Supplementary Information: Distinguishing Adolescents With ADHD From Their Unaffected Siblings and Healthy Comparison Subjects by Neural Activation Patterns During Response Inhibition

Daan van Rooij, MSc.; Pieter J. Hoekstra, M.D.; Maarten Mennes, PhD.; Daniel von Rhein, MSc.; Andrieke Thissen, MSc.; Dirk Heslenfeld, PhD.; Marcel P. Zwiers, PhD.; Stephen V. Faraone, PhD.; Jaap Oosterlaan, PhD.; Barbara Franke, PhD.; Nanda Rommelse, PhD.; Jan K. Buitelaar, M.D.; Catharina A. Hartman, M.D.

Supplementary Methods

Determining diagnostic status in the NeuroIMAGE sample

To determine ADHD diagnosis, as well as possible comorbid oppositional defiant disorder (ODD) and conduct disorder (CD) at the time of participation in NeuroIMAGE, all participants in the study were assessed using a semi-structured diagnostic interview (the Dutch translation of the Schedule for Affective Disorders and Schizophrenia for School-Age Children - Present and Lifetime Version (K-SADS (1)), conducted by trained professionals). ADHD diagnosis was further supplemented by the Conners' ADHD questionnaires. For participants using medication, ratings were collected on their functioning off medication. Each child was assessed with a parent-rated questionnaire (Conners' Parent Rating Scale - Revised: Long version (CPRS-R:L); combined with either a teacher-rating (Conners' Teacher Rating Scale - Revised: Long version (CTRS-R:L (2)) for children younger than 18 years) or a self-report (Conners' Adult ADHD Rating Scales - Self-Report:Long Version (CAARS-S:L) (3) for children over 18 years. Parents, and

children if they were at least 12 years, were interviewed separately and were initially only administered the ADHD screening interview. Participants with elevated scores on any of the screen items of the K-SADS were administered the full ADHD section. A diagnostic algorithm was applied to determine diagnostic status based on a combination of symptom counts on the K-SADS and CTRS-R:L or CAARS-S:L, both providing operational definitions of each of the 18 behavioral symptoms defined by the DSM-IV (5). Symptom counts of the CTRS-R:L or CAARS-S:L were only used when at least two symptoms were reported on this questionnaire. The following scales of the Conners' ADHD questionnaires were used: DSM Inattentive behavior (scale L of the CPRS-R:L/CTRS-R:L; scale E of the CAARS-S:L), DSM Hyperactive/Impulsive behavior (scale M of the CPRS-R:L/CTRS-R:L; scale F of the CAARS-S:L), and DSM Total (scale N of the CPRS-R:L/CTRS-R:L; scale G of the CAARS-S:L). Participants with a symptom count of less than six symptoms of hyperactive/impulsive behavior and/or inattentive behavior were diagnosed with ADHD, provided they: a) met the DSM-IV criteria for pervasiveness and impact of the disorder (measures derived from the K-SADS), b) showed an age of onset before 12 years (following the proposed changes for the DSM-V; (2)), derived from the K-SADS, and c) the child received a T-score higher than 63 on at least one of the DSM ADHD scales on either one of the Conners' ADHD questionnaires. Unaffected participants were required to exhibit a T-score below 63 on each of the scales of each of the Conners' ADHD questionnaires, and have less than three symptoms derived from the combined symptom counts of the K-SADS and CTRS-R:L/CAARS-S:L. Criteria were slightly adapted for young adults, such that a combined symptom count of five symptoms was sufficient for a diagnosis (3). Young adults were considered unaffected when they exhibited less than two symptoms on the combined symptom counts. Cases that remained unclassifiable using the criteria above (N=64; 6%) were evaluated by a team of experts to derive a consensus diagnosis. Participants that did not fulfill criteria for either ADHD or unaffected status were labeled 'subthreshold ADHD'.

Participants were diagnosed with ODD if they exhibited four or more of the DSM-IV symptoms derived from the K-SADS. Likewise, CD was determined if a participant exhibited three symptoms or more DSM-IV symptoms derived from K-SADS interviews. Reading disorder was not diagnosed directly within the NeuroIMAGE project, but pre-existing diagnosis of reading disorder by a recognized medical institution were incorporated in the study design.

Participant inclusion

Inclusion criteria for participation in the MRI experiments consisted of the absence of claustrophobia and any metal in the body (e.g., braces). In accordance with ethics regulations, we obtained informed consent from participants above 16 years of age and of the parents of participating children under age 12. For participants between 12 and 16 consent was obtained from the participants as well as their parents.

Two hundred and eight participants with ADHD, 116 of their unaffected siblings, and 129 adolescents from control families subsequently successfully performed the Stop Signal Task in the fMRI scanner. Participants with a 'Subthreshold ADHD' diagnosis could not be unequivocally attributed to either group (n=53) and were therefore excluded from this study. Patients were required to withhold any form of psychoactive medication for 48 hours before the test day. Twenty participants with an ADHD diagnosis were entirely medication naive at time of testing. None of the siblings or controls was currently using stimulant medication.

Only participants who had completed three or four SST runs were included in further analyses. In total, 21 subjects who completed only three runs were included (twelve patients and six unaffected siblings). In addition, six participants did not reach an accuracy exceeding 70% on the go-trials and were excluded (four patients and two healthy controls); 11 participants were removed from the analysis due to excessive movement (defined as movement of more than one voxel in any direction within a run)

during scanning (nine patients and one unaffected sibling). Finally, 16 participants showed incidental neuroradiological findings (e.g., enlarged ventricles, subarachnoid cysts; ten patients and four unaffected siblings), and were excluded from further analysis. Accordingly, we included 185 patients, 111 unaffected siblings, and 124 controls in our analyses (see Main Text, Table 1.).

IQ estimation

The estimated IQ measure used during the NeuroIMAGE follow up was based on two subtests of the Wechsler Intelligence Scale for Children (WISC-III) or Wechsler Adult Intelligence Scale (WAIS-III), namely the vocabulary and block-design subtests. These subtests were selected to estimate verbal and spatial IQ respectively, with the mean of these two scores used as an estimate of the subject's full IQ score. Given the estimated nature of these IQ scores, they were not subsequently used for diagnostic or exclusion criteria, as performance on a full IQ test was already implemented to establish exclusion criteria during recruitment (4).

Stop signal task

The Stop Signal task was used to measure response inhibition in our participants (5). In this task, participants are required to respond as quickly as possible to visually presented GO stimuli (in this case, an aircraft pointing left or right). In 25% of trials the GO stimulus was followed by a visual stop signal, requiring participants to withhold their response. The stop signal consisted of a white cross superimposed on the GO stimulus. Importantly, the delay between presentation of the GO and STOP stimulus (the stop signal delay or SSD) was varied based on the participant's performance, to ensure each participant reached successful inhibition on 50% of Stop-trials. At the onset of the practice run the

SSD was set to 250 ms, after each successful inhibition the SSD was increased with 50 ms (making successful inhibition harder). In contrast, after each failed inhibition the SSD was decreased with 50 ms to facilitate inhibition on the next STOP trial. The SSD reached at the end of each run was forwarded to the next run. We administered one practice run of 60 trials (48 go- and 12 stop-trials) outside of the scanner and four runs of 60 trials during fMRI acquisition. If subject performance was below 25% after the second practice run, additional practice runs were obtained until they reached 25% successful inhibitions. All subjects entered into the analysis were verified to perform at around 50% accuracy during inhibitory trials.

After completion of the task we estimated the length of the inhibition process using the Stop-Signal Reaction Time (SSRT), which was calculated by subtracting the mean SSD from mean reaction time. Reaction time variability (RTV) and number of total omission and commission errors on GO-trials (Errors) were additional outcome measures of the stop signal task, and used to provide an indication of general task performance not necessarily related to the response inhibition process (6).

fMRI acquisition

Data were acquired at two sites on similar 1.5 Tesla Siemens scanners (Siemens Sonata at VU UMC in Amsterdam; Siemens Avanto at Donders Centre for Cognitive Neuroimaging in Nijmegen) using identical protocols. The SST was collected in four fMRI runs of 60 trial using a T2*weighted echo planar imaging sequence (TR=2340 ms, TE=40 ms, FOV=224x224 mm, 37 slices, voxel size=3.5x3.5x3.5 mm, 94 volumes per run). For spatial localization and normalization, we included each participant's high resolution MPRAGE T1 scan (TR=2730ms, TE=2.95ms, TI=1000ms, voxel size=1x1x1mm, FOV=256mm, 176 slices).

fMRI preprocessing

fMRI data were processed using FSL FEAT (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl; fMRI Expert Analysis Tool, version 6.0). Preprocessing included removal of the first four volumes of each run, within run motion correction to the middle volume, slice-timing correction, spatial smoothing using a 6mm Gaussian kernel, and highpass temporal filtering (0.01 Hz). For all runs we calculated transformation to the participant's T1 anatomical image using linear, boundary-based registration implemented in FSL-FLIRT.

fMRI single subject analysis

First, general linear models were constructed for each participant. Factors of interest were successful stop, failed stop and successful go-trials. Failed go-trials, movement trials (trials within an 8 second interval before movements exceeding 1 mm), signal from cerebral spinal fluid and white matter, and 24 realignment parameters were added as covariates. Single-subject beta-maps were transformed to participant-level anatomical space (3 mm isotropic resolution) using linear transformation matrices obtained via boundary-based registration in FSL, and combined across runs using a fixed effects model. This resulted in three participant-level contrast maps: (1) successful stop - go and (2) failed stop - go to isolate activation of successful and failed inhibition respectively, using go trial activity as an implicit baseline; as well as a (3) failed-successful stop contrast to model activation unique to the failed inhibition process. Custom group templates were used to calculate an anatomically neural 'midspace' for all group comparisons, to account for possible structural differences between diagnostic and gender groups.

