nature portfolio

Peer Review File

Brain Regions with Gestational Age Differences Mediate Cognition in Adolescents Born Very Premature

Corresponding Author: Dr Dawn Jensen

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)
Thank you for the opportunity to review this manuscript, which examines neuronal alteration associated with moderate and very preterm birth compared to full-term birth, their mediating role on cognitive outcomes at 9 years of age, and their trajectories from 9 to 13 years. The authors leverage a large longitudinal sample and use an interesting statistical approach, applying canonical correlation analysis to derive CVs from ROI-based GM volumes and from both ROI- and tract-based WM FA measures. The results are clearly presented and well discussed. I really enjoyed reading this manuscript.

I have a few comments, mostly aimed at helping the reader follow the different steps of the study more easily.

Introduction
Page 2 - “While these most recent studies provide evidence that neurodevelopmental effects associated with premature birth seen in infants are still present in adolescence, they fail to elucidate the possible connection between these regional brain differences and the cognitive deficits seen in premature subjects.”- It would be helpful to go a bit deeper into the existing literature, especially regarding socio-emotional functioning (since CBCL measures are included in this study), for example: *Montagna, A., & Nosarti, C. (2016). Socio-emotional development following very preterm birth: pathways to psychopathology. Frontiers in psychology, 7, 80.

Page 2 - “These subjects had both grey matter volume (GM) and fractional anisotropy (FA) measures available from the Freesurfer analyses performed by the ABCD study (total n = 2128) from three timepoints (9-14 yrs old), as well as cognitive and behavioral measures from the child behavioral checklist and the NIH toolbox” - Since methods are already introduced here, it would help the reader to specify more clearly what GM and FA refer to in this study (e.g., GM volume; FA from ROI-based analyses and tract-based analyses).

Results
- It might be due to the submission format, but the supplementary tables are missing captions.
- It would be very helpful to include brain illustrations showing the five brain regions (GM regions and FA regions/tracts) with the highest loadings for each CV.
- The FA results are not always clear in terms of whether they refer to ROI-based or tract-based measures. Please clarify this throughout.
- Figure 1: I would suggest specifying what higher or lower scores mean (e.g., higher contribution to the CV?).
- There is no clear reporting of results related to the CBCL. If there were no significant group differences, this should be explicitly stated and reported.
- For clarity, I suggest adding “baseline” or “at 9 years” in the sections subtitle that refer only to the baseline assessment.

Discussion
The mediation analyses show quite lateralised results, with a left-dominant pattern. Could you please elaborate on this point?
In addition, it seems that there were no significant group differences on the CBCL, although many studies report socio-emotional difficulties in adolescents born very preterm. Could you please discuss this in more detail?

Reviewer #2

(Remarks to the Author)
The study investigates the relation between prematurity, co-varying cortical volume and white matter microstructure, cognition, and broad mental health dimensions in late childhood. It also tests whether prematurity is associated with longitudinal changes in brain structure across late childhood and adolescence. The findings include cross-sectional and longitudinal structural differences between term-born and prematurely born youth, as well as indirect effects of prematurity on cognition via specific grey matter ROIs. I find the overall topic of pre-/perinatal links to brain and cognitive outcomes timely, and the paper is for the most part well structured and well written. I have several comments that I hope will be helpful, and please correct me wherever I have misunderstood. Finally, I would strongly recommend pre-registering future studies involving a similar level of exploratory analyses.

1. Sample-specific information on nationality, age range, and sex distribution is missing from the abstract. ABCD is a well-known cohort, and not specifying it in the abstract initially gave the impression that the sample was prematurity-enriched or otherwise selected.

2. As cognition and broader mental health dimensions are central to the paper, it would be helpful to include information in the introduction on how prematurity relates to these outcomes. Particularly for mild to moderate prematurity (i.e., individuals born “only” a few weeks early), which is relevant for the current sample.

3. Do I understand correctly that, rather than focusing on effects most strongly associated with prematurity per se, the authors specifically target co-varying patterns between cortical volume and white matter microstructure, discarding potentially stronger modality-specific effects? Even within T1w morphometry, different imaging metrics are only moderately correlated, and similar considerations likely apply for white matter diffusion. This choice becomes more difficult to interpret given that, after identifying the strongest cross-matter couplings, these patterns are “uncoupled” i.e. GM and FA regions are assessed separately for the majority of analyses. It is unclear why, unless this reflects practical difficulties in mapping CCA weights back to brain space. As currently presented, the role of performing CCA in the first place is difficult to interpret.

4. Could the authors clarify these CCA-related methodological aspects: 1. Were imaging variables scaled prior to CCA? If not, the results may primarily reflect variance differences rather than genuine GM–FA covariation. 2. In the split-half procedure, were canonical weights first estimated in the discovery sample and then applied to the replication sample to compute canonical scores? If not could the authors please present such analyses? My understanding is that re-running CCA independently in the replication set would not test generalizability/out-of-sample validity of the identified canonical patterns. (3) How was statistical significance of the CCs assessed, and was correction applied for the number of canonical pairs?

5. Given that GM volume is influenced by sex and head size, did the authors consider residualizing ROIs for such covariates prior to CCA? As expected it appears the CCA loadings show stronger effects with sex than prematurity (Supplementary table 1). In my opinion the ROIs should have been adjusted for scanner bias before CCA, could the authors present the effect of included scanners (not scan-site) for the CCs in Supplementary Table 1? Finally, please present all effect sizes and p values within the supplementary tables, now I believe only statistically significant findings are reported.

6. The neuroimaging methods section would benefit from better structure and detail. Please specify scanner field strength, manufacturers, and specifically the total number of individual scanners used in the current study, as each introduce bias. In addition, information on quality control is missing, including how many participants were excluded based on ABCD or study-specific QC criteria. Finally, I belive MMPS is not a standard pipeline but created by ABCD specifically, and it would be helpful to still present the processing steps in order. Also it is stated: “Desikan atlas [14]. Subcortical and cortical regional values were then extracted from the atlas for a total of 94 regions.” Were subcortical white matter desikan regions also used in the study?

7. Major: It is unexpected that broad mental health dimensions are introduced in the methods section, as these outcomes are not motivated or addressed in the abstract or introduction. Moreover, if I understand correctly, results for these analyses are not presented. This may give an unfortunate impression of selective reporting. Please document the findings of all analyses performed and discuss null and positive findings equally.

8. With regards to reference 8, the authors state that: “Previous research using this cohort... Due to the univariate methods they applied, they failed to find significant effects”. If I understand correctly, the current study also relies largely on univariate analyses of individual ROIs. Also what is meant by “failing to find significant effects”?

9. Minor: In addition to ref 8 it might be relevant to discuss findings in relation to our paper which also used the ABCD sample to assess prematurity and multivariate imaging, combined by linked-ICA: “Linking pregnancy- and birth-related risk factors to a multivariate fusion of child cortical structure” (https://doi.org/10.1073/pnas.2422281122).

10. Minor: As statistical significance is documented in the results section and implied for reported findings, it may improve readability to reduce the extensive usage of the term “significant(ly)”. e.g., “Adolescents born very premature have significant differences within correlated patterns of GM and FA.” and “significantly less well”.

11. Minor: It is stated that “Supplementary Table 1 contains the full details of the CVs”. Consider adding a table legend and instead state that is shows certain effect sizes or the like.

12. Minor: Consider removing summation of actual findings currently presented in introduction: “Our findings reveal persistent structural brain differences in regions of the brain that are critical to attentional processes, memory, cognitive control, problem solving, and information processing. These differences not only strongly mediated the relationship with cognitive deficits related to GA, but also persisted across four years of development with little change in trajectory, suggesting that people born very premature may experiences differences in neuronal structure throughout their lifetime.”

