의학박사 학위논문

Genetic alterations of NK/T-cell lymphoma

NK/T 림프종의 유전변이

2017 년 2 월

서울대학교 대학원

의과학과 의과학 전공

박 하 영

NK/T 림프종의 유전변이

지도교수 김 종 일

이 논문을 의학박사 학위논문으로 제출함
2017년 2월

서울대학교 대학원
의과학과 의과학전공
박 하 영

박하영의 의학박사 학위논문을 인준함
2017년 1월

위 원 장 _____ (인)
부위원장 _____ (인)
위 원 _____ (인)
위 원 _____ (인)
위 원 _____ (인)

Genetic alterations of NK/T-cell lymphoma

by

Ha Young Park

A thesis submitted to the Department of Biomedical Science in partial fulfillment of the requirement of the Degree of Doctor of Philosophy in Medical Science at Seoul National University College of Medicine

January 2017

Approved by Thesis Committee:

Professor _____Chairman

Professor _____Vice chairman

Professor _____

Professor _____

Professor _____

ABSTRACT

Genetic alterations of NK/T-cell lymphoma

Ha Young Park

Major in Biomedical Science

Department of Biomedical Science

Seoul National University Graduate School

Introduction: Extranodal NK/T-cell lymphoma nasal type (ENKL) is a rare type of non-Hodgkin lymphoma that more frequently occurs in East Asia and Latin America. Even though its molecular background has been discussed in the last few years, the current knowledge does not explain the disease pathogenesis in most cases of ENKL.

Methods: To investigate the ENKL-specific genetic variants, we performed multiple types of next-generation sequencing on 34 ENKL samples, including whole-exome sequencing (9 cancer tissues and 4 cancer cell lines), targeted sequencing (21 cancer tissues), and RNA sequencing (3 cancer tissues and 4 cancer cell lines).

Results: Mutations were found most frequently in 3 genes, *STAT3*, *BCOR*, and

MLL2 (which were present in 9, 7, and 6 cancer samples, respectively), whereas there were only 2 cases of *JAK3* mutation. In total, JAK/STAT pathway- and histone modification-related genes accounted for 55.9% and 38.2% of cancer samples, respectively, and their involvement in ENKL pathogenesis was also supported by gene expression analysis. In addition, we provided 177 genes upregulated only in cancer tissues, which appear to be linked with angiocentric and angiodestructive growth of ENKL.

Conclusions: In this study, we propose several novel driver genes of ENKL, and show that these genes and their functional groups may be future therapeutic targets of this disease.

* This work is published in Oncotarget. (1)

Keywords: extranodal NK/T-cell lymphoma nasal type, next-generation sequencing, JAK-STAT pathway, chromatin modification, somatic mutation, transcriptome

Student number: 2013-30609

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ENKL : Extranodal NK/T-cell lymphoma nasal type

NGS : Next-generation sequencing

EBV : Epstein–Barr virus

TSGs : tumor suppressor genes

FFPE : formalin-fixed paraffin-embedded

WES : Whole exome sequencing

nsSNVs : nonsynonymous single nucleotide variants

indels : insertions/deletions

INTRODUCTION

Extranodal NK/T-cell lymphoma nasal type (ENKL)

Extranodal natural killer (NK)/T-cell lymphoma nasal type (ENKL) is a rare histopathological subtype of lymphoma with unique clinical features and geographic variation. This lymphoma usually occurs in the upper airway tract, mostly in the nasal and paranasal area; however, around 20% of cases can occur in other tissues, including the skin, soft tissue, gastrointestinal tract, and testis (2-5). Although most of the cases of ENKL are diagnosed in the early stage of the disease, patients usually have poor response to combination chemotherapy, resulting in 46%–60% long-term survival (4, 5). Only half of the patients with advanced-stage disease can survive for 1 year, despite improvements in treatment (6, 7).

Genetic variations of ENKL in previous studies

ENKL is strongly associated with Epstein–Barr virus (EBV) infection. EBV infection is thought to be an early event in the pathogenesis of the disease (8), and additional genetic alterations are essential to induce lymphomagenesis. Mutations in well-known tumor suppressor genes (TSGs), including *TP53* (20%–60% of cases) and *FAS* (around 50% of cases), have been reported (9), and array-based comparative genomic

hybridization studies reported variable genetic changes, including gains in 1q21–q44, 2q, and 7q, and loss of 6q16–27 and 17p15–22. Among these lesions, the 6q region includes several TSGs, such as *PRDM1*, *FOXO3*, and *HACE1* (10, 11). In particular, recent studies reported frequent *JAK3*-activating mutations in ENKL patients using next-generation sequencing (NGS), which suggests that the JAK/STAT signaling pathway is a key molecular factor in the pathogenesis of this disease (12, 13). *DDX3X* encoding RNA helicase frequently mutated in Chinese population, which was also identified by NGS most recently (14).

Previous gene expression profiling of NK cell malignancies led to the association between apoptosis, cell adhesion molecules/extracellular matrix (ECM) receptor interaction, and signal transduction pathways (including JAK/STAT, mTOR, etc.) with ENKL tissues (15). MicroRNA dysregulation, which is significantly enriched among genes involved in cell cycle-related, p53, and MAPK signaling, was also suggested as a mechanism of lymphomagenesis (16). However, the current knowledge on ENKL does not adequately explain the disease pathogenesis in most cases

Genomic sequencing in cancer study

Genomic sequencing based on target enrichment is a cost-effective manner of observing cancer genome alterations in nucleotide level. In addition to whole-exome sequencing (WES), customized targeted sequencing can be used for

screening specific genomic regions of interest. Besides them, the whole-transcriptomic approach using RNA-Seq has also become one of the common initial methods to discover genetic and molecular alterations in cancer (17). In addition to gene expression profiling, RNA-Seq enables the identification of genomic variants, including fusion gene mutations.

Purpose and design of study

Although several studies have suggested a carcinogenic mechanism for ENKL, its molecular pathogenesis remains under-recognized. Because of the rarity of the disease and limitations in obtaining sufficient amounts of fresh tissues for molecular studies, to date, its study has been restricted to a few research groups. Here, we report genomic information on ENKL using the NGS method, and suggest several novel genes or pathways involved in the pathogenesis of this disease. In addition, we provide comparative data on gene expression profiles in primary ENKL tissues, NK-cell lymphoma cell lines, and normal NK cells, which support the genetic alterations discovered in the genomic sequencing experiments.

MATERIALS AND METHODS

NGS study subjects

In this study, we collected 34 cancer samples for sequencing, which included 9 fresh-frozen ENKL tissues (CT1–9), 4 NK-cell lymphoma cell lines (CC1–4), and 21 formalin-fixed paraffin-embedded (FFPE) ENKL tissues (TT1–21). The diagnosis of ENKL was established according to the 2008 World Health Organization classification. In all cases, the immunohistochemical study for CD3, CD56, TIA-1, and granzyme B, as well as EBER in situ hybridization was carried out using FFPE tissues. More than 70% of the infiltrated cells in the samples were positive for CD56. At least more than 70% of CD56-positive cells were positive for EBER in situ hybridization. In the CD56-negative case, the case was included for the study if EBV was positive in more than 50% of the infiltrating cells. For the frozen samples, frozen section was made for evaluation of cellularity and necrosis. As controls, 4 matched normal blood samples (NB1–4, each paired with CT1–4) and 3 NK cell samples (NC1–3) were obtained from the peripheral blood of volunteers (11). For NGS, we extracted DNA from all study samples and RNA from some tissues (CT1, 2, and 5) and cell lines (CC1-4 and NC1–3). Table 1 summarizes the information of the study subjects who were enrolled in this study. All patient samples were obtained in Samsung Medical Center, Seoul, Korea and this study was approved by the Institutional Review Board of the Samsung Medical Center, in accordance with the Declaration of Helsinki (approval number, 2013-12-076)

Table 1. Study subjects for sequencing and clinical information

Sample ID	Cell line ID	Gender	Age at Diagnosis	Tissue origin or original diagnosis of cell line	Stage
CC1	NK92	Male	50	ANKL	
CC2	NKYS	Female	19	ENKL	
CC3	SNK6	Male	62	ENKL	
CC4	HANK1	Female	46	ENKL	
CT1		Male	47	nasal cavity	II
CT2		Male	45	nasal cavity	IV
CT3		Male	43	nasal cavity	II
CT4		Female	55	nasal cavity	IV
CT5		Male	33	nasal cavity	IAE
CT6		Female	46	nasal cavity	II
CT7		Male	44	nasal cavity	II
CT8		Male	60	nasal cavity	II
CT9		Female	82	nasal cavity	I
TT1		Male	83	nasal cavity	IV
TT2		Male	39	Ileum	I
TT3		Female	53	colon	I
TT4		Male	39	nasal cavity	IV
TT5		Female	68	nasal cavity	IB
TT6		Male	74	nasal cavity	I
TT7		Male	55	nasal cavity	I
TT8		Male	48	palate	II
TT9		Male	43	nasal cavity	II
TT10		Male	49	nasal cavity	I
TT11		Male	64	skin	III
TT12		Male	61	nasal cavity	I
TT13		Male	42	nasal cavity	IV
TT14		Female	51	nasal cavity	I
TT15		Female	59	skin	IV
TT16		Female	49	nasal cavity	I
TT17		Female	53	nasal cavity	I
TT18		Male	57	nasal cavity	II
TT19		Female	44	nasal cavity	I
TT20		Male	50	nasal cavity	IIB
TT21		Male	47	nasal cavity	II

DNA and RNA sequencing

DNA samples from 13 cancers (CT1–9 and CC1-4) and 7 normals (NB1–4 and NC1–3) underwent WES using SureSelect Human All Exon 50M (Agilent Inc., Palo Alto, CA). For the FFPE tissue samples, we applied targeted amplicon sequencing using the HaloPlex target enrichment kit (Agilent Inc.). The target regions were designed to include coding exons of ENKL-associated genes (18), cancer-driver genes (19), JAK/STAT signaling pathway genes, and BCOR network genes, together with those of genes with ≥ 2 somatic variants in 4 tumor-normal WES pairs. The full list of target genes is available in the Table 9. Sequencing was performed on a HiSeq 2000 machine (Illumina, San Diego, CA), and the reads that were generated were aligned to the National Center for Biotechnology Information (NCBI) human reference genome (hg19) using Bowtie2 (20). The mutated genes in GO group, "chromatin modification" (GO:0016568) and JAK/STAT signaling pathway of KEGG from targeted sequencing were confirmed to be somatic by Sanger sequencing on paired-normal DNAs, which were extracted from uninvolved bone marrow tissues. In all cases, bone marrow biopsy was performed as a staging work up. Bone marrow involvement was analyzed by FACS analysis for aspirate and EBER in situ hybridization for trephine biopsy. Mutual exclusivities of the genes were tested by Gitools (21). We also checked for known *JAK3* mutations (p.Ala572Val, p.Ala573Val) by Sanger sequencing.

