1 **Molecular heterogeneity in peripheral T-cell lymphoma, not otherwise specified**
2 **revealed by comprehensive genetic profiling.**
3

4 Yosaku Watatani[1], Yasuharu Sato [2,3], Hiroaki Miyoshi[4], Kana Sakamoto [5], Kenji Nishida [2],
5 Yuka Gion [3], Yasunobu Nagata [1], Yuichi Shiraishi [6], Kenichi Chiba [6], Hiroko Tanaka [6],
6 Lanying Zhao[1], Yotaro Ochi [1], Yasuhide Takeuchi [1], June Takeda [1], Hiroo Ueno[1], Yasunori
7 Kogure[1,16], Yusuke Shiozawa [1], Nobuyuki Kakiuchi[1], Tetsuichi Yoshizato [1], Masahiro M.
8 Nakagawa[1], Yasuhito Nanya [1], Kenichi Yoshida [1], Hideki Makishima[1], Masashi Sanada [7],
9 Mamiko Sakata-Yanagimoto[8], Shigeru Chiba [8], Ryota Matsuoka [9], Masayuki Noguchi[9],
10 Nobuhiro Hiramoto[10], Takayuki Ishikawa [10], Junichi Kitagawa [11], Nobuhiko Nakamura[11],
11 Hisashi Tsurumi [11], Tatsuhiko Miyazaki [12], Yusuke Kito [13], Satoru Miyano[6], Kazuya
12 Shimoda[14], Kengo Takeuchi [5, 15], Koichi Ohshima [4], Tadashi Yoshino [2], Seishi Ogawa [1],
13 Keisuke Kataoka [1,16]
14

15 [1] Department of Pathology and Tumor Biology, Graduate School of Medicine, Kyoto
16 University, Kyoto, Japan.
17 [2] Department of Pathology, Okayama University Graduate School of Medicine, Dentistry

and Pharmaceutical Sciences, Okayama, Japan.

[3] Division of Pathophysiology, Okayama University Graduate School of Health Sciences, Okayama, Japan.

[4] Department of Pathology, Kurume University School of Medicine, Kurume, Japan.

[5] Pathology Project for Molecular Targets, The Cancer Institute, Japanese Foundation for Cancer Research, Tokyo, Japan.

[6] Laboratory of DNA Information Analysis, Human Genome Center, Institute of Medical Science, The University of Tokyo, Tokyo, Japan.

[7] Department of Advanced Diagnosis, Clinical Research Center, Nagoya Medical Center, Nagoya, Japan.

[8] Department of Hematology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan.

[9] Department of Pathology, Faculty of Medicine, University of Tsukuba, Tsukuba, Japan.

[10] Department of Hematology, Kobe City Medical Center General Hospital, Kobe, Japan.

[11] Department of Hematology, Gifu University Graduate School of Medicine, Gifu, Japan.

[12] Pathology Division, Gifu University Hospital, Gifu, Japan.

1 [13] Department of Pathology and Translational Research, Gifu University Graduate
2 School of Medicine, Gifu, Japan.
3 [14] Department of Gastroenterology and Hematology, Faculty of Medicine, University of
4 Miyazaki, Miyazaki, Japan.
5 [15] Division of Pathology, The Cancer Institute, Japanese Foundation for Cancer
6 Research, Tokyo, Japan
7 [16] Division of Molecular Oncology, National Cancer Center Research Institute, Tokyo,
8 Japan.
9
10 These authors contributed equally: Yasuharu Sato, Hiroaki Miyoshi, and Kana
11 Sakamoto.
12

13 **Running title:**

14 Molecular heterogeneity in peripheral T-cell lymphoma, not otherwise specified.
15

***Contact information:**

Seishi Ogawa, M.D., Ph.D.
Address: Yoshida Konoe-cho, Sakyo-ku, Kyoto, Kyoto, 606-8501, Japan
Phone: +81-75-753-9284 Fax: +81-75-753-9282
E-mail: sogawa-tky@umin.ac.jp

Keisuke Kataoka, M.D., Ph.D.
Address: 5-1-1 Tsukiji, Chuo-ku, Tokyo 104-0045, Japan
Phone: +81-3-3542-2511 Fax: +81-3-5565-0727
e-mail: kkataoka-tky@umin.ac.jp

Competing Interests Statement:

The authors declare no potential conflicts of interest.

Abstract

Peripheral T-cell lymphoma, not otherwise specified (PTCL, NOS) is a diagnosis of exclusion, being the most common entity in mature T-cell neoplasms, and its molecular pathogenesis remains significantly understudied. Here, combining whole-exome and targeted-capture sequencing, gene expression profiling, and immunohistochemical analysis of tumor samples from 133 cases, we have delineated the entire landscape of somatic alterations, and discovered frequently affected driver pathways in PTCL, NOS, with and without a T follicular helper (TFH) cell phenotype. In addition to previously reported mutational targets, we identified a number of novel recurrently altered genes, such as *KMT2C*, *SETD1B*, *YTHDF2*, and *PDCD1*. We integrated these genetic drivers using hierarchical clustering and identified a previously undescribed molecular subtype characterized by *TP53* and/or *CDKN2A* mutations and deletions in non-TFH PTCL, NOS. This subtype exhibited different prognosis and unique genetic features associated with extensive chromosomal instability, which preferentially affected molecules involved in immune escape and transcriptional regulation, such as*HLA-A/B* and *IKZF2*. Taken together, our findings provide novel insights into the molecular pathogenesis of PTCL, NOS by highlighting their genetic heterogeneity. These results

1 should help to devise a novel molecular classification of PTCLs and to exploit a new

2 therapeutic strategy for this group of aggressive malignancies.

1 **Introduction**

2 PTCLs represent a clinically, histologically, and molecularly heterogeneous group of
3 non-Hodgkin lymphomas (NHLs) derived from mature post-thymic T cells.[1, 2] Among
4 them, the most common entity is PTCL, NOS, accounting for approximately 30% of all
5 PTCLs.[3] Patients with PTCL, NOS generally demonstrate aggressive clinical course
6 and are often refractory to standard therapy. By definition, PTCL, NOS includes cases
7 that do not meet the criteria for any specific PTCL subtypes and has been considered a
8 "wastebasket" category.

9 It has been recognized that a subset of PTCLs classified as PTCL, NOS has a
10 T follicular helper (TFH) cell phenotype (i.e. positive for CD4, PD-1, CD10, CXCL13,
11 BCL6, and so on) and some pathological features of angioimmunoblastic T-cell
12 lymphoma (AITL).[4-6] In addition, recent genetic studies revealed that these cases
13 share some of the recurrent genetic alterations found in AITL, such as mutations
14 affecting *TET2*, *DNMT3A*, and *RHOA*.[7-11] Among these, the *RHOA* G17V mutation is
15 highly specific to both PTCL subtypes and, when expressed in mouse T cells, induces
16 TFH cell specification and, together with *TET2* loss, results in the development of
17 AITL-like tumors.[12] On the basis of these findings, the revised World Health

Organization (WHO) classification of haematological malignancies recommended that this subset of PTCL, NOS should be classified as PTCL with a TFH cell phenotype as a provisional entity (referred to as "TFH PTCL, NOS").[5] However, the molecular pathogenesis of the remaining cases in the PTCL, NOS category is still poorly understood. The currently available genetic data from several small series reported different recurrent mutations and copy number alterations (CNAs),[13-16] which preclude a solid conclusion as to the genomic landscape of the tumor. Systematic characterization of genetic alterations should significantly contribute to refining the molecular classification, improving prognostication, and identifying candidate therapeutic targets in this entity, as demonstrated in other lymphomas.[6,17]

Here, we conducted a comprehensive genetic analysis to determine the spectrum of mutations, CNAs, and structural variations (SVs) in PTCL, NOS with and without TFH cell phenotype. In particular, our efforts focused on genetically dissecting the molecular pathogenesis and identifying a new molecular subgroup of PTCL, NOS, showing unique genetic and clinicopathological features.

1 **Materials and methods**

2 **Patient samples**

3 A total of 142 patients diagnosed with PTCL, NOS at six institutions were enrolled in this
4 study according to the protocols approved by the Institutional Review Boards. This
5 study was approved by the institutional ethics committees of the Graduate School of
6 Medicine, Kyoto University and other participating institutes. All cases were reviewed
7 and a consensus diagnosis was made by expert hematopathologists according to the
8 criteria of the 2008 WHO classification,[18] of which 94 cases examined for tumor content.
9 HTLV-1 infection was examined by anti-HTLV-1 antibody detection and/or Southern
10 blotting for HTLV-1 proviral DNA. HTLV-1 -positive cases were considered as adult
11 T-cell leukemia/lymphoma (ATL) and excluded from this study before analysis. Based
12 on the recent revision of the WHO classification,[5] TFH markers, including PD-1, CD10,
13 CXCL13, and BCL6, were evaluated, and PTCL, NOS cases positive for at least two
14 TFH markers were diagnosed as TFH PTCL, NOS. Because the minimum criteria for
15 assignment of TFH phenotype is not well established, we considered PTCL, NOS cases
16 expressing only one TFH marker as unclassifiable PTCL, NOS. Age, sex, and other
17 clinical characteristics are summarized in **Supplemental Table S1**. Genomic DNA

was extracted from fresh frozen tumor tissues or buccal swabs (as normal control) using the QIAamp DNA Mini kit (QIAGEN, Hilden, Germany) or commercially prepared by SRL Inc. (Tokyo, Japan). RNA was extracted from fresh frozen or formalin-fixed paraffin-embedded (FFPE) tumor tissues with RNeasy Mini Kit (QIAGEN).

Whole-exome sequencing (WES)

SureSelect Human All Exon v5 kits (Agilent Technologies, Santa Clara, CA, USA) were used for exome capture according to the manufacturer's instructions. Sequencing data were generated using the Illumina HiSeq 2500 platform with a standard 125-bp paired-end read protocol, as previously described.[19] WES data for 3 PTCL, NOS, 3 AITL, and 81 ATL cases were described in our previous reports.[10, 19] Publicly available WES data for PTCL, NOS [accession number phs000689.v1.p1 [9]], AITL [phs000689.v1.p1 [9] and SRP029591 [20]], anaplastic large cell lymphoma [ALCL, SRP044708 [21]], and extranodal NK/T-cell lymphoma [ENKTL, SRP057085 [22]] were obtained from the National Center for Biotechnology Information Sequence Read Archive. Sequence alignment and mutation calling were performed using the Genomon pipeline (https://github.com/Genomon-Project), as previously described,[19, 23]

with minor modifications. Putative somatic mutations with (i) Fisher's exact P value < 0.01; (ii) > 4 variant reads in tumor; (iii) allele frequency in tumor > 0.025; and (iv) sequencing depth in tumor ≥ 30 were adopted and filtered by excluding (i) synonymous single-nucleotide variants (SNVs); (ii) variants only present in unidirectional reads; and (iii) variants occurring in repetitive genomic regions. These candidate mutations were further filtered by removing known variants listed in the 1000 Genomes Project (October 2014 release), NCBI dbSNP build 131, National Heart, Lung, and Blood Institute (NHLBI) Exome Sequencing Project (ESP) 6500, the Human Genome Variation Database (version 2.0), the Exome Aggregation Consortium (ExAC), or our in-house single nucleotide polymorphism (SNP) database, unless they were listed in the COSMIC database (v70). Moreover, recurrently altered genes, including *RHOA*, *TET2*, *IDH2*, *DNMT3A*, and *TP53*, were manually reviewed for additional mutations. Finally, mapping errors were removed by visual inspection with Integrative Genomics Viewer (IGV).

