Supplementary figures

Figure S1

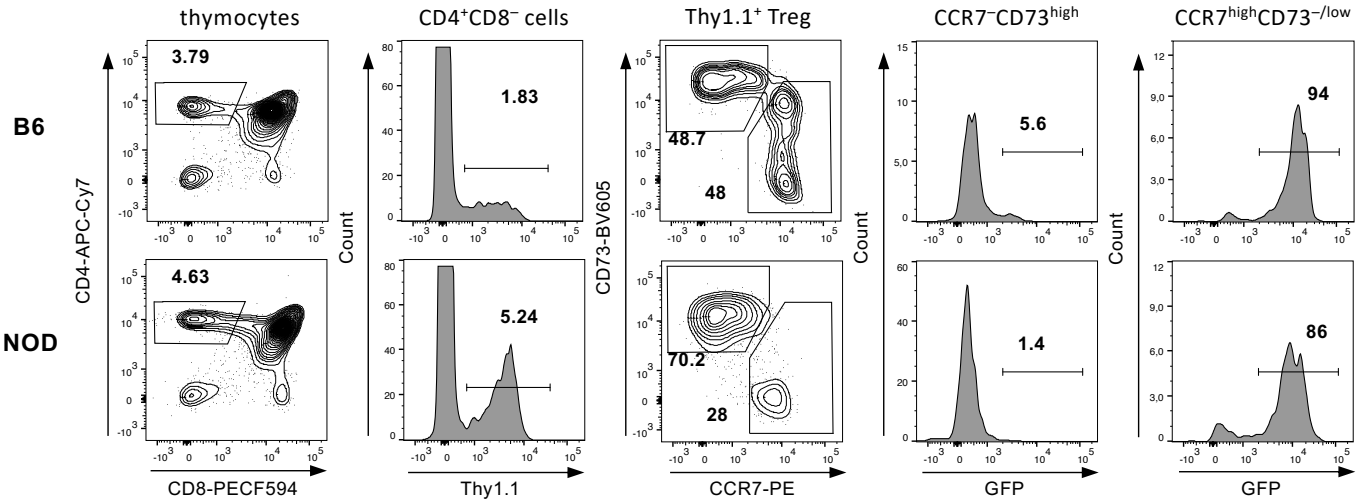

Figure S1. *A CCR7^{high}CD73$^{-/low}$ phenotype reliably identifies newly developing Treg in the thymus*. Thymocytes from eight-week-old B6 and NOD *Foxp3-Thy1^{a} Rag2-Gfp* females were analyzed by flow-cytometry using antibodies to CD4, CD8, Thy1.1, CD73, and CCR7. Depicted are the consecutive gates used to identify CCR7^{-}CD73^{high} and CCR7^{high}CD73$^{-/low}$ Treg and the GFP-expression on these two populations. Whereas CCR7^{-}CD73^{high} cells are GFP^{-} and therefore recirculating, CCR7^{high}CD73$^{-/low}$ cells are mostly GFP^{+} and therefore newly developed. Values indicate % of cells within depicted gates in this representative experiment out of more than three performed.