RNA was assessed for quality and was quantified using an RNA 6000 Nano LabChip on a 2100 Bioanalyzer (Agilent Inc.). The sequencing libraries were

prepared as described previously and sequenced on a HiSeq 2000 machine (Illumina) (17). The sequenced reads were aligned to the NCBI human reference genome (hg19) using the STAR 2-pass method (22, 23).

Table 2. Candidates of targeted sequencing

Groups	Genes
Somatic >=2 samples	FGF10, FRY, MLL2, TP53
Cancer driver	ABL2, AKT1, ARID1A, ARID1B, ARNT, ASXL1, AXIN1, BAP1, BCL11A, BCOR, BRAF, BRCA2, C15orf55, CASP8, CBL, CCND2, CD274, CHEK2, CIITA, DAXX, DICER1, EGFR, EPS15, ERBB2, ERG, ETV4, FANCA, FANCC, FANCD2, FANCG, FAS, FGFR2, FGFR3, FLCN, FNBP1, FOXO1, FOXP1, FUBP1, GPHN, HAS2, HIP1, HOXA13, IDH1, IDH2, IKZF1, ITK, JAK1, JAK3, KDM5A, KDM5C, KDM6A, KIT, LASP1, LCK, LIFR, MAFB, MAP3K1, MKL1, MLH1, MLL, MLLT10, MLLT4, MSH6, MYB, MYH11, NCOR1, NFKB2, NIN, NPM1, NSD1, NTRK3, NUMA1, NUP214, PAX3, PAX5, PAX8, PDE4DIP, PPARG, PRDM16, PTCH1, PTEN, RANBP17, RARA, ROS1, SETBP1, SETD2, SMAD4, TAL2, TET2, TNFAIP3, TPM3, TSC2, TSHR, WAS, WHSC1, WT1, ZNF521
JAK/STAT pathway	CBLB, FES, IL4R, IL7R, STAM, STAT3, IFNAR1, PTPN2
BCOR network	CPNE4, HDAC3, HDAC5, HDAC6, ZBTB7A
Known ENKL-associated	ABCB1, ABL1, ACTN4, AIM1, AKAP1, AKT2, AKT3, ALL, ARAP1 , ARHGDIA, ATG5, AURKA, BAIAP2, BCAR1, BCL2L1, BLK, CALM3, CAMKK2, CASP9, CAV1, CCL2, CCNA2, CDH5, CDK1, CDKN1A, CDKN1B, CDKN2A, CDKN2B, CHUK, CREB1, CRK, CRKL, CSNK2A1, CTNNB1, CX3CL1, CXCL10, CXCL9, ELK1, FIGF, FLT1 , FLT4, FOS, FOXO3, GSK3A, GSK3B, HACE1, HCK, HRAS, IL6R, IQGAP1, ITGAV, JAK2, JUN, KDR, KLRD1, KPNA1, MAP2K1 , MAP2K2 , MAP2K3 , MAP2K4 , MAP2K6 , MAP2K7 , MAP3K5, MAPK1, MAPK10, MAPK3, MAPK8, MAPK9, MET, MYC, MYCN, MYOF, NRAS, PDGFA, PDGFB, PDGFC, PDGFRA, PDPK1, PIK3CB, PIK3CD, PIK3CG, PIK3R1, PIK3R2, PIK3R3, PIK3R6 , PIN1, PLA2G4A, PLCG1, PLK1, PRDM1, PRKAA1, PRKAB1, PRKACA, PRKAG1, PRKCA, PRKCB , PRKCD, PRKCE, PRKDC, PTK2B, PTPN11, PTPN6, PTPRJ, PVR, PXN, RAF1, RAPGEF1, RARB, RASA1, RB1, ROCK1, RPS6KA1, RPS6KA3, S1PR1 , SERPINB9, SFN, SH2D2A, SHB, SHC1, SHF, SLA, SLC2A4, SLC9A3R1 , SLC9A3R2 , SPHK1, SRC, SRF, STAT1, STAT5A, STAT5B, TAGLN, TBC1D4, TP73, USP6NL, VAV2, YES1, YWHAB, YWHAE, YWHAG, YWHAH, YWHAQ, YWHAZ

Sequence variation analysis

In WES, reads marked as PCR duplicates were removed from downstream analysis (Picard, http://broadinstitute.github.io/picard/), and GATK was used to perform indel realignment and base quality score recalibration (24). For somatic variant calling from 4 tumor/normal pairs, we used muTect and GATK SomaticIndelDetector to call SNVs and indels, respectively (25). In the case of unpaired DNA samples, including those from FFPE tissues, GATK UnifiedGenotyper was used to call mutation candidates. All variants called were annotated in several genomic databases using ANNOVAR (26) and further narrowed down to driver candidates, as follows: (1) nonsilent SNVs or coding indels; (2) allele frequency = 0.0 in the 1000 Genomes Project, Exome Sequencing Project, and Complete Genomics sequencing data (27, 28); (3) not shown in an additional 919 Korean exomes including NC1–3 (unpublished); and (4) located in genes other than those reported previously as having many false positives (29).

For variant calling from RNA sequencing data, we followed the best-practice recommendations of GATK, which include indel realignment, base recalibration, and a variant-calling process using HaplotypeCaller (24). As a validation step, called variants were compared with those obtained in the genomic sequencing experiment described above.

Fusion gene analysis

We used the deFuse tool to discover fusion transcripts in RNA-sequenced samples

(30). In addition to the default filtration, we tried to filter false-positive calls according to the following criteria: (1) every gene pair shown in NC samples was excluded; (2) adjacent gene fusions and those with gene distances < 200 kb were removed, unless they were predicted to be inversions or eversions; (3) fusions called as alternative splicing events were ignored; (4) fusions were supported by ≥ 5 spanning reads and ≥ 8 spanning mate pairs across breakpoints; (5) at least 1 gene in pairs was included in the RefSeq gene set; and (6) pairs of gene fusions were not paralogs of each other based on Ensembl version 72 (31, 32). The fusions selected above were confirmed using the TopHat-Fusion caller with default options (33), and their genes were compared with the cancer gene census of COSMIC (34).

Gene expression profiles of ENKL

We applied the HTSeq and DESeq2 tools to our RNA-Seq reads for gene expression analysis (35, 36). Hierarchical clustering of samples was conducted with gene expression levels using Cluster 3.0, the results of which were visualized using Java Treeview (37, 38).

DEGs were defined as those with a q-value < 0.05 and $|\log_2(\text{fold change})| \geq 1$, thus differentiating 1 sample group from another. For further filtration, we also called DEGs for each cancer sample in comparison with NC. The genes selected here underwent gene set enrichment analysis (GSEA) (39). Among the Molecular Signatures Database, which is a collection of gene sets for use with GSEA, we selected the following gene databases for analysis: curated gene sets including

chemical and genetic perturbations (CGP) and KEGG gene sets. For gene ontology (GO) analysis, we used the ClueGO program, which is implemented in Cytoscape (40, 41).

RESULTS

ENKL samples exhibited genomic heterogeneity and frequent mutations in TSGs

We performed WES on 9 ENKL tissue samples (cancer tissue, CT1-9) and 4 NK-cell lymphoma/leukemia cell line samples (cancer cell, CC1-4), with 4 paired normal blood (normal blood, NB1-4) and 3 unpaired NK cells from healthy volunteers (normal cell, NC1-3) as controls. The sequencing summary are listed up in Table 3. A total of 251 somatic mutations, including 220 nonsynonymous single nucleotide variants (nsSNVs) and 31 coding insertions/deletions (indels), were identified from 4 paired samples. However, each mutation was identified in just 1 sample, and only 5 genes (*FGF10*, *KRAS*, *MLL2*, *FRY*, and *TP53*) were shared by ≥ 2 samples. These somatic variants were not detected further in 5 ENKL tissues and 4 NK-cell lymphoma cell line samples. Therefore, we extended our variant analysis to include RNA-Seq and targeted sequencing of 21 paraffin-embedded samples, which were analyzed together with 13 WES samples (Tables 1 and 2).

When all platforms were considered (WES, targeted sequencing, and RNA-Seq), *STAT3* was the most frequently mutated gene (9/34 cases, 26.5%), followed by *BCOR* (7/34 cases, 20.6%), and *MLL2* (6/34 cases, 17.6%) (Figure 1). Interestingly, their variants were nearly mutually exclusive of each other (P = 0.000044). Among the known ENKL-associated genes, mutations were most frequent in *TP53* (4/34 cases, 11.8%), followed by *KRAS* (2/34 cases, 5.9%), and *IL6R* (2/34 cases, 5.9%) (18). In

Table 3. Whole exome sequencing and targeted sequencing summary

Sample ID	**# of total read**	**% of reads mapped to reference**	**% of reads aligned on target regions**	**Mean read depth on target regions (x)**
Whole exome sequencing				
CT1	91623886	87.97	42.33	55.31
CT2	106566657	87.43	44.73	67.8
CT3	83573414	86.42	43.91	51.28
CT4	100528275	88.6	41.95	60.54
CT5	61658328	88.32	45.46	40.09
CT6	115484318	88.7	43.83	72.95
CT7	107550158	87.72	43.17	66.09
CT8	79072891	88.29	45.36	51.27
CT9	56206504	88.66	41.32	33.32
NB1	53961299	96.59	24.42	11.03
NB2	78163556	97.37	28.53	18.86
NB3	59571859	96.84	24.79	12.36
NB4	82030376	97.55	31.41	21.89
CC1	70225407	95.67	29.24	17.28
CC2	71155111	96.15	31.14	18.7
CC3	67066008	96.27	31.14	17.68
CC4	75820509	96.15	30.24	19.38
NC1	63114578	96.25	29.83	15.89
NC2	71838616	96.12	30.48	18.49
NC3	88579148	96.38	28.47	21.33

