Supplementary Notes

In vivo yeast screening for high-specificity SpCas9 variants

To generate the reporter yeast strains (yACMO-off1/-off4) we modified the *TRP1* (chromosome IV) and *ADE2* (chromosome XV) genomic loci. An EGFP-derived on-target sequence (**Supplementary Table 2**) for Cas9 was introduced in the *TRP1* gene, while similar sequences containing single mismatches (off-targets, **Supplementary Table 2**) were incorporated into the *ADE2* locus. A stop codon was added immediately after each on/off-target site to ensure premature interruption of translation, while a 100bp duplication was positioned on both sides to favour homology-driven scarless repair after cleavage of the target sequence. The knockout of the two genes by reporter cassette insertion produces defects in the tryptophan and adenine metabolic pathways. This metabolic alteration leads to growth arrest in the absence of tryptophan and to the accumulation of a red pigment in low adenine medium leading to the formation of red colonies. Following double strand breaks induced by Cas9, each locus can be efficiently repaired by single strand annealing favoured by the two homology regions, thus restoring prototrophy for the two nutrients (corresponding to cell growth and the generation of white colonies, for *TRP1* and *ADE2* targeting, respectively) (**Fig. 1a**). Therefore, the successful editing event at each of the two loci can be visualized using appropriate reporter plates: no tryptophan and low adenine concentrations (SDluta$_5$ plates). The assay readout consists in a two-step process: first, on-target cleavage efficiency can be measured by dividing the total number of colonies (red + white) obtained on reporter plates without tryptophan and with low adenine (SDluta$_5$ plates) over those recovered on plates selecting for total transformants incorporating Cas9 and sgRNA plasmids (SDlu plates); second, on- vs. off-target activity can be evaluated by counting the number of red (*TRP1*$^+$*/ade2*$^-$) and white (*TRP1*$^+$*/ADE2*$^+$, corresponding to off-target cleavages) colonies in the same reporter plates.

To perform the screening, the yACMO-off4 reporter strain was stably transformed with a plasmid coding for a sgRNA perfectly matching the on-target site (*TRP1* locus). The off-target sequence (off4) located in the *ADE2* locus is characterized by a single mismatch with the sgRNA spacer in position 15 from the PAM. The library of mutated SpCas9 molecules was generated by co-transformation of PCR-mutated fragments deriving from the open reading frame of the REC3 domain (**Supplementary Figure 1a**) and a plasmid encoding a galactose-inducible REC3-deleted SpCas9. The mutagenized REC3 fragments and the REC3-deleted SpCas9 plasmid reconstitute *in vivo* the full-length SpCas9 open reading frame by homologous recombination. The transformed colonies containing the new nuclease variants were recovered after an overnight incubation in non-selective medium followed by SpCas9 induction and plating on several reporter plates. We employed a short induction time (5 hours) to select against variants with poor on-target activity which might emerge with longer induction periods (>5 hours). Red colonies were streaked on reporter plates containing galactose to maintain SpCas9 constantly expressed and exacerbate any off-target cleavage. After 48 hours, the DNA coding for the selected SpCas9 variants was recovered from the most red-pigmented streaks and subsequently used for re-challenging in yACMO-off4, in order to eliminate false positive clones containing non-sense mutations or frameshifts (see schematics in **Fig. 1c**). In addition, this allowed to measure more precisely the cleavage activity of each variant, discard those less catalytically active compared to wild-type SpCas9 and rank the remaining ones according to on-target cleavage efficiency and their ability to discriminate off-targets (**Fig. 1d**).

Effect of extra 5' guanines in spacers on evoCas9 and evo-dCas9-VP64 activity

It has been reported that high-fidelity SpCas9 variants are incompatible with the addition of a mismatched guanine at the 5'-end of guide RNA spacers to favor transcription from the Pol III[1,2]. We thus evaluated evoCas9 activity in these experimental conditions by using

sgRNAs containing 5' mismatched guanines as well as guides characterized by longer spacers (**Supplementary Fig. 4b-d**). We observed complete abrogation of cleavage activity in all the experimental conditions where the spacers did not correspond to a canonical 20 bp matching sequence, including those where fully matching spacers longer than 20 nucleotides were employed. Interestingly, the same drop in on-target activity was measured also for a spacer containing 19 matching nucleotides and a mismatched G in the 20th PAM-distal position. This feature may reduce the total number of targetable sites, since spacer sequences natively starting with a guanine can exclusively be used. However, this limitation can be circumvented by using other methods for guide RNAs synthesis (e.g. chemical synthesis and electroporation or expression through tRNA-flanked sgRNA[3,4]), or by designing sgRNAs starting with cytidine or adenine instead of guanine, which are compatible with U6-driven transcription systems, as demonstrated by our (**Supplementary Fig. 4e**) and others' results[1].

