Check for updates

ARTICLE OPEN

Identifying rare genetic variants in 21 highly multiplex autism families: the role of diagnosis and autistic traits

Ravi Prabhakar More [1,3✉], Varun Warrier [1,3], Helena Brunel[1], Clara Buckingham[1], Paula Smith[1], Carrie Allison[1], Rosemary Holt[1], Charles R. Bradshaw [2,4] and Simon Baron-Cohen [1,4✉]

Autism is a highly heritable, heterogeneous, neurodevelopmental condition. Large-scale genetic studies, predominantly focussing on simplex families and clinical diagnoses of autism have identified hundreds of genes associated with autism. Yet, the contribution of these classes of genes to multiplex families and autistic traits still warrants investigation. Here, we conducted whole-genome sequencing of 21 highly multiplex autism families, with at least three autistic individuals in each family, to prioritise genes associated with autism. Using a combination of both autistic traits and clinical diagnosis of autism, we identify rare variants in genes associated with autism, and related neurodevelopmental conditions in multiple families. We identify a modest excess of these variants in autistic individuals compared to individuals without an autism diagnosis. Finally, we identify a convergence of the genes identified in molecular pathways related to development and neurogenesis. In sum, our analysis provides initial evidence to demonstrate the value of integrating autism diagnosis and autistic traits to prioritise genes.

INTRODUCTION

Autism is a group of neurodevelopmental conditions characterised by difficulties communicating socially in addition to engaging in unusually restricted and repetitive behaviours and sensory difficulties. Autism is substantially heritable with a prevalence of 0.9–2.6%, with twin and familial heritability ranging from 60 to 90% [1, 2]. Variants across the allelic spectrum are associated with autism: de novo protein-truncating and missense variants in highly constrained genes [3, 4] at one end of the allelic frequency spectrum and common single nucleotide polymorphisms at the other end [5]. Considerable progress has been made in identifying genes, with information largely coming from de novo protein-truncating variants in highly constrained genes [3, 4, 6, 7]. However, de novo variants together account for less than approximately 3% of the total attributable variance [3, 8].

Most of the existing evidence comes from the study of simplex (one autistic individual per family) families, and recent studies have identified the role of rare inherited variation in autism. For instance, Simons Simplex Collection data shows that autistic individuals have a higher frequency of private, inherited truncating single nucleotide variants (SNVs) and copy number variants (CNVs) compared to non-autistic siblings [9]. The results of another study indicated that families with multiple autistic individuals (multiplex families) are enriched for both inherited and de novo rare CNVs [10]. Finally, a recent study identified an excess of inherited protein-truncating and missense variation in constrained genes in multiplex families [9, 11]. Rare inherited variants, missense and

Protein Truncating Variant (PTV) together account for up to 6% of the total likelihood of autism [8, 9]. Given that having a family relative with an autism diagnosis substantially increases an individual's likelihood of autism [1], studying the genetics of autism using highly multiplex families may aid in identifying genes associated with autism. Furthermore, given that family relatives of autistic individuals have elevated scores on measures of subclinical autism features - termed autistic traits [12] - incorporating information from measures of autistic traits may further aid in identifying genes associated with autism. Recent studies have investigated the effects of rare variants (inherited and de novo) associated with autism on autistic traits [13–15], but to our knowledge, no study has integrated autistic trait information with autism diagnosis to prioritise genetic variants.

To understand the contribution of rare inherited single-nucleotide variants (SNVs), Insertions/Deletions (INDELs) and de novo variants in highly multiplex autism families, we conducted Whole Genome Sequencing (WGS) on 21 families with three or more autistic individuals in the immediate family. We collected information on autistic traits as measured using the Autism Spectrum Quotient [16] (adult version or age equivalent measure for those under 16 years of age) for all individuals. We first investigated the association between de novo and inherited rare variants in autistic individuals. We then incorporated information regarding autistic traits where individuals were classified into subgroups based on their autistic traits, and investigated the relative contribution of de novo and inherited rare variants in these groups.

[1]Autism Research Centre, Department of Psychiatry, University of Cambridge, Cambridge, United Kingdom. [2]Gurdon Institute, University of Cambridge, Cambridge, United Kingdom. [3]These authors contributed equally: Ravi Prabhakar More, Varun Warrier. [4]These authors jointly supervised this work: Charles R. Bradshaw, Simon Baron-Cohen. ✉email: rm975@medschl.cam.ac.uk; sb205@cam.ac.uk

MATERIALS AND METHODS
Participants
We recruited 21 highly multiplex families into the study ($n = 112$ participants). We define highly multiplex families as families with three or more individuals with an autism diagnosis in the immediate family. These families were recruited between 2014 to 2018. Ten families had four or more individuals in the immediate family with an autism diagnosis. For one family (Family 10), we were unable to recruit all autistic individuals with an autism diagnosis and thus had genetic information only on two autistic members in the family. In total, 50 of the 112 individuals were females, and 76 had a diagnosis of autism (34 autistic females) from a qualified professional. Four individuals (three females) were suspected to have an autism diagnosis but had not been formally assessed at the time of participation in the study. Diagnosis of autism was provided by qualified clinicians independent of the study team. The study team obtained copies of the diagnostic report to verify the assessment of the autism diagnosis. Pedigree diagrams of all families are provided in Supplementary Fig. 1.

All participants completed a measure of autistic traits: The adult Autism Spectrum Quotient [16], the adolescent Autism Spectrum Quotient [17], the child Autism Spectrum Quotient [18], or the Quantitative Checklist for Autism in Toddlers [19]. The details of autistic trait information are provided in Supplementary Fig. 2. All participants provide one or more samples of saliva using Oragene-500 saliva kit. DNA was quantified and extracted by LGC using protocols designed to extract DNA from Oragene-500 saliva kits (GEN-9300-330).

The study obtained ethical approval from the Human Biology Research Ethics Committee (HBREC.2015.02). Informed consent was obtained from all participants.

Whole genome sequencing and quality control
For all samples, Whole Genome Sequencing was done using the Illumina HiSeq 4000 platform with paired-end reads of 150 base pairs. In read preprocessing, reads with low quality (less than Q20) were filtered out and low quality bases were trimmed from the ends of the reads. Adaptors were removed from both pairs using the Trim Galore tool (https://zenodo.org/badge/latestdoi/62039322). Shorter fragments where the reads overlapped were joined to provide fragments of >150 bases [20]. After pre-processing, the filtered data were aligned to the human genome reference assembly GRh37/hg19 using BWA-MEM 0.7.12 [21]. Next, PCR duplicates were removed using Picard tools (https://broadinstitute.github.io/picard/) and Base Quality Score Recalibration, as well as INDEL realignment, was performed using GATK 3.7 [22]. Using the Qualimap 2 tool, summary statistics were obtained to assess the effectiveness of read mapping and alignment quality mainly the total proportion of mapped reads and coverage of the reference genome [23].

Single-nucleotide polymorphisms (SNPs) and Insertions/Deletions (INDELs) were called using GATK 4.1 Haplotypecaller [22]. Finally, the called SNPs and INDELs were annotated with the Annovar tool [24, 25]. We retained rare non-synonymous variants (minor allele frequency <1% according to the Non-Finnish European (NFE) population from the Genome Aggregation Database (gnomAD) [26] along with variants considered to be pathogenic by SIFT [27] and PolyPhen-2 [28] algorithms, stop-gain and stop-loss variants. For disruptive INDELs, we considered frameshift insertion/deletion, stop-gain, and stop-loss variants. All SNP/INDELs variants were supported by at least 10 reads, and heterozygous variants had at least 33% reads corresponding to the alternate allele. A separate variant filtering analysis was performed for each of the 21 families for autistic traits and diagnosis approaches.

