A preprocessed open diffusion derivatives dataset from the Healthy Brain Network

Adam Richie-Halford[*,1], Matthew Cieslak[*,2], Alexandre R. Franco[3], Valerie J. Sydnor[2], Jason Yeatman[4], Lei Ai[3], Michael Milham[3], the Fibr raters, Theodore D. Satterthwaite[**,2], Ariel Rokem[**,1]

1. University of Washington, 2. University of Pennsylvania, 3. Child Mind Institute, 4. Stanford University

OHBM 2021 Poster 1349

Contact: richford@uw.edu

(* These authors contributed equally, ** These authors contributed equally)

Introduction

- The Healthy Brain Network (HBN; `http://fcon_1000.projects.nitrc.org/indi/cmi_healthy_brain_network/`) is collecting data from 10,000 children and adolescents together with extensive behavioral and health assessments [1].
- This includes diffusion MRI (dMRI) data, allowing for analysis of the physical properties of white matter.
- However, dMRI analysis is complicated by the need for complicated preprocessing steps. Misapplication of these steps can induce bias in subsequent interpretation [2].
- Once preprocessing is done correctly, there is little need for downstream researchers to repeat these steps.

Challenge: Provide an openly available preprocessed diffusion derivatives dataset that applies best practices in a robust and transparent way.

- We introduce the HBN Preprocessed Open Diffusion Derivatives (HBN-POD2), a large dataset for the analysis of structural brain connectivity and pediatric mental health.
- Preprocessing was performed in QSIPrep [3] version 0.12.2. QSIPrep is A BIDS App that automatically builds preprocessing pipelines based on BIDS inputs.

