

Integrated genomic analysis identifies recurrent mutations and evolution patterns driving the initiation and progression of follicular lymphoma

Jessica Okosun^{1,14}, Csaba Bödör^{1,2,14}, Jun Wang^{3,14}, Shamzah Araf¹, Cheng-Yuan Yang⁴, Chenyi Pan^{5,6}, Sören Boller⁴, Davide Cittaro⁷, Monika Bozek⁸, Sameena Iqbal¹, Janet Matthews¹, David Wrench¹, Jacek Marzec³, Kiran Tawana¹, Nikolay Popov¹, Ciaran O’Riain¹, Derville O’Shea¹, Emanuela Carlotti¹, Andrew Davies⁹, Charles H. Lawrie¹⁰, Andras Matolcsy², Maria Calaminici¹, Andrew Norton¹¹, Richard J. Byers¹², Charles Mein⁸, Elia Stupka⁷, T. Andrew Lister¹, Georg Lenz¹³, Silvia Montoto¹, John G. Gribben¹, Yuhong Fan^{5,6}, Rudolf Grosschedl⁴, Claude Chelala³ & Jude Fitzgibbon¹

¹Centre for Haemato-Oncology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom.

²1st Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary.

³Centre for Molecular Oncology, Barts Cancer Institute, Queen Mary University of London, London, United Kingdom.

⁴Department of Cellular and Molecular Immunology, Max Planck Institute of Immunobiology and Epigenetics, Freiburg, Germany.

⁵School of Biology, Georgia Institute of Technology, Atlanta, USA.

⁶Parker H. Petit Institute for Bioengineering and Biosciences, Georgia Institute of Technology, Atlanta, USA.

⁷Centre for Translational Genomics and Bioinformatics, San Raffaele Scientific Institute, Milano, Italy.

⁸Genome Centre, Barts and the London School of Medicine and Dentistry, London, United Kingdom.

⁹Cancer Sciences Division, University of Southampton, Southampton, United Kingdom.

¹⁰Oncology Department, Biodonostia Research Institute, San Sebastian, Spain.

¹¹Department of Histopathology, The Christie NHS Foundation Trust, Manchester, United Kingdom.

¹²Department of Histopathology, Manchester Royal Infirmary, Manchester, United Kingdom.

¹³Department of Hematology, Oncology and Tumor Immunology, Charité Universitätsmedizin, 13353 Berlin, Germany.

¹⁴These authors contributed equally to this work.

Correspondence should be addressed to J.O. (j.e.okosun@qmul.ac.uk) or C.B. (bodor.csaba1@med.semmelweis-univ.hu)

Supplementary Note

High density SNP arrays. DNA samples (26 paired FL-tFL and 3 diagnostic FL DNA samples, including 7 of our discovery cases) were hybridized to the Affymetrix SNP 6.0 microarrays according to the manufacturer's protocols (see URLs). Copy number analysis was performed using the PennCNV-Affy⁵⁹ workflow. First, the apt-probeset-genotype program in Affymetrix Power Tools (APT) was used to generate genotyping calls from the raw CEL files using the Birdseed v2 algorithm. The APT software apt-probeset-summarize program was then used to extract allele-specific signal values for each SNP, after applying quantile normalization and median polish on the data. The Log R Ratio (LRR) values and the B-allele frequencies (BAF) values were subsequently calculated using the default clustering file "gw6.genocluster" provided in the PennCNA-afly package. For the tumor samples with paired GL samples, the LRR values were further adjusted relative to those of matching GL. For the detection of CNAs for each tumor sample, we used the combination of PennCNA (the "agre.hmm" file provided by the package were used) and genoCN⁶⁰ algorithms. The overlap calls between the two algorithms were selected, and any calls supported by less than 10 SNPs, or overlapping with immunoglobulin and T-cell receptor loci, centromeric and telomeric regions were excluded. The final individual CNA calls were used as the input for the frequency plots of amplifications and deletions in a 50-kb window along the genome.

The vegaMC⁶¹ tool was applied to determine significantly recurrent CNAs. To highlight the CNA regions common and specific to both the FL or tFL group, we performed the vegaMC test to each group (FL vs tFL) separately. A significance of $p < 0.05$, segment size larger than 5-kb and frequency of > 4 out of 31 cases (approx 12.9%) was selected as thresholds for the analysis and CNA plots. Any regions identified to be significant for both amplifications and deletions were excluded.

To determine regions of copy neutral loss of heterozygosity (cnLOH) also known as aUPD, the ASCAT algorithm (version 2.1) was used, taking into account tumor ploidy and purity. The paired and unpaired analyses were performed accordingly. For samples with ASCAT ploidy $< 2.7n$, cnLOHs were identified when the corrected integral allele-specific copy number was estimated to be 2 for allele A and 0 for allele B. In a fraction of samples with ASCAT ploidy $\geq 2.7n$, all cnLOHs appeared to have the corrected allele-specific copy number to be 3 or 4 for allele A but zero for allele B.

References:

59. Wang, K. *et al.* PennCNV: an integrated hidden Markov model designed for high-resolution copy number variation detection in whole-genome SNP genotyping data. *Genome Res* **17**, 1665-74 (2007).

60. Sun, W. et al. Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic Acids Res* **37**, 5365-77 (2009)
61. Morganella, S. & Ceccarelli, M. VegaMC: a R/bioconductor package for fast downstream analysis of large array comparative genomic hybridization datasets. *Bioinformatics* **28**, 2512-4 (2012).

List of Supplementary Figures and Tables

Supplementary Figure 1: Summary of study workflow

Supplementary Figure 2: Base substitution patterns

Supplementary Figure 3: Summary of non-synonymous mutations

Supplementary Figure 4: Copy number status of 6 WGS cases

Supplementary Figure 5: Phylogenetic trees based on non-synonymous mutations

Supplementary Figure 6: Phylogenetic trees based on all variants in the 6 WGS cases

Supplementary Figure 7: Mutation spectrum in 32 paired FL-tFL cases

Supplementary Figure 8: Sequence alignment of linker histones

Supplementary Figure 9: *HIST1H1C* and WT ESC lines

Supplementary Figure 10: Schematics of the domains and location of mutations

Supplementary Figure 11: Integrative analysis of genetic events

Supplementary Figure 12: Evolution of genetic events during disease course

Supplementary Figure 13: Copy number alterations from SNP6.0 arrays

Supplementary Figure 14: Comparison of variants detected by the various sequencing approaches

Supplementary Table 1: Clinical features and treatment timelines

Supplementary Table 2: Coverage data for WGS and WES

Supplementary Table 3: Summary of somatic variants identified in the 10 discovery cases

Supplementary Table 4: List of all non-synonymous mutations in 10 discovery cases