Targeted sequencing				
TT1	11159591	97.43	90.5	404.72
TT2	11702018	96.14	91.4	419.84
TT3	11845989	96.96	89.8	429.68
TT4	15508183	96.27	89.6	559.46
TT5	13276718	97.86	89.2	488.96
TT6	9120711	99.1	91.4	339.84
TT7	12611991	98.52	90.8	467.83
TT8	10353037	97.98	90	383.27
TT9	11824189	99.52	91.3	447.29
TT10	6094752	98.45	91.1	227.21
TT11	5975530	98.38	90.6	221.64
TT12	7901997	99.01	89.9	297.68
TT13	6523981	99	90.8	244.03
TT14	14435633	98.97	91.1	535.23
TT15	14323547	99.35	90.2	541.2
TT16	10131464	99.18	91.1	380.76
TT17	13024576	99.41	89.3	494.81
TT18	14612968	99.25	91	547.23
TT19	12598787	98.78	93	463.78
TT20	13182418	99.08	91.8	492.41
TT21	10592149	99.26	91.3	397.35

Table 4. RNA-Seq summary

Sample	# of input read pairs	# of uniquely mapped read pairs	% of uniquely mapped reads	# of splices	Mismatch rate per base (%)	Deletion rate per base (%)	Deletion average length (bp)	Insertion rate per base (%)	Insertion average length (bp)	# of read pairs mapped to multiple loci	% of reads mapped to multiple loci
CT1	134,608,830	96,217,766	71.48%	20,027,595	1.76%	0.01%	1.69	0.17%	1.03	6,727,932	5.00%
CT2	67,178,728	51,770,240	77.06%	13,957,240	1.68%	0.01%	1.67	0.17%	1.02	2,622,296	3.90%
CT5	60,416,976	48,604,024	80.45%	13,432,784	1.29%	0.01%	1.7	0.17%	1.02	2,873,880	4.76%
CC1	82,124,472	68,529,218	83.45%	21,988,419	1.23%	0.01%	1.59	0.01%	1.31	2,164,360	2.64%
CC2	58,914,806	48,426,878	82.20%	17,649,500	1.30%	0.01%	1.56	0.01%	1.33	2,053,480	3.49%
CC3	57,018,962	46,214,554	81.05%	14,541,826	1.28%	0.01%	1.57	0.01%	1.31	2,252,324	3.95%
CC4	87,302,558	71,680,848	82.11%	23,989,492	1.10%	0.01%	1.62	0.01%	1.32	3,328,890	3.81%
NC1	66,248,234	59,676,512	90.08%	18,230,789	1.00%	0.01%	1.59	0.01%	1.33	2,334,814	3.52%
NC2	61,463,934	53,245,120	86.63%	15,540,434	1.09%	0.01%	1.63	0.01%	1.32	2,060,568	3.35%
NC3	66,318,492	56,943,484	85.86%	16,695,473	1.12%	0.01%	1.62	0.01%	1.35	2,129,828	3.21%

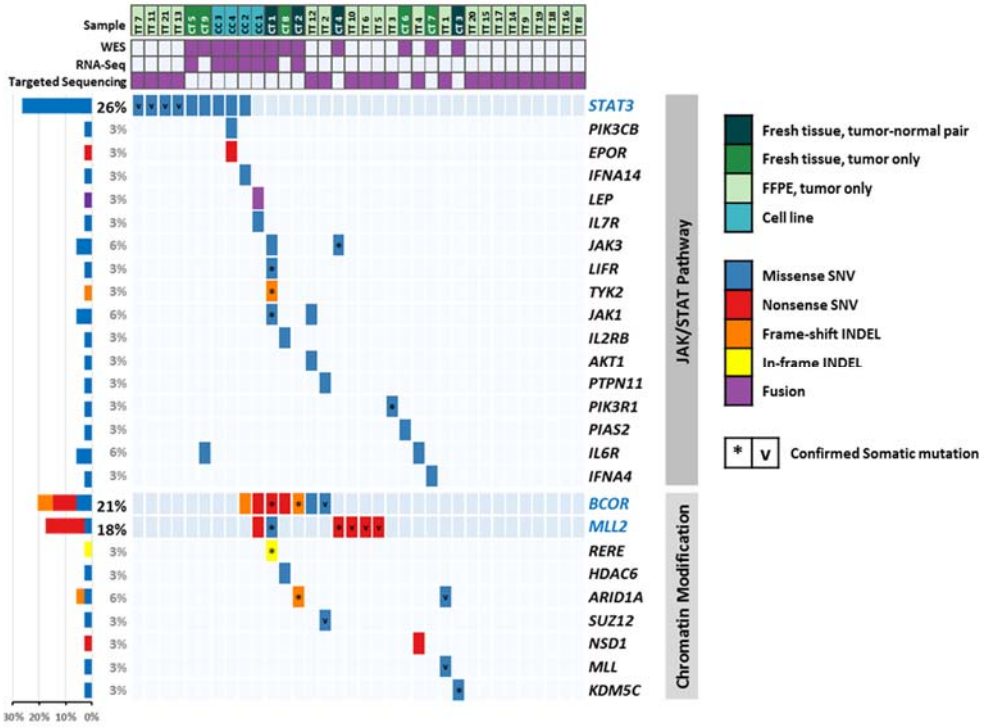

Figure 1. Distribution of mutations in ENKL. Each column represents an ENKL case, which is shown in different colors according to the platform used. Each row represents a gene in the JAK/STAT pathway or chromatin modification gene ontology. Five mutation types are also distinguished by different colors. Asterisks (*) indicate somatic mutations in four tumor-normal pairs of exome sequencing. 'v' marks indicate mutations in targeted sequencing which confirmed to be somatic by Sanger sequencing of matched normal (uninvolved bone marrow) samples. Bars (left) represent the mutation rates in the 34 samples. ENKL, extranodal NK/T-cell lymphoma nasal-type.

total, 22 cases had mutations in TSGs, including *MLL2*, *BCOR*, and *TP53*, 45% of which (10/22 cases) had at least 1 loss-of-function (LOF) mutation (nonsense SNV/frameshift indel), which implies that inactivation of TSGs may play an important role in ENKL pathogenesis.

Genome analysis of ENKL revealed enrichment of alterations in the JAK/STAT pathway, among which *STAT3* was the most frequently mutated gene

In JAK/STAT cascade genes, all mutations occurred at different positions, with the exception of 1 nsSNV of *STAT3* encoding p.Glu616Gly, which was identified in 2 out of 34 cases. Another 2 cases showed the same amino acid alteration (*STAT3* p.Ser614Arg), differing only at the nucleotide level. Interestingly, all mutations occurred at the SRC homology 2 (SH2) domain of STAT3 (Figure 2).

A recent study reported that activating mutations of *JAK3* (p.Ala572Val and p.Ala573Val) may play an important role in the pathogenesis of ENKL (12). Because of the low coverage of WES in this region, we confirmed these mutations by Sanger sequencing. A total of 39 samples were used for checking *JAK3* hotspots, which included a subset of the samples used in exome and targeted sequencing and new 17 FFPE samples (ST1-17, Table 5). Only 2 cases (2/39 cases, 5.1%) exhibited p.Ala527Val mutations, a frequency that was quite different from that mentioned in the report described above, which discovered either of 2 hotspot mutations in up to 35.4% of ENKL patients. We thoroughly examined the coding

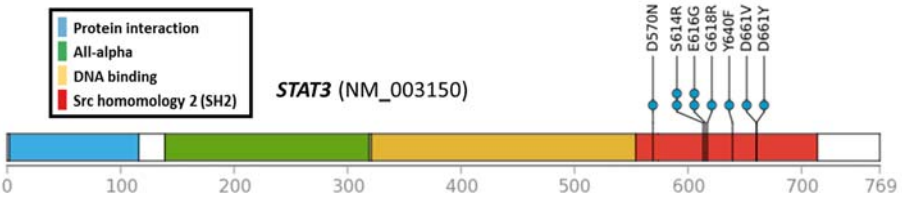

Figure 2. Locations of STAT3 mutations. All 9 missense SNVs were accumulated in the SH2 domain.

Table 5. Sanger sequencing result of JAK3 hotspots

Sample ID	**chr19:17948006 (C>T) p.Ala572Val**	**chr19:17948009 (C>T) p.Ala573Val**
CT1	**T**	C
CT2	C	C
CT8	C	C
CT9	C	C
ST1	C	C
ST2	C	C
ST3	C	C
ST4	C	C
ST5	C	C
ST6	C	C
ST7	**T**	C
ST8	C	C
ST9	C	C
ST10	C	C
ST11	C	C
ST12	C	C
ST13	C	C
ST14	C	C
ST15	C	C
ST16	C	C
ST17	C	C
TT1	C	C
TT3	C	C
TT4	C	C
TT5	C	C
TT6	C	C

TT7	C	C
TT8	C	C
TT9	C	C
TT10	C	C
TT11	C	C
TT12	C	C
TT13	C	C
TT14	C	C
TT15	C	C
TT16	C	C
TT17	C	C
TT18	C	C
TT21	C	C

regions of *JAK3* by targeted deep sequencing again, but no additional mutation was found. However, genes included in the JAK/STAT signaling pathway (42) were mutated in about 55.9 % (19/34) of all cases (Figure 1).

Mutations in *MLL2* and *BCOR*, which are related to epigenetic regulation, were also frequent in ENKL

Mutations in *MLL2* were the second most common mutation (7/37 cases), and all except one were nonsense SNVs. *MLL2* is also known as *KMT2D* and encodes a lysine-specific methyltransferase that methylates the Lys-4 position of histone H3. In addition, *BCOR* mutations occurred in 7 cases, about two-thirds of which were LOF mutations in exon 4 (1 frameshift deletion and 3 nonsense SNVs). Specific histone deacetylases (*HDAC1*, *HDAC3*, and *HDACB/5*) have been reported to interact with *BCOR* (43). Both *MLL2* and *BCOR* are classified into the same GO group, "chromatin modification" (GO:0016568) (44); cases with mutated genes belonging to this category accounted for a total of 45.9% of cancers (17/37 cases) (Figure 1). Considering the mutual exclusivity between mutations in *MLL2* and *BCOR* and those of *STAT3* (Figure 1), epigenetic dysregulation might be another important feature of ENKL pathogenesis, together with the alterations of the JAK/STAT cascade.