Figure S2

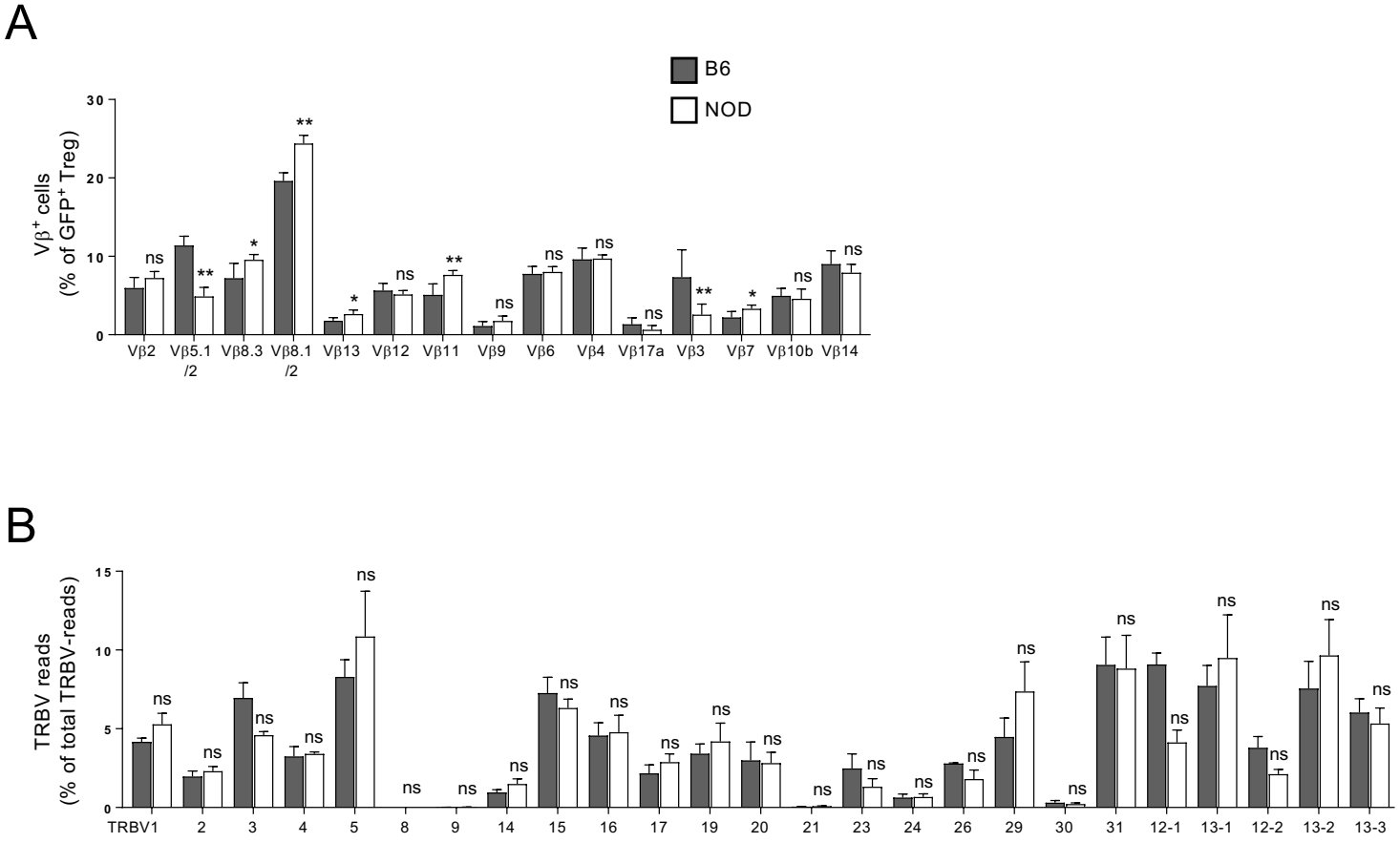

Figure S2. *TRBV usage by newly developed Treg*. Newly developed thymic Treg from eight-week-old wt B6 and NOD females were analyzed by flow-cytometry (A) and TCRseq (B) as described in the legends to Fig. 1 and 2. Usage of TRBV found in NOD (white bars) and B6 mice (black bars) is depicted as mean values ± SD, n=6 mice per mouse-strain for flow-cytometry and n=3 pools of eight to ten mice per strain for TCRseq. ns = not significant; *P < 0.05; **P < 0.01 (Mann-Whitney test)

