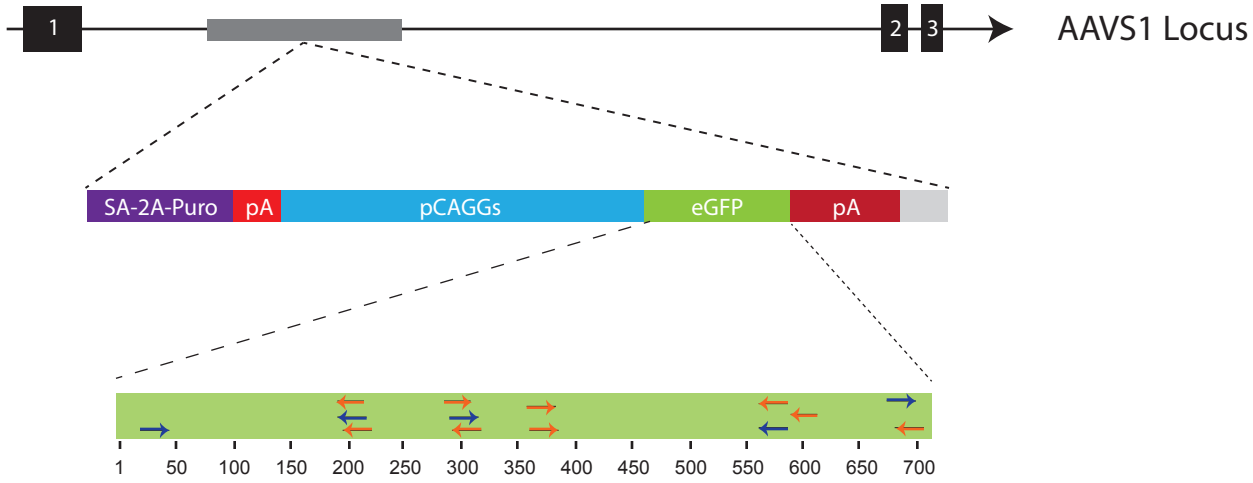
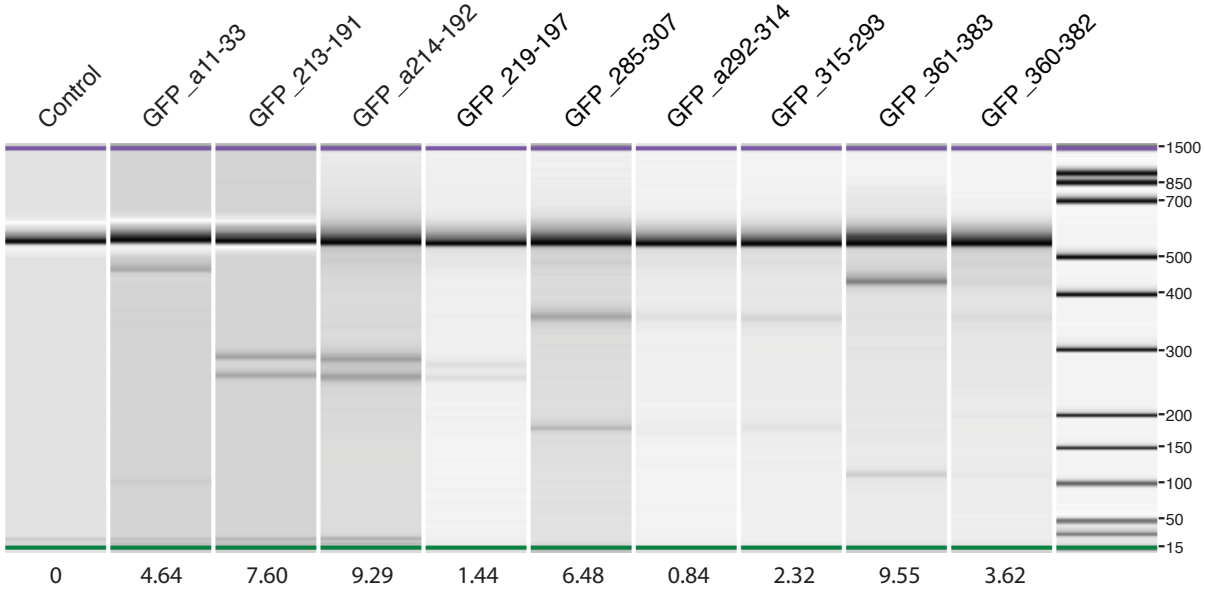
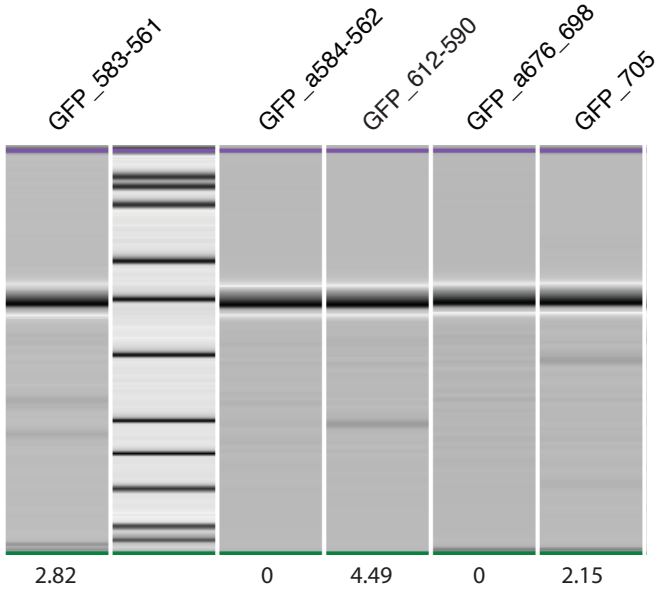
Supplementary Figure 1





