File S1

Supplementary Online Methods

Nematode culture

Strains were cultured according to standard methods (BRENNER 1974). The *C. elegans* strains N2 (Bristol), CB120 *unc-4(e120)*, and PS252 *dpy-11(e224)* were used.

DNA cloning

The *hCas9* nuclease gene was amplified from plasmid hCas9 (MALI *et al.* 2013) (ID# 41815; Addgene, Cambridge, MA), modified to include 5' and 3' untranslated regions previously reported to function well in the *C. elegans* germline (WOOD *et al.* 2011), bounded on the 5' side with an SP6 phage RNA polymerase binding site and on the 3' side with a *Kpn*I restriction site. The resulting construct was cloned into a Bluescript vector (Agilent Technologies, Santa Clara, CA) to generate the plasmid SP6-hCas9-Ce-mRNA; this was cleaved with *Kpn*I and then used for the *in vitro* synthesis of *hCas9* mRNA using mMessage mMachine SP6 and Poly(A) tailing kits (Life Technologies, Grand Island, NY).

A scaffold for the cloning of targeting sequences to generate sgRNAs was constructed using as a template the plasmid "gRNA Empty Vector" (MALI *et al.* 2013) (ID# 41824; Addgene), The downstream gRNA fusion part of this vector was retained, along with the *Afl*II site for recombinational cloning, bounded on its 5' end with an SP6 phage RNA polymerase sequence on its 3' end with a *Kpn*I restriction site (replacing the PolIII terminator sequence found in the template plasmid). The resulting construct was cloned into Bluescript to generate the plasmid SP6-sgRNA-scaffold. Sequences intended for use in targeting CRISPR-Cas-mediated cleavage were amplified by PCR with flanking sequences suitable for recombination into the *Afl*II site of this plasmid using Gibson cloning (New England Biolabs, Beverly, MA). Clones were confirmed by sequencing, cleaved with *Kpn*I, and used for the *in vitro* synthesis of sgRNAs using MEGAscript SP6 (Life Technologies).

Sequences were selected for use in targeting for cleavage by CRISPR-Cas on the basis of their position within the gene (preferably in an early exon) and the absence of strong BLAST hits elsewhere in the target genome for the 3' half of the 20mer sequence. Sequences used were: *dpy-11*: GAGCTGGGCACCATGGAGCA; *unc-4*: GATATCGTCATCCGGTGACG.

CRISPR-Cas mutagenesis and mutant analysis

The germline syncytia of P_0 animals were injected with the synthetic hCas9 mRNA at a final concentration of 200 – 500 ng/µl together with the sgRNA of choice at a final concentration of 40 – 100 ng/µl, in water, using established microinjection techniques (MELLO and FIRE 1995). One to three injected animals were placed on a 6 cm or 10 cm Petri dish containing NGM agar and OP50 bacteria. Animals were transferred to new plates as food supplies became exhausted. F_2 progeny of injected P_0s were screened for the presence of phenotypic animals. Mutant strains were analyzed using PCR and DNA sequencing (Laragen, Culver City, CA) to determine the number

and nature of independent isolates, especially in the case of isolates arising from the same injected P_0 or the same pool of injected P_0s. In the case of strains where no amplicon was recovered using primers that amplified from the wild type a 400 – 600 bp product centered on the targeted cleavage site, additional primer pairs were used to test for the presence of sequences at intervals of 2 – 3 kb from the targeted cleavage site, in each direction. PCRs were also performed that in the wild type would amplify larger (5 or 10 kb) genomic sequences centered on the targeted cleavage site, using Expand Long polymerase mix (New England Biolabs). Selected candidates were tested for the presence of a reciprocal translocation by outcrossing and examining the self-progeny of animals heterozygous for the mutation for the presence of dead embryos that would result from aneuploid zygotes. In particular, *sy745* mutants contain all tested sequences near to the target site, but attempts to amplify across the target site using PCR were not successful; the self-progeny of *sy745*/+ heterozygotes did not include dead (aneuploid) embryos, indicating *sy745* is likely to be an inversion or a large insertion, rather than a reciprocal translocation. *sy750* mutants fail to complement *dpy-11(e224)*, but no mutation was found near the site targeted for cleavage by CRISPR-Cas. There may be an as-yet undiscovered mutation away from the cleavage site; homology-directed repair of double-strand breaks involves error-prone DNA synthesis, causing mutations at a significant remove from the site of the break (Strathern *et al.* 1995; Deem *et al.* 2011); however, no coding change was found in *dpy-11* in *sy750* animals. Alternatively, the *dpy-11(sy750)* mutant may contain a deletion-duplication, including a wild-type copy of the locus surrounding the targeted cleavage site in the context of genomic rearrangements that disrupt the function of the *dpy-11* gene.