13. Minor: Why was scan-site also added as a random effect to models that did not include imaging?

14. Minor: What is meant by “disrupted” cortical folding?

15. Minor: I believe Figure 1 and 2 should be moved to the Results section.

16. Minor: Were limiting the amounts of scanners considered when matching full term and pre-term youths?

17. Minor: What is meant by: “The ABCD Study has... (dMRI) with well calibrated scanning parameters?”

18. Minor: For readability, consider reporting p-values in decimal rather than exponential (“p<8.05e-6”) notation.

Reviewer #3

(Remarks to the Author)
The manuscript Brain Regions with Gestational Age Differences Mediate Cognition in Adolescents Born Very Premature by Jensen and colleagues describes a study in which the authors leverage data from the ABCD study to investigate whether being born prematurely has lasting effects on brain structure and cognitive function. It has several strengths including the sample size and a very comprehensive analysis. There are a number of issues that came up while reading the manuscript, including a lack of detail or clarity in the methods, missing results, and conclusions that cannot be made based on the results found. I also feel that the manuscript needs heavy revising in terms of writing. It's difficult to assess whether a revision would lead to a publishable manuscript, because that depends on clarification of the methods. I've outlined my specific comments below.

1. In the methods, the authors state: Two summary cognitive measures, fluid and crystalized cognition, were derived from the battery of NIH Toolbox tests given to subjects. These outcomes spanned cognitive domains such as working memory, executive function, attention, episodic memory, language, and processing speed [16]. The fluid cognition summary score represents the ability to learn new information, while the crystalized cognition summary score represents already acquired knowledge [16]." How were the summary cognitive measures computed? Did all tests contribute to both, or did some contribute to fluid and others contribute to crystalized cognition?

2. The definition for fluid cognition (which I'm assuming is similar to fluid intelligence) doesn't seem correct. Fluid cognition generally refers to the ability to reason, manipulate information, and solve problems including in the abstract without prior knowledge. It would be assesed with tests of attention, working memory, etc. The authors definition that it is the ability to learn new information seems at odds with the accepted definition. Could the authors explain why they feel that fluid cognition captures learning, instead of how it is usually operationalized?

3. In the methods, the authors define age as the following: "GA status was the independent variable, sex and age (in months to account for birth differences) were the fixed effects, and site was the random effect." I assume the authors are referring to the age of the participant at the time of the brain scan, rather than at birth. If so, how would reporting age in months account for birth differences differently to reporting them as years (i.e., what is the difference between 10.25 years and 123 months)? If this is the case, wouldn't that not account for birth differences, but instead it would account for age at scan? If the authors are referring to postmenstrual age rather than age at scan, then this is moot, but right now I'm completely unclear on how age is being included in the models.

4. Was CCA performed on all time points' imaging data, or just the 9-10 year old time point data? If all data, how was shared variance across time points accounted for? And under the following heading, "Cross-Sectional Analyses of Baseline Measures", is it correct to assume that this is only 9-10 year old data? It's stated in the heading that it's a cross-sectional analysis but the age group is not explicitly described in the text, and there are three sets of cross-sectional data and it's unclear which data the CCA was performed on in the first place.

5. I'd recommend formatting the equations for easier reading.

6. The results do not describe the null findings for mediating effects of FA on cognitive outcomes.

7. Figure 3B should clarify that it's referring to grey matter volume.

8. In the discussion, the authors state that "Deficits in these regions, both for GM and FA would disrupt and degrade problem solving skills, as well as make retention and recall of prior knowledge more difficult." However, they did not find a relationship between FA and cognitive function, despite finding group differences in FA. Can the authors include discussion of why (as they state) deficits in FA would disrupt and degrade cognitive skills, yet they didn't find that here?

9. The authors state "These GA-dependent differences were reduced GM and FA in regions of the brain that are critical to

attentional processes, memory, cognitive control, problem solving, and information processing These differences not only strongly mediated the relationship with cognition but also persisted across four years of development with little change in trajectory." I don't feel like this is a valid interpretation of the results. There were reductions in FA, however they did not mediate the relationship with cognition. Additionally, calling a partial mediation of 15% "strong" feels overstated and would generally be classified as a weak or low level of variance accounted for.

10. As an aside, the manuscript would greatly benefit from editing for spelling and grammar. As a few examples, I've shared some sentences below that made the manuscript difficult to read:
Grey matter volume (GM) and fractional anisotropy (FA) derivatives provided by the ABCD study.
Based on the current research, we believe that investigating correlated patterns of grey and white matter in adolescents born premature will revealed brain regions with GA-dependent differences.
GM measures were obtained from T1-weighted sMRI images underwent gradient distortion correction
We implemented a discovery and method, where the data was split randomly 50/50, to correct for overinflation [12]
Subject's GM and FA measures from three time points (baseline, 2nd year, and 4th year follow ups) were the dependent variables
in CV5, where very premature adolescents significantly different in both GM ($t=3.93$, $p<8.93e-5$) and FA ($t=3.17$, $p<0.0015$) compared to full-term
While able to leverage a large diverse cohort and perform both longitudinal and cross-sectional analyses, there are some limitations to this research

Version 1:

Reviewer comments:

Reviewer #2

(Remarks to the Author)
I have no further comments.

Reviewer #3

(Remarks to the Author)
The authors have address my concerns.

Reviewer 1

Thank you for the opportunity to review this manuscript, which examines neuronal alteration associated with moderate and very preterm birth compared to full-term birth, their mediating role on cognitive outcomes at 9 years of age, and their trajectories from 9 to 13 years. The authors leverage a large longitudinal sample and use an interesting statistical approach, applying canonical correlation analysis to derive CVs from ROI-based GM volumes and from both ROI- and tract-based WM FA measures. The results are clearly presented and well discussed. I really enjoyed reading this manuscript.

I have a few comments, mostly aimed at helping the reader follow the different steps of the study more easily.

Thank you for the positive feedback and comments.

1) Page 2 - “While these most recent studies provide evidence that neurodevelopmental effects associated with premature birth seen in infants are still present in adolescence, they fail to elucidate the possible connection between these regional brain differences and the cognitive deficits seen in premature subjects.”- It would be helpful to go a bit deeper into the existing literature, especially regarding socio-emotional functioning (since CBCL measures are included in this study), for example:
*Montagna, A., & Nosarti, C. (2016). Socio-emotional development following very preterm birth: pathways to psychopathology. Frontiers in psychology, 7, 80.

We appreciate the opportunity to rectify this oversight. The introduction as well as discussion have been changed to address the reviewer’s concerns:

A comprehensive review from 2016 highlights the socio-emotional deficits found in adolescents born very premature [38]. We did not detect differences or deficits in behavioral problems, which may be attributable to two factors. Firstly, this study has a much larger number of subjects than most of the previous studies. This could lead to a smaller effect size across a much more heterogeneous cohort, while studies with a smaller number of subjects may have inflated effect sizes [42]. We also limited our investigation of socio-emotional deficits to the summary scores of internalizing and externalizing behaviors. Future work should investigate more specific behaviors, for example depression, social anxiety, or oppositional disorders.

