BMC Genomics

RESEARCH ARTICLE Open Access

Identification of candidate intergenic risk loci in autism spectrum disorder

Susan Walker[1] and Stephen W Scherer[1,2*]

Abstract

Background: Copy number variations (CNVs) and DNA sequence alterations affecting specific neuronal genes are established risk factors for Autism Spectrum Disorder (ASD). In what is largely considered a genetic condition, so far, these mutations account for ~20% of individuals having an ASD diagnosis. However, non-coding genomic sequence also contains functional elements introducing additional disease risk loci for investigation.

Results: We have performed genome-wide analyses and identified rare inherited CNVs affecting non-genic intervals in 41 of 1491 (3%) of ASD cases examined. Examples of such intergenic CNV regions include 16q21 and 2p16.3 near known ASD risk genes *CDH8* and *NRXN1* respectively, as well as novel loci contiguous with *ZHX2*, *MOCS1*, *LRRC4C*, *SEMA3C,* and other genes.

Conclusions: Rare variants in intergenic regions may implicate new risk loci and genes in ASD and also present useful data for comparison with coming whole genome sequence datasets.

Keywords: Autism spectrum disorder, Copy number variation, Non-coding DNA

Background

Newer genomic technologies like high-resolution microarrays and next generation exome sequencing have enabled the identification of many clinically relevant genetic variants for both Mendelian and complex disorders. Yet for many conditions the identified genes account for only a proportion of heritability. This observation coupled with the recognition of the functional relevance of non-genic regions [1] target these genomic segments as candidates for investigation for a role in disease.

ASD encompasses a range of neurodevelopmental disorders characterised by social impairment, communication difficulties and restricted, repetitive behavioural patterns. ASD, which is clinically and genetically heterogeneous, demonstrates high heritability, familial clustering and ~4:1 male to female bias. While there has been progress identifying risk genes, most are still unknown [2]. Analyses of rare (<1% population frequency) CNVs, insertions and deletions (indels) and point mutations have most convincingly identified synaptic genes such as members of the Neuroligin (*NLGN3*, *NLGN4)* [3], Neurexin (*NRXN1* [4], *NRXN2* [5], *NRXN3* [6]), SHANK (*SHANK1* [7], *SHANK2* [8], *SHANK3* [9]) families and Gephyrin [10] as highly-penetrant risk loci [2]. ASD subjects with multiple genetic risk factors for ASD and associated medical conditions are also known [11]. In addition, there are a few examples of mutations in ASD cases identified in non-genic segments of DNA [12] and non-coding RNAs [13]. Similar findings are even better documented in studies of intellectual disability [14,15], which is observed in ~40% of cases of ASD. Focusing on the intergenic intervals of the genome, we performed a systematic genome-wide investigation to identify rare CNVs enriched in cases compared with controls [16] to identify known and novel ASD susceptibility loci.

Methods

A collection of 1491 unrelated ASD cases were genotyped using either the Illumina 1M (993) or the Affymetrix SNP 6.0 platforms (498). The ASD subjects, all diagnosed using gold-standard instruments including Autism Diagnostic Interview and Autism Diagnostic Observation Schedule, are described elsewhere [16,17]. Informed written consent was obtained from all participants, as approved by the Research Ethics Boards at The Hospital for Sick Children

* Correspondence: Stephen.Scherer@sickkids.ca
[1]Program in Genetics and Genome Biology, The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario M5G 1L7, Canada
[2]McLaughlin Centre, University of Toronto, Toronto, Ontario M5S 1A1, Canada

