Efficient Homology Directed Repair by Cas9:Donor Localization and Cationic Polymeric Transfection in Mammalian Cells

Authors:
Philip J.R. Roche[1], Heidi Gytz[2], Faiz Hussain[1], Christopher J.F. Cameron [1, 3, 4], Denis Paquette[1], Mathieu Blanchette[4], Josée Dostie[1, 3], Bhushan Nagar[2], Uri David Akavia[1, 3]

Affiliations:
1 Department of Biochemistry, McIntyre Medical Building, Room 815, 3655 Promenade Sir William Osler, Montreal, Quebec H3G 1Y6
2 Department of Biochemistry and Groupe de Recherche Axé sur la Structure des Protéines, Francesco Bellini Life Sciences Building, Room 464, 3649 promenade Sir-William-Osler, Montreal, Quebec H3G 0B1
3 Rosalind and Morris Goodman Cancer Research Centre, 1160 Pine Avenue, Montreal, QC Canada H3A 1A3
4 School of Computer Science and Centre for Bioinformatics, McConnell Engineering Building. Room 318, 3480 University, Montreal, QC, Canada, H3A 0E9

Abstract:
Homology directed repair (HDR) induced by site specific DNA double strand breaks (DSB) with CRISPR/Cas9 is a precision gene editing approach that occurs at low frequency in comparison to indel forming non homologous end joining repair. In order to obtain high HDR frequency in mammalian cells, we delivered donor DNA to a DSB site by engineering a Cas9 protein fused to a high-affinity monoavidin domain to accept biotinylated DNA donors. In addition, we used the cationic polymer, polyethylenimine, to efficiently deliver our Cas9-DNA donor complex into the nucleus, thereby avoiding drawbacks such as cytotoxicity and limited in vivo translation associated with the more commonly used nucleofection technique. Combining these strategies led to an improvement in HDR rates of up to 90% on several test loci (CXCR4, EMX1, TLR). Our approach offers a cost effective, simple and broadly applicable editing method, thereby expanding the CRISPR/Cas9 genome editing toolbox.

Introduction:
CRISPR/Cas9 is a gene editing approach directed by short guide RNA (sgRNA), which induces double strand breaks (DSBs) at specific genomic locations. DSBs are usually repaired in the cell by non-homologous end joining (NHEJ) or the rarer homology directed repair (HDR). NHEJ often generates deletions and insertions (indels) at the target site whereas HDR utilizes a donor DNA (dDNA) template to repair the lesion (Mali *et al*, 2013; Cong *et al*, 2013), resulting in single-point mutations or larger insertions (Schwank *et al*, 2013). However, the isolation of low frequency HDR-edited cell clones is time consuming and difficult. Current approaches often select Cas9 expressing cells via fluorescent markers and/or antibiotic resistance (Mali *et al*, 2013; Cong *et al*, 2013; Schwank *et al*, 2013; Van Trung Chu *et al*, 2015), to subsequently isolate only HDR-edited clones from the population of selected cells. While selection raises the percentage of successfully modified clones, it reduces the overall number of cells, and requires insertion of a relevant

selection marker, usually via plasmid. Without chemical or biochemical interventions, HDR rates of below 1% are considered normal (Miyaoka *et al*, 2016).

Pharmaceutical interventions that block NHEJ can also improve HDR. Chu et al. inhibited the action of DNA ligase IV or Ku70, which mediate NHEJ, using a variety of methods – including pharmacological (SCR7), protein-based (expression of viral Ad4) and/or shRNA. Using these approaches, decreases in NHEJ (and corresponding increases in HDR) of 4-fold, 8-fold and 5-fold were achieved, respectively (Van Trung Chu *et al*, 2015). Other studies have shown that cell cycle phase can influence DNA repair, where NHEJ occurs more often in the G1 phase and HDR only increases in late S/G2. For example, nocodazole arrested cells in S/G2 phase can exhibit 38% HDR with Cas9 ribonucleoprotein (RNP) (Lin *et al*, 2014). However, the applicability of this method is cell line dependent, since in the MDA-MB-4687 breast cancer cell modelline, nocodazole, vincristine or colchicine cause a G1 arrest (Holt *et al*, 1997).

HDR can be improved by optimisation of the dDNA homology arm (HA) length and structure. Zhang et al. demonstrated 26% HDR with double cut plasmid donors with long (600 bp) HA length (Zhang *et al*, 2017). Designing dDNA reflecting aspects of the Cas9 DSB mechanism also impacts HDR frequency. The Cas9 complex releases the distal strands (single strand or double) almost immediately post cut, while the proximal strand is released later (Shibata *et al*, 2017; Richardson *et al*, 2016). Richardson et al. (Richardson *et al*, 2016) benefitted from dDNA that was antisense to the released genomic DNA as well as from asymmetric HA for capturing resected genomic regions. These approaches improved HDR by 2.6-fold compared to symmetric single strand dDNA and 4-fold over double stranded dDNA; thus the overall rate of HDR (57%) surpassed chemical and genetic interventions (Richardson *et al*, 2016) (further summarized in Supplemental Table 1).

In summary, HDR efficiency ranges from fractions of 1% (Van Trung Chu *et al*, 2015; Lin *et al*, 2014; Miyaoka *et al*, 2016; Gutschner *et al*, 2016; Davis & Maizels, 2016) to ~60% (Richardson *et al*, 2016; Lee *et al*, 2017b), influenced by cell cycle (Gutschner *et al*, 2016), drug treatment (Van Trung Chu *et al*, 2015; Lin *et al*, 2014; Li *et al*, 2017), upregulation of DNA repair proteins (Rad51/52) (Shao *et al*, 2017), siRNA knockdown of NHEJ factors (BRCA1) (Davis & Maizels, 2016), donor design (Zhang *et al*, 2017; Richardson *et al*, 2016), timed protein degradation loci, cell type and transfection method. However, a consistent and common approach has not yet been defined. Many pharmacological interventions are possible in cell culture, however these interventions may not be appropriate *in vivo* (Lee *et al*, 2017a; Yanik *et al*, 2017).

Transfection methods also influence the percentage of modified cells, where the current state-of-the-art method is nucleofection. Here, the Cas9 RNP complex is delivered directly into the cell nucleus, which results in lower off-target DSBs (Kim *et al*, 2014; Kouranova *et al*, 2016) and reduced nuclease lag time (Lin *et al*, 2014), thereby improving both NHEJ and HDR. The advantage of using a pre-formed Cas9 RNP complex versus plasmid delivery is that it avoids the

lag time associated with mRNA and sgRNA transcription, Cas9 mRNA translation, and formation of intracellular RNP, balanced against the rates of sgRNA and protein degradation. Questions remain as to whether nucleofection is applicable in clinical settings other than ex-vivo editing and transplantation (Lin *et al*, 2017).

Cationic polymer RNP delivery of negatively charged conjugates is a potentially flexible delivery platform that has thus far drawn limited attention in CRISPR/Cas9 applications. To date, there have been only two novel attempts using this approach - sgRNA:donor chimeras and CRISPR-GOLD applying PAsp(DET) polyplexes (Lee *et al*, 2017b; 2017a). Taking into account cost, biodegradation (Wen *et al*, 2009), and proven nucleic acid transfection (Boussif *et al*, 1995), we decided to use Cas9 RNP delivery by polyethylenimine (PEI). PEI enters the cell by size dependent entry mechanisms (clathrin, calnexin or macropinocytosis) (Khalil *et al*, 2006), escaping endosomes via a proton pump effect and translocating to the nucleus (Oh *et al*, 2002). PEI has been observed to be favourable with regards to nuclear localization over poly-L lysine polyplexes (Pollard *et al*, 1998). Additional benefits of PEI are microtubular active transport, which increases Cas9 concentration at the peri-nucleus (Drake & Pack, 2008; Suh *et al*, 2003). Editing by Cas9 could be improved by an increase in intranuclear or perinuclear RNP benefitting from both PEI and importin mediated (protein nuclear localization sequences [NLS]) nuclear trafficking concurrently (Mali *et al*, 2013; Cong *et al*, 2013).

In this paper we describe a proof-of-concept HDR system that may provide a simple and broadly applicable gene editing solution. The keys to our system of improved HDR efficiency are combining Cas9 engineered donor colocalization (Cas9 mono-avidin:biotin dDNA) and application of dual nuclear localization modes (PEI and NLS), such that the probability of colocalizing active RNP with dDNA at the time of DSB is increased.

Results:

Protein Engineering and Characterisation of PEI:Cas9MAV Polyplexes

We speculate that the current low HDR frequency by CRISPR/Cas9 is primarily due to the inefficient delivery of Cas9 to the nucleus and/or inefficient delivery of dDNA to the DSB. To enhance Cas9 delivery to the nucleus, we used PEI. In order to optimize delivery of dDNA to the DSB, we forced the conjugation of the dDNA to Cas9 through the high affinity binding system. The avidin-biotin system is facile to implement compared to multi-step chemical conjugation of sgRNA and dDNA to create chimeras, as the biotin modification can be introduced by a PCR primer extension. Previous studies (Lee *et al*, 2017b) reported that colocalizing the dDNA using a chimeric crRNA improved HDR, but not significantly compared to cycle arrest strategies (10-35%) (Lee *et al*, 2017b; Jinek *et al*, 2012). In addition, the crRNA:dDNA chimera was exceptionally short in length (crRNA length 42, dDNA length 87 including the HAs), which could hinder the ability of the dDNA to interact with the DSB site. We hypothesized that HDR can be further improved by increasing the linker distance between the conjugated dDNA and the protein,

thereby providing additional flexibility for the dDNA to capture the released genomic DNA. Thus, we engineered a Cas9-monoavidin (Cas9MAV) construct, whereby Cas9 already containing dual NLS sequences, was further modified by MAV at its C-terminus with a 17-amino acid linker between Cas9 and MAV (Fig.1A). The detailed procedure for the formation of the modified Cas9 is described in the methods section. Combined with biotinylated dDNA, Cas9MAV should increase the local concentration of the dDNA with improved flexibility. The rationale for choosing MAV as opposed to streptavidin, which exists as a tetramer and could deliver up to four dDNA components, was to maintain tight control on number of dDNAs being delivered (i.e. exactly one dDNA per Cas9) to the DSB.

The size and molecular weight (MW) of Cas9MAV was determined by size-exclusion chromatography-multi-angle light scattering (SEC-MALS) (calculated MW: 175.474 kDa, measured MW: 174.5 kDa (Fig.1A)). The hydrodynamic radius was estimated to be 9.5 nm which is in accordance with the presence of the flexible NLS-linker-MAV tether in Cas9MAV and a previously published Cas9 hydrodynamic radii (Carlson-Stevermer *et al*, 2017; Jinek *et al*, 2014; Nishimasu *et al*, 2014).

Before attempting colocalization intracellularly, formation of a Cas9MAV complex bound to biotinylated dDNA must be first demonstrated *in vitro*. Analysis by dynamic light scattering (DLS) (Fig.1) established that near complete conjugation of donor to protein was achieved, ensuring that the predominant protein species is conjugated and ready to provide dDNA for HDR repair. DLS was also used to validate polyplex formation. DLS measures light scattering generated by particles representing size, distribution and polydispersity. Sample mean intensity distribution can be transformed to a mean volume distribution (sample population spatial occupancy), and a mean number distribution (number of particles per population per volume). DLS determined that Cas9MAV:sgRNA is monodisperse with a 9 nm radius correlating well with SEC-MALS results of Cas9MAV alone (above, Fig. 1C and 1D). The donor DNA tested here was the 800 bp long CF3b (5' biotinylated) and has an approximate radius of 30 nm. Preformed Cas9MAV:sgRNA RNP was mixed with CF3b in a 1:1 molar ratio and transferred back to the cuvette for DLS measurements. CF3b was able to complex completely with Cas9MAV:sgRNA as no residual free CF3b dDNA could be detected upon RNP addition (Fig 1C and 1D). It was observed that the radius of the final Cas9MAV:sgRNA:CF3b RNP complex was smaller with a hydrodynamic radius reduced to 6 nm, although a small population was observed at ~ 16 nm. We speculate that the reduction in the RNP radius of gyration is due to the conjugation of the negative phosphodiester backbone to the MAV portion of Cas9MAV, which presumably restricts the free rotation of the dDNA. In all cases the 1:1 molar ratio of the interactions of the sequential RNP formulation is confirmed by the absence of the free macromolecule populations (i.e. CF3b and Cas9MAV:sgRNA populations disappear when sgRNA:Cas9MAV:CF3b forms).

DLS confirmed the qualitative assessment by agarose gel shift of Cas9MAV:biotin donor complexation. Fig.1B shows the donors with and without Cas9MAV (purple and green boxes, respectively). When Cas9MAV and the biotin donors were loaded on an agarose gel, no migration was detected due to the increase in molecular weight of the complex.

Next, DLS confirmed encapsulation of the RNP complex by PEI (Fig.1E and 1F). PEI was added directly to the sample described in Figure 1C and 1D and incubated at room temperature for 10 min before DLS measurements. By intensity and volumetric measurements, we observed two species of 134 nm ($^{+}$/-10) and 530 nm ($^{+}$/-10) radii. The Cas9MAV:sgRNA:CF3b population was no longer detectable, indicating that it formed a polyplex with PEI. Considering the number of species in each population, PEI polyplexes of 134 nm radius were the most prevalent. Thus, based on this radius and prior knowledge of PEI, it is most likely that calnexin-mediated and macropinocytosis mechanisms (Khalil *et al*, 2006) are employed by the Cas9MAV:sgRNA:dDNA:PEI polyplex to cross the cell membrane in transfection experiments, since these mechanisms are more compatible with objects of 130 nm.

High transfection efficiency with PEI:Cas9MAV, resulting in high indel frequency in human and murine cell lines

High transfection efficiency must be obtained for the proposed Cas9:MAV system to function. To track transfection of the PEI:Cas9MAV complex, we used fluorescent PEI. Fig.1G, left panel shows fluorescent light microscopy of non-transfected and transfected cells with green fluorescent PEI:Cas9MAV complex with no dDNA (see supplemental methods). This confirms that transfection has occurred qualitatively (positive in the green fluorescence imaging channel). Quantitative measurements were performed using flow cytometry and 95% of cells were shown to have been transfected by the (carboxyfluorescein) FAM-labelled PEI:Cas9MAV complex (Fig.1G, right). With the demonstration of successful cellular uptake and DLS experiments verifying intact RNP formation, transfection and nuclease activity can now be investigated in multiple cell lines.

Six cell lines (Human: HEK293T, MCF-7, HeLa, U2SO and Mouse: 4T1, 3T3) were chosen to reflect both easy- (HEK293T, HeLa) and difficult-to-transfect (4T1, MCF7) cells, two species, and a variation in tissue of origin/disease (kidney, breast cancer, cervical cancer, osteosarcoma). To examine multiple loci, RNPs were formed complexing Cas9MAV with sgRNA for FLCN, ACC1 and EMX1 for human cell lines (Fig.1H) and FLCN & ACC1 for mouse cell lines (Fig.1I). NHEJ indels were introduced in all cell lines by transfection with PEI:Cas9MAV with no dDNA, and the frequency was determined by a T7 endonuclease assay. Indel frequencies (50-70%) showed broad equivalence across the human cell lines (Fig. 1H). In mouse 4T1 and 3T3 cell lines, two loci were tested (ACC1 and FLCN) and indel frequencies of 50-60% were observed (Fig. 1I). These results demonstrate that the PEI:Cas9MAV could be successfully introduced to other cell lines besides HEK293T.

High Rates of HDR achieved by combining PEI and Cas9MAV

To investigate the impact of PEI nuclear co-delivery and the impact of Cas9MAV upon donor design, we outlined a simple comparison experiment at the CXCR4 loci, focusing on four double stranded dDNA designs (Fig. 2A), with biotinylated and non-biotinylated versions. Each donor was designed with a HindIII restriction digest sequence within 5 bp of the DSB. The intention was to explore the design rules for double stranded donors using equal HAs of 600 bp (CF1), 400 bp (CF3) and asymmetric short donors (Fas1 and Fas2), the latter being sub-optimal designs as their HAs are significantly truncated. In our initial experiments (Fig. 2B) we assessed the indel frequency of non-biotinylated donors, finding that with short asymmetric dDNA, indel formation persisted. This was observed to occur for both Cas9MAV and a version of Cas9 lacking MAV, but still possessing dual NLSs (Cas9NLS), as well as for cells treated with the drug, nocodazole (200 µg/ml). Overall Fas1 and Fas2 averaged 57% and 54% indels, respectively, whereas CF1 and CF3 averaged only 5% and 24%. CF1 dDNA saw no observable indels with Cas9MAV and nocodazole treated variants.

Nocodazole is intended to aid HDR by cell cycle arrest. For non-biotinylated donors (CF1, CF3, Fas1 and Fas2) using RNPs formed with Cas9NLS or CAS9MAV (Fig. 2C), we observed no significant difference between drug treated and non-drug treated cells. Shorter non-biotinylated donors generally result in lower HDR percentages, as they are less likely to form stable, hybridised species with genomic DNA than longer donors;. however, these donors still benefit from PEI nuclear deliver as shown in (Fig. 2B).

An improved HDR percentage was observed for all donor constructs that were biotinylated, including the shorter Fas1b and Ras2 (biotinylated version of Fas2) (Fig.2D). Having excluded the need for drug treatments, we conclude that the HDR improvement we observe is a consequence of Cas9MAV and PEI transfection. Figure 2D shows an improvement in HDR percentage by virtue of colocalization when using shorter donor constructs. Figure 2E summarizes the overall impact of localization by Cas9MAV and shows that at the CXCR4 loci, Cas9MAV utilisation contributes a 20% improvement in HDR rates for all donors in comparison with Cas9NLS.