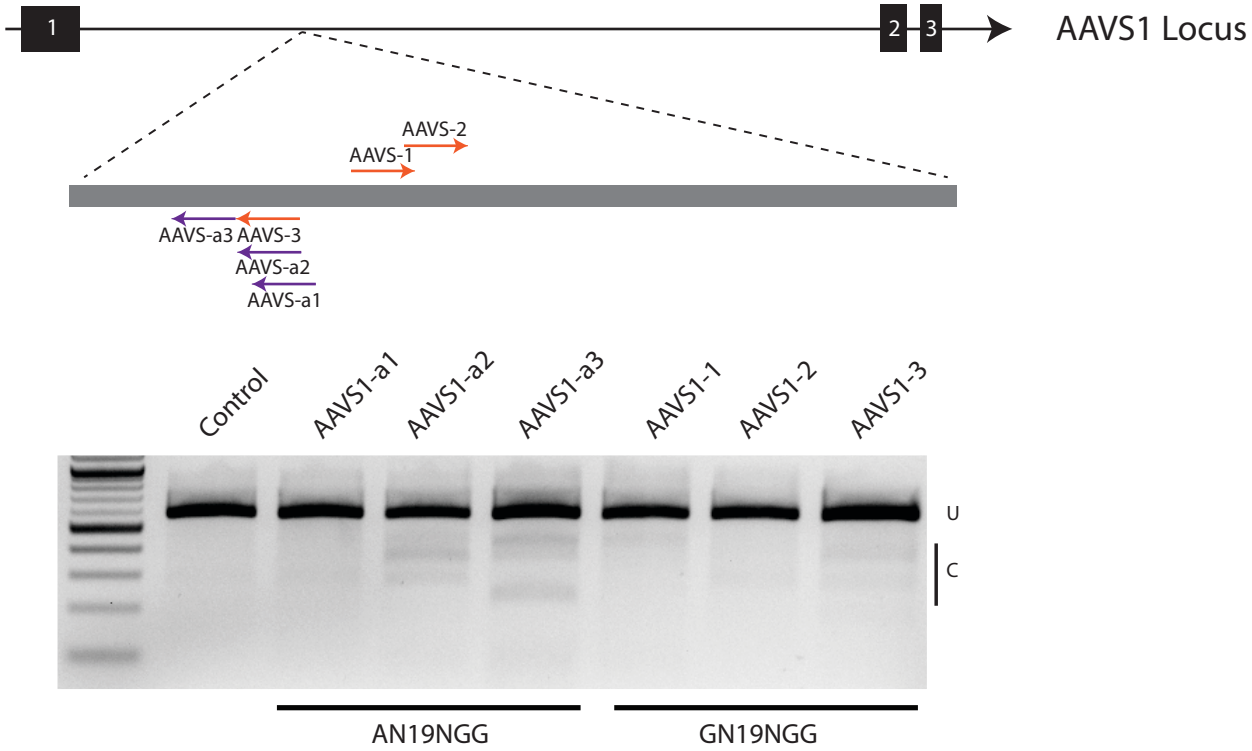


Supplementary Figure Legends

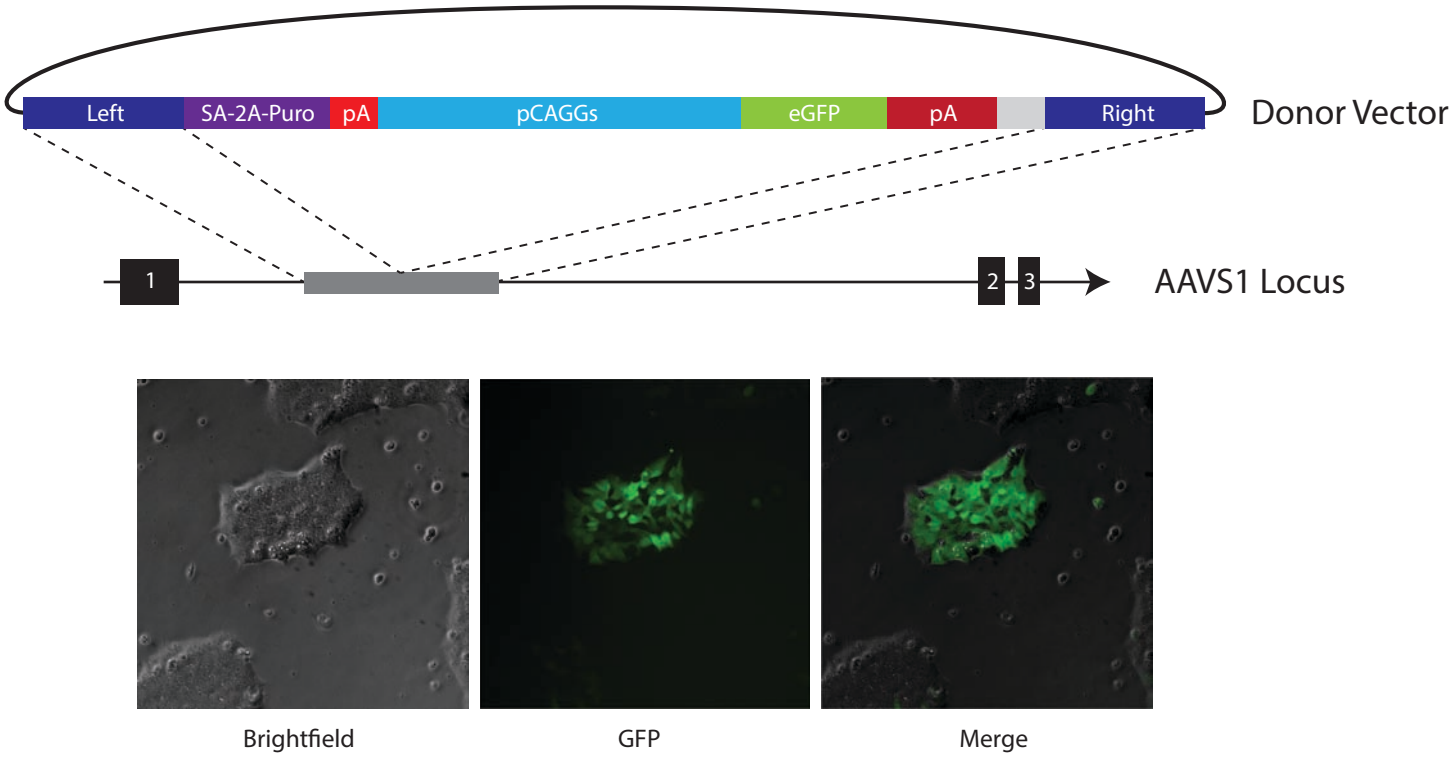
Supplementary Figure 1. Surveyor analysis and quantification of NHEJ in HEK-293 cells. Shown above is an eGFP schematic with arrows indicating the targeting sites. Target sites on the plus strand are indicated pointing to the right, and minus strand targets are indicated pointing to the left; blue arrows indicate H1 promoter gRNAs and orange arrows indicate U6 promoter gRNAs. Shown below is the Bioanalyzer gel from the Surveyor assay. The target site coordinates are listed above, and the calculated % indel is indicated below.

