Efficient ablation of genes in human hematopoietic stem and effector cells using CRISPR/Cas9

Pankaj K. Mandal[1,2,14], Leonardo M. R. Ferreira[1,3,14], Ryan Collins[4], Torsten B. Meissner[1], Christian L. Boutwell[5], Max Friesen[1], Vladimir Vrbanac[5,6], Brian S. Garrison[1,2,7], Alexei Stortchevoi[4], David Bryder[8], Kiran Musunuru[1,9,10,11], Harrison Brand[4], Andrew M. Tager[5,6], Todd M. Allen[5], Michael E. Talkowski[4,11,12], Derrick J. Rossi[1,2,7,9,15] and Chad A. Cowan[1,9,11,13,15]

SUPPLEMENTAL DATA

EXPERIMENTAL PROCEDURES

Molecular Biology

We subcloned a human-codon-optimized *Cas9* gene with a C-terminal nuclear localization signal (Mali et al., 2013) into a CAG expression plasmid with 2A-GFP (Ding et al., 2013). The guide RNAs (gRNAs) were separately expressed from a plasmid with the human *U6* polymerase III promoter (Mali et al., 2013). Each gRNA sequence was introduced in this plasmid using BbsI restriction sites. All guides were designed using the online optimized design tool at http://crispr.mit.edu. gRNA and primer sequences are enlisted under **List of reagents** at the end of experimental procedures.

Primary blood cell isolation

Primary $CD4^+$ T cells were isolated from peripheral blood (Leukopacs, MGH) using RosetteSep $CD4^+$ T cell enrichment cocktail (STEMCELL Technologies). $CD34^+$ cells from G-CSF mobilized peripheral blood were purchased from AllCells.

Cell culture

HEK293T, K562 and T cells were cultured in RPMI-1640 medium supplemented with 10% FBS. $CD34^+$ HSPCs were cultured in DMEM/F12 medium supplemented with 10% FBS, β-mercaptoethanol, GlutaMax, Pencillin-Streptomycin, minimum non-essential

amino acid and human cytokine cocktails (GM-CSF, SCF, TPO, Flt3 ligand, IL3, IL6). Cell lines were passaged every 3-4 days.

Transfection of Cells

Human primary $CD4^+$ T cells and $CD34^+$ HSPCs were transfected with Cas9-2A-GFP and gRNA encoding plasmids using respective Amaxa Nucelofection kits (Human CD34 cell Nucleofector kit #VPA-1003 for $CD34^+$ HSPCs, Human T cell Nucleofector kit #VPA-1002 for $CD4^+$ T cells, and Cell Line Nucleofector kit V #VCA-1003 for K562 cells) and cell-specific Nucleofection program (U-008 for $CD34^+$ HSPCs, U-014 for CD4+ T cells, and T-016 for K562 cells) with an Amaxa Nucleofection II device as per manufacturers instructions with minor modifications. HEK293T cells were seeded in 6-well plates the day before transfection and transfected using Fugene 6 (Promega). For dual gRNA combinations in $CD34^+$ HSPCs, individual gRNAs were used at half the amount of single gRNA conditions, keeping total gRNA amount the same across the experimental settings.

Cell sorting

For the *CCR5* targeting experiments in $CD34^+$ HSPCs, cells were thawed and cultivated for 6-8 hours in complete DMEM/F12 medium prior to transfection. Following transfection, cells were plated in antibiotic free medium. 24 hours post-transfection, cells were harvested in sample medium (2% FBS and 2 mM EDTA in PBS without Ca^{2+} and Mg^{2+}) and HSPCs were stained with anti-CD34-PE/Cy7 (clone: 581, Biolegend, 1:100) for 20 min on ice. Live, GFP^+ $CD34^+$ HSPCs were sorted using an Aria II sorter (BD Bioscience) and plated in complete DMEM/F12 medium supplemented with human cytokine cocktail and culture for 72 hours prior to analysis. For the B2M experiments, cells were stained with mouse monoclonal anti-B2M-APC antibody (clone: 2M2, Biolegend) 48 or 72 hours post-transfection to estimate loss of B2M expression. FACS data were analyzed using FlowJo software.

Colony forming cell (CFC) assay

1500 sorted $CD34^+$ HSPCs were plated in 1.5 ml of methylcellulose (MethoCult™ H4034 Optimum, Stem Cell Technologies) on a 35 mm cell culture dish and cultured for two weeks at 37 °C in a 5% CO2 incubator. Colonies were then counted and scored.

Surveyor/CEL assay

Amplicons spanning the different targeted regions were PCR amplified using the Phusion polymerase and HF Buffer (New England Biolabs) and CEL assay was carried out using the Surveyor Mutation detection kit (Transgenomic) according to the manufacturer's instructions, with minor modifications.

Clonal analysis

Colonies grown in MethoCult^{TM} H4034 Optimum were individually picked and lysed in 50 µl of lysis buffer containing detergent and Proteinase K buffer (van der Burg et al., 2011). Samples were digested at 56 °C for 1 h followed by Proteinase K inactivation at 95 °C for 15 min. 50 µl of water with RNase A were added to the samples. 2 µl of samples were use for PCR. A 436 bp amplicon spanning the targeted region was PCR amplified using GoTaq® Green Master Mix (Promega) as per manufacturer's instructions. For single gRNA experiments, PCR products were analyzed by Sanger sequencing (Macrogen). For dual gRNA experiments, PCR products were analyzed by agarose gel electrophoresis.

