Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press



Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line

Patrick Essletzbichler, Tomasz Konopka, Federica Santoro, et al.

Genome Res. published online November 4, 2014
Access the most recent version at doi:10.1101/gr.177220.114

P<P Published online November 4, 2014 in advance of the print journal.

Creative Commons License This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**



To subscribe to *Genome Research* go to:
https://genome.cshlp.org/subscriptions

© 2014 Essletzbichler et al.; Published by Cold Spring Harbor Laboratory Press

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Method

Megabase-scale deletion using CRISPR/Cas9 to generate a fully haploid human cell line

Patrick Essletzbichler,[1] Tomasz Konopka,[2] Federica Santoro,[1,4] Doris Chen,[2] Bianca V. Gapp,[2] Robert Kralovics,[2] Thijn R. Brummelkamp,[2,3] Sebastian M.B. Nijman,[1,2] and Tilmann Bürckstümmer[1]

[1]*Haplogen GmbH, 1030 Vienna, Austria;* [2]*Research Center for Molecular Medicine of the Austrian Academy of Sciences (CeMM), 1090 Vienna, Austria;* [3]*The Netherlands Cancer Institute, 1066 CX Amsterdam, The Netherlands*

Near-haploid human cell lines are instrumental for genetic screens and genome engineering as gene inactivation is greatly facilitated by the absence of a second gene copy. However, no completely haploid human cell line has been described, hampering the genetic accessibility of a subset of genes. The near-haploid human cell line HAP1 contains a single copy of all chromosomes except for a heterozygous 30-megabase fragment of Chromosome 15. This large fragment encompasses 330 genes and is integrated on the long arm of Chromosome 19. Here, we employ a CRISPR/Cas9-based genome engineering strategy to excise this sizeable chromosomal fragment and to efficiently and reproducibly derive clones that retain their haploid state. Importantly, spectral karyotyping and single-nucleotide polymorphism (SNP) genotyping revealed that engineered-HAPloid (eHAP) cells are fully haploid with no gross chromosomal aberrations induced by Cas9. Furthermore, whole-genome sequence and transcriptome analysis of the parental HAP1 and an eHAP cell line showed that transcriptional changes are limited to the excised Chromosome 15 fragment. Together, we demonstrate the feasibility of efficiently engineering megabase deletions with the CRISPR/Cas9 technology and report the first fully haploid human cell line.

[Supplemental material is available for this article.]

Most eukaryotic organisms are diploid, inheriting one genome copy from each parent. Deleterious mutations present in one copy are usually buffered by the other copy. This fail-safe mechanism represents a major complication for the study of gene function in most eukaryotes: When one allele is altered, the phenotypic consequences of this alteration are often masked by the other intact allele.

In vertebrates, haploidy—the presence of a genome in a single copy—is naturally confined to the stage of the gametes. Experimentally, however, haploid somatic cells can be derived from a number of organisms including medaka (Yi et al. 2009), mouse (Elling et al. 2011; Leeb and Wutz 2011) and rat (Li et al. 2013). In humans, near-haploid somatic cells have been found in certain tumors including leukemias (Oshimura et al. 1977; Andersson et al. 1995) and chondrosarcomas (Bovee et al. 1999). Importantly, a near-haploid human cell line was isolated from a chronic myeloid leukemia patient and stably cultured over several months (Kotecki et al. 1999). This cell line, referred to as KBM-7, contains one copy of most chromosomes with the exception of chromosome 8 and a portion of Chromosome 15, which are disomic.

Near-haploidy of KBM-7 cells has been exploited to perform large-scale loss-of-function screens in human cells (Carette et al. 2009). Such screens have been employed to study a variety of processes in KBM-7 cells, ranging from host–pathogen interactions to signaling and drug mechanism of action (Carette et al. 2011a,b; Birsoy et al. 2013; Jae et al. 2013). In addition, KBM-7 cells were used to assemble a large library of human isogenic cell lines (Burckstummer et al. 2013), thereby enabling both forward and reverse genetics experiments in human cells.

KBM-7 cells can be reprogrammed to induced pluripotent stem cells by overexpression of *POU5F1*, *SOX2*, *KLF4*, and *MYC* (Takahashi and Yamanaka 2006; Carette et al. 2010). These reprogramming experiments also yielded a near-haploid cell line with a fibroblast-like morphology termed HAP1 (Carette et al. 2011b). In contrast to KBM-7 cells, HAP1 cells are adherent and lack the second copy of Chromosome 8. However, HAP1 cells are not fully haploid as they retain two copies of a fragment of Chromosome 15, one of which is fused to Chromosome 19.

Cas9 is an endonuclease that was first isolated from *Streptococcus pyogenes*. It can be programmed by short guide RNAs (gRNAs) to cleave any genomic locus that is complementary to the gRNA and followed by a protospacer adjacent motif (PAM; NGG for *Streptococcus pyogenes* Cas9) (Cho et al. 2013; Cong et al. 2013; Jinek et al. 2013; Mali et al. 2013b). Cleavage of genomic DNA by Cas9 triggers endogenous repair mechanisms such as nonhomologous end joining (NHEJ) that lead to imprecise repair of the breakpoint, thereby causing mutations at the specific locus of interest. Cas9-mediated genome engineering has been used in a variety of organisms from yeast to man (for a review, see Mali et al. 2013a). Importantly, Cas9 has also been used to delete loci of interest (Xiao et al. 2013), thereby offering enticing perspectives for synthetic biology. However, until now, only kilobase-size deletions have been reported (Fujii et al. 2013; Horii et al. 2013; Xiao et al. 2013), raising the question if larger deletions can be engineered with high enough efficiency to obtain genetically modified clones. In this study, we employed the versatile CRISPR/Cas9 system to delete one copy of the disomic portion of Chromosome 15 that is present in

[4]Present address: Department of Cell and Molecular Biology, Karolinska Institutet, 17177 Stockholm, Sweden
Corresponding author: tibu@haplogen.com
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.177220.114.

© 2014 Essletzbichler et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

HAP1 cells. The deleted fragment encompasses 30 million base pairs, encoding 330 human genes, and corresponds to roughly a third of the total size of Chromosome 15.

