Neuron
NeuroView

Simons Variation in Individuals Project (Simons VIP): A Genetics-First Approach to Studying Autism Spectrum and Related Neurodevelopmental Disorders

The Simons VIP Consortium[1,*,**]
[1]Membership of the Consortium is provided in Table S5
*Correspondence: jspiro@simonsfoundation.org (J.E. Spiro)
**Correspondence: wkc15@columbia.edu (W.K. Chung)
DOI 10.1016/j.neuron.2012.02.014

We describe a project aimed at studying a large number of individuals (>200) with specific recurrent genetic variations (deletion or duplication of segment 16p11.2) that increase the risk of developing autism spectrum (ASD) and other developmental disorders. The genetics-first approach augmented by web-based recruitment, multisite collaboration and calibration, and robust data-sharing policies could be adopted by other groups studying neuropsychiatric disorders to accelerate the pace of research.

Motivation for the Collection

A key need for human genetic studies, not only in neuroscience but also in other disease areas, is access to a large number of individuals who have been reliably and thoroughly characterized. Rigorous clinical phenotyping is critical, and lack thereof can be a major bottleneck to progress. For many neuropsychiatric disorders such as ASD, bipolar disorder, and schizophrenia, this can be a particular challenge given the heterogeneity and complexity of the symptomatology for these disorders, which are diagnosed using inherently subjective behavioral criteria.

For ASD, a number of initiatives have been developed to fill this need for well-characterized individuals and biospecimens. Many of these involve consortium programs and large-scale collaborations between multiple institutes and investigators. Projects such as the Autism Genetic Resource Exchange (AGRE), which was initiated first by Cure Autism Now and is now administered by Autism Speaks, the Simons Simplex Collection (SSC), the Autism Genome Project, and the NIMH repository (Fischbach and Lord, 2010; Geschwind et al., 2001) have provided the backbone for new discoveries in ASD genetics over the last several years (State, 2010).

In this NeuroView, we discuss a new initiative, the Simons VIP, which was launched to fulfill a complementary need: in contrast to the existing genetic collections for ASD, where recruitment of patients is based on clinical diagnosis, the Simons VIP project takes a "genetics first" approach. The logic behind this approach is based on the increasing evidence suggesting that the genetics underlying neuropsychiatric disorders are complex and may involve mutations in hundreds of genes, each of which is relatively infrequent. Such heterogeneity makes it extremely challenging to perform patient cohort studies because there may be characteristics specific to certain subsets that are not common to all individuals. Nevertheless, certain highly penetrant copy-number variations (CNVs) or mutations in single genes are observed recurrently in cohorts ascertained by psychiatric diagnosis. For example, in the case of ASD see Levy et al. (2011), Marshall et al. (2008), and Sanders et al. (2011).

Given the relatively low recurrence (and therefore the small number of subjects typically studied with any specific genetic event), it is as yet unknown whether brain structure, neurophysiological responses, and associated clinical phenotypes will differ meaningfully between genetic etiologies and/or ultimately whether interventions will be generalizable to all patients or will be more effective if tailored to the underlying genetic etiology.

By recruiting and studying large numbers of families with deletions or duplications of 16p11.2, without regard to clinical diagnosis or age, we aim to address this question by studying the cross-sectional diversity and early longitudinal course of this genetically well-defined group of individuals at the behavioral and neurocognitive level. Although the Simons VIP project is initially focused on 16p11.2, the structure of the project should have broader applications for other complex genetic disorders. Longer-term goals are to use these data to develop targeted interventions and focused clinical care for individuals with specific genetic events.

Why 16p11.2?

Of the recurrent genetic causes of neuropsychiatric disorders, recent work has highlighted deletions and duplications of a shared region on 16p11.2 as being of particular interest given their relatively high frequency in large clinical cytogenetic microarray databases (Cooper et al., 2011; Kaminsky et al., 2011) and the evidence implicating 16p11.2 in a number of neuropsychiatric disorders.

