SUPPLEMENTAL MATERIAL

Zhao et al., http://www.jem.org/cgi/content/full/jem.20160670/DC1

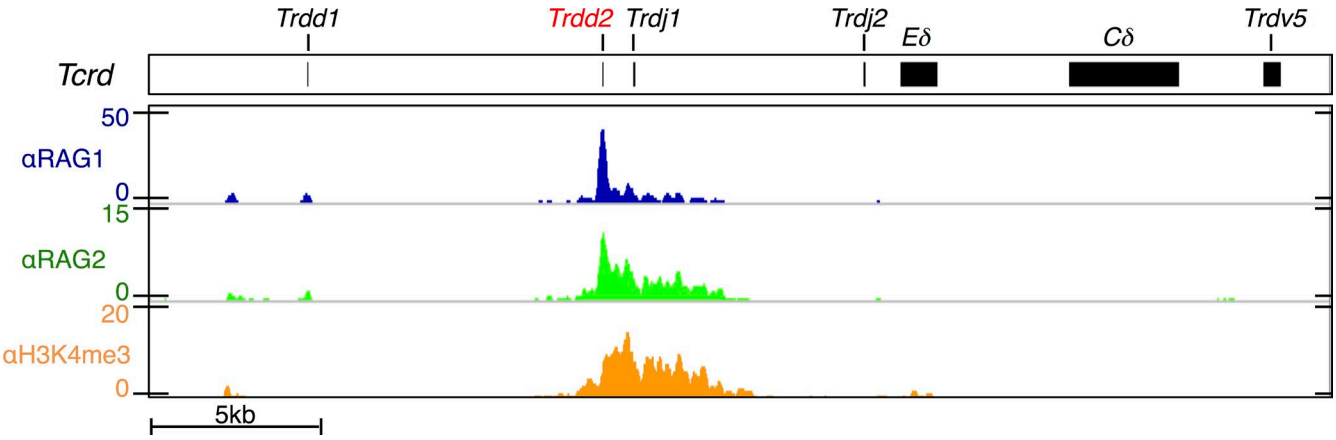

Figure S1. **High-level enrichment of RAG binding and H3K4me3 at *Trdd2*.** IGV plot displaying RAG1, RAG2, and H3K4me3 ChIP-seq profile from Dβ $CD4^+CD8^+$ pre–T cell thymocytes. The Dβ mouse is deficient for endogenous RAG1, but harboring a catalytically dead RAG1 (D708A) mutant that retains its DNA-binding properties, and also expresses a Tcrβ transgene.

