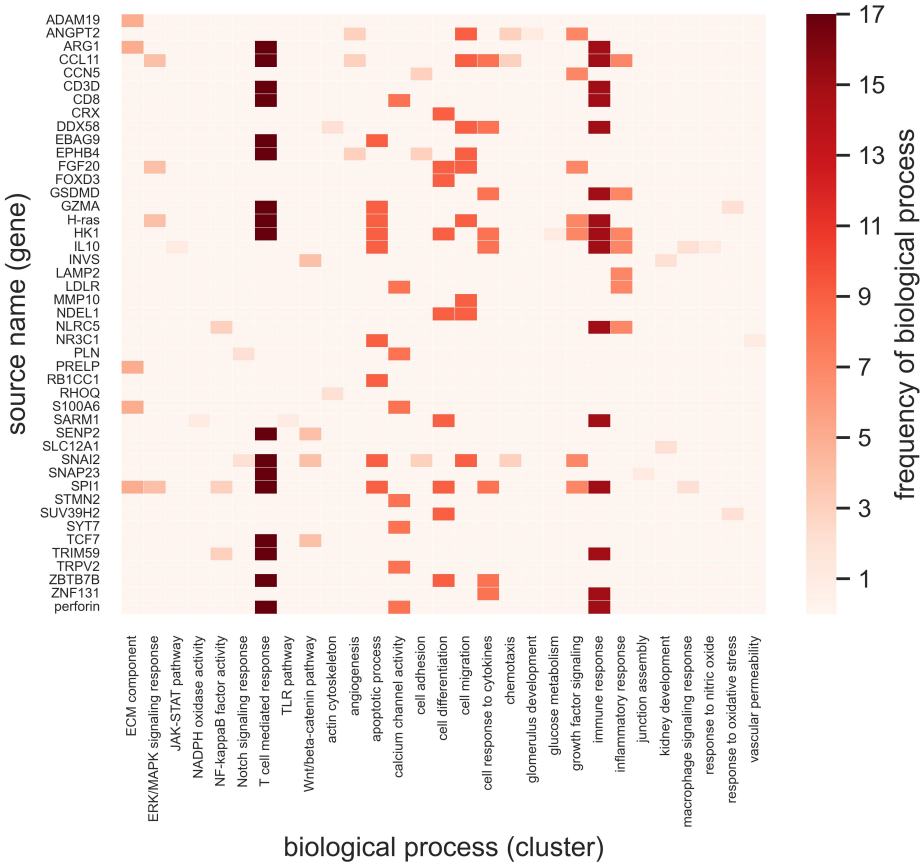

Figure 6. Source node (gene) names and their biological processes (term label) in the diabetic kidney disease (DKD), immune response (IR), and glomerular endothelial cells (GEC) domains. The frequency (term count) of each biological process (*x*-axis) is color-coded in the range of 0–17. ECM: extracellular matrix. ERK: extracellular signal-regulated kinase. MAPK: mitogen-activated protein kinase. JAK: Janus kinase. STAT: signal transducer and activator of transcription. NADPH: nicotinamide adenine dinucleotide phosphate. NF: nuclear factor. TLR: toll-like receptor. Wnt: wingless/integrated. Definitions of the source node abbreviations are provided in the Supplementary File S2.

The mean HeteSim scores of the identified genes (Figure A1) and proteins were provided and (Figure A2) mapped to their unique ontology IDs. The SPI1, SNAP23, STMN2, and ZNF131 genes and TYK2, NFKBIA, and CREG1 proteins had the highest mean HeteSim scores and were closely related to the user-specified targets. The CD3D, CD8, LAMP2, SUV39H2, TCF7, and ZBTB7B genes (Figure A3) frequently recurred in the cross-domain analysis and were associated with the T-cell-mediated response, the Wnt/β-catenin pathway, cell differentiation, the immune response, the response to cytokines, and the response to oxidative stress. The CCL1, CD226, HEY2, TAP2, and TMSB10 proteins (Figure A4) frequently occurred at the intersection of the DKD, GEC, and IR domains and were associated with biological processes like the JAK/STAT pathway, the Notch signaling response, the T-cell-mediated response, cell migration, the response to cytokines, chemotaxis, and the immune response.

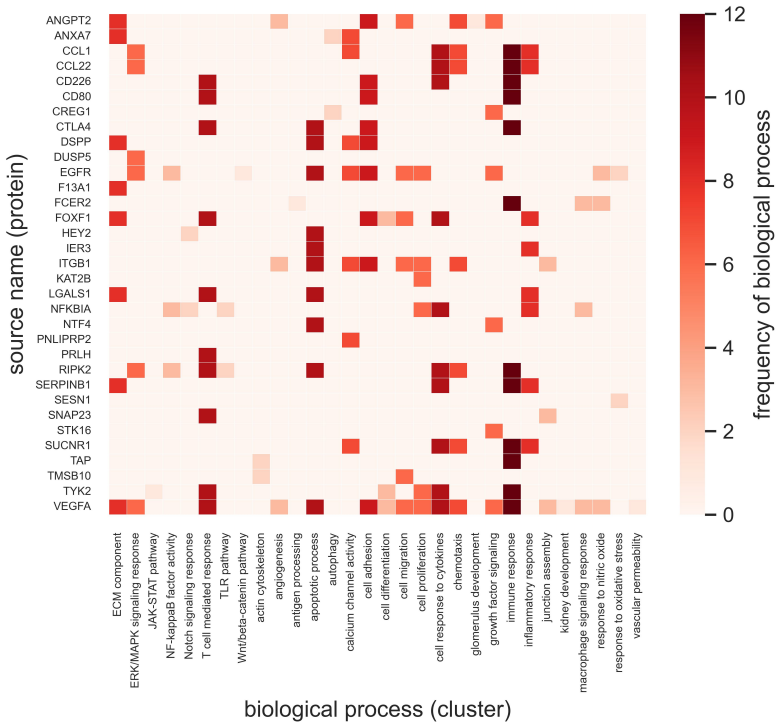

Figure 7. Source node (protein) names and their biological processes (term label) in the diabetic kidney disease (DKD), immune response (IR), and glomerular endothelial cells (GEC) domains. The frequency (term count) of each process (*x*-axis) is color-coded in the range 0–12. ECM: extracellular matrix. ERK: extracellular signal-regulated kinase. MAPK: mitogen-activated protein kinase. JAK: Janus kinase. STAT: signal transducer and activator of transcription. NADPH: nicotinamide adenine dinucleotide phosphate. NF: nuclear factor. TLR: toll-like receptor. Wnt: wingless/integrated. Definitions of the source node abbreviations are provided in the Supplementary File S2.

2.3. Interaction Network in Cytoscape

The interaction network between source genes or proteins and their respective regulatory relationships, generated in Cytoscape [38,39], is shown in Figure 8. The interaction network assembles the annotated outputs from the semantic network together. Figure 8 shows a positive regulatory relationship and negative regulatory relationship, through green arrows and red lines with a flat end, respectively, between source proteins and genes. The source genes and proteins are shown with a blue node and a yellow node, respectively. Upon analyzing the interaction network, it was observed that NF-κB, a transcription factor, is involved in numerous signaling events, including the inflammatory response; it is positively regulated by the PRDX3, EGFR, RIPK2, ABL2, and IRF3 genes and negatively regulated by the SPI1, TRIM59, NLRC5, and PSMD10 proteins and the NFKBIA gene. The VEGFA gene is central to the regulation of VEGF receptor 1 (VEGFR1), VEGF receptor 2 (VEGFR2), the adherens junction, nitric oxide, p38/MAPK signaling kinase, and ERK1/ERK2 signaling kinase. Vascular endothelial development and growth in endothelial cells heavily rely on nitric oxide [9,17,42], which is positively influenced by the VEGFA, EGFR, and FCER2 genes and negatively influenced by the the IL-10 protein. RIPK2 and DDX58, as well as GSDMD, RIPK2, and HK1, play positive roles in regulating pro-inflammatory cytokines like IL-6 and IL-1β. The anti-inflammatory IL-10 gene was correlated with the response to inflammatory cytokines, IL-1β, IL-6, IL-17, and IL-12. SUCNR1 is involved in both glucose homeostasis and macrophage activation and is a potential link in understanding glucose-mediated macrophage cell polarization [43]. The adherens junction proteins are responsible for regulating the endothelial cell–cell junction and vascular permeability in healthy and diseased states [9,44–47]. Our analysis identified that adherens junction proteins are positively regulated by the SNAP23 protein and negatively regulated

by the VEGFA gene. The NR3C1 protein encodes a glucocorticoid receptor and was found to play a role in reducing vascular permeability within endothelial cells. The FCER2 gene was one of the identified genes involved in positive regulation of macrophage activation.

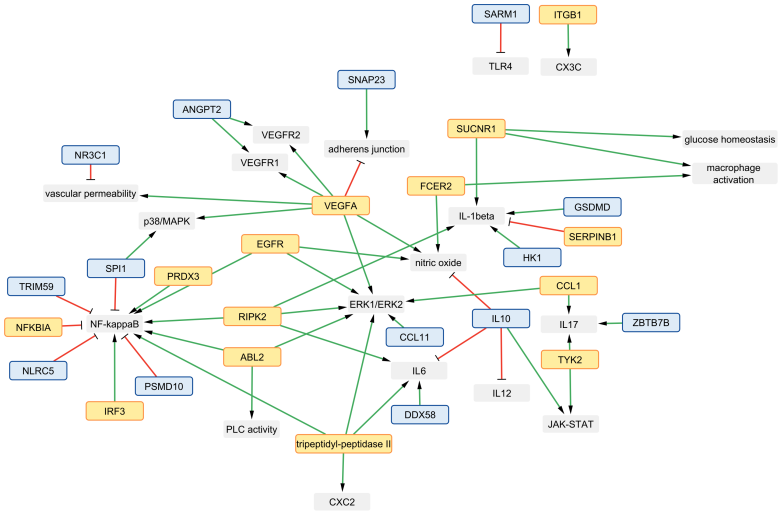

Figure 8. Regulatory relationship between top 10% of identified source nodes and signaling molecules generated in Cytoscape [38,39]. The grey blocks are signaling molecules/outcomes. The yellow blocks with an orange border are source proteins, and the blue blocks with a dark blue border are genes obtained from SemNet 2.0 analysis. A green arrow indicates a positive regulatory relationship between two nodes, and a red line with a flat end indicates a negative regulatory relationship. Signaling molecules (grey blocks) are defined as follows: TLR: toll-like receptor. NF-kappaB: nuclear factor-κB. ERK: extracellular signal-regulated kinase. MAPK: mitogen-activated protein kinase. JAK: Janus kinase. STAT: signal transducer and activator of transcription. VEGF: vascular endothelial growth factor. VEGFR: vascular endothelial growth factor receptor. CX3C: CX3C-chemokine. CX2C: CX2C-chemokine. IL: interleukin. PLC: phospholipase. Additional definitions for other sources are provided in the Supplementary File S2.

