General Functional Connectivity: shared features of resting-state and task fMRI drive reliable individual differences in functional brain networks

Maxwell L. Elliott[1*]
Annchen R. Knodt[1]
M. Justin Kim[1]
Tracy R. Melzer[2,3]
Ross Keenan[2,4]
David Ireland[5]
Sandhya Ramrakha[5]
Richie Poulton[5]
Avshalom Caspi[1,6-8]
Terrie E. Moffitt[1,6-8]
Ahmad R. Hariri[1]

1. *Department of Psychology & Neuroscience, Duke University Box 104410, Durham, NC 27708, USA*
2. *New Zealand Brain Research Institute, Christchurch, New Zealand*
3. *Department of Medicine, University of Otago, Christchurch, New Zealand*
4. *Christchurch Radiology Group, Christchurch, New Zealand*
5. *Dunedin Multidisciplinary Health and Development Research Unit, Department of Psychology, University of Otago, 163 Union St E, Dunedin, 9016, NZ*
6. *Social, Genetic, & Developmental Psychiatry Research Centre, Institute of Psychiatry, Psychology, & Neuroscience, King's College London, De Crespigny Park, Denmark Hill, London SE5 8AF, UK*
7. *Department of Psychiatry & Behavioral Sciences, Duke University School of Medicine, Durham, NC 27708, USA*
8. *Center for Genomic and Computational Biology, Duke University Box 90338, Durham, NC 27708, USA*

*Correspondence: maxwell.elliott@duke.edu

Summary

Intrinsic connectivity, commonly measured using resting-state fMRI, has emerged as a fundamental feature of the brain. However, largely due to practical limitations, many studies do not collect enough resting-state data to generate reliable measures of intrinsic connectivity necessary for studying individual differences. Here we present general functional connectivity (GFC) as a method for leveraging shared features across resting-state and task fMRI and demonstrate that GFC is substantially more reliable than intrinsic connectivity estimated from typically available resting-state data alone (i.e., 5-10 minutes). We further demonstrate that the increase in reliability gained through GFC boosts predictive utility by accounting for upwards of 200% more variance in the domain of intelligence. Our work suggests that GFC represents an opportunity to improve the reliability and predictive utility of fMRI for mapping individual differences, and that collection of resting-state and task scans need not be a zero-sum competition when designing future studies.

Introduction

Functional magnetic resonance imaging (fMRI) has proven invaluable for identifying the neural architecture of human behavior and cognition (Betti et al., 2013; Fox et al., 2007; Huth et al., 2016). Accordingly, fMRI has been widely adopted in myriad studies seeking to deepen our understanding of the human brain in both health and disease (Cole et al., 2011; Power et al., 2013; Satterthwaite et al., 2016). fMRI studies have recently expanded in both scale and scope, often collecting functional neuroimaging data in thousands of individuals in an effort to adequately power the search for neural correlates of complex human traits and predictive biomarkers for mental illness (Casey et al., 2018; Elliott et al., 2018; Miller et al., 2016; Swartz et al., 2015; Thompson et al., 2014). In this context, most studies have focused on the acquisition of resting-state functional MRI to map the intrinsic connectivity of neural networks. This choice is often prompted by three considerations. First, intrinsic connectivity networks appear to be more heritable than task-elicited activity (Elliott et al., 2017; Ge et al., 2017; Glahn et al., 2010). Second, resting-state data are relatively easy to collect from informative populations of children, the elderly, and the mentally ill (Fox, 2010; Greicius, 2008; Shehzad et al., 2009). Third, intrinsic connectivity plays a primary role in shaping task-based brain activity and associated behaviors (Cole et al., 2014, 2016; Krienen et al., 2014; Tavor et al., 2016).

While analyses of resting-state data have revealed insights about average, group level features of the human brain (Buckner et al., 2008; Power et al., 2011; Yeo et al., 2011), progress has lagged in identifying individualized signatures, which are critical if ongoing large-scale studies are to be successful in the search for risk and disease biomarkers. This lack of progress partially reflects

the generally poor test-retest reliability of intrinsic connectivity estimates in many resting-state studies (Birn et al., 2013; Cannon et al., 2018). The reliability of intrinsic connectivity measures must be improved, as unreliability limits the ability to predict individual differences (Nunnally Jr., 1970; Vul et al., 2009). In this paper we propose general functional connectivity (GFC), a simple method to combine resting-state and task fMRI to improve the reliability of individual differences in intrinsic connectivity.

Historically researchers have collected 5-10 minutes of resting-state data in response to both published recommendations (Van Dijk et al., 2010; Shehzad et al., 2009) and practical limitations (e.g., scan time, cost and subject stamina). While such data may be adequate for detecting patterns of intrinsic connectivity common across individuals, more than 25 minutes of data is needed to reliably detect individual differences in these same patterns (Anderson et al., 2011; Hacker et al., 2013; Laumann et al., 2015). This suggests that intrinsic connectivity may simply be a noisy measure of stable individual differences whose 'true' signal requires lengthier measurement (Braga and Buckner, 2017; Gordon et al., 2017; Gratton et al., 2018; Mueller et al., 2013). However, due to the high cost of scan time and the limited ability of many individuals including children, the elderly, and the mentally ill to lie still without falling asleep for tens of minutes of scanning, this level of in-depth measurement is simply not feasible for a majority of studies (Power et al., 2012; Satterthwaite et al., 2013; Tagliazucchi and Laufs, 2014). To illustrate the point, two recent meta-analyses of resting-state intrinsic connectivity in depression (Kaiser et al., 2015) and schizophrenia (Dong et al., 2018) revealed average resting-state scan times of 6.53 minutes (k = 23 studies) and 6.24 minutes (k = 36 studies), respectively. It is likely that

intrinsic connectivity based on these brief scans would have low reliability, reducing the ability to discover reliable individual differences in disease biomarkers. If measures of intrinsic connectivity are to realize their potential as a clinical tool (Fox, 2010; Matthews et al., 2006) and contribute to personalized medicine (Collins and Varmus, 2015; Hamburg and Collins, 2010), there must first be substantial improvement in their reliability, and this should be ideally achieved within the practical limitations of study design and participant burden.

Although few studies collect enough resting-state data to achieve high levels of reliability, many collect multiple task scans in addition to a resting-state scan. While it has been implicitly assumed that task and resting-state scans are two separate measures of brain function, to be analyzed and studied independently, a growing body of evidence suggests that the intrinsic connectivity measured by each may have substantial overlap, sharing over 80% of the same variance (Cole et al., 2014; Geerligs et al., 2015; Tavor et al., 2016). Further, intrinsic connectivity from task scans may even accentuate individual differences (Finn et al., 2017; Satterthwaite et al., 2018), and resting-state and task data can be complimentary in deriving reliable components of individual differences in intrinsic connectivity (Gratton et al., 2018).

If task and resting-state data converge in their measurements of individual differences in intrinsic connectivity, and length of data acquisition is the primary limiting factor in establishing a reliable measure, then combining these scans could achieve higher reliability. Here we hypothesized that measures of intrinsic connectivity will be more reliable when task and resting-state scans are combined in comparison with short resting-state scans alone. Additionally, we hypothesized that

this boost in reliability will increase the predictive utility of intrinsic connectivity in mapping the neural correlates of individual differences in complex human traits. We leveraged 3 existing datasets to test our hypotheses. First, to determine how much data are needed for a reliable estimate of intrinsic connectivity we investigated the influence of a broad range of scan lengths on reliability using data from a single individual who underwent over 20 hours of fMRI scanning (Poldrack et al., 2015). Second, we investigated the test-retest reliability of different combinations of task and resting-state data in both the Human Connectome Project (HCP) (Van Essen et al., 2013), which uses state-of-the-art fMRI hardware and acquisition parameters in a highly educated, healthy sample, and in the Dunedin Longitudinal Study (Poulton et al., 2015), which uses more common fMRI hardware and acquisition parameters in a population representative birth cohort. Finally, we determined the relative gain in explaining individual differences in the complex human trait of intelligence resulting from the increased reliability of intrinsic connectivity estimates in both the HCP and Dunedin Study.

Results

How much data are needed for converging estimates of intrinsic connectivity?

An influential recent study suggested that approximately 25 minutes of resting-state data are needed to adequately measure an individual's intrinsic connectivity (Laumann et al., 2015). We began our investigation by replicating these findings using a general preprocessing framework and then extending these analyses by adopting a broader definition of intrinsic connectivity by combining resting-state and task scans. While the original study employed procedures optimized for single-subject analyses, we were able to generally replicate the relationship between amount

of data and reliability using a typical group processing and parcellation scheme (Figure 1; red curve). Split-half reliability was calculated using 125 minutes of data from one half of the data pool to define the "true" connectivity matrix, correlated with increasing amounts of data from the other half. 125 minutes was used in order to have comparable amounts of data between rest and task analyses (see supplemental figure S1 for direct replication using 400 minutes). Specifically, we found that as the amount of resting-state data used to create the intrinsic connectivity maps increases, the split half convergence between the "true" connectivity and the estimated connectivity sharply increases, from r = .64 with half a session totaling 4.97 minutes to r = .87 with 3 sessions of data totaling 29.81 minutes. This curve begins to level off with around 50 minutes of data (r = .91), and by 80 minutes of data has mostly converged (r = .93), peaking at r = .94 with 124.22 minutes of data (see supplemental for graph of entire range).

We next investigated reliability using a broadened definition of intrinsic connectivity, which we term General Functional Connectivity (GFC), defined as an equal amount of data from resting-state and task scans. Here, the "true" connectivity matrix was defined from 124.93 minutes of data, which represents the total time of task scans. As seen in the blue curve in Figure 1, GFC showed a similar pattern of convergence with the "true" intrinsic connectivity measures as observed above using only resting-state data. The curve rises sharply from r = .62 with 6.31 minutes of data to r = .83 with 31.27 minutes of data. This curve begins to level off with 49.51 minutes of data (r = .87) and continues to slowly rise, peaking when the entire subset of task data is used (124.93 minutes of data) at r = .93.

