nature portfolio

Peer Review File

Spatiotemporal Brain Transcriptomics Reveal Risk Gene Hot-Spots in Major Neuropsychiatric Disorders

Corresponding Author: Professor Weiqing Liu

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript looks at risk gene sets for different neuropsychiatric disorders (obtained from other GWAS studies) and uses rnaseq and scrnaseq data to identify spatial and temporal expression patterns to establish neuronal hot spots for these disorders.

Fig 3 doesn't really seem to support the interpretations they are making. They say "Across the 15 traits examined, risk gene sets exhibited consistent enrichment across cortical subregions, while subcortical structures showed disorder-specific enrichment profiles, especially for OCD, ADHD, Panic disorder, and mood-related traits, including Neuroticism, BIP, and MDD. In contrast, risk gene sets for IQ and ASD did not show prominent subcortical expression enrichment." The patterns on the heatmaps do not strongly indicate this and the zscores sseem to alternate between high (Red) and low (blue) within each brain region, potentially indicating disease samples vs controls. They also mention using "qualitative evaluation" to determine hotspots. A different way of visualizing this or a quantitative measure may be more convincing.

For Fig 3, statement like this "we infer that the first t-SNE dimension (tSNE1) primarily captures temporal variation, while the second (tSNE2) reflects spatial differences." is misleading because the
dimensions produced by tsne cannot be interpreted in this way (like a PCA can). They may want to use PCA/DAPC for this. Overall, there seems to be an incorrect application of tsne/incorrect language around it. tsne is not a clustering algorithm, so using a clustering algorithm to identify those 3 clusters would be more meaningful. The hierarchical clustering results in S1 make more sense to include in the main text, but they may not be showing the temporal clustering the authors want to show. While I can see what the attemp wa with Fig 3b, the way the points have been grouped to make 3 groups seems subjective and bit arbitrary.

For fig 5, I'd recommend putting the disorders in an order such that disorders showing similar patterns in spatial expression are next to each. Also, all regionss talked about in the figure description hsould be labeled (like striatum). What exactly is meant by "Regional enrichment “hot-spots” are labeled accordingly. "

WGCNA:
Fig S5 a and b are too low resolution to zoom in and see the beta and pvalues.Highlighting the 32 moduless that met the cutoffs on this figure would be useful. S5c seems to show how significantlly enriched each module is with risk genes. This way of representing it is not very clear + the size of these gene modules is not communicated here. Instead, consider doing a bar chart where height of the bar chart indicates size of module and color indicates pvalue. Same recommendation for Fig 6C.

Is there a justification for these somewhat arbitrary cutoffs for selecting 16 modules: "To further prioritize biologically relevant modules, we focused on 16 that exhibited the strongest enrichment for disorder-associated risk genes by selecting the top three modules per trait based on odds ratios"

Overall, lot of text refers to supplementary figures. If these figures are this pertinent, the authors should consider including some parts of the supplementary figures in the main manuscript.

Reviewer #2

(Remarks to the Author)
I have read the manuscript with great interest and appreciate this work. This study presents a comprehensive and ambitious analysis of neuropsychiatric disorder risk gene expression patterns and hot-spots, providing valuable insights into potential disease mechanisms. However, I have several major and minor comments that need to be addressed before the manuscript can be considered for publication.

Major Comments

1. Introduction (Line 51 - "unbiased"): The term "unbiased" is too strong. Even analyses derived purely from hypothesis-free, data-driven approaches are subject to various forms of bias (e.g., data quality, selection bias, methodological choices). I would recommend replacing "unbiased" with a more accurate term such as "data-driven" or "hypothesis-free".

2. Results: "Spatial enrichment patterns of neuropsychiatric risk gene sets implicate disease-relevant brain regions and circuits" section:
a. Consistency of Enrichment: The claim, "Across all the 15 traits, risk gene sets exhibited consistent enrichment patterns within cortical sub-regions," is not visually supported by Fig3. Specifically, the areas emphasized by the dashed boxes do not appear "consistent" upon visual inspection. Please provide statistical evidence (e.g., a formal measure of correlation or similarity) to rigorously support this claim of "consistent patterns."
b. Statistical Analysis for Enrichment: When mentioning specific enrichments, such as the ADHD risk genes in the thalamus and striatum (Line 121), the authors should clarify if corresponding statistical analyses were performed to confirm the significance of these findings.
c. Underreported Signals: Figure 3 visually suggests a stronger enrichment signal for ADHD risk genes in the cerebellum compared to the thalamus and striatum. Similarly, MDD risk genes show a strong signal in the cerebellum, potentially stronger than the amygdala and striatum mentioned in the text. Please address these strong, yet unmentioned, signals in the text.

3. Results: "Spatiotemporal expression patterns of risk gene sets differentiate neuropsychiatric disorder subgroups" section:
a. Missing Traits in Classification: The text describes a t-SNE classification resulting in three categories, yet states that some diseases, such as ADHD, were not included in the initial t-SNE analysis. However, a later result mentions ADHD as being part of the "Prenatal subcortical" cluster. Please clarify this discrepancy and ensure all diseases discussed in the clustering analysis were properly included and accounted for.

4. Results: "Assessing the biological relevance of risk gene expression patterns using ENIGMA brain morphometry" section:
a. Multiple Testing Correction: The analysis correlating risk gene expression patterns with ENIGMA brain morphometry changes (Fig 5) needs a clear statement on multiple testing correction. Given that several correlation p-values in Figure 5 are very close to the p=0.05 threshold, it is critical to confirm if and how multiple testing correction (e.g., Bonferroni or FDR) was applied to control the false discovery rate for this analysis.

5. Results: "WGCNA-derived co-expression modules and their functional implications in neuropsychiatric disorders" section:
a. Reference Category Justification: The choice of "thalamus during young adulthood was set as the reference category" for this analysis (Line 234) must be rigorously justified. Please cite appropriate literature or provide a clear, biologically or statistically sound rationale for selecting this specific reference category.
b. Undiscussed Modules: Figure S5 indicates enrichment signals for some modules in the thalamus that were not discussed. The authors should address the relevance of these additional modules in the Results and Discussion sections.

6. Results (Lines 234-237) and Figure S6 Discrepancy: There appears to be a direct contradiction between the text and Figure S6.
The text states "modules overrepresented during early developmental stages were enriched in fetal cell types... In contrast, modules predominant in the perinatal and early childhood stages showed enrichment in glial cells." However, if I understand it correctly, Figure S6 visually suggests the opposite for the latter part: modules overrepresented in early childhood are enriched in glial cells, while modules predominant in the perinatal and early childhood stages (ME magenta, ME yellow, ME green, ME purple, ME darkgreen, and ME turquoise) appear to show enrichment in fetal cell types. Please re-verify the text against the figure and correct the description to accurately reflect the results shown in Figure S6.

7. Results: Glial vs. Neuronal Enrichment: The authors classify a group as "Glia-enriched" (e.g., Line 237), yet this group contains "Neurons" cells. Could the observed categorization or downstream results be dominated by the neuronal cell data? To confirm the specificity of the glia enrichment, I suggest performing a sensitivity analysis where the neuronal cell data is removed or down-weighted and the categorization/enrichment analysis is repeated to assess its robustness.

8. Results (Line 252 - "significantly overrepresented"): When stating that "M1 and M4 were significantly overrepresented in endothelial cells," the authors must provide statistical evidence (e.g., p-value, corrected p-value, or other relevant metric) to support the use of the term "significantly overrepresented."

9. Methodology: Data Integration and Harmonization: The Methods section indicates the use of multiple datasets and mentions analyses "across samples," suggesting data integration. Please provide a detailed explanation in the Methods section regarding how different datasets were combined (if they were) and, crucially, the specific harmonization techniques (e.g., batch effect correction) employed to ensure comparability and minimize bias between datasets.

10. Methodology: WGCNA Module Size: In the WGCNA section (Line 541), the authors specify a minimum module size of 40. Was a maximum module size also set? Please comment on whether significant differences in module size exist across the identified modules and if this variation could potentially impact or bias the downstream analysis results.

11. Methodology: Risk Gene Set Currency: The risk gene sets used for the analysis may not be the most current. While I appreciate the extensive effort involved in the current work, it would significantly strengthen the manuscript if the authors could replicate a core analysis for one or two key disorders using the most recently published risk gene sets. This would help determine the sensitivity of the main findings to the choice of input gene lists.

Minor Comments

12. Consistency: In the Results section, please ensure consistency in naming brain regions. Use either the full name (e.g., striatum) or the abbreviation (e.g., Str) uniformly throughout the text.

13. Disease Name Capitalization: Please standardize the capitalization of disease names (e.g., ADHD, MDD, SCZ) throughout the entire manuscript. Currently, there is an inconsistent mix of all-caps and title-case/lowercase initial letters (e.g., Panic disorder).

I look forward to reviewing a revised version of the manuscript that addresses these points. The core findings are potentially important, and addressing these concerns will significantly enhance the clarity, rigor, and impact of the work.

Reviewer #3

(Remarks to the Author)
This study presents a comprehensive and novel spatiotemporal analysis of genome-wide risk gene sets across 15 major neuropsychiatric disorders. By integrating bulk transcriptomics, single-cell RNA-seq data, GWAS, and ENIGMA data, the authors establish a multidimensional framework linking gene expression dynamics to disease mechanisms and clinical onset timing.

Though these analyses, the study identifies distinct spatiotemporal “hot-spots” of risk gene expression that correspond to critical developmental periods and disease-relevant brain regions. These patterns enable the classification of disorders into three biological meaningful subgroups: prenatal cortical, prenatal subcortical, and postnatal. Moreover, the authors uncover co-expression modules enriched for disease risk genes and categorize them into glia-enriched, adult neuron-enriched, and fetal-enriched groups. Notably, pleiotropic modules such as M14 and M16 implicates early developmental processes, including forebrain synapse organization and chromatin regulation, suggesting share pathogenic mechanisms across multiple disorders.

The study provides comprehensive and valuable analyses, offering important new insights. Although I have a few quibbles and believe the manuscript would benefit from some refinement, overall it represents a meaningful contribution to the scientific literature and is likely to be well-received.

Minor concerns

1) The overall structure of the abstract and results should be improved, as the key message is too broad and lacks specificity. First, it does not clearly explain the motivation of the study or how it addresses a gap compared to previous research, and the analytical methods are presented in a list-like manner. The results do not specify what the three clusters represent or what the “hot-spots” refer to. A description of the associated cell types and biological functions is also needed.

2) Figure 2b: A more detailed explanation is needed regarding which specific early-onset disorders and neurological disorders are referred to in Fig.2B, what their typical ages of onset are, and how these onset ages are related to the results.

3) General visualization of figures:
Emphasizing specific disorders and their associated expression values directly within the heatmap may limit the reader’s ability to interpret the data objectively (Fig.3). It would be more appropriate to present expression values for the disorders of interest separately, for instance using bar plots or dot plots.
As in Fig.4A, please modify the t-SNE visualization so that grouped clusters are indicated based on precise values, such as a 95% confidence interval.

4) Figure S1: It is difficult to conclude that there is a clear difference between the two groups on the heatmap in panel (a) because certain stage doesn’t show a distinct separation. For example, the young adulthood stage does not exhibit a clear distinction. Statistical comparisons between the early-expression cluster and the late-expression cluster would be necessary.
In addition, late infancy shows high expression across the entire disease spectrum. Therefore, a discussion focusing on late

infancy, which appears to be most strongly associated with the disease, is also needed to be explained.
It is necessary to explain in the figure legend for panel (c) what each stage represents in terms of developmental phases. In addition, since the values for many disorders are presented within a single graph, it is difficult to distinguish the data. It would be preferable to redraw the figure using an alternative visualization, such as a dot plot.

4) Figure 4b:
Multiple GWAS datasets for the same disorder may not exhibit the highest r value in the same brain region: instead, each dataset can show its strongest correlation in different regions. For example, in the case of epilepsy, the highest r values are observed in the CB, AMY and MGE regions, whereas for ASD, they are found in the A1C, M1C, and DFC regions. It is therefore important to discuss whether the regions showing high r values share functional relevance related to disease, and what underlying reasons should explain these results.

6) Figure S5: The number of modules identified by the authors is quite large, and some of them exhibit highly similar expression profiles. For example, the darkturquoise, darkolivegreen, and plum1 modules display similar patterns. To further condense the clusters into more biologically meaningful modules, it would be appropriate to cluster the modules based on pairwise eigengene correlations and merge those that share similar expression profiles.
Additionally, after such merging, the authors should clearly address why the ADHD x MEdarkolivegreen combination shows the highest odds ratio (OR) yet a non-significant P value. While this might be due to the small size of the ADHD risk gene set, other possible explanations should also be discussed explicitly.

7) Figures 6, S7
The authors divided the modules into three groups: Glia-enriched, Adult neuron-enriched, and Fetal-enriched. However, it is unclear whether these groups are truly appropriate to be clustered together as shown in the heatmap, and whether each group exhibits a clear difference from the others. Statistical metrics that support and justify this grouping are needed.
In addition, beyond the odds ratios presented in Fig.6C, statistical significance values such as P values should also be provided.
Moreover, in module M4, the odds ratio is higher for OCD than for AD. An interpretation of this observation should be included.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
This manuscript looks at risk gene sets for different neuropsychiatric disorders
(obtained from other GWAS studies) and uses rnaseq and scrnaseq data to identify spatial
and temporal expression patterns to establish neuronal hot spots for these disorders. It is of great value in this field.

I greatly appreciate the changes the authors have made which clarify the results greatly.

The authors have addressed my concerns about figure 3 and the quantitative analysis is appreciated. The quantitative analysis is currently presented mainly in supplementary figures s3a and b. s3 b does a reasonable job summarizing th results, but s3a is confusing. It's blurry, the asterisks representing the pvalues are very difficult to see and the meaning behind the colored bars in some of the scatterplots in unclear. So, I'd suggest fixing this or removing it.

I appreciate that the authors switched to pca for figure 4 and the results, especially with the correlation plots are clear here.

The changes to figure 5 also clear up my questions.

The changes to figure s7 lso make the wgcna results much clearer.

Thank you!

Reviewer #2

(Remarks to the Author)
I appreciate the authors' careful revisions and believe that the manuscript is now much stronger.

Overall, the majority of the concerns I raised in the previous review have been adequately addressed, and the manuscript has been improved considerably. The analyses are now clearer, and the presentation has been strengthened.

One issue that remains, however, is the inconsistency in Fig 3 between the rebuttal letter and the main text regarding the classification results for Panic Disorder. I recommend that the authors carefully review this point and ensure that the description and the figure are fully consistent.

Open Access This Peer Review File is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.
In cases where reviewers are anonymous, credit should be given to 'Anonymous Referee' and the source.
The images or other third party material in this Peer Review File are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.
To view a copy of this license, visit https://creativecommons.org/licenses/by/4.0/

Reviewers' comments:

Reviewer #1 (Remarks to the Author):

This manuscript looks at risk gene sets for different neuropsychiatric disorders (obtained from other GWAS studies) and uses rnaseq and scrnaseq data to identify spatial and temporal expression patterns to establish neuronal hot spots for these disorders.

1. Fig 3 doesn't really seem to support the interpretations they are making. They say "Across the 15 traits examined, risk gene sets exhibited consistent enrichment across cortical subregions, while subcortical structures showed disorder-specific enrichment profiles, especially for OCD, ADHD, Panic disorder, and mood-related traits, including Neuroticism, BIP, and MDD. In contrast, risk gene sets for IQ and ASD did not show prominent subcortical expression enrichment." The patterns on the heatmaps do not strongly indicate this and the zscores sseem to alternate between high (Red) and low (blue) within each brain region, potentially indicating disease samples vs controls. They also mention using "qualitative evaluation" to determine hotspots. A different way of visualizing this or a quantitative measure may be more convincing.

Response:

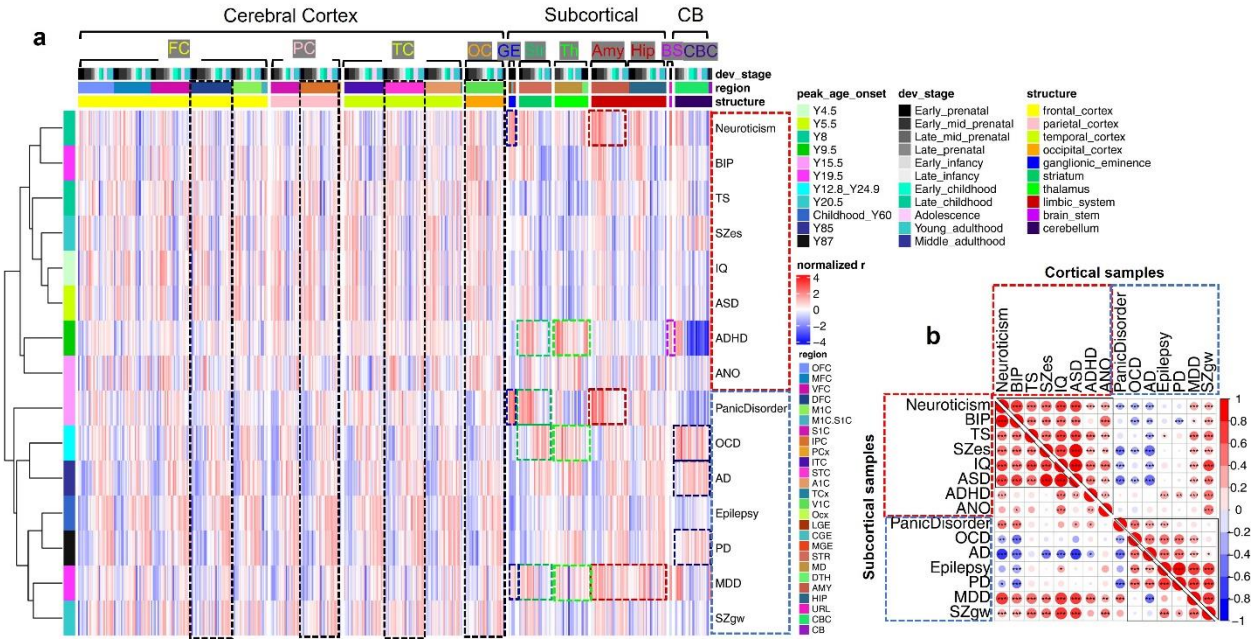
We thank the reviewer for highlighting the limitations of qualitative interpretation based on heatmap visualization alone. In response, we have added a quantitative analysis that formally evaluates risk gene expression enrichment across major brain regions, and we have revised the figures and text accordingly.

Specifically, we compared risk gene expression enrichment between cerebral cortical regions and subcortical structures for each of the 15 traits (revised Fig. S3). The frontal cortex was used as the reference structure due to its largest sample size in the BrainSpan dataset and its well-established relevance to cognition and psychiatric disorders. For each trait, enrichment levels in all other brain structures were statistically compared with those in the frontal cortex using t-tests, followed by FDR correction across structurs.

This quantitative analysis demonstrates that cerebral cortical subregions exhibit broadly comparable enrichment levels across most traits, supporting the notion of relatively conserved cortical enrichment patterns. In contrast, subcortical structures show markedly greater heterogeneity and trait specificity. Notably, the ganglionic eminence displays significantly higher enrichment for neuroticism, panic disorder, and major depressive disorder relative to cortical regions, while the amygdala shows selective enrichment for mood-related traits. Although based on a limited number of samples, the brainstem consistently shows elevated enrichment for neuroticism, ADHD, and MDD, and the cerebellum exhibits modest but significant enrichment for OCD, Alzheimer's disease, and Parkinson's disease.