We wished to validate the performance of the biotinylated dDNA and Cas9MAV by investigating whether the genomic edits persisted at the mRNA level. Using Fas1b and its associated HindIII site to edit cells, which were split for DNA and RNA extraction at the point of cell harvesting. DNA extraction and PCR were performed and insert validated again by restriction digest (Fig. 2F). Genomic DNA was degraded by DNAse1 treatment before reverse transcription (see supplemental methods). Two primer sets for two separate extractions demonstrated the presence of the edit on the mRNA level (Fig. 2G, H). This enabled us to determine by cDNA derived PCR products, that restrictive cleavage was present and the edited DNA sequence had been transcribed to mRNA, demonstrating the persistence of the edit. Furthermore the edits persist after significant dilution

and re-growth of cells post passaging (Supplemental Figure S2). Samples edited with CF1, CF3 and Fas1/Ras2b were prepared for Sanger sequencing and all samples returned the correct sequence (Fig. 2I).

Traffic Light Reporter Assay

Given the high HDR frequencies observed on the CXCR4 loci, we wanted to demonstrate the versatility of our PEI:Cas9MAV system by applying it to other loci. It was important not to choose loci and commensurate sgRNAs of equivalent performance as this would not give a clear evaluation of the methodologic generality of the system, nor provide insight on the influence of variables such as loci accessibility and sgRNA performance. We thus chose EMX1 and a Traffic Light Reporter (TLR) system (Certo *et al*, 2011), to obtain a more realistic appraisal of performance.

The CXCR4 loci is usually associated with high indel frequency. In seeking to check if our method was applicable at loci with low indel frequency, the TLR assay appeared an appropriate choice, since indel frequency observed in the TLR assay is usually historically low (12%) (Robert *et al*, 2015). Evaluation was performed using a Cas9NLS plasmid control as an indel forming positive control (Supplemental Figure S1). We observed a decline in NHEJ (0%) (Fig. S1) at the 48 hr time point (5'b and 3'b dDNA variants (Supplemental Table S2)) and HDR frequencies of 1-3%. At the 96 hr time point complete HDR of DSBs will have occurred, and there has been sufficient time for the Cas9 plasmid NHEJ control to reach an analytical signal (RFP expression) for flow cytometry. A frequency of 17.1% for NHEJ was observed at 96 hrs for the Cas9 plasmid control reflecting previous NHEJ frequencies with this assay (Robert *et al*, 2015), and a fluorescent microscope image was included for reference in Figure 3 as a qualitative assessment of expression. We observed that it was possible to detect with both fluorescence microscopy and flow cytometry the mixed NHEJ/HDR population in cells edited with the PEI:Cas9MAV system (Fig. 3A, 3B, 3E and supplemental Figure S1). NHEJ frequency for PEI:Cas9MAV at 48 hrs ranges from 0.01-0.14%, but increases to 0.23-0.61% at 96 hrs representing a small heterogeneous population in edited cells. HDR frequency for 5'b, 3'b and dual biotin modified 5'3' donors with the PEI:Cas9MAV system by 96 hrs gave a range of 20-32% HDR with an average HDR percentage of 25% for all donors. NHEJ was a minor contributor to edits with less than 2% for all biotinylated donors. Comparison of 5'b donor (96 hrs) and control (96 hrs) are given as representative examples in Figure 3. Full plots are presented in supplemental Figure S1A,B and tabular summary of results are presented in supplementary Fig. S1C.

EMX1

The HDR frequencies achieved by our system in both CXCR4 and TLR experiments represented an improvement on previous pharmacological interventions (Van Trung Chu *et al*, 2015; Certo *et al*, 2011). In keeping with the trends observed for CXCR4 gene loci (Fig. 2), we surmise that EMX1 indel frequencies (Fig. 3I) are a metric for sgRNA quality and genomic loci accessibility.

Based on the indel frequencies we observed, we predict a lower HDR percentage than observed for CXCR4, but greater than TLR. We considered FLCN as an alternative, but chose the loci:sgRNA combination with lower indel frequency (Fig. 3I). An additional stringency was introduced to the dDNA by truncating the equal HAs to 200 bp, on either side of a 6-repeat stop codon sequence and BamHI site. The donor is compromised in comparison to the ideal HAs of 400 and 600 bp used for CF3/CF1. BamHI was used to rapidly assess the percentage of HDR occurring (Fig. 3K). The experiment involved two donors (Fig. 3K), where biotinylation at the 5' end of the forward primer is denoted by 'b'.

The non-biotinylated donor generated an HDR percentage of 60% with high (<70%) indel formation, suggesting mono-allelic editing; partial incomplete HDR leading to indel formation and reflecting the semi-quantitative inaccuracies of gel-based assessments. This is similar to the observation of the short double stranded donors used in CXCR4 investigations (Fas1/Fas2 137 bp) and in the TLR study (111 bp total) generating lower HDR occurrence. Biotinylated EMX1 donor was used with the PEI:Cas9MAV system, resulting in indels dropping to less than 10% and HDR occurrence increased to 90% (Fig. 3 L&M).

We also wished to consider the impact of donor DNA being inserted randomly where the Cas9 DSBs were generated. A simple experiment was performed where the sgRNA for EMX1 was mismatched with a biotinylated CF3 CXCR4 donor. EMX1 retained its capacity for indel generation (Fig. 3I) and the absence of insert was validated as HindIII digestion did not result in cleavage (Fig. 3J). While not a complete proof of non-insertion where Cas9MAV cleaves DNA, it offers some indication that non-matched donors do not result in HDR.

PEI Transfection Method and Cas9MAV impacts to HDR: A comparison with Lipofectamine 3000

Central to our system are PEI transfection and MAV modification offering an alternative nuclear trafficking approach and dDNA-Cas9 co-localization at the site of the DSB, respectively. Nucleofection was shown to result in equal indel formation at CXCR4 loci (Fig. S3) and an improvement in indel formation at EMX1. Comparing Cas9NLS and Cas9MAV with nucleofection suggests that nucleofection renders the NLS sequence redundant. Early experiments with the lipofection technique suggested that nuclear localization is a minor aspect of lipofection with cytosolic delivery being most common for DNA payloads (Akita *et al*, 2004). We analyzed transfection by Lipofectamine 3000 as a substitute method for PEI. Lipofection shares many characteristics with PEI, such as payload (DNA/protein) protection from metabolic degradation, high transfection rates and low cytotoxicity (Fig. S6). Liposomes escape lysosomal degradation by diffusing freely within cells, rather than engaging in microtubule transport (Cardarelli *et al*, 2016). However, lipofection lacks nuclear localization and microtubule transport, and thus provides an ideal comparison transfection method to PEI along with other derivatives such as CRISPRmax that achieve high transfection and HDR editing (~20%) (Yu *et al*, 2016).

Figure 4A and 4B detail the lipofectamine experiments conducted with the CXCR4 loci using that same donors and editing system as in Figure 2. Each plate was transfected with RNP:LP3K, and biotin or non-biotinylated donors, respectively. DNA extraction from Fas2 donor treated cells failed, so comparison of this donor and its biotinylated variant, Ras2 were excluded. Figure 4A compares indel frequencies between biotin and non-biotinylated donors. It is observed as before (Figure 2 and 3 for CXCR4 and EMX1, respectively), that indels decline where biotinylated donors are used with CF1 being an aberration in this instance with persistent indels. Decreased indel formation is indicative of increased HDR. This was confirmed by Fig. 4A and 4B, where biotinylated donors outperform non-biotinylated donors (CF1:CF1b = 85% to 89%; Fas1:Fas1b = 55% to 90%; CF3:Cf3b = 46% to 92%), but do not achieve the same HDR percentage as those cells transfected by PEI (~99% Fig. 2). The difference between the biotin/non-biotinylated donors suggests that donor DNA colocalization to DSB increase HDR. Finally, the impact of PEI was considered. LP3K effectively enables the donor delivery to the cell's cytoplasm, and the substantially lower HDR rates observed for non-biotinylated donors compared to those observed with PEI suggests that the nuclear localization potential of PEI plays a significant role. If sufficient donor DNA is present in the nucleus, the chance of interaction between donor template and distally released genomic DNA from the Cas9 RNP increases, but would likely be governed by a random walk diffusion regime as tethered localization through Cas9MAV is not enabled. Further mechanistic evaluation is beyond the scope of this paper, but both PEI and MAV modifications appear to have statistically relevant effects on the HDR frequencies achieved in this work (Fig.2).

NGS sequencing Confirmation of Editing Frequencies

To consider the frequency of HDR at a greater depth of sequencing we designed an Ion torrent sequencing approach. We took 7 CXCR4 sample experiments covering HDR (with and without biotin donor), and unedited controls (Tables S5, S6). Figure 4C presents an overview of the libraries sequenced and Supplemental Table S4 details the variables of each experiment sequenced. 'FControl' returned a clear unedited population of sequences, replicating the canonical genomic sequence for CXCR4. Our principal interest was to evaluate the sequences returned for complete edits (full insertions of donor sequence) and partial edits (combination insert and indel formation). When compared to all other libraries (FBest, F60, BC4,5,6), the rate of HDR edits becomes apparent showing that HDR has occurred at a 90% or greater rate (Fig. 4C&D). One concern with using non-sequencing approaches such as restriction digests for HDR or T7 for indel formation is the under- or overestimation of the editing that has occurred. The FBest library (CXCR4 with Cf1b donor) is a good indication of correlation with gel-based analysis given in Figure 2. These results suggest that suppression of indels and high editing percentage was achieved for the complete insert, with less than 5% being partial inserts/indels. In this case, the restriction digest provides a good guide for donor design performance, mirroring the indel suppression and high HDR percentage observed by NGS. When Fas1, a non-biotinylated asymmetric donor was analysed (F60 library), we observe an unedited fraction persisting, but at a lower percentage (~2%) than was

estimated from the average restriction digest value (Fig. 2C). It is worth considering that the potential for higher editing exists for Fas1 as the standard deviation margin of error suggests that even by gel analysis, the editing frequency ranges broadly (~20%).

The existence of a persistent partial HDR insert in all libraries excluding control suggests logically that no process is 100% perfect in its ability to insert at a loci. While the analytical approach (Ion Torrent) is prone to generation of indels, as are other sequencing by synthesis approaches, the partial insert frequency is greater than either enzymatic amplification errors or misreads. Partial insert reads are only observed in HDR experiment libraries suggesting that these reads are due to incomplete templating in HDR, though the mechanism cannot be speculated upon without further investigation.

Discussion:

We demonstrated the potential for a cationic polymer delivery of PEI:RNP approach for a streamlined and efficient Cas9 HDR methodology. The potential to utilize polymeric delivery leveraging specifically the nuclear localization and high transfection rates appear to be a contributing factor in the improvement of HDR rate observed at the CXCR4, TLR and EMX1 loci tested, while generating high frequency NHEJ at other loci (FLCN, 2 sgRNA targeting ACC1, and EMX1) when dDNA is not present. Further work is required to unravel the percentage contributions of the nuclear localization effect in HEK293T and other cell lines, and how generalizable the principle is to other cell lines such as embryonic stem cells with respect to RNP delivery. Additional studies should be performed on RNP delivery relating to polymer derivatization and polyplex size (Ogris *et al*, 1998). These polyplex properties have been shown as key determinants of DNA transfection in different cell types broadening the PEI:RNP transfection approaches with the CRISPR/Cas9 system.

With respect to Cas9MAV and the 20% HDR efficiency increase observed over Cas9NLS without donor biotin mediated co-localization (Figure 2), PEI improves non-biotinylated dDNA delivery to the nucleus, but the monoavidin:biotin conjugate appears to relax further the design rules that usually operate for template donor during HDR. In previous studies, the necessity for double stranded dDNA with HAs of 400-600 bp has been required to achieve high HDR frequencies (Lin *et al*, 2014; Zhang *et al*, 2017). Using Cas9MAV, asymmetric short double stranded dDNA (137 bp, HA 37 bp and 93 bp, respectively) can achieve equal efficiency of HDR, mitigating the suboptimal design.

While the purpose of this research was to avoid the application of nucleofection/electroporation techniques that induce increased cell death in the cell line studied, the Cas9MAV RNP could be delivered by nucleofection as an alternative method (Supplemental Figure S3), however, at the cost of cell viability. Nucleofection may lead to problems in the potential for translating the transfection method into *in vivo* studies, since animals have a long history of electroporation

(Adachi *et al*, 2002), but human studies are still under development (Jiang *et al*, 2015) in comparison to viral and non-viral transfection methodologies.

To codify the experimental results, the loci:sgRNAs combinations chosen were loosely termed the 'good', the 'bad' and the 'ugly', by indel frequency without HDR. The 'good' refers to a loci where the sgRNA scores highly in design programs and indel frequency is in excess of 60-80% as indicated by T7 endonuclease assay. We hypothesize that for HDR, a consistently high occurrence of DSBs is required with high recognition of target by sgRNA. We acknowledge that T7 assays cannot reflect conservative DSB repairs, but offer only an approximate measure of cut efficiency. This approximation refers specifically to the performance at CXCR4. The 'bad' is a loci (EMX1) where the indel frequency is around 60% denoting lower DSB frequency or higher conservative DSB repair. Success was determined by achieving HDR above 60% to be consistent with the hypothesis that indel and HDR frequencies in this system invert. The 'ugly' is an HDR assay using the traffic light reporter system (Certo *et al*, 2011), that in literature examples exhibits very low HDR frequency ranging from 1-4%. Correspondingly, TLR experiments show comparatively low NHEJ at 12% or less, even with Cas9 plasmid systems, reflective of the loci:sgRNA combined inefficiency. In this case, the determination of success was the suppression of NHEJ in favour of HDR and achieving HDR in excess of 4% without drug treatment, siRNA knockdowns, or other interventions (Robert *et al*, 2015). In the three loci tested, we observed high percentages of HDR occurrence and suppression of NHEJ, with varying degrees of HDR between the loci due to experimental factors. We caution that any Cas9 experiment's final editing efficiency will be governed by a combination of two factors: 1) loci accessibility (condensed/uncondensed chromatin) and sgRNA quality (sequence recognition) will govern the ability to create DSBs, whereas 2) donor sequence and insert size influence the templating repair itself. While the latter is partially mitigated by application of the PEI:Cas9MAV system, we observed through the TLR experiment the importance of efficient sgRNA, since in this experiment the global HDR percentage was only 25% reflecting the combined effect of loci and sgRNA quality.

The ability of the Cas9-MAV system to remove many of the donor design constraints (length, asymmetric, single/double strand, HA length), offers an intriguing opportunity we hope to explore by using other loci, as well as the size of non-genomic DNA spliced between the HAs. Our results do not stand in isolation. Recent publications exploring the role of donor localization to the nucleus and importantly, DSB site (Carlson-Stevermer *et al*, 2017; Ma *et al*, 2017), offer support that our system is affecting the mechanisms of DSB repair. It is uncertain at this point which repair pathway and protein effectors of repair, such as DNA-ligase, Ku heterodimer, Rad51/51 among others, are involved and at what stage. Figure 4D summarises our current model, focused on the benefit derived from biotin mediated donor localization with two nuclear delivery approaches.

In contrast to other CRISPR/Cas9 methodologies involving plasmid or mRNA delivery, our system does not require the cell to transcribe sgRNA and translate protein, which can impede RNP

intracellular assembly. By coupling donor to Cas9, we remove the issues of a freely diffusing donor (plasmid or linear) and compartmentalisation of the donor in the cytosol. If the donor is not present, it is more likely that canonical DSB repair (NHEJ and conservative blunt end repair(Bétermier *et al*, 2014)) will occur rather than HDR. Recent work using aptamer:streptavidin localisation introduced ssODN to nuclei (Carlson-Stevermer *et al*, 2017), achieving ~6% HDR. This rate is lower than what could have been achieved with extracellular RNP assembly and hence an increased concentration of functional RNP:donor conjugates within the cell. In future work we hope to extend the transfection principle and Cas9 methodology to difficult-to-transfect cells other than MCF-7 & 4T1, and develop a cell-type specific mechanism for transfection amelioration of cationic polymeric delivery.

Materials and Methods:

Cell culture

HEK293T, MCF-7, 3T3, HeLa and U2OS cell lines were all cultured in DMEM supplemented with 10% FBS, 100 U/mL penicillin, and 100 U/mL streptomycin and were maintained at 37°C and 5% CO2. The exception being 4T1 which was cultured in DMEM with 10% FBS, 10mM HEPES, 1mM Sodium Pyruvate, 36mM $NaHCO_3$, 5 µg/mL amphotericin, 0.5% gentamicin and 1% L-glutamine. Media, trypsin and FBS were supplied by Wisent. Cells were kept at low passage for experimentation, not exceeding 10 passages before starting fresh cultures from frozen stocks.

For transfections, seeding density was $3x10^5$ cells per 6 well and $1x10^5$ for 12 well plates, prepared on the day prior to PEI transfection. Confluency of 60-70% at time of transfection was our objective, so for fast growing (HeLa) or larger cells (U2SO), seeding densities of $1x10^5$ and $0.8x10^5$ respectively were used for preparation for 6 well plates.

Cell Viability

Cell viability was determined by trypan blue staining in a 1:1 v/v ratio to 5µl sample of harvested cells. Staining was verified by Countess II or visually by haemocytometer. All cells were maintained at 97% viability post transfection.

sgRNA

Full sequences and synthesis by in vitro transcription are described in Supplemental Methods.

PEI preparation

100mg of linear PEI (polyscience) was dissolved in 100 ml nuclease free milliQ water. Solution was magnetically stirred for 4hrs at room temperature in a sterile duran bottle. Once the polymer was dissolved pH was adjusted to pH7.4 using concentrated HCl. Secondary amine protonation was assessed by ninhydrin assay by taking aliquots. Solution was filtered (0.22um) and aliquoted to 1.5 ml eppendorf tubes and stored at -20°C. Aliquots were removed as required for experiments and stored at 4°C for polyplex formation, with a maximum usage period of two weeks. PEI

fluorescent labelling (FAM-PEI synthesis), secondary amine assay and transfection evaluation are described in Supplemental Methods.

