



ARTICLE

Neurocognitive and functional heterogeneity in depressed youth

Erica B. Baller ¹, Antonia N. Kaczurkin ^{1,2}, Aristeidis Sotiras ^{3,4,5}, Azeez Adebimpe ¹, Danielle S. Bassett ^{1,6,7,8,9,10,11}, Monica E. Calkins ^{1,12}, Ganesh B. Chand ^{3,5}, Zaixu Cui ¹, Raquel E. Gur ^{1,3,9,12}, Ruben C. Gur ^{1,3,12}, Kristin A. Linn ¹³, Tyler M. Moore ^{1,12}, David R. Roalf ^{1,12}, Erdem Varol ^{3,5}, Daniel H. Wolf ^{1,5,12}, Cedric H. Xia ¹, Christos Davatzikos ^{3,5} and Theodore D. Satterthwaite ^{1,5,12}

Depression is a common psychiatric illness that often begins in youth, and is sometimes associated with cognitive deficits. However, there is significant variability in cognitive dysfunction, likely reflecting biological heterogeneity. We sought to identify neurocognitive subtypes and their neurofunctional signatures in a large cross-sectional sample of depressed youth. Participants were drawn from the Philadelphia Neurodevelopmental Cohort, including 712 youth with a lifetime history of a major depressive episode and 712 typically developing (TD) youth matched on age and sex. A subset (MDD $n = 368$, TD $n = 200$) also completed neuroimaging. Cognition was assessed with the Penn Computerized Neurocognitive Battery. A recently developed semi-supervised machine learning algorithm was used to delineate neurocognitive subtypes. Subtypes were evaluated for differences in both clinical psychopathology and brain activation during an *n*-back working memory fMRI task. We identified three neurocognitive subtypes in the depressed group. Subtype 1 was high-performing (high accuracy, moderate speed), Subtype 2 was cognitively impaired (low accuracy, slow speed), and Subtype 3 was impulsive (low accuracy, fast speed). While subtypes did not differ in clinical psychopathology, they diverged in their activation profiles in regions critical for executive function, which mirrored differences in cognition. Taken together, these data suggest disparate mechanisms of cognitive vulnerability and resilience in depressed youth, which may inform the identification of biomarkers for prognosis and treatment response.

Neuropsychopharmacology (2021) 46:783–790; <https://doi.org/10.1038/s41386-020-00871-w>

INTRODUCTION

Depressive disorders are common, and are consistently ranked among the leading causes of disability world-wide [1, 2]. Major depressive disorder (MDD) is often difficult to treat, with one-third of patients remaining symptomatic despite treatment [3, 4]. Depression frequently starts in adolescence, with a 3–5% prevalence of MDD in youth [5, 6]. Like adults, there is significant heterogeneity in response to treatment in youth with MDD [7, 8]. The high variability in treatment response suggests that MDD is a heterogeneous illness, with multiple pathophysiologic pathways that converge on a similar clinical phenotype [9–12]. However, the Diagnostic and Statistical Manual of Mental Health Disorders continues to rely solely on clinical symptom classification [13].

Although mood symptoms define MDD, deficits in cognition are consistently reported in studies of depressed youth [14–16]. There is also substantial cognitive heterogeneity in depression among youth; some experience profound mood symptoms with cognitive resilience, whereas others demonstrate marked cognitive impairment [17, 18]. Early studies suggest that cognition may have

prognostic value as well. A large longitudinal cohort study demonstrated that baseline neuropsychological profiles best predicted functional outcomes in depressed youth, even surpassing prediction from baseline mood symptoms alone [19]. Neurocognitive limitations have also been found to negatively impact recovery from MDD [20].

In particular among cognitive domains, executive function undergoes protracted development during adolescence, a period that coincides with increased vulnerability to mood disorders [21–23]. Networks that subservise executive functioning have emerged as important targets in the study of youth depression. However, the few neuroimaging studies that have evaluated cognitive control in depressed youth have yielded mixed results. Whereas some studies have shown less prefrontal cortex activation in depressed youth as compared to healthy controls [24, 25], other studies have shown greater activation [26, 27]. Of note, none of these studies characterized or evaluated cognitive heterogeneity, which may account for conflicting findings. Given the high degree of cognitive heterogeneity in depression and the important relationship between cognitive

¹Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; ²Department of Psychology, Vanderbilt University, Nashville, TN 37235, USA; ³Department of Radiology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA; ⁴Department of Radiology, Washington University, St. Louis, MO 63130, USA; ⁵Center for Biomedical Image Computing and Analytics, University of Pennsylvania, Philadelphia, PA 19104, USA; ⁶Department of Bioengineering, University of Pennsylvania, Philadelphia, PA 19104, USA; ⁷Penn Center for Neuroimaging and Therapeutics, University of Pennsylvania, Philadelphia, PA 19104, USA; ⁸Department of Electrical and Systems Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA; ⁹Department of Neurology, University of Pennsylvania, Philadelphia, PA 19104, USA; ¹⁰Department of Physics and Astronomy, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA; ¹¹Santa Fe Institute, Santa Fe, NM 87501, USA; ¹²Lifespan Brain Institute, Penn Medicine and Children's Hospital of Philadelphia, Philadelphia, PA 19104, USA and ¹³Department of Biostatistics, Epidemiology, and Informatics, University of Pennsylvania, Philadelphia, PA 19104, USA
Correspondence: Theodore D. Satterthwaite (sattertt@penmedicine.upenn.edu)

Received: 12 March 2020 Revised: 8 September 2020 Accepted: 10 September 2020
Published online: 2 October 2020

function and functional outcome, we sought to identify neurocognitive subtypes in youths with a history of depression.

Machine learning tools are increasingly used for uncovering more biologically homogenous subtypes within heterogeneous conditions like MDD [28]. In this study, we used a recently developed semi-supervised machine learning algorithm called Heterogeneity through Discriminative Analysis (HYDRA) [29, 30]. We then evaluated the cognitively defined subgroups on independent measures that were not used in the subtype identification process, including clinical symptoms and brain activation during an *n*-back working memory task [31]. We selected the *n*-back because it reliably recruits brain networks that are relevant for cognitive control, are developmentally sensitive, and implicated in mood disorders [32–35]. We predicted that we would identify cognitive subtypes that had distinct neural signatures that would provide information beyond the clinical symptomatology of MDD.