In addition, to capture global pattern similarity among traits beyond individual region-wise comparisons, we examined cross-trait relationships based on enrichment profiles in cortical and subcortical samples separately. As shown in the revised Fig. 3b, cortical

samples exhibit higher intra-cluster homogeneity than subcortical samples, consistently observed in both prenatally and postnatally enriched disorder groups.
Together, these analyses replace the previously qualitative identification of enrichment "hotspots" with explicit statistical comparisons and similarity-based metrics, providing quantitative support for our interpretation that cortical enrichment patterns are relatively consistent across traits, whereas subcortical enrichment profiles are more disorder-specific.



Revised Fig. 3 Spatial enrichment patterns of neuropsychiatric risk gene sets highlight disease-relevant brain regions and circuits.
Changes made: Correlation matrices illustrating the relationships among the 15 traits based on enrichment patterns in cortical samples (upper right triangle) and subcortical samples (lower left triangle) was added as Fig. 3b.

2. For Fig 4, statement like this "we infer that the first t-SNE dimension (tSNE1) primarily captures temporal variation, while the second (tSNE2) reflects spatial differences." is misleading because the dimensions produced by tsne cannot be interpreted in this way (like a PCA can). They may want to use PCA/DAPC for this. Overall, there seems to be an incorrect application of tsne/incorrect language around it. tsne is not a clustering algorithm, so using a clustering algorithm to identify those 3 clusters would be more meaningful. The hierarchical clustering results in S1 make more sense to include in the main text, but they may not be showing the temporal clustering the authors want to show. While I can see what the attemp wa with Fig 3b, the way the points have been grouped to make 3 groups seems subjective and bit arbitrary.

   Response:

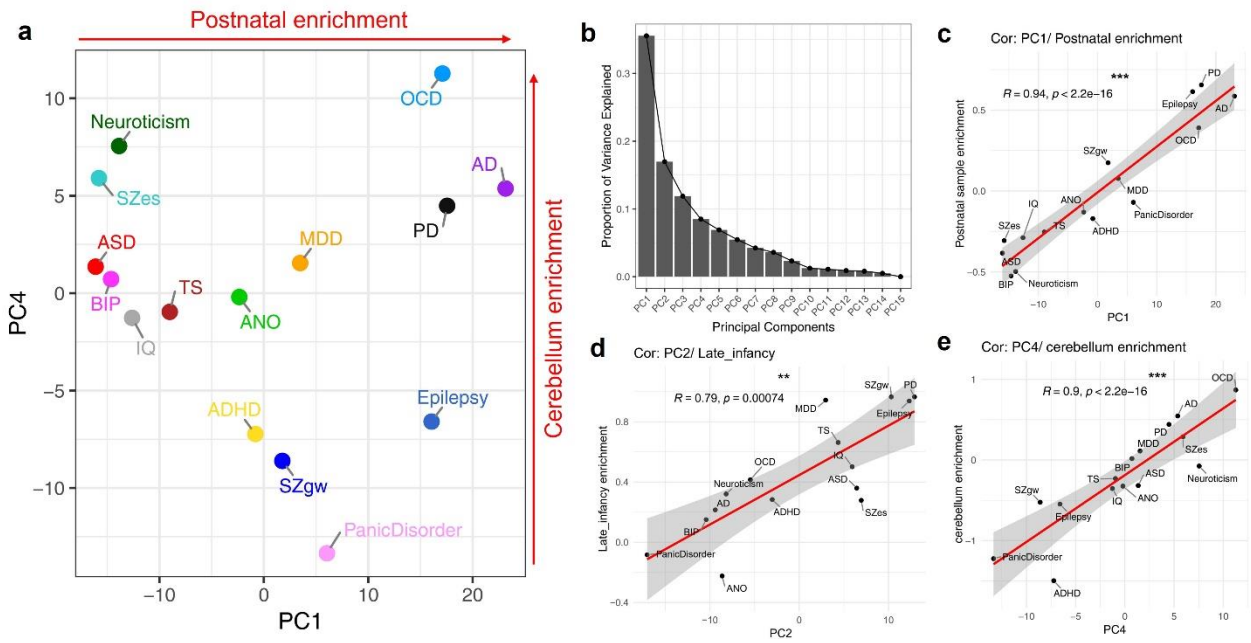
We thank the reviewer for this important methodological comment. We fully agree that t-SNE dimensions are not directly interpretable in terms of specific biological axes, and that our previous description risked overstating the meaning of individual t-SNE coordinates. Moreover, t-SNE is a visualization method rather than a clustering or inferential tool, and grouping traits based on visual proximity can be subjective.

To address these concerns, we have replaced the t-SNE–based analysis with a principal component analysis (PCA), which provides orthogonal, variance-decomposed components that are explicitly interpretable and statistically testable. PCA was performed on the Z-scored rank biserial correlation coefficients of the 15 disorder-associated risk gene sets across all BrainSpan samples (revised Fig. 4 and Fig. S4).

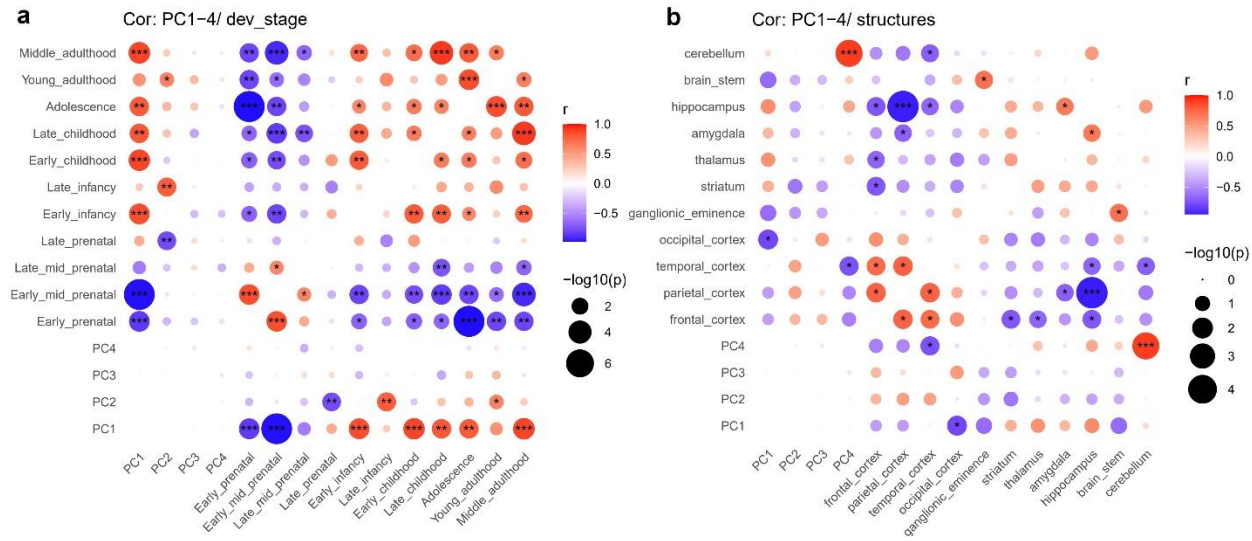
The first two principal components (PC1 and PC2) and a higher-order component (PC4) together explain more than 72% of the total variance in spatiotemporal enrichment patterns. Importantly, the biological interpretation of these components was not assumed a priori. Instead, we quantitatively assessed their associations with developmental stage–specific and structure-specific enrichment variables. These analyses demonstrate that PC1 is strongly associated with prenatal–postnatal enrichment differences, PC2 captures enrichment dynamics specific to late infancy and young adulthood, and PC4 is selectively associated with cerebellar enrichment (revised Fig. 4c–e and Fig. S4), with key associations surviving FDR correction.

We now present the relationships among the 15 traits as a continuous distribution in the space defined by temporally and spatially interpretable principal components (PC1 and PC4), rather than assigning traits to discrete clusters. Traits that appear proximal in this space therefore share similar spatiotemporal enrichment profiles in a quantitative sense, without implying hard cluster membership. This approach avoids subjective grouping and aligns more closely with the continuous nature of neurodevelopmental and neuroanatomical variation.

Together, these revisions correct the inappropriate interpretation of t-SNE dimensions, remove arbitrary trait grouping, and replace qualitative visualization with statistically grounded dimensional decomposition and correlation analyses.



Revised Fig. 4 Principal Component Analysis of 15 neuropsychiatric traits based on the spatiotemporal enrichment patterns of risk gene sets.



Revised Fig. S4 Correlation matrix between the PCA components and individual spatiotemporal enrichment variables across 15 neuropsychiatric traits.

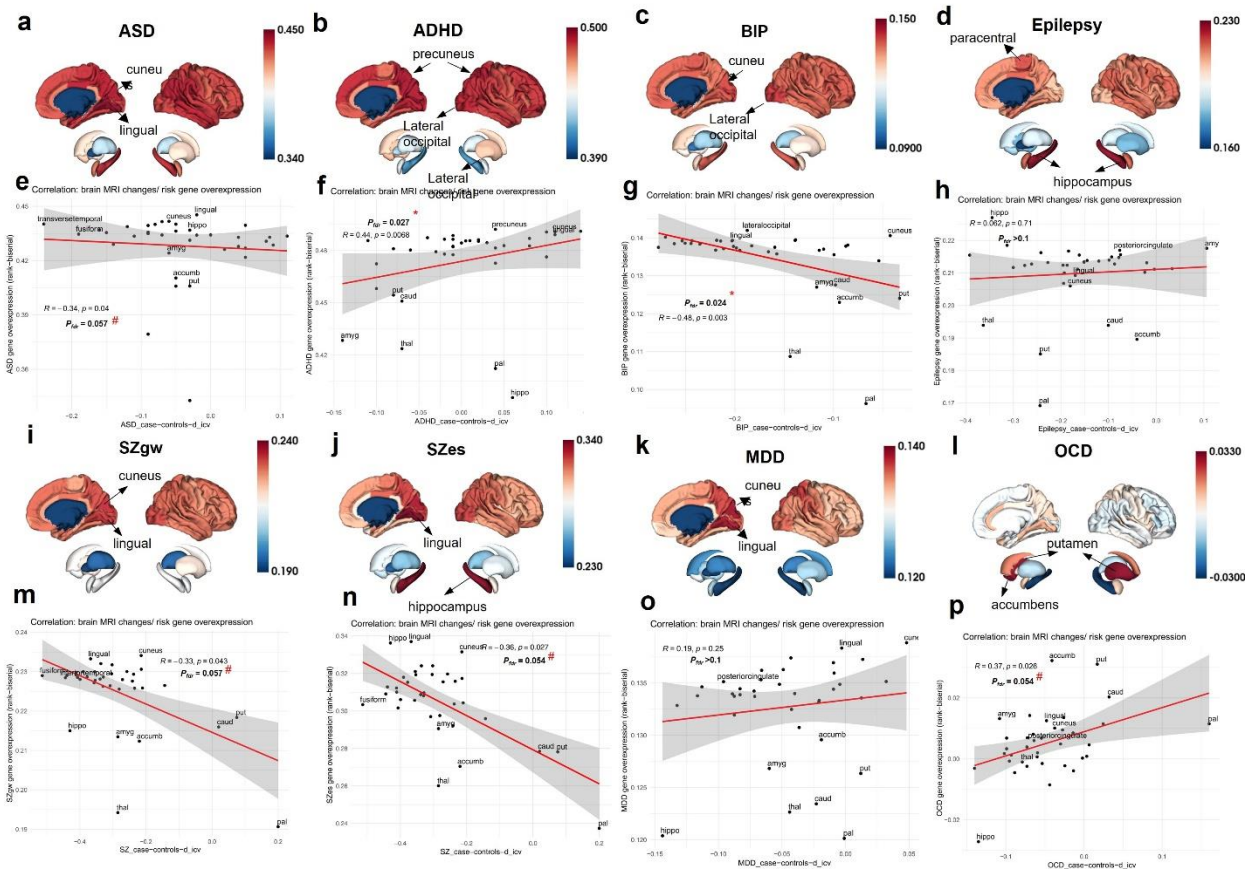
3. For fig 5, I'd recommend putting the disorders in an order such that disorders showing similar patterns in spatial expression are next to each. Also, all regionss talked about in the figure description hsould be labeled (like striatum). What exactly is meant by "Regional enrichment "hot-spots" are labeled accordingly. "

Response:

We thank the reviewer for these helpful suggestions. In the revised Fig. 5, we have reordered the disorders so that those exhibiting similar spatial enrichment patterns are placed adjacent to one another, facilitating visual comparison across traits.

In addition, all brain regions referenced in the figure description (including subcortical structures such as the putamen, nucleus accumbens and hippocampus) are now explicitly labeled in the figure.

We also clarify the definition of regional enrichment "hot-spots." Specifically, for each disorder, hot-spots refer to the two brain regions showing the highest rank-biserial correlation coefficients for risk gene expression enrichment among the 38 ENIGMA-aligned regions. These labels are intended to highlight regions with the strongest relative enrichment and do not imply discrete boundaries or qualitative categorization.



Revised Fig.5 Spatial enrichment of risk gene sets in ENIGMA brain regions and their association with MRI-derived structural alterations in major psychiatric disorders.

4. WGCNA:

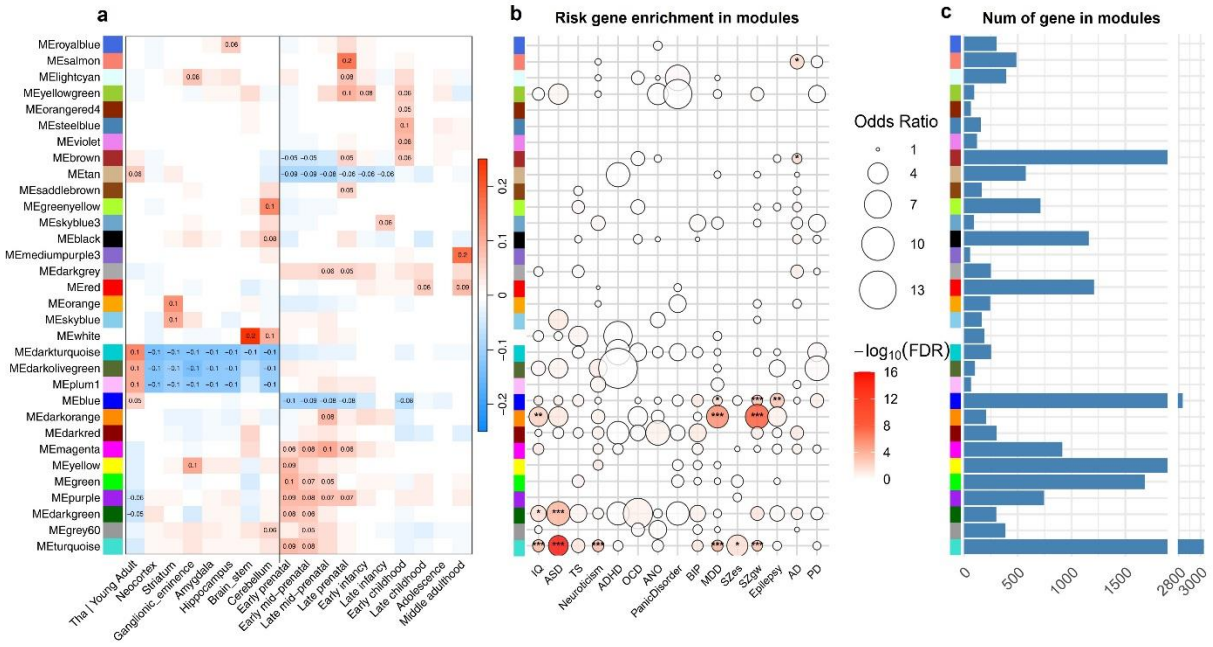
Fig S5 a and b are too low resolution to zoom in and see the beta and pvalues. Highlighting the 32 moduless that met the cutoffs on this figure would be useful. S5c seems to show how significantlly enriched each module is with risk genes. This way of representing it is not very clear + the size of these gene modules is not communicated here. Instead, consider doing a bar chart where height of the bar chart indicates size of module and color indicates pvalue. Same recommendation for Fig 6C.

Response:

We thank the reviewer for these constructive suggestions regarding figure clarity and information content. We agree that the original heatmaps in Fig. S5a and b were difficult to interpret at high resolution, particularly with respect to regression coefficients and P values, and that module size information was not clearly conveyed.

To address these issues, we have implemented several revisions. First, all regression coefficients (β values) and associated P values underlying Fig. S5 are now provided in full in Supplementary Table 10 (Data supporting Fig. S7, Spatiotemporal expression and risk gene enrichment-minModuleSize=40). To improve visualization, the original Fig. S5 is now revised as Fig. S7, and focuses on the 32 modules that showed the strongest spatiotemporal expression enrichment, defined as having at least one spatial or temporal regression coefficient exceeding $\beta > 0.05$, thereby highlighting biologically meaningful expression peaks while avoiding visual clutter. Second, we replaced the original OR/P-value heatmaps in Fig. S5b and c with more interpretable visualizations. Specifically, enrichment of neuropsychiatric risk gene sets across modules is now shown using a bubble plot (Fig. S7b), in which bubble size represents the odds ratio and color encodes -log10(FDR-adjusted P values). In addition, we added a bar plot explicitly showing the size of each co-expression module, allowing effect size, statistical significance, and module size to be interpreted jointly.

The same visualization strategy has been applied to Fig. 6C to maintain consistency across analyses. Together, these changes substantially improve figure readability and more clearly communicate the relationships among spatiotemporal expression patterns, risk gene enrichment, and module size.



Revised Fig. S7 (corresponding to previous Fig. S5) Module eigengene (min size=40) enrichment in spatiotemporal levels and risk gene sets.

5. Is there a justification for these somewhat arbitrary cutoffs for selecting 16 modules: "To further prioritize biologically relevant modules, we focused on 16 that exhibited the strongest enrichment for disorder-associated risk genes by selecting the top three modules per trait based on odds ratios"

Response:

To address this concern, we note that the selection of modules was guided by both statistical significance and effect size considerations. Specifically, enrichment analyses were performed between each disorder-associated risk gene set and all 32 gene modules, followed by false discovery rate (FDR) correction across modules for each disorder. After correction, we observed that for each disorder no more than three modules reached FDR-significant enrichment. For some disorders, fewer than three modules survived FDR correction, likely due to the relatively small size of the corresponding risk gene sets and the resulting limited statistical power. Importantly, several of these non-significant modules nevertheless exhibited relatively large odds ratios, indicating potentially meaningful effect sizes despite not meeting stringent multiple-testing thresholds. Given that odds ratios provide complementary information to p-values by reflecting the magnitude of enrichment rather than statistical confidence alone, we considered these high-OR modules to be biologically relevant candidates.

To facilitate consistent cross-disorder comparisons of spatiotemporal expression enrichment patterns, we therefore selected the top three modules per disorder ranked by odds ratio for downstream analyses. This approach balances statistical rigor with

biological interpretability and has been widely adopted in systems genetics and network-based analyses when comparing enrichment patterns across multiple traits.