Supplementary Figure 2

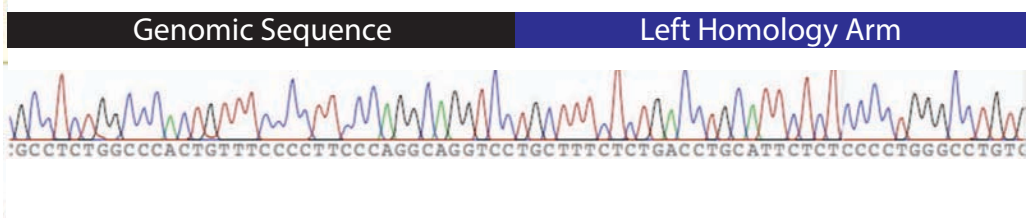
a



b

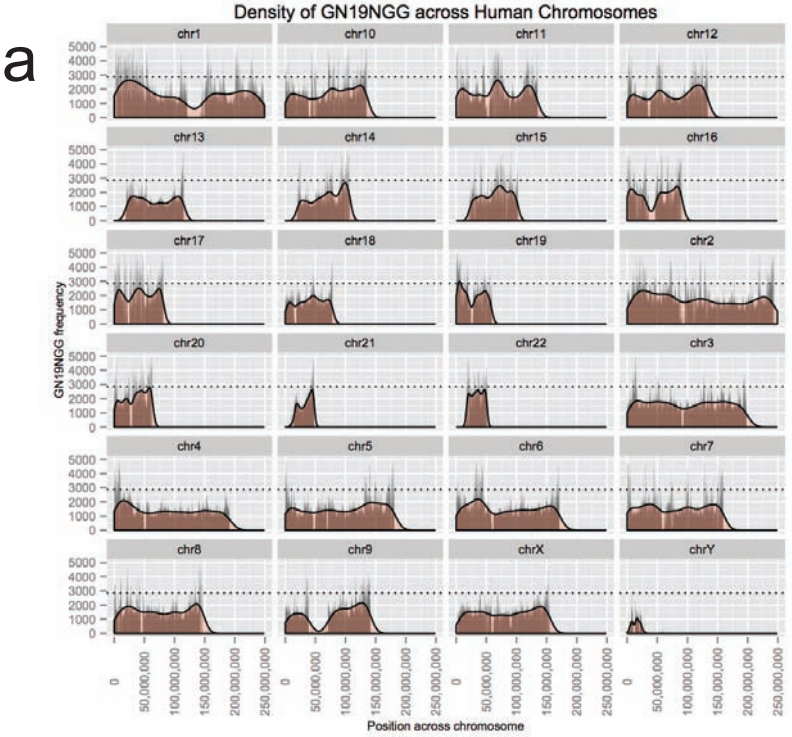


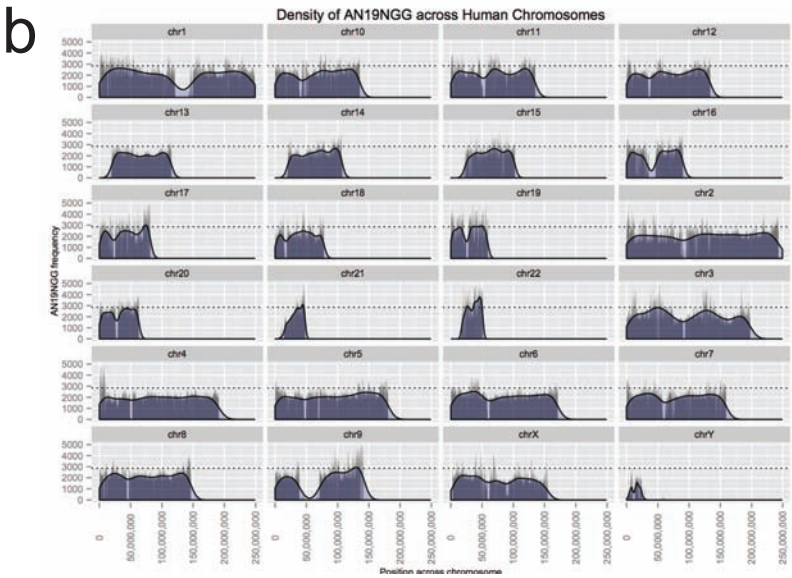
c

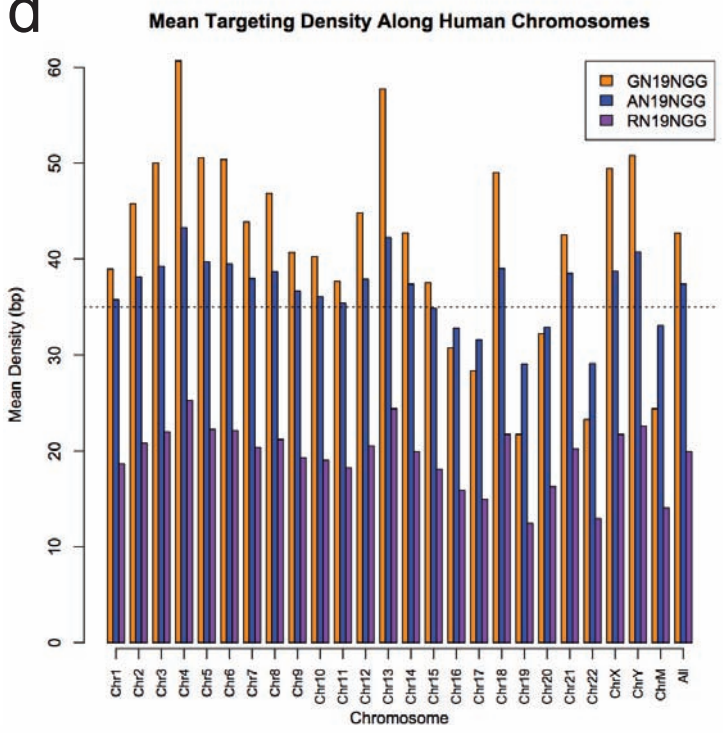


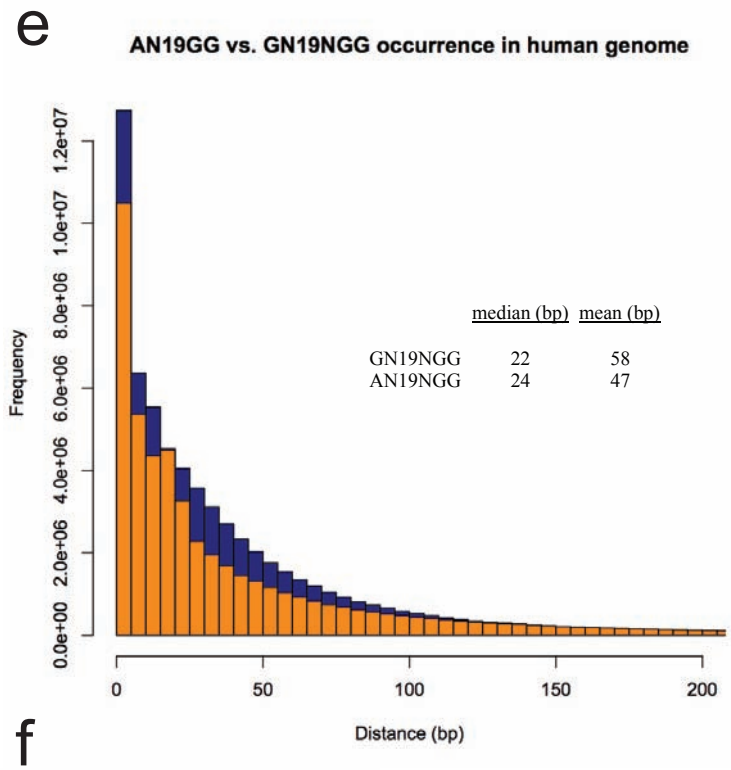
Supplementary Figure 2. Targeting and homologous recombination at the AAVS1 locus. **(a)** Surveyor analysis of three gRNAs expressed by the H1 promoter (AAVS1-1a through -1-3a) and three gRNAs expressed by the U6 promoter (AAVS-1-1 through -1-3, and a control nontargeting gRNA. **(b)** Schematic of AAVS-1 targeting donor vector shown above, and cell imaging of an GFP-positive H7 ES cell colony following electroporation with H1::AAVS1-3a gRNA and the AAVS-1 targeting vector. **(c)** Sanger sequencing of the targeting junction region indicating correct integration by homologous recombination.

