

A conserved structural organization of the human public TCR β repertoire persists throughout exceptional longevity

Luigi Buonaguro

`l.buonaguro@istitutotumori.na.it`

Istituto Nazionale Tumori IRCCS "Fondazione G. Pascale"

Beatrice Cavalluzzo

Istituto Nazionale Tumori IRCCS "Fondazione G. Pascale"

Angela Mauriello

Istituto Nazionale Tumori IRCCS "Fondazione G. Pascale"

Concetta Ragone

Istituto Nazionale Tumori IRCCS "Fondazione G. Pascale"

Biancamaria Cembrola

Istituto Nazionale Tumori IRCCS "Fondazione G. Pascale"

Simona Mangano

Istituto Nazionale Tumori IRCCS "Fondazione G. Pascale"

Ana Mendizabal-Sasieta

Omniscope, Inc. US

Juan Melero

Omniscope, Inc. US

Marta Soto

Omniscope, Inc. US

Marta Grzelak

Omniscope Inc <https://orcid.org/0000-0002-2653-2895>

Mariana Capelli

Omniscope, Inc. US

Holger Heyn

Centro Nacional de Análisis Genómico (CNAG)

Maria Tagliamonte

Istituto Nazionale Tumori IRCCS "Fondazione G. Pascale"

Keywords:

Posted Date: September 2nd, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10827420/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: There is **NO** Competing Interest.

A conserved structural organization of the human public TCR β repertoire persists throughout exceptional longevity

Beatrice Cavalluzzo¹, Angela Mauriello¹, Concetta Ragone¹, Biancamaria Cembrola¹, Simona Mangano¹, Ana Mendizabal-Sasieta², Juan Luis Melero², Marta Soto², Marta Grzelak², Mariana Capelli², Holger Heyn^{2,3}, Maria Tagliamonte¹, Luigi Buonaguro^{1*}

¹Innovative Immunological Models Unit, National Cancer Institute, IRCCS - "Fond G. Pascale", Naples, Italy;

²Omniscope, Inc., Delaware, USA; ³Centro Nacional de Análisis Genómico (CNAG), Barcelona, Spain

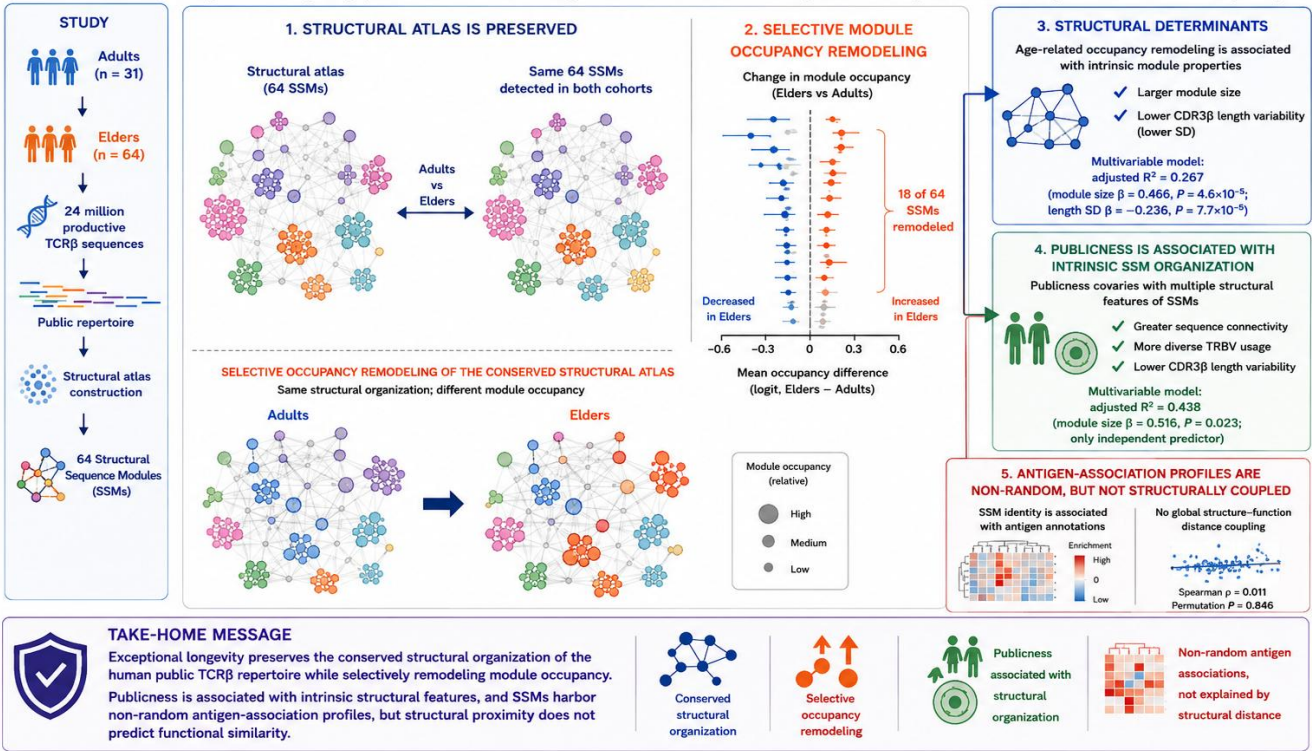
*** Corresponding Author:**

Luigi Buonaguro,
Lab of Innovative Immunological Models,
Istituto Nazionale per lo Studio e la Cura dei Tumori,
"Fondazione Pascale" - IRCCS,
Via Mariano Semmola, 52
Tel.: +39-081-1777 0587;
l.buonaguro@istitutotumori.na.it

Graphical Abstract

A conserved structural organization underlies the human public TCR β repertoire throughout exceptional longevity

Exceptional longevity preserves structural organization while selectively remodeling Structural Sequence Module occupancy.



Abstract

The human public T-cell receptor (TCR) repertoire provides a shared immunological framework across individuals, yet whether it possesses a conserved higher-order structural organization and how this organization is affected by lifelong immune ageing remain unknown.

Here, we constructed a structural atlas of the human public TCR β repertoire from 1.33 million unique public CDR3 β sequences derived from healthy Adults (<70 years) and Elders (>80 years). The atlas comprised 64 reproducible Structural Sequence Modules (SSMs) representing sequence-coherent neighborhoods supported by quantitative model selection and independent topological validation. Projection of individual repertoires onto this common framework demonstrated complete preservation of all SSMs across both cohorts. Although 18 modules exhibited significant differences in unweighted occupancy, the global module hierarchy remained highly conserved, indicating that exceptional longevity is associated primarily with quantitative redistribution among pre-existing structural compartments rather than loss or replacement of repertoire organization. Rarefaction analysis demonstrated that unequal cohort size strongly influenced public-clonotype discovery but had minimal impact on the direction and magnitude of module-level occupancy differences, whereas frequency-weighted analysis indicated that remodeling was not primarily driven by expansion of highly abundant public clonotypes. Susceptibility to remodeling was associated mainly with module size and within-module CDR3 β length variability. Module publicness was highly conserved across age groups and associated with coordinated sequence, germline and network properties, while TRBJ usage was substantially more strongly associated with SSM identity than TRBV usage. Finally, curated antigen annotations were non-randomly distributed across SSMs, demonstrating module-level functional biases, although structural proximity did not generally predict similarity in antigen-recognition profiles.

Together, these findings identify higher-order structural organization as a robust property of the human public TCR β repertoire. Exceptional longevity preserves this organization while selectively remodeling the occupancy of pre-existing Structural Sequence Modules, distinguishing age-associated quantitative redistribution from cohort-size effects and clonal expansion. The structural atlas therefore provides a reproducible framework for investigating the organization and remodeling of adaptive immune repertoires across ageing and immune-mediated conditions.

Introduction

The adaptive immune system maintains immune memory for decades despite continuous antigenic exposure, recurrent infections and lifelong turnover of T-cell populations. Ageing progressively reshapes the T-cell compartment through thymic involution, contraction of the naïve pool, expansion of antigen-experienced clones and changes in T-cell receptor (TCR) diversity (1-4). These processes contribute to immunosenescence and influence susceptibility to infection, vaccine responsiveness and age-associated disease (3, 5). Yet whether lifelong immune remodeling also alters the higher-order organization of the TCR repertoire remains poorly understood.

The public TCR repertoire, defined by clonotypes shared across unrelated individuals, provides a reproducible window into this organization. Public clonotypes arise through convergent V(D)J recombination, non-random generation probabilities, thymic selection and recurrent exposure to common antigens (6 – 9). Most repertoire studies, however, have characterized public TCRs primarily through clonotype sharing, frequency, diversity, clonality or germline-gene usage. Although these metrics describe repertoire composition, they provide limited information about how structurally related receptors are organized within sequence space (10, 11).

TCR sequence space is unlikely to be randomly populated (12). Receptors can converge toward related CDR3 sequence features despite distinct nucleotide origins and germline configurations (13, 14), and recent computational approaches indicate that TCR sequences can occupy biologically meaningful neighborhoods in sequence or representation space (12, 15 – 17). Sequence-similarity networks, embedding approaches and repertoire-level functional units therefore support the possibility that adaptive immune repertoires possess organizational structure beyond individual clonotypes. Whether such higher-order organization exists reproducibly within the human public TCR repertoire, however, and whether it remains stable throughout physiological ageing are unresolved questions.

Exceptional longevity provides a particularly stringent setting in which to examine these questions. Individuals reaching advanced age have experienced decades of antigenic stimulation and immune remodeling while retaining sufficient immune competence for long-term survival (1, 18). If the public TCR repertoire possesses a stable higher-order organization, its principal structural components might remain detectable despite extensive lifelong remodeling, with ageing affecting their relative utilization rather than requiring wholesale reorganization. Conversely, substantial alteration of the underlying architecture would be expected to affect the composition, representation or relationships of its structural components.

Here, we investigated whether the human public TCR β repertoire can be represented within a reproducible structural atlas and how this organization differs between Adults and exceptionally long-lived Elders. We constructed and independently validated a common atlas of Structural Sequence Modules (SSMs) and projected individual public repertoires onto this fixed reference to quantify module occupancy. We then assessed the robustness of age-associated differences to cohort-size imbalance and clonotype abundance, examined the intrinsic module properties associated with occupancy remodeling and population publicness, and used curated antigen-specific TCR annotations to determine whether functional associations are non-randomly distributed across the structural atlas and whether structural proximity is related to functional similarity.

