Engineering microdeletions and microduplications by targeting segmental duplications with CRISPR

Derek J C Tai1–4, Ashok Ragavendran1,3,4, Poornima Manavalan1,3,4, Alexei Stortchevoi1,3,4, Catarina M Seabra1–5, Serkan Erdin1,3,4, Ryan L Collins1,3,4, Ian Blumenthal1, Xiaoli Chen6, Yiping Shen7, Mustafa Sahin8, Chengsheng Zhang9, Charles Lee9,10, James F Gusella1,11,12 & Michael E Talkowski1–4,11

Recurrent, reciprocal genomic disorders resulting from non-allelic homologous recombination (NAHR) between near-identical segmental duplications (SDs) are a major cause of human disease, often producing phenotypically distinct syndromes. The genomic architecture of flanking SDs presents a challenge for modeling these syndromes; however, the capability to efficiently generate reciprocal copy number variants (CNVs) that mimic NAHR would represent a valuable modeling tool. We describe here a CRISPR/Cas9 genome engineering method, single-guide CRISPR/Cas9 targeting of repetitive elements (SCORE), to model reciprocal genomic disorders and demonstrate its capabilities by generating reciprocal CNVs of 16p11.2 and 15q13.3, including alteration of one copy-equivalent of the SDs that mediate NAHR in vivo. The method is reproducible, and RNA sequencing reliably clusters transcriptional signatures from human subjects with in vivo CNVs and their corresponding in vitro models. This new approach will provide broad applicability for the study of genomic disorders and, with further development, may also permit efficient correction of these defects.

Recurrent microdeletion and microduplication syndromes (rMDS) are among the most common causes of human neurodevelopmental and psychiatric disorders. These recurrent rearrangements are mediated by NAHR, which occurs between two highly homologous SDs that flank a genomic segment and can result in either copy loss (microdeletion) or the reciprocal copy gain (microduplication) of this segment.1 In one study, the prevalence of five common NAHR-mediated CNVs was estimated to be 0.47% of consecutive newborns2, and the rate of recurrent rMDS in prenatal samples referred for diagnostic screening has been estimated to be 1.5%, emphasizing the relevance of the rearrangements to developmental abnormalities3.

Reciprocal CNV of a small segment of chromosome 16p11.2 (OMIM 611913) is a common rMDS that has been associated with intellectual disability, autism spectrum disorder, schizophrenia and other neuropsychiatric disorders, as well as anthropometric traits, including obesity4. Like all such NAHR-mediated genomic disorders, 16p11.2 rMDS involves gain or loss of a unique genic segment and one copy-equivalent of the SD. The unique genic segment of the 16p11.2 CNV spans 593 kb (ref. 5) containing 47 genes, of which 28 are annotated as protein coding (based on Ensembl GRCh37 V71 (ref. 6)). It is flanked by parallel and highly homologous (~99% identity) SDs, each spanning 147 kb (ref. 5) and containing 6 duplicated genes (5 annotated as protein coding6). The mechanism of NAHR-mediated CNV formation in vivo involves the mispairing of the flanking SDs, which can result in either the loss or gain of a 740-kb segment encompassing one copy of the unique genic segment and one copy-equivalent of the SD7. Disentangling the effects of the entire CNV and of individual genes in it are confounded by several factors in studies of humans and their cells, including variable genetic background, the uncertain relevance of available peripheral tissues and the inability to collect large cohorts. Meanwhile, animal models may be limited in their generalizability to humans8,9. New methods to model rMDS alterations in induced pluripotent stem cells (iPSCs) that can then be differentiated into relevant tissues, including neural precursors and derivative neural cells, would open access to innovative approaches to understanding the impact of rMDS in human neurodevelopment.

The clustered regularly interspaced short palindromic repeats (CRISPR)/CRISPR-associated proteins (Cas) complex was discovered as a ribonucleoprotein interfering complex involved in bacterial adaptive immunity10, and it has now evolved into a programmable and broadly applicable genome engineering tool11. Like zinc finger nucleases (ZFNs)12 and transcription activator-like effector nucleases (TALENs)13, CRISPR/Cas has inherent endonuclease activity

1Molecular Neurogenetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. 2Department of Neurology, Harvard Medical School, Boston, Massachusetts, USA. 3Psychiatric and Neurodevelopmental Genetics Unit, Center for Human Genetic Research, Massachusetts General Hospital, Boston, Massachusetts, USA. 4Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts, USA. 5Institute of Biomedical Sciences Abel Salazar, University of Porto, Porto, Portugal. 6Beijing Municipal Key Laboratory of Child Development and Nutriomics, Capital Institute of Pediatrics, Beijing, China. 7Department of Laboratory Medicine, Boston Children’s Hospital, Boston, Massachusetts, USA. 8Department of Neurology, Boston Children’s Hospital, Boston, Massachusetts, USA. 9The Jackson Laboratory for Genomic Medicine, Farmington, Connecticut, USA. 10Department of Graduate Studies – Life Sciences, Ewha Womans University, Seoul, Korea. 11Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, Massachusetts, USA. 12Department of Genetics, Harvard Medical School, Boston, Massachusetts, USA. Correspondence should be addressed to M.E.T. (talkowski@chrg.mgh.harvard.edu) or J.F.G. (gusella@helix.mgh.harvard.edu).

