Infant functional connectivity fingerprints predict long-term language and pre-literacy outcomes

Xi Yu[1,2*], Silvina Ferradal[3,4*], Danielle D. Sliva[2,5], Jade Dunstan[2], Clarisa Carruthers[2], Joseph Sanfilippo[2], Jennifer Zuk[2,7], Lilla Zöllei[6], Emma Boyd[6], Borjan Gagoski[4,7], P. Ellen Grant[4,7], Nadine Gaab[2,7]

[1]State Key Laboratory of Cognitive Neuroscience and Learning, Beijing Normal University, Beijing, 100875, China;

[2]Laboratories of Cognitive Neuroscience, Division of Developmental Medicine, Department of Medicine, Boston Children's Hospital, MA, 02115, USA;

[3]Department of Intelligent Systems Engineering, Indiana University, Bloomington, IN, 47408, USA;

[4]Fetal Neonatal Neuroimaging and Developmental Science Center, Boston Children's Hospital, Boston, MA, 02115, USA;

[5]Department of Neuroscience, Brown University, Providence, RI, 02912, USA;

[6]A.A. Martinos Center for Biomedical Imaging, Massachusetts General Hospital, Boston, MA, 02129, USA;

[7]Harvard Medical School, Boston, MA, 02115, USA.

*Authors contributed equally to this work

Corresponding author:

Xi Yu, PhD

State Key Laboratory of Cognitive Neuroscience and Learning

Beijing Normal University,

Beijing 100875, China

xi.yu@bnu.edu.cn

Acknowledgments.

This study was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development #R01HD65762-01 (awarded to N.G. and P.E.G), the Charles H. Hood Foundation (awarded to N.G.), and the Boston Children's Hospital Pilot Grant (awarded to N.G.).

We sincerely thank all the families for their participation in this longitudinal study.

Abstract

Functional connectivity analyses can delineate brain organization in infancy, enabling the characterization of early brain characteristics associated with subsequent behavioral outcomes. Previous studies have identified specific functional networks of infant brains that underlie cognitive abilities and pathophysiology subsequently observed in toddlers. However, it is unknown whether and how functional networks emerging in infancy contribute to the development of higher-order complex functions of language and literacy. Employing a five-year longitudinal imaging project starting in infancy, the current study utilizes resting-state functional connectivity MRI to examine prospective associations between infant functional connectivity fingerprints (FCF) and subsequent school-age language and foundational literacy skills. Utilizing multivariate pattern analyses, we demonstrate that FCF patterns associated with language-related brain regions established during the first year of life are prospectively associated with oral language and phonological skills at a mean age of 6.5. Moreover, these longitudinal associations are preserved when key environmental influences are controlled for and are independent of emergent language abilities in infancy, suggesting early development of functional network characteristics in supporting the acquisition of high-order language and pre-literacy skills. Altogether, the current results highlight the importance of functional organization established in infancy as a neural scaffold underlying the protracted development of complex cognitive functions.

Significance statement

Functional brain networks undergo extensive development within the first few years of life. Although functional connectivity has been linked to cognitive abilities in toddlerhood, little is known regarding the long-term relevance of functional connections established in infancy for the protracted development of higher-order abilities such as language. Employing a five-year infant longitudinal imaging project, we demonstrate, for the first time, prospective associations between functional connectivity patterns in infancy and language and pre-literacy abilities after the onset of formal schooling. These results provide insights into the far-reaching developmental significance of early neural network mechanisms on the learning process of higher-order cognitive functions.

Introduction

Human brains undergo extensive changes within the first years of life, establishing critical foundations of neuroarchitecture that underlies the protracted maturational process of perceptual and higher-order cognitive functions. Recent advances in pediatric neuroimaging techniques refine functional mapping of the infant brain, primarily through resting-state functional connectivity (FC) MRI analyses (1). FC characterizes functional brain organization by examining synchronized fluctuations in the blood oxygen dependent (BOLD) signals among different brain regions (2). For instance, FC analyses conducted in late pre-term and full-term neonates have reported adult-like functional connectivity patterns in regions underlying motor and primary sensory processing (e.g., primary auditory and visual networks) at birth (3-6). As maturation progresses, long-distance functional connectivity patterns among association cortices develop, coupled with the emergence of higher-order cognitive functions such as executive function and language (5-8). To date, only a few studies have directly examined the cognitive developmental significance of functional connectivity emerging in infancy, revealing associations between infant FC patterns and cognitive abilities and pathophysiology in toddlers and preschoolers (9-11). However, it is still unknown whether and how FC in infancy can account for individual differences in the long-term trajectory of language and foundational literacy abilities into school-age.

Early language acquisition starts in utero and is significantly shaped by the linguistic environment (12). Newborns demonstrate selective preferences for human speech compared to other sounds, such as sine-wave analogs with comparable spectral and temporal parameters as those of real speech (e.g., (13, 14), and are able to recognize and categorize different aspects of speech, such as phonetic and stress patterns (15-17). Infants typically develop an understanding

of familiar words as early as the second half of the first year, building a critical foundation for long-term language development into school-age (18-21). Early linguistic experiences further serve as critical building blocks for the development of emergent literacy skills. These skills include phonological processing, the ability to recognize and manipulate the internal structure of speech sounds, one of the key predictors of subsequent reading development (e.g., (22-24). Moreover, this developmental time course is sensitive to environmental characteristics, including language and early literacy exposure. For example, it has been shown that the quality and quantity of child-directed speech are highly correlated with vocabulary gain in toddlers (25-27). In addition to language exposure, the quality and quantity of home literacy practices are also correlated with subsequent verbal and literacy skills during preschool and school-age (e.g., (28-31), highlighting the important role of the environment. Moreover, emergent literacy skills are then further deepened through instruction once the child starts school and reciprocally influence language development, such as vocabulary (32-35). Overall, the development of language and pre-literacy skills starts in infancy and continues through childhood and beyond, constantly shaped by the quantity and quality of input provided in the child's home, school, and other environmental contexts.

Crucially, at the time of birth, infant brains have already developed the essential neurobiological scaffold for language and literacy acquisition. It has been repeatedly demonstrated that, when listening to human speech compared to control sounds or speech played backward, newborns and preverbal babies activate language-related areas (36, 37), largely located in the temporal lobe (38-41). Electrophysiological responses to speech perception in infancy are prospectively associated with language and literacy outcomes and trajectories at preschool and school-age, suggesting early brain function has a long-lasting influence on

subsequent language and literacy acquisition (e.g., (42-48). Moreover, white matter characteristics have been shown to correlate with the concurrent verbal skills in the first five years of life (49), and white matter organization at term age has been shown to predict preterm infants' language development at age two (50), implying the early emergence of structural connectivity underlying language development. Furthermore, ***large-scale functional topologies*** of infant/toddler brains have been linked to subsequent language development at age 4 (51). More specifically, among 23 children, the developmental trajectories of FC within the first two years of life were found to be associated with language abilities at four years of age, highlighting the significance of functional brain characteristics in the earliest years of life. However, it is still an open question as to what extent functional networks established in infancy are prospectively associated with individual differences in language and literacy development at school-age, after the onset of formal literacy instruction. Moreover, it is unknown whether (pre)language skills and early linguistic environmental variables mediate this association.