High-throughput sequencing and analysis

Genomic DNA was isolated from *dpy-11(sy740)* and *dpy-11(sy745)* mutant strains by thorough digestion using Proteinase K (Life Technologies) in the presence of β-mercaptoethanol, one round of phenol-chloroform extraction, chloroform extraction, and spooling from ethanol, followed by RNase treatment (Qiagen, Valencia, CA) and then a second round of extractions and spooling. Genomic DNA libraries were built using Illumina's standard paired-end protocol (Bentley *et al.* 2008), and 50mer unpaired reads were obtained from each library using an Illumina HiSeq (Illumina, San Diego, CA).

The sequence data were aligned to the *C. elegans* genome (WormBase release 235; ftp://ftp.wormbase.org/pub/wormbase/species/c_elegans) using BWA (Li and Durbin 2009). SNPs and small insertions and deletions (indels) were identified using the GATK pipeline (DePristo *et al.* 2011) following standard practices for variant detection. Larger deletions and insertions were identified using a custom pipeline implemented in Perl based on the split-read approach to define candidate indel locations followed by a refinement step based on the Smith-Waterman local alignment algorithm (Smith and Waterman 1981) to determine precise indel structures.

Mismatches and small deletions and insertions predicted by GATK analysis to be unique to either the *dpy-11(sy740)* or the *dpy-11(sy745)* mutant strain and deletions and insertions predicted by split-read analysis to be present in either or both strains were manually curated by examining an alignment of the sequencing reads to the *C. elegans* genome generated using the Burrows-Wheeler

aligner (LI and DURBIN 2009) in Integrative Genomic Viewer (IGV) (ROBINSON *et al.* 2011). Examination of the reads aligned in this fashion could demonstrate the presence of wild-type sequence lacking the deletion; a predicted deletion could still be present in a subset of the sequenced DNA. The algorithm used to align reads for manual curation would not be able to align reads that indicated the presence of a large deletion, and so if such a deletion were present but not homozygous, inspection of the aligned reads would detect wild-type sequence at the locus and would not display reads substantiating the presence of the deletion. For this reason, we used PCR to test a subset of the predicted deletions for which inspection of the aligned reads demonstrated the presence of wild-type sequence. PCRs were performed using primers spanning selected predicted deletions shown by examination of the aligned reads not to be homozygous, to test for the possible presence of smaller bands from template carrying the deletion. The oligonucleotide sequences used in these PCRs are shown in Table S4. These PCR assays did not provide evidence for off-target effects of CRISPR-Cas mutagenesis (Table S2); the unconfirmed deletions may represent computational or sequencing artifacts.

Estimates of the false-negative rate for mutation detection for the GATK and split-read analyses were made by repeating the analyses using as reference genomes versions of the *C. elegans* genome with sequences inserted at known positions, such that a comparison should show deletions compared to the reference sequence. The results of these analyses are presented in Table S3.

Literature Cited

BENTLEY, D. R., S. BALASUBRAMANIAN, H. P. SWERDLOW, G. P. SMITH, J. MILTON *et al.*, 2008 Accurate whole human genome sequencing using reversible terminator chemistry. Nature **456:** 53-59.

BRENNER, S., 1974 The genetics of *Caenorhabditis elegans*. Genetics **77:** 71-94.

DEEM, A., A. KESZTHELYI, T. BLACKGROVE, A. VAYL, B. COFFEY *et al.*, 2011 Break-induced replication is highly inaccurate. PLoS Biol **9:** e1000594.

DEPRISTO, M. A., E. BANKS, R. POPLIN, K. V. GARIMELLA, J. R. MAGUIRE *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet **43:** 491-498.

LI, H., and R. DURBIN, 2009 Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics **25:** 1754-1760.

MALI, P., L. YANG, K. M. ESVELT, J. AACH, M. GUELL *et al.*, 2013 RNA-guided human genome engineering via Cas9. Science **339:** 823-826.

MELLO, C., and A. FIRE, 1995 DNA transformation. Methods Cell Biol **48:** 451-482.

ROBINSON, J. T., H. THORVALDSDOTTIR, W. WINCKLER, M. GUTTMAN, E. S. LANDER *et al.*, 2011 Integrative genomics viewer. Nat Biotechnol **29:** 24-26.

SMITH, T. F., and M. S. WATERMAN, 1981 Identification of common molecular subsequences. J Mol Biol **147:** 195-197.

STRATHERN, J. N., B. K. SHAFER and C. B. MCGILL, 1995 DNA synthesis errors associated with double-strand-break repair. Genetics **140:** 965-972.

WOOD, A. J., T. W. LO, B. ZEITLER, C. S. PICKLE, E. J. RALSTON *et al.*, 2011 Targeted genome editing across species using ZFNs and TALENs. Science **333:** 307.

