IL32 Western Blot. The extracted cellular proteins were fractionated and transferred onto nitrocellulose membranes by electro-blotting. After being blocked in 5% milk in TBS supplemented with 0.05% Tween 20 for 1 hour at RT, the membranes were incubated with purified mouse anti-human IL-32αβδγ antibody IgG1, κ (Clone KU32-52, 1:1,000, Biolegend, San Diego, CA) overnight at 4 °C in 5% milk in TBST. The membranes were washed with TBST and then incubated with 1:1,000 HRP-conjugated goat anti-mouse polyclonal secondary antibodies (Cell Signaling, Beverly, MA) for 1 hour at RT. After washing with TBS, antibody binding was visualized using enhanced chemiluminescence (the SuperSignal™ West Pico Chemiluminescence Substrate, Thermo Fisher Scientific, Rockford, IL). The equivalent loading of proteins in each well was confirmed by Ponceau staining and actin control.

METHODS-ONLY REFERENCES

SUPPLEMENTARY FIGURE LEGENDS

Supplementary Fig. 1. Expression profiles of cell-specific surface markers and T-helper cell cytokines. Th, T-helper cells; T-reg, regulatory T cells.

Supplementary Fig. 2. Mutation validated by transcriptional expression in RNA-seq. a) Summary of the validation status for all somatic mutations identified from 32 patients for which RNA-seq data were available. **b**) A breakdown of the validation status by mutation types. The percent of total values were labeled for all pie charts. Gene not expressed, no RNA-seq reads were mapped at the particular mutation site; Mutant allele

not expressed, there were RNA-seq reads mapped at the particular mutation site but no variant supporting reads were detected; Mutant allele expressed, at least one variant supporting RNA-seq reads were detected; VAF, variant allele fraction.

Supplementary Fig. 3. The mutation rate in SS patients. The somatic non-silent mutation rate estimated by whole-exome sequencing.

Supplementary Fig. 4. The UVB mutation signature in SS patients. a) The heatmap on the left shows the mutation signatures across 37 SS patients in the discovery cohort. Samples are ordered by the fraction of UVB signatures. Three prominent signatures, out of 21 cancer mutation signatures (Covington and Wheeler, in preparation) were observed. The top two tracks indicate total mutation frequency (0-10.5, from black to white) and the race: Orange: Caucasian; Brown: African American. The figure on the right shows the relative proportions of each base change that characterize each of the signatures shown in the heatmap. Each signature is displayed according to the 32 sequence contexts immediately 3' and 5' to the mutated base. The numbers on the y-axis indicate the fraction of mutations attributed to a specific mutation type. **b**) The number of mutations in each patient's tumor that are attributable to CpG and UVB signatures. Blue: UVB, Red: CpG. See Online Methods (Mutation Signatures) for computation of mutation signature count. Virtually all SS samples exhibited some level of CpG (NCG > NTG) mutation at a nearly constant rate with respect to overall mutation rate. Many SS samples had mutations corresponding to UVB exposure, which was the principal factor driving increasing mutation rate. Anova analysis indicated that 85% of the variability in mutation

counts is explained by the UVB signature component. **c**) Correlation analysis of UVB signatures in SS patients with prior radiation therapies**.** DNM, di-nucleotide mutations.

Supplementary Fig. 5. Schematic representation of somatic mutations identified in other potential drivers in combined cohorts (n = 105).

Supplementary Fig. 6. Conservation analysis of *CARD11* mutation hotspots. The protein sequences of all species were downloaded and aligned using Uniprot (see URLs). Two sequence segments (**a, b**) were selected to show the mutation hotspots. The mutations sites were marked and the amino acid changes were indicated. The two numbers in the parentheses indicate the number of mutations identified in the discovery cohort and extension cohort, respectively.

Supplementary Fig. 7. The *PLCG1* mutations in the combined cohorts. The allele fractions were shown for all mutations.

Supplementary Fig. 8. Survival analysis for recurrent somatic alterations that have potential pathogenic significance. The Kaplan-Meier survival plots were shown and the P values were calculated by Log-Rank test.

Supplementary Fig. 9. Correlation analysis between DNA copy number alterations and gene expression levels. One-way ANOVA was used to compare the means of groups and student *t*-test was used to calculate the P values between two groups. The

DNA copy number status was inferred by Nexus analysis of the SNP array data. The expression levels were normalized FPKM values calculated by Cufflinks using RNA-seq data. N.S., not statistically significant; LOH, loss of heterozygosity; HD, homozygous deletion; UPD, copy-neutral LOH.

Supplementary Fig. 10. Correlation analysis between expression of *IL32* and *IL2RG, CARD11*. RNA expression data was used for analysis. Numbers on Y axis are the normalized gene expression levels (FPKM values) calculated by the Cufflinks algorithm using RNA-seq data. Bivariate fit was made by JMP algorithm.

Supplementary Fig. 11. The landscape of TCR repertoire. RNA-seq coverage tracks were shown for the TCR-Vβ (**a**) and Vα (**b**) loci. The T-cell samples from five healthy donors were also included for comparison purpose. The recurrently rearranged TCR-Vβ and Vα genes were indicated by arrows and their names were displayed underneath. The scale was adjusted for a better visualization. The data range for all control samples was set to 0–100; the data range for samples with polyclonal TCR-Vβ or Vα expression was set to 0–500; the data range for samples with dominant TCR-Vβ or Vα expression was set to 0–5,000.

Supplementary Fig. 12. Survival analysis for the combined cohorts. The Kaplan-Meier survival plots were shown and the P values were calculated by Log-Rank test.