PNAS
www.pnas.org

Supplementary Information for

# Thymus-derived B cell clones persist in the circulation after thymectomy in myasthenia gravis

Ruoyi Jiang, Kenneth B. Hoehn, Casey S. Lee, Minh C. Pham, Robert Homer, Frank C. Detterbeck, Inmaculada Aban, Leslie Jacobson, Angela Vincent, Richard J. Nowak, Henry J. Kaminski, Steven H. Kleinstein, Kevin C. O'Connor

Kevin C. O'Connor, Steven H. Kleinstein
Email: kevin.oconnor@yale.edu, steven.kleinstein@yale.edu

**This PDF file includes:**

Figures S1 to S10
Tables S1 to S3



www.pnas.org/cgi/doi/10.1073/pnas.2007206117

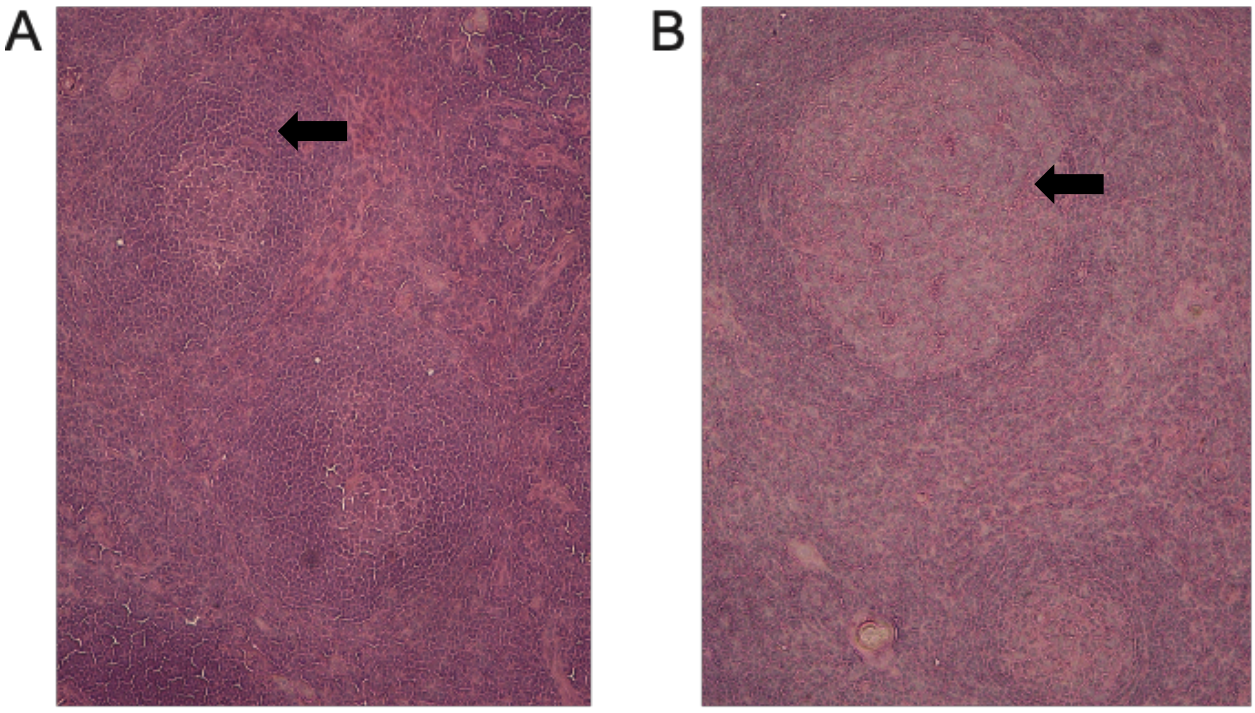
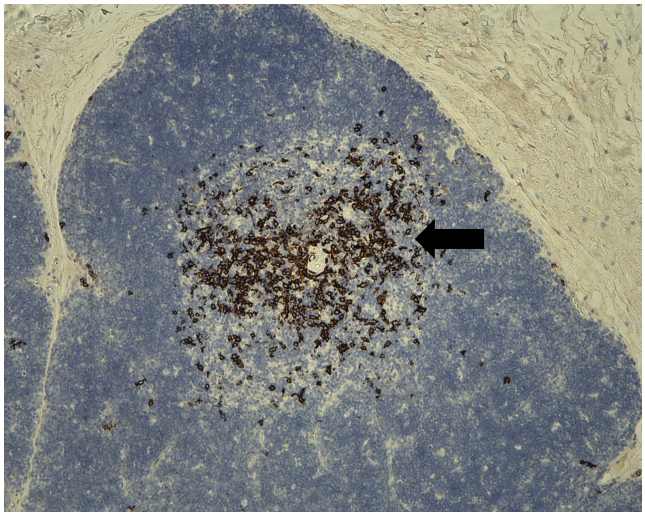



**Fig. S1.** Histology section showing multiple germinal centers (arrows) in a frozen section from thymus (specimen from patient THY2). Specimen stained with H&E (A, B).