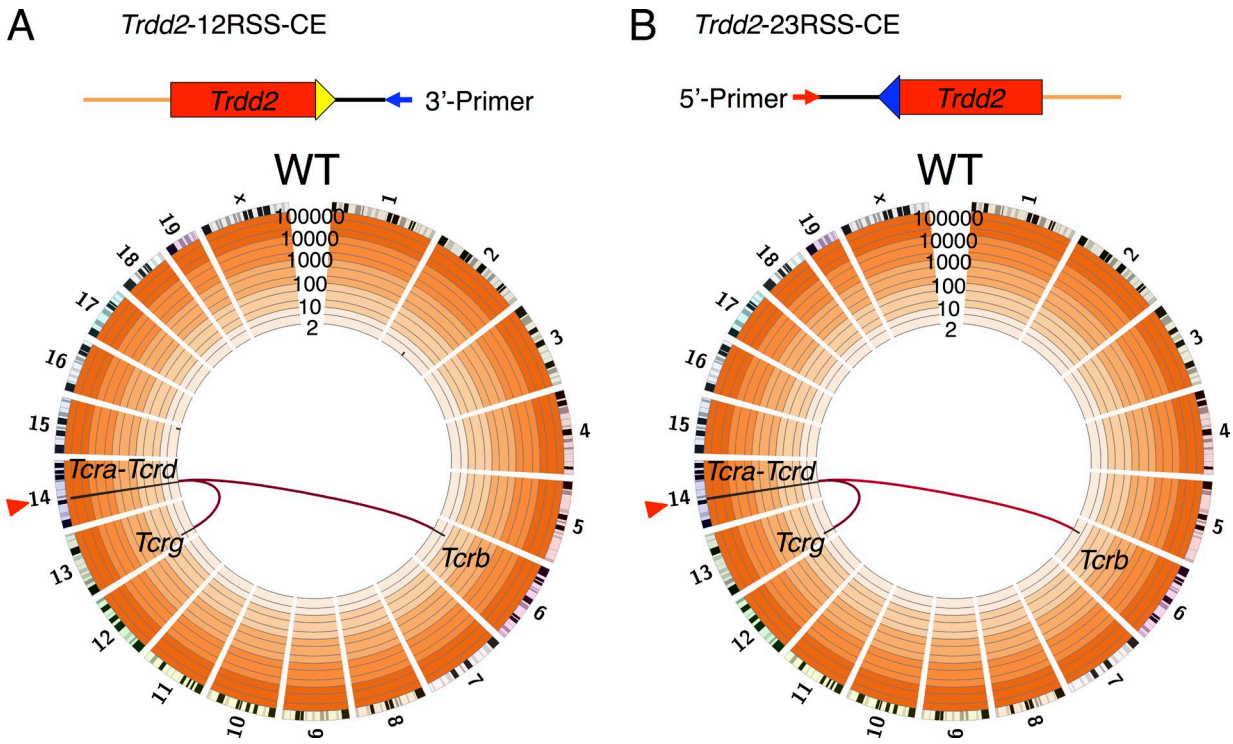

Figure S2. **Genome-wide distribution of prey junctions from *Trdd2* 12RSS and 23RSS CE libraries.** (A and B, top) Schematic of junction reads from 3′-Primer (A) or 5′-Primer (B). Bait primer (red or blue arrows), bait sequence (black lines), prey sequence (orange lines) are indicated. The relative position of 12RSS (blue triangle), 23RSS (yellow triangle) and *Trdd2* gene segment (red rectangle) are labeled. (bottom) Circos plots showing pooled WT *Trdd2*-12RSS-CE libraries ($n = 3$; A) and WT *Trdd2*-23RSS-CE libraries ($n = 4$; B). Red triangles indicate the *Trdd2* bait-site. Red lines link the *Trdd2* bait to *Tcr* loci hotspots. Bin size is 5 Mb (black bars). *Trdd2*-12RSS-CE and 23RSS-CE libraries were normalized to 76,966 and 106,193, respectively.