Results

Spectral karyotyping of HAP1 cells revealed that the Chromosome 15 fragment present in the parental KBM-7 cell line is retained in HAP1 (Fig. 1A; Supplemental Fig. 1). It is fused to the long arm of Chromosome 19. To map the boundaries of the disomic region on Chromosome 15, we analyzed single-nucleotide polymorphism (SNP) array data from KBM-7 cells (Burckstummer et al. 2013). These data indicate that the disomic fragment encompasses almost 30 million base pairs (from around Chr 15: 61,105,000 to around Chr 15: 89,890,000) (Fig. 1B). The presence of heterozygous SNPs from this region shows that the disomic region did not arise by duplication. Instead, it represents a remnant of its diploid heterozygous origin.

To generate a fully haploid somatic human cell line, we aimed to eliminate one copy of the disomic region from Chromosome 15 by CRISPR/Cas9-mediated genome engineering. We reasoned that coapplication of two gRNAs positioned at the boundaries of the disomic region would result in the simultaneous cleavage of both ends by Cas9, leading to the elimination of the intervening fragment. As the precise sequence of the Chromosome 19/15 fusion was not known, we opted for gRNAs cleaving within the disomic region from Chromosome 15. We designed four gRNAs (Fig. 1C), two for each end. Cas9 and the gRNA expression plasmids were introduced by transient transfection of HAP1 cells. As transfection efficiency was ~10% (Supplemental Fig. 2), we included a plasmid bearing a blasticidin-resistance gene and eliminated untransfected cells using blasticidin selection. Following transient transfection of HAP1 cells, we monitored Cas9 cleavage at the four loci using the T7 endonuclease assay (Kim et al. 2009). Cleavage events could occur either on the Chromosome 15/19 fusion or on the authentic Chromosome 15. We found that gRNAs 1, 3, and 4 elicited efficient cleavage at the designated loci, while gRNA 2 did not cause any detectable genome modification (Fig. 2A; Supplemental Fig. 3). Next, we assessed whether the deletion could be detected in the pool of transfected cells. We designed two primer pairs to flank the 30 million base pair region from Chromosome 15 that we aimed to delete. As expected, no PCR band was detectable in wild-type HAP1 cells. However, we readily detected a PCR product in the sample in which gRNAs 1 and 3 had been combined (Fig. 2B), indicating that the DNA ends exposed by double-strand breaks had been joined together.

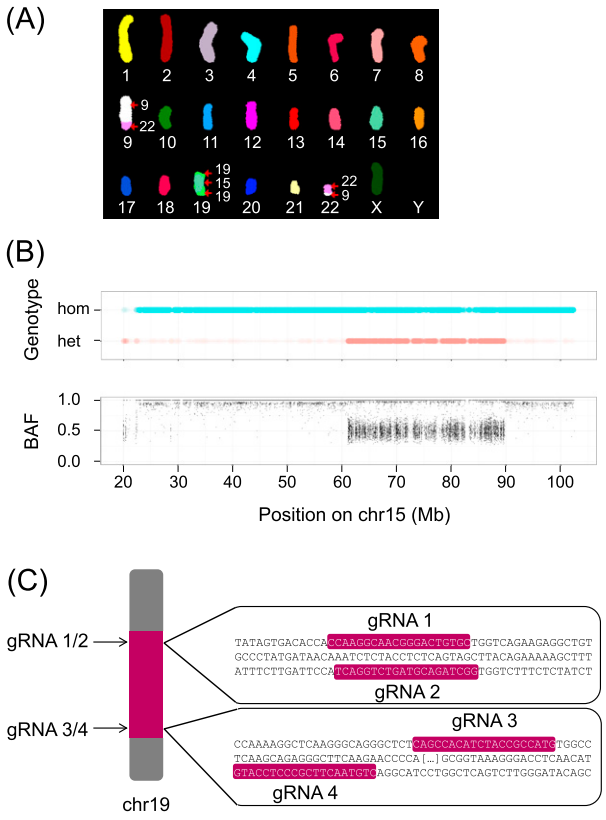


Figure 1. Genomic makeup of HAP1 and strategy for deletion of the Chromosome 15 fragment. (*A*) HAP1 cells were subjected to spectral karyotyping to characterize their chromosomal landscape and identify disomic regions. (*B*) Single-nucleotide polymorphism data from KBM-7 cells were analyzed to identify chromosomal segments that are heterozygous and hence disomic. (*Top*) Heterozygous SNPs are depicted in red; homozygous SNPs, in turquoise. (*Bottom*) B-allele frequency (BAF) on Chromosome 15. A BAF of 0.5 is indicative of heterozygosity. (*C*) Schematic representative of the gRNA design. Two gRNAs (colored in pink) were placed at the boundaries of the disomic region from Chromosome 15.

While it was encouraging to detect the deletion in the pool of transfected cells, it may still represent a very rare event, complicating the isolation of a clonal cell line carrying the deletion. We isolated clones from the sample in which gRNAs 1 and 3 had been combined and assessed the presence of the deletion by PCR. A first batch of 200 clones contained four clones (designated A8, A11, B11, and F6) in which the deletion was detectable by PCR (Supplemental Fig. 4). A second batch of 200 clones that was independently derived contained one additional such clone (designated E9) (Fig. 3A). This indicates that deletions or rearrangements occurred in ~1% of the transfected cells.

To exclude the possibility that the excised region of Chromosome 15 had been retained by integrating somewhere else in the genome, we analyzed heterozygous SNPs. We observed that clones A8, B11, and F6 were still heterozygous for the given SNPs (Supplemental Fig. 5), while clones A11 and E9 were homozygous (Fig. 3B). This indicates that proper deletion only occurred in clones A11 and E9. Sanger sequencing of the deletion PCR products obtained from clones A11 and E9 showed different breakpoints in the two clones (Supplemental Fig. 6), demonstrating that these clones originated from independent editing events.

