INTRODUCTION

Hoarding Disorder (HD) is one of the most recent mental disorders to be included in the DSM-5 [1] and ICD-11 [2]. Individuals with HD experience persistent difficulties parting with possessions, regardless of their value, due to a perceived need to save the items and distress associated with discarding them. This results in the accumulation of possessions that clutter active living areas and substantially compromise their intended use, causing clinically significant distress or impairment. Most people with HD also excessively acquire items that they do not need and experience distress if they are unable or are prevented from acquiring items (excessive acquisition specifier). Critically, these symptoms are not attributable to another medical or mental disorder, such as obsessive-compulsive disorder (OCD), psychosis or dementia [1, 2].

The prevalence of HD in the population is estimated to be approximately 1% to 2.5% for both men and women [3–6], but a much larger proportion of the population experience symptoms at various levels of severity, with estimates up to 6.7% [7] and 9% [8] in some studies. It is widely believed that the liability to hoarding
symptoms (HS) is continuous in the general population, with clinically relevant HD at the extreme end of the spectrum [9].

HS typically appear in early-to-middle-adolescence and, in contrast to many other psychiatric disorders, symptom severity increases with age [10–13]. Psychiatric comorbidity is common in HD, with up to 70% of individuals having at least one additional disorder, most commonly anxiety and/or depression (DEP [14, 15]). Attentional problems are also common in individuals with HD [14, 16–18].

The etiology of HD is largely unknown, though likely to be multifactorial in nature and related to a complex interplay of genetic, neurobiological, and psychosocial factors. Family studies have consistently shown that HD run in families [15, 19–22]. Population-based twin studies have estimated the heritability of hoarding symptoms based on self-report questionnaires [3, 4, 7, 23–26]. In adults, heritability estimates range from 26 to 49%; the remaining variance was due to unique environmental factors and measurement error, whilst shared environmental factors appear to play a negligible role. In young people, large population-based samples of twins (N = 3974 twins, [3]; 

\( N = 25,523, [4] \)) reported higher heritability of hoarding symptoms amongst 15-year-old boys than in girls (33% vs 17%) and significant shared environment influences (22%) among female twins only, while Burton et al. [8] reported a heritability of 61% with no shared environment effect (221 twin pairs). Thus, it is possible that genetic and environmental influences on hoarding symptoms change across development, with shared environmental factors being more important in young people (particularly in girls).

Linkage and GWAS studies of HS have been rare thus far and have largely been conducted in small samples of OCD or Tourette syndrome patients. Candidate gene studies in individuals with OCD have suggested (largely non-replicated) associations between HS and a number of candidate variants [27–31]. Three previous modestly sized genome-wide linkage studies of HS in OCD or Tourette syndrome samples resulted in either no significant or conflicting results [21, 32, 33]. One study in OCD patients found linkage between HS and a region on chromosome 14q23–32 [21], and another linkage study in OCD patients found evidence for interaction with a region on chromosome 9q that houses SLC1A1, a glutamate transporter gene [32]. One GWAS of OCD symptom dimensions reported SETD3, a gene highly expressed in the brain and involved in apoptotic processes and transcriptional changes, to be associated with HS [34]. Another GWAS focused on HS in a British twin cohort [35]. The sample included 3304 twins from the TwinsUK cohort, predominantly female (91.8%), with a mean age of 56.8 years. All participants completed the Hoarding Rating Scale Self-Report (HRS-SR; [36]), a brief self-administered instrument consisting of five items (clutter, difficulty discarding, excessive acquisition, distress, and impairment). While no genome-wide significant loci were identified, two genomic loci on chromosomes 5 and 6 showed suggestive evidence for association with HS.

Larger samples are needed to increase power to detect significant genetic effects. Therefore, the current study aimed to conduct a GWAS meta-analysis of several large international cohorts from Sweden, the Netherlands, England, and Canada that included parent- or self-report hoarding scale data. We pooled data from seven population-based cohorts that together include 27,651 individuals (including 7012 twin-pairs), representing a more than eightfold increase in sample size compared to the previous study by Perraud et al. [35]. We followed up the results with gene-based and gene-set analyses, as well as leave-one-out hoarding symptom polygenic risk score (PRS) analyses and cross-trait PRS analyses to examine a possible genetic association between other phenotypes and HS.

