

Review

Reckoning the Dearth of Bioinformatics in the Arena of Diabetic Nephropathy (DN)—Need to Improvise

Jae-Wook Oh ¹, Manikandan Muthu ², Steve W. Haga ³, Vimala Anthonydhasan ⁴, Piby Paul ⁵ and Sechul Chun ^{2,*}

¹ Department of Stem Cell and Regenerative Biotechnology, Konkuk University, Seoul 143-701, Korea; ohjw@konkuk.ac.kr

² Department of Environmental Health Sciences, Konkuk University, Seoul 143-701, Korea; bhagatmani@gmail.com

³ Department of Computer Science and Engineering, National Sun Yat Sen University, Kaohsiung 804, Taiwan; stevewhaga@yahoo.com

⁴ Department of Microbiology and Immunology, Institute for Biomedicine, Gothenburg University, 413-90 Goteborg, Sweden; vimalalisha@gmail.com

⁵ St. Jude Children's Cancer Research Hospital, 262 Danny Thomas Place, Memphis, TN 38105, USA; piby.paul@stjude.org

* Correspondence: scchun@konkuk.ac.kr; Tel.: +82-24500574

Received: 2 June 2020; Accepted: 1 July 2020; Published: 9 July 2020



Abstract: Diabetic nephropathy (DN) is a recent rising concern amongst diabetics and diabetologist. Characterized by abnormal renal function and ending in total loss of kidney function, this is becoming a lurking danger for the ever increasing population of diabetics. This review touches upon the intensity of this complication and briefly reviews the role of bioinformatics in the area of diabetes. The advances made in the area of DN using proteomic approaches are presented. Compared to the enumerable inputs observed through the use of bioinformatics resources in the area of proteomics and even diabetes, the existing scenario of skeletal application of bioinformatics advances to DN is highlighted and the reasons behind this discussed. As this review highlights, almost none of the well-established tools that have brought breakthroughs in proteomic research have been applied into DN. Laborious, voluminous, cost expensive and time-consuming methodologies and advances in diagnostics and biomarker discovery promised through beckoning bioinformatics mechanistic approaches to improvise DN research and achieve breakthroughs. This review is expected to sensitize the researchers to fill in this gap, exploiting the available inputs from bioinformatics resources.

Keywords: diabetes; diabetic nephropathy; bioinformatics; proteomics; microalbumin

1. Introduction

Diabetic nephropathy (DN), also known as diabetic kidney disease (DKD), is a marked rise in the urinary albumin excretion (UAE) rate leading to aberrant kidney function. This results in variations in plasma creatinine, glomerular filtration rate (GFR) and creatinine levels [1]. While there are other macrovascular and microvascular complications resulting from diabetes, DN is the most predominant microvascular complication [2]. Complications arising due to diabetes involve pathologic changes affecting blood vessels, nerves, skin and the eye. Macrovascular complications profoundly influence the blood vessels in brain, heart and extremities and microvascular complications of diabetes involve retinopathy and nephropathy (Figure 1). Approximately 30% of patients with either of type 1 or type 2 diabetes are reported to end up with nephropathy [3]. DN is one of the major consequences of type 1 and type 2 diabetes, which culminates in end-stage renal disease. Hyperglycemia is the driving force behind diabetic nephropathy, operating through production of free radicals that trigger

oxidative stress. Oxidative stress significantly contributes towards accelerating the severity of DN (mechanisms undisclosed). Thus, DN has impacted public health and social economy significantly and has become a well-known cause of concern amidst diabetics [4].



Figure 1. An overview of the two categories of diabetes-based complications and associated organs.

Proteomics is a field that promises to enable the fundamental understanding in various biological processes. Although proteomics began with establishing the identity and estimating the expression levels of proteins, analytical instrumentation, automation of large-scale analytical tools and emergence of bioinformatics tools has broadened the scope and vision. This has enabled researchers to analyze biologically relevant proteins, PTMs, interactions, foldings, etc. To start with, we begin with complex protein mixtures consisting of proteins of different molecular mass, solubilities and modifications. These are then separated into rather less complex mixtures. Generally, the separation is carried out by either two-dimensional gel electrophoresis (2-DE) or various chromatography-based approaches and peptides analyzed by MALDI-TOF-MS or ESI-MS. The data generated by MS are identified based on available databases using various bioinformatics tools. This is how the identity of a protein is established from the original complex protein mixture. Kumar et al. (2017) extensively reviewed the available proteomic tool options for clinical analytes [5]. These tools include: two-dimensional gel electrophoresis (2-DE), two-dimensional differential gel electrophoresis (2D-DIGE), mass spectrometry (MS) including MALDI- and ESI-based ionization techniques, liquid chromatography and multi-dimensional separation technology, surface-enhanced laser desorption/ionization (SELDI), proteins and antibody microarrays, isotope-coded affinity tags (ICAT), isobaric tag for relative and absolute quantitation (iTRAQ), stable isotope labeling with amino acids in cell culture (SILAC), affinity tagging and mass spectroscopy and proteogenomics [5].

Protein analysis/proteomics permits for the rapid assessment of the proteins expressed in biological samples. Biological samples are systematically analyzed for identifying, quantifying and differentiating various proteins that hold significance in pathology. Biofluid proteomics has led to the identification of disease biomarkers [6] for non-invasive renal disease diagnosis [7,8]. The routine tests for urine proteins can only estimate proteinuria or a single protein at a time, whereas proteomic technologies allow simultaneous examine multiple urinary and plasma proteins in a single run [9]. High resolution followed by instrumentation sensitivity and throughput capabilities are the prerequisites of any ideal analytical technique in the field of proteomic investigations. Although various other advanced proteomic analytical techniques have evolved with the rise of quantitative and shotgun proteomics, 2-DE is still the most predominantly used technique with respect to diabetic nephropathy. This is

followed by capillary electrophoresis coupled with mass spectrometry (CE-MS). In the case of diabetic nephropathy, 2-DE is the most commonly used. 2-DE separates proteins based on differential pH in the first dimension and it separates based on variations in molecular masses in the second dimension [10], allowing the characterization of the renal proteome, renal cortex [11,12], glomerular cells [13] and tubular epithelial cells [14,15]. Significant understanding has evolved through the use of proteomic tools and resources. Kim et al. [16] studied the serum proteomes of type 2 diabetic patients with end stage renal disease (ESRD) and normal patients. In 2012, Zürlbig et al. [17] analyzed urine samples from type 1 and 2 diabetic patients to assess peptides in urine samples for signs of DN. Papale et al. [18] identified urinary protein biomarkers using surface enhanced laser desorption/ionization (SELDI)-time of flight (TOF)/mass spectrometry (MS). A few other reports have successfully classified the variation between DN [19–23] and other chronic kidney diseases [24]. Early detection of changes in the urinary proteome of diabetics can help stop progressive renal function loss. Owing to its direct clinical implications, urinary proteomics is considered as a valuable tool for premature detection of patients having kidney- and heart-related complications well ahead of actual manifestation of renal dysfunction [21].

Bioinformatics is the synergistic application of computer science and information technology (IT) to biology and medicine. Its rich reservoir of inputs arises from its algorithms, artificial intelligence databases, soft computing, data mining, image processing, modeling and simulation systems. Bioinformatics laid the groundwork in proteomics via the genome sequencing projects and microarray-based [25] and phenotypic profiles at the cell level as well as in organisms as such [26]. It focuses on gene expression [27]. While biological data rest on experimental methods and analytical instrumentation for data collection and analysis, bioinformatic tools enable achieving ambitious goals easily. Computational methods work beside analytical methods for acquisition of large quantities of data in order to arrive at useful scientific conclusions (Figure 2). Biological databases help scientists to elucidate biological phenomena based on the structure and functionality of biomolecules and their interaction. This leads to a proper understanding of diseases and reinforces the fight against diseases. Modeling and Simulation assists in understanding the interaction of individual systems and the system on the whole. Thus, computer modeling and simulation methodologies play a pivotal role in alignment of sequences, gene identification, drug design and discovery and elucidation of protein structures for medical applications [28]. A drug development process can take 10–15 years [29], and is expensive as well. A recent Forbes analysis estimated the cost at \$5 billion [30,31]. The final cost of any drug is arrived at taking into account the total costs from discovery to approval and the cost of absorbing all the clinical trial failures [32]. In this regard, bioinformatics tools certainly enhance the efficiency of target discovery and validation processes, reducing time and being cost-effective.