What is the test-retest reliability of GFC?

As demonstrated above, the reliability of intrinsic connectivity is substantially influenced by the amount of data used to derive the estimates. Next, we used data from the HCP to evaluate the test-retest reliability of both intrinsic connectivity derived from resting-state data alone and of GFC derived from combinations of task and resting-state data in the HCP. Test-retest reliability, as measured by Intra-Class Correlation (ICC), ranges from 0-1, and is commonly classified according to the following cutoffs: 0 - .4 = poor, .4 - .6 = moderate, .6 - .8 = good and .8 - 1 = excellent (Cicchetti, 1994). In the HCP, when 5 minutes (post-censoring) of resting-state data were used to measure individual differences in intrinsic connectivity defined within a common atlas of 264 regions (Power et al., 2011), the reliability was generally poor, (mean ICC = .30; 67% of edges were poor , 27% moderate, 5% good, and less than 1% excellent). As more resting-state data were added, the test-retest reliability continued to increase up to the limits of the dataset at 40 minutes (mean ICC = .63; 7% of edges poor, 31% moderate, 53% good, and 9% excellent) (see Figure 3).

Critically, this general pattern held when task data was incrementally added to resting-state data to generate estimates of GFC in the HCP. The mean test-retest reliability of 5 minutes of resting-state data showed a 73% increase when 25 minutes of task was added (mean ICC = .52; 23% of edges were poor, 39% moderate, 33% good, and 4% excellent). Pure GFC, defined as exactly equal parts of each task and resting-state scan, followed this pattern as well: 5 minutes of data exhibited poor reliability (mean ICC = .33, 10% of edges were poor, 30% moderate, 8% good, and

less than 1% excellent) but 40 minutes of data showed good reliability (mean ICC = .61, 10% of edges were poor, 32% moderate, 49% good, and 9% excellent) (see figure 3).

This general improvement in reliability with increasing data was not unique to the HCP, but also observed in the population representative Dunedin Study. The base protocol Dunedin resting-state scan of 8:16 minutes showed generally low reliability (mean ICC = .27, 68% of edges were poor, 23% moderate, 9% good, and 1% excellent). The reliability of the intrinsic connectivity estimates, however, improved by 104% when the 8:16 minutes of resting-state data was combined with 27 minutes (before censoring) of task data (mean ICC = .55, 20% of edges were poor, 35% moderate, 38% good and 6% excellent).

Does better reliability of GFC increase predictive utility of individual differences?

Connectome Predictive Mapping (CPM) was used to predict intelligence from both resting-state only-derived intrinsic connectivity and combined resting-state and task GFC in both the HCP and the Dunedin study. CPM predicts a phenotype from a linear combination of edges that are related ($p<.01$) to your phenotype of interest. Here, intelligence was used as the variable of interest because it is a reliability individual difference measure that has a history in prediction studies using CPM (Dubois et al., 2018a, 2018b; Finn et al., 2015; Noble et al., 2017; Weschler, 2008). In the HCP, we restricted analyses to the first 8 minutes of resting-state data to approximate the data available in the Dunedin Study as well as many studies with limited scan time. When doing so, we found that individual differences in intelligence were significantly predicted from 8 minutes of resting-state data ($r = .12$, $p = .032$). However, the predictive utility increased when

25 minutes of task data was added to estimate GFC ($r = .33$, $p < .001$) (see figure 4). Similar advantages in predictive utility were observed in the Dunedin Study when using GFC. Notably, in the Dunedin Study intrinsic connectivity estimates derived from 8:16 minutes of resting-state data alone predicted intelligence ($r = .12$, $p = .015$). When 27 minutes (before censoring) of task data was added to estimate GFC, the predictive utility increased ($r = .22$, $p < .001$). In the HCP, the addition of task data resulted in a 679% increase in variance explained ($r^2 = .014$ to $r^2 = .109$), while in the Dunedin Study the addition of task data resulted in a 243% increase in variance explained ($r^2 = .014$ to $r^2 = .048$).

Discussion

Here we present a novel method for improving the reliability of individual differences in intrinsic connectivity by combining resting-state and task data to generate an estimate of intrinsic connectivity that we call general functional connectivity (GFC). Across multiple datasets we found that the addition of task data to short, but representative amounts of resting-state data substantially increases the reliability of intrinsic connectivity measures. Moreover, this increase in reliability manifests as significant gains in predictive utility of individual differences in complex human traits, namely intelligence. These findings have several consequences for current and future neuroimaging research because they demonstrate that (1) reliable measurement of individual differences in the functional architecture of the brain are not only achievable in niche, specialty datasets with hours of resting-state data, but also in many existing datasets if GFC is adopted; (2) this gain in reliability has consequences for the study of individual differences in behavior and cognition, as demonstrated in the increase in predictive ability of intelligence; (3)

to the extent that replicable gains in reliability may be achieved in other datasets, our results provide a reference that researchers can use to roughly estimate the gain in reliability they may achieve by adopting GFC (figure 2 and 3). We elaborate on each of these points, next.

Data aggregation across task and resting-state fMRI boosts reliability

We found that GFC, derived from the combination of resting-state and task fMRI is a reliable measure of individual differences in intrinsic connectivity. Consistent with previous studies (Birn et al., 2013; Finn et al., 2015; Laumann et al., 2015), we showed that the reliability of intrinsic connectivity is highly dependent on the amount of data collected. This is true of resting-state functional connectivity as well as GFC. Across a broad range of scan lengths (figures 1 and 3), GFC is comparable to resting-state functional connectivity in its reliability. While the split half-reliability of resting-state connectivity (r= .64) and GFC (r= .62) is moderate with only 5 minutes of data, this relatively high split-half correspondence with the "true" intrinsic connectivity of an individual's brain does not lead to a reliable measure of individual differences. Instead, in 2 independent samples, 5-10 minutes of data produced poor reliability in both the HCP and Dunedin Study data sets. Only with 30-40 minutes of data do resting-state connectivity and GFC begin to display good reliability. This finding follows directly from past studies that have shown reliability depends on scan length (Anderson et al., 2011; Birn et al., 2013; Hacker et al., 2013), that task and resting-state data share a large proportion of variance (Cole et al., 2014; Geerligs et al., 2015), and that both task and resting-state data measure common individual differences in intrinsic connectivity (Gratton et al., 2018). However, to our knowledge, this is the first

demonstration of the tangible boost in reliability that can be achieved by combining task and resting-state data to generate GFC.

Importantly, most neuroimaging datasets only have 5-10 minutes of resting-state data, which are limited by sampling variability, motion-artifacts, and other sources of noise (Gordon et al., 2017; Gratton et al., 2018; Power et al., 2014). Given the reliability estimates reported here, it is likely that in many datasets, individual differences in intrinsic connectivity will be unmeasurable with rest-state alone (Anderson et al., 2011; Hacker et al., 2013). However, it is common practice for researchers to collect 10-40 minutes of task fMRI data in addition to resting-state. Therefore, if researchers adopt a broader definition of intrinsic connectivity, such as GFC, they can combine task and resting-state data to achieve a reliable measure of individual differences in intrinsic connectivity.

General functional connectivity has better predictive utility

Previous research has shown that the predictive utility of intrinsic connectivity data depends on the amount of data available (Dubois et al., 2018a; Finn et al., 2015). This follows directly from the fact that the observed correlation between any two measures is limited by the reliability of each measure (Nunnally Jr., 1970; Vul et al., 2009). Directly in line with these findings, we show that when task data is added to short (5 - 10 minutes) resting-state data the predictive ability increases substantially (679% in HCP and 243% in the Dunedin Study). While this is a single observation, this result provides a proof-of-principal that increasing the reliability of intrinsic connectivity estimates through adopting a GFC framework can have benefits for predictive utility

in mapping individual differences. However, future research is needed to extend these findings to other phenotypes. While neural correlates of behavior in health and disease have reported with 5-10 minutes of resting-state data, these findings have failed to power a cumulative literature of replicable effects (Button et al., 2013; Szucs and Ioannidis, 2017). This is likely due in part to the low reliability (ICCs $\leq$ .4) demonstrated here. However, with GFC, many functional connections display good to excellent reliability (ICCs $\geq$ .6). And, with this higher reliability, specific functional connections can be expected to have higher replicability and greater promise as biomarkers in personalized medicine.

How general is general functional connectivity?

An understandable reluctance in implementing GFC could be that it would introduce heterogeneity in the estimation of intrinsic connectivity. Whereas resting-state intrinsic connectivity is thought to be a common framework with generalizability across datasets, the combination of different sets of tasks in different datasets may introduce additional variability. Below, we provide 3 reasons why we think this is less of a problem than it may appear and point to features of our analyses that support the validity of GFC as a robust measure of individual differences in intrinsic connectivity.

First, resting-state data is by nature heterogenous, and the resting-state is its own type of task (Buckner et al., 2013). Researchers collect resting-state data under different conditions in which participants close their eyes, have their eyes open, or fixate on a screen. These differences in resting-state conditions influence intrinsic connectivity (Patriat et al., 2013). Despite these

differences, however, all are collectively referred to as resting-state scans. In addition, factors like thought content (Christoff et al., 2009; Hurlburt et al., 2015), caffeine intake (Wong et al., 2012), recent cognitive tasks (Waites et al., 2005), and falling asleep (Deco et al., 2014) can introduce further heterogeneity biasing intrinsic connectivity estimates in many resting-state datasets. It has even been estimated that approximately 30% of subjects fall asleep within the first 3 minutes of a resting-state scan (Tagliazucchi and Laufs, 2014). Despite this heterogeneity, resting-state fMRI measures a common set of functional networks (Yeo et al., 2011), and displays trait-like features driven by stable factors such as genetics (Glahn et al., 2010) and structural connectivity (Honey et al., 2009). For these reasons resting-state data represents substantial promise as an individual differences measure, but only if enough data are collected to average out sampling variability (Gratton et al., 2018). Connectivity estimates from task data are also shaped by many of these stable factors (Cole et al., 2014; Krienen et al., 2014), and task data predominantly measure overlapping individual differences that are only weakly influenced by task demands (Gratton et al., 2018).