3. Discussion

The biomedical literature is a continuously growing repository of complex and deeply interconnected information. Despite powerful, user-friendly scientific databases, it is difficult for scientists and clinicians to extract useful information in their niche from these large and complex databases [34]. SemNet 2.0, an open-source literature-based discovery technique applied in this study, assists scientists and clinicians by leveraging the power of biomedical text mining to guide their research and development efforts. In this study, novel cross-domain text mining with SemNet 2.0 identified signaling molecules and pathways that are often studied in relation to diabetes, the immune response, kidney disease, and dysfunction of glomerular endothelial cells. The cross-domain analyses determined the relatedness between five pathological events by identifying significant source nodes that are mutually shared by these pathological events. The pairwise cross-domain analyses also determined the distribution of these source nodes across different semantic types (Figure 3).

3.1. Top-Ranked Intersecting GEC-IR-DKD Nodes

Among the top 10% of predicted source nodes, 77 source nodes were common among the GEC, IR, and DKD domains (Figure 4). The source genes or proteins with relatively high mean HeteSim scores were highly associated with and prevalent in the GEC, DKD, and IR domains (Figure 4). The observed intersection (Figure 5) indicates the importance of studying the synergistic interaction between the immune system and glomerular endothelial cells to better understand the early stage of DKD progression.

3.1.1. Top-Ranked Intersecting Cellular Functions and Signaling Processes

The top 10% of intersecting source nodes were associated with various cellular functions and cellular signaling responses. The cellular functions include angiogenesis, glucose metabolism, cell apoptosis, cell–cell junction integrity, and cell adhesion. The cellular signaling responses include growth factor signaling, the response to nitric oxide, the response to oxidative stress, the cytokine response, macrophage signaling, the TLR pathway, the T-cell-mediated response, NFκB factor activity, ERK/MAPK signaling, the JAK/STAT pathway, the WNT/β-catenin pathway, and NADPH oxidase activity. Several of these pathways are known to be implicated in GEC injury, inflammation, and fibrosis associated with DKD [6,8,11,17]. Similarly, the analyses suggested that the apoptotic processes, the response to cytokines, the T-cell-mediated response, the immune response, calcium channel activity, and growth factor signaling were the most frequent and active biological processes in DKD (Figures 6 and 7).

- Role of T-cells. This finding is consistent with other studies that report T-cells as being the most studied immune cells that infiltrate kidney tissues and trigger inflammatory responses in DKD [48–50]. Given the prevalent role of T-cells and the T-cell-mediated cytokine response in DKD [28,51,52], their highly ranked importance by the unsupervised ranking algorithm in SemNet 2.0 was expected.
- Role of calcium. Increased calcium channel activity exerts significant vascular and tubular effects on the kidneys, which leads to the enhancement of glomerular filtration rate (GFR) and renal blood flow (RBF) [53–55].
- Role of VEGF. Growth factor signaling via VEGF, fibroblast growth factors, transforming growth factor-β, and insulin-like growth factors in diabetes and diabetic kidney disease has been studied in detail [56,57]. VEGF is a potent angiogenic and vascular permeability factor and is responsible for endothelial cell proliferation and differentiation and increased permeability [9,58]. VEGF also maintains endothelial cell homeostasis, and a disturbance in basal VEGF levels is implicated in diabetes-related complications, including kidney disease [59]. Specifically, VEGF-A is associated with macrophage or monocyte differentiation, which suggests its role in the macrophage response in pathological conditions [9,42,57,58].
- Role of TGF. An increase or decrease in the production of transforming growth factor -β1 (TGF-β1) has been associated with diabetic nephropathy and retinopathy [56]. The expression of TGF-β1 is increased in endothelial cells, which, in turn, triggers the activation of TGF receptors, namely TGFBR2 and TGFBR3, on B lymphocytes, podocytes, glomerular endothelial cells, and mesangial cells, leading to epithelial–mesenchymal transition and fibrosis in the development of diabetic nephropathy [17,57,59]. Insulin-like growth factor-I is a naturally occurring single-chain polypeptide that has been widely used in the treatment of diabetic glomerular and renal tubular injuries [56,57,60].

3.1.2. Top-Ranked Intersecting Genes

The identified genes were involved in multiple functional roles. Highlights for some of the top-ranked genes are discussed below in the context of the literature.

- Role of succinate receptor 1 (SUCNR1). SUCNR1 is involved in both glucose homeostasis and macrophage activation. SUCNR1 is an extracellular receptor activated by succinate, and SUCNR1 accumulation in macrophages is known to activate the pro-inflammatory response [43]. Moreover, the role of SUCNR1 has been suggested in the development of fibrosis in diabetes mellitus and other diabetes-related complications such as diabetic retinopathy and metabolic syndrome [61]. SUCNR1 can serve as a potential link in understanding glucose-mediated macrophage cell polarization.
- Role of hexokinase 1 (HK1). The HK1 gene encodes a ubiquitous form of hexokinase, which localizes to the outside membrane of mitochondria. Mutations in HK1 have been associated with hemolytic anemia due to hexokinase deficiency. However, its role in DKD is more likely linked to dysregulated glucose metabolism. HK1 is also

associated with the cytokine response, the inflammatory response, and growth factor signaling [62].

- Role of ephrin. Ephrin receptors make up the largest subgroup of the receptor tyrosine kinase family, which have a key role in vascular regulation. SemNet 2.0 has previously highlighted the role of tyrosine kinase pathways in resistant hypertension [36]. The protein encoded by this gene binds to ephrin-B2 and plays an essential role in vascular development. Ephrin receptor EPHB4 is associated with angiogenesis in DKD, the immune response, and GEC. EPHB4 receptor interactions between endothelial cells and monocytes/macrophages are relevant for vascular development [57,59]. The inhibition of proteins in the ephrin B family prevents endothelial cell sprouting and initiates disorders in endothelial cell assembly [59].
- Role of serpin family B member 1 (SERPINB1). SERPINB1 is among the serpin protein families that are found in GEC. SERPINB1 acts primarily to protect the cells from proteases released into the cytoplasm during stress [63]. The results presented suggest SERPINB1's involvement in the inflammatory cytokine response. Previous research has suggested serpin proteins to be associated with macrophage motility as well [59].
- Role of integrins. Integrins regulate many biological processes, such as cell growth, migration, and signaling and cytokine activation, thereby contributing to inflammation and angiogenesis [64]. In the present study, integrin ITGB1 was associated with several cellular functions: angiogenesis, apoptosis, cell–cell junction integrity, and cell adhesion (Figure 5). Studies have also suggested combined treatment strategies through the inhibition of both ITGB and ITGA integrins to reduce macrophage filtration into the glomeruli [59,65]. Such compelling evidence suggests that ITGB1 may have the potential to be a clinical marker for the prognosis of glomerular diseases, immune cell infiltration, and glomerular endothelial viability [59].

3.1.3. Top-Ranked Intersecting Proteins

Among the top 10% of predicted source proteins, the TYK2, CREG1, NFKBIA, and SNAP23 proteins were highly associated with the user-specific target nodes based on the calculated mean HeteSim scores (Figure A2).

- Role of tyrosine kinsase 2 (TYK2). Previous studies found an association of TYK2 candidate with type 1 diabetes mellitus and a role of TYK2 in regulating apoptotic and pro-inflammatory pathways in pancreatic β-cells through modulation of the type I interferon signaling pathway [66,67]. Likewise, tyrosine kinase inhibitor drugs were previously predicted by SemNet 2.0 to be associated with hyperglycemia in patients who were not initially diabetic [29].
- Role of cellular repressor of E1A stimulated genes 1 (CREG1). CREG1 has been studied rigorously in relation to glucose uptake, renal dysfunction, angiogenesis, and diabetes-related comorbidity [68–70].
- Role of NFKBIA. NFKBIA regulates the activity of NFκB, which plays a role in processes such as the accumulation of advanced glycation end products and activation of the renin–angiotensin system pathways, protein kinase C, and oxidative stress in diabetic nephropathy [71].
- Role of synaptosome-associated protein 23 (SNAP23). Our analyses also identified SNAP23 associated with adherens junction assembly in correlation to GEC, the immune response, and DKD [72]. SPI1 and SNAP23 were genes highly associated with the immune response, cell differentiation, cell migration, the response to cytokines, and apoptotic processes (Figures 6 and A1). Some studies have previously identified SPI1 gene involvement in regulatory mechanisms in DKD, but this may need more experimental verification [73]. The SNAP23 gene is relatively abundant in the kidney and primarily involved in exocytosis [74]. SNAP23 has been shown to reduce proteinuria, reduce podocyte foot process fusion, and reduce endothelial cell damage upon the inhibition of SNAP23-mediated exocytosis [75].

3.2. Visualization of Literature-Based Discovery Network Predictions

Here the SemNet 2.0 analysis was conducted, and the prediction was visualized as an interaction network using Cytoscape [38,39]. Such a representation of source nodes using Cytoscape allowed for a straightforward interpretation of regulatory relationships in our generated data set. The comprehensive text mining analysis provided potential candidates involved in dysregulated signaling events (Figure 8) that can be used to address the limitations of our existing network model [9]. Relevant similarities were identified between the SemNet 2.0 findings and the previous network (Figure 1). Both networks share some common pathophysiological outcomes and signaling nodes involving TLR4, VEGF-A, VEGFR1, VEGFR2, IL-1β, IL-6, NO, PLC, NF-κB, the adherens junction, vascular permeability, and macrophage activation (Figures 1 and 8). SemNet 2.0 analysis is useful for recognizing proteins or genes that regulate these signaling nodes and/or pathophysiological events of interest. The predicted source genes and proteins can be further studied through network-based computational approaches and mechanistic modeling. The regulatory relationship visualized in Cytoscape could be a useful starting point to build a network-based model or identify interconnections between genes or proteins that may enable us to overcome the knowledge gaps or limitations of existing or published network models. There are various open-source and paid alternatives to Cytoscape, including Gephi [76], Tableau [77], NodeXL [78], and Neo4j [79], which can be used to visualize data obtained from text mining techniques. CompositeView has many similarities to Cytoscape and has been successfully implemented and customized to examine SemNet and SemNet 2.0 results [80]. A detailed comparison of the strengths and limitations of CompositeView against other similar software has been provided previously [80].

3.3. Comparing LBD Networks to Traditional Bioinformatics Networks

Including comprehensive biomedical literature in the cross-domain analysis provided breadth in the mechanistic understanding of disease progression, which is often not achievable through manual literature searches. This study demonstrated the compatibility and ease of use of the LBD tool SemNet 2.0, with various pieces of open-source bioinformatics software, to efficiently gather and assemble information that can be useful in the field of systems biology. Specifically, the present study outlined a process and case study by which to compare the similarities and knowledge gaps between signaling motifs obtained from SemNet 2.0 and previously published traditional signaling networks.