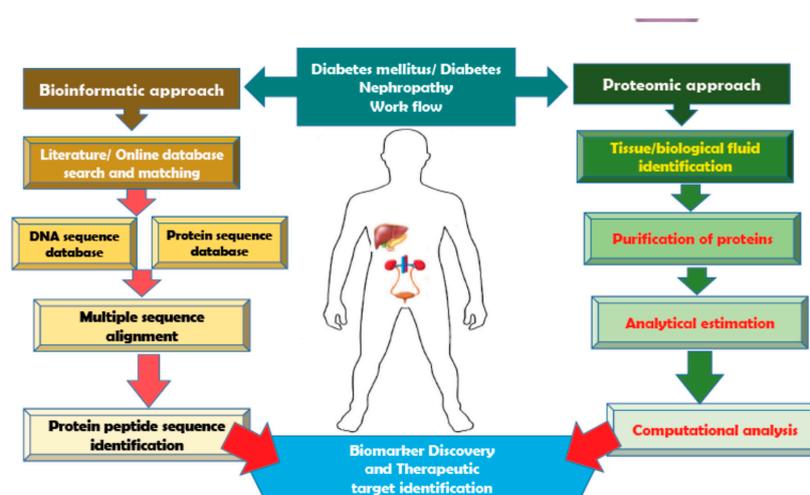


Figure 2. Schematic showing the general proteomic work flow and Bioinformatics work flow.

This current review presents the raising concern around DN. A brief snapshot highlighting the tools that are being applied for diabetes-based research is presented. Proteomic advancements made in the area of DN are discussed. While bioinformatics has been used successfully in various fields of proteomics, and in diabetes related research, this review highlights the gap in bioinformatics applications in the area of DN. This review is expected to persuade and draw the attention of scientists working in this area to adopt bioinformatics-based approaches for unraveling mysteries surrounding DN.

2. A Snapshot of Bioinformatic Tools Used for Diabetes Mellitus

Diabetes mellitus is characterized by aberrant glucose metabolism resulting in high blood glucose level in blood and eventually in urine too. This condition of high glucose level in the body is medically defined as hyperglycemia. This can happen due to insulin secretion deficiency, insulin resistance or both. Symptoms of diabetes include palpitations, polyuria, weight loss, headache, polydipsia, tachycardia, blurred vision and polyphagia [33]. Diabetes mellitus is a significant health disorder and the World Health Organization (WHO) states “diabetes is a world-wide epidemic” [34]. Thus, the resources devoted to screening of diabetes as well as primary prevention have gained paramount importance.

The use of bioinformatics tools in diabetes research has aided in biomarker identification studies. The use of bioinformatics resources enables the prediction of functional regions of the protein/domains, structures and phylogenetic/evolutionary relationships of the sequences examined. It further leads to scripting of algorithms and statistics that solve biological issues. Bioinformatics is thus an interdisciplinary field of exploiting areas of biology, computer science, chemistry, math and engineering for analyzing and interpreting various aspects of life science [35,36]. The information acts as a guideline for the practical scientific work, saving time and giving an access to various tools available online for free. Currently, major trends in proteomics-based bioinformatics research include: ligand-based drug design for modulating metabolic pathways and for elucidating the structure of protein and molecular docking studies that play a vital role in drug discovery [37] and omics approaches such as bulk transcriptomics, epigenetics, single cell transcriptomics, metagenomics and epigenetics. Table 1 give a snap shot of the various bioinformatics tools that have been thus far applied to diabetes research.

Various Human protein–protein interaction (PPI) databases are also reported. Human Annotated and Predicted Protein Interaction (HAPPI) database helps explore PPI data needed for network biology studies. This information has been used for understanding biological processes such as those involved in insulin signaling in type 2 diabetes [65]. The University of California Santa Cruz (UCSC) Genome Browser is a well-known web-based tool for dynamic display of any required genome portion. Using this option, users can upload their own datasets (in any format) as custom annotation tracks [66]. One can visualize, browse and display information from type 2 diabetes for genomic information. Recently, large-scale meta-analyses of GWAS (Genome Wide Association Studies) has improved the existing understanding on genetic factors behind type 2 diabetes. These studies have identified that type 2 diabetes holds risk genes located in localized parts of the genome (i.e., chromosome 20). This knowledge has evolved from using the type 2 diabetes genetic network [67]. Gene annotation information can be obtained from the VEGA (Vertebrate Genome Annotation) database. Genetic data are also available on type 2 diabetes from gene bank databases [68].

Table 1. Snapshot of Bioinformatic tools used for Diabetes mellitus.

Sl. No	Function	Bioinformatic Tool	Location	References
1	Proteomics Identification database	PRIDE	https://www.ebi.ac.uk/pride/	Perez-Riverol et al., 2019 [38]
2	Medical Calculators, Clinical Resources for diabetic data	diabetic-database	https://globalrph.com/medcalcs/diabetic-database/	[39]
3	Protein sequence Database	UniProt	www.uniprot.org	Chen, C. et al., 2017 [40]
4	Genetic sequence database	GenBank	https://www.ncbi.nlm.nih.gov/genbank/	Benson, D.A. et al., 2013 [41]
5	Proteomic fragments analysis	MASCOT	https://www.sanger.ac.uk/science/tools	Brosch, M. et al., 2009 [42]
6	Human Metabolome Database	hmdb	https://hmdb.ca/	Wishart, D.S. et al., 2018 [43]
7	Structural, functional annotation of proteins	Gene3D	http://gene3d.biochem.ucl.ac.uk/Gene3D/	Yeats et al., 2006 [44]
8	miRNA, mRNA, protein, phosphoproteins and metabolite expression data sets	KUPKB	http://www.kupkb.org/	Klein et al., 2012 [45]
9	Protein Motif fingerprinting	PRINTS database	http://www.bioinf.manchester.ac.uk/dbbrowser/PRINTS/	Attwood et al., 2003 [46]
10	Gene expression	GEO dataset	https://www.ncbi.nlm.nih.gov/geo/	Zhao et al., 2013 [47]
11	Knowledge portal for type 2 diabetics	Type 2 Diabetes	http://www.type2diabetesgenetics.org/	Jeyaraman, M.M. et al., 2020 [48]
12	Resources for genomic and epigenetic studies of type 2 diabetes and associated issues	Diabetes Epigenome Atlas	https://www.diabetesepigenome.org/	Khetan et al., 2018 [49]
13	Database for diabetes based diagnostic methods	pima	https://www.kaggle.com/uciml/pima-indians-diabetes-database	Barale, M. and Shirke, D.T., 2016 [50]
14	National Diabetes Information	NIDDK	http://diabetes.niddk.nih.gov/about/	Whetzel et al., 2015 [51]

Table 1. Cont.

Sl. No	Function	Bioinformatic Tool	Location	References
15	Protein identification using peptide information from MS/MS	ProteinProphetTM	http://proteinprophet.sourceforge.net/	Nesvizhskii, A.I. et al., Anal Chem. 2003 [52]
16	2D Gel Database	SWISS-2DPAGE	http://world-2dpage.expasy.org/swiss-2dpage/	Hoogland, C. et al., 2014 [53]
17	Protein 3D structure database	PDB	www.rcsb.org	wwPDB Consortium, 2019 [54]
18	Protein–Protein interaction networks	STRING	http://string-db.org	Szklarczyk, D. et al., 2015 [55]
19	For structural and functional Annotation	SUPFAM Database	http://supfam.org	Wilson, D. et al., 2009 [56]
20	Repository for chemical substances and their biological activities	PUBCHEM	https://pubchem.ncbi.nlm.nih.gov	Kim, S. et al., 2016 [57]
21	For Visualizing and interpreting metabolomic data	Cytoscape MetScape 3.1	http://metscape.med.umich.edu/	Karnovsky, A. et al. 2012 [58]
22	Literature search based on disease related terms mapped to PubChem compounds for annotating compound networks	MetDisease	http://metdisease.ncibi.org/	Duren, W. et al., 2014 [59]
23	Metabolic Pathway Database	KEGG	https://www.kegg.jp/	Kanehisa, M. et al., 2016 [60]
24	Gene ontology	PANTHER	http://www.pantherdb.org/	Thomas, P.D. et al., 2013 [61]
25	Gene Set Enrichment Analysis	GSEA	http://www.webgestalt.org/	Wang, J. et al., 2017 [62]
26	Database for Single Polymorphic data	dbSNP	http://www.ncbi.nlm.nih.gov/SNP	Smigielski, M. et al., 2000 [63]
27	Database for target genes of potential miRNAs	mirnet	https://www.mirnet.ca/	Fan, Y. et al., 2018 [64]