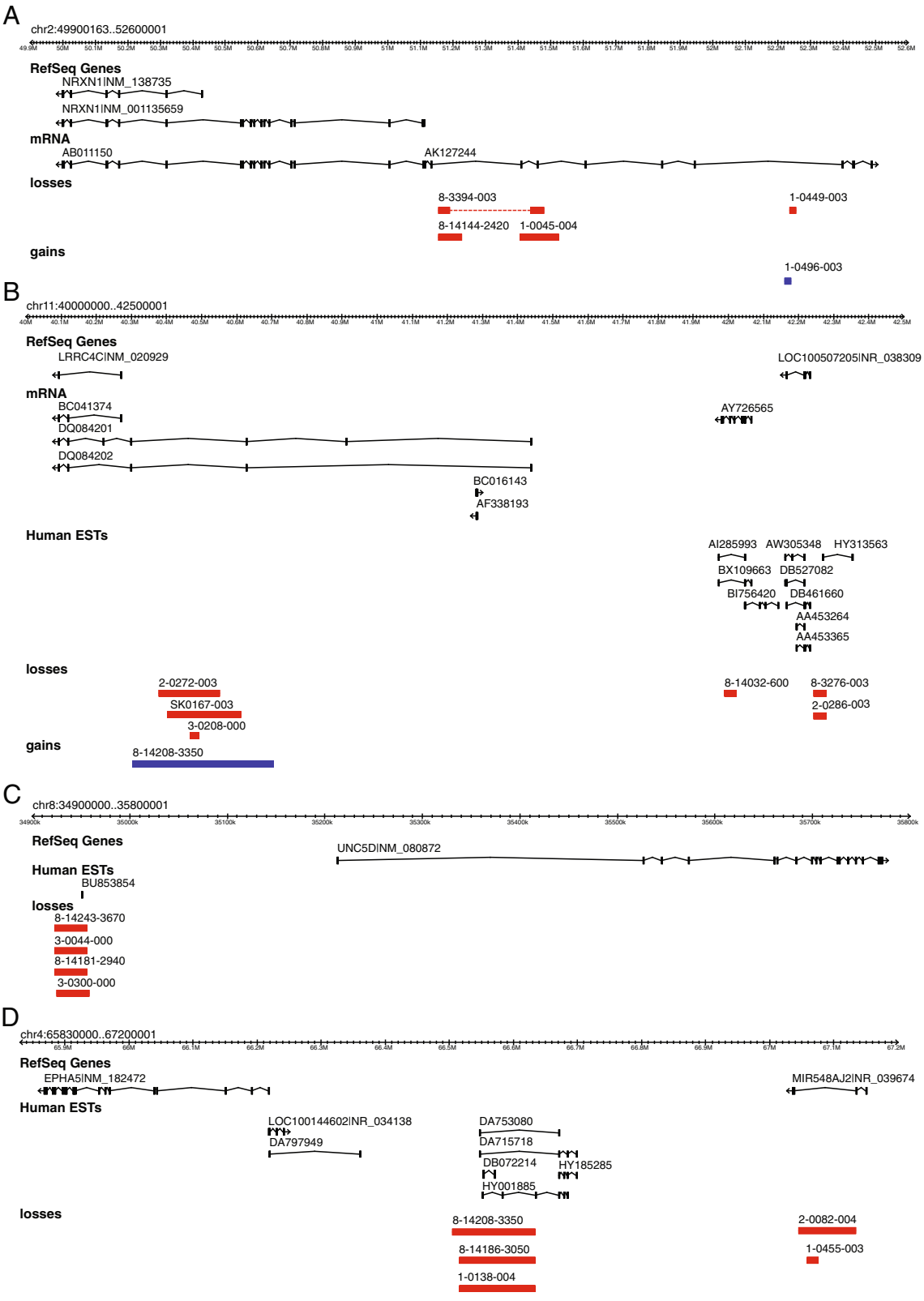

Figure 1 (See legend on next page.)

