Peer Review File

Article information: http://dx.doi.org/10.21037/atm-20-1087

Response to Reviewer A

Comment 1: The authors claim this method referenced to tumors. But a quick PubMed search by the reviewer shows many studies in the CKD field itself. Do the authors have a grasp on what the current strategies/studies are there in CKD/DKD.

Reply1: Thank you for the valuable comment. We are very sorry for our incorrect statement " WGCNA is an effective method to screen out key genes, which is wildly used in diseases such as tumors " . We do agree with you that WGCNA (as well as other bioinformatics analyses) are not only used for researches concerning tumors. We have pay close attention to similar studies and know that WGCNA is applied in various diseases, such as tumors, chronic disease, immunity disease, mental disease, also in kidney disease. Our improper statement may mislead readers. Thank you again for pointing it out.

Changes in the text: We modified our text and added the corresponding references (see Page 5-6, line 67-70).

Comment 2: Throughout the manuscript, there is a lack of clear and in-depth explanation of the method used. A major flaw in data processing. The authors claim they converted the probe expression in (.txt) file to gene expression on line 60 under Data processing heading. The reviewer looked at that .txt file in the GEO website for that database GSE104948 and see that the data is already at the gene level. So, raises the question about the processing of the data where they claim removal of probes and taking median. And since this study is dependent on this data it's important for the authors to clarify this as this raises a major concern on the validity of this study. Therefore, the authors need to explain this in detail on the data processing.

Reply 2: Thank you for your comments. We are sorry for our unclear description of data procession. As we know, the chip technology refers to hybridizing a large number of probes with labeled sample molecules, then obtaining the corresponding gene expression of the sample molecules by detecting the intensity of the hybridization signal of each probe molecule. So, the expression matrix file is often presented in the form of probes. In our study, the expression matrix files of dataset

GSE104948 was downloaded from the Gene Expression Omnibus (GEO) database (download link: https://ftp.ncbi.nlm.nih.gov/geo/series/GSE104nnn/GSE104948/matrix/). Reviewer 1 claimed that the data is already at gene level, but we double check the expression matrix file and confirmed again that the data presented in probe form. Therefore, the probe annotation is necesssary to correspond probes to genes symbols, as other similar studies (1-4). In all probes of the microarray, a certain probe can point to multiple genes, meanwhile a certain gene can be measured by multiple probes, so we removed the probe matching multiple genes, and as to the gene matching with multiple probes, we took the average value of probes as the final expression value of the corresponding gene. In order to better show, we take a screenshot of the expression matrix file:

Data download

Download family	Format
SOFT formatted family file(s)	SOFT
MINiML formatted family file(s)	MINiML
Series Matrix File(s)	TXT

Supplementary file	Size	Download	File type/resource
GSE104948_RAW.tar	738.6 Mb	(http)(custom)	TAR (of CEL)