A



B

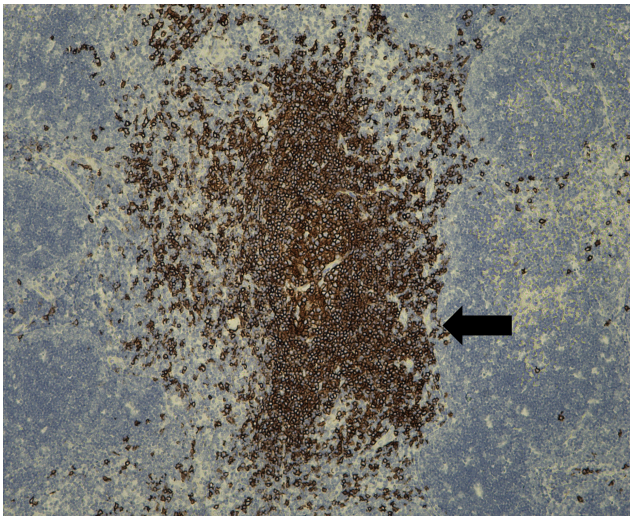


**Fig. S2.** Identification of a B cell infiltrate in specimen THY-Y; a sample from a patient who was not associated with the MGTX trial. (A, B) CD20 staining is shown showing cortex with B cell infiltration of the medulla (arrows).

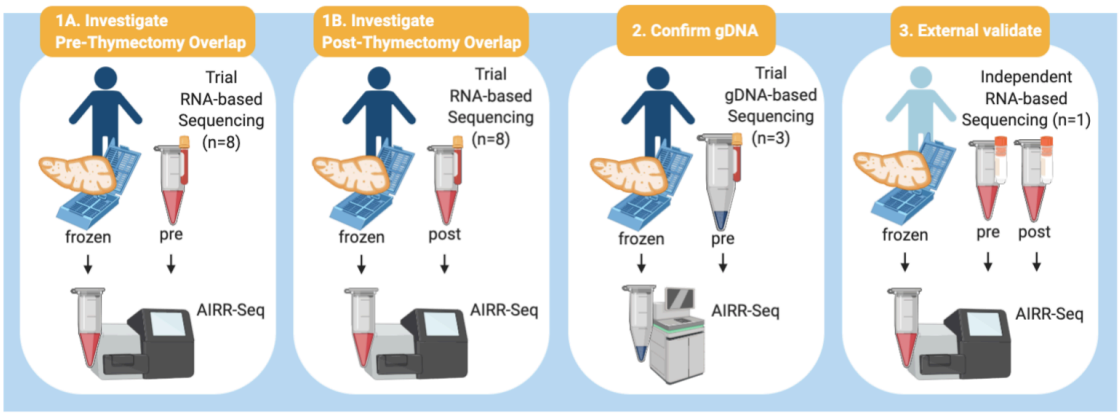



**Fig. S3.** Experimental outline illustrating the three approaches used to investigate the presence of clonal overlap between the thymus and the circulation. The goal of the first approach was to identify overlap between the circulation and the thymus before and after thymectomy in MGTX samples and characterize the features of that overlap. This was performed using an RNA-based B cell receptor AIRR-Seq approach. The goal of the second approach was to confirm the presence of overlap between the circulation and the thymus before thymectomy using gDNA-based B cell receptor AIRR-Seq. The goal of the final approach was to confirm the presence of overlap between the circulation and the thymus before and after thymectomy in a sample from an independent patient sample collected at a different institution.

