STUDY PROTOCOL

Stage 1 Registered Report: Variation in neurodevelopmental outcomes in children with sex chromosome trisomies: protocol for a test of the double hit hypothesis [version 2; referees: 3 approved]

Dianne F. Newbury, Nuala H. Simpson, Paul A. Thompson, Dorothy V. M. Bishop

1Department of Biological and Medical Sciences, Oxford Brookes University, Oxford, Oxfordshire, OX3 0BP, UK
2Department of Experimental Psychology, University of Oxford, Oxford, Oxfordshire, OX1 3UD, UK

Abstract

Background: The presence of an extra sex chromosome is associated with an increased rate of neurodevelopmental difficulties involving language. Group averages, however, obscure a wide range of outcomes.

Hypothesis: The 'double hit' hypothesis proposes that the adverse impact of the extra sex chromosome is amplified when genes that are expressed from the sex chromosomes interact with autosomal variants that usually have only mild effects. *Neuroligin-4* genes are expressed from X and Y chromosomes; they play an important role in synaptic development and have been implicated in neurodevelopment. We predict that the impact of an additional sex chromosome on neurodevelopment will be correlated with common autosomal variants involved in related synaptic functions.

We describe here an analysis plan for testing this hypothesis using existing data. The analysis of genotype-phenotype associations will be conducted after this plan is published and peer-reviewed.

Methods: Neurodevelopmental data and DNA are available for 130 children with sex chromosome trisomies (SCTs: 42 girls with trisomy X, 43 boys with Klinefelter syndrome, and 45 boys with XYY). Children from a twin study using the same phenotype measures will form two comparison groups (Ns = 184 and 186). Three indicators of a neurodevelopmental phenotype will be used: (i) Standard score on a test of nonword repetition; (ii). A language factor score derived from a test battery; (iii) A general scale of neurodevelopmental challenges based on all available information. Autosomal genes were identified by literature search on the basis of prior association with (a) speech/language/reading phenotypes and (b) synaptic function. Preselected regions of two genes scoring high on both criteria, *CNTNAP2* and *NRXN1*, will be tested for association with neurodevelopmental outcomes using Generalised Structural Component Analysis. We predict the association with one or both genes will be detectable in children with SCTs and stronger than in the comparison samples.
Keywords
Language disorder, sex chromosome trisomy, autism, genetics, neurexin, neurexin
Amendments from Version 1

The changes all concern clarification. We have not made changes to the analysis plan, for reasons explained in responses to reviewers. Reviewers raised some other possible hypotheses that we could test, but, given that our small sample limits the number of reasonably-powered analyses we can undertake, we decided to stay with the original plan for testing the double hit hypothesis, specifically with CNTNAP2 and NRXN1 – while accepting that NRXN1, in particular, is a long shot.

We have added references to justify statements about synaptic role of CNTNAP2:

“The role of the CNTNAP2 protein in developing brain is not fully understood, and it is likely to play multiple roles at different time-points. While early functional studies of the CNTNAP2 protein indicated that it localises to nodes of Ranvier in axonal membranes, it is now recognised to have key functions at the synapse (Lu et al., 2016; Zweier et al., 2009).”

When introducing our ‘double hit’ account, we have noted other papers proposing a double hit aetiology of neurodevelopmental disorders in the context of microdeletions:

“The notion of a ‘double hit’ aetiology has been proposed previously to account for cases where a microdeletion is inconsistently associated with neurodevelopmental disorder (Girirajan et al., 2010; Newbury et al., 2013): the idea is that a severe phenotype may be seen when there are two copy number variants or mutations, each of which may be relatively innocuous on its own. Here, we extend that idea to argue that the effect of altered neuroligin gene dosage may depend on the genetic background provided by autosomes”

We have added clarification to questions about how we handled linkage disequilibrium.

We have made some changes in wording when describing the Generalised Structural Component Analysis, to give additional clarification.

See referee reports

Introduction

Developmental language disorder (DLD), a condition in which there are unexplained and persistent difficulties with language acquisition, affects around 7% of children (Norbury et al., 2016). Family studies show that DLD runs in families (Bishop, 2008), yet it has proved hard to identify any genetic or environmental factors that substantially increase risk. One reason is that DLD appears to be a complex multifactorial disorder where influences of individual genetic variants (alleles) are typically of small effect, and may interact with other genetic factors and with the environment. Indeed, the ways in which disorders pattern in families suggest that common genetic variants that confer risk of language disorder may lead to an autistic phenotype when they occur with other genetic risk factors (Bishop, 2010). Thus the specific phenotype can depend on the constellation of genetic variants, rather than there being separate risk factors for DLD and autism spectrum disorder (ASD).

Rather than recruiting increasingly large numbers to try to find reliable associations between language disorders and genetic variants in genome-wide studies, one way forward is to study rare disorders that have a large impact on the phenotype, which may point to functional pathways involved in more common forms of disorder. One instance of a striking association between a genetic condition and language disorder in children of normal intelligence is provided by the sex chromosome trisomies (SCTs), each of which affects 1–1.5 per 1000 children (Nielsen & Wohlet, 1991). In the 1960s, research was initiated to investigate neurodevelopmental outcomes of children with SCT detected on neonatal screening. A systematic review of these studies showed that in all three trisomies there were high rates of speech and language impairment, motor problems, and educational difficulties, despite IQ being within normal limits in most cases (Leggett et al., 2010). Furthermore, studies of samples who have developmental language disorder of unknown cause find an increased prevalence of sex chromosome trisomies (Simpson et al., 2014).

In a study of children with sex chromosome trisomies identified on prenatal screening, Bishop et al. (2011) found that 7 of 30 (24%) girls with karyotype 47,XXX, 9 of 19 (47%) boys with 47, XXY and 15 of 21 (71%) boys with 47,XYY had a history of speech and language-therapy, compared with rates of 4% in sisters and 18% in brothers. Furthermore, this same study found that 2 of 19 (11%) boys with 47.XXY, and 4 of 21 (20%) boys with 47.XYY had received a diagnosis of ASD, compared with an estimated national prevalence rate of 0.2% in girls and 0.6% in boys. In addition, many children with SCTs who were not diagnosed with ASD had evidence of communication difficulties on parental report, including pragmatic (autistic-like) problems, in all three karyotypes. More recent research has provided further evidence of a link with autism as well as other neurodevelopmental disorders in boys with a sex chromosome trisomy (Ross et al., 2012).

The impact of a trisomy is influenced by distinctive characteristics of the sex chromosomes. In most cases, the phenotypic effects of SCTs are much less severe than the impact of an autosomal trisomy: Down syndrome (trisomy 21) usually leads to intellectual disability, and most other trisomies are lethal. Viable trisomies usually involve small chromosomes with a low gene count (for example the Y chromosome), where the effects associated with altered gene dosage are less severe. An exception to this rule is the X chromosome. The X chromosome has a relatively high gene count, but the impact of a duplication is relatively mild because mechanisms of inactivation have evolved, such that in typical females, only one copy is active, and in effect, both males and females have one set of functional genes from this chromosome. In trisomies that involve the X chromosome, two copies are inactivated, largely negating the presence of additional genetic material. There are, however, exceptions to this rule, with between 12–20% genes escaping inactivation to some extent: These include genes in the pseudo-autosomal region, and other genes that have homologues on the Y chromosome (Carrel & Brown, 2017).

