Supplementary Methods

Whole exome sequencing and bioinformatics

Paris. For patients A1 and G1, exome capture was performed according to the manufacturer's protocol using the Illumina TruSeq exome enrichment kit and sequencing of 100 bp paired end reads on an Illumina HiSeq. Approximately 10 Gb of sequence were obtained for each subject such that 90% of the coding bases of the exome defined by the consensus coding sequence (CCDS) project were covered by at least 10 reads. Adaptor sequences and quality trimmed reads were removed using the Fastx toolkit (http://hannonlab. cshl.edu/fastx_toolkit/) and a custom script was then used to ensure that only read pairs with both mates present were subsequently used. Reads were aligned to hg19 with BWA31, and duplicate reads were marked using Picard (http://picard.sourceforge.net/) and excluded from downstream analyses. Single nucleotide variants (SNVs) and short insertions and deletions (indels) were determined using samtools (http://samtools.sourceforge.net/) pileup and varFilter32 with the base alignment quality (BAQ) adjustment disabled, they were then quality filtered to require at least 20% of reads supporting the variant call. Variants were annotated using both ANNOVAR33 and custom scripts to identify whether they affected protein coding sequences, and whether they had previously been seen in the public data bases of exomes and the 9723 exomes previously sequenced at our center.

US. For family C, WES was performed using the Agilent SureSelect^{XT} Clinical Reseearch Exome kit to target known disease-associated exonic regions of the genome (coding sequences and splice junctions of known protein-coding genes associated with disease, as well as an exomic backbone) using gDNA. The targeted regions were sequenced using the Illumina NextSeq® 500 System with 150 bp paired ends reads. Using NextGENe® software and a proprietary bioinformatics pipeline, the DNA sequence was aligned and compared to the human genome build 19(hg19/NCBI build 37). The average depth of coverage across the targeted

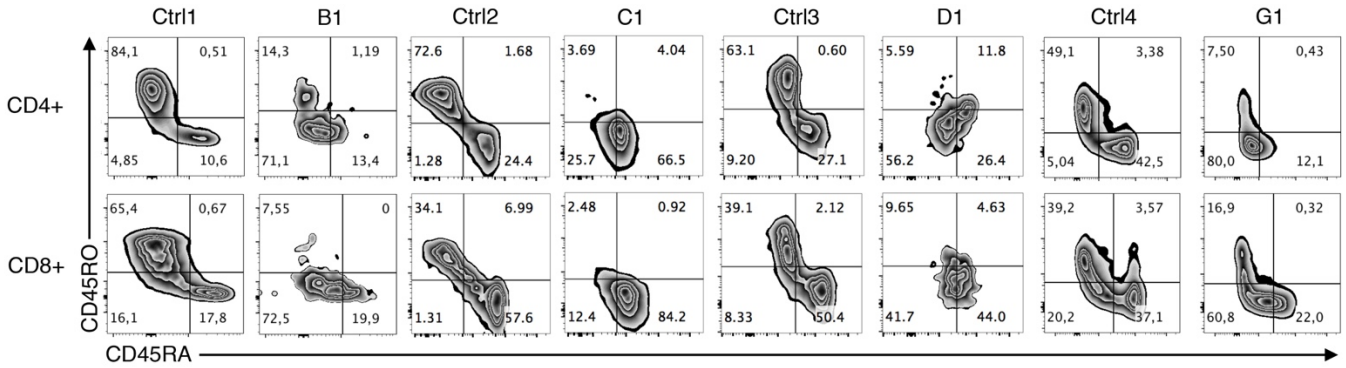
regions was calculated. The Cartagenia Bench Lab NGS software was used to filter and analyze sequence variants identified in the patient and compare them to the sequences of family members. All reported variants were confirmed by Sanger sequencing for the patient and the family members. For family D, WES was performed using the SeqCap EZ Human Exome v3.0 (Roche) enrichment kit followed by sequencing using the Illumina HiSeq 2500 sequencer (100bp paired end). Causal variants were identified using Ingenuity ® Variant Analysis ™ software (www.ingenuity.com/variants) from Qiagen and NextGENe software from SoftGenetics (StateCollege, PA). The sequence reads were filtered for quality, then matched to GRCh37/hg19 genome reference sequence. Average target coverage of 100X and 96% of target of 10X depth was obtained. In the target region 89,694 non-redundant variants in 17,099 genes were identified in the family. Variants were then filtered based on frequency, nucleotide and amino acid sequence conservation, *in silico* pathogenicity predictions and models of genetic inheritance. For patient E1, WES was performed using the Ion Torrent AmpliSeq RDY Exome Kit (Life Technologies) and the Ion Chef and Proton instruments (Life Technologies). Briefly, 100 ng of gDNA was used as the starting material for the AmpliSeq RDY Exome amplification step following the manufacturer's protocol. Library templates were clonally amplified and enriched using the Ion Chef and the Ion PI Hi-Q Chef Kit (Chef package version IC.4.4.2, Life Technologies), following the manufacturer's protocol. Enriched, templated Ion Sphere Particles were sequenced on the Ion Proton sequencer using the Ion PI chip v3 (Life Technologies). Read mapping and variant calling were performed using the Ion Torrent Suite software v4.4.2. In short, sequencing reads were mapped against the UCSC hg19 reference genome using the Torrent Mapping Alignment Program (TMAP) map4 algorithm. SNPs and INDELS were called by the Torrent Variant Caller plugin (v.4.414-1) using the 'Generic-Proton-Germ Line – Low Stringency' configuration. Only reads that were unambiguously mapped were used for variant calling. Variants were annotated using ANNOVAR (http://annovar.openbioinformatics.org/).

