Fast Connectivity Gradient Approximation: Maintaining spatially fine-grained connectivity gradients while reducing computational costs

Karl-Heinz Nenning[1]*, Ting Xu[2], Arielle Tambini[1], Alexandre R. Franco[1], Daniel S. Margulies[3], Stanley J. Colcombe[1], Michael P. Milham[1,2]

1 Nathan S. Kline Institute for Psychiatric Research, Orangeburg, NY, USA
2 Child Mind Institute, New York, NY, USA
3 CNRS & Université de Paris, Paris, France

* Correspondence: karl-heinz.nenning@nki.rfmh.org

Abstract

Brain connectome analysis suffers from the high dimensionality of connectivity data, often forcing a reduced representation of the brain at a lower spatial resolution or parcellation. However, maintaining high spatial resolution can both allow fine-grained topographical analysis and preserve subtle individual differences otherwise lost. This work presents a computationally efficient approach to estimate spatially fine-grained connectivity gradients and demonstrates its application in improving brain-behavior predictions.

Main

Graph-based models have become mainstream tools for the visualization and characterization of relationship structures within complex systems in machine learning, computer vision, and bioinformatics. Unfortunately, their application is not without challenges - particularly when dealing with representations at higher resolutions, as their computational and storage demands can exceed the capacity of most computing resources. Data reduction is indispensable when the feasibility of graph-based models is impacted. Such reductions can be achieved through statistical techniques (e.g., principal component, factor analyses), data downsampling approaches (e.g., local averaging based on spatial or topographic structure), or both. Here, using graph-based models in neuroimaging as an example, we demonstrate the need for careful consideration of the order in which these two reduction strategies are applied, suggesting an optimized workflow when both are pursued.

In neuroimaging, graph-based representations are increasingly used to characterize topographic patterns of gradual transitions and boundaries between systems - i.e., connectivity gradients.[1–3] But maintaining a full spatial representation of the connectome is challenging. The number of connections scales exponentially with the number of nodes, and a common spatial resolution of the brain comprising 60,000 cortical nodes (vertices) constitutes 3,600,000,000 connections (edges). This makes it virtually impossible to evaluate the full-scale connectivity structure on consumer grade hardware. A common practice to overcome this challenge is to reduce the spatial resolution of the data, typically by averaging across vertices within spatial regions of interest (ROIs) defined using a parcellation template.[4] However, the loss of individual-specific detail in such parcellation approaches may result in the loss of meaningful information; there is also little agreement on the appropriate choice of parcellation template in

connectome studies.[5,6] Moreover, the capability to maintain a fine-grained spatial resolution allows researchers to fully characterize the spatial layout of functional regions, an important and often overlooked aspect of modeling brain connectivity.[7]

Here, to facilitate connectivity gradients at a fine-grained spatial resolution, we propose *Fast Connectivity Gradient Approximation* (FCGA). Instead of using a full-scale vertex-by-vertex connectivity matrix, at its core, FCGA leverages a set of landmarks to approximate the underlying connectivity structure at full spatial resolution, with an efficiency that makes it possible to run on common computer hardware. These landmarks are flexible, and can, for example, be based on individual vertices or predefined ROIs.

A schematic workflow of our proposed approach is displayed in Figure 1a. After (i) establishing a set of k landmarks, the (ii) functional connectivity between all n vertices (or voxels) and the k landmarks is calculated, resulting in a n-by-k connectivity matrix ($CM_{n\text{-}by\text{-}k}$) between all pairs of vertices and landmarks. Following current practices when calculating connectivity gradients, we quantify the spatial similarity of vertex-wise connectivity profiles. To do so, (iii) a k-by-k (landmark-by-landmark) connectivity matrix ($CM_{k\text{-}by\text{-}k}$) that characterizes the functional connectivity between all pairs of landmarks is established. Then, row-wise thresholding, retaining the 10% strongest positive connections, is applied to both $CM_{n\text{-}by\text{-}k}$ and $CM_{k\text{-}by\text{-}k}$. Subsequently, (iv) the row-wise cosine similarity between all pairs of vertices in $CM_{n\text{-}by\text{-}k}$ and $CM_{k\text{-}by\text{-}k}$ is calculated. This results in an n-by-k affinity matrix ($W_{n\text{-}by\text{-}k}$) that characterizes the spatial similarities between the voxel-wise connectivity profiles with the landmark regions. Finally, (v) we use Principal Component Analysis for dimensionality reduction of $W_{n\text{-}by\text{-}k}$, establishing the approximated connectivity gradients.