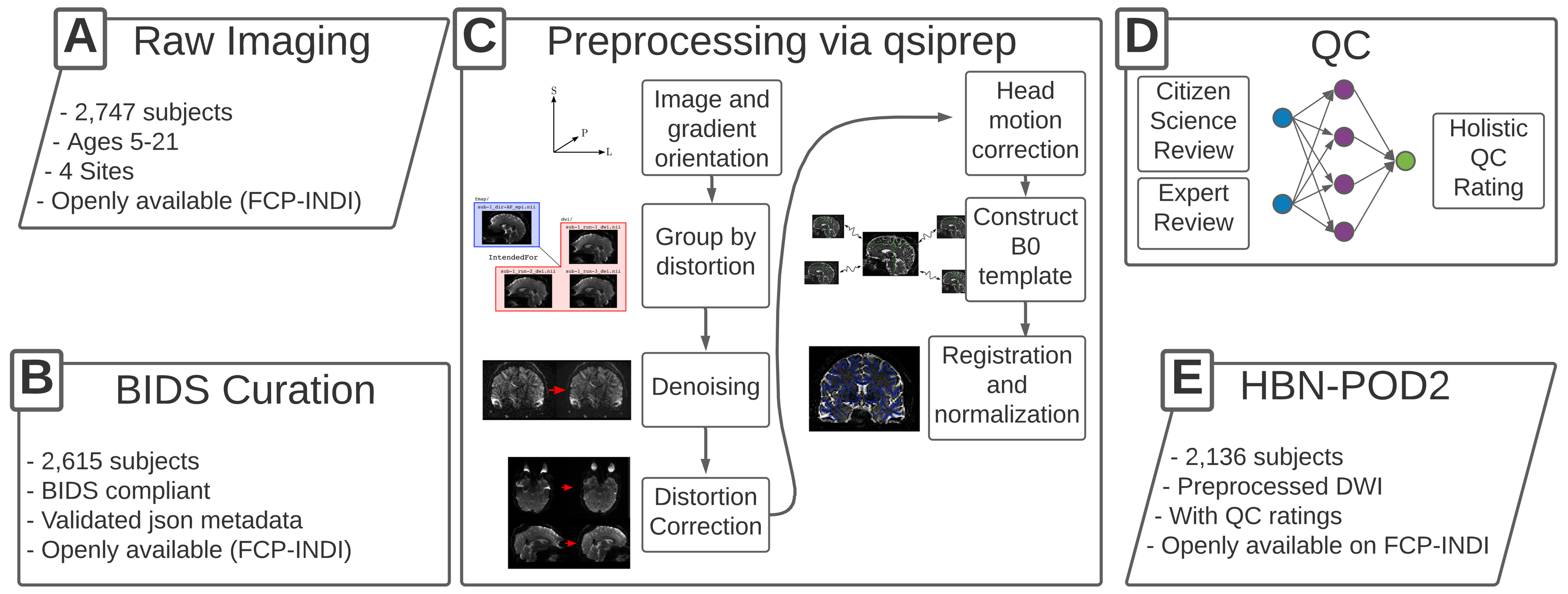

Figure 1: HBN-POD2 data provenance: **(A)** Imaging data for 2,747 subjects, aged 5-21 years and collected at four sites in the New York City area, was made available through the Functional Connectomes Project and the International Neuroimaging Data-Sharing Initiative (FCP-INDI). **(B)** These data were curated for compliance to the Brain Imaging Data Structure (BIDS) specification and availability of imaging metadata in json format. 2,615 subjects met this specification. **(C)** Imaging data was preprocessed using qsiprep to group, distortion correct, motion correct, denoise, coregister and resample MRI scans. 2,136 subjects passed this step, with the majority of failures coming from subject with missing dMRI scans. **(D)** A image classification algorithm was trained on a combination of both expert and citizen scientist reviews (see Figure 2) to assign either passing or failing quality control (QC) scores to each individual. **(E)** The HBN-POD2 dataset, including QC scoring, is openly available through FCP-INDI.

Conclusion and future work

- We present a large preprocessed dMRI dataset for the study of pediatric mental health and structural brain connectivity.
  - 2,136 subjects, aged 5-21
  - preprocessed DWI
  - including QC ratings created from a combination of expert ratings and citizen scientists
  - openly available on FCP-INDI
- HBN-POD2 is amenable to different dMRI analysis, including tractometry, graph theoretical analysis, and combinations with fMRI.
- Future work:
  - HBN-POD2 will grow as HBN reaches its 10,000 participant goal.
  - Develop a deep learning model for QC of future releases without the need for citizen science ratings.

Quality Control

- Six dMRI experts rated 200 participants' processed data on a 1-5 scale using the dmriprep-viewer interface `https://www.nipreps.org/dmriprep-viewer`)
- We found excellent inter-rater reliability: ICC3k=0.93 (95% CI: [0.91, 0.94]) Ratings are available in `https://github.com/richford/hbn-pod2-expert-qc`.
- We call this subset the "gold standard" QC dataset.
- But the size of HBN-POD2 makes expert QC on the full dataset prohibitive, we estimate QC of the full dataset would take each FTE expert rater 200 hours.

Challenge: Generate QC labels for a large dMRI dataset when expert ratings are prohibitively time consuming.

- Based on the SwipesForScience framework (https://swipesforscience.org/; [4]), we created a citizen science web-app that presents static/dynamic views of b=0 maps and directionally-encoded FA maps for a binary (pass/fail) decision for 1,653 HBN subjects.
- 374 citizen scientists provided $587,778$ ratings (of these, 133 citizen scientists provided $> 3,000$ votes each) for a mean of > 50 ratings per slice (or > 200 ratings per subject).
- We then trained a gradient boosting model (XGB) to predict expert scores based on a combination of citizen science ratings and automated QSIPrep QC metrics, attaining $R^2 = 0.80$ on a held-out test set from the "gold standard."
- A separate model trained on only the QSIPrep QC metrics attained $R^2 = 0.68$, demonstrating the added value of the citizen science raters.