Our findings reveal a highly conserved higher-order organization of the human public TCR β repertoire across age groups. Exceptional longevity is associated predominantly with quantitative redistribution among pre-existing structural modules rather than loss or replacement of the atlas, while individual modules exhibit reproducible structural properties and non-random antigen associations. These results establish Structural Sequence Modules as an intermediate organizational level linking individual public clonotypes to repertoire-wide architecture and provide a framework for investigating structural remodeling across ageing and other immune contexts.

Materials and Methods

1. Study population and sample processing

Peripheral blood samples were obtained from healthy younger adults (Adults, <70 years) and exceptionally long-lived individuals (Elders, >80 years). The final analysis included 31 Adults and 64 Elders. All participants provided written informed consent prior to enrollment, and the study was conducted in accordance with the Declaration of Helsinki following approval by Institutional Ethics Committee. Demographic and clinical characteristics are summarized in Supplementary Table 1.

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood by density-gradient centrifugation, cryopreserved in fetal bovine serum containing dimethyl sulfoxide, and stored in liquid nitrogen until analysis. Frozen PBMCs were rapidly thawed at 37°C, washed in complete culture medium and phosphate-buffered saline, and processed for RNA extraction. Total RNA was isolated using the RNeasy Plus Micro Kit (QIAGEN, Hilden, Germany) on the QIAcube Connect automated platform according to the manufacturer's instructions. RNA quantity and integrity were assessed before library preparation.

2. TCR β repertoire sequencing and public-clonotype definition

TCR β sequencing libraries were generated using the Omniscope single-molecule TCR β assay (osTCR; Omniscope Inc., Wilmington, DE, USA). Raw reads were processed using the proprietary Omniscope analysis pipeline to reconstruct complete TCR β rearrangements, assign TRBV and TRBJ germline genes, identify CDR3 nucleotide and amino-acid sequences, and quantify clonotype abundance.

Productive TCR β clonotypes were defined by the unique combination of CDR3 β amino-acid sequence, TRBV gene and TRBJ gene. Public clonotypes were identified independently within each cohort as clonotypes detected in at least two unrelated individuals. Adult and Elder public repertoires were then combined to generate the public-union dataset used for atlas construction.

Because multiple TRBV/TRBJ configurations could encode the same CDR3 β amino-acid sequence, structural sequence-space analyses were performed on unique CDR3 β amino-acid sequences after public-clonotype identification. The union comprised 1,332,323 unique public CDR3 β sequences, including 350,884 found in both cohort-specific public repertoires, 260,179 detected only in Adults and 721,260 detected only in Elders. Because classification as public or cohort-exclusive depends on sample size, these counts were considered descriptive and their sensitivity to cohort-size imbalance was assessed by rarefaction.

3. Construction and validation of the structural atlas

The structural atlas was constructed from the union of unique public CDR3 β amino-acid sequences. To reduce the contribution of relatively conserved terminal residues, three amino acids were removed from each terminus and the resulting central CDR3 β regions were decomposed into overlapping 3-, 4- and 5-mers. K-mer composition was represented using a term-frequency/inverse-document-frequency (TF-IDF) matrix with sublinear term-frequency scaling and L2 normalization; k-mers occurring in fewer than two sequences were excluded. The sparse feature matrix was projected into a 30-dimensional latent sequence space by truncated singular-value decomposition (SVD),

followed by L2 normalization. Similarity in clustering space therefore reflected the angular geometry of normalized latent k-mer representations rather than raw amino-acid identity.

Sequences were partitioned using MiniBatch k-means clustering with a fixed random seed and repeated centroid initializations. Atlas resolution was evaluated by comparing 32-, 64- and 96-module solutions generated using identical preprocessing. Cluster separation was quantified by cosine silhouette score and reproducibility across independent initializations by adjusted Rand index (ARI). Mean silhouette scores were 0.383, 0.389 and 0.356 for 32, 64 and 96 modules, respectively, with corresponding mean ARI values of 0.695, 0.766 and 0.758. The 64-module solution was therefore selected as the best-supported compromise between separation, reproducibility and atlas resolution and defined the 64 Structural Sequence Modules (SSMs) used for subsequent analyses.

For visualization, module centroids in normalized 30-dimensional latent space were projected into two dimensions using principal component analysis (PCA). PCA was used exclusively for graphical representation and did not contribute to clustering or statistical testing.

Module cohesion was independently evaluated using edit-distance sequence networks. Within each SSM, CDR3 β sequences were connected at Levenshtein distance ≤ 2 , permitting substitutions, insertions and deletions. Cohesion was quantified using the fraction of connected sequence pairs and nearest-neighbor sequence distance and compared with size-matched sequence sets sampled from the complete public repertoire. A substitution-only Hamming analysis restricted to equal-length sequence pairs was performed as a sensitivity analysis. Greater-than-random internal connectivity was observed for all 64 SSMs under both definitions.

4. Subject-level projection and module occupancy analyses

Subject-specific public repertoires were projected onto the fixed 64-SSM atlas according to CDR3 β amino-acid sequence. Unweighted occupancy of each SSM was defined as the number of unique public CDR3 β sequences assigned to that module divided by the total number of unique public CDR3 β sequences in the individual's public repertoire. Adult–Elder differences were evaluated using two-sided Wilcoxon rank-sum tests with Benjamini–Hochberg correction across the 64 modules; signed effect sizes were retained to characterize the direction and magnitude of remodeling.

Global occupancy architecture was assessed from normalized 64-dimensional individual occupancy profiles using pairwise Jensen–Shannon distances. Principal coordinate analysis (PCoA) was used for visualization, and cohort differences were tested by PERMANOVA with 9,999 label permutations. Subject-level architecture was additionally characterized by the number of occupied SSMs, Shannon occupancy diversity, Pielou evenness and maximum single-module occupancy. Cohort differences were tested using Wilcoxon rank-sum tests followed by Benjamini–Hochberg correction. Conservation of the module-occupancy hierarchy was evaluated by Spearman correlation between mean Adult and Elder occupancy across the 64 SSMs.

5. Cohort-size and frequency-weighted sensitivity analyses

To determine whether occupancy differences were influenced by clonal expansion, a complementary frequency-weighted analysis was performed in which each public clonotype contributed according to its observed repertoire frequency. Weighted occupancy was defined as the fraction of total public-

clonotype abundance assigned to each SSM, and Adult–Elder comparisons used the same statistical framework as the primary unweighted analysis.

The influence of unequal cohort size was assessed by repeated rarefaction. In each of 1,000 iterations, 31 of the 64 Elders were sampled without replacement to match the Adult cohort size, after which public clonotypes were re-identified de novo using the primary definition of detection in at least two unrelated individuals. Public-sequence discovery and SSM occupancy were recalculated for each balanced 31-versus-31 comparison. Robustness was assessed by comparing full-cohort Elder-minus-Adult effects with the distributions and mean effects obtained after rarefaction and by evaluating preservation of effect direction across all 64 SSMs.

To assess the potential influence of HLA background on age-associated occupancy differences, analyses were repeated in individuals carrying HLA-A02:01. The HLA-A02:01-positive subset comprised 10 Adults and 22 Elders. Module-level Elder-minus-Adult occupancy effects were recalculated across all 64 SSMs and compared with full-cohort effects using Spearman rank correlation. Directional concordance was additionally evaluated for the 18 SSMs identified as significantly remodeled in the primary analysis. Global occupancy architecture within the HLA-A*02:01-positive subset was assessed using Jensen–Shannon distances, PCoA and PERMANOVA, and subject-level Shannon diversity, Pielou evenness and maximum module occupancy were compared between cohorts using two-sided Wilcoxon rank-sum tests.

6. Module publicness, structural descriptors and remodeling analyses

Within each cohort, module publicness was defined as the mean, across public CDR3 β sequences assigned to an SSM, of the fraction of cohort individuals carrying each sequence. Publicness was normalized by the number of individuals in the corresponding cohort (31 Adults or 64 Elders), permitting comparison despite unequal cohort sizes. Mean normalized module publicness was calculated across cohorts, and conservation of relative publicness between Adults and Elders was evaluated by Spearman rank correlation.

Each SSM was characterized using complementary descriptors including module size, CDR3 β length statistics and length entropy, TRBV and TRBJ diversity, dominant germline-gene fractions, and Levenshtein-network properties including network density and nearest-neighbor sequence distance. Germline diversity was quantified by Shannon entropy. Associations of these descriptors with module publicness and the absolute magnitude of Adult–Elder occupancy remodeling were evaluated using Spearman correlations with Benjamini–Hochberg correction. Associations between SSM identity and TRBV or TRBJ usage were evaluated using contingency tables and summarized by Cramér's V.

Because candidate descriptors were correlated, multivariable analyses used Elastic Net regularization for feature selection. Predictors and outcomes were standardized before fitting. ElasticNetCV used eight-fold cross-validation across l1_ratio values of 0.10, 0.25, 0.50, 0.75, 0.90, 0.95 and 1.00 and 100 candidate regularization strengths, with random seed 20260825 and a maximum of 200,000 iterations. Predictors with non-zero coefficients were subsequently entered into ordinary least-squares models with HC3 heteroscedasticity-consistent standard errors; model fit was summarized by adjusted R². The framework was applied separately to remodeling magnitude and mean normalized publicness. Cross-validation selected l1_ratio = 1.0 for remodeling and l1_ratio = 0.10 for publicness.

7. Antigen-specific annotation and functional analyses

Functional annotation was performed using curated antigen-specific TCR β records from VDJdb. Public sequences were considered annotated when CDR3 β amino-acid sequence, TRBV gene and TRBJ gene matched the corresponding VDJdb record after removal of allelic suffixes. Duplicate records representing the same CDR3 β –annotation combination were collapsed. Under this stringent definition, 11,065 unique public CDR3 β sequences received annotations, spanning all 64 SSMs and 917 distinct exact epitopes. Functional profiles were constructed at the exact-epitope, antigen-gene and antigen-species levels.