An emerging theme in the field is that the same CNV may increase risk for multiple cognitive and neuropsychiatric disorders. For example, the 7q11.23 deletion causes Williams Syndrome, whereas duplication of the same region is associated with ASD (Sanders et al., 2011). Similarly, the chromosomal region 16p11.2 has been associated with ASD (Weiss et al., 2008), as well as schizophrenia, bipolar disorder, developmental delay, and body weight regulation (see Supplemental References available online). Also, the specific dosage effects of the genes in this region may regulate components of the anthropometric phenotype in a reciprocal fashion: early

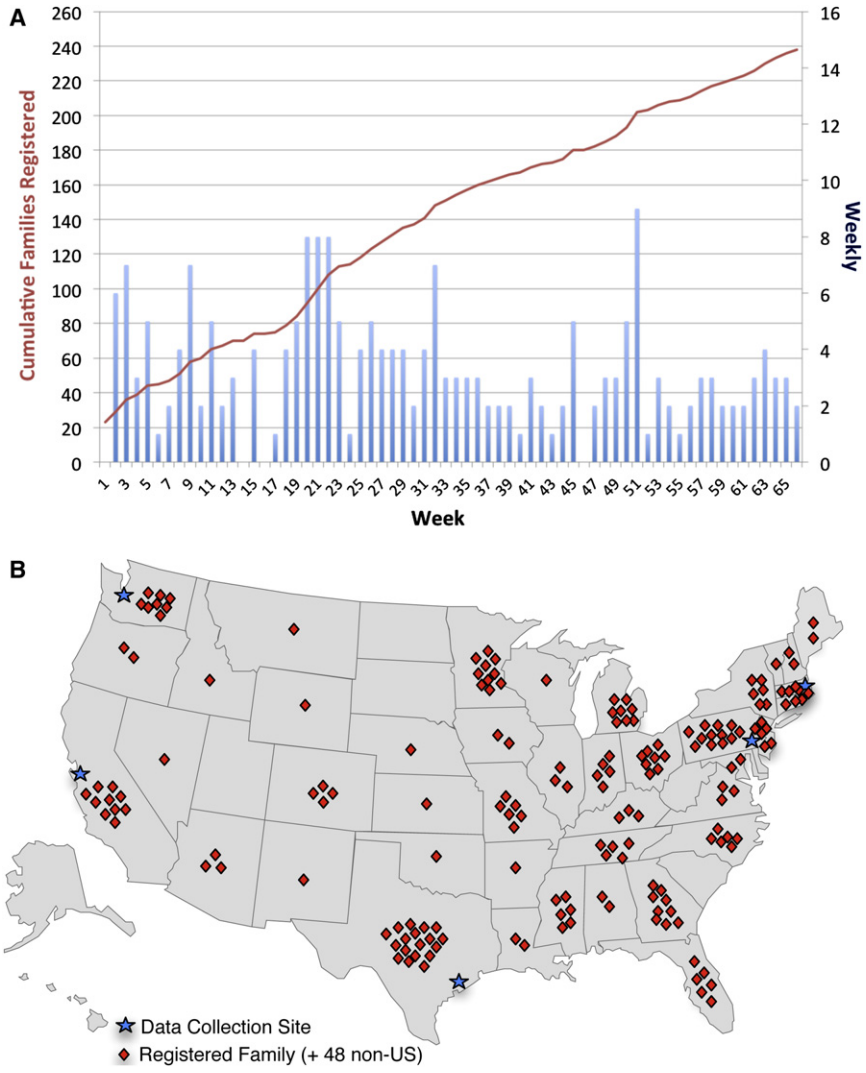


Figure 1. Families Registered on Simons VIP Connect
(A) Weekly (blue) and cumulative (red) number of worldwide families registered during the first 66 weeks. (B) Distribution of registered families in the United States (red). Study centers are indicated with blue stars.

evidence suggests that people with 16p11.2 deletions may have high body mass index (BMI) and large head circumferences whereas duplications have low BMI and small head circumferences (Bochukova et al., 2010; Jacquemont et al., 2011; McCarthy et al., 2009; Shinawi et al., 2010; Walters et al., 2010).

