Multi-reporter selection for the design of active and more specific zinc-finger nucleases for genome editing

Benjamin L. Oakes1,*, Danny F. Xia2,*, Elizabeth F. Rowland1,*, Denise J. Xu1, Irina Ankoudinova2, Jennifer S. Borchardt2, Lei Zhang2, Patrick Li2, Jeffrey C. Miller2, Edward J. Rebar2 & Marcus B. Noyes1,3,4,5

Engineered nucleases have transformed biological research and offer great therapeutic potential by enabling the straightforward modification of desired genomic sequences. While many nuclease platforms have proven functional, all can produce unanticipated off-target lesions and have difficulty discriminating between homologous sequences, limiting their therapeutic application. Here we describe a multi-reporter selection system that allows the screening of large protein libraries to uncover variants able to discriminate between sequences with substantial homology. We have used this system to identify zinc-finger nucleases that exhibit high cleavage activity (up to 60% indels) at their targets within the CCR5 and HBB genes and strong discrimination against homologous sequences within CCR2 and HBD. An unbiased screen for off-target lesions using a novel set of CCR5-targeting nucleases confirms negligible CCR2 activity and demonstrates minimal off-target activity genome wide. This system offers a straightforward approach to generate nucleases that discriminate between similar targets and provide exceptional genome-wide specificity.

1 The Lewis-Sigler Institute for Integrative Genomics, Princeton University, Princeton, New Jersey 08544, USA. 2 Sangamo BioSciences Inc., Richmond, California 94804, USA. 3 Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA. 4 Institute for Systems Genetics, NYU School of Medicine, New York, New York 10016, USA. 5 Department of Biochemistry and Molecular Pharmacology, NYU School of Medicine, New York, New York 10016, USA. * These authors contributed equally to this work. Correspondence and requests for materials should be addressed to M.B.N. (email: marcus.noyes@nyumc.org).
Desginer nucleases have revolutionized studies of gene structure and function in higher organisms by enabling the simple and efficient manipulation of specified genomic sequences. While several different nuclease platforms have been developed for this purpose, the strategy they employ is the same: direct a double-strand break (DSB) to a desired genomic locus and allow repair mechanisms to edit the sequence. If the break is repaired by mutagenic non-homologous end joining (NHEJ), the insertions or deletions (indels) that often result can lead to frame shifts and functional knockouts. If the break is repaired by homology-directed repair (HDR), the sequence can be rewritten if an appropriate donor template of DNA is provided. This approach has proven successful across diverse species and platforms.

As applications for designer nucleases have multiplied and extended into more sensitive areas such as stem cell biology and therapeutic, the focus of design efforts has shifted from improving activity to optimizing genome-wide specificity. A key element of this shift has been the emerging realization that nucleases designed for genome-editing applications will frequently need to discriminate against highly similar off-target sequences, due to the high frequency of repeat elements, gene duplications and pseudogenes in eukaryotic genomes. Further complicating matters, successful HDR can be limited by the distance between the DSB and the sequence to be modified. For these reasons, the precise location of a DSB may be critical, and similarity with potential off-target sequences unavoidable.

The crux of a high-fidelity designer nuclease are the abilities to both strongly specify a desired sequence and to avoid others. While strong in vivo on-target activity has been demonstrated for all the designer nuclease platforms (zinc-finger nucleases (ZFNs), homing endonucleases, transcription activator-like effector domains (TALENs) and RNA-guided CRISPR-Cas9 systems (RGEns)), fidelity remains uncertain. Early-stage ZFNs, TALENs and RGENs can, in certain cases, show substantial activity at alternative sequences similar to the designed target as well as unanticipated sequences. Several approaches have been employed to address these issues. Nicking RGENs, Cas9-Fok1 fusions, RNA guide length and alternative TALE domains have all been employed to address concerns of off-target activity. Furthermore, algorithms that allow the user to choose a specific target sequence based on its reduced risk of off-target activity have been developed. However, these approaches limit therapeutic applications that require precise modification of sequences with high degrees of homology to other regions of the genome.

To discriminate between similar targets, a complex protein–DNA interface with great engineering potential, coupled with a powerful selection system for identifying new designs that exhibit desired binding properties, may be required. Therefore, the complexities of intra- and inter-finger interactions offered by the zinc-finger domain may provide advantages over the more modular TALEN and RGEN platforms. However, until recently diverse pools of zinc fingers able to bind all 3-nt (nucleotide) targets were unavailable, and few selection systems allowed for the recovery of proteins with discriminatory attributes. Prior ZFN engineering for in vivo application by both bacterial one and two-hybrid assays have been primarily limited to pools of binding zinc fingers and do not present a counter-selective element that would produce zinc fingers able to discriminate between closely related sequences. Therefore, this new complete zinc-finger pool set coupled with a system that provides designer nucleases, zinc finger or otherwise, that can distinguish between two closely related sequences would be of great utility.

We recently described a new resource of zinc-finger pools with significantly greater depth than previously available. Here we describe a multi-reporter selection system to screen libraries of these zinc fingers for combinations that can discriminate between two similar targets. We apply this system to identify ZFNs that provide excellent discrimination between highly homologous sequences for two targets in vivo: CCR5 versus CCR2 and haemoglobin beta (HBB) versus haemoglobin delta (HBD). In each case we recover ZFN pairs that manifest strong on-target activity with minimal or no detectible activity at the homologous off-target sequence. Finally, in an unbiased screen of genome-wide off-target activity for a subset of the CCR5 ZFN pairs, we identify just a small number of low-frequency off-target loci. For one of these ZFN pairs, we identify just a single off-target locus modified above 0.1% lesion frequency at this locus demonstrating exceptionally genome-wide fidelity. This work demonstrates the ability of the multi-reporter selection system to uncover proteins with fine-tuned specificity and the potential for pristine genome editing.