Non-random distribution of functional annotations across SSMs was evaluated using contingency tables and summarized by Cramér's V. For antigen-species profiles, standardized Pearson residuals were used to identify individual SSM–antigen associations, with cell-wise tests corrected by Benjamini–Hochberg FDR. Functional richness, Shannon diversity and dominance were calculated for each SSM and compared among Adult-increased, Elder-increased and stable modules using Kruskal–Wallis tests.

To assess structure–function relationships, pairwise cosine distances between SSM centroids in normalized 30-dimensional structural space were compared with Jensen–Shannon distances between normalized functional profiles. Analyses were performed independently at the exact-epitope, antigen-gene and antigen-species levels. Associations between structural and functional distance matrices were quantified using Spearman correlation across all unique SSM pairs, with significance evaluated by 9,999 permutations of functional-profile identities relative to the fixed structural-distance matrix.

8. Statistical analysis

Unless otherwise specified, statistical tests were two-sided. Adult–Elder comparisons used Wilcoxon rank-sum tests, associations between continuous module-level variables used Spearman rank correlation, and comparisons involving more than two module categories used Kruskal–Wallis tests. Multiple testing was controlled by the Benjamini–Hochberg procedure, with FDR-adjusted $P < 0.05$ considered significant. Multivariate occupancy differences were assessed by PERMANOVA with 9,999 permutations, and structure–function distance associations by 9,999 module-label permutations. Multivariable models used standardized variables, Elastic Net feature selection and OLS refitting with HC3 heteroscedasticity-consistent standard errors. A fixed random seed was used for stochastic computational procedures where applicable.

Results

Construction and validation of the public TCR β structural atlas

To establish a common structural framework for subsequent analyses, we reconstructed the public TCR β repertoire by integrating public clonotypes identified independently in 31 Adults (<70 years) and 64 Elders (>80 years). The resulting union comprised 1,332,323 unique public CDR3 β amino-acid sequences. Central CDR3 β regions were represented using overlapping 3–5 amino-acid k-mers, transformed by TF-IDF, projected into a 30-dimensional latent sequence space by truncated SVD and partitioned using MiniBatch k-means (Fig. 1A).

We next evaluated atlas resolution by comparing clustering solutions containing 32, 64 and 96 modules. The 64-module solution provided the best-supported combination of sequence-space separation and clustering reproducibility, with a cosine silhouette score of 0.389 and mean adjusted Rand index (ARI) of 0.766, compared with 0.383/0.695 for 32 modules and 0.356/0.758 for 96 modules (Fig. 1B). The 64-module solution was therefore adopted as the fixed structural atlas for all subsequent analyses.

Projection of module centroids into two-dimensional principal-component space revealed a distributed organization of the 64 Structural Sequence Modules (SSMs) within the common latent sequence space (Fig. 1C). Module sizes were markedly heterogeneous and followed a long-tailed distribution, with a limited number of large SSMs accounting for a substantial fraction of the public repertoire and progressively smaller modules representing the remaining sequence diversity (Fig. 1D).

We independently tested whether SSMs represented sequence-coherent neighborhoods rather than partitions generated solely by the clustering algorithm. Under both Levenshtein distance ≤ 2 and a substitution-only Hamming criterion restricted to equal-length sequences, all 64 SSMs exhibited greater internal connectivity than size-matched randomized controls (Fig. 1E). Germline architecture also showed an asymmetric relationship with module identity: TRBV usage was only weakly associated with SSM assignment (Cramér's $V = 0.046$), whereas TRBJ usage showed a substantially stronger association ($V = 0.663$) (Fig. 1F).

Together, these analyses establish a reproducible 64-module structural atlas of the human public TCR β repertoire supported by quantitative model selection and independent sequence-topology validation.

A

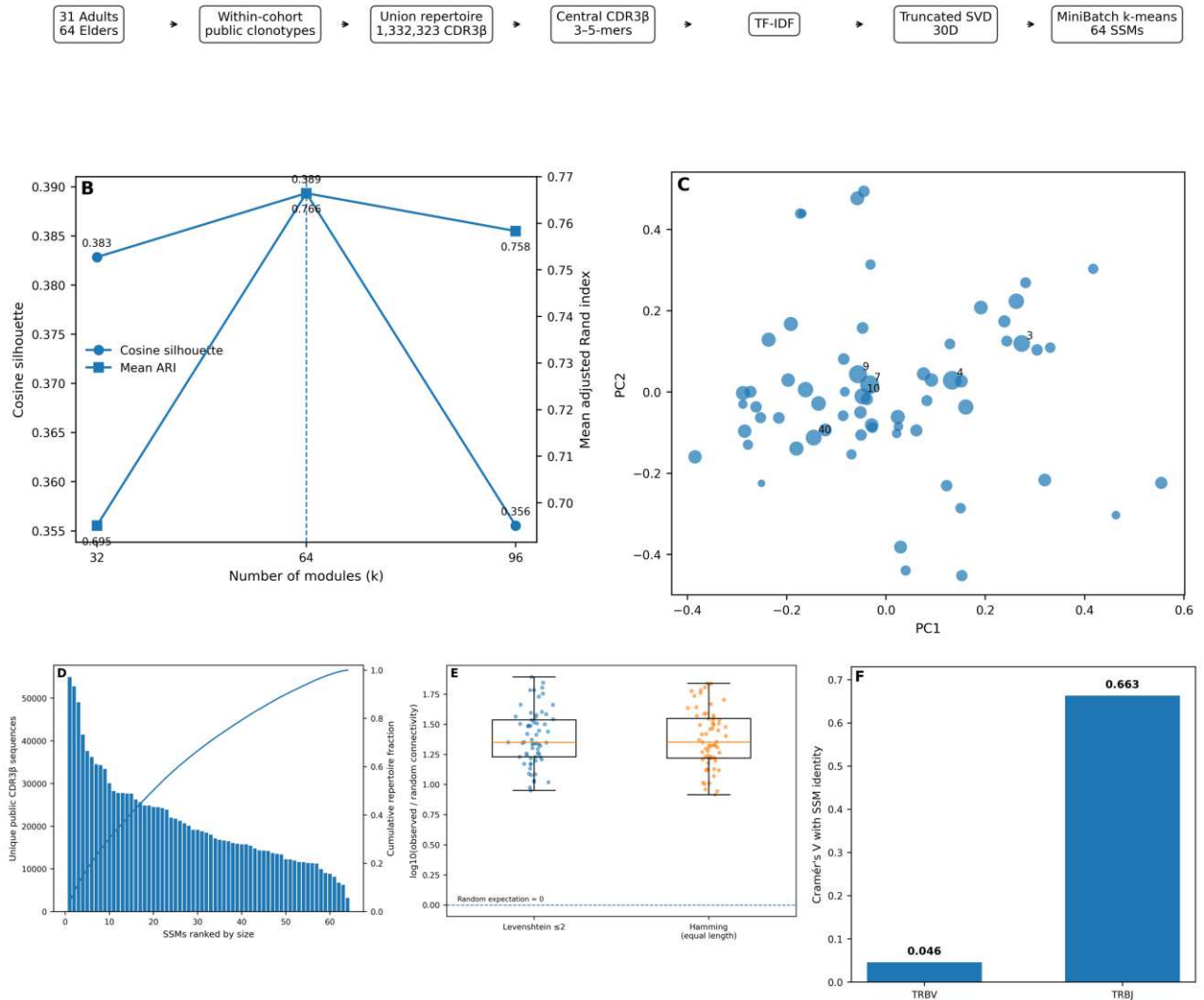


Figure 1 | Construction and quantitative validation of the structural atlas of the human public TCRβ repertoire.

(A) Atlas-construction workflow. Public clonotypes identified independently in 31 Adults and 64 Elders were combined into a union of 1,332,323 unique public CDR3β sequences. Central CDR3β 3–5-mer composition was represented by TF-IDF, projected into a 30-dimensional latent space by truncated SVD and partitioned by MiniBatch k-means into 64 Structural Sequence Modules (SSMs). (B) Selection of atlas resolution. Solutions with 32, 64 and 96 modules were compared using cosine silhouette score and mean adjusted Rand index (ARI) across independent initializations; k=64 provided the best-supported combination of separation and reproducibility. (C) PCA projection of the 64 SSM centroids calculated in the normalized 30-dimensional clustering space. PCA was used for visualization only. (D) SSM size distribution and cumulative fraction of the public repertoire, with modules ranked by decreasing number of unique public CDR3β sequences. (E) Independent topological validation of SSM cohesion using Levenshtein distance ≤ 2 and substitution-only Hamming connectivity among equal-length sequences. Values show $\log_{10}(\text{observed}/\text{random connectivity})$; zero indicates random expectation. (F) Association between SSM identity and germline-gene usage, quantified by Cramér's V (TRBV = 0.046; TRBJ = 0.663).

Structural Sequence Modules define biologically coherent sequence neighborhoods

Having established the atlas, we next examined whether individual SSMs displayed reproducible sequence and biological characteristics beyond their definition in latent clustering space.

SSMs differed substantially in their CDR3 β length architecture. Modules with similar mean CDR3 β lengths frequently displayed different degrees of within-module length variability, indicating that module identity was not reducible to receptor length alone (Fig. 2A).

Module publicness was remarkably conserved between age groups. After normalization for cohort size, the relative publicness of individual SSMs was almost identical in Adults and Elders (Spearman $\rho = 0.992$, $P = 4.03 \times 10^{-57}$) (Fig. 2B). Thus, although the absolute number of sequences satisfying the publicness criterion depends on cohort size, the relative hierarchy of more and less public structural neighborhoods is highly stable.

SSMs also exhibited distinct germline and sequence characteristics. TRBJ composition varied substantially across modules, generating module-specific germline profiles (Fig. 2C). Representative sequence logos revealed characteristic position-specific amino-acid patterns within individual SSMs (Fig. 2D), while positional entropy profiles demonstrated heterogeneous distributions of conserved and variable residues across normalized CDR3 β positions (Fig. 2E).

Together with the independent topological validation in Figure 1, these observations indicate that SSMs represent reproducible sequence neighborhoods characterized by distinct combinations of CDR3 β architecture, germline composition and sequence conservation.