Employing a five-year longitudinal imaging dataset from infancy to school-age, the current study utilizes resting-state functional connectivity MRI methods to investigate prospective associations between infant functional brain networks and subsequent measures of language and emergent literacy skills at age 6. We first conducted multivariate pattern analysis (MVPA) to identify functional networks in infancy that are associated with subsequent school-age performance while controlling for critical environmental factors. Further analyses were performed to assess whether the inclusion of receptive and expressive language performance measured in infancy leads to changes in these observed associations. The results of the current study offer the potential to inform our understanding of how early-emerging functional network mechanisms can contribute to individual differences in language and reading abilities later in life.

Results

Infant FCFs can predict school-age language and foundational literacy skills

We first evaluated whether functional connectivity fingerprints (FCF) in infancy were prospectively associated with the subsequent development of language and foundational literacy skills at school-age. The outcome measures included oral language (OL) and two key pre-literacy skills commonly recognized as crucial cognitive precursors of reading – phonological processing and rapid automatized naming (RAN, (52-54). To probe the functional networks specifically associated with long-term language and literacy development, special consideration was taken to minimize the impact of other factors known to influence children's verbal skills. Specifically, key environmental factors, including socio-economic status (SES; indicated through parental educational background) and home literacy environment (HLE), were collected during the infant imaging time point (see SI Methods) and included as covariates along with the infant age at the scan. We further measured the children's general cognitive abilities subsequently at school-age, which were included as a control variable to partial out individual variance in general cognitive maturation. Meanwhile, given the close associations between the resting-state FC pattern of one region and the functional relevance (i.e., task-based fMRI activation) of this region (55-57), region-specific FC patterns were characterized for each infant brain using the FCF approach that computed whole-brain functional connectivity pattern associated with each anatomical region of interest (ROI).

To identify infant functional networks prospectively associated with school-age performance of language and pre-literacy skills, Support Vector Regression (SVR) analyses were performed for each pair of the region-specific FCFs and each measure of language and pre-literacy skill (see Figure 1 for a full description of the analysis). For every infant, region-specific

FCF of each ROI was generated using Pearson's correlations between BOLD time series of this region and that of all cortical ROIs spanning the entirety of the cortex derived from the one-year-old infant AAL atlas (58); see a full list of all 78 regions in table S2). Given the important role of temporal lobe regions in the development of speech perception and language processing (38, 39, 59-61), the current analysis focused on the FCFs of 16 seed ROIs located in the bilateral temporal lobes. The FCF associated with a particular seed region was entered into the SVR models as the input feature with performance on each of the three assessments (OL, phonological processing, RAN) as the outcome measures in order to identify infant functional networks important for development in each specific cognitive domain. Moreover, to minimize overfitting, the SVR model was first trained using 80% of the dataset and then tested using the remaining 20% of the data. This process was repeated 1000 times, generating a distribution of the SVR performance. Permutation tests were subsequently performed following the same procedure except that each of the tasks performed was randomly assigned to every participant in order to generate a random distribution. Both distributions, obtained by the real and shuffled performance, were compared using a Kolmogorov-Smirnov test, and the effect size was further calculated using Cohen's d (62). Here, we only focus on strong associations whose distributions were significantly higher than the respective randomized distributions (corrected for multiple comparisons) with a large effect size (Cohen's $d > 0.8$; see a similar thresholding method in (63), and the full results in Table S3).

The SVR analyses revealed a strong association between the FCF of the middle portion of the left temporal pole (midLTP) in infancy and oral language skills measured at school-age (Figure 2, Table 1). We further identified three regions whose FCF patterns exhibited strong associations with phonological processing skills at the school-age: the left fusiform gyrus

(LFFG), right fusiform gyrus (RFFG), and the right inferior temporal gyrus (RITG, Figure 2, Table 1, see similar results based on a leave-one-out cross-validation approach in SI Methods and Table S4). No region in the temporal lobe showed significant associations between infant FCFs and subsequent RAN abilities. Lastly, correlation analyses were computed between motion parameters (number of outliers and mean FD) and language/literacy assessments at school-age. No significant results were observed (all $p_{\text{uncorrected}} > 0.1$, Table S5), indicating that head movement during image acquisition cannot explain the observed associations.

Early functional network correlates of long-term verbal development are established within the first year of life

Our first analyses revealed strong prospective associations of school-age performance in the FC patterns of infants 4 and 18 months old. Since functional neural networks develop rapidly during the first two years of life (64), the age range of the imaging collection time point might lead to the assumption that the observed relationship could be driven by the more established functional networks among the older infants (>12 months old). Therefore, to further evaluate the importance of the early-emerged functional connectivity patterns in the observed associations, we reran the SVR analyses in the four identified FCFs using a subset of 29 infants who completed the imaging session before one year of age (mean age = 8.6 ± 1.8 months). Based on the same criteria, strong associations were still present between the infant FCF of midLTP and oral language skills at school-age (mean $r = 0.46 \pm 0.37$, $p_{\text{corrected}} = 1.3*10^{-85}$, Cohen's $d = 1.2$) as well as those of left fusiform gyri (mean $r = 0.39 \pm 0.42$, $p_{\text{corrected}} = 1.1*10^{-52}$, Cohen's $d = 0.88$) and RITG (mean $r = 0.33 \pm 0.43$, $p_{\text{corrected}} = 8.8*10^{-47}$, Cohen's $d = 0.82$) for the phonological skills measured at age 6. However, the association between the infant FCF of RFFG and the subsequent phonological skills (mean $r = 0.36 \pm 0.43$, $p_{\text{corrected}} = 3.8*10^{-40}$, Cohen's $d = 0.74$)

failed to outperform permutation results by large effect (Cohen's $d > 0.8$) and therefore was not considered in the subsequent analyses.

Prospective associations between infant FCFs and school-age language and phonological performance are independent of early language milestones

Foundational language skills have been demonstrated to emerge early in infancy, which lay a critical foundation for (and are often associated with) the development of more complex language functions (see (65) for a review). It is, therefore, critical to address the question as to whether the emergent language skills of these infants might explain the observed longitudinal links between the FCFs in infancy and school-age performance. To address this, we further selected a subset of 24 infants from the original sample, whose receptive and expressive language skills were assessed at the infant time point. We then examined whether the inclusion of language skills measured in infancy change the prospective associations between infant functional networks and school-age performance identified in previous analyses; namely, the infant FCF of midLTP and oral language skills, and the infant FCFs of LLF and ITG and phonological skills. Specifically, two SVR models were run for each of the three identified associations. While both models included regular covariates similar to the ones included in the main analyses, only the second model further considered the receptive and expressive language skills measured in infancy. The association distributions derived from the two models were compared using the Kolmogorov-Smirnov test. If the identified prospective associations are *partially* mediated by language skills in infancy, we should expect significant decreases in association performances of the model controlling for infant language skills compared to the model that doesn't control for infant language skills.

Results showed no reduction in the association strength between the identified FCFs and school-age oral language (midLTP: $p_{uncorrected}$ = 0.21) and phonological skills (LFFG: $p_{uncorrected}$ = 0.28; RITG: $p_{uncorrected}$ = 0.93) when the infant language skills were included as covariates compared to without (Figure 3). Lastly, correlation analyses were also run between each of the school-age language and pre-literacy scores and their infancy MSEL scores for the receptive and expressive language skills, which did not reveal any significant associations, even before correction for multiple comparisons (all $p_{\text{uncorrected}} > 0.1$, Table S5).