To identify de novo variants in trios (father, mother, and child), we analysed 44 trios identified in 9 of the 21 families. De novo variant calling was performed using the VarScan2 [29] trio option with parameters (minimal coverage 10, minimal variant frequency 0.20, p-value 0.05, adjacent variant frequency 0.05, and adjacent p-value 0.15). For the variant filtering step, we considered the following main criteria; 1) Reads depth greater than 20 in all three samples; 2) Alternate allele reads greater than or equal to 8 in the child; 3) Heterozygous de novo status of the variant in the child with alternate allele frequencies greater than 40% and absent in parents (alternate allele frequency less than 5%); 4) Non-synonymous and rare variants less than or equal to 1% in gnomAD NFE population; 5) Disruptive variants were considered based on SNV (Non-synonymous, SIFT and PolyPhen2 HDIV), frameshift In/Dels, stop gain/loss, and splice junctions. The resultant variants encoding genes are prioritised based on pLI greater than or equal to 0.9 and considered for the gene candidates.

Identifying and prioritising inherited and de novo variants within a family
Given that autism may be underdiagnosed in older individuals [30], we used information from both autism diagnosis and autistic traits to identify potential genetic variants in the pedigree. For autistic phenotypic traits, we divided individuals into four groups based on their scores on the Autism Spectrum Quotient (AQ), which was developed in a previous study of over 3000 individuals: [31] 1) Broad Autistic Phenotype (BAP) (between 1 - 2 standard deviations from the mean AQ score); 2) Medium Autistic Phenotype (MAP, 2 - 3 standard deviations from the mean); 3) Narrow Autistic Phenotype (NAP, 3+ standard deviations from the mean); and 4) Average or low autistic traits, which encompassed all other participants [31]. In the variant filtering process, we considered variants based on their presence in phenotypic traits in the family: 1) BAP variants were present in individuals of three groups (BAP, MAP, and NAP); 2) MAP variants were present in two groups (MAP and NAP); 3) NAP variants were present in only in individuals from the NAP group. The total number of individuals considered for the filtering criteria for autistic traits in each family is given in Supplementary Table 1. In the diagnosis-level variant filtering approach, we prioritised variants based on the autistic diagnosis of individuals in the pedigree. For this purpose, we applied two main criteria; 1) The variant had to be shared among at least two individuals diagnosed with autism; 2) The variant had to be absent in family members who were not autistic. We used the same criteria to identify variants associated with autistic traits. Specifically, for each autistic trait subgroup, the variant had to be present in at least two individuals in that autistic trait subgroup, or a higher autistic trait subgroup, and must be absent from other individuals. Variant prioritisation for each family was done using autistic traits in all groups and diagnosis status using bespoke scripts available at: https://github.com/ravimore8386/VarFilter-PD. We restricted our analyses to variants in constrained genes identified by using two related metrics: a probability of loss-of-function intolerance score greater than 0.9, and a LOEUF in the first decile (less than 0.37) set of genes from gnomAD.

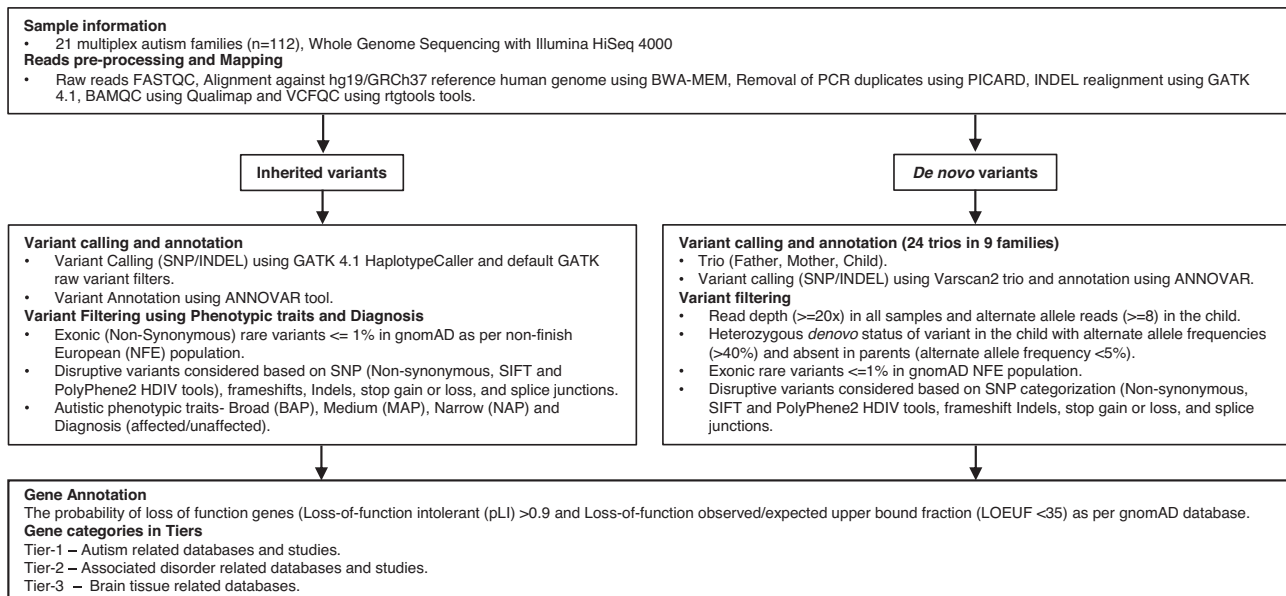### Identifying candidate genes in published studies and databases
Based on previous studies and existing databases, we curated four groups of genes associated with autism, co-occurring conditions, and human brain tissue. First, to identify genes that are robustly associated with autism or other severe neurodevelopmental conditions (Tier 1), we investigated whether they are present in one of the following databases: SFARI Gene (Categories S, 1–3) [32], SPARK Gene [33], Satterstrom et al., 2020 (99 genes) [3], De Rubeis et al., (2014) 107 genes [34], DDG2P developmental disorders database (monoallelic confirmed, probable, or possible) [35], or Developmental Brain Disorder Gene Database (DBDGD) (https://dbd.geisingeradmi.org/) [36]. Tier 2 consists of genes with less robust evidence linking them to autism, intellectual disability, or co-occurring conditions such as epilepsy, and ADHD. In Tier 2, we included genes with evidence from the following databases: AutDB (1225 genes) [37], Gene4Denovo (106 genes) [38], autismKB (171 genes) [39], epilepsy gene-sets of EpilepsyGene (499 genes) [40] and Wang et al., 2017 (621 genes) [41], Intellectual Disability databases IDGenetics (496 genes) [42] and sysID (1224 genes) [43], and ADHDgene (359 genes) [44]. Finally, we defined a third tier (Tier 3), which consists of brain tissue-related genes present in the databases Allen Brain Atlas (1000 genes) [45] and The Human Protein Atlas (1460 genes) [46], SynaptomeDB (1886 genes) [47]. Figure 1 shows the flowchart of the analysis performed in this study.

Analysis of the overrepresentation of genes
The genes prioritised in Tiers 1-3 from inherited and denovo analysis were inputted into PANTHER (Protein ANalysis THrough Evolutionary Relationships) to perform gene set overrepresentation and enrichment analysis for specific molecular functions, biological processes, cellular components, and protein class [48]. Protein-protein interaction networks were predicted using the STRING database (STRING 168 v9.1; http://string-db.org/) and only interactions with a high confidence score (≥0.7) were considered [49].