**MATERIALS AND METHODS**

**Cohorts and phenotype assessment**

The Hoarding Symptom (HS) GWAS meta-analysis included individuals from seven different European-ancestry cohorts. Four cohorts are part of the Swedish Twin Registry (STR, [37]), namely different age groups of the Child and Adolescent Twin Study in Sweden (CATSS15, CATSS18, and CATSS24 [38]), and the Young Adult Twins in Sweden Study (YATSS [39]). CATSS is a prospective, longitudinal study of all twins born in Sweden since 1992. For CATSS, one measurement time point per individual was selected, preferring the measurement at age 24 over age 18 over age 15 if more than one measurement was completed. The other cohorts are from the Netherlands Twin Register (NTR, [40]); Spit for Science (SfS, [41, 42]), and TwinsUK (see Supplementary Material for detailed descriptions). Data from the TwinsUK cohort were included in a previous GWAS [35]. The cohorts are all population-based, predominantly including twins (except SfS), with a mean age-range between 11 and 57 (Table 1). Participants, or their parents, answered one of two questionnaires assessing HS.

In STR, NTR, and TwinsUK, HS were assessed using four (STR-YATSS, STR-CATSS24) to five items (STR-CATSS15, STR-CATSS18, NTR, TwinsUK) of the Hoarding Rating Scale Self-Report (HRS-SR; [43]), while in SfS parent- or self-reported hoarding traits were assessed using two items from the Toronto Obsessive Compulsive Scale (TOCS), a 21-item questionnaire described elsewhere [44, 45] (see Supplementary Table S1 for questionnaire details). Though the TOCS was originally designed to measure obsessive-compulsive symptoms, the two questions used here effectively reflect two core components of hoarding, namely acquisition of objects and difficulty discarding. To summarize across HRS-SR items, four items of the HRS-SR were used to calculate a one-factor model using a latent variable analysis with the R package lavaan [46] confirmatory-factor-analysis function. For STR item 4 of the HRS-SR was not assessed (To what

### Table 1. Overview of cohorts included in the GWAS meta-analysis of HS.

|                  | STR-CATSS15 | STR-CATSS18 | STR-CATSS24 | STR YATSS | NTR | SfS | TwinsUK |
|------------------|-------------|-------------|-------------|-----------|-----|-----|---------|
| N                | 3605        | 3286        | 2313        | 2947      | 6839| 5218| 3443    |
| N MZ twin pairs  | 241         | 256         | 191         | 552       | 866 | –   | 317     |
| N DZ twin pairs  | 1219        | 1137        | 685         | 348       | 425 | –   | 775     |
| N siblings       | –           | –           | –           | –         | 438 | –   | –       |
| N parents        | –           | –           | –           | –         | 1361| –   | –       |
| % female         | 50%         | 56%         | 57%         | 63%       | 66% | 48% | 92%     |
| Mean age ± SD    | 15.47 ± 0.36| 18.56 ± 0.33| 23.84 ± 0.32| 23.93 ± 1.78| 41.49 ± 15.19| 10.92 ± 2.79| 56.7 ± 12.6|