Cloning of SpCas9-NLS-mono avidin

SpCas9 constructs with a C-terminal 17 aa linker followed by monoavidin were cloned by an overlap extension PCR. Primers are listed in Supplemental Info, Table S2. Fragment 1 was amplified using an internal EcoRI site in pMJ915 (Lin *et al*, 2014) and primers SpCas9_EcoRI_445_for and NLS_18linker_rev using iProof High-Fidelity DNA Polymerase (Biorad, USA). Fragment 2 was amplified from pRSET-mSA encoding monoavidin for high affinity binding of one biotin (Lim *et al*, 2013) using primers 18linker_avidin_for and avidin_NotI_rev, and VENT DNA polymerase (NEB, USA). Fragments 1 and 2 were combined and subjected to 10 x overlap extension, before primers SpCas9_EcoRI_445_for and avidin_NotI_rev were added for amplification. The resulting fragment was digested with EcoRI and NotI, and ligated into pMJ806(36) with Kanamycin resistance.
pMJ915 and pMJ806 were gifts from Jennifer Doudna (Addgene plasmid # 69090 and #39312)
pRSET-mSA was a gift from Sheldon Park (Addgene plasmid # 39860)

The resulting fusion construct contained an N-terminal hexahistidine-maltose binding protein (His-MBP) tag, followed by a tobacco etch virus (TEV) protease cleavage site and wild type SpCas9 with two C-terminal SV40 nuclear localization signals, and lastly an 17 aa linker to monoavidin. Plasmid Map of Cas9MAV included in supplemental data SD2

Purification of Cas9 proteins

SpCas9 fusion constructs were expressed and purified essentially as described previously (36). Briefly, proteins were expressed in BL21(DE3) Rosetta2 cells grown in LB media at 18°C for 16 h following induction with 0.2 mM IPTG at OD_{600} = 0.8. The cell pellet was lysed in 500 mM NaCl, 5 mM imidazole, 20 mM Tris-HCl pH 8, 1 mM PMSF and 2 mM B-me, and disrupted by sonication. The cleared lysate was subjected to Ni affinity chromatography using two prepacked 5 mL HisTrap columns/1-2 L cell culture. The columns were extensively washed first in 20 mM Tris pH 8.0, 250 mM NaCl, 5 mM imidazole pH 8.0, 2 mM B-me and after, in 20 mM HEPES pH 7.5, 200 mM KCl, 10 % glycerol, 0.5 mM DTT, before elution with 250 mM imidazole. The His-MBP affinity tag was removed by overnight TEV protease cleavage w/o dialysis.

The cleaved Cas9 protein was separated from the tag and co-purifying nucleic acids by purification on a 5 mL Heparin HiTrap column eluting with a linear gradient from 200 mM - 1 M KCl over 12 CV. Lastly, a gel filtration on Superdex 200 Increase 10/300 GL in 5% glycerol, 250 mM KCl, 20 mM HEPES pH 7.5 separated the nucleic-acid bound protein from the clean SpCas9 protein. Eluted protein was concentrated to ~10 mg/mL, flash-frozen in liquid nitrogen and stored at -80°C.

RNP Formation

2.1µl of 1x phosphate buffered saline (sterile and 0.22um filtered), 1.7µl of Cas9 (11mg/ml) or Cas9MAV, 2.1µl of 300nM to 900nM sgRNA (concentration varied with respect to Cas9 molarity to maintain 1:1 ratio) were combined in a sterile PCR tube, vortexed gently and incubated for 20 minutes at 25°C. For HDR experiments 5-20µl of 300ng/µl biotinylated or non-biotinylated donors are added at 15 minutes into incubation, with the concentration added adapted to concentration of Cas9MAV used. Biotin:monoavidin association occurs within 5 minutes.

Donor DNA

Donor DNA was created by PCR amplification from gblocks (CXCR4 and EMX1 loci) and for TLR donor from a plasmid kindly donated by Francis Robert from the Pelletier Lab (Department of Biochemistry, McGill). Details of primers and gblocks are found in Supplemental Data. For biotinylated donors, substitution of forward or reverse primers with 5' biotinylated modifications was performed. All oligos were quantified after PCR using nanodrop and concentration of working stocks were 300 ng/µl in all cases. Oligos were used without any further purification. Oligonucleotides used in this research were purchased from IDT (Integrated DNA technologies, USA) and BioCorp (Montreal, Quebec).

PEI:RNP Polyplex Formation

After formation of RNP, 30µl of PEI (25 kDa 1mg/ml) was added to the RNP and vortexed. The resulting solution was incubated for 20 minutes at 25°C. After incubation 71.5 µl of DMEM was added and tubes incubated at 37°C for 20 minutes to temperature equilibrate with cells to be transfected. Polyplex solution was added dropwise to cells. The above protocol functions for transfecting one well of a 6 well plate, or three wells of a 12 well plate. In the case of larger transfections, quantities were scaled by the estimated cell density of culture vessel. In the case of 6 well plates we estimate 1 million cells at point of transfection.

SEC-MALS

A size-exclusion chromatography multi-angle light scattering (SEC-MALS) experiment for Cas9MAV was prepared by injecting 50µl sample at 3 mg/ml onto a Superose 6 increase 10/300 GL column (Sigma) equilibrated in 20 mM HEPES, pH 7.5 and 150 mM KCl at a flow rate of 0.3 ml/min. The eluted peak was analyzed using a Wyatt miniDAWN TREOS multi-angle light scattering instrument and a Wyatt Optilab rEX differential refractometer. Data were evaluated using the ASTRA 5.3.4 software with BSA at 5 mg/ml as a reference.

Dynamic Light Scattering (DLS)

Cas9MAV/sgRNA/dDNA complex hydrodynamic radius and PEI polyplex dimensions were evaluated using a Malvern Nanosizer S dynamic light scattering instrument. Measurements were performed with biotinylated donor DNA concentrations of 2 uM, and preformed Cas9/sgRNA concentrations were adjusted accordingly in a 1:1 molar ratio. All samples were measured in the same buffer (150mM KCL, 20mM HEPES pH 7.5) at 20 µl in a Zen2112 cuvette. The flow cell

was equilibrated to 37°C to mirror the conditions of cell culture that polyplexes and Cas9 variant would exist during transfection. Measurements were performed for 1 min with a minimum of 15 measurements. PEI conc was 1mg/ml.

DNA extraction

Cells were harvested using 0.025% trypsin (0.5ml per well for 12 well and 1ml for 6 well plates) after a 1xPBS wash and centrifuged in 1.5ml eppendorfs at 8000k for 5 minutes. Supernatant was discarded and pellet washed with 0.5ml PBS. Cells were spun again and supernatant discarded. Full protocol for extraction is detailed in Supplemental Methods. Homemade lysis, column binding, wash and elution buffers compositions are in Supplemental Data. Proteinase K and RNase 1 were sourced from Sigma Aldrich.

RNA extraction

For cellular RNA Trizol extraction following manufacturer's protocol (Thermofisher, USA) and Direct-zol clean up columns (Zymo Research) were used for preparation of RNA for RT-PCR validation of genomic edits. DNAse1 treatment for enzymatic degradation of genomic DNA was performed following manufacturer's protocol. DNA degradation was verified by gel electrophoresis. DNase1 and buffer was supplied by NEB, USA.

PCR

Genomic PCR was performed with either Q5 2X master mix or Phusion polymerase (NEB, USA). PCR was used for donor DNA creation using IDT gBlocks as templates. All oligos used are specified in Suppplemental Table S2.

RT-PCR

Reverse transcription was performed using the High reverse transcriptase kit (Biobasic, Canada). For PCR step 2 microliters of solution were used in the standard Phusion master mix protocol (NEB, USA).

Gel Electrophoresis

For analysis of PCR products, and enzymatic digests, 2% agarose (Biobasic, Montreal) gels were prepared and a Fluoro-loading dye (Zmtech Montreal) was used to visualize the DNA (2µl per 25µl volume). Gels were run at 90 volts in 1x TAE buffer. For RNA analysis urea polyacrylamide gels (12%) were used and run at 120V for one hour with a pre-run prior to loading of half an hour at 130V in 1X TBE buffer. RNA gels were stained with a Ethidium bromide staining solution (10µl EtBr 10mg/ml in 30ml TBE). Gel images were recorded using UV illumination and imaging gel dock.

T7 Endonuclease Indel assay

T7 Indel assay was performed according to the protocol of Guschin et al (37) with products being quantified by agarose gel electrophoresis. PCR amplification of genomic edited and unedited DNA was performed using Q5 polymerase. All T7 endonuclease assay samples were performed with minimum of 3 biological replicates. ImageJ was used for analysis and quantification of band intensity detailed in Supplemental Methods.

Restriction Digest HDR assay

HDR frequency was assessed by restriction digest analysis. Genomic loci chosen were analysed by NEBcutter 2.0 to identify any restriction digest sequences in endogenous DNA. Excluding those enzymes with palindromic sequences present, specific enzymes without palindromic sequences were used to create a donor inserts. In the case of CXCR4 and EMX1, the donor sequence TAG-HindIII and BamHI (respectively, see Figure 2 and 3) were inserted into gblocks with homology arms matching the genomic loci. Details of gblock sequence and subsequent PCR method for creating donors of various lengths are included in the supplemental data. Restriction digest was performed after amplification from DNA extracted from edited and unedited cells following the manufacturer's protocol (NEB) for cut smart High fidelity restriction enzymes. All HDR restriction digest assays were performed with a minimum of 3 biological replicates. In brief, 5 µl of PCR amplicon was diluted in 35µl nuclease free water and 5µl cut smart buffer. 1µl of high fidelity restriction enzyme (HindIII or BamHI) was added and vortex, volume was adjusted to a final volume of 50µl. Reaction was incubated at 37°C for 15 minutes and quenched with Fluoro-DNA loading dye 6X (Zmtech Scientific, Canada), before analysis directly on 2% agarose gels. Gel images were analysed in ImageJ and percentage HDR was calculated according to equation 2 (Supplemental Methods).

TLR HDR assay

TLR assay was performed using cells transfected by AAV virus with pCVL-TL-dsRed Reporter 2.1 (VF2468 ZFN target) EF1a inserted at a safe harbour loci (Robert *et al*, 2015). The TLR transfected cell line was a kind gift from Francis Robert in the Pelletier lab. Cells were plated into 12 well plates and split into 6 well plates 24 hours after transfection. Cell were then grown for 5 days with samples. Control Cas9 plasmid transfections were performed by PEI transfection (1ug plasmid DNA, 30µl PEI and 300µl DMEM for a 2 well transfection). PEI:Cas9MAV was transfected into cells with 5' b donor, 3' b donor and 5'3' b donor. Cells were prepared for flow cytometry at 48 hrs and 96 hrs. Further details in supplemental methods.
The Cas9 plasmid was a gift from Jerry Pelletier (McGill University).

Fluorescent Microscopy of TLR assay Cells

For TLR assay cell plates (12 well and 6 well) GFP and RFP fluorescent images were collected using a Zeiss AX10 inverted fluorescence microscope and standard GFP and RFP filters. Image acquisition required a AxioCam MRm camera and Axiovision software. TIFF images were processed in ImageJ.

Cloning and Sanger sequencing

PCR products for sanger sequencing were cloned into bacterial blunt end cloning vectors (pCR, pUCM) as part of blunt end cloning kits (Thermo Fisher ZeroBlunt kit and Biobasic pUCM-T cloning vector kit). Plasmids were transformed into XL1 Blue competent cells. Cells were grown on selective plates and colonies picked for overnight broth culture and subsequent mini-prepping (kit). Where pUCM-T vector was used a 1 -TAQ (NEB, USA) polymerase was used to give the adenosine overhang. Sanger sequencing was performed by the McGill University and Génome Québec Innovation Centre (Montreal, Quebec). Sequence analysis for Sanger sequencing was performed in CLC Main Workbench (QIAGEN Bioinformatics, Redwood City, CA, USA) for sequence alignment and identification of donor insert sequences.

Ion Torrent

Ion Torrent sequencing of DNA samples was performed on a ThermoFisher Scientific Ion Torrent Personal Genome Machine ™ (PGM) apparatus using Ion 316 chip kit v2 (ThermoFisher Scientific, 4483324), with potentially 6 million reads. Samples were multiplexed upon the 316 chip by barcoding of 9 samples by PCR prior to the Ion Torrent work flow. Details of barcoding primers is included in supplemental data. After barcoding PCR, samples were run on a 2% gel and each band pertaining to the library amplicons (approximately 200bp) were excised and gel extraction performed (Qiagen gel extraction kit). DNA was quantified by nanodrop and diluted to 8 pmol prior to Ion Torrent work flow.

Ion Torrent CRISPR/Cas9 read processing

Sequenced products were filtered for a minimum length of 61 nucleotides (minimum expected distance to target subsequences of interest – described below) and mapped to a custom BLAST database (v2.5.0) consisting of 14 sequences (a combination of seven barcodes [plus primers] and two expected sequences [no insertion or the insertion of a stop codon and HindIII restriction enzyme site]). BLAST mapped libraries were then analyzed for two target subsequences: 1) the presence of an insertion (TAGAAGCTT) or 2) absence of any change in the canonical sequence near the CRISPR/Cas9 insertion site (CCAGGAT). Aligned reads that did not completely overlap with the target subsequences were removed ('truncated' reads). Any alignments that resembled a partial insertion (e.g., TAGA, TAGAA, TAG-A, etc.) or non-insertion (based on BLAST alignments) were then analyzed for their frequency of insertions or deletions (indels) at a given position of the target site, including five nucleotides up-/down-stream of the target subsequence (see Supplemental Data). A total of 793,117 sequenced CRISPR/Cas9 products were produced, where 2.522%, 0.104%, and 97.374% of products were too short to be aligned to expected product sequences, unmappable by the BLAST algorithm, or able to be aligned by BLAST to the expected product sequences, respectively. Of the 772,286 mappable products: 728,664 (91.873% of total sequenced products) contained a perfect match to the target subsequences; 16,083 (2.028%) produced alignments that did not overlap with expected target subsequences; and the remaining 27,539 (3.472%) contained partial insertions/non-insertions.

Acknowledgements:
The authors would like to acknowledge the financial support of the Faculty of Medicine, McGill University (Akavia), Canadian Institutes of Health Research (CIHR grant MOP-133535 to Nagar; CIHR grant MOP-142451 to Dostie), and the Natural Sciences and Engineering Research Council of Canada (NSERC Discovery grant to Blanchette). We would like to thank Francis Robert and Jerry Pelletier (helpful discussions, loan of equipment, and reagents), Bozena Samborska from Russell Jones' lab (helpful discussions, gift of reagents and cell lines), Sawn McGuirk from Julie St. Pierre's lab, Jutta Steinberger from Jerry Pellietier's lab, Sidong Huang (material gifts of cell lines), Alba Guarne and Kalle Gehring (providing access to DLS and SEC-MALS, respectively).

Competing Interests:
The authors have no competing financial interests.

Figure Legends:

Figure 1 **A**. SEC-MALS experiments were performed to determine the molecular weight and size of the modified Cas9MAV. The left y-axis shows the differential refractive index of the eluting protein peak (black), indicating perfect monodispersity. The calculated MW is 175.5 kDa and the measured MW across the peak is denoted on the right y-axis (red). An illustration of the Cas9MAV protein construct is included above the SEC-MALS figure. **B**. Migration of biotinylated DNA donors was tested on an agarose gel in the absence (left, magenta), or presence (right, green) of the Cas9MAV RNP. **C**. Mean volume distribution of three consecutive samples of the complex formation assessed by DLS. The scattering of 2uM CF3 donor DNA (light gray) was firstly measured, followed by 2uM preformed Cas9MAV/sgRNA complex (black). Cas9MAV RNP was subsequently added to the CF3 sample in a 1:1 molar ratio and measured again (dark grey). **D.** Mean number distribution of the experiment shown in C. **E**. Mean volume distribution of the full polyplex. PEI was added directly to the preformed Cas9MAV/sgRNA/CF3 sample analysed in C and D and incubated for polyplex formation before being transferred back to the cuvette. DLS was measured as before, the peaks observed in C and D were undetectable, but two new populations of larger size are observed. The x-axis is adjusted accordingly. **F**. Mean number distribution of the experiment shown in E. DLS experiment was performed once with a minimum of 15 measurements per sample. **G**. Transfection efficiency of PEI. FAM-labelled PEI was transfected into HEK293T cells and fluorescence of control (left) and transfected (middle) cells was assessed using flow cytometry (right, biological replicates n=3, each repeat involves 5000 individual cell measurements). **H**. Evaluation of indel formation frequency using PEI:Cas9MAV in human and murine cells. Three genes EMX1, FLCN and ACC1 were targeted in human cells, while ACC1

and FLCN were targeted in murine cells. Each experiment for indel formation was evaluated as 3 biological repeats for each gene in each cell line (human or murine).

Figure 2 Summary of CXCR4 HDR Experimentation in HEK293T cells

A. Overview of the donor lengths and designs used, biotinylated and non biotinylated. "b" enclosed in a circle denotes position of Biotin post PCR synthesis. Orientation of sense strand arrow indicated by 5' and 3' label. Green line denotes the central position and provides guidance as to the location of start of Stop codon (TAG) and HindIII sequence that follows. Lengths quoted in base pairs represents the total homology arm lengths used in each donor. The length of arrows either side of the dashed line indicates either equal homology arm length or asymmetric donor construction. Figure 2A also includes some donor designs not used in experimentation (CR1b, CR3b), but that could be formed by PCR donor creation detailed in supplemental.

B and **C.** Indel frequency (E) and its corresponding suppression in HDR experiments (F) with non-biotinylated donors. Significance was determined between the performance of Fas1/2 donors and longer CF1 and CF3 donors. Legend for panels E, F and G: MAV = Cas9MAV, NOC = Cas9 NLS with nocodazole treatment, and NLS = Cas9 with nuclear localization signal.

