Title

From rare Copy Number Variations to biological processes in ADHD

Authors

Benjamin Harich[1#], Monique van der Voet[1#], Marieke Klein[1], Pavel Čížek[2,‡], Michaela Fenckova[1], Annette Schenck[1,*], Barbara Franke[1,3,*]

[1]Department of Human Genetics, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, The Netherlands

[2]Centre for Molecular and Biomolecular Informatics, Radboud Institute for Molecular Life Sciences, Radboud university medical center, Nijmegen, the Netherlands

[3]Department of Psychiatry, Donders Institute for Brain, Cognition and Behaviour, Radboud university medical center, Nijmegen, The Netherlands

[‡]Former address

[*] Co-corresponding authors

[#]These authors contributed equally to this work

Abstract

Aim: Attention-deficit/hyperactivity disorder (ADHD) is a highly heritable psychiatric disorder. The objective of this study was to define ADHD-associated candidate genes, and their associated molecular modules and biological themes, based on the analysis of rare genetic variants.

Methods: We combined data from 11 published copy number variation (CNV) studies in 6176 individuals with ADHD and 25026 controls and prioritized genes by applying an integrative strategy based on criteria including recurrence in ADHD individuals, absence in controls, complete coverage in copy number gains, and presence in the minimal region common to overlapping CNVs, as well as on protein-protein interactions and information from cross-species genotype-phenotype annotation.

Results: We localized 2241 eligible genes in the 1532 reported CNVs, of which we classified 432 as high-priority ADHD candidate genes. The high-priority ADHD candidate genes were significantly co-expressed in the brain. A network of 66 genes was supported by ADHD-relevant phenotypes in the cross-species database. In addition, four significantly interconnected protein modules were found among the high-priority ADHD genes. A total of 26 genes were observed across all applied bioinformatic methods. Look-up in the latest genome-wide association study for ADHD showed that among those 26, *POLR3C* and *RBFOX1* were also supported by common genetic variants.

Conclusions: Integration of a stringent filtering procedure in CNV studies with suitable bioinformatics approaches can identify ADHD candidate genes at increased levels of credibility. Our pipeline provides additional insight in the molecular mechanisms underlying ADHD and allows prioritization of genes for functional validation in validated model organisms.

41

Introduction

Attention-deficit/hyperactivity disorder (ADHD) is one of the most common neuropsychiatric disorders, with a prevalence of 5–6% in children.(1) The disorder persists into adulthood in a significant proportion of affected individuals, resulting in a prevalence of 2.5–4.9% in adults.(2) The clinical symptoms of ADHD include age-inappropriate inattention, hyperactivity, and impulsivity.(3) Twin and adoption studies estimated a high heritability of 76% for ADHD.(3)

Identification of the genes implicated in ADHD and their molecular functions offers opportunities to understand the neurobiological mechanisms leading to ADHD and facilitates the development of diagnostic tools and new treatments. However, despite the high heritability, identification of ADHD risk genes has been difficult, mainly due to ADHD's complex genetic architecture.(2,4,5) To date, mainly genetic variants that frequently occur in the population have been investigated for their role in ADHD, either through studies of candidate genes or hypothesis-free genome-wide association studies (GWAS).(6,7) A recent GWAS meta-analysis identified the first 12 loci harbouring ADHD risk variants.(8) Another type of GWAS has focused on the association of rare copy number variants (CNVs) with ADHD. Such CNV GWASs have largely concentrated on rare events (primarily) observed in individuals diagnosed with ADHD. We analysed the 11 studies published to date that have detected rare CNVs in ADHD cases.(9,10,19,11–18) Those CNV GWASs implicated more than 2200 candidate genes in ADHD, though most have investigated rather limited sample sizes, and most of the CNVs were detected only in single patients. Based on the average mutation rate in-between individuals in the general population, single-patient rare findings have a high chance to be false positive; it is thus important to concentrate on repeatedly occurring copy number events.(20)

Data integration from various sources is an important strategy to move from genes to biologically meaningful modules. Examples for this come from publications on ADHD and related disorders.(21),(22) A publication on autism spectrum disorders showed how data integration enabled the identification of highly conserved gene clusters that improve our understanding of neuropsychiatric disorders.(21) Similarly, a recent study found a significant overlap of ADHD case CNVs with targets of the Fragile-X mental retardation protein (FMRP), a gene cluster involved in neurodevelopmental disorder risk.(22) In many cases, data integration currently takes place only across data modalities derived from studies in humans. This neglects the wealth of phenotypic information that can be derived from model organisms such as monkey, rat, mouse, zebrafish, and fruit fly.(7)

In this study, we surveyed and integrated data on CNVs associated with ADHD from existing publications, aiming to define robustly ADHD-associated genes, molecular modules, and biological themes underlying this disorder. We combined data from the 11 published ADHD CNV studies and applied an integrative strategy using redundancy criteria, data on protein-protein interactions, and – most innovative – employing information from cross-species genotype-phenotype annotation to prioritize candidate genes.(23,24) We classified 432 high-priority ADHD candidate genes, supported by co-expression, cross-species phenotype, and protein interaction information, with 26 genes highlighted across all approaches. Integration with data on common genetic variants showed that, amongst these 26 genes, *POLR3C* and *RBFOX1* were significantly associated with ADHD in the largest SNP-based GWAS meta-analysis to date.

Materials and Methods

Identification of genes affected by ADHD CNVs

Coordinates of rare CNVs occurring in individuals diagnosed with ADHD (cases) were retrieved from 11 studies published until now (9,10,19,11–18) Discovery samples of in total 6176 ADHD cases and 25026 controls provided the bases for the analysed studies (see **Supplementary Table 1** for study characteristics). Coordinates of 1532 CNVs were retrieved for our analysis. These were mapped to the same reference human genome (hg19) using UCSC Lift Genome Annotations; the minimal ratio of bases that must remap was set to 0.95.(25) The CNV coordinates were used to retrieve RefGene information from the UCSC MySQL database, using a Structured Query Language (SQL) query (see **Supplementary Text 1**).(26) Information retrieved included the overlap with coding sequence, transcriptional direction, the total gene and CNV size in base pairs, exact gene start/end position, and percentage of gene coding sequence (CDS) represented by the CNV (**Supplementary Table 1**, Tab 1).

Selection of genes recurrently affected in ADHD CNVs

Transcript variants and biotypes were extracted for each CNV through a batch NCBI nucleotide query (**Supplementary Table 2**, Tab 1 for biotypes). For gene copy-number losses, we included all those genes entirely deleted or partially truncated. We also annotated, whether the N- or C-terminal region of transcripts were affected (**Supplementary Table 2**, Tab 1). For gene copy-number gains, we considered transcripts for which both a 2kb promoter and the coding region were entirely duplicated. For overlapping CNVs, the minimal region of overlap was identified to

narrow down the putative region involved in ADHD. Only mRNA-coding genes affected by CNVs in at least two cases with ADHD were selected for subsequent analyses (high-priority catalogue), given interpretability and possibility to perform cluster and protein-protein interaction analyses.

Co-expression network analysis

The BrainSpan developmental transcriptome data set (RNA-Seq Gencode v10) was used to investigate the overrepresentation of co-expressed genes of our high priority gene-set across all brain regions and developmental time points (embryo to adult), relative to the rest of the genome.(27) The expression coefficients for each mRNA-coding gene in all time points and brain regions in the BrainSpan data set were concatenated. The co-expression correlation score was calculated for each gene-pair. Gene-pairs with correlation score >0.3 were assigned to a co-expression network, each node representing a single gene and each connection representing the correlation score. The sum of the correlation scores for the investigated gene-set and for 10000 random gene-sets of the same size was calculated. An enrichment score was calculated by dividing the sum of the correlation scores per gene-set by the mean of the 10000 random gene-sets. P-values were calculated by comparing how many of the 10000 correlation scores of the random gene-sets were equal or higher than those of the investigated gene-set.

