Supplementary Information

Supplementary Methods

Supplementary References

Supplementary Figure Legends

Supplementary Table Legends

Supplementary Methods

WES and processing of variants

WES was performed using nine FFPE samples from eight patients with SPTCLs (including one sample of recurrence from patient SP03) diagnosed at SNUH (discovery set). Matched non-neoplastic tissue samples were available from two patients (SP01 and SP04). Genomic DNA was extracted from the samples and quantified using a Qubit fluorometer (Invitrogen, Eugene, OR, USA). Libraries were generated using the hybrid capture method and SureSelect^{XT} Human All Exon V5 (Agilent Technologies, Santa Clara, CA, USA). Libraries were sequenced on an Illumina HiSeq system (Illumina, San Diego, CA, USA) using the paired-end 2 × 101-bp read option (Theragene ETEX Bio Institute; Suwon, Republic of Korea). We used Burrows-Wheeler Aligner (BWA) v.0.7.12[1] to map sequencing reads from tumors and germline samples to the reference genome hg19/NCBI GRCh 37; Picard (v1.92) was used for deduplication, and the Genome Analysis Tool Kit (GATK, v2.3-9)[2] was used for local re-alignment. Sequencing metrics are summarized in **Supplementary Table S3**.

Single nucleotide variants and small insertions and deletions (indels) were identified by MuTect (v1.1.4)[3] and Indelocator (v2.3-9), followed by variant annotation using SnpEff (v.4.2). We discarded all low-quality variants, including those with variant allele frequency (VAF) of < 5% or supporting reads ≤ 10; only non-synonymous variants were included for further analysis.

For seven samples without matched germline controls, an additional variant filtering process was implemented to reduce contamination by plausible germline variants and to focus on biologically meaningful alterations. Variants with a VAF of 45–55% or > 95% were excluded, and those with a minor allele frequency (MAF) ≥ 0.01% in gnomAD East Asian (gnomAD_EAS) databases, or MAF ≥ 0.1% in the Korean Reference Genome Database

(KRGDB), were filtered out; however, we manually reviewed the whole coding sequence of *HAVCR2* to search for pathogenic variants. Genes with a large size, very low expression, or very late replication times (e.g., *MUC16, MUC5B, TTN)* were removed.[4] In addition, we generated an unmatched panel of normal (PoN) samples by merging the WES results from two non-neoplastic tissue samples (from patients SP01 and SP04), with available in-house sequencing data from three additional germline tissue samples of patients with other types of lymphoid diseases. Then, we discarded all variants found in both the PoN and the discovery set. For missense mutations with available functional prediction data (by PolyPhen-2 HVAR, PolyPhen-2 HDIV, SIFT, MutationTaster, and LRT), only variants strongly predicted to have a deleterious effect were accepted (e.g., predicted deleterious by more than two out of three tools, three out of four tools, or four out of five tools).

Targeted sequencing and processing of variants

To unveil the mutational landscape of SPTCLs, MFs, PGDTCLs, and LPs, we created a customized panel comprising 208 genes (**Supplementary Table S4**) selected by the following criteria: genes with mutations found in more than two patients in the discovery set of this study or previously published SPTCL cohorts[5-8] (e.g., *HAVCR2, PIAS3, PLCG2*); genes with mutations found in at least one patient in the discovery set and known to have functional implications in inflammatory responses or T-cell biology (e.g., *IFNL2, F5, GDF1*); genes with mutations previously reported in CTCLs[9]; and other genes implicated in the pathogenesis of lymphoid neoplasms (e.g., *RHOA, TET2, MYD88*).

TGS was performed for 32 patients: 20 with SPTCLs, 8 with MF, 3 with PCGDTCLs, and 1 with LP. DNA extraction and library preparation and sequencing were performed at Macrogen Incorporate (Seoul, Republic of Korea). The quantity and purity of DNA were assessed using Qubit (Invitrogen) and NanoDrop (Thermo Fisher Scientific,

Waltham, MA, USA). Sequencing libraries were prepared using a SureSelect^{XT} Target Enrichment Kit (Agilent Technologies), and paired-end sequencing (150 bp × 2) was performed on a NextSeq 500 system (Illumina), reaching a mean coverage depth of 589× (**Supplementary Table S5**). Sequencing reads were mapped to the reference genome hg19/NCBI GRCh 37 using BWA, and variants were called using GATK.

