

Brain-expressed exons under purifying selection are enriched for *de novo* mutations in autism spectrum disorder

Mohammed Uddin¹, Kristiina Tammimies^{1,2}, Giovanna Pellecchia¹, Babak Alipanahi³, Pingzhao Hu¹, Zhuozhi Wang¹, Dalila Pinto⁴⁻⁶, Lynette Lau¹, Thomas Nalpathamkalam¹, Christian R Marshall^{1,7}, Benjamin J Blencowe^{8,9}, Brendan J Frey^{3,8}, Daniele Merico¹, Ryan K C Yuen¹ & Stephen W Scherer^{1,7,9}

A universal challenge in genetic studies of autism spectrum disorders (ASDs) is determining whether a given DNA sequence alteration will manifest as disease. Among different population controls, we observed, for specific exons, an inverse correlation between exon expression level in brain and burden of rare missense mutations. For genes that harbor *de novo* mutations predicted to be deleterious, we found that specific critical exons were significantly enriched in individuals with ASD relative to their siblings without ASD ($P < 1.13 \times 10^{-38}$; odds ratio (OR) = 2.40). Furthermore, our analysis of genes with high exonic expression in brain and low burden of rare mutations demonstrated enrichment for known ASD-associated genes ($P < 3.40 \times 10^{-11}$; OR = 6.08) and ASD-relevant fragile-X protein targets ($P < 2.91 \times 10^{-157}$; OR = 9.52). Our results suggest that brain-expressed exons under purifying selection should be prioritized in genotype-phenotype studies for ASD and related neurodevelopmental conditions.

Rare genomic variants (with a population frequency of <0.05) are being identified as risk factors in ASD¹. ASD exhibits extensive clinical and genetic heterogeneity with high heritability, and chromosomal abnormalities, copy number variations (CNVs), insertion-deletions (indels), single-nucleotide variations (SNVs) and combinations thereof have been described as being etiological²⁻⁴. Indeed, >100 ASD susceptibility genes with variants of variable (and largely undetermined) penetrance and expressivity are known, with several hundred other genes estimated to exist⁵. Common genetic variants may also exert effects in ASD⁶. The importance of rare variants in other brain disorders such as schizophrenia, intellectual disability and epilepsy has also been recognized^{7,8}.

Proving the causality of particular DNA sequence variants in ASD can be complicated by the exceptional rarity of the mutations involved

(often unique to individuals or families) and the rate of incidence of new mutations, which is often the same in controls. For example, the rate at which *de novo* SNVs and indels occur (~ 1 per exome) is similar in individuals affected by ASD and in unaffected individuals (population controls or unaffected siblings)^{4,9}. For CNVs, a slightly greater burden of *de novo* events has been observed in ASD cases compared to controls³, but these deletions and duplications often affect multiple genes, complicating genotype-phenotype correlations. There has been some progress in identifying the exons critical for the clinical manifestation of ASD in *NRXN1* (ref. 10), *NRXN3* (ref. 11), *ASTN2* (ref. 12) and *AUTS2* (ref. 13). However, a robust approach is needed to find the rare etiological alterations among the thousands of other variants identified in genome scanning experiments.

We sought to exploit transcriptome maps of developing brain in humans¹⁴ coupled with data from population genetics studies of mutational burden¹⁵⁻¹⁸ to examine the impact of genetic variants in ASD and other disorders (Fig. 1 and Supplementary Fig. 1). Exome sequencing has identified an excess of damaging variants that arose over the past 5,000 years among individuals of European ancestry^{17,18}, but how this variation relates to gene expression in the developing brain or to disease is little understood. Our initial examination of the *SHANK1*, *SHANK2*, *FMRI*, *AFF2* and *SCN2A* ASD-associated genes in control samples showed an inverse correlation between the burden of rare missense mutations (defined as the number of rare missense variants divided by exon length) and the expression levels for individual exons in prenatal brain (Fig. 2a and Supplementary Fig. 2). We then hypothesized that, to prevent adverse functional consequences, the accumulation of potentially deleterious mutations is suppressed for 'critical exons' that demonstrate higher expression in brain. We undertook global testing to determine whether rare mutations within such exons were more likely to cause brain disorders than those within other exons where recent mutational burden was relaxed, regardless of gene expression. We also considered entire gene

¹The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Ontario, Canada. ²Center of Neurodevelopmental Disorders (KIND), Neuropsychiatric Unit, Department of Women's and Children's Health, Karolinska Institutet, Stockholm, Sweden. ³Department of Electrical and Computer Engineering, University of Toronto, Toronto, Ontario, Canada. ⁴Seaver Autism Center for Research and Treatment, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ⁵Department of Psychiatry, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ⁶Department of Genetics and Genomics Sciences, Icahn School of Medicine at Mount Sinai, New York, New York, USA. ⁷McLaughlin Centre, University of Toronto, Toronto, Ontario, Canada. ⁸Donnelly Centre, University of Toronto, Toronto, Ontario, Canada. ⁹Department of Molecular Genetics, University of Toronto, Toronto, Ontario, Canada. Correspondence should be addressed to S.W.S. (stephen.scherer@sickkids.ca).

Received 6 December 2013; accepted 10 April 2014; published online 25 May 2014; doi:10.1038/ng.2980