Raw data provided as supplementary file

Processed data included within Sample table

Probe matrix file

ID_REF	GSM2810770	GSM2810771	GSM2810772	GSM2810773	GSM2810774	GSM2810775	GSM2810776	GSM2810777	GSM2810778	GSM2810779	G
10000_at	8.39481969	8.23478755	8.76976578	8.26023993	8.103905	8.48508761	8.18712804	8.81460385	8.75550842	8.41514146	8
10001_at	6.34464612	6.43266426	6.6570519	5.98792413	6.4401251	6.19897231	6.13597054	5.96638904	6.49565627	6.20562171	6
10002_at	4.88331236	4.88938413	4.80764787	5.09508871	4.72988802	4.83117502	4.73581121	4.90228066	5.05884804	4.66435555	4
10003_at	4.68823301	4.7009386	4.69278283	4.46076973	4.69675615	4.63236931	4.92895556	5.0096698	5.05492978	4.1044057	4
100048912_at	5.3597172	5.30850919	5.58147197	5.6132042	5.29369159	5.46283973	5.37138569	5.52217101	5.72493917	5.76942344	
10004_at	6.23205842	6.08483024	6.50884043	6.6430278	5.6781874	6.30232093	5.97011055	6.44962961	6.73420541	6.33033416	6
10005_at	9.01351751	9.21961894	9.72167143	9.42859185	8.24797658	9.4777493	8.69948285	9.32561028	9.3845257	9.75301926	9
10006_at	9.14910701	8.88469612	9.20803985	8.43240675	9.01590982	8.977687	8.59295486	9.25148368	9.13637959	9.28971115	9
10007_at	9.62493326	9.74779884	9.10979298	10.4100418	9.92915885	9.60720089	9.91792362	9.74678166	9.26589095	10.4096502	9
100093698_at	4.5410667	4.29126815	4.51481453	4.80141618	4.35775216	4.62230546	4.65915764	4.5366047	4.45551275	4.62068656	
10009_at	5.81542663	5.93477218	5.94687085	5.69292631	5.83271148	5.62363201	5.79830777	6.03490918	5.69621423	5.54421584	5
1000_at	7.24484812	7.4448125	6.70576973	7.21224535	8.2346145	7.13454498	7.26945205	6.92424054	6.24775843	7.202958	
10010_at	7.54936732	7.52625611	7.51398216	7.01783193	7.85851438	7.73820366	7.22486012	7.17897953	7.2585671	7.31484789	7
100126791_at	5.49147182	5.69218587	5.45477418	6.6315585	5.73125084	5.57646316	5.77880066	5.77010847	5.37237588	6.80429072	5
100128124_at	5.09605656	5.5107021	5.14013659	5.50101277	5.12528964	5.41767517	5.27532317	5.28363098	5.36524354	5.1793568	5
100128640_at	4.76087388	5.48717502	4.90336009	5.56576144	5.36698189	5.72038436	5.19365465	5.01951112	5.10580018	5.40326564	5
100129128_at	4.40396283	4.67280686	4.57699911	4.8056846	4.66351107	4.62100522	4.34246935	4.68809211	4.6853519	4.58474685	4
100129250_at	8.01794286	8.30087189	8.38425749	8.67396356	7.82153778	8.7262959	7.90505974	8.03927841	8.33976677	8.27953691	8
100129271_at	5.32450105	5.25608608	5.1265717	5.32550661	5.14556507	5.72781887	5.44275578	5.27211422	5.74021174	5.76080629	5
100129361_at	8.70241877	8.33999479	9.043749	8.68311859	8.08871914	7.90923794	7.85070276	8.80803278	8.92799627	8.83854817	8
100129460_at	4.99347269	5.17271105	5.0786618	5.17906371	4.92755438	5.08910314	5.30004291	5.33948043	5.04174013	5.09723291	4
100129482_at	5.52420388	5.60334539	5.78555097	5.34692498	5.38957362	5.81680014	5.66699643	5.6211446	5.57025116	5.62461603	
100129503_at	7.21427417	6.58935697	6.86360743	6.9995351	7.07027394	6.66508439	7.04532576	6.85075794	6.88562447	6.95887291	6
100129973_at	4.1640739	4.29985715	4.016999	4.67183945	4.10243845	3.88773847	4.34010241	4.52409387	4.45230307	4.70998542	4
100130100_at	5.07353298	5.92574592	5.39433999	5.49794481	6.72258532	5.50044221	6.12030188	5.28378531	5.80818199	5.97717205	5
100130331_at	4.66395037	4.70647641	4.82909029	4.83536784	4.82195391	4.87961706	4.89179999	4.81766322	4.79333174	4.87692757	4
100130449_at	5.06630706	4.54555152	5.10870119	4.63022682	4.58333351	4.87589159	4.42582224	4.43363409	4.80233769	4.3282193	4
100131532_at	5.68400465	5.48080825	6.06866653	5.96352593	5.47746209	5.64925448	5.75636449	5.90602058	5.74750456	5.79955974	5
100131755_at	5.42703224	5.27780446	5.53158543	5.15609246	5.2535714	5.03659885	5.41509221	5.48841109	5.54531576	5.20773118	5

We can provide the probe matrix file in TXT format before and after the annotation for your reference if necessary. Thank you again for your patience and valuable

comment.

Changes in the text: We have re-written this part according to your kind suggestions (see Page 7, line 85-89).

Comment 3: In figure 1 its says DEGs for DKD/normal related module. Again, there is a major flaw in the implementation/interpretation of the WGCNA method. Basically, the two modules that the authors identified as DKD and normal in fact can be considered as modules that are upregulated/downregulated in DKD. The authors need to invest a little more on the implementation of this method on how they are associating the modules to the phenotypes of DKD and normal and thereby the interpretation of the findings.

Reply 3: Thank you for your advice. The purpose of module-trait correlation analysis is to find the co-expression modules that are significantly related to a certain trait. In our study, DKD and normal are the clinical traits of samples. After constructing weighted gene co-expression network, genes were divided into several co-expression modules. The expression value of each gene in the modules is expressed as a continuous value. When exploring the relationship between the trait and module, it is difficult to analyze the entire module. So, a new concept module eigengenes (MEs) was introduced, which was the chief component of a certain module and can represent the expression pattern of the module. Then, we calculated the correlation coefficients between MEs and clinical trait by Pearson's test to find trait-related modules. In order to verify the MEs can represent the module well, we calculated the correlation analysis between GS and MM (the definitions of GS and MM are available in page 9-10, line 122-126 of the manuscript). When a module had a high correlation between GS and MM, and most points in the scatter plot diagram are distributed in the upper right corner, the genes in the module had a close relationship with both MEs and traits related to modules. That is, MEs can represent the whole module well and the major genes in the module had high relationship with clinical traits. Modules meet these criteria are considered to be worthy of further analyses (Figure 4 of the manuscript).

Changes in the text: We have re-written this part (see Page 9, line 114-127).

Comment 4: Several places talk about DKD and normal. But what is normal. Define normal. And actually contradicts at couple places. In line 178 they claim that the red

module is ."......metabolic process, which uncovered the metabolism disorder in DKD." and in line 186 mentioning ".....most genes in red module may have a great impact in maintaining normal function".