In light of the current model where mismatches located in more PAM-distal positions are more easily tolerated by SpCas9, this result is unexpected since a mismatch in position 18 of the guide cannot be completely discriminated by evoCas9 (**Fig. 2b**). This, together with the complete loss of activity observed with spacer sequences longer than 20 bp and containing or not mismatched 5'-Gs, indicates that evoCas9 does not tolerate alterations of the RNA:DNA heteroduplex in positions that are structurally located at the end of the cavity that harbours the duplex itself. In addition, it has been consistently reported that the inclusion of two extra guanines at the 5'-terminus of sgRNAs increases targeting specificity[5,6]. This effect, still lacking a functional explanation, is possibly justified by the behaviour we observed with our mutant.

Additional data on the effect of 5' mismatched guanines were obtained using catalytically inactive SpCas9-based transcriptional activators (dCas9-VP64[7], **Supplementary Fig. 11**). This experimental system can be considered a good indicator of the strength of SpCas9

binding to the target DNA, as transcriptional activation will be proportional to the time spent by the activator on the target promoter. Interestingly, we did not observe any important difference in EGFP transcriptional activation when comparing samples transfected with evo-dCas9-VP64 together with the on-target TetO-on guide RNA or the same on-target guide with an added extra 5'-G (**Supplementary Fig. 11**). This result, together with previous data showing a consistent loss of editing activity when using sgRNA with an additional initial G (**Supplementary Fig. 4b-e**), suggests that evoCas9 is indeed able to bind target sites using sgRNAs characterized by a mismatched extra 5'-guanine, but is then unable to cut the bound DNA. Accordingly, when we compared the EGFP fold-activation obtained using mismatched guides, the increased specificity observed using evo-dCas9-VP64 was modest when compared with the original dCas9-VP64, further reinforcing the idea that evoCas9 binds to mismatched targets, even though less efficiently, but is then unable to complete the cleavage reaction (**Supplementary Fig. 11b-c**). Finally, the lower background activation observed in the presence of our high-specificity variant might be due to a lower propensity of evo-Cas9 to bind stably DNA (**Supplementary Fig. 11b-c**)**.**

Supplementary References

1. Kleinstiver, B. P. *et al.* High-fidelity CRISPR-Cas9 nucleases with no detectable genome-wide off-target effects. *Nature* **529,** 490–495 (2016).
2. Slaymaker, I. M. *et al.* Rationally engineered Cas9 nucleases with improved specificity. *Science* **351,** 84–88 (2016).
3. Xie, K., Minkenberg, B. & Yang, Y. Boosting CRISPR/Cas9 multiplex editing capability with the endogenous tRNA-processing system. *Proc. Natl. Acad. Sci. U.S.A.* **112,** 3570–3575 (2015).
4. Mefferd, A. L., Kornepati, A. V. R., Bogerd, H. P., Kennedy, E. M. & Cullen, B. R.

Expression of CRISPR/Cas single guide RNAs using small tRNA promoters. *RNA* **21,** 1683–1689 (2015).

5. Cho, S. W. *et al.* Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases. *Genome Research* **24,** 132–141 (2014).
6. Kim, D., Kim, S., Kim, S., Park, J. & Kim, J.-S. Genome-wide target specificities of CRISPR-Cas9 nucleases revealed by multiplex Digenome-seq. *Genome Research* **26,** 406–415 (2016).
7. Perez-Pinera, P. *et al.* RNA-guided gene activation by CRISPR-Cas9–based transcription factors. *Nat. Methods* **10,** 973–976 (2013).
8. Chen, B. *et al.* Dynamic Imaging of Genomic Loci in Living Human Cells by an Optimized CRISPR/Cas System. *Cell* **155,** 1479–1491 (2013).
9. Tsai, S. Q. *et al.* GUIDE-seq enables genome-wide profiling of off-target cleavage by CRISPR-Cas nucleases. *Nat. Biotechnol.* **33,** 187–197 (2014).
10. Chen, J. S. *et al.* Enhanced proofreading governs CRISPR-Cas9 targeting accuracy. *Nature* (2017). doi:10.1038/nature24268