RNA-Seq revealed novel fusion genes and inactivation of *BCOR* as driver candidates

After the stepwise filtration of ambiguous calls from deFuse (described in the Methods), we selected 13 fusion candidates from cancer samples (Table 6). These included PDE4DIP (PARP8–PDE4DIP in CC2), which was reported previously as fusion gene partners in hematologic malignancy [19]. We discovered an in-frame candidate among fusion genes, SND1–LEP of CC1, which was validated by Sanger sequencing. Even though its function in carcinogenesis remains unclear, the coverage depth pattern around breakpoints was distinctive from those of other samples. LEP is a member of the JAK/STAT signaling pathway and the fusion conserves most of the leptin domain. Figure 3A shows SND1-LEP fusion, which were called again in TopHat-Fusion, and Figure 3B and C shows the result of validation by RT-PCR and Sanger sequencing.

In addition to the fusion candidates mentioned above, we found BCOR inactivation in CC2, which was called in deFuse but predicted to be adjacent, alternative splicing, and deletion. CC2 is an NK-cell lymphoma cell line that was established from the circulating malignant cells of a 19-year-old female diagnosed with ENKL. The coverage pattern of WES also supported this mutation (Figure 4); it was shown to be a homozygous deletion, although BCOR is located on X chromosome and this was a female sample.

We screened the expression levels of the TSGs that were selected from the cancer gene census, and found that 2 genes, BCOR and SH2B3, showed near-

Table 6. Fusion gene candidates

Chr1	Chr2	Gene name1	Gene name2	Gene location1	Gene location2	Genomic break position1	Genomic break position2	Sample
7	3	*AC005682.5*	*MAP4*	intron	intron	22896632	47992191	CT2
1	2	*HFM1*	*AC064836.3*	intron	upstream	91853017	203211035	CT1
17	9	*SEC14L1*	*AL161626.1*	intron	upstream	75158305	79186654	CT5
16	16	*C16orf59*	*CCNF*	coding	utr3p	2512476	2508158	CC2
17	17	*CCDC144NL*	*UBBP4*	intron	upstream	20771998	21709712	CC2
7	2	*CTB-111H14.1*	*CIR1*	intron	coding	106131456	175216453	CT2
22	22	*IGLL5*	*IGLV1-44*	coding	coding	23235960	22735712	CT5
22	22	*IGLL5*	*IGLV1-51*	coding	coding	23235964	22677325	CT5
19	4	*ZNF543*	*LDB2*	utr3p	intron	57841135	16749626	CT1
7	7	*SND1*	*LEP*	coding	coding	127292505	127894457	CC1
5	1	*PARP8*	*PDE4DIP*	intron	coding	49977280	144952200	CC2
17	4	*SEC14L1*	*PPEF2*	intron	intron	75158374	76807301	CT1
5	3	*SERINC5*	*TRIM42*	utr3p	downstream	79409395	140476573	CC2

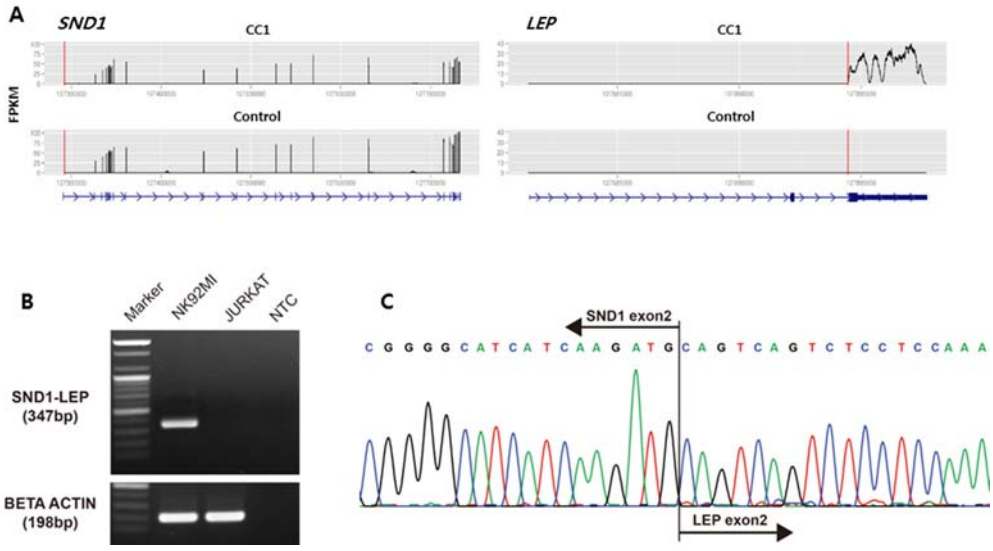

Figure 3. In-frame fusion gene candidate in ENKL. (A) Coverage pattern of the SND1–LEP fusion found in CC1, which was compared with the mean coverage of NC samples (Control). We assessed their expression levels in FPKM, and the predicted breakpoints are presented as red vertical lines. Their expression patterns around breakpoints were quite different from those of normal controls. FPKM, fragments per kilobase of transcript per million fragments mapped. (B) RT-PCR detection of 347bp of SND1-LEP transcripts using primer set of SND1 F and LEP R. A specific band was amplified in NK92MI(CC1) cell line. JURKAT cell line is a negative control. NTC(No Template Control) is RT-PCR products, which lack temeplate. (C) The sequencing analysis of the SND1-LEP revealed that SND1 exon2 was fused in-frame with LEP exon2.

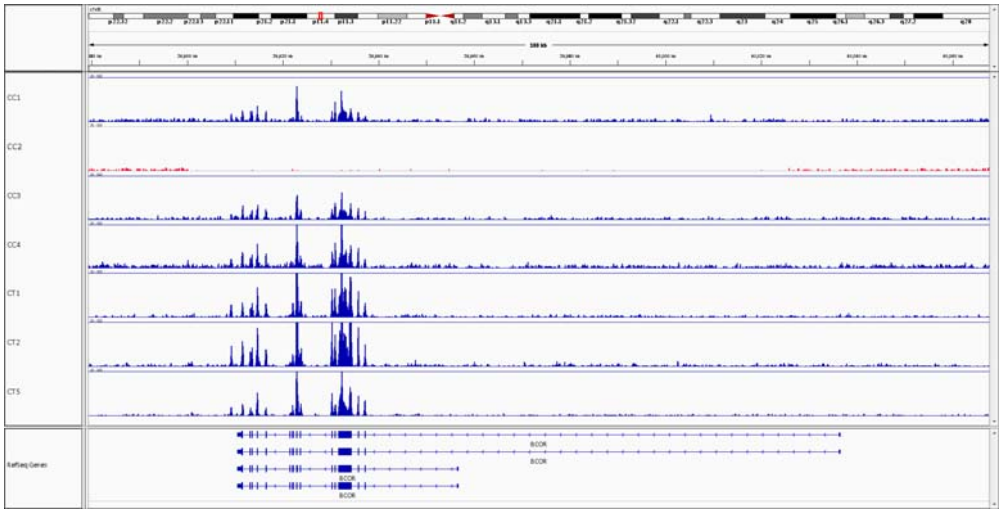

Figure 4. **Coverage patterns of *BCOR* in WES.** This figure displays the read depth of *BCOR* region in WES samples. CC2 was predicted to have *BCOR* deletion in the RNA-Seq analysis, and now it shows the very low read depth throughout the region in WES. It suggests that CC2 has the homozygous deletion of *BCOR* region.

complete suppression (fragments per kilobase of transcript, per million fragments sequenced (FPKM) < 0.1), whereas every NC sample had FPKM < 1 (Figure 5A and B). As previous studies reported that SH2B3 provides negative feedback on the JAK/STAT pathway [20], its expressional suppression might also affect the development of ENKL.

The gene expression profiles of both ENKL tissues and cell line samples reflect JAK/STAT cascade dysregulation and epigenetic alteration

We conducted sample clustering analyses to compare the overall gene expression profiles of each gene, and found that samples were well clustered by their sample groups, CT, CC, and NC. However, CT was clustered together with NC, and not CC, even though they both represent ENKL cancer samples (Figure 6A). One previous report listed differentially expressed genes (DEGs) of ENKL using array-based methods (15). We compared their $\log_2$(fold change) values with those found here for CT and CC samples, respectively (Figure 6B and C). Among DEGs that satisfied $|\log_2(\text{ratio})| \geq 1$ in both studies, 86.63% (175 out of 202) had the same direction in case of CT ($r^2 = 0.32$), whereas only 68.84% (148 out of 215) had changes with the same direction in CC samples ($r^2 = 0.11$); this pattern was concordant with the results of clustering, showing that CC samples were distant from CT. Upregulated genes tended to be well correlated with CT samples (CT, 150 out of 155; CC, 86 out of 148), whereas downregulated genes were correlated with CC samples (CT, 25 out of 47; CC, 62 out of 67).

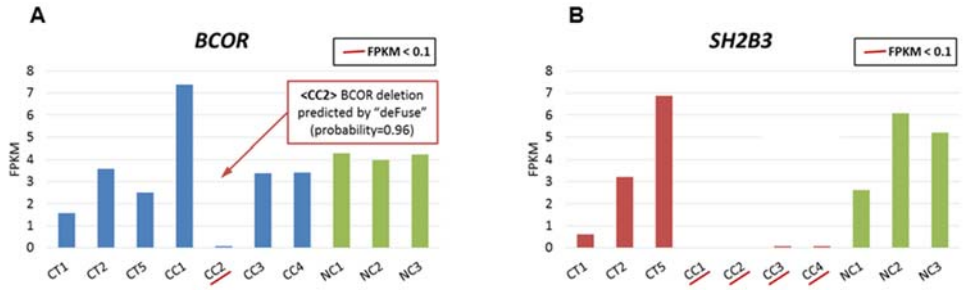

Figure 5. Inactivation of TSGs in ENKL. (A) Near-complete suppression of BCOR in CC2. The expression of BCOR was FPKM < 0.1 in CC2, and deFuse predicted the deletion of BCOR in this sample. (B) Near-complete suppression of SH2B3 in multiple cancer samples. All CC samples had SH2B3 expression < 0.1 in FPKM. Downregulation of this gene was also shown in CT1.