RESULTS
Overall, 84% of the WGS reads from saliva samples could be mapped to the reference human genome hg19 with 90.05% of bases at a coverage of at least 20x for 112 individuals. From this data, approximately 4,144,188 variants per individual were called

Sample information
- 21 multiplex autism families (n=112), Whole Genome Sequencing with Illumina HiSeq 4000

Reads pre-processing and Mapping
- Raw reads FASTQC, Alignment against hg19/GRCh37 reference human genome using BWA-MEM, Removal of PCR duplicates using PICARD, INDEL realignment using GATK 4.1, BAMQC using Qualimap and VCFQC using rtgtools tools.

Inherited variants

De novo variants

Variant calling and annotation
- Variant Calling (SNP/INDEL) using GATK 4.1 HaplotypeCaller and default GATK raw variant filters.
- Variant Annotation using ANNOVAR tool.

Variant Filtering using Phenotypic traits and Diagnosis
- Exonic (Non-Synonymous) rare variants <= 1% in gnomAD as per non-finish European (NFE) population.
- Disruptive variants considered based on SNP (Non-synonymous, SIFT and PolyPhene2 HDIV tools), frameshifts, Indels, stop gain or loss, and splice junctions.
- Autistic phenotypic traits- Broad (BAP), Medium (MAP), Narrow (NAP) and Diagnosis (affected/unaffected).

Variant calling and annotation (24 trios in 9 families)
- Trio (Father, Mother, Child).
- Variant calling (SNP/INDEL) using Varscan2 trio and annotation using ANNOVAR.

Variant filtering
- Read depth (>=20x) in all samples and alternate allele reads (>=8) in the child.
- Heterozygous *denovo* status of variant in the child with alternate allele frequencies (>40%) and absent in parents (alternate allele frequency <5%).
- Exonic rare variants <=1% in gnomAD NFE population.
- Disruptive variants considered based on SNP categorization (Non-synonymous, SIFT and PolyPhene2 HDIV tools, frameshift Indels, stop gain or loss, and splice junctions.

Gene Annotation
The probability of loss of function genes (Loss-of-function intolerant (pLI) >0.9 and Loss-of-function observed/expected upper bound fraction (LOEUF <35) as per gnomAD database.

Gene categories in Tiers
Tier-1 – Autism related databases and studies.
Tier-2 – Associated disorder related databases and studies.
Tier-3 – Brain tissue related databases.

Fig. 1 Overview of the rare inherited SNP/INDEL and de novo variants analysis pipeline used in this study. The separate variant calling and filtering were conducted for de novo and inherited variants. After this, variants were grouped into three tiers. SNV single nucleotide variant, Indel insertion or deletion, NFE Non-Finish European, GATK The Genome Analysis Toolkit, pLI Loss-of-function intolerant, LOEUF Loss-of-function observed/expected upper bound fraction, SIFT Sorting Intolerant From Tolerant, PolyPhen-2 Polymorphism Phenotyping v2, BAP Broad Autistic Phenotype, MAP Medium Autistic Phenotype, NAP Narrow Autistic Phenotype.

with a transitions (Ti) to transversions (Tv) ratio of 2.02 to 2.03. On average, we detected 24,704 exonic SNPs per individual (synonymous: 12,448; nonsynonymous: 12,253). For INDELs the average was 1,045,875 variants per individual; 423 Frameshifts; 37 Stop gains/losses, and 136 Splicing variants. Detailed variant statistics are given in Supplementary Fig. 2.

De novo variants in constrained genes

As a class, protein-truncating de novo variants in constrained genes are strongly linked to autism [3, 4]. Therefore, we initially focussed on this class of variants in 10 families where de novo variant calling was possible. After variant filtering, we identified six variants in six constrained genes (*DYNC1H1, KDM3A, ZBTB18, TAF4, DGKI,* and *PHLPP1*) with pLI > 0.9, which were detected in seven autistic individuals from six families. Three of these genes (*DYNC1H1, KMD3A,* and *ZBTB18*) have been robustly associated with autism [3] and severe developmental disorders (results presented in Table 1). Furthermore, *TAF4* has been previously associated with autism [33, 50] and a missense mutation in *TAF4* (A124P) was identified in two autistic individuals from two different families (Families 4 and 21) in our study. Despite observing this twice in our datasets, it is not observed in gnomAD non-Finnish European (NFE) populations. This is in accordance with previous studies that demonstrate the enrichment of ultra-rare variants (i.e., variants that are not observed in the ExAC/gnomAD population database) in autism [4, 51]. Finally, we also identified a frameshift deletion variant in *PHLPP1*, a gene with known structural variants (18q21.2) that have been associated with intellectual disability [52] but not with autism.

Segregation of inherited variants based on autistic traits and diagnoses within families

We then focused on inherited variants, primarily focusing on individuals with an autism diagnosis. We examined rare variants in constrained genes that are shared by at least two diagnosed autistic individuals in a family but are absent in all non-autistic individuals in the same family. Using this approach, we identified a total of 37 variants encoding 35 genes (Table 1). Fifteen of these genes have previously been linked to either autism or severe

undiagnosed developmental disorders. Examples of these genes include *CHD7, ANKRD11,* and *SON* which are robustly associated with both autism and severe developmental disorders. In addition, we identified three genes with less robust evidence linking them to autism or other co-occurring conditions (Tier 2), and four further genes enriched in brain tissue.

Given that many individuals with autism may be undiagnosed, especially in older generations, we further expanded our analyses to include traits associated with autism. We defined three groups for the autistic traits (NAP, MAP, and BAP; see the Methods section for more details) using self- or parent-reported measures of autistic traits (in the case of children). This classification was done agnostic to an autism diagnosis, but based on autistic trait scores. This also allowed for the classification of family members who had not sought a diagnosis for various reasons. It resulted in 47 individuals in the NAP group, 21 in the MAP group, 25 in the BAP group, and the remaining were outside of these three groups.

Based on the analysis of autistic traits, we found three additional genes with robust evidence for association with autism and severe developmental disorders. This includes a rare missense variant in *MTOR*, that is shared by individuals in the NAP group including individuals without an autism diagnosis. We also identified seven additional Tier 2 genes and two additional Tier 3 genes (Table 1). This illustrates the potential utility of considering a trait-based approach in addition to a diagnosis-based approach for gene discovery.