Matched germline samples were not subjected to TGS, and only a single variant caller algorithm was applied. Therefore, an additional variant filtering process was carried out to eliminate possible false positive calls. Only non-synonymous point mutations were included in further analyses, and we filtered out variants with allele frequency below 3% or with altered read counts of less than 10. Other than *HAVCR2*^{Y82C}, variants with MAF ≥ 0.1% in the ExAC Database[10] were removed. Identical point mutations recurrently detected on more than 5 out of 32 tested samples were considered putative false positives and were discarded. Finally, only missense variants predicted to be benign by more than three out of four tools (PolyPhen-2 HDIV, SIFT, MutationTaster, and PROVEAN) were removed.

RNA-seq

RNA-seq was performed on eight samples in the discovery set, which included four patients with the *HAVCR2*^{Y82C} genotype and four with *HAVCR2*^{WT} genotype. Sequencing libraries were prepared using a TruSeq^{TM} RNA Exome Kit (Illumina), and were sequenced on the Illumina HiSeq platform (Theragene ETEX Bio Institute) using the paired-end 2 × 100-bp option. After quality assessment of raw FASTQ files using FastQC (v.0.11.5), adapter sequences in sequencing reads were trimmed using Trimmomatic (v.0.36)[11]. Reads were aligned using STAR aligner (v.2.6.0a)[12] and were quantified using RSEM (v.1.3.1)[13] as transcripts per million (TPM). Sequencing statistics are described in **Supplementary Table S6**.

Direct sequencing

For those who were not suitable for high-throughput sequencing, direct sequencing of *HAVCR2* exon 2 was performed, covering all of the previously reported variants in patients with SPTCL (Y82C, I97M, T101I).[5-7] DNA was extracted from FFPE tissue samples using a Maxwell 16 FFPE Plus Tissue LEV DNA Purification Kit (Promega, Madison, WI, USA). Polymerase chain reaction (PCR) primers specific for *HAVCR2* exon 2 were designed using Primer-BLAST[14] (**Supplementary Table S7**), and nested PCR was performed using Ex Taq (TaKaRa Bio, Shiga, Japan). Direct sequencing was performed using an ABI3730xl DNA analyzer (Applied Biosystems, Waltham, MA, USA).

Immunohistochemistry and T-cell clonality test

IHC results were retrieved from the pathology report of each participating institution. During a central review process conducted SNUH, immunostainings for TCRβF1 (TCR1151, Thermo Fisher Scientific), TCRγ (gamma 3.20, Thermo Fisher Scientific), T-cell clonality test using an IdentiClone *TCRG* Gene Clonality Assay (Invivoscribe Technologies Inc., San Diego, CA, USA) were performed if necessary.

To validate the findings from the gene expression analysis, we performed IHC on the 4-μm-thick whole sections of FFPE tissue samples: CCR4 (HPA031613, Atlas Antibodies, Stockholm, Sweden) on Bond-Max Autostainer (Leica Biosystems, Wetzlar, Germany), FoxP3 (236 A/E7, Abcam, Cambridge, UK) and pSTAT3 (#9145, Cell Signaling Technology, Danvers, MA) on Ventana Benchmark XT automated stainer (Ventana Medical System, AZ, USA). CCR4, Foxp3, and pSTAT3 immunostains were digitally scanned using an Aperio AT2 (Leica Biosystems), and positive cells in representative tumor areas were quantified by

Nuclear V9 algorithm of ImageScope (Aperio Technologies, Vista, CA, USA). The positivity for each marker was defined as the percentage of positive cells in the analyzed area.

Double-staining

Double-staining was carried out in selected cases (n = 7) for whole tumor sections using antibodies for CCR4 (HPA031613, Atlas Antibodies, Stockholm, Sweden), FoxP3 (236 A/E7, Abcam, Cambridge, UK) and GATA3 (L50-823, Cell Marque, Rocklin, CA, USA). We used a double-staining oDAB-uRed kit (Ventana) on the BenchMark XT Slide automated system (Ventana) according to the manufacturer's protocol. Foxp3 and GATA3 were visualized with brown color using an OptiView DAB IHC detection kit (Ventana), and CCR4 was visualized with red color using ultraView Universal AP Red kit (Ventana).