Another modality that this study's overall analysis can be qualitatively compared to is the pathway enrichment analysis of differentially expressed genes. Recent meta-analyses of gene expression datasets for diabetic nephropathy obtained the following top terms from their pathway enrichment analyses: immune system, extracellular matrix organization, hemostasis, signal transduction, and platelet activation by Hojjati et al. [81] and immune activation, T-cell activation, and cell adhesion by Zhong et al. [28]. While these are not exactly the terms our analysis yielded, they are related. For example, hemostasis broadly encompasses several of the cellular functions we listed in Section 3.1.1. Likewise, signal transduction lumps together many of the cellular signaling responses we itemized in Section 3.1.1. Immune system includes effects that resulted from our analysis, such as the response to cytokines, the T-cell-mediated response, and the immune response. Extracellular matrix organization is highly related to TGF and fibrosis. Among the previously identified diagnostic markers for DKD from Zhong et al. [28], tenascin C (TNC), tissue inhibitor metalloproteinase 1 (TIMP1), and tropomyosin 1 (TPM1) were also identified by the cross-domain analyses here (refer to raw data "combined.csv" in [82]). All 15 of the hub differentially expressed genes listed in Hojjati et al. [81] were also identified here (see raw data "combined.csv" in [82]).

3.4. Limitations and Future Directions

Biomedical text mining and similarity-based clustering analyses have their limitations. The clustering of these biomedical concepts or nodes based on similarity repre-

sents similarity in the patterns of associations with the user-specified target node. Thus, the similarity-based association of source to target depends on the amount and quality of literature data [83]. The implementation of additional link prediction algorithms with SemNet 2.0, as was performed by McCoy and colleagues to use SemNet 2.0 to predict COVID-19 drugs while the virus was new and was the subject of minimal studies [34], is one way to overcome this limitation. Regardless, a larger sample size of data reduces any bias from any lesser-quality publications. On the other hand, the user can control the loss of information when less evidence for a subject is available in the literature [31,83].

SemNet 2.0 is a methodology for ranking relevance and relatedness among nodes in a knowledge graph. This is similar to link prediction models, which enable the inference of novel relationships from existing edges and nodes in a knowledge graph. Performing link prediction leads to more nuanced search queries that build on SemNet simulation results. The Python library Pykeen has built-in functionality that can predict the head, relation, or tail (h, r, t) for an incomplete triple [84]. For example, given the incomplete triple (?, r, t), candidates for the head are scored and ranked based on KG embedding models such as RotatE, TransE, and ComplEx. Link prediction can be used for future research with the SemNet simulation results presented here.

Moreover, with advancements in single-cell transcriptomics, near-single-cell proteomics, and spatial metabolomics, there is emerging evidence and data for kidney tissues and infiltrating macrophages [57,85–87]. The widespread availability of such integrated and high-quality datasets [85] will enable better information gain when using LBD techniques. Future work could involve validating or comparing the findings from SemNet 2.0 to single-cell RNA sequencing data [57,85–87]. Integrating data within and across domains remains a big challenge due to heterogeneity. Recent and ongoing progress towards the collection, standardization, and integration of various metadata variables from data resources, including Kidney Tissue Atlas Ontology, Precision Medicine Metadata Ontology, and the Human Reference Atlas, has proven effective in identifying kidney-specific gene biomarkers and cell types [88].

Currently, a wide knowledge gap exists between biology and drug development, which results in sub-optimal treatment options against DKD [89]. Although glycemic control treatments are useful in the management of DKD to some extent, there is still potential for the discovery of new treatment strategies targeting inflammation, oxidative stress, fibrosis, and other pathological events [15,90]. Future applications may involve applying cross-domain analysis for the identification of pathological mechanisms, treatment strategies, and plausible hypotheses for DKD treatment and management. The proposed LBD technique can aid in bridging the knowledge gaps between DKD etiology and treatment. Applying LBD techniques to DKD research will enable faster processing of novel and actionable knowledge from vast, diverse, and seemingly disconnected fragments of information and the utilization of processed information towards treatment design [91]. The Kidney Precision Medicine Project (KPMP) has extensively contributed to the representation of kidney phenotype terms for acute and chronic kidney disease and increased the ability to improve personalized treatment. Future analysis of the KPMP data using LBD may be feasible in extracting information by profiling and integrating clinical, pathological, cellular, and molecular characteristics associated with the increasing pool of patients with specific diseases [88]. The highly associated genes or proteins observed at the intersection of the DKD, IR, and GEC domains could be used in future assessment through either experimental validation or a mathematical model.

4. Materials and Methods

The present study used advanced artificial intelligence-based text mining techniques to identify relevant signaling molecules and their relation to glucose-mediated inflammation in DKD. The general workflow is shown in Figure 2. First, SemNet 2.0 simulations were performed to identify the top-ranked nodes across multiple relevant domains using a knowledge graph of semantic relationships extracted from 33+ PubMed articles

(Figure 2A). The top-ranked nodes represent the most relevant concepts to DKD and its related pathophysiology. Next, using the top-ranked nodes, a GO network analysis was performed to obtain a functional map of the biological processes (Figure 2B). This functional map provides an intuitive means to summarize the thousands of relevant nodes into a format that enables mechanistic hypothesis formulation and testing. Finally, a regulatory interaction network was visualized in Cytoscape [38,39] using the top-ranked nodes and their corresponding semantic relationships from the SemNet knowledge graph (Figure 2C). In short, the regulatory interaction network provides the granular information necessary for the construction of subsequent biochemical models or protein interaction networks.

Specifically, SemNet 2.0 was used to perform a cross-domain analysis across the following five disease domains: diabetes (DB), kidney disease (KD), immune response (IR), glomerular endothelial cells (GEC), and DKD (Figure 9). The user-specified target nodes for each domain were chosen from observed interactions in the network model and published studies [9]. For instance, the binding of toll-like receptors (TLRs) is one of the key determinants of the immune response. Therefore, it was considered as one of the target nodes in this study. A complete list of target nodes is provided in the Supplementary File S1.

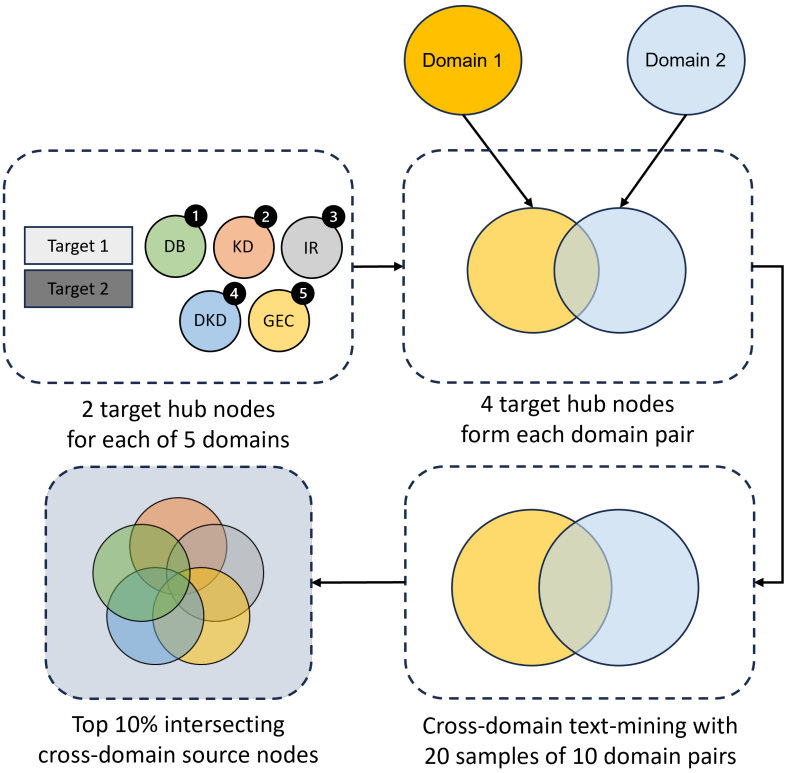

Figure 9. Workflow of cross-domain analyses in SemNet 2.0 performed to identify intersecting source nodes across five domains: diabetes (DB), kidney disease (KD), immune response (IR), diabetic kidney disease (DKD), and glomerular endothelial cells (GEC). Colored circles represent different domains, which are also numbered here. Illustrated in the first box, the first step was to randomly select two target nodes from each of the five domains. The specifics of the first step is illustrated in the next box, where each domain pair formed a combination of four target hub nodes. This information was provided as input to SemNet 2.0 to perform runs on each of these domain pairs. The process of generating four domain pairs was repeated 50 times, such that each of the ten unique domain pairs was sampled 20 times for a total of 200 SemNet simulations. This cross-domain text-mining process is represented by the third box. The final box illustrates that the SemNet simulations yielded intersecting cross-domain source nodes, and the top 10% of these source nodes were analyzed further. See Section 4.2 for details in the text.

4.1. SemNet 2.0

SemNet 2.0 [31] is a piece of open-source software that ingests publicly available text relationships from PubMed and the National Library of Medicine [33] to perform LBD tasks. SemNet 2.0 uses a heterogeneous semantic network to provide consistent and valuable categorization of all concepts represented in the UMLS metathesaurus, and the UMLS provides a universal ontology to relate concepts from the biomedical literature [92]. More information on the semantic networks and their types can be found in the UMLS reference manual [92]. SemNet 2.0 queries a biomedical knowledge graph composed of semantic triples extracted from PubMed's 33 million abstracts. The original SemNet was proposed in 2019 by Mitchell and Sedler [83] and was later optimized by Kirkpatrick and colleagues in 2022 [31]. Each semantic triple consists of a head, a relation, and a tail, where the head and tail entities are the nodes, and the relation is a directed edge. The underlying mathematical theory and computational implementation of SemNet 2.0 can be found in the original publication [31]. SemNet 2.0 is summarized below in the context of its use in the present study.

SemNet 2.0 is available as Python code and uses natural language processing to identify source nodes relevant to user-specified target nodes. The source nodes are the set of nodes that the target nodes share in common; that is, they are reachable within the search depth and metapath length, which are search parameters defined in the next section. Each node is a biomedical concept, as defined in the UMLS, with a type such as "disease or syndrome" (dsyn), "amino acid, peptide, or protein" (aapp), etc. There are 133 types and 54 relations. Each directed edge encodes a relation, such as treats, affects, inhibits, etc.

The user defines four inputs: the target nodes, source node types, search depth, and metapath length. Target nodes are the nodes of interest, and SemNet 2.0 queries surrounding nodes that are connected to those nodes. The source node types can be restricted to certain semantic types, such as dsyn or aapp. Search depth is the number of hops away from the source node (Figure 10). For a given target node, a search depth of 1 finds all adjacent nodes directly connected to it. It is ideal to increase the search depth to find novel results, as the connections of neighbors to target nodes are more prominent and commonly acknowledged in the scientific literature. Metapath length is the total path from a target to a source node. Multiple paths can be consolidated into a single metapath based on the types of source nodes. Hypothetically, an infinite number of paths can be used to connect a target to a source node, and the metapath length can add a constraint for identifying relevant or innovative pathways. Increasing both metrics can drastically expand the scope of the search. As such, the metapath length and search depth inputs are actually key model parameters that can change the model results.