Correcting for structural brain differences using custom MRI templates

For all group contrasts in fMRI activation, we used custom MRI templates to account for possible structural brain differences between our groups. Structural differences may be present

between ADHD patients and healthy controls (7–9), as well as between males and females (10; 11). By transforming all functional activation data to standard MNI152 space to enable group comparisons, it is possible that the needed transformation and non-linear warping differs between groups of interest. This can introduce perceived group differences in neural activation that are actually related to underlying structural brain differences and associated registration biases. To counteract such biases in our sample, we constructed specific templates for each subgroup. In addition, we transformed all participants' brains to a diagnosis- and sex-neutral 'midspace' (4). We transformed each participant's beta maps to one of four templates created in MNI152 space that matched with the participant's diagnostic and gender characteristics, i.e., ADHD-male, ADHD-female, unaffected-male, unaffected-female. Group level differences in functional activation were assessed by bringing all participants to a neutral 'midspace' using these four specific templates. Participant contrast maps were transformed to the midspace by concatenating the transformation from participant-level native space maps to that participant's specific sub template with the transformation of the specific template to our general study template and a weighted average of the transformation between the general study template and each specific template. The weighting of the latter transformation was done in accordance with the number of participants in each subgroup. This procedure effectively neutralizes possible registration biases between populations.

A-priori power estimates

Based on previous literature, power analyses has been conducted to estimate the number of subjects at which significant effects may be expected on behavioral and neural measures in the current study.

The power analysis for SSRT analyses were based on the review by Lipszyc (2010) and a large sample study by Bidwell (2007). These studies have reported effect sizes of G=0.62 and d=0.73 respectively when looking at SSRT differences between patients with ADHD and healthy controls, leading to an a-priori sample size of between 24 and 33 participants per diagnostic group to find a significant SSRT difference between patients with ADHD and controls. Unaffected siblings were also included in the Bidwell paper, with a reported d=0.51 effect size of SSRT differences between siblings and healthy controls, leading to a minimum group estimate of 49 participants per group. A-priori sample size estimation for fMRI data is unfortunately more complicated, since effect sizes are rarely mentioned in fMRI studies. Instead, only test statistics like z-values or even just p-values are reported in earlier literature on response inhibition. Since only peak statistics are usually reported, using these values for power calculations will necessarily underestimate the number of subjects required, and will not render more information about necessary sample sizes beyond the number of subjects that particular study included to detect a significant result (12). Based on previous publications on response inhibition in ADHD (13–16), sample sizes of between seven and 15 participants per group would be required to find a significant difference in neural activation between patients with ADHD and healthy controls. Earlier comparisons between unaffected siblings and healthy controls (16) indicate sample sizes between six and 16 subjects per group would be required to find significant differences in neural activation between these groups.

However, as the (17) paper demonstrates, the subject numbers in these studies do not provide enough power to sufficiently guarantee a lack of false positive results. Therefore, the incorporation of at least one hundred subjects in each group of current study should provide us with more than enough power to detect any difference between patients with ADHD, siblings and healthy controls on both behavioral and neural measures, and assure the validity of these results.

Supplementary Sensitivity Analyses

A series of additional sensitivity analyses were conducted to examine possible effects of differences between diagnostic groups in baseline neural activation, as well as for the additional factors of IQ, gender, comorbid disorders, task performance and familial relatedness within our sample. To ensure that none of the reported effects in the main body of the manuscript were dependant on differences between the groups in these factors, the main diagnostic group F-contrasts of the fMRI analysis were repeated with matched groups on each of these factors. Since this matching lead to the exclusion of a significant part of our sample, in particular the most severe patients with ADHD, the corrected threshold for which the effects are reported in these sensitivity analysis has been set at $p<.05$ instead of the $p<.01$ threshold reported in the main body of the manuscript.

Medication use, duration of use, and scan-site effects were also investigated in further sensitivity analyses. For these factors, the post-hoc Generalized Estimating Equations analyses reported in Table 2 of the main text have been repeated. In case of medication use and duration, these factors and their interaction terms were added to the GEE models to ensure medication status did not influence the main results. In case of scan-site, the GEE models have been repeated split by scan-site, to demonstrate that not one site was driving the reported results.

The outcome of these sensitivity analyses are all detailed in the supplementary results.

Supplementary Results

fMRI task activation

Table S1. Brain regions activated during the stop-signal task

Area	Color	Side [a]	Peak voxel			BA	# voxels [b]	Max F-value [c]
			x	y	z			
Stop-success contrast								
Supramarginal gyrus		R	58	-40	34	40,39	12658	11.5
Insular cortex, inferior frontal gyrus		R	32	20	-10	45,44, 13	11996	11.5
Occipital cortex		L	-32	-90	-2	18,19,22,40	9301	11.4
Insular cortex, inferior frontal gyrus		L	-30	18	-10	13, 47	1436	10.3
Frontal pole		L	-34	58	16	10	567	6.35
Anterior cingulate gyrus, preSMA		L	2	-24	28	23,6,8,9	459	7.06
Precuneus		R/L	8	-70	46	7	258	7.06
Temporal pole		R	48	8	-38	21	163	6.32
Stop-fail contrast								
Supramarginal gyrus		R	58	-40	34	40,22,18,19	23878	11.5
Inferior frontal gyrus		R	46	18	-2	13,44, 9, 10	7695	12.7
Anterior cingulate gyrus, preSMA		R/L	4	28	30	23, 24, 6, 8	6488	12.7
Insular cortex, inferior frontal gyrus		L	-30	18	-10	13, 45	2496	12
Frontal pole		L	-28	48	26	9, 10	864	7.94
Thalamus/caudate nucleus		R	12	-10	10		849	7.04
Precuneus		R/L	10	-72	44	19	713	7.91
Temporal pole		R	48	6	-34	21	298	7.44
Stop-fail – Stop-success contrast								
Calcarine occipital cortex		R/L	18	-66	8	17, 18	6163	7.01
Inferior frontal gyrus		L	-46	12	0	44, 45	4555	7.2
Anterior cingulate gyrus, preSMA		R/L	-2	18	36	32, 6	4429	8.3

Note: preSMA = pre-supplementary motor area; BA = Brodmann area.
[a] Side indicates the hemisphere (left/right).
[b] # voxels indicates the number of voxels in a cluster.
[c] Correction for multiple comparisons applied using a cluster threshold of $z > 2.6$ and significance threshold of $p<.01$ corrected.

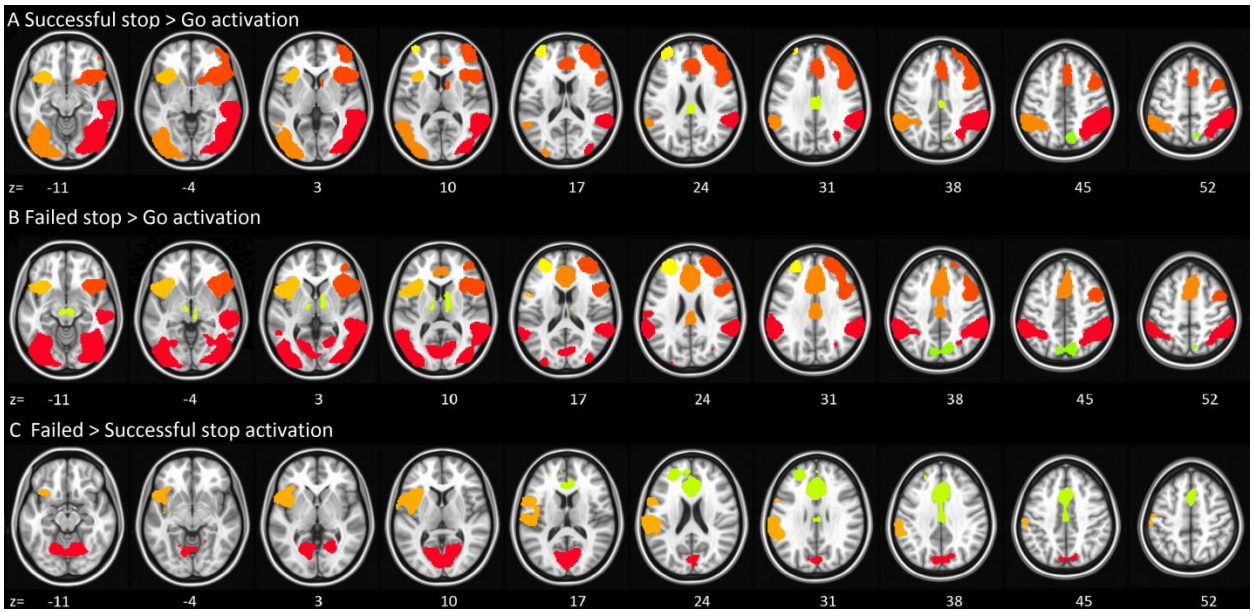

Figure S1: (A) Successful stop > go network: neural activation across all groups. (B) Failed stop > go network: neural activation across all groups. (C). Failed stop > successful stop network: neural activation across all groups. Distinct colors refer to distinct statistically significant clusters, correspondence between colors and anatomical labels can be found in table 2.

Associations between group differences in brain activation and stop-task outcome measures

Table S2. Associations between brain activation and task outcome measures

Area	side	RTV				SSRT				Errors			
Stop-success network		*B* [a]	Wald-χ^2 [a]	Cohen's d [a]	p-value [a]	*B*	Wald-χ^2	Cohen's d	p-value	*B*	Wald-χ^2	Cohen's d	p-value
Inferior frontal cortex	L	-.055	2.357	.15	.125	**-.068**	**7.885**	**.277**	**.005**	-.174	.498	.069	.481
Superior frontal gyrus	L	-.117	4.736	.214	.030	**-.119**	**13.831**	**.369**	**<.001**	.403	1.464	.118	.226
Supramarginal gyrus	L	.037	.721	.083	.396	.042	1.578	.123	.209	.143	.283	.052	.595
Postcentral gyrus	R	-.029	.466	.067	.495	-.010	3.337	.179	.660	.602	.092	.03	.762
Temporal-parietal junction	R	.007	.066	.025	.797	-.014	.549	.072	.459	-.231	1.270	.110	.260
Stop-fail network													
Inferior frontal cortex	L	-.036	1.635	.125	.201	-.053	6.282	.246	.012	-.511	4.278	.203	.039
Temporal-parietal Junction	L	-.011	1.151	.105	.283	-.020	1.601	.124	.206	-.350	.255	.049	.613
Temporal-parietal Junction	R	-.026	.105	.032	.746	-.020	.889	.092	.346	-.104	2.318	.149	.128
Superior frontal gyrus	L	-.041	1.266	.11	.260	.004	.030	.017	.862	-.397	3.981	.196	.046
Anterior cingulate cortex	L/R	.014	.177	.041	.674	-.012	.186	.042	.666	-.524	2.264	.147	.132
Supramarginal gyrus	L	-.077	3.426	.181	.064	-.053	2.574	.157	.109	-.036	.015	.012	.904

Note: SSRT = Stop-signal reaction time; RTV = Reaction time variance; Errors = Number of errors on go-trials.
Bolded values indicate significant effects.
[a] Reported statistics indicate the association between stop-task outcome measures and neural activation. All measures derived from a single generalized estimating equations model for familial dependency between siblings, as well as for covariates age, gender, IQ, and scan site. A Bonferroni-Holm corrected threshold of adjusted p-values was used to correct for multiple comparisons.

