

Integration-defective lentiviral vector mediates efficient gene editing through homology-directed repair in human embryonic stem cells

Yebo Wang^{1,2}, Yingjia Wang³, Tammy Chang², He Huang^{1,*} and Jiing-Kuan Yee^{2,*}

¹Bone Marrow Transplantation Center, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang 310003, China, ²Department of Diabetes and Metabolic Diseases Research, Beckman Research Institute, City of Hope National Medical Center, Duarte, CA 91010, USA and ³Department of Pathology, School of Medicine, Hangzhou Normal University, Hangzhou, Zhejiang 310036, China

Received May 12, 2016; Revised October 09, 2016; Editorial Decision October 20, 2016; Accepted October 24, 2016

ABSTRACT

Human embryonic stem cells (hESCs) are used as platforms for disease study, drug screening and cell-based therapy. To facilitate these applications, it is frequently necessary to genetically manipulate the hESC genome. Gene editing with engineered nucleases enables site-specific genetic modification of the human genome through homology-directed repair (HDR). However, the frequency of HDR remains low in hESCs. We combined efficient expression of engineered nucleases and integration-defective lentiviral vector (IDLV) transduction for donor template delivery to mediate HDR in hESC line WA09. This strategy led to highly efficient HDR with more than 80% of the selected WA09 clones harboring the transgene inserted at the targeted genomic locus. However, certain portions of the HDR clones contained the concatemeric IDLV genomic structure at the target site, probably resulted from recombination of the IDLV genomic input before HDR with the target. We found that the integrase protein of IDLV mediated the highly efficient HDR through the recruitment of a cellular protein, LEDGF/p75. This study demonstrates that IDLV-mediated HDR is a powerful and broadly applicable technology to carry out site-specific gene modification in hESCs.

INTRODUCTION

Human pluripotent stem cells (hPSCs), such as embryonic stem cells (ESCs) and induced pluripotent stem cells (iPSCs), are used to establish surrogate models to study development biology and disease pathogenesis in a dish. They can also serve as platforms for drug validation, screen-

ing and cell-based therapy. To facilitate these applications, it is frequently necessary to genetically manipulate the hPSC genome. Recent progress in the development of gene editing technologies has significantly advanced the ability to insert, disrupt or repair a gene in hPSCs. These gene editing tools, including zinc-finger nucleases (ZFNs), transcription activator-like effector nucleases (TALENs) and clustered regularly interspaced short palindromic repeat (CRISPR)/CRISPR-associated (Cas) system, bind their genomic target and create double-strand DNA breaks (DSBs) (1). DSBs instigate DNA repair through non-homologous end-joining (NHEJ), which is error-prone and frequently leads to gene inactivation (2). Alternatively, in the presence of the sister chromatid or an exogenous donor DNA template, DSBs stimulate homology-directed repair (HDR), resulting in targeted genome modification (3). Both NHEJ and HDR are active in nearly all cell types and organisms. HDR occurs only during S and G2 phases whereas NHEJ occurs throughout the entire cell cycle (4). The frequency of HDR is generally lower and more variable than NHEJ, likely caused by the competition with NHEJ for repairing DSBs and specific suppression of HDR in M and early G1 phases to prevent telomere fusion (5). The obstacle to carry out HDR is exacerbated in hPSCs due to the poor efficiency of introducing the donor template. Intrinsic properties unique to hPSCs also contribute to the poor HDR frequency as much higher HDR frequencies were observed in murine ESCs (6). The HDR frequency in hPSCs could also be affected by the activity of the targeted gene. Several studies have shown relatively poor HDR frequencies when a silent gene in hPSCs was targeted when compared with an active gene (7–9). These problems in conjunction with the difficulty to maintain stable hPSCs in culture hamper broad application of hPSCs in basic research and clinical application. Although ongoing efforts have made some progress to boost the HDR frequency in human ESCs with small

*To whom correspondence should be addressed. Tel: +1 6262 037 860; Fax: +1 6263 018 280; Email: jyee@coh.org

Correspondence may also be addressed to He Huang. Tel: +86 136 0571 4822; Fax: +86 8820 8277; Email: huanghe@zju.edu.cn

Disclaimer: The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

molecules of NHEJ inhibitors (10–13), the potential effect of these small molecules on self-renewal and pluripotency of hESCs remains unknown. Data from several studies suggests that the low HDR frequency in hESCs is not likely due to a failure of the engineered nuclease to recognize its genomic target (14,15), since the frequency of NHEJ-induced genetic modifications is significantly higher than that of the HDR-induced genetic modifications with the same nuclease in hESCs. The major limitation seems to be due to the lack of efficient engagement of the HDR-mediated repair pathway in hESCs. Finding strategies to overcome this obstacle and improve the efficiency of precise gene editing in hPSCs therefore pose a major challenge.

To boost the HDR frequency in human cells, several groups used non-integrating viral vectors to deliver the gene editing system and the donor template. Viral vectors are attractive owing to their high efficiency in transducing human cells. Since these vectors are non-integrating, their genomes would not persist in actively proliferating cells. In addition, viral transduction permits fine control over the number of DNA copies that reach the nucleus to mediate HDR. Vectors derived from adeno-associated virus (AAV) have been shown to achieve high efficiencies of gene targeting (up to 34% with selection) with ZFN in human U2OS osteosarcoma line (16). Donor template-carrying AAV6 in conjunction with ZFN mRNAs was shown to mediate HDR at frequencies ranging from 19 to 43% with fetal liver-derived hematopoietic progenitor cells (HPCs) in the absence of selection (17). AAV-mediated donor template delivery in the absence of an engineered nuclease also generated highly efficient HDR (up to 30% with selection) in hESCs and iPSCs (18,19). However, these studies were limited only to active loci in human HPCs or PSCs. Whether silent loci can be edited with the AAV vector at a similar frequency remains unclear. Using adenoviral (Ad) vectors to introduce DSBs, although the overall frequency of getting clones expressing the GFP marker was low, nearly all the GFP⁺ clones were derived from HDR (20). However, Coluccio *et al.* showed in a side-by-side comparison that IDLV, in the presence of ZFN, was more efficient than the Ad vector to mediate HDR in human cells (21). Ad vector mediated HDR without an engineered nuclease was also shown to work in hESCs and iPSCs (22,23). However, long homology arms (>10 kb) were required to mediate such HDR, making it more time consuming for the donor template assembly and its insertion into the Ad vector backbone. Several studies used integration-defective lentiviral vector (IDLV) for gene editing in human cells. IDLV is derived from mutations in the integrase (IN) of human immunodeficiency virus-1 (HIV-1) that impair the ability of HIV-1 to integrate into the host genome. Although IDLV was shown to be more efficient than AAV and Ad vectors to mediate HDR in some studies (20,21), the overall frequency remained relatively low in hPSCs. In studies using IDLVs to deliver both the ZFNs and the donor template, the HDR frequency was ~3.5% in human ESCs and ~0.6% in human iPSCs in the absence of selection (24,25). Since IDLV does not support efficient transgene expression due to epigenetic silencing of the unintegrated vector genome (26), the low HDR frequency in hPSCs may be enhanced by using differ-

ent strategies to increase the expression of the engineered nuclease.

We chose to use nucleofection to introduce the expression plasmids for TALEN pairs and CRISPR/Cas9 into hESCs. The study by Zou *et al.* showed that ZFNs could be efficiently expressed with nucleofection and achieved significant gene correction through HDR in hESCs (27). We found that, under this condition, the donor template delivered via IDLV transduction led to significantly improved HDR frequencies in hESC line WA09. More than 80% of the WA09 clones screened were derived from homologous recombination of the shared homology regions between the donor template and the targeted genomic locus. Importantly, the high HDR frequency with IDLV occurred despite the two genetic loci we tested, the genes encoding hematopoietic-specific Wiskott-Aldrich Syndrome (WAS) protein and motor neuron-specific transcription factor HB9, remained silent in undifferentiated WA09 cells. We also found that highly efficient HDR mediated by IDLV was partly due to the recruitment of the host factor, lens epithelium-derived growth factor p75 splice variant (LEDGF/p75) (28), by the IN protein in IDLV. LEDGF/p75 promotes the repair of cellular DSBs through the endogenous HDR pathway (29). Active recruitment of host HDR proteins by the IDLV IN protein in the pre-integration complex could therefore account for the unique advantage of using IDLV to deliver the donor template for HDR in hESCs. This strategy overcomes a major hurdle in the gene editing technology and shows that IDLV-mediated HDR is a powerful and broadly applicable technology to carry out site-specific gene editing in hESCs.