SemNet 2.0 calculates a metric called HeteSim to quantify the relevance between a source node and target node [31]. HeteSim was developed to quantify relevance in heterogeneous networks [93]. There are two ways to calculate HeteSim: deterministic and randomized. The deterministic HeteSim was used in all SemNet 2.0 simulations in the present study to enhance accuracy at the expense of computational speed. HeteSim can be further characterized by an exact (deterministic) mean or approximate mean. The exact mean is found by aggregating the HeteSim scores of multiple target nodes to the same source node. The approximate mean has a performance advantage over the exact mean, especially for metapaths of greater length. For the simulations here, the exact HeteSim mean was used.

HeteSim is calculated by determining the cosine similarity between two probability vectors. Let x be defined as the path length between a given target node and a source node. HeteSim takes the middle layer of nodes or the nodes at an $x/2$ path length away from the target. From the target and source nodes, weights of 1 are distributed evenly across nodes. Each subsequent layer continually redistributes the weights until the middle layer is reached. A left probability vector and a right probability vector are generated from either side. The cosine similarity is calculated between them, which is the HeteSim score. When combining the results from multiple SemNet 2.0 simulations, the mean HeteSim scores are

normalized and the percentiles ranked to adjust for differences in node count, path count, etc. The normalization process allows the simulation results to be directly compared.

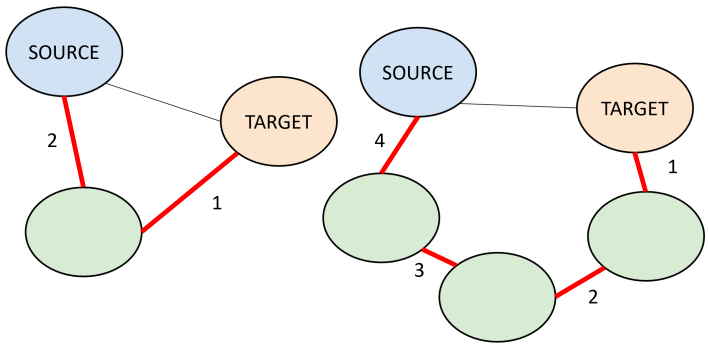

Figure 10. Two subgraphs with nodes (ovals) connected by edges (lines). The blue ovals are source nodes, the orange ovals are target nodes, and unlabeled green ovals represent nodes in the subgraph that are neither source nor target nodes. Red lines show the HeteSim metapath between target and source nodes, and the numbers near the lines provide running counts of the HeteSim metapath length that increases with each green node between the target and the source. In the subgraph on the left, the source node is at a depth of 1 away from the target, but the HeteSim metapath length is 2. On the right, the same source node is also at a depth of 1 away, but the HeteSim metapath length is 4.

As described above, SemNet 2.0 uses unsupervised learning rank aggregation to calculate the importance of a source node using the HeteSim score. Because SemNet 2.0 is an unsupervised model, there is no ground truth set of rankings by which to explicitly compare them. Domain knowledge provides an important layer of validation. The results of SemNet 2.0 studies have been successfully evaluated in COVID-19 [34,36], Parkinson's disease [35], chronic myeloid leukemia [29], and pediatric acute leukemia [37]. In fact, the majority of repurposed drug candidates recommended by SemNet early in the pandemic advanced to successful clinical trials as adjuvant therapies [94].

Note that the data for SemNet 2.0 simulations are based on processed semantic relationships from SemMedDB [31]. All data in the knowledge graph are available to be queried. However, the identified source nodes are limited by the search parameters, namely the search depth and metapath length parameters, as described above. The ranking results produced for a given query will remain consistent unless (1) a different (updated) version of the knowledge graph is deployed that contains additional new literature relationships or (2) the user changes the target node input(s) or specified search parameters.

SemNet 2.0 contains around 100 million semantic predictions (subject, object, predicate triples) extracted from PubMed articles. SemNet 2.0 separates the papers from their links and aggregates relation triples. SemMedDB is the basis of SemNet's corpus of entities and relations [33].

Other biomedical-domain knowledge graphs exist, such as KnowLife, which uses UMLS as its dictionary and 13 binary relations [95]. It encompasses entities in health and life sciences, built from various web sources, including online communities. PubMed Knowledge Graph comprises entities from 29 million PubMed abstracts, in addition to providing granular-level details about the articles themselves, including each author's educational background and affiliation history [96]. PrimeKG draws from only 20 resources, but it describes over 17,000 diseases and around 4 million relationships, with a focus on precision medicine analysis [97]. This knowledge graph allows the user to examine helpful indications and contraindications for drugs and how they impact disease progression. For drug repurposing, the model DREAMwalk—Drug Repurposing through Exploring Associations using Multi-layer random walk—uses "guilt-by-association" between drugs and diseases to generate hypothetical drug and disease node sequences, using a novel multi-layer approach that leverages node semantic neighbors [98]. Nonetheless, the unique properties of SemNet 2.0 made it the best choice for the present study.

4.2. SemNet 2.0 Simulations for DKD

Several relevant user-specified target nodes for each domain are given in the Supplementary File S1. Each of these targets was assumed to be a "hub node", which is a singular node that is well connected in a graph. These were based on prior domain knowledge. Networks of hub nodes enable improved cross-domain analysis by functionally increasing the search depth in areas of the knowledge graph of chief interest [29,36].

In general, a SemNet 2.0 run is performed on cross-domain pairs to find the source nodes in common between four target nodes. To sample the user-specified target nodes between pairs of domains, the first step was to randomly select two target hub nodes from each of the five domains, resulting in each domain pair having a combination of four target hub nodes (Figure 9). Five domains with two hub nodes each yielded ten hub nodes and four domain pairs. A SemNet run was performed on each of these domain pairs. The process of generating four domain pairs was repeated 50 times, such that each of the ten unique domain pairs was sampled 20 times (Figure 9) for a total of 200 SemNet simulations. The search depth was 2, and the metapath length was 3. Due to computational limitations, if the number of source nodes exceeded 1000, a random sample of the source nodes was taken such that no more than 1000 nodes had their HeteSim scores calculated and ranked. If the number of source nodes exceeded 1500 or was less than 10, the random combination of target nodes was deemed unproductive, and a new simulation with a different set of target nodes was performed.

Some simulations had several hundreds of nodes within the <1500 range, but they were limited by a certain target node (e.g., "Disorder of mineral metabolism"). Throughout this study, each node or concept's name and semantic type follow the UMLS ontology. SemNet 2.0 predicted source nodes associated with the target nodes and ranked these source nodes based on the optimized HeteSim similarity metric [31]. The association of the source nodes with user-specified target nodes in each pairwise domain was measured using a mean HeteSim score, calculated by averaging the HeteSim score of recurring nodes.

An exhaustive list of predicted source nodes was obtained at the intersection of each pairwise domain analysis. These identified source nodes were categorized into semantic types such as genes, proteins, and enzymes. These source nodes appeared more than once in each pairwise domain analysis. Each unique source node's occurrence was counted as the source node's frequency and denoted as count. A mean HeteSim score for each source node was generated and used for further analysis of the source nodes. The top 10% of normalized and highly ranked source nodes were aggregated from the simulations (Figure 9). The top 10% of the source nodes were chosen based on the overall predicted relevance using the mean HeteSim score.

4.3. Analysis of Source Nodes Revealed by SemNet 2.0

4.3.1. Functional Ontology Mapping

The biological process or function of the source nodes revealed by SemNet 2.0 simulations is not always available. Due to the vast simulation data, understanding the biological role of these source nodes through a literature survey may not be feasible. The SemNet 2.0 simulation generated a large amount of data. Thus, data-mining techniques were employed to map the source concepts to their biological functions. A common way of searching shared functions among genes is to incorporate the biological knowledge provided by biological ontologies [99–101]. The Gene Ontology resource is a major bioinformatics initiative that provides tools to annotate genes to their biological processes [102,103]. The mouse genome informatics (MGI) term mapper was used to provide ontologies or biological processes of the top 10% of genes or proteins [40,41]. Specifically, the list of identified source node names was input into the MGI Batch Query, and the respective functional ontology terms available were retrieved from the GO database. However, this method generated multiple functional classifications for a unique source node. Therefore, these functional ontologies were grouped by a unique ontology ID when duplicates or similar biological functions were listed. A unique

numeric label was generated for each unique ontology ID (term label), and the frequency of each ontology term mapped to a source node was recorded (term count).

4.3.2. Cytoscape

Cytoscape [38,39] is an open-source software platform for visualizing complex associations and integrating these with any attribute data. Cytoscape can build network models of interaction and tools for annotating and analyzing the connections or relationships in a data set [104]. The architecture is flexible, and the input data can include genes, proteins, chemicals, or enzymes [104]. Cytoscape [38,39] was used to generate a linked protein–protein network using the top 10% of identified source nodes and their mapped ontologies. The functional ontologies of each source node were analyzed. The functional ontologies that indicated a positive or negative relationship with a signaling molecule were considered. These relationships between source nodes and signaling molecules were selected by searching for specific keywords: "positive regulation" or "negative regulation". A negative regulation in GO database terms refers to any process that stops, prevents, or reduces the frequency or rate of covalent alteration of one or more amino acid residues within a protein [103]. A positive regulation suggests any process that activates or increases the frequency or rate of chemical reactions and pathways involving a protein [103]. A protein–protein interaction (PPI) file was created to store these source nodes as inputs and signaling molecules as outputs using a simple interaction file (sif) format. The sif-formatted file consisted of three main column entries: (1) inputs, (2) interaction type, and (3) outputs. The interaction type used in Cytoscape was protein–protein interaction. The interaction edges between source nodes and signaling molecules were based on the specified positive or negative relationship using a $+1$ or -1 relation index, respectively. The PPI file was used to create the interaction network in Cytoscape. Two additional column entries were added to the PPI file—the edge relation indices and input node type—to visually distinguish between regulatory relationships, genes, and proteins. More information on generating the PPI file and the format can be found in the Cytoscape user manual [39].