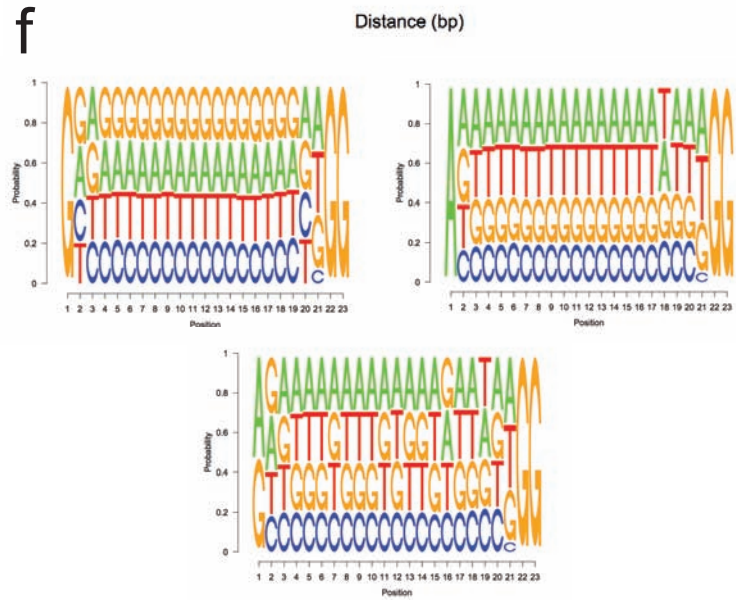
Supplementary Figure 3







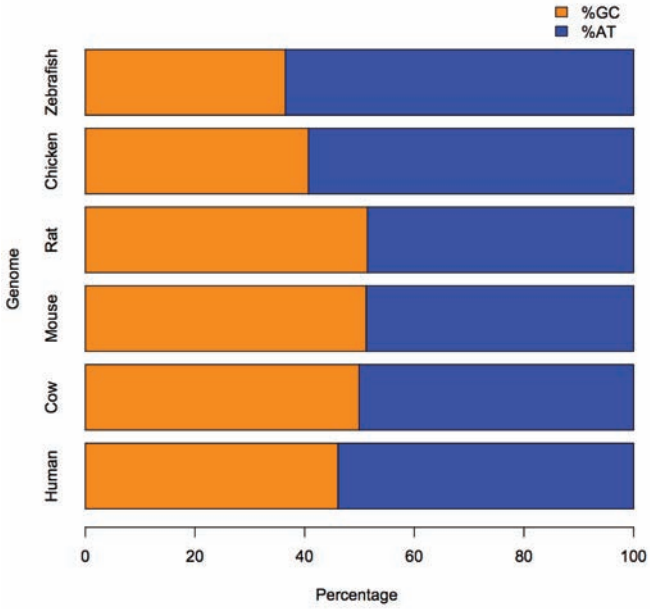




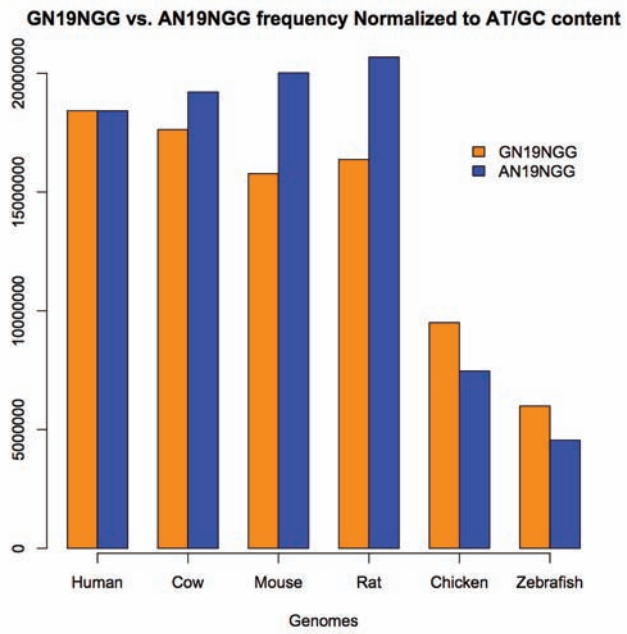
Supplementary Figure 3. Bioinformatic analysis of $GN_{19}NGG$ and $AN_{19}NGG$ sites in the genome. (**a-c**) Three panels depicting the density of each gRNA sites in the human genome: $GN_{19}NGG$ **(a)**, $AN_{19}NGG$ **(b)**, and $RN_{19}NGG$ **(c)**. Within each plot, the density of CRISPR sites is plotted along each chromosome. Overlaid in semi-transparent (orange, blue, or purple) is the density curve calculated as a smooth Gaussian kernel. The dotted line indicates 35 bp; as a reference, on average, TALEN targeting sites are estimated to occur every 35 base pairs and ZFN sites occur every couple hundred base pairs[3,49]. (d) Barplot of the cumulative mean CRISPR targeting density per human chromosome. $GN_{19}NGG$ (orange), $AN_{19}NGG$ (blue), and $RN_{19}NGG$ (purple) indicate the respective CRISPR sites. The dotted line indicates the 35 bp reference. **(e)** Frequency and distance between adjacent CRISPR sites in the genome. Barplot of the frequency and distance of adjacent $GN_{19}NGG$ (orange) and $AN_{19}NGG$ (blue) sites is in the genome. The mean and median values are inset within the plot. **(f)** SeqLogo of all $GN_{19}NGG$ (top left), $AN_{19}NGG$ (top right), and $RN_{19}NGG$ (bottom) sites in the human genome.

