Supporting Information

Gopalakrishnan et al. 10.1073/pnas.1304048110

SI Materials and Methods

CD4-FITC (561835), CD8-FITC (553031), CD4-biotin (553044), CD8a-biotin (553028), CD44-PE (553134), and CD25-APC (557192) antibodies were purchased from BD Biosciences and used for cell staining and sorting. H3K4me2 (07-030, Millipore), H3K27ac (ab4729), and P300 (C-20) (SC-585×) for H3ac (06-599; Millipore), H3K4me3 (39159; Active Motif), and CTCF (07-729; Millipore) antibodies were purchased and used for ChIP experiments.
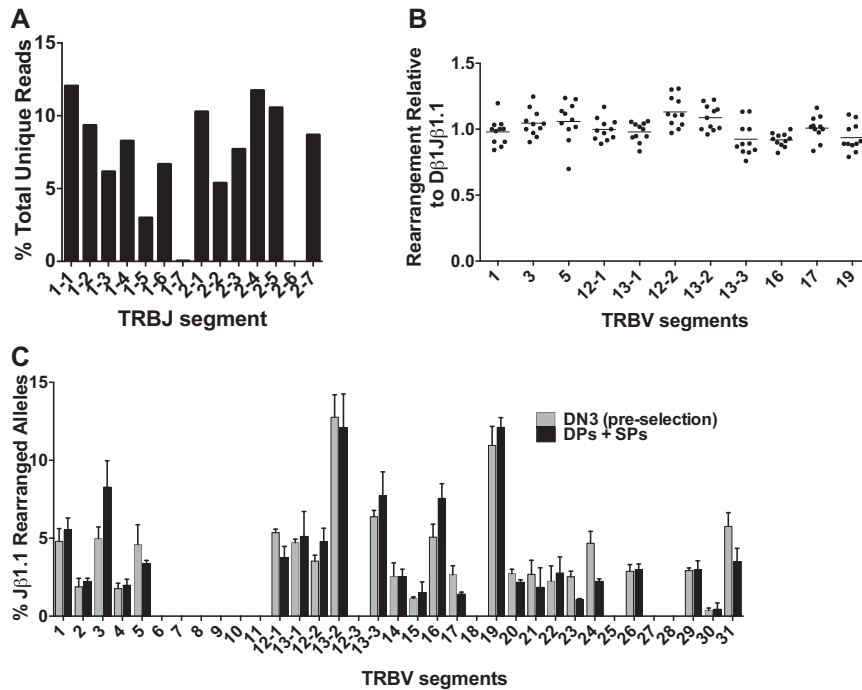

Fig. S1. Variable (V)β repertoire comparisons. (*A*) Joining (J)β use profile from high-throughput sequencing [mean (*n* = 3),15,000–20,000 unique reads per sample]. (*B*) Distribution of rearrangements from high-throughput sequencing involving Vβ segments and each of the 11 functional Jβ segments. Shown are distributions for rearrangements of Vβ segments yielding at least 1,000 unique reads. Data are represented relative to the distribution of Vβ-Jβ1.1, where percent total Vβ-Jβ1.1 is set to a value of 1 (Fig. 1). Each circle represents a data point for a given Jβ segment. (*C*) Comparison of Vβ use in preselection and postselection thymocytes measured by the genomic DNA (gDNA) assay described in Fig. 1C (mean ± SEM, *n* = 3).

Fig. S2. Role of spatial proximity in shaping the *Tcrb* repertoire. (*A*) Chromosome conformation capture (3C) analysis of Rag-deficient double negative (DN) thymocytes showing relative cross-linking frequencies between a diversity (D)β2 anchor and HindIII fragments spanning Vβ gene segments. Data are presented as mean ± SEM (*n* = 3). (*B*) 3C analysis using an Eβ anchor.

Fig. S3. Luciferase assays. Promoter activity assay for upstream regions of Vβ regions. Relative promoter strengths for select upstream Vβ regions were assessed using luciferase reporter constructs. pGL3 constructs containing the Eβ enhancer and the respective Vβ promoter regions were transfected into a pre–T-cell line (T3). Promoter activities were assessed at 24 h posttransfection and normalized to Renilla (RLU). Data are represented as averages ± SEM (*n* = 3) relative to the control PDβ1 promoter. ψ, pseudo-Vβ gene segments; CTRL, promoterless construct.