Supplementary Table 5: Validated non-synonymous (including indels) mutations

Supplementary Table 6: Structural variants and promoters

Supplementary Table 7: Targeted resequencing coverage

Supplementary Table 8: Sample details

Supplementary Table 9: Somatic variants from targeted resequencing

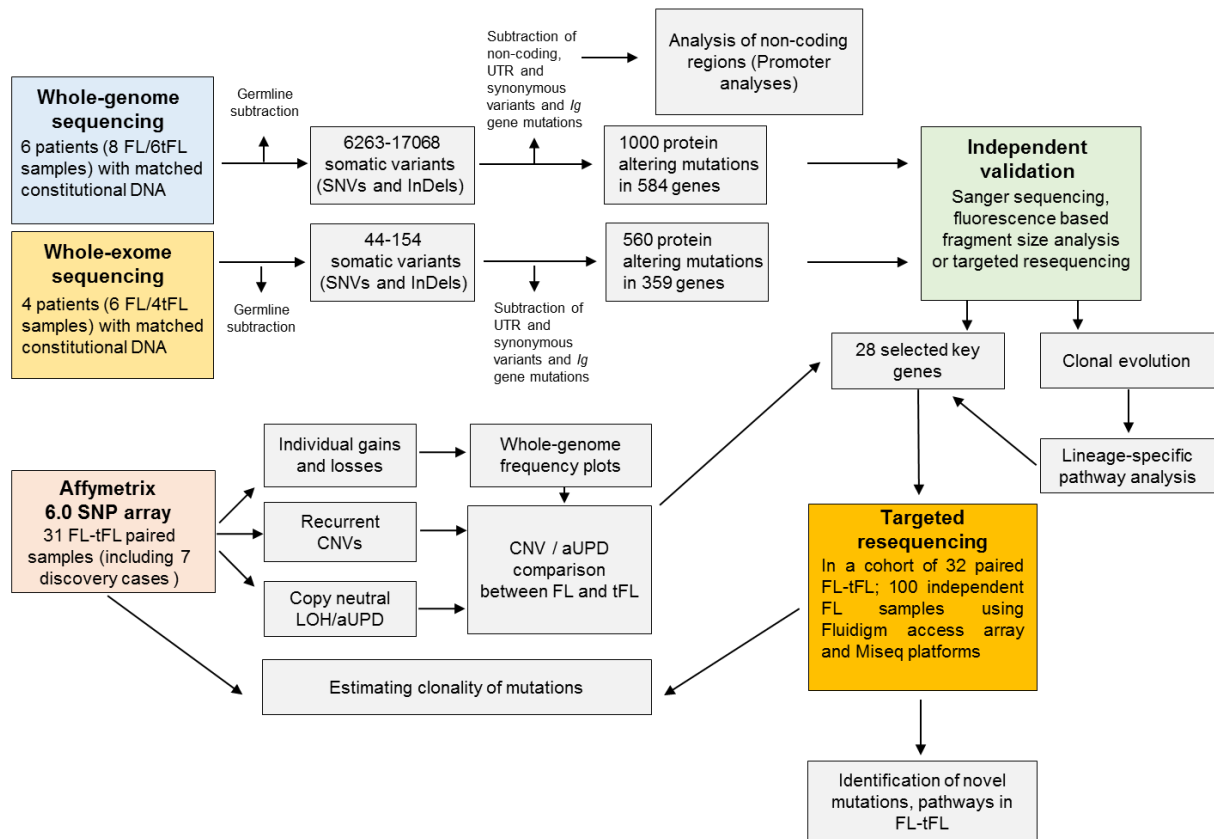
Supplementary Table 10: Copy number alterations from SNP6.0 arrays

Supplementary Table 11: Purity and ploidy values

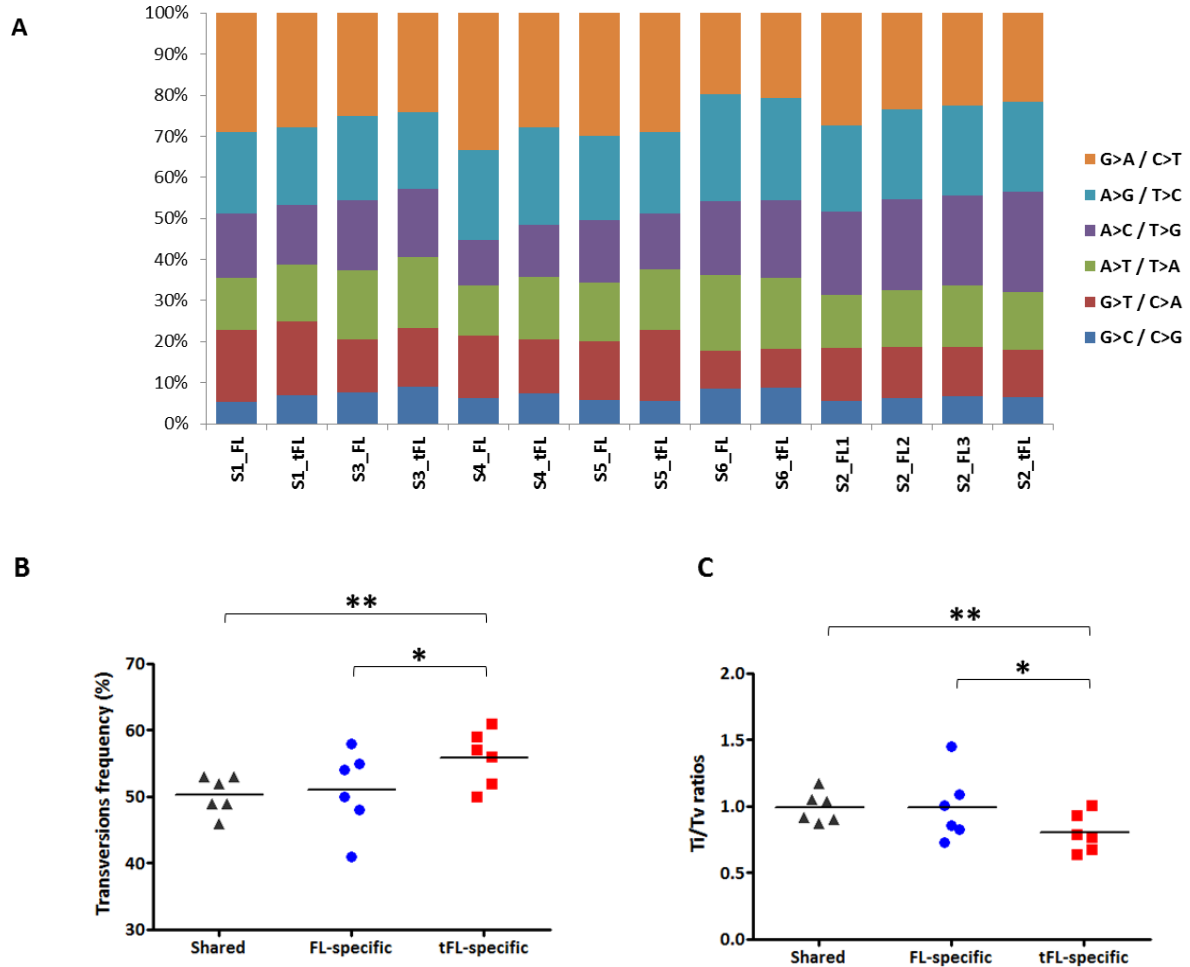
Supplementary Table 12: SNVs analyzed by ABSOLUTE

