Identification of disease-associated traits and clonotypes in the T-cell receptor repertoire of monozygotic twins affected by inflammatory bowel diseases

Elisa Rosati[1], Mikhail V. Pogorelyy[2,8], C. Marie Dowds[1], Frederik T. Moller[3,4], Signe B. Sorensen[5,6], Yuri B. Lebedev[2], Norbert Frey[7], Stefan Schreiber[8], Martina E. Spehlman[7#], Vibeke Andersen[5,6,9#], Ilgar Z. Mamedov[2,10,11,12#] and Andre Franke[1*#]

[1]Institute of Clinical Molecular Biology, Kiel University, Kiel, Germany
[2]Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russian Federation
[3]Department of Infectious Disease Epidemiology and Prevention, Statens Serum Institut, Copenhagen, Denmark
[4]Department of Epidemiology Research, Statens Serum Institut, Copenhagen, Denmark
[5]Focused Research Unit for Molecular Diagnostic and Clinical Research, University Hospital of Southern Denmark, Aabenraa, Denmark
[6]Institute of Molecular Medicine, University of Southern Denmark, Odense, Denmark
[7]Department of Internal Medicine III, University Hospital S.-H., Kiel, Germany,
[8]Department of Internal Medicine I, University Hospital S.-H., Kiel, Germany,
[9]IRS-Center Sønderjylland, University of Southern Denmark, Odense, Denmark
[10]Pirogov Russian National Research Medical University (RNRMU)
[11]Dmitry Rogachev National Research Center of Pediatric Hematology, Oncology and Immunology
[12]CEITEC, Masaryk university, Brno, Czech Republic

#these authors contributed equally

*correspondence to:
Prof. Andre Franke
Institute of Clinical Molecular Biology
Christian-Albrechts-University of Kiel
Rosalind-Franklin-Str. 12
D- 24105 Kiel
Germany

E-mail: a.franke@mucosa.de
Tel.: +49 179 485 1891

Short title: TCR repertoire analysis in IBD monozygotic twins

Estimation of Mucosal associated invariant T (MAIT) and Natural Killer T cells (NKT) CDR3α frequency

In this study we used MAIT and NKT invariant α chains to track frequency of these subsets in bulk repertoire sequencing (RepSeq) data. Because it is possible that conventional T cells recombine the same invariant TCRα chain by chance and thus confound the analysis, we first estimated the probability of each clonotype to originate from a MAIT or NKT cell. The evaluation was performed given the MAIT or NKT invariant TCRα chain recombined by chance using the Bayes theorem, P(MAIT | αM) and P(NKT | αN), respectively.

αM and αN are sets of known invariant alpha chains for MAIT and invariant NKT (iNKT). The probability of a MAIT/NKT cell using one of these invariant chains is

$$P(\alpha M \mid MAIT)=P(\alpha N \mid NKT) = 1$$

For a conventional T cell (Tconv), the probability of recombining any TCRα chain from αM and αN is given by the sum of the recombination probabilities for the TCRα chains in these respective sets, which was estimated using OLGA tool[35]:

$$P(\alpha M \mid Tconv)=4.4e^{-4}$$

$$P(\alpha N \mid Tconv)= 5.8e^{-6}$$

For P(MAIT) and P(NKT), the proportion of these cell subsets in the blood, we used the average frequency of these subsets in the Caucasian population:

$$P(MAIT)=3\%$$ [36]

$$P(NKT)=0.076\%$$ [37]

We calculated the probability that, when detecting one of these invariant TCRα sequences, it actually originates from a MAIT cell as:

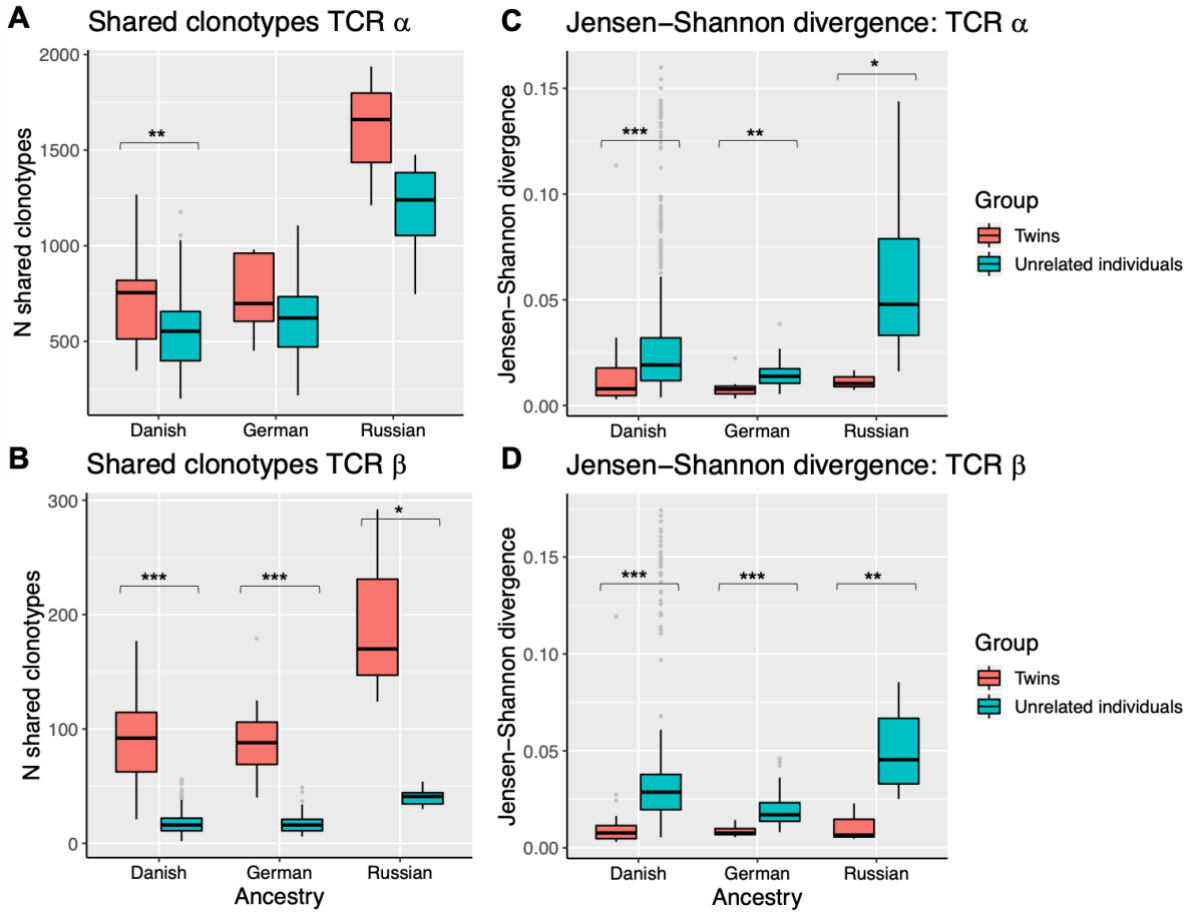
$$P(MAIT \mid \alpha M) =P(\alpha M \mid MAIT) \cdot P(MAIT) / (P(\alpha M \mid MAIT) \cdot P(MAIT)$$