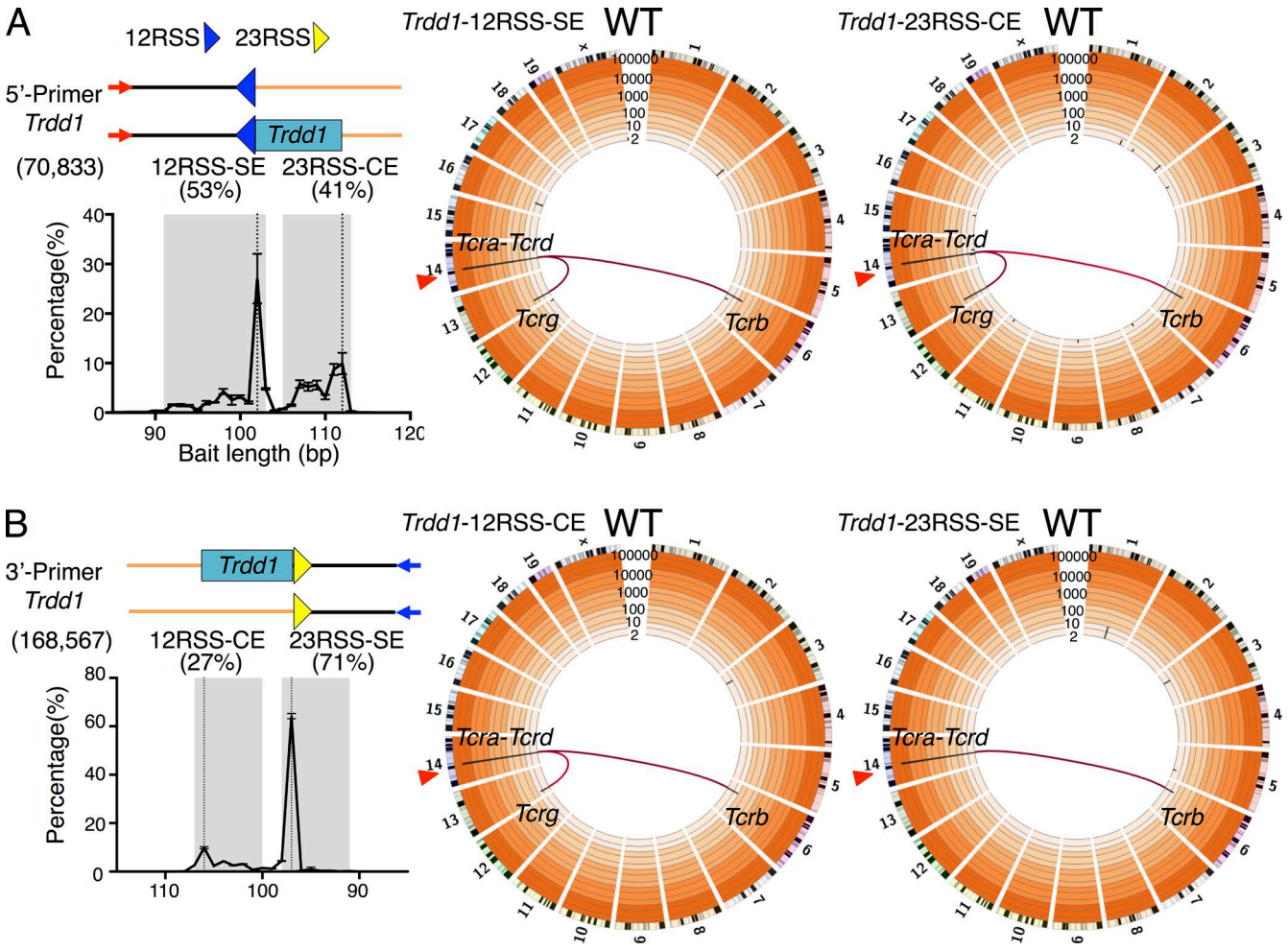

Figure S3. **Bait length and genome-wide prey junction distributions baited from RAG-initiated DSBs *Trdd1*.** (A and B, top left) Schematic of junction reads from 5′-Primer (A) or 3′-Primer (B). Bait primer (red or blue arrows), bait sequence (black lines), and prey sequence (orange lines) are indicated. The relative position of 12RSS (blue triangle), 23RSS (yellow triangle), and *Trdd1* gene segment (blue rectangle) are marked. (bottom left) Length distribution of bait sequence, plotted as the percentage of total junctions, corresponds to the relative position of the predicted bait break-sites (dotted lines). Total junctions analyzed are shown in parentheses. Ranges of bait sequence lengths isolated for SE and CE bait libraries are marked in gray (see Materials and methods for details). Percentages of total junctions in indicated range is shown above. (middle and right) Circos plots displaying genome-wide prey junction distribution from *Trdd1*-12RSS-SE (B; $n = 3$; A, middle), *Trdd1*-23RSS-CE (B; $n = 3$; A, right), *Trdd1*-12RSS-CE (C; $n = 3$; B, middle), *Trdd1*-23RSS-SE (C; $n = 3$; B, right) libraries. Bin size is 5 Mb (black bars); red lines linking the *Trdd1* bait (red triangle) to *Tcr* loci hotspots. *Trdd1*-12RSS-SE and *Trdd1*-12RSS-CE libraries are normalized to 70,833. *Trdd1*-23RSS-SE and *Trdd1*-23RSS-CE libraries are normalized to 168,567.