We evaluated the validity of FCGA on the group and individual level, using high resolution fMRI data (59,412 cortical vertices) from the Human Connectome Project (HCP)[8] and the Nathan Kline Institute-Rockland Sample (NKI-RS).[9] We parametrically varied the amount of downsampling (number of landmarks) across sampling choices (vertex- and ROI-level).

At the group level, we used Spearman rank correlation to compare the spatial similarity between the approximated connectivity gradients (G_{FCGA}) and the connectivity gradients based on the full 59,412-by-59,412 connectivity matrix (G_{full_fc}). The spatial similarity between G_{FCGA} and G_{full_fc} increased with the number of landmarks used to calculate the approximated gradients (Figure 1b). Remarkably, using 1000 ROIs as landmarks (~1.7% of the full connectivity matrix), the average spatial similarity across 25 connectivity gradients reached ρ=0.86. Using 3000 (~5%) uniformly distributed vertices, a spatial similarity of ρ=0.98 was achieved with <10% of the computational time and memory usage as compared to the calculation of G_{full_fc} (Figure 1c). The spatial topography of the top connectivity gradients, shown in Figure 1d, further illustrates the high spatial similarity between G_{FCGA} and G_{full_fc} across a number of sampling choices.

At the individual level, reliability and discriminability analysis confirmed the repeatability and the preservation of individual features with G_{FCGA}. To quantify agreement between G_{FCGA} and G_{full_fc}, we calculated the intraclass correlation coefficient (ICC) and discriminability for the same

acquisition. We observed an average vertex-wise ICC of >0.9 from 1000 (~1.7%) landmarks onwards, and discriminability was already nearly perfect with 300 (<1%) landmarks (Figure 2a). Next, we compared the ICC and the discriminability of G_{FCGA} and G_{full_fc} across sessions (Figure 2b). We observed that landmarks based on single vertices (uniformly or randomly distributed) yielded slightly lower ICC and discriminability for G_{FCGA} than for G_{full_fc}. However, landmarks based on predefined parcels on the group or individual level yielded a slightly better discriminability and a similar ICC for G_{FCGA} when compared to G_{full_fc} (Figure 2b).

Finally, we evaluated the practical implications of our FCGA approach on brain-wide association studies by predicting age and full-scale intelligence (FSIQ) across the lifespan. Specifically, we compared the predictive performance of coarse-grained gradients calculated from parcellated time-series to FCGA constructed parcel-wise summarized fine-grained gradients (Figure 2c). Importantly, we observed that fine-grained gradients (i.e. G_{FCGA}) subjected to parcel-averaging outperformed coarse-grained gradients for predicting both age and FSIQ across all tested parcellations ($p_{Bonferroni}<0.05$ in 32/32 comparisons) (Figure 2c). Notably, for FSIQ we observed that gradients calculated on a low number of parcels did not perform better than chance. This indicates that constructing connectivity gradients from parcellated data might lose meaningful information in the process that is otherwise preserved in the fine-grained gradient topography.

Taken together, our results demonstrate the feasibility of establishing spatially fine-grained connectivity gradients while mitigating the computational burden of vertex-level functional connectivity data by reducing computational time and memory usage. The ability to fully appreciate the spatial layout of functional regions in gradient representations at full resolution could improve associations between functional imaging data and cognitive traits,[7] and inform structure-function relationships.[10] Importantly, the advantage of preserving informative individual features in gradients of spatially fine-grained connectivity data was emphasized by the improved prediction of age and intelligence over gradients based on coarse-grained connectivity data. Overall, FCGA strengthens the core concept of connectivity gradients by maintaining the full spatial resolution to study spatial transitions of brain organization. Finally, its efficiency also removes computational barriers and can enable the widespread application of connectivity gradients to capture signatures of the connectome.