Data mining, biological interpretation, and candidate gene discovery were performed using various online tools, including The Database for Annotation, Visualization and Integrated Discovery (DAVID, https://david.ncifcrf.gov), and GeneCards (http://www.genecards.org). Target coverage was evaluated using the Torrent Coverage Analysis plugin (v.4.414-1), and the output was further evaluated using in-house, custom Perl scripts.

Japan. WES details have been previously provided elsewhere (2, 20).

DNA-seq exome data have been deposited to NCBI Sequence Read Archive (SRA) data base accessible through the SRA accession number SRP136278 (https://www.ncbi.nlm.nih.gov/sra/?term=SRP136278).

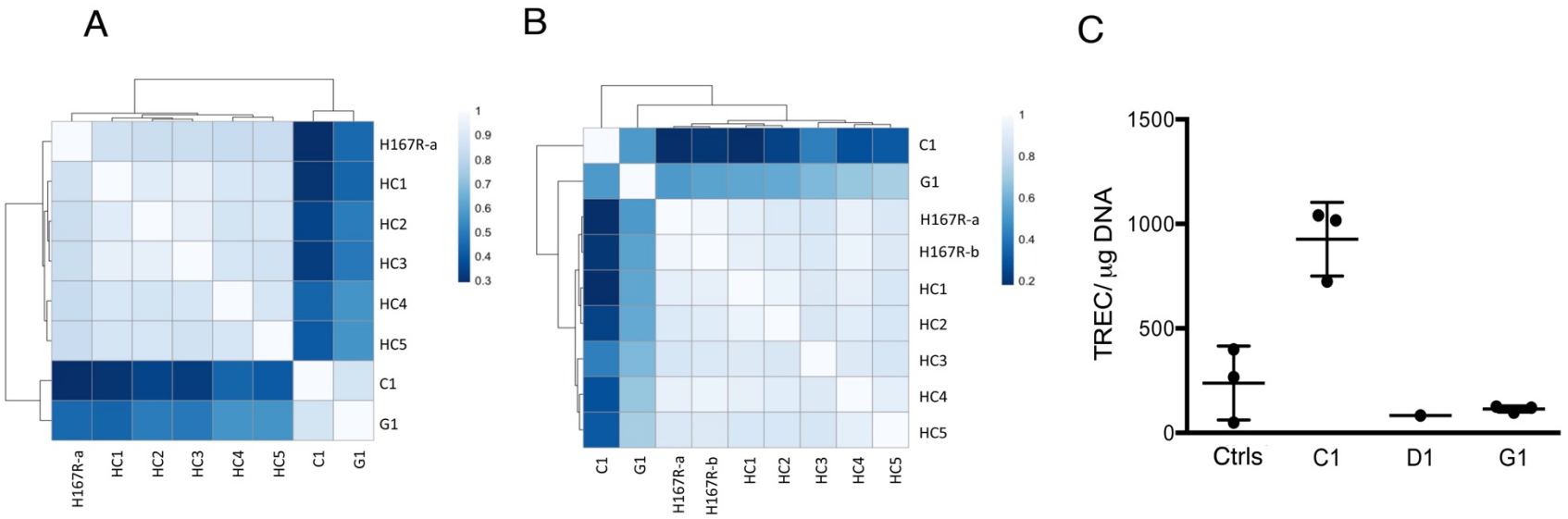
Supplemental Figure 1



Supplemental Figure 1. CD45RA and CD45RO expression in T cells from PBMCs of patients with *IKZF1*$^{N159S/T}$ mutations

Dot-plots from flow cytometric analysis showing the frequency of naïve ($CD45RA^+CD45RO^-$) and memory cells ($CD45RA^-CD45RO^+$) among $CD4^+$ and $CD8^+$ T cells from patients (B1, C1, D1 and G1) and healthy donor controls (Ctrl1, Ctrl2, Ctrl3 and Ctrl4). Data are representative of one experiment for patient B1 and two experiments for patient C1, D1 and G1.

Supplemental Figure 2



Supplemental Figure 2. Correlation analysis of targeted RNASeq data on naïve CD4⁺ T cells and monocytes from healthy donor controls and patients with IKZF1^{H167R}, IKZF1^{N159S} and IZKF1^{N159T} mutations and T cell receptor excision circles (TREC) analysis in PBMCs of patients with IKZF1N159T and IZKF1N159S mutations.
(A,B) The correlation analysis was performed between two siblings with IKZF1^{H167R} mutation (H167R-a and H167R-b) versus a patient with IKZF1^{N159S} mutation (C1) and a patient with IKZF1^{N159T} mutation (G1) along with five healthy controls (HC) on enriched unstimulated naïve CD4⁺ T cells (A) and LPS-stimulated monocytes (B). Scale bar indicates Pearson correlation coefficient for variance stabilized data for differentially expressed mRNAs.
(C) Real-time quantitative PCR for detecting T cell receptor excision circles (TREC) was performed on 25ng of genomic DNA from total PBMCs of 3 healthy age-matched pediatric controls (Ctrls) and patients with IKZF1^{N159S} (C1, D1) or IZKF1^{N159T} (G1) mutations. Housekeeping β-actin gene was tested to ensure the DNA samples were adequate (not shown). Data are means ± SD of three replicates for C1 and G1, and one experiment for D1. Each experiment was performed in triplicate.