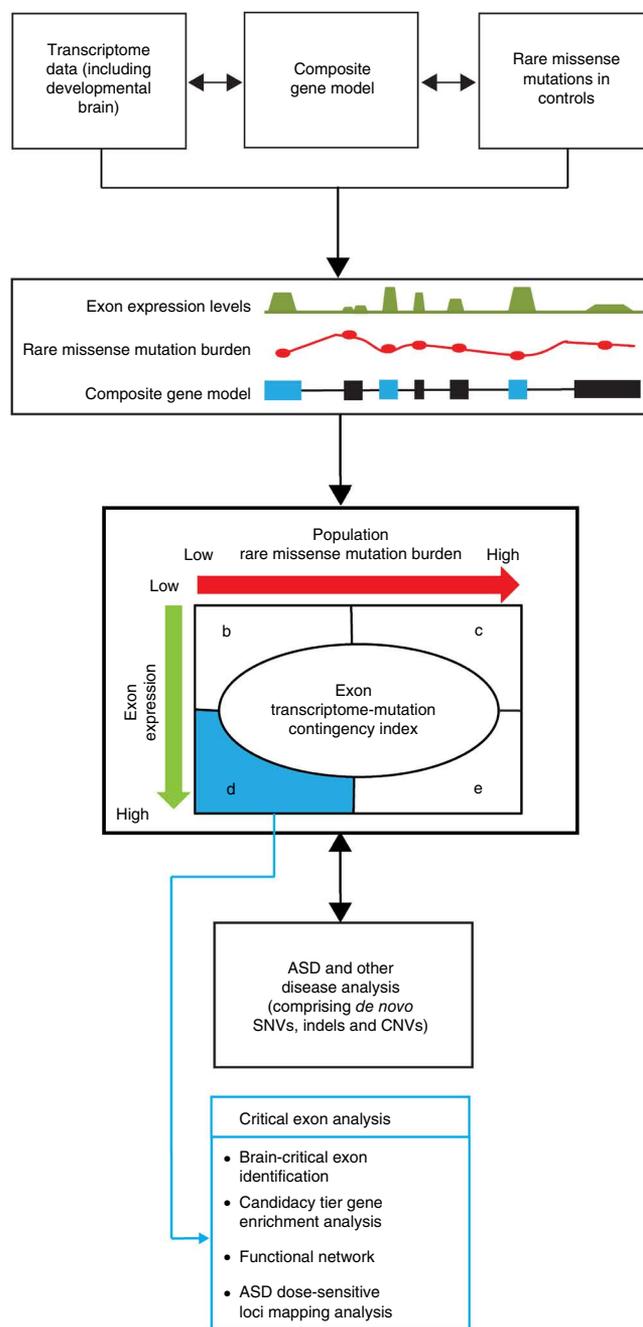
Figure 1 Assembling an exon transcriptome-mutation contingency index. A composite gene model (**Supplementary Note**) was generated and used to integrate exon expression and rare missense mutation levels in controls. Using this model, an exon transcriptome-mutation contingency index was generated. A critical exon (index “d” in blue) was defined within the contingency index as one that was highly expressed relative to other exons while showing suppressed accumulation of missense mutations in population controls (based on 75th-percentile thresholds from the data). A tolerated exon (index “e”) was defined as corresponding to a high brain expression profile and high accumulation of rare missense mutations in the population. Using this transcriptome-mutation contingency index, enrichment analyses were performed for genes implicated in ASD and other disorders. The set of brain-critical exons (3,955 in 1,744 genes) was further analyzed for functional enrichment and disease association.

units, but this analysis did not produce the same significant findings as observed with critical exons.

We classified exons in two ways to construct an exon transcriptome-mutation contingency index (**Supplementary Note**): (i) exon expression levels were discretized into classes with high and low expression (above or below the 75th percentile of expression for the entire data set, respectively), and (ii) the burden of rare missense mutations was defined as the number of rare missense variants divided by exon length and was classified as high or low in the same manner (**Fig. 1**). Using expression data from 11 tissues (**Supplementary Fig. 3**) for 16,713 RefSeq genes, we tested genome-wide association between burden of rare missense mutations and exonic expression, finding that brain cerebellum showed a strong association (Bonferroni-corrected Fisher’s exact test, $P < 2.53 \times 10^{-67}$) (**Fig. 2b**). Exons classified as critical also showed significant enrichment ($P < 4.44 \times 10^{-16}$) for Pfam (database of known protein families) domains compared to non-critical exons in cerebellum transcriptome data.

To examine the relevance of our observations with respect to brain disorders, we tested whether genes found (by sequencing or microarray) in selected data sets to carry *de novo* mutations (SNVs and indels or CNVs, respectively) predicted to be deleterious exhibited a similar inverse correlation with spatiotemporal exonic expression. Using exome or genome sequence data sets from index cases of ASD, schizophrenia or intellectual disability (**Supplementary Note**), we identified 243 genes containing 256 *de novo* mutations predicted to be deleterious^{2,4,7,8,19–23}. For the ASD cases from the Simon’s Simplex Collection (SSC), equivalent sequence data exist for unaffected sibling controls, identifying 82 genes containing 84 *de novo* exonic deleterious mutations^{4,19}. Exon expression data were obtained from both microarray analysis and RNA sequencing of 196 normal brain samples (**Supplementary Tables 1 and 2**).

Our initial analysis comparing entire genes with *de novo* mutations in ASD cases and siblings without ASD from SSC showed no significant difference in conservation scores, the distribution of the burden of rare missense mutations (**Supplementary Figs. 4 and 5**), deleterious variant predictions from *in silico* tools (PolyPhen, SIFT and MutationTaster) or the induction of aberrant splicing²⁴ (**Supplementary Fig. 6**). However, the inverse correlation between the burden of rare missense mutations and spatiotemporal expression in the brain in microarray-based transcriptome data was significant for genes with *de novo* mutations ascertained in ASD cases (Bonferroni-corrected Spearman’s correlation test, $P < 2.0 \times 10^{-16}$). No significant correlation was detected for genes ascertained in unaffected siblings (**Fig. 2c**). We also observed a similar pattern of inverse correlation for genes with *de novo* mutations found in intellectual disability and schizophrenia cases (**Fig. 2c**). When analyzing genes mutated in ASD cases, we observed the strongest association in prenatal transcriptome



samples (**Fig. 3a,b**), especially for prenatal orbital frontal cortex (Bonferroni-corrected Fisher’s exact test, $P < 1.13 \times 10^{-38}$; OR = 2.40). No association was observed for any brain expression data in genes mutated in the siblings of individuals with ASD. Results with microarray expression data were replicated using RNA sequencing data (**Supplementary Fig. 7**).