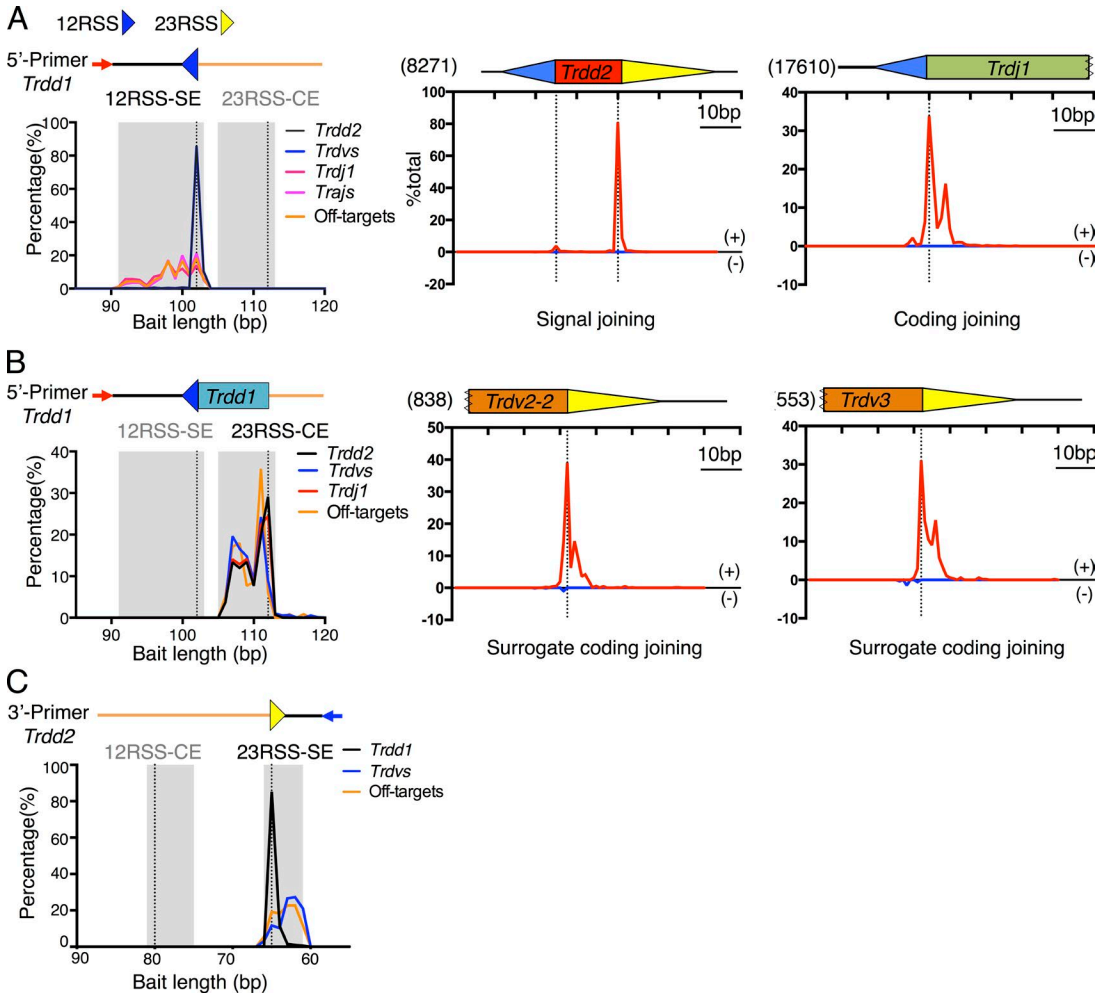

Figure S4. **Precise signal join and processed surrogate coding join examples.** (A–C, top left) Schematic of junction reads baited from *Trdd1*-12RSS-SE (A), *Trdd1*-23RSS-CE (B), or *Trdd2*-23RSS-SE (C). Bait primer (red or blue arrows), bait sequence (black lines), and corresponding prey sequence (orange lines) are indicated. The relative position of 12RSS (blue triangle), 23RSS (yellow triangle), and *Trdd1* gene segment (blue rectangle) are denoted. (bottom left) Length distribution of bait sequence with corresponding preys clustered within selected hotspot regions recovered from *Trdd1*-12RSS-SE (A), *Trdd1*-23RSS-CE (B), or *Trdd2*-23RSS-SE (C) libraries. Bait sequence length reflects processing of several nucleotides from bait BEs before end-joining. (A, left) Distribution of bait sequence length reveal precise signal joining patterns for prey ends at *Trdd2* and *Trdvs*. Prey ends at *Trdj1*, *Traj* cluster and off-targets downstreamof *Trdd1*, are lead by bait sequences with end-processing reminiscent of coding joining patterns. Linear distribution of prey junctions at *Trdd2* (middle) or *Trdj1* (right) recovered from *Trdd1*-12RSS-SE. (B, left) End-processing at bait sequences with corresponding preys at selected regions from *Trdd1*-23RSS-CE libraries. Distribution of prey junctions in *Trdv2-2* (middle) and *Trdv3* (right) from *Trdd1*-23RSS-CE bait libraries reflects end-processing from prey BE. (C) End-processing at bait sequences distinguishes corresponding preys at *Trdvs* and RAG off-targets resulting from secondary recombination events that occurred at chromosomally positioned *Trdd1*-12RSS/*Trdd2*-23RSS fusions (see Fig. 3 A for additional information). Red and blue lines in A and B (middle and right) represent + and — joining orientation, respectively.

