

A high definition picture of key genes and pathways mutated in pediatric follicular lymphoma

Federica Lovisa,^{1,2} Andrea Binatti,³ Alessandro Coppe,^{1,3} Simona Primerano,^{1,2} Elisa Carraro,¹ Marta Pillon,¹ Marco Pizzi,⁴ Vincenza Guzzardo,⁴ Salvatore Buffardi,⁵ Fulvio Porta,⁶ Piero Farruggia,⁷ Raffaella De Santis,⁸ Pietro Bulian,⁹ Giuseppe Basso,^{1,2} Elena Lazzari,¹⁰ Emanuele S.G. d'Amore,¹⁰ Stefania Bortoluzzi³ and Lara Mussolin^{1,2}

FL and AB contributed equally to this work

LM and SB are co-senior authors

¹Clinic of Pediatric Onco-Hematology, Department of Women's and Children's Health, University of Padova, Padova; ²Istituto di Ricerca Pediatrica Città della Speranza, Padova; ³Department of Molecular Medicine, University of Padova, Padova; ⁴Surgical Pathology and Cytopathology Unit, Department of Medicine, University of Padova, Padova; ⁵Santobono-Pausilipon Hospital, Napoli; ⁶Children's Hospital, Spedali Civili, Brescia; ⁷A.R.N.A.S. Ospedali Civico Di Cristina Benfratelli, Palermo; ⁸IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia; ⁹Clinical and Experimental Onco-Hematology Unit, IRCCS Centro di Riferimento Oncologico, Aviano and ¹⁰Department of Pathological Anatomy, San Bortolo Hospital, Vicenza, Italy

Correspondence: LARA MUSSOLIN - lara.mussolin@unipd.it
doi:10.3324/haematol.2018.211631

A high definition picture of key genes and pathways mutated in pediatric follicular lymphoma

Federica Lovisa^{1,2}, Andrea Binatti³, Alessandro Coppe^{1,3}, Simona Primerano^{1,2}, Elisa Carraro¹, Marta Pillon¹, Marco Pizzi⁴, Vincenza Guzzardo⁴, Salvatore Buffardi⁵, Fulvio Porta⁶, Piero Farruggia⁷, Raffaella De Santis⁸, Pietro Bulian⁹, Giuseppe Basso^{1,2}, Elena Lazzari¹⁰, Emanuele S.G. d'Amore¹⁰, Stefania Bortoluzzi³, Lara Mussolin^{1,2}

¹Clinic of Pediatric Onco-Hematology, Department of Women's and Children's Health, University of Padova, Italy

²Istituto di Ricerca Pediatrica Città della Speranza, Padova, Italy

³Department of Molecular Medicine, University of Padova, Padova, Italy

⁴Surgical Pathology and Cytopathology Unit, Department of Medicine, University of Padova, Italy

⁵Santobono-Pausilipon Hospital, Napoli, Italy

⁶Children's Hospital, Spedali Civili, Brescia, Italy

⁷A.R.N.A.S. Ospedali Civico Di Cristina Benfratelli, Palermo, Italy

⁸IRCCS Casa Sollievo della Sofferenza, San Giovanni Rotondo, Foggia, Italy

⁹Clinical and Experimental Onco-Hematology Unit, IRCCS Centro di Riferimento Oncologico, Aviano, Italy

¹⁰Department of Pathological Anatomy, San Bortolo Hospital, Vicenza, Italy

SUPPLEMENTARY METHODS AND RESULTS

Summary

Supplementary Methods	2
Table S1. Clinico-pathological and cytogenetic features of the study population.	5
Figure S1. Representative histological features of PTNFL.	6
Figure S2. Bioinformatic workflow for detection of high confident somatic SNPs and reconstruction of a mutated pathway-derived network.	7
Figure S3. Sequence coverage profile per patient.	8
Figure S4. Somatic variants detected.	9
Table S2. Details on selected somatic variants detected in 9 cases of FL of the pediatric age and on involved transcripts with impact scoring (H, high; M, moderate) and amino acid change.	10
Table S3. Sequences of PCR primers used for validation of selected variants detected by WES.	10
Figure S5. Validation of somatic variants obtained by Sanger sequencing.	11
Figure S6. Impact of somatic variants to protein products.	14
Figure S7. KEGG and Reactome pathways derived meta-network of genes somatically mutated in FL of the pediatric age.	17

Supplementary Methods

Study population

The study population included 21 patients with lymphoproliferative disorders belonging to the spectrum of follicular lymphoid neoplasms of the pediatric age (2008 WHO Classification of Tumors of the Haematopoietic and Lymphoid Tissues and its 2016 update) (**Table S1**). In compliance with the Helsinki Declaration, informed written consent was obtained from parents or legal guardians on behalf of the children enrolled in the study. The median age at diagnosis was 14.2 years (range: 4.0-17.7 years), with a male to female ratio of 10:1. Of them, 18 cases were originally diagnosed as pediatric-type nodal follicular lymphoma (PTNFL) and 3 cases as primary testicular follicular lymphoma (PFLT). The immunohistochemical profile disclosed positivity for pan-B cell markers, Bcl6 and CD10, with negative (19/21 cases) or weak positive (2/21 cases) Bcl2 in all the cases. The mean proliferation index (Ki-67 immunostaining) was greater than 40%. MUM1 immunostaining was negative in all tested cases (18/18). A representative PTNFL case showing partial architectural effacement by large follicles is reported in **Figure S1**. B-cell monoclonality was confirmed by polymerase chain reaction of the immunoglobulin genes, according to BIOMED-2 guidelines (van Dongen *et al.*, Leukemia 2003). As for the clinical management, 15/21 cases were treated with surgical excision alone, while 6/21 cases were also treated with 2 to 4 cycles of chemotherapy according to the AIEOP-LNH97 protocol. A single case experienced a second malignancy after 4 courses of chemotherapy. At present all patients are alive and in good clinical condition (mean follow-up: 41.7 months, range 8.3-124.8 months).