Other bioinformatics-based studies have reported the identification of biomarkers that influence vascular calcification processes as a basis of diagnosis and treatment of type 2 diabetes in its subclinical stages. The LC MSMS extracted and identified proteins were analyzed through gene ontology and compared with other specific proteins using DisGeNET database. Cytoscape software analyzed the network of biological functions of the proteins. These bioinformatics tools revealed the involvement of ideal biomarkers of vascular calcification in patients with type 2 diabetes. Other researchers have used other bioinformatics tools for the identification of biomarkers that aid in therapeutic targeting in case of early diabetes [69] or premature-onset of type 2 diabetes using the DIRECT (Diabetes Research on Patient Stratification) Consortium. The DIRECT Consortium under the Innovative Medicines Initiative (IMI) combines the European Union (EU), European academic institutions and pharmaceutical companies and aims at identifying biomarkers that will aid in diabetes related drug development, as well as helping in developing a stratified approach for treatment of type 2 diabetes.

3. Proteomic Advances Made in the Area of Diabetic Nephropathy

Lee and Choi [70] reported that their PubMed search using the keywords “proteomics” or “proteomic” or “proteome” with “diabetic nephropathy” hit 86 articles published between 2002 and 2010. Van et al. [4] hit 155 search results as of 2017. We also performed a PubMed search and the results as of 2020 show 222 articles including reviews. This number underscores the impact of proteomics in DN, as observed by the consistent need driven enthusiasm in research published in this area.

DKD is also described as a glomerular related pathology, since injury to the glomerulus precedes and outweighs progressive injury to the tubulo-interstitium [71–73]. Stages involved in DN have been clearly demarcated. Uncomplicated diabetes is the preliminary stage of DN/DKD, wherein modifications in hyperfiltration and hypertrophy act as indicators [74]. Between 20% and 30% of patients with diabetes proceed to the next stage, which is incipient diabetic nephropathy [1–3], where there is persistent microalbuminuria, with progressive decline in GFR [6]. The onset of proteinuria marks the manifestation of late-stage diabetic kidney disease. Satirapoj and Adler (2014) [75]. in their extensive review on diabetic nephropathy. addressed the risk factors, clinical stages, pathogenesis, renal pathology, diagnostic criteria and available therapeutic interventions with respect to DN. It is well known that multiple factors are involved in DN and the molecular mechanisms are poorly understood.

Microalbuminuria is currently the most trustworthy diabetic nephropathy predictor [1–3], although there are other arguments refuting the credibility of using this as the sole marker. Proteomic methods allow rapid understanding of a proteome (meaning the complete set of proteins expressed within urine, plasma or serum samples) [6]. This paves way for effective non-invasive diagnosis of renal diseases [7]. The proteomic tools that help analyze such protein traces in urine and kidney, aid in early and premature diagnosis. Mass spectrometry has enabled the identification of thousands of proteins and peptides and urinary markers in one go. Urinary proteomics and renal proteomics have had important clinical implications in DN, given that urine can be collected noninvasively and easily and is output from kidneys [75]. Urinary biomarker discovery improves disease diagnosis, prognosis and treatment [21]. Serum can also be used for proteomic analysis, but, since the urinary proteins are more stable and it is easier to handle urine than blood, urine is a more advantageous option [18,23]. Analytical techniques need to possess high resolution, high sensitivity and high throughput abilities for identifying novel DN biomarkers. Proteomics-focused understanding of DN requires information on expression of individual proteins, their function and modifications in their functions as a consequence of diabetes. On the other hand, individual proteins or pathways altered by diabetes are also of interest [76].

Non-invasive diagnostics for diabetic nephropathy gains eminence from the aspect that they could aid in prognosis prior to progression of the disease. At present, the most sensitive and non-invasive indicator of the progression of diabetic nephropathy is microalbuminuria. Senatorski et al. [77] reported that TGF-beta 1 and IL-6 levels in urine are effective prognostic factors for identifying the onset of DN during the course of diabetes mellitus. An ideal renal biomarker should be easy to measure

and noninvasive. Accuracy, reproducibility, sensitivity, specificity and cost effectiveness, leading to accurate prediction, prognosis and progression of disease, are trademarks of an efficient renal marker. Proteomic approaches have discovered biomarkers, such as collagen fragments [18,78], β 2 MG [15], proinflammatory cytokines [79] and retinol-binding protein (RBP) [80]. Such proteomic biomarkers can help differentiate between DN affected patients and DN-free diabetics. Proteomic investigations in the area of DN have made consistent and considerable breakthrough in the area of biomarker discovery.

Kim et al. [16] reported changes in 26 proteins specific to type 2 diabetics and Zürlbig et al. [17] reported peptides in urine samples for signs of DN, employing a CKD biomarker classifier from a longitudinal cohort of diabetics. The application of this classifier enabled early detection. This was much ahead of the routine urinary diagnosis. Collagen fragments could help in early diagnosis (3–5 years) before macroalbuminuria set in. Papale et al. [20] evaluated proteomic profiles of urine generated by SELDI-TOF MS and could reliably identify biopsy-proven DN and establish lineages with renal injury. Using the model, normoalbuminuria patients, microalbuminuria patients and those with DN were classified in diabetic and non-diabetic patients. Ubiquitin and β 2 MG are the predominant biomarkers evolved using the classification model. Overgaard et al. [22] analyzed plasma type 1 diabetic patients and have differentiated various stages of kidney disease and classified them accordingly. Using SELDI-TOF-MS analysis, four peaks of transthyretin, apolipoprotein A1, apolipoprotein C1 and cystatin C were identified. Rossing et al. [18] also identified biomarkers to differentiate and identify diabetic patients with and without DN [18,19,77]. Recently, Roscioni et al. [21] developed a urinary proteomic risk score-based prediction system that could evidence the progression of microalbuminuria in type 2 diabetic patients. Good et al. [81] reported on a classifier that was specifically indicative of CKD. Few other reports confirm that proteome analysis of urine could distinguish between DN and CKD patients [18,19,24]. These reports show that proteomic approaches have a lot to offer towards early novel and noninvasive predictors during DN diagnosis. Speculating on the limitations of DN proteomic studies, small sample size and reproducibility, as well as need for extensive data collection and analysis and time and energy and cost effectiveness may be worth mentioning.

4. Future Perspective: Bioinformatics Applications into DN

With a large amount of proteomic inputs in DN, we delve into how far bioinformatic resources have been invested into harnessing molecular data towards computational rationalization of DN proteomic data. With the inputs of bioinformatics resources being enormous in proteomics research and given the fact that DN has also extensive proteomic aspects to it, this review starts off on a high note that we will be consolidating on voluminous advances and reporting tools that have been used in DN. It was interesting to observe that, while bioinformatics aspects have been applied to diabetes-based research, our search could only arrive at a handful of publications in this area (Table 2), with scant reports of few databases alone being used for DN investigations. Recently, with the development of high-throughput technologies, it is expected that much of these deliverables would be harnessed and applied. The few scattered reports on bioinformatics applications in DN research are presented in this section.

Single nucleus RNA sequencing technology is able to bring about changes in gene expression in early DN patients, enhance urinary potassium secretion and decrease calcium and magnesium reabsorption. These changes identify biomarkers leading to early detection and treatment [3]. To identify the factors influencing diabetic nephropathy (DN), Fan et al. [82] reported RNA sequencing of renal biopsy samples of patients with early and advanced DN and normal kidney tissue from nephrectomy samples. RNA-seq of 28 DN and 9 control samples was conducted, where the total RNA was extracted and evaluated by an Agilent 2100 Bioanalyzer. cDNA libraries were constructed and sequenced using a HiSeq 4000 system. The good quality reads were identified based on human reference databases employing the use of STAR alignment algorithm [83]. After filtering and normalization, differential analysis by limma test [84] identified significantly dysregulated genes. The output was analyzed using Gene Ontology (GO) function [85] and various pathways such as Ingenuity

Pathway Analysis, Reactome, Kyoto Encyclopedia of Genes and Genomes, Naba, BioCarta, Panther, Pathway Interaction Database and WikiPathways. As for deconvolution analysis, CIBERSORT [86] was used and single-cell data belonging to human kidney were extracted from the Gene Expression Omnibus (GEO) database (GSE109564) and SEURAT identified 16 cell types [87]. The results obtained employing bioinformatic resources show significant increase in fibroblasts, monocytes and myofibroblasts in DN-advanced kidneys.