All the prioritised genes (de novo and inherited variants) obtained from both approaches (phenotypic traits and diagnosis) were plotted across chromosomes using PhenoGram (Fig. 2A). We identified a total of 17 genes that are implicated using both approaches (phenotypic traits and diagnosis), 12 were identified only using the trait-based approach, and 11 were identified only using the diagnosis-based approach (Fig. 2B). Of the 17 genes implicated using both approaches, nine (*ANKRD11, SETD2, CUX1, BRPF1, YTHDC1, DYNC1H1, KDM3A, WDR7,* and *MACF1*) are Tier 1 genes and are highlighted in red. Of the 11 identified only using the diagnosis-based approach, eight were Tier 1 (*CHD7, SON, PREX1, PTPRT, RIMS2, KDM5A, PTCH1, ZBTB18* and *PSMD1*). In contrast, only three Tier 1 genes (*KMT2C, ATP2B2,* and *MTOR*) were

**Table 1.** Summary of rare genetic variants encoding genes that are either inherited or de novo.

Gene symbol	Gene name	Inherited/de novo	Family	Trait/Diagnosis	Chr: Start	Ref/Alt	AA change	dbSNP	MAF NFE	SIFT	P2 HDIV	Notes	Tier
KMT2C	Lysine Methyltransferase 2C	Inherited	19	MAP	7:151932996	C/T	p.G892E	rs372408170	0.0001	D	D	DDG2P mono-allelic probable	1
ANKRD11	Ankyrin Repeat Domain 11	Inherited	10	NAP/Diagnosis	16:89346774	G/T	p.P2059H	rs117997391	0.0075	D	D	DDG2P mono-allelic confirmed	1
SETD2	SET Domain Containing 2, Histone Lysine Methyltransferase	Inherited	14	MAP/Diagnosis	3:47165380	G/A	p.S205F	.	.	D	P	DDG2P mono-allelic probable	1
			10	NAP/Diagnosis	3:47164711	C/T	p.R428H	rs201984344	0.0004	D	D		1
ATP2B2	ATPase Plasma Membrane Ca2+ Transporting 2	Inherited	15	BAP	3:10392216	G/A	p.R683C	rs754680596	.	D	D	SFARI gene 2	1
CUX1	Cut Like Homeobox 1	Inherited	16	MAP/Diagnosis	7:101921292	A/C	p.K530Q	rs118010189	0.0071	D	D	DDG2P possible	1
JADE2	Jagged Canonical Notch Ligand 1	Inherited	14	NAP	5:133901944	C/T	p.R370C	rs137957798	0.0005	D	D		2
MTOR	Mechanistic Target Of Rapamycin Kinase	Inherited	25	NAP	1:11272478	T/C	p.Y1151C	rs151082401	0.0003	D	D	DDG2P mono-allelic confirmed	1
BRPF1	Bromodomain And PHD Finger Containing 1	Inherited	10	NAP/Diagnosis	3:9783082	G/A	p.V605M	.	.	D	P	DDG2P mono-allelic confirmed	1
ESR1	Estrogen Receptor 1	Inherited	10	NAP/Diagnosis	6:152265352	C/T	p.R269C	rs142712646	0.0008	D	D		2
MAP2	Microtubule Associated Protein 2	Inherited	16	MAP/Diagnosis	2:210559050	G/A	p.G719D	rs148922251	0.0051	D	D		2
SEMA3F	Semaphorin 3F	Inherited	22	MAP	3:50225538	C/T	p.P684L	rs754678639	.	D	D		2
YTHDC1	YTH Domain Containing 1	Inherited	28	BAP/Diagnosis	4:69203301	C/A	p.D150Y	rs186920853	.	D	D	DBDGDB Score 2	1
CHD7	chromodomain helicase DNA binding protein 7	Inherited	19	Diagnosis	8:61707624	G/T	p.D726Y	rs748119797	.	D	D	DDG2P mono-allelic confirmed	1
SON	SON DNA and RNA binding protein	Inherited	19	Diagnosis	21:34923617	A/G	p.T694A	rs141608426	0.0034	D	D	DDG2P mono-allelic confirmed	1
PREX1	Phosphatidylinositol-3,4,5-trisphosphate dependent Rac exchange factor 1	Inherited	3	Diagnosis	20:47309282	C/T	p.G322S	rs202212653	.	D	D	SFARI gene 2	1
PTPRT	Protein tyrosine phosphatase receptor type T	Inherited	10	Diagnosis	20:40735430	G/C	p.T1129R	rs201830301	6.48E-05	D	D	SFARI gene 3	1
RIMS2	Regulating synaptic membrane exocytosis 2	Inherited	21	Diagnosis	8:104897660	C/T	p.S86F	rs17854256	0.0062	D	P	DDG2P mono-allelic probable	1
KDM5A	Lysine demethylase 5A	Inherited	4	Diagnosis	12:430192	G/A	p.P837L	rs745469846	6.48E-05	D	D	DDG2P possible	1
DYNC1H1	Dynein cytoplasmic 1 heavy chain 1	de novo	15	BAP/Diagnosis	14:102509021	C/T	p.P4150L	.	.	D	D	DDG2P mono-allelic confirmed	1
KDM3A	Lysine demethylase 3A	de novo	4	MAP/Diagnosis	2:86716780	A/G	p.K1191E	.	.	D	D	DBDGDB Score 3	1

Table 1. continued

Gene symbol	Gene name	Inherited/de novo	Family	Trait/Diagnosis	Chr: Start	Ref/Alt	AA change	dbSNP	MAF NFE	SIFT	P2 HDIV	Notes	Tier
ZBTB18	zinc finger and BTB domain containing 18	de novo	22	N/Diagnosis	1:244217437	G/C	p.V112L	.	0	D	D	DDG2P mono-allelic confirmed	1
TAF4	TATA-box binding protein associated factor 4	de novo	4/21	MAP/BAP/Diagnosis	20:60640497	C/G	p.A124P	.	.	D	P		2
CLOCK	Clock Circadian Regulator	Inherited	22	BAP	4:56319253	C/G	p.E392Q	rs373421741	.	D	P		2
JAG1	Jagged Canonical Notch Ligand 1	Inherited	19	MAP	20:10621462	T/G	p.R105S	rs146006022	0	D	P		2
NCOR2	Nuclear Receptor Corepressor 2	Inherited	10	NAP/Diagnosis	12:124832388	C/T	p.R1342H	rs36081651	0.0066	D	D		2
STXBP5L	Syntaxin Binding Protein 5 Like	Inherited	15	BAP	3:120952486	G/A	p.V379M	rs61996323	0.0084	D	D		2
TNS3	Tensin 3	Inherited	15	BAP	7:47331594	G/A	p.T1296M	rs41280696	0.0086	D	D		2
TTC28	Tetratricopeptide Repeat Domain 28	Inherited	14	NAP	22:28426233	G/A	p.R1352C	rs201500299	7.00E-04	D	D		2
PTCH1	patched 1	Inherited	19	Diagnosis	9:98242733	G/A	p.P295L	rs370755364	.	D	D	DBDGDB Score 3	1
PHLPP1	PH domain and leucine rich repeat protein phosphatase 1	de novo	7	NAP/Diagnosis	18:60506115	C/-	p.Y624fs	.	.	NA	NA		2
CASKIN1	CASK Interacting Protein 1	Inherited	19	MAP/Diagnosis	16:2231252	G/A	p.S706L	rs201599923	3.00E-04	D	D		3
WDR7	WD Repeat Domain 7	Inherited	10	NAP/Diagnosis	18:54547385	C/G	p.A1139G	rs773660280	.	D	D	DDG2P probable	1
FAT3	FAT Atypical Cadherin 3	Inherited	14	MAP/Diagnosis	11:92534043	G/T	p.V2622F	rs17615477	0.0054	D	D		3
MACF1	Microtubule Actin Crosslinking Factor 1	Inherited	3	NAP/Diagnosis	1:39853641	G/T	p.A2981S	rs144760259	0.0052	D	D	DDG2P mono-allelic probable	1
		Inherited	10	NAP/Diagnosis	1:39763365	G/T	p.C815F	rs148207245	0.0069	D	P	DDG2P mono-allelic probable	1
PCDH1	Protocadherin 1	Inherited	15	BAP	5:141243507	G/A	p.R418C	rs143703336	6.48E-05	D	D		3
VCAN	Versican	Inherited	14	NAP	5:82286147	G/T	p.V101L	rs758097509	.	D	D		3
PCLO	Piccolo presynaptic cytomatrix protein	Inherited	11	Diagnosis	7:82584378	G/A	p.T1964M	rs148432464	0	D	P		3
PSMD1	Proteasome 26S subunit, non-ATPase 1	Inherited	19	Diagnosis	2:232010978	A/G	p.M675V	rs201118764	0	D	D	SFARI 1	1
TP53BP1	Tumour protein p53 binding protein 1	Inherited	22	Diagnosis	15:43714134	C/A	p.G1340V	rs570488146	.	D	P		3
DGKI	Diacylglycerol kinase iota	de novo	4	BAP/Diagnosis	7:137531385	C/A	p.G75V	.	.	D	P		3