2) Page 2 - “These subjects had both grey matter volume (GM) and fractional anisotropy (FA) measures available from the Freesurfer analyses performed by the ABCD study (total n = 2128) from three timepoints (9-14 yrs old), as well as cognitive and behavioral measures from the child behavioral checklist and the NIH toolbox” - Since methods are already introduced here, it would help the reader to specify more clearly what GM and FA refer to in this study (e.g., GM volume; FA from ROI-based analyses and tract-based analyses)

The introduction was changed to include the further clarification of the Freesurfer parcellations:

These subjects had both grey matter volume (GM) and fractional anisotropy (FA) measures from the Freesurfer cortical and subcortical parcellations, and in the case of FA, additional tract-

based parcellations from Atlas Tracts [39]. These measures were calculated by the ABCD study (total n = 2128)…

[39] Hagler DJ Jr, Hatton S, Cornejo MD, Makowski C, Fair DA, Dick AS, Sutherland MT, Casey BJ, Barch DM, Harms MP, Watts R, Bjork JM, Garavan HP, Hilmer L, Pung CJ, Sicat CS, Kuperman J, Bartsch H, Xue F, Heitzeg MM, Laird AR, Trinh TT, Gonzalez R, Tapert SF, Riedel MC, Squeglia LM, Hyde LW, Rosenberg MD, Earl EA, Howlett KD, Baker FC, Soules M, Diaz J, de Leon OR, Thompson WK, Neale MC, Herting M, Sowell ER, Alvarez RP, Hawes SW, Sanchez M, Bodurka J, Breslin FJ, Morris AS, Paulus MP, Simmons WK, Polimeni JR, van der Kouwe A, Nencka AS, Gray KM, Pierpaoli C, Matochik JA, Noronha A, Aklin WM, Conway K, Glantz M, Hoffman E, Little R, Lopez M, Pariyadath V, Weiss SR, Wolff-Hughes DL, DelCarmen-Wiggins R, Feldstein Ewing SW, Miranda-Dominguez O, Nagel BJ, Perrone AJ, Sturgeon DT, Goldstone A, Pfefferbaum A, Pohl KM, Prouty D, Uban K, Bookheimer SY, Dapretto M, Galvan A, Bagot K, Giedd J, Infante MA, Jacobus J, Patrick K, Shilling PD, Desikan R, Li Y, Sugrue L, Banich MT, Friedman N, Hewitt JK, Hopfer C, Sakai J, Tanabe J, Cottler LB, Nixon SJ, Chang L, Cloak C, Ernst T, Reeves G, Kennedy DN, Heeringa S, Peltier S, Schulenberg J, Sripada C, Zucker RA, Iacono WG, Luciana M, Calabro FJ, Clark DB, Lewis DA, Luna B, Schirda C, Brima T, Foxe JJ, Freedman EG, Mruzek DW, Mason MJ, Huber R, McGlade E, Prescot A, Renshaw PF, Yurgelun-Todd DA, Allgaier NA, Dumas JA, Ivanova M, Potter A, Florsheim P, Larson C, Lisdahl K, Charness ME, Fuemmeler B, Hettema JM, Maes HH, Steinberg J, Anokhin AP, Glaser P, Heath AC, Madden PA, Baskin-Sommers A, Constable RT, Grant SJ, Dowling GJ, Brown SA, Jernigan TL, Dale AM. Image processing and analysis methods for the Adolescent Brain Cognitive Development Study. Neuroimage. 2019 Nov 15;202:116091. doi: 10.1016/j.neuroimage.2019.116091. Epub 2019 Aug 12. PMID: 31415884; PMCID: PMC6981278.

3) It might be due to the submission format, but the supplementary tables are missing captions.

Supplementary Materials have been updated with comprehensive model information.

4) It would be very helpful to include brain illustrations showing the five brain regions (GM regions and FA regions/tracts) with the highest loadings for each CV.

Thank you for the excellent suggestion. Figure 1 (all updated figures are at the end of reviewer comments) now reflects the CV results and comprehensive model details have been included in Supplementary Materials 1.

5) The FA results are not always clear in terms of whether they refer to ROI-based or tract-based measures. Please clarify this throughout.

We apologize for the confusion. The manuscript has been updated to clarify which atlases were used for both GM and FA measures.

6) Figure 1: I would suggest specifying what higher or lower scores mean (e.g., higher contribution to the CV?).

The caption for Figure 1 was amended to include more clarity regarding the meaning of CV scores:

Figure 1: Canonical Variates with GA-dependent differences, regions with GM volume differences highlighted in blue, regions with FA differences highlighted in yellow, seen in **A**, **E**, **H**, and **K**. Each CVs top 5 regions with loadings as shown in **B**, **F**, **I**, and **L**. Dotplots showcase GA-dependent differences in CVs, where the asterisk highlights significance. Full term in red, moderately premature are in grey, and very premature in orange. In CV2, (plot **C**, GM (t=3.50, p<0.0005) and FA (t=5.19, p<2.25^-7)), CV5 (plot **G**, FA (t=3.17, p<0.002)), and CV7 (plot **J**, GM (t=3.38, p<0.002)), very premature

adolescents contributed more to the correlated patterns of GM and/or FA than their peers. In CV8 (plot **M**, FA ($t=-4.48$, $p<0.0007$)), moderately premature adolescents contributed less than their peers to the correlated patterns of that CV.

7) There is no clear reporting of results related to the CBCL. If there were no significant group differences, this should be explicitly stated and reported.

We apologize for the oversight. Results were updated to reflect the lack of significant group differences in CBCL measures. Also Supplementary Materials 3 contains the comprehensive model results for behavioral measures:

There were no significant GA differences found for subjects' internalizing and externalizing scores.

8) For clarity, I suggest adding "baseline" or "at 9 years" in the sections subtitle that refer only to the baseline assessment.

Headings for sections subtitles were updated to more accurately reflect relevant ages:

Adolescents born very premature have significant differences within correlated patterns of GM and FA at 10 years old:

Brain region-specific GA-dependent differences at 10 years old:

GA-dependent differences in behavior and cognition at 10 years old:

Brain regions with GA-dependent differences mediate the relationship between GA and cognition at 10 years old:

GA-dependent regional brain differences persist from ages 10-14:

9) The mediation analyses show quite lateralised results, with a left-dominant pattern. Could you please elaborate on this point?

Thank you for the astute observation. Due to another reviewer's suggestions, updated results after the addition of the intercranial volume variable for GM models removed this lateralization.

10) In addition, it seems that there were no significant group differences on the CBCL, although many studies report socio-emotional difficulties in adolescents born very preterm. Could you please discuss this in more detail?

The discussion has been amended to include the following:
A comprehensive review from 2016 highlights the socio-emotional deficits found in adolescents born very premature [38]. This was not the case for this cohort, which may be for two reasons. Firstly, this study has a much larger number of subjects than most

of the previous studies. This could lead to a smaller effect size for the behavioral linear models than in studies with smaller subjects, which can inflate effect sizes [42]. We also limited our investigation of socio-emotional deficits to the summary scores of internalizing and externalizing behaviors. Future work should investigate more specific behaviors, for example depression, social anxiety, or oppositional disorders.

[38] Montagna A, Nosarti C. Socio-Emotional Development Following Very Preterm Birth: Pathways to Psychopathology. Front Psychol. 2016 Feb 12;7:80. doi: 10.3389/fpsyg.2016.00080. PMID: 26903895; PMCID: PMC4751757.

[42] Rochefort-Maranda, G. Inflated effect sizes and underpowered tests: how the severity measure of evidence is affected by the winner's curse. Philos Stud 178, 133–145 (2021). https://doi.org/10.1007/s11098-020-01424-z

Reviewer 2

The study investigates the relation between prematurity, co-varying cortical volume and white matter microstructure, cognition, and broad mental health dimensions in late childhood. It also tests whether prematurity is associated with longitudinal changes in brain structure across late childhood and adolescence. The findings include cross-sectional and longitudinal structural differences between term-born and prematurely born youth, as well as indirect effects of prematurity on cognition via specific grey matter ROIs. I find the overall topic of pre-/perinatal links to brain and cognitive outcomes timely, and the paper is for the most part well structured and well written. I have several comments that I hope will be helpful, and please correct me wherever I have misunderstood. Finally, I would strongly recommend pre-registering future studies involving a similar level of exploratory analyses.