Supplementary Table 13: Aberrant somatic hypermutation patterns

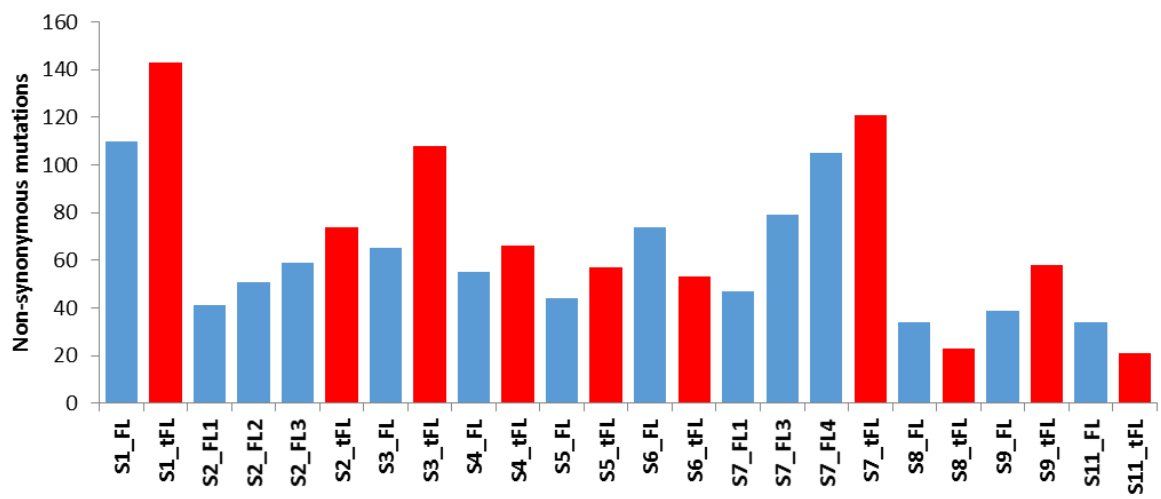
OVERVIEW OF THE WORKFLOW



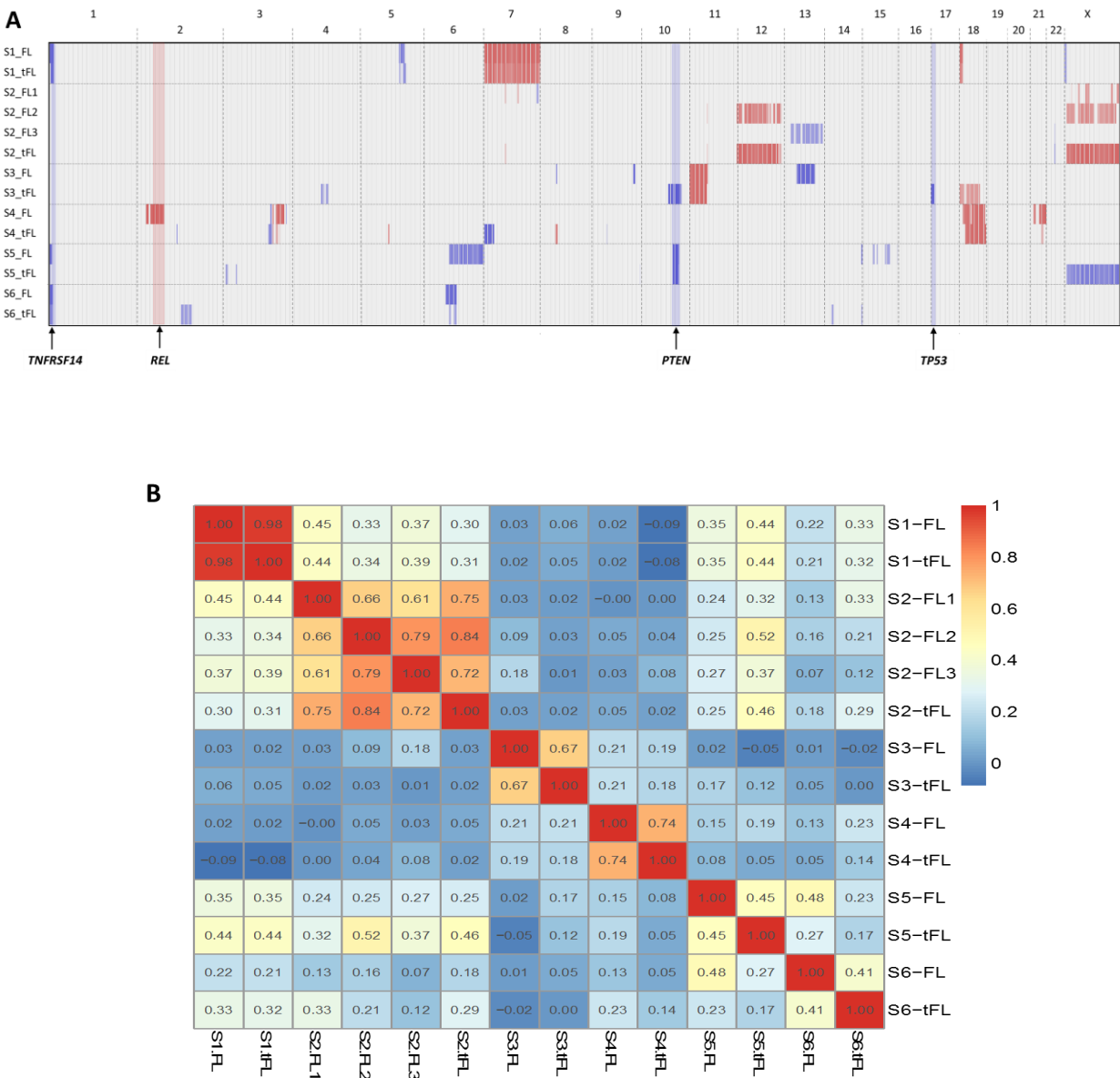
Supplementary Figure 1. Summary of study workflow. The schema shows the main analytical approaches used in this study: whole genome or exome sequencing; copy number profiling using Affymetrix SNP 6.0 arrays and targeted resequencing of candidate genes for prevalence screening.



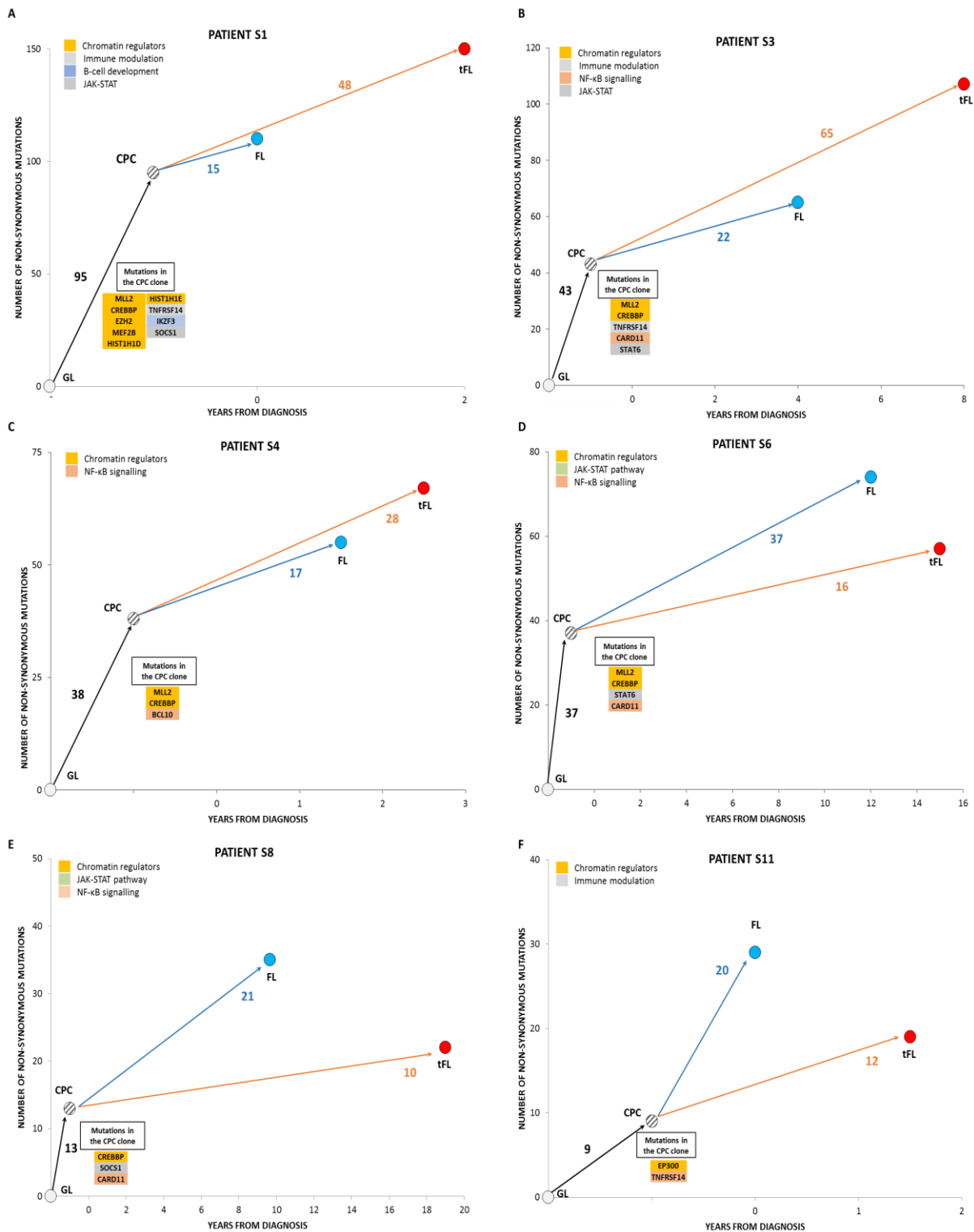
Supplementary Figure 2. Base substitution patterns. (a) Distribution of base substitutions considering all somatic variants in each of the 6 WGS cases. (b) Comparison of transversion frequencies in shared, FL-specific and tFL-specific variants (paired t-test; $P = * 0.016$; $** 0.0055$). (c) Transition/Transversion (Ti/Tv) ratios ($P = * 0.028$; $** 0.008$).