To characterize the chromosomal landscape in the various clones, we subjected each clone to spectral karyotyping. Clones A11 and E9 were found to be fully haploid and Chromosome 19 was intact (Fig. 3C), indicating that the defective Chromosome 19/15 fusion had been repaired as intended. These two clones can therefore be considered fully haploid human cell lines. In contrast, clones A8, B11, and F6 were found to be diploid (Supplemental Fig. 7), presumably as a consequence of endoreduplication prior to the editing event. In these clones we observed a deletion on one allele of the "authentic" Chromosome 15, while the Chromosome 19/15 fusion was unaffected. This explains why these clones were positive in the deletion PCR (the signal originated from the authentic Chromosome 15 that now bears a deletion), while we did not see loss-of-heterozygosity (both the authentic Chromosome 15 and the Chromosome 19/15 fusion were still present). Based on these observations, we focused our future efforts on clones A11 and E9 and decided to name them engineered HAPloid (eHAP) E9 and eHAP A11.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

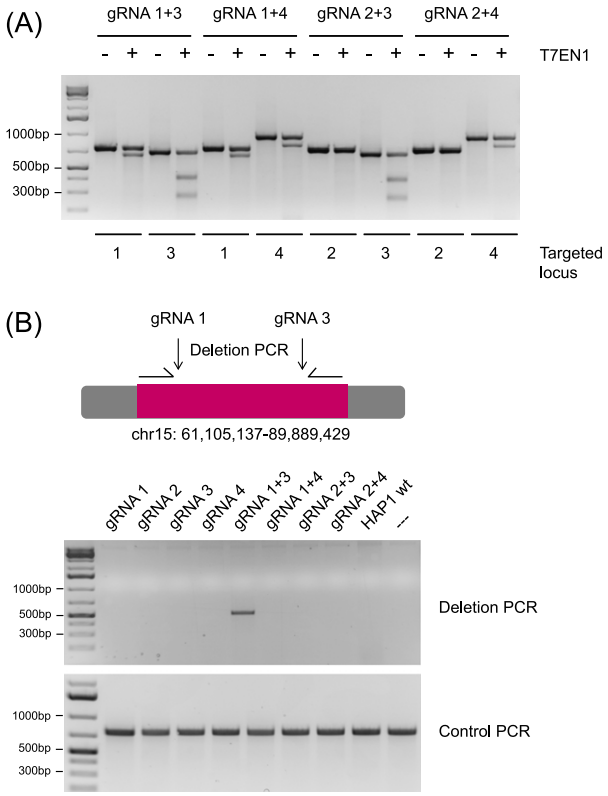


Figure 2. Combination of gRNAs 1 and 3 causes deletion of the Chromosome 15 fragment. HAP1 cells were transfected with various combinations of gRNAs (as indicated). Around 10 d post transfection, genomic DNA was isolated from pools of transfected cells. (*A*) The regions targeted by individual gRNAs were amplified by PCR using suitable primer pairs. Digestion of these PCR products by T7 endonuclease provides a semiquantitative measure for Cas9 editing efficiency. (*B*) To assess whether the fragment between gRNAs 1 and 3 had been excised following Cas9 cleavage, we performed a deletion PCR using a forward primer (HG6090) that binds to position Chr 15: 61,105,055 and a reverse primer (HG6093) that binds to position Chr 15: 89,889,818. We also included a control PCR (primer pair HG6090/HG6091) to confirm that every sample contained genomic DNA, suitable for PCR.

As HAP1 cells can spontaneously convert to a diploid state, we assessed the stability of the clones in which deletion had occurred. Propidium iodide staining showed that both clones eHAP A11 and eHAP E9 were haploid (Supplemental Fig. 8) and indistinguishable from haploid HAP1 cells. When passaged for 20 passages, HAP1 cells as well eHAP E9 retained perfect haploidy, while eHAP A11 had partially converted to diploidy, possibly by endoreduplication. This indicates that fully haploid human cells are viable and can stably maintain their haploid karyotype.

We decided to further characterize eHAP A11 and eHAP E9 by performing whole-genome sequencing, and we included the parental HAP1 cells in these analyses. Whole-genome sequencing was performed at ~20× coverage in HAP1 and ~6× coverage in eHAP A11 and eHAP E9, respectively. A global comparison between the three cell lines showed similar relative coverage across the genome, with the exception of the fragment from Chromosome 15, for which HAP1 cells showed twice the coverage in comparison to eHAP E9 (Fig. 4A; Supplemental Fig. 9). This observation supports the notion that the complete disomic region from Chromosome 15 had been deleted in eHAP A11 and eHAP E9 cells. Inspection of the gRNA target sites on Chromosome 15 revealed that both eHAP A11 and eHAP E9 may harbor small indels (eHAP A11: 13-bp deletion at 61M breakpoint and 150-bp deletion at 89M breakpoint; eHAP E9: 4-bp deletion at 89M breakpoint), presumably arising from Cas9 cleavages that were subsequently repaired by NHEJ. As the gRNAs target noncoding regions on Chromosome 15, these minor alterations are unlikely to affect cellular physiology. We also analyzed putative CRISPR/Cas9 off-target sites for each of the two gRNAs in the two clones (Supplemental Table 1). Importantly, we did not find any evidence for indels (data not shown), indicating that off-target cleavage during the genome editing process was negligible. Global analysis of single-nucleotide substitutions and short indels revealed that the majority of the single-nucleotide substitutions that were lost in the eHAP clones with regard to their HAP1 parent could be attributed to the Chromosome 15 fragment that had been deleted (Supplemental Figs. 10, 11). Both clones also acquired some mutations due to genetic drift (Supplemental Figs. 10, 11). Overall, the whole-genome sequencing confirmed the data obtained by spectral karyotyping, highlighting the deletion of the chromosome 15 fragment and suggesting only minor additional alterations in these cells.

Finally, we compared near-haploid HAP1 cells to the fully haploid eHAP A11 and E9 cells by RNA sequencing. We also included KBM-7 cells as a control. To control for variability due to culturing conditions, we included two replicates derived from parallel cultures. Remarkably, the Spearman's correlations between HAP1 and eHAP E9 were just as strong as those between replicates of the same cell line (Fig. 4B). This indicates that the global expression profile of HAP1 and eHAP E9 cells is nearly identical. Of note, the number of genes that differed between HAP1 and eHAP E9 by at least twofold was around 600, whereas this number was over 3000 genes compared to KBM-7 cells (Fig. 4C). Furthermore, the majority of genes that were differentially expressed between HAP1 and eHAP E9 clustered on the deleted fragment of chromosome 15 (Fig. 4D). Altogether, our RNA sequencing data indicate no major differences between eHAP A11 or eHAP E9 and HAP1 and shows mainly differences arising from the large chromosomal deletion we engineered.