Supplemental Figure 3

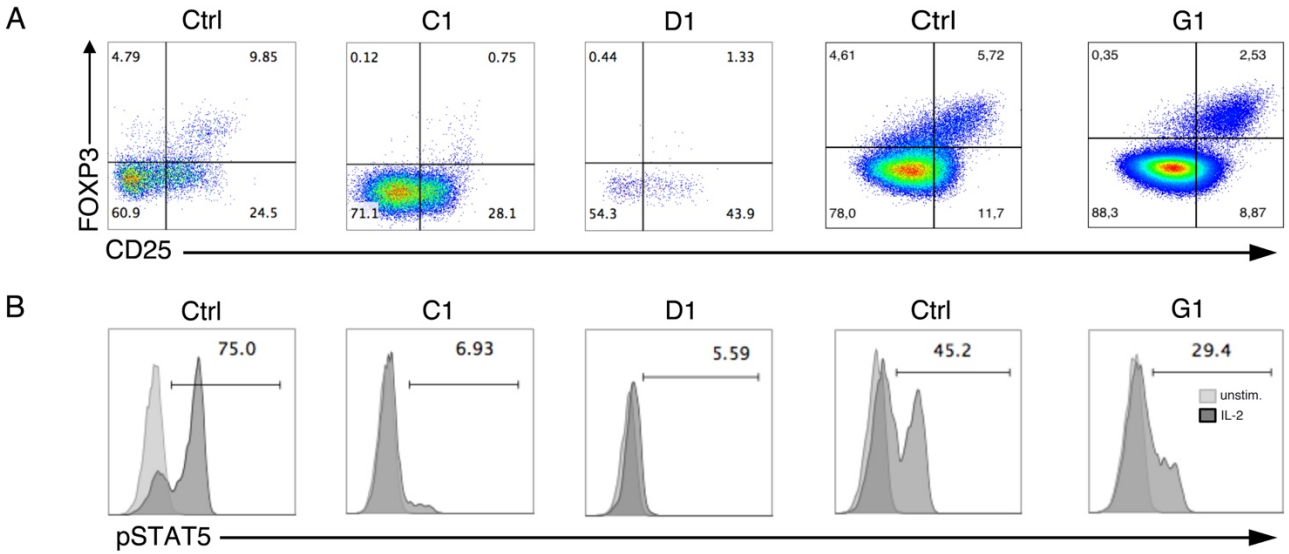
