1. Collaborative Award in Science application

Reference number	214322/Z/18/Z
Applicant name	Prof Simon Baron-Cohen
Title of application	Common Variant Genetics of Autism and Autistic Traits (GWAS) Consortium
Total amount requested	£3,283,002.00

2. Application summary

Application title
Common Variant Genetics of Autism and Autistic Traits (GWAS) Consortium

Proposed duration of funding (months)
60

Proposed start date	15/01/2019

Name of administering organisation
University of Cambridge

Lead applicant's address at administering organisation	
Department/Division	Autism Research Centre, Department of Psychiatry
Organisation	University of Cambridge
Street	18B Trumpington Road
City/Town	Cambridge
Postcode/Zipcode	CB2 8AH
Country	United Kingdom

Research funding area Please select from the drop-down list the funding area that you consider your research falls under
Genetics, Genomics and Population Research

3. Lead applicant

Lead applicant details	
Full Name	Prof Simon Baron-Cohen

Department	Department of Psychiatry
Division	
Organisation	University of Cambridge
Address Line 1	Autism Research Centre, Section of Developmental Psychiatry
City/Town	CAMBRIDGE
Postcode	CB2 8AH
Country	United Kingdom
Telephone No.	[redacted]
Email Address	[redacted]

ORCID iD	
ORCID iD	0000-0001-9217-2544

Career history (current/most recent first)

From	To	Position	Organisation
01/2001	01/2025	Professor in Developmental Psychopathology and Director of the Autism Research Centre	University of Cambridge
01/1999	01/2001	Reader in Developmental Psychopathology	University of Cambridge
01/1994	01/1999	Lecturer in Developmental Psychopathology	University of Cambridge
01/1991	01/1994	Senior Lecturer in Developmental Psychopathology	Institute of Psychiatry, London
01/1988	01/1991	New Blood Lecturer	Institute of Psychiatry, London
01/1987	01/1988	Lecturer	University College London

Education/training

From	To	Qualification	Subject	Organisation
09/1985	09/1987	Master of Philosophy (MPhil)	Clinical Psychology	Institute of Psychiatry, London
09/1982	09/1985	Doctor of Philosophy (PhD;DPhil)	Psychology	University College London
09/1978	07/1981	Bachelor of Arts (BA)	Human Sciences	University of Oxford

Source(s) of personal salary support
University of Cambridge

Clinical status Do you have a medical/veterinary degree?	No

Career breaks	No

Have you had any career breaks or periods of part-time work, for example parental or long-term sick leave?	

Do you wish to undertake this award part time?	No

Career contributions
What are your most important research-related contributions to date? These may include contributions to health policy or practice, or to technology or product discovery and development.

(1) The mindblindness theory of autism (1985). To explain the social-communication deficits in autism, Baron-Cohen and colleagues formulated the ‘theory of mind’ (ToM) hypothesis and confirmed this using the False Belief test. He coined the term ‘mindblindness’ to describe autism. He showed that absence of the precursors to ToM predicts autism diagnosis as early as 18 months old, using the Checklist for Autism in Toddlers (CHAT). He outlined a model of ToM that has its roots in infant ‘shared attention’. He conducted the first neuroimaging study of ToM and demonstrated lesions in prefrontal cortex and amygdala can impair ToM. He also conducted the first fMRI study to show atypical amygdala function in autism. Finally, he showed that children with autism can be taught to 'mindread' emotion recognition via specialist educational software.

(2) The prenatal sex steroid theoryof autism (1997). To understand why autism is more common in males, he tested if prenatal sex steroid hormones play a role in the aetiology of autism, since animal research shows that prenatal sex steroids play a causal role in the sexual differentiation of the brain. He instigated a longitudinal study demonstrating for the first time in humans how normative variation in amniotic prenatal testosterone (pT) levels correlate with postnatal brain and behavioural development. This entailed studying children of women who had undergone amniocentesis in pregnancy. To date, these typically developing children have been tested postnatally at 6 time points. He discovered that in typical toddlers, amount of eye contact and vocabulary development are inversely correlated with pT levels, and in school age children, quality of social relationships, ToM performance, and scores on the EQ are also inversely correlated with pT levels. In contrast, he found that scores on the Embedded Figures Test (of attention to detail), on the Systemizing Quotient (SQ), and measures of narrow interests are positively correlated with pT levels. He conducted the first human neuroimaging studies of the correlates of pT. In 2015 he collaborated with the Danish Biobank to show that people with autism have elevated levels of pT and the Δ4 sex steroid precursors to pT.

Experience relevant to this proposal.
Please summarise your key achievements and experience which are relevant to this proposal.

Simon Baron-Cohen is Professor of Developmental Psychopathology, University of Cambridge, Fellow at Trinity College, Cambridge and Director of the Autism Research Centre in Cambridge. He is author of Mindblindness, The Essential Difference, Prenatal Testosterone in Mind, and Zero Degrees of Empathy. He has edited scholarly anthologies including Understanding Other Minds. He has written books for parents and teachers including Autism and Asperger Syndrome: The Facts. He is author of the DVDs Mind Reading and The Transporters, to help children with autism learn emotion recognition, both nominated for BAFTA awards. He formulated the ‘mindblindness’ theory of autism (1985) and the ‘prenatal sex steroid’ theory of autism (1997). He has also made contributions to many fields of autism research, to typical cognitive sex differences, and synaesthesia research. He created the first UK clinic for adults with suspected Asperger Syndrome (1999) that has helped over 1,000 patients to have their disability recognized. He gave a keynote address to the United Nations in New York on Autism Awareness Day 2017 on the topic of Autism and Human Rights. He is a Fellow of the British Psychological Society, the British Academy, and the American Psychological Association. He is Vice-President of the National Autistic Society, and President of the International Society for Autism Research (INSAR). He was Chair of the NICE Guideline Development Group for Autism (Adults) and is Chair of the Psychology Section of the British Academy. He is co-editor in chief of the journal *Molecular Autism* and is a National Institute

of Health Research (NIHR) Senior Investigator. See www.autismresearchcentre.com. In the field of autism genetics, he has co-authored recent papers with Dr Varun Warrier using GWAS of relevant phenotypic traits such as the Reading the Mind in the Eyes (80,000 genotyped participants, via 23andMe) in Molecular Psychiatry and the Empathy Quotient (50,000 genotyped participants) in Translational Psychiatry. He also leads the Templeton World Charitable Foundation program on autism genetics studying rare variants in 20 highly multiplex families, which is moving into its final year of 3 years, which is a collaboration with Illumina Inc and is discovering novel rare genetic mutations in autism.

Research outputs
List up to 20 of your most significant research outputs, ensuring that at least five of these are from the last five years. For 10 of these outputs, provide a statement describing their significance and your contribution (up to 50 words per output).

Research outputs may include (but are not limited to):

- Peer-reviewed publications and preprints
- Datasets, software and research materials
- Inventions, patents and commercial activity

For original research publications indicate those arising from Wellcome-funded grants in **bold**, and provide the PubMed Central ID (PMCID) reference for each of these. Please refer to guidance notes.

Please give citation in full, including title of paper and all authors Citations to preprints should state "Preprint", the repository name and the articles persistent identifier (e.g DOI).*

*(*All authors, unless more than 10, in which case please use 'et al', ensuring that your position as author remains clear.)*

Top 20 Publications in chronological order, from Baron-Cohen:

1. Baron-Cohen, S, Leslie, A.M., & Frith, U, (1985) Does the autistic child have a "theory of mind?" Cognition, 21, 37-46. [SBC collected the data as part of his PhD. This paper stimulated hundreds of subsequent studies into theory of mind in autism].

2. Baron-Cohen, S, (1995) Mindblindness: an essay on autism and theory of mind. MIT Press/Bradford Books. [SBC wrote this as a monograph summarising his PhD and postdoc research. This was the first time the theory of mind literature in autism had been integrated].

3. Stone, V, Baron-Cohen, S, & Knight, R, (1998) Frontal lobe contributions to theory of mind. Journal of Cognitive Neuroscience, 10, 640-656. [SBC supervised the postdoc and co-designed the experiments. This paper identified both orbitofrontal cortex and amygdala lesions in theory of mind].

4. Baron-Cohen, S, Ring, H, Wheelwright, S, Bullmore, E, Brammer, M, Simmons, A, & Williams, S, (1999) Social intelligence in the normal and autistic brain: an fMRI study. European Journal of Neuroscience, 11, 1891-1898. [SBC designed the experiment and wrote the paper. This paper was the first to report atypical amygdala function in autism].

5. Baron-Cohen, S, Ring, H, Bullmore, E, Wheelwright, S, Ashwin, C, & Williams, S, (2000) The amygdala theory of autism. Neuroscience and Behavioural Reviews, 24, 355-364. [SBC was lead author and proposed the theory. This was the first paper to highlight the importance of the

amygdala in autism].

6. Baron-Cohen, S, & Wheelwright, S, Skinner, R, Martin, J, & Clubley, E, (2001) The Autism-Spectrum Quotient: Evidence from Asperger Syndrome/high-functioning autism, males and females, scientists, and mathematicians. Journal of Autism and Developmental Disorders, 31, 5-17. [SBC co-designed the experiment and wrote the paper. This was the first paper to provide evidence for the association between mathematical talent and autistic traits].

7. Baron-Cohen, S, Wheelwright, S, & Hill, J, (2001) The 'Reading the Mind in the Eyes' Test Revised Version: A study with normal adults, and adults with Asperger Syndrome or High-Functioning Autism. Journal of Child Psychiatry and Psychiatry,42, 241-252. [SBC designed the measure and wrote the paper. The measure has now been used in hundreds of other studies].

8. Baron-Cohen, S, (2002) The extreme male brain theory of autism. Trends in Cognitive Sciences, 6, 248-254. [SBC wrote the paper and proposed the theory. This paper stimulated 15 years of research into prenatal sex steroid hormones in autism].

9. Nunn, J, Gregory, L, Morris, R, Brammer, M, Bullmore, E, Harrison, J, Williams, S, Baron-Cohen, S, and Gray, J, (2002) Functional magnetic resonance imaging of synaesthesia: activation of colour vision area V4/V8 by spoken words. Nature Neuroscience, 5, 371-375. [SBC co-designed the study. This was the first demonstration using fMRI of the brain basis of synaesthesia].

10. Baron-Cohen, S, Knickmeyer, R, & Belmonte, M (2005) Sex differences in the brain: implications for explaining autism. Science, 310, 819-823. [SBC wrote the paper. This review paper pulled together all the research relevant to the link between autism and gender].

11. Chakrabarti, B, Hill-Cawthorne, G, Dudridge, F, Kent, L, Wheelwright, S, Allison, C, Banerjee-Basu, S, & Baron-Cohen, S, (2009) Genes related to sex-steroids, neural growth and social-emotional behaviour are associated with autistic traits, empathy and Asperger Syndrome. Autism Research, 2,157-177. [SBC co-designed the study. This was the first candidate gene study of Asperger Syndrome].

12. Baron-Cohen, S, Lombardo, M, Auyeung, B, Ashwin, E, Chakrabarti, B, & Knickmeyer, R, (2011) Why are Autism Spectrum Conditions more prevalent in males? PLOS Biology, 9, 1-10. PMC3114757. [SBC co-authored the paper. This was a major overview of why being male is a risk factor for autism].

13. van Honk, J, Schuttera, D, Bosa, P, Kruijtc, A, Lentjes, E, & Baron-Cohen, S, (2011) Testosterone administration impairs cognitive empathy in women depending on second-to-fourth digit ratio. Proceedings of the National Academy of Sciences of the USA, 108, 3448-52. PMC3044405 [SBC contributed to the design of the study. This was the first study to show that manipulation of testosterone changes social cognition].

14. Schwarz, E, Guest P, Rahmoune, H, Wang, L, Levin, Y, Ingudomnukul, E, Ruta, L, Kent, L, Spain, M, Baron-Cohen, S, & Bahn, S, (2011) Sex-specific serum biomarker patterns in adults with Asperger's Syndrome. Molecular Psychiatry,16, 1213-20. [SBC supervised the PhD student who collected the samples. This was the first biomarker study to show distinct profiles in autism as a function of sex].

15. Lombardo, M, Ashwin, E, Auyeung, B, Chakrabarti, Taylor, K, Hackett, G, Bullmore, E, & Baron-Cohen, S, (2012) Fetal testosterone influences sexually dimorphic gray matter in the human brain. Journal of Neuroscience, 32, 674-80. PMC3306238. [SBC co-authored the paper and codesigned the experiment. This Wellcome Trust funded study was the first human demonstration

that fetal testosterone changes brain structure developmentally].

16. Lai, M-C, Lombardo, M, & Baron-Cohen, S, (2013) Autism. The Lancet,383(9920), 896-910. [SBC co-authored the paper. This was a major update on what we know about autism].

17. Lai, M-C, Lombardo, M, Suckling, J, Ruigrok, A, Chakrabarti, B, Ecker, C, Deoni, S, Craig, M, Murphy, D, Bullmore, E, MRC AIMS Consortium, & Baron-Cohen, S, (2013) Biological sex affects the neurobiology of autism. Brain, 136, 2799-2815. PMC3754459 [SBC co-supervised the PhD student, co-designed the study and co-authored the paper. This Wellcome Trust funded study showed how autistic women are masculinised in brain structure (both grey and white matter].

18. Baron-Cohen, S, Auyeung, B, Nørgaard-Pedersen, B, Hougaard, D.M, Abdallah, M.W, Melgaard, L, Cohen, A.S, Chakrabarti, B, Ruta, L, Lombardo, M.V, (2014) Elevated fetal steroidogenic activity in autism. Molecular Psychiatry,1-8. PMC4184868 [SBC provided funding for the study (MRC program grant and Wellcome Trust project grant) and co-designed the experiment and coauthored the paper. It was the first demonstration of prenatal sex steroids being elevated in children who later show autism].

19. Cassidy, S, Bradley, P, Robinson, J, Allison, C, McHugh, M, & Baron-Cohen, S, (2014) Suicidal ideation and suicide plans or attempts in adults with Asperger's syndrome attending a specialist diagnostic clinic: a clinical cohort study. The Lancet Psychiatry, 1, 142–147. [SBC provided access to the data in the NHS, and conceived of the study. This was the first major clinical study showing autistic adults have elevated rates of suicidal plans and attempts].

20. Warrier, V, Grasby, KL, Uzefovsky, F, Toro, R, Smith, P, Chakrabarti, B, Khadake, J, Mawbey-Adamson, E, Litteman, N, Hottenga, J-J, Lubke, G, Boomsma, DI, Martin, NG, Hatemi, PK, Medland, SE, Hinds, DA, Bourgeron, T, & Baron-Cohen, S, (2017) Genome-wide meta-analysis of cognitive empathy: heritability, and correlates with sex, neuropsychiatric conditions and cognition. Molecular Psychiatry, (DOI: 10.1038/mp.2017.122). PMC5656177. [SBC initiated the collaboration with 23andMe, suggested the measure, and co-wrote the paper. This was the first big data test (80,000 volunteers) for genes associated with cognitive empathy].

Total number of peer-reviewed publications which you have authored/co-authored. Please exclude abstracts and literature reviews.	513

Current and recent research funding (including Wellcome Trust grants)
Please list all held in the last five years and any key prior grants (list the most recent first). State the name of the awarding body, name(s) of grantholder(s), title of project, amounts awarded, your role in the project, and start and end dates of support. For all active grants, indicate the number of hours per week that are spent on each project.

1. Action autism innovative medicine studies – 2 – Trials (AIMS-2-TRIAL) (funded by Innovative Medicines Initiative)
Professor Simon Baron-Cohen (Principal Investigator - 0.1fte)
Awarded €2,808,025
Dates of the grant – 01.05.2018 – 30.04.2023

2. Investigating the role of NRXN1 in autism (funded by the Autism Research Trust)
Professor Simon Baron-Cohen (Principal Investigator - 0.1fte)
Awarded £475,000
Dates of the grant - 01.01.2018 to 31.08.2020

3. Autism and the criminal justice system (funded by the Autism Research Trust)
Professor Simon Baron-Cohen (Principal Investigator - 0.1fte)

Awarded £164,896
Dates of the grant - 01.01.2018 to 31.12.2020

4. Hormones and Biomarkers in Pregnancy (funded by the Autism Research Trust)

Professor Simon Baron-Cohen (Principal Investigator - 0.1fte)
Awarded £40,000
Dates of the grant - 01.01.2017 to 30.09.2020

5. Vulnerability in adults with autism spectrum conditions (funded by the Autism Research Trust and Autistica)
Professor Simon Baron-Cohen (Principal Investigator - 0.1fte)
Awarded £80,000
Dates of the grant - 01.10.2016 to 31.05.2018

6. NIHR Senior investigator award (funded by NIHR)
Professor Simon Baron-Cohen (Award holder)
Awarded £45,000
Dates of the grant - 01.04.2016 to 31.03.2019

7. Suicide, autism and autistic traits (funded by Collaboration for Leadership in Applied Health Research & Care (CLAHRC))
Professor Simon Baron-Cohen (Principal Investigator)
Awarded £44,688
Dates of the grant - 01.04.2016 to 30.06.2017

8. Investigating the link between mathematical ability and autism using genetics and epigenetics (IMAGE) (funded by the Templeton World Charity Foundation, Inc.)
Professor Simon Baron-Cohen (Principal Investigator – 0.1 fte)
Awarded £1,805,657
Dates of the grant – 15.12.2015 to 14.12.2018

9. Oxytocin inhalation (funded by the Autism Research Trust)
Professor Simon Baron-Cohen (Principal Investigator – 0.1 fte)
Awarded £81,880
Dates of the grant – 23.01.2014 to 31.12.2018

10. Hormones and biomarkers (funded by the Autism Research Trust)
Professor Simon Baron-Cohen (Principal Investigator)
Awarded £32,000
Dates of the grant – 23.01.2014 to 31.12.2017

11. Autism gene sequencing (funded by the Autism Research Trust)
Professor Simon Baron-Cohen (Principal Investigator - 0.1fte)

Awarded £112,000
Dates of the grant – 06.07.2013 to 14.12.2018

12. European Autism Interventions (EU-AIMS) (funded by Innovative Medicines Initiative)
Professor Simon Baron-Cohen (Principal Investigator)

Awarded £528,572
Dates of the grant – 01.04.2012 to 31.03.2018

13. Foetal testosterone effects on brain structure and function (funded by the Wellcome Trust)
Professor Simon Baron-Cohen (Principal Investigator)
Awarded £289,602
Dates of the grant – 01.10.2010 to 30.09.2014

14. Autistic traits, autism spectrum conditions, and foetal testosterone (funded by the Medical Research
Council)
Professor Simon Baron-Cohen (Principal Investigator)
Awarded £943,891
Dates of the grant - 01.01.2007 - 13.05.2012

Please describe how the currently active grants listed above relate to this application
Some of the above grants will provide for sample collection, DNA extraction, and genotyping of autistic individuals included in the GWAS meta-analysis. These are outline below: 1. *The AIMS-2-TRIAL* will provide some of the samples (N = 700, cases and controls) included in the autism GWAS meta-analysis. 2. *The IMAGE* study funded by the Templeton World Charity Foundation, Inc. has provided funds for collecting samples and extracting DNA from 1000 autistic individuals. This study has also allowed us to develop methods to collect saliva samples through the postal system and the development of an online phenotyping and registration website. In addition, the *vulnerability in adults with autism spectrum conditions* study has provided funds to investigate vulnerabilities in autistic individuals. We will use the phenotypic measures developed by this study to investigate vulnerabilites in autistic individuals in the UK Autism Biobank.

Time spent on research How many hours per week do you spend on research?
40

How many hours per week will be spent on this project?
4

4. Applicants

1

Applicant	
Full Name	Prof Daniel Geschwind
Department	
Division	
Organisation	University of California, Los Angeles
Address Line 1	

City/Town	
Postcode	
Country	
Telephone No.	
Email Address	[redacted]

Career history (current/most recent first)			
From	**To**	**Position**	**Organisation**
05/2016	06/2025	Director and Sr Assoc Dean and Assoc VC	UCLA School of Medicine
01/2005	06/2025	Professor	UCLA School of Medicine
01/2005	06/2025	Gordon and Virginia MacDonald Distinguished Chair	UCLA School of Medicine
05/2003	06/2025	Director	UCLA School of Medicine
09/1997	06/2025	Co-Director	UCLA School of Medicine

Education/training				
From	**To**	**Qualification**	**Subject**	**Organisation**
09/1984	05/1991	Doctor of Medicine (MD)	School of Medicine	Yale University
09/1984	05/1991	Doctor of Philosophy (PhD;DPhil)	PhD Neurobiology	Yale University
08/1979	06/1982	Bachelor of Arts (BA)	Chemistry and Psychology	Dartmouth College

Source(s) of personal salary support
UCLA

Career contributions
What are your most important research-related contributions to date? These may include contributions to health policy or practice, or to technology or product discovery and development.

I began large-scale collaborative studies of ASD genetics and, in collaboration with the Cure Autism Now foundation, developed AGRE, a public family resource for the study of autism genetics. My group played a leading role in the first major whole exome sequencing studies performed in ASD, applied transcriptional profiling to determine the functional impact of genome-wide structural chromosomal variation in ASD, and recently, contrasted the contribution of de novo and inherited CNV to ASD in multiplex and simplex families.

In 2001, we hypothesized a portion of risk for ASD was on a continuum with common genetic risk for quantitative component phenotypes in the general population, and there are intermediates between genes and disease, endophenotypes. We tested this model, identifying loci for quantitative impairment in social responsiveness/communication, non-verbal communication, and language, and loci related to male/female genetic risk and quantitative functional imaging endophenotypes, to

connect genetic risk to brain circuits.

My laboratory works in neurogenetics through genome wide transcriptomics in normal brain and development of systems biology methods for analysis and data integration. We performed the first genome-wide analysis of human cortical patterning and demonstrated brain transcriptome has a reproducible structure that can be used to identify drivers of human brain evolution and disease. We showed how network frameworks facilitate hypothesis generation and permit evolving beyond network predictions to experimental validation, and extended our analysis of transcriptomic networks to epigenetic mechanisms of gene regulation via integrative analysis of 3-dimensional chromatin structure and enhancer marks. This work identified novel mechanisms underlying neuropsychiatric disorders such as schizophrenia and ASD.

My laboratory identified the first genome wide evidence for convergent biology in ASD via transcriptomic studies of both lymphoblasts and post mortem brain, validated this in independent samples using RNA sequencing and used co-expression network analysis to demonstrate circuit and pathway convergence in ASD. We characterized the human specific transcriptional network downstream of the neurodevelopmental gene, FoxP2, which causes a speech/language disorder.

Finally, my laboratory has led neurobiological studies of ASD-related common and rare genetic variants in humans and model systems to elucidate their mechanisms, providing proof of principles for the field.

Experience relevant to this proposal
Please summarise your key achievements and experience which are relevant to this proposal.

Following medical school, training as a PhD in neurobiology and neuroanatomy, post-graduate training in genetics, and clinical neurology residency, I founded the neurogenetics program at UCLA in 1997, and soon after, the Center for Autism Research and Treatment (CART). Since then, my laboratory has focused on integrating genetics and genomics with basic neurobiology to develop a more systematic understanding of neurodevelopmental and neurodegenerative conditions. The over-arching goal is to develop new therapeutics for nervous system disorders in which disease-altering therapies are not currently available. Our major focus has been advancing our understanding of autism spectrum disorder (ASD) and developing shared scientific resources for ASD research, including inclusion of under-represented groups in our genetics research. I played a leading scientific role in the development of the Autism Genetic Resource Exchange (www.AGRE.org) and have led several large-scale collaborative research efforts in ASD genetics. Another key aspect of my efforts has been the application of functional genomics to complex CNS disease, as well as the development and application of new analytic methods that elucidate the underlying network organization in multi-dimensional data, permitting integration of genomic and genetic data with phenotype data on a large scale. Overall I have more than 370 publications and am an elected member of the Institute of Medicine, which recognizes contributions to genetics and genomics of neurological disorders.

I am the Gordon and Virginia MacDonald Distinguished Professor of Human Genetics, Neurology and Psychiatry at UCLA. In the capacity as Senior Associate Dean and Associate Vice Chancellor of Precision Health, I lead the Institute for Precision Health at UCLA and have put considerable effort into fostering large-scale collaborative patient resources for genetic research and data sharing, beginning with AGRE and continuing with work in the form of an Autism Center of Excellence Network Award from NIH, now in its 11th year, focusing on genetic studies in underrepresented populations. The importance of large scale, collaborative, patient oriented research motivates my leadership of the Institute for Precision Health at UCLA, which will serve as a hub for mobilizing patient data for research, biobanking, and development and integration of new

digital health initiatives.

Research outputs
List up to 20 of your most significant research outputs, ensuring that at least five of these are from the last five years. For 10 of these outputs, provide a statement describing their significance and your contribution (up to 50 words per output).

Research outputs may include (but are not limited to):

- Peer-reviewed publications and preprints
- Datasets, software and research materials
- Inventions, patents and commercial activity

For original research publications indicate those arising from Trust-funded grants in **bold**, and provide the PubMed Central ID (PMCID) reference for each of these. Please refer to guidance notes.

Please give citation in full, including title of paper and all authors Citations to preprints should state "Preprint", the repository name and the articles persistent identifier (e.g DOI).*

*(*All authors, unless more than 10, in which case please use 'et al', ensuring that your position as author remains clear.)*

1. Yonan AL, Alarcón M, Cheng R, Magnusson PK, Spence SJ, Palmer AA, Grunn A, Juo SH, Terwilliger JD, Liu J, Cantor RM, **Geschwind DH***, Gilliam TC. (2003). A genomewide screen of 345 families for autism-susceptibility loci. *Am J Hum Genet.* 73: 886-97. (*Corresponding author.) No PMCID.
2. Sanders SJ, Murtha MT, Gupta AR, Murdoch JD, Raubeson MJ, Willsey AJ, Ercan-Sencicek AG, DiLullo NM, Parikshak NN, Stein JL, Walker MF, Ober GT, Teran NA, Song Y, El-Fishawy P, Murtha RC, Choi M, Overton JD, Bjornson RD, Carriero NJ, Meyer KA, Bilguvar K, Mane SM, Sestan N, Lifton RP, Günel M, Roeder K, **Geschwind** DH*, Devlin B*, State MW* (2012). De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature*. 485(7397): 237-41. (*senior authors, co-supervised this study). PMC3667984.
3. Luo R, Sanders SJ, Tian Y, Voineagu I, Huang N, Chu SH, Klei L, Cai C, Ou J, Lowe JK, Hurles ME, Devlin B, State MW, **Geschwind DH** (2012). Genome-wide transcriptome profiling reveals the functional impact of rare de novo and recurrent CNVs in Autism Spectrum Disorder. *Am J Hum Genet*. 91(1): 38-55. PMC3397271.
4. Leppa VM, Kravitz SN, Martin CL, Andrieux J, Le Caignec C, Martin-Coignard D, DyBuncio C, Sanders SJ, Lowe JK, Cantor RM, **Geschwind DH** (2016). Rare inherited and de novo CNVs reveal complex contributions to ASD risk in multiplex families. *Am J Hum Genet*. 99(3): 540-5 PMC5011063.
5. Alarcón M, Stone JL, Duvall JA, Abrahams BS, Sebat J, Wigler M, Nelson SF, Cantor RM, **Geschwind DH.** (2008). Linkage, association, and gene-expression analyses identify CNTNAP2 as an autism-susceptibility gene. *Am J Hum Genet*. 82(1): 150-9. PMC225395
6. Scott-Van Zeeland AA, Abrahams BS, Alvarez-Retuerto AI, Sonnenblick LI, Rudie JD, Ghahremani D, Mumford J, Poldrack RA, Dapretto M, **Geschwind DH***, Bookheimer SY* (2010). Altered functional connectivity in frontal lobe circuits is associated with variation in the autism risk gene CNTNAP2. *Sci Transl Med.* 2(56): 56ra80. (*Senior authors, co-supervised this study). PMC3065863.
7. Lowe JK, Werling DM, Constantino JN, Cantor RM, **Geschwind DH** (2015). Social responsiveness, an autism endophenotype: genomewide significant linkage to two regions on chromosome 8. *Am J Psychiatry*. 172(3): 266-75. PMC4523091.
8. Oldham MC, Konopka G, Iwamoto K, Langfelder P, Kato T, Horvath S, **Geschwind DH.** (2008). Functional organization of the transcriptome in human brain. *Nat Neurosci*. 11(11): 1271-82. PMC2756411.