5. Conclusions

This LBD study comprehensively and efficiently identified and prioritized relevant signaling molecules and pathways associated with DKD. Cross-domain relationships were queried and ranked from 33+ million PubMed articles using SemNet 2.0. The SemNet 2.0 analysis yielded two valuable outcomes: (1) the relatedness between source genes or proteins that intersect the DKD-GEC-IR pathophysiology; (2) the creation of a protein/gene interaction network using objective, comprehensive LBD findings in place of an inherently more limited manual literature review. EPHB4, SERPINB1, ITGB1 TYK2, CREG1, NFKBIA, SPI1, and SNAP23 were among the most highly ranked concepts at the intersection of the GEC, IR, and DKD domains. These findings corroborate the relevance of studying the synergistic interaction between the immune system and glomerular endothelial cells to better understand the early stages of DKD progression. The results support the use of LBD to aid in the prioritization of multi-scalar pathological mechanisms and drug targets, the development of protein–protein interactions and biochemical models, the testing of hypotheses through experiments, and the advancement of biomedical decision-making.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/ijms25084503/s1.

Author Contributions: Conceptualization, K.P., C.S.M. and A.N.F.V.; Methodology, K.P., J.H.D. and C.S.M.; Software, J.H.D. and C.S.M.; Validation, K.P., J.H.D. and C.S.M.; Formal Analysis, K.P., J.H.D. and C.S.M.; Investigation, K.P., J.H.D., C.S.M. and A.N.F.V.; Resources, K.P., J.H.D., C.S.M. and A.N.F.V.; Data Curation, K.P., J.H.D., C.S.M. and A.N.F.V.; Writing—Original Draft Preparation, K.P.; Writing—Review and Editing, K.P., J.H.D., C.S.M. and A.N.F.V.; Visualization, K.P. and J.H.D.; Supervision, C.S.M. and A.N.F.V.; Project Administration, A.N.F.V.; Funding Acquisition, C.S.M. and A.N.F.V. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Institutes of Health grant (R35GM133763) to A.N.F.V., the National Science Foundation CAREER grant (2133411) to A.N.F.V., the National Science Foundation CAREER grant (1944247) to C.S.M., the National Institute of Health grant (R35GM152245) to C.S.M., the National Institute of Health grant (U19-AG056169) sub-awarded to C.S.M., and the Chan Zuckerberg Initiative grant (253558) to C.S.M.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data and analysis files for this study are provided in a repository [82], which is available here: https://github.com/ashleefv/DKD_CaseStudy_SemNet2 (accessed on 13 March 2024). The code leverages the open-source software SemNet 2.0 [31,32], which is available in the following repository: https://github.com/pathology-dynamics/semnet-2 (accessed on 9 January 2024).

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

Appendix A

Table A1 provides a summary of previously published mathematical models to study inter-disease relatedness for DKD and the kidney more generally.

Table A1. Summary of mathematical models used to study inter-disease relatedness for DKD and the kidney more generally.

Topic	Data Type	Model Type	Main Contribution	References
DKD kidney and immune crosstalk	In vitro	Logic-based ODE model	Identified critical interactions and chemical species associated with the pathophysiology of glucose and immune-mediated glomerular endothelial dysfunction using a network-based model.	[9]
Glomerular fibrosis in DKD	Mice, in vitro	Kinetic ODE model	Clarified mechanisms behind diabetic glomerular fibrosis and the reasons behind the failure of certain drugs to guide the creation of better diabetes-related kidney damage therapies.	[15]
Hyperglycemia-induced podocyte injury in DKD	Human	Kinetic ODE model	Integrated mathematical modeling, optimization, and sensitivity analyses to simulate the effects of glucose on the local RAS in podocytes in diabetic nephropathy to achieve physiologically plausible model outputs.	[18]
Podocytes in kidney diseases	In vitro	Multi-compartment ODE model	Analyzed feedforward motifs, spatial dynamics, and the roles of the PKA and MAPK pathways and provided insights into factors that could drive proliferation or differentiation in cultured podocytes.	[19]
Tubulointerstitial inflammation and fibrosis in lupus nephritis	Human	PDE model	Modeled disease progression and effects of dosing levels for novel therapeutics by simulating various levels of inhibition necessary to attenuate the inflammation to the fibrosis pathway.	[20]
Kidney and cardiac disorders	Human	Kinetic ODE model	Used QSP approach to understand the complex interactions between cardiac and renal functions in HF-rEF and the effects of pharmacological interventions such as ACEi and SGLT2i therapies.	[22]
Kidney and bone metabolism	Human	QSP model	Network topology enabled more effective targeting of key components and better prediction of perturbation effects.	[24]
CKD and bone	Mixed	QSP model	Conjugated individual existing cardiac and renal models to study pathophysiological interplay and disease relatedness. This model can be used in personalizing therapy and research settings and generating new hypotheses.	[25]
Multiple myeloma and kidney disorders	Human	Kinetic ODE model	Captured the qualitative behavior of the cell and protein populations and their interaction with the proximal tubule of the kidney, free light chains, renal fibroblasts, and myeloma cells. This model may support better patient prognosis in patients with multiple myeloma and renal impairment.	[23]

ACE: angiotensin-converting enzyme. CKD: chronic kidney disease. DKD: diabetic kidney disease. HF-rEF: heart failure with reduced ejection fraction. i: inhibitor drug classes. ODE: ordinary differential equations. PDE: partial differential equations. QSP: quantitative systems pharmacology. RAS: renin–angiotensin system. SGLT: sodium–glucose co-transporters.

Figures A1–A4 are provided to show further analysis results.

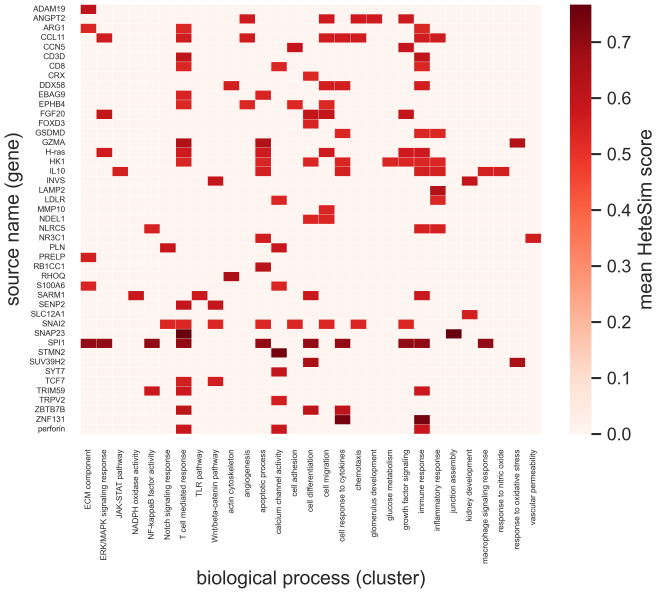

Figure A1. Source node (gene) names and their biological processes (term label) in the diabetic kidney disease (DKD), immune response (IR), and glomerular endothelial cells (GEC) domains. The mean HeteSim score of each gene from the SemNet 2.0 simulation is represented by the color bar on the right. ECM: extracellular matrix. ERK: extracellular signal-regulated kinase. MAPK: mitogen-activated protein kinase. JAK: Janus kinase. STAT: signal transducer and activator of transcription. NADPH: nicotinamide adenine dinucleotide phosphate. NF: nuclear factor. TLR: toll-like receptor. Wnt: wingless/integrated. Definitions of the source node abbreviations are provided in the Supplementary File S2.

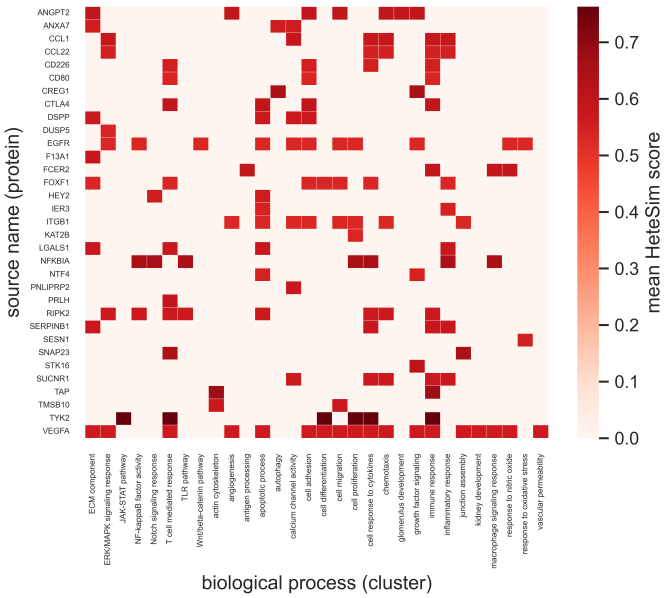

Figure A2. Source nodes (protein) names and their biological processes (term label) in the diabetic kidney disease (DKD), immune response (IR), and glomerular endothelial cells (GEC) domains. The mean HeteSim score of each protein from the SemNet 2.0 simulation is represented by the color bar on the right. ECM: extracellular matrix. ERK: extracellular signal-regulated kinase. MAPK: mitogen-activated protein kinase. JAK: Janus kinase. STAT: signal transducer and activator of transcription. NADPH: nicotinamide adenine dinucleotide phosphate. NF: nuclear factor. TLR: toll-like receptor. Wnt: wingless/integrated. Definitions of the source node abbreviations are provided in the Supplementary File S2.

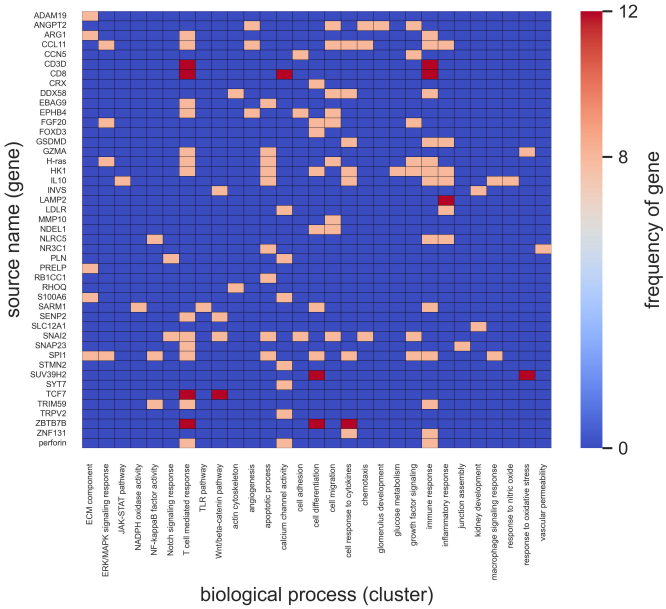

Figure A3. Source node (gene) names and their biological processes (term label) in the diabetic kidney disease (DKD), immune response (IR), and glomerular endothelial cells (GEC) domains. The frequency (count) of each source gene (*y*-axis) is color-coded. ECM: extracellular matrix. ERK: extracellular signal-regulated kinase. MAPK: mitogen-activated protein kinase. JAK: Janus kinase. STAT: signal transducer and activator of transcription. NADPH: nicotinamide adenine dinucleotide phosphate. NF: nuclear factor. TLR: toll-like receptor. Wnt: wingless/integrated. Definitions of the source node abbreviations are provided in the Supplementary File S2.