METHODS

Participants

The Philadelphia Neurodevelopmental Cohort (PNC), funded by the National Institute of Mental Health Grand Opportunity (GO) mechanism of the American Recovery and Reinvestment Act, was designed to characterize clinical and neurobehavioral phenotypes of genotyped youths. As previously described in two dedicated publications, a total of 9498 participants aged 8–22 years received cognitive assessment and clinical phenotyping, and a subset of 1601 youths also completed neuroimaging as part of the PNC [36, 37]. We excluded participants with missing data or those with medical disorders that could impact brain function. Assessment of lifetime psychopathology was conducted using GOASSESS, a structured screening interview based on a modified version of the K-SADS [38]. Using this instrument, 712 youth met screening criteria for a lifetime history of a major depressive episode as defined by DSM-IV-TR, and 2310 were typically developing (TD) youth with no psychiatric diagnosis [39]. The proportion of depressed youth in this sample is consistent with the general population [40]. We refer to youths with a history of a major depressive episode as *depressed youth* (DY). Given the extensive literature documenting the effects of age and sex on brain development, and the fact that youths with a lifetime history of MDD were more likely to be older and female, we selected a sample of TD youths that were matched to the DY on age and sex. This matching procedure was implemented in R using the “MatchIt” package, and yielded a final sample of 712 DYs and 712 TDs (Table 1). A subset of these youth (TD = 200, DY = 168; Table 1) also completed the *n*-back working memory task during functional magnetic resonance imaging (fMRI) and passed strict quality control criteria [41]. Our multistep matching procedure, as detailed in the Supplementary Material, ensured that the TD and MDD group were demographically matched, while preferentially including TDs who had completed neuroimaging. The institutional review boards of both the University of Pennsylvania and the Children’s Hospital of Philadelphia approved all study procedures.

Measures of clinical psychopathology

As in prior work, to provide a dimensional summary of the diverse clinical data for all participants, we used a confirmatory bifactor analysis to model four orthogonal dimensions of psychopathology (anxious-misery, psychosis, externalizing, and fear) plus a general factor, overall psychopathology [29, 41, 42]. To avoid analytic circularity, our factor analysis excluded all items from the depression section of the interview that were used as part of inclusion criteria for the DY group (see Supplementary Material). As the depression group was identified based on a lifetime history of depression irrespective of current mood state, but mood state may impact cognitive performance, participants completed the

Table 1. Sample demographics for the whole group (A) as well as for the imaging subsample (B).

	Typically developing	Depressed	<i>P</i>
(A) Demographics of whole group			
<i>n</i>	712	712	
% White	64	55	0.001
% Female	67	67	1.00
Maternal education (mean [sd])	14.93 [2.52]	14.12 [2.29]	<0.001
Age (mean [sd])	16.21 [2.91]	16.13 [2.90]	0.94
(B) Demographics of imaging subsample			
<i>n</i>	200	168	
% White	61	46.4	<0.001
% Female	59	64	0.34
Maternal education (mean [sd])	14.93 [2.58]	13.98 [2.36]	<0.001
Age (mean [sd])	16.49 [2.84]	16.87 [2.20]	0.57
Typically developing and depressed youth were matched on age and sex prior to subtyping.			

State-Trait Anxiety Inventory (STAI) during the neuroimaging session. Previous work has shown that the STAI assesses broad anxious-misery spectrum symptoms, including both anxiety and depression, rather than anxiety specifically [43–45].

Cognitive assessment

Cognition was assessed using the University of Pennsylvania Computerized Neurocognitive Battery (CNB) [46]. Twenty-six measures obtained from 14 neurocognitive tests of performance were assessed (12 for accuracy, 14 for speed). Domains included executive functioning (three tests), episodic memory (three tests), social cognition (three tests), complex reasoning (three tests), and sensorimotor speed (two tests) as detailed in the Supplementary Material. Verbal intelligence was estimated with the Wide Range Achievement Test, 4th Edition reading subscale with total subscale scores reported as T-scores (mean = 100, SD = 15) [47].

Parsing cognitive heterogeneity with semi-supervised machine learning

To identify cognitive subtypes among our sample of DY, we used a semi-supervised machine learning tool: HYDRA [29, 30]. HYDRA compares a reference group (e.g., controls) to a target group (e.g., patients) to identify *k* subtypes (clusters) within the target group [30]. In contrast to fully supervised learning techniques, which cannot distinguish between subtypes of patients, HYDRA simultaneously performs classification and clustering (Fig. 1A). Unlike unsupervised clustering techniques (such as *k*-means or community detection), the semi-supervised algorithm clusters the differences between the two groups, rather than clustering the groups themselves, thereby parsing phenotypic heterogeneity of underlying neurobiological processes. Rather than coercing participant data points into a single common discriminative pattern, HYDRA allows for the separation of distinct groups distinguished by multiple decision boundaries. The result is a data-driven approach to identifying subtypes of DY that can be further evaluated on independently measured clinical and imaging characteristics.

HYDRA was used to define cognitive subtypes using the 26 accuracy and speed measures from the cognitive battery. Given known developmental and sex differences in cognition, both age and sex were included as covariates in HYDRA. Running HYDRA on

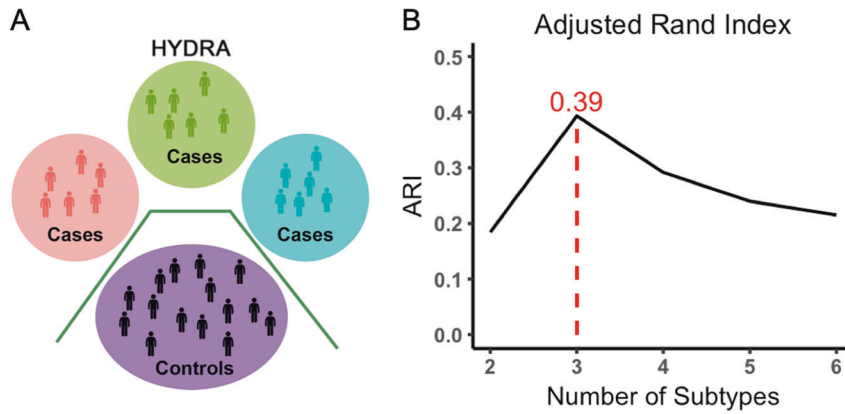


Fig. 1 Heterogeneity through Discriminative Analysis (HYDRA) algorithm and subtype selection. **A** HYDRA is a semi-supervised machine learning algorithm that reveals homogenous subtypes within a clinical group by maximizing subtype-specific margins between patient subtypes and controls, while adjusting for covariates. **B** The stability of the clustering solution after cross-validation was evaluated over a resolution range of 2–10 clusters (2–6 shown here), and was quantified by the adjusted rand index (ARI). The maximum ARI was seen with three subtypes.

the cognitive measures (as opposed to the imaging measures) allowed us to leverage the large sample size of the cognitive dataset, while using the imaging measure as an independent data type not used in clustering. Consistent with prior studies using this technique, we derived multiple clustering solutions requesting two to ten clusters in order to obtain a range of possible solutions [29, 30]. The adjusted Rand index (ARI) was calculated using tenfold cross-validation to evaluate the stability of each solution; the solution with the highest ARI value was selected for subsequent analyses. Permutation testing was used to statistically evaluate the stability of observed ARI values in comparison to a null distribution (see Supplementary Material). Clinical symptomatology and imaging data were not used for clustering, allowing them to serve as independent validators of the subtypes.