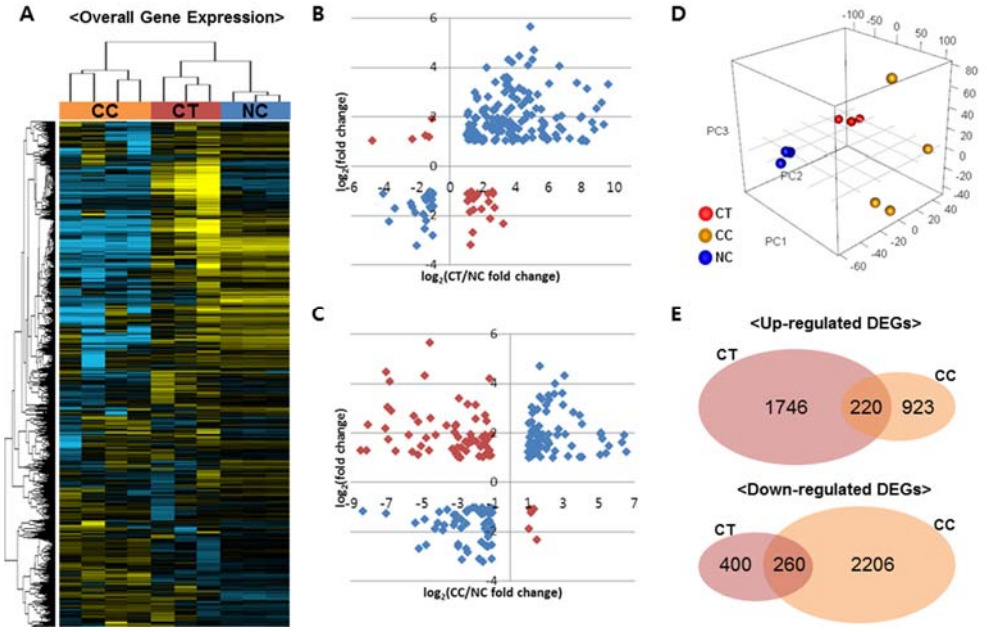

Figure 6. Gene expression profiles of cancer tissues and cancer cell line samples.

(A) Hierarchical clustering resulted in a good categorization of study samples according to their cellular identity. CT, cancer tissue; CC, cancer cell line; and NC, normal cell line. (B) The expression levels of DEGs reported by a previous study were consistent with those of CT samples. (C) However, CC samples differed largely regarding upregulated genes, among which approximately half were downregulated in CC. Here, blue squares represent genes showing concordant expression patterns between the two, while red ones indicate discordant gene expressions. (D) Principal component analysis separating study samples. PC2 differentiated cancer samples from normal samples. (E) Comparison of CT and CC samples regarding DEGs. There were some shared components in gene expression between CT and CC. DEG, differentially expressed gene.

In the principal component analysis, CC was also separated from the other groups in the PC1 axis. However, in PC2, CT and CC were clustered together and were located apart from NC. Despite their huge difference, CT and CC samples shared 220 upregulated and 260 downregulated DEGs compared with NC samples (Figure 6D and E, Figure 7A). Upregulated DEGs were enriched in the GOs (biological process) regulating phosphatidylinositol 3-kinase (PI3K) activity, which include the JAK/STAT cascade, whereas downregulated DEGs were among the GOs (immune system process) related to T-cell immunity (Figure 7B, Table 7). These results suggest that gene expression patterns also support major roles for the PI3K or JAK/STAT pathways in ENKL pathogenesis. Among the Kyoto Encyclopedia of Genes and Genomes (KEGG) gene sets, several pathways related to cancer were ranked in the top 10, in addition to "JAK/STAT signaling pathway" and "mTOR signaling pathway," which is located downstream of the JAK/STAT cascade (Figure 7C). We also found that several CGP gene sets related to the polycomb complex PRC2, which is involved in H3K27 trimethylation, were enriched among these upregulated genes (including the *Suz12* and *Eed* target genes). These findings support the contention that epigenetic dysregulation might be a key molecular factor in ENKL pathogenesis, as suggested in the mutation section (Figure 7D).

Table 7. Gene ontologies enriched with the DEGs common or different in CT and CC samples

GO ID	GO term	# of genes	% of associated genes	P-value (Bonferroni corrected)
Commonly up-regulated in CT and CC				
GO:0048706	embryonic skeletal system development	9	7.03	1.32E-04
GO:0010801	negative regulation of peptidyl-threonine phosphorylation	3	33.33	3.64E-03
GO:0060711	labyrinthine layer development	5	9.62	6.08E-03
GO:0003714	transcription corepressor activity	9	4.29	6.61E-03
GO:0043551	regulation of phosphatidylinositol 3-kinase activity	4	12.50	1.10E-02
GO:0014066	regulation of phosphatidylinositol 3-kinase signaling	5	7.14	2.23E-02
GO:0048704	embryonic skeletal system morphogenesis	6	6.32	1.21E-02
GO:0043524	negative regulation of neuron apoptotic process	7	5.15	1.25E-02
GO:2001243	negative regulation of intrinsic apoptotic signaling pathway	3	4.00	2.82E-02
GO:0051101	regulation of DNA binding	3	4.00	2.82E-02
GO:0042523	positive regulation of tyrosine phosphorylation of Stat5 protein	3	16.67	2.89E-02
GO:0060021	palate development	5	6.10	4.47E-02
Commonly down-regulated in CT and CC				
GO:0002562	somatic diversification of immune receptors via germline recombination within a single locus	6	10.34	4.28E-04
GO:0046006	regulation of activated T cell proliferation	4	12.90	6.08E-03

GO:0030217	T cell differentiation	8	4.32	6.34E-03
GO:1901215	negative regulation of neuron death	7	4.83	8.04E-03
GO:0010770	positive regulation of cell morphogenesis involved in differentiation	4	8.51	2.47E-02
GO:0048546	digestive tract morphogenesis	4	7.41	3.87E-02
GO:0043604	amide biosynthetic process	4	7.27	3.89E-02
GO:0060993	kidney morphogenesis	4	6.90	4.55E-02
Up-regulated only in CT				
GO:0001944	vasculature development	27	4.58	1.33E-10
GO:0003158	endothelium development	10	12.99	1.26E-07
GO:0032835	glomerulus development	6	9.84	1.40E-03
GO:0034329	cell junction assembly	10	4.78	1.40E-03
GO:0031589	cell-substrate adhesion	11	4.21	1.70E-03
GO:0030165	PDZ domain binding	7	7.22	2.02E-03
GO:0040013	negative regulation of locomotion	9	4.86	3.21E-03
GO:0007626	locomotory behavior	9	4.69	4.20E-03
GO:0045444	fat cell differentiation	9	5.45	1.37E-03
GO:0070328	triglyceride homeostasis	4	13.33	9.94E-03
GO:0032103	positive regulation of response to external stimulus	9	4.62	4.58E-03
GO:0019838	growth factor binding	7	5.74	7.82E-03
GO:0022617	extracellular matrix disassembly	7	5.47	1.00E-02

GO:0030204	chondroitin sulfate metabolic process	5	8.47	1.23E-02
GO:0030857	negative regulation of epithelial cell differentiation	4	12.50	1.24E-02
GO:0032963	collagen metabolic process	6	5.17	3.32E-02
GO:0050709	negative regulation of protein secretion	4	8.89	3.82E-02
GO:0050729	positive regulation of inflammatory response	5	6.17	4.18E-02
GO:0060041	retina development in camera-type eye	6	4.84	4.34E-02
GO:0010811	positive regulation of cell-substrate adhesion	5	5.95	4.68E-02
GO:0030193	regulation of blood coagulation	4	5.19	1.33E-01

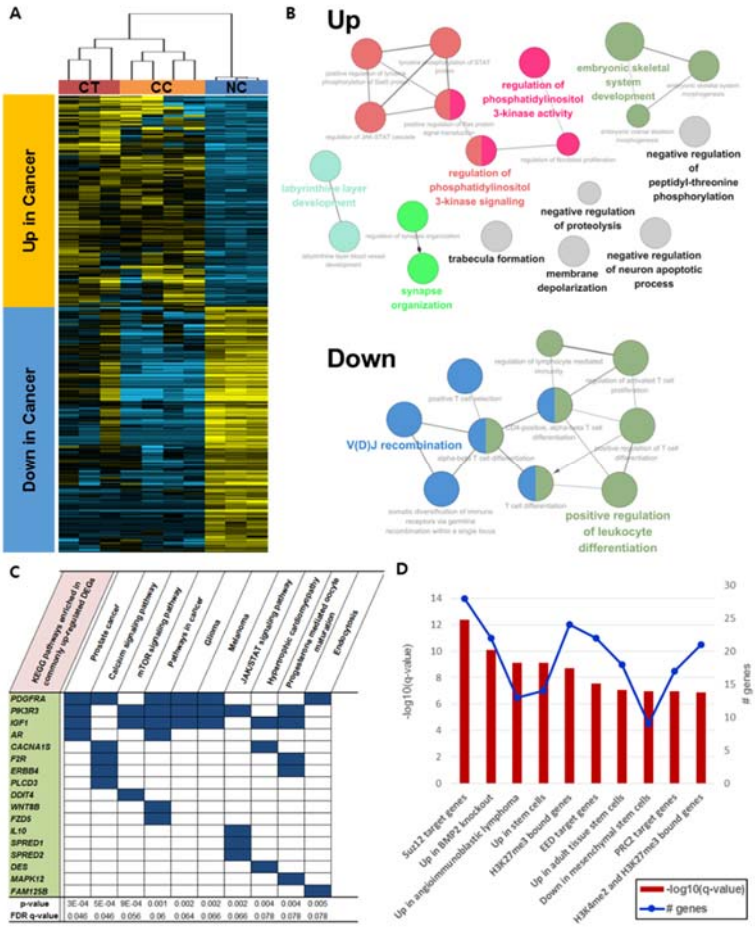

Figure 7. Functional enrichment of commonly upregulated or downregulated genes in both CT and CC samples. (A) Hierarchical clustering and heat map of common DEGs. (B) Gene ontologies enriched in up- or downregulated genes. (C) Top 10 KEGG pathways and (D) top 10 CGP gene sets enriched in upregulated genes.

