Peer Review File

Article information: https://dx.doi.org/10.21037/pm-23-27

Reviewer A

Developmental disorders result from impaired prenatal development, often attributed to deleterious genetic variations. In the manuscript "The role of gene expression trajectories in developmental disorders", authors emphasized the crucial role of gene expression trajectories in advancing the understanding of developmental disorders.

Couple questions are required to be answered before it will be accepted.

Comment 1: The gene expression trajectory was the crucial topic in the study. How to define the gene expression trajectory? Please state in the introduction.

Reply 1: *Thank you for this comment. We have acknowledged this point and added a definition of gene expression trajectory in the Introduction section.*

Changes in the text: *We added a definition of gene expression trajectory as* "Developmental expression trajectories refer to the dynamic patterns of gene expression throughout the process of development." *on lines 80-81.*

Comment 2: There were massive causes involved in the developmental disorders, such as mutation or deletion. Why to focus on the gene expression trajectories in the study? Please state in the introduction.

Reply 2: *Thank you for this comment. Developmental disorders are primarily caused by deleterious genetic variations in genes that are crucial for proper development. The knockout mouse was a common research tool, but 49.53% of genes linked to various disorders exhibited no overlapping phenotype between the knockout mouse model and the corresponding human disease. In addition, previous studies have revealed that there are no differential gene expression levels in orthologous genes between humans and mice within the same tissue. Therefore, we focused on gene expression trajectories and tried to provide a potential explanation for the divergent phenotypes resulting from null mutations in human and mouse by comparing the similarity of gene expression trajectories and phenotypic similarity.*

Changes in the text: *In the Introduction section, we present a revised description of our study using more precise data as "*Previous studies using knockout mouse models have shown that a considerable percentage (49.53%) of genes linked to disorders do not exhibit a shared phenotype between the mouse model and the corresponding human disease.*" on lines 68-71. We have also provided more detailed comments as "*Genes associated with the developmental process are precisely

regulated(5). Species with comparable developmental trajectories often display shared patterns in conserved regulatory networks and molecular processes that contribute to observed phenotypic similarities. Previous studies have revealed that half of human genes showed distinct temporal trajectories comparing to their mouse orthologs(6). Zheng et al, found no significant contribution of expression level divergence between human and mouse orthologous genes across homologous tissues of the two species from microarray expression datasets(7). It's unclear whether the similarity of developmental expression trajectories could serve as an informative index for phenotypic similarity as widely assumed.*" in the Introduction section on lines 72-79.*

Comment 3: How to obtain the calculated formula of spearman correlation ρ? Please state in the methods.

Reply 3: *Thank you for this comment. Spearman correlation ρ is calculated by assigning ranks to the data points for each variable, converting the data into ordinal form. Then, the difference between the ranks for each pair of observations is computed. The correlation coefficient is derived from these differences and ranges between -1 and 1. In this study, the Spearman correlation coefficient was calculated using the 'cor' function in R (version 4.2.2).*

Changes in the text: *We have added a description of spearman correlation ρ as "*ρ is ranges from -1 to 1 with the values closer to 1 indicating a stronger positive association between human and mouse RPKMs. *on lines 134-135. In addition, we mentioned that "*Gene expression trajectory similarities were calculated using Spearman correlation using the 'cor' function in R (version 4.2.2)*" on lines 187-188.*

Comment 4: The human and mouse gene temporal expression data was analyzed in the study. How about the spatial gene expression trajectories in developmental disorders?

Reply 4: *Thanks for your comment. Spatial gene expression trajectories require the use of single-cell RNA sequencing and spatial transcriptomics. However, in this study, we utilized bulk transcriptome data of human and mouse to analysis the tissue-specific temporal expression patterns of genes, which did not allow for spatial relevant analysis. Spatial gene expression trajectory is helpful in understanding the spatial specificity of gene expression in tissues or cell populations, we will explore these aspects in future investigations.*

Changes in the text: *We have clearly defined gene expression trajectory as "*Developmental expression trajectories refer to the dynamic patterns of gene expression throughout the process of development. Similarity of expression trajectories can be quantified by comparing RNA sequencing (RNA-seq) data from

homologous tissues of human and mouse across calibrated developmental ages(8)*" on lines 80-83. We also provided more information of gene expression data as "*We utilized gene temporal expression data from bulk transcriptomes of humans and mice across developmental stages.*" on lines 123-124.*

Comment 5: Missing experimental data was the biggest short board in the study. It was better to validate the analytical data by experiments.