When restricting our analysis to the exons with *de novo* mutations ascertained in ASD cases and in siblings without ASD, 37.3% and 32.2% of the 196 brain samples, respectively, showed high expression (above the 75th percentile) (**Fig. 4a**). The presence of a roughly equal rate of *de novo* mutations in cases and siblings within these exons that are highly expressed in the brain makes it difficult to distinguish actual mutational effects. However, when we applied our transcriptome-mutation contingency index, a difference was found (predominantly for prenatal transcriptome samples), whereby 116 of

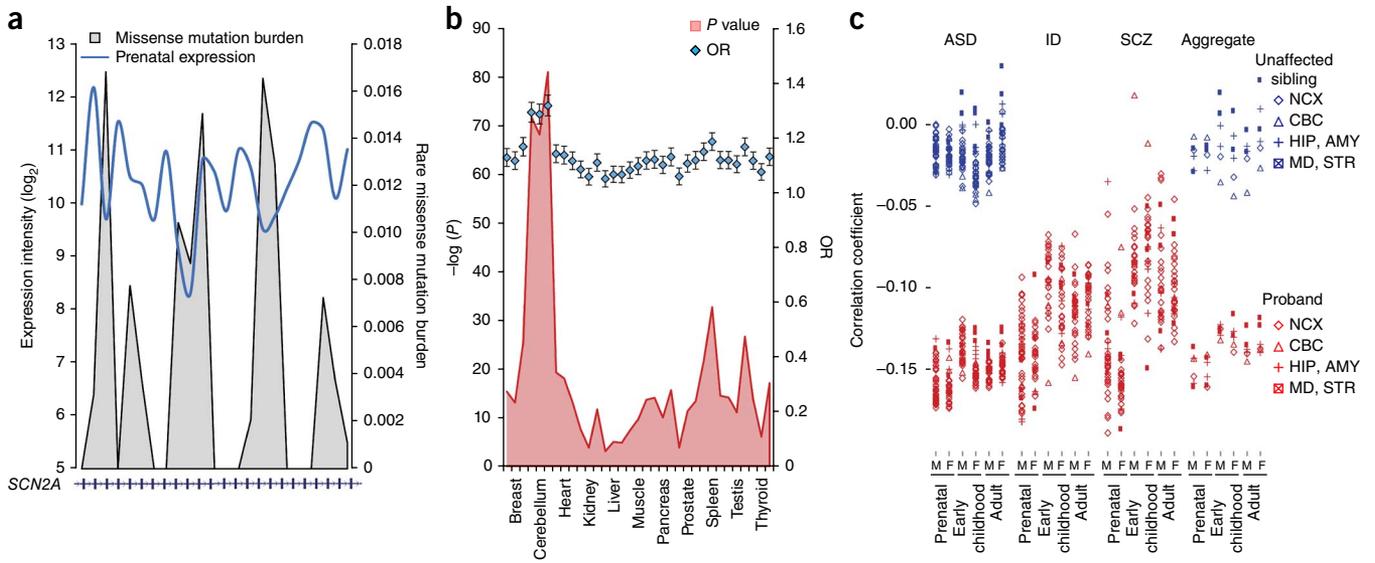


Figure 2 Inverse correlation between the burden of rare missense mutations and exon expression levels. **(a)** Integration of exon expression levels and the burden of rare missense mutations across the *SCN2A* gene. **(b)** Genome-wide association between the burden of rare missense mutations and exon expression levels in 11 human tissues (each analyzed in triplicate). Ranges for OR values represent 95% confidence intervals (CIs). **(c)** Exon expression levels in 196 postmortem brain samples (controls) from 13 male (M) and 13 female (F) donors in 3 developmental periods (prenatal, early childhood and adult) for 16 brain regions (11 neocortex (NCX) regions, cerebellar cortex (CBC), hippocampus (HIP), amygdala (AMY), mediodorsal nucleus of thalamus (MD) and striatum (STR)). The spatiotemporal correlation coefficient shown was computed for genes reported to have *de novo* mutations (SNVs or indels) in ASD cases or unaffected siblings or in cases of intellectual disability (ID) or schizophrenia (SCZ). The last panel shows aggregate spatiotemporal correlation coefficients (average expression from samples for a developmental period) for genes ascertained in the unaffected siblings of individuals with ASD or in any of the subject groups for the three diseases.

256 (45%) exons affected by *de novo* mutation in cases and 24 of 84 (28%) exons affected by *de novo* mutation in unaffected siblings were classified as critical exons. In contrast, a twofold excess of *de novo*

mutations classified as ‘tolerated’ (**Fig. 1**) was found in siblings (**Fig. 4a**). Considering loss-of-function mutations only, the enrichment was threefold for critical exons in cases (**Supplementary Fig. 8a**).

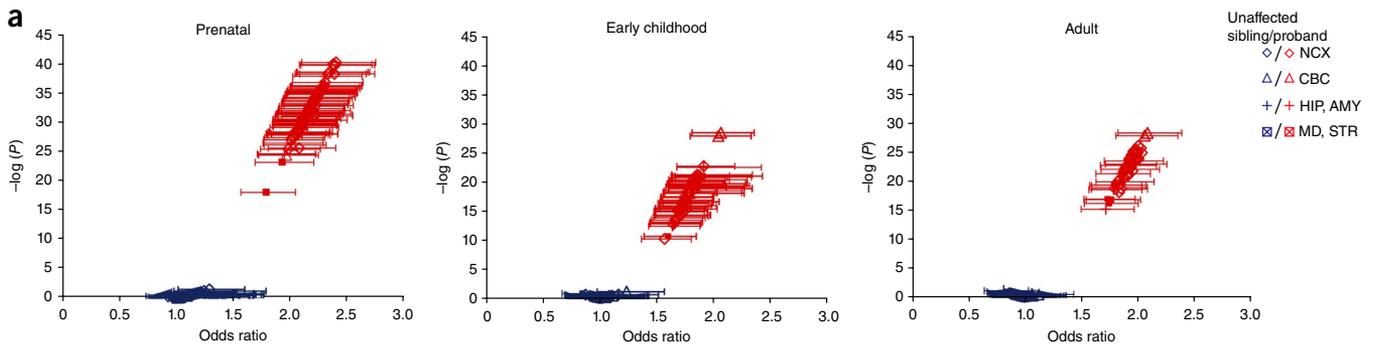
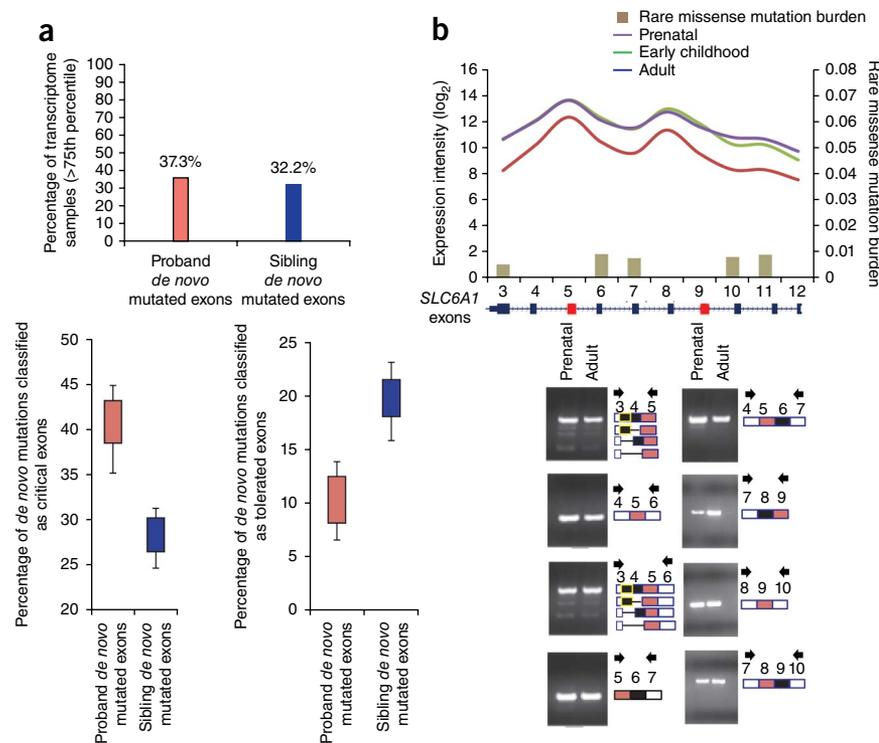