Cancer tissue-specific DEGs are linked to the pathophysiologic features of ENKL

As there was a huge difference between CT and CC regarding gene expression, we selected genes that were upregulated only in CT, which might reflect the tumor microenvironment of ENKL. As shown in Figure 8A and 8B, many gene sets involved in blood coagulation and vasculature development were significantly enriched in these DEGs, which was consistent with the results of a previous study that reported the activation of angiogenesis-related genes in ENKL tissues (Table 7) (15). Enriched pathways such as “focal adhesion” and “ECM receptor interaction” also suggest that these genes are activated during the interaction between cancer and the adjacent environment. In particular, gene sets such as “pathways in cancer” were also among the most significant gene sets, and might promote the development of ENKL. Interestingly, cancer-related CGP gene sets were even more significant in CT-specific DEGs than were those of commonly upregulated ones, such as genes that are activated in angioimmunoblastic lymphoma (Figures 7D and 8C).

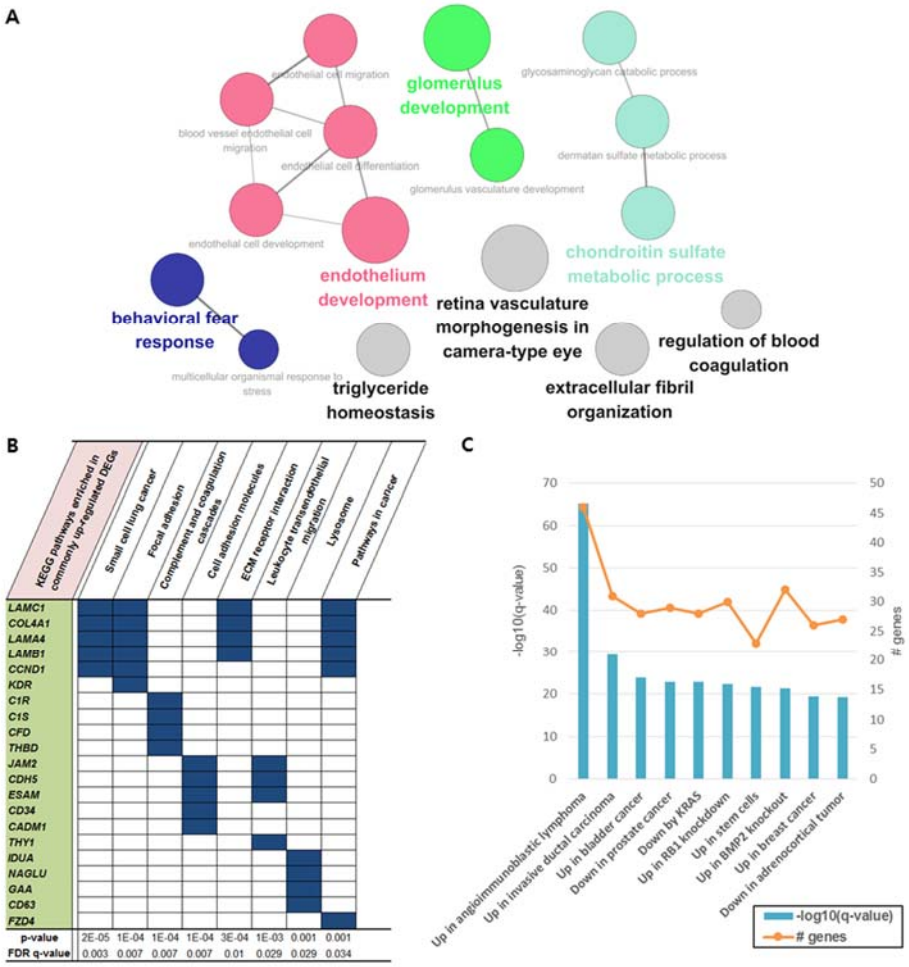

Figure 8. Functional enrichment of genes that were upregulated only in CT samples. (A) Gene ontologies enriched in CT-specific upregulated genes. There were several vasculature-development- or endothelium-development-related ontologies, which were the most significant. (B) Top 10 KEGG pathways and (C) top 10 CGP gene sets enriched in these genes.

DISCUSSION

In this study, the molecular background of ENKL was explored using genome and transcriptome sequencing, which enabled us to discover novel drivers and therapeutic targets. We designed the study to screen exonic regions of known cancer genes, and mutations were shown to be enriched in *STAT3*, *BCOR*, and *MLL2*. In terms of functional gene groups, the JAK/STAT cascade and histones modification-related gene covered 55.9% and 38.2% of samples, respectively. Gene expression analysis also supported these functional groups in ENKL pathogenesis; moreover, we reported the cancer tissue-specific alterations at the molecular level.

Previous studies have suggested that the JAK/STAT and MAPK signaling pathways are crucial for disease development (15, 16). In particular, frequent *JAK3*-activating mutations have shown that the JAK/STAT signaling pathway is a key molecular factor of ENKL (12). However, the mutation frequency of *JAK3* (p.Ala572Val and p.Ala573Val) was much lower in our cases compared with previous reports (5.1% vs 35.4%). In a recent study of French population (13), *JAK3* mutations encoding p.Ala573Val were observed in the 15.8% of ENKL samples (3/19). In that study, most of the cases expressed phosphorylated JAK3 regardless of their *JAK3* mutation status (100% of mutant cases and 84.2% of wild type cases). Such result was also supported by a study of Japanese population that reported frequent *JAK3* phosphorylation and low *JAK3* mutation (5.0%). Another Japanese group even showed that no *JAK3* mutations (p.Ala572Val or p.Ala573Val) was found in a total of 49 samples which included 17 ENKL cases (45). These results suggest that the

function of *JAK3* might be altered mainly due to phosphorylation than mutation, and molecular backgrounds of JAK3-activation in ENKL might be different according to ethnicity. Instead of *JAK3* mutations, we identified several alterations in other components of the JAK/STAT pathway. About a quarter (26.5%) of our ENKL cases had nsSNVs in *STAT3*, which were clustered on the SH2 domain. Its pro-oncogenic role has been demonstrated by stabilizing via the formation of dimers with other STAT proteins through reciprocal SH2 domain interactions (46). Frequent mutations in this domain can be presumed to be gain-of-oncogenic function mutations; however, additional studies are needed to confirm these mutations and their functional effects.

STAT3 can be activated by extrinsic ways as well as by intrinsic ways. EBV, the well-known pathogen of ENKL, is one of the extrinsic sources of activation of *STAT3*; the latent membrane protein 1 of EBV was suggested as a STAT-activating agent in a mouse model (47). Another study demonstrated that the Epstein–Barr nuclear antigen 2 also was a coactivator for the transcriptional enhancer of *STAT3* (48). Our results can be partly interpreted as intrinsic components of *STAT3* activation. We found a novel fusion gene of *LEP* in 1 sample, the protein product of which (leptin) plays a role at the top of the JAK/STAT pathway as a stimulator of *STAT3* phosphorylation (49). Loss of *SH2B3* has also been reported to increase *STAT3* phosphorylation (50), the expression of which was nearly completely suppressed in half of the cases (Figure 5B). In addition to the inactivation of *SH2B3*, we found that *PIK3R3* was upregulated in all ENKL samples; this gene is also known as a downstream element of the JAK/STAT cascade and leads to cell proliferation and

survival (KEGG pathways). Together with *STAT3* mutations, or independently, these alterations may play roles in ENKL pathogenesis.

Epigenetic dysregulation is an emerging part of cancer genomics, which is attributed to new genes discovered by NGS technology. The protein encoded by *MLL2* is a histone methyltransferase that targets the H3K4 site, and LOF mutation of *MLL2* has been reported in hematologic malignancy. In one recent study that was performed using NGS, B cell lymphomas had frequent somatic mutations in *MLL2*, most of which (83.3%) were LOF mutations, similar to ours (85.7%) (51). Some types of solid tumors also had *MLL2* mutations, although those exhibited a somewhat lower proportion of LOF (11.8% of renal cell carcinoma (52) and 66.7% of medulloblastoma (53) cases). The effect of altered MLL2 might vary according to tumor type.

BCOR was first reported as a corepressor of BCL-6 that is critical for germinal center formation (43), and its role in hematologic malignancy has been studied previously. Interestingly, most *BCOR* alterations occurred in male ENKL cases (6 mutations in men and 1 mutation in women), although this association was not statistically significant (27.3% of men vs 8.3% of women, $P = 0.378$). We tried to check whether mutations in *BCOR* in the X chromosome are related to the male predominance. However, *BCOR* mutation rates of male patients were lower in other male predominat malignancies (myelodysplastic syndrome (MDS), 3.8% of men vs 6.5% of women (54); acute myeloid leukemia (AML), 5.5% of men vs 6.9% of women (55); EBV-positive gastric cancer (EBV-GC), 14.2% of men vs 20% of women (56)). Next, we reviewed several public data to check whether BCOR

mutations are associated with EBV infection-related malignancies. EBV-GCs, one of 4 molecular subtypes of gastric cancer, had frequent *BCOR* mutations (3 nonsense and 1 splicing) compared with the other 3 types (15.4% vs 4.8%, $P = 0.051$) (56). With reference to the cBioPortal for Cancer Genomics (http://www.cbioportal.org), we selected several other solid tumors that are among the top 10 *BCOR*-mutated cancers and have available supplementary data sets (uterine endometrial carcinoma (UECA) (57), lung adenocarcinoma (LUAD) (58), melanoma (59), and colorectal adenocarcinoma (CRC) (60)), along with those of 2 hematologic malignancies, AML (55) and MDS (54). Figure 9 shows the *BCOR* mutation rate and the proportion of LOF mutations for each cancer. EBV-associated malignancies (ENKL and EBV-GC) showed high percentages for both properties. UECA, which is not associated with EBV infection, exhibited a rather high *BCOR* mutation rate, but a low LOF proportion in *BCOR*, similar to LUAD and melanoma. The hematologic malignancies showed opposite patterns to those observed in UECA. In summary, EBV infection might be related to high mutation rates of *BCOR*, and LOF mutations seem to account for their majority.