Received 8 May 2015; accepted 7 January 2016; published online 1 February 2016; doi:10.1038/nn.4235
that can generate in DNA a double-strand break when it encounters its target along with a proto-spacer-associated motif (PAM) sequence, 5′-NGG, where N is any nucleotide. Most current applications of CRISPR/Cas rely on a single Cas protein in the type II CRISPR/Cas system, Cas9, which in its host bacteria forms a complex with CRISPR RNA (crRNA) and trans-activating crRNA. In the laboratory, a single engineered RNA (a guide RNA, or gRNA) as a chimer of a crRNA and a fixed trans-activating crRNA can direct Cas9 to a specified target sequence that is complementary to crRNA. One of the most common applications for the CRISPR/Cas system is targeted genome editing in cells and model organisms. To generate point mutations or deletions of a genic segment, Repair of gRNA-directed Cas9 nuclease-induced double-strand breaks through non-homologous end-joining (NHEJ) can induce indel mutations that abolish normal expression of the target gene. To introduce precise point mutations or insertions at a target gene locus cleaved byCas9, a donor template carrying a mutation can be delivered into the cell to permit repair by homologous recombination. In addition, pairs of gRNAs can be used in a dual-guide RNA (dgRNA) strategy to produce deletions or other genomic rearrangements.

In this study, we present a method, SCORE, for efficiently producing microdeletions and microduplications comparable to those associated with reciprocal genomic disorders in humans as a consequence of NAHR between flanking SDs. To accomplish this, SCORE uses a single-guide RNA (sgRNA) strategy to directly target perfectly homologous sequences in the SDs. We demonstrate the capability of the method to rapidly and efficiently generate reciprocal CNVs using two proof-of-principle rMDS regions of differing size and composition, 16p11.2 and 15q13.3. In each region, the method generates reciprocal CNVs that mimic the products of the NAHR mechanism, including deletion or duplication of one copy-equivalent of the Sds, with reproducible rates of efficiency. These data suggest that SCORE can open widespread access to modeling these common causes of human developmental anomalies.

RESULTS

Generation of reciprocal 16p11.2 CNVs using CRISPR/Cas9

We postulated that optimizations in emerging genome engineering technologies could be applied to human iPSCs to generate reciprocal CNV directly comparable to those in recurrent NAHR-mediated rMDS. We initially tested this hypothesis on the common 16p11.2 rMDS. An overview of our experimental design is shown in Figure 1a. All studies used a previously described control iPSC line (8330) as the test sample. As positive controls, we used iPSC lines generated from subjects harboring 16p11.2 rMDS that were previously detected from clinical screening (four subjects with microdeletion, three subjects with microduplication) as well as family-based controls (see Online Methods; Supplementary Fig. 1). We refer herein to the distal SD (closest to the telomere) as SDA and the proximal SD (closest to the centromere) as SDB. It has been previously shown that the CRISPR/Cas system can efficiently generate predictable deletions and inversions using a dgRNA strategy. We first used this proven approach to create heterozygous deletions of a contiguous 575-kb sequence containing all genes in the unique genic segment as an initial comparison group (Fig. 1b). This initial experiment used two independent and uniquely mapping dgRNAs targeted within the contiguous genic segment between the Sds (see Online Methods). This experiment generated the predicted 575,414-bp microdeletion while leaving the flanking Sds fully intact (Ensembl GRCh37 V7.1) chromosome 16: 29,624,445–30,199,859 (Fig. 1b). Sanger sequencing confirmed the precise breakpoint and revealed NHEJ as the likely repair mechanism (Supplementary Fig. 2). However, this dgRNA strategy for deleting the unique genic segment does not accurately model reciprocal rMDS in humans, where NAHR alters this unique segment and one copy-equivalent of the Sds.

To create a method that would mimic the consequences of the in vivo NAHR mechanism, we hypothesized that transfecting a sgRNA that targets perfectly homologous sites in each flanking SD could generate a 740-kb microdeletion of the 16p11.2 rMDS that is mediated by NAHR; that is, one encompassing the unique genic segment and one copy-equivalent of each gene in the Sds, mirroring the size of the CNV in humans. We evaluated the potential of this new single-guide method, SCORE, to model rMDS lesions (Fig. 1b). We targeted a homologous site in each of SDA and SDB to promote simultaneous DNA breaks in both Sds, but not elsewhere in the genome (Fig. 1b; see Online Methods for details of sgRNA design). We predicted this strategy would promote formation of a 739,346-bp deletion (chromosome 16: 29,487,574–30,226,919) comprising all of the 593-kb unique genic segment and one copy-equivalent of the Sds, as is produced by the in vivo mechanism. We used the sgRNA CRISPR transfection methods described in Online Methods, followed by fluorescence-activated cell sorting (FACS) to isolate single iPSCs into individual wells of Matrigel-coated 96-well plates. Once clonal iPSC colonies were formed, relevant gene dosage was measured by copy number screening. This screening readily identified CRISPR/Cas-modified clones that showed decreased dosage consistent with the expected microdeletion. The assay also identified clones with increased 16p11.2 dosage consistent with apparent microduplication, suggesting that our SCORE method can promote predictable reciprocal dosage imbalances that mimic the in vivo consequences of NAHR.

Replication and characterization of 16p11.2 CNV iPSCs

To estimate the efficiency of the approach, the entire experiment was repeated and 114 clones were systematically screened by copy number analysis (Fig. 1c). In this replicate experiment, 12 clones harbored the expected 740-kb deletion, suggesting an overall efficiency of 10.5% for microdeletion, and 4.3% of clones harbored a putative microduplication (12 and 5, respectively, of 114 clones screened). Further analyses of select dosage-altered clones were then performed by strand-specific transcriptome sequencing (RNAseq; 23 total lines) and genome-wide DNA microarray analysis to screen for on-target and off-target CNVs (Supplementary Fig. 3). Dosage microarray and RNAseq confirmed the observed microdeletions and microduplications (Figs. 1 and 2). Notably, no off-target CNVs were detected from genome-wide analyses (Supplementary Fig. 4). We did not obtain homozygous deletions of the 16p11.2 rMDS region, suggesting that total absence of one or more genes in this segment may be incompatible with iPSC viability in our system.