Fig. S4. Computational analysis of Vβ use determinants. (*A*) Scatter plot of overall correlation between natural log values of observed and fitted frequencies using the complete set of 13 features. Each circle represents one rearranging Vβ segment. The line indicates the best fit between measured and fitted re-arrangement frequencies reflect a strong correlation (Pearson correlation coefficient, 0.779; $P = 0.47$). (*B*) Relative contribution of the minimal eight features to the accuracy of fit as computed by three different approaches (Img, last, first) and the corresponding linear regression coefficients. The best fit formula is as follows:

$$\sum_{i=1}^{8} \text{Coefficient}_i \times \text{Feature}_i.$$

The raw values and coefficients corresponding to each feature are provided in Tables S5–S8.

Table S1.   3C ranks for and rearrangement frequencies

3C HindIII fragments	Rank Eβ	Rank Dβ1	Rank Dβ2	Average rank	Percent recombination
V1	21	20	25	22	4.80392
V2	24	23	17	21.3	1.890142
V3	24	23	17	21.3	4.983436
V4	1	5	2	2.67	1.766591
V5	1	5	2	2.67	4.590954
V6	13	3	12	9.33	0
V7	18	8	22	16	0
V8	18	8	22	16	0
V9	5	7	10	7.33	0
V10	10	16	3	9.67	0
V11	9	12	6	9	0
V12-1	6	14	19	13	5.360448
V13-1	8	9	11	9.33	4.710525
V12-2	8	9	11	9.33	3.534854
V13-2	8	9	11	9.33	12.76473
V12-3	7	4	8	6.33	0
V13-3	2	19	16	12.3	6.387671
V14	3	1	4	2.67	2.558942
V15	17	10	14	13.7	1.14465
V16	11	11	21	14.3	5.066513
V17	12	6	1	6.33	2.652322
V18	4	2	23	9.67	0
V19	4	2	23	9.67	10.95472
V20	15	21	7	14.3	2.722579
V22	22	18	5	15	2.254725
V23	23	13	20	18.7	2.528768
V24	25	17	24	22	4.680041
V25	25	17	24	22	0
V26	19	24	18	20.3	2.878938
V27	20	22	9	17	0
V28	14	25	15	18	0
V29	16	15	13	14.7	2.927078

Vβs present in the same fragment are given identical ranks.

Table S2. Primers and probes for Vβ utilization assay

Taqman probes (5′FAM and 3′ TAMRA from Sigma Life Sciences)	Sequences
Jβ 1.1 probe	5′FAM-TGTGAGTCTGGTTCCTTTACCAA-3′TAMRA
Jβ 2.1 Probe	5′HEX-TAGGACGGTGAGTCGTGTCC-3′TAMRA