Discussion

In this study, we provide the first evidence for prospective associations between functional network characteristics in infancy and school-age development of language and foundational literacy skills. Utilizing a five-year longitudinal imaging dataset, we identify three temporal regions whose functional connectivity fingerprint patterns within the first year of life could reliably account for individual differences in oral language (left temporal pole) or phonological skills (left fusiform gyrus and right inferior temporal gyrus) measured at school-age in the same group of children. These strong associations are observed when the critical environmental factors of SES and HLE in infancy, as well as general cognitive abilities at school-age, were controlled, suggesting the early emergence of neural networks for long-term language and phonological development. We further show that these longitudinal links are independent of (pre)language skills in infancy, suggesting that the neural mechanisms underlying subsequently developing language and pre-literacy skills might be, at least partially, rooted early in the emergent functional network architecture of infant brains. Altogether, these findings uncover a critical role of functional connectivity patterns established in infancy on the protracted development of high-order cognitive functions of language and literacy.

The current study utilizes functional connectivity fingerprints to identify brain regions associated with subsequent cognitive development, highlighting the importance of inter-regional interactions in regional functional specialization through development. It has been long hypothesized that the functional development of individual regions is facilitated and shaped by earlier developing connectivity patterns (57, 66). In line with this account, it has been shown that the functional connectivity pattern of one region at rest is associated with task-based fMRI activation of this region (67), and changes in the FC patterns parallel the anatomical boundaries between functionally distinct neighbors (68). Moreover, it has been recently demonstrated that the (structural) connectivity fingerprints at age five before children learned to read are predictive of word-specific functional activation after 3-years' reading instruction (69), providing direct evidence for the developmental significance of network mechanisms for neural functional specialization. Following the same line, our current study demonstrates that functional connectivity pattern established in infancy for a given region is associated with the cognitive skill this region subsequently supports (see more discussion below), suggesting that functional topologies that emerge within the first year of life lay a critical neural foundation for subsequent long-term development. However, future studies are needed to further examine the role of early-developed functional connectivity in shaping the neural correlates of cognitive functions.

The infant FCF pattern of the left temporal pole is associated with the school-age language performance, shedding light on an early functional network mechanism that could underlie the acquisition of complex language skills. Previous fMRI studies have revealed neural activation in the left temporal pole during speech perception in infants (e.g., (38, 39, 60). In adults, the temporal pole is involved in word- and picture-based semantic tasks (e.g., (70-73), and typically viewed as a hub for lexico-semantic representation (74, 75). Therefore, our

working hypothesis is that the temporal pole might serve as a core component of a wide-spread neural substrate that enables the learning of lexico-semantic representation via early language and speech learning/processing. Moreover, the fact that this association is observed when key environmental factors and the school-age general cognitive skills are controlled for and is also independent of (pre)language abilities in infancy further highlights that the early-developed neural networks underlie high-order language acquisition. Our hypothesis is also consistent with Dehaene-Lambertz and colleagues' (76) recent proposal of an early symbolic system associated with speech/language learning observed in infants from the first months of life onward. In summary, the current finding provides empirical evidence of the far-reaching developmental significance of this early-developing functional network on the trajectory of language acquisition. Future work is warranted to further investigate the network characteristics and specific roles of their constituent brain regions in supporting this complex learning process.

We further identify prospective associations between pre-literacy skill, i.e., phonological processing and the infant FCF patterns of the left fusiform gyrus and right inferior temporal cortex, regions relevant for literacy and language processing. The left fusiform gyrus is central to literacy development. It hosts the visual word form area, which is developed after the onset of learning to read and underlies the orthographic processing of words (77). Moreover, the fusiform gyrus is also activated during phonological processing (e.g., (78) and its neural activation pattern encodes both phonological and orthographic information of words (79). Here, it is demonstrated that functional connectivity patterns of the left fusiform gyrus established with the first year of life are prospectively associated with the development of phonological skills in school-age children, suggesting a functional role in the development of phonological and orthographic processing. Moreover, the bilateral inferior temporal gyri (and fusiform to a lesser extent) are

also important components of the ventral stream within the neuroanatomy of speech processing, which serve as an interface linking speech code (i.e., phonological representation) with the corresponding semantic information (80, 81). Brain lesion sizes to this area (inferior temporal and fusiform gyri) have been shown to associate with the auditory word comprehension performance among stroke patients, even after controlling for the conceptual knowledge as indexed by performance in object recognition (82). Greater activation of this area has been observed when participants listened to semantically-ambiguous sentences compared to unambiguous sentences (83, 84). Therefore, the observed relationship between the functional network of the ventral speech processing pathway and phonological development is consistent with the longitudinal associations between the development of speech perception and oral language comprehension at infant and toddler stages and literacy skills at school-age (44, 85-87). Lastly, it is also interesting to note that rapid automatized naming (RAN) abilities measured at school-age failed to be reliably associated with FCF of any temporal region. The RAN assessment measures the efficiency of integrating multiple language-related (e.g., access and retrieval of phonological and semantic information) and general cognitive (e.g., attention, vision, and motor planning) processes (54). Therefore, the null results might indicate that the RAN skills reflecting the integration abilities might be minimally associated with early-emerged functional networks in infancy.

Broadly speaking, our study illustrates the application of infant resting-state imaging techniques along with a longitudinal design to characterize the initial “blueprints” of the brain’s functional organization that support the protracted development of complex cognitive function. Functional network architecture bridging brain structure and function undergoes transformative changes after birth, especially among connectivity profiles of association regions responsible for

high-order functions (5-8). Echoing this maturational trajectory, complex cognitive abilities that emerge during the critical period of infancy are generally thought to have a protracted and dynamic progression that is sensitive to both genetic and environmental influences (88). Surveying the relationship between infant brain functional topologies and the subsequent development of cognitive abilities provides a framework for characterizing the cortical scaffold upon which cognitive abilities emerge. The acquisition of resting-state functional neuroimaging, requiring no active task engagement, thus presents a feasible solution to characterizing brain functional organization in infancy. Combined with a longitudinal design, this approach provides a valuable opportunity to identify early brain functional characteristics that could account for individual variation in subsequent behavioral development, thus highlighting the potential of developing imaging markers to identify at-risk populations for early intervention (64).

Our study represents the first evidence of prospective associations between infant functional topologies and long-term development of language and foundational literacy skills measured at school-age. However, the results should be interpreted with some caution. The infants investigated in the current study had a wide age range at our initial scan. Due to challenges associated with the recruitment and retention of a longitudinal infant MRI cohort, participants with a wide range of age (4-18 months old) at the initial scan were included. This imposes challenges for both data processing and result interpretation. Based on the mean age of the infant sample (10.2 ± 3.5 months), the one-year-old AAL atlas was utilized in the current study (58). Due to the significant growth and myelination of the human brain within the first two years of life (e.g., (89-91), applying one atlas to an infant sample of 4-18 months old could potentially introduce registration errors. To minimize this risk, visual inspection was conducted to ensure the correct mapping of every cortical region to the structural image of each infant.

Moreover, despite the risk of overlooking functional subdivisions within each cortical region, the current approach of adopting the mean time series of cortical regions defined by anatomical parcellation is more resistant to registration errors than a more fine-grained parcellation derived from a different age group. Ultimately, the development of age- /developmental stage- specific atlases is needed and will be crucial to better characterize and understand the functional network mechanisms underlying cognitive development in infants (see (92, 93) for reviews). Furthermore, several approaches were adopted to minimize the potential confounding effects of age-at-scan, including modeling age-at-scan and the infant's language environments as covariates in the main analyses, performing conformational analyses in a subgroup of younger infants, and directly evaluating the contribution of early foundational language skills on observed associations. Nevertheless, the current project design limits a clear distinction between neural mechanisms established at birth and contributions of postnatal growth/experiences. Future longitudinal investigation starting in neonates, along with behavioral and neural characterization at multiple developmental stages in a longitudinal design, are needed to unravel neural trajectories underlying language and literacy development and key components that shape this process.