9. Sun W, Poschmann J, Cruz-Herrera del Rosario R, Parikshak NN, Hajan HS, Kumar V, Ramaswamy R, Belgard TG, Elanggovan B, Wong CCY, Mill J, **Geschwind DH**, Prabhakar S (2016). Histone acetylome-wide association study of autism spectrum disorder. *Cell*. 167(5): 1385-97. PMCID in process.
10. Won H, de la Torre-Ubieta L, Stein JL, Parikshak NN, Huang J, Opland CK, Gandal M, Sutton G, Hormozdiari F, Lu D, Lee CH, Eskin E, Voineagu I, Ernst J, **Geschwind DH.** (2016). Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature.* 538(7626): 523-7. PMCID in process.
11. Konopka G, Bomar JM, Winden K, Coppola G, Jonsson ZO, Gao F, Peng S, Preuss TM, Wohlschlegel JA, **Geschwind DH.** (2009). Human-specific transcriptional regulation of CNS development by FOXP2. *Nature*. 462(7270): 213-7. *One of the top ten breakthroughs in mental illness related research by NARSAD investigators of 2009, Highlighted in* Nature News and Views*: Dominguez MH and Rakic P (Nov 12, 2009) 462:169-70.* PMC2778075.
12. Voineagu I, Wang X, Johnston, P, Lowe J, Tian, Y, Horvath S, Mill J, Cantor R, Blencowe BJ, **Geschwind DH.** (2011). Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature.* 474(7351): 380-4. (*Highlighted in *Nature Review Neuroscience.* Whalley, K. (June 8) 12(7): 372. ***Autism Speaks: Top Ten Science, Autism Research Achievements of 2011.*) PMC3607626.
13. Parikshak NN, Luo, R, Zhang A, Won H, Lowe JK, Chandran V, Horvath SH, **Geschwind DH.** (2013). Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. 155(5): 1008-21. PMC3934107.
14. Parikshak NN, Swarup V, Belgard TG, Irimia M, Ramaswami G, Gandal MJ, Hartl C, Leppa V, de la Torre Ubieta L, Huang J, Lowe JK, Blencowe BJ, Horvath S, **Geschwind DH** (2016). Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*. 540(7633): 423-7. PMCID in process.
15. Konopka G, Wexler E, Rosen E, Mukamel Z, Osborn GE, Chen L, Lu D, Gao F, Gao K, Lowe JK, Geschwind DH (2012). Modeling the functional genomics of autism using human neurons. Mol Psychiatry. 17(2): 202-14. PMC3170664.
16. Peñagarikano O, Abrahams BS, Herman E, Dong H, Almajano J, Bragin A, Peles E, **Geschwind DH.** (2011). Absence of CNTNAP2 leads to epilepsy, neuronal migration abnormalities, and core autism-related deficits. *Cell*. 147 (1): 235-46. *(*Selected research highlight in* Disease Models and Mechanisms*. *Autism Speaks: Top Ten Science, Autism Research Achievements of 2011)* PMC3390029.
17. Peñagarikano O, Lázaro MT, Lu XH, Gordon A, Dong H, Lam HA, Peles E, Maidment NT, Murphy NP, Yang XW, Golshani P, **Geschwind DH.** (2015). Exogenous and evoked oxytocin restores social behavior in the Cntnap2 mouse model of autism. *Sci Transl Med*. 7 (271): 271-8. PMC4498445.
18. Short PJ, McRae JF, Gallone G, Sifrim A, Won H, **Geschwind DH**, Wright CF, Firth HV, FitzPatrick DR, Barrett JC, Hurles ME. (2018). De novo mutations in regulatory elements in neurodevelopmental disorders. Nature. 555(7698): 611-16.
19. de la Torre Ubieta L, Stein JL, Won H, Opland CK, Liang D, Lu D, **Geschwind DH** (2018). The dynamic landscape of open chromatin during human corical neurogenesis. Cell. 172: 1-16.
20. Gandal MJ, Haney JR, Parikshak NN, Leppa V, Ramaswami G, Hartl C, Schork AJ, Appadurai V, Buil A, Werge TM, Liu C, White KP, CommonMind Consorium, PsychENCODE Consoritum, iPasych-BROAD Working Group, Horvath S, **Geschwind DH**. (2018). Shared molecular neuropathology across major psychiatric disorders parallels polygenic overlap. Science. 359(6376) 693-697.

Total number of peer-reviewed publications which you have authored/co-authored. Please exclude abstracts and literature reviews.	399

Current and recent research funding (including Wellcome Trust grants)
Please list all held in the last five years and any key prior grants (list the most recent first). State the

name of the awarding body, name(s) of grantholder(s), title of project, amounts awarded, your role in the project, and start and end dates of support. For all active grants, indicate the number of hours per week that are spent on each project.
Neural circuitry linking oxytocin deficiency and social impairment in autism. SFARI Pilot Award (2016). Role of CASPR2(CNTANP2) in brain circuits. SFARI Research Award (2012). Autism Genetics, Phase II: Increasing Representation of Human Diversity. NIH (2013 - 2018). Defining transcriptional networks in maternal immune activation models and schizophrenia. NIH (2018 - 2019). Building integrative CNS networks for genomic analysis of autism. NIH (2016 - 2021). Integrative Genomic Analysis of Human Brain Development and Autism. NIH (2016 - 2020) Genomic Strategies to Identify High-impact Psychiatric Risk Variants. NIH (2014 - 2018) Defining cell types, lineage, and connectivity in developing human fetal cortex. NIH (2014 - 2017) Genetic and genomic analyses to connect genes to brain to cognition in ASD. NIH (2014 - 2016)

Please describe how the currently active grants listed above relate to this application
The above grants will contribute to the GWAS of autism, and elucidate the functional impact of the GWAS loci. Specifically, the '*Autism Genetics, Phase II: Increasing Representation of Human Diversity*' study has sequenced the genomes of more than 2000 individuals from autism families (see: https://doi.org/10.1101/338855). This includes more than 500 autism probands. Genotype data will be available for this which can be included in the autism meta-analysis. This dataset will also allow for investigating the combined contribution of common and rare variants to autism. Additionally, several of the above projects aim to understand the functional consequences of these loci using data from single cell sequencing of specific cell types in the developing brain, Hi-C data to annotate non-genic regions, and transcriptome sequencing of the autism post-mortem brain to identify enrichment in gene networks.

Time spent on research How many hours per week do you spend on research?
40

How many hours per week will be spent on this project?
4

2

Applicant	
Full Name	Dr Matthew Hurles
Department	
Division	

Organisation	Wellcome Trust Sanger Institute
Address Line 1	Wellcome Trust Genome Campus
City/Town	CAMBRIDGE
Postcode	CB10 1SA
Country	United Kingdom
Telephone No.	
Email Address	[redacted]

Career history (current/most recent first)

From	To	Position	Organisation
09/2003	03/2023	Senior Group Leader	Wellcome Trust Sanger Institute

Education/training

From	To	Qualification	Subject	Organisation
09/1996	10/1999	Doctor of Philosophy (PhD;DPhil)	Genetics	University of Leicester

Source(s) of personal salary support

Wellcome Trust Sanger Institute

Career contributions

What are your most important research-related contributions to date? These may include contributions to health policy or practice, or to technology or product discovery and development.

I am dedicated to applying new genetic technologies to improve the diagnosis of patients with rare genetic conditions. I lead the Deciphering Developmental Disorders (DDD) Study, a collaboration between 14,000 families with children with severe, undiagnosed developmental disorders, all 24 clinical genetic centres in the UK and Ireland, and the Wellcome Sanger Institute. Together we are understanding the diverse genetic landscape of these disorders, and applying this knowledge to achieve improved diagnostic testing. As a part of the DDD study, my lab has identified several novel genes associated with DDD, quantified recessive causes of DDD, demonstrated a role for de novo mutations in regulatory elements, and quantified the relative contribution of mutations at splice sites for DDD.

I also lead the Prenatal Assessment of Genomes and Exomes (PAGE) Study, a collaboration between pregnant
mothers and their partners, a network of UK Fetal Medicine Centres caring for these pregnant women and the Wellcome Sanger Institute. Together we are investigating the genetic causes of developmental anomalies that are identified during prenatal ultrasound screening, with the aim of improving the prognostic information that can be provided to patients.

The work of my research group has been characterized by rapid adoption of new technologies for assaying genetic variation, development of novel analytical strategies for making sense from large datasets and thoughtful application of these tools for advancing our understanding of human genetic diseases. More recently, our highly
collaborative research has had increasing translational impact, resulting in genetic diagnoses for over a thousand children with previously undiagnosed developmental disorders, and leading to the

founding of a startup company (Congenica Ltd) to provide sustainable genetic diagnostic services to the NHS and other healthcare providers.

I believe we have a moral imperative to give patients and their families the opportunity to share their genetic data anonymously, to enable them to benefit from having the greatest possible number of researchers and clinicians analysing their data. Together with Helen Firth, I lead the DECIPHER initiative that is enabling rare disease patients to share anonymised genetic and clinical data globally.

Experience relevant to this proposal
Please summarise your key achievements and experience which are relevant to this proposal.

I am the head of the human genetics programme at the Wellcome Sanger Institute. My research has provided new insights into the contributions of structural variants including deletions and duplications to rare and common diseases, the factors infleuncing de novo mutation rates , and the delineation of the genetic causes of severe developmental disorders, including neurodevelopmental conditions and congenital heart defects.

In particular, I lead the Deciphering Developmental Disorders study, a UK-wide study on genetically undiagnosed developmental study. As a part of this we have brought together doctors in the 24 Regional Genetics Services, throughout the UK and Republic of Ireland to recruit participants into the study. Whole exome sequencing has been conducted in more than 13,600 probands from more than 13,500 families (total ~33,000 individuals including parents). Our research published in 2017 has identified 14 novel genes involved in developmental disorders providing genetic diagnosis for many participants in the DDD. We have since more than doubled the cohort size identifying dozens of novel genes. Our research has also demonstrated a role for de novo variants in regulatory regions and splice sites in developmental disorders. Tied to this, we have also developed a website (DECIPHER) that allows for sharing of information in individuals with rare disorders, accelerating gene discovery.

I'm also collaborating on the PAGE study (Prenatal Assessment of Genomes and Exomes), which conducts whole exome sequence of foetuses with developmental abnormalities apparent through fetal ultrasound in 1000 fetuses across the UK. This additional information will help to acquire new knowledge about the genetic variation causing the observed abnormalities. The new insights gained by this study will be used to improve diagnostic methods, allowing better genetics-derived prognoses and more informed parental counselling as well as future management of pregnancy and childbirth.

I also co-founded Congenica Ltd, to provide genetic analysis services to the NHS and other healthcare providers, which will also analyse patients' DNA sequenced in the UK 100,000 Genomes Project. I will bring my knowledge from the DDD project to advise on autism gene discovery, and make the bioinformatics expertise of the Sanger Centre available

Research outputs
List up to 20 of your most significant research outputs, ensuring that at least five of these are from the last five years. For 10 of these outputs, provide a statement describing their significance and your contribution (up to 50 words per output).

Research outputs may include (but are not limited to):

- Peer-reviewed publications and preprints
- Datasets, software and research materials
- Inventions, patents and commercial activity

For original research publications indicate those arising from Trust-funded grants in **bold**, and provide the PubMed Central ID (PMCID) reference for each of these. Please refer to guidance

notes.

Please give citation in full, including title of paper and all authors Citations to preprints should state "Preprint", the repository name and the articles persistent identifier (e.g DOI).*

*(*All authors, unless more than 10, in which case please use 'et al', ensuring that your position as author remains clear.)*

De novo mutations in regulatory elements in neurodevelopmental disorders. Short PJ, McRae JF, Gallone G, Sifrim A, Won H, Geschwind DH, Wright CF, Firth HV, FitzPatrick DR, Barrett JC, Hurles ME.Nature. 2018 Mar 29;555(7698):611-616. doi: 10.1038/nature25983.

Distinct genetic architectures for syndromic and nonsyndromic congenital heart defects identified by exome sequencing. Sifrim A, Hitz MP, Wilsdon A, Breckpot J, Turki SH, Thienpont B, McRae J, Fitzgerald TW, Singh T, Swaminathan GJ, Prigmore E, Rajan D, Abdul-Khaliq H, Banka S, Bauer UM, Bentham J, Berger F, Bhattacharya S, Bu'Lock F, Canham N, Colgiu IG, Cosgrove C, Cox H, Daehnert I, Daly A, Danesh J, Fryer A, Gewillig M, Hobson E, Hoff K, Homfray T; INTERVAL Study, Kahlert AK, Ketley A, Kramer HH, Lachlan K, Lampe AK, Louw JJ, Manickara AK, Manase D, McCarthy KP, Metcalfe K, Moore C, Newbury-Ecob R, Omer SO, Ouwehand WH, Park SM, Parker MJ, Pickardt T, Pollard MO, Robert L, Roberts DJ, Sambrook J, Setchfield K, Stiller B, Thornborough C, Toka O, Watkins H, Williams D, Wright M, Mital S, Daubeney PE, Keavney B, Goodship J; UK10K Consortium, Abu-Sulaiman RM, Klaassen S, Wright CF, Firth HV, Barrett JC, Devriendt K, FitzPatrick DR, Brook JD; Deciphering Developmental Disorders Study, Hurles ME. Nat Genet. 2016 Sep;48(9):1060-5. doi: 10.1038/ng.3627. Epub 2016 Aug 1.

Discovery of four recessive developmental disorders using probabilistic genotype and phenotype matching among 4,125 families. Akawi N, McRae J, Ansari M, Balasubramanian M, Blyth M, Brady AF, Clayton S, Cole T, Deshpande C, Fitzgerald TW, Foulds N, Francis R, Gabriel G, Gerety SS, Goodship J, Hobson E, Jones WD, Joss S, King D, Klena N, Kumar A, Lees M, Lelliott C, Lord J, McMullan D, O'Regan M, Osio D, Piombo V, Prigmore E, Rajan D, Rosser E, Sifrim A, Smith A, Swaminathan GJ, Turnpenny P, Whitworth J, Wright CF, Firth HV, Barrett JC, Lo CW, FitzPatrick DR, Hurles ME; DDD study. Nat Genet. 2015 Nov;47(11):1363-9. doi: 10.1038/ng.3410. Epub 2015 Oct 5.

Large-scale discovery of novel genetic causes of developmental disorders. Deciphering Developmental Disorders Study. Nature. 2015 Mar 12;519(7542):223-8. doi: 10.1038/nature14135. Epub 2014 Dec 24. Variation in genome-wide mutation rates within and between human families. Conrad DF, Keebler JE, DePristo MA, Lindsay SJ, Zhang Y, Casals F, Idaghdour Y, Hartl CL, Torroja C, Garimella KV, Zilversmit M, Cartwright R, Rouleau GA, Daly M, Stone EA, Hurles ME, Awadalla P; 1000 Genomes Project. Nat Genet. 2011 Jun 12;43(7):712-4. doi: 10.1038/ng.862.

A map of human genome variation from population-scale sequencing. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, Gibbs RA, Hurles ME, McVean GA. Nature. 2010 Oct 28;467(7319):1061-73. doi: 10.1038/nature09534.

Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF, Giannoulatou E, Holmes C, Marchini JL, Stirrups K, Tobin MD, Wain LV, Yau C, Aerts J, Ahmad T, Andrews TD, Arbury H, Attwood A, Auton A, Ball SG, Balmforth AJ, Barrett JC, Barroso I, Barton A, Bennett AJ, Bhaskar S, Blaszczyk K, Bowes J, Brand OJ, Braund PS, Bredin F, Breen G, Brown MJ, Bruce IN, Bull J, Burren OS, Burton J, Byrnes J, Caesar S, Clee CM, Coffey AJ, Connell JM, Cooper JD, Dominiczak AF, Downes K, Drummond HE, Dudakia D, Dunham A, Ebbs B, Eccles D, Edkins S,

Edwards C, Elliot A, Emery P, Evans DM, Evans G, Eyre S, Farmer A, Ferrier IN, Feuk L, Fitzgerald T, Flynn E, Forbes A, Forty L, Franklyn JA, Freathy RM, Gibbs P, Gilbert P, Gokumen O, Gordon-Smith K, Gray E, Green E, Groves CJ, Grozeva D, Gwilliam R, Hall A, Hammond N, Hardy M, Harrison P, Hassanali N, Hebaishi H, Hines S, Hinks A, Hitman GA, Hocking L, Howard E, Howard P, Howson JM, Hughes D, Hunt S, Isaacs JD, Jain M, Jewell DP, Johnson T, Jolley JD, Jones IR, Jones LA, Kirov G, Langford CF, Lango-Allen H, Lathrop GM, Lee J, Lee KL, Lees C, Lewis K, Lindgren CM, Maisuria-Armer M, Maller J, Mansfield J, Martin P, Massey DC, McArdle WL, McGuffin P, McLay KE, Mentzer A, Mimmack ML, Morgan AE, Morris AP, Mowat C, Myers S, Newman W, Nimmo ER, O'Donovan MC, Onipinla A, Onyiah I, Ovington NR, Owen MJ, Palin K, Parnell K, Pernet D, Perry JR, Phillips A, Pinto D, Prescott NJ, Prokopenko I, Quail MA, Rafelt S, Rayner NW, Redon R, Reid DM, Renwick, Ring SM, Robertson N, Russell E, St Clair D, Sambrook JG, Sanderson JD, Schuilenburg H, Scott CE, Scott R, Seal S, Shaw-Hawkins S, Shields BM, Simmonds MJ, Smyth DJ, Somaskantharajah E, Spanova K, Steer S, Stephens J, Stevens HE, Stone MA, Su Z, Symmons DP, Thompson JR, Thomson W, Travers ME, Turnbull C, Valsesia A, Walker M, Walker NM, Wallace
C, Warren-Perry M, Watkins NA, Webster J, Weedon MN, Wilson AG, Woodburn M, Wordsworth BP, Young AH, Zeggini E, Carter NP, Frayling TM, Lee C, McVean G, Munroe PB, Palotie A, Sawcer SJ, Scherer SW, Strachan DP, Tyler-Smith C, Brown MA, Burton PR, Caulfield MJ, Compston A, Farrall M, Gough SC, Hall AS, Hattersley AT, Hill AV, Mathew CG, Pembrey M, Satsangi J, Stratton MR, Worthington J, Deloukas P, Duncanson A, Kwiatkowski DP, McCarthy MI, Ouwehand W, Parkes M, Rahman N, Todd JA, Samani NJ, Donnelly P. Nature. 2010 Apr 1;464(7289):713-20. doi: 10.1038/nature08979.

Origins and functional impact of copy number variation in the human genome. Conrad DF, Pinto D, Redon R, Feuk L, Gokcumen O, Zhang Y, Aerts J, Andrews TD, Barnes C, Campbell P, Fitzgerald T, Hu M, Ihm CH, Kristiansson K, Macarthur DG, Macdonald JR, Onyiah I, Pang AW, Robson S, Stirrups K, Valsesia A, Walter K, Wei J; Wellcome Trust Case Control Consortium, Tyler-Smith C, Carter NP, Lee C, Scherer SW, Hurles ME. Nature. 2010 Apr 1;464(7289):704-12. doi: 10.1038/nature08516. Epub 2009 Oct 7.

Large, rare chromosomal deletions associated with severe early-onset obesity. Bochukova EG, Huang N, Keogh J, Henning E, Purmann C, Blaszczyk K, Saeed S, Hamilton-Shield J, Clayton-Smith J, O'Rahilly S, Hurles ME, Farooqi IS. Nature. 2010 Feb 4;463(7281):666-70. doi: 10.1038/nature08689. Epub 2009 Dec 6.

Global variation in copy number in the human genome. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H, Shapero MH, Carson AR, Chen W, Cho EK, Dallaire S, Freeman JL, González JR, Gratacòs M, Huang J, Kalaitzopoulos D, Komura D, MacDonald JR, Marshall CR, Mei R, Montgomery L, Nishimura K, Okamura K, Shen F, Somerville MJ, Tchinda J, Valsesia A, Woodwark C, Yang F, Zhang J, Zerjal T, Zhang J, Armengol L, Conrad DF, Estivill X, Tyler-Smith C, Carter NP, Aburatani H, Lee C, Jones KW, Scherer SW, Hurles ME. Nature. 2006 Nov 23;444(7118):444-54.

Total number of peer-reviewed publications which you have authored/co-authored. Please exclude abstracts and literature reviews.	133

Current and recent research funding (including Wellcome Trust grants)
Please list all held in the last five years and any key prior grants (list the most recent first). State the name of the awarding body, name(s) of grantholder(s), title of project, amounts awarded, your role in the project, and start and end dates of support. For all active grants, indicate the number of hours per week that are spent on each project.

The Deciphering Developmental Disorder Study - Wellcome and the Department of Health, and the Wellcome
Sanger Institute [grant number WT098051].

The Deciphering Developmental Disorder study - Health Innovation Challenge Fund [grant number HICF-1009-003].

Please describe how the currently active grants listed above relate to this application
As a part of the Deciphering Developmental Disorders (DDD), we have conducted WES, WGS, and genotyping of several individuals (N ~ 14,000) with developmental disorders. Many of these have co-morbid autism. We will include these individuals in the autism GWAS. Further, we have expertise in collecting DNA samples through postal saliva kits, genotyping and analysis of the genotypes, and feedback of relevant findings as a part of the DDD study. This will be useful for the current study.

Time spent on research How many hours per week do you spend on research?
40

How many hours per week will be spent on this project?
4

3

Applicant	
Full Name	Prof David Rowitch
Department	Paediatrics
Division	School of Clinical Medicine
Organisation	University of Cambridge
Address Line 1	Hills Road
City/Town	Cambridge
Postcode	CB20QQ
Country	
Telephone No.	[redacted]
Email Address	[redacted]

Career history (current/most recent first)			
From	**To**	**Position**	**Organisation**
03/2016	05/2026	Professor and Head	University of Cambridge
07/2006	02/2016	Professor and Chief of Neonatology	University of California San Francisco
07/1999	06/2006	Assisitant, then Associate Professor	Harvard Medical School
07/1996	06/1999	Postdoctoral Fellow	Harvard University

Education/training

From	To	Qualification	Subject	Organisation
07/1989	06/1996	Intern, Resident and Neonatology Fellow	Paediatrics and Neonatology	Children's Hospital Boston
07/1982	06/1989	Doctor of Medicine (MD)	Medicine	University of California, Los Angeles
07/1984	06/1987	Doctor of Philosophy (PhD;DPhil)	Biochemistry	University of Cambridge

Source(s) of personal salary support

University of Cambridge

Career contributions
What are your most important research-related contributions to date? These may include contributions to health policy or practice, or to technology or product discovery and development.

My laboratory was amongst the first to investigate the developmental genetics of glial lineages of the CNS. Olig1/2 genes are master regulators of motor neuron and oligodendrocyte precursor (OPC) specification (Lu et al, 2002, Cell). More recently, we described a novel function for OPCs in regulating white matter angiogenesis through expression of Wnt ligands (Yuen et al, 2014, Cell). We have proposed that astrocytes comprise a functionally heterogeneous population and have shown that they are specified in the embryo according to a spatial template (Muroyama et al., 2005, Nature; Tsai et al., 2012, Science) and optimized to support local neuronal circuits Molofsky et al., 2014, Nature, Kelly et al., 2018, Neuron). I have led or participated in studies showing Olig2 is a useful biomarker in human neonatal white matter lesions (Billards et al., 2008, Brain Path) and malignant gliomas (Ligon et al, 2004, JNEN, Ligon et al., 2007, Neuron, Griveau et al., 2018, Cancer Cell).

Clinical: I am PI of a first-in-man Phase I clinical study of oligodendrocyte progenitor transplant in boys with the rare leukodystrophy, Pelizeaus Mersbacher Disease (PMD), which reported safety and putative MRI myelin signals at 1-year post-transplant (Gupta et al., 2012, Sci Trans Med). I co-founded the Newborn Brain Research Institute with neurologist Donna Ferriero at UCSF in 2006. As Head of Neonatology (2006-2015), I began a neonatal brain bank to study human development and injuries leading to cerebral palsy. With colleagues, we have shown evidence for late migration of young neurons in term neonates (Sanai et al., 2011, Nature), which might comprise a target of injury in hypoxic-ischemic encephalopathy. Although I was recently been appointed as Professor of Paediatrics at Cambridge (2016) I have made a significant contribution to NIHR in a number of ways. I support the NHS England East of England Genomics Medicine Centre by promoting recruitment of paediatrics patients with rare diseases and their families. Overall recruitment for rare diseases within Cambridge, and the Centre as a whole, is ahead of trajectory.

Summary: My total publications number 176 with 29256 citations and H-index = 86.

Experience relevant to this proposal
Please summarise your key achievements and experience which are relevant to this proposal.

My laboratory program focuses on the developmental genetics of glial cells of the CNS called astrocytes and myelinating oligodendrocytes. As a neonatologist physician-scientist, I look for extensions of neuroscience that can impact our understanding of human neurological diseases, particularly cerebral palsy and neurogenetic diseases. I have worked productively with neurobiologists in mammalian and invertebrate systems as well as computer and imaging scientists. In my current capacity in Cambridge University, I am developing pediatric applications of

genomic and stem cell medicine in the neonatal ICU and for childhood rare brain disorders.

On this collaborative grant my role is as Head of the Department of Paediatrics in Cambridge University and Head of the Paediatric Theme of the NIHR Biomedical Research Centre (BRC) in Cambridge where I have organised a network of Consultant Paediatricians across the East of England who can help us recruit autism cases to the UK Autism Biobank that is part of this grant. In addition, I have had 20 years of research experience in studying gene expression in the brain and how this affects risk of neurodevelopmental disorders, so can advise on the bioinformatics as autism gene discovery progresses.

Research outputs
List up to 20 of your most significant research outputs, ensuring that at least five of these are from the last five years. For 10 of these outputs, provide a statement describing their significance and your contribution (up to 50 words per output).

Research outputs may include (but are not limited to):

- Peer-reviewed publications and preprints
- Datasets, software and research materials
- Inventions, patents and commercial activity

For original research publications indicate those arising from Trust-funded grants in **bold**, and provide the PubMed Central ID (PMCID) reference for each of these. Please refer to guidance notes.

Please give citation in full, including title of paper and all authors Citations to preprints should state "Preprint", the repository name and the articles persistent identifier (e.g DOI).*

*(*All authors, unless more than 10, in which case please use 'et al', ensuring that your position as author remains clear.)*

1. Lu QR, Sun T, Zhu Z, Ma N, Garcia M, Stiles CD, Rowitch DH. Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte lineage connection. Cell, 2002; 109, 75-86. In prior work (Lu et al., 2000, Neuron), we reported Olig transcription factors in the oligodendrocyte lineage. This paper established that Olig2 function regulates pattern formation and embryonic oligodendrocyte and motor neuron development. Olig2 was the first transcription factor described that was essential for glial development.

2. Muroyama Y, Fujiwara Y, Orkin SH, Rowitch DH. Specification of astrocytes by bHLH protein SCL in a restricted region of the neural tube. Nature. 2005 Nov 17;438(7066):360-3. Here we showed that bHLH protein SCL/TAL1 engaged in cross-antagonistic interactions with Olig2 to regulate astrocyte versus oligodendrocyte subtype specification. The paper showed a segmentally restricted mechanism for gliogenesis regulated by CNS pattern formation.

3. Ligon KL, Huillard E, Mehta S, KesariS, Liu H, Alberta J, Bachoo RM, Kane M, Louis DN, DePinho RA, Anderson DJ, Stiles CD, Rowitch DH. Olig2-regulated lineage-restricted pathway controls replication competence in neural stem cells and malignant glioma. Neuron. 2007; 53:503-17. As Olig2 is expressed--and has an essential role-- in high-grade glioma, this paper provided a demonstration that CNS developmental regulatory factors can also play important roles in brain cancer. PMCID: PMC1810344

4. Sanai N, Nguyen T, Ihrie RA, Mirzadeh Z, Tsai HH, Wong M, Gupta N, Berger MS, Huang E, Garcia-Verdugo JM, Rowitch DH, Alvarez-Buylla A. Corridors of migrating neurons in the human brain and their decline during infancy. Nature. 2011 Sep 28;478(7369):382-6. This paper

reported a novel population of late migrating interneurons in the post-natal human forebrain, raising the possibility that this late developmental mechanism might be affected adversely by neonatal brain injury. PMCID: PMC3197903

5. Fancy SPJ, Harrington EP, Yuen T, Silbereis JC, Zhao C, Baranzini SE, Bruce C, Otero JJ, Huang EJ, Nusse R, Franklin RJM and Rowitch DH. Axin2 as regulatory and therapeutic target in newborn brain injury and remyelination. Nat. Neurosci. 2011 Jun 26;14(8):1009-16. We previously showed the Wnt pathway was an important modulator of oligodendrocyte maturation (Fancy et al., 2009, Genes Dev). Here, we reported Wnt pathway activation in human neonatal white matter injury. We found that the small molecule Wnt inhibitor XAV939 accelerated remyelination by stabilizing levels of Axin2. PMCID: PMC3145042

6. Heine VM, Griveau A, Chapin C, Ballard PL, Chen JK, Rowitch DH. A small-molecule smoothened agonist prevents glucocorticoid-induced neonatal cerebellar injury. Sci Transl Med. 2011 Oct 19;3(105):105. About 20% of extremely preterm infants will show cerebellar hypoplasia associated with severely impaired long-term neurodevelopmental outcomes. Here, we showed that a small molecule agonist of Sonic Hedgehog (SAG) can protect against glucocorticoid-induced cerebellar injury in mouse neonatal brain. PMCID: PMC3694585

7. Tsai HH, Li H, Fuentealba LC, Molofsky AV, Taveira-Marques R, Zhuang H, Tenney A, Murnen AT, Fancy SP, Merkle F, Kessaris N, Alvarez-Buylla A, Richardson WD, Rowitch DH. Regional Astrocyte Allocation Regulates CNS Synaptogenesis and Repair. Science. 2012 337(6092):358-62. This paper showed astrocytes are generated in multiple progenitor domains and are allocated according to a region-restricted template. It suggested that spatially diverse programmes of astrogenesis might confer long-term specialized functions. PMCID: PMC4059181

8. Gupta N, Henry RG, Strober J, Kang SM, Lim DA, Bucci M, Caverzasi E, Gaetano L, Mandelli ML, Ryan T, Perry R, Farrell J, Jeremy RJ, Ulman M, Huhn SL, Barkovich AJ, Rowitch DH. Neural stem cell engraftment and myelination in the human brain. Sci Transl Med. 2012 4, 155. In the fatal leukodystrophy Pelizaeus-Merzbacher disease (PMD), endogenous oligodendrocytes are unable to produce myelin due to mutation of PLP1. This first in man Phase I study of neural stem cells for treatment of white matter disorders has shown long-term safety promoting later Phase clinical investigation. PMCID: PMC3893824

9. AV Molofsky, KW Kelley, H-HTsai, SA Redmond, SM Chang, L Madireddy, JR Chan, SE Baranzini, EM Ullian, and DH Rowitch. Astrocyte positional signals maintain sensorimotor circuit integrity. Nature, 2014 509:189-94. This paper is the first to show that region-specific astrocyte-encoded function is required for neural circuit activity and integrity. This paradigm may extend to other aspects of the CNS in the regulation of normal neurological activity and in disease states. PMCID: PMC4057936

10. TJ Yuen, JC Silbereis, A Griveau, SM Chang, R Daneman, SPJ Fancy, H Zahed, E Maltepe and DH Rowitch. Oligodendrocyte-encoded HIF function couples postnatal myelination and white matter angiogenesis, Cell, 2014 158:383-96. In this paper, we show that oligodendrocyte precursors express angiogenic Wnts under control of the HIF pathway and that Wnt activity promotes white matter vascularization, coordinating this with the energy-expensive process of myelination PMCID: PMC4149873

Total number of peer-reviewed publications which you have authored/co-authored. Please exclude abstracts and literature reviews.	150

Current and recent research funding (including Wellcome Trust grants)

Please list all held in the last five years and any key prior grants (list the most recent first). State the name of the awarding body, name(s) of grantholder(s), title of project, amounts awarded, your role in the project, and start and end dates of support. For all active grants, indicate the number of hours per week that are spent on each project.
Completed/Significant: Howard Hughes Medical Institute Investigator. Pathogenesis and Rational Treatment of Cerebral Palsy 04/01/2008-09/01/2015 $5M Paul G. Allen Family Foundation-Allen Distinguished Investigators (ADI) Program (Ullian, Rowitch, co-PIs). Matching Regional Diversity with Function: Unique Astrocyte Signals Mature Regionally Matched Neurons 05/01/2015-4/30/2018 $750K Active: NIH/NINDS (P01NS083513; PI-Rowitch) Regulation of Cellular Pathways in Human Brain Development 07/01/14-06/30/19 $5M Wellcome Trust Senior Investigator Award (PI-Rowitch) Understanding Astrocyte Regional and Functional Diversity 12/01/15-11/31/21 £3M Action Medical Research (PI-Rowitch). Personalised cell-based models of hypomyelinating leukodystrophy 09/01/15-10/31/17 £160K NIHR Cambridge Biomedical Research Centre (Theme co-lead Paediatrics-Rowitch) Women's Health and Paediatrics Theme 04/01/17-03/31/22 £1.8M

Please describe how the currently active grants listed above relate to this application
I am the co-lead of the NIHR Cambridge Biomedical Research Centre (Paediatrics). Our consent is designed for recontact and recall-by-genotype. This will be very useful for designing the UK Autism Biobank study.