Primers for cloning VβJβ template plasmids	Sequences
Jβ 1.1 F	5′-GACAGACGGATCCTGGCACTGTGCAAACACAGAAGTC-3′
Jβ 1.1 R	5′-TACATCGCGGCCGCACTCGAATATGGACACGGAGGACA-3′
Jβ 2.1 F	5′GACAGACGGATCCGTAACTATGCTGAGCAGTTCTTCGGACC-3′
Jβ 2.1 R	5′-TACATCGCGGCCGCAGTCCTGGAAATGCTGGCACAAAC-3′
V1-F	5′-TATCTCGAGCTGGAGCAAAACCCAAGGTGG-3′
V1-R	5′-CGAGAAGCTTTGCAGTACAAGGTTCTGCCCT-3′
V2-F	5′-TATCTCGAGCGAAAATTATCCAGAAACCAA-3′
V2-R	5′-CGAGAAGCTTGCACAGAAGTATGTGGCCGAG-3′
V3-F	5′-TATCTCGAGCAGATGGTGACCCTCAATTGT-3′
V3-R	5′-TAGCGAAGCTTTAAGCTGCTGGCACAGAAG-3′
V4-F	5′-TATCTCGAGGACGGCTGTTTTCCAGACTC-3′
V4-R	5′-CGAGAAGCTTTGGCACAGAGATACACAGCAG-3′
V5-F	5′-TATCTCGAGGATATCTAATCCTGGGAAGAGC-3′
V5-R	5′-CGAGAAGCTTCTGCCGTGGATCCAGAAGACT-3′
V6-F	5′-TATCTCGAGGTTACAGACATGGGACAGAATGTCA-3′
V6-R	5′-CGAGAAGCTTAGCTGCTGGCATACATAGTGGAGT-3′
V7-F	5′-TATCTCGAGAGCAGGCTCTGTCTTCTGACTTGT-3′
V7-R	5′-CGAGAAGCTTAGAACAGTGCAGAGTCCTTTGGCT-3′
V8-F	5′-TAGCCTCGAGCATTCAGACTCCCAAATCAT-3′
V8-R	5′-TAGCGAAGCTTTCTGTGCATGATCTGGAGAC-3′
V9-F	5′-TATCTCGAGGTGACACAATTTCTGGTCCTACTGG-3′
V9-R	5′-CGAGAAGCTTCTTCTGGCACAGAGATAGATGCCT-3′
V10-F	5′-TATCTCGAGGGTGGAATCACCCAGACACCTAGATA-3′
V10-R	5′-CGAGAAGCTTAGTACATGGAGGTCTGGTTGGAACTG-3′
V11-F	5′-TATCTCGAGAGGCACTTCTGATATGTGGCCTCT-3′
V11-R	5′-CGAGAAGCTTAGTTAGAAACCATGGCTCTTGCCC-3′
V12-1-F	5′-TATCTCGAGCTGACGTGTATTCCCATCTCT-3′
V12-1-R	5′-TAGCGAAGCTTTCCAGTTCCAAGGCACTCATG-3′
V13-1-F	5′-TATCTCGAGTGGTTAGCCCAAGTGTGCTTCTCT-3′
V13-1-R	5′-CGAGAAGCTTAAGCCAATTCCAGCAGGAGGAAGA-3′
V12-2-F	5′-TATCTCGAGCATTGCTGCTGCTGCTGCTGC-3′
V12-2-R	5′-TAGCGAAGCTTACACGGCAGAGTCCTCTAG-3′
V13-2-F	5′-TATCTCGAGTCCTGTGTTCAAGTGAGTGCTGGT-3′
V13-2-R	5′-CGAGAAGCTTTTGGTCTGGAGGCCTTGTATCCAT-3′
V12-3-F	5′-TAGCCTCGAGCCTTCTCCCCAGGTTCAGC-3′
V12-3-R	5′-TAGCGAAGCTTCACAGTAAAGTCCTCTAGGTCC-3′
V13-3-F	5′-TATCTCGAGGACGATATGATCAGGCTTTG-3′
V13-3-R	5′-TAGCGAAGCTTAGAAATATACAGCTGTCTGAG-3′
V14-F	5′-TATCTCGAGTATGCAGTCCTACAGGAAGGGCAA-3′
V14-R	5′-CGAGAAGCTTAAACTGCTGGCACAGAGATAGGTG-3′
V15-F	5′-TATCTCGAGCAGACACCCAGACATGAGGT-3′
V15-R	5′-CGAGAAGCTTACAGCTGAGTCCTTGGGTTCTG-3′
V16-F	5′-TATCTCGAGCACCTAGGCACAAGGTGACA-3′
V16-R	5′-TAGCGAAGCTTCAGGACTCAGCGGTGTATCT-3′
V17-F	5′-TATCTCGAGGGATACTACGGTTAAGCAGAAC-3′
V17-R	5′-TAGCGAAGCTTAGCACAGAGGTACATGGCAG-3′
V18-F	5′-TAGCCTCGAGGCTGGTGTCACCACGAACCT-3′
V18-R	5′-TAGCGAAGCTTCTCTGCATCTTCCAGATCTGC-3′
V19-F	5′-TATCTCGAGCTCAGACACCCAAATTCCTGA-3′
V19-R	5′-CGAGAAGCTTGCTATACTGCTGGCACAGAGA-3′
V20-F	5′-TATCTCGAGCGTCTATCAATATCCCAGAAG-3′
V20-R	5′-CGAGAAGCTTAGCACCACAGAGATATAAGCC-3′
V21-F	5′-TAGCCTCGAGGTTGTCCAGAATCCTAGACAT-3′
V21-R	5′-TAGCGAAGCTTGTACACAGCTGAATCTGTTAG-3′
V22-F	5′-TATCTCGAGCCAAGTTATCCAGACTCCAT-3′
V22-R	5′-TAGCGAAGCTTATAACACTGAGTCTCCAGCCTC-3′
V23-F	5′-TATCTCGAGGAAAGGCCAGGAAGCAGAGAT-3′
V23-R	5′-CGAGAAGCTTGCTGGAGCACAAGTACAGTGC-3′

Table S2. Cont.