(See figure on previous page.)
Figure 1 Genome browser views of ASD specific CNVs at A) 2p16.3 B) 11p12 C) 8p12 and D) 4q13.1. In each case, representative isoforms of known RefSeq genes, mRNA and/or Expressed Sequence Tags are shown. Deletions and duplications are represented by red and blue bars, respectively. In Figure 1A) a dashed line indicates a diploid region located between two adjacent deletions in the same individual. Additional browser views from other loci shown in Table 1 are included in Additional file 1 A-J. In all cases where parental DNA was available, the CNVs shown were found to be inherited. Additional case SK0167-003 found in Marshall *et al.* [19].

and McMaster University. For controls, 1287 samples from the SAGE cohort were genotyped on with the Illumina 1M and 1234 samples from the Ottawa Heart Institute (OHI) and 1123 from the POPGEN collections were genotyped on the Affymetrix SNP 6.0. CNV discovery was performed using previously described pipelines [16-18]. Three CNV detection tools were used for each platform (Birdsuite, iPattern and Genotyping Console for Affymetrix 6.0 and iPattern, QuantiSNP & PennCNV for Illumina 1 M). A subset of CNVs in both cases and controls were considered rare if they were present in <1% of the overall dataset and these were further analysed if they failed to intersect or fall within a known gene (according to the NCBI Reference Sequence (RefSeq), August 2011). Rare genic CNVs identified from these data have been reported previously and from these data approximately 10% of cases carry a *de novo* or rare inherited CNV thought to contribute to ASD in that individual [16,17,19,20]. All CNVs discussed were validated where DNA was available using independent laboratory methods such as long range or quantitative PCR and the mode of inheritance determined (Additional files 1 and 2).

Results and discussion

Microarray data from a cohort of 1491 unrelated ASD probands were analysed for rare copy number variants as described previously [16,17] and CNVs falling outside of known coding sequence were identified. A total of 212 non-coding genomic regions were determined as harboring overlapping CNVs in two or more unrelated ASD cases that were absent in control samples. Each region was examined for plausible biological function by comparison with multiple databases. Data was collated for evidence of expressed sequences from mRNA or EST data at GenBank or evolutionary conservation as well as functional predictions from the VISTA enhancer browser (http://enhancer.lbl.gov/) and Rfam (http://rfam.sanger.ac.uk/). The Database of Genomic Variants (http://dgvbeta.tcag.ca/dgv/app/home) was used to eliminate additional regions as non-ASD specific CNVs and regions with >80% masked as repetitive sequences were removed. Loci were also prioritised as being of potential clinical significance in ASD due to proximity to genes considered known or candidate ASD risk genes [17].

Fifteen intergenic regions emerged as plausible candidate ASD risk loci and in all instances the defining CNV events were inherited. In one of these regions, an additional case (SK0167-003) was found with an overlapping CNV described by Marshall *et al.* (2008) [19] (Table 1, Figure 1 and Additional files 1 and 2). In 14 of 15, the intergenic interval identified has not been described before and in three regions the CNV neighboured a known ASD gene, namely, *CDH8* [21], *C3orf58* [22] and *NRXN1* [4]. In the case of the *NRXN1* gene, upstream CNVs found in five individuals impact the same mRNA (AK127244) reported elsewhere with a CNV in a family with ASD (Table 1, Figure 1A) [23]. Examples of other intergenic CNVs identified highlight regions at 8q24.12 upstream of *ZHX2*, 6p21.2 upstream of *MOCS1*, 11p12 upstream of *LRRC4C* (Figure 1B) and 7q21.11 upstream of *SEMA3C*, as putative novel ASD rearrangements. In one case (8-14208-3350), deletions were identified at three separate loci; 4q13.1 upstream of *EPHA5*, 11p14.3 upstream of *LUZP2* and 11p12 upstream of *LRRC4C* and another case (3-0496-003) carried a 46, XXY sex chromosome imbalance. Other CNVs found in these 41 cases are shown in Additional file 3 and any or all of these may be contributing to the genetic load for ASD [11,17]. Interestingly, all the CNVs identified through our analysis are inherited events. The significance of this observation is still to be determined but suggests incomplete and/or variable penetrance of phenotype, which is something often observed in ASD [6,7,17].