Supplementary References

1. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
2. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res*. 2010;20(9):1297–1303.
3. Cibulskis K, Lawrence MS, Carter SL, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat Biotechnol*. 2013;31(3):213–219.
4. Lawrence MS, Stojanov P, Polak P, et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. *Nature*. 2013;499(7457):214–218.

5. Li Z, Lu L, Zhou Z, et al. Recurrent mutations in epigenetic modifiers and the PI3K/AKT/mTOR pathway in subcutaneous panniculitis-like T-cell lymphoma. *Br J Haematol*. 2017;181(3):406–410.
6. Gayden T, Sepulveda FE, Khuong-Quang D-A, et al. Germline HAVCR2 mutations altering TIM-3 characterize subcutaneous panniculitis-like T cell lymphomas with hemophagocytic lymphohistiocytic syndrome. *Na. Genet.* 2018;50(12):1650–1657.
7. Polprasert C, Takeuchi Y, Kakiuchi N, et al. Frequent germline mutations of HAVCR2 in sporadic subcutaneous panniculitis-like T-cell lymphoma. *Blood Adv*. 2019;3(4):588–595.
8. Fernandez-Pol S, Costa HA, Steiner DF, et al. High-throughput Sequencing of Subcutaneous Panniculitis-like T-Cell Lymphoma Reveals Candidate Pathogenic Mutations. *Appl Immunohistochem Mol Morphol*. 2019;27(10):740–748.
9. da Silva Almeida AC, Abate F, Khiabanian H, et al. The mutational landscape of cutaneous T cell lymphoma and Sézary syndrome. *Nat Genet.* 2015;47(12):1465–1470.
10. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536(7616):285–291.
11. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014;30(15):2114–2120.
12. Dobin A, Davis CA, Schlesinger F, et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics*. 2012;29(1):15–21.
13. Li B, Dewey CN. RSEM: accurate transcript quantification from RNA-Seq data with or without a reference. *BMC Bioinformatics*. 2011;12(1):1–16.
14. Ye J, Coulouris G, Zaretskaya I, et al. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012;13(1):134.

Supplementary Figure legends

Supplementary Figure S1. Flow chart describing the study design and sequencing methods.

Supplementary Figure S2. Mutational landscape assessed by WES in the discovery set.
WES was performed in nine samples from eight patients, including a sample of recurrence from patient SP03. Comparison of SP03-1 (initial sample) and SP03-2 (sample of recurrence) revealed no significant sequential acquisition of somatic mutations; all mutations were shared between two samples, albeit in low allele frequency (asterisk).

Supplementary Figure S3. Detection of *DDX11* mutations in patients with SPTCL.
(a) SP4 harbored two point mutations in *DDX11* on the same allele. **(b)** Distribution of *DDX11* mutations discovered in the current study population.

Supplementary Figure S4. Summary of the clinicopathological characteristics of patients with SPTCL.
Characteristics of the 53 SPTCLs included in this study are summarized.

Supplementary Figure S5. Survival analyses according to bone marrow status and hematopoietic stem cell transplantations status.
No significant differences in RFS were observed according to bone marrow status and hematopoietic stem cell transplantation status by log-rank test.

Supplementary Figure S6. Overall survival according to various clinicopathological

features and risk stratification using the score system.

Overall survival of patients were compared, which showed no significant differences according to various clinicopathological features.

Supplementary Table Legends

Supplementary Table S1. Clinicopathological features of patients with subcutaneous panniculitis like T-cell lymphoma in Republic of Korea

Supplementary Table S2. Patients with SPTCL complicated by HLH or HLH-like systemic illness

Supplementary Table S3. Whole exome sequencing statistics and quality metrics

Supplementary Table S4. List of genes selected for targeted gene sequencing (TGS)

Supplementary Table S5. Targeted gene sequencing statistics and quality metrics

Supplementary Table S6. mRNA sequencing statistics

Supplementary Table S7. *HAVCR2* polymerase chain reaction primer and experimental condition