Image acquisition and processing

Task paradigm, image acquisition, and preprocessing methods are as previously detailed [41] and described in the Supplementary Material. A fractal version of the *n*-back task was used to probe working memory function. As in previous studies, we selected the 2-back versus 0-back contrast as the primary contrast of interest because it robustly indexes working memory load [32, 41, 48]. The mean percent signal change on the primary contrast of interest (2-back vs. 0-back) was extracted from 21 a priori regions of interest (ROIs) within the executive system defined in a previously published study (Supplementary Fig. 1) [32]. As prior, behavioral performance during the fMRI task was summarized using the signal detection measure *d'* [49, 50].

Group-level statistical analyses

Having identified subtypes of DY, we sought to understand the characteristics of these subtypes. As our subtypes were defined using cognitive performance data, we first sought to describe the cognitive profiles of each subtype. Notably, statistical testing of cognitive performance between subtypes was not performed; as the cognitive data were used in the clustering procedure, subtypes differed in cognitive performance by construction. In contrast, clinical symptomatology and neuroimaging were independent data types that were not used in the clustering procedure, and thus were appropriate for statistical testing. Accordingly, as a first step we evaluated the clinical profiles of subtypes and controls. Finally, we evaluated whether subtypes displayed differential brain activity in the *n*-back working memory task within the 21 executive system ROIs.

For all analyses, we used a general linear model to test how well subtypes predicted the outcome of interest (clinical or imaging

measures), where subtype was modeled as a factor. When evaluating differences in activation during the *n*-back task, we included mean in-scanner motion as an additional covariate to control for the potentially confounding effects of motion on image quality. An omnibus ANOVA testing for group differences was corrected for multiple comparisons by controlling the false discovery rate (FDR, $Q < 0.05$). For measures that passed FDR correction, we then conducted pairwise post hoc tests to determine which subtypes significantly differed from each other; these post hoc tests were corrected for multiple comparisons using the Tukey method. Age-by-sex, age-by-group, and *n*-back motion-by-group interactions in the ROIs were evaluated separately, but were not significant ($P_{fdr} > 0.05$) and not evaluated further.

To conclude our study, we further evaluated between-subtype differences in resting-state functional connectivity (see Supplementary Material). Last, we performed sensitivity analyses excluding participants who were taking psychoactive medications at the time of the clinical assessment. Given the known effects psychoactive substances can have on mood, cognition, and brain activity, we sought to ensure that our results were not driven by medication effects [51, 52]. Throughout, effect sizes are reported using the Cohen's *d* statistic.

RESULTS

Of the ten possible clustering solutions generated by HYDRA, a well-defined peak at $k = 3$ emerged (ARI = 0.39, permutation-based $P_{fdr} = 0.011$), suggesting the presence of three distinct neurocognitive subtypes of DY (Fig. 1B and Supplementary Fig. 2). Each subtype had a similar number of participants (Subtype 1: $n = 264$; Subtype 2: $n = 237$; Subtype 3: $n = 211$). As an initial step, we evaluated the demographics of the neurocognitive subtypes. As expected, the subtypes did not differ in age or sex. However, Subtype 2 had a lower percentage of white patients and lower levels of maternal education. While significant, this difference was relatively modest: on average, mothers had some college education, and differed at most by ~1.5 years (Subtype 1 vs. Subtype 2).

Subtypes show distinct cognitive profiles

We next characterized the subtypes based on their overall cognitive accuracy and speed (Fig. 2A). Across all accuracy domains, Subtype 1 consistently outperformed both other depressed subtypes as well as TDs (Fig. 2B). Large effect sizes were noted (Subtype 1 vs. Subtype 2, Cohen's $d = 1.58$; Subtype 1 vs. Subtype 3, Cohen's $d = 1.49$; Supplementary Table 1). In

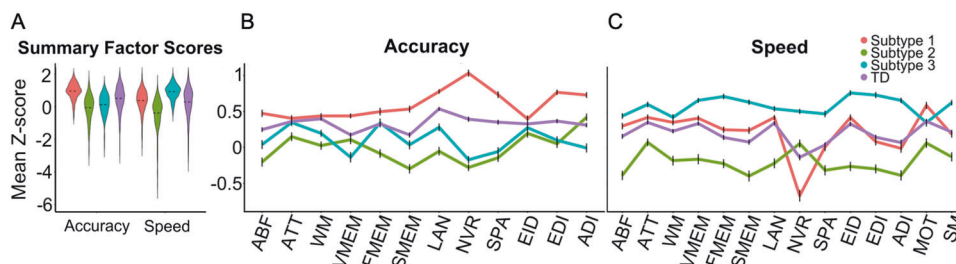


Fig. 2 Subtypes revealed by HYDRA differ in their neurocognitive profiles. **A** Three neurocognitive signatures emerged in depressed youth: High-performing Subtype 1 had preserved cognition, with high accuracy and speed; Impaired Subtype 2 had low accuracy and speed; Impulsive Subtype 3 had high speed but low accuracy. Patterns were largely consistent for all measures of accuracy (**B**) and speed (**C**). Horizontal dashed lines reflect the mean. Error bars reflect standard error of the mean. HYDRA Heterogeneity through Discriminative Analysis, ABF abstraction/mental flexibility, ATT attention, WM working memory, VMEM verbal memory, FMEM face memory, SMEM spatial memory, LAN language/verbal reasoning, NVR nonverbal reasoning, SPA spatial reasoning, EID emotion recognition, EDI emotion discrimination, ADI age discrimination, MOT motor, SM sensorimotor.

contrast, when cognitive speed was evaluated, Subtype 1 performed similarly to TDs (Cohen’s $d = -0.11$), with faster speed than Subtype 2 (Cohen’s $d = 0.97$) and slower speed than Subtype 3 (Cohen’s $d = -0.93$; Fig. 2C). Effect sizes for individual measures of speed reflected a similar pattern (Supplementary Table 2). Of note, these effect sizes are likely inflated given that cognitive data were used for clustering, thus guaranteeing differences between subtypes on these measures.