**Table S1.** Table of V(D)J sequence counts and numbers of clones identified from mRNA-based AIRR sequencing of the B cell receptor.

| PATIENT | TIME | CLONES | IgA V(D)J SEQUENCES | IgG V(D)J SEQUENCES | IgM V(D)J SEQUENCES |
|---|---|---|---|---|---|
| THY1 | 0 Months | 4967 | 1740 | 1317 | 3391 |
| THY1 | 12 Months | 3978 | 918 | 737 | 2941 |
| THY1 | Thymus | 12418 | 1616 | 16225 | 689 |
| THY2 | 0 Months | 20756 | 6982 | 7518 | 13879 |
| THY2 | 12 Months | 7997 | 4959 | 4505 | 4649 |
| THY2 | Thymus | 38252 | 7735 | 58372 | 9753 |
| THY3 | 0 Months | 105 | 41 | 50 | 46 |
| THY3 | 12 Months | 90 | 37 | 35 | 47 |
| THY3 | Thymus | 29693 | 4309 | 59637 | 8762 |
| THY4 | 0 Months | 130 | 110 | 57 | 18 |
| THY4 | 12 Months | 157 | 142 | 54 | 44 |
| THY4 | Thymus | 3440 | 485 | 4520 | 274 |
| THY5 | 0 Months | 48 | 49 | 15 | 21 |
| THY5 | 12 Months | 305 | 194 | 108 | 108 |
| THY5 | Thymus | 13303 | 2299 | 25397 | 3831 |
| THY6 | 0 Months | 7866 | 4437 | 1661 | 4921 |
| THY6 | 12 Months | 4684 | 5236 | 1443 | 1560 |
| THY6 | Thymus | 11827 | 2226 | 20058 | 1256 |
| THY7 | 0 Months | 342 | 97 | 181 | 95 |
| THY7 | 12 Months | 257 | 86 | 54 | 130 |
| THY7 | Thymus | 13885 | 1152 | 20230 | 924 |
| THY8 | 0 Months | 63 | 37 | 26 | 29 |
| THY8 | 12 Months | 114 | 87 | 24 | 34 |
| THY8 | Thymus | 3704 | 182 | 4778 | 87 |
| THY-Y | 0 Months | 42041 | 40447 | 13407 | 33711 |
| THY-Y | 6 Months | 7230 | 3997 | 2239 | 2313 |
| THY-Y | Thymus | 38541 | 14547 | 41682 | 11373 |