Time spent on research How many hours per week do you spend on research?
40

How many hours per week will be spent on this project?
4

Applicant details summary
Please also complete brief summary details for the lead applicant and each applicant listed above.

Name	**Organisation**	**Role in project**
Prof Daniel Geschwind	UCLA	Co-I: Integrating functional data to understand the biology of the loci identified
Dr Matthew Hurles	Wellcome Trust Sanger Institute	Co-I: Statistical analysis, CNV

Applicant details summary Please also complete brief summary details for the lead applicant and each applicant listed above.		
Name	**Organisation**	**Role in project**
		analysis, and overall project management
Professor David Rowitch	University of Cambridge	Co-I: Overall project management, and evaluating clinical impact of outcomes
Professor Simon Baron-Cohen	University of Cambridge	Lead investigator: Overall project management

5. Collaborators

Will you require any key collaborators for this proposal?	Yes

Please list any key collaborators* (name and organisation) and provide a very brief outline of their role in the proposed research.
**The collaborators named may be replaced with suitable alternatives should it be necessary or appropriate to do so.*

Collaborators include: Professors Thomas Bourgeron (Pasteur Institut, Paris), Jakob Grove (Aarhus University, Denmark), Sarah Medland (QIMR - Berghofer), Nick Martin (QIMR - Berghofer), Dorret Boomsma (Vrije University, Amsterdam), John Perry (MRC Epidemiology Unit, University of Cambridge), Wendy Chung (Columbia University) for the SFARI SPARK consortium, Anders Borglum (Aarhus University) and Mark-Daly (MIT-Harvard Broad Institute) for the PGC-iPSYCH consortium, Joseph Buxbaum (Mount Sinai Hospital, NY) for the Autism Sequencing Consortium, Declan Murphy (KCL, London) for the EU-AIMS consortium, Stephan Sanders (UCSF), and Drs Carrie Allison (ARC, Cambridge) and Varun Warrier (ARC, Cambridge).

The data collection and analytical methods required to identify common genetic variants associated with autism are immense. The data cannot be generated by a single lab and concerted efforts are required on a global scale to recruit, phenotype, and genotype participants. This series of studies could not be achieved without collaborating with other researchers.

Murphy and **Bourgeron** lead the EU-AIMS genetics sample collection and statistical analysis for more than 800 autistic individuals. **Daly** leads the PGC-autism consortium. **Buxbaum** leads the Autism Sequencing Consortium. **Grove** and **Borglum** lead the iPSYCH-autism consortium. **Medland** and **Martin** lead the efforts to recruit, genotype and analyse data from 10,000 autistic individuals in Australia. They will also collect data on autistic traits in the QIMR cohorts. Chung is the Scientific Director of the Simon's SPARK collection, and leads the effort to collect 50,000 autism cases in the United States. **Boomsma** is the PI of the Netherlands Longitudinal Twin Registry and a leading behavioural geneticist. She has been phenotyping participants using the Children's Behaviour Checklist and will contribute to the autistic traits GWAS. **Sanders** is an expert autism geneticist and has expertise in the analysis of structural variants, and more recently, variation in the non-coding genome. **Perry** is an expert on GWAS and has conducted several GWAS of reproductive behaviour, and more recently, loneliness. **Warrier** will coordinate with existing genotyped cohorts to re-phenotype participants using measures of autistic traits. **Allison** will coordinate the phenotyping of the UK autism cohort.

I confirm that the collaborators named above have agreed to be involved, as described, in the proposed research and are willing for their details to be included as part of this application.

Confirmed

6. Related applications

Is this or a similar application for funding currently under consideration elsewhere?	Yes

Please provide name(s) of funding organisation(s) and decision date(s)
Yes, the autistic traits GWAS is under consideration as [redacted] has made significant contributions to the Collaborative grant application that [redacted] should be a co-applicant. However, on the advice of the [redacted] since [redacted] is also a candidate for the [redacted] is listed as a collaborator here. If both are funded, we will use the salary to hire a second statistical analyst to analyse the case-control autism GWAS as [redacted] will embark on the [redacted].

Is this a resubmission of an application submitted to the Trust within the last 24 months?	No

7. Research summary

Research summary
Please provide a summary of your proposed research, including key goals, for an expert audience

Autism is a life-long developmental condition with a prevalence of approximately 1%, and heritability estimates of between 64-92%. It is polygenic, with variants across the frequency spectrum (from rare to common) contributing to risk. Considerable progress has been made in identifying rare variants in autism, but the largest genome-wide association study (GWAS) of autism (18,000 cases, 28,000 controls) identified only five loci associated with autism, compared to 179 in schizophrenia (40,000 cases, 65,000 controls). Polygenic scores account for 2.45% of the variance in autism, despite 30-50% of the variance in risk for autism being attributable to common variants, suggesting many more common variants remain to be found. Additionally, autistic traits are normally distributed in the general population, but there is no well-powered GWAS of autistic traits. The proposed research aims to accelerate the discovery of common, low frequency, and copy number variants in autism and autistic traits. Specifically, we will: (1) Establish a UK-wide Autism Biobank (N = 10,000 cases), with links to electronic health records and who can be recalled by genotype; (2) Conduct a GWAS of autism in 100K cases from around the world; and (3) Conduct a GWAS of autistic traits in the population in 250K individuals

Lay summary
Please provide a summary of your proposed research that people who may not be familiar with the subject can understand. We may edit your summary and then use it to describe your research on our website and elsewhere.

Autism is a lifelong developmental condition and approximately 1% of the population are thought to be autistic. It is largely genetic, and between 400-1000 genes are thought to contribute to autism. However, fewer than 100 genes have been identified that are linked to autism. We propose to accelerate gene discovery by collecting DNA samples from 10,000 autistic individuals in the UK and their immediate families. We will combine this genetic information with genetic information from 90,000 additional autistic individuals already ascertained across the world. This large-scale

resource will enable us to identify several genetic variants that contribute to the development of autism. This information will allow us to better understand the biology of autism, improve on existing methods for diagnosing autism, and investigate if there are genetically-defined subgroups within autistic people.

8. Details of research project

Please provide full details of your proposal. These should include:
(a) Aims and research questions;
(b) Essential background;
(c) Approach;
(d) Expected outcomes;
(e) Timetable and milestones, where appropriate.

No more than **5,000** words should be used to describe the proposal.

Common Variant Genetics of Autism and Autistic Traits (GWAS) Consortium

Principal Applicants: Professors Simon Baron-Cohen (Autism Research Centre/ARC, Cambridge), Matthew Hurles (Wellcome Sanger Institute, Hinxton), David Rowitch (Pediatrics Department, Cambridge), Daniel Geschwind (UCLA)

Other Collaborators: Professors Thomas Bourgeron (Pasteur Institut, Paris), Jakob Grove (Aarhus University, Denmark), Sarah Medland (QIMR - Berghofer), Nick Martin (QIMR - Berghofer), Dorret Boomsma (Vrije University, Amsterdam), John Perry (MRC Epidemiology Unit, University of Cambridge), Wendy Chung (Columbia University) for the SFARI SPARK consortium, Anders Borglum (Aarhus University) and Mark Daly (MIT-Harvard Broad Institute) for the PGC-iPSYCH consortium, Joseph Buxbaum (Mount Sinai Hospital, NY) for the Autism Sequencing Consortium, Declan Murphy (KCL, London) for the EU-AIMS consortium, Stephan Sanders (UCSF), and Drs Carrie Allison (ARC, Cambridge) and Varun Warrier (ARC, Cambridge).

Aims and objectives

Objective: To accelerate gene-discovery, genetic stratification, and biomarker identification in autism.

Aims:

1. To recruit 10,000 autistic individuals from the UK and where possible, their families;
2. Where possible, to deeply-phenotype the 10,000 autistic individual (UK Autism Biobank), link to Electronic Health Records (EHRs), and provide a rich resource for gene discovery and recall-by-genotype studies in autism;
3. To conduct a GWAS and CNV meta-analysis of 100,000 autistic individuals with data collected from the UK (10K), the US (SPARK, N = 50K), Australia (N = 10K), the iPSYCH and Psychiatric Genomics Consortium (N = 20K), the Autism Sequencing Consortium (5K) and other cohorts (5K) (Autism meta-analysis);
4. To conduct a parallel GWAS of autistic traits (250,000, UK Biobank, and other sources), and conduct a multi-trait GWAS of autism and autistic traits;
5. To perform fine-mapping of significant loci and identify functional genes by integrating gene expression (e.g., integration of human cell atlas and BRAIN initiative databases), meQTL, eQTL and chromatin interactions data from neural tissues;
6. To investigate how polygenic scores for autism alter normative developmental trajectories, and brain structure and function in adolescents and adults;
7. To investigate if polygenic scores can identify sources of heterogeneity based on sex, IQ, social and non-social domains of autism, and related co-morbidities;
8. To identify modifiable risk factors for autism using Mendelian Randomization and related methods.

Importance

Autism is highly heritable, with twin and family-based heritability estimates between 64-92%[1–3]. Autism is a neurodevelopmental condition with a marked sex-difference[4,5] (3:1, male to female ratio), and high mental and physical health co-morbidities[6]. Autism has a prevalence of around 1 to 2%[7], with approximately 700,000 people on the autism spectrum in the UK[a]. 70% of autistic individuals have co-morbid epilepsy, anxiety, depression, or learning disability[8]. These and other difficulties result in a 2.5-fold increased odds of mortality[9] or 18 years of reduced life-expectancy. 35% of autistic individuals report planned or attempted suicide[10], and 85% are not in full-time employment[11]. In the UK alone, the cost of supporting autistic individuals is £32 billion annually[11], which is more than cardiovascular disease, stroke, and cancer combined, making it the most costly medical condition in the UK. In contrast, research funding for autism and learning disability combined is £4 million per year[b] in the UK compared to £521 million for cancer[c]. There is an urgent need to both evaluate interventions to improve clinical and quality of life outcomes, and to better understand the biology of autism to improve detection, diagnosis and precision interventions, based on gene discovery and well-defined subgrouping. The proposed program aims to accelerate our knowledge of the biology of autism.

Autistic people without intellectual disability (ID), who may constitute at least two-thirds of all autistic people[7], score at the extreme high end of a normally distributed continuum of autistic traits in the general population[12,13]. Autistic traits are milder, subclinical manifestations of the symptoms of clinically diagnosed autism. Large-scale research efforts have advanced our understanding of autism genetics in three ways. First, it is clear that autism is highly polygenic, with recent estimates suggesting that between 400-1,000 genes may be involved[14–16]. Second, variation across the allele-frequency spectrum contributes to autism: rare genetic variants, including large copy number variations, are estimated to account for 10-

[a] https://digital.nhs.uk/catalogue/PUB05061

[b] http://nationalautismproject.org.uk/wp-content/uploads/2017/01/autism-dividend-report.pdf

[c] http://b.3cdn.net/joinmq/1f731755e4183d5337_apm6b0gll.pdf

30%[17–19] of the total risk, with large per-variant effect sizes, while common, inherited variation accounts for 15-50%[20–22] of the risk *en masse*, with each common variant contributing only modestly to the total risk[23]. Third, the phenotypic heterogeneity evident in autism can in part be explained at a genetic level: Whole exome sequencing (WES) has identified specific clinical and behavioural phenotypes of autism attributable to variants in specific genes (e.g., *CHD8*[24], *ADNP*[25], and *PTEN*[26]). In contrast, emerging evidence from GWAS of autistic traits suggest at least two distinct sources of shared risk with autism. One emerges from the social domain (difficulties in social interaction and communication), and the other from the non-social domain (unusually repetitive and unusually restricted behaviours and interests)[27–29]. Currently only 78 genes enriched for *de novo* protein truncating variants[d], 11 CNVs, and 5 SNPs[23] have been associated with autism with 'high confidence' (i.e., reliably replicable). Clearly therefore, this means only a fraction of the total genetic risk for autism has been discovered. These genetic discoveries have provided insights into pharmacological targets, with current clinical trials for mTOR inhibitors[30] for *PTEN*[e]*, NF1, TSC1 and TSC2*, mGluR antagonists for *FMR1*[31], and rescue of autistic-like phenotypes in animal models for *SHANK3*[32,33] and *MECP2*[34].

A GWAS is a very useful method to investigate the contribution of both common and low-frequency genetic variants to autism (accounting for 15-50%[20–22] of the total variance). Additionally, compared to other methods like WES or whole genome sequencing (WGS), GWAS is cost effective (£30 per genotyping chip compared to £1,000 for WGS), can interrogate variants in the non-coding genome compared to WES, and there are well-developed methods to interpret these findings, and identify downstream functional enrichment in cell types, tissues, and pathways. Further, well-powered GWAS provide well-powered polygenic scores which are relevant to all individuals seemingly independent of the contribution of *de novo* variants[35,36]. Large GWAS in psychiatric conditions have demonstrated that these conditions are highly polygenic, and that with *sufficient sample size*, replicable genetic loci can be identified. For adult psychiatric conditions such as schizophrenia[37], and more recently, major depressive disorder[38] and bipolar disorder[39], GWAS have helped to identify relevant cell[40] and tissue types[37–39], to repurpose drugs[41,42], to understand heterogeneity in the condition[38,43,44], and to identify how polygenic scores alter typical development, illuminating pre-clinical developmental trajectories[45–48]. Importantly, for these adult psychiatric conditions, GWAS have reached a stage where "there is a sample-size-dependent *inflection point* beyond which the number of genome-wide significant loci increases linearly"[49]. For autism, despite considerable SNP-heritability (11-50%)[20,23,50], only 5 genome-wide significant loci have been identified[23], which is far fewer than in other psychiatric conditions with similar twin heritability such as schizophrenia (145 genome-wide significant loci),[37] depression (44 genome-wide significant loci),[38] and bipolar disorder (30 genome-wide significant loci)[51] **(Figure 1)**. The current sample size of the autism GWAS limits identification of associated loci, the statistical power of polygenic scores, and our understanding of biological risk mechanisms, neural and developmental correlates of autism and autistic traits. Thus, there is an urgent need to increase sample sizes, to progress our understanding of autism genetics. In the UK, there is no large-scale genetic study investigating the genetics of autism. Autism is under-represented in large cohorts such as the UK Biobank and ALSPAC, primarily due to the healthy volunteer bias. Autism is not included in large-scale sequencing studies such as the 100,000 Genomes Project[f], and barring seven autism related loci, is not included in the NHS Directory of Genetic Disorders & Genes for Diagnostic Testing[g].

[d] http://spark-sf.s3.amazonaws.com/SPARK_gene_list.pdf

[e] https://www.ptenresearch.org/news/posts/2017/june/launch-of-clinical-trial-to-investigate-the-efficacy-on-neurocognition-and-behaviour-of-everolimus-in-children-and-adolescents-with-pten-mutations/

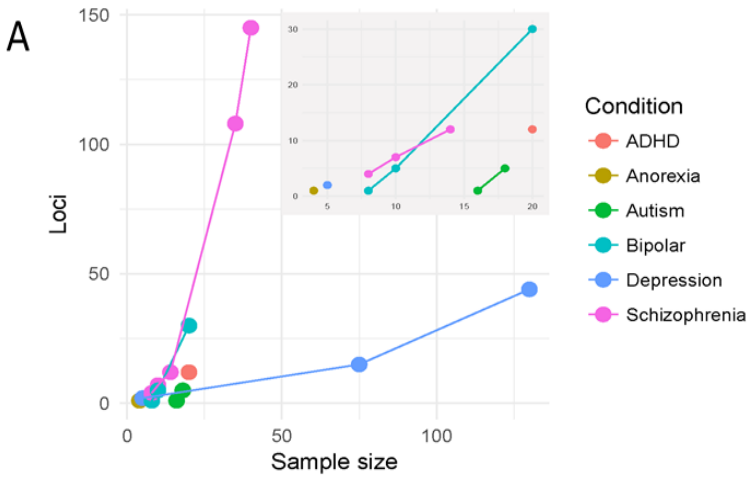
[f] https://www.genomicsengland.co.uk/about-gecip/nominating-rd-or-tumour-type/

[g] https://ukgtn.nhs.uk/fileadmin/uploads/ukgtn/Documents/Resources/Library/Reports_Guidelines/NHS_Directory_of_Genetic_Testing/UKGTN_Directory_of_Genetic_Testing_version_v14_design.pdf

We argue there are 6 outstanding challenges to understand the polygenic architecture of autism: (a) We need **larger sample sizes** to robustly identify genetic variants associated with autism **(Figure 1)** (Aims 1 – 3); (b) We need to systematically integrate functional information from developing and adult neural tissues and cell types, and to fine-map to **identify causal variants** and prioritize functional genes for further analysis (Aim 5); (c) We need **well-powered polygenic scores** to integrate genetic data into diagnosis and to better understand developmental, neural, and gene-environment effects in autism (Aim 6); (d) We need to disentangle the underlying **genetic heterogeneity** in autism (Aim 7); (e) We need to identify **modifiable risk factors** for autism severity (Aim 8). Finally, (f) many cases of 'idiopathic' autism represent the extreme end of a continuum of autistic traits that are distributed across the population, and yet there has not yet been a well-powered **GWAS of autistic traits** (Aim 4).

To address these challenges, we will recruit 10,000 autistic individuals, and where possible their families, to establish a UK-wide autism resource, and use previously genotyped controls from the UK Biobank. We will genotype all autistic individuals. Our primary aim in recruiting these 10,000 autistic individuals is to increase the sample size of existing autism GWAS. In addition, we will link their genetic data to self- and care-giver report phenotypic data collected during the project as well as Electronic Health Records (EHRs), where possible. The UK Autism Biobank will allow for future **recall-by-genotype** studies for targeted phenotyping and enable better causal inference by minimizing confounding[52]. We will combine these individuals with existing genotype data to achieve a total sample size of 100,000 autistic individuals, so that we can conduct **the largest GWAS of autism**. Calculations based on the progress in autism GWAS so far and comparisons with other psychiatric conditions suggest our GWAS of 100K cases will identify between **155 to 165 new loci** associated with autism. In parallel, for the first time, we will conduct **a genome-wide meta-analysis of autistic traits**, and use multivariate methods to meta-analyse the two closely related phenotypes.

Figure 1: Progress in GWAS analysis of psychiatric conditions



Sample size (number of cases, in thousands) to loci identified in psychiatric condition. Inset provides greater resolution for studies that have identified 20 or fewer loci. Note the 'inflection point' in gene discovery in schizophrenia and depression, but not yet in autism.

2. Methods

2.1 UK Autism Biobank:

We will collaborate with a team of world-class international scientists to conduct the largest GWAS of autism, with 100,000 cases, of which 90,000 will be obtained from other cohorts. To achieve this goal, we will collect DNA from 10,000 autistic individuals and their immediate families from the UK (creating the **UK Autism Biobank**). Our research methods to achieve this goal need to be scalable, and convenient for both the participant and the research team. Balancing this, we also aim to collect deeper medical

information using EHRs where possible. In this phase of the study, we will genotype only autistic individuals, and utilize controls from the UK Biobank, to keep costs down.

2.1.1 Rationale: The additional 10,000 cases of autism in the UK are necessary and useful because:

1. We need larger sample sizes to identify genetic variants, develop polygenic scores that predict a sizeable proportion of the variance, and identify biologically relevant gene networks, pathways, cell types and tissues. With larger sample sizes, a small increase in sample size provides greater returns. For instance, a recent GWAS of schizophrenia identified 50 new loci after including 11,000 new cases and 25,000 new controls[37]. Based on this, estimate that every 1,000 cases above 70,000 will identify 4 additional loci. In sum, our additional 10,000 can identify up to 40 new loci, markedly improving the statistical power of polygenic scores.
2. Where possible and based on consent, we will collect deeper phenotypic records using self- and caregiver-report measures. Given the immense heterogeneity of autism, this will provide greater resolution into the specific autism-related phenotypes and co-morbidities at an individual level. The utility of such a design cannot be underestimated. Two such cohorts for autism exist (Autism Genetic Resource Exchange (AGRE), and Simons Simplex Collection (SSC)), both of which have relatively small sample sizes (N < 2,500) but have contributed immensely to gene discovery and understanding the biology of autism[19,21,23,36,53–57]. For example, the role of CNVs were first discovered in the AGRE[58]. The UK cohort will be four times the size of either of these cohorts, and will provide much larger sample sizes for investigating the genetics of specific autism domains, investigating sources of heterogeneity, and other analyses that require deeper phenotypic data. Further, the availability of a **recall-by-genotype** option in the consent form will allow for detailed downstream investigation based on specific genotypes.
3. This will be the largest autism cohort with links to EHRs, enabling investigation of deeper genotype-phenotype relationships, and investigation of the combined contribution of rare CNVs and polygenic risk to autism and co-morbidities. This high-level clinical phenotyping is absent in most existing large-scale cohorts of autism.
4. A substantial fraction of the cohort will be recruited from Child Development Centres, and longitudinal phenotypic data will be collected for these individuals, enabling deeper investigation into gene x environment effects. Given that autism is a neurodevelopmental condition, it is important to understand investigate the condition from a developmental perspective to identify contributors to severity, co-morbidity, and poor subjective wellbeing at specific developmental stages. To date, there is no autism cohort that has longitudinal phenotypic data from participants.
5. While the first phase of the study will only genotype probands, in later phases of the study we will genotype parents and siblings, and conduct WES and/or WGS of probands, parents and siblings to contribute to ongoing international efforts for gene discovery and annotating variants in the non-coding genome[59].
6. This resource will be made available to other researchers who wish to use the anonymized phenotypic and genetic data or who want to re-contact participants to conduct further downstream analysis, including neuroimaging and cognitive testing, providing an excellent resource for multi-modal investigation of autism.
7. Finally, within the UK, there is no large-scale genetic study focusing on autism. Cohorts like the UK Biobank and ALSPAC are under-represented for autism due to the healthy-volunteer bias. The 100,000 Genomes Project does not currently sequence DNA from autistic individuals. Whilst there are UK-specific cohorts for schizophrenia (CLOZUK[37]), eating disorders (Genomics England and Charlotte's Helix[h]), depression (Generation Scotland[38]), Bipolar Disorder (UK Bipolar Disorder Research Network[44]), and ADHD (Cardiff sample[60]), to our knowledge there is no single autism-

[h] https://www.charlotteshelix.net/

specific genetic cohort in the UK. This makes the UK under-represented in international genetic studies of autism in comparison with countries like the United States (SSC[61], AGRE[62]), Denmark (iPSYCH[23]), and Australia (Australian Autism Biobank[i]).

In sum, the extra 10,000 cases will add substantially to gene discovery and polygenic prediction, and will enable deeper analysis of the genotype-phenotype effects and opportunity for recall-by-genotype.

2.1.2 Recruitment and stakeholder engagement

We will recruit 10,000 autistic individuals, and where possible, their families (simplex, multiplex, and multigenerational). Given the number of participants, we will recruit over all five years of the grant, and analyse the first freeze of the genotyped cohort (Year 3). All our recruitment will be done using a combination of 7 strategies (see below), via an online, easy-to-access website, and by posting saliva kits to collect DNA. These are not independent strategies, so we will ensure that participants are not recruited twice into the same study by verifying their name, sex, date of birth, and postal address. The strategies include:

1. **Autism assessment centres**: Autism assessment in the UK is under the remit of the NHS. There are varying models of local services: young children tend to be seen in Child Development Centres (CDCs), and older children are seen within Community Paediatric and/or CAMH services. Adults are seen in specialist clinics, or adult Mental Health, or in Learning Disability Partnerships. Across the UK there are 346 autism assessment centres (NHS and private[j]). NHS Clinical services assess individuals with suspected autism diagnosis across the life-span. We will collaborate closely with relevant clinical services to advertise and recruit participants into the study. In addition, several of these centres have records and contact details of individuals who have previously received an autism diagnosis, so can advertise the study to these individuals. In Cambridgeshire, approximately 400 children are seen for assessment per year across 3 NHS child assessment centres, with 60-70% receiving an autism diagnosis (personal communication, Dr Angharad Walters, Consultant Community Paediatrician, Cambridgeshire). Assuming 10-15% participate in the study, between 90-180 children will be recruited from Cambridgeshire alone over a period of 3-5 years. Extrapolating this to across the UK we estimate that we will recruit between 6,000-8,000 children over 3-years, conservatively assuming a 10% recruitment. **Rowitch** is a consultant paediatrician with excellent contacts with CDCs to facilitate this first recruitment strategy.
2. **Schools for autistic children**: There are 125 special education schools that provide education for children with autism across the UK[k]. Each school typically supports at least 100 children, making a potential number of cases of 12,500. We will advertise the study in these schools to recruit participants into the study. Again, assuming 10% of these schools recruit their pupils, this would result in 1,250 autistic children.
3. **Schools with Special Educational Needs (SEN):** Across the UK, we have identified 1,515 SEN schools, with 1,244,255 with SEN support. 26.9% of these students (334,074) have autism as their primary need[l]. We will contact all these schools to advertise the study to parents of these children. Assuming a 10% participation, we estimate this can recruit approximately 33,000 individuals, far above our target number.
4. **Local Branches of the National Autistic Society**: There are over a 110 local branches of the National Autistic Society (NAS). Each branch typically has 100 families, making a potential number of 11,000 families. We will advertise our project to these. Assuming 10% of NAS

[i] https://www.autismcrc.com.au/get-involved/participate-study/australian-autism-biobank-study

[j] https://www.autism.org.uk/directory/browse/cid=6~aid=1.aspx

[k] http://www.specialneedsuk.org/results.asp?specialityid=1&CurrentPage=1

[l] https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/633031/SFR37_2017_Main_Text.pdf

branches recruit their autistic family members, this would result in 1100 autistic individuals. **Baron-Cohen** is a Vice President of the NAS.

5. **Research databases** including the Cambridge Autism Research Database (CARD, www.autismresearchcentre.com), the UK autism spectrum database (http://www.asd-uk.com/), and the Wales Autism Research Database (www.walesautismresearchcentre.com), DASLNE (http://daslne.org/), the DISCOVER network, BASIS network (http://www.basisnetwork.org/). Together, these research databases have details of more than 10,000 autistic individuals and families. CARD alone has 400 people whose DNA is already banked and available, which is the result of contacting 1,445 autistic individuals over 3 years (recruitment rate of 27%) **(Figure 2)**. CARD will therefore provide an additional 400 autistic people, as the DNA is under the control of Baron-Cohen as PI, and we are likely to recruit autistic individuals from the other research databases and networks.
6. **Other sources** of recruitment will include: Support services for people with autism; statutory services such as Local Authorities and Special Educational Needs teams, as well as professional bodies representing those involved in the assessment of those with autism such as the British Academy of Childhood Disability, ACAMH etc. These latter already advertise research studies to their members.
7. **Media campaign and advertisement** in local and national newspapers: As a part of this, we will employ a PR firm to conduct a media campaign. This will include emailing and calling broadcast and print journalists to offer them a story to publicise the study, advertise in local newspapers, and advertise in autism specific publications such as the Autism Eye (circulation = 13,000) and Your Autism Magazine (circulation = 25,000). This will parallel a social media strategy using platforms such as Twitter and Facebook, with targeted adds for mothers of autistic individuals and autistic individuals. Similar methods have been used elsewhere to recruit participants into studies on autism[63], anorexia[64], and depression (personal communication – Professor Nick Martin). It is estimated that the media campaign increased recruitment by 5-fold in a recent genetic study of depression (personal communication – Professor Nick Martin). We will also ask the Wellcome Trust and the University of Cambridge to recruit via their Science Communication outreach teams. This media recruitment strategy is a safety net for additional autistic individuals but is not critical to achieving our target of 10,000, which is comfortably achievable from the above strategies.