In vivo transplantation of CD34$^+$ HSPCs

NOD/SCID/IL2R$\gamma^{-/-}$ (NSG) mice (The Jackson Laboratory) were housed in a pathogen-free facility, maintained in microisolator cages, and fed autoclaved food and water. Adult (6-8 weeks of age) NSG mice were conditioned with sub-lethal (2 Gy) whole-body irradiation. The conditioned recipients were transplanted with 75,000-sorted CD34$^+$ HSPCs expressing Cas9 alone (control group, n=2) or Cas9 with crCCR5_D+Q gRNAs (experimental group, n=5). At 12 weeks post-transplantation, all mice were euthanized and blood, bone marrow, and spleen samples were taken for characterization of human hematopoietic cell chimerism. Human CD45$^+$ cells were sorted for DNA isolation and analysis of *CCR5* deletion.

Single Cell PCR assay

48 h after electroporation with Cas9 and different gRNA combinations, GFP$^+$ primary CD4$^+$ T cells were sorted into 384-well plates (Twin tec skirted PCR plate, Eppendorf) containing 4 µl of prepGEM Tissue (ZyGEM) per well. Cells were lysed and digested following the manufacturer's instructions to release the genomic DNA. A multiplexed

nested PCR was then carried out in the same plate with the primer combinations represented in Supplemental Figures S2C and S2F. The resulting DNA was then used in two subsequent PCR reactions, one amplifying a positive control region, to determine successful genomic DNA isolation from a single cell, and another one amplifying a region lying between the two gRNA binding sites, allowing us to quantify the percentage of cells homozygous for the dual gRNA induced deletion (Supplemental Figures S2D and S2G). Cells were scored based on the melting curves of the PCR amplicons. PCR reactions were performed using an Applied Biosystems ViiA 7 real-Time PCR System (Life Technologies).

Off-Target Prediction and Capture Sequencing

Degenerate gRNA off-target sequences were predicted for each gRNA targeting *CCR5* using the CRISPR Design off-target prediction tool (Hsu et al., 2013). Off-target sequences were further supplemented by alignment of each gRNA to the human genome using BOWTIE of which all results up to and including 3 mismatches were added to the total off-target list (Langmead et al., 2009). All instances of each predicted off-target sequence existent in the human genome reference build GRCh37v71 were recorded (Supplementary Table T1). Each guide RNA target site (n=5) and predicted off-target site (n=126) was selected for capture sequencing using the Agilent SureSelectXT Target Enrichment System. Capture intervals were expanded by approximately 500 bp in both the 5' and 3' directions to ensure exhaustive capture of the targeted region and detection of any genetic lesion occurring at or near a predicted gRNA on- or off-target site, as we have previously shown accurate capability to detect translocations and inversions using targeted capture of probes in proximity to a rearrangement breakpoint using a CapBP procedure as described (Talkowski et al., 2011). Probes were tiled with 60-fold greater density over each predicted 23bp on- or off-target gRNA binding site than the flanking kilobase of sequence. Isogenic $CD34^+$ HSPCs-mPB were transfected with CRISPR/Cas9 plasmids (one Cas9 only-treated control group, three treatment groups transfected with a single gRNA, and three treatment groups transfected with dual gRNAs). Sorted $CD34^+$ genome edited HSPCs were cultured for two weeks prior to DNA isolation. Capture libraries were prepared from DNA extracted from seven treatment groups. Capture libraries were sequenced as 101 bp paired-end reads on an Illumina HiSeq2000 platform.

NGS Data Processing and Computational Analysis

Read pairs were aligned to GRCh37v71 with Bwa-MEM v0.7.10-r789 (Li, arXiv 2013). Alignments were processed using PicardTools and SAMBLASTER (Faust and Hall, 2014). The Genome Analysis Toolkit (GATK) v3.1-1-g07a4bf8 was applied for base quality score recalibration, insertion/deletion (InDel) realignment, duplicate removal, and single nucleotide variant (SNV) and InDel discovery and genotyping per published best-practice protocols (McKenna et al, Genome Res 2010; DePristo et al, Nat Genet 2011; Van der Auwera et al, 2013). SNVs and InDels were annotated using ANNOVAR (Wang et al., 2010). Structural variants (SVs) were detected with LUMPY v0.2.5 considering both anomalous pair and split read evidence at a minimum call weight threshold of 7 and an evidence set score ≤ 0.05 (Layer et al., 2014). Candidate copy number variants (CNVs) were further statistically assessed by Student's t-test for a concomitant change in depth of coverage across the putative CNV. As a final exhaustive measure, each on- and off-target site was manually scrutinized in each capture library for evidence supporting predictable mutagenesis that is not detectable by the computational algorithms due to low levels of mosaicism in the sequenced population.

Evaluation of Off-Target Mutation Frequency

A statistical framework was developed to assess off-target mutational burden for each gRNA. For each off-target site (n=126), all reads with at least one nucleotide of overlap with that 23bp off-target site were collected and their CIGAR information was tabulated into categories as follows: reads representing small InDels (CIGAR contains at least one "I" or "D"), reads potentially representative of other rearrangements (CIGAR contains at least one "S" or "H"), and reads reflecting reference sequence (CIGAR did not match either of the two former categories). Such counts were gathered at all 126 sites in all seven libraries and were further pooled to form comparison groups of "treatment" libraries (transfected gRNA matches corresponding off-target site gRNA) and "control" libraries (transfected gRNA does not match corresponding off-target site gRNA). Next, at each off-target site, relative n-fold enrichment of each read classification between treatment and control libraries was evaluated. Finally, a one-tailed Fisher's Exact Test was performed to assess the statistical significance of enrichment of variant reads in treatments versus controls at each off-target site, followed by Bonferroni correction to retain an experiment-wide significance threshold of $\alpha = 0.05$.

List of Reagents

Guides targeting CCR5.

