Supplementary Online Content

Miller CH, Hamilton JP, Sacchet MD, Gotlib IH. Meta-analysis of functional neuroimaging of major depressive disorder in youth. *JAMA Psychiatry*. Published online September 2, 2015. doi:10.1001/jamapsychiatry.2015.1376.

eAppendix. Methods: Additional Explanation of Data Analysis.
eFigure. Neurosynth Features Associated with Whole-Brain Meta-Analytic Maps.
eTable 1. Clinical Characteristics and Comorbidity of Included Samples of MDD Youth.
eTable 2. Rank-Ordered List of Neurosynth Features Associated with Whole-Brain Meta-Analytic Maps.

This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix. Methods: Additional Details Concerning Data Analysis

We applied multilevel kernel density analysis (MKDA; Wager et al., 2007) to published fMRI studies that compared neural activation in groups of youth diagnosed with MDD to age-matched healthy controls. This analytic procedure involved conducting an exhaustive literature search for relevant primary studies and using reported coordinates from these studies for whole-brain, voxel-wise comparisons, corrected for multiple comparisons. We applied this analytic procedure to five subgroups of studies, decomposed by task conditions, and conducted follow-up analyses to establish more directly whether regional activity patterns were task-specific or robust across different experimental stimuli. Finally, we decoded whole-brain maps from each of the five groupings of primary studies using the Neurosynth database to provide a large-scale, quantitative basis for reverse inference.

We first identified potential primary studies in the PubMed and Web of Science databases. For each database, we entered search terms into either the “Title/Abstract” field (PubMed) or “Topic” field (Web of Science) to locate primary studies that involved fMRI (<”fMRI” OR "functional MRI" OR "functional magnetic resonance">) of MDD (<depress* OR “MDD”>) in youth (<child* OR teen* OR adolescen* OR kid* OR juvenile* OR pediatric OR "early onset" OR "early-onset" OR “youth”>). We also examined the reference lists and study tables of relevant review articles to identify for inclusion other primary studies that may have been missed in the original search. We then inspected each article generated by this search process and retained only those that satisfied the following inclusion criteria: (1) used an fMRI-based voxel-wise whole-brain analysis (WBA) of task-based activation data; (2) compared a group of participants ages 4-24 years (mean=18.36 years) diagnosed with MDD according to DSM criteria (APA, 2000) at the time of scan and age-matched healthy controls; and (3) reported coordinates of

brain regions with abnormal activations in standard space, using the Talairach atlas or Montreal Neurological Institute (MNI) template.

We included studies that assessed other groups, such as anxiety disorders, if the study reported neuroimaging results from at least one contrast of MDD youth compared to non-disordered controls. Similarly, we included studies that used an alternative approach, such as an ROI analysis, only if they also reported neuroimaging results from a WBA. We assured the inclusion of independent samples by contacting relevant investigators to verify sample independence. We retained studies with overlapping samples only if they used different experimental paradigms that could be included in distinct analyses (e.g., positive valence vs. negative valence). For each analysis in which multiple studies with overlapping samples could have contributed, we included only the study with the largest sample size.

These searches generated 325 and 719 articles from PubMed and Web of Science, respectively. From this pool of 1044 articles, we excluded 743 articles because they did not select participants based on a diagnosis of MDD, 81 articles because they did not include whole-brain data, 76 articles because they did not include task-based activation data, 19 articles because they included participants outside of our specified age range, 5 articles because they did not include a healthy control group and, consequently, lacked relevant contrasts, 85 articles because they were conference abstracts, and 7 articles because they were literature reviews and not empirical papers. This left a total of 14 primary studies, each of which was independently identified in both database searches, that met our inclusion criteria.

Next, we extracted published coordinates, expressed in Talairach or MNI space, from each of these 14 primary studies. In instances of a single study with coordinates from multiple contrasts, we included each set of findings involving MDD vs. non-disordered controls if they were obtained from different experimental conditions (e.g.,

happy vs. neutral faces and sad vs. neutral faces). If findings were drawn from similar contrasts, we chose only a single contrast with the greatest intensity (e.g., we selected high-intensity happy over medium/low-intensity happy vs. neutral faces) or the largest sample (e.g., suicide-non-attempters over suicide attempters) or the purest definition (e.g., pure MDD over MDD with comorbid anxiety).