$$+ P(\alpha M \mid Tconv) \cdot (1-P(MAIT)))=98.6\%$$

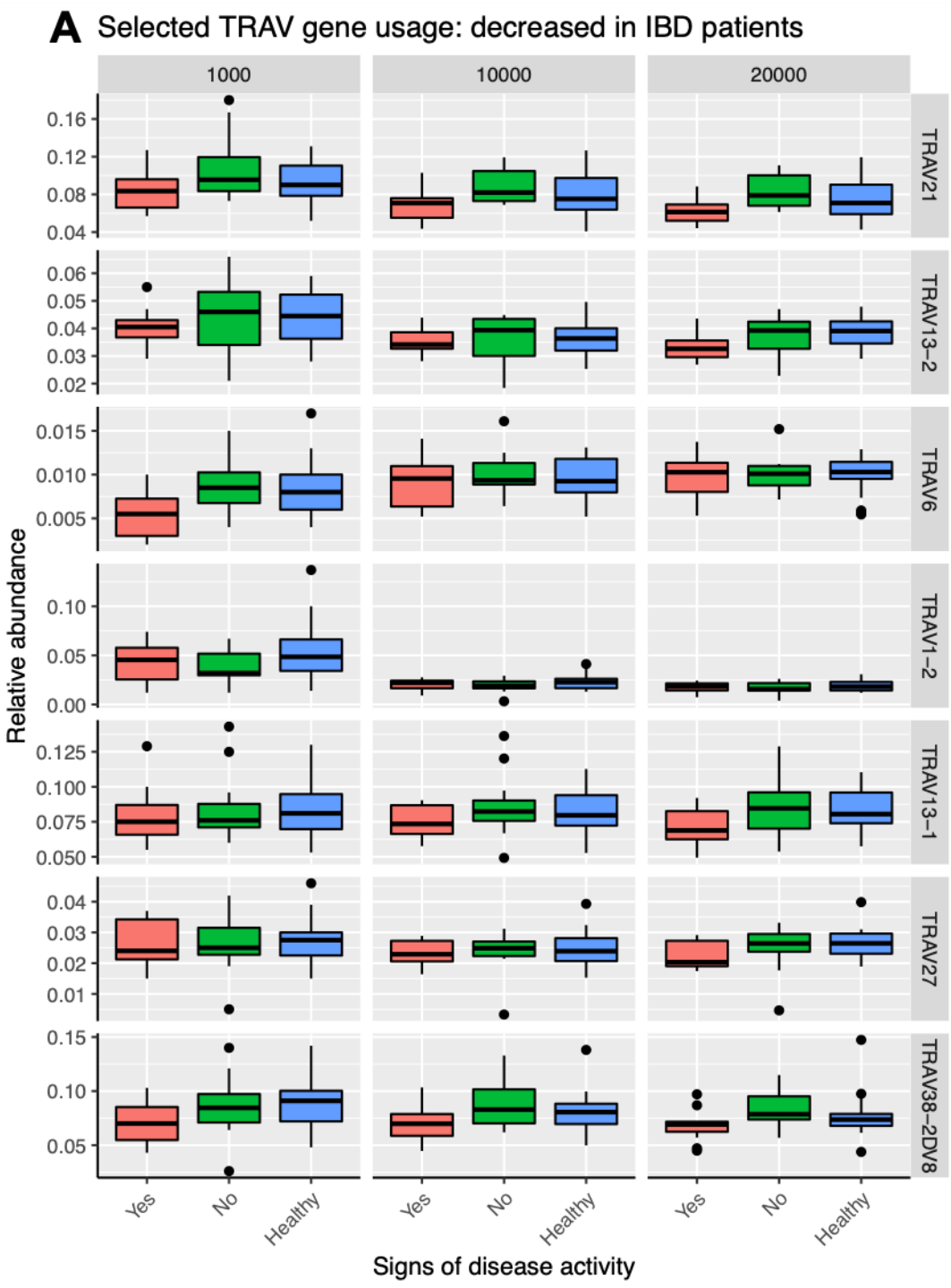
In the same way for NKT cells:

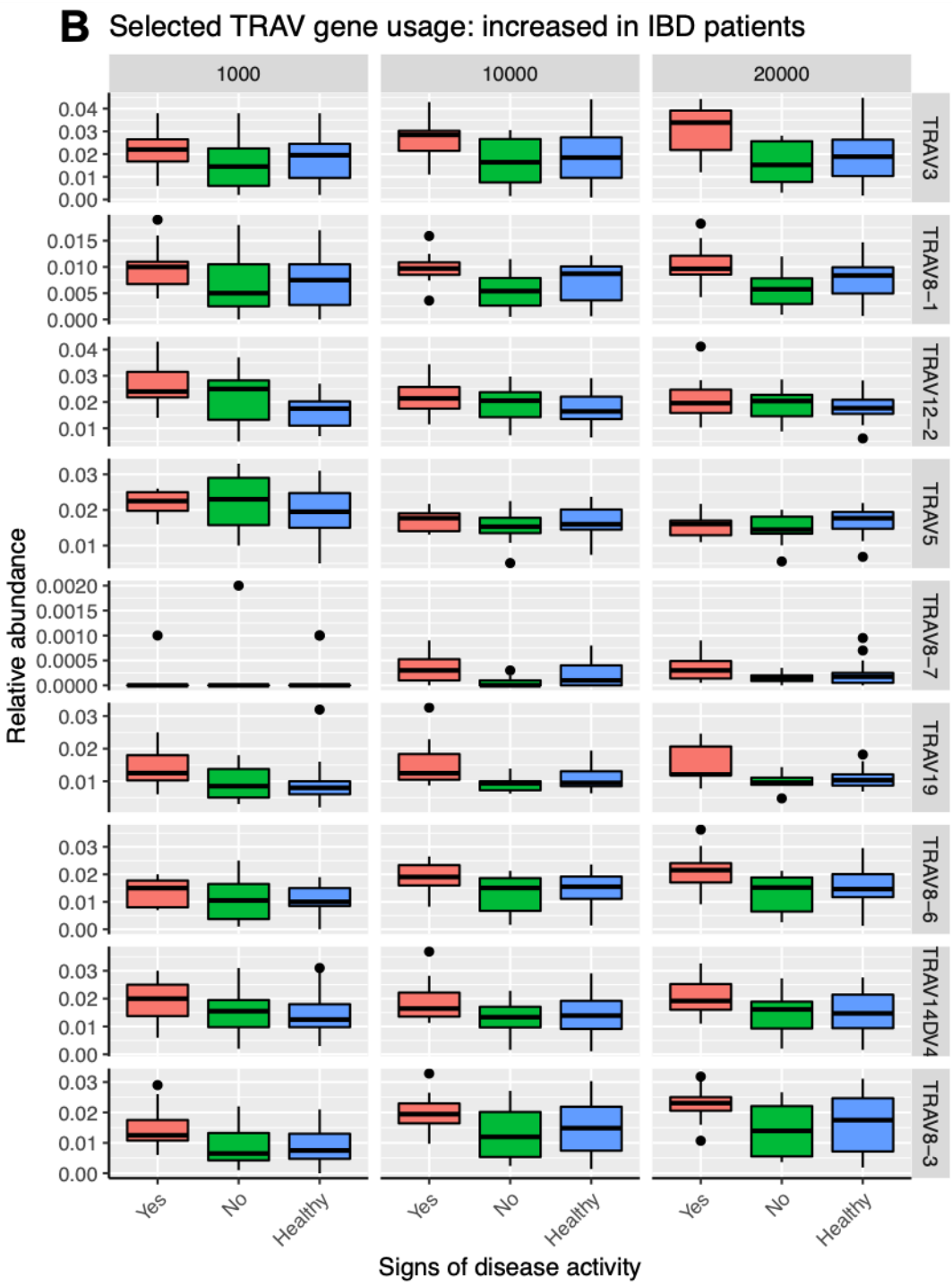
$$P(NKT \mid \alpha N)=99.2\%$$

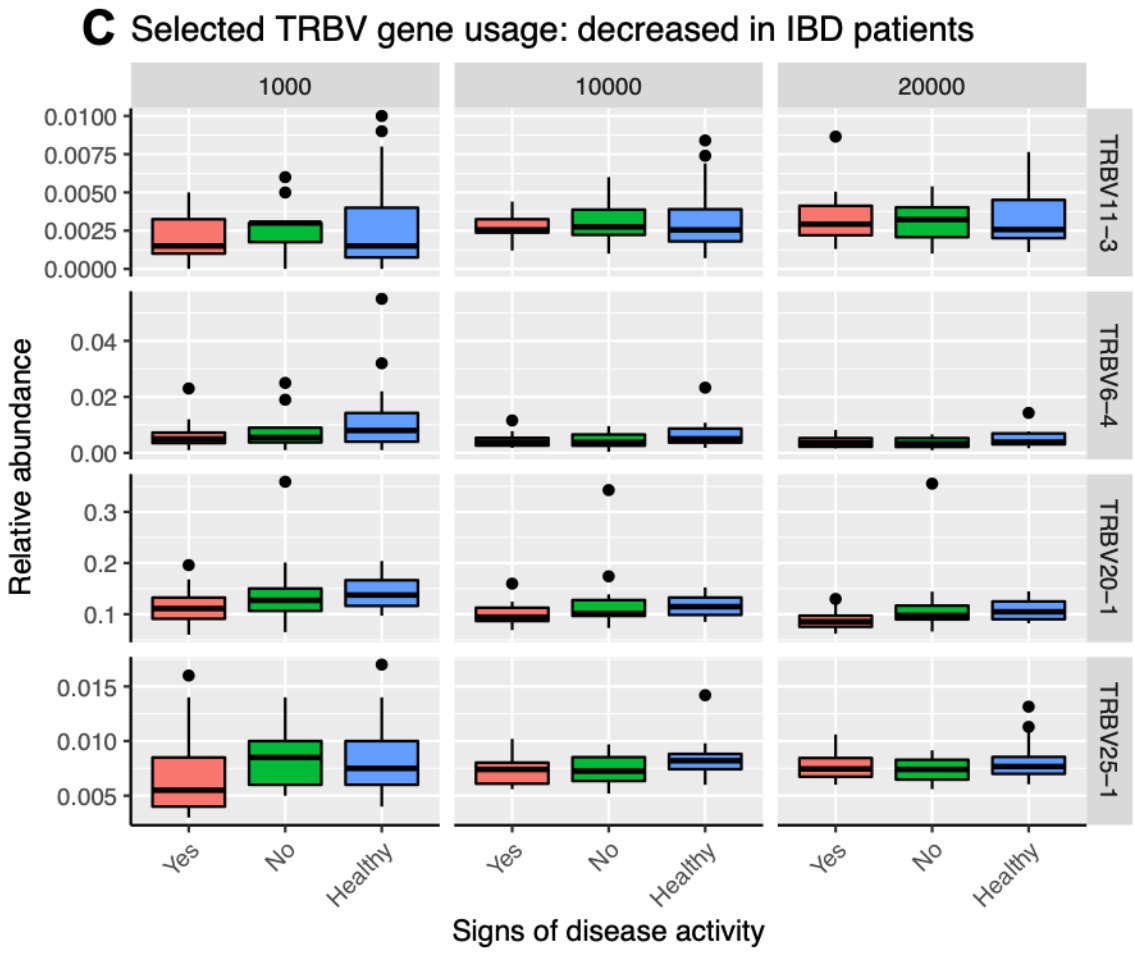
Therefore, if a clonotype carries one of these invariant TCRα chains, the probability that this clonotype is a conventional T cell is lower than 1.5% both for NKT and MAIT TCRα sequences. For analysis of MAIT β TCRs, we employed TCR sequences published in literature, particularly from Howson *et al*.[38] Additionally, we used in house produced single-cell data to match known MAIT α TCRs with unknown MAIT β TCRs. A list of all used sequences is available as **Supplementary data 2**.