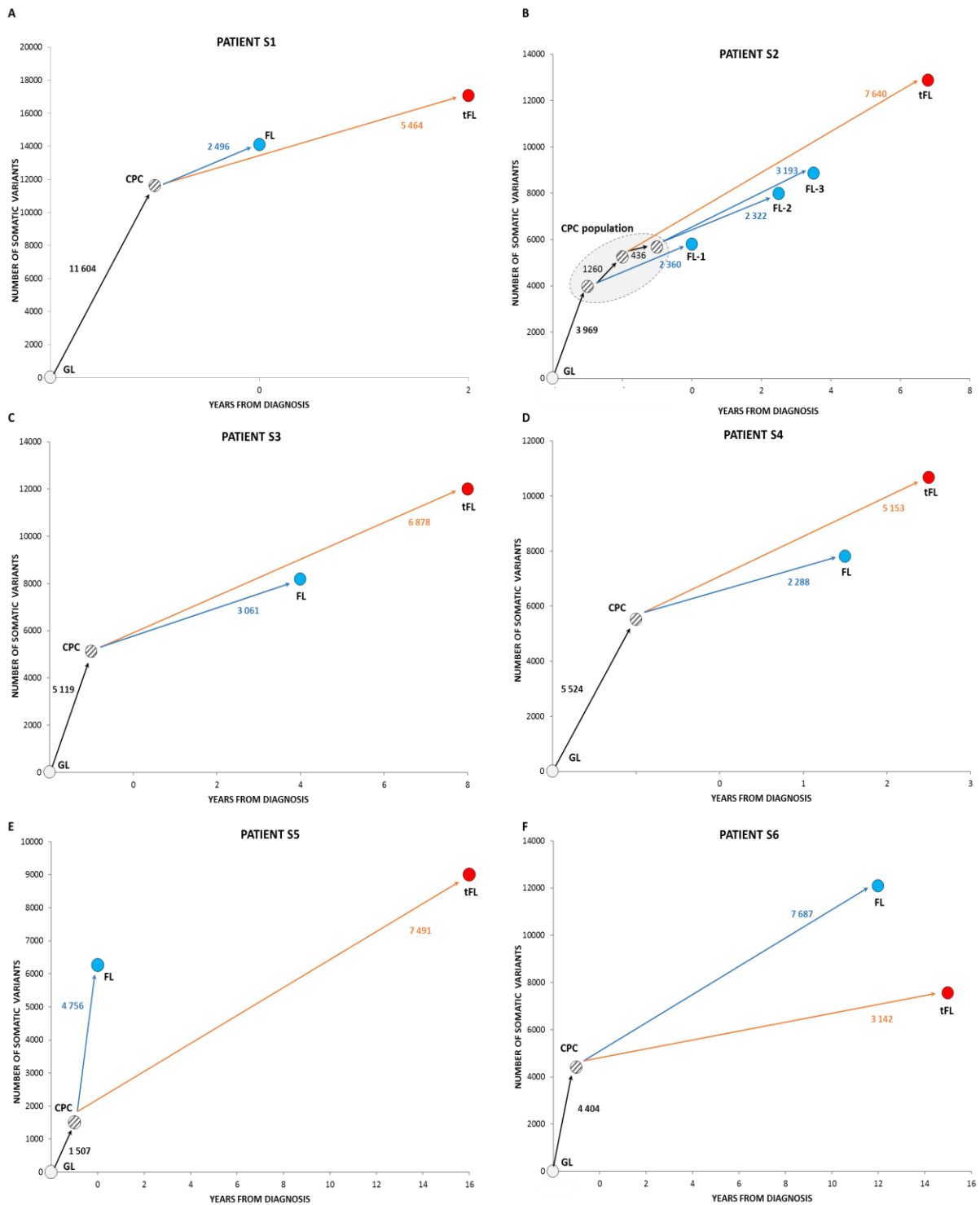
Supplementary Figure 3. Summary of non-synonymous mutations for the 10 discovery cases.



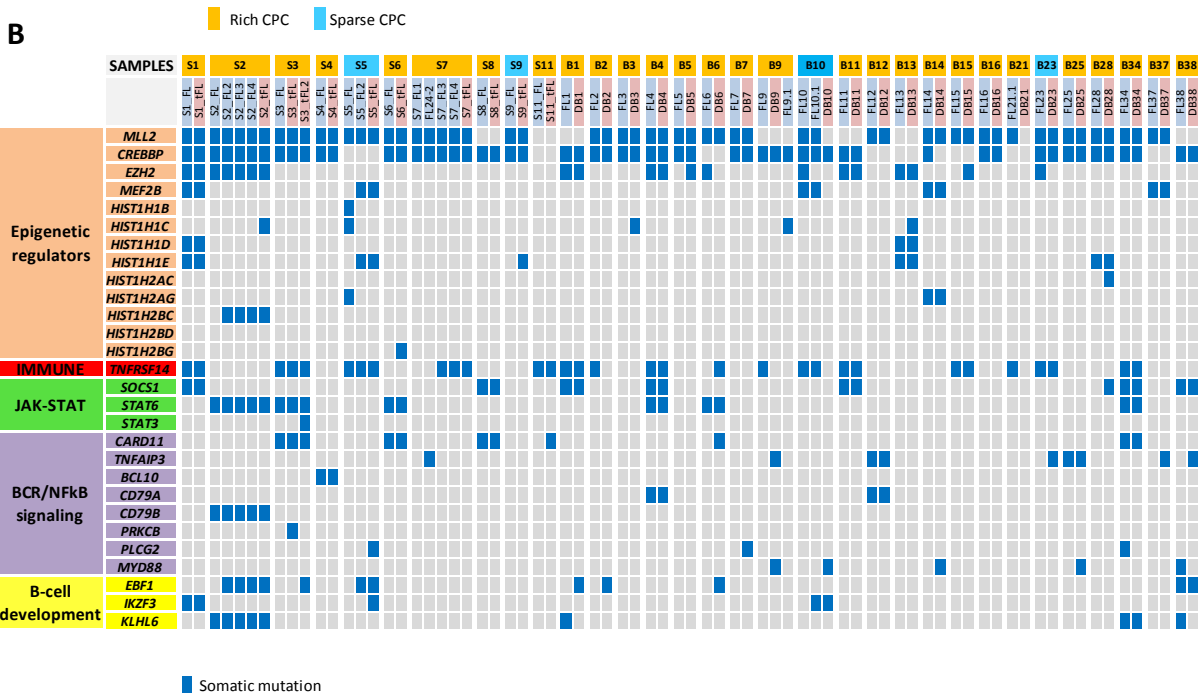
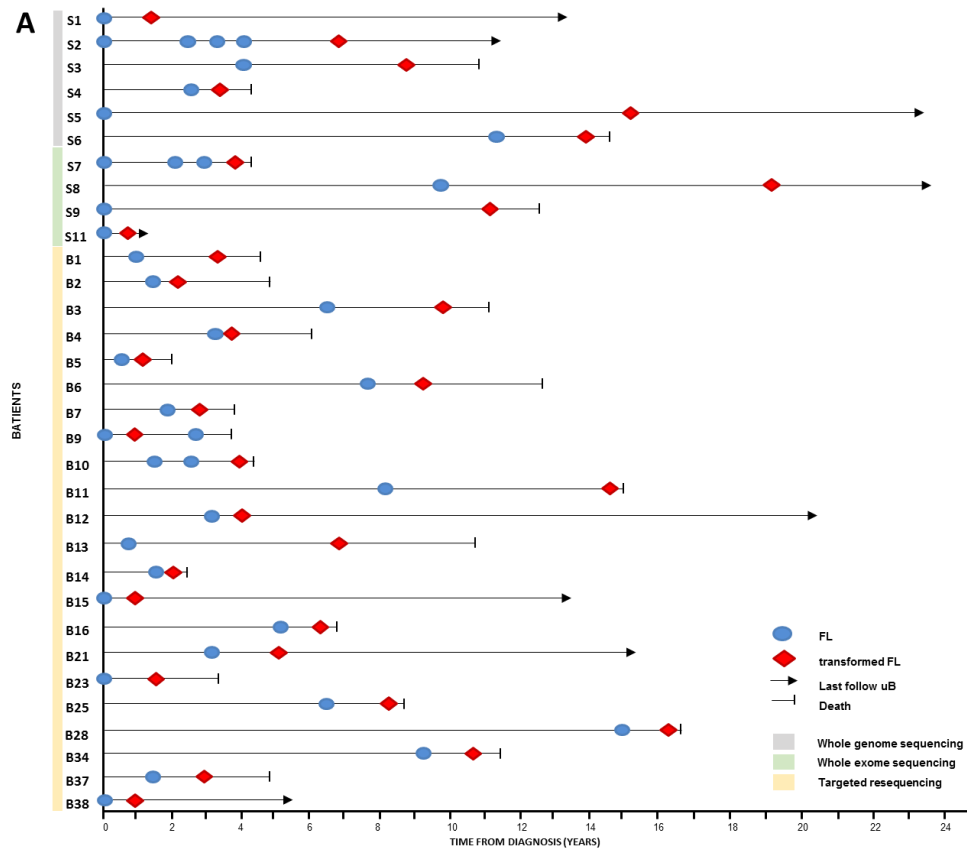
Supplementary Figure 4. Copy number status of 6 WGS cases. (a) Genome-wide copy number alterations in the 6 WGS cases indicating specific gene deletions (*TNFRSF14*, *PTEN* and *TP53*) and amplifications (*REL*) that have been reported previously. (b) Correlation between copy number alterations of individual samples. Cases S1-S4 show a high correlation between their FL and tFL biopsies in comparison to case S5 and S6.



