SUPPLEMENTAL MATERIAL

Moffitt et al., https://doi.org/10.1084/jem.20160894

The Journal of Experimental Medicine

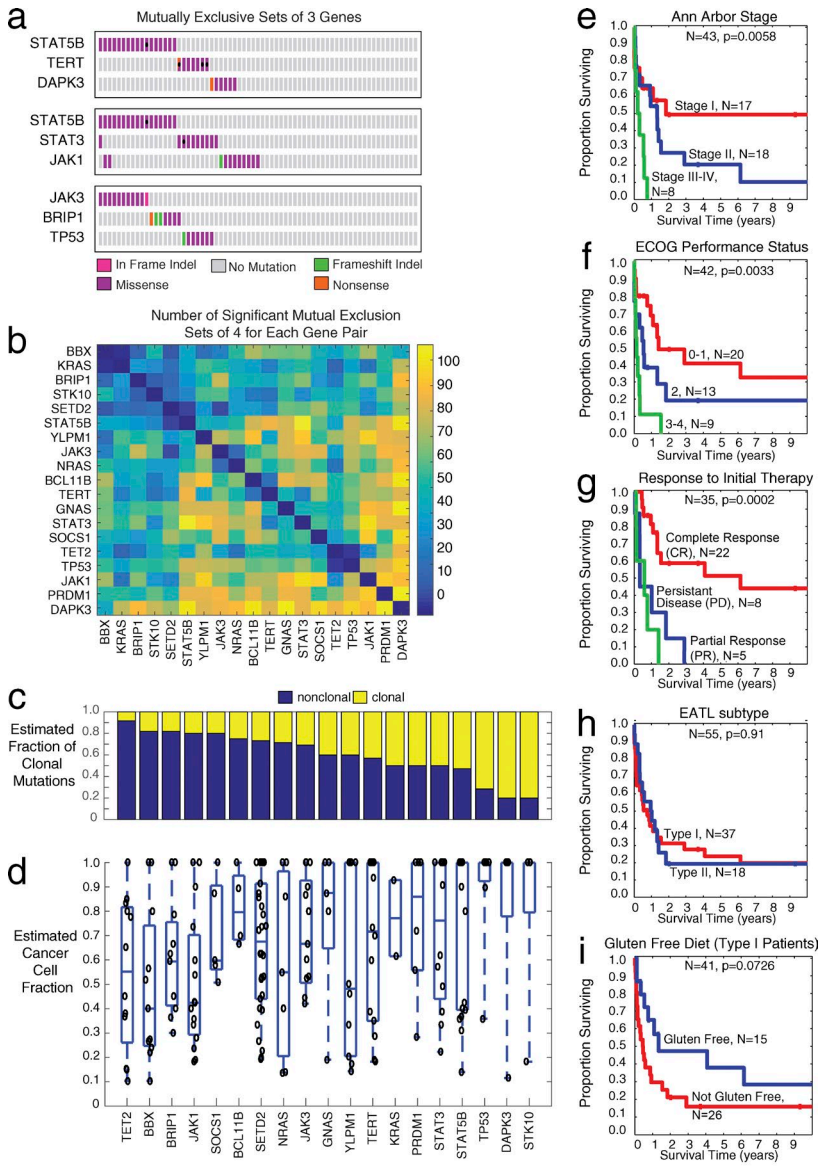

Figure S1. **Analysis of mutual exclusion, allele frequencies, and clinical outcomes in EATL.** (a) Representative mutually exclusive sets of three genes. Rows are individual genes, and columns represent EATL samples. Mutation type is indicated by the box color. Samples with multiple mutations in the same gene of different types are colored with both colors. Black dots indicate multiple mutations in the same sample. Synonymous mutations were removed before analysis. All sets have FDR <0.05 in the weighted row exclusivity test. Additional sets are listed in Table S2. Mutations from 69 EATL samples were tested. (b) Gene sets of four genes were tested for mutual exclusion with the weighted row exclusivity test. Each gene pair from the EATL driver genes was counted for how many times it occurred among the significant sets (FDR <0.05). Yellow indicates higher evidence for mutual exclusion, and blue indicates less evidence for mutual exclusion, as shown in the color bar. Mutations from 69 EATL samples were tested. (c) Analysis of cancer cell fraction for driver gene mutations. Cancer cell fraction is calculated as observed allele frequency, corrected for tumor purity estimate, with assumption of heterozygous somatic mutation. Fraction of observed mutations in a gene