As summarized above, *BCOR* mutation tended to occur significantly more frequently in EBV-associated malignancy (odds ratio = 3.74, $P = 0.001$) and mostly in the form of LOF. Previously, the oncogenic effect of EBV was explained by various epigenetic regulation mechanisms (61). Considering that *BCOR* regulates chromatin modification by interacting with some histone-deacetylase-family genes, we were able to link the 2 major features, EBV infection and *BCOR* mutation, via an epigenetic mechanism. Overall, the mutations found in *MLL2* and *BCOR* led to the association

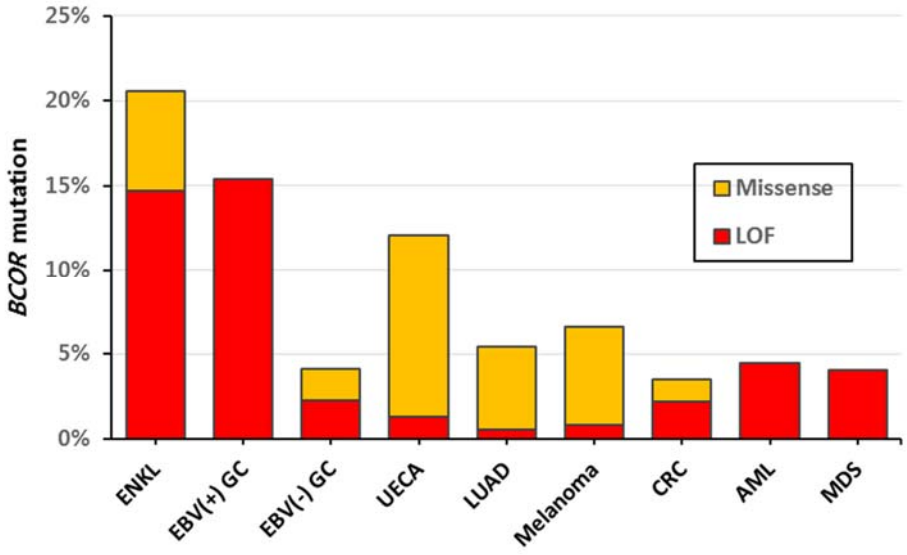

Figure 9. Mutation rates of BCOR according to tumor type. EBV-associated malignancies (ENKL and EBV(+) GC) showed high percentages for both BCOR mutation rate and LOF proportion. LOF, loss-of-function; GC, gastric carcinoma; UECA, uterine endometrial carcinoma; LUAD, lung adenocarcinoma; CRC, colorectal adenocarcinoma; AML, acute myeloid leukemia; MDS, myelodysplastic syndrome.

between epigenetic dysregulation and ENKL pathogenesis. To the best of our knowledge, the present study was the first to identify mutations in chromatin-modifying genes in this disease. Together with the mutational features, RNA-Seq data also support this finding; upregulated genes were enriched in PRC2-related gene sets in gene expression analysis.

Recurrent *DDX3X* mutations were identified in Chinese population by mutation screening using WES and confirmed in targeted sequencing of extended validation group (14). *DDX3X* encodes RNA helicase and play a role in cell cycle regulation (62). In the functional study of the report, DDX3X mutants showed decreased RNA-unwinding activity, which resulted in unregulated cell-cycle progression in NK cells. We checked mutation status of *DDX3X* in four tumor-normal paired cases to identify confident somatic mutations of the gene. In the default results from a mutation caller MuTect, none of four cases had *DDX3X* mutation. However, in the results including mutations failed to be accepted as somatic in MuTect , one of four had homozygous *DDX3X* missense mutation. MuTect rejected the result because of normal pair had alternate allele with low frequency, which could not pass the criteria for somatic mutation. When we selected candidates for targeted sequencing, genes with ≥ 2 somatic variants in 4 tumor-normal WES pairs were included, *DDX3X* mutations were not confirmed in extended validation group. Chinese group first screened 25 cases using WES and identified DDX3X mutations in 24% (6/25) of all. The mutation frequency are silimilar as ours (25%, 1/4). Small sample size to identify recurrent mutations was one of limitations of our study.

Until now, cell line samples have been used broadly in most types of cancer studies. For gene expression profiling, we designed a study to include both CT and CC samples, and tried to discover the common and different points between them. We found that CT and CC samples were quite different at the level of gene expression for both up and downregulated genes. The unsupervised clustering revealed even closer distance between CT and NC than that of CT and CC. As we selected study subjects which satisfied sufficient tumor cell contents, the difference could not be explained by normal cell contamination of tumor tissues. Due to the accumulated genetic alterations over division, the cancer cell lines may exhibit more prominent expression pattern far from primary tissue. Since we used the genes showing most variable expression patterns throughout samples for clustering analysis, these alterations specific for CC might affect greatly to the results. Although only some parts of DEGs were shared between the two groups, it seems meaningful that the DEGs included genes of JAK/STAT pathway or chromatin modification, supporting our main findings derived from mutational profiles.

Regarding genes that were upregulated only in CT, GO analysis showed that they were enriched in gene sets associated with angiogenesis, which is consistent with a previous study (15). One possible explanation is that this might be a compensation for the vascular destruction caused by the angiocentric and angiodestructive growth of ENKL (63). In addition, the pathways that are enriched among these genes include “focal adhesion” and “ECM receptor interaction,” which include genes that mostly overlapped with those of “pathways in cancer.” As these types of pathways are associated with aggressive features of cancer (64), we suggest that the CT-specific

changes might be linked to the characteristics of ENKL. These findings would have been missed if we analyzed cell line samples only, which do not include the effect of interaction between cancer and adjacent tissues, such as tumor microenvironment.

Here, we proposed several molecular candidates that can be applied as new therapeutic approaches using targeted agents. There seems the ethnic specificity of genetic backgrounds of ENKL, and more evaluations are required for these candidates in Asian populations. In addition to the validation of these candidates, it is necessary to subclassify ENKL patients according to their molecular backgrounds and check whether clinical features differ according to these groups, which will be feasible if a large sample size is used. We believe that the findings of this study will contribute substantially to the design of future studies of ENKL and of personalized approaches for cancer patients.

REFERENCES

1. Lee S, Park HY, Kang SY, Kim SJ, Hwang J, Lee S, et al. Genetic alterations of JAK/STAT cascade and histone modification in extranodal NK/T-cell lymphoma nasal type. Oncotarget. 2015;6(19):17764-76.
2. Au WY, Weisenburger DD, Intragumtornchai T, Nakamura S, Kim WS, Sng I, et al. Clinical differences between nasal and extranasal natural killer/T-cell lymphoma: a study of 136 cases from the International Peripheral T-Cell Lymphoma Project. Blood. 2009;113(17):3931-7.
3. Lee J, Kim WS, Park YH, Park SH, Park KW, Kang JH, et al. Nasal-type NK/T cell lymphoma: clinical features and treatment outcome. Br J Cancer. 2005;92(7):1226-30.
4. Lee J, Suh C, Park YH, Ko YH, Bang SM, Lee JH, et al. Extranodal natural killer T-cell lymphoma, nasal-type: a prognostic model from a retrospective multicenter study. J Clin Oncol. 2006;24(4):612-8.
5. Suzuki R, Suzumiya J, Yamaguchi M, Nakamura S, Kameoka J, Kojima H, et al. Prognostic factors for mature natural killer (NK) cell neoplasms: aggressive NK cell leukemia and extranodal NK cell lymphoma, nasal type. Ann Oncol. 2010;21(5):1032-40.
6. Yamaguchi M, Kwong YL, Kim WS, Maeda Y, Hashimoto C, Suh C, et al. Phase II study of SMILE chemotherapy for newly diagnosed stage IV, relapsed, or refractory extranodal natural killer (NK)/T-cell lymphoma, nasal type: the NK-Cell Tumor Study Group study. J Clin Oncol. 2011;29(33):4410-6.

7. Jaccard A, Gachard N, Marin B, Rogez S, Audrain M, Suarez F, et al. Efficacy of L-asparaginase with methotrexate and dexamethasone (AspaMetDex regimen) in patients with refractory or relapsing extranodal NK/T-cell lymphoma, a phase 2 study. Blood. 2011;117(6):1834-9.

8. Kanavaros P, Briere J, Emile JF, Gaulard P. Epstein-Barr virus in T and natural killer (NK) cell non-Hodgkin's lymphomas. Leukemia. 1996;10 Suppl 2:s84-7.

9. Aozasa K, Takakuwa T, Hongyo T, Yang WI. Nasal NK/T-cell lymphoma: epidemiology and pathogenesis. Int J Hematol. 2008;87(2):110-7.

10. Nakashima Y, Tagawa H, Suzuki R, Karnan S, Karube K, Ohshima K, et al. Genome-wide array-based comparative genomic hybridization of natural killer cell lymphoma/leukemia: different genomic alteration patterns of aggressive NK-cell leukemia and extranodal Nk/T-cell lymphoma, nasal type. Genes Chromosomes Cancer. 2005;44(3):247-55.

11. Karube K, Nakagawa M, Tsuzuki S, Takeuchi I, Honma K, Nakashima Y, et al. Identification of FOXO3 and PRDM1 as tumor-suppressor gene candidates in NK-cell neoplasms by genomic and functional analyses. Blood. 2011;118(12):3195-204.

12. Koo GC, Tan SY, Tang T, Poon SL, Allen GE, Tan L, et al. Janus kinase 3-activating mutations identified in natural killer/T-cell lymphoma. Cancer discovery. 2012;2(7):591-7.

13. Bouchekioua A, Scourzic L, de Wever O, Zhang Y, Cervera P, Aline-Fardin A, et al. JAK3 deregulation by activating mutations confers invasive growth

advantage in extranodal nasal-type natural killer cell lymphoma. Leukemia. 2014;28(2):338-48.

14. Jiang L, Gu ZH, Yan ZX, Zhao X, Xie YY, Zhang ZG, et al. Exome sequencing identifies somatic mutations of DDX3X in natural killer/T-cell lymphoma. Nat Genet. 2015;47(9):1061-6.

15. Huang Y, de Reynies A, de Leval L, Ghazi B, Martin-Garcia N, Travert M, et al. Gene expression profiling identifies emerging oncogenic pathways operating in extranodal NK/T-cell lymphoma, nasal type. Blood. 2010;115(6):1226-37.

16. Ng SB, Yan J, Huang G, Selvarajan V, Tay JL, Lin B, et al. Dysregulated microRNAs affect pathways and targets of biologic relevance in nasal-type natural killer/T-cell lymphoma. Blood. 2011;118(18):4919-29.