Guide ID	Sequence
crCCR5_A	GCTGCCGCCCAGTGGGACTT
crCCR5_B	GATCTGGTAAAGATGATTCC
crCCR5_C	ACAATGTGTCAACTCTTGAC
crCCR5_D	TCACTATGCTGCCGCCCAGT
crCCR5_O	GGTGACAAGTGTGATCACTT
crCCR5_P	GACAAGTGTGATCACTTGGG
crCCR5_Q	GCTGTGTTTGCGTCTCTCCC

Guides targeting B2M

Guide ID	Sequence
crB2M_1	GATGTCTCGCTCCGTGGCCT
crB2M_2	CTCGCGCTACTCTCTCTTTC
crB2M_3	GACTCACGCTGGATAGCCTC
crB2M_4	CCAGAAAGAGAGAGTAGCGC
crB2M_5	CACAGCTAAGGCCACGGAGC
crB2M_6	GGCCGAGATGTCTCGCTCCG
crB2M_7	TTGCGGGAGCGCATGCCTTT
crB2M_8	CCACCTCTTGATGGGGCTAG
crB2M_9	ATACCTTGGGTTGATCCACT
crB2M_10	CGTGAGTAAACCTGAATCTT
crB2M_11	AAGTCAACTTCAATGTCGGA
crB2M_12	CATAGATCGAGACATGTAAG
crB2M_13	GCTACTCTCTCTTTCTGGCC
crB2M_14	ACCCAAACCAAGCCTTTCTA
crB2M_15	TATAAGTGGAGGCGTCGCGC

PCR primers used for CEL assay

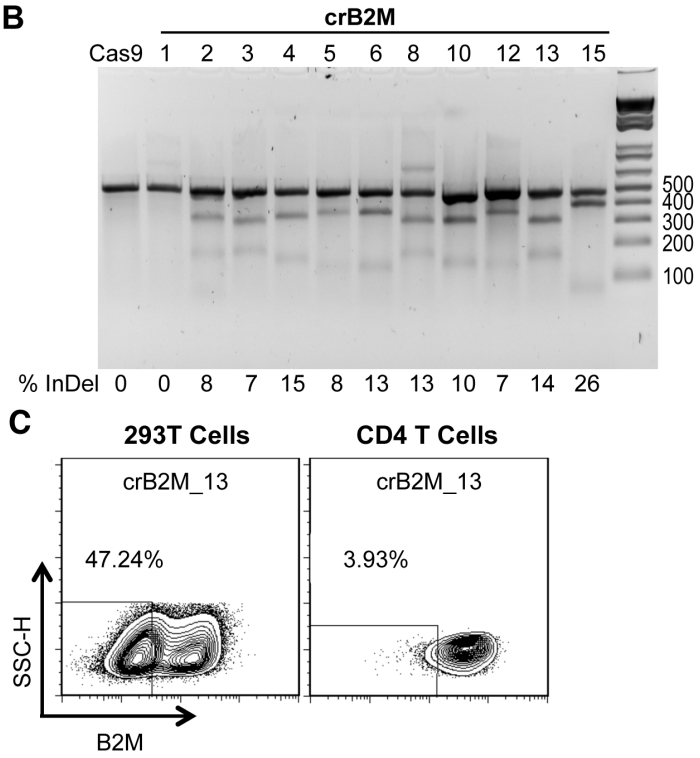
Fw: CCR5_CEL_F	CTGCAAAAGGCTGAAGAGCA	For all guides targeting CCR5
Rev: CCR5_CEL_R	CCCCAAGATGACTATCTTTAATGTC	
Fw: Le277	CTGGCTTGGAGACAGGTGAC	For crB2M_6 and crB2M_13
Rev: Le679	GACGCTTATCGACGCCCTAA	
Fw: Le680	CAAAATCTTGCCGCCTTCCC	For crB2M_8
Rev: Le681	ACTTTCCAAAATGAGAGGCATGA	
Fw: Le682	CCAGAGTGGAAATGGAATTGGGA	For crB2M_10
Rev: Le683	ACTCATACACAACTTTCAGCAGCTT	
Fw: Le684	TCATGGGTAGGAACAGCAGC	For crB2M_12
Rev: Le685	TCTCCTCAGCAGAGATGTCC	

FIGURES

A

% B2M negative cells

crB2M	Mean±SEM
None	1.5±0.2
1	6.9±1.0
2	24.0±3.8
3	10.0±1.5
4	47.1±7.3
5	22.9±5.5
6	46.6±0.6
7	7.0±0.3
8	10.2±0.6
9	11.0±3.2
10	30.4±2.8
11	13.9±2.5
12	15.5±1.6
13	48.0±1.8
14	8.6±0.2
15	25.1±2.9



D

% B2M negative cells

crB2M	Donors	Mean±SEM
None	n=6	1.1±0.3
6	n=6	1.4±0.2
7	n=3	3.4±0.7
8	n=6	2.3±0.3
10	n=6	3.2±1.2
11	n=3	2.9±0.6
12	n=6	4.0±0.9
13	n=6	4.7±0.9