To provide insight into the rearrangement mechanism associated with SCORE-generated iPSCs, we attempted to clone and sequence the breakpoints generated by CRISPR/Cas within the Sds. To our knowledge, identification of the precise exchange points in Sds of 16p11.2 rMDS subjects has not been accomplished owing to the inability to distinguish the SDA and SDB sequence at most sites. However, in our experiments, we were able to explore the predicted sgRNA targeting sites for potential alterations by performing cloning and Sanger sequencing of the region. We found clear junctions among the 12 CRISPR-treated 740-kb microdeletion lines and 5 microduplication lines (Supplementary Fig. 5a,b). Three different types of indel mutation were found in three microdeletion lines (Del 1–3), two mutations were found in one deletion line (Del 4), a single-nucleotide T insertion was found in four deletion lines (Del 5–8),
and four deletion junctions contained no evidence of new mutations at the breakpoints, suggesting potential homology-mediated repair (Del 9–12) (Supplementary Fig. 5a).

Intriguingly, only a single-nucleotide T insertion was found in four of five CRISPR-treated 740-kb microduplication lines (Supplementary Fig. 5b). These data suggest disparate mechanisms of repair for these CRISPR-treated microdeletion and microduplication lines, which do involve nonhomologous or microhomology-mediated mechanisms but likely also involve homology-mediated mechanisms, potentially including NAHR.

RNAseq was performed for 23 lines, including CRISPR-treated 575- and 740-kb microdeletion lines, CRISPR-treated 740-kb microduplication lines, 16p11.2 patient lines with either microdeletion23 or microduplication, and two types of control lines: those from CRISPR treatment that did not produce a CNV and those from family member controls of 16p11.2 CNV carriers (Supplementary Fig. 3). RNAseq analyses revealed reduced expression and overexpression of all genes in the unique genic segment for the 740-kb SCORE-generated microdeletion and microduplication lines, respectively (Fig. 2a). Moreover, we observed expression patterns reflecting the loss of one copy of the genes localized to the SDs from the 740-kb deletion lines (reduction from four to three total copies of the SDs across the two homologous chromosomes) and the gain of one extra copy of the SDs in the microduplication lines (increase from four to five copies) (Fig. 2a). These in vitro results were consistent with the transcriptional patterns described previously in lymphoblastoid cell lines from families harboring the 16p11.2 rMDS CNV9. As expected, there was no reduction of genes in the SDs from RNAseq of the 575-kb microdeletion line. Hierarchical clustering from the transcriptome data of the 16p11.2 rMDS region revealed that all CRISPR-treated deletion and duplication lines clustered together with the corresponding patient deletion and duplication lines, respectively, while control lines clustered together with the control family member (Fig. 2b). In addition, western blot analysis of MAPK3, encoded in the unique genic segment, also confirmed altered levels of protein expression in our CRISPR-treated iPSC lines with 16p11.2 CNVs (Supplementary Fig. 6).

We applied statistical analyses based on general linear models to compare each genotype (deletion or duplication) to controls at the gene level. We also assessed results from a linear model requiring gene dosage to be positively or negatively correlated with copy state (for example, a linear trend in gene expression must be observed with copy state = 1, 2 or 3), as in a previous study in families harboring 16p11.2 CNV and in the cortex of mouse models6. We observed largely identical results in the CRISPR-treated iPSCs to those previously detected; all 32 genes detected by RNAseq (25 protein coding) in the region showed a linear increase in expression with copy number. In the 740-kb microdeletion and microduplication lines, 28 genes (22 protein coding) and 25 genes (20 protein coding) were significantly downregulated and upregulated, respectively, relative to controls (nominal P < 0.05) (Fig. 2a). Genome-wide differential expression was most significant for genes in the 16p11.2 rMDS region (Fig. 2c), again replicating previous findings in humans and mouse models5,9,24. These results imply that *in vitro* models yield comparable expression profiles to those of human iPSCs harboring the *in vivo* 16p11.2 CNV.

**Replication of the SCORE approach in the 15q13.3 rMDS**

To demonstrate generalizability of the SCORE method, we performed a second study, targeting another common, yet much larger, rMDS region encompassing a 1.771-Mb unique genic segment and 218-kb flanking SDs in 15q13.3. Recurrent rMDS of this region have been associated with mental retardation, epilepsy, schizophrenia, autism, obesity and variable facial and digital dysmorphisms (OMIM 612001)25–27. The unique genic segment contains seven protein encoding genes (Ensembl GRCh37 chromosome 15: 30,910,306–32,445,406) and, when combined with one copy-equivalent of the flanking SVs, the
Figure 2  Gene expression characterization of putative CNV lines. (a) Relative fold change of gene expression based on RNAseq in the 16p11.2 rMDS and SD regions is shown for the selected CRISPR/Cas9 575-kb deletion line (open red circles), CRISPR/Cas9 740-kb deletion lines (filled red circles), and CRISPR/Cas9 740-kb microduplication lines (filled blue circles). Fold changes were calculated as the mean difference of expression for each gene between the CNV lines and all controls based on contrasts from the generalized linear model; error bars represent the back-transformed mean difference ± s.e.m. (b) Heat map of expression, estimated as counts per million mapped reads, for all genes in the 16p11.2 rMDS and SD regions and hierarchical clustering, using average linkage, for all CRISPR/Cas9 treated lines and patient iPSC lines. (c) Genome wide P-value distribution of genes from linear model of expression as a linear function of the CNV. Chromosome 16 is highlighted in red; we see the strongest signal from the genes in the 16p11.2 rMDS region, as reported in previous studies.8