The fact that there is an increase in problems affecting speech, language and communication in all three sex chromosome trisomies suggests there is an adverse impact of an additional copy of a gene that is expressed and has homologous forms on the X and Y chromosomes. Neuroligin-4 (NLGN4) is a strong
candidate for such a gene, for several reasons (Bishop et al., 2011). First, NLGN4X, located on Xp22, escapes inactivation (Berletch et al., 2011). Second there is a homologous gene, NLGN4Y on the Y chromosome at Yq11.2. Third, neuroligins are expressed in brain, as well as other tested tissues (Abrahams & Geschwind, 2010; Jamain et al., 2003). Fourth, as reviewed by Cao & Tabuchi (2017), mutations of NLGN4 have been linked to ASD (Jamain et al., 2003; Laumonnier et al., 2004; Lawson-Yuen et al., 2008; Marshall et al., 2008; Pampanos et al., 2009; Talebizadeh et al., 2006; Yan et al., 2008) – although this finding is inconsistent (Bishop et al., 2008, 2010), Bishop proposed that a defect in NLGN4 may abolish formation or function of synapses involved in communication. Note that these authors also implicated another X-chromosome neuroligin, NLGN3, in autism, but this is located at Xq13, where one copy would be inactivated, and there is no homologue on the Y-chromosome. Therefore, unlike NLGN4, NLGN3 would not be over-expressed in those with an extra X or Y chromosome.

For the reasons described above, we may hypothesise that an extra copy of NLGN4 could be implicated in neurodevelopmental problems. However, we also need to explain within- karyotype variation. Although there is a substantial increase in rates of speech, language and social communication problems in children with SCTs, the additional chromosome does not cause language impairment or ASD in a deterministic fashion. A minority of children have no evidence of developmental difficulties, a minority are severely affected with disabilities extending across many domains, and most have mild to moderate impairments (Linden & Bender, 2002).

The wide variation in outcomes suggests that the extra gene dosage could act as a multiplier of other risk factors, which interact with the sex chromosome genes in a dosage-dependent manner and so only assume importance in the subset of individuals who have other genetic or environmental risk factors (Bishop & Scerif, 2011). This explanation is consistent with rodent research comparing the effect of a NLGN3 mutation between different strains of mouse, suggesting the impact is dependent on the genetic background (Jaramillo et al., 2017). It also is compatible with evidence from studies of mutations in NLGN4 in humans, which found that the same mutation may be associated with different phenotypes within one family (Jamain et al., 2003; Laumonnier et al., 2004; Lawson-Yuen et al., 2008; Yan et al., 2005). As well as autism, NLGN4 associations have been described with intellectual disability, language disorder and Tourette syndrome (Lawson-Yuen et al., 2008; Yan et al., 2005).

**Hypothesis**

Our pre-planned analysis is designed to test the ‘double hit’ hypothesis.

The ‘double hit’ hypothesis: Neuroligins act as a multiplier of effects of neurexins

The notion of a ‘double hit’ aetiology has been proposed previously to account for cases where a microdeletion is inconsistently associated with neurodevelopmental disorder (Girirajan et al., 2010; Newbury et al., 2013): the idea is that a severe phenotype may be seen when there are two copy number variants or mutations, each of which may be relatively innocuous on its own. Here, we extend that idea to argue that the effect of altered neuroligin gene dosage may depend on the genetic background provided by autosomes (Bishop & Scerif, 2011). In this regard, it is of particular interest to note that neuroligin proteins form part of the same functional network as a group of presynaptic transmembrane proteins, known as neuroligins; their interactions play a key role in synaptogenesis (Hussain & Sheng, 2005). CNTNAP2 encodes a member of the neurexin superfamily whose polymorphisms have been associated with common forms of language impairment (Graham & Fisher, 2015), though the effect size is relatively small (Vernes et al., 2008). The role of the CNTNAP2 protein in developing brain is not fully understood, and it is likely to play multiple roles at different time-points. While early functional studies of the CNTNAP2 protein indicated that it localises to nodes of Ranvier in axonal membranes, it is now recognised to have key functions at the synapse (Lu et al., 2016; Zweier et al., 2009). This raises the possibility that a CNTNAP2 gene variant that has a modest effect in individuals of normal karyotype might have a much larger impact in the context of overexpression of a neuroligin. This hypothesis predicts that presence of an additional sex chromosome will amplify the impact of common genetic variants that have two characteristics: (a) they have been associated with DLD or ASD, and (b) they are in the same functional network as neuroligins. Figure 1 shows a schematic showing two genes of interest to our current study, CNTNAPs and Neurexins, interacting with neuroligins in the synaptic cleft.

**Methods**

We report how we determined our sample size, all data exclusions, all manipulations, and all measures in the study (Simmons et al., 2012).

**Power analysis and impact of ascertainment bias**

We aimed to recruit sufficient children with trisomies to detect an effect size of d = 0.5 for each copy of a given genetic variant on a phenotype, equivalent to a standardized regression slope of 0.25. The anticipated effect size is hard to judge, but the average impact of a sex chromosome trisomy on verbal IQ is more than one SD from the general population mean (Leggett et al., 2010), suggesting that if the trisomy acts as a multiplier of effects of autosomal variants, this effect could be large. When testing variants with a prior association with disorder, we can make a directional prediction. We aimed to recruit 150 children with trisomies, which would have given 94% power to detect a slope of 0.25 on one-tailed test. However, we recruited only 140 children and had missing data on some variables, so numbers, and consequently power, are lower than this. In addition, we have to take into account that the sample is not representative of children with sex chromosome trisomies, because around 50% had the trisomy discovered in childhood when developmental difficulties were being investigated (see below). We devised a simulation to check the impact of these factors...
Neurexins (such as NRXN1), neuroligins (such as NLGN4) and contactin-associated proteins (such as CNTNAP2) all form part of the synaptic scaffolding system.

Comparison group: Comparison data came from a sample of children aged from 6 years 0 months to 11 years 11 months who had completed the same test battery, who were taking part in a twin study of language and laterality (Wilson & Bishop, 2018a), and whose first language at home was English. Although twinning is a risk factor for early language delay, this effect appears to wash out with age, and by school age, genetic factors play a major role in the aetiology of language disorder (Bishop, 2006; Rice et al., 2018). In this sample, we aimed for an over-representation of twin pairs in which one or both twins had language or literacy problems that might be indicative of DLD. This was coded on the basis of parental response on a telephone interview: any mention of language delay, history of speech and language therapy, current language problems or dyslexia was coded as ‘parental concern’. We aimed to recruit 180 pairs selected on the basis of having language or literacy problems (60 MZ, 60 DZ opposite sex and 60 DZ same sex), and 60 unselected pairs (20 of each type): we fell short of this goal as seen in Figure 3. For the current analysis, we grouped together all twins, regardless of zygosity and parental concern,
Reason for testing
Time of diagnosis

Medical
Prenatal: N = 49
Postnatal: N = 34

Neurodevelopmental disorder
Postnatal: N = 59

Excluded: DNA quality
N = 4

Trisomy

XXX
N = 31

XYY
N = 23

EXXY
N = 22

Figure 2. Flowchart showing characteristics of children recruited to SCT group.

Parental concern re language:
Neither twin: N = 53 pairs

One or both twins: N = 141 pairs

Exclude: No DNA
N children = 2

N children = 16

Total N twin children

MZ:
N = 14 boys
44 girls

DZ:
N = 19 boys
27 girls

MZ:
N = 79 boys
45 girls

DZ:
N = 77 boys
65 girls

Figure 3. Flowchart showing characteristics of children recruited to comparison groups. Information about zygosity, gender and parental concern is shown for information, but was not used in the analysis. Because twins are not independent, the final sample was divided into two subgroups of 184 and 186 children respectively, each containing one member from each pair, selected at random. (Ns not equal because some twins had missing DNA from just one member of the pair).
and then divided them into two subsamples by selecting one twin from each pair at random, after excluding 18 cases with missing or insufficient DNA. This means we can replicate the analysis for twins with a diploid (typical) karyotype. Note that this replication sample is not independent, as the genotype for the MZ twins is the same in the two subsamples, and is related for DZ twins.