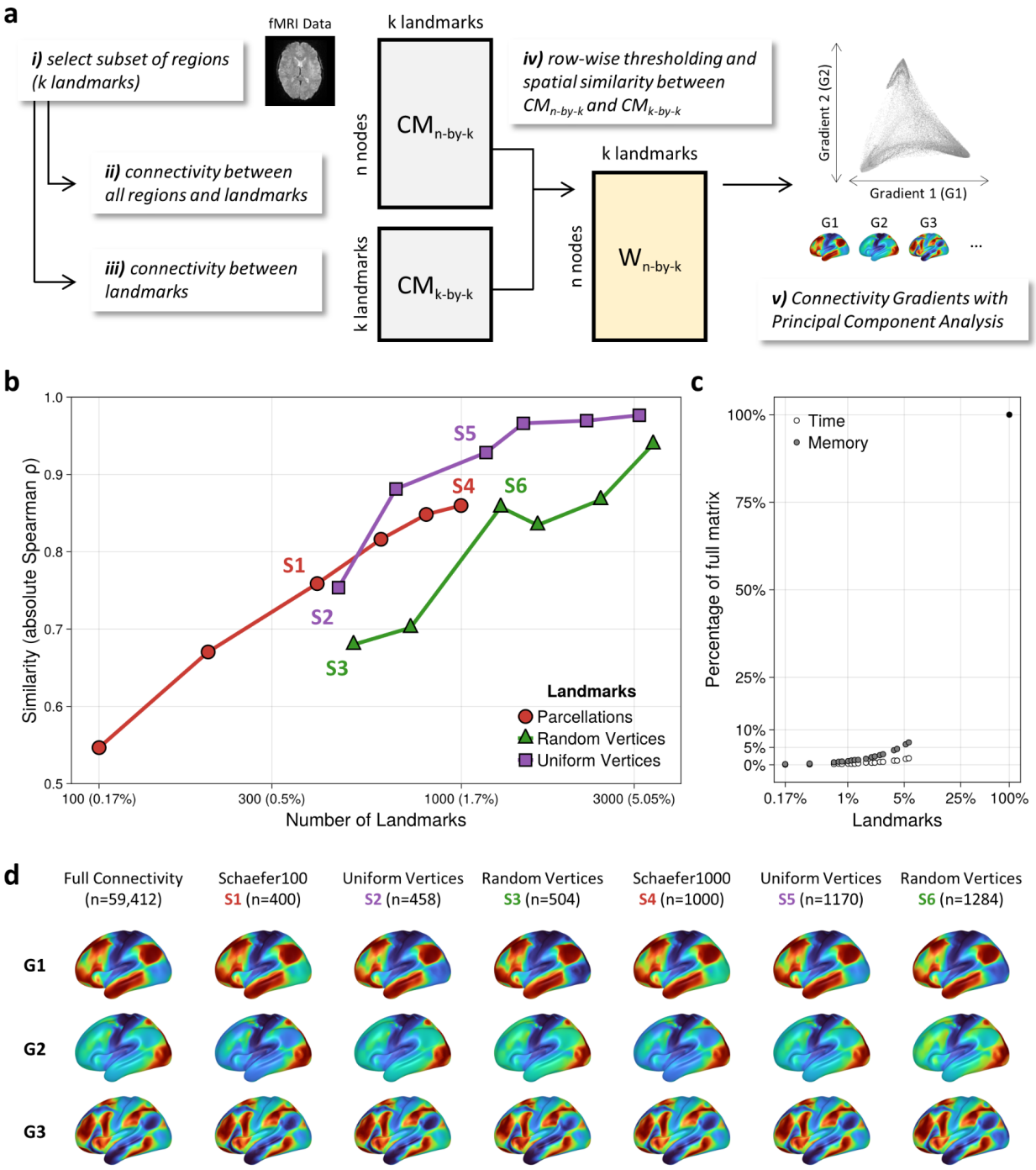

Figure 1. a) A schematic workflow of the proposed Fast Connectivity Gradient Approximation (FCGA) approach. b) FCGA provides high fidelity to the full connectivity structure with only a fraction of landmarks and c) <10% computational time and memory requirements. d) The core topography is similar across a different number of landmarks and sampling choices.