that are clonal. Mutations are considered "clonal" if cancer cell fraction is >0.9. Genes are sorted by fraction of observed mutations that are clonal, with *STK10* having the highest fraction of clonal events, and *TET2* having the lowest fraction. Mutations from 69 EATL samples are included. (d) Distribution of cancer cell fractions for mutations in each driver gene. Each point in the boxplot represents one observed mutation in one patient. Genes are sorted by fraction of observed mutations that are clonal. Mutations from 69 EATL samples are included. (e) Kaplan–Meier plot separating the cohort by Ann Arbor stage. $n = 43$. Log-rank test, P = 0.0058. Stage I: $n = 17$; stage II: $n = 18$; stage III–IV: $n = 8$. (f) Kaplan–Meier plot separating the cohort by Eastern Cooperative Oncology Group (ECOG) performance status. $n = 42$. Log-rank test, P = 0.0033. 0–1: $n = 20$; 2: $n = 13$; 3–4: $n = 9$. (g) Kaplan–Meier plot separating the cohort by response to initial therapy. $n = 35$. Log-rank test, P = 0.0002. Complete response (CR): $n = 22$; partial response (PR): $n = 5$; persistent disease (PD): $n = 8$. (h) Kaplan–Meier plot separating the cohort by EATL subtype. $n = 55$. Log-rank test, P = 0.91 (not significant). Type I: $n = 37$; type II, $n = 18$. (i) Kaplan–Meier plot separating the type I cohort by gluten-free diet. $n = 41$. Log-rank test, P = 0.0726. Gluten free: $n = 15$; not gluten free: $n = 26$.