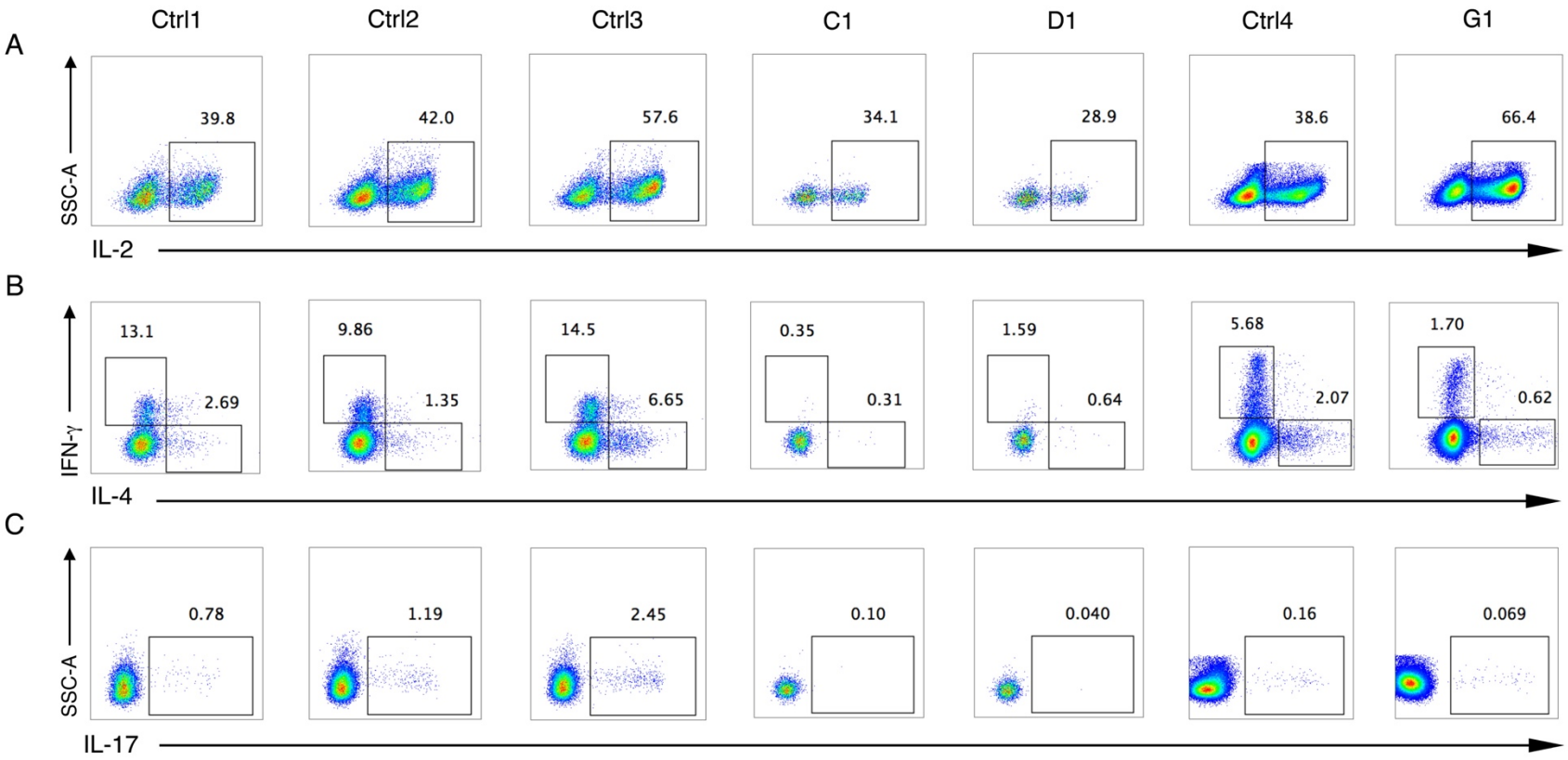