Whole Exome Sequencing

Purified genomic DNA from paired tumor samples and peripheral blood were outsourced for Whole Exome Sequencing (WES) analysis (Biodiversa srl, Rovereto (TN), Italy). Samples were enriched in protein coding sequences using the SureSelect Human All Exon V5 (Agilent Technologies, Santa Clara, CA), following the manufacturer's instructions. The resulting libraries were subjected to paired-end sequencing (2 x 150 bp) on an Illumina HiSeq4000 system, with a theoretical coverage of 150X for tumor samples and 100X for paired germline samples.

Somatic variants detection and prioritization

Reads were aligned to human genome version hg19 using BWA. Picard and GATK Genome Analysis Tool Kit were used to process the files generated during alignment to the reference genome and to remove reads mapping to multiple genomic positions and putative PCR duplicates in order to reduce the risk of false positive variant calling. *RealignerTargetCreator* and *IndelRealigner* commands from

GATK suite were used to perform a local realignment around indels considering the information from all reads mapping to the region. To improve reads base quality scores accuracy using all reads aligned to the considered region, base quality score recalibration was obtained using the *BaseRecalibrator* command from GATK. Somatic SNPs were detected by MuTect and then annotated by SnpEff and SnpSift. SnpEff predicted functional and putative impact of detected variants. SnpSift's *annotate* command provided the association of known variants to dbSNP (v. 150) and COSMIC (v. 82) identifiers, clinical significance from Clinvar (updated on 05/09/2017). Known variants annotated in Clinvar as benign or likely benign were discarded and only variants with SnpEff predicted impact *HIGH* or *MODERATE* were further considered. Finally, variants with non-Finnish European population allele frequency >5% according to ExAC (Exome Aggregation Consortium; Cambridge, MA, <http://exac.broadinstitute.org>) data were discarded.

Mutated gene network analysis

R Graphite Bioconductor package (v. 1.20.1) was used to map and convert pathway topologies annotated in KEGG and Reactome into pathway-derived gene networks, using appropriate biology-driven rules to transform different types of direct and indirect relations between genes and genes products annotated in pathways (i.e. regulatory relations, participation to molecular complexes and biosynthetic pathways, also with compound intermediates) into pairwise gene connections. Only pathways including at least one mutated gene were converted into networks and merged in order to build up a unique non-redundant KEGG-Reactome meta-network of 299,387 pairwise interactions between 10,672 genes. Two types of interactions were considered with different relevance: direct connections according to pathway-topology and participation to same pathway. Visualization, optimization and annotation of the network were performed using Cytoscape v3.5.1.

Sanger sequencing

Selected somatic variants with variant allele frequency (VAF) >15% were validated by direct Sanger sequencing of both control and tumor samples. Based on exome sequencing results, specific primers were designed for candidate mutations by using Primer Express 3.0 (Applied Biosystems, Foster City, CA). PCR products were purified by the Illustra ExoProStar kit (GE Healthcare, Little Chalfont, UK) and then sequenced with BigDye™ v.1.3 Cycle Sequencing kit and the Applied Biosystems 3500 Dx Series Genetic Analyzer (Applied Biosystems). PCR products of variants with VAF < 15% were cloned into TOPO TA cloning vector (Invitrogen, Carlsbad, CA) and at least 20 colonies were sequenced. Mutations on exon 1 of TNFRSF14 were evaluated in an extended series of other 12 patients with FL of the pediatric age (11 PTNFL, 1 PFLT), either by direct sequencing or PCR cloning and sequencing of at least 20 colonies.

Immunohistochemical analysis

Immunohistochemistry for phosphorylated extracellular signal-regulated kinase protein (pERK) was performed on 4 µm-thick formalin-fixed paraffin-embedded sections, using a primary anti-pERK rabbit polyclonal antibody (Cell Signaling, Leiden - NL, product ID: #9101). Heat/EDTA-based antigen retrieval methods were applied, as previously described (Pizzi *et al.*, Hum Pathol 2018). Antigen detection was performed with the Bond Polymer Refine Detection kit in an automated immunostainer (Bond maX, Menarini, Florence - Italy). Histological pictures were acquired by imaging digital techniques (Leica DFC295 digital microscope color camera and software; Milan - Italy).

Table S1. Clinico-pathological and cytogenetic features of the study population.

Patient ID	Gender	Age	Stage	Anatomic site	CD10	Bcl6	Ki67	MUM1	Bcl2	FISH BCL2	FISH BCL6	Tumor cells %	TNFRSF14 mutations	Treatment Protocol	Outcome
W1	M	10.9	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	neg	neg	neg	50	pos	complete resection, w&w	alive, CR
W2	M	15.8	II	peripheral lymph nodes	pos	pos	≥ 40%	na	neg	na	na	30-40	pos	LNH97 4 courses	alive, CR
W3	M	17.4	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	neg	neg	neg	60-70	neg	complete resection, w&w	alive, CR
W6	M	5.0	I	testis	pos	pos	≥ 40%	neg	neg	neg	neg	30	neg	complete resection, w&w	alive, CR
W7	M	16.7	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	neg	neg	neg	70	neg	LNH97 2 courses	alive, CR
W8	M	13.4	I	Ear-nose-throat	pos	pos	≥ 40%	neg	neg	neg	neg	30-40	neg	LNH97 4 courses	alive, CR2
W10	M	6.7	I	testis	pos	pos	≥ 40%	neg	neg	neg	neg	20-30	neg	complete resection, w&w	alive, CR
W13	M	14.2	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	neg	neg	neg	80	neg	complete resection, w&w	alive, CR
W14	F	14.5	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	weak	neg	neg	30-40	neg	complete resection, w&w	alive, CR
S1	M	7.8	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	neg	neg	neg	40	neg	complete resection, w&w	alive, CR
S2	M	15.0	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	neg	neg	neg	70	pos	complete resection, w&w	alive, CR
S3	M	14.8	II	peripheral lymph nodes	pos	pos	≥ 40%	neg	neg	neg	neg	30-40	neg	LNH97 4 courses	alive, CR
S4	M	16.8	I	Ear-nose-throat	pos	pos	≥ 40%	neg	neg	neg	neg	20	pos	complete resection, w&w	alive, CR
S5	M	5.0	I	testis	pos	pos	≥ 40%	neg	neg	neg	neg	10-15	pos	complete resection, w&w	alive, CR
S6	M	15.4	I	Ear-nose-throat	pos	pos	≥ 40%	neg	neg	neg	neg	40-50	neg	complete resection, w&w	alive, CR
S7	F	17.7	I	na	pos	pos	≥ 40%	neg	neg	neg	neg	70	neg	complete resection, w&w	alive, CR
S8	M	17.1	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	neg	neg	neg	60	pos	complete resection, w&w	alive, CR
S9	M	5.9	II	peripheral lymph nodes	pos	pos	≥ 40%	na	neg	na	na	80	neg	LNH97 4 courses	alive, CR
S10	M	10.1	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	weak	na	na	na	neg	complete resection, w&w	alive, CR
S11	M	4.0	I	peripheral lymph nodes	pos	pos	≥ 40%	neg	neg	neg	neg	30	pos	complete resection, w&w	alive, CR
S12	M	4.5	I	Ear-nose-throat	pos	pos	≥ 40%	na	neg	na	na	80	pos	LNH97 2 courses	alive, CR