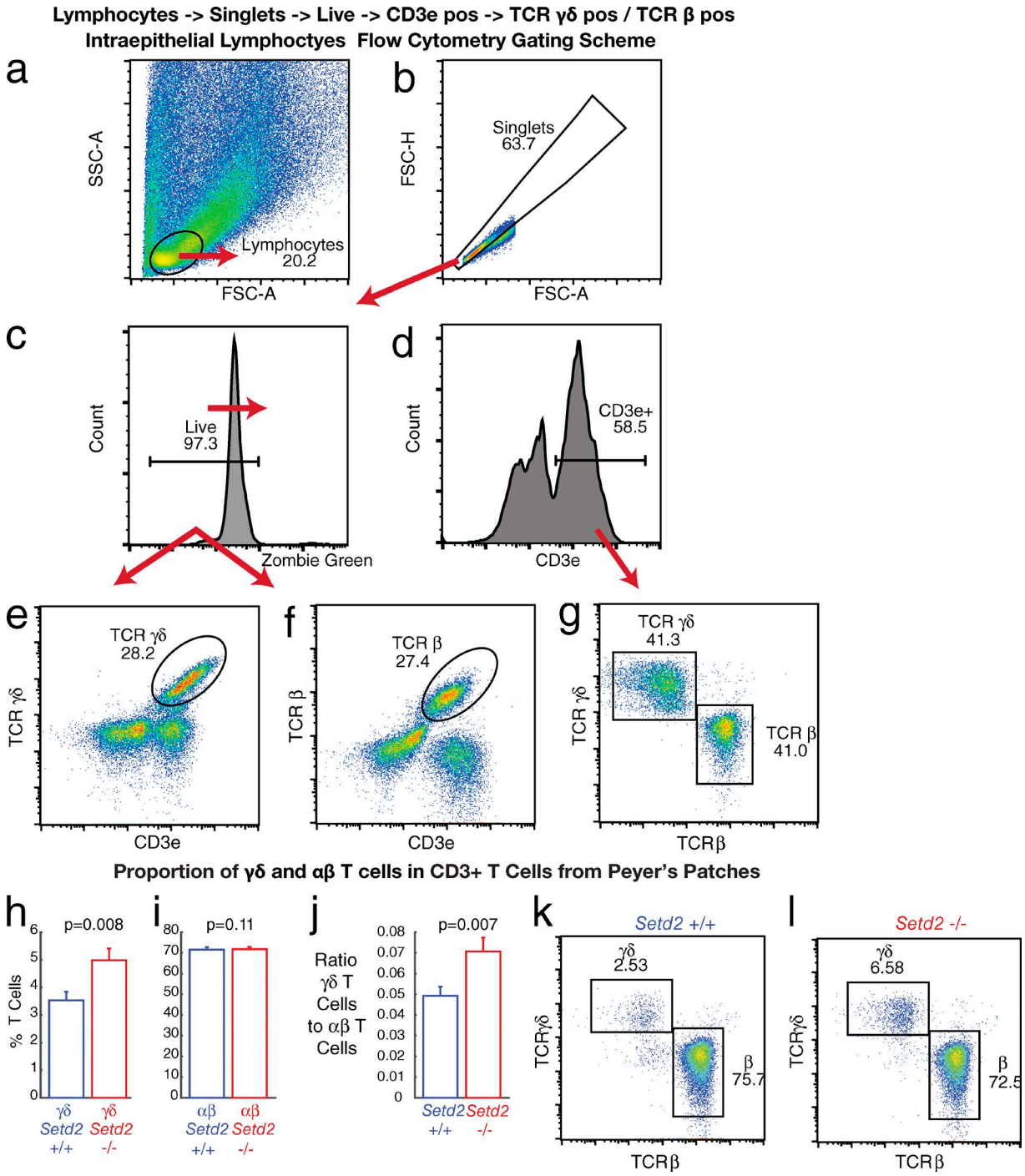

Figure S2. **Flow cytometry scheme for IELs and proportion of αβ and γδ cells in Peyer's patches.** (a–g) Diagram shows the flow cytometry scheme for IEL T cell populations in the *Setd2* wild-type and knockout mice. (a) Gating on the lymphocyte population. (b) Gating on singlets. (c) Gating on live cells. (d) Gating on CD3e-positive cells. (e and f) Gating on γδ and αβ T cells using both the TCR marker and CD3e (for quantification). (g) Gating on γδ and αβ T cells after gating on CD3e for all live lymphocytes (for sorting and visualization). (h–l) Proportion of αβ and γδ T cells in Peyer's patches. (h and i) Percentage of T cells that are $CD3e^{pos}$ and $TCR\gamma\delta^{pos}$ (h) or $CD3e^{pos}$ and $TCR\beta^{pos}$ (i) from Peyer's patch tissue in *Setd2*$^{+/+}$ (left, blue) and *Setd2*$^{-/-}$ (right, red) mice. (j) Ratio of $\gamma\delta^{+}$ to $\alpha\beta^{+}$ T cell proportions in *Setd2*$^{+/+}$ (left, blue) and *Setd2*$^{-/-}$ (right, red) mice. Bars show standard error of the mean. Two-way ANOVA tests were done for comparisons across genotype, correcting also for experimental batch. P-values for genotype differences are 0.008 (h), 0.11 (i), and 0.007 (j). Data are shown from nine different experimental batches with a mean of five mice per batch. Total number of wild-type *Setd2*$^{+/+}$ mice = 26. Total number of knockout *Setd2*$^{-/-}$ mice = 23. (k and l) Representative flow cytometry plots showing the increase in γδ T cells and decrease in αβ T cells in Peyer's patches from *Setd2*$^{+/+}$ (k) mice to *Setd2*$^{-/-}$ (l) mice. Plots are representative of results found comparing 26 *Setd2* wild-type mice and 23 *Setd2*-null mice.

