nature portfolio

Peer Review File

Linking Changes in Sulcal Morphometry to Cognitive Development from Childhood to Adolescence

Corresponding Author: Professor Shuyu Li

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)
The manuscript entitled “Linking Changes in Sulcal Morphology to Cognitive Development from Childhood to Adolescence” is an interesting, comprehensive examination of several sulcation metrics in youth. Using a longitudinal sample of structural T1-weighted data, they first find age-related changes in several sulcal structural metrics. They then compare known developmental sulcal timing to their data, finding stronger changes in their sample for sulci with earlier emergence. Next, they examine brain-behavior relationships via multivariate analysis and find a few statistically significant associations. Finally, they explore genetic associations by correlating sulcal changes in their data with expression data derived from the Allen Human Brain Atlas (AHBA), finding several statistically-significant gene sets associated with sulcation measures.

There are several strengths to the study. In general, the introduction and discussion sections are well-written. The sample is reasonably large, and the use of a longitudinal design further increases power to detect age-related effects. The image acquisition and preprocessing (e.g. pulse sequence, FreeSurfer) are widely-accepted in the neuroimaging field. Sulcation metrics are relatively understudied in the literature, and this study will make a meaningful contribution in this area. The authors have meticulously described image processing QA procedures and excluded spurious data, which is greatly appreciated. In broad strokes, the statistical tools used (GAMMs, LASSO, AHBA analytics, gene enrichment analysis) are appropriate. The authors provide multiple interesting analyses (i.e. age-related change, brain-behavior associations, genetic associations) and present a wealth of results that will increase the study’s impact.

There are also a few weaknesses:
-First and probably most importantly, the majority of the described associations throughout the manuscript are fairly small. For example, the largest main effect of age on any sulcal measure has an adjusted R2 of 0.114. In other words, although the age-related changes are highly statistically significant, the effect sizes are modest at best.
-The Results described in the “Sulcal emergence timing” section will be confusing for the casual reader. The explanation for Tapp, T10, and T50 measures is sparse and understanding them is central to this analysis. In particular, I think more details about the definition of Tapp would be useful. It is also unclear why F-statistic is being used for T10 associations with SA, as the Methods (including Supplementary Methods) do not include any description of this portion of the analyses, nor mention of using p-values as a regressor; this begs the question what other correlations were tested for this section that aren’t mentioned? It also is unclear why SA should have positive correlations with both R2 and p-value for the same measure, as generally larger effects should have smaller p-values? Further, I think that the legend for Figure 3 adds to confusion about these analyses; I would make it clear that Tapp, T10, and T50 are effectively three different measures of sulcal emergence (albeit with similar patterns), all assessed across gestational age, and not developmental timepoints tested in and of themselves (the word “stratification” in Figure 3 title adds to the confusion since it implies categorization, and the only categorical portions of Figure 3 are the distinction between Tapp, T10, and T50 brain maps). Many of these issues could be reduced by including a section in the Methods explicitly describing the analyses for sulcal emergence.
-The reported brain-behavior associations are particularly weak. Although the analytic approach is largely valid, it is notable that several massively multivariate analyses resulted in only three statistically significant brain-behavior associations. Although the authors do describe literature supporting evidence for their findings in the Discussion, the overall the face validity of these results is relatively low given the risk of stochastic errors and overfitting.
-The gene enrichment analysis is interesting, and I particularly appreciated the care to avoid spatial autocorrelations in the data. There are a couple of other potential pitfalls to this analysis, perhaps the most obvious is that AHBA data is based on older adults, which requires the assumption that the biological processes influencing sulcation metrics are the comparable

throughout the lifespan -- and it is doubtful that this is entirely true. Another question is the spatial localization of AHBA to the data; although co-registrations to AHBA are not uncommon in many recent MRI papers, the fact that the current study focuses on sulci raises the question of how many AHBA probes were actually sampled from adjacent gyri (i.e. are not sulcal)? The authors do note that anatomically ambiguous measures were excluded, but do not provide details, for example what % of probes were excluded. More information on alignment procedures would be necessary for replication, and this probably should be included in the Methods.
-The Discussion does not include a Limitations section. Although there are many interesting aspects to this paper, each section has its own limitations and caveats that should be communicated to the reader. For example, 1) the main effects of age are based on relatively limited longitudinal data, with >75% of the sample having 2 or fewer timepoints. Effect sizes also are small, 2) somewhat discordant results on sulcal emergence based on metric, 3) fairly weak brain-behavior associations, and 4) numerous unstated assumptions in AHBA analyses.

In summary, this is an thought-provoking, multifaceted study on a relatively understudied topic that presents fairly convincing evidence of age-related changes in sulcation metrics in youth, as well as interesting results in several other related domains. I think the study would benefit from more methodological clarity to facilitate future replication, as well as a more explicit descriptions of the inherent limitations to the complex analytic approaches used.

Minor points:
-Please Confirm/explicitly state that the scan was MP-RAGE.
-Please describe the rationale for why some sulcal measures were TIV-controlled and others were not.
-Much of the presented results in the section “Morphological Changes of Each Sulcus Throughout the Brain” listing structures and their associated R2 and p-values could potentially be better presented in a table, including nonsignificant results. Some other data that aren’t presented in detail (i.e. sex effects) could also be included in a table.
-Page 4, line 151: I don’t think I would consider R2 of -0.040 a “pronounced decrease.” Given the fairly large N, I would focus more on effect sizes than p-values when interpreting results.
-Figure 2a: It would be useful to mention in the figure legend that these plots are for S.O.T.lat_left specifically, and not a global measure.
-Figure 3: please check to make sure that the y-label (R2) is correct since the scale is substantially different than the adjacent panel.
-When presenting beta-weights in the text, it is difficult for the reader to get a sense of effect size, since the scale of these measures (e.g. mm) is not explicitly defined in the manuscript (although scales can be inferred based on the figure labels). Consider reporting standardized beta weights instead in order to improve readability. Also, the number of significant figures should be consistent throughout the manuscript.
-In general, the visibility of many axis labels could be improved, primarily by using larger fonts for numeric values. The number of significant figures also is inconsistent from panel to panel.

Typographical Errors:
-Page 4, line 145, ‘ ” ‘ .
-Page 9, line 332, extra space “(exclude”

Reviewer #2

(Remarks to the Author)
Brief summary of the manuscript:
The authors analyze longitudinal MRI from 312 children (6–14 years; 490 scans) to quantify developmental changes in sulcal morphometry (surface area, depths, width, length, cortical thickness) across several whole brain sulci and relate these to cognition. They report systematic age-related trajectories (e.g., thinning and sulcal widening), a link between fetal emergence timing and sulcal area change, and associations between changes in specific sulci and executive control/working memory performance. A transcriptomic enrichment analysis further implicates synaptic and cytoskeletal processes in sulcal development. Overall, the study offers a comprehensive, multimodal view of how sulcal morphometry relates to cognitive maturation.

Overall impression of the work:
This is a solid, carefully executed study with notable strengths: a large developmental cohort, rigorous longitudinal modeling, appropriate TIV adjustment, and a thoughtful integration of morphometric, cognitive, and gene-expression analyses. I particularly appreciated the use of LASSO to link longitudinal sulcal change to cognition and the explicit treatment of brain-size correction, which is essential in developmental work. Beyond documenting trajectories, the joint morphometry–cognition–gene framework points to plausible biological pathways and functional consequences, advancing a field that has long sought mechanistic accounts of cortical folding. If adopted widely, the authors’ modeling choices and systematic TIV adjustment could help set more rigorous standards for developmental morphometry. With clearer terminology (morphology vs. morphometry), resolution of a few interpretational inconsistencies, explicit reporting of sex effects, and a more integrative Discussion that connects findings to mechanisms and future directions, the manuscript will be substantially strengthened and broadly informative to the community.

MAJOR POINTS:

1. Use the term “morphometry” consistently (not “morphology”) for quantitative measures. Throughout the paper, when referring to quantitative indices (sulcal area, depth, cortical thickness, length, width), use morphometry/morphometric rather than morphology/morphological. The (qualitative) folding pattern is longitudinally stable after birth. You already hint at this

distinction (e.g., “morphometric features” in line 229); please adopt consistent terminology across the manuscript and add a brief sentence in the Introduction clarifying morphology vs. morphometry, with appropriate citations you will find in:

Cachia, A. et al. Longitudinal stability of the folding pattern of the anterior cingulate cortex during development. Dev. Cogn. Neurosci. 19, 122–127 (2016).

2. Direction of the “emergence timing” effect needs to be clarified and aligned across Results, Fig. 3 caption, and Discussion. In the Results (lines 185–186) you say: “These findings suggest that earlier emergence generally associated with stronger age-related effects in surface area measures.”, in the Fig.3 caption you say: “Earlier-emerging sulci demonstrate more pronounced developmental changes in sulcal area during childhood and adolescence.” And also in the discussion (lines 276-277): “We found that sulci appearing later in fetal development exhibited less pronounced changes in surface area from childhood to adolescence.”. Indeed, the discussion (lines 277-280) ends like this: “This finding aligns with the hypothesis that earlier-developing sulci are subject to more robust genetic regulation throughout the lifespan, possibly due to the brain's prioritization of regions vital for essential functons” which seems like a fair way to argue, as more robust genetic expression supports the result of few changes during postnatal development in the early-forming sulci. The correct direction of the effect and its correct interpretation should be clarified at the indicated points.

3. Report and discuss sex effects explicitly. You note robust sex main effects across 214 models (lines 167–169). Please (i) present a concise table (main text or Supplement) summarizing which features/sulci showed sex effects and effect directions, and (ii) add a short paragraph in the Discussion interpreting these findings in light of developmental literature.

4. Expand the Discussion on the sulcus of Jensen (secondary intermediate ramus of IPS) and cognition. In lines 285–287, you link this region to executive control within the multiple-demand system; please strengthen this with targeted sulcal-morphology/cognition literature on IPS, e.g.:

Santacroce F, Cachia A, Fragueiro A, Grande E, Roell M, Baldassarre A, Sestieri C, Committeri G. Human intraparietal sulcal morphology relates to individual differences in language and memory performance. Commun Biol. 2024 May 2;7(1):520. doi: 10.1038/s42003-024-06175-9. PMID: 38698168; PMCID: PMC11065983. e

Roell, M. et al. Sulcation of the intraparietal sulcus is related to symbolic but not non-symbolic number skills. Dev. Cogn. Neurosci. 51, 100998 (2021).

5. Speculate more concretely about candidate white-matter pathways (lines 293–295). Where you discuss structure–function coupling, consider hypothesizing tract involvement consistent with the implicated sulci: e.g., fronto-parietal SLF II/III for IPS–prefrontal links (executive control), ILF/VOF for posterior intralingual/occipito-temporal contributions (visual memory/working memory interfaces), and fronto-orbital connections (e.g., anterior thalamic radiations/frontostriatal pathways) for orbital sulcus and alerting. These are framed as plausible mechanisms guiding future work rather than definitive claims.

6. In general, broaden and deepen the Discussion across key findings. Several sections currently read as concise lists. For example, lines 300–302 propose that wider sulci may reflect neural reorganization underlying better performance; please add mechanistic justifications and insert citations consistent with your enrichment results to strengthen these claims. Apply the same treatment to the other principal results so each has (i) speculation about the effects’ cause, (ii) citation in ligne or in contrast with the results, and (iii) a forward-looking implication for future research/interventions.

7. Methods transparency on model counts and longitudinal change computation.
- Please state explicitly the total number of GAMMs tested (e.g., N sulci × N features).
- Clarify if longitudinal change analyses (Δ metrics, LASSO) included only participants with ≥ 2 timepoints or the overall sample (312). In the second case, it needs to be better explained how subjects who have only one MRI scan are considered in longitudinal analyses.

MINOR POINTS:

- Clarify Hypothesis 1 (lines 96–97). As written, it is somewhat vague.
- Define better the term “Tapp” on first use.
- Consider adding a concise Supplement for QC outcomes. You excluded sulci with < 75% extraction success and removed three specific sulci (lines 383–387). A supplemental table listing success rates per sulcus and an explanatory figure showing a bad sulcal segmentation would improve reproducibility.

Reviewer #3

(Remarks to the Author)
Overview: Thank you for the opportunity to review this manuscript! It was very well written and the figures are clear and informative. Briefly, the authors examined a cohort of children, aged 6-14 years, at multiple time points, acquiring both brain structural and cognitive data. The authors initially examined the age-related development of several sulci; and then linked age-related change in sulci morphometry to cognitive measures and gene expression data (derived from the Allen Human Brain Atlas). While the manuscript is relevant and timely, several aspects remain to be addressed, as outlined below. This particularly includes emphasising/clarifying the rationale for your analysis at the sulcus level; and clarifying how you accounted for the different ages, inter-scan intervals, and numbers of acquired time points (especially since around 168 of your participants only had cross-sectional MRI data).

Abstract:
- Could the authors include the inter-scan interval in the abstract
- Would it be possible to include statistical results

Introduction:
- It would be helpful to further elaborate on the biological/mechanistic relevance of sulcal morphological changes, including to clinical phenotypes
- Line 76: this sentence is a bit confusing; can you clarify which sulcus/gyrus would relate to eg fine motor skills, reasoning/memory etc.?
- Overall very well written and clear

Results:
- Again, can you please include inter-scan interval times
- The number of abbreviations is quite high and at times difficult to follow – I would consider writing out some of the terms you are currently abbreviating (e.g. sentence starting in line 128 includes 6 abbreviations)
- As the results come before the methods, can you talk a bit more about how you examined the age-related changes, especially since you had 3 timepoints
- Are changes modelled linearly or non-linearly?
- Paragraph starting in line 146: this is VERY hard to follow – can you instead move this information to a table?
- Paragraph starting in line 170: this is very interesting but comes a bit out of the blue; can you integrate this more and explain, why/how this analysis added to the rest of the work? Also, there are again a lot of terms and abbreviations in this paragraph which, if you haven't yet read the methods, are difficult to follow.
- When you say "earlier emergence associated with stronger age-related effects", does this mean that sulci formed earlier show a stronger change over time during childhood?
- Regression results – can you include significance levels?

Discussion:
- How do you account for the wide age range of your participants? Surely someone aged 6 years will show a different development compared to someone aged 14 years?
- What about regions where you do not observe age-effects? How would you interpret their 'stability' over time?
- Do 'non-changing' sulci also predict clinical performance? And how are they underpinned by molecular mechanisms?
- Overall, the concept of studying sulci and gyri, rather than e.g., vertices or regions, is very interesting and important – but could you emphasize more (in the introduction/discussion) why that is the case, and what this approach adds to the aforementioned analyses at the region- or vertex-level?
- Line 276: I cannot follow your argument here – if earlier developing sulci change more during childhood, would that not indicate that those sulci are in fact more susceptible to environmental regulation?
- Line 304: by alterations, do you mean reductions?
- Would you consider adding strengths and limitations, and a conclusion?

Methods:
- Line 332: correct "with cognitively abnormal"
- You should be clearer throughout the manuscript that only 47 participants had all three scans and that 186 participants had cross-sectional data only
- How did you examine 'change' in individuals with one scan only?
- Why were children recruited based on their performance on a cognitive test?
- How did you account for the very large number of models (presumably ~50 sulci by six features)?
- How did you account for the different times of scanning, the difference inter-scan intervals, and the fact that some individuals have one, two, or three scans?

- Were cognitive performances age-normed? Would it be helpful to examine percentage-wise change rather than delta?

- When you examined association of anatomical change and clinical profiles, did you base your analysis on the regression betas of change? And if so, how did you accommodate individuals with 1 scan? And for individuals with three scans, did you examine change between T1 and T2, and T2 and T3 separately?

- For the gene decoding and enrichment analysis: can you clarify a bit more how you downsampled AHBA gene values to your sulci / gyri regions of interest? (ie what do you mean by "selected sulcal regions were systematically merged according to predetermined grouping criteria"? What do you mean by "regions deemed irrelevant to the analysis")

- Figures: helpful, but text is (partly) too small; for instance, axis labels of Figure 2 are difficult to decipher.

- Figure (no number) with Subject Age distribution – what is the x-axis here?; also, can you plot separately for individuals with one/two/three scans; colors are very difficult to make out in this graph (including the green/purple outline); also, what do ANT/WM mean? Please provide a figure legend

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
The authors have addressed the majority of my questions. This is an interesting, comprehensive manuscript that improves our understanding of sulcal development and will be of interest to many readers. In particular, I appreciate the expansion of the methods to improve replication and the addition of a limitations section.

As a minor point, there remains some inconsistencies with significant digits which should be corrected prior to publication.

Reviewer #2

(Remarks to the Author)
The authors have responded adequately to my feedback from the last review and made the necessary revisions to their manuscript.

Reviewer #3

(Remarks to the Author)
I thank the authors for their careful response to my comments; and I believe, in doing so, they have much improved the manuscript.

RESPONSE TO REVIEWER COMMENTS

Reviewer #1:

The manuscript entitled “Linking Changes in Sulcal Morphology to Cognitive Development from Childhood to Adolescence” is an interesting, comprehensive examination of several sulcation metrics in youth. Using a longitudinal sample of structural T1-weighted data, they first find age-related changes in several sulcal structural metrics. They then compare known developmental sulcal timing to their data, finding stronger changes in their sample for sulci with earlier emergence. Next, they examine brain-behavior relationships via multivariate analysis and find a few statistically significant associations. Finally, they explore genetic associations by correlating sulcal changes in their data with expression data derived from the Allen Human Brain Atlas (AHBA), finding several statistically-significant gene sets associated with sulcation measures.

There are several strengths to the study. In general, the introduction and discussion sections are well-written. The sample is reasonably large, and the use of a longitudinal design further increases power to detect age-related effects. The image acquisition and preprocessing (e.g. pulse sequence, FreeSurfer) are widely-accepted in the neuroimaging field. Sulcation metrics are relatively understudied in the literature, and this study will make a meaningful contribution in this area. The authors have meticulously described image processing QA procedures and excluded spurious data, which is greatly appreciated. In broad strokes, the statistical tools used (GAMMs, LASSO, AHBA analytics, gene enrichment analysis) are appropriate. The authors provide multiple interesting analyses (i.e. age-related change, brain-behavior associations, genetic associations) and present a wealth of results that will increase the study’s impact.