Figure 3  Generation of 15q13.3 rMDS in human iPSC by CRISPR/Cas9. (a) Illustration of targeted 15q13.3 rMDS segment and flanking SDs, with all protein-coding genes shown (Ensembl GRCh37). The sgRNA targeting the SDs to promote a model of NAHR-mediated CNV is indicated in red (target sequence: 5′-CTTATGGATTGCAGGAC-3′, chromosome 15: 30,792,593–30,792,611 and 32,799,503–32,799,521). (b) Copy number screening of all genes in the segment identified eleven iPSC lines harboring 15q13.3 rMDS (deletion and duplication). Experiments are in triplicate. Data are mean ± s.e.m. (c) Efficiency of CRISPR/Cas9 generation of this 2-Mb microdeletion and microduplication using the SCORE approach for this larger, ~2 Mb CNV was 5.5% for microdeletion (9 of 164 clones screened) and 1.2% for microduplication (2 of 164 clones screened; Fig. 3c). Notably, these relatively high efficiencies for isolation and clonal expansion. Copy-number screening of 164 CRISPR-treated iPSC colonies that were expanded from single cells in the initial experiment revealed 9 microdeletions and 2 microduplications (Fig. 3b). In contrast to the 16p11.2 CNV, we identified two clones that were consistent with homozigous microdeletion of the 15p13.3 segment. We observed no dosage increase beyond a single duplication, and no dosage changes were detected for genes flanking the CRISPR-targeted region (Supplementary Fig. 7). The overall efficiency of the SCORE approach for this larger, ~2 Mb CNV was 5.5% for microdeletion (9 of 164 clones screened) and 1.2% for microduplication (2 of 164 clones screened; Fig. 3c). Notably, these relatively high efficiencies for...
such large CNVs were detected based on a single initial experiment without optimization. Further development in this or other regions could result in still higher efficiencies.

**DISCUSSION**

We show here that carefully designed CRISPR/Cas9 genome engineering can emulate the genetic architecture of NAHR-mediated recurrent genomic disorders by generating lesions that mimic those produced in vivo. Previous studies have shown that dual-guide CRISPR/Cas9 genome editing can generate multiple classes of structural variation mediated by non-homologous repair within unique genomic segments across a range of sizes\(^\text{18-20}\). This SCORE approach demonstrates that it is possible to model the outcome of NAHR-mediated rMDS with relatively high efficiency, including reciprocal dosage imbalance of one copy-equivalent of the flanking SDs. SCORE thereby opens access to genome-wide modeling of these common causes of human congenital anomalies, including interrogation of the unique genic segment and the SDs in any rMDS region.

The junction sequences observed in the CRISPR/Cas-generated CNVs suggest that different molecular mechanisms may predominate in the generation of microdeletions and microduplications using SCORE (Supplementary Fig. 5). A mechanism based purely on NAHR that produced coincident deletion and duplication events in a single cell would not be detected in our screen for altered dosage. However, if NAHR-mediated alterations affected sister chromatids, segregation of the events in mitosis before FACS sorting would then lead to individual cells with either deletion or duplication, which might be expected to be recovered in roughly equal numbers. The junction sequences in fact recovered suggest that non-NAHR mechanisms also operate in this system. Among the SCORE-generated microduplication lines, we observed evidence of both NHEJ and microhomology-mediated repair for the microdeletions, as well as wild-type sequence for some microdeletions or a single insertion of a T nucleotide at the breakpoint (Supplementary Fig. 5b). These data suggest that homology-mediated repair, possibly including NAHR, may be important in the production of microduplications and of a fraction of microdeletions generated using this method.

To minimize the potential for mosaicism, we performed FACS sorting to obtain single cells, followed by clonal expansion. Nonetheless, sister chromatid exchange could occur during clonal expansion, leading to low levels of mosaicism in the final cell population. This would be challenging to detect. Scrutiny of the RNAseq and microarray results did not suggest mosaicism in a meaningful fraction of cells, though we cannot exclude this possibility. It is also possible that deletion may occur on one chromosome and duplication on the other, leading to a balanced dosage state that would not be detected by our dosage screening, though such events would be expected to be rare. Further experiments will be needed to tease apart these varied mechanisms of repair and their relative efficiencies.

The generation of 15q13.3 rMDS CNV also confirmed that relatively large reciprocal CNV models can be derived with the SCORE approach with an efficiency that is comparable to that of the much smaller 740-kb 16p11.2 CNVs initially generated. Previous studies suggest that dual-guide CRISPR that is reliant upon NHEJ has less than 1% efficiency for deletions greater than 100 kb (ref. 19). By contrast, we found that the 10.5% SCORE efficiency of generating the 740-kb 16p11.2 microdeletion with a single cut site was reduced only modestly to 5.5% when the deletion size increased to the 2 Mb of 15q13. For microduplication, the corresponding efficiency decrease was from 4.3% to 1.2%. This suggests that relatively efficient modeling of rMDS will be tractable across a broad size distribution. It may also indicate that the participation of homology-mediated repair contributes to a greater efficiency of CNV formation than in methods relying on non-homologous repair alone. The precise size limitations of the SCORE approach remains to be determined, but we note that the 15q13.3 CNV is one of the larger recurrent reciprocal genomic disorders commonly observed in humans (1 in 40,000 for population incidence\(^\text{23}\), 0.48% of cases of neurodevelopmental abnormalities\(^\text{20}\)).

In conclusion, we describe here a method using a sgRNA that targets highly homologous SDs flanking unique genomic segments and thereby derives recurrent, reciprocal dosage imbalances. This method models the consequences of NAHR, which drives many of the most common causes of human congenital anomalies. Our experiments reveal that SCORE can be rapidly implemented to produce rMDS CNVs of different size and copy state; we observed 0, 1, 2 and 3 copies of the rMDS regions. Genome-wide analyses suggest that the method generates models with high fidelity, as we did not observe off-target CNVs following genome-wide screening, and transcriptome studies of the in vitro edited lines recapitulated previous findings in rMDS families and mouse models. We provide complete details on the design and application of this method, which will enable modeling of rMDS in multiple tissue types and, with further development and optimization, could provide a tractable route to in vitro correction of these common genomic imbalances.

**METHODS**

Methods and any associated references are available in the online version of the paper.