Results Establishment of a multi-reporter selection system. The omega-based bacterial one-hybrid (B1H) system has proven a simple and extremely sensitive method for the investigation of protein–DNA interactions (see overview, Supplementary Note 1). This system differentiates itself from other bacterial hybrid assays through the employment of the omega subunit of RNA polymerase (rpmZ) as the fusion partner to the protein of interest. In this way omega acts as an activation domain through recruitment of the polymerase. Omega is a non-essential component of the core holoenzyme allowing selections to be carried out in an rpmZ knockout strain and therefore in the absence of competition from endogenous omega. The lack of competition allows activation of the reporter even at low levels of fusion protein expression and the recovery of interactions with a large range of affinities. As a result, the method has allowed the characterization of transcription factor DNA-binding specificities for most common DNA-binding domain families as well as the selection of synthetic homeodomains and zinc fingers with novel specificities.

For these studies, we improved this method and adapted it to the challenge of providing highly effective DNA-binding proteins that could discriminate homologous sequences. To accomplish this, we first created a variation of the B1H system in which its two selectable markers (HIS3 and URA3; Fig. 1a) are expressed from independent promoters. Our goal in doing this was to enable simultaneous selection for or against activation of these two reporter genes. To test whether our new system accomplished this, we compared the activity of the reporters while varying known protein–DNA interactions that drive their expression. Here we expressed the three-finger Zif268 zinc-finger protein as a direct fusion to omega. We fixed the Zif268 consensus target sequence upstream of the promoter that drives the URA3 reporter and paired this with alternative Zif268 targets that exhibit a range of affinities upstream of the HIS3 reporter. Bacteria bearing each Zif268-binding site combination were grown to log phase, titred and plated on selective media. Only cells that offered a functional interaction to drive URA3 survived the presence of 6-azauracil (6-aza) while the affinity of the interaction that drives HIS3 determined survival at various stringency 3-amino triazole (3-AT) concentrations (Fig. 1b). The reverse of this is also true when the sequences in sites 1 and 2 are switched (Supplementary Fig. 1). We confirmed these observations in follow-up studies of doubling time in liquid media (Supplementary Fig. 2). In sum, while maintaining a desired protein–DNA interaction that drives one reporter, the
interaction that drives the secondary reporter can be selected for, or against, through the addition of inhibitors in the media.

By separating the HIS3 and URA3 reporters we are able to investigate two independent interactions simultaneously. Still, survival alone does not allow us to differentiate between the attributes of multiple, functional interactions, only indicating that they all surpass the survival threshold required of the selection. Therefore, as a second improvement to the original B1H system, we added fluorescent reporters to each of the selectable markers to provide a secondary and more graded measure of activity (henceforth referred to as the multi-reporter bacterial one-hybrid system (MR-B1H)). A green fluorescent protein (GFP) cassette is expressed from the same promoter that drives HIS3 expression, and mCherry expressed from the promoter that drives URA3 (Fig. 1a). To demonstrate the utility of this approach, we again turned to Zif268. While fixing the interaction that drives URA3 expression, we tested the same set of binding sites as above to drive HIS3 and therefore, GFP expression (Fig. 1c). With this system we observed that under varying conditions of URA3 expression, survival is related to the affinity of the protein–DNA interaction that drives its expression and unrelated to the selection conditions impacting the competing binding site and reporter. The same cells tested in c, as well as a complete set of the site 1 sequences paired with the negative ‘b’ sequence, were grown in either rich media or selective media that impacts only URA3 expression. While HIS3 expression is not selected for, GFP output is related to the affinity of the interaction that drives the HIS3/GFP reporter and unrelated to the growth conditions that impact site 2 and URA3/mCherry.

results confirm that selection conditions designed to influence URA3 expression do not impact the HIS3/GFP transcription. Therefore, activation of the reporter is a function of the protein–DNA interaction that drives its expression and the fluorescent output is a secondary measure of the activity that interaction offers.

Selection of proteins by target discrimination. Having shown that the MR-B1H system can function as a reporter to compare the relative strengths of two interactions, we next sought to demonstrate that this system can be used to select proteins with new binding properties. We first prepared libraries of four-finger zinc-finger DNA-binding proteins using our recently reported, complete set of selected zinc-finger pools38,52–54 as templates for library assembly (Supplementary Fig. 4). This approach was guided by prior strategies for assembly of zinc-finger pools36,32–34. By selecting four-finger variants from diverse pool assemblies we circumvent inter-finger complications that often arise when neighbouring fingers of known specificity are designed next to one another. Rather, we select combinations of zinc fingers from four-finger libraries that are most compatible with one another.

To expand the utility of the MR-B1H system, we screened our libraries, created from our individual pools, to uncover proteins able to discriminate between related sequences. For our initial
within a recognition helix may influence one another. Therefore, we took a conservative approach and chose only complete, selected ZFPs to follow-up in all experiments reported here. The SELEX results revealed generally good specificity by the selected ZFPs, with several exhibiting a marked preference for the intended target base at those positions differing between CCR5 and CCR2 (Fig. 3a,b). Next, we expressed these candidates as ZFN pairs in K562 cells and quantified their activity at the CCR5 and CCR2 targets (Fig. 3c). For comparison, a published set of CCR5-targeted ZFNs55 were tested in parallel with the MR-B1H-produced ZFNs. We found that our selected ZFNs were highly active, yielding up to 50% indels at CCR5. Moreover, many were also highly selective for CCR5 versus CCR2, with half yielding a modification ratio of >12. Interestingly, while the left monomers have some influence, the CCR2 activity of the MR-B1H-produced ZFNs appears to be primarily related to the specificity of the right monomer and its ability to bind the adenine that differentiates CCR2 from CCR5 (see SELEX results, Fig. 3b). Those with no SELEX evidence of binding adenine at this position (candidates 46696 and 46697) offer CCR5 to CCR2 indel ratios that range from 12.9 to 41. For comparison, under these experimental conditions the previously published ZFNs yielded 73% modification of CCR5 in this study, with a threefold preference versus cleavage of CCR2.