Some twin children had evidence of autism spectrum disorder (N = 15) or intellectual disability (N = 3), and twelve failed a hearing screen on the day of testing, although none of them had any known sensorineural hearing loss. For the current study, because we were interested in a broader phenotype than pure DLD, these cases were retained in the sample.

Test battery

Psychiatric evaluation. In an initial telephone interview, parents were asked about the child’s medical and educational history, including a question about whether anyone had diagnosed the child with a neurodevelopmental disorder such as ASD, developmental language disorder (DLD) or specific language impairment, dyslexia or dyspraxia. In addition, one or both parents were asked to complete the online Development and Well-being Assessment (DAWBA) (Goodman et al., 2000) in their own time. 84 parents of SCT cases and 133 parents of twins complied with this request. The DAWBA gives information on likelihood of the child meeting criteria for a range of psychiatric diagnoses; a final diagnosis is made by a trained rater who assimilates all the information and evaluates it against DSM5 criteria (American Psychiatric Association, 2013).

Language, literacy and cognitive assessments. All children were seen at home or in a quiet space in their school for a neurocognitive assessment, using the battery of language and nonverbal ability tests shown in Table 1. Hearing was screened in left and right ears using a DSP Pure Tone Audiometer (Micro Audiometric Corporation). The child was familiarised with the task of raising their hand on hearing a tone using 40 dB (HL) tones. They were then tested with 25 dB pure tones at frequencies of 500, 1000, 2000 and 4000 Hz. Louder tones were presented in 5 dB steps to establish a threshold at any frequency where a 25 dB tone was not detected. Children with an average threshold greater than 30 dB in the better ear were categorized as failing the screen. The battery also included tests of literacy: the Picture and Digit naming tests from the Phonological Assessment Battery (Frederickson et al., 1997), the Test of Word Reading Efficiency (Torgesen et al., 1999) and the Neale Analysis of Reading Ability -2 (Neale, 1999), but these are not included in the current analysis as there was much missing data from the youngest children. In addition, handedness and language laterality were assessed. Results from laterality assessments were unremarkable and are not considered further here (Wilson & Bishop, 2018a; Wilson & Bishop, 2018b).

Phenotypes

We will consider three quantitative phenotypes that range from a specific measure of a heritable language skill, through a more general language measure, to a measure that potentially indexes a wide range of neurodevelopmental problems:

A) Nonword repetition, which is regarded as a measure of phonological short-term memory. This was singled out as an individual measure because it has previously been identified in twin studies as a good marker of heritable language problems (Bishop et al., 1996) and has also been associated with genetic variants linked to language/literacy in the CNTNAP2, CMIP, ATP2C2, KIAA0319, and DCDC2 genes (Carrión-Castillo et al., 2017; Marino et al., 2012; Newbury et al., 2009; Newbury et al., 2011; Scerri et al., 2011; Vernes et al., 2008). In the current study, we used scaled scores from Repetition of Nonsense Words from the NEPSY (Korkman et al., 1998).

B) A general language factor derived from the four other language tests (Verbal Comprehension, Oromotor Sequences, Sentence Repetition and Vocabulary. As

### Table 1. Assessment battery.

<table>
<thead>
<tr>
<th>Instrument</th>
<th>Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Woodcock Johnson III Tests of Cognitive Abilities (Woodcock et al., 2007)</td>
<td>Verbal Comprehension</td>
</tr>
<tr>
<td>NEPSY: A Developmental Neuropsychological Assessment (Korkman et al., 1998)</td>
<td>Repetition of Nonsense Words</td>
</tr>
<tr>
<td>NEPSY: A Developmental Neuropsychological Assessment (Korkman et al., 1998)</td>
<td>Oromotor Sequences</td>
</tr>
<tr>
<td>NEPSY: A Developmental Neuropsychological Assessment (Korkman et al., 1998)</td>
<td>Sentence Repetition</td>
</tr>
<tr>
<td>Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999)</td>
<td>Vocabulary</td>
</tr>
<tr>
<td>Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999)</td>
<td>Block Design</td>
</tr>
<tr>
<td>Wechsler Abbreviated Scale of Intelligence (WASI) (Wechsler, 1999)</td>
<td>Matrices</td>
</tr>
<tr>
<td>Parental questionnaires</td>
<td></td>
</tr>
<tr>
<td>The Children’s Communication Checklist-2 (CCC-2) (Bishop, 2003)</td>
<td></td>
</tr>
<tr>
<td>The Social Responsiveness Scale (SRS) (Constantino, 2005)</td>
<td></td>
</tr>
</tbody>
</table>

For the NEPSY tests, norms extend only to age 12 yr 11 months, and so we used extrapolated scores, as documented in Appendix 1.
documented in Appendix 2, the decision to combined these measures into a single language factor was made after exploring the factor structure of the available phenotypic measures, with the goal of obtaining a reliable indicator of overall language function.

C) A global measure of burden of neurodevelopmental problems extending beyond language, including autistic features. This was developed on an ad hoc basis, using all available information from parental report (see Appendix 3).

DNA collection and analysis

Oragene kits (OG-500, DNA Genotek Inc, Ontario, Canada) were used to collect saliva for DNA analysis from children with SCTs and their parents and available twin pairs. DNA extraction was performed using an ethanol precipitation protocol as detailed in the standard protocol (DNA genotek). All extracted DNA was genotyped on the Infinium ‘Global Screening Array-24 (v1)’, which includes 692,824 SNPs including rare and common variations. Data were processed in the Illumina BeadStudio/GenomeStudio software (v. 2.03) and all SNPs with a GenTrain (quality) score of < 0.5 were excluded at this stage. All genotypes were further filtered using PLINK software v1.07 (Purcell et al., 2007); as recommended by Anderson et al. (2010), samples with a genotype success rate below 95% or a heterozygosity rate ±2 SD from the mean were removed, as were SNPs with a Hardy-Weinberg equilibrium P < 0.00001 or a minor allele frequency of less than 1%. Identity data within families and twin-pairs were used to exclude samples with unexpected gender or relationships. SNPs that showed an inheritance error rate > 1% or skewed missing rates between genotype plates were also excluded. Control data (CEU, YRI, CHB, JPT, Hapmap release #3) were employed through a principal component analysis within Eigenstrat (Price et al., 2006) to identify individuals with divergent ancestry. Sixteen individuals (6 twin pairs and 4 SCT cases) were identified as having African ancestry and 21 individuals (6 twin pairs and nine SCT family members) were identified as having Asian ancestry. Any SNPs that showed a significant association with non-European ancestry (P < 0.0001) were excluded. The final genome-wide dataset consisted of 500 individuals (370 twins, divided into two subgroups, and 130 independent SCT cases) and 451,093 autosomal SNPs with a genotyping rate of 99.78%.

Procedure

Ethical approval was obtained for the study in 2011 from the Berkshire NHS Research Ethics Committee (reference 11/SC/0096), and data collection started in August of that year, finishing in October 2016. Information sheets, consent forms and ethics approval documents are available on Open Science Framework. Families who had expressed interest in the study were interviewed by telephone to assess whether the child met inclusion criteria, and if so, an appointment was made to see the child at home or at school, depending on parental preference. Families were widely dispersed around the UK, including Northern Ireland, Scotland, Wales and England. During the course of recruitment a total of eight research assistants as well as the senior author were involved in assessing children. The assessment was conducted in a single session lasting between 2–3 hours per child, with breaks where needed.

Analysis plan

Study data are analysed using R software (R Core Team, 2016), with the main database managed using REDCap electronic data capture tools hosted at the University of Oxford (Harris et al., 2009).

Potentially, there is a very large number of genotypes and phenotypes that could be analysed to test our hypothesis, as well as different ways of creating subgroups. This consideration, coupled with the small sample size, makes it important to control adequately for multiple testing to guard against type I error (Grabitz et al., 2018). For this reason, we stored phenotype and genotype data separately and specified an analysis plan in detail, as reported here. The analysis of genotype-phenotype associations will be conducted after this plan is registered and peer-reviewed.