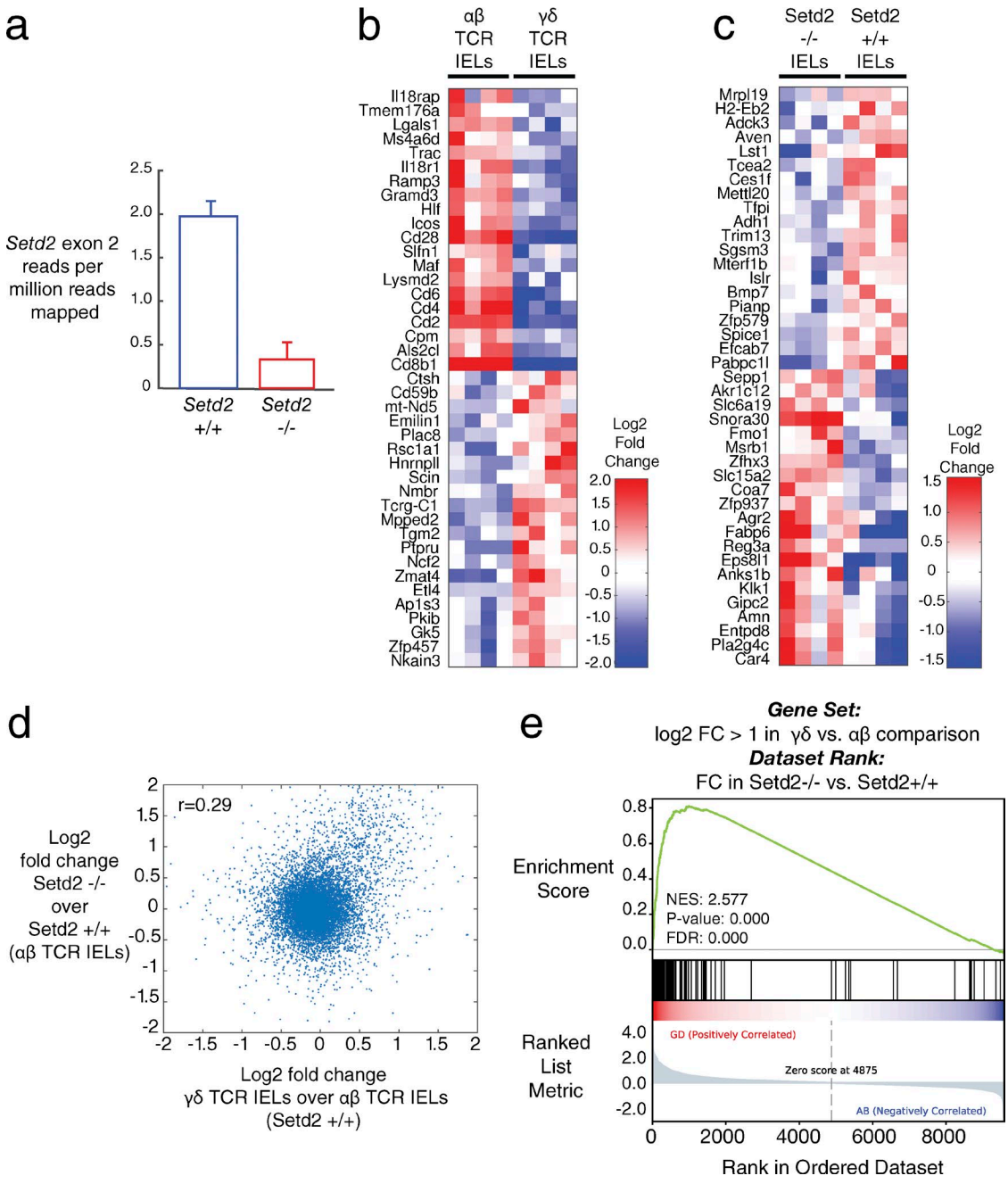

Figure S3. **RNA-sequencing analysis of IELs.** Comparisons of γδ and αβ T cell populations and *Setd2*$^{+/+}$ and *Setd2*$^{-/-}$ mice (n = 4 mice, eight total samples). (a) Normalized read count for exon 2 of *Setd2*, which is deleted in *Setd2*-null mice. Reads overlapping the exon were counted and normalized by total number of mapped reads/1,000,000. Bars show standard error of the mean. Four samples from each genotype were averaged. $P < 0.001$ (Student's t test). (b and c) Heat map of RNA-sequencing log2 FPKM data for eight intraepithelial samples divided by TCR type (b) and genotype (c). Top 20 genes in each direction are shown, as determined by filtering for $P < 0.1$ between the two groups and top genes selected by greatest fold change between two groups. (d) Scatterplot of log2 fold change between γδ and αβ IELs in *Setd2*$^{+/+}$ versus log2 fold change between *Setd2*$^{-/-}$ and *Setd2*$^{+/+}$ αβ IELs. r = 0.29, correlation test, $P \ll 0.001$. (e) Gene set enrichment plot showing dataset ranked by fold change between genotypes, and gene set selected as genes with log2 fold change >1 in γδ versus αβ IELs. Green line shows enrichment statistic as gene list is traversed. Black lines indicate the rank of the gene set of interest in the ranked list. GSEA KS-test $P \ll 0.001$.

Tables S1–S10 are available as tabs in an Excel file. Table S1 lists the variants in EATL significantly mutated genes. Table S2 lists the mutual exclusion sets. Table S3 lists the copy-number alterations in EATL. Table S4 lists the HLA genotypes for DQA1 and DQB1 genes. Table S5 lists the clinical and pathological features of the EATL cases. Table S6 lists the differentially expressed genes between type I and type II. Table S7 lists the RNA-sequencing gene expression for the IELs. Table S8 lists the individual T cell proportions for IELs and Peyer's patches. Table S9 lists the summary statistics for exome and RNA sequencing. Table S10 lists the oligo sequences used for rRNA depletion.