Word count: 1422

ADHD: progressing from genetic discoveries to biological insights

Anita Thapar

Anita Thapar FRCPsych, PhD, FMedSci

MRC Centre for Neuropsychiatric Genetics and Genomics and Division of Psychological Medicine and Clinical Neurosciences, Cardiff University School of Medicine, Hadyn Ellis Building, Maindy Road, Cardiff CF24 4HQ, UK.

thapar@cf.ac.uk

Attention Deficit/Hyperactivity Disorder (ADHD) is one of the most heritable of all neuropsychiatric

disorders (1). Despite being so highly disruptive and impairing, relatively few effective treatments

are available and outcomes tend to be poor. Gaining insight into the pathophysiology that underlies

ADHD represents a crucial first step towards developing novel and effective modes of diagnosis and

treatment. However, the complex multi-factorial nature of psychiatric disorders coupled with

inaccessibility to live brain tissue poses a serious challenge to investigations of biology. Genetic

studies provide a useful strategy for gaining insight into biology and this understanding has provided

a strong impetus for conducting genome-wide discovery studies of many different psychiatric

disorders, including ADHD (2).

Recent genome-wide association studies have shown that ADHD is associated with multiple rare and

common genetic variants (1,3,4). Deleterious mutations contributing to large effect sizes tend to be

rare because they are more rapidly removed from the population by natural selection while

common variants by themselves have small effect size. Rare mutations thus are especially

interesting from the perspective of offering a window into biology. However, discoveries from

genome-wide discovery studies represent only the first step. That is because an association between

ADHD and a gene variant does not necessarily tell us which specific gene /genes are involved, or the

biological mechanisms that might be involved (2). Typically, genome-wide association studies only

identify genomic regions in which potential risk genes lie. Thus, further work is needed.

The paper by Harich et al. (5) sets out to prioritise genes implicated by genome-wide association

studies of rare copy number variants (CNV) in ADHD. Copy number variants are one class of rare

gene variant that involve sub-microscopic alterations to chromosomal structure. Deletions and

duplications of DNA have been observed to be associated with ADHD, other neurodevelopmental

disorders such as ASD, Intellectual Disability and Tourette Syndrome as well as psychiatric disorders

such as schizophrenia and depression (6). Like all classes of genetic variation, CNVs show marked pleiotropy across different psychiatric disorders and there is also prominent variation in the severity and nature of clinical features manifest by carriers (7,8). For example, CNVs associated with ADHD overlap with those implicated in ASD and schizophrenia (9–11).

Although CNVs at different locations on the genome have been reported to be associated with ADHD, to date no study has been large enough to implicate individual CNVs at a genome-wide significant level with the exception of one pooled analysis of 15q 13.3 duplications (12). Another challenge is that CNVs typically encompass multiple genes. Thus, the first tranche of case-control CNV studies of ADHD have not led yet to the identification of risk genes. The paper by Harich et al. addresses this gap and moves the field of ADHD genetics forward. The authors identify "high-priority" genes by taking the CNVs reported in 11 published studies of ADHD and using a comprehensive set of bio-informatic approaches to analyse these. They capitalize on publicly available biological data resources and, also examine convergence of results across different methods. Triangulation of evidence across different designs and methods is increasingly recognised as crucial for epidemiological and etiological research (13) and here we see its application in genetics.

The authors initially identified 2241 potential genes within CNVs reported to be associated with ADHD from the 11 published studies. This is a prohibitively large number of genes if the aim is to investigate gene function in the laboratory. The aim of the authors was to select the most plausible genes from this very large list in different ways. They first included only genes that were completely deleted or partially truncated or were entirely duplicated. Some CNVs overlap so cover even larger stretches of DNA and the degree of overlap is variable in different individuals. The authors narrowed down the region of interest by selecting the minimal region common to all these overlaps. They then

2

selected messenger RNA coding genes that were in at least 2 cases and excluded those observed in controls. This created a list of 432 high-ranking genes.

Next the authors used an atlas known as Brain Span to test the biological plausibility of these high-ranking genes. Genes contain the instructions for manufacturing proteins, but this process first requires DNA to be transcribed to (expressed as) messenger RNA. This atlas maps the RNA profile of human brain across development including during the prenatal period (14). The authors found significant evidence of enrichment for co-expressed genes among the high-ranking genes; that is, the high-ranking genes appeared to be expressed together in the brain at the same time. Further analysis of these genes using a protein-protein interaction database identified several protein networks that involved 62 proteins.

One interesting aspect of this paper is that the authors further conducted a set of alternative cross-species analyses to test plausibility of the 432 high-ranking genes. Using cross-species data, they observed that 18 of the 432 priority genes were associated with cross-species phenotypes that resemble features of ADHD (e.g. hyperactivity). Examination of a cross-species biological database of interactions enabled the authors to identify an additional 48 genes that interact with these 18 genes providing a total of 66 priority genes implicated by cross-species data.

The authors finally tested for convergence between the human and cross-species approaches and identified a final list of 26 "highest confidence" ADHD genes that were observed across all these approaches (human and cross-species). These highest confidence genes also showed significant evidence of being co-expressed in brain, in the prenatal period and later in development, and over half of the genes showed broad expression patterns across different tissue types (i.e. not just brain).

2

They also highlight that the higher confidence genes link to several inter-connected protein networks that cross human and other species.