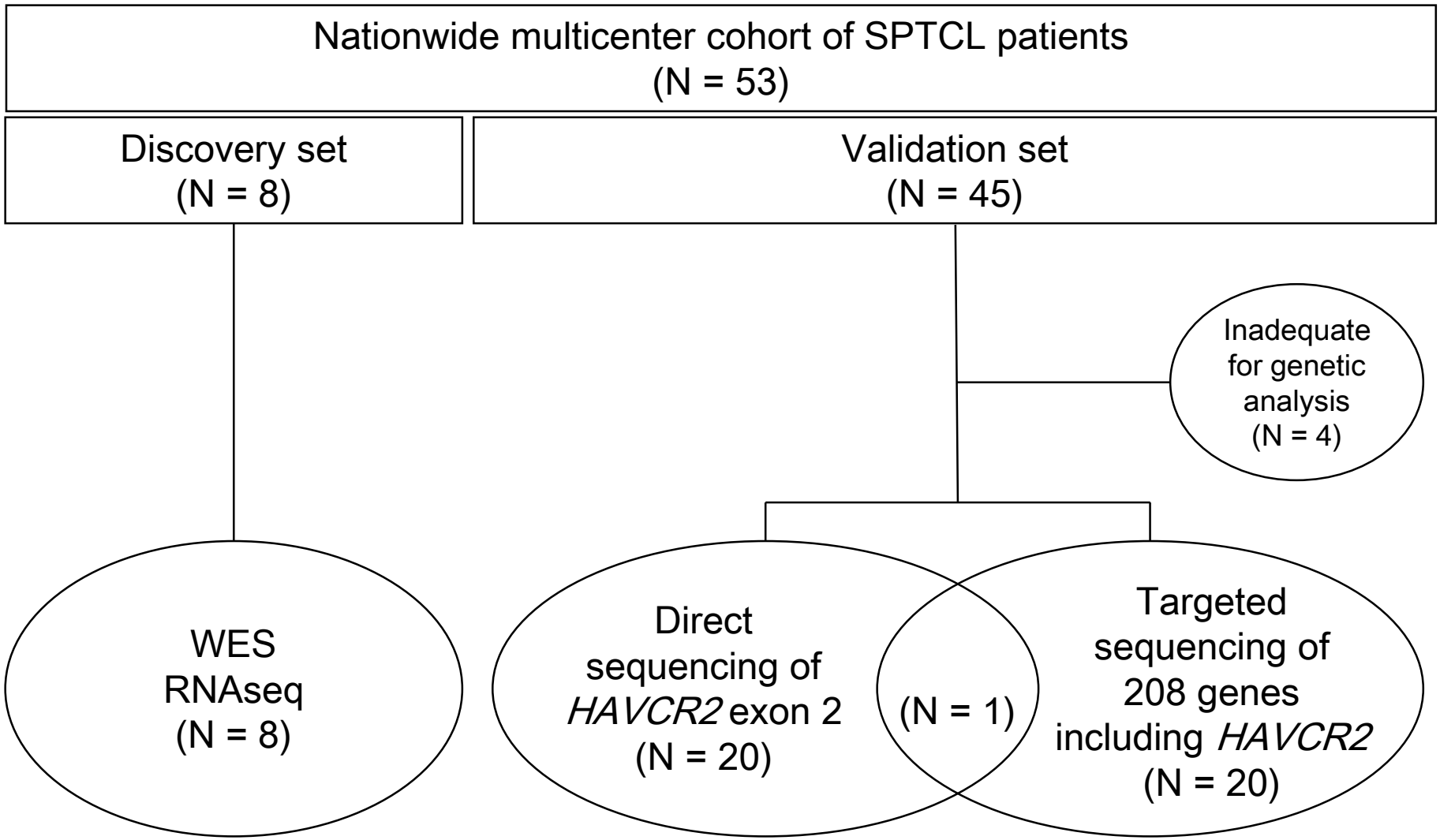
Supplementary Table S8. Variants detected by whole exome sequencing in the discovery set

Supplementary Table S9. Variants detected by targeted gene sequencing in the validation set