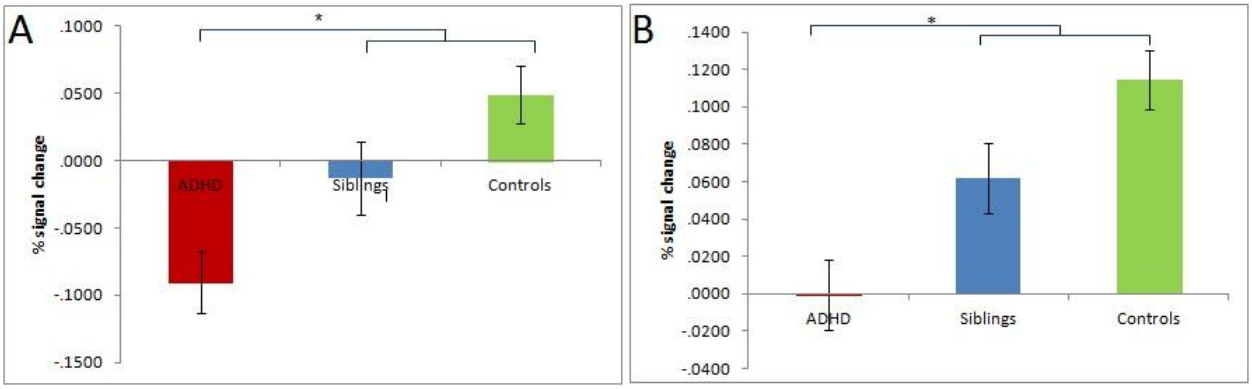

Figure S2: Percentage signal change differences between diagnostic groups during successful stop condition in left superior frontal gyrus (A) and left inferior frontal gyrus (B).

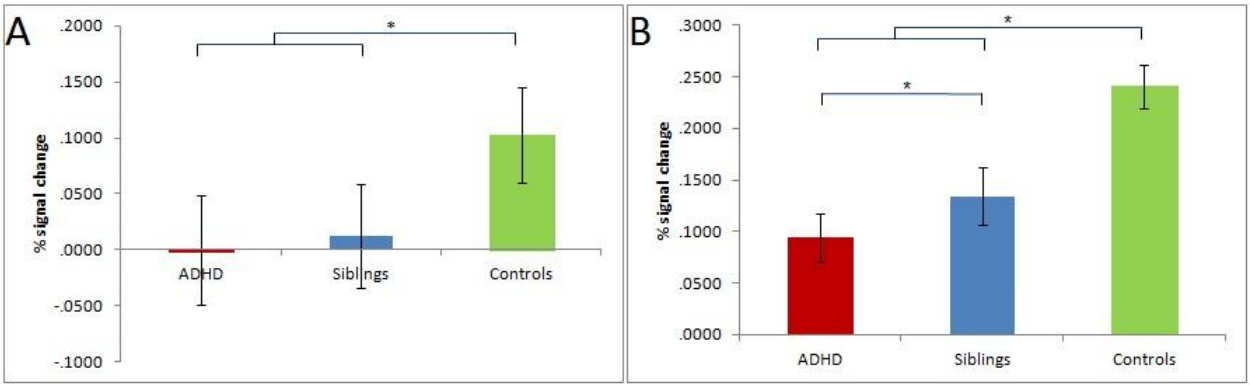

Figure S3: Percentage signal change differences between diagnostic groups during failed stop condition in left superior frontal gyrus (A) and left inferior frontal gyrus (B).

Group differences in fMRI task activation in the GO network

The neural activation during successful go trials was used as a baseline condition in the main contrasts of interest, i.e. successful stop – go and failed stop – go. To ensure that there were no structural differences in go network activation that underlie the results from our contrasts of interest, we tested the difference between ADHD patients, unaffected siblings and healthy controls on successful go trials. At a cluster level threshold of p<.01 we detected one significant cluster in the frontal pole (see Figure S1). This cluster did not overlap with any of the group effects found in the successful or failed stop contrasts reported in the main text, and did not survive correction for Family Wise Errors (FWE).

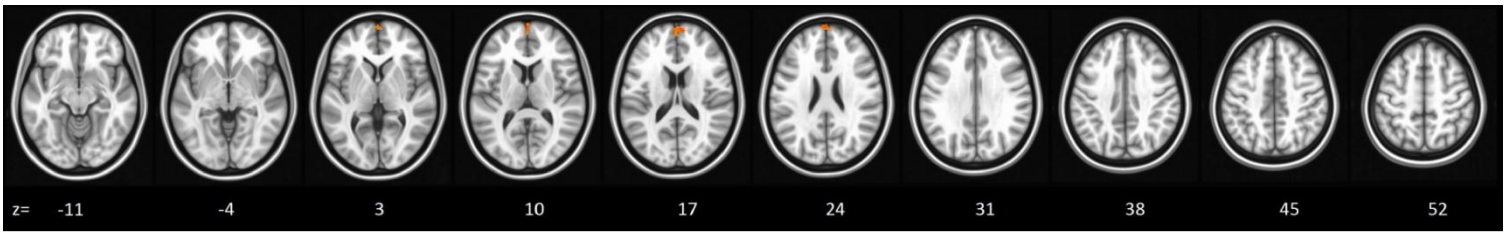

Figure S4. Go network: Brain activation differences between controls and siblings or ADHD patients. Red hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain.

Additionally, the main between group differences observed in the superior and inferior frontal gyri during both successful and failed inhibition conditions (as presented in Figures 3 and 5 of the main text) are presented below split within the constituent 'go' and 'stop' trials. The presented raw beta-values are exported from the same group level model as the contrast conditions of interests presented in the main text, and therefore differ in the precise statistical estimation of the averages and variance from the contrast conditions. Nevertheless, these Supplementary Figures 2 and 3 demonstrate that the group differences are mostly due to decreased deactivation of these regions during stop trials in the control group, which is mostly absent in the probands with ADHD.

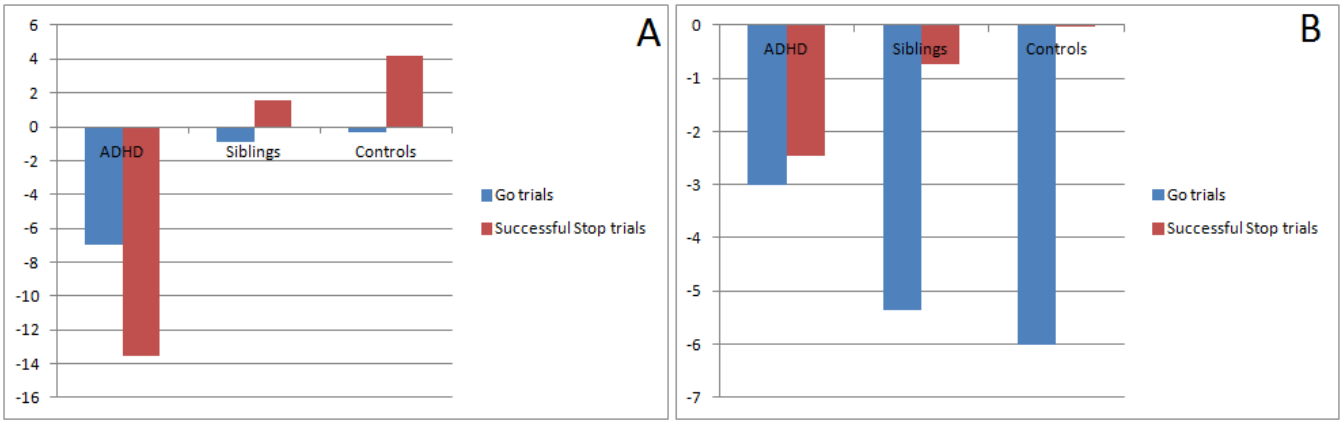

Figure S5. Between group neural activation differences in superior frontal gyrus (A) and inferior frontal gyrus (B) as observed in the successful stop > go contrast depicted in Figure 3 of the main text. Blue bars represent the average beta value during go trials; red bars represent successful stop trials.