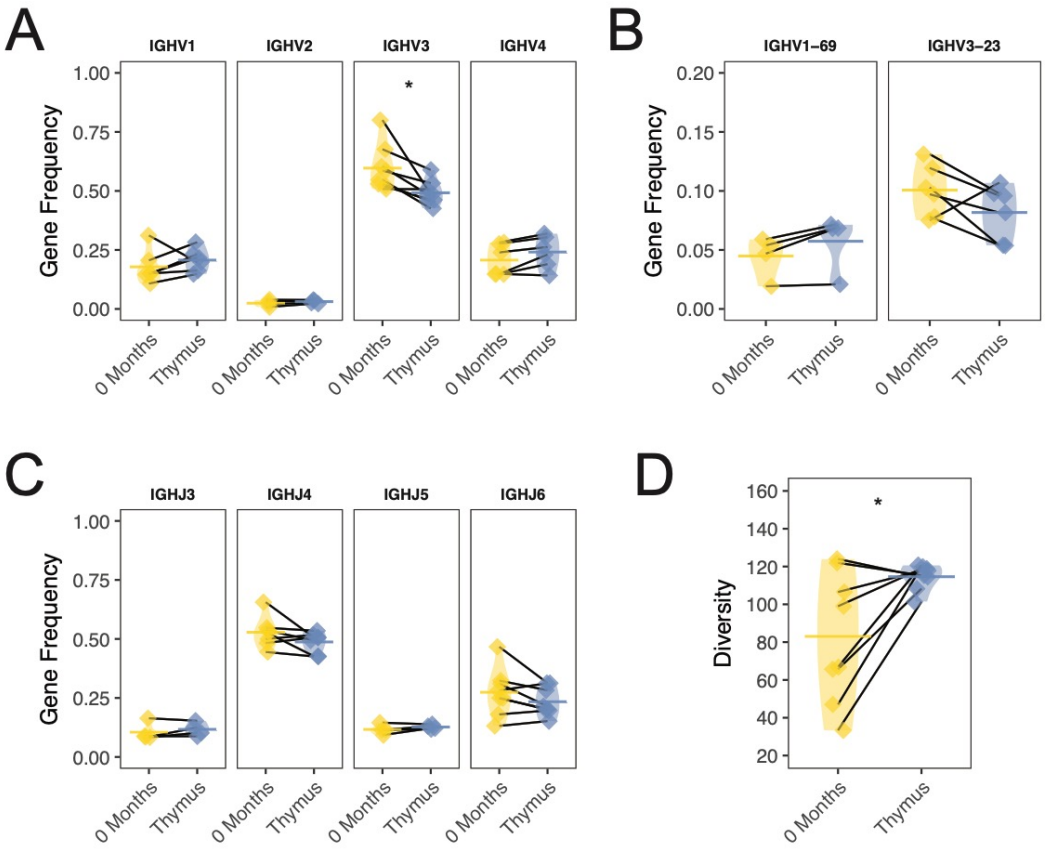



**Fig. S4.** Thymus IgG-expressing B cells are distinct from those in paired circulating peripheral blood. (A) VH family usage differences of circulating repertoires pre-thymectomy and in the thymus for IgG-switched V(D)J sequences. (B) Differences in the usage of VH1-69 and VH3-23 gene segments of circulating repertoires pre-thymectomy and in the thymus for IgG-switched V(D)J sequences. (C) Differences in the use of JH gene segments of circulating repertoires pre-thymectomy and in the thymus for IgG-switched V(D)J sequences. Gene segment usage was only computed for VH family, VH gene segments and JH genes with more than three V(D)J sequences. Horizontal bars show the average gene usage frequency for a given cluster. Frequencies belonging to the same patient are paired with a black line. (D) The clonal distribution of pre-thymectomy and thymus IgG-switched V(D)J sequences is presented as diversity values using a Hill diversity index at q=2 corresponding to the numbers equivalent of Simpson's diversity. Horizontal bars show the average diversity for a given cluster. Frequencies belonging to the same patient are paired with a black line. Statistical differences are shown only when significant (****P < 0.0001; ***P < 0.001; **P < 0.01; *P<0.05).