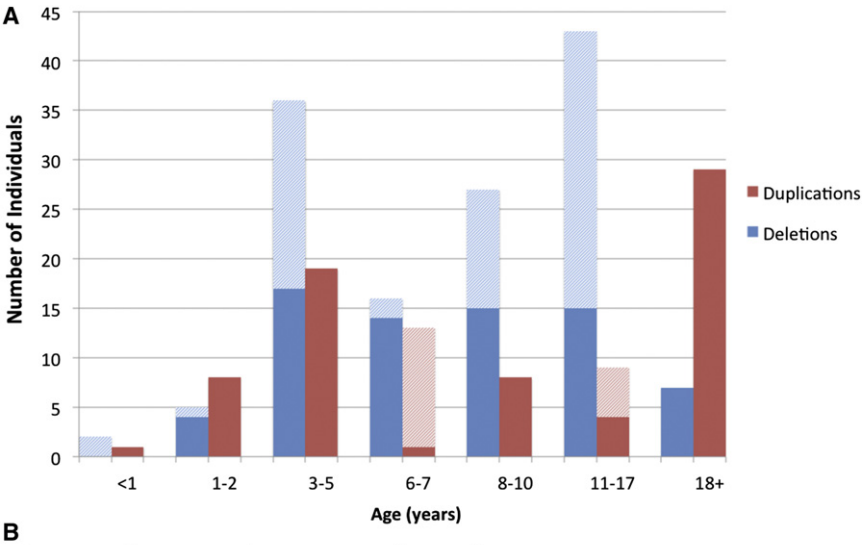
The 16p11.2 CNV is ~600 kb, with consistent breakpoints and contains approximately 29 genes. Studying a genetic lesion that is consistent among individuals increases the homogeneity of the cohort in a manner that is not usually possible for most single-gene disorders in which a wide variety of mutations are observed that may have varying effects on function and lead to different phenotypes associated with various genotypes.

Recruitment and Retention of Subjects: Simons VIP Connect

Family engagement is a key factor for success of any long-term genetic project. For Simons VIP this begins with recruitment and core phenotyping but does not end there. For example, we have actively engaged families with forums in which families can learn about 16p11.2 and interact with one another and webinars in which families can learn about the latest research.

Launched in September 2010, Simons VIP Connect (http://www.simonsvipconnect.org/) is designed to support an online community for individuals worldwide with 16p11.2 deletions and duplications and their families and is the primary means of recruiting families for the Simons VIP. Many families find Simons VIP Connect on their own via internet search for 16p11.2 after they receive results from a clinical genetic test screening for CNVs. Our recruitment strategies have also included directed traffic from Google ads and Facebook and links from other chromosomal disorder patient advocacy websites (Unique, http://www.rarechromo.org, and CDO, http://www.chromodisorder.org/CDO/). We also established collaborations with clinical molecular cytogenetics laboratories (notably but not limited to those sites participating in the International Standards for Cytogenomic Arrays Consortium, [ISCA]; https://www.iscaconsortium.org/) to notify treating physicians to refer patients who meet study eligibility. We also sought out referrals from medical professionals including genetic counselors, geneticists, child neurologists, and developmental pediatricians who were informed through direct mailings. Families who previously participated in the SSC who were found to have a 16p11.2 deletion or duplication were also invited to enter the Simons VIP study. In addition, as chromosome microarray testing is entering into the prenatal area, fetuses with 16p11.2 deletions/duplications are beginning to be identified and provide the opportunity to understand fetal and early childhood brain development in this population. Within the first year after launch, over 200 families from around the world have joined the online community of Simons VIP Connect. We have registered approximately four new families/week with a broad regional and age distribution (Figure 1).