Discussion

We present here a megabase-scale deletion, engineered by CRISPR/Cas9, that greatly exceeds published deletions (Fujii et al. 2013; Horii et al. 2013; Xiao et al. 2013), thus demonstrating the feasibility of chromosome-scale genome engineering. Importantly, engineered chromosomally stable clones could be obtained using a standard subcloning and PCR screening procedure. Large deletions up to 24 megabases have been reported using zinc finger nucleases (Lee et al. 2010, 2012) or TALENs (Gupta et al. 2013; Kim et al. 2013; Xiao et al. 2013), but the efficiency cannot always be inferred because few studies reported the isolation of single clones carrying these deletions. In cases when single clones were isolated, they were retrieved at a surprisingly high frequency (~0.5%) (Kim et al. 2013), given the fact that TALENs are generally believed to be less efficient than the CRISPR/Cas9 system. When finalizing this manuscript, we noticed several recent reports showing large deletions (Canver et al. 2014) or chromosomal rearrangements (Choi and Meyerson 2014) as a consequence of paired CRISPR/Cas9 cleavage. Efficiencies reported for monoallelic, megabase-scale deletions were in the range of 1% (Canver et al. 2014) and are thus comparable to our findings.

We based our excision strategy on available SNP array data, which depend on the presence of heterozygous SNPs and thus provide a lower resolution than whole-genome sequencing. The

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

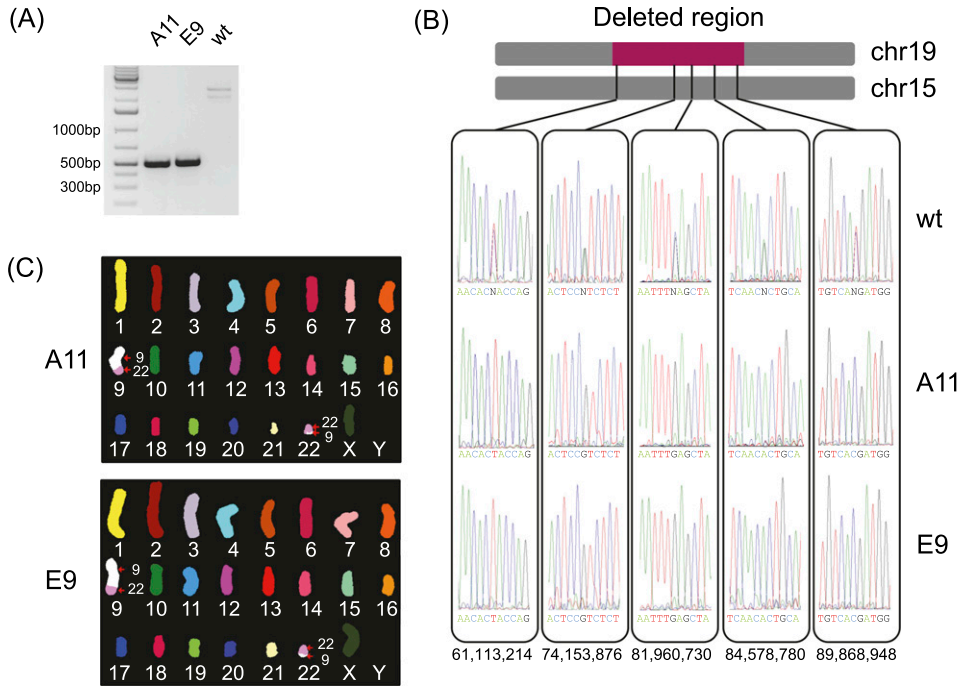


Figure 3. Clones A11 and E9 are fully haploid human cell lines. (*A*) Deletion PCR (for experimental details, see Fig. 2B) for clones A11 and E9, as well as HAP1 wild-type cells. (*B*) To assess loss-of-heterozygosity in clones A11 and E9, we isolated genomic DNA and selected five genomic loci containing SNPs that were heterozygous in HAP1 cells. Each locus was amplified by suitable PCR primers and the PCR products were sent for Sanger sequencing. (*C*) Clones A11 and E9 were analyzed by spectral karyotyping to assess the global genomic landscape of these clones.

whole-genome sequencing data we obtained in the meantime revealed that the disomic region from Chromosome 15 comprises the following positions: Chr 15: 61,103,219–89,893,074. As a consequence, we did not eliminate the Chromosome 15 fragment completely. In fact, in eHAP cells, Chromosome 19 still contains a fragment of a few kilobases from Chromosome 15 (~2 kb from around Chr 15: 61M and ~4 kb from around Chr 15: 89M).

We noticed that only one of our gRNA pairs (gRNAs 1 and 3) gave rise to clones bearing the deletion on Chromosome 15, while at least one other combination of gRNAs (gRNAs 1 and 4) was effective as judged from the T7 endonuclease assay and failed to induce the deletion (Fig. 2A,B). The primer pair that was used to detect the deletion induced by gRNAs 1 and 3 is incompatible with the deletion induced by gRNAs 1 and 4 because gRNA 4 lies downstream from the reverse primer. However, we tried one additional primer pair and failed to detect deletion induced by gRNAs 1 and 4 (data not shown). We speculate that ligation of the ends liberated by gRNAs 1 and 4 was somehow unfavorable, possibly due to missing microhomologies that promote NHEJ (Brissett et al. 2007).

One major concern with the use of the CRISPR/Cas9 system is the cleavage at undesired sites that are closely related to the on-target site, and several recent reports highlight the possibility of off-target editing (Fu et al. 2013; Hsu et al. 2013; Pattanayak et al. 2013; Cho et al. 2014). However, the frequencies at which off-target editing is observed vary from low (Cho et al. 2014; Veres et al. 2014) to considerable (Fu et al. 2013; Hsu et al. 2013; Kuscu et al. 2014). Some of these apparent differences may be attributed to the expression level of Cas9 that varies between different cell types, depending on transfection efficiency (Pattanayak et al. 2013). Other such differences may be attributed to the method of detection. We observe very limited, if any, off-target editing in eHAP A11 or eHAP E9 cells. However, of note, we observe some additional chromosomal aberrations in clones A8, B11, and F6. Many of these alterations involve Chromosome 13 and pieces thereof. The fact that the gRNAs that were used for the deletion also have off-target sites on Chromosome 13 (Supplemental Table 1) may suggest that these alterations are the consequence of off-target editing.