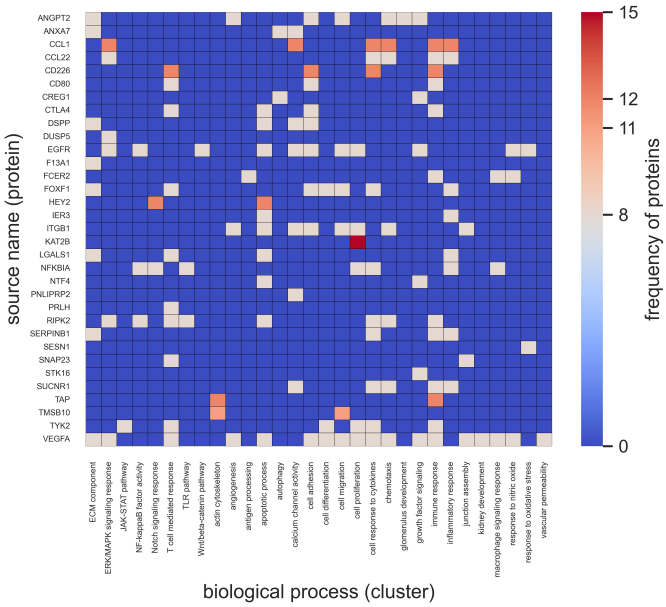

Figure A4. Source node (protein) names and their biological processes (term label) in the diabetic kidney disease (DKD), immune response (IR), and glomerular endothelial cells (GEC) domains. The frequency (count) of each source protein (*y*-axis) is color-coded. ECM: extracellular matrix. ERK: extracellular signal-regulated kinase. MAPK: mitogen-activated protein kinase. JAK: Janus kinase. STAT: signal transducer and activator of transcription. NADPH: nicotinamide adenine dinucleotide phosphate. NF: nuclear factor. TLR: toll-like receptor. Wnt: wingless/integrated. Definitions of the source node abbreviations are provided in the Supplementary File S2.

References

1. Hoogeveen, E.K. The Epidemiology of Diabetic Kidney Disease. *Kidney Dial.* **2022**, *2*, 433–442. [CrossRef]
2. Yang, J.; Liu, Z. Mechanistic Pathogenesis of Endothelial Dysfunction in Diabetic Nephropathy and Retinopathy. *Front. Endocrinol.* **2022**, *13*, 816400. [CrossRef] [PubMed]
3. Ghaderian, S.B.; Hayati, F.; Shayanpour, S.; Beladi Mousavi, S.S. Diabetes and End-Stage Renal Disease; A Review Article on New Concepts. *J. Renal. Inj. Prev.* **2015**, *4*, 28–33. [CrossRef]
4. Gembillo, G.; Ingrasciotta, Y.; Crisafulli, S.; Luxi, N.; Siligato, R.; Santoro, D.; Trifirò, G. Kidney Disease in Diabetic Patients: From Pathophysiology to Pharmacological Aspects with a Focus on Therapeutic Inertia. *Int. J. Mol. Sci.* **2021**, *22*, 4824. [CrossRef]
5. Wan, K.S.; Hairi, N.N.; Mustapha, F.; Mohd Yusoff, M.F.; Mat Rifin, H.; Ismail, M.; Moy, F.M.; Ahmad, N.A. Prevalence of Diabetic Kidney Disease and the Associated Factors among Patients with Type 2 Diabetes in a Multi-Ethnic Asian Country. *Sci. Rep.* **2024**, *14*, 7074. [CrossRef]
6. Tuttle, K.R.; Agarwal, R.; Alpers, C.E.; Bakris, G.L.; Brosius, F.C.; Kolkhof, P.; Uribarri, J. Molecular Mechanisms and Therapeutic Targets for Diabetic Kidney Disease. *Kidney Int.* **2022**, *102*, 248–260. [CrossRef] [PubMed]
7. Alicic, R.Z.; Rooney, M.T.; Tuttle, K.R. Diabetic Kidney Disease: Challenges, Progress, and Possibilities. *Clin. J. Am. Soc. Nephrol.* **2017**, *12*, 2032–2045. [CrossRef] [PubMed]
8. Mazzieri, A.; Porcellati, F.; Timio, F.; Reboldi, G. Molecular Targets of Novel Therapeutics for Diabetic Kidney Disease: A New Era of Nephroprotection. *Int. J. Mol. Sci.* **2024**, *25*, 3969. [CrossRef]
9. Patidar, K.; Ford Versypt, A.N. Logic-Based Modeling of Inflammatory Macrophage Crosstalk with Glomerular Endothelial Cells in Diabetic Kidney Disease. *bioRxiv* **2023**. [CrossRef]
10. Navarro-González, J.F.; Mora-Fernández, C.; Muros de Fuentes, M.; García-Pérez, J. Inflammatory molecules and pathways in the pathogenesis of diabetic nephropathy. *Nat. Rev. Nephrol.* **2011**, *7*, 327–340. [CrossRef]
11. Wang, N.; Zhang, C. Recent Advances in the Management of Diabetic Kidney Disease: Slowing Progression. *Int. J. Mol. Sci.* **2024**, *25*, 3086. [CrossRef] [PubMed]
12. Ayala, T.S.; Tessaro, F.H.G.; Jannuzzi, G.P.; Bella, L.M.; Ferreira, K.S.; Martins, J.O. High glucose environments interfere with bone marrow-derived macrophage inflammatory mediator release, the TLR4 pathway and glucose metabolism. *Sci. Rep.* **2019**, *9*, 11447. [CrossRef] [PubMed]
13. Tessaro, F.H.G.; Ayala, T.S.; Bella, L.M.; Martins, J.O. Macrophages from a type 1 diabetes mouse model present dysregulated PI3K/AKT, ERK 1/2 and SAPK/JNK levels. *Immunobiology* **2019**, *225*, 151879.
14. Thomas, H.Y.; Ford Versypt, A.N. Pathophysiology of Mesangial Expansion in Diabetic Nephropathy: Mesangial Structure, Glomerular Biomechanics, and Biochemical Signaling and Regulation. *J. Biol. Eng.* **2022**, *16*, 19. [CrossRef] [PubMed]
15. Thomas, H.Y.; Ford Versypt, A.N. A Mathematical Model of Glomerular Fibrosis in Diabetic Kidney Disease to Predict Therapeutic Efficacy. *bioRxiv* **2023**. [CrossRef]
16. Lassén, E.; Daehn, I.S. Molecular Mechanisms in Early Diabetic Kidney Disease: Glomerular Endothelial Cell Dysfunction. *Int. J. Mol. Sci.* **2020**, *21*, 9456. [CrossRef]
17. Hu, S.; Hang, X.; Wei, Y.; Wang, H.; Zhang, L.; Zhao, L. Crosstalk Among Podocytes, Glomerular Endothelial Cells and Mesangial Cells in Diabetic Kidney Disease: An Updated Review. *Cell Commun. Signal.* **2024**, *22*, 136. [CrossRef] [PubMed]
18. Pilvankar, M.R.; Higgins, M.A.; Ford Versypt, A.N. Mathematical Model for Glucose Dependence of the Local Renin–Angiotensin System in Podocytes. *Bull. Math. Biol.* **2018**, *80*, 880–905. [CrossRef]
19. Azeloglu, E.U.; Hardy, S.V.; Eungdamrong, N.J.; Chen, Y.; Jayaraman, G.; Chuang, P.Y.; Fang, W.; Xiong, H.; Neves, S.R.; Jain, M.R.; et al. Interconnected Network Motifs Control Podocyte Morphology and Kidney Function. *Sci. Signal.* **2014**, *7*, ra12. [CrossRef]
20. Hao, W.; Rovin, B.H.; Friedman, A. Mathematical Model of Renal Interstitial Fibrosis. *Proc. Natl. Acad. Sci. USA* **2014**, *111*, 14193–14198. [CrossRef]
21. Ma, T.K.; Kam, K.K.; Yan, B.P.; Lam, Y.Y. Renin-Angiotensin-Aldosterone System Blockade for Cardiovascular Diseases: Current Status. *Br. J. Pharmacol.* **2010**, *160*, 1273–1292. [CrossRef] [PubMed]
22. Yu, H.; Basu, S.; Hallow, K.M. Cardiac and Renal Function Interactions in Heart Failure with Reduced Ejection Fraction: A Mathematical Modeling Analysis. *PLoS Comput. Biol.* **2020**, *16*, e1008074. [CrossRef] [PubMed]
23. Walk, J.C.; Ayati, B.P.; Holstein, S.A. Modeling the Effects of Multiple Myeloma on Kidney Function. *Sci. Rep.* **2019**, *9*, 1726. [CrossRef] [PubMed]
24. Peterson, M.C.; Riggs, M.M. A Physiologically Based Mathematical Model of Integrated Calcium Homeostasis and Bone Remodeling. *Bone* **2010**, *46*, 49–63. [CrossRef] [PubMed]
25. Gaweda, A.E.; McBride, D.E.; Lederer, E.D.; Brier, M.E. Development of a Quantitative Systems Pharmacology Model of Chronic Kidney Disease: Metabolic Bone Disorder. *Am. J. Physiol. Ren. Physiol.* **2020**, *320*, F203–F211. [CrossRef]
26. Zhou, A.X.; Jeansson, M.; He, L.; Wigge, L.; Tonelius, P.; Tati, R.; Cederblad, L.; Muhl, L.; Uhrbom, M.; Liu, J.; et al. Renal Endothelial Single-Cell Transcriptomics Reveals Spatiotemporal Regulation and Divergent Roles of Differential Gene Transcription and Alternative Splicing in Murine Diabetic Nephropathy. *Int. J. Mol. Sci.* **2024**, *25*, 4320. [CrossRef]
27. Shao, X.; Shi, Y.; Wang, Y.; Zhang, L.; Bai, P.; Wang, J.; Aniwan, A.; Lin, Y.; Zhou, S.; Yu, P. Single-Cell Sequencing Reveals the Expression of Immune-Related Genes in Macrophages of Diabetic Kidney Disease. *Inflammation* **2024**, *47*, 227–243. [CrossRef] [PubMed]