For each individual cohort included in the HS meta-analysis (STR-CATSS15, STR-CATSS18, STR-CATSS24, STR-YATSS, NTR, SfS, TwinsUK), the table lists the total sample size included (N), the number of monozygotic twin pairs (N MZ twin pairs), the number of dizygotic twin pairs (N DZ twin pairs), the number of siblings (N siblings), the number of parents (N parents), the percentage of females and males in the total N (% females [males]), and the mean age with standard deviations (SD). Twins where only one twin participated were not counted as twins. NTR twin pairs include 10 multiplets, (31 individuals), 185 spouses of twin probands and 1153 twins without any other family member participating. Note that CATSS samples were later pooled across the three cohorts (CATSS15, CATSS18, CATSS24) for GWAS analysis, depending on the platform they were genotyped on (GSA, PsychChip).
extent do you experience impairment in your life (daily routine, job/school, social activities, family activities, financial difficulties) because of clutter, difficulty discarding, or problems with buying or acquiring things?), while for NTR and TwinsUK item 5 was not included (“To what extent do you experience emotional distress because of clutter, difficulty discarding or problems with buying or acquiring things?”). Fit indices of the one-factor model were compared to their corresponding gene(s) based on their position in the same construct for all cohorts. Standardized individual factor scores were calculated for the common factor model using the lavPredict function. Standardized individual factor scores were calculated for the common factor model using the lavPredict function. In case an individual was missing one item, the mean of the remaining three items was used to impute the missing value. If more than one item was missing, the individual was removed from the analysis. Two SFS Hoarding items of the TOCS were summed and standardized into a Z-score. To ensure reliable and valid symptom reporting, SFS participants <12 years of age with self-reported HS were excluded.

**Genome-wide association analysis**

All participants were genotyped on SNP-arrays based on DNA from saliva or blood. One part of the STR-CATSS samples was genotyped on the PsychChip array (N = 8539), another part was genotyped on the GSA genotyping platform (GSA, PsychChip), forming two separate CATSS datasets (STR-CATSS-GSA and STR-CTASS-PC). Each of the six datasets (STR-CATSS-GSA, STR-CATSS-PC, STR-YATSS, NTR, SFS, and TwinsUK) underwent stringent quality-control (QC), including the removal of non-European ancestry outliers based on PCA and imputation using the HRC [47] (STR, NTR) or the 1000 G [48] (SFS, TwinsUK) reference sets (see Supplementary Material for more details). After genotyping, quality control, and imputation of each cohort, STR included 12,151, NTR 6839, SFS 5218, and TwinsUK 3443 (total N = 26,728) individuals with complete genotypic and phenotypic information.

A linear mixed modeling GWAS was conducted within each cohort using GCTA-fastGWA [49, 50]. For STR, NTR, and TwinsUK a sparse Genetic Relatedness Matrix (GRM) was calculated and the first 10 principal components, sex, age, and genotyping batches were used as covariates. In a sparse GRM all off-diagonal values below 0.05 are set to 0, thereby accounting for the close relatedness of individuals in the data. For SFS, analyses were performed on unrelated individuals; the first enrolled sibling from each family was selected for further analysis. The analysis was performed with a full GRM and sex, age, respondent (parent vs. child reporting), genotyping array, principal components 1–3 and projected principal components 1–3 (see Supplement for details) as covariates.

Next, the resulting GWAS summary statistics were cleaned and harmonized. All variants were filtered on minor allele frequency (MAF) > 1%, and imputation-quality score > 0.8. All datasets were aligned to the HRC-reference. In case alleles were reported on different strands, they were flipped to the orientation in the HRC reference. Strand ambiguous A/T and C/G SNPs were removed if their MAF was ≥0.4. Remaining ambiguous SNPs were strand aligned by comparing MAF to the HRC reference [47]. We then used METAL [51] within the Rapid Imputation for CONsortial Pipeline (Ricopili) [52] to conduct an inverse variance weighted meta-analysis. The genomic control factor (Lambda and Lambda1000) was inspected for each individual cohort to detect any residual population stratification or systematic technical artifacts. Also, the linkage disequilibrium (LD) score regression (LDSC) [53] intercept was inspected as an alternative measure of test statistic inflation. The genome-wide significance threshold was set at $5 \times 10^{-8}$.

**Heritability**

Heritability estimates of each individual cohort were extracted from the GCTA association output. GCTA uses a restricted maximum likelihood (REML) approach [54] to estimate heritability in the GRM that is supplied to correct for relatedness in the linear association test. This means that for the twin cohorts (STR, NTR, and TwinsUK), the heritability was based on the sparse GRM, while for the unrelated SFS cohort, heritability was based on the full GRM. For all heritability estimates, the same covariates as in the GWAS analyses were used. We further used LDSC [53] to calculate the SNP heritability of the HS GWAS meta-analysis. The SNP heritability in LDSC is based on the estimated slope from the regression of the SNP effect sizes from the GWAS on the LD score.