Thank you for your consideration and feedback. Your suggestions were very helpful and made the analysis and manuscript better.

1) Sample-specific information on nationality, age range, and sex distribution is missing from the abstract. ABCD is a well-known cohort, and not specifying it in the abstract initially gave the impression that the sample was prematurity-enriched or otherwise selected.

We apologize for the oversight. The abstract has been updated to identify the ABCD cohort:

Using 2100 subjects from the Adolescent Brain and Cognitive Development study, a longitudinal cohort of adolescents (aged 10-14),

2) As cognition and broader mental health dimensions are central to the paper, it would be helpful to include information in the introduction on how prematurity relates to these outcomes. Particularly for mild to moderate prematurity (i.e., individuals born “only” a few weeks early), which is relevant for the current sample.

Thank you for the feedback, the introduction was amended accordingly:

Previous research has also established that there are socio-emotional as well as cognitive deficits associated with premature birth that seem to persist into adulthood [38, 43]. These behavioral problems can manifest in adolescence as attention problems, anxiety, or poor emotional regulation [38]. The cognitive domains often impacted by very premature birth include memory, executive function, inhibitory control, and cognitive flexibility [43].

While these most recent studies provide evidence that neurodevelopmental effects associated with premature birth seen in infants are still present in adolescence, they fail to elucidate the possible connection between these regional brain differences and the behavioral and cognitive deficits seen in premature subjects [38].

3) Do I understand correctly that, rather than focusing on effects most strongly associated with prematurity per se, the authors specifically target co-varying patterns between cortical volume and white matter microstructure, discarding potentially stronger modality-specific effects? Even within T1w morphometry, different imaging metrics are only moderately correlated, and similar considerations likely apply for white matter diffusion. This choice becomes more difficult to interpret given that, after identifying the strongest cross-matter couplings, these patterns are “uncoupled” i.e. GM and FA regions are assessed separately for the majority of analyses. It is unclear why, unless this reflects practical difficulties in mapping CCA weights back to brain space. As currently presented, the role of performing CCA in the first place is difficult to interpret.

We apologize for the lack of clarity. The introduction has been updated to include more specifics regarding the rationale:

To extend the current understanding of GA-dependent neuronal differences in early adolescence and their relationship to cognitive and behavioral deficits, we have chosen an a approach that allows consideration of the developmental relationship between changes in GM volume and FA measures during adolescence. While not in lockstep with one another, there is research demonstrating that maturation-driven decreases in GM volume occur in coordination with increases in FA [27, 28]. To the end, we applied a canonical correlation analysis (CCA), that will allow us to first focus on GA-dependent differences in highly correlated patterns of grey matter and white matter measures. A CCA finds maximally correlated linear combinations of two variables, highlighting data-driven multivariate relationships [11]. Since both grey matter volume and white matter integrity have been shown to be changing in synchrony, if diametrically, during adolescence [27, 28], leveraging the shared information between the two modalities would highlight developmentally relevant brain regions to then investigate longitudinally for GA-driven neuronal differences and their role in cognitive and behavioral performance.

[11] Zhuang, X., Yang, Z., & Cordes, D. (2020). A technical review of canonical correlation analysis for neuroscience applications. *Human Brain Mapping*, *41*(13), 3807–3833. https://doi.org/10.1002/hbm.25090

[27] Winkler, A. M., Kochunov, P., Blangero, J., Almasy, L., Zilles, K., Fox, P. T., Duggirala, R., & Glahn, D. C. (2010). Cortical thickness or grey matter volume? The importance of selecting the phenotype for imaging genetics studies. *NeuroImage*, *53*(3), 1135–1146. https://doi.org/10.1016/j.neuroimage.2009.12.028

[28] Bava, S., Thayer, R., Jacobus, J., Ward, M., Jernigan, T. L., & Tapert, S. F. (2010). Longitudinal characterization of white matter maturation during adolescence. *Brain Research*, *1327*, 38–46.

4) 1. Were imaging variables scaled prior to CCA? If not, the results may primarily reflect variance differences rather than genuine GM–FA covariation. 2. In the split-half procedure, were canonical weights first estimated in the discovery sample and then applied to the replication sample to compute canonical scores? If not could the authors please present such analyses? My understanding is that re-running CCA independently in the replication set would not test generalizability/out-of-sample validity of the identified canonical patterns. (3) How was statistical significance of the CCs assessed, and was correction applied for the number of canonical pairs?

Thank you for the feedback, more specifics regarding the CCA have now been included in the method section to answer your questions:

1) The imaging variables were scaled prior to the CCA

2) Canonical weights were first estimated in the discovery sample and applied to the replication sample to compute the canonical scores
3) A Barlett's test was used to determine the significance of the CCs and then corrected for the number of CVs.

5) Given that GM volume is influenced by sex and head size, did the authors consider residualizing ROIs for such covariates prior to CCA? As expected it appears the CCA loadings show stronger effects with sex than prematurity (Supplementary table 1). In my opinion the ROIs should have been adjusted for scanner bias before CCA, could the authors present the effect of included scanners (not scan-site) for the CCs in Supplementary Table 1? Finally, please present all effect sizes and p values within the supplementary tables, now I believe only statistically significant findings are reported.

We appreciate the thoughtful feedback on our analysis. We decided to preserve as much biological variance as possible in the imaging measures, which would preclude regressing sex differences out before the CCA. Site was included rather than scanner to keep the correction consistent between variables and models (cognition and behavior

U1scores ~ group + GestAge + ICs + Sex + (1|scanner)

	U 1 scores		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-1.00	-1.68 – -0.31	**0.004**
Sex	0.07	-0.02 – 0.15	0.141
groupnew [pre A]	0.10	0.02 – 0.19	**0.017**
groupnew [pre B]	0.04	-0.12 – 0.19	0.640
GestAge	0.01	0.00 – 0.01	**0.021**
ICs	0.34	0.29 – 0.38	**<0.001**
Random Effects			
σ^2	0.82		
$\tau_{00\ scanner}$	0.05		
ICC	0.06		
$N_{scanner}$	3		
Observations	2091		
Marginal R^2 / Conditional R^2	0.116 / 0.168		

U1scores ~ group + GestAge + ICs + Sex + (1|site)

	U 1 scores		
Predictors	*Estimates*	*CI*	*p*
(Intercept)	-0.81	-1.45 – -0.16	**0.015**
Sex	0.08	-0.01 – 0.17	0.071
groupnew [pre A]	0.05	-0.04 – 0.14	0.259
groupnew [pre B]	0.01	-0.15 – 0.16	0.931
GestAge	0.01	-0.00 – 0.01	0.053
ICs	0.33	0.29 – 0.38	**<0.001**
Random Effects			
σ^2	0.81		
$\tau_{00\ site}$	0.05		
ICC	0.05		
N_{site}	22		
Observations	2130		
Marginal R^2 / Conditional R^2	0.111 / 0.158		

Model Comparison Figure: Random effects residuals show that site and scanner account for similar amounts of variance in the model.

also require a correction for multiple sites). A comparison of random effects for a model that includes scanner rather than site shows that the site correction is comparable to the scanner correction (see included figure below). Results have been updated to include intercranial volume in the models that include grey matter volume (see below for manuscript changes). Supplementary Materials have been updated to include all models (see Supplementary Materials 1-4).