Targeted-capture sequencing

Targeted capture sequencing was performed using a custom SureSelect library (Agilent

1 Technologies), for which 140 genes (**Supplemental Table S2**) reported to be
2 recurrently mutated in PTCL, NOS, AITL, ATL, ALCL, ENKTL, cutaneous T-cell Lymphoma,
3 and major subtypes of B-cell lymphomas (**Supplemental Table S3;** refs. [5, 6, 9, 10, 17, 19-22,
4 24-31]). Additional probes for 1,999 SNPs were included to calculate genomic copy
5 numbers.[32] Mutation calling was performed with Empirical Bayesian Mutation Calling
6 (EBCall).[33] Candidate mutations were filtered in the same manner as for WES
7 analysis, except for the inclusion of (i) P value < 10^{-4}; (ii) > 4 variant reads in tumor; and
8 (iii) allele frequency in tumor > 0.025, and (iv) the exclusion of missense SNVs with
9 allele frequency of 0.35-0.65 in copy-neutral regions, unless they were listed in the
10 COSMIC database (v70).
11
12 **RNA sequencing (RNA-seq)**
13 Libraries for RNA-seq were prepared from total RNA extracted from fresh frozen tumor
14 tissues using the NEBNext Ultra RNA Library Prep kit for Illumina (New England
15 BioLabs, Beverly, MA, USA), and subjected to sequencing using the HiSeq 2500
16 instrument with a standard 125-bp paired-end read protocol. The sequencing reads
17 were aligned to the human reference genome (hg19) using STAR (v2.5.3).[34] The

mapped reads per gene were counted with featureCounts (v1.5.3) from the R-package 'Rsubread', and normalized to counts per million (CPM) (R-package 'edgeR').[35] To identify significantly enriched pathways in each group, Gene Set Enrichment Analysis (GSEA: v2.2.4) with the Molecular Signatures Database-curated gene sets (hallmark and C2: v6.1) was performed for genes expressed at > 1 CPM in two or more samples.

nCounter gene expression assay

Details of the nCounter assay (NanoString Technologies, Seattle, WA) have been reported previously.[36] Briefly, the *GATA3*, *TBX21*, and 20 housekeeping gene probes (NanoString Technologies) were hybridized to 300 ng of total RNA for 16 hours at 65 °C, and applied to the nCounter Preparation Station for automated removal of excess probe and immobilization of probe-transcript complexes on a streptavidin-coated cartridge. Data were analyzed by using the nSolver 4.0 software (NanoString). To test the validity of nCounter analysis, a linear regression analysis was performed between normalized counts and CPM for GATA3 and TBX21 expressions, respectively.

1 **SV and CNA analysis**

2 SVs and CNAs were detected using the Genomon pipeline and the CNACS algorithm,

3 respectively, as previously described.[19, 32, 37] Putative SVs were manually curated and

4 further filtered by removing those (i) with Fisher's exact *P* value > 0.01; (ii) with ≥ 6

5 supporting reads in tumor; (iii) with allele frequency in tumor < 0.02; or (iv) present in

6 any of control samples. SV breakpoints were visually inspected using IGV.

7 Candidate focal CNAs (shorter than half a chromosome arm, except for 17p deletions

8 involving *TP53*) were assessed in genomic regions where sequencing coverage was

9 sufficient in unmatched control samples, and then manually reviewed and further filtered

10 by removing those with < 3 probes. Frequency of focal CNAs were calculated for 49

11 genes (i) with recurrent mutations or SVs (found in ≥ 3 cases) in our cohort (47 genes)

12 and/or (ii) with focal homozygous deletions or high-level (copy number ≥ 4)

13 amplifications in at least two samples (2 genes: *CDKN2A* and *ARID2*). To confirm

14 CNAs detected by the CNACS algorithm, we conducted SNP array karyotyping for 24

15 samples using the Affymetrix GeneChip Human Mapping 250K NspI array (Affymetrix,

16 Santa Clara, CA, USA), as previously described.[19, 23] Microarray data were analyzed

17 to estimate total and allele-specific copy numbers using CNAG/AsCNAR algorithms.

1 Significantly recurrent arm-level CNAs were identified using a binomial distribution test,

2 as previously described.[38]

3

4 **Mutation analysis**

5 Pairwise correlations between alterations (present in ≥ 10 cases) were assessed by

6 Fisher's exact test with Benjamini-Hochberg correction ($q < 0.1$). Mutational signature

7 was determined by pmsignature (version 0.2.1), as previously described.[39] The *RHOA*

8 G17V mutations were separately analyzed, because they behaved differently from other

9 *RHOA* mutations.

10

11 **Hierarchical clustering**

12 Unsupervised hierarchical clustering of recurrent somatic alterations, including 49 genes

13 (affected by mutations and/or focal CNAs) and 14 arm-level CNAs, was performed with

14 Spearman's rank correlation and Ward's linkage algorithm (R-package ' heatmap.2').

15

1 **Detection of HTLV-1 genome**

2 For the detection of HTLV-1 sequence, after sequencing reads were mapped to the
3 HTLV-1 genome (AB513134), the number of the HTLV-1-aligned reads were
4 enumerated and divided by the number of total reads mapped to the human reference
5 genome (GRCh37). Then, the obtained ratio was evaluated for the cut-off value of
6 0.01%, which was determined so that all confirmed ATL cases were included (data not
7 shown).
8
9 **Survival analysis**

10 Survival data were available for 46 patients with PTCL, NOS. Observations were
11 censored at the last follow-up. The median follow-up was 22.6 months in surviving
12 patients, and 24 patients were alive at the last follow-up. The Kaplan-Meier method
13 was used to estimate overall survival, and the log-rank test was used to assess
14 differences in overall survival between patient groups (R-package 'survival').
15
16 **Immunohistochemical analysis (IHC)**

17 IHC for PD-1, CD10, CXCL13, BCL6, GATA3, and TBX21 was performed on FFPE

1 tissue sections using antibodies directed against PD-1 (NAT105, Abcam, Cambridge,
2 UK), CD10 (56C6, Leica Biosystems, Newcastle, UK), CXCL13 (polyclonal, R&D
3 Systems, Minneapolis, MN, USA), BCL6 (EP529Y, Abcam; PG-B6P, Dako, Glostrup,
4 Denmark; and LN22, Novocastra, Newcastle, UK), GATA3 (L50-823, Nichirei
5 Bioscience, Tokyo, Japan), and TBX21 (4B10, Abcam). The antigen–antibody
6 complexes were visualized with Histofine Simple Stain MAX PO (Nichirei Bioscience),
7 Bond polymer Refine Detection kit (Leica Biosystems), or the REAL EnVision Detection
8 system (Dako).
9
10 **CRISPR-mediated gene targeting**
11 Human IRF2BP2 sgRNA targeted sites were designed manually and checked in silico.
12 The pSpCas9(BB)-2A-GFP (pX458) vector expressing Cas9 (Addgene plasmid 48138)
13 was digested with BbsI and ligated to annealed and phosphorylated sgRNA
14 oligonucleotides. Jurkat cells, obtained from the RIKEN Cell Bank, were transfected
15 with indicated vectors using the Amaxa Nucleofector system (Lonza, Bazel,
16 Switzerland) system according to the manufacturer's instructions.
17 CRISPR/Cas9-mediated targeting was confirmed by PCR-based deep sequencing and

1 expression analysis by real-time quantitative PCR, as previously described.[37] The

2 sgRNA sequences and PCR primers are listed in **Supplemental Table S4**.

3

4 **Real-time quantitative PCR**

5 cDNA synthesis from total RNA was performed with ReverTra Ace qPCR RT Kit

6 (TOYOBO, Tokyo, Japan), and subjected to quantitative reverse transcription PCR with

7 SYBR Premix Ex TaqII (Tli RNaseH Plus) (TaKaRa, Shiga, Japan) and LightCycler 480

8 System (Roche, Basel, Switzerland) according to the manufacturer's instructions. All

9 assays were performed in three technical replicates for each biological replicate and

10 relative expression was normalized for 18S rRNA.

11

12 **Immunoblot analysis**

13 Cells were lysed, subjected to SDS–PAGE, and transferred to a PVDF membrane

14 (Millipore, Bedford, MA, USA). The blot was incubated with the antibodies listed in

15 **Supplemental Table S5**, and visualized by Immobilon Western Chemiluminescent HRP

16 Substrate (Millipore).

17

Luciferase assay

Jurkat cells were collected 48 h after transfection with pX458 and pGL4.30 (luc2P/NFAT-RE/Hygro, Promega, Madison, WI, USA) vectors and assayed for NFAT luciferase activity using the Dual-Luciferase Reporter Assay System (Promega) and Wallac ARVO SX 1420 Multilabel Counter (PerkinElmer, Waltham, MA, USA). Firefly luciferase activity was normalized by Renilla luciferase activity (phRL-TK vector, Promega) in each sample and is presented with a logarithmic scale relative to the activity in mock-transfected cells.

Cell proliferation assay

Five thousand cells transduced with indicated vectors were inoculated into 96-well culture plates, and their growth was monitored using Cell Counting Kit-8 (DOJINDO LABORATORIES, Kumamoto, Japan) according to the manufacturer's protocol.

Statistical analysis

Statistical analyses were performed with R 3.1.3 software (The R Foundation for Statistical Computing). Comparisons between groups were based on the Wilcoxon

1 rank-sum test for continuous data with Bonferroni correction (if necessary) and the
2 Fisher's exact test with Benjamini–Hochberg correction (if necessary) for categorical
3 data. For functional assays, normality of data distribution and homogeneity of variance
4 were assessed by the Shapiro–Wilk's test and F-test, respectively. Student's
5 two-tailed t-test was used to compare two groups and a Welch's correction was applied
6 when comparing groups with unequal variance (F-test $P < 0.05$). In box plots, the
7 center line and lower and upper hinges correspond to the median, and the first and third
8 quartiles (25 and 75 percentiles), respectively. The upper and lower whiskers extend
9 from the upper and lower hinges to the largest or smallest values no further than 1.5×
10 inter-quartile range from the hinges.