The mechanism of action of these rare CNVs in the pathogenesis of ASD could be (i) through altering the necessary copy number or positional context of key DNA sequence elements required for regulating the proper expression of nearby genes [1], (ii) affecting still undiscovered genes or non-coding RNAs residing in the CNV regions and (iii) disrupting uncharacterized isoforms of the adjacent annotated genes. In the first scenario, we find CNVs both upstream (e.g. *UNC5D* (Figure 1C), *MOCS1*, *ASTN2*, *SEMA3C*, *ZHX2*, *LUZP2*, *CDH8*) and downstream (*C3orf58*, *RXRA*, *MRGPRD*) of known ASD risk genes and putative novel loci. For at least three regions (4q13.1, 6p21.2 and 11p12 (shown in Figure 1D, Additional file 1C and Figure 1B respectively)), our CNV mapping data in fact identify two distinct clusters of CNVs at the same locus, all overlapping spliced ESTs and thus with a possible regulatory role. Secondly, three independent CNV deletions interrupting a collection of spliced expressed sequenced tags approximately 330 kb proximal to *EPHA5* highlight a potentially newly discovered ASD

Table 1 ASD specific CNVs in intergenic regions

Locus	Gene	Sample	CNV	Start	End	Size	Furthest distance from gene	Bin
2p16.3	*NRXN1* AK127244 mRNA	1-0045-004	loss	51405882	51524684	118802	1124	ii
		8-3394-003	loss	51439897	51479683	39786		
		8-3394-003	loss	51157414	51189362	31948		
		8-14144-2420	loss	51157414	51225851	68437		
		1-0496-003	gain	52220120	52238172	18052		
		1-0449-003	loss	52237072	52253660	16588		
3p22.3	*ARPP21*	2-1213-003	loss	34984049	35102773	118724	563	ii
		3-0100-000	gain	35086691	35094736	8045		
3q24	*C3orf58 ZIC1, ZIC4*	1-0007-003	loss	146168760	146934953	766193	1383 1955, 1979	i
		8-3093-004	loss	146575437	146631141	55704		
4q13.1	*EPHA5*	8-14208-3350	loss	66505324	66633530	128206	840	i
		8-14186-3050	loss	66515708	66633530	117822		
		1-0138-004	loss	66515708	66633530	117822		
		2-0082-004	loss	67045815	67134170	88355		
		1-0455-003	loss	67058506	67075558	17052		
6p21.2	*MOCS1*	3-0139-000	gain	40021898	40078515	56617	168	i or ii
		2-0139-003	gain	40023327	40062155	38828		
		1-0381-003	loss	40174188	40209324	35136		
		2-1368-003	loss	40174188	40210694	36506		
7q21.11	*SEMA3C*	8-6258-03	loss	80431202	80512022	80820	96	i
		1-0345-005	loss	80482597	80517630	35033		
8p12	*UNC5D NRG1*	8-14243-3670	loss	34923482	34956067	32585	256 2183	i
		3-0044-000	loss	34923482	34956067	32585		
		3-0300-000	loss	34925149	34957854	32705		
		8-14181-2940	loss	34923482	34956067	32585		
8q24.13	*ZHX2*	8-3317-003	gain	123572785	123625681	52896	237	i or ii
		3-0186-000	loss	123583028	123639417	56389		
9q33.1	*ASTN2*	8-3055-004	loss	119254497	119374796	120299	98	i
		3-0115-000	loss	119314967	119319559	4592		
9q34.2	*OLFM1 RXRA*	2-1272-003	gain	136479329	136604233	124904	508 8	i
		2-1189-003	gain	136480334	136598491	118157		
11p14.3	*LUZP2*	8-14175-2820	loss	24177612	24316053	138441	160	i or ii
		8-14059-1020	loss	24262511	24303132	40621		
		8-14208-3350	loss	24262511	24303132	40621		
11p12	*LRRC4C*	8-14208-3350	gain	40304880	40703298	398418	196	iii
		2-0272-003	loss	40379668	40550356	170688		
		SK0167-003	loss	40417554	40610400	192846		
		3-0208-000	loss	40468058	40492541	24483		
11p12	*LRRC4C*	8-14032-600	loss	41990280	42021250	30970	1738	i or ii
		8-3276-003	loss	42243624	42279094	35470		
		2-0286-003	loss	42243624	42279094	35470		
11q13.2	*MRGPRD*	4-0023-003	loss	68486121	68493638	7517	10	i
		2-1075-003	loss	68486121	68500238	14117		
16q21	*CDH8*	8-14251-3750	loss	61650435	61787984	137549	1030	i or ii
		2-1175-003	loss	61658675	61755232	96557		