**Gene-based and gene-set analyses**

We carried out a Multi-marker Analysis of Genomic Annotation (MAGMA) [55] v1.08 as implemented in the web-based tool Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) [56] v.3.7 to test genetic associations at the gene level for the combined effect of SNPs in or near protein coding genes. Gene-based $p$-values were computed by mapping the corresponding gene(s) based on their position in the genome. Positional mapping was based on ANNOVAR annotations and the maximum distance between SNPs and genes was set to the default setting of 10 kb. Based on the results of gene analysis, competitive gene-set analysis was performed with default parameters. The 15,496 gene-sets were obtained from MsigDB v7.0, including “Curated gene sets” consisting of nine data resources including KEGG, Reactome, and BioCarta, and GO terms consisting of three categories (biological processes, cellular components, and molecular functions).

**Cross-trait polygenic risk score (PRS) analyses**

To explore the genetic relationship between HS and other phenotypes, we calculated a range of PRSs based on large-scale GWAS summary statistics. We selected mainly studies of psychiatric disorders, i.e., OCD [57], DEP [58], schizophrenia (SCZ) [59], autism spectrum disorder (ASD) [60], attention-deficit hyperactivity disorder (ADHD) [61], and educational attainment (EA) [62]. PRS were computed in PRSice2 [63] for each cohort. The PRS scores were calculated as the weighted sum of the risk allele dosages at preselected $p$-value thresholds based on the reported thresholds in the primary publications (EA: $P = 1$; ADHD, ASD, OCD, SCZ: $P = 0.01$; DEP: $P = 0.5$). For STR, PRS analyses were conducted separately for the three datasets (STR-CATSS-GSA, STR-CATSS-PC, STR-YATSS) and were subsequently merged using the replication module of the Ricopili pipeline, sign tests on the top SNPs (inclusion threshold of $P < 0.05$) and $R^2 > 0.5$ were performed between each pair-wise combination of cohorts as well as between LOO meta-analyses and carried out a set of sensitivity analyses. With the replication module of the Ricopili pipeline, sign tests on the top SNPs (inclusion threshold of $P < 0.05$) and $R^2 > 0.5$ were performed between each pair-wise combination of cohorts as well as between LOO meta-analyses and the left-out cohort, we calculated LOO PRS with HS. If there was considerable observed heterogeneity across study sites ($I^2 > 50\%$ and/or $P_{\text{Het}} < 0.05$), we further calculated a random effects model.

**Compatibility of cohorts**

To identify if the summary statistics from any of the included cohorts substantially deviated from the others, we performed leave-one-out (LOO) GWAS meta-analyses and carried out a set of sensitivity analyses. With the replication module of the Ricopili pipeline, sign tests on the top SNPs (inclusion threshold of $P < 0.05$) and $R^2 > 0.5$ were performed between each pair-wise combination of cohorts as well as between LOO meta-analyses and the left-out cohort, we conducted LOO PRS analyses, following the same procedure as for the cross-trait PRS analyses described above (see previous method-section on cross-trait PRS analyses for details).
RESULTS
Phenotype normalization
The distributions of each cohorts’ sum-scores have been maximally normalized (distribution of the item- and total raw hoarding scores are shown in Supplementary Figs. S1–S6, distribution of the one-factor model scores (STR, NTR, and TwinsUK) and standardized scores (SfS) are shown in Supplementary Figs. S7–S10). An over-representation of zero sum-scores indicates censoring. The model-fit of the one-factor models in the latent variable analysis of the HRS-SR for NTR (CFI = 0.999, SRMR = 0.014) and STR (STR-GSA: CFI = 0.995, SRMR = 0.032; STR-PC: CFI = 0.997, SRMR = 0.020; STR-YATSS: CFI = 0.997, SRMR = 0.020) were excellent, thereby demonstrating that all items of the HRS-SR highly load onto one common factor and psychometrically measure the same construct, allowing us to drop one item of the HRS-SR per analysis. The two TOCS hoarding items in the SfS data could not be assessed for model fit, however, the two items were highly comparable in wording to items 2 and 3 of the HRS.