Figure 3 Spatiotemporal association analysis of exon expression levels and burden of rare missense mutations for genes with *de novo* SNVs or indels predicted to be deleterious. **(a)** Spatiotemporal association analysis for genes reported to have deleterious *de novo* mutations in cases or unaffected siblings. Ranges for OR values represent 95% CIs. **(b)** Spatiotemporal heat maps (with the locations of brain tissue samples depicted on the medial surface) include information on 16 brain regions that were outlined in 3 developmental human brain stages (prenatal, early childhood and adult). For each brain region, (AMY, amygdaloid complex; CBC, cerebellar cortex; V1C, primary visual cortex; STC, posterior (caudal) superior temporal cortex; IPC, posterior inferior parietal cortex; A1C, primary auditory cortex; S1C, primary somatosensory cortex; M1C, primary motor cortex; STR, striatum; DFC, dorsolateral prefrontal cortex; MFC, medial prefrontal cortex; VFC, ventrolateral prefrontal cortex; OFC, orbital frontal cortex; MD, mediodorsal nucleus of thalamus; ITC, inferolateral temporal cortex; HIP, hippocampus) the color gradient reflects the association *P* value between expression levels and the burden of rare missense mutations.



Figure 4 Highly expressed exons with *de novo* mutations. **(a)** Top, percentages of brain tissues that have high expression (>75th percentile) of the exons comprising *de novo* mutations that were ascertained in cases and unaffected siblings. Bottom, box plots show the percentages of *de novo* mutations (for cases and unaffected siblings) that can be classified as critical exons (left) or tolerated exons (right) using all transcriptome samples. Each confidence interval represents the fraction of variability in the number of critical or tolerated exons obtained from different transcriptome samples (interquartile range (IQR) is between the first and third quartiles). **(b)** Top, the burden of rare missense mutations and expression in human developing brain of the *SLC6A1* gene, for which mutated exons (red boxes) have been reported in ASD cases. Bottom, to capture the exon inclusion/exclusion event for the mutated exons in prenatal brain and adult brain. RT-PCR assays were designed considering various combinations of the flanking exons. Exclusion events (defined by a line for an exon) were confirmed by Sanger sequencing (**Supplementary Table 3**). Partial exclusion of exon 3 (yellow boxes) of the *SLC6A1* gene was also captured owing to the variable length of this exon in different isoforms.



Without applying a cutoff on expression, a significant dichotomy in expression can still be observed with respect to high and low burdens of rare missense mutations (**Supplementary Fig. 8b**), predominantly for prenatal samples (Bonferroni-corrected Wilcoxon test, $P < 6.06 \times 10^{-40}$). We predict that the 116 critical exons from ASD cases are constitutive and are part of most transcripts for the corresponding genes (with low mutational burden) and that their mutation is likely to have adverse effects (**Fig. 4a**). In one example, we confirmed the inverse correlation between the burden of missense mutations and expression levels for exons 5 and 9 of *SLC6A1* (in which *de novo* mutations are reported in ASD cases), showing that these exons are not subject to splicing (**Fig. 4b** and **Supplementary Table 3**). However, the highly expressed exons within *EPB41L3*, which was reported to have a loss-of-function mutation in an unaffected sibling, showed exon skipping (**Supplementary Fig. 9**).

Our analysis of a schizophrenia data set²⁵ showed strong association of *de novo* mutations in genes identified in cases for prenatal samples, consistently in prefrontal cortex regions (**Supplementary Fig. 10** and **Supplementary Table 4**). Next, we analyzed a set of 89 genes comprising loss-of-function mutations that were ascertained to be tolerated in humans²⁶. As expected, loss-of-function mutations ascertained from ASD cases showed an inverse correlation between the burden of rare missense mutations and expression, unlike loss-of-function mutations that were apparently tolerated (in individuals without ASD) (**Supplementary Fig. 11**). We also tested a more general grouping of genes involved in brain disorders in Online Mendelian Inheritance in Man (OMIM), and a positive association was observed ($P < 1.2 \times 10^{-9}$; OR = 1.73) in the cerebellum for genes involved in autosomal dominant disease (**Supplementary Fig. 12**). Further analysis identified associations in other genes under purifying selection (**Supplementary Fig. 13** and **Supplementary Table 5**).