Figure S3

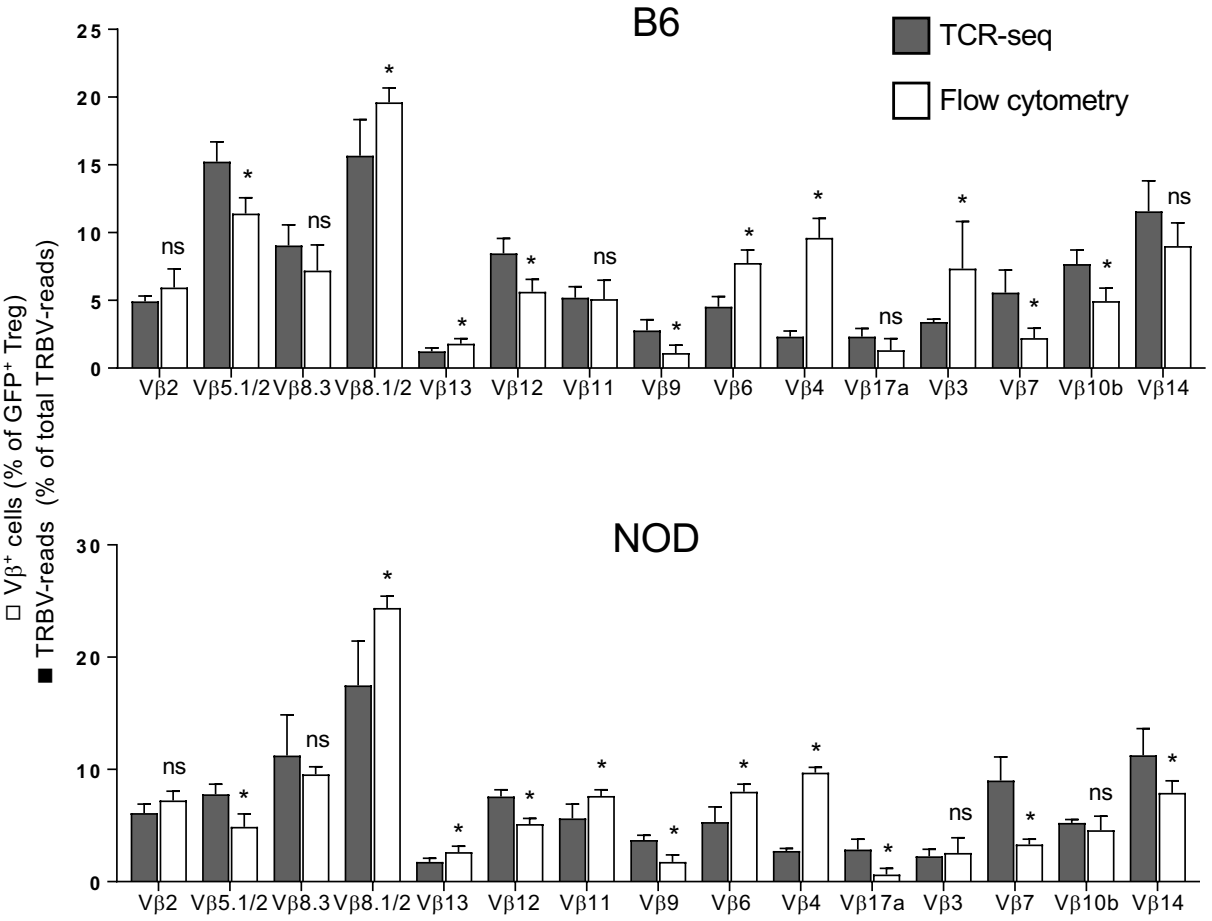

Figure S3. *Comparison of the TRBV usage by newly developed Treg as determined by flow-cytometry and TCRseq in indicated mouse-strains*. Newly developed thymic Treg from eight-week-old wt B6 and NOD females were analyzed by flow-cytometry and TCRseq as described in the legends to Fig. 1 and 2. Usage of TRBV found by TCRseq (black bars) and by flow-cytometry (white bars) is depicted as mean values ± SD, n=6 mice per mouse-strain for flow-cytometry and n=3 independently analyzed pools of eight to ten mice per strain for TCRseq. For TCRseq, this analysis was limited to the TRBV segments to which antibodies were available, the sum of which is 100% for flow-cytometry and for TCRseq. For TCR-Vβ *vs.* TRBV correspondence, see http://www.imgt.org. ns = not significant; *P < 0.05 (Mann-Whitney test)

Figure S4

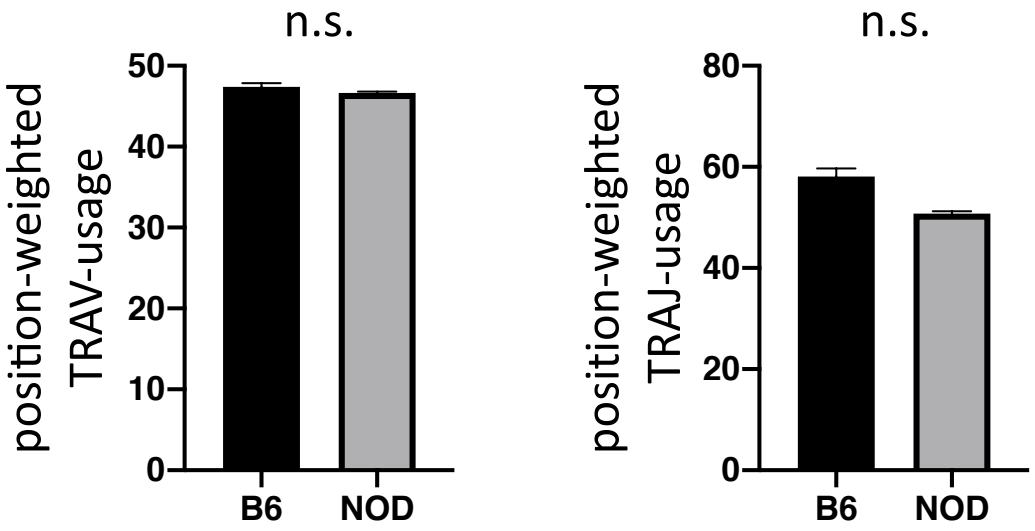

Figure S4. *TRAV and TRAJ usage according to genome-position*. $CD4^+CD8^-Thy1.1^+GFP^+$ thymic Treg from eight-week-old *Rag2-Gfp Foxp3-Thy1^a* mutant B6 and NOD female mice were analyzed by TCRseq. "position-weighted TRAV-usage" = $(\sum_{i=1}^{n}((\%\ \text{usage TRAVi})\ \text{x i})) \div n$, in which "i" is the TRAV number according to genome position, numbered 1 to n from 5' to 3'. Same for "position-weighted TRAJ-usage". Mean values ± SD, n=3 pools of eight to ten mice per strain. n.s., not significant (Mann-Whitney test).