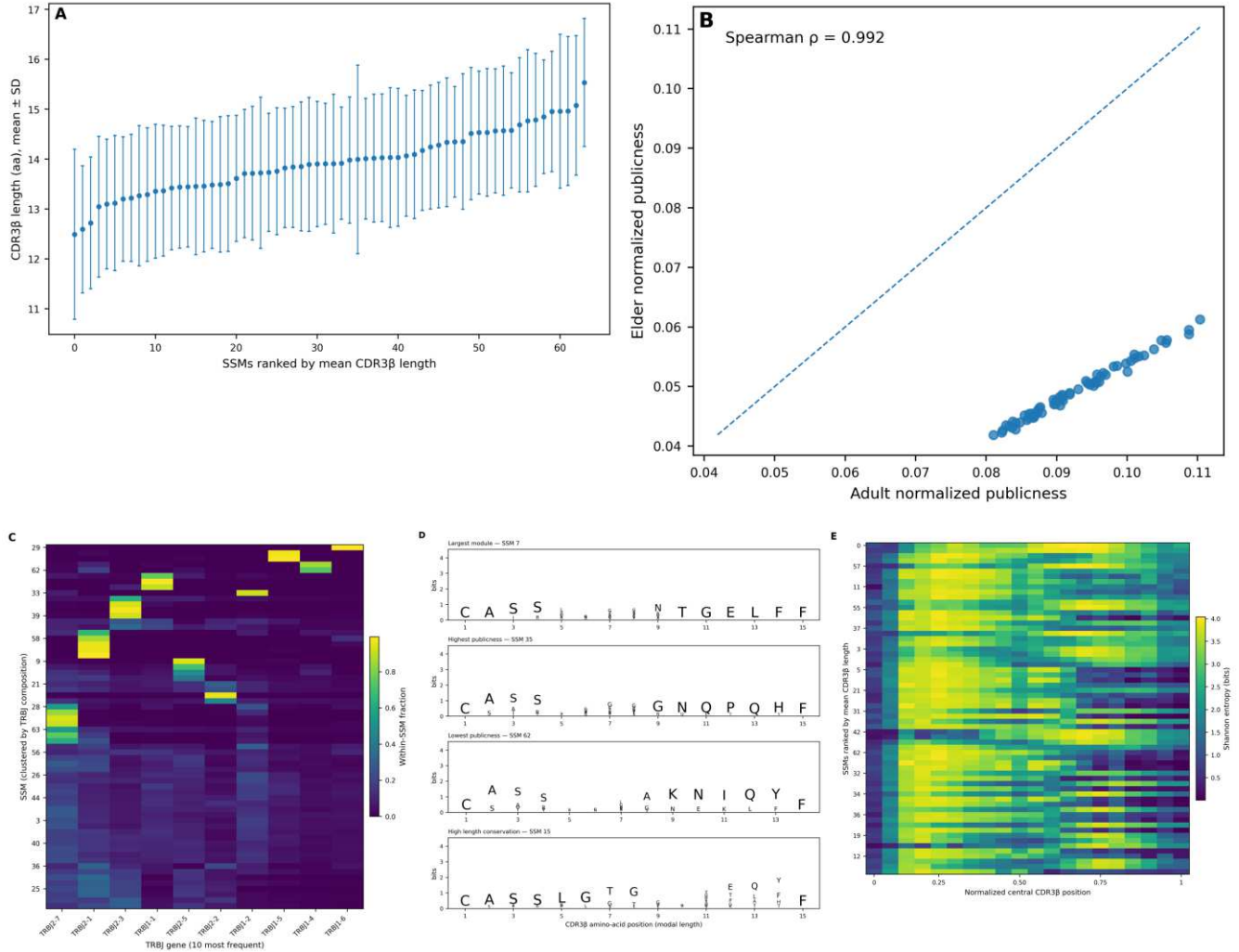


Figure 2 | Structural Sequence Modules represent biologically coherent units of the public TCRβ repertoire. (A) CDR3β length architecture across the 64 SSMs. Points indicate module means and error bars the corresponding standard deviations. (B) Conservation of normalized module publicness between Adults and Elders (Spearman $\rho = 0.992$, $P = 4.03 \times 10^{-57}$); the dashed line indicates identity. (C) Within-SSM relative frequencies of the ten most frequent TRBJ genes. Modules are ordered by similarity in TRBJ composition. (D) Sequence logos for representative SSMs selected by predefined module-level criteria. Logos were generated from sequences of modal CDR3β length; letter height represents position-specific information content. (E) Positional Shannon entropy across normalized central CDR3β positions. Lower values indicate greater sequence conservation.

Exceptional longevity preserves the structural atlas while selectively remodeling module occupancy

We next asked whether decades of immune ageing alter the structural repertoire itself or instead modify the relative utilization of pre-existing SSMs.

Projection of subject-specific public repertoires onto the common atlas demonstrated complete preservation of structural breadth: all 64 SSMs were detected in every individual from both cohorts (Fig. 3A). Despite this preservation, 18 of 64 SSMs exhibited significant Adult–Elder differences in

unweighted occupancy after Benjamini–Hochberg correction ($\text{FDR} < 0.05$) (Fig. 3B). Age-associated differences therefore reflected redistribution of clonotypic representation among pre-existing modules rather than loss or replacement of structural components.

Global occupancy architecture remained highly conserved. Jensen–Shannon distance analysis followed by PCoA showed extensive overlap between cohorts, although PERMANOVA detected a modest cohort effect (pseudo- $F = 2.58$, $P = 0.040$) (Fig. 3C). Subject-level analysis showed identical SSM richness in both groups, whereas Shannon occupancy diversity and Pielou evenness were slightly higher in Elders ($\text{FDR} = 0.049$ for both); maximum module occupancy did not differ significantly ($\text{FDR} = 0.725$) (Fig. 3D). The overall module hierarchy was nevertheless nearly identical between Adults and Elders (Spearman $\rho = 0.996$, $P = 3.33 \times 10^{-66}$) (Fig. 3E).

Because the Elder cohort was larger, we tested whether unequal cohort size could account for the observed occupancy differences. Across 1,000 rarefaction iterations in which Elders were subsampled to 31 individuals and public clonotypes re-identified *de novo*, cohort balancing substantially altered the absolute number of public sequences detected but had minimal impact on module-level occupancy effects. Full-cohort and rarefied effect sizes were almost perfectly correlated ($\rho \approx 0.999$), with mean effect direction preserved across all 64 SSMs (Fig. 3F).

A complementary frequency-weighted analysis identified no individual SSM remaining significant after FDR correction, indicating that the unweighted occupancy differences were not primarily attributable to expansion of a limited number of highly abundant public clonotypes.

To assess whether age-associated occupancy differences were influenced by HLA background, we performed a sensitivity analysis restricted to HLA-A02:01-positive individuals (10 Adults and 22 Elders). Despite the reduced sample size, module-level occupancy effects remained strongly concordant with those observed in the complete cohort (Spearman $\rho = 0.651$, $P = 5.59 \times 10^{-9}$), and all 18 SSMs identified as significantly remodeled in the primary analysis retained the same direction of effect (Supplementary Fig. 1). Global occupancy profiles did not differ significantly between Adults and Elders within the HLA-A02:01-positive subset (PERMANOVA pseudo- $F = 1.418$, $P = 0.187$), and no significant differences were observed in Shannon occupancy diversity, Pielou evenness or maximum module occupancy (all $P > 0.36$). Thus, restriction to a shared HLA background preserved the direction of age-associated module remodeling, whereas the reduced sample size limited statistical power for detecting cohort-level differences.

Together, these analyses indicate that exceptional longevity is associated with quantitative redistribution of unique public clonotypes within a strongly conserved structural atlas, with module-level effects remaining robust to cohort-size imbalance and restriction to a shared HLA-A*02:01 background.