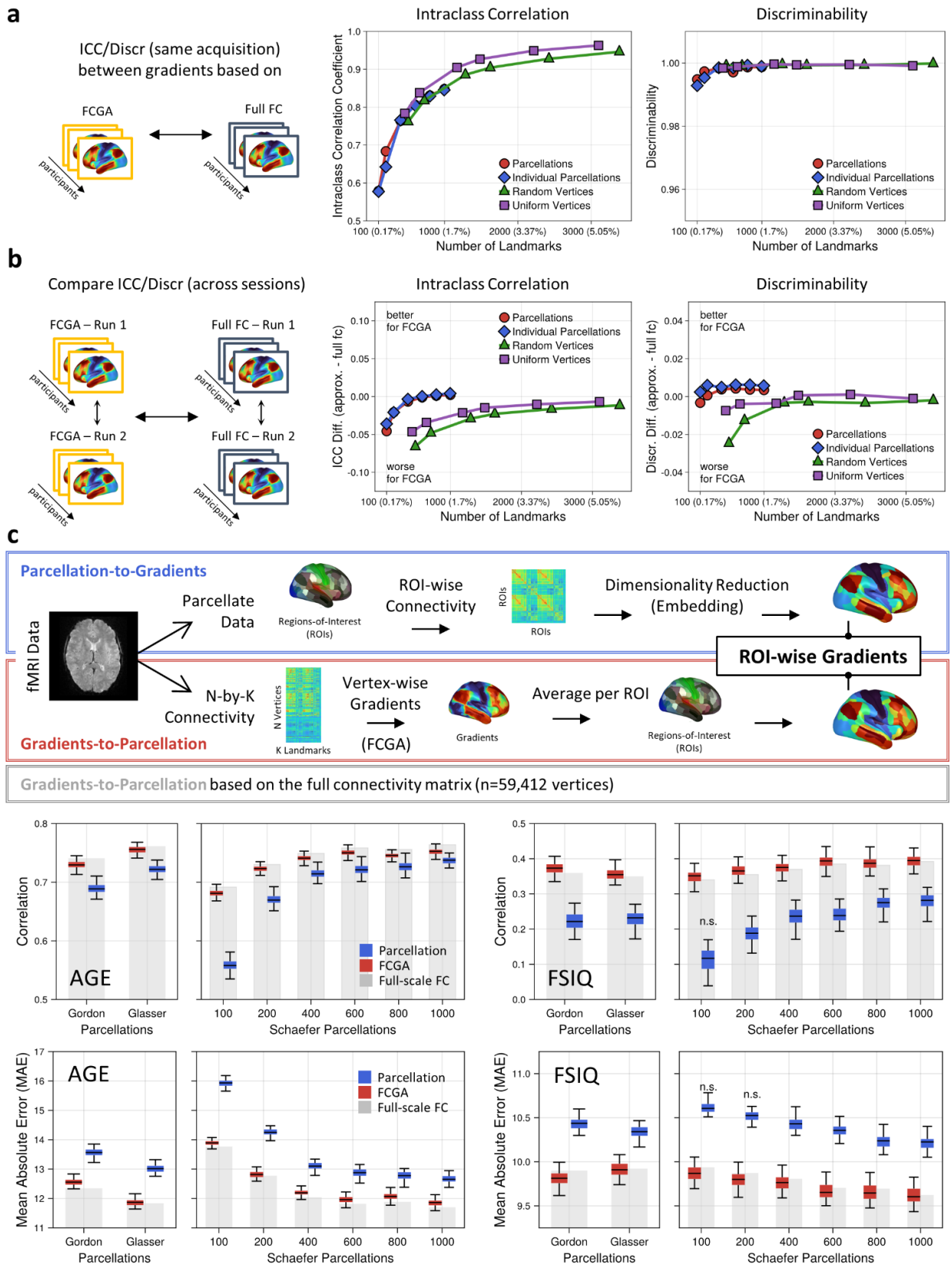

Figure 2. Intraclass correlation and discriminability within the same session confirms the

repeatability and the preservation of individual features within (a) and across (b) sessions. c) Parcel-averaging of spatially fine-grained gradients outperforms gradient calculation on a parcel level for both age and FSIQ prediction. Boxplot show 500 predictions, colored boxes interquartile range (iqr) with whiskers spanning 1.5*iqr. Predictions that are not significantly better than prediction with randomly shuffled labels are denoted as n.s.

Methods

Data

To evaluate the validity of FCGA on the group and individual level, we used resting-state fMRI data from the Human Connectome Project (HCP).[8] The HCP data was acquired at Washington University at St. Louis on a customized Siemens 3T Connectome Skyra scanner. Resting-state fMRI was acquired with a multiband factor of 8, 2mm isotropic resolution, and a repetition time of 0.72 seconds for a duration of 14.4 minutes, resulting in 1200 volumes per acquisition. Participants were asked to relax, keep eyes open and fixated on a crosshair, and not to fall asleep. Four resting-state runs were collected in different sessions across two days (REST1 and REST2), where each session comprised two runs with different phase encoding directions (LR and RL). We used the minimally preprocessed fMRI data provided by the HCP.[11] Briefly, the resting-state data has been motion corrected, minimally spatially smoothed (2mm), high-pass filtered (2000s cutoff), denoised for motion-related confounds and artifacts using independent component analysis,[12] and spatially aligned to the 2 mm standard CIFTI grayordinates space.[11,13]