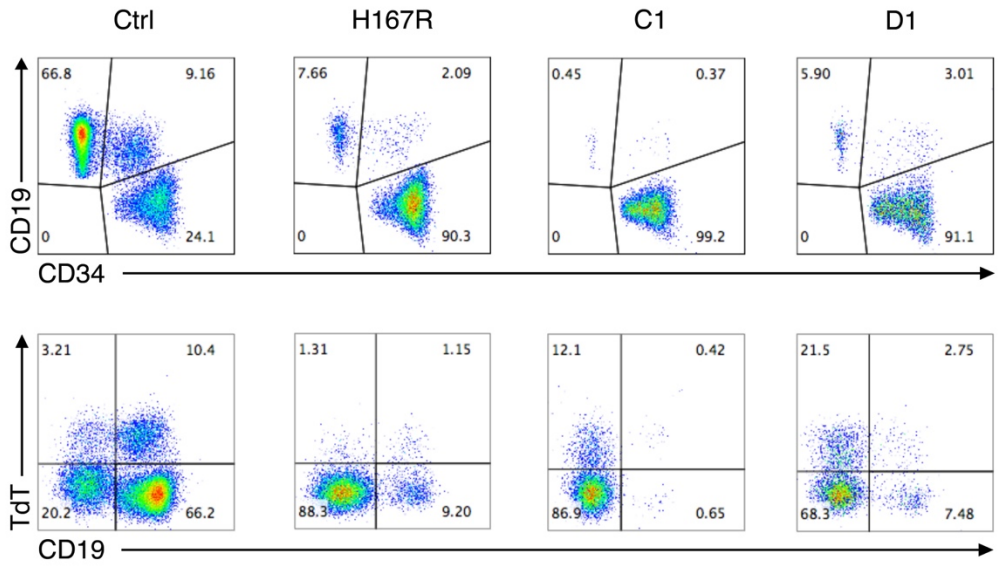
Figure 3. Defective STAT5 phosphorylation and low natural regulatory T cells in CD4+ T cells of patients with IKZF1N159S/T mutations
(A) Dot plots of FoxP3 and CD25 expression in CD4+ T cells from PBMCs of patients (C1, D1 and G1) and two paired healthy donor controls (Ctrl). Natural regulatory T cells are defined by the percentage of FoxP3+CD25+ cells among CD4+ T cells.
(B) Histograms in gated CD4+ T cells from PBMCs stimulated with (dark grey) or without (pale grey) IL-2 (10ng/ml) for 20 minutes. STAT5 (Y694) phosphorylation was analyzed by intracellular staining in patients (C1, D1 and G1) and two paired healthy donor controls (Ctrl). Data shown are representative of two experiments from C1 and D1 and one experiment from patient G1.

Supplemental Figure 4



Supplemental Figure 4. Normal IL-2 production and defective Th1/Th2/Th17 generation in CD4$^+$ T cells of patients with *IKZF1*$^{N159S/T}$ mutations
This figure is the dot-plot representation of the main figure 3F experiment and shows among CD4$^+$ T cells the percentage of cells positive for the indicated cytokine after 6h-PMA/ionomycin stimulation. Th1/Th2/Th17 flow cytometric phenotypes from three patients (C1, D1 and G1) were compared to three healthy donor controls (Ctrl).