Subgroups

In our main pre-specified analysis we will treat all three trisomies together. This is because the double hit hypothesis postulates a common mechanism that would apply regardless of karyotype. If we find an association between genotype and phenotype, we will carry out exploratory analyses to consider whether this is moderated by karyotype. In particular, we will be in a position to test a prediction by Skuse (2018) that there is more variable expressivity of NLGN4X than NLGN4Y, which should lead to lower phenotypic variability in XYY compared to the other karyotypes. Note, however, that the ascertainment bias in the sample is problematic for making cross-karyotype comparisons, and the focus would have to be just on those who were not diagnosed because of neurodevelopmental problems (see Figure 2). This is a small sample and so there would be a high risk of missing a true effect (type II error).

Prioritising genotypes for analysis

We conducted a series of literature searches to prioritise autosomal genes for analysis, focusing on genes that had an association with childhood speech and language disorders and that were relevant for synaptic function (see Appendix 4). This led us to select two candidates; CNTNAP2 and NRXN1. Both of these genes are large (>1 MB) and included over 100 SNPs from the genotyping array. In order to avoid false positives with our small sample size, we chose to focus our analysis on regions that have previously been associated with neurodevelopmental disorder, analysing all genotyped SNPs (after quality control steps described above (see “DNA collection and analysis”) within these selected regions.

In CNTNAP2 (NM_014141), we focused on a region spanning exons 13-14 (chr7:147,514,390-147,612,852 (hg19)). This region includes a cluster of 9 SNPs previously associated with language disorder (Vernes et al., 2008; Whitehouse et al., 2011). We had direct genotype data for 22 SNPs across this region. In addition, we used imputation to obtain genotypes for SNPs
rs2710102 and rs7794745. These were the first SNPs reported to be associated with ASD, and represent the two main SNPs used in the majority of association studies in neurodevelopmental disorders (Alarcón et al., 2008; Arking et al., 2008). These two SNPs were not directly genotyped on the Illumina arrays and were therefore imputed for all individuals. Imputation was performed on the Michigan Imputation Server, an online server which generates phased and imputed genotypes using high-density reference panels. Variant Call Files (VCF) were uploaded for 15,936 SNPs genotyped on chromosome 7. Genotypes were phased within Eagle and imputed by Minimac against the Human Reference Panel hrc.r1.1.2016, which includes 64,940 haplotypes of predominantly European ancestry. In total, genotypes were generated for 2,289,829 SNPs across chromosome 7,513,970 of which had quality scores 0.9. The two SNPs of interest, rs2710102 and rs7794745 had quality scores of 0.9938 and 0.94127 respectively.

The second candidate is NRXN1 (NM_004801). Although this gene met our criteria of being relevant for both synaptic function and neurodevelopmental phenotypes, the studies showing this link involved deletions rather than common variants (Ching et al., 2010). A recent analysis of clinical microarray data showed that deletions in exons near the 5’ end of NRXN1 were specifically implicated in neurodevelopmental disorders (Lowther et al., 2017). Accordingly, we focused on 23 SNPs in this region of the gene chr2:51,141,501-51,280,121 (hg19). These SNPs covered exons 1–4 plus 20 Kb upstream (5’) of the gene as this region includes important regulatory sequences.

The SNPs within the chosen regions were filtered for minor allele frequency and Hardy Weinberg equilibrium (as outlined in DNA collection and analysis) but were not pruned for linkage disequilibrium. Previous simulations indicate that the GSCA method is not greatly affected by linkage disequilibrium (see Romdhani et al., 2014). Across the CNTNAP2 region, six pairwise combinations of SNPs had R2 > 0.8. Across the NRXN1 region, 8 pairwise combinations of SNPs had R2 > 0.8. Details of the SNPs included in the analysis and a table of correlations between SNPs are shown with Appendix material.

Statistical methods

CNTNAP2 and NRXN1 genes from the trisomy sample will be analysed for association with a latent variable based on the three phenotypes using a structural equation modelling (SEM) approach adapted for genetic analysis (Romdhani et al., 2015). The model specification for our analysis is shown in Figure 4.

Romdhani et al. (2015) used the Generalized Structured Component Analysis (GSCA) developed by Hwang and Takane (2004). This method uses component-based path modelling rather than traditional covariance-based SEM, allowing adequate model fit to be achieved when using smaller samples (Chin & Newstead, 1999; Tenenhaus et al., 2005). The measurement models in the SEM framework are not typical regression format using latent factors; instead they are fitted using alternating least squares to estimate the weights and parameters, which is similar to principal components analysis. The advantage of this approach is that it does not attempt to fit the whole covariance matrix for observed and latent variables, but rather fits a separate measurement model for the contribution of observed variables to each latent factor, as well as a covariance model for the latent factors. Hence, we do not estimate the contribution of individual SNPs in each gene to the phenotype; rather, their influence is represented via the weighted sum. Similarly, the latent phenotypic factor (termed Neuro in Figure 4) is a weighted sum of the three measures of the phenotype. We will estimate the significance of one direct pathway from the CNTNAP2 gene to the latent phenotype, and one from NRXN1 to the latent phenotype. This method thus gives a single estimate of the overall impact of SNPs in a region of the phenotype.

We conducted simulations that indicated that this method is feasible with the number of SNPs and phenotypes in our sample (see Appendix 5): the permutation method, used by this approach to effectively quantify the test statistic distribution, generates p-values independently for each path, and a correction is required to take this into account. Because the evidence of association of common variants was stronger for CNTNAP2 than for NRXN1, we used a sequential approach to setting a significance level (alpha), using a critical p-value of .05 to test the pathway from CNTNAP2 to the Neuro factor, and .025 for the pathway from NRXN1 to the Neuro factor.

In addition, we will do the same analyses with children from the two comparison samples.

We predict that one or both paths from CNTNAP2 and NRXN1 to the Neurodev factor will indicate significant association in the sex chromosome trisomy sample. We further predict that any associations in the comparison samples will be similar in direction, but smaller in size and may not reach statistical significance.

Any additional analyses of subgroups or phenotypes suggested by inspecting the data will be treated as exploratory and in need of replication in another sample.

Self-certification statement

The authors confirm that they had no prior access to the full dataset proposed for analysis.

The neurodevelopmental data has already been processed by DB and PT to derive the phenotypes to be used in the analysis; DFN has processed the DNA data separately to decide on the genotypes. The key tests proposed in this protocol will involve putting the two strands of data together, which has deliberately not been done, so that predictions can be derived without being aware of the data.
Figure 4. Structural equation diagram for analysis. The GSCA analysis estimates the path from each gene to the Neurodev factor, with significance calculated by permutation analysis.
Competing interests
No competing interests were disclosed.

Grant information
This work was funded by Wellcome Trust Programme Grants 082498, and European Research Council Advanced Grant 694189.

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Acknowledgements
We offer warmest thanks to the families who took part in the study, and school staff who helped facilitate assessment arrangements. We greatly appreciate the help we had with recruitment from Unique (the Rare Chromosome Disorder Support Group), the Klinefelter’s Syndrome Association, and staff from Clinical Genetics Centres in the following NHS Trusts: Royal Devon and Exeter, The Northwest London Hospitals, University Hospitals Bristol, Sheffield Children’s, Liverpool Women’s, University Hospital Southampton, Birmingham Women’s Hospital, Great Ormond Street Hospital for Children, Leeds Teaching Hospitals, Newcastle upon Tyne Hospitals, Oxford University Hospitals, Central Manchester University Hospitals, University Hospitals of Leicester, Nottingham University Hospitals, Belfast Health and Social Care Trust and the NIHR Clinical Research Network. We extend special thanks to Dr Gillian Borthwick, National Research Coordinator NIHR CRN: Genetics, for helping us navigate the complexities of working with multiple NHS sites under the NIHR UK Rare Genetic Disease Research Consortium Agreement. The study would not have been possible without the generous support of Prof Simon Fisher, Director of the Max Planck Institute for Psycholinguistics, who offered facilities for analysis of genetic material during the data collection phase of this study. We thank Katerina Kucera and Arianna Vino for their substantial work in processing DNA samples.

