Supplemental Method

Diagnosis. Diagnosis of remitted MDD or HC status was determined using the Diagnostic Interview for Genetic Studies (Nurnberger et al., 1994), and was confirmed using a modified Family Interview for Genetic Studies completed with a parent, guardian, or close sibling (Nurnberger et al., 1994). Both diagnostic measures have demonstrated good validity and reliability (Bucholz et al., 1994; Maxwell, 1992).

PGNG. The task involves responding as quickly and as accurately as possible to certain target types (letters rapidly presented on a computer screen), while inhibiting prepotent responses to targets that repeat, and static non-targets. The PGNG is completed during scanning using a front display and consists of three runs, including 52 lure trials and 240 target trials. For all trials, a serial stream of letters is presented (black letters in 40-point Times font on a white background computer screen), each letter for 500-ms intervals with a 0-ms inter-stimulus interval. Responses are made as quickly as possible with the index finger of the dominant hand by five-button response 'claw' (Psychology Software Tools). The task yields a variable called percent correct inhibition (PCI), which was used for analysis (behavioral accuracy). The PGNG has demonstrated good construct validity and retest reliability in healthy and depressed samples (Langenecker et al., 2007b,c; Votruba and Langenecker, 2013, Peters et al., 2015; Stange et al., 2017).

FEPT. In each trial, a fixation cross is presented for 500 ms, followed by a face for 300ms, followed by a visual grey-scale mask that was matched to the facial stimuli for brightness and contrast for 100ms to prevent visual afterburn phenomena. Participants identified which emotion they perceived, and responded in the subsequent 3100ms using the 5-button response claw (Psychology Software Tools). The dependent behavioral variables were accuracy for each

emotion. The face stimuli consisted of color pictures from the MacBrain Foundation set (Tottenham et al., 2009). A pre-scanner computerized practice version used face stimuli from the Ekman and Friesen (1976) set. Blocks of faces were interspersed with rest blocks across five runs of 4 min, 20 s each. The emotions were counterbalanced across participants so that the response for each emotion had a different finger for different participants.

fMRI Acquisition. The UM scanner was a 3T Signa (release VH3, General Electric, USA) and used a reverse spiral sequence to acquire 36 slices, 3.5-mm thick. The UIC scanner was a 3T GE Discovery and acquired 44 slices, 3-mm thick with a gradient-echo axial echoplanar imaging sequence. UM data were despiked using FSL by the technicians at the scanner. Data from UIC were despiked using AFNI at the beginning of our preprocessing protocol. After that, all data were slice-time corrected in SPM, then realigned to the 10th volume in FSL using MCFLIRT (Jenkinson et al., 2002). Brain extraction of anatomical images was performed with FSL's Brain Extraction Tool (Smith, 2002) then co-registered to functional images and spatially normalized to Montreal Neurological Institute (MNI) space in SPM, with a final reconstructed spatial resolution of 2x2x2. Smoothing was completed in SPM with a full-width at half-maximum filter of 5mm. The FEPT and PGNG were processed as separate tasks. First-level model contrast images were carried forward for correct rejections (PGNG), sad faces and fearful faces (FEPT). The second-level group-by-task condition models were also built in SPM8.

Supplemental Results

Site x Group Differences in height or extent of activation. ANOVAs were conducted to examine whether height or extent of fMRI activation was influenced differentially by study site. For the clusters identified by main effect of group contrasts in SPM, the site x group

interaction was the key analysis of interest for extent of height (degree) comparisons. For regions identified by group x task contrasts in SPM, the site x group x task interaction was the key analysis of interest. Neither of the interactions between group and site predicting activation in the two main effect clusters were significant. Of the 14 clusters identified by the group x task interaction within the sad faces and rejections model, only one (temporal lobe, fusiform gyrus, in the CCN) significantly differed by group ($p = .035$); the group x task interaction was significant only in the UM sample. Of the five additional clusters identified by the group x task interaction within the fearful faces and rejections model, only one (parietal lobe, supramarginal gyrus, in the SEN) significantly differed by group ($p < .001$); the group x task interaction was significant only in the UIC sample. Finally, within the ANOVAs predicting extent of fMRI activation across the two networks and three task conditions (sad faces, fearful faces, rejections), the site x group x task interaction was non-significant.