We then used extracted coordinate data to construct indicator maps in Talairach space for each individual study, with hyperactive (MDD>control) and hypoactive (control>MDD) regions displayed separately. These indicator maps were constructed in Talairach space at 1mm isotropic voxel resolution with the AFNI statistical package (Cox, 1996), and MNI coordinates were converted to Talairach space using the MATLAB script mni2tal.m (http://imaging-mrc-cbu.cam.ac.uk/imaging/MniTalairach). Each indicator map was composed of binary values, with 1s and 0s used to represent voxels with and without reported between-group differences, respectively. Because primary studies frequently failed to report observed cluster sizes, it was often necessary to assign a standard sphere size to reported coordinates; for this analysis, we applied a standard radius of 10mm. This value has been used in other neuroimaging meta-analyses and is considered a suitable standard value for the desired sensitivity and spatial resolution of fMRI (e.g., Etkin et al., 2012; Hamilton et al., 2012).

The indicator maps from each primary study were then merged to create a meta-analytic statistical map that displayed global activation values at each voxel. Specifically, activation values were obtained by computing the weighted proportion of primary studies that reported statistically significant activation differences between MDD and control groups according to the following formula:

$$\hat{P}_V = \sum_{n=1}^{N} w_n I_n$$

where the weighted activation proportion at each voxel ($\hat{P}_V$) is calculated from the square root of the number of subjects (w_n), across both MDD and control groups, multiplied by the binary indicator value (I_n) of the nth of N studies. Global activation values were then obtained by calculating the difference between activation values from areas of hyperactivity versus hypoactivity, in order to quantitatively assess global activity across studies that reported opposite directionality in activation:

$$\hat{P}_{VG} = \hat{P}_{VI} - \hat{P}_{VD}$$

where $\hat{P}_{VG}$ is the global activation value and $\hat{P}_{VI}$ and $\hat{P}_{VD}$ are the activation values for increased and decreased levels, respectively, for MDD compared to control participants.

This meta-analytic statistical map was then compared against a null-hypothesis density distribution in order to retain only clusters with global activation values above those expected by statistical chance; remaining regions of abnormal activity were thresholded by cluster size in order to further minimize false positives. Specifically, null hypothesis distributions were computed at $\alpha<0.005$ by performing 10,000 Monte Carlo simulations in MATLAB, limited to a gray matter mask (plus 8-mm border) in the standard brain (SPM2 segmented avg152tl.img with 9-mm Gaussian smoothing).

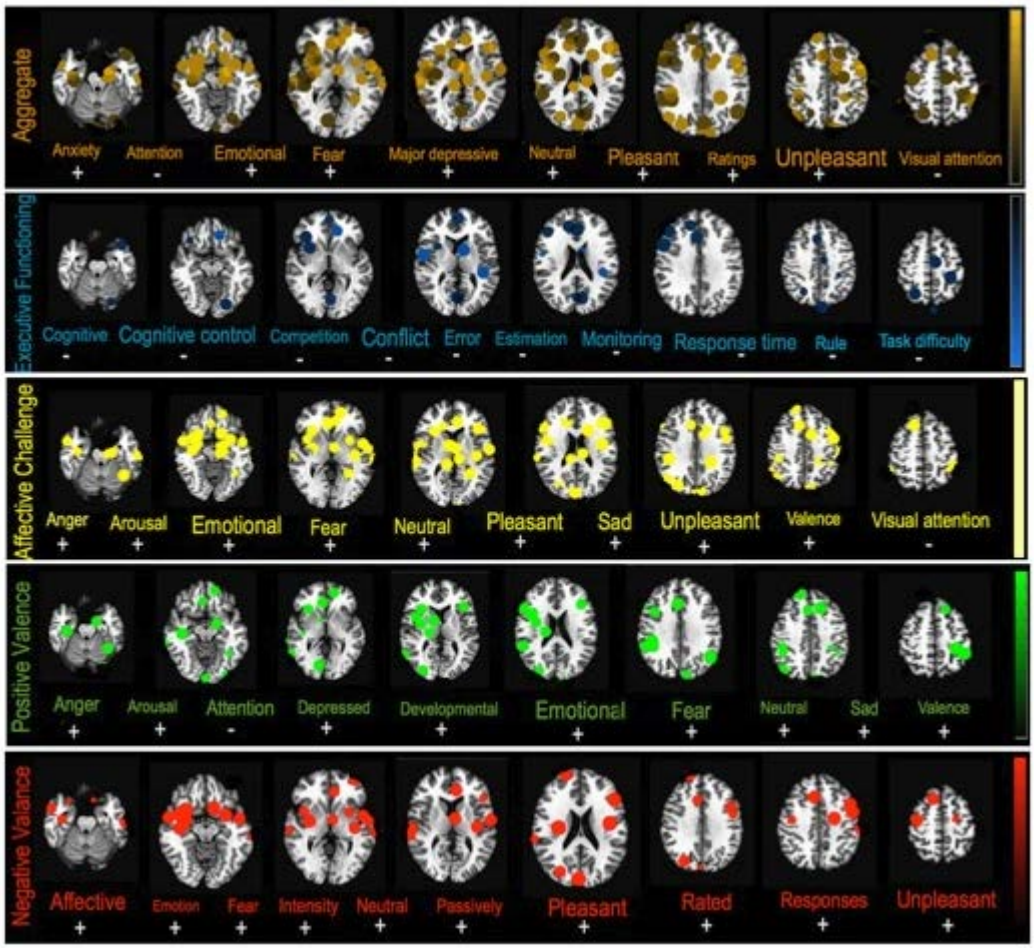
The AFNI program AlphaSim was then used to determine the cluster size necessary to obtain family-wise-error-rate (FWE) control at $\alpha<0.05$ for comparison of multiple voxels across the whole brain. Because each combination of p_f/ p_v values is accompanied by a particular meta-analytic statistic and cluster threshold, however, choosing only one p_v/p_f pair necessarily results in a detection bias that favors either larger or smaller clusters. For example, in our aggregate analysis, using the most lenient thresholds of $\alpha=0.05$ requires a cluster size of 8292 contiguous voxels with a meta-analytic statistic $\hat{P}_{VG}>5.71$, whereas the most stringent threshold of $\alpha=0.0001$ requires a cluster size of only 661 contiguous voxels with $\hat{P}_{VG}>16.95$. Consequently, in order to