1 **Results**

2 **WES of patients with PTCL, NOS.**

3 To delineate the entire picture of genetic alterations in PTCL, NOS, we initially
4 performed WES analysis of tumor and normal samples from 20 PTCL, NOS patients
5 (**Supplemental Figure 1A-B**), including 7 TFH, 3 unclassifiable, and 4 non-TFH PTCL,
6 NOS cases from our cohort. In total, we detected 1,068 somatic mutations (1.5
7 mutations/Mb/sample), including 971 SNVs and 97 insertions and deletions (indels), as
8 well as 42 SVs (**Figure 1A and Supplemental Tables S6-S8**). These mutations
9 mainly consisted of age-related C > T transitions at CpG sites, followed by C > A
10 substitutions at the CpCpT context, whose etiology has been unknown (**Figure 1B**).
11 Approximately a half of patients exhibited a low overall mutation frequency (< 0.5
12 mutations/Mb), while there were four samples showing a moderate to high mutation rate
13 (2-10 mutations/Mb) (**Figure 1A**). Additional targeted sequencing not only validated
14 the somatic mutations detected by WES sequencing, but also captured multiple
15 previously reported mutations with low allele frequencies, such as those involving *TET2*
16 and *RHOA* (colored in red **Figure 1C**),[7-11], suggesting a possibility that some driver
17 mutations overlooked with WES analysis. In addition to TFH-related mutations, the

1 observed alterations included recurrent mutations and deletions of *TP53* (n = 7), a

2 well-known tumor suppressor gene,[40] which were associated with a higher tumor

3 mutation burden (**Figure 1D**). Moreover, a number of mutations frequently observed in

4 other subtypes of lymphomas, such as those in *CDKN2A*, *VAV1*, and *TBL1XR1*,[5, 17, 19, 41]

5 were also detected (**Figure 1A, C, Supplemental Figure S1C and Supplemental**

6 **Table S6**). These results suggest a potential role of lymphoma-associated mutations,

7 particularly those affecting *TP53*, as driver alterations in the molecular pathogenesis of

8 PTCL, NOS.

9

10 **Overview of PTCL, NOS genomes revealed by deep targeted-capture sequencing.**

11 On the basis of these results, we then carried out deep targeted-capture

12 sequencing that covered 140 lymphoma-associated genes (**Supplemental Table S2**) in

13 a cohort of 142 patients with PTCL, NOS (including 11 WES cases), with a mean depth

14 of 627× (range, 399-830×) (**Supplemental Figure S2A-B**). Unexpectedly, 18 cases

15 had a substantial number of sequencing reads mapped to the HTLV-1 proviral genome

16 (**Supplemental Figure S2C**). After excluding these cases, who were considered to

17 have ATL, we analyzed the remaining 124 cases and identified 438 non-silent somatic

mutations (333 SNVs and 105 indels), with a median of 3 per sample (range 0-13) (**Supplemental Table S9**). These included numerous driver mutations which did not appear to be readily identifiable by WES due to low allele frequency (**Supplemental Figure S2D**), suggesting that deep targeted-sequencing would be required to delineate the entire landscape of driver alterations in PTCL, NOS.

When the results from targeted-capture sequencing and WES were combined, a total of 41 genes were found to be recurrently mutated (in ≥ 3 cases), of which 12 (such as *TET2*, *TP53*, *RHOA*, and *DNMT3A*) were affected in more than 5% of a total of 133 cases (**Supplemental Figure S2E**). Copy number analysis based on the sequencing method identified 222 focal CNAs in 41 recurrently altered genes (6 amplified and 35 deleted genes) (**Supplemental Figure S2F and Supplemental Table S10**). Among them, 10 (such as *TP53*, *CDKN2A*, *CD28*, *HLA-B*, and *IKZF2*) were affected in more than 5% of the cases, some of which showed high-level amplifications or homozygous deletions (**Supplemental Figure S2F and S3A-B**). Additionally, 251 arm-level CNAs were detected in 14 significantly altered chromosome arms (7 gains and 7 losses) (**Supplemental Figure S3C**). SNP array karyotyping was also performed for 24 samples, in which 25 out of 27 focal CNAs (93%) and 136 of 141

arm-level CNAs (96%) were confirmed (**Supplemental Figure S3D and Supplemental Table S10**).

We also identified 32 SVs (21 deletions, 5 inversions, 5 tandem duplications, and 1 translocation) in recurrently affected genes, including *IKZF2* and *CD274* (**Supplemental Figure S2F and Supplemental Table S11**). Overall, 107 (80%) and 69 (52%) of 133 PTCL, NOS patients carried at least one driver mutation and CNA/SV, respectively, which belong to a wide spectrum of T-cell-related biological processes (**Figure 2 and 3**). When evaluated together, 116 (87%) patients harbored at least one somatic alteration, and 49 genes, including 25 previously unreported genes (*HLA-A/B*, *KMT2C*, *NOTCH1*, *ARID1A*, and so on), were recurrently affected (in ≥ 3 cases), including 10 genes affected in more than 10% of the cases (**Figure 2**). Among these 133 cases, three or more TFH makers were ev aluated by IHC in 98 cases, of which 37, 25, and 36 were considered to have TFH, unclassifiable, and non-TFH PTCL, NOS, respectively (**Supplemental Table S1**).

***RHOA* G17V and *IDH2* R172 mutations are highly specific for TFH PTCL, NOS.**

In accordance with previous reports,[7-11] *TET2* (44%), *RHOA* (26%), and

DNMT3A (12%) were frequently altered in this cohort (**Figure 2, 3 and Supplemental Figure S4A**). Although *IDH2* mutations had not previously been reported in PTCL, NOS, including that with TFH phenotype,[42, 43] 10 (8%) cases harbored *IDH2* mutations, mostly consisting of R172 substitutions (**Figure 2 and Supplemental Figure S4A**). Immunohistochemical evaluation revealed significant associations of *TET2* and *RHOA* mutations with the expression of TFH markers, such as PD-1, CD10, CXCL13, and BCL6 (**Figure 4A**). Although *DNMT3A* and *IDH2* mutations tended to occur more frequently in TFH PTCL, NOS, many types of alterations, such as *CD28* mutations and amplifications, were present irrespective of TFH marker status, suggesting partially overlapping genetic mechanisms involved in TFH and non-TFH PTCL, NOS (**Figure 4A and Supplemental Figure S4B**). Interestingly, at least one TFH marker was positive in all cases with *RHOA* G17V mutations, which were almost invariably accompanied by *TET2* mutations with higher variant allele frequencies (**Figure 2B, 4A and Supplemental Figure S4C-D**). By contrast, none of five cases with other *RHOA* mutations were TFH PTCL, NOS, suggesting the G17V substitution is pathognomonic for TFH-related PTCL. As expected, unclassifiable PTCL, NOS cases exhibited a genetic feature intermediate between TFH and non-TFH PTCL, NOS, suggesting this entity

consisted of a mixed population of TFH and non-TFH PTCL, NOS cases (**Supplemental Figure S4B, C**).

A distinct molecular subtype characterized by *TP53* and *CDKN2A* alterations in non-TFH PTCL, NOS.

Although rarely reported in the previous literatures,[13, 14, 44, 45] *TP53* mutations and deletions were found in as many as 37 cases (28%) with PTCL, NOS, where the majority (51%) had a biallelic lesion (**Figure 2, 3 and Supplemental Figure S4A, E**). *CDKN2A*, another tumor suppressor, was focally deleted in 17 cases (13%), of which 11 had homozygous deletions (**Figure 2, 3 and Supplemental Figure S3A**). Remarkably, *TP53* and *CDKN2A* represented two leading targets of genetic alterations in non-TFH PTCL, NOS, and their alterations negatively correlated with TFH marker expression (**Figure 4A**). Prompted by the inverse correlation between *TP53* and *CDKN2A* and TFH-related alterations, we investigated co-occurrence and exclusion between somatic alterations (**Figure 4B**). TFH-related abnormalities, including *TET2* alterations and *RHOA* G17V and *IDH2* mutations, showed a strong tendency to co-occur, whereas the *RHOA* G17V mutation were mutually exclusive with *TP53* and *CDKN2A* alterations

(**Figure 4B**). Moreover, the latter two alterations significantly co-occurred with somatic aberrations involving the *HLA-A*, *HLA-B*, *CD58*, and *IKZF2* genes (**Figure 4B**). Taken together, these observations clearly depicted two molecular subtypes in PTCL, NOS: subtypes characterized by TFH-related alterations and *TP53* and *CDKN2A* alterations, respectively (**Figure 4C**).

***TP53/CDKN2A*-altered PTCL, NOS shows marked chromosome instability.**

In consistent with these findings, hierarchical clustering of recurrent somatic alterations revealed three molecular subtypes with discrete genetic features: those with TFH-related alterations (*TET2*, *RHOA* G17V, and *IDH2*) (group 1), those with *TP53/CDKN2A* alterations (group 2) and, those lacking any of the above alterations (group 3) (**Figure 2**). While group 1 shows similar immunophenotype and genetic alterations to TFH PTCL, NOS,[5] group 2 is supposed to represent a novel molecular subtype in PTCL, NOS (**Figure 5A**). As revealed by genome-wide copy number profiling, almost all group 2 cases exhibited extensive chromosomal abnormalities, which were rarely seen in other groups (**Figure 5B**), pointing to a discrete genetic feature of group 2 tumors. This difference was quantitatively substantiated by a higher

1 number of abnormal genomic segments in *TP53*- or *CDKN2A*-altered cases than those
2 harboring TFH-related or other alterations (**Figure 5C and Supplemental Figure**
3 **S5A-B**). Although group 3 showed a lower number of genetic alterations, *ATM*
4 mutations and deletions were detected in a subset of group 3, some of which had
5 extensive CNAs, similar to group 2 cases (**Figure 2B, 5B and Supplemental Figure**
6 **S5A-B**). Given that ATM regulates the ARF-TP53 tumor suppressor pathway in
7 response to DNA damage[46] (**Figure 3**), *ATM*-altered tumors may exploit a shared
8 oncogenic mechanism with group 2, which is characterized by tumor suppressor
9 inactivation.

10 GSEA analysis with RNA-seq data from 16 fresh frozen tumor tissues using
11 curated gene sets showed that AITL-related genes were the second most enriched
12 signature in group 1, confirming the validity and reliability of expression analysis (**Figure**
13 **5D and Supplemental Tables S12**). Among hallmark gene sets, genes associated
14 with stromal response and inflammation were enriched in group 1, whereas cell
15 cycle-related genes were overrepresented in group 2 (**Figure 5D and Supplemental**
16 **Tables S12**), suggesting that differences of genetic features among molecular subtypes
17 are reflected in gene expression profiles. Although it has been reported that PTCL,

1 NOS can be classified into two subgroups by *GATA3* and *TBX21* expressions,[47] these
2 expressions measured by RNA-seq, nCounter analysis, or IHC were similar among
3 subtypes, (**Figure 5E and Supplemental Figure S5C-F**).
4 Histologically, group 2 showed a higher tumor content than group 1 (**Figure 6A**),
5 which is consistent with gene expression profiling data showing strong immune or stromal
6 cell-related signatures in group 1. With regard to clinical outcome, group 2 showed the
7 worst survival, followed by group 1, whereas group 3 had an excellent outcome (**Figure**
8 **6B**), suggesting that *TP53/CDKN2A* alterations and associated chromosomal instability
9 confer an adverse prognostic impact. When compared with other PTCL subtypes,
10 group 1, corresponding to TFH PTCL, NOS, showed a similar pattern of mutations and
11 CNAs to AITL (**Figure 6C, Supplemental Figure S6 and Supplemental Tables**
12 **S13-S21**), consistent with previous reports.[9-11] By contrast, group 2 showed a unique
13 profile of somatic alterations, although itshared a number of aberrations with other
14 PTCL subtypes (**Figure 6C**). These findings indicate that *TP53/CDKN2A*-altered
15 cases have a molecular pathogenesis distinct from other PTCL subtypes, which may
16 underlie their different clinical behavior.
17

Frequent genetic alterations associated with immune evasion in *TP53/CDKN2A*-altered PTCL, NOS.

Recurrent alterations in different PTCL, NOS subtypes affected a number of discrete functional pathways. Among these, uniquely overrepresented in group 2 was the pathways involved in immune surveillance (**Figure 6D and Supplemental Figure S7**), which include the components of the class I major histocompatibility complex (MHC) (*HLA-A* and *HLA-B*), the transactivator of MHC class II (*CIITA*), immune checkpoints (*CD274*) and molecules engaged in cell adhesion (*CD58*) and death signaling (*FAS*) (**Figure 2 and 3**). In group 2, most of these genes were affected by loss-of-function alterations, particularly by focal deletions (**Figure 7A and Supplemental Figure S8A-B**), suggesting a possible link to genomic instability characteristic of this subgroup. Intriguingly, we identified recurrent loss-of-function mutations involving the *PDCD1* gene (3% in the entire cohort), the gene encoding an inhibitory receptor, PD-1 (**Figure 7A**). In addition to focal deletions found in ATL and other T-cell lymphomas, which were recently reported to induce T-cell malignancies in mice,[48] frameshift and nonsense mutations of *PDCD1* were observed in PTCL, NOS, suggesting that loss of PD-1 function take places through multiple mechanisms.

