Release the Krakencoder: A unified brain connectome translation and fusion tool

Keith W. Jamison[1,2], Zijin Gu[3], Qinxin Wang[4], Ceren Tozlu[2], Mert R. Sabuncu[2,3], and Amy Kuceyeski[1,2]

[1]Department of Computational Biology, Cornell University, Ithaca, NY, USA
[2]Department of Radiology, Weill Cornell Medicine, New York, New York, USA
[3]School of Electrical and Computer Engineering, Cornell University and Cornell Tech, New York, NY, USA
[4]Department of Biomedical Engineering, Tsinghua University, Beijing, 100084, China
[*]Corresponding author. E-mail address: kwj2001@med.cornell.edu

ABSTRACT

Brain connectivity can be estimated in many ways, depending on modality and processing strategy. Here we present the Krakencoder, a joint connectome mapping tool that simultaneously, bidirectionally translates between structural (SC) and functional connectivity (FC), and across different atlases and processing choices via a common latent representation. These mappings demonstrate unprecedented accuracy and individual-level identifiability; the mapping between SC and FC has identifiability 42-54% higher than existing models. The Krakencoder combines all connectome flavors via a shared low-dimensional latent space. This "fusion" representation i) better reflects familial relatedness, ii) preserves age- and sex-relevant information and iii) enhances cognition-relevant information. The Krakencoder can be applied without retraining to new, out-of-age-distribution data while still preserving inter-individual differences in the connectome predictions and familial relationships in the latent representations. The Krakencoder is a significant leap forward in capturing the relationship between multi-modal brain connectomes in an individualized, behaviorally- and demographically-relevant way.

1 Introduction

The brain is a vastly complex, interconnected network of neurons (and other types of cells) whose proper function allows us to move, think, feel, and observe as well as interact with our environment. Studying how the brain's connections relate to behavior, so-called brain connectome-behavior mapping, is crucial not only for understanding how the brain works generally but also for identifying biomarkers of disease, predicting outcomes in neurological disorders, and designing personalized interventions. Brain connectivity can be probed through various structural and functional neuroimaging techniques. Structural connectivity (SC) can be captured with diffusion MRI (dMRI) measures of white matter, or the anatomical wiring, that connects different brain regions. Functional connectivity (FC), as measured with blood oxygen level dependent functional MRI (BOLD-fMRI), quantifies the temporal similarity between brain regions' dynamic neural activity patterns.The association between SC and FC is a subject of intense research, as it is believed that structural pathways provide the backbone over which functional activation flows. Despite this, correlations of SC and FC are moderate at best. Understanding the relationship between SC and FC is key to deciphering how the brain's physical structure supports its dynamic functions and may offer insights into the neural mechanisms underlying behavior, injury or disease, and recovery[1–4].

There are many approaches to modeling the relationship between SC and FC, most of which focus on using SC to predict FC. These methods include i) biophysical models, e.g. linked neural mass models[5–10], ii) graph theoretical models, e.g. communication strategies[11–15], iii) statistical models, i.e. that model SC and FC topological metrics[16–19], and iv) network control theory[20–23]. More recently, machine learning has been used to predict FC from SC[24–27], predict SC from FC[28], or to translate a connectivity estimate between parcellations[29,30]. Many of these studies have shown that metrics of SC-FC relationships vary with age, sex, and behavioral variables[16,27,31], or disease/injury and recovery[4,32,33].

Among the challenges and considerations in modeling this relationship, the assumption that estimated SC is an objective, fixed constraint on possible FC patterns is increasingly questioned. Evidence suggests that the brain's functional dynamics can be highly flexible and adaptive, and may be driven by geometric factors not measured with SC[34]. Variability in SC estimation techniques, particularly in their sensitivity to detect faint or indirect connections, further complicates the interpretation of connectivity data. Moreover, the performance metrics and loss functions currently employed in connectivity studies often fail to adequately account for inter-subject variability, potentially

skewing the results and interpretations.

Despite the utility of modeling the relationship between connectomes, there is strong methodological disagreement in the literature as to how to extract the connectomes themselves from brain images[35–37]. Studies have shown that biases inherent in different processing choices, referred to here as "flavors", can lead to different or even contradictory conclusions[38,39]. Each connectome flavor may differ in how the brain is parcellated into regions, how potential noise, artifacts, or confounds are removed, and how the connectivity between regions is quantified, among other methodological choices. Furthermore, connectome data from different studies are often shared only in certain flavors, making comparisons between datasets challenging.

In this work, our fundamental assumption is that each set of choices in the imaging and processing pipelines provides a different view of the same underlying system. Drawing inspiration from recent advances in multi-view learning[40], we develop and share a tool called "Krakencoder" that provides a way to combine these choices (i.e. connectome fusion), and thus create a more comprehensive representation of the brain's connectivity. Deriving a unified latent representation from within-modality and/or across modality connectome estimates may allow reconciliation of differing views of the same underlying network in order to overcome limitations and combine benefits from various processing choices. Krakencoder is a novel "universal" brain connectivity encoding architecture and is capable of transforming connectivity estimates between parcellations, estimation techniques, and modalities with unprecedented precision. Importantly, the Krakencoder is able to preserve inter-individual differences in a way that is behaviorally and demographically relevant. This element is crucial if we are to use the Krakencoder to improve our understanding of how brain structure and function and their interplay maps to behavior, impairment, or recovery after disease or injury.

2 Results

2.1 Krakencoder model architecture and training

The Krakencoder architecture comprises a set of encoders and decoders that allow transformation of each connectivity type into every other connectivity type via a common low-dimensional latent representation, see **Fig. 1** . The Krakencoder was trained to optimize for both reconstruction accuracy and preservation of individual differences. The latter was achieved by enforcing latent-space representations of different connectivity types from the same subject to be close to one another and far from other subjects' latent space connectome representations. The model was trained on resting-state FC and white-matter SC data from 683 healthy young adults (366 female, aged 28.6$\pm$3.8) from the Human Connectome Project, validated during training on 79 held-out subjects (43 female, aged 29.1$\pm$3.8), and finally tested on 196 subjects (107 female, aged 29.1$\pm$3.6). Care was taken to ensure that no siblings spanned the divisions, as this data leakage can lead to over-optimistic model performance. Connectivity flavors include 3 parcellations (with 86, 268, and 439 regions), 3 common FC estimates (Pearson correlation, Pearson correlation following global signal regression, and regularized partial correlation), and 2 SC estimates (volume-normalized streamline counts from deterministic and probabilistic tractography), for a total of 15.

We create a separate encoder and decoder for each of the 15 connectivity flavors and then train the model to predict each flavor from each other flavor. There are 225 total training paths, including 15 autoencoders that map between the same flavor (See **Fig. 1a**) and 210 transcoders that predict one flavor from another (See **Fig. 1b**). Each encoder is a single fully-connected linear layer, followed by L^2 normalization that constrains the latent representation to the surface of a 128-dimensional hypersphere. Each decoder is a single fully-connected linear layer. Dimensionality of the connectivity flavors varies between parcellations, which could lead to an imbalance in the training to favor higher dimension connectivity flavors. We balanced the contributions of each flavor by reducing each connectivity input to 256 dimensions using a precomputed PCA, derived using only training data to avoid leakage between train and test sets. Reconstruction loss for each training path includes both Pearson correlation and Euclidean distance. We include a contrastive component to the correlation and Euclidean loss functions, to explicitly preserve inter-subject variability of predicted connectomes. A contrastive loss term is applied to the latent representation as well to further promote inter-subject variability of the representations. The term "contrastive" here refers to a penalty that attempts to force the same subject's reconstructed and latent space connectomes close to each other and far apart from other individuals' representations. Finally, we include an intra-subject latent space consistency loss to maximize the similarity of the latent representations from each connectome flavor for a given subject (See **Fig. 1c**). This combination of loss functions allowed us to fine-tune the tradeoff between reconstruction accuracy and inter-subject variability (See **Fig. S2**), and to condition the latent space representation. See Methods for details on training procedure and loss functions.

During model inference, the Krakencoder predicts each of the 15 connectivity flavors from any other, transforming

predictions that exist in the reduced 256-dimensional space to their higher native dimensions using the inverse of the precomputed PCA transforms. The model can also combine information from multiple connectomes by averaging their latent space representations to obtain a “fusion” representation; this fusion representation can then be decoded to predict all 15 connectome flavors. The encoding and averaging step in the fusion process can include all available connectome flavors, or a subset of flavors to, for instance, predict FC flavors from the fusion latent representation of all SC flavors or vice versa (“fusionSC” or “fusionFC”, **Fig. 1d**), or to predict one parcellation using inputs from the averaged latent representations of all other parcellations (“fusion-parc”, **Fig. 1d**). The latter fusion version is particularly informative if one is using the Krakencoder to translate connectome data to a desired atlas on which it had not been computed.

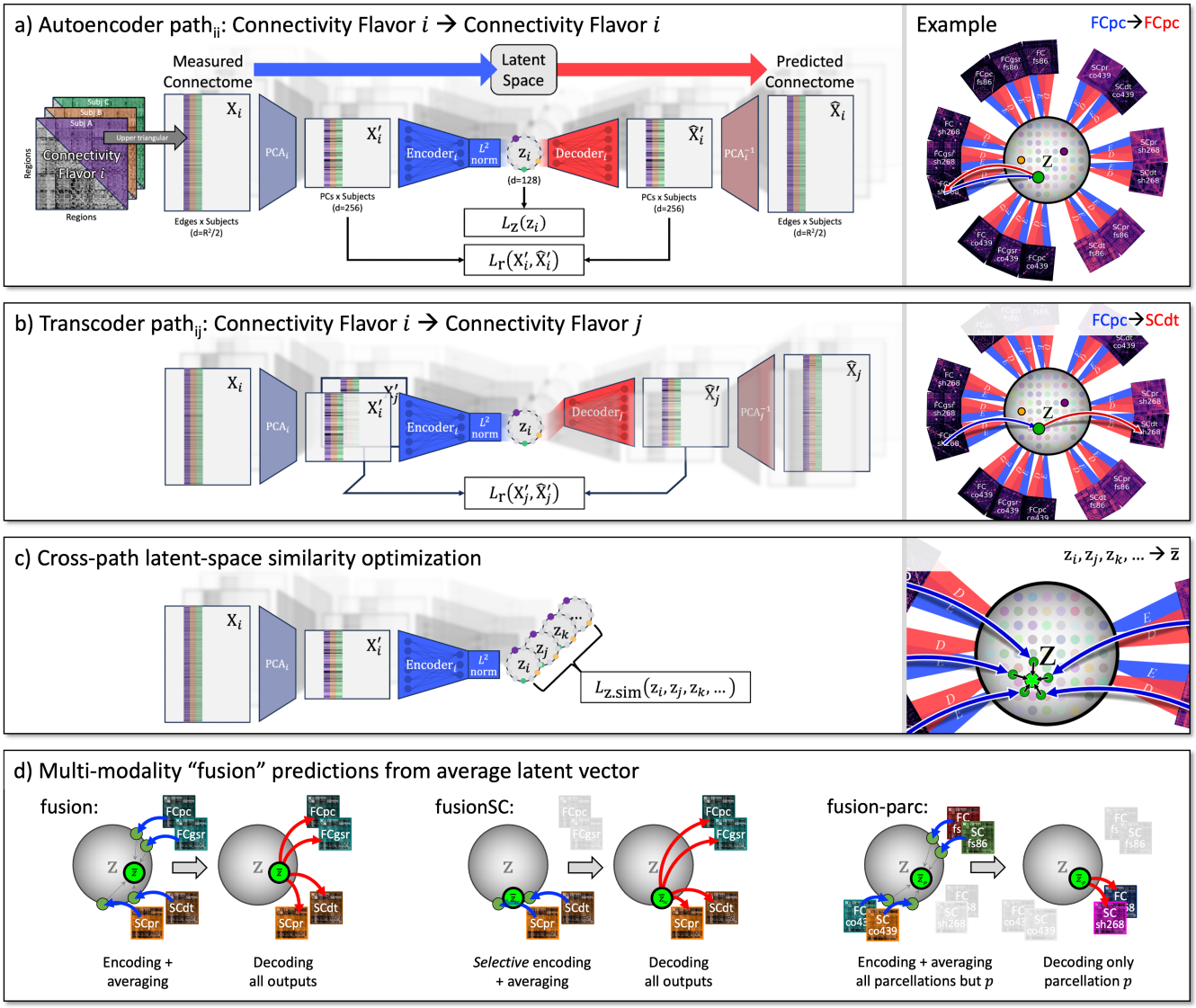

Figure 1. Key features of the Krakencoder architecture. a. Autoencoder for connectivity flavor i (i.e., $path_{ii}$) begins by stacking the upper triangular portion of each subject's connectivity matrix into input $X_i \in \mathbb{R}^{n_{edges} \times n_{subj}}$. A precomputed, fixed PCA transformation normalizes the data and reduces dimensionality to $X_i' \in \mathbb{R}^{256 \times n_{subj}}$, equalizing the size of disparate input flavors. A single fully-connected layer $Encoder_i$ followed by L^2 normalization transforms X_i' into a latent hypersphere surface $z_i \in \mathbb{R}^{128 \times n_{subj}}$. Batch-wise encoding loss $L_z(z_i)$ controls inter-subject separation in latent space. A single fully-connected layer $Decoder_i$ transforms z_i to $\hat{X}_i'$, and batch-wise reconstruction loss $L_r(X_i', \hat{X}_i')$ and $L_z(z_i)$ are backpropagated to optimize $Encoder_i$ and $Decoder_i$. **b.** Transcoder $path_{ij}$ converting input flavor i to output flavor j begins the same as $path_{ii}$, transforming $X_i \rightarrow X_i' \rightarrow z_i$, then $Decoder_j$ transforms $z_i \rightarrow \hat{X}_j'$, and reconstruction loss $L_r(X_j', \hat{X}_j')$ is backpropagated to optimize $Encoder_i$ and $Decoder_j$. **c.** Latent similarity loss $L_{z.sim}(z_i, z_j, z_k, \ldots)$ provides explicit control to ensure the latent representations of each subject are consistent across connectivity flavors. See **Table 1** for details about loss terms. **d.** Illustration of "fusion" predictions from averaged latent space vectors. For these predictions, we average the encoded latent vector from all input flavors (as in the "fusion" model), or a subset of input flavors (as in the "fusionSC" model which only averages latent vectors from SC inputs), and then decodes this average vector to all output flavors. The "fusion-parc" model demonstrates cross-parcellation prediction, by averaging only inputs from the *other* parcellations.