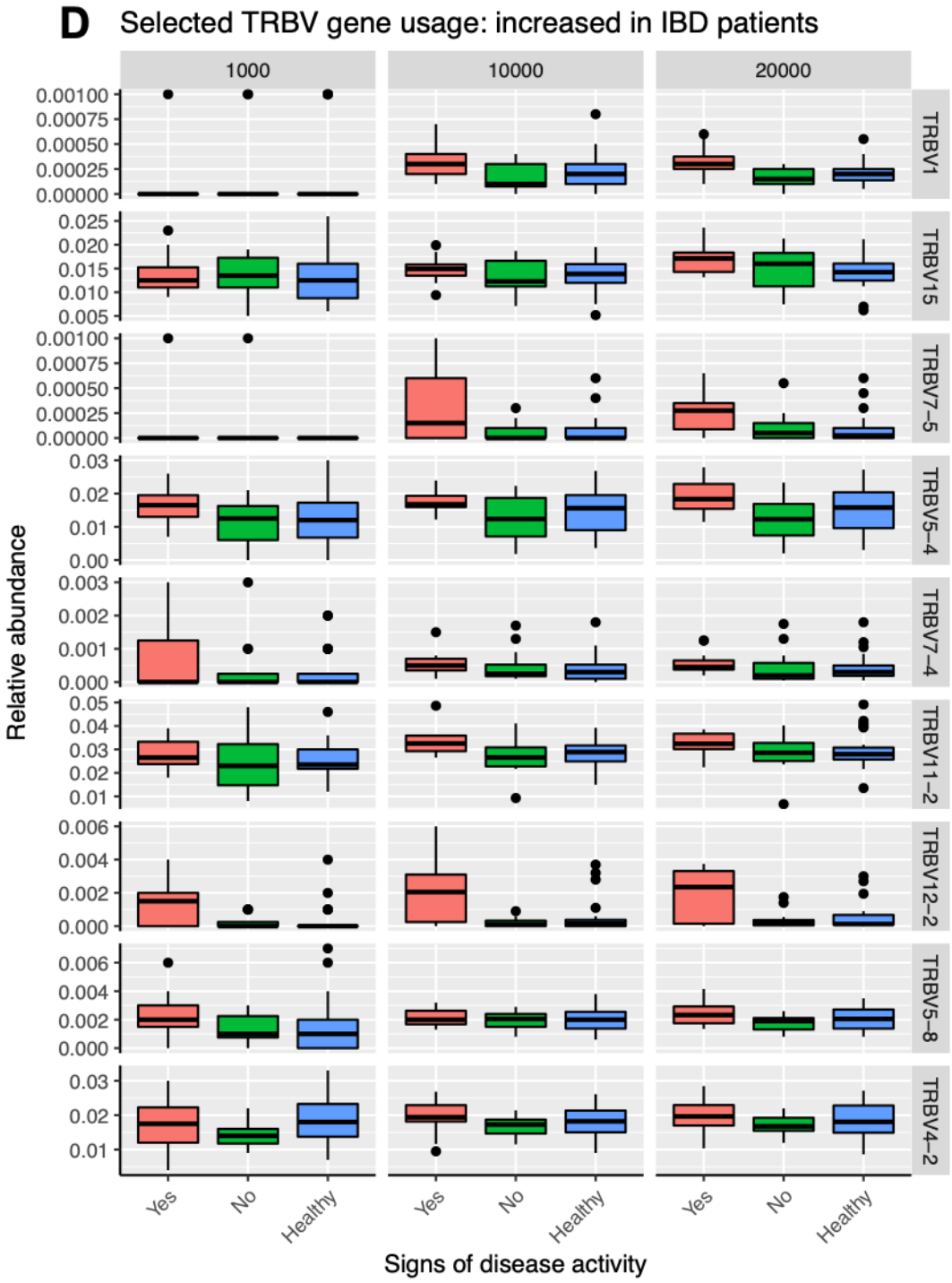


Supplementary figure 1: Twin specific repertoire features in the 10,000 most abundant clonotypes. Pairwise number of shared clonotypes for **[A]** TCRα [Danish p = 0.0038, German p = 0.07, Russian p = 0.1] and **[B]** TCRβ [Danish $p = 9.6 \times 10^{-5}$, German $p = 6 \times 10^{-4}$, Russian p = 0.009]. Pairwise Jensen-Shannon divergence for **[C]** TCRα [Danish $p = 5.7 \times 10^{-13}$, German $p = 5.4 \times 10^{-7}$, Russian p = 0.01] and **[D]** TCRβ [Danish $p = 7 \times 10^{-9}$, German $p = 6.6 \times 10^{-6}$, Russian p = 0.004] V gene usage. (*) p-value < 0.05, (**) p-value < 0.005, (***) p-value < 0.0005.

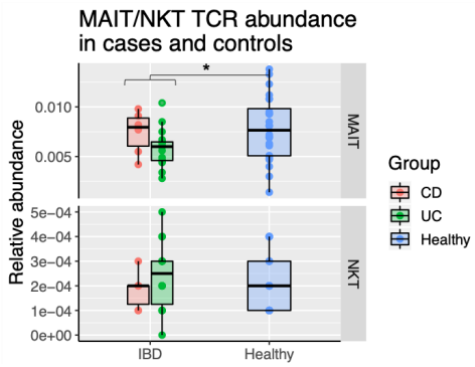

A
Selected TRAV gene usage: decreased in IBD patients
1000
10000
20000
TRAV21
TRAV13–2
TRAV6
TRAV1–2
TRAV13–1
TRAV27
TRAV38–2DV8
Relative abundance
Yes
No
Healthy
Signs of disease activity


B Selected TRAV gene usage: increased in IBD patients
1000
10000
20000
TRAV3
TRAV8–1
TRAV12–2
TRAV5
TRAV8–7
TRAV19
TRAV8–6
TRAV14DV4
TRAV8–3
Relative abundance
Yes
No
Healthy
Signs of disease activity