Enrichment of somatic lesions in transcriptional and post-transcriptional regulators in *TP53/CDKN2A*-altered PTCL, NOS.

Another significant finding was the enrichment of somatic alterations in transcriptional and post-transcriptional regulators in group 2 (**Figure 6D**). These regulators included transcription factors (*IKZF2*, *PRDM1*, and *ETV6*), transcriptional co-repressors (*TBL1XR1* and *IRF2BP2*), and RNA-binding proteins (*DDX3X* and *YTHDF2*) (**Figure 2 and 3**). IKZF2, also known as HELIOS, is one of the major regulators of T-cell development and affected in 29% of group 2 tumors exclusively through SV/CNAs, such as intragenic deletions, duplications, and inversions, most likely leading to dominant-negative spliced variants [49] (**Supplemental Figure S7 and S8C**).

IRF2BP2, which encodes an IRF2-dependent transcriptional co-repressor,[50, 51] was another common genetic target in group 2 tumors (**Supplemental Figure S7**) and also affected in other lymphoma types,[19, 52] in which frequent non-sense or frameshift mutations are thought to lead to loss of function of IRF2BP2 (**Figure 7A**). To assess the functional consequence of IRF2BP2 mutations on T-cell lymphomagenesis, we evaluated the effect of *IRF2BP2* disruption on cellular growth and the transcriptional

activity of NFAT, a major downstream target of T-cell receptor (TCR) signaling, using CRISPR/Cas9-mediated gene editing (**Figure 7B-D**). Although *IRF2BP2* disruption did not affect cell proliferation, it caused an enhanced transcription from an NFAT response element in a human T-cell line (Jurkat), regardless of co-transfection with NFAT1 or CD3/CD28 stimulation, suggesting that loss-of-function alterations of IRF2BP2 lead to TCR signaling activation (**Figure 7E and Supplemental Figure S8D**). By contrast, recurrent deteriorating deletions and mutations in *YTHDF2* were detected in both group 1 and 2 (8 % of the entire cohort)(**Figure 7A and Supplemental Figure S7**). This gene encodes a reader protein that recognizes N^6-methyladenosine, the most abundant internal modification in mammalian mRNA, and reduces the stability of target transcripts, [53] which may suggest the functional importance of deregulated mRNA stability in the pathogenesis of T-cell lymphoma.

Other commonly affected pathways and molecules in *TP53/CDKN2A*-altered PTCL, NOS.

Signal transduction molecules were also common mutational targets in group 2 tumors, including *NOTCH1* and *SOCS1* (**Figure 6D and Supplemental Figure S7**).

Although activating mutations in genes related to TCR signaling are reported in TFH cell-derived lymphomas,[11] in our cohort, more than two thirds of both group 1 and 2 cases harbored somatic changes in the components of TCR–NF-κB signaling and their downstream pathways (**Figure 6D**). However, the spectrum of target genes substantially differed between group1 and 2 tumors. In group 1, *RHOA* mutations represented by far the most predominant alterations. By contrast, the alterations in group 2 involved a broader spectrum ofgenes than those in group 1, such as *CD28*, *PLCG1*, *CARD11*, *TNFAIP3*, and *PTPRC*, which were frequently affected by focal CNAs, including high-level amplifications or homozygous deletions, rather than missense mutations (**Supplemental Figure S3B and S7**).

In addition to TFH-related mutations, such as those affecting *TET2*, *IDH2*, and *DNMT3A*, recurrent mutations and CNAs/SVs were also present in a variety of epigenetic regulators, including histone modifiers (*KMT2C* (*MLL3*), *KMT2D* (*MLL2*), *SETD1B*, *SETD2*, and *CREBBP*) and SWI/SNF-mediated chromatin remodelers (*ARID1A* and *ARID2*), in our entire cohort (**Figure 2 and 3**). Among these, two histone 3 lysine 4 methyltransferases, KMT2C and SETD1B, were frequently inactivated by loss-of-function mutations or focal deletions in group 2 (**Supplemental Figure S7 and**

1 **S8E**). The remaining group of molecules affected in PTCL, NOS were G protein–

2 coupled receptors involved in T-cell trafficking, such as *CCR4* and *CCR7*, which are

3 also commonly mutated in other T-cell neoplasms (**Figure 2 and 3**).

Discussion

Through extensive genetic analyses using high-throughput sequencing, we have delineated a comprehensive registry of genetic alterations in PTCL, NOS. It includes not only known mutational targets in PTCL, NOS and other lymphoma subtypes, but also novel recurrently altered genes previously unreported in this tumor type, such as *KMT2C*, *SETD1B*, *YTHDF2*, and *PDCD1*. As expected from highly variable clinical presentation and prognosis as well as pathological findings, PTCL, NOS is shown to be a heterogeneous entity in terms of genetic profile.[3] However, it should be underscored that PTCL, NOS does not represent a mere waste basket category, but comprises several discrete subtypes of mature T-cell neoplasms on the basis of unique genetic profiles.

Group 1 tumors, characterized by TFH-related mutations, such as *TET2*, *RHOA* G17V, and *IDH2* mutations, correspond to a provisional entity of TFH PTCL, NOS, according to the revised WHO classification.[5] These tumors also exhibit a variety of somatic alterations at low frequencies, such as *VAV1*, *CD28*, and *YTHDF2*, most of which are shared by other PTCL, NOS subtypes, suggesting overlapping mechanisms of lymphomagenesis. Group 2 tumors are a previously unrecognized

molecular subtype, which harbors frequent *TP53* and/or *CDKN2A* alterations. This subtype shows the unique genetic features characterized by an increased burden of CNAs, which preferentially targeted molecules involved in immune surveillance and transcriptional regulation, including *HLA-A/B* and *IKZF2*. The high prevalence of *TP53/CDKN2A* alterations demonstrated the biological relevance of tumor suppressor inactivation and resultant genomic instability during T-cell lymphomagenesis, which is supported by the fact that T-cell lymphoma is one of the most common malignancies observed in p53-deficeint mice.[40] Except for *ATM* alterations in their subset, group 3 tumors lack a subtype-defining alteration, suggesting the necessity for further molecular investigation in this subtype.

Many efforts have been undertaken to further molecularly characterize and subdivide the heterogeneous group of tumors classified as this category. Microarray analysis of gene expression identified a biologically distinct entity showing a proliferation signature associated with a shorter survival.[54] More recently, large-scale gene expression profiling enabled the characterization of two different molecular subgroups related to high expression of either TBX21 or GATA3.[47] However, the molecular categorization of PTCL, NOS still remains controversial due to the

1 inadequate understanding of genetic landscape of the tumor. Therefore, the
2 identification of the *TP53/CDKN2A*-altered molecular subtype with different genetic
3 features and prognosis can offer a clue to understand the genetic heterogeneity of
4 PTCL, NOS and provide novel insights into its molecular classification and patient
5 stratification, hopefully leading to the improvement of diagnostic and therapeutic
6 strategy for this deadly disease.

Acknowledgments

This work was supported by Grant-in-Aid from the Japan Agency for Medical Research and Development [Medical Research and Development Programs Focused on Technology Transfer (15im0210102h0001) and Project for Cancer Research and Therapeutic Evolution (16cm0106501h0001)] and the Naito Foundation. We thank Miki Sagou and Tsuyoshi Shirahari for technical assistance. The supercomputing resources were provided by the Human Genome Center, the Institute of Medical Science, The University of Tokyo.

Author's contributions:

W.Y. and K.K. designed the study, analyzed the data, and wrote the manuscript; Y.Shiraishi, K.C., H.Tanaka, and S.Miyano developed sequence data processing pipelines; Y.Nagata, Y.O., Y.T., J.T., H.U., Y.Kogure, Y.Shiozawa, N.K., T.Yoshizato, M.Nakagawa, Y.Nanya, K.Y., H.Makishima, M.S. and S.O. performed sequencing data analysis; Y.Sato, K.N., Y.G., H.Miyoshi., K.S., L.Z., M.S-Y., M.Noguchi, R.M., S.C., N.H., T.I., N.N., J.K., T.M., Y.kito, H.Tsurumi, K.S., K.T., K.O., T.Yoshino managed patients and prepared samples; and all

1 authors reviewed the manuscript during its preparation.

2 **Supplementary information**

3 Supplementary information is available at Leukemia's website.

4

References

1. Vose J, Armitage J, Weisenburger D, International TCLP. International peripheral T-cell and natural killer/T-cell lymphoma study: pathology findings and clinical outcomes. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2008 Sep 1; **26**(25): 4124-4130.

2. Foss FM, Zinzani PL, Vose JM, Gascoyne RD, Rosen ST, Tobinai K. Peripheral T-cell lymphoma. *Blood* 2011 Jun 23; **117**(25): 6756-6767.

3. Broccoli A, Zinzani PL. Peripheral T-cell lymphoma, not otherwise specified. *Blood* 2017 Mar 2; **129**(9): 1103-1112.

4. de Leval L, Rickman DS, Thielen C, Reynies A, Huang YL, Delsol G, *et al.* The gene expression profile of nodal peripheral T-cell lymphoma demonstrates a molecular link between angioimmunoblastic T-cell lymphoma (AITL) and follicular helper T (TFH) cells. *Blood* 2007 Jun 1; **109**(11): 4952-4963.

5. Swerdlow SH, Campo E, Harris NL, Jaffe ES, Pileri SA, Stein H, *et al. WHO classification of tumours of haematopoietic and lymphoid tissues*. France: IARC Press, 2017.

6. Moffitt AB, Dave SS. Clinical Applications of the Genomic Landscape of Aggressive Non-Hodgkin Lymphoma. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2017 Mar 20; **35**(9): 955-962.

7. Couronne L, Bastard C, Bernard OA. TET2 and DNMT3A mutations in human T-cell lymphoma. *The New England journal of medicine* 2012 Jan 5; **366**(1): 95-96.

8. Lemonnier F, Couronne L, Parrens M, Jais JP, Travert M, Lamant L, *et al.* Recurrent TET2 mutations in peripheral T-cell lymphomas correlate with TFH-like features and adverse clinical parameters. *Blood* 2012 Aug 16; **120**(7): 1466-1469.

9. Palomero T, Couronne L, Khiabanian H, Kim MY, Ambesi-Impiombato A, Perez-Garcia A, *et al.* Recurrent mutations in epigenetic regulators, RHOA and FYN kinase in peripheral T cell lymphomas. *Nature genetics* 2014 Feb; **46**(2): 166-170.

10. Sakata-Yanagimoto M, Enami T, Yoshida K, Shiraishi Y, Ishii R, Miyake Y, *et al.* Somatic RHOA mutation in angioimmunoblastic T cell lymphoma. *Nature genetics* 2014 Feb; **46**(2): 171-175.

11. Vallois D, Dobay MP, Morin RD, Lemonnier F, Missiaglia E, Juilland M, *et al.* Activating mutations in genes related to TCR signaling in angioimmunoblastic and other follicular helper T-cell-derived lymphomas. *Blood* 2016 Sep 15; **128**(11): 1490-1502.

12. Cortes JR, Ambesi-Impiombato A, Couronne L, Quinn SA, Kim CS, da Silva Almeida AC, *et al.* RHOA G17V Induces T Follicular Helper Cell Specification and Promotes Lymphomagenesis. *Cancer cell* 2018 Feb 12; **33**(2): 259-273 e257.