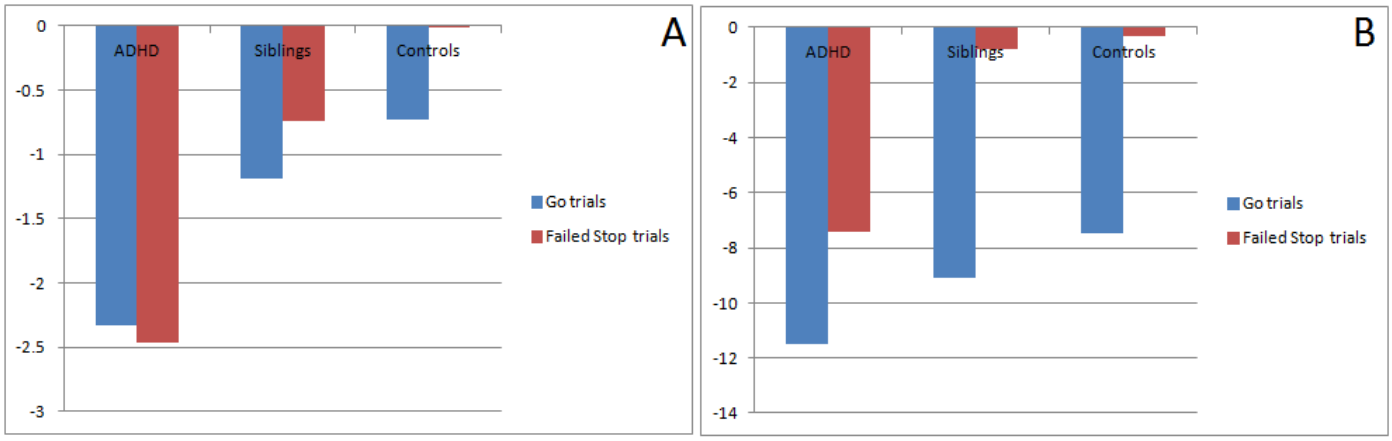

Figure S6. Between group neural activation differences in superior frontal gyrus (A) and inferior frontal gyrus (B) as observed in the failed stop > go contrast depicted in Figure 3 of the main text. Blue bars represent the average beta value during go trials; red bars represent failed stop trials.

fMRI task activation per diagnostic group

In Figures 2 and 4 of the main text we showed patterns of decreased activation in ADHD patients and their unaffected siblings in several nodes within the response inhibition and ventral attention networks. To demonstrate that these patterns were indeed caused by hypoactivation of these nodes and are not representative of a difference in topographical structuring of the underlying networks, the task activation maps for the two conditions of interest are presented per diagnostic group (see Figures S4, S5). At a FWE corrected threshold of p<.01 these analyses showed qualitatively similar distributions of the general task activation networks for each of the diagnostic groups, with activation nodes in bilateral inferior frontal, superior frontal, supramarginal and temporal/parietal areas, temporal pole, occipital and cingulate cortices.

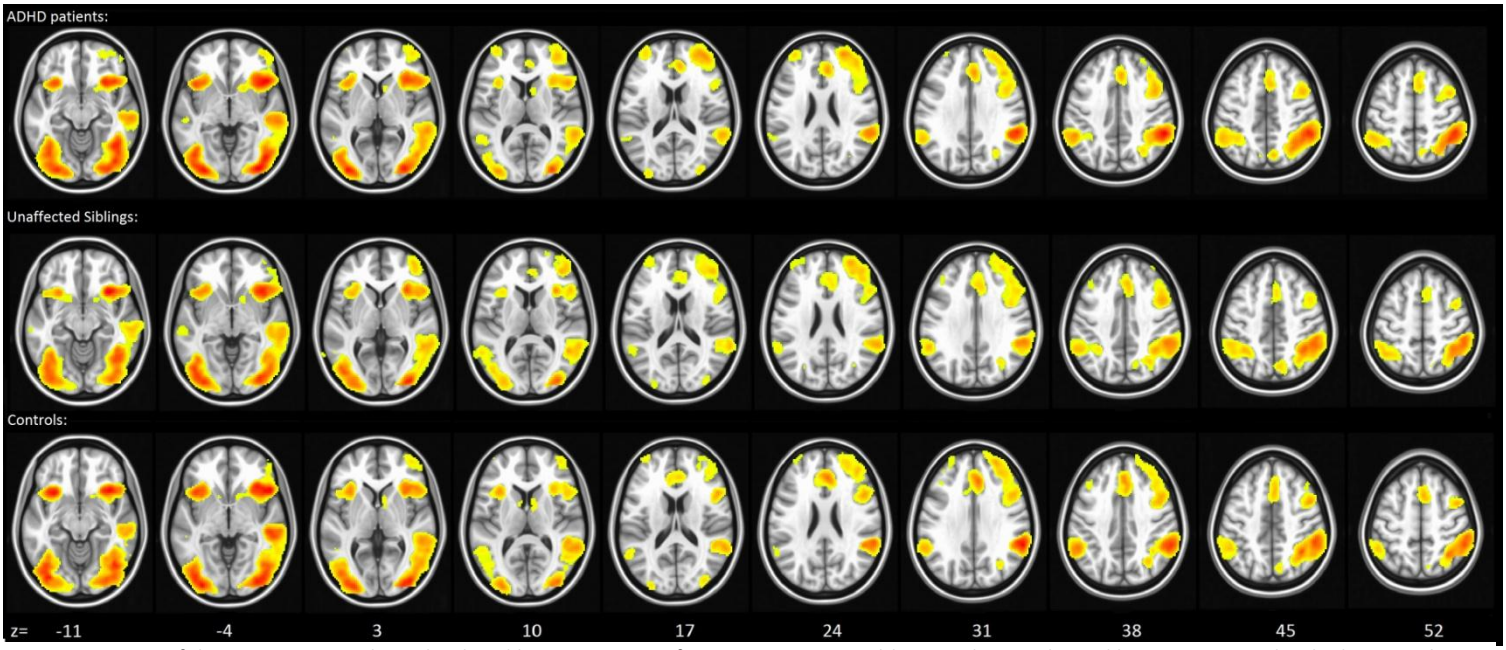

Figure S7. Successful stop - go network: Task related brain activation for ADHD patients, siblings and controls. Red hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain.

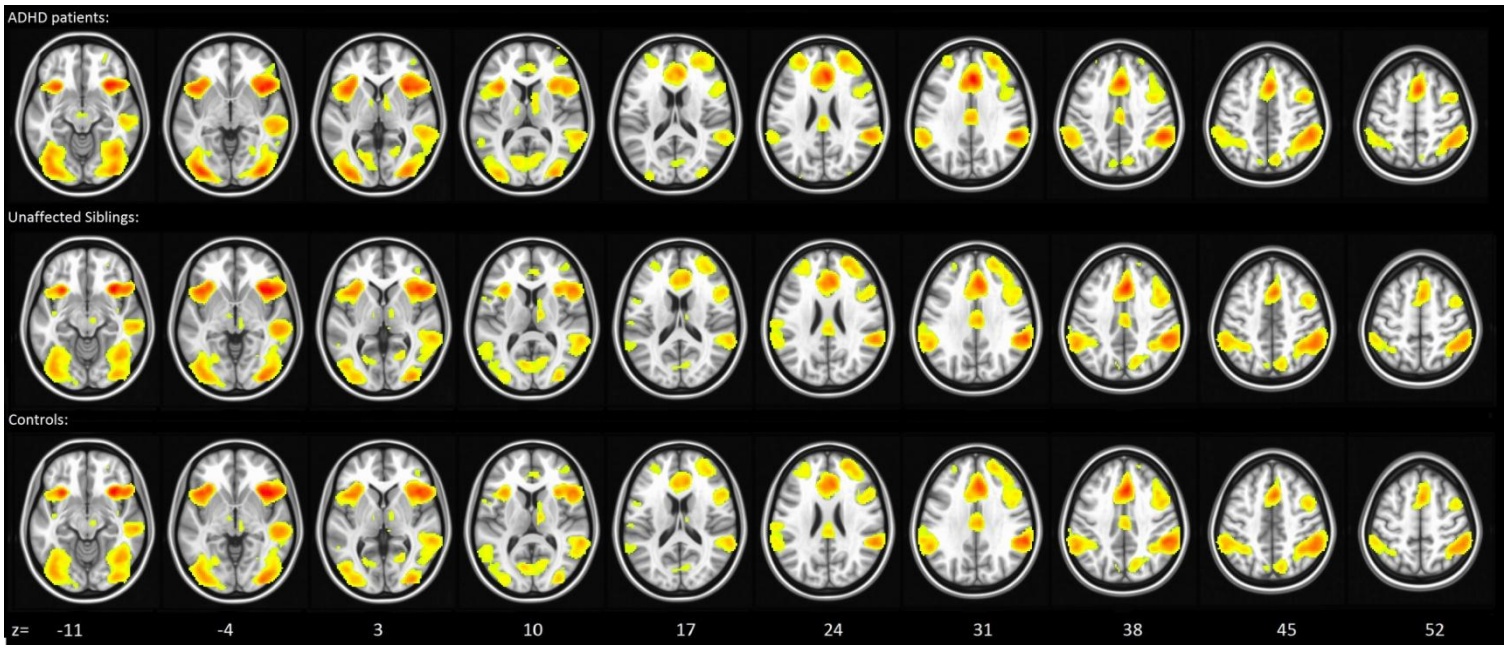

Figure S8. Failed stop - go network: Task related brain activation for ADHD patients, siblings and controls. Red hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain.

Group differences in fMRI task activation, matched for IQ and gender and scan-site

The distribution of IQ and gender differed between patients, unaffected siblings and healthy controls in this study. In addition to covarying IQ and gender in all main analyses, we repeated the main fMRI analyses of the diagnostic group contrast of interest in subgroups of our sample matched on IQ and gender, respectively. IQ matched groups were achieved by subdividing the IQ scores into six bins of 15 points each. Subsequently, the ratio of patients, siblings, and controls was equalized to the mean of the entire sample for each bin. This led to the exclusion of 46 patients and 16 siblings from the lower three bins, as well as 30 controls and 12 siblings from the upper three. Gender matched groups were achieved by equalizing the ratio of females in the unaffected sibling and controls groups to the patients. This lead to the exclusion of 42 females from the unaffected siblings and 45 females from the controls. At a corrected threshold of $p<.05$ these analyses showed activation in the same clusters as in our main analyses with unmatched groups, with significant clusters in the left inferior frontal, superior frontal,

supramarginal and temporal/parietal areas (see Figures S6, S7 for IQ and gender matched analyses respectively).

Demographic factors were not a-priori matched across scan sites (4); even though no main effects of scan-site or any interaction effects between scan-site and diagnostic group effects were found in any of the neural activation analyses, there may still have been sub-threshold effects of scan-site that influenced the group effects reported in the main text. In order to account for the influence of possible mechanical or demographic differences between scan-site, all the post-hoc tests detailing the diagnostic effects on neural activation were repeated separated by scan site (see Table S1). These results showed that the direction of the neural effects is generally identical across sites, and the large majority remains significant in both samples despite the loss of power by this split.