Location and size of all CNVs discovered are listed with the proposed associated candidate gene. Bin denotes possible mechanism of action by i) altering sequence elements required for regulating expression of neighboring genes ii) affecting undiscovered genes or non-coding RNAs iii) disrupting uncharacterised isoforms of adjacent genes. Genome browser views of all loci are shown in Figure 1 and Additional file 1. All pedigrees are shown in Additional file 2. Additional sample SK0167-003 identified in reference [19].

risk gene (Figure 1D). Finally, longer isoforms of *LRRC4C* likely exist given the discovery of mRNAs DQ084201 and DQ084202. There are, of course, other functional DNA elements or modifications that need to be considered [24] as the mapping resolution increases.

Conclusions

Given the challenges faced in interpreting the clinical significance of multitudes of genetic variants found in for example, whole genome sequencing [25], accruing evidence across multiple studies will advocate loci outside of known genes or other regulatory elements for further study, particularly for rare variants. In this light, these data provide a useful resource for comparison as new data sets of both CNVs and nucleotide-level variants become available to help fine-map additional discover new ASD risk loci. This general research strategy can also be applied to other disease gene studies.

Additional files

Additional file 1: Genome Browser views of loci with ASD specific CNVs.

Additional file 2: Pedigree structure for all families listed in Table 1.

Additional file 3: Table of all rare CNVs detected in the individuals described herein.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

SW and SWS conceived the project and wrote the manuscript. SW designed the analysis, interpreted the data and conducted laboratory validation experiments. Both authors read and approved the final manuscript.

Acknowledgements

Supported by NeuroDevNet, Genome Canada, the Ontario Genomics Institute, The Government of Ontario, the Canadian Institute of Health Research (CIHR), the Canadian Institute for Advanced Research, The McLaughin Centre, the Canadian Foundation for Innovation and the Ontario Ministry of Research and Innovation. SW holds a joint CIHR Autism Research Training and NeuroDevNet post-doctoral Fellowship. SWS holds the GlaxoSmithKline-CIHR Chair in Genome Sciences at the University of Toronto and The Hospital for Sick Children. We thank The Centre for Applied Genomics for technical contributions. We also acknowledge the assistance of Ines Sousa, Astrid Vicente, Alistair Pagnamenta, Richard Holt, Anthony Monaco, Catalina Betancur and Sylvia Lamoureux for assistance with CNV validation experiments.