Rare CNVS are not the only class of genetic variation associated with ADHD risk. Rare sequence variants and common gene variants have also been shown to be associated with ADHD (1). The most recent and largest published GWAS identified 12 genome-wide significant loci (15). Given the polygenic nature of ADHD, a key question was whether any of the highest confidence CNV-genes identified by Harich et al. overlap with genes implicated by common gene variants. The authors examine this and identify two of their priority genes, POLR3C and RBFOX1, that map within genomic regions showing common variant associations.

This paper is an important as it moves the field forward, one step closer to biological insights from ADHD genetic discovery. The authors conduct an elegant and comprehensive set of bioinformatic and statistical analyses using published CNV data to infer genes that are most plausibly involved in ADHD pathogenesis. The next steps will be to conduct laboratory -based animal and cellular studies of the prioritised genes to fully understand the molecular and developmental mechanisms involved. Laboratory work examining the function of genes is expensive and time consuming and thus prioritising genes in this way is invaluable.

As the authors point out there are limitations. First, studies of ADHD rare mutations trail behind other neurodevelopmental disorders (e.g. ASD, intellectual disability (ID)) and many psychiatric disorders (e.g. schizophrenia), despite being so heritable. Discovery sample sizes remain small when very large samples are needed to identify genome-wide significant loci (because of allowing for multiple testing). The CNVs examined in this paper are not genome-wide significant. Also, the

authors had to rely on case-control CNV data which means the CNVs will include ones that are inherited from the affected offspring's parents and those that arise *de novo*. Rare *de novo* mutations in offspring using parent-offspring trios are especially powerful because these mutations are more likely pathogenic and thus can provide important biological insights. Yet, unlike for ASD, ID and schizophrenia there have been very few such studies for ADHD. Another challenge faced by the authors is that the original 11 published CNV studies all used different approaches in defining their CNVs. Furthermore, CNVs do not work in isolation. ADHD is multi-factorial in nature; multiple genetic variants as well as environmental risk factors also contribute. Finally, while the approach used by the authors is important and the highlighted genes help prioritize ones of high interest, the study cannot confirm that these genes are causally involved in ADHD. This means that much more work including, further laboratory work will be needed before findings result in clinical translation.

Some may question, how soon will this type of genetics research help clinicians and patients? The answer is that while there have been rapid and enormous advances with genetic discovery, initial genetic association findings do not tell us about what specific genes and mechanisms are involved so the approaches used by Harich et al represent an important step in understanding biology. Continued gene variant discovery in larger patient samples, further bio-informatic analyses laboratory- based and clinical studies will all be needed not only to uncover pathophysiology and test interventions but also to move towards evaluating potential clinical applications.

1.   Thapar A. Discoveries on the Genetics of ADHD in the 21st Century: New Findings and Their Implications. Am J Psychiatry. 2018;175(10):943–50.

2.   Sullivan PF, Geschwind DH. Defining the Genetic, Genomic, Cellular, and Diagnostic Architectures of Psychiatric Disorders. Cell. 2019;177(1):162–83.

3.   Demontis D, Walters RK, Martin J, Mattheisen M, Als TD, Agerbo E, et al. Discovery Of The First Genome-Wide Significant Risk Loci For ADHD. Nat Genet. 2018 Jun 3;IN PRESS.

4.   Satterstrom FK, Walters RK, Singh T, Wigdor EM, Lescai F, Demontis D, et al. Autism spectrum disorder and attention deficit hyperactivity disorder have a similar burden of rare protein-truncating variants. Nat Neurosci. 2019 Dec 25;22(12):1961–5.

5.   Harich, B; van der Voet, M; Klein, M; Sizek, P; Fenckova, M; Shenck, A; Franke B. From rare Copy Number Variations to biological processes in ADHD. Am J Psychiatry. 2020;

6.   Sullivan PF, Agrawal A, Bulik CM, Andreassen OA, Børglum AD, Breen G, et al. Psychiatric genomics: An update and an Agenda. Am J Psychiatry. 2018 Jan 1;175(1):15–27.

7.   Rees E, Owen MJ. Translating insights from neuropsychiatric genetics and genomics for precision psychiatry. Genome Med. 2020 Apr 29;12(1):43.

8.   Cross-Disorder Group of the Psychiatric Genomics Consortium. Electronic address: plee0@mgh.harvard.edu, Cross-Disorder Group of the Psychiatric Genomics Consortium. Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. Cell. 2019;179(7):1469-1482.e11.

9.   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. 2011 Aug 10;3(95):95ra75.

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. 2010 Oct 23;376(9750):1401–8.

11. Gudmundsson OO, Walters GB, Ingason A, Johansson S, Zayats T, Athanasiu L, et al. Attention-deficit hyperactivity disorder shares copy number variant risk with schizophrenia and autism spectrum disorder. Transl Psychiatry. 2019;9.

12. Williams NM, Franke B, Mick E, Anney RJL, Freitag CM, Gill M, et al. 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 Mar;169(2):195–204.

13. Lawlor DA, Tilling K, Davey Smith G. Triangulation in aetiological epidemiology. Int J Epidemiol. 2017 Jan 20;45(6):dyw314.

14. Miller JA, Ding S-L, Sunkin SM, Smith KA, Ng L, Szafer A, et al. Transcriptional landscape of the prenatal human brain. Nature. 2014 Apr 2;508(7495):199–206.

15. 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. 2019 Jan 26;51(1):63–75.