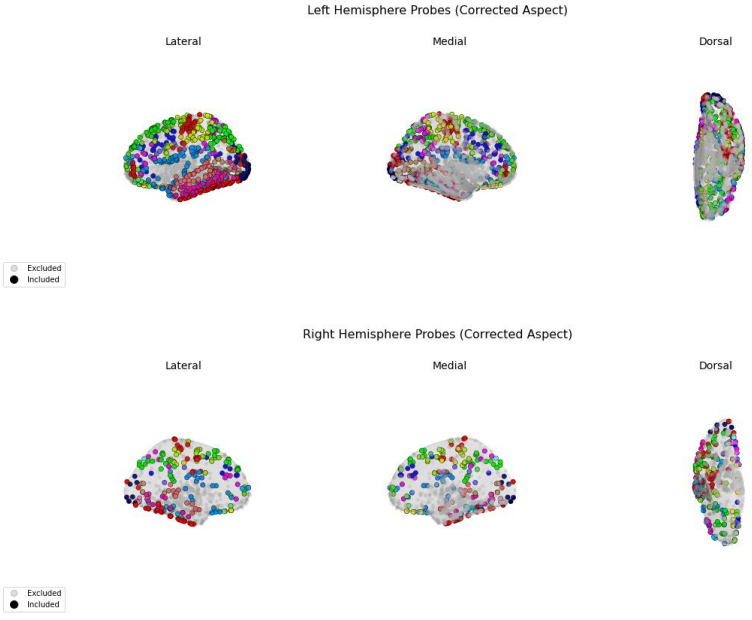
There are also a few weaknesses:

1.First and probably most importantly, the majority of the described associations throughout the manuscript are fairly small. For example, the largest main effect of age on any sulcal measure has an adjusted R2 of 0.114. In other words, although the age-related changes are highly statistically significant, the effect sizes are modest at best.	**Response：** We are sincerely grateful to the reviewer for raising this point regarding the magnitude of the effect sizes. It highlights a crucial nuance in developmental neuroimaging that is often overlooked. We fully acknowledge that the reported effect sizes (adjusted R^2 up to ~0.11) appear modest at first glance. However, after carefully benchmarking our results against a latest large-scale reference datasets and considering the specific biological context of adolescent sulcal dynamics, we respectfully submit that these effects represent robust and biologically substantial developmental trends, supported by the following three lines of evidence: 1. Benchmarking against "New Normals" in Large-Scale Studies: We contextualize our findings within the recent paradigm shift established by landmark studies such as Marek et al. (2022, Nature). In their analysis of ~50,000 individuals, they demonstrated that reproducible brain-wide associations are significantly smaller than historically assumed. Specifically, they reported that the median univariate effect size ($\|r\|$) was only 0.01, and even the top 1% of largest associations reached an $\|r\|$ of just >0.06 ($R^2 \approx 0.0036$). In stark contrast, our observed maximal adjusted R^2 of 0.114 (corresponding to an $\|r\| \approx$ 0.33) stands well above these benchmarks. This suggests that the "modest" effects we report are, in fact, comparatively strong and statistically robust signals for specific morphological features. 2. Biological "Subtlety" of Sulcal Tuning (vs. Thickness): The magnitude of our observed effects aligns perfectly with the biological reality of this age window (6 – 14 years). Unlike the dramatic structural establishment seen in infancy, this period is characterized by a gradual "fine-tuning" or "flattening" of the sulcal landscape. Quantitative longitudinal data from Tamnes et al. (2017) confirm that this process is inherently subtle: they found that the annual percentage change (APC) for Surface Area (mean APC ≈ -0.55%) is significantly smaller than that for Cortical Thickness (mean APC ≈ -1.03%) or Volume (mean APC ≈ -1.43%). Our results

	mirror this biological pattern—showing gradual sulcal widening alongside more pronounced cortical thinning—indicating that the smaller effect sizes in sulcal measures reflect a true biological property (structural stability) rather than measurement noise. 3. Enhanced Specificity via TIV Adjustment: Crucially, our analytical strategy prioritized anatomical specificity over raw variance explanation. By utilizing TIV-adjusted residuals, we rigorously regressed out the "global scaling" effects of brain growth. While leaving global volume uncorrected would certainly yield higher raw R^2 values, it would conflate specific regional remodeling with general somatic growth. The variance captured in our models (up to 11%) therefore represents pure, localized topological reshaping (e.g., specific deepening or widening of a sulcus) independent of head size, providing a conservative but highly specific characterization of neural development. 4. Methodological Rigor and Robust Statistical Framework: Finally, to ensure that these modest effect sizes represent genuine developmental signals rather than statistical artifacts, we employed a rigorous dual-verification approach using Generalized Additive Mixed Models (GAMMs). As detailed in our Methods section, we did not rely solely on the significance of the smooth age term. Instead, we implemented a strict nested model comparison strategy, testing the full model (Age + Sex) against a null model (Sex only) using Likelihood Ratio Tests. We defined a developmental effect as significant only if it simultaneously satisfied two criteria: (1) significant smooth terms for age, and (2) a significant improvement in model fit (Likelihood Ratio Test, $p<0.05$) compared to the null model. Furthermore, all results were subjected to False Discovery Rate (FDR) correction to control for multiple comparisons across the whole brain. This conservative statistical framework ensures that the reported associations, while modest in magnitude, are statistically reliable and robust against Type I errors. **Revisions in Manuscript:** [Page 8, Lines 304] *"In contrast to the dramatic cortical expansion observed in infancy, the sulcal changes in school-age children were modest in magnitude (annual rate < 1%) and effect size ($R^2 \approx 0.017$). These results align with recent large-scale neuroimaging evidence indicating that reproducible brain-phenotype associations are typically characterized by small effects(*Marek et al., 2022)*."* **References:** Marek, S. et al. Reproducible brain-wide association studies require thousands of individuals. Nature 603, 654–660 (2022). Alemán-Gómez, Y. et al. The human cerebral cortex flattens during adolescence. J. Neurosci. 33, 15004–15010 (2013). Tamnes, C. K. et al. Development of the cerebral cortex across adolescence: A multisample study of inter-related longitudinal changes in cortical volume, surface area, and thickness. J Neurosci 37, 3402–3412 (2017).
2.The Results described in the "Sulcal emergence timing" section will be confusing for the casual reader. The explanation for Tapp, T10, and T50 measures is sparse and understanding them is central to this analysis. In particular, I think more details about the definition of Tapp would be useful. It is also unclear why F-statistic is being used for T10 associations with SA, as the	**Response:** We appreciate the reviewer's request for clarification regarding the sulcal emergence definitions and statistical metrics. We agree that the previous description was confusing. We have extensively revised this section to clarify our methodology and findings: 1. Clarified Definitions (T_{app}, T_{10}, T_{50}): We have added a dedicated definition section in the Methods. We clearly distinguish between T_{app} (visual appearance time based on anatomical consensus, e.g., Nishikuni & Ribas,

Methods (including Supplementary Methods) do not include any description of this portion of the analyses, nor mention of using p-values as a regressor; this begs the question what other correlations were tested for this section that aren't mentioned? It also is unclear why SA should have positive correlations with both R2 and p-value for the same measure, as generally larger effects should have smaller p-values? Further, I think that the legend for Figure 3 adds to confusion about these analyses; I would make it clear that Tapp, T10, and T50 are effectively three different measures of sulcal emergence (albeit with similar patterns), all assessed across gestational age, and not developmental timepoints tested in and of themselves (the word "stratification" in Figure 3 title adds to the confusion since it implies categorization, and the only categorical portions of Figure 3 are the distinction between Tapp, T10, and T50 brain maps). Many of these issues could be reduced by including a section in the Methods explicitly describing the analyses for sulcal emergence.	2013), and T_{10}/T_{50} (quantitative folding milestones derived from curvature modeling, Snyder et al., 2024). 2. Standardized Statistics (Removing F-stats): We have replaced the confusing F-statistics with 2 metrics in standard Generalized Additive Mixed Models (GAMMs): **Significance: FDR-corrected p-values (P_{FDR}).** **Signed Effect Size (Signed Δadj R^2):** To capture both the magnitude and direction of developmental changes, we assigned the sign of the age coefficient to the $\Delta \text{adj } R^2$. Since SA and maxD generally decrease with age in our sample, early-emerging sulci exhibit large negative values (strong decrease), while late-emerging sulci exhibit values closer to zero (weaker decrease). 3. Resolving the Correlation Logic: The reviewer correctly questioned the relationship between effect sizes and p-values. With our refined "Signed R^2" metric, we observed a positive correlation between emergence time (T_{app}) and Signed R^2. This indicates that earlier-emerging sulci (lower T_{app}) show stronger negative developmental effects (more negative R^2) and higher statistical significance (lower p-values). This resolves the apparent contradiction and suggests that structures formed earlier in gestation undergo more substantial morphological remodeling (reduction) during childhood. 4. Revised Figure 3: We have updated Figure 3 and its legend to strictly separate our plotted metrics (Panel A) from literature-derived reference maps (Panel B & C) and the statistical correlations (Panel D). **Revisions in Manuscript:** [Page 13, Lines 529] Definition of Sulcal Emergence Timing: *To characterize the chronological hierarchy of cortical folding, we defined sulcal emergence using three complementary metrics...* *We first defined the Sulcal Appearance Time (Tapp) as the gestational age (in weeks) at which a sulcus first becomes visibly identifiable on the cortical surface... derived from fetal autopsy studies and in utero MRI assessments...* *To complement these discrete visual observations, we utilized continuous quantitative milestones... the onset of folding (T10), defined as the gestational age where local mean curvature reaches 10% of its maximum projected adult value, and the midpoint of folding (T50), defined as the gestational age corresponding to 50% of curvature maturation.* **References:** 1. Snyder, W. E. et al. A bimodal taxonomy of adult human brain sulcal morphology related to timing of fetal sulcation and trans-sulcal gene expression gradients. Neuron 112, 3396-3411.e6 (2024). 2. Nishikuni, K. & Ribas, G. C. Study of fetal and postnatal morphological development of the brain sulci: Laboratory investigation. Journal of Neurosurgery: Pediatrics 11, 1 – 11 (2013). 3. Dubois, J. et al. Mapping the early cortical folding process in the preterm newborn brain. Cerebral Cortex (New York, N.Y.: 1991) 18, 1444 – 1454 (2008). 4. Chi, J. G., Dooling, E. C. & Gilles, F. H. Gyral development of the human brain. Ann. Neurol. 1, 86 – 93 (1977).
3.The reported brain-behavior associations are particularly weak. Although the analytic approach is largely valid, it is notable that several massively multivariate analyses resulted in only three statistically significant brain-behavior associations.	**Response:** We sincerely appreciate the reviewer's rigorous assessment of the validity of our findings. We agree that in high-dimensional neuroimaging analyses, the risk of overfitting and stochastic errors is a primary concern. To address this and bolster the face validity of our results, we have made the following major revisions:

Although the authors do describe literature supporting evidence for their findings in the Discussion, the overall the face validity of these results is relatively low given the risk of stochastic errors and overfitting.	1. Implementation of Hierarchical Validation (Triangulation): We have moved beyond relying on a single analytical stream. As detailed in the revised Methods and Results sections, we now employ a "triangulation" strategy. We first used LASSO regression with nested cross-validation specifically to handle high-dimensionality and prevent overfitting by enforcing sparsity (L1 penalty). Crucially, we then validated the features selected by LASSO using two complementary approaches: ✓ Multivariate PLS Correlation: To confirm that these features are part of a robust global covariance pattern (validated via permutation testing and bootstrapping). ✓ Univariate Partial Correlation (control age and gender): To confirm region-specific effects using strict False Discovery Rate (FDR) correction ($q < 0.05$) while controlling for confounders. The convergence of these three distinct statistical frameworks on the same key regions (e.g., the left orbital frontal sulcus for alerting, the IPS for executive control) strongly suggests that these are robust biological signals rather than stochastic noise. 2. Demonstration of Developmental Specificity: To further rule out spurious correlations, we performed the exact same analytical pipeline on the baseline (static) data (see Method "*Baseline brain – behavior associations.*"). This analysis yielded null results (no significant PLS or FDR-corrected associations). This dichotomy highlights that our significant longitudinal findings are not random artifacts, but are specifically sensitive to the dynamic changes in brain structure occurring over time. 3. Strict Feature Selection and Rigorous Statistical Validation: To explicitly address the risks of stochastic errors and overfitting, we implemented a conservative statistical framework designed to minimize Type I errors. For predictive modeling (LASSO): We utilized nested 5-fold cross-validation to optimize the regularization parameter (λ) and evaluated predictive performance on an independent held-out test set (20% of the sample). This ensures that the reported associations are generalizable and not artifacts of the training data1. We further validated the necessity of this sparse feature selection by comparing LASSO against non-sparse models (e.g., Ridge regression), which failed to generalize (Test set r=0.07, P=0.731), confirming that the identified associations are driven by specific topographical changes rather than global noise. For multivariate covariance (PLS): Significance was established via 1,000 permutation tests (overall P<0.05) to test the latent modes against random chance, and feature stability was assessed using 1,000 bootstrap iterations. For univariate analysis: All reported partial correlations were subjected to strict False Discovery Rate (FDR) correction (q<0.05) across multiple comparisons.
4.The gene enrichment analysis is interesting, and I particularly appreciated the care to avoid spatial autocorrelations in the data. There are a couple of other potential pitfalls to this analysis, perhaps the most obvious is that AHBA data is based on older adults, which requires the assumption that the biological processes influencing sulcation metrics are the comparable throughout the lifespan -- and it is doubtful that this is entirely true. Another question is the spatial localization of AHBA to the data; although co-registrations to AHBA are	**Response:** We thank the reviewer for raising these critical points regarding the limitations of the AHBA dataset and the spatial specificity of our gene expression analysis. We have addressed these concerns by adding a detailed "Limitations" section and expanding the "Methods" section to clarify our rigorous probe selection procedure. 1. **On the use of adult gene expression data for developmental studies:** We acknowledge that the AHBA is derived from adult donors, which is a limitation when studying brain development. However, this is a widely accepted approach in the field of imaging transcriptomics due to the scarcity of high-resolution developmental gene expression atlases. We rely on the assumption that the spatial topography of gene expression (i.e., the relative expression gradients across brain regions) is established early in development and remains relatively conserved into adulthood. As demonstrated Kang et al. (2011), the vast majority of regional transcriptional divergence occurs prenatally. In the

not uncommon in many recent MRI papers, the fact that the current study focuses on sulci raises the question of how many AHBA probes were actually sampled from adjacent gyri (i.e. are not sulcal)? The authors do note that anatomically ambiguous measures were excluded, but do not provide details, for example what % of probes were excluded. More information on alignment procedures would be necessary for replication, and this probably should be included in the Methods.	postnatal lifespan (including adolescence and adulthood), the spatial topography of gene expression remains relatively stable. In addition, the spatial patterning of cortical gene expression is stable between late childhood/adolescence and adulthood (Whitaker et al., 2016). While absolute expression levels may change, the regional differences—which drive our correlation analyses—are robust. We have now explicitly stated this limitation and its underlying assumption in the Discussion. **2. On the spatial localization of probes to sulcal regions:** The reviewer rightfully asks whether the probes attributed to sulci might be "contaminated" by adjacent gyral signals. To ensure anatomical specificity, we employed a strict probabilistic filtering pipeline: **Probabilistic Mapping:** Instead of simple binary masks, we utilized a probabilistic sulcal atlas (derived from BrainVISA) and applied a strict probability threshold (0.2) to define the "core" sulcal regions. This method effectively filters out the ambiguous boundaries where sulci transition into gyral crowns. Our filtering process was rigorous. Out of the 3702 probes mapped to the vicinity of our regions of interest by abagen toolbox, we excluded approximately 76% of the samples because they fell into the probabilistic background or ambiguous zones (see Methods). **Anatomical Verification:** We quantitatively analyzed the anatomical labels of the included probes provided by the AHBA ontology. Among the 881 retained samples, although only 56 samples explicitly contained the term "sulcus", a substantial majority (609 samples, ~69%) were annotated as located on a "bank" (e.g., inferior bank of gyrus, bank of the central sulcus). In neuroanatomy, a "bank" refers to the sulcal wall, confirming that these probes capture the biological signal of the sulcal folding geometry rather than the gyral crests. To further validate this spatial specificity, we visualized the probe distribution on the fsaverage surface. Retained probes are color-coded by their assigned sulcal regions, while excluded probes are displayed in gray.  **Revisions in Manuscript:** Location: Methods – Gene Enrichment Analysis [Page 17, Lines 678] *"The resulting 3D NIFTI file was processed using the abagen toolbox (version 0.1.3; https://github.com/rmarkello/abagen) to generate regional gene expression profiles. A rigorous filtering process was applied to explicitly exclude probes not belonging to the target sulcal regions. Specifically, out of 3,702 probes initially mapped to the vicinity of our regions of interest, 2,821 (approximately 76.2%) were excluded as they fell onto the background or gyral*

	regions. Consequently, 881 probes located directly within the sulci were retained for analysis, ensuring that the expression profiles accurately reflected sulcal folding geometry rather than gyral signals." Location: Discussion – Limitations [Page 11, Lines 432] *"Finally, the gene enrichment analysis relied on the adult AHBA atlas. Although the spatial topology of cortical gene expression is relatively conserved postnatally, developmental-specific dynamic expression patterns may be overlooked."* **References:** 1. Kang, H. J. et al. Spatio-temporal transcriptome of the human brain. Nature 478, 483 – 489 (2011). 2. Whitaker, K. J. et al. Adolescence is associated with genomically patterned consolidation of the hubs of the human brain connectome. Proceedings of the National Academy of Sciences 113, 9105 – 9110 (2016).
5.The Discussion does not include a Limitations section. Although there are many interesting aspects to this paper, each section has its own limitations and caveats that should be communicated to the reader. For example, 1) the main effects of age are based on relatively limited longitudinal data, with >75% of the sample having 2 or fewer timepoints. Effect sizes also are small, 2) somewhat discordant results on sulcal emergence based on metric, 3) fairly weak brain-behavior associations, and 4) numerous unstated assumptions in AHBA analyses.	**Response:** We thank the reviewer for this insightful suggestion. We agree that a transparent discussion of the study's limitations is crucial for a balanced interpretation of the findings. Accordingly, we have added a dedicated "Limitations" section in the Discussion. In this new section, we explicitly address the four points raised: 1. Longitudinal constraints: We acknowledge that despite the large sample size, the longitudinal depth per subject is limited, which constrains the modeling of complex non-linear trajectories. 2. Effect sizes: We discuss the modest effect sizes and the need for caution in clinical translation. 3. Discordance in metrics: We clarify the heuristic nature of the sulcal emergence metrics (T10, T50, Tapp) and the lack of "ground truth" biological validation, which explains potential discordance. 4. AHBA assumptions: We discuss the limitations of using an adult gene atlas for developmental analysis. **Revisions in Manuscript:** *Limitations*[Page 11, Lines 417] *In contrast to traditional vertex-wise analyses, the sulcal morphometry strategy employed in this study is anchored to topologically invariant anatomical units86, thereby promoting the biological homology of the observed developmental trajectories. However, several limitations warrant consideration when interpreting these results. First, despite the substantial total sample size, the longitudinal sampling density was limited, with the majority of subjects contributing only 2-3 time points (>75%). This constrains our capacity to fully characterize complex, non-linear developmental trajectories. Second, the observed effect sizes were modest. While this is consistent with typical patterns in large-scale neuroimaging association studies, it warrants caution regarding the immediate translation of these findings to clinical contexts. Third, the metrics employed to define the timing of sulcal emergence (i.e., T10, T50, and Tapp) were derived from heterogeneous literature sources and were largely based on visual inspection standards or cross-sectional inferences. Crucially, these heuristic thresholds currently lack rigorous biological validation regarding developmental onset and have not been verified through individual-level longitudinal tracking. Consequently, they may represent statistical approximations rather than precise anatomical milestones, highlighting the need for future studies to establish more physiologically grounded standards. Finally, the gene enrichment analysis relied on the adult AHBA atlas. Although the spatial topology of cortical gene expression is relatively conserved*

	postnatally, developmental-specific dynamic expression patterns may be overlooked. Future studies should integrate pediatric transcriptomic data and Diffusion Tensor Imaging (DTI) to further validate the direct links between sulcal morphological remodeling, white matter microstructure, and specific gene expression profiles.