Second, we provide converging evidence for benefits of GFC across 3 very different samples. The data from the 3 samples were collected on 3 different scanners, with a broad range of scanning parameters. The samples ranged from the MyConnectome, which consists of a single Stanford professor who is likely the most brain-imaged individual in history (Poldrack et al., 2015), to a large healthy sample of highly educated individuals in the HCP (Van Essen et al., 2013), to a population-representative birth cohort with a broad range of mental and physical health conditions, a broad range of socioeconomic status and full representation of variability in many

complex traits (Poulton et al., 2006, 2015). Across samples, the resting-state data came from eyes closed and eyes opened conditions and the task data came from 16 different tasks unevenly distributed across datasets (5 in MyConnectome, 7 in HCP and 4 in Dunedin Study). The data were processed with different preprocessing schemes and software ("minimally preprocessed" in HCP and custom scripts in MyConnectome and Dunedin Study). In addition, amount of scan time was tightly controlled in MyConnectome and HCP (equal time after censoring), whereas in the Dunedin dataset motion censoring led to unequal scan lengths across participants as is the case in traditional individual differences analyses. Despite these many differences between samples, our results show striking convergence in the reliability gained through the GFC framework, with average ICCs at .52 in HCP and .55 in Dunedin with approximately 25 minutes of data. Altogether, these results demonstrate that GFC has good test-retest reliability, inter-rater reliability (different preprocessing schemes), out-of-sample reliability (multiple datasets) and parallel forms reliability (different tasks in each sample) (Dubois and Adolphs, 2016). While we present evidence in favor of the generalizability of GFC, future research is still needed to formally investigate the heterogeneity in GFC and find optimal combinations of tasks that most efficiently measure individual differences in intrinsic connectivity (Finn et al., 2017; Shah et al., 2016).

Third, with the 5-10 minutes of resting-state data available to many researchers, investigations of individual differences in intrinsic connectivity will be tenuous. Reliability fundamentally limits the ability to detect associations between any two measures (Nunnally Jr., 1970; Vul et al., 2009). Therefore, any investigation mapping intrinsic connectivity to individual differences in behavior,

cognition, or disease is fundamentally limited by the reliability of intrinsic connectivity measures. Power to detect true effects depends on the anticipated effect size, which is in turn dependent on reliability of each measurement (Kanyongo et al., 2007; Williams and Zimmerman, 1989). Given that many datasets only have 5-10 minutes of resting-state data and consequently poor reliability of intrinsic connectivity measures, without GFC many studies will have drastically reduced statistical power (Button et al., 2013; Ioannidis, 2005, 2008), and be limited to finding tenuous correlates of individual differences. In this way, GFC provides one practical step for imaging research to confront the power and reliability issues identified in neuroscience (Button et al., 2013; Szucs and Ioannidis, 2017) and beyond (Aarts et al., 2015; Errington et al., 2014; Ioannidis, 2005; Ioannidis et al., 2001; Munafò et al., 2017) and move towards a cumulative science of individual differences (Dubois and Adolphs, 2016). For these reasons, we think that the gains in reliability offered by GFC may outweigh the potential task-induced "heterogeneity."

Conclusion

Here we propose general functional connectivity (GFC) as a method for improving the reliability of estimating individual differences in the intrinsic architecture of the brain's functional networks. We demonstrate that, when the amount of data available for analysis is held constant, the combination of resting-state and task scans is nearly as reliable as resting-state alone. Furthermore, adding task data to typically short runs of resting-state data is substantially more reliable than the resting-state data alone. This increase in reliability subsequently boost the predictive utility of intrinsic connectivity in mapping individual differences. Many researchers who have less 25 minutes of resting-state data but have task fMRI as well, can immediately boost

reliability and consequently power, in studies that use intrinsic connectivity. Our findings also suggest that future data collection should consider naturalistic fMRI (Hasson et al., 2004, 2010; Huth et al., 2016; Lahnakoski et al., 2012; Vanderwal et al., 2017) and engaging tasks in order to limit motion (Power et al., 2012; Yuan et al., 2009), in addition to resting-state scans when measuring intrinsic connectivity. This is especially true when studying patient populations, children, and the elderly as they often cannot lie still for 25 minutes of resting-state (Satterthwaite et al., 2013). Researchers may need to think of collecting resting-state and task data as a zero-sum competition for scan time when designing studies. Instead, we show that task and resting-state data provide complementary measures of intrinsic connectivity that together can lead to reliable measurement of general functional connectivity. Collecting adequate amounts (> 25 minutes) of high-quality fMRI data should be emphasized over the traditional task-rest dichotomy. Reliability and measurement must continue to be at the forefront of fMRI studies if the method is to continue to mature into a cumulative neuroscience of individual differences with clinical relevance.

Methods

Datasets

MyConnectome. This dataset consists of extensive brain imaging and behavioral data from a single 45-year-old male (Poldrack et al., 2015). This study was designed to push the limits of precision mapping of a single individual brain and has been described extensively elsewhere (Laumann et al., 2015; Poldrack et al., 2015). Briefly, several high resolution T1-weighted structural images, 104 sessions of resting-state fMRI (10:00 minute scan length), and 51 sessions of task fMRI were collected over the course of 532 days. We used 88 sessions of resting-state fMRI in our analyses. Task data involving language working memory, spatial working memory, n-back, motion/stop-signal, and object localizer were used because each had at least 8 sessions (after combining spatial and language working memory).

Human Connectome Project. This is a publicly available data set that includes 1206 participants with extensive MRI imaging and behavioral measurement. Participants were all free of current psychiatric or neurologic illness and between 25 and 35 years of age. Much of the dataset consists of twins and family members. To avoid bias due to family confounding, only one individual per family was retained (family member with the least motion during scanning). Additionally, participants were excluded from prediction analyses if they had less than 40 minutes of total task data, less than 310 timepoints (3:53 seconds) of any of the fMRI tasks or less than 50 minutes of resting-state data after motion censoring. This resulted in 300 subjects left to be used in our prediction analyses. Additionally, 45 subjects completed the entire scanning protocol again at a second timepoint as part of the HCP test-retest dataset. These data were

used in the reliability portion of our analyses (see below). 8 subjects were removed from these analyses because they had at least one fMRI scan with truncated or missing data. 6 subjects were removed because they did not have sufficient data (> 40 minutes of rest and > 40 minutes of task) after censoring. Therefore, the test-retest reliability analyses included data from 31 participants.

The acquisition parameters and minimal preprocessing of these data have been described extensively elsewhere (Glasser et al., 2013). In our analysis we used the minimally preprocessed data in volumetric MNI space ("fMRIVolume" pipeline). Briefly, participants underwent extensive MRI measurement that included T1 and T2 weighted structural imaging, diffusion tensor imaging and nearly 2 hours of resting-state and task fMRI. One hour of resting-state fMRI was collected on each participant in four 15-minute scans (1200 time-points each) split-up into two scanning sessions over two days. In each scan session the two resting-state scans were followed by task fMRI (Smith et al., 2013). Across the two sessions each participant completed 7 fMRI tasks. Task fMRI is described extensively elsewhere (Barch et al., 2013). Briefly, tasks were designed to identify functionally relevant "nodes" in the brain and included working memory (810 timepoints, 10:02 minutes), gambling (506 timepoints, 6:24 minutes), motor function (568 timepoints, 7:06 minutes), language (632 timepoints, 7:54 minutes), social cognition (548 timepoints, 6:54 minutes), relational processing (464 timepoints, 6:52 minutes) and emotional processing (352 timepoints, 4:32 minutes). Altogether, 4800 timepoints totaling 60 minutes of resting-state fMRI and 3880 timepoints totaling 48:30 minutes of task fMRI were collected on each participant.

Dunedin Longitudinal Study. This is a longitudinal investigation of health and behavior in a complete birth cohort. Study members (N=1,037; 91% of eligible births; 52% male) were all born between April 1972 and March 1973 in Dunedin, New Zealand (NZ) and eligible based on residence in the province and participation in the first assessment at age 3. The cohort represents the full range of socioeconomic status on NZ's South Island and matches the NZ National Health and Nutrition Survey on key health indicators (e.g., BMI, smoking, GP visits) (Poulton et al., 2015). The cohort is primarily white; fewer than 7% self-identify as having non-Caucasian ancestry, matching the South Island (Poulton et al., 2015). Assessments were carried out at birth and ages 3, 5, 7, 9, 11, 13, 15, 18, 21, 26, 32, and 38 years, when 95% of the 1,007 study members still alive took part. Neuroimaging data collection is ongoing in participants who are now aged 45 years. We currently have collected completed neuroimaging data for 599 subjects. Subjects were excluded these analyses if they did not complete the rest scan and all 4 task scans or if they did not have sufficient data after motion censoring was applied (see below for censoring details) or if they had. Data for 542 subjects survived these fMRI exclusion criteria. Of these 542 subjects 413 had completed IQ testing at age 45 at the time of data analysis and were included in the final prediction analyses. Additionally, 20 subjects completed the entire scan protocol a second time. These subjects were used to define test-retest reliability in our analysis of the Dunedin study.

Each participant was scanned using a Siemens Skyra 3T scanner equipped with a 64-channel head/neck coil at the Pacific Radiology imaging center in Dunedin, New Zealand. High resolution structural images were obtained using a T1-weighted MP-RAGE sequence with the following

parameters: TR = 2400 ms; TE = 1.98 ms; 208 sagittal slices; flip angle, 9°; FOV, 224 mm; matrix =256×256; slice thickness = 0.9 mm with no gap (voxel size 0.9×0.875×0.875 mm); and total scan time = 6 minutes and 52 seconds. Functional MRI was collected during resting-state and 4 tasks. Functional MRI was collected with a series of 72 interleaved axial T2-weighted functional slices were acquired using a 3-fold multi-band accelerated echo planar imaging sequence with the following parameters: TR = 2000 ms, TE = 27 msec, flip angle = 90°, field-of-view = 200 mm, voxel size = 2mm isotropic, slice thickness = 2 mm without gap.