improve detection sensitivity without favoring either large clusters (small $\hat{P}_{VG,}$ large cluster size) or small clusters (large $\hat{P}_{VG,}$ small cluster size), we evaluated results at several regular levels of significance combinations, ranging from α= 0.05–0.0001, for both p_f and p_v. Importantly, this approach limits the search for significant clusters that always remain under the established threshold of α<0.05 for both p_f and p_v, but avoids the problem of favoring clusters with a particular size.

Each of the four study groupings used in our task-specific analyses—executive functioning and affective processing (superordinate categories) and positive and negative valence (subordinate categories of affective processing)—represents a major neuropsychological construct that has been examined extensively in the literature and includes the same experimental tasks that were used in the primary studies that comprise this meta-analysis. Importantly, in research with healthy adults and with individuals diagnosed with MDD, there are commonalities in patterns of neural activation across these four categories of psychological processing, thereby warranting an examination of an “aggregate” grouping; there are also, however, distinct patterns of activation in these categories, warranting their separation into the groupings examined in this meta-analysis. Conducting both aggregated and specific analyses enabled us to identify not only task-general effects that are observed across a variety of experimental conditions (analyzing all of the studies together), but also task-specific effects that are associated selectively with specific experimental tasks (analyzing and contrasting specific groups of studies/tasks). In order to ensure reliable study/task classification, each study contrast was independently categorized into one or more of these groupings by three researchers familiar with this field (inter-rater reliability, kappa=0.884); the few differences among the raters were subsequently resolved through discussion and consensus. All final analyses were conducted on the consensus groupings.

One study, Gaffrey et al. (2013), examined a sample of MDD children with a substantially different mean age (5.04 years) than the other 13 primary studies (13.42–20.37 years). Therefore, we re-analyzed the extracted brain coordinates, excluding the Gaffrey et al. (2013) study, and confirmed that 11 of the 12 brain regions that we identified remained statistically significant. One region, the right ventrolateral prefrontal cortex (Talairach coordinates=-47,3,14), failed to reach statistical significance in this re-analysis, which we think may be attributable in part to its more noticeable hyperactivation during early development and possible switch to hypoactivation around adulthood, as noted in other studies (see p.12).

eFigure. Neurosynth Features Associated with Whole-Brain Meta-Analytic Maps



Note: All brain maps are non-thresholded contrasts of MDD – healthy control participants. Each of the five panels displays brain maps from a single grouping of studies, indicated at the left of each panel, and is composed of images taken at regular intervals from axial slices that conform to neurological convention. For each of the study groupings, the ten most strongly correlated Neurosynth features are displayed; the relative sizes of the terms are scaled according to the magnitude of their correlation coefficient and the direction of the correlation for each term is indicated as either positive (+) or negative (-).

eTable 1. Clinical Characteristics and Comorbidity of Included Samples of MDD Youth