Figure S5

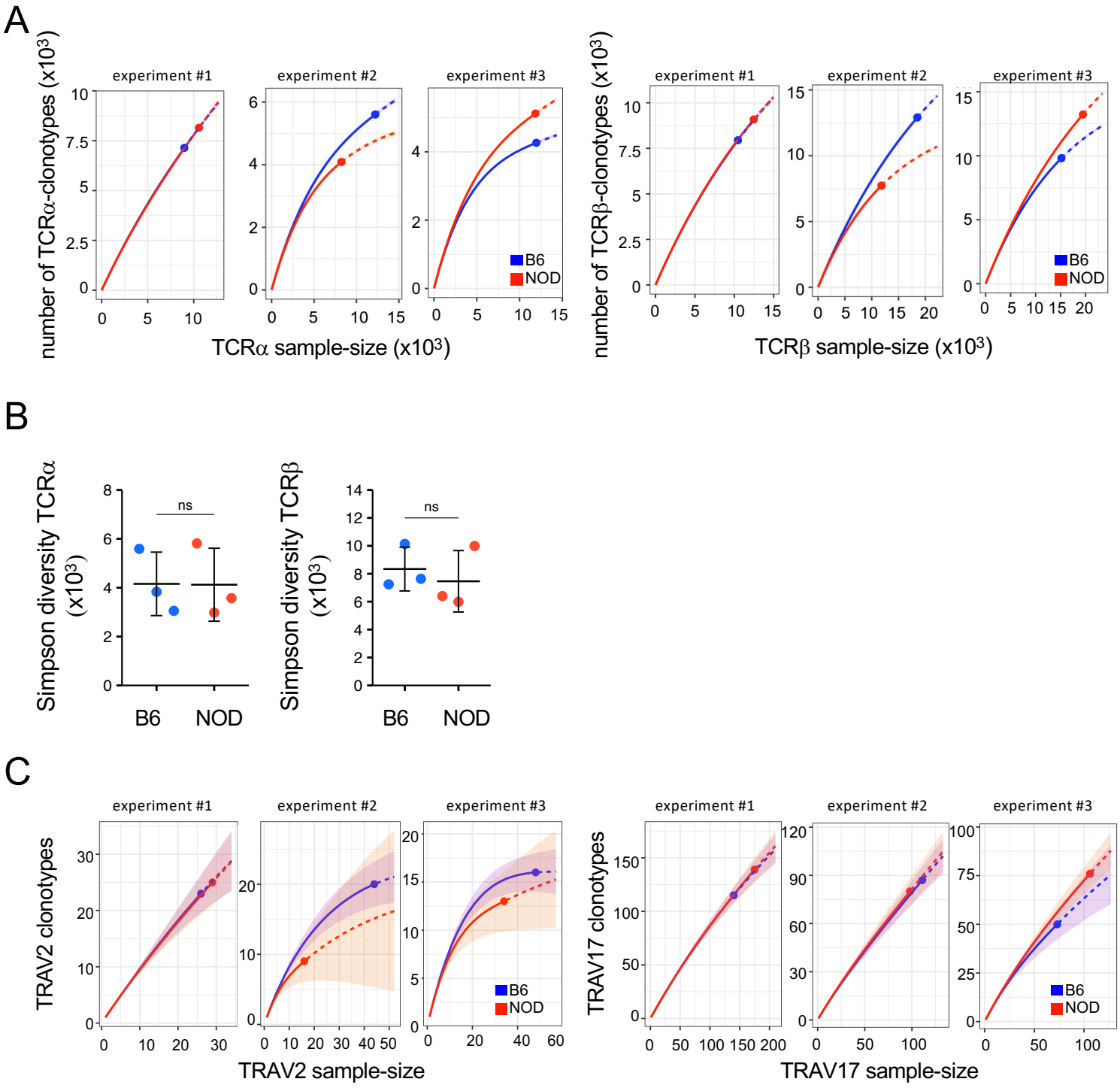

Figure S5. *Diversity of TCRα and TCRβ clonotypes expressed by newly developed Treg*. CD4⁺CD8⁻Thy1.1⁺GFP⁺ thymic Treg from eight-week-old *Rag2-Gfp Foxp3-Thy1ᵃ* mutant B6 and NOD female mice were analyzed by TCRseq. (A) Rarefaction-plots of TCRα and TCRβ-clonotypes, assessed in the three experiments performed. (B) Simpson-diversity of the TCRα and TCRβ-clonotypes. Mean values ± SD, n=3 pools of eight to ten mice per strain. ns, not significant (Mann-Whitney test). (C) Rarefaction-plots of TRAV2 and TRAV17-clonotypes. Shadows indicate 95% confidence intervals.