In summary, this is an thought-provoking, multifaceted study on a relatively understudied topic that presents fairly convincing evidence of age-related changes in sulcation metrics in youth, as well as interesting results in several other related domains. I think the study would benefit from more methodological clarity to facilitate future replication, as well as a more explicit descriptions of the inherent limitations to the complex analytic approaches used.

Minor points:

6.Please Confirm/explicitly state that the scan was MP-RAGE.	**Response:** We confirm that the T1-weighted structural images were acquired using an MP-RAGE sequence. We have explicitly stated this in the "*Image Acquisition*" subsection of the Methods to ensure clarity . **Revisions in Manuscript:** [Page 12, Lines 470] *…For each subject, T1-weighted structural images were acquired using a MPRAGE sequence with the following parameters…*
7.Please describe the rationale for why some sulcal measures were TIV-controlled and others were not.	**Response:** We appreciate the opportunity to clarify the specific definitions of our metrics and the rationale for our TIV correction strategy. 1. Definitions of Morphological Metrics: First, as detailed in our methods, we computed six metrics in native space to capture distinct aspects of sulcal morphology: ✓ Cortical Thickness (CT): The distance from the gray/white matter interface to the brain surface. ✓ Sulcal Width (SW): A proxy measure representing the ratio between CSF volume and surface area. ✓ Surface Area (SA): The calculated mesh area of the sulcal regions. ✓ Maximum Depth (maxD): The distance from the deepest point within the sulcus to the brain surface. ✓ Mean Depth (meanD): The average distance from all points within the sulcus to the surface. ✓ Sulcus Length (SL): The length of the junction line between the sulcus and the external brain hull. 2. Rationale for TIV Correction: Our decision to control for Total Intracranial Volume (TIV) was driven by a correlation analysis aimed at distinguishing global brain growth from specific regional changes. We observed two distinct patterns in our dataset: ✓ Metrics Independent of TIV (Raw Measures): Consistent with the view that local microstructural or shape features are biologically distinct from overall head size, CT and SW exhibited negligible correlations with TIV (absolute mean $r < 0.1$). Consequently, we analyzed the raw measures for CT and SW to preserve their biological interpretability without introducing noise from unnecessary correction. ✓ Metrics Dependent on TIV (Corrected Measures): In contrast, "size-related" metrics (SA and SL) showed moderate-to-strong correlations with TIV (mean $\|r\| > 0.25$). Furthermore, while depth metrics (maxD and meanD) are often categorized as shape features, our data revealed they also scaled significantly with brain size in this developmental cohort (mean $\|r\| \approx 0.19 - 0.24$). To isolate specific

	morphological maturation from global expansion, we utilized TIV-adjusted residuals for SA, SL, maxD, and meanD.
8.Much of the presented results in the section "Morphological Changes of Each Sulcus Throughout the Brain" listing structures and their associated R2 and p-values could potentially be better presented in a table, including nonsignificant results. Some other data that aren't presented in detail (i.e. sex effects) could also be included in a table.	**Response:** We agree that the original text was overly dense with statistical values, which made it difficult to compare patterns across the brain. In accordance with your recommendation, we have made the following major revisions: 1. Streamlined Text: We have removed the exhaustive lists of R^2 and p-values from the main text in the "Morphological Changes of Each Sulcus Throughout the Brain" section. The text now focuses on describing the broad developmental patterns and highlighting key findings. 2. New Table 1 (Age Effects): We have synthesized the developmental results into a comprehensive matrix-style table (Table 1). As requested, this table includes all sulcal structures and features — reporting the effect size ($\Delta adjR^2$) for each. To ensure completeness, non-significant associations are also listed (indicated by grey color), allowing readers to see where effects are absent. Significant effects surviving FDR correction are clearly marked with asterisks. 3. New Supplementary Table S1 (Sex Effects): We have aggregated the significant sex effect statistics into a table (Supplementary Table S1). This refined analysis identified 9 significant sex effects, clearly outlining the direction and significance of dimorphism across sulcal features. We believe these changes have significantly improved the readability of the manuscript and provided a more systematic presentation of the data. **Table 1. Age-Related Developmental Changes in Sulcal Morphometry.** **Structure** | **SA (adj)** | **maxD (adj)** | **meanD (adj)** | **SL (adj)** | **CT** | **SW** F.C.L._left | -0.045* | 0.002 | -0.005 | 0.001 | -0.072* | 0.058* F.C.L._right | -0.040* | 0.001 | -0.001 | -0.012 | -0.034 | 0.018 F.C.M.ant._left | -0.013* | -0.015 | -0.028 | 0.003 | -0.037* | 0.023 F.C.M.ant._right | -0.004 | -0.017 | -0.003 | 0.002 | -0.029* | -3.27e-4 F.C.M.post._left | -0.004 | -0.012* | -0.012* | -0.002 | -0.027* | 0.012 F.C.M.post._right | 0.001 | -0.019* | -0.019* | -0.002 | -0.047* | 0.030* F.Cal.ant.-Sc.Cal._left | -0.027* | -0.014 | -0.01 | -0.013* | -0.059* | 0.04 F.Cal.ant.-Sc.Cal._right | -0.002* | -0.009 | -0.018* | -0.003 | -0.024 | 0.017 F.Coll._left | -0.017* | -0.035* | -0.024 | 0.002 | -0.030* | 0.013 F.Coll._right | -0.015* | -0.032* | -0.031* | -1.37e-4 | -0.03 | 0.01 F.I.P._left | -0.041* | -0.023* | 0.005 | -0.030* | -0.091* | 0.030* F.I.P._right | -0.013* | -0.017* | -0.01 | 0.001 | -0.071* | 0.033* F.I.P.r.int.2_left | 0.003 | -0.002 | -0.002 | 0.002 | -0.025 | 0.004 F.P.O._left | -0.013* | -0.016 | -0.012 | -0.002 | -0.081* | 0.024* F.P.O._right | -0.025* | -0.006* | -0.004 | 0.002 | -0.099* | 0.041* OCCIPITAL_left | -0.002 | 0.003 | 0.004 | 0.001 | -0.034* | 0.022* OCCIPITAL_right | -0.009 | -0.009 | -0.005 | 0.001 | -0.032 | 0.018 S.C._left | -0.008* | -0.011 | 0.002 | 0.001 | -0.01 | 0.006

S.C._right	-0.013*	-0.012*	0.005	-0.003	-0.013	0.007
S.Call._left	-0.005*	-0.013	-0.02	0.006	-0.018	0.01
S.Call._right	0.001	0.004	0.003	-0.001	-0.024	-0.001
S.Cu._left	-0.010*	-0.011	-0.007	-0.002	-0.058*	0.022
S.Cu._right	-0.015*	0.001	0.003	-0.012*	-0.052*	0.03
S.F.inf._left	-0.004	0.005	0.003	-0.002	-0.069*	0.033
S.F.inf._right	-0.004	-0.002	0.001	2e-4	-0.051*	0.054*
S.F.int._left	-0.006	-0.003	-0.003	-0.004	-0.005	0.01
S.F.int._right	-0.008	-0.005	0.002	-0.01	-0.027	0.005
S.F.inter._left	-0.001	-0.001*	-6.04e-5	0.003	-0.036*	0.015
S.F.inter._right	0.008	0.003	-0.008	0.004	-0.054*	0.016*
S.F.orbitaire._left	3.24e-4	-0.001	-0.011	0.003	-0.011	0.033
S.F.sup._left	-0.012*	-0.031*	-0.009	-0.006	-0.013	0.011
S.F.sup._right	-0.002*	-0.001	-0.003	0.005	-0.018*	0.024*
S.Li.ant._left	-0.003	-0.002	-0.003	-0.002	-0.026	0.005
S.Li.ant._right	0.002	0.002	0.002	0.002	-0.025	0.015
S.Li.post._left	-0.007	-0.006	-0.019	0.002	-0.008	0.015
S.Li.post._right	-0.009	-0.021	-0.005	0.002	-0.036*	0.009
S.O.T.lat._left	-0.019*	-0.012*	-0.011*	-0.005	-0.025*	0.037*
S.O.T.lat._right	-0.011*	-3.29e-5	-0.008	-0.019*	-0.051*	-0.004
S.Olf._left	-0.005	-0.027	-0.036*	0.001	-0.083*	0.053*
S.Olf._right	-0.028*	-0.045*	-0.040*	1.72e-4	-0.073*	0.080*
S.Or._left	-0.023*	-0.007	-0.016	-0.002	-0.048*	0.011
S.Or._right	-0.037*	-0.023*	-0.015*	-0.009	-0.040*	0.031*
S.Pa.int._left	-0.009*	1.12e-4	0.015*	-0.003	-0.081*	0.026*
S.Pa.int._right	-0.040*	-0.017*	-0.007*	-0.031*	-0.057*	0.022*
S.Pa.t._right	-0.004	-0.004	-0.005	-0.01	-0.023	-1.63e-4
S.Pe.C._left	-0.017*	0.001*	-0.01	-0.012*	-0.001	0.005
S.Pe.C._right	0.003*	0.018*	0.012	-0.002	-0.009	0.007
S.Po.C._left	-0.009*	-0.016*	-0.003	0.001	-0.057*	0.018
S.Po.C._right	0.001	0.004	-0.002	3.14e-5	-0.047*	0.022*
S.T.i._left	-0.008*	-0.002	3.71e-4	-0.006	-0.053*	0.114*
S.T.i._right	-0.009*	0.007	0.015	-0.011*	-0.047*	0.067*
S.T.s._left	0.001	0.002	-0.003	-0.006	-0.065*	0.037*
S.T.s._right	-0.004*	3.91e-4	-0.001	-0.006	-0.039*	0.034*

Note: Values represent the change in adjusted R^2 derived from Generalized Additive Mixed Models (GAMMs). To isolate specific regional morphometric changes from global brain growth, Surface Area, Sulcal Length, Max Depth, and Mean Depth were analyzed using TIV-adjusted residuals. Cortical Thickness and Sulcal Width are presented as raw measures. * indicate significant age-related effects that survived FDR correction

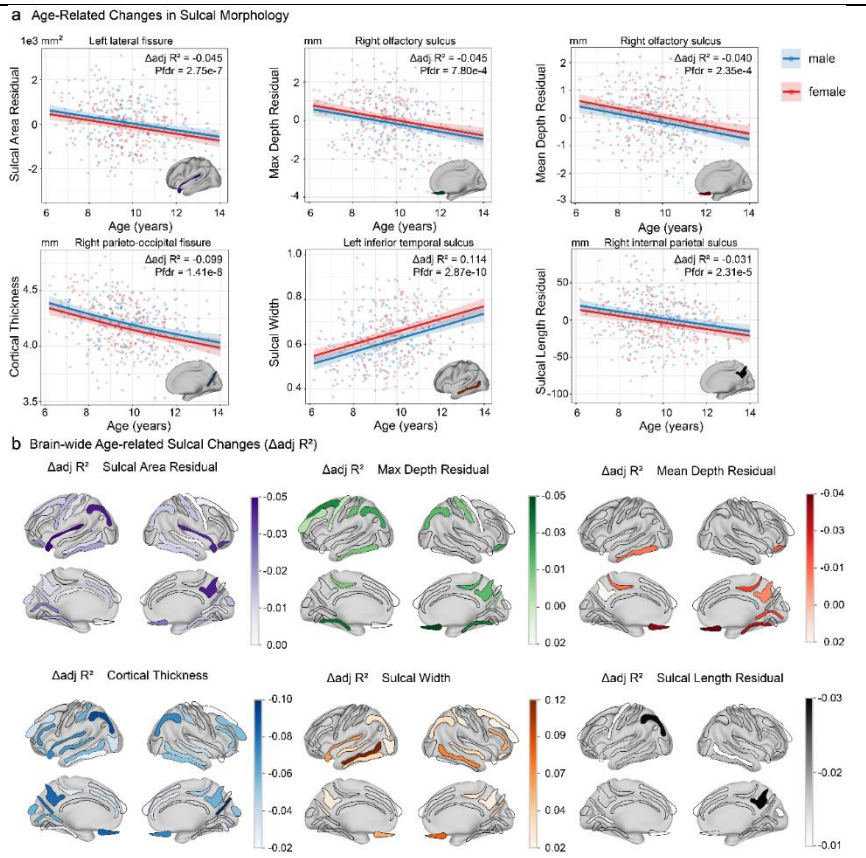
Abbreviations: adj: TIV-adjusted residuals; SA: Surface Area; maxD: Maximum Depth; meanD: Mean Depth; SL: Sulcal Length; SW: Sulcal Width; CT: Cortical Thickness.

Supplementary Table S2. Sex Differences in Sulcal Morphometry.

Label	CT	maxD (adj)	meanD (adj)	SL (adj)	SW	SA (adj)
OCCIPITAL_left	2.24	0.36	0.56	1.08	-3.71*	1.59
S.C._right	1.36	-0.74	-1.27	3.32*	-2.8	2.46
S.F.inf._left	1.89	-0.99	-1.94	1.27	-3.71*	-0.08
S.F.inf._right	0.08	-0.28	-0.4	1.29	-3.54*	0.58
S.F.inter._left	0.32	0.99	0.57	0.25	-3.42*	0.65
S.F.inter._right	2.1	0.77	0.03	0.84	-3.91*	1
S.O.T.lat._right	0.29	2.12	1.88	1.12	-3.43*	1.62
S.Pe.C._right	1.57	-0.77	0.66	0.71	-3.74*	1.45
S.T.s._right	0.94	0.17	0.01	0.13	-3.25*	0.62

Note: Values represent the T-statistics for the main effect of sex derived from the GAMMs. Positive values indicate significantly greater values in males compared to females, while negative values indicate greater values in females. Similar to the main analysis, Surface Area, Max Depth, and Mean Depth were adjusted for TIV.* indicate significant sex effects that survived FDR correction.
Abbreviations: adj: TIV-adjusted residuals; SA: Surface Area; maxD: Maximum Depth; meanD: Mean Depth; SW: Sulcal Width; CT: Cortical Thickness.

9.Page 4, line 151: I don't think I would consider R2 of -0.040 a "pronounced decrease." Given the fairly large N, I would focus more on effect sizes than p-values when interpreting results.	**Response:** We agree with the reviewer's assessment. We have revised the text to avoid subjective qualifiers like "pronounced decrease" when describing small effect sizes (e.g., $R^2 = -0.040$). In the revised manuscript, we focus on objectively reporting the effect sizes and acknowledging that in large samples, statistically significant results may still reflect subtle biological effects.
10.Figure 2a: It would be useful to mention in the figure legend that these plots are for S.O.T.lat_left specifically, and not a global measure.	**Response:** We thank the reviewer for this helpful suggestion. We have revised Figure 2a to make it more representative and comprehensive. Instead of focusing on a single sulcus, we now present the developmental trajectories of the six sulcal features (TIV-adjusted Surface Area residuals, Maximum Depth residuals, Mean Depth residuals, Sulcal Length residuals, Sulcal Width, and Cortical Thickness). For each feature, we selected and plotted the specific sulcus that exhibited the largest developmental effect (i.e., the maximum absolute $\Delta \text{adj } R^2$). The figure legend has been updated to reflect these changes and clearly state the specific sulci being visualized, ensuring it is no longer mistaken for a global measure.

	 **Fig. 2 Quantitative analysis of age-related changes in sulcal morphometry.** **(a)** Developmental trajectories of key sulcal features. The developmental characteristics for each sulcus were modeled using a generalized additive mixed model (GAMM). To represent the most prominent developmental trends, this panel displays the trajectories of the sulci showing the largest developmental effects (maximum absolute Δ adj R^2) for each of the six morphometric features (Surface Area, Maximum Depth, Mean Depth, Sulcal Length, Sulcal Width, and Cortical Thickness). The Δ adj R^2 and p-values from the GAMM are provided, with blue indicating males and red indicating females. **(b)** Regional distribution of developmental effects. Δ adj R^2 values associated with age-related changes across six features for various sulci. All sulci displayed in the figure exhibit significant age-related developmental effects ($p < 0.05$, FDR corrected $p < 0.05$).
11.Figure 3: please check to make sure that the y-label (R2) is correct since the scale is substantially different than the adjacent panel.	**Response:** We thank the reviewer for noticing this. We have verified and corrected the y-axis labels in Figure 3 to ensure they accurately reflect the scale and metric (R^2) presented, ensuring consistency with the adjacent panels.

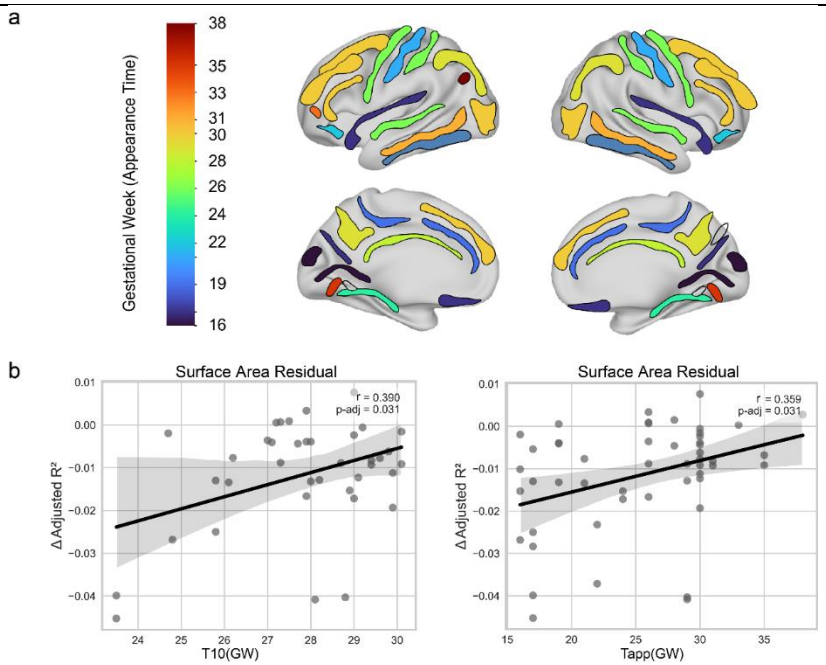


Fig. 3 Impact of sulcal emergence timing on postnatal morphometric remodeling. (a) Surface map visualization of Sulcal Appearance Time (T_{app}), defined as the gestational week when a sulcus first becomes visually identifiable based on anatomical consensus (see Methods). (b) Scatter plots illustrating the relationship between sulcal emergence (T_{app} and T_{10}) and the magnitude of postnatal development for adjusted surface area. The y-axis represents the signed Δ adj R^2, where more negative values indicate a stronger age-related reduction. The significant positive correlations indicate that earlier-emerging sulci (lower T_{app}) exhibit more profound morphometric reductions (larger negative effect sizes) and higher statistical significance during childhood and adolescence.