Reply 4: Thank you for your comments. Sorry for our unclear expression. In our study, the clinical traits of samples included DKD and normal. The kidney tissue was got from DKD patients and healthy living donors correspondingly. Therefore, normal is referred to the healthy living donors.

Module with positive relationship with DKD played a role in the pathogenesis of DKD, while modules positively correlating with normal trait were important in maintaining normal biological function. Based on our results, the red module is significantly related to the normal trait (r = 0.8, P = 2e–07). What is more, GO and KEGG enrichment analyses on red module showed that most genes in this module were related to the metabolic process. So, major genes in red module took part in various metabolic processes, which indicated a role in maintaining normal biological functions. The results also indirectly reflected the possible metabolic disorder of DKD, which were in accordance with previous reports (5-7).

Changes in the text: We have modified our text as advised (see Page 6, line 80-83; Page 16, line 214-221; Page 20-21, line 270-275)"

Comment 5: Throughout the manuscript there is a lack of clarity and poor explanation. The methods section is written very sparsely. The authors do not provide enough material to convince the reviewer that the finding is robust from a computational perspective and neither for a biomarker claim.

Reply 5: We'd like to thank you for your careful readings and constructive comments. We apologize for unclear description of our manuscript. We have carefully modified our manuscript according to your suggestions and added more details to improve clarity for better understanding. We hope our modifications can meet the requirements of ATM.

Changes in the text: We have modified our manuscript (Revised portion are marked in red)"

Response to Reviewer B

Comment 1: The authors analyze Transcriptome data from the GEO database to identify bub genes associated with DKD using WGCNA. An advantage of WGCNA is that it provides powerful module preservation statistics that assess the density (how tight interconnections among genes in a module are) and connectivity patterns of individual modules. Similar research has been performed before by Tang et al. (Tang W, et al. Eur Rev Med Pharmacol Sci. 2012;16(14):1967–1973). This should be discussed. The difference in results between these two studies might be explained by the batch effect. A new hub-gene, FCER1G, was identified and validated in an independent GEO dataset. I think this study still adds something new to the literature.

Reply1: Thank you for your encouraging and helpful comments. In fact, we have carefully studied the research you mentioned before we conducted our study. Both Tang W and we found that DKD was related to immunity. However, the key genes we discovered are not the same. There are several reasons we want to explain:

Firstly, we agree with what you said, the batch effect (different chip platforms, operators, reagents, or instruments) can make the experimental results not exactly consistent. Both expression data mining and basic experiments (such as western blot, immunehistochemical et al.) would face these similar problems. Therefore, it is very important to verify the robustness of the result. In our study, we verified the differential expression level of the key gene between normal and DKD tissues in three other external datasets and found FCER1G overexpressed in DKD compared with normal in each dataset. Furthermore, we explored the relationship between FCER1G and clinical traits of DKD and found a negative correlation between the expression of FCER1G in DKD glomeruli and glomerular filtration rate (GFR), which further confirmed the reliability of our results.

Secondly, different analytic methods may lead to the difference of the most vital genes. In our study, we constructed the weighted gene co-expression network, and analyzed Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) enrichment, finally got the key gene from the top 2 most significant terms from terms of GO-BP enrichment. while the methods of other similar researches are different from ours, the key genes could also be different. We think this is not contradictory, but complementary.

Lastly, the pathogenesis behind the disease is complex and must be the result of interaction of multiple genes. Therefore, the distinction of results from different

research can be mutually corroborated and also contributed to mining more useful information. What is more, both Tang W and we found that DKD was related to immunity, which further verified a high credibility of our result.

Once again, thank you very much for your comments.

References

1. Yang J, Wang L, Xu Z, et al. Integrated Analysis to Evaluate the Prognostic Value of Signature mRNAs in Glioblastoma Multiforme. Front Genet 2020;11:253.

2. Guan X, Guan Z, Song C. Expression profile analysis identifies key genes as prognostic markers for metastasis of osteosarcoma. Cancer Cell Int 2020;20:104.

3. Zhang H, Guo L, Zhang Z, et al. Co-Expression Network Analysis Identified Gene Signatures in Osteosarcoma as a Predictive Tool for Lung Metastasis and Survival. J Cancer 2019;10:3706-16.

4. Xu W, Li J, Li J, et al. An Investigation about Gene Modules Associated with hDPSC Differentiation for Adolescents. Stem Cells Int 2019;2019:8913287.

5. Niewczas MA, Mathew AV, Croall S, et al. Circulating Modified Metabolites and a Risk of ESRD in Patients With Type 1 Diabetes and Chronic Kidney Disease. Diabetes Care 2017;40:383-90.

6. Tofte N, Vogelzangs N, Mook-Kanamori D, et al. Plasma metabolomics identifies markers of impaired renal function: A meta-analysis of 3,089 persons with type 2 diabetes. J Clin Endocrinol Metab 2020.

7. Afshinnia F, Nair V, Lin J, et al. Increased lipogenesis and impaired beta-oxidation predict type 2 diabetic kidney disease progression in American Indians. JCI Insight 2019;4.