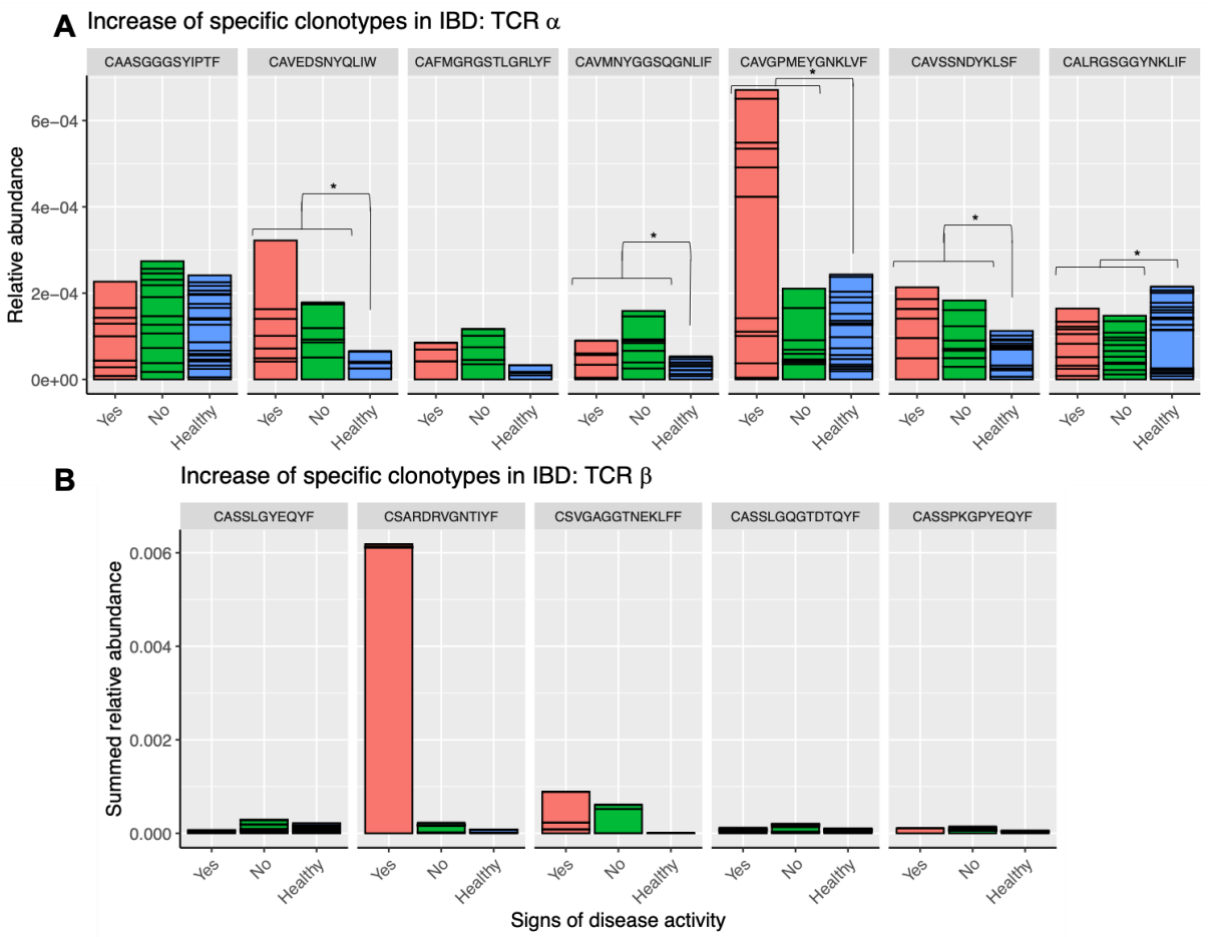

C Selected TRBV gene usage: decreased in IBD patients
1000
10000
20000
TRBV11–3
TRBV6–4
TRBV20–1
TRBV25–1
Relative abundance
0.0100
0.0075
0.0050
0.0025
0.0000
0.04
0.02
0.00
0.3
0.2
0.1
0.015
0.010
0.005
Yes
No
Healthy
Signs of disease activity