Chr chromosome, *Start* start chromosomal position according to hg19/GRCh37, *Ref* reference allele, *Alt* alternative allele, *AA change* alteration on the protein level, *MAF NFE* minor allele frequency in Non-Finish European (NFE) population as per GnomAD database, *P2* HDIV-PolyPhen-2, *HDIV* HumDiv-trained, *D* damaging/probably damaging/deleterious/disease-causing, *P* possibly damaging, *NA* not applicable.
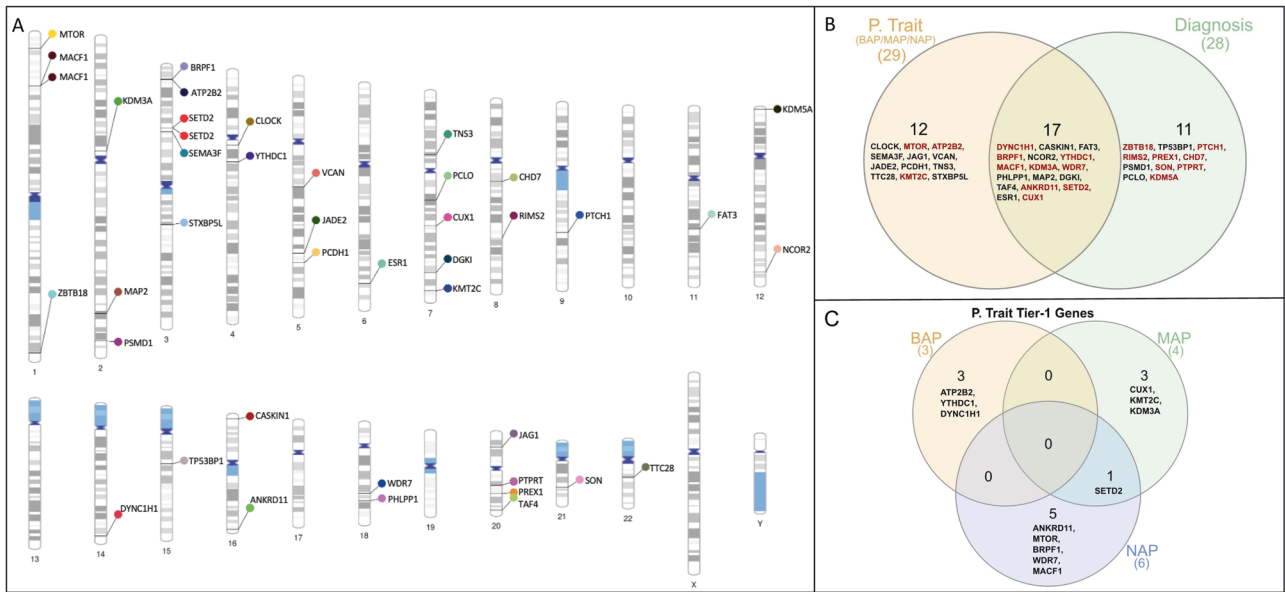
Fig. 2 Identified prioritised genes in the phenotype trait and diagnosis approaches in multiplex autism families. A Variants harbouring genes across all chromosomes. **B** Venn diagram showing the common and unique genes identified in phenotypic traits and diagnosis. Tier-1 genes are highlighted in red colour. **C** The distribution of the Tier 1 genes identified in BAP, MAP, and NAP phenotypic trait categories.

identified in the phenotypic trait-based approach. The distribution of the Tier 1 genes identified in the phenotypic trait-based (BAP, MAP, and NAP) are depicted in Fig. 2C. Six genes were identified in the NAP group, four in the MAP group, and three in the BAP group. This suggests that a greater number of Tier 1 genes are identified with increasing stringency of the AQ cut-offs.

Whilst the above two approaches highlight genes based on sharing, we also investigated how many individuals with an autism diagnosis were carriers of rare variants in a Tier 1 gene. In total, 50 out of the 74 individuals with autism diagnosis were carriers of rare variants in a Tier 1 gene. In comparison, only 11 of the 34 individuals without an autism diagnosis and who did not suspect they were undiagnosed autistic, were carriers of a rare variant in a Tier 1 gene (χ^2 (1, $N = 108$) = 11.75, $p = 6.07 \times 10^{-4}$). Altogether, this suggests that autistic individuals are enriched for rare genetic variants in genes previously linked to autism and/or neurodevelopmental conditions.

Analysis of functional and gene ontology (GO) terms
We next conducted functional and gene ontology (GO) analysis using genes in the three Tiers, to identify shared molecular processes. We identified significant enrichment in terms pertaining to developmental and neuronal processes (Table 2).

Additionally, we used the STRING database to search for protein-protein interactions, as shown in Fig. 3. Analysis showed that *KDM5A* and *KDM3A* strongly interact and co-express together in the HDMs demethylate histones (HSA-3214842) pathway (FDR 0.0345). Interestingly, we found an inherited SNP variant (rs745469846) harboured in *KDM5A* and a de novo SNP variant (chr2:86716780) in the *KDM3A* gene in Family 4. Histone demethylation genes have previously been implicated in autism [53], and these variants suggest that perturbation of their function may result in the autism phenotype.

DISCUSSION
In this study, we conducted whole-genome sequencing on 21 highly multiplex autism families and used both autistic trait and diagnosis status to identify potential genetic variants associated with autism. Our study highlights the value of autistic traits in identifying potential genetic variants associated with autism.

We first focused on genes with robust evidence linking them to autism or severe developmental conditions (Tier 1 genes) to identify molecular genetic diagnosis in these 21 families. In individuals where de novo variant calling was possible, we identified three autistic individuals who had a rare, protein-altering variant in highly constrained genes previously implicated in autism or severe developmental conditions, suggesting that even in multiplex families, de novo variants may be a genetic contributor to autism. For rare, inherited variants, we used a different approach. We focussed on genes shared in at least two autistic individuals and absent in any individual without an autism diagnosis. Overall, the approach identified rare variants in twenty different genes known to be associated with autism or other developmental conditions.

Two genes (SETD2 and MACF1) were identified in two separate families. Both genes are associated with both diagnosis and high autistic traits (NAP and MAP) and are classified as Tier 1 genes. Rare, protein altering genetic variants in *SETD2* were identified in MAP and NAP individuals in two families, indicating variants in this gene may increase autistic traits. *SETD2* is on chromosome 3, and codes for the SET domain containing 2 protein. This protein is a histone methyltransferase that trimethylates lysine 36 of histone H3 in nucleosomes. Genetic variants in *SETD2* are associated with autism [54]. According to van Rij et al., (2018) [55], de novo frameshift mutations were found in the *SETD2* gene in two people with Luscan-Lumish syndrome, who were diagnosed with intellectual disability, speech delay, macrocephaly, facial dysmorphism, and autism. Additionally, protein altering variants in *MACF1* (Microtubule-actin crosslinking factor 1) were found in NAP individuals in two families. The *MACF1* gene is known to be involved in multiple neural processes during development, including neurite outgrowth and neuronal migration [56].