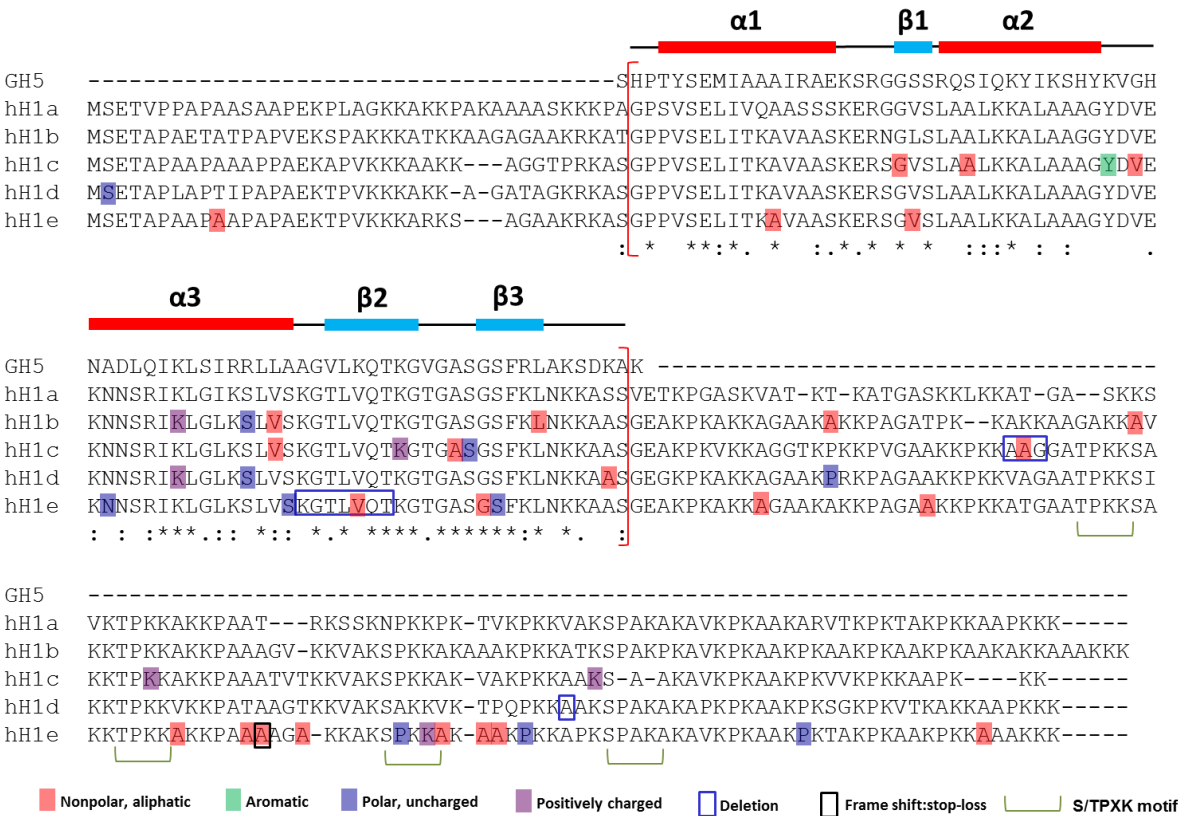
Supplementary Figure 5. Phylogenetic trees based on non-synonymous mutations. (a) S1. (b) S3. (c) S4. (d) S6. (e) S8. (f) S11. All trees depict a divergent evolution pattern with early clonal gene mutations occurring within the 'trunk' – the putative CPC.



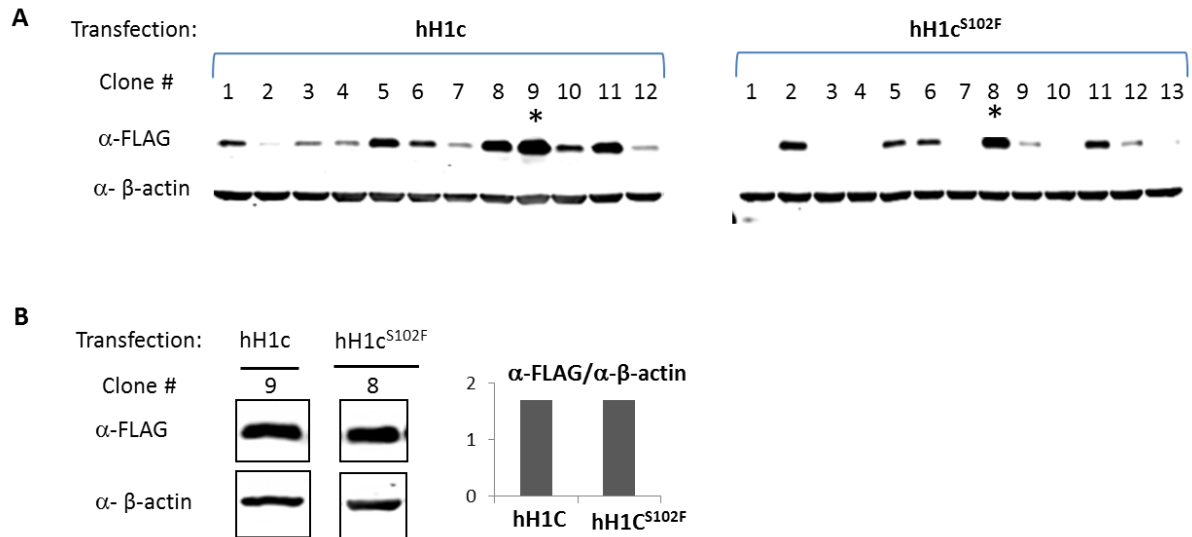
Supplementary Figure 6. Phylogenetic trees based on all somatic variants. (a) S1. (b) S2. (c) S3. (d) S4. (e) S5. (f) S6. All cases with the exception of S5 have evolved from a 'rich' CPC.



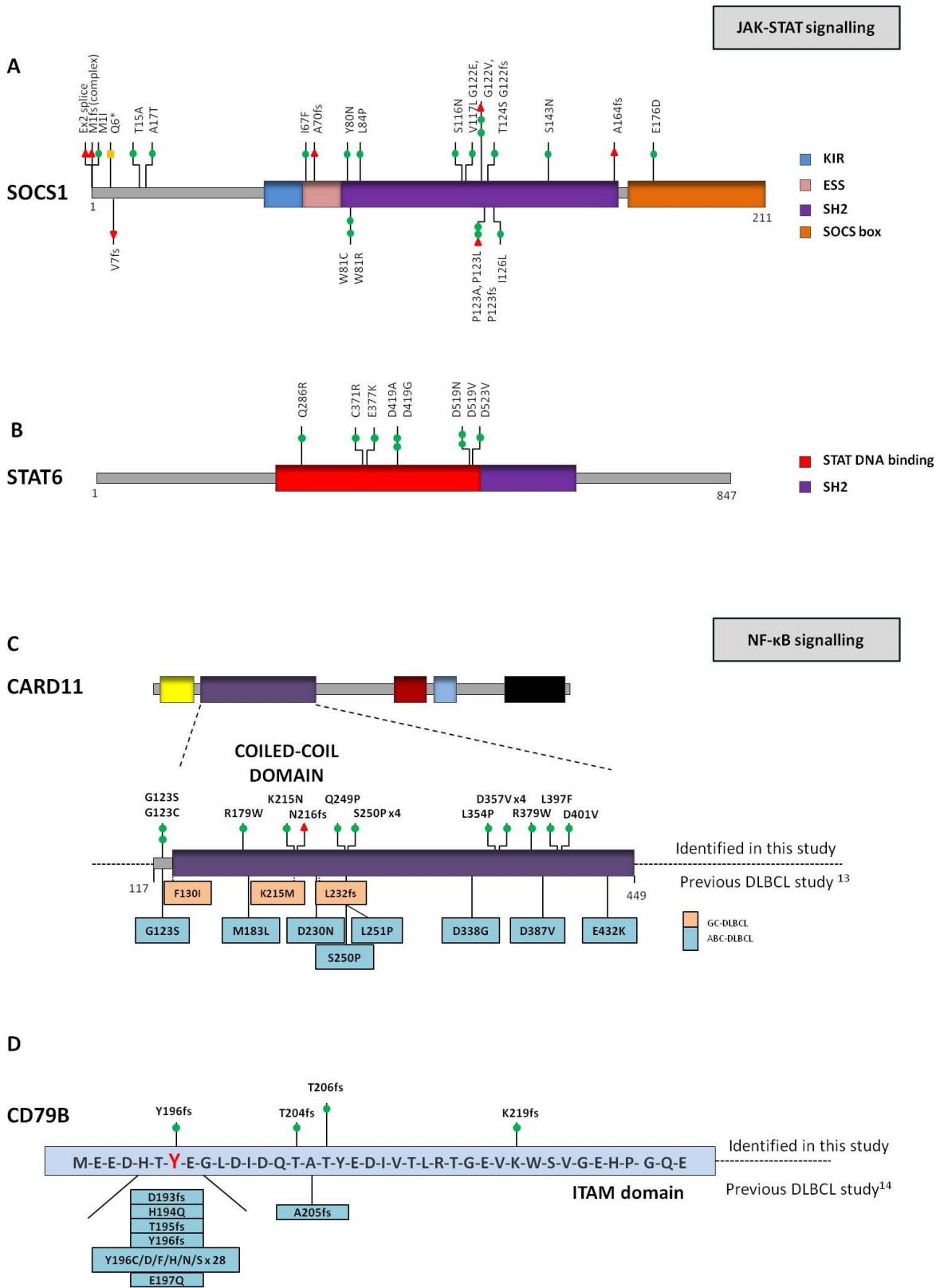
Supplementary Figure 7. Mutation spectrum in 32 paired FL-tFL cases. (a) Timelines illustrating the intervals from the FL biopsies screened (blue) to the tFL biopsies (pink) in each of the 32 cases. (b) Heatmap of the mutation spectrum of the selected 28 genes. Based on the pattern of mutations, the majority of cases evolve through a ‘rich’ CPC with only 4 cases adopting a ‘sparse’ pattern of evolution (S5, S9, B10 and B23) (also see Supplementary Fig. 12b).