The use of the CRISPR/Cas9 system for deletion of specific genomic regions of interest paves the way for the functional characterization of promoters, enhancers, and other regulatory regions in the human genome. Moreover, the inactivation of entire gene clusters by deletion will enable the study of gene families composed of individual members with redundant functions. An enticing application of this technology is the creation of a minimal-essential genome that is sufficient for a human cell to survive and proliferate in culture.

Moreover, we present the first fully haploid human cell line that we term eHAP. Although haploid cells from other organisms had been isolated previously (Yi et al. 2009; Elling et al. 2011; Leeb and Wutz 2011; Li et al. 2013), only near-haploid human cell lines had been reported. Nonetheless, the near-haploid cell lines KBM-7 and HAP1 have demonstrated the value of subdiploid karyotypes for functional genomics (Carette et al. 2009, 2011a). The ploidy of a cell is a critical determinant for the success of any genome editing technology: Unlike in diploid or polyploid cell lines, editing events can be easily traced by PCR and Sanger sequencing. Indeed, the modal gene copy number of frequently used cell lines such as HeLa and A549 far exceeds that of a diploid cell line, which may severely hamper genome editing. Thus, from a genome engineering and functional genomics perspective, haploid cells are highly attractive. In line with this consideration, one of the first genome-wide CRISPR/Cas9 screens was conducted in KBM-7 cells (Wang et al. 2014). We therefore anticipate that this fully haploid human cell line will become the workhorse for genome engineering and screening in the broader scientific community.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

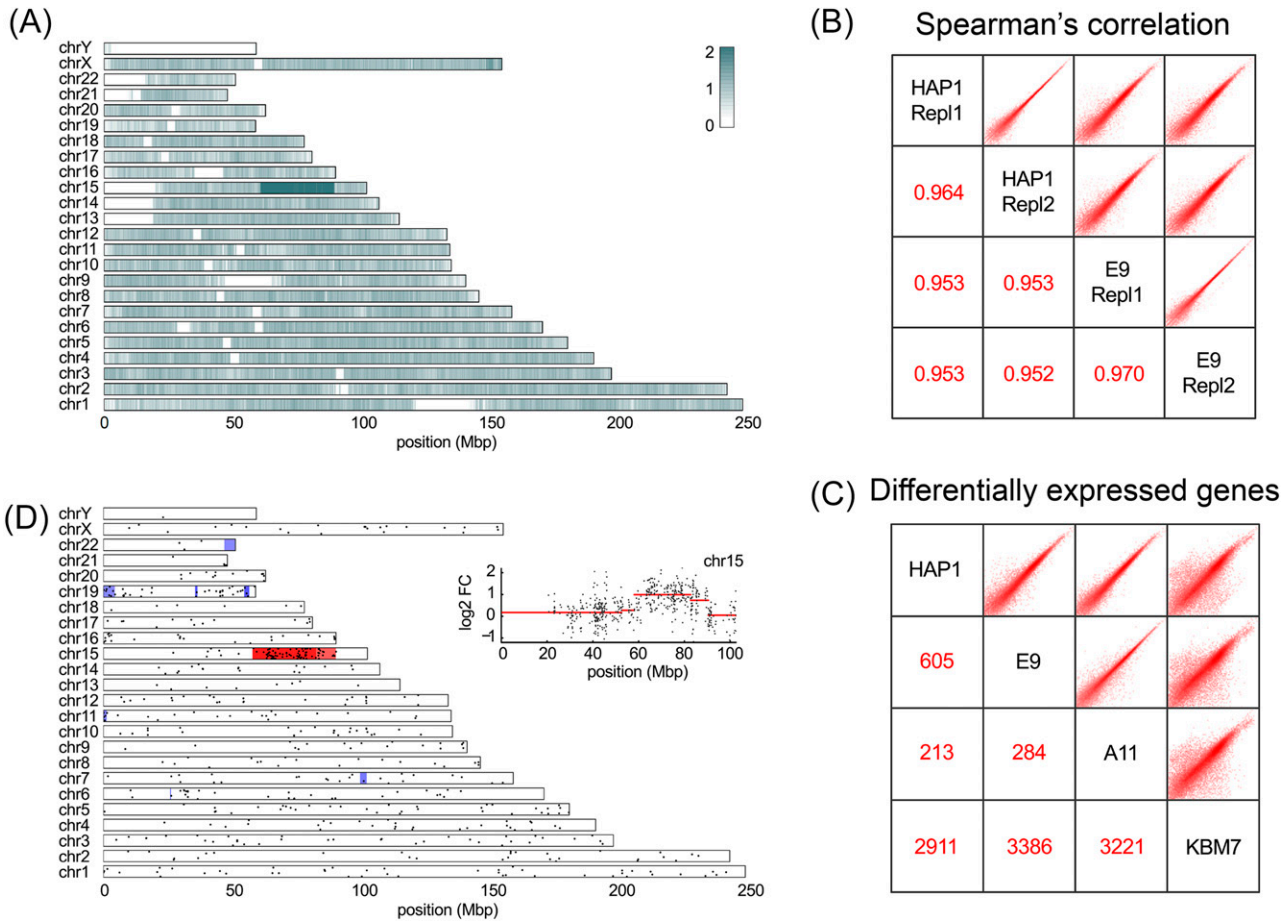


Figure 4. Genomic and transcriptomic changes in eHAP cells are largely confined to Chromosome 15. (*A*) Whole-genome sequencing was performed on parental HAP1 cells and the A11 and E9 eHAP clones. In this panel, relative coverage between HAP1 and E9 data reveals a copy number loss restricted to the edited Chr 15 fragment. Large white regions correspond to unassembled pieces of the human genome. (*B*) Two biological replicates of HAP1 cells and two technical replicates of the E9 clone were subjected to RNA sequencing. Spearman's correlation between the samples shows that overall expression is consistent between the parental line and the edited clones. (*C*) Two replicates of each cell line were compared pairwise. The number of highly expressed (FPKM > 5) and twofold differentially expressed genes are indicated. (*D*) Expression ratios between HAP1 and E9 cells were subjected to segmentation analysis. Red/blue bands show segments for which the expression level is consistently higher/lower in the parental cell line. (*Inset*) Details of the segmentation.