Figure S1. Representative histological features of PTNFL.

A-B. At low magnification, the lymph node architecture was partially effaced by large, serpiginous follicles (**A**), composed of small to medium-sized blastoid elements with occasional large centroblasts (*insert*). **B-G.** The neoplastic follicles were diffusely positive for CD20 (**B**) with a rich CD3-positive T-cell background (**C**). Follicular B-cells expressed the germinal center markers, CD10 (**D**) and Bcl6 (**E**), and were consistently negative for both Bcl2 (**F**) and MUM1 (**G**). (H&E and peroxidase stain; panel A-C: bars indicate 1000 μm ; panel A, insert: bar indicates 50 μm ; panel D-G: bars indicate 100 μm).

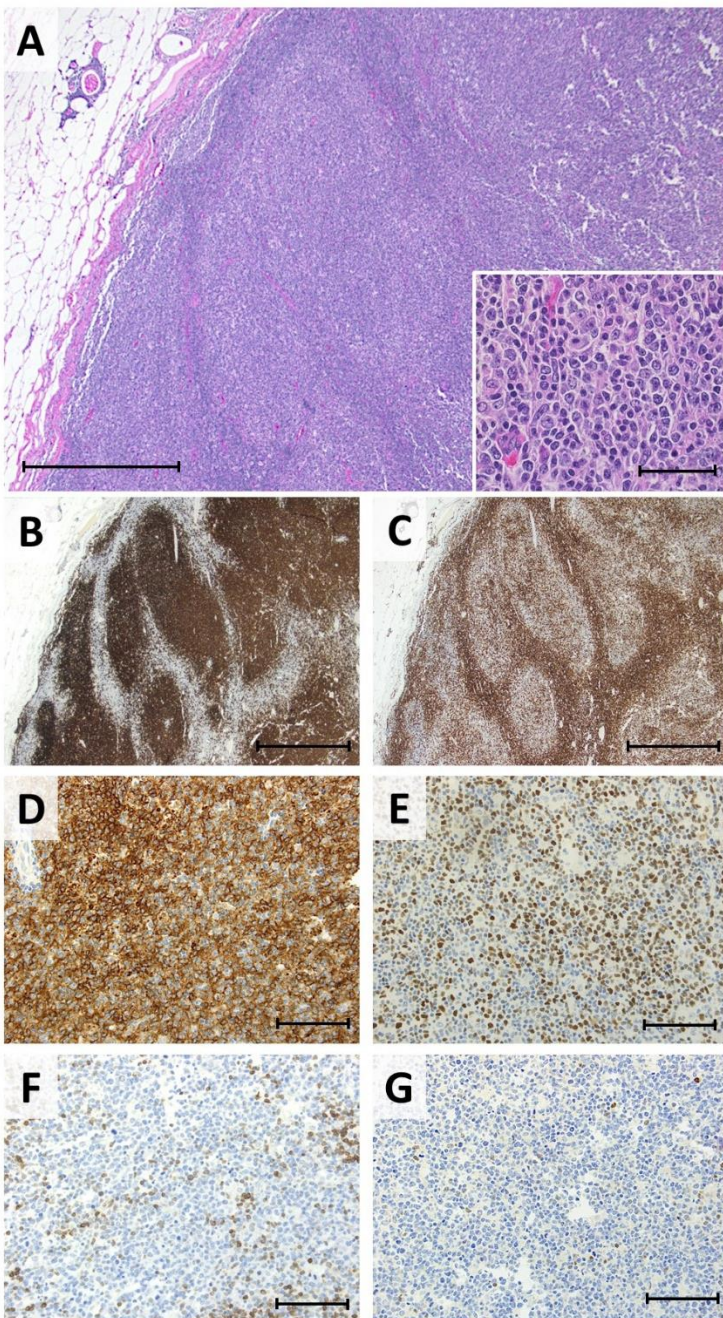


Figure S2. Bioinformatic workflow for detection of high confident somatic SNPs and reconstruction of a mutated pathway-derived network.

Reads mapping, reads processing and variant calling followed by annotation were automated by using SCons software building tool.