Supplementary Figure 4

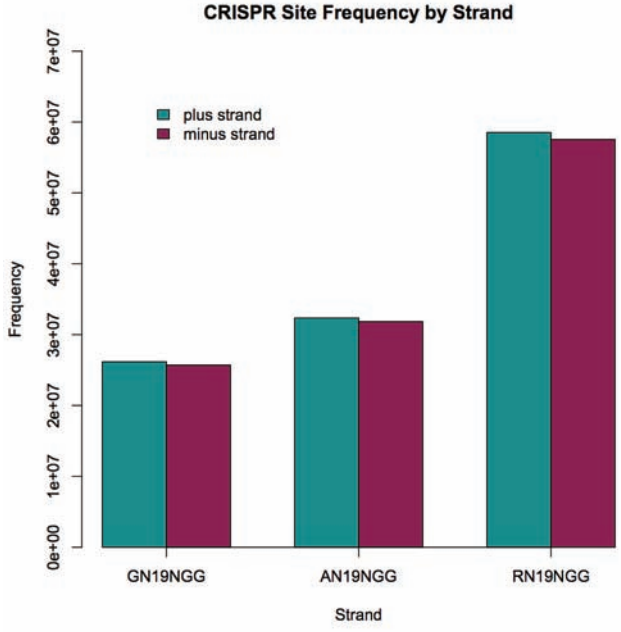
a



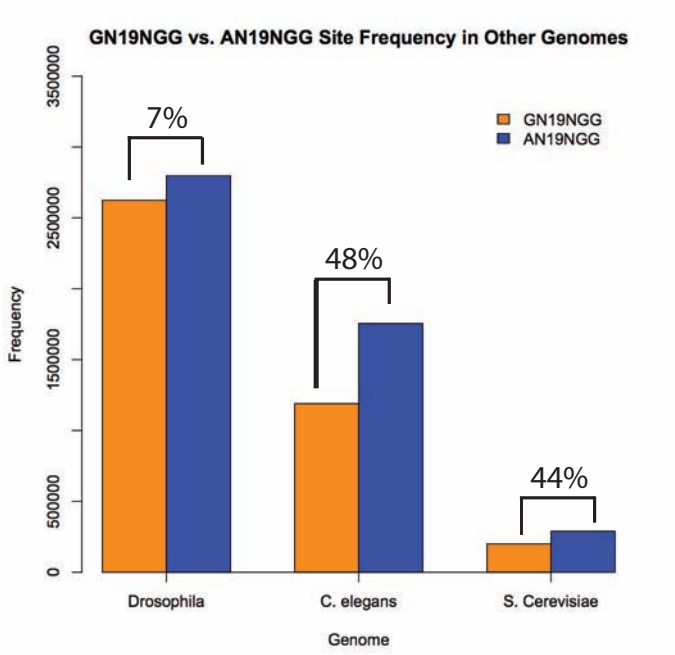
b



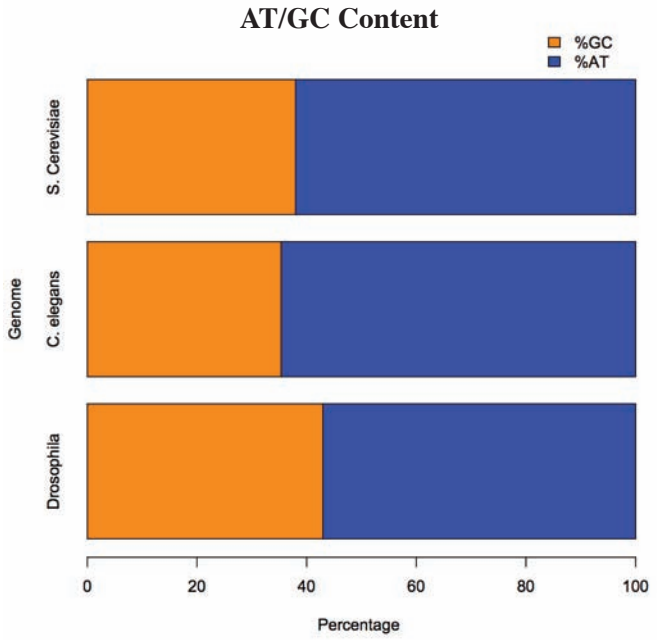
c



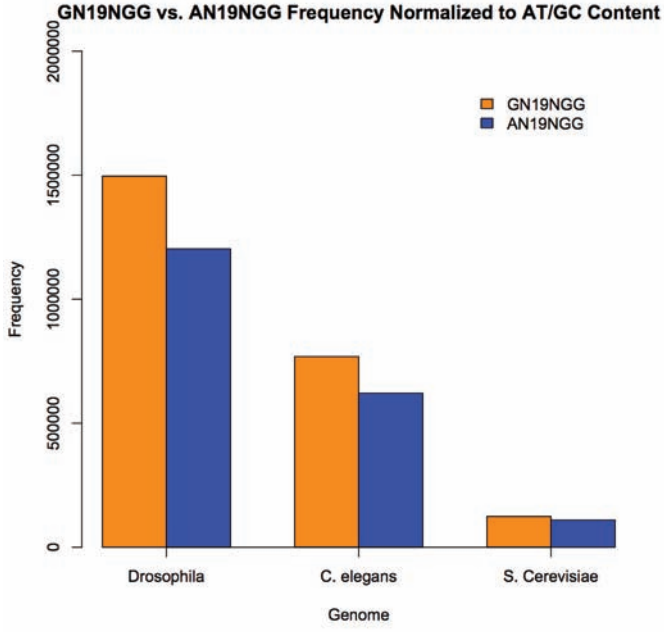
d



e



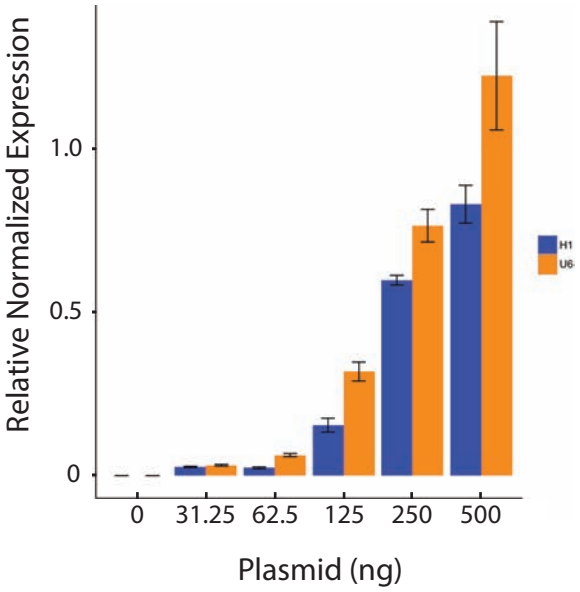
f



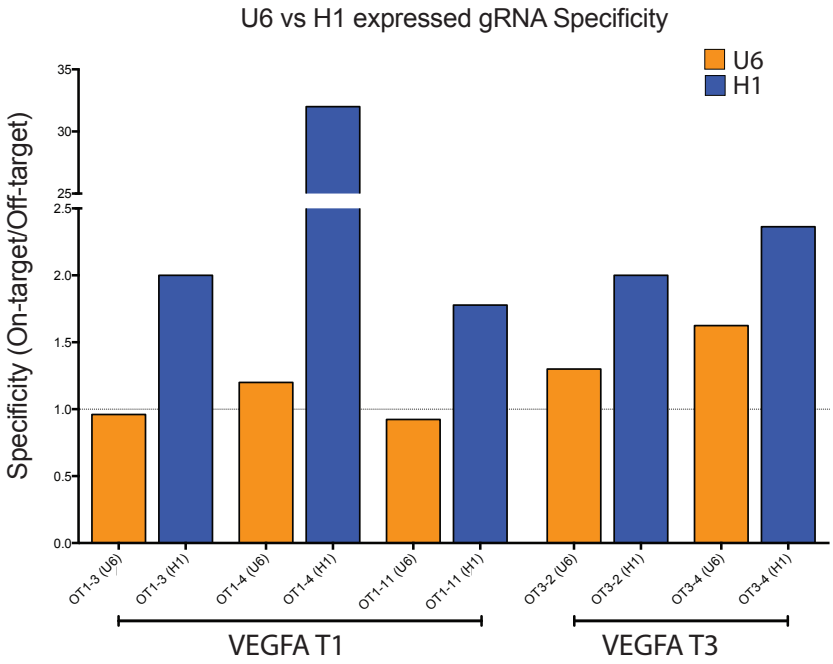
Supplementary Figure 4. AT/GC genome content and CRISPR site frequency. **(a)** The percent AT (blue) or GC (orange) is indicated for human, cow, mouse, rat, chicken, and zebrafish genomes. **(b)** The frequency of GN19NGG (orange) and AN19NGG (blue) sites normalized to AT/GC content are indicated. **(c)** CRISPR site frequency by strand for $GN_{19}NGG$ (left), $AN_{19}NGG$ (middle), and $RN_{19}NGG$ (right) sites. The plus strand (left column) is indicated by blue-green, and minus strand (right column) in purple-red. **(d)** GN19NGG (orange) and AN19NGG (blue) site frequency in Drosophila, C. elegans, and S. cerevisiae are indicated. **(e)** The percent AT (blue) or GC (orange) content and, **(f)** the normalized frequency of CRISPR sites.