12.When presenting beta-weights in the text, it is difficult for the reader to get a sense of effect size, since the scale of these measures (e.g. mm) is not explicitly defined in the manuscript (although scales can be inferred based on the figure labels). Consider reporting standardized beta weights instead in order to improve readability. Also, the number of significant figures should be consistent throughout the manuscript.

Response:

We thank the reviewer for this helpful suggestion to improve readability and interpretability.

1. Standardization of Beta Weights: We agree that reporting raw beta weights makes it difficult to compare effect sizes across metrics with different physical units (e.g., cortical thickness in mm vs. surface area in mm^2). In the revised manuscript, we have standardized all brain morphology metrics and cognitive scores using z-score normalization prior to entering them into any statistical models (LASSO, PLS, or linear regression). Consequently, the coefficients (β) reported in the text and figures now represent standardized beta weights. This allows readers to directly interpret the strength of associations in standard deviation units.
2. Consistency of Significant Figures: We have carefully proofread the entire manuscript and standardized the reporting of significant figures. We now consistently report correlation coefficients (r) and beta weights (β) to two decimal places, Δ adj R^2 and p-values to three decimal places (or in scientific notation for very small values), ensuring uniformity throughout the text and tables.

Revisions in Manuscript:

[Page 15, Lines 609]

In the *"Statistical analysis of brain – behavior associations"* section in Method, we added the statement: *"…Prior to analysis, all morphometric and cognitive metrics were standardized (Z-score) to facilitate cross-scale comparisons…"*

-In general, the visibility of many axis labels could be improved, primarily by using larger fonts for numeric values. The number of significant figures also is inconsistent from panel to panel.	We have improved the visibility of all figures. Specifically, we have: 1. Increased the font size for axis labels and numeric values to ensure they are legible. 2. Standardized the number of significant figures across all panels for consistency.
Typographical Errors: -Page 4, line 145, ‘ ” ‘ . -Page 9, line 332, extra space “(exclude”	We apologize for these oversights. We have corrected the typographical error on Page 4 (line 145) and removed the extra “(” on Page 9 (line 332). We have also thoroughly proofread the manuscript to eliminate similar errors.

Reviewer #2 (Remarks to the Author):

Brief summary of the manuscript:

The authors analyze longitudinal MRI from 312 children (6–14 years; 490 scans) to quantify developmental changes in sulcal morphometry (surface area, depths, width, length, cortical thickness) across several whole brain sulci and relate these to cognition. They report systematic age-related trajectories (e.g., thinning and sulcal widening), a link between fetal emergence timing and sulcal area change, and associations between changes in specific sulci and executive control/working memory performance. A transcriptomic enrichment analysis further implicates synaptic and cytoskeletal processes in sulcal development. Overall, the study offers a comprehensive, multimodal view of how sulcal morphometry relates to cognitive maturation.

Overall impression of the work:

This is a solid, carefully executed study with notable strengths: a large developmental cohort, rigorous longitudinal modeling, appropriate TIV adjustment, and a thoughtful integration of morphometric, cognitive, and gene-expression analyses. I particularly appreciated the use of LASSO to link longitudinal sulcal change to cognition and the explicit treatment of brain-size correction, which is essential in developmental work. Beyond documenting trajectories, the joint morphometry–cognition–gene framework points to plausible biological pathways and functional consequences, advancing a field that has long sought mechanistic accounts of cortical folding. If adopted widely, the authors' modeling choices and systematic TIV adjustment could help set more rigorous standards for developmental morphometry. With clearer terminology (morphology vs. morphometry), resolution of a few interpretational inconsistencies, explicit reporting of sex effects, and a more integrative Discussion that connects findings to mechanisms and future directions, the manuscript will be substantially strengthened and broadly informative to the community.

MAJOR POINTS: :

1. Use the term "morphometry" consistently (not "morphology") for quantitative measures. Throughout the paper, when referring to quantitative indices (sulcal area, depth, cortical thickness, length, width), use morphometry/morphometric rather than morphology/morphological. The (qualitative) folding pattern is longitudinally stable after birth. You already hint at this distinction (e.g., "morphometric features" in line 229); please adopt consistent terminology across the manuscript and add a brief sentence in the Introduction clarifying morphology vs. morphometry, with appropriate citations you will find in: Cachia, A. et al. Longitudinal stability of the folding pattern of the anterior cingulate cortex during development. Dev. Cogn. Neurosci. 19, 122–127 (2016).	**Response:** We sincerely thank the reviewer for this insightful comment regarding terminology. We agree that distinguishing between qualitative "morphology" (folding patterns) and quantitative "morphometry" is crucial for precision, especially given the evidence that folding patterns remain stable postnatally while quantitative dimensions continue to change. In accordance with your suggestion, we have: ✓ Revised the terminology throughout the manuscript: We have systematically replaced "morphology/morphological" with "morphometry/morphometric" whenever referring to quantitative indices such as sulcal area, depth, thickness, length, and width. This change has been applied to the Title, Abstract, Introduction, Results, and Discussion sections. ✓ Clarified the distinction in the Introduction: We have inserted a statement in the Introduction to explicitly define these terms and acknowledge the longitudinal stability of folding patterns, citing Cachia et al. (2016). **Revisions in Manuscript:** [Page 2, Lines 59] *" While qualitative sulcal patterns (morphology) remain longitudinally stable from childhood to adulthood (Cachia et al., 2016), sulcal morphometry (quantitative dimensions) undergoes dynamic remodeling alongside micro- and macrostructural organization across the postnatal lifespan."*
2. Direction of the "emergence timing" effect needs to be clarified and aligned across Results, Fig. 3 caption, and Discussion. In the Results (lines 185–	**Response:** We thank the reviewer for identifying this crucial ambiguity. We acknowledge that the phrasing in the original manuscript led to a seeming contradiction

186) you say: “These findings suggest that earlier emergence generally associated with stronger age-related effects in surface area measures.”, in the Fig.3 caption you say: “Earlier-emerging sulci demonstrate more pronounced developmental changes in sulcal area during childhood and adolescence.” And also in the discussion (lines 276-277): “We found that sulci appearing later in fetal development exhibited less pronounced changes in surface area from childhood to adolescence.”. Indeed, the discussion (lines 277-280) ends like this: “This finding aligns with the hypothesis that earlier-developing sulci are subject to more robust genetic regulation throughout the lifespan, possibly due to the brain’s prioritization of regions vital for essential functons” which seems like a fair way to argue, as more robust genetic expression supports the result of few changes during postnatal development in the early-forming sulci. The correct direction of the effect and its correct interpretation should be clarified at the indicated points.	between the Results ("earlier emergence = stronger age-related effects") and the Discussion ("earlier emergence = robust genetic regulation"). We have now clarified the text across the Results and Discussion to ensure a consistent interpretation. Specifically, we have refined the Discussion to explain why robust genetic regulation leads to stronger statistical age-effects (i.e., clearer developmental trends) in our group-level analysis. The logic is based on the trade-off between developmental consistency and inter-individual variability: ✓ Canalization and Shared Functional Demands (Early Sulci): Early-emerging sulci (primary folds) are not only under stronger genetic control (Waddington et al., 1942; Lohmann et al., 1999) but also anchor primary sensorimotor regions that undergo universal functional maturation. During childhood, all individuals experience shared developmental milestones (e.g., refinement of basic motor skills, speech articulation). These shared, ubiquitous environmental demands likely drive consistent structural remodeling across the entire cohort. This high uniformity—driven by both genetic blueprints and shared functional experiences—results in a high signal-to-noise ratio and robust statistical age effects. ✓ Plasticity and Inter-individual Variability (Late Sulci): In contrast, late-emerging regions (often corresponding to high-expansion cortical zones) are evolutionarily more recent and functionally related to higher-order cognition. Hill et al. (2010) showed that these regions undergo the greatest postnatal expansion and are highly sensitive to environmental inputs and individual experiences. Consequently, the developmental remodeling of late-emerging sulci is driven by unique, experience-dependent factors, leading to high inter-individual variability (or "noise"). This heterogeneity in individual trajectories likely dilutes the group-level mean trend, resulting in the "weaker" statistical associations we observed. Therefore, the "stronger effect" in early sulci reflects a shared, genetically programmed developmental milestone, whereas the "weaker effect" in late sulci reflects divergent, experience-dependent plasticity. **Revisions in Manuscript:** Results Section [Page 5, Lines 188] *“These findings suggest a chronological gradient where the earliest-forming primary sulci undergo the most profound and consistent morphological remodeling during childhood, whereas later-emerging sulci exhibit more attenuated group-level changes.”* Discussion Section [Page 9, Lines 326] *...This finding is explained by developmental "Canalization", a mechanism that buffers essential phenotypes against variation. Early sulci are governed by strong genetic programs and subserve fundamental functions with shared developmental milestones. This commonality minimizes within-group variance, yielding high signal-to-noise ratios in population analyses. By contrast, later-developing association sulci show less significant group-level changes...* **References** 1. Van Essen, D. C. A tension-based theory of morphogenesis and compact wiring in the central nervous system. Nature 385, 313 – 318 (1997). 2. Van Essen, D. C. A 2020 view of tension-based cortical morphogenesis. Proc. Natl. Acad. Sci. 117, 32868 – 32879 (2020). 3. Hill, J. et al. Similar patterns of cortical expansion during human development and evolution. Proc. Natl. Acad. Sci. U. S. A. 107, 13135 – 13140 (2010). 4. Lohmann, G., von Cramon, D. Y. & Steinmetz, H. Sulcal variability of twins. Cereb Cortex 9, 754 – 763 (1999). 5. Waddington, C. H. Canalization of development and the inheritance of acquired characters. Nature 150, 563 – 565 (1942).

3. Report and discuss sex effects explicitly. You note robust sex main effects across 214 models (lines 167–169). Please (i) present a concise table (main text or Supplement) summarizing which features/sulci showed sex effects and effect directions, and (ii) add a short paragraph in the Discussion interpreting these findings in light of developmental literature.	**Response:** We thank the reviewer for this constructive suggestion. We agree that explicitly reporting sex effects significantly enhances the interpretability of our developmental models. To distinguish specific sexual dimorphism from global brain size effects, we performed a refined analysis focusing on Cortical Thickness, Sulcal Width, and TIV-adjusted measures for Surface Area and Mean Depth. **This analysis identified 9 significant sex effects models. This lower count compared to previous findings is due to the exclusion of models such as surface area without adjustment for Total Intracranial Volume (TIV).** In response to your specific suggestions: (i) Summary of Sex Effects: We have added a new Supplementary Table 2, which explicitly lists the t-statistics and directions of sex effects for all sulcal features. We found significant sex effects of sulcal width in 22 out of 53 sulci, all showing the pattern Males < Females (narrower sulci in males). This finding aligns with established allometric principles. As demonstrated by Toro et al. (2008), there is a strong positive correlation between brain size and the Gyrification Index (GI). Since males typically possess larger brain volumes (approx. 10 – 15% larger), they exhibit a higher degree of cortical folding complexity. This increased packing density mechanically results in narrower sulci. Crucially, after adjusting for global scaling (TIV), sex differences largely disappeared. Significant effects were found in only 1/53 regions for Cortical Thickness, 3/53 for Adjusted Surface Area, and 6/53 for Adjusted Mean Depth. (ii) Revisions in Discussion: We have added a new paragraph in the Discussion to interpret these findings. We cite Jäncke (2014) to support the consensus that most sex differences in brain morphology are statistically explained by the global difference in brain size, rather than specific sexual dimorphism. **Revisions in Manuscript:** Discussion Section [Page 8, Lines 320] *"…observed sex differences (deeper sulci in males vs. wider in females) largely vanished after correcting for total intracranial volume (TIV). This supports the "Allometry" hypothesis65, indicating that sexual dimorphism arises from physical constraints —where larger brains require more compact folding — rather than divergent neurodevelopmental programs."* **Supplementary Table S2. Sex Differences in Sulcal Morphometry.** | Label	CT	maxD (adj)	meanD (adj)	SL (adj)	SW	SA (adj)
---	---	---	---	---	---	---
OCCIPITAL_left	2.24	0.36	0.56	1.08	-3.71*	1.59
S.C._right	1.36	-0.74	-1.27	3.32*	-2.8	2.46
S.F.inf._left	1.89	-0.99	-1.94	1.27	-3.71*	-0.08
S.F.inf._right	0.08	-0.28	-0.4	1.29	-3.54*	0.58
S.F.inter._left	0.32	0.99	0.57	0.25	-3.42*	0.65
S.F.inter._right	2.1	0.77	0.03	0.84	-3.91*	1
S.O.T.lat._right	0.29	2.12	1.88	1.12	-3.43*	1.62
S.Pe.C._right	1.57	-0.77	0.66	0.71	-3.74*	1.45
S.T.s._right	0.94	0.17	0.01	0.13	-3.25*	0.62	 Note: Values represent the T-statistics for the main effect of sex derived from the GAMMs. Positive values indicate significantly greater values in males compared to females, while negative values indicate greater values in females. Similar to the main analysis, Surface Area, Max Depth, and Mean Depth were adjusted for TIV.* indicate significant sex effects that survived FDR correction. Abbreviations: adj: TIV-adjusted residuals; SA: Surface Area; maxD: Maximum Depth; meanD: Mean Depth; SW: Sulcal Width; CT: Cortical Thickness. **References:** 1. Sotiras, A. et al. Patterns of coordinated cortical remodeling during adolescence and their associations with functional specialization and

	evolutionary expansion. Proceedings of the National Academy of Sciences 114, 3527 – 3532 (2017). 2. Toro, R. et al. Brain size and folding of the human cerebral cortex. Cereb Cortex 18, 2352 – 2357 (2008). 3. Eliot, L., Ahmed, A., Khan, H. & Patel, J. Dump the 'dimorphism': Comprehensive synthesis of human brain studies reveals few male-female differences beyond size. Neurosci Biobehav Rev 125, 667 – 697 (2021). 4. Jäncke, L., Mérillat, S., Liem, F. & Hänggi, J. Brain size, sex, and the aging brain. Hum Brain Mapp 36, 150 – 169 (2015).
4. Expand the Discussion on the sulcus of Jensen (secondary intermediate ramus of IPS) and cognition. In lines 285–287, you link this region to executive control within the multiple-demand system; please strengthen this with targeted sulcal-morphology/cognition literature on IPS, e.g.: Santacroce F, Cachia A, Fragueiro A, Grande E, Roell M, Baldassarre A, Sestieri C, Committeri G. Human intraparietal sulcal morphology relates to individual differences in language and memory performance. Commun Biol. 2024 May 2;7(1):520. doi: 10.1038/s42003-024-06175-9. PMID: 38698168; PMCID: PMC11065983. E Roell, M. et al. Sulcation of the intraparietal sulcus is related to symbolic but not non-symbolic number skills. Dev. Cogn. Neurosci. 51, 100998 (2021).	**Response:** We are grateful to the reviewer for this insightful comment and for directing us to the highly relevant work by Roell et al. (2021) and Santacroce et al. (2024). These references have been instrumental in strengthening the theoretical framework of our discussion regarding the intraparietal sulcus (IPS) and its ramifications. In the revised Discussion we have expanded our interpretation to integrate these findings. Specifically: ✓ Roell et al. (2021) demonstrated that the sulcation patterns of the IPS are specifically related to symbolic number skills, suggesting that structural differentiation in this region reflects the specialization of cortical networks for abstract processing. ✓ Santacroce et al. (2024) further extended this by linking IPS morphology to individual differences in language and memory, reinforcing the role of parietal sulcal variability in high-level cognitive domains. These studies provide crucial empirical support for our hypothesis. They suggest that the morphological variability of the IPS (including its secondary intermediate ramus, the Jensen sulcus) is not random but functionally significant. This aligns perfectly with our finding that the widening of the Jensen sulcus facilitates conflict resolution. We argue that this structural feature represents an enhanced "anatomical segregation" within the multiple-demand system — creating a clearer physical boundary between the Supramarginal Gyrus and the Angular Gyrus — thereby reducing neural interference and optimizing executive control efficiency. By incorporating these citations, we have firmly anchored our findings regarding the Jensen sulcus within the broader context of IPS structure-function relationships. **Revisions in Manuscript:** Discussion section [Page 10, Lines 386] *"Within the Executive Control Network, the widening of the left F.I.P.r.int.2 (corresponding to the Jensen Sulcus) emerged as the most robust predictor of improvements in conflict resolution. This region resides within the Multiple-demand network, a system critical for complex task processing and problem-solving. This functional relevance is supported by recent evidence showing that the sulcal morphometry of the IPS is significantly associated with individual differences in cognitive domains, including symbolic number skills and language and memory performance. The Jensen Sulcus serves as a key anatomical landmark delineating the Supramarginal Gyrus (SMG) from the Angular Gyrus (AG). While acknowledging that this sulcus exhibits inter-individual variability, cytoarchitectonic evidence indicates that when morphologically present, it marks the precise functional transition within its depth. Consequently, we postulate that the widening of this sulcus signifies an enhanced "anatomical segregation" between the SMG and AG. This structural adaptation aligns with our transcriptomic findings: the widening likely results from a localized reduction in synaptic tension, which allows the sulcal walls to relax and separate. While the Superior Longitudinal Fasciculus may exert the necessary mechanical constraint to maintain the fold's deep integrity, the superficial*