Eight minutes and 16 seconds (248 TRs) of resting-state fMRI was collected immediately before the four task fMRI scans. During the resting-state scan participants were instructed to stay awake with their eyes open and while a gray cross presented against a black background. Participants completed an emotion processing task (6:40 minutes, 200 TRs), a color Stroop task (6:58 minutes, 209 TRs), the monetary incentive delay task (7:44 minutes, 232 TRs) and an episodic memory task (5:44 minutes, 172 TRs). Tasks are described in detail in the supplemental.

fMRI pre-processing

Minimal preprocessing was first applied to all data. For the HCP dataset this was done with the minimal preprocessing pipeline (Glasser et al., 2013) and included correction for B0 distortion, realignment to correct for motion, registration to the participant's structural scan, normalization to the 4D mean, brain masking, and non-linear warping to MNI space.

Similar "minimal preprocessing" steps were applied to both the MyConnectome and Dunedin datasets using custom processing scripts. Anatomical images were skull-stripped, intensity-normalized, and nonlinearly warped to a study-specific average template in the standard stereotactic space of the Montreal Neurological Institute template using the ANTs SyN registration algorithm (Avants et al., 2008; Klein et al., 2009). Time-series images were despiked, slice-time-corrected, realigned to the first volume in the time-series to correct for head motion using AFNI tools (Cox, 1996), coregistered to the anatomical image using FSL's Boundary Based Registration (Greve and Fischl, 2009), spatially normalized into MNI space using the non-linear ANTs SyN warp from the anatomical image, and resampled to 2mm isotropic voxels. Dunedin images were additionally corrected for B0 distortions.

Time-series images for each dataset were further processed to limit the influence of motion and other artifacts. Voxel-wise signal intensities were scaled to yield a time-series mean of 100 for each voxel. Motion regressors were created using each participant's 6 motion correction parameters (3 rotation and 3 translation) and their first derivatives (Jo et al., 2013; Satterthwaite et al., 2013) yielding 12 motion regressors. Five components from white matter and cerebrospinal fluid were extracted using CompCorr (Behzadi et al., 2007) and used as nuisance regressors. In the MyConnectome and Dunedin datasets, images were bandpass filtered to retain frequencies between .008 and .1 Hz. In the HCP dataset, images only underwent highpass filtering with a cutoff of .008 Hz. High frequency signals were retained because removing high frequency signals would have resulted in excessive loss of degrees of freedom due to the very low TR (.75 seconds) (Bright et al., 2017; Caballero-Gaudes and Reynolds, 2017). In the MyConnectome dataset, volumes exceeding 0.25mm frame-wise displacement or 1.55

standardized DVARS (Nichols, 2017; Power et al., 2014) were censored. In the HCP dataset, we followed the empirically derived thresholds of .39mm frame-wise displacement or 4.9 units above the median DVARS as recommended (Burgess et al., 2016). In the Dunedin dataset, we investigated a range of framewise-displacement cutoffs using QC-RSFC plots in order to derive the optimal threshold for removing motion artifact as recommended (Power et al., 2014). This investigation led to thresholds of 0.35mm framewise-displacement and 1.55 standardized DVARS. In all datasets, nuisance regression, bandpass filtering and censoring for each time-series was performed in a single processing step using AFNI's 3dTproject. Identical time-series processing was applied to resting-state and task time-series with one exception. To remove functionally connectivity predominantly driven be task-evoked coactivation, activation due to task structure was added as an addition nuisance covariate to all task scans and removed from their time-series (Fair et al., 2007) (see supplemental for details on task modeling and regression).

Functional Connectivity processing

We investigated whole brain intrinsic connectivity using 264 brain regions from a parcellation scheme derived in a large independent dataset (Power et al., 2011). BOLD data were averaged within 5mm spheres surrounding each of the 264 coordinates in the parcellation. These average time-series were extracted independently from each scan session in every dataset. This allowed the time-series to be flexibly recombined. Correlation matrices were derived from these time-series using Pearson correlation.

Reliability estimation

We first used the MyConnectome dataset to recreate and extend previous findings regarding measurement reliability (Laumann et al., 2015). First, data from all 84 sessions of resting-state scans were used to create correlation matrices. Sessions were randomly split into two equal halves of 44 sessions (average 398.51 minutes of data post-censoring), and all the time-series from one half of the data were concatenated together and cross correlation of this time-series was computed to define the "true" resting-state correlation matrix. Then, using a subset of the other resting-state sessions, a second correlation matrix was created. The upper-triangle from each correlation matrix was then extracted and fisher-transformed to increase normality. The Pearson correlation between the upper triangle of the "true" correlation matrix and a subset of data (ranging from ½ a session to 40 sessions) was computed to define reliability. For each subset of data, sessions were randomly resampled 1000 times into the "true" and subset halves of the data to get robust estimates of mean reliability.

This resampling procedure was then repeated with our more general definition of intrinsic connectivity. In this analysis, all task and resting-state scans were combined to define general functional connectivity (GFC). To ensure correlation matrices were approximately equally representative of all task and resting-state data, 40 scan sessions (made up of 8 sessions from each of 4 tasks and 8 from rest) were used. As in the resting-state only analyses, scan sessions were evenly split into 2 pools of 20 scan sessions. Each pool of 20 scans consisted of 4 randomly selected scan sessions from each of the 4 tasks and rest. The "true" correlation matrix was created by concatenating and cross-correlating the time-series from one pool of 20 scans. Then reliability was calculated by calculating the Pearson correlation between the Fisher transformed

upper triangle of the "true" correlation matrix with a second correlation matrix computed from randomly selecting a subset (1 to 20 scans) from the second pool of scans. The sampling procedure was repeated 1000 times for each number of scan sessions used to create the subset.

Intraclass correlation (ICC) was used to quantify the test-retest reliability of intrinsic connectivity measures in the HCP and Dunedin test-retest datasets. ICC (3,1) was used in all analyses (Chen et al., 2018). In the HCP dataset, the influence of the amount of data, as well as type of data (rest-state and task) on ICC was investigated. In the resting-state analyses, intrinsic connectivity matrices were calculated across a range of scan lengths (5, 10, 15, 20, 25, 30, 35 and 40 minutes of post-censoring data) mirrored in the test and retest samples. ICC was then calculated for each edge of this matrix, yielding 34716 (264*(264-1)/2) unique ICC values for each scan length. A more general definition of intrinsic connectivity was derived by including task data in the intrinsic connectivity matrices. The influence of adding task to rest was quantified across a range of scan lengths. Increasing amounts of task (5, 10, 15 ,20 ,25 ,30 ,35 and 40 minutes) was added to 5, 10, 15 and 20 minutes of rest to test the influence of reliability within a practical range of data availability. In addition, a pure form of GFC was formally investigated (figures 1,2 and 3). This involved constructing a dataset from an equal amount of each task and rest. This pure GFC was investigated over a variety of scan lengths (5, 10, 15, 20, 25, 30, 35 and 40 minutes of post-censoring data). In all analyses, an equal amount of data from all scan types (each task and rest) were combined for up to 25 minutes of data. After this point, equal amounts of each task could no longer be added together because some tasks were significantly shorter. Above 25 minutes, data was selected at random from the remaining task TRs to avoid any bias. In the Dunedin study,

2 ICC matrices were constructed. The first was built from each participant's single resting-state scan and the second from all task and resting-state scans.

Connectome predictive mapping

The predictive utility of increasing reliability through our GFC measure was evaluated by testing the ability of the different intrinsic connectivity matrices in predicting intelligence in the HCP and Dunedin datasets. Intelligence was measured in the HCP using the Raven's Progressive Matrices (PMAT24_A_CR) (Raven, 1941). This measure was adopted because it is a proxy for cognitive ability that has been shown to be predicted by intrinsic connectivity in the HCP (Dubois et al., 2018a; Finn et al., 2015; Noble et al., 2017). In the Dunedin study, intelligence was measured using the WAIS-IV (Weschler, 2008). The WAIS-IV is a well-established and validated measure of individual differences in cognitive ability.

In both the HCP and Dunedin datasets, intelligence was predicted using Connectome Predictive Mapping (CPM) (Shen et al., 2017). This framework provides a general method to predict any measure from intrinsic connectivity matrices. In this approach, predictors are filtered by selecting edges that have a $p < .01$ correlation with the measure of interest. Three linear regression predictive models are then built, one from the positive features, one from the negative features and one from the combination of features (Shen et al., 2017). Here we report results from the combined model (see supplemental for results from all models). A leave-one-out cross validation scheme was adopted in which the models were trained with all participants except one and used to predict the measure in the left-out participant. This is repeated until all

participants have been left out, and the correlation between predicted and true scores is adopted as an unbiased effect size measure of predictive utility.

Acknowledgement. The Dunedin Multidisciplinary Health and Development Study is supported by the NZ HRC, NZ MBIE, National Institute on Aging grant R01AG032282, R01AG049789, and UK Medical Research Council grant MR/P005918/1. MLE is supported by the National Science Foundation Graduate Research Fellowship under Grant No. NSF DGE-1644868. The authors declare no competing financial interests.

References

Aarts, A.A., Anderson, J.E., Anderson, C.J., Attridge, P.R., Attwood, A., Axt, J., Babel, M., Bahník, Š., Baranski, E., Barnett-Cowan, M., et al. (2015). Estimating the reproducibility of psychological science. Science (80-.). *349*.

Anderson, J.S., Ferguson, M.A., Lopez-Larson, M., and Yurgelun-Todd, D. (2011). Reproducibility of single-subject functional connectivity measurements. Am. J. Neuroradiol. *32*, 548–555.