**Extending fine-tuned ZFNs reduces competitive activity.** The MR-B1H system can uncover ZFPs with strong discrimination between two targets that differ by a single base pair, even when we have restricted ourselves to duplicate an exact target in the literature. However, with a complete C2H2 zinc-finger pool set we are not limited by sequence and have the flexibility to slightly shift and expand the zinc-finger target, if advantageous, and still remain in close proximity to the sequence to be modified. Therefore, we reasoned that by increasing the number of fingers per monomer and maximizing the counter selection by focusing on the divergence between homologous targets, we could further improve the discrimination that our ZFNs are able to offer.

To test this approach we shifted the CCR5 target 6bp 3′ (Supplementary Fig. 6). We also increased the number of mismatches between the CCR5 and CCR2 targets by extending each ZFP to contain six fingers per monomer (Fig. 4a and Supplementary Fig. 7). To limit library size and maintain the complexity for each finger, zinc fingers were selected from two overlapping four-finger libraries to bind these targets (selected ZFPs shown in Fig. 4b and Supplementary Fig. 7). In this way, the two C-terminal and two N-terminal fingers of the overlapping libraries recognize the same sequences (purple ovals in Fig. 4b). Common fingers recovered in the ‘overlap positions’ of both libraries were used as guides to design six-finger proteins. However, four-finger proteins recovered to bind the 12-nt target proximal to the cut site can also be used directly. In this way, both four- and six-finger proteins were selected using the counter-selection assay detailed above.

Two four- and two six-finger proteins that target the shifted CCR5 sequence were tested for their function as ZFNs in K562 cells. These ZFNs are related in that the four C-terminal fingers of the six-finger proteins are identical to one of the four-finger variants (see Fig. 4, table). Therefore, any differences in activity are due to the addition of two N-terminal fingers. As above, the indel frequencies at both the CCR5 and CCR2 loci were measured for each ZFN pair (Fig. 4c). Interestingly, the number of fingers has a large impact on the right target but not the left. Only ZFNs that utilize right monomers with six fingers provide high CCR5 indel frequencies, ranging from 19 to 61%. What is more, for each of these eight combinations the CCR2 indel frequency ranges from 0.09 to 0.14%, similar to the background frequency found

**Fine-tuned ZFNs improve CCR5–CCR2 discrimination in vivo.** The MR-B1H system is able to provide target discrimination in human cells. To test our selected ZFPs outside of bacteria, we first repeated the procedure above to select ZFPs able to bind the published 5′ CCR5 target (left target) while discriminating against the CCR2 sequence (Supplementary Fig. 5). Next, we assessed the DNA-binding specificity of a set of selected ‘left’ and ‘right’ ZFPs using systematic evolution of ligands by exponential enrichment (SELEX). For this study, candidates were chosen as intact proteins rather than designed from a consensus to preserve context-dependent aspects of performance. As shown in our previous studies41 all amino acids

---

study, we targeted a well-characterized sequence within the CCR5 gene, at which indels can mediate efficient functional inactivation and cellular resistance to HIV infection22,53–57. This locus is provided an attractive initial test of our selection system, since it exhibits substantial sequence identity with a second sequence in the human genome (within the homologous CCR2 gene), and the availability of highly active and specific published reagents22 for this target would provide a benchmark against which to gauge the performance of any selected ZFNs. To utilize this tool, we modified the MR-B1H system to select zinc-finger arrays able to bind the CCR5 target and provide discrimination against the CCR2 target (Fig. 2a).

The 12-nt sequence targeted by the published 3′ CCR5 ZFN monomer22 (right target) was installed upstream of the HIS3/GFP reporter (Fig. 2a). The homologous CCR2 sequence (matching at 11/12 bases) was installed upstream of the URA3/mcherry reporter. ZFP array libraries were expressed as omega fusions and paired with this CCR5-focused MR-B1H reporter vector (Supplementary Fig. 4). A low-stringency, HIS3-positive selection was performed to remove non-functional arrays from the library. The surviving library members were pooled, again paired with the reporters and selected for activation of HIS3 (CCR5 target), with a secondary selection for, against or neutral for URA3 (CCR2 target). Classification of these cells as GFP or mCherry positive was defined by setting gates to established background-level fluorescence using a negative control that should not actively express either fluorescent reporter (see Fig. 2c as an example). With experimental samples, when activation of URA3 is selected against, the number of cells above the GFP background but below the mCherry threshold increased by 2.5-fold over cells grown with HIS3 selection alone (Fig. 2b, top, quadrant 4). Moreover, these conditions produced a stringent population of high GFP and low mCherry activities (Fig. 2b, pop 3) enriched by 10-fold. Conversely, when activation of URA3 is selected for, 80% of the enriched cells are both GFP and mCherry positive (Fig. 2b, bottom), a 4-fold enrichment over the HIS3 selection alone. Populations of these cells were recovered and the ZFPs they harbour were sequenced (Fig. 2b, right). The amino acids enriched in the alpha helix of the N-terminal finger are noticeably different depending on the URA3 selection conditions (Fig. 2b, green box). On the basis of the canonical model of ZFP–DNA recognition, this first helix should bind to the AAG and AAA triplets that differentiate BS1 from BS2 of the reporter, respectively. Candidate ZFPs that represent these populations were tested again with the reporters to confirm the fluorescent activity, and thus their DNA-binding attributes, in the absence of selective pressure (Fig. 2c). These studies confirmed that we can select new ZFP arrays from combinatorial libraries that offer fine-tuned attributes by modifying selection conditions and recovering fluorescent populations that represent the characteristics we desire.