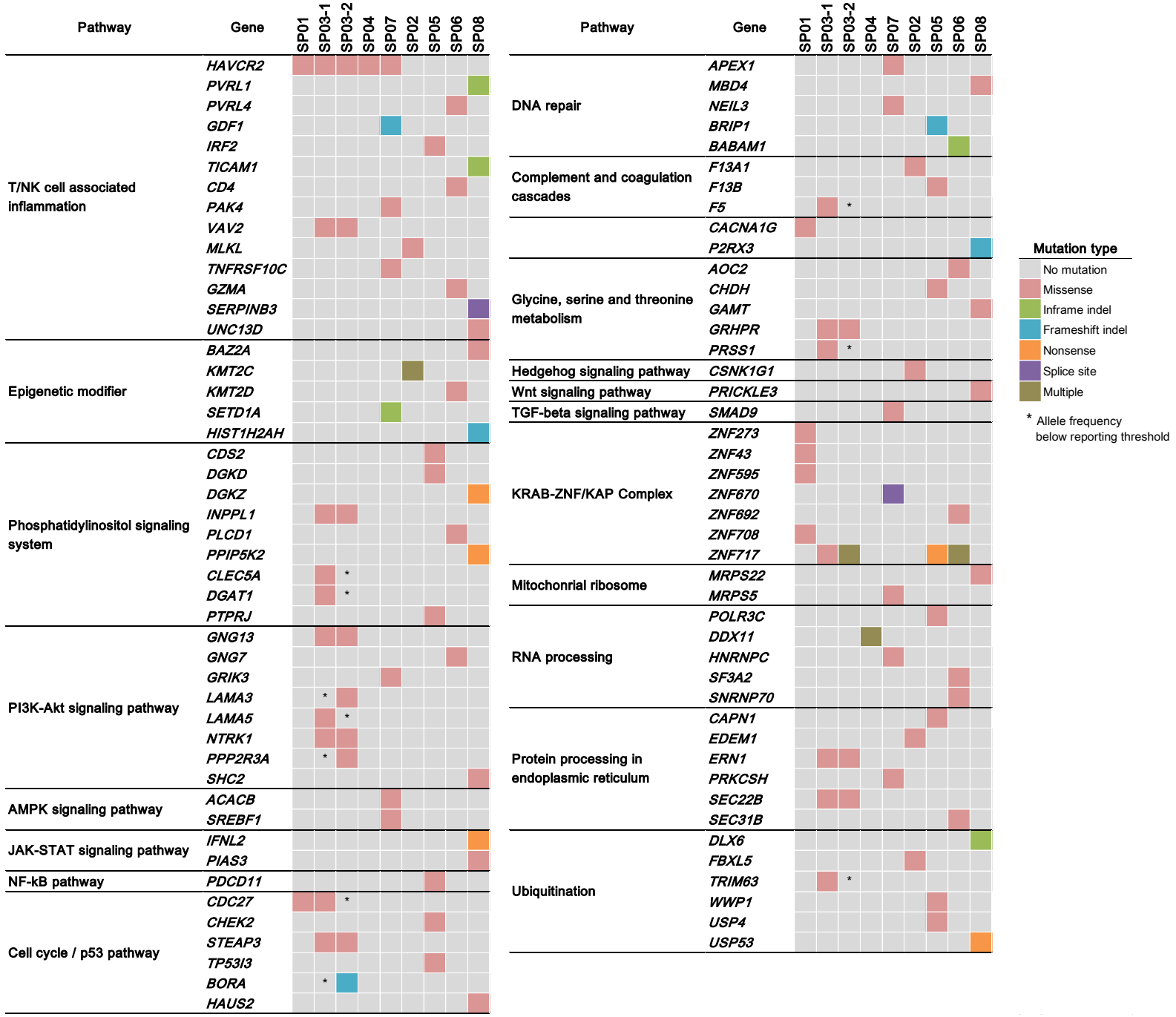
Supplementary Table S10. Result of gene set enrichment analysis

Supplementary Table S11. Detailed descriptive statistics of pSTAT3, CCR4 and Foxp3 expression