28. Zhong, M.; Zhu, E.; Li, N.; Gong, L.; Xu, H.; Zhong, Y.; Gong, K.; Jiang, S.; Wang, X.; Fei, L.; et al. Identification of Diagnostic Markers Related to Oxidative Stress and Inflammatory Response in Diabetic Kidney Disease by Machine Learning Algorithms: Evidence from Human Transcriptomic Data and Mouse Experiments. *Front. Endocrinol.* **2023**, *14*, 1134325. [CrossRef]
29. Mehra, N.; Varmeziar, A.; Chen, X.; Kronick, O.; Fisher, R.; Kota, V.; Mitchell, C.S. Cross-Domain Text Mining to Predict Adverse Events from Tyrosine Kinase Inhibitors for Chronic Myeloid Leukemia. *Cancers* **2022**, *14*, 4686. [CrossRef]
30. Swanson, D.R. Fish Oil, Raynaud's Syndrome, and Undiscovered Public Knowledge. *Perspect. Biol. Med.* **1986**, *30*, 7–18. [CrossRef]
31. Kirkpatrick, A.; Onyeze, C.; Kartchner, D.; Allegri, S.; Nakajima An, D.; McCoy, K.; Davalbhakta, E.; Mitchell, C.S. Optimizations for Computing Relatedness in Biomedical Heterogeneous Information Networks: SemNet 2.0. *Big Data Cogn. Comput.* **2022**, *6*, 27. [CrossRef] [PubMed]
32. Kirkpatrick, A.; Onyeze, C.; Kartchner, D.; Allegri, S.; Nakajima An, D.; McCoy, K.; Davalbhakta, E.; Mitchell, C.S. SemNet 2.0. Available online: https://github.com/pathology-dynamics/semnet-2 (accessed on 9 January 2024).
33. Kilicoglu, H.; Shin, D.; Fiszman, M.; Rosemblat, G.; Rindflesch, T.C. SemMedDB: A PubMed-Scale Repository of Biomedical Semantic Predications. *Bioinformatics* **2012**, *28*, 3158–3160. [CrossRef]
34. McCoy, K.; Gudapati, S.; He, L.; Horlander, E.; Kartchner, D.; Kulkarni, S.; Mehra, N.; Prakash, J.; Thenot, H.; Vanga, S.V.; et al. Biomedical Text Link Prediction for Drug Discovery: A Case Study with COVID-19. *Pharmaceutics* **2021**, *13*, 794. [CrossRef] [PubMed]
35. Tandra, G.; Yoone, A.; Mathew, R.; Wang, M.; Hales, C.M.; Mitchell, C.S. Literature-based Discovery Predicts Antihistamines are a Promising Repurposed Adjuvant Therapy for Parkinson's disease. *Int. J. Mol. Sci.* **2023**, *24*, 12339. [CrossRef] [PubMed]
36. Kartchner, D.; McCoy, K.; Dubey, J.; Zhang, D.; Zheng, K.; Umrani, R.; Kim, J.J.; Mitchell, C.S. Literature-Based Discovery to Elucidate the Biological Links between Resistant Hypertension and COVID-19. *Biology* **2023**, *12*, 1269. [CrossRef]
37. Al-Hussaini, I.; White, B.; Varmeziar, A.; Mehra, N.; Sanchez, M.; Lee, J.; DeGroote, N.P.; Miller, T.P.; Mitchell, C.S. An Interpretable Machine Learning Framework for Rare Disease: A Case Study to Stratify Infection Risk in Pediatric Leukemia. *J. Clin. Med.* **2024**, *13*, 1788. [CrossRef] [PubMed]
38. Shannon, P.; Markiel, A.; Ozier, O.; Baliga, N.; Wang, J.T.; Ramage, D.; Amin, N.; Schwikowski, B.; Ideker, T. Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **2003**, *13*, 2498–2504. [CrossRef] [PubMed]
39. Cytoscape Contributors. Cytoscape User Manual. Available online: https://manual.cytoscape.org (accessed on 9 January 2024).
40. Baldarelli, R.; Smith, C.; Finger, J.; Hayamizu, T.; McCright, I.; Xu, J.; Shaw, D.; Beal, J.; Blodgett, O.; Campbell, J.; et al. The Mouse Gene Expression Database (GXD): 2021 Update. *Nucleic Acids Res.* **2020**, *49*, D924–D931. [CrossRef]
41. Blake, J.; Baldarelli, R.; Kadin, J.; Richardson, J.; Smith, C.; Bult, C.; Anagnostopoulos, A.; Beal, J.; Bello, S.; Blodgett, O.; et al. Mouse Genome Database (MGD): Knowledgebase for Mouse–Human Comparative Biology. *Nucleic Acids Res.* **2020**, *49*, D981–D987. [CrossRef]
42. Wu, Q.; Finley, S. Mathematical Model Predicts Effective Strategies to Inhibit VEGF-eNOS Signaling. *J. Clin. Med.* **2020**, *9*, 1255. [CrossRef]
43. Keiran, N.; Ceperuelo-Mallafré, V.; Calvo, E.; Hernández-Alvarez, M.I.; Ejarque, M.; Núñez-Roa, C.; Horrillo, D.; Maymó-Masip, E.; Rodríguez, M.M.; Fradera, R.; et al. SUCNR1 Controls an Anti-Inflammatory Program in Macrophages to Regulate the Metabolic Response to Obesity. *Nat. Immunol.* **2019**, *20*, 581–592. [CrossRef] [PubMed]
44. Stan, R.V.; Tse, D.; Deharvengt, S.J.; Smits, N.C.; Xu, Y.; Luciano, M.R.; McGarry, C.l.; Buitendijk, M.; Nemani, K.V.; Elgueta, R.; et al. The Diaphragms of Fenestrated Endothelia: Gatekeepers of Vascular Permeability and Blood Composition. *Dev. Cell* **2012**, *23*, 1203–1218. [CrossRef]
45. Finch, N.C.; Neal, C.R.; Welsh, G.I.; Foster, R.R.; Satchell, S.C. The Unique Structural and Functional Characteristics of Glomerular Endothelial Cell Fenestrations and Their Potential as a Therapeutic Target in Kidney Disease. *Am. J. Physiol. Ren. Physiol.* **2023**, *325*, F465–F478. [CrossRef] [PubMed]
46. Becker, S.; Wandel, E.; Wobus, M.; Schneider, R.; Amasheh, S.; Sittig, D.; Kerner, C.; Naumann, R.; Hamann, J.; Aust, G. Overexpression of CD97 in Intestinal Epithelial Cells of Transgenic Mice Attenuates Colitis by Strengthening Adherens Junctions. *PLoS ONE* **2010**, *5*, e8507. [CrossRef]
47. Satchell, S.; Tooke, J. What is the Mechanism of Microalbuminuria in Diabetes: A Role for the Glomerular Endothelium? *Diabetologia* **2008**, *51*, 714–725. [CrossRef]
48. Cai, F.; Zhou, X.; Jia, Y.; Yao, W.; Lv, J.; Liu, G.; Yang, L. Identification of Key Genes of Human Advanced Diabetic Nephropathy Independent of Proteinuria by Transcriptome Analysis. *Biomed. Res. Int.* **2020**, *2020*, 7283581. [CrossRef]
49. Lim, A.K.H.; Tesch, G.H. Inflammation in Diabetic Nephropathy. *Mediat. Inflamm.* **2012**, *2012*, 146154. [CrossRef] [PubMed]
50. Klessens, C.Q.F.; Zandbergen, M.; Wolterbeek, R.; Bruijn, J.A.; Rabelink, T.J.; Bajema, I.M.; IJpelaar, D.H.T. Macrophages in Diabetic Nephropathy in Patients with Type 2 Diabetes. *Nephrol. Dial. Transplant.* **2017**, *32*, 1322–1329. [CrossRef]
51. Liu, Y.; Lv, Y.; Zhang, T.; Huang, T.; Lang, Y.; Sheng, Q.; Liu, Y.; Kong, Z.; Gao, Y.; Lu, S.; et al. T cells and their Products in Diabetic Kidney Disease. *Front. Immunol.* **2023**, *14*, 1084448. [CrossRef]
52. Roep, B.O. The Role of T-cells in the Pathogenesis of Type 1 Diabetes: From Cause to Cure. *Diabetologia* **2003**, *46*, 305–321. [CrossRef]
53. Chan, L.; Schrier, R.W. Effects of Calcium Channel Blockers on Renal Function. *Annu. Rev. Med.* **1990**, *41*, 289–302. [CrossRef] [PubMed]
54. Hao, Y.; Wang, Z.; Frimpong, F.; Chen, X. Calcium-Permeable Channels and Endothelial Dysfunction in Acute Lung Injury. *Curr. Issues Mol. Biol.* **2022**, *44*, 2217–2229. [CrossRef]
55. Dalal, P.J.; Muller, W.A.; Sullivan, D.P. Endothelial Cell Calcium Signaling During Barrier Function and Inflammation. *Am. J. Pathol.* **2020**, *190*, 535–542. [CrossRef]