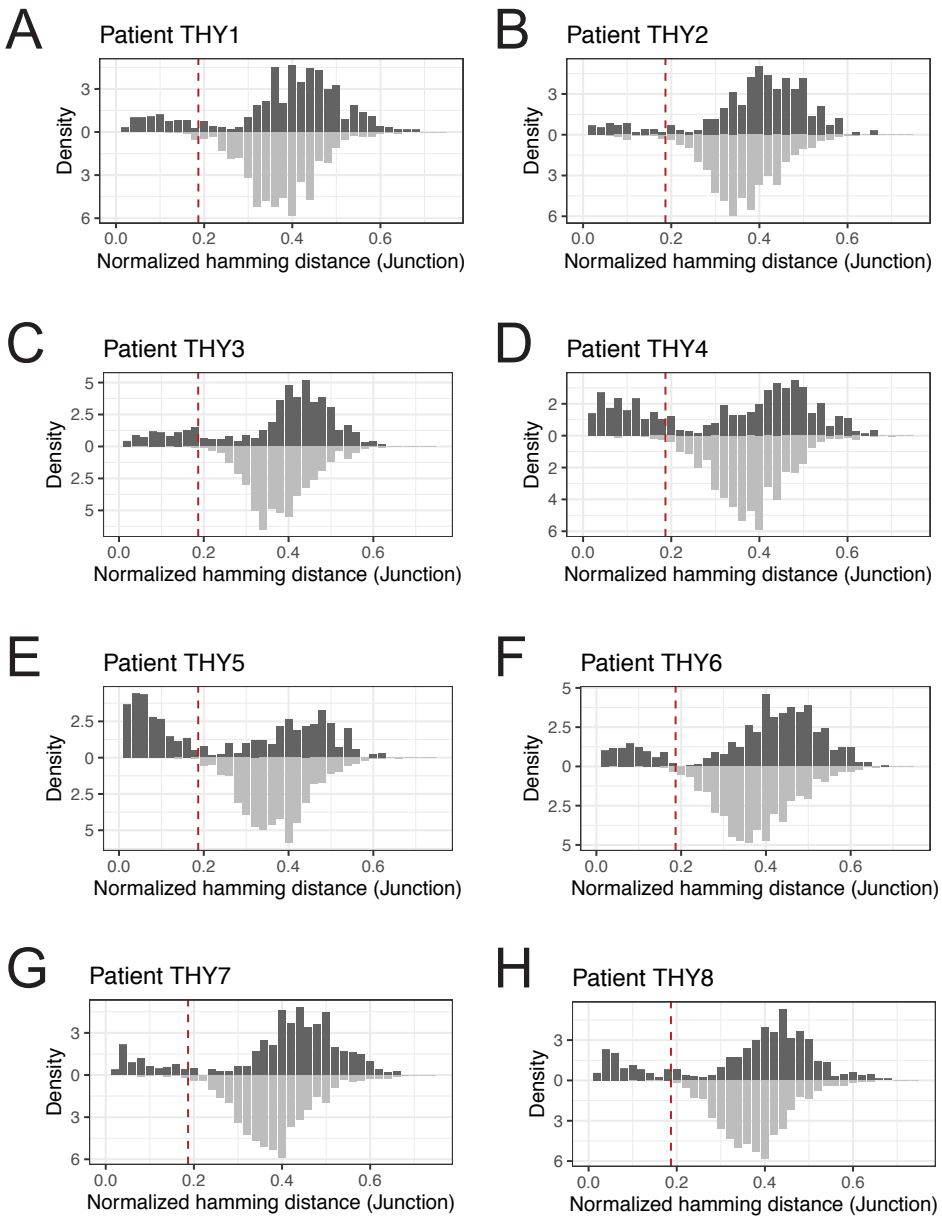



**Fig. S5.** Distance-to-nearest plots used to identify threshold for identifying clonal members from mRNA-based AIRR-Seq of the B cell receptor. The graphs show the distribution of junctional distance-to-nearest distances among sequences using the same V gene and J gene with the same junction length. The minimum between the two modes of this histogram is used to identify the threshold for assigning clusters of sequences as clones. Upper histograms within each plot represent within patient nearest-neighbor hamming distances, while lower histograms within each plot represent between patient distances (which are used to identify the expected distribution of junctional distance-to-nearest distances for sequences that do not belong to clones). The dashed vertical red line indicates the threshold used for clonal clustering.