17. Ju YS, Lee WC, Shin JY, Lee S, Bleazard T, Won JK, et al. A transforming KIF5B and RET gene fusion in lung adenocarcinoma revealed from whole-genome and transcriptome sequencing. Genome Res. 2012;22(3):436-45.

18. Huang Y, de Leval L, Gaulard P. Molecular underpinning of extranodal NK/T-cell lymphoma. Best practice & research Clinical haematology. 2013;26(1):57-74.

19. Vogelstein B, Papadopoulos N, Velculescu VE, Zhou S, Diaz LA, Jr., Kinzler KW. Cancer genome landscapes. Science. 2013;339(6127):1546-58.

20. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nature methods. 2012;9(4):357-9.

21. Perez-Llamas C, Lopez-Bigas N. Gitools: analysis and visualisation of

genomic data using interactive heat-maps. PloS one. 2011;6(5):e19541.

22. Engstrom PG, Steijger T, Sipos B, Grant GR, Kahles A, Consortium R, et al. Systematic evaluation of spliced alignment programs for RNA-seq data. Nature methods. 2013;10(12):1185-91.

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

24. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome research. 2010;20(9):1297-303.

25. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nature biotechnology. 2013;31(3):213-9.

26. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010;38(16):e164.

27. Abecasis GR, Auton A, Brooks LD, DePristo MA, Durbin RM, Handsaker RE, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012;491(7422):56-65.

28. Drmanac R, Sparks AB, Callow MJ, Halpern AL, Burns NL, Kermani BG, et al. Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. Science. 2010;327(5961):78-81.

29. Lawrence MS, Stojanov P, Polak P, Kryukov GV, Cibulskis K, Sivachenko A,

et al. Mutational heterogeneity in cancer and the search for new cancer-associated genes. Nature. 2013;499(7457):214-8.

30. McPherson A, Hormozdiari F, Zayed A, Giuliany R, Ha G, Sun MG, et al. deFuse: an algorithm for gene fusion discovery in tumor RNA-Seq data. PLoS Comput Biol. 2011;7(5):e1001138.

31. Steidl C, Shah SP, Woolcock BW, Rui L, Kawahara M, Farinha P, et al. MHC class II transactivator CIITA is a recurrent gene fusion partner in lymphoid cancers. Nature. 2011;471(7338):377-81.

32. Kangaspeska S, Hultsch S, Edgren H, Nicorici D, Murumagi A, Kallioniemi O. Reanalysis of RNA-sequencing data reveals several additional fusion genes with multiple isoforms. PLoS One. 2012;7(10):e48745.

33. Kim D, Salzberg SL. TopHat-Fusion: an algorithm for discovery of novel fusion transcripts. Genome Biol. 2011;12(8):R72.

34. Forbes SA, Bindal N, Bamford S, Cole C, Kok CY, Beare D, et al. COSMIC: mining complete cancer genomes in the Catalogue of Somatic Mutations in Cancer. Nucleic acids research. 2011;39(Database issue):D945-50.

35. Anders S, Pyl PT, Huber W. HTSeq — A Python framework to work with high-throughput sequencing data. bioRxiv. 2014.

36. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-Seq data with DESeq2. bioRxiv. 2014.

37. de Hoon MJ, Imoto S, Nolan J, Miyano S. Open source clustering software. Bioinformatics. 2004;20(9):1453-4.

38. Saldanha AJ. Java Treeview--extensible visualization of microarray data.

Bioinformatics. 2004;20(17):3246-8.

39. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A. 2005;102(43):15545-50.

40. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics. 2009;25(8):1091-3.

41. Saito R, Smoot ME, Ono K, Ruscheinski J, Wang PL, Lotia S, et al. A travel guide to Cytoscape plugins. Nature methods. 2012;9(11):1069-76.

42. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research. 2000;28(1):27-30.

43. Huynh KD, Fischle W, Verdin E, Bardwell VJ. BCoR, a novel corepressor involved in BCL-6 repression. Genes & development. 2000;14(14):1810-23.

44. Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, Cherry JM, et al. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics. 2000;25(1):25-9.

45. Kimura H, Karube K, Ito Y, Hirano K, Suzuki M, Iwata S, et al. Rare occurrence of JAK3 mutations in natural killer cell neoplasms in Japan. Leukemia & lymphoma. 2014;55(4):962-3.

46. Yu H, Pardoll D, Jove R. STATs in cancer inflammation and immunity: a leading role for STAT3. Nature reviews Cancer. 2009;9(11):798-809.

47. Shair KH, Bendt KM, Edwards RH, Bedford EC, Nielsen JN, Raab-Traub N. EBV latent membrane protein 1 activates Akt, NFkappaB, and Stat3 in B cell lymphomas. PLoS pathogens. 2007;3(11):e166.

48. Muromoto R, Ikeda O, Okabe K, Togi S, Kamitani S, Fujimuro M, et al. Epstein-Barr virus-derived EBNA2 regulates STAT3 activation. Biochemical and biophysical research communications. 2009;378(3):439-43.

49. Sanchez-Margalet V, Martin-Romero C. Human leptin signaling in human peripheral blood mononuclear cells: activation of the JAK-STAT pathway. Cellular immunology. 2001;211(1):30-6.

50. Perez-Garcia A, Ambesi-Impiombato A, Hadler M, Rigo I, LeDuc CA, Kelly K, et al. Genetic loss of SH2B3 in acute lymphoblastic leukemia. Blood. 2013;122(14):2425-32.

51. Morin RD, Mendez-Lago M, Mungall AJ, Goya R, Mungall KL, Corbett RD, et al. Frequent mutation of histone-modifying genes in non-Hodgkin lymphoma. Nature. 2011;476(7360):298-303.

52. Dalgliesh GL, Furge K, Greenman C, Chen L, Bignell G, Butler A, et al. Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature. 2010;463(7279):360-3.

53. Parsons DW, Li M, Zhang X, Jones S, Leary RJ, Lin JC, et al. The genetic landscape of the childhood cancer medulloblastoma. Science. 2011;331(6016):435-9.

54. Damm F, Chesnais V, Nagata Y, Yoshida K, Scourzic L, Okuno Y, et al. BCOR and BCORL1 mutations in myelodysplastic syndromes and related disorders.

Blood. 2013;122(18):3169-77.

55. Grossmann V, Tiacci E, Holmes AB, Kohlmann A, Martelli MP, Kern W, et al. Whole-exome sequencing identifies somatic mutations of BCOR in acute myeloid leukemia with normal karyotype. Blood. 2011;118(23):6153-63.

56. Cancer Genome Atlas Research N. Comprehensive molecular characterization of gastric adenocarcinoma. Nature. 2014;513(7517):202-9.

57. Cancer Genome Atlas Research N, Kandoth C, Schultz N, Cherniack AD, Akbani R, Liu Y, et al. Integrated genomic characterization of endometrial carcinoma. Nature. 2013;497(7447):67-73.

58. Imielinski M, Berger AH, Hammerman PS, Hernandez B, Pugh TJ, Hodis E, et al. Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. Cell. 2012;150(6):1107-20.

59. Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, et al. A landscape of driver mutations in melanoma. Cell. 2012;150(2):251-63.

60. Cancer Genome Atlas N. Comprehensive molecular characterization of human colon and rectal cancer. Nature. 2012;487(7407):330-7.

61. Tempera I, Lieberman PM. Epigenetic regulation of EBV persistence and oncogenesis. Seminars in cancer biology. 2014;26:22-9.

62. Soto-Rifo R, Ohlmann T. The role of the DEAD-box RNA helicase DDX3 in mRNA metabolism. Wiley Interdiscip Rev RNA. 2013;4(4):369-85.

63. Steven H. Swerdlow EC, Nancy Lee Harris, Elaine S.Jaffe, Stefano A. Pileri, Harald Stein, Jurgen Thiele, James W. Vardiman. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. 4th ed. International

Agency for Reasearch on Cancer: WHO PRESS; 2008.

64. Shan Z, Li G, Zhan Q, Li D. Gadd45a inhibits cell migration and invasion by altering the global RNA expression. Cancer biology & therapy. 2012;13(11):1112-22.

국문초록

NK/T 림프종의 유전변이

서울대학교 대학원 의과학과 의과학 전공

박 하 영

서론 : NK/T 림프종 (extranodal NK/T-cell lymphoma nasal type , ENKL) 은 드물게 발생하는 비호지킨 림프종의 일종이며, 주로 동아시아와 라틴아메리카에서 많이 발생한다. 발병에 관여하는 분자 유전학적 기전에 관해 여러 연구가 발표되었으나, 현재까지도 정확한 발병 원인은 밝히지 못했다.

방법 : ENKL 에서 특이적으로 발생하는 유전 변이를 찾기 위해 총 34명의 ENKL 환자 샘플에 대하여 다양한 차세대 시퀀싱 기법을 적용하였다. 이는 9명 환자에서 유래한 암종 조직과 4개의 암 세포주에 대한 엑솜 시퀀싱, 21 명의 환자에서 유래한 암종 조직에 대한 타겟 시퀀싱, 3명의 환자 유래 암종 조직과 4개의 세포주에 대한 전사체 시퀀싱을 포함한다.

결과 : *STAT3*, *BCOR*, *MLL2* 의 세 유전자에 가장 많은 유전변이가 있었고 각각 9, 7, 6 개의 샘플에서 발견되었다. 이전 연구에서 빈번한 변이가 발견된다고 알려진 유전자인 *JAK3* 에서는 2개의 유전변이만이 발견되었다. 전체적으로는 JAK/STAT 신호 전달 경로와 히스톤 변형에 관련된 유전자군에서 많은 변이가 발견되었고, 이는 각각 55.9% 와 38.2% 에 해당한다. 이러한 경향은 유전자 발현량 분석에서도 동일하게 관찰되었다. 또한 암종 조직에서 과발현된 177개의 유전자는 혈관중심 성장패턴을 보이면서 혈관파괴를 잘 일으키는 ENKL 의 병리학적 소견에 부합하는 결과였다.

결론 : 본 연구를 통해 ENKL 발생에 관여하는 새로운 유전 변이를 발견하였으며, 이는 향후 치료의 타겟으로 큰 역할을 할 수 있을 것으로 기대된다.

* 이 논문은 Oncotarget 에 출간되었다. (1)

주요어 : NK/T 림프종, 차세대 시퀀싱, JAK-STAT 신호 전달 경로, 염색질 변화, 체세포 변이, 전사체
학번 : 2013-30609