Overall, Subtype 1 was a high-performing subset of depressed participants, who were able to efficiently maximize the trade-off between accuracy and speed. Accordingly, we call Subtype 1 “High-performing.” In contrast to the high-performing Subtype 1, Subtype 2 showed globally impaired cognition, with the lowest accuracy and slowest speed of all subtypes; we call Subtype 2 “Impaired.” Finally, Subtype 3 had poor accuracy performance but fast speed, suggesting that Subtype 3 was impulsive, and was unable to accurately balance the competing demands of accuracy and speed. As such, we named this final subtype “Impulsive.”

Clinical symptoms are similar across cognitive subtypes

Next, we evaluated differences in the clinical symptom profiles of the subtypes, using dimensions of psychopathology defined using factor analysis. Notably, this independent clinical data were not used in the clustering process. Omnibus testing revealed between group differences in the domains of anxious-misery ($F_{3,1419} = 75.3$, $P_{fdr} < 0.0001$), externalizing behavior ($F_{3,1419} = 34.6$, $P_{fdr} < 0.0001$), fear ($F_{3,1419} = 23.9$, $P_{fdr} < 0.0001$), and overall psychopathology ($F_{3,1419} = 345.7$, $P_{fdr} < 0.0001$). As expected, all subtypes had higher psychopathology compared to TDs across these dimensions, which largely drove the ANOVA results. The psychosis factor did not differ across TDs and DY subtypes.

Despite such clear differences from controls, there were very few significant differences in clinical symptoms between the subtypes. Across the clinical measures evaluated, the subtypes only differed on the fear dimension (Impaired Subtype 2 > High-performing Subtype 1, $T(1419) = -4.7$, $P < 0.0001$, $d = -0.39$; Impaired Subtype 2 > Impulsive Subtype 3, $T(1419) = 4.48$, $P < 0.0001$, $d = 0.40$; see Supplementary Tables 3 and 4). High-performing Subtype 1 also had slightly more anxious-misery symptoms than Impaired Subtype 2 with a small effect size ($T(1419) = 2.8$, $P = 0.03$; $d = 0.24$). Factor analysis with all item-level symptom questions (including the depression items) was performed for comparison and was remarkably consistent (Supplementary Table 5). Similarly, there were no differences between the neurocognitive subtypes in state or trait anxiety (Supplementary Tables 6 and 7), indicating that the neurocognitive subtypes did not simply reflect the current burden of clinical symptoms.

Cognitive subtypes display distinct patterns of activation during a working memory task

Next, we tested the hypothesis that neurocognitive subtypes reflected distinct neural profiles. To do this, we evaluated activation during the n -back working memory task for the subsample of participants who completed imaging (High-performing Subtype 1: $n = 68$; Impaired Subtype 2: $n = 53$; Impulsive Subtype 3: $n = 47$; TD = 200). Specifically, we examined the signal change in 21 executive system ROIs defined a priori with an omnibus ANOVA. Of these 21 regions, six showed significant differences between groups ($P_{fdr} < 0.05$; Fig. 3A), including the left anterior dorsolateral prefrontal cortex ($F_{3,363} = 4.20$, $P_{fdr} = 0.0427$), anterior cingulate ($F_{3,363} = 3.58$, $P_{fdr} = 0.0496$), left dorsal frontal cortex ($F_{3,363} = 3.92$, $P_{fdr} = 0.0427$), right precuneus ($F_{3,363} = 4.65$, $P_{fdr} = 0.0427$), left precuneus ($F_{3,363} = 3.97$, $P_{fdr} = 0.0427$), and right crus II ($F_{3,363} = 3.82$, $P_{fdr} = 0.0427$). Five regions mapped onto well-known cortical networks: the frontoparietal network (left anterior dorsolateral prefrontal cortex, bilateral precuneus) and the cingulo-opercular network (dorsal anterior cingulate and dorsal frontal cortex).

Post hoc analyses revealed that the greatest number of differences were observed between High-performing Subtype 1 and Impaired Subtype 2, although Subtype 1 and 3 also differed in several regions (Table 2 and Fig. 3B). Specifically, subtype-by-ROI post hoc analyses confirmed that Subtype 1 had higher activation magnitude than Subtypes 2 and 3 in all six regions, with moderate effect sizes for all regions. Subtype 1 had higher activation magnitude than Subtype 3 in right crus II and left dorsal frontal cortex with moderate effect sizes. There were no pairwise differences between Impaired Subtype 2 and Impulsive Subtype 3 (Table 2 and Supplementary Table 8). In-scanner behavioral performance reflected this pattern as well, with Impaired Subtype 2 having the lowest mean d' score, followed by Impulsive Subtype 3, TD, and High-performing Subtype 1 (Supplementary Fig. 3). In sum, neurocognitive subtypes appear to have neural signatures that in part reflect in-scanner cognitive performance, despite the similar clinical symptomatology of these subtypes.

In contrast to our n -back results, our analyses of resting-state functional connectivity did not demonstrate statistically significant differences between subtypes. This suggests that specific task probes (like the n -back working memory task used in our study) may be more sensitive to differences between cognitive subtypes of DY.

Sensitivity analyses in medication-free participants provide convergent results

Finally, we performed sensitivity analyses that excluded participants ($n = 308$) who were treated with psychoactive medications at the time of study. In the remaining participants ($n = 1116$),