---
Figure 2 | Selection of zinc fingers that can discriminate between similar targets. (a) Diagram depicts the selection process. The desired and counter targets are placed in front of the promoters that drive HIS3 and URA3 expression, respectively. Zinc-finger pools previously selected to bind each 3 bp subsite of the desired target are used as PCR templates to assemble a four-finger library, illustrated as rainbow-coloured ovals. This four-finger library is expressed as an omega fusion. To select four-finger members of this library, which are able to discriminate between the desired targets, cells are grown under conditions that are inhibited by URA3 expression but required HIS3 activation. A workflow of the procedure is shown below. (b) Selection conditions influence the enriched amino acids that correspond to the target mismatch. Using the library described in a, selection for HIS3 activation but against URA3 activation increases the fraction of the population in the GFP-positive (Pos), mCherry-negative (Neg) quadrant 4 in comparison with a HIS3-positive selection alone (top). Using the same library, selection for both HIS3 and URA3 activations increases the fraction of the population in the GFP-positive, mCherry-positive quadrant 2 in comparison to a HIS3-positive selection alone (bottom). Sequencing the zinc fingers recovered from stringent populations of these selection conditions reveal a stark difference in the amino acids enriched in the helix that corresponds to the difference in the desired and counter target (green box). (c) The binding attributes are confirmed for zinc-finger candidates that represent the recovered pools in b. Zinc-finger candidates are paired with the CCR5 versus CCR2 reporter vector and grown without selection. The GFP versus mCherry attributes are complementary to the selection conditions from which the candidate protein was derived. The sequence of the tested zinc-finger helices are shown (N to C) above the dot plots with the discriminator helix shown in colour.
The indel frequency recovered from a control that expressed GFP rather than a nuclease pair is shown below. The published candidates 8266–20505 were tested in parallel for reference. The ratio of CCR5 to CCR2 indel frequency is shown in the right table. boxed at the top. (Those that tolerate binding to the selected-against base are shown as red numbers. The previously published helices for each left and right monomers are green. The recovered percentage of sequences that correspond to the CCR5 or CCR2 base at this critical position are listed to the right of each plot.

From a GFP-transfected control sample. As a result, we are able to leverage differences in the CCR5 and CCR2 sequences, while remaining in close proximity to the desired cut site, by slightly shifting and expanding our target. By doing so we have created ZFNs that offer strong CCR5 activity and little if any activity at CCR2.

A second disease-related target complicated by the need for discrimination against a highly conserved homologue is HBB. Mutations in the HBB sequence lead to sickle cell anaemia as well as other blood-borne diseases. The coding sequences for HBB and HBD are 93% identical in human making it difficult to modify HBB without altering HBD. Using the same approach as outlined above, we focused the selection of zinc fingers to bind directly at a mutant HBB sequence that causes sickle cell anaemia (Fig. 5a and Supplementary Fig. 8B). From these results, a four- and six-finger protein that bind the right and left targets were paired (see Fig. 5, table) and tested as ZFNs in K562 cells. For each pair, indel frequencies were measured at both the HBB and HBD loci (Fig. 5c). Interestingly, while these ZFNs produce high HBB indel frequencies regardless of finger number, the HBD indel frequency is largely dependent on the number of fingers in the left monomer. In both cases, the six-finger left monomers increase HBD indel frequencies from background levels to low percentages. As the extension of the left monomer does not pick up additional differences between the HBB and HBD sequences, it is possible the extended recognition increases the affinity at the HBD target while lessening the consequence of the single mismatch present in the left monomer targets. These results imply that the extension of a nuclease target that does not
increase the number of mismatches relative to a similar sequence in the genome may lead to an increase in off-target activity. Regardless, we have produced ZFNs that target the sickle cell mutation in HBB with high activity and in some cases, background levels of HBD activity.

**Unbiased genomic screen shows negligible off-target activity.** The ability of a ZFP to discriminate between closely related sequences is a powerful indicator of *in vivo* specificity but does not eliminate the possibility that these fine-tuned nucleases bind and modify other, less-predictable sequences. This is apparent from unbiased screens of genome-wide cutting for ZFNs\(^{22}\), TALENs\(^{24,25}\) and RGENs\(^{25,28}\), which in some cases have revealed cleavage at sequences that are more diverged from the desired target relative to other, uncleaved sequences, demonstrating that genome-wide fidelity can be more complicated than sequence similarity alone. Therefore, to map off-target loci for our different CCR5 ZFN pairs, we used a modified version of this integrase-defective lentiviral vector (IDLV) capture and mapping protocol, which we have previously described\(^22\). To increase the sensitivity of the assay, we used the high sequencing depth provided by Illumina sequencing and performed the assay in biologic triplicate to minimize apparent clusters of IDLV integrations caused simply by obtaining more sequence reads. We ranked clusters of IDLV integrations (defined as IDLV integrations within 1,000 bp of other IDLV integrations) based on the total number of integrations, the number of replicates containing those integrations and the ratio of integrations in the ZFN-treated samples to the control samples (see Methods). We then designed PCR primers to amplify the top 20 ranked clusters for each ZFN pair and characterized the indel frequency in K562 cells at each of these 20 potential off-target sites per ZFN pair; to provide additional information we also characterized indel frequencies at loci corresponding to the top 20 ranked clusters for other CCR5 ZFNs that target similar sequences (that is, we tested the activity of 46698:46705 at clusters obtained with either 46698:46705 or 46700:46705). For the previously described 8266:20505 ZFN pair, this process revealed nine active off-target sites with an aggregate activity of 33.3% (Table 1). For the two new ZFN pairs that targeted the same sequence, 46693:46696 and 46693:46697, this process yielded 11 active off-target sites with an aggregate activity of 21.2% and 10 active off-target sites with an aggregate off-target activity of 6.6%, respectively. However, it is difficult to discern whether the decrease in aggregate off-target activity is due to improved genome-wide specificity compared with the previously described ZFN pair or simply a reflection of their lower overall activity under these conditions (Table 1).

