



# Network Analyses Based on Machine Learning Methods to Quantify Effects of Peptide–Protein Complexes as Drug Targets Using Cinnamon in Cardiovascular Diseases and Metabolic Syndrome as a Case Study

Yingying Wang<sup>1,2</sup>, Lili Wang<sup>3</sup>, Yinhe Liu<sup>3</sup>, Keshen Li<sup>1,2\*</sup> and Honglei Zhao<sup>3\*</sup>

<sup>1</sup>Department of Neurology and Stroke Center, The First Affiliated Hospital of Jinan University, Guangzhou, China, <sup>2</sup>Clinical Neuroscience Institute, The First Affiliated Hospital of Jinan University, Guangzhou, China, <sup>3</sup>Fuwai Hospital Chinese Academy of Medical Sciences, Shenzhen, China

## OPEN ACCESS

### Edited by:

Ruiquan Ge,  
Hangzhou Dianzi University, China

### Reviewed by:

Shihua Zhang,  
Wuhan University of Science and  
Technology, China  
Lei Liu,  
Harbin Medical University, China

### \*Correspondence:

Keshen Li  
likeshen1971@126.com  
Honglei Zhao  
939924240@qq.com

### Specialty section:

This article was submitted to  
Computational Genomics,  
a section of the journal  
Frontiers in Genetics

**Received:** 16 November 2021

**Accepted:** 07 December 2021

**Published:** 24 December 2021

### Citation:

Wang Y, Wang L, Liu Y, Li K and  
Zhao H (2021) Network Analyses  
Based on Machine Learning Methods  
to Quantify Effects of Peptide–Protein  
Complexes as Drug Targets Using  
Cinnamon in Cardiovascular Diseases  
and Metabolic Syndrome as a  
Case Study.  
Front. Genet. 12:816131.  
doi: 10.3389/fgene.2021.816131

Peptide–protein complexes play important roles in multiple diseases such as cardiovascular diseases (CVDs) and metabolic syndrome (MetS). The peptides may be the key molecules in the designing of inhibitors or drug targets. Many Chinese traditional drugs are shown to play various roles in different diseases, and comprehensive analyses should be performed using networks which could offer more information than results generated from a single level. In this study, a network analysis pipeline was designed based on machine learning methods to quantify the effects of peptide–protein complexes as drug targets. Three steps, namely, pathway filter, combined network construction, and biomarker prediction and validation based on peptides, were performed using cinnamon (CA) in CVDs and MetS as a case. Results showed that 17 peptide–protein complexes including six peptides and four proteins were identified as CA targets. The expressions of AKT1, AKT2, and ENOS were tested using qRT-PCR in a mouse model that was constructed. AKT2 was shown to be a CA-indicating biomarker, while E2F1 and ENOS were CA treatment targets. AKT1 was considered a diabetic responsive biomarker because it was down-regulated in diabetic but not related to CA. Taken together, the pipeline could identify new drug targets based on biological function analyses. This may provide a deep understanding of the drugs' roles in different diseases which may foster the development of peptide–protein complex–based therapeutic approaches.

**Keywords:** peptide–protein complexes, network analyses, metabolic syndrome, cardiovascular, cinnamon

## INTRODUCTION

Peptide–protein complexes are the key components of protein–protein interaction (PPI) networks. Nearly 15–40% PPIs are mediated by these short linear peptides (Neduva et al., 2005). The peptide–protein complexes are proven to play important roles predominantly in both signaling and regulatory pathways, implicating that the peptides are involved in many human diseases

(Pawson and Nash, 2003). As a result, the peptides are attracting more attention in drug research fields since they may be the key molecules in the designing of inhibitors or drug targets (Parthasarathi et al., 2008; London et al., 2010).

Due to the characters of peptide-protein interactions, it is reasonable to perform network analyses based on machine learning methods since the relationships between those peptides and proteins could be illustrated clearly in the form of graphs (Zhu et al., 2020; Ji et al., 2021; Yingying et al., 2021). Similarly, some complex diseases are found to be similar based on network analyses, indicating that more relationships between different diseases could be predicted using bioinformatics pipelines (Wang et al., 2019). Many Chinese traditional drugs are shown to play various roles in different diseases, and comprehensive analyses should be performed using networks which may offer more information than results generated from a single level. However, no pipeline aiming to predict the peptide-protein complex as drug targets in different diseases had been proposed. In this study, a network analysis pipeline was designed based on machine learning methods to quantify the effects of peptide-protein complexes as drug targets.

In this pipeline, diseases with at least 20 related genes and drugs with at least one related biological functional term could be used as analysis objects. Diseases that are similar to each other on at least one level (such as medical or biological level) are recommended. The candidate drugs do not need to be proven useful in the diseases analyzed since predicting new roles of the candidate drugs is also one application of the pipeline. Based on the abovementioned concerns, two types of diseases (cardiovascular diseases (CVDs) and metabolic syndrome (MetS)) and a Chinese traditional drug (cinnamon) as a case were chosen.

Cinnamon (*Cinnamomum zeylanicum* and Cinnamon cassia, CA) is one of the most important spices used daily (Hariri and Ghiasvand, 2016). Cinnamaldehyde is one of the main resinous ingredients found in CA, which is commonly used as a Chinese medicine for blood circulation disturbance and inflammation (Sheng et al., 2008; Cao et al., 2010; Yang et al., 2015). It was shown that cinnamaldehyde played important roles in both CVDs and MetS (patients suffering from type 2 diabetes (T2D), and glucose/insulin metabolism disturbance or insulin resistance, and was involved with at least two of the following four items: hypertension, dyslipidemia, obesity, and microalbuminuria defined by the WHO criteria) (Mollazadeh and Hosseinzadeh, 2016). CVDs and MetS are not independent since MetS is one of the most undeniable reasons of CVDs. Besides, there are multiple types of biomarkers identified as common features of CVDs and MetS, such as non-coding RNAs, proteins, and metabolites (Das et al., 2020).

It is of great importance to explore the mechanism of CA since this drug could participate in both of the disease types at the same time (Sheng et al., 2008; Yang et al., 2015). One possible reason may be its antidiabetic action by modulating the insulin and insulin-like growth factor (IGF1) signaling pathways (Schriner et al., 2014) since insulin resistance was

proven to play a fundamental key role for MetS complications (Khan et al., 1990). Besides, CA was shown to retard the progression of cardiac hypertrophy and fibrosis *via* blocking the ERK signaling pathway (Zhang et al., 2015; Xiao et al., 2017). However, functional analysis for CA in a system way, especially based on biological pathways, is still lacking.

As an integration of molecular interaction; genetic, cellular, and environmental information processing; and metabolism reactions, biological pathways are often used in systematic analyses of complex diseases such as CVDs, T2D, and cancers (Salt and Hardie, 2017; Kakiuchi-Kiyota et al., 2019; Kaku, 2019). Peptide-protein complexes were also proven to be the key components in pathways. It was postulated that there may be associations between the common pathways shared by CVD/MetS and CA which could be detected based on peptide-protein complex analyses. In this study, a new network analysis pipeline was proposed based on machine learning methods to identify common drug targets in different diseases.

## MATERIALS AND METHODS

The analyses were performed using the following three steps: (as shown in **Figure 1**).

Step1: Pathway filter. The similarity between any two selected diseases was calculated and used to filter the disease pairs. Enrichment analyses were performed for the related genes of the disease pairs. Meanwhile, the CA-related pathways were found through literature searching. Common pathways were then filtered and used as the inputs for step 2.

Step2: Combined network construction. All the common pathways were then converted into networks. The network structure similarities were calculated using two types of machine learning methods, and an integrate score was designed to measure the similarity between any two common pathways on the structural level in order to explore the potential correlations of these pathways. The pathways were then merged into a combined pathway network. Proteins in the pathways were merged into a combined protein network.