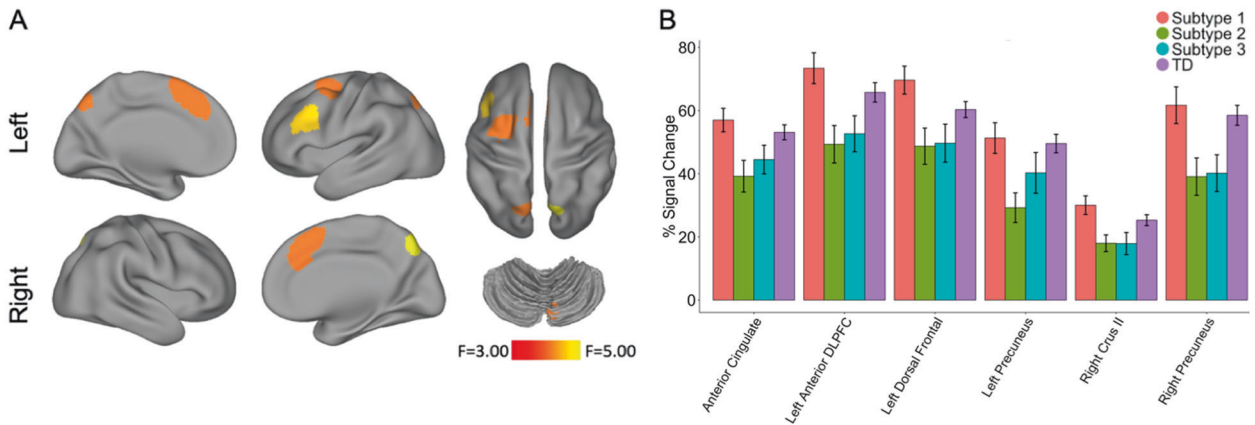


Fig. 3 Neurocognitive subtypes differ in activation of executive regions during an *n*-back working memory paradigm. A Group differences ($P_{\text{fdr}} < 0.05$) in *n*-back activation between subtypes were present in six functionally defined regions of interest, which were defined a priori in prior published work [32]. See Supplementary Fig. 1 for all twenty-one regions of interest. **B** Group differences were driven by a consistent pattern across regions, with greater activation in High-performing Subtype 1 and TDs than in Impaired Subtype 2 or Impulsive Subtype 3. Error bars reflect standard error of the mean.

Table 2. Post hoc pairwise contrasts for regions where differential activation during the *n*-back task were found.

Pairwise contrasts for <i>n</i> -back activation		
	<i>P</i>	Cohen's <i>d</i>
Subtype 1 vs. Subtype 2		
Right crus II	0.031	0.54
Left anterior DLPFC	0.017	0.58
Dorsal anterior cingulate	0.031	0.53
Left dorsal frontal	0.022	0.54
Left precuneus	0.028	0.59
Right precuneus	0.045	0.49
Subtype 1 vs. Subtype 3		
Right crus II	0.036	0.50
Left dorsal frontal	0.031	0.52
Subtype 2 vs. Subtype 3		
	NS	

Pairwise contrasts are reported as *P* values, and were adjusted via the Tukey method. Effect sizes are reported as Cohen's *d*.

cognitive profiles were virtually identical to the main analysis that considered the full group (Supplementary Fig. 4). Similar to the full group, clinical differences between groups were isolated to higher levels of fear in Impaired Subtype 2 (Supplementary Tables 9 and 10); no differences in state or trait anxiety were observed (Supplementary Tables 11 and 12). Finally, additional ROI showed significant differences between subtypes in the medication-free subsample, despite reduced statistical power. Specifically, in addition to the six executive regions that differed among groups in the full sample, the right crus I and left parietal cortex also displayed significant differences in activation ($P_{\text{fdr}} = 0.046$ for both; Supplementary Tables 13 and 14).

DISCUSSION

Using a recently developed semi-supervised machine learning algorithm and a large sample of youth with a history of depression, we identified three distinct neurocognitive subtypes of depression. Subtype 1 (High-performing) had globally

preserved cognition, and outperformed the TD youth on all domains. Subtype 2 had globally impaired cognition, while Subtype 3 was impulsive, sacrificing accuracy for speed. The activation profiles of each subtype during the *n*-back task generally reflected their neurocognitive signatures. This concordance between cognitive and neuroimaging results suggest that our data-driven approach identified biologically relevant subtypes. Importantly, these subtypes were not clearly distinguishable based on their clinical symptoms, with the exception of small differences in the fear domain. The significantly more robust differences in quantitative cognitive and neural measures are relevant given that psychiatric illnesses and treatment recommendations are currently based solely on observed clinical symptoms. Overall, our study highlights both the important heterogeneity of cognitive dysfunction in depression, and the broader promise of machine learning for parsing heterogeneity in psychiatric disorders.

Although subtypes were defined using a cognitive battery administered out of the scanner, we were able to evaluate differences between them using independent fMRI data not used in clustering. The subset of regions that showed differences between subtypes were located within the frontoparietal (dorsolateral prefrontal cortex, precuneus) and cingulo-opercular networks (dorsal anterior cingulate, dorsal frontal cortex), which are of particular developmental relevance. The frontoparietal network balances cognitive flexibility with cognitive control, both within and between separate distributed networks [53, 54]. Throughout healthy adolescent brain development, there is increased connectivity within the frontoparietal network, and the brain spends progressively more time in a frontoparietal-dominant state [55]. Dysfunctional development of this network is a risk factor for psychopathology [41]. Brain imaging studies in adults with affective disorders show abnormalities in frontoparietal network activity as well [56]. Regions within the cingulo-opercular network, which regulates salience and inhibitory control, also showed differences between the subtypes. TD adolescent brains show progressive strengthening of the cingulo-opercular network, reflecting the ability to process salient information and to engage in impulse control when selecting behaviors. Abnormal functioning of the cingulo-opercular network has been associated with anhedonia in youth as well as attention-deficit hyperactivity disorder [57–59]. In our study, youths with MDD with preserved cognition had consistently higher activation in several frontoparietal and cingulo-opercular regions even as compared to TD youth. Both the Impaired and Impulsive groups had lower activity

in these regions, suggesting that failure to effectively recruit these networks can result in distinct cognitive deficits.

Given the differences in reaction time between Subtypes 2 and 3, we expected to see the groups differ more during the imaging task. Although Subtype 2 generally had numerically lower mean percent signal change than Subtype 3, we did not find statistically significant differences when we directly compared Subtypes 2 and 3. As the main difference between these groups lies in the domain of impulsivity, which is not directly measured in the *n*-back, the *n*-back task might be less suited to demonstrate neural differences between these two groups. We hypothesize that tasks that test impulsivity and response inhibition specifically (such as a Go/No-go task) may better highlight the differences between these two subgroups.

Despite differences in cognition and neural activity in the neurocognitive subtypes, the subtypes had generally similar clinical profiles, indicating that the cognitive and neural differences observed between subtypes did not merely reflect differences in clinical status. Although Subtype 2 had higher fear scores than both Subtypes 1 and 3, the effect sizes of these differences were small. This pattern of results aligns with data suggesting that patients with similar symptomatic presentations may have divergent cognitive deficits, prognosis, and response to treatment [60]. Furthermore, this finding aligns with results from a previous meta-analysis in adults with MDD that was unable to find reliable subtypes based on symptoms alone [61].