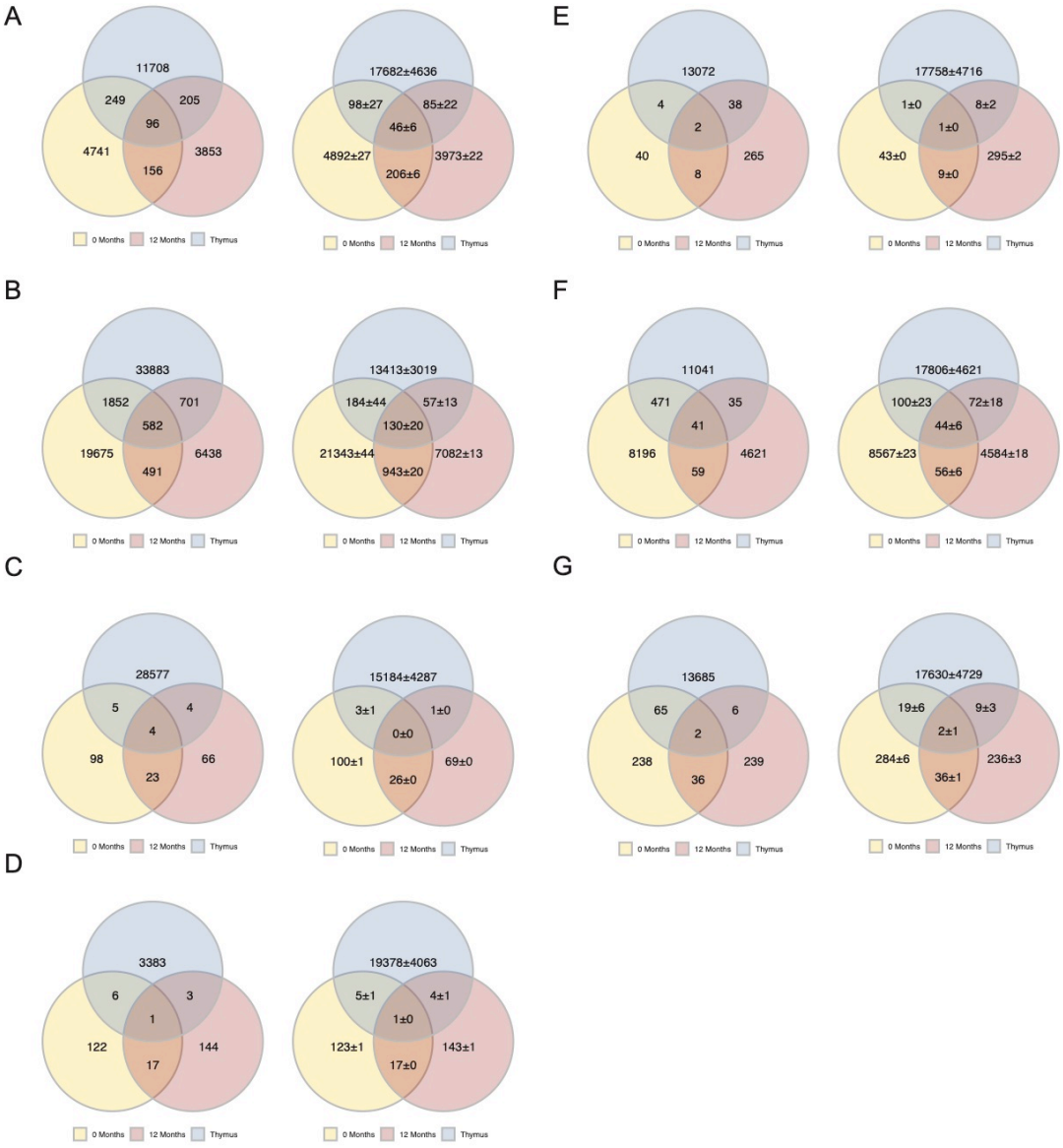



**Fig. S6.** Quantification of shared clones between the thymus, pre-thymectomy and post-thymectomy circulation. Venn-diagrams of shared clones between the thymus, pre-thymectomy and post-thymectomy circulation V(D)J repertoires for patient (A) THY1, (B) THY2, (C) THY3, (D) THY4, (E), THY5, (F) THY6, and (G) THY7. The quantification of observed clonal sharing is shown on the left and corresponding quantification of background clonal sharing is shown to the right. Background sharing was computed by permuting the identity of the thymus used to assess for clonal sharing, that is, the observed clonal sharing of clones from thymus samples derived from different patients with the circulation was calculated. Uncertainty was computed as the standard deviation of values from the corresponding background overlap and is shown using the ± symbol**.**

**Table S2.** Table of V(D)J sequence counts and numbers of clones identified from genomic DNA (gDNA)-based B cell receptor AIRR-Seq.

| PATIENT | TIME | V(D)J SEQUENCES | CLONES |
|---|---|---|---|
| THY6 | 0 Months | 7825 | 7683 |
| THY6 | Thymus | 2413 | 2266 |
| THY7 | 0 Months | 2341 | 2290 |
| THY7 | Thymus | 3430 | 3171 |
| THY2 | 0 Months | 640 | 609 |
| THY2 | Thymus | 1826 | 1749 |