Integrated cross-species phenotype and protein-protein interaction network

We used the Monarch Initiative cross-species phenotype database to retrieve genes associated with an ADHD-related phenotype.(28) We defined core phenotypes, by selecting terms based on attention-deficit, hyperactivity, and impulsivity (**Supplementary Table 3**). The genes

connected to the cross-species core phenotypes of ADHD were subsequently superimposed to inter-species BioGRID protein interaction data. The interaction plot was visualised with Cytoscape 3.4.0.(29)

Identification of enriched-protein interaction modules

Networks of physical interactions in the gene-set were assessed with DAPPLE 0.17(30,31) based on InWeb data with the following parameter settings: Number of permutations: 1000 with adaptive permutation function, Plot: true, Seed File: genes. A set of 432 high-priority genes formed the input for this analysis. The modules were visualised with Cytoscape 3.4.0.(29)

Gene-based and gene-set association analyses of prioritized modules with ADHD risk

We used data from the recent meta-analysis of genome-wide association studies (GWAS) of 20,183 patients with ADHD and 35,191 controls as performed by the Psychiatric Genomics Consortium (PGC) ADHD Working Group and the Danish iPSYCH Initiative.(8) Details on the samples and quality control can be found in the **Supplementary Methods** and in Demontis et al..(8)

Gene-based association analyses were performed using the Multi-marker Analysis of GenoMic Annotation (MAGMA) software package (version 1.05; for details see **Supplementary Text 1**).(32) Genes were considered gene-wide significant if they reached the Bonferroni correction threshold-adjusted for the number of genes tested ($p<0.05/26$).

Results

Identification of genes located in ADHD-associated CNVs and definition of high-priority gene-set

We extracted data from the 11 studies reporting rare CNVs in a total discovery sample of 6176 ADHD cases and 25026 controls (see **Supplementary Table 1** for study characteristics). Coordinates of 1532 CNVs were retrieved, containing 2241 mRNA-coding candidate genes.

To identify the genes with an increased likelihood of contributing to ADHD pathology, we removed all genes duplicated with an incomplete promoter or coding sequence, or those aberrations found in controls across all studies (**Figure 1**). Genes identified in at least two rare CNVs were placed among the high-ranking candidates, due to their recurring nature. In addition, we calculated the minimal region common to overlapping CNVs, to narrow down the region of interest. Together, this resulted in a high-ranking list of 432 genes (**Supplementary Table 2**, **High-priority gene list**). The remaining 1316 genes, observed in only a single patient, were considered low-ranking (**Supplementary Table 2**, **Low-priority gene list**).

High-priority ADHD candidate genes show increased co-expression in the brain

It has been shown that proteins encoded by genes implicated in a genetically heterogeneous disorder tend to operate in common molecular pathways and processes.(22,33–36) To evaluate biological coherence of high-priority ADHD candidate genes in an unbiased way, we assessed their co-expression, a prerequisite for genes to jointly act in biological and developmental

processes, during the development of the most relevant tissue, the brain. We used the BrainSpan data set to test for gene co-expression and found a significant enrichment (E) of co-expressed genes in the high-priority gene-list (n=432; E=1.04, p=0.0044). The low-priority genes (n=1316; E=1.01, p=0.28) did not show significant co-expression enrichment.

Cross-species phenotypes link a network of 66 high-priority candidate genes to ADHD core symptoms

We used the Monarch Initiative cross-species genotype–phenotype database, which contains phenotypic information from 58 species, to i) evaluate our gene prioritisation, ii) retrieve independent evidence for the relevance of our high-priority candidate gene-set for core ADHD features, and iii) identify functionally associated networks of high-priority genes with the ADHD core symptoms hyperactivity, attention-deficit, and impulsivity (for exact search terms see **Supplementary Table 3**).(28) Eighteen of the 432 high-priority ADHD CNV candidate genes were associated with cross-species terms related to attention and hyperactivity: attention-deficit/hyperactivity disorder, hyperactivity, increased vertical activity, hyperactive, and abnormally increased process quality locomotory exploration behaviour **(Figure 2)**.

Based on the 18 genes validated by the cross-species approaches, we mined the cross-species Biological General Repository for Interaction Datasets (BioGRID) for interactors. This approach connected 48 additional genes of our high-priority catalogue to the cross-species terms **(Figure 2)**.

CNV-derived high-priority ADHD candidate genes form molecular modules implicating specific biological processes in ADHD pathology

In addition to the cross-species approach, we also used the DAPPLE algorithm to identify significantly connected proteins among the 432 high-priority genes, based on the integration of human data from protein–protein, genetic, and pathway interactions.(30) This algorithm depicted 17 modules of connected proteins, each comprising of 2–15 ADHD high-priority candidates that directly interacted with each other (**Figure 3**). Taking both direct and indirect interactions into account, the hubs with significant connectivity contained 20 proteins (**Figure 3** and **Supplementary Table 4**). Of those, eight significant proteins were found in four direct protein-interaction modules (**Figure 3, Module 1–4**). Of those modules, ten proteins, WWOX, PPM1F, PARK2, TUBA3C, MAPK1, MYC, SEPT5, POLR2B, POLR1A, and POLR3C were found connected to cross-species core ADHD phenotypes (**Figure 2**).

Two genes were found significantly associated with ADHD in common variant GWA studies

A set of 26 genes amongst the original 432 high-priority CNV-derived ADHD candidate genes was consistently observed in all the different approaches we employed (**Figure 4, Table 1**). We performed gene-based association analyses for all 26 genes to evaluate whether these are also implicated in ADHD risk through common genetic variants, using the largest meta-analytic GWAS data for ADHD currently available (n=55374, PGC-iPSYCH ADHD working groups). Two genes were significantly associated with ADHD after correction for multiple testing. These genes were *POLR3C* (p= 0.000020373) and *RBFOX1* (p= 0.00018202) (see **Supplementary Table 5 and**

Supplementary Figure 1). Interestingly, the *POLR3C* gene was among the top 0.43% of the most strongly associated genes (rank 80 out of 18411).

Discussion

Here we present an integrated analysis of ADHD-associated CNVs identified in the first 11 published studies. The limited power of the individual studies has been a crucial bottleneck in the definition of high priority ADHD candidate genes for further studies. From the 1532 CNVs described in the 11 studies, using strict criteria, we extracted 2241 mRNA-coding genes; this number is likely to contain many false positives due to the individually low occurrence of the rare CNVs (20), but the number is too large to allow each CNV to be studied in (animal) models for validation and mechanistic insights. Here, we aimed to prioritize genes linked to ADHD among the 2241 based on robustness of findings across different bioinformatics approaches used in an integrative manner and including both human and animal model-derived data. For this, we selected only those genes that were recurrently affected by CNVs in the patients and focussed on the minimal overlapping region of different CNVs in a region. Furthermore, we removed all genes that were affected by a CNV in healthy controls of other studies, and those that were only partially duplicated (lacking a full coding sequence and promoter). Those stringent criteria substantially reduced the number of candidate genes from the CNVs to one fifth of the original number. We showed that the selected 432 high-priority genes were significantly more co-expressed in the developing brain in comparison to random gene groups; this is evidence that our selection enriches for biologically coherent genes that are expressed at

the same time in the same tissue, a prerequisite for them to be involved in the same biological processes and cause similar phenotypes when disturbed.

Studies in model organisms can provide a wealth of phenotypic information, due to the high level of functional conservation across species. For ADHD, several model organisms have been shown to provide valid phenotypes. Those include monkey, rat, mouse, zebrafish, and fruit fly, where relevant phenotypes can be observed upon genetic manipulation or drug administration.(7) To identify biological processes underlying the selection of genes, we therefore took a novel approach that has not yet been applied to the field of neuropsychiatric disorders. We mined the Monarch Initiative database, which integrates genotype–phenotype relations across species, and found that 18 of the 432 high-priority genes, when manipulated in animal models, cause phenotypes that are face-valid to the ADHD core phenotypes attention-deficit, hyperactivity, and impulsivity. It is highly likely that the identified 18 genes whose protein-products form functional biological connections with genes/proteins for which detailed functional characterization, in particular information on ADHD-related phenotypes, is still lacking. We therefore retrieved direct interactors with the products of the 18 genes and identified an interconnected network of 66 proteins linked directly or indirectly to disease-relevant phenotypes **(Figure 2)**. We also tested, whether our high-priority gene list itself formed protein-interaction modules with significantly interconnected proteins, which could provide us with information on biological processes. Indeed, we identified four modules comprising significantly connected proteins from our selection. Of these, the major modules, module 1 and module 4, connected to and were supported by ADHD-related phenotypes across species, showing the added value of cross-species analysis **(Figure 2)**.