Methods 1-6 will be piloted in Cambridgeshire for a six-month period before being implemented in other counties (see **Timeline**). This will enable us to develop a *best practice* guideline that can be implemented in other counties. To participate, participants will register and provide consent (see below), provide a saliva sample using an Oragene OG-500 DNA kit, and complete phenotypic questionnaires (see below). We have previously collected saliva and extracted DNA for 400 autistic individuals from CARD. These DNA samples have excellent quality and quantity (**Figure 2**). Similarly, the DDD study (**PI: Hurles**), recruited ~40,000 saliva samples by post, yielding DNA that has been used for SNP genotyping, exome and genome sequencing. We will not conduct any in-person testing, and participants will not need to travel to Cambridge. All phenotyping will be completed electronically via a dedicated website that is already created at http://dna.autismresearchcentre.net/ and saliva collection kits will be sent by post. We have used this method to recruit more than 2,000 participants into genetic studies at the Autism Research Centre, including one currently funded by the Templeton World Charity Foundation (TWCF) investigating rare genetic variants in autism multiplex families. Participants will not be paid for their participation, but will be entered into a prize draw as an incentive to participate (£100 for 100 participants).

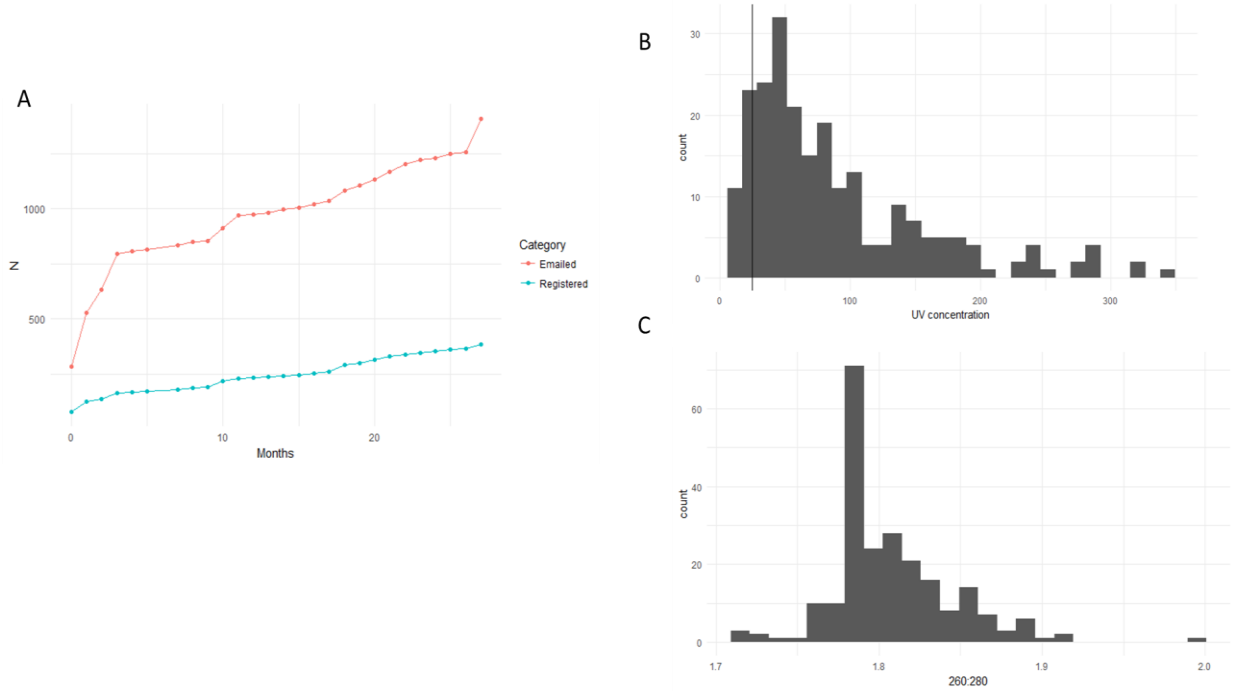
Inclusion and exclusion criteria: Probands (autistic individuals) must have an ICD-9, ICD-10, ICD-11, DSM-III, DSM-IV or DSM-5 diagnosis of autism, which we will confirm by asking for a copy of their clinical report. We have successfully used this diagnostic phenotyping in our current TWCF project. In that study we confirmed 100% accuracy of diagnosis in a sub-sample (i.e., that autistic people are able to

provide their clinical report, and that the diagnosis was based on the use of expert clinicians using validated diagnostic methods [including the ADOS, ADI-R, 3Di, AAA, etc.,] together with clinical judgement). Given the size of the samples recruited, it is not feasible to confirm the autism diagnosis via in-person testing. Furthermore, previous GWAS of neuropsychiatric conditions have identified high genetic correlations between self-report and clinically ascertained cohorts[38,60]. Because this is a genetic study, we will exclude anyone who we suspect may have a non-genetic cause of autism. We will thus exclude participants who have a history of: 1. Preterm birth (less than 36 weeks), and weighed <2kg at birth; 2. Extensive prenatal or perinatal complications or brain injury, including being in NICU for more than 3 days post-birth; 3. A developmental or early acquired injury, disease, or abnormality with known effects on the brain (e.g. meningitis or septicaemia leading to acquired brain injury); and 4. Severe early malnutrition or other potentially environmental contributing cause. Participants of any age can participate in the study. Family members will be asked to confirm that they are biologically related to the proband. We will not exclude participants on the grounds of a known genetic cause of autism as this will not be a large fraction and there is good evidence that polygenic risk also contribute to autism in these cases[35,36].

Stakeholder engagement: To ensure continued participation, we will form a stakeholder advisory committee, comprising autistic individuals, parents of autistic individuals, researchers, and community paediatricians and clinicians. The advisory committee will meet twice annually to: 1. Examine the progress of the study; 2. Prioritize downstream research; and 3. Develop strategies for communicating the research outcomes. All participants and community paediatricians involved in the study will be updated about the study bi-annually via dedicated newsletters. We are already working with Autistica, a UK autism research charity with expertise in working with autistic individuals in shaping research, as part of the AIMS-2-TRIALS consortium. Baron-Cohen is the Lead for WP5 ('Outreach') in this European consortium and has been working with Autism Europe to ensure the views of autistic people are heard in autism research. He also spoke on the topic of autism and human rights at the UN on Autism Awareness Day 2017[m].

[m] http://cambridgeneurodiversityhub.co.uk/what-does-simon-baren-cohen-say-at-the-un-in-2017/

Figure 2: Pilot data from 400 individuals recruited from the CARD



A: Progress in recruitment. Red line represents participants emailed over a 27-month period, and the blue line represents the participants who have registered and provided saliva samples over the same period.
B: Histogram of the DNA concentration of the samples. The black vertical line represents a concentration of 25 ng/uL. 89% of the samples obtained had a concentration greater than 25 ng/uL.
C: Histogram of nucleic acid quantitation using spectrophotometry. Almost all samples had ratios (260:280) around 1.8.

2.1.3 Phenotyping: Previous studies have demonstrated that statistical power of GWAS depends on sample size rather than detailed phenotyping[56,65]. Our recruitment strategy will prioritize number of participants recruited over representativeness of the cohort, or collecting deep phenotypic information. To this end, participants will be encouraged to complete the phenotypic measures, but this will not be compulsory as our primary aim is increase sample size. We will use self- or informant-report and EHR linkage to phenotype participants. To ensure maximum participation we will not collect time-consuming phenotypic measures. Most of our baseline phenotypic measures are designed to be: 1. Easy to administer and brief; 2. Administered across multiple ages; and 3. Self- or informant-report, to allow individuals with moderate or severe learning difficulties, or no learning difficulties, to participate. Further, we will prioritize measures with good psychometric properties including internal consistency and reliability. We will collect phenotypic information of autistic traits, repetitive behaviour and restricted interests, sensory difficulties, cognitive aptitude, and developmental and medical history. All phenotypic measures will be administered to probands and their first degree relatives who are participating in this study. We will collect the following baseline phenotypes:

1. **The Autism Spectrum Quotient** (AQ)[66]: This is a 50-item questionnaire of broad autistic traits. It measures both social and non-social traits and takes 5-10 minutes to complete. The AQ has excellent internal consistency (Cronbach's alpha > 0.65) and good test-retest reliability ($r > 0.7$)[66]. Finally, versions of the AQ can be completed across a range of ages (18 months +)[67–69] , and is correlated with other measures of autistic traits[70]. This is online at www.autismresearchcentre.com.

2. **The Repetitive Behaviour Scale-Revised**[71] is a comprehensive measure of repetitive and restrictive behaviour in autism. It consists of 43 questions and can be used across a range of ages. It has good internal consistency (0.78-0.91)[71], including good test-retest reliability (r = 0.71)[71].
3. **The Short Sensory Profile**[72] is a 38-item caregiver based questionnaire of sensory difficulties, and takes 5-10 minutes to complete. It has excellent psychometric properties and has been used widely in research.
4. **The Developmental and Medical History Questionnaire**: This includes questions on developmental history including language, communication, and motor profiles; genetic screening tests; antenatal and postnatal complications; co-morbid neuropsychiatric diagnosis; and other co-morbidities relevant to autism such as sleep and GI complications. This has already been designed and is online at www.autismresearchcentre.com
5. **The Short Raven's Matrices Test**[73]: We will use the abbreviated (9-item) Raven's Standard Progressive Matrices Test to quantify non-verbal IQ. This is highly correlated with long form of the test (r = 0.98), and has good psychometric properties. We note the floor effect in this test and that it will not be suitable for individuals with severe learning difficulties. However, we cannot collect in-person IQ information from individuals with severe and profound learning difficulties as this is time-consuming, and may discourage some participants. Rather, we will accept educational attainment and type of school (information provided by the parent) as a proxy measure for severe learning difficulties. This test is online at www.autismresearchcentre.com
6. **Brief demographic data**: This is designed to collect the basic information we need to stratify both cases and controls and includes: biological sex, date of birth, post-code, ethnicity, singleton/multiple birth, handedness, educational attainment, type of school attended, family history of autism, and NHS number or equivalent. These are all part of the brief registration at www.autismresearchcentre.com
7. **Consent**: All 10,000 autistic individuals will be asked to consent for their anonymised data to be shared with collaborators, and to consent to be re-contacted for longitudinal studies and to hear about future studies. We have provided more details of consent and the ethical framework for this study under the "Research involving human participants, human biological material and identifiable data" section in the application.

In addition, we will collect updated annual phenotyping information focusing on co-morbidities and development. This will include standardised brief measures of psychiatric comorbidities (e.g., depression, anxiety, eating disorders, psychosis, etc), medical conditions (e.g., gastrointestinal pain, diabetes, cancer, etc.,), life-style characteristics (e.g., sleep difficulties, alcohol and smoking consumption, exercise, etc.), psychological profiles (e.g., communication skills, motor skills, peer-relationship difficulties, and vulnerability, etc). The Autism Research Centre has developed and used these measures in earlier studies. These phenotypic measures will allow for gene-environment analyses, and for sub-typing risk groups.

In addition to phenotyping, we will also link data to EHRs for probands and first degree relatives to investigate the impact of parental mental health on the phenotype of the proband. EHRs (Primary and Secondary Care) represent a highly valuable resource for gaining deeper insights into genotype-phenotype correlations. Linking EHRs to genetics has been used to stratify psychiatric conditions, to investigate the genetics of stratified subgroups and dimensional domains, and conduct phenome-wide association studies[65,74–76]. We will obtain consent from all participants to link their individual data to EHRs. Primary Care data will be obtained from GP clinical system suppliers such as EMIS, TPP, Microtest and INPS and/or through service providers such as Apollo, for which we have requested funds. We will extract full patient records including encounters, diagnoses, medication, lab results and appointments data. Access to secondary data including Hospital Episode Statistics (HESS), will be made to the Independent Group

Advising on the Release of Data (IGARD) using informed consent. We will request data from HESS, Admitted Patient Care, Diagnostic Imaging Dataset, Critical Care, Outpatients and Accident & Emergency data as well as Mental Health Services Data. We will request information on NHS number, date of birth, post-code, GP practice, and sex from the participants. We will link the cohort to the NHS spine via NHS digital (Personal Demographics Services) or similar services (e.g. eDRIS and SPIRE in Scotland). Where NHS number is not available from participants, we will use the Personal Demographics Services to obtain the NHS number via triangulation. Data from EHRs will be extracted using natural language processing and other methods. We will further collaborate closely with HDR-UK (**Hurles, Co-PI, Cambridge**) for expertise on mining information from the linked medical records.

Ethics will include possibility of recontacting, data sharing, and informed consent. The UK Autism Biobank will share data (individual and summary) with other PIs, but will not share resources due to the finite nature of the resources. Clinically relevant findings (in this Phase, known autism CNVs) will be relayed to the participants through a nominated clinician. See "Research involving human participants, human biological material and identifiable data" for ethics, and "Outputs management and sharing" sections for further details on ethics and data availability, respectively.

2.1.4 Data storage: Identifiable information will be stored in the University of Cambridge Secure Data Hosting Service. The SDHS provides a dedicated network, separated from the production network by a firewall, for storing sensitive personal data and hosting computers involved in its management and analysis.

2.2 GWAS and CNV analysis of autism and autistic traits

2.2.1 Autistic traits: Rationale and phenotypes

Autistic traits represent subclinical manifestations of the autism phenotype, are normally distributed in the general population, and are heritable ($0.4 <$ twin $h^2 < 0.9$)[77]. Autistic traits can be measured using multiple different instruments including the Autism Spectrum Questionnaire-10 (10 item version of the AQ, or AQ-10)[78], the Social and Communication Disorders Checklist[79], and the Social Responsiveness Scale[80]. These measures are moderately to highly correlated with each other[70], and previous GWAS of subjective wellbeing and neuroticism have demonstrated the utility of combining correlated phenotypic measures of the same underlying latent for gene discovery[81]. Further, autistic traits capture greater intrinsic variation in the latent trait than a case-control design and have high genetic correlation with autism[13,45,53]. We will thus complement the case-control GWAS of autism with a GWAS of autistic traits and meta-analyse, an approach similar to the latest GWAS of ADHD[60]. We have approached several collaborators who have measures of autistic traits (**Table 1**), and obtained in-principle approval for this study. In addition, we have obtained in-principle approval from the UK Biobank, the Lifelines Cohort, and the Generation Scotland to phenotype participants using the AQ-10.

Table1: Sample sizes and phenotypes for the autistic traits GWAS

Name	Max N	Phenotype
UK Biobank*	500,000	AQ-10
BLTS	1,5000	AQ-10 + SRS
LifeLines cohort*	5,0000	AQ-10
ALSPAC	4,800	SCDC
Netherlands Longitudinal Twin Register	3,7900	CBCL
CATSS	13,000	A-TAC
RATSS	5,000	AQ
RAINE	1,100	AQ
Radboud University	3,500	AQ (12 item)
Generation Scotland*	20,000	AQ-10

Table 1 above provides the maximum sample size and the phenotype for the cohorts that have been contacted and in principle have agreed to participate.
** indicates cohorts where phenotypes have not been already collected.*
SRS = Social Responsiveness Scale[80]; A-TAC = The Autism-Tics, AD/HD and other Comorbidities inventory (A-TAC)[82]; CBCL = Child Behaviour Checklist[83]; AQ = Autism Spectrum Quotient; AQ-10 and AQ-12 = 10-item and 12-item version of the AQ. Maximum N = 650,300.
We have conservatively estimated that data will be available for a third of participants (N ~ 250,000) based on genotyping, quality control, and participation rate in cohorts where data has not already been collected.

2.2.2 Primary statistical analyses
We will conduct two complementary GWAS analysis: GWAS of autism (case-control GWAS) and GWAS of autistic traits (**Figures 3 and 4**). We estimate a final sample size of 100,000 cases and roughly equal number of controls for the case-control GWAS. All cases from the UK will be genotyped in the UK Biobank axiom array to enable the use of participants in the UK Biobank as population controls. In this phase of the study, we will not genotype parents, but will bank their DNA for future analyses. For both the GWAS, we will conduct common and low-frequency variant GWAS by combining data from multiple different cohorts. In all cohorts, imputation will be conducted using data from the 1,000 Genomes[84], the Haplotype Reference Consortium[85], or the UK-10K reference panel[86]. If data becomes available from larger consortium (e.g. TOPMED, and 100,000 Genomes Project), we will use these panels for imputation. Data will be quality controlled for allele frequency, Hardy Weinberg equilibrium, per SNP and per-individual genotyping rate, and imputation accuracy. Population stratification will be controlled for by including ancestry principal components as covariates or by utilizing a linear mixed-effects model. Replication-by-proxy will be conducted by investigating concordance of effect direction of the lead SNPs from one GWAS in the other GWAS, given the high twin and SNP genetic correlation between autism and autistic traits[45,53,87]. In addition, for both GWAS we will conduct '*in sample replication*' by assessing concordance of effect direction of lead SNPs in all cohorts. We will use a conservative threshold of $P < 5x10^{-9}$ to account for the greater number of statistical tests conducted when investigating low-frequency variants. Statistical power calculations suggest that the final sample has 80% power to identify variants with an odds ratio = 1.04, and calculations suggest that the study will identify approximately 155-165 novel variants (**Section 2.2.4** provides more details). Using multivariate techniques, we will meta-analyse the GWAS of autism and autistic traits, an approach that has been used for ADHD and depression to further increase sample size, identify additional variants, and increase the predictive power of polygenic scores derived from the meta-analyses.

We will conduct fine-mapping analyses to identify potential causal variants, and integrate eQTL, methylation, and Hi-C data from the adult[88] and developing brain[89] to prioritize genes for functional analyses. Integrating data from multiple different genomic functional categories, we will fine-map and prioritize causative variants and genes for downstream functional analysis.

We will conduct CNV association meta-analysis of autism and autistic traits. CNVs will be called from quantile normalized raw probe-set intensity values. Probabilistic CNV calls will be generated from Penn CNV[90], after merging adjacent CNVs with small gaps, and removing samples with more than 200 CNVs, and had a call rate < 96%. Individual CNVs will be excluded if they are covered by <10 probes or had poor density of coverage. We will focus on rare CNVs (population allele frequency < 0.1%) based on the Database of Genomic Variation[91]. Analyses will be conducted at a probe level after translating CNV calls and quality to a probe-level. We will include sex, genetic principle components, and age as covariates. Meta-analyses across cohorts will be conducted based on previously described procedures[92].

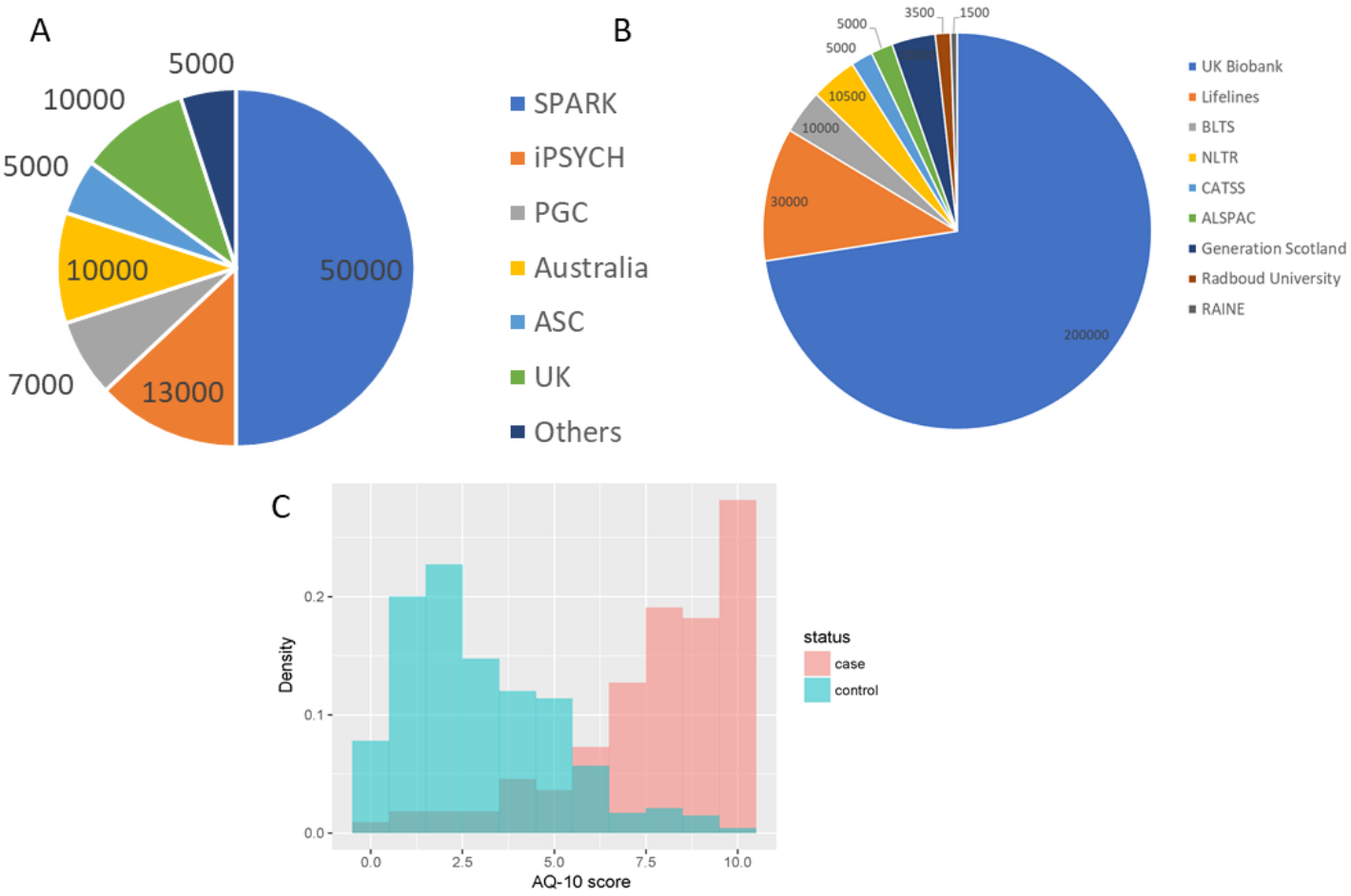
Following the primary-analysis we will conduct the following analyses:

a. To understand the neural underpinnings of the autism GWAS, we will investigate enrichment[93,94] in genes highly expressed in specific human[95–99] and mouse neuronal cell types [100–102], genes with specific expression in different brain tissues[88,103], and genetic pathways.
b. To understand if the autism GWAS is enriched for specific functional categories, we will investigate enrichment in genomic regions defined by functional properties (e.g. open chromatin regions[94], conserved regions, genes intolerant to loss-of-function mutations[104], transcriptionally dysregulated genes in autism[105,106]).
c. To understand the shared genetics between autism and other phenotypes, we will conduct genetic correlation analyses at both a global[107,108] and local genomic level[109], using phenotypes from the UK Biobank and other large GWAS[110]. We will use Structural Equation Modelling[111] using summary GWAS data to further delineate the effect of mediator phenotypes on autism. To identify shared loci between autism and other neuropsychiatric conditions, we will conduct Bayesian co-localization analyses[112].
d. To identify modifiable risk factors for autism and other co-morbid conditions, we will conduct two-sample summary Mendelian Randomization analyses and Latent Causal Variable[113] analyses.
e. To identify the correlation between autism and normative development, we will investigate how polygenic scores contribute to cross-sectional and longitudinal speech, communication and motor development in Lifelines Cohort (N ~ 3000)[114], and ALSPAC (N ~ 5000)[115]. We will further investigate the role of polygenic scores in the development of typical peer networks and social relationships[116,117] in the Add Health Cohort (N ~ 4,000) [118].
f. To investigate the indirect effects[119] of autistic traits on adaptive behaviour and overall autism symptom severity, we will test if polygenic scores of autism in the untransmitted parental alleles are associated with scores on the Vineland Adaptive Behaviour Scale and ADOS-G[120] scores in the children in the Simon's Simplex Collection (N ~ 2,500 families)[61] and the Autism Genetic Resource Exchange (N ~ 1,500 families)[62].
g. To understand the effects of autism on typical brain development, we will investigate the association of autism polygenic scores in cross-sectional and longitudinal structural and functional brain networks in the ABCD (N = 10,000)[121] and the UK Autism Biobank (N = 100,000)[122].
h. To investigate secondary medical phenotypes associated with autism, we will investigate if polygenic scores for autism (developed excluding the UK Autism Biobank) are associated with clusters of medical phenotypes identified in EHRs in the UK Autism Biobank.
i. To investigate the combined contribution of common and rare-variants with autism severity, we will investigate if polygenic scores for autism excluding the Simon's Simplex, MSSNG, and the

AGRE cohorts, interact with *de novo* LGD variation in genes intolerant to loss-of-function mutations in the three cohorts (SSC, MSSNG, and AGRE) to modify the severity of the condition.

j. To reposition drugs in autism. First, using eQTL data, we will impute transcriptomes for the autism GWAS. This, in combination with drug-induced transcriptome will help identify enriched drug targets[42]. Second, we will generate 'drug pathways' by using data on drug-gene interactions, and test for enrichment of these pathways in the autism GWAS[41].

Figure 3: How the samples will be assembled from existing cohorts and distribution of the AQ-10 scores in 700,000 individuals



3A:Samples included in the case-control autism GWAS (cases only). Of these, samples from the Autism Sequencing Consortium (ASC), iPSYCH, and PGC have been collected and genotyped. Samples collection in SPARK is in progress. New samples will include the UK 10K and the Australian 10K. The remaining 5K will be from other cohorts such as the EU-AIMS, DECODE, 23andMe, and Kaiser Permanente, and ALSPAC. Not all samples from SPARK have been collected (18,000 probands collected so far), and the SPARK team anticipate completion of DNA collection in 18 months.
3B: Samples included in the autistic traits GWAS. Samples will primarily be from the UK Biobank. 3C: Distribution of AQ-10 scores in cases (pink) and controls (turquoise) in 700,000 individuals. The distribution histogram above shows the scores on the Autism Spectrum Quotient-10 (AQ-10) in autistic individuals and the general population in 700,000 individuals. An AQ-10 score greater than 6 is an excellent discriminator of case-control diagnosis (positive predictive power = 0.85).

2.2.3 Secondary analyses:

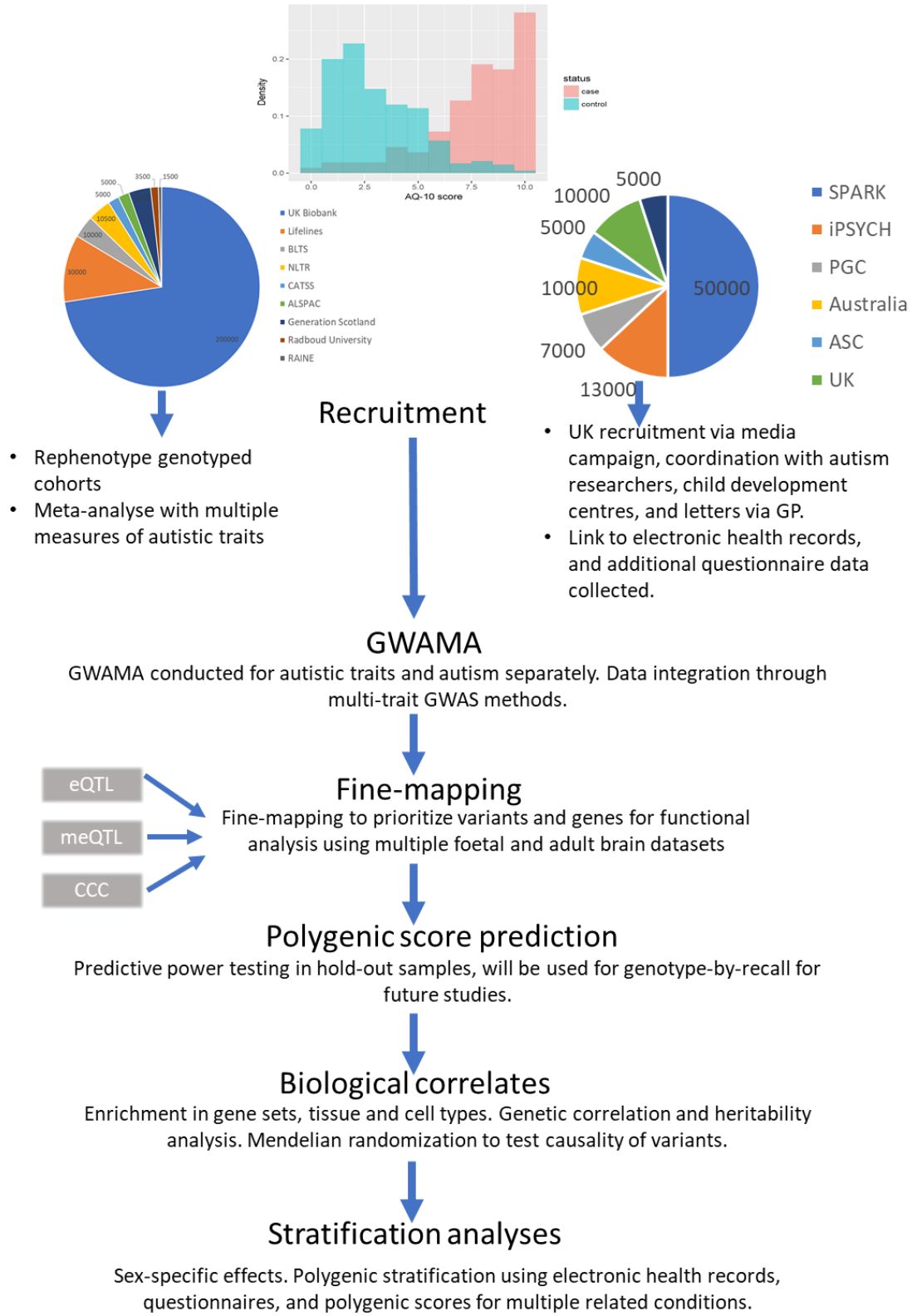
The unprecedented sample size will also allow us to conduct several secondary GWAS analyses to investigate sources of heterogeneity **(Figure 5)**. We will conduct the following additional analyses:

1. **Sex-interaction and stratification analyses (Figure 5A)**: Given the sex differences in prevalence of autism (between 3:1, males to females)[5,6], it is important to investigate the sex-specific effects of common variants in autism. We will conduct a sex-interaction GWAS of both autism and autistic traits to identify loci with differential effects based on sex. As several downstream analyses are not developed for an interaction model, we will further conduct sex-stratified analyses to identify sex-specific genes, pathways, and enriched tissues, and sex-specific genetic correlations. Sex information will be available for all cohorts included in the primary analyses.