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

**ACKNOWLEDGMENTS**

We thank S. Haggarty and S. Sheridan (Center for Human Genetic Research, Massachusetts General Hospital) for generously providing the control iPSC line. These studies were supported by funding from the Simons Foundation for Autism Research (SFARI 328656 and 346042 (M.E.T.) and SFARI 308955 (J.F.G.)), the Nancy Lurie Marks Family Foundation (J.F.G. and M.E.T.), the US National Institutes of Health (R01NS093200 (J.F.G. and M.E.T.), R00MH095867 (M.E.T.), P01GM061354 (J.F.G. and M.E.T.)), the March of Dimes (M.E.T.), NARSAD (M.E.T.) and Autism Speaks (J.F.G.). Research reported in this publication was also partially supported by US National Cancer Institute award P30CA034196 (C.L.) and National Human Genome Research Institute award U41HG007497 (C.L.). We gratefully acknowledge the resources provided by the AGRE consortium and the participating AGRE families. AGRE is a program of Autism Speaks and is supported, in part, by grant 1U24MH081810 from the US National Institute of Mental Health to C.M. Lajonchere. C.L. is supported by an Ewha Womans University Distinguished Professorship. D.J.C.T. is a recipient of the Postdoctoral Research Abroad Program sponsored by Ministry of Science and Technology of Taiwan (102-2917-I-564-012).

**AUTHOR CONTRIBUTIONS**

M.E.T., J.F.G., D.J.C.T., A.S. and I.B. conceived and designed the studies. D.J.C.T., P.M., C.M.S. and A.S. performed molecular studies. A.R., R.L.C. and S.E. generated lines recapitulated previous findings in rMDS families and mouse models. We provide complete details on the design and application of this method, which will enable modeling of rMDS in multiple tissue types and, with further development and optimization, could provide a tractable route to in vitro correction of these common genomic imbalances.

**COMPETING FINANCIAL INTERESTS**

The authors declare no competing financial interests.

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

1. Stankiewicz, P. & Lupski, J.R. Genome architecture, rearrangements and genomic disorders. Trends Genet. 18, 74–82 (2002).
2. Tucker, T. et al. Prevalence of selected genomic deletions and duplications in a French-Canadian population-based sample of newborns. Mol. Genet. Genomic Med. 1, 87–97 (2013).

3. Konialis, C. et al. Uncovering recurrent microdeletion syndromes and subtelomeric deletions/duplications through non-selective application of a MLPA-based extended prenatal panel in routine prenatal diagnosis. Prenat. Diagn. 31, 571–577 (2011).

4. Maillard, A.M. et al. Expression of the genome in memory. N. Engl. J. Med. 368, 208–222 (2013).

5. Weiss, L.A. & Joung, J.K. CRISPR-Cas systems for editing, regulating and targeting genomes. Nat. Rev. Genet. 16, 752–764 (2015).

6. Golzio, C. & Delsied, G. Modeling of the human genome in cancer cells. Mol. Cancer 14, 1–15 (2015).

7. Walters, R.G. et al. The genetic basis of autism spectrum disorder. Trends Genet. 32, 449–460 (2016).

8. Flicek, P. et al. Ensembl 2013. Nucleic Acids Res. 41, D803–D811 (2013).

9. Gasiunas, G., Barrangou, R., Horvath, P. & Siksnys, V. Cas9-crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria. Proc. Natl. Acad. Sci. USA 109, 8239–8244 (2012).

10. Sander, J.D. et al. CRISPR-Cas systems: a consensus framework for genome editing. Nat. Rev. Genet. 17, 255–268 (2016).

11. Sharp, A.J. & Joung, J.K. CRISPR-Cas systems for genome editing. Nat. Biotechnol. 33, 1122–1139 (2015).

12. Canver, M.C. et al. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. 289, 21312–21324 (2014).

13. Mandal, P.K. et al. Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. Cell Stem Cell 15, 643–652 (2014).

14. Platt, R.J. et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell 159, 440–455 (2014).

15. Mandai, P.K. et al. Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. Cell Stem Cell 15, 643–652 (2014).

16. Vassos, E. et al. Penetration for copy number variants associated with schizophrenia. Hum. Mol. Genet. 19, 3477–3481 (2010).

17. Platt, R.J. et al. CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell 159, 440–455 (2014).

18. Mandai, P.K. et al. Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9. Cell Stem Cell 15, 643–652 (2014).

19. Canver, M.C. et al. Characterization of genomic deletion efficiency mediated by clustered regularly interspaced palindromic repeats (CRISPR)/Cas9 nuclease system in mammalian cells. J. Biol. Chem. 289, 21312–21324 (2014).

20. Xiao, A. et al. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. Nucleic Acids Res. 41, e141 (2013).

21. Veres, A. et al. Low incidence of off-target mutations in individual CRISPR-Cas9 and TALEN targeted human stem cell clones detected by whole-genome sequencing. Cell Stem Cell 15, 27–30 (2014).

22. Sheridan, S.D. et al. Epigenetic characterization of the FMR1 gene and aberrant neurodevelopment in human induced pluripotent stem cell models of fragile X syndrome. PLoS One 6, e26203 (2011).

23. Shen, Y. et al. Intra-family phenotypic heterogeneity of 16p11.2 deletion carriers in a three-generation Chinese family. Am. J. Med. Genet. B Neuropsychiatr. Genet. 156, 226–232 (2011).

24. Kumar, R.A. et al. Recurrent 16p11.2 microdeletions in autism. Hum. Mol. Genet. 17, 628–638 (2008).

25. Sharp, A.J. et al. A recurrent 15q13.3 microdeletion syndrome associated with mental retardation and seizures. Nat. Genet. 40, 322–328 (2008).

26. van Bon, B.W. et al. Further delineation of the 15q13 microdeletion and duplication syndromes: a clinical spectrum varying from non-pathogenic to a severe outcome. J. Med. Genet. 46, 511–523 (2009).

27. Stefansson, H. et al. Large recurrent microdeletions associated with schizophrenia. Nature 455, 232–236 (2008).

28. Antonacci, F. et al. Palindromic GOLGA8 core duplicons promote chromosome 15q13.3 microdeletion and evolutionary instability. Nat. Genet. 46, 1293–1302 (2014).