We thank the High-Throughput Genomics Group at the Wellcome Trust Centre for Human Genetics (funded by Wellcome Trust grant reference 090532/Z/09/Z) for the generation of the Sequencing data.

Finally, the study would not have been possible without the hard work and dedication of eight research assistants who travelled around the UK to conduct assessments, including Nicola Gratton, Georgina Holt, Annie Brookman, Elaine Gray, Louise Atkins, Holly Thornton, Sarah Morris and Eleanor Paine.

Supplementary material
Appendices 1–5 are available on Open Science Framework

References


Reference Source


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


Publisher Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


Publisher Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


Publisher Full Text


Reference Source


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text


Reference Source


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text

Wilson AC, Bishop DV: Sex chromosome trisomies are not associated with atypical lateralisation for language (Preprint). Dev Med Child Neurol. in press.


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text


PubMed Abstract | Publisher Full Text | Free Full Text


PubMed Abstract | Publisher Full Text | Free Full Text
Open Peer Review

Current Referee Status: ✔️ ✔️ ✔️

Version 2

Referee Report 16 May 2018

doi:10.21956/wellcomeopenres.15855.r32984

✔️ Armin Raznahan
Developmental Neurogenomics Unit (DNU), National Institute of Mental Health (NIMH), Bethesda, MD, USA

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Referee Report 04 May 2018

doi:10.21956/wellcomeopenres.15855.r33083

✔️ David H. Skuse
Behavioural & Brain Sciences Lab, UCL Great Ormond Street Institute of Child Health, London, UK

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.

Version 1

Referee Report 06 April 2018

doi:10.21956/wellcomeopenres.15031.r31601

David H. Skuse

Background:
The proposed analysis is predicated on a clinical observation that individuals with sex chromosome
duplication aneuploidies (47,XXX; 47,XXY; 47,YYY) have an increased rate of social communication dysfunction, including language. Intra-aneuploidy variability in the severity of that phenotype remains unexplained. Newbury and colleagues hypothesize that functionally significant variants of certain autosomal genes (CNTNAP2 and NRXN1), which have a well-established role in neurodevelopment, could serve to amplify the detrimental impact of sex chromosome aneuploidy.

The conceptual basis of this study can be found in Bishop and Scerif. The authors noted there are similarities between the phenotype of Klinefelter syndrome (XXY) and that of individuals who suffer from a Specific Language Impairment. They suggested that neuroligin genes on the sex chromosomes (specifically NLGN3 and NLGN4) would be expressed in excessive dosage in XXY syndrome, with a potential impact on the risk of neurodevelopmental syndromes (NDS). They proposed an epistatic interaction occurs between NLGN3/4 and a specific autosomal gene on chromosome 7 (CNTNAP2), variants of which had been shown to associate with language problems in previous studies. Neither genetic risk factor (sex-linked or autosomal) would be sufficient to cause SLI alone; not all XXY males have language problems, and many CNTNAP2 risk variants are inherited from an apparently unaffected parent. However, in combination (excessive NLGN3/4 expression plus risk variants of CNTNAP2), there could be some multiplicative impact of genotypic risk that resulted in a clinical (language-related) phenotype.

In their updated hypothesis, Bishop and her colleagues hypothesise that the sex-linked genetic risk factor in the aneuploidy is most likely to be NLGN4X. Not only is there an equivalent gene on the Y chromosome (NLGNY) but NLGN4X escapes X-inactivation (unlike NLGN3X), and therefore could be expressed in increased dosage in duplication syndromes such as XXY. Does NLGN4X/Y impact language development? A recent review of genomic influences on language development Graham and Fisher (2015) do not mention NLGN4, but of course that does not negate the possibility that major expression changes due to aneuploidy could have influence outside the normal range of their activity.

**Neuroligins:**
Neuroligins are post-synaptic transmembrane proteins and form complexes with neurexins (as shown in Figure 1). There are four neuroligins expressed in humans, NLGN1 (3q26.3), NLGN2 (17p13) on the autosomes and NLGN3 (Xq13.1) & NLGN 4 (Xp22.32) on the X chromosome.

NLGN4Y on the Y chromosome is almost identical structurally to NLGN4X. Neuroligins NLGN4X and NLGN4Y are both expressed in the brain. Whilst there is no specific study on the influence of X-chromosome duplication in risk of ASD, there is evidence that NLGN4Y overexpression has a role in the pathogenesis of ASD in XYY syndrome.

Note that there could be a difference in X and Y-linked neuroligins’ role in neurodevelopmental processes, both in terms of where in the brain the genes are expressed and in the degree to which they are expressed, despite the fact that the two structural variants of the NLGN4 gene are very similar to one another.

**Neurexins:**
Neurexins are in the same functional network as neuroligins, and form complexes with them at neuronal synapses. There are three NRXN genes, NRXN1, NRXN2 and NRXN3. Their trans-synaptic cell-adhesion role mediates essential signaling between presynaptic and postsynaptic domains. Mutations in neurexins have been widely implicated in cognitive impairment and predisposition to psychiatric disorders.
CNTNAP2 encodes a neurexin-like cell adhesion molecule (Caspr2) but the impact of specific CNTNAP2 genetic variants (and the functional integrity of Caspr2) on brain development is still not well understood. Interestingly, the expression of CNTNAP2 is regulated by the FOXP2 transcription factor (itself linked to speech/language disorders). Whilst the neurexin-neurexin transmembrane complex has substantial supporting evidence, Figure 1 implies there is an equivalent transmembrane connection between ‘CNTNAPs’ and Contactins/Catenins, but the evidence for that is weak. A recent review indicated the main role of contactin-associated proteins (CNTNAPs) is outside the synapse, in neuron–glia interactions in myelinated axons.

Experimental procedure:
The experimental procedure proposed is to examine genetic risk for a disordered language phenotype in an already collected unbiased sample of X-chromosome aneuploidies, encompassing girls with XXX trisomy, and boys with XXY and XYY syndromes. A comparison sample of twins participating in a study of language and laterality has been chosen (two separate samples of 184/186 children). The phenotypic assessment battery is comprehensive and sensibly treats phenotypes as quantitative on the whole (although the classification from the DAWBA psychiatric phenotypes is nominal).

Cognitive deficits and ASD risk in sex chromosome aneuploidies:
A recent review of sex-chromosome aneuploidies and their cognitive consequences points out that XXY and XYY boys tend to have lower verbal than nonverbal IQ. In XXX syndrome both nonverbal and verbal abilities tend to be equally impaired, and verbal skills are equivalent to those of XXY males. In fact, in all three aneuploidies the degree of verbal impairment is similar (0.7-1.0 SDS), which implies a common mechanism may be influencing language development. It is worth noting, however, that in 45,X (X0-Turner) syndrome there is no significant verbal IQ/speech/language deficit; it seems duplication of a sex-chromosome risk factor is more detrimental than haploinsufficiency in respect of speech/language deficits.

Both XXY and XYY syndromes are associated with a high risk of ASD, though not XXX. There is a substantially increased prevalence of ASD in association with X0 (up to 30%), which indicates a dissociation between the processes underlying ASD risk and language impairment in sex chromosome aneuploidy. To be specific, XXX females have relatively low risk of ASD despite language impairments that are equivalent in severity to XXY and XXY (where ASD risk is high): X0 females have a high risk of ASD but they do not typically experience any significant language impairments. There is a possible confounding impact of hormonal variables: XXX females are fertile, X0 females are not. Many XXY males are androgen deficient, XYY males are not.