This study adds new insights to the growing body of research that uses machine learning to understand heterogeneity in psychiatry [62]. Previous studies have primarily used either unsupervised or supervised machine learning algorithms, both of which have limitations [28, 58]. Unsupervised machine learning algorithms allow subjects to be clustered into subtypes, but do not account for important data like clinical diagnosis. Subtypes from unsupervised methods typically include both cases and controls, which is less clinically useful. Alternatively, it is possible to use unsupervised methods on patients alone. However, this approach fails to identify features that differentiate patients from controls, which are likely to be of the greatest biological relevance. In contrast, supervised machine learning algorithms can be used to directly differentiate controls and patients. However, supervised algorithms require the group label to be provided, and thus cannot assess heterogeneity. Our study overcomes these limitations by using a semi-supervised method that simultaneously performs classification and clustering. In this process, we identified subtypes of DY using features that also discriminated clusters from controls.

Machine learning analyses of neuroimaging data are becoming increasingly popular, but there are inherent difficulties in relying solely on imaging to define subtypes. Neuroimaging scans are expensive to obtain and as such, generating large datasets can be challenging [63]. Youth imaging studies are even more challenging, especially due to reduced data quality resulting from in-scanner motion [36, 64]. In our study, we were able to leverage a much larger dataset by evaluating cognitive data with HYDRA, and were subsequently able to link cognitive subtypes to patterns of brain activation. Understanding heterogeneity in cognitive performance—and using neuroimaging as an external validation—provides an alternative approach to defining biotypes.

Two limitations should be noted. First, we evaluated a cross-sectional sample, precluding estimates of within-individual change that are critical for studying neurodevelopment. Our study was also limited by an assessment that evaluated only a lifetime history of a major depressive episode, rather than diagnosis at the time of study participation. However, state measures of anxious-misery were not different between subtypes, suggesting that there is a low likelihood that current affective state drove the observed between-subtype differences. In addition, in sensitivity analyses, which excluded youth currently

taking psychoactive medications, our findings across all clinical and neuroimaging studies remained robust.

These limitations notwithstanding, our results suggest several clear next steps. First, moving forward, it will be important to link cognitive heterogeneity in depression to disease progression and functional outcomes in youth in longitudinal studies. Second, understanding how heterogeneous cognitive and neural deficits moderate treatment response is a critical next step. Finally, these data could help inform next-generation personalized neuromodulatory therapies that are tailored to the deficits present in an individual patient [65].

FUNDING AND DISCLOSURES

This work was supported by grants from the National Institute of Mental Health (NIMH; Grant Numbers: R01MH120482, R01MH107703, R01MH112847, and R01MH113550 to TDS; 2T32MH019112-29A1 to EBB; K99MH117274 to ANK; R01MH107235 to RCG; R01MH13565 to DHW; and R01MH11207 to CD). Additional support was provided by the Lifespan Brain Institute at the Children's Hospital of Philadelphia and Penn Medicine. The PNC was funded by RC2 Grants MH089983 and MH089924 to REG from the NIMH. Support for developing multivariate pattern analysis software (AS & TDS) was provided by a seed grant by the Center for Biomedical Computing and Image Analysis (CBICA) at Penn. Support was also provided by a NARSAD Young Investigator Award (ANK) as well as a Penn PROMOTES Research on Sex and Gender in Health grant (ANK) awarded as part of the Building Interdisciplinary Research Careers in Women's Health (BIRCWH) Grant (K12 HD085848) at the University of Pennsylvania. The authors declare no competing interests.

DATA AVAILABILITY

See https://github.com/PennBBL/baller_heterogen_2019 for an overview and all data analysis code used in this manuscript. Data from the Philadelphia Neurodevelopmental Cohort can be accessed at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000607.v3.p2.

CODE AVAILABILITY

All code for HYDRA can be found at <https://github.com/evanol/HYDRA>.

AUTHOR CONTRIBUTIONS

EBB, M.D., M.S. was responsible for conceptual design, and did the primary data analysis, interpretation of results, manuscript preparation, and graphics. ANK, Ph.D. provided a template for data analysis, supervised data analysis, and verified the accuracy of the final results. AS provided support for HYDRA data analysis. AA, Ph.D. provided support for neuroimaging data analysis. DSB, Ph.D. aided in conceptual design. MEC, Ph.D. provided expert guidance in clinical phenotyping and clinical symptom factor analysis. GBC, Ph.D. performed the permutation analysis. ZC, Ph.D. provided expert guidance in data analysis, graphics support, and verified the accuracy of the final results. REG, M.D., Ph.D. leads the development of the PNC dataset. RCG, Ph.D. leads the development of the CNB. KAL, Ph.D. provided expert statistical guidance, specifically with the usage of the MatchIt R-package. TMM, Ph.D. performed the specialized clinical factor analysis. DRR, Ph.D. provided expert guidance on the interpretation of the CNB. EV, Ph.D. provided expert support in HYDRA usage. DHW, M.D., Ph.D. aided in conceptual design and supervised the processing of the neuroimaging data. CHX, Ph.D. provided data analysis and graphics support. CD, Ph.D. was involved in conceptual design, specifically with respect to HYDRA, which was developed in his lab. TDS, M.D., M.A. was the supervising mentor. He was involved in the conceptual design, data analysis, manuscript drafting, and review.

ADDITIONAL INFORMATION

Supplementary Information accompanies this paper at (<https://doi.org/10.1038/s41386-020-00871-w>).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