29. Li, J. et al. Efficient inversions and duplications of mammalian regulatory DNA elements and gene clusters by CRISPR/Cas9. J. Mol. Cell Biol. 7, 284–298 (2015).

30. Vassos, E. et al. Prevalence of selected genomic deletions and duplications in a French-Canadian population-based sample of newborns. Mol. Genet. Genomic Med. 1, 87–97 (2013).
ONLINE METHODS

Guide RNA design and preparation. We used the CRISPR Design Tool (http://tools.genome-engineering.org/) to obtain the sequences of guide RNAs targeting the 16p11.2 rMDS segment. The design of guide RNAs and reference sequence are based on genome assembly GRCh37. To generate the 575-kb microdeletion that includes all unique genes in the rMDS segment, the sequences of dual-guide RNAs were as follows:

- gRNA427: 5′-GCAGTTGCAAGCCATGTAGCT-3′, chromosome 16: 29,624,428–29,624,447
- gRNA841: 5′-GCCTGACACGGGCCAGG-3′, chromosome 16: 30,199,842–30,199,861

To design the sgRNAs targeting 16p11.2 SDs and 15q13.3 SDs with the SCORE approach, we first identified all possible 18–25mer guides with Jellyfish and performed a degenerate BLAST search to identify sequences that would uniquely target the 16p11.2 and 15q13.3 SDs, respectively, with no predicted off-target effects. Additionally, for the 15q13.3 guides, we refined our original guide design methodology by using a recently released tool, Off-Spotter, to screen out candidate guides with less than four mismatches across nontarget regions, further mitigating any possible off-target effects. Finally, we cross-verified the target sites and regions with the new assembly (GRCh38) to ensure there were no artifacts arising from updates to the reference genome.

For generation of the 740-kb microdeletion and microduplication in 16p11.2, the following single-guide RNA was used:

- sg569: 5′-GACATGCTCTATGCGCAT-3′, chromosome 16: 29,487,572–29,487,590 and 30,226,917–30,226,935

For generation of the 2-Mb microdeletion and microduplication in 15q13.3, the following single-guide RNA was used:

- sg387: 5′-GCCCTAGGGGATTGCGGAC-3′, chromosome 15: 30,792,593–30,792,611 and 32,799,503–32,799,521

All gRNAs were cloned into pSpCas9(BB)-2A-Puro plasmid with a puromycin selection site. gRNA841 was cloned into gRNA Cloning Vector (pGuide, Addgene 104124) using a BbsI restriction site. Validation of the guide sequence in the gRNA vector was confirmed by Sanger Sequencing. Before transfection, all plasmids were purified from EndoFree Plasmid Maxi Kit according to the manufacturer’s instruction (Qiagen).

Cell culture and DNA transfection. Human induced pluripotent stem cells (iPSCs), derived from adult fibroblasts, were maintained on Matrigel-coated dish (Corning) with Essential 8 medium (Invitrogen) and incubated at 37 °C in a humidified atmosphere with 5% CO2. For the dual-guide approach, we transfected human iPSCs with pSpCas9(BB)-2A-Puro plasmid and gRNA Cloning Vector, which carry both guide RNAs and collected individual iPSC colonies following the exact modification that occurred (for the 575-kb deletions) (~14 d after sorting), the genomic DNA from those colonies was characterized by copy number assay.

Single-cell isolation by fluorescence-activated cell sorting (FACS). To obtain isogenic iPSC colonies following CRISPR/Cas9 treatment, single cells were isolated by FACS. At 72 h after nucleofection, the iPSCs were dissociated into a single-cell suspension with Accutase and resuspended in PBS with 10 μM ROCK inhibitor (Santa Cruz Biotech). All samples were filtered through 5-μm polystyrene tubes with 35-μm mesh cell strainer caps (BD Falcon 352235) immediately before being sorted. After adding the viability dye TO-PRO-3 (Invitrogen), the GFP+TO-PRO-3+ iPSCs were sorted and plated, with one cell placed in each well of Matrigel-coated 96-well plates by a BD FACSaria II sorter with a 100-μm nozzle under sterile conditions. Once multicellular colonies were clearly visible (2–3 d after sorting), they were collected into individual wells of Matrigel-coated 96-well plates by manual picking. Once individual iPSC colonies were available (~14 d after sorting), the genomic DNA from those colonies was characterized by copy number assay.

Genomic DNA extraction. For high-throughput extraction in 96-well format, the genomic DNA was isolated from cells transfected by the Cas9-gRNA expression vector (pX459) using a Rapid DNA extraction method. Cells were lysed by adding DNA extraction buffer containing Proteinase K (0.2 mg/mL). Samples were digested at 55 °C for 30 min followed by Proteinase K inactivation at 95 °C for 10 min, followed by RNase A incubation at room temperature for 10 min. For genomic DNA extraction from 16p11.2 patient iPSC lines, we performed extraction by DNeasy Blood & Tissue Kit (Qiagen) following the manufacturer’s instruction. Genomic DNA for 15q13.3 patients was obtained from the Simons Foundation Autism Research Initiative.

Screening of individual iPSC colonies. To isolate genomic DNA from the iPSC colonies, iPSCs were detached with ReLeSR (Stem Cell Technologies) and then prepared by using the Rapid DNA extraction method. For detection of deletion, the genomic region flanking the CRISPR target site was amplified by PCR (Supplementary Fig. 2a). The primers were synthesized by Integrated DNA Technologies (IDT) with the following sequences:

- Forward: 5′-CACTGGAATTTCCTCCTTTCA-3′
- Reverse: 5′-CCCTCAGGTATCCCTCTCAT-3′

PCR reactions were performed using 2 μg of genomic DNA and Phusion High Fidelity Master Mix (NEB), with the following cycling conditions: 95 °C for 2 min; 95 °C for 15 s, 63 °C for 30 s, 72 °C for 1 min (45 cycles); 72 °C for 2 min. PCR products were visualized, which was followed by Sanger sequencing to determine the exact modification that occurred (for the 575-kb deletions) (Supplementary Fig. 2b). To determine the heterozygosity of 16p11.2 region in CRISPR targeting iPSC lines, we followed the same PCR protocol performed in the detection of deletion to amplify part of the 16p11.2 region that is deleted in 16p11.2 rMDS. Also, the PCR was performed using the same forward primer as in detection of deletion and a newly designed reverse primer targeting the unique rMDS segment.