Conclusion.

Utilizing a five-year longitudinal neuroimaging project starting in infancy, the current study provides the first empirical evidence of prospective associations between functional connectivity fingerprints in infants and the development of oral language and phonological skills in school-age children. Moreover, we further show that these longitudinal associations with school-age language and pre-literacy skills are not explained by foundational language skills in infancy, suggesting a unique role of functional networks in the development of language and foundational literacy abilities. Overall, this body of work reveals early-emerged functional

network mechanisms in infancy that underlie the long-term development of complex cognitive abilities with a protracted neuro-trajectory that starts early in life or even in utero.

Data availability.

All documentation and code used in these analyses can be shared with any research team. Due to Institutional Review Board regulations at Boston Children's Hospital at the time of consent, our data cannot presently be uploaded to a permanent third-party archive. However, data sharing can be initiated through a Data Usage Agreement. We are strongly committed to working closely on such a sharing plan with anyone who would like to re-analyze our data and will provide templates and procedural outlines.

Methods

Participants and study overview

All infants were recruited at Boston Children's Hospital (BCH) as part of an ongoing longitudinal project which aims to characterize brain changes underlying language and reading development from infancy to school-age. The current protocol was reviewed and approved by the institutional review board at BCH. Before participation, informed written consent was obtained from a parent or legal guardian of each infant.

A total of 60 infants (4-18 months) successfully completed resting-state functional and structural MR imaging between the years of 2011 and 2015. Among them, six subjects were excluded from further analysis due to poor image quality (n = 2) or excessive motion (n = 4, see details below). Among 54 subjects with usable imaging data, 42 (21 females, 10.2 ± 3.5 months) were successfully followed over five years (5.6 ± 1.0 years) until school-age (mean age = 6.5 ±

0.96 years). Infant participants who were successfully followed (n=42) and those were not (n=12) did not differ significantly by age, gender, head movement during scanning, emergent language skills in infancy, or environmental measures (all $p_{uncorrected} > 0.05$, table S6). The longitudinal sample, characteristically representing the whole sample, was thus included in the current prospective association analyses. All 42 children were from English-speaking households and were exposed to the English language since birth. They further demonstrated typical general cognitive ability based on their performance (standard scores (SS) > 85, mean SS = 107 ± 13) on the matrices subtest of the Kaufman Brief Intelligence Test-II at school-age (KBIT-II, Kaufman & Kaufman, 2004. Moreover, none of the subjects had birth complications, neurological trauma, or reported developmental delay. All subjects had MRIs that were clinically interpreted as normal by a pediatric neuroradiologist at BCH.

Psychometric and environmental measures

At the infant stage, emergent receptive and expressive languages were evaluated using the Mullen Scales of Early Learning assessment (MSEL, (94) in 29 infants. Raw scores were converted into T scores with a mean of 50 and SD of 10 for the subsequent analyses. Furthermore, environmental factors critical for language and literacy development were also collected through parental questionnaires, which included socio-economic status (SES, i.e., parental education, n = 41) and home literacy environment (HLE, adapted from (95), n = 38, see SI Methods).

At the school-age stage, a comprehensive assessment battery was administered to evaluate each child's general cognitive ability, language, literacy, reading, and spelling skills. Since most of the child participants had not yet developed reading and spelling skills during the second longitudinal timepoint (mean age = 6.5 ± 0.96 years), the present study focused on language and

foundational literacy skills that could be reliably measured within this age range. This included assessments of oral language (OL, the OL subtest of the Woodcock-Johnson-IV assessment (WJ-IV, (96), rapid automatized naming (RAN, Comprehensive Test of Phonological Processing-II (CTOPP-2), (97) and phonological processing skills (the phonological processing subtest of WJ-IV, see detailed descriptions of assessments in the SI Methods).

Image acquisition

All images were acquired during natural sleep without sedation on a Siemens 3T Trio MRI scanner with a 32-channel adult head coil (see (98) for a detailed infant imaging procedure). A motion-compensated T1 weighted multi-echo MPRAGE sequence was applied to collect high-quality structural images for each infant (slice number = 128, TR = 2520 ms, flip angle = 7°, field of view = 192 mm^2, voxel size =1 x 1 x 1 mm^3). Moreover, an 8-minute EPI-BOLD sequence was applied to collect resting-state fMRI data for every infant (TR = 3000 ms, TE = 30 ms, flip angle = 60°, voxel size =3 x 3 x 3 mm^3).

Image analysis

a. Resting-state fMRI preprocessing

Functional MRI images were first corrected for slice timing using FSL/SLICETIMER (and an in-house function for images with multi-slice acquisition). Rigid body alignment for motion correction and estimation of motion regressors was performed using FSL/MCFLIRT. Anatomical and functional MRI volumes were spatially normalized to the UNC 1-year old infant template (58) using affine transformations (FSL/FLIRT). After spatial normalization, linear regression was performed using the following nuisance regressors: (1) six motion regressors from realignment estimates (accounting for translations and rotations in three directions), (2) mean cerebral spinal fluid (CSF) and white matter (WM) signals extracted from subject-specific

anatomical masks (computed with FSL/FAST), and (3) spikes regressors based on motion outliers (see §Artifact detection). Temporal band-pass filtering (0.01-0.1Hz) and spatial smoothing (Gaussian filter, FWHM = 6 mm) were applied after nuisance regression.

Artifact detection

An in-house script was used to identify artifacts in the fMRI data, including motion and spiking (https://github.com/xiyu-bnu/infant_restingstate_prediction). Framewise displacement (FD) was estimated from translations and rotations obtained from rigid-body alignment after motion correction (99). Volumes with FD > 0.3 mm were considered outliers. In order to minimize potential artifacts in neighboring volumes due to temporal filtering or spin history, one preceding and two subsequent frames were also marked as outliers. A binary vector indexing all the outliers was built for each subject and used as a motion regressor to remove contaminated frames from the time series prior to filtering. Only subjects with more than 5-minute usable volumes were considered for further analysis.

b. Seed-based connectivity analysis

Seed regions were based on 78 cortical regions of interest (ROIs) defined by the 1-year-old Infant Brain Atlases (58) derived from the Automated Anatomical Labeling (AAL) atlas (100). Seed-based time series were extracted by averaging the BOLD signal across all the voxels in each ROI. The functional connectivity fingerprint (FCF) pattern associated with each seed ROI was generated by correlating the BOLD time series of the ROI with that of all other ROIs using the Pearson product-moment correlation formula. Correlation coefficients were normalized to z-scores using Fisher's transformation.

c. Prediction and classification

A Support Vector Regression (SVR) classifier was applied for the prospective association analyses using the LIBSVM package (101). Relationships between the FCF pattern of each anatomical region in the temporal lobe and the school-age performance of each assessment (i.e., OL, phonological processing, and RAN) were examined. A cross-validation approach was applied in the SVR analyses, in which 20% of the datasets were held out (testing dataset) while the remaining datasets were used to train an SVR model. Moreover, to minimize potential confounding effects on the association analysis, task performance (standardized scores) in the training dataset was first fit to a linear regression model with environmental factors (parental education level and HLE), age at the infant scan, and general cognitive ability at school-age as independent variables. The residual values of each regression were then used as outcome measures in the SVR analyses. The regression model was further applied to the testing dataset in order to obtain the residual values corresponding to the task performance for the testing participants. The trained SVR model was then applied to the testing sample to produce the predicted values, which were further correlated (Spearman) with the real performance (residual values) as the SVR performance. This process was repeated 1000 times to generate the true distributions of the SVR performance. The distribution of the prediction performance was tested against a null or random distribution, which was generated following the same procedure except that the test performance used during training was randomly permuted in each iteration. Statistical significance against the null distribution was tested with a Kolmogorov-Smirnov test. In addition, the effect size between the true and null distributions was measured using Cohen's *d* measure (62), where larger magnitudes indicate higher predictive power of the infant FCFs on school-age performance when compared to the null distribution. Only associations significantly (corrected for the comparison comparisons) outperforming the null distribution with a large

effect size (Cohen's $d > 0.8$, (62) were considered in the current analyses (see Figure 1 for a general pipeline of the prediction analyses).

