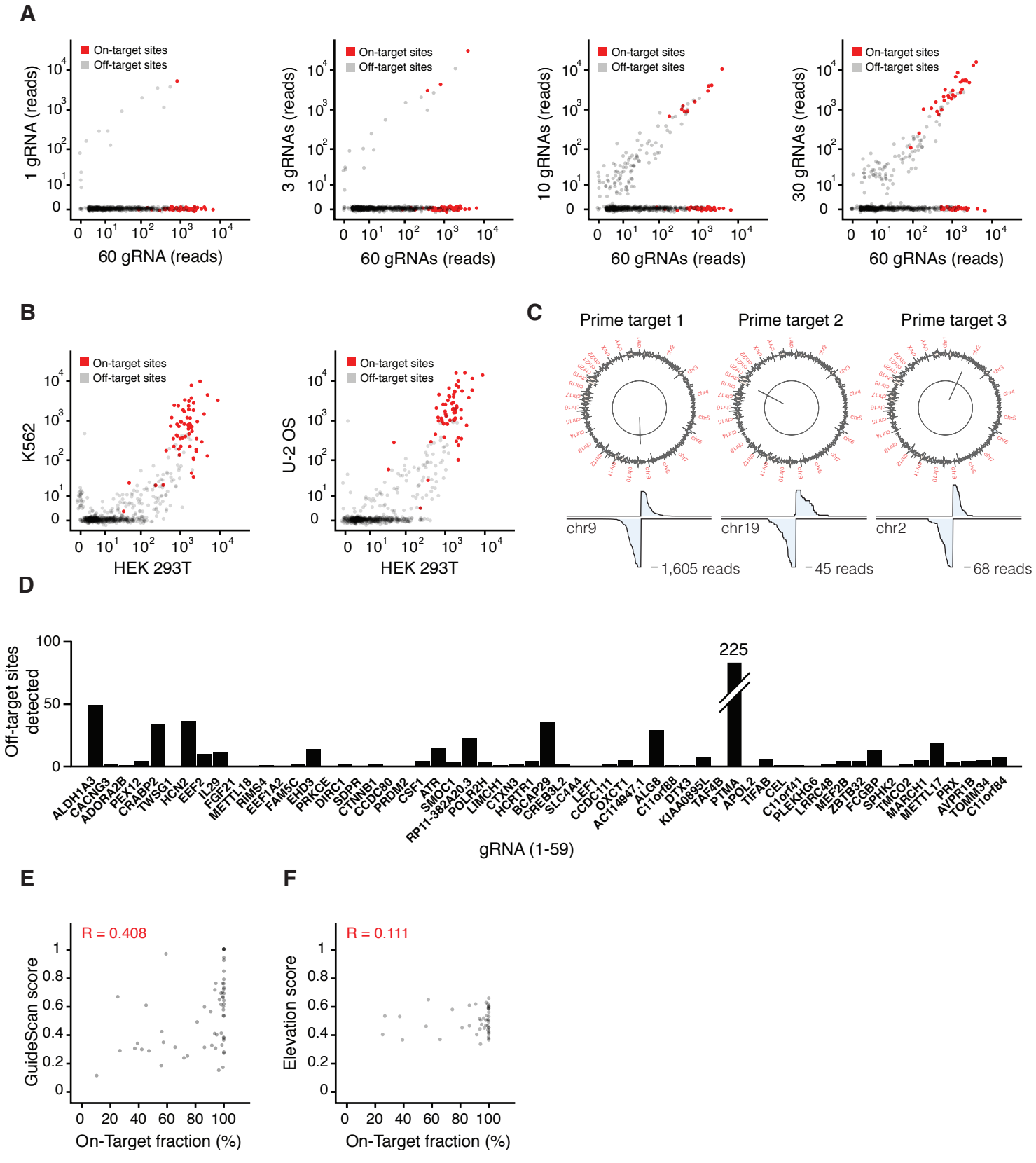
Supplementary Information

Highly parallel profiling of Cas9 variant specificity

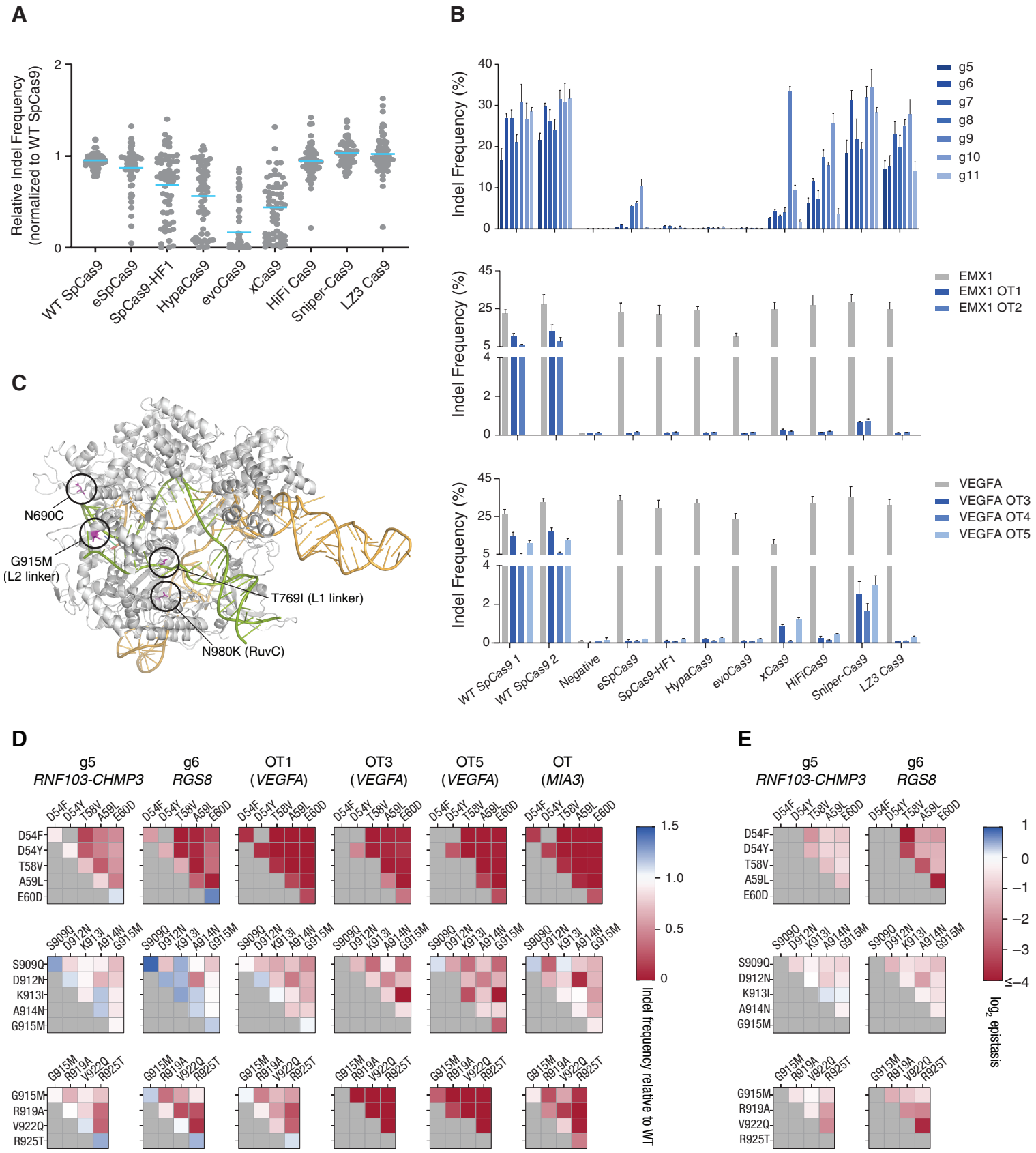
Jonathan L. Schmid-Burgk[1,2,3,4,5], Linyi Gao[1,2,3,4,5], David Li[1,2,3,4,5], Zachary Gardner[1,2,3,4], Jonathan Strecker[1,2,3,4], Blake Lash[1,2,3,4], Feng Zhang[1,2,3,4,5,6,7,*]