As much as this collection will provide data to researchers, the project also has a component aimed at providing information to families. The site content is actively curated by a team of genetic counselors who maintain up-to-date summaries about publications on 16p11.2, publish a newsletter for families, and host a series of webinars by Simons VIP scientists/physicians and outside



B

	Deletions	Duplications	Total:
Males	36	29	65
Females	28	27	55
TOTAL:	64	56	120

Figure 2. Consented and Potential Participants as of December 2011

(A) Consented individuals are in solid colors. Shaded regions indicate additional potential families for study who are not consented because they live outside of North America or are unable to participate in the study at the present time. While the solid bars indicate individuals, the shaded bars are numbers of families and therefore are a conservative estimate of the number of individuals as cascade testing in registered families often identifies additional individuals with 16p11.2 deletions (average 1.3 per family) or duplications (average 2.1 per family).

(B) Distribution of Simons VIP participants by gender and deletion/duplication status.

experts on topics of interest to families. The website also offers the option to "ask an expert" that has been used by patients and health care providers. Starting in the summer of 2012, Simons VIP will organize a meeting at which families can interact directly with each other and Simons VIP researchers. The feedback of aggregate research results to the patient community has been a strong motivation to keep families engaged and actively participating (see Supplemental Experimental Procedures).

Inclusion in Simons VIP Study and Ascertainment Biases

Our goal is to create a large cohort of subjects who were as genetically similar as possible. Below we describe steps taken, which involve including as many family members as possible beyond the individual who had the original genetic test. Also described is how we are dealing with ascertainment biases that complicate ours and all efforts in this field.

When families register for Simons VIP Connect online (or by calling a toll-free telephone number), they are provided with a description of the Simons VIP study and, if they are in the United States or Canada, are asked if they would like to participate. Genetic test reports are reviewed to confirm eligibility, which consists of having the canonical deletion or duplication (~600 kb, chr16: 29,557,497-30,107,356; hg18), or a smaller CNV at the locus. Exclusion criteria include any other pathogenic CNVs or other neurogenetic or neurological diagnoses unrelated to 16p11.2 (e.g., tuberous sclerosis).

Blood samples provided by participants are used to test for the 16p11.2 deletion/duplication by fluorescent in situ hybridization (FISH) or comparative genomic hybridization to determine who in the nuclear family carries the deletion/duplication, with cascade testing of additional members in the family as far as the family is willing or able to allow. (Approximately 50% of families registering on Simons VIP Connect already know whether the deletion or duplication is inherited.) All deletion/duplication carriers in the family are eligible for participation (see Supplemental Experimental Procedures). Within one year, we have obtained consent from over 100 individuals with 16p11.2 deletions or duplications to participate in research (Figure 2). This represents extremely efficient recruitment for a rare genetic disorder.

A limitation of our recruitment strategy is the possible ascertainment bias to individuals who were clinically significantly affected enough for a parent or provider to seek an etiologic diagnosis and have access to a clinical chromosome microarray. This could bias the study toward ascertainment of more severely affected probands. By performing cascade genetic testing within the families, we have intentionally attempted to increase recruitment and include individuals in familial cases who may not have come to clinical attention. We are also pursuing strategies to enroll children who are enrolled in other research studies not targeted at developmental disabilities in which genomic CNV analysis identified 16p11.2 deletions and duplications and who consented to recontact. However, given that our study is not a population-based epidemiological study, our results will probably only define the range of clinical phenotypes associated with 16p11.2 deletions and duplications but cannot be used to predict the probability of any particular phenotype in an asymptomatic fetus or infant.