	widening acts as a physical barrier to reduce neural interference between adjacent functional areas." **References:** 1. Roell, M. et al. Sulcation of the intraparietal sulcus is related to symbolic but not non-symbolic number skills. Developmental Cognitive Neuroscience 51, 100998 (2021). 2. Santacroce, F. et al. Human intraparietal sulcal morphology relates to individual differences in language and memory performance. Commun Biol 7, 520 (2024). 3. Van Essen, D. C. A 2020 view of tension-based cortical morphogenesis. Proc. Natl. Acad. Sci. 117, 32868 – 32879 (2020).
5. Speculate more concretely about candidate white-matter pathways (lines 293–295). Where you discuss structure–function coupling, consider hypothesizing tract involvement consistent with the implicated sulci: e.g., fronto-parietal SLF II/III for IPS–prefrontal links (executive control), ILF/VOF for posterior intralingual/occipito-temporal contributions (visual memory/working memory interfaces), and fronto-orbital connections (e.g., anterior thalamic radiations/frontostriatal pathways) for orbital sulcus and alerting. These are framed as plausible mechanisms guiding future work rather than definitive claims.	**Response:** We thank the reviewer for this insightful suggestion. We agree that explicitly discussing candidate white-matter pathways adds significant mechanistic depth to our structure-function interpretations. Although our study focuses on cortical surface morphology, we acknowledge that these folding patterns are intrinsically linked to the underlying connectivity. In the revised Discussion, we have integrated the speculation regarding specific fiber tracts directly into the interpretation of the observed sulcal-behavior associations. We frame these underlying structural connections as plausible mechanistic hypotheses to guide future tractography studies: <u>1. Visual Working Memory and the Inferior Longitudinal Fasciculus (ILF)</u> For the left hemisphere, our LASSO results showed a "node-specificity," where specific posterior markers (e.g., the posterior intralingual sulcus/posterior lingual gyrus) acted as decisive predictors—a "bottleneck" effect. Anatomically, the posterior lingual gyrus serves as a critical cortical anchor for the Inferior Longitudinal Fasciculus (ILF) (Latini et al., 2017; Panesar et al., 2018). The ILF is the core pathway connecting the occipital visual cortex to the anterior temporal memory systems, supporting the rapid transfer and short-term storage of visual representations (Herbet et al., 2018). We speculate that the left hemisphere adopts a "precision pruning" strategy: morphological optimization in these key "gating" sulci likely reflects the maturation of ILF terminals, enhancing the efficiency of visual input entry into memory buffers. Conversely, the right hemisphere showed a broader pattern of structural reorganization involving the transverse parietal (S.Pa.t) and inferior frontal (S.F.inf) sulci, suggesting a synergistic redistribution of resources within the dorsal and ventral attention networks, likely supported by long-range fronto-parietal connections. <u>2. Executive Control and the Superior Longitudinal Fasciculus (SLF)</u> Regarding the executive control network, the widening of the left secondary intermediate ramus of the IPS (sulcus of Jensen) was the strongest predictor of conflict resolution. Linking this to the "tension-based hypothesis" (Van Essen, 2020), we propose that this specific sulcal remodeling is driven by the mechanical tension of the Superior Longitudinal Fasciculus (SLF II/III). As the core skeleton of the fronto-parietal network, the maturation of SLF fibers (e.g., increased axon diameter or myelination) would exert altered tension on the inferior parietal lobule, leading to the "opening up" of the Jensen sulcus. This morphological change physically segregates the supramarginal gyrus from the angular gyrus, potentially minimizing neural interference and optimizing top-down control signals from the prefrontal cortex. <u>3. Alerting and the Anterior Thalamic Radiations (ATR)</u> Finally, the efficiency of the alerting network was primarily linked to the expansion of the left orbital frontal sulcus. The maintenance of tonic alertness relies heavily on thalamo-cortical arousal. We hypothesize that the observed surface area expansion in the orbitofrontal cortex (OFC) relates to the Anterior Thalamic Radiations (ATR), which connect the dorsomedial thalamus to the prefrontal cortex (Sadaghiani & D'Esposito, 2015). An increase in cortical

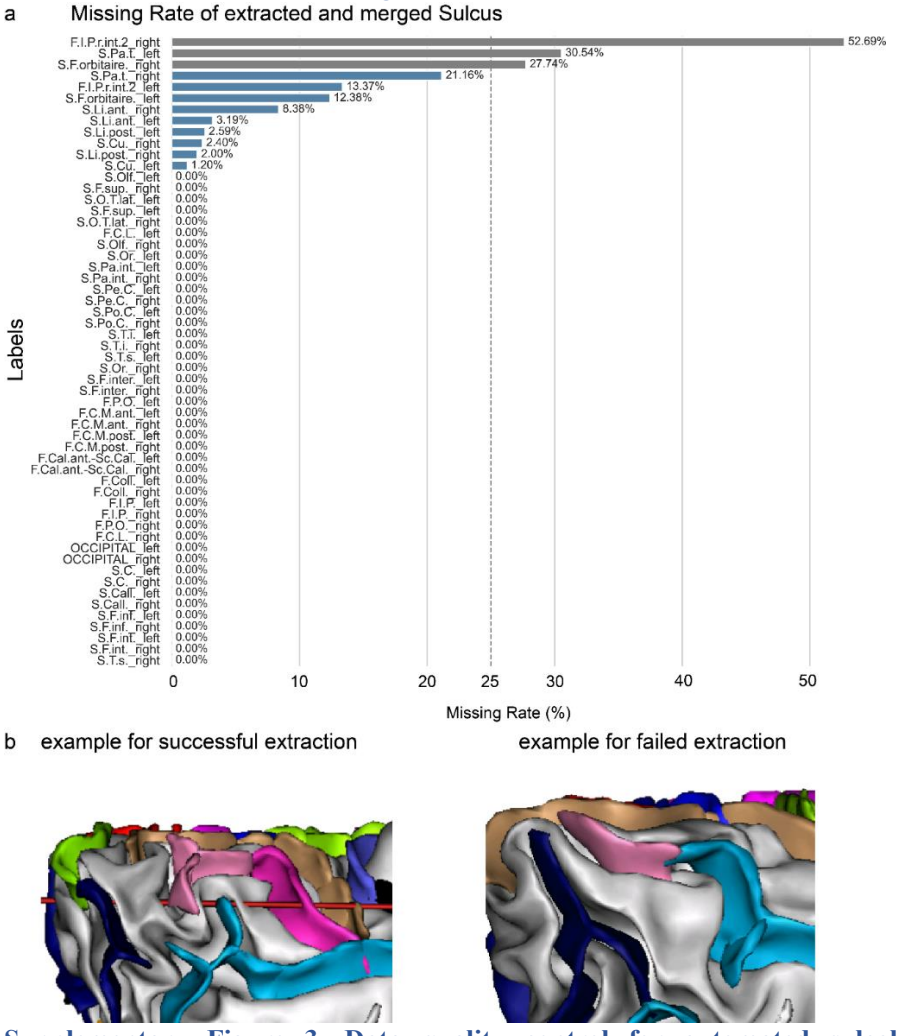
	surface area typically reflects an increase in the number of discrete processing units (cortical minicolumns). Therefore, the expansion of the orbital frontal sulcus may represent an increased capacity to receive and process ascending arousal signals from the thalamus via the ATR, thereby supporting sustained attention. **Revisions in Manuscript:** Discussion section [Page 10, Lines 368] (Regarding ILF and Visual Working Memory): *"Furthermore, the posterior lingual gyrus serves as a critical cortical anchor for the Inferior Longitudinal Fasciculus (ILF). Given that the ILF is a core pathway connecting the occipital visual cortex to the anterior temporal memory system, supporting visual object recognition and short-term storage, we postulate that the morphometric remodeling (e.g., widening) of these specific ILF-anchored nodes may facilitate more efficient neural transmission of visual representations into memory systems."* (Regarding SLF and Jensen Sulcus): *"This structural adaptation aligns with our transcriptomic findings: the widening likely results from a localized reduction in synaptic tension, which allows the sulcal walls to relax and separate. While the Superior Longitudinal Fasciculus may exert the necessary mechanical constraint to maintain the fold's deep integrity, the superficial widening acts as a physical barrier to reduce neural interference between adjacent functional areas."* (Regarding ATR and Alerting Network): *"Gains in Alerting network efficiency were primarily associated with adjusted surface area expansion in the left orbital frontal sulcus. As a critical node within the Salience Network, the expansion of the orbital cortex likely corresponds to an enlargement of the cortical projection zones of the Anterior Thalamic Radiation. This structural change may reflect an enhanced processing capacity for thalamic arousal signals within this region."* **References** 1. Sadaghiani, S. & D'Esposito, M. Functional Characterization of the Cingulo-Opercular Network in the Maintenance of Tonic Alertness. Cereb Cortex 25, 2763–2773 (2015). 2. Panesar, S. S., Yeh, F.-C., Jacquesson, T., Hula, W. & Fernandez-Miranda, J. C. A quantitative tractography study into the connectivity, segmentation and laterality of the human inferior longitudinal fasciculus. Front Neuroanat 12, 47 (2018). 3. Latini, F. et al. Segmentation of the inferior longitudinal fasciculus in the human brain: A white matter dissection and diffusion tensor tractography study. Brain Research 1675, 102–115 (2017). 4. Herbet, G., Zemmoura, I. & Duffau, H. Functional anatomy of the inferior longitudinal fasciculus: From historical reports to current hypotheses. Front Neuroanat 12, 77 (2018). 5. Van Essen, D. C. A 2020 view of tension-based cortical morphogenesis. Proc. Natl. Acad. Sci. 117, 32868–32879 (2020).
6. In general, broaden and deepen the Discussion across key findings. Several sections currently read as concise lists. For example, lines 300–302 propose that wider sulci may reflect neural reorganization underlying better performance; please add mechanistic justifications and insert citations consistent with your enrichment	**Response:** We sincerely thank the reviewer for this constructive suggestion to strengthen our Discussion. We agree that providing deeper mechanistic interpretations and forward-looking implications is essential to contextualize our findings. In the revised Discussion, we have expanded the analysis of our three core findings (global fine-tuning, temporal principle/canalization, and gene-cognition associations) following the reviewer's recommended structure:

results to strengthen these claims. Apply the same treatment to the other principal results so each has (i) speculation about the effects' cause, (ii) citation in ligne or in contrast with the results, and (iii) a forward-looking implication for future research/interventions.	1. For the global "fine-tuning" pattern: We have explicitly linked the observed "sulcal widening" to the "Tension-based Morphogenesis" hypothesis and our own transcriptomic findings regarding synaptic signaling enrichment. We added implications suggesting sulcal width as a potential macroscopic proxy for synaptic pruning. 2. For the "Temporal Principle": We elaborated on the interplay between genetic constraints and environmental plasticity to explain the stability of primary sulci versus the variability of association sulci. We concluded with an implication for selecting specific sulci as biomarkers for different research purposes (normative tracking vs. plasticity studies). 3. For Brain-Behavior relationships: We integrated the specific anatomical roles of the identified sulci (e.g., as anchors for white matter tracts like ILF/SLF) to explain their predictive power for Working Memory and Attention. We added implications regarding their utility in tracking cognitive maturation. These revisions transform the previous descriptions into a more comprehensive interpretation of the developmental mechanisms. Please refer to the revised Discussion section in the manuscript. **Revisions in Manuscript:** Discussion section [Page 8, Lines 309] (Regarding Global "Fine-tuning" and Synaptic Mechanisms): *Specifically, "sulcal widening" may be mechanically coupled with cortical thinning via "Tension-based Morphogenesis"1. Our transcriptomic analysis supports this mechanical hypothesis: we found that sulcal width variations are enriched for genes related to synaptic signaling. As synaptic pruning reduces gray matter volume and dendritic density, the consequent decline in local axonal tension allows the sulcal walls to relax, manifesting as widening. This process accommodates the maturation of short-range U-fibers1, marking a shift from generalized to specialized neural connectivity1. Crucially, this remodeling holds functional significance during the peak of structural-functional coupling1. We postulate that increased sulcal width reflects neural reorganization aimed at optimizing transmission efficiency, providing the biophysical basis for efficient cognitive networks. Consequently, future longitudinal studies are warranted to further investigate how macroscopic sulcal width variations reflect underlying microstructural processes in vivo.* (Regarding Temporal Principle and "Canalization"): *This finding is explained by developmental "Canalization"1, a mechanism that buffers essential phenotypes against variation. Early sulci are governed by strong genetic programs1 and subserve fundamental functions with shared developmental milestones. ... By contrast, later-developing association sulci show less significant group-level changes1, as they are highly sensitive to unique environmental experiences and cognitive abilities1. ... This dissociation implies that future research could explore the distinct roles of these regions, with primary sulci potentially reflecting conserved maturation and association sulci offering insights into neuroplasticity.* (Regarding Brain-Behavior Mechanisms and Tractography Anchors): *Furthermore, the posterior lingual gyrus serves as a critical cortical anchor for the Inferior Longitudinal Fasciculus (ILF). Given that the ILF is a core pathway connecting the occipital visual cortex to the anterior temporal memory system... we postulate that the morphometric remodeling (e.g., widening) of these specific ILF-anchored nodes may facilitate more efficient neural transmission of visual representations into memory systems.* (Regarding Executive Control and Anatomical Segregation): *Consequently, we postulate that the widening of this sulcus signifies an enhanced "anatomical segregation" between the SMG and AG. This structural*

	adaptation aligns with our transcriptomic findings: the widening likely results from a localized reduction in synaptic tension, which allows the sulcal walls to relax and separate... Future research combining functional connectivity MRI with sulcal morphology could validate whether these structural segregations directly correlate with reduced functional crosstalk between the SMG and AG.
7. Methods transparency on model counts and longitudinal change computation. - Please state explicitly the total number of GAMMs tested (e.g., N sulci × N features).	**Response:** We thank the reviewer for this suggestion regarding statistical transparency. We have now explicitly stated the total number of models tested in the Methods section under "Modeling the development of sulcal morphology." Specifically, we analyzed 53 sulci across 6 morphological features (Cortical Thickness, Sulcal Width, TIV-adjusted Surface Area, TIV-adjusted Sulcal Length, TIV-adjusted Maximum Depth, and TIV-adjusted Mean Depth), resulting in a total of 318 GAMM models (53 sulci × 6 features).
- Clarify if longitudinal change analyses (Δ metrics, LASSO) included only participants with ≥ 2 timepoints or the overall sample (312). In the second case, it needs to be better explained how subjects who have only one MRI scan are considered in longitudinal analyses.	**Response:** We appreciate the reviewer allowing us to clarify the sample composition for the different analyses. We have distinguished between the developmental trajectory modeling (GAMMs) and the cognitive-association modeling (LASSO) as follows: ✓ For the GAMM developmental trajectories: We utilized the full dataset (312 participants, 490 scans), which includes subjects with 1, 2, or 3 time points. The advantage of Mixed Models (GAMMs/LMEs) is their ability to handle unbalanced longitudinal data. They utilize subjects with a single time point to robustly estimate the intercepts (baseline population means) and variance components, while subjects with longitudinal data (≥ 2 scans) contribute to estimating the slopes (within-subject trajectories). Excluding single-timepoint data would reduce statistical power for estimating the fixed effects of age and sex. ✓ For the cognitive analysis: We explicitly restricted this analysis to participants with ≥ 2 time points. Subjects with only a single MRI scan were excluded from these longitudinal difference analyses. This is because our analysis relied on calculated difference scores (Δ metrics = Time 2 - Time 1) to link morphological changes to cognitive changes. **Revisions in Manuscript:** Methods Section[Page 15, Lines 611] *"The statistical analysis proceeded in two stages: first, we examined baseline associations between sulcal morphometry and cognitive function; second, for participants with at least two time points, we quantified the coupling between longitudinal morphometric and cognitive changes. Longitudinal change (Δ) was defined as the difference between follow-up (T2) and baseline (T1) measurements…"*

MINOR POINTS:

- Clarify Hypothesis 1 (lines 96–97). As written, it is somewhat vague.	**Response:** We have addressed all minor points as requested: Hypothesis 1: We have refined the hypothesis to be more specific regarding the directionality and nature of the expected changes. Revised Text [Page 3, Lines 99]: *"Based on this evidence, we hypothesize that: (1) sulci emerging earlier in fetal development will exhibit more profound morphological remodeling during childhood and adolescence compared to later-emerging sulci..."*
- Define better the term "Tapp" on first use.	Definition of Tapp: We have clarified the definition upon first mention.

	Revised Text [Page 5, Lines 177]: *"... T_{app} (the gestational week of visual appearance based on anatomical consensus), and T_{10}/T_{50} (the gestational ages marking the onset and midpoint of folding, derived from curvature modeling) ..."*
- Consider adding a concise Supplement for QC outcomes. You excluded sulci with <75% extraction success and removed three specific sulci (lines 383–387). A supplemental table listing success rates per sulcus and an explanatory figure showing a bad sulcal segmentation would improve reproducibility.	QC Supplement: We have a revised Supplementary Figure 3, which lists the extraction missing rates for all sulci. Furthermore, we add an additional subplot, which visualizes the difference between a successful extraction and a failed extraction of left anterior intralingual sulcus.  **Supplementary Figure 3. Data quality control for automated sulcal extraction.** **(a)** Sulcal Extraction Missing Rates. Sulcal identification was performed across the cohort; the three specific sulci depicted with grey bars were excluded from subsequent analyses due to an extraction success rate below 75% (equivalent to a missing/error rate exceeding 25%). **(b)** Representative Examples of Sulcal Extraction. Visualization of successful (left) and failed (right) sulcal identification instances for the left anterior intralingual sulcus (Deep Pink), illustrating the criteria for quality control and anatomical inclusion.

Reviewer #3 (Remarks to the Author):

Overview: Thank you for the opportunity to review this manuscript! It was very well written and the figures are clear and informative. Briefly, the authors examined a cohort of children, aged 6-14 years, at multiple time points, acquiring both brain structural and cognitive data. The authors initially examined the age-related development of several sulci; and then linked age-related change in sulci morphometry to cognitive measures and gene expression data (derived from the Allen Human Brain Atlas). While the manuscript is relevant and timely, several aspects remain to be addressed, as outlined below. This particularly includes emphasising/clarifying the rationale for your analysis at the sulcus level; and clarifying how you accounted for the different ages, inter-scan intervals, and numbers of acquired time points (especially since around 168 of your participants only had cross-sectional MRI data).