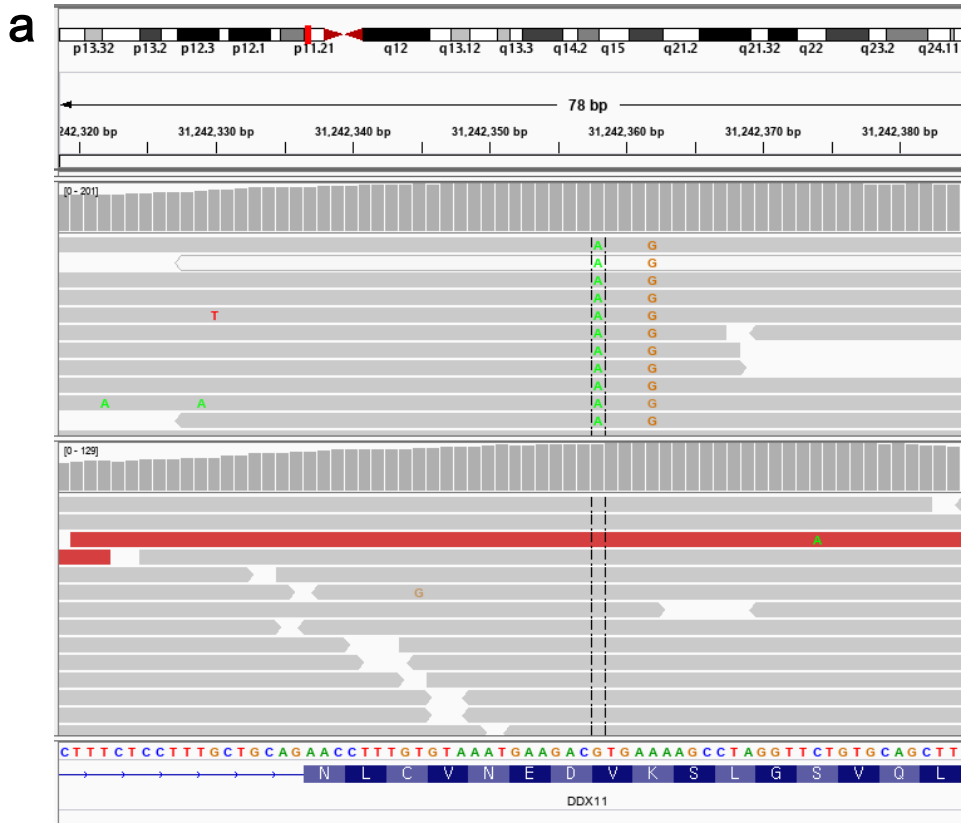
Supplementary Figure S1



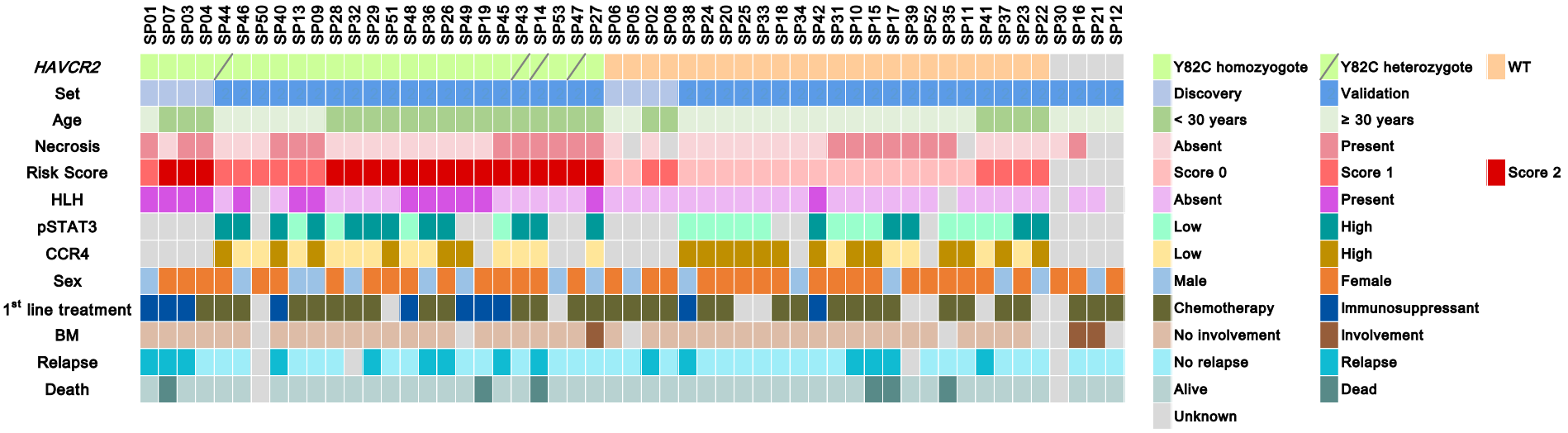
Supplementary Figure S2



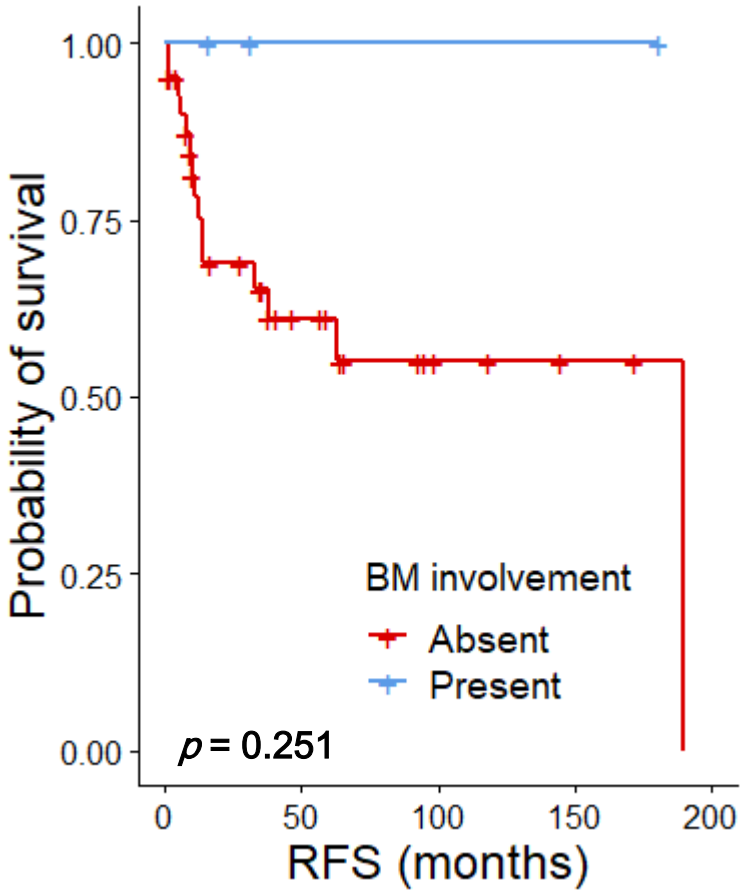