Methods

KBM-7 cells and HAP1 cells were cultured in IMDM supplemented with 10% FCS. Cells were passaged every 48 h. A microscopic image of the two cell lines is shown in Supplemental Figure 1.

Spectral karyotyping was performed and analyzed by WiCell Cytogenetics. In brief, metaphase chromosome spreads were prepared on slides according to standard cytogenetic procedures. These slide preparations were then hybridized according to the DNA spectral karyotyping hybridization and detection protocol provided by Applied Spectral Imaging (ASI) using the supplied SKY probes and detection system. Images were captured and analyzed using the HiSKYV spectrum imaging system from ASI.

Selection of SNPs

Heterozygous SNPs were selected from the SNP list published in Burckstummer et al. (2013) derived from whole-genome sequencing of the KBM-7 parental cell line. Equal spacing of the SNPs and visualization of the genotypes was performed by custom R scripts (R version 3.0.1; package "ggplot2") based on the cumulative variant list from both whole-genome and exome sequencing experiments (see also Burckstummer et al. 2013).

CRISPR/Cas9-mediated genome engineering

The sequences that were selected as gRNAs can be found in Table 1. HAP1 cells were transiently transfected with expression plasmids for *Streptococcus pyogenes* Cas9 and suitable gRNAs using TurboFectin (OriGene) according to manufacturer's instructions. To enrich for transfected cells, we cotransfected a plasmid encoding a blasticidin-resistance gene and subjected cells to transient selection with 20 μg/mL blasticidin. Transfected cells were expanded either for genomic DNA isolation or for limiting dilution.

T7 endonuclease assay

To assess Cas9-mediated genome editing at specific loci, we expanded cells and isolated DNA using the QIAamp DNA mini kit

Table 1. Sequences selected as gRNAs

Guide RNA	Genomic location	Sequence (5′-3′)
gRNA1	Chr 15: 61,105,137	CCAAGGCAACGGGACTGTGC
gRNA2	Chr 15: 61,105,238	TCAGGTCTGATGCAGATCGG
gRNA3	Chr 15: 89,889,429	CAGCCACATCTACCGCCATG
gRNA4	Chr 15: 89,889,982	GTACCTCCCGCTTCAATGTC

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

Table 2. Primer pairs used for the T7 endonuclease assay

PCR primers	Genomic location	Sequence (5′ to 3′)
HG6090	Chr 15: 61,105,055 fwd	AGGTGAACATCATCCAGAAGGGGCA
HG6091	Chr 15: 61,105,783 rev	CAATGGCAGGCTGATGGGAAGATGG
HG6092	Chr 15: 89,889,153 fwd	AACTGTTATCTGCCACCCTGGTCCC
HG6093	Chr 15: 89,889,818 rev	AAGCAGGCACAAAACCAAAGCCTCT
HG6094	Chr 15: 89,889,174 fwd	TCCCTGGGACAGAGTTAAAGTGGCA
HG6095	Chr 15: 89,890,119 rev	AAAATAAGGGGACCGGACAGGCTGG

(Qiagen). For each of the four gRNAs, we designed one specific primer pair to amplify the edited locus using GoTaq polymerase (Promega) according to manufacturer's instructions (Table 2).

PCR products were used for the T7 endonuclease assay as described previously (Kim et al. 2009). In brief, PCR products were subjected to another round of denaturation (for 2 min at 94°C) and annealing (94°C to 85°C at 2°C per second, 85°C to 25°C at 0.1°C per second), followed by digestion with the mismatch-sensitive T7 endonuclease I (NEB). Digested products were resolved on a 2% agarose gel.

Deletion PCR

To assess the deletion of the fragment encompassing Chr 15: 61,105,002–89,890,003, we isolated genomic DNA using the QIAamp DNA mini kit (Qiagen) and subjected it to PCR using GoTaq polymerase (Promega) and the oligonucleotides HG6090 and HG6093 (see sequences above).

Isolation of single-cell clones

Single HAP1 clones were obtained by limiting dilution. To this end, cells were trypsinized and serially diluted to a concentration of 15 cells per mL. Fifty microliters of this suspension were seeded in each well of a 384-well plate. Individual wells were inspected by microscopy to exclude polyclonal cell lines. Monoclonal cell lines were expanded.

PCR and Sanger sequencing to detect heterozygous SNPs

The primer pairs that were used to amplify heterozygous SNPs from HAP1 cells (or corresponding clones) using GoTaq polymerase (Promega) can be found in Table 3. Each forward primer contained an M13 primer binding site labeled in bold. PCR products were purified and directly subjected to Sanger sequencing using the M13 sequencing primer (TGTAAAACGACGGCCAG).

Table 3. Primer pairs used to amplify heterozygous SNPs from HAP1 cells using GoTaq polymerase

PCR primers	Genomic location	Sequence
HG5740	Chr 15: 61,113,214	**TGTAAAACGACGGCCAG**GCTGCTTAAAATGCAGATTCCAAAG
HG5741	Chr 15: 61,113,214	GGAGAGATAGGGGATAAAAGTGGTG
HG6690	Chr 15: 74,153,876	**TGTAAAACGACGGCCAG**AGAAGGATTAAGCTCCACTTACCAA
HG6691	Chr 15: 74,153,876	TGAAAGGTATGGGAGAATGTAACAA
HG6696	Chr 15: 81,960,730	**TGTAAAACGACGGCCAG**TGCAGATGAAAGTTGTTTGCCTAAT
HG6697	Chr 15: 81,960,730	GGAGGGATAGCATTGGGAGATATAC
HG6698	Chr 15: 84,578,780	**TGTAAAACGACGGCCAG**ACTTAGTCCCTTAAAATTGTGCTTCA
HG6699	Chr 15: 84,578,780	ATACATTGAAAGAGCCATGTCTGGG
HG5742	Chr 15: 89,868,948	**TGTAAAACGACGGCCAG**CAGACTCTTGAACCCAAACTCTTTC
HG5743	Chr 15: 89,868,948	GGACTGACTTAGTGTCTTTGCTTTT

The M13 primer binding sites are shown in bold.