[1] Broad Institute of MIT and Harvard
Cambridge, MA 02142, USA
[2] McGovern Institute for Brain Research
[3] Department of Brain and Cognitive Sciences
[4] Department of Biological Engineering
Massachusetts Institute of Technology, Cambridge, MA 02139, USA
[5] These authors contributed equally
[6] Howard Hughes Medical Institute, Cambridge, MA 02139, USA
[7] Lead Contact

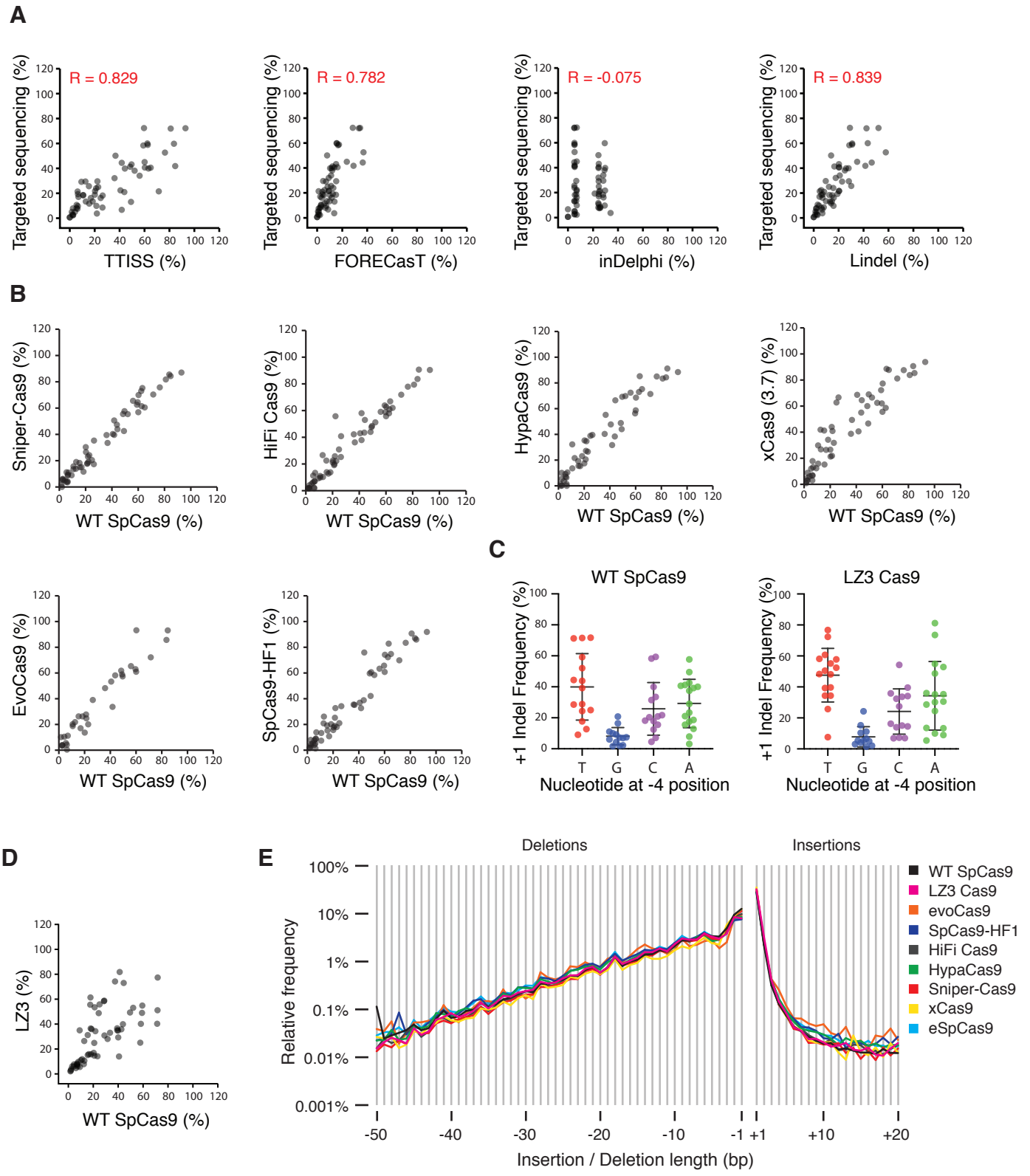
*Correspondence should be addressed to F.Z. (zhang@broadinstitute.org)