Reply 5: *I agree with your viewpoint that the lack of experimental data provided by this study is its biggest limitation. However, the data of iGluRs genes presented in our study is derived from experimental results reported in the literature, ensuring its reliability and validity as a trustworthy data source.*

Changes in the text: *We added detailed screening methods in the Method section as "*The phenotypes of patients harbouring PTVs were independently evaluated by two physicians, as per the assessment using the HPO terms. Relevant HPO terms associated with the gene in question were obtained from the official HPO website, and those aligning with the clinical descriptions found in the literature were carefully chosen for analysis. We also curated the MPO terms for gene knockout mice with documented literature reports in the MGI database and verified them within the manuscript.*" on lines 164-169. We hope this makes it more reliability and validity.*

Comment 6: It was better to add reference (DOI: 10.21037/qims.2018.08.08) about the genotype-phenotype correlation.

Reply 6: *Thank you for this comment. In our study, we found that comparable phenotypes may arise from different gene expression trajectories between human and mouse. In cases where a gene becomes dysfunctional, the remaining genes of the same pathway might have the capacity to assume compensatory functions, thereby influencing the observed differences in phenotypes between human and mouse. This reference implied that the mechanisms underlying disease-associated genes possess the potential to align with molecular processes and pathways. So, we have added this reference to our Discussion section.*

Changes in the text: *We added to our Discussion section on lines 370-374 as"*The other possible explanation is that the mechanisms underlying disease-associated genes possess the potential to align with molecular processes and pathways, revealing intricate interdependencies(31). In cases where a gene becomes dysfunctional, the remaining genes have the capacity to assume compensatory functions, thereby influencing the observed differences in phenotypes between human and mouse." *and added in the reference into our citations.*

Comment 7: What were the roles of GRIA genes in development? Please state in the discussion.

Reply 7: *Thanks for your comments. We have provided additional insights into the functions of the GRIA genes to enhance reader comprehension in the Discussion section.*

Changes in the text: *On lines 383-386, we have added a description of function of AMPAR (composed of GluA subunits encoded by the GRIA gene) as "*The long-term plasticity of synaptic transmission in the cerebellum is recognized as the key mechanism underlying motor learning. It is hypothesized that the activation of presynaptic AMPARs could be a molecular event involved in this fundamental process(32).*" Additionally, we have also supplemented the role of the GRIA gene in sleep regulation as "*In rats, the transitions between wakefulness and sleep are accompanied by alterations in extracellular glutamate concentrations of cerebral cortex(34). Recent findings provide compelling evidence indicating a notable role for GluA3 channel activity in the regulation of sleep behavior in both mice and humans(33).*" on lines 393-399.*

Comment 8: It is the common sense that knockout mouse models are good tools for predicting disease-causing genes in developmental disorders. What were the significance of the study? Please supplement in the discussion.

Reply 8: *Thanks for your comments. The knockout mouse model was an important experimental system for the biomedical science. But 49.53% of genes linked to various disorders exhibited no overlapping phenotype between the knockout mouse model and the corresponding human disease. By comparing the similarity of gene expression trajectories and phenotypic similarity, we provided a potential explanation for the divergent phenotypes resulting from null mutations in humans and mice. Furthermore, the utilization of gene expression trajectories assists in predicting potential pathogenic genes, enhancing the reliability of the results.*

Changes in the text: *As advised, we have modified our text in the Discussion part on lines 433-440 as "*Overall, the current work is the first to demonstrate the relationship between expression trajectory similarity and phenotypic similarity and provided a potential explanation for the distinct phenotypes resulting from null mutations in humans and mice from a novel perspective. In addition, comparing expression trajectories provides a solid basis for forthcoming investigations involving knockout mouse models. Furthermore, the use of gene expression trajectories might improve the accuracy and reliability of predicting potential pathogenic genes. This approach holds promise for advancing our understanding of gene function and disease mechanisms. Consequently, this study opens up new avenues for future research endeavors of developmental disorders."

Reviewer B

Comment 1: First, the title needs to indicate the research design of this study and needs to be specific to similarities in gene expression trajectories between human and mouse.

Reply 1: *Thanks for your comments. Based on your recommendation, we will ensure that the new title is more specific and accurately describes the focus and objectives of the study.*

Changes in the text: *We have revised the title as "*Gene expression trajectories aiding the unveiling of phenotypic similarities in developmental disorders: insights from human-mouse comparative analysis*".*

Comment 2: Second, the abstract needs to indicate the knowledge gap on this research focus and the objectives of this study. The methods need to describe the variables and genetic data in the databases and how phenotype of developmental disorders were ascertained. The results need to quantify the findings by providing statistics and P values, not the P values only. The conclusion is still repeating the significance of this research focus and I suggest the authors to make the "valuable insights" more clear and have comments for the clinical implications of the findings.

Reply 2: *Thanks for your comments. We have carefully considered your suggestions and have made the necessary revisions in the following aspects:*

1. Regarding the Abstract, we have revised the background of the Abstract section to indicate the knowledge gap and stated the objectives of the study.