V24-F	5′-TATCTCGAGGAGTAACCCAGACTCCACGAT-3′
V24-R	5′-CGAGAAGCTTGACTGCTGGCACAGAGCTACA-3′
V25-F	5′-TAGCCTCGAGCTAGCTTCAAGGCTCTTCTA-3′
V25-R	5′-TAGCGAAGCTTATGTAGAATCTCCTGCTTCT-3′
V26-F	5′-TATCTCGAGCAGACTCCAAGATATCTGGTG-3′
V26-R	5′-CGAGAAGCTTCTGCTGGCACAGAGGTACAGT-3′
V27-F	5′-TAGCCTCGAGCTCCAAAGTACTCTATTATG-3′
V27-R	5′-TAGCGAAGCTTGAGGTAGGATTCATTCTCTG-3′
V28-F	5′-TAGCCTCGAGCATCCAAATCGCAAGACACC-3′
V28-R	5′-TAGCGAAGCTTAGGTGCACACATGCCTGGTCG-3′
V29-F	5′-TATCTCGAGCTGATCAAAAGAATGGGAGAG-3′
V29-R	5′-CGAGAAGCTTCTAGCACAGAAGTACACAGATG-3′
V30-F	5′-TATCTCGAGTGCTTGCCTCATGGATCTCTGTCT-3′
V30-R	5′-CGAGAAGCTTGAACTACAGAAATAGATACTGC-3′
V31-F	5′-TAGCCTCGAGCTGAGACTGATTACATGTAA-3′
V31-R	5′-TAGCGAAGCTTAGAAGCCAGAGTGGCTGAGA-3′
qPCR primers for Taqman assay	**Sequences**
qV1F	5′-GCCACACGGGTCACTGATAC-3′
qV2F	5′-GTTCAAAGAAAAACCATTTAG-3′
qV3F	5′-GATGGTTCATATTTCACTCT-3′
qV4F	5′-CAGATAAAGCTCATTTGAAT-3′
qV5F	5′-GCCCAGACAGCTCCAAGCTAC-3′
qV6F	5′-CAGAGATGCCTGATGGATTGTT-3′
qV7F	5′-CAGCACACCAATTTGGTGACT-3′
qV8F	5′-GAGGTCTCTAAGGGGTAC-3′
qV9F	5′-CTTCTCCATGTTGAAGAGCCAA-3′
qV10F	5′-AGAAATGAGATACAGAGCTTTCC-3′
qV11F	5′-AGTTAGAAACCATGGCTCTTGC-3′
qV12-1F	5′-'TAGCAATGTGGTCTGGTACCAG-3′
qV13-1F	5′-GGTACAAGGCCACCAGAACA-3′
qV12-2F	5′-TCTCTCTGTGGCCTGGTATCAA-3′
qV13-2F	5′-GCTGGCAGCACTGAGAAAGGA-3′
qV12-3F	5′-CCTGAGTGCCTTGGACCT-3′
qV13-3F	5′-TTCCCTTTCTCAGACAGCTGTA-3′
qV14F	5′-TATCAGCAGCCCAGAGACCAG-3′
qV15F	5′-CACTCTGAAGATTCAACCT-3′
qV16F	5′-CTCAGCTCAGATGCCCAAT-3′
qV17F	5′-CAATCCAGTCGGCCTAACA-3′
qV18F	5′-CCACGAACCTAAGATACAT-3′
qV19F	5′-CTCGAGAGAAGAAGTCATCT-3′
qV20F	5′-CAGTCATCCCAACTTATCCT-3′
qV21F	5′-GCTAAGAAACCATGTACCAT-3′
qV22F	5′-CAGTTCCTCTGAGGCTGGA-3′
qV23F	5′-CTGTGTGCCCCTCCAGCTCA-3′
qV24F	5′-CTCAGCTAAGTGTTCCTCGA-3′
qV25F	5′-CTATGTGGCATATTACTGGT-3′
qV26F	5′-CCTTCAAACTCACCTTGCAGC-3′
qV27F	5′-CATTGTTCATATGGCATT-3′
qV28F	5′-CTCTGATAGATATATCAT-3′
qV29F	5′-CTGATTCTGGATTCTGCTA-3′
qV30F	5′-CAATGCAAGGCCTGGAGACA-3′
qV31F	5′-AAATCAAGCCCTAACCTCTAC-3′
qJβ1.1R	5′-CTCGAATATGGACACGGAGGACATGC-3′
qJβ2.1R	5′-CCTGATACAGGGCCTTGGATAGTTA-3′