D. Demonstration of HDR frequencies with biotinylated donors..

E. Summative comparison of Cas9NLS and Cas9MAV with biotinylated and non biotinylated donors. NLS is out performed by Cas9MAV. Statistical analysis was performed using one way Anova and Tukey test with p-values: * = 0.05, **=0.01, ***=0.001.

F. 800 bp amplicon expressed from genomic DNA and the commensurate HDR derived HindIII cleavage.

G and **H.** RT PCR products (800 bp and 500 bp) and their HindIII products. At both genomic and mRNA levels the edit was shown to persist. Each experiment was repeated with biological replicates of 12 initial samples (1 failed extraction of DNA for genomic evaluation), 11 successful RNA extractions gave 8 successful RT-PCRs, Fig. 2G shows the first 3 lanes were amplification failed to be specific. Fig. 2H shows the second RT-PCR gel were 8 amplifications were successful, and resulted in 8 digests.

I. Summary of Sanger sequences from different Cas9MAV experiments with the 4 biotinylated donors illustrated in A. Base sequence modification caused by insertion of donor DNA is marked with red triangles. Full list of experimental conditions for samples sequenced and donors used to induce insert included in supplemental data.

Figure 3 HDR Valuations using Traffic Light reporter, EMX1 loci.

A. Demonstration of HDR with GFP fluorescence, while **D** shows no RFP emission from cells (no NHEJ). We observed that GFP and RFP emissions could be concurrent, hence NHEJ and HDR both occur within single cells.

Representative cells are present in panels **B** and **E. C** and **F** show NHEJ (RFP emission) and no HDR in the usage of Cas9 plasmid.

G shows the control population used to gate for flow cytometry, and **H** shows HDR caused by using Cas9MAV with donor 5'b, as an indicative example of HDR observed as occurring at the 96hr time point, occurring in the TLR assay with PEI:Cas9MAV. Further analysis of TLR assay at 48hrs and 96hrs is provided in Supplemental Figure S1 detailing full experiment.
I considers indel frequency at EMX1 and FLCN loci after transfection with a mismatched biotin donor for CXCR4. Indel frequency remained constant.
J is an initial investigation as to whether a donor not matched to genomic sequence can insert at a locus cleaved by Cas9MAV using our method. EMX1 loci PCR amplicon which does not form digest, showing mismatched donor was not inserted.
K shows the two donor designs for EMX1, central to the design is the inserted BamH1 sequence, between equal homology arms of EMX1 locus with matched non-biotinylated and biotinylated donors and PEI:Cas9MAV. Donor was 400bp (+ BamH1 & quadruple stop codon repeat) and contained a BamH1 site for validation of insertion. T7 and HDR assay samples are summations of a minimum of 3 biological replicates, p-value determined by one sided student's T-test and error bar represents standard deviation.

Figure 4 Deeper consideration of the PEI:Cas9MAV system towards a proposed model. Implementation of Lipofectamine 3000 with non biotin and biotin donors performed in HEK293T cell line. Experiment is a direct comparison to Figure 2, where Cas9MAV:PEI system produced high HDR, to evaluate the impact of PEI absence and non-biotin/biotin donor performance via an alternative transfection method.
A details the indel average percentage evaluated by T7 endonuclease assay for non-biotinylated/biotinylated donors, minimum of 3 biological replicates, p-value determined by one sided student's T-test and error bar represent standard deviation.
B represents the results of HindIII digest HDR assay, to evaluate homology directed repair by non-biotin/biotin donors.
C is a high level evaluation of library from Ion Torrent Sequencing. Each sequence was subdivided firstly on the basis of unedited and edited sequences and grouped by barcode to give total reads per library. Library name and commensurate donor variant used with PEI:Cas9MAV are detailed in Supplemental Data. NGS provides a more quantitative method of evaluating HDR than either TLR or digest assays, as direct sequence analysis can provide information relating to the knockin variation or partial insertions.
D is a table describing the complete and partial insertion sequence numbers (with percentages) from Ion torrent sequencing analysis. Total reads implies total reads per library. Complete refers to exact alignment to either no insert control sequence or inserted HDR sequence.
E Schematics of the hypothetical PEI:Cas9MAV system based on the experiments conducted. Two variants of canonical Cas9 are considered. Cas9NLS is a commonly used variant enabling nuclear localization, Cas9-NLS-monoavidin (Cas9MAV) is a new variant developed for this work containing a C-terminal peptide linker and monoavidin domain. The cationic polymer used to form polyplexes is represented in its general monomeric form and in cartoon lines with (+) representing

charge. Using both variants, Cas9NLS can exist in association in a PEI polyplex with DNA or without as there is no physical link. Cas9MAV applies biotin linkage to co-localize donor DNA with the MAV domain. It is hypothesized that PEI:RNP could enter the cell via endocytosis and undergo endosomal escape. Nuclear transport could happen via two mechanisms: PEI mediated nuclear transport, or entrance by the C-terminal nuclear localization sequence recruiting importins for internalization. Once in the nucleus, Cas9 and donor are available to perform HDR.

Figures

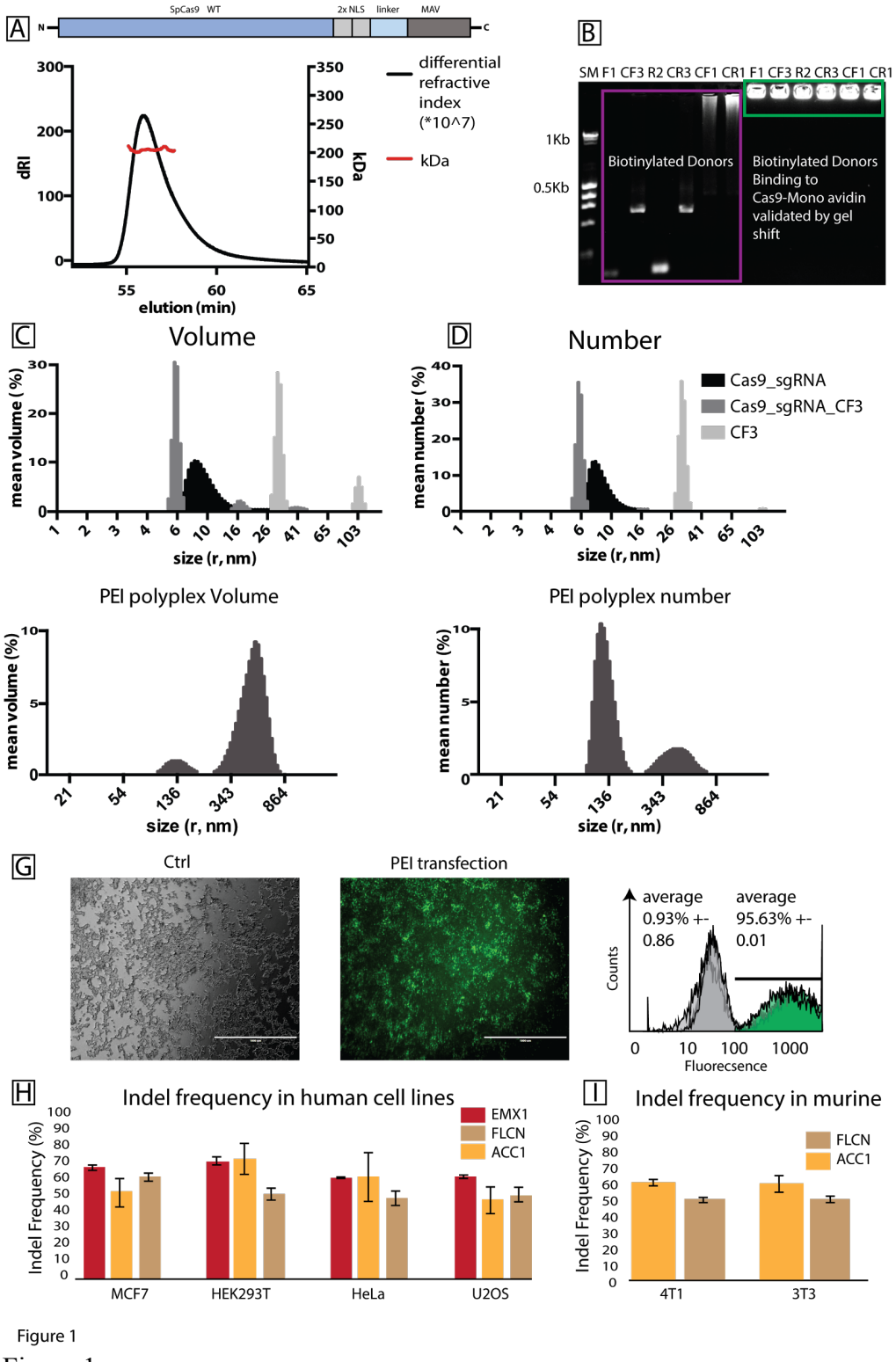


Figure 1

Figure 1

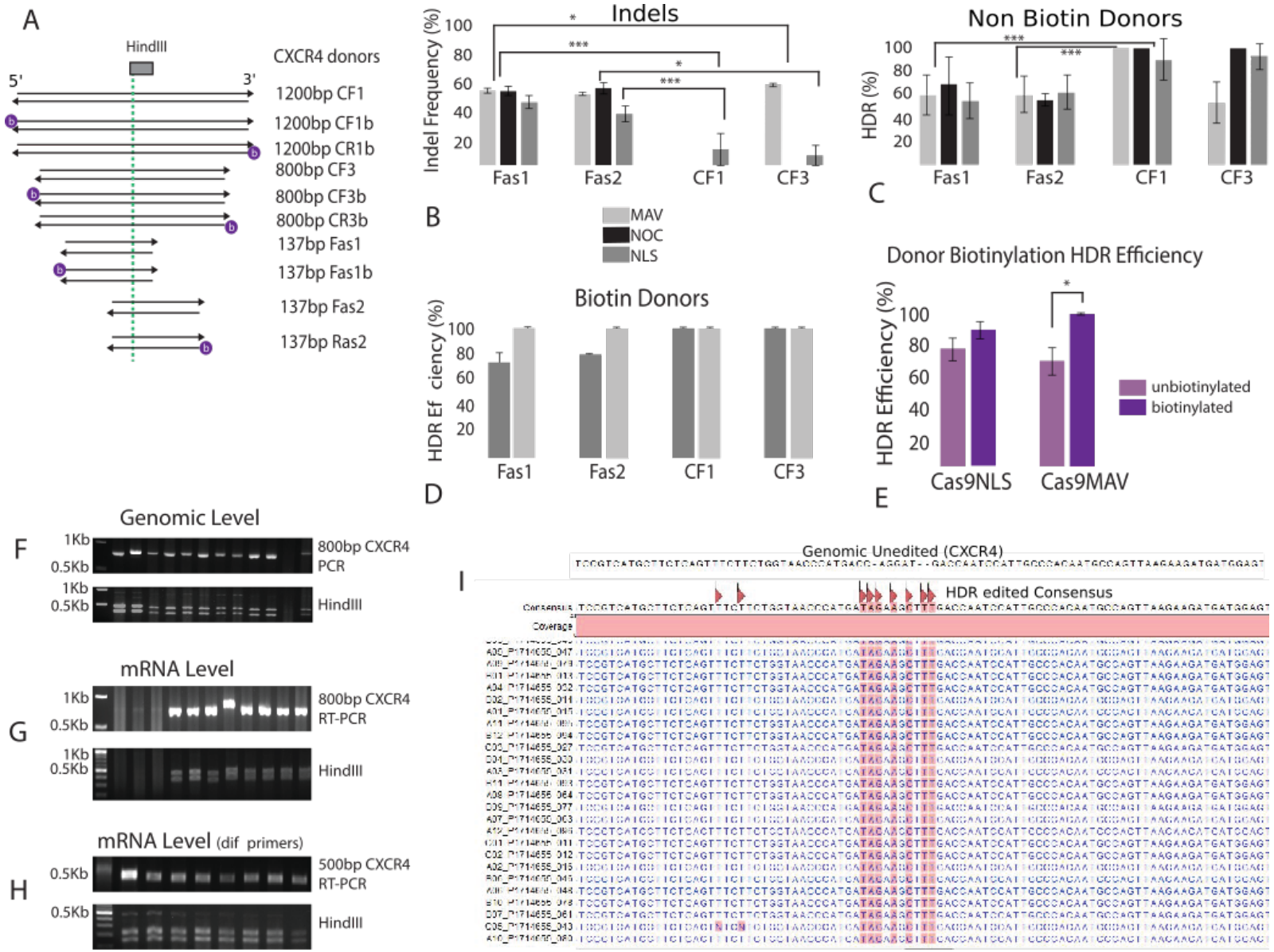

A
HindIII
5'
3'
CXCR4 donors
1200bp CF1
1200bp CF1b
1200bp CR1b
800bp CF3
800bp CF3b
800bp CR3b
137bp Fas1
137bp Fas1b
137bp Fas2
137bp Ras2
Indels
Indel Frequency (%)
Fas1
Fas2
CF1
CF3
B
MAV
NOC
NLS
Non Biotin Donors
HDR (%)
C
Biotin Donors
HDR Efficiency (%)
D
Donor Biotinylation HDR Efficiency
HDR Efficiency (%)
unbiotinylated
biotinylated
Cas9NLS
Cas9MAV
E
Genomic Level
F
1Kb
0.5Kb
800bp CXCR4 PCR
HindIII
mRNA Level
G
800bp CXCR4 RT-PCR
mRNA Level (dif primers)
H
500bp CXCR4 RT-PCR
Genomic Unedited (CXCR4)
HDR edited Consensus
I

Figure 2

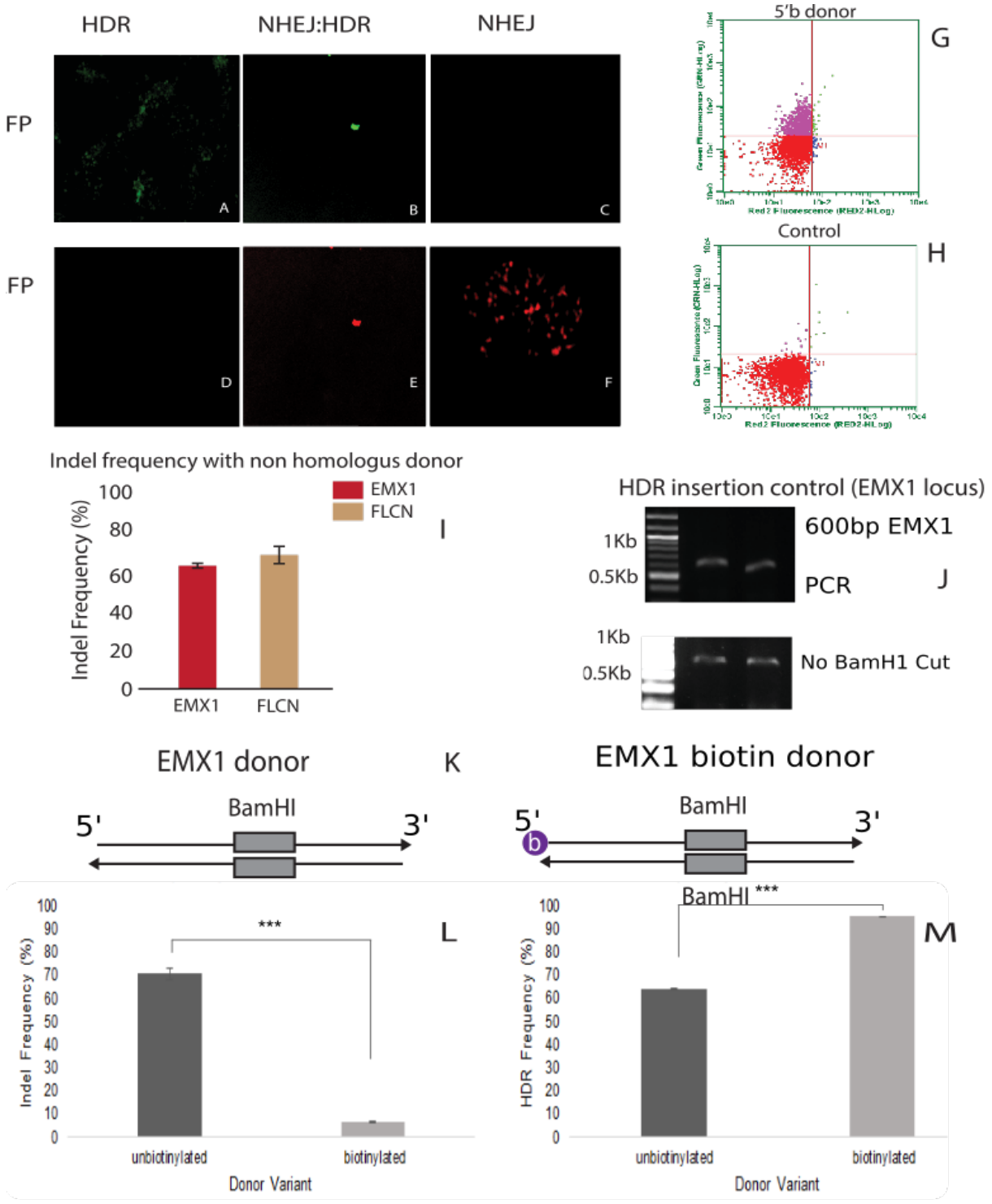
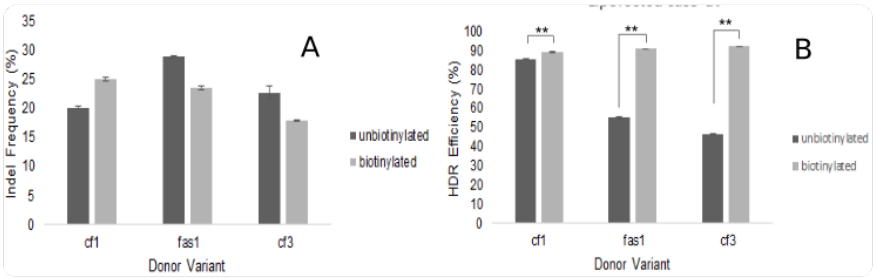
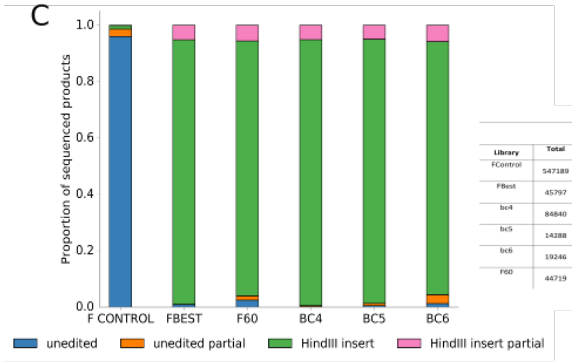