MATERIALS AND METHODS

Cell culture

HEK293T cells (CRL 3216, ATCC, Manassas, VA, USA) and HEK293 cells (CRL1573, ATCC) were cultured in Dulbecco's Modified Eagle medium (DMEM) supplemented with 10% fetal bovine serum (Hyclone, Logan, UT) and 2 mM GlutaMAX (Life Technologies, Gaithersburg, MD) at 37°C with 5% CO₂ incubation. Human embryonic stem cell line WA09 (WiCell, Madison, WI, USA) and derived HDR clones were maintained in ESC medium consisting of DMEM/F12 (1:1) with 20% Knockout Serum Replacement, 2 mM GlutaMAX, 0.1 mM MEM NEAA, 100 U/ml penicillin, 100 mg/ml streptomycin and 0.1 mM 2-mercaptoethanol (Sigma, St Louis, MO, USA), supplemented with 4 ng/ml basic fibroblast growth factor (bFGF). Cells were cultured on irradiated mouse embryonic fibroblasts (MEF) CF-1 (GlobalStem, Rockville, MD, USA) for regular culture and DR4 MEF for G418 or puromycin selection. WA09 cells were transferred to Matrigel (BD Biosciences, San Jose, CA, USA)-coated dishes and cultured with mTeSR1 (Stemcell Technologies, Vancouver, BC, Canada) for at least one passage prior to nucleofection followed by daily culture medium change.

TALEN and CRISPR/Cas9 design and plasmid construction TAL effector

Nucleotide Targeter 2.0 (<https://tale-nt.cac.cornell.edu/>) was used to identify the TALEN target site. TALEN was assembled and the expression plasmid constructed as described (14,30). The assemble kit was purchased from Addgene (Cambridge, MA) (31). To construct pU6-CRISPR, the fragment containing the U6 promoter and tracrDNA was polymerase chain reaction (PCR) amplified from pX330 (32) (Addgene) and cloned into pBluescript SK (-). To construct the CRISPR expression plasmid, complementary oligonucleotides encoding each sgRNA were annealed and cloned into the BbsI sites in pU6-CRISPR. The design of the sgRNA was according to the recommendation on the Zhang laboratory website (<http://crispr.genome-engineering.org>). To construct the Cas9 expression plasmid, the CBh-hCas9 fragment was PCR amplified from pX330 and cloned into pBluescript SK (-).

Evaluation of the cleavage efficiency of the engineered nuclease

HEK293 cells were grown to 40% confluence in 48-well culture dishes and transfected with 0.14 μg of each TALEN expression plasmid or 0.08 μg pU6-sgRNA together and 0.2 μg pCBh-hCas9 using Lipofectamine 2000 (Invitrogen, San Diego, CA, USA). Forty-eight hours after transfection, the genomic DNA was extracted with Epicentre QuickExtract solution (Epicentre Biotechnologies, Madison, WI, USA). Approximately 4000 genome equivalents of the transfected cells were used in a 25 μl PCR reaction to amplify the region of interest. Primers used for monitoring HDR were WAS-F (AAAGGAAGTTGGGCAGAGGTGAGT), WAS-R (CCATCGATTGTGTGTTGGATGGTCATGGAGGT) and HB9-F (AACGCA GCAAAAAGGCCAAA), HB9-R (ACGCTCGTGACA TAATCCCC). PCR amplification was carried out with Hotstar Taq (Qiagen, Valencia, CA, USA) using the following cycling conditions: 95°C for 15 min for initial denaturation; 35 cycles at 94°C for 30 s, 60°C for 30 s, 72°C for 30 s and a final extension at 72°C for 5 min. The activity of the engineered nuclease was measured using the Surveyor nuclease following the manufacturer's instruction (Transgenomic, Omaha, NE, USA). ImageJ was used to measure the indel efficiency by quantifying the intensity of the band separated by gel electrophoresis. The following formula was used to calculate the percent of indel formation (33): % indel formation = $100 \times [1 - (1 - \text{fraction cleaved})^{1/2}]$; fraction cleaved = $100 \times \text{sum of the cleavage product peaks} / (\text{cleavage products} + \text{parent peak})$.

Construction of donor IDLV plasmids

To construct the donor template-containing IDLV for the *WAS* gene, the *WAS* genomic sequence spanning the nuclease cleavage site was PCR amplified using two primer sets: WASTL-F (CCGGAATTCGTGTGGAGAGGAGATGGGAAAGTT) and WASTL-R (CCGGAATTCATCCATCCAGAGACACAGGGAAG) for the upstream homology arm; WASTR-F (CCATCGATGGTAAGAGTGGATGGAGGAATGAG) and WASTR-R (CCATCGAT

TGTGTGTTGGATGGTCATGGAGGT) for the downstream homology arm. The *neo^R* gene controlled by the phosphoglycerate kinase (pgk) promoter was inserted between the two homology arms to generate pWAS-T. To generate the pHIV7/*WAS* for making IDLV, the 3 kb fragment containing the *neo^R* gene flanked by the two homology arms derived from the *WAS* locus was isolated from pWAS-T by EcoRI and XhoI digestion, blunted with Klenow (New England Biolabs, Ipswich, MA, USA), and cloned into the unique BamHI site in pHIV7. To construct the donor template-containing IDLV for the *HB9* gene, the *HB9* genomic sequence spanning the nuclease cleavage site was PCR amplified using two primer sets: HB9TL-F (TGCTCTAGAGTCTCACCTGAGGC CATT) and HB9TL-R (CCATCGATCTCGAGCTGGGCGCGGGCTGGTGGC) for the upstream homology arm, HB9TR-F (CCATCGATGAGCCCCGCGGC CCAGCAGGTGCGGC) and HB9TR-R (ACGCGTCGACTCTAGAGTTTGAACGCTCGTGACATAATC) for the downstream homology arm. The *pur^R* gene encoding the puromycin N-acetyl-transferase controlled by the pgk promoter was inserted between the two homology arms to generate pHB9-T. To generate the pHIV7/*HB9*-T for making IDLV, the 2.48 kb fragment containing the *pur^R* gene flanked by the two homology arms derived from the *HB9* locus was isolated from pHB9-T by XbaI digestion, blunted with Klenow, and cloned into the unique BamHI site in pHIV7.

IDLV production and IDLV-mediated HDR in HEK293 cells

To generate infectious IDLV, the lentiviral construct was transfected into HEK293T cells with pC-Help or p8.91.Q168A and pCMV-G by calcium phosphate coprecipitation as described previously (34). The vector harvested was concentrated by precipitation with polyethylene glycol (PEG). To determine the relative IDLV titer, HEK293 cells were transduced with serially diluted viral stocks, selected in 800 $\mu\text{g}/\text{ml}$ G418- or 1 $\mu\text{g}/\text{ml}$ puromycin-containing medium. Colony-Forming Unit (CFU) was scored after two weeks in selection. To carry out HDR, 2 μg of each TALEN plasmid or 1 μg pU6-sgRNA together, 3 μg pCBh-hCas9, with or without 1 μg IDLV construct were co-transfected into 6×10^5 HEK293 cells using Lipofectamine 2000 in 6-well dishes. This was followed by IDLV transduction 16 h later at a dose of 2000 CFUs. The treated cells were serially diluted, plated and selected in G418- or puromycin-containing medium 48 h after transduction. The genomic DNA from individual *neo^R* or *pur^R* colonies was isolated after two weeks using Quick Extract solution and subjected to PCR screening. The PCR primers used for measuring the HDR frequency for the *WAS* gene were WF1 (TGCTCCCTGCCAGCTAACAA) and WR1 (CCATCTTGTTCATGGCCGATCCC) for the genomic region upstream from the nuclease cleavage site, WF2 (AGCGCATCGCCTTCTATCGCC) and WR2 (TCTGCCATCCTTCCATTCACTCA) for the genomic region downstream from the nuclease cleavage site. WF1 and WR2 were located outside of the homology arms to avoid the amplification of the donor template-containing IDLV genome. The amplification was carried out with PCR

Master Mix (Promega, Madison, WI, USA), using the following cycling conditions: 95°C for 5 min for initial denaturation; 34 cycles of 94°C for 30 s, 58°C for 30 s and 72°C for 90 s; and a final extension at 72°C for 7 min. The primer set for H-T concatemer detection was R-LTR (CGCACCCA TCTCTCTCCTTCTA), LTR-WPRE (TCCTTCTGCTAC GTCCCTTC). The primer set for DNA loading control in Figures 1C and 2C was Loading Ctrl-F (AGGGTTTCAC TATGAAGGGAGGGA) and Loading Ctrl-R (AGTGGA CCAGAACGACCCTTGTTA).

IDLV-mediated HDR in WA09 cells

WA09 cells were pre-treated with 10 μ M Rock Inhibitor for at least 2 h and dissociated into single cell suspension with 1 mg/ml Accutase (Invitrogen). Approximately 1.5×10^6 cells were mixed with 2.5 μ g of each TALEN plasmid or 1.5 μ g pU6-sgRNA and 3.5 μ g pCBh-hCas9, nucleofected using program B-016 with Nucleofector (Lonza, Basel, Switzerland) and cultured in Matrigel coated 24-well dishes in mTeSR1 supplemented with 10 μ M Rock Inhibitor. This was followed by IDLV transduction at 2000 CFUs 16 h later. Forty-eight hours after transduction, cells were dissociated by Accutase and seeded at a concentration of 1.5×10^4 cells/well into 6-well dishes with irradiated DR4 feeder cells. G418 at a concentration of 600 μ g/ml or puromycin at a concentration of 0.5 μ g/ml was used for selection 3 days after cell re-plating. Two weeks after drug selection, half of each colony was manually picked and transferred into Quick Extract DNA solution for genomic DNA extraction. PCR screening was carried out as mentioned above. The primers for identifying homozygous HDR clones in Figure 2A were homo Ctrl-F (AAGTCCCCTCTCATGGTCT) and homo Ctrl-R (GCTCGTCCATCCACATACCT). This primer set flanked the nuclease cleavage site in the *WAS* locus (Figure 1A) and would generate a 359-bp PCR fragment if homozygous HDR did not occur. The PCR primer set used for measuring the HDR frequency in the *HB9* locus were 9F1 (GAGGAGGGAAACATCCTGAGA) and 9R1 (GGGGAACCTTCCTGACTAGGG) for the genomic region upstream from the nuclease cleavage site, 9F2 (GC AACCTCCCCTTCTACGAG) and 9R2 (CCCTCCCCAT CATTCTACA) for the genomic region downstream from the nuclease cleavage site.