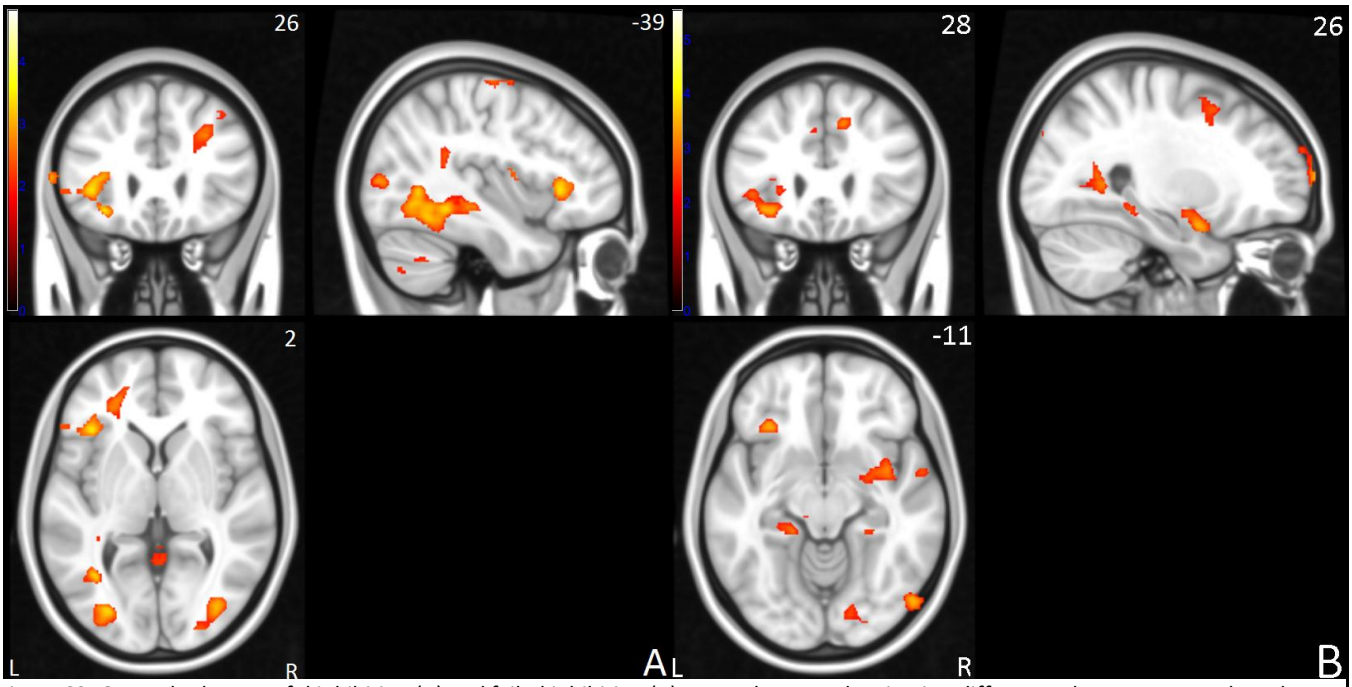

Figure S9.IQ matched successful inhibition (A) and failed inhibition (B) networks: Neural activation differences between controls and siblings or ADHD patients matched for IQ. (corrected p-value <.05). Yellow hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain

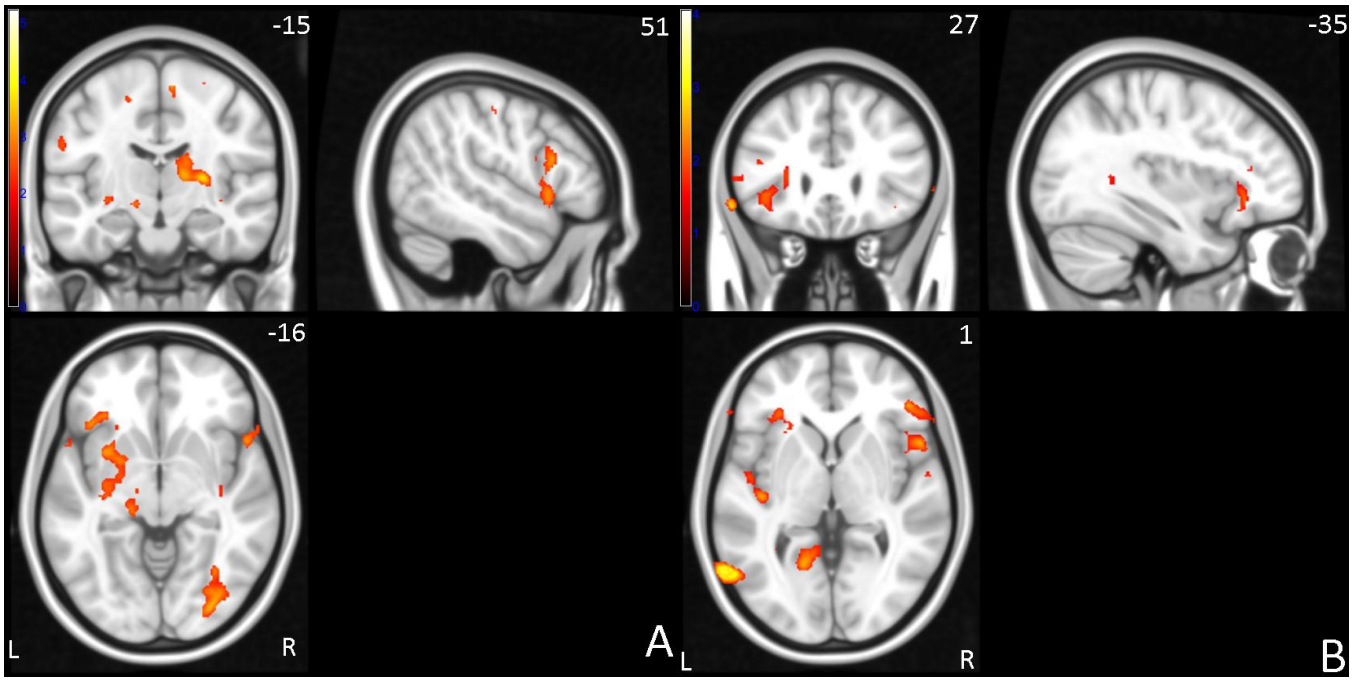

Figure S10. Gender matched successful inhibition (A) and failed inhibition (B) networks: Neural activation differences between controls and siblings or ADHD patients matched for gender (corrected p-value <.05). Yellow hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain

Table S3. Group differences in neural activation split by scanner location

Nijmegen

Stop-success contrast	side [a]	Wald-chi^2 [b]	p-value [b]	between group effects	*B* values (Controls vs. ADHD) [c]
Inferior Frontal gyrus	L	21.039	<.001	Controls = Sibs > ADHD	-17.13
Superior Frontal gyrus	L	12.46	0.002	Controls = Sibs > ADHD	-22.54
Supramarginal gyrus	L	10.369	0.006	Controls = Sibs > ADHD	-13.99
Postcentral gyrus	R	2.7	0.259		-6.81
Temporal-parietal junction	R	10.7	0.005	Controls = Sibs > ADHD	-9.92
Stop-fail contrast					
Inferior frontal gyrus	L	21.119	<.001	Controls > Sibs > ADHD	-16.28
Temporal-parietal junction	L	8.298	0.016	Controls = Sibs > ADHD	-9.72
Temporal-parietal junction	R	19.716	<.001	Controls > Sibs > ADHD	-10.71
Superior frontal gyrus	L	6.317	0.042	Controls > Sibs = ADHD	-10.64
Anterior cingulate cortex	L/R	8.046	0.018	Controls > Sibs = ADHD	-11.57
Supramarginal gyrus	L	6.275	0.043	Controls = Sibs > ADHD	-12.64

Amsterdam

Stop-success contrast	side	Wald-chi^2	p-value	between group effects	
Inferior Frontal gyrus	L	3.031	0.22		-6.5
Superior Frontal gyrus	L	0.997	0.607		-5.8
Supramarginal gyrus	L	9.99	0.007	Controls = Sibs > ADHD	-17.43
Postcentral gyrus	R	12.32	0.002	Controls = Sibs > ADHD	-19.28
Temporal-parietal junction	R	9.433	0.009	Controls = Sibs > ADHD	-10.5
Stop-fail contrast					
Inferior frontal gyrus	L	20.019	<.001	Controls > Sibs > ADHD	-13.15
Temporal-parietal junction	L	26.954	<.001	Controls = Sibs > ADHD	-20.85
Temporal-parietal junction	R	17.831	<.001	Controls = Sibs > ADHD	-10.24
Superior frontal gyrus	L	8.425	0.015	Controls > Sibs = ADHD	-10.96
Anterior cingulate cortex	L/R	4.931	0.085		-9.75
Supramarginal gyrus	L	7.245	0.027	Sibs > ADHD	-8.45

Note: ADHD = attention deficit/hyperactivity disorder

[a] Side indicates the hemisphere (left/right).

[b] Significant clusters are derived from the diagnostic group F-contrast, Wald-χ^2 and p-values reflect the effect of diagnostic group in each region as derived from generalized estimating equations model, corrected for familial dependency, as well as for covariates age, gender, IQ and scan site.

[c] *B* values indicate the direction and size of post-hoc Control vs. ADHD contrast

Group differences in fMRI task activation, corrected for medication duration

The influence of medication use on the reported diagnostic groups effects on neural activation was investigated using two sets of Generalized Estimating Equations models, both repeating all the post-hoc analyses detailing the group effects on neural activation. The first set of models included a categorical factor for medication use status (currently using stimulant medication or no current medication use). This analysis indicated no main effects of medication status, nor any interaction effects between medication status and diagnostic group. The second set of models included a continuous variable detailing the duration of medication use in months for each subject. This analysis showed a main effect of medication duration on anterior cingulate activation during failed stop-trials (B=-.174, p<.014). However, there was no interaction between the effect of medication duration and diagnostic group, and the effect of diagnostic group remained present with the addition of the medication variable.

Group differences in fMRI task activation, matched for ODD, CD and RD

The comorbidity of ADHD with other disorders may provide additional difficulties when interpreting group differences between patients with ADHD, unaffected siblings and healthy controls, several reviews have reported on response inhibition deficits in these disorders, either on their own (18) or in combination with ADHD (19). However, since these comorbid conditions may share a neurobiological and genetic basis with ADHD (20), and also in order to retain the generalisability of our sample, we chose not to exclude patients with ADHD and oppositional defiant disorder (ODD), conduct disorder (CD) or reading disorder (RD). To ensure that our reported effects were not driven solely by the presence of these disorders, three distinct sensitivity analyses are reported, in which we repeat the main fMRI analyses of the diagnostic group contrast with all participants showing comorbid ODD, CD or RD removed from the sample, respectively.