HDR
NHEJ:HDR
NHEJ
FP
FP
A
B
C
D
E
F
5'b donor
G
Control
H
Indel frequency with non homologus donor
Indel Frequency (%)
100
80
60
40
20
0
EMX1
FLCN
I
HDR insertion control (EMX1 locus)
1Kb
0.5Kb
600bp EMX1
PCR
J
1Kb
0.5Kb
No BamH1 Cut
EMX1 donor
K
BamHI
5'
3'
EMX1 biotin donor
BamHI
5'
3'
b
Indel Frequency (%)
***
L
unbiotinylated
biotinylated
Donor Variant
BamHI ***
HDR Frequency (%)
M
unbiotinylated
biotinylated
Donor Variant

Figure 3





D

		No insertion		Stop codon + HindIII insertion	
Library	Total	Complete	Partial	Complete	Partial
FControl	547189	525531 (96.00%)	14259 (02.60%)	6856 (01.30%)	543 (00.10%)
FBest	45797	412 (00.90%)	45 (00.10%)	42989 (93.87%)	2351 (05.13%)
bc4	84840	131 (00.15%)	321 (00.38%)	80057 (94.36%)	4331 (05.11%)
bc5	14288	74 (00.52%)	117 (00.82%)	13391 (93.72%)	706 (04.94%)
bc6	19246	244 (01.27%)	603 (03.13%)	17283 (89.80%)	1116 (05.80%)
F60	44719	1127 (02.52%)	654 (01.46%)	40457 (90.47%)	2481 (05.55%)

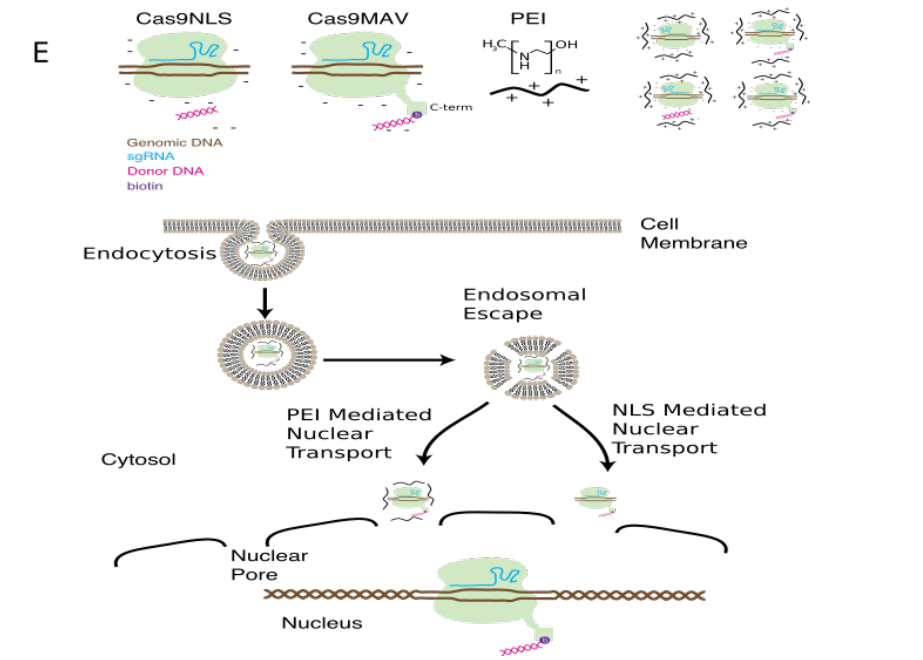


Figure 4

References:

Adachi O, Nakano A, Sato O, Kawamoto S, Tahara H, Toyoda N, Yamato E, Matsumori A, Tabayashi K & Miyazaki J (2002) Gene transfer of Fc-fusion cytokine by in vivo electroporation: application to gene therapy for viral myocarditis. *Gene Ther.* **9:** 577–583

Akita H, Ito R, Khalil IA, Futaki S & Harashima H (2004) Quantitative three-dimensional analysis of the intracellular trafficking of plasmid DNA transfected by a nonviral gene delivery system using confocal laser scanning microscopy. *Molecular Therapy* **9:** 443–451

Bétermier M, Bertrand P & Lopez BS (2014) Is Non-Homologous End-Joining Really an Inherently Error-Prone Process? *PLoS Genet* **10:** e1004086

Boussif O, Lezoualc'h F, Zanta MA, Mergny MD, Scherman D, Demeneix B & Behr JP (1995) A versatile vector for gene and oligonucleotide transfer into cells in culture and in vivo: polyethylenimine. *Proc. Natl. Acad. Sci. USA* **92:** 7297–7301

Cardarelli F, Digiacomo L, Marchini C, Amici A, Salomone F, Fiume G, Rossetta A, Gratton E, Pozzi D & Caracciolo G (2016) The intracellular trafficking mechanism of Lipofectamine-based transfection reagents and its implication for gene delivery. *Sci Rep* **6:** 25879

Carlson-Stevermer J, Abdeen AA, Kohlenberg L, Goedland M, Molugu K, Lou M & Saha K (2017) Assembly of CRISPR ribonucleoproteins with biotinylated oligonucleotides via an RNA aptamer for precise gene editing. *Nat Comms* **8:** 1711

Certo MT, Ryu BY, Annis JE, Garibov M, Jarjour J, Rawlings DJ & Scharenberg AM (2011) Tracking genome engineering outcome at individual DNA breakpoints. *Nat Meth* **8:** 671–676

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

Davis L & Maizels N (2016) Two Distinct Pathways Support Gene Correction by Single-Stranded Donors at DNA Nicks. *Cell Rep* **17:** 1872–1881

Drake DM & Pack DW (2008) Biochemical investigation of active intracellular transport of polymeric gene-delivery vectors. *J Pharm Sci* **97:** 1399–1413

Godbey WT, Wu KK & Mikos AG (1999) Tracking the intracellular path of poly(ethylenimine)/DNA complexes for gene delivery. *Proc. Natl. Acad. Sci. USA* **96:** 5177–5181

Gutschner T, Haemmerle M, Genovese G, Draetta GF & Chin L (2016) Post-translational Regulation of Cas9 during G1 Enhances Homology-Directed Repair. *Cell Rep* **14:** 1555–1566

Holt SE, Aisner DL, Shay JW & Wright WE (1997) Lack of cell cycle regulation of telomerase activity in human cells. *Proc. Natl. Acad. Sci. USA* **94:** 10687–10692

Jiang C, Davalos RV & Bischof JC (2015) A review of basic to clinical studies of irreversible electroporation therapy. *IEEE Trans Biomed Eng* **62:** 4–20

Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA & Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337:** 816–821

Jinek M, Jiang F, Taylor DW, Sternberg SH, Kaya E, Ma E, Anders C, Hauer M, Zhou K, Lin S, Kaplan M, Iavarone AT, Charpentier E, Nogales E & Doudna JA (2014) Structures of Cas9 endonucleases reveal RNA-mediated conformational activation. *Science* **343:** 1247997–1247997

Khalil IA, Kogure K, Akita H & Harashima H (2006) Uptake pathways and subsequent intracellular trafficking in nonviral gene delivery. *Pharmacol. Rev.* **58:** 32–45

Kim S, Kim D, Cho SW, Kim J & Kim J-S (2014) Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins. *Genome Res.* **24:** 1012–1019

Kouranova E, Forbes K, Zhao G, Warren J, Bartels A, Wu Y & Cui X (2016) CRISPRs for Optimal Targeting: Delivery of CRISPR Components as DNA, RNA, and Protein into Cultured Cells and Single-Cell Embryos. *Human Gene Therapy* **27:** 464–475

Lee K, Conboy M, Park HM, Jiang F, Kim HJ, DeWitt MA, Mackley VA, Chang K, Rao A, Skinner C, Shobha T, Mehdipour M, Liu H, Huang W-C, Lan F, Bray NL, Li S, Corn JE, Kataoka K, Doudna JA, et al (2017a) Nanoparticle delivery of Cas9 ribonucleoprotein and donor DNA in vivo induces homology-directed DNA repair. *Nature Biomedical Engineering* **337:** 1

Lee K, Mackley VA, Rao A, Chong AT, DeWitt MA, Corn JE & Murthy N (2017b) Synthetically modified guide RNA and donor DNA are a versatile platform for CRISPR-Cas9 engineering. *eLife Sciences* **6:** 1479

Li G, Zhang X, Zhong C, Mo J, Quan R, Yang J, Liu D, Li Z, Yang H & Wu Z (2017) Small molecules enhance CRISPR/Cas9-mediated homology-directed genome editing in primary cells. *Sci Rep* **7:** 8943

Lim KH, Huang H, Pralle A & Park S (2013) Stable, high-affinity streptavidin monomer for protein labeling and monovalent biotin detection. *Biotechnol. Bioeng.* **110:** 57–67

Lin MI, Paik E, Mishra B, Burkhardt D & Kernytsky A (2017) CRISPR/Cas9 Genome Editing to Treat Sickle Cell Disease and B-Thalassemia: Re-Creating Genetic Variants to Upregulate Fetal Hemoglobin Appear Well-Tolerated, Effective and Durable. *Blood* **130:** 284

Lin S, Staahl BT, Alla RK & Doudna JA (2014) Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery. *eLife Sciences* **3:** e04766 Available at: https://elifesciences.org/articles/04766

Ma M, Zhuang F, Hu X, Wang B, Wen X-Z, Ji J-F & Xi JJ (2017) Efficient generation of mice carrying homozygous double-floxp alleles using the Cas9-Avidin/Biotin-donor DNA system. *Cell Res* **27:** 578–581

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

Miyaoka Y, Berman JR, Cooper SB, Mayerl SJ, Chan AH, Zhang B, Karlin-Neumann GA & Conklin BR (2016) Systematic quantification of HDR and NHEJ reveals effects of locus, nuclease, and cell type on genome-editing. *Sci Rep* **6:** 23549

Nishimasu H, Ran FA, Hsu PD, Konermann S, Shehata SI, Dohmae N, Ishitani R, Zhang F & Nureki O (2014) Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell* **156:** 935–949

Ogris M, Steinlein P, Kursa M, Mechtler K, Kircheis R & Wagner E (1998) The size of DNA/transferrin-PEI complexes is an important factor for gene expression in cultured cells. *Gene Ther.* **5:** 1425–1433

Oh Y-K, Suh D, Kim JM, Choi H-G, Shin K & Ko JJ (2002) Polyethylenimine-mediated cellular uptake, nucleus trafficking and expression of cytokine plasmid DNA. *Gene Ther.* **9:** 1627–1632

Pollard H, Remy JS, Loussouarn G, Demolombe S, Behr JP & Escande D (1998) Polyethylenimine but not cationic lipids promotes transgene delivery to the nucleus in

mammalian cells. *J. Biol. Chem.* **273:** 7507–7511
Richardson CD, Ray GJ, DeWitt MA, Curie GL & Corn JE (2016) Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. *Nat Biotechnol* **34:** 339–344
Robert F, Barbeau M, Éthier S, Dostie J & Pelletier J (2015) Pharmacological inhibition of DNA-PK stimulates Cas9-mediated genome editing. *Genome Med* **7:** 93
Schwank G, Koo B-K, Sasselli V, Dekkers JF, Heo I, Demircan T, Sasaki N, Boymans S, Cuppen E, van der Ent CK, Nieuwenhuis EES, Beekman JM & Clevers H (2013) Functional repair of CFTR by CRISPR/Cas9 in intestinal stem cell organoids of cystic fibrosis patients. *Cell Stem Cell* **13:** 653–658
Shao S, Ren C, Liu Z, Bai Y, Chen Z, Wei Z, Wang X, Zhang Z & Xu K (2017) Enhancing CRISPR/Cas9-mediated homology-directed repair in mammalian cells by expressing Saccharomyces cerevisiae Rad52. *International Journal of Biochemistry and Cell Biology* **92:** 43–52
Shibata M, Nishimasu H, Kodera N, Hirano S, Ando T, Uchihashi T & Nureki O (2017) Real-space and real-time dynamics of CRISPR-Cas9 visualized by high-speed atomic force microscopy. *Nat Comms* **8:** 1430
Suh J, Wirtz D & Hanes J (2003) Efficient active transport of gene nanocarriers to the cell nucleus. *Proc. Natl. Acad. Sci. USA* **100:** 3878–3882
Van Trung Chu, Weber T, Wefers B, Wurst W, Sander S, Rajewsky K & Kühn R (2015) Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells. *Nat Biotechnol* **33:** 543–548
Wen Y, Pan S, Luo X, Zhang X, Zhang W & Feng M (2009) A biodegradable low molecular weight polyethylenimine derivative as low toxicity and efficient gene vector. *Bioconjug. Chem.* **20:** 322–332
Yanik M, Müller B, Song F, Gall J, Wagner F, Wende W, Lorenz B & Stieger K (2017) In vivo genome editing as a potential treatment strategy for inherited retinal dystrophies. *Prog Retin Eye Res* **56:** 1–18
Yu X, Liang X, Xie H, Kumar S, Ravinder N, Potter J, de Mollerat du Jeu X & Chesnut JD (2016) Improved delivery of Cas9 protein/gRNA complexes using lipofectamine CRISPRMAX. *Biotechnol. Lett.* **38:** 919–929
Zhang J-P, Li X-L, Li G-H, Chen W, Arakaki C, Botimer GD, Baylink D, Zhang L, Wen W, Fu Y-W, Xu J, Chun N, Yuan W, Cheng T & Zhang X-B (2017) Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage. *Genome biology* **18:** 35
Zhou R, Geiger RC & Dean DA (2004) Intracellular trafficking of nucleic acids. *Expert Opin Drug Deliv* **1:** 127–140

Supplemental Materials and Methods

Table S1 Review of Cas9 HDR Improvement strategies is found in supplemental data and document SD5.

Classification	Name	Sequence
gBLOCK template for CXCR4 dDNA formation by PCR	Cxcr4 gblocHindIII	GCTGATCCCAATGTAGTAAGGCAGCCAACAGGCGAAGAAAGCCAGGATG AGGATGACTGTGGTCTTGAGGGCCTTGCGCTTCTGGTGGCCCTTGGAGTGT GACAGCTTGGAGATGATAATGCAATAGCAGGACAGGATGACAATACCAG GCAGGATAAGGCCAACCATGATGTGCTGAAACTGGAACACAACCACCCAC AAGTCATTGGGGTAGAAGCGGTCACAGATATATCTGTCATCTGCCTCACTG ACGTTGGCAAAGATGAAGTCGGGAATAGTCAGCAGGAGGGCAGGGATCC AGACGCCAACATAGACCACCTTTTCAGCCAACAGCTTCCTTGGCCTCTGAC TGTTGGTGGCGTGGACGATGGCCAGGTAGCGGTCCAGACTGATGAAGGCC AGGATGAGGACACTGCTGTAGAGGTTGACTGTGTAGATGACATGGACTGC CTTGCATAGGAAGTTCCCAAAGTACCAGTTTGCCACGGCATCAACTGCCC AGAAGGGAAGCGTGATGACAAAGAGGAGGTCGGCCACTGACAGTGCAGC CTGTACTTGTCCGTCATGCTTCTCAGTTTCTTCTGGTAACCCATGATAGAAG CTTTGACCAATCCATTGCCCACAATGCCAGTTAAGAAGATGATGGAGTAG ATGGTGGGCAGGAAGATTTTATTGAAATTAGCATTTTCTTCACGGAAACAG GGTTCCTTCATGGAGTCATAGTCCCCTGAGCCCATTTCCTCGGTGTAGTTA TCTGAAGTGTATATCTGCAAAAGAGGCAAAGGAATGGACATTCACTTCCA ATTCAGCAAGCATTAACCCAGTTAAAAAAAATTTTTAAAGCAATTTAAAA AACCAATTCAGGCTTGCTTTCTTCAGGAAATTCTGAAGTAGTGGGCTAAGG GCACAAGAGAATTAATGTAGAATCCTACAACTCTCCTCCCCATCTTTTCCC ATAGTGACTTCATTATATCCTTCTTTGGTAGAACCAATTACAAAATTCTTTG TTTAGAACAAAAGGGCACTGAGACGCTGAGGGTTTCAAAGTCACATCTTG GCTAACTCCTCTGCCCCGCCCACTAGAGGGAAGAAAAAAAACCTTCCTTA GGAGGAAAAAAAAAATACACACAAAGAGGCCACTCCCAGGCGGCGTGGG GGGTGGGGTGGGTGCTGCTAGGAGGGGCAGTGATTAACTTTTTGTA
Gblock template for EMX1 dDNA formation by PCR	EMX1 BamH1	GCCATGACTCCAGGCCTCCCCAAAGCCTGGCCAGGGAGTGGCCAGAGTCC AGCTTGGGCCCACGCAGGGGCCTGGCCAGCAGCAAGCAGCACTCTGCCCT CGTGGGTTTGTGGTTGCCCACCCTAGTCATTGGAGGTGACATCGATGTCCT CCCCATTGGCCTGCTTCGTGGCAATGCGCCACCGGTTGATGTGATGGGAGG GGATCCTAGTAGTAATGATCGGACTCAGGCCCTTCCTCCTCCAGCTTCTGC CGTTTGTACTTTGTCCTCCGGTTCTGGAACCACACCTTCACCTGGGCCAGG GAGGGAGGGGCACAGATGAGAAACTCAGGAGGCCCCCAGAGCAGCCACT GGGGCCTCAACACTCAGGCTGAGCTGAGAGCCTGATGGGAAGACTGAGGC TACATAGGGTTAGGGGCCCCAGGCCGGGGTCCCCTCTGACCAGCTGCTCC CATGGGTCTAACATTCACAGAAGGGGATGGCAGGGCAGGAAGAGGACAC AG
Donor	TLR/TLRb	ACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGGCAAGCTGCCCGTG CCATGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGC CGCTACCCCG
Engen sgRNA synthesis	CXCR4:	TAATACGACTCACTATAGGG*GGGCAATGGATTGGTCATCC* GTTTTAGAGCTAGA
	EMX1:	*TAATACGACTCACTATAGGG*GAGTCCGAGCAGAAGAAGAA GTTTTAGAGCTAGA
	Acc1:1	*TAATACGACTCACTATAGGG*AAGCCCTTCGAACATACACCGTTTTAGAGCTA GA