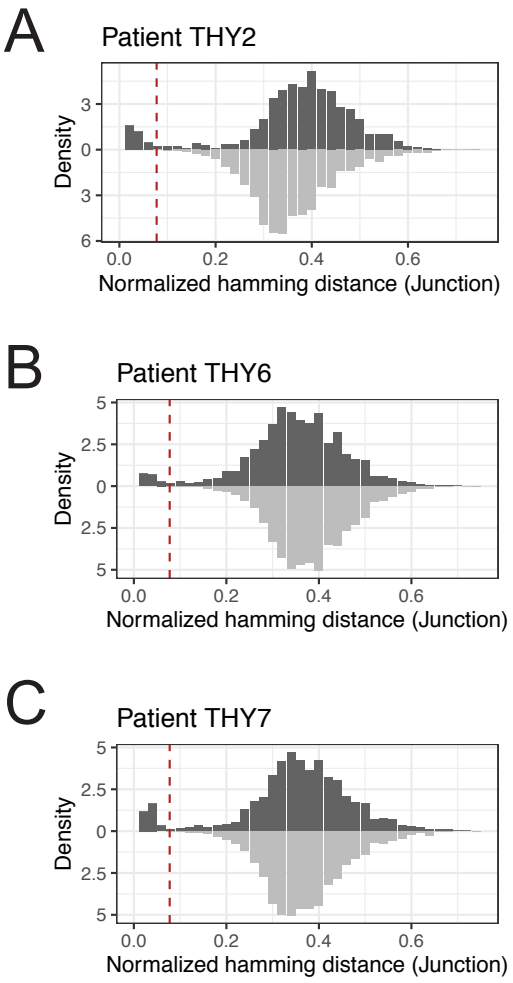



**Fig. S7.** Distance-to-nearest plots used to identify a threshold for assigning clonal members from gDNA-based B cell receptor AIRR-Seq. The graphs (A-C) show the distribution of junctional distance-to-nearest distances among sequences using the same V gene and J gene with the same junction length. The minimum between the two modes of this histogram is used to identify the threshold for assigning clusters of sequences as clones. Upper histograms within each plot represent within patient nearest-neighbor hamming distances, while lower histograms within each plot represent between patient distances (which are used to identify the expected distribution of junctional distance-to-nearest distances for sequences that do not belong to clones). The dashed vertical red line indicates the threshold used for single linkage hierarchical clonal clustering.

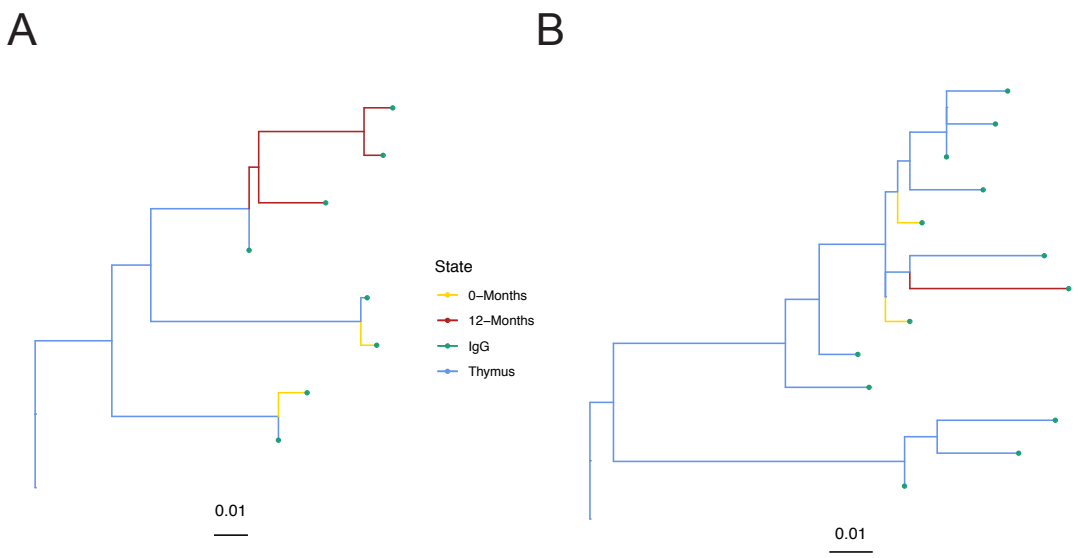
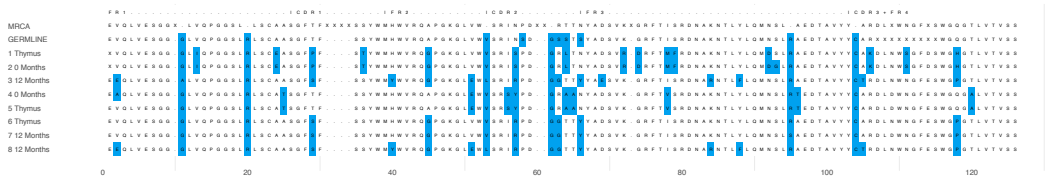



**Fig. S8.** Lineage trees displaying two examples of IgG-expressing B cell clones found in the thymus and circulation before, and 12 months after thymectomy. Tree topologies and branch lengths were estimated using maximum parsimony. Edge lengths are quantified based on intervening somatic hypermutations per site between observed V(D)J sequences per the scale. Tips are colored by antibody isotype, and each internal branch is colored by whether its descendant node was predicted to have occurred in the thymus or pre/post thymectomy circulation using a maximum parsimony algorithm (see Methods).