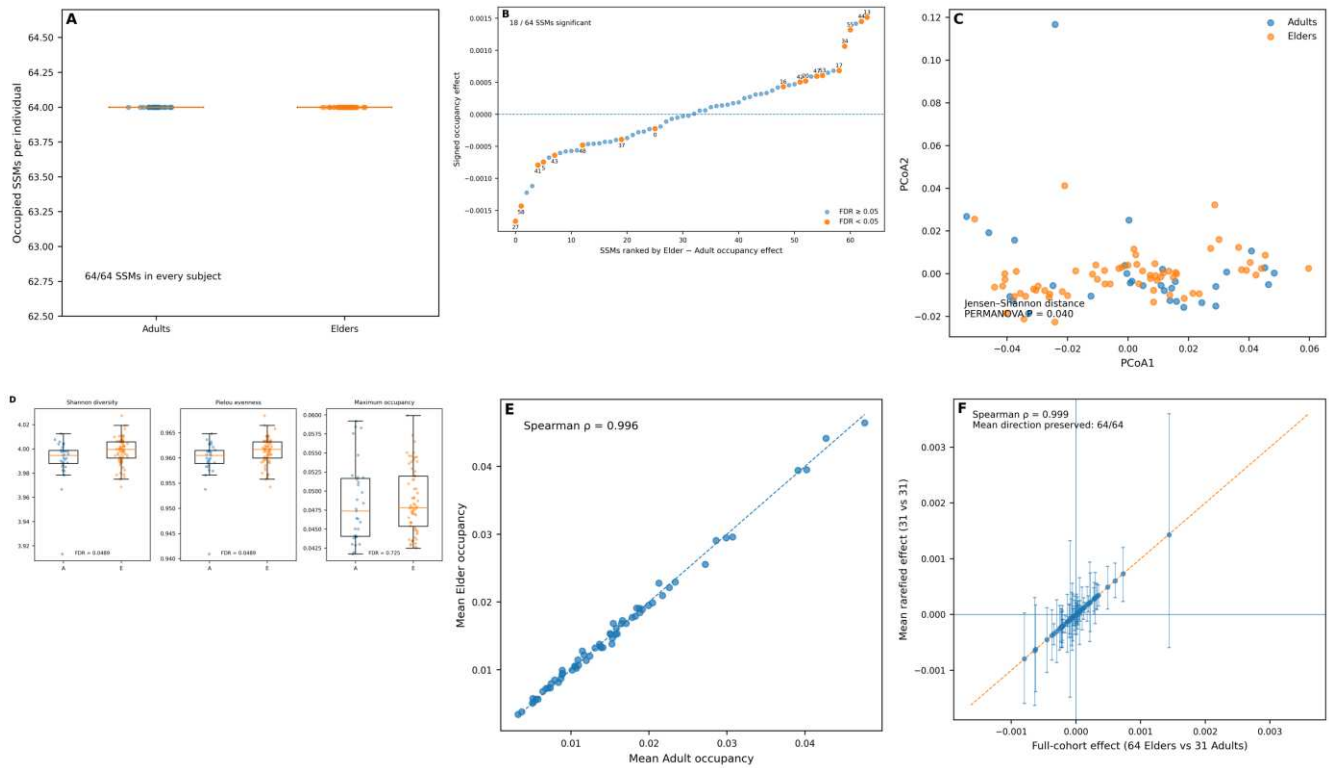


Figure 3 | Exceptional longevity preserves the structural atlas while selectively remodeling module occupancy. (A) Number of occupied SSMs per individual. All Adults and Elders contained 64/64 SSMs. (B) Elder-minus-Adult differences in unweighted SSM occupancy. Eighteen of 64 SSMs were significant after Benjamini–Hochberg correction (FDR < 0.05). (C) PCoA of individual occupancy profiles based on Jensen–Shannon distance (PERMANOVA $P = 0.040$). (D) Shannon occupancy diversity, Pielou evenness and maximum module occupancy in Adults and Elders. FDR-adjusted P values are shown. (E) Mean SSM occupancy in Adults versus Elders (Spearman $\rho = 0.996$, $P = 3.33 \times 10^{-66}$); the dashed line indicates identity. (F) Cohort-size rarefaction. Full-cohort occupancy effects are compared with mean effects from 1,000 balanced 31-versus-31 resamplings ($\rho = 0.999$); error bars indicate rarefaction intervals.

Intrinsic module properties are associated with age-associated occupancy remodeling

We next asked whether susceptibility to age-associated remodeling was related to intrinsic properties of individual SSMs.

Across candidate module descriptors, the magnitude of Adult–Elder occupancy differences showed its strongest positive association with module size (Spearman $\rho = 0.405$, FDR = 0.0116) (Fig. 4A,B). Remodeling magnitude was inversely associated with within-module CDR3 β length variability, measured both as CDR3 β length standard deviation ($\rho = -0.373$, FDR = 0.0155) and length entropy ($\rho = -0.359$, FDR = 0.0156) (Fig. 4A,C,D). Larger modules with more constrained CDR3 β length distributions therefore tended to undergo greater quantitative occupancy shifts.

In contrast, module publicness was not significantly associated with remodeling magnitude ($\rho = 0.197$, $P = 0.120$), nor was Levenshtein network density ($\rho = -0.080$, $P = 0.532$). These findings do not

support the earlier hypothesis that lower publicness or local network density is a major determinant of age-associated remodeling.

To identify independent contributions, candidate descriptors were analyzed using Elastic Net feature selection followed by OLS regression with HC3 heteroscedasticity-consistent standard errors. Module size remained positively associated with remodeling (standardized $\beta = 0.466$, $P = 4.6 \times 10^{-5}$), whereas CDR3 β length SD showed an independent inverse association ($\beta = -0.236$, $P = 7.7 \times 10^{-5}$). The final model explained 26.7% of between-module variation (adjusted $R^2 = 0.267$; Fig. 4E).

Thus, age-associated occupancy remodeling is non-uniform across the atlas and is associated primarily with module size and within-module CDR3 β length constraint.

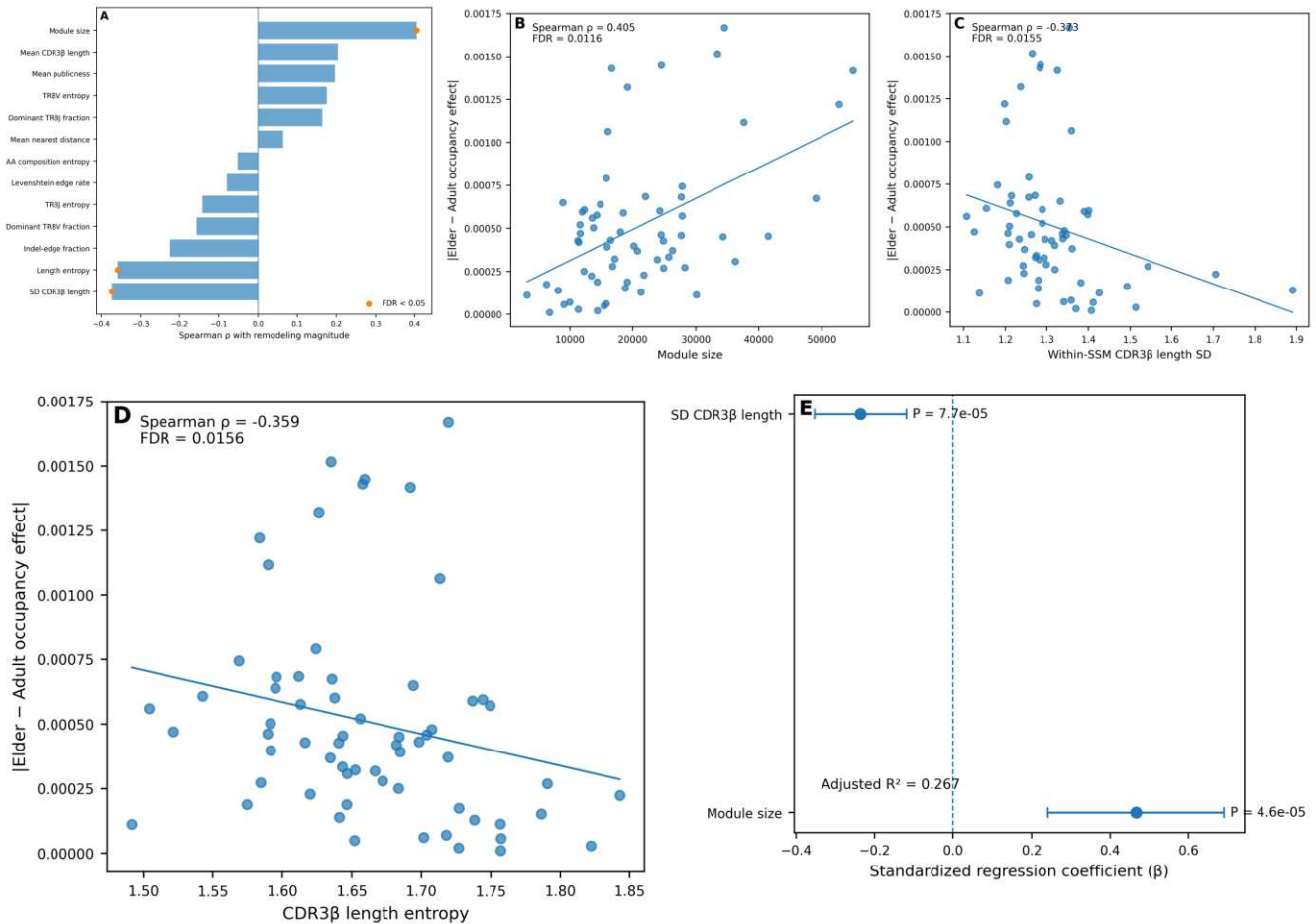


Figure 4 | Intrinsic module properties are associated with age-related occupancy remodeling. (A) Spearman associations between intrinsic SSM descriptors and absolute Elder-minus-Adult occupancy effects. Significant associations after Benjamini–Hochberg correction are indicated. (B–D) Relationships of remodeling magnitude with module size ($\rho = 0.405$, FDR = 0.0116), within-SSM CDR3 β length SD ($\rho = -0.373$, FDR = 0.0155) and CDR3 β length entropy ($\rho = -0.359$, FDR = 0.0156), respectively. Lines are shown for visualization only. (E) Multivariable model after Elastic Net feature selection and OLS refitting with HC3 standard errors. Points show standardized coefficients and 95% confidence intervals. Module size ($\beta = 0.466$, $P = 4.6 \times 10^{-5}$) and CDR3 β length SD ($\beta = -0.236$, $P = 7.7 \times 10^{-5}$) were independently associated with remodeling (adjusted $R^2 = 0.267$).

Publicness is associated with the intrinsic organization of Structural Sequence Modules

The striking conservation of module publicness between Adults and Elders prompted us to investigate the structural properties associated with population sharing.

Across the 64 SSMs, publicness was significantly associated with multiple sequence-, germline- and topology-based descriptors (Fig. 5A). Positive associations included TRBV entropy ($\rho = 0.572$), Levenshtein network density ($\rho = 0.486$) and dominant TRBJ fraction ($\rho = 0.469$), whereas publicness was inversely associated with within-module CDR3 β length variability ($\rho = -0.456$), length entropy ($\rho = -0.446$), TRBJ entropy ($\rho = -0.427$) and nearest-neighbor sequence distance ($\rho = -0.417$) (Fig. 5B–D). Module size showed a weaker positive association ($\rho = 0.256$, FDR = 0.049) (Fig. 5E).

These relationships define a coordinated structural phenotype: more public SSMs tend to display greater local sequence connectivity, more constrained CDR3 β length distributions and distinct germline architecture. Because many of these descriptors were themselves correlated, however, their univariable associations should not be interpreted as independent effects.

Accordingly, Elastic Net feature selection followed by multivariable regression produced a model explaining 43.8% of between-module variation in publicness (adjusted $R^2 = 0.438$). After accounting for covariance among predictors, module size was the only independently significant predictor (standardized $\beta = 0.516$, 95% CI 0.071–0.961, $P = 0.023$) (Fig. 5F).

Module publicness should therefore be regarded as a highly conserved module-level characteristic associated with coordinated structural, sequence and germline organization, rather than as evidence for multiple independent structural determinants of sharing.