Role of the Y-chromosome and partial X-inactivation:
Newbury and colleagues report that in their sample the proportion of XYY males with normal-range cognition was just 14% but in XXX syndrome was four times greater (55%). The figure for XXY males was intermediate (37%). Speech and language problems were also more common along the same gradient, 24% in XXX, 47% in XXY and 71% in XYY.

What could account for that gradation of risk?
First, NLGN4Y may have a different expression pattern in the brain from NLGNX and therefore over-expression may be more detrimental to key regions of the ‘social brain’. There is emerging evidence that Y-linked genes are subject to increased expression in duplication syndromes. Hence, the effect of having an additional Y chromosome may be particularly impairing, and different in its impact on brain...
development to an additional X-chromosome.

Second, there is evidence that although there is homology between the PAR1 genes in males and females (i.e. between X and Y allelic variants), in general pseudoautosomal genes in the PAR1 region of males are expressed at a higher level than those in females implying that aneuploidy for the Y chromosome could be more detrimental than aneuploidy for the X-chromosome.

Third, although a key element of the Bishop et al hypothesis is that NLGN4X escapes X-inactivation (XCI) and therefore will be over-expressed in XXX females and XXY males, the gene lies outside PAR1 and its escape from inactivation is only partial (average 26%). This partial expression is substantially lower than most X-linked genes that escape inactivation. More significantly, in respect to its implications for the Bishop/Scerif hypothesis, expression of NLGN4X from the 'inactive' X-chromosome is highly variable between typical females. The proportion of females in which escape of this gene from XCI has been recorded is only 38%. However, we must acknowledge that the situation in aneuploidies has not been studied.

An alternative hypothesis:

Accordingly, without the need to evoke an epistatic relationship between NLGN4X and Y and neurexin-associated SNVs, there is good evidence that within-karyotype variation could be explained on the basis of variability in expression of genes from the sex chromosomes, particularly within the female sample with XXX, and XXY males. On the basis of this alternative hypothesis, we would expect to find the least variability in language-related phenotype in the XYY male sample, and arguably the most variable phenotypes in the XXX sample. These data are already available to Newbury et al, hence the alternative hypothesis could easily be tested.

The role of CNTNAP2 and NRXN1 SNPs:

Coming onto the proposed interaction between the NLGNs and genotypic variants of NRXN1 and CNTNAP2, what is the evidence that the SNPs under investigation are themselves risk factors for language problems (with or without associated ASD)? Two particular SNPs of interest, according to the proposal, are rs2710102 and rs7794745 (both of which are in introns); they have been extensively researched in association studies of neurodevelopmental disorder.

Support for the Bishop et al hypothesis would derive from a finding that ‘risk variants’ of the autosomal genes CNTNAP2 and NRXN1 should have some phenotypic impact on brain structure/function in neurotypical populations; compelling evidence could be adduced from neuroimaging, comparing brain structure/function between populations with the variants of interest.

Uddén et al (2017) attempted to confirm findings made by earlier neuroimaging studies that claimed the rs7794745 variant of CNTNAP2 impacted grey matter in regions of the ‘social’ brain plausibly associated with ASD. Using a substantially larger sample of typical subjects they found that the risk variant associated with rs7794745 did link significantly to grey matter changes in the visual dorsal stream area of the occipital lobe. However, they could not find evidence of changes to volume changes in regions more strongly implicated in autistic traits (such as frontal or temporal areas). They express caution about potential misinterpretation of neuroimaging findings when small sample studies are performed.

The evidence supporting the other main CNTNAP2 variant of interest to Newbury et al, rs2710102, is not strong on the basis of evidence from typical populations. In a replication study, Dennis et al (2011) concluded that it may be a polymorphism that, when combined with others, could increase the risk for autism by enhancing the susceptibility to language disorders but there are countervailing findings too.
The relevance of NRXN1 to the hypothesis comes, as Bishop et al acknowledge, from evidence that exonic deletions at the 5’ end of the gene are associated with a substantially increased risk of neurodevelopmental dysfunction in terms of cognition and psychiatric risk\textsuperscript{17}. In the proposed study the region will be reasonably well covered with 23 SNPs.

On the other hand, if polymorphic variation in this gene were associated with neurodevelopmental problems in a non-clinical population (as opposed to deletion of exons in clinical cases) there should be some published evidence. Wang et al, (2018\textsuperscript{18}) recently found no association between two SNPs in NRXN1 and a clinical phenotype among a Chinese mixed population of ASD and controls (although neither of those SNPs is listed in Appendix 7 of the proposal). Where is the evidence that, in a typical population, SNP variants at the 5’ end of NRXN1 are associated with language development?

**Statistical analysis:**
Finally, there are novel statistical methods to be used for the analysis of this sample, which have been devised to get around the problem of limited power (due to small sample size). The Romdhani et al analysis claims to permit adequate model fit using smaller samples than conventional methods of association analysis. A weighted influence of a sum of SNPs is represented (Figure 4), implying the impact of a concatenation of variants within the critical region of candidate genes can be treated as a ‘weighted score’, irrespective of the contribution of individuals SNPs in each gene to the phenotype. This seems highly speculative, whereas the application of similar techniques to the phenotypic measures is uncontroversial. The Romdhani method’s application to multiple SNPs in a relatively small sample has, so far as I can ascertain, not been applied in other studies to date.

**Summary:**
Bishop and her colleagues are to be congratulated for the work they have done in assembling a relatively large sample of X-chromosome aneuploidies and for the careful phenotyping they have conducted to date.

Copy number variants (CNV), both duplications and deletions, can be detrimental to neurodevelopment and some (such as those affecting the 16p11.2 region) can impact on language development\textsuperscript{19}. It would have been worth checking the samples of aneuploidy collected by this team with a microarray for possible ‘second hits’\textsuperscript{20}. It is also the case that many potentially pathogenic autosomal CNV have variable expression phenotypically and may be inherited from an apparently unaffected parent. Polygenic background variation (as well as environmental factors) are often cited as a modifying factor, accounting for the variable penetrance\textsuperscript{21}.

In summary, it is frequently the case that inappropriate gene dosage syndromes involving duplication or deletion of critical regions of the genome to have variable phenotypic consequences on neurodevelopment between affected individuals. The proposal from Newbury et al aims to test a specific hypothesis, which entails making the assumption that the risk of a particular phenotype (detriment to language) is raised to a consistent(subclinical) threshold by X- or Y chromosome duplication. The biological mechanism is assumed to be NLGN4X/Y overexpression. The impact on an affected individual’s observable phenotype is, they propose, further modified by functional variants (SNPs) on the autosomal genes CNTNAP2 and NRXN1. It is these SNPs that contribute a range of variable risk (risk that contributes to the polygenic background of the general population) that could, in certain specific combinations, tip the balance toward a phenotype characterized by a clinically significant disorder of language.
The proposal that inter-aneuploidy variation on the sex chromosomes could be strongly influenced by NLGN4X/Y expression is a plausible one. There are novel elements to the Bishop et al hypothesis that are worth testing, in particular the possibility that Romdhani’s approach to measuring aggregated SNP risk in small samples is valid. However, on the basis of recently discovered evidence concerning the characteristics of homologous X- and Y-linked genes, it seems to me that intra-aneuploidy modifying factors on language and social communication skills are more likely to be associated with the characteristics of variable sex chromosome gene expression than with SNPs in the two autosomal candidates, CNPNAP2 and NRXN1.

Have the authors pre-specified sufficient outcome-neutral tests for ensuring that the results obtained can test the stated hypotheses, including positive controls and quality checks?

Partly.