Table S1 Detailed results of GATK mutation detection

Position	Found in:	Mutant reads / total reads	# reads of other strain	wild-type sequence	mutant sequence
II: 4101403	*sy740*	4/18	10	CACCT**C$_3$AC$_7$**TC	CACCT**C$_{11}$**TC
III: 8457740	*sy740*	13/23	10	TAGGGGAA**G**TGTATTTG	TAGGGGAA**C**TGTATTTG
IV: 13503822	*sy740*	6/13	3	CCCCCA**A**TTGGA**C**ATCCCC	CCCCCA**G**TTGGA**T**ATCCCC
IV: 13503828		6/15	5		
IV: 8719829	*sy740*	8/14	6	ACAGT$_5$**G$_{13}$AG$_5$**TCTAAC	ACAGT$_5$**G$_{19}$**TCTAAC
X: 17375284	*sy740*	8/36	38	GATTGCGT**G**AAGCAAAG	GATTGCGT**A**AAGCAAAG
V: 13647424-13647432	*sy745*	6/22	15	ATCCT**(TCG)$_9$**TC(TCG)$_5$CG	ATCCT**(TCG)$_6$**TC(TCG)$_5$CG
I: 3075678	*sy745*	8/28	21	GTTTTAATT**A$_{13}$**CTGA$_7$GT	GTTTTAATT**A$_{14}$**CTGA$_7$GT
X: 14728375	*sy745*	5/17	12	CGTTAGA**G$_{14}$AG$_3$**TGAAGA	CGTTAGA**G$_{18}$**TGAAGA

Comparison of the high-throughput sequencing output generated using the GATK pipeline identified 1419 predicted changes between *dpy-11(sy740)* and the *C. elegans* reference genome, and 1441 predicted changes between *dpy-11(sy745)* and the *C. elegans* reference genome. Of these predicted changes, 151 were unique to *dpy-11(sy740)* and 173 to *dpy-11(sy745)*, totaling 324 candidates to be strain-specific changes. Mutations predicted to be unique to either strain were manually curated by inspection of the reads aligned to the reference genome: of 324 predicted mutations, 313 were present in both strains and 3 were observed in neither strain. The remaining 8 are detailed above: the position of each is given, the representation of the mutation among reads from the strain bearing the mutation is given, the number of reads at that site for the other strain is given, and the nature of the mutation is shown, with the affected nucleotide(s) bolded and underlined. Note that none of these sites shows homology to the targeting sequence used in the sgRNA to direct Cas9 nuclease activity, GAGCTGGGCACCATGGAGCA.

Table S2 Detailed results of split-read mutation detection

Linkage Group	Start	Stop	Size	Score	Predicted in:	Homozygous in:
I	1580676	1580681	5	10	Both	Both
I	6151523	6151526	3	10	Both	Both
I	10948442	10948685	243	2	Both	Both
II	563219	563256	37	2	*sy745*	Both
II	4117187	4117193	6	2	*sy740*	Both
II	4611574	4611595	21	4	*sy745*	Both
IV	3036896	3036902	6	2	Both	Both
IV	8121709	8121716	7	5	*sy740*	Both
IV	8578126	8580609	2483	2	Both	Both
V	931548	931552	4	2	Both	Both
V	1645712	1647498	1786	9	Both	Both
V	5625703	5625715	12	12	Both	Both
V	5755542	5756050	508	3	Both	Both
V	6081823	6094837	13014	7	Both	Both
V	7725877	7725881	4	2	Both	Both
V	9026726	9026729	3	6	Both	Both
V	9063330	9063335	5	13	Both	Both
V	15434910	15434919	9	10	Both	Both
V	19820304	19820370	66	2	Both	Both
X	2002626	2002632	6	4	Both	Both
X	4938588	4938592	4	8	Both	Both
X	8941405	8941409	4	12	Both	Both
X	8941405	8941409	4	8	Both	Both
X	14432312	14432326	14	5	Both	Both
I	230840	231919	1079	7	Both	Neither
I	232747	237780	5033	2	Both	Neither
I	238430	238468	38	4	Both	Neither
I	246119	246175	56	3	*sy745*	Neither*
I	3812704	4548148	735444	13	Both	Neither
I	13156287	13156436	149	4	Both	Neither
I	14169356	14386795	217439	2	Both	Neither
II	2220320	2221389	1069	2	Both	Neither
II	3775873	7422381	3646508	2	*sy745*	Neither
II	6187749	6187758	9	2	*sy740*	Neither*
II	12009805	12009852	47	2	*sy740*	Neither*
II	12572308	12573728	1420	4	Both	Neither
III	13032636	13032832	196	4	*sy740*	Neither
IV	7727245	7727296	51	3	Both	Neither
IV	11071120	11072356	1236	11	Both	Neither
IV	14320741	14356442	35701	2	*sy745*	Neither
IV	15438235	16899372	1461137	4	*sy745*	Neither
V	3707494	3707683	189	2	Both	Neither
V	13646108	13646149	41	8	Both	Neither*
V	17344382	17344476	94	2	*sy745*	Neither*
X	1614748	1615141	393	2	*sy740*	Neither*
X	1614997	1615141	144	4	Both	Neither*
X	7077853	7077873	20	3	*sy740*	Neither
X	16014052	16014197	145	2	*sy740*	Neither*