Functional knockdown of LEDGF/p75 with a shRNA and the IBD domain

A short hairpin RNA (shRNA), L1, targeting LEDGF/p75 was designed as described (29). Two single-stranded oligodeoxyribonucleotides with the sequences of GATCCCCGCAATGAGGATGTGACTAATTCA AGAGATTAGTCACATCCTCATTGCTTTTTGGAAA and AGCTTTTCCAAAAGCAATGAGGATGTGACT AATCTCTTGAATTAGTCACATCCTCATTGCGGG were synthesized, annealed and cloned into an expression vector containing the H1 promoter. A similar expression plasmid containing the shRNA specific for the firefly luciferase (*luc*) gene was used as the negative control. To test the knockdown efficiency of L1, 1 or 4 μ g of the shRNA plasmid and 1 μ g of pmaxGFP plasmid as the

transfection efficiency control were co-transfected into HEK293 cells using Lipofectamine 2000 in 6-well dishes. Seventy-two hours after transfection, the transfected cells were harvested and lysed for protein extraction. Approximately 20 μ g of the total protein were separated by gel electrophoresis and subjected to western blot analysis using the rabbit anti-LEDGF/p75 antibody (1:1000, Bethyl Laboratories, Montgomery, TX, USA) and the rabbit polyclonal anti-GAPDH antibody (Abcam, Cambridge, MA, USA). Signals were detected with Odyssey Imaging System (LI-COR, Lincoln, NE, USA) for LEDGF/p75 and X-ray film for GAPDH. Plasmid pCHMWS/GFP-IBD-IRES-Puro used for the IBD competition study and the negative control plasmid pCHMWS/GFP-IRES-Puro were kindly provided by Dr Debyser (KU Leuven and IRC KULAK, Belgium). For the LEDGF/p75 knockdown study, 0.5 μ g WASCR-2, 1.5 μ g pCBh-hCas9 and 2 μ g L1 or the luc shRNA control were co-transfected into 6×10^5 HEK293 cells using Lipofectamine 2000 in 6-well dishes. For the IBD competition study, pCHMWS/GFP-IBD-IRES-Puro and pCHMWS/GFP-IRES-Puro instead of L1 and the luc shRNA were used. This was followed by the IDLV transduction at a dose of 2,000 CFUs 16 h after transfection. Treated cells were serially diluted, plated and selected in G418. After 2 weeks in selection, neo^R colonies were stained with methylene blue (2% in methanol) and counted by the ImageJ system.

Statistics

Statistical analysis was performed with the GraphPad Prism 6.0 software. Error bars indicate mean \pm s.d.. Comparisons between groups were carried out using two-sided *t*-test. Significance was considered when $P < 0.05$.

RESULTS

Favored donor template delivery via IDLV for HDR

We first determined whether using IDLV to deliver the donor template for HDR had any advantage over conventional DNA transfection in human cells. Several TALEN pairs were designed to target a genomic region in intron 6 of the Wiskott-Aldrich Syndrome (*WAS*) gene (Figure 1A). Mutations in this gene cause profound immunodeficiency in patients (35). The TALEN pairs were targeted to the 5' end of intron 6 as mutations in this region were frequently found to affect normal splicing of the *WAS* transcript (36). Three upstream and six downstream TALENs were designed to cleave the targeted region (Figure 1A). A study of mix-matching the upstream and downstream TALENs in HEK293 cells showed that the combination of W3L and W4R generated the highest efficiency in insertion deletion (indel) formation (Figure 1B). To construct the donor template for HDR, a cassette containing the *WAS* sequences of the targeted region and the gene encoding neomycin phosphotransferase (*neo*^R) gene was inserted into a lentiviral vector for IDLV production (Figure 1A). For TALEN expression, we used lipofectamine to transfect the W3L/W4R expression plasmids into HEK293 cells. Transfected cells were transduced with the IDLVs 16 h later and selected for neo^R colonies. In the absence of TALEN, few neo^R colonies

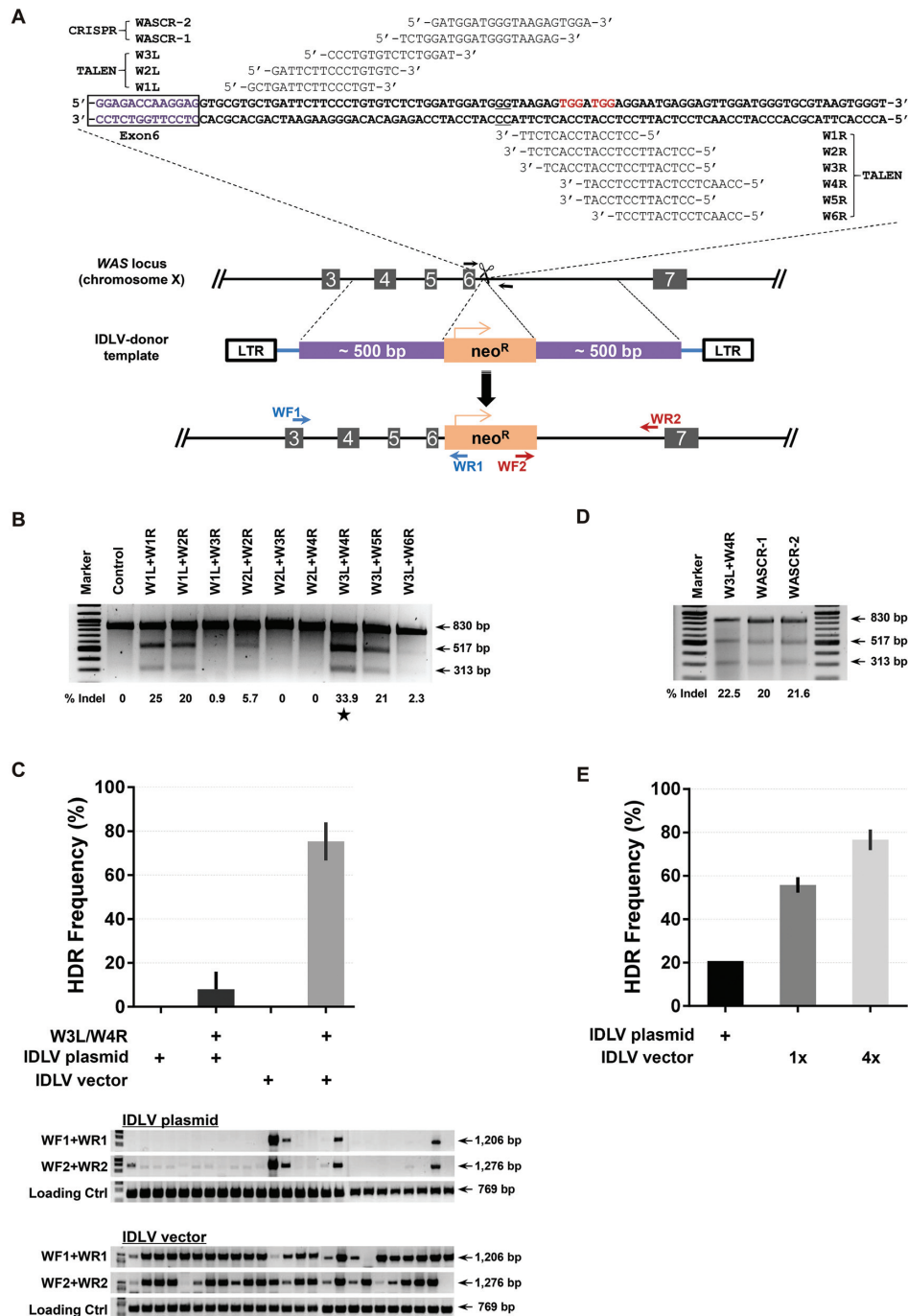


Figure 1. IDLV mediates efficient HDR at the *WAS* locus in HEK293 cells. (A) Schematic display of the targeting strategy at the *WAS* locus. The sequences of CRISPR and TALEN targeting intron 6 of the *WAS* gene are shown. PAM sequences for the two CRISPRs are labeled in red. The *WAS* genomic regions used as the homology arms (~500 bp in length) for HDR are indicated. *WAS* exons are shown as numbered boxes. Two unlabeled dark arrows flanking the nuclease cleavage site in the *WAS* gene denote the PCR primers used to determine whether homologous HDR occurs. Orange arrow in the *neo^R* gene indicates the direction of transcription. Neo^R colonies were screened by PCR using WF1 and WR1 to detect homologous recombination at the upstream region of the nuclease cleavage site and WF2 and WR2 to detect homologous recombination at the downstream region. (B) TALEN cleavage efficiency determined by the Surveyor assay. HEK293 cells were transfected with different TALEN combinations. Efficiency of indel formation is shown underneath each lane. Expected fragment sizes after Surveyor cleavage are indicated on the right. (C) HDR frequency in IDLV transfected (lanes 1 and 2) and transduced (lanes 3 and 4) HEK293 cells. Data represents the mean of three independent experiments ± s.d.. Gels representing some screening data to detect HDR-derived colonies are also shown. The upper panel shows colonies from transfection of the IDLV and W3L/W4R plasmids and the lower panel shows the colonies from a combination of W3L/W4R plasmid transfection and IDLV transduction. Primers used for PCR are indicated on the left and the expected fragment sizes are indicated on the right. (D) CRISPR cleavage efficiency determined by the Surveyor assay. HEK293 cells were transfected with the indicated CRISPR/Cas9 plasmids and the efficiency of indel formation determined by the Surveyor assay. (E) HDR frequency mediated by the co-transfection of the WASCR-2/Cas9 plasmids and the IDLV construct or the transfection of the WASCR-2/Cas9 plasmids and the transduction with IDLV in HEK293 cells. Except the co-transfection experiment, data represents the mean of three independent experiments ± s.d.