Avants, B.B., Epstein, C.L., Grossman, M., and Gee, J.C. (2008). Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain. Med. Image Anal. *12*, 26–41.

Barch, D.M., Burgess, G.C., Harms, M.P., Petersen, S.E., Schlaggar, B.L., Corbetta, M., Glasser, M.F., Curtiss, S., Dixit, S., Feldt, C., et al. (2013). Function in the human connectome: Task-fMRI and individual differences in behavior. Neuroimage *80*, 169–189.

Behzadi, Y., Restom, K., Liau, J., and Liu, T.T. (2007). A component based noise correction method (CompCor) for BOLD and perfusion based fMRI. Neuroimage *37*, 90–101.

Betti, V., DellaPenna, S., de Pasquale, F., Mantini, D., Marzetti, L., Romani, G.L., and Corbetta, M. (2013). Natural scenes viewing alters the dynamics of functional connectivity in the human brain. Neuron *79*, 782–797.

Birn, R.M., Molloy, E.K., Patriat, R., Parker, T., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., and Prabhakaran, V. (2013). The effect of scan length on the reliability of resting-state fMRI connectivity estimates. Neuroimage *83*, 550–558.

Braga, R.M., and Buckner, R.L. (2017). Parallel Interdigitated Distributed Networks within the Individual Estimated by Intrinsic Functional Connectivity. Neuron *95*, 457–471.e5.

Bright, M.G., Tench, C.R., and Murphy, K. (2017). Potential pitfalls when denoising resting state fMRI data using nuisance regression. Neuroimage *154*, 159–168.

Buckner, R.L., Andrews-Hanna, J.R., and Schacter, D.L. (2008). The Brain's Default Network. Ann. N. Y. Acad. Sci. *1124*, 1–38.

Buckner, R.L., Krienen, F.M., and Yeo, B.T.T. (2013). Opportunities and limitations of intrinsic functional connectivity MRI. Nat. Neurosci. *16*, 832–837.

Burgess, G.C., Kandala, S., Nolan, D., Laumann, T.O., Power, J.D., Adeyemo, B., Harms, M.P., Petersen, S.E., and Barch, D.M. (2016). Evaluation of Denoising Strategies to Address Motion-Correlated Artifacts in Resting-State Functional Magnetic Resonance Imaging Data from the Human Connectome Project. Brain Connect. *6*, 669–680.

Button, K.S., Ioannidis, J.P.A., Mokrysz, C., Nosek, B.A., Flint, J., Robinson, E.S.J., and Munafò, M.R. (2013). Power failure: Why small sample size undermines the reliability of neuroscience. Nat. Rev. Neurosci. *14*, 365–376.

Caballero-Gaudes, C., and Reynolds, R.C. (2017). Methods for cleaning the BOLD fMRI signal. Neuroimage *154*, 128–149.

Cannon, T.D., Cao, H., Mathalon, D.H., and Gee, D.G. (2018). Reliability of an fMRI paradigm for emotional processing in a multisite longitudinal study: Clarification and implications for statistical power. Hum. Brain Mapp. *39*, 599–601.

Casey, B.J., Cannonier, T., Conley, M.I., Cohen, A.O., Barch, D.M., Heitzeg, M.M., Soules, M.E., Teslovich, T., Dellarco, D. V., Garavan, H., et al. (2018). The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. Dev. Cogn. Neurosci.

Chen, G., Taylor, P.A., Haller, S.P., Kircanski, K., Stoddard, J., Pine, D.S., Leibenluft, E., Brotman, M.A., and Cox, R.W. (2018). Intraclass correlation: Improved modeling approaches and applications for neuroimaging. Hum. Brain Mapp. *39*, 1187–1206.

Christoff, K., Gordon, A.M., Smallwood, J., Smith, R., and Schooler, J.W. (2009). Experience sampling during fMRI reveals default network and executive system contributions to mind wandering. Proc. Natl. Acad. Sci. *106*, 8719–8724.

Cicchetti, D. V. (1994). Guidlines, Criteria, and Rules of Thumb for Evalauting Normed and Standardized Assessment Instruments in Psychology. Psychol. Assess. *6*, 284–290.

Cole, M.W., Anticevic, A., Repovs, G., and Barch, D. (2011). Variable global dysconnectivity and individual differences in schizophrenia. Biol. Psychiatry *70*, 43–50.

Cole, M.W., Bassett, D.S., Power, J.D., Braver, T.S., and Petersen, S.E. (2014). Intrinsic and task-evoked network architectures of the human brain. Neuron *83*, 238–251.

Cole, M.W., Ito, T., Bassett, D.S., and Schultz, D.H. (2016). Activity flow over resting-state networks shapes cognitive task activations. Nat. Neurosci. *19*, 1718–1726.

Collins, F.S., and Varmus, H. (2015). A New Initiative on Precision Medicine. N. Engl. J. Med. *372*, 793–795.

Cox, R.W. (1996). AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. *29*, 162–173.

Deco, G., Hagmann, P., Hudetz, A.G., and Tononi, G. (2014). Modeling resting-state functional networks when the cortex falls asleep: Local and global changes. Cereb. Cortex *24*, 3180–3194.

Van Dijk, K.R.A., Hedden, T., Venkataraman, A., Evans, K.C., Lazar, S.W., and Buckner, R.L. (2010). Intrinsic Functional Connectivity As a Tool For Human Connectomics: Theory, Properties, and Optimization. J. Neurophysiol. *103*, 297–321.

Dong, D., Wang, Y., Chang, X., Luo, C., and Yao, D. (2018). Dysfunction of Large-Scale Brain Networks in Schizophrenia: A Meta-analysis of Resting-State Functional Connectivity. Schizophr. Bull. *44*, 168–181.

Dubois, J., and Adolphs, R. (2016). Building a Science of Individual Differences from fMRI. Trends Cogn. Sci. *20*, 425–443.

Dubois, J., Galdi, P., Han, Y., Paul, L.K., Adolphs, R., Angeles, L., Systems, I., and Dubois, J. (2018a). Resting-state functional brain connectivity best predicts the personality dimension of openness to experience. BioRxiv 215129.

Dubois, J.C., Galdi, P., Paul, L.K., and Adolphs, R. (2018b). A distributed brain network predicts general intelligence from resting-state human neuroimaging data. BioRxiv 257865.

Elliott, L., Sharp, K., Alfaro-Almagro, F., Douaud, G., Miller, K., Marchini, J., and Smith, S. (2017). The genetic basis of human brain structure and function: 1,262 genome-wide associations found from 3,144 GWAS of multimodal brain imaging phenotypes from 9,707 UK Biobank participants. BioRxiv.

Elliott, M.L., Romer, A., Knodt, A.R., and Hariri, A.R. (2018). A Connectome Wide Functional Signature of Transdiagnostic Risk for Mental Illness. Biol. Psychiatry 1–8.

Errington, T.M., Iorns, E., Gunn, W., Tan, F.E. lisabet., Lomax, J., and Nosek, B.A. (2014). An open investigation of the reproducibility of cancer biology research. Elife *3*.

Van Essen, D.C., Smith, S.M., Barch, D.M., Behrens, T.E.J., Yacoub, E., and Ugurbil, K. (2013). The WU-Minn Human Connectome Project: An overview. Neuroimage *80*, 62–79.

Fair, D.A., Schlaggar, B.L., Cohen, A.L., Miezin, F.M., Dosenbach, N.U.F., Wenger, K.K., Fox, M.D.,

Snyder, A.Z., Raichle, M.E., and Petersen, S.E. (2007). A method for using blocked and event-related fMRI data to study "resting state" functional connectivity. Neuroimage *35*, 396–405.
Finn, E.S., Shen, X., Scheinost, D., Rosenberg, M.D., Huang, J., Chun, M.M., Papademetris, X., and Constable, R.T. (2015). Functional connectome fingerprinting: Identifying individuals using patterns of brain connectivity. Nat. Neurosci. *18*, 1664–1671.
Finn, E.S., Scheinost, D., Finn, D.M., Shen, X., Papademetris, X., and Constable, R.T. (2017). Can brain state be manipulated to emphasize individual differences in functional connectivity? Neuroimage *160*, 140–151.
Fox, M.D. (2010). Clinical applications of resting state functional connectivity. Front. Syst. Neurosci. *4*.
Fox, M.D., Snyder, A.Z., Vincent, J.L., and Raichle, M.E. (2007). Intrinsic Fluctuations within Cortical Systems Account for Intertrial Variability in Human Behavior. Neuron *56*, 171–184.
Ge, T., Holmes, A.J., Buckner, R.L., Smoller, J.W., and Sabuncu, M.R. (2017). Heritability analysis with repeat measurements and its application to resting-state functional connectivity. Proc. Natl. Acad. Sci. *114*, 5521–5526.
Geerligs, L., Rubinov, M., Cam-CAN, and Henson, R.N. (2015). State and Trait Components of Functional Connectivity: Individual Differences Vary with Mental State. J. Neurosci. *35*, 13949–13961.
Glahn, D.C., Winkler, A.M., Kochunov, P., Almasy, L., Duggirala, R., Carless, M.A., Curran, J.C., Olvera, R.L., Laird, A.R., Smith, S.M., et al. (2010). Genetic control over the resting brain. Proc Natl Acad Sci U S A *107*, 1223–1228.
Glasser, M.F., Sotiropoulos, S.N., Wilson, J.A., Coalson, T.S., Fischl, B., Andersson, J.L., Xu, J., Jbabdi, S., Webster, M., Polimeni, J.R., et al. (2013). The minimal preprocessing pipelines for the Human Connectome Project. Neuroimage *80*, 105–124.
Gordon, E.M., Laumann, T.O., Gilmore, A.W., Newbold, D.J., Greene, D.J., Berg, J.J., Ortega, M., Hoyt-Drazen, C., Gratton, C., Sun, H., et al. (2017). Precision Functional Mapping of Individual Human Brains. Neuron *95*, 791–807.e7.
Gratton, C., Laumann, T.O., Nielsen, A.N., Greene, D.J., Gordon, E.M., Gilmore, A.W., Nelson, S.M., Coalson, R.S., Snyder, A.Z., Schlaggar, B.L., et al. (2018). Functional Brain Networks Are Dominated by Stable Group and Individual Factors, Not Cognitive or Daily Variation. Neuron 439–452.
Greicius, M. (2008). Resting-state functional connectivity in neuropsychiatric disorders. Curr. Opin. Neurol. *21*, 424–430.
Greve, D.N., and Fischl, B. (2009). Accurate and robust brain image alignment using boundary-based registration. Neuroimage *48*, 63–72.
Hacker, C.D., Laumann, T.O., Szrama, N.P., Baldassarre, A., Snyder, A.Z., Leuthardt, E.C., and Corbetta, M. (2013). Resting state network estimation in individual subjects. Neuroimage *82*, 616–633.
Hamburg, M.A., and Collins, F.S. (2010). The Path to Personalized Medicine - Perspective. N. Engl. J. Med. *363*, 301–304.
Hasson, U., Nir, Y., Levy, I., Fuhrmann, G., and Malach, R. (2004). Intersubject Synchronization of Cortical Activity during Natural Vision. Science (80-.). *303*, 1634–1640.
Hasson, U., Malach, R., and Heeger, D.J. (2010). Reliability of cortical activity during natural stimulation. Trends Cogn. Sci. *14*, 40–48.