Step3: Biomarker prediction and validation based on peptides. The nodes in the combined protein network were first ranked according to the network topological characters. Then protein-peptide complexes containing these top proteins as receptors were selected, and the peptides were then clustered. The top genes with peptides clustered into the same clusters were selected as candidate biomarkers and validated using qRT-PCR in a mouse diabetic model that was constructed.

### Disease Similarity Calculation

Methods that can calculate the distances between any two diseases based on any biological or medical level could be used. In this case, a module-based method (Menche et al., 2015) was used. The similarity  $S_{ij}$  between two diseases  $i$  and  $j$  was calculated as follows:

$$S_{ij} \equiv \langle d_{ij} \rangle - \frac{\langle d_{ii} \rangle + \langle d_{jj} \rangle}{2}.$$

Of which,  $\langle d_{ii} \rangle$  and  $\langle d_{jj} \rangle$  represented the average shortest distances inside diseases  $i$  and  $j$ , respectively, while  $\langle d_{ij} \rangle$  represented the pairwise average shortest distance between disease  $i$  and  $j$ . The shortest distances were calculated for any protein pairs inside/between diseases using the relationships integrated from multiple molecular interaction levels including protein, regulatory, and metabolic pathways, and kinase substrate.

A z-score was calculated based on the random control networks by 1000 permutations of disease lists preserving randomization. A  $p$ -value for each  $S_{ij}$  score was calculated using the Mann-Whitney U test. Then FDR was used to obtain the  $q$ -values.

### Information Converting From Genes to Biological Pathways

The information conversions from genes to biological pathways were performed using the DAVID EASE score (Huang et al., 2009a; Huang et al., 2009b), which was a modified Fisher exact  $p$ -value. For any disease-related gene list  $l_i$  and biological pathway  $w_a$ , the EASE score was calculated as follows:

$$e(l_i, w_a) = 1 - \sum_{i=0}^{GH-1} \frac{\binom{OH}{i} \binom{OT-OH}{GT-i}}{\binom{OT}{GT-1}}.$$

Of which, the calculation methods of GH (gene hits), GT (gene total), OH (genome hits), and OT (genome total) are shown in the following 2\*2 table:

An  $e$ -value not above the threshold supported the alternative hypothesis that the probability of the first cell in the 2\*2 table was actually greater than that expected under the null hypothesis that the two variables were independent. The conclusion was that there was an association between the row and the column variables in the table, which meant the proportions of those genes falling into each category were different among groups.

### From Biological Pathways to Graphs

The information conversions from biological pathways to protein-protein networks were performed using the R package “graphite” (Sales et al., 2012). The algorithm in this package kept the information of protein complexes, gene families, and

removing chemical compounds from the final graphs, which was especially important in the peptide complex analyses of this study.

### Network Structure Similarity Calculation

The network structure similarity calculation algorithms could be divided into two types: alignment-free and alignment-based network comparison (Frigo et al., 2021). In this pipeline, it was recommended to use at least one alignment-free algorithm and one alignment-based algorithm to compare the different networks and combine the scores together.

### Alignment-Free Algorithm Based on Graphlet Degree Distribution Agreement

The alignment-free network comparison algorithms performed the network similarity analyses by quantifying the overall topological similarity between networks, irrespective of node mappings between the networks, and without any conserved edges or subgraph identification. In this pipeline, the algorithm named GDD agreement was chosen, which performed the structural similarity (SS) between networks based on the graphlet degree distribution as follows (Przulj, 2007):

The similarity between any two networks  $G$  and  $W$  was calculated as follows:

$$S_{GDD}(G, W) = \frac{1}{n} \sum_{j=0}^{n-1} S_{GDD}^j(G, W).$$

Of which,

$$S_{GDD}^j(G, W) = 1 - \left( \sum_{k=1}^{\infty} \left[ \frac{d_G^j(k)}{k} / \sum_{k=1}^{\infty} \frac{d_G^j(k)}{k} - \frac{d_W^j(k)}{k} / \sum_{k=1}^{\infty} \frac{d_W^j(k)}{k} \right]^2 \right)^{\frac{1}{2}}.$$

Of which,  $d_G^j(k)$  is the sample distribution of the number of nodes in network  $G$  touching the appropriate graphlet  $k$  times. The range of  $S_{GDD}$  is [0,1]; a higher score meant the two networks compared were more similar to each other.

### Alignment-Based Algorithm Based on the Hungarian Method

The alignment-based network comparison methods referred to a series of algorithms aiming to find a mapping between the nodes of at least two networks that preserved edges and a large subgraph between the networks. In this pipeline, an alignment-based algorithm was chosen based on a Hungarian method as follows:

	Number of genes in $l_i$	Number of genes in the genome
Number of genes in $w_a$	GH-1	OH-GH+1
Number of genes not in $w_a$	GT-GH	OT-GT-(OH-GH)

The network alignment scores, that is,  $S_{AE}(G, W)$  (between any two networks  $G$  and  $W$ ), were performed using the Hungarian method (Kuhn, 1955) on a square distance matrix  $C$  (if the sizes of the two networks were different, the larger number of nodes was used), which was calculated as follows:

$$C_{ab} = \sqrt{\sum_{t \in T} (M_{G_{at}} - M_{W_{bt}})^2}.$$

Of which,  $M_{G_{at}} = \frac{-\sum_{j=1}^{N_G} A_{G_{a,j}}^{(t)} \ln(A_{G_{a,j}}^{(t)})}{H(I^{N_G})}$ , where  $A_{G_{a,j}}^{(t)}$  is a transition matrix of network  $G$ , which was constructed by converting the raw square transition matrix into Markov processes by normalizing each row sum to unity.  $A_{G_{a,j}}^{(t)}$  contained probabilities of edges transferring information from the  $i$ th to the  $j$ th member of the system in exactly  $t$  units of time. For  $t \in \{2^y\} \forall y \in \mathbb{N}$ ,  $t_{\max} \geq 2D$  and  $t_{\max-1} < 2D$ , where  $D$  is the max diameter of the two networks  $G$  and  $W$  being compared, and the R packages “igraph” and “netcom” were used to perform the calculation.

## Integrated Network Similarity Score

The integrated network similarity scores between the two networks  $G$  and  $W$  were calculated as follows:

$$S(G, W) = S_{GDD}(G, W) + S_{AE}(G, W).$$

A higher  $S$  score indicated that the two networks compared were more similar to each other on the structural level.

## Biomarker Prediction and Validation

The prediction and validation of the biomarkers were performed using the following steps:

- 1) The proteins in the combined protein network were ranked according to the network topological characters. For each node, degree and node betweenness were calculated. The edge betweenness was calculated for each edge using the R package “igraph.”
- 2) The protein-peptide complexes containing these top proteins as receptors were selected, and the peptides were then clustered. The high-resolution structures of protein-peptide complexes containing genes in the combined network as receptors were downloaded from the Protein Data Bank (PDB).
- 3) The top proteins with peptides clustered into the same clusters were selected as candidate biomarkers. The peptide sequences of these complexes were then classified using Hammock (1.2.0) (Krejci et al., 2016), which used hidden Markov model profiles for peptide sequence clustering. The consensus sequence for each cluster was generated using ClustalW (Thompson et al., 1994) and WebLogo (Crooks et al., 2004).
- 4) The candidate biomarkers were validated using qRT-PCR in a mouse diabetic model constructed as follows:

Fifty-nine male C57 mice (14–16g/28–35 days) were purchased from Guangdong Medical Experimental Animal

Center (Certificate No.: 44007200062167, License No.: scxk (Guangdong) 2018-0002, SPF clean grade).