Abstract:	
1. Could the authors include the inter-scan interval in the abstract	**Response:** We thank the reviewer for this helpful suggestion to improve the clarity of our study design. We have revised the Abstract to explicitly state that the data were collected with an inter-scan interval of approximately one year. Detailed information regarding the sampling distribution has been provided in the Methods section and Supplementary Figure 1. **Revisions in Manuscript:** Abstract [Page 1, Lines 35] *"…This study analyzed structural MRI data from 312 children (6-14 years), with longitudinal data (inter-scan interval ~1 year) available for a subset. Results revealed…"*
2. Would it be possible to include statistical results	**Response:** We fully agree with the reviewer that quantitative evidence is essential for supporting our claims. We carefully considered including specific statistical values in the Abstract. However, given the scope of our whole-brain analysis—which involves 53 distinct sulci and multiple morphometric features (e.g., width, depth, thickness)—selecting a few isolated statistical parameters would not accurately reflect the global nature of the developmental patterns we observed. Additionally, listing the necessary statistics to provide a complete picture would compromise the readability and exceed the strict word count limits of the Abstract. Therefore, we have prioritized a concise conceptual summary in the Abstract, while providing comprehensive statistical details (including effect sizes and FDR-corrected p-values) in the Results section and Tables (e.g., Table 1). We hope this presentation strategy effectively balances empirical rigor with the need for a concise summary.
Introduction:	
3. It would be helpful to further elaborate on the biological/mechanistic relevance of sulcal morphological changes, including to clinical phenotypes	**Response:** We appreciate the reviewer's suggestion to strengthen the biological and clinical context of our study. In the revised Introduction, we have expanded the description of the cellular mechanisms underlying sulcal changes (e.g., radial glia proliferation, synaptic pruning) and explicitly linked sulcal morphometric deviations to clinical phenotypes to highlight their potential translational value. **Revisions in Manuscript:** Introduction [Page 3, Lines 89]: "...Cortical surface area is driven by radial glia proliferation and relates to minicolumn number and myelination processes, while thickness is influenced by intermediate progenitor cells and associated with synaptic pruning and myelination. Regional differences in cellular proliferation, coupled with axonal tension and biomechanical mechanisms, influence sulcal patterns including folding degree, depth, length, and width. Crucially, deviations in these sulcal morphometric trajectories are closely linked to atypical neural connectivity and have been identified as potential biomarkers for neurodevelopmental disorders, such as autism spectrum disorder (ASD) and schizophrenia. Since

	these processes vary temporally and spatially, changes in sulcal dimensions throughout development should reflect corresponding gene expression patterns..."
4. Line 76: this sentence is a bit confusing; can you clarify which sulcus/gyrus would relate to eg fine motor skills, reasoning/memory etc.?	**Response:** We thank the reviewer for this insightful comment. We agree that the specific associations between sulcal morphology and cognitive domains were not clearly specified in the original text. We have revised the sentence to explicitly link specific cognitive functions to their corresponding sulcal landmarks based on recent literature. Specifically, we now mention the paracingulate sulcus in relation to motor/executive control, the left intraparietal sulcus (IPS) pattern linked to arithmetic ability (Schwizer Ashkenazi et al., 2024), and the tertiary sulci in the lateral prefrontal cortex associated with reasoning performance (Voorhies et al., 2021; Willbrand et al., 2022). **Revisions in Manuscript:** Introduction [Page 2, Lines 73]: *"...For instance, the morphological variability of the paracingulate sulcus has been associated with executive control. In the domain of numeracy, the sulcal pattern of the left intraparietal sulcus (IPS) —specifically the presence of a sectioned IPS—is significantly related to arithmetic performance in children and adolescents. Furthermore, recent studies highlight the role of late-developing structures in higher-order cognition, showing that the depth and presence of tertiary sulci in the lateral prefrontal cortex, such as the para-intermediate frontal sulcus, are predictive of reasoning abilities..."* **References** 1. Schwizer Ashkenazi, S. et al. Are numerical abilities determined at early age? A brain morphology study in children and adolescents with and without developmental dyscalculia. Developmental Cognitive Neuroscience 67, 101369 (2024). 2. Voorhies, W. I., Miller, J. A., Yao, J. K., Bunge, S. A. & Weiner, K. S. Cognitive insights from tertiary sulci in prefrontal cortex. Nat Commun 12, 5122 (2021). 3. Willbrand, E. H., Ferrer, E., Bunge, S. A. & Weiner, K. S. Development of human lateral prefrontal sulcal morphology and its relation to reasoning performance. J Neurosci 43, 2552 – 2567 (2023).
Results:	
5. Again, can you please include inter-scan interval times	**Response:** We thank the reviewer for this important reminder. We have now clarified in the revised manuscript that the inter-scan interval for the longitudinal participants was approximately one year. Specifically, we have added this information to both the Methods (Participants section) and Results sections to ensure clarity. For a detailed visualization of the age distribution and sampling time points, please refer to Supplementary Figure 1. **Revisions in Manuscript:** Results Section [Page 4, Lines 123] *...contributing a total of 490 MRI scans across three time points with an inter-scan interval of approximately one year (N=47 with 3 scans, N=97 with 2 scans, and N=168 with 1 scan ; see supplementary Fig. 1).* Methods Section [Page 12, Lines 459] *...For longitudinal participants, the inter-scan interval was approximately one year. Specifically, three MRI scans were available for 47 children (31 females and 16 males), two MRI scans were available for 97 children...*
6. The number of abbreviations is quite high and at times difficult to follow – I would consider writing out some of the	**Response:** We apologize for the difficulty caused by the excessive use of abbreviations. We have thoroughly reviewed the manuscript to improve readability. We have:

terms you are currently abbreviating (e.g. sentence starting in line 128 includes 6 abbreviations)	✓ Reduced the use of unnecessary abbreviations where possible. ✓ Ensured that all terms are defined upon their first appearance in each section. ✓ Provided a comprehensive Abbreviation Table (Supplementary Table 1) to serve as a quick reference for the anatomical nomenclature used in the study **Revisions in Manuscript:** We ensured that statistical abbreviations are explicitly defined upon their first use in each major section. For example, "FDR" is now written as "False Discovery Rate (FDR)" in the updated results text (See Results, paragraph 2). [Page 4, Lines 143] *"...we observed systematic morphometric changes characterized by decreases in adjusted surface area, adjusted maximum and mean depths, and adjusted sulcal length, accompanied by cortical thinning and sulcal widening across the brain..." (See Results, Morphometric Changes of Each Sulcus Throughout the Brain).*
7. As the results come before the methods, can you talk a bit more about how you examined the age-related changes, especially since you had 3 timepoints	**Response:** We agree that a brief methodological explanation at the beginning of the Results section is necessary for clarity given the structure of the manuscript. We have added a paragraph at the start of the Results section explicitly stating that we employed Generalized Additive Mixed Models (GAMMs) to guide the reader. To provide a more detailed explanation regarding our strategy for handling the 3 timepoints: We selected GAMMs specifically because they are ideally suited for identifying developmental trajectories in accelerated longitudinal designs like ours, where the number of timepoints varies across participants (N=47 with 3 scans, N=97 with 2 scans, and N=168 with 1 scan). Our approach addresses the complexities of this "unbalanced" data structure in two key ways: 1. Handling Longitudinal Dependencies (The "Mixed" Component): We included a subject-specific random intercept (u_i)in the model. This accounts for the correlation between repeated measures within the same individual (controlling for the fact that a child's second scan is not independent of their first). 2. Utilization of All Available Data: Crucially, mixed models allow us to utilize data from all participants, not just those with complete longitudinal sets. Participants with only one timepoint contribute valuable information to the estimation of the population-level mean (β_0) and fixed effects (e.g., sex), while participants with multiple timepoints contribute to estimating the within-subject variance and trajectory shape.
8. Are changes modelled linearly or non-linearly?	**Response:** To address the question regarding the shape of the developmental trajectories: The changes were modeled non-linearly. Developmental changes in childhood are often complex and do not necessarily follow a straight line. Instead of forcing a linear fit, we utilized the "Additive" component of Generalized Additive Mixed Models (GAMMs), which employs a non-parametric smooth function, f(Age), to model the relationship between age and sulcal morphology. Specifically, our modeling strategy involved: **Thin-Plate Regression Splines:** We used thin-plate regression splines (with k=3 basis dimensions) to fit the age term. This allows the model to flexibly capture linear, quadratic, or more complex local variations in development without imposing a rigid functional form a priori. **Model Structure:** The final model structure used was:

	$$Y_{ij} = \beta_0 + \beta_1 \cdot Sex_i + f(Age_{ij}) + u_i + \epsilon_{ij}$$ Where Y_{ij} is the morphological measure for subject i at timepoint j, and u_i is the random effect ($\sim N(0, \sigma_u^2)$). **Data-Driven Smoothness:** To ensure we did not overfit the data (e.g., fitting a curve where a line would suffice), the degree of smoothness was estimated from the data itself using Restricted Maximum Likelihood (REML). Additionally, we used nested model comparisons to confirm that the inclusion of the non-linear age term provided a better fit than a null model.
9. Paragraph starting in line 146: this is VERY hard to follow – can you instead move this information to a table?	**Response:** We completely agree with the reviewer that the original text listing statistical values was dense and difficult to read. We have removed the exhaustive list of statistics from the text and synthesized these findings into a new, comprehensive Table 1 ("Age-Related Developmental Changes in Sulcal Morphology"). This table now clearly presents the direction of change, Adjusted R^2 values, and significance levels (FDR-corrected) for key morphological features (Cortical Thickness, Sulcal Width, and TIV-adjusted Surface Area/Length/Depth) across all studied sulci. We believe this significantly improves the clarity and accessibility of the results.

Table 1. Age-Related Developmental Changes in Sulcal Morphometry.

Label	SA (adj)	maxD (adj)	meanD (adj)	SL (adj)	CT	SW
F.C.L._left	-0.045*	0.002	-0.005	0.001	-0.072*	0.058*
F.C.L._right	-0.040*	0.001	-0.001	-0.012	-0.034	0.018
F.C.M.ant._left	-0.013*	-0.015	-0.028	0.003	-0.037*	0.023
F.C.M.ant._right	-0.004	-0.017	-0.003	0.002	-0.029*	-3.27e-4
F.C.M.post._left	-0.004	-0.012*	-0.012*	-0.002	-0.027*	0.012
F.C.M.post._right	0.001	-0.019*	-0.019*	-0.002	-0.047*	0.030*
F.Cal.ant.-Sc.Cal._left	-0.027*	-0.014	-0.01	-0.013*	-0.059*	0.04
F.Cal.ant.-Sc.Cal._right	-0.002*	-0.009	-0.018*	-0.003	-0.024	0.017
F.Coll._left	-0.017*	-0.035*	-0.024	0.002	-0.030*	0.013
F.Coll._right	-0.015*	-0.032*	-0.031*	-1.37e-4	-0.03	0.01
F.I.P._left	-0.041*	-0.023*	0.005	-0.030*	-0.091*	0.030*
F.I.P._right	-0.013*	-0.017*	-0.01	0.001	-0.071*	0.033*
F.I.P.r.int.2_left	0.003	-0.002	-0.002	0.002	-0.025	0.004
F.P.O._left	-0.013*	-0.016	-0.012	-0.002	-0.081*	0.024*
F.P.O._right	-0.025*	-0.006*	-0.004	0.002	-0.099*	0.041*
OCCIPITAL_left	-0.002	0.003	0.004	0.001	-0.034*	0.022*
OCCIPITAL_right	-0.009	-0.009	-0.005	0.001	-0.032	0.018
S.C._left	-0.008*	-0.011	0.002	0.001	-0.01	0.006
S.C._right	-0.013*	-0.012*	0.005	-0.003	-0.013	0.007
S.Call._left	-0.005*	-0.013	-0.02	0.006	-0.018	0.01
S.Call._right	0.001	0.004	0.003	-0.001	-0.024	-0.001
S.Cu._left	-0.010*	-0.011	-0.007	-0.002	-0.058*	0.022
S.Cu._right	-0.015*	0.001	0.003	-0.012*	-0.052*	0.03
S.F.inf._left	-0.004	0.005	0.003	-0.002	-0.069*	0.033
S.F.inf._right	-0.004	-0.002	0.001	2e-4	-0.051*	0.054*
S.F.int._left	-0.006	-0.003	-0.003	-0.004	-0.005	0.01
S.F.int._right	-0.008	-0.005	0.002	-0.01	-0.027	0.005
S.F.inter._left	-0.001	-0.001*	-6.04e-5	0.003	-0.036*	0.015
S.F.inter._right	0.008	0.003	-0.008	0.004	-0.054*	0.016*
S.F.orbitaire._left	3.24e-4	-0.001	-0.011	0.003	-0.011	0.033
S.F.sup._left	-0.012*	-0.031*	-0.009	-0.006	-0.013	0.011
S.F.sup._right	-0.002*	-0.001	-0.003	0.005	-0.018*	0.024*
S.Li.ant._left	-0.003	-0.002	-0.003	-0.002	-0.026	0.005
S.Li.ant._right	0.002	0.002	0.002	0.002	-0.025	0.015
S.Li.post._left	-0.007	-0.006	-0.019	0.002	-0.008	0.015

S.Li.post._right	-0.009	-0.021	-0.005	0.002	-0.036*	0.009
S.O.T.lat._left	-0.019*	-0.012*	-0.011*	-0.005	-0.025*	0.037*
S.O.T.lat._right	-0.011*	-3.29e-5	-0.008	-0.019*	-0.051*	-0.004
S.Olf._left	-0.005	-0.027	-0.036*	0.001	-0.083*	0.053*
S.Olf._right	-0.028*	-0.045*	-0.040*	1.72e-4	-0.073*	0.080*
S.Or._left	-0.023*	-0.007	-0.016	-0.002	-0.048*	0.011
S.Or._right	-0.037*	-0.023*	-0.015*	-0.009	-0.040*	0.031*
S.Pa.int._left	-0.009*	1.12e-4	0.015*	-0.003	-0.081*	0.026*
S.Pa.int._right	-0.040*	-0.017*	-0.007*	-0.031*	-0.057*	0.022*
S.Pa.t._right	-0.004	-0.004	-0.005	-0.01	-0.023	-1.63e-4
S.Pe.C._left	-0.017*	0.001*	-0.01	-0.012*	-0.001	0.005
S.Pe.C._right	0.003*	0.018*	0.012	-0.002	-0.009	0.007
S.Po.C._left	-0.009*	-0.016*	-0.003	0.001	-0.057*	0.018
S.Po.C._right	0.001	0.004	-0.002	3.14e-5	-0.047*	0.022*
S.T.i._left	-0.008*	-0.002	3.71e-4	-0.006	-0.053*	0.114*
S.T.i._right	-0.009*	0.007	0.015	-0.011*	-0.047*	0.067*
S.T.s._left	0.001	0.002	-0.003	-0.006	-0.065*	0.037*
S.T.s._right	-0.004*	3.91e-4	-0.001	-0.006	-0.039*	0.034*

Note: Values represent the change in adjusted R^2 derived from Generalized Additive Mixed Models (GAMMs). To isolate specific regional morphometric changes from global brain growth, Surface Area, Sulcal Length, Max Depth, and Mean Depth were analyzed using TIV-adjusted residuals. Cortical Thickness and Sulcal Width are presented as raw measures. * indicate significant age-related effects that survived FDR correction
Abbreviations: adj: TIV-adjusted residuals; SA: Surface Area; maxD: Maximum Depth; meanD: Mean Depth; SL: Sulcal Length; SW: Sulcal Width; CT: Cortical Thickness.

10. Paragraph starting in line 170: this is very interesting but comes a bit out of the blue; can you integrate this more and explain, why/how this analysis added to the rest of the work? Also, there are again a lot of terms and abbreviations in this paragraph which, if you haven't yet read the methods, are difficult to follow.

Response:

We thank the reviewer for this insightful comment. We agree that the introduction of the sulcal emergence analysis appeared abrupt in the original text and that the density of abbreviations (T_{app}, T_{10}, T_{50}) created a barrier to readability.

The inclusion of these metrics was not arbitrary but designed to bridge our postnatal findings with fundamental theories of cortical morphogenesis. Specifically, we aimed to test whether the chronological order of folding acts as a biological constraint on postnatal plasticity. Our hypothesis is grounded in the "Protomap" theory (Rakic, 1988), which posits that the fundamental cytoarchitectonic map is specified in the ventricular zone during early embryogenesis. Complementing this, Van Essen (1997) proposed that mechanical tension along early-forming axon pathways drives the formation of primary folds. Therefore, early emergence times are a proxy for this "hard-wired" genetic and structural blueprint.

To quantify this, we referenced established benchmarks of sulcal emergence derived from in vivo MRI studies of preterm cortical folding and post-mortem fetal anatomical examinations (see Figure 3c), specifically using the time of appearance (T_{app}) to label sulci along a developmental timeline. This analysis was crucial for interpreting the heterogeneity in our results (i.e., why some sulci show strong age-effects while others do not). By integrating the literature on variability, we provided a mechanistic explanation: As noted by Lohmann et al. (2008), early-forming deep sulci exhibit high stability and low inter-individual variability across the population. This structural consistency—driven by the genetic Protomap—results in a high signal-to-noise ratio in our group analysis, manifesting as "stronger" statistical age-effects. In contrast, Hill et al. (2010) demonstrated that evolutionarily recent, late-expanding regions (association cortex) are highly sensitive to environmental inputs. Consequently, the development of late-emerging sulci is driven by unique, experience-dependent factors rather than a shared genetic timeline. This high inter-individual variability "dilutes" the group-level mean trend, explaining the "weaker" statistical associations we observed.

To address your concerns about flow and clarity, we have made the following changes: We have added a paragraph at the beginning of this section explicitly outlining the theoretical motivation described above. We have rewritten the text to define all acronyms immediately upon first use. We have added a subsection in the Methods to rigorously define the derivation of T_{app}, T_{10}, T_{50}, as well as their biological significance.

Revisions in Manuscript:

Results Section [Page 5, Lines 167]

"Association between Sulcal Emergence Timing and Developmental Magnitude To investigate whether the chronological order of sulcal formation organizes postnatal development, we examined the relationship between the timing of sulcal emergence in utero and the magnitude of age-related changes observed in our cohort. We hypothesized that, primary folds are the first to emerge under strong genetic control, establishing stable anatomical landmarks that persist from birth. In contrast, secondary and tertiary folds emerge later and are increasingly susceptible to mechanical and environmental influences. To investigate whether the chronological order of sulcal formation organizes childhood development, we examined the relationship between the timing of sulcal emergence and the magnitude of age-related changes observed in our cohort. Sulcal emergence was parameterized using three complementary metrics (detailed in Methods): Tapp (the gestational week of visual appearance based on anatomical consensus) and T10 / T50 (the gestational ages marking the onset and midpoint of folding, derived from curvature modeling)."

Method Section [Page 13, Lines 528]

Definition of Sulcal Emergence Timing

To characterize the chronological hierarchy of cortical folding, we defined sulcal emergence using three complementary metrics that integrate classical anatomical consensus with modern computational modeling (individual values for all sulci are listed in Supplementary Table 1). We first defined the Sulcal Appearance Time (Tapp) as the gestational age (in weeks) at which a sulcus first becomes visibly identifiable on the cortical surface. Values for Tapp were collated from a comprehensive review of established anatomical literature, including fetal autopsy studies and in utero MRI assessments, serving as a benchmark for the stage where primitive sulcal indentations become macroscopically consistent across the population. To complement these discrete visual observations, we utilized continuous quantitative milestones derived from the trajectory of surface curvature changes, as formally described by Snyder et al. (2024). In this framework, the maturation of mean curvature for each sulcus was modeled using a logistic growth function across gestation to extract two keys physiological timepoints: the onset of folding (T10), defined as the gestational age where local mean curvature reaches 10% of its maximum projected adult value, and the midpoint of folding (T50), defined as the gestational age corresponding to 50% of curvature maturation.

References:

1. Rakic, P. Specification of cerebral cortical areas. Science 241, 170 – 176 (1988).
2. Van Essen, D. C. A tension-based theory of morphogenesis and compact wiring in the central nervous system. Nature 385, 313 – 318 (1997).
3. Dubois, J. et al. The dynamics of cortical folding waves and prematurity-related deviations revealed by spatial and spectral analysis of gyrification. NeuroImage 185, 934 – 946 (2019).
4. Nishikuni, K. & Ribas, G. C. Study of fetal and postnatal morphological development of the brain sulci: Laboratory investigation. Journal of Neurosurgery: Pediatrics 11, 1 – 11 (2013).
5. Chi, J. G., Dooling, E. C. & Gilles, F. H. Gyral development of the human brain. Ann. Neurol. 1, 86 – 93 (1977).