Genome-wide association results
The final dataset included 27,651 individuals with complete phenotypic and genotypic data and 6,541,342 autosomal SNPs. No significant inflation was observed (λ = 1.024, λ1000 = 1.001, LDSC intercept = 1.0173, see Supplementary Fig. S11 for QQ plot). No SNP exceeded the genome-wide significance threshold (see Fig. 1 for a Manhattan-plot including the Manhattan-plot of the GWAS in the upper panel). The SNPs with the lowest $P$-values ($<1 \times 10^{-8}$) were rs117321479 ($P = 3.12 \times 10^{-7}$) and rs78426839 ($P = 7.70 \times 10^{-7}$) on chromosome 2, and rs79297972 ($P = 8.90 \times 10^{-7}$) on chromosome 18 (see Supplementary Figs. S12–S15 for regional association plots and forest plots). The region tagged by rs117321479 spans 57.6 kb (LD $r^2 > 0.6$) and entails the gene SOX5. The region tagged by rs78426839 spans 22.1 kb (LD $r^2 > 0.6$) and entails the genes TUBA4B, DNAJB2, PTPRN, MIR153-1, RESP18, and DNPEP. The region tagged by rs7567224 spans 24.40 kb (LD $r^2 > 0.6$) and entails the gene CNTNAP5, while the region tagged by rs78426839 spans 94.3 kb (LD $r^2 > 0.6$). In addition, 19 independent SNPs with $P < 1 \times 10^{-8}$ were identified (see Supplementary Table S2 for a list of corresponding association results).

Heritability
For the twin cohorts, the additive genetic variance of HS, estimated based on the sparse genetic relatedness matrices, ranged between 0.26 (NTR) and 0.48 (TwinsUK), with estimates for the STR cohorts in between (STR-CATSS15: 0.47, STR-CATSS18: 0.29, STR-CATSS24: 0.35, STR-YATSS: 0.28). Note that the SNP-heritability estimates based on the twin cohorts are largely driven by the twin resemblance (~0.5 between DZ twins and siblings, 1.0 for MZ twins, and 0 between unrelated individuals). The heritability for SfS, only including unrelated individuals, was 0.11 (SE = 0.057, $P = 0.0303$). The SNP-based heritability estimate of the GWAS meta-analysis using LDSC resulted in a total observed-scale heritability of 0.019 (SE = 0.016, $Z = 1.18$, $P = 0.235$).

Gene-based analyses
We conducted gene-based tests to determine whether any protein-coding gene carries a load of common variation associated with HS. SNPs were mapped to 18,646 protein coding genes obtained from Ensembl build 85. No gene reached the Bonferroni-corrected significance threshold of $p = 0.05/18,646 = 2.682 \times 10^{-6}$ (see Fig. 1 for a Manhattan-plot including the Manhattan-plot of the gene-based test in the lower panel, and
We conducted sign-test analyses between LOO GWAS analyses of age-separated STR subcohorts (Supplementary Table S5), age-separated NTR subcohorts (Supplementary Table S6) and the respective left-out cohort. The STR cohorts were divided into four age groups, pertaining to their division into separate phenotyping rounds (CATSS15 mean age = 15.46; CATSS18 mean age = 18.56; CATSS24 mean age = 23.84; YATSS mean age = 23.93), while the NTR data was separated into three age groups (group1 < 30 years; group2 30–45 years; group3 > 45 years). The sign test results at the p-value threshold of $p = 1 \times 10^{-6}$ identified very few independent genomic regions (0–6) and are therefore difficult to interpret. While the ratios of the sign tests for the other two p-value thresholds ($p = 1 \times 10^{-5}$ and $p = 1 \times 10^{-4}$) varied between 0.2 and 0.8, there is no apparent pattern indicating a systematic deviation of one cohort from the rest.

The LOO PRS analyses did not show a significant association between any of the PRSs and the HS score of the left-out cohort. While this could suggest heterogeneity across the cohorts, it is likely indicating a lower power for this analysis considering that each discovery cohort had a rather low sample size and we only saw significant associations with PRS based on GWASs with considerably higher sample sizes (SCZ, DEP, EA). See Supplementary Tables S3 and 2 for results.