2. The variables collected included temporal expression data and phenotypes of human and mouse. The temporal expression data was obtained from bulk transcriptomes provided in an article published in Nature. In this study, the phenotypes of human and mouse were represented using HPO terms and MPO terms, respectively. The HPO and MPO annotations for orthologous genes were respectively sourced from the HPO website and MGI database. To enhance reader comprehension, we have added a description of the sources for HPO terms and MPO terms in the Methods section of the main text. The HPO terms for genes with PTVs in iGluRs genes were manually selected from the retrieved literature, and the MPO terms for gene knockout mice reported in the literature were screened from the MGI database. Detailed screening methods will be provided in the Methods section of the main text.

3. To enhance the presentation of our results, we have supplemented statistical measures to quantify our findings.

4. In the conclusion, we have rephrased our statements and include comments on the clinical implications of our findings.

Changes in the text:

*1. We have revised the background of abstract as "*Developmental disorders result from impaired prenatal development, often attributed to deleterious genetic variations. Knockout mouse models are critical tools in diseases research, but null mutations of orthologous genes in human and mouse often resulted in different phenotypes. As phenotypes are described during developmental processes, we explored the temporal patterns of gene expression during development.*" on lines 24-28.*

*2. As advise, we had revised the sentences as "*We utilized Spearman correlation analysis to compute the expression similarity of orthologous genes in temporal expression data of humans and mice.*" on lines 30-31. To explain the acquisition of phenotype similarity scores, we have revised the sentence in the Abstract on line 34-36 as follows "*The phenotypic similarity of orthologous genes can be quantified using Phenodigm scores derived from the calculations based on Human Phenotype Ontology (HPO) and Model Phenotype Ontology (MPO) terms.*". In addition, we have added a description of the sources for HPO and MPO terms in the Methods section of the main text as "*The HPO terms were initially created by developing an ontology using information from the Clinical Synopsis of the OMIM database and gradually collaborating with clinicians in workshops to enhance and expand the clinical terminology(9). Phenotypes of mouse for each gene were described using MPO terms obtained from MGI database(17) (www.informatics.jax.org/). The MPO serves as a structured vocabulary, allowing the consistent annotation of mouse phenotypic data from various sources (e.g., published literature, large-scale mutagenesis centers, individual research laboratories) using standardized phenotype terminology(18).*" on lines 152-159. We have also provided detailed screening methods in the Methods section of the main text as "*The phenotypes of patients harbouring PTVs were independently evaluated by two physicians, as per the assessment using the HPO terms. Relevant HPO terms associated with the gene in question were obtained from the official HPO website, and those aligning with the clinical descriptions found in the literature were carefully chosen for analysis. We also curated the MPO terms for gene knockout mice with documented literature reports in the MGI database and verified them within the manuscript.*" on lines 164-169.*

3. We have added standardized test statistic Z as "Additionally, in DDG2P genes, viable with phenotype (VP) genes with similar expression trajectories in the brain exhibited analogous phenotypes across humans and mice (AD genes: Wilcoxon test: Z = 11, P = 0.02; AR genes: Wilcoxon test: Z = 9, P = 0.003)." on lines 42-45. We also supplemented statistics in the main text.

*4. The Conclusions of Abstract have been revised as "*The gene expression trajectory offered a potential elucidation for the disparate phenotypes observed due to null

mutations in human and mouse. Leveraging gene expression trajectories might enhance the precision in predicting candidate pathogenic genes*." on lines 51-53.*

Comment 3: Third, in the introduction of this main text, the sentence "provide a new idea for investigating phenotypes of developmental disorders" is unclear. The authors need to provide more detailed comments and insights on the possible clinical significance and implications of this research focus. Please also have comments on the knowledge gaps on this research focus in the literature.

Reply 3: *Thanks for your comments. The knockout mouse model was an important experimental system for the biomedical science. But 49.53% of genes linked to various disorders exhibited no overlapping phenotype between the knockout mouse model and the corresponding human disease. Previous studies have revealed that there are no differential gene expression levels in orthologous genes between human and mouse within the same tissue. Therefore, by comparing the similarity of gene expression trajectories and phenotypic similarity, we have provided a potential explanation for the divergent phenotypes resulting from null mutations in humans and mice. Furthermore, the utilization of gene expression trajectories assists in predicting potential pathogenic genes, enhancing the reliability of the results.*