Updated Results:

Adolescents born very premature have significant differences within correlated patterns of GM and FA at 10 years old: Eight pairs of significantly correlated CVs, after Bonferroni correction, replicated for this cohort, with the minimum correlation (r) of 0.49 and a maximum correlation (r) of 0.74. **Supplementary Materials 1** contains the full details of the CVs. Of the replicated CVs, four showed significant group differences. These differences were seen in CV2, where very premature adolescents contributed more to the correlated patterns in both GM (t=3.50, p<0.0005) and FA (t=5.19, p<2.25^-7); in CV5, where very premature adolescents contributed more to the FA (t=3.17, p<0.002); in CV7, where very premature adolescents contributed more to the patterns of GM (t=3.38m p<0.0007); and in CV8, where moderately premature adolescents contributed less to the patterns of FA (t=-4.48, p<8.11^-6). CV2 highlighted correlated patterns between regions of GM (highest 5 contributors) in the left thalamus (DSK), left putamen (DSK), left caudate (DSK), right medial orbitofrontal gyrus (DSK), left anterior cingulate (DSK) and FA (highest five contributors) in the right superior longitudinal fasiculus (slf) (AT), right superior corticostriate (DSK), left thalamus (DSK), right inferior frontal fasiculus (AT), and left orbitofrontal cortex (DSK). The top five regions of FA in CV5 consisted the right superior cotricostriate (DSK), right insula (DSK), left superior corticostriate (DSK), right anterior cingulate (DSK), and left slf (AT). The top five regions of GM in CV7 were right superior frontal cortex (DSK), left orbitofrontal cortex (DSK), right cerebellar cortex (DSK), right putamen (DSK), and right pars triangularis (DSK).The top five contributors to the FA patterns highlighted in CV8 were right superior corticostriate (DSK), right slf (AT), left slf (AT), left superior corticostriate (DSK), and right anterior cingulate (AT). **Figure 1** shows the significant correlated patterns GM and FA of these CVs, as well as the region loadings and GA-related differences in the subject contributions to these patterns.

Brain region-specific GA-dependent differences at 10 years old: Results of the linear regression investigating group differences in GM and FA within the regions with that contributed to the correlated patterns in CV2, CV5, CV7, and CV8 were seen primarily in GM after correction for multiple testing. Significant differences were found in the GM of the left thalamus the moderately (t=-2.95, p<0.003) and very premature group (t=-3.26, p<0.001), the GM of the left anterior cingulate for the moderately premature (t=-3.47, p<0.0005), and the GM of the right cerebellar cortex for the very premature (T=-3.17, p<0.003). Significantly higher FA was found in the very premature group in the right anterior cingulate (DSK) (t=3.66, p<0.0003). **Figures 2 A-B** highlight the cross-sectional differences between GA groups in the left anterior cingulate and right cerebellar cortex. Complete model results are found in **Supplementary Materials 2**.

GA-dependent differences in behavior and cognition at 10 years old: Significant group differences were identified in cognitive performance after multiple testing corrections, with moderately premature (t=-4.1, p<3.2e-5) and very premature (t=-6.9, p<7.2e-12) adolescents both performing significantly less well than their full-term peers in crystalized cognition. With regard to fluid cognition, only the very premature (t=-6.00, p<2.3e-9) adolescents performed significantly less well. **Figure 3A** shows the differences in the cross-sectional performance of the GA groups on the fluid and crystalized cognition metric. There were no significant GA differences found for subjects' internalizing and externalizing scores. **Supplementary Materials 3** contains the full model results.

Brain regions with GA-dependent differences mediate the relationship between GA and cognition at 10 years old: To determine if these group differences in the brain might mediate cognitive performance, we first performed linear regressions to test for significant relationships between these brain regions and fluid and crystalized cognition. Mediation analyses were only performed on brain regions that were significantly linearly associated with cognition. The GM volume in the right cerebellar cortex mediated the relationship between GA and fluid cognition

by 11% and crystalized cognition by 15%. The GM volume in the left anterior cingulate mediated the relationship between GA and crystalized cognition by 11%. **Figure 3B-C** highlights the mediated effects found in both regions.

GA-dependent regional brain differences persist from ages 10-14: Of the four brain regions with significant cross-sectional GA-dependent differences, two showed significant longitudinal differences in the repeated measures mixed-effects models. The GM volume in the left thalamus, very premature adolescents had lower gray matter volume, ($t=-2.99$, $p<0.003$), across time. Conversely, in the right anterior cingulate, very premature adolescents had higher FA, ($t=3.89$, $p<0.0001$), than their peers. These GA-dependent differences can be seen in **Figures 2C-D**. Complete model details are found in **Supplementary Materials 4**.

6) Please specify scanner field strength, manufacturers, and specifically the total number of individual scanners used in the current study, as each introduce bias. In addition, information on quality control is missing, including how many participants were excluded based on ABCD or study-specific QC criteria. Finally, I belive MMPS is not a standard pipeline but created by ABCD specifically, and it would be helpful to still present the processing steps in order. Also it is stated: “Desikan atlas [14]. Subcortical and cortical regional values were then extracted from the atlas for a total of 94 regions.” Were subcortical white matter desikan regions also used in the study?

Thank you for the suggestions. Atlas specifics were included per the first reviewer to clarify which atlases were used for both the GM and FA. For FA, both AtlasTract regions as well as the cortical and subcortical desikan regions were used. The methods were updated to include more information regarding the scanners and preprocessing:

The ABCD Study has collected structural MRI (sMRI), and diffusion MRI (dMRI), across 22 sites with scanning parameters as harmonized as possible with three different scanner types (GE, Philips, and Siemens). GM measures were obtained from T1-weighted collected with the following parameters: slices = 176–256, field of view = 256 × 240–256, TE = 2.8 ms, flip angle = 8° [14]. FA measures were calculated from dMRI collected with the following parameters: slices = 81, field of view = 240 x 240, TE = 88 ms, flip angle = 90°, diffusion directions = 96, b-values = 3000 [14]. These images were then processed using the Multimodal Processing Stream (MMPS), software designed to process large-scale multimodal neuroimaging data [14]. This pre-processing for the T1-weighted images included gradient distortion correction, intensity normalization, tissue segmentation, and then rigid alignment with a FreeSurfer (v7.1.1) Desikan atlas [14]. The dMRI images were corrected for eddy current distortion, head motion, and B0 distortions, and hen resampled to standard orientation [15]. Tensor values were calculated from the full shell, using all b values, for microstructural tissue measures that included FA [15]. GM volume measures for subcortical and cortical regional values were then extracted from the FreeSurfer Desikan atlas for a total of 94 regions, while FA values were extracted from both the Desikan (DSK) atlas for cortical and subcortical regions as well as the AtlasTract (AT) for tract specific measures for a total of 127 regions [15]. Image quality ratings were calculated from pial and white

matter overestimation, artifacts, and inhomogeneity. Image artifacts were also quantified, as was head motion [14]. Subjects who had ratings that indicated excessive motion, artifacts, or sub-par image quality were excluded. These quality control and FreeSurfer derivate measure were provided by the ABCD study in the 5.0 Release.

7) Major: It is unexpected that broad mental health dimensions are introduced in the methods section, as these outcomes are not motivated or addressed in the abstract or introduction. Moreover, if I understand correctly, results for these analyses are not presented. This may give an unfortunate impression of selective reporting. Please document the findings of all analyses performed and discuss null and positive findings equally.

Thank you for pointing out this oversight. The introduction was amended to reflect the following:

Previous research has also established that there are socio-emotional as well as cognitive deficits associated with premature birth that seem to persist into adulthood [38, 43]. These behavioral problems can manifest in adolescence as attention problems, anxiety, or poor emotional regulation [38]. The cognitive domains often impacted by very premature birth include memory, executive function, inhibitory control, and cognitive flexibility [43].