For CNV analysis, we used *de novo* calls from the Autism Genome Project (AGP) (stages 1 (ref. 3) and 2 (ref. 27)) and two SSC cohorts (SSC1 and SSC2)^{28,29}. The combined AGP stage 1 and stage 2 data (2,446 ASD cases and 2,640 controls) included 101 *de novo* CNVs

(losses and gains; affecting 886 genes in cases and 544 genes in controls). SSC1 included 64 *de novo* calls (affecting 497 genes)²⁹, and SSC2 had 75 *de novo* CNVs (affecting 993 genes)²⁸. As with our findings for *de novo* SNV and indel mutations, we observed associations between microarray-based brain expression levels and the burden of rare missense mutations for genes ascertained through the identification of *de novo* CNVs in ASD cases (**Fig. 5a** and **Supplementary Fig. 14**). Interestingly, consistent with *de novo* mutation findings, the strongest association from CNV data was found predominantly with prefrontal cortex (Bonferroni-corrected Fisher's exact test, $P < 8.06 \times 10^{-16}$; OR = 1.47). When we further restricted the analysis to CNVs involving single-gene deletions, an overall higher OR was observed for CNVs ascertained in cases than for those ascertained in controls (**Supplementary Fig. 15**). We replicated these results using RNA sequencing data (**Supplementary Fig. 16** and **Supplementary Table 6**).

To develop a resource for the discovery of neuronal disease-associated genes, we applied the inverse correlation framework to multiple transcriptome data sets and identified 3,955 'brain-critical exons' (from 1,744 genes) with high expression specific to the brain and a low burden of rare mutations (2.01-fold more tissue-specific critical exons in brain than in 9 other tissues) (**Supplementary Fig. 17** and **Supplementary Table 7**). For example, *RBFOX1* and its targets *ATP2B1* and *GRIN1* were identified in this brain-critical exon data set, and this pathway can be dysregulated in ASD³⁰.

We tested for potential enrichment of the 3,955 brain-critical exons in candidacy tier data sets comprising fragile-X mental retardation protein (FMRP) targets³¹, the postsynaptic proteome (PSP)³², genes associated with ASD risk (autosomal dominant and X linked)⁵ and genes affected by predicted deleterious *de novo* mutations identified in ASD trio studies (considering 2 backgrounds: all RefSeq exons and only brain-expressed exons). Indeed, highly significant enrichment of brain-critical exons was detected for the FMRP targets, 41.8% (Bonferroni-corrected Fisher's exact test, $P < 2.91 \times 10^{-157}$; OR = 9.52); PSP, 28.11% ($P < 7.190 \times 10^{-54}$; OR = 4.43); genes associated with ASD risk (autosomal dominant or X linked), 34.56% ($P < 3.40 \times 10^{-11}$; OR = 6.08);

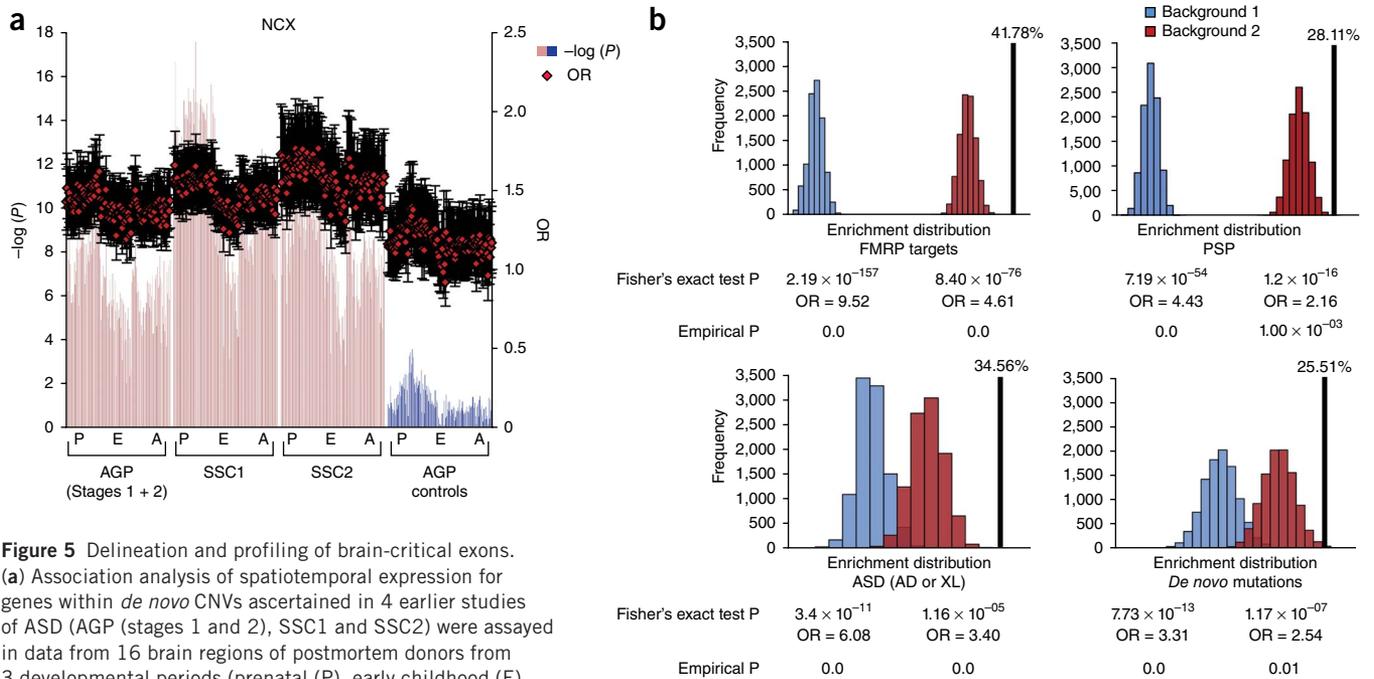


Figure 5 Delineation and profiling of brain-critical exons. (a) Association analysis of spatiotemporal expression for genes within *de novo* CNVs ascertained in 4 earlier studies of ASD (AGP (stages 1 and 2), SSC1 and SSC2) were assayed in data from 16 brain regions of postmortem donors from 3 developmental periods (prenatal (P), early childhood (E) and adult (A)). The plot shows spatiotemporal association results between exon expression levels (for 11 neocortex regions) and burden of rare missense mutations for genes within *de novo* CNVs reported in ASD cases (red bars) and in controls (blue bars). Ranges for OR values represent 95% CIs. (See **Supplementary Figs. 14 and 15** for amygdala, cerebellar cortex, hippocampus, mediodorsal nucleus of thalamus and striatum transcriptome association results and RNA sequencing validation.) (b) Brain-critical exon enrichment analysis on FMRP targets, PSP, ASD-associated genes (autosomal dominant (AD) or X linked (XL)) and genes with *de novo* mutations represented by black vertical lines and *P* values (Fisher's exact test, Benjamini-Hochberg corrected) relative to the two corresponding backgrounds (entire RefSeq exon set (blue) and exons with high expression in the brain (red)). Empirical *P* values were obtained from the permutation test for each background (10,000 random draws of an equal number of exons).

and genes with *de novo* mutation in ASD cases, 25.5% ($P < 7.73 \times 10^{-13}$; OR = 3.31) (**Fig. 5b**). As a negative control, we analyzed an equal number of exons (3,955) that are enriched for rare missense mutations and are highly expressed in the brain. Reanalysis of enrichment showed no significance after conducting permutation tests on the two backgrounds (**Supplementary Table 8**).