Table 1. Summary of screening neo^R colonies derived from HDR at the *WAS* locus

Cells	Engineered nucleases	Plasmids	IDLV donor template	Total colonies screened	HDR colonies at		
					Both arms		
					Heterozygous	Homozygous	Single arm
HEK293	TALEN	W3L + W4R	Plasmid	70	5	-	1
		W3L + W4R	Virus	70	52	-	8
	CRISPR/Cas9	WASCR-2/Cas9	Plasmid	48	10	-	10
		WASCR-2/Cas9	Virus (1x)	105	59	-	11
		WASCR-2/Cas9	Virus (4x)	60	45	1	2
WA09	TALEN	W3L + W4R	Virus	81	40	1	9
		WASCR-2/Cas9	Virus	97	78	1	10

“-”Not done.

appeared (15 ± 2 , $n = 3$). As the IN mutation (D116N) in IDLV completely abolished its enzymatic activity (37), emergence of these colonies could be attributed to the preferential incorporation of the linear IDLV genome by naturally occurring DSBs in HEK293 cells. In the presence of W3L/W4R, the number of neo^R colonies increased significantly (193 ± 64 , $n = 3$). This increase could be attributed to one of the two possibilities. First, DSBs generated by the TALENs at the genomic target stimulated the incorporation of the linear IDLV genome via its two free ends. Second, DSBs stimulated HDR between the genomic locus and the donor template in the IDLV genome. To differentiate these two possibilities, we used genomic PCR to screen the neo^R colonies for HDR. Out of 70 colonies screened, 52 (75%) scored positive for HDR at both homology arms whereas 8 (11%) scored positive at only one arm (Figure 1C and Table 1). Thus, up to 86% of the neo^R colonies were derived from HDR with a majority of them containing the expected configuration. In contrast, co-transfection of HEK293 cells with the W3L/W4R expression plasmids and the IDLV construct followed by PCR screening of the neo^R colonies exhibited a HDR frequency of lower than 10% (Figure 1C and Table 1). Thus, a minimum of 8-fold enrichment in the clones derived from homologous recombination was achieved when the donor template was delivered through IDLV transduction instead of DNA transfection. Since HEK293 cells can be transfected with high efficiency and multiple copies of the donor template plasmid are expected to be introduced via transfection, it is unlikely that the difference in the HDR frequency between IDLV transduction and plasmid transfection can simply be attributed to the availability of the donor template. One possibility is that the donor template in the context of an IDLV genome has a favored configuration to serve as the substrate for HDR or simply because the IDLV genome in the pre-integration complex of lentiviral vectors is more stable than transfected naked plasmid DNA in the cell.

To ensure that the elevated HDR frequency observed with IDLV transduction is not limited to the engineered nuclease used, we designed two CRISPRs (WASCR-1 and WASCR-2) targeting the same genomic region as the TALEN pairs (Figure 1A). A comparison of the indel frequency in HEK293 cells showed a similar efficiency between the CRISPRs and W3L/W4R (Figure 1D). HEK293 cells were transfected with the WASCR-2/Cas9 plasmids and transduced with the same donor template-containing IDLV shown in Figure 1A. PCR analysis of neo^R colonies indicated that out of 105 colonies screened, 70 (67%) of them

scored positive for HDR. Among them, 59 (56%) scored positive at both homology arms (Table 1). Co-transfection of HEK293 cells with the WASCR-2/Cas9 and the IDLV plasmids led to the emergence of the neo^R colonies, and out of 48 colonies screened, 10 (21%) were derived from HDR at both arms (Table 1). To determine whether an increase in IDLV input can boost the HDR frequency further, we transduced HEK293 cells with 4-fold more IDLV. PCR analysis of neo^R colonies showed that 46 (77%) out of 60 neo^R colonies screened were positive for HDR at both arms (Table 1). An increase in the IDLV input therefore boosted the likelihood of gaining HDR-derived clones with the expected configuration at both homology arms (Figure 1E). These results strongly suggest that the donor template delivered through IDLV transduction has a significant advantage over DNA transfection to mediate HDR in HEK293 cells. This is irrespective of the engineered nuclease used to generate DSBs as long as the engineered nuclease is expressed efficiently in the target cell.

Efficient HDR mediated by IDLV in WA09 cells

To determine whether IDLV can mediate highly efficient HDR in hESCs, WA09 cells were nucleofected with the expression plasmids for W3L/W4R followed by IDLV transduction. Neo^R colonies were screened for HDR by PCR. Out of 81 neo^R colonies screened, 50 (62%) scored positive for HDR. Among them, 41 (51%) were derived from HDR at both homology arms including one colony with both *WAS* alleles altered by HDR (Table 1). Since previous studies showed that the genome of IDLV was prone to the formation of concatemers through illegitimate recombination in transduced cells (20,21,24), we measured the fraction of HDR colonies containing the head-to-tail (H-T) IDLV concatemers at the *WAS* locus. We used a PCR primer pair specific for the IDLV genome to identify those clones containing the H-T structure (Figure 2A). Among 30 randomly picked colonies with HDR at either one or both arms, 15 of them exhibited the H-T concatemer configuration at the targeted *WAS* locus (Figure 2A, upper panel). The presence of additional sequences derived from the IDLV vector would interfere with DNA repair or transgene expression from the targeted genomic locus. Excluding those colonies with the concatemeric structure reduces the overall frequency of HDR-derived colonies with the expected genomic modification but still remains well above the HDR frequency achieved by plasmid transfection or viral vector transduction shown in other studies. A similar study was carried