2. **Social and non-social autistic traits GWAS (Figure 5B)**: Several lines of research have demonstrated that social and non-social domains of autism are phenotypically and genetically dissociable[29,123,124]. A few studies have investigated the genetic architecture of these domains in small cohorts of autistic individuals[57,125], but results have been limited by the small sample size and low statistical power. We will conduct GWAS of social and non-social domains in the autism cohorts, and, in parallel, GWAS of social and non-social autistic traits in the general population by dividing items in the AQ-10 and other autistic trait measures into those that capture social and non-social traits. We will additionally investigate if these domains are genetically dissociable using genetic correlation, if polygenic scores predict different outcomes, and enrichment in different genes, tissues, and cell types. Finally, using transcriptomic data from the developing brain, we will investigate shared and distinct enrichment of polygenic signals of the social and non-social GWAS in different spatio-temporal gene expression modules. Using EHRs we will investigate if polygenic scores from the social and non-social domains are associated with different co-morbidities. We expect domain specific information from 3 autism cohorts (N ~ 60K, Simon's Simplex, UK Autism Biobank, and AGRE), and social and non-social autistic traits information on all participants included in the autistic traits GWAS (N ~ 250K).

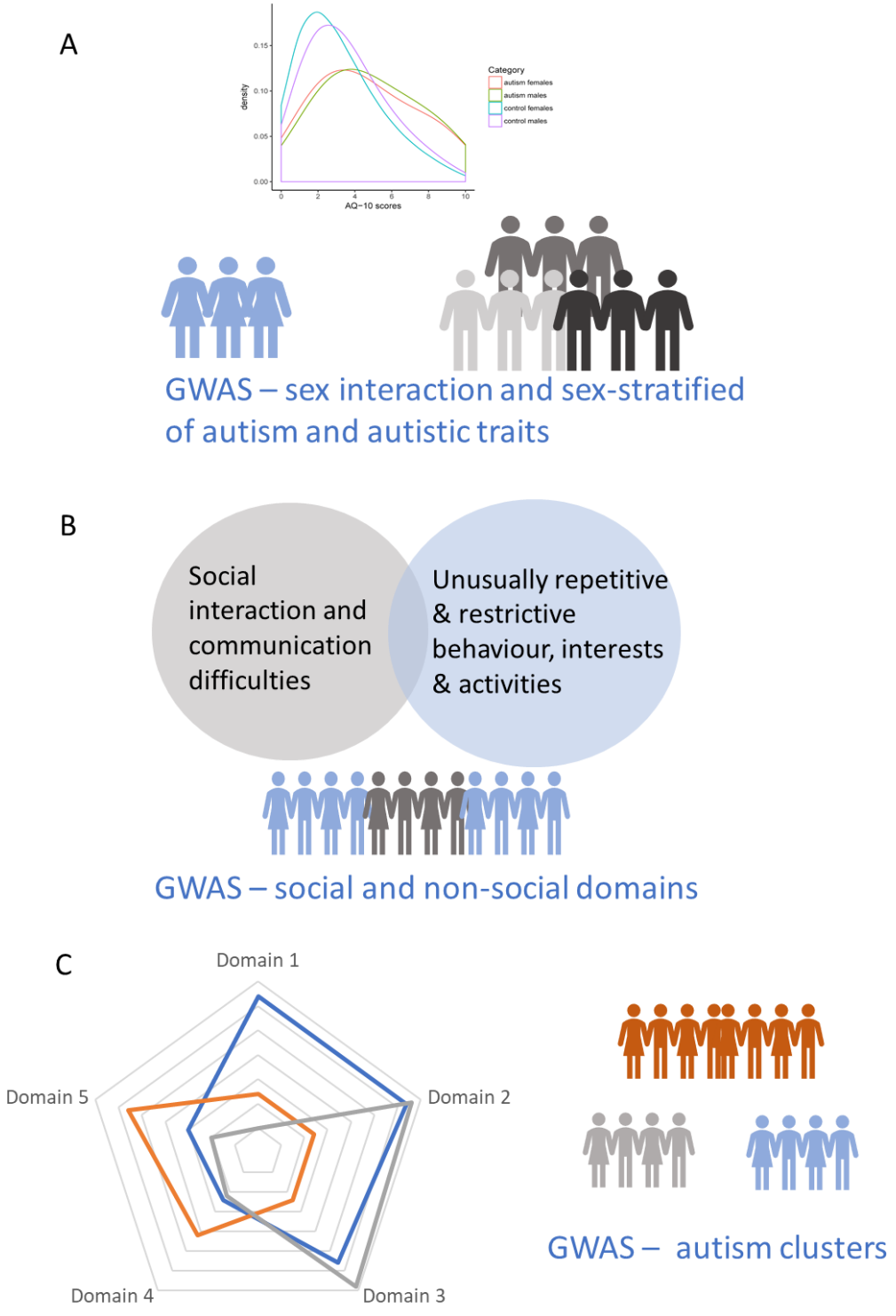
3. **GWAS of autism subgroups (Figure 5C)**: To better understand heterogeneity within the autism spectrum, we will conduct phenotypic clustering (K-means, Finite Mixture Modelling) within the autism spectrum (N > 60K - Simon's Simplex, AGRE, SPARK, and the UK Autism Biobank), and conduct GWAS of autism in the clustered cohorts. Genetic correlation analyses will be conducted to investigate if the polygenic architecture is dissimilar across the clusters. Using EHRs in the UK cohort, we will investigate if these clusters are enriched for different co-morbidities.

Figure 4: Analytical plan of the GWAS analyses



Flowchart of the analytical plan for the study

Figure 5: Schematic diagram of secondary analyses



Schematic diagram of secondary analyses.
5A:Sex-stratified and sex-interaction GWAS analysis to investigate sex-specific effects. The kernel density diagram shows differences in AQ-10 scores between autistic and non-autistic males and females in 700,000 individuals (Greenberg, Warrier et al., submitted).
5B: Heterogeneity of social and non-social domains in the autism GWAS, and in parallel, social and non-social autistic traits, followed by meta-analysis.
Figure 5C: GWAS of clusters of autistic individuals based on clustering of phenotype scores.

2.2.4: Power calculations **(Figure 6)**:
Power calculations for the GWAS analyses were conducted using Genetic Power Calculator[126]. The primary autism GWAS will have 80% power to detect a variant with Odds Ratio = 1.04. The autistic traits GWAS will have 80% power to detect a variant with a variance explained (R^2) of 0.01% roughly corresponds to an odds ratio of 1.02 for common variants (minor allele frequency > 0.05). For polygenic scores, power calculations were conducted in line with the theory laid out elsewhere[127]. Statistical power graph for polygenic score regression analysis is dependent on the heritability of the testing and the training phenotypes, and the genetic correlations between the two phenotypes, which is equivalent to the square root of the variance explained (R^2) by polygenic scores in an infinitely large sample. Phenotype 2 represents one of the various testing phenotypes (developmental trajectories, imaging metrics etc). Power calculations have been conducted separately for the autism and the autistic traits GWAS. At a SNP heritability > 0.15 for the autistic traits GWAS, the Lifelines cohort will have sufficient statistical power to detect an effect that's equivalent to $r_g > 0.3$ between the two phenotypes in an infinitely large sample. Power calculations for investigating the combined contribution of common and rare variants were conducted using G*Power (http://www.gpower.hhu.de/). At a sample size of 3,000, the study will have 80% power to significantly ($P < 0.05$) identify a predictor that explains 0.25% of the variance ($R^2 = 0.0025$). Using the entire SPARK and MSSNG dataset (N = 5,000), the study will have 80% power to detect PTV-scores that explain 0.15% of the total variance in the phenotype. At an N ~ 30,000, the study will have 80% power to identify a predictor that explains 0.025% of the variance. Power calculations were done using pilot analysis in the Simon's Simplex Collection (N = 2,221), using a linear regression.

We estimate the case-control GWAS will identify between 155-165 new loci. This estimate was arrived at using two methods. In the first method, we calculated the slope of the loci discovered to the total number of autism cases included in the GWAS using existing GWAS. Extrapolating it to N = 100,000 cases yielded 165 new loci. This follows from observations from other GWAS studies that there is a largely linear relationship between sample size and loci identified in GWAS. In the second method[128], we accounted for population prevalence of autism under the hypothesis that cases in rarer conditions will have a greater liability, and thus, will have a greater difference in liabilities between cases and unselected controls. Using a population prevalence of 1% for autism, 100,000 cases and twice as many controls, we estimate that the study will discover 155 new loci.

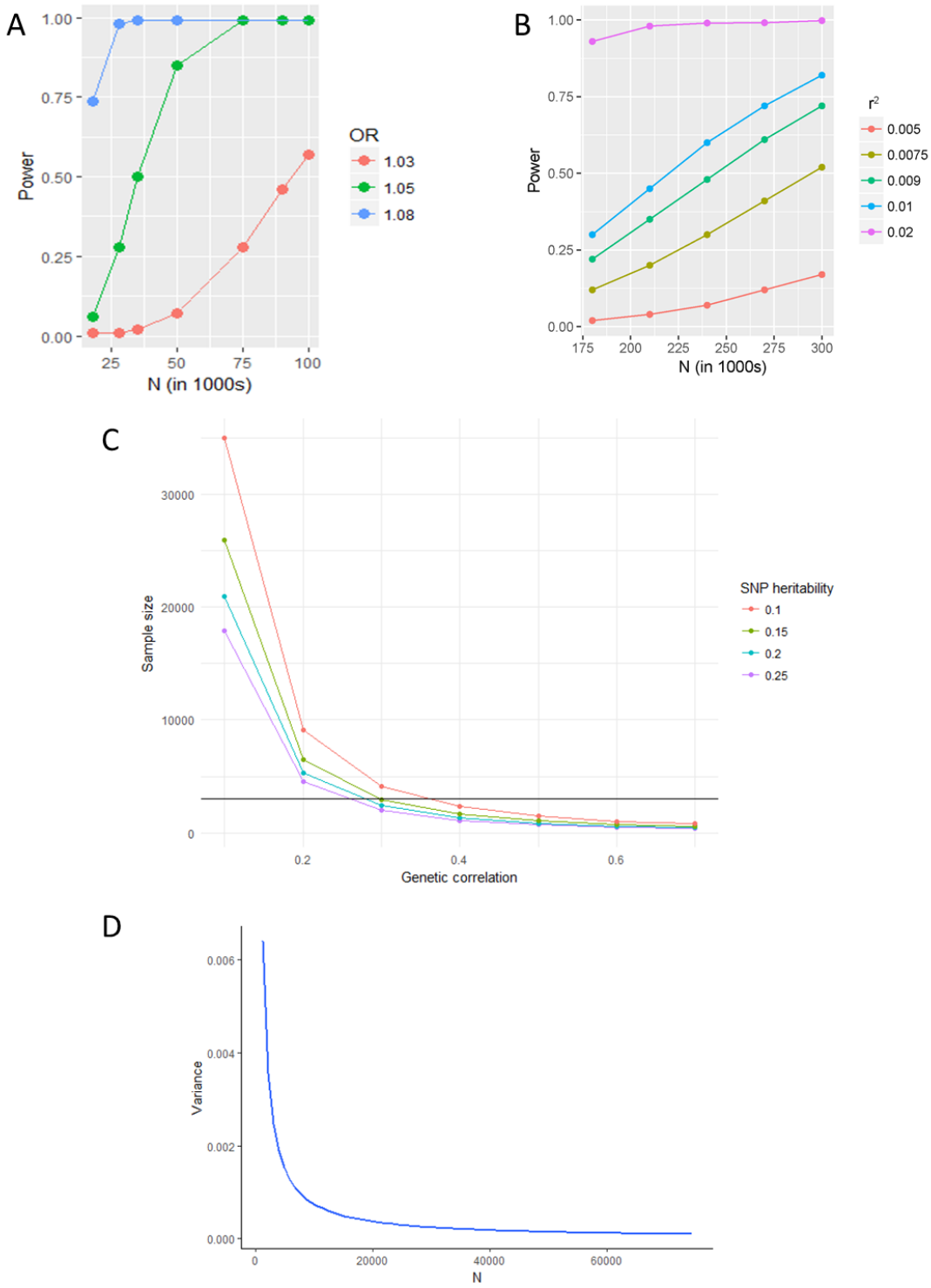
3. Impact, novelty, and expected outcomes:
This will be the largest GWAS of autism and autistic traits to date, and will accelerate the identification of genetic loci and whole-genome genotyping of the entire cohort. We predict that this study will identify at between 155-165 new significant genetic loci and genes, assuming a similar slope of discovery to the current autism GWAS studies. Meta-analysis using the GWAS of autistic traits will increase this number, but this depends on the genetic correlation between the autism GWAS and the autistic traits GWAS. This will accelerate biomarker discovery and stratification using deep phenotyping, electronic health records, and polygenic scores, and will provide a central platform for autism research in the UK, accelerating discoveries by integrating data from genetics. This study will generate well-powered polygenic scores for autism and we will investigate the predictive power of the polygenic scores in independent samples. This can be used to inform diagnosis, particularly the impact on patient management in the UK cohort (specifically, using genetic scores to prioritize therapeutic strategies for autism and related co-morbidities), conduct recall-by-genotype studies, and investigate gene-by-environment effects. We will investigate the biological correlates of autism: which tissues, gene-sets, cell types, and developmental periods are enriched for common genetic risk for autism. We will further investigate heritability across subtypes, sex-specific effects, and effects of social and non-social domains of autism. In addition, we anticipate the following outputs:

1. **3+ open access scientific articles** in high impact journals. This will include the GWAS of autism, the GWAS of autistic traits, CNV analyses, effects of polygenic scores, and subgroup and subdomain analyses of autism. This is a highly conservative estimate and we anticipate more publications from the core team, collaborators and other PIs who request access to the data, as is evident from the several publications from other, smaller autism cohorts such as Simon's Simplex Collection and Autism Genetic Resource Exchange. Specifically, the GWAS of autism and the GWAS of autistic traits will be the largest GWAS of these two phenotypes upon publication.
2. **4+ poster presentations/talks** of the findings in top genetics, psychiatry, and autism conferences such as ASHG, ESHG, INSAR, WCPG, and SfN.
3. **Access to data and code.** This includes both summary statistics of the full GWAS which will be made openly available to all researchers, and individual level data of the UK Autism Biobank, which will be available after a short application process.
4. Public dissemination of principle findings through **social media, press release, public engagement talks** to non-specialists.

4. Timeline

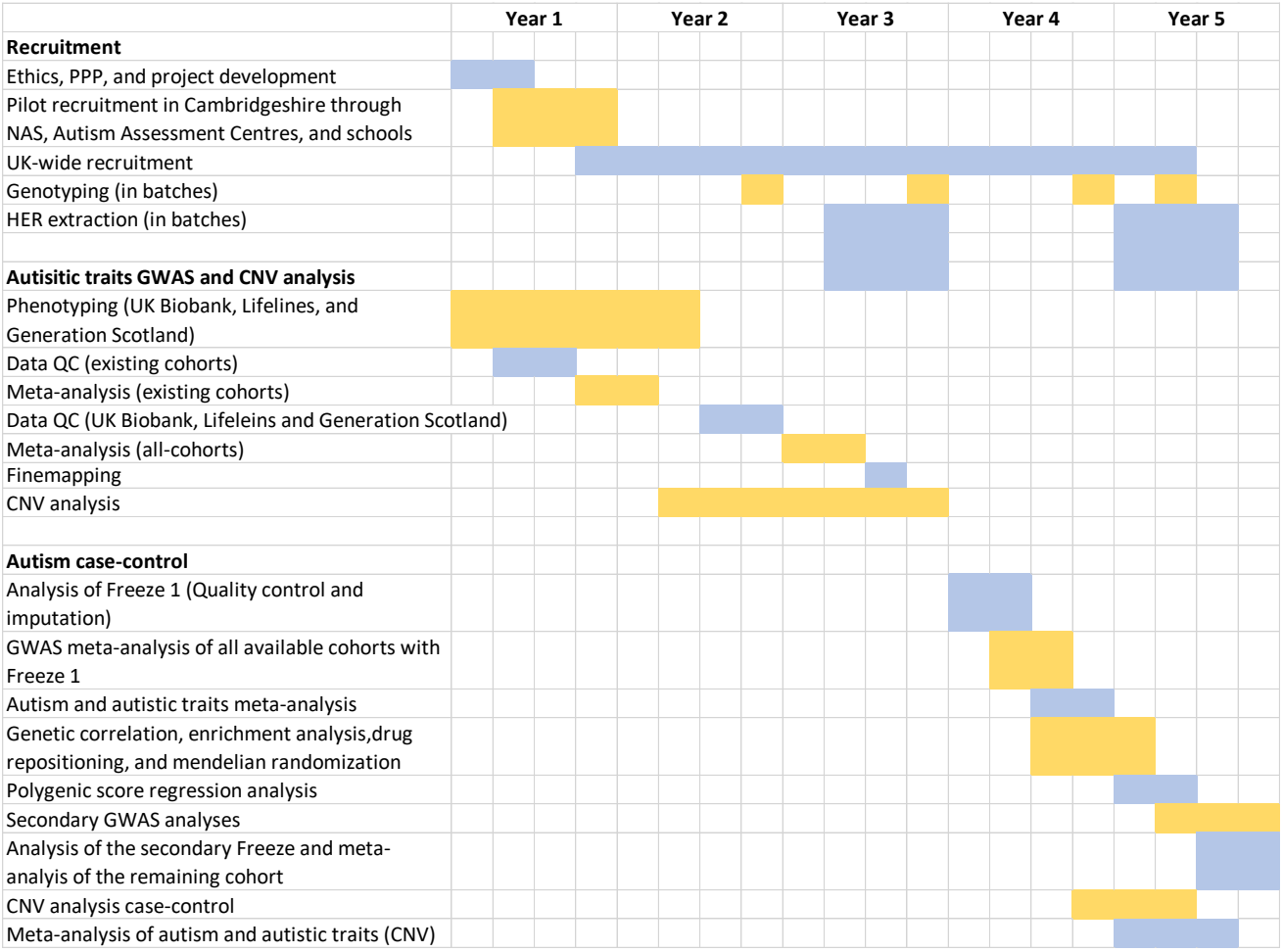
Figure 7 is a Gantt chart that provides the timeline of the proposed activities.

Figure 6: Power calculations



A: Statistical power of the autism GWAS at various OR and sample sizes. B: Statistical power of the autistic traits GWAS at various r^2 and sample sizes. C: Statistical power for the polygenic score regression analyses at various SNP heritabilities (coloured lines), sample sizes, and genetic correlations between the primary phenotype (autism or autistic traits), and the various secondary phenotypes. D: Statistical power graph for the standardized PTV-score on autism phenotypes.

Figure 7: Gantt Chart with specific timelines



5. Future studies:

In addition to accelerating gene discovery in autism, this will create the largest autism biobank in the UK. All participants will be registered in a separate database and their biological material stored for future analysis. DNA will be collected from immediate (1st degree) family members of autistic individuals. The current phase of the study will genotype all probands and contribute to international efforts for variant discovery and dissecting heterogeneity. In the future, this resource will allow for:

1. WGS and WES: While the current phase of the study investigates common and low-frequency variants alongside structural variants, future phases will investigate the contribution of rare variants, in both exonic and non-exonic regions by conducting WES and WGS respectively of probands and their immediate parents.
2. Recall-by-genotype studies: This will allow to conduct targeted phenotyping based on genetic risk scores (single or multi-variant) at extremes to test specific hypothesis. This will be particularly useful to understand the genetic and environmental contributions to modifiable co-morbidities such as anxiety, depression and sleep disorders, all of which are common in autistic individuals.
3. Longitudinal phenotype trajectories: Continued phenotyping using self-report and EHRs will provide an optimal resource to understand the genetic contributors to mental-health trajectories. Such a resource is not available anywhere. The significant number of participants that will recruited at a young age through child autism assessment centres and specific schools will be invaluable for assessing trajectories of mental and physical health conditions in childhood, adolescence and early adulthood.

4. Familial effects on autism severity and mental health: It is currently unclear how familial environment contributes to mental health condition in autistic individuals. Genotyping parents and sibling to investigate will allow for investigating indirect effects[119] of polygenic scores on the probands.
5. Multi-modal investigation of autism: Finally, future studies will combine genetic with information derived from other methods such as cognitive testing, neuroimaging, transcriptomics, and methylation to better understand the contributors to autism.

References

1. Tick B, Bolton PF, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: A meta-analysis of twin studies. *J Child Psychol Psychiatry Allied Discip*. 2016;57(5):585-595. doi:10.1111/jcpp.12499

2. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The Heritability of Autism Spectrum Disorder. *JAMA*. 2017;318(12):1182. doi:10.1001/jama.2017.12141

3. Wang K, Gaitsch H, Poon H, Cox NJ, Rzhetsky A. Classification of common human diseases derived from shared genetic and environmental determinants. *Nat Genet*. 2017;49(9):1319-1325. doi:10.1038/ng.3931

4. Halladay AK, Bishop S, Constantino JN, et al. Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. *Mol Autism*. 2015;6(1):36. doi:10.1186/s13229-015-0019-y

5. Loomes R, Hull L, Mandy WPL. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J Am Acad Child Adolesc Psychiatry*. 2017;56(6):466-474. doi:10.1016/j.jaac.2017.03.013

6. Lai M-C, Lombardo M V., Baron-Cohen S. Autism. *Lancet*. September 2013. doi:10.1016/S0140-6736(13)61539-1

7. Baio J, Wiggins L, Christensen DL, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveill Summ*. 2018;67(6):1-23. doi:10.15585/mmwr.ss6706a1

8. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry*. 2008;47(8):921-929. doi:10.1097/CHI.0b013e318179964f

9. Hirvikoski T, Mittendorfer-Rutz E, Boman M, Larsson H, Lichtenstein P, Bölte S. Premature mortality in autism spectrum disorder. *Br J Psychiatry*. 2016;208(3):232-238. doi:10.1192/bjp.bp.114.160192

10. Cassidy S, Bradley P, Robinson J, Allison C, McHugh M, Baron-Cohen S. Suicidal ideation and suicide plans or attempts in adults with Asperger's syndrome attending a specialist diagnostic clinic: a clinical cohort study. *The Lancet Psychiatry*. 2014;1(2):142-147. doi:10.1016/S2215-0366(14)70248-2

11. Buescher AVS, Cidav Z, Knapp M, Mandell DS. Costs of Autism Spectrum Disorders in the United Kingdom and the United States. *JAMA Pediatr*. 2014;168(8):721. doi:10.1001/jamapediatrics.2014.210

12. Ronald A, Happé F, Price TS, Baron-Cohen S, Plomin R. Phenotypic and genetic overlap between autistic traits at the extremes of the general population. *J Am Acad Child Adolesc Psychiatry*. 2006;45(10):1206-1214. doi:10.1097/01.chi.0000230165.54117.41

13. Robinson EB, Koenen KC, McCormick MC, et al. Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Arch Gen Psychiatry*. 2011;68(11):1113-1121. doi:10.1001/archgenpsychiatry.2011.119

14. Krishnan A, Zhang R, Yao V, et al. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat Neurosci*. 2016;19(11):1454-1462. doi:10.1038/nn.4353

15. He X, Sanders SJ, Liu L, et al. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. Williams SM, ed. *PLoS Genet*. 2013;9(8):e1003671. doi:10.1371/journal.pgen.1003671

16. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet*. 2014;15(2):133-141. doi:10.1038/nrg3585

17. Gratten J, Wray NR, Peyrot WJ, McGrath JJ, Visscher PM, Goddard ME. Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations. *Nat Genet*. 2016;48(7):718-724. doi:10.1038/ng.3577

18. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*. 2014;515(7526):216-221. doi:10.1038/nature13908

19. Sanders SJ, He X, Willsey AJ, et al. Insights into Autism Spectrum Disorder genomic architecture and biology from 71 risk loci. *Neuron*. 2015;87(6):1215-1233. doi:10.1016/j.neuron.2015.09.016

20. Gaugler T, Klei LL, Sanders SJ, et al. Most genetic risk for autism resides with common variation. *Nat Genet*. 2014;46(8):881-885. doi:10.1038/ng.3039

21. Klei LL, Sanders SJ, Murtha MT, et al. Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism*. 2012;3(1):9. doi:10.1186/2040-2392-3-9

22. Lee SH, Ripke S, Neale BM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45(9):984-994. doi:10.1038/ng.2711

23. Grove J, Ripke S, Als TD, et al. Common risk variants identified in autism spectrum disorder. *bioRxiv*. November 2017:224774. doi:10.1101/224774

24. Bernier R, Golzio C, Xiong B, et al. Disruptive CHD8 mutations define a subtype of autism aarly in development. *Cell*. 2014;158(2):263-276. doi:10.1016/j.cell.2014.06.017

25. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat Genet*. 2014;46(4):380-384. doi:10.1038/ng.2899

26. Frazier TW, Embacher R, Tilot AK, Koenig K, Mester J, Eng C. Molecular and phenotypic abnormalities in individuals with germline heterozygous PTEN mutations and autism. *Mol Psychiatry*. 2015;20(9):1132-1138. doi:10.1038/mp.2014.125

27. Warrier V, the 23andMe Research team, Bourgeron T, Baron-Cohen S. Genome-wide association study of social relationship satisfaction: significant loci and correlations with psychiatric conditions. *bioRxiv*. October 2017:196071. doi:10.1101/196071

28. Warrier V, Toro R, Chakrabarti B, et al. Genome-wide analyses of self-reported empathy: correlations with autism, schizophrenia, and anorexia nervosa. *bioRxiv*. November 2017:050682. doi:10.1101/050682

29. Warrier V, Toro R, Chakrabarti B, et al. Systemizing is genetically correlated with autism and is genetically distinct from social autistic traits. *bioRxiv*. December 2017:228254. doi:10.1101/228254

30. Vorstman JAS, Spooren W, Persico AM, et al. Using genetic findings in autism for the development of new pharmaceutical compounds. *Psychopharmacology (Berl)*. 2014;231(6):1063-1078. doi:10.1007/s00213-013-3334-z

31. Vorstman JAS, Parr JR, Moreno-De-Luca D, Anney RJL, Nurnberger Jr JI, Hallmayer JF. Autism genetics: opportunities and challenges for clinical translation. *Nat Rev Genet*. 2017;18(6):362-376. doi:10.1038/nrg.2017.4

32. Duffney LJ, Zhong P, Wei J, et al. Autism-like deficits in Shank3-deficient mice are rescued by targeting actin regulators. *Cell Rep*. 2015;11(9):1400-1413. doi:10.1016/j.celrep.2015.04.064

33. Mei Y, Monteiro P, Zhou Y, et al. Adult restoration of Shank3 expression rescues selective autistic-like phenotypes. *Nature*. 2016;530(7591):481-484. doi:10.1038/nature16971

34. Sztainberg Y, Chen H, Swann JW, et al. Reversal of phenotypes in MECP2 duplication mice using genetic rescue or antisense oligonucleotides. *Nature*. 2015;528(7580):123-126. doi:10.1038/nature16159

35. Niemi ME, Martin HC, Rice DL, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *bioRxiv*. May 2018:309070. doi:10.1101/309070

36. Weiner DJ, Wigdor EM, Ripke S, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet*. May 2017. doi:10.1038/ng.3863

37. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet*. February 2018:1. doi:10.1038/s41588-018-0059-2

38. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681. doi:10.1038/s41588-018-0090-3

39. Stahl E, Breen G, Forstner A, et al. Genomewide association study identifies 30 loci associated with bipolar disorder. *bioRxiv*. January 2018:173062. doi:10.1101/173062

40. Skene NG, Bryois J, Bakken TE, et al. Genetic identification of brain cell types underlying schizophrenia. *Nat Genet*. 2018;50(6):825-833. doi:10.1038/s41588-018-0129-5

41. Gaspar HA, Breen G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci Rep*. 2017;7(1):12460. doi:10.1038/s41598-017-12325-3

42. So H-C, Chau CK-L, Chiu W-T, et al. Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat Neurosci*. 2017;20(10):1342-1349. doi:10.1038/nn.4618

43. Ruderfer DM, Ripke S, McQuillin A, et al. Genomic dissection of bipolar disorder and schizophrenia, including 28 Subphenotypes. *Cell*. 2018;173(7):1705-1715.e16. doi:10.1016/j.cell.2018.05.046

44. Allardyce J, Leonenko G, Hamshere M, et al. Association between schizophrenia-related polygenic liability and the occurrence and level of mood-incongruent psychotic symptoms in bipolar disorder. *JAMA Psychiatry*. 2018;75(1):28. doi:10.1001/jamapsychiatry.2017.3485

45. St Pourcain B, Robinson EB, Anttila V, et al. ASD and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties. *Mol Psychiatry*. January 2017. doi:10.1038/mp.2016.198

46. Hubbard L, Tansey KE, Rai D, et al. Evidence of common genetic overlap between schizophrenia and cognition. *Schizophr Bull*. 2016;42(3):832-842. doi:10.1093/schbul/sbv168

47. Riglin L, Collishaw S, Richards A, et al. Schizophrenia risk alleles and neurodevelopmental outcomes in childhood: a population-based cohort study. *The lancet Psychiatry*. 2017;4(1):57-62. doi:10.1016/S2215-0366(16)30406-0

48. Jones HJ, Stergiakouli E, Tansey KE, et al. Phenotypic manifestation of genetic risk for schizophrenia during adolescence in the general population. *JAMA Psychiatry*. 2016;73(3):221. doi:10.1001/jamapsychiatry.2015.3058

49. Smoller JW, Andreassen OA, Edenberg HJ, Faraone S V., Glatt SJ, Kendler KS. Psychiatric genetics and the structure of psychopathology. *Mol Psychiatry*. January 2018. doi:10.1038/s41380-017-0010-4

50. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism*. 2017;8(1):21. doi:10.1186/s13229-017-0137-9

51. Stahl E, Breen G, Forstner A, et al. Genomewide association study identifies 30 loci associated with bipolar disorder. *bioRxiv*. January 2018:173062. doi:10.1101/173062

52. Corbin LJ, Tan VY, Hughes DA, et al. Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference. *Nat Commun*. 2018;9(1):711. doi:10.1038/s41467-018-03109-y

53. Robinson EB, St Pourcain B, Anttila V, et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet*. 2016;48(5):552-555. doi:10.1038/ng.3529

54. Robinson EB, Samocha KE, Kosmicki JA, et al. Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc Natl Acad Sci*.

2014;111(42):15161-15165. doi:10.1073/pnas.1409204111

55. Pinto D, Delaby E, Merico D, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet*. 2014;94(5):677-694. doi:10.1016/j.ajhg.2014.03.018

56. Chaste P, Klei LL, Sanders SJ, et al. A genome-wide association study of autism using the Simons Simplex Collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? *Biol Psychiatry*. 2015;77(9):775-784. doi:10.1016/j.biopsych.2014.09.017

57. Cantor RM, Navarro L, Won H, Walker RL, Lowe JK, Geschwind DH. ASD restricted and repetitive behaviors associated at 17q21.33: genes prioritized by expression in fetal brains. *Mol Psychiatry*. 2018;23(4):993-1000. doi:10.1038/mp.2017.114

58. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007;316(5823):445-449. doi:10.1126/science.1138659

59. Sanders SJ, Neale BM, Huang H, et al. Whole Genome Sequencing in Psychiatric Disorders: the WGSPD Consortium. *bioRxiv*. July 2017:160499. doi:10.1101/160499

60. Demontis D, Walters RK, Martin J, et al. Discovery of the first genome-wide significant risk loci for ADHD. *bioRxiv*. 2017.