At the group level, we used the dense group-average functional connectome provided as part of the HCP S1200 data release. In brief, this connectivity matrix was constructed from the minimal preprocessed data across 1003 individuals that each had ~1 hour (4x14.4 minutes) of resting-state fMRI acquisitions. The dense connectivity matrix of size 91,282 x 91,282 grayordinates (59,412 cortical vertices and 31,870 subcortical voxels) was calculated based on components of an incremental group-pca. For the group-average connectome of the HCP, we averaged the precomputed functional connectivity measures for each landmark instead of recalculating landmark-based connectivity. At individual-level, we used the 100 unrelated subjects cohort from the HCP (54F/46M, age = 29 ± 3.7 years) and two repeated resting-state fMRI runs (REST1_LR and REST2_LR) for each individual. We used the minimally preprocessed data as provided by the HCP. For both the group and individual level, we focused our analysis on the 59,412 cortical vertices.

We used a cross-sectional lifespan sample with 313 healthy participants (214 female, age: 6-85 years, 42.2 ± 22.4 years) to evaluate the practical implications of our approach on brain-wide association studies. Participants were selected from the Nathan Kline Institute-Rockland Sample (NKI-RS) [9], who have no diagnosis of any mental or neurological disorders, and passed quality control of a head motion criteria (mean framewise displacement < 0.25mm). The NKI-RS data was acquired at the Nathan Kline Institute on a Siemens TrioTim 3 Tesla scanner. Resting-state fMRI was acquired with a multiband factor of 4, 3mm isotropic resolution, and a repetition time of 0.645 seconds for a duration of 9.7 minutes, which resulted in 900 volumes per run. Preprocessing was performed with the Connectome Computational System,[14] and included discarding the first five timepoints, compressing temporal spikes, slice timing correction, motion correction, 4D global mean intensity normalization, nuisance regression (Friston's 24 model, cerebrospinal fluid and white matter), linear and quadratic detrending, band-pass filtering (0.01-0.1Hz), as well as global signal regression. The preprocessed data were then projected on the 32k fsLR surface template with 32,492 vertices per hemisphere.

Connectivity landmarks

We used a varying number of landmarks across different sampling strategies. We utilized randomly and uniformly distributed vertices across the cortex, where the selected vertices were consistent across individuals. On the ROI-level, connectivity landmarks were defined by the average time-series based on group parcellations[15–17] and individualized parcellations for the HCP sample.[18]

Comparing gradients

For the group level analysis and comparisons, no alignment or reordering of the gradients (PCA components) was performed. For the analysis on the individual level, we used one set of reference gradients based on the HCP dense connectome for both the approximated and full-scale gradients. For each individual, sampling, and gradient construction approaches, the gradients were aligned to this reference with orthogonal Procrustes alignment.[19] Orthogonal Procrustes finds the optimal linear transformation so that two sets of gradients have a matched component order and coefficient signs.

Intraclass correlation and discriminability

We compared the reliability and discriminability of the approximated gradients (G_{FCGA}) and the full-scale gradients (G_{full_fc}) within and across sessions, using 2 repeated resting-state acquisitions from the 100 unrelated individuals sample of the HCP. Within-session analysis treated G_{FCGA} and G_{full_fc} of the same acquisition as repeated measures. Across-session analysis evaluated differences in the reliability and reproducibility between G_{FCGA} and G_{full_fc}. Discriminability[20] was used to measure the similarity of connectivity gradients. It is a nonparametric multivariate statistic that quantifies the degree to which repeated measurements (e.g., gradients of different approaches or sessions) are relatively similar to each other. At the vertex level, we quantified the intraclass correlation coefficient (ICC), a univariate measure of the degree of absolute agreement.[21,22]

Prediction of individual-specific measures

We used the NKI-RS lifespan sample to evaluate implications of our proposed approach for the prediction of individual-specific measures such as age and FSIQ. As described above, connectivity gradients are often calculated on a reduced data representation on a parcel-level. In this study, we compared the predictive performance of connectivity gradients calculated on parcellated time-series (parcellation-to-gradients) to a parcel-wise averaging of fine-scale gradients constructed with FCGA (gradients-to-parcellation). Connectivity gradients of the parcellated data are established as in prior work,[1] using all parcels as landmarks instead of a subset. We evaluated distinct, commonly used parcellations,[15–17] and focused on the first 5 gradients as features to avoid excessively increasing the feature space. For prediction, we used ridge regression with a L2 regularization as implemented with glmnet,[23] and a nested 10-fold cross-validation scheme for hyperparameter (lambda) selection. The predictions of each fold were aggregated, and performance was measured with mean absolute error (MAE) and correlation to the true age and FSIQ values. We repeated the 10-fold cross-validation runs 500 times, with random splits for each fold. Additionally, we tested the prediction results against a

baseline of 500 prediction runs with randomly shuffled labels to evaluate if the predictive performance is greater than chance.

