INVESTIGATION OF SPATIAL AND TEMPORAL HETEROGENEITY IN THE GENETIC BACKGROUND OF GERMINAL CENTER LYMPHOMAS

PhD thesis

Bence Bátai, MD

Doctoral School of Pathology

Semmelweis University

Supervisor: Csaba Bödör, DSc

Official reviewers: Zoltán Wiener, DSc
Ferenc Magyari, MD, PhD

Head of the Complex Examination Committee:
Judit Demeter, MD, DSc
Members of the Complex Examination Committee:
Janina Kulka, MD, DSc
Erika Tóth, MD, PhD

Budapest
2024

1. INTRODUCTION

B-cell non-Hodgkin lymphomas (NHL) are the most prevalent lymphoid malignancies in adult patients, with an estimated 20 new diagnoses per 100 000 inhabitants per year in developed countries. Understanding the heterogeneity of this disease group through molecular studies including the investigation of genetic, transcriptional and expression changes have improved current diagnostic classification, risk stratification and therapeutic decision making. Although the outcome of B-cell NHL-s has improved during the past two decades due to the approval of anti-CD20 antibodies, the development of relapsed/refractory disease, as well as tailoring treatment intensity to avoid unnecessary toxicity are unmet needs of clinical management in most subtypes, warranting further research to understand the biological determinants of disease progression and response to therapy.

This thesis focuses on the genetic background of a rare B-cell lymphoma subtype, primary cutaneous follicle center lymphoma (PCFCL) and the most prevalent indolent lymphoma subtype, follicular lymphoma (FL).

Recently developed methods opened the field of new sampling methods providing limited nucleic acid input and small tumor fractions like liquid-biopsies and broadened the scope of genetic studies with limited quality tumor tissue samples available like low-coverage whole genome sequencing (lcWGS). The new methods together with multi region sequencing and/or using concomitant cell-free DNA (cfDNA) samples representing the whole tumor burden provide possibility to deepen our knowledge about disease evolution and progression.

2. OBJECTIVES

During our research we aimed to:

1. Investigate the genetic background of PCFCL including gene mutations of *EZH2* and *TNFRSF14* and assess their association with clinicopathologic parameters, as well as structural genetic alterations
2. Analyze the genome-wide copy number profile of PCFCL using low-coverage whole-genome sequencing
3. Compare the copy number profile of PCFCL to FL to identify biomarkers facilitating differential diagnosis
4. Investigate the association of copy number burden metrics, as well as copy number alterations with clinicopathological parameters in PCFCL
5. Analyze evolutionary trajectories of PCFCL
6. Investigate genetic heterogeneity using paired tumor tissue DNA (tisDNA) and cfDNA samples in R/R FL
7. Analyze the role of spatial heterogeneity in disease evolution in R/R FL
8. Investigate the prognostic role of ctDNA levels in R/R FL

3. METHODS

3.1. Investigation of the genetic background of PCFCL

3.1.1. Patient selection

These studies were conducted in accordance with the Declaration of Helsinki and approved by the Ethics Committee of the Hungarian Medical Research Council (45371-2/2016/EKU and IV/5495-3/2021/EKU). As both studies were performed using similar methodology, we will discuss them together in the *Methods* referring to them as *first* and *second* study, where necessary.

We collected and reevaluated 28 tissue samples from 21 cases of PCFCL for the *first* study and 28 tissue samples from 20 patients for the second study, respectively.

For the *second* study nodal follicular lymphoma (FL) samples were also collected from 64 patients.

3.1.2. Mutation analysis

In the *first* study, Sanger sequencing of Y641 (exon 16), A682, and A692 (exon 18) hotspots of the *EZH2* gene and the whole coding sequence of the *TNFRSF14* gene were carried out using custom primers.

3.1.3. Low-coverage whole genome sequencing

For the second study, sequencing libraries were prepared using the NEBNext Ultra II DNA Library Prep Kit for Illumina with Unique Dual Index UMI Adaptors (New England Biolabs) and sequenced to a target yield of 10 M bases per library.

3.1.4. Analysis of low-coverage whole genome sequencing data

After primary analysis, relative copy number alterations were identified with QDNASeq (v.1.26.0) using a custom 500 kilobase bin size mappability file. Absolute copy numbers were determined using a custom R script.