Focusing again on rare variants in Tier 1 genes, we identified a modest excess of this class of variants in individuals with an autism diagnosis compared to those without an autism diagnosis. Approximately two-thirds of all autistic individuals in our study carried a rare variant in a Tier 1 gene. Whilst we are unable to provide a genetic diagnosis in this study, this excess of rare variants in autistic individuals suggests a genetically complex aetiology.

Table 2. A list of overrepresented functional categories based on GO enrichment analysis of genes.

Gene ontology	GO term	Fold enrichment	p-value	False discovery rate (FDR)	#Genes	Genes symbol
Biological process	Sex differentiation (GO:0007548)	6.78	2.83E−04	4.51E−02	7	PKD1, KDM5A, TAF4, ESR1, CHD7, PATZ1
	Regulation of neurogenesis (GO:0050767)	6.78	2.60E−05	9.54E−03	8	MAP2, SEMA3F, MTOR, RELA, MACF1, CUX1, CHD7, STK11,
	Regulation of developmental growth (GO:0048638)	6.57	1.04E−04	2.29E−02	7	MAP2, SEMA3F, PTCH1, MTOR, RIMS2, MACF1, CHD7
	Regulation of neuron projection development (GO:0010975)	6.13	1.72E−05	7.13E−03	9	MAP2, PREX1, FAT3, PKD1, SEMA3F, MTOR, MACF1, CUX1, STK11
	Developmental process involved in reproduction (GO:0003006)	4.37	3.18E−06	1.86E−03	14	KDM3A, PKD1, SETD2, PTCH1, KDM5A, TAF4, MTOR, BPTF, CLOCK, ESR1, CHD7, PATZ1, STK11, ATP2B2
	Positive regulation of developmental process (GO:0051094)	3.32	6.69E−05	1.70E−02	14	PREX1, ACIN1, PKD1, PTCH1, SP1, TP53BP1, MTOR, JAG1, RELA, RIMS2, MACF1, CUX1, CHD7, STK11
Molecular function	Transcription regulatory region sequence-specific DNA binding (GO:0000976)	2.93	0.000144	0.0457	15	SP1, KDM5A, TAF4, MTOR, ZBTB18, BPTF, RELA, CLOCK, BAZ2A, NCOR2, HIC2, CUX1, ESR1, CHD7, PATZ1
Cellular component	Presynaptic cytoskeleton (GO:0099569)	66.33	0.000593	0.0514	2	PCLO, RIMS2
	Presynaptic active zone (GO:0048786)	15.92	0.000142	0.0177	4	PCLO, DGKI, RIMS2, ATP2B2
	Dendritic growth cone (GO:0044294)	66.33	0.000593	0.0493	2	MAP2, PTCH1
	Glutamatergic synapse (GO:0098978)	7.83	2.51E−06	0.000717	9	ATP2B2, PTPRT, SEMA3F, PCLO, MTOR, DGKI, ACAN, RELA, ATP2B2

Finally, we conducted pathway analyses to identify convergence at the level of biological mechanisms, using genes identified across all three tiers. This analysis highlighted the role of both neuronal development and synaptic activity in the genes that we have identified so far, which is in line with biological mechanisms implicated by previous studies of genes associated with autism.

Previous studies have investigated the effects of genes associated with autism on autistic traits. However, these studies have not integrated autistic traits and autism diagnosis to prioritise genes, possibly because: 1) a large number of participants in existing cohorts come from simplex families (e.g., Simon's Simplex Collection); 2) these cohorts are not consistently measured for autistic traits; and 3) autistic traits measures are predominantly used for children with almost none of the large cohorts having used the same measure of autistic traits in both children and parents, making comparisons difficult. In our study, we were not statistically well-powered to identify new genes, and so investigated if rare protein altering variants in constrained genes already linked to autism, related conditions, or highly expressed in the brain are observed individuals with a diagnoses or individuals with high autistic traits. We identified individuals without an autism diagnosis who are carriers of protein altering rare variants in constrained genes previously implicated in autism. Whilst autistic traits measure are not a proxy for undiagnosed autism, this approach highlights the value of using trait-based information in addition to clinical diagnosis. This approach may be particularly relevant for older individuals in families with diagnosed autistic individuals, as they may be autistic but not have a clinical diagnosis.

This is one of the largest WGS studies focussing exclusively on highly multiplex autism families, however, this study does have some limitations. Whilst WGS does help to identify genetic variants, the current study is not statistically well-powered to identify novel genes. However, our approach of dividing genes into tiers of evidence helps highlight potential candidate genes that may reach statistical significance using larger sample sizes. In addition, the ability to measure autistic traits in our study demonstrates that at least some individuals without an autism diagnosis are carriers of rare variants in genes that have been implicated in autism. Nevertheless, these findings need to be validated and extended in other cohorts.

To this end, there is a need to collect measures of autistic traits in parents, harmonise different measures of autistic traits using different approaches (e.g., factor analysis), and investigate to what extent these findings are generalisable to different family types (e.g., families with just one or two individuals who have an autism diagnosis). Whilst we use the AQ in the current study, it is unclear to what extent these findings are generalisable to other measures of autistic traits. Furthermore, recent studies have demonstrated that autistic traits are multi-dimensional [14, 57], so future research needs to identify whether specific dimensions of autistic traits are enriched for different genetic variants.