The candidate gene approach was employed to search for genes that are involved in the susceptibility of DN, before the development of genome-wide association studies. Recently, Van et al. [4] investigated the implications of differentially excreted urinary proteins in DN based published, relevant data. Candidate markers in DN patients and controls were extracted from 31 published studies. All candidates were weighted equally for network analyses. To conduct the localization analysis, the screened candidates were mapped onto specific nephron segments based on their renal expression (in normal tissues) using the existing literature on Human Protein Atlas. For the functional analysis, biological processes Biological Networks Gene Ontology [88] implicated in the progression of DKD were used. Enrichment Map plug-ins of Cytoscape software helped in mapping the proteins associated with each stage of DN [89]. PPI networks equipped for the retrieval of interacting genes were constructed [90]. Using the bioinformatics tools, the authors successfully and comprehensively identified crucial biological processes in DKD and proteins associated with a particular stage of disease. Despite genome-wide association on unique and common genetic variants from exome sequencing, the genetics behind DN are poorly understood [91–93], although proteome- and transcriptome-based studies have been utilized for diabetes-based biomarker identification [94]. It is emphasized that DN would benefit from these studies as well. Figure 3 displays the room for development and improvisation of bioinformatics options into DN research. Next generation bioinformatics resources such as those listed below are those that we propose for use in DN-related research to achieve expected breakthroughs.

Table 2. Bioinformatic inputs into the area of Diabetic Nephropathy.

Software Function/Application	Bioinformatics Resources	Reference
Investigating implications of proteins in urine samples from DN patients	using protein–protein interactions (PPI) network analysis-STRINGv10.	Van et al., 2017 [4]; Szklarczyk, D. et al., [55]
Biomarkers for DN	PPI for determining interactions within proteins involved in progression of diabetes	Abedi and Gheisari, 2015 [95]; Saito et al., 2016 [96]; Varemo et al., 2015 [93]
DN urinary biomarkers in the various nephrons were elucidated and mapping of protein biomarkers in nephron segments	Human Protein Atlas https://www.proteinatlas.org/ determined differences in protein expressions in renal tissues vs. normal tissues	Uhlen et al., 2010 [97]; Van et al. 2017 [4]
Identify DN in Type 2 diabetic patients	decision tree-based prediction tool to identify DN in patients with type 2 diabetes	Huang et al., 2015 [98]
DN prediction	applied machine learning for early prediction of DN via risk factor analysis	Cho et al., 2008 [99]
DN related Factors	random forest learning algorithm (Breiman, 2001) for understanding factors behind diabetic peripheral neuropathy (DPN).	DuBrava et al. 2017 [100]; Chadinee et al., 2018 [37]
RNA sequencing of biopsy kidney samples from early DN and advanced DN patients and that from normal kidney tissue	Gene ontology http://geneontology.org/ CIBERSORT https://cibersortx.stanford.edu/ Gene expression Omnibus https://www.ncbi.nlm.nih.gov/geo/ SEURAT https://satijalab.org/seurat/	The Gene Ontology, 2017 [85]; Newman, A.M. et al., 2015 [86]; Zhao, M. et al., 2017 [47]; Clough, E. and Barrett, T. 2016 [87]
Identification of enriched biological processes, 76 differentially expressed proteins in diabetes affected kidneys identified	Cytoscape using plug ins Biological Networks Gene Ontology Enrichment Map	Maere et al., 2005 [88]; Merico et al., 2010 [89]; Van et al., 2017 [4]

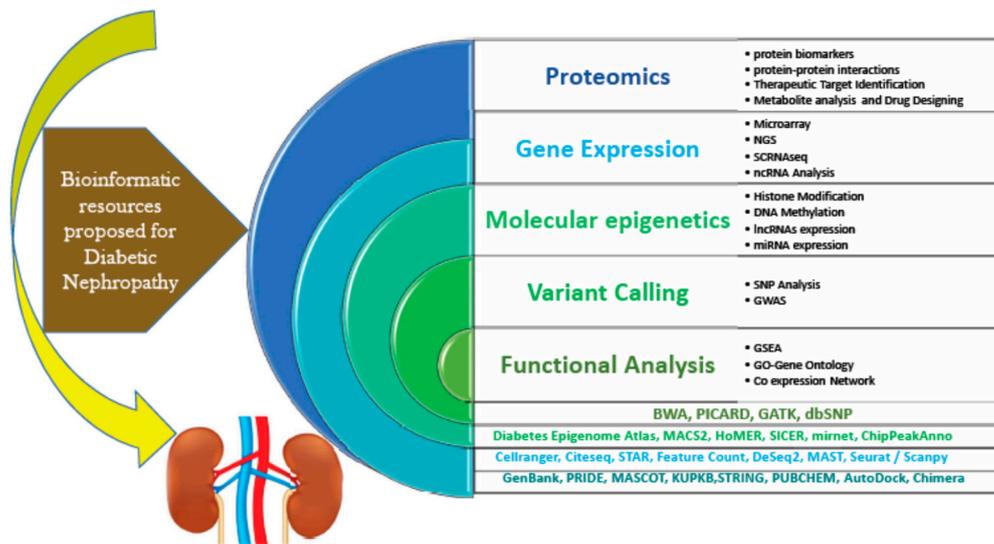


Figure 3. The future perspective of DN research: through improvisation from various bioinformatic resources, categories and areas where bioinformatics resources will come handy are listed.

These resources include: BWA (Burrows–Wheeler Aligner) aligns genomic regions; PICARD is used to manipulate the high-throughput sequence (HTS) data and sam/bam/vcf file processing; GATK (Genome Analysis Toolkit) is for variant analysis of HTS data; dbSNP (Database of single nucleotide polymorphism) is a repository for nucleotide variations; MACS2 (Model-based Analysis of ChIP-seq) is for identifying significantly enriched genomic regions in ChIP/DNase-seq data; HoMER (Hypergeometric Optimization of Motif EnRichment) predicts the known/de-novo motifs for NGS sequence analysis; SICER (spatial clustering approach for the identification of ChIP-enriched regions) calls the broad peaks from chipseq data and identifies the enriched and differentially binding region; Cell ranger is a software used for primary analysis of the chromium 10× single cell data; DeSeq2 (differential gene expression analysis of RNA-seq data) finds the significantly differentially expressed genes between disease and control sample; MAST (Motif Alignment & Search Tool) is used to annotate the matched motif; Seurat/Scanpy is a single cell RNA seq downstream analysis pipeline; GenBank is a DNA sequence database; proteomic tools such as PRIDE (PRoteomics IDentifications database and core tools) analyze and visualize the MS based proteomics data; MASCOT (predicting protein sequences from mass spectral data) is an integrated tool for Peptide mass fingerprint, sequence query; MS/MS Ion Search uses uninterpreted MS/MS data from one or more peptides; Sys-BodyFluid is a database for human body fluid proteome research generally used for body fluid related proteomic study; Human urinary proteomic fingerprint database (UPdb); Urinary Exosome Protein Database is used for primary data derived from MALDI-, MELDI-, SELDI-, CE-, LC- and other TOF-MS analyses in urinary research; Human Kidney and Urine Proteome Project (HKUPP) is used for proteomic analysis of urine samples and human kidney tissues; The Kidney & Urinary Pathway Knowledge Base (KUPKB) contains multi-omics data such as mRNA, miRNA, metabolites and proteins; STRING (Search Tool for the Retrieval of Interacting Genes/Proteins) predicts interaction network for protein–protein functional enrichment analysis; PUBCHEM is a molecular chemical information database used to retrieve the drug/compound structure; AutoDock (Automated Docking Tool) is for molecular docking; and Chimera is an interactive visualization and analysis tool for three-dimensional molecular structures and related data. All these resources are well established in urinary-, kidney- and diabetes-related proteomics and metabolomics; DN would no doubt benefit from these assets. When bioinformatics tools could lead to unequivocal progress in DN biomarker studies leading to early diagnostics and for DN drug discovery based on the receptor protein, it is rather strange that given all these assets that still bioinformatics is kept at bay.