The mice were divided into four groups as follows: 1) Group A (Control + vehicle): 5 mice were given solvent control (0.5% carboxymethyl cellulose solution (CMC)) by gavage; 2) Group B (Control + CA): 6 mice were given CA by gavage (the dose was 20 mg/kg/BW); 3) Group C (T2D + vehicle): 24 diabetic mice were given solvent by gavage; 4) Group D (T2D + CA): 24 diabetic mice were given CA by gavage.

Of which, the models of 48 diabetic mice were constructed using streptozotocin (STZ) using the following steps: 1) pretreatment: all the mice were made to starve 12 h before modeling; 2) model construction: STZ was intraperitoneally injected at a dose of 150 mg/kg/BW; 3) model test: the blood glucose value was measured continuously after 3 days of STZ injection. If the random blood glucose was  $>16.7$  mmol/L, the model was considered successful. Otherwise, another injection of STZ was administered until the random blood glucose was  $>16.7$  mmol/L.

Drug treatment (Groups C and D) was started 5 weeks after modeling. After 7 weeks of administration, all animals were killed, and the hearts of mice were treated with TRIzol and stored at  $-80^\circ\text{C}$ . Then qRT-PCR was performed for the candidate genes (the top proteins were mapped to their coding genes). The animal experiment was approved and recognized by the experimental Animal Ethics Committee of Shenzhen Sun Yat sen Cardiovascular Hospital (Approval No.: rye2019102806).

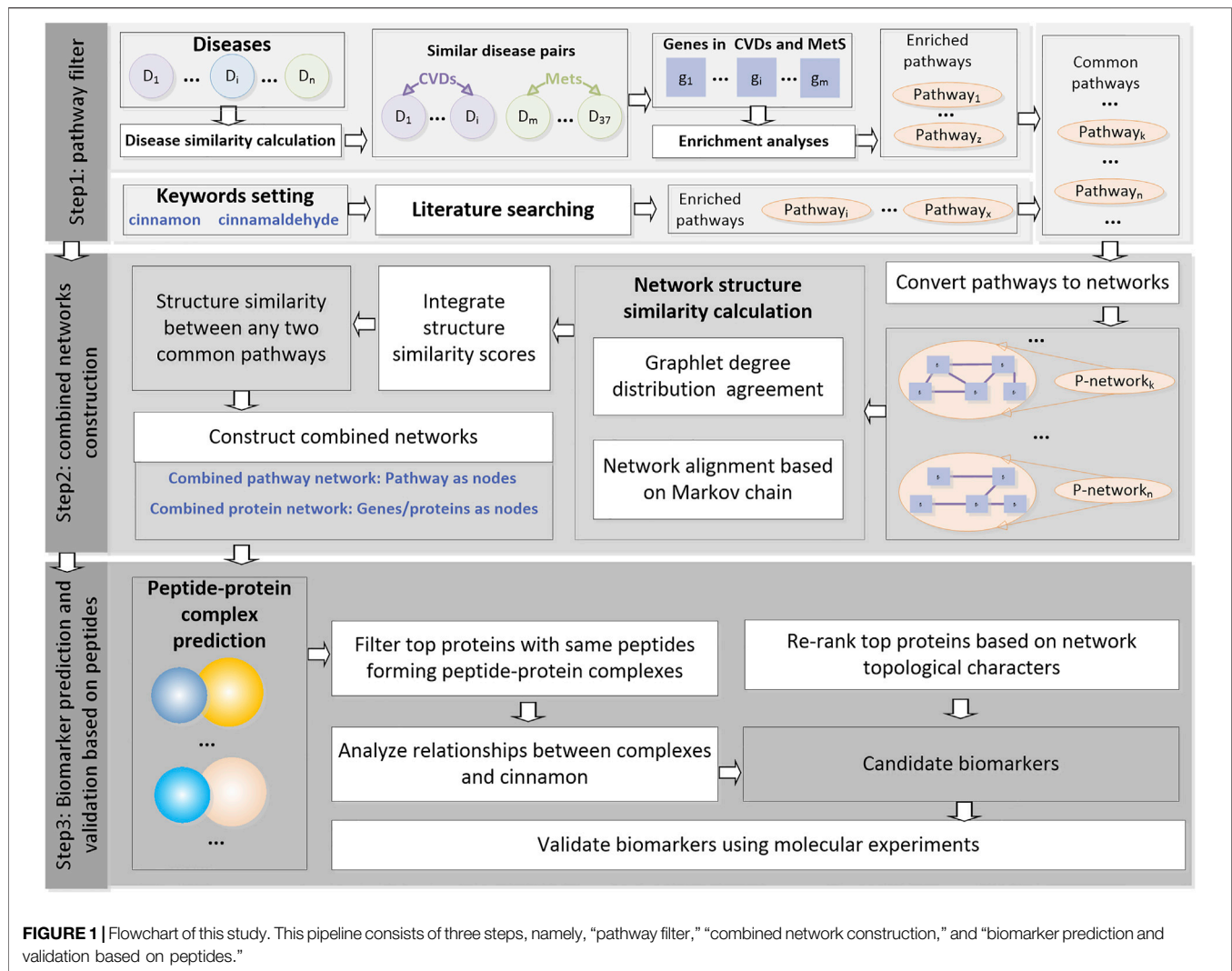
## RESULTS

### Pathway Filtrations

CVD (such as coronary disease) and MetS (such as diabetes mellitus) lists were extracted from Medical Subject Heading (MeSH) ontology with at least 20 disease-related genes from either OMIM or GWAS (listed in **Supplementary Table S1**). 553 disease pairs were shown to be similar with each other with a  $z$ -score  $\geq 1.6$  and  $q$ -value  $\leq 0.001$ . The 19 CVDs and 18 MetS comprising the 553 disease pairs were selected as HM (HeartMetS) datasets. As shown in **Figure 2A**, the average numbers of genes related to MetS (179.0556) were 2-fold of CVDs (86.42105). This indicated that MetS may be more complex than CVDs since these diseases involve the abnormality of multiple systems, such as endocrine, digestive, and immune systems.

The gene lists of each disease were then used as inputs of the information converting calculation. 179 pathways in KEGG (Kanehisa et al., 2017) and Biocarta with at least one  $e$ -value not above 0.05 were selected as HM-enriched pathways (see **Supplementary Table S2**). The common pathways in the two databases were named using KEGG ID. Otherwise, if there exists any difference between the two pathways, both of the pathways were kept.

CA was shown to play important roles through biological pathways in reducing metabolic syndrome complications and CVDs as reviewed in the former research (Yang et al., 2015;



Mollazadeh and Hosseinzadeh, 2016). The words “cinnamon” and “cinnamaldehyde” were used for literature searching through the NCBI PubMed to find the related pathways since 32 CA-related pathways were selected and binned into two groups according to their effects on diseases types: antidiabetic (including 28 pathways) and antihypertensive (including four pathways).