3.2. Investigation of spatial genetic heterogeneity in R/R FL analyzing concurrent cell-free DNA - tissue DNA pairs

3.2.1. Patient and sample selection

We investigated 114 samples from 22 relapsed/refractory follicular lymphoma patients who had either paired liquid-biopsy and tumor biopsy sample available at the time of disease progression and/or paired tissue biopsies at diagnosis or relapse to determine intrapatient heterogeneity in follicular lymphoma.

3.2.2. Sequencing library preparation

Library preparation was performed using the SureSelectXT HS Reagent Kit (Agilent, USA). For target-enrichment, a custom 1.17 Mb SureSelectXT HS (Agilent, USA) panel was designed covering the whole coding region 173 genes, as well as clonotype determination were included. Libraries were sequenced on a NovaSeq6000 or NextSeq2000 instrument (Illumina, USA) to a median depth of 6695 (1872-16003) and 2435 (591-3676), respectively.

3.2.4. Processing of sequencing data

Primary analysis of sequencing data was performed to leverage single molecular barcode (MBC) information to retrieve the highest number of biologically different DNA molecules. Variants were called and filtered using GATK 4.1.7.0. All variants per patient were kept that met filtering criteria in at least one sample of the patient.

3.2.5. Phylogenetic reconstruction

SciClone was used to cluster variants, which were processed in ClonEvol to infer clonal relationship of clusters and create phylogenetic trees. All clonal models were investigated and curated manually to select the most likely model.

4. RESULTS

4.1. Investigation of the role of TNFRSF14 and EZH2 in the pathogenesis of PCFCL

We assessed the mutation status of *TNFRSF14* by Sanger sequencing in 17 PCFCL patients with available DNA samples. *TNFRSF14* mutations were detected in 4/17 (23.5%) patients with three nonsense mutations (c.35G > A; p.Trp12*) previously reported in FL and one previously unreported missense variant (c.157 T > G; p.Cys53Gly) targeting the extracellular domain. Two cases with *TNFRSF14* mutations harbored concomitant 1p36 deletions resulting in complete loss of functional *TNFRSF14*. Significant impact of mutations on clinical features and disease outcome was not observed. The presence of gain of function mutations affecting the *EZH2* gene were also assessed at three recurrent mutation hotspots (Y646, A682, and A692), and no mutations were detected in our PCFCL cohort.

4.2. Analysis of copy number alterations in PCFCL using low-coverage WGS

We performed low-coverage whole genome sequencing on 28 samples of 20 PCFCL patients. At least one CNA was detected in 26/28 (92.9%) samples with median 4 CNAs (IQR: 2-8) per sample. The most frequently identified copy number losses at the cytoband level were the deletions of the 19p12-q13.11 region (28.6%), 1p36.33-p36.23 and 6q24.1 regions (25.0%) as well as the deletions in the 6q16.1-q23.3 regions (21.4%). The most frequently identified copy number gains were spanning the 2p22.2-p15 (32.1%), 2p22.3 (28.6%), 12q13.13-q14.3 (25.0%) and 1q21.2-q25.3 (25.0%) regions (Figure 1.).

Amplifications were more often conferred by arm level changes, while deletions were often focal.