While these most recent studies provide evidence that neurodevelopmental effects associated with premature birth seen in infants are still present in adolescence, they fail to elucidate the possible connection between these regional brain differences and the behavioral and cognitive deficits seen in premature subjects [38].

The discussion was amended as well:

A comprehensive review from 2016 highlights the socio-emotional deficits found in adolescents born very premature [38]. This was not the case for this cohort, which may be for two reasons. Firstly, this study has a much larger number of subjects than most of the previous studies. This could lead to a smaller effect size for the behavioral linear models than in studies with smaller subjects, which can inflate effect sizes [42]. We also limited our investigation of socio-emotional deficits to the summary scores of internalizing and externalizing behaviors. Future work should investigate more specific behaviors, for example depression, social anxiety, or oppositional disorders.

Supplementary Materials 1-4 now contain comprehensive details about all models.

8) With regards to reference 8, the authors state that: “Previous research using this cohort... Due to the univariate methods they applied, they failed to find significant effects”. If I understand correctly, the current study also relies largely on univariate analyses of individual ROIs. Also what is meant by “failing to find significant effects”?

These statements were removed for clarity and the discussion updated as follows:

Previous research using this cohort to investigate the neuronal impacts of gestational age considered cortical thickness and surface area of grey matter [8] at baseline, 10 years of age. Our study replicated their cross-sectional results, even though we applied a CCA to initially select ROIs based on correlated patterns rather

than a whole brain investigation. Both studies found that adolescents born very premature had lower GM volume than moderately premature and full term subjects, in particular in the left anterior cingulate, the right cerebellar cortex, and the left thalamus.

9) Minor: In addition to ref 8 it might be relevant to discuss findings in relation to our paper which also used the ABCD sample to assess prematurity and multivariate imaging, combined by linked-ICA: “Linking pregnancy- and birth-related risk factors to a multivariate fusion of child cortical structure” (https://doi.org/10.1073/pnas.2422281122).

Thank you for pointing out this recent study. The introduction has been updated to include the following:

Another study using the same cohort found that maternal pregnancy complications, low birth weight, as well as premature birth were associated with global altered surface area and cortical thickness [51].

10) Minor: As statistical significance is documented in the results section and implied for reported findings, it may improve readability to reduce the extensive usage of the term “significant(ly)”. e.g., “Adolescents born very premature have significant differences within correlated patterns of GM and FA.” and “significantly less well”.

Results updated to reflect suggestions

11) Minor: It is stated that “Supplementary Table 1 contains the full details of the CVs”. Consider adding a table legend and instead state that is shows certain effect sizes or the like.

Supplementary Materials 1-4 now contain comprehensive details about all models.

12) Minor: Consider removing summation of actual findings currently presented in introduction: “Our findings reveal persistent structural brain differences in regions of the brain that are critical to attentional processes, memory, cognitive control, problem solving, and information processing. These differences not only strongly mediated the relationship with cognitive deficits related to GA, but also persisted across four years of development with little change in trajectory, suggesting that people born very premature may experiences differences in neuronal structure throughout their lifetime.”

The introduction has been amended to reflect the reviewers suggestions:

This comprehensive investigation will further our understanding of the implications of GA-dependent GM and FA differences, their impact on cognition, and their persistence across adolescence.

13) Minor: Why was scan-site also added as a random effect to models that did not include imaging?

Site still is a confounding variable for cognitive and behavioral testing, so it was included in non-imaging models.

14) Minor: What is meant by “disrupted” cortical folding?

The authors of the following study [3] report that the patterns of cortical folding in infants born very premature not only differs from full term infants, but is highly correlated to connectivity deficits in white matter. The introduction has been amended for clarity:

These deficits include, but are not limited to, disrupted cortical folding patterns of grey matter and connectivity of white matter within the first year of development [3],…

[3] Andrew Melbourne, Giles S. Kendall, M. Jorge Cardoso, Roxanna Gunny, Nicola J. Robertson, Neil Marlow, Sebastien Ourselin, Preterm birth affects the developmental synergy between cortical folding and cortical connectivity observed on multimodal MRI, NeuroImage, Volume 89, 2014, Pages 23-34, ISSN 1053-8119

15) Minor: I believe Figure 1 and 2 should be moved to the Results section.

Updated figures have been moved to the results section.

16) Minor: Were limiting the amounts of scanners considered when matching full term and pre-term youths?

That is an interesting question. We feel that limiting the number of scanners would not have significantly limited the number of sites, which carries a similar amount of variance, but would have reduced the number of subjects, simultaneously reducing the statistical power.

17) Minor: What is meant by: “The ABCD Study has... (dMRI) with well calibrated scanning parameters?”

Methods were updated as above.

18) Minor: For readability, consider reporting p-values in decimal rather than exponential (“p<8.05e-6”) notation.

Results were updated as above.

Reviewer 3

The manuscript Brain Regions with Gestational Age Differences Mediate Cognition in Adolescents Born Very Premature by Jensen and colleagues describes a study in which the authors leverage data from the ABCD study to investigate whether being born prematurely has lasting effects on brain structure and cognitive function. It has several strengths including the sample size and a very comprehensive analysis. There are a number of issues that came up while reading the manuscript, including a lack of detail or clarity in the methods, missing results, and conclusions that cannot be made based on the results found. I also feel that the manuscript needs heavy revising in terms of writing. It's difficult to assess whether a revision would lead to a publishable manuscript, because that depends on clarification of the methods. I've outlined my specific comments below.

We appreciate the feedback on our manuscript, hopefully the revisions undertaken based the reviewers comments has increased clarity into the methods as well as results and discussion. Additionally, the Supplementary Materials have been expanded to include complete model results.

1) In the methods, the authors state: Two summary cognitive measures, fluid and crystalized cognition, were derived from the battery of NIH Toolbox tests given to subjects. These outcomes spanned cognitive domains such as working memory, executive function, attention, episodic memory, language, and processing speed [16]. The fluid cognition summary score represents the ability to learn new information, while the crystalized cognition summary score represents already acquired knowledge [16]." How were the summary cognitive measures computed? Did all tests contribute to both, or did some contribute to fluid and others contribute to crystalized cognition?

Thank you for your comment. The methods were updated to include clarification of the fluid and crystalized cognitive summary scores:

The fluid cognition summary score is calculated from the results of the Dimensional Change Card Sort, Flanker Inhibitory Control and Attention, Picture Sequence Memory (Form A), List Sorting Working Memory, and Pattern Comparison tests. This fluid cognition score is considered to represent capacity information processing in novel situations [16, 48]. The crystalized cognition summary score is calculated from the results of the Picture Vocabulary and Oral Reading Recognition tests, representing acquired knowledge [16, 48].

2) The definition for fluid cognition (which I'm assuming is similar to fluid intelligence) doesn't seem correct. Fluid cognition generally refers to the ability to reason, manipulate information, and solve problems including in the abstract without prior knowledge. It would be assesed with tests of attention, working memory, etc. The authors definition that it is the ability to learn new information seems at odds with the accepted definition. Could the authors explain why they feel that fluid cognition captures learning, instead of how it is usually operationalized?

We apologize for the confusion. Per this paper *NIH Toolbox Cognitive Function Battery (CFB): Composite Scores of Crystallized, Fluid, and Overall Cognition* [48], the fluid cognition summary score reflects “capacity for new learning and information processing in novel situations” and crystalized cognitive summary reflects “past learning experiences”.