Supplementary Figure S3

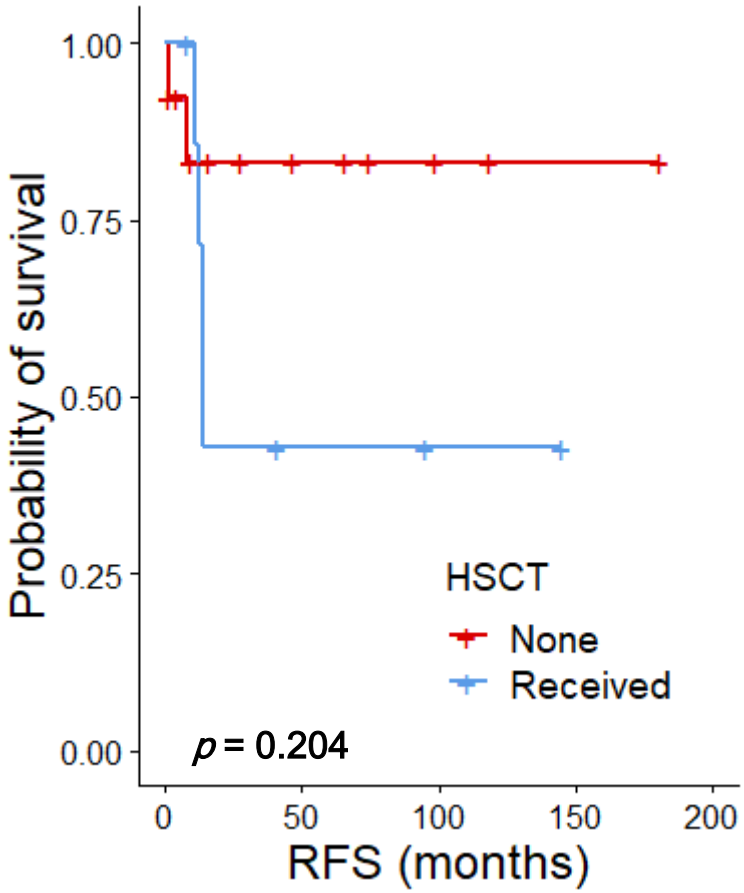


Supplementary Figure S4



Supplementary Figure S5





Supplementary Figure S6