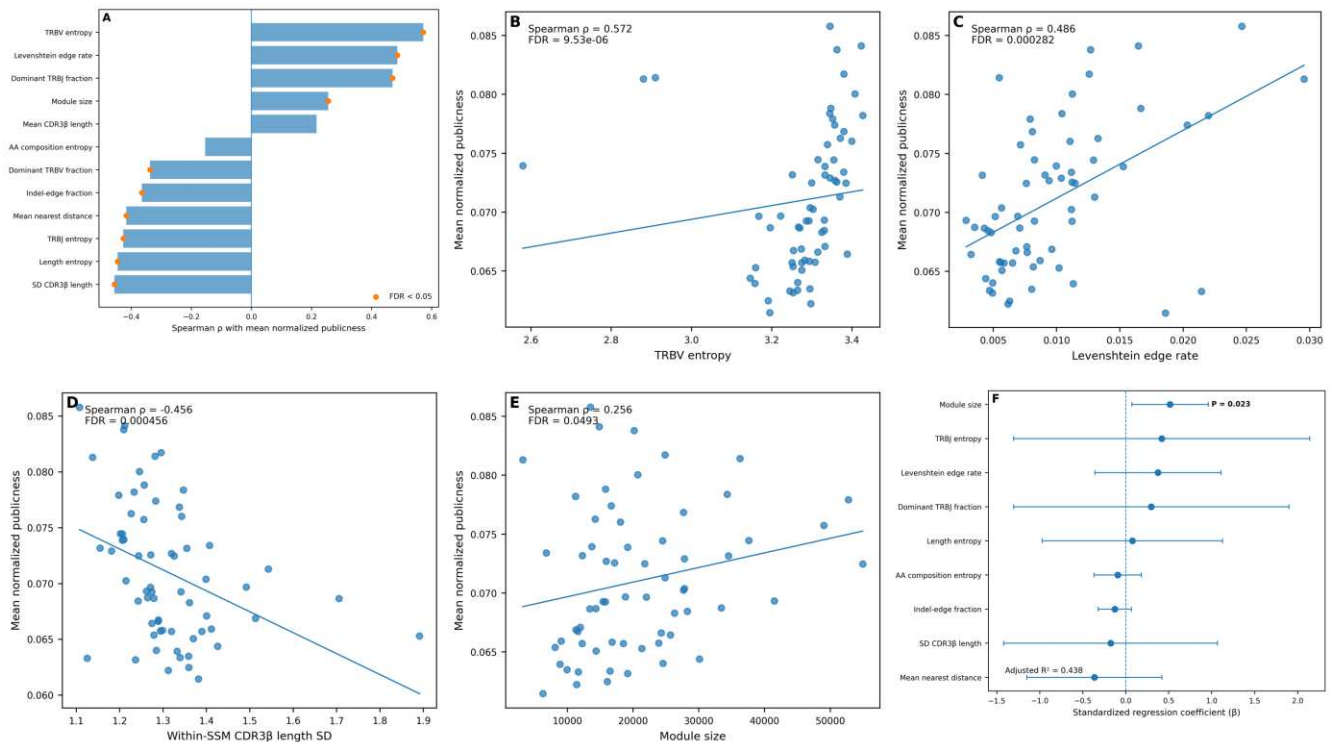


Figure 5 | Publicness is associated with the intrinsic organization of Structural Sequence Modules. (A) Spearman associations between mean normalized module publicness and intrinsic SSM descriptors. FDR-significant associations are indicated. (B–E) Relationships of publicness with TRBV entropy ($\rho = 0.572$), Levenshtein edge rate ($\rho = 0.486$), within-SSM CDR3 β length SD ($\rho = -0.456$) and module size ($\rho = 0.256$), respectively. Corresponding FDR values are shown in each panel; fitted lines are for visualization only. (F) Multivariable model after Elastic Net feature selection and OLS refitting with HC3 standard errors. Points show standardized coefficients and 95% confidence intervals. The model explained 43.8% of between-module variation (adjusted $R^2 = 0.438$); module size was the only independently significant predictor ($\beta = 0.516$, 95% CI 0.071–0.961, $P = 0.023$).

Structural Sequence Modules exhibit non-random antigen-association profiles without global structure–function distance coupling

We finally investigated whether SSM organization was associated with antigen-recognition properties using curated VDJdb annotations. In the primary stringent analysis, matches were required to agree for CDR3 β amino-acid sequence, TRBV and TRBJ. This identified 11,065 unique public CDR3 β sequences representing 917 distinct epitopes, with annotated sequences present in all 64 SSMs (Fig. 6A).

Antigen annotations were non-randomly distributed across the atlas. SSM identity was significantly associated with exact epitope identity (Cramér's $V = 0.279$), antigen gene ($V = 0.141$) and antigen species ($V = 0.092$) (Fig. 6B,C). At the antigen-species level, 32 of 64 SSMs showed at least one significant enrichment or depletion after FDR correction, indicating that individual modules possess distinct antigen-association profiles.

These functional differences were not systematically related to age-associated occupancy remodeling. Antigen-species richness ($P = 0.242$), Shannon diversity ($P = 0.464$) and dominance ($P = 0.578$) did not differ significantly among Adult-increased, stable and Elder-increased SSMs (Kruskal–Wallis tests; Fig. 6D).

Importantly, non-random antigen association did not imply that neighboring SSMs had similar functional profiles. In the primary stringent-matching analysis, structural distance was not associated with exact-epitope functional distance (Spearman $\rho = -0.053$, permutation $P = 0.409$; Fig. 6E). No association was detected at the antigen-species level ($\rho = -0.066$, $P = 0.265$), whereas antigen-gene analysis yielded only a weak inverse association ($\rho = -0.144$, $P = 0.023$). Sensitivity analyses using the broader CDR3 β -only matching scheme similarly failed to support a consistent structure–function distance relationship (Fig. 6F).

Thus, SSMs are functionally non-random but not functionally deterministic: individual modules show reproducible biases in curated antigen associations, whereas geometric proximity between modules does not provide a general predictor of functional similarity.

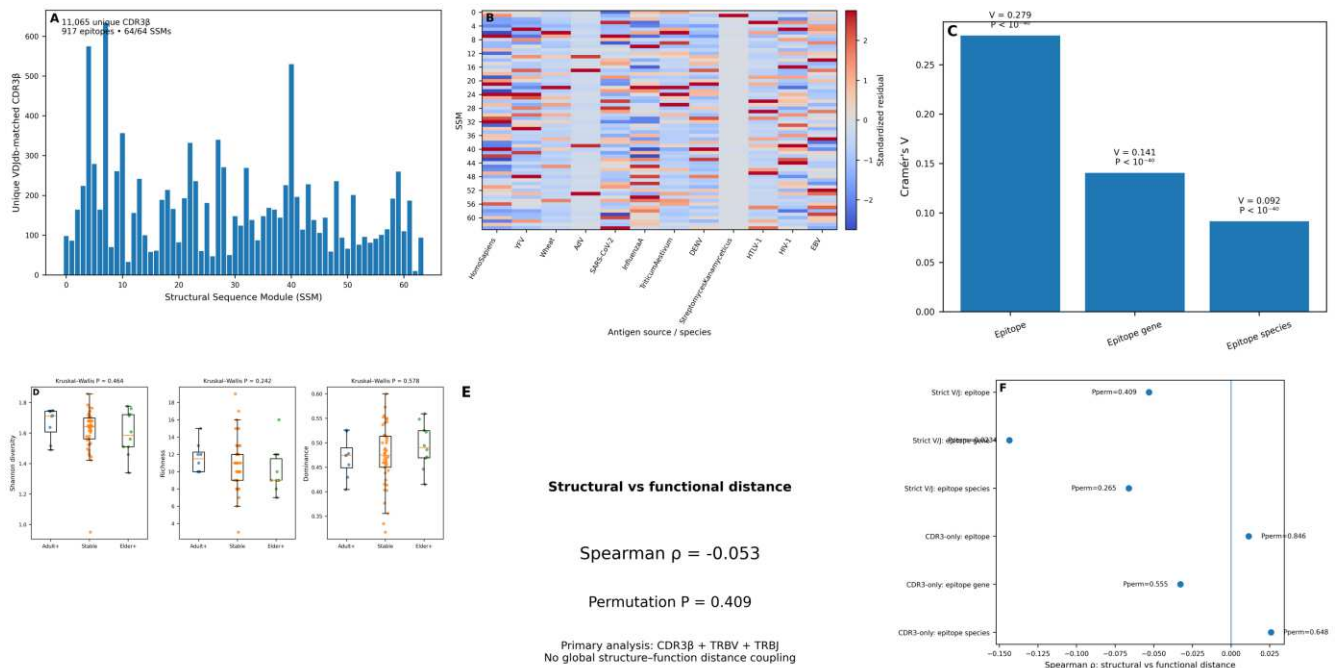


Figure 6 | Structural Sequence Modules exhibit non-random antigen-association profiles without global structure-function distance coupling. (A) VDJdb annotation coverage using stringent CDR3β/TRBV/TRBJ matching. 11,065 unique public CDR3β sequences, representing 917 exact epitopes, were distributed across all 64 SSMs. (B) Standardized residuals describing enrichment or depletion of selected antigen-source/species annotations across SSMs. (C) Association between SSM identity and functional annotation, quantified by Cramér's V at the epitope, epitope-gene and epitope-species levels. (D) Functional Shannon diversity, richness and dominance among Adult-increased, stable and Elder-increased SSMs; Kruskal-Wallis P values are shown. (E) Relationship between structural and functional distances in the primary stringent-matching analysis. No global association was detected at the epitope level (Spearman $\rho = -0.053$, permutation $P = 0.409$). (F) Sensitivity analyses across annotation levels and matching definitions. Points show Spearman correlations between structural and functional distances; permutation P values are indicated.

Discussion

The present study identifies a higher-order organizational structure within the human public TCR β repertoire and examines its stability in exceptional longevity. By representing more than 1.3 million unique public CDR3 β sequences within a common structural framework, we defined 64 reproducible Structural Sequence Modules (SSMs) supported by quantitative model selection and independent topological validation. Projection of individual repertoires onto this atlas revealed a striking combination of structural conservation and quantitative remodeling: all 64 modules were retained across Adults and Elders, whereas a subset showed age-associated differences in occupancy. These findings suggest that exceptional longevity is accompanied primarily by redistribution within a conserved repertoire architecture rather than by loss or replacement of its major structural components.