Changes in the text: *We have modified our text as advised on lines 72-79 as "*Genes associated with the developmental process are precisely regulated(5). Species with comparable developmental trajectories often display shared patterns in conserved regulatory networks and molecular processes that contribute to observed phenotypic similarities. Previous studies have revealed that half of human genes showed distinct temporal trajectories comparing to their mouse orthologs(6). Zheng et al, found no significant contribution of expression level divergence between human and mouse orthologous genes across homologous tissues of the two species from microarray expression datasets(7). It's unclear whether the similarity of developmental expression trajectories could serve as an informative index for phenotypic similarity as widely assumed." *Additionally, we have added possible clinical significance in the end of the Introduction section as "*Leveraging gene expression trajectories can facilitate predicting candidate pathogenic genes, offering valuable opportunities for understanding the developmental disease and the discovering the potential disease-related genes." *on lines 103-107.*

Comment 4: Fourth, in the methodology of the main text, please have a brief overview of the research procedures and the questions to be answered by these procedures at the beginning of this section. Please also describe the research methodology of this study. Please have detailed information on the databases

including the gene data and how the phenotype of developmental disorders were measured in both human and mouse.

Reply 4: *Thank you for your valuable feedback. We appreciate your suggestions to improve the Methods section of our manuscript. In response to your comments, we have made the following revisions:*

1. We have added a brief overview of the research procedure, including the questions addressed by these procedures and the research methodology of this study at the beginning of the Methods section.

2. This study involves gene data comprising 17150 one-to-one human to mouse orthologous gene pairs. Our primary focus is on 2,187 orthologous gene pairs associated with developmental disorders, sourced from the DDG2P database. These gene pairs were classified using the FUSIL categorization. We have added more detailed information on the DDG2P database and FUSIL categorization.

3. In this study, the phenotypes of human and mouse were represented using HPO terms and MPO terms, respectively. The HPO and MPO annotations for orthologous genes were respectively sourced from the HPO website and MGI database. We have added a description of the sources for HPO terms and MPO terms in the Methods section of the main text. The HPO terms for genes with PTVs in iGluRs genes were manually selected from the retrieved literature, and the MPO terms for gene knockout mice reported in the literature were screened from the MGI database. Detailed screening methods will be provided in the Methods section.

Changes in the text:

*1. As advised by the reviewer, we have added the Research procedure at the beginning of Methods part on lines 111-120 as "*In order to investigate the feasibility of utilizing mouse models to simulate human developmental disorders, we employed the HPO and MPO terms to compute original Phenodigm scores, serving as indicators of phenotypic similarity. Furthermore, we conducted Spearman correlation coefficient analysis on temporal gene expression data from corresponding tissues in humans and mice to assess the similarities in gene expression trajectories. To mitigate potential confounding effects from variants with diverse functionalities, we meticulously screened cases involving protein truncating variants (PTVs) in the iGluRs gene family, utilizing the HPO terms. Furthermore, we employed the MPO terms in gene knockout mouse models. Through the computation of refined Phenodigm scores based on HPO terms and MPO terms, our study aimed to elucidate the feasibility of harnessing gene expression trajectory similarity as a valuable tool to facilitate the prediction of phenotypic similarities."

*2.We have provided detailed description of DDG2P database as "*The database was created by Deciphering Developmental Disorders (DDD) Study, which has

recruited >13,400 individuals with undiagnosed severe and/or extreme developmental disorders from the UK and Ireland. The main objective of the project is to define the genetic architecture of developmental disorders, while also aiming to identify new developmental disorders loci(19).*" on lines 174-178. We have also supplemented the specific classification criteria of the categorization as "*Genes were classified according to full spectrum of intolerance to loss-of-function (FUSIL) bin(20), which took advantage of the comprehensive organismal viability screen performed by the International Mouse Phenotyping Consortium (IMPC) database (https://www.mousephenotype.org/understand/data-collections/essential-genes-portal/) and the cellular viability studies conducted by Project Achilles Avana data set (https://depmap.org/portal/achilles/)*" on lines 179-183.*

*3. We have added a description of the sources for HPO and MPO terms in the Methods section as "*The HPO terms were initially created by developing an ontology using information from the Clinical Synopsis of the OMIM database and gradually collaborating with clinicians in workshops to enhance and expand the clinical terminology(9). Phenotypes of mouse for each gene were described using MPO terms obtained from MGI database(17) (www.informatics.jax.org/). The MPO serves as a structured vocabulary, allowing the consistent annotation of mouse phenotypic data from various sources (e.g., published literature, large-scale mutagenesis centers, individual research laboratories) using standardized phenotype terminology(18).*" on lines 152-159. We have also provided detailed screening methods in the Methods section as "*The phenotypes of patients harbouring PTVs were independently evaluated by two physicians, as per the assessment using the HPO terms. Relevant HPO terms associated with the gene in question were obtained from the official HPO website, and those aligning with the clinical descriptions found in the literature were carefully chosen for analysis. We also curated the MPO terms for gene knockout mice with documented literature reports in the MGI database and verified them within the manuscript." *on lines 164-169.*