	Acc1:3	*TAATACGACTCACTATAGGG*GCTTAAGGACAACACCTGTGGTTTTAGAGCTAGA
	FLCN	*TAATACGACTCACTATAGGG*ATCCACAGACAGGTTCTGGTGTTTTAGAGCTAGA
	TLRd	TTCTAATACGACTCACTATAGAGCAGCGTATTCGAGAGTGGTTTTAGAGCTAGA
Generic S.pyo reverse complement backbone for sgRNA		*5'AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGACTAGCCTTATTTTAACTTGCTATTTCTAGCTGTAAAAC3'*
Cxcr4 primers genomic	Cxcr4 F	TGAGGACACTGCTGTAGAGGT
Cxcr4 primers genomic	Cxcr4 R	TGCTTGCTGAATTGGAAGTGA
	CF1	CCCAATGTAGTAAGGCAGCCA
	CR1	TGGGAGTGGCCTCTTTGTGTG
	FAS1	CCGTCATGCTTCTCAGTTTCTT (only fas1 needs a biotin version, name Fas1b)
	RAS1	AAGGAACCCTGTTTCCGTGA
	FAS2	Fas2: CCAGTTTGCCACGGCATC
	RAS2	GTCAAAGCTTCTATCATGGGTTAC (only Ras2 gets the biotin version, names Ras2b)
	CF3	ACGTTGGCAAAGATGAAGTCG (biotin version: CF3b)
	CR3	GACTTTGAAACCCTCAGCGTC (biotin version CR3b
EMX1 Primers	Donor construction-Forward	CCATGACTCCAGGCCTCCC (biotin version EMX1b)
EMX1 Primers	Donor construction-Reverse	TGCCATCCCCTTCTGTGAAT (biotin version EMX1b)
EMX1 primers genomic	EMX1-F	GACTCCAGGCCTCCCCAAAG
EMX1 primers genomic	EMX1-R	GCCCCTAACCCTATGTAGCC
Ion torrent Primers	F-control	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAAGGCCACACCAGCCTGTACTTGTCCGTCAT
	F60	CCATCTCATCCCTGCGTGTCTCCGACTCAGTGTGCCTGTCCAGCCTGTACTTGTCCGTCAT

	FBEST	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGATCGGTTCCAGCCTGTACTTGTCCGTCAT
	R-Ionseq	CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGATTCCTGCCCACCATCTACTCC
	ion-bc4	CCATCTCATCCCTGCGTGTCTCCGACTCAGTCAGGAATACCAGCCTGTACTTGTCCGTCAT
	ion-bc5	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGGAAGAACCTCCAGCCTGTACTTGTCCGTCAT
	ion-bc6	CCATCTCATCCCTGCGTGTCTCCGACTCAGCGAAGCGATTCCAGCCTGTACTTGTCCGTCAT
	ion-bc7	CCATCTCATCCCTGCGTGTCTCCGACTCAGCAGCCAATTCTCCAGCCTGTACTTGTCCGTCAT
Primers	TLR donor-F	ACGGCAAGCTGACCCTGAAGT
	TLR donor-R	GGGGTAGCGGCTGAAGCACTGCAC
Cas9 expression plasmid primers	SpCas9_1025_seq_for	TGAGCAAGAAATAGGCAAAGCA
	17linker_avidin_for	GAGACCCCAGGAACTAGCGAGAGCGCTACACCGGAATCGGCGGAAGCGGGTATCACC
	avidin_NotI_rev	AATATGCGGCCGCttaAGACGCCGCAGACGGTTTAA
	SpCas9_EcoRI_445_for	TTTTCGAATTCCTTATTATGTTGGTCC
	NLS_18linker_rev	TGTAGCGCTCTCGCTAGTTCCTGGGGTCTCTGAACCGCTATCAACTTTTCGTTTCTTTTTAGGA

Table S2 Oligonucleotide sequences used for experimentation.

Plasmid Maps

Plasmid Map for Cas9MAV found in SD2 attachment

Statistic Analysis and Raw data for Figure 2, 3 and 4

Graphs and p-values are detailed in supplementary document SD1

TLR Flow analysis:

The second loci we investigated is a variant of the Traffic light reporter assay. The cell line was a gift from Prof Jerry Pelletier and Francis Robert (Biochemistry, McGill University). Donor template was provided on a plasmid backbone and amplified using TLR donor F and R primers for biotinylation (see oligo table in supplemental). Cells were transfected using PEI:Cas9MAV and biotinylated donor DNA to induce a frame shift resulting in GFP expression, or if only indels are formed mCherry.

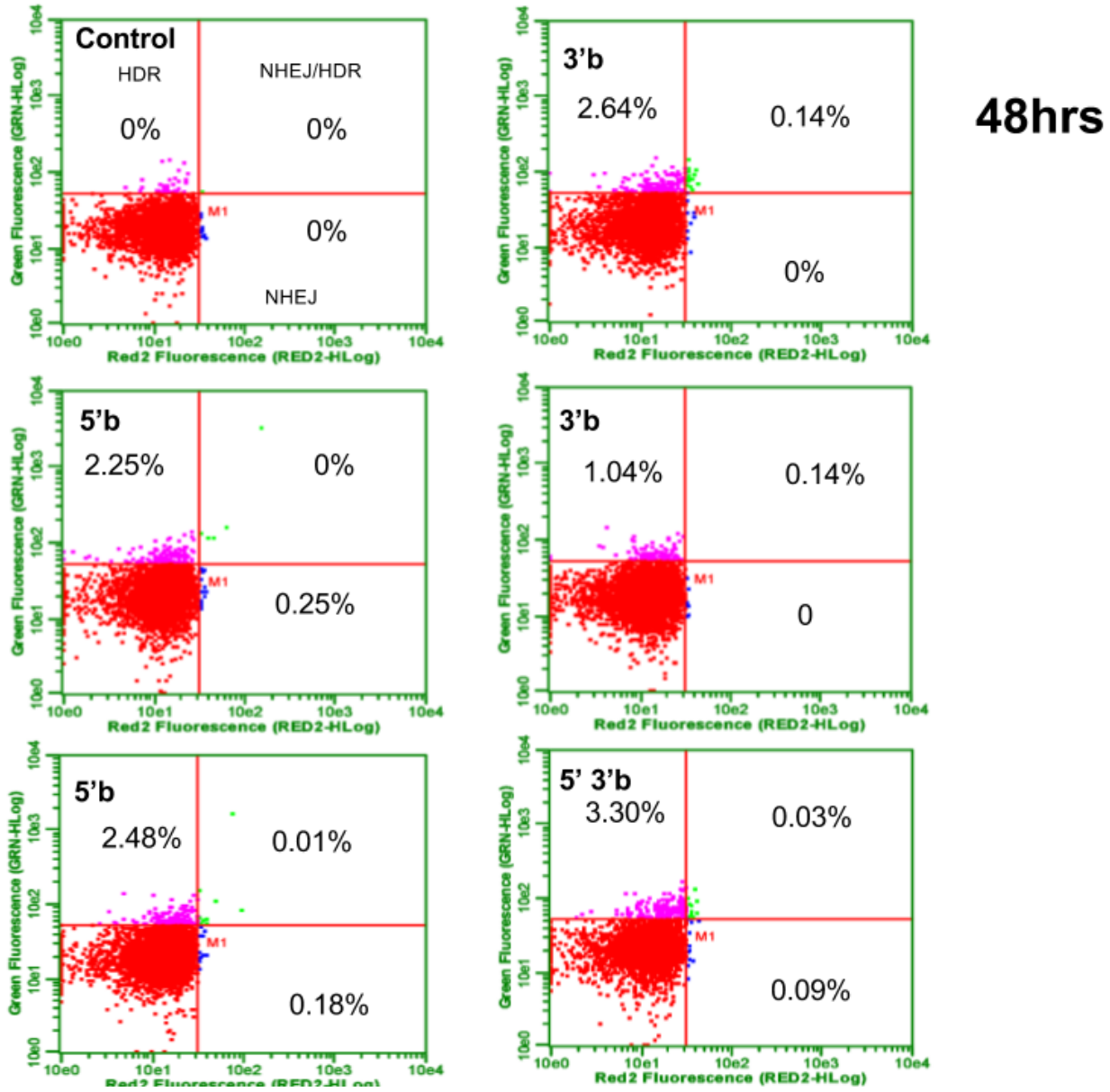
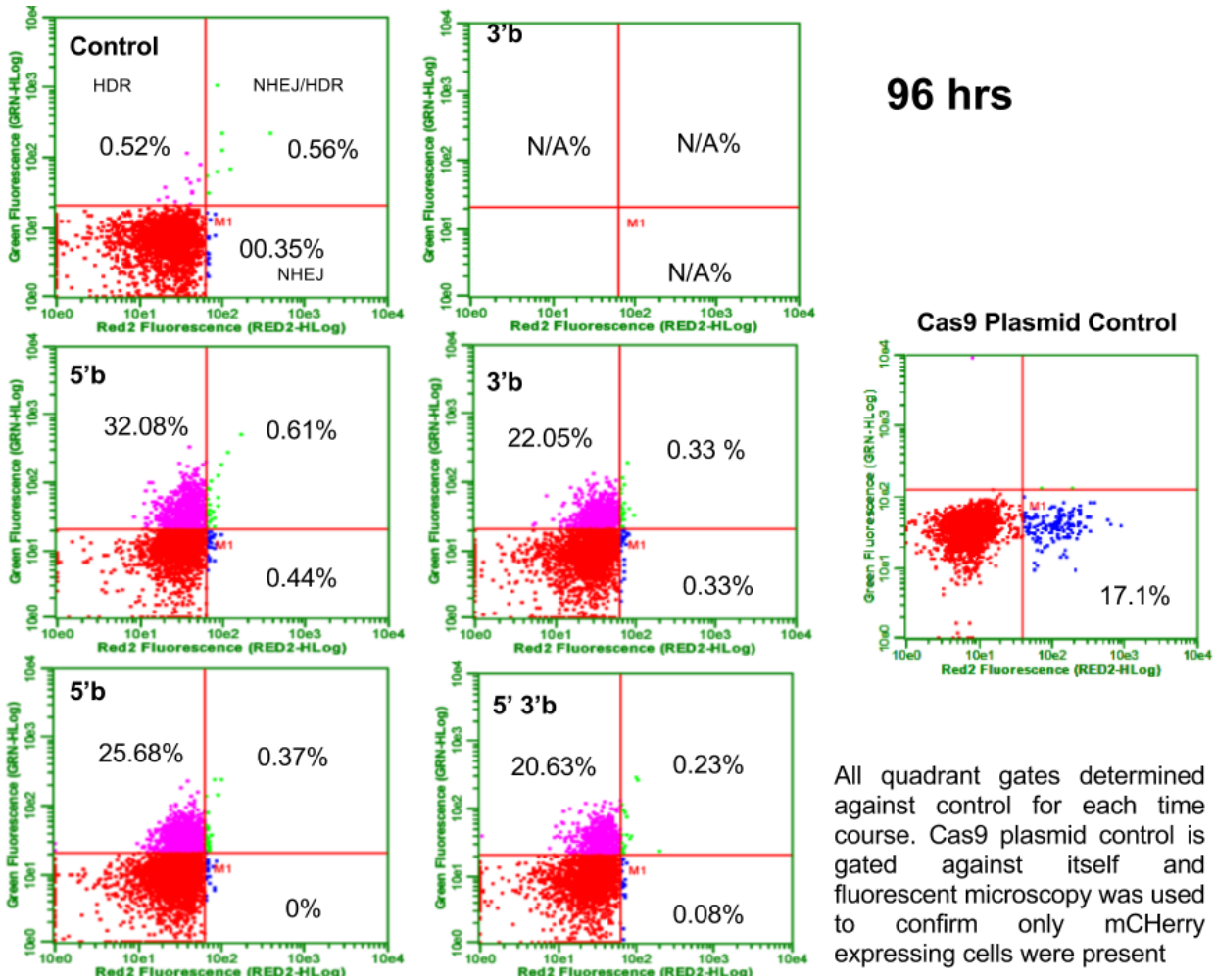


Figure S1A. 48hr flow data



All quadrant gates determined against control for each time course. Cas9 plasmid control is gated against itself and fluorescent microscopy was used to confirm only mCHerry expressing cells were present

Figure S1B 96hr flow data

		48hr			96hrs	
	LR/NHEJ	**UL/HDR**	**NHEJ/HDR**	**LR/NHEJ**	**UL/HDR**	**NHEJ/HDR**
Control	0.26	0.88	0.23	0.56	0.52	0.35
5'b	0.07	2.25	0	0.44	32.08	0.61
5'b	0.06	2.48	0	0	25.68	0.37
3'b	0	2.64	0.14	N/A	N/A	N/A
3'b	0	1.04	0.14	0.33	22.05	0.33
5'3'b	0.03	3.3	0.03	0.08	20.63	0.23
mean Average	0.032	2.342	0.062	0.2125	25.11	0.385
sd	0.03271085447	0.826026634	0.07224956747	0.2067808824	5.110114154	0.1611417595

Control values for each quadrant were used to normalise controls and subtracted from each quadrant of test samples.

In the case of control value exceeding test sample quadrant value, hence becoming negative value is set to zero, to avoid negative numbers

LR/NHEJ = Lower right quadrant Nonhomologous end joining

UL/HDR = Upper left quadrant homology directed repair

NHEJ/HDR = Combined non-homologous end joining and HDR (mono allelic for both)

Figure S1C Summary table for TLR experiments

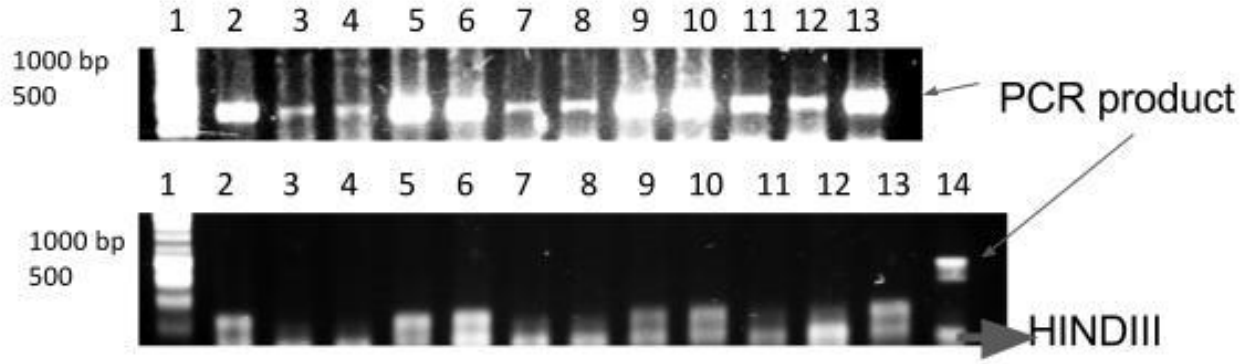
Figure S1 Traffic Light report assay for HDR and NHEJ assessed at 48hr (A) and 96hrs (B) post transfection with PEI:Cas9MAV. Data is presented using the quadrant gating approach used previously by Certo et al (Certo *et al*, 2011). Control Sample quadrant values used to normalise each time period and set the gating to exclude dead cells. Cells in lower left quadrant represent the unedited population. Each experiment was repeated twice for flow analysis apart from combined 5’3’ donor. 5’ in the context of this experiment means Forward primer biotin generated donor. 3’ means reverse primer with modified biotin. 5’3’b refers to a donor generated as a combination of both biotinylated primers.

Cells in upper left quadrant express GFP, hence HDR. Those cells in upper right are co-expressing GFP and RFP, are both HDR and NHEJ. Cells found in the lower right quadrant are expressing RFP representing NHEJ. For each variable 2000-5000 live cells were gated and analysed for RFP and GFP expression using the Guava flow cytometer. Sample 3'b at 96hrs had few to no surviving cells post harvesting by trypsin and PBS washes.

Cas9 plasmid control is included to give a measure of the standard performance of plasmid vectors expressing Cas9, but also the NHEJ frequency normally observed in the absence of donor. It should be noted Cas9 plasmid requires 4-7days to produce a reliable signal and only on the 4th day did the plasmid system record a indicative signal and hence it’s inclusion only in the 96hr data set. Over all performance of Cas9MAV was assessed by averaging the performance of all donors (n=4 repeats, 2 for each donor type) and for each time period (48hrs and 96hrs).

CXCR4 Regrowth experiments:
An important hypothesis considered was if cells are edited at high frequency (99%) then the edit should persist in cell population as the most numerous sequence. If the edit is lower frequency after a period of re-growth, edited cells are regularly out competed and this presents a significant challenge to single cell cloning for formation of stable cell lines. If the results of the CXCR4 experiment as confirmed in figure 3 at genomic and mRNA levels are consistent, the edit should persist after cell passaging and aggressive dilution of total cell number for the seeding of subsequent plates.