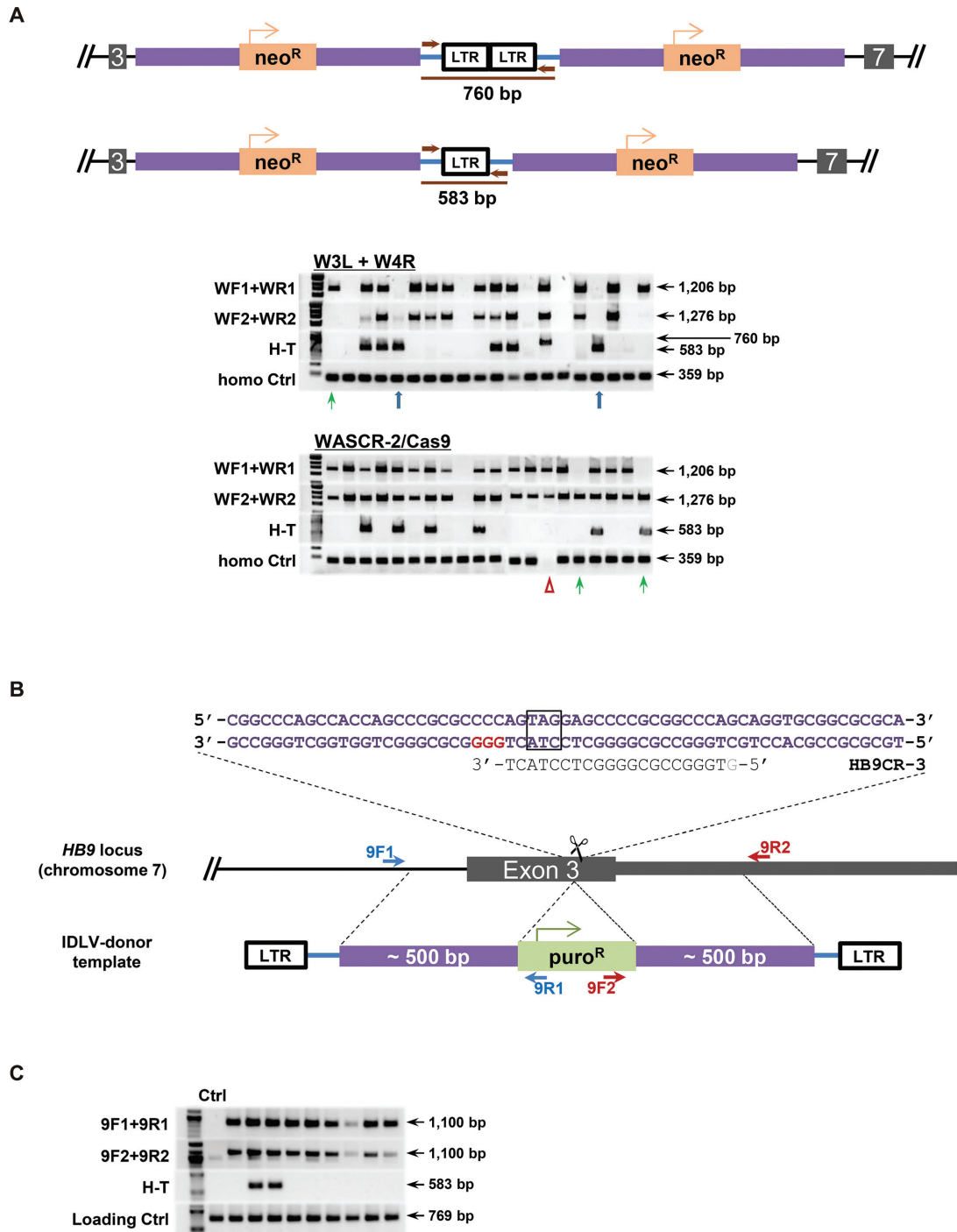


Figure 2. IDLV mediates efficient HDR at the *WAS* locus in WA09 cells. (A) Schematic display of the head-to-tail (H-T) IDLV concatemer with either 1 or 2 LTRs. Red arrows indicate the PCR primers used in the detection of the H-T structure and the expected sizes of the PCR fragments are also shown. Gels demonstrate PCR screening of HDR-derived *neo^R* WA09 colonies and those with the H-T structure. Primers used for PCR screening and H-T detection are indicated on the left and the expected fragment sizes are indicated on the right. Green arrows indicate those colonies with HDR at only one homology arm. Blue arrows indicate non-HDR colonies with the H-T structure; red triangle indicates a colony with HDR at both alleles as determined by the unlabeled PCR primers (homo Ctrl) shown in Figure 1A. (B) Schematic display of the targeting strategy at the *HB9* locus. Exon 3 is shown as the gray box and the sequence spanning the CRISPR cleavage site is shown. The stop codon for the *HB9* coding region is boxed. *HB9CR-3* sequence and position are indicated and the PAM sequence is labeled in red. The *HB9* genomic regions used as the homology arms for HDR in the IDLV construct are indicated by the dashed lines. Colonies derived from HDR were screened using 9F1 and 9R1 to detect homologous recombination at the upstream region of the nuclease cleavage site and 9F2 and 9R2 to detect homologous recombination at the downstream region. (C) Detection of HDR-derived WA09 colonies with the insertion of the *puro^R* gene into the *HB9* locus.

Table 2. Summary of screening *pur*^R colonies derived from HDR at the *HB9* locus

Cell	Plasmids	IDLV donor template	Total colonies screened	HDR colonies at		
				Both arms	Single arm	H-T Concatemer
WA09	HB9CR-3/Cas9	Virus (1×)	9	9	0	2
	HB9CR-3/Cas9	Virus (2×)	30	14	1	14
	HB9CR-3/Cas9	Virus (4×)	60	29	11	38

out using WASCR-2/Cas9 in WA09 cells. Out of 97 *neo*^R colonies screened, 89 (92%) scored positive for HDR and 79 (81%) of those exhibited HDR at both homology arms (Figure 2A and Table 1). It is interesting that the HDR frequency derived from WASCR-2/Cas9 was significantly higher than that from W3L/W4R in WA09 cells (92 versus 62%) whereas W3L/W4R generated HDR-derived colonies more efficiently than WASCR-2/Cas9 in HEK293 cells (86 versus 67%) (Table 1). This result is consistent with two previous studies demonstrating more efficient indel formation at the genomic target with CRISPR/Cas9 than that with TALEN in human ESCs (15,38). Examination of 30 randomly picked HDR-derived colonies with WASCR-2/Cas9 showed the presence of the H-T structure in 15 of them (Figure 2A, lower panel). Concatemer formation is therefore unavoidable when IDLV is used to deliver the donor template.

To ensure that the high HDR frequency mediated by IDLV is not limited to the *WAS* locus, we designed a CRISPR, HB9CR-3, targeting a site in exon 3 of the motor neuron-specific *HB9* gene (39). A donor template containing the puromycin-resistant (*pur*^R) gene flanked by the genomic sequence spanning the HB9CR-3 cleavage site was inserted into a lentiviral vector for IDLV production (Figure 2B). WA09 cells were nucleofected with the HB9CR-3/Cas9 expression plasmids followed by the IDLV transduction. In the first round of screening, all nine *pur*^R colonies were positive for HDR at both homology arms with two of them exhibiting the H-T concatemer (Figure 2C). Repeated studies with increased IDLV input showed the HDR frequency near 50% or higher at the *HB9* locus (Table 2). However, the frequency of colonies containing the H-T concatemer also increased significantly. This data suggests that the high frequency of HDR mediated by IDLV is not limited to a specific genomic locus but the high-level IDLV input may lead to concatemerization of the IDLV genome not suitable for site-specific gene modification. To apply the IDLV strategy in gene editing, the donor template-containing IDLV input needs to be titrated carefully to optimize the HDR frequency at the targeted genomic locus.

LEDGF/p75 contributed to IDLV-mediated HDR

One possibility to explain the significant advantage of using IDLV transduction to mediate HDR over plasmid transfection is that the pre-integration complex of IDLV actively recruits cellular protein(s) involved in the normal process of HDR, leading to the observed increase in clones derived from homologous recombination. We tested this hypothesis by focusing on the HIV-1 IN protein. The cellular factor LEDGF/p75 is known to associate with HIV-1

IN and involved in HIV pathogenesis (28). LEDGF/p75 binds to chromatin via its PWFP domain at the N terminus and to HIV-1 IN through its integrase binding domain (IBD) at the C terminus (Figure 3A) (40). It was recently shown that LEDGF/p75 promoted the repair of DSBs through the normal homologous recombination pathway (29). LEDGF/p75 carries out this activity by recruiting C-terminal binding protein interacting protein (CtIP) in a DNA damage-dependent manner. CtIP acts together with the MRE11-RAD50-NBS1 complex to promote DNA end resection and generate single-stranded DNA crucial for HDR (41). Thus, it is likely that IDLV, through the interaction of IN with LEDGF/p75, recruits cellular proteins such as CtIP to facilitate HDR (Figure 3A). To test this hypothesis, we first determined whether LEDGF/p75 was involved in IDLV-mediated HDR. As shown in Figure 3B, downregulation of LEDGF/p75 in HEK293 cells with a small hairpin RNA (shRNA) resulted in significant reduction in the number of *neo*^R colony, and genomic PCR of the pooled *neo*^R colonies confirmed that LEDGF/p75 reduction indeed lowered IDLV-mediated HDR frequency (Figure 3B, lower panel). Over-expression of the LEDGF/p75 IBD that competes with LEDGF/p75 for IN binding (42) also led to a similar reduction in the *neo*^R colony number and the HDR frequency confirmed by genomic PCR (Figure 3C). The data is consistent with the hypothesis that LEDGF/p75 plays an important role in IDLV-mediated HDR. These studies, however, cannot exclude the possibility that LEDGF/p75 is directly involved in the cellular HDR pathway and a reduction in its expression by the shRNA or disruption of its activity by IBD leads to a universal reduction in the HDR frequency in general. To directly address the role of the IN protein in IDLV-mediated HDR, we generated IDLV with a packaging plasmid containing the IN mutant Q168A (43). This mutant loses the ability to bind LEDGF/p75 but continues to support viral DNA replication and migration of the pre-integration complex into the nucleus (43,44). This is in contrast to the D116N IN mutant we used to produce the IDLV for HDR in all the studies described above. The D116N mutant contains a mutation in the IN catalytic core domain (37). This mutation abolishes the enzymatic activity of IN but supports viral DNA replication, the migration of the pre-integration complex into the nucleus and, most importantly, the interaction with LEDGF/p75 (44,45). Side-by-side comparison in HEK293 cells transduced with IDLVs containing different IN mutations showed a significant reduction in the HDR frequency at the *WAS* locus with IDLV(Q168A) relative to that with IDLV(D116N) (Figure 3D). This data suggested that the direct interaction between the IN protein and LEDGF/p75 was important for IDLV-