Module 1 contained two significantly linked proteins: WW-domain containing oxidoreductase (WWOX) and DNA-directed RNA polymerase I subunit RPA1 (POLR1A). WWOX, involved in autosomal recessive cerebellar ataxia-epilepsy-intellectual disability syndrome,(37) but no direct connections with ADHD are known. However, WWOX interacts with Protein phosphatase 1F (PPM1F), a Ser/Thr protein phosphatase that modulates RhoA and Ca^{2+}/calmodulin-dependent protein kinase II pathways.(38,39) Other members of this phosphatase protein family are involved in mediating dopaminergic signalling via G-protein-coupled receptors.(40) In addition, miss-expression of PPM1F in the substantia nigra in Parkinson's patients may implicate PPM1F more directly in dopaminergic biology.(41) Dopamine signalling pathways have been found altered repeatedly in ADHD patients and form the basis for the most widely used pharmacological treatment approach.(42,43) As we can directly connect PPM1F to the cross-species term hyperactivity and indirectly link *WWOX, POLR1A, POLR2B,* and *POLR3C* to ADHD core phenotypes, we can extend the network with genes potentially regulating dopaminergic signalling (**Figure 2**). The RNA polymerase II subunits (POLR1A, POLR2B, and POLR3C) are involved in the regulation and finetuning of transcription.

Module 2 clusters proteins required for blood–brain barrier formation, which function in cell–cell junctions and communication. This module contains one significantly connected protein: Catenin alpha-3 (CTNNA3). This adherence junction protein, also known to be associated with ASD, likely modulates cerebral and ependymal regions through GABA-A receptor activation.(44) Tight junction protein ZO-1 (TJP1) forms the connection to the other three proteins in this module. This gene is affected by CNVs in 28 ADHD cases, being the most frequently occurring gene affected by copy number alterations in our survey of the ADHD CNV

studies **(Supplementary Table 1)**. TJP1, together with Claudin-5 (CLDN5, affected in 12 ADHD cases) represents an important constituent of the blood–brain barrier.(45,46) Protein kinase C eta type (PRKCH) regulates TJP1.(47) Genes regulating neuronal cell adhesion are also significantly associated with ASD, schizophrenia, and bipolar disorder, raising the hypothesis that this mechanism plays a role across different neuropsychiatric disorders.(48,49) Due to the high number of ADHD cases with a CNV in this module, we postulate that cell–cell junctions play an important role in ADHD.

Module 3 contains two proteins that directly interact with each other: the significantly interconnected Density-regulated protein (DENR) and the Eukaryotic elongation factor 2 kinase (EEF2K), both involved in regulation and initiation of translation.(50) Regulation and initiation of translation has been linked to neuropsychiatric disorders including ADHD, through the regulation of brain-derived neurotrophic factor (BDNF).(51–53) Based on the repeated association and the described functional work, we suggest the involvement of transcriptional regulation as one of the mechanisms that modulate the ADHD risk.

Module 4 contains four significantly connected proteins: BIRC6, RAB15, SEPT5, and PHKB. Baculoviral IAP repeat-containing protein 6 (BIRC6 or Bruce) is an inhibitor of apoptosis involved in prostate cancer progression, but also acts in neuronal protection against apoptosis.(54,55) Ras-related protein Rab-15 (RAB15) is a direct connector of BIRC6, it plays a role in regulating synaptic vesicle membrane flow in nerve terminals.(56,57) Septin-5 (SEPT5) is involved in the binding of SNARE complexes, inhibiting synaptic vesicle exocytosis.(58) Recent studies have shown that manipulation of this gene in mice leads to altered social interaction and altered affective behaviours.(59,60) Phosphorylase b kinase regulatory subunit beta (PHKB) is

involved in glycogen metabolism and has been linked to neuronal plasticity.(61) Other proteins in the hub link to the cross-species phenotype term hyperactivity and attention-deficit/hyperactivity disorder: MAPK1 is directly connected, and SEPT5, PARK2, MYC, and TUBA3C are indirectly connected. They are thus prime candidates for further evaluation in functional assays.

While this study started from rare CNVs, we found corroborating evidence for several of the genes implicated in ADHD also in studies of common genetic variants. Among the 26 CNV-affected genes most consistently observed across the different bioinformatics approaches applied in this study, we also found *RBFOX1* and *POLR3C* to be associated with the disorder in the largest SNP-based GWAS meta-analysis to date. In a recent study of Lee et al., *RBFOX1* was discovered to be the second most pleiotropic locus of the genome-wide meta-analysis amongst eight psychiatric disorders.(62) *RBFOX1* encodes a splice regulator which is regulating several genes involved in neuronal development and which is mainly expressed in the brain.(62–65) Animal models have shown that *RBFOX1* is involved in mouse corticogenesis and aggressive behaviours.(62,64,66,67) *POLR3C* is included in a well-known small CNV located at 1q21.1, which contributes to a broad spectrum of phenotypes in addition to ADHD, including morphological features and autism spectrum disorders (ASD).(68,69) Genes listed in **Table 1** may be viewed as having, individually, the highest credibility as ADHD candidate genes. We therefore recommend to prioritize these genes in future studies searching for rare (single nucleotide) variants in ADHD and for functional characterization of gene-disease pathways. **Table 1** shows that also among these 26 genes, common biological themes are present, such as

transcription (already highlighted by the module analyse), mitochondria biology, mRNA metabolism, and cytoskeleton.

Our study should be viewed in the light of some strengths and limitations. We show that filtering based on recurrent CNVs restricted to ADHD cases in conjunction with complementary bioinformatics methods bares great potential to prioritise ADHD candidate genes. For the first time, cross-species phenotypes were used to identify candidate genes linked to ADHD core phenotypes, pointing to high-priority candidate genes. Several of those form significantly connected protein networks characterized by shared functions. The selection of genes analysed in this study is extensive, but is by no means exhaustive. First, it is important to note that by only analysing recurrent CNVs and focussing on the minimal regions of overlap amongst CNVs in the same region, genes with relevance for ADHD may be overlooked; on the other hand, the stringent evidence-based filtering holds high potential to uncover the most ADHD-relevant biological pathways. In addition, neither the surveyed CNV studies nor our study consider effects of CNVs onto the surrounding genetic landscape. A duplicated CNV translocation, for example, can have an impact on the expression of genes in the chromosomal region at the site of insertion, which alone or together with the duplicated genes can contribute to ADHD phenotypes.(70) Lastly, the criteria used for selection of CNVs were not identical across the 11 source studies.

In conclusion, our study shows that stringent filtering in CNV studies in combination with a complementary battery of bioinformatics approaches can identify ADHD candidate genes at increased levels of credibility. We suggest that testing genes operating in the identified modules and those 26 consistently found in all approaches employed here in animal models like rat,

346 mouse, zebrafish, and fruit fly,(7) for their ability to modulate behaviour will provide further

347 insights into the mechanistic pathways and biology of ADHD.

348

349

Legends to figures

Figure 1. CNV processing scheme to establish high- and low-priority ADHD candidate genes.

(**a**) Collection: CNV coordinates were extracted from 11 different studies reporting rare CNVs in ADHD cohorts (see **Supplementary Table 1**). (**b**) Cleaning: CNVs identified in controls and copy-number gains with an incomplete promoter and coding sequences were excluded. (**c**) Prioritisation: a high-priority gene list was generated by excluding genes found only in a single individual with ADHD and genes not present in the minimal overlapping region.

Figure 2. Cross-species phenotypes link an interaction network of 66 high-priority candidate genes to ADHD core symptoms

Of the 432 high-priority ADHD CNV candidate genes, 18 were found in the Monarch Initiative database to have a core ADHD phenotype relating to hyperactivity, attention deficit, and impulsivity. These 18 genes were used as seeds to create a BioGRID interaction network, using cross-species information on direct protein–protein interactions, genetic interactions and predicted interactions. We found a highly interconnected network with 48 secondary interactors from the high-priority gene list. Dashed lines show whether a connection was found once (¨) or multiple times (--) in the BioGRID database.

Figure 3. DAPPLE interaction network identifies four molecular modules with significantly connected proteins.

A DAPPLE analysis was performed to identify highly related protein–protein interaction clusters based on direct- and indirect connectivity. Four modules containing significantly connected proteins were identified.

Figure 4. Visual representation of the gene-selection process.

A summary of the analysis process undertaken in the paper, starting from 2241 genes prioritised to 432 high-priority candidates (**Figure 1**), followed by a cross-species phenotype and BioGRID analysis of 66 genes (**Figure 2**) and DAPPLE analysis with 62 candidates (**Figure 3**). In total 26 candidates overlap across analyses (**Table 1**).