**DISCUSSION**

With 27,651 included individuals we conducted the largest GWAS study of HS in the population to date. Although we could not report any genome-wide significant SNPs, we found a significant contribution of common genetic factors to HS as indicated by a substantial genetic SNP heritability of 11% ($P = 0.0303$) in one of our cohorts (SFS) with unrelated samples. It suggests that, with sufficient power, specific genetic variants that are associated with HS will be eventually identified. SNP-based heritability of the meta-analysis as calculated with LDSC was low ($h^2 = 0.019$, SE = 0.016, $Z = 1.18$) and non-significant. We therefore did not conduct genetic correlation analyses as it is suggested to have a heritability Z-score of above 1.5 (optimal > 4) in order for the analysis to be meaningful [53, 67]. We also found a significant genetic variance component in the twin family cohorts, ranging from 28 to 48%. These heritability estimates may be largely driven by the relatedness between the samples and are indeed more
power to discover associations [74, 75]. Nevertheless, the quantitative ADHD symptom GWAS by Middeldorp et al. [75] showed a high genetic correlation and strong concordance at individual loci with clinical ADHD [61], thereby supporting the hypothesis that clinically diagnosed cases are the extreme of a quantitative symptom trait and further demonstrating the usefulness and importance of quantitative assessment of symptomatics in the population.

A further limitation is that hoarding symptom scales may reflect heterogeneous disorders. It is known that many different psychiatric disorders can cause hoarding-like symptoms, such as schizophrenia, OCD, or severe depression. HD is essentially a diagnosis of exclusion (DSM-5, ICD-11). As in any study based on self-administered instruments, it was not possible to rule out these other causes of HS. It will further be of interest to determine the risk of which to HS in the population and clinical HD share the same genetic susceptibility. However, no case-control GWAS of HD exists, nor is HD assessed in large datasets like UK BioBank [76, 77], which is unfortunate given the relatively high prevalence of around 2.5% for HD [6] and high individual and societal cost [78, 79]. Thus, for the time being, the study of HS in the population may be the only feasible approach to understanding the genetics of HD.

To conclude, we found that HS are heritable, confirming and extending previous twin studies. Nevertheless, we had limited power to detect any genome-wide significant loci. Much larger samples will be needed to further extend our findings and reach a “gene discovery zone”. Further, additional samples should be more ethically diverse to ensure that results are relevant to individuals of non-European ancestry [80]. Future research should include the collection of DNA samples from individuals with HS, as well as samples from strictly diagnosed HD patients.

DATA AVAILABILITY
The meta-analyzed summary statistics will be made available via the Psychiatric Genomics Consortium Download page (https://www.med.unc.edu/pgc/download-results/).