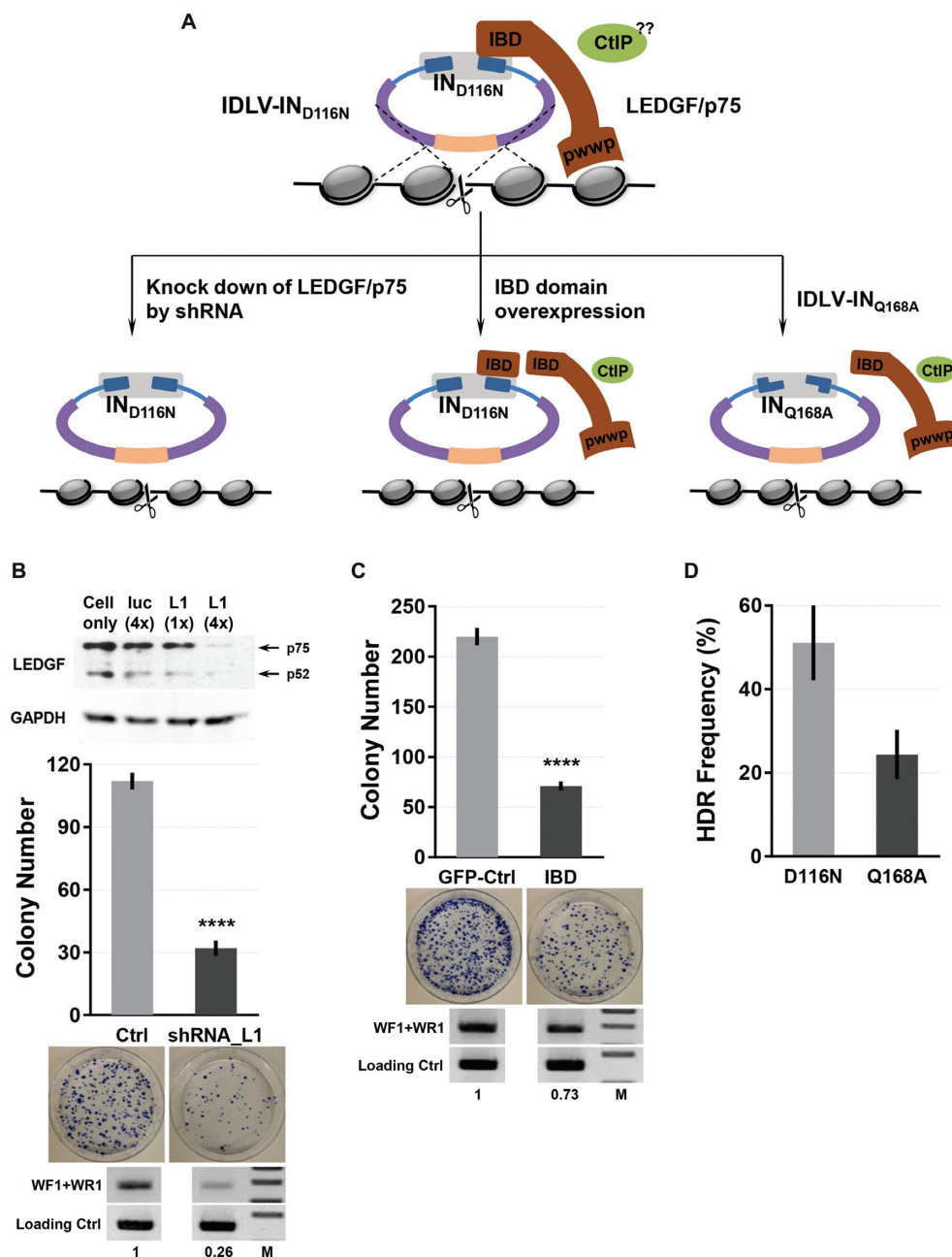


Figure 3. LEDGF/p75 recruitment is important for IDLV-mediated HDR. (A) Schematic display of the interaction between the IN protein and LEDGF/p75. Knockdown of LEDGF/p75 by shRNA or over-expression of the IBD domain would disrupt this interaction and reduce the IDLV-mediated HDR. Alternatively, using an IN mutant, Q168A, to abolish its interaction with LEDGF/p75 should also reduce the IDLV-mediated HDR if LEDGF/p75 is involved in this activity. Although CtIP directly interacts with LEDGF/p75 to mediate DSB repair, it is not clear whether CtIP plays a role in IDLV-mediated HDR. (B) Knockdown of LEDGF/p75 by shRNA L1 reduced the frequency of IDLV-mediated HDR. HEK293 cells were transfected with either L1 at two different amounts (1 and 4×) or a control shRNA, luc, at a fixed amount (4×). Western blot analysis was carried out with the cell extract from the transfected culture 72 h later using a LEDGF/p75-specific antibody and a GAPDH-specific antibody for the loading control. To determine the effect of LEDGF/p75 knockdown on IDLV-mediated HDR, HEK293 cells were co-transfected with the L1 or luc shRNA, and the WASCR-2/Cas9 expression plasmids followed by IDLV transduction and selection in G418-containing medium. The transfected culture was remained in selection for additional two weeks and the neo^R colonies were stained and counted. Data represents the mean of three independent experiments. $P < 0.0001$. Genomic PCR was performed with the pooled neo^R colonies to determine the HDR frequency. The PCR primers used are indicated on the left of the gel. For quantification, the PCR band intensity of the control shRNA was set at 1. 'M' indicates the 1 kb ladder. (C) IBD over-expression reduced the frequency of IDLV-mediated HDR. The experiment was carried out using a similar procedure as described in (B) except that the L1 and luc shRNA expression plasmids were replaced with pCHMWS/GFP-IBD-IRES-Puro expressing IBD and pCHMWS/GFP-IRES-Puro as the empty vector control. Data represents the mean of three independent experiments. $P < 0.0001$. Genomic PCR and quantification were similarly performed as described in (B). (D) IDLV with IN_{Q168A} reduced the HDR frequency. HEK293 cells were co-transfected with the WASCR-2/Cas9 expression plasmids, followed by the transduction with either IDLV_{D116N} or IDLV_{Q168A} at the same multiplicity of infection and selected in G418. Neo^R colonies were screened for HDR by PCR as described in Figure 1C. Data represents the mean of three independent experiments. $P = 0.0665$.

mediated HDR. Based on these studies, we concluded that the interaction between IN in the IDLV pre-integration complex and LEDGF/p75 might facilitate the recruitment of other cellular proteins, such as CtIP, to repair DSBs generated by the engineered nuclease. The ability of the lentiviral IN protein to recruit LEDGF/p75 therefore explain the unique advantage of IDLV over AD and AAV vectors to deliver the donor template for HDR in human cells including hESCs.

DISCUSSION

In the current study, we showed highly efficient HDR mediated by IDLV in human HEK293 cells and hESC WA09 cells. The HDR frequency, including recombination at one or both homology arms, reached more than 80% among the antibiotics-selected colonies in both cell lines. Our previous study and a study by Yang *et al.* showed that the HDR frequency in hPSCs was not limited by the cleavage activity of an engineered nuclease as the nuclease generated NHEJ at higher frequencies than HDR at the same genomic locus (14,15). Rather, the frequency of HDR seems to depend on the availability of the donor template at the DSB. The donor template present in the context of the IDLV genome should be well protected by the viral and cellular proteins in the pre-integration complex from degradation by various cellular nucleases. In contrast, a transfected donor plasmid is more likely under constant exposure to cellular nucleases which can lead to the degradation of the donor template. This hypothesis is consistent with our observation that the same IDLV construct for making infectious vector, when introduced as a plasmid into target cells, performed HDR much less efficiently than the infectious IDLV (Figure 1C). The other potential reason that accounts for the observed high HDR frequency with IDLV could be due to the presence of the HIV IN protein in the pre-integration complex. Under the normal process of lentiviral infection, the viral pre-integration complex is tethered to the host chromatin through the interaction between the IN protein and LEDGF/p75 to facilitate host genome integration (28). Although the enzymatically inactive D116N IN mutant we used to prepare the IDLV for HDR loses its ability to catalyze viral DNA integration, mutation at this residue in the IN protein continues to support the tethering of the viral DNA to host chromatin (44). In contrast, the Q168A mutation completely abolishes the ability of the IN protein to bind host chromatin (43,44). The result in Figure 3 demonstrates the importance of chromatin tethering of the viral pre-integration complex through the interaction between IN and LEDGF/p75 for IDLV-mediated HDR. The other possibility for the observed high HDR frequencies could be due to the active recruitment of the cellular CtIP protein, an important component for homologous recombination, by the chromatin-associated pre-integration complex of IDLV to stimulate HDR. It was shown that LEDGF/p75 tethered CtIP to the sites of DNA damage via a direct interaction between the two proteins (29). CtIP is involved in DNA end resection and generation of single-strand DNA (ssDNA) at DSBs (46). The ssDNA at DSBs serves as a key intermediate for the assembly of the signaling machinery for homologous recombination-mediated DSB repair. Importantly,

these possibilities are not mutually exclusive: the stability of the IDLV pre-integration complex, the chromatin tethering of the pre-integration complex and the active recruitment of HDR proteins may all contribute to the observed high HDR frequency mediated by IDLV. Our observation that disruption of the interaction between the IN protein and LEDGF/p75 only reduced the HDR frequency by half is consistent with this hypothesis (Figure 3D). One caveat is whether LEDGF/p75 can bind both IN and CtIP simultaneously. Daugaard *et al.* reported that a LEDGF/p75 mutation D366N abolished IN binding but continued to support the CtIP binding (29), suggesting that these two proteins did not compete for LEDGF/p75 binding.

