Appendix A. ADHD diagnostic algorithm

To determine psychiatric diagnoses, all participants (children and parents alike) were assessed with a combination of ADHD rating scales and a semi-structured diagnostic interview. In order to determine ADHD diagnoses, a diagnostic algorithm was applied based on the behavioral questionnaires (typically filled in by parents as well as a second observer) and the diagnostic interview, using DSM-IV criteria (American Psychiatric Association, 2000). Inconsistent cases were reviewed by a team of trained experts, in order to derive a consensus diagnosis.

Measures

Children were assessed with a parent rating scale (CPRS-R:L; 1998a), and either a teacher rating scale (CTRS-R:L; 1998b), applied for children < 18 years, or a self-report (CAARS-S:S; 1999), applied for children ≥ 18 years. A semi-structured diagnostic interview (KSADS-PL; Kaufman et al., 1997) was administered to both the children (if ≥ 12 years old) and their parents separately. Initially, all participants were only administered the screening interview. Participants with elevated scores on any of the screen items were administered the full ADHD section.

Parents were assessed similarly with an observer ADHD rating scale (CAARS-O:SV; 1999), typically filled in by their partner. The KSADS-PL was administered to all parents, who were, if possible, interviewed together with their partner.

Of the Conners' ADHD questionnaires the following scales were used:

- DSM Inattentive behavior
- DSM Hyperactive/Impulsive behavior
- DSM Total

For all participants using medication, ratings were done of the participant's functioning off medication.

The diagnostic algorithm

The diagnostic algorithm applied to all participants was based on a combination of symptom counts on the ADHD rating scales and the KSADS-PL, both providing operational definitions of each of the 18 behavioral symptoms of ADHD defined by the DSM-IV. Combined counts for each symptom were determined based on the KSADS-PL scores combined with scores on either the teacher rating scale (for children <18 years), the self-report (for children ≥18), or the observer rating (for parents).

Based on the algorithm, participants were given either an 'affected' (ADHD diagnosis) status or 'unaffected' status.

The following criteria were used to classify ADHD ('affected' status):

- Combined symptom count of ≥ 6 symptoms of inattentive or hyperactive/impulsive behavior
- T-score ≥ 63 on at least one of the ADHD subscales on at least one of the available Conners' ADHD rating scales
- Age of onset before 12

- Symptoms cause clinical impairment
- Symptoms are not better accounted for by another disorder

For children ≥18 years and parents, criteria were slightly adapted, such that a combined symptom count of 5 symptoms and age of onset before 15 years were sufficient for an 'affected' status.

Participants were labelled 'unaffected' if they received a T<63 on each of the scales of the Conners' rating scales, and if they had ≤ 3 symptoms (or ≤ 2 symptoms for children of ≥18 years and parents), derived from the combined symptom counts.

For analysis purposes, participants who did not meet criteria for either affected or unaffected status, were labeled 'subthreshold ADHD'.

Comorbid disorders

Participants were diagnosed with ODD if they exhibited four or more of the DSM-IV symptoms derived from the K-SADS. Likewise, conduct disorder (CD) was determined if a participant exhibited three symptoms or more DSM-IV symptoms derived from K-SADS interviews.

For internalizing disorders, we used the anxiety and depression module of the K-SADS, which was administered if the participants had elevated scores on the screening section. Diagnoses were made based on the instructions given therein, in accordance with DSM-IV-TR criteria.

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.

References

American Psychiatric Association. (2000). *Diagnostic and statistical manual of mental disorders: DSM-IV-TR*: American Psychiatric Publishing, Inc.

Conners, C. K., Erhardt, D., & Sparrow, E. P. (1999). *Conner's Adult ADHD Rating Scales: CAARS*: Multi-Health Systems, North Tonawanda, NY.

Conners, C. K., Sitarenios, G., Parker, J. D. A., & Epstein, J. N. (1998a). The revised Conners' Parent Rating Scale (CPRS-R): factor structure, reliability, and criterion validity. *Journal of abnormal child psychology, 26*(4), 257-268.

Conners, C. K., Sitarenios, G., Parker, J. D. A., & Epstein, J. N. (1998b). Revision and restandardization of the Conners Teacher Rating Scale (CTRS-R): factor structure, reliability, and criterion validity. *Journal of abnormal child psychology, 26*(4), 279-291.

Kaufman, J., Birmaher, B., Brent, D., Rao, U., Flynn, C., Moreci, P., et al. (1997). 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, 36*(7), 980-988.

Appendix B: Assessment of stress exposure

The long-term difficulties questionnaire

This questionnaire was filled in by the participants' parents with the instruction to indicate which situations are currently applicable to their child.

1. Your child has a chronic illness or handicap.
2. Someone else in the immediate family has a chronic illness or handicap.
3. Your child has a very high work pressure at school.
4. There are issues with your house (for instance too small, noisy, or busy).
5. There are issues with your neighborhood (for instance vandalism, unsafe).
6. Someone in your immediate family lost their job or became unfit for work for longer than three months.
7. Your immediate family has financial difficulties.
8. Your child has less friends than he/she would like.
9. Your child is being bullied at school or in the neighborhood.
10. Your child can't get along with someone in your immediate family.
11. Your child can't get along with someone else.
12. Other immediate family members can't get along with each other.
13. Your partner and you are separated.