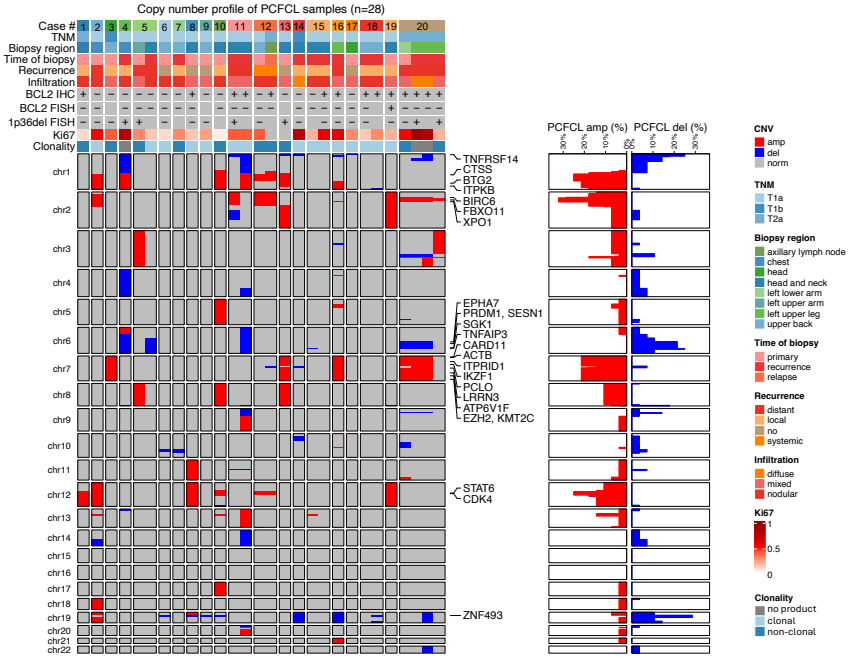

Figure 1. *Copy number profile of all primary cutaneous follicle center lymphoma (PCFCL) samples analyzed in the study (n = 28) displaying associated clinicopathological characteristics and frequency of copy number alterations (CNA). Genes in regions harboring a copy number alteration in at least 20% of the samples are displayed on the right side of the heatmap.*
Abbreviations: amp: amplification, del: deletion, FISH: fluorescence in situ hybridization, IHC: immunohistochemistry, norm: normal, TNM: tumor, node, metastasis stage of cutaneous lymphomas other than mycosis fungoides and Sezary syndrome.

Subsequently, we compared the copy number profiles of PCFCL samples to data generated from a matched cohort of diagnostic, pre-treatment nodal follicular lymphoma (FL) samples (n=64).

Our results indicated relatively similar copy number profiles of PCFCL and FL. Most frequent CNAs in both cohorts were deletions on 1p and 19p chromosome arms affecting *TNFRSF14* and *ZNF493*, respectively, and amplifications on 1q, 2p and 12p chromosome arms spanning the coding regions of *CTSS*, *BIRC6*, *FBXO11*, *XPO1*, *STAT6* and *CDK4*.

Interestingly, out of the most frequent alterations known in FL, amplifications on chromosome 18q peaking at the 18q21.33 cytoband and affecting the locus of the *BCL2* oncogene showed significant enrichment in FL samples (31.3% (20/64), while it was detected in only one PCFCL sample (5.0% (1/20), p=0.018, Fisher's exact test). Significant enrichment of amplifications in the 13q14.11-q14.2 region was observed in PCFCL (PCFCL: 15.0% (3/20) vs. FL: 1.6% (1/64), p=0.040, Fisher's exact test) spanning the coding region of *FOXO1*.

Analyzing all PCFCL samples, copy number burden was significantly higher in samples of patients' developing distant cutaneous spread both in terms of proportion of genome altered (localized: 0.02 (IQR: 0.01-0.12) vs. distant: 0.13 (IQR: 0.07-0.16), p=0.033) and number of CNAs (localized: 2 (IQR: 1-5) vs. distant: 9 (3-11), p=0.017) pointing to higher genomic instability in samples more readily involving distant disease sites (Figure 2.). Including all samples, 2p22.2-p15 and 3q23-q24 amplifications, as well as 6q16.1-q23.3 and 9p21.3 deletions were detected as potentially prognostic biomarkers of the disease course in PCFCL (Figure 2.).

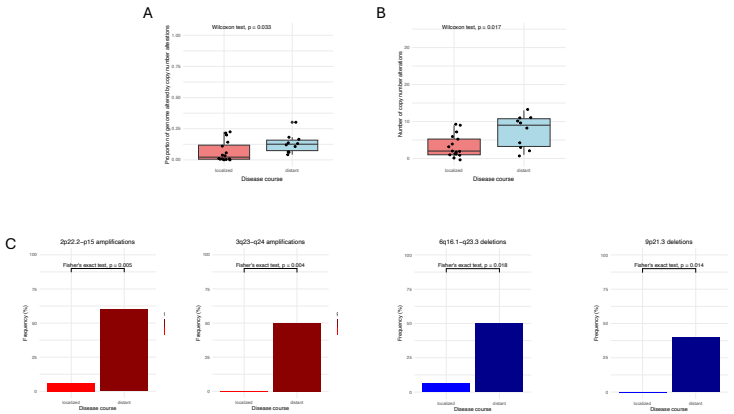

Figure 2. *Distribution of copy number burden metrics and alterations in distinctive regions between primary cutaneous follicle center lymphoma (PCFCL) patients showing localized disease course compared to patients developing distant cutaneous spread or systemic disease. A-B) Proportion of genome altered and the number of identified copy number alterations were significantly higher in patients with distant disease spread suggesting increased genomic instability. C) Including all samples, 2p22.2-p15 and 3q23-q24 amplifications, as well as 6q16.1-q23.3 and 9p21.3 deletions were detected as potentially prognostic biomarkers of the disease course in PCFCL.*

Comparing the copy number profiles of sequential tissue samples from 6 patients, overlapping alterations were observed in 3 patients during the disease course with additional emerging and diminishing alterations underlining the role of spatiotemporal heterogeneity in the pathogenesis of PCFCL (Figure 1.).