High HDR frequency relies on efficient expression of the engineered nuclease. Compared with previous studies using IDLV to express the engineered nuclease (24,25), our study used plasmid nucleofection to express the engineered nuclease in hESCs. This procedural change is expected to elevate the expression of the engineered nuclease, most likely due to the introduction of multiple copies of the nuclease expression plasmids. A previous study comparing the effect on HDR with I-SceI expressed from either an integrating lentiviral vector or an IDLV showed a much higher HDR frequency in several different human cell lines with the integrating lentiviral vector (47). Together, these studies confirm that IDLV is not suitable for high-level nuclease expression (24,25). Other studies used a combination of Ad vectors to express the engineered nuclease and IDLV to provide the donor template for HDR in human cells. Holkers *et al.* used this combination to show efficient HDR (~87% with selection) at the *AAVSI* locus in human myoblasts (20). Coluccio *et al.* used a similar strategy and showed a HDR frequency of 10–20% at the *AAVSI* locus in two human keratinocyte cell lines but only 0.2–0.3% in primary human keratinocytes (21). Thus, the HDR frequency varies significantly in different cell types even with the Ad vector system capable of robust nuclease expression. Although the Ad vector is able to mediate efficient nuclease expression in different cell types, our study shows that nucleofection of the nuclease expression plasmid is sufficient to mediate efficient HDR in hESCs. It avoids the time consuming steps of constructing the Ad nuclease expression vector and infectious vector production. Our data also shows that a significant fraction of the HDR colonies derived from the IDLV strategy described here contains the concatemeric IDLV structure at the targeted genomic site. Several studies using IDLV for HDR have detected the similar structure (20,21,24). The concatemer is most likely derived from illegitimate recombination among the linear IDLV genome in the transduced cells. The hypothesis is consistent with the observation that an increase in the IDLV input resulted in higher frequencies of WA09 colonies containing the concatemer (Table 2). Although the formation of the concatemer does not seem to interfere with HDR, these tandem repeats are not expected to restore a reading frame in gene repair nor will they yield the expected transgene expression pattern in the context of gene insertion. To minimize the production of hESC clones with such a structure, the level of the input IDLV needs to be carefully titrated to maximize the generation of HDR clones with the desired gene modification.

Multiple studies carried out in hPSCs have indicated that the HDR frequency at silent genetic loci was generally lower than that at active loci. Zhou *et al.* used ZFN to correct the missense mutation in the β -globin gene and obtained a single gene-corrected clone out of 300 clones screened in a sickle cell anemia patient-derived iPSC line (8). Sebastiano *et al.* carried out a similar study in a different sickle cell anemia iPSC line and achieved \sim 10% HDR frequency at the β -globin locus (48). Soldner *et al.* used ZFN-mediated HDR to either introduce mutations into the neuron-specific gene encoding α -synuclein in several hESC lines or correct the gene mutation in a patient-derived iPSC lines and achieved HDR frequencies ranging from 0 to 2% (49). In contrast to these studies that targeted silent genes, Zhou *et al.* used ZFN to target the constitutively active gene encoding phosphatidylinositol glycan class A (PIG-A) and showed much higher HDR frequencies (up to 50%) in both ESC and iPSC lines (27). Hockemeyer *et al.* used ZFN to target two active loci in hPSCs, *Oct4* and *AAVS1*, and showed HDR frequencies ranging from 39 to $>$ 90% in a hESC line. In contrast, targeting a silent gene encoding transcription factor PITX3 resulted in a HDR frequency of only 11% in the same hESC line (7). Thus, these studies showed a trend of lower HDR frequency with silent loci in hPSCs. The two loci we evaluated in the current study, *WAS* and *HB9*, are active in hematopoietic lineages and motor neurons, respectively, but remain silent in undifferentiated WA09 cells. Yet with the IDLV strategy, we obtained a HDR frequency of $>$ 60% at both genetic loci. Our study was carried out with a donor template containing only \sim 500 bp homology arm on each side of the neo^R marker. The size of the homology arm is the smallest among all the reported studies. The size of the homology arm has been reported to be proportional to the HDR frequency (6). The short homology arm length in the context of the IDLV genome is therefore not rate-limiting to mediate efficient HDR in both HEK293 and WA09 cells. Together, our studies demonstrate the advantages of using IDLV to mediate HDR for site-specific gene modification in hPSCs.

In summary, we show that IDLV-mediated delivery of the donor template represents a highly efficient strategy to mediate gene editing in hESCs. HDR clones with the desired gene modification could be established at much higher frequencies than other strategies reported in the literature. We discover that the IN protein in the pre-integration complex of IDLV, through its interaction with LEDGF/p75, contributes significantly to the high HDR frequency in human cells. With the demonstrated HDR frequency, the burden to carry out large-scale gene editing in hESCs can be overcome by this strategy. Although a significant fraction of the HDR-derived hESC clones contains the concatemeric structure, the clones with the desired gene modification can easily be identified by PCR-based assays. As this strategy generates high frequency of HDR-derived clones, it would be less time consuming and labor intensive to isolate the correct gene-edited clones. Future direction would be to identify strategies to reduce hESC clones with the concatemeric structure.

ACKNOWLEDGEMENTS

We thanks Drs Debyser and Serguera for providing the IBD and the Q168A packaging plasmids used in the study. Research reported in this publication included work performed in the Analytical Cytometry Core and Integrative Genomics Core of City of Hope National Medical Center.

FUNDING

National Key Basic Research Program of China [2015CB964900, in part]; Projects of International Cooperation and Exchanges [81520108002]; Zhejiang Provincial Natural Science Foundation of China [LQ14H080001]; Major Program of National Natural Science Foundation of China [81230014]; National Cancer Institute of the National Institutes of Health [P30CA033572]. Funding for open access charge: National Key Basic Research Program of China [2015CB964900].

Conflict of interest statement. None declared.

REFERENCES

- Kim, H. and Kim, J.S. (2014) A guide to genome engineering with programmable nucleases. *Nat. Rev. Genet.*, **15**, 321–334.
- Lieber, M.R. (2010) The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. *Annu. Rev. Biochem.*, **79**, 181–211.
- Shrivastav, M., De Haro, L.P. and Nickoloff, J.A. (2008) Regulation of DNA double-strand break repair pathway choice. *Cell Res.*, **18**, 134–147.
- Heyer, W.D., Ehmsen, K.T. and Liu, J. (2010) Regulation of homologous recombination in eukaryotes. *Annu. Rev. Genet.*, **44**, 113–139.
- Orthwein, A., Fradet-Turcotte, A., Noordermeer, S.M., Canny, M.D., Brun, C.M., Strecker, J., Escobedo-Diaz, C. and Durocher, D. (2014) Mitosis inhibits DNA double-strand break repair to guard against telomere fusions. *Science*, **344**, 189–193.
- Zwaka, T.P. and Thomson, J.A. (2003) Homologous recombination in human embryonic stem cells. *Nat. Biotechnol.*, **21**, 319–321.
- Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKelver, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B. *et al.* (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.*, **27**, 851–857.
- Zou, J., Mali, P., Huang, X., Dowey, S.N. and Cheng, L. (2011) Site-specific gene correction of a point mutation in human iPS cells derived from an adult patient with sickle cell disease. *Blood*, **118**, 4599–4608.
- Forster, R., Chiba, K., Schaeffer, L., Regalado, S.G., Lai, C.S., Gao, Q., Kiani, S., Farin, H.F., Clevers, H., Cost, G.J. *et al.* (2014) Human intestinal tissue with adult stem cell properties derived from pluripotent stem cells. *Stem Cell Rep.*, **2**, 838–852.
- Yu, C., Liu, Y., Ma, T., Liu, K., Xu, S., Zhang, Y., Liu, H., La Russa, M., Xie, M., Ding, S. *et al.* (2015) Small molecules enhance CRISPR genome editing in pluripotent stem cells. *Cell Stem Cell*, **16**, 142–147.
- Maruyama, T., Dougan, S.K., Truttmann, M.C., Bilate, A.M., Ingram, J.R. and Plough, H.L. (2015) Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. *Nat. Biotechnol.*, **33**, 538–542.
- Chu, V.T., Weber, T., Wefers, B., Wurst, W., Sander, S., Rajewsky, K. and Kuhn, R. (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat. Biotechnol.*, **33**, 543–548.
- Singh, P., Schimenti, J.C. and Bolcun-Filas, E. (2015) A mouse geneticist's practical guide to CRISPR applications. *Genetics*, **199**, 1–15.
- Wang, X., Wang, Y., Huang, H., Chen, B., Chen, X., Hu, J., Chang, T., Lin, R.J. and Yee, J.K. (2014) Precise gene modification mediated by TALEN and single-stranded oligodeoxynucleotides in human cells. *PLoS One*, **9**, e93575.