The specificity of the new CCR5 ZFN pairs that target a slightly shifted and extended sequence, 46698:46705 and 46700:46705, allow a more meaningful comparison because they have similar activity to the previously described ZFN pair at the intended CCR5 site. For 46698:46705, four active off-target sites were identified with an aggregate activity of 1.5% (Table 2). The best results were obtained with 46700:46705 that had three active off-target sites and an aggregate activity of 0.5% (Table 2). This represents 22- and 67-fold decreases in aggregate off-target activity, respectively, relative to the published CCR5 ZFNs.

![Diagram](image)

**Figure 4 | Extending zinc-finger targets to increase CCR2 discrimination.** (a) The CCR5 target is shifted 6 nt 3′ relative to the target in Fig. 3 (bold and underlined). Mismatches with the CCR2 sequence are shown as red letters. Each ZFN monomer is increased from four to six fingers. (b) For each target, two overlapping four-finger libraries are produced. The right monomer pools are shown here while the left monomer pools are shown in Supplementary Fig. 7. Pools of these libraries are colour coded to emphasize that the overlapping zinc fingers target the same sequences. Zinc fingers are selected from each pool by selection for the CCR5 sequence but against the CCR2 sequence. Targets are shown 3′ to 5′ to emphasize the overlap in the targets of the four-finger selections. From each of these selections, 10 of the selected ZFPs are shown. Candidates (Cand.) used to design the four- and six-finger monomers employed as nucleases are bold and underlined. All enriched amino acids for each of the four-finger selections are shown below as a sequence logo with the overlapping two fingers boxed in purple. (c) Candidate ZFN pairs were expressed *in vivo* and the percentage of indels at CCR5 and CCR2 measured. Indel frequencies recovered at either target from cells that did not express a nuclease (GFP) are shown below each table. The ratio of CCR5 to CCR2 indel frequency is shown below. A table of the four- and six-finger zinc-finger helices used in the nuclease studies, shown N term to C term, is provided.
Together these results indicate that the MR-B1H selection system allows the recovery of zinc fingers that can distinguish between homologous targets and at least in this case, offer excellent genome-wide fidelity. This stringent level of specificity was further supported by SELEX (Supplementary Fig. 9).

Discussion

Here we describe the development of a multi-reporter selection system that enables the assay and selection of zinc fingers able to discriminate between similar targets. Zinc fingers were selected and employed as nucleases in human cells to differentiate between the CCR5 and CCR2 genes, as well as the HBB and HBD should be italicized in this context genes. These ZFNs provide significant discrimination between these targets offering strong on-target activity and in many cases negligible activity at the designed against off-target sequence. For the most promising CCR5 candidates an unbiased screen of genome-wide nuclease activity uncovered only a small number of low-frequency off-targets and in the best case, aggregate off-target activity of just 0.5%.

A number of other strategies for improving the cellular specificity of targeted nucleases have also been developed, including the use of dual nickases, alternative delivery procedures, and target-choosing algorithms that minimize homology between the target site and the rest of the genome. While strategies to measure off-target activity genome wide have been developed and demonstrate unanticipated levels of off-target activity in some nuclease systems, previously described specificity optimization approaches have generally been tested at limited numbers of bioinformatically predicted off-target sites and still failed to completely abolish off-target activity. Given this, it seems likely that ultimate specificity will require the application of multiple orthogonal approaches. We note that our strategy for improving specificity—selection-based engineering of the protein–DNA interface—may be combined with these other methods. In this work, however, to demonstrate the utility of our methods we have deliberately avoided combining approaches and have taken on a ‘worst case scenario’ selecting nucleases to target sequences complicated by homology at other sites in the genome. We are able to produce ZFNs that discriminate between these similar targets and still return genome-wide fidelity that is similar to or superior to any reported nuclease.

Our studies tested both four- and six-finger proteins, which allowed for an initial assessment of how target length influences activity, homologue discrimination and global specificity. Complex behaviour was observed. In the CCR5 studies, extra fingers substantially and consistently improved the activity of the right ZFNs (Fig. 4c, compare 46704 versus 46702 and 46705 versus 46703), but similar extensions to the left ZFNs yielded more modest effects (both increased and decreased activity). However, increasing the length of the left ZFNs did reduce off-target activity in our unbiased genome screen from 1.46 to 1.05% aggregate cleavage (Table 1). In our HBB studies, extension of the left ZFN—but not the right one—negatively impacted homologue discrimination. One possible explanation for this result is that under our study conditions the left ZFN plays a relatively more important role in distinguishing HBB from HBD, and that extension of the its target lessens the impact of binding the single mismatch that distinguishes the HBB and HBD
### Table 1 | Indel frequencies recovered at genomic loci as predicted across IDLV screens using four-fingered zinc-finger monomers.

| chr | CCR5 | KRR1 | RR1F1 | CCR9 | ZCCHC4 | Indel | Total | IDLV | pval | Total | IDLV | pval | Total | IDLV | pval |
|-----|------|------|-------|------|--------|-------|-------|------|------|-------|------|------|-------|------|------|
| 12  | 52455623 | 14755623 | 22445623 | 12445623 | 2055623 | 46693 | 46697 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 14  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 17  | 132054562 | 20304562 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 19  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
|     | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 2  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 9  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 2  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 9  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |

Loci are colour coded by the IDLV screen from which they were recovered.