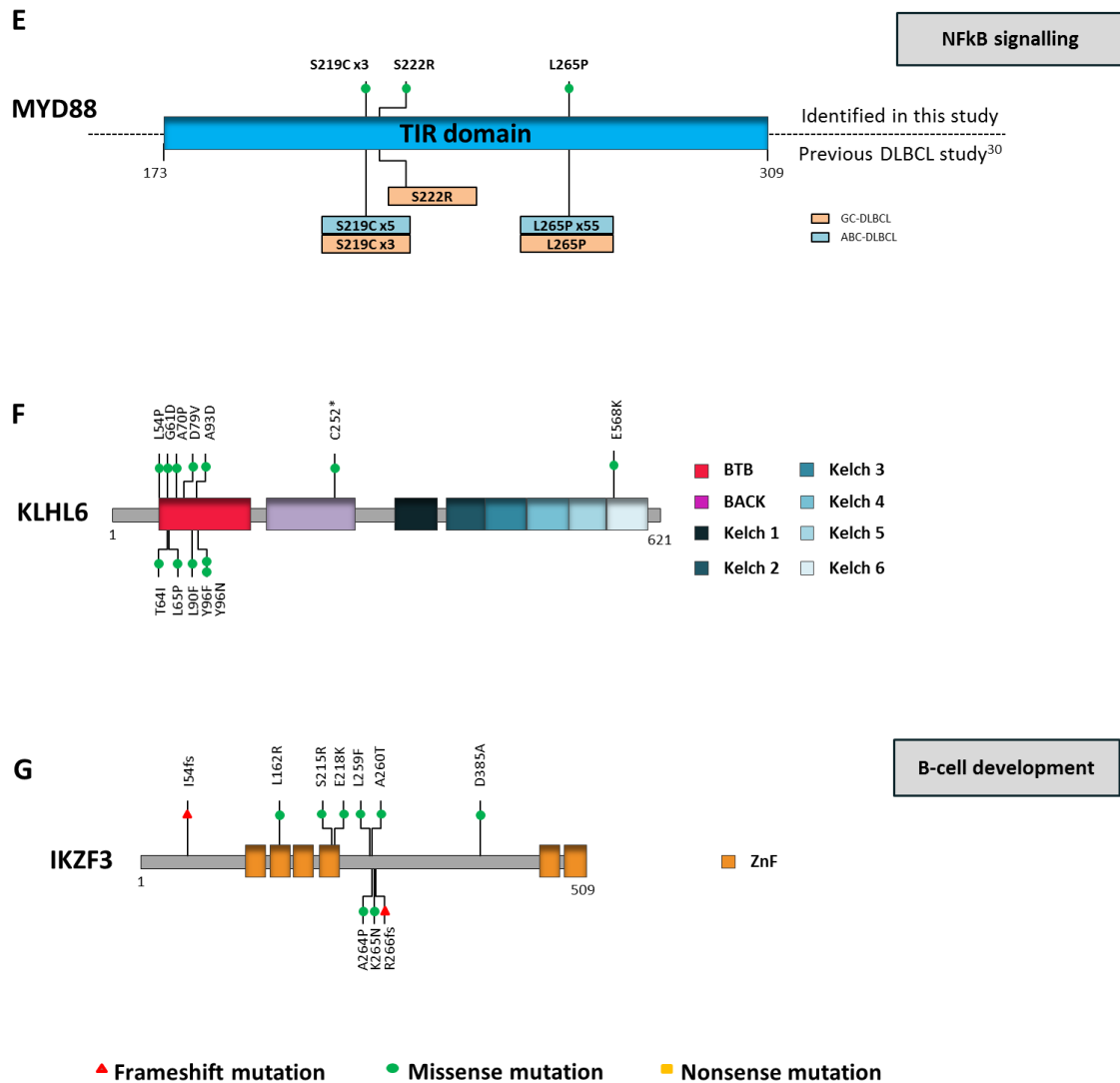


Supplementary Figure 8. hH1a –H1e sequence alignment and the distribution of mutations. GH5: globular domain of chicken histone H5. H1 globular domain is marked with the red bracket.

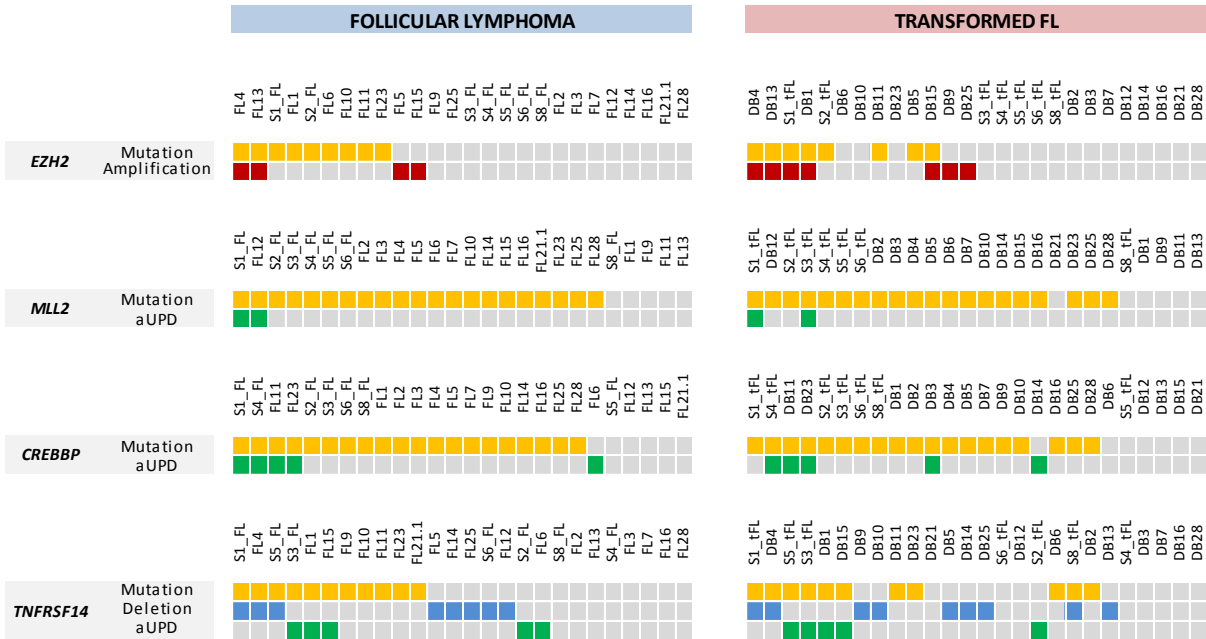


Supplementary Figure 9. *HIST1H1C* and WT ESC lines. (a) Expression of FLAG tagged human H1c (hH1c) or hH1c^{S102F} mutant in H1c/H1d/H1e triple null embryonic stem cells (H1 TKO ESCs). H1 TKO ESCs were transfected with vectors expressing FLAG-hH1c or FLAG-hH1c/S102F mutant. Twenty-four stable ESC clones were picked for each transfection and screened using an anti-FLAG antibody. Representative immunoblots of FLAG-hH1c or FLAG-hH1c/S102F over-expressing cell clones are shown. Immunoblots with anti-b-ACTIN antibody were included as loading controls. Clones indicated with asterisks, demonstrating high expression levels of hH1c or hH1c/S102F were used in subsequent analysis. (b) FLAG-hH1c and FLAG-hH1c^{S102F} are expressed at the same levels in selected clones shown.