Core Phenotyping and Structural Neuroimaging

The Simons VIP represents the first large-scale effort to study the natural history of individuals with specific genetic events associated with nonsyndromic ASD and related disorders. The psychological phenotyping battery is intentionally broad enough to capture the spectrum of possible diagnoses of individuals with 16p11.2 (and also appropriate for individuals with other genetic events, thus enabling studies that directly compare different defined genetic conditions), yet streamlined enough to allow for completion of the protocol in a two-day evaluation.

Families travel to one of three participating core phenotyping centers: Baylor College of Medicine, Houston; Children's Hospital Boston, Harvard University, Boston; or University of Washington, Seattle. The protocol includes a comprehensive, age-appropriate battery of psychological tests and interviews, a neurological exam, growth measurements, standard and three dimensional craniofacial surface images (3dMD Inc., Atlanta & London) for dysmorphology, a structural brain MRI for participants who can complete the study without the use of sedation, and collection of biospecimens including blood and an optional skin biopsy to harvest fibroblasts for future generation of induced pluripotent stem cells (iPSCs). For a more detailed description of the phenotyping and imaging protocols, see Tables S1 and S2.

To avoid a common pitfall where the same individual is reported in multiple studies, as is often the case for rare disorders, all participants are assigned a global unique identifier (Johnson et al., 2010). Data are entered into a custom database designed and maintained by Prometheus Research LLC, as previously described (Fischbach and Lord, 2010). Biospecimens are processed and stored at the Rutgers University Cell and DNA Repository (RUCDR) for use by the research community. Nuclear family members who do not carry the 16p11.2 deletion/duplication are also encouraged but not required to participate and are evaluated with a limited number of psychological tests to serve as controls. These family controls serve as an important control for other familial factors as measures such as IQ can be compared not only to population controls but also the unaffected family controls.

As diagnostic differences across clinical sites have often been a challenge for human genetic studies, we have developed the phenotyping protocols with an aim for consistency and reliability. Diagnoses are based on standardized measures applied to DSM-IV-TR criteria (see Supplemental Experimental Procedures). Children age 4 years and younger will be assessed longitudinally with a combination of parental interviews every 6 months and serial psychometric testing at ages 6, 12, 18, 24, 36, and 48 months.

The structural brain MRI protocol, which also includes sequences typically included in a clinical scan, is identical across sites, and the scanners are carefully cross-calibrated (see Supplemental Experimental Procedures).

Functional Neuroimaging and Neurophysiology

Many studies report signatures of brain activity that correlate with neuropsychiatric disease status. Given the phenotypic and genetic heterogeneity of these disorders, including ASD, it is perhaps not surprising that these measures are often quite noisy, they can result in contradictory results, and even the most salient features usually require substantial cohorts to validate. Two important research questions are to determine whether these signatures would be less noisy in a more genetically homogeneous population, and, because there is clinical diversity within this genetically homogeneous cohort, whether these brain activity signatures correlate with phenotype (ASD) or genetic etiology (16p11.2 del/dup). Toward this end we have added multiple measures of brain function.

Participants who are able to complete the structural MRI (without evidence of significant motion artifact) and who are 7 years old or older are asked to participate in an additional component of the study that involves functional imaging. The purpose of this part of the study is to address whether detailed structural and functional imaging coupled with a comprehensive neuropsychological battery on both 16p11.2 participants and controls can identify robust correlations between 16p11.2 deletions and duplications and brain function. The two imaging modalities, fMRI and MEG, complement each other with regard to tradeoffs in temporal versus spatial resolution. The protocols for fMRI and MEG interrogate a broad array of cognitive domains, including language, executive function, and face and motor processing and incorporate both resting state and task related protocols (for more detailed imaging protocols, see Tables S3 and S4). To insure consistency of measures, the functional imaging studies are performed over 2 days at the University of California, San Francisco or at Children's Hospital of Philadelphia, where the two sites have nearly identical MRI and MEG hardware and software implementations that have been calibrated for data pooling (see Supplemental Experimental Procedures).