Figure S6

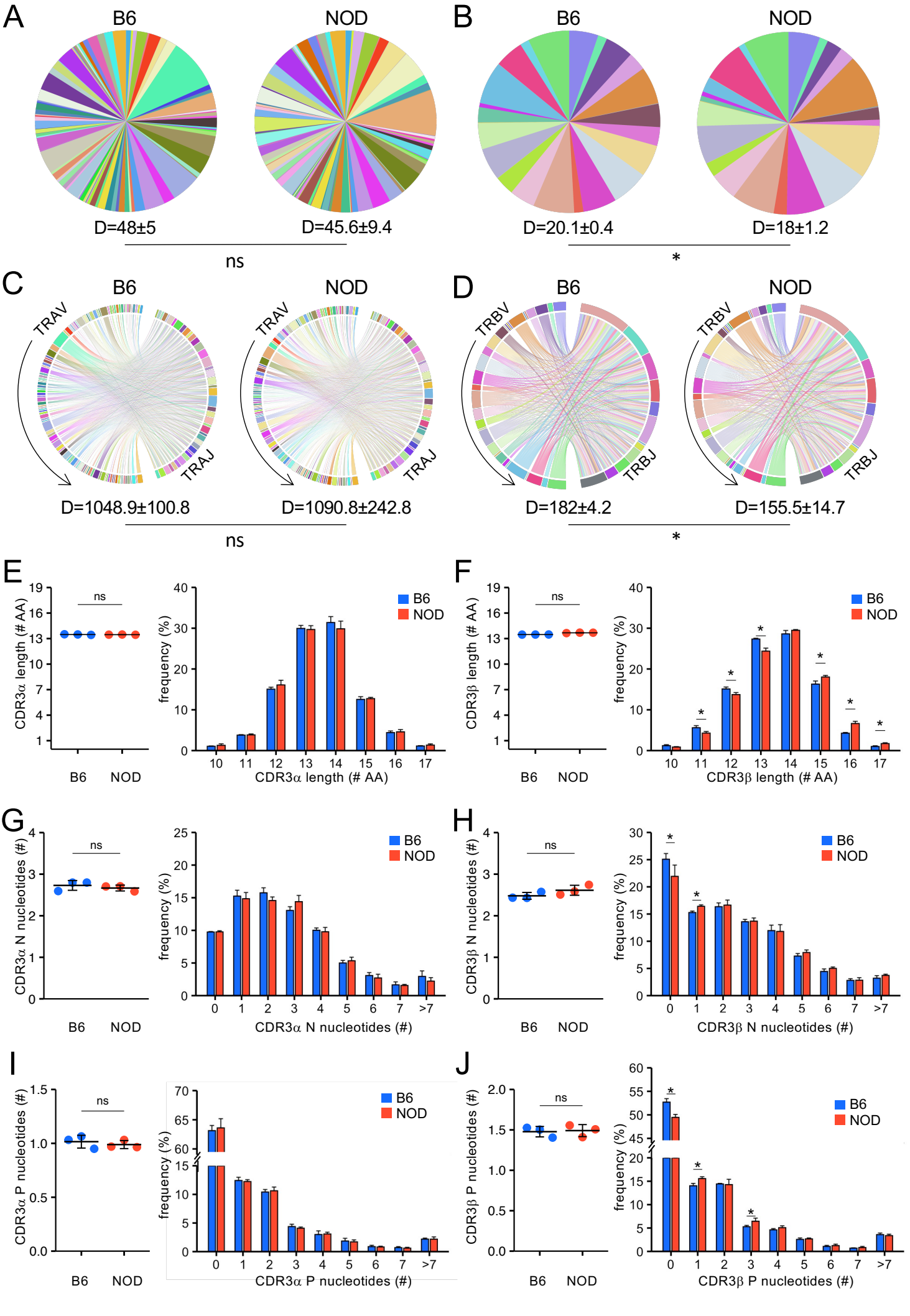
A
B6
NOD
D=48±5
D=45.6±9.4
ns
B
B6
NOD
D=20.1±0.4
D=18±1.2
*
C
B6
NOD
TRAV
TRAJ
D=1048.9±100.8
D=1090.8±242.8
ns
D
B6
NOD
TRBV
TRBJ
D=182±4.2
D=155.5±14.7
*
E
CDR3α length (# AA)
ns
B6
NOD
frequency (%)
CDR3α length (# AA)
F
CDR3β length (# AA)
CDR3β length (# AA)
G
CDR3α N nucleotides (#)
CDR3α N nucleotides (#)
H
CDR3β N nucleotides (#)
CDR3β N nucleotides (#)
I
CDR3α P nucleotides (#)
CDR3α P nucleotides (#)
J
CDR3β P nucleotides (#)
CDR3β P nucleotides (#)

Figure S6. *Similar V and J segment usage and CDR3-lengths of TCRα and TCRβ expressed by peripheral Treg that had recirculated back to the NOD vs. B6 thymi*
Recirculating $CD4^+CD8^-Thy1.1^+GFP^-$ Treg isolated from thymi of eight-week-old *Rag2-Gfp Foxp3-Thy1^a* mutant B6 and NOD female mice were FACS-sorted and analyzed by UMI-based TCRseq. (A) TRAV-, (B) TRBV-, (C) TRAV/TRAJ, and (D) TRBV/TRBJ-usage, mean values of the three samples. In (C) and (D), V and J segments are ordered as in the genome (arrows). The indicated diversity-measure "D" is the Shannon-diversity (*i.e.* the exponential of the Shannon entropy), represented as mean ± SD. CDR3α (E) and CDR3β (F) lengths in amino-acids. CDR3s starts with conserved Cys and Ala and end with conserved Phe. (G,H) Number of (non-germline-encoded) N-nucleotides and (I,J) (palindromic) P-nucleotides in the CDR3α (G,I) and CDR3β (H,J). In E-J, left panels indicate mean values ± SD, right panels distribution (mean values ± SD). n=3 independently analyzed pools of eight to ten thymi per strain. ns = not significant, *$P < 0.05$ (Mann-Whitney test).