Identification of the junction sequences in 16p11.2 rMDS iPSC lines. For isolation the sequences, the genomic region around the CRISPR target site was amplified by nested PCR. The primers were synthesized by IDT with the following sequences:

- Forward: 5′-TGAAGTTTCAGGAAACATTGG-3′
- Reverse-1: 5′-CGTGTTGAGGACATGCAAC-3′
- Reverse-2: 5′-TGAAGCGGCGTGGCGGAC-3′

The first PCR reaction was performed using 1 μl of genomic DNA, forward and reverse-1 primers and Phusion High Fidelity Master Mix (NEB), with the following cycling conditions: 95 °C for 2 min; 95 °C for 15 s, 61 °C for 30 s, 72 °C for 1 min (45 cycles); 72 °C for 2 min. The second PCR reaction was performed using 1 μl of the first PCR product, forward and reverse-2 primers and Phusion High Fidelity Master Mix (NEB), with the following cycling conditions: 95 °C for 2 min; 95 °C for 15 s, 61 °C for 30 s, 72 °C for 1 min (45 cycles); 72 °C for 1 min. PCR products were visualized, followed by Zero Blunt PCR cloning (ThermoFisher Scientific) and Sanger sequencing to determine the exact mutation that occurred (Supplementary Fig. 5).

Copy number analysis: quantitative real-time PCR (qRT-PCR). For initial copy number screening of CRISPR-treated iPSC lines, qRT-PCR was performed using six sets of TaqMan probes targeting six genes at 16p11.2 microdeletion region.
and a reference primer/probe set targeting the RPPH1 gene on chromosome 14 (TaqMan Copy Number Reference Assay RNAsP Applied Biosystems). Reactions were carried out in triplicate in 20 μl total volume containing 1 μl of genomic DNA sample, LightCycler 480 Probes Master (Roche) and 20x primer/probe mix. The 16p11.2 region primers and FAM-labeled probes were synthesized by IDT with the following sequences:

QRT-F: 5'-CTAACAACGGAAGAGGATGAC-3'
QRT-R: 5'-CATTTGGGACTCATCAGCC-3'
QRT-probe: 5'-CTGCAGAAGGACGGACAGC-3'
KCTD13-F: 5'-CCTCAAGGTGGTGTTAAGGAGA-3'
KCTD13-R: 5'-CAACCTGGTGTGTTATAGGATT-3'
KCTD13-probe: 5'-CACACCTGGTGCTTGCAAGACG-3'
CDPT-F: 5'-GGCCCCAGCGCAAGAATA-3'
CDPT-R: 5'-AAGCACCTCCACGCTCTC-3'
CDPT-probe: 5'-CCGAGTGAATGATAGGTCGCA-3'
BOLA2-F: 5'-ATCGAGCGGTCAGGAGG-3'
BOLA2-R: 5'-TGGAGGGCGGTGTTGAAA-3'
BOLA2-probe: 5'-TACACATGGACTCGGTCGCCG-3'
SLX1A-F: 5'-AGCGTTGCCTCAGGATAGC-3'
SLX1A-R: 5'-GGTGATTCGGCTGTAAC-3'
SLX1A-probe: 5'-CACCAGAGCTTGTTTGAGGAC-3'
SULT1A4-F: 5'-TTTCCCTTCACAAGCCTTCGC-3'
SULT1A4-R: 5'-TCTGAGTCGGAAGATTGGCTGCCC-3'

For the copy number screening in the 15q13.3 CRISPR work, qRT-PCR was performed using six sets of TaqMan probes targeting six genes (FANI, MTTM10, TRPM1, KLF13, OTUD7A and CHRNA7) in the 15q13.3 microdeletion region, two sets of TaqMan probes targeting two genes (SCG5 and TPI1) outside the 15q13.3 CRISPR targeting SDs and a reference primer/probe set targeting the RPPH1 gene. All primer/probe sets were purchased from ThermoFisher Scientific, with the following assay identifiers: FANI, Hs00942380_cn; MTTM10, Hs0916555_cn; TRPM1, Hs00942380_cn; KLF13, Hs02255264_cn; OTUD7A, Hs00989444_cn; CHRNA7, Hs03900067_cn; SCG5, Hs03900067_cn; and TPI1, Hs0393363_cn.

Amplification was performed by using the LightCycler 480 Real Time PCR system (Roche) under the following cycling conditions: preincubation step, 95 °C for 10 min; amplification step, 95 °C for 10 s, 55 °C for 30 s and 72 °C for 1 s (50 cycles); cooling step, 40 °C for 10 s. Relative copy number of genes was determined by using the 2ΔΔCₚ method46, where ΔCₚ is the difference of cycle threshold values for the test gene and reference gene RPPH1. For each gene, experiments are in triplicate. Statistical analysis for copy number was performed using one-way analysis of variance (ANOVA) followed by Dunnett’s adjustment for multiple comparisons.

Chromosomal microarray analysis. Array-based comparative genomic hybridization (aCGH) was performed on the Agilent 4 × 180K SurePrint G3 Human CGH Microarray (design number 022060) according to the protocol provided by the manufacturer. The assay tests for imbalances (that is, gains or losses) in the genomic DNA sample being tested by comparing the test DNA to a Promega male control DNA sample obtained from a pool of karyotypically normal individuals. This array platform contains 180,880 probes taken from throughout the human genome. A genomic imbalance is noted when six or more oligonucleotides show a minimum average log ratio of 0.25 for one-copy gains and ≤0.50 for one-copy losses; oligonucleotide information is based on the March 2006, NCBI 36.3 (hg18) build of the Human Genome (UCSC Genome Browser, http://genome.ucsc.edu/cgi-bin/hgGateway/). However, the data reports were generated using the genome build hg19. This assay does not exclude chromosome anomalies smaller than the assay’s effective resolution. The assay is also not specifically designed to detect mosaicism, uniparental disomy, methylation abnormalities or other chromosomal rearrangements (including chromosomal translocations, insertions and inversions).