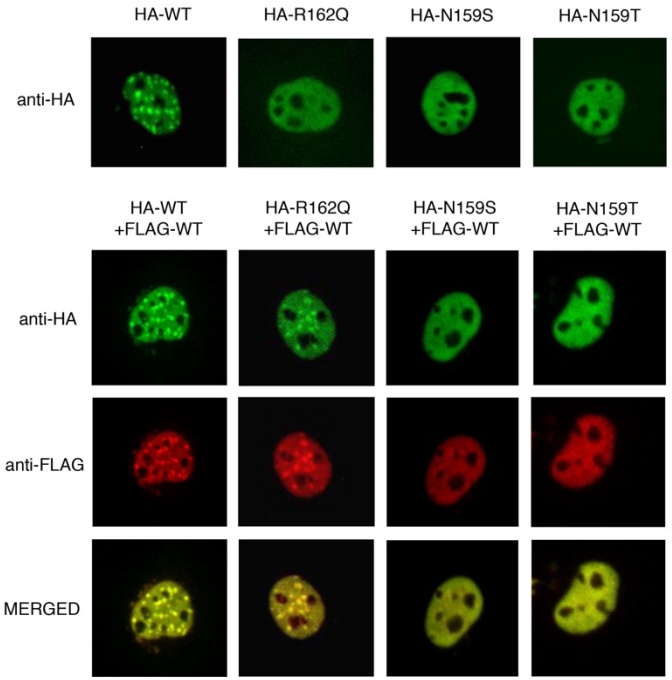
Supplemental Figure 5



Supplemental Figure 5. Bone marrow aspirates from patients with IKZF1^{N159S} and IKZF1^{H167R} mutations

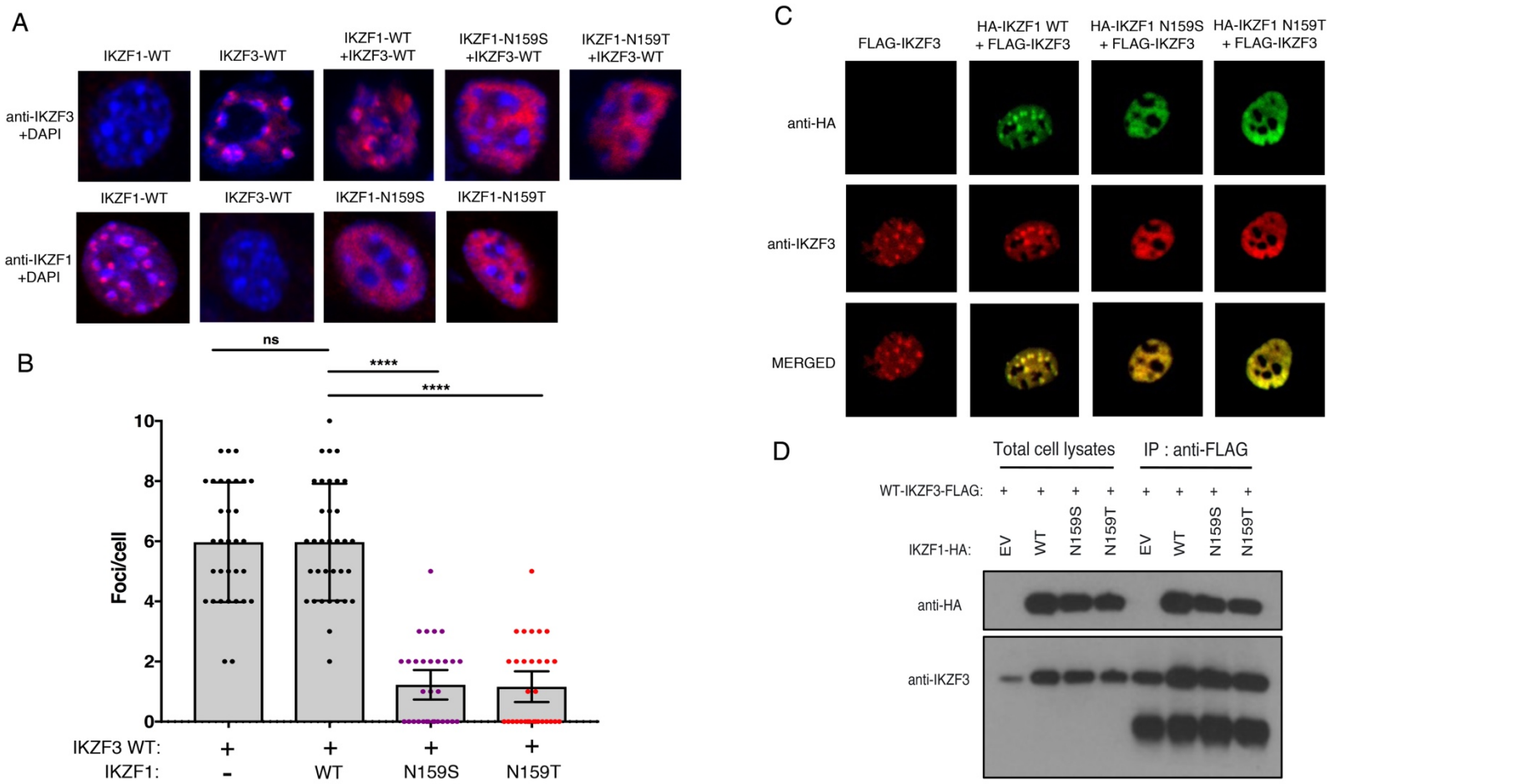
Flow-cytometric analysis of bone marrow aspirates from patients (C1 and D1) were compared to a healthy donor control (Ctrl) and to a previously published patient with IKZF1^{H167R} haplo-insufficient mutation (H167R). A gate was used that collected all cells that were positive for CD34, CD19, or both. Pro-B cells were defined as CD34^{+}CD19+ cells and precursors of pro-B cells as CD34^{+}cTdT^{+}CD19^{-} cells.