Received: 22 January 2013 Accepted: 20 July 2013
Published: 24 July 2013

References

1. Klopocki E, Mundlos S: **Copy-number variations, noncoding sequences, and human phenotypes.** *Annu Rev Genomics Hum Genet* 2011, **12**(1):53–72.
2. Devlin B, Scherer SW: **Genetic architecture in autism spectrum disorder.** *Curr Opin Genet Dev* 2012, **22**(3):229–237.
3. Jamain S, Quach H, Betancur C, Rastam M, Colineaux C, Gillberg IC, Soderstrom H, Giros B, Leboyer M, Gillberg C, *et al*: **Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism.** *Hum Genet* 2003, **34**(1):27–29.
4. Szatmari P, Paterson AD, Zwaigenbaum L, Roberts W, Brian J, Liu XQ, Vincent JB, Skaug JL, Thompson AP, Senman L, *et al*: **Mapping autism risk loci using genetic linkage and chromosomal rearrangements.** *Nat Genet* 2007, **39**(3):319–328.
5. Gauthier J, Siddiqui T, Huashan P, Yokomaku D, Hamdan F, Champagne N, Lapointe M, Spiegelman D, Noreau A, Lafrenière R, *et al*: **Truncating mutations in NRXN2 and NRXN1 in autism spectrum disorders and schizophrenia.** *Hum Genet* 2011, **130**(4):563–573.
6. Vaags Andrea K, Lionel Anath C, Sato D, Goodenberger M, Stein Quinn P, Curran S, Ogilvie C, Ahn Joo W, Drmic I, Senman L, *et al*: **Rare deletions at the neurexin 3 locus in autism spectrum disorder.** *Am J Hum Genet* 2012, **90**(1):133–141.
7. Sato D, Lionel Anath C, Leblond Claire S, Prasad A, Pinto D, Walker S, O'Connor I, Russell C, Drmic Irene E, Hamdan Fadi F, *et al*: **SHANK1 Deletions in males with autism spectrum disorder.** *Am J Hum Genet* 2012, **90**(5):879–887.
8. Berkel S, Marshall CR, Weiss B, Howe J, Roeth R, Moog U, Endris V, Roberts W, Szatmari P, Pinto D, *et al*: **Mutations in the SHANK2 synaptic scaffolding gene in autism spectrum disorder and mental retardation.** *Nat Genet* 2010, **42**(6):489–491.
9. Moessner R, Marshall CR, Sutcliffe JS, Skaug J, Pinto D, Vincent J, Zwaigenbaum L, Fernandez B, Roberts W, Szatmari P, *et al*: **Contribution of SHANK3 mutations to autism spectrum disorder.** *Am J Hum Genet* 2007, **81**(6):1289–1297.
10. Lionel AC, Vaags AK, Sato D, Gazzellone MJ, Mitchell EB, Chen HY, Costain G, Walker S, Egger G, Thiruvahindrapuram B, *et al*: **Rare exonic deletions implicate the synaptic organizer gephyrin (GPHN) in risk for autism, schizophrenia and seizures.** *Hum Mol Genet* 2013, **22**(10):2055–2066.
11. Leblond CS, Heinrich J, Delorme R, Proepper C, Betancur C, Huguet G, Konyukh M, Chaste P, Ey E, Rastam M, *et al*: **Genetic and functional analyses of *SHANK2* mutations suggest a multiple Hit model of autism spectrum disorders.** *PLoS Genet* 2012, **8**(2):e1002521.
12. Noor A, Whibley A, Marshall CR, Gianakopoulos PJ, Piton A, Carson AR, Orlic-Milacic M, Lionel AC, Sato D, Pinto D, *et al*: **Disruption at the PTCHD1 locus on Xp22.11 In autism spectrum isorder and intellectual disability.** *Sci Transl Med* 2010, **2**(49):49ra68.
13. Kerin T, Ramanathan A, Rivas K, Grepo N, Coetzee GA, Campbell DB: **A noncoding RNA antisense to moesin at 5p14.1 In autism.** *Sci Transl Med* 2012, **128:**128ra140.
14. Bonnet C, Masurel-Paulet A, Khan AA, Béri-Dexheimer M, Callier P, Mugneret F, Philippe C, Thauvin-Robinet C, Faivre L, Jonveaux P: **Exploring the potential role of disease-causing mutation in a gene desert: duplication of noncoding elements 5′ of GRIA3 is associated with GRIA3 silencing and X-linked intellectual disability.** *Hum Mutat* 2012, **33**(2):355–358.
15. Huang L, Jolly Lachlan A, Willis-Owen S, Gardner A, Kumar R, Douglas E, Shoubridge C, Wieczorek D, Tzschach A, Cohen M, *et al*: **A noncoding, regulatory mutation implicates HCFC1 in nonsyndromic intellectual disability.** *Am J Hum Genet* 2012, **91**(4):694–702.
16. Lionel AC, Crosbie J, Barbosa N, Goodale T, Thiruvahindrapuram B, Rickaby J, Gazzellone M, Carson AR, Howe JL, Wang Z, *et al*: **Rare copy number variation discovery and cross-disorder comparisons identify risk genes for ADHD.** *Sci Transl Med* 2011, **3**(95):95ra75.
17. Pinto D, Pagnamenta AT, Klei L, Anney R, Merico D, Regan R, Conroy J, Magalhaes TR, Correia C, Abrahams BS, *et al*: **Functional impact of global rare copy number variation in autism spectrum disorders.** *Nature* 2010, **466**(7304):368–372.
18. Pinto D, Darvishi K, Shi X, Rajan D, Rigler D, Fitzgerald T, Lionel AC, Thiruvahindrapuram B, MacDonald JR, Mills R, *et al*: **Comprehensive assessment of array-based platforms and calling algorithms for detection of copy number variants.** *Nat Biotech* 2011, **29**(6):512–520.
19. Marshall CR, Noor A, Vincent JB, Lionel AC, Feuk L, Skaug J, Shago M, Moessner R, Pinto D, Ren Y, *et al*: **Structural variation of chromosomes in autism spectrum disorder.** *Am J Hum Genet* 2008, **82**(2):477–488.
20. Prasad A, Merico D, Thiruvahindrapuram B, Wei J, Lionel AC, Sato D, Rickaby J, Lu C, Szatmari P, Roberts W, *et al*: **A discovery resource of rare copy number variations in individuals with autism spectrum disorder.** *G3 (Bethesda)* 2012, **2**(12):1665–1685.
21. Pagnamenta AT, Khan H, Walker S, Gerrelli D, Wing K, Bonaglia MC, Giorda R, Berney T, Mani E, Molteni M, *et al*: **Rare familial 16q21 microdeletions under a linkage peak implicate cadherin 8 (CDH8) in susceptibility to autism and learning disability.** *J Med Genet* 2011, **48**(1):48–54.