2.2 Measuring connectome prediction accuracy

A key consideration when assessing the performance of a connectome prediction model is the high degree of similarity of measured connectomes with the population mean. The Pearson correlation between connectomes of unrelated individuals can exceed 0.9, depending on the flavor, and a model that simply predicts the population mean for every subject can appear to have a very high prediction accuracy if it uses Pearson correlation as its metric (See **Fig. 2d** , "pop. mean" in bottom subpanel). One simple solution to this issue is to subtract the population mean calculated from training subjects before correlating measured and predicted connectomes; here we refer to this metric as $\mathrm{avgcorr}_{\mathrm{demean}}$. However, this metric still cannot distinguish whether predictions faithfully capture individual variation, particularly when there is clustering within the population (see **Fig. S1b**). Our goal here is to create individualized connectome predictions to understand the sources and consequences of across-subject variability. Thus, we also consider the identifiability of our predictions: whether our measured connectome for a subject is closer to the predicted connectome for that subject than the predicted connectome of any other subject. The "top-1 accuracy" is a winner-take-all measure that reflects the fraction of subjects for which this is true. For a more continuous measure of identifiability, more suitable to assessing imperfect predictions, we use an "average rank percentile" measure ($\mathrm{avgrank}$), which is the fractional rank of the similarity (Pearson correlation) of a subject's measured and predicted connectomes compared to all other predicted connectomes, averaged across subjects (**Eq. 4**). Taken together, prediction accuracy and identifiability, as measured by $\mathrm{avgcorr}_{\mathrm{demean}}$ and $\mathrm{avgrank}$, convey a more complete picture of how well a model explains individuals' connectomes.

In addition to these measures of prediction accuracy for the connectivity matrix values, we compute several common graph theoretical metrics to assess how faithfully the predicted connectomes capture the network properties of the measured connectomes. We use the Brain Connectivity Toolbox[41] to compute average node strength, average betweenness centrality, characteristic path length, and modularity (Louvain community detection) for each measured and predicted connectome, and compute the Spearman rank correlation between the predicted and measured values across all test subjects for each metric.

2.3 Connectome prediction performance

The accuracy of the Krakencoder's connectome predictions was assessed on 196 held-out test subjects via identifiability (top-1 accuracy and $\mathrm{avgrank}$) and (de-meaned) correlation ($\mathrm{avgcorr}_{\mathrm{demean}}$). The heatmaps in **Fig. 2a-c** illustrate the various metrics for every pair of the 15 connectome flavors, as well as "fusion" predictions decoded from the averaged latent vectors from groups of encoded connectome flavor inputs (See **Fig. 1d**). The "*average*" column and row show the average performance metric for each source and target flavor, respectively.

Top-1 identifiability (**Fig. 2a**) measures the fraction of subjects whose predicted connectome is more similar (higher Pearson correlation) to their measured connectome than any other subject's measured connectome (random chance $1/n_{subj}$ = 0.005). Average rank percentile (**Fig. 2b**) is the fractional rank of the similarity (Pearson correlation) of a subject's predicted and measured connectomes compared to all other measured connectomes, averaged across subjects (random chance 0.5). The $\mathrm{avgcorr}_{\mathrm{demean}}$ metric (**Fig. 2c**) is the average Pearson correlation between a subject's predicted and measured connectomes, after subtracting the population average of the measured connectome flavor from the predicted and measures connectomes. The average rank percentile for FCs predicted from FCs of different parcellations or connectivity estimations (FC $\rightarrow$ FC) shows near-perfect identifiability (**Fig. 2b** , upper section, upper left block mean=1.0). Similarly, SC $\rightarrow$ SC predictions have near-perfect identifiability (upper section, lower right block mean=0.99). Inter-modality predictions performed worse than within-modality, but still far exceeding chance, with the overall mean of SC $\rightarrow$ FC (lower left block) and FC $\rightarrow$ SC (upper right block) average rank identifiability of 0.82 and 0.85, respectively. In general, we find better $\mathrm{avgrank}$ prediction identifiability coincides with connectome sparsity, with higher dimensional parcellations (Coco439) and more sparse FC estimates ($\mathrm{FC}_{\mathrm{pcorr}}$) being more easily identifiable. The $\mathrm{avgcorr}_{\mathrm{demean}}$ metric (**Fig. 2c**) shows an average of 0.38 for FC $\rightarrow$ FC, and an average of 0.35 for SC $\rightarrow$ SC. Inter-modality predictions are again more difficult, with average FC $\rightarrow$ SC=0.16, and SC $\rightarrow$ FC=0.09. Lower dimensional parcellations generally had higher $\mathrm{avgcorr}_{\mathrm{demean}}$. Higher dimensional atlases are likely better at capturing inter-individual variability in the functional boundaries within an atlas and thus have higher identifiability, while lower dimensional atlases likely benefit from higher SNR within the measured regional SC and FC and thus have smaller reconstruction error when mapping between flavors.

Predictions of regularized partial correlation FC ($\mathrm{FC}_{\mathrm{pcorr}}$) had the highest $\mathrm{avgrank}$, but the lowest $\mathrm{avgcorr}_{\mathrm{demean}}$ among FC flavors, which reflects the noise inherent in this estimate, even with regularization, but also an increased sparsity that drives identifiability. SC predicted from $\mathrm{FC}_{\mathrm{pcorr}}$ yielded consistently better performance than other FC inputs, likely due to the increased sparsity and removal of non-specific correlations. The best single FC input for predicting all SC flavors was the Shen268 $\mathrm{FC}_{\mathrm{pcorr}}$, which predicts Coco439 $\mathrm{SC}_{\mathrm{pr}}$ with an $\mathrm{avgrank}$ of 0.99 and $\mathrm{avgcorr}_{\mathrm{demean}}$ of 0.19. The best single SC input for predicting all FC flavors was the Coco439 $\mathrm{SC}_{\mathrm{pr}}$.

The bottom subpanel of **Fig. 2a-c** shows connectome predictions using the "fusion" latent space average from multiple input flavors. The "fusion-parc" predictions exclude inputs of the same parcellation from the averaged latent vectors, and their relatively high values demonstrate the level of shared information across flavors within the Krakencoder's unified latent space.

Finally, we show that the Krakencoder's fusion prediction also preserves the network properties of the measured connectomes. **Fig. S6** shows the Spearman correlation between the measured connectomes' network properties (mean node strength, mean betweenness centrality, characteristic path length and modularity) and the Krakencoder's predicted connectomes' network properties, across 196 held-out individuals in the test set. The three fusion representations are used as input and each of the 15 connectome flavors as output; correlations between measured network metrics from the various familial relatedness groups are shown as comparison. The Krakencoder's predicted connectomes generally preserve the network properties of the measured connectomes as well as or better than the correlation between network properties extracted from identical twins (MZ) or two scans of the same individual (self-retest).

2.4 Connectome similarity of related individuals

The "fusionSC-parc" and "fusionFC-parc" rows in **Fig. 2a-c** represent the accuracy of the Krakencoder when predicting an SC or FC connectome flavor from all other SC or FC flavors (excluding those from the predicted parcellation). If we compare this directly to **Fig. 2d** , we can see how the Krakencoder's predicted connectome accuracy compares to several baseline estimates: measured connectomes from age- and sex-matched individuals who are unrelated, family members (non-twin siblings, dizygotic [DZ] and monozygotic [MZ] twins), retest sessions from the same individual (self retest), or the population mean connectome. For instance, the "self retest" row in the **Fig. 2d** heatmaps can be considered a measurement reliability ceiling, while the "unrelated" row shows the performance that can be explained by age- and sex-matching alone. Within a modality, top-1 accuracies (ranging from 0.93-1 for "fusionFC-parc"$\rightarrow$FC and 0.83-0.97 for "fusionSC-parc"$\rightarrow$SC) and average rank percentiles (all

equal to 1 for the "fusionFC/SC-parc") in **Fig. 2a** and **Fig. 2b** are similar to the self retest identifiability (when connectomes derived from imaging the same person twice are used to perform the identification) and better than the identifiability of individuals who are genetically identical (MZ twins). Similarly, the diagonal entries in **Fig. 2c** , representing the autoencoder $\text{avgcorr}_{\text{demean}}$ for each flavor, range from 0.13-0.74, which is comparable to the "self retest" noise ceiling (0.15-0.81), and are larger than the "MZ twin" row (0.07-0.47, a 40-110% increase). This indicates that the autoencoders with a 128-dimensional bottleneck preserve more inter-subject connectome variability than can be explained by identical genetics, and about the same amount of variability as expected when measuring the connectome from the same individual twice. The "fusionSC" and "fusionFC" rows in **Fig. 2c** represent the $\text{avgcorr}_{\text{demean}}$ for the unified SC (or FC) $\rightarrow$FC (or SC) connectome predictions, with values ranging from 0.03 to 0.16 for SC$\rightarrow$FC and 0.16-0.26 depending on the flavor. These exceed the similarity observed in the "sibling" or "DZ twin" row from the $\text{avgcorr}_{\text{demean}}$ subpanel in **Fig. 2d** , indicating that predicting FC from our unified SC representation preserves more intersubject FC variability than can be explained by age, sex, and 50% genetic similarity.

2.5 Spatial variation in edge prediction accuracy

The primary performance metrics assess global connectome similarity. We further assessed the spatial variability of prediction accuracy to determine whether certain brain regions are more accurately predicted than others. For fusion-parc, fusionFC-parc, and fusionSC-parc (fusion predictions made without the parcellation being predicted), the prediction accuracy for each edge is the Pearson correlation between the predicted and measured values for the 196 held-out test subjects. To simplify visualization and enable comparison across varied size parcellations, each region is assigned to one of 7 cortical resting state networks[42] or subcortical and cerebellar networks. The edge-wise accuracies (region $\times$ region) are then grouped and averaged into a 9×9 network representation for each connectome flavor. In the top block of **Fig. S3a** , the network-averaged edge prediction accuracy is presented for each of the fusion types as input (rows) and each connectivity flavor as output (columns). The bottom block shows the baseline network-averaged edge similarity comparing measured connectomes of unrelated age- and sex-matched subjects and related individuals. The figure demonstrates that the Krakencoder's prediction accuracy is relatively uniform throughout the cortex and subcortex. Networks with lower Krakencoder predictability (e.g., the limbic network), also show low self-retest measurement reliability (bottom row). In **Fig. S3b** , each network $\times$ network heatmap in **Fig. S3a** is correlated with the corresponding self-retest measurement reliability (bottom row), to quantify the consistency of the spatial pattern of edge-wise accuracies with the reliability of the edge-wise measurements. We see that the spatial pattern of the edge-wise accuracy of the Krakencoder's fused intra- and inter-modality predictions is as similar to the spatial pattern of the edge-wise self-retest measurement reliability as the MZ twins' spatial patterns of edge-wise reliability. **Fig. S4** summarizes the edge-wise prediction accuracy at a region level for each connectivity flavor in its own parcellation, showing a similar agreement between prediction and measurement reliability.