Table S3. Primers and probes for 3C assay

3C anchor primers and Taqman probes	Sequences
(5'FAM and 3' TAMRA from Sigma Life Sciences)	
Dβ_1 HindIII probe	5'-AAGGCATTGTTGCATGATCCT-3'
Dβ_2 HindIII probe	5'-AAATGCTGGGCCTCTGTAGA-3'
Eβ_ HindIII probe	5'-CATAAGCATTGTCATGTTTGTGACA-3'
ERCC3 HindIII probe	5'-AAAGCTTGCACCCTGCTTTAGTGGCC-3'
Dβ_1 HindIII primer	5'-TGAAATTTTTCTGCCGAAAGGAC-3'
Dβ_2 HindIII primer	5'-GCGGGATCCAAGAGAACTCA-3'
Eβ_ HindIII primer	5'-GAAAATTGGCATCGGTTTGC-3'

HindIII primers	Sequences
V1	5'-TATCTCTGTGGGGCATGCAG-3'
V2	5'-TTTCATTCACAGCCGACCAG-3'
V3	5'-TTTCATTCACAGCCGACCAG-3'
V4	5'-AGCTCGACACAGAAAGCAAGTT-3'
V5	5'-AGCTCGACACAGAAAGCAAGTT-3'
V6	5'-GGTTCCCTTCACTTCCCACA-3'
V7	5'-GTCCGCTAGCAGCCAGAGTT-3'
V8	5'-GTCCGCTAGCAGCCAGAGTT-3'
V9	5'-ACCAGAGGGCAGCTGAAAAT-3'
V10	5'-GTGCCTGTACCATGCTGTGG-3'
V11	5'-TTCAGCAAGTAGGTGCGAAGA-3'
V12-1	5'-TGGTGGGATCCTGACAGCTTATA-3'
V13-1	5'-CCATCTGCATGAACACCTTCTT-3'
V12-2	5'-CCATCTGCATGAACACCTTCTT-3'
V13-2	5'-CCATCTGCATGAACACCTTCTT-3'
V12-3	5'-GGATCTTGGTCTCGGGAGGT-3'
V13-3	5'-CTCAGCTGCACCCTCACAAC-3'
V14	5'-CAGGCTTTTGAGTGGCATGT-3'
V15	5'-AGGCAGGAGGTGAGTCTTGG-3'
V16	5'-TATCATGCCCAGCTGCATTC-3'
V17	5'-GTTAGGCCGACTGGATTGGA-3'
V18	5'-GGCAGTGTTACAGAACCCAGTG-3'
V19	5'-GGCAGTGTTACAGAACCCAGTG-3'
V20	5'-TGTGATGGGTTGTCATCTGGA-3'
V22	5'-CCAAGGGATGATGTCACAGG-3'
V23	5'-TACACCGGCCAGGAGAGACT-3'
V24	5'-ACTAGGCCAGCAGAGGATGC-3'
V25	5'-ACTAGGCCAGCAGAGGATGC-3'
V26	5'-AGCATAGGATTGGGCCTCAG-3'
V27	5'-CATCACTGCGCCTAGCAATC-3'
V28	5'-GCGTGTGCCACGTTTTTGTA-3'
V29	5'-CTCTAGCAATCCCCCTGTGC-3'
V31	5'-AAGGAGAGAGCAGGCCACAG-3'
Dβ_1	5'-AAGGCATTGTTGCATGATCC-3'
Dβ_2	5'-TGGGGCCCTCACTTTTCTTA-3'
Eβ	5'-TCCTAAGGAGAGGCAGAGTGG-3'
ERCC3	5'-GACTTCTCACCTGGGCCTACA-3'