61. Fischbach GD, Lord C. The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron*. 2010;68(2):192-195. doi:10.1016/J.NEURON.2010.10.006

62. Geschwind DH, Sowinski J, Lord C, et al. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet*. 2001;69(2):463-466. doi:10.1086/321292

63. The SPARK Consortium. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron*. 2018;97(3):488-493. doi:10.1016/j.neuron.2018.01.015

64. Thornton LM, Munn-Chernoff MA, Baker JH, et al. The Anorexia Nervosa Genetics Initiative: Overview and methods. *bioRxiv*. December 2017:234013. doi:10.1101/234013

65. Hall LS, Adams MJ, Arnau-Soler A, et al. Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Transl Psychiatry*. 2018;8(1):9. doi:10.1038/s41398-017-0034-1

66. Baron-Cohen S, Wheelwright SJ, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists

and mathematicians. *J Autism Dev Disord*. 2001;31(1):5-17.

67. Allison C, Baron-Cohen S, Wheelwright S, et al. The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): A Normally Distributed Quantitative Measure of Autistic Traits at 18–24 Months of Age: Preliminary Report. *J Autism Dev Disord*. 2008;38(8):1414-1425. doi:10.1007/s10803-007-0509-7

68. Baron-Cohen S, Hoekstra RA, Knickmeyer R, Wheelwright S. The Autism-Spectrum Quotient (AQ)—Adolescent Version. *J Autism Dev Disord*. 2006;36(3):343-350. doi:10.1007/s10803-006-0073-6

69. Auyeung B, Baron-Cohen S, Wheelwright S, Allison C. The Autism Spectrum Quotient: Children's Version (AQ-Child). *J Autism Dev Disord*. 2008;38(7):1230-1240. doi:10.1007/s10803-007-0504-z

70. Charman T, Loth E, Tillmann J, et al. The EU-AIMS Longitudinal European Autism Project (LEAP): clinical characterisation. *Mol Autism*. 2017;8(1):27. doi:10.1186/s13229-017-0145-9

71. Lam KSL, Aman MG. The Repetitive Behavior Scale-Revised: Independent validation in individuals with Autism Spectrum Disorders. *J Autism Dev Disord*. 2007;37(5):855-866. doi:10.1007/s10803-006-0213-z

72. Tomchek SD, Dunn W. Sensory processing in children with and without autism: A comparative study using the Short Sensory Profile. *Am J Occup Ther*. 2007;61(2):190-200. doi:10.5014/ajot.61.2.190

73. Bilker WB, Hansen JA, Brensinger CM, Richard J, Gur RE, Gur RC. Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*. 2012;19(3):354-369. doi:10.1177/1073191112446655

74. Howard DM, Clarke T-K, Adams MJ, et al. The stratification of major depressive disorder into genetic subgroups. *bioRxiv*. May 2017:134601. doi:10.1101/134601

75. McCoy TH, Castro VM, Hart KL, et al. Genome-wide association study of dimensional psychopathology using electronic health records. *Biol Psychiatry*. 2018;83(12):1005-1011. doi:10.1016/j.biopsych.2017.12.004

76. Prieto ML, Ryu E, Jenkins GD, et al. Leveraging electronic health records to study pleiotropic effects on bipolar disorder and medical comorbidities. *Transl Psychiatry*. 2016;6(8):e870-e870. doi:10.1038/tp.2016.138

77. Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: A decade of new twin

studies. *Am J Med Genet Part B Neuropsychiatr Genet*. 2011;156(3):255-274. doi:10.1002/ajmg.b.31159

78. Allison C, Auyeung B, Baron-Cohen S, Bolton PF, Brayne C. Toward brief "Red Flags" for autism screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist for Autism in toddlers in 1,000 cases and 3,000 controls [corrected]. *J Am Acad Child Adolesc Psychiatry*. 2012;51(2):202-212.e7. doi:10.1016/j.jaac.2011.11.003

79. Skuse DH, Mandy WPL, Scourfield J. Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *Br J Psychiatry*. 2005;187(6):568-572. doi:10.1192/bjp.187.6.568

80. Constantino JN, Davis SA, Todd RD, et al. Validation of a brief quantitative measure of autistic traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *J Autism Dev Disord*. 2003;33(4):427-433. doi:10.1023/A:1025014929212

81. Okbay A, Baselmans BML, De Neve J-E, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet*. 2016;48(6):624-633. doi:10.1038/ng.3552

82. Larson T, Anckarsäter H, Gillberg C, et al. The autism--tics, AD/HD and other comorbidities inventory (A-TAC): further validation of a telephone interview for epidemiological research. *BMC Psychiatry*. 2010;10:1. doi:10.1186/1471-244X-10-1

83. Mazefsky CA, Anderson R, Conner CM, Minshew N. Child Behavior Checklist scores for school-aged children with Autism: Preliminary evidence of patterns suggesting the need for referral. *J Psychopathol Behav Assess*. 2011;33(1):31-37. doi:10.1007/s10862-010-9198-1

84. Gibbs RA, Boerwinkle E, Doddapaneni H, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

85. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

86. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun*. 2015;6(1):8111. doi:10.1038/ncomms9111

87. Robinson EB, Koenen KC, McCormick MC, et al. Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Arch Gen Psychiatry*. 2011;68(11):1113. doi:10.1001/archgenpsychiatry.2011.119

88. Ardlie KG, Deluca DS, Segre A V., et al. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (80-).* 2015;348(6235):648-660. doi:10.1126/science.1262110

89. Won H, de la Torre-Ubieta L, Stein JL, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. 2016;538(7626):523-527. doi:10.1038/nature19847

90. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007;17(11):1665-1674. doi:10.1101/gr.6861907

91. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*. 2014;42(Database issue):D986-D992. doi:10.1093/nar/gkt958

92. Macé A, Tuke MA, Deelen P, et al. CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits. *Nat Commun*. 2017;8(1):744. doi:10.1038/s41467-017-00556-x

93. 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. doi:10.1371/journal.pcbi.1004219

94. Finucane HK, Bulik-Sullivan BK, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235. doi:10.1038/ng.3404

95. Lake BB, Ai R, Kaeser GE, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science (80-).* 2016;352(6293):1586-1590. doi:10.1126/science.aaf1204

96. Lake BB, Chen S, Sos BC, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol*. 2017;36(1):70-80. doi:10.1038/nbt.4038

97. Nowakowski TJ, Bhaduri A, Pollen AA, et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science (80-).* 2017;358(6368):1318-1323. doi:10.1126/science.aap8809

98. Zhong S, Zhang S, Fan X, et al. A single-cell RNA-seq survey of the developmental landscape of the human prefrontal cortex. *Nature*. 2018;555(7697):524-528. doi:10.1038/nature25980

99. Ecker JR, Geschwind DH, Kriegstein AR, et al. The BRAIN Initiative Cell Census Consortium:

Lessons learned toward generating a comprehensive brain cell atlas. 2017. doi:10.1016/j.neuron.2017.10.007

100. Romanov RA, Zeisel A, Bakker J, et al. Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat Neurosci*. 2017;20(2):176-188. doi:10.1038/nn.4462

101. Zeisel A, Muñoz-Manchado AB, Codeluppi S, et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*. 2015;347(6226):1138-1142. doi:10.1126/science.aaa1934

102. Tasic B, Menon V, Nguyen TN, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci*. 2016;19(2):335-346. doi:10.1038/nn.4216

103. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653

104. Lek M, Karczewski KJ, Minikel E V., et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291. doi:10.1038/nature19057

105. Parikshak NN, Swarup V, Belgard TG, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*. 2016;540(7633):423-427. doi:10.1038/nature20612

106. Parikshak NN, Luo R, Zhang A, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. 2013;155(5):1008-1021. doi:10.1016/j.cell.2013.10.031

107. Bulik-Sullivan BK, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406

108. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295. doi:10.1038/ng.3211

109. Shi H, Mancuso N, Spendlove S, Pasaniuc B. Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *Am J Hum Genet*. 2017;101(5):737-751. doi:10.1016/j.ajhg.2017.09.022

110. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. *LD Hub: A Centralized Database and Web Interface to Perform LD Score Regression That Maximizes the Potential of Summary Level GWAS Data for SNP Heritability and Genetic Correlation Analysis*. Vol 33. Cold Spring Harbor

Labs Journals; 2016. doi:10.1093/bioinformatics/btw613

111. Grotzinger AD, Rhemtulla M, Vlaming R de, et al. Genomic SEM Provides Insights into the Multivariate Genetic Architecture of Complex Traits. *bioRxiv*. April 2018:305029. doi:10.1101/305029

112. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016;48(7):709-717. doi:10.1038/ng.3570

113. O'Connor L, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *doi.org*. October 2017:205435. doi:10.1101/205435

114. Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2015;44(4):1172-1180. doi:10.1093/ije/dyu229

115. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The "Children of the 90s"—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013;42(1):111-127. doi:10.1093/ije/dys064

116. Rotheram-Fuller E, Kasari C, Chamberlain B, Locke J. Social involvement of children with autism spectrum disorders in elementary school classrooms. *J Child Psychol Psychiatry*. 2010;51(11):1227-1234. doi:10.1111/j.1469-7610.2010.02289.x

117. Kasari C, Locke J, Gulsrud A, Rotheram-Fuller E. Social Networks and Friendships at School: Comparing Children With and Without ASD. *J Autism Dev Disord*. 2011;41(5):533-544. doi:10.1007/s10803-010-1076-x

118. Resnick MD, Bearman PS, Blum RW, et al. Protecting adolescents from harm. Findings from the National Longitudinal Study on Adolescent Health. *JAMA*. 1997;278(10):823-832. http://www.ncbi.nlm.nih.gov/pubmed/9293990. Accessed May 31, 2018.

119. Kong A, Thorleifsson G, Frigge ML, et al. The nature of nurture: Effects of parental genotypes. *Science*. 2018;359(6374):424-428. doi:10.1126/science.aan6877

120. Lord C, Rutter M, Goode S, et al. Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord*. 1989;19(2):185-212. http://www.ncbi.nlm.nih.gov/pubmed/2745388. Accessed May 8, 2015.

121. Casey BJ, Cannonier T, Conley MI, et al. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci*. March 2018. doi:10.1016/J.DCN.2018.03.001

122. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19(11):1523-1536. doi:10.1038/nn.4393

123. Happé F, Ronald A, Plomin R. Time to give up on a single explanation for autism. *Nat Neurosci*. 2006;9(10):1218-1220. doi:10.1038/nn1770

124. Dworzynski K, Happé F, Bolton P, Ronald A. Relationship Between Symptom Domains in Autism Spectrum Disorders: A Population Based Twin Study. *J Autism Dev Disord*. 2009;39(8):1197-1210. doi:10.1007/s10803-009-0736-1

125. Tao Y, Gao H, Ackerman B, Guo W, Saffen D, Shugart YY. Evidence for contribution of common genetic variants within chromosome 8p21.2-8p21.1 to restricted and repetitive behaviors in autism spectrum disorders. *BMC Genomics*. 2016;17(1):163. doi:10.1186/s12864-016-2475-y

126. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003;19(1):149-150. http://www.ncbi.nlm.nih.gov/pubmed/12499305. Accessed June 9, 2016.

127. Dudbridge F. Power and predictive accuracy of polygenic risk scores. Wray NR, ed. *PLoS Genet*. 2013;9(3):e1003348. doi:10.1371/journal.pgen.1003348

128. Bacanu S-A, Kendler KS. Method to estimate the approximate samples size that yield a certain number of significant GWAS signals in polygenic traits. *Genet Epidemiol*. 2018;42(5):488-496. doi:10.1002/gepi.22125

Edward Bullmore PhD FRCP FRCPsych FMedSci
Professor of Psychiatry
Head of Department

UNIVERSITY OF CAMBRIDGE

Department of Psychiatry

6th April 2018

To Whom It May Concern

Letter of support

This is to confirm that I am happy to support **the Common Variant Genetics of Autism and Autistic Traits Consortium** collaborative grant proposal to the Wellcome Trust.

Twin and familial studies have clearly demonstrated that autism is heritable. While studies have identified rare genetic variants associated with autism, only a handful of common genetic variants have been associated with autism. This is a critical gap in our understanding of the biology of autism. To address this, Simon Baron-Cohen and colleagues have developed an ambitious proposal to conduct a genome-wide meta-analysis using data from multiple autism cohorts world-over and, in parallel, cohorts with information on autistic traits. This will considerably advance our understanding of the biology of autism and autistic traits.

Power calculations have suggested that the study will identify more than 160 genetic loci associated with autism, providing a significant increase in the statistical power of polygenic scores. In addition, the development of a UK-wide genotyped autism cohort with links to electronic medical health records will be of great significance to the scientific and autistic community. This will allow for recall by genotype approaches in a number of downstream studies such as exome or whole-genome sequencing, neuroimaging, or cognitive testing. It will also enable deeper interrogating of genotype-phenotype correlations by integrating data from a number of questionnaires measuring traits related to autism, electronic medical health records, and genotypes.

The Autism Research Centre has a long-standing interest in investigating the dimensional nature of autism. Over the last few years, it has been active in investigating the genetic correlates of autism and autistic traits, and has developed close collaborations with leading researchers in the field of psychiatric genetics both in the UK and abroad. The ARC also closely works with industry experts in the field – Illumina, Inc., and 23andMe, Inc. The Department of Psychiatry has been rated one of the UK's nationally leading research groups in the three most recent Research

Herchel Smith Building for Brain and Mind Sciences
Forvie Site Robinson Way
Cambridge Biomedical Campus
CB2 0SZ UK
[redacted]

Assessment Exercises, and it plays a leading role in the internationally excellent Cambridge Neuroscience community of researchers in neuroscience and mental health.

Yours sincerely

Ed Bullmore
Head of Department

Herchel Smith Building for Brain and Mind Sciences
Forvie Site Robinson Way
Cambridge Biomedical Campus
CB2 0SZ UK
[redacted]

Institut Pasteur

Paris, le 23/07/18

Thomas Bourgeron
Institut Pasteur
25, Rue du Docteur Roux
75724 Paris Cedex 15, France
[redacted]

To whom it may concern,

I am writing in my capacity as the Director of the Human Genetics and Cognitive Functions Unit from the Neuroscience Department of the Institut Pasteur. My group gathers psychiatrists, neurobiologists and geneticists to better understand the social brain and the conditions that affect it such as autism.

It is my great pleasure to support the grant application for the Wellcome Trust proposed by Prof. Simon Baron-Cohen and Varun Warrier. I think that this project is crucial if we want to accelerate the research on the genetic architecture of autism and autistic traits. The establishment of a UK-wide autism biobank that's linked to electronic medical health records is mandatory if we want to tackle the genetic and phenotypic heterogeneity of autism. This project will also foster the collaboration with other autism researchers worldwide.

Their proposal includes a large-genome wide meta-analysis association (GWAMA) study of autism in 100K individuals, and recruit 10K new cases of autism from the UK into this; they will also conduct a GWAMA of autistic traits in the general population in 250K individuals. Following these two GWAMAs, they will identify causal variants for functional analysis using fine mapping. Their analysis pipeline will investigate enrichment in tissues, cell types, and neural networks as well as a deep investigation of the clinical heterogeneity within the autism cohort, in particular, mechanisms that contribute to sex differences and large variances in cognitive ability within the autism spectrum.

My group is already collaborating with Prof. Simon Baron-Cohen and Varun Warrier on different projects and this new application will clearly addresses key questions that are still not understood. For example, we have, still very few information on the combination of *de novo*, rare and common variants that will impact on the severity of the clinical symptoms. It has now been repeatedly observed that the same genes can be risk factors for different neuropsychiatric conditions. What we do not understand, however, is how the same genetic 'risk' can have such divergent outcomes.

As the PI of the genetics Work Package of the EU-AIMS funded by the European Union, I consider this application as an remarkable opportunity to have a complementary approach to this complex field of research.

Sincerely yours,

[redacted]

Prof. Thomas Bourgeron

Member of the French Academy of Sciences

Professor at the University Paris Diderot

Director of the Unit "Human Genetics and Cognitive Functions "

Supporting autism research at Cambridge
www.autismresearchtrust.org

The Autism Research Trust
19-21 Cookridge Street
Leeds LS2 3AG
www.autismresearchtrust.org
info@autismresearchtrust.org

Wellcome Trust
Gibbs Building
215 Euston Road
London NW1 2BE

17 April 2018

To whom it may concern,

The Common Variant Genetics of Autism and Autistic Traits Consortium

I am writing on behalf of the Autism Research Trust in support of the Common Variant Genetics of Autism and Autistic Traits Consortium. The study aims to conduct a GWAS on 100,000 autistic individuals and a GWAS on autistic traits.

We believe this is a very important project which will significantly advance our understanding of the genetics of autism, and therefore the Autism Research Trust would look to support the project by funding the data collection and acquisition for the GWAS of autistic traits (£50,000), should this application be funded by the Wellcome Trust.

Yours faithfully,

[redacted]

Charlotte Anderson
Chief Executive

Mark J. Daly, Ph.D.
Chief
Analytic and Translational Genetics Unit
Massachusetts General Hospital

Associate Professor of Medicine
Harvard Medical School
Massachusetts General Hospital

Institute Member
Co-Director, Program in Medical and Population Genetics
Broad Institute of Harvard and MIT

Department of Medicine
Analytic and Translational Genetics Unit
Massachusetts General Hospital
Simches Research Center
185 Cambridge Street
CPZN-6818
Boston, MA 02114

[redacted]

April 10, 2018

Dear Review Committee:

This letter is written to confirm my strong support for **the Common Variant Genetics of Autism and Autistic Traits Consortium** collaborative grant proposal to the Wellcome Trust.

I am very enthusiastic to collaborate on this study as it is important and complementary to existing efforts and will move the field ahead. As you know, the Psychiatric Genomics Consortium has been actively involved in delineating the polygenic risk for autism. However, we simply do not have the existing collections nor convenient strategies from biobanks to advance the autism effort forward in the way it needs to be. To this end, this is a timely and significant proposal that will advance our understanding of autism biology. Further, as the Director at the Institute for Molecular Medicine Finland (FIMM) I plan to initiate activities in Finland that can make meaningful contributions to this effort as well.

I strongly support this important proposal and look forward to contributing in whatever way possible.

Sincerely,

[redacted]

Mark Daly, PhD

Icahn School of Medicine at Mount Sinai

Joseph D. Buxbaum, Ph.D.
G. Harold and Leila Y. Mathers Professor
Director, Seaver Autism Center for Research and Treatment
Deputy Chair, Department of Psychiatry
Vice Chair for Mentoring, Department of Psychiatry
Chief, Center of Excellence in Neurodevelopmental Disorders, Freidman Brain Institute

Departments of Psychiatry, Genetics and Genomic Sciences, and Neuroscience
Icahn School of Medicine at Mount Sinai
Mount Sinai Health System
One Gustave L. Levy Place, Box 1230
Annenberg 22-24A
New York, NY 10029-6574

[redacted]

April 16th, 2018

Prof. Simon Baron-Cohen
Dr. Varun Warrier
Autism Research Centre
Cambridge University
Cambridge, UK

Dear Simon and Varun,

This is to confirm that I am happy to support the **Common Variant Genetics of Autism and Autistic Traits Consortium** collaborative grant proposal to the Wellcome Trust.

As my group has shown, it is clear that, by far, the largest fraction of the genetic risk for autism can be attributed to polygenic inheritance (Gaugler et al, Nat Genet 2014, 46(8):881-5, “Most genetic risk for autism resides with common variation”). The current proposal to conduct a GWAS of autism in 100,000 individuals with autism, followed by a GWAS of autistic traits, will provide considerable insights into the biology of autism. Collecting deeper phenotypic measures, including links to medical records will certainly help in studying the underlying heterogeneity in autism.

As you know, I am the founder and communicating PI for the Autism Sequencing Consortium (ASC) which currently has whole exome data on close to 40,000 autism-related samples (quads, trios, and cases with ancestry matched controls). About 14,000 of these WES samples were completed by the ASC since our 2015 publication, and, through our collaboration with iPSYCH and Broad (Borglum, Daly, Grove, also collaborators on your proposal), we are incorporating about 10,000 WES samples from the Danish blood spot program, so much of the data is at yet unpublished. In the remaining 4 years of the grant we anticipate having at least 60,000 samples analysed by WES or WGS. For many of the samples in the ASC there is genotype data available through the ASC or other databases, however, for over half the recent samples that genotyping is in process. In short, we expect to be able to add GWAS data for ~5,000 samples in the next year and additional GWAS data for another ~10,000 samples within four years. All data will be available to you. We have additional phenotypic data (beyond a categorical autism diagnosis) for most ASC samples.

Furthermore, we have, as one of our three current ASC aims, the plan to: “Use results from common and rare variant studies to describe the interplay of such variation in ASD risk.” In this Aim, we will integrate WES variants with data from WGS and GWAS to produce a complete picture of the genetic architecture of ASD, to improve gene discovery, and to refine clinical interpretation. This is an area where we would be happy to collaborate further with you. For many of the samples in existing cohorts, we have already organized the WES and GWAS data. In addition, going forward, we will run the Global Screening Array (GSA) for all ASC sample that are to be run with WES, and we even have some capacity to run WES on additional samples.

I look forward to a strong collaboration and to being part of your very exciting and critical study. I will be available for regular calls and annual meetings. In short, I strongly support this endeavour and will be delighted to collaborate on this study.

Sincerely,

Joseph Buxbaum, PhD

Institute of Psychiatry, Psychology & Neuroscience
Department of Forensic & Neurodevelopmental Sciences (FANS)

Prof Declan Murphy
Head of Department

Box PO23
De Crespigny Park
Denmark Hill
London SE5 8AF
[redacted]
kcl.ac.uk/fans

KING'S College LONDON

Sackler INSTITUTE FOR TRANSLATIONAL NEURODEVELOPMENT
translating basic science into treatment advances

9 April 2018

Letter of Support

This is to confirm that I am happy to support **the Common Variant Genetics of Autism and Autistic Traits Consortium** collaborative grant proposal to the Wellcome Trust.

This highly collaborative initiative will help us identify many more common genetic variants and CNVs associated with autism. Polygenic scores developed from this study can be used to further investigate heterogeneity, gene-environment interactions and biological correlates.

The EU-AIMS and the IMI will be very happy to collaborate in this project where possible.

Sincerely,

[redacted]

Declan Murphy

Professor of Psychiatry and Brain Maturation.
Mortimer D Sackler Professor of Translational Neurodevelopment.
Director of the Sackler Institute of Translational Neurodevelopment, Institute of Psychiatry, King's College London.
Head of Department of Forensic and Neurodevelopmental Sciences, Institute of Psychiatry, King's College London.

VU VRIJE UNIVERSITEIT AMSTERDAM

2500

Dr V. Warrier & prof S. Baron-Cohen

DATE	OUR REFERENCE	YOUR LETTER DATED	YOUR REFERENCE
March 30, 2018			

E-MAIL	TELEPHONE	ENCLOSURE(S)
[redacted]	[redacted]	

Re: Application

Dear Prof Baron-Cohen,

It is my great pleasure to strongly support your application “Common Variant Genetics of Autism and Autistic Traits Consortium”.

Furthermore, as you know, the Netherlands Twin Register has collected information using a continuous scale to assess autism, in a general population sample and we will be delighted to contribute to analyses of phenotype-genotype associations using this resource.

I very much hope you may be successful with this application, which will greatly further our insights into the etiology of autism and autism spectrum disorder.

Yours sincerely,

Prof Dorret I. Boomsma, Ph.D.

[redacted]

Netherlands Twin Register, Dept Biological Psychology,
Vrije Universieit, Amsterdam
Van der Boechorststraat 1, 1081 BT, Amsterdam, The Netherlands
[redacted]

MRC | Epidemiology Unit

investigating the causes and prevention of diabetes and obesity

16th April 2018

Dear Professor Baron-Cohen

Re: Common Variant Genetics of Autism and Autistic Traits Consortium application

I am writing in support of this highly collaborative grant that aims to conduct a Genome-Wide Association Study (GWAS) on autism and related quantitative traits at unprecedented scale. As you will be aware, only a tiny fraction of the heritable component of autism has been discovered to date, hugely limiting efforts to better understand the aetiology, disease sub-classification and prediction of this disorder. The proposed study will address this by collaboratively conducting the largest GWAS of autism to date, five times larger than the current GWAS of autism. In addition, there is a strong need to establish a UK-wide autism resource with deep-phenotyping and links to electronic health records, a resource which will enable future recall-by-genotype studies and the next generation of functional studies in this area.

I strongly support this endeavour and will be delighted to collaborate on this study.

[redacted]

John R.B Perry, Programme Leader (Growth and development)
MRC Epidemiology Unit, University of Cambridge.

DEPARTMENT OF BIOMEDICINE
THE FACULTY OF HEALTH SCIENCES
AARHUS UNIVERSITY

iPSYCH

To Whom It May Concern

Letter of Support of the proposal "Common Variant Genetics of Autism and Autistic Traits Consortium"

We are writing to express our full support of the proposal "Common Variant Genetics of Autism and Autistic Traits Consortium" and our enthusiasm for collaborating on this.

We are both part of iPSYCH (the Lundbeck Foundation Initiative for Integrative Psychiatric Research) as, respectively, PI (Anders Børglum) and lead analyst of the autism genetic investigations (Jakob Grove). In iPSYCH we are conducting genetic investigations of a Danish nation-wide birth cohort including 80,000 individuals of whom more than 50,000 suffer from psychiatric disorders, with a main focus on autism spectrum disorder (ASD) and other neurodevelopmental disorders.

Our recent GWAS paper (currently under review at Nature Genetics) which included data from iPSYCH and the PGC (Psychiatric Genomics Consortium) reported the first robustly identified common variants associated with ASD.

We have collaborated with Simon Baron-Cohen and Varun Warrier previously to conduct GWAS of traits related to autism. This includes a GWAS on empathy[1], social relationship satisfaction (under review), and on systemizing (under review).