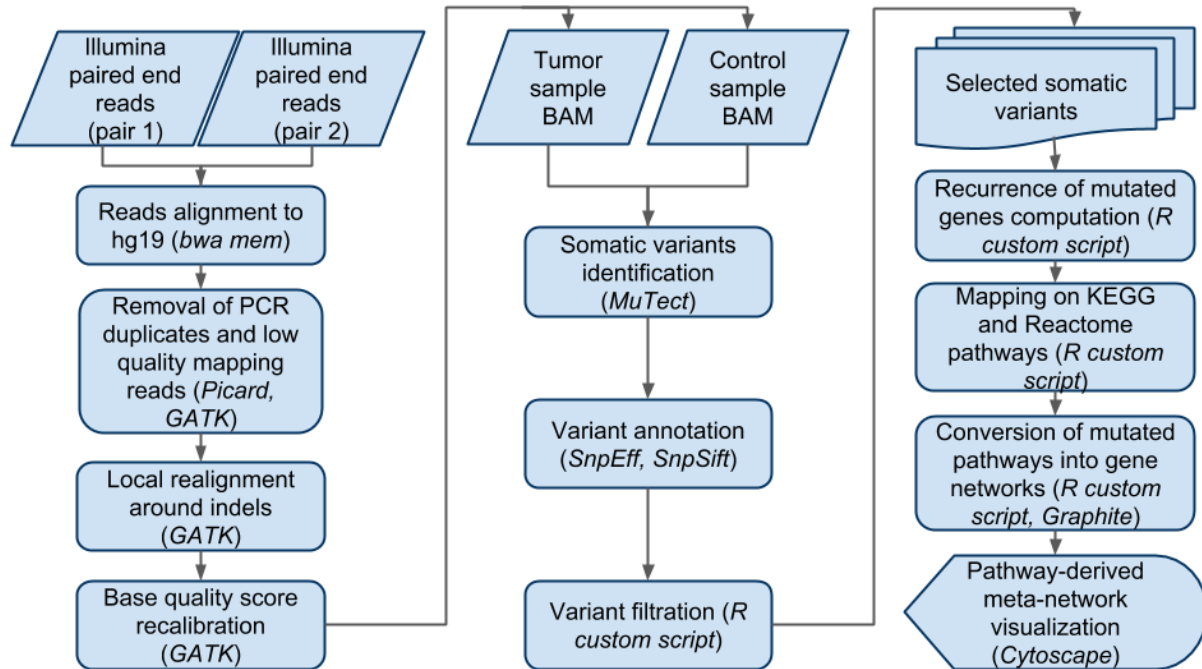


Figure S3. Sequence coverage profile per patient.

Each bar on the x-axis represents a patient; on the y-axis the percentage of targeted bases coverage is shown; different colors represent different coverage ranges; T, tumor sample, N, paired peripheral blood.

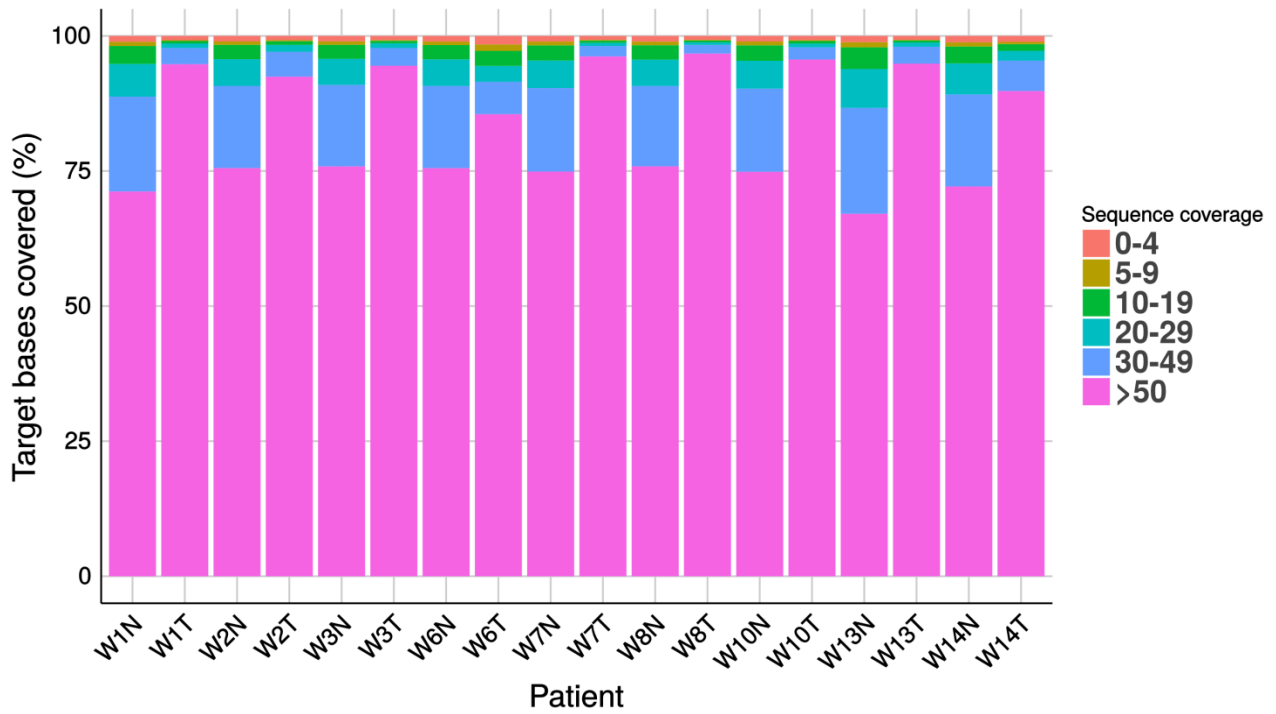


Figure S4. Somatic variants detected.

A) Mutation spectrum in each patient describing the proportions of the 6 transition and transversion categories observed among the selected high confidence somatic nucleotide substitutions; **B)** Number of somatic variants detected in each sample in relation with the mean coverage in the sample (no correlation is observed, rho: -0.35, p-value: 0.36).