Honey, C.J., Honey, C.J., Sporns, O., Sporns, O., Cammoun, L., Cammoun, L., Gigandet, X., Gigandet, X., Thiran, J.P., Thiran, J.P., et al. (2009). Predicting human resting-state functional connectivity from structural connectivity. Proc. Natl. Acad. Sci. U. S. A. *106*, 2035–2040.
Hurlburt, R.T., Alderson-Day, B., Fernyhough, C., and Kï¿½hn, S. (2015). What goes on in the resting-state? A qualitative glimpse into resting-state experience in the scanner. Front. Psychol. *6*.
Huth, A.G., De Heer, W.A., Griffiths, T.L., Theunissen, F.E., and Gallant, J.L. (2016). Natural speech reveals the semantic maps that tile human cerebral cortex. Nature *532*, 453–458.
Ioannidis, J.P.A. (2005). Why most published research findings are false. PLoS Med. *2*, 0696–0701.
Ioannidis, J.P.A. (2008). Why Most Discovered True Associations Are Inflated. Epidemiology *19*, 640–648.
Ioannidis, J.P.A., Ntzani, E.E., Trikalinos, T.A., and Contopoulos-Ioannidis, D.G. (2001). Replication validity of genetic association studies. Nat. Genet. *29*, 306–309.
Jo, H.J., Gotts, S.J., Reynolds, R.C., Bandettini, P.A., Martin, A., Cox, R.W., and Saad, Z.S. (2013). Effective preprocessing procedures virtually eliminate distance-dependent motion artifacts in resting state FMRI. J. Appl. Math. *2013*.
Kaiser, R.H., Andrews-Hanna, J.R., Wager, T.D., and Pizzagalli, D.A. (2015). Large-Scale Network Dysfunction in Major Depressive Disorder A Meta-analysis of Resting-State Functional Connectivity. JAMA Psychiatry *72*, 603–611.
Kanyongo, G.Y., Brook, G.P., Kyei-Blankson, L., and Gocmen, G. (2007). Reliability and Statistical Power: How Measurement Fallibility Affects Power and Required Sample Sizes for Several Parametric and Nonparametric Statistics. J. Mod. Appl. Stat. Methods *6*, 81–90.
Klein, A., Andersson, J., Ardekani, B.A., Ashburner, J., Avants, B., Chiang, M.C., Christensen, G.E., Collins, D.L., Gee, J., Hellier, P., et al. (2009). Evaluation of 14 nonlinear deformation algorithms applied to human brain MRI registration. Neuroimage *46*, 786–802.
Krienen, F.M., Yeo, B.T.T., and Buckner, R.L. (2014). Reconfigurable task-dependent functional coupling modes cluster around a core functional architecture. Philos. Trans. R. Soc. Lond. B. Biol. Sci. *369*, 20130526-.
Lahnakoski, J.M., Glerean, E., Salmi, J., Jääskeläinen, I.P., Sams, M., Hari, R., and Nummenmaa, L. (2012). Naturalistic fMRI Mapping Reveals Superior Temporal Sulcus as the Hub for the Distributed Brain Network for Social Perception. Front. Hum. Neurosci. *6*.
Laumann, T.O., Gordon, E.M., Adeyemo, B., Snyder, A.Z., Joo, S.J., Chen, M.Y., Gilmore, A.W., McDermott, K.B., Nelson, S.M., Dosenbach, N.U.F., et al. (2015). Functional System and Areal Organization of a Highly Sampled Individual Human Brain. Neuron *87*, 658–671.
Matthews, P.M., Honey, G.D., and Bullmore, E.T. (2006). Applications of fMRI in translational medicine and clinical practice. Nat. Rev. Neurosci. *7*, 732–744.
Miller, K.L., Alfaro-Almagro, F., Bangerter, N.K., Thomas, D.L., Yacoub, E., Xu, J., Bartsch, A.J., Jbabdi, S., Sotiropoulos, S.N., Andersson, J.L.R., et al. (2016). Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat. Neurosci. *19*, 1523.
Mueller, S., Wang, D., Fox, M.D., Yeo, B.T.T., Sepulcre, J., Sabuncu, M.R., Shafee, R., Lu, J., and Liu, H. (2013). Individual Variability in Functional Connectivity Architecture of the Human Brain. Neuron *77*, 586–595.
Munafò, M.R., Nosek, B.A., Bishop, D.V.M., Button, K.S., Chambers, C.D., Percie Du Sert, N.,

Simonsohn, U., Wagenmakers, E.J., Ware, J.J., and Ioannidis, J.P.A. (2017). A manifesto for reproducible science. Nat. Hum. Behav. *1*, 1–9.
Nichols, T.E. (2017). Notes on Creating a Standardized Version of DVARS. 1–5.
Noble, S., Spann, M.N., Tokoglu, F., Shen, X., Constable, R.T., and Scheinost, D. (2017). Influences on the Test–Retest Reliability of Functional Connectivity MRI and its Relationship with Behavioral Utility. Cereb. Cortex 1–15.
Nunnally Jr., J.C. (1970). Introduction to psychological measurement. (New York, NY, US: McGraw-Hill).
Patriat, R., Molloy, E.K., Meier, T.B., Kirk, G.R., Nair, V.A., Meyerand, M.E., Prabhakaran, V., and Birn, R.M. (2013). The effect of resting condition on resting-state fMRI reliability and consistency: A comparison between resting with eyes open, closed, and fixated. Neuroimage *78*, 463–473.
Poldrack, R.A., Laumann, T.O., Koyejo, O., Gregory, B., Hover, A., Chen, M.Y., Gorgolewski, K.J., Luci, J., Joo, S.J., Boyd, R.L., et al. (2015). Long-term neural and physiological phenotyping of a single human. Nat. Commun. *6*.
Poulton, R., Honcox, R.J., Milne, B.J., Baxter, J.M., Scott, K., and Wilson, N. (2006). The Dunedin multidisciplinary health and development study: Are its findings consistent with the overall New Zealand population? N. Z. Med. J. *119*.
Poulton, R., Moffitt, T.E., and Silva, P.A. (2015). The Dunedin Multidisciplinary Health and Development Study: overview of the first 40 years, with an eye to the future. Soc. Psychiatry Psychiatr. Epidemiol. *50*, 679–693.
Power, J.D., Cohen, A.L., Nelson, S.M., Wig, G.S., Barnes, K.A., Church, J.A., Vogel, A.C., Laumann, T.O., Miezin, F.M., Schlaggar, B.L., et al. (2011). Functional Network Organization of the Human Brain. Neuron *72*, 665–678.
Power, J.D., Barnes, K.A., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2012). Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. Neuroimage *59*, 2142–2154.
Power, J.D., Schlaggar, B.L., Lessov-Schlaggar, C.N., and Petersen, S.E. (2013). Evidence for hubs in human functional brain networks. Neuron *79*, 798–813.
Power, J.D., Mitra, A., Laumann, T.O., Snyder, A.Z., Schlaggar, B.L., and Petersen, S.E. (2014). Methods to detect, characterize, and remove motion artifact in resting state fMRI. Neuroimage *84*, 320–341.
Raven, J.C. (1941). Standardization of progressive matrices. Br. J. Med. Psychol. *19*, 137–150.
Satterthwaite, T.D., Elliott, M.A., Gerraty, R.T., Ruparel, K., Loughead, J., Calkins, M.E., Eickhoff, S.B., Hakonarson, H., Gur, R.C., Gur, R.E., et al. (2013). An improved framework for confound regression and filtering for control of motion artifact in the preprocessing of resting-state functional connectivity data. Neuroimage *64*, 240–256.
Satterthwaite, T.D., Cook, P.A., Bruce, S.E., Conway, C., Mikkelsen, E., Satchell, E., Vandekar, S.N., Durbin, T., Shinohara, R.T., and Sheline, Y.I. (2016). Dimensional depression severity in women with major depression and post-traumatic stress disorder correlates with fronto-amygdalar hypoconnectivty. Mol. Psychiatry *21*, 894–902.
Satterthwaite, T.D., Xia, C.H., and Bassett, D.S. (2018). Personalized Neuroscience: Common and Individual-Specific Features in Functional Brain Networks. Neuron *98*, 243–245.
Shah, L.M., Cramer, J.A., Ferguson, M.A., Birn, R.M., and Anderson, J.S. (2016). Reliability and