In light of these recent successes, we are delighted to continue our fruitful collaboration on investigating the genetic basis of ASD.

With our warmest recommendations, sincerely,

[redacted]

Anders Børglum

[redacted]

Jakob Grove

Department of Biomedicine

Centre for Integrative Sequencing, iSEQ

Lundbeck Foundation Initiative for Integrative Psychiatric Research, iPSYCH

Anders D. Børglum
MD, PhD
Professor
Centre Director

Dato: 10.04.2018

[redacted]

Afs. CVR-nr.: 31119103

[1] V. Warrier et al., Transl. Psychiatry. 8 (2018), doi:10.1038/s41398-017-0082-6.

Department of Biomedicine
Centre for Integrative Sequencing, iSEQ
Aarhus University
Bartholins Allé 6
8000 Aarhus C, Denmark

Tlf.: +45 8716777
Fax: +45 86123173
http://www.biomed.au.dk
http://www.iseq.au.dk
http://www.ipsych.dk

SPARK
Igniting autism research
Improving lives

April 12th, 2018

Professor Simon Baron-Cohen, FBA
Fellow, Trinity College, Cambridge,
President, International Society for Autism Research,
Director, Autism Research Centre
Psychiatry Department
Cambridge University
Douglas House
18B Trumpington Road
Cambridge CB2 8AH UK
[redacted]
www.autismresearchcentre.com

Dear Simon,

I am pleased to write this letter of support for your Common Variant Genetics of Autism and Autistic Traits Consortium grant that aims to conduct a GWAS on 100,000 autistic individuals, and in parallel, a GWAS of autistic traits. This study will develop a resource of 10,000 autistic individuals and their families in the UK. This effort will complement the efforts of SPARK (https://sparkforautism.org/), which aims to recruit 50,000 autistic families in the United States.

Given the immense genetic heterogeneity in autism, a world-wide collaborative effort is needed to identify genetic variants that contribute to autism risk. In parallel, we need deeper phenotyping of autistic individuals to understand the source of the heterogeneity and link this to genetics. The autism common variant consortium aims to do precisely this by pooling together genetic data collected from 100,000 autistic individuals world-wide. This will be a significant and vital step towards using genetic scores for aiding diagnosis and informing therapeutics. As the principal investigator of SPARK, I will be delighted to collaborate on this study, and will be glad to share our genome wide genotyping and exome sequencing data and expertise for this study.

Best of luck with your application.

Sincerely,

[redacted]

Wendy Chung MD PhD
Director of Clinical Research, SFARI

Queensland Institute of Medical Research

Genetic Epidemiology Laboratory
Division of Genetics and Population Health
Professor Nick Martin
[redacted]

29 March 2018

Professor Simon Baron-Cohen, Dr Varun Warrier
Autism Research Centre
University of Cambridge

Dear Simon and Varun,

Re: Wellcome Trust Grant on GWAS for Autism Spectrum traits

Just a short note to say how strongly we support your application and how keen we are to collaborate with you on its aims. While there are some single major gene effects on autism, it is clear that much of the risk comes from polygenic inheritance which also affects endophenotypes which are normally distributed in the population. It follows that GWAS for those traits in large, unselected samples (much easier to collect than clinically defined case samples) can elucidate the genetics of autism, just as GWAS for blood pressure in the normal range has illuminated the genetics of hypertension. We in Australia have access to just such large unselected samples, already genotyped, and are already well embarked on collecting your AQ instrument for them, as well as the Eyes test and Social Reciprocity Scale on other subsets. We shall be delighted to contribute these data to your effort and look forward to collaborating.

We wish you success in your application.

Yours sincerely

[redacted]

Nicholas G. Martin, PhD, FAA
Professor and Senior Scientist

Sarah Medland, PhD
Professor and Principal Research Fellow

300 Herston Road, Herston Q 4006 Australia | QIMR Locked Bag 2000 Herston Q 4029
T (617) 3362 0222 F (617) 3362 0111 W www.qimr.edu.au

Does your proposal involve a clinical trial?	No

You may submit up to two A4 pages of additional information (such as graphs, figures, tables and essential unpublished data).

Common Variant Genetics of Autism and Autistic Traits (GWAS) Consortium

Risk assessment

We provide a risk assessment for the study below. For each risk, we provide the description of the risk, the potential impact, the probability of risk (high – H, medium – M, and low -L), and the strategies to manage and mitigate the risk.

1. **Low number of participants recruited into the UK Autism Biobank (M)**: The total number of participants recruited into the UK Autism Biobank is lower than anticipated. This will affect all downstream analysis, which will reduce the statistical power. To mitigate this, we have detailed 7 different strategies to recruit participants into the study. This includes targeted recruitment strategies through the Child Development Centres and other autism assessment centres, Special Education Needs Schools etc, and more global recruitment strategy through a dedicated PR campaign. We have discussed each of these methods with local paediatricians, support groups, and PR agencies. Further, we will recruit participants into the study for the entirety of the five-year period of the grant, with the initial set of analyses focussing on individuals recruited in the first three years. These recruitment strategies were developed after conversations with PIs and research managers of cohorts in the UK (ALSPAC, Generation Scotland), and abroad (SPARK, the BLTS). We believe this approach will help us meet the recruitment target as a similar approach was used in SPARK, where more than 100,000 autistic individuals and their families have registered to date. If we are unable to meet our recruitment target of 8,000 probands by year 4, we will: 1. Run a second PR campaign to boost numbers; 2. Recruit from outside the UK (in Europe) using the network of research collaborators as a part of the EU-AIMS initiative (at no extra cost to the Wellcome).

2. **Low number of participants in the GWAS (M)**: The total number of participants included in the two GWAS studies is lower than anticipated. This will affect all downstream analysis, which will reduce the statistical power. This is likely if participants are lost after quality control or if phenotypic data is not available for enough participants. The estimates provided here are conservative. In general, approximately 46% of the emailed UK Biobank cohort have responded to the mental health questionnaire[1]. In the current study design, we have conservatively estimated a third of the participants in each cohort will respond the questionnaire in the autistic traits GWAS. However, for all studies, we have conducted statistical power calculations keeping in mind multiple scenarios. For the autistic traits GWAS, this can be mitigated by applying new methods to include related individuals[2], conduct multi-ethnic GWAS meta-analysis (MAMA, Turley et al., in preparation), and conduct multivariate GWAS using related phenotypes[3]. For the polygenic score analyses, this can be mitigated by using improved methods such as multivariate polygenic score techniques[4,5] and by better modelling the linkage disequilibrium[6]. This will improve the predictive accuracy of the polygenic scores. Further, the estimate of 155-165 loci identified is calculated using only the autism GWAS. Integrating the GWAS from autistic traits will increase the total number of loci identified. Overall, the combination of the autistic traits GWAS and the autism GWAS, when meta-analysed, should comfortably allow us to identify 155-165 loci.

3. **Delays in data access (M)**: There are delays in obtaining the data, which delays the overall project. Overall, while this will not affect the results of the project, it will delay the project. To mitigate this, we have signed MTAs with various PIs already with the aim of getting existing data by the end 2018. In addition, we have already downloaded WES, WGS, and genotype data from the Simon's Simplex Collection and AGRE. In Year 0, we will also develop pipelines for calculating polygenic risk scores.

4. **Effect sizes are lower than anticipated (M):** This is similar to risk #2, and suggests that the effect sizes are smaller than anticipated. Power calculations have been conducted conservatively to allow for a range of effect sizes. In addition, statistical power calculations for the GWAS and the polygenic score

analyses are in line with observed effect sizes for neuropsychiatric conditions. Further, aspects of this can be mitigated using the steps outlined in risk #2.

References

1. Davis KAS, Coleman JRI, Adams M, et al. Mental health in UK Biobank: development, implementation and results from an online questionnaire completed by 157 366 participants. *BJPsych Open*. 2018;4(03):83-90. doi:10.1192/bjo.2018.12

2. Loh P-R, Kichaev G, Gazal S, Schoech AP, Price AL. Mixed-model association for biobank-scale datasets. *Nat Genet*. June 2018:1. doi:10.1038/s41588-018-0144-6

3. Turley P, Walters RK, Maghzian O, et al. Multi-trait analysis of genome-wide association summary statistics using MTAG. *Nat Genet*. 2018;50(2):229-237. doi:10.1038/s41588-017-0009-4

4. Maier RM, Zhu Z, Lee SH, et al. Improving genetic prediction by leveraging genetic correlations among human diseases and traits. *Nat Commun*. 2018;9(1):989. doi:10.1038/s41467-017-02769-6

5. Krapohl E, Patel H, Newhouse S, et al. Multi-polygenic score approach to trait prediction. *Mol Psychiatry*. 2018;23(5):1368-1374. doi:10.1038/mp.2017.163

6. Vilhjálmsson BJ, Yang J, Finucane HK, et al. Modeling Linkage Disequilibrium Increases Accuracy of Polygenic Risk Scores. *Am J Hum Genet*. 2015;97(4):576-592. doi:10.1016/j.ajhg.2015.09.001

References
You should give the citation in full, including title of paper and all authors.

References

1. Tick B, Bolton PF, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: A meta-analysis of twin studies. *J Child Psychol Psychiatry Allied Discip*. 2016;57(5):585-595. doi:10.1111/jcpp.12499

2. Sandin S, Lichtenstein P, Kuja-Halkola R, Hultman C, Larsson H, Reichenberg A. The Heritability of Autism Spectrum Disorder. *JAMA*. 2017;318(12):1182. doi:10.1001/jama.2017.12141

3. Wang K, Gaitsch H, Poon H, Cox NJ, Rzhetsky A. Classification of common human diseases derived from shared genetic and environmental determinants. *Nat Genet*. 2017;49(9):1319-1325. doi:10.1038/ng.3931

4. Halladay AK, Bishop S, Constantino JN, et al. Sex and gender differences in autism spectrum disorder: summarizing evidence gaps and identifying emerging areas of priority. *Mol Autism*. 2015;6(1):36. doi:10.1186/s13229-015-0019-y

5. Loomes R, Hull L, Mandy WPL. What Is the Male-to-Female Ratio in Autism Spectrum Disorder? A Systematic Review and Meta-Analysis. *J Am Acad Child Adolesc Psychiatry*. 2017;56(6):466-474. doi:10.1016/j.jaac.2017.03.013

6. Lai M-C, Lombardo M V., Baron-Cohen S. Autism. *Lancet*. September 2013. doi:10.1016/S0140-6736(13)61539-1

7. Baio J, Wiggins L, Christensen DL, et al. Prevalence of Autism Spectrum Disorder Among Children Aged 8 Years — Autism and Developmental Disabilities Monitoring Network, 11 Sites, United States, 2014. *MMWR Surveill Summ*. 2018;67(6):1-23. doi:10.15585/mmwr.ss6706a1

8. Simonoff E, Pickles A, Charman T, Chandler S, Loucas T, Baird G. Psychiatric disorders in children with autism spectrum disorders: prevalence, comorbidity, and associated factors in a population-derived sample. *J Am Acad Child Adolesc Psychiatry*. 2008;47(8):921-929. doi:10.1097/CHI.0b013e318179964f

9. Hirvikoski T, Mittendorfer-Rutz E, Boman M, Larsson H, Lichtenstein P, Bölte S. Premature mortality in autism spectrum disorder. *Br J Psychiatry*. 2016;208(3):232-238. doi:10.1192/bjp.bp.114.160192

10. Cassidy S, Bradley P, Robinson J, Allison C, McHugh M, Baron-Cohen S. Suicidal ideation and suicide plans or attempts in adults with Asperger's syndrome attending a specialist diagnostic clinic: a clinical cohort study. *The Lancet Psychiatry*. 2014;1(2):142-147. doi:10.1016/S2215-0366(14)70248-2

11. Buescher AVS, Cidav Z, Knapp M, Mandell DS. Costs of Autism Spectrum Disorders in the United Kingdom and the United States. *JAMA Pediatr*. 2014;168(8):721. doi:10.1001/jamapediatrics.2014.210

12. Ronald A, Happé F, Price TS, Baron-Cohen S, Plomin R. Phenotypic and genetic overlap between autistic traits at the extremes of the general population. *J Am Acad Child Adolesc Psychiatry*. 2006;45(10):1206-1214. doi:10.1097/01.chi.0000230165.54117.41

13. Robinson EB, Koenen KC, McCormick MC, et al. Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Arch Gen Psychiatry*. 2011;68(11):1113-1121. doi:10.1001/archgenpsychiatry.2011.119

14. Krishnan A, Zhang R, Yao V, et al. Genome-wide prediction and functional characterization of the genetic basis of autism spectrum disorder. *Nat Neurosci*. 2016;19(11):1454-1462. doi:10.1038/nn.4353

15. He X, Sanders SJ, Liu L, et al. Integrated model of de novo and inherited genetic variants yields greater power to identify risk genes. Williams SM, ed. *PLoS Genet*. 2013;9(8):e1003671. doi:10.1371/journal.pgen.1003671

16. Ronemus M, Iossifov I, Levy D, Wigler M. The role of de novo mutations in the genetics of autism spectrum disorders. *Nat Rev Genet*. 2014;15(2):133-141. doi:10.1038/nrg3585

17. Gratten J, Wray NR, Peyrot WJ, McGrath JJ, Visscher PM, Goddard ME. Risk of psychiatric illness from advanced paternal age is not predominantly from de novo mutations. *Nat Genet*. 2016;48(7):718-724. doi:10.1038/ng.3577

18. Iossifov I, O'Roak BJ, Sanders SJ, et al. The contribution of de novo coding mutations to autism spectrum disorder. *Nature*. 2014;515(7526):216-221. doi:10.1038/nature13908

19. Sanders SJ, He X, Willsey AJ, et al. Insights into Autism Spectrum Disorder genomic

architecture and biology from 71 risk loci. *Neuron*. 2015;87(6):1215-1233. doi:10.1016/j.neuron.2015.09.016

20. Gaugler T, Klei LL, Sanders SJ, et al. Most genetic risk for autism resides with common variation. *Nat Genet*. 2014;46(8):881-885. doi:10.1038/ng.3039

21. Klei LL, Sanders SJ, Murtha MT, et al. Common genetic variants, acting additively, are a major source of risk for autism. *Mol Autism*. 2012;3(1):9. doi:10.1186/2040-2392-3-9

22. Lee SH, Ripke S, Neale BM, et al. Genetic relationship between five psychiatric disorders estimated from genome-wide SNPs. *Nat Genet*. 2013;45(9):984-994. doi:10.1038/ng.2711

23. Grove J, Ripke S, Als TD, et al. Common risk variants identified in autism spectrum disorder. *bioRxiv*. November 2017:224774. doi:10.1101/224774

24. Bernier R, Golzio C, Xiong B, et al. Disruptive CHD8 mutations define a subtype of autism aarly in development. *Cell*. 2014;158(2):263-276. doi:10.1016/j.cell.2014.06.017

25. Helsmoortel C, Vulto-van Silfhout AT, Coe BP, et al. A SWI/SNF-related autism syndrome caused by de novo mutations in ADNP. *Nat Genet*. 2014;46(4):380-384. doi:10.1038/ng.2899

26. Frazier TW, Embacher R, Tilot AK, Koenig K, Mester J, Eng C. Molecular and phenotypic abnormalities in individuals with germline heterozygous PTEN mutations and autism. *Mol Psychiatry*. 2015;20(9):1132-1138. doi:10.1038/mp.2014.125

27. Warrier V, the 23andMe Research team, Bourgeron T, Baron-Cohen S. Genome-wide association study of social relationship satisfaction: significant loci and correlations with psychiatric conditions. *bioRxiv*. October 2017:196071. doi:10.1101/196071

28. Warrier V, Toro R, Chakrabarti B, et al. Genome-wide analyses of self-reported empathy: correlations with autism, schizophrenia, and anorexia nervosa. *bioRxiv*. November 2017:050682. doi:10.1101/050682

29. Warrier V, Toro R, Chakrabarti B, et al. Systemizing is genetically correlated with autism and is genetically distinct from social autistic traits. *bioRxiv*. December 2017:228254. doi:10.1101/228254

30. Vorstman JAS, Spooren W, Persico AM, et al. Using genetic findings in autism for the development of new pharmaceutical compounds. *Psychopharmacology (Berl)*. 2014;231(6):1063-1078. doi:10.1007/s00213-013-3334-z

31. Vorstman JAS, Parr JR, Moreno-De-Luca D, Anney RJL, Nurnberger Jr JI, Hallmayer JF. Autism genetics: opportunities and challenges for clinical translation. *Nat Rev Genet*. 2017;18(6):362-376. doi:10.1038/nrg.2017.4

32. Duffney LJ, Zhong P, Wei J, et al. Autism-like deficits in Shank3-deficient mice are rescued by targeting actin regulators. *Cell Rep*. 2015;11(9):1400-1413. doi:10.1016/j.celrep.2015.04.064

33. Mei Y, Monteiro P, Zhou Y, et al. Adult restoration of Shank3 expression rescues selective autistic-like phenotypes. *Nature*. 2016;530(7591):481-484. doi:10.1038/nature16971

34. Sztainberg Y, Chen H, Swann JW, et al. Reversal of phenotypes in MECP2 duplication mice using genetic rescue or antisense oligonucleotides. *Nature*. 2015;528(7580):123-126. doi:10.1038/nature16159

35. Niemi ME, Martin HC, Rice DL, et al. Common genetic variants contribute to risk of rare severe neurodevelopmental disorders. *bioRxiv*. May 2018:309070. doi:10.1101/309070

36. Weiner DJ, Wigdor EM, Ripke S, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. *Nat Genet*. May 2017. doi:10.1038/ng.3863

37. Pardiñas AF, Holmans P, Pocklington AJ, et al. Common schizophrenia alleles are enriched in mutation-intolerant genes and in regions under strong background selection. *Nat Genet*. February 2018:1. doi:10.1038/s41588-018-0059-2

38. Wray NR, Ripke S, Mattheisen M, et al. Genome-wide association analyses identify 44 risk variants and refine the genetic architecture of major depression. *Nat Genet*. 2018;50(5):668-681. doi:10.1038/s41588-018-0090-3

39. Stahl E, Breen G, Forstner A, et al. Genomewide association study identifies 30 loci associated with bipolar disorder. *bioRxiv*. January 2018:173062. doi:10.1101/173062

40. Skene NG, Bryois J, Bakken TE, et al. Genetic identification of brain cell types underlying schizophrenia. *Nat Genet*. 2018;50(6):825-833. doi:10.1038/s41588-018-0129-5

41. Gaspar HA, Breen G. Drug enrichment and discovery from schizophrenia genome-wide association results: an analysis and visualisation approach. *Sci Rep*. 2017;7(1):12460. doi:10.1038/s41598-017-12325-3

42. So H-C, Chau CK-L, Chiu W-T, et al. Analysis of genome-wide association data highlights candidates for drug repositioning in psychiatry. *Nat Neurosci*. 2017;20(10):1342-1349. doi:10.1038/nn.4618

43. Ruderfer DM, Ripke S, McQuillin A, et al. Genomic dissection of bipolar disorder and schizophrenia, including 28 Subphenotypes. *Cell*. 2018;173(7):1705-1715.e16. doi:10.1016/j.cell.2018.05.046

44. Allardyce J, Leonenko G, Hamshere M, et al. Association between schizophrenia-related polygenic liability and the occurrence and level of mood-incongruent psychotic symptoms in bipolar disorder. *JAMA Psychiatry*. 2018;75(1):28. doi:10.1001/jamapsychiatry.2017.3485

45. St Pourcain B, Robinson EB, Anttila V, et al. ASD and schizophrenia show distinct developmental profiles in common genetic overlap with population-based social communication difficulties. *Mol Psychiatry*. January 2017. doi:10.1038/mp.2016.198

46. Hubbard L, Tansey KE, Rai D, et al. Evidence of common genetic overlap between schizophrenia and cognition. *Schizophr Bull*. 2016;42(3):832-842. doi:10.1093/schbul/sbv168

47. Riglin L, Collishaw S, Richards A, et al. Schizophrenia risk alleles and neurodevelopmental outcomes in childhood: a population-based cohort study. *The lancet Psychiatry*. 2017;4(1):57-62. doi:10.1016/S2215-0366(16)30406-0

48. Jones HJ, Stergiakouli E, Tansey KE, et al. Phenotypic manifestation of genetic risk for schizophrenia during adolescence in the general population. *JAMA Psychiatry*. 2016;73(3):221. doi:10.1001/jamapsychiatry.2015.3058

49. Smoller JW, Andreassen OA, Edenberg HJ, Faraone S V., Glatt SJ, Kendler KS. Psychiatric genetics and the structure of psychopathology. *Mol Psychiatry*. January 2018. doi:10.1038/s41380-

017-0010-4

50. The Autism Spectrum Disorders Working Group of The Psychiatric Genomics Consortium. Meta-analysis of GWAS of over 16,000 individuals with autism spectrum disorder highlights a novel locus at 10q24.32 and a significant overlap with schizophrenia. *Mol Autism*. 2017;8(1):21. doi:10.1186/s13229-017-0137-9

51. Stahl E, Breen G, Forstner A, et al. Genomewide association study identifies 30 loci associated with bipolar disorder. *bioRxiv*. January 2018:173062. doi:10.1101/173062

52. Corbin LJ, Tan VY, Hughes DA, et al. Formalising recall by genotype as an efficient approach to detailed phenotyping and causal inference. *Nat Commun*. 2018;9(1):711. doi:10.1038/s41467-018-03109-y

53. Robinson EB, St Pourcain B, Anttila V, et al. Genetic risk for autism spectrum disorders and neuropsychiatric variation in the general population. *Nat Genet*. 2016;48(5):552-555. doi:10.1038/ng.3529

54. Robinson EB, Samocha KE, Kosmicki JA, et al. Autism spectrum disorder severity reflects the average contribution of de novo and familial influences. *Proc Natl Acad Sci*. 2014;111(42):15161-15165. doi:10.1073/pnas.1409204111

55. Pinto D, Delaby E, Merico D, et al. Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am J Hum Genet*. 2014;94(5):677-694. doi:10.1016/j.ajhg.2014.03.018

56. Chaste P, Klei LL, Sanders SJ, et al. A genome-wide association study of autism using the Simons Simplex Collection: Does reducing phenotypic heterogeneity in autism increase genetic homogeneity? *Biol Psychiatry*. 2015;77(9):775-784. doi:10.1016/j.biopsych.2014.09.017

57. Cantor RM, Navarro L, Won H, Walker RL, Lowe JK, Geschwind DH. ASD restricted and repetitive behaviors associated at 17q21.33: genes prioritized by expression in fetal brains. *Mol Psychiatry*. 2018;23(4):993-1000. doi:10.1038/mp.2017.114

58. Sebat J, Lakshmi B, Malhotra D, et al. Strong association of de novo copy number mutations with autism. *Science*. 2007;316(5823):445-449. doi:10.1126/science.1138659

59. Sanders SJ, Neale BM, Huang H, et al. Whole Genome Sequencing in Psychiatric Disorders: the WGSPD Consortium. *bioRxiv*. July 2017:160499. doi:10.1101/160499

60. Demontis D, Walters RK, Martin J, et al. Discovery of the first genome-wide significant risk loci for ADHD. *bioRxiv*. 2017.

61. Fischbach GD, Lord C. The Simons Simplex Collection: A resource for identification of autism genetic risk factors. *Neuron*. 2010;68(2):192-195. doi:10.1016/J.NEURON.2010.10.006

62. Geschwind DH, Sowinski J, Lord C, et al. The autism genetic resource exchange: a resource for the study of autism and related neuropsychiatric conditions. *Am J Hum Genet*. 2001;69(2):463-466. doi:10.1086/321292

63. The SPARK Consortium. SPARK: A US Cohort of 50,000 Families to Accelerate Autism Research. *Neuron*. 2018;97(3):488-493. doi:10.1016/j.neuron.2018.01.015

64. Thornton LM, Munn-Chernoff MA, Baker JH, et al. The Anorexia Nervosa Genetics Initiative: Overview and methods. *bioRxiv*. December 2017:234013. doi:10.1101/234013

65. Hall LS, Adams MJ, Arnau-Soler A, et al. Genome-wide meta-analyses of stratified depression in Generation Scotland and UK Biobank. *Transl Psychiatry*. 2018;8(1):9. doi:10.1038/s41398-017-0034-1

66. Baron-Cohen S, Wheelwright SJ, Skinner R, Martin J, Clubley E. The autism-spectrum quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J Autism Dev Disord*. 2001;31(1):5-17.

67. Allison C, Baron-Cohen S, Wheelwright S, et al. The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): A Normally Distributed Quantitative Measure of Autistic Traits at 18–24 Months of Age: Preliminary Report. *J Autism Dev Disord*. 2008;38(8):1414-1425. doi:10.1007/s10803-007-0509-7

68. Baron-Cohen S, Hoekstra RA, Knickmeyer R, Wheelwright S. The Autism-Spectrum Quotient (AQ)—Adolescent Version. *J Autism Dev Disord*. 2006;36(3):343-350. doi:10.1007/s10803-006-0073-6

69. Auyeung B, Baron-Cohen S, Wheelwright S, Allison C. The Autism Spectrum Quotient: Children's Version (AQ-Child). *J Autism Dev Disord*. 2008;38(7):1230-1240. doi:10.1007/s10803-007-0504-z

70. Charman T, Loth E, Tillmann J, et al. The EU-AIMS Longitudinal European Autism Project (LEAP): clinical characterisation. *Mol Autism*. 2017;8(1):27. doi:10.1186/s13229-017-0145-9

71. Lam KSL, Aman MG. The Repetitive Behavior Scale-Revised: Independent validation in individuals with Autism Spectrum Disorders. *J Autism Dev Disord*. 2007;37(5):855-866. doi:10.1007/s10803-006-0213-z

72. Tomchek SD, Dunn W. Sensory processing in children with and without autism: A comparative study using the Short Sensory Profile. *Am J Occup Ther*. 2007;61(2):190-200. doi:10.5014/ajot.61.2.190

73. Bilker WB, Hansen JA, Brensinger CM, Richard J, Gur RE, Gur RC. Development of abbreviated nine-item forms of the Raven's standard progressive matrices test. *Assessment*. 2012;19(3):354-369. doi:10.1177/1073191112446655

74. Howard DM, Clarke T-K, Adams MJ, et al. The stratification of major depressive disorder into genetic subgroups. *bioRxiv*. May 2017:134601. doi:10.1101/134601

75. McCoy TH, Castro VM, Hart KL, et al. Genome-wide association study of dimensional psychopathology using electronic health records. *Biol Psychiatry*. 2018;83(12):1005-1011. doi:10.1016/j.biopsych.2017.12.004

76. Prieto ML, Ryu E, Jenkins GD, et al. Leveraging electronic health records to study pleiotropic effects on bipolar disorder and medical comorbidities. *Transl Psychiatry*. 2016;6(8):e870-e870. doi:10.1038/tp.2016.138

77. Ronald A, Hoekstra RA. Autism spectrum disorders and autistic traits: A decade of new twin studies. *Am J Med Genet Part B Neuropsychiatr Genet*. 2011;156(3):255-274. doi:10.1002/ajmg.b.31159

78. Allison C, Auyeung B, Baron-Cohen S, Bolton PF, Brayne C. Toward brief "Red Flags" for autism screening: The Short Autism Spectrum Quotient and the Short Quantitative Checklist for Autism in toddlers in 1,000 cases and 3,000 controls [corrected]. *J Am Acad Child Adolesc Psychiatry*. 2012;51(2):202-212.e7. doi:10.1016/j.jaac.2011.11.003

79. Skuse DH, Mandy WPL, Scourfield J. Measuring autistic traits: heritability, reliability and validity of the Social and Communication Disorders Checklist. *Br J Psychiatry*. 2005;187(6):568-572. doi:10.1192/bjp.187.6.568

80. Constantino JN, Davis SA, Todd RD, et al. Validation of a brief quantitative measure of autistic traits: Comparison of the Social Responsiveness Scale with the Autism Diagnostic Interview-Revised. *J Autism Dev Disord*. 2003;33(4):427-433. doi:10.1023/A:1025014929212

81. Okbay A, Baselmans BML, De Neve J-E, et al. Genetic variants associated with subjective well-being, depressive symptoms, and neuroticism identified through genome-wide analyses. *Nat Genet*. 2016;48(6):624-633. doi:10.1038/ng.3552

82. Larson T, Anckarsäter H, Gillberg C, et al. The autism--tics, AD/HD and other comorbidities inventory (A-TAC): further validation of a telephone interview for epidemiological research. *BMC Psychiatry*. 2010;10:1. doi:10.1186/1471-244X-10-1

83. Mazefsky CA, Anderson R, Conner CM, Minshew N. Child Behavior Checklist scores for school-aged children with Autism: Preliminary evidence of patterns suggesting the need for referral. *J Psychopathol Behav Assess*. 2011;33(1):31-37. doi:10.1007/s10862-010-9198-1

84. Gibbs RA, Boerwinkle E, Doddapaneni H, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393

85. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48(10):1279-1283. doi:10.1038/ng.3643

86. Huang J, Howie B, McCarthy S, et al. Improved imputation of low-frequency and rare variants using the UK10K haplotype reference panel. *Nat Commun*. 2015;6(1):8111. doi:10.1038/ncomms9111

87. Robinson EB, Koenen KC, McCormick MC, et al. Evidence that autistic traits show the same etiology in the general population and at the quantitative extremes (5%, 2.5%, and 1%). *Arch Gen Psychiatry*. 2011;68(11):1113. doi:10.1001/archgenpsychiatry.2011.119

88. Ardlie KG, Deluca DS, Segre A V., et al. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science (80-)*. 2015;348(6235):648-660.