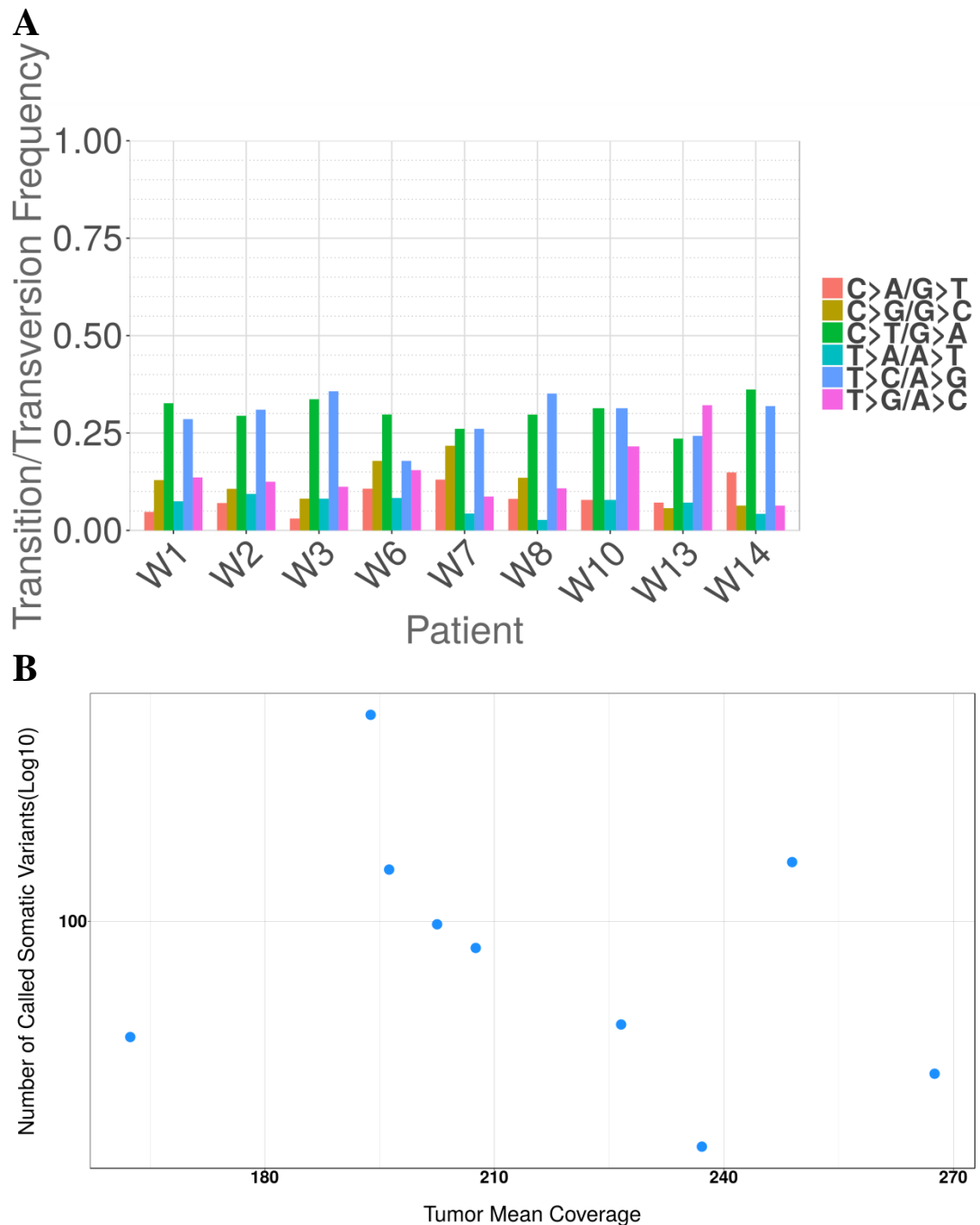


Table S2. Details on selected somatic variants detected in 9 cases of FL of the pediatric age and on involved transcripts with impact scoring (H, high; M, moderate) and amino acid change. (see separate .xlsx file)

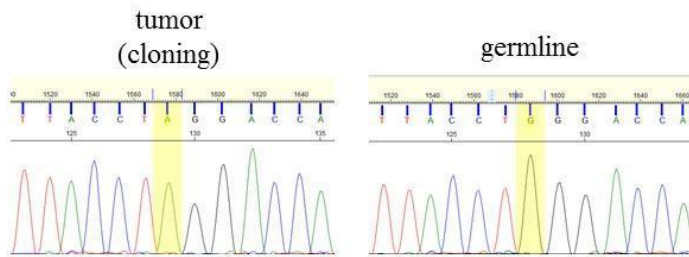
Table S3. Sequences of PCR primers used for validation of selected variants detected by WES.

Gene	Primer	Nucleotide sequence (5'-3')	PCR product
ARHGEF1	ARHGEF1_W746*_F	TCCCCCTCTCCCTGCAGATCA	100 bp
	ARHGEF1_W746*_R	CCGCTCCGACACAGTCTGT	
	ARHGEF1_R566H_F	CCCAAAGCCTGGGCCA	113 bp
	ARHGEF1_R566H_R	AGCCGCTGCATCTCCGT	
ATG7	ATG7_spldonor6-7_F	GGTTTCCTTGCTTAAACACTACAGTGA	123 bp
	ATG7_spldonor6-7_R	GTCTCTTGTAACATGAACTTCAGA	
GNA13	GNA13_E311*_F	CTGGTCCC GGCGTTTGT	141 bp
	GNA13_E311*_R	AACAAGACAGACTTGCTTGAGGAGA	
MAP2K1	MAP2K1_ex2_F	GCAGAAGAAGCTGGAGGAGCTA	128 bp
	MAP2K1_ex2_R	CCCCCAGCTCACTGATCTTCT	
RSF1	RSF1_Q1211*_F	CGTCGCAA ACTCTTCTGGGA	143 bp
	RSF1_Q1211*_R	TTTAAAATAGAAAGTGACTTCAGTGATGA	
TNFRSF14	TNFRSF14ex1_F	TGCCGGTCTGAGCCTGA	108 bp
	TNFRSF14ex1_R	AGGAGGCTCGGGGGCT	
UBAP2	UBAP2_Q38*_F	GCTTCAA AATCTGAATCATTCTTATCA	113 bp
	UBAP2_Q38*_R	CTTATGCTTTTGGTGTTCAAATCCT	
ZNF608	ZNF608_spldonor3-4_F	AATCTCTCTACAGTTTCCCTGCA	136 bp
	ZNF608_spldonor3-4_R	CGTGGAGGAACAAAACGTACGT	