13. Schatz JH, Horwitz SM, Teruya-Feldstein J, Lunning MA, Viale A, Huberman K, *et al.* Targeted mutational profiling of peripheral T-cell lymphoma not otherwise specified highlights new mechanisms in a heterogeneous pathogenesis. *Leukemia* 2015 Jan; **29**(1): 237-241.

14. Simpson HM, Khan RZ, Song C, Sharma D, Sadashivaiah K, Furusawa A, *et al.* Concurrent Mutations in ATM and Genes Associated with Common gamma Chain Signaling in Peripheral T Cell Lymphoma. *PLoS One* 2015; **10**(11): e0141906.

15. Fujiwara SI, Yamashita Y, Nakamura N, Choi YL, Ueno T, Watanabe H, *et al.* High-resolution analysis of chromosome copy number alterations in angioimmunoblastic T-cell lymphoma and peripheral T-cell lymphoma, unspecified, with single nucleotide polymorphism-typing microarrays. *Leukemia* 2008 Oct; **22**(10): 1891-1898.

16. Hartmann S, Gesk S, Scholtysik R, Kreuz M, Bug S, Vater I, *et al.* High resolution SNP array genomic profiling of peripheral T cell lymphomas, not otherwise specified, identifies a subgroup with chromosomal aberrations affecting the REL locus. *British journal of haematology* 2010 Feb; **148**(3): 402-412.

17. Reddy A, Zhang J, Davis NS, Moffitt AB, Love CL, Waldrop A, *et al.* Genetic and Functional Drivers of Diffuse Large B Cell Lymphoma. *Cell* 2017 Oct 5; **171**(2): 481-494 e415.

18. Swerdllow S, Campo E, Harris NL. *WHO classification of tumours of haematopoietic and lymphoid tissues*. France: IARC Press, 2008.

1

2 19. Kataoka K, Nagata Y, Kitanaka A, Shiraishi Y, Shimamura T, Yasunaga J, *et al.*
3 Integrated molecular analysis of adult T cell leukemia/lymphoma. *Nature genetics* 2015
4 Nov; **47**(11): 1304-1315.

5

6 20. Yoo HY, Sung MK, Lee SH, Kim S, Lee H, Park S, *et al.* A recurrent inactivating mutation
7 in RHOA GTPase in angioimmunoblastic T cell lymphoma. *Nature genetics* 2014 Apr;
8 **46**(4): 371-375.

9

10 21. Crescenzo R, Abate F, Lasorsa E, Tabbo F, Gaudiano M, Chiesa N, *et al.* Convergent
11 mutations and kinase fusions lead to oncogenic STAT3 activation in anaplastic large cell
12 lymphoma. *Cancer cell* 2015 Apr 13; **27**(4): 516-532.

13

14 22. Jiang L, Gu ZH, Yan ZX, Zhao X, Xie YY, Zhang ZG, *et al.* Exome sequencing identifies
15 somatic mutations of DDX3X in natural killer/T-cell lymphoma. *Nature genetics* 2015
16 Sep; **47**(9): 1061-1066.

17

18 23. Kataoka K, Iwanaga M, Yasunaga JI, Nagata Y, Kitanaka A, Kameda T, *et al.* Prognostic
19 relevance of integrated genetic profiling in adult T-cell leukemia/lymphoma. *Blood* 2018
20 Jan 11; **131**(2): 215-225.

21

22 24. Odejide O, Weigert O, Lane AA, Toscano D, Lunning MA, Kopp N, *et al.* A targeted
23 mutational landscape of angioimmunoblastic T-cell lymphoma. *Blood* 2014 Feb 27;
24 **123**(9): 1293-1296.

25

25. da Silva Almeida AC, Abate F, Khiabanian H, Martinez-Escala E, Guitart J, Tensen CP, *et al.* The mutational landscape of cutaneous T cell lymphoma and Sezary syndrome. *Nature genetics* 2015 Dec; **47**(12): 1465-1470.

26. Kiel MJ, Sahasrabuddhe AA, Rolland DC, Velusamy T, Chung F, Schaller M, *et al.* Genomic analyses reveal recurrent mutations in epigenetic modifiers and the JAK-STAT pathway in Sezary syndrome. *Nature communications* 2015 Sep 29; **6:** 8470.

27. Choi J, Goh G, Walradt T, Hong BS, Bunick CG, Chen K, *et al.* Genomic landscape of cutaneous T cell lymphoma. *Nature genetics* 2015 Sep; **47**(9): 1011-1019.

28. de Miranda NF, Georgiou K, Chen L, Wu C, Gao Z, Zaravinos A, *et al.* Exome sequencing reveals novel mutation targets in diffuse large B-cell lymphomas derived from Chinese patients. *Blood* 2014 Oct 16; **124**(16): 2544-2553.

29. Okosun J, Bodor C, Wang J, Araf S, Yang CY, Pan C, *et al.* Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma. *Nature genetics* 2014 Feb; **46**(2): 176-181.

30. Okosun J, Wolfson RL, Wang J, Araf S, Wilkins L, Castellano BM, *et al.* Recurrent mTORC1-activating RRAGC mutations in follicular lymphoma. *Nature genetics* 2016 Feb; **48**(2): 183-188.

31. Mansouri L, Noerenberg D, Young E, Mylonas E, Abdulla M, Frick M, *et al.* Frequent NFKBIE deletions are associated with poor outcome in primary mediastinal B-cell

1 lymphoma. *Blood* 2016 Dec 8; **128**(23)**:** 2666-2670.

2

3 32. Yoshizato T, Nannya Y, Atsuta Y, Shiozawa Y, Iijima-Yamashita Y, Yoshida K, *et al.*
4 Genetic abnormalities in myelodysplasia and secondary acute myeloid leukemia: impact
5 on outcome of stem cell transplantation. *Blood* 2017 Apr 27; **129**(17)**:** 2347-2358.

6

7 33. Shiraishi Y, Sato Y, Chiba K, Okuno Y, Nagata Y, Yoshida K, *et al.* An empirical Bayesian
8 framework for somatic mutation detection from cancer genome sequencing data. *Nucleic*
9 *Acids Res* 2013 Apr; **41**(7)**:** e89.

10

11 34. Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, *et al.* STAR: ultrafast
12 universal RNA-seq aligner. *Bioinformatics* 2013 Jan 1; **29**(1)**:** 15-21.

13

14 35. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for
15 assigning sequence reads to genomic features. *Bioinformatics* 2014 Apr 1; **30**(7)**:**
16 923-930.

17

18 36. Kulkarni MM. Digital multiplexed gene expression analysis using the NanoString
19 nCounter system. *Curr Protoc Mol Biol* 2011 Apr; **Chapter 25:** Unit25B 10.

20

21 37. Kataoka K, Shiraishi Y, Takeda Y, Sakata S, Matsumoto M, Nagano S, *et al.* Aberrant
22 PD-L1 expression through 3'-UTR disruption in multiple cancers. *Nature* 2016 Jun 16;
23 **534**(7607)**:** 402-406.

24

38. Cancer Genome Atlas Research N. Integrated genomic analyses of ovarian carcinoma. *Nature* 2011 Jun 29; **474**(7353): 609-615.

39. Shiraishi Y, Tremmel G, Miyano S, Stephens M. A Simple Model-Based Approach to Inferring and Visualizing Cancer Mutation Signatures. *PLoS Genet* 2015 Dec; **11**(12): e1005657.

40. Bieging KT, Mello SS, Attardi LD. Unravelling mechanisms of p53-mediated tumour suppression. *Nat Rev Cancer* 2014 May; **14**(5): 359-370.

41. Abate F, da Silva-Almeida AC, Zairis S, Robles-Valero J, Couronne L, Khiabanian H, *et al.* Activating mutations and translocations in the guanine exchange factor VAV1 in peripheral T-cell lymphomas. *Proceedings of the National Academy of Sciences of the United States of America* 2017 Jan 24; **114**(4): 764-769.

42. Wang C, McKeithan TW, Gong Q, Zhang W, Bouska A, Rosenwald A, *et al.* IDH2R172 mutations define a unique subgroup of patients with angioimmunoblastic T-cell lymphoma. *Blood* 2015 Oct 8; **126**(15): 1741-1752.

43. Cairns RA, Iqbal J, Lemonnier F, Kucuk C, de Leval L, Jais JP, *et al.* IDH2 mutations are frequent in angioimmunoblastic T-cell lymphoma. *Blood* 2012 Feb 23; **119**(8): 1901-1903.

44. Vasmatzis G, Johnson SH, Knudson RA, Ketterling RP, Braggio E, Fonseca R, *et al.* Genome-wide analysis reveals recurrent structural abnormalities of TP63 and other

1 p53-related genes in peripheral T-cell lymphomas. *Blood* 2012 Sep 13; **120**(11):
2 2280-2289.

3

4 45. Matsushima AY, Cesarman E, Chadburn A, Knowles DM. Post-thymic T cell lymphomas
5 frequently overexpress p53 protein but infrequently exhibit p53 gene mutations. *The*
6 *American journal of pathology* 1994 Mar; **144**(3): 573-584.

7

8 46. Sherr CJ. Divorcing ARF and p53: an unsettled case. *Nat Rev Cancer* 2006 Sep; **6**(9):
9 663-673.

10

11 47. Iqbal J, Wright G, Wang C, Rosenwald A, Gascoyne RD, Weisenburger DD, *et al.* Gene
12 expression signatures delineate biological and prognostic subgroups in peripheral T-cell
13 lymphoma. *Blood* 2014 May 8; **123**(19): 2915-2923.

14

15 48. Wartewig T, Kurgyis Z, Keppler S, Pechloff K, Hameister E, Ollinger R, *et al.* PD-1 is a
16 haploinsufficient suppressor of T cell lymphomagenesis. *Nature* 2017 Dec 7; **552**(7683):
17 121-125.

18

19 49. Asanuma S, Yamagishi M, Kawanami K, Nakano K, Sato-Otsubo A, Muto S, *et al.* Adult
20 T-cell leukemia cells are characterized by abnormalities of Helios expression that
21 promote T cell growth. *Cancer science* 2013 Aug; **104**(8): 1097-1106.

22

23 50. Childs KS, Goodbourn S. Identification of novel co-repressor molecules for Interferon
24 Regulatory Factor-2. *Nucleic Acids Res* 2003 Jun 15; **31**(12): 3016-3026.

51. Carneiro FR, Ramalho-Oliveira R, Mognol GP, Viola JP. Interferon regulatory factor 2 binding protein 2 is a new NFAT1 partner and represses its transcriptional activity. *Molecular and cellular biology* 2011 Jul; **31**(14): 2889-2901.

52. Bruno A, Boisselier B, Labreche K, Marie Y, Polivka M, Jouvet A , *et al.* Mutational analysis of primary central nervous system lymphoma. *Oncotarget* 2014 Jul 15; **5**(13): 5065-5075.

53. Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D , *et al.* N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* 2014 Jan 2; **505**(7481): 117-120.

54. Cuadros M, Dave SS, Jaffe ES, Honrado E, Milne R, Alves J , *et al.* Identification of a proliferation signature related to survival in nodal peripheral T-cell lymphomas. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology* 2007 Aug 1; **25**(22): 3321-3329.

Figure Legends

Figure 1 WES analysis for 20 PTCL, NOS cases.