Table 1. The 26 genes consistently identified across all bioinformatics approaches employed in the current study.

Supplementary Figure 1: Regional association of genes most strongly associated with ADHD. Regional association plots showing association signal for ADHD in the PGC-iPSYCH GWAS meta-analysis data for the two most strongly associated high priority ADHD candidate genes, including flanking regions of 100kb. (A) *POLR3C* locus with the top-SNP (rs376814422) indicated by the black arrow. (B) *RBFOX1* locus with the top-SNP (rs6500945) indicated by the purple dot. Results are shown as -log (p-value) for genotyped and imputed SNPs. The colour of each marker reflects its LD (r2) with the SNP indicated by the purple dot. The recombination rate is plotted in blue. cm/Mb, centimorgan/megabase. Chr, chromosome.

392 **Supplementary Table 1:** Full lists of raw input data points extracted from 11 ADHD CNV studies
393 (Tab 1), high-priority gene selection (Tab 2) and low-priority gene selection (Tab 3).
394
395 **Supplementary Table 2:** Full list of all selected ADHD core phenotype labels in the Monarch
396 Initiative database query.
397 **Supplementary Table 3:** Overview of all included studies reporting rare CNVs in an ADHD
398 cohort.
399 **Supplementary Table 4:** Significant connected genes of the high-priority gene list after DAPPLE
400 analysis and the assigned modules in **Figure 3**.
401 **Supplementary Table 5:** Results MAGMA analysis of the 26 genes consistently observed in all
402 the different approaches
403 **Supplementary Methods 1:** SQL query script for the UCSC database, GWAS meta-analyses data
404 set for ADHD and Gene-based association analyses for ADHD GWAS meta-analyses data
405
406
407

Author contribution

BH: performed data analysis and created interaction networks, contributed to manuscript writing. MV: conceived the study, performed data analysis, processed CNVs using SQL, curated cross-species Monarch Initiative and BioGRID information, contributed to supervision and manuscript writing. MK: Performed gene-based association analyses and helped with data interpretation. PC and MF: performed the co-expression analysis. BF: contributed to supervision and manuscript writing. AS: conceived the study, contributed to supervision and manuscript writing. All authors proof-read the manuscript.

Acknowledgements

Benjamin Harich was funded by a Radboud University Medical Center PhD grant. Monique van der Voet's research was supported by the Netherlands Organization for Scientific Research (NWO) VENI grant (grant 91.614.084). Marieke Klein was supported by a grant from the Netherlands Science Agenda (NWA) for the NeuroLabNL project (grant 400.17.602). Barbara Franke's contribution was supported by a personal Vici grant from the Netherlands Organization for Scientific Research (NWO; grant 016-130-669). This work also received support from the European Community's Horizon 2020 (H2020/2014 – 2020) European Trainings Network Programme MiND under grant agreement n° 643051 to Barbara Franke and Annette Schenck. Additional support was received from the ECNP Network ADHD across the Lifespan and the European Community's Horizon 2020 Programme under grant agreements n° 667302 (CoCA) and n° 728018 (Eat2beNICE). This report reflects only the authors' views, and the European Union is not liable for any use that may be made of the information contained therein.

Conflict of interest

Barbara Franke has received educational speaking fees from Medice. None of the other authors report conflicts of interest.

Bibliography

1. Polanczyk G, De Lima MS, Horta BL, Biederman J, Rohde LA. The worldwide prevalence of ADHD: A systematic review and metaregression analysis. Am J Psychiatry [Internet]. 2007 Jun;164(6):942–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/17541055
2. Franke B, Faraone S V, Asherson P, Buitelaar J, Bau CHD, Ramos-Quiroga JA, et al. The genetics of attention deficit/hyperactivity disorder in adults, a review. Mol Psychiatry [Internet]. 2012;17(10):960–87. Available from: http://dx.doi.org/10.1038/mp.2011.138%5Cnpapers2://publication/doi/10.1038/mp.2011.138
3. Faraone S V, Asherson P, Banaschewski T, Biederman J, Buitelaar JK, Ramos-Quiroga JA, et al. Attention-deficit/hyperactivity disorder. Vol. 1, Nature reviews. Disease primers. England; 2015. p. 15020.
4. Gizer IR, Ficks C, Waldman ID. Candidate gene studies of ADHD: a meta-analytic review. Hum Genet [Internet]. 2009 Jul [cited 2014 May 24];126(1):51–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19506906
5. Franke B, Neale BM, Faraone S V. Genome-wide association studies in ADHD. Hum Genet [Internet]. 2009 Jul [cited 2012 Mar 14];126(1):13–50. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19384554

6. Faraone S V., Asherson P, Banaschewski T, Biederman J, Buitelaar JK, Ramos-Quiroga JA, et al. Attention-deficit/hyperactivity disorder. Nat Rev Dis Prim [Internet]. 2015;1:15020. Available from: http://www.nature.com/articles/nrdp201520

7. Klein M, Onnink M, van Donkelaar M, Wolfers T, Harich B, Shi Y, et al. Brain imaging genetics in ADHD and beyond - Mapping pathways from gene to disorder at different levels of complexity. Neurosci Biobehav Rev. 2017 Jan;80:115–55.

8. Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery of the first genome-wide significant risk loci for attention deficit/hyperactivity disorder. Nat Genet [Internet]. 2019;51(1):63–75. Available from: https://doi.org/10.1038/s41588-018-0269-7

9. Williams N, Franke B. Genome-wide analysis of copy number variants in attention deficit hyperactivity disorder: the role of rare variants and duplications at 15q13. 3. Am J Psychiatry. 2012;169(2):195–204.

10. Williams NM, Zaharieva I, Martin A, Langley K, Mantripragada K, Fossdal R, et al. Rare chromosomal deletions and duplications in attention-deficit hyperactivity disorder: a genome-wide analysis. Lancet [Internet]. 2010 Oct 23 [cited 2013 Nov 15];376(9750):1401–8. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2965350&tool=pmcentrez&rendertype=abstract

11. Stergiakouli E. Investigating the contribution of common genetic variants to the risk and pathogenesis of ADHD. Am J Psychiatry [Internet]. 2012 [cited 2014 Feb 11];169:186–94. Available from: http://journals.psychiatryonline.org/article.aspx?articleid=483662

12. Ramos-Quiroga J-A, Sánchez-Mora C, Casas M, Garcia-Martínez I, Bosch R, Nogueira M, et al. Genome-wide copy number variation analysis in adult attention-deficit and hyperactivity disorder. J Psychiatr Res [Internet]. 2014 Feb [cited 2014 Aug 11];49:60–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24269040

13. Martin J, Cooper M, Hamshere ML, Pocklington A, Scherer SW, Kent L, et al. Biological overlap of attention-deficit/hyperactivity disorder and autism spectrum disorder: evidence from copy number variants. J Am Acad Child Adolesc Psychiatry [Internet]. 2014 Jul [cited 2014 Jul 13];53(7):761-770.e26. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=4074351&tool=pmcentrez&rendertype=abstract

14. Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J, et al. Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD. Sci Transl Med [Internet]. 2011 Aug 10 [cited 2014 Aug 9];3(95):95ra75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/21832240

15. Lesch K-P, Selch S, Renner TJ, Jacob C, Nguyen TT, Hahn T, et al. Genome-wide copy number variation analysis in attention-deficit/hyperactivity disorder: association with neuropeptide Y gene dosage in an extended pedigree. Mol Psychiatry [Internet]. 2011 May [cited 2012 Oct 26];16(5):491–503. Available from: http://www.ncbi.nlm.nih.gov/pubmed/20308990

16. Jarick I, Volckmar A-L, Pütter C, Pechlivanis S, Nguyen TT, Dauvermann MR, et al. Genome-wide analysis of rare copy number variations reveals PARK2 as a candidate gene for attention-deficit/hyperactivity disorder. Mol Psychiatry [Internet]. 2014 Jan [cited

2014 Jul 19];19(1):115–21. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3873032&tool=pmcentrez&rendertype=abstract

17. Elia J, Glessner JT, Wang K, Takahashi N, Shtir CJ, Hadley D, et al. Genome-wide copy number variation study associates metabotropic glutamate receptor gene networks with attention deficit hyperactivity disorder. Nat Genet [Internet]. 2012 Jan [cited 2012 Oct 4];44(1):78–84. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22138692

18. Elia J, Gai X, Xie HM, Perin JC, Geiger E, Glessner JT, et al. Rare structural variants found in attention-deficit hyperactivity disorder are preferentially associated with neurodevelopmental genes. Mol Psychiatry [Internet]. 2010 Jun [cited 2012 Oct 26];15(6):637–46. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2877197&tool=pmcentrez&rendertype=abstract

19. Yang L, Neale BM, Liu L, Lee SH, Wray NR, Ji N, et al. Polygenic transmission and complex neuro developmental network for attention deficit hyperactivity disorder: Genome-wide association study of both common and rare variants. Am J Med Genet Part B Neuropsychiatr Genet [Internet]. 2013;162(5):419–30. Available from: http://doi.wiley.com/10.1002/ajmg.b.32169

20. Gratten J, Visscher PM, Mowry BJ, Wray NR. Interpreting the role of de novo protein-coding mutations in neuropsychiatric disease. Nat Genet [Internet]. 2013;45(3):234–8. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23438595

21. Hormozdiari F, Penn O, Borenstein E, Eichler EE. The discovery of integrated gene

networks for autism and related disorders. Genome Res. 2015 Jan;25(1):142–54.