### Combined Pathway Network Analyses

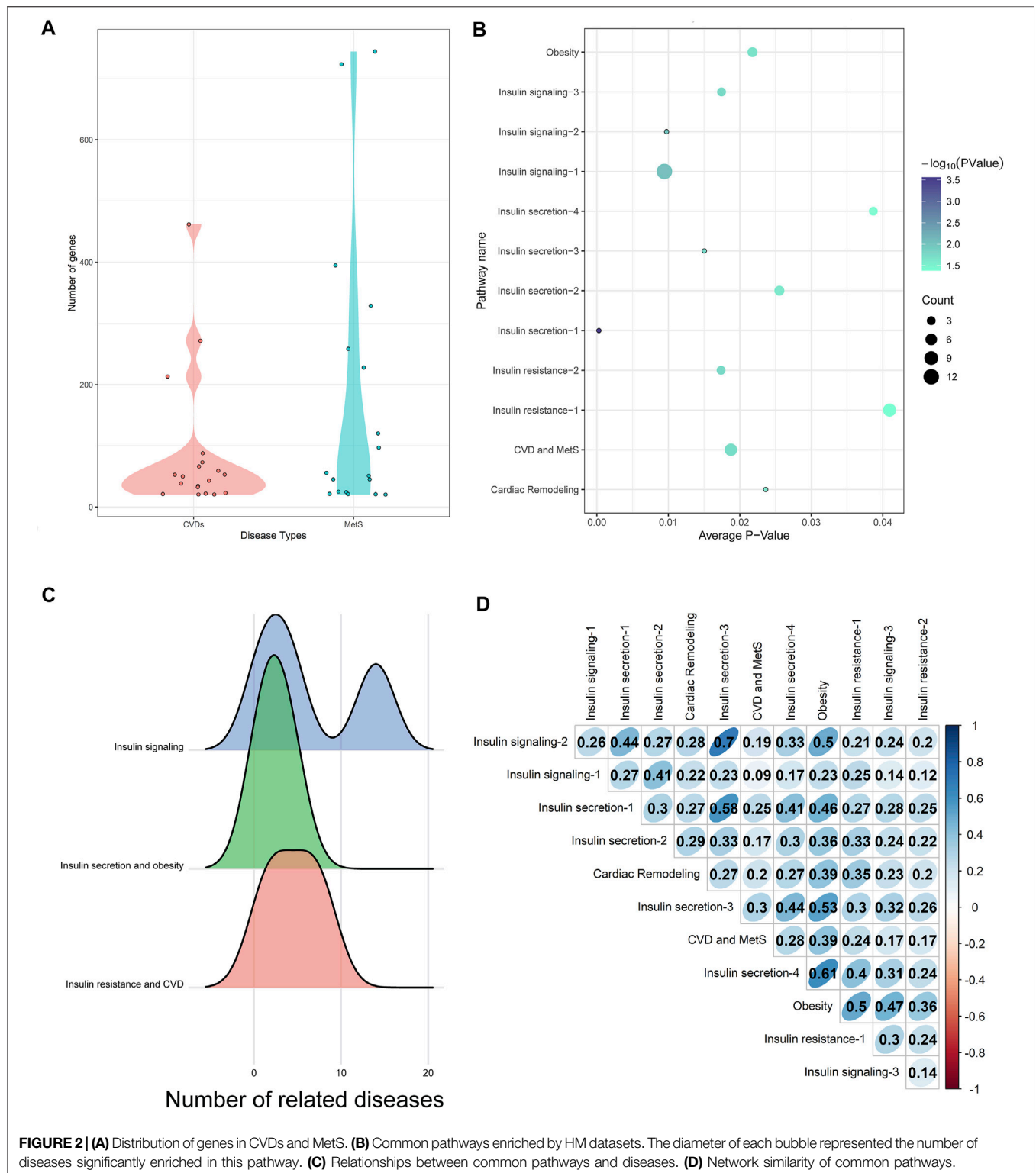
As shown in **Figure 2B**, there were 12 common pathways between the 179 HM-enriched pathways and 32 CA-related pathways. This indicated the dynamical roles CA played in different diseases or disease stages, including diabetes mellitus, obesity, MetS, and CVDs. The combined pathway network was built using the 12 common pathways as nodes. Sixty-seven connections were built if the two pathways shared at least one gene/protein.

The 12 common pathways were enriched by different numbers of diseases in the HM datasets. Of which, the insulin signaling pathway (hsa04910, marked as “Insulin

signaling-1”) was enriched by 12 diseases, while the following three pathways were only enriched by one disease: IL-2 receptor beta chain in T-cell activation (h\_il2rbPathway, marked as “Cardiac Remodeling”), the IGF-1 receptor and longevity (h\_longevityPathway, marked as “Insulin secretion-1”), multiple antiapoptotic pathways from IGF-1r signaling lead to bad phosphorylation (h\_igf1rPathway, marked as “Insulin secretion-3”), and sprouty regulation of tyrosine kinase signals (hsa04911, marked as “Insulin signaling-2”).

The 12 common pathways could be divided into three types according to their contributions to CVD and MetS (as shown in **Figure 2C**).

- 1) Insulin signaling: CA could enhance the insulin signaling pathway in the skeletal muscle by increasing the tyrosine phosphorylation level (Qin et al., 2003). Three pathways were involved in this stage, including the insulin signaling pathway (enriched by MetS and CVDs), tyrosine metabolism (enriched by MetS), and sprout regulation of



tyrosine kinase signals (enriched by CVD). It was interesting to see that the “insulin signaling pathway” was closely connected not only to MetS such as diabetes mellitus but also to CVDs, such as heart diseases. This may be explained by the fact that insulin signaling was an integral pathway

regulating the life span of laboratory organisms (Schriner et al., 2014).  
 2) Insulin secretion and obesity: Since impaired insulin secretion was one of the pathophysiological abnormalities in type 2 diabetes, IGF (insulin-like growth factors)-I, which was

**TABLE 1** | Pathway similarity results of different pathways.

Type of pathway sets	Number of pathways	$S_{GDD}$	$S_{AE}$
Common	12	0.257095099	0.047247541
CA-related	32	0.355600743	0.091677581
HM-related	179	0.324632623	0.061221968

**TABLE 2** | List of top 10 nodes and edges in the combined protein network.

Topological character	Protein symbols/protein-protein pairs
Degree	IRS1, MAOA, MAOB, AMPK1, AMPK2, PRKAB1, PRKAB2, PRKAG1, PRKAG2, and PRKAG3
Node betweenness	IRS1, OGT, AKT2, INS, AKT1, RAPGEF4, INSR, PDE3B, PTPA, and GNAS
Edge betweenness	PTPA-AKT2, IRS1-IGF1R, PPARGC1A-OGT, AKT1-E2F1, IGF1R-RAF1, AKT2-PDE3B, OGT-AKT1, E2F1-IL2RA, PRKCE-INSR, and NOS3-IRS1

shown to inhibit insulin secretion, would play a key role in the process (Leahy and Vandekerckhove, 1990; Pørksen et al., 1997). CA could increase the phosphorylation levels of the IGF-I receptor and its downstream signaling molecules (Takasao et al., 2012). It was interesting that binding IGF-I to its receptor could cause the activation of the tyrosine kinase, leading to autophosphorylation of the intrinsic tyrosines, which transduced the IGF-I signal to a complex network that was ultimately responsible for cell proliferation, modulation of tissue differentiation, and protection from apoptosis (Laviola et al., 2007).

- 3) Insulin resistance and CVD: The study showed that the insulin action on cAMP was severely impaired in insulin-resistant patients (Laviola et al., 2007). The cyclic-AMP signaling pathway was shown to be modulated by CA to exhibit antidiabetic action (Schriner et al., 2014). “Regulation of lipolysis in adipocytes” (marked as “Obesity”) was closely linked to MetS since the variations of insulin resistance severity may be related to the regulation of lipolysis in adipocytes (Guilherme et al., 2008). The “AMPK signaling pathway” was proven to be a master regulator of key molecular effectors involved in both metabolic processes and cardiovascular homeostasis by modulating the mTOR signaling and IGF-1 pathway (Salminen and Kaarniranta, 2012). The pathway “IL-2 receptor beta chain in T-cell activation” was proven to significantly attenuate ventricular remodeling by reducing infarct size and improving left ventricular (LV) function (Zeng et al., 2016).

The  $S_{GDD}$  and  $S_{AE}$  were calculated for all the 32 CA-related pathways and the 179 HM-enriched pathways. Overall, the average intra-similarity (pathways of the same types including “insulin signaling,” “insulin secretion and obesity,” and “insulin resistance and CVD” as illustrated above) in either CA-related or HM-enriched pathways was similar: higher  $S_{GDD}$  score and lower  $S_{AE}$  scores (see **Table 1** for details). This indicated that these pathways may have small similar structures instead of the whole network. Each pathway may be an up or downstream event in a disease since the biological processes inducing diseases were complex. There may be local similar

structures between two pathways, especially the adjacent ones, that may help transform the information quickly.