A, Number of coding mutations (top), frequency of mutational signature (middle), and a heat map showing the distribution of mutations in *TET2*, *RHOA*, *IDH2*, *TP53*, and *CDKN2A* alterations, TFH markers, sample source, and molecular subtype are depicted. All panels are aligned, with the vertical tracks representing 20 PTCL, NOS cases. **B,** Two mutational signatures identified by pmsignature algorithm in PTCL, NOS. Signature 1 was predominated by age-related C > T transitions at CpG dinucleotides, while signature 2 (of unknown etiology) consisted of C > A substitutions at CpCpT context. **C,** Hierarchy of somatic mutations is shown with their allele frequencies in four representative PTCL, NOS cases. Lymphoma-associated alterations are shown in green, and mutations that were undetectable by WES but were identified later by targeted-capture sequencing are shown in red. Mutations located in non-amplified regions and *TP53/CDKN2A* deletions are depicted. **D,** Boxplots showing the number of somatic mutations identified by WES in cases with or without *TP53* alterations. $^{**}P < 0.005$, Wilcoxon rank-sum test.

1

2 **Figure 2 Landscape of somatic alterations in PTCL, NOS.**

3 **A,** Frequency and type of somatic alterations in 49 recurrently altered genes (found in ≥
4 3 cases) for 133 PTCL, NOS cases, including 127 cases from our series and 6 cases
5 from a previous study. Genes not previously reported as altered in PTCL, NOS are
6 shown in red. **B,** Co-mutation plot showing the spectrum of somatic alterations in
7 recurrently altered genes (n = 49) and chromosome arms (n = 14) across 133 PTCL,
8 NOS cases. Samples were organized by hierarchical clustering with Spearman's rank
9 correlation and Ward's linkage algorithm. Molecular subtype, experimental platform, IHC
10 (TFH markers, GATA3, and TBX21) as well as related functional pathways (right) are
11 also shown. Other *RHOA* mutations (3 and 2 cases in group 2 and 3, respectively) are
12 shown in a different color from G17V mutation.

13

14 **Figure 3 Commonly affected functional pathways in PTCL, NOS.**

15 Driver alterations, including mutations, CNAs, and SVs, are summarized according to

1 their functionalities. Frequencies of mutations (left) and CNAs/SVs (right) are
2 expressed as the percentage of altered cases in 133 PTCL, NOS cases. Major
3 determinants of the molecular classificationare highlighted by green (group 1) or red
4 (group 2) boxes. Gain-of-function mutations and activating CNAs/SVs shown in red,
5 and loss-of-function mutations and disrupting CNAs/SVs are shown in blue.
6
7 **Figure 4 Co-occurring and mutually exclusive associations define two**
8 **molecular subtypes in PTCL, NOS.**

9 **A,** Comparison of frequencies of recurrent somatic alterations between patients with 37
10 TFH and 36 non-TFH PTCL, NOS. Recurrently altered genes (n = 19) present in ≥ 10
11 cases (7%) in the entire cohort are shown. $^{*}q < 0.1$, $^{**}q < 0.01$, Fisher's exact test with
12 Benjamini–Hochberg correction. **B,** Pairwise associations among 19 recurrently
13 altered genes found in ≥ 10 cases (7%) in the entire cohort. Only significant
14 correlations (OR > 10 and $q < 0.1$, Fisher's exact test with Benjamini–Hochberg
15 correction) are shown with their odds ratios (OR). Orange colors depict gene pairs that
16 are co-mutated more than expected by chance, and blue colors depict mutually

exclusive gene pairs. **C,** Two molecular subtypes [group 1(TFH-related) and group 2 (*TP53/CDKN2A*)] and their major determinants in PTCL, NOS. Orange and blue lines represent co-occurring and mutually exclusive associations, respectively. The line width is proportional to the statistical significance (q value) of the association. Genes showing at least 2 significant associations in **(B)** are shown.

Figure 5 *TP53/CDKN2A*-altered PTCL, NOS shows a distinct genetic features characterized by chromosome instability and cell cycle dysregulation.

A, Comparison of frequencies of TFH marker positivity among molecular subtypes. Fisher's exact test. **B,** The heat map shows somatic CNA segments (copy number gains/amplifications, losses/deletions, and uniparental disomies (UPD) with ≥ 10 probes) in each sample (horizontal axis) plotted by chromosomal location (vertical axis). Samples are vertically aligned in the same order as in **Figure 2B**. **C,** Number of abnormal chromosomal segments identified in cases with indicated alterations, regardless of presence or absence of other alterations. Each dot represents a single case. ** $P < 0.005$, *** $P < 0.0005$, Wilcoxon rank-sum test with Bonferroni correction.

1 **D,** Significantly enriched gene signatures for group 1 (left) and 2 (right) in GSEA
2 analysis with RNA-seq data from 16 cases (4, 6, and 6 cases in group 1, 2, and 3,
3 respectively) using the hallmark (top) and C2 (bottom) gene sets. **E,** Dot plots of
4 normalized counts of *GATA3* (left) and *TBX21* (right) expressions measured by
5 nCounter analysis in each group (10, 13, and 11 cases in group 1, 2, and 3, respectively) .
6 Wilcoxon rank-sum test.

7

8 **Figure 6 Clinical and genetic differences among three molecular subtypes**
9 **of PTCL, NOS.**

10 **A,** Comparison of tumor cell fraction among three molecular subtypes of PTCL, NOS.
11 Fisher's exact test. **B,** Kaplan-Meier survival curves of overall survival of 46 PTCL,
12 NOS cases stratified by molecular subtype. The prognostic impact on survival was
13 evaluated by log-rank test. **C,** Comparison of frequencies of somatic alterations
14 among the entire cohort (n = 133) and each molecular subtype (n = 50, 42, and 41 for
15 group 1, 2, and 3, respectively) of PTCL, NOS, AITL (n = 26), ATL (n = 81) [19], ALCL (n =
16 23) [21], and ENKTL (n = 25). [22] Diagonal lines represent no data available. **D,**

1 Proportion of the number of somatic alterations belonging to each functional pathway

2 among three molecular subtypes of PTCL, NOS. **q < 0.01, ***q < 0.001, Fisher's

3 exact test with Benjamini–Hochberg correction.

4

5 **Figure 7 Genetic alterations in molecules associated with immune evasion and**

6 **transcriptional regulation.**

7 **A,** Positions and types of somatic mutations encoded in *HLA-A* (NM_002116), *PDCD1*

8 (NM_005018), *IRF2BP2* (NM_001077397), and *YTHDF2* (NM_001173128) detected in

9 133 PTCL, NOS cases. *IRF2BP2* mutations observed in 81 ATL cases are also shown.

10 **B,** Validation of CRISPR/Cas9-mediated targeting of *IRF2BP2* gene by amplicon

11 sequencing. Representative sequencing data for mock- (left) and sgIRF2BP2-1-

12 (right) transfected samples were visualized with IGV. **C,** Relative expression of

13 *IRF2BP2* mRNA in Jurkat cells transfected with the indicated sgRNA vectors (n = 4).

14 Expression values were normalized to the mock-transfected control. Data represent

15 means ± s.d., ****P* < 0.0005, Student's t test with Welch's correction. **D,** Immunoblot

16 analysis of IRF2BP2 in Jurkat cells transfected with the indicated sgRNA vectors.

Representative result of two independent experiments. **E,** Luciferase assays of NFAT transcriptional activity in Jurkat cells transfected with the indicated sgRNA vectors with or without NFAT1 expression vector, in the presence or absence of CD3/CD28 stimulation (n = 3-5 biological replicates, respectively). RLU, relative luminometer units. Data represent means ± s.d. *$P < 0.05$, **$P < 0.005$, Student's t test with Welch's correction.