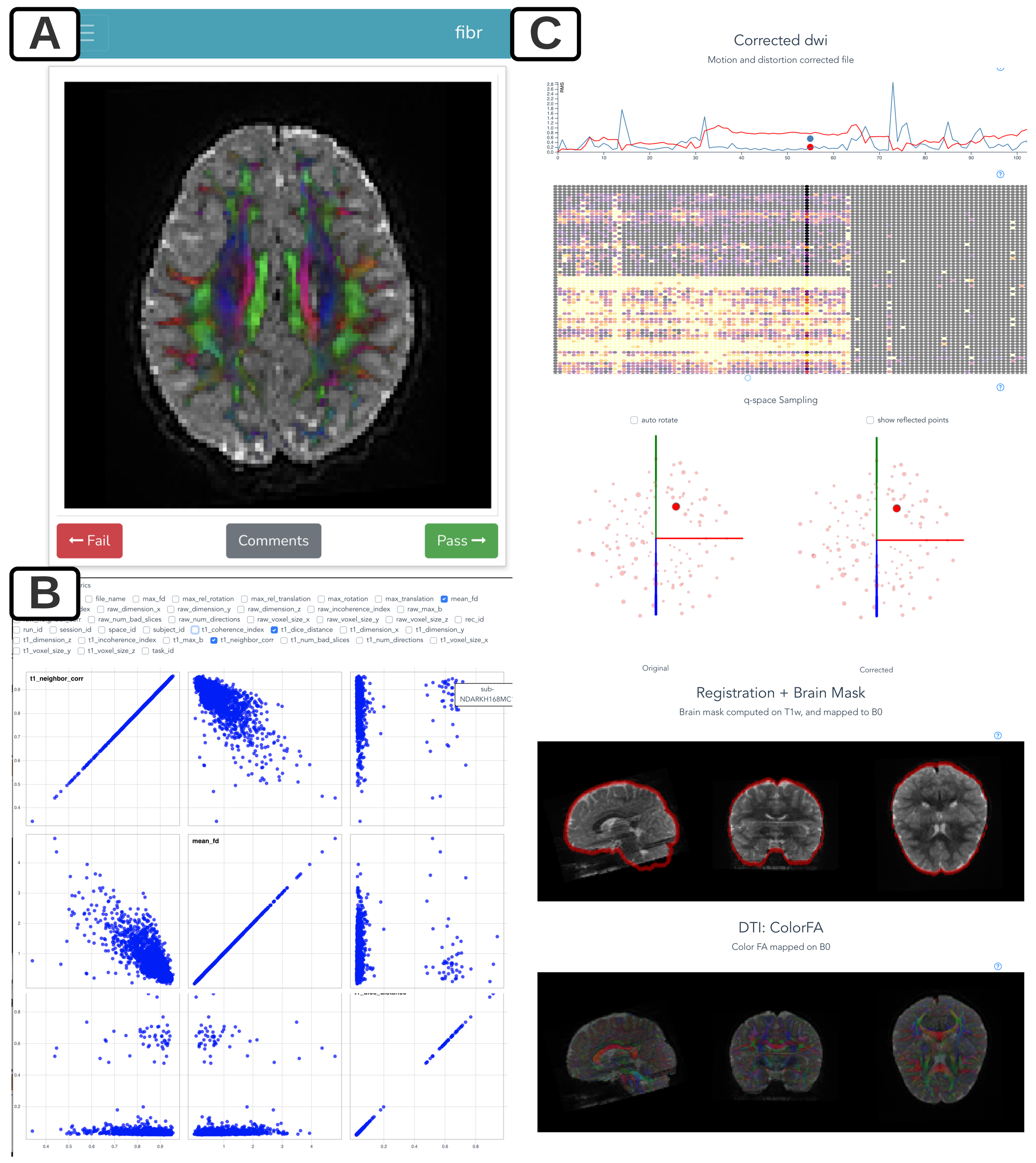

Figure 2: HBN-POD2 quality control (QC) instruments: **(A)** The user interface for citizen science QC. After a tutorial, users are asked to give binary pass/fail ratings to each subject's color scale fractional anisotropy image. The intuitive swipe or click interface allows citizen scientists to review more images than is practical for expert review. Expert reviewers use the more advanced qsiprep-viewer interface, where they can **(B)** view the distribution of QC metrics for the entire study using interactive scatterplots and violin plots, and **(C)** inspect Individual subjects' preprocessing results, including corrected dMRI images, frame displacement, q-space sampling distributions, registration information, and a DTI model.