References

1. Gratton C, *et al.* (2018) Functional brain networks are dominated by stable group and individual factors, not cognitive or daily variation. *Neuron* 98(2):439-452. e435.
2. Biswal B, Zerrin Yetkin F, Haughton VM, & Hyde JS (1995) Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. *Magnetic resonance in medicine* 34(4):537-541.
3. Doria V, *et al.* (2010) Emergence of resting state networks in the preterm human brain. *Proceedings of the National Academy of Sciences* 107(46):20015-20020.
4. Fransson P, *et al.* (2007) Resting-state networks in the infant brain. *Proceedings of the National Academy of Sciences* 104(39):15531-15536.
5. Gao W, *et al.* (2015) Functional network development during the first year: relative sequence and socioeconomic correlations. *Cerebral cortex* 25(9):2919-2928.
6. Gao W, Alcauter S, Smith JK, Gilmore JH, & Lin W (2015) Development of human brain cortical network architecture during infancy. *Brain Structure and Function* 220(2):1173-1186.
7. Fair DA, *et al.* (2009) Functional brain networks develop from a "local to distributed" organization. *PLoS comput biol* 5(5):e1000381.
8. Gu S, *et al.* (2015) Emergence of system roles in normative neurodevelopment. *Proceedings of the National Academy of Sciences* 112(44):13681-13686.
9. Alcauter S, *et al.* (2014) Development of thalamocortical connectivity during infancy and its cognitive correlations. *Journal of Neuroscience* 34(27):9067-9075.
10. Chen Y, *et al.* (2020) The Subgrouping Structure of Newborns with Heterogenous Brain–Behavior Relationships. *Cerebral Cortex.*
11. Rogers CE, *et al.* (2017) Neonatal amygdala functional connectivity at rest in healthy and preterm infants and early internalizing symptoms. *Journal of the American Academy of Child & Adolescent Psychiatry* 56(2):157-166.
12. Skeide MA & Friederici AD (2016) The ontogeny of the cortical language network. *Nature Reviews Neuroscience* 17(5):323-332.
13. Moon C, Cooper RP, & Fifer WP (1993) Two-day-olds prefer their native language. *Infant behavior and development* 16(4):495-500.
14. Vouloumanos A & Werker JF (2007) Listening to language at birth: Evidence for a bias for speech in neonates. *Developmental science* 10(2):159-164.
15. Eimas PD, Siqueland ER, Jusczyk P, & Vigorito J (1971) Speech perception in infants. *Science* 171(3968):303-306.
16. Kuhl PK (2004) Early language acquisition: cracking the speech code. *Nature reviews neuroscience* 5(11):831-843.
17. Sansavini A, Bertoncini J, & Giovanelli G (1997) Newborns discriminate the rhythm of multisyllabic stressed words. *Developmental Psychology* 33(1):3.
18. Marchman VA & Fernald A (2008) Speed of word recognition and vocabulary knowledge in infancy predict cognitive and language outcomes in later childhood. *Developmental science* 11(3):F9-F16.
19. Newman R, Ratner NB, Jusczyk AM, Jusczyk PW, & Dow KA (2006) Infants' early ability to segment the conversational speech signal predicts later language development: A retrospective analysis. *Developmental psychology* 42(4):643.

20. Swingley D (2009) Contributions of infant word learning to language development. *Philosophical Transactions of the Royal Society B: Biological Sciences* 364(1536):3617-3632.
21. Fernald A, Perfors A, & Marchman VA (2006) Picking up speed in understanding: Speech processing efficiency and vocabulary growth across the 2nd year. *Developmental psychology* 42(1):98.
22. Duff FJ, Reen G, Plunkett K, & Nation K (2015) Do infant vocabulary skills predict school-age language and literacy outcomes? *Journal of Child Psychology and Psychiatry* 56(8):848-856.
23. Pierce LJ, Genesee F, Delcenserie A, & Morgan G (2017) Variations in phonological working memory: Linking early language experiences and language learning outcomes. *Applied Psycholinguistics* 38(6):1265-1300.
24. Scarborough HS, Neuman S, & Dickinson D (2009) Connecting early language and literacy to later reading (dis) abilities: Evidence, theory, and practice. *Approaching difficulties in literacy development: Assessment, pedagogy and programmes* 10:23-38.
25. Huttenlocher J, Haight W, Bryk A, Seltzer M, & Lyons T (1991) Early vocabulary growth: relation to language input and gender. *Developmental psychology* 27(2):236.
26. Romeo RR, *et al*. (2018) Beyond the 30-million-word gap: Children's conversational exposure is associated with language-related brain function. *Psychological science* 29(5):700-710.
27. Rowe ML (2012) A longitudinal investigation of the role of quantity and quality of child-directed speech in vocabulary development. *Child development* 83(5):1762-1774.
28. Ece Demir-Lira Ö, Applebaum LR, Goldin-Meadow S, & Levine SC (2019) Parents' early book reading to children: Relation to children's later language and literacy outcomes controlling for other parent language input. *Developmental science* 22(3):e12764.
29. Evans MA, Shaw D, & Bell M (2000) Home literacy activities and their influence on early literacy skills. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale* 54(2):65.
30. Hood M, Conlon E, & Andrews G (2008) Preschool home literacy practices and children's literacy development: A longitudinal analysis. *Journal of Educational Psychology* 100(2):252.
31. Sénéchal M (2006) Testing the home literacy model: Parent involvement in kindergarten is differentially related to grade 4 reading comprehension, fluency, spelling, and reading for pleasure. *Scientific studies of reading* 10(1):59-87.
32. Cunningham AE & Stanovich KE (1997) Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental psychology* 33(6):934.
33. Hatcher PJ, Hulme C, & Snowling MJ (2004) Explicit phoneme training combined with phonic reading instruction helps young children at risk of reading failure. *Journal of child Psychology and Psychiatry* 45(2):338-358.
34. Nation K & Hulme C (2011) Learning to read changes children's phonological skills: Evidence from a latent variable longitudinal study of reading and nonword repetition. *Developmental science* 14(4):649-659.
35. Piasta SB & Wagner RK (2010) Developing early literacy skills: A meta-analysis of alphabet learning and instruction. *Reading research quarterly* 45(1):8-38.