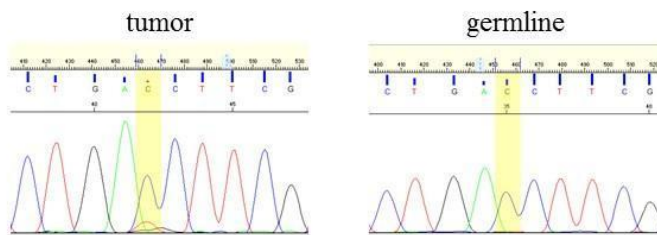
Figure S5. Validation of somatic variants obtained by Sanger sequencing.

For each patient, the sequence chromatograms of the appropriate tumor and paired germline samples are shown.

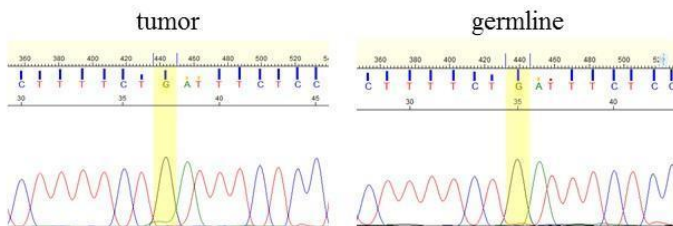
ARHGEF1 (chr19:42409131 G>A, p.Trp746*) - Patient W1



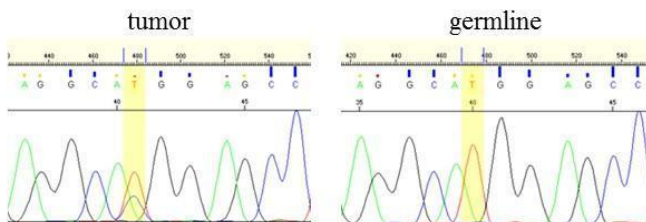
ATG7 (chr3:11356968 G>A, -475) - Patient W1



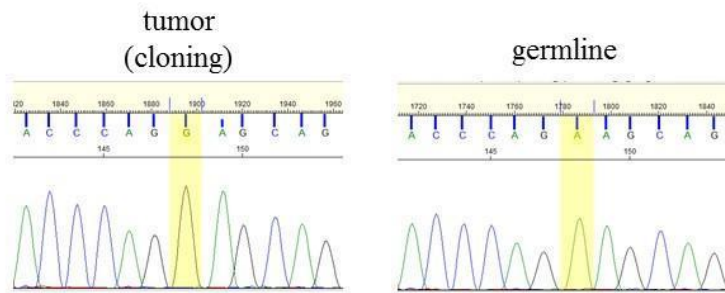
RSF1 (chr11:77383207 G>A, p.Gln1211*) - Patient W1



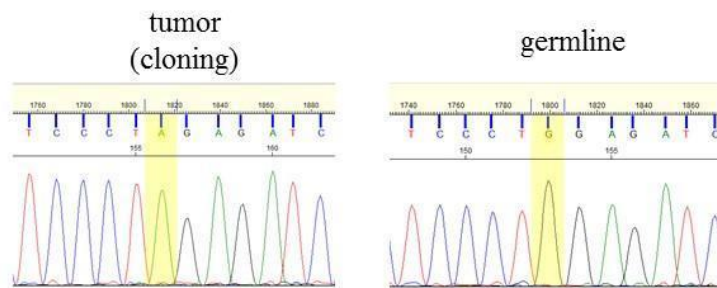
TNFRSF14 (chr1:2488105 T>C, p.Met1Thr) - Patient W1



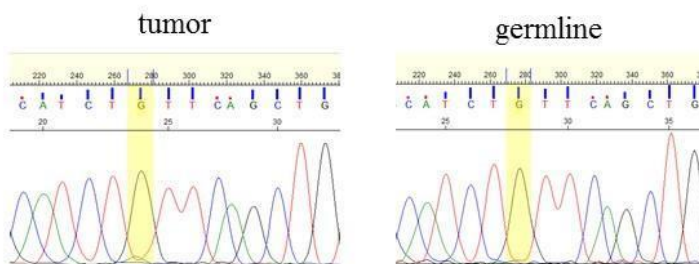
MAP2K1 (chr15:66727453 A>G, p.Lys57Glu) - Patient W2



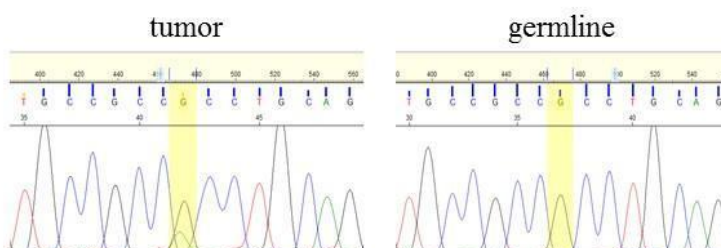
TNFRSF14 (chr1:2488138 G>A, p.Trp12*) - Patient W2



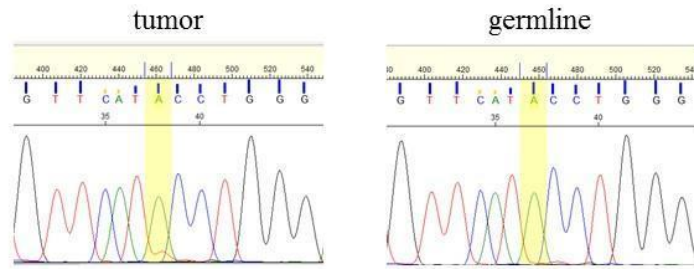
UBAP2 (chr:33998850 G>A, p.Gln38*) - Patient W2