Table S4. Luciferase assay cloning primers

Primer name	Sequences
Eβ-F	5′-ATTGGATCCGTTAACCAGGCACAGTAGGACC-3′
Eβ-R	5′-ATTGGATCCCCATGGTGCATACTGAAGGCTTC-3′
Pro-V1F	5′-TAGCCTCGAGGAGTGACTAGTTACTTCTGC-3′
Pro-V1R	5′-TAGCGAAGCTTCTCTGAGACCTCAGGTTCTC-3′
Pro-V3-F	5′-TATCTCGAGGGGACTCAGTTCAGTAGTC-3′
Pro-V3-R	5′-CGAGAAGCTTAGTAGGGTCACGGCAGGAA-3′
Pro-V4F	5′-TAGCCTCGAGTGTGCTAAGGGCACCAATGAAT-3′
Pro-V4R	5′-TAGCGAAGCTTGTTGGGTCAAGGCAGGGCAAAT-3′
Pro V5-FX	5′-TAGCCTCGAGTATCCATTGTATGCTCTGTTTG-3′
Pro V5-RH	5′-TAGCGAAGCTTGGTGGAATCAGGCTCCAGACG-3′
Pro V6-FX	5′-TAGCCTCGAGCTACAAGCTCCCAAGAGAGAG-3′
Pro V6-RH	5′-TAGCGAAGCTTCTCTGGAGAAGACAGAGGAC-3′
Pro-V7F	5′-TAGCCTCGAGGCTGCTGAATAGCAAGTTTCCAG-3′
Pro-V7R	5′-TAGCGAAGCTTTTGGAGGTTTGGATCTGTAGTCT-3′
Pro V9-FX	5′-TAGCCTCGAGGGAACTTTCATGTGAGGAGA-3′
Pro V9-RH	5′-TAGCGAAGCTTCTGCAAAAATATAAGTTGTGAACAG-3′
Pro V10-FX	5′-TAGCCTCGAGGGGATATCTCTATGCTTTAATG-3′
Pro V10-RH	5′-TAGCGAAGCTTCTGGAGAAGGAGGCATAAGGA-3′
Pro-V11F	5′-TAGCCTCGAGTTCCCTACAGTGTCAAGGGCTG-3′
Pro-V11R	5′-TAGCGAAGCTTTGTACCCACAGGGTTGTTCTCA-3′
Pro-V12-2-F	5′-TAGCCTCGAGCAACTGACTCAGAGAAAAAC-3′
Pro-V12-2-F	5′-TAGCGAAGCTTCCTCTCAGGATACTGGTCTCT-3′
Pro-V14F	5′-TACATCGCTAGCCATTTATGTGTACCATAATAAT-3′
Pro-V14R	5′-TAGCCTCGAGGGCAGATTGAGGGCAGAGGAG-3′
Pro-V16F	5′-TAGCCTCGAGTTGCAATCTACCTCTGCTGCTC-3′
Pro-V16R	5′-TAGCGAAGCTTTTGTGATGACACCACTGTCTCCG-3′
Pro V17-FX	5′-TAGCCTCGAGGCAGGTGTGACCTACGATAAC-3′
Pro V17-RH	5′-TAGCGAAGCTTGGATGGTCCAGAACAGGAAA-3′
Pro-V19-F	5′-TATCTCGAGCATTTGAGAAAGACAACAA-3′
Pro-V19-R	5′-CGAGAAGCTTAGTTTGGAGGGACTTTCTT-3′
Pro-V20F	5′-TAGCCTCGAGGATAAGGTAACTGAAGCGGGA-3′
Pro-V20R	5′-TAGCGAAGCTTCTTCAGTGTTGACTTCACACC-3′
Pro-V22F	5′-TAGCCTCGAGGATGAAATATGGTAACAAGG-3′
Pro-V22R	5′-TAGCGAAGCTTAGGAGATAAAGGGCTACATA-3′
Pro-V24F	5′-TACATCGCTAGCCCAATGATATGTGCAGAGATGA-3′
Pro-V24R	5′-TAGCCTCGAGGATCACACTAGGCCAGCAGAG-3′
Pro-V25F	5′-TAGCCTCGAGCAATTGGGCCATCTTCTGCCAC-3′
Pro-V25R	5′-TAGCGAAGCTTCAGGTGGATACTTCATTCC-3′
Pro-V28F	5′-TAGCCTCGAGAGTTGTCTTGTGGGCAACTCTG-3′
Pro-V28R	5′-TAGCGAGATCTGCTAGATAGCCTCAAGGCTGCAAA-3′