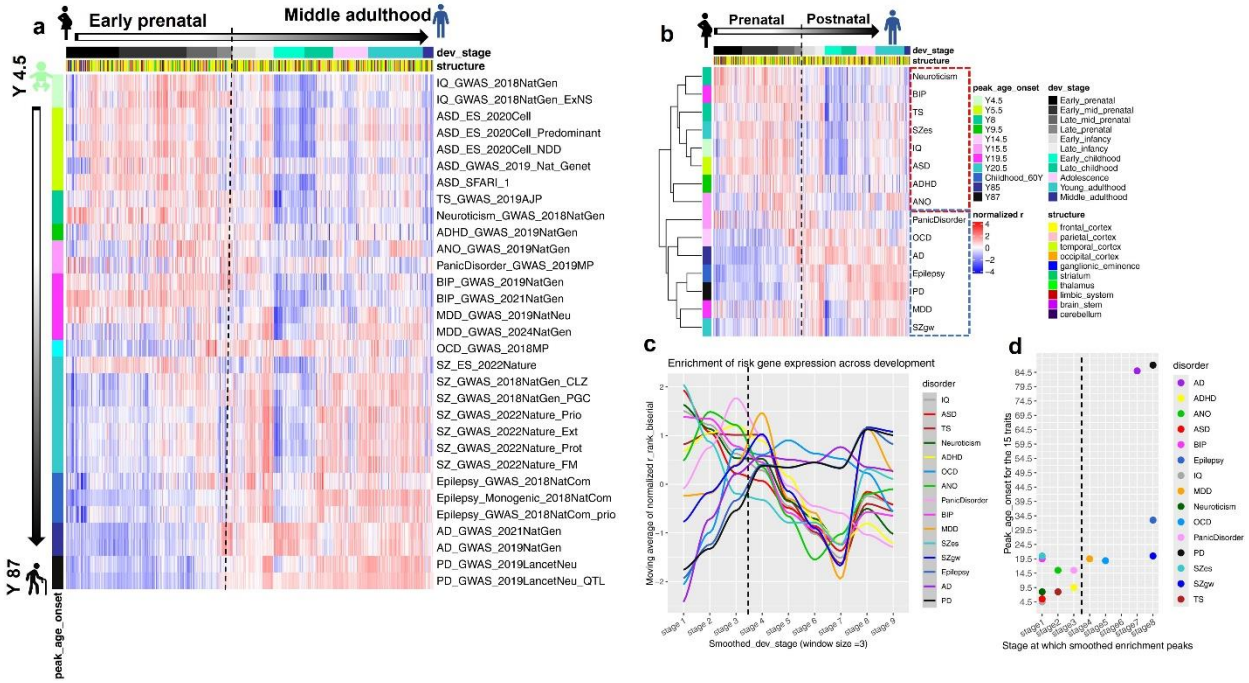
6. Overall, lot of text refers to supplementary figures. If these figures are this pertinent, the authors should consider including some parts of the supplementary figures in the main manuscript.

Response:

We thank the reviewer for this helpful suggestion. We agree that several analyses originally presented in the Supplementary Materials are central to the main conclusions and therefore warrant inclusion in the main manuscript.

In response, we have reorganized the figures by promoting the original Fig. S1—which presents the core developmental clustering of neuropsychiatric disorders based on risk gene expression dynamics—to the main text as Fig. 2. This figure now appears early in the Results section and establishes the fundamental prenatal–postnatal stratification framework that underlies subsequent spatial and integrative analyses.

We believe that this reorganization improves the logical flow of the manuscript, reduces reliance on Supplementary Figures for key interpretations, and enhances accessibility of the central findings for readers.



Revised Fig. 2 (composed of the original Fig. 2b and Fig.S1b-c, and an additional scatter plot showing the relationship between the developmental stage at which the moving-averaged r values peak and the peak age of onset for the 15 traits in panel d) Differential developmental dynamics of risk gene expression in relation to peak age of onset.

Reviewer #2 (Remarks to the Author):
I have read the manuscript with great interest and appreciate this work. This study presents a comprehensive and ambitious analysis of neuropsychiatric disorder risk gene expression patterns and hot-spots, providing valuable insights into potential disease mechanisms. However, I have several major and minor comments that need to be addressed before the manuscript can be considered for publication.

Major Comments

1. Introduction (Line 51 - "unbiased"): The term "unbiased" is too strong. Even analyses derived purely from hypothesis-free, data-driven approaches are subject to various forms of bias (e.g., data quality, selection bias, methodological choices). I would recommend replacing "unbiased" with a more accurate term such as "data-driven" or "hypothesis-free".

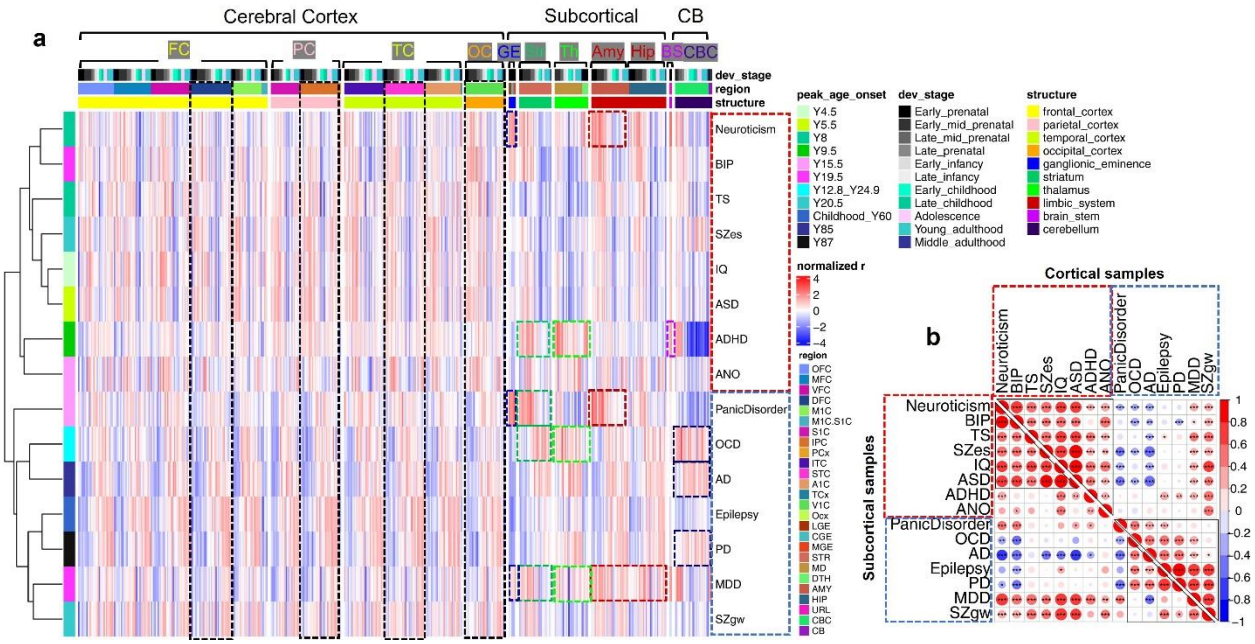
Response:
We thank the reviewer for this important clarification. We agree that the term "unbiased" may be overly strong in this context. Accordingly, we have revised the text to replace "unbiased" with "data-driven, hypothesis-free", which more accurately reflects the nature of genome-wide studies while acknowledging inherent sources of bias. This change has been made in the Introduction (Line 61).

2. Results: "Spatial enrichment patterns of neuropsychiatric risk gene sets implicate disease-relevant brain regions and circuits" section:

   a. Consistency of Enrichment: The claim, "Across all the 15 traits, risk gene sets exhibited consistent enrichment patterns within cortical sub-regions," is not visually supported by Fig 3. Specifically, the areas emphasized by the dashed boxes do not appear "consistent" upon visual inspection. Please provide statistical evidence (e.g., a formal measure of correlation or similarity) to rigorously support this claim of "consistent patterns."

Response:
We thank the reviewer for this important comment. We agree that visual inspection alone is insufficient to support claims of consistency across traits. To address this concern, we performed pairwise correlation analyses among the 15 traits based on their enrichment profiles, separately for cortical and subcortical samples.
As shown in the revised Fig. 3b, cortical samples exhibit significantly higher cross-trait similarity, reflected by greater intra-cluster homogeneity and a higher proportion of statistically significant correlations after Bonferroni correction, in both prenatal and postnatal groups. In contrast, subcortical enrichment patterns show greater heterogeneity across disorders.

We have revised the text accordingly to replace the qualitative description of "consistent" patterns with a statistically supported interpretation based on correlation analyses (Lines 163-165).



Revised Fig. 3 Spatial enrichment patterns of neuropsychiatric risk gene sets highlight disease-relevant brain regions and circuits.

Changes made: Correlation matrices illustrating the relationships among the 15 traits based on enrichment patterns in cortical samples (upper right triangle) and subcortical samples (lower left triangle) was added as Fig. 3b.

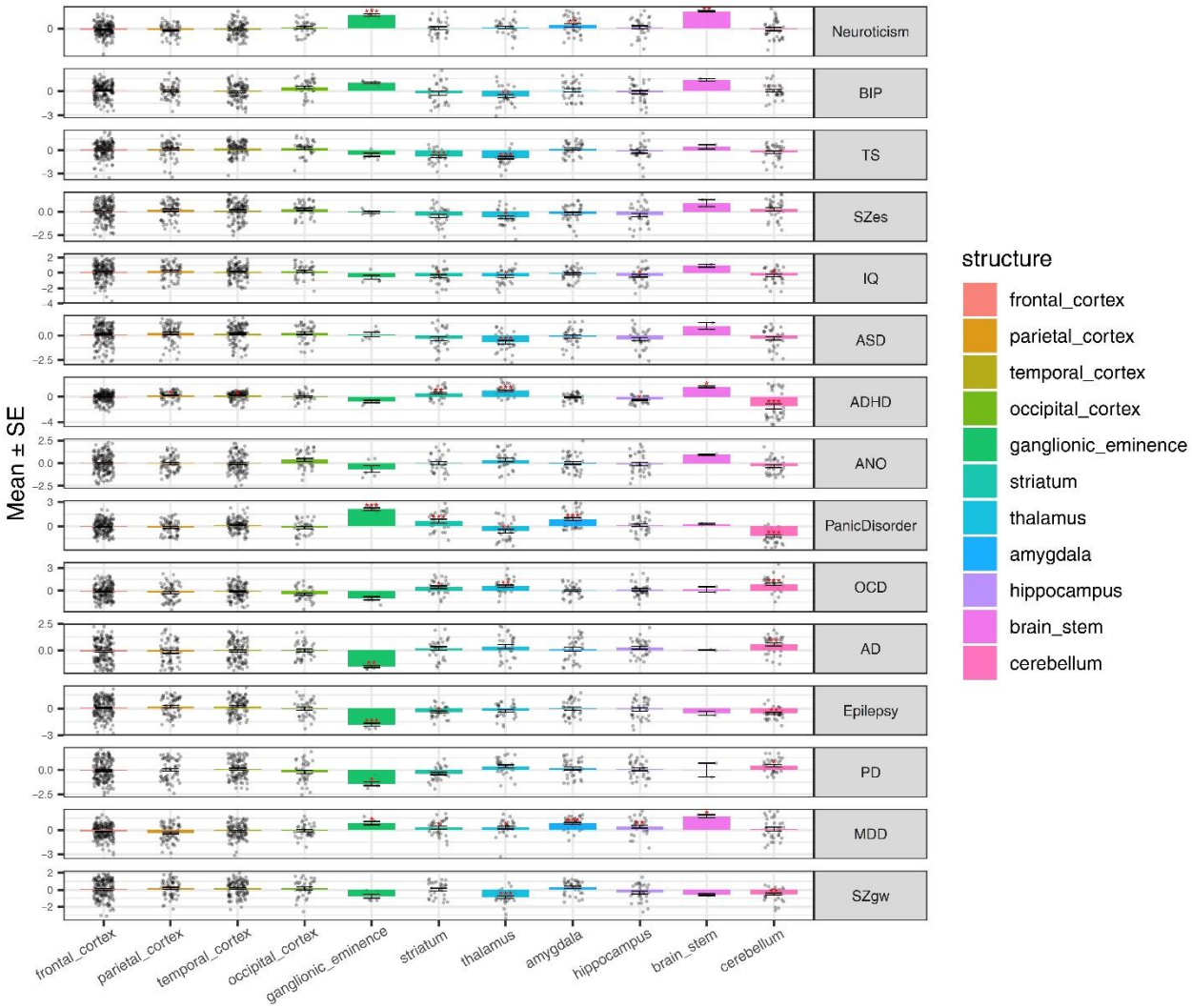
b. Statistical Analysis for Enrichment: When mentioning specific enrichments, such as the ADHD risk genes in the thalamus and striatum (Line 121), the authors should clarify if corresponding statistical analyses were performed to confirm the significance of these findings.

Response:

We thank the reviewer for this important point. To formally evaluate the statistical significance of region-specific enrichment patterns, we performed quantitative comparisons of risk gene expression enrichment across major brain structures for all 15 traits (revised Fig. S3).

Specifically, enrichment levels in each brain region were compared with those in the frontal cortex using t-tests, followed by FDR correction across regions. These analyses confirmed that ADHD risk genes show significantly higher enrichment in the thalamus and striatum, and that panic disorder risk genes are significantly enriched in the ganglionic eminence, striatum, and amygdala (Fig. S3).

We have revised the Results text accordingly to explicitly reference these statistical analyses and avoid reliance on qualitative descriptions alone (Lines 168-169).



Revised Fig. S3a Comparison of risk gene expression enrichment across major brain regions. (a) Bar plots with individual sample points illustrate the relative enrichment of risk gene set expression across major brain structures.

Disorder	FC	GE	Str	Th	Amy	Hip	BS	CB
Neuroticism	Ref	↑↑↑			↑↑		↑↑	
BIP	Ref			↓↓				
TS	Ref		↓↓↓	↓↓↓				
SZes	Ref			↓				
IQ	Ref		↓	↓		↓		↓
ASD	Ref			↓↓↓				
ADHD	Ref		↑↑	↑↑↑		↓	↑	↓↓↓
Panic disorder	Ref	↑↑↑	↑↑↑	↓↓	↑↑↑			↓↓↓
OCD	Ref		↑	↑↑				↑↑↑
AD	Ref	↓↓						↑↑
Epilepsy	Ref	↓↓↓						↓↓
PD	Ref	↓						↑
MDD	Ref	↑	↑	↑	↑↑↑	↑↑	↑	
SZgw	Ref			↓↓↓				↓

Revised Fig. S3b Comparison of risk gene expression enrichment across major brain regions. (b) Summary of brain structures showing significant differences in risk gene enrichment relative to the frontal cortex for each trait.

c. Underreported Signals: Figure 3 visually suggests a stronger enrichment signal for ADHD risk genes in the cerebellum compared to the thalamus and striatum. Similarly, MDD risk genes show a strong signal in the cerebellum, potentially stronger than the amygdala and striatum mentioned in the text. Please address these strong, yet unmentioned, signals in the text.

Response:

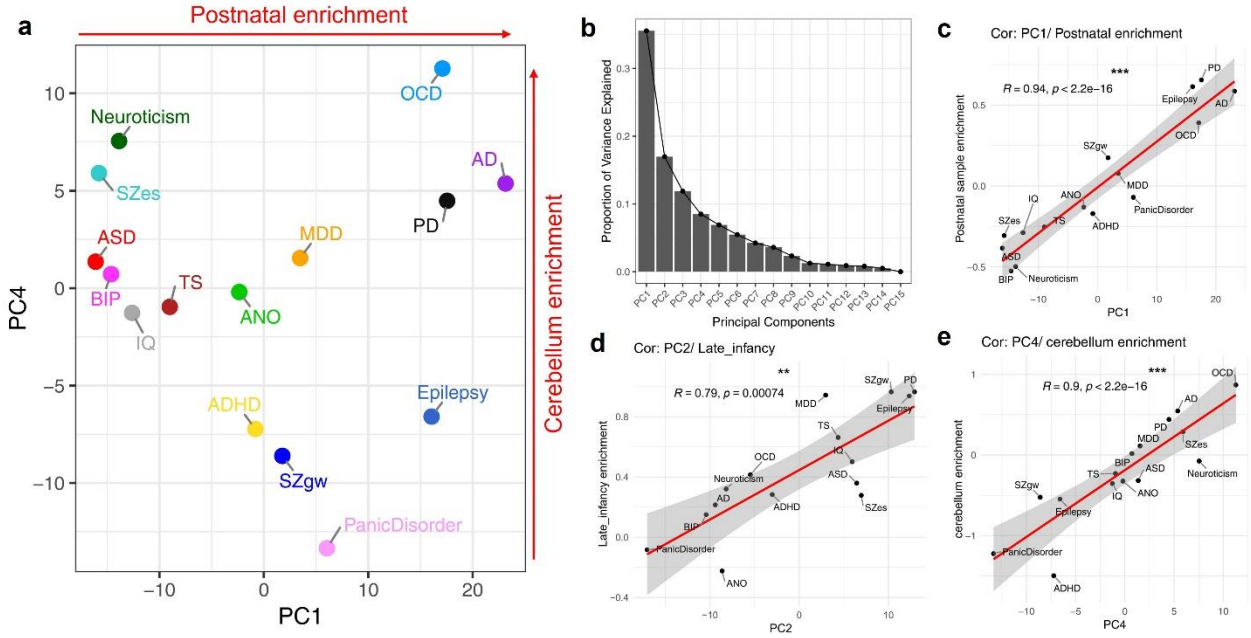
We thank the reviewer for this careful observation. We agree that visual inspection of Fig. 3 may give the impression of relatively strong cerebellar enrichment signals for ADHD and MDD. To formally evaluate these apparent signals, we performed quantitative, region-wise statistical comparisons of risk gene expression enrichment across major brain structures (Fig. S3).

These analyses indicate that, despite visually elevated values in some samples, cerebellar enrichment for ADHD and MDD does not reach statistical significance after FDR correction when compared with the frontal cortex. In contrast, we observed modest but statistically significant cerebellar enrichment for OCD, AD, and PD (Fig. S3).
In addition, principal component analyses further support a selective association between cerebellar enrichment and a subset of traits, as reflected by the strong correlation between PC4 and cerebellar enrichment that survives FDR correction (Fig. 4e), rather than a general cerebellar involvement across all disorders.
We have revised the Results section to explicitly clarify this distinction between visually apparent signals and statistically supported enrichment, and to describe cerebellar enrichment specifically for OCD, AD, and PD.