Data- and Biospecimen-Sharing Plans

A fundamental principle of this project is that in addition to an active research project, it is also intended to provide the broad scientific community with a valuable resource for future research. Data will be made available through a web-based portal to approved investigators in raw and processed forms to allow for further analyses and comparison to other cohorts. Researchers involved in the creation of the Simons VIP resource should be suitably acknowledged but will not restrict access to the biospecimens, phenotype, or neuroimaging data. Researchers can use SFARI Base (http://base.sfari.org), the online Simons Foundation Autism Research Initiative (SFARI) data repository, to review specific and aggregate characteristics, to identify interesting subsets of cases, and to request biospecimens and/or data in raw and processed forms for further analyses and comparison to other cohorts.

We believe that our data-sharing policy is ideal in that it allows rapid access to data and biospecimens to the community but acknowledges that others may wish to analyze or publish on their data before releasing it to the community. As the ability to recontact subjects can be useful for ongoing studies, a procedure will be in place to determine whether and when the Simons VIP research team will recontact subjects for follow-up research or ancillary studies that require data or biospecimens beyond those already collected. This procedure will balance research priorities with potential burden on families.

Future Plans

We envision that the Simons VIP will produce immediate information about the medical, cognitive, and neuroimaging profiles of subjects with deletions and duplications of 16p11.2 that should be of considerable value to families and their clinicians. Phenotyping and neuroimaging protocols will be evaluated on a regular basis and additions or subtractions will be made depending on the

value of the data. Future experiments may also involve recontacting families to ask for their participation in collecting additional data, such as EEG, quantitative sleep data, or additional biospecimens. Furthermore, the availability of biospecimens linked to the other types of data will provide many future opportunities for follow-up study. For example, we expect that targeted resequencing of the remaining or extra 16p11.2 allele and/or expression studies may help to narrow down which genes in the interval are of particular relevance to the phenotypes. Exome or full-genome sequencing of the samples may identify other genetic alterations, in addition to 16p11.2, that may be relevant to the phenotype and differential expressivity between individuals; analysis of epigenetics may be similarly informative.

Studies that transform the banked fibroblasts into iPSCs and then differentiate them into neurons may help to understand how deletions or duplications of 16p11.2 influence early neural development or neuronal function and may provide an effective platform for high throughput screening of drugs that could potentially be tested for efficacy in this genetically homogeneous cohort.

In addition to supporting the core Simons VIP effort directly, SFARI will continue to entertain proposals through the regular Request for Applications (RFA) for work related to the analysis of phenotype or neuroimaging data and specimens. SFARI is prepared to join with other funding agencies in supporting further studies utilizing the resources generated by the Simons VIP effort.

In closing, the genetics-first, multisite, and highly collaborative nature of this project, combined with clear data-sharing policies, allows for ready scalability. We think it will have broad applicability to other efforts to understand genotype-phenotype relationships. This approach is especially warranted in neuropsychiatric disorders, where the clinical heterogeneity of disorders, diagnosed behaviorally, present special problems, but should extend beyond as well.

SUPPLEMENTAL INFORMATION

Supplemental Information includes five tables (including the extended author list) and Supplemental Experimental Procedures and can be found with this article online at doi:10.1016/j.neuron.2012.02.014.

ACKNOWLEDGMENTS

We are extremely grateful to the families who are participating in this study. We thank Cathy Lord and Helen Tager-Flusberg for advice on the phenotyping protocols, Prometheus Research for work on data management, RUCDR for help with managing biospecimens, and Alexandra Bowe for help with preparation of the figures. The Simons VIP Connect website is hosted by PatientCrossroads (http://www.patientcrossroads.com/), which provides registry systems that connect communities of people with rare diseases and scientists studying those conditions. Timothy P.L. Roberts has a consulting relationship with Prism Clinical Imaging; David H. Ledbetter has consulting relationships with Roche Nimblegen, Combimatrix, and Celula. Both report no overlap with the Simons VIP. Arthur L. Beaudet is Chair of the Department of Molecular and Human Genetics at Baylor College of Medicine (BCM) which offers extensive genetic laboratory testing, and BCM derives revenue from this activity. The Simons VIP Connect website and Simons VIP were funded by the Simons Foundation as part of SFARI.