Although massive strides are being made in the area of DN, several notable limitations remain. There is a need to identify more potential drug targets using computer-aided drug designing/virtual screening. More novel biomarkers improve the therapeutic efficacy. Since there are no DN-based databases/repositories available, the information is scattered everywhere. Epigenetic mechanisms are assumed to play a role in the regulatory element mechanism of DN. This area of study works hand in hand with bioinformatics resources. Epigenetic regulations such as DNA methylation, noncoding RNAs and histone modifications, have a profound influence on DKD. Using quantitative proteomics (2DE, pSILAC, SILAC, iTRAQ, 2D-DIGE and iCAT), it is possible to find the microRNA-based targets and their protein products, which are the key facts for identifying potential targets. These bioinformatics options contribute to developing novel therapies for DN and understanding the DN driven genetic and epigenetic changes. Epigenetic changes being reversible can be used successfully for DN therapeutics. Very few studies have reported on gene expression profiling of genome wide association studies such as transcriptome analysis, epigenetic profiling, single cell RNA sequence analysis and metagenomics studies leading to identifying disease targets. The integration of bioinformatics with proteomics has led to revolutionizing breakthroughs in various fields. Such a benefit has been reaped in diabetes research. While this is the case, there is no doubt that diabetic nephrology will also take advantage of this technique. There is no realistic answer as to why almost nothing has been applied from bioinformatics to DN research, but this could be as simple as a gap in the collaboration of researchers from diverse fields, such as bioinformatics and proteomics as well as by nephrologists and diabetologist. This review sees a need for such an exploration and combinatorial approach in this area.

5. Conclusions

The rising concern regarding diabetic nephropathy is addressed and the breakthroughs achieved through bioinformatic inputs in the area of diabetes research are presented. The voluminous proteomics-based progress made in the area of DN is presented. Except for a handful of databases and tools that this review assembles, almost no technology transfer has been achieved from bioinformatics information and resources to DN research. This review highlights this void and aims are initiating some thought and attention in this direction. The lack of bioinformatics applications into DN research is emphasized and further integration of bioinformatics resources is presented as the future of DN research.

Author Contributions: M.M., V.A. and S.W.H. conceived and participated in the preparation of the manuscript. J.-W.O., P.P., S.C. engaged in discussions and reviewing the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

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

References

1. Gross, J.L.; de Azevedo, M.J.; Silveiro, S.P.; Canani, L.H.; Caramori, M.L.; Zelmanovitz, T. Diabetic nephropathy: Diagnosis, prevention, and treatment. *Diabetes Care* **2005**, *28*, 164–176. [[CrossRef](#)] [[PubMed](#)]
2. Felehgari, V.; Rahimi, Z.; Mozafari, H.; Vaisi-Raygani, A. ACE gene polymorphism and serum ACE activity in Iranians type II diabetic patients with macroalbuminuria. *Mol. Cell. Biochem.* **2011**, *346*, 23–30. [[CrossRef](#)] [[PubMed](#)]
3. Molitch, M.E.; DeFronzo, R.A.; Franz, M.J.; Keane, W.F.; Mogensen, C.E.; Parving, H.H.; Steffes, M.W.; American Diabetes Association. Nephropathy in diabetes. *Diabetes Care* **2004**, *27*, S79–S83. [[PubMed](#)]
4. Van, J.A.; Scholey, J.W.; Konvalinka, A. Insights into Diabetic Kidney Disease Using Urinary Proteomics and Bioinformatics. *J. Am. Soc. Nephrol.* **2017**, *28*, 1050–1061. [[CrossRef](#)]
5. Kumar, S.; Gaur, V.; Khurana, S.; Bose, S.; Kiran, M.; Kiran, M.; Sharawat, S.K. Proteomics Tools—An Update. *Clin. Oncol.* **2017**, *2*, 1358.
6. Ben Ameer, R.; Molina, L.; Bolvin, C.; Kifagi, C.; Jarraya, F.; Ayadi, H.; Molina, F.; Granier, C. Proteomic approaches for discovering biomarkers of diabetic nephropathy. *Nephrol. Dial. Transplant.* **2010**, *25*, 2866–2875. [[CrossRef](#)]

7. Jungblut, P.; Wittmann-Liebold, B. Protein analysis on a genomic scale. *J. Biotechnol.* **1995**, *41*, 111–120. [[CrossRef](#)]
8. Moresco, R.N.; Sangoi, M.B.; De Carvalho, J.A.M.; Tatsch, E.; Bochi, G.V. Diabetic nephropathy: Traditional to proteomic markers. *Clin. Chim. Acta* **2013**, *421*, 17–30. [[CrossRef](#)]
9. Champion, C.G.; Sanchez-Ferras, O.; Batchu, S.N. Potential Role of Serum and Urinary Biomarkers in Diagnosis and Prognosis of Diabetic Nephropathy. *Can. J. Kidney Health Dis.* **2017**, *4*, 2054358117705371. [[CrossRef](#)]
10. Thongboonkerd, V. Study of Diabetic Nephropathy in the Proteomic Era. *Contrib. Nephrol.* **2011**, *170*, 172–183.
11. Cutillas, P.R.; Biber, J.; Marks, J.; Jacob, R.; Stieger, B.; Cramer, R.; Waterfield, M.; Burlingame, A.L.; Unwin, R.J. Proteomic analysis of plasma membrane vesicles isolated from the rat renal cortex. *Proteomics* **2005**, *5*, 101–112. [[CrossRef](#)] [[PubMed](#)]
12. Magni, F.; Sarto, C.; Valsecchi, C.; Casellato, S.; Bogetto, S.F.; Bosari, S.; Di Fonzo, A.; Perego, R.A.; Corizzato, M.; Doro, G.; et al. Expanding the proteome two-dimensional gel electrophoresis reference map of human renal cortex by peptide mass fingerprinting. *Proteomics* **2005**, *5*, 816–825. [[CrossRef](#)] [[PubMed](#)]
13. Yoshida, Y.; Miyazaki, K.; Kamiie, J.; Sato, M.; Okuizumi, S.; Kenmochi, A.; Kamijo, K.; Nabetani, T.; Tsugita, A.; Xu, B.; et al. Two-dimensional electrophoretic profiling of normal human kidney glomerulus proteome and construction of an extensible markup language (XML)-based database. *Proteomics* **2005**, *5*, 1083–1096. [[CrossRef](#)]
14. Barile, M.; Pisitkun, T.; Yu, M.-J.; Chou, C.-L.; Verbalis, M.J.; Shen, R.-F.; Knepper, M.A. Large-Scale Protein Identification in Intracellular Aquaporin-2 Vesicles from Renal Inner Medullary Collecting Duct. *Mol. Cell. Proteom.* **2005**, *4*, 1095–1106. [[CrossRef](#)] [[PubMed](#)]
15. Dihazi, H.; Asif, A.R.; Agarwal, N.K.; Doncheva, Y.; Müller, G.A. Proteomic Analysis of Cellular Response to Osmotic Stress in Thick Ascending Limb of Henle’s Loop (TALH) Cells. *Mol. Cell. Proteom.* **2005**, *4*, 1445–1458. [[CrossRef](#)] [[PubMed](#)]
16. Kim, H.-J.; Cho, E.-H.; Yoo, J.-H.; Kim, P.-K.; Shin, J.-S.; Kim, M.-R.; Kim, C.W. Proteome Analysis of Serum from Type 2 Diabetics with Nephropathy. *J. Proteome Res.* **2007**, *6*, 735–743. [[CrossRef](#)]
17. Zürbig, P.; Jerums, G.; Hovind, P.; MacIsaac, R.J.; Mischak, H.; Nielsen, S.E.; Panagiotopoulos, S.; Persson, F.; Rossing, P. Urinary Proteomics for Early Diagnosis in Diabetic Nephropathy. *Diabetes* **2012**, *61*, 3304–3313. [[CrossRef](#)]
18. Rossing, K.; Mischak, H.; Dakna, M.; Zürbig, P.; Novak, J.; Julian, B.A.; Good, D.M.; Coon, J.J.; Tarnow, L.; Rossing, P.; et al. Urinary proteomics in diabetes and CKD. *J. Am. Soc. Nephrol.* **2008**, *19*, 1283–1290. [[CrossRef](#)]
19. Rao, P.V.; Lu, X.; Standley, M.; Pattee, P.; Neelima, G.; Girishes, G.; Dakshinamurthy, K.; Roberts, J.C.T.; Nagalla, S.R. Proteomic Identification of Urinary Biomarkers of Diabetic Nephropathy. *Diabetes Care* **2007**, *30*, 629–637. [[CrossRef](#)]
20. Papale, M.; Di Paolo, S.; Magistroni, R.; Lamacchia, O.; Di Palma, A.M.; De Mattia, A.; Rocchetti, M.T.; Furci, L.; Pasquali, S.; De Cosmo, S.; et al. Urine Proteome Analysis May Allow Noninvasive Differential Diagnosis of Diabetic Nephropathy. *Diabetes Care* **2010**, *33*, 2409–2415. [[CrossRef](#)]
21. Roscioni, S.S.; de Zeeuw, D.; Hellemons, M.E.; Mischak, H.; Zürbig, P.; Bakker, S.J.; Gansevoort, R.T.; Reinhard, H.; Persson, F.; Lajer, M.; et al. A urinary peptide biomarker set predicts worsening of albuminuria in type 2 diabetes mellitus. *Diabetologia* **2012**, *56*, 259–267. [[CrossRef](#)] [[PubMed](#)]
22. Overgaard, A.J.; Hansen, H.G.; Lajer, M.; Pedersen, L.; Tarnow, L.; Rossing, P.; McGuire, J.N.; Pociot, F. Plasma proteome analysis of patients with type 1 diabetes with diabetic nephropathy. *Proteome Sci.* **2010**, *8*, 4. [[CrossRef](#)] [[PubMed](#)]
23. Fliser, D.; Novak, J.; Thongboonkerd, V.; Argilés, A.; Jankowski, V.; Girolami, M.A.; Jankowski, J.; Mischak, H. Advances in Urinary Proteome Analysis and Biomarker Discovery. *J. Am. Soc. Nephrol.* **2007**, *18*, 1057–1071. [[CrossRef](#)] [[PubMed](#)]
24. Varghese, S.A.; Powell, T.B.; Budisavljevic, M.N.; Oates, J.C.; Raymond, J.R.; Almeida, J.S.; Arthur, J. Urine biomarkers predict the cause of glomerular disease. *J. Am. Soc. Nephrol.* **2007**, *18*, 913–922. [[CrossRef](#)] [[PubMed](#)]
25. Shoemaker, D.D.; Linsley, P.S. Recent developments in DNA microarrays. *Curr. Opin. Microbiol.* **2002**, *5*, 334–337. [[CrossRef](#)]
26. Gerlai, R. Phenomics: Fiction or the future? *Trends Neurosci.* **2002**, *25*, 506–509. [[CrossRef](#)]
27. Tyers, M.; Mann, M. From genomics to proteomics. *Nature* **2003**, *422*, 193–197. [[CrossRef](#)]