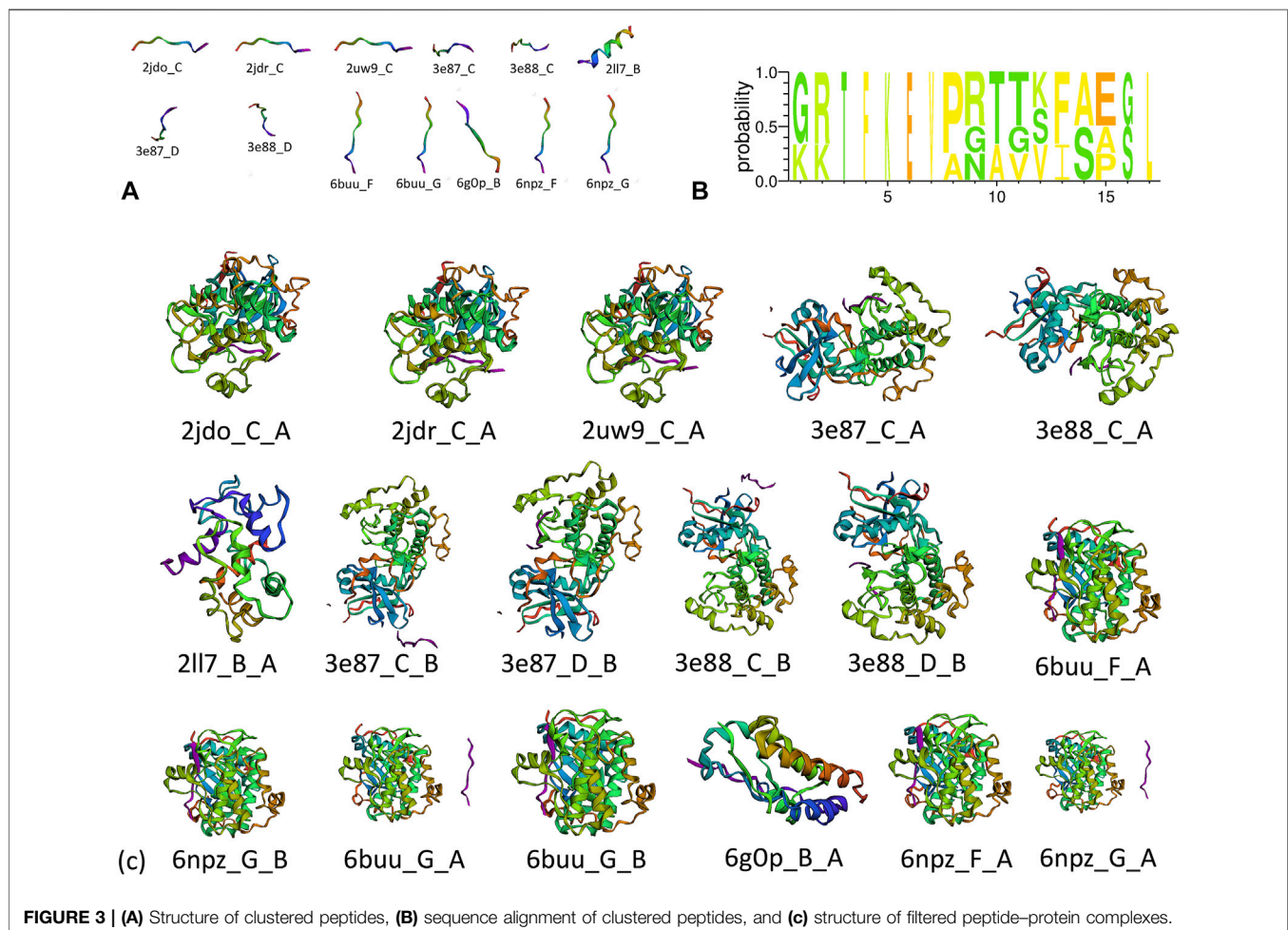
The combined pathway network similarity scores between the 12 common pathways are shown in **Figure 2D**. Of which, the pathway “Regulation of lipolysis in adipocytes (hsa04923)” (marked as “Obesity” in **Figure 2D**) got the highest average combined network similarity score (0.438) in the 12 common pathways. As illustrated above, this pathway was involved in the “Insulin resistance and CVD” processes of CVD and MetS, which was the downstream event of CVD and MetS, indicating that more cross-talks may exist between this pathway and the upstream events through the similarity network structures. Compared with this, the pathway “Insulin signaling pathway (hsa04910/h\_insulinPathway)” (marked as “Insulin signaling-1” in **Figure 2D**) got the smallest average combined network similarity score (0.218). Interestingly, this pathway was the node with the highest degree 15 in the combined pathway network. Considering the biological character of this pathway, these indicated that this upstream event in MetS and CVD may play a triggering role regardless of structure similarities to other downstream pathways.

## Peptide-Protein-Based Drug Targets Selection

The combined protein network was built using all the proteins of the 12 common pathways. The network comprised 335 nodes and 1793 edges. The proteins with top 10 degree, node betweenness, and edge betweenness are listed in **Table 2** and selected as raw candidate biomarkers. The degree of a node indicated the importance of a node in the network. A higher degree meant more connections with other nodes; thus, the proteins with higher degree may be the key targets of CA. Five of the top 10 degree proteins had been proven to be regulated by CA, including IRS1, AMPK1, AMPK2, PRKAB1, and PRKAB2. The other five proteins could be divided into two groups: monoamine oxidase (MAOA and MAOB) and protein kinase AMP-activated non-catalytic subunit gamma (PRKAG1, PRKAG2, and PRKAG3) which may be the candidate targets of CA. Cinnamon extracts (CEs) were shown to increase insulin sensitivity by increasing the mRNA expression of INSR (insulin receptor) (Anderson et al.,