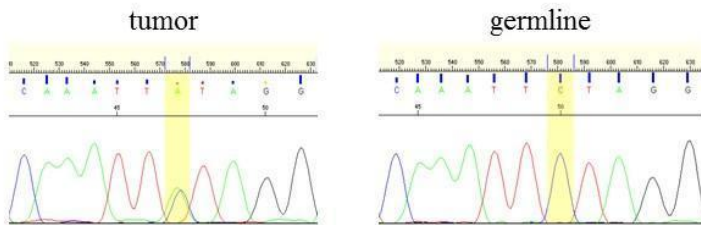
ARHGEF1 (chr19:42406962 G>A, p.Arg566His) - Patient W3



ZNF608 (chr5:123985301 A>T, -1016) - Patient W6



GNA13 (chr17:63010578 C>A, p.Glu311*) - Patient W13



MAP2K1 (chr15:66727463 T>G, p.Val60Gly) - Patient W13

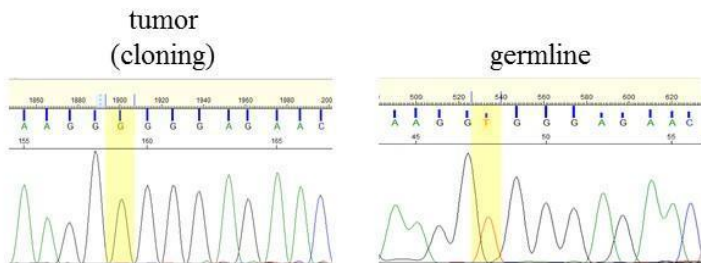
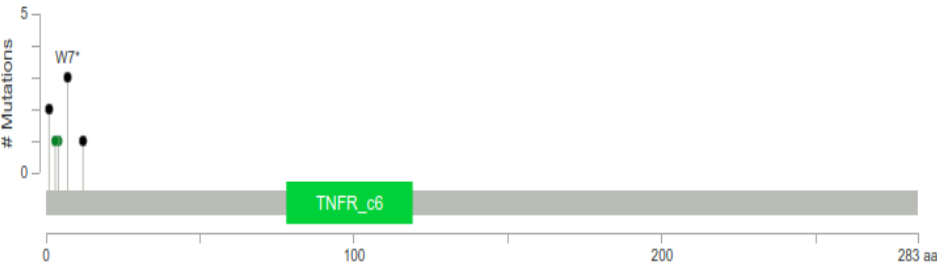
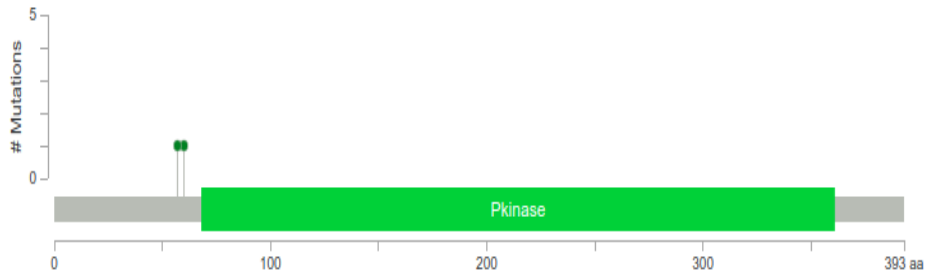
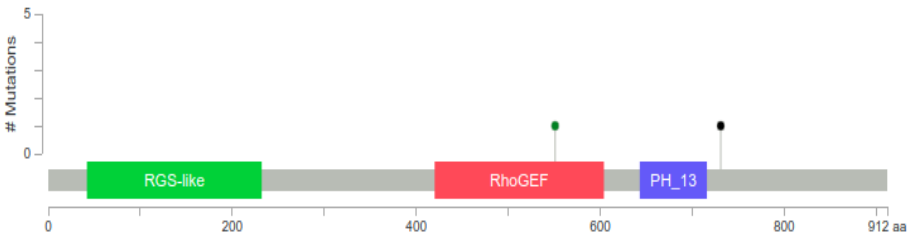
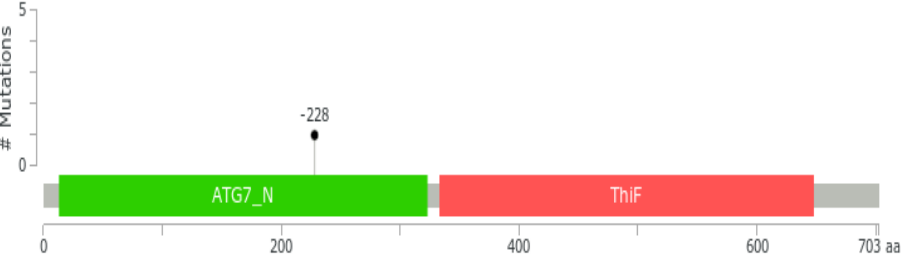
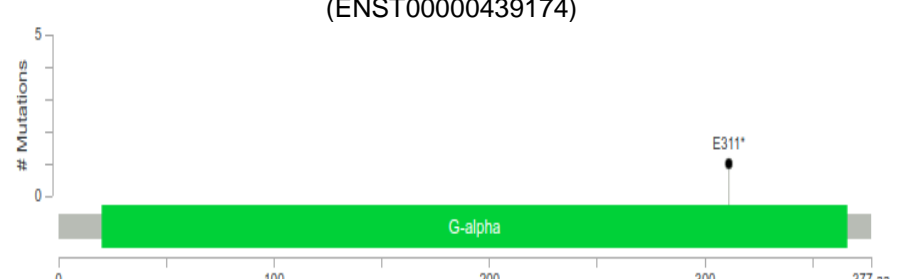


Figure S6. Impact of somatic variants to protein products.