22. Thapar A, Martin J, Mick E, Vásquez AA, Langley K, Scherer SW, et al. Psychiatric gene discoveries shape evidence on ADHD ' s biology. 2015;21(9):1–6.

23. Doelken SC, Köhler S, Mungall CJ, Gkoutos G V, Ruef BJ, Smith C, et al. Phenotypic overlap in the contribution of individual genes to CNV pathogenicity revealed by cross-species computational analysis of single-gene mutations in humans, mice and zebrafish. Dis Model Mech [Internet]. 2013 Mar [cited 2014 Feb 11];6(2):358–72. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3597018&tool=pmcentrez&rendertype=abstract

24. Köhler S, Doelken SC, Ruef BJ, Bauer S, Washington N, Westerfield M, et al. Construction and accessibility of a cross-species phenotype ontology along with gene annotations for biomedical research. F1000Research [Internet]. 2013 Jan [cited 2014 Feb 3];2:1–12. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3799545&tool=pmcentrez&rendertype=abstract

25. UCSC Lift Genome Annotations [Internet]. Available from: https://genome.ucsc.edu/cgi-bin/hgLiftOver

26. UCSC MySQL database [Internet]. Available from: genome-mysql.cse.ucsc.edu

27. Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, Shen EH, Ng L, Miller JA, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. Nature [Internet]. 2012 Sep 20;489(7416):391–9. Available from: http://dx.doi.org/10.1038/nature11405

28. Mungall CJ, McMurry JA, Kohler S, Balhoff JP, Borromeo C, Brush M, et al. The Monarch Initiative: An integrative data and analytic platform connecting phenotypes to genotypes across species. Nucleic Acids Res. 2017;45(D1):D712–22.

29. Cline M, Smoot M, Cerami E. Integration of biological networks and gene expression data using Cytoscape. Nat Protoc [Internet]. 2007;2(10):2366–82. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3685583&tool=pmcentrez&rendertype=abstract%5Cnhttp://www.nature.com/nprot/journal/v2/n10/abs/nprot.2007.324.html

30. Rossin EJ, Lage K, Raychaudhuri S, Xavier RJ, Tatar D, Benita Y, et al. Proteins encoded in genomic regions associated with immune-mediated disease physically interact and suggest underlying biology. PLoS Genet [Internet]. 2011 Jan [cited 2014 Jul 22];7(1):e1001273. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3020935&tool=pmcentrez&rendertype=abstract

31. Reich M, Liefeld T, Gould J, Lerner J, Tamayo P, Mesirov JP. GenePattern 2.0. Nat Genet [Internet]. 2006;38(5):500–1. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16642009

32. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol. 2015;11(4):1–19.

33. O'Dushlaine C, Rossin L, Lee PH, Duncan L, Parikshak NN, Newhouse S, et al. Psychiatric genome-wide association study analyses implicate neuronal, immune and histone pathways. Nat Neurosci [Internet]. 2015;18(2):199–209. Available from:

http://www.nature.com/doifinder/10.1038/nn.3922%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/25599223%5Cnhttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC4378867%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/25599223%5Cnhttp://www.pubmedcentral.nih.gov/articlerender

34. Mooney MA, McWeeney SK, Faraone S, Hinney A, Hebebrand J, Nigg JT, et al. Pathway analysis in attention deficit hyperactivity disorder: An ensemble approach. Am J Med Genet Part B Neuropsychiatr Genet [Internet]. 2016;171:815–26. Available from: http://doi.wiley.com/10.1002/ajmg.b.32446

35. Subramanian M, Timmerman CK, Schwartz JL, Pham DL, Meffert MK. Characterizing autism spectrum disorders by key biochemical pathways. 2015;9(September):1–18.

36. Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F, et al. Systematic Phenomics Analysis Deconvolutes Genes Mutated in Intellectual Disability into Biologically Coherent Modules. Am J Hum Genet [Internet]. 2016;98(1):149–64. Available from: http://dx.doi.org/10.1016/j.ajhg.2015.11.024

37. Mallaret M, Synofzik M, Lee J, Sagum CA, Mahajnah M, Sharkia R, et al. The tumour suppressor gene WWOX is mutated in autosomal recessive cerebellar ataxia with epilepsy and mental retardation. Brain. 2014;137(2):411–9.

38. Harvey BP, Banga SS, Ozer HL. Regulation of the multifunctional Ca2+/calmodulin-dependent protein kinase II by the PP2C phosphatase PPM1F in fibroblasts. J Biol Chem. 2004;279(23):24889–98.

39. Xie Y, Tan E-J, Wee S, Manser E, Lim L, Koh C-G. Functional interactions between phosphatase POPX2 and mDia modulate RhoA pathways. J Cell Sci. 2008;121(Pt 4):514–

21.

40. Marion S, Urs NM, Peterson SM, Sotnikova TD, Beaulieu J-M, Gainetdinov RR, et al. Dopamine D2 Receptor Relies upon PPM/PP2C Protein Phosphatases to Dephosphorylate Huntingtin Protein. J Biol Chem [Internet]. 2014 Apr 25;289(17):11715–24. Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4002081/

41. Simunovic F, Yi M, Wang Y, Macey L, Brown LT, Krichevsky AM, et al. Gene expression profiling of substantia nigra dopamine neurons: further insights into Parkinson's disease pathology. Brain. 2009;132(7):1795–809.

42. Swanson J., Flodman P, Kennedy J, Spence MA, Moyzis R, Schuck S, et al. Dopamine genes and ADHD. Neurosci Biobehav Rev [Internet]. 2005 Jan [cited 2017 Aug 22];24(1):21–5. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0149763499000627

43. Volkow ND, Wang G-J, Fowler JS, Ding Y-S. Imaging the Effects of Methylphenidate on Brain Dopamine: New Model on Its Therapeutic Actions for Attention-Deficit/Hyperactivity Disorder. Biol Psychiatry [Internet]. 2005 Jun [cited 2017 Aug 22];57(11):1410–5. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0006322304011631

44. Folmsbee SS, Wilcox DR, Tyberghein K, De Bleser P, Tourtellotte WG, van Hengel J, et al. αT-catenin in restricted brain cell types and its potential connection to autism. J Mol Psychiatry [Internet]. 2016;4(1):2. Available from: http://jmolecularpsychiatry.biomedcentral.com/articles/10.1186/s40303-016-0017-9

45. Lee J, Freeman JL. Zebrafish as a Model for Developmental Neurotoxicity Assessment: The Application of the Zebrafish in Defining the Effects of Arsenic, Methylmercury, or Lead on