**TABLE 3** | Peptide in the CA-related cluster.

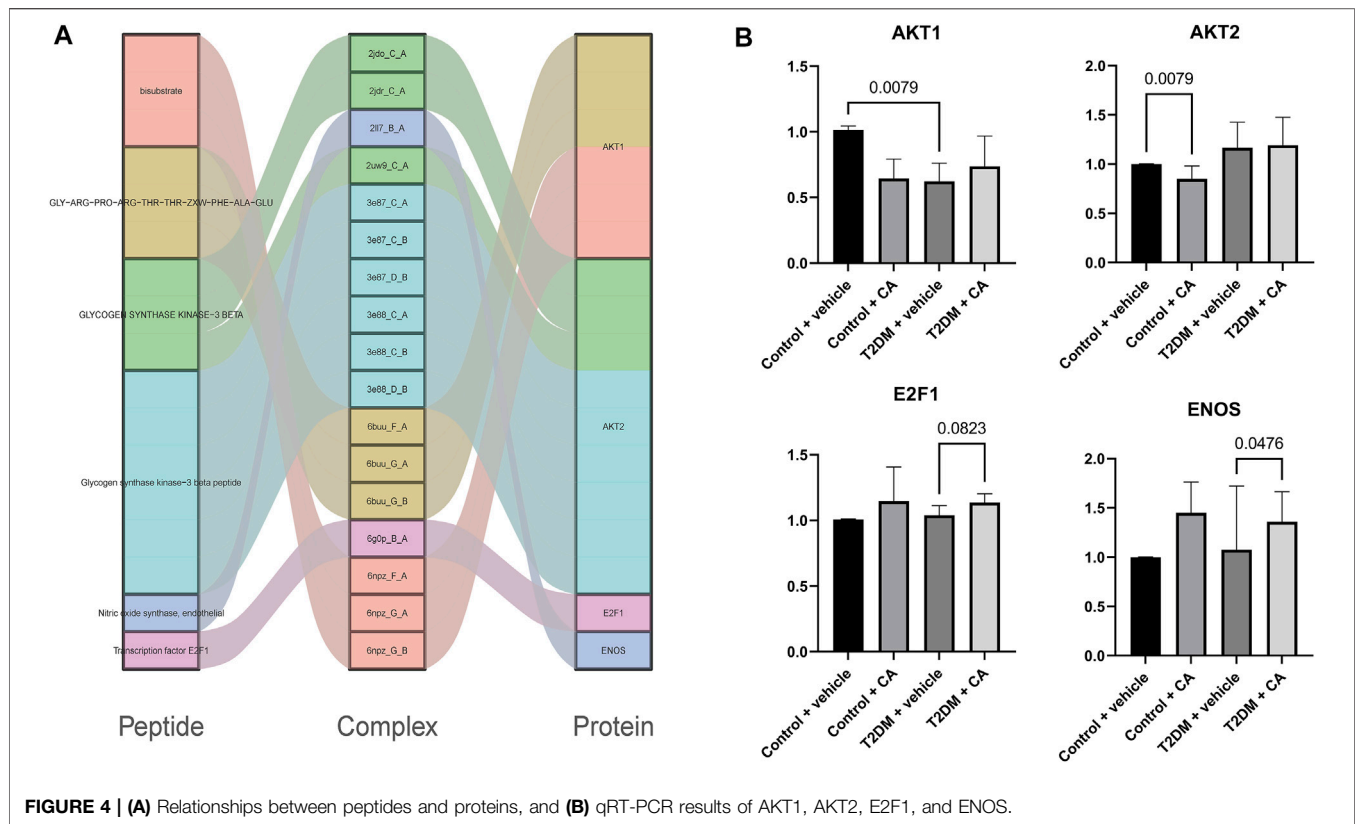
PDB	Peptide chain	Peptide size	Peptide sequence	Peptide description	Peptide molecular weight	Peptide aromaticity	Peptide instability	Peptide isoelectric point
6buu	F	11	GRPRTTXFAEX	GLY-ARG-PRO-ARG-THR-THR-ZXW-PHE-ALA-GLU	–	0.09	–	9.6
6buu	G	11	GRPRTTXFAEX	GLY-ARG-PRO-ARG-THR-THR-ZXW-PHE-ALA-GLU	–	0.09	–	9.6
6npz	F	11	GRPRTTXFAEX	Bisubstrate	–	0.09	–	9.6
6npz	G	11	GRPRTTXFAEX	Bisubstrate	–	0.09	–	9.6
2jdo	C	10	GRPRTTSFAE	Glycogen synthase kinase-3 beta	1121.2	0.1	20.72	9.6
2jdr	C	10	GRPRTTSFAE	Glycogen synthase kinase-3 beta	1121.2	0.1	20.72	9.6
2uw9	C	10	GRPRTTSFAE	Glycogen synthase kinase-3 beta	1121.2	0.1	20.72	9.6
3e87	C	10	GRPRTTSFAE	Glycogen synthase kinase-3 beta peptide	1121.2	0.1	20.72	9.6
3e87	D	10	GRPRTTSFAE	Glycogen synthase kinase-3 beta peptide	1121.2	0.1	20.72	9.6
3e88	C	10	GRPRTTSFAE	Glycogen synthase kinase-3 beta peptide	1121.2	0.1	20.72	9.6
3e88	D	10	GRPRTTSFAE	Glycogen synthase kinase-3 beta peptide	1121.2	0.1	20.72	9.6
6g0p	B	9	PGXGVXSPG	Transcription factor E2F1	–	0	–	5.96
2l17	B	17	KKTFKEVANAVKISASL	Nitric oxide synthase, endothelial	1834.16	0.06	1.14	10

**FIGURE 3** | (A) Structure of clustered peptides, (B) sequence alignment of clustered peptides, and (C) structure of filtered peptide-protein complexes.

2013), promoting IRS1 (insulin receptor substrate 1) phosphorylation (Liu et al., 2016), and activating AMPK1/2 (protein kinase AMP-activated catalytic subunit alpha 1/2) (Hu

et al., 2013). On the contrary, CE was shown to decrease the expression of genes encoding insulin signaling pathway proteins, including IGF1R (Cao et al., 2010). INS-encoded insulin and trimer





procyanidins in CE were shown to contribute to the INS-1 pancreatic  $\beta$ -cell protection (Sun et al., 2016).

Compared with degree, the measure “betweenness” reflected the importance of proteins/protein–protein pairs in the interplays between different pathways/diseases. Three of the top 10 betweenness proteins, including IRS1, INS, and INSR, were validated to be regulated by CA. Six of the 10 betweenness edges contained at least one validated CA target. It was found that the two nodes forming the edge IRS1-IGF1R were CA targets; however, IRS1 was upregulated, while IGF1R was downregulated, indicating there may exist complex interactions between CA targets.

A total of 67 protein–peptide complexes containing these top proteins as receptors were selected, and the peptides in these were then aligned and clustered. In total, 13 peptides were grouped in the CA-related cluster, their characters are listed in **Table 3**, and the structures are shown in **Figure 3A,B**. In total, 17 peptide–protein complexes were then filtered (see **Figure 3C** for the complexes’ structures), see **Figure 4A** for the relationships between these peptides and proteins.

Four of the raw candidate biomarkers (AKT1, AKT2, E2F1, and ENOS) were receptors of the abovementioned 17 peptide–protein complexes. qRT-PCR was performed on the four genes (see **Supplementary Table S2** for details).

The candidate biomarkers were divided into three groups according to their expression changing pattern in the qRT-PCR results as follows: see **Figure 4B** for details 1) The genes differentially expressed between Group A (Control + vehicle) and Group B (Control + CA) were named as CA-indicating biomarkers since the two groups were under normal condition, while the only

difference between the two groups was the drug CA. 2) The genes differentially expressed between Group A (Control + vehicle) and Group C (T2D + vehicle) were named as T2D responsive biomarkers since these genes were significantly differentially expressed between T2DM and controls but were not related to the drug CA. 3) The genes differentially expressed between Group C (T2D + vehicle) and Group D (T2D + CA) were named as CA treatment targets since the samples of the two groups were all T2D, while the only difference between them was the treatment of CA. Of which, AKT2 was a CA-indicating biomarker and AKT1 was a T2D responsive biomarker, while E2F1 and ENOS were CA treatment targets. E2F1 and ENOS were shown to cooperate with each other in the treatment of hypertension (Li et al., 2019). Combined with results from this study, the two genes might also cooperate with each other in T2D and become the targets of CA. Besides, the two genes were found to be targeted by SARS-CoV-2–encoded miRNAs in recent research (Aydemir et al., 2021). As a result, CA may be a potential candidate drug to help reduce or prevent the complications since CVDs were one of the most common complications in COVID-19 patients.

## DISCUSSION

The analysis pipeline that was proposed in this study was based on the related genes of multiple diseases. In this study, these genes were collected from OMIM and GWAS results; however, the updates of the gene lists might only influence the results slightly since the analyses were performed on pathway levels. The information conversion from genes to pathways could capture most of the

functional characters of the disease, which may not be changed by adding or deleting a small number of genes. CA was shown to play roles in a wide disease spectrum, which was the character of many Chinese traditional medicines. Thus, the drug targets of these diseases may share some similar characters reflected by peptide clusters. The pipeline proposed in this study could be applied to other diseases and drugs. Pathways were commonly used in biological and medical analyses which could gain deep understanding of diseases. However, other biological terms that could be converted into networks could also be used in this pipeline.