[48] Akshoomoff N, Beaumont JL, Bauer PJ, Dikmen SS, Gershon RC, Mungas D, Slotkin J, Tulsky D, Weintraub S, Zelazo PD, Heaton RK. VIII. NIH Toolbox Cognition Battery (CB): composite scores of crystallized, fluid, and overall cognition. Monogr Soc Res Child Dev. 2013 Aug;78(4):119-32. doi: 10.1111/mono.12038. PMID: 23952206; PMCID: PMC4103789.

3) In the methods, the authors define age as the following: "GA status was the independent variable, sex and age (in months to account for birth differences) were the fixed effects, and site was the random effect." I assume the authors are referring to the age of the participant at the time of the brain scan, rather than at birth. If so, how would reporting age in months account for birth differences differently to reporting them as years (i.e., what is the difference between 10.25 years and 123 months)? If this is the case, wouldn't that not account for birth differences, but instead it would account for age at scan? If the authors are referring to postmenstrual age rather than age at scan, then this is moot, but right now I'm completely unclear on how age is being included in the models.

Thank you pointing out the need for more accurate verbiage. Current research suggests that including a corrected age should be included when investigating impacts of premature birth, as differences between subjects born premature and their full term peers are attenuated when you also consider their corrected ages [49,50]. The age at scan provided by the ABCD study is given in months, so weeks premature were converted to months to perform the correction. The methods sections was updated for clarity as follows:

GA status was the independent variable, sex and age (corrected to account for birth differences [49, 50])

[49] Elmrayed S, Dai S, Lodha A, Kumar M, Fenton TR. Preterm growth assessment: the latest findings on age correction. J Perinatol. 2025 May;45(5):607-615. doi: 10.1038/s41372-024-02202-z. Epub 2025 Jan 16. PMID: 39820765; PMCID: PMC12221983.

[50] Doyle, Lex W.Anderson, Peter J. et al., Do We Need to Correct Age for Prematurity When Assessing Children? The Journal of Pediatrics, Volume 173, 11 - 12

4) Was CCA performed on all time points' imaging data, or just the 9-10 year old time point data? If all data, how was shared variance across time points accounted for? And under the following heading, "Cross-Sectional Analyses of Baseline Measures", is it correct to assume that this is only 9-10 year old data? It's stated in the heading that it's a cross-sectional analysis but the age group is not explicitly described in the text, and there are three sets of cross-sectional data and it's unclear which data the CCA was performed on in the first place.

All results were updated to reflect age of subjects at time of tests. The CCA was only performed on Baseline data.

5) I'd recommend formatting the equations for easier reading

We appreciate you recommendation. The methods were updated with formatted equations:

These datasets, $X \epsilon \mathbb{R}^{n x p}$ and $Y \epsilon \mathbb{R}^{n x p}$, where n is the number of subjects and p and q represent the number of variables in the datasets X and Y, the CCA will identify projection vectors u and v that solve [12] the following: $max_{u,v} u^{\intercal} \Sigma_{xy} v = min_{u,v} \| Xu - Yv \|_2^2$ subject to $\| Xu \|_2^2 = \| Yv \|_2^2 = 1$, where ΣXY is the covariance matrix of XY.

6) The results do not describe the null findings for mediating effects of FA on cognitive outcomes.

Supplementary Materials 1-4 now contain comprehensive details of all models.

7) Figure 3B should clarify that it's referring to grey matter volume.

We apologize for the omission. The figures have been updated to reflects suggestions and updated results and can be found at the end of reviewer comments.

8) In the discussion, the authors state that "Deficits in these regions, both for GM and FA would disrupt and degrade problem solving skills, as well as make retention and recall of prior knowledge more difficult." However, they did not find a relationship between FA and cognitive function, despite finding group differences in FA. Can the authors include discussion of why (as they state) deficits in FA would disrupt and degrade cognitive skills, yet they didn't find that here?

Thank you for the feedback. The discussion was updated to be more specific. References to FA with regards to cognition were removed.

9) The authors state "These GA-dependent differences were reduced GM and FA in regions of the brain that are critical to attentional processes, memory, cognitive control, problem solving, and information processing These differences not only strongly mediated the relationship with cognition but also persisted across four years of development with little change in trajectory." I don't feel like this is a valid interpretation of the results. There were reductions in FA, however they did not mediate the relationship with cognition. Additionally, calling a partial mediation of 15% "strong" feels overstated and would generally be classified as a weak or low level of variance accounted for.

The discussion was updated to be more specific. References to FA with regards to cognition were removed.

10) As an aside, the manuscript would greatly benefit from editing for spelling and grammar. As a few examples, I've shared some sentences below that made the manuscript difficult to read:
Grey matter volume (GM) and fractional anisotropy (FA) derivatives provided by the ABCD study.
Based on the current research, we believe that investigating correlated patterns of grey and white matter in adolescents born premature will revealed brain regions with GA-dependent differences.
GM measures were obtained from T1-weighted sMRI images underwent gradient distortion correction
We implemented a discovery and method, where the data was split randomly 50/50, to correct for overinflation [12]
Subject's GM and FA measures from three time points (baseline, 2nd year, and 4th year follow ups) were the dependent variables
in CV5, where very premature adolescents significantly different in both GM ($t=3.93$, $p<8.93e-5$) and FA ($t=3.17$, $p<0.0015$) compared to full-term
While able to leverage a large diverse cohort and perform both longitudinal and cross-sectional analyses, there are some limitations to this research

We appreciate your feedback on these errors in grammar and language. The manuscript was updated per reviewer's suggestions.