REFERENCES

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Lond Engl*. 2017;390:1211–59.
2. Friedrich MJ. Depression is the leading cause of disability around the world. *JAMA*. 2017;317:1517.
3. Sinyor M, Schaffer A, Levitt A. The sequenced treatment alternatives to relieve depression (STAR*D) trial: a review. *Can J Psychiatry Rev Can Psychiatr*. 2010;55:126–35.
4. McLachlan G. Treatment resistant depression: what are the options? *BMJ*. 2018;363:k5354. <https://www.bmj.com/content/363/bmj.k5354>.
5. Kessler RC, Berglund P, Demler O, Jin R, Merikangas KR, Walters EE. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the National Comorbidity Survey Replication. *Arch Gen Psychiatry*. 2005;62:593–602.
6. Ghandour RM, Sherman LJ, Vladutiu CJ, Ali MM, Lynch SE, Bitsko RH, et al. Prevalence and treatment of depression, anxiety, and conduct problems in US children. *J Pediatr*. 2019;206:256–267.e3.
7. Brent D, Emslie G, Clarke G, Wagner KD, Asarnow JR, Keller M, et al. Switching to another SSRI or to venlafaxine with or without cognitive behavioral therapy for adolescents with SSRI-resistant depression. *JAMA J Am Med Assoc*. 2008;299:901–13.
8. Kennard BD, Emslie GJ, Mayes TL, Nakonezny PA, Jones JM, Foxwell AA, et al. Sequential treatment with fluoxetine and relapse prevention CBT to improve outcomes in pediatric depression. *Am J Psychiatry*. 2014;171:1083–90.
9. Melchior M, Ziad A, Courtin E, Goldberg M, Zins M, van der Waerden J. Intergenerational socioeconomic mobility and adult depression: the CONSTANCES study. *Am J Epidemiol*. 2018;187:260–9.
10. Lee C-H, Giuliani F. The role of inflammation in depression and fatigue. *Front Immunol*. 2019;10:1–12. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6658985/>.
11. Chapman DP, Whitfield CL, Felitti VJ, Dube SR, Edwards VJ, Anda RF. Adverse childhood experiences and the risk of depressive disorders in adulthood. *J Affect Disord*. 2004;82:217–25.
12. Flint J, Kendler KS. The genetics of major depression. *Neuron*. 2014;81:484–503.
13. American Psychiatric Association. Diagnostic and statistical manual of mental disorders (DSM-5[®]). Arlington, VA: American Psychiatric Pub; 2013.
14. Wagner S, Müller C, Helmreich I, Huss M, Tadić A. A meta-analysis of cognitive functions in children and adolescents with major depressive disorder. *Eur Child Adolesc Psychiatry*. 2015;24:5–19.
15. Maalouf FT, Brent D, Clark L, Tavitian L, McHugh RM, Sahakian BJ, et al. Neurocognitive impairment in adolescent major depressive disorder: state vs. trait illness markers. *J Affect Disord*. 2011;133:625–32.
16. Allott K, Fisher CA, Amminger GP, Goodall J, Hetrick S. Characterizing neurocognitive impairment in young people with major depression: state, trait, or scar? *Brain Behav*. 2016;6:e00527.
17. Hermens DF, Hodge MAR, Naismith SL, Kaur M, Scott E, Hickie IB. Neuropsychological clustering highlights cognitive differences in young people presenting with depressive symptoms. *J Int Neuropsychol Soc*. 2011;17:267–76.
18. Barzilay R, Calkins ME, Moore TM, Boyd RC, Jones JD, Benton TD, et al. Neurocognitive functioning in community youth with suicidal ideation: gender and pubertal effects. *Br J Psychiatry J Ment Sci*. 2019;3:1–7.
19. Lee RSC, Hermens DF, Redoblado-Hodge MA, Naismith SL, Porter MA, Kaur M, et al. Neuropsychological and socio-occupational functioning in young psychiatric outpatients: a longitudinal investigation. *PLoS ONE*. 2013;8:e58176. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3585793/>.
20. Goodall J, Fisher C, Hetrick S, Phillips L, Parrish EM, Allott K. Neurocognitive functioning in depressed young people: a systematic review and meta-analysis. *Neuropsychol Rev*. 2018;28:216–31.
21. Favre T, Hughes C, Emslie G, Stavinoha P, Kennard B, Carmody T. Executive functioning in children and adolescents with major depressive disorder. *Child Neuropsychol J Norm Abnorm Dev Child Adolesc*. 2009;15:85.
22. Blakemore S-J, Choudhury S. Development of the adolescent brain: implications for executive function and social cognition. *J Child Psychol Psychiatry*. 2006;47:296–312.
23. Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, et al. Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci*. 1999;2:861.
24. Halari R, Simic M, Pariante CM, Papadopoulos A, Cleare A, Brammer M, et al. Reduced activation in lateral prefrontal cortex and anterior cingulate during attention and cognitive control functions in medication-naïve adolescents with depression compared to controls. *J Child Psychol Psychiatry*. 2009;50:307–16.
25. Chantiluke K, Halari R, Simic M, Pariante CM, Papadopoulos A, Giampietro V, et al. Fronto-striato-cerebellar dysregulation in adolescents with depression during motivated attention. *Biol Psychiatry*. 2012;71:59–67.
26. Pan LA, Batezati-Alves SC, Almeida JRC, Segreti A, Akkal D, Hassel S, et al. Dissociable patterns of neural activity during response inhibition in depressed adolescents with and without suicidal behavior. *J Am Acad Child Adolesc Psychiatry*. 2011;50:602–611.e3.
27. Yang TT, Simmons AN, Matthews SC, Tapert SF, Frank GK, Bischoff-Grethe A, et al. Depressed adolescents demonstrate greater subgenual anterior cingulate activity. *Neuroreport*. 2009;20:440–4.
28. Drysdale AT, Grosenick L, Downar J, Dunlop K, Mansouri F, Meng Y, et al. Resting-state connectivity biomarkers define neurophysiological subtypes of depression. *Nat Med*. 2017;23:28.
29. Kaczurkin AN, Sotiras A, Baller EB, Barzilay R, Calkins ME, Chand GB, et al. Neurostructural Heterogeneity in Youths With Internalizing Symptoms. *Biol Psychiatry*. 2020;87:473–82.
30. Varol E, Sotiras A, Davatzikos C, Alzheimer's Disease Neuroimaging Initiative. HYDRA: revealing heterogeneity of imaging and genetic patterns through a multiple margin discriminative analysis framework. *Neuroimage*. 2017;145:346–64.
31. Ragland JD, Turetsky BI, Gur RC, Gunning-Dixon F, Turner T, Schroeder L, et al. Working memory for complex figures: an fMRI comparison of letter and fractal n-back tasks. *Neuropsychology*. 2002;16:370–9.
32. Satterthwaite TD, Wolf DH, Erus G, Ruparel K, Elliott MA, Gennatas ED, et al. Functional maturation of the executive system during adolescence. *J Neurosci*. 2013;33:16249–61.
33. Goldman-Rakic PS. Regional and cellular fractionation of working memory. *Proc Natl Acad Sci*. 1996;93:13473–80.
34. Perlman SB, Huppert TJ, Luna B. Functional near-infrared spectroscopy evidence for development of prefrontal engagement in working memory in early through middle childhood. *Cereb Cortex*. 2016;26:2790–9.
35. Yüksel D, Dietsche B, Konrad C, Dannlowski U, Kircher T, Krug A. Neural correlates of working memory in first episode and recurrent depression: an fMRI study. *Prog Neuropsychopharmacol Biol Psychiatry*. 2018;84:39–49.
36. Satterthwaite TD, Elliott MA, Ruparel K, Loughhead J, Prabhakaran K, Calkins ME, et al. Neuroimaging of the Philadelphia Neurodevelopmental Cohort. *Neuroimage*. 2014;86:544–553.
37. Calkins ME, Merikangas KR, Moore TM, Burstein M, Behr MA, Satterthwaite TD, et al. The Philadelphia Neurodevelopmental Cohort: constructing a deep phenotyping collaborative. *J Child Psychol Psychiatry*. 2015;56:1356–69.
38. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for affective disorders and schizophrenia for school-age children-present and lifetime version (K-SADS-PL): initial reliability and validity data. *J Am Acad Child Adolesc Psychiatry*. 1997;36:980–8.
39. American Psychiatric Association. Diagnostic and statistical manual of mental disorders: DSM-IV-TR. Washington, DC: American Psychiatric Association; 2000.
40. Merikangas KR, He J, Burstein M, Swanson SA, Avenevoli S, Cui L, et al. Lifetime prevalence of mental disorders in U.S. adolescents: results from the National Comorbidity Survey Replication–Adolescent Supplement (NCS-A). *J Am Acad Child Adolesc Psychiatry*. 2010;49:980–9.
41. Shanmugan S, Wolf DH, Calkins ME, Moore TM, Ruparel K, Hopson RD, et al. Common and dissociable mechanisms of executive system dysfunction across psychiatric disorders in youth. *Am J Psychiatry*. 2016;173:517–26.
42. Kaczurkin AN, Park SS, Sotiras A, Moore TM, Calkins ME, Cieslak M, et al. Evidence for dissociable linkage of dimensions of psychopathology to brain structure in youths. *Am J Psychiatry*. 2019;176:1000–9.
43. Bados A, Gómez-Benito J, Balaguer G. The state-trait anxiety inventory, trait version: does it really measure anxiety? *J Pers Assess*. 2010;92:560–7.
44. Bieling PJ, Antony MM, Swinson RP. The state-trait anxiety inventory, trait version: structure and content re-examined. *Behav Res Ther*. 1998;36:777–88.
45. Nitschke JB, Heller W, Imig JC, McDonald RP, Miller GA. Distinguishing dimensions of anxiety and depression. *Cogn Ther Res*. 2001;25:1–22.
46. Moore TM, Reise SP, Gur RE, Hakonarson H, Gur RC. Psychometric properties of the Penn computerized neurocognitive battery. *Neuropsychology*. 2015;29:235.
47. Wilkinson GS, Robertson GJ. WRAT 4: wide range achievement test. Lutz, FL: Psychological Assessment Resources; 2006.
48. Wolf DH, Satterthwaite TD, Calkins ME, Ruparel K, Elliott MA, Hopson RD, et al. Functional neuroimaging abnormalities in youth with psychosis spectrum symptoms. *JAMA Psychiatry*. 2015;72:456–65.
49. Snodgrass JG, Corwin J. Pragmatics of measuring recognition memory: Applications to dementia and amnesia. *J Exp Psychol Gen*. 1988;117:34–50.