Supplemental Figure 1 | Extended validation and application of TTISS, Related to Figure 1 (A) TTISS results for multiplexing of 1, 3, 10, 30, and 60 gRNAs. The number of reads for each detected genomic locus is plotted. On-target sites are marked in red. **(B)** Quantitative TTISS results from three cell lines using 59 guides. **(C)** Detection of donor integration sites using prime editing targeting three genomic loci in HEK 293T cells. Spacer and extension sequences are provided in Supplemental Table 3. **(D)** Distribution of off-target sites per gRNA across 59 gRNAs detected by TTISS using WT SpCas9. **(E)** Comparison of GuideScan-predicted specificity scores to TTISS measured on-target fractions for 59 guides. **(F)** Comparison of Elevation specificity scores to TTISS measured on-target fractions for 47 guides which could be scored by the CRISPR ML online interface.



Supplemental Figure 2 | On-target and off-target activity of selected SpCas9 variants, Related to Figures 1 and 2. All indel frequencies were quantified by targeted deep sequencing. **(A)** Normalized indel frequencies for 59 target sites for WT, LZ3 Cas9, and seven previously reported SpCas9 specificity-enhancing variants. Each dot represents a different guide (mean of n = 2 replicates). The teal lines show the median activity for each Cas9 variant. Target sites were selected from the GeCKO library (Shalem et al. Science 2014), each targeting a different gene, without prior knowledge of activity. **(B)** Activity of SpCas9 variants at additional on-target and off-target sites. Guides g5-g11 were selected based on prior knowledge of low activity for eSpCas9(1.1) and SpCas9-HF1. **(C)** Crystal structure of SpCas9 (PDB ID: 5F9R) showing the position of the four mutations in LZ3. **(D)** Activity of double mutants of selected specificity-enhancing single mutants. **(E)** Epistasis plots of the variants in (D) for guides g1 and g2, where epistasis was calculated as fAB/(fA x fB), where fAB is the normalized indel frequency of the double mutant, and fA and fB are the normalized indel frequencies of the corresponding single mutants.



Supplemental Figure 3 | Extended assessment of +1 indel frequencies using TTISS, Related to Figure 3. (A) +1 insertion frequencies measured by TTISS or predicted by FORECasT, inDelphi, or Lindel are correlated to +1 frequencies measured by targeted indel sequencing for WT SpCas9 across 58 gRNAs. **(B)** TTISS-predicted +1 frequencies for SpCas9 variants calculated for 58 gRNAs plotted against TTISS-predicted +1 frequencies for WT SpCas9. **(C)** +1 indel frequencies measured by targeted sequencing for WT SpCas9 and LZ3 Cas9 across 59 guides, grouped by the nucleotide identity at the -4 position relative to the PAM. **(D)** Plot of +1 frequencies for LZ3 against +1 frequencies for WT SpCas9 as measured by targeted sequencing for 59 gRNAs. **(E)** Insertion and deletion length distributions of Cas9 variants across 59 guides from targeted sequencing. Indel length frequencies relative to total indels are shown on logarithmic scale.