28. Gedela, S.; Rao, A.A.; Medicherla, N.R. Identification of Biomarkers for Type 2 Diabetes and Its Complications: A Bioinformatic Approach. *Int. J. Biomed. Sci. IJBS* **2007**, *3*, 229–236.
29. Tamimi, N.A.; Ellis, P. Drug Development: From Concept to Marketing! *Nephron Clin. Pract.* **2009**, *113*, c125–c131. [CrossRef]
30. Profile: Pharmaceutical Research Industry. PhRMA Website. 2013. Available online: <http://www.phrma.org/sites/default/files/pdf/PhRMA%20Profile%202013.pdf> (accessed on 6 March 2014).
31. Herper, M. The Cost of Creating a New Drug Now \$5 Billion, Pushing Big Pharma to Change. Forbes Website. Available online: <http://www.forbes.com/sites/matthewherper/2013/08/11/how-the-staggering-cost-of-inventing-new-drugs-is-shaping-the-future-of-medicine/> (accessed on 6 March 2014).
32. Katara, P. Role of bioinformatics and pharmacogenomics in drug discovery and development process. *Netw. Model. Anal. Health Inform. Bioinform.* **2013**, *2*, 225–230. [CrossRef]
33. Expert Committee on the, D.; Classification of Diabetes, M. Report of the expert committee on the diagnosis and classification of diabetes mellitus. *Diabetes Care* **2002**, *26*, S5–S20.
34. World Health Organization. WHO/CONRAD Technical Consultation on Nonoxynol-9, World Health Organization, Geneva, 9–10 October 2001: Summary Report. *Reprod. Health Matters* **2002**, *10*, 175–181. [CrossRef]
35. Diabetes: The Cost of Diabetes. Retrieved. Available online: <http://www.who.int/mediacentre/factsheets/fs236/en/> (accessed on 16 January 2005).
36. Spengler, S.J. Techview: Computers and biology. Bioinformatics in the information age. *Science* **2000**, *287*, 1221–1223. [CrossRef]
37. Thippakorn, C.; Schaduangrat, N.; Nantasenamat, C. Proteomic and bioinformatic discovery of biomarkers for diabetic nephropathy. *EXCLI J.* **2018**, *17*, 312–330. [PubMed]
38. Perez-Riverol, Y.; Csordas, A.; Bai, J.; Bernal-Llinares, M.; Hewapathirana, S.; Kundu, D.J.; Inuganti, A.; Griss, J.; Mayer, G.; Eisenacher, M.; et al. The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* **2018**, *47*, D442–D450. [CrossRef] [PubMed]
39. The Diabetes Control and Complications Trial Research Group. The Effect of Intensive Treatment of Diabetes on the Development and Progression of Long-term Complications in Insulin-dependent Diabetes Mellitus. *N. Engl. J. Med.* **1994**, *14*, 286–287.
40. Chen, C.; Huang, H.; Wu, C.H. Protein Bioinformatics Databases and Resources. *Methods Mol. Biol.* **2017**, *1558*, 3–39.
41. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Lipman, D.J.; Ostell, J.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2013**, *42*, D32–D37. [CrossRef]
42. Brosch, M.; Yu, L.; Hubbard, T.; Choudhary, J. Accurate and Sensitive Peptide Identification with Mascot Percolator. *J. Proteome Res.* **2009**, *8*, 3176–3181. [CrossRef]
43. Wishart, D.S.; Feunang, Y.D.; Marcu, A.; Guo, A.C.; Liang, K.; Vázquez-Fresno, R.; Sajed, T.; Johnson, D.; Li, C.; Karu, N.; et al. HMDB 4.0: The human metabolome database for 2018. *Nucleic Acids Res.* **2018**, *46*, D608–D617. [CrossRef]
44. Yeats, C.; Maibaum, M.; Marsden, R.; Dibley, M.; Lee, D.; Addou, S.; Orengo, C.A. Gene3D: Modelling protein structure, function and evolution. *Nucleic Acids Res.* **2006**, *34*, D281–D284. [CrossRef]
45. Klein, J.; Jupp, S.; Moulos, P.; Fernandez, M.; Buffin-Meyer, B.; Casemayou, A.; Chaaya, R.; Charonis, A.; Bascands, J.L.; Stevens, R.; et al. The KUPKB: A novel Web application to access multiomics data on kidney disease. *FASEB J.* **2012**, *26*, 2145–2153. [CrossRef] [PubMed]
46. Attwood, T.K. PRINTS and its automatic supplement, prePRINTS. *Nucleic Acids Res.* **2003**, *31*, 400–402. [CrossRef]
47. Zhao, M.; Li, M.; Yang, Y.; Guo, Z.; Sun, Y.; Shao, C.; Li, M.; Sun, W.; Gao, Y. A comprehensive analysis and annotation of human normal urinary proteome. *Sci. Rep.* **2017**, *7*, 3024. [CrossRef]
48. Jeyaraman, M.M.; Al-Yousif, N.S.H.; Singh Mann, A.; Dolinsky, V.W.; Rabbani, R.; Zarychanski, R.; Abou-Setta, A.M. Resveratrol for adults with type 2 diabetes mellitus. *Cochrane Database Syst. Rev.* **2020**, *1*, CD011919. [CrossRef] [PubMed]
49. Khetan, S.; Kursawe, R.; Youn, A.; Lawlor, N.; Jillette, A.; Marquez, E.J.; Ucar, D.; Stitzel, M.L. Type 2 Diabetes-Associated Genetic Variants Regulate Chromatin Accessibility in Human Islets. *Diabetes* **2018**, *67*, 2466–2477. [CrossRef] [PubMed]