### Table 2 | Indel frequencies recovered at genomic loci as predicted across IDLV screens using six-fingered zinc-finger monomers.

| chr | CCR5 | KRR1 | RR1F1 | CCR9 | ZCCHC4 | Indel | Total | IDLV | pval | Total | IDLV | pval | Total | IDLV | pval |
|-----|------|------|-------|------|--------|-------|-------|------|------|-------|------|------|-------|------|------|
| 12  | 52455623 | 14755623 | 22445623 | 12445623 | 2055623 | 46693 | 46697 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 14  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 17  | 132054562 | 20304562 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 19  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
|     | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 2  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 9  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 2  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |
| 9  | 3055623 | 4565623 | 2145623 | 125623 | 46697 | 46693 | 140537415 | 48768 | 0.000 | 14738 | 1 | 0.01 |

Loci are colour coded by the IDLV screen from which they were recovered.
sequences. In addition, here we report examples of nucleases that offer both high specificity and activity as well as examples where specificity is achieved at a small cost of activity. It is clear there is much left to be learned about how target length and composition influences the performance of ZFNs and other designed nucleases.

Finally, we have focused the MR-B1H system on zinc fingers because we reasoned the complex intra- and inter-finger interactions that lead to the context-specific activity of many zinc fingers may provide the sensitive ‘all or nothing’ behaviour required to bind one target and discriminate against all similar sequences. Without context dependence, the modular TALE domain required to bind one target and discriminate against all similar interactions that lead to the context-specific activity of many because we reasoned the complex intra- and inter-finger selection pressure. Cells were diluted 1:150 into a optically clear flat bottom 96-well (Sigma, Z380059) and placed in a plate reader. The plates were grown shaking at 16 °C overnight (minimum of 16 h). The ligaton was ethanol precipitated and resuspended in 1 μl of water per 1 μg of vector backbone used in the ligation. Detailed zinc-finger and vector sequences are provided in Supplementary Notes 2–4.

**Zinc-finger protein selections.** To select four-finger proteins able to bind 12 nt of interest, the library assemblies described above were paired with the appropriate multi-reporter vector and transformed into our selection strain. Detailed binding site and vector sequences are provided in Supplementary Notes 2–5. For each transformation 1 μl of the ligation (loosely representing 1 μg of library vector) was paired with 1 μl of multi-reporter vector (500 ng–1 μg) and transformed into our ArpsZΔ cell line. Each pool was cultured without uracil and histidine. A volume of 20 μl of ligation was ethanol precipitated and resuspended in 1 μl of media that contained kanamycin and ampicillin. The cells were again expanded for 1 h at 37 °C. The cells were pelleted and washed four times in NM without uracil or histidine. The cells were resuspended in 1 μl of NM without uracil and histidine. 20 μl of this resuspension was used in 10-fold dilutions on rich media plates while the remaining 980 μl were stored at 4 °C overnight. The following day, cell titres provide a cell count per volume. A total of 2 × 10^6 cells were plated on NM plates containing 5 mM 3-AT to provide a low-stringency positive selection and remove non-functional zinc fingers. These plates were incubated at 37 °C for 36–48 h. After incubation, 1 μl of each pool was used to start a 15-mL culture of minimal media containing 100 μM IPTG and 10 mM 3-AT and 2 mM 5-FOA. These cultures were grown for 24–30 h at 30 °C but not allowed to reach OD600 above 0.8. Cells were recovered and resuspended in PBS plus 0.1% Tween for sort preparation.

**Fluorescence-activated cell sorting and recovery.** For each experiment a negative control was grown in minimal but non-selective media. This control did not express a DNA-binding domain (DBD) able to activate either reporter and could thus be used to establish background levels of GFP and mCherry expression using the exact reporter vectors and bacterial strain as the experimental samples. These levels were used as limits for cell gating and classification of GFP- and/or mCherry-positive cells (see Fig. 2c as an example). Therefore, desired experimental samples were sorted as described above into rich media (SOC) using a BD FACSAria Vantage SE w/DivA instrument (BD Biosciences) at 16 psi with a 70-μm nozzle using sterile PBS as sheath fluid. Cells were characterized using forward and side

Fluorescence-activated cell sorting and recovery. For each experiment a negative control was grown in minimal but non-selective media. This control did not express a DNA-binding domain (DBD) able to activate either reporter and could thus be used to establish background levels of GFP and mCherry expression using the exact reporter vectors and bacterial strain as the experimental samples. These levels were used as limits for cell gating and classification of GFP- and/or mCherry-positive cells (see Fig. 2c as an example). Therefore, desired experimental samples were sorted as described above into rich media (SOC) using a BD FACSAria Vantage SE w/DivA instrument (BD Biosciences) at 16 psi with a 70-μm nozzle using sterile PBS as sheath fluid. Cells were characterized using forward and side
scatter parameters and GFP and mCherry fluorescent proteins were excited via 488 and 568 nm laser lines, respectively. Emitted fluorescence was collected using a 530/30 bandpass filter for GFP and a 600/70 bandpass filter for mCherry. Data were acquired and analysed using FACSdiva software (BD Biosciences). In all, 30,000 events were collected for each sorted population. A volume of the recovered events were plated on rich media to recover 250–500 colonies of bacteria and grown overnight at 37 °C. From 24–48 individual colonies, the zinc-finger coding sequences were amplified by PCR and sequenced for each target selection. Coding sequences were translated enriched amino acids compared for analysis.

