Supporting Information

Wei et al. 10.1073/pnas.1500788112

```
                                                                  FW1          ─
           ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·
     1   GAGGTGCAGCTGTTGGAGTCTGGGGGAGGCTTGGTACAGCCTGGGGGGTCCCTGAGACTCTCCTGTGC   68
                              CDR1                                    FW2
           ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ─────────  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·
    69   AGCCTCTGGATTCACCTTTAGCAGCTATGCCATGAGCTGGGTCCGCCAGGCTCCAGGGAAGGGGCTGGA  137
                         ─────────  CDR2  ─────────
           ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·
   138   GTGGGTCTCAGCTATTAGTGGTAGTGGTGGTAGCACATACTACGCAGACTCCGTGAAGGGCCGGTTCAC  206
                                            FW3
           ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ─────────────  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·
   207   CATCTCCAGAGACAATTCCAAGAACACGCTGTATCTGCAAATGAACAGCCTGAGAGCCGAGGACACGGC  275
           ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·  ·
   276   CGTATATTACTGT  288
```

Fig. S1. Germ line *IGHV3-23*01* DNA sequence. FWs and CDRs are marked on top of each row of the DNA sequences. Bolded motifs are overlapping hotspots (WGCW) AGCT sites. On each side, the nucleotide positions are labeled.
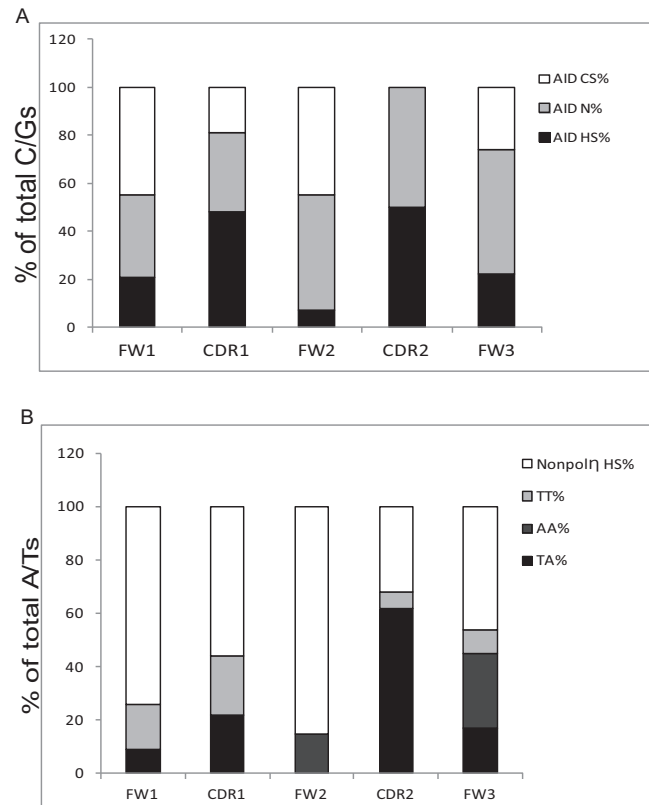

Fig. S2. AID and Polη DNA motifs in *IGHV3-23*01*. (*A*) Percentage of AID hot/cold/neutral spots. Each bar shows the total Cs on both DNA strands in each FW or CDR, as labeled on the bottom. The black part in the bar demonstrates the percentage of AID hotspots (AID HS%) in the total Cs of each FW or CDR. The dark gray part shows the percentage of AID neutral spots (AID N%), and the white part shows the percentage of AID cold spots (AID CS%). (*B*) Polη hotspot percentage. Each bar represents the total As and Ts in each FW or CDR. The black part in the bar tells the percentage of ATs in TA motifs of the total ATs (TA%) in each FW or CDR, as labeled on the bottom. The dark gray part shows the percentage of A/Ts in AA motifs (AA%), and the light gray part shows the percentage of A/Ts in TT motifs (TT%). The white part is the percentage of A/Ts in non-Polη hotspots (nonPolη HS%). Fisher test is done for analyzing the significances of the comparisons.

Fig. S3. Structural view of CDR1 and CDR2 in 3-23 heavy chain (PDB ID code 3QOS). CDR1 and CDR2 are in cyan, and the rest are in gray. Pymol was used to generate the graphs. Gln3 and Leu4 are encoded by CAG and CTG, which include AGCT at residues 9 and 10. Ser31 is encoded by AGC in motif AGCT of OHS1. Ser35 is encoded by AGC in motif AGCT at 104/105 in CDR1. Ala50 is encoded by GCT in motif AGCT of OHS2. Tyr59 is encoded by TAC, which is a WRC AID hotspot in the end of CDR2.