For the analysis without any ODD presence, 55 patients with ADHD and four unaffected siblings were removed (see Figure S8). For the analysis without any CD presence, twelve patients with AHD were removed (see Figure S9). For reading disorder, 34 patients with ADHD, eleven unaffected siblings and eleven healthy controls were removed (see Figure S10). At a corrected threshold of p<.05 these analyses showed activation in the same clusters as in our main analyses with unmatched groups, with significant activation in the left inferior frontal, superior frontal, supramarginal and temporal/parietal areas.

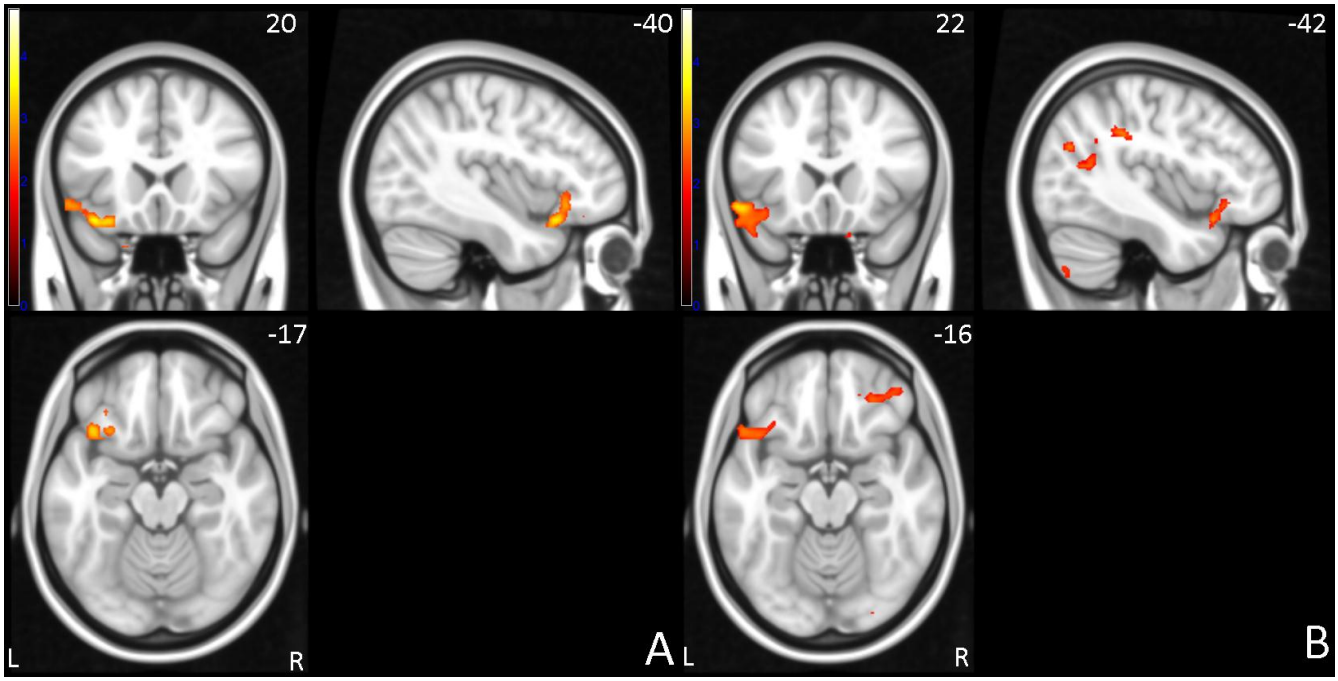

Figure S11. ODD matched Successful inhibition (A) and failed inhibition (B) networks: Neural activation differences between controls and siblings or ADHD patients with all comorbid oppositional defiant disorder (ODD) patients removed (corrected p-value <.05). Yellow hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain

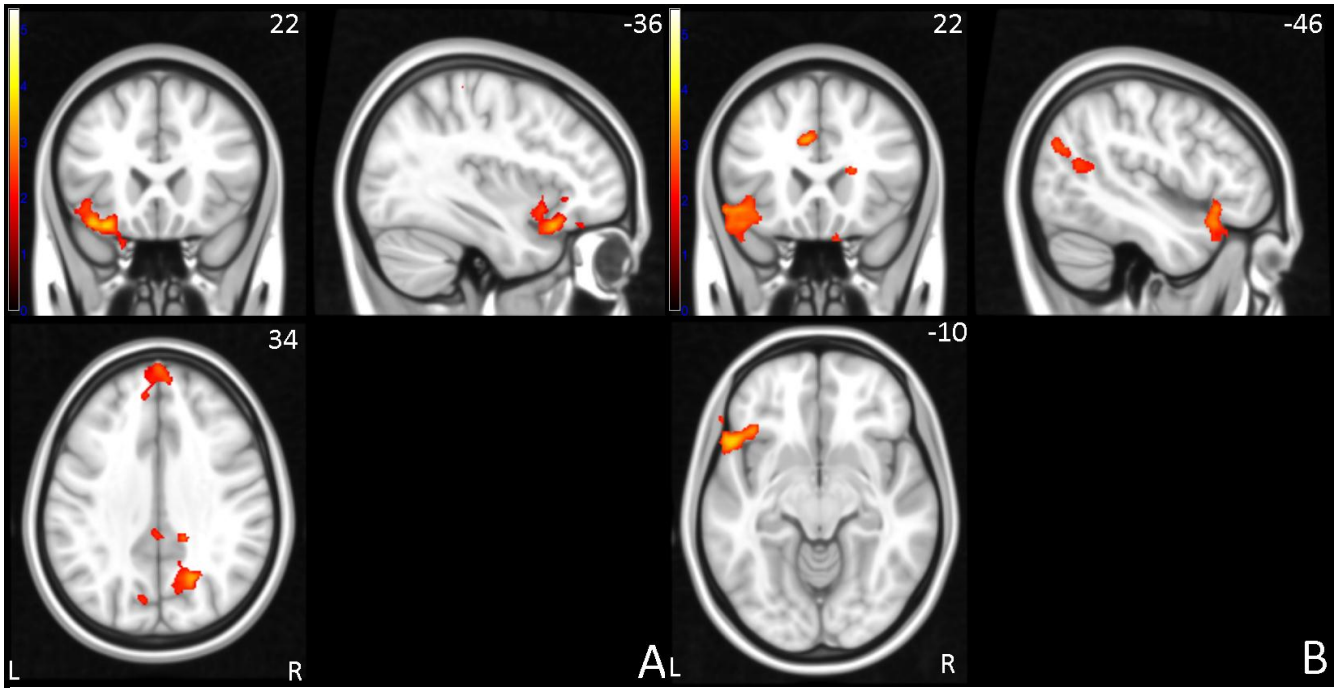

Figure S12. CD matched successful inhibition (A) and failed inhibition (B) networks: Neural activation differences between controls and siblings or ADHD patients with all comorbid conduct disorder (CD) patients removed (corrected p-value <.05). Yellow hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain

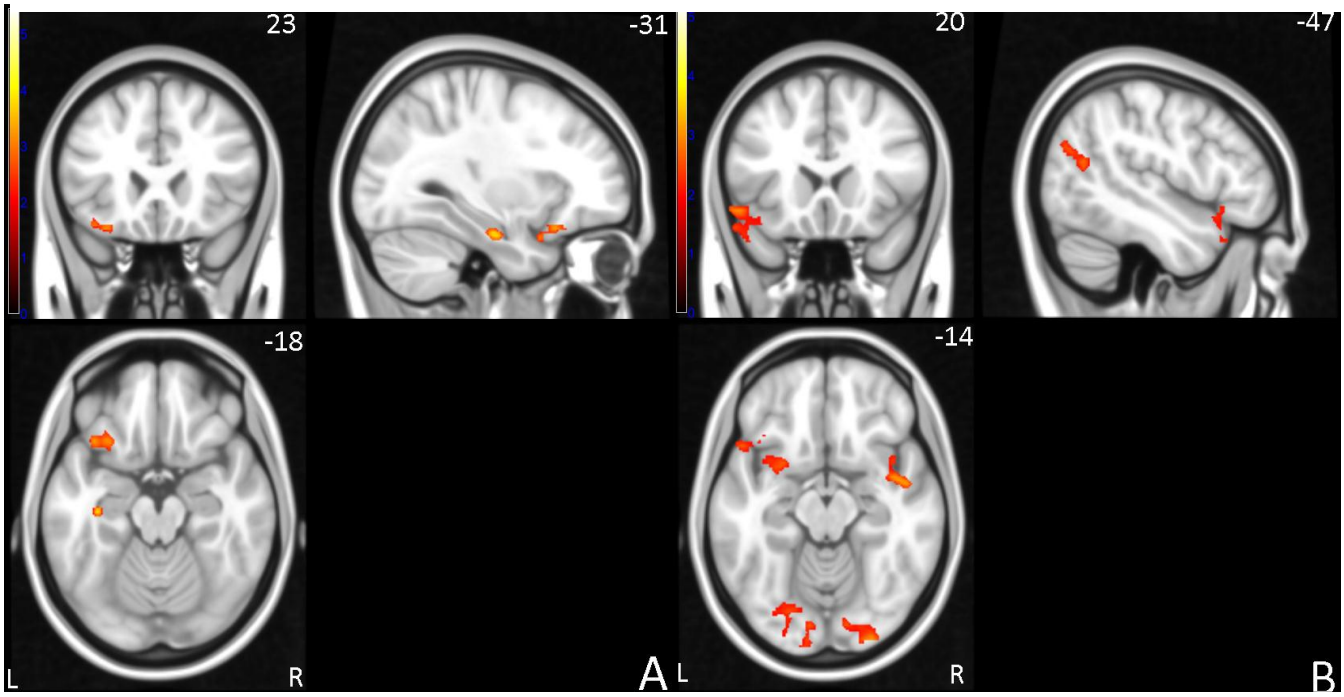

Figure S13. RD matched successful inhibition (A) and failed inhibition (B) networks: Neural activation differences between controls and siblings or ADHD patients with all comorbid reading disorder (RD) patients removed (corrected p-value <.05). Yellow hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain

Group differences in fMRI task activation, after correction for SSRT length

To demonstrate that neural measures of response inhibition are more sensitive to detect susceptibility for ADHD, it should additionally be shown that the neural differences between diagnostic groups remain when correcting for SSRT differences. Three additional sensitivity analyses were performed to support this. First, the relation between stop task outcome measures and neural activation measures were repeated split by diagnostic group (see SI Table 2). These analyses showed that the relation between SSRT and inferior frontal activation holds across patients with ADHD and healthy controls in both the successful stop (B=-.422, p<.001; B=-.402, p<.05 respectively) and failed stop (B=-.422, p<.001; B=-.402, p<.05 respectively) conditions.