Figure S7

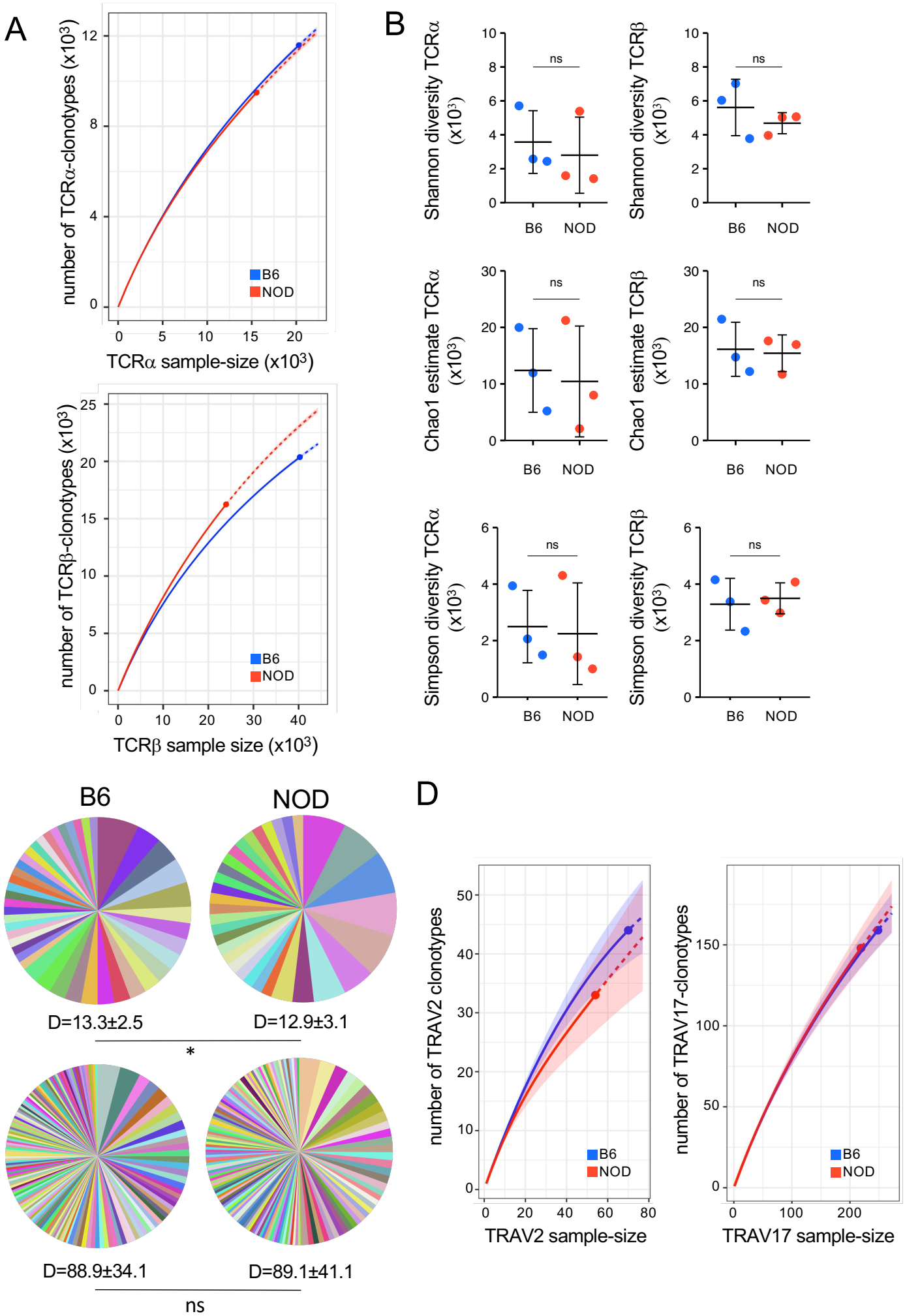
A
number of TCRα-clonotypes (x10³)
TCRα sample-size (x10³)
B6
NOD
number of TCRβ-clonotypes (x10³)
TCRβ sample size (x10³)
B
Shannon diversity TCRα (x10³)
Shannon diversity TCRβ (x10³)
Chao1 estimate TCRα (x10³)
Chao1 estimate TCRβ (x10³)
Simpson diversity TCRα (x10³)
Simpson diversity TCRβ (x10³)
ns
B6
NOD
C
B6
NOD
TRAV-2
D=13.3±2.5
D=12.9±3.1
*
TRAV-17
D=88.9±34.1
D=89.1±41.1
ns
D
number of TRAV2 clonotypes
TRAV2 sample-size
number of TRAV17-clonotypes
TRAV17 sample-size
B6
NOD