REFERENCES
1. DSM-5. Diagnostic and statistical manual of mental disorders: DSM-5. 2013:237–42.
2. World Health Organization. ICD-11: International classification of diseases (11th revision). Retrieved from https://www.who.int/ 2019.
3. Ivanov VZ, Mataix-Cols D, Serlachius E, Lichtenstein P, Anckarsäter H, Chang Z, et al. Prevalence, comorbidity and heritability of hoarding symptoms in adolescence: a population based twin study in 15-year olds. PLoS ONE. 2013;8:1–7.
4. Ivanov VZ, Nordsletten AE, Mataix-Cols D, Serlachius E, Lichtenstein P, Lundström S, et al. Heritability of hoarding symptoms across adolescence and young adulthood: A longitudinal twin study. PLoS ONE. 2017;12.https://doi.org/10.1371/JOURNAL.PONE.0179541.
5. Nordsletten AE, Reichenberg A, Hatch SL, Fernández De La Cruz L, Pertusa A, Hotopf M, et al. Epidemiology of hoarding disorder. Br J Psychiatry. 2013;203:445–52.
6. Postlethwaite A, Kellett S, Mataix-Cols D. Prevalence of Hoarding Disorder: a systematic review and meta-analysis. J Affect Disord. 2019;256:309–16.
7. Mathews CA, Delucchi K, Cath DC, Willemsen G, Boomsma DI. Partitioning the etiology of hoarding and obsessive-compulsive symptoms. Psychol Med. 2014;44:2867–76.
8. Burton CL, Park LS, Corfield EC, Forget-Dubois N, Dupuis A, Sinopoli VM, et al. Heritability of obsessive–compulsive trait dimensions in youth from the general population. Transl Psychiatry. 2018;8:191.
9. Timpano KR, Bromann-Fulks JJ, Gaesmer H, Enner C, Rief W, Olatunji BO, et al. A taxometric exploration of the latent structure of hoarding. Psychol Assess. 2013;25:194–203.
10. Ayers CR, Saxena S, Golshan S, Wetherell JL. Age at onset and clinical features of late life compulsive hoarding. Int J Geriatr Psychiatry. 2010;25:142–9.
11. Cath DC, Nizam K, Boomsma D, Mathews CA. Age-specific prevalence of hoarding and obsessive compulsive disorder: a population-based study. Am J Geriatr Psychiatry. 2017;25:245–55.
33. Zhang H, Leckman JF, Pauls DL, Tsai CP, Kidd KK, Rosario-Campos M. Genome-wide association study of pediatric obsessive-compulsive traits: shared genetic risk between traits and disorder. Transl Psychiatry. 2021;11:11–10.

37. Zagai U, Lichtenstein P, Pedersen NL, Magnusson PKE. The Swedish Twin Registry: a database of Swedish twins. J Psychopathol Behav Assess. 2005;27:29–38.

52. Lam M, Awasthi S, Watson HJ, Goldstein J, Panagiotaropoulou G, Trubetskoy V, et al. ROCIPiL: Rapid Imputation for CNOrtisas PipeLine. Bioinformatics. 2020;36:930–933.

59. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and Quantitative Trait Locus analysis of the BDNF gene in relation to two phenotypes: hoarding and obesity. J Abnorm Psychol. 2009;118:293–301.

76. Magnusson PKE, Almqvist C, Rahman I, Ganna A, Viktorin A, Walum H, et al. The Swedish twin registry: establishment of a biobank and other recent developments. Twin Res Hum Genet. 2013;16:317–32.

84. Anckarsäter H, Lundström S, Kollberg L, Kerekes N, Palm C, Carlström E, et al. The Child and Adolescent Twin Study in Sweden (CATSS). Twin Res Hum Genet. 2011;14:495–508.

98. Rosen S, Lavaan: an R package for structural equation modeling. J Stat Softw. 2016;70:1–36.

12. Grisham JR, Frost RO, Steketee G, Kim HJ, Hood S. Age of onset of compulsive hoarding. J Anxiety Disorder. 2006;20:675–86.

18. Tolin DF, Frost RO, Steketee G, Tolin DF. Comorbidity in hoarding disorder. Depress Anxiety. 2011;28:876–84.

29. Sinopoli VM, Erdman L, Burton CL, Park LS, Dupuis A, Shan J, et al. Serotonin system genes and hoarding with and without other obsessive-compulsive disorder symptoms: a survey of the Australian Twin Registry. Am J Med Genet B Neuropsychiatr Genet. 2011;156:240–51.

34. Alemany-Navarro M, Cruz R, Real E, Segalàs C, Bertolín S, Rabionet R, et al. Looking into the genetic bases of OCD dimensions: a pilot genome-wide association study of pediatric obsessive-compulsive traits: shared genetic risk between traits and disorder. Transl Psychiatry. 2021;11:11–10.

36. Auyeung B, Cromwell J, Liddle PF, Dazzan P, Garety PA, Jones PB, et al. ADHD in children associated with lifetime hoarding symptoms? An epidemiological study. Depress Anxiety. 2013;30:74–81.

42. Torres JL, Martin J, Khan N, Soderberg G, Rosenberg A, Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale applications. Int J Epidemiol. 2020;49:1279–83.

47. Magistad KD, Oktay A, Cleveland MJ, Peacock J, Turecki G, et al. Annotation of 150 genome-wide significant loci identifies putative causal variants and highlights the importance of the prefrontal brain regions. Nature. 2019;567:343–52.