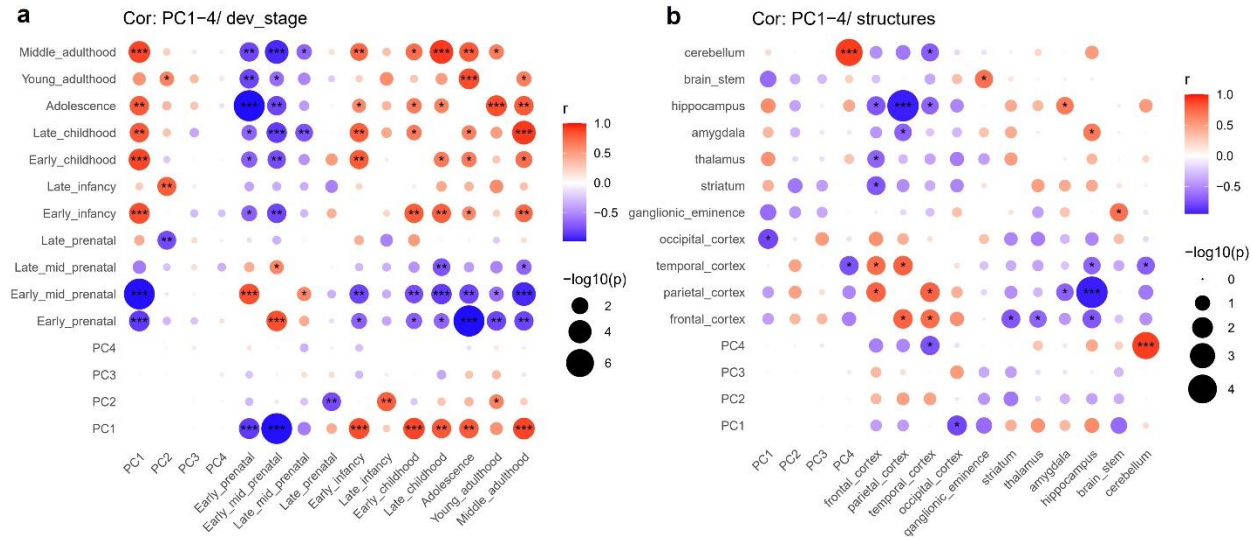
3. Results: "Spatiotemporal expression patterns of risk gene sets differentiate neuropsychiatric disorder subgroups" section:

   a. Missing Traits in Classification: The text describes a t-SNE classification resulting in three categories, yet states that some diseases, such as ADHD, were not included in the initial t-SNE analysis. However, a later result mentions ADHD as being part of the "Prenatal subcortical" cluster. Please clarify this discrepancy and ensure all diseases discussed in the clustering analysis were properly included and accounted for.

Response:
We thank the reviewer for raising this point. In the revised manuscript, we have substantially restructured this section and replaced the t-SNE–based analysis with a principal component analysis (PCA) framework. PCA was performed on the spatiotemporal enrichment profiles of all 15 neuropsychiatric traits, allowing the extraction of interpretable temporal and spatial dimensions that systematically capture similarities and differences across disorders (Revised Fig. 4).
The PCA results revealed that the temporal components PC1 and PC2 together effectively segregate the 15 traits into prenatal and postnatal clusters. PC1 primarily reflects a broad prenatal- postnatal contrast, whereas PC2 captures rapid perinatal transitions in gene expression, characterized by a shift from low expression in late prenatal stages to elevated expression during late infancy. Disorders such as epilepsy, PD, MDD, and SZgw showed prominent loadings along this perinatal trajectory, consistent with their temporal enrichment patterns observed in Fig. 2a-b and Fig. S4a.
In addition, the spatial component PC4 was strongly associated with cerebellar enrichment. Using the frontal cortex as a reference, risk genes for OCD, AD, and PD showed relatively higher enrichment in the cerebellum, whereas panic disorder and epilepsy exhibited comparatively lower cerebellar enrichment. These PCA-derived dimensions provide a unified and biologically interpretable framework for stratifying neuropsychiatric disorders based on shared spatiotemporal risk gene expression patterns.



Revised Fig. 4 Principal Component Analysis of 15 neuropsychiatric traits based on the spatiotemporal enrichment patterns of risk gene sets.



Revised Fig. S4 Correlation matrix between the PCA components and individual spatiotemporal enrichment variables across 15 neuropsychiatric traits.

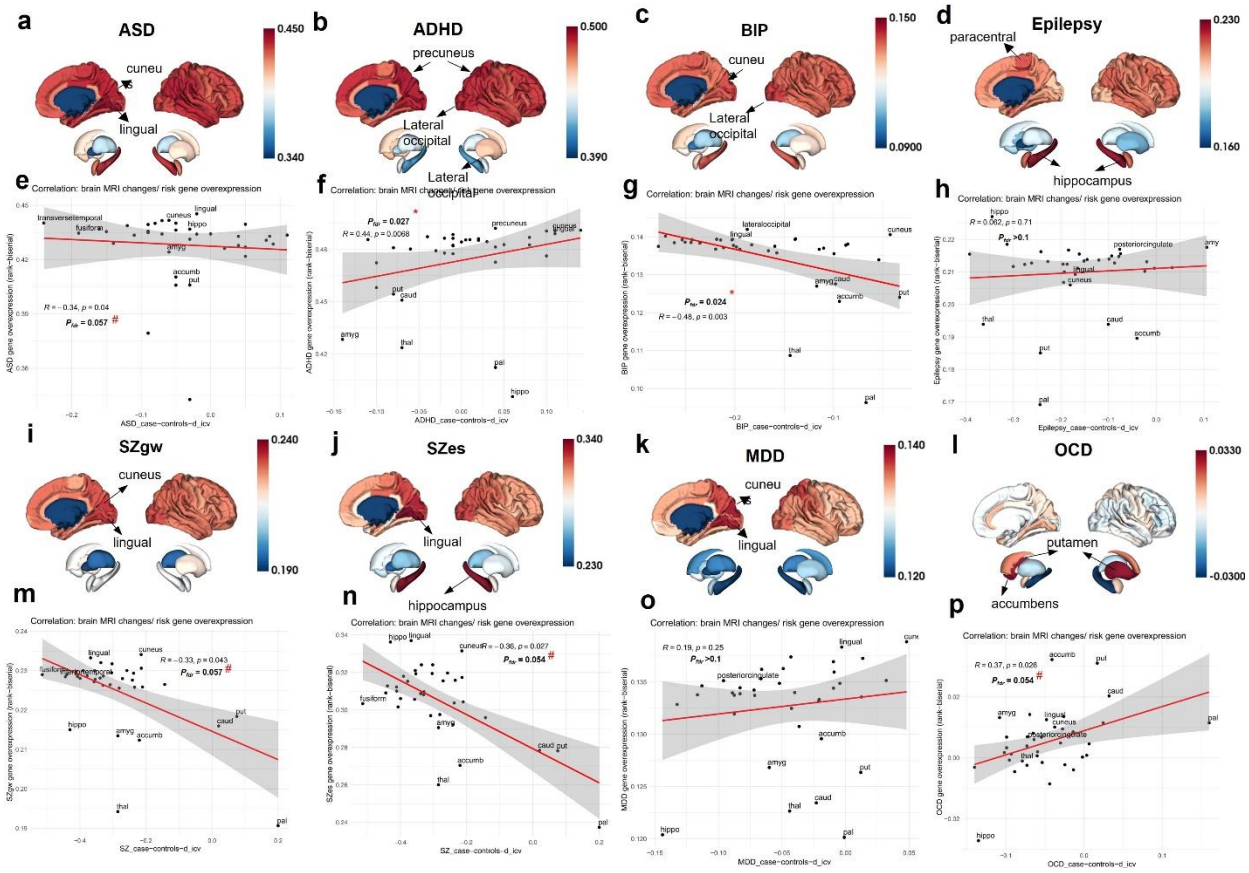
4. Results: "Assessing the biological relevance of risk gene expression patterns using ENIGMA brain morphometry" section:

a. Multiple Testing Correction: The analysis correlating risk gene expression patterns with ENIGMA brain morphometry changes (Fig 5) needs a clear statement on multiple testing correction. Given that several correlation p-values in Figure 5 are very close to the p=0.05 threshold, it is critical to confirm if and how multiple testing correction (e.g., Bonferroni or FDR) was applied to control the false discovery rate for this analysis.

Response:

We thank the reviewer for highlighting the importance of multiple testing correction in this analysis. In the revised manuscript, we have clarified that all correlations between regional risk gene enrichment and ENIGMA-derived structural alterations were assessed using Spearman correlation and corrected for multiple comparisons using the false discovery rate (FDR) method across disorders.

Specifically, associations surviving FDR correction ($P_{fdr} < 0.05$) are now explicitly marked with asterisks in Fig. 5, and P_{fdr} values are reported in the figures and figure legends. Several associations that were close to the nominal significance threshold ($P \approx 0.05$) in the original version are now appropriately described as trends rather than statistically significant findings (marked with hashes in Fig. 5). We have revised the Results text accordingly to distinguish FDR-significant effects from marginal associations and to avoid overinterpretation (Lines 244-245, Lines 248-250).



Revised Fig.5 Spatial enrichment of risk gene sets in ENIGMA brain regions and their association with MRI-derived structural alterations in major psychiatric disorders.

5. Results: "WGCNA-derived co-expression modules and their functional implications in neuropsychiatric disorders" section:

a. Reference Category Justification: The choice of "thalamus during young adulthood was set as the reference category" for this analysis (Line 234) must be rigorously justified. Please cite appropriate literature or provide a clear, biologically or statistically sound rationale for selecting this specific reference category.

Response:

We thank the reviewer for raising this point. In the linear regression framework using dummy variables, the choice of reference category is statistically arbitrary and does not affect the estimated differences between categories or the overall inference of spatiotemporal expression patterns. Any spatial region or developmental stage could, in principle, be selected as the reference without altering the relative relationships among coefficients.

We selected the thalamus during young adulthood as the reference category for two practical reasons. First, this combination represents a well-sampled and developmentally mature baseline in the BrainSpan dataset, thereby providing a stable intercept for model estimation. Second, using a mid-range developmental stage and a broadly connected subcortical structure facilitates intuitive visualization and interpretation of relative expression changes across both brain regions and developmental stages. Importantly, the identification of spatiotemporal expression "hot-spots" depends on the relative differences among regression coefficients and is invariant to the choice of reference level. We have clarified this rationale in the revised Methods section (Lines 293-295, and lines 611-617).

b. Undiscussed Modules: Figure S5 indicates enrichment signals for some modules in the thalamus that were not discussed. The authors should address the relevance of these additional modules in the Results and Discussion sections.

Response:

We appreciate the reviewer's observation regarding additional modules showing enrichment in the thalamus in Fig. S5. We note that although certain modules (e.g., plump1) exhibited pronounced thalamic enrichment, they did not show significant enrichment for risk-associated genes across the 15 neuropsychiatric traits examined. Given the primary focus of this study on disease-relevant molecular networks, these modules were not subjected to further functional analyses in the original manuscript.

In the revised version, we now explicitly acknowledge the presence of these thalamus-enriched but disease-unassociated modules and discuss their potential interpretation as

molecular networks underlying region-specific physiological or homeostatic functions rather than neuropsychiatric disease mechanisms (Lines 289-293).

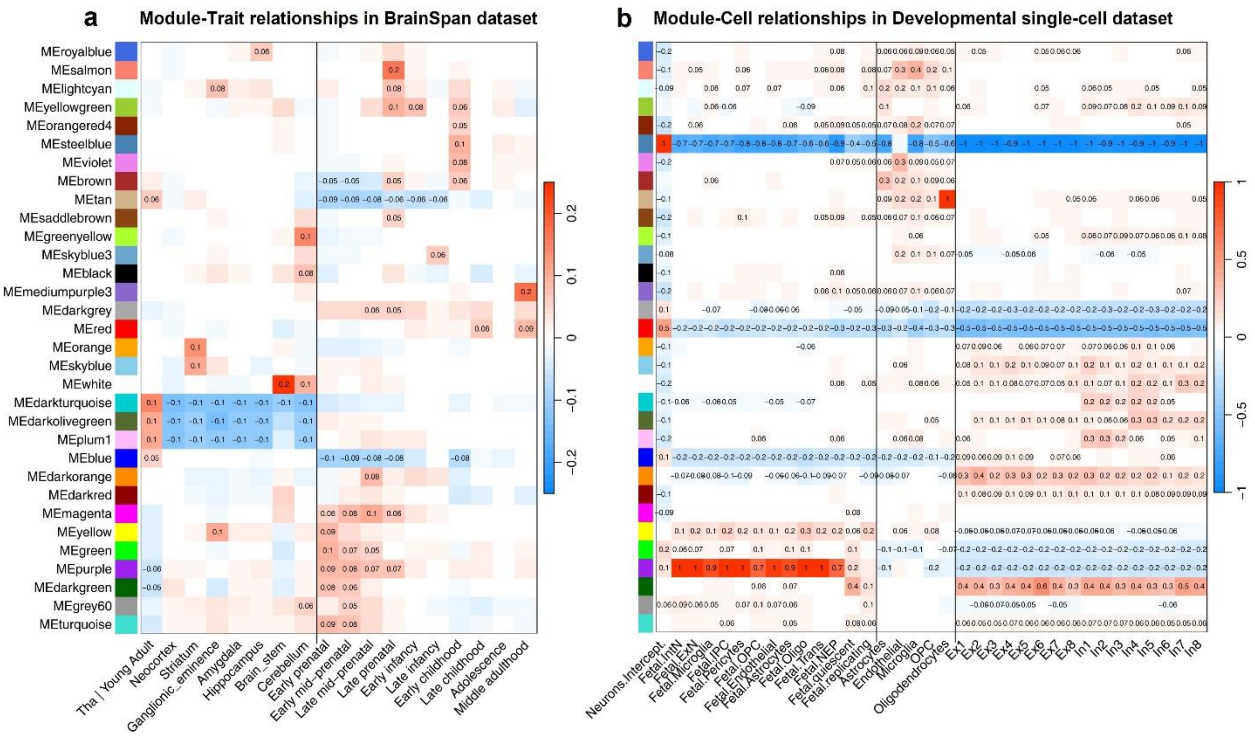
6. Results (Lines 234-237) and Figure S6 Discrepancy: There appears to be a direct contradiction between the text and Figure S6.

The text states "modules overrepresented during early developmental stages were enriched in fetal cell types... In contrast, modules predominant in the perinatal and early childhood stages showed enrichment in glial cells." However, if I understand it correctly, Figure S6 visually suggests the opposite for the latter part: modules overrepresented in early childhood are enriched in glial cells, while modules predominant in the perinatal and early childhood stages (ME magenta, ME yellow, ME green, ME purple, ME darkgreen, and ME turquoise) appear to show enrichment in fetal cell types. Please re-verify the text against the figure and correct the description to accurately reflect the results shown in Figure S6.

Response:

We thank the reviewer for this careful observation. After thoroughly re-examining the text in relation to the original Fig. S6 (now revised as Fig. S11), we confirm that the underlying results are consistent; however, we agree that the original wording could be misinterpreted due to the use of the broader term “perinatal”. To improve clarity and more accurately reflect the spatiotemporal enrichment patterns shown in revised Fig. S11, we have revised the text by replacing “perinatal” with the more precise term “late prenatal and early childhood periods” (Lines 331-332, and lines 336-337). This modification better captures the developmental window during which the corresponding modules exhibit peak overrepresentation and glial cell-type enrichment.

These revisions do not alter the original results or conclusions but enhance the precision and interpretability of the presentation. We appreciate the reviewer’s comment, which helped improve the clarity of the manuscript.



Revised Fig. S11 Parallel comparison of module–spatiotemporal and module–cell-type associations.

(b) Cell-type–specific enrichment of the same modules. Linear regression models were used to assess the influence of major brain cell types (encoded as dummy variables) on ME expression levels. Adult neurons were used as the reference category and absorbed into the model intercept, representing baseline module expression. Predicted ME expression levels for other cell types were obtained by adding the corresponding regression coefficients (β) to the intercept. Only the 32 modules that showed the most significant spatial or temporal enrichment in panel (a) were retained for visualization in both panels. Heatmaps display only module–cell-type combinations with an absolute regression coefficient $|\beta| \geq 0.05$ and a false discovery rate (FDR)–corrected P value < 0.05. Red indicates relative enrichment of module gene expression, whereas blue indicates relative depletion. Modules on the y-axis are ordered according to eigengene similarity distances derived from WGCNA hierarchical clustering.

7. Results: Glial vs. Neuronal Enrichment: The authors classify a group as "Glia-enriched" (e.g., Line 237), yet this group contains "Neurons" cells. Could the observed categorization or downstream results be dominated by the neuronal cell data? To confirm the specificity of the glia enrichment, I suggest performing a sensitivity analysis where the neuronal cell data is removed or down-weighted and the categorization/enrichment analysis is repeated to assess its robustness.

Response:

We thank the reviewer for this insightful comment. To address the concern that the classification of “glia-enriched” modules might be influenced by the presence of neuronal cell types, we performed an explicit sensitivity analysis. Specifically, neuronal cell types were removed from the glial category and treated as a separate reference group in a linear regression model to compare module expression levels across distinct cell-type categories, including fetal cell types, adult glial cells, and adult neurons (see revised Fig. S11).
The results of this sensitivity analysis consistently supported the original results, revealing distinct enrichment patterns of co-expression modules preferentially enriched in fetal cell types, adult glial cells, and adult neurons, respectively (defined by the cell type exhibiting the largest regression coefficient among all cell types). These findings confirm the robustness and cell-type specificity of the enrichment patterns reported in the manuscript (Lines 298-306).

8. Results (Line 252 - "significantly overrepresented"): When stating that "M1 and M4 were significantly overrepresented in endothelial cells," the authors must provide statistical evidence (e.g., p-value, corrected p-value, or other relevant metric) to support the use of the term "significantly overrepresented."

   Response:

Thank you for this important comment. We agree that the use of the term “significantly overrepresented” requires explicit statistical support. In the revised manuscript, we updated the analysis in Fig. 6b by replacing the original M-W U-test with a linear regression framework (revised as Fig. S11), in which the mean expression level of genes in each ME was regressed on cell-types derived from single-cell transcriptomic data. This approach allowed us to directly estimate the association strength between each module and specific cell types, thereby improving statistical power and interpretability.
For each cell type, regression coefficients (B values) and corresponding p-values were obtained, followed by FDR correction performed separately within each cell type across modules. In the revised Fig. S11, only ME - cell type pairs satisfying $|B| \geq 0.05$ and FDR-adjusted $p < 0.05$ are displayed. Therefore, all module - cell type associations shown in the heatmap are statistically significant by definition.
Accordingly, the revised Fig. S11 legend has been updated to clarify the regression-based analysis and multiple-testing correction procedure, and the Results text has been revised to explicitly state the criteria underlying the term “significantly overrepresented” and “cell-type enriched” (Lines 298-306).

9. Methodology: Data Integration and Harmonization: The Methods section indicates the use of multiple datasets and mentions analyses "across samples," suggesting data integration. Please provide a detailed explanation in the Methods section regarding how different datasets were combined (if they were) and, crucially, the specific

harmonization techniques (e.g., batch effect correction) employed to ensure comparability and minimize bias between datasets.

Response:

Thank you for raising this important point. We apologize for any ambiguity in the original wording. We would like to clarify that no direct integration or merging of different transcriptomic datasets was performed at the raw or normalized expression level.

Instead, each transcriptomic dataset (BrainSpan, AHBA, and PsychENCODE single-cell datasets) was analyzed independently using a unified analytical framework, specifically gene set expression enrichment analysis and linear regression. The resulting enrichment metrics (e.g., rank-biserial correlation coefficients or regression coefficients) were then compared qualitatively across datasets, with the goal of assessing the robustness and biological plausibility of convergent patterns across independent transcriptomic modalities. All statistical inferences reported in this study are therefore dataset-specific, and cross-dataset comparisons should be interpreted as descriptive rather than inferential.

Within each dataset, all preprocessing steps, normalization procedures, and batch effect corrections followed the standardized pipelines provided by the original data generators, as documented in the corresponding technical white papers and primary publications. Given the scope and length constraints of the manuscript, these procedures were summarized rather than reimplemented or re-estimated.

To avoid confusion, we have revised the Methods section to explicitly state that datasets were processed and analyzed separately (Lines 576-579).

10. Methodology: WGCNA Module Size: In the WGCNA section (Line 541), the authors specify a minimum module size of 40. Was a maximum module size also set? Please comment on whether significant differences in module size exist across the identified modules and if this variation could potentially impact or bias the downstream analysis results.