The 48 candidate deletions predicted by split-read analysis in the *dpy-11(sy740)* and/or *dpy-11(sy745)* strains on the basis of two or more reads ("Score" in the table) are listed by linkage group, start site, end site, and size. Each was manually curated by examination of reads aligned to the reference sequence using the Burrows-Wheeler aligner. 24 of the 48 candidate deletions were homozygous in both strains; the other 24 had reads consistent with the presence of wild-type sequence at these coordinates in both strains, and so are marked as

being homozygous in neither strain. Predicted deletions in this latter class could have been represented by sequencing reads if they were present as heterozygotes, but such reads would not have been mapped to the corresponding site in the genome by the Burrows-Wheeler aligner. The candidate deletions marked with an asterisk (*) were tested using PCR to seek smaller products as predicted (see Table S4); of these, only one was present, the predicted deletion starting at 13646108 on LGV; that deletion was predicted in both strains and was detectable by PCR in both strains, indicating that it did not result from CRISPR-Cas nuclease activity.

Table S3 Estimation of false-negative rates for mutation-detection algorithms

A. False-negative rate of GATK mutation detection

Deletion size (bp)	Overall (n)	False-negative frequency: Within repetitive regions (n)	Outside repetitive regions (n)
1	32% (1000)	41% (311)	28% (689)
2	32% (1000)	37% (325)	30% (675)
3	28% (1000)	36% (291)	25% (709)
5	34% (1000)	42% (328)	30% (672)
10	35% (500)	45% (147)	31% (353)
20	100% (500)	100% (148)	100% (352)
50	100% (250)	100% (72)	100% (178)

B. False-negative rate of split-read mutation detection

Deletion size (bp)	Overall (n)	False-negative frequency: Within repetitive regions (n)	Outside repetitive regions (n)
1	36% (1000)	65% (311)	23% (689)
2	35% (1000)	68% (325)	20% (675)
3	34% (1000)	66% (291)	21% (709)
5	35% (1000)	62% (328)	22% (672)
10	31% (500)	63% (147)	18% (353)
20	35% (500)	68% (148)	22% (352)
50	38% (250)	69% (72)	25% (178)

The whole-genome sequencing output from the *dpy-11(sy740)* and *dpy-11(sy745)* strains were tested for the detection of deletions against versions of the *C. elegans* reference genome sequence into which small insertions had been made, of known position and sequence, using the same mutation-detection methods used to seek off-target effects of CRISPR-Cas-mediated mutagenesis. The frequencies at which each method failed to detect these insertions as being apparent homozygous deletions in the genome of the sequenced strain is shown for each analysis method. In each case, the results are further broken down between insertion sites within regions noted using RepeatMasker (www.RepeatMasker.org) as being highly repetitive, and insertion sites not determined to be within highly repetitive regions.

Table S4 Oligonucleotides used to test deletions predicted by split-read mutation detection. The linkage groups (LGs), positions, and extents of deletions predicted by split-read analysis and known from examination of Illumina sequence not to be present as homozygotes, along with the sequences of oligonucleotides used in attempts to detect these deletions by PCR.

LG	Start	Stop	Size	Forward Primer	Reverse Primer
I	246119	246175	56	TTTTCAAAAGTTACAGATGTTTTCG	TCCAGACAGTGCCGAATATG
II	6187749	6187758	9	GTCGTCTCGTCCCGATCC	CAAAACTCTGTGCAATGGATG
II	12009805	12009852	47	TAACGCGAATATGGCCTACG	GTGGCCTGGGAAGAGTTAGG
V	13646108	13646149	41	GCGCCCGCGTATATAAATT	AAAAAAGTTCTCCGCTGCAA
V	17344382	17344476	94	TGCCCGAAAGTACGAGTTTT	GTGTCGCGTCTTTGTCTCAA
X	1614748	1615141	393	TCGGTTCATACCGATCACAA	AGAACGGCCAAATTCTTCCT
X	16014052	16014197	145	GCTGTCAAGTCCGGTAGAGC	AAAGTCGCCAAACACCAAAG