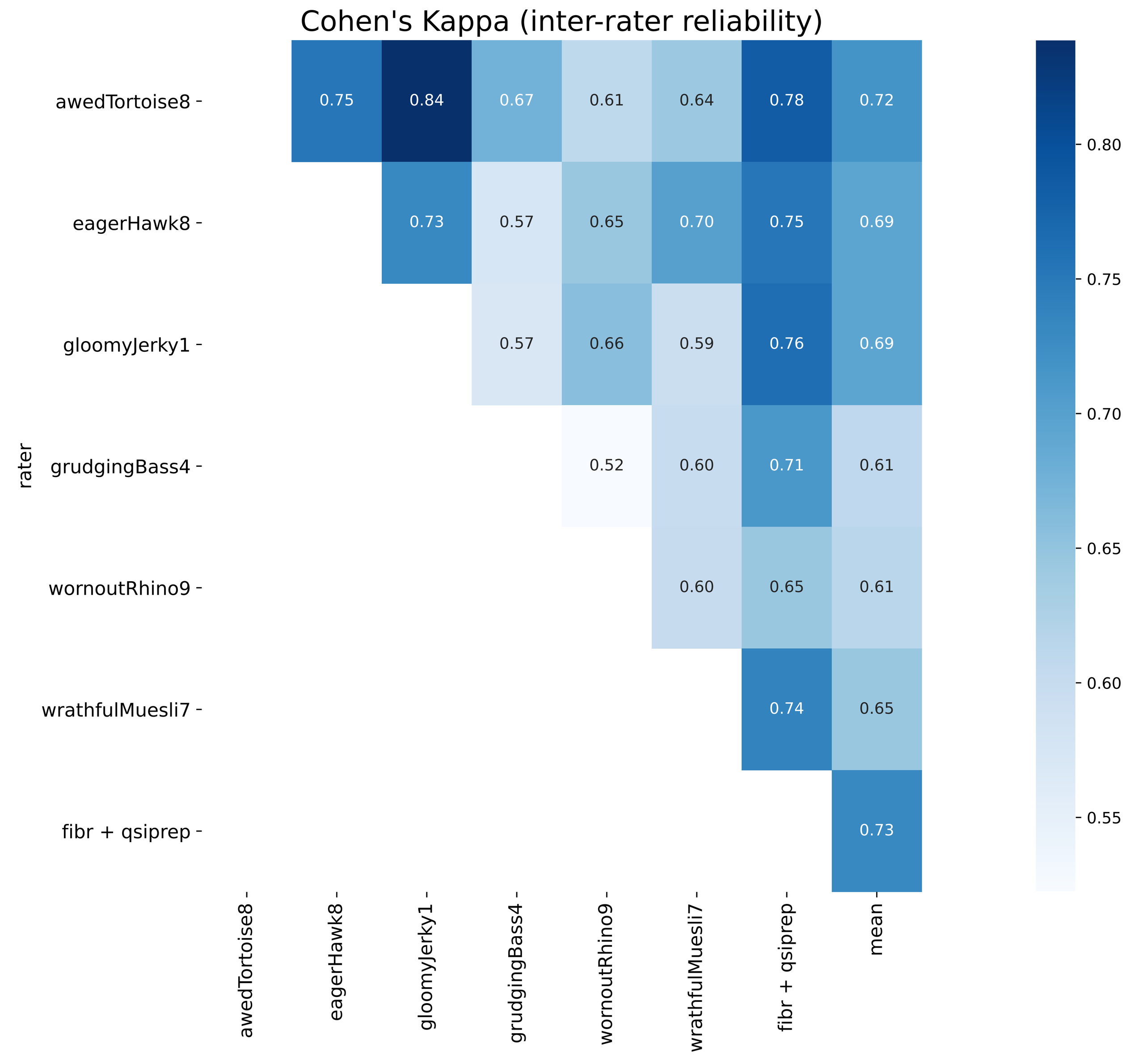

Figure 3: HBN-POD2 quality control (QC) results: Six dMRI experts (with anonymized names) rated a subset of 200 subjects, with acceptable inter-rater reliability: **ICC3k** $= 0.930 (95\% CI : [0.91, 0.94])$, **ICC3** $= 0.688 (95\% CI : [0.64, 0.74])$, $\langle \kappa \rangle = 0.648$. We trained a gradient boosting model (XGB) on the average expert rating, obtaining $R^2 = 0.80$ on a held-out test set. Treating XGB as a seventh expert rater, preserves inter-rater reliability: **ICC3k** $= 0.945 (95\% CI : [0.93, 0.96])$, **ICC3** $= 0.709 (95\% CI : [0.66, 0.75])$. This entitles us to extend it's ratings to the entire HBN-POD2 dataset.

* **ICC3** and **ICC3k** are appropriate when a fixed set of k raters rate each subject (i.e. a fully crossed design). When all subjects are coded by multiple raters and the average of their ratings is used for hypothesis testing, **ICC3k** is appropriate. When a subset of subjects is coded by multiple raters and the reliability of their ratings is meant to generalize to other subjects rated by only one coder, the single-measure **ICC3** must be used.

- Treating the new XGB model as an additional expert rater, we attain **ICC3k** $= 0.945 (95\% CI : [0.93, 0.96])$ and **ICC3** $= 0.709 (95\% CI : [0.66, 0.75])$.
- The reliability of the XGB model entitles us to apply the XGB ratings to the remaining HBN-POD2 scans, generating QC labels for the entire dataset.
- HBN-POD2 offers open QC metrics, eliminating the need for investigators to repeat quality assessment for over 2,000 scans.