Regarding different family types, previous research has indicated that both the genetic architecture of autism and the phenotypic variation in autistic traits differ between simplex and multiplex families. Compared to multiplex families, autistic individuals have enriched for de novo protein-truncating variants in constrained genes [11, 58], and non-autistic family members have scores on measures of autistic traits closer to the general population [59]. It is unclear to what extent our findings are relevant for simplex families, who are over-represented in large autism cohorts. Finally, we did not have further phenotypic information such as motor coordination, adaptive behaviour, or intellectual functioning of the participants of this study. Therefore, we were unable to investigate if these factors also contribute.
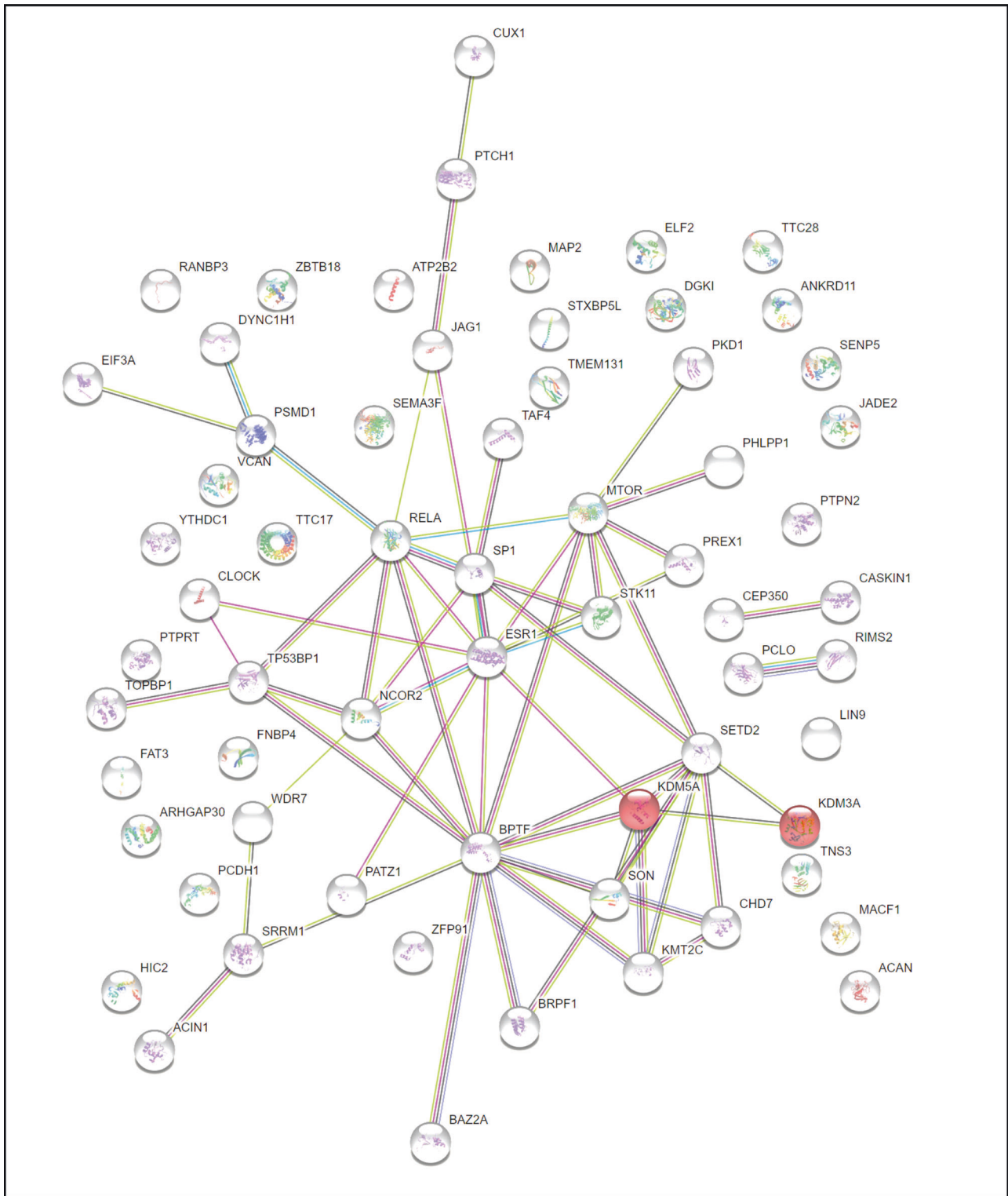
Fig. 3 Protein–protein interaction (PPI) Gene network analysis based on the STRING database of prioritised genes in pedigrees. Protein–protein interaction (PPI) analysis was conducted using all genes identified in the current study across all tiers (both inherited and de novo).

In conclusion, we report the results of whole-genome sequencing of 21 highly multiplex autism families. By combining both autistic trait and diagnostic information, we identify several rare, protein altering variants in constrained genes linked to autism or related conditions in these families. We demonstrate a modest excess of rare variants in genes associated with autism and/or neurodevelopmental conditions in autistic individuals compared to non-autistic family members. Our study demonstrates the value of combining both diagnosis and trait-based information to prioritise candidate genes for further investigation.