reproducibility of individual differences in functional connectivity acquired during task and resting state. Brain Behav. *6*, 1–15.
Shehzad, Z., Kelly, A.M.C., Reiss, P.T., Gee, D.G., Gotimer, K., Uddin, L.Q., Lee, S.H., Margulies, D.S., Roy, A.K., Biswal, B.B., et al. (2009). The resting brain: Unconstrained yet reliable. Cereb. Cortex *19*, 2209–2229.
Shen, X., Finn, E.S., Scheinost, D., Rosenberg, M.D., Chun, M.M., Papademetris, X., and Constable, R.T. (2017). Using connectome-based predictive modeling to predict individual behavior from brain connectivity. Nat. Protoc. *12*, 506–518.
Smith, S.M., Beckmann, C.F., Andersson, J., Auerbach, E.J., Bijsterbosch, J., Douaud, G., Duff, E., Feinberg, D.A., Griffanti, L., Harms, M.P., et al. (2013). Resting-state fMRI in the Human Connectome Project. Neuroimage *80*, 144–168.
Swartz, J.R., Knodt, A.R., Radtke, S.R., and Hariri, A.R. (2015). A neural biomarker of psychological vulnerability to future life stress. Neuron *85*, 505–511.
Szucs, D., and Ioannidis, J.P.A. (2017). Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature. PLoS Biol. *15*.
Tagliazucchi, E., and Laufs, H. (2014). Decoding Wakefulness Levels from Typical fMRI Resting-State Data Reveals Reliable Drifts between Wakefulness and Sleep. Neuron *82*, 695–708.
Tavor, I., Parker Jones, O., Mars, R.B., Smith, S.M., Behrens, T.E., and Jbabdi, S. (2016). Task-free MRI predicts individual differences in brain activity during task performance. Science *352*, 216–220.
Thompson, P.M., Stein, J.L., Medland, S.E., Hibar, D.P., Vasquez, A.A., Renteria, M.E., Toro, R., Jahanshad, N., Schumann, G., Franke, B., et al. (2014). The ENIGMA Consortium: Large-scale collaborative analyses of neuroimaging and genetic data. Brain Imaging Behav. *8*, 153–182.
Vanderwal, T., Eilbott, J., Finn, E.S., Craddock, R.C., Turnbull, A., and Castellanos, F.X. (2017). Individual differences in functional connectivity during naturalistic viewing conditions. Neuroimage *157*, 521–530.
Vul, E., Harris, C., Winkielman, P., and Pashler, H. (2009). Puzzlingly High Correlations in fMRI Studies of Emotion, Personality, and Social Cognition1. Perspect. Psychol. Sci. *4*, 274–290.
Waites, A.B., Stanislavsky, A., Abbott, D.F., and Jackson, G.D. (2005). Effect of prior cognitive state on resting state networks measured with functional connectivity. Hum. Brain Mapp. *24*, 59–68.
Weschler, D. (2008). Wechsler Adult Intelligence Scale - Fourth Edition. Stat. Solut. 1–3.
Williams, R.H., and Zimmerman, D.W. (1989). Statistical power analysis and reliability of measurement. J. Gen. Psychol. *116*, 359–369.
Wong, C.W., Olafsson, V., Tal, O., and Liu, T.T. (2012). Anti-correlated networks, global signal regression, and the effects of caffeine in resting-state functional MRI. Neuroimage *63*, 356–364.
Yeo, B.T.T., Krienen, F.M., Sepulcre, J., Sabuncu, M.R., Lashkari, D., Hollinshead, M., Roffman, J.L., Smoller, J.W., Zollei, L., Polimeni, J.R., et al. (2011). The organization of the human cerebral cortex estimated by intrinsic functional connectivity. J. Neurophysiol. *106*, 1125–1165.
Yuan, W., Altaye, M., Ret, J., Schmithorst, V., Byars, A.W., Plante, E., and Holland, S.K. (2009). Quantification of head motion in children during various fMRI language tasks. Hum. Brain Mapp. *30*, 1481–1489.

Figures

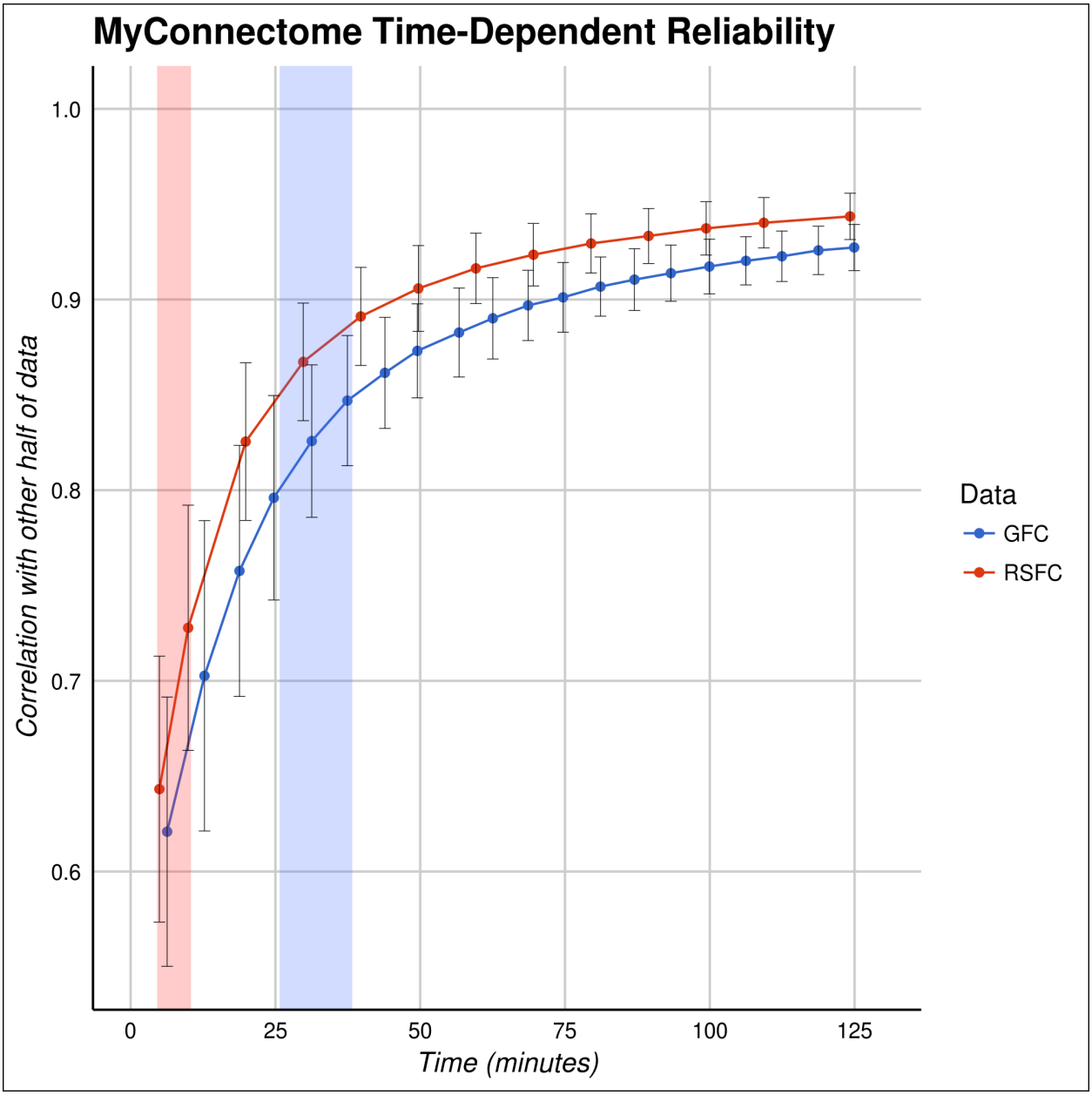

Figure 1. Reliability of individual differences in intrinsic connectivity directly reflects the amount of data collected either from resting-state scans alone or combined with task scans. Pearson correlations with the second-half "true" correlation matrix (125 minutes of data) as a function of smaller subsets of data. The split-half reliability curve for resting-state functional connectivity (RSFC) is in red and for general functional connectivity (GFC) in blue. The red bar denotes 5 – 10 minutes of resting-state data alone typically collected in the majority of studies, while the blue bar denotes 25-35 minutes of resting-state plus task scans also typically available in these studies.