36. Dehaene-Lambertz G & Pena M (2001) Electrophysiological evidence for automatic phonetic processing in neonates. *Neuroreport* 12(14):3155-3158.
37. Mahmoudzadeh M, *et al.* (2013) Syllabic discrimination in premature human infants prior to complete formation of cortical layers. *Proceedings of the National Academy of Sciences* 110(12):4846-4851.
38. Dehaene-Lambertz G, Dehaene S, & Hertz-Pannier L (2002) Functional neuroimaging of speech perception in infants. *science* 298(5600):2013-2015.
39. Dehaene-Lambertz G, *et al.* (2006) Functional organization of perisylvian activation during presentation of sentences in preverbal infants. *Proceedings of the National Academy of Sciences* 103(38):14240-14245.
40. Pena M, *et al.* (2003) Sounds and silence: an optical topography study of language recognition at birth. *Proceedings of the National Academy of Sciences* 100(20):11702-11705.
41. Perani D, *et al.* (2011) Neural language networks at birth. *Proceedings of the National Academy of Sciences of the United States of America* 108(45):18566.
42. Choudhury N & Benasich AA (2011) Maturation of auditory evoked potentials from 6 to 48 months: prediction to 3 and 4 year language and cognitive abilities. *Clinical Neurophysiology* 122(2):320-338.
43. Guttorm TK, *et al.* (2005) Brain event-related potentials (ERPs) measured at birth predict later language development in children with and without familial risk for dyslexia. *Cortex* 41(3):291-303.
44. Guttorm TK, Leppänen PH, Richardson U, & Lyytinen H (2001) Event-related potentials and consonant differentiation in newborns with familial risk for dyslexia. *Journal of Learning Disabilities* 34(6):534-544.
45. Leppänen P, *et al.* (2012) Infant brain responses associated with reading-related skills before school and at school age. *Neurophysiologie Clinique/Clinical Neurophysiology* 42(1-2):35-41.
46. Leppänen PH, *et al.* (2010) Newborn brain event-related potentials revealing atypical processing of sound frequency and the subsequent association with later literacy skills in children with familial dyslexia. *Cortex* 46(10):1362-1376.
47. Molfese DL (2000) Predicting dyslexia at 8 years of age using neonatal brain responses. *Brain and language* 72(3):238-245.
48. Van Zuijen TL, *et al.* (2012) Temporal auditory processing at 17 months of age is associated with preliterate language comprehension and later word reading fluency: an ERP study. *Neuroscience letters* 528(1):31-35.
49. Deoni SC, *et al.* (2016) White matter maturation profiles through early childhood predict general cognitive ability. *Brain Structure and Function* 221(2):1189-1203.
50. Vassar R, *et al.* (2020) Neonatal brain microstructure and machine-learning-based prediction of early language development in children born very preterm. *Pediatric Neurology*.
51. Emerson RW, Gao W, & Lin W (2016) Longitudinal study of the emerging functional connectivity asymmetry of primary language regions during infancy. *Journal of Neuroscience* 36(42):10883-10892.
52. Hulme C & Snowling MJ (2013) Learning to read: What we know and what we need to understand better. *Child development perspectives* 7(1):1-5.

53. Melby-Lervåg M, Lyster S-AH, & Hulme C (2012) Phonological skills and their role in learning to read: a meta-analytic review. *Psychological Bulletin* 138(2):322.
54. Norton ES & Wolf M (2012) Rapid automatized naming (RAN) and reading fluency: Implications for understanding and treatment of reading disabilities. *Annual review of psychology* 63:427-452.
55. Cohen AL, *et al.* (2008) Defining functional areas in individual human brains using resting functional connectivity MRI. *Neuroimage* 41(1):45-57.
56. Gordon EM, *et al.* (2016) Generation and evaluation of a cortical area parcellation from resting-state correlations. *Cerebral cortex* 26(1):288-303.
57. Mars RB, Passingham RE, & Jbabdi S (2018) Connectivity fingerprints: from areal descriptions to abstract spaces. *Trends in cognitive sciences* 22(11):1026-1037.
58. Shi F, *et al.* (2011) Infant brain atlases from neonates to 1-and 2-year-olds. *PloS one* 6(4):e18746.
59. Sato H, *et al.* (2012) Cerebral hemodynamics in newborn infants exposed to speech sounds: A whole-head optical topography study. *Human brain mapping* 33(9):2092-2103.
60. Shultz S, Vouloumanos A, Bennett RH, & Pelphrey K (2014) Neural specialization for speech in the first months of life. *Developmental Science* 17(5):766-774.
61. Telkemeyer S, *et al.* (2009) Sensitivity of newborn auditory cortex to the temporal structure of sounds. *Journal of Neuroscience* 29(47):14726-14733.
62. Cohen J (2013) *Statistical power analysis for the behavioral sciences* (Academic press).
63. Rudolph MD, *et al.* (2018) Maternal IL-6 during pregnancy can be estimated from newborn brain connectivity and predicts future working memory in offspring. *Nature neuroscience* 21(5):765-772.
64. Gilmore JH, Knickmeyer RC, & Gao W (2018) Imaging structural and functional brain development in early childhood. *Nature Reviews Neuroscience* 19(3):123.
65. Cristia A, Seidl A, Junge C, Soderstrom M, & Hagoort P (2014) Predicting individual variation in language from infant speech perception measures. *Child development* 85(4):1330-1345.
66. Johnson MH (2001) Functional brain development in humans. *Nature Reviews Neuroscience* 2(7):475-483.
67. Cole MW, Ito T, Bassett DS, & Schultz DH (2016) Activity flow over resting-state networks shapes cognitive task activations. *Nature neuroscience* 19(12):1718-1726.
68. Johansen-Berg H, *et al.* (2004) Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. *Proceedings of the National Academy of Sciences* 101(36):13335-13340.
69. Saygin ZM, *et al.* (2016) Connectivity precedes function in the development of the visual word form area. *Nature neuroscience* 19(9):1250.
70. Lambon Ralph MA, Pobric G, & Jefferies E (2009) Conceptual knowledge is underpinned by the temporal pole bilaterally: convergent evidence from rTMS. *Cerebral Cortex* 19(4):832-838.
71. Striem-Amit E, Wang X, Bi Y, & Caramazza A (2018) Neural representation of visual concepts in people born blind. *Nature communications* 9(1):1-12.
72. Visser M, Jefferies E, & Lambon Ralph M (2010) Semantic processing in the anterior temporal lobes: a meta-analysis of the functional neuroimaging literature. *Journal of cognitive neuroscience* 22(6):1083-1094.