We examined the 1,744 genes encompassing these 3,955 brain-critical exons for enrichment in specific biological pathways (with stringent significance cutoffs of $P < 1.0 \times 10^{-3}$ and false discovery rate (FDR) < 0.01), identifying a network involving synapse regulation, neuron differentiation, signaling complexes and synaptic vesicles (**Supplementary Fig. 18** and **Supplementary Table 9**). These findings, which were made without previous knowledge of ASD-associated genes, overlap with results obtained using other data sets derived from known ASD-relevant genes^{3,27,29}.

We speculate that these 1,744 genes represent candidate susceptibility loci for ASD and/or related disorders. Of these genes, 518 (29.7%) have already been implicated as candidate genes for neuropsychiatric diseases. We therefore examined whether our critical exon index might help to differentiate the etiological gene(s) from bystanders colocalized within genomic disorder-associated loci involved in ASD. In the CNV at 16p11.2 (~600-kb segment encompassing 29 genes), 4 genes (*PRRT2*, *SEZ6L2*, *ASPHD1* and *KCTD13*) were found to harbor critical exons (**Supplementary Fig. 19** and **Supplementary Table 10**); 3 of these genes have been implicated in atypical ASD cases with smaller deletions³³ (**Supplementary Fig. 16**). Other examples are shown in **Supplementary Figures 20 and 21**.

Our exon transcriptome-mutation contingency index provides a new approach to characterize the penetrance of putative mutations. Two studies that couple gene-centric mutational data with brain

coexpression networks also implicate prenatal brain in ASD^{34,35}. However, these analyses did not exploit data for mutational burden and the full complement of *de novo* mutations³⁵, which we now show to add substantially to the assessment of exon expression. Our most significant data thus far come from studies of ASD, most likely because of the greater maturity of the data sets available, but also because ASD seems to be the neuropsychiatric condition most associated with rare and *de novo* mutations^{4,19,20,22}. From an evolutionary perspective, there seems to be selective pressure on specific isoforms of neuronal genes to contain particular exons and maintain the integrity of their expression. With more accurate lists of genes that confer susceptibility, as well as population control and gene expression data, studies of genotype and phenotype correlation for ASD will have increased power to model the effects of single and multiple variants.

METHODS

Methods and any associated references are available in the [online version of the paper](#).

Accession codes. Raw data from ASD CNV genotyping are available in the NCBI database of Genotypes and Phenotypes (dbGaP) under accession [phs000267.v4.p2](#).

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank the Centre for Applied Genomics for informatics support, the Allen Institute of Brain Science, the National Heart, Lung, and Blood Institute (NHLBI)

and the Autism Genome Project for sharing data. We thank J. Buchanan for critical review and editing of the manuscript. This work was supported by grants from the University of Toronto McLaughlin Centre, NeuroDevNet, Genome Canada and the Ontario Genomics Institute (project 4445), the Canadian Institutes for Health Research (CIHR) (FRN 74527 and FRNXGG818), the Canadian Institute for Advanced Research, the Canada Foundation for Innovation, the government of Ontario (GL2-01-013), the Ontario Brain Institute and Autism Speaks. R.K.C.Y. holds an Autism Speaks Meixner Fellowship in Translational Research. K.T. holds a fellowship from the Swedish Research Council. S.W.S. holds the GlaxoSmithKline-CIHR Chair in Genome Sciences at the University of Toronto and the Hospital for Sick Children.

AUTHOR CONTRIBUTIONS

M.U. and S.W.S. conceived the project, designed its components and contributed to the original concept of the project. K.T. performed quantitative PCR and RT-PCR analysis on brain tissues. M.U., G.P., D.M., P.H., R.K.C.Y., Z.W., D.P., L.L., T.N. and C.R.M. helped perform different components of the transcriptome and exome mutation analyses. M.U., G.P. and D.M. designed and performed the pathway analysis. B.A., B.J.F. and B.J.B. conducted the splicing code analysis. M.U., K.T. and S.W.S. coordinated the entire study and wrote the manuscript.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the [online version of the paper](#).

Reprints and permissions information is available online at <http://www.nature.com/reprints/index.html>.