We observe that higher level association cortex (e.g., default mode) tends to have higher SC$\rightarrow$FC prediction accuracy than sensory-motor cortex (e.g., visual). In **Fig. S5a** , we further quantify this relationship by correlating regional prediction accuracy with an estimate of the cortical hierarchy gradient derived from functional connectivity[43], and find a significant positive correlation (Spearman r=0.249, p_{spin}=0.019, 10000 rotations[44]). Thus, our SC$\rightarrow$FC predictions explain more of the population variability in higher level cortex than in lower. Conversely, if we correlate corresponding rows in the square measured and predicted connectome matrices, we find that predicted connections to high-level cortex are less coupled to their measured rows than low-level cortex (Spearman r=-0.191, p_{spin}=0.116; see **Fig. S5b**). This latter relationship is consistent with previous work on structure-function coupling[16,19], but primarily reflects the relationship between population averaged SC and FC rather than predicting inter-individual variability.

2.6 Demographic and behavioral prediction from the Krakencoder's latent space representations

We next tested if the Krakencoder latent space preserves inter-subject variation related to family structure as well as demographic and behavioral features. **Fig. 3a** shows the inter-subject similarity of latent space representations for pairs of age- and sex-matched unrelated individuals, non-twin siblings, DZ and MZ twins, and self-retest, along with inter-subject similarity of observed connectomes. By comparing ROC separability of each pair of violin plots, we see that separation is greater for the Krakencoder latent space than for the observed connectomes, with statistical significance for unrelated vs sibling, DZ vs MZ twins, and MZ vs unrelated ($p_{perm} < .05$, 10000 permutations, FDR-corrected).

To demonstrate the preservation of demographically- and behaviorally-relevant features in the Krakencoder's

latent space, we trained linear models to predict individuals' age, sex, and cognitive scores ("total", "fluid", "crystallized", and "early childhood" composite scores from the NIH toolbox) from the Krakencoder latent space, and compared them to the same predictions based on observed connectome data. Linear kernel ridge regression models for each variable were built from the mean latent vectors (fusionFC, fusionSC, fusionFCSC, each of which is $n_{subj} \times 128$) or from the average inter-subject cosine similarity of observed data (i.e. the average of 9 FC, 6 SC, or 15 SC+FC $n_{subj} \times n_{subj}$ matrices). **Fig. 3b** shows Krakencoder and observed connectome data predict sex with a similar balanced accuracy of ≥ 0.93. The Krakencoder's latent space predicts age with Pearson $r \geq 0.5$, (fusionFC r=0.50, MAE=2.35; fusionSC r=0.51, MAE=2.57; fusionFCSC r=0.58, MAE=2.33), similar to the predictions from observed data (observed FC r=0.57, MAE=2.31; observed SC r=0.49, MAE=2.58; observed FC+SC r=0.59, MAE=2.33). For the age predictions, fusionSC outperformed observed SC and observed FC outperformed fusionFC ($p_{perm} < 10^{-3}$, FDR-corrected), but the observed data and fusion SCFC models performed similarly. Interestingly, the Krakencoder's latent space predictions significantly outperformed the observed data model predictions for predicting all cognitive composite scores. See Methods for details on inter-subject similarity calculations and model fitting.

2.7 Sensitivity analyses

Sensitivity analyses were performed to determine which brain areas' connections played the largest role in i) the accuracy of the Krakencoder's connectome mapper, and ii) the accuracy of the models predicting age, sex, or total cognition from the Krakencoder's latent space. These analyses demonstrate how the Krakencoder may be used to understand the biological underpinnings of, for example, SC to FC relationships or what brain networks may play the largest role in predicting cognition. For simplicity, regions were grouped into 7 cortical resting state networks[42] or subcortical and cerebellar networks. To perform the sensitivity analyses, every connectome edge was set to the population mean of the training set, except for the columns/rows corresponding to the regions within the network of interest, which were left untouched, see **Fig. 4a** . The fusion latent space representations (fusion, fusionSC, and fusionFC) were re-calculated for each individual and used to predict i) each connectome flavor and ii) age, sex and total cognition using re-fitted KRR models.

Fig. 4b shows the mean $\mathrm{avgrank}$ and $\mathrm{avgcorr}_{\mathrm{demean}}$ when only individual-level values of those networks' connections' were retained. The accuracy metrics for the fully intact, non-masked connectomes are shown on the right; values that are lower indicate that network's connections do not contribute as much to the accuracy of the Krakencoder's mapping. The top row shows the mean $\mathrm{avgrank}$ over the SC flavors predicted from fusionFC, the second row shows the mean $\mathrm{avgrank}$ over the FC flavors predicted from fusionSC and the third row shows the mean $\mathrm{avgrank}$ across all flavors of predicted SC and FC. We see that default mode network connections appear important for mapping accuracy, particularly for the fusion→all and fusionFC→SC. Connections to/from regions in the somatomotor network and subcortex seem to be more informative for fusionSC→FC than for fusionFC→SC.

In **Fig. 4c** , we see the largest retention in accuracy for predicting sex when functional and structural connections to/from regions in the default mode network and structural connections to/from the subcortex are retained. Age was best predicted using functional connections to/from the salience/ventral attention network and structural connections to/from the subcortex. Total cognition was best predicted by functional connections to/from regions in default mode and dorsal attention networks, and by structural connections to/from regions in the visual network.

2.8 Krakencoder performance on out-of-sample, out-of-distribution data

To assess generalizability of the Krakencoder to new datasets, we applied our frozen, pretrained Krakencoder model to data from the multi-site HCP Lifespan study[45–47], including both the Development cohort (HCP-D) and the Aging cohort (HCP-A). These two studies differ from the original HCP young adult study in both subject demographics (ages 8-21 for HCP-D and 36-100+ for HCP-A vs 22-37 for HCP-YA) and acquisition parameters, with voxel sizes and scan durations more similar to those used in typical neuroimaging studies (see Methods for more details). To account for domain shift, we linearly mapped the population mean of each connectome flavor to the population mean of the HCP-YA training subjects before applying the HCP-YA derived PCA dimensionality reduction and Krakencoder latent space projection. Despite the demographic and acquisition differences, the Krakencoder still predicts connectome flavors with high accuracy and precision. **Fig. 5a-d** shows $\mathrm{avgrank}$ and $\mathrm{avgcorr}_{\mathrm{demean}}$ for the HCP-D (age 8-21) and HCP-A (age 36-100+) cohorts. Within-modality connectome prediction identifiability remained high for both HCP-D and HCP-A ($\mathrm{avgrank}$ > 0.96). Inter-modal prediction identifiability was reduced 6-12% compared to held-out HCP-YA subjects, though still well above chance ($\mathrm{avgrank}$ > 0.75). Intra-modal $\mathrm{avgcorr}_{\mathrm{demean}}$ was reduced by 10-26%, while inter-modal $\mathrm{avgcorr}_{\mathrm{demean}}$ was 30-50% lower than held-out HCP-YA. Thus, despite significant differences in age and acquisition reducing prediction accuracy, the Krakencoder model still performs well above chance and preserves inter-individual differences when mapping between connectome flavors.

The HCP-D dataset contains 76 sibling pairs (50 same-sex, 26 opposite sex, Δage 0-9yr), so we performed a familial similarity analysis similar to that for HCP-YA. In **Fig. 5e** , we show that siblings were more separable from age- and sex-matched individuals in latent space than in the observed data for all three fusion types (fusionFC, fusionSC, fusionFCSC) ($p_{perm} < 10^{-3}$, FDR-corrected). Finally, we performed a regression analysis to predict age, sex, and cognition in the HCP-D and HCP-A datasets (See **Fig. S7**). New predictive models using the same approach as the HCP-YA predictions were fit and cross-validated for each of the two new datasets. As with HCP-YA, sex was predicted very accurately by both latent and observed data models (balanced held-out test accuracy = 0.86-0.95). Age was predicted with much higher accuracy than in HCP-YA (HCP-D r=0.84-0.95, MAE=1.45-1.72; HCP-A r=0.76-0.85, MAE=6.13-6.52 on held-out test data), likely due the much larger age range and larger age-related changes during development and aging. Cognitive score prediction performance for the HCP-A data was similar to or better than HCP-YA, perhaps due to larger age-related cognitive effects, but performed considerably worse for the HCP-D dataset. This decrease in predictive ability was especially pronounced when predicting cognition from SC, where accuracy (Pearson's r) from latent and observed data models was not significantly greater than 0. For all other predictions, observed data models performed significantly better than those on the Krakencoder latent space ($p_{perm} < 10^{-3}$, FDR-corrected), though the Krakencoder predictions remained well above chance.

We aim to further test Krakencoder's biological plausibility by applying it to connectomic data from individuals with pathological changes in the brain's structural and functional connections. We extracted SC and FC from 100 individuals with multiple sclerosis (age: 22-71, 45.5$\pm$11.4 years; 66% female; disease duration: 12.97$\pm$8.07 years), an autoimmune disease that results in lesions primarily in the brain's white matter[48]. This disease has also been associated with changes in SC and FC that are related to individuals' disability levels[49,50]. The fMRI and dMRI acquisitions for these patients had significantly lower spatial, temporal, and angular resolution than the HCP-YA or HCP-Lifespan data (see **Table S1**). We fed the observed, pathological SC and FC into the corresponding arms of the frozen Krakencoder to obtain a fusionSC and fusionFC, from which we predicted the various FC or SC flavors. We found that the average rank identifiability of the Krakencoder's predicted FC (from the fusionSC) was up to 71%, which was higher than the identifiability of the observed SC to observed FC (calculated within the same atlas), which was at most 54%, see **Fig. S8** . Mapping the fusionFC, calculated from fMRI in the MS patients, to the various SC flavors using the Krakencoder had around the same identifiability (up to 63%) as the mapping between observed FC to SC. These results demonstrate that even though the Krakencoder was trained on young, healthy individuals, it can still represent with good fidelity the relationship between connectomes in individuals with pathology. Importantly, it may indeed better capture the mapping from SC to FC than the observed connectome data itself. This provides evidence of the biological plausibility of the Krakencoder's architecture and hints at its possible clinical significance in capturing low-dimensional representations of disease pathology.

2.9 Comparison with existing techniques

While predicting FC from SC is not the only application of our model, this route of prediction is among the most commonly studied problems in brain connectivity mapping. To compare our current Krakencoder model with this previous literature, we gathered a few recent studies that used different approaches to predict FC from SC. The first is a deep neural network proposed by Sarwar et. al[24] ("deepnet") and the second is a graph neural network from Neudorf et. al[25] ("graphnet"). The Krakencoder model incorporates 6 flavors of SC (2 per parcellation) and 9 flavors of FC (3 per parcellation). These two previous SC-FC prediction frameworks are designed to predict a single flavor of FC from SC. Thus, we trained 12 separate models for each of the two frameworks for all combinations of SC_{dt}/SC_{pr} and FC/FC_{pcorr}. FC_{gsr} was excluded due to computational constraints and relative redundancy with FC.

By both $\mathrm{avgrank}$ and $\mathrm{avgcorr}_{\mathrm{demean}}$ metrics, the predictions from our model outperform both "deepnet" and "graphnet" models across all connectome flavors and parcellations (See **Fig. S9**). Specifically, the Krakencoder $\mathrm{avgrank}$ was 17-86% higher (average 42%) than "deepnet" or "graphnet" models for the same single SC inputs, and 18-90% higher (average 54%) using the fusionSC input to predict each FC flavor. Krakencoder $\mathrm{avgcorr}_{\mathrm{demean}}$ was 1-890% higher (average 335%) than either model using single SC inputs, and 25-980% higher (average 450%) using fusionSC. Similarly, the accuracy of network properties for FC predicted by Krakencoder exceeded that from either alternative SC to FC framework. These particular results highlight the special emphasis of the Krakencoder on preservation of individual differences when mapping from SC to FC.

A note on computational efficiency: training all 12 "deepnet" models required approximately 140 GPU hours on an Nvidia A100 (4x5 hours for FS86, 4x10 hours for Shen268, and 4x20 hours for Coco439), and all 12 "graphnet" models required 1200 GPU hours (4x50 hours for FS86, 4x100 hours for Shen268, 4x150 hours for Coco439). For comparison, the Krakencoder required only 30 GPU hours to train a model for all 15x15=225 prediction paths simultaneously. When applying the Krakencoder model to new connectome data, the computational requirements

are minimal and can even be done using free cloud-based platforms like Google Colab.