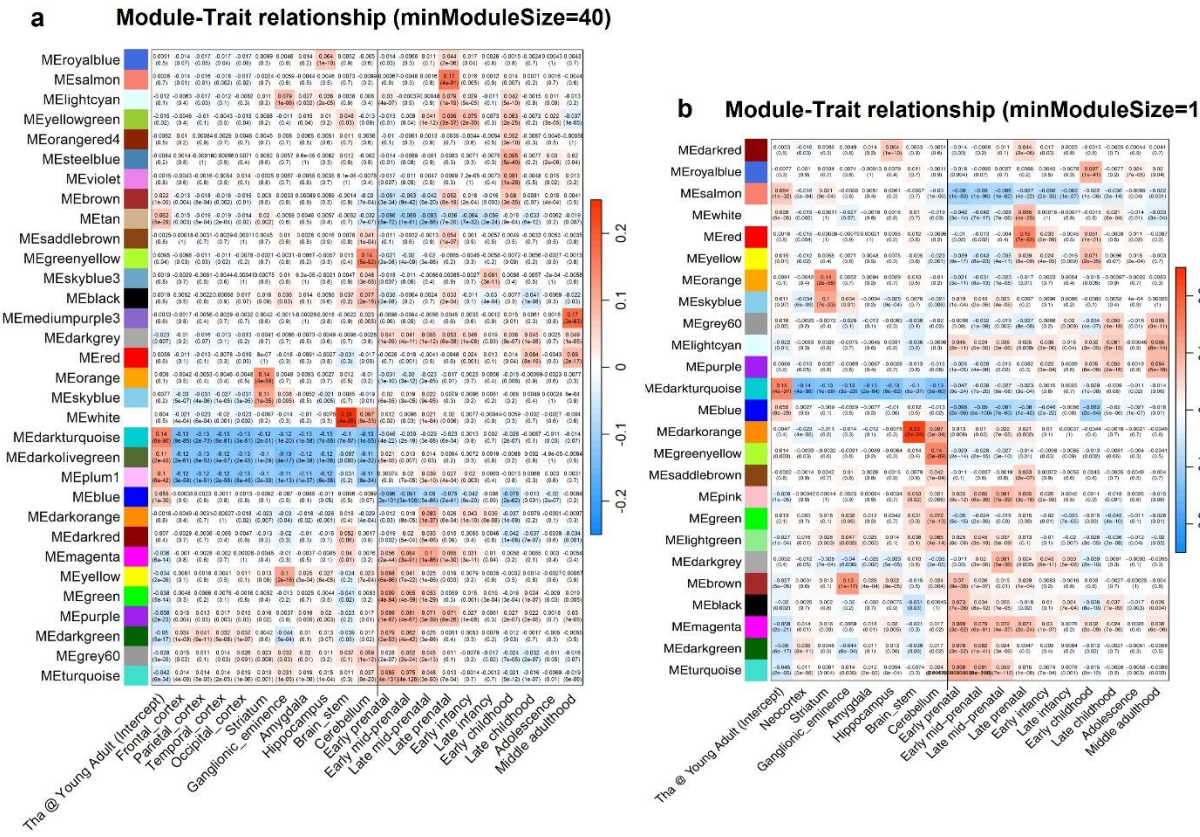
    Response:

We thank the reviewer for raising this methodological point. In our original WGCNA analysis, we specified a minimum module size of 40, but did not set an explicit maximum module size. This strategy follows standard practice in WGCNA analyses, in which only a lower bound is defined to avoid unstable small modules, while larger modules are allowed to emerge naturally based on the intrinsic co-expression structure of the transcriptome (e.g.,Radulescu et al., 2018).

As expected, the resulting modules showed variability in size. To evaluate whether this variation could influence downstream analyses, we performed a sensitivity analysis using a more stringent minimum module size of 120. Notably, the modules identified under this condition showed substantial concordance with those obtained using a minimum size of 40, particularly with respect to their spatiotemporal expression specificity and

developmental patterns (as in the revised Fig. S8). This indicates that the major biologically relevant co-expression structures are robust to reasonable changes in module size parameters.

Importantly, our downstream analyses were conducted primarily at the Module Eigengene level, which captures the dominant expression pattern of each module and reduces dependence on absolute module size. Therefore, although differences in module size exist, they are unlikely to systematically bias the downstream association and enrichment analyses. Overall, we believe that the observed variation in module size reflects biologically meaningful transcriptomic organization rather than a technical artifact, and does not materially affect the robustness or interpretation of our results (Lines 277-288).

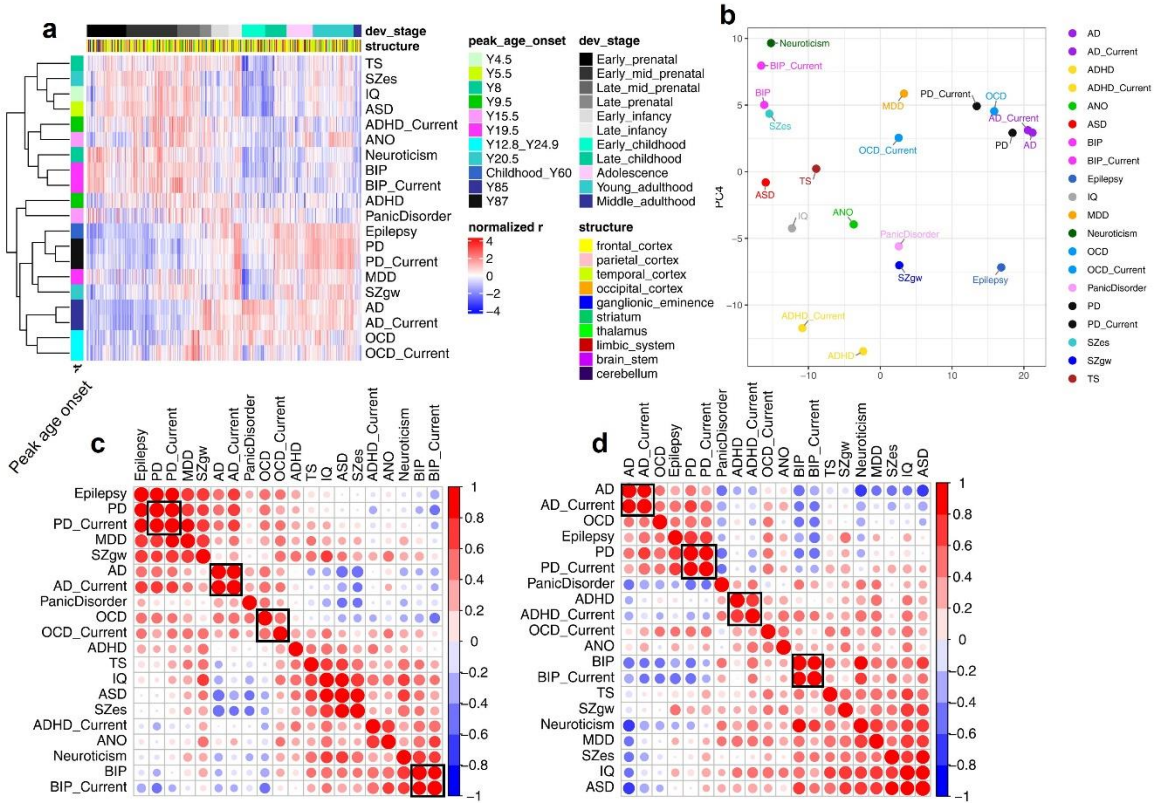


Revised Fig. S8 Module identification under different minimum module size settings.

11. Methodology: Risk Gene Set Currency: The risk gene sets used for the analysis may not be the most current. While I appreciate the extensive effort involved in the current work, it would significantly strengthen the manuscript if the authors could replicate a core analysis for one or two key disorders using the most recently published risk gene sets. This would help determine the sensitivity of the main findings to the choice of input gene lists.

Response:

We appreciate the reviewer's concern regarding the currency of the risk gene sets. To evaluate the sensitivity of our findings to updated genetic inputs, we assembled recently published GWAS-derived risk gene sets for five disorders (BIP, ADHD, OCD, AD, and PD) and incorporated them into our analysis alongside the original 15 traits (Revised Fig. S2). Using the identical enrichment and clustering procedures, we observed that disease risk gene enrichment patterns and grouping results were largely preserved. This indicates that our core conclusions are robust to updates in risk gene annotations and are not dependent on a specific version of the input gene sets (Lines 148-150).



Revised Fig. S2 Validation of the spatiotemporal enrichment and clustering patterns with the most current published GWAS datasets.

Minor Comments

12. Consistency: In the Results section, please ensure consistency in naming brain regions. Use either the full name (e.g., striatum) or the abbreviation (e.g., Str) uniformly throughout the text.

Response:

We have revised the manuscript to ensure consistent naming of brain regions throughout the Results section. All brain structures are now referred to using either their full names or standardized abbreviations defined at first mention (e.g., ganglionic eminence (GE)), and this convention is applied consistently across the text.

13. Disease Name Capitalization: Please standardize the capitalization of disease names (e.g., ADHD, MDD, SCZ) throughout the entire manuscript. Currently, there is an inconsistent mix of all-caps and title-case/lowercase initial letters (e.g., Panic disorder).

Response:

We have standardized the capitalization of disease names throughout the entire manuscript. Abbreviated disease names (e.g., ADHD, MDD, SZ) are consistently presented in all capital letters, while full disease names are written in lowercase (e.g., panic disorder, epilepsy), except when appearing at the beginning of a sentence. In figures and figure legends, disease names used as category labels are capitalized for clarity, and this convention is applied consistently.

14. I look forward to reviewing a revised version of the manuscript that addresses these points. The core findings are potentially important, and addressing these concerns will significantly enhance the clarity, rigor, and impact of the work.

Response:

We sincerely thank the reviewer for this positive and constructive assessment. We have carefully addressed all the concerns raised and believe that the revisions have substantially improved the clarity, rigor, and overall impact of the manuscript.

Reviewer #3 (Remarks to the Author):

This study presents a comprehensive and novel spatiotemporal analysis of genome-wide risk gene sets across 15 major neuropsychiatric disorders. By integrating bulk transcriptomics, single-cell RNA-seq data, GWAS, and ENIGMA data, the authors establish a multidimensional framework linking gene expression dynamics to disease mechanisms and clinical onset timing.

Though these analyses, the study identifies distinct spatiotemporal “hot-spots” of risk gene expression that correspond to critical developmental periods and disease-relevant brain regions. These patterns enable the classification of disorders into three biological meaningful subgroups: prenatal cortical, prenatal subcortical, and postnatal. Moreover, the authors uncover co-expression modules enriched for disease risk genes and categorize them into glia-enriched, adult neuron-enriched, and fetal-enriched groups. Notably, pleiotropic modules such as M14 and M16 implicates early developmental processes, including forebrain synapse organization and chromatin regulation, suggesting share pathogenic mechanisms across multiple disorders.

The study provides comprehensive and valuable analyses, offering important new insights. Although I have a few quibbles and believe the manuscript would benefit from some refinement,

overall it represents a meaningful contribution to the scientific literature and is likely to be well-received.

Minor concerns

1. The overall structure of the abstract and results should be improved, as the key message is too broad and lacks specificity. First, it does not clearly explain the motivation of the study or how it addresses a gap compared to previous research, and the analytical methods are presented in a list-like manner. The results do not specify what the three clusters represent or what the “hot-spots” refer to. A description of the associated cell types and biological functions is also needed.

   Response:

We thank the reviewer for this insightful comment. We have substantially revised the abstract and Results section to clarify the motivation, analytical framework, and biological interpretation of our findings. Specifically, we now explicitly state the knowledge gap addressed by this study—namely, the lack of a systematic spatiotemporal and cell-type–resolved characterization of GWAS risk genes across brain development—and emphasize how our integrative framework extends beyond prior adult or region-restricted analyses.

We further reorganized the Results to follow a coherent analytical progression, moving from spatiotemporal enrichment to cluster definition, anatomical “hot-spots,” and finally cell-type and functional interpretations.

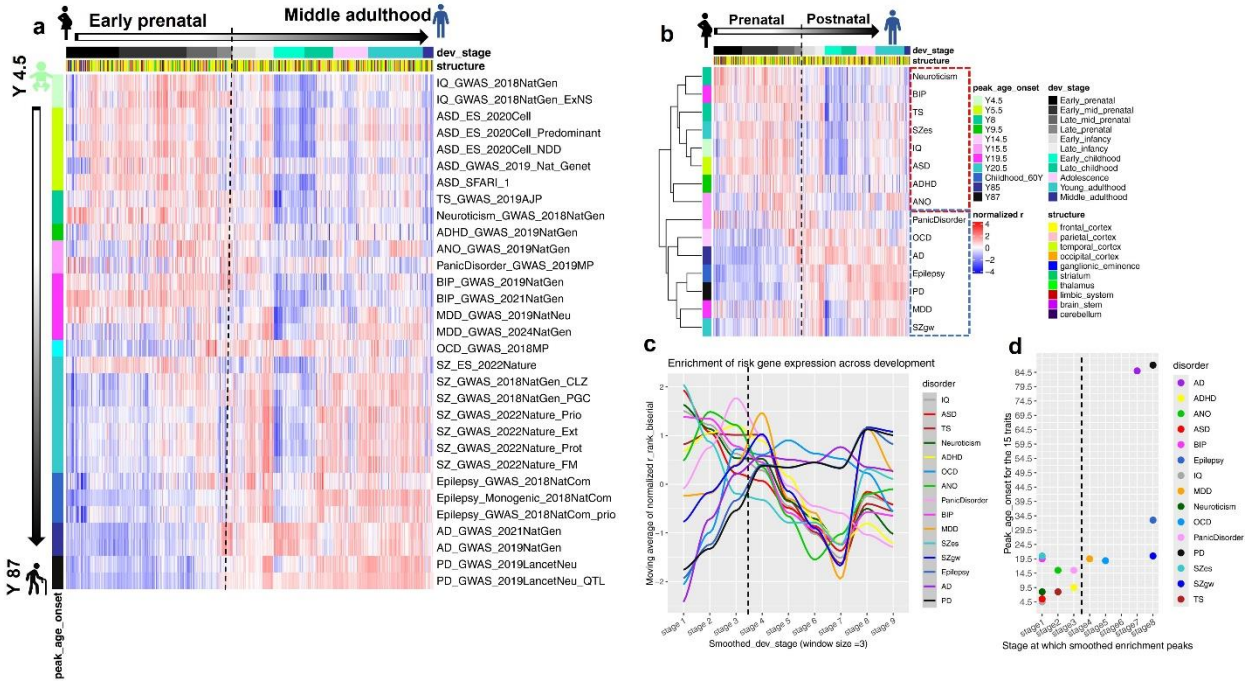
2. Figure 2b: A more detailed explanation is needed regarding which specific early-onset disorders and neurological disorders are referred to in Fig.2B, what their typical ages of onset are, and how these onset ages are related to the results.

Response:

We thank the reviewer for this helpful suggestion. We have revised both the Results section and the Figure 2 legend to explicitly specify which disorders are considered early-onset psychiatric traits and which are neurological disorders, together with their typical peak ages of onset and their relationship to the observed developmental enrichment patterns (Lines 123-131).

Specifically, early-onset psychiatric traits in Fig. 2 include ASD, ADHD, TS, and Neuroticism, all of which typically exhibit peak onset in childhood or adolescence. These traits show prenatal or early postnatal enrichment of risk gene expression, with peak enrichment occurring before or during early childhood. In contrast, neurological disorders such as AD, PD, and epilepsy, which generally have later peak ages of onset, display later developmental enrichment, spanning late prenatal stages through adulthood. To directly link developmental enrichment timing with clinical onset, we added Fig. 2d, which quantitatively demonstrates a positive relationship between the developmental stage of peak risk gene expression enrichment and the peak age of onset across all 15 traits.

Note that although accumulating evidence suggests that SZ associated with rare exonic variants may have an earlier age of onset than SZ driven predominantly by common GWAS variants, a well-defined peak age of onset for this subgroup is currently lacking. Therefore, in the present study, the peak age of onset for SZes was conservatively set to be the same as that for SZgw.



Revised Fig. 2 (composed of the original Fig. 2b and Fig.S1b-c, and an additional scatter plot showing the relationship between the developmental stage at which the moving-averaged r values peak and the peak age of onset for the 15 traits in panel d) Differential developmental dynamics of risk gene expression in relation to peak age of onset.

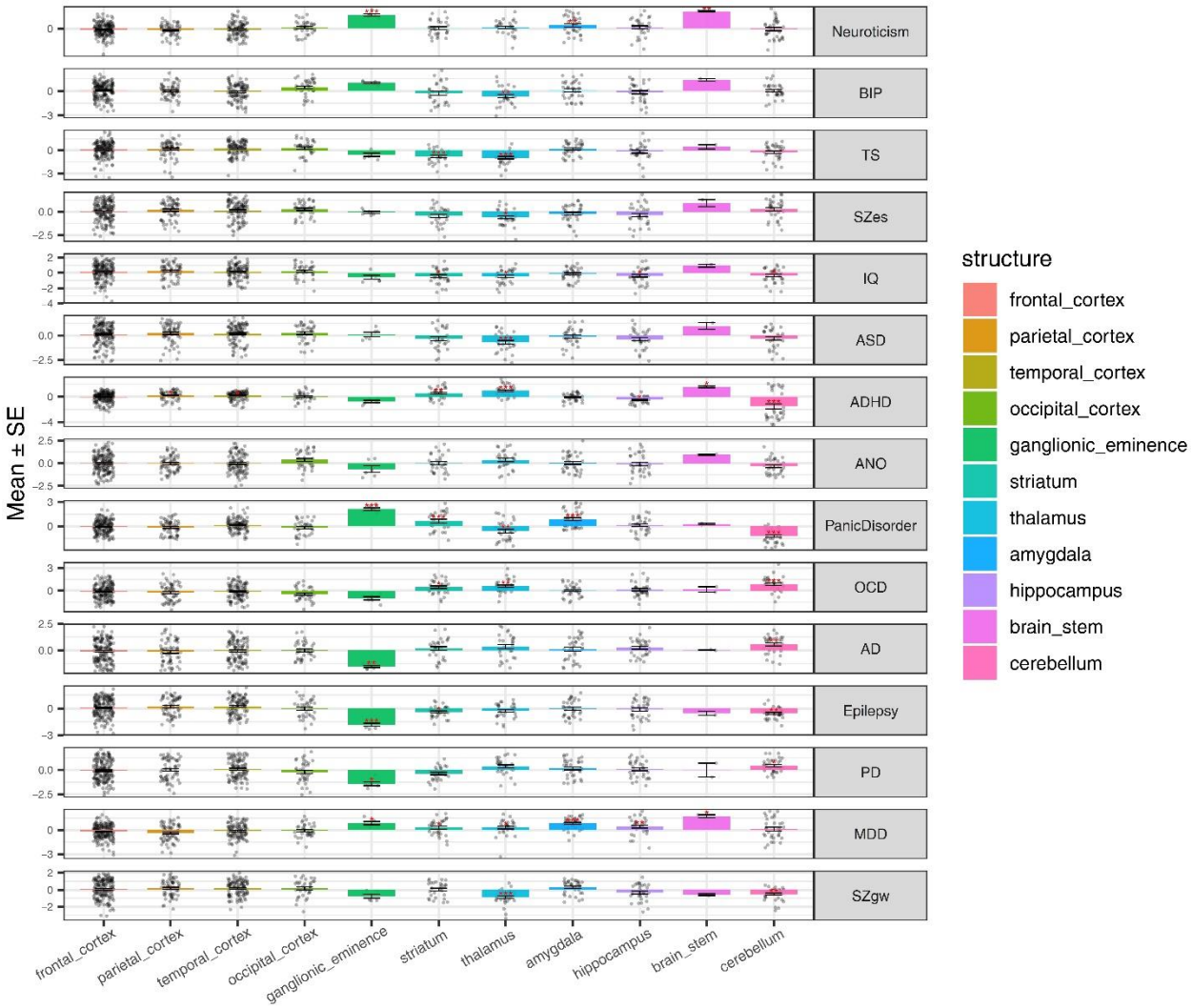
3. General visualization of figures:
   Emphasizing specific disorders and their associated expression values directly within the heatmap may limit the reader's ability to interpret the data objectively (Fig.3). It would be more appropriate to present expression values for the disorders of interest separately, for instance using bar plots or dot plots.