References
8. Peñagarikano O: CNTNAP2 Mutations in Autism. 2016. 177-188 Publisher Full Text


**Is the rationale for, and objectives of, the study clearly described?**
Yes

**Is the study design appropriate for the research question?**
Partly

**Are sufficient details of the methods provided to allow replication by others?**
Yes

**Are the datasets clearly presented in a useable and accessible format?**
Yes

**Competing Interests:** No competing interests were disclosed.

**Referee Expertise:** Child Psychiatry, Genomic influences on Cognition & Behaviour, Sex chromosomes
I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.

Author Response 16 Apr 2018

Dorothy Bishop, University of Oxford, UK

This review gave an excellent overview of background material relevant to our study. Here we respond just to comments that raised questions about our planned approach.

P 13, para 1. The Graham and Fisher review provides a masterly overview of associations with language-related disorders found in GWAS, and we now add a mention of this. For two reasons, we adopted a different approach to those authors to identify candidates for analysis: first, we had a specific hypothesis that led us to focus on genes involved in synaptic function, and second, we were conscious that if we analysed a very large number of candidate SNPs, we would compromise our analysis, given that the small sample limited our statistical power. Accordingly, we took a systematic approach to homing in on candidates, searching for genes that met the dual criteria of being mentioned in association with language-related phenotypes, and being involved in synaptic function. This threw up some genes that have not traditionally been regarded as involved in language function because our search term picked up studies that included language phenotypes beyond SLI or dyslexia: notably autism or schizophrenia, where language is affected. These seemed worth exploring, because, as noted by Graham and Fisher, the boundaries between neurodevelopmental disorders are fuzzy, especially in the context of sex chromosome trisomies (SCTs), where a wide range of phenotypes can be seen.

P 14, para 4. We have modified our language to indicate that much about how CNTNAP2 operates is still not understood. Nevertheless, we argue that there is supportive evidence for a role of CNTNAP2 in synaptic development. See, for instance, Lu, Z., Reddy, M. V. V. S., Liu, J., Kalichava, A., Liu, J., Zhang, L., . . . Rudenko, G. (2016). Molecular architecture of Contactin-associated Protein-like 2 (CNTNAP2) and its interaction with Contactin 2 (CNTN2). Journal of Biological Chemistry, 291(46), 24133-24147. doi:10.1074/jbc.M116.748236

P 14, para 6-7. We would query the idea that there is such a clear division between risk for language impairment and risk for ASD in different trisomies. The study cited by Printzlau et al cites Bishop et al as finding no increase in risk of ASD in girls with XXX. That was so, but the ASD rates were based solely on parental report of a diagnosis, and it is increasingly recognised that girls are underdiagnosed. Furthermore, in that same study, parental report on CCC-2, which assesses autistic-like language features, the girls with XXX showed evidence of impairment in pragmatic skills, scoring well below the unaffected sibling group. We are in the process of writing up a detailed account of the phenotypic data that we have on our current sample; one conclusion that we will draw is that variation within a karyotype is considerably greater than variation between karyotypes. Overall, our data are consistent with the idea that all three karyotypes create an increased risk for language difficulties and ASD symptoms. The fact that language is typically unaffected in X0 girls, despite increased risk for ASD, is indeed interesting in suggesting that the impact of deletion differs from that of duplication. This is a point we think relevant for a later Discussion section.

P 15, para 2. The estimates of proportion of cases with normal-range cognition do not come from Newbury et al – they do correspond to figures reported by Bishop & Scerif that relate to the Bishop
et al (2011) sample. However, the figures do *not* refer to ‘normal range cognition’ but rather to cases where there was no neurodevelopmental or educational diagnosis, from parental report. Many of the problems that were reported concerned behaviour rather than cognition, notably ADHD and ASD. For data on cognitive function, a better source is the systematic review by Leggett et al (2010), which focused on neonatally identified samples and found verbal ability impaired in all three karyotypes, but lower nonverbal IQ in the XXX females. So while we agree that there is potential interest in comparing the three karyotypes as they could potentially throw light on underlying mechanisms, we think it is justifiable to treat them together for this specific analysis to ascertain whether genetic background can explain phenotypic variation in language phenotypes.

P 15, paras 3-4. These potential explanations for greater impact of NLGN4Y compared to NLGN4X are plausible and interesting. If the genetic background from candidate genes exerts any impact, we can consider whether any further variance is explained by karyotype (see Analysis Plan, Subgroups), but note we are underpowered to detect karyotype differences unless they are very large, and the ascertainment bias in our sample means that we cannot meaningfully interpret main effects on phenotype (see end para 1, participants).

P 15, para 4. We included genes from PAR1 in our literature search and found none was associated with language phenotypes.

P 15, para 5-6. We agree that variation in escape from inactivation for NLGN4X could explain phenotypic variation; we are not in a position to assess variable gene expression, but as Skuse notes, if this mechanism is important, then we should see greater phenotypic heterogeneity in karyotypes with an extra X than in the XYY group. We can indeed report those results, but we again would note that interpretation would have to be cautious, given the ascertainment bias. A higher percentage of boys with XYY had the trisomy identified because they were having problems (see Figure 2, p 5); if we restrict consideration only to the unbiased group, the numbers are very small. We have now mentioned this alternative hypothesis in the text, but also pointed out that testing it with our data carries a high risk of type II error.

P 15-16. SNPs as risk factors
The exact definition of “risk variant” is a very hard thing! As Skuse points out, even when particular SNPs are robustly implicated, we often see flipping between risk alleles. Our approach departs from that usually adopted, because we are not convinced that a focus on specific SNPs will be illuminating. We therefore have chosen to perform an analysis of all SNPs across risk regions capturing maximum information across these regions rather than relying on specific SNPs which may well be proxies for functional variants.

P 16, para 3 and 4. We agree with the reviewer that prior literature on SNPs associated with language has not always led to convincing results and strong claims have sometimes been based on woefully underpowered studies. Our review of the neurogenetics field (Grabitz et al, 2017) is one reason why we decided we should preregister our proposed analysis, as it is just too easy to find something in a dataset if you have enough potential genotypes and phenotypes and a small sample.

P 16, para 5. We agree that the predicted association of NRXN1 with language phenotypes is a long shot; the reviewer is correct – we are aware of no prior reports of association with common variants. This is why this is our second hypothesis. However, while the evidence for language association is far less convincing than for CNTNAP2, in terms of its role in neural circuits, NRXN1
is a strong candidate. We noted in the text that there was evidence that deletions near the 5’ end of NRXN1 were specifically implicated in neurodevelopmental disorders (Lowther et al., 2017). Potentially, the SNP analysis will tag variations and deletions in this region: if a SNP falls within a region of deletion then we would expect to see a loss of heterozygosity across that region which will alter the allele frequencies of the SNPs accordingly. Furthermore, surrounding SNPs may be in LD with the presence of deletions. To be honest, we’ll be surprised if we find anything with NRXN1, but it emerged from our literature search and we think it’s worth looking at.

P 16, para 7: comment on the analysis. We agree that our approach to analysis has not been applied to other studies in this area, and this is a high risk approach. We decided to adopt this method because it appeared to be the best way of optimising power in an analysis with such a small sample. If we do find that it reveals results of interest, it would suggest it might be worth using this method more widely: we do think the traditional approach of looking at many SNPs and correcting for the number of SNPs may obscure genuine associations.

P 17. Para 1: We will be looking at CNVs in a separate analysis; this is a natural extension to the work of Newbury and Simpson, who have already looked at samples with SLI. Although CNVs can be looked at in the context of a double hit model, this is complicated by an alternative possibility, that those with a trisomy may also carry an additional burden of CNVs. We feel therefore that this merits a separate report.