73. Wang X, Men W, Gao J, Caramazza A, & Bi Y (2020) Two Forms of Knowledge Representations in the Human Brain. *Neuron*.
74. Patterson K, Nestor PJ, & Rogers TT (2007) Where do you know what you know? The representation of semantic knowledge in the human brain. *Nature reviews neuroscience* 8(12):976-987.
75. Schwartz MF, *et al*. (2009) Anterior temporal involvement in semantic word retrieval: voxel-based lesion-symptom mapping evidence from aphasia. *Brain* 132(12):3411-3427.
76. Dehaene-Lambertz G, Fló A, & Peña M (2020) Infants' early competence for language and symbols. in *The social brain: a developmental perspective*.
77. Dehaene S & Cohen L (2011) The unique role of the visual word form area in reading. *Trends in cognitive sciences* 15(6):254-262.
78. Dietz NA, Jones KM, Gareau L, Zeffiro TA, & Eden GF (2005) Phonological decoding involves left posterior fusiform gyrus. *Human brain mapping* 26(2):81-93.
79. Zhao L, *et al*. (2017) Orthographic and phonological representations in the fusiform cortex. *Cerebral Cortex* 27(11):5197-5210.
80. Hickok G & Poeppel D (2004) Dorsal and ventral streams: a framework for understanding aspects of the functional anatomy of language. *Cognition* 92(1-2):67-99.
81. Hickok G & Poeppel D (2007) The cortical organization of speech processing. *Nature reviews neuroscience* 8(5):393-402.
82. Bonilha L, *et al.* (2017) Temporal lobe networks supporting the comprehension of spoken words. *Brain* 140(9):2370-2380.
83. Rodd JM, Davis MH, & Johnsrude IS (2005) The neural mechanisms of speech comprehension: fMRI studies of semantic ambiguity. *Cerebral Cortex* 15(8):1261-1269.
84. Rodd JM, Johnsrude IS, & Davis MH (2012) Dissociating frontotemporal contributions to semantic ambiguity resolution in spoken sentences. *Cerebral Cortex* 22(8):1761-1773.
85. Flax JF, Realpe-Bonilla T, Roesler C, Choudhury N, & Benasich A (2009) Using early standardized language measures to predict later language and early reading outcomes in children at high risk for language-learning impairments. *Journal of learning disabilities* 42(1):61-75.
86. Puolakanaho A, Poikkeus A-M, Ahonen T, Tolvanen A, & Lyytinen H (2004) Emerging phonological awareness differentiates children with and without familial risk for dyslexia after controlling for general language skills. *Annals of Dyslexia* 54(2):221-243.
87. Torppa M, Lyytinen P, Erskine J, Eklund K, & Lyytinen H (2010) Language development, literacy skills, and predictive connections to reading in Finnish children with and without familial risk for dyslexia. *Journal of learning disabilities* 43(4):308-321.
88. Karmiloff-Smith A, *et al.* (2012) Genetic and environmental vulnerabilities in children with neurodevelopmental disorders. *Proceedings of the National Academy of Sciences* 109(Supplement 2):17261-17265.
89. Deoni SC, Dean III DC, O'Muircheartaigh J, Dirks H, & Jerskey BA (2012) Investigating white matter development in infancy and early childhood using myelin water faction and relaxation time mapping. *Neuroimage* 63(3):1038-1053.
90. Groeschel S, Vollmer B, King M, & Connelly A (2010) Developmental changes in cerebral grey and white matter volume from infancy to adulthood. *International Journal of Developmental Neuroscience* 28(6):481-489.
91. Sanchez CE, Richards JE, & Almli CR (2012) Neurodevelopmental MRI brain templates for children from 2 weeks to 4 years of age. *Developmental psychobiology* 54(1):77-91.

92. Turesky T, Vanderauwera J, & Gaab N (2020) Imaging the rapidly developing brain: Current challenges for MRI studies in the first five years of life.
93. Zhang H, Shen D, & Lin W (2019) Resting-state functional MRI studies on infant brains: A decade of gap-filling efforts. *Neuroimage* 185:664-684.
94. Mullen EM (1995) *Mullen scales of early learning* (AGS Circle Pines, MN).
95. Denney MK, English JP, Gerber MM, Leafstedt J, & Ruz ML (2001) Family and Home Literacy Practices: Mediating Factors for Preliterate English Learners at Risk.
96. Schrank FA & Wendling BJ (2015) Woodcock-Johnson® IV Tests of Early Cognitive and Academic Development.
97. Wagner R, Torgesen J, Rashotte C, & Pearson NA (1999) *CTOPP-2: Comprehensive Test of Phonological Processing–Second Edition* (Pro-ed Austin, TX).
98. Raschle N, *et al.* (2012) Pediatric neuroimaging in early childhood and infancy: challenges and practical guidelines. *Annals of the New York Academy of sciences* 1252(1):43-50.
99. Power JD, Barnes KA, Snyder AZ, Schlaggar BL, & Petersen SE (2012) Spurious but systematic correlations in functional connectivity MRI networks arise from subject motion. *Neuroimage* 59(3):2142-2154.
100. Tzourio-Mazoyer N, *et al.* (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. *Neuroimage* 15(1):273-289.
101. Chang C-C & Lin C-J (2011) LIBSVM: A library for support vector machines. *ACM transactions on intelligent systems and technology (TIST)* 2(3):1-27.

Table 1. Longitudinal associations between infant functional connectivity fingerprints and school-age performance

FCF seed region	School-age performance	SVR performance		Comparisons against null distribution	
		mean *r*	S.D.	$p_{corrected}$	Effect size (Cohen's *d*)
midLTPO	Oral language	0.42	0.36	$8.5*10^{-78}$	1.1
LFFG	Phonological processing	0.40	0.37	$2.7*10^{-62}$	0.96
RFFG	Phonological processing	0.31	0.39	$3.3*10^{-48}$	0.82
RITG	Phonological processing	0.41	0.36	$1.2*10^{-60}$	0.94

Note. SVR = Support vector regression; midLTPO = Left temporal pole (the middle portion); LFFG = Left fusiform gyrus; RFFG = Right fusiform gyrus; RITG = Right inferior temporal gyrus.

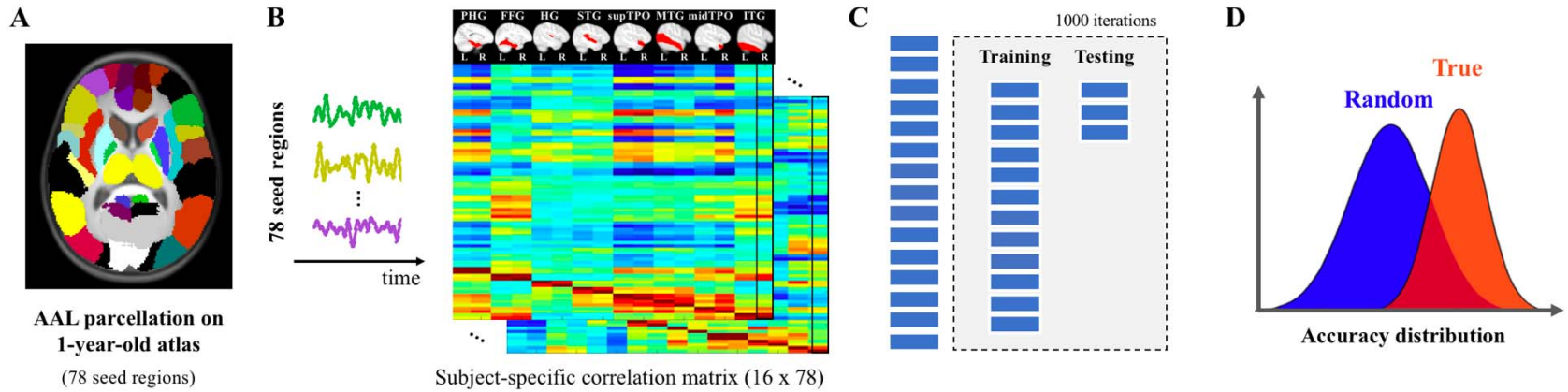


Figure 1. General pipeline of the prediction and classification analyses. **A.** Seed regions were defined as 78 cortical parcellations derived from the 1-year-old infant AAL atlas (Shi et al., 2011). **B.** Time series were extracted for every participant by averaging the BOLD signal across all the voxels within each seed region. The functional connectivity fingerprints (FCF) were obtained by pair-wise Pearson's correlations. The current analyses focused on 16 seed regions residing in the bilateral temporal lobes given the previously reported engagement of this region in speech and language processing in infancy (Dehaene-Lambertz, Dehaene, & Hertz-Pannier, 2002; Dehaene-Lambertz et al., 2006; Sato et al., 2012; Shultz, Vouloumanos, Bennett, & Pelphrey, 2014; Telkemeyer et al., 2009). Thus, a 16 x 78 correlation matrix was generated for each subject, where each column represents an FCF pattern associated with a specific seed region in the temporal lobe as shown in the top sagittal views. **C.** Prediction analyses were performed to identify regions whose FCFs in infancy were prospectively associated with the long-term development of language and pre-literacy skills at school-age. Each FCF pattern was entered into a Support Vector Regression (SVR) model as features for the prediction analyses on every school-age performance. To objectively evaluate the prediction performance, the whole sample was randomly divided into training (80%) and testing (20%) datasets. The task performance (standard scores) from the training set was first submitted to a linear regression model with environmental measures (SES and HLE), general cognitive skill at school-age, and age in months at the infant imaging time point as independent regressors. The residuals of the models, representing individual differences for each skill after controlling for environmental and general cognitive influences, were then entered into the SVR analyses. The trained SVR model was then applied to the FCF patterns of the testing datasets, which generated the predicted task performance. The predicted values were then correlated with the true performance (i.e., residual values derived from the linear regression model based on the standard scores of the testing set),