Cells were edited at CXCR4 using PEI-Cas9 system with biotinylated donors as per standard protocol and cell density. After 48hrs they were split 10:1 into a 6 well plate. Hypothesis is that if edit is 100% or near, the edit will still persist after aggressive dilution and 7 days regrowth, validating that the edit is stable and inheritable. Figure S2, confirms the edit persisted after a 10:1 passage into 6 cell plates and 7 days regrowth. The experiment confirms that all donors induce persistent, high level HDR at this loci.



CXCR4 HINDIII digests were run on an agarose gel at 115 V for 20 minutes. Lane 1 is the DNA ladder (100 bp). Lanes 2-13 are for the CXCR4 HINDIII tests. Lane 2 is CF1b, Lane 3 is Fas1b, Lane 4 is Ras1b, Lane 5 is CF3b. Lanes 6-13 are the repeats (triplicates were run). Lane 14 is the uncut PCR product CF1b (as a control).

Figure S2 - Growth of HDR edited cells to evaluate the stability of HDR edit after aggressive 10 to 1 dilution of cells.

Nucleofection Transfection of the Cas9MAV
Transfection of the Cas9MAV RNP was evaluated by the performance on amaxa nucleofector II, using the T7 indel assay as a means to confirm nuclease performance.

Briefly HEK-293T cells were transfected with the Amaxa Nucleofector® Kit V (Program Q-001) with a GFP positive control performed to confirm nucleofection instrument functioned. Cells were

harvested and counted. 1 million cells were resuspended in Kit V buffer with RNP prepared as per RNP transfection method.

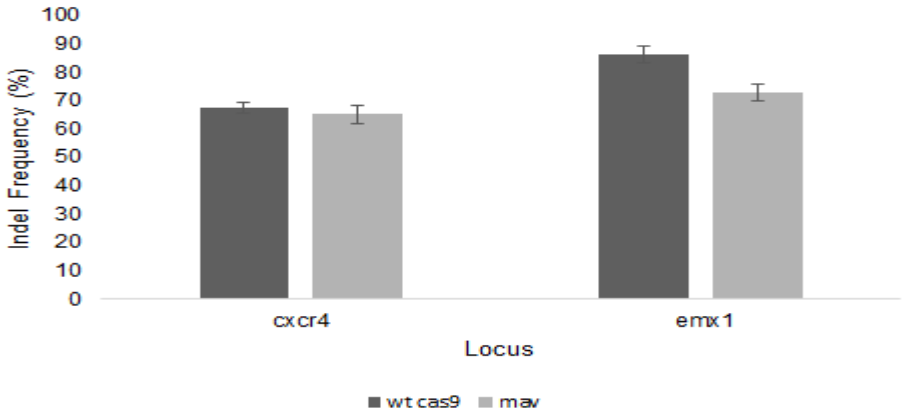


Figure S3 Indel frequency at CXCR4 and EMX1 loci nucleofection of Cas9MAV RNP. Comparison of Wild type Cas9 and Cas9MAV

The indel frequency was assessed as near equivalent to PEI transfection, for both Wild type Cas9 (no NLS) and Cas9MAV at both loci. In addition it confirms the Cas9MAV system can be introduced by alternative transfection systems.

Suppression of NHEJ using PEI:Cas9MAV RNP system

Suppression of NHEJ was observed where HDR frequency was found to be in high order by HINDIII restriction site insertion into genomic loci (CXCR4). Figure S4 demonstrates an example where in combination with the RNP system, Nocodazole was administered to investigate the effects of cell cycle arrest. Principally the effect was negligible, as HDR rates were statistically irrelevant from using the RNP system without cell cycle arrest.

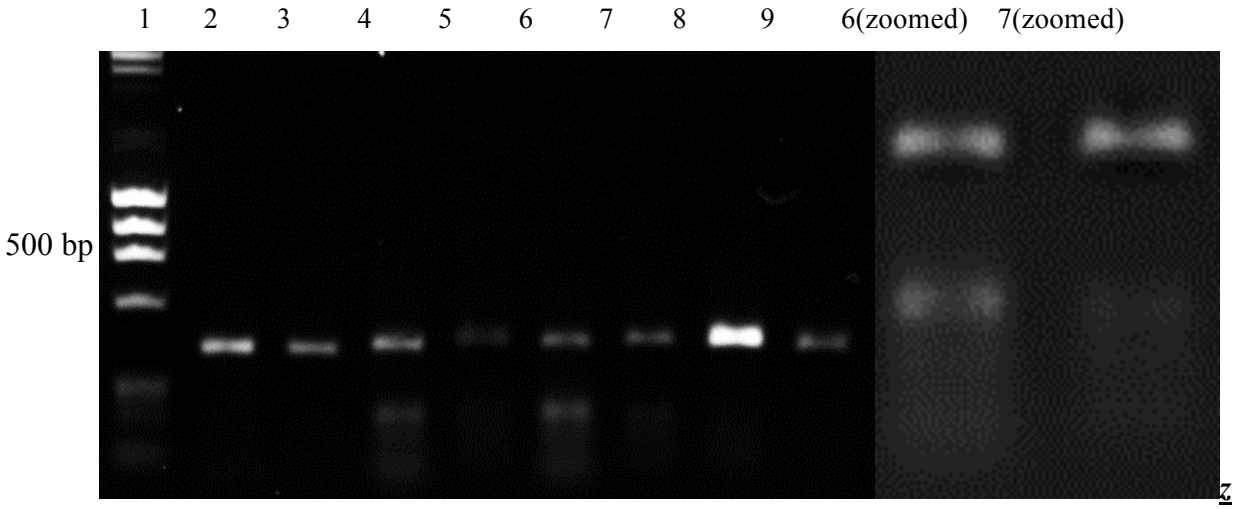


Figure S4: T7 Endonuclease Assay of Nocodazole Treated Samples from CXCR4 locus. A 2% TAE agarose gel was run at 90V for 20 minutes. In lane 1, 5 µL of DNA ladder was run. The following lanes contain T7 digest products of the nocodazole treated variants (with unbiotinylated donors) from the CXCR4 locus (5 µL in each lane). The lanes 2-9 show alternating T7 cut products and uncut samples. Lanes 2 and 3 show Cas9-noc using the cf1 donor. Lanes 4 and 5 show Cas9-noc using the fas1 donor. Lanes 6 and 7 show Cas9-noc using the fas2 donor. Lanes 8 and 9 show Cas9-noc using the cf3 donor. On the right is a zoomed version of lanes 6 and 7 for easier visualization of the bands.

The top bands in lanes 2-9 correspond to the uncut sample, whereas the bottom bands are the cleaved products, and slightly lower are the primer-dimers. As seen in the gel, different donors led to different amounts of NHEJ. For example in lane 2, it is seen that nocodazole treated samples with the cf1 donor produced no NHEJ (no bottom band is visible) whereas in the zoomed lanes 6 and 7 it is clear that T7 produced a cut with the fas2 donor. Technical replicates were performed and further analysis was done in ImageJ to determine the extent of NHEJ with all the variants. Using the intensities of the bands, the following equation was used to calculated indel frequency:

Equation 1:

$$\% \, indel = 100 \, x \, \frac{intensity \, of \, digested \, band}{intensity \, of \, digested \, band \, + \, intensity \, of \, undigested \, band}$$

Assessing HDR frequency by insertion of restriction digest palindromic sequences

As a rapid means of assessing HDR frequency EMX1 and CXCR4 donors were designed with a restriction digest site not present in the genomic sequence. The loci were assessed using NEB cutter restriction digest predictor before selection of enzymes was performed. IN the case of EMX1 BamH1 was chosen and CXCR4 HindIII was used. Once HDR was completed after 48hrs comparison of samples against an unedited control were performed. An example of the results and analysis method are shown below:

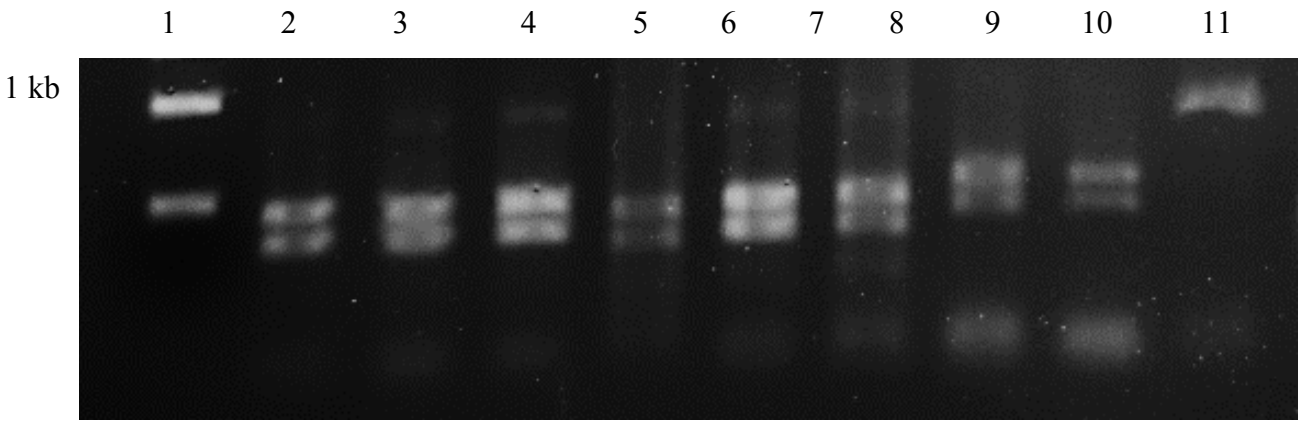


Figure S5: HindIII Digest Assay of samples from CXCR4 locus. A 2% TAE agarose gel was run at 90V for 20 minutes. In lane 1, 5 µL of DNA ladder was run. The following lanes contain HINDIII digest products of the Cas9-nls variants (with unbiotinylated donors) from the CXCR4 locus (5 µL in each lane). Lanes 2-5 show Cas9NLS using the cf1, fas1, fas2 and cf3 donors respectively. Lanes 6 and 7 are repeats of Cas9NLS fas1 and fas2, and lanes 8 and 9 are repeats of cf1 and cf3. Lane 10 is a positive control. Lane 11 is the unedited HEK genomic band.

The top bands in lanes 2-10 correspond to the uncut sample, whereas the middle two bands are the cleaved products, and slightly lower are the primer-dimers. Similar to the results from the T7 tests, the various donors led to different amounts of HDR. For example in lane 2, it is seen that Cas9-nls samples with the cf1 donor produced an apparent 100% HDR (no top band is visible) whereas in the lanes 3 and 4 (Cas9NLS with Fas1 and Fas2) HINDIII did not completely cleave

the upper band (so there is less than 100% HDR). Both biological and technical replicates were performed and ImageJ was used to determine the HDR efficiency with all the variants.

Equation 2:

$$\% HDR = 100\ x\ \frac{(Digested\ product\ first\ band\ \ intensity)\ +\ (Digested\ product\ 2nd\ band\ intensity)}{(Digested\ product\ first\ band\ intensity\ +\ Digested\ second\ band\ intensity)\ +\ (Undigested\ band\ intensity)}$$

Comparison of Cell Viability when transfecting Cas9MAV

Cell viability of transfection method was considered an important factor in the choice of PEI as the means of introducing Cas9MAV to the cells. Maintaining a high cell viability (greater than 80%) means a great number of cells can be potentially edited. Figure S6 details the percentage live cells assessed by trypan blue staining after 48hrs. Three methods are compared, PEI, Lipofectamine 3000 and Amaxa nucleofection using kit V for the transfection of Cas9MAV in HEK293T cells. None of the methods in our hands using the protocols detailed significantly affecting cell viability.

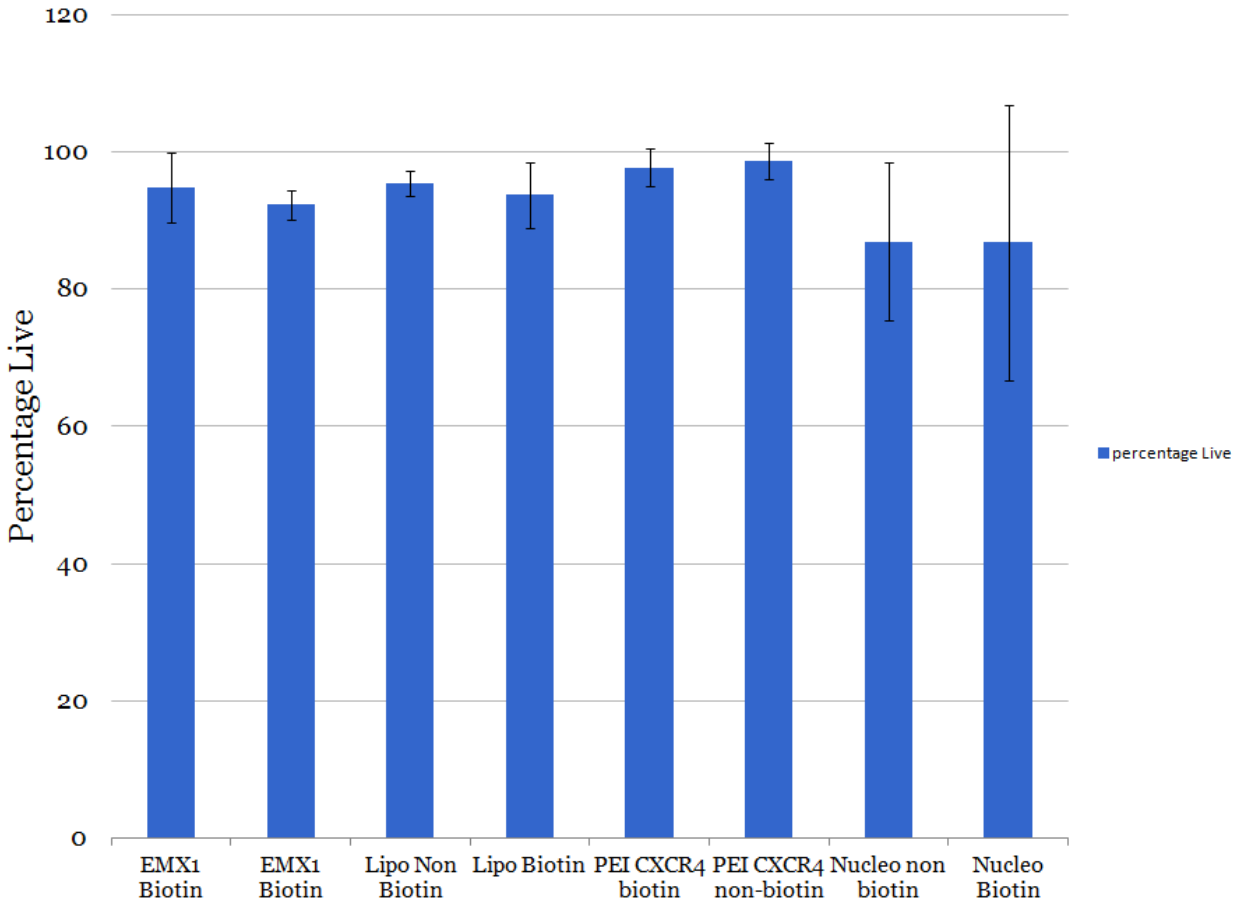


Figure S6 Comparison of Cell viability with each of the transfection methodologies

Deeper Evaluation of HDR by Ion Torrent Sequencing

A key facet of investigating the high frequency of HDR was to explore the direct sequence of libraries derived from our CXCR4 HDR experiments (fig.2 main paper). A criticism of HDR assay by restriction digest insertion is that it can under and overestimate HDR based on band intensity, blurring efficiency of cut and experimental error. While best experimental practice can be followed and the approach has been applied widely by Doudna et al (Lin *et al*, 2014) among others. In brief the introduction of a palindromic sequence enables rapid identification of edits. The degree to which the PCR product from genomic DNA is cleaved based on the introduction gives a measure of the HDR events. Alternative approaches such as reporter assays, can be open to other variables such as donor length and whether the insert is inframe and transcribed correctly. To delve into the occurrence of HDR in this manuscript, Ion torrent sequencing was chosen. By using next generation sequencing, we can explore frequencies for complete HDR inserts, partial inserts/indels and un-edited sequences. With greater coverage than Sanger, which is biased towards the most dominant sequences, a more accurate evaluation of the HDR frequency than the gels was attempted.

Supplemental Methods:

Genomic PCR

All primers were ordered from either IDT or Biocorp and diluted in Rnase free water to 100uM stocks for storage at -20°C. Working stocks of 10uM for all primers were prepared and stored at -20°C until required. PCR was performed with Phusion, Q5 and One Taq (NEB) in 25 µl volumes according to the manufacturer's instructions.

PCR generation of DNA donors and Biotinylated donors.

Generation of donor DNA was performed from gBLOCKS design with restriction digest sites specific to the gene we hoped to edit. All donors of different lengths were performed using match forward and reverse primers to generate asymmetric, equal homology arms variants and biotinylated versions created by substitution of either reverse or forward primer as per diagram in figure 3. Dependent upon the quality of PCR band, gel extraction was performed for clean up.

Donor design was based on a series of conclusions from papers listed in Supplemental Dataset SD1. CF1 and CF3 were designed with 600 and 400bp homology arms that were seen to be optimal for larger double stranded donors. For Fas1 and Fas2, a short sub-optimal double stranded donors. We applied asymmetric HA based on the 5' 37bp arm being shorter than the distal 93bp arm after the insert. Insertion was design to break the PAM and part of the guide sequence to disrupt re-cutting that could mitigate HDR identification. This design rule was incorporated in to CXCR4 and EMX1 donor design templates.