**Competing Interests:** None

Referee Report 26 March 2018

doi:10.21956/wellcomeopenres.15031.r31717

Beate St Pourcain

Language and Genetics Department, Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Newbury and colleagues submitted a Wellcome Open Research Stage 1 registered report studying neurodevelopmental outcomes in children with sex chromosome trisomies (SCTs), who show an increased rate of neurodevelopmental difficulties involving language. Adopting a double-hit hypothesis, the authors assume that the presence of an additional sex chromosome will amplify the impact of common autosomal genetic variants residing within genes involved in synaptic functionality. The authors specifically focus on the functional network of neuroligins involving CNTNAPs and Neurexins, as encoded by the selected candidate genes CNTNAP2 and NRXN1 and will study common variation at these loci.

I have the following comments:

1. XXX females and XXY males (Klinefelter's syndrome) show X chromosome inactivation, while XYY males do not, suggesting additional mechanisms that may affect gene dosage. Also there seems to be a possibility for differences among SCTs groups with respect to the proportion of children affected by a history of speech and language-therapy (24% for XXX girls vs 71% for XYY boys). Thus, the statement that common mechanism will apply regardless of karyotype may not be as unequivocally applicable as assumed. Could the authors clarify their statements on karyotype? Why could individuals with karyotypes involving X chromosome inactivation not be lumped
together as part of a sensitivity analysis? Also, the authors may consider whether the severity of the outcome depends on common variation that influences (random) X chromosome inactivation e.g. within \textit{XIST} and \textit{TSIX} genes.

2. The authors assume that risk to developmental language disorder (DLDs) as carried by common variation, observed in individuals with a normal karyotype, is exacerbated in the context of SCTs. This is a novel and very provocative hypothesis that can provide deeper insight into the aetiology of DLDs. However, it is also known that common variation contributing to risk of developing ASD in children with and without \textit{de novo} structural variation might act through slightly different pathways (Weiner et al., 2017\textsuperscript{1}). This suggests that also common variation acting on the background of chromosomal abnormalities may invoke partially different aetiological mechanisms. If possible, within the realm of adequate power, I would therefore recommend extending the scope of common variation tested. Nonetheless, the hypotheses proposed in this report are biologically plausible and, more importantly, testable with adequate power and will provide an improved understanding of the genetic mechanisms contributing to DLD.

3. The statistical analysis using partial least squares path modelling is sophisticated and adequate to detect joint common variant effects. The description of variants selected could be slightly improved, however, especially with respect to handling LD patterns.

In conclusion, this report proposes a very interesting and valuable study that can provide insight into the aetiological mechanisms underlying DLDs using a sophisticated and thoughtful analysis approach studying a DLD high-risk population of children with SCTs. Investigating established candidate genes has furthermore implications for the understanding of the phenotype in the wider population of children with DLDs without karyotypic abnormalities.

References

\textbf{Is the rationale for, and objectives of, the study clearly described?}
Yes

\textbf{Is the study design appropriate for the research question?}
Partly

\textbf{Are sufficient details of the methods provided to allow replication by others?}
Yes

\textbf{Are the datasets clearly presented in a useable and accessible format?}
Yes

\textbf{Competing Interests:} No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard.
Many thanks for the thoughtful comments on our protocol.

1. The point about possible variation between karyotypes agrees with one of the main points raised by Skuse (see above). It makes sense that there are some variations across karyotypes that will need another explanatory account. Nevertheless, we think it worth starting by testing a general mechanism, for two reasons. First, language impairment seems to be more like height – just as there is a tendency to be tall in all three karyotypes, so there is a tendency to have poor language skills. On our quantitative measures of language, the three groups do not look very different (though there is large variance around means). So there appears to be a general influence affecting all three karyotypes, even if there are karyotype-specific effects superimposed. NLGN4X does, at least in part, escape inactivation, so provides a plausible mechanism. Second, we have limited statistical power. If we take seriously correction for multiple comparisons, then we can either investigate several potential explanations with a stringent alpha, or reduce the number of tests we do and have a less stringent alpha. We have opted for the latter approach.

2. As noted in the text, focusing on two genes so we can optimise power will not preclude us from doing further exploratory analyses, which would include investigating karyotype-specific effects, but we would then need to replicate findings in another sample. It would, of course, be great to extend the analyses and not just confine it to two candidates, but we would rapidly lose power if we applied appropriate corrections. More minor point: inclusion of variants on XIST and TSIX would pose problems as they are on the X chromosome.

3. Improved account of LD patterns – please see response to Raznahan below.

Competing Interests: None
hypothesis is based on the idea that gene dosage abnormality of Neuroligins will multiply the effects of functional variation in neurexins (CNTNAP2 and NRXN1) on language outcomes.

The ideas motivating this report are biologically plausible, although the cited examples of epistasis do not involve copy number variations. The effects of gene dosage variation on gene function, and on the potential for interaction between functional variations in other genes are not necessarily the same as the effects of SNPs or SNVs.

The proposed methods are suitable for testing the specific hypotheses to be tested. The authors take a thorough approach to the challenging issues to ascertainment bias in studies of SCT - adjusting their methods for power analysis and participant selection. The comparison group is twins rather than singletons, but the authors give reasons for why this might not bias their analyses. The behavioral/developmental phenotypes to be used are fit for purpose. The authors use transparent and systematic criteria for selecting their two autosomal genes of interest, but I could not locate details for their selection of regions within the genes besides the statement that they “focused their analysis on regions that have previously been associated with neurodevelopment disorder”. The genome addresses provided suggested that they are using linkage peaks perhaps, or CNV boundaries? It might be good to clarify this. Also once the regions are set, was linkage disequilibrium taken into account as part of selecting SNPS from target regions? Were SNPs entered into the SEM as linear ordinal variables, or were more complex SNP effects considered? It would be helpful to clarify these issues.

In summary, notwithstanding the suggested additions of information above, this report describes a sound test of a plausible hypothesis for an important observation - marked inter-individual variability in phenotype severity within SCT. Addressing this question in SCT would have implications for the more general phenomenon of phenotypic variability in copy number variation syndromes.

Have the authors pre-specified sufficient outcome-neutral tests for ensuring that the results obtained can test the stated hypotheses, including positive controls and quality checks?

Partly.

Is the rationale for, and objectives of, the study clearly described?
Yes

Is the study design appropriate for the research question?
Partly

Are sufficient details of the methods provided to allow replication by others?
Yes

Are the datasets clearly presented in a useable and accessible format?
Yes

Competing Interests: No competing interests were disclosed.

I have read this submission. I believe that I have an appropriate level of expertise to confirm that it is of an acceptable scientific standard, however I have significant reservations, as outlined above.
Thank you for your encouraging comments. Here are our responses to specific points.

P 21, para 5. Effects of gene dosage variation may not be same as SNPs or SNVs.
This is a fair point, though previous reports have implicated the double-hit hypothesis in relation to copy number variants encompassing other genes (for example, Newbury et al 2013 (PMID: 22909776), Girirajan et al 2009 (PMID: 20154674), and we now cite this work. In addition to the evidence from SNVs and biological plausibility, we feel this model is worth exploring in sex chromosome trisomies.

P 21, para 5 – need for more details for selection of regions
The regions selected were based upon previous association analyses of CNTNAP2 and NRXN1 in relation to language disorder. We have now updated the “Prioritising genotypes for analysis” section to make this clearer.

P 21, linkage disequilibrium taken into account?
We have now added this to the text:
“The SNPs within the chosen regions were filtered for minor allele frequency and Hardy Weinberg Equilibrium (as outlined in DNA collection and analysis) but were not pruned for linkage disequilibrium. Previous simulations indicate that the GCSA method is not greatly affected by linkage disequilibrium (see Romdhani et al 2014). A table of correlations between SNPs can be found in the Appendix. Across the CNTNAP2 region, six pairwise combinations of SNPs had R² > 0.8. Across the NRXN1 region, 8 pairwise combinations of SNPs had R² > 0.8.”

Competing Interests: None