- Devlin, B. & Scherer, S.W. Genetic architecture in autism spectrum disorder. *Curr. Opin. Genet. Dev.* **22**, 229–237 (2012).
- Neale, B.M. *et al.* Patterns and rates of exonic *de novo* mutations in autism spectrum disorders. *Nature* **485**, 242–245 (2012).
- Pinto, D. *et al.* Functional impact of global rare copy number variation in autism spectrum disorders. *Nature* **466**, 368–372 (2010).
- Sanders, S.J. *et al.* *De novo* mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
- Betancur, C. Etiological heterogeneity in autism spectrum disorders: more than 100 genetic and genomic disorders and still counting. *Brain Res.* **1380**, 42–77 (2011).
- Anney, R. *et al.* Individual common variants exert weak effects on the risk for autism spectrum disorders. *Hum. Mol. Genet.* **21**, 4781–4792 (2012).
- Rauch, A. *et al.* Range of genetic mutations associated with severe non-syndromic sporadic intellectual disability: an exome sequencing study. *Lancet* **380**, 1674–1682 (2012).
- Xu, B. *et al.* Exome sequencing supports a *de novo* mutational paradigm for schizophrenia. *Nat. Genet.* **43**, 864–868 (2011).
- Gratten, J., Visscher, P.M., Mowry, B.J. & Wray, N.R. Interpreting the role of *de novo* protein-coding mutations in neuropsychiatric disease. *Nat. Genet.* **45**, 234–238 (2013).
- Marshall, C.R. *et al.* Structural variation of chromosomes in autism spectrum disorder. *Am. J. Hum. Genet.* **82**, 477–488 (2008).
- Vaags, A.K. *et al.* Rare deletions at the neurexin 3 locus in autism spectrum disorder. *Am. J. Hum. Genet.* **90**, 133–141 (2012).
- Lionel, A.C. *et al.* Disruption of the *ASTN2/TRIM32* locus at 9q33.1 is a risk factor in males for autism spectrum disorders, ADHD and other neurodevelopmental phenotypes. *Hum. Mol. Genet.* **23**, 2752–2768 (2014).
- Beunders, G. *et al.* Exonic deletions in *AUTS2* cause a syndromic form of intellectual disability and suggest a critical role for the C terminus. *Am. J. Hum. Genet.* **92**, 210–220 (2013).
- Kang, H.J. *et al.* Spatio-temporal transcriptome of the human brain. *Nature* **478**, 483–489 (2011).
- Khurana, E. *et al.* Integrative annotation of variants from 1092 humans: application to cancer genomics. *Science* **342**, 1235587 (2013).
- Petrovski, S., Wang, Q., Heinzen, E.L., Allen, A.S. & Goldstein, D.B. Genic intolerance to functional variation and the interpretation of personal genomes. *PLoS Genet.* **9**, e1003709 (2013).
- Fu, W. *et al.* Analysis of 6,515 exomes reveals the recent origin of most human protein-coding variants. *Nature* **493**, 216–220 (2013).
- Tennesen, J.A. *et al.* Evolution and functional impact of rare coding variation from deep sequencing of human exomes. *Science* **337**, 64–69 (2012).
- Iossifov, I. *et al.* *De novo* gene disruptions in children on the autistic spectrum. *Neuron* **74**, 285–299 (2012).
- Jiang, Y.H. *et al.* Detection of clinically relevant genetic variants in autism spectrum disorder by whole-genome sequencing. *Am. J. Hum. Genet.* **93**, 249–263 (2013).
- O’Roak, B.J. *et al.* Exome sequencing in sporadic autism spectrum disorders identifies severe *de novo* mutations. *Nat. Genet.* **43**, 585–589 (2011).
- O’Roak, B.J. *et al.* Sporadic autism exomes reveal a highly interconnected protein network of *de novo* mutations. *Nature* **485**, 246–250 (2012).
- Vissers, L.E. *et al.* A *de novo* paradigm for mental retardation. *Nat. Genet.* **42**, 1109–1112 (2010).
- Barash, Y. *et al.* Deciphering the splicing code. *Nature* **465**, 53–59 (2010).
- Gulsuner, S. *et al.* Spatial and temporal mapping of *de novo* mutations in schizophrenia to a fetal prefrontal cortical network. *Cell* **154**, 518–529 (2013).
- MacArthur, D.G. *et al.* A systematic survey of loss-of-function variants in human protein-coding genes. *Science* **335**, 823–828 (2012).
- Pinto, D. *et al.* Convergence of genes and cellular pathways dysregulated in autism spectrum disorders. *Am. J. Hum. Genet.* **94**, 677–694 (2014).
- Levy, D. *et al.* Rare *de novo* and transmitted copy-number variation in autistic spectrum disorders. *Neuron* **70**, 886–897 (2011).
- Sanders, S.J. *et al.* Multiple recurrent *de novo* CNVs, including duplications of the 7q11.23 Williams syndrome region, are strongly associated with autism. *Neuron* **70**, 863–885 (2011).
- Voineagu, I. *et al.* Transcriptomic analysis of autistic brain reveals convergent molecular pathology. *Nature* **474**, 380–384 (2011).
- Darnell, J.C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
- Bayés, A. *et al.* Characterization of the proteome, diseases and evolution of the human postsynaptic density. *Nat. Neurosci.* **14**, 19–21 (2011).
- Crepel, A. *et al.* Narrowing the critical deletion region for autism spectrum disorders on 16p11.2. *Am. J. Med. Genet. B. Neuropsychiatr. Genet.* **156**, 243–245 (2011).
- Parikshak, N.N. *et al.* Integrative functional genomic analyses implicate specific molecular pathways and circuits in autism. *Cell* **155**, 1008–1021 (2013).
- Willsey, A.J. *et al.* Coexpression networks implicate human midfetal deep cortical projection neurons in the pathogenesis of autism. *Cell* **155**, 997–1007 (2013).

ONLINE METHODS

Exome data. We used data from the Exome Sequence Project (ESP) (February 2013) initiated by the National Heart, Lung, and Blood Institute (NHLBI) to calculate the burden of rare missense mutations in human populations^{17,18}. Variants were obtained from 6,503 exomes (2,203 African Americans and 4,300 European Americans). We used composite RefSeq gene annotation (which includes all exons from annotated isoforms) for our analysis, and, after excluding genes with no variant calls, a total of 20,417 genes (including protein-coding genes and pseudogenes) remained for analysis (**Supplementary Note**). For any exons within the composite gene model (hg19) (**Fig. 1**), variants annotated as missense (Genome Analysis Toolkit (GATK) annotation) for any gene isoform were considered. We calculated a normalized burden of mutations from the rare missense mutation count for each exon in the genome.

Transcriptome data. We used multiple transcriptome data sets. The first data set included expression data from the Affymetrix GeneChip Human Exon 1.0 ST array for 11 normal human tissues, each analyzed in triplicate³⁶. Tissues included cerebellum, breast, heart, liver, muscle, kidney, thyroid, pancreas, prostate, spleen and testis. The Partek Genomics Suite was used to compute core meta-probe set signal intensity from the raw .CEL files. Exons were kept for analysis only if they were also covered in ESP for variant calls (**Supplementary Note**).

The second data set used consisted of spatiotemporal expression profiles from 196 samples from the developmental human BrainSpan database¹⁴. Donors were selected so that each developmental period included at least two age- and sex-matched donors (**Supplementary Table 1**). The developmental periods were categorized into three groups: prenatal (8 to 21 weeks post-conception), early childhood (4 months to 3 years) and adulthood (≥ 13 years). For each donor, we obtained expression data from at least 13 brain regions, and 16 brain regions are common within the 3 developmental periods (**Supplementary Tables 1 and 2**). Spatiotemporal correlation coefficients are the Pearson's correlation coefficients for the burden of rare missense mutations with respect to the expression levels of exons in human brain tissues defined by brain region (spatial) and developmental stage (temporal).