2.10 Extending the Krakencoder to new connectivity flavors

To broaden the applicability of our model, we demonstrate a means to extend the pre-trained Krakencoder model beyond the 15 flavors shown here. As depicted in **Fig. S10a** , given a new connectivity flavor computed on the same training subjects as the pre-trained model, we train a new autoencoder with an additional loss term to force the new latent representation into alignment with the pre-computed fusion latent representation. To illustrate this capability, we extended our model to new data from a previously unseen parcellation with bandpass-filtered time series for FC (original flavors use a high-pass filter) and un-normalized streamline count for SC (original flavors use pairwise volume normalization). As seen in **Fig. S10b** , prediction identifiability and reconstruction accuracy are comparable to the original 15 flavors. Furthermore, unlike training the original 225 path model, which takes up to 30 hours on a GPU, adding each new flavor takes only 1-2 minutes without a GPU. In this example, we used the original 683 HCP-YA training subjects, but this model-extension technique can use any population for which have already computed the Krakencoder latent representation.

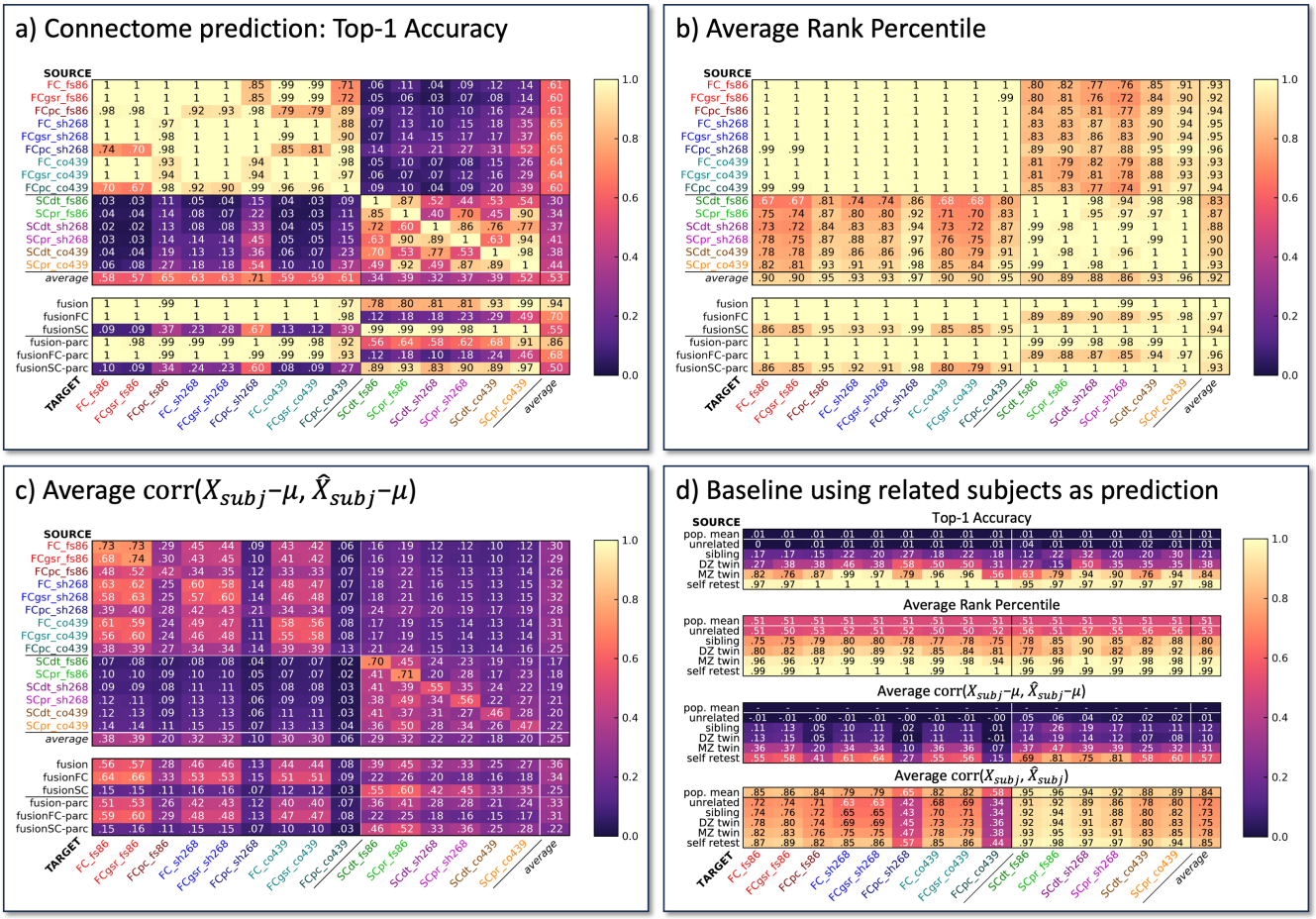

Figure 2. Connectome prediction performance for all input and output flavors on 196 held-out HCP-YA subjects. Heatmaps in **a**-**c** show connectome prediction performance from each *source* flavor (row) to each *target* flavor (column). Flavors include 3 parcellations (fs86, sh268, co439), 3 FC estimates (FC=Pearson, FCgsr=global signal regressed Pearson, and FCpc=regularized partial correlation), and 2 SC estimates (SCdt=deterministic tractography, SCpr=probabilistic tractography). The "*average*" row is the mean of performance metrics for all sources in the column above. The "fusion" row shows predictions from the average latent vector from all 15 connectome flavors (See **Fig. 1d**). "fusionFC" and "fusionFC" are predicted from the average of 9 FC or 6 SC latent vectors, respectively. "fusion-parc" predicts each parcellation from the average of all parcellations *except* itself. **a.** Top-1 accuracy measures the fraction of test subjects whose predicted connectome is closer to their own measured connectome than any other subject's measured connectome. Chance level is 1/196=0.005. **b.** Average rank percentile measures the fraction of *other* subjects' measured connectomes that are further from the target subject's predicted connectome than its own measured connectome. This rank percentile is then averaged across subjects. Chance level is 0.5. **c.** Average correlation of each subject's measured and predicted connectomes, after subtracting the population average ($\text{avgcorr}_{\text{demean}}$). **d.** For each subject s_a, we select an age- and sex-matched individual s_b of a given relatedness (unrelated, sibling, DZ/MZ twin, or self retest) and consider the measured connectome from s_b as the prediction for s_a. For each relatedness level, we compare the set of pseudo-predictions to measured connectomes using each performance metric, as though they were predictions from our model. Thus, we can observe the reliability and utility of each flavor and performance metric, and provide multiple baselines for evaluating model performance. We also show performance using the population average for each flavor ("pop. mean") as pseudo-prediction, which for the avgcorr metric (bottom) exceeds test-retest, but for all other metrics is zero or chance.

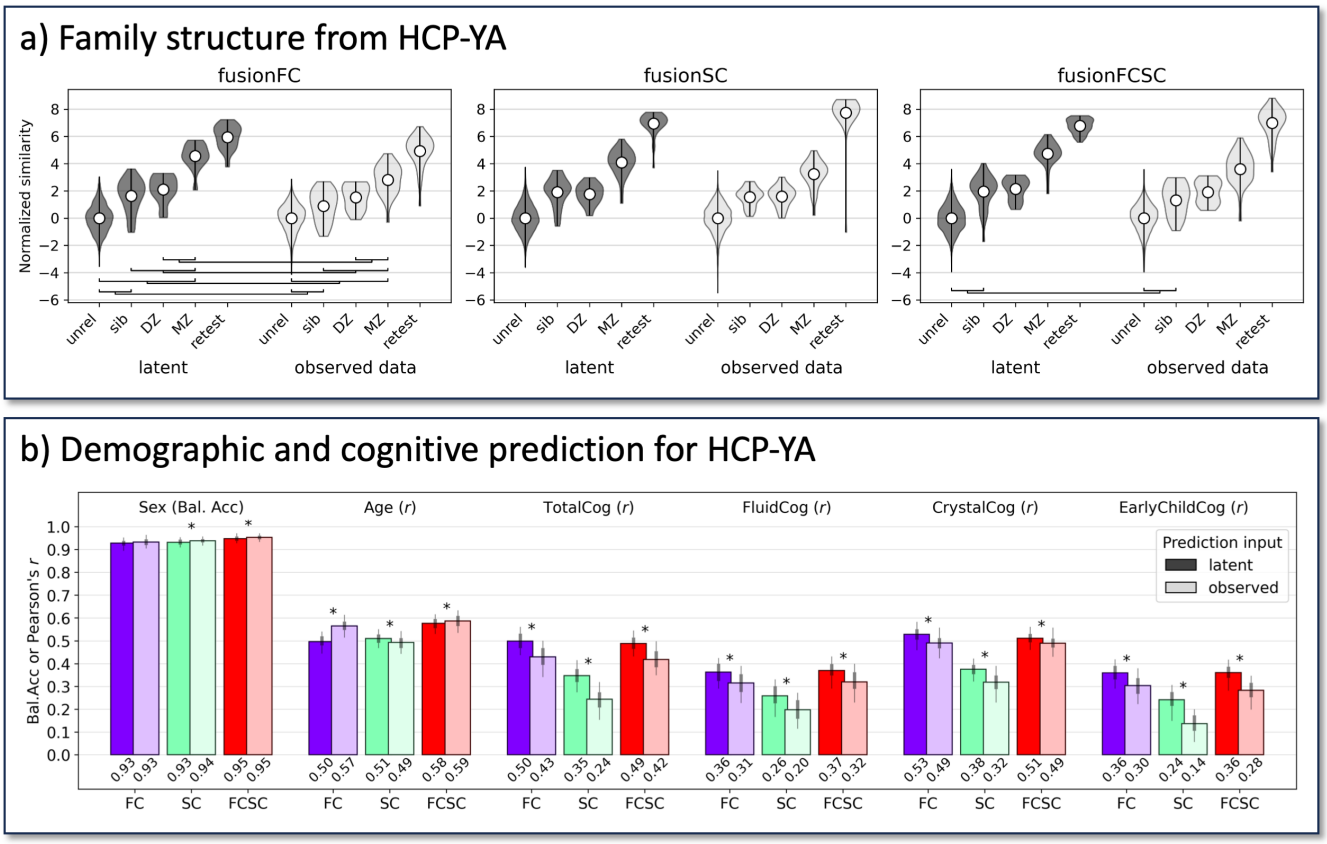

Figure 3. Familial connectome similarity and demographic prediction for held-out HCP-YA subjects. a. Inter-subject similarity of latent space representations and observed SC/FC data, stratified by family relationship. Each violin plot shows the distribution of inter-subject similarity for age- and sex-matched pairs for each familial relationship, as well as unrelated pairs and self-retest. Due to the high baseline similarity in observed data, latent and observed data similarity measures were independently z-scored by the mean and standard deviation of the similarity of unrelated subjects for visual comparison. Separability of each pair of violins within a modality was computed as the area under the receiver operating characteristic (ROC) curve. Separability estimates for a given relationship pair from the latent space and observed data were statistically compared using a permutation test, where bars show family-level pairs where latent space separability exceeds observed data separability ($p_{perm} < .05$, 10000 permutations, FDR-corrected). **b.** Prediction accuracy of demographic and cognitive metrics from different subsets of the Krakencoder's latent space and observed SC and FC data. Variability of prediction accuracy was assessed through bootstrap resampling (N=100). $*=$ significant difference between prediction accuracy from latent space and observed data. ($p_{perm} < 10^{-3}$, 10000 permutations, FDR-corrected).