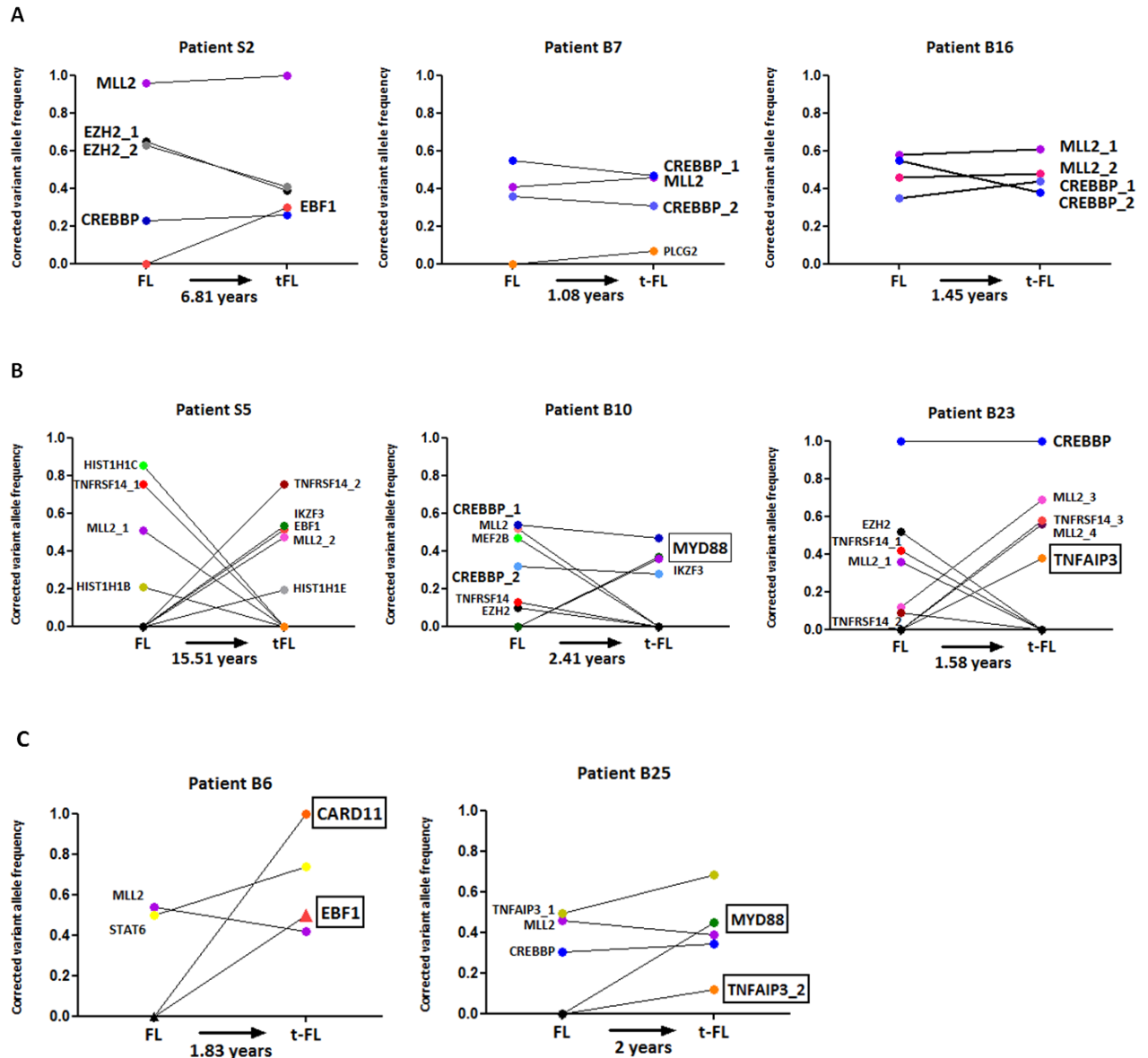




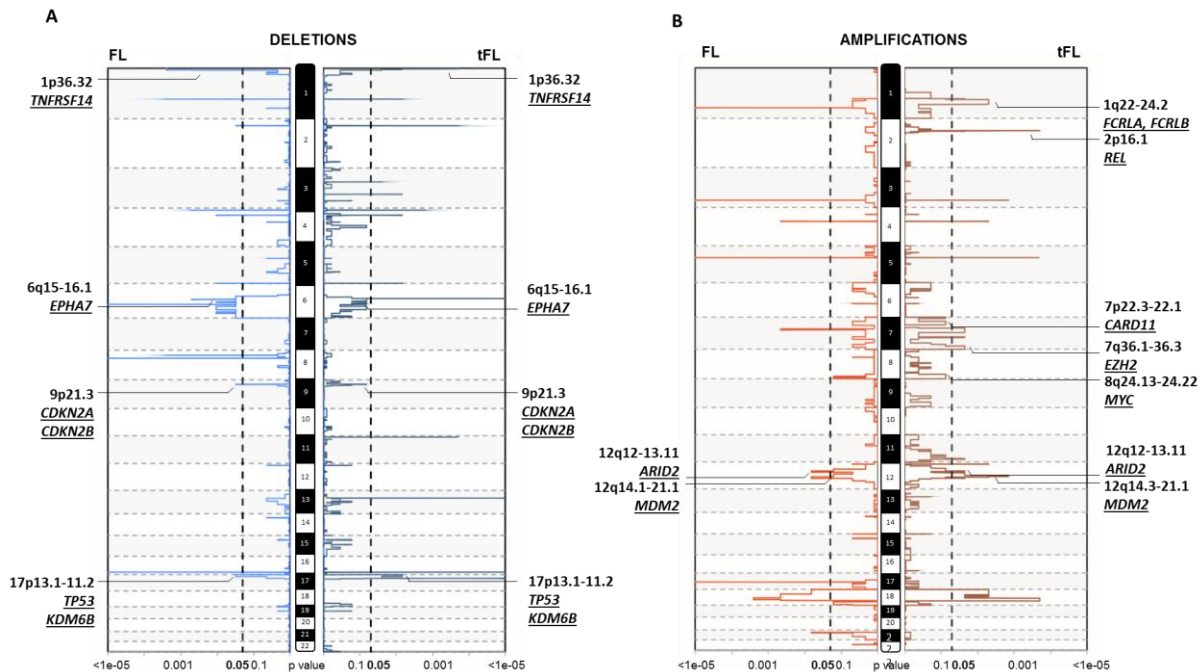
Supplementary Figure 10. Schematics of the domains and location of mutations. (a) SOCS1. (b) STAT6. (c) CARD11. (d) CD79B. (e) MYD88. (f) KLHL6 and (g) IKZF3. Multiple mutations at different amino acid residues within the same sample are depicted. Plots for genes involved in NF-κB signaling (*CARD11*, *CD79B* and *MYD88*) are compared to previously published sites of mutations in ABC and GC DLBCL. KIR, kinase inhibitory region; SH2, Src homology 2; ESS, extended SH2 subdomain; SOCS box, suppressors of cytokine signaling box; ITAM, immunoreceptor tyrosine-based activation motif; TIR, toll/interleukin-1 receptor; BTB, broad-complex, tramtrack and bric a brac; BACK, BTB And C-terminal Kelch; ZnF, zinc finger.