We opted for gBLOCK design templates and PCR reaction of donors, based on ease of experimentation and speed. It also allowed the flexibility to mix 5' biotinylated primers into the reactions as substitutes to create biotinylated versions rapidly.

Double stranded donors were applied in all cases, deliberately for positive reason (CF1/CF3) and negative reasons (Fas1/Fas2) and the choice was borne out in the relatively poor performance of shorter donors when non-biotinylated versions were applied in fig.3. In choosing biotinylated versions we opted for CF1b, F denotes the forward primer carries the biotin, and this was the case with Fas1b and CF3b but Ras1b applied a reverse primer modification.

For donors with TLR, we used a 111 base pair sequence designed by Francis Robert of the Pelletier lab. In a different approach we used a double stranded donor using with 5' or 3' biotinylation, or dual biotinylation, to study the impact. 5' and 3' labelling denotes the orientation of the original single stranded donor design. The 3' mod implies the location of the biotin modification on the antisense strand, as opposed to the 5' label which is upon the sense strand. The donor design is asymmetric and is shorter than the recommended 137 bp used by Richardson et al (Richardson *et al*, 2016).

For EMX1 we deliberately mitigated the HA to below 200bp (total donor length 400bp plus tag repeat sequence and BamH1 site). The intention was to evaluate whether another sub-optimal design as was performed with TLR, could be recovered by using the PEI transfection approach. In achieving ~70% HDR, it was indicative that PEI nuclear location can rescue or mitigate the conventional donor design rules. We have tested the biotinylated version of the EMX1 donor, and found it followed the performance trend of the Fas1b and Ras2b dDNA, recovering a higher HDR frequency by benefiting from donor co-localisation to DSB.

An aspect unexplored in the mechanistic sense in this work is the impact of removing diffusion and nuclear translocation from the dDNA by attachment to Cas9MAV. With other transfection methodologies, nucleofection excepted, the challenges of nuclear import remain and it is possible less than 30% reaches the nucleus from studies of plasmids and presuming free donor DNA diffusion occurs slowly ($3x10^{-8}$cm/s) (Zhou *et al*, 2004), unless under the influence of active transport and importins. This consideration lead us to evaluate the importance of donor:RNP co-

localisation and concentration. In future work we hope to evaluate accurate quantities of Cas9MAV:donor conjugate within the nucleus.

sgRNA synthesis:

IVT one piece is made from attaching T7 promoter upstream of guide sequence to be tested, and a overlap sequence. Overlap sequence matches the reverse complement of the S.pyg scaffold.

Example for CXCR4:

TAATACGACTCACTATAGGG -	GGGCAATGGATTGGTCATCC -	GTTTTAGAGCTAGA
T7 promoter	guide	overlap

Protocol follows the NEB Engen protocol, with one minor alteration. All sgRNA were prepared by Engen sgRNA synthesis by the appending of a T7 promoter sequence and 3' hybridisation sequence to the reverse complement strand. Using the combination of bacterial polymerases and T7 transcriptase in the kit, DNA strand elongation occurs to create a double strand DNA template for the T7. Differing from the standard NEB protocol, reactions were carried out over 12-18hrs, after which DNAse1 treatment was performed for 30mins. DNase1 was heat deactivated in the presence of EDTA to remove MgCl2. sgRNA can be cleaned up using a rapid RNeasy spin columns or similar kits such as RNA clean and concentrator (Zymo Research USA). sgRNA was confirmed for molecular weight using 6M Urea polyacylamide gel electrophoresis. RNA was then quantified using a nanodrop and molarity calculated. Working stocks of sgRNA were prepared at 300-1000uM, with addition of RNase inhibitors and kept at -20°C. Storage solutions were kept at -80°C with Rnase inhibitor at manufacturer's recommended concentration.

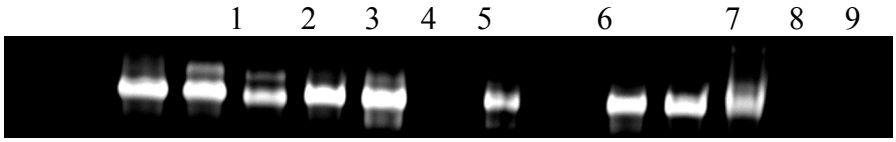


Figure S6 Representative figure for sgRNA synthesis. sgRNA visualised by EtBr and UV illumination. 1 FLCN1 2= FLCN 2, 3= FLCN3, 4=LBK1, 5=ACC1:1, 6=ACC1:2, 7=ACC1:3, 7=Emx1, 8=CXCR4, 9=CD81

DNA extraction methodology:

Pellet was resuspended in 0.2ml PBS, 0.2 ml Lysis buffer with 20mg/ml Proteinase K and incubated for 2-12hrs at 54°C on a dry heating block. After, samples were incubated with 2µl RNAse (10mg/ml) for 30 minutes. 0.2ml 4M GuHCl binding buffer and 0.2ml EtOH was added and the solution inclusive of precipitation was added to econo-spin silica columns. The samples were centrifuged for 3 mins (8000K) and supernatant discarded. 500µl of wash buffer was added to column, which was spun (3min, 8000K). The wash step was repeated again and supernatant discarded. Column was supn for 1 minute to dry silica membrane of column and replace collection tube with an 1.5ml eppendorf tube. Sample was eluted with 50-100µl of elution buffer or DEPC

water by centrifugation (1 minute, 8000K). Quantification of extracted DNA was performed using a Nanodrop spectrometer.
Buffer compositions:

Lysis: 10mM Tris 2mM EDTA 1% SDS,
Binding buffer: 3M GuHCl 3.75M NH_4Ac pH 6,
Wash buffer: 10 mM Tris-HCl pH 7.5, 80% ethanol
Elution: MilliQ water or 10mM Tris:EDTA buffer at pH 8

PEI preparation:

The degree of protonation of PEI has an impact upon polyplex formation and transfection efficiency. We applied a ninhydrin assay to assess the protonation of secondary amine. At pH 7.4 it is suggested that the ideal 40-50% protonation occurs and using the assay an absorbance of 0.025 to 0.03 for PEI incubated with ninhydrin is considered the goal of PEI solution preparation.

In brief a 100mM Ninhydrin solution was prepared in milliQ water. 333µl of Ninhydrin stock and 666µl of PEI solution were vortexed and heated to 75°C for 20 minutes in a dry bath, followed by quenching on ice for 5 minutes. Solution was measured by UV/VIS (487nm) and the absorbance recorded. For evaluation of PEI consistency, a calibration curve was prepared with PEI solutions adjusted from pH1 to pH 14. 40-50% protonation of secondary amines equates to pH7-7.4.

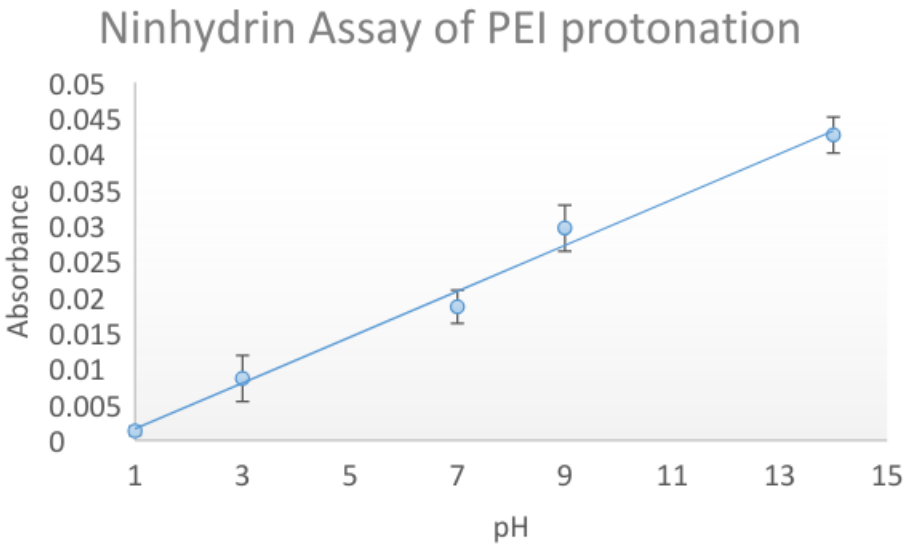


Figure S7 Ninhydrin calibration curve for assessment of secondary amine protonation

PEI labelling

Adapting the method of Godbey et al (Godbey *et al*, 1999), FAM-nhs ester was substituted for oregon green. Briefly PEI (10mg/ml, supplier) was dissolved in 0.1M Sodium bicarbonate and reaction of the primary/secondary amines with FAM-NHS ester (50µl per ml 10mg/ml supplier) was performed at standard laboratory conditions (1atm, 25°C) and magnetically stirred. Vessel

was enclosed to exclude light and photobleaching of fluorescent dye. Resulting FAM-labelled PEI was aliquoted into 200 µl volumes and stored at -20°C until required. For Fam labelling of polyplexes, aliquots are defrosted and vortexed, before adjustment to room temperature. 2-5µl of total 30µl (10mg/ml) unlabelled PEI used to form RNP polyplexes is substituted.

Transfection with FAM-labelled PEI:Cas9MAV

Polyplexes were formed using the standard PEI:Cas9MAV protocol (supplemental) substitute 5µl of unlabelled PEI for FAM:PEI. HEK293T were seeded into a 6 well plate 24hrs before to adhere (seeding density $3x10^5$) to achieve 60-70% confluence at point of transfection. After 24hrs cells were washed with 1x PBS and media refreshed before fluorescent image acquisition using an EVOS cell imager. ImageJ was used for basic processing and image normalisation.

Liposomal delivery of Cas9MAV

As a comparative experiment we performed lipofection using Lipofectamine 3000 (LP3K). The basic protocol for lipofection was adapted from Yu et al [10.1007/s10529-016-2064-9]. To maintain equivalence of experimentation, the same RNP preparation (Cas9MAV, sgRNA concentrations) as for PEI transfection was used for RNP formation and donor addition (biotinylated donors or non-biotinylated donors). 0.3 µl per well of LP3K was used and after ten minutes incubation at 25°C, transfection solutions were added to cells (5.5µl per well). The RNP transfection solution followed the recipe below:

RNP mixture	vol. 5.9µl
LP3K	vol. 0.9µl
Donor	vol. 5µl
OptiMEM	vol. 4.7µl

On two 12 well plate 3 repeats were created for each non-biotinylated donor (CF1, Fas1, Fas2, CF3) and biotinylated variant (CF1b, Fas1b, Ras1b, CF3b). The other alteration to the PEI transfection protocol reflects LP3K stability in serum rich media. Prior to transfection standard DMEM media was removed and replaced with OptiMEM (Gibco). Two hours after transfection the OptiMEM media was replaced by DMEM. Cells were allowed to grow for 48hrs prior to analysis.

Sanger Sequencing and Experimental Conditions of Samples

Table S3 details the experimental conditions (Cas9 variants) and donor used to induce HDR inserts. Number of clones sequenced is also listed. Full sequences are included PDF alignment performed in CLC workbench. Code provided associates sequences to experiments and donors used.

Samples	Cas9 Variant	Sequencing code	Donor Variant
Cas9MAV biotin	Cas9MAV	3	CF1b
Cas9MAV biotin	Cas9MAV	4	CF3b
CF1 non biotin	Cas9 NLS	5	CF1
CF3 non biotin	Cas9 NLS	6	CF3
Empty Vector/controls	N/A	1,2	N/A

Table S3 Sanger sequencing sample Experimental Details and sequencing code

Ion Torrent Sequencing and Preparation:

For ion torrent sequencing design of new primers for a 119 bp region bridging the insert site were designed. Figure s8 shows the genomic location with HDR edit validated by sanger sequencing. Primer 1 and Primer 2 denote the genomic designed for CXCR4.

Figure S8 Design of genomic primers for ion torrent sequencing. The sequence has been edited to show the location of the proposed HinDIII site.

In the context of Ion torrent sequencing addition sequences needed to be appended to the forward and reverse primers.

Name	Primer Position	Sequence 5'-3'	Notes

A-Key	Forward	CCATCTCATCCCTGCGTGTCTCCGACTCAG	Barcode sequence should follow this Key
P1-Extended Key	Reverse	CCACTACGCCTCCGCTTTCCTCTCTATGGGCAGTCGGTGAT	Extended version of P1-Key for TAQMAN quantification

Table S4 Sequences for A-key and P1 for addition to forward and reverse primers

At the 5' end of the forward primer a 10-12 ntp barcoding sequence was added, unique to each sample. The reverse primer was appended at the 5' end with the P1 key, containing a taqman probe target sequence, as important as part of the ion torrent work flow.

An initial amplification is performed on all samples using One Taq from genomic DNA samples extracted from un-edited and HDR edited cells (different donors, state of biotinylation). Once confirmed by gel electrophoresis (2% agarose), confirming product size as 200 bp and forming a single product. The products for each sample were gel extracted and purified with a Qiagen QIAquick gel extraction kit. Each extraction had between 2-5ng/µl. 5µl of each sample were combined in a clean 1.5ml eppendorf, average sample concentration was given a 4ng/ul and the quantity of total DNA in 35µl was estimated to be 0.14µg, equating to a molarity of 30.46nM. Samples were then combined and diluted to 8 pmol total DNA ready for emulsion PCR as part of the ion torrent work flow. Templating and sequencing were performed according to the Ion torrent ION PGM template OT2 200 kit and ION PGM Hi-Q sequencing protocols.

Samples	Donor used	Cas9 used	Description	Primer Code
1	CF1 biotin	Cas9MAV	HDR	FBEST
2	Fas1 non biotin	Cas9MAV	HDR	F60
3	CF3 non biotin	Cas9MAV	HDR	BC4
4	CF1 non biotin	NLS	HDR	BC5
5	Fas1 biotin	Cas9 MAV	HDR	BC6
6	No donor	Cas9 MAV	Indel	BC7
7	No Donor	Cas9 (no sgRNA)	Control unedited	F CONTROL

Table S5 Description of samples selected for Ion torrent sequencing.

BC7 library sequencing was not included in data analysis due to minimal sequences completed in comparison to other libraries. Sequences are included in the FastQ data file but not used for further analysis.

Library name	A key	Barcode (black) Primer (blue
F CONTROL	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGAAGGCCACAC CAGCCTGTACTTGTCCGTCAT
FBEST	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGATCGGTTC CAGCCTGTACTTGTCCGTCAT
F60	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCTGCCTGTC CAGCCTGTACTTGTCCGTCAT
BC4	CCATCTCATCCCTGCGTGTCTCCGACTCAG	TCAGGAATAC CAGCCTGTACTTGTCCGTCAT
BC5	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGGAAGAACCTC CAGCCTGTACTTGTCCGTCAT
BC6	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CGAAGCGATTCCAGCCTGTACTTGTCCGTCAT
BC7	CCATCTCATCCCTGCGTGTCTCCGACTCAG	CAGCCAATTCTC CAGCCTGTACTTGTCCGTCAT

Table S6 Libraries for Ion Torrent and association to Barcode. Libraries derive their name from forward primer employed for each separate sample.

Provided below are tables of alignment and frequency of occurrence for each nucleotide that differs from the predicted sequence. Predicted sequences are either the un-edited genomic (length 17bp), or the expected edited sequence (length 19bp), focusing upon the sequence comparison tables were generated to show where the frequency of alternative base assignments after BLAST analysis.

Attached as supplemental data are the FastQ raw reads, and excel spreadsheet investigating knockin variants by analysis of alternative base pair assignments. The most common knock in variant assessed results from a single nucleotide insertion. Given the nature of Ion torrent sequencing being prone to indel formation this sub-population was assigned separate from complete and accurate knockins, as uncertainty regarding assigning it as a genuine knockin variant was precluded.

Mycoplasma Testing:

For cell culture contamination testing through the period of experimentation (September to December 2017) we applied a PCR based mycoplasma test (Zmtech, Montreal). Cells and culture media were sampled bi weekly and PCR was performed according to manufacturer's protocol.

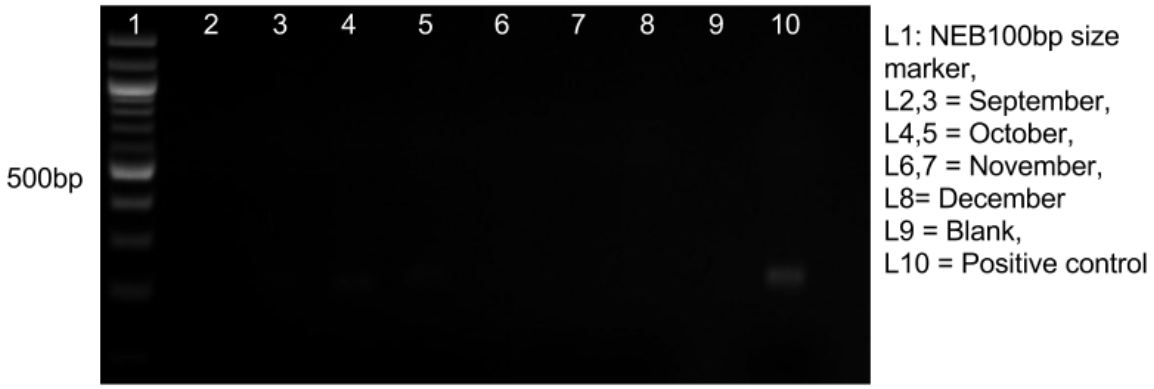


Figure S9. Summary of Mycoplasma testing of Cell cultures over period of experimentation

Summary of Supplemental Data sets:

SD1 - Tabular summary of HDR improvement strategies
SD2 - Plasmid Map of Cas9MAV
SD3 - Excel spreadsheet for gel quantifications for T7 and Restriction Digest HDR assay
SD4 - TLR flow cytometry statistics
SD5 - PDF file of Sanger sequenced clone raw reads and chromatograms
SD6 - Analysis of Ion torrent sequencing
SD7 - Fastq file for the ion torrent libraires
SD8 - Ion torrent sequencing data processing code