For each mutated gene and considering the protein isoform encoded by the main transcript of the gene, Lollipop plots show the type and the position of selected somatic variants in relation to the protein sequence and domain structure. The columns on the right provide details on identified variants, as variant type, protein change, and possible functional relevance. Previously unknown variants are indicated in red.

GENE SYMBOL, PROTEIN (TRANSCRIPT ENSEMBL ID)	VARIANT (VAF)	AA CHANGE	SAMPLES	COMMENT
<p><i>TNFRSF14</i>, Tumor necrosis factor receptor superfamily member 14 (ENST00000355716)</p> 	chr1:2488105 T>C (0.36)	M1?	W1, S4	Start lost variant inducing loss of function
	chr1:2488111 C>T ()	P3L	S12	Missense variant in N-terminal signal peptide
	chr1:2488113 C>T()	P4S	S5	Missense variant in N-terminal signal peptide
	chr1:2488123 – 2488124 G>A	W7*	S2, S8, S11	Stop gain variant inducing loss of function
	chr1:2488138 G>A (0.10)	W12*	W2	Stop gain variant inducing loss of function
<p><i>MAP2K1</i>, Dual specificity mitogen-activated protein kinase 1 (ENT00000307102)</p> 	chr15:66727453 A>G (0.10)	K57E	W2	Missense variant lying closely to protein kinase domain (COSM1315807 and COSM5369532)
	chr15:66727463 T>G (0.22)	V60G	W13	Missense variant lying closely to protein kinase domain

<p>ARHGEF1, Rho guanine nucleotide exchange factor 1 (ENST00000354532)</p> 	chr19:42406962 G>A (0.26)	R551H	W3	Missense variant hitting RhoGEF domain (COSM6206921)
	chr19:42409131 G>A (0.16)	W731*	W1	Stop gain variant cutting the phosphotyrosine site in 738 position and a phosphoserine in 863 position
<p>ATG7, Autophagy-related protein 7 (ENST00000354449)</p> 	chr3:11356968 G>A (0.22)	-228	W1	Stop gain variant inducing loss of function
<p>GNA13, Guanine nucleotide-binding protein subunit alpha 13 (ENST00000439174)</p> 	chr17:63010578 C>A (0.47)	E311*	W13	Stop gain variant already detected in colorectal cancer (COSM1680008)
<p>RSF1, Remodeling and spacing factor 1 (ENST00000308488)</p> 	chr11:77383207 G>A (0.13)	Q1211*	W1	Stop gain variant inducing loss of function cutting 13 phosphoserine, 2 phosphothreonine and 2 N6-acetyllysine sites

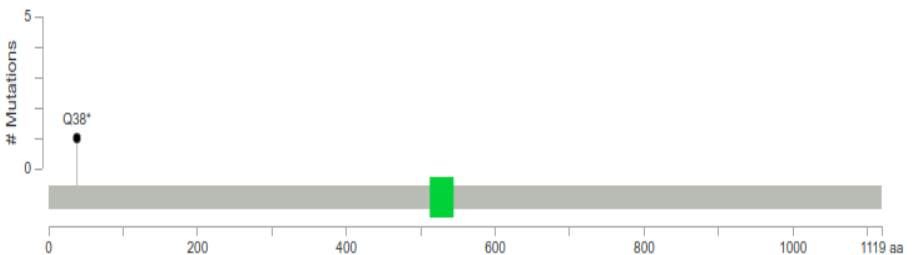
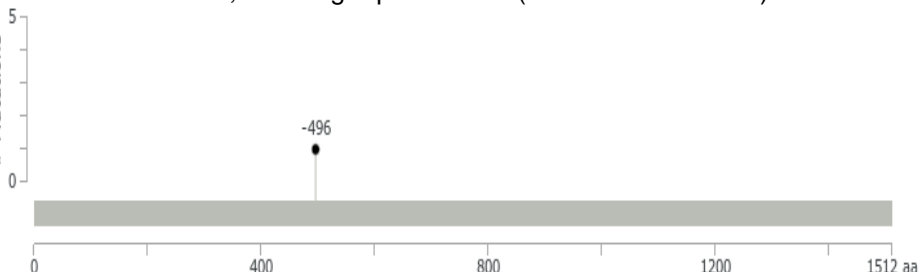
<p><i>UBAP2</i>, Ubiquitin-associated protein 2 (ENST00000360802)</p> 	<p>chr9:33998850 G>A (0.11)</p>	<p>Q38*</p>	<p>W2</p>	<p>Stop gain variant inducing loss of function cutting the most of protein including Ubiquitin associated domain</p>
<p><i>ZNF608</i>, Zinc finger protein 608 (ENST00000306315)</p> 	<p>chr5:123985301 A>T (0.11)</p>	<p>-496</p>	<p>W6</p>	<p>Stop gain variant probably inducing loss of function cutting 5 phosphoserine and 11 interaction sites for covalent linkages with SUMO2</p>

Figure S7. KEGG and Reactome pathways derived meta-network of genes somatically mutated in FL of the pediatric age.

A) Net nodes indicate genes somatically mutated in the cohort, with node size proportional to recurrence in the cohort analyzed by WES; genes involved in the negative regulation of MAPK, G-protein coupled receptor pathways, chromatin modifying enzymes and B-cell differentiation are indicated in different colors, using panel B as color legend; direct functional links between mutated genes are depicted and small grey nodes indicate pairs of mutated genes connected through a single non mutated gene in the network; isolated nodes located close to the connected components are associated to them according to annotated function and pathway participation; red node labels indicate the genes for which new mutations were identified. **B)** Mutation table showing which patient carries mutations in the genes of the main identified pathways. Patients are grouped according to the primary site of the tumor. Patients W7 and W8, negative for mutations connected with those represented in the figure, are not shown (PTNFL= pediatric-type nodal follicular lymphoma; PFLT= primary follicular lymphoma of the testis).