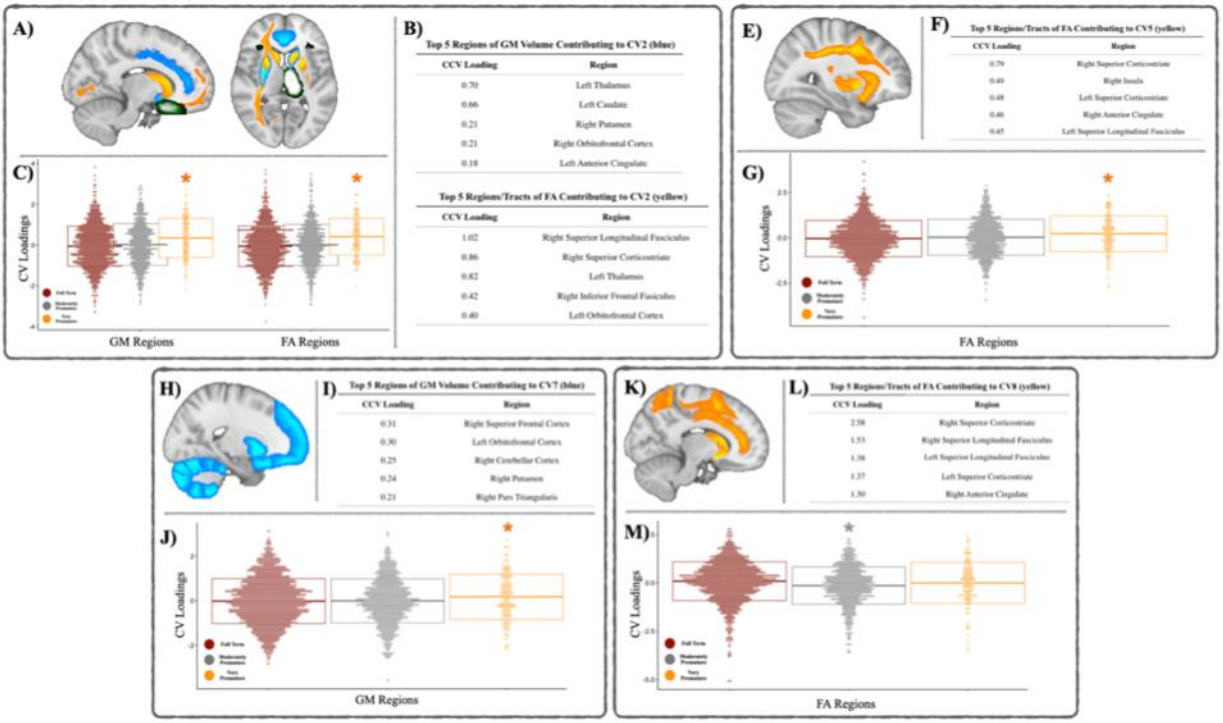

Figure 1: Canonical Variates with GA-dependent differences, regions with GM volume differences highlighted in blue, regions with FA differences highlighted in yellow, seen in **A**, **E**, **H**, and **K**. Each CVs top 5 regions with loadings as shown in **B**, **F**, **I**, and **L**. Dotplots showcase GA-dependent differences in CVs, where the asterisk highlights significance. Full term in red, moderately premature are in grey, and very premature in orange. In CV2, (plot **C**, GM (t=3.50, p<0.0005) and FA (t=5.19, p<2.25^-7)), CV5 (plot **G**, FA (t=3.17, p<0.002)), and CV7 (plot **J**, GM (t=3.38, p<0.002)), very premature adolescents contributed more to the correlated patterns of GM and/or FA than their peers. In CV8 (plot **M**, FA (t=-4.48, p<0.0007)), moderately premature adolescents contributed less than their peers to the correlated patterns of that CV.

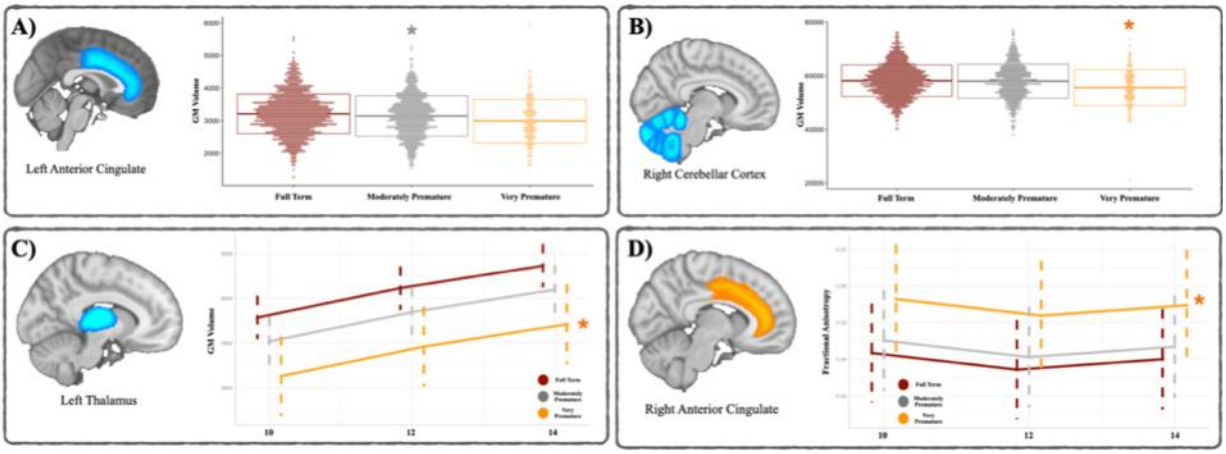

Figure 2: Regional GA-dependent differences - Baseline results at ten years of age: **A)** Left anterior cingulate, where moderately premature adolescents had lower GM volume than their peers (t=-3.47, p<0.0005). **B)** Right cerebellar cortex, where very premature adolescents had lower GM volume (t=-3.17, p<0.003). Longitudinal results, 10-14 years of age: **C)** Left thalamus, where very premature subjects had lower GM volume than their peers (t=-2.99, p<0.003). **D)** Right anterior cingulate, where very premature subjects had higher FA (t=3.89, p<0.0001) across time. GM is highlighted in blue, FA in yellow. In both dotplots and line plots, full term subjects are red, moderately premature subjects in grey, and very premature subjects in orange. Significance is shown by asterisk.

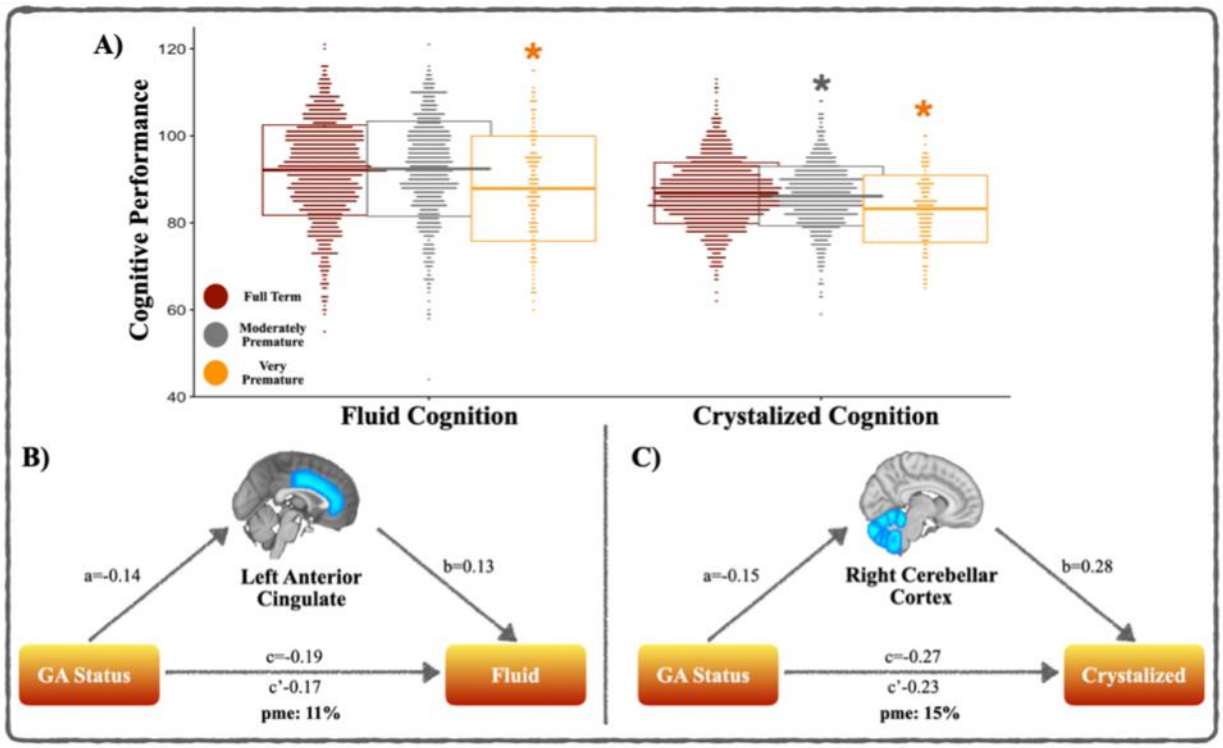

Figure 3: GA-dependent differences in cognition, mediated by regional GM volume differences - **A)** moderately premature adolescent performed less well in crystalized cognition than their full term peers (t=-4.10, p<3.2^-5), while very premature adolescents performed less well in fluid (t=-6.00, p<2.3^-9) as well as crystalized (t=-6.9, p<7.2^-12) cognition than their peers. **B)** The left anterior cingulate mediated the relationship between GA and fluid cognition by 11%. **C)** The right cerebellar cortex mediated the relationship between GA and crystalized cognition by 15%. GM is highlighted in blue. In the dotplot, full term subjects are red, moderately premature subjects in grey, and very premature subjects in orange. Significance is shown by asterisk.