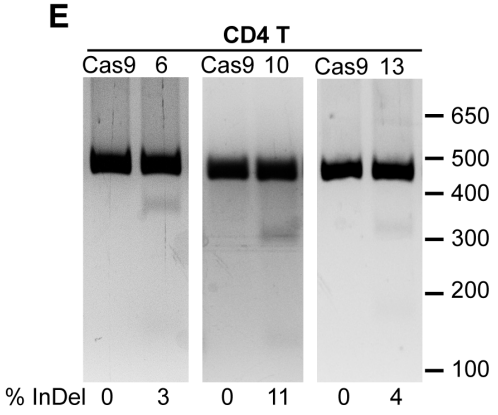


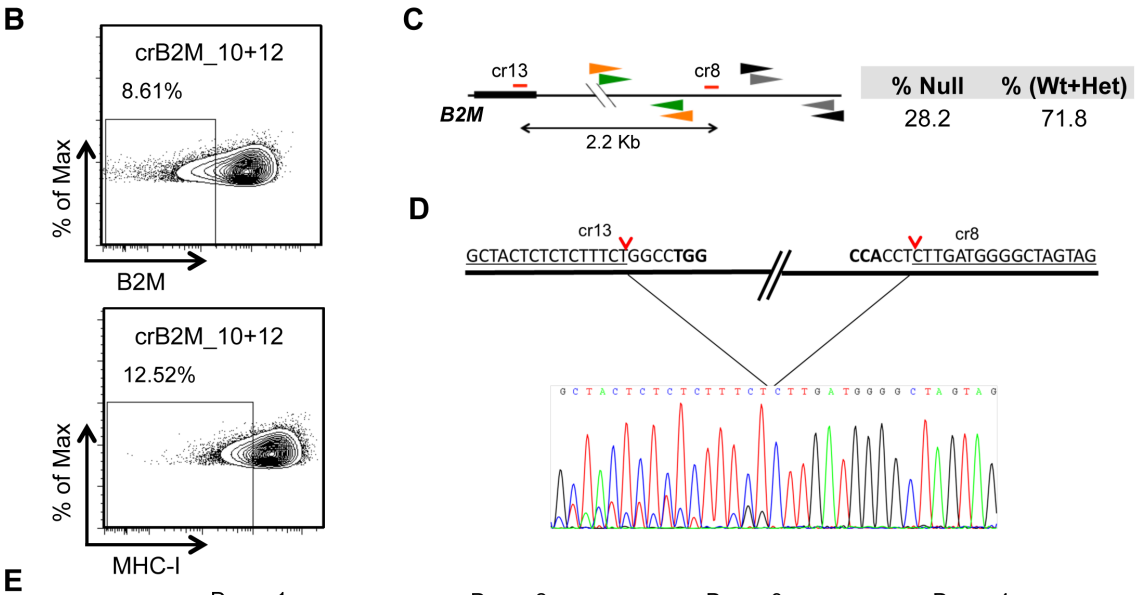
Figure S1. Evaluation of on target mutational efficiencies of various gRNAs targeting *B2M* (Related to Figure 1).

A) B2M deletion efficiency for all gRNAs targeting *B2M* locus in HEK293T cells as measured by flow cytometry. Pooled data from 3 independent experiments shown as mean±SEM. B) B2M deletion efficiencies of selected guides in HEK293T cells, measured as % InDels by CEL Surveyor assay. C) Comparison of B2M surface expression in HEK293T cells and primary $CD4^+$ T cells when transfected with Cas9 and guide crB2M_13. D) B2M deletion efficiency for selected guides targeting the *B2M* locus in primary $CD4^+$ T-cells, as measured by flow cytometry. E) B2M deletion efficiencies of selected guides in primary CD4+ T cells, measured as % InDels by CEL Surveyor assay.

A

% B2M negative cells

crB2M	Donor 1	Donor 2	Donor 3	Mean±SEM
6+7	2.14	4.47	2.99	3.2±0.68
6+8	2.60	3.31	3.25	3.05±0.23
10+11	2.81	4.01	3.96	3.59±0.39
10+12	8.61	5.69	5.82	6.71±0.95
13+8	34.15	6.22	13.64	18.0±8.35
8+10	0.09	3.34	2.95	2.13±1.02



E

	Donor 1			Donor 2			Donor 3			Donor 4		
crCCR5	# clones	+/- (%)	-/- (%)	# clones	+/- (%)	-/- (%)	# clones	+/- (%)	-/- (%)	# clones	+/- (%)	-/- (%)
A+B	83	11.5	9	96	27	8	189	10	23	92	5.4	37
D+Q	88	6.8	9	84	25	42	154	5.8	33	61	23	23
C+D	87	12.6	4.6	NA	-	-	189	13	29	96	32.3	32.3

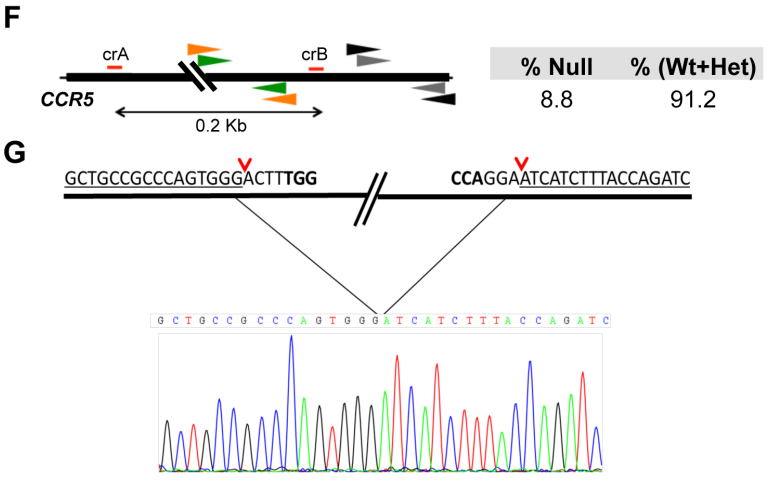


Figure S2. Targeting efficiency of dual gRNA combinations (Related to Figure 2).

A) B2M deletion efficiency for 6 dual gRNA combinations from three independent donors as measured by flow cytometry. B) FACS plots showing loss of MHC class I surface expression (bottom panel) following *B2M* deletion (top panel). C) Schematic of the single cell nested PCR

strategy for the *B2M* locus (left panel), black and gray arrowheads: control primer pairs, orange and green arrowheads: primer pairs flanking targeting region. % B2M null single cells is shown (right panel, n=301). D) Sanger sequencing chromatogram showing predicted deletion of targeted region at *B2M* locus. E) Clonal *CCR5* deletion efficiency for three dual gRNA combinations in $CD34^+$ HSPC-mPB obtained from multiple donors. DNA isolated from individual colony was analyzed by PCR and gel electrophoresis. F) Schematic of the single cell nested PCR strategy (left panel) for determining deletion of *CCR5* in primary $CD4^+$ T cells. % *CCR5* null single cells is shown (right panel, n=363). G) Sanger sequencing chromatogram shows predicted deletion at targeted region.