22. Morrow EM, Yoo S-Y, Flavell SW, Kim T-K, Lin Y, Hill RS, Mukaddes NM, Balkhy S, Gascon G, Hashmi A, *et al*: **Identifying autism loci and genes by tracing recent shared ancestry.** *Science* 2008, **321**(5886):218–223.
23. Ching MSL, Shen Y, Tan W-H, Jeste SS, Morrow EM, Chen X, Mukaddes NM, Yoo S-Y, Hanson E, Hundley R, *et al*: **Deletions of NRXN1 (neurexin-1) predispose to a wide spectrum of developmental disorders.** *Am J Med Genet B Neuropsychiatr Genet* 2010, **153B**(4):937–947.
24. The ENCODE Project Consortium: **An integrated encyclopedia of DNA elements in the human genome.** *Nature* 2012, **489**(7414):57–74.
25. Jiang Y-h, Yuen Ryan KC, Jin X, Wang M, Chen N, Wu X, Ju J, Mei J, Shi Y, He M, *et al*: **Detection of Clinically Relevant Genetic Variants in Autism Spectrum Disorder by Whole-Genome Sequencing.** *Am J Hum Genet* 2013, http://dx.doi.org/10.1016/j.ajhg.2013.06.012.

doi:10.1186/1471-2164-14-499
Cite this article as: Walker and Scherer: **Identification of candidate intergenic risk loci in autism spectrum disorder.** *BMC Genomics* 2013 **14**:499.