and the correlation coefficient was considered as performance of the SVR model. This procedure was repeated 1000 times to produce a distribution of prediction performance for each pair of seed region and task performance. **D.** The distribution of prediction performance based on the true language and literacy outcomes was compared against a null distribution generated by the same procedure, but assigning random scores of school-age performance (permutation tests) using a Kolmogorov-Smirnov test. An effect size was further calculated for the prediction performance of each seed region.

Abbreviations: PHG = ParaHippocampal gyrus; FFG = Fusiform gyrus; HG = Heschl's gyrus; STG = Superior temporal gyrus; supTPO = Temporal pole (the superior portion); MTG = Middle temporal gyrus; midTPO = Temporal pole (the middle portion); ITG = Inferior temporal gyrus.

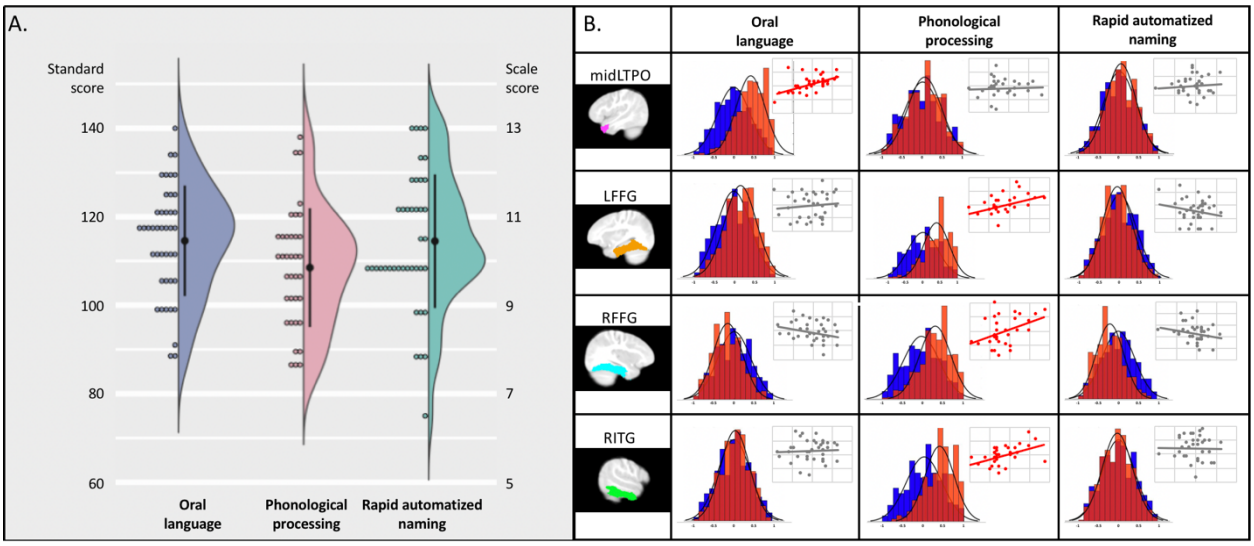
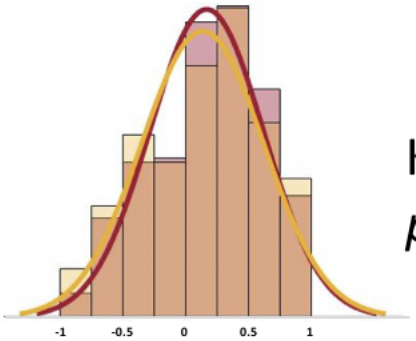


Figure 2. Prospective associations between functional connectivity fingerprints (FCF) in infancy and language and foundational literacy development at school-age. **A.** Distribution of the oral language (standard scores), phonological processing (standard scores), and rapid automatized naming (scale scores) skills measured at school-age. **B.** The association results between school-age skill performance and each of the infant FCFs of the four identified seed regions, including the left temporal pole (the middle portion, midLTPO), left fusiform gyrus (LFFG), right fusiform gyrus (RFFG) and right inferior temporal gyrus (RITG). The distribution maps for each pair show the performances of the true prediction analyses (red) against the null distribution (blue) based on permutation results. The identified longitudinal associations were further revaluated using a leave-one-out cross-validation approach where one subject's dataset was selected for testing while the remaining subjects were applied to train the support vector regression model. The obtained SVR results are presented as a scatter plot in the upper right corner of each cell to illustrate the fitness of true performance and predicted outcomes across the whole sample (see a detailed analysis procedure and the statistical results in SI Methods and Table S4). Significant association results (Bonferroni corrected) with a large effect size (Cohen's d $\geq$ 0.8) were highlighted in red.

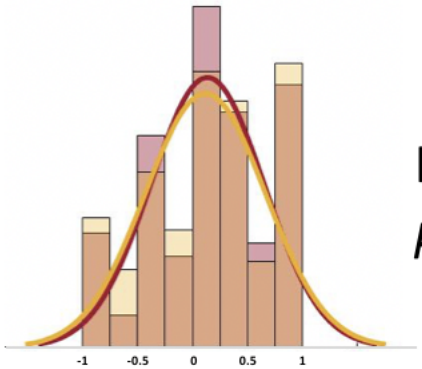
midLTPO → Oral language



KS statistics = 0.047

$p_{\text{uncorrected}} = 0.21$

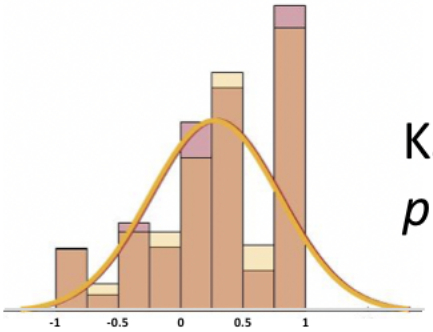
LFFG → Phonological processing



KS statistics = 0.044

$p_{\text{uncorrected}} = 0.28$

RITG → Phonological processing



KS statistics = 0.024

$p_{\text{uncorrected}} = 0.93$

Figure 3. No significant differences in SVR performance with (yellow) and without (red) infant language milestones included as control variables. To evaluate whether (pre)language skills in infancy might contribute to the identified prospective associations, we further compared two models with and without receptive and expressive language skills in infancy included as covariates (in addition to the other measures included in the previous analyses) respectively. A Kolmogorov–Smirnov (KS) test was used to compare the distributions of association strength between the two models. No significant results were obtained before correcting for multiple comparisons.
Abbreviations: midTPO = Temporal pole (the middle portion); LFFG = Left fusiform gyrus; RITG = Right inferior temporal gyrus.