References

1. Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci. U. S. A.* **113**, (2016).
2. Bernhardt, B. C., Smallwood, J., Keilholz, S. & Margulies, D. S. Gradients in brain organization. *Neuroimage* **251**, (2022).
3. Haak, K. V., Marquand, A. F. & Beckmann, C. F. Connectopic mapping with resting-state fMRI. *Neuroimage* **170**, (2018).
4. Arslan, S. *et al.* Human brain mapping: A systematic comparison of parcellation methods for the human cerebral cortex. *Neuroimage* **170**, (2018).
5. Domhof, J. W. M., Jung, K., Eickhoff, S. B. & Popovych, O. V. Parcellation-induced variation of empirical and simulated brain connectomes at group and subject levels. *Network Neuroscience* **5**, 798 (2021).
6. Bryce, N. V. *et al.* Brain parcellation selection: An overlooked decision point with meaningful effects on individual differences in resting-state functional connectivity. *Neuroimage* **243**, 118487 (2021).
7. Bijsterbosch, J. D. *et al.* The relationship between spatial configuration and functional connectivity of brain regions. (2018) doi:10.7554/eLife.32992.
8. Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: an overview. *Neuroimage* **80**, (2013).
9. Nooner, K. B. *et al.* The NKI-Rockland Sample: A Model for Accelerating the Pace of Discovery Science in Psychiatry. *Front. Neurosci.* **6**, 32787 (2012).
10. Haenelt, D. *et al.* High-resolution quantitative and functional MRI indicate lower myelination of thin and thick stripes in human secondary visual cortex. *Elife* **12**, (2023).
11. Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *Neuroimage* **80**, (2013).

12. Salimi-Khorshidi, G. *et al.* Automatic denoising of functional MRI data: combining independent component analysis and hierarchical fusion of classifiers. *Neuroimage* **90**, (2014).
13. Robinson, E. C. *et al.* MSM: a new flexible framework for Multimodal Surface Matching. *Neuroimage* **100**, (2014).
14. Xu, T., Yang, Z., Jiang, L., Xing, X.-X. & Zuo, X.-N. A Connectome Computation System for discovery science of brain. *Sci Bull. Fac. Agric. Kyushu Univ.* **60**, 86–95 (2015).
15. Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178 (2016).
16. Gordon, E. M. *et al.* Generation and Evaluation of a Cortical Area Parcellation from Resting-State Correlations. *Cereb. Cortex* **26**, 288–303 (2016).
17. Schaefer, A. *et al.* Local-Global Parcellation of the Human Cerebral Cortex from Intrinsic Functional Connectivity MRI. *Cereb. Cortex* **28**, 3095–3114 (2018).
18. Kong, R. *et al.* Individual-Specific Areal-Level Parcellations Improve Functional Connectivity Prediction of Behavior. *Cereb. Cortex* **31**, 4477 (2021).
19. Langs, G., Golland, P. & Ghosh, S. S. Predicting Activation Across Individuals with Resting-State Functional Connectivity Based Multi-Atlas Label Fusion. *Medical Image Computing and Computer-Assisted Intervention -- MICCAI 2015* 313–320 (2015).
20. Bridgeford, E. W. *et al.* Eliminating accidental deviations to minimize generalization error and maximize replicability: Applications in connectomics and genomics. *PLoS Comput. Biol.* **17**, e1009279 (2021).
21. Shrout, P. E. & Fleiss, J. L. Intraclass correlations: uses in assessing rater reliability. *Psychol. Bull.* **86**, (1979).
22. Xu, T. *et al.* ReX: an integrative tool for quantifying and optimizing measurement reliability for the study of individual differences. *Nat. Methods* (2023) doi:10.1038/s41592-023-01901-3.

23. Friedman, J., Hastie, T. & Tibshirani, R. Regularization Paths for Generalized Linear Models via Coordinate Descent. *J. Stat. Softw.* **33**, (2010).