Table S5. Recombination substrate oligoes

Primer name	Sequences
RS V1 F	TAGCCTCGAGATACGGAGCTGAGGCTGCAAG
RS V1 R	TACATCGCGGCCGCAGTCACCTTATAACTCATGCA
RS V15F	TAGCCTCGAGCCTTCTCCACTCTGAAGATTC
RS V15R	TACATCGCGGCCGCTTCCACCCAAGATTTCTTAA
RS V16F	TAGCCTCGAGACTCAACTCTGAAGATCCAGA
RS V16R	TACATCGCGGCCGCTAATGTAATACTCGTTACCAT
RS V18F	TAGCCTCGAGCCCAACATCCTAAAGTGGG
RS V18R	TACATCGCGGCCGCTTCCTCCGTAAGCATGGTG
RS V20F	TAGCCTCGAGCAGTCATCCCAACTTATCCT
RS V20R	TACATCGCGGCCGCCTCCTGGGTACCCTCCCATTTC
RS V23F	TAGCCTCGAGCACTCTGCAGCCTGGGAATC
RS V23R	TACATCGCGGCCGCTGACTTGGTCTGGGTGTGCTG
RS V24F	TAGCCTCGAGAGTGCATCCTGGAAATCCTAT
RS V24R	TACATCGCGGCCGCAGACCTGGCCTGTTTCTCATG
RS V26F	TAGCCTCGAGCAAGAAGTTCTTCAGCAAATA
RS V26R	TACATCGCGGCCGCGATACAGGTTTCAGTTAGTT

Table S6. Computational analysis coefficients for determinants of Vβ frequencies (all *Tcrb* V gene segments): Classifer step, three features

Features	Estimate	SE	t	Pr (>|t|)
Intercept	1.09059	1.52205	0.717	0.47903
Recombination signal information content (RIC) score	0.08803	0.02619	3.362	0.00207
Formaldehyde-assisted isolation of regulatory elements (FAIRE)	0.03185	0.01639	1.944	0.06105
RNA Pol II	0.65913	0.26654	2.473	0.01909

Table S7. Computational analysis coefficients for determinants of Vβ frequencies (all *Tcrb* V gene segments): Combinatorial analysis of 13 features and their correlation to recombination frequency

Number of features	Pearson correlation coefficient	P value
13	0.77954	0.4707
8	0.74191	0.1015
7	0.72604	0.07434
6	0.71277	0.04925
5	0.68818	0.03779
4	0.66405	0.0265
3	0.64982	0.01359
2	0.60304	0.01089
1	0.53998	0.00782

Table S8. Coefficients for determinants of Vβ frequencies (rearranging Vβ segments)

Features	Estimate	SE	t	Pr (>|t|)
All *Tcrb* V gene segments (regressor step, 13 features)				
Intercept	0.08707	5.81E+00	0.015	0.9882
RIC score	0.08817	3.72E−02	2.373	0.0273
3C cross-linking	−2.17745	3.14E+00	−0.693	0.4961
Transcription	−0.1299	1.56E−01	−0.83	0.4156
CCCTC-binding factor (CTCF)	0.9394	1.24E+00	0.756	0.4579
FAIRE	0.01538	2.46E−02	0.625	0.5384
H3ac	−0.31847	5.05E−01	−0.63	0.5353
H3K27ac	0.03124	3.86E−02	0.81	0.4271
H3K4me1	0.16488	6.88E−01	0.24	0.813
H3K4me2	0.0194	1.67E−02	1.159	0.2595
H3K4me3	−0.08483	3.68E−01	−0.231	0.8197
H3K9me2	0.74873	1.27E+00	0.59	0.5618
P300	−0.03168	3.03E−02	−1.047	0.3069
RNA Pol II	1.10351	5.21E−01	2.119	0.0462
All *Tcrb* V gene segments (Regressor step, 5 features)				
Intercept	0.62866	0.35047	1.794	0.0907
Transcription	−0.0613	0.0467	−1.313	0.2066
CTCF	0.31418	0.25634	1.226	0.2371
H3K4me2	0.00779	0.00365	2.137	0.0475
H3K4me3	0.16139	0.0666	2.423	0.0268
P300	−0.0066	0.00646	−1.027	0.319

Other Supporting Information Files

Dataset S1 (XLSX)