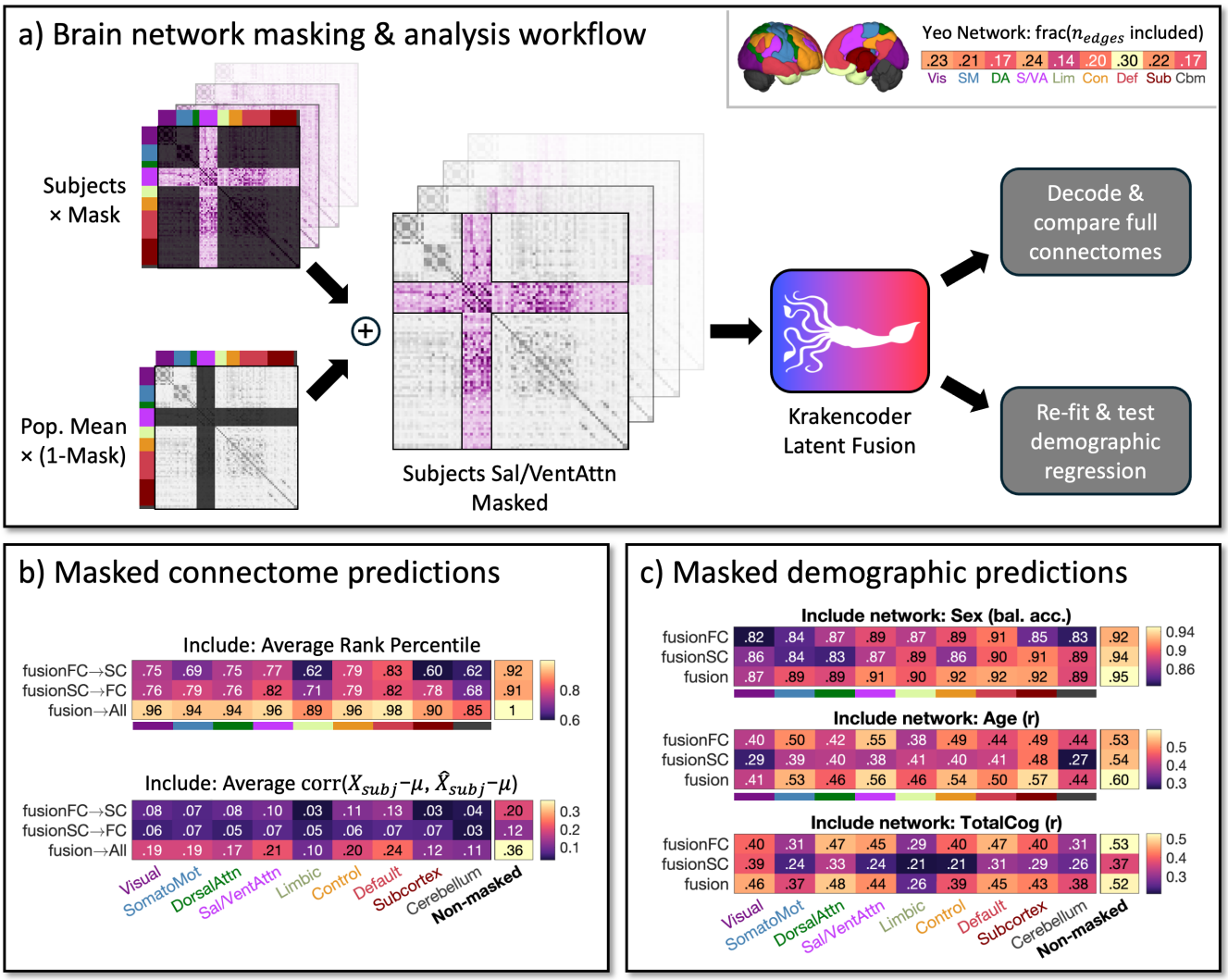

Figure 4. Spatial input sensitivity analysis of Krakencoder predictions. a. We estimate the sensitivity of the Krakencoder's prediction accuracy to regions within a given network by replacing all connectome edge values not in that network with the population mean and feeding these masked connectomes into the connectome and demographic predictions. The accuracy after masking reveals the amount of information that particular network's regional connections have in the mapping and demographic/behavioral predictions. **b.** Connectome prediction identifiability (avgrank, top) and reconstruction accuracy ($\text{avgcorr}_{\text{demean}}$, bottom) using only the connections to/from regions within each brain network shows how much information the model is utilizing from each network when predicting whole-brain connectomes. The "non-masked" column on the right shows performance using original connectomes without masking. **c.** Similarly, predicting subject sex (top), age (middle), and total cognition (bottom) from masked input data shows the relative amount of information the connections to/from regions in each network contain about those demographic and cognitive features.

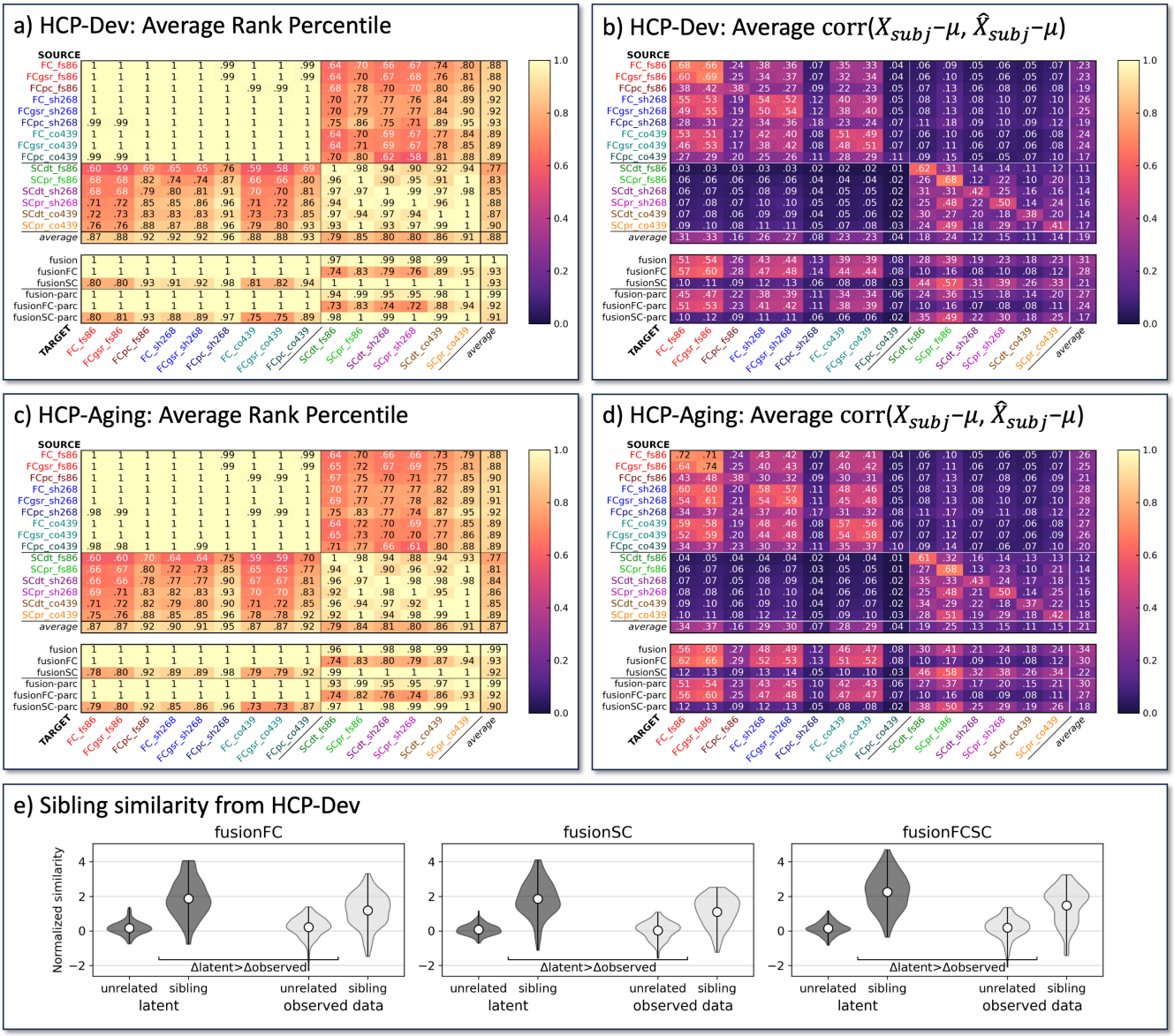

Figure 5. Out-of-sample performance of the Krakencoder on individuals from across the lifespan. Heatmaps in **a**-**b** show connectome prediction performance for subjects from the HCP-Development study (ages 8-21) and heatmaps in **c**-**d** show connectome prediction performance for subjects from the HCP-Aging study (ages 36-100). The Krakencoder model applied was trained on HCP-YA data (ages 22-37) and frozen; no re-training or fine-tuning was performed. **a,c.** Average rank percentile (**Fig. 2b**) is the fractional rank of the similarity (Pearson correlation) of a subject's predicted and measured connectomes compared to all other measured connectomes, averaged across subjects (random chance 0.5). **b,d.** Average correlation of each subject's measured and predicted connectomes, after subtracting the population average. **e.** Connectomes from sibling pairs in HCP-D are more similar than age- and sex-matched unrelated pairs, with latent space similarity significantly more separable than observed data. Each violin plot shows the distribution of inter-subject similarity for sibling pairs or pairs of unrelated individuals with matching age and sex disparity. As in **Fig. 3a** , latent and observed connectome similarity were independently normalized for visualization. Separability was measured by ROC analysis, and significance was assessed using permutation testing. For all three fusion types (fusionFC, fusionSC, fusionFCSC), siblings were more similar in the Krakencoder's latent space than in the observed data ($p_{perm} < 10^{-3}$).

3 Discussion

Here, we present the Krakencoder, a joint connectome encoder/decoder tool, capable of simultaneously, bidirectionally translating between structural and functional connectivity, and connectomes from different atlases and processing choices. The Krakencoder's predicted connectomes match measured connectomes with an unprecedented level of accuracy and individual-level identifiability, while also preserving their network properties. The Krakencoder's across modality mapping (from SC to FC and vice versa), which is arguably the most commonly performed mapping, had identifiability that was 42-54% higher than existing deep learning models. In addition to translating between connectome flavors, the Krakencoder can fuse connectome representations to create a single low-dimensional representation of all measured connectome flavors, each of which provide a different view of the true underlying connectome. The latent space of the Krakencoder appears to i) better reflect familial relatedness (compared to observed connectome data), ii) preserve age- and sex- relevant information and iii) enhance cognition-relevant information (again, compared to the observed connectome data). Finally, we demonstrated that the Krakencoder can be applied without any retraining to completely new, out-of-age-distribution lifespan data and still preserve inter-individual differences in the connectome predictions and familial relatedness in the fused latent space representations. Overall, we believe the Krakencoder has several novel aspects and applications, and is a significant leap forward in quantifying the relationship between multi-modal brain connectomes and how these relationships map to behavior.

Previous work in mapping between connectomes, mostly predicting FC from SC, has largely focused on optimizing accuracy in the model predictions by maximizing the correlation between predicted and measured FC. However, we argue here that over-emphasis on this metric can result in a model that predicts FC that is close to the population mean and/or does not preserve individual differences. The "deepnet" model did incorporate into its loss function a term that aimed to reduce inter-subject similarity of predicted FC[24], however it appears that this only pushed the individual FC predictions away from one another and not toward the particular individual's measured connectomes as its identifiability was just above chance. The Krakencoder's connectome predictions, on the other hand, achieve perfect average rank identifiability within the same connectome modality, and consistently high average rank identifiability predicting FC from SC. Other modeling work has revealed that regional SC-FC coupling follows a functional gradient wherein the coupling is highest in visual and somatomotor networks and lowest in higher-order association networks[16,19]. The Krakencoder predictions of FC from fusionSC follow this trend where the regional, row-wise correlation of observed FC and the Krakencoder's predicted FC from fusionSC appears to be linearly decreasing with increasing regional hierarchy (see **Fig. S5**). This metric of row-wise correlation is very different from the across-subject, network-level correlations we are illustrating in **Fig. S3** that appear to be inversely correlated with regional hierarchy. As the Krakencoder architecture predicts each edge from all other regions and edges, rather than a single row, the edge-wise accuracy appears to be more aligned with measurement reliability than regional coupling. (see **Fig. S3** and **Fig. S4**).

The Krakencoder has many novel aspects, including its ability to translate between connectome flavors without requiring raw data. This is useful when researchers only have access to a connectome from a particular pipeline or atlas and need it in another. One previously developed tool called CAROT does allows translation of a connectome from a different atlas into another without requiring the raw data[30]; however, this tool does not allow translating connectomes between different processing pipelines or mapping between SC and FC. There is not one FC or SC processing pipeline that has been agreed upon by the field; many researchers chose one pipeline and replicate on others. The ability of the Krakencoder to predict other flavors within the same modality suggests a common, latent connectomic backbone that gives rise to each FC or SC flavor when a given parcellation or connectome estimation strategy is employed. The Krakencoder's fusion functionality removes the burden of having to choose one pipeline or replicate findings on several different pipelines and attempt to make general conclusions. More than that, we believe the fusion of various snapshots of the underlying connectome increases the SNR of the connectome representations and perhaps becomes more than the sum of its parts. This latter statement is supported by the finding that models predicting cognition from the fused latent space representations outperformed models based on the observed connectome data; age and sex predictions were similar between the models but these are generally easier to predict compared to cognitive scores. The Krakencoder's SNR boosting effect will be especially useful when the tool is applied to connectomes extracted from lower-quality data, which is often the case in clinical populations. Here, using lower-quality MRI data from individuals with multiple sclerosis, we demonstrate that the fusionSC to FC mapping has better identifiability (and the fusionFC to SC mapping similar identifiability) than the observed connectome data.

The Krakencoder is robust when applied to completely new, out-of-distribution data from individuals across the lifespan (8-100+) and from adults with pathological changes to their connectomes (individuals with multiple sclerosis, or MS). The frozen model (trained on young adult data) still retained a high level of identifiability and explained

variance when mapping between connectomes from the HCP-D, HCP-A, and MS datasets. The Krakencoder appears somewhat impervious to the sweeping connectomic changes that occur throughout human development, aging, and, to a certain extent, disease processes[51,52]. The drop in demographic predictions of the latent space for the HCP-D, HCP-A, and MS data may mean there may be a benefit of the Krakencoder from further training (perhaps fine tuning) using lifespan and /or disease data.