Supplementary Figure 5

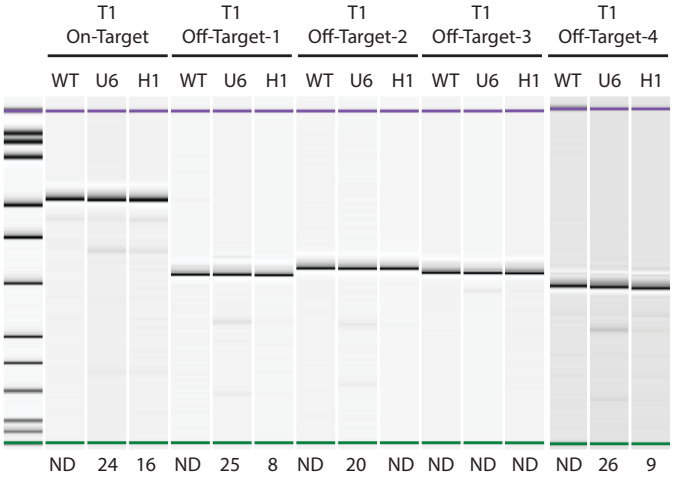
a



d



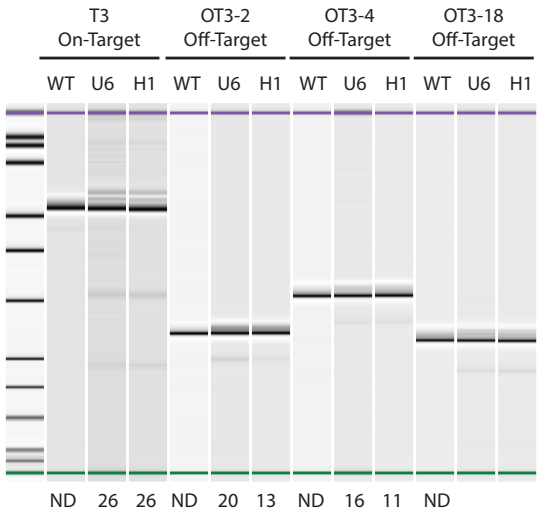
b



Target 1: VEGFA

T1	GGGTGGGGGGAGTTTGCTCCTGG
OT1-3	GG**A**TGG**A**GGGAGTTTGCTCCTGG
OT1-4	GGG**A**GGG**T**GGAGTTTGCTCCTGG
OT1-6	**C**GG**G**GG**A**GGGAGTTTGCTCCTGG
OT1-11	GGG**GA**GGGG**A**AGTTTGCTCCTGG

c



Target 3: VEGFA

T3	GGTGAGTGAGTGTGTGCGTGTGG
OT3-1	GGTGAGTGAGTGTGTG**T**GTGAGG
OT3-2	**A**GTGAGTGAGTGTGTG**T**GTGGGG
OT3-4	G**C**TGAGTGAGTGT**A**TGCGTGTGG
OT3-18	**T**GTG**G**GTGAGTGTGTGCGTGAGG

Supplementary Figure 5. Analysis off-target hits induced at on-target and off-target sites by U6 or H1 expressed gRNAs. **(a)** qRT-PCR analysis of the VEGFA T1 gRNA expression levels from titrating amounts of either the H1 promoter (blue) or U6 promoter (orange). **(b)** On-target and off-target analysis of the VEGFA T1. Surveyor analysis indicated on the left and the target sequences on the right with mismatches indicated in red. **(c)** Same as **(b)** with the VEGFA T3 target. **(d)** On-target to off-target specificity of VEGFA T1. The ratio of the on-target mutagenesis/off-target mutagenesis between the H1 promoter (blue) or U6 promoter (orange). Values below the dotted line at 1.0 indicates greater off-target mutagenesis than on-target mutatgenesis. For all parts, the on-target and off-target sites are labeled as in Fu et al. (2013) and Cho et al. (2014)[27,29].

Supplementary Figure 6

a

Endogenous U6

U6::GN19NGG

U6::AN19NGG

Downstream initiation/truncated gRNA

U6::CN19NGG

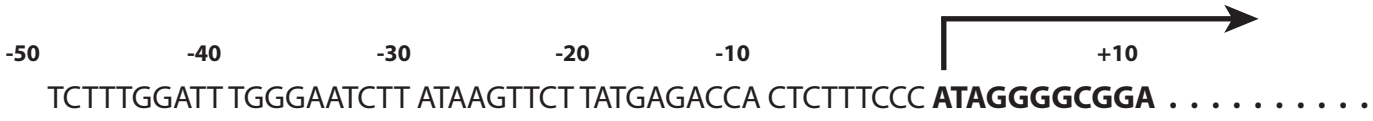
Downstream initiation/truncated gRNA

U6::TN19NGG

Downstream initiation/truncated gRNA

b

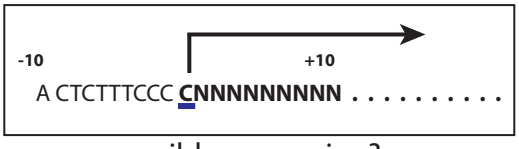
Endogenous H1



H1::GN19NGG

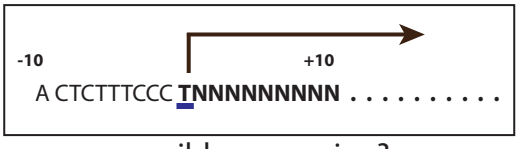
H1::AN19NGG

H1::CN19NGG



possible expression?