The portability of the pipeline was shown in all the three steps. In step 1 (pathway filter), the similarity calculation methods between different disease pairs could be replaced by any suitable distance measures. The disease-related and drug-related pathways could be selected using any suitable scores or ways. Other functional resources and transcriptional information such as GO terms, transcriptional factors–targets, or miRNA targets could also be used. However, pathways were recommended as the primary choice because the biological pathways were widely used in biological and medical analyses since they could reflect the molecular connections in the form of graphs, which could be analyzed using multiple computational methods. Besides, the correlations between pathways and peptides were closer than those between other types of functional resources. In step 2 (combined network construction), the network structure similarities could be measured using one alignment-free and one alignment-based algorithms. In step 3 (biomarker prediction and validation based on peptides), the peptide clustering algorithms could be replaced by any other suitable alignment method.

## CONCLUSION

In this study, a new pipeline was proposed to discover drug targets based on peptides. The network analyses based on machine learning methods could quantify the effects of peptide–protein complexes with similar structures as drug targets in multiple diseases.

## REFERENCES

- Anderson, R. A., Qin, B., Canini, F., Poulet, L., and Roussel, A. M. (2013). Cinnamon Counteracts the Negative Effects of a High Fat/high Fructose Diet on Behavior, Brain Insulin Signaling and Alzheimer-Associated Changes. *PLoS One* 8 (12), e83243. doi:10.1371/journal.pone.0083243
- Aydemir, M. N., Aydemir, H. B., Korkmaz, E. M., Budak, M., Cekin, N., and Pinarbasi, E. (2021). Computationally Predicted SARS-COV-2 Encoded microRNAs Target NFKB, JAK/STAT and TGFB Signaling Pathways. *Gene Rep.* 22, 101012. doi:10.1016/j.genrep.2020.101012
- Cao, H., Graves, D. J., and Anderson, R. A. (2010). Cinnamon Extract Regulates Glucose Transporter and Insulin-Signaling Gene Expression in Mouse Adipocytes. *Phytomedicine* 17 (13), 1027–1032. doi:10.1016/j.phymed.2010.03.023
- Crooks, G. E., Hon, G., Chandonia, J.-M., and Brenner, S. E. (2004). WebLogo: A Sequence Logo Generator. *Figure 1. Genome Res.* 14 (6), 1188–1190. doi:10.1101/gr.849004
- Das, S., Shah, R., Dimmeler, S., Freedman, J. E., Holley, C., Lee, J. M., et al. (2020). Noncoding RNAs in Cardiovascular Disease: Current

## DATA AVAILABILITY STATEMENT

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

## ETHICS STATEMENT

The animal study was reviewed and approved by the experimental Animal Ethics Committee of Shenzhen Sun Yat sen Cardiovascular Hospital.

## AUTHOR CONTRIBUTIONS

YW, KL, and HZ contributed to the conception and design of the study. YW performed the bioinformatics analysis. LW constructed the animal model. YL performed the illustration of results. All authors contributed to manuscript revision, and read and approved the submitted version.

## FUNDING

This study was supported by the China Postdoctoral Science Foundation (2020M683188), the Fund of “Sanming” Project of Medicine in Shenzhen (SZSM201911019), and the National Natural Science Foundation of China (81971079 and 61702496).

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2021.816131/full#supplementary-material>

Knowledge, Tools and Technologies for Investigation, and Future Directions: A Scientific Statement from the American Heart Association. *Circ. Genom Precis Med.* 13, e000062–372. doi:10.1161/HCG.0000000000000062

- Frigo, M., Cruciani, E., Coudert, D., Deriche, R., Natale, E., and Deslauriers-Gauthier, S. (2021). Network Alignment and Similarity Reveal Atlas-Based Topological Differences in Structural Connectomes. *Netw. Neurosci.* 5 (3), 711–733. doi:10.1162/netn\_a\_00199
- Guilherme, A., Virbasius, J. V., Puri, V., and Czech, M. P. (2008). Adipocyte Dysfunctions Linking Obesity to Insulin Resistance and Type 2 Diabetes. *Nat. Rev. Mol. Cel Biol* 9 (5), 367–377. doi:10.1038/nrm2391
- Hariri, M., and Ghiasvand, R. (2016). Cinnamon and Chronic Diseases. *Adv. Exp. Med. Biol.* 929, 1–24. doi:10.1007/978-3-319-41342-6\_1
- Hu, N., Yuan, L., Li, H. J., Huang, C., Mao, Q. M., Zhang, Y. Y., et al. (2013). Anti-Diabetic Activities of Jiaotaiwan in Db/db Mice by Augmentation of AMPK Protein Activity and Upregulation of GLUT4 Expression. *Evid. Based Complement. Alternat Med.* 2013, 180721. doi:10.1155/2013/180721
- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Bioinformatics Enrichment Tools: Paths toward the Comprehensive Functional Analysis of Large Gene Lists. *Nucleic Acids Res.* 37 (1), 1–13. doi:10.1093/nar/gkn923