A




B

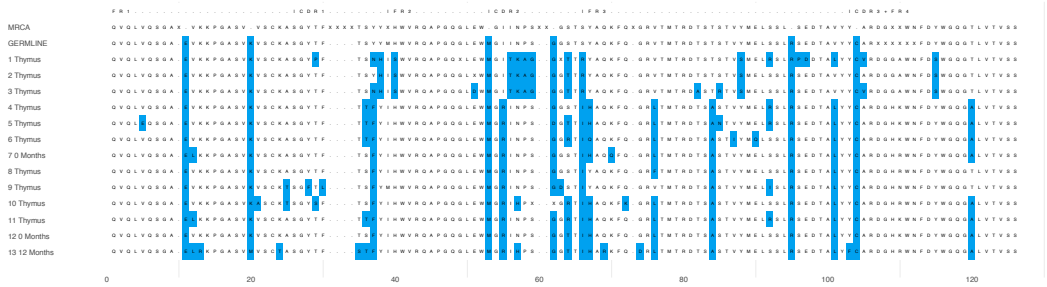



**Fig. S9.** Amino acid sequence alignments of two representative examples of IgG-expressing B cell clones found in thymus before thymectomy and at 12 months after thymectomy. The clones (A, B) were both derived from patient THY2. Replacement mutations are highlighted (blue) relative to the most recent common ancestor (MRCA).

**Table S3.** Phylogenetic analysis of IgG sequences from thymus and circulation. "Included clones" indicates the number of clones which contained both thymus and circulation IgG sequences at the time of thymectomy and contained more than two sequences that were either distinct or found in different samples. "Thymus/circ. ratio" is the ratio of total sequences obtained from the thymus vs. circulation at the date of thymectomy. "Observed SP" is the proportion of predicted changes that occurred from the thymus to the circulation in observed lineage trees (detailed in Ref. 60). "Random SP" is the proportion of predicted changes that occurred from thymus to circulation in lineage trees with permuted tip locations. "Mean $\delta$" shows the mean difference between observed and random SP values. The significance of this difference (P value) is the proportion of 1000 permutation replicates in which $\delta \leq 0$. All numbers reported except "Included Clones" are rounded to two significant digits.

| PATIENT | INCLUDED CLONES | THYMUS/CIRC. RATIO | OBSERVED SP | RANDOM SP | MEAN $\delta$ | P-VALUE |
|---|---|---|---|---|---|---|
| THY1 | 44 | 2.7 | 0.61 | 0.63 | -0.015 | 0.63 |
| THY2 | 550 | 2.4 | 0.68 | 069 | -0.011 | 0.83 |
| THY5 | 2 | 260 | 0.8 | 0.61 | 0.19 | 0.27 |
| THY6 | 80 | 1.9 | 0.67 | 0.61 | 0.068 | 0.011 |
| THY7 | 37 | 58 | 0.93 | 0.87 | 0.056 | 0.01 |

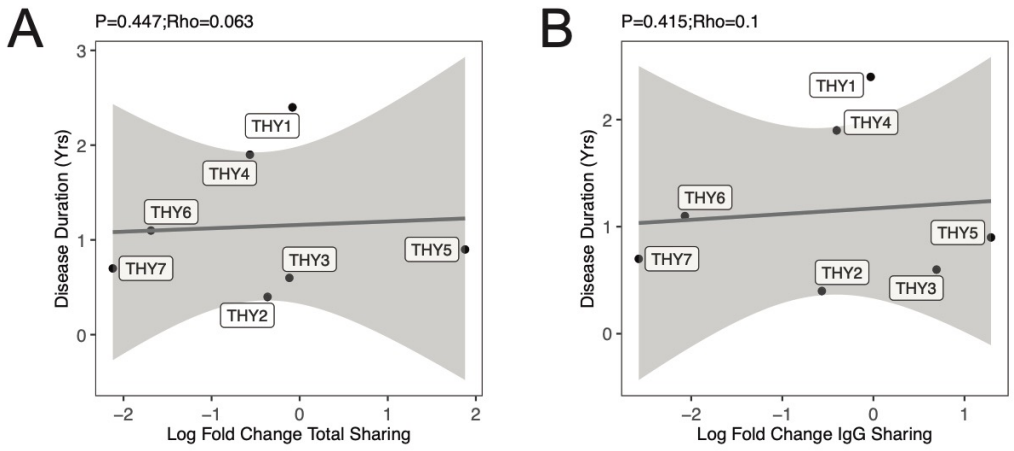



**Fig. S10.** Correlation of the log fold change in clonal sharing between the thymus and circulation as assessed with a Bray Curtis index (on the x-axis) and duration of disease prior to thymectomy in years (y-axis). Correlation is shown for the total repertoire (A) or the IgG-switched repertoire only (B).