A

crCCR5_A: GCTGCCGCCCAGTGGGACTTTGG
CCR2: **A**CTG**T**C**T**CCC**T**GT**A**G**AA**C**T**GG

crCCR5_B: GATCTGGTAAAGATGATTCCTGG
CCR2: **C**AT**T**T**A**GTAAAGATGATTCCTGG

crCCR5_C: ACAATGTGTCAACTCTTGACAGG
CCR2: **G**CA**T**T**T**C**T**C**GTT**CTC-TGA-AG**T**

crCCR5_D: TCACTATGC-TGCCGCCCAGTGG
CCR2: TCACTA**G**GCATGC**T**GCC-AG**A**G**C**

crCCR5_Q: GCTGTGTTTGCGTCTCTCCCAGG
CCR2: GCTGTGTTTGC**T**TCT**G**TCCCAGG

B

Mutation	crCCR5 treatment								
	B			A+B			A		
	Reads Supporting Mutation	*Total Reads at Site*	*Frequency*	*Reads Supporting Mutation*	*Total Reads at Site*	*Frequency*	*Reads Supporting Mutation*	*Total Reads at Site*	*Frequency*
One Base Insertion	30	5,963	0.50%	2	5,339	0.04%	0	4,678	0.00%
Two Base Insertion	0	5963	0.00%	1	5,339	0.02%	0	4,678	0.00%
One Base Deletion	5	5,963	0.08%	9	5,339	0.17%	4	4,678	0.09%
Two Base Deletion	1	5,963	0.02%	1	5,339	0.02%	0	4,678	0.00%
Total	**36**	5,963	**0.60%**	**13**	**5,339**	**0.24%**	**4**	**4,678**	**0.09%**

Figure 3. Potential off-target sites identified in *CCR5* homologue *CCR2* and analysis of events detected at the single off-target site in which mutagenesis was significantly detected above background (Related to Figure 4).

A) Sequence alignment of *CCR5* gRNAs utilized in this study in relation to the closest homologous sequence in *CCR2* showing mismatched nucleotides in bold. Noteworthy is the fact that gRNA crCCR5_B, which yielded the sole significantly detected off-target mutagenesis in *CCR2* (detailed in panel B), has 3 nucleotide mismatches, which are distal to the PAM (underlined) and seed (grey box) sequences. B) In-depth analyses of all sequence reads at the single off-target site in which mutagenesis was significantly detected above background in both capture libraries treated with the associated gRNA (B; libraries treated with single gRNA crCCR5_B & dual-gRNA crCCR5_A+B), as well as the library treated with gRNA crCCR5_A as a comparison. Total off-target mutation frequency at this site was 0.6% in the single gRNA treatment (crCCR5_B) and notably decreased to 0.24% in the dual gRNA treatment (crCCR5_A+B) in which gRNA plasmid concentration of each gRNA was half of that utilized in single gRNA treatments.

SUPPLEMENTAL TABLES

Table S1. Predicted gRNA mapping in Ensembl GRCh37v71 (related to Figure 4). See the spread sheet.

Table S2. Guide Pair crCCR5_A+B On-Target Alleles, Related to Figure 4.

Allele	Cas9 Guide Site	Sequence	Split Reads			Estimated Allele Frequency
			(+) Strand	*(-) Strand*	***TOTAL***	
Reference	A (Distal)	TATGCTGCCGCCCAGTGGGACTTTGGAAATACAATGTGTCAACTC	1836	1728	**3564**	76.78%
Reference	B (Proximal)	GGCTGTGTTTGCGTCTCTCCCAGGAATCATCTTTACCAGATCTCA	1340	1753	**3093**	73.80%
206bp deletion	AB (Both)	TATGCTGCCGCCCAGTGGGA | ATCATCTTTACCAGATCTCAAAAAG	411	411	**822**	18.61%
205bp inversion	AB (Both)	GAGTTGACACATTGTATTTCCAAAG | ATCATCTTTACCAGATCTCA TATGCTGCCGCCCAGTGGGA | TCCTGGGAGAGACGCAAACACAGCC	60	78	**138**	3.12%
1bp deletion	B (Proximal)	TGGCTGTGTTTGCGTCTCTCCCAGG | ATCATCTTTACCAGATCTCA	23	27	**50**	1.19%
206bp deletion with C insertion at break	AB (Both)	TATGCTGCCGCCCAGTGGGA | CATCATCTTTACCAGATCTCAAAAA	19	8	**27**	0.61%
1bp deletion	A (Distal)	TATGCTGCCGCCCAGTGGGA | TTTGGAAATACAATGTGTCAACTCT	14	11	**25**	0.54%
207bp deletion	AB (Both)	TATGCTGCCGCCCAGTGGGA | TCATCTTTACCAGATCTCAAAAAGA	10	8	**18**	0.41%
A insertion	B (Proximal)	GCTGTGTTTGCGTCTCTCCCAGGAA | ATCATCTTTACCAGATCTCA	7	8	**15**	0.36%
A insertion	A (Distal)	TATGCTGCCGCCCAGTGGGA | ACTTTGGAAATACAATGTGTCAACT	4	7	**11**	0.24%
3bp deletion	A (Distal)	TATGCTGCCGCCCAGTGGGA | TGGAAATACAATGTGTCAACTCTTG	7	0	**7**	0.15%
2bp deletion	B (Proximal)	GTGGCTGTGTTTGCGTCTCTCCCAG | ATCATCTTTACCAGATCTCA	4	2	**6**	0.14%
TC insertion	A (Distal)	TATGCTGCCGCCCAGTGGGA | TCCTTTGGAAATACAATGTGTCAAC	3	3	**6**	0.13%
209bp deletion	AB (Both)	TATGCTGCCGCCCAGTGGGA | ATCTTTACCAGATCTCAAAAAGAAG	1	4	**5**	0.11%
4bp deletion	B (Proximal)	TGGTGGCTGTGTTTGCGTCTCTCCC | ATCATCTTTACCAGATCTCA	4	0	**4**	0.10%
205bp deletion	AB (Both)	TATGCTGCCGCCCAGTGGGA | AATCATCTTTACCAGATCTCAAAAA	2	2	**4**	0.09%
206bp deletion with A insertion at break	AB (Both)	TCACTATGCTGCCGCCCAGTGGGAA | ATCATCTTTACCAGATCTCA	2	2	**4**	0.09%
12bp deletion	A (Distal)	TATGCTGCCGCCCAGTGGGA | AATGTGTCAACTCTTGACAGGGCTC	4	0	**4**	0.09%
2bp deletion	A (Distal)	TATGCTGCCGCCCAGTGGGA | TTGGAAATACAATGTGTCAACTCTT	2	1	**3**	0.06%
Unidentifiable novel sequence insertion	B (Proximal)	GAGTTACATGATCCCCCATGTTGTG | ATCATCTTTACCAGATCTCA	2	0	**2**	0.05%
5bp deletion	B (Proximal)	GTGGTGGCTGTGTTTGCGTCTCCCC | ATCATCTTTACCAGATCTCA	1	0	**1**	0.02%
A->T transversion	B (Proximal)	GGCTGTGTTTGCGTCTCTCCCAGGTATCATCTTTACCAGATCTCA	0	1	**1**	0.02%
208bp deletion	AB (Both)	TATGCTGCCGCCCAGTGGGA | CATCTTTACCAGATCTCAAAAAGAA	1	0	**1**	0.02%
7bp deletion	A (Distal)	TATGCTGCCGCCCAGTGGGA | AATACAATGTGTCAACTCTTGACAG	0	1	**1**	0.02%
8bp deletion	A (Distal)	TATGCTGCCGCCCAGTGGGA | ATACAATGTGTCAACTCTTGACAGG	0	1	**1**	0.02%
C->T transition	A (Distal)	TATGCTGCCGCCCAGTGGGATTTTGGAAATACAATGTTTCAACTC	1	0	**1**	0.02%