Response:

We appreciate the reviewer's suggestion regarding figure visualization. In response, we have generated dot/point plots (Fig. S3) to illustrate the enrichment of risk gene expression across major brain regions for all 15 traits, rather than emphasizing specific disorders directly within a heatmap. In these plots, individual sample points are displayed, and the frontal cortex was used as the reference structure due to its largest sample size in the BrainSpan dataset and its well-established relevance to cognition and psychiatric disorders.

For each trait, enrichment levels in other brain structures were quantitatively compared with the frontal cortex using t-tests, and significance was indicated with FDR-corrected P values. This approach allows readers to objectively assess the spatial specificity of gene expression across all traits. As shown in Fig. S3, cerebral cortical subregions generally exhibit comparable enrichment across traits, whereas subcortical structures show greater heterogeneity: the ganglionic eminence is selectively enriched for neuroticism, panic disorder, and MDD; the amygdala shows enrichment for mood-related traits; and the cerebellum demonstrates modest enrichment for OCD, AD, and PD. The limited brainstem samples also suggest consistent enrichment for neuroticism, ADHD, and MDD.

We believe that these dot plots address the reviewer's concern by providing a more objective and quantitative visualization of trait-specific risk gene enrichment across brain regions.



Revised Fig. S3a Comparison of risk gene expression enrichment across major brain regions. (a) Bar plots with individual sample points illustrate the relative enrichment of risk gene set expression across major brain structures.

Disorder	FC	GE	Str	Th	Amy	Hip	BS	CB
Neuroticism	Ref	↑↑↑			↑↑		↑↑	
BIP	Ref			↓↓				
TS	Ref		↓↓↓	↓↓↓				
SZes	Ref			↓				
IQ	Ref		↓	↓		↓		↓
ASD	Ref			↓↓↓				
ADHD	Ref		↑↑	↑↑↑		↓	↑	↓↓↓
Panic disorder	Ref	↑↑↑	↑↑↑	↓↓	↑↑↑			↓↓↓
OCD	Ref		↑	↑↑				↑↑↑
AD	Ref	↓↓						↑↑
Epilepsy	Ref	↓↓↓						↓↓
PD	Ref	↓						↑
MDD	Ref	↑	↑	↑	↑↑↑	↑↑	↑	
SZgw	Ref			↓↓↓				↓

Revised Fig. S3b Comparison of risk gene expression enrichment across major brain regions. (b) Summary of brain structures showing significant differences in risk gene enrichment relative to the frontal cortex for each trait.

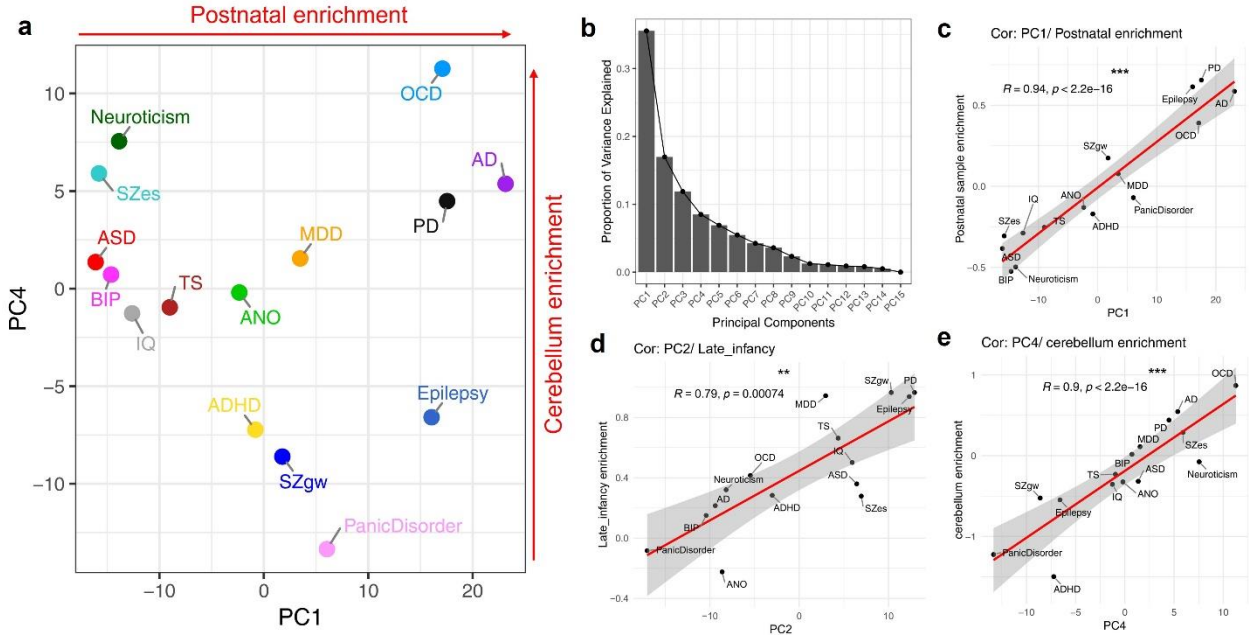
4. As in Fig.4a, please modify the t-SNE visualization so that grouped clusters are indicated based on precise values, such as a 95% confidence interval.

Response:

We thank the reviewer for the suggestion regarding t-SNE visualization. Following feedback from other reviewers, we have replaced the t-SNE plots with PCA-based visualization to classify the 15 traits, as PCA allows for a more quantitative representation of group separation and variance explained.

In the updated figure (Fig. 4a), the relative positions of traits are displayed along the two principal components, providing an objective view of clustering without relying on arbitrary 2D embeddings. The figure legend now describes the variance captured by each component and the spatial relationships among traits, allowing readers to interpret the clustering quantitatively.

We believe this PCA-based approach addresses the reviewer's concern regarding precise quantification of cluster separation while maintaining clarity of visualization.



Revised Fig. 4 Principal Component Analysis of 15 neuropsychiatric traits based on the spatiotemporal enrichment patterns of risk gene sets.

(a) Principal component analysis (PCA) was performed on the Z-scored rank biserial correlation coefficients of the 15 risk gene sets across all BrainSpan samples. The scatter plot shows the distribution of the 15 traits in the space defined by the temporal (PC1) and spatial (PC4) principal components. Positions along PC1 and PC4 reflect similarities in trait-specific expression enrichment, with traits exhibiting comparable spatiotemporal patterns clustering together.

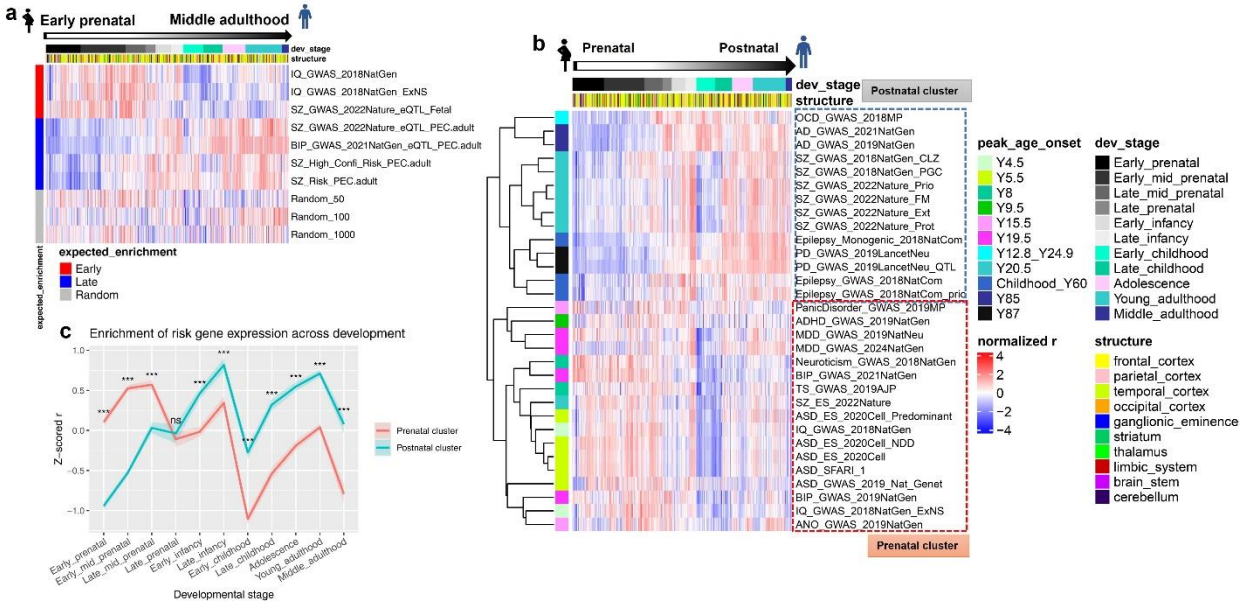
5. Figure S1: It is difficult to conclude that there is a clear difference between the two groups on the heatmap in panel (a) because certain stage doesn't show a distinct separation. For example, the young adulthood stage does not exhibit a clear distinction. Statistical comparisons between the early-expression cluster and the late-expression cluster would be necessary.

Response:

We thank the reviewer for pointing out that visual inspection of the heatmap in the initial Fig. S1a (now in Fig. S1b) alone is insufficient to support a clear separation

between early- and late-expression clusters at all developmental stages, particularly during young adulthood. We fully agree that formal statistical comparisons are necessary. To address this concern, we added panel (c) in the revised Fig. S1, which provides a stage-wise quantitative comparison between the prenatally and postnatally enriched disorder clusters. As shown in Fig. S1c, the two clusters exhibit distinct and statistically significant differences in expression enrichment across development, assessed using Wilcoxon rank-sum tests with FDR correction.

Specifically, the prenatally enriched cluster shows significantly higher expression enrichment during early and mid-prenatal stages, whereas the postnatally enriched cluster becomes significantly more enriched after birth, including childhood and adulthood stages. Notably, the enrichment trajectories of the two clusters intersect around late prenatal stage, marking a transitional period during which neither group shows a dominant enrichment advantage. These results demonstrate that the apparent overlap observed in late prenatal reflects a biologically meaningful developmental transition, and that the overall distinction between prenatally- and postnatally-enrichment clusters is statistically supported when assessed longitudinally.



Revised Fig.S1 Validation and developmental stratification of spatiotemporal expression enrichment patterns of neuropsychiatric risk gene sets

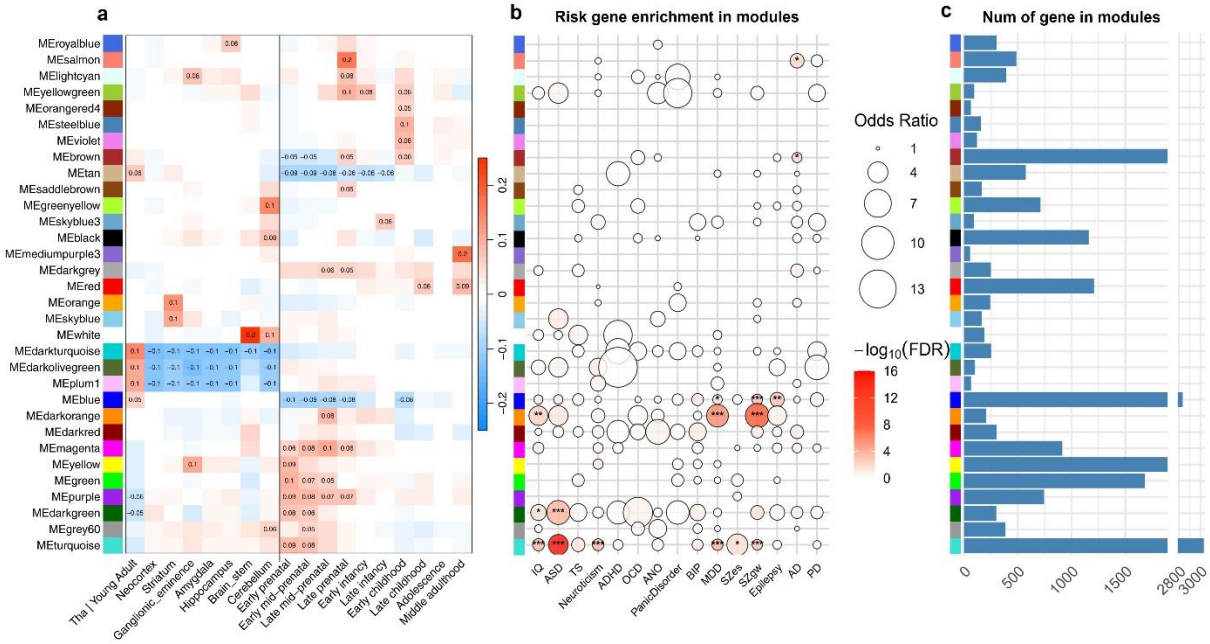
(c) Line plots comparing the developmental dynamics of risk gene expression enrichment between the prenatally and postnatally overrepresented disorder clusters. For each developmental stage, lines represent the mean Z-scored r value across datasets within each cluster. Shaded ribbons indicate the 95% confidence interval of the mean (mean ± 1.96 × SE). The two trajectories intersect during late prenatal development, marking a transitional phase between prenatal and postnatal dominance. Statistical differences between the two clusters at each developmental stage were assessed using Wilcoxon rank-

sum tests, with false discovery rate (FDR)-corrected P values indicated above the curves (FDR-adjusted P < 0.05, 0.01, and 0.001 denoted by *, **, and ***, respectively).

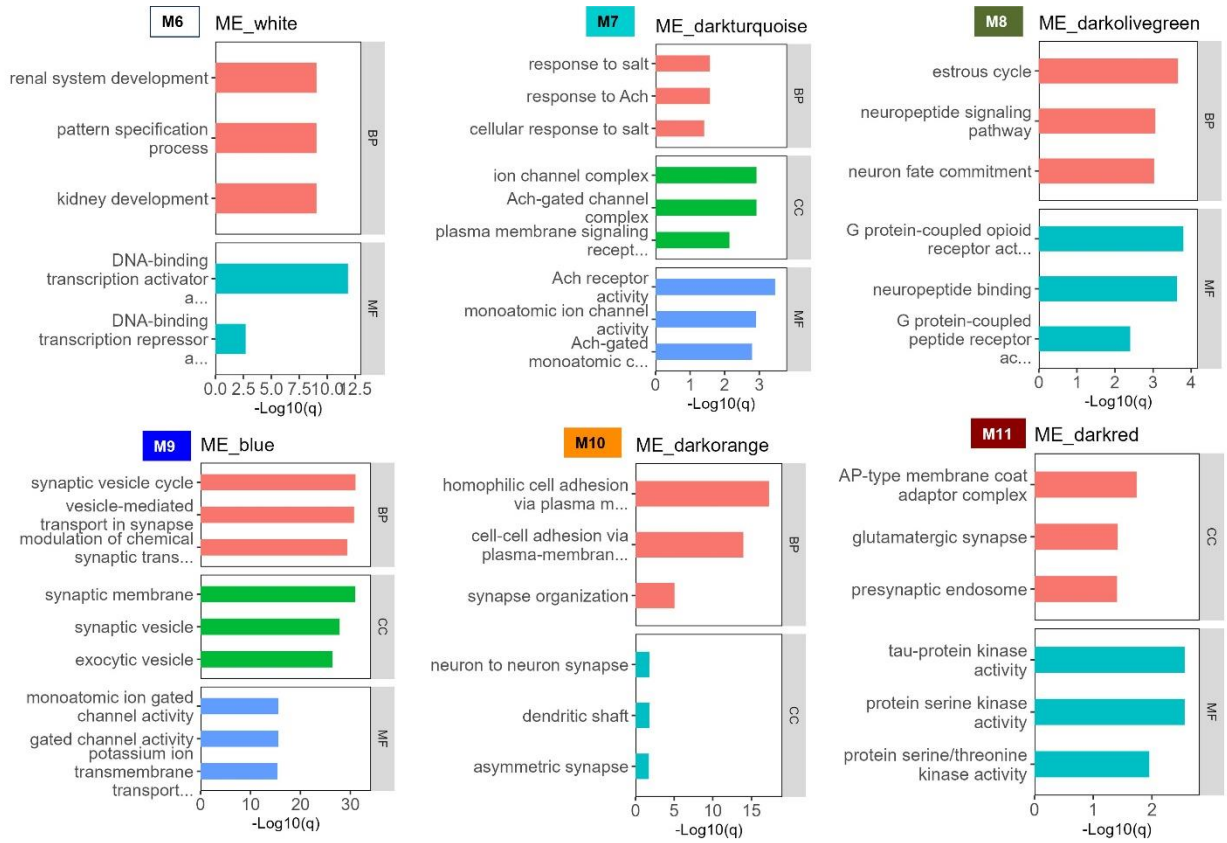
6. In addition, late infancy shows high expression across the entire disease spectrum. Therefore, a discussion focusing on late infancy, which appears to be most strongly associated with the disease, is also needed to be explained.

Response:

We appreciate the reviewer's suggestion to discuss the high risk gene expression observed during late infancy. Indeed, as shown in Fig. 2b and Fig. S1b, late infancy exhibits elevated expression of risk genes across both prenatal and postnatal disorders. This pattern is likely related to the elevated expression of genes in the ME blue, ME darkorange, and ME skyblue3 modules during this stage (Fig. S7). These modules are enriched for risk genes from multiple prenatal and postnatal disorders, indicating that late infancy serves as a transitional stage from infancy to childhood, during which many disease-associated genes and functional pathways may show overall high expression across multiple cell types. Notably, these include genes involved in synaptic structure and functional regulation (Fig. S13), providing a mechanistic explanation for the overall enrichment of multiple disease risk gene sets at this stage. Therefore, late infancy appears to be a critical developmental window in which diverse neuropsychiatric risk genes are simultaneously active, potentially contributing to vulnerability across multiple disorders. We have discussed this point in the revised manuscript (Lines 417-428).



Revised Fig. S7 (corresponding to previous Fig. S5) Module eigengene (min size=40) enrichment in spatiotemporal levels and risk gene sets.