REFERENCES

Bochukova, E.G., Huang, N., Keogh, J., Henning, E., Purmann, C., Blaszczyk, K., Saeed, S., Hamilton-Shield, J., Clayton-Smith, J., O'Rahilly, S., et al. (2010). Nature *463*, 666–670.

Cooper, G.M., Coe, B.P., Girirajan, S., Rosenfeld, J.A., Vu, T.H., Baker, C., Williams, C., Stalker, H., Hamid, R., Hannig, V., et al. (2011). Nat. Genet. *43*, 838–846.

Fischbach, G.D., and Lord, C. (2010). Neuron *68*, 192–195.

Geschwind, D.H., Sowinski, J., Lord, C., Iversen, P., Shestack, J., Jones, P., Ducat, L., and Spence, S.J.; AGRE Steering Committee. (2001). Am. J. Hum. Genet. *69*, 463–466.

Jacquemont, S., Reymond, A., Zufferey, F., Harewood, L., Walters, R.G., Kutalik, Z., Martinet, D., Shen, Y., Valsesia, A., Beckmann, N.D., et al. (2011). Nature *478*, 97–102.

Johnson, S.B., Whitney, G., McAuliffe, M., Wang, H., McCreedy, E., Rozenblit, L., and Evans, C.C. (2010). J. Am. Med. Inform. Assoc. *17*, 689–695.

Kaminsky, E.B., Kaul, V., Paschall, J., Church, D.M., Bunke, B., Kunig, D., Moreno-De-Luca, D., Moreno-De-Luca, A., Mulle, J.G., Warren, S.T., et al. (2011). Genet. Med. *13*, 777–784.

Levy, D., Ronemus, M., Yamrom, B., Lee, Y.H., Leotta, A., Kendall, J., Marks, S., Lakshmi, B., Pai, D., Ye, K., et al. (2011). Neuron *70*, 886–897.

Marshall, C.R., Noor, A., Vincent, J.B., Lionel, A.C., Feuk, L., Skaug, J., Shago, M., Moessner, R., Pinto, D., Ren, Y., et al. (2008). Am. J. Hum. Genet. *82*, 477–488.

McCarthy, S.E., Makarov, V., Kirov, G., Addington, A.M., McClellan, J., Yoon, S., Perkins, D.O., Dickel, D.E., Kusenda, M., Krastoshevsky, O., et al; Wellcome Trust Case Control Consortium. (2009). Nat. Genet. *41*, 1223–1227.

Sanders, S.J., Ercan-Sencicek, A.G., Hus, V., Luo, R., Murtha, M.T., Moreno-De-Luca, D., Chu, S.H., Moreau, M.P., Gupta, A.R., Thomson, S.A., et al. (2011). Neuron *70*, 863–885.

Shinawi, M., Liu, P., Kang, S.H., Shen, J., Belmont, J.W., Scott, D.A., Probst, F.J., Craigen, W.J., Graham, B.H., Pursley, A., et al. (2010). J. Med. Genet. *47*, 332–341.

State, M.W. (2010). Neuron *68*, 254–269.

Walters, R.G., Jacquemont, S., Valsesia, A., de Smith, A.J., Martinet, D., Andersson, J., Falchi, M., Chen, F., Andrieux, J., Lobbens, S., et al. (2010). Nature *463*, 671–675.

Weiss, L.A., Shen, Y., Korn, J.M., Arking, D.E., Miller, D.T., Fossdal, R., Saemundsen, E., Stefansson, H., Ferreira, M.A., Green, T., et al; Autism Consortium. (2008). N. Engl. J. Med. *358*, 667–675.