doi:10.1126/science.1262110

89. Won H, de la Torre-Ubieta L, Stein JL, et al. Chromosome conformation elucidates regulatory relationships in developing human brain. *Nature*. 2016;538(7626):523-527. doi:10.1038/nature19847

90. Wang K, Li M, Hadley D, et al. PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res*. 2007;17(11):1665-1674. doi:10.1101/gr.6861907

91. MacDonald JR, Ziman R, Yuen RKC, Feuk L, Scherer SW. The Database of Genomic Variants: a curated collection of structural variation in the human genome. *Nucleic Acids Res*. 2014;42(Database issue):D986-D992. doi:10.1093/nar/gkt958

92. Macé A, Tuke MA, Deelen P, et al. CNV-association meta-analysis in 191,161 European adults reveals new loci associated with anthropometric traits. *Nat Commun*. 2017;8(1):744. doi:10.1038/s41467-017-00556-x

93. 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. doi:10.1371/journal.pcbi.1004219

94. Finucane HK, Bulik-Sullivan BK, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235. doi:10.1038/ng.3404

95. Lake BB, Ai R, Kaeser GE, et al. Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain. *Science (80-)*. 2016;352(6293):1586-1590. doi:10.1126/science.aaf1204

96. Lake BB, Chen S, Sos BC, et al. Integrative single-cell analysis of transcriptional and epigenetic states in the human adult brain. *Nat Biotechnol*. 2017;36(1):70-80. doi:10.1038/nbt.4038

97. Nowakowski TJ, Bhaduri A, Pollen AA, et al. Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science (80-)*. 2017;358(6368):1318-1323. doi:10.1126/science.aap8809

98. Zhong S, Zhang S, Fan X, et al. A single-cell RNA-seq survey of the developmental

landscape of the human prefrontal cortex. *Nature*. 2018;555(7697):524-528. doi:10.1038/nature25980

99. Ecker JR, Geschwind DH, Kriegstein AR, et al. The BRAIN Initiative Cell Census Consortium: Lessons learned toward generating a comprehensive brain cell atlas. 2017. doi:10.1016/j.neuron.2017.10.007

100. Romanov RA, Zeisel A, Bakker J, et al. Molecular interrogation of hypothalamic organization reveals distinct dopamine neuronal subtypes. *Nat Neurosci*. 2017;20(2):176-188. doi:10.1038/nn.4462

101. Zeisel A, Muñoz-Manchado AB, Codeluppi S, et al. Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*. 2015;347(6226):1138-1142. doi:10.1126/science.aaa1934

102. Tasic B, Menon V, Nguyen TN, et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nat Neurosci*. 2016;19(2):335-346. doi:10.1038/nn.4216

103. Lonsdale J, Thomas J, Salvatore M, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013;45(6):580-585. doi:10.1038/ng.2653

104. Lek M, Karczewski KJ, Minikel E V., et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285-291. doi:10.1038/nature19057

105. Parikshak NN, Swarup V, Belgard TG, et al. Genome-wide changes in lncRNA, splicing, and regional gene expression patterns in autism. *Nature*. 2016;540(7633):423-427. doi:10.1038/nature20612

106. Parikshak NN, Luo R, Zhang A, et al. Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell*. 2013;155(5):1008-1021. doi:10.1016/j.cell.2013.10.031

107. Bulik-Sullivan BK, Finucane HK, Anttila V, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236-1241. doi:10.1038/ng.3406

108. Bulik-Sullivan BK, Loh P-R, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.

doi:10.1038/ng.3211

109. Shi H, Mancuso N, Spendlove S, Pasaniuc B. Local Genetic Correlation Gives Insights into the Shared Genetic Architecture of Complex Traits. *Am J Hum Genet*. 2017;101(5):737-751. doi:10.1016/j.ajhg.2017.09.022

110. Zheng J, Erzurumluoglu AM, Elsworth BL, et al. *LD Hub: A Centralized Database and Web Interface to Perform LD Score Regression That Maximizes the Potential of Summary Level GWAS Data for SNP Heritability and Genetic Correlation Analysis*. Vol 33. Cold Spring Harbor Labs Journals; 2016. doi:10.1093/bioinformatics/btw613

111. Grotzinger AD, Rhemtulla M, Vlaming R de, et al. Genomic SEM Provides Insights into the Multivariate Genetic Architecture of Complex Traits. *bioRxiv*. April 2018:305029. doi:10.1101/305029

112. Pickrell JK, Berisa T, Liu JZ, Ségurel L, Tung JY, Hinds DA. Detection and interpretation of shared genetic influences on 42 human traits. *Nat Genet*. 2016;48(7):709-717. doi:10.1038/ng.3570

113. O'Connor L, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *doi.org*. October 2017:205435. doi:10.1101/205435

114. Scholtens S, Smidt N, Swertz MA, et al. Cohort Profile: LifeLines, a three-generation cohort study and biobank. *Int J Epidemiol*. 2015;44(4):1172-1180. doi:10.1093/ije/dyu229

115. Boyd A, Golding J, Macleod J, et al. Cohort Profile: The "Children of the 90s"—the index offspring of the Avon Longitudinal Study of Parents and Children. *Int J Epidemiol*. 2013;42(1):111-127. doi:10.1093/ije/dys064

116. Rotheram-Fuller E, Kasari C, Chamberlain B, Locke J. Social involvement of children with autism spectrum disorders in elementary school classrooms. *J Child Psychol Psychiatry*. 2010;51(11):1227-1234. doi:10.1111/j.1469-7610.2010.02289.x

117. Kasari C, Locke J, Gulsrud A, Rotheram-Fuller E. Social Networks and Friendships at School: Comparing Children With and Without ASD. *J Autism Dev Disord*. 2011;41(5):533-544. doi:10.1007/s10803-010-1076-x

118. Resnick MD, Bearman PS, Blum RW, et al. Protecting adolescents from harm. Findings from

the National Longitudinal Study on Adolescent Health. *JAMA*. 1997;278(10):823-832. http://www.ncbi.nlm.nih.gov/pubmed/9293990. Accessed May 31, 2018.

119. Kong A, Thorleifsson G, Frigge ML, et al. The nature of nurture: Effects of parental genotypes. *Science*. 2018;359(6374):424-428. doi:10.1126/science.aan6877

120. Lord C, Rutter M, Goode S, et al. Autism diagnostic observation schedule: a standardized observation of communicative and social behavior. *J Autism Dev Disord*. 1989;19(2):185-212. http://www.ncbi.nlm.nih.gov/pubmed/2745388. Accessed May 8, 2015.

121. Casey BJ, Cannonier T, Conley MI, et al. The Adolescent Brain Cognitive Development (ABCD) study: Imaging acquisition across 21 sites. *Dev Cogn Neurosci*. March 2018. doi:10.1016/J.DCN.2018.03.001

122. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. *Nat Neurosci*. 2016;19(11):1523-1536. doi:10.1038/nn.4393

123. Happé F, Ronald A, Plomin R. Time to give up on a single explanation for autism. *Nat Neurosci*. 2006;9(10):1218-1220. doi:10.1038/nn1770

124. Dworzynski K, Happé F, Bolton P, Ronald A. Relationship Between Symptom Domains in Autism Spectrum Disorders: A Population Based Twin Study. *J Autism Dev Disord*. 2009;39(8):1197-1210. doi:10.1007/s10803-009-0736-1

125. Tao Y, Gao H, Ackerman B, Guo W, Saffen D, Shugart YY. Evidence for contribution of common genetic variants within chromosome 8p21.2-8p21.1 to restricted and repetitive behaviors in autism spectrum disorders. *BMC Genomics*. 2016;17(1):163. doi:10.1186/s12864-016-2475-y

126. Purcell S, Cherny SS, Sham PC. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics*. 2003;19(1):149-150. http://www.ncbi.nlm.nih.gov/pubmed/12499305. Accessed June 9, 2016.

127. Dudbridge F. Power and predictive accuracy of polygenic risk scores. Wray NR, ed. *PLoS Genet*. 2013;9(3):e1003348. doi:10.1371/journal.pgen.1003348

128. Bacanu S-A, Kendler KS. Method to estimate the approximate samples size that yield a

certain number of significant GWAS signals in polygenic traits. *Genet Epidemiol*. 2018;42(5):488-496. doi:10.1002/gepi.22125

Are there any papers listed in your 'References' section as being "in press" that you wish to submit to us?	No

Team composition and management
Please describe why a collaborative approach is necessary for this project; the roles of all applicants; and how the project will be managed and led.

Please describe why a collaborative approach is necessary for this project; the roles of all applicants; and how the project will be managed and led.

The data collection and analytical methods required to identify common genetic variants associated with autism are immense. The data cannot be generated by a single lab and concerted efforts are required on a global scale to recruit, phenotype, and genotype participants. The co-investigators on this study are **Baron-Cohen**, **Rowitch**, **Hurles**, and **Geschwind**. This series of studies could not be achieved without the Cambridge teams collaborating with the Sanger Centre and UCLA. Cambridge (Baron-Cohen and Rowitch) has expertise on autism and neurodevelopmental disorders, and will lead cohort recruitment and phenotyping. The Sanger Centre will lead the efforts to genotype and analyse the data. UCLA will lead on interpreting the biological consequences of the loci and investigating the downstream, functional consequences of the autism GWAS.

Baron-Cohen will supervise the funded personnel on this study (Allison, Warrier, Smith, a project manager, a research assistant, an administrator and a webmaster). Together, they will design the study, obtain ethical approval, design the webpage, recruit participants into the study, collect phenotypic measures including linking to the EHRs, collect saliva samples and extract DNA from it, and send the samples for genotyping. **Warrier** will coordinate between the centres and will lead the statistical analysis, including coordinating between the various PIs. **Allison** will work with the research manager, the administrator, and the research assistant to coordinate ethics, recruitment, phenotyping, collecting saliva samples and genotyping. **Smith** will work with the webmaster to design the website, obtain the EHRs, and curate the data. This team have successfully worked together on another challenging autism genetics grant funded by the Templeton World Charity Foundation, and have a highly efficient pipeline for DNA collection, extraction (at LGC), sequencing (at Illumina, Inc), so are highly experienced.

The team will collaborate with a team of world-class international scientists (see the list of collaborators in the previous section) who have expertise in autism and/or human genetics. These collaborators will provide access to additional data including genotypic information on 90,000 additional autistic individuals and an equal number of controls, and genotypic information on individuals who have completed a measure of autistic traits. See the previous section for their specific contributions to this study and their expertise.

Host organisation(s)
Describe the commitment/contribution, if any, that the host organisation(s) will make to this project.

Describe the commitment/contribution, if any, that the host organisation(s) will make to this project.

The host organization(s) will contribute a proportion of the salaries of the lead investigator and the three co-investigators. **Baron-Cohen** has secured a grant from the Templeton World Charitable Foundation (TWCF) to collect DNA samples from 1,000 autistic individuals. The TWCF grant will fund the salaries of Smith, Allison, and Warrier for the first year of the grant. Baron-Cohen has also secured funding from the Autism Research Trust to support the project by funding the data collection and acquisition for the GWAS of autistic traits (£50,000, which is therefore not included in the application for the Wellcome Collaborative Grant).

9. Outputs management and sharing

Will the proposed research generate outputs of data, software, materials or intellectual property that hold significant value as a resource for the wider research community?	Yes

Which approach do you intend to use to maximise the impact of your significant research outputs to improve health and benefit the wider research community?
A combination of both approaches

Please provide an outputs management plan. Ensure this describes any significant data, software, materials or intellectual property outputs, their management, and resources required (refer to guidance).
Please provide an outputs management plan. Ensure this describes any significant data, software, materials or intellectual property outputs, their management, and resources required (refer to guidance). *Data availability*: Anonymized individual-level genetic and phenotypic data from the UK Autism Biobank will be made available to PIs to conduct specific research projects subject to approval from the PIs of the UK Autism Biobank. PIs will need to apply with details of the research project providing: 1. A signed Researcher Distribution Agreement with legal details about research data and use, and requires institutional approval; 2. Copy of the IRB research protocol and approval letter from the local IRB. Summary GWAS data will be made available to researchers to download. EHRs will currently not be made available to PIs as there is no consensus on the anonymity of these. However, the research team will endeavour to make these data available when methods become available to anonymize them. PIs may also request to recontact participants for specific studies provided participants have consented to this. This includes recall by genotype. All approved research projects and summaries will be made publicly available. If a participant withdraws, PIs will be informed of their withdrawal and be asked to delete the data linked to the participant ID. Pipelines used to process the data will be made available on Github or similar pages. This will be included in the Autism Research Centre's github page: https://github.com/autism-research-centre/ *Research communication*: All research publications will be made open access to ensure accessibility of the results to the wider scientific community. We will disseminate the results through conferences, press releases, radio and print interviews, and social media including Twitter. The Autism Research Centre (ARC) The ARC also publishes about 30 peer reviewed scientific articles per year, and it publishes in practitioner journals for GPs, etc., All stakeholders involved in the study (participants, clinicians, and advisory groups), will be updated through bi-annual newsletters. This will include a summary of the findings and implications of the research written for a non-scientific audience.

10. Public engagement

Do you have plans for engaging with the non-academic public about your work?	Yes

Please provide a brief outline of your public engagement plans.
All participants and community paediatricians involved in the study will be updated about the study bi-annually using dedicated newsletters. All research will be published in open access journals. In addition, we will make the summary of the results and their implications available to the general public through press releases, radio, television, and print interviews, social media platforms such as Twitter, and in talks to the general public such as *Pint of Science*. The Autism Research Centre (ARC) has a long history of actively engaging with the general public to summarize their research findings. We are already working with Autistica, a UK autism research charity with substantial expertise in working with autistic individuals in shaping research, as part of the AIMS-2-TRIALS consortium. In addition, we issue press releases for our significant work, including GWAS of empathy (https://www.cam.ac.uk/research/news/study-finds-that-genes-play-a-role-in-empathy), GWAS of cognitive empathy (https://www.cam.ac.uk/research/news/genes-influence-ability-to-read-a-persons-mind-from-their-eyes), and imaging-genetics studies of autistic children (https://www.cam.ac.uk/research/news/scientists-link-genes-to-brain-anatomy-in-autism). The ARC has an excellent track record of science communication. For example, Baron-Cohen has spoken at Science Festivals (Cheltenham, Edinburgh, Cambridge), Literary Festivals (Hay), and Music Festivals (Wellcome Trust Outreach at Latitude), as well as taking part in television documentaries (Channel 4's 2017 Are you autistic? 2017; BBC2 2016 Employable Me) and radio (Radio 4 Today Programme). Baron-Cohen also regularly gives outreach scientific presentations to autism charities, secondary schools, etc., The ARC is also working with policy influencers such as Autism Europe and the All Party Parliamentary Group for Autism, to translate research into policy.

Please note that we provide support for Wellcome Trust funded researchers to engage with the non-academic public. Do you wish to receive information about training, funding and other public engagement opportunities?	Yes

11. Location of activity

Will the funded activity take place at more than one location?	Yes

For each location, select the country and, where applicable, state the organisation (please include the administering organisation). Indicate the approximate percentage of the total funds that will be spent in each location, entering zero for locations where activity will take place but no significant funds will be spent. Salary costs, if requested, should be attributed to the employing organisation.

Country	Organisation	Percentage of funds
United Kingdom	University of Cambridge	100
United Kingdom	Wellcome Trust Sanger Institute	0
United States	University of California, Los Angeles	0

Will you require funds to be awarded directly to more than one location?	No

12. Costs requested and justification

Please select the currency in which you wish to apply.

GBP - Pound Sterling

Is the selected currency your local currency?	Yes

Salaries Are you requesting salaries? Please refer to guidance notes and definition of terms for further details.	Yes

Salaries

Staff category	Name (if known)	Basic starting salary (p.a.)	Salary grade / scale	Period on project (months)	% time	Total (£)
Postgraduate research assistant	Dr Varun Warrier	[redacted]	7 / [redacted]	49	100	[redacted]
Other research staff - Research Assistant	Research Assistant - to be appointed	29,799	5 / 039	60	100	207,681
Technician	Web Master	£39,992	8 / 49	60	20	57,906
Technician - Database Manager	Paula Smith	[redacted]	7 / [redacted]	49	100	[redacted]
Senior postdoctoral research assistant	Dr Carrie Allison	[redacted]	9 / [redacted]	49	100	[redacted]
Postgraduate research assistant	Project Manager - to be appointed	£35,550	7 / 045	60	100	253,620
Administrative support staff	Procurement and Finance Coordinator	£28,098	6 / 37	60	100	207,189

Justification for personnel requested, specifying their role and responsibilities.
1. Project Manager: 1.0 FTE for 60 months who will manage and coordinate all aspects of the project. Post-holder will have experience in managing large projects. Duties will include: developing the ethical and logistical framework of the project, recruiting participants, handling contracts, ensuring the deliverables are met, liaise with the Autism Assessment Centres, the NAS, and schools to ensure recruitment targets are met, liaise with the stake-holder advisory panel and supervise the research assistant. 2. Research Manager (Dr Carrie Allison) is requested at 1.0 FTE for 49* months, who will be responsible for phenotyping. She will work alongside the Project Manager in liaising with participants and collaborators, the Research Office that governs all research projects in the University, and with the Clinical School that provides sponsorship and indemnity to research projects. She will assist with PPI and impact activities. Dr Allison has more than a decade of experience managing research projects at the Autism Research Centre. She is Co-Investigator on the £1.8 million autism multiplex gene sequencing study (Templeton World Charity Foundation), has expertise in setting up and running genetic studies, is an expert on autism phenotyping measures and developed the short version of the AQ (AQ-10), which will be used in the current

study, and the quantitative checklist for autism in toddlers (Q-CHAT).

3. Research Associate (Dr Varun Warrier): 1.0 FTE for 49* months, who will lead the genetic and statistical analysis. He will coordinate with existing genotyped cohorts to re-phenotype participants using autistic traits measures. Warrier's PhD was on the genetics of traits related to autism. He has expertise in GWAS and GWAS meta-analysis on phenotypes related to autism including cognitive empathy, systemizing, social relationship satisfaction, and theory of mind. He is involved in the autism multiplex family sequencing study, developed resources and pipelines at the High-performance computing cluster in Cambridge to conduct GWAS analyses.

4. Data Manager (Paula Smith): 1.0 FTE for 49* months, who will be responsible for managing the online data collection system, linking to EHRs, storing EHRs in data safe havens, and using natural language processing to extract data from these. She will work closely with the webmaster in study website development. Smith has 4 years experience working on the Cambridge Autism Research Database (CARD) as Database Manager. She also has relevant cancer genetics experience.

5. Research Assistant: 1.0 FTE for 60 months to support the team, responsible for day-to-day participant communication, coordinating saliva kit delivery and returns to the National Biosample Centre for extraction, storage and genotyping, and assisting the Project Manager in recruitment.

6. Webmaster: 0.2 FTE for 60 months, who will be responsible for database maintenance, data storage, and creating a project specific landing webpage.

7. Procurement and Finance Coordinator: 1.0 FTE for 60 months, to assist on HR and finance activities required to support the study, provide day-to-day administrative support. This is essential for purchasing, travel booking and participant payments.

*posts are funded from existing grants for the first 11 months of the project, therefore representing excellent value for money.

Materials and consumables Are you requesting materials and consumables?	Yes

Materials and consumables

Description	Total (£)
High-performance computing	25,000
Oragene DNA kits (including shipping and handling)	376,250
Computing and internet access and account fees	10,480
Computers	7,690
Computer remote access fee	1,490
Genotyping 10,000 samples	300,000
DNA aliquoting and long-term DNA storage (10-years)	108,000
DNA extraction (30,000 samples)	670,680
Electronic Health Record data linkage and download	192,000
PR campaign and advertising costs	46,770
Participant prize draw	10,000
Postage of DNA kits (outward and return)	100,000

Justification for materials and consumables requested
1. **Recruitment**: We request costs for a PR campaign **(£40,625**), including contacting top journalists, writing articles for magazines and newspapers, placing advertisements in autism related media, running a social media campaign through 'social media influencers'. Quotes are from Keep it Successfully Simple PR company. Additionally, we request **£6,145** to advertise on social media platforms and target ad campaigns. 2. **Phenotyping**: We request **£192,000** for obtaining NHS numbers via triangulation, linking participants' information to primary/secondary care records. These costs are based on quotes from NHS digital and GP records service providers (e.g. TPP). All other phenotyping will be done online through self- or caregiver-report measures. 3. **Participant Incentives**: We request **£10,000** in the form of a prize draw (£100 for 100 individuals). 4. **Sample collection**: We request **£375,000** for 30,000 Oragene kits (£12.50 per kit), to collect saliva from 10,000 autistic individuals, and biological family members. We request **£1,250** for shipping of Oragene kits from DNA Genotek to the Autism Research Centre . We request **£100,000** postage costs (outward and return), as £3.33 per participant using Royal Mail first class large-letter format, for 30,000 individuals. 5. **DNA extraction and storage**: We request **£670,680** for DNA extraction, quanititation using spectrophotometry, normalizing, and two aliquots (approximately £22.36 (inclusive of VAT) per sample for 30,000 samples). We request **£108,000** for storage for 10 years for 30,000 samples. Extraction, aliquoting and storage will be done at the NIHR National Biosample Centre (https://ukbiocentre.com/). These are estimates based on quotes from various service providers, representing the least expensive costs. 6. **DNA genotyping**: We will genotype 10,000 DNA samples (all autistic probands) using the UK Biobank axiom array at the NIHR National Biosample Centre at £30.00 per sample including VAT **(£300,000** in total) and is a very competitive quote from the NIHR National Biosample Centre. It will be advantageous to extract, genotype, and store the DNA at a single location. 7. **High-performance computing and storage**: We request **£25,000** for converting the raw CEL files to genotype calls, conducting the genotype and CNV analysis in the UK dataset, conducting the GWAS and CNV meta-analysis (N = 100K cases for autism, N = 250K for the autistic traits GWAS). This will include quality control, functional annotation, polygenic score analyses, and other downstream analyses. We will store data from the UK cohort for a period of 10 years. 8. **Clinical School Computing Service and remote access**: We request **£10,280** for access to the Clinical School Computing Service (£23 per month per user) for five staff for 60 months and three staff for 49 months ** This provides an email account, support, and network access for all members of staff. We request **£200** for setting up computational services for the Project Manager and Research Assistant (£100 per person) and **£1,490** for remote access fees for the computers at a cost of £4 per month per person. 9. **Computers**: We request **£7,690** to purchase computers for 8 staff which are essential for admin, contacting participants, data entry and analyses.

Animals Are you requesting animals?	No

Equipment Are you requesting equipment or equipment maintenance?	No

Synchrotron radiation sources Will the proposed research require access to a synchrotron source?	No

Access charges Are you requesting access charges?	No

Travel and subsistence Are you requesting travel and subsistence?	Yes

Travel and subsistence

Description	Total (£)
Conference attendance (Warrier, Allison, and Research Assistant)	15,000
Travel and subsistence for Warrier to go to the United States (UCLA)	5,275
Travel for research collaborotors to attend meetings	5,000
Travel and Subsistence for conference / scientific meeting attendance for Baron-Cohen	10,000

Justification for travel and subsistence costs requested.

1. We request **£10,000** for conference attendance/scientific meetings for Baron-Cohen (£2,000 per annum for five years).
2. We request **£15,000** for conference attendance/scientific meetings for Allison, Warrier, and the Research Assistant (£1,000 per annum per person for 5 years).
3. Given the highly collaborative nature of the study, we request **£5,275** for Warrier to travel to UCLA (Geschwind lab) to collaborate on the downstream functional annotation of the GWAS loci (based on the University of Cambridge recommended rate of £142.50 per day to provide funds for up to 30 days, totalling £4,275). We request a further £1,000 to enable Warrier to travel to the US from the UK.
4. We request **£5,000** for a travel allowance for researchers and stakeholders to travel to Cambridge and other participating sites including the Wellcome Trust Sanger Institute, various Autism Assessment Centres, and local NAS bodies. This is £1000 per annum, for 5 years.

Miscellaneous costs Are you requesting miscellaneous costs as part of this application?	Yes

Miscellaneous costs

Type	Description	Total (£)
Other	Professional development [redacted]	4,085
Other	Miscellaneous (printing, stationary, software)	5,000
Running conferences, meetings, workshops	Stakeholder engagement meetings	8,200
Other	Visa fees and health surcharge [redacted]	2,560

Justification for miscellaneous costs requested.

1. **Other - Professional Development of Research Staff** [redacted] for 49 months):

We request **£4,085**.

2. **Other - Miscellaneous**: We request £1,000 per annum (**£5,000 in total)** to cover costs including printing, software, and stationary.
3. **Running conferences, meetings, workshops** – 10 Stakeholder engagement meetings: To facilitate the study, we request £1,000 for refreshments (£100 per meeting) for the advisory group to meet biannually for five years. We request £640 per annum as payment for 8 people to attend the meetings (£40 each, over 5 years, total: £3,200). We request a travel budget of £4,000 for this purpose (£50 per person per biannual meeting over 5 years). Total requested: **£8,200**.
4. **Other – Visa fees and health surcharges:** [redacted] Total: **£2,560**.

Are you requesting research management costs under the miscellaneous costs heading? (for applicants from low- and middle-income countries only)	No

Summary of financial support requested	
	Total (£)
Salaries / Stipends	1,379,522
Materials and consumables	1,848,360
Animals	0
Associated animals costs	0
Equipment	0
Maintenance for existing equipment	0
Access charges	0
Travel and subsistence	35,275
Miscellaneous other	19,845
Total	**3,283,002**

13. Full economic costing

Is your organisation based in the UK?	Yes

Is your organisation calculating the full economic cost of this proposal?	Yes

What is the total full economic cost (£)?	4639191

14. Research involving human participants, human biological material and identifiable data

Does your project involve human participants, human biological material, or identifiable/potentially identifiable data?	Yes

Please confirm that you have read the Trust's guidance on the feedback of health-related findings in research and that you are in the process of considering your approach to this.
Confirmed

Please state by whom and when the ethics of the project has been, or will be, reviewed and specify any other regulatory approvals that have been obtained, or will be sought. We reserve the right to see relevant approval documents at any point during the lifetime of the grant, in accordance with our policy position on research involving human participants.
Ethical approval will be obtained from the East of England NHS Health Research Authority. Participants will be informed that: 1. This is a study to investigate the genetic and non-genetic contributors of autism; 2. Their de-identified data will be shared with approved researchers; 3. Their DNA will be stored and used for a number of genetic analyses including genotyping, and potentially whole-genome sequencing/whole-exome sequencing; 4. We will obtain their EHRs and education records where possible; 5. Participants can choose to be re-contacted for future studies; 6. Autism-relevant findings in this (CNVs) or future phases of the study (rare variants) will be provided to them if they consent to it via their nominated clinician; and 6. We will not provide any other genetic information back to the participants as we are not a lab with the required approval to do so. Participants will be provided with details of the study, steps to participate, and potential risks and benefits. All adults (above 16 years of age) with mental capacity must provide informed consent to participate. Individuals between 10 and 16 years of age with mental capacity must provide assent to their legal caregiver (parent or legal guardian) who will then provide consent to participate. Consent from the legal caregiver will be obtained for children below 10 years. For adults without mental capacity (as assessed by legal caregiver), consent will be obtained from the legal caregiver. Participants will be able to withdraw at any point during the study. If participants withdraw in the first 30 days after initially registering into their study, all their data will be removed from the study. Subsequent withdrawals will be conducted prospectively i.e. they will be removed from all future studies from that point. Consent will include permission to share data with collaborators anonymously, and to be recontacted to hear about new studies, and to be invited to hear about being recalled by genotype for new studies.

In the course of your project, do you propose to use facilities within the National Health Service (NHS) or to involve patients being cared for by the NHS?	Yes

Is a formal sponsor required for the project, for example under the Medicines for Human Use (Clinical Trials) Regulations or the Research Governance Framework for Health and Social Care and equivalent guidance?	Yes

Please indicate which organisation(s) has/have agreed to fulfil this role. Please note that the Wellcome Trust cannot act as sponsor.
The University of Cambridge will provide formal sponsorship.

If any potentially commercially exploitable results may be based upon tissues or samples derived from human participants, please confirm that there has been appropriate informed consent for such use.
We do not anticipate any potentially commercially exploitable results will emerge from this study. However, participants will be informed of all potential risks and benefits of the study, and informed consent will be obtained from the participant or care-giver if the participant is underage or lacks capacity to consent.

15. Proposals involving animals

Please indicate which of the following apply:
Neither of the above

16. Risks of research misuse

Please confirm that you have considered whether your proposed research could generate outcomes that could be misused for harmful purposes.
Confirmed

Have you identified any tangible risks of this type?	No

17. Freedom to operate/conflicts of interest

Describe any freedom to operate issues or potential conflicts of interest that have been identified or that might arise and how these will be or have been addressed.

In particular, please consider the following:

- Do any of the individuals involved in the project hold any consultancies or equities in, or directorships of, companies or other organisations that might have an interest in the results of the proposed research?
- Will the proposed research use technology, materials or other inventions that are subject to any patents or other form of intellectual property protection?
- Will any element of the research be subject to agreements with commercial, academic or other organisations, including arrangements with collaborators named in the grant application, that might lead to intellectual property issues or restrictions?

We do not think there are any potential conflicts of interest.

18. Wellcome Trust supported facilities

Will the project be based in one of the following Wellcome Trust supported facilities:

- the Wellcome Trust Sanger Institute
- a Wellcome Trust Centre
- an Africa and Asia Programme
- the Francis Crick Institute?

Yes

Please specify
Wellcome Trust Sanger Institute