	6. Lohmann, G., von Cramon, D. Y. & Steinmetz, H. Sulcal variability of twins. Cereb Cortex 9, 754 – 763 (1999). 7. Hill, J. et al. Similar patterns of cortical expansion during human development and evolution. Proc. Natl. Acad. Sci. U. S. A. 107, 13135 – 13140 (2010).
11. When you say "earlier emergence associated with stronger age-related effects", does this mean that sulci formed earlier show a stronger change over time during childhood?	**Response:** Yes, your interpretation is entirely correct. Our findings indicate that sulci formed earlier in gestation exhibit more profound morphological remodeling during the childhood-to-adolescence period. In the revised Results section ("Association between Sulcal Emergence Timing and Developmental Magnitude"), we explain that surface area generally decreases with age in our sample (yielding negative Signed Δ adj R^2 values). Consequently, a "stronger change" corresponds to a larger negative value. We observed a chronological gradient where the earliest-forming primary sulci (lower Tapp) exhibit significantly larger negative values compared to later-emerging sulci, which show effects closer to zero. This confirms that the earliest anatomical landmarks undergo the most significant quantitative restructuring during this developmental stage.
12. Regression results – can you include significance levels?	**Response:** Thank you for this suggestion. We have updated the Results section to explicitly include the correlation coefficients and significance levels for the regression analyses linking sulcal emergence timing to developmental magnitude. Specifically, as detailed in the revised manuscript: For adjusted Surface Area (SA), we report a significant positive correlation between Appearance Time (Tapp) and the magnitude of change (Signed$\Delta adj\ R^2$) with *r=0.36* and adjusted *p=0.031*.When using the folding onset metric (T10), we observed a similar significant correlation with *r=0.39* and adjusted *p=0.031*. For adjusted Maximum Depth, earlier emergence also predicted stronger age-related changes with *r=0.39* and adjusted *p=0.034*. **Revisions in Manuscript:** [Page 5, Lines 181] *"… we observed significant positive correlations between T_{app} and the Signed $\Delta adj\ R^2$ ($r = 0.36$, $adjusted\ p = 0.031$, Fig. 3D). Similarly, T_{10} showed a positive correlation with the Signed $\Delta adj\ R^2$ ($r = 0.39, adjusted\ p = 0.031$). …Similar patterns were observed for adjusted maximum depth, where earlier emergence predicted stronger age-related changes and higher statistical significance ($r = 0.39$, $adjusted\ p = 0.034$)."*
Discussion:	
13. How do you account for the wide age range of your participants? Surely someone aged 6 years will show a different development compared to someone aged 14 years?	**Response:** We appreciate the reviewer raising this important point regarding the developmental heterogeneity across our age range (6–14 years). We fully agree that a 6-year-old brain is distinct from a 14-year-old brain. This is precisely why we employed Generalized Additive Mixed Models (GAMMs) rather than simple linear models or group-based comparisons. Our modeling approach accounts for this wide age range in two key ways: 1. Continuous Trajectory Modeling: Instead of treating age as a categorical variable or assuming a uniform linear change across the entire range, GAMMs utilize non-parametric thin-plate regression splines to fit a smooth function of age, denoted as $f(Age)$. This allows the model to capture non-linear developmental trajectories. It can flexibly characterize distinct rates of change at different stages—for example, detecting rapid changes at age 6 that might plateau by age 14—without imposing a rigid linear assumption. 2. Longitudinal Design: By including subject-specific random effects, the model leverages the longitudinal nature of our data (repeated measures). This ensures that the estimated trajectory reflects valid within-subject

	developmental changes over time, rather than merely cross-sectional differences between younger and older children. Therefore, our approach explicitly models how development differs between age 6 and 14 by mapping the continuous trajectory connecting these timepoints.
14. What about regions where you do not observe age-effects? How would you interpret their 'stability' over time?	**Response:** We thank the reviewer for this insightful question regarding the interpretation of null results. We agree that "stability" (i.e., the absence of statistically significant age-related changes) requires careful biological contextualization, as it may stem from distinct underlying mechanisms depending on the cortical region's ontological status. We have re-examined the specific sulci that showed no significant age effects in our dataset. Based on their anatomical classification and our statistical results, we attribute this observed stability to two distinct mechanisms: 1. Structural Plateau of Primary Folds (The "Anchor" Hypothesis) For primary sensorimotor folds, the observed stability likely reflects that these regions have reached a developmental plateau prior to the age range of our cohort (6 – 14 years). In our dataset, the Central Sulcus (S.C.) and Callosal Sulcus (S.Call.) — which are among the earliest folds to develop—showed no significant age-related changes in key morphological metrics. Specifically, the left Central Sulcus showed stable surface area and depth, while the Callosal Sulcus exhibited stability in width and depth across the midline. Biologically, these primary sulci serve as genetically constrained "biomechanical anchors" for the cortical landscape (Lohmann et al., 2008). Given that longitudinal mapping has established that the sensorimotor cortex matures earliest (Gogtay et al., 2004), the absence of linear age effects in these regions suggests they function as invariant anatomical landmarks during childhood and adolescence, maintaining structural stability while adjacent association areas continue to reorganize. 2. High Anatomical Variability in Tertiary Sulci In contrast, null results observed in tertiary frontal regions likely stem from high inter-subject heterogeneity rather than a maturational plateau. Our data revealed that the Orbitofrontal Sulcus (S.F.orbitaire) and the Intermediate Frontal Sulcus (S.F.inter) exhibited high p-values with no discernible developmental trend (e.g., left S.F.orbitaire Surface Area: $P > 0.05$; left S.F.inter mean depth: $P > 0.05$). Unlike primary folds, these tertiary sulci are phylogenetically recent and highly variable across individuals (Ono et al., 1990). The morphology of the orbitofrontal cortex is known to be notoriously heterogeneous, with significant individual differences in branching patterns (Chiavaras & Petrides, 2000). This high individual variability likely masks group-level developmental trends, distinguishing this form of statistical stability from the maturational plateau observed in primary regions. Thus, the null results in our study are not uniform: for the Central and Callosal sulci, they likely represent a structural plateau of primary folds; for the Orbitofrontal regions, they likely reflect the anatomical heterogeneity inherent to the association cortex. **References:** 1. Chiavaras, M. M. & Petrides, M. (2000). Orbitofrontal sulci of the human and macaque monkey brain. Journal of Comparative Neurology, 422, 35 – 54. 2. Lohmann, G., von Cramon, D. Y. & Colchester, A. C. F. (2008). Deep sulcal landmarks provide an organizing framework for human cortical folding. Cerebral Cortex, 18, 1415 – 1420.

	3. Gogtay, N. et al. (2004). Dynamic mapping of human cortical development during childhood through early adulthood. Proc. Natl. Acad. Sci., 101, 8174 – 8179. 4. Ono, M., Kubik, S. & Abernathey, C. D. (1990). Atlas of the Cerebral Sulci. Thieme.
15. Do 'non-changing' sulci also predict clinical performance? And how are they underpinned by molecular mechanisms?	**Response:** We thank the reviewer for this thought-provoking series of questions. We agree that the absence of statistical age-effects (i.e., "stability") requires careful biological contextualization. To address this, we integrated our null-result findings with our cognitive prediction models (LASSO/PLS) and consulted the relevant literature to interpret the underlying molecular landscape. **Our analysis reveals that "non-changing" sulci are significantly predictive of cognitive performance,** but they operate through two distinct mechanistic pathways—serving as either "Stable Scaffolds" or "Variable Substrates"—depending on their anatomical hierarchy and evolutionary history. Regarding functional relevance, our data challenges the notion that only developmentally dynamic regions are functionally significant. First, we found that the Central Sulcus (S.C.), which showed a structural plateau, served as a robust predictor of Working Memory performance ($r \approx 0.23$). We interpret this as a "Neural Efficiency" model where the sulcus acts as a "stable scaffold." Consistent with the literature describing early folds as invariant landmarks, the morphological integrity of this primary region—established early in development—provides the necessary sensorimotor anchor for the efficient execution of higher-order executive loops (Im & Grant, 2019). Conversely, the Orbitofrontal Sulcus (S.F.orbitaire) exhibited null age effects driven by high inter-subject variability ($p = 0.98$) yet emerged as a top predictor for the Alerting Network ($r = 0.25$). This reflects a "Computational Capacity" model where the lack of a unified age trend stems from the anatomical heterogeneity inherent to tertiary sulci. A larger surface area in this region suggests a richer, albeit highly variable, cortical substrate available to support complex, top-down attentional control (Hill et al., 2010). Mechanistically, we interpret these two forms of stability through the lens of the Protomap and Radial Unit Hypothesis, supported by evidence of differential evolutionary and postnatal expansion. The structural invariance of primary folds (like the S.C.) appears underpinned by strong spatiotemporal genetic control. These folds are hypothesized to be biologically determined by a highly conserved "protomap" of gene expression that defines the boundaries of functional areas early in fetal life. As noted in reviews of early cortical folding, these primary landmarks form during the radial growth phase and show high spatial invariance, suggesting they are tightly regulated by intrinsic genetic programs to ensure the "anchor" remains fixed. In contrast, the "stability" of tertiary folds (like the S.F.orbitaire) manifests as high variability, which is consistent with regions of recent evolutionary expansion that are less genetically constrained. Comparative studies between humans and macaques indicate that these high-expansion association areas (frontal/temporal) undergo "protracted" development and are more sensitive to environmental inputs. While primary folds are linked to the stable protomap, later-developing tertiary folds are influenced significantly by tangential expansion and mechanical factors, leading to more variable and chaotic folding patterns. This suggests that the morphology of these "plastic zones" is less rigidly determined by the early genetic template, allowing them to act as flexible substrates that adapt to individual experiences rather than following a fixed species-wide trajectory during childhood. **References:** 1. Lohmann, G. Sulcal variability of twins. Cereb. Cortex 9, 754 – 763 (1999). 2. Chiavaras, M. M. & Petrides, M. Orbitofrontal sulci of the human and

	macaque monkey brain. Journal of Comparative Neurology 422, 35 – 54 (2000). 3. Glasser, M. F. et al. A multi-modal parcellation of human cerebral cortex. Nature 536, 171 – 178 (2016). 4. Hill, J. et al. Similar patterns of cortical expansion during human development and evolution. Proc. Natl. Acad. Sci. U. S. A. 107, 13135 – 13140 (2010). 5. Im, K. & Grant, P. E. Sulcal pits and patterns in developing human brains. NeuroImage 185, 881 – 890 (2019).
16. Overall, the concept of studying sulci and gyri, rather than e.g., vertices or regions, is very interesting and important – but could you emphasize more (in the introduction/discussion) why that is the case, and what this approach adds to the aforementioned analyses at the region- or vertex-level?	**Response:** Thank you for this valuable suggestion. We have elaborated on three key advantages of this level of analysis: **Biological Validity & Developmental Anchoring:** Unlike vertices, which are arbitrary discretization points, sulci and gyri represent biologically meaningful units rooted in neurodevelopment. While cortical folding is mechanically driven by the differential expansion of the cortex relative to the underlying white matter (Tallinen et al., 2016), the locations of primary folds are not random. They are widely considered to be "anchored" by genetically determined sulcal roots (Regis et al., 2005), ensuring anatomical homology across individuals. Furthermore, sulcal fundi frequently align with cytoarchitectonic and functional boundaries (e.g., the central sulcus delineating motor from somatosensory cortex) (Fischl et al., 2008). Thus, analyzing data at the level of these structures captures intrinsic biological organization better than geometric grids. **Topological Alignment:** Standard vertex-wise analyses rely on spherical registration, which often fails to perfectly align variable folding patterns across subjects. Since human cortical folding exhibits high variability, a vertex at a fixed coordinate may correspond to a gyrus in one subject and a sulcus in another. By extracting and labeling specific sulci, our approach aligns data based on anatomical features rather than spatial coordinates, thereby directly addressing the issue of structural misalignment and ensuring that comparisons are made between homologous anatomical regions. **Statistical Power:** Analyzing meaningful anatomical units (dozens of sulci) rather than high-dimensional meshes (hundreds of thousands of vertices) drastically reduces the multiple comparison correction burden, thereby enhancing the sensitivity to detect robust biological effects. **Revisions in Manuscript:** In discussion: [Page 11, Lines 418] *"In contrast to traditional vertex-wise analyses, the sulcal morphology strategy employed in this study is anchored to topologically invariant anatomical units [Regis et al., 2005], thereby ensuring the biological homology of the observed developmental trajectories."* **References:** 1. Tallinen, T. et al. On the growth and form of cortical convolutions. Nature Phys 12, 588 – 593 (2016). 2. Régis, J. et al. 'sulcal root' generic model: A hypothesis to overcome the variability of the human cortex folding patterns. Neurol. Med. Chir. (tokyo) 45, 1 – 17 (2005). 3. Fischl, B. et al. Cortical Folding Patterns and Predicting Cytoarchitecture. Cereb Cortex 18, 1973 – 1980 (2008).
17. Line 276: I cannot follow your argument here – if earlier developing sulci change more during childhood, would that not indicate that those sulci	**Response:** We thank the reviewer for this insightful comment. We agree that the relationship between "magnitude of change" and "genetic vs. environmental regulation" is nuanced and was not clearly articulated in our original discussion. Intuitively, high plasticity (environmental susceptibility) implies greater change. However, our argument rests on the distinction between consistent

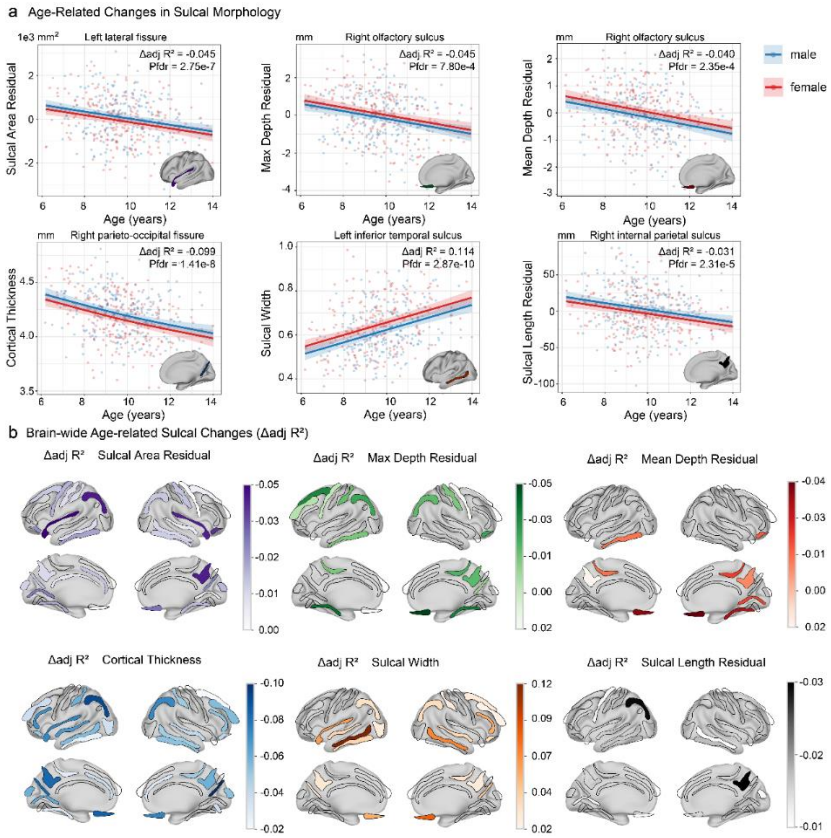
are in fact more susceptible to environmental regulation?	developmental remodeling (driven by genetic programs or universal environmental demands) and experience-dependent plasticity (driven by individual-specific environmental factors).We have revised the manuscript (specifically the Discussion) to clarify this mechanism. Our interpretation is based on the following framework: 1. Early-Emerging Sulci (Canalization and Shared Trajectories): These primary folds are under stronger genetic control (Lohmann et al., 1999). During childhood and adolescence, these regions undergo substantial remodeling, but this remodeling is likely driven by universal developmental milestones shared by all individuals (e.g., basic sensorimotor maturation). Because these "environmental" demands are ubiquitous and the genetic blueprint is robust, the developmental trajectory is highly consistent across the cohort. This results in a strong statistical signal and a steep developmental slope at the group level. 2. Late-Emerging Sulci (Plasticity and Inter-individual Variability): Late-emerging regions are indeed more plastic and susceptible to environmental regulation, as noted by Hill et al. (2010). However, this susceptibility relates to individual-specific experiences (learning unique skills, distinct social environments). Consequently, while these regions may change significantly within a single individual, the direction and magnitude of change vary widely between individuals. This high inter-individual variability (noise) dilutes the group-level mean trend, resulting in what appears to be "weaker" age-related effects in our statistical models. **References:** 1. Lohmann, G., von Cramon, D. Y. & Steinmetz, H. Sulcal variability of twins. Cereb Cortex 9, 754 – 763 (1999). 2. Hill, J. et al. Similar patterns of cortical expansion during human development and evolution. Proc. Natl. Acad. Sci. U. S. A. 107, 13135 – 13140 (2010).
18. Line 304: by alterations, do you mean reductions?	**Response:** We appreciate the opportunity to clarify this point. By "alterations," we specifically refer to deviations from the optimal size in either direction (both reductions and expansions), rather than reductions alone. In the referenced study by Leingärtner et al. (2007), the authors genetically manipulated Emx2 levels in mice. They found that both low levels (leading to expanded sensorimotor areas) and high levels (leading to reduced sensorimotor areas) resulted in significant behavioral deficits. This suggests that there is an evolutionarily "optimal" size range for cortical areas, and any significant morphological deviation—whether an increase or a decrease—can impair functional performance. We have explicitly clarified this interpretation in the revised Discussion **Reference** Leingärtner, A. et al. Cortical area size dictates performance at modality-specific behaviors. Proc. Natl. Acad. Sci. 104, 4153 – 4158 (2007).
19. Would you consider adding strengths and limitations, and a conclusion?	**Response:** We fully agree with this suggestion as it enhances the structural completeness of the manuscript. 1. Limitations: We have added a dedicated "Limitations" section at the end of the Discussion. In this section, we critically evaluate our sample characteristics (e.g., the number of time points), the effect sizes of the observed changes, and the methodological constraints associated with using adult gene expression atlases (AHBA) for pediatric data. 2. Conclusion: We have also added a concise "Conclusion" section to summarize the main findings regarding the "fine-tuning" developmental model, the "canalization" principle, and the structure-function-gene couplings.