50. Barale, M.S.; Shirke, D.T. Cascaded Modeling for PIMA Indian Diabetes Data. *Int. J. Comput. Appl.* **2016**, *139*, 1–4. [[CrossRef](#)]
51. Whetzel, P.L.; Grethe, J.S.; Banks, D.E.; Martone, M.E. The NIDDK Information Network: A Community Portal for Finding Data, Materials, and Tools for Researchers Studying Diabetes, Digestive, and Kidney Diseases. *PLoS ONE* **2015**, *10*, e0136206. [[CrossRef](#)]
52. Nesvizhskii, A.I.; Keller, A.; Kolker, E.; Aebersold, R. A Statistical Model for Identifying Proteins by Tandem Mass Spectrometry. *Anal. Chem.* **2003**, *75*, 4646–4658. [[CrossRef](#)]
53. Hoogland, C.; Mostaguir, K.; Sanchez, J.-C.; Hochstrasser, D.F.; Appel, R.D. SWISS-2DPAGE, ten years later. *Proteomics* **2004**, *4*, 2352–2356. [[CrossRef](#)]
54. Burley, S.K.; Berman, H.M.; Bhikadiya, C.; Bi, C.; Chen, L.; Di Costanzo, L.; Christie, C.; Duarte, J.M.; Dutta, S.; Feng, Z.; et al. Protein Data Bank: The single global archive for 3D macromolecular structure data. *Nucleic Acids Res.* **2019**, *47*, D520–D528.
55. Szklarczyk, D.; Franceschini, A.; Wyder, S.; Forslund, K.; Heller, D.; Huerta-Cepas, J.; Simonovic, M.; Roth, A.; Santos, A.; Tsafou, K.P.; et al. STRING v10: Protein-protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* **2014**, *43*, D447–D452. [[CrossRef](#)] [[PubMed](#)]
56. Wilson, D.; Pethica, R.; Zhou, Y.; Talbot, C.; Vogel, C.; Madera, M.; Chothia, C.; Gough, J. SUPERFAMILY—sophisticated comparative genomics, data mining, visualization and phylogeny. *Nucleic Acids Res.* **2009**, *37*, D380–D386. [[CrossRef](#)] [[PubMed](#)]
57. Kim, S.; Thiessen, P.A.; Bolton, E.E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B.A.; et al. PubChem Substance and Compound databases. *Nucleic Acids Res.* **2015**, *44*, D1202–D1213. [[CrossRef](#)]
58. Karnovsky, A.; Weymouth, T.; Hull, T.; Tarcea, V.G.; Scardoni, G.; Laudanna, C.; Sartor, M.A.; Stringer, K.A.; Jagadish, H.V.; Burant, C.; et al. Metscape 2 bioinformatics tool for the analysis and visualization of metabolomics and gene expression data. *Bioinformatics* **2012**, *28*, 373–380. [[CrossRef](#)]
59. Duren, W.; Weymouth, T.; Hull, T.; Omenn, G.S.; Athey, B.; Burant, C.; Karnovsky, A. MetDisease—Connecting metabolites to diseases via literature. *Bioinformatics* **2014**, *30*, 2239–2241. [[CrossRef](#)]
60. Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462. [[CrossRef](#)]
61. Thomas, P.D.; Campbell, M.J.; Kejariwal, A.; Mi, H.; Karlak, B.; Daverman, R.; Diemer, K.; Muruganujan, A.; Narechania, A. PANTHER: A Library of Protein Families and Subfamilies Indexed by Function. *Genome Res.* **2003**, *13*, 2129–2141. [[CrossRef](#)]
62. Wang, J.; Vasaiakar, S.V.; Shi, Z.; Greer, M.; Zhang, B. WebGestalt 2017: A more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* **2017**, *45*, W130–W137. [[CrossRef](#)]
63. Smigielski, E.M.; Sirotkin, K.; Ward, M.; Sherry, S.T. dbSNP: A database of single nucleotide polymorphisms. *Nucleic Acids Res.* **2000**, *28*, 352–355. [[CrossRef](#)]
64. Fan, Y.; Habib, M.; Xia, J. Xeno-miRNet: A comprehensive database and analytics platform to explore xeno-miRNAs and their potential targets. *PeerJ* **2018**, *6*, e5650. [[CrossRef](#)] [[PubMed](#)]
65. Chen, J.; Mamidipalli, S.; Huan, T. HAPPY: An online database of comprehensive human annotated and predicted protein interactions. *BMC Genom.* **2009**, *10*, S1–S16. [[CrossRef](#)] [[PubMed](#)]
66. Karolchik, D.; Hinrichs, A.S.; Kent, W.J.; Speir, M.L.; Zweig, A.S.; Rosenbloom, K.R.; Raney, B.J.; Paten, B.; Nejad, P.; Lee, B.T.; et al. The UCSC Genome Browser. *Curr. Protoc. Bioinform.* **2012**, *1*, 1–4. [[CrossRef](#)] [[PubMed](#)]
67. Gutiérrez-Aguilar, R.; Benmezroua, Y.; Vaillant, E.; Balkau, B.; Marre, M.; Charpentier, G.; Sladek, R.; Froguel, P.; Neve, B. Analysis of KLF transcription factor family gene variants in type 2 diabetes. *BMC Med Genet.* **2007**, *8*, 53. [[CrossRef](#)] [[PubMed](#)]
68. Ashurst, J.L.; Chen, C.K.; Gilbert, J.G.; Jekosch, K.; Keenan, S.; Meidl, P.; Searle, S.M.; Stalker, J.; Storey, R.; Trevanion, S.; et al. The Vertebrate Genome Annotation (Vega) database. *Nucleic Acids Res.* **2005**, *33*, D459–D465. [[CrossRef](#)]
69. Koivula, R.W.; Heggie, A.; Barnett, A.; Cederberg, H.; Hansen, T.H.; Koopman, A.D.; Ridderstråle, M.; Rutters, F.; Vestergaard, H.; Gupta, R.; et al. Discovery of biomarkers for glycaemic deterioration before and after the onset of type 2 diabetes: Rationale and design of the epidemiological studies within the IMI DIRECT Consortium. *Diabetologia* **2014**, *57*, 1132–1142. [[CrossRef](#)]