Figure S7. *Similar diversities of TCRα and TCRβ clonotypes expressed by peripheral Treg that had recirculated back to NOD vs. B6 thymi*

Recirculating $CD4^+CD8^-Thy1.1^+GFP^-$ Treg isolated from the thymi of eight-week-old *Rag2-Gfp Foxp3-Thy1^a* mutant B6 and NOD female mice were FACS-sorted and analyzed by UMI-based TCRseq.(A) Rarefaction plots of data pooled from the three experiments for TCRα (top) and TCRβ (bottom) clonotypes (as determined by V-usage, CDR3-sequence, and J-usage) in the indicated strains. Shadows indicate 95% sampling-confidence intervals. (B) Shannon diversities (*i.e*. the exponential of the Shannon entropy, top), Chao1 estimates (middle), and Simpson-diversities (bottom) for TCRα (left) and TCRβ (right)-clonotypes. Bars indicate mean values + SD. (C) Mean clonotype frequencies for TCRα using TRAV2 or TRAV17 in indicated strains. The diversity measure D indicates mean Shannon-diversity ± SD. n=3 experiments performed on pools of eight to ten thymi per mouse-strain. ns = not significant, *P < 0.05 (Mann-Whitney test). (D) Rarefaction plots of pooled data for TCRα using TRAV2 or TRAV17 in indicated strains. Shadows indicate 95% sampling-confidence intervals.

Supplemental research design and methods

High throughput sequencing of *Tcra* and *Tcrb* mRNA

RNA extraction

Total RNA was extracted by Nucleospin RNA XS (Macherey-Nagel) according to the manufacturer´s instructions and was quality controlled (RIN > 8) using Agilent 2100 BioAnalyzer.

cDNA synthesis

cDNA synthesis was performed as previously described (1) with some modifications. The primers used to generate the *Tcra* cDNA were as follows:

TRAC_RT1: 5'-CTCAGCGTCATGAGCAGGTTAAAT-3'

TRAC_RT2: 5'-CAGGAGGATTCGGAGTCCCATAA-3'

TRAC_RT3: 5'-TTTTACAACATTCTCCAAGA-3'

TRAC_RT4: 5'-TTCTGAATCACCTTTAATGA-3'

TRAC_RT5: 5'-ATGAGATAATTTCTACACCT-3'

TRAC_RT6: 5'-TTTGGCTTGAAGAAGGAGCG-3'

TRAC_RT7: 5'-TTCAAAGCTTTTCTCAGTCA-3'

TRAC_RT9: 5'-TGGTCTCTTTGAAGATATCT-3'

The primers used to generate the *Tcrb* cDNA were as follows:

TRBC_RT1: 5'-GGTAGCCTTTTGTTTGTTTG-3' (54-73 of Trbc1/2)

TRBC_RT2: 5'-CCCCTGGCCAAGCACACGAG-3' (84-103)

TRBC_RT3: 5'-TGCCATTCACCCACCAGCTC-3' (128-147)

TRBC_RT4: 5'-GCTATAATTGCTCTCCTTGT-3' (182-201)

TRBC_RT5: 5'-TTGCGAGGATTGTGCCAGAA-3' (243-262)

TRBC_RT6: 5'-CTTGTCCTCCTCTGAAAGCC-3' (292-311)

TRBC_RT7: 5'-GCCTCTGCACTGATGTTCTG-3' (342-361)

The sequence of the template-switch oligonucleotide was the following:

5’-TACACGACGCTCTTCCGATCUNNNNUNNNNUNNNNUCTTrGrGrGrGrG-3’.

Tcra library generation

Library preparation was performed as previously described (1) with some modifications. The first PCR reaction was performed with oligonucleotides UNIV5_P12v2 (5’-TACACGACGCTCTTCCGATC-3’) and TRAC3_P1v2 (5’-AAGTCGGTGAACAGGCAGAG-3’)(2’ 95°C, 10 cycles of 20” 95°C, 15” 59°C, 45” 70°C, final incubation 3.5’ 70°C). The amplicons were then purified using Agenecourt AMPure XP beads. The second, semi-nested PCR was performed with oligonucleotides UNIV5_P12v2 and TRAC3_P2v2 (5’-AGCAGGTTCTGGGTTCTGGA-3’)(2’ 95°C; 20 cycles of 20” 95°C, 15” 59°C, 45” 70°C, final incubation 3.5’ 70°C). The third PCR was performed using oligonucleotides UNIV5_P3v2 (5’-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC-3’) and TRAC3_P3v2-index (5’-CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTAGCAGGTTCTGGGTTCTGGA-3’, in which XXXXXXXX is a sequencing index)(2’ 95°C; 1 cycle of 20” 95°C, 15” 59°C, 45” 70°C, 5 cycles of 20” 95°C, 15” 75°C, 45” 70°C, final incubation 3.5’ 70°C)). For the fourth PCR, oligonucleotides UNIV5_P4v2 (5’-AATGATACGGCGACCACCGA-3’) and UNIV3_P4v2 (5’-CAAGCAGAAGACGGCATACGA-3’) were used (2’ 95°C, 5 cycles of 20” 95°C, 15” 60°C, 45” 70°C; final incubation of 3.5’ at 70°C). The quality of each library was verified using Agilent 2100 BioAnalyzer (mean peak size 640 bp). The samples were indexed and sequenced with 300pb paired-end on MiSeq sequencer (Illumina).