Future work will apply the Krakencoder to predict connectomes under other clinical conditions, and further explore the relationship between the common latent representation and factors of neuroscientific or clinical significance. We will also create a fine-tuning procedure to further increase the robustness of the Krakencoder when applied to novel datasets in order to promote its wider applicability, particularly in datasets that contain pathology like stroke and Alzheimer's disease. The Krakencoder could also be adapted to have encoding and decoding arms for data from other modalities (MEG or EEG) or even to explicitly capture demographic/behavioral information so that the latent space may be used more effectively for brain-behavior mapping. While we purposefully chose atlases that were derived in varied ways, i.e. using anatomical vs functional information, there still may be disagreement in the labeling of regions across subjects. However, by translating between and fusing SC and FC derived from all atlases, the Krakencoder attempts to identify a latent organizational representation that underlies those observed in different overlapping parcellation schemes. In the context of SC-FC coupling, we do not think these atlas choices prevent its accurate evaluation within an individual, since we are using the same parcellation for both SC and FC. Finally, while this first proof-of-concept work focused on applying the Krakencoder to static FC, we know that brain activity is flexible and adapts to task demands, varied states, and pharmacological manipulations. Further work will investigate adapting the Krakencoder to better reflect dynamic changes that occur in brain networks.

4 Methods

4.1 Data description

The data for this study come from the original Human Connectome Project (HCP)[53] and the HCP Lifespan studies[45–47], which together includes high-resolution 3T MRI data for 2490 healthy participants aged 8-100+. Primary model training and testing was performed using data from 958 subjects from the original "young adult" HCP study (HCP-YA), aged 22-37, which includes monozygotic and dizygotic twin pairs as well as non-twin siblings. For benchmarking performance, we include a subset of 40 HCP-YA subjects who returned for a complete "retest" session 1-11 months after their initial session. The HCP-Development (HCP-D, 608 subjects, aged 8-21, including 76 pairs of non-twin siblings) and HCP-Aging (HCP-A, 716 subjects aged 36-100) cohorts from HCP Lifespan were used to evaluate our pre-trained model on out-of-sample, out-of-distribution data. HCP Lifespan data differs from HCP-YA in both age range and acquisition parameters, with Lifespan acquiring less than half as much fMRI and dMRI data as HCP-YA. HCP Lifespan studies were also collected across four different sites using standard MRI scanners while the HCP-YA data were all acquired on a custom MRI scanner at a single site. See **Table S1** for demographic and acquisition details. For all three datasets, we excluded subjects with incomplete resting-state fMRI or diffusion MRI, and subjects with known acquisition or QC issues.

4.2 Construction of Functional Connectomes

All resting-state fMRI timeseries were preprocessed by HCP using the Minimal Preprocessing Pipeline[54], which includes motion correction, susceptibility distortion correction, coregistration to anatomy and nonlinear template space (FSL MNI152 template, 6th generation), as well as automated denoising using ICA-FIX[55]. We used a custom post-processing pipeline to identify motion and global signal outlier timepoints (global signal>5σ, motion derivative >.9mm), regress out tissue-specific nuisance time series (5 eigenvectors each from eroded white matter and CSF masks[56]) and motion-related time series (24 total, including 6 motion parameters, their backward derivatives, and the squares of both[57]), temporally filter (high-pass >0.01 Hz for HCP-YA, >0.008 Hz for HCP-A/D, using DCT projection), and parcellate time series using each atlas. Outlier timepoints were excluded from nuisance regression and temporal filtering. Regional time series for each brain region were obtained by averaging voxels in the denoised time series data. Regional time series for each of the four (for HCP-YA) or two (for HCP-Lifespan) scans were variance-normalized and concatenated, and Pearson correlation between regional time series (excluding outlier timepoints) resulted in an region$\times$region resting-state functional connectivity (FC) matrix for each subject. We also computed an FC matrix with global signal regression, FC_{gsr}, by regressing the mean gray matter time series and its temporal derivative from each regional time series before computing region$\times$region Pearson correlation. Finally, we computed a Tikhonov-regularized partial correlation FC_{pcorr} (or FCpc) that minimizes the average distance between subject FC_{pcorr} and the unregularized population average partial correlation[36]; see **Supplemental Methods** for regularization details.

4.3 Construction of Structural Connectomes

Diffusion MRI were preprocessed by HCP using the Minimal Preprocessing Pipeline, which jointly corrected for motion, EPI distortion, and eddy current distortion using FSL's "topup" and "eddy" tools[58,59], before being linearly coregistered to the anatomical image. Preprocessed diffusion MRI were further processed using MRtrix3[60], including bias correction, constrained spherical deconvolution (multi-shell, multi-tissue FOD estimation, lmax=8[61]), and whole-brain tractography. We performed both deterministic (SD_STREAM[62], or SCdt) and probabilistic, anatomically-constrained tractography (iFOD2+ACT[63,64], or SCpr) using dynamic white-matter seeding[65] for both methods, resulting in 5 million total streamlines per subject for each tractography algorithm. Structural connectivity (SC) matrices were constructed for each atlas by counting the number of streamlines that ended in each pair of gray matter regions, normalized by the total volume of each region pair.

4.4 Parcellations

We used three different whole-brain atlases, covering a range in size and construction methodology. The 86-region FreeSurfer atlas (FS86) combines 68 cortical gyri from Desikan-Killiany and 18 subcortical gray matter regions (aparc+aseg[66,67]). The 268-region Shen atlas (Shen268 or sh268) is an MNI-space cortical and subcortical volumetric atlas based on resting-state fMRI clustering[68,69]. The 439-region CocoHCP439 atlas (Coco439 or co439) combines 358 cortical regions from the HCP multimodal parcellation[70], defined using anatomical and functional connectivity gradients, with 12 subcortical regions from FreeSurfer "aseg" (further modified by FSL FIRST[71]), 30 thalamic nuclei derived from FreeSurfer7[72] (50 original outputs included many small nuclei, which were merged into

the final set of 30), 27 cerebellar regions from the SUIT atlas[73], and 12 additional subcortical nuclei from AAL3[74]. The FS86 and Coco439 atlases were defined based on each individual subject's FreeSurfer surface and subcortical parcellations, whereas the Shen268 atlas was applied directly to MNI-resampled data for each subject. The FS86, Shen268, and Coco439 parcellations result in 3655, 35778, and 96141 pairwise connectivity estimates, respectively.

4.5 Evaluating connectome predictions

As previously discussed, and as illustrated in **Fig. S1** , we consider several complementary assessments of how predicted connectomes (denoted $\hat{X}_s$ for subject s) reflect measured connectomes (denoted X_s). Based on these considerations, we use the following metrics for evaluating model performance:

$$\text{avgcorr} = \frac{1}{n_{subj}} \sum_{s=1}^{n_{subj}} \text{corr}(X_s, \hat{X}_s) \tag{1}$$

$$\text{avgcorr}_{\text{demean}} = \frac{1}{n_{subj}} \sum_{s=1}^{n_{subj}} \text{corr}(X_s - \mu_{\text{tr}}, \hat{X}_s - \mu_{\text{tr}}), \text{ where } \mu_{\text{tr}} = \left(\frac{1}{n_{subj}} \sum_{s=1}^{n_{subj}} X_s\right) \text{ for training subjects} \tag{2}$$

$$\text{top1acc} = \frac{1}{n_{subj}} \sum_{s=1}^{n_{subj}} [\text{argmax}_{a \in 1 \ldots n_{subj}}(\text{corr}(X_s, \hat{X}_a)) = s], \text{ where } [\cdot] \text{ represents an indicator function} \tag{3}$$

$$\text{avgrank} = \frac{1}{n_{subj}} \sum_{s=1}^{n_{subj}} \left(\frac{1}{n_{subj}} \sum_{a \neq s}^{n_{subj}} [\text{corr}(X_s, \hat{X}_a) < \text{corr}(X_s, \hat{X}_s)]\right) \tag{4}$$

4.6 Model training

Each model training epoch comprises two stages: 1) path-wise reconstruction and inter-subject dispersion and 2) latent space consistency. For a given path $i \rightarrow j$, we loop through batches of subjects ($n_{batch} = 41$), apply $Encoder_i$ to transform input flavor X'_i ($256 \times n_{batch}$) to latent representation z_i ($128 \times n_{batch}$), and $Decoder_j$ to transform z_i to predicted output flavor $\hat{X}'_j$ ($256 \times n_{batch}$). Path-wise reconstruction loss and inter-subject dispersion terms ($L_{r.corrI}$, $L_{r.mse}$, $L_{r.marg}$) are computed per batch from X'_j and $\hat{X}'_j$. These reconstruction losses collectively penalize reconstruction error in both absolute MSE and correlation terms, and promote inter-subject separation in the reconstruction space. Path-wise latent space dispersion loss ($L_{z.corrI}$, $L_{z.dist}$), computed from z_i, promotes inter-subject separation in the latent space. This forward pass and backpropagation is repeated for each batch of training subjects, before moving on to the next training path. The order in which paths are selected is randomly reshuffled each epoch. The second stage of each training epoch, latent space consistency, enforces similarity of the latent space representations from each input flavor. We re-compute z_i for all input flavors from the encoders updated in the path-wise stage, and then compute $L_{z.sim}$ as the total MSE between the set of z_i matrices for all flavors. Explicit weights were applied to the reconstruction MSE loss term $L_{r.mse}$ ($w_{rm} = 1000$), the combined path-wise latent space dispersion loss L_z ($w_z = 10$), and the latent space intra-subject consistency loss $L_{z.sim}$ ($w_{zs} = 10000$). See **Algorithm 1** for pseudocode depiction of training procedure, and **Table 1** for details of loss function terms.

After each training epoch, we compute predicted connectomes for all paths on a held-out set of 80 validation subjects, and the inverse PCA transform (pre-computed on the training data only) is applied to compute $\hat{X}_j \in \mathbb{R}^{n_{edges}}$ from $\hat{X}'_j \in \mathbb{R}^{256}$. Reconstruction accuracy and identifiability are computed in this high-dimensional data space. Loss term weights were tuned empirically to balance the reconstruction ($\text{avgcorr}_{\text{demean}}$) and identifiability ($\text{avgrank}$) performance metrics in the validation dataset. **Fig. S2** illustrates this trade-off when tuning MSE reconstruction loss.

The model was implemented in PyTorch 1.10, using the AdamW optimizer with learning rate 10^{-4}. Model weights were regularized by weight decay 10^{-2}, as well as dropout in all encoders and decoders (dropout rate 0.5). The model was trained for 5000 epochs, which took approximately 36 hours on an Nvidia A100 GPU. Due to the regularization and the largely linear nature of the model, we did not observe significant overfitting. Validation performance plateaued after 500-1000 epochs, and we selected the 2000 epoch checkpoint for further evaluation.

Algorithm 1 Krakencoder model training procedure. (See **Table 1** for loss function details)

Initialization: $X'_i \leftarrow \text{PCA}(X_i)$ for all $i \in 1 \ldots n_{flav}$ # Transform inputs into $\mathbb{R}^{256 \times n_{subj}}$ PCA space

for *epoch* = 1, . . . , n_{epochs} **do**

  # Stage 1: Path-wise reconstruction loss
  for each $path_{ij} \in \text{shuffle}((i,j) \in 1 \ldots n_{flav})$ **do** # Loop through each path $i \rightarrow j$
    for each *subject_batch* **do**
      $z_i \leftarrow Encoder_i(X'_i)$
      $\hat{X}'_j \leftarrow Decoder_j(z_i)$
      $L_{path} \leftarrow L_r(X'_j, \hat{X}'_j) + w_z * L_z(z_i)$ # Latent space and reconstruction loss for $X'_i \rightarrow z_i \rightarrow \hat{X}'_j$
      backprop(L_{path}) # Update $Encoder_i$ and $Decoder_j$ for this *subject_batch*
    end for
  end for

  # Stage 2: Latent space consistency
  for each *subject_batch* **do**
    $z_i \leftarrow Encoder_i(X'_i)$ for all $i \in 1 \ldots n_{flav}$ # All input flavors $\rightarrow$ latent space using updated $Encoder_i$
    $L_{z.sim} \leftarrow \sum_{i,j \neq i} \|z_i - z_j\|_2^2$ # Similarity of latent space across flavors for each subject
    backprop($L_{z.sim}$) # Update all $Encoder_i$ for this *subject_batch*
  end for

  Evaluation: $\hat{X}_i \leftarrow \text{PCA}^{-1}(\hat{X}'_i)$ for all $i \in 1 \ldots n_{flav}$ # Predicted PCA outputs $\rightarrow$ high dimensional $\mathbb{R}^{n_{edges} \times n_{subj}}$
end for

4.7 Procedure for comparing family structure and predicting demographics

To examine how the connectomes (predicted and measured) reflect family structure, we first compute the averaged latent vector in the Krakencoder for each subject from either all 15 flavors ("fusion") or the 9 FC ("fusionFC") and 6 SC ("fusionSC") separately, and then compute a $n_{subj} \times n_{subj}$ inter-subject latent similarity matrix using Pearson correlation. These pairwise similarities are grouped according to the inter-subject relationship. For comparison, we also compute the inter-subject similarity for all 15 flavors of the high-dimensional measured connectome data, average these 9, 6, or 15 $n_{subj} \times n_{subj}$ similarity matrices, and group them according to relatedness. Due to the high baseline similarity in observed data, latent and observed data similarity measures were independently z-scored by the mean and standard deviation of the similarity of unrelated subjects for visual comparison. Differences of ROC separability of the distributions (e.g., separability of DZ and MZ distributions in latent space, compared to separability of DZ and MZ in observed data) was assessed through permutation testing (10000 matched permutations).