Second, the main fMRI analysis of the diagnostic group contrast was repeated for subgroups matched on SSRT performance. To obtain these groups, the SSRT values were divided into six bins of one SD width. Within each bin, the ratio of controls, unaffected siblings and patients with ADHD was equalized to the mean of the entire sample. This lead to the exclusion of 15 control subjects and 16 unaffected siblings from the three bins with lowest SSRTs, and the exclusion of 22 patients with ADHD from the three bins with highest SSRT scores. The results of these analyses are shown in Figure S11, at a threshold of p<.05 these showed a similar pattern of results as the main analysis, with significant group differences in nodes in the inferior frontal, superior frontal and temporal/parietal areas (see Figure S11). This suggest that neural activation is indeed more closely related to familial risk factors underlying ADHD than just the behavioral measures, and is therefore a more sensitive measure for dissociating diagnostic groups.

Lastly, we compared the tail of the SSRT distribution between the three diagnostic groups, by describing the percentage of patients with ADHD and their siblings with a score below the 10^{th} percentile of the score distribution in healthy controls. This analysis indicated that 12% of patients with

ADHD and 7% of their siblings have a score below the 10th percentile of healthy controls. These data indicate that the spread of SSRT values is relatively homogenous across groups and that the lower SSRT scores in ADHD are not the result of a skewed tail of the SSRT distribution, i.e., only a subsample of ADHD patients appears severely impaired on response inhibition performance. These data support the previous findings of large inter-individual variation of SSRT scores within patient groups (19), and argue against the utility of SSRT as an independent endophenotype for ADHD. The neural activation data were consistently more skewed, an average of 22% of patients and 19% of the siblings showed activation below the 90th percentile of controls, indicating a more robust deviation in neural activation in probands with ADHD (see SI Table 3).

Both reaction time variability and error rates were also more consistently affected in patients with ADHD, to a similar degree as the neural activation measures. However, for each of the investigated measures only a subsample of patients with ADHD shows scores strongly deviant from the control sample.

Table S4. Relations between brain activation and SSRT length split by diagnostic group

		ADHD			Siblings			Controls		
		B [a]	Wald-χ^2 [a]	p-value [a]	B	Wald-χ^2	p-value	B	Wald-χ^2	p-value
Inferior Frontal cortex	L	-0.422	**8.744**	**0.004**	-0.78	0.24	0.625	-0.402	**3.278**	0.05
Superior Frontal Gyrus	L	-0.143	2.583	0.108	-0.136	1.008	0.315	-0.294	**5.496**	0.019

Note: ADHD = attention deficit/hyperactivity disorder; SSRT = stop-signal reaction time;
[a] Reported B, Wald-χ^{22} and p-values reflect the effects the effect of neural activation measures on SSRT length, derived from generalized estimated equations models corrected for familial dependency between siblings, as well as for covariates age, gender, IQ and scan site.

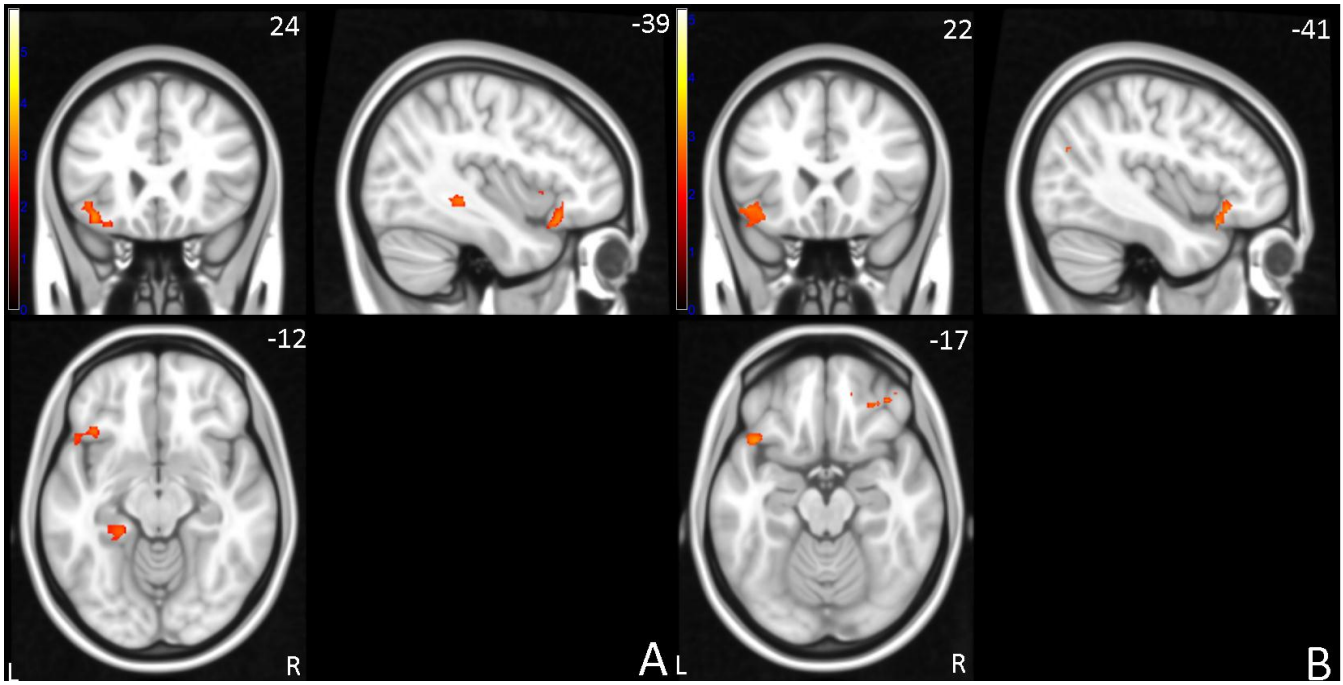

Figure S14. SSRT matched successful inhibition (A) and failed inhibition (B) networks: Neural activation differences between controls and siblings or ADHD patients matched on SSRT performance (corrected p-value <.05). Yellow hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain

Table S5. Percentage of patients with ADHD scoring above the 90th percentile of controls

Stop-task outcome	ADHD %	Siblings %
SSRT	12	7
Errors	27	16
ICV	31	12
Area		
Inferior frontal gyrus	25	23
Superior frontal gyrus	24	25
Supramarginal gyrus	18	18
Postcentral gyrus	23	14
Temporal-parietal junction	23	14
Inferior frontal gyrus	18	20
Temporal-parietal junction	16	14
Temporal-parietal junction	22	17
Superior frontal gyrus	21	23
Anterior cingulate cortex	31	28
Supramarginal gyrus	25	18

Note: ADHD = attention deficit/hyperactivity disorder; SSRT = Stop-signal reaction time; RTV = Reaction time variance; Errors = Number of errors on go-trials.

Group differences in fMRI task activation, corrected for familial influences.

The NeuroIMAGE sample is a cohort of families with and without a history of ADHD. Multiple siblings per family were included. Sibling pairs were either designated to the same diagnostic group, or divided between patients with ADHD and unaffected siblings. Since siblings are more similar than unrelated participants, this can lead to an overestimation of the similarity between the patients with ADHD and unaffected siblings, and an overestimation of the difference between the latter two groups and healthy controls. FSL (as well as other fMRI analysis tools) do not enable direct correction for these familial relations as other statistical packets do. Therefore, in the main text of the manuscript only the outcome values of corrected post-hoc statistical tests were reported. Nonetheless, in order to ensure these familial relations did not influence the results in the main text, all fMRI analyses of the main

diagnostic group contrast were repeated within a subsample where only one child per family was selected at random from the full sample. This lead to the exclusion of 49 patients with ADHD, 65 unaffected siblings and 46 healthy controls. The results of these analyses are shown in figure S12, and show similar patterns of activation as the main analyses in both conditions, with between-group activation differences in inferior frontal, superior frontal and temporal/parietal areas (results use a corrected threshold of p<.05).