50. Shamosh NA, DeYoung CG, Green AE, Reis DL, Johnson MR, Conway ARA, et al. Individual differences in delay discounting: relation to intelligence, working memory, and anterior prefrontal cortex. *Psychol Sci*. 2008;19:904–11.
51. Wandschneider B, Koepp MJ. PharmacofMRI: determining the functional anatomy of the effects of medication. *NeuroImage Clin*. 2016;12:691–7.
52. Kraus C, Castrén E, Kasper S, Lanzenberger R. Serotonin and neuroplasticity—links between molecular, functional and structural pathophysiology in depression. *Neurosci Biobehav Rev*. 2017;77:317–26.
53. Power JD, Schlaggar BL, Lessov-Schlaggar CN, Petersen SE. Evidence for hubs in human functional brain networks. *Neuron*. 2013;79:798–813.
54. Marek S, Dosenbach NUF. The frontoparietal network: function, electrophysiology, and importance of individual precision mapping. *Dialogues Clin Neurosci*. 2018;20:133–40.
55. Medaglia JD, Satterthwaite TD, Kelkar A, Ciric R, Moore TM, Ruparel K, et al. Brain state expression and transitions are related to complex executive cognition in normative neurodevelopment. *NeuroImage*. 2018;166:293–306.
56. Petersen SE, Posner MI. The attention system of the human brain: 20 years after. *Annu Rev Neurosci*. 2012;35:73–89.
57. Pornpattananangkul N, Leibenluft E, Pine DS, Stringaris A. Association between childhood anhedonia and alterations in large-scale resting-state networks and task-evoked activation. *JAMA Psychiatry*. 2019;76:624–33.
58. Costa Dias TG, Iyer SP, Carpenter SD, Cary RP, Wilson VB, Mitchell SH, et al. Characterizing heterogeneity in children with and without ADHD based on reward system connectivity. *Dev Cogn Neurosci*. 2015;11:155–74.
59. Roy A, Bennett R, Posner J, Hulvershorn L, Castellanos F, Klein R. Altered intrinsic functional connectivity of the cingulate cortex in children with severe temper outbursts. *Dev Psychopathol*. 2017;30:1–9.
60. Gorlyn M, Keilp JG, Grunebaum MF, Taylor BP, Oquendo MA, Bruder GE, et al. Neuropsychological characteristics as predictors of SSRI treatment response in depressed subjects. *J Neural Transm*. 2008;115:1213–9.
61. van Loo HM, de Jonge P, Romeijn J-W, Kessler RC, Schoevers RA. Data-driven subtypes of major depressive disorder: a systematic review. *BMC Med*. 2012;10:156.
62. Kircanski K, White LK, Tseng W-L, Wiggins JL, Frank HR, Sequeira S, et al. A latent variable approach to differentiating neural mechanisms of irritability and anxiety in youth. *JAMA Psychiatry*. 2018;75:631–9.
63. Perlman SB. Neuroimaging in child clinical populations: considerations for a successful research program. *J Am Acad Child Adolesc Psychiatry*. 2012;51:1232–5.
64. Satterthwaite TD, Wolf DH, Loughhead J, Ruparel K, Elliott MA, Hakon H, et al. Impact of in-scanner head motion on multiple measures of functional connectivity: relevance for studies of neurodevelopment in youth. *NeuroImage*. 2012;60:623–32.
65. Kim TD, Hong G, Kim J, Yoon S. Cognitive enhancement in neurological and psychiatric disorders using transcranial magnetic stimulation (TMS): a review of modalities, potential mechanisms and future implications. *Exp Neurobiol*. 2019;28:1–16.