The stressful live events questionnaire

This questionnaire was filled in by the participants themselves with the instructions to indicate whether they experienced any of the following events in the past five years.

1. A romantic relationship ended against your will.
2. An important friendship ended.
3. You failed on something important to you.
4. A change occurred affecting your immediate family, making it a lot less pleasant there.
5. A change occurred at school or work, making it a lot less pleasant there.
6. A group of friends which you spent a lot of time with wanted nothing to do with you anymore.
7. You left a church or religious community because you didn't feel at home there anymore.
8. A loved one died.
9. You were so seriously ill that there were concerns of permanent consequences or death.
10. You were physically abused.
11. You were raped or sexually assaulted.

Composite stress measure

If participants filled in less than half the items on both questionnaires, they were excluded from further analysis; (2.1% missing data for the SLE score; 3.3% for the LTD score). If more than half the items were filled in, missing items were imputed with 'no', i.e. we assumed the major life event had not occurred if not reported; 0.6% of the items were imputed for the SLE and 3.6% of the items for the LTD questionnaire. After this, the scores on the questionnaires were transformed to Z-values and averaged according to common practice for aggregating similar measures.

Appendix C. Sensitivity Analyses

Interaction between the gene-environment interaction and location, and ADHD diagnosis

In order to test whether the gene-environment interaction (GxE) effect on grey matter volume (GMV) and ADHD symptom count was different between testing locations (Amsterdam vs. Nijmegen), we extracted the mean GMV value of clusters of voxels with a significant correlation in the main analysis, and reran the analysis with a three-way interaction between *5-HTTLPR* genotype, stress exposure, and testing location in R. The results are shown in the top part of Table S2. As can be seen, this three-way interaction was non-significant for all clusters.

The same approach was used to test whether the effect of the GxE was different for those with and without a full ADHD diagnosis. These results are summarized in the bottom part of Table S2. As with testing location¸ there were no significant three-way interactions.

Direction of effects within-diagnostic group and within-location subsamples

As an additional sensitivity analysis, we reran the mediation analysis, with the GxE as predictor, GMV as mediator and ADHD symptom count as outcome measure, within diagnostic groups (control subjects and those with subthreshold ADHD combined, separately from those with an ADHD diagnosis), as well as within both locations (Amsterdam and Nijmegen) separately. The purpose of this analysis was not to determine significance, as the above described approach with an additional interaction terms is better suited to answer the question whether there are significant differences between groups. Rather, this analysis was to check whether the direction of effects was the same between groups. The results of these analyses are summarized in Table S3. The direction of effects is the same across subsamples, i.e. there is no reason to suspect the results are driven by one diagnostic group or scanning location.

Table S2. Analysis of a three-way interaction between the GxE and location (top part of the table) and diagnosis (bottom part) on the mean GMV of clusters found to be significant in the main analysis. The columns with statistics (B, SE, P-value) refer to the three-way interaction term.

Location (peak, other regions in cluster)	X	Y	Z	B	SE	P-value
Paracingulate Gyrus, Superior Frontal Gyrus	0	32	50	0.00025	0.009	.98
Middle Frontal Gyrus, Frontal Pole	-51	21	33	0.0074	0.011	.51
Precentral Gyrus	5	-23	54	-0.0011	0.0074	.88
Frontal Pole	-30	54	20	-0.0065	0,011	.57
Anterior Cingulate Gyrus, Paracingulate Gyrus, Superior Frontal Gyrus	0	35	44	-0.0016	0.012	.89
Location (peak, other regions in cluster)	X	Y	Z	B	SE	P-value
Paracingulate Gyrus, Superior Frontal Gyrus	0	32	59	0.0023	0.0067	.73
Middle Frontal Gyrus, Frontal Pole	-51	21	33	0.0028	0.0084	.74
Precentral Gyrus	5	-23	54	-0.0055	0.0055	.32
Frontal Pole	-30	54	20	-0.0097	0.0085	.26
Anterior Cingulate Gyrus, Paracingulate Gyrus, Superior Frontal Gyrus	0	35	44	-0.0040	0.0087	.64

Note: X, Y, Z coordinates are in MNI-space. The anatomical labels are according to the Harvard-Oxford atlas. Abbreviations: MNI=Montreal Neurological Institute; B=Regression coefficient; SE=standard error

Table S3. Direction of effects within the subsamples for the mean grey matter volume of clusters found to be significant in the main analysis. Path A represents the significant results from the regression of grey matter volume on the gene-environment interaction, and path AB represents the location of significant mediation effects. The regression coefficients refer to that of the gene-environment interaction term for each subset.