Table S3. Guide Pair crCCR5_C+D On-Target Alleles, Related to Figure 4.

Allele	Cas9 Guide Site	Sequence	Split Reads			Estimated Allele Frequency
			(+) Strand	*(-) Strand*	***TOTAL***	
Reference	C (Proximal)	ATACAATGTGTCAACTCTTGACAGGGCTCTATTTTATAGGCTTCT	1704	1457	**3161**	79.72%
Reference	D (Distal)	GGCTCACTATGCTGCCGCCCAGTGGGACTTTGGAAATACAATGTG	1659	1459	**3118**	78.64%
35bp deletion	CD (Both)	GGCTCACTATGCTGCCGCCC | GACAGGGCTCTATTTTATAGGCTTC	310	270	**580**	14.63%
34bp deletion	CD (Both)	GGCTCACTATGCTGCCGCCC | TGACAGGGCTCTATTTTATAGGCTT	97	99	**196**	4.94%
33bp deletion	CD (Both)	GGCTCACTATGCTGCCGCCC | TTGACAGGGCTCTATTTTATAGGCT	23	11	**34**	0.86%
1bp deletion	C (Proximal)	AATACAATGTGTCAACTCT | GACAGGGCTCTATTTTATAGGCTTCT	6	3	**9**	0.23%
T->G transversion 1bp 5' of PAM	D (Distal)	GGCTCACTATGCTGCCGCCCAGGGGGACTTTGGAAATACAATGTG	3	0	**3**	0.08%
3bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | GGGACTTTGGAAATACAATGTGTCA	3	0	**3**	0.08%
Unidentifiable novel sequence insertion	C (Proximal)	CCGGCAAACAAACCACCGC | GACAGGGCTCTATTTTATAGGCTTCT	3	0	**3**	0.08%
3bp deletion	C (Proximal)	GAAATACAATGTGTCAACT | GACAGGGCTCTATTTTATAGGCTTCT	3	0	**3**	0.08%
5bp deletion	C (Proximal)	TCACTATACAATGTGTCAAGAC | AGGGCTCTATTTTATAGGCTTCT	2	0	**2**	0.05%
2bp deletion	C (Proximal)	AAATACAATGTGTCAACTC | GACAGGGCTCTATTTTATAGGCTTCT	0	2	**2**	0.05%
34bp deletion; breaks offset 1bp 3' of both Cas9 sites	CD (Both)	GGCTCACTATGCTGCCGCCC | AACAGGGCTCTATTTTATAGGCTTC	1	0	**1**	0.03%
G->A transition middle base of PAM	C (Proximal)	ATACAATGTGTCAACTCTTGACAAGGCTCTATTTTATAGGCTTCT	1	0	**1**	0.03%
19bp deletion	C (Proximal)	CCCAGTGGGACTTTGGAAAA | ACAGGGCTCTATTTTATAGGCTTCT	1	0	**1**	0.03%
T->C transition 2bp 5' of Cas9 site	C (Proximal)	ATACAATGTGTCAACTCCT | GACAGGGCTCTATTTTATAGGCTTCT	1	0	**1**	0.03%
T->C transition	C (Proximal)	ATACAATGTGTCAACTCTC | GACAGGGCTCTATTTTATAGGCTTCT	0	1	**1**	0.03%

Table S4. Guide Pair crCCR5_D+Q On-Target Alleles, Related to Figure 4.