Propidium iodide staining

Cells were treated with 100 ng/mL KaryoMax (Gibco) for 6 h or left untreated, harvested by trypsinization, and washed twice with PBS. Cells were simultaneously lysed and stained using Nicoletti buffer (0.1% sodium citrate, 0.1% Triton X-100, 0.5 unit/mL RNase A, 20 units/mL RNase T1, 50 μg/mL propidium iodide). Haploid and diploid reference cell lines were included as controls. Propidium iodide staining was quantified by flow cytometry.

Whole-genome sequencing and analysis

Genomic DNA was subjected to library preparation for whole-genome sequencing using Illumina TruSeq DNA PCR-free sample preparation kit according to manufacturer's instructions. Libraries were sequenced on an Illumina HiSeq 2000 using paired-end, 100-bp-read chemistry. We aligned the data onto the human genome (hg19) using Bowtie 2 (Langmead and Salzberg 2012) and analyzed variants using Bamformatics (sourceforge.net/projects/bamformatics) and custom R scripts (R Core Team 2013). We checked for artifactual alignments using secondary alignments produced by GSNAP (Wu and Nacu 2010).

RNA sequencing and analysis

Total RNA (1 μg) was subjected to library preparation using an Illumina TruSeq RNA sample preparation kit according to manufacturer's instructions. DNA libraries were sequenced on an Illumina HiSeq 2000 using 50-bp, single-read chemistry. We aligned reads with TopHat (Trapnell et al. 2012) using GENCODE (V19) gene annotations (Harrow et al. 2012). We evaluated expression on genes by counting reads and weighting by mapping multiplicity. For over/under expression analysis, we considered genes with a minimum expression level (FPKM 5), a fold change of two, and clearance on the uncertainty levels. For the segmentation analysis, we computed the logarithm of fold changes for all expressed genes (FPKM 0.2), arranged them in chromosomal order, and then applied piecewise-flat segmentation (Tarabichi et al. 2012). The segmentation values were then used to compile a genome-wide amplification map.

Data access

Whole-genome sequencing data from this study have been submitted to the NCBI Sequence Read Archive (SRA; http://www.ncbi.nlm.nih.gov/sra) under accession numbers SRP044390 (HAP1) and SRP044387 (eHAP A11 and E9). RNA sequencing data have also been submitted to the SRA under accession number SRP044391. All cell lines mentioned in the manuscript (KBM-7, HAP1, eHAP A11, and eHAP E9) can be obtained from Haplogen Genomics (Vienna, Austria).

Competing interest statement

The authors declare that Haplogen has filed a patent application (EP 13194940.6.) describing the generation of a fully haploid somatic human cell line.

Acknowledgments

We thank the entire Haplogen team for helpful suggestions and discussions.

Downloaded from genome.cshlp.org on November 20, 2026 . Published by Cold Spring Harbor Laboratory Press

We thank Wolfgang Fischl for critical reading of the manuscript. T.K. is supported by an ERC Starting grant (ERC-StG-2012-311166) and B.G. by a Boehringer Ingelheim fellowship. We thank the Biomedical Sequencing Facility (BSF) for whole-genome and RNA sequencing.

References

Andersson BS, Collins VP, Kurzrock R, Larkin DW, Childs C, Ost A, Cork A, Trujillo JM, Freireich EJ, Siciliano MJ, et al. 1995. KBM-7, a human myeloid leukemia cell line with double Philadelphia chromosomes lacking normal c-ABL and BCR transcripts. *Leukemia* **9:** 2100–2108.

Birsoy K, Wang T, Possemato R, Yilmaz OH, Koch CE, Chen WW, Hutchins AW, Gultekin Y, Peterson TR, Carette JE, et al. 2013. MCT1-mediated transport of a toxic molecule is an effective strategy for targeting glycolytic tumors. *Nat Genet* **45:** 104–108.

Bovee JV, Cleton-Jansen AM, Kuipers-Dijkshoorn NJ, van den Broek LJ, Taminiau AH, Cornelisse CJ, Hogendoorn PC. 1999. Loss of heterozygosity and DNA ploidy point to a diverging genetic mechanism in the origin of peripheral and central chondrosarcoma. *Genes Chromosomes Cancer* **26:** 237–246.

Brissett NC, Pitcher RS, Juarez R, Picher AJ, Green AJ, Dafforn TR, Fox GC, Blanco L, Doherty AJ. 2007. Structure of a NHEJ polymerase-mediated DNA synaptic complex. *Science* **318:** 456–459.

Burckstummer T, Banning C, Hainzl P, Schobesberger R, Kerzendorfer C, Pauler FM, Chen D, Them N, Schischlik F, Rebsamen M, et al. 2013. A reversible gene trap collection empowers haploid genetics in human cells. *Nat Methods* **10:** 965–971.

Canver MC, Bauer DE, Dass A, Yien YY, Chung J, Masuda T, Maeda T, Paw BH, Orkin SH. 2014. Characterization of genomic deletion efficiency mediated by CRISPR/Cas9 in mammalian cells. *J Biol Chem* **289:** 21312–21324.

Carette JE, Guimaraes CP, Varadarajan M, Park AS, Wuethrich I, Godarova A, Kotecki M, Cochran BH, Spooner E, Ploegh HL, et al. 2009. Haploid genetic screens in human cells identify host factors used by pathogens. *Science* **326:** 1231–1235.

Carette JE, Pruszak J, Varadarajan M, Blomen VA, Gokhale S, Camargo FD, Wernig M, Jaenisch R, Brummelkamp TR. 2010. Generation of iPSCs from cultured human malignant cells. *Blood* **115:** 4039–4042.

Carette JE, Guimaraes CP, Wuethrich I, Blomen VA, Varadarajan M, Sun C, Bell G, Yuan B, Muellner MK, Nijman SM, et al. 2011a. Global gene disruption in human cells to assign genes to phenotypes by deep sequencing. *Nat Biotechnol* **29:** 542–546.