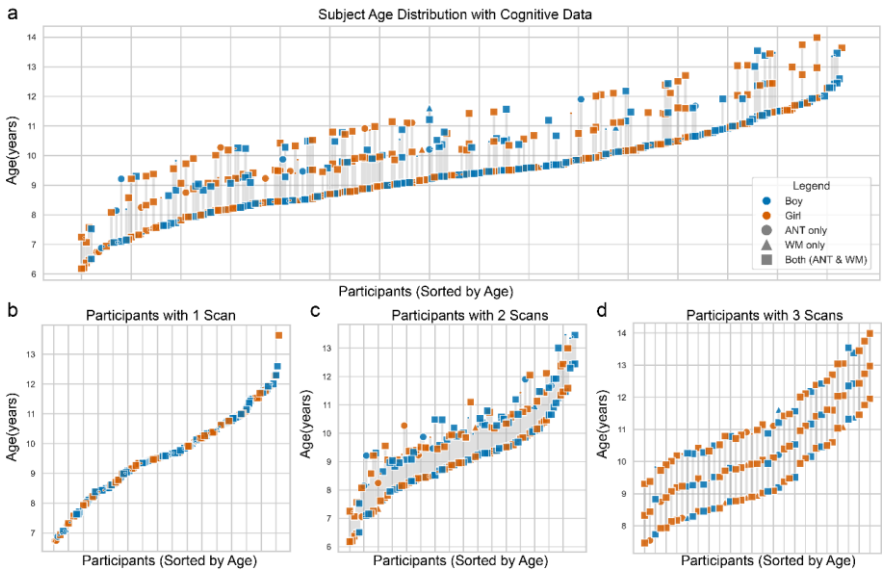
	Revisions in Manuscript: *Limitations* [Page 11, Lines 417] *In contrast to traditional vertex-wise analyses, the sulcal morphometry strategy employed in this study is anchored to topologically invariant anatomical units88, thereby promoting the biological homology of the observed developmental trajectories. However, several limitations warrant consideration when interpreting these results. First, despite the substantial total sample size, the longitudinal sampling density was limited, with the majority of subjects contributing only 2-3 time points (>75%). This constrains our capacity to fully characterize complex, non-linear developmental trajectories. Second, the observed effect sizes were modest. While this is consistent with typical patterns in large-scale neuroimaging association studies, it warrants caution regarding the immediate translation of these findings to clinical contexts. Third, the metrics employed to define the timing of sulcal emergence (i.e., T10, T50, and Tapp) were derived from heterogeneous literature sources and were largely based on visual inspection standards or cross-sectional inferences. Crucially, these heuristic thresholds currently lack rigorous biological validation regarding developmental onset and have not been verified through individual-level longitudinal tracking. Consequently, they may represent statistical approximations rather than precise anatomical milestones, highlighting the need for future studies to establish more physiologically grounded standards. Finally, the gene enrichment analysis relied on the adult AHBA atlas. Although the spatial topology of cortical gene expression is relatively conserved postnatally89, developmental-specific dynamic expression patterns may be overlooked. Future studies should integrate pediatric transcriptomic data and Diffusion Tensor Imaging (DTI) to further validate the direct links between sulcal morphological remodeling, white matter microstructure, and specific gene expression profiles.* *Conclusion* [Page 11, Lines 439] *In summary, by integrating large-scale longitudinal neuroimaging, cognitive-behavioral data, and transcriptomics, this study systematically characterizes the developmental patterns of sulcal morphometry from childhood to adolescence at the whole-brain level. Our results indicate that sulcal morphometric remodeling acts as a "fine-tuning" process, characterized by cortical thinning and sulcal widening. This observation aligns with the principle of developmental "canalization"—suggesting that early-forming primary folds exhibit robust developmental trajectories. Furthermore, we revealed an intrinsic link between these macroscopic structural changes and the expression of genes related to synaptic signaling and the cytoskeleton. We also confirmed that specific sulcal morphometric features serve as potential morphometric signatures for predicting gains in higher-order cognitive functions, such as working memory and attention control. These findings provide novel biological insights into structural brain maturation during typical development.*
Methods:	
20. Line 332: correct “with cognitively abnormal”	**Response:** We apologize for this grammatical error. We have corrected the phrase in the revised Methods section ("Participants") to read: *"...exclude participants with cognitive abnormalities..."* [Page 12, Lines 456]
21. Why were children recruited based on their performance on a cognitive test?	**Response:** We appreciate the opportunity to clarify our recruitment procedures. The "standardized cognitive ability test" mentioned in the Participants section refers to a general aptitude screening (assessing basic mathematical and linguistic abilities) that served strictly as an inclusion criterion during the initial recruitment phase. The purpose of this screening was to ensure that the cohort represented typically developing children with normal intelligence levels, capable of

	understanding instructions and completing the subsequent experimental tasks (such as the N-back and Attention Network Test). Children with intellectual disabilities or those unable to comprehend the screening tasks were excluded. Importantly, this screening score was used solely for cohort selection and was not used as a behavioral variable in our subsequent structure-function analyses.
22. You should be clearer throughout the manuscript that only 47 participants had all three scans and that 186 participants had cross-sectional data only - How did you examine 'change' in individuals with one scan only?	**Response:** We acknowledge the need for absolute transparency regarding the dataset structure. First, we have explicitly reiterated the sample composition in both the Participants and Results sections: "$N = 47$ *with 3 scans,* $N = 97$ *with 2 scans, and* $N = 168$ *with 1 scan (see Supplementary Figure 1).*" Second, regarding the modeling of "change" with single-scan participants: It is important to clarify that in a Mixed Model (GAMM) framework, **we do not calculate "change" for single-scan individuals in isolation**. Instead, these participants contribute to the model in the following ways: 1. Anchoring the Trajectory: Single time-points contribute to estimating the population-level intercept (β_0) and the overall shape of the developmental curve ($f(Age)$). They help "anchor" the trajectory, providing high-density data points across the age span (6–14 years). 2. Improving Precision: By including these subjects, we increase the effective sample size for estimating the fixed effects (Age, Sex), which in turn improves the precision of the estimated longitudinal slope for the entire population. The model borrows strength from the entire dataset to infer the developmental trend, effectively utilizing both cross-sectional (between-subject) and longitudinal (within-subject) information.
23. How did you account for the very large number of models (presumably ~50 sulci by six features)?	**Response:** We rigorously controlled for multiple comparisons to minimize Type I errors. Given the large number of models (53 sulci $\times$ 6 morphological features $=$ 318 models), we applied the Benjamini-Hochberg False Discovery Rate (FDR) correction. The p -values for the age smooth terms (approximate significance of smooth terms) were corrected across all models. We have explicitly stated in the Methods section that only effects surviving an FDR threshold of $q < 0.05$ are reported as significant.
24. How did you account for the different times of scanning, the difference inter-scan intervals, and the fact that some individuals have one, two, or three scans?	**Response:** This complexity is exactly why we chose Generalized Additive Mixed Models (GAMMs) over traditional repeated-measures ANOVA. Our model accounts for these variations through three specific mechanisms: 1. Continuous Age Variable: GAMMs treat Age as a continuous variable rather than discrete "time bins." This means the exact age of each child at each scan (e.g., 7.2 years vs. 7.8 years) is used directly, naturally handling variable scanning times. 2. Random Intercepts: We included a subject-specific random intercept (u_i) to account for the correlation between repeated measures within the same individual. This handles the unbalanced design (1, 2, or 3 scans) by allowing the model to estimate subject-specific baselines while estimating a common developmental trajectory. 3. Robustness to Intervals: In mixed models, the "inter-scan interval" does not need to be uniform. The model fits the trajectory based on the available data points in the continuous age space. We have ensured the Methods section clearly describes this statistical strategy.
25. Were cognitive performances age-normed? Would it be helpful to	**Response:**

examine percentage-wise change rather than delta?	We thank the reviewer for these crucial methodological questions regarding our cognitive measures. We have carefully re-examined our choice of metrics and performed additional analyses to validate our approach. **1. On Age-Norming vs. Raw Scores:** We explicitly used raw (absolute) performance scores (standardized Z-scores) derived from standardized cognitive paradigms — specifically the classic numerical N-back task and the Attention Network Test (ANT). These well-validated tasks provide stable metrics that allow us to reliably distinguish and quantify the cognitive improvements explicitly driven by age-related development. Our primary research goal is to track the absolute developmental trajectory of cognitive functions in relation to brain maturation. Age-normed scores represent a child's relative standing among peers. If a child develops at a normal rate, their age-normed score remains constant, effectively "regressing out" the very developmental gains we aim to study. By using absolute scores, we preserve the longitudinal growth signal essential for linking morphological changes to functional improvements. To address the concern about baseline age bias, we examined the correlation between Baseline Age and Baseline Performance. For Working Memory, Alerting and Orienting networks, we found no significant correlation ($p > 0.05$), indicating that baseline age did not bias baseline performance. For Executive Control, we observed a weak but expected correlation ($r \approx 0.17, p < 0.05$). Since improvement in this metric corresponds to negative values (reduction in conflict cost), this positive correlation indicates that younger children showed larger improvements compared to older children, consistent with the steeper developmental curve in early childhood. Crucially, preserving this age-related variance is necessary to model the coupling between brain growth and cognitive gain. Using age-normed scores would have artificially removed this biological signal. **2. On Percentage-wise Change vs. Delta (Δ):** We chose the absolute difference (Δ) to establish a direct mechanistic link between the absolute magnitude of morphological change and the absolute magnitude of cognitive gain. Percentage change is highly sensitive to the baseline denominator. Subjects with lower baseline scores (smaller denominators) would exhibit mathematically inflated percentage changes for the same absolute gain. Our analysis shows that using percentage change introduces a strong artificial negative correlation with baseline scores, masking true biological associations. From a measurement theory perspective, our metrics (Z-scores and RT differences) operate on an Interval Scale, which inherently lacks a biological 'absolute zero.' Consequently, ratio-based calculations may not be appropriate in this context (similar to how a temperature change from 10° C to 20° C is not a "100% increase in heat"). While normalization strategies could address the issue of negative values, they often rely on sample-dependent transformations that may introduce unintended biases. For these reasons, we determined that the absolute Delta (Δ) is the most robust and theoretically consistent metric to quantify the 'functional gain per unit of structural growth' in our longitudinal model.
26. When you examined association of anatomical change and clinical profiles, did you base your analysis on the regression betas of change? And if so, how did you accommodate individuals with 1 scan? And for individuals with three scans, did you examine change between T1 and T2, and T2 and T3 separately?	**Response:** We appreciate the opportunity to clarify our analytical strategy for the brain-behavior association, which differs from the GAMM developmental modeling. **1. Exclusion of Single-Scan Participants:** For the analysis of associations between morphological change and cognitive change, we strictly included only participants with longitudinal data (i.e., at least two time points). Participants with only one scan ($N = 168$) were excluded from this specific part of the analysis because morphological "change" cannot be computed for them. **2. Calculation of Change Scores (Δ):**

	We did not use regression betas to represent change for individual subjects. Instead, we calculated the direct longitudinal change scores (Δ) for each subject i by subtracting baseline (T_k) values from follow-up (T_{k+1}) values: $$\Delta\text{Metric}_i = \text{Metric}_{i,T_{k+1}} - \text{Metric}_{i,T_k}$$ For individuals with three scans (T1, T2, T3), we maximized the use of the longitudinal data by examining changes between consecutive time points (i.e., treating T1→T2 and T2→T3 as available longitudinal instances), provided the corresponding cognitive data were also available and valid. **3. Statistical Control:** To account for the variability introduced by this approach, our statistical models (Partial Correlation and LASSO Regression) rigorously controlled for baseline age, sex, and crucially, the inter-scan interval (the specific time elapsed between the two scans). This ensures that the observed associations reflect true developmental coupling rather than artifacts of differing time intervals. We have revised the Methods section ("Statistical analysis of brain – behavior associations") to explicitly describe this procedure.
27. For the gene decoding and enrichment analysis: can you clarify a bit more how you downsampled AHBA gene values to your sulci / gyri regions of interest? (ie what do you mean by "selected sulcal regions were systematically merged according to predetermined grouping criteria"? What do you mean by "regions deemed irrelevant to the analysis")	**Response:** We appreciate the opportunity to clarify our gene mapping pipeline. To generate reliable gene expression profiles for our specific sulcal regions, we processed the BrainVISA probabilistic atlas through a rigorous customization step before mapping the AHBA data. 1. Clarification on "Systematic Merging": The native BrainVISA atlas is highly granular, often segmenting a single anatomical sulcus into multiple small fragments based on local discontinuities. To ensure anatomical coherence and sufficient tissue sample coverage from the AHBA dataset, we merged these fragments into cohesive anatomical units. ✓ We followed standard anatomical definitions (Ono et al., 1990). For example, fragmented segments of the inferior frontal sulcus (S.F.inf) or the lateral fissure (F.C.L) were merged into single, continuous structural entities (as detailed in Supplementary Figure 2).This merging was essential to create robust Regions of Interest (ROIs) that matched the standard neuroanatomical nomenclature used in our morphological analysis and ensured that each region was large enough to contain valid AHBA tissue samples. 2. Clarification on "Regions Deemed Irrelevant": This phrase refers to two specific types of exclusions applied during the atlas customization phase: ✓ Non-Sulcal Structures: The original probabilistic atlas includes probability maps for non-cortical structures (e.g., ventricles, corpus callosum) or non-sulcal landmarks (e.g., the interhemispheric fissure). These were excluded as they are not the focus of our cortical sulcal analysis. ✓ Unreliable Extractions: As detailed in our quality control procedures (Methods and Supplementary Figure 3), certain small or highly variable sulci (e.g., F.I.P.r.int.2_right, S.Pa.t._left) had extremely low extraction success rates (< 75%) in our pediatric cohort. To maintain consistency between the morphological analysis and the gene expression analysis, these regions were also excluded from the transcriptomic mapping. 3. Downsampling Procedure: After defining these 53 valid sulcal ROIs, we used the abagen toolbox to map the AHBA microarray data. • We did not simply "downsample" the high-resolution AHBA data. Instead, abagen assigns tissue samples to regions based on MNI coordinates. • Crucially, to ensure spatial specificity, we used the modified atlas where only voxels with a sulcal probability > 0.2 were retained (as described in the Methods section).

	• Tissue samples falling within these thresholded, merged sulcal masks were then aggregated (averaged) to generate the expression profile for each ROI. We have revised the Gene Enrichment Analysis section in the Methods to explicitly describe this merging and filtering logic, ensuring the procedure is transparent and reproducible.
28. Figures: helpful, but text is (partly) too small; for instance, axis labels of Figure 2 are difficult to decipher.	**Response:** We thank the reviewer for pointing this out. We have carefully revised Figure 2 and all other figures to increase the font size of axis labels, legends, and annotations, ensuring they are clearly legible at standard publication size.  **Fig. 2 Quantitative analysis of age-related changes in sulcal morphometry. (a)** Developmental trajectories of key sulcal features. The developmental characteristics for each sulcus were modeled using a generalized additive mixed model (GAMM). To represent the most prominent developmental trends, this panel displays the trajectories of the sulci showing the largest developmental effects (maximum absolute Δadj R^2) for each of the six morphometric features (Surface Area, Maximum Depth, Mean Depth, Sulcal Length, Sulcal Width, and Cortical Thickness). The Δadj R^2 and p-values from the GAMM are provided, with blue indicating males and red indicating females. **(b)** Regional distribution of developmental effects. Δadj R^2 values associated with age-related changes across six features for various sulci. All sulci displayed in the figure exhibit significant age-related developmental effects ($p < 0.05$, FDR corrected $p < 0.05$).
29. Figure (no number) with Subject Age distribution – what is the x-axis here?; also, can you plot separately for individuals with one/two/three scans; colors are very difficult to make out in this graph (including the green/purple	**Response:** Response: We apologize for the lack of clarity in the original version of this figure. Following your constructive suggestions, we have significantly revised the figure (now Supplementary Figure 1) to improve its clarity and informativeness:

outline); also, what do ANT/WM mean? Please provide a figure legend	**X-axis Clarification:** We have clearly labeled the x-axis. In these plots, the x-axis represents individual participants, who are sorted by their baseline age (from youngest to oldest) to provide a clear view of the sample's age span. Separate Panels: As suggested, we have added three sub-panels (b, c, and d) to separately visualize individuals with one, two, and three or more scans. This avoids the visual clutter of the full sample plot. **Enhanced Visualization:** To improve distinguishability beyond color alone, we now use different markers: Circles for participants with ANT data only, Triangles for WM data only, and Squares for those with both. We have adopted a high-contrast, color-blind-friendly palette (Blue for boys, Vermillion for girls).Sub-panels b – d are plotted for easier comparison between scan-frequency groups. A detailed legend has been added. We have also explicitly defined the abbreviations: ANT stands for Attention Network Test, and WM stands for Working Memory (N-back task).  **Supplementary Figure 1. Age distribution of participants and cognitive data availability.** The x-axis represents individual participants sorted by their baseline age, with gray lines connecting repeated scans of the same individual. Colors denote gender (blue for boys; orange for girls). Different marker shapes indicate the availability of cognitive data at each specific time point, including data from the Attention Network Test (ANT) and/or Working Memory (WM; N-back task). (a) Full sample distribution: Age distribution for the entire longitudinal dataset (n = 490 scans from 312 participants). (b – d) Sub-group distributions: Separate panels for participants with (b) one scan, (c) two scans, and (d) three or more scans.

RESPONSE TO REVIEWER COMMENTS

Reviewer #1: The authors have addressed the majority of my questions. This is an interesting, comprehensive manuscript that improves our understanding of sulcal development and will be of interest to many readers. In particular, I appreciate the expansion of the methods to improve replication and the addition of a limitations section. As a minor point, there remains some inconsistencies with significant digits which should be corrected prior to publication.	**Response：** We thank the reviewer for this careful observation. We have thoroughly re-examined the entire manuscript and implemented consistent numerical reporting standards throughout. Specifically, correlation coefficients (r, r_partial) and regression weights (β) are reported to two decimal places, while Δadj R^2 and p-values are reported to three decimal places (with values below 0.001 expressed as < 0.001, and exceptionally small p_FDR values retained in scientific notation). In response to this comment, two main corrections were made. First, several entries in Table 1 were previously expressed in scientific notation (e.g., -3.27×10^{-4}, 2×10^{-4}) due to negligible, non-significant effect sizes. These have now been standardized to three decimal places (i.e., 0.000), consistent with all other entries in the table. Second, the subscript notation for FDR-corrected p-values was inconsistently capitalized across the manuscript, appearing alternately as p_fdr and p_FDR. All instances have now been unified to p_FDR throughout the text. We believe these corrections fully address the reviewer's concern, and we once again thank the reviewer for this valuable and constructive feedback.
Reviewer #2: The authors have responded adequately to my feedback from the last review and made the necessary revisions to their manuscript. **Reviewer #3:** I thank the authors for their careful response to my comments; and I believe, in doing so, they have much improved the manuscript.	**Response：** We sincerely thank both reviewers for their positive assessment of the revised manuscript and for confirming that their previous concerns have been adequately addressed.