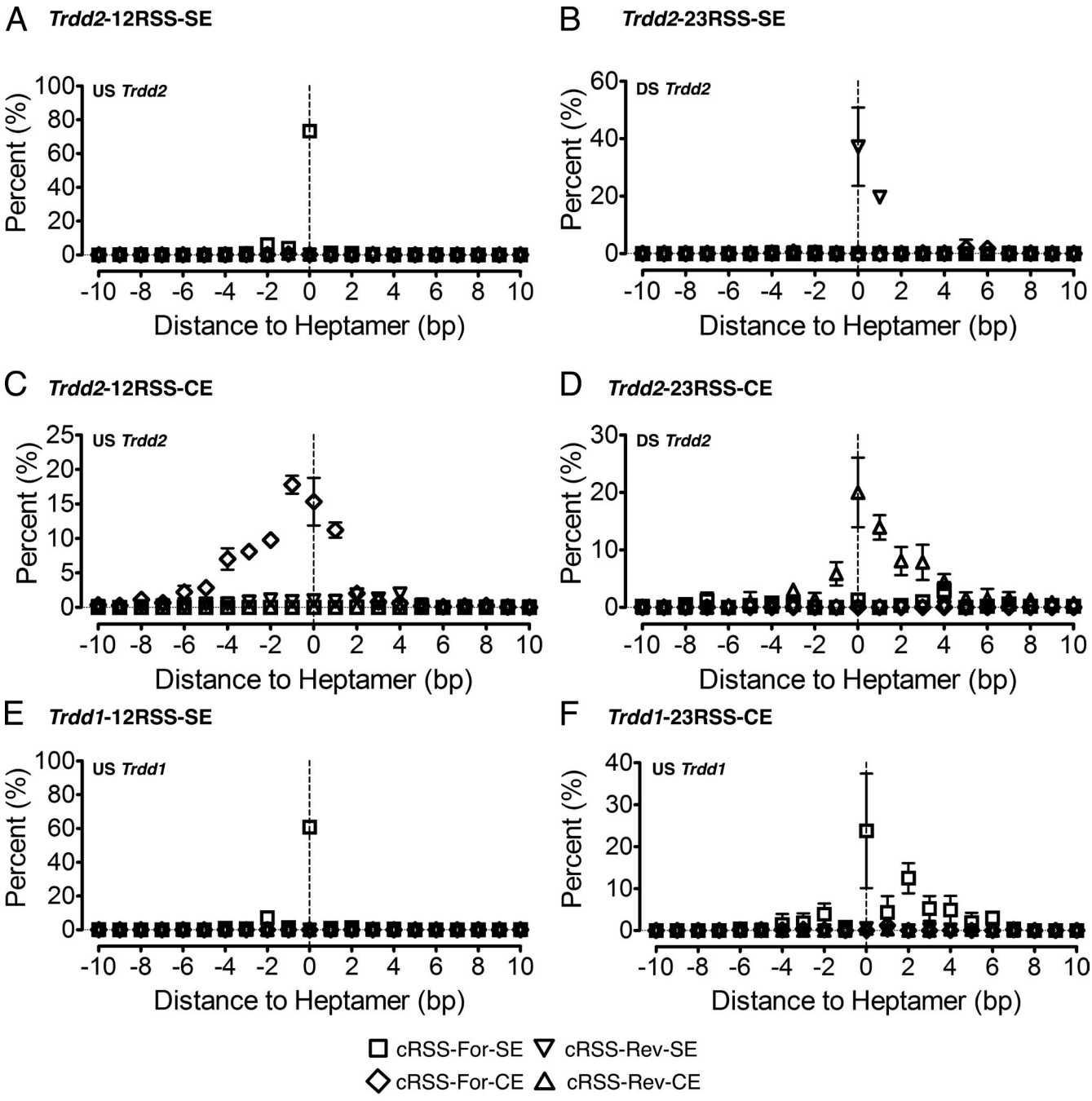

Figure S5. **Junctions not corresponding to bona fide RSS sites are associated with RAG off-targets.** (A–F) Distance of the junctions not corresponding to bona fide RSS sites to cRSS (CAC) in selected regions; junction percentages are from total junctions that are not corresponding to bona fide RSS sites. *Trdd2*-12RSS-SE (A; $n = 4$); *Trdd2*-23RSS-SE (B; $n = 3$); *Trdd2*-12RSS-CE (C; $n = 3$); *Trdd2*-23RSS-CE (D; $n = 4$); *Trdd1*-12RSS-SE (E; $n = 3$); and *Trdd1*-23RSS-CE (F; $n = 3$). US *Trdd2*, upstream region of *Trdd2* in CIL; DS *Trdd2*, downstream region of *Trdd2* in CIL; US *Trdd1*, upstream region of *Trdd1* in CIL; DS *Trdd1*, downstream region of *Trdd1* in CIL. cRSS-For-SE, + orientation joining at forward cRSS; cRSS-For-CE, − orientation joining at forward cRSS; cRSS-Rev-SE, − orientation joining at reverse cRSS; cRSS-Rev-CE, + orientation joining at reverse cRSS. Mean ± SD from at least three independent libraries.