For demographic predictions, we used a linear support vector classifier (SVC) to predict sex, and kernel ridge regression with linear kernel to predict age and cognition[75]. For the Krakencoder, models were fit using cross-validation on the average latent vectors for the 683 subjects used to train the Krakencoder model, and tested on the 196 held-out subjects. For the observed connectome data, we computed the inter-subject cosine similarity matrix of each connectome flavor, and used the averaged similarity matrices as input into the model, using the same training and testing splits as for the Krakencoder. Regression hyperparameters were selected by nested cross-validation grid search. Variability of prediction accuracy was assessed through bootstrap resampling (N=100), and significant differences between Krakencoder and observed data models was assessed using permutation testing of matched bootstrap samples (10000 permutations).

Data and code availability

Code is available at https://github.com/kjamison/krakencoder. Post-processing code for FC can be found here: https://github.com/kjamison/fmriclean. Additional data files are available at https://osf.io/dfp92.

Preprocessed data for this study are available for download from the Human Connectome Project (www.humanconnectome.org). Users must agree to data use terms for the HCP before being allowed access to the data and ConnectomeDB, details are provided at https://www.humanconnectome.org/study/hcp-

Loss category		Equation	Description
Path loss	Reconstruction accuracy and inter-subject dispersion	$L_{r.corrI}(X_i', \hat{X}_i') = \lVert \mathrm{corr}(X_i', \hat{X}_i') - I_{n_{subj} \times n_{subj}} \rVert_2$	$[n_{subj} \times n_{subj}]$ correlation matrix of measured→predicted should be high along the diagonal and low off-diagonal
		$L_{r.mse}(X_i', \hat{X}_i') = \lVert X_i' - \hat{X}_i' \rVert_2^2$	MSE between measured→predicted should be low
		$L_{r.marg}(X_i', \hat{X}_i') = -\min_{a,b \neq a}(\lVert X_i'(s_a) - \hat{X}_i'(s_b) \rVert_2)$	Predicted $\hat{X}'$ for subject s_a should be far from nearest measured X' for $s_{b \neq a}$
	Latent space inter-subject dispersion	$L_{z.corrI}(z_i) = \lVert \mathrm{corr}(z_i, z_i) - I_{n_{subj} \times n_{subj}} \rVert_2$	$[n_{subj} \times n_{subj}]$ correlation matrix of latent space vectors (for the same path) should be low off-diagonal
		$L_{z.dist}(z_i) = -\sum_{a,b \neq a}^{n_{subj}} \lVert z_i(s_a) - z_i(s_b) \rVert_2$	Mean inter-subject latent space distances should be high
	Combined path loss	$L_r(path_{ij}) = L_{r.corrI}(X_j', \hat{X}_j') + L_{r.marg}(X_j', \hat{X}_j') + w_{rm} * L_{r.mse}(X_j', \hat{X}_j')$ $(w_{rm} = 1000)$ $L_z(path_{ij}) = L_{z.corrI}(z_i) + L_{z.dist}(z_i)$ $L_{path} = L_r(path_{ij}) + w_z * L_z(path_{ij})$ $(w_z = 10)$	
	Latent space intra-subject consistency	$L_{z.sim}(z) = \sum_{a=1}^{n_{subj}} \left(\sum_{i,j \neq i}^{n_{flav}} \lVert z_i(s_a) - z_j(s_a) \rVert_2^2 \right)$	Total MSE between all latent vectors from all flavors for each subject should be low

Table 1. Loss function summary

young-adult/data-use-terms. Data from the HCP Aging and HCP Development studies can be downloaded as part of the HCP-Lifespan 2.0 release, distributed by the NIMH Data Archive (https://nda.nih.gov). See https://www.humanconnectome.org/study/hcp-lifespan-aging/data-releases and https://www.humanconnectome.org/study/hcp-lifespan-development/data-releases for more information about data use terms.

Post-processed connectomes and related input files can be made available upon reasonable request from corresponding author K.J., subject to HCP data use restrictions described above.

References

[1] Demertzi, A. *et al.* Human consciousness is supported by dynamic complex patterns of brain signal coordination. *Sci. Adv.* **5**, eaat7603, DOI: 10.1126/sciadv.aat7603 (2019). Publisher: American Association for the Advancement of Science.

[2] Hellyer, P. J., Scott, G., Shanahan, M., Sharp, D. J. & Leech, R. Cognitive Flexibility through Metastable Neural Dynamics Is Disrupted by Damage to the Structural Connectome. *J. Neurosci.* **35**, 9050–9063, DOI: 10.1523/JNEUROSCI.4648-14.2015 (2015). Publisher: Society for Neuroscience Section: Articles.

[3] Olafson, E. R. *et al.* Functional connectome reorganization relates to post-stroke motor recovery and structural and functional disconnection. *NeuroImage* **245**, 118642, DOI: 10.1016/j.neuroimage.2021.118642 (2021).

[4] Tozlu, C., Card, S., Jamison, K., Gauthier, S. A. & Kuceyeski, A. Larger lesion volume in people with multiple sclerosis is associated with increased transition energies between brain states and decreased entropy of brain activity. *Netw. Neurosci.* **7**, 539–556, DOI: 10.1162/netn_a_00292 (2023).

[5] Honey, C. J. *et al.* Predicting human resting-state functional connectivity from structural connectivity. *Proc. Natl. Acad. Sci.* **106**, 2035–2040, DOI: 10.1073/pnas.0811168106 (2009). Publisher: Proceedings of the National Academy of Sciences.

[6] Ritter, P., Schirner, M., McIntosh, A. R. & Jirsa, V. K. The Virtual Brain Integrates Computational Modeling and Multimodal Neuroimaging. *Brain Connect.* **3**, 121–145, DOI: 10.1089/brain.2012.0120 (2013).

[7] Sanz Leon, P. *et al.* The Virtual Brain: a simulator of primate brain network dynamics. *Front. Neuroinformatics* **7**, DOI: 10.3389/fninf.2013.00010 (2013). Publisher: Frontiers.

[8] Luppi, A. I. *et al.* Dynamical models to evaluate structure–function relationships in network neuroscience. *Nat. Rev. Neurosci.* **23**, 767–768, DOI: 10.1038/s41583-022-00646-w (2022). Publisher: Nature Publishing Group.

[9] Kringelbach, M. L. & Deco, G. Brain States and Transitions: Insights from Computational Neuroscience. *Cell Reports* **32**, 108128, DOI: 10.1016/j.celrep.2020.108128 (2020).

[10] Martí-Juan, G. *et al.* Using The Virtual Brain to study the relationship between structural and functional connectivity in patients with multiple sclerosis: a multicenter study. *Cereb. Cortex* **33**, 7322–7334, DOI: 10.1093/cercor/bhad041 (2023).

[11] Avena-Koenigsberger, A., Misic, B. & Sporns, O. Communication dynamics in complex brain networks. *Nat. Rev. Neurosci.* **19**, 17–33, DOI: 10.1038/nrn.2017.149 (2018). Publisher: Nature Publishing Group.

[12] Seguin, C., van den Heuvel, M. P. & Zalesky, A. Navigation of brain networks. *Proc. Natl. Acad. Sci.* **115**, 6297, DOI: 10.1073/pnas.1801351115 (2018).

[13] Betzel, R. F., Faskowitz, J., Mišić, B., Sporns, O. & Seguin, C. Multi-policy models of interregional communication in the human connectome, DOI: 10.1101/2022.05.08.490752 (2022). Pages: 2022.05.08.490752 Section: New Results.

[14] Abdelnour, F., Voss, H. U. & Raj, A. Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *NeuroImage* **90**, 335–347, DOI: 10.1016/j.neuroimage.2013.12.039 (2014).

[15] Rosenthal, G. *et al.* Mapping higher-order relations between brain structure and function with embedded vector representations of connectomes. *Nat. Commun.* **9**, 2178, DOI: 10.1038/s41467-018-04614-w (2018). Number: 1 Publisher: Nature Publishing Group.

[16] Gu, Z., Jamison, K. W., Sabuncu, M. R. & Kuceyeski, A. Heritability and interindividual variability of regional structure-function coupling. *Nat. Commun.* **12**, 4894, DOI: 10.1038/s41467-021-25184-4 (2021).

[17] Ghosh, S., Raj, A. & Nagarajan, S. S. A joint subspace mapping between structural and functional brain connectomes. *NeuroImage* **272**, 119975, DOI: 10.1016/j.neuroimage.2023.119975 (2023).

[18] Fouladivanda, M., Iraji, A., Wu, L. & Calhoun, V. A spatially constrained independent component analysis jointly informed by structural and functional connectivity. *bioRxiv* (2023).

[19] Vázquez-Rodríguez, B. *et al.* Gradients of structure–function tethering across neocortex. *Proc. Natl. Acad. Sci.* **116**, 21219–21227, DOI: 10.1073/pnas.1903403116 (2019). Publisher: Proceedings of the National Academy of Sciences.

[20] Gu, S. *et al.* Controllability of structural brain networks. *Nat. Commun.* **6**, 8414, DOI: 10.1038/ncomms9414 (2015). Publisher: Nature Publishing Group.

[21] Cornblath, E. J. *et al.* Temporal sequences of brain activity at rest are constrained by white matter structure and modulated by cognitive demands. *Commun. Biol.* **3**, 1–12, DOI: 10.1038/s42003-020-0961-x (2020). Publisher: Nature Publishing Group.

[22] Parkes, L. *et al.* Network Controllability in Transmodal Cortex Predicts Positive Psychosis Spectrum Symptoms. *Biol. Psychiatry* **90**, 409–418, DOI: 10.1016/j.biopsych.2021.03.016 (2021).

[23] Singleton, S. P. *et al.* Receptor-informed network control theory links LSD and psilocybin to a flattening of the brain's control energy landscape. *Nat. Commun.* **13**, 5812, DOI: 10.1038/s41467-022-33578-1 (2022). Publisher: Nature Publishing Group.

[24] Sarwar, T., Tian, Y., Yeo, B. T. T., Ramamohanarao, K. & Zalesky, A. Structure-function coupling in the human connectome: A machine learning approach. *NeuroImage* **226**, 117609, DOI: 10.1016/j.neuroimage.2020.117609 (2021).

[25] Neudorf, J., Kress, S. & Borowsky, R. Structure can predict function in the human brain: a graph neural network deep learning model of functional connectivity and centrality based on structural connectivity. *Brain Struct. & Funct.* **227**, 331–343, DOI: 10.1007/s00429-021-02403-8 (2022).

[26] Levakov, G., Faskowitz, J., Avidan, G. & Sporns, O. Mapping individual differences across brain network structure to function and behavior with connectome embedding. *NeuroImage* **242**, 118469, DOI: 10.1016/j.neuroimage.2021.118469 (2021).

[27] Li, Y., Wei, Q., Adeli, E., Pohl, K. M. & Zhao, Q. Joint Graph Convolution for Analyzing Brain Structural and Functional Connectome. In Wang, L., Dou, Q., Fletcher, P. T., Speidel, S. & Li, S. (eds.) *Medical Image Computing and Computer Assisted Intervention – MICCAI 2022*, Lecture Notes in Computer Science, 231–240, DOI: 10.1007/978-3-031-16431-6_22 (Springer Nature Switzerland, Cham, 2022).

[28] Zhang, L., Wang, L. & Zhu, D. Predicting brain structural network using functional connectivity. *Med. Image Analysis* **79**, 102463, DOI: 10.1016/j.media.2022.102463 (2022).