ZFN construct for cellular studies. ZFN protein sequences and expression constructs used for cellular studies are provided in Supplementary Note 6. Unless otherwise noted all constructs were generated using standard molecular biology methods.

SELEX studies. An oligonucleotide target library was synthesized bearing the sequence: 5′-CAGGGATCCATGTACGCCCNNNNNNNNNNNNNNNNNNNNGGCCACTGTTACGCCGATCTCGG-3′ where ‘N’ denotes a mixture of all four bases. The library was converted to double-stranded duplex by annealing with 8 nM of library primer 5′-CCAGG ATCCGGACTACGGTTG-3′ in 100 μl 1× PCR Master (Roche) supplemented to 1.2 mM of each dNTP and 5 mM MgSO₄, followed by incubation at 95 °C for 2 min, 94 °C for 5 min, 58 °C for 5 min and 72 °C for 15 min.

For the first assay cycle, ZFNs were expressed directly from plasmid templates using a 1:1 reaction to couple the trans-spacer-target system (Promega) and the manufacturer’s recommended conditions with buffers supplemented to 10 mM ZnCl₂. Expressed ZFNs contained a triple Flag tag fused to their N terminus. A volume of 12 μl of TnT reaction mix was then mixed with 200 pmol of library duplex in a total volume of 100 μl of binding buffer (5 mM dithiothreitol, 10 mM ZnCl₂, 50 mM MgCl₂, 0.01% BSA Fraction V and 100 mM NaCl in PBS (calcium free)). After incubation for 50 min protein–DNA complexes were captured on anti-FLAG M2 magnetic beads (Sigma) and washed five times with wash buffer (5 mM dithiothreitol, 10 μM ZnCl₂, 5 μM MgCl₂, 0.01% BSA Fraction V and 100 mM NaCl in PBS (calcium free)). Bound target was PCR amplified using the 5′ library primer (above) and a 5′ library primer (5′-GAGGATCCATGCGATCAGTACG-3′) and the resulting amplicon was used as input for additional cycles of enrichment. Protein expression and binding conditions for these subsequent cycles were identical to the conditions used in the first round. After three cycles, recovered DNA fragments were sequenced using an Illumina MiSeq system. The protocol for adaptation of the Illumina sequencing primers and sequencing is as described in reference 64.

SELEX FASTQ sequences from the MiSeq were adapter trimmed using SeqPrep (J. St. John, unpublished, https://github.com/jstjohn/SeqPrep). SELEX library sequences were further filtered by custom python scripts for correct length and fixed flanking region composition (exact match). In all, 200 randomly sampled filtered sequences were used as input to the GADMET discovery programme with options maskR = 0 fullscan = 0 gen = 3. Position frequency matrices discovered by GADMET were then aligned to the intended sequence and reverse complemented if necessary. Matrices longer than the intended sequence were trimmed to only those regions overlapping the intended sequence according to the highest-scoring alignment, yielding the final matrices provided in Fig. 3.

Gene modification of endogenous CCR5 and CCR2. To screen ZFN pairs for NHEJ-mediated gene modification, K562 cells were cultured in RPMI 1640 media (Invitrogen) supplemented with 10% (v/v) FBS, 2 mM l-glutamine, 100 U/ml penicillin and 100 mg ml⁻¹ streptomycin. Cells (1×10⁵) were transfected with 400 or 800 ng of each plasmid for HBB-targeted ZFNs. After 1 day cultures were transferred to a six-well plate. On day 14 and day 28 post-transfection, genomic DNA was isolated (Qiagen DNeasy Blood & Tissue Kit) and processed to isolate insert-genome junctions as described48 (steps 1–38), except for the use of an 8-s extension time, and annealing temperatures of 53, 47 and 50 °C for each amplification step. Candidate products were then processed for high throughput sequencing via MiSeq using standard methods.

DNA sequence reads were then processed as follows: first, nonidentical reads were filtered for correct priming and adapter sequences, and the resulting sequences mapped to the genomes reported within the section ‘Gene modification of endogenous CCR5, CCR2’. Ambiguities from candidate off-target loci were then amplified using the PCR primers shown in Supplementary Data 1 (designed using Primer3 with the following optimal conditions: amplicon size of 200 nucleotides, a Tm of 60 °C, primer length of 20 nucleotides and GC content of 50%). Adapters were added for a second PCR reaction to add the Illumina library sequences (ACACGAGCTCTTCCGAATTC forward primer and GACGTTGCTCTTCCGATAT reverse primer), followed by MiSeq sequencing using standard methods. For off-targets where evidence of indels were found, the putative binding sites for each set of zinc fingers are shown in Supplementary Table 1.

In detail, genomic DNA was purified with the Qiagen DNeasy Blood and Tissue Kit. Genomic regions of interest were amplified in 50 μl using 250 ng of genomic DNA with Phusion (NEB) in 200 μM dNTPs. Primers were used at a final concentration of 0.5 μM and the following cycling conditions: initial melt of 98 °C for 30 s, followed by 30 cycles of 98 °C for 10 s, 60 °C for 30 s and 72 °C for 15 s, followed by a final extension 72 °C for 10 min. PCR products were diluted 1:200 in H₂O. A volume of 1 μl diluted PCR product was used in a 10-μl PCR reaction to add the Illumina library sequences with Phusion (NEB) in Buffer GC with 200 μM dNTPs. Primers were used at a final concentration of 0.5 μM and the following cycling conditions: initial melt of 98 °C for 30 s, followed by 12 cycles of 98 °C for 10 s, 60 °C for 30 s, 72 °C for 15 s, followed by a final extension 72 °C for 10 min. PCR products were pooled and purified using the Qiagen Qiaquick PCR Purification Kit (Qiagen). Samples were quantified with the Qubit dsDNA HS Assay Kit (Life Technologies). Samples were diluted to 2 nM and sequenced on an Illumina MiSeq Instrument (Illumina) with a 300 cycle sequencing kit. To compare indels across samples and controls, the statistical test described in ref. 21 was applied to the number of sequences scored as indels and the total number of sequences for both the ZFN-treated sample and the cognate control. P values were then adjusted for multiple comparisons with a Bonferroni correction.