606 Early Neurodevelopment. Toxics [Internet]. 2014;2(3):464–95. Available from:
607 http://www.mdpi.com/2305-6304/2/3/464/
608 46. Greene C, Campbell M. Tight junction modulation of the blood brain barrier: CNS delivery
609 of small molecules. Tissue Barriers [Internet]. 2016 Jan 2;4(1):e1138017. Available from:
610 http://dx.doi.org/10.1080/21688370.2015.1138017
611 47. Suzuki T, Elias BC, Seth A, Shen L, Turner JR, Giorgianni F, et al. PKC eta regulates occludin
612 phosphorylation and epithelial tight junction integrity. Proc Natl Acad Sci [Internet].
613 2009;106(1):61–6. Available from:
614 http://eutils.ncbi.nlm.nih.gov/entrez/eutils/elink.fcgi?dbfrom=pubmed&id=19114660&re
615 tmode=ref&cmd=prlinks%5Cnpapers3://publication/doi/10.1073/pnas.0802741106
616 48. Marui T, Funatogawa I, Koishi S, Yamamoto K, Matsumoto H, Hashimoto O, et al.
617 Association of the neuronal cell adhesion molecule (NRCAM) gene variants with autism.
618 Int J Neuropsychopharmacol. 2009;12(1):1–10.
619 49. O'Dushlaine C, Kenny E, Heron E, Donohoe G, Gill M, Morris D, et al. Molecular pathways
620 involved in neuronal cell adhesion and membrane scaffolding contribute to schizophrenia
621 and bipolar disorder susceptibility. Mol Psychiatry. 2011;16(3):286–92.
622 50. Smith EM, Proud CG. cdc2–cyclin B regulates eEF2 kinase activity in a cell cycle- and
623 amino acid-dependent manner. EMBO J [Internet]. 2008 Mar 13;27(7):1005 LP – 1016.
624 Available from: http://emboj.embopress.org/content/27/7/1005.abstract
625 51. Lipton JO, Sahin M. The neurology of mTOR. Neuron [Internet]. 2014;84(2):275–91.
626 Available from: http://www.ncbi.nlm.nih.gov/pubmed/25374355
627 52. Autry AE, Adachi M, Nosyreva E, Na ES, Los MF, Cheng P, et al. NMDA receptor blockade

at rest triggers rapid behavioural antidepressant responses. Nature [Internet]. 2011;475(7354):91–5. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3172695&tool=pmcentrez&rendertype=abstract

53. Hawi Z, Cummins TDR, Tong J, Arcos-Burgos M, Zhao Q, Matthews N, et al. Rare DNA variants in the brain-derived neurotrophic factor gene increase risk for attention-deficit hyperactivity disorder: a next-generation sequencing study. Mol Psychiatry [Internet]. 2017;22(4):580–4. Available from: http://europepmc.org/abstract/MED/27457811

54. Low CG, Luk ISU, Lin D, Fazli L, Yang K, Xu Y, et al. BIRC6 Protein, an Inhibitor of Apoptosis: Role in Survival of Human Prostate Cancer Cells. PLoS One. 2013;8(2):e55837.

55. Sokka AL, Mudo G, Aaltonen J, Belluardo N, Lindholm D, Korhonen L. Bruce/apollon promotes hippocampal neuron survival and is downregulated by kainic acid. Biochem Biophys Res Commun. 2005;338(2):729–35.

56. Elferink LA, Anzai K, Scheller RH. rab15, a novel low molecular weight GTP-binding protein specifically expressed in rat brain. J Biol Chem. 1992;267(9):5768–75.

57. Stenmark H. Rab GTPases as coordinators of vesicle traffic. Nat Rev Mol Cell Biol [Internet]. 2009;10(8):513–25. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19603039

58. Beites CL, Campbell KA, Trimble WS. The septin Sept5/CDCrel-1 competes with alpha-SNAP for binding to the SNARE complex. Biochem J [Internet]. 2005;385(Pt 2):347–53. Available from: http://www.biochemj.org/content/385/2/347.abstract

59. Suzuki G, Harper KM, Hiramoto T, Sawamura T, Lee M, Kang G, et al. Sept5 deficiency

650 exerts pleiotropic influence on affective behaviors and cognitive functions in mice. Hum
651 Mol Genet. 2009;18(9):1652–60.
652 60. Harper KM, Hiramoto T, Tanigaki K, Kang G, Suzuki G, Trimble W, et al. Alterations of
653 social interaction through genetic and environmental manipulation of the 22q11.2 gene
654 sept5 in the mouse brain. Hum Mol Genet. 2012;21(15):3489–99.
655 61. Tadi M, Allaman I, Lengacher S, Grenningloh G, Magistretti PJ. Learning-Induced gene
656 expression in the hippocampus reveals a role of neuron -Astrocyte metabolic coupling in
657 long term memory. PLoS One. 2015;10(10):1–21.
658 62. Lee PH, Anttila V, Won H, Feng Y-CA, Rosenthal J, Zhu Z, et al. Genome wide meta-
659 analysis identifies genomic relationships, novel loci, and pleiotropic mechanisms across
660 eight psychiatric disorders. bioRxiv [Internet]. 2019 Jan 1;528117. Available from:
661 http://biorxiv.org/content/early/2019/01/26/528117.abstract
662 63. Gehman LT, Stoilov P, Maguire J, Damianov A, Lin C-H, Shiue L, et al. The splicing
663 regulator Rbfox1 (A2BP1) controls neuronal excitation in the mammalian brain. Nat
664 Genet. 2011 May;43(7):706–11.
665 64. Hamada N, Ito H, Iwamoto I, Morishita R, Tabata H, Nagata K. Role of the cytoplasmic
666 isoform of RBFOX1/A2BP1 in establishing the architecture of the developing cerebral
667 cortex. Mol Autism [Internet]. 2015;6(1):56. Available from:
668 https://doi.org/10.1186/s13229-015-0049-5
669 65. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, et al. Shared
670 molecular neuropathology across major psychiatric disorders parallels polygenic overlap.
671 Science. 2018 Feb;359(6376):693–7.

672 66. Hamada N, Ito H, Nishijo T, Iwamoto I, Morishita R, Tabata H, et al. Essential role of the
673 nuclear isoform of RBFOX1, a candidate gene for autism spectrum disorders, in the brain
674 development. Sci Rep [Internet]. 2016 Aug 2;6:30805. Available from:
675 https://doi.org/10.1038/srep30805
676 67. Fernandez-Castillo N, Gan G, van Donkelaar MMJ, Vaht M, Weber H, Retz W, et al.
677 RBFOX1, encoding a splicing regulator, is a candidate gene for aggressive behavior. Eur
678 Neuropsychopharmacol. 2017 Nov;
679 68. Busè M, Cuttaia HC, Palazzo D, Mazara M V, Lauricella SA, Malacarne M, et al. Expanding
680 the phenotype of reciprocal 1q21.1 deletions and duplications: a case series. Ital J Pediatr
681 [Internet]. 2017 Jul;43(1):61. Available from: https://doi.org/10.1186/s13052-017-0380-x
682 69. Goodbourn PT, Bosten JM, Bargary G, Hogg RE, Lawrance-Owen AJ, Mollon JD. Variants in
683 the 1q21 risk region are associated with a visual endophenotype of autism and
684 schizophrenia. Genes Brain Behav [Internet]. 2014;13:144–51. Available from:
685 https://doi.org/10.1111/gbb.12096
686 70. Zhang F, Lupski JR. Non-coding genetic variants in human disease. Hum Mol Genet. 2015
687 Oct;24(R1):R102-10.
688