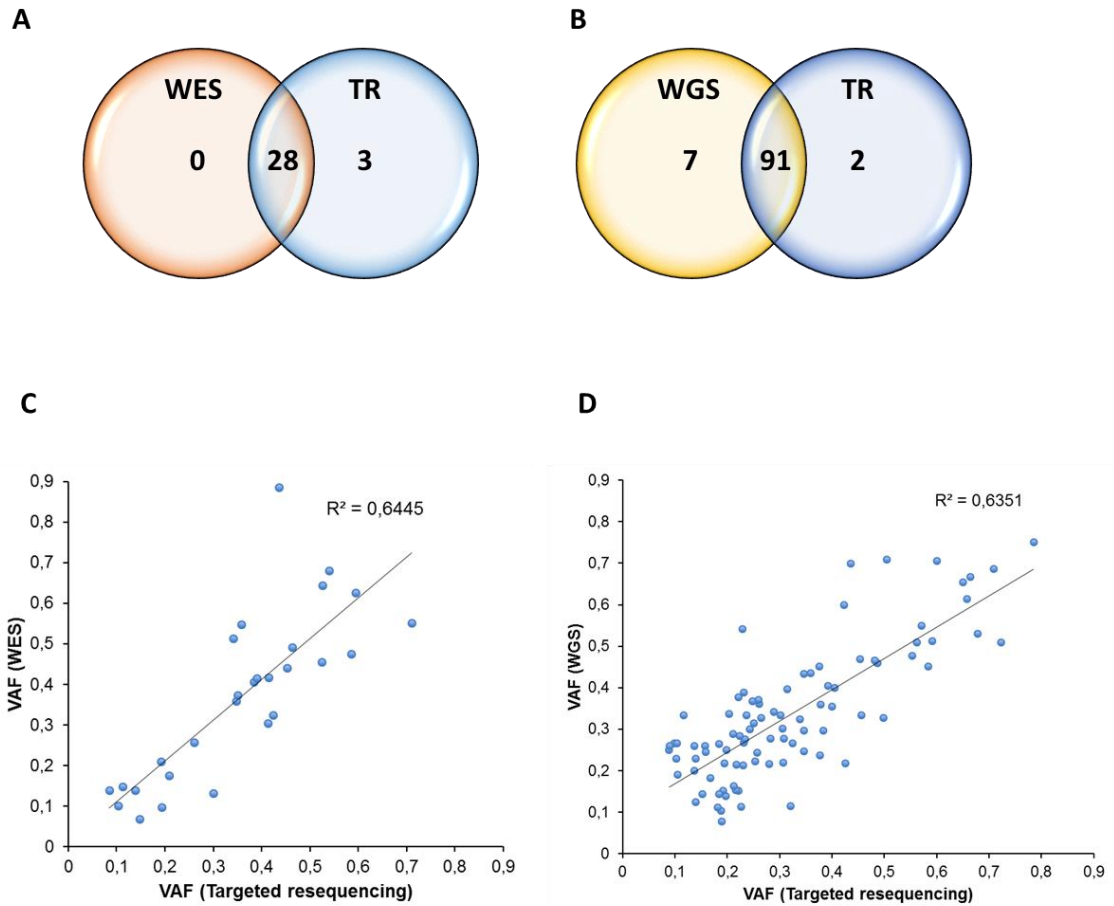
Supplementary Figure 11. Integrative analysis of genetic events based on select genes in 26 paired FL-tFL where all information was available. Mutations, CNA (amplifications), CNA (deletions) and aUPD events are colored yellow, red, blue and green, respectively. Mutation events (SNVs and InDels) were identified by whole genome sequencing and targeted resequencing, while CNAs and aUPDs were identified using Affymetrix 6.0 SNP arrays.



Supplementary Figure 12. Evolution of genetic events during the disease course. Graphical representation of corrected variant allele frequencies (VAFs) of specific genes across dual time points (FL and tFL) based on read counts obtained from deep sequencing. (a) Three cases demonstrating concomitant mutations in chromatin regulators (*MLL2*, *CREBBP* and/or *EZH2*) showing relatively unchanged clone sizes as the disease progressed from FL to tFL. (b) Three cases illustrating evolution from a ‘sparse’ precursor clone over varying intervals. In patient B10, two different *CREBBP* mutations are dominant during the disease course whereas other mutations are gained (*MYD88* and *IKZF3*) or lost (*MLL2*, *MEF2B*, *TNFRSF14*, *EZH2*). Likewise, patient B23 represents a ‘sparse’ CPC evolutionary pattern with clone harboring a *CREBBP* mutation. (c) Further examples depicting the acquisition of mutations in *EBF1* and genes regulating of NF-κB signaling (*CARD11*, *MYD88* and *TNFAIP3*). The annotations after specific genes for example *EZH2_1* and *EZH2_2* indicate different gene mutations within the individual case.



Supplementary Figure 13. Recurrent copy number changes in paired FL-tFL. (a) Copy number loss (deletion) plots displaying significantly altered regions by chromosomal position in FL (left panel) and tFL (right panel). Significance thresholds are denoted by vertical dash line (p -value < 0.05 computed from the VegaMC analysis). (b) Amplification plots. The most frequent significant peaks of interest are labeled with the candidate genes mapped from UCSC Genome Browser (GRCh37/hg19).



Supplementary Figure 14. Comparison of variants detected by the various sequencing approaches. High concordance between variants detected by Fluidigm-MiSeq based targeted resequencing (TR) versus (a) whole exome sequencing (WES); and (b) whole genome sequencing (WGS). Correlation of VAF estimates detected by TR compared to (c) WES; and (d) WGS.

Supplementary Table 1: Clinical features and treatment timelines for the ten discovery cases analysed by whole genome and exome sequencing.

Patient ID	Age at Diagnosis (Years)	Sex	Tumor Site	t(14;18)	Stage at Diagnosis	FLIPI at Diagnosis	Treatment
GENOME SEQUENCING CASES							
S1	48	M	LN	Y	3	2	Diagnosis (Nov 98): CHOP FL-1 (Jul 99): Cytarabine + Etoposide, followed by BEAM autograft Transformation (Jun 00): Bexxar
S2	57	M	LN	Y	4	2	Diagnosis (Apr 01): Chl, then CHOP FL-1 (Dec 03): Expectant management FL-2 (Oct 04): Rituximab FL-3 (Feb 07): Velcade + rituximab Transformation (Jan 08): R-CHOP followed by RIC-SCT
S3	48	F	LN	Y	3	Unknown	Diagnosis (Jan 94): Chl + Prednisolone FL-1 (Nov 95): Chl FL-2 (Feb 97): Rituximab, then FMD FL-3 (Nov 98): Bexxar FL-4 (Nov 99): CHOP + RT Transformation (Nov 01): Cytarabine + Etoposide, followed by BEAM autograft FL-5 (Nov 02): Expectant management FL-6 (May 03): Velcade Transformation (Apr 04): No therapy
S4	45	F	LN	Y	1	Unknown	Diagnosis (Jul 99): RT FL-1 (Nov 00): Chl Transformation (Nov 02): CHOP, then LD-56 (trial), then Rituximab
S5	48	M	LN	Y	1	Unknown	Diagnosis (Jun 88): RT FL-1 (Mar 92): Surgical excision FL-2 (Aug 00): Expectant management Transformation (Mar 04): R-CHOP
S6	44	F	LN	Y	Unknown	Unknown	Diagnosis (Dec 93): Chl + Prednisolone FL-1 (Oct 99): Chl + Prednisolone FL-2 (Feb 01): Chl + Prednisolone FL-3 (Jan 02): Rituximab FL-4 (Aug 03): Chl + Prednisolone, then R-CHOP FL-5 (May 05): Velcade + Rituximab FL-6 (Feb 07): Expectant management Transformation (Nov 07): Cyclophosphamide, then R-CHOP
EXOME SEQUENCING CASES							
S7	40	M	LN	Y	4	Unknown	Diagnosis (May 96): FMD FL-1 (Jul 97): Rituximab + BEAM autograft FL-2 (Jul 98): Chl + Prednisolone, then CHOP

SUPPLEMENTARY INFORMATION

							FL-3 (Jul 99): Bexxar FL-4 (Jan 00): IEP Transformation (Jun 00): Methotrexate
S8	52	F	LN	Y	4	Unknown	Diagnosis (Nov 88): Chlorambucil FL-1 (Nov 98): Expectant management Transformation (Nov 07): R-CHOP tFL-1 (May 11): RT
S9	44	M	LN	Y	4	Unknown	Diagnosis (May 91): Chl + IFN FL-1 (Jan 95): Chl Transformation (Oct 02): CHOP FL-2 (Apr 03): Idarubicin + Cytarabine FL-3 (Oct 03): Palliative Transformation (Jun 10): R-ICE then palliative
S11	29	M	LN	Y	4	2	Diagnosis (Jul 11): Expectant management Transformation (Feb 12): R-CHOP

LN: lymph node; FL-1: first relapse; CHOP: cyclophosphamide, doxorubicin, vincristine, prednisolone; R: rituximab; BEAM: carmustine, etoposide, cytarabine, melphalan; Chl: chlorambucil; FMD: fludarabine, mitoxantrone, dexamethasone; RIC-SCT: reduced intensity conditioning stem cell transplantation; RT: radiotherapy; IEP: ifosfamide, etoposide, prednisolone; IFN: interferon α ; ICE: ifosfamide, cisplatin, etoposide