Carette JE, Raaben M, Wong AC, Herbert AS, Obernosterer G, Mulherkar N, Kuehne AI, Kranzusch PJ, Griffin AM, Ruthel G, et al. 2011b. Ebola virus entry requires the cholesterol transporter Niemann-Pick C1. *Nature* **477:** 340–343.

Cho SW, Kim S, Kim JM, Kim JS. 2013. Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease. *Nat Biotechnol* **31:** 230–232.

Cho SW, Kim S, Kim Y, Kweon J, Kim HS, Bae S, Kim JS. 2014. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Res* **24:** 132–141.

Choi PS, Meyerson M. 2014. Targeted genomic rearrangements using CRISPR/Cas technology. *Nat Commun* **5:** 3728.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, et al. 2013. Multiplex genome engineering using CRISPR/Cas systems. *Science* **339:** 819–823.

Elling U, Taubenschmid J, Wirnsberger G, O'Malley R, Demers SP, Vanhaelen Q, Shukalyuk AI, Schmauss G, Schramek D, Schnuetgen F, et al. 2011. Forward and reverse genetics through derivation of haploid mouse embryonic stem cells. *Cell Stem Cell* **9:** 563–574.

Fu Y, Foden JA, Khayter C, Maeder ML, Reyon D, Joung JK, Sander JD. 2013. High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells. *Nat Biotechnol* **31:** 822–826.

Fujii W, Kawasaki K, Sugiura K, Naito K. 2013. Efficient generation of large-scale genome-modified mice using gRNA and CAS9 endonuclease. *Nucleic Acids Res* **41:** e187.

Gupta A, Hall VL, Kok FO, Shin M, McNulty JC, Lawson ND, Wolfe SA. 2013. Targeted chromosomal deletions and inversions in zebrafish. *Genome Res* **23:** 1008–1017.

Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22:** 1760–1774.

Horii T, Morita S, Kimura M, Kobayashi R, Tamura D, Takahashi RU, Kimura H, Suetake I, Ohata H, Okamoto K, et al. 2013. Genome engineering of mammalian haploid embryonic stem cells using the Cas9/RNA system. *Peer J* **1:** e230.

Hsu PD, Scott DA, Weinstein JA, Ran FA, Konermann S, Agarwala V, Li Y, Fine EJ, Wu X, Shalem O, et al. 2013. DNA targeting specificity of RNA-guided Cas9 nucleases. *Nat Biotechnol* **31:** 827–832.

Jae LT, Raaben M, Riemersma M, van Beusekom E, Blomen VA, Velds A, Kerkhoven RM, Carette JE, Topaloglu H, Meinecke P, et al. 2013. Deciphering the glycosylome of dystroglycanopathies using haploid screens for lassa virus entry. *Science* **340:** 479–483.

Jinek M, East A, Cheng A, Lin S, Ma E, Doudna J. 2013. RNA-programmed genome editing in human cells. *eLife* **2:** e00471.

Kim HJ, Lee HJ, Kim H, Cho SW, Kim JS. 2009. Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res* **19:** 1279–1288.

Kim Y, Kweon J, Kim A, Chon JK, Yoo JY, Kim HJ, Kim S, Lee C, Jeong E, Chung E, et al. 2013. A library of TAL effector nucleases spanning the human genome. *Nat Biotechnol* **31:** 251–258.

Kotecki M, Reddy PS, Cochran BH. 1999. Isolation and characterization of a near-haploid human cell line. *Exp Cell Res* **252:** 273–280.

Kuscu C, Arslan S, Singh R, Thorpe J, Adli M. 2014. Genome-wide analysis reveals characteristics of off-target sites bound by the Cas9 endonuclease. *Nat Biotechnol* **32:** 677–683.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9:** 357–359.

Lee HJ, Kim E, Kim JS. 2010. Targeted chromosomal deletions in human cells using zinc finger nucleases. *Genome Res* **20:** 81–89.

Lee HJ, Kweon J, Kim E, Kim S, Kim JS. 2012. Targeted chromosomal duplications and inversions in the human genome using zinc finger nucleases. *Genome Res* **22:** 539–548.

Leeb M, Wutz A. 2011. Derivation of haploid embryonic stem cells from mouse embryos. *Nature* **479:** 131–134.

Li W, Li X, Li T, Jiang MG, Wan H, Luo GZ, Feng C, Cui X, Teng F, Yuan Y, et al. 2013. Genetic modification and screening in rat using haploid embryonic stem cells. *Cell Stem Cell* **14:** 404–414.

Mali P, Esvelt KM, Church GM. 2013a. Cas9 as a versatile tool for engineering biology. *Nat Methods* **10:** 957–963.

Mali P, Yang L, Esvelt KM, Aach J, Guell M, DiCarlo JE, Norville JE, Church GM. 2013b. RNA-guided human genome engineering via Cas9. *Science* **339:** 823–826.

Oshimura M, Freeman AI, Sandberg AA. 1977. Chromosomes and causation of human cancer and leukemia. XXIII. Near-haploidy in acute leukemia. *Cancer* **40:** 1143–1148.

Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. 2013. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol* **31:** 839–843.

R Core Team. 2013. *R: a language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. http://www.R-project.org/.

Takahashi K, Yamanaka S. 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126:** 663–676.

Tarabichi M, Detours V, Konopka T. 2012. Piecewise polynomial representations of genomic tracks. *PLoS ONE* **7:** e48941.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc* **7:** 562–578.

Veres A, Gosis BS, Ding Q, Collins R, Ragavendran A, Brand H, Erdin S, Talkowski ME, Musunuru K. 2014. 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.

Wang T, Wei JJ, Sabatini DM, Lander ES. 2014. Genetic screens in human cells using the CRISPR-Cas9 system. *Science* **343:** 80–84.

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* **26:** 873–881.

Xiao A, Wang Z, Hu Y, Wu Y, Luo Z, Yang Z, Zu Y, Li W, Huang P, Tong X, et al. 2013. Chromosomal deletions and inversions mediated by TALENs and CRISPR/Cas in zebrafish. *Nucleic Acids Res* **41:** e141.

Yi M, Hong N, Hong Y. 2009. Generation of medaka fish haploid embryonic stem cells. *Science* **326:** 430–433.

Received April 22, 2014; accepted in revised form September 3, 2014.