References

[1] Alexander et al. An open resource for transdiagnostic research in pediatric mental health and learning disorders. *Scientific Data*, 4:170181, December 2017.

[2] Derek K. Jones and Mara Cercignani. Twenty-five pitfalls in the analysis of diffusion mri data. *NMR in Biomedicine*, 23(7):803–820, 2010.

[3] Matthew Cieslak, Philip A. Cook, Xiaosong He, Fang-Cheng Yeh, Thijs Dhollander, Azeez Adebimpe, Geoffrey K. Aguirre, Danielle S. Bassett, Richard F. Betzel, Josiane Bourque, Laura M. Cabral, Christos Davatzikos, John Detre, Eric Earl, Mark A. Elliott, Shreyas Fadnavis, Damien A. Fair, Will Foran, Panagiotis Fotiadis, Eleftherios Garyfallidis, Barry Giesbrecht, Ruben C. Gur, Raquel E. Gur, Max Kelz, Anisha Keshavan, Bart S. Larsen, Beatriz Luna, Allyson P. Mackey, Michael Milham, Desmond J. Oathes, Anders Perrone, Adam R. Pines, David R. Roalf, Adam Richie-Halford, Ariel Rokem, Valerie J. Sydnor, Tinashe M. Tapera, Ursula A. Tooley, Jean M. Vettel, Jason D. Yeatman, Scott T. Grafton, and Theodore D. Satterthwaite. Qsiprep: An integrative platform for preprocessing and reconstructing diffusion mri. *bioRxiv*, 2020.

[4] Anisha Keshavan, Jason D. Yeatman, and Ariel Rokem. Combining citizen science and deep learning to amplify expertise in neuroimaging. *Frontiers in Neuroinformatics*, 13:29, 2019.

Acknowledgments

NIH National Institute of Mental Health

eScience Institute ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

Alfred P. Sloan FOUNDATION

GORDON AND BETTY MOORE FOUNDATION