Study	Clinical Characteristics			Comorbidity		
Study	**BDI/CDI Score[1]**	**Medicated (%)**	**First Episode (%)**	**Bipolar Disorder (%)**	**Anxiety Disorder (%)**	**Axis I Disorder[2] (%)**
Chantiluke et al. (2012)	36.00 (8.0)	0.00	100.00	n/a	n/a	n/a
Colich et al. (2015)	26.67 (8.04)*	38.99	n/a	0.00	44.44	n/a
Davey et al. (2011)	33.90 (11.3)	52.90	52.90	0.00	29.40	0.00
Diler et al. (2013)	n/a	60.00	100.00	n/a	60.00	n/a
Diler et al. (2014)	73.80 (12.50)	60.00	n/a	n/a	60.00	n/a
Gaffrey et al. (2013)	n/a	0.00	n/a	n/a	n/a	56.5
Halari et al. (2009)	35.40 (7.70)	0.00	100.00	n/a	n/a	n/a
Hall et al. (2014)	28.26 (12.00)	0.00	n/a	0.00	43.75	56.0
Roberson-Nay et al. (2006)	n/a	0.00	n/a	n/a	40.00	n/a
Sharp et al. (2014)	n/a	0.00	n/a	n/a	n/a	n/a
Tao et al. (2012)	n/a	0.00	n/a	0.00	31.60	0.00
Yang et al. (2009)	n/a	0.00	n/a	n/a	n/a	n/a
Yang et al. (2010)	n/a	0.00	100.00	n/a	n/a	n/a
Zhong et al. (2012)	n/a	0.00	100.00	n/a	n/a	n/a

[1] BDI: Beck Depression Inventory; CDI: Children's Depression Inventory (indicated with asterisk)
[2] Percentage of MDD participants with any other comorbid DSM Axis I disorder, where reported.

eTable 2. Rank-Ordered List of Neurosynth Features Associated with Whole-Brain Meta-Analytic Maps

Aggregate Brain Map		Affective Challenge Brain Map		Executive Function Brain Map		Positive Valence Brain Map		Negative Valence Brain Map	
Feature	CC	Feature	CC	Feature	CC	Feature	CC	Feature	CC
unpleasant	0.142	emotional	0.1262	conflict	-0.1267	emotional	0.0957	pleasant	0.1162
pleasant	0.1275	pleasant	0.1229	cognitive control	-0.1143	fear	0.091	affective	0.1131
emotional	0.1136	unpleasant	0.122	monitoring	-0.11	attention	-0.081	unpleasant	0.1124
fear	0.1112	fear	0.1098	error	-0.1051	anger	0.0807	rated	0.1088
visual attention	-0.105	neutral	0.1093	response time	-0.1044	sad	0.0756	emotion	0.1022
anxiety	0.1021	visual attention	-0.1073	cognitive	-0.0976	depressed	0.0707	responses	0.0999
neutral	0.1011	arousal	0.104	estimation	-0.0962	developmental	0.0691	neutral	0.0923
major depressive	0.1002	anger	0.0956	rule	-0.096	valence	0.0689	intensity	0.0915
attention	-0.0996	valence	0.0953	competition	-0.0914	neutral	0.0688	passively	0.0894
ratings	0.0994	sad	0.0945	task difficulty	-0.0909	arousal	0.0684	fear	0.0858
happy	0.0978	happy	0.0931	likelihood estimation	-0.0893	threat	0.0668	anxiety	0.0839
negative	0.0969	affective	0.0895	solving	-0.0888	response time	-0.0666	valence	0.0833
reactivity	0.096	major depressive	0.087	shifting	-0.0866	adaptive	-0.0632	imagery	-0.0817
angry	0.0947	threat	0.0841	inhibition	-0.0835	implicit	0.0625	disgust	0.0808
arousal	0.0901	ratings	0.084	response conflict	-0.0811	pleasant	0.0624	psychiatric disorders	0.0767
valence	0.0901	reward	0.0811	attention task	-0.0793	visual attention	-0.0623	arousal	0.0762
efficient	-0.0899	anxiety	0.0808	adaptive	-0.079	happy	0.0602	ratings	0.0754
fearful	0.0889	reactivity	0.0807	integrate	-0.0789	conflicting	0.06	efficiency	-0.0753
performance	-0.0886	sexual	0.0806	performance	-0.0787	mood	0.0565	performance	-0.073
motivation	0.0878	motivation	0.0803	executive	-0.0758	social	0.0545	visual attention	-0.0726
intensity	0.0868	imagery	-0.0775	demands	-0.0757	low level	-0.0543	feelings	0.07

rewarding	0.0866	psychiatric disorders	0.0773	distraction	-0.0749	children	0.053	speech production	0.0694
sexual	0.0862	negative	0.0759	response inhibition	-0.0694	hallucinations	-0.0517	reward	0.0688
affective	0.0853	discriminate	-0.071	compensatory	-0.0687	task relevant	-0.0516	avoidance	0.0687
abnormality	0.0847	mood	0.0706	decision making	-0.0687	traumatic	0.0514	integration	0.0678

Note: All brain maps are non-thresholded contrasts of MDD – healthy control participants. For each of the five whole- brain maps, Neurosynth features are ordered vertically by decreasing magnitude of the correlation coefficient (CC). Green shading represents positive correlations between brain maps and features, whereas red shading indicates negative correlations between brain maps and features.