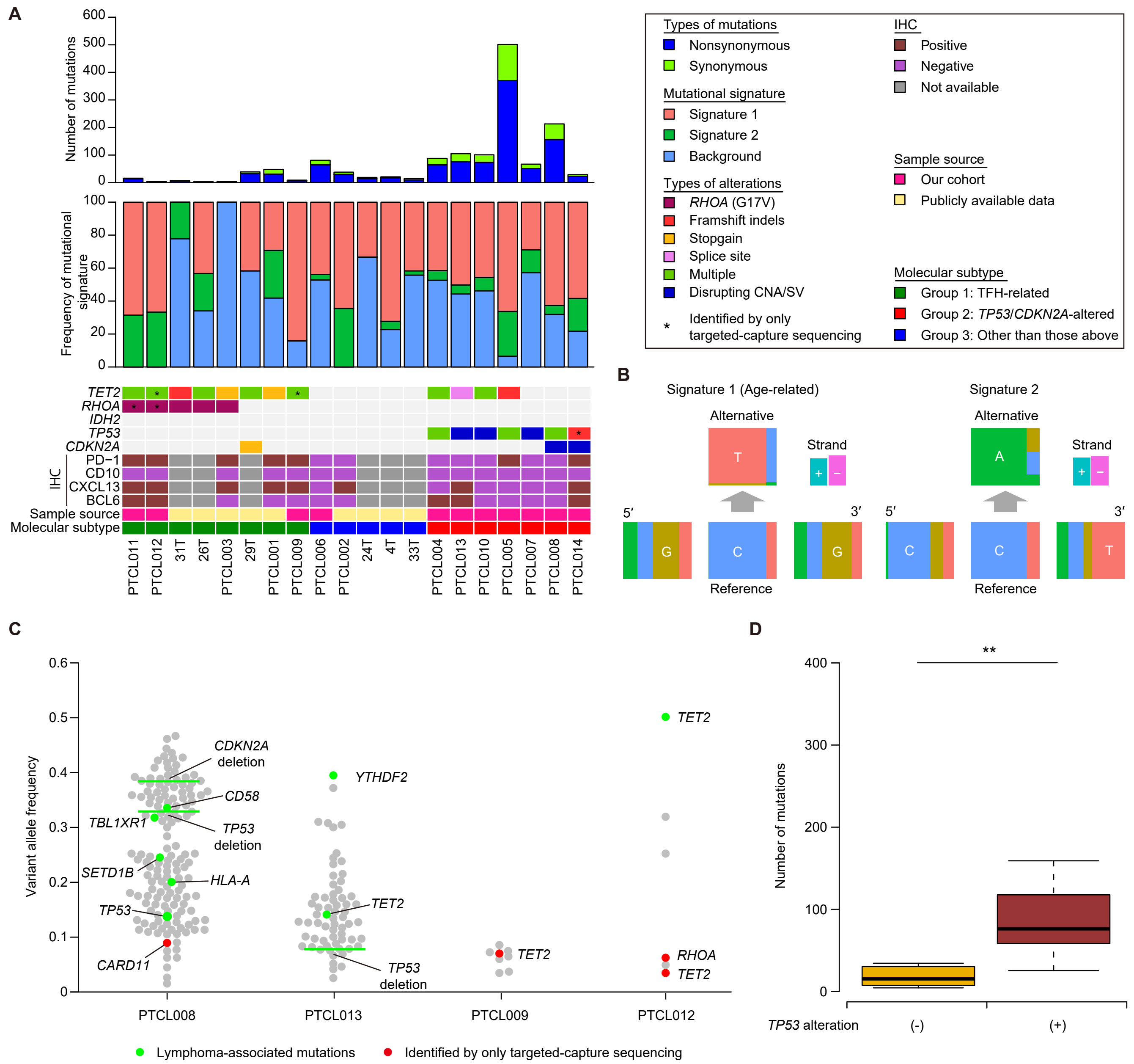


Figure 1

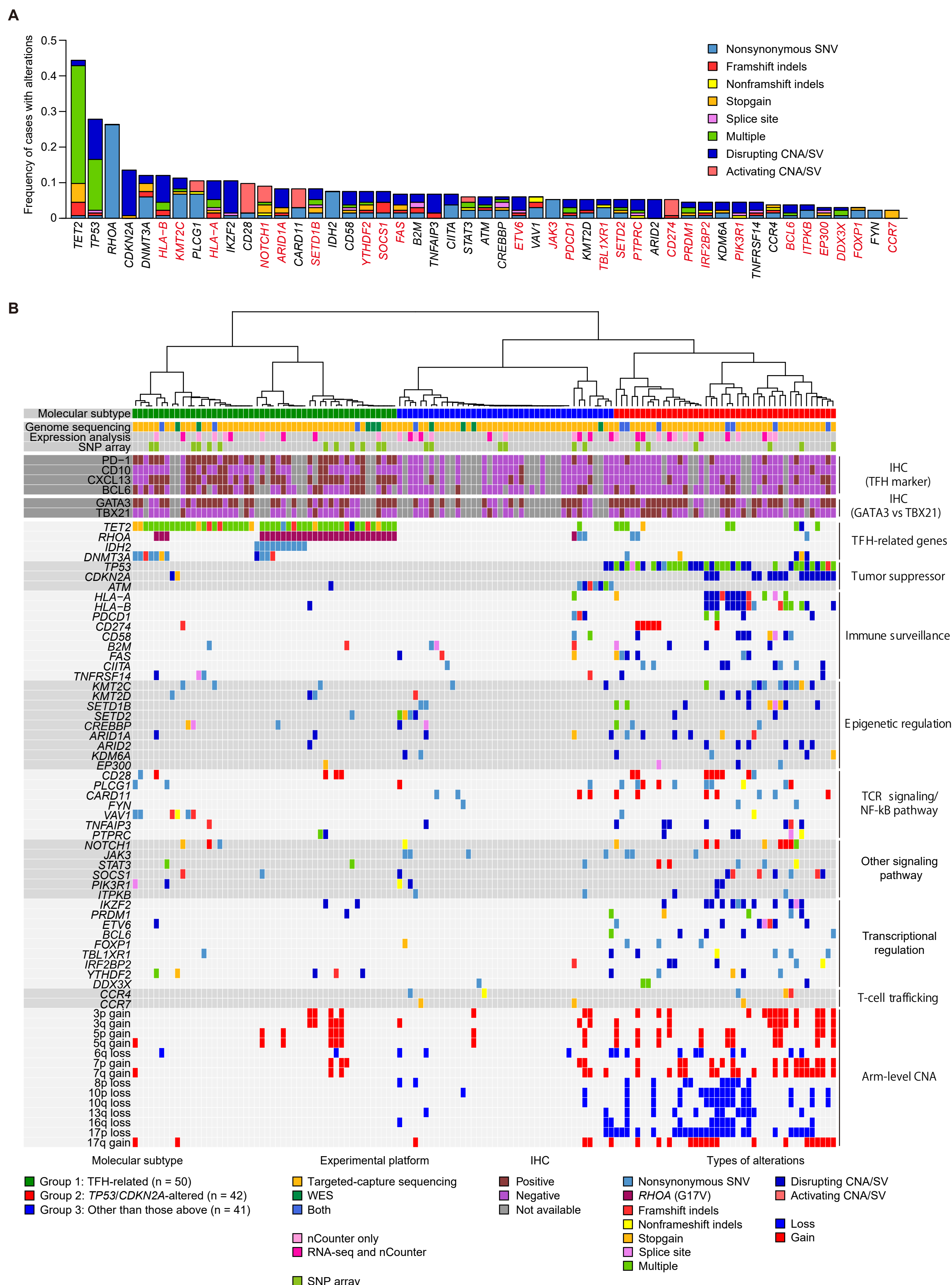


Figure 2

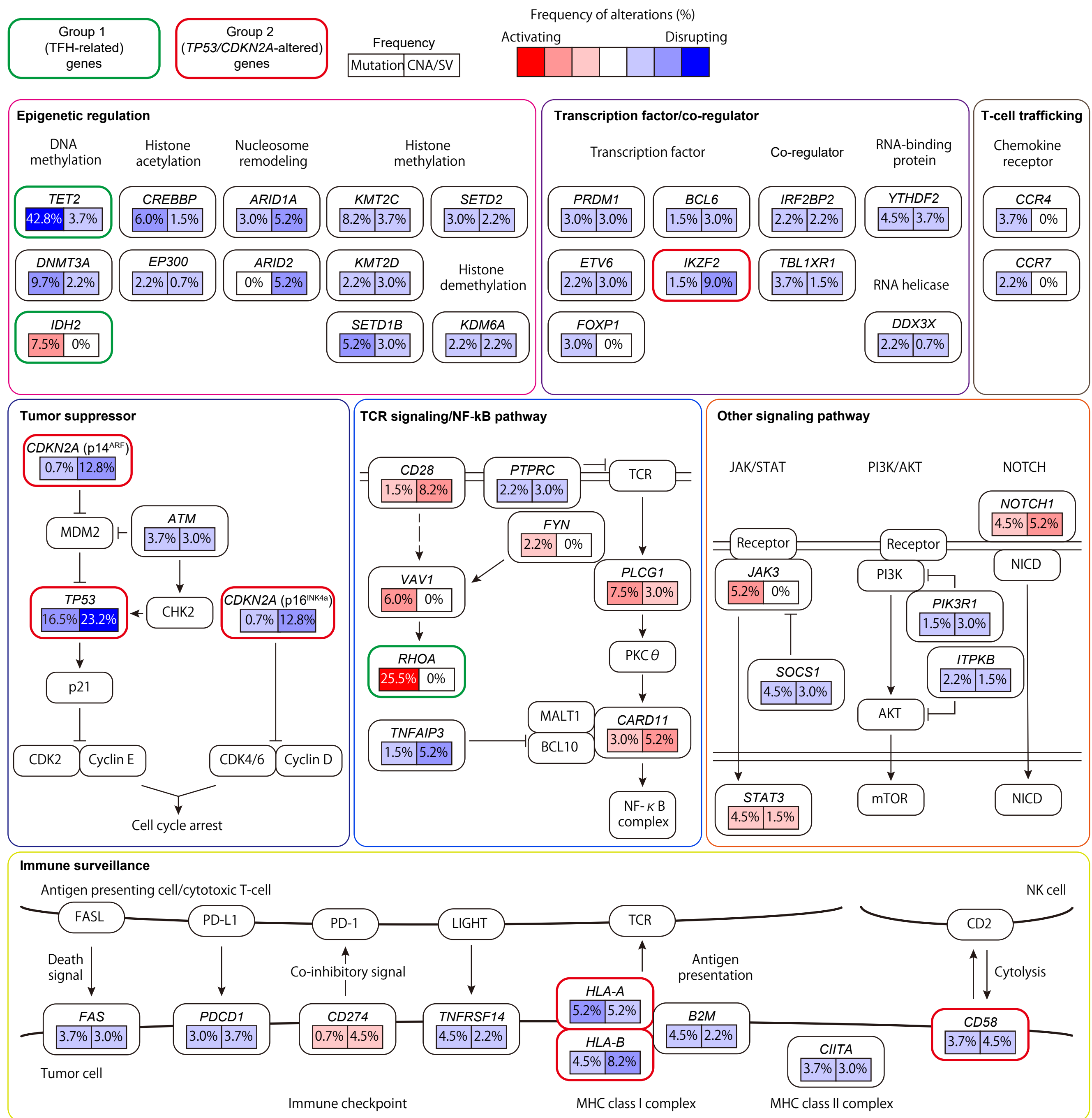
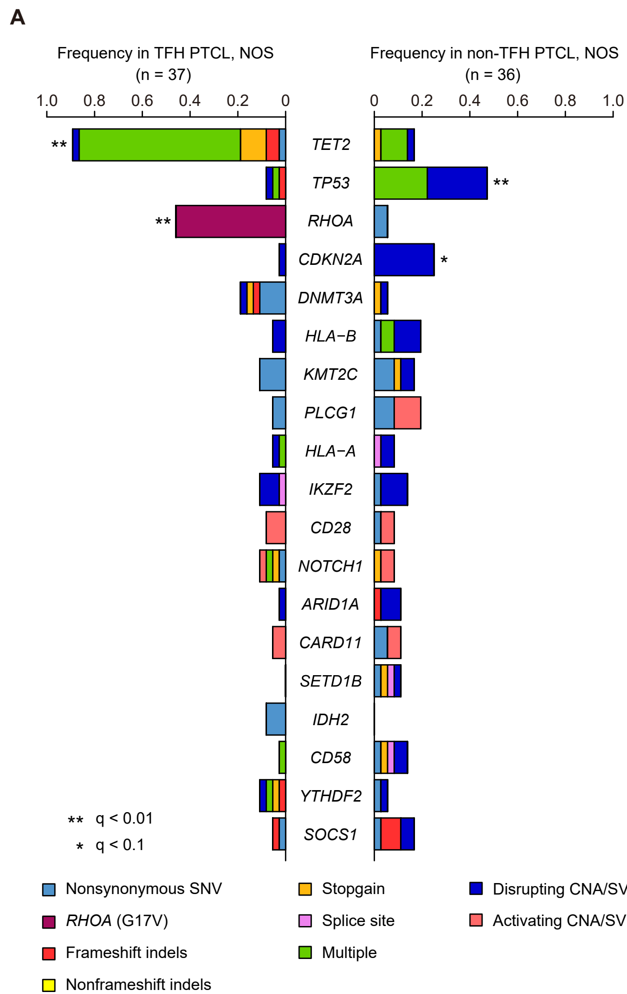
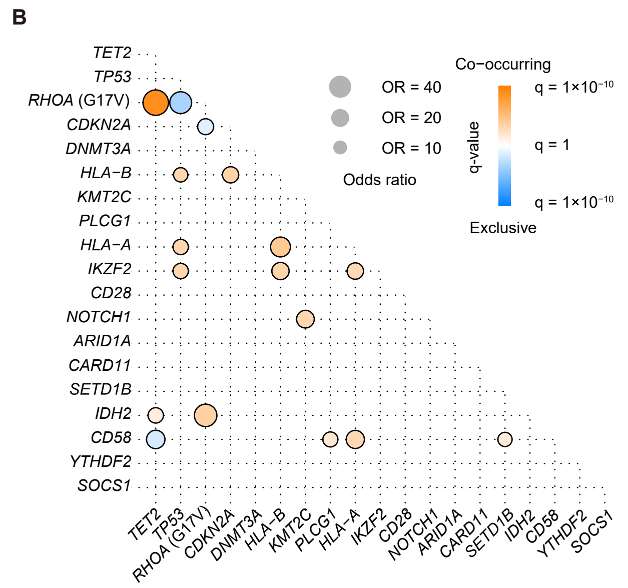