Tcrb library generation

Library preparation was performed as previously described for *Tcra* libraries (1), adapted to the *Tcrb* locus. The first PCR reaction was performed with oligonucleotides UNIV5_P12v2 (5'-TACACGACGCTCTTCCGATC-3') and TRBC3_P1v2 (5'-TGATGGCTCAAACAAGGAGACC-3')(2' 95°C, 10 cycles of 20" 95°C, 15" 59°C, 45" 70°C, final incubation 3.5' 70°C). The amplicons were then purified using Agenecourt AMPure XP beads. The second, semi-nested PCR was performed with oligonucleotides UNIV5_P12v2 and TRBC3_P2v2 (5'-GGGTGGAGTCACATTTCTCAGAT-3')(2' 95°C; 20 cycles of 20" 95°C, 15" 59°C, 45" 70°C, final incubation 3.5' 70°C). The third PCR was performed using oligonucleotides UNIV5_P3v2 (5'-AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATC-3') and TRBC3_P3v2-index (5'-CAAGCAGAAGACGGCATACGAGATXXXXXXXXGTGACTGGAGTTCAGACGTGTGCTC TTCCGATCTGGGTGGAGTCACATTTCTCAGAT-3', in which XXXXXXXX is a sequencing index)(2' 95°C; 1 cycle of 20" 95°C, 15" 59°C, 45" 70°C, 5 cycles of 20" 95°C, 15" 75°C, 45" 70°C, final incubation 3.5' 70°C)). For the fourth PCR, oligonucleotides UNIV5_P4v2 (5'-AATGATACGGCGACCACCGA-3') and UNIV3_P4v2 (5'-CAAGCAGAAGACGGCATACGA-3') were used (2' 95°C, 5 cycles of 20" 95°C, 15" 60°C, 45" 70°C; final incubation of 3.5' at 70°C). The quality of each library was verified using Agilent 2100 BioAnalyzer (mean peak size 640 bp). The samples were indexed and sequenced with 300pb paired-end on MiSeq sequencer (Illumina).

Processing of TCRseq data

The reads were preprocessed with the toolkit pRESTO (2) as follows. Using FilterSeq, reads with a quality > 20 were selected. Using MaskPrimers and PairSeq algorithms, the sequences corresponding to the *TCR*α constant region (AGCAGGTTCTGGGTTCTGGA) or *TCR*β constant region (GGGTGGAGTCACATTTCTCAGAT) and indicating location of the UMI (CTTGGGGG) were searched for and indexed to the head of the paired reads. Using BuildConsensus, consensus-sequences of the reads with the same UMI were constructed.

Next, the forward and reverse reads were aligned to assemble the *Tcra* and the *Tcrb* sequences (AssemblePairs) and the UMI groups containing at least two reads were selected. The sequenced fragments from each selected UMI were aligned to the *Tcra* or *Tcrb* genomic region using the toolkit MiXCR (3) with the tools "align" and "assemble". The aligned fragments were exported as "clonotype-tables" using the tool "exportClones".

Using VDJtools (4) and customized R-scripts these clonotype-tables were then processed and graphs generated. The V-J-usage plots were generated using PlotFancyVJUsage command. The CDR3 lengths were calculated based on the information in the clonotype-table. The numbers of "P" and "N" nucleotides were extracted from the clonotype-tables using the command "exportClones VGeneWithP". Rarefaction plots and diversity statistics were calculated and graphed using the R package "iNEXT". All custom scripts used are available at: https://github.com/arielgalindoalbarran/NOD_TCR_repertoire.git.

Data generated during the current study are available in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/gds) under accession number GSE159001

References

1. Feng Y, van der Veeken J, Shugay M, Putintseva E V, Osmanbeyoglu HU, Dikiy S, et al. A mechanism for expansion of regulatory T-cell repertoire and its role in self-tolerance. Nature. 2015;528(7580):132–6.
2. Vander Heiden JA, Yaari G, Uduman M, Stern JNH, O'connor KC, Hafler DA, et al. PRESTO: A toolkit for processing high-throughput sequencing raw reads of lymphocyte receptor repertoires. Bioinformatics. 2014;30(13).
3. Bolotin DA, Poslavsky S, Mitrophanov I, Shugay M, Mamedov IZ, Putintseva E V., et al. MiXCR: Software for comprehensive adaptive immunity profiling. Nat Methods. 2015;12(5):380–1.
4. Shugay M, Bagaev D V., Turchaninova MA, Bolotin DA, Britanova O V., Putintseva E V., et al. VDJtools: Unifying Post-analysis of T Cell Receptor Repertoires. PLoS Comput Biol. 2015;11(11):1–16.