This framework extends conventional descriptions of public T-cell repertoires, which have largely focused on clonotype sharing, diversity, abundance, germline-gene usage and antigen specificity. Such measures remain essential for characterizing repertoire composition but generally treat clonotypes as individual analytical units (10, 11, 19). The SSM framework introduces an intermediate organizational level in which related public CDR3 β sequences are considered collectively as reproducible neighborhoods in sequence space (12 – 15). Importantly, these modules were not defined solely by a single sequence characteristic. The selected 64-module resolution was reproducible across clustering initializations, all modules showed greater-than-random internal connectivity under independent Levenshtein- and Hamming-based criteria, and individual SSMs displayed distinct CDR3 β length, sequence-conservation and germline profiles. Together, these observations support the interpretation of SSMs as reproducible structural units rather than arbitrary partitions of the public repertoire.

One of the principal findings is the extent to which this organization is preserved across age groups. Every SSM was detected in every individual, and the rank hierarchy of mean module occupancy was almost identical between Adults and Elders. At the same time, 18 of 64 modules showed significant differences in unweighted occupancy, and multivariate analysis detected a modest cohort effect on the overall occupancy distribution. Thus, conservation and remodeling are not mutually exclusive: the repertoire retains the same structural components and nearly the same global hierarchy while redistributing a fraction of its public clonotypes among them. This distinction is important because it argues against describing healthy immune ageing either as complete structural invariance or as wholesale repertoire reorganization. Instead, exceptional longevity appears compatible with limited quantitative remodeling within a strongly conserved structural framework.

Three sensitivity analyses further refine this interpretation. First, balancing the Elder cohort to the Adult sample size substantially altered the number of clonotypes classified as public but had minimal effect on module-level occupancy differences, demonstrating that cohort size strongly influences public-clonotype discovery but does not account for the observed remodeling pattern. Second, frequency-weighted analysis produced no FDR-significant module differences, indicating that the primary unweighted signal is not simply a consequence of expansion of highly abundant public clonotypes. Third, restriction to HLA-A*02:01-positive individuals preserved the overall pattern of module-level effects: effect sizes remained strongly concordant with the complete cohort, and all 18 SSMs identified as remodeled in the primary analysis retained the same direction of effect. Although

cohort-level significance was lost in this smaller subset, this is consistent with the substantial reduction in sample size and should be interpreted together with the preservation of effect direction. Collectively, these analyses indicate that the observed remodeling is robust to cohort-size imbalance and restriction to a shared HLA background and primarily reflects redistribution of distinct public clonotypes rather than differences in clonal abundance.

The structural characteristics associated with remodeling also differed from our initial expectation. Remodeling magnitude was not significantly related to module publicness or local Levenshtein network density. Instead, larger modules and modules with more constrained CDR3 β length distributions showed greater occupancy differences between age groups. These associations persisted in multivariable analysis, in which module size and CDR3 β length variability contributed independently to remodeling magnitude. The resulting model explained approximately 27% of between-module variation, indicating that intrinsic module architecture contributes to susceptibility to remodeling but does not fully determine it. Other influences not captured by the present structural descriptors—including antigenic history, cellular phenotype, HLA background or stochastic clonal dynamics—are therefore likely to contribute to the remaining variation.

Module publicness showed a different pattern. Relative publicness was almost perfectly conserved between Adults and Elders and was associated with multiple aspects of SSM organization, including germline diversity, local sequence connectivity and CDR3 β length architecture. These observations support the concept that population sharing is organized non-randomly across structural sequence space. However, the multivariable analysis substantially attenuated several strong univariable associations, with module size emerging as the only independently significant predictor in the final model. Publicness should therefore not be interpreted as being independently determined by each of the correlated structural properties examined here. A more conservative interpretation is that highly and weakly public modules occupy different regions of a coordinated structural landscape, within which several sequence, germline and topological characteristics covary (6 – 9).

The pronounced difference between TRBV and TRBJ associations with SSM identity provides an additional clue to the organization of this landscape. TRBV usage showed only a weak relationship with module assignment, whereas TRBJ usage was substantially more strongly associated with SSM identity. Because the atlas was constructed from central CDR3 β sequence representations rather than germline-gene labels, this asymmetry suggests that the modular organization captures sequence features that are differentially related to the germline contributions surrounding the CDR3 β region. At the same time, neither association should be interpreted as demonstrating that SSM identity is determined by germline usage: modules contain heterogeneous germline configurations and were defined independently of TRBV/TRBJ annotation. The biological basis of the stronger TRBJ association will require dedicated mechanistic analysis (15, 20, 21).

Functional annotation provided a complementary view of SSM organization. Using stringent CDR3 β /TRBV/TRBJ matching to VDJdb, antigen-specific sequences were identified in all 64 modules, and SSM identity was significantly associated with exact epitope, antigen-gene and antigen-species annotations. Individual modules therefore exhibit non-random antigen-association profiles. Importantly, however, this result is distinct from a global relationship between structural and functional distances. Structurally neighboring SSMs did not consistently display more similar epitope profiles, and the absence of a general distance relationship was reproduced across alternative annotation levels and the broader CDR3 β -only sensitivity analysis. The isolated weak association

observed for antigen-gene profiles under stringent matching was not reproduced across the other definitions and therefore does not provide evidence for general structure–function distance coupling (13, 14, 22 – 24).

This distinction helps define what the structural atlas does—and does not—represent biologically. SSM membership contains information about antigen-association patterns at the population level, but structural proximity alone is insufficient to infer antigen specificity (15, 25, 26). This is compatible with the many-to-many relationship between TCR sequence and antigen recognition, as well as with the incomplete and non-uniform coverage of current antigen-specific TCR databases. The atlas should therefore be viewed as an organizational framework for grouping related public receptors rather than as a deterministic map from sequence geometry to antigen specificity. Functional information can be enriched within particular structural neighborhoods without being arranged along a simple continuous gradient across the atlas.

More broadly, the SSM framework provides an intermediate scale between individual clonotypes and repertoire-wide summary statistics. At the clonotype level, immune repertoires are extraordinarily diverse and sensitive to individual exposure history. At the repertoire level, global diversity indices can obscure localized changes. Structural modules offer a complementary representation in which conservation, remodeling and biological annotation can be examined within the same fixed reference system. This may be particularly useful for longitudinal or comparative studies in which the central question is not whether identical clonotypes are shared between groups, but whether different repertoires occupy common regions of sequence space in systematically different ways.

Several limitations should be considered. First, the atlas was constructed exclusively from public clonotypes and therefore does not represent the much larger private component of individual TCR repertoires. The extent to which comparable modular organization exists in private sequence space remains unknown. Second, public-clonotype discovery is intrinsically dependent on cohort size. Although rarefaction demonstrated that the principal occupancy effects were robust to the imbalance between Adults and Elders, absolute numbers of public and cohort-exclusive clonotypes should not be interpreted independently of sampling depth. Third, the study is cross-sectional and therefore cannot determine when the observed occupancy differences arise or whether individual modules change progressively within the same person during ageing. Fourth, functional analyses depend on curated VDJdb annotations, which represent only a fraction of antigen-specific TCR diversity and are affected by experimental, HLA and literature biases. The stringent matching strategy increases annotation specificity at the cost of reduced coverage, although the broader CDR3 β -only sensitivity analysis led to the same overall conclusion regarding the absence of general structure–function distance coupling. Finally, the structural atlas is based on TCR β sequences alone and therefore does not incorporate paired TCR α information, cellular phenotype or the complete HLA context of antigen recognition.

Collectively, our findings support a model in which the human public TCR β repertoire is organized into reproducible structural neighborhoods that remain strongly conserved in exceptional longevity. Age-associated differences are expressed primarily as quantitative redistribution of distinct public clonotypes among pre-existing modules, rather than as loss or replacement of the underlying atlas, and the susceptibility of individual modules to remodeling is partly associated with their intrinsic architecture. SSMs also exhibit highly conserved differences in population publicness and non-random antigen associations, while structural proximity does not provide a general predictor of functional

similarity. The structural atlas therefore provides a reproducible framework for investigating how adaptive immune repertoires preserve global organization while accommodating quantitative change across ageing and, potentially, other immune states.

Acknowledgements

N/A

Declarations

1. **Ethics approval**

The study has been approved by the Institutional Ethical Committee of the Istituto Nazionale Tumori - IRCCS - "Fond G. Pascale" – NAPOLI IT (Protocol n. 57/22 OSS "HepAnt – Tumor antigen discovery for innovative cancer immunotherapies in HCC: from benchside to bedside" e subsequent amendments)

2. **Consent for publication**

The corresponding author has received consent for publication.

3. **Consent to participate**

All the participants have signed the consent to participate.

4. **Availability of data and material**

The processed AIRR-seq format TCRB contig datasets generated and analysed during the current study are not publicly available due to intellectual property and confidentiality restrictions.

5. **Competing interests**

The authors declare no potential conflicts of interest.

6. **Funding**

The study was funded by the Italian Ministry of Health through Institutional "Ricerca Corrente" (Projects L2/30_25 to LB; L2/35_25 to MT); the PNRR Ministero Salute PNRR-POC-2022-12375769 "Molecular mimicry to improve liver cancer immunotherapy" (2023 – 2025) (to LB).