15. Yang, L., Guell, M., Byrne, S., Yang, J.L., De Los Angeles, A., Mali, P., Aach, J., Kim-Kiselak, C., Briggs, A.W., Rios, X. *et al.* (2013) Optimization of scarless human stem cell genome editing. *Nucleic Acids Res.*, **41**, 9049–9061.
16. Rahman, S.H., Bobis-Wozowicz, S., Chatterjee, D., Gellhaus, K., Pars, K., Heilbronn, R., Jacobs, R. and Cathomen, T. (2013) The nontoxic cell cycle modulator indirubin augments transduction of adeno-associated viral vectors and zinc-finger nuclease-mediated gene targeting. *Hum. Gene Ther.*, **24**, 67–77.
17. Wang, J., Exline, C.M., DeClercq, J.J., Llewellyn, G.N., Hayward, S.B., Li, P.W., Shivak, D.A., Surosky, R.T., Gregory, P.D., Holmes, M.C. *et al.* (2015) Homology-driven genome editing in hematopoietic stem and progenitor cells using ZFN mRNA and AAV6 donors. *Nat. Biotechnol.*, **33**, 1256–1263.
18. Khan, I.F., Hirata, R.K., Wang, P.R., Li, Y., Kho, J., Nelson, A., Huo, Y., Zavaljevski, M., Ware, C. and Russell, D.W. (2010) Engineering of human pluripotent stem cells by AAV-mediated gene targeting. *Mol. Ther.*, **18**, 1192–1199.
19. Li, L.B., Chang, K.H., Wang, P.R., Hirata, R.K., Papayannopoulou, T. and Russell, D.W. (2012) Trisomy correction in Down syndrome induced pluripotent stem cells. *Cell Stem Cell*, **11**, 615–619.
20. Holkers, M., Maggio, I., Henriques, S.F., Janssen, J.M., Cathomen, T. and Goncalves, M.A. (2014) Adenoviral vector DNA for accurate genome editing with engineered nucleases. *Nat. Methods*, **11**, 1051–1057.
21. Coluccio, A., Miselli, F., Lombardo, A., Marconi, A., Malagoli Tagliazucchi, G., Goncalves, M.A., Pincelli, C., Maruggi, G., Del Rio, M., Naldini, L. *et al.* (2013) Targeted gene addition in human epithelial stem cells by zinc-finger nuclease-mediated homologous recombination. *Mol. Ther.*, **21**, 1695–1704.
22. Aizawa, E., Hirabayashi, Y., Iwanaga, Y., Suzuki, K., Sakurai, K., Shimoji, M., Aiba, K., Wada, T., Tooi, N., Kawase, E. *et al.* (2012) Efficient and accurate homologous recombination in hESCs and hiPSCs using helper-dependent adenoviral vectors. *Mol. Ther.*, **20**, 424–431.
23. Liu, G.H., Suzuki, K., Qu, J., Sancho-Martinez, I., Yi, F., Li, M., Kumar, S., Nivet, E., Kim, J., Soligalla, R.D. *et al.* (2011) Targeted gene correction of laminopathy-associated LMNA mutations in patient-specific iPSCs. *Cell Stem Cell*, **8**, 688–694.
24. Lombardo, A., Cesana, D., Genovese, P., Di Stefano, B., Provasi, E., Colombo, D.F., Neri, M., Magnani, Z., Cantore, A., Lo Riso, P. *et al.* (2011) Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat. Methods*, **8**, 861–869.
25. Lombardo, A., Genovese, P., Beausejour, C.M., Colleoni, S., Lee, Y.L., Kim, K.A., Ando, D., Urnov, F.D., Galli, C., Gregory, P.D. *et al.* (2007) Gene editing in human stem cells using zinc finger nucleases and integrase-defective lentiviral vector delivery. *Nat. Biotechnol.*, **25**, 1298–1306.
26. Joglekar, A.V., Stein, L., Ho, M., Hoban, M.D., Hollis, R.P. and Kohn, D.B. (2014) Dissecting the mechanism of histone deacetylase inhibitors to enhance the activity of zinc finger nucleases delivered by integrase-defective lentiviral vectors. *Hum. Gene Ther.*, **25**, 599–608.
27. Zou, J., Maeder, M.L., Mali, P., Pruetz-Miller, S.M., Thibodeau-Beganny, S., Chou, B.K., Chen, G., Ye, Z., Park, I.H., Daley, G.Q. *et al.* (2009) Gene targeting of a disease-related gene in human induced pluripotent stem and embryonic stem cells. *Cell Stem Cell*, **5**, 97–110.
28. Ciuffi, A., Llano, M., Poeschla, E., Hoffmann, C., Leipzig, J., Shinn, P., Ecker, J.R. and Bushman, F. (2005) A role for LEDGF/p75 in targeting HIV DNA integration. *Nat. Med.*, **11**, 1287–1289.
29. Daugaard, M., Baude, A., Fugger, K., Povlsen, L.K., Beck, H., Sorensen, C.S., Petersen, N.H., Sorensen, P.H., Lukas, C., Bartek, J. *et al.* (2012) LEDGF (p75) promotes DNA-end resection and homologous recombination. *Nat. Struct. Mol. Biol.*, **19**, 803–810.
30. Doyle, E.L., Booher, N.J., Standage, D.S., Voytas, D.F., Brendel, V.P., Vandyk, J.K. and Bogdanove, A.J. (2012) TAL Effector-Nucleotide Targeter (TALE-NT) 2.0: tools for TAL effector design and target prediction. *Nucleic Acids Res.*, **40**, W117–W122.
31. Cermak, T., Doyle, E.L., Christian, M., Wang, L., Zhang, Y., Schmidt, C., Baller, J.A., Somia, N.V., Bogdanove, A.J. and Voytas, D.F. (2011) Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting. *Nucleic Acids Res.*, **39**, e82.
32. Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A. *et al.* (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, **339**, 819–823.
33. Guschin, D.Y., Waite, A.J., Katibah, G.E., Miller, J.C., Holmes, M.C. and Rebar, E.J. (2010) A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol. (Clifton, N.J.)*, **649**, 247–256.
34. Wang, X., Wang, Y., Wu, X., Wang, J., Wang, Y., Qiu, Z., Chang, T., Huang, H., Lin, R.J. and Yee, J.K. (2015) Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. *Nat. Biotechnol.*, **33**, 175–178.
35. Thrasher, A.J. and Burns, S.O. (2010) WASP: a key immunological multitasker. *Nat. Rev. Immunol.*, **10**, 182–192.
36. Imai, K., Morio, T., Zhu, Y., Jin, Y., Itoh, S., Kajiwara, M., Yata, J., Mizutani, S., Ochs, H.D. and Nonoyama, S. (2004) Clinical course of patients with WASP gene mutations. *Blood*, **103**, 456–464.
37. Mochizuki, H., Schwartz, J.P., Tanaka, K., Brady, R.O. and Reiser, J. (1998) High-titer human immunodeficiency virus type 1-based vector systems for gene delivery into nondividing cells. *J. Virol.*, **72**, 8873–8883.
38. Ding, Q., Regan, S.N., Xia, Y., Oostrom, L.A., Cowan, C.A. and Musunuru, K. (2013) Enhanced efficiency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs. *Cell Stem Cell*, **12**, 393–394.
39. Arber, S., Han, B., Mendelsohn, M., Smith, M., Jessell, T.M. and Sockanathan, S. (1999) Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. *Neuron*, **23**, 659–674.
40. Llano, M., Vanegas, M., Hutchins, N., Thompson, D., Delgado, S. and Poeschla, E.M. (2006) Identification and characterization of the chromatin-binding domains of the HIV-1 integrase interactor LEDGF/p75. *J. Mol. Biol.*, **360**, 760–773.
41. You, Z., Shi, L.Z., Zhu, Q., Wu, P., Zhang, Y.W., Basilio, A., Tonnu, N., Verma, I.M., Berns, M.W. and Hunter, T. (2009) CtIP links DNA double-strand break sensing to resection. *Mol. Cell*, **36**, 954–969.
42. De Rijck, J., Vandekerckhove, L., Gijsbers, R., Hombrouck, A., Hendrix, J., Vercammen, J., Engelborghs, Y., Christ, F. and Debysier, Z. (2006) Overexpression of the lens epithelium-derived growth factor/p75 integrase binding domain inhibits human immunodeficiency virus replication. *J. Virol.*, **80**, 11498–11509.
43. Qamar Saeed, M., Dufour, N., Bartholmae, C., Sieranska, U., Knopf, M., Thierry, E., Thierry, S., Delelis, O., Grandchamp, N., Pilet, H. *et al.* (2014) Comparison between several integrase-defective lentiviral vectors reveals increased integration of an HIV vector bearing a D167H mutant. *Mol. Ther. Nucleic Acids*, **3**, e213.
44. Emiliani, S., Mousnier, A., Busschots, K., Maroun, M., Van Maele, B., Tempe, D., Vandekerckhove, L., Moisan, F., Ben-Slama, L., Witvrouw, M. *et al.* (2005) Integrase mutants defective for interaction with LEDGF/p75 are impaired in chromosome tethering and HIV-1 replication. *J. Biol. Chem.*, **280**, 25517–25523.
45. Gupta, K., Diamond, T., Hwang, Y., Bushman, F. and Van Dyne, G.D. (2010) Structural properties of HIV integrase. Lens epithelium-derived growth factor oligomers. *J. Biol. Chem.*, **285**, 20303–20315.
46. Sartori, A.A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J. and Jackson, S.P. (2007) Human CtIP promotes DNA end resection. *Nature*, **450**, 509–514.
47. Cornu, T.I. and Cathomen, T. (2007) Targeted genome modifications using integrase-deficient lentiviral vectors. *Mol. Ther.*, **15**, 2107–2113.
48. Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F. *et al.* (2011) In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells*, **29**, 1717–1726.
49. Soldner, F., Laganier, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S. *et al.* (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell*, **146**, 318–331.