56. van Nieuwenhoven, F.A.; Jensen, L.J.N.; Flyvbjerg, A.; Goldschmeding, R. Imbalance of Growth Factor Signalling in Diabetic Kidney Disease: Is Connective Tissue Growth Factor (CTGF, CCN2) the Perfect Intervention Point? *Nephrol. Dial. Transplant.* **2005**, *20*, 6–10. [CrossRef] [PubMed]
57. Song, Y.; Liu, Y.; Guo, F.; Zhao, L.; Qin, G. Single-Cell RNA Sequencing Reveals Rac1 Involvement in Macrophages Efferocytosis in Diabetic Kidney Disease. *Inflammation* **2023**. [CrossRef]
58. Majumder, S.; Advani, A. VEGF and the Diabetic Kidney: More than too much of a good thing. *J. Diabetes Complicat.* **2017**, *31*, 273–279. [CrossRef] [PubMed]
59. Li, T.; Shen, K.; Li, J.; Leung, S.W.S.; Zhu, T.; Shi, Y. Glomerular Endothelial Cells are the Coordinator in the Development of Diabetic Nephropathy. *Front. Med.* **2021**, *8*, 655639. [CrossRef]
60. Mohamed-Ali, V.; Pinkney, J. Therapeutic Potential of Insulin-Like Growth Factor-1 in Patients with Diabetes Mellitus. *Treat. Endocrinol.* **2002**, *1*, 399–410. [CrossRef]
61. Ariza, A.C.; Deen, P.M.; Robben, J.H. The Succinate Receptor as a Novel Therapeutic Target for Oxidative and Metabolic Stress-Related Conditions. *Front. Endocrinol.* **2012**, *3*, 22. [CrossRef] [PubMed]
62. Codocedo, J.F.; Landreth, G.E. The Intersection of Metabolism and Inflammation Is Governed by the Intracellular Topology of Hexokinases and the Metabolic Fate of Glucose. *Immunometabolism* **2022**, *4*, e00011. [CrossRef]
63. Torriglia, A.; Martin, E.; Jaadane, I. The Hidden Side of SERPINB1/Leukocyte Elastase Inhibitor. *Semin. Cell Dev. Biol.* **2017**, *62*, 178–186. [CrossRef] [PubMed]
64. Mezu-Ndubuisi, O.J.; Maheshwari, A. The Role of Integrins in Inflammation and Angiogenesis. *Pediatr. Res.* **2021**, *89*, 1619–1626. [CrossRef] [PubMed]
65. Wu, X.; Tiwari, A.K.; Issekutz, T.B.; Lefkowith, J.B. Differing Roles of CD18 and VLA-4 in Leukocyte Migration/Activation during Anti-GBM Nephritis. *Kidney Int.* **1996**, *50*, 462–472. [CrossRef] [PubMed]
66. Pellenz, F.; Dieter, C.; Duarte, G.; Canani, L.; Souza, B.; Crispim, D. The rs2304256 Polymorphism in TYK2 Gene Is Associated with Protection for Type 1 Diabetes Mellitus. *Diabetes Metab. J.* **2021**, *45*, 899. [CrossRef] [PubMed]
67. Mori, H.; Takahashi, H.; Mine, K.; Higashimoto, K.; Inoue, K.; Kojima, M.; Kuroki, S.; Eguchi, T.; Ono, Y.; Inuzuka, S.; et al. TYK2 Promoter Variant Is Associated with Impaired Insulin Secretion and Lower Insulin Resistance in Japanese Type 2 Diabetes Patients. *Genes* **2021**, *12*, 400. [CrossRef]
68. Yan, C.; Fang, P.; Zhang, H.; Tao, J.; Tian, X.; Li, Y.; Zhang, J.; Sun, M.; Li, S.; Wang, H.; et al. CREG1 Promotes Angiogenesis and Neovascularization. *Front. Biosci.* **2014**, *19*, 1151–1161. [CrossRef] [PubMed]
69. Goto, A.; Endo, Y.; Yamashita, H. CREG1 Stimulates AMPK Phosphorylation and Glucose Uptake in Skeletal Muscle Cells. *Biochem. Biophys. Res. Commun.* **2022**, *641*, 162–167. [CrossRef] [PubMed]
70. Ghobrial, G.; Araujo, L.; Jinwala, F.; Li, S.; Lee, L. The Structure and Biological Function of CREG. *Front. Cell Dev. Biol.* **2018**, *6*, 136. [CrossRef] [PubMed]
71. Yang, M.; Zhang, C. The Role of Innate Immunity in Diabetic Nephropathy and Their Therapeutic Consequences. *J. Pharm. Anal.* **2024**, *14*, 39–51. [CrossRef]
72. Guo, Z.; Neilson, L.J.; Zhong, H.; Murray, P.S.; Zanivan, S.; Zaidel-Bar, R. E-Cadherin Interactome Complexity and Robustness Resolved by Quantitative Proteomics. *Sci. Signal.* **2014**, *7*, rs7. [CrossRef]
73. Li, B.; Zhao, X.; Xie, W.; Hong, Z.; Zhang, Y. Integrative Analyses of Biomarkers and Pathways for Diabetic Nephropathy. *Front. Genet.* **2023**, *14*, 1128136. [CrossRef] [PubMed]
74. Inoue, T.; Nielsen, S.; Mandon, B.; Terris, J.; Kishore, B.K.; Knepper, M.A. SNAP-23 in Rat Kidney: Colocalization with Aquaporin-2 in Collecting Duct Vesicles. *Am. J. Physiol. Ren. Physiol.* **1998**, *275*, F752–F760. [CrossRef] [PubMed]
75. Caster, D.J.; Korte, E.A.; Tan, M.; Barati, M.T.; Tandon, S.; Creed, T.M.; Salant, D.J.; Hata, J.L.; Epstein, P.N.; Huang, H.; et al. Neutrophil Exocytosis Induces Podocyte Cytoskeletal Reorganization and Proteinuria in Experimental Glomerulonephritis. *Am. J. Physiol. Ren. Physiol.* **2018**, *315*, F595–F606. [CrossRef] [PubMed]
76. Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. In Proceedings of the International AAAI Conference on Web and Social Media, Oxford, UK, 19 March 2009; Volume 3, pp. 361–362. [CrossRef]
77. Batt, S.; Grealis, T.; Harmon, O.; Tomolonis, P. Learning Tableau: A Data Visualization Tool. *J. Econ. Ed.* **2020**, *51*, 317–328. [CrossRef]
78. Smith, M.; Ceni, A.; Milic-Frayling, N.; Shneiderman, B.; Mendes Rodrigues, E.; Leskovec, J.; Dunne, C. NodeXL: A Free and Open Network Overview, Discovery and Exploration Add-in for Excel 2007/2010/2013/2016. Available online: https://www.smrfoundation.org (accessed on 9 January 2024).
79. Neo4j, Inc. Neo4j-The World's Leading Graph Database. Available online: http://neo4j.org/ (accessed on 9 January 2024).
80. Allegri, S.A.; McCoy, K.; Mitchell, C.S. CompositeView: A Network-Based Visualization Tool. *Big Data Cogn. Comput.* **2022**, *6*, 66. [CrossRef] [PubMed]
81. Hojjati, F.; Roointan, A.; Gholaminejad, A.; Eshraghi, Y.; Gheisari, Y. Identification of Key Genes and Biological Regulatory Mechanisms in Diabetic Nephropathy: Meta-Analysis of Gene Expression Datasets. *Nefrologia* **2023**, *43*, 575–586. [CrossRef] [PubMed]
82. Patidar, K.; Deng, J.H.; Mitchell, C.S.; Ford Versypt, A.N. DKD_CaseStudy_SemNet2. Available online: https://github.com/ashleefv/DKD_CaseStudy_SemNet2 (accessed on 13 March 2024). [CrossRef]
83. Sedler, A.; Mitchell, C.S. SemNet: Using Local Features to Navigate the Biomedical Concept Graph. *Front. Bioeng. Biotechnol.* **2019**, *7*, 156. [CrossRef]

84. Ali, M.; Berrendorf, M.; Hoyt, C.T.; Vermue, L.; Sharifzadeh, S.; Tresp, V.; Lehmann, J. PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings. *J. Mach. Learn. Res* **2021**, *22*, 1–6.
85. Stewart, A.P.; Clatworthy, M.R. Single-Cell Genomics Sheds Light on Kidney Tissue Immunity. *Nat. Rev. Nephrol.* **2024**, *20*, 71–72. [CrossRef]
86. Lake, B.B.; Menon, R.; Winfree, S.; Hu, Q.; Melo Ferreira, R.; Kalhor, K.; Barwinska, D.; Otto, E.A.; Ferkowicz, M.; Diep, D.; et al. An Atlas of Healthy and Injured Cell States and Niches in the Human Kidney. *Nature* **2023**, *619*, 585–594. [CrossRef]
87. Hansen, J.; Sealfon, R.; Menon, R.; Eadon, M.T.; Lake, B.B.; Steck, B.; Anjani, K.; Parikh, S.; Sigdel, T.K.; Zhang, G.; et al. A Reference Tissue Atlas for the Human Kidney. *Sci. Adv.* **2022**, *8*, eabn4965. [CrossRef]
88. He, Y.O.; Barisoni, L.; Rosenberg, A.Z.; Robinson, P.; Diehl, A.D.; Chen, Y.; Phuong, J.P.; Hansen, J.; Herr II, B.W.; Börner, K.; et al. Ontology-Based Modeling, Integration, and Analysis of Heterogeneous Clinical, Pathological, and Molecular Kidney Data for Precision Medicine. *bioRxiv* **2024**. [CrossRef]
89. Bhatia, D.; Srivastava, S.P. Editorial: Diabetic Kidney Disease: Routes to Drug Development, Pharmacology and Underlying Molecular Mechanisms. *Front. Pharmacol.* **2023**, *14*, 1252315. [CrossRef] [PubMed]
90. Kim, M.K. Treatment of Diabetic Kidney Disease: Current and Future Targets. *Korean J. Intern. Med.* **2017**, *32*, 622–630. [CrossRef]
91. Sebastian, Y.; Smalheiser, N.R. Editorial: Emerging Areas in Literature-Based Discovery. *Front. Res. Metr. Anal.* **2023**, *8*, 1122547. [CrossRef]
92. National Library of Medicine (US). UMLS Reference Manual. Available online: https://www.ncbi.nlm.nih.gov/books/NBK9676/ (accessed on 9 January 2024).
93. Shi, C.; Kong, X.; Huang, Y.; Yu, P.S.; Wu, B. HeteSim: A General Framework for Relevance Measure in Heterogeneous Networks. *IEEE Trans. Knowl. Data Eng.* **2014**, *26*, 2479–2492. [CrossRef]
94. Pires, C. A Systematic Review on the Contribution of Artificial Intelligence in the Development of Medicines for COVID-2019. *J. Pers. Med.* **2021**, *11*, 926. [CrossRef]
95. Ernst, P.; Siu, A.; Weikum, G. KnowLife: A Versatile Approach for Constructing a Large Knowledge Graph for Biomedical Sciences. *BMC Bioinform.* **2015**, *16*, 157. [CrossRef] [PubMed]
96. Xu, J.; Kim, S.; Song, M.; Jeong, M.; Kim, D.; Kang, J.; Rousseau, J.F.; Li, X.; Xu, W.; Torvik, V.I.; et al. Building a PubMed Knowledge Graph. *Sci. Data* **2020**, *7*, 205. [CrossRef]
97. Chandak, P.; Huang, K.; Zitnik, M. Building a Knowledge Graph to Enable Precision Medicine. *Sci. Data* **2023**, *10*, 67. [CrossRef]
98. Bang, D.; Lim, S.; Lee, S.; Kim, S. Biomedical Knowledge Graph Learning for Drug Repurposing by Extending Guilt-by-Association to Multiple Layers. *Nat. Commun.* **2023**, *14*, 3570. [CrossRef] [PubMed]
99. Yu, G.; Wang, L.G.; Han, Y.; He, Q.Y. clusterProfiler: An R Package for Comparing Biological Themes Among Gene Clusters. *OMICS* **2012**, *16*, 284–287. [CrossRef] [PubMed]
100. Hu, Y.; Zhou, W.; Ren, J.; Dong, L.; Wang, Y.; Jin, S.; Cheng, L. Annotating the Function of the Human Genome with Gene Ontology and Disease Ontology. *Biomed. Res. Int.* **2016**, *2016*, 4130861. [CrossRef] [PubMed]
101. Asif, M.; Martiniano, H.F.M.C.M.; Vicente, A.M.; Couto, F.M. Identifying Disease Genes Using Machine Learning and Gene Functional Similarities, Assessed through Gene Ontology. *PLoS ONE* **2018**, *13*, e0208626. [CrossRef] [PubMed]
102. Gene Ontology Consortium.; Ashburner, M.; Ball, C.; Blake, J.; Botstein, D.; Butler, H.; Cherry, J.; Davis, A.P.; Dolinski, K.; Dwight, S.; et al. Gene Ontology: Tool for the Unification of Biology. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]
103. Gene Ontology Consortium. The Gene Ontology Resource: Enriching a GOld Mine. *Nucleic Acids Res.* **2020**, *49*, D325–D334. [CrossRef]
104. Cox, B.J. Systems Biology Analyses of the Placenta. In *The Guide to Investigation of Mouse Pregnancy*; Academic Press: Cambridge, MA, USA, 2014; pp. 259–274.

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