Supplemental Figure 6



Supplemental Figure 6. Interference of tagged IKZF1N159S/T mutants with pericentromeric targeting of WT IKZF1
In upper panel, NIH3T3 cells were transfected with HA-tagged WT or mutant IKZF1 (R162Q, N159S, N159T) and were labeled with anti-HA antibody and an Alexa488-conjugated secondary antibody (green). Cells were visualized by fluorescence microscopy. In bottom panels, NIH3T3 cells were co-transfected with HA-tagged WT or mutant IKZF1 (green) plus a FLAG-tagged vector expressing WT IKZF1 and developed using an anti-Flag antibody followed by Alexa568-conjugated secondary antibody (red). Cells were visualized by fluorescence microscopy. Magnification X175. Data shown are representative of 3 experiments.

Supplemental Figure 7



Supplemental Figure 7. Inteference of IKZF1N159S/T mutants with pericentromeric targeting of WT IKZF3

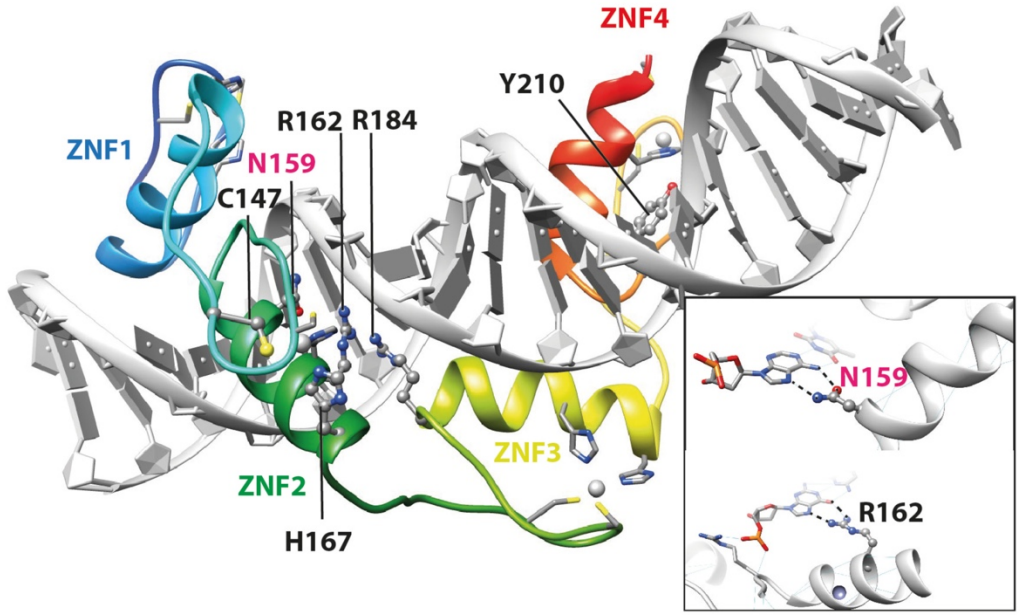
(A) Interference of IKZF1N159S/T mutants with pericentromeric targeting of WT IKZF3. NIH3T3 cells transfected with WT IKZF3 alone or co-transfected with WT or mutant IKZF1 (1:1 ratio). After 36 hours, cells were labeled with anti-IKZF3 mAb plus Alexa546-conjugated secondary antibody (red). Cells transfected with WT IKZF1 alone used as negative control for IKZF3 staining and cells transfected with WT IKZF3 alone used as negative control for IKZF1 staining. Cells transfected with WT or mutant IKZF1 and stained with anti-IKZF1 used as internal controls for IKZF1 transfection. Cells were visualized by fluorescence microscopy. Magnification X190. Data shown representative of 2 experiments.