Figure 3





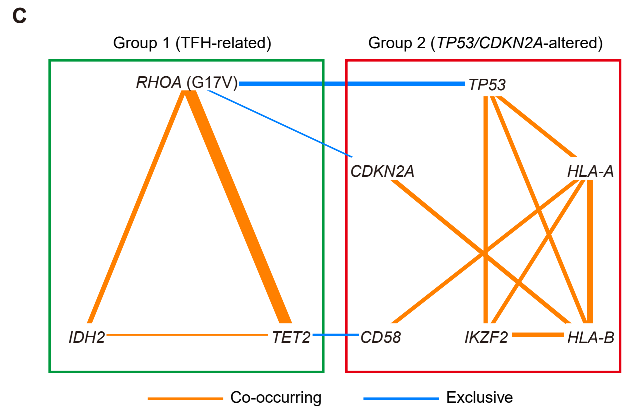


Figure 4

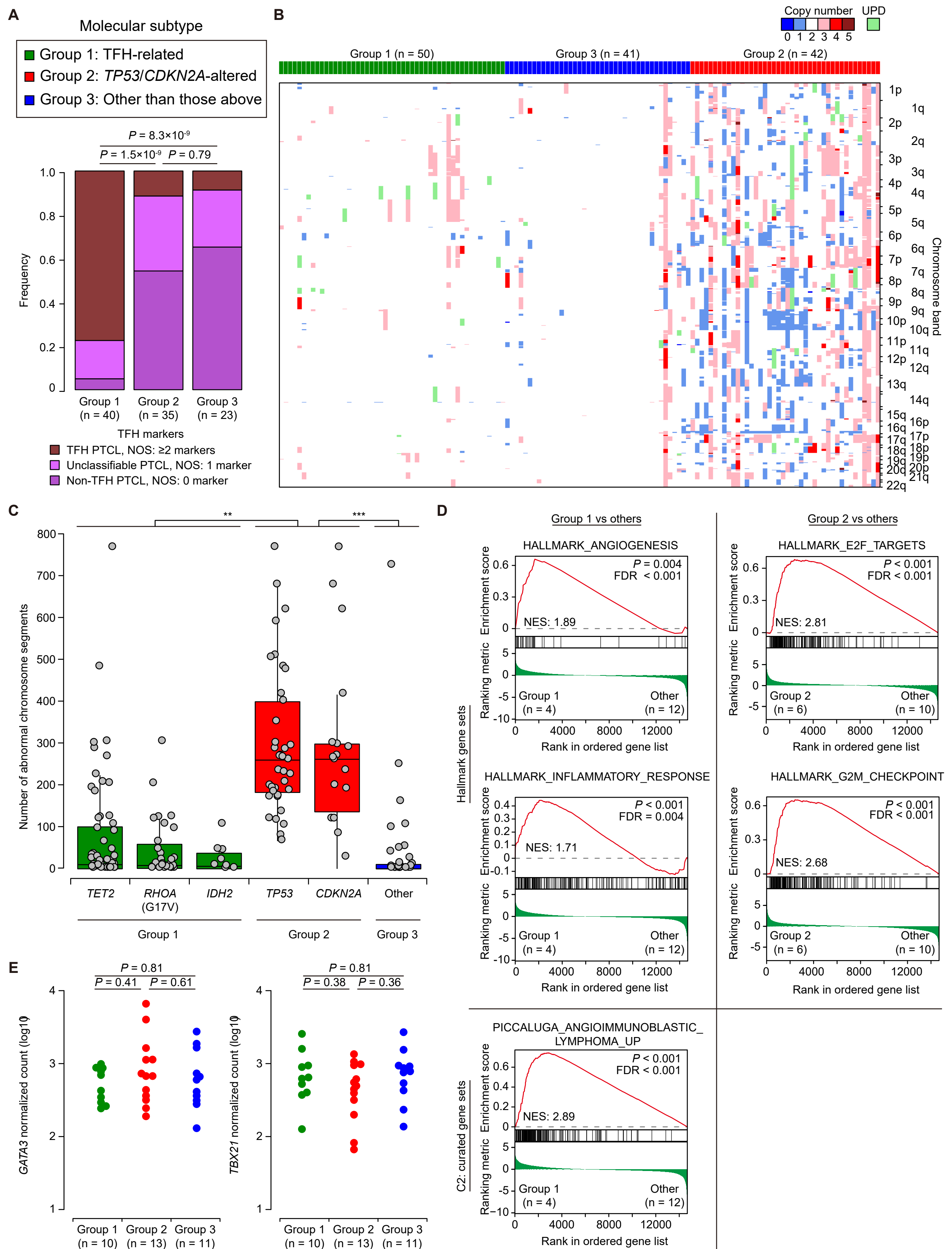


Figure 5

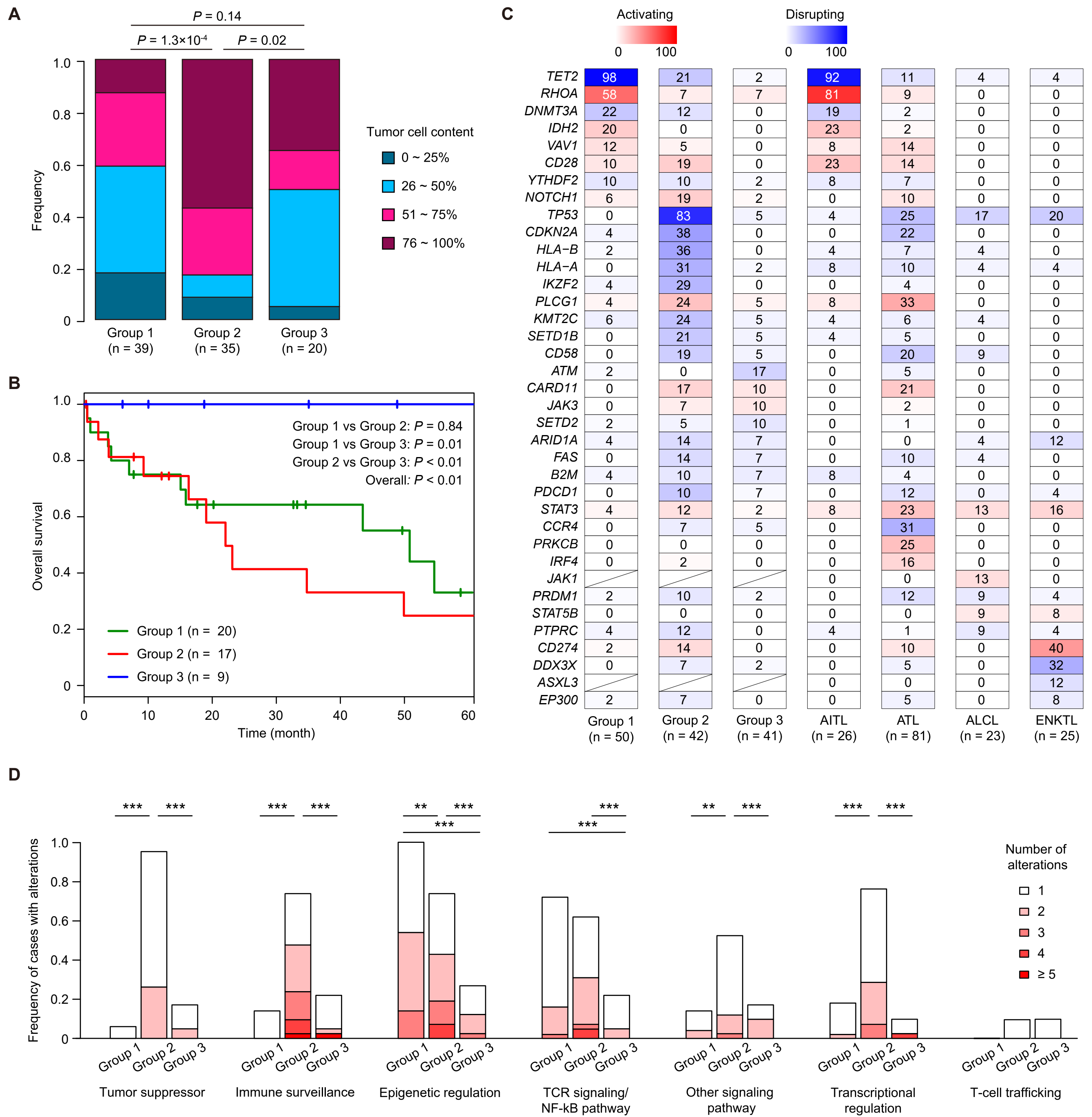


Figure 6

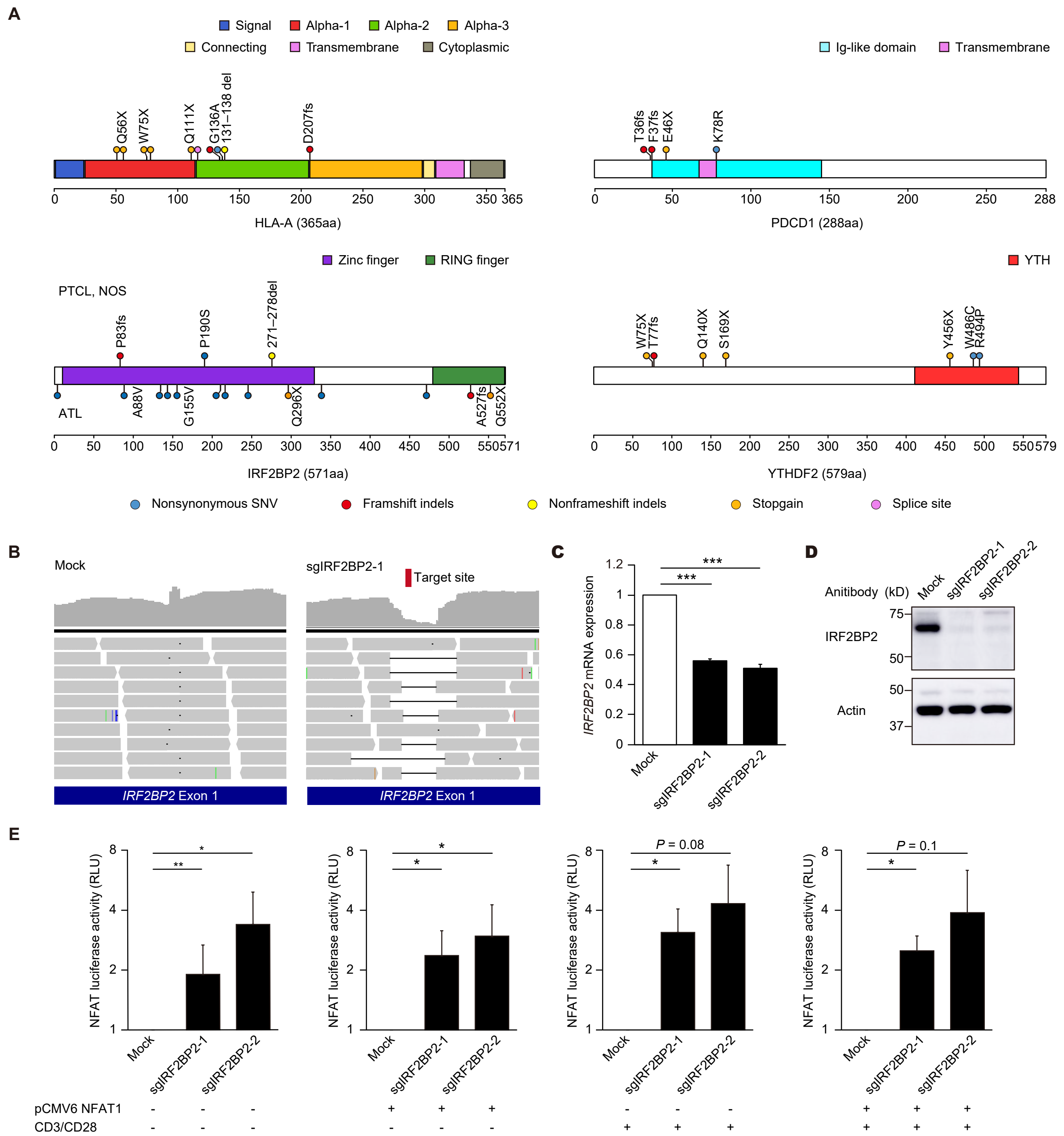


Figure 7