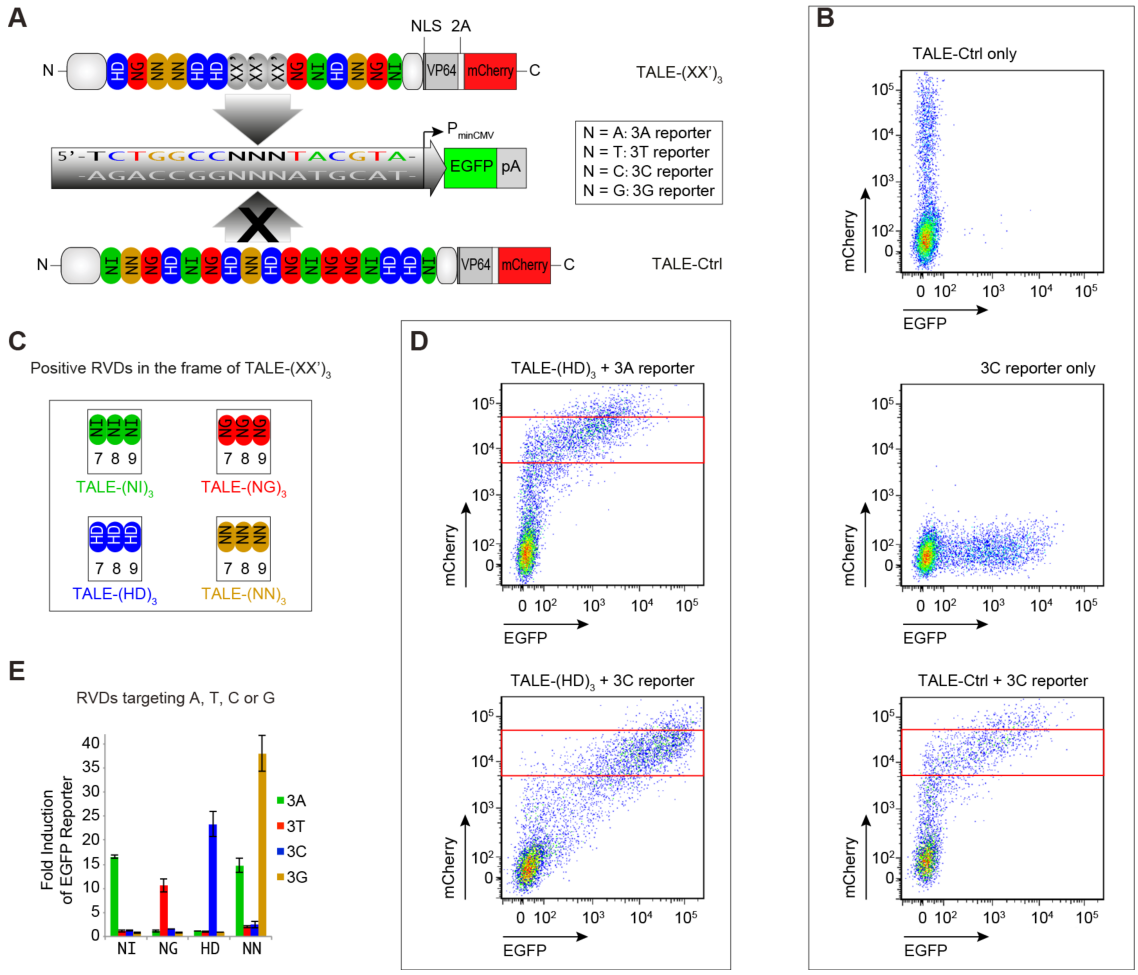

Supplementary information, Figure S1 Screening system for the assessment of TALE RVD efficiencies and specificities. (**A**) Design of the screening system for novel TALE RVDs. The customized TALEs used for RVD screening contained 14.5 repeats fused with the VP64 trans-activation domain and 2A peptide-linked mCherry. The variable diresidues (XX') for testing were placed in the 7th - 9th repeat modules, and the customized TALE was designated as TALE-(XX')$_3$. X and X' represents the 12th and 13th amino acids in the 7th - 9th repeat modules, respectively. To determine the DNA recognition specificity of variable RVDs, four reporters were constructed, consisting of TALE-(XX')$_3$ binding sites with three consecutive nucleotides (A, T, C or G) substituted at positions 7 - 9 in front of a minimal CMV promoter (P_{minCMV}) and its downstream EGFP gene. Construct encoding TALE-Ctrl has the identical backbone as TALE-(XX')$_3$ except that its TALE repeat region is different

as indicated, which does not match with any reporters. (**B**) FACS analysis of HEK293T cells transfected with TALE-Ctrl only, 3C reporter only and TALE-Ctrl plus 3C reporter (from top to bottom). Red box indicates the region of data collection. (**C**) Customized TALE-$(XX')_3$ for testing the DNA binding activity of commonly used RVDs (NI, NG, HD and NN). (**D**) Representative FACS analysis of HEK293T cells co-transfected with TALE-$(HD)_3$ and the 3A (top) or 3C reporter (bottom). Red boxes indicate the region of data collection. (**E**) Base binding activity of the common RVDs in (**C**). The horizontal axis labels indicate the variable RVD (XX') for testing in TALE-$(XX')_3$. The color-coded bars represent the fold induction levels of the different reporters (A, green; T, red; C, blue; and G, yellow) in this and the subsequent figures. Data are means ± s.d., n = 3.