Figure 1

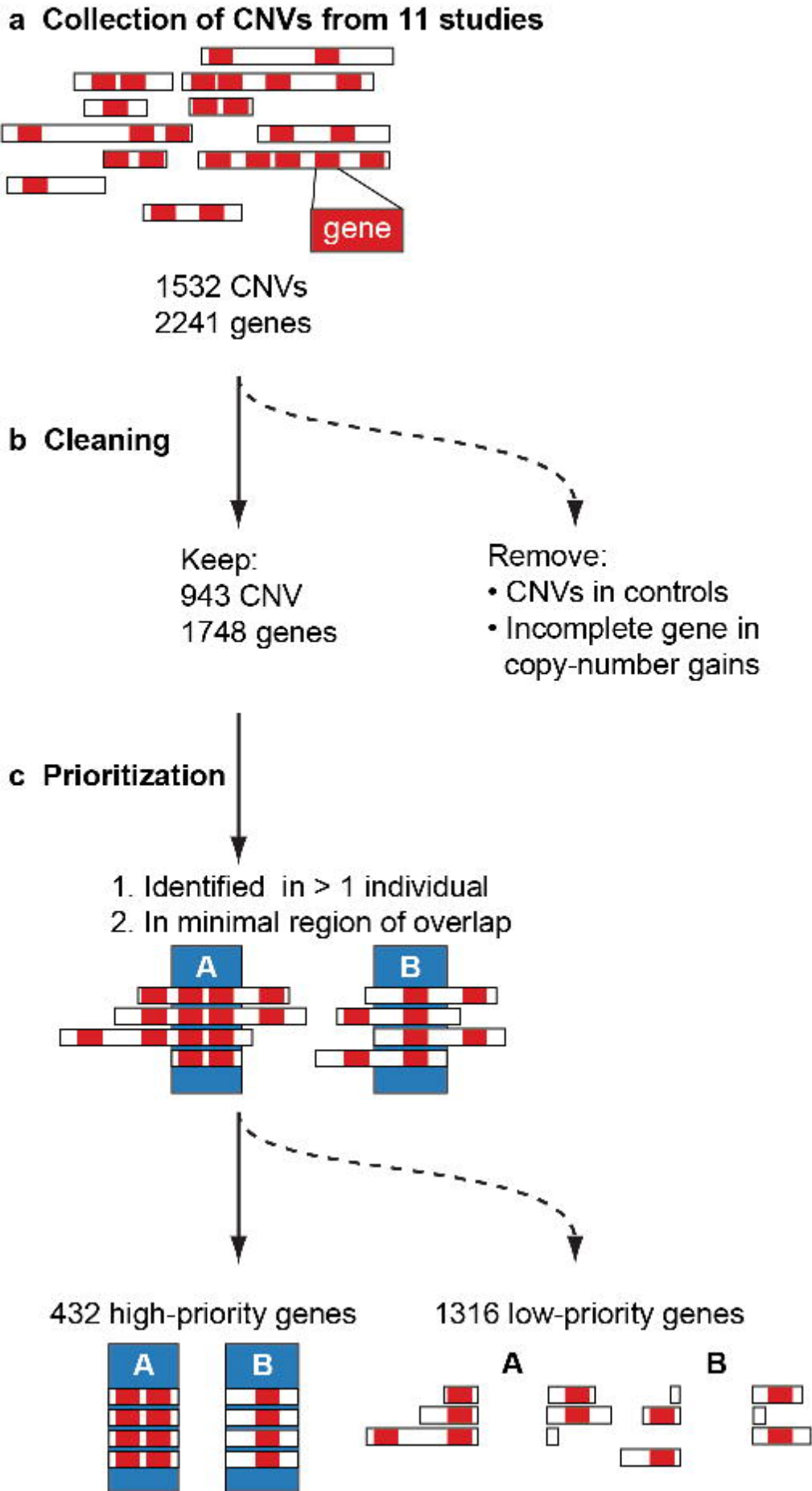
a Collection of CNVs from 11 studies
gene
1532 CNVs
2241 genes
b Cleaning
Keep:
943 CNV
1748 genes
Remove:
• CNVs in controls
• Incomplete gene in copy-number gains
c Prioritization
1. Identified in > 1 individual
2. In minimal region of overlap
A
B
432 high-priority genes
1316 low-priority genes
A
B
A
B

Figure 2

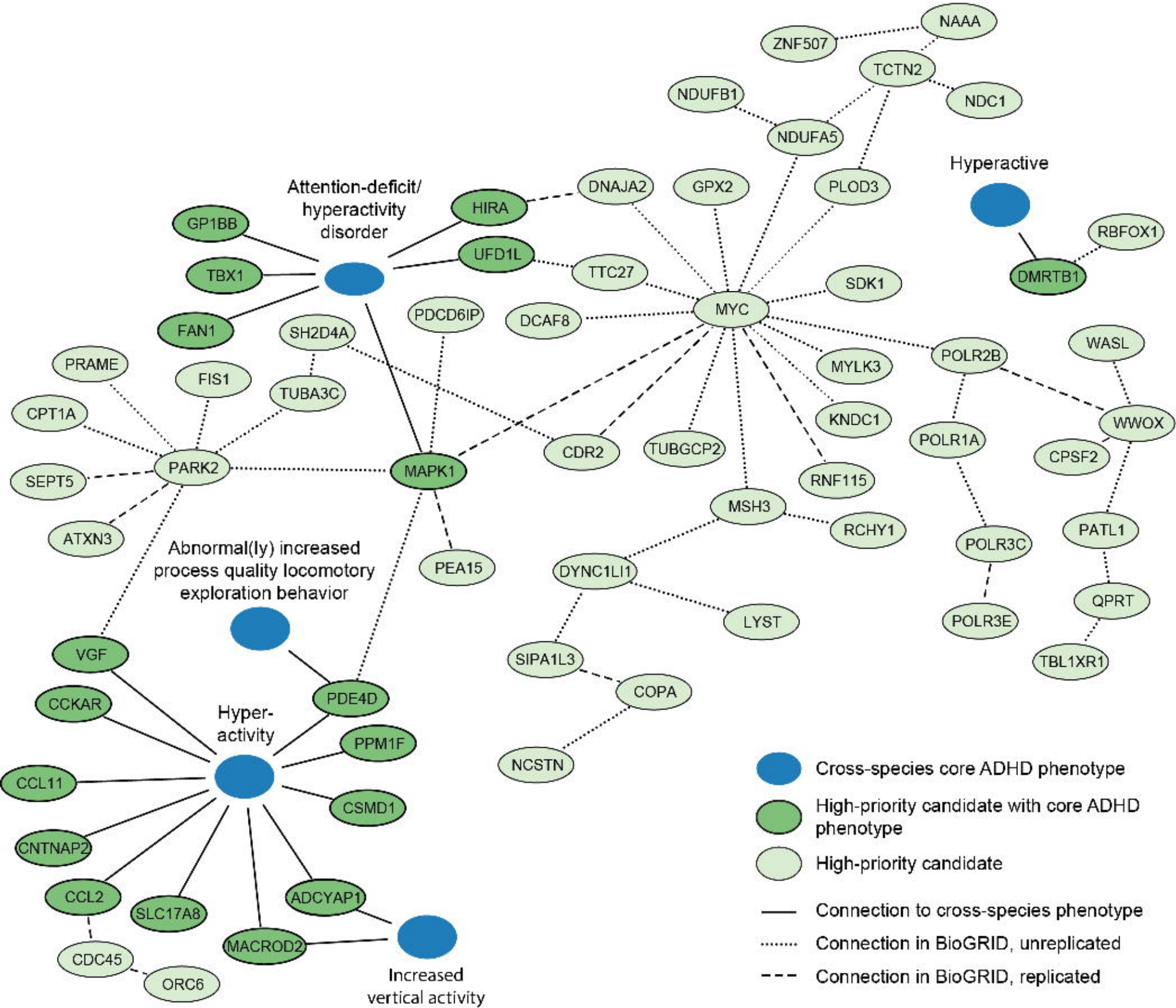
NAAA
ZNF507
TCTN2
NDUFB1
NDC1
NDUFA5
Hyperactive
Attention-deficit/
hyperactivity
disorder
DNAJA2
GPX2
PLOD3
HIRA
GP1BB
RBFOX1
UFD1L
TBX1
TTC27
DMRTB1
SDK1
MYC
PDCD6IP
DCAF8
FAN1
SH2D4A
WASL
POLR2B
PRAME
MYLK3
FIS1
TUBA3C
KNDC1
CPT1A
WWOX
POLR1A
CDR2
TUBGCP2
CPSF2
PARK2
MAPK1
SEPT5
RNF115
MSH3
RCHY1
ATXN3
PATL1
POLR3C
Abnormal(ly) increased
process quality locomotory
exploration behavior
PEA15
DYNC1LI1
QPRT
LYST
POLR3E
VGF
SIPA1L3
TBL1XR1
COPA
PDE4D
CCKAR
Hyper-
activity
PPM1F
NCSTN
CCL11
CSMD1
CNTNAP2
CCL2
SLC17A8
ADCYAP1
MACROD2
CDC45
ORC6
Increased
vertical activity
Cross-species core ADHD phenotype
High-priority candidate with core ADHD phenotype
High-priority candidate
Connection to cross-species phenotype
Connection in BioGRID, unreplicated
Connection in BioGRID, replicated