REFERENCES

1. Bai D, Yip BHK, Windham GC, Sourander A, Francis R, Yoffe R, et al. Association of genetic and environmental factors with autism in a 5-country cohort. JAMA Psychiatry. 2019;76:1035–43.
2. Tick B, Bolton P, Happé F, Rutter M, Rijsdijk F. Heritability of autism spectrum disorders: a meta-analysis of twin studies. J Child Psychol Psychiatry. 2016;57:585–95.
3. Satterstrom FK, Kosmicki JA, Wang J, Breen MS, De Rubeis S, An J-Y, et al. Large-scale exome sequencing study implicates both developmental and functional changes in the neurobiology of autism. Cell. 2020;180:568–84.e23.
4. Kosmicki JA, Samocha KE, Howrigan DP, Sanders SJ, Slowikowski K, Lek M, et al. Refining the role of de novo protein-truncating variants in neurodevelopmental disorders by using population reference samples. Nat Genet. 2017;49:504–10.
5. Grove J, Ripke S, Als TD, Mattheisen M, Walters RK, Won H, et al. Identification of common genetic risk variants for autism spectrum disorder. Nat Genet. 2019;51:431–44.
6. Sanders SJ, He X, Willsey AJ, Ercan-Sencicek AG, Samocha KE, Cicek AE, et al. Insights into autism spectrum disorder genomic architecture and biology from 71 risk loci. Neuron. 2015;87:1215–33.
7. Sebat J, Lakshmi B, Malhotra D, Troge J, Lese-Martin C, Walsh T, et al. Strong association of de novo copy number mutations with autism. Science. 2007;316:445–9.
8. Gaugler T, Klei L, Sanders SJ, Bodea CA, Goldberg AP, Lee AB, et al. Most genetic risk for autism resides with common variation. Nat Genet. 2014;46:881–5.
9. Krumm N, Turner TN, Baker C, Vives L, Mohajeri K, Witherspoon K, et al. Excess of rare, inherited truncating mutations in autism. Nat Genet. 2015;47:582–8.
10. Leppa Virpi M, Kravitz Stephanie N, Martin C, Andrieux J, Le Caignec C, Martin-Coignard D, et al. Rare inherited and de novo CNVs reveal complex contributions to ASD risk in multiplex families. Am J Hum Genet. 2016;99:540–54.
11. Ruzzo EK, Pérez-Cano L, Jung J-Y, Wang L, Kashef-Haghighi D, Hartl C, et al. Inherited and de novo genetic risk for autism impacts shared networks. Cell. 2019;178:850–866.e26.
12. Frazier TW, Youngstrom EA, Hardan AY, Georgiades S, Constantino JN, Eng C. Quantitative autism symptom patterns recapitulate differential mechanisms of genetic transmission in single and multiple incidence families. Mol Autism. 2015;6:58.
13. Antaki D, Guevara J, Maihofer AX, Klein M, Gujral M, Grove J, et al. Publisher correction: a phenotypic spectrum of autism is attributable to the combined effects of rare variants, polygenic risk and sex. Nat Genet. 2022;54:1284–92.
14. Warrier V, Zhang X, Reed P, Havdahl A, Moore TM, Cliquet F, et al. Genetic correlates of phenotypic heterogeneity in autism. Nat Genet. 2022;54:1293–304.
15. Trost B, Thiruvahindrapuram B, Chan AJS, Engchuan W, Higginbotham EJ, Howe JL, et al. Genomic architecture of autism from comprehensive whole-genome sequence annotation. Cell. 2022;185:4409–27.e18.
16. Baron-Cohen S, Wheelwright S, 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;31:5–17.
17. Baron-Cohen S, Hoekstra RA, Knickmeyer R, Wheelwright S. The autism-spectrum quotient (AQ)—adolescent version. J Autism Dev Disord. 2006;36:343–50.
18. Auyeung B, Baron-Cohen S, Wheelwright S, Allison C. The autism spectrum quotient: children's version (AQ-Child). J Autism Dev Disord. 2008;38:1230–40.
19. Allison C, Baron-Cohen S, Wheelwright S, Charman T, Richler J, Pasco G, 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;38:1414–25.
20. Magoc T, Salzberg SL. FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics. 2011;27:2957–63.
21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
22. Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline. Curr Protoc Bioinformatics. 2013;43:11.10.1–11.10.33.
23. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016;32:292–4.
24. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc. 2015;10:1556–66.
25. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:e164–e164.
26. Karczewski KJ, Francioli, Laurent C, Tiao G, Cummings BB, Alföldi J, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434–43.
27. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc. 2009;4:1073–81.
28. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, et al. A method and server for predicting damaging missense mutations. Nat Methods. 2010;7:248–9.
29. Koboldt DC, Larson DE, Wilson RK. Using VarScan 2 for germline variant calling and somatic mutation detection. Curr Protoc Bioinformatics. 2013;44:15.4.1–17.
30. Lai M-C, Baron-Cohen S. Identifying the lost generation of adults with autism spectrum conditions. Lancet Psychiatry. 2015;2:1013–27.
31. Wheelwright S, Auyeung B, Allison C, Baron-Cohen S. Defining the broader, medium and narrow autism phenotype among parents using the Autism Spectrum Quotient (AQ). Mol Autism. 2010;1:10.
32. Abrahams BS, Arking DE, Campbell DB, Mefford HC, Morrow EM, Weiss LA, et al. SFARI Gene 2.0: a community-driven knowledgebase for the autism spectrum disorders (ASDs). Mol Autism. 2013;4:36.
33. Feliciano P, Daniels AM, Green Snyder L, Beaumont A, Camba A, Esler A, et al. SPARK: a US cohort of 50,000 families to accelerate autism research. Neuron. 2018;97:488–93.
34. De Rubeis S, He X, Goldberg AP, Poultney CS, Samocha K, Ercument Cicek A, et al. Synaptic, transcriptional and chromatin genes disrupted in autism. Nature. 2014;515:209–15.
35. Firth HV, Richards SM, Bevan AP, Clayton S, Corpas M, Rajan D, et al. DECIPHER: database of chromosomal imbalance and phenotype in humans using ensembl resources. Am J Hum Genet. 2009;84:524–33.
36. Gonzalez-Mantilla AJ, Moreno-De-Luca A, Ledbetter DH, Martin CL. A cross-disorder method to identify novel candidate genes for developmental brain disorders. JAMA Psychiatry. 2016;73:275.
37. Basu SN, Kollu R, Banerjee-Basu S. AutDB: a gene reference resource for autism research. Nucleic Acids Res. 2009;37:D832–6.
38. Zhao G, Li K, Li B, Wang Z, Fang Z, Wang X, et al. Gene4Denovo: an integrated database and analytic platform for de novo mutations in humans. Nucleic Acids Res. 2019;48:D913–26.
39. Yang C, Li J, Wu Q, Yang X, Huang AY, Zhang J, et al. AutismKB 2.0: a knowledgebase for the genetic evidence of autism spectrum disorder. Database. 2018;2018:bay106.
40. Ran X, Li J, Shao Q, Chen H, Lin Z, Sun ZS, et al. EpilepsyGene: a genetic resource for genes and mutations related to epilepsy. Nucleic Acids Res. 2014;43:D893–9.
41. Wang J, Lin Z-J, Liu L, Xu H-Q, Shi Y-W, Yi Y-H, et al. Epilepsy-associated genes. Seizure. 2017;44:11–20.
42. Chen C, Chen D, Xue H, Liu X, Zhang T, Tang S, et al. IDGenetics: a comprehensive database for genes and mutations of intellectual disability related disorders. Neurosci Lett. 2018;685:96–101.
43. Kochinke K, Zweier C, Nijhof B, Fenckova M, Cizek P, Honti F, et al. Systematic phenomics analysis deconvolutes genes mutated in intellectual disability into biologically coherent modules. Am J Hum Genet. 2016;98:149–64.
44. Zhang L, Chang S, Li Z, Zhang K, Du Y, Ott J, et al. ADHDgene: a genetic database for attention deficit hyperactivity disorder. Nucleic Acids Res. 2012;40:D1003–9.
45. Sunkin SM, Ng L, Lau C, Dolbeare T, Gilbert TL, Thompson CL, et al. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. Nucleic Acids Res. 2013;41:D996–1008.
46. Uhlen M, Fagerberg L, Hallstrom BM, Lindskog C, Oksvold P, Mardinoglu A, et al. Tissue-based map of the human proteome. Science 2015;347:1260419–1260419.
47. Pirooznia M, Wang T, Avramopoulos D, Valle D, Thomas G, Huganir RL, et al. SynaptomeDB: an ontology-based knowledgebase for synaptic genes. Bioinformatics 2012;28:897–9.
48. Mi H, Muruganujan A, Huang X, Ebert D, Mills C, Guo X, et al. Protocol Update for large-scale genome and gene function analysis with the PANTHER classification system (v.14.0). Nat Protoc. 2019;14:703–21.
49. Szklarczyk D, Gable AL, Nastou KC, Lyon D, Kirsch R, Pyysalo S, et al. The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. Nucleic Acids Res. 2020;49:D605–12.
50. Ji X, Kember RL, Brown CD, Bućan M. Increased burden of deleterious variants in essential genes in autism spectrum disorder. Proc Natl Acad Sci. 2016;113:15054–9.
51. Weiner DJ, Wigdor EM, Ripke S, Walters RK, Kosmicki JA, Grove J, et al. Polygenic transmission disequilibrium confirms that common and rare variation act additively to create risk for autism spectrum disorders. Nat Genet. 2017;49:978–85.

52. Kaminsky EB, Kaul V, Paschall J, Church DM, Bunke B, Kunig D, et al. An evidence-based approach to establish the functional and clinical significance of CNVs in intellectual and developmental disabilities. Genet Med. 2011;13:777–84.
53. LaSalle J. Autism genes keep turning up chromatin. OA Autism. 2013;1:14.
54. Lumish HS, Wynn J, Devinsky O, Chung WK. Brief report: SETD2 mutation in a child with autism, intellectual disabilities and epilepsy. J Autism Dev Disord. 2015;45:3764–70.
55. van Rij MC, Hollink IHIM, Terhal PA, Kant SG, Ruivenkamp C, van Haeringen A, et al. Two novel cases expanding the phenotype of SETD2-related overgrowth syndrome. Am J Med Genet A. 2018;176:1212–5.
56. Moffat JJ, Ka M, Jung E-M, Smith AL, Kim W-Y. The role of MACF1 in nervous system development and maintenance. Semin Cell Dev Biol. 2017;69:9–17.
57. Warrier V, Toro R, Won H, Leblond CS, Cliquet F, Delorme R, et al. Social and non-social autism symptoms and trait domains are genetically dissociable. Commun Biol. 2019;2:328.
58. Chang TS, Cirnigliaro M, Arteaga SA, Pérez-Cano L, Ruzzo EK, Gordon A, et al. The contributions of rare inherited and polygenic risk to ASD in multiplex families. medRxiv. 2022. https://doi.org/10.1101/2022.04.05.22273459.
59. Gerdts JA, Bernier R, Dawson G, Estes A. The broader autism phenotype in simplex and multiplex families. J Autism Dev Disord. 2013;43:1597–605.

ACKNOWLEDGEMENTS

AUTHOR CONTRIBUTIONS
SB-C, VW, CRB, CA, RH, and PS designed the study. PS, RH, CB, and VW collected the data. RPM, CRB, HB, and VW analysed the data. RPM, CRB, VW, and SB-C wrote the first draft of the manuscript. All authors read, commented on, and edited the manuscript.

COMPETING INTERESTS
The authors declare no competing interests.

ADDITIONAL INFORMATION
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41380-022-01938-4.

Correspondence and requests for materials should be addressed to Ravi Prabhakar More or Simon Baron-Cohen.

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

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