Allele	Cas9 Guide Site	Sequence	Split Reads			Estimated Allele Frequency
			(+) Strand	*(-) Strand*	*TOTAL*	
Reference	D (Distal)	GGCTCACTATGCTGCCGCCCAGTGGGACTTTGGAAATACAATGTG	1662	1261	**2923**	54.53%
Reference	Q (Proximal)	GGGTGGTGGCTGTGTTTGCGTCTCTCCCAGGAATCATCTTTACCA	1296	1535	**2831**	52.82%
205bp deletion	DQ (Both)	TTCTGGGCTCACTATGCTGCCGCCC | CCCAGGAATCATCTTTACCA	332	269	**601**	11.21%
205bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | CCCAGGAATCATCTTTACCAGATCT	321	248	**569**	10.62%
204bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | TCCCAGGAATCATCTTTACCAGATC	235	278	**513**	9.57%
204bp deletion	DQ (Both)	TCTGGGCTCACTATGCTGCCGCCCT | CCCAGGAATCATCTTTACCA	223	273	**496**	9.25%
205bp inversion	DQ (Both)	GGCTCACTATGCTGCCGCCC | AGAGACGCAAACACAGCCACCACCC CACATTGTATTTCCAAAGTCCCACT | CCCAGGAATCATCTTTACCA	48	41	**89**	1.66%
19bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | AATGTGTCAACTCTTGACAGGGCTC	41	24	**65**	1.21%
206bp inversion	DQ (Both)	GGCTCACTATGCTGCCGCCC | AGAGACGCAAACACAGCCACCACCC ACATTGTATTTCCAAAGTCCCACTT | CCCAGGAATCATCTTTACCA	26	18	**44**	0.82%
T insertion	D (Distal)	GGCTCACTATGCTGCCGCCC | TAGTGGGACTTTGGAAATACAATGT	8	9	**17**	0.32%
T insertion	Q (Proximal)	GGTGGTGGCTGTGTTTGCGTCTCTT | CCCAGGAATCATCTTTACCA	7	8	**15**	0.28%
206bp deletion	DQ (Both)	CTTCTGGGCTCACTATGCTGCCGCC | CCCAGGAATCATCTTTACCA	4	9	**13**	0.24%
Unidentifiable novel sequence insertion	Q (Proximal)	GCTGTTTCCTGTGTGAAATTGTTAT | CCCAGGAATCATCTTTACCA	5	7	**12**	0.22%
208bp deletion	DQ (Both)	CCCTTCTGGGCTCACTATGCTGCCG | CCCAGGAATCATCTTTACCA	4	6	**10**	0.19%
208bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | AGGAATCATCTTTACCAGATCTCAA	4	5	**9**	0.17%
207bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | CCAGGAATCATCTTTACCAGATCTC	2	7	**9**	0.17%
109bp deletion w/ TTA insertion	D (Distal)	GGCTCACTATGCTGCCGCCC | TTACTGTCGTCCATGCTGTGTTTGC	5	3	**8**	0.15%
205bp deletion	DQ (Both)	TTCTGGGCTCACTATGCTGCCGCC | TCCCAGGAATCATCTTTACCA	4	3	**7**	0.13%
219bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | TTACCAGATCTCAAAAAGAAGGTCT	3	4	**7**	0.13%
Unidentifiable novel sequence insertion	D (Distal)	GGCTCACTATGCTGCCGCCC | TGTATACCGTCGACCTCTAGCTAGA	3	3	**6**	0.11%
C insertion	D (Distal)	GGCTCACTATGCTGCCGCCC | CAGTGGGACTTTGGAAATACAATGT	2	3	**5**	0.09%
2bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | TGGGACTTTGGAAATACAATGTGTC	2	3	**5**	0.09%
14bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | AATACAATGTGTCAACTCTTGACAG	0	5	**5**	0.09%
202bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | TCTCCCAGGAATCATCTTTACCAGA	0	5	**5**	0.09%
202bp deletion	DQ (Both)	TGGGCTCACTATGCTGCCGCCCTCT | CCCAGGAATCATCTTTACCA	0	5	**5**	0.09%
215bp deletion	DQ (Both)	TACTGTCCCCTTCTGGGCTCACTAT | CCCAGGAATCATCTTTACCA	2	2	**4**	0.07%
1bp deletion	Q (Proximal)	TGGGTGGTGGCTGTGTTTGCGTCTT | CCCAGGAATCATCTTTACCA	2	2	**4**	0.07%
4bp deletion	Q (Proximal)	ACTTGGGTGGTGGCTGTGTTTGCGT | CCCAGGAATCATCTTTACCA	1	3	**4**	0.07%
A insertion	D (Distal)	GGCTCACTATGCTGCCGCCC | AAGTGGGACTTTGGAAATACAATGT	3	0	**3**	0.06%
1bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | GTGGGACTTTGGAAATACAATGTGT	3	0	**3**	0.06%
72bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | TCTTCATCATCCTCCTGACAATCGA	3	0	**3**	0.06%
48bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | TTATAGGCTTCTTCTCTGGAATCTT	3	0	**3**	0.06%
Unidentifiable novel sequence insertion	Q (Proximal)	TTCTTCGATCAGTCTAAAAATGGCT | CCCAGGAATCATCTTTACCA	3	0	**3**	0.06%
214bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | CATCTTTACCAGATCTCAAAAAGAA	2	1	**3**	0.06%
50bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | AATAGGCTTCTTCTCTGGAATCTTC	0	3	**3**	0.06%
152bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | TTTGGGGTGGTGACAAGTGTGATCA	0	3	**3**	0.06%
181bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | TTAAAAGCCAGGACGGTCACCTTTG	0	3	**3**	0.06%
192bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | TGTTTGCGTCTCTCCCAGGAATCAT	1	1	**2**	0.04%
13bp deletion	Q (Proximal)	AGTGTGATCACTTGGGTGGTGGCTG | CCCAGGAATCATCTTTACCA	1	1	**2**	0.04%
62bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | TCTGGAATCTTCTTCATCATCCTCC	0	2	**2**	0.04%
A->T transversion	D (Distal)	GGCTCACTATGCTGCCGCCCTGTGGGACTTTGGAAATACAATGTG	0	2	**2**	0.04%
8bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | TTTGGAAATACAATGTGTCAACTCT	1	0	**1**	0.02%
195bp deletion	DQ (Both)	GGCTCACTATGCTGCCGCCC | CCCAAGAATCATCTTTACCAGATCT	1	0	**1**	0.02%
207bp deletion	DQ (Both)	CCTTCTGGGCTCACTATGCTGCCGC | CCCAGGAATCATCTTTACCA	1	0	**1**	0.02%
1bp deletion	Q (Proximal)	TGGGTGGTGGCTGTGTTTGCGTCTC | CCCAGGAATCATCTTTACCA	1	0	**1**	0.02%
17bp deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | ACAATGTGTCAACTCTTGACAGGGC	0	1	**1**	0.02%
A deletion	D (Distal)	GGCTCACTATGCTGCCGCCC | GTGGGACTTTGGAAATTCAATGTGT	0	1	**1**	0.02%
205bp deletion w/ G insertion	DQ (Both)	GGCTCACTATGCTGCCGCCC | GCCCAGGAATCATCTTTACCAGATC	0	1	**1**	0.02%
187bp deletion w/ 1bp deletion 1bp 5' of proximal guide PAM	DQ (Both)	GGCTCACTATGCTGCCGCCC | GGCTGTGTTTGCGTCTCTCCAGGAA	0	1	**1**	0.02%
218bp deletion	DQ (Both)	TCTTACTGTCCCCTTCTGCGTCTCT | CCCAGGAATCATCTTTACCA	0	1	**1**	0.02%
C insertion	Q (Proximal)	GGTGGTGGCTGTGTTTGCGTCTCTC | CCCAGGAATCATCTTTACCA	0	1	**1**	0.02%