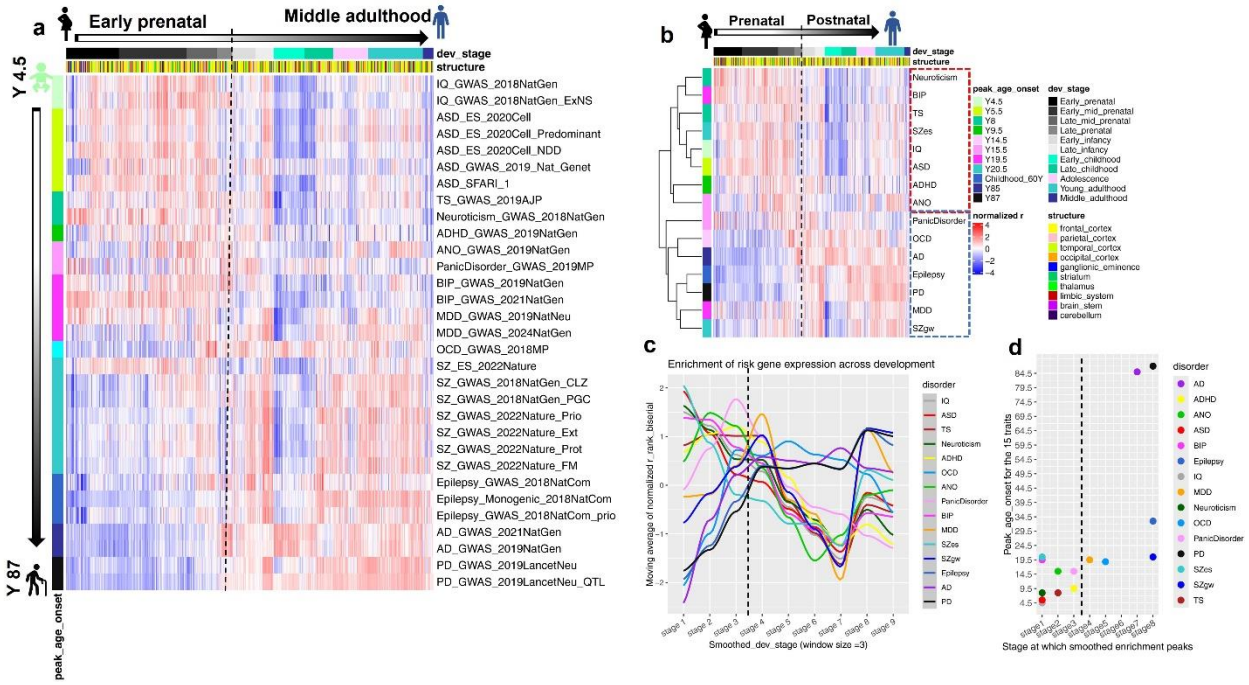
Revised Fig. S13 Functional annotation of M6-M11 co-expression module genes. Gene Ontology (GO) enrichment analysis of co-expressed genes in M6-M11 highlighted distinct Biological Processes (BP), Cellular Components (CC), and Molecular Functions (MF) implicated in neuropsychiatric disorders. For each module, the top three significantly enriched GO terms in each category are shown (q value < 0.05).

7. It is necessary to explain in the figure legend for panel (c) what each stage represents in terms of developmental phases. In addition, since the values for many disorders are presented within a single graph, it is difficult to distinguish the data. It would be preferable to redraw the figure using an alternative visualization, such as a dot plot.

Response:

We thank the reviewer for pointing out the need for clearer explanation of developmental stages and improved visualization. In the revised manuscript, we have expanded the figure legend for Fig. 2c to explicitly describe what each summarized stage represents in terms of developmental phases. Specifically, the nine developmental stages correspond to major prenatal and postnatal periods derived from the original 11 BrainSpan time points by applying a moving-average filter (window size = 3), thereby smoothing temporal fluctuations while preserving biologically meaningful transitions across development.

To improve the readability of the multiple trajectories shown in Fig. 2c, we applied smoothing to the line plots and increased line thickness, thereby highlighting the dynamic trends of risk gene expression enrichment for each trait. This adjustment reduces visual noise caused by closely overlapping curves and allows the overall developmental patterns to be more clearly distinguished across traits.



Revised Fig. 2 Differential developmental dynamics of risk gene expression in relation to peak age of onset.

(c) Smoothed line plots illustrate the temporal expression dynamics of risk gene sets, highlighting divergent developmental trajectories between prenatally and postnatally enriched disorders. For traits enriched prenatally, normalized *r* values peaked during early developmental stages, whereas postnatally enriched disorders showed maximal expression enrichment after birth. Most psychiatric disorders or traits—except OCD, MDD, and SZ_{GW}—demonstrated prenatal enrichment of risk gene expression. Notably, risk genes for MDD and SZ_{GW} exhibited two distinct peaks during postnatal development, suggesting multiple windows of vulnerability. The nine summarized developmental stages were generated by applying a moving-average filter (window size = 3) to the original 11 BrainSpan time points.

8. Figure 4b:

Multiple GWAS datasets for the same disorder may not exhibit the highest r value in the same brain region: instead, each dataset can show its strongest correlation in different regions. For example, in the case of epilepsy, the highest r values are observed in the CB, AMY and MGE regions, whereas for ASD, they are found in the A1C, M1C, and DFC regions. It is therefore important to discuss whether the regions

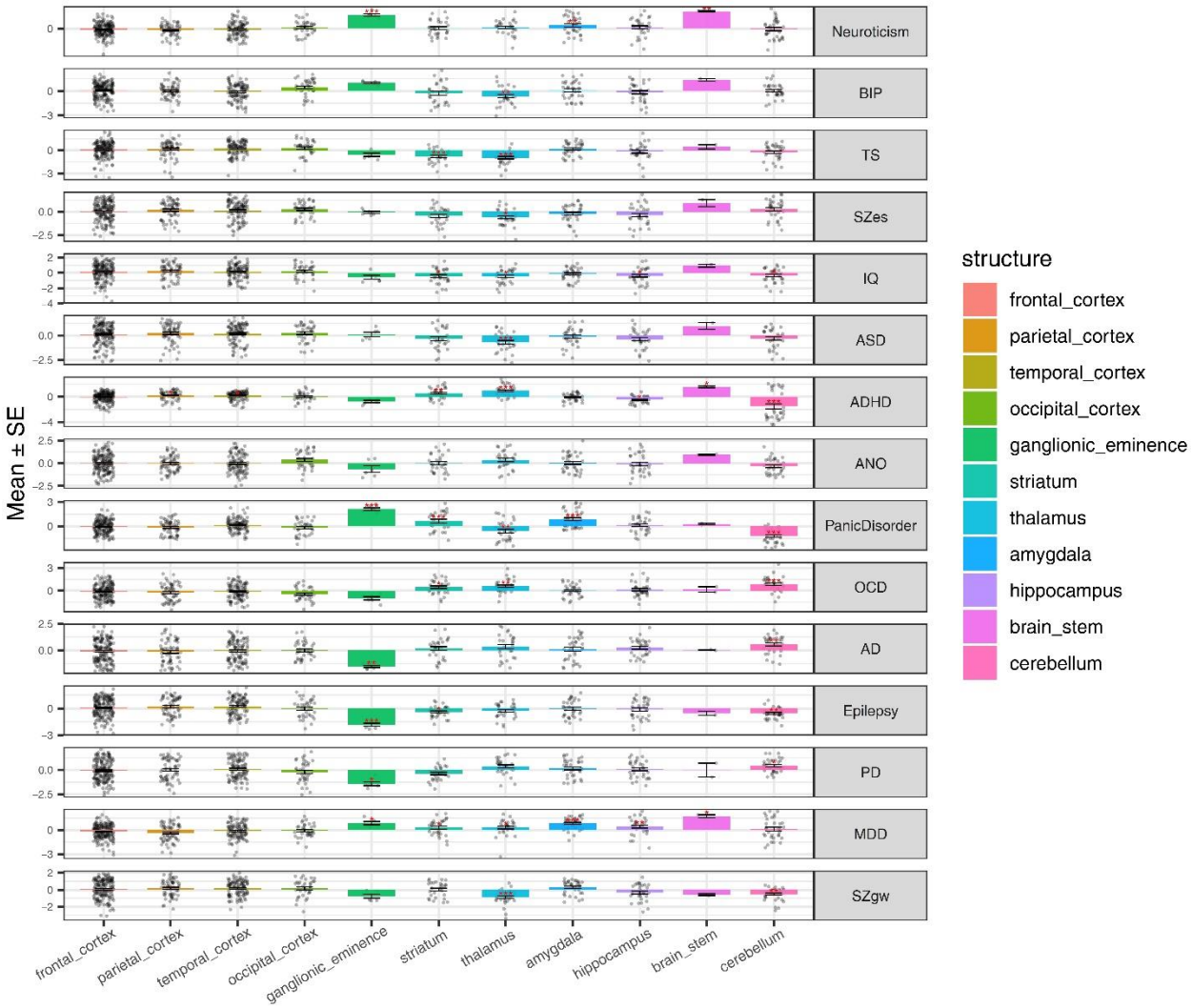
showing high r values share functional relevance related to disease, and what underlying reasons should explain these results.

Response:

We thank the reviewer for this insightful comment regarding the regional heterogeneity observed across multiple GWAS datasets for the same disorder. Following suggestions from other reviewers, we have replaced the original Fig. 4b with a quantitative scatter-plot–based analysis (Fig. S3), which allows a more objective comparison of enrichment patterns across brain regions and traits. Correspondingly, the Results section and figure legends have been revised (Lines 166-176).

Using this updated analysis, we now explicitly describe that risk genes for several disorders show preferential enrichment in functionally relevant brain regions, for example, enrichment of OCD, ADHD, MDD, and neuroticism risk genes in the CB, thalamus, amygdala, and hippocampus, whereas risk genes for ASD, IQ, and SZ exhibit broad enrichment across multiple cortical regions. These observations are consistent with prior neurobiological and neuroimaging evidence implicating these regions in the respective cognitive, emotional, and behavioral phenotypes.

Importantly, we note that the presence of high r values in multiple brain regions for the same disorder likely reflects the polygenic and systems-level nature of neuropsychiatric diseases, rather than inconsistency across GWAS datasets. From a broader neurobiological perspective, complex psychiatric disorders are increasingly understood to arise from dysfunction in large-scale neural networks spanning multiple brain regions, as well as from interactions among distinct cell types. For example, excitation–inhibition microcircuit imbalance has been implicated in schizophrenia, while coordinated contributions of microglia and astrocytes to excitatory neuronal vulnerability are thought to underlie Alzheimer's disease pathology. Such network- and cell-type–level mechanisms provide a conceptual framework for why risk gene enrichment is often distributed across multiple, functionally connected brain regions rather than confined to a single anatomical structure.



Revised Fig. S3a Comparison of risk gene expression enrichment across major brain regions. (a) Bar plots with individual sample points illustrate the relative enrichment of risk gene set expression across major brain structures.

Disorder	FC	GE	Str	Th	Amy	Hip	BS	CB
Neuroticism	Ref	↑↑↑			↑↑		↑↑	
BIP	Ref			↓↓				
TS	Ref		↓↓↓	↓↓↓				
SZes	Ref			↓				
IQ	Ref		↓	↓		↓		↓
ASD	Ref			↓↓↓				
ADHD	Ref		↑↑	↑↑↑		↓	↑	↓↓↓
Panic disorder	Ref	↑↑↑	↑↑↑	↓↓	↑↑↑			↓↓↓
OCD	Ref		↑	↑↑				↑↑↑
AD	Ref	↓↓						↑↑
Epilepsy	Ref	↓↓↓						↓↓
PD	Ref	↓						↑
MDD	Ref	↑	↑	↑	↑↑↑	↑↑	↑	
SZgw	Ref			↓↓↓				↓

Revised Fig. S3b Comparison of risk gene expression enrichment across major brain regions. (b) Summary of brain structures showing significant differences in risk gene enrichment relative to the frontal cortex for each trait.

9. Figure S5: The number of modules identified by the authors is quite large, and some of them exhibit highly similar expression profiles. For example, the darkturquoise, darkolivegreen, and plum1 modules display similar patterns. To further condense the clusters into more biologically meaningful modules, it would be appropriate to cluster the modules based on pairwise eigengene correlations and merge those that share similar expression profiles.

Response:

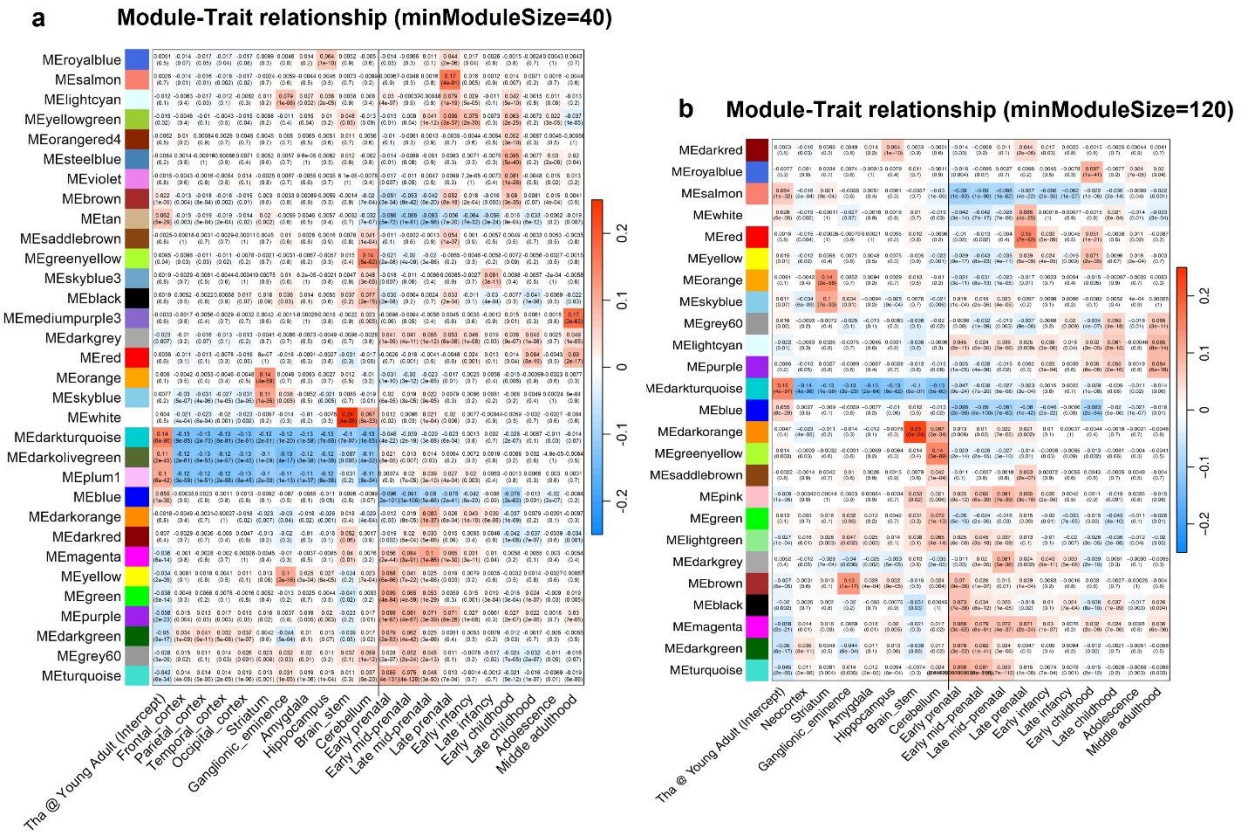
We thank the reviewer for the thoughtful comments regarding module size, module similarity, and potential redundancy in the WGCNA results. As noted above, in our original WGCNA analysis we specified a minimum module size of 40, without imposing an explicit maximum module size. This approach follows standard WGCNA practice, in which only a lower bound is

defined to avoid unstable small modules, while allowing larger modules to emerge naturally from the intrinsic co-expression structure of the transcriptome. As expected, this resulted in modules with a broad size distribution.

To address concerns about module redundancy and to explore whether a more condensed module structure would yield more biologically meaningful results, we performed an additional sensitivity analysis using a more stringent minimum module size of 120 (Fig. S8). Under this condition, most modules identified at the larger size threshold could be readily mapped to corresponding modules from the original analysis, indicating that the major co-expression structures are robust to changes in module size parameters (Fig. S9, Fig. S14). However, we also observed that some module – disorder enrichment signals that were significant under the original setting were no longer significant when the minimum module size was increased (MEturquoise – Neuroticism, MDD, and SZgw). This suggests that while increasing module size may improve statistical power for some analyses, it can simultaneously reduce sensitivity to modules with strong spatiotemporal specificity, which are particularly relevant for developmental and region-specific disease risk.

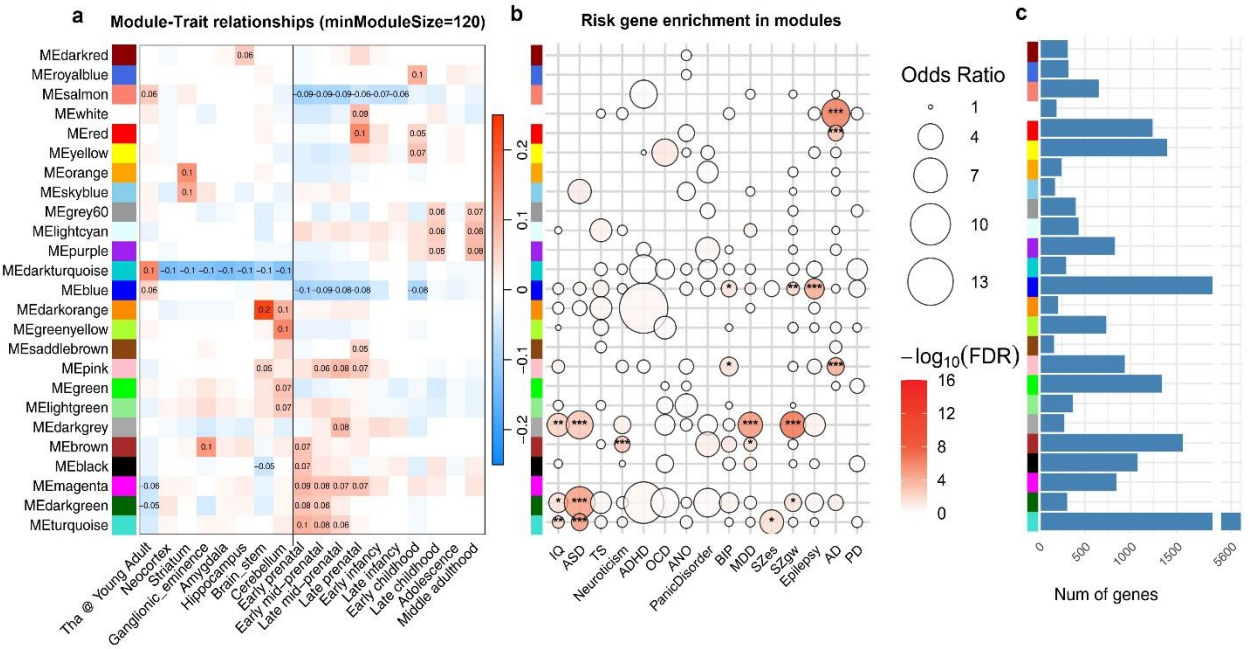
In response to the suggestion to merge modules with similar expression profiles (e.g., darkturquoise, darkolivegreen, and plum1), we carefully re-examined their properties. Although these modules show similar spatial enrichment patterns, they exhibit distinct developmental trajectories and different cell-type enrichment profiles. Merging these modules would therefore increase statistical power at the cost of reduced spatiotemporal and cellular resolution, potentially obscuring biologically meaningful distinctions in risk gene enrichment.