References

1. Kim, H. & Kim, J. S. A guide to genome engineering with programmable nucleases. Nat. Rev. Genet. 15, 321–334 (2014).
2. Bibikova, M., Golic, M., Golic, K. G. & Carroll, D. Targeted chromosomal cleavage and mutagenesis in Drosophila using zinc-finger nucleases. Genetics 161, 1169–1175 (2002).
3. Rouet, P., Smith, F. & Jasmin, M. Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells. Proc. Natl Acad. Sci. USA 91, 6064–6068 (1994).
4. Rouet, P., Smith, F. & Jasmin, M. Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease. Mol. Cell. 14, 135–152 (1999).
5. Uzun, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. Nat. Rev. Genet. 11, 636–646 (2010).
6. Segal, D. J. & Meckler, J. F. Genome engineering at the dawn of the golden age. Annu. Rev. Genomics Hum. Genet. 14, 135–152 (2013).
7. Gaj, T., Gersbach, C. A. & Barsab, 3rd C. F. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol. 31, 397–405 (2013).
Joung, J. K. & Sander, J. D. TALENs: a widely applicable technology for targeted genome editing. Nature reviews. Mol. Cell Biol 14, 49–55 (2013).

Niu, Y. et al. Generation of gene-modified cytomolgous monkey via Cas9/RNA-mediated gene targeting in one-cell embryos. Cell 156, 836–843 (2014).

Mali, P. et al. RNA-guided human genome engineering via Cas9. Science 339, 823–826 (2013).

Gratz, S. J. et al. Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease. Genetics 194, 1029–1035 (2013).

Friedland, A. E. et al. Heritable genome editing in C. elegans via a CRISPR-Cas9 system. Nat. Methods 10, 741–743 (2013).

Miller, J. C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. Nat. Biotechnol. 25, 778–785 (2007).

Bhakta, M. S. et al. Highly active zinc-finger nucleases by extended modular assembly. Genome Res. 23, 530–538 (2013).

Guilinger, J. P. et al. Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleaveage specificity. Nat. Methods 11, 429–435 (2014).

Cong, L., Zhou, R., Kuo, Y. C., Cunniff, M. & Zhang, F. Comprehensive analysis of Cas9 to FokI nuclease improves the specificity of genome modification. Nat. Biotechnol. 33, 876–790 (2015).

Gabriel, R. et al. An unbiased genome-wide analysis of zinc-finger nuclease specificity. Nat. Biotechnol. 29, 816–823 (2011).

Mussolino, C. et al. A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. Nucleic Acids Res. 39, 9283–9293 (2011).

Wang, X. et al. Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors. Nat. Biotechnol. 33, 175–178 (2015).

Frock, R. L. et al. Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. Nat. Biotechnol. 33, 179–186 (2014).

Hsu, P. D. et al. DNA targeting specificity of RNA-guided Cas9 nucleases. Nat. Biotechnol. 32, 789–790 (2014).

Ying, J., Zhou, R., Kuo, Y. C., Cunniff, M. & Zhang, F. The CRISPR-Cas9 system for efficient RNA-directed genome engineering. Cell 157, 1255–1269 (2014).

Cho, S. W. et al. Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. Genome Res. 24, 1020–1027 (2014).

Li, C. et al. Simultaneous gene editing by injection of mRNAs encoding transcription activator-like effectors into mouse embryos. Nat. Biotechnol. 32, 1649–1658 (2014).

Qu, Y. et al. A TALE nuclease architecture for efficient genome editing. Nat. Biotechnol. 29, 143–148 (2011).

Li, L. et al. CRISPR/Cas-mediated genome editing in human cells. Cell 157, 1269–1280 (2014).

Anderson, D. M. et al. CRISPR-Cas9 as a versatile tool for engineering mammalian genome. Cell 163, 889–898 (2015).

Najafabadi, H. S. et al. CHOPCHOP: a CRISPR/Cas9 and TALEN web tool for genome editing. Nucleic Acids Res. 43, W376–81 (2015).
performed the IDLV capture study and characterized ZFN activity in cellular studies at the intended targets and the anti-selected off-target sites; E.F.R. processed zinc-finger pools used in selections, and assembled zinc-finger array libraries for selection; D.J.X. processed selections and cell sorting for HBB zinc-finger selections; I.A. performed SELEX studies; J.S.B. performed cell studies of ZFN activity at sites identified via IDLV capture analysis; L.Z. designed expression vectors and studies of ZFN activity and assisted with data analysis; P.L. performed capture study data analysis; J.C.M. designed SELEX and IDLV capture studies and performed data analysis and contributed to writing the manuscript; E.J.R. designed cellular, SELEX and IDLV capture studies, and contributed to writing the manuscript; M.B.N. designed the multi-reporter selection system, chose zinc-finger targets and assembly strategies, tested controls to validate the system and contributed to writing the manuscript.

Additional information
Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

Competing Financial Interest The following are employees of Sangamo Biosciences: D.F.X., I.A., J.S.B., L.Z., P.L., J.C.M., E.J.R.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Oakes, B. L. et al. Multi-reporter selection for the design of active and more specific zinc-finger nucleases for genome editing. Nat. Commun. 7:10194 doi: 10.1038/ncomms10194 (2016).

This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/