Library preparation and RNA-sequencing (RNAseq). RNA samples were prepared from 23 iPSC lines (Supplementary Fig. 3). There were 5 iPSC lines derived from parents harboring 16p11.2 microdeletion (two families) as well as 3 lines derived from patients with 16p11.2 microduplication (one family) (Supplementary Fig. 3). Other lines were created using CRISPR/Cas9 system with or without guide RNAs. All RNA samples were extracted with Trizol reagent according to the manufacturer’s instruction (Invitrogen). RNAseq libraries were prepared from 200 ng of total RNA using a TrueSeq Stranded mRNA Sample Prep Kit (Illumina cat. no. RS-122-2102) Each library also included 1 μl of a 1:100 dilution of ERCC RNA Control Spike-In (Ambion) containing 92 synthetic RNA standards of known concentration and sequence. These synthetic RNAs cover a 10⁶ range of concentration, as well as varying in length and GC content to allow validation of dose response and the fidelity of the procedure in downstream analyses37. Libraries were multiplexed, pooled and sequenced on multiple lanes of an Illumina HiSeq2500, generating an average of 40 million paired-end 50-cycle reads for each sample.

RNAseq data analysis. RNAseq data was processed using a standard workflow which includes quality control of fastq reads using FastQC (http://www.bioinformatics. babraham.ac.uk/projects/fastqc/). Reads were aligned to the Ensembl GRCh37 human reference using GSNAP v 2014/12/19 (ref 38) and quality control of the alignments was performed with custom script wrappers for multiple utilities; namely, PicardTools, the RNAseqQC40 module from GenePattern and RSeqQC40. Read counts were quantified using the BedTools suite41, taking into account strand specificity (described in ref. 9) and ERCC spike-ins were used to threshold values within the lower limits of detection (described in ref. 9), and genes with >3 reads across all samples were chosen for subsequent analysis and normalized by the library size. We also filtered out rRNA and rRNA genes as well as genes <250 nt in length, which is the minimum length for the ERCC-spike-in transcripts. Read counts were analyzed in the R environment using generalized linear models (GLMs) assuming a negative binomial distribution, and significance was assessed based on contrasts for hypothesis testing using both nominal and FDR P-value cut-offs. For details of the model fitting procedures and the statistical analysis procedures, readers are referred to our previous study49.

iPSC whole-cell lysate preparation and western blot. iPSC lysate was prepared in RIPA buffer (50 mM Tris-HCl (pH 7.4), 150 mM NaCl, 2 mM EDTA, 1% IGEPA CA-630, 1 mM phenylmethylsulfonyl fluoride, 20 mg/ml pepstatin A, 20 mg/ml leupeptin, 20 mg/ml aprotinin, 50 mM NaF and 1 mM Na3VO4). The lysate was resolved by 10% Bis-Tris PAGE (ThermoFisher Scientific). The proteins resolved by SDS-PAGE were transferred to a nitrocellulose membrane (cat. no. 11036-41BL, Sartorius) and western blot analysis was conducted by using the following antibodies: mouse anti-MAPK3 (12D11, cat. no. MA1-13041, ThermoFisher Scientific), rabbit anti-β-actin (cat. no. ab8227, Abcam)42. The secondary antibodies used were HRP-conjugated goat-anti mouse IgG antibody (cat. no. ab97250, Abcam) and HRP-conjugated goat-anti rabbit IgG antibody (cat. no. ab6721, Abcam). The membrane was developed by reacting with chemiluminiscence HRP substrate (Millipore) and exposed to Amersham Hyperfilm ECL (GE Healthcare) for visualization of protein bands. The protein bands were quantified using NIH Image J software. Statistical analysis for western blot was performed using one-way analysis of variance (ANOVA) followed by Dunnett’s adjustment for multiple comparisons. Full-length blots are shown in Supplementary Figure 8.

A Supplementary Methods Checklist is available.

31. Ran, F.A. et al. Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. 8, 2281–2308 (2013).
32. Margais, G. & Kingsford, C. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27, 764–770 (2011).
33. Plisatska, V. & Rigoutsos, I. “OffSpotter”: very fast and exhaustive enumeration of genomic lookalikes for designing CRISPR/Cas guide RNAs. Bioinformatics 30, 2184–2185 (2012).
34. Maiti, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823–826 (2013).
35. McClive, P.J. & Sinclair, A.H. Rapid DNA extraction and PCR-sexing of mouse embryos. Mol. Reprod. Dev. 80, 225–236 (2011).
36. Yaffe, M.B. et al. A motif-based profile scanning approach for genome-wide prediction of signaling pathways. Nat. Biotechnol. 19, 348–353 (2001).
37. Jiang, L. et al. Synthetic spike-in standards for RNA-seq experiments. Genome Res. 21, 1543–1551 (2011).
38. Wu, T.D. & Nieuw, S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics 26, 873–881 (2010).
39. DeLuca, D.S. et al. RNA-SeqC: RNA-seq metrics for quality control and process optimization. Bioinformatics 28, 1530–1532 (2012).
40. Wang, L., Wang, S. & Li, W. RSeQC: quality control of RNA-seq experiments. Bioinformatics 28, 2184–2185 (2012).
41. Quinlan, A.R. & Hall, I.M. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842 (2010).
42. Verbovšek, U. et al. Expression analysis of all protease genes reveals cathepsin K to be overexpressed in glioblastoma. PLoS One 9, e111819 (2014).