After systematically comparing results across different module size thresholds and evaluating the trade-off between statistical power and biological resolution, we chose to retain the original WGCNA settings for downstream analyses. We believe this approach better preserves the developmental, regional, and cellular specificity of co-expression modules relevant to neuropsychiatric disease risk, while maintaining robustness of the main findings.



Revised Fig. S8 Module identification under different minimum module size settings. (a, b) WGCNA results obtained using different minimum module size thresholds, with a minimum module size of 40 in panel A and 120 in panel B. In both panels, only modules whose module eigengenes (MEs) exhibited at least one regression coefficient (B) ⩾ 0.05 for any spatial or temporal variable level are shown. Modules on the y-axis are ordered according to their eigengene similarity distance derived from WGCNA hierarchical clustering.

Overall, a high degree of similarity was observed between the two analyses. Several modules with spatiotemporally specific expression patterns were consistently identified under both parameter settings, exhibiting highly comparable enrichment profiles across developmental stages and brain regions. These modules likely represent corresponding co-expression module pairs detected under different minimum size constraints. Representative paired modules include turquoise – turquoise, darkgreen – darkgreen, yellow – brown, purple – magenta, darkturquoise – darkturquoise, blue – blue, orange – orange, greenyellow – greenyellow, white – darkorange, royalblue – darkred, and salmon – red (panel a – panel b, respectively).

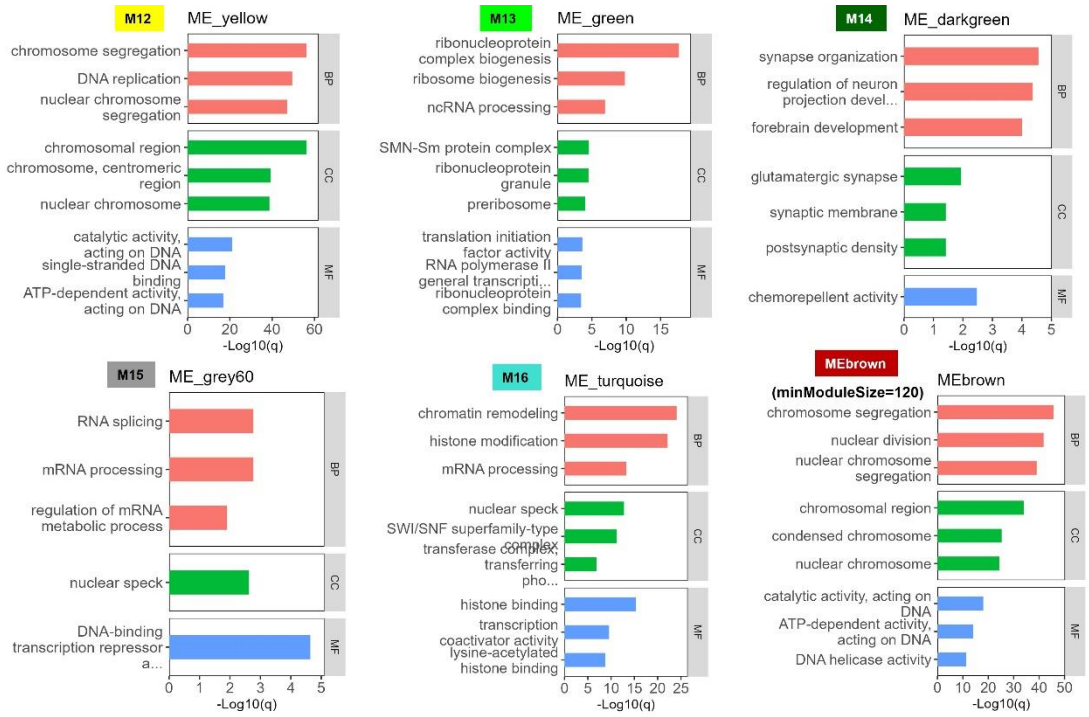


Revised Fig. S9 Module eigengene (min size=120) enrichment in spatiotemporal levels and risk gene sets.

(a) Linear regression models were used to assess the influence of brain region and developmental stage (encoded as dummy variables) on ME expression levels, when the minimum size was set as 120. The thalamus during young adulthood was set as the reference category and absorbed into the model intercept, which thus represents the baseline expression of each module. Predicted ME expression levels for other regions and developmental stages were computed by adding corresponding β coefficients to the intercept, allowing identification of spatiotemporal expression peaks ("hot-spots") for each module. Only 25 modules that showed the most significant spatial or temporal expression enrichment (with at least one $\beta > 0.05$ in the regression model) were retained for visualization in the heatmap. Within these modules, only module–spatiotemporal variable level combinations with an absolute regression coefficient $|B| \geqslant 0.05$ and a false discovery rate (FDR)–corrected P value < 0.05 are displayed, with red colors indicate enrichment of module gene expression and blue indicate deprivation. Modules on the y-axis are ordered according to their eigengene similarity distance derived from WGCNA hierarchical clustering.

(b) Enrichment of neuropsychiatric disorder risk gene sets across the 25 spatiotemporal modules. Odds ratios (ORs) that higher than 1 and the corresponding P values from chi-square tests are displayed in the Trait × Module matrix, with the size of bubbles indicating OR values and color gradient indicating –log10 (FDR) magnitude. P values across all the original 31 WGCNA detected modules for each of the disorders were FDR corrected (P_{fdr}<0.05: *; P_{fdr}<0.01: **; P_{fdr}<0.001: ***).

(c) Horizontal bar plot showing the number of genes in each of the 25 co-expression modules. Due to the substantial variability in module size (ranging from 166 to 5705 genes), the x-axis was truncated between 1800-5500 using a broken scale to improve visualization of both small and large modules.



Revised Fig. S14 Functional annotation of co-expression module M12-M16.
GO enrichment analysis of co-expressed genes in M12-M16 revealed distinct BP, CC, and MF implicated in neuropsychiatric disorders. For each module, the top three significantly enriched GO terms in each category are shown (q value < 0.05); Notably, an additional module identified when the minimum module size was increased to 120 (ME brown) exhibited highly consistent GO enrichment patterns with module M12 derived using a minimum module size of 40, supporting the robustness of the WGCNA in our analysis with different minimum module size.

10. Additionally, after such merging, the authors should clearly address why the ADHD x MEdarkolivegreen combination shows the highest odds ratio (OR) yet a non-significant P value. While this might be due to the small size of the ADHD risk gene set, other possible explanations should also be discussed explicitly.

Response:
We thank the reviewer for requesting clarification regarding the observation that the ADHD × MEdarkolivegreen combination shows a relatively high odds ratio (OR) but does not reach statistical significance.
First, we note that even after re-running WGCNA with a more stringent minimum module size of 120, the newly derived modules likewise did not show FDR-significant enrichment for

ADHD risk genes. This consistency across module size settings suggests that the non-significant result is not driven by module definition or redundancy, but rather reflects intrinsic properties of the ADHD risk gene set. The most direct explanation is the relatively small size of the ADHD risk gene set, which limits statistical power in enrichment analyses and makes it difficult for even moderate-to-large effect sizes (ORs) to survive multiple-testing correction.

Beyond sample size considerations, several additional explanations are plausible and are now explicitly discussed in the revised manuscript. One possibility is that ADHD risk genes do not converge strongly onto a single dominant co-expression module, but are instead distributed across multiple modules, each contributing a small number of risk genes. These modules may represent distinct and partially independent pathogenic pathways, resulting in elevated ORs for specific module–disorder pairs without achieving statistical significance.

Another potential explanation relates to the genetic architecture of ADHD. Compared with disorders such as MDD that are well described by a common-variant–common-disease model, ADHD risk may be more strongly influenced by a smaller number of variants with moderate-to-large effects, including rare or low-frequency variants. Under such a genetic architecture, gene-set–based enrichment methods that rely on aggregation of many small-effect genes may have reduced power, leading to non-significant P values despite biologically meaningful effect sizes.

Taken together, the combination of limited gene set size, distributed module involvement, and disorder-specific genetic architecture likely explains why the ADHD × MEdarkolivegreen association shows a high OR but does not reach statistical significance (Lines 498-508).

11. Figures 6, S7

The authors divided the modules into three groups: Glia-enriched, Adult neuron-enriched, and Fetal-enriched. However, it is unclear whether these groups are truly appropriate to be clustered together as shown in the heatmap, and whether each group exhibits a clear difference from the others. Statistical metrics that support and justify this grouping are needed.

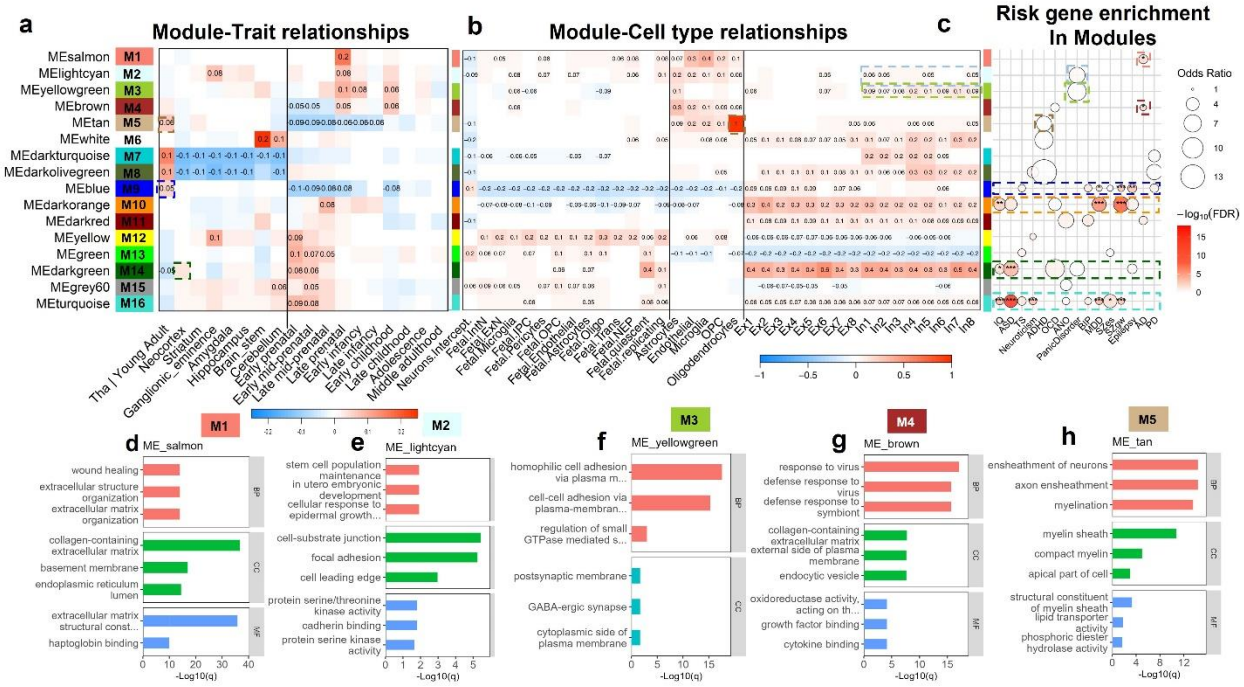
Response:

We thank the reviewer for raising this important point regarding the grouping of co-expression modules. In the revised manuscript, we have removed the categorical division of the top modules into “glia-enriched,” “adult neuron-enriched,” and “fetal-enriched” groups. Instead, we now treat each module individually and characterize its enrichment profile across cell types and developmental stages.

Although visual inspection of the heatmaps in the original version suggested that certain modules appeared preferentially enriched in glial cells, adult neurons, or fetal cell populations, we recognize that this coarse grouping was not sufficiently supported by quantitative metrics. Importantly, the apparent clustering largely reflects the ordering of modules based on their co-expression similarity in the BrainSpan WGCNA analysis, rather than a statistically defined separation at the single-cell level.

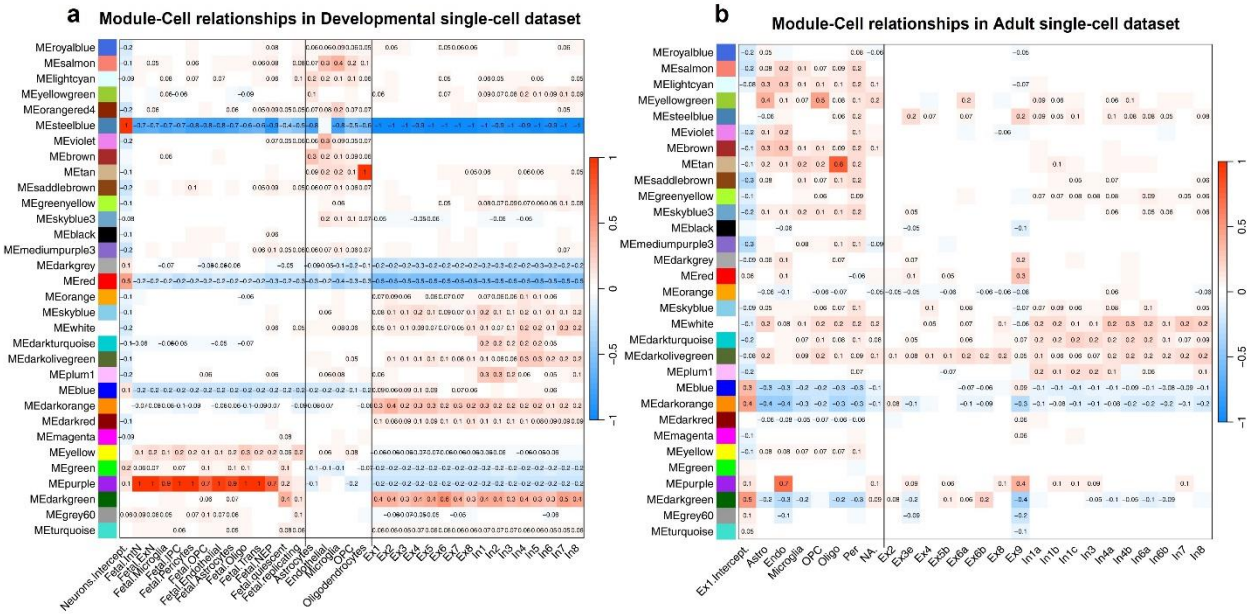
Our updated analyses (revised Fig. 6 and Fig. S12) demonstrate that cell-type enrichment patterns are not strictly mutually exclusive across modules. Several modules show significant enrichment in multiple cell types (e.g., M14 in both fetal quiescent cells and adult neocortical neurons; M2 and M3 in both astrocytes and interneurons), indicating overlapping functional modalities or shared developmental origins. This pleiotropy argues against forcing modules into discrete cell-type–based clusters.

Therefore, we conclude that module-wise analysis of cell-type enrichment provides a more accurate and biologically informative framework than grouping modules into broad categories. The revised Results section and corresponding figures have been updated accordingly to emphasize continuous and overlapping enrichment patterns, rather than categorical module classes (Lines 298-306).



Revised Fig. 6 Spatiotemporal co-expression module-based risk gene enrichment and extrapolated pathogenic mechanisms.

(c) Bubble plot showing the module-level risk gene enrichment for individual traits, with the top three modules per trait selected based on P values. The size of bubbles indicating OR values and color gradient indicating –log10 (FDR) magnitude. P values across all the original 41 WGCNA detected modules for each of the disorders were FDR corrected (P_{fdr} < 0.05: *; P_{fdr} < 0.01: **; P_{fdr} < 0.001: ***).



Revised Fig. S12 Module–cell type relationships in developmental and adult single-cell transcriptomic datasets.

12. In addition, beyond the odds ratios presented in Fig.6C, statistical significance values such as P values should also be provided. Moreover, in module M4, the odds ratio is higher for OCD than for AD. An interpretation of this observation should be included.

Response:

We thank the reviewer for this constructive suggestion. In the revised manuscript, we have updated Fig. 6c to explicitly present both effect sizes and statistical significance for the enrichment of disorder risk genes across spatiotemporal modules. Specifically, enrichment results are now visualized using a bubble plot, in which bubble size represents the odds ratio (OR) and color intensity encodes the corresponding -log10 (FDR-adjusted P value) derived from chi-square tests. P values were corrected for multiple testing across all the initial 41 WGCNA detected modules for each disorder using the false discovery rate (FDR) procedure, and significance thresholds are now clearly indicated in the figure legend. This revised presentation enables a more transparent and quantitative comparison of enrichment strength and statistical support across traits and modules (Revised Fig. 6).

Regarding the observation that module M4 shows a higher OR for OCD than for AD but does not reach statistical significance, we believe this primarily reflects limited statistical power due to the relatively small size of the OCD risk gene set, rather than a lack of biological relevance. Indeed, when we repeated the WGCNA analysis using a more stringent minimum module size (minModuleSize = 120), we identified a module (ME yellow) that closely recapitulates the spatiotemporal enrichment pattern and Gene Ontology functional annotations of the original M4 module. Notably, in this alternative module definition, OCD

risk genes exhibited marginally significant enrichment (P_{fdr} = 0.058, Fig. S9), supporting a consistent, albeit modest, association.

Together, these results highlight that interpretation of enrichment analyses should consider both effect size (OR) and statistical significance, particularly for disorders with smaller risk gene sets. The observed pattern for OCD suggests a potentially meaningful association with M4-related biological processes that may be underpowered in the primary analysis but remains biologically plausible (Lines 277-288, lines 498-508).

Response to Reviewers

Response to Reviewer #1:

We thank the reviewer for the positive evaluation of our work and for the constructive suggestions, which have helped us further improve the clarity of the manuscript and figures.

Comment:

The quantitative analysis is currently presented mainly in supplementary figures S3a and S3b. S3b does a reasonable job summarizing the results, but S3a is confusing. It's blurry, the asterisks representing the p values are very difficult to see and the meaning behind the colored bars in some of the scatterplots is unclear. So, I'd suggest fixing this or removing it.

Response:

We thank the reviewer for this helpful suggestion. Because the matrix composed of 15 disorders × 11 brain regions is relatively large, each boxplot and the corresponding significance annotations became quite small in the original Supplementary Fig. S3a, making the asterisks difficult to recognize, particularly at lower resolution.

To address this issue and improve readability, we excluded the “brainstem” data from statistical comparisons due to the small sample size and increased the size of the asterisks to enhance visibility. Furthermore, we merged the information from the original Figures S3a and S3b into a revised single Figure S3. Specifically, the blue and red arrows that summarized the statistical comparisons in the original S3b are now placed at the lower right of the corresponding boxplots. This revised layout enables readers to more intuitively interpret both the direction and statistical significance of each pairwise comparison while directly visualizing the underlying data distributions.

We believe this revised presentation substantially improves the clarity and interpretability of the results.

Response to Reviewer #2

We sincerely thank the reviewer for the positive feedback and for acknowledging the improvements made in the revised manuscript.

Comment:

One issue that remains, however, is the inconsistency in Fig. 3 between the rebuttal letter and the main text regarding the classification results for Panic Disorder.

Response:

We thank the reviewer for carefully pointing out this inconsistency. The issue indeed occurred due to an oversight during the previous revision process. While the revised versions of Figures 2 and 3 were correctly updated in the main text, the corresponding figures in the rebuttal letter were not replaced in time, which resulted in the inconsistency noted by the reviewer.

We have now carefully rechecked all figures, results, and descriptions throughout the main text and supplementary materials. We confirm that the results and descriptions presented in the main manuscript and supplementary materials are now correct and internally consistent.