7. **Authors' contributions**

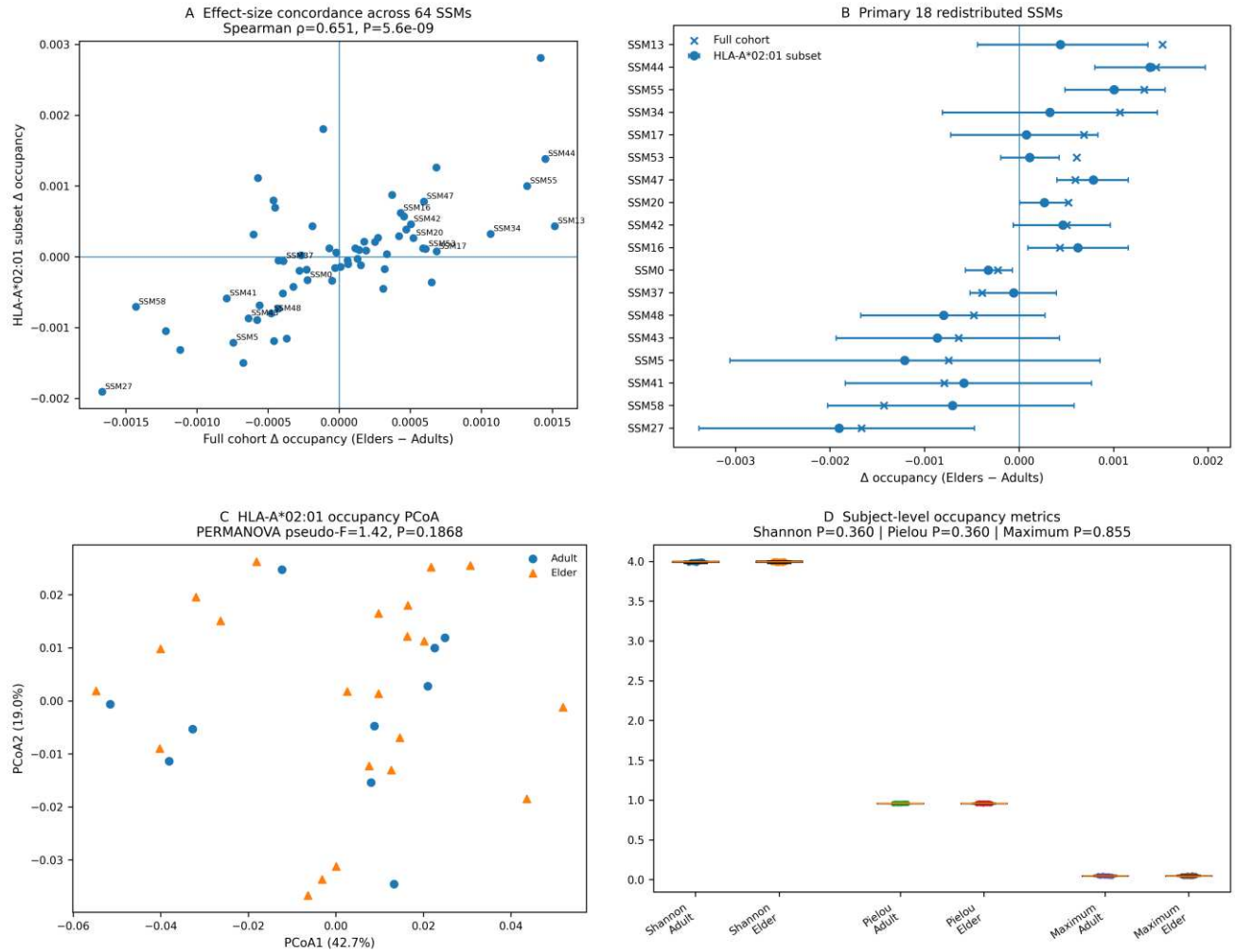
BC, AM, CR, BMC, SM, performed all the wet lab analyses and bioinformatics analyses; AMS, JLM, MS, MG, MC, HH performed all the NGS analyses and additional confirmatory statistical analyses; MT and LB designed and coordinated the study; BC, MT and LB drafted the manuscript. All Authors approved the manuscript.

References

1. Weng NP. Numbers and odds: TCR repertoire size and its age changes impacting on T cell functions. *Semin Immunol.* 2023;69:101810. doi:10.1016/j.smim.2023.101810.
2. Britanova OV, Putintseva EV, Shugay M, et al. Age-related decrease in TCR repertoire diversity measured with deep and normalized sequence profiling. *J Immunol.* 2014;192:2689–2698. doi:10.4049/jimmunol.1302064.
3. Bektas A, Schurman SH, Sen R, Ferrucci L. Human T cell immunosenescence and inflammation in aging. *J Leukoc Biol.* 2017;102:977–988. doi:10.1189/jlb.3RI0716-335R.
4. Appay V, Sauce D. Naive T cells: the crux of cellular immune aging? *Exp Gerontol.* 2014;54:90–93. doi:10.1016/j.exger.2014.01.003.
5. Weiskopf D, Weinberger B, Grubeck-Loebenstien B. The aging of the immune system. *Transpl Int.* 2009;22:1041–1050. doi:10.1111/j.1432-2277.2009.00927.x.
6. Venturi V, Chin HY, Price DA, Douek DC, Davenport MP. Sharing of T cell receptors in antigen-specific responses is driven by convergent recombination. *Proc Natl Acad Sci USA.* 2006.
7. Quigley MF, Greenaway HY, Venturi V, et al. Convergent recombination shapes the clonotypic landscape of the naive T-cell repertoire. *Proc Natl Acad Sci USA.* 2010.
8. Murugan A, Mora T, Walczak AM, Callan CG Jr. Statistical inference of the generation probability of T-cell receptors from sequence repertoires. *Proc Natl Acad Sci USA.* 2012.
9. Elhanati Y, Sethna Z, Marcou Q, Callan CG Jr, Mora T, Walczak AM. Predicting the spectrum of TCR repertoire sharing with a data-driven model of recombination. *Immunol Rev.* 2018
10. Mhanna V, Bashour H, Lê Quý K, et al. Adaptive immune receptor repertoire analysis. *Nat Rev Methods Primers.* 2024;4:6. doi:10.1038/s43586-023-00284-1.
11. Chiffelle J, Genolet R, Perez MAS, Coukos G, Zoete V, Harari A. T-cell repertoire analysis and metrics of diversity and clonality. *Curr Opin Biotechnol.* 2020;65:284–295. doi:10.1016/j.copbio.2020.07.010.
12. Madi A, Poran A, Shifrut E, et al. T cell receptor repertoires of mice and humans are clustered in similarity networks around conserved public CDR3 sequences. *eLife.* 2017;6:e22057. doi:10.7554/eLife.22057.
13. Glanville J, Huang H, Nau A, et al. Identifying specificity groups in the T cell receptor repertoire. *Nature.* 2017;547:94–98. doi:10.1038/nature22976.
14. Dash P, Fiore-Gartland AJ, Hertz T, et al. Quantifiable predictive features define epitope-specific T cell receptor repertoires. *Nature.* 2017;547:89–93. doi:10.1038/nature22383.
15. Mayer-Blackwell K, Schattgen S, Cohen-Lavi L, et al. TCR meta-clonotypes for biomarker discovery with tcridist3 enabled identification of public, HLA-restricted clusters of SARS-CoV-2 TCRs. *eLife.* 2021;10:e68605. doi:10.7554/eLife.68605.

16. Sidhom JW, Larman HB, Pardoll DM, Baras AS. DeepTCR is a deep learning framework for revealing sequence concepts within T-cell repertoires. *Nat Commun.* 2021;12:1605. doi:10.1038/s41467-021-21879-w.
17. Hu J, Pan M, Reid B, Tworoger S, Li B. Quantifiable blood TCR repertoire components associate with immune aging. *Nat Commun.* 2024;15:8171. doi:10.1038/s41467-024-52522-z.
18. Britanova OV, Shugay M, Merzlyak EM, et al. Dynamics of individual T cell repertoires: from cord blood to centenarians. *J Immunol.* 2016;196:5005–5013. doi:10.4049/jimmunol.1600005.
19. Miho E, Yermanos A, Weber CR, Berger CT, Reddy ST, Greiff V. Computational strategies for dissecting the high-dimensional complexity of adaptive immune repertoires. *Front Immunol.* 2018;9:224. doi:10.3389/fimmu.2018.00224.
20. Emerson RO, DeWitt WS, Vignali M, et al. Immunosequencing identifies signatures of cytomegalovirus exposure history and HLA-mediated effects on the T cell repertoire. *Nat Genet.* 2017;49:659–665. doi:10.1038/ng.3822.
21. DeWitt WS III, Smith A, Schoch G, Hansen JA, Matsen FA IV, Bradley P. Human T cell receptor occurrence patterns encode immune history, genetic background, and receptor specificity. *eLife.* 2018;7:e38358. doi:10.7554/eLife.38358.
22. Shugay M, Bagaev DV, Zvyagin IV, et al. VDJdb: a curated database of T-cell receptor sequences with known antigen specificity. *Nucleic Acids Res.* 2018;46:D419–D427. doi:10.1093/nar/gkx760.
23. Goncharov M, Bagaev D, Shcherbinin D, et al. VDJdb in the pandemic era: a compendium of T cell receptors specific for SARS-CoV-2. *Nat Methods.* 2022;19:1017–1019. doi:10.1038/s41592-022-01578-0.
24. Huang H, Wang C, Rubelt F, Scriba TJ, Davis MM. Analyzing the *Mycobacterium tuberculosis* immune response by T-cell receptor clustering with GLIPH2 and genome-wide antigen screening. *Nat Biotechnol.* 2020;38:1194–1202. doi:10.1038/s41587-020-0505-4.
25. Valkiers S, et al. T cell receptor sequence clustering and antigen specificity. *Comput Struct Biotechnol J.* 2020.
26. Pogorelyy MV, Shugay M. A framework for annotation of antigen specificities in high-throughput T-cell repertoire sequencing studies. *Front Immunol.* 2019;10:2159. doi:10.3389/fimmu.2019.02159.

SUPPL FIG. 1



Suppl. Figure 1 | HLA-A*02:01 sensitivity analysis of age-associated SSM occupancy remodeling. (A)

Concordance between Elder-minus-Adult occupancy effects across all 64 SSMs in the complete cohort and the HLA-A02:01-positive subset (10 Adults and 22 Elders; Spearman $\rho = 0.651$, $P = 5.59 \times 10^{-9}$). (B) Comparison of full-cohort and HLA-A02:01-subset occupancy effects for the 18 SSMs significantly remodeled in the primary analysis. Points show subset effect estimates with 95% bootstrap confidence intervals; crosses indicate full-cohort effects. All 18 SSMs retained the same direction of effect. (C) PCoA of Jensen–Shannon distances between individual SSM occupancy profiles in the HLA-A02:01-positive subset (PERMANOVA pseudo-F = 1.418, $P = 0.187$). (D) Shannon occupancy diversity, Pielou evenness and maximum module occupancy in HLA-A02:01-positive Adults and Elders; no significant cohort differences were detected.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Suppl.table1.docx](#)