Table S5. Off-target Sites with Statistically Significant Mutational Burden and their Comparison, Related to Figure 4.

Guide	Degenerate Sequence	Predicted Binding Site			Fisher's Exact Test *p*		N-Fold Enrichment	
		Chr	*Start*	*End*	*InDel*	*Split*	*InDel*	*Split*
B	CCAGGAATCATCTTTACTAAATG	3	46,399,538	46,399,561	4.4x10^-13	1.0000	3.51	0.47
B	CCAGGCATCTTCTTTACCAGCTC	8	104,258,547	104,258,570	0.0165	1.0000	1.93	0.45
C	AAAATGTGTCAACTCTTGATTAG	12	12,976,312	12,976,335	0.0361	1.0000	1.56	0.27
C	AAAATGTGTCAACTCTTGATTAG	12	12,976,312	12,976,335	0.0361	1.0000	1.56	0.27
C	CCCACTGGGCTGCAGAATACAGA	10	129,722,982	129,723,005	0.0338	0.9996	1.82	0.50
Q	TCTGTGTTTGCCTCTCTCTCAGG	12	19,270,661	19,270,684	0.0086	1.0000	2.08	0.20
Q	CCCGGGAGGGAGGCAAAAACAGC	17	11,833,418	11,833,441	0.0083	1.0000	2.24	0.20

gRNA	gRNA Combinations Tested	On-Target Site		Most Significant Off-Target Site	
		Mutation Frequency	Variant Read Enrichment (Treatment vs. Control)	Variant Read Enrichment (Treatment vs. Control)	Fisher's Exact *p* (Treatment vs Control)
A	A, AB	5.79%	58.59	1.84	0.1616
B	B, AB	16.65%	51.02	3.51	4.416x10^-13
C	C, CD	17.02%	30.88	1.56	0.0361
D	D, CD, DQ	26.23%	57.64	1.82	0.0338
Q	DQ	21.19%	46.84	2.24	0.0083

REFERENCES

Ding, Q., Lee, Y.K., Schaefer, E.A., Peters, D.T., Veres, A., Kim, K., Kuperwasser, N., Motola, D.L., Meissner, T.B., Hendriks, W.T.*, et al.* (2013). A TALEN genome-editing system for generating human stem cell-based disease models. Cell stem cell *12*, 238-251.

Faust, G.G., and Hall, I.M. (2014). SAMBLASTER: fast duplicate marking and structural variant read extraction. Bioinformatics *30*, 2503-2505.

Hsu, P.D., Scott, D.A., Weinstein, J.A., Ran, F.A., Konermann, S., Agarwala, V., Li, Y., Fine, E.J., Wu, X., Shalem, O.*, et al.* (2013). DNA targeting specificity of RNA-guided Cas9 nucleases. Nature biotechnology *31*, 827-832.

Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome biology *10*, R25.

Layer, R.M., Chiang, C., Quinlan, A.R., and Hall, I.M. (2014). LUMPY: a probabilistic framework for structural variant discovery. Genome biology *15*, R84.

Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823-826.

Talkowski, M.E., Ernst, C., Heilbut, A., Chiang, C., Hanscom, C., Lindgren, A., Kirby, A., Liu, S., Muddukrishna, B., Ohsumi, T.K.*, et al.* (2011). Next-generation sequencing strategies enable routine detection of balanced chromosome rearrangements for clinical diagnostics and genetic research. American journal of human genetics *88*, 469-481.

Van der Auwera *et al*. (2013). From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. *Current Protocols in Bioinformatics.* 43:11.10.1-11.10.33.

van der Burg, M., Kreyenberg, H., Willasch, A., Barendregt, B.H., Preuner, S., Watzinger, F., Lion, T., Roosnek, E., Harvey, J., Alcoceba, M.*, et al.* (2011). Standardization of DNA isolation from low cell numbers for chimerism analysis by PCR of short tandem repeats. Leukemia *25*, 1467-1470.

Wang, K., Li, M., and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research *38*, e164.