Path	*Location (peak main analysis)*	*B Full*	*B Nijmegen*	*B Amsterdam*	*B Controls*	*B ADHD*
Path A	Paracingulate Gyrus, Superior Frontal Gyrus	-0.015	-0.016	-0.015	-0.018	-0.014
	Middle Frontal Gyrus, Frontal Pole	-0.022	-0.016	-0.024	-0.028	-0.012
	Precentral Gyrus	-0.017	-0.017	-0.015	-0.013	-0.022
Path AB	Frontal Pole	-0.019	-0.28	-0.11	-0.09	-0.22
	Anterior Cingulate Gyrus, Paracingulate Gyrus, Superior Frontal Gyrus	-0.019	-0.28	-0.14	-0.10	-0.24

Note: The anatomical labels are according to the Harvard-Oxford atlas. Abbreviations: B=Regression coefficient; R= right hemisphere; L=left hemisphere.

Appendix D. Full results of the whole-brain voxel-based morphometry analyses

Table S1. Summary of the clusters where genotype, stress exposure, and the gene-environment interaction are significantly correlated with grey matter volume at p=.05, as determined by Random Field Theory.

Predictor: Stress

Path	*Direction*	*Location*	*X*	*Y*	*Z*	*Cluster size*	Z_{Max}
A	Negative	Anterior cingulate gyrus	0	10.5	40.5	6678	3.93
	Negative	Middle frontal gyrus	-30	28.5	37.5	1439	3.55
	Negative	Temporal fusiform cortex, Parahippocampal gyrus	-25.5	-42	-9	803	3.52
	Positive	Lateral occipital cortex	-39	-82.5	-3	409	3.55
AB	N/A	N/A					

Predictor: Genotype

Path	*Direction*	*Location*	*X*	*Y*	*Z*	*Cluster size*	Z_{Max}
A	L > S	Cerebellum I-IV	3	-49.5	-7.5	1909	3.46
	L > S	Lingual gyrus	-13.5	-75	-7.5	617	3.45
	S > L	Postcentral gyrus, Precentral gyrus	-55.5	-10.5	25.5	574	3.41
	S > L	Inferior frontal gyrus	60	21	15	554	3.47
	S > L	Anterior cingulate gyrus	-1.5	4.5	31.5	468	3.46
	L > S	Cerebellum VI	-16.5	-52.5	-25.5	339	3.1
AB	Negative	Inferior frontal gyrus	57	22.5	19.5	233	3.15

Predictor: GxE

Path	*Direction*	*Location*	*X*	*Y*	*Z*	*Cluster size*	Z_{Max}
A	Negative	Anterior cingulate gyrus, Paracingulate gyrus, Superior frontal gyrus	0	32	50	334	3.52
	Negative	Middle frontal gyrus, Frontal pole	-51	21	33	1232	3.57
	Negative	Precentral gyrus	5	-23	54	1097	3.45
B*	Negative	Frontal pole, Anterior cingulate gyrus, Inferior frontal gyrus, Superior frontal gyrus, Insular cortex	21	63	6	65251	3.69
	Negative	Inferior temporal gyrus, Middle temporal gyrus	49.5	-45	-6	5044	3.75
	Negative	Occipital fusiform gyrus, Lateral occipital cortex	-43.5	-76.5	-15	2068	3.61
AB	Positive	Anterior cingulate gyrus, Paracingulate gyrus, Superior frontal gyrus	0	34.5	43.5	363	3.28
	Positive	Frontal pole	-30	54	19.5	390	3.4

Note: X, Y, Z coordinates are in MNI-space and represent the peak of the cluster. The anatomical labels are according to the Harvard-Oxford atlas. MNI=Montreal Neurological Institute; Z_{max}= Z-score at the

*peak of the cluster. Path A represents the correlation between the predictors and GMV, *path B represents the correlation between GMV and ADHD symptom count (this path is the same for genotype, stress, and the GxE term, and is therefore only displayed in the GxE part of the table), and path AB represents the mediation analysis.*

Appendix E: Three-group variant of the demographics table

Table S1. Demographics table of the participants split by genotype, separately for homozygotes and heterozygotes.

Variable	*S-allele homozygotes*	*SD*	*Heterozygotes*	*SD*	*L-allele homozygotes*	*SD*	*Test-statistic*	*DF*	*P-value*
Participants	100		356		245				
Covariates									
Amsterdam location	46.0%		53.1%		51.4%		X=1.57	2	.46
Male gender	54.0%		53.9%		59.2%		X=1.77	2	.41
Age in years	17.69	3.48	16.71	3.55	17.14	3.63	F=3.20	699	.04
Parents' years of education	12.29	2.53	11.93	2.50	12.18	2.46	F=1.20	699	.30
Stress Z-score	0.03	1.12	-0.06	0.95	0.07	1.01	F=1.36	699	.26
Number of stressful live events	2.12	1.71	1.98	1.46	2.19	1.54	F=1.36	699	.26
Number of long-term difficulties	0.91	1.42	1.14	1.06	1.27	1.50	F=2.83	687	.06

Note: Differences between genotypes in the categorical variables 'location' and 'gender' were analyzed with a Chi-square test; for the other, continuous variables we performed an analysis of variance.SD= standard deviation. DF= degrees of freedom.