Principal-component analysis. Principal-component analysis (PCA; R statistical software) was used to capture overall variance between components that might show different temporal expression patterns (using 196 microarray samples). The loadings of PCA are correlation coefficients between principal-component scores and the original variable. Principal-component loadings measure the importance of each variable in accounting for variability in the principal component. The PCA box plot (**Supplementary Fig. 1**) was constructed by extracting loading vectors for the gene sets from genome-wide spatiotemporal PCA. PCA was also used for tissue-wise clustering for the 11 tissue replicates.

ASD candidacy tier data set. This data set included a comprehensive curation of *de novo* damaging mutations (loss of function, frameshift, splicing and missense) from published whole-exome or whole-genome sequencing studies primarily on ASD and related neuropsychiatric conditions^{2,4,7,8,19–23}. Similarly, *de novo* damaging mutations were curated for unaffected siblings from two SSC studies^{4,19}. As a negative control, genes with previously predicted loss-of-function mutations that are systematically ascertained to be tolerated in humans were also used. We have tested association on *de novo* CNV data sets obtained from four large independent ASD studies (**Supplementary Note**). We further analyzed genes associated with ASD risk (autosomal dominant and X-linked)⁵, FRMP and FOXP2 targets^{31,37}, PSP and OMIM neuropsychiatric disease genes, both autosomal dominant and recessive (**Supplementary Note**).

Splicing code analysis. To analyze the implications of *de novo* SNVs (ascertained in cases and siblings) for alternative splicing patterns of the affected genes, we applied the 'splicing code' algorithm²⁴ on the variants. We first mapped the *de novo* mutations onto canonical transcripts and determined the target exon and its flanking introns and immediate neighboring exons. Then, we extracted all splicing code *cis* features for the mutation and its wild-type counterpart, applied the code and computed the predicted change in the inclu-

sion of the target exon. Exon inclusion, denoted by 'percent spliced-in' (Ψ), is defined as the fraction of times an exon is included in the final transcript³⁸. For each *de novo* mutation, the splicing code predicts a $\Delta\Psi$ value; if the $\Delta\Psi$ value of a mutation was greater than 95% of the $\Delta\Psi$ values for common polymorphisms (in the dbSNP database), it was considered significant. The splicing code analysis found only 73 *de novo* mutations (from cases and siblings) passing the $\Delta\Psi$ threshold of -0.01 (**Supplementary Table 11**). Code-predicted $\Delta\Psi$ values showed no significant difference for the *de novo* mutations identified in cases and siblings (Kolmogorov-Smirnov (K-S) test and AUC (area under the curve); **Supplementary Fig. 6**).

Next, applying two additional normalizations, we again analyzed whether exons that had variable expression (according to developmental time period or tissue type) were the exons that exclusively contributed to association between spatiotemporal expression and burden of rare missense mutations. For each gene (g), we computed differential expression by applying two transformations on each exon, $e_{d,i}$: (i) subtraction of exon expression e_i from the mean expression for the gene

$$e_{d,i} = e_i - \frac{1}{n} \sum_i^n g(e_i)$$

and (ii) subtraction of the total expression from the median expression for the gene, $e_{d,i} = e_i - \text{median}(g(e))$. After applying both normalizations independently to the BrainSpan data, the association test for genes with *de novo* SNVs (as ascertained from cases or unaffected siblings) using spatiotemporal data from brain samples was not significant. This finding suggested that these exons are not the only ones contributing to the association signal, which is consistent with our splicing code analysis.

Detection of 'tissue-specific exons'. To identify tissue-specific exons that were highly expressed only in a particular tissue and showed suppressed accumulation of rare missense mutations, we used microarray transcriptome data from 11 human tissues and human developing brain data from BrainSpan (**Supplementary Note**).

RT-PCR and quantitative RT-PCR. mRNA expression levels and alternative splicing events in selected genes were analyzed by RT-PCR and quantitative RT-PCR (qRT-PCR). To detect splicing events, primer pairs were designed to prime from nearby flanking exons (**Supplementary Table 3**). To quantify brain-critical exons by qRT-PCR, we designed primers within the specific exon or within the specific exon and the flanking exon (**Supplementary Table 10**). cDNA synthesis, RT-PCR and qRT-PCR were performed using standard conditions (**Supplementary Note**).

Permutation test. To show enrichment of brain-critical exons within four candidacy tier gene sets (FMRP, PSP, autosomal dominant and X-linked genes in ASD, and genes with *de novo* mutations in ASD), a permutation test was performed in two backgrounds. The first background included exons from the entire genome (RefSeq). The second background was used to test an even narrower hypothesis, and only exons with high brain expression were used (**Supplementary Note**) as background. For each test, a set of 3,955 exons was drawn 10,000 times from each background and used to compute the significance of brain-critical exon enrichment (**Fig. 5b**). As a negative control, we selected the top 3,955 exons in the genome that had a highly enriched burden of rare missense mutations and were expressed in brain for similar enrichment analysis (Fisher's exact test and permutation test) within the four candidacy tier gene sets (**Supplementary Table 8**).

Pathway enrichment analysis and network construction. Pathway analysis was performed to identify biologically meaningful pathways. Gene sets from Gene Ontology (GO) and pathways from the National Cancer Institute at the US National Institutes of Health (NCI-NIH), the Kyoto Encyclopedia of Genes and Genomes (KEGG) and Reactome were used for analysis, and significance was determined by applying Fisher's exact test. A permutation test (10,000 permutations) was also used to assess the significance of pathway analysis for each gene set, and FDR was computed using the Benjamini-Hochberg

procedure. The significantly enriched set (Fisher's exact test $P < 1.0 \times 10^{-3}$ and FDR < 0.01) was used for network construction.

36. Gardina, P.J. *et al.* Alternative splicing and differential gene expression in colon cancer detected by a whole genome exon array. *BMC Genomics* **7**, 325 (2006).
37. Ayub, Q. *et al.* FOXP2 targets show evidence of positive selection in European populations. *Am. J. Hum. Genet.* **92**, 696–706 (2013).
38. Katz, Y., Wang, E.T., Airoldi, E.M. & Burge, C.B. Analysis and design of RNA sequencing experiments for identifying isoform regulation. *Nat. Methods* **7**, 1009–1015 (2010).