(B) Foci counts from (A) experiment. IKZF3 foci were counted with the naked eye in 30-35 transfected 3T3 cells for each condition. Observer bias was eliminated by coding the slides prior to inspection. Statistical analysis was performed between the different groups using one-way ANOVA test. Horizontal lines represent the median ± 95% confidence interval (CI). **** indicates p value <0.0001.

(C) Interference of tagged IKZF1N159S/T mutants with pericentromeric targeting of WT IKZF3. NIH3T3 cells transfected with a FLAG-tagged WT IKZF3 alone or co-transfected with HA-tagged WT or mutant IKZF1. Cells were labeled with anti-IKZF3 mAb plus Alexa568-conjugated secondary antibody (red) and with anti-HA antibody plus Alexa488-conjugated secondary antibody (green). Cells were visualized by fluorescence microscopy. Magnification X175. Data shown representative of 3 experiments.

(D) Preserved interaction of tagged IKZF1N159S/T mutants with WT IKZF3. HEK/293T cells transfected with WT IKZF3-FLAG and WT or mutant IKZF1-HA. Immunoprecipitations (IP) were performed using anti-FLAG. Data represent western blot of whole cell lysates and IP samples (right panel) with anti-HA (upper panels) and anti-IKZF3 antibodies (lower panels). 5% of total protein lysates used for IP as control for total cell lysates to test protein expression (left panel). Data shown representative of 3 experiments.

Supplemental Figure 8



Supplemental Figure 8. Modeling structures of the four zinc fingers of the DNA binding domain of IKZF1
Model of the 3D structure of the IKZF1 ZF1-4 DNA-binding domains, built on the PRMD9ZF8-11, whose 3D structure has been solved in complex with a hot spot oligonucleotide (pdb 5egb) (17). The IKZF1 ZF arrays are predicted to follow the right handed twist of the DNA, with the helices occupying the major groove. Mutated amino acids are shown in a ball-and- stick representation. H167 is one of the four amino acids involved in ZF2 Zn 2+ coordination; the H167R mutation is thus likely to disturb the ZF2 fold. Y210 is also involved in the hydrophobic core of ZF4; the Y210C mutation thus also probably disturbs the fold of this ZF module, although cysteine residues are observed in this position in ZF1 and ZF3 (see Figure 1C corresponding to the ZF1-4 alignment). The three other amino acids (N159, R162, R184) are clustered in a unique region, in contact with DNA. The predicted interactions are shown in details at right, for the mutations reported in this study (N159, R162). N159 (at position -4 in ZF2) is predicted to make H-bonds with A, respectively. Mutations of these two amino acids (N159T, N159S) might affect the binding specificity of the ZF modules, rather than directly disrupting the interaction, as the substituted amino acids are also observed to interact with DNA. In contrast, other mutations (R162L, R162Q (at position -1 in ZF2) and R184Q (at position -7 in ZF3)) are predicted to interfere with DNA binding.