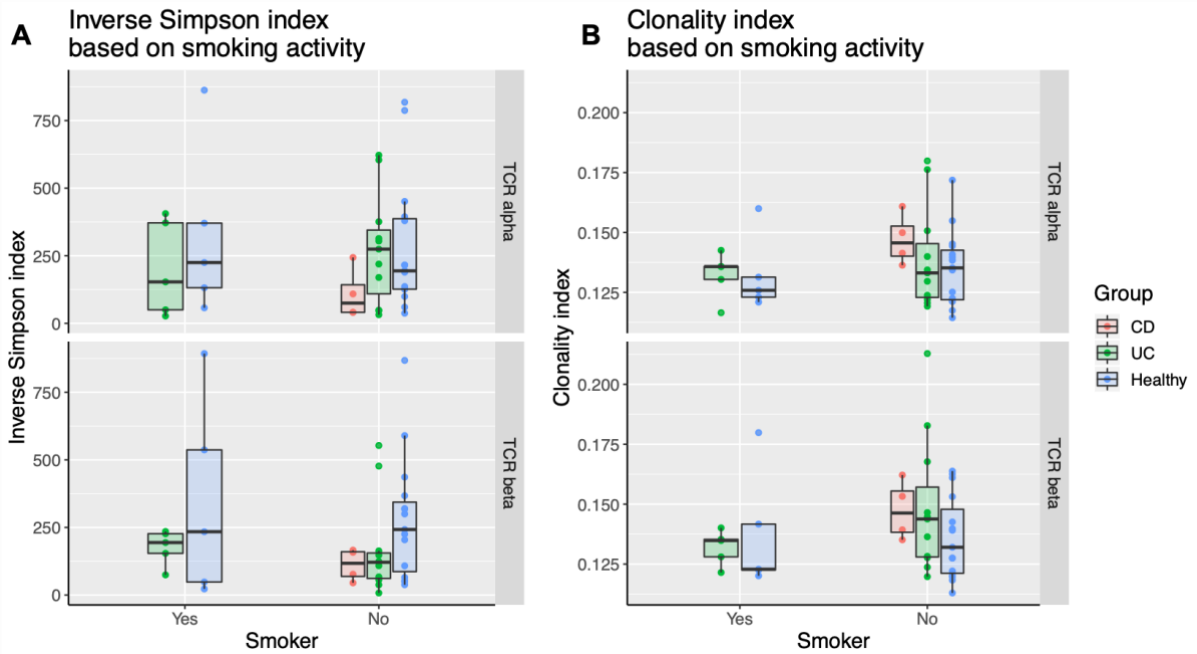
Supplementary Figure 2: Selected V genes differentially used in IBD. The figure shows specific TCR V genes which were showing to be decreased or increased in IBD individuals compared to their healthy twins. The analysis was repeated by looking at the most abundant 1000, 10000, 20000 clonotypes and the reported genes are the ones which were showing differential abundance in at least one of the three analysis. [A] TCRα V genes decreased in IBD and [B] increased in IBD. [C] TCRβ V genes decreased in IBD and [D] increased in IBD.



Supplementary figure 3: MAIT and NKT cell TCRs abundance among the 10,000 most abundant clonotypes. Cumulative abundance of TCR sequences originating from MAIT [IBD-healthy p = 0.02, CD-healthy p = 0.21,UC-healthy p= 0.08, UC-CD p=0.17] and NKT cells [IBD-healthy p = 0.8, CD-healthy p = 0.09,UC-healthy p= 0.19, UC-CD p=0.3] in IBD patients (CD and UC) and healthy individuals. Abundance of MAIT TCRs seems to be decreased in IBD patients. (*) p-value < 0.05.



Supplementary figure 4: Clonotypes which abundance is increased in IBD patients. Plot is divided in patients with active and inactive IBD and their healthy co-twins. [A] TCRα [IBD-healthy: CAASGGGSYIPTF p = 0.06, CAVEDSNYQLIW p = 0.046, CAFMGRGSTLGRLYF p = 0.1, CAVMNYGGSQGNLIF p = 0.046, CAVGPMEYGNKLVF p = 0.046, CAVSSNDYKLSF p = 0.046, CALRGSGGYNKLIF p = 0.046] and [B] TCRβ [IBD-healthy: CSARDRVGNTIYF p = 0.1, CSVGAGGTNEKLFF p = 0.15, CASSLGQGTDTQYF p = 0.1; active-inactive: CASSLGYEQYF p = 0.21, CASSPKGPYEQYF p = 0.21].



Supplementary figure 5: Impact of smoking behaviour on peripheral TCR repertoire diversity. Smokers (yes) and non-smokers (no) are compared as well as healthy individuals and IBD patients. **[A]** Inverse Simpson diversity index for TCRα (top panel) and TCRβ (bottom panel) **[B]** TCR clonality (inverse of Shannon entropy) for TCRα (top panel) and TCRβ (bottom panel). Diversity values calculated on downsampled data. There were no statistically significant differences between the groups.