- Huang, D. W., Sherman, B. T., and Lempicki, R. A. (2009). Systematic and Integrative Analysis of Large Gene Lists Using DAVID Bioinformatics Resources. *Nat. Protoc.* 4 (1), 44–57. doi:10.1038/nprot.2008.211
- Ji, C., Chen, H., Wang, R., Cai, Y., and Wu, H. (2021). Smoothness Sensor: Adaptive Smoothness-Transition Graph Convolutions for Attributed Graph Clustering. *IEEE Trans Cybern.* doi:10.1109/TCYB.2021.3088880
- Kakiuchi-Kiyota, S., Schutten, M. M., Zhong, Y., Crawford, J. J., and Dey, A. (2019). Safety Considerations in the Development of Hippo Pathway Inhibitors in Cancers. *Front. Cell Dev. Biol.* 7, 156. doi:10.3389/fcell.2019.00156
- Kaku, K. (2019). A New Concept of GLP-1 Signaling Pathway on Pancreatic Insulin Secretion. *J. Diabetes Investig.* 11, 265–267. doi:10.1111/jdi.13136
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., and Morishima, K. (2017). KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* 45 (D1), D353–D361. doi:10.1093/nar/gkw1092
- Khan, A., Bryden, N. A., Polansky, M. M., and Anderson, R. A. (1990). Insulin Potentiating Factor and Chromium Content of Selected Foods and Spices. *Biol. Trace Elem. Res.* 24 (3), 183–188. doi:10.1007/BF02917206
- Krejci, A., Hupp, T. R., Lexa, M., Vojtesek, B., and Muller, P. (2016). Hammock: a Hidden Markov Model-Based Peptide Clustering Algorithm to Identify Protein-Interaction Consensus Motifs in Large Datasets. *Bioinformatics* 32 (1), 9–16. doi:10.1093/bioinformatics/btv522
- Kuhn, H. W. (1955). The Hungarian Method for the Assignment Problem. *Naval Res. Logistics Q.* 2 (1-2), 83–97. doi:10.1002/nav.3800020109
- Laviola, L., Natalicchio, A., and Giorgino, F. (2007). The IGF-I Signaling Pathway. *Cpd* 13 (7), 663–669. doi:10.2174/138161207780249146
- Leahy, J. L., and Vandekerkhove, K. M. (1990). Insulin-Like Growth Factor-I at Physiological Concentrations Is a Potent Inhibitor of Insulin Secretion\*. *Endocrinology* 126 (3), 1593–1598. doi:10.1210/endo-126-3-1593
- Li, H., Li, Q., Zhang, Y., Liu, W., Gu, B., Narumi, T., et al. (2019). Novel Treatment of Hypertension by Specifically Targeting E2F for Restoration of Endothelial Dihydrofolate Reductase and eNOS Function under Oxidative Stress. *Hypertension* 73 (1), 179–189. doi:10.1161/hypertensionaha.118.11643
- Liu, Y., Li, X., Xie, C., Luo, X., Bao, Y., Wu, B., et al. (2016). Prevention Effects and Possible Molecular Mechanism of Mulberry Leaf Extract and its Formulation on Rats with Insulin-Insensitivity. *PLoS One* 11 (4), e0152728. doi:10.1371/journal.pone.0152728
- London, N., Raveh, B., Movshovitz-Attias, D., and Schueler-Furman, O. (2010). Can Self-Inhibitory Peptides Be Derived from the Interfaces of Globular Protein-Protein Interactions? *Proteins* 78 (15), 3140–3149. doi:10.1002/prot.22785
- Menche, J., Sharma, A., Kitsak, M., Ghiassian, S. D., Vidal, M., Loscalzo, J., et al. (2015). Uncovering Disease-Disease Relationships through the Incomplete Interactome. *Science* 347 (6224), 1257601. doi:10.1126/science.1257601
- Mollazadeh, H., and Hosseinzadeh, H. (2016). Cinnamon Effects on Metabolic Syndrome: a Review Based on its Mechanisms. *Iran J. Basic Med. Sci.* 19 (12), 1258–1270. doi:10.22038/ijbms.2016.7906
- Nedeva, V., Linding, R., Su-Grand, I., Stark, A., Masi, F. d., Gibson, T. J., et al. (2005). Systematic Discovery of New Recognition Peptides Mediating Protein Interaction Networks. *Plos Biol.* 3 (12), e405. doi:10.1371/journal.pbio.0030405
- Parthasarathi, L., Casey, F., Stein, A., Aloy, P., and Shields, D. C. (2008). Approved Drug Mimics of Short Peptide Ligands from Protein Interaction Motifs. *J. Chem. Inf. Model.* 48 (10), 1943–1948. doi:10.1021/ci800174c
- Pawson, T., and Nash, P. (2003). Assembly of Cell Regulatory Systems through Protein Interaction Domains. *Science* 300 (5618), 445–452. doi:10.1126/science.1083653
- Pørksen, N., Hussain, M. A., Bianda, T. L., Nyholm, B., Christiansen, J. S., Butler, P. C., et al. (1997). IGF-I Inhibits Burst Mass of Pulsatile Insulin Secretion at Supraphysiological and Low IGF-I Infusion Rates. *Am. J. Physiol.* 272 (3 Pt 1), E352–E358. doi:10.1152/ajpendo.1997.272.3.E352
- Przulj, N. (2007). Biological Network Comparison Using Graphlet Degree Distribution. *Bioinformatics* 23 (2), e177–e183. doi:10.1093/bioinformatics/btl301
- Qin, B., Nagasaki, M., Ren, M., Bajotto, G., Oshida, Y., and Sato, Y. (2003). Cinnamon Extract (Traditional Herb) Potentiates *In Vivo* Insulin-Regulated Glucose Utilization via Enhancing Insulin Signaling in Rats. *Diabetes Res. Clin. Pract.* 62 (3), 139–148. doi:10.1016/s0168-8227(03)00173-6
- Sales, G., Calura, E., Cavaliere, D., and Romualdi, C. (2012). Graphite - a Bioconductor Package to Convert Pathway Topology to Gene Network. *BMC Bioinformatics* 13, 20. doi:10.1186/1471-2105-13-20
- Salminen, A., and Kaamiranta, K. (2012). AMP-activated Protein Kinase (AMPK) Controls the Aging Process via an Integrated Signaling Network. *Ageing Res. Rev.* 11 (2), 230–241. doi:10.1016/j.arr.2011.12.005
- Salt, I. P., and Hardie, D. G. (2017). AMP-activated Protein Kinase. *Circ. Res.* 120 (11), 1825–1841. doi:10.1161/circresaha.117.309633
- Schriner, S. E., Kuramada, S., Lopez, T. E., Truong, S., Pham, A., and Jafari, M. (2014). Extension of *Drosophila* Lifespan by Cinnamon through a Sex-specific Dependence on the Insulin Receptor Substrate chico. *Exp. Gerontol.* 60, 220–230. doi:10.1016/j.exger.2014.09.019
- Sheng, X., Zhang, Y., Gong, Z., Huang, C., and Zang, Y. Q. (2008). Improved Insulin Resistance and Lipid Metabolism by Cinnamon Extract through Activation of Peroxisome Proliferator-Activated Receptors. *PPAR Res.* 2008, 581348. doi:10.1155/2008/581348
- Sun, P., Wang, T., Chen, L., Yu, B.-w., Jia, Q., Chen, K.-x., et al. (2016). Trimer Procyanidin Oligomers Contribute to the Protective Effects of Cinnamon Extracts on Pancreatic  $\beta$ -cells *In Vitro*. *Acta Pharmacol. Sin* 37 (8), 1083–1090. doi:10.1038/aps.2016.29
- Takasao, N., Tsuji-Naito, K., Ishikura, S., Tamura, A., and Akagawa, M. (2012). Cinnamon Extract Promotes Type I Collagen Biosynthesis via Activation of IGF-I Signaling in Human Dermal Fibroblasts. *J. Agric. Food Chem.* 60 (5), 1193–1200. doi:10.1021/jf2043357
- Thompson, J. D., Higgins, D. G., Gibson, T. J., and Clustal, W. (1994). CLUSTAL W: Improving the Sensitivity of Progressive Multiple Sequence Alignment through Sequence Weighting, Position-specific gap Penalties and Weight Matrix Choice. *Nucl. Acids Res.* 22 (22), 4673–4680. doi:10.1093/nar/22.22.4673
- Wang, Y., Huang, X., Liu, J., Zhao, X., Yu, H., and Cai, Y. (2019). A Systems Analysis of the Relationships between Anemia and Ischemic Stroke Rehabilitation Based on RNA-Seq Data. *Front. Genet.* 10, 456. doi:10.3389/fgene.2019.00456
- Xiao, Y., Wu, Q. Q., Jiang, X. H., and Tang, Q. Z. (2017). Cinnamaldehyde Attenuates Pressure Overload-Induced Cardiac Fibrosis via Inhibition of Endothelial Mesenchymal Transition. *Zhonghua Yi Xue Za Zhi* 97 (11), 869–873. doi:10.3760/cma.j.issn.0376-2491.2017.11.015
- Yang, L., Wu, Q. Q., Liu, Y., Hu, Z. F., Bian, Z. Y., and Tang, Q. Z. (2015). Cinnamaldehyde Attenuates Pressure Overload-Induced Cardiac Hypertrophy. *Int. J. Clin. Exp. Pathol.* 8 (11), 14345–14354.
- Yingying, W., Yu, Y., Jianfeng, L., and Keshen, L. (2021). Construction of Anatomical Structure-specific Developmental Dynamic Networks for Human Brain on Multiple Omics Levels. *Curr. Bioinformatics* 16 (9), 1133–1142. doi:10.2174/15748936166666210331115659
- Zeng, Z., Yu, K., Chen, L., Li, W., Xiao, H., and Huang, Z. (2016). Interleukin-2/ Anti-Interleukin-2 Immune Complex Attenuates Cardiac Remodeling after Myocardial Infarction through Expansion of Regulatory T Cells. *J. Immunol. Res.* 2016, 8493767. doi:10.1155/2016/8493767
- Zhang, L. Q., Zhang, Z. G., Fu, Y., and Xu, Y. (2015). Research Progress of Trans-cinnamaldehyde Pharmacological Effects. *Zhongguo Zhong Yao Za Zhi* 40 (23), 4568–4572.
- Zhu, R., Ji, C., Wang, Y., Cai, Y., and Wu, H. (2020). Heterogeneous Graph Convolutional Networks and Matrix Completion for miRNA-Disease Association Prediction. *Front. Bioeng. Biotechnol.* 8, 901. doi:10.3389/fbioe.2020.00901

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

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

Copyright © 2021 Wang, Wang, Liu, Li and Zhao. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.