Figure 3

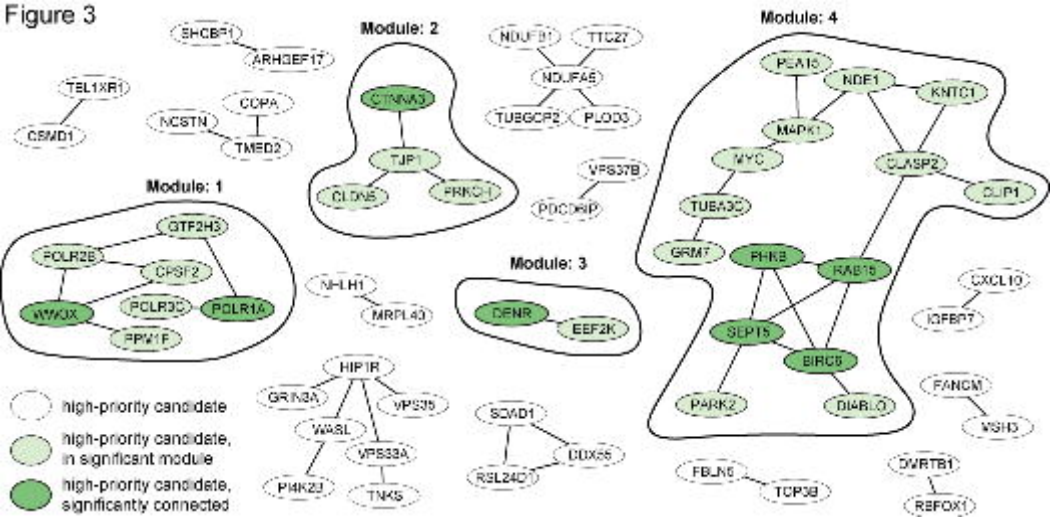
SHCBP1
ARHGEF17
TBL1XR1
CSMD1
NCSTN
COPA
TMED2
Module: 2
CTNNA3
TJP1
CLDN5
FRK
NDUFB1
TTC27
NDUFA5
TUBGCP2
PLOD3
VPS37B
PDCD6IP
Module: 4
PEAK1
NDE1
KNTC1
MAPK1
MYC
CLASP2
CLIP1
TUBA3C
GRM7
PHKB
RAB15
SEPT5
BIRC5
PARK2
DIABLO
CXCL10
IGFBP7
FANCM
WSH3
DVRTB1
REFOX1
Module: 1
GTF2H3
POLR2E
CPSF2
WWOX
POLR3D
POLR1A
PPWD1
NHLH1
MRPL43
Module: 3
CPNE
EEF2K
HIP1R
GRM3A
VPS35
WASL
VPS33A
PIK3CJ
TNKS
SCAD1
DDX55
RSL24D1
FBLN5
TCP38
high-priority candidate
high-priority candidate, in significant module
high-priority candidate, significantly connected

Figure 4

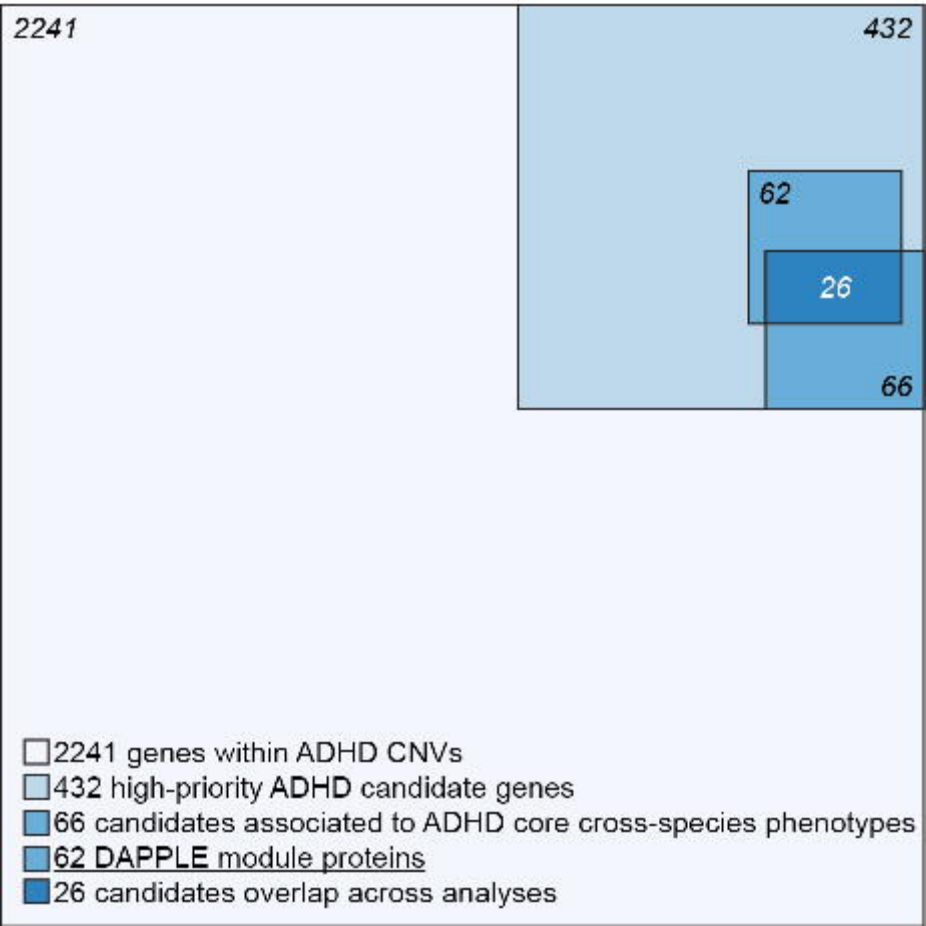
2241
432
62
26
66
2241 genes within ADHD CNVs
432 high-priority ADHD candidate genes
66 candidates associated to ADHD core cross-species phenotypes
62 DAPPLE module proteins
26 candidates overlap across analyses

Table 1

Gene symbol	Full gene name	Gene function	Biological themes
COPA	coatomer protein complex subunit alpha	protein transport between the endoplasmic reticulum and Golgi compartments	
CPSF2	cleavage and polyadenylation specific factor 2	pre-mRNA 3-end formation	mRNA metabolism
CSMD1	CUB And Sushi Multiple Domains 1	unknown	
DMRTB1	DMRT Like Family B With Proline Rich C-Terminal 1	transcription factor	transcription
MAPK1	Mitogen-Activated Protein Kinase 1	wide range of cellular processes like proliferation, differentiation, transcription regulation and development	
MSH3	MutS Homolog 3	DNA mismatch repair	
MYC	MYC Proto-Oncogene, BHLH Transcription Factor	transcription factor	transcription
NCSTN	Nicastrin	type I transmembrane glycoprotein	
NDUFA5	NADH:Ubiquinone Oxidoreductase Subunit A5	subunit of complex I of the mitochondrial respiratory chain	mitochondria
NDUFB1	NADH:Ubiquinone Oxidoreductase Subunit B1	subunit of complex I of the mitochondrial respiratory chain	mitochondria
PARK2	Parkin RBR E3 Ubiquitin Protein Ligase	component of a multiprotein E3 ubiquitin ligase complex, degradation of defective mitochondria	mitochondria
PDCD6IP	Programmed Cell Death 6 Interacting Protein	functions within the ESCRT pathway	
PEA15	Phosphoprotein Enriched In Astrocytes 15	negative regulator of apoptosis	
PLOD3	Procollagen-Lysine,2-Oxoglutarate 5-Dioxygenase 3	homodimeric enzyme in the rough enodplasmatic reticulum	
POLR1A	RNA Polymerase I Subunit A	subunit of the RNA polymerase I complex	mRNA metabolism
POLR2B	RNA Polymerase II Subunit B	subunit of the RNA polymerase II complex	mRNA metabolism
POLR3C	RNA Polymerase III Subunit C	subunit of the RNA polymerase III complex	mRNA metabolism
PPM1F	Protein Phosphatase, Mg2+/Mn2+ Dependent 1F	Ser/Thr protein phosphatase	
RBFOX1	RNA Binding Protein, Fox-1 Homolog 1	regulates tissue-specific alternative splicing	
SEPT5	Septin 5	regulation of cytoskeletal organization and SNARE complexes	cytoskeleton
TBL1XR1	Transducin Beta Like 1 X-Linked Receptor 1	F-box-like protein involved in the recruitment of the ubiquitin/19S proteasome complex to nuclear receptor-regulated transcription units.	transcription
TTC27	Tetratricopeptide Repeat Domain 27	unknown	
TUBA3C	Tubulin Alpha 3c	major component of microtubules	cytoskeleton
TUBGCP2	Tubulin Gamma Complex Associated Protein 2	microtubule nucleation at the centrosome	cytoskeleton
WASL	Wiskott-Aldrich Syndrome Like	Regulates actin polymerization	cytoskeleton
WWOX	WW Domain Containing Oxidoreductase	WW Domain Containing Oxidoreductase	