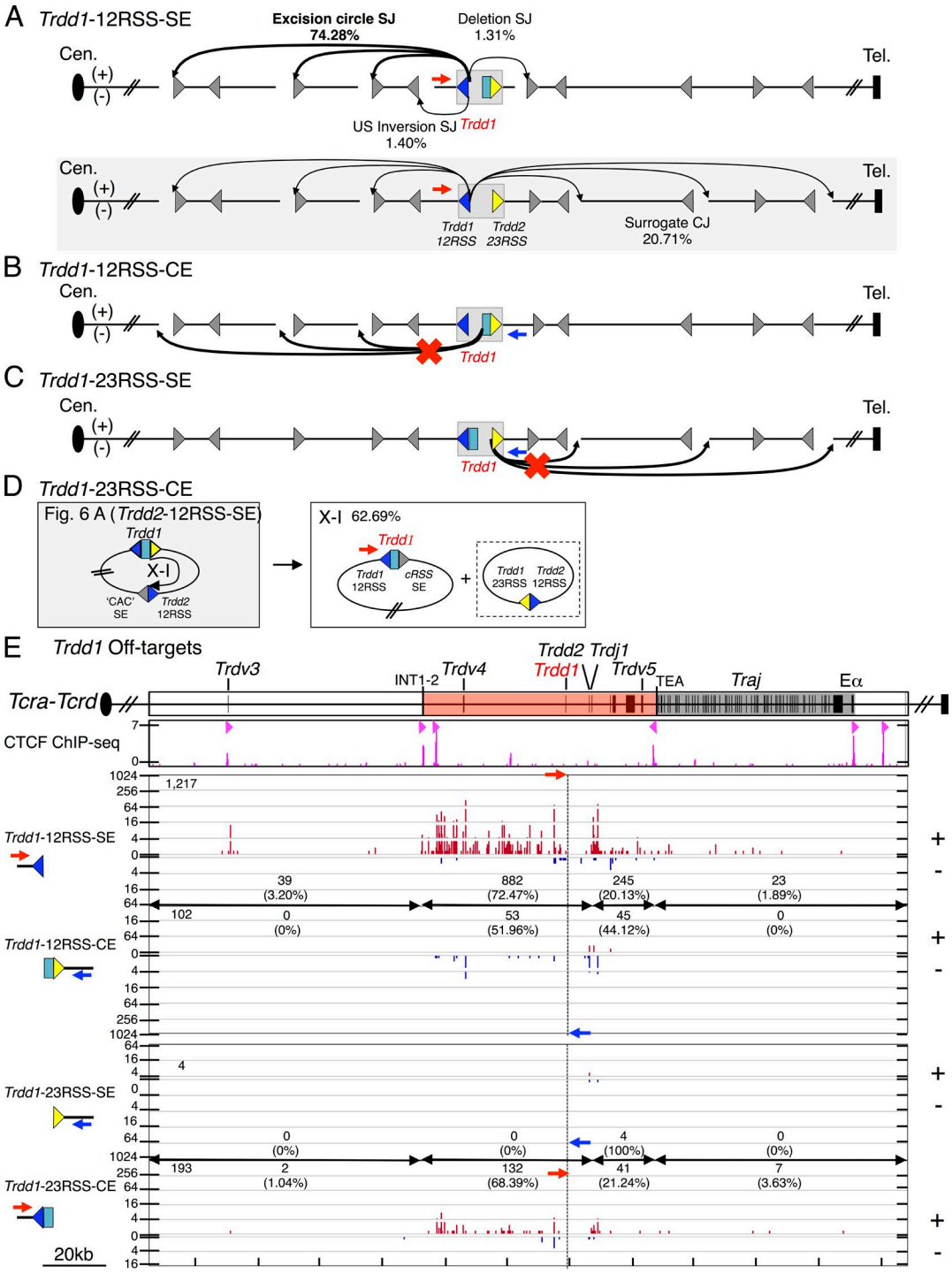

Figure S6. **Profiles of off-targets baited from RAG-initiated DSBs at *Trdd1*.** (A–D) Diagram illustrating joining outcomes to cRSS (CAC) sites (gray triangle) from *Trdd1*-12RSS-SE (A; *n* = 3), -CE (B; *n* = 3), *Trdd1*-23RSS-SE (C; *n* = 3), and *Trdd1*-23RSS-CE (D; *n* = 3) libraries. Joining outcomes baited from *Trdd1*-12RSS-SE in the fused *Trdd1*-12RSS/*Trdd2*-23RSS configuration to cRSSs is shaded in gray (A). Few RAG off-targets are detected by *Trdd1*-12RSS-CE (B) or *Trdd1*-23RSS-SE (C). RAG off-targets detected by *Trdd1*-23RSS-CE are secondary joining events that occurred in excision circle harboring *Trdd2*-12RSS-SE and cRSS signal joins (gray box; D). Junction percentages of total RAG off-targets in excision circle, deletion, upstream (US) inversion, and downstream (DS) inversion joining events are shown. 5′- (red arrow) and 3′-Primers (blue arrow) are indicated. Dotted line indicates the position of bait BEs. (E) Gene segment organization of the region between *Trdv3* and Eα (3′ portion of the *Tcra-Tcrd* locus). CIL (red) and *Traj* cluster (dark gray) is marked. (middle) ChIP-seq profile of CTCF. CBE orientation is marked by light red triangles. (bottom) IGV plots of *Trdd1* RSS bait RAG off-target junction distributions. Total RAG off-targets are listed in the top left corner for each bait. Junction numbers and percentages as of total RAG off-targets in indicated regions are shown.