H1::TN19NGG



possible expression?

Supplementary Figure 6. Properties of U6 versus H1 promoters in expressing gRNAs for CRISPR targeting. **(a)** Top diagram shows the endogenous human U6 promoter and transcriptional start site. Bottom diagram indicates the use of the U6 promoter to drive gRNAs with different +1 nucleotides. Because U6 requires a G to initiate (top left), the panels that start with A (top right), C (bottom left), or T (bottom right) will likely initiate the first downstream G leading to a truncated gRNA. **(b)** Top diagram shows the endogenous human H1 promoter and transcriptional start site. Bottom diagram indicates the use of the H1 promoter to drive gRNAs with different +1 nucleotides. H1 can initiate with a G (top left) or an A (top right) leading to full-length gRNAs. Also, H1 has been reported to allow for transcription initiating at C and T nucleotides, which would allow for full-length transcripts for any +1 nucleotide downstream of the H1 promoter.

Supplementary Table 1

Cas9	Target Site	Frequency (unmasked)	Frequency (masked)
S. pyogenes			
	GN19NGG	69,041,571	33,076,776
	AN19NGG	81,077,137 **(17%)**	37,795,743 **(14%)**
N. Meningitis			
	GN23NNNNGATT	4,055,280	3,227,027
	AN23NNNNGATT	6,942,105 **(71%)**	1,966,548 **(64%)**
T. thermophilus			
	GN17NNAGAAW	5,400,222	2,723,164
	AN17NNAGAAW	10,383,453 **(92%)**	4,593,021 **(69%)**

Supplementary Table 1. Bioinformatic analysis of alternative Cas9 targeting sites in the human genome. Columns moving from left to right indicate the Cas9 species of origin, the CRISPR target site, the frequency of occurrence in the unmasked human genome, and the frequency of occurrence in the repeat-masked human genome. The percent increase is indicated next the appropriate values in bold.

Supplementary Table 2

a

	Promoter	5' nucleotide	Strand	PAM	GC*	Tm*	3' Stability (kcal/mol)**
GFP_213-191	U6	G	-	GGG	65%	68.0°C	7.9 ΔG
GFP_a214-192	H1	A	-	AGG	65%	66.0°C	7.6 ΔG
GFP_219-197	U6	G	-	AGG	65%	69.4°C	11.1 ΔG
GFP_285-307	U6	G	+	AGG	55%	63.8°C	7.0 ΔG
GFP_a292-314	H1	A	+	CGG	45%	57.3°C	8.1 ΔG
GFP_315-293	U6	G	-	TGG	55%	62.8°C	6.7 ΔG
GFP_360-382	U6	G	+	AGG	60%	67.0°C	8.2 ΔG
GFP_361-383	U6	G	+	GGG	55%	64.8°C	7.0 ΔG
GFP_583-561	U6	G	-	GGG	80%	78.9°C	8.6 ΔG
GFP_a584-562	H1	A	-	GGG	75%	76.9°C	9.8 ΔG
GFP_612-590	U6	G	-	CGG	55%	57.6°C	6.4 ΔG
GFP_a676_698	H1	A	+	CGG	70%	72.5°C	6.1 ΔG
GFP_705_683	U6	G	-	CGG	60%	63.0°C	7.8 ΔG

* Calculated based on 20bp target sequence

** Calculated for the five 3' nucleotides based on predicted DNA:DNA hybridization values

b

	Promoter	5' nucleotide	Strand	PAM	GC*	Tm*	3' Stability (kcal/mol)**
AAVS1-g1	U6	G	+	GGG	70%	67.3°C	6.7 ΔG
AAVS1-g2	U6	G	+	TGG	65%	64.7°C	7.8 ΔG
AAVS1-g3	U6	G	-	GGG	60%	65.5°C	10.9 ΔG
AAVS1-a1	H1	A	-	CGG	45%	54.3°C	6.0 ΔG
AAVS1-a2	H1	A	-	TGG	60%	65.5°C	12.4 ΔG
AAVS1-a3	H1	A	-	CGG	45%	55.3°C	8.2 ΔG

c

	CRISPR target
GFP_213-191	5'-GCACTGCACGCCGTAGGTCA-3'
GFP_a214-192	5'-AGCACTGCACGCCGTAGGTC-3'
GFP_219-197	5'-GCTGAAGCACTGCACGCCGT-3'
GFP_285-307	5'-GGAGCGCACCATCTTCTTCA-3'
GFP_a292-314	5'-ACCATCTTCTTCAAGGACGA-3'
GFP_315-293	5'-GCCGTCGTCCTTGAAGAAGA-3'
GFP_360-382	5'-GGTGAACCGCATCGAGCTGA-3'
GFP_361-383	5'-GTGAACCGCATCGAGCTGAA-3'
GFP_583-561	5'-GCACGGGGCCGTCGCCGATG-3'
GFP_a584-562	5'-AGCACGGGGCCGTCGCCGAT-3'
GFP_612-590	5'-GGTGCTCAGGTAGTGGTTGT-3'
GFP_a676_698	5'-ACCGCCGCCGGGATCACTCT-3'
GFP_705_683	5'-GTCCATGCCGAGAGTGATCC-3'

Supplementary Table 2. gRNA targeting sequences and properties. **(a)** eGFP targeting constructs indicating the eGFP coordinates, gRNA promoter, 5' nucleotide, targeting strand, PAM motif, GC content, Tm, and thermodynamic stability. **(b)** AAVS-1 targeting sequences indicating the gRNA promoter, 5' nucleotide, targeting strand, PAM motif, GC content, Tm, and thermodynamic stability. **(c)** Sequence of the 20 base gRNA constructs targeting eGFP.