70. Lee, S.; Choi, M.E. Urinary biomarkers for early diabetic nephropathy: Beyond albuminuria. *Pediatr. Nephrol.* **2015**, *30*, 1063–1075. [[CrossRef](#)]
71. Pagtalunan, M.E.; Miller, P.L.; Jumping-Eagle, S.; Nelson, R.G.; Myers, B.D.; Rennke, H.G.; Coplon, N.S.; Sun, L.; Meyer, T.W. Podocyte loss and progressive glomerular injury in type II diabetes. *J. Clin. Investig.* **1997**, *99*, 342–348. [[CrossRef](#)]
72. Mauer, S.M. Structural-functional correlations of diabetic nephropathy. *Kidney Int.* **1994**, *45*, 612–622. [[CrossRef](#)]
73. Chavers, B.M.; Bilous, R.W.; Ellis, E.N.; Steffes, M.W.; Mauer, S.M. Glomerular Lesions and Urinary Albumin Excretion in Type I Diabetes without Overt Proteinuria. *N. Engl. J. Med.* **1989**, *320*, 966–970. [[CrossRef](#)]
74. Mogensen, C.E.; Christensen, C.K.; Vittinghus, E. The Stages in Diabetic Renal Disease: With Emphasis on the Stage of Incipient Diabetic Nephropathy. *Diabetes* **1983**, *32*, 64–78. [[CrossRef](#)]
75. Satirapoj, B.; Adler, S.G. Comprehensive approach to diabetic nephropathy. *Kidney Res. Clin. Pr.* **2014**, *33*, 121–131. [[CrossRef](#)] [[PubMed](#)]
76. Starkey, J.M.; Tilton, R.G. Proteomics and systems biology for understanding diabetic nephropathy. *J. Cardiovasc. Transl. Res.* **2012**, *5*, 479–490. [[CrossRef](#)] [[PubMed](#)]
77. Senatorski, G.; Paczek, L.; Kropiewnicka, E.; Bartłomiejczyk, I. [Cytokines in noninvasive diagnostics of diabetic nephropathy progression]. *Polski Merkur. Lek. Organ Polskiego Towar. Lek.* **2002**, *13*, 28–32.
78. Merchant, M.L.; Perkins, B.A.; Boratyn, G.M.; Ficociello, L.H.; Wilkey, D.W.; Barati, M.T.; Bertram, C.C.; Page, G.P.; Rovin, B.H.; Warram, J.H.; et al. Urinary peptidome may predict renal function decline in type 1 diabetes and microalbuminuria. *J. Am. Soc. Nephrol.* **2009**, *20*, 2065–2074. [[CrossRef](#)]
79. Wołkow, P.P.; Niewczas, M.A.; Perkins, B.; Ficociello, L.H.; Lipinski, B.; Warram, J.H.; Krolewski, A.S. Association of urinary inflammatory markers and renal decline in microalbuminuric type 1 diabetics. *J. Am. Soc. Nephrol.* **2008**, *19*, 789–797. [[CrossRef](#)] [[PubMed](#)]
80. Mischak, H.; Kaiser, T.; Walden, M.; Hillmann, M.; Wittke, S.; Herrmann, A.; Knueppel, S.; Haller, H.; Fliser, D. Proteomic analysis for the assessment of diabetic renal damage in humans. *Clin. Sci.* **2004**, *107*, 485–495. [[CrossRef](#)]
81. Good, D.M.; Zürbig, P.; Argilés, A.; Bauer, H.W.; Behrens, G.; Coon, J.J.; Dakna, M.; Decramer, S.; Delles, C.; Dominiczak, A.; et al. Naturally Occurring Human Urinary Peptides for Use in Diagnosis of Chronic Kidney Disease *. *Mol. Cell. Proteom.* **2010**, *9*, 2424–2437. [[CrossRef](#)]
82. Fan, Y.; Yi, Z.; D’Agati, V.D.; Sun, Z.; Zhong, F.; Zhang, W.; Wen, J.; Zhou, T.; Li, Z.; He, L.; et al. Comparison of Kidney Transcriptomic Profiles of Early and Advanced Diabetic Nephropathy Reveals Potential New Mechanisms for Disease Progression. *Diabetes* **2019**, *68*, 2301–2314. [[CrossRef](#)]
83. Dobin, A.; Gingeras, T.R. Mapping RNA-seq Reads with STAR. *Curr. Protoc. Bioinform.* **2015**, *51*, 1–19. [[CrossRef](#)]
84. Costa-Silva, J.; Domingues, D.; Lopes, F.M. RNA-Seq differential expression analysis: An extended review and a software tool. *PLoS ONE* **2017**, *12*, e0190152. [[CrossRef](#)] [[PubMed](#)]
85. The Gene Ontology, C. Expansion of the Gene Ontology knowledgebase and resources. *Nucleic Acids Res.* **2016**, *45*, D331–D338.
86. Newman, A.M.; Liu, C.L.; Green, M.R.; Gentles, A.J.; Feng, W.; Xu, Y.; Hoang, C.D.; Diehn, M.; Diehn, M. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* **2015**, *12*, 453–457. [[CrossRef](#)] [[PubMed](#)]
87. Clough, E.; Barrett, T. The Gene Expression Omnibus Database. *Breast Cancer* **2016**, *1418*, 93–110.
88. Maere, S.; Heymans, K.; Kuiper, M. BiNGO: A Cytoscape plugin to assess overrepresentation of Gene Ontology categories in Biological Networks. *Bioinformatics* **2005**, *21*, 3448–3449. [[CrossRef](#)] [[PubMed](#)]
89. Merico, D.; Isserlin, R.; Stueker, O.; Emili, A.; Bader, G.D. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLoS ONE* **2010**, *5*, e13984. [[CrossRef](#)]
90. Bonomo, J.A.; Guan, M.; Ng, M.C.; Palmer, N.D.; Hicks, P.J.; Keaton, J.M.; Lea, J.P.; Langefeld, C.D.; Freedman, B.I.; Bowden, D.W. The ras responsive transcription factor RREB1 is a novel candidate gene for type 2 diabetes associated end-stage kidney disease. *Hum. Mol. Genet.* **2014**, *23*, 6441–6447. [[CrossRef](#)]
91. Pezzolesi, M.G.; Poznik, G.D.; Mychaleckyj, J.C.; Paterson, A.D.; Barati, M.T.; Klein, J.B.; Ng, D.P.; Placha, G.; Canani, L.H.; Bochenski, J.; et al. Genome-Wide Association Scan for Diabetic Nephropathy Susceptibility Genes in Type 1 Diabetes. *Diabetes* **2009**, *58*, 1403–1410. [[CrossRef](#)] [[PubMed](#)]

92. Sandholm, N.; Salem, R.M.; McKnight, A.J.; Brennan, E.P.; Forsblom, C.; Isakova, T.; McKay, G.J.; Williams, W.W.; Sadlier, D.M.; Mäkinen, V.-P.; et al. New Susceptibility Loci Associated with Kidney Disease in Type 1 Diabetes. *PLoS Genet.* **2012**, *8*, e1002921. [[CrossRef](#)]
93. Väre, L.; Scheele, C.; Broholm, C.; Mardinoglu, A.; Kampf, C.; Asplund, A.; Nookaew, I.; Uhlen, M.; Pedersen, B.K.; Nielsen, J. Proteome- and Transcriptome-Driven Reconstruction of the Human Myocyte Metabolic Network and Its Use for Identification of Markers for Diabetes. *Cell Rep.* **2015**, *11*, 921–933. [[CrossRef](#)]
94. Marinković, T.; Oresic, M. Modeling strategies to study metabolic pathways in progression to type 1 diabetes—Challenges and opportunities. *Arch. Biochem. Biophys.* **2016**, *589*, 131–137. [[CrossRef](#)]
95. Abedi, M.; Gheisari, Y. Nodes with high centrality in protein interaction networks are responsible for driving signaling pathways in diabetic nephropathy. *PeerJ* **2015**, *3*, e1284. [[CrossRef](#)] [[PubMed](#)]
96. Saito, R.; Rocanin-Arjo, A.; You, Y.-H.; Darshi, M.; Van Espen, B.; Miyamoto, S.; Pham, J.; Pu, M.; Romoli, S.; Natarajan, L.; et al. Systems biology analysis reveals role of MDM2 in diabetic nephropathy. *JCI Insight* **2016**, *1*, e87877. [[CrossRef](#)] [[PubMed](#)]
97. Uhlen, M.; Oksvold, P.; Fagerberg, L.; Lundberg, E.; Jonasson, K.; Forsberg, M.; Zwahlen, M.; Kampf, C.; Wester, K.; Hober, S.; et al. Towards a knowledge-based Human Protein Atlas. *Nat. Biotechnol.* **2010**, *28*, 1248–1250. [[CrossRef](#)] [[PubMed](#)]
98. Huang, G.-M.; Huang, K.-Y.; Lee, T.-Y.; Weng, J.T.-Y. An interpretable rule-based diagnostic classification of diabetic nephropathy among type 2 diabetes patients. *BMC Bioinform.* **2015**, *16*, S5. [[CrossRef](#)]
99. Cho, B.H.; Yu, H.; Kim, K.-W.; Kim, T.H.; Kim, I.Y.; Kim, S.I. Application of irregular and unbalanced data to predict diabetic nephropathy using visualization and feature selection methods. *Artif. Intell. Med.* **2008**, *42*, 37. [[CrossRef](#)]
100. Dubrava, S.; Mardekian, J.; Sadosky, A.; Bienen, E.J.; Parsons, B.; Hopps, M.; Markman, J. Using Random Forest Models to Identify Correlates of a Diabetic Peripheral Neuropathy Diagnosis from Electronic Health Record Data. *Pain Med.* **2016**, *18*, 107–115. [[CrossRef](#)]



© 2020 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 (<http://creativecommons.org/licenses/by/4.0/>).