64. Cropley M, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.

65. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agero E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet. 2019;51:63–75.

70. Lee SH, Wray NR, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet. 2011;88:76–82.

75. Yildiz N, Strom et al. A resource panel of 64,976 haplotypes for genotype imputation. Nat Genet. 2016;48:1279–83.

79. Magnusson PKE, Almqvist C, Rahman I, Ganna A, Viktorin A, Walum H, et al. The Swedish twin registry: establishment of a biobank and other recent developments. Twin Res Hum Genet. 2013;16:317–32.

83. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.
8

ACKNOWLEDGEMENTS

We acknowledge The Swedish Twin Registry (STR) for data access. The STR is managed by Karolinska Institutet and receives funding through the Swedish National Infrastructure for Computing (SNIC) at Uppmax partially funded by the Swedish Research Council through grant agreement no. 2018-05973. The Netherlands Twin Register (NTR) warmly thanks all twin families for their participation. NTR is supported by multiple grants from the Netherlands Organisation for Scientific Research (NWO) and Medical Research (ZonMW): Netherlands Twin Registry Repository (NWO 480-15-001/674); the Biobank-based integrative omics study (BIOS) funded by BMBF-NL (NWO projects 184.021.007 and 184.033.111); the European Science Council (ERC) Genetics of Mental Illness (ERC Advanced, 230374, PI Boomsma); the Royal Netherlands Academy of Science Professor Award (PAR/6635) to DIB; Rutgers University Cell and DNA Repository (NIMH U24 MH08457-06); the Avera Institute, Sioux Falls, South Dakota (USA) and the National Institutes of Health (NIH R01 HD042157-01A1). Part of the genotyping was funded by the Genetic Association Information Network (GAIN) of the Foundation for the National Institutes of Health and Grand Opportunity grants (1RC2 MH089551). TwinsUK is funded by the Wellcome Trust, Medical Research Council, Versus Arthritis, European Union Horizon 2020, Chronic Disease Research Foundation (CDRF), Zoe Ltd, the National Institute for Health and Care Research (NIHR) Clinical Research Network (CRN) and Biomedical Research Centre based at Guy’s and St Thomas’ NHS Foundation Trust in partnership with King’s College London. Spit for Science was supported by the Canadian Institutes of Health Research (RGS, MOP-93696 and PDA, MOP-106573). Dr. Arnold is supported by the Alberta Innovates Translational Health Chair in Child and Youth Mental Health. We acknowledge support by the Open Access Publication Fund of Humboldt-Universität zu Berlin.

AUTHOR CONTRIBUTIONS

NIS, DJAS, DC, DMC, DIB, and MM contributed to the conception of the over-all study design: CLJ, RP, ML, JC, JIH, VZI, HL, PL, PM, CR, RS, HMW, JC, PDA, DIB, SMM, DMC, and DC contributed to the data collection of the individual datasets and/or provided code; NIS, DJAS, TS, and CI conducted all primary data analyses; NIS, DS, and DMC drafted the manuscript; all authors provided critical edits and discussions and approved the submitted version of the manuscript.

FUNDING

Open Access funding enabled and organized by Projekt DEAL.

COMPETING INTERESTS

David Mataix-Cols receives royalties for contributing articles to UpToDate, Wolters Kluwer Health, outside of the submitted work. Henrik Larsson reports receiving grants from Shire Pharmaceuticals; personal fees from and serving as a speaker for Medice, Shire/Takeda Pharmaceuticals and Evolan Pharma AB; and sponsorship for a conference on attention-deficit/hyperactivity disorder from Shire/Takeda Pharmaceuticals and Evolan Pharma AB, all outside the submitted work. Henrik Larsson is editor-in-chief of JCPP Advances. Russell J. Schachar has consulted to E. Lilly, Highland Therapeutics and eHave. All other authors report no potential conflict of interest.

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41398-022-02248-7.

Correspondence and requests for materials should be addressed to Nora I. Strom.

Reprints and permission information is available at http://www.nature.com/reprints

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