[29] Liang, Q. *et al.* Transforming Connectomes to "Any" Parcellation via Graph Matching. In Manfredi, L. *et al.* (eds.) *Imaging Systems for GI Endoscopy, and Graphs in Biomedical Image Analysis*, Lecture Notes in Computer Science, 118–127, DOI: 10.1007/978-3-031-21083-9_12 (Springer Nature Switzerland, Cham, 2022).

[30] Dadashkarimi, J. *et al.* Cross Atlas Remapping via Optimal Transport (CAROT): Creating connectomes for different atlases when raw data is not available. *Med. Image Analysis* 102864, DOI: 10.1016/j.media.2023.102864 (2023).

[31] Zhu, Y., Cui, H., He, L., Sun, L. & Yang, C. Joint Embedding of Structural and Functional Brain Networks with Graph Neural Networks for Mental Illness Diagnosis. Tech. Rep. arXiv:2107.03220, arXiv (2022). DOI: 10.48550/arXiv.2107.03220. ArXiv:2107.03220 [physics, q-bio] type: article.

[32] Kuceyeski, A. *et al.* The application of a mathematical model linking structural and functional connectomes in severe brain injury. *NeuroImage: Clin.* **11**, 635–647, DOI: 10.1016/j.nicl.2016.04.006 (2016).

[33] Kuceyeski, A. F., Jamison, K. W., Owen, J. P., Raj, A. & Mukherjee, P. Longitudinal increases in structural connectome segregation and functional connectome integration are associated with better recovery after mild TBI. *Hum. Brain Mapp.* **40**, 4441–4456, DOI: 10.1002/hbm.24713 (2019). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.24713.

[34] Pang, J. C. *et al.* Geometric constraints on human brain function. *Nature* **618**, 566–574, DOI: 10.1038/s41586-023-06098-1 (2023).

[35] Murphy, K. & Fox, M. D. Towards a consensus regarding global signal regression for resting state functional connectivity MRI. *NeuroImage* **154**, 169–173, DOI: 10.1016/j.neuroimage.2016.11.052 (2017).

[36] Pervaiz, U., Vidaurre, D., Woolrich, M. W. & Smith, S. M. Optimising network modelling methods for fMRI. *NeuroImage* **211**, 116604, DOI: 10.1016/j.neuroimage.2020.116604 (2020).

[37] Sarwar, T., Ramamohanarao, K. & Zalesky, A. Mapping connectomes with diffusion MRI: deterministic or probabilistic tractography? *Magn. Reson. Medicine* **81**, 1368–1384, DOI: 10.1002/mrm.27471 (2019).

[38] Li, J. *et al.* Global signal regression strengthens association between resting-state functional connectivity and behavior. *NeuroImage* **196**, 126–141, DOI: 10.1016/j.neuroimage.2019.04.016 (2019).

[39] Messé, A. Parcellation influence on the connectivity-based structure–function relationship in the human brain. *Hum. Brain Mapp.* **41**, 1167–1180, DOI: 10.1002/hbm.24866 (2020). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/hbm.24866.

[40] Radford, A. *et al.* Learning Transferable Visual Models From Natural Language Supervision. In *Proceedings of the 38th International Conference on Machine Learning*, 8748–8763 (PMLR, 2021). ISSN: 2640-3498.

[41] Rubinov, M. & Sporns, O. Complex network measures of brain connectivity: Uses and interpretations. *NeuroImage* **52**, 1059–1069, DOI: 10.1016/j.neuroimage.2009.10.003 (2010).

[42] Yeo, B. T. T. *et al.* The organization of the human cerebral cortex estimated by intrinsic functional connectivity. *J. Neurophysiol.* **106**, 1125–1165, DOI: 10.1152/jn.00338.2011 (2011).

[43] Margulies, D. S. *et al.* Situating the default-mode network along a principal gradient of macroscale cortical organization. *Proc. Natl. Acad. Sci.* **113**, 12574–12579, DOI: 10.1073/pnas.1608282113 (2016).

[44] Alexander-Bloch, A. F. *et al.* On testing for spatial correspondence between maps of human brain structure and function. *NeuroImage* **178**, 540–551, DOI: 10.1016/j.neuroimage.2018.05.070 (2018).

[45] Somerville, L. H. *et al.* The Lifespan Human Connectome Project in Development: A large-scale study of brain connectivity development in 5–21 year olds. *NeuroImage* **183**, 456–468, DOI: 10.1016/j.neuroimage.2018.08.050 (2018).

[46] Bookheimer, S. Y. *et al.* The Lifespan Human Connectome Project in Aging: An overview. *NeuroImage* **185**, 335–348, DOI: 10.1016/j.neuroimage.2018.10.009 (2019).

[47] Harms, M. P. *et al.* Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *NeuroImage* **183**, 972–984, DOI: 10.1016/j.neuroimage.2018.09.060 (2018).

[48] Tozlu, C. *et al.* The sequence of regional structural disconnectivity due to multiple sclerosis lesions. *Brain Commun.* **5**, fcad332, DOI: 10.1093/braincomms/fcad332 (2023).

[49] Tozlu, C., Jamison, K., Gu, Z., Gauthier, S. A. & Kuceyeski, A. Estimated connectivity networks outperform observed connectivity networks when classifying people with multiple sclerosis into disability groups. *NeuroImage. Clin.* **32**, 102827, DOI: 10.1016/j.nicl.2021.102827 (2021).

[50] Schoonheim, M. M., Broeders, T. A. A. & Geurts, J. J. G. The network collapse in multiple sclerosis: An overview of novel concepts to address disease dynamics. *NeuroImage: Clin.* **35**, 103108, DOI: 10.1016/j.nicl.2022.103108 (2022).

[51] Pines, A. R. *et al.* Dissociable multi-scale patterns of development in personalized brain networks. *Nat. Commun.* **13**, 2647, DOI: 10.1038/s41467-022-30244-4 (2022).

[52] Filippi, M. *et al.* Age-related vulnerability of the human brain connectome. *Mol. Psychiatry* 1–9, DOI: 10.1038/s41380-023-02157-1 (2023). Publisher: Nature Publishing Group.

[53] Van Essen, D. C. *et al.* The WU-Minn Human Connectome Project: An overview. *NeuroImage* **80**, 62–79, DOI: 10.1016/j.neuroimage.2013.05.041 (2013).

[54] Glasser, M. F. *et al.* The minimal preprocessing pipelines for the Human Connectome Project. *NeuroImage* **80**, 105–124, DOI: 10.1016/j.neuroimage.2013.04.127 (2013).

[55] Salimi-Khorshidi, G. *et al.* Automatic denoising of functional MRI data: Combining independent component analysis and hierarchical fusion of classifiers. *NeuroImage* **90**, 449–468, DOI: 10.1016/j.neuroimage.2013.11.046 (2014).

[56] Behzadi, Y., Restom, K., Liau, J. & Liu, T. T. A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. *NeuroImage* **37**, 90–101, DOI: 10.1016/j.neuroimage.2007.04.042 (2007).

[57] Power, J. D. *et al.* Methods to detect, characterize, and remove motion artifact in resting state fMRI. *NeuroImage* **84**, 320–341, DOI: 10.1016/j.neuroimage.2013.08.048 (2014).

[58] Smith, S. M. *et al.* Advances in functional and structural MR image analysis and implementation as FSL. *NeuroImage* **23 Suppl 1**, S208–219, DOI: 10.1016/j.neuroimage.2004.07.051 (2004).

[59] Andersson, J. L. R. & Sotiropoulos, S. N. An integrated approach to correction for off-resonance effects and subject movement in diffusion MR imaging. *NeuroImage* **125**, 1063–1078, DOI: 10.1016/j.neuroimage.2015.10.019 (2016).

[60] Tournier, J.-D. *et al.* MRtrix3: A fast, flexible and open software framework for medical image processing and visualisation. *NeuroImage* **202**, 116137, DOI: 10.1016/j.neuroimage.2019.116137 (2019).

[61] Jeurissen, B., Tournier, J.-D., Dhollander, T., Connelly, A. & Sijbers, J. Multi-tissue constrained spherical deconvolution for improved analysis of multi-shell diffusion MRI data. *NeuroImage* **103**, 411–426, DOI: 10.1016/j.neuroimage.2014.07.061 (2014).

[62] Tournier, J.-D., Calamante, F. & Connelly, A. MRtrix: Diffusion tractography in crossing fiber regions. *Int. J. Imaging Syst. Technol.* **22**, 53–66, DOI: 10.1002/ima.22005 (2012). _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/ima.22005.

[63] Tournier, J. D., Calamante, F. & Connelly, A. Improved probabilistic streamlines tractography by 2nd order integration over fibre orientation distributions. In *Proceedings of the international society for magnetic resonance in medicine*, vol. 1670 (John Wiley & Sons, Inc. New Jersey, USA, 2010).

[64] Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. Anatomically-constrained tractography: improved diffusion MRI streamlines tractography through effective use of anatomical information. *NeuroImage* **62**, 1924–1938, DOI: 10.1016/j.neuroimage.2012.06.005 (2012).

[65] Smith, R. E., Tournier, J.-D., Calamante, F. & Connelly, A. SIFT2: Enabling dense quantitative assessment of brain white matter connectivity using streamlines tractography. *NeuroImage* **119**, 338–351, DOI: 10.1016/j.neuroimage.2015.06.092 (2015).

[66] Desikan, R. S. *et al.* An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. *NeuroImage* **31**, 968–980, DOI: 10.1016/j.neuroimage.2006.01.021 (2006).

[67] Fischl, B. *et al.* Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. *Neuron* **33**, 341–355, DOI: 10.1016/s0896-6273(02)00569-x (2002).

[68] Shen, X., Tokoglu, F., Papademetris, X. & Constable, R. T. Groupwise whole-brain parcellation from resting-state fMRI data for network node identification. *NeuroImage* **82**, 403–415, DOI: 10.1016/j.neuroimage.2013.05.081 (2013).

[69] Finn, E. S. *et al.* Functional connectome fingerprinting: identifying individuals using patterns of brain connectivity. *Nat. Neurosci.* **18**, 1664–1671, DOI: 10.1038/nn.4135 (2015).

[70] Glasser, M. F. *et al.* A multi-modal parcellation of human cerebral cortex. *Nature* **536**, 171–178, DOI: 10.1038/nature18933 (2016). Number: 7615 Publisher: Nature Publishing Group.

[71] Patenaude, B., Smith, S. M., Kennedy, D. N. & Jenkinson, M. A Bayesian model of shape and appearance for subcortical brain segmentation. *NeuroImage* **56**, 907–922, DOI: 10.1016/j.neuroimage.2011.02.046 (2011).

[72] Iglesias, J. E. *et al.* A probabilistic atlas of the human thalamic nuclei combining ex vivo MRI and histology. *NeuroImage* **183**, 314–326, DOI: 10.1016/j.neuroimage.2018.08.012 (2018).

[73] Diedrichsen, J., Balsters, J. H., Flavell, J., Cussans, E. & Ramnani, N. A probabilistic MR atlas of the human cerebellum. *NeuroImage* **46**, 39–46, DOI: 10.1016/j.neuroimage.2009.01.045 (2009).

[74] Rolls, E. T., Huang, C.-C., Lin, C.-P., Feng, J. & Joliot, M. Automated anatomical labelling atlas 3. *NeuroImage* **206**, 116189, DOI: 10.1016/j.neuroimage.2019.116189 (2020).

[75] Ooi, L. Q. R. *et al.* Comparison of individualized behavioral predictions across anatomical, diffusion and functional connectivity MRI. *NeuroImage* **263**, 119636, DOI: 10.1016/j.neuroimage.2022.119636 (2022).

Acknowledgements

This work was supported by the following grants: NIH RF1 MH123232 (AK), Ann S. Bowers Foundation (AK), NIH R01 AG053949 (MS), NSF CAREER 1748377 (MS).

Data were provided by the Human Connectome Project, WU-Minn Consortium (Principal Investigators: David Van Essen and Kamil Ugurbil; 1U54MH091657) funded by the 16 NIH Institutes and Centers that support the NIH Blueprint for Neuroscience Research; and by the McDonnell Center for Systems Neuroscience at Washington University.

Research reported in this publication was supported by the National Institute Of Mental Health of the National Institutes of Health under Award Number U01MH109589, National Institute On Aging of the National Institutes of Health under Award Number U01AG052564. The HCP-Development 2.0 Release data used in this report came from DOI: 10.15154/1520708. The HCP-Aging 2.0 Release data used in this report came from DOI: 10.15154/1520707.

The authors would like to thank Fang-Chang Yeh for sharing pre-processed diffusion MRI data for HCP-Development and HCP-Aging subjects.

Author contributions statement

K.W.J. conceived the study and developed the software with help from Z.G.. K.W.J., A.K., and M.R.S. analyzed results, and Q.W. and C.T. helped with validation. K.W.J., A.K., and M.R.S. wrote and edited the manuscript. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.