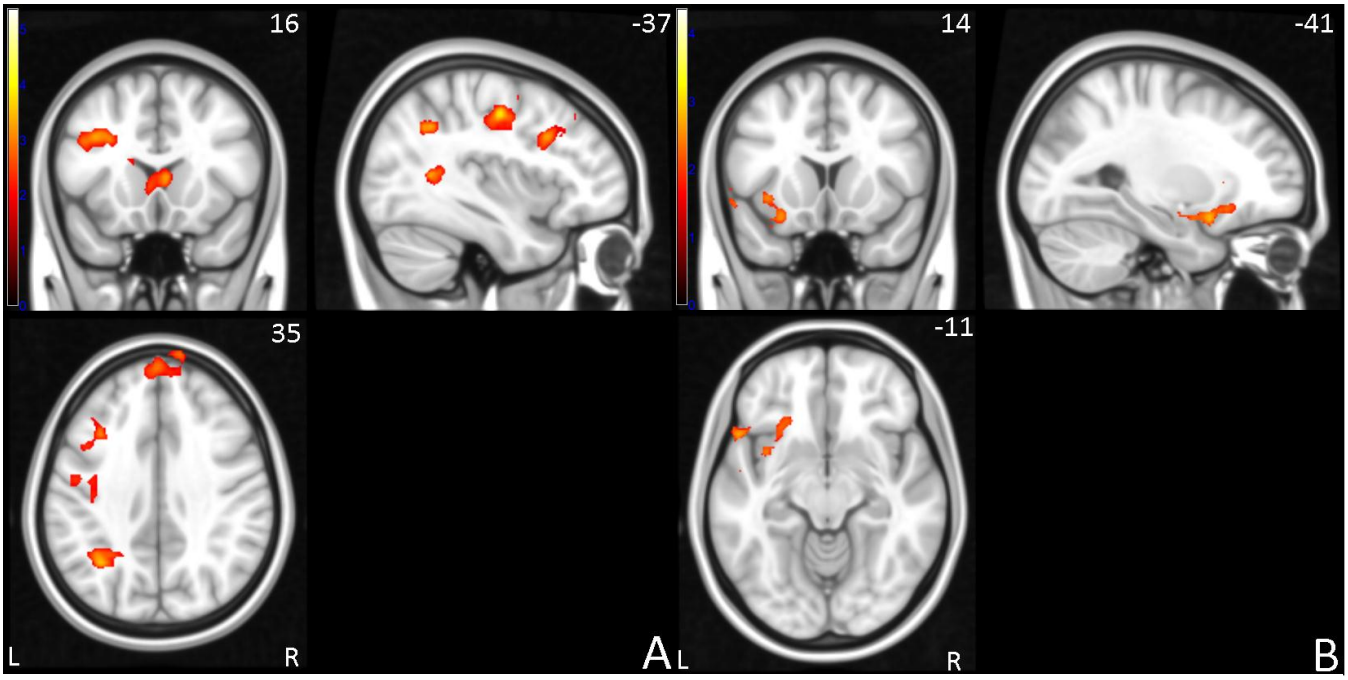

Figure S15. Sibling matched successful inhibition (A) and failed inhibition (B) networks: Neural activation differences between controls and siblings or ADHD patients with only one sibling per family included (corrected p-value <.05). Yellow hues correspond to higher signal in control subjects. Right side of the image corresponds to the right hemisphere of the brain

<u>Differences between unaffected siblings and healthy controls, corrected for subthreshold symptoms.</u>

Unaffected siblings and controls are both defined by the presence of two or less DSM-IV symptoms for ADHD, as measured by interviews and questionnaires. Nevertheless, there may have been features of ADHD within the unaffected siblings that did not reach the threshold to be counted as an official DSM-IV symptom. These subthreshold symptoms may have been picked up by elevated scores

on the Conners' Questionnaires. In order to account for the possibility that the differences between siblings and healthy controls are due to the presence of subthreshold ADHD symptoms, an additional sensitivity analysis was run and added to the SI. These analyses suggest a significantly higher average Conners z-score for unaffected siblings than healthy controls (mean z=-0.59, sd=0.45 for siblings, mean z=-0.74, sd=0.32 for controls; B=.153, p=.003) (see SI Table 4). Subsequently, average Conners scores were added to all neural and behavioral analyses where differences between siblings and controls were detected. The post-hoc contrasts between controls and siblings have been replicated for all these analyses (see SI Table 5). Most effects reported in the manuscript did not change in terms of statistical significance after these corrections. Only the difference between unaffected siblings and controls found in the left supramarginal and right temporal/parietal areas during the successful stop condition dropped below significance. The outcome of this analysis is reported in the results section of the manuscript. It is therefore unlikely that the reported differences between unaffected siblings and controls are influenced by subthreshold symptoms of ADHD

Table S6. Group differences in total Conners scores

	Mean z-score [a]	SD	95% confidence interval	
ADHD	0.78	0.51	0.68	0.88
Siblings	-0.59	0.45	-0.63	-0.44
Controls	-0.74	0.32	-0.78	-0.66

Note: ADHD = attention deficit/hyperactivity disorder; SD = standard deviation.

[a] mean z-scores indicate the combined scores on the Conner's Parent Rating Scales (CPRS) and Conners' Adult ADHD Rating Scale (CAARS) or Conners' Teacher Rating Scales (CTRS).

Table S7. Unaffected siblings versus healthy control effects controlled for Conners scores.

	Siblings		Controls			
	Mean	SD	Mean	SD	B [a]	P [a]
RTV	95.18	3.73	86.06	3.38	9.12	**.027**
Errors	4.64	0.67	3.44	0.56	1.2	**.047**
Stop-success contrast Activation						
L Supramarginal gyrus	-6.18	3.37	0.27	3.52	-6.46	.109
R Postcentral gyrus	-17.81	3.26	-8.97	3.64	-8.842	**.041**
R Temporal-parietal junction	18.57	2.42	22.3	2.08	-3.72	.174
Stop-fail contrast Activation						
L Inferior frontal gyrus	14.19	2.21	25.14	2.4	-10.95	**<.001**
L Temporal-parietal junction	15.95	2.59	25.33	2.58	-9.38	**0.005**
R Temporal-parietal junction	13.23	1.62	20.01	1.57	-6.77	**<.001**
L Superior frontal gyrus	-0.3	2.59	9.84	2.71	-10.146	**0.004**
Anterior cingulate cortex	20.26	2.63	28.5	2.99	-8.24	**0.014**

Note: RTV = Reaction time variance; Errors = Number of errors on go-trials; SD = standard deviation.

[a] B and p-values are derived from generalized estimating equations models corrected for familiality and including IQ, Age, gender and scan-site as covariates.

References

1. Kaufman C, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, Williamson D, Ryan N: 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

2. American Psychiatric Association: Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition, Text Revision (DSM-IV-TR). Arlington, VA, American Psychiatric Association; 2000.

3. Kooij SJJ, Buitelaar JK, Van den Oord EJ, Furer JW, Rijnders CAT, Hodiamont PPG: Internal and external validity of Attention-Deficit Hyperactivity Disorder in a population-based sample of adults. Psychol Med 2005; 35:817–827

4. Von Rhein D, Mennes M, Van Ewijk H, Groenman AP, Zwiers M, Oosterlaan J, Heslenfeld D, Franke B, Hoekstra PJ, Faraone S V, Hartman C, Buitelaar J: The NeuroIMAGE study: a prospective phenotypic, cognitive, genetic and MRI study in children with attention-deficit hyperactivity disorder. Design and descriptives. Eur Child Adolesc Psychiatry in press

5. Logan GD, Cowan WB, Davis KA: On the ability to inhibit simple and choice reaction time responses: a model and a method. J Exp Psychol Hum Percept Perform 1984; 10:276–291

6. Overtoom CCE, Bekker EM, Van Der Molen MW, Verbaten MN, Kooij JJS, Buitelaar JK, Kenemans JL: Methylphenidate restores link between stop-signal sensory impact and successful stopping in adults with attention-deficit/hyperactivity disorder. Biol Psychiatry 2009; 65:614–619

7. Seidman LJ, Valera EM, Makris N: Structural brain imaging of attention-deficit/hyperactivity disorder. Biol Psychiatry 2005; 57:1263–72

8. Shaw P, Lerch J, Greenstein D, Sharp W, Clasen L, Evans A, Giedd J, Castellanos FX, Rapoport J: Longitudinal mapping of cortical thickness and clinical outcome in children and adolescents with attention-deficit/hyperactivity disorder. Arch Gen Psychiatry 2006; 63:540–549

9. Nakao T, Radua J, Rubia K, Mataix-Cols D: Gray matter volume abnormalities in ADHD: voxel-based meta-analysis exploring the effects of age and stimulant medication. Am J Psychiatry 2011; 168:1154–63

10. Sowell ER, Peterson BS, Kan E, Woods RP, Yoshii J, Bansal R, Xu D, Zhu H, Thompson PM, Toga AW: Sex differences in cortical thickness mapped in 176 healthy individuals between 7 and 87 years of age. Cereb Cortex 2007; 17:1550–1560

11. Leonard CM, Towler S, Welcome S, Halderman LK, Otto R, Eckert MA, Chiarello C: Size matters: cerebral volume influences sex differences in neuroanatomy. Cereb Cortex 2008; 18:2920–2931

12. Mumford JA: A power calculation guide for fMRI studies. Soc Cogn Affect Neurosci 2012; 7:738–742

13. Durston S, Tottenham NT, Thomas KM, Davidson MC, Eigsti I-M, Yang Y, Ulug AM, Casey BJ: Differential patterns of striatal activation in young children with and without ADHD. Biol Psychiatry 2003; 53:871–878

14. Pliszka SR, Glahn DC, Semrud-Clikeman M, Franklin C, Perez R 3rd, Xiong J, Liotti M: Neuroimaging of inhibitory control areas in children with attention deficit/hyperactivity disorder who were treatment naive or in long-term treatment. Am J Psychiatry 2006; 163:1052–1060

15. Sebastian A, Gerdes B, Feige B, Klöppel S, Lange T, Philipsen A, Tebartz L, Elst V, Lieb K, Tüscher O: Neural correlates of interference inhibition , action withholding and action cancelation in adult ADHD. Psychiatry Res 2012; 202:132–141

16. Durston S, Mulder M, Casey BJ, Ziermans T, Van Engeland H: Activation in ventral prefrontal cortex is sensitive to genetic vulnerability for attention-deficit hyperactivity disorder. Biol Psychiatry 2006; 60:1062–1070

17. Button KS, Ioannidis JPA, Mokrysz C, Nosek B a, Flint J, Robinson ESJ, Munafò MR: Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 2013; 14:365–376

18. Rubia K, Cubillo A, Smith AB, Woolley J, Heyman I, Brammer MJ: Disorder-specific dysfunction in right inferior prefrontal cortex during two inhibition tasks in boys with attention-deficit hyperactivity disorder compared to boys with obsessive-compulsive disorder. Hum Brain Mapp 2010; 31:287–299

19. Lipszyc J, Schachar R: Inhibitory control and psychopathology: a meta-analysis of studies using the stop signal task. J Int Neuropsychol Soc 2010; 16:1064–1076

20. Taylor J, Allan N, Mikolajewski AJ, Hart SA: Common genetic and nonshared environmental factors contribute to the association between socioemotional dispositions and the externalizing factor in children. J Child Psychol Psychiatry 2013; 54:67–76