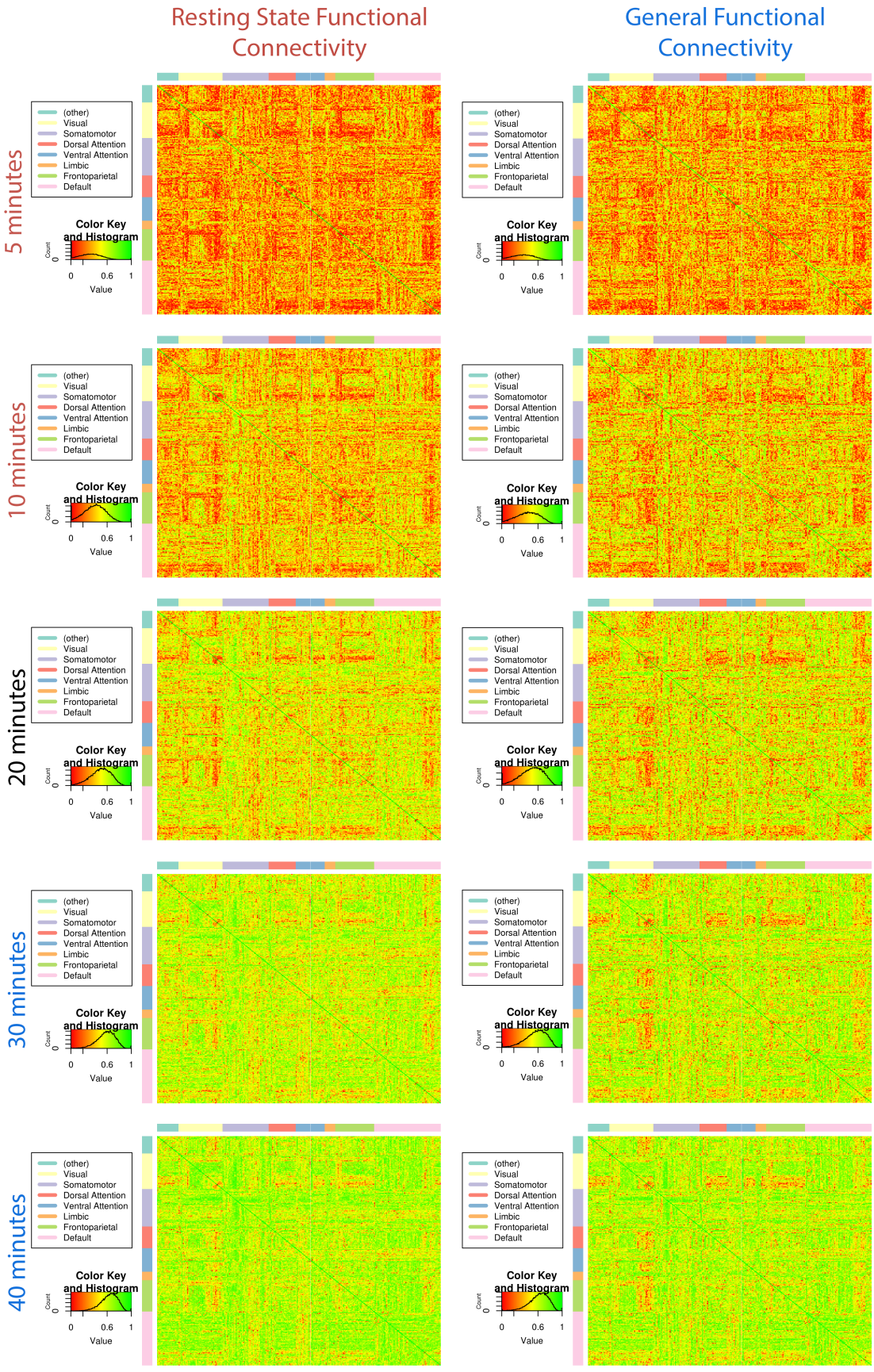

Figure 2. Test-retest reliability of individual differences in both RSFC and GFC scales with the amount of the data available for analysis. Intra-Class Correlation (ICC) matrices for RSFC (left panel) and GFC (right panel) demonstrate comparable gains in reliability across common neural networks as defined by (Power et al., 2011). 5 and 10 minutes are written in red because these are common scan lengths for resting-state scans. 30 and 40 minutes are written in blue because many researchers have collected this amount of fMRI data when resting-state and task fMRI is combined. To the bottom left of ICC matrices is the color key for the ICC values with a histogram indicating the density of ICCs for the corresponding graph.

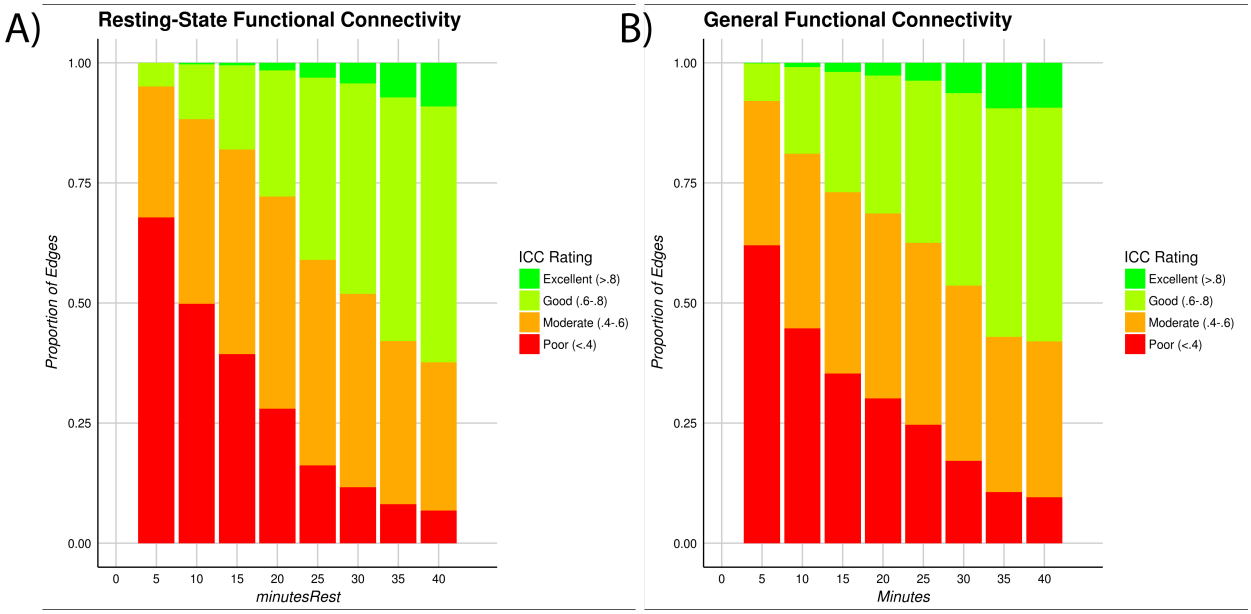

Figure 3. Stacked bar plots displaying the proportion of functional connections (i.e., edges) across neural networks as defined by Power et al. (Power et al., 2011) that meet criteria for excellent, good, moderate and poor reliability as indexed by ICCs as a function of how much data is used to estimate resting-state functional connectivity (A) and general functional connectivity (B).

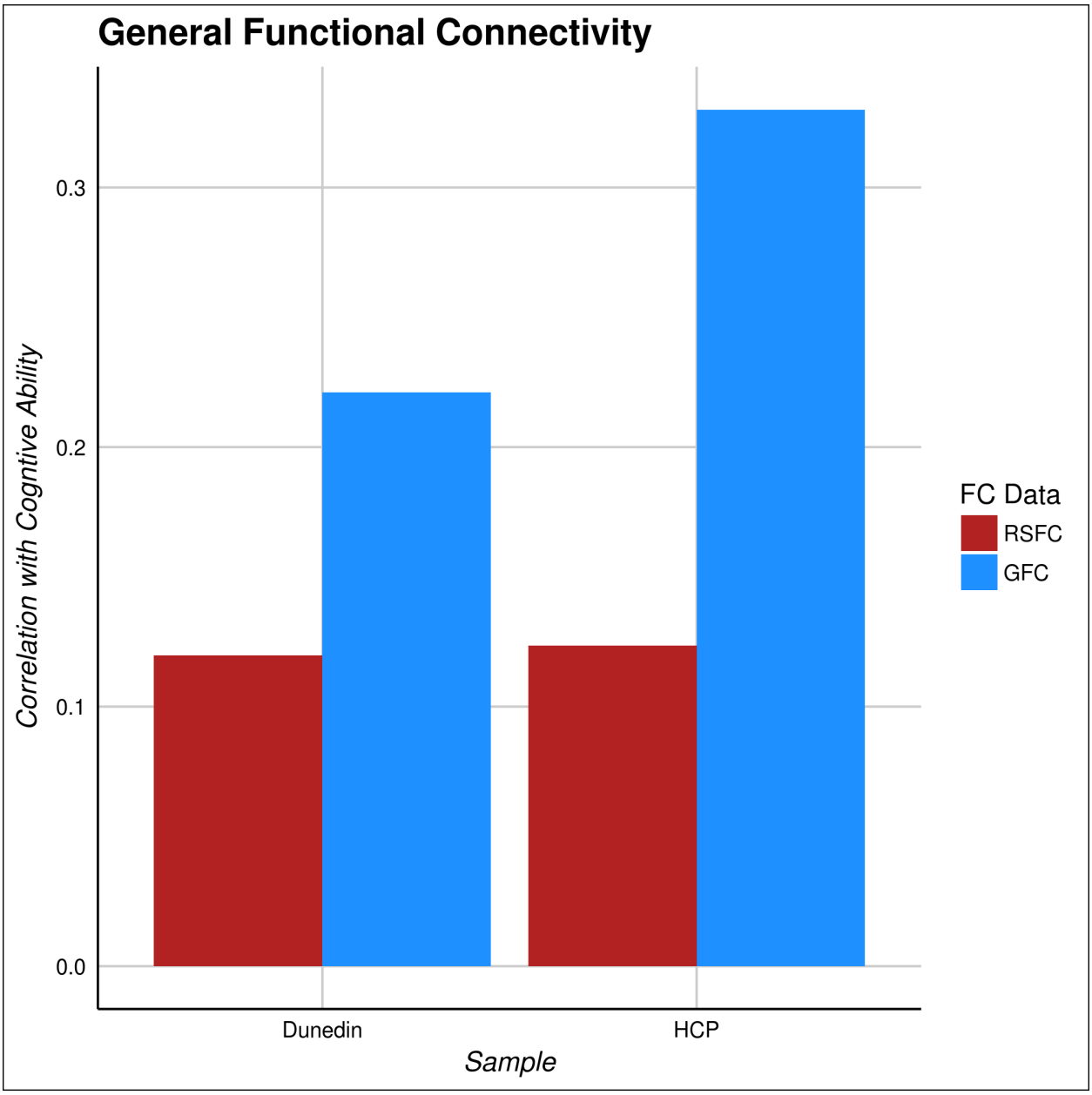

Figure 4. Connectome predictive mapping (CPM) analyses of individual differences in intelligence reveal significant boosts in variance explained when adopting general functional connectivity (GFC) in comparison with resting-state functional connectivity (RSFC). 8 minutes of rest-state data combined with 25 minutes of task data were used for the GFC estimates in both samples. These amounts were chosen so that scan lengths could be the same across samples and because these scan lengths are commonly found in fMRI imaging datasets.