With regard to the clinical outcome, two patients with available sequential samples displayed an unfavorable clinical course, one with development of systemic spread (case #12), or distant skin recurrence and underlying soft tissue infiltration requiring chemoimmunotherapy (case #20), with both patients harbouring a 2p amplification in the earliest common progenitor highlighting its role in disease progression and pathogenesis.

4.3. Investigation of spatial genetic heterogeneity in R/R FL analyzing concurrent cell-free DNA – tissue DNA pairs

Molecular profiling of recurrently mutated genes, copy number alterations and immunoglobulin gene rearrangements were performed on 91 tisDNA or cfDNA samples of 22 R/R FL patients.

Non-silent single nucleotide variants and small indels were identified in 21 of 24 (87.5%) pretreatment cfDNA samples. In positive cfDNA samples median 7 variants (range: 1-34) were identified in 75 out of 173 genes in the sequencing panel previously shown to harbor mutations in FL. Considering 20 patients who had a pretreatment cfDNA sample at relapse 18/20 patients (90.0%) harbored at least one detectable alteration at relapse with three patients harboring mutant cfDNA samples before two treatment lines during the disease course.

Non-silent small variants were detected in 59 of 62 tumor tissue samples, with median 9 variants per sample (range: 1-32).

Considering all samples, the most frequent alterations were *KMT2D* (65%), *CREBBP* (54%), *TNFRSF14* (41%), *EZH2* (24%) and *EP300* (20%) mutations (Figure 3.). Cytosine to thymine transitions were the most frequent events both in

cfDNA and tisDNA samples, with missense variants dominating the mutation landscape (61.7%, 523/847 variants). Focusing on the genetic heterogeneity between pretreatment cfDNA samples and concurrent tumor tissue samples before the same treatment line, we compared variants in 20 tumor tissue samples of 15 patients to corresponding cfDNA samples. Performing tumor-informed filtering, retaining called variants in all samples meeting filtering quality metrics in at least one of the patient samples resulted in a variant overlap of median 63.6% (IQR: 28.2%-86.2%). Furthermore, using tumor informed filtering more unique variants were recovered in cfDNA (14.8%, IQR: 0.0%-33.1%) than tisDNA samples (9.8%, IQR: 0.0%-24.7%).

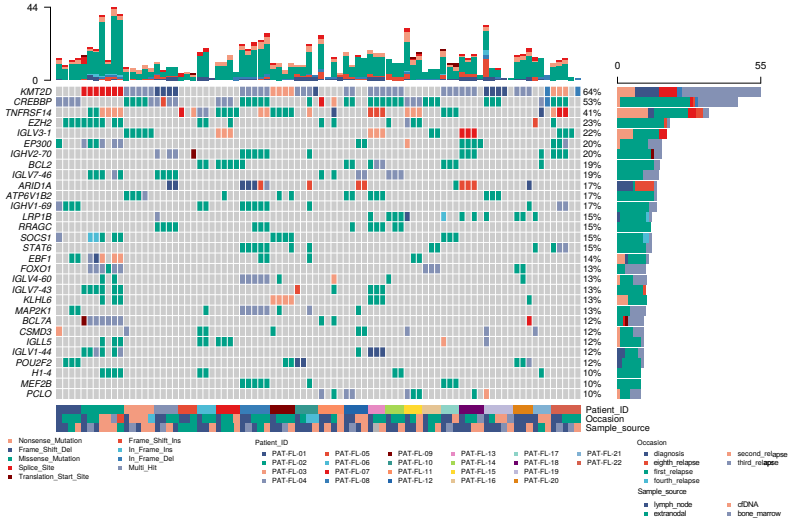

Figure 3. *Oncoplot of identified variants in all samples using tumor-informed filtering*

When comparing spatial compartmentalization of variants to tisDNA pairs from concomitant nodal, or extranodal-nodal pairs in 5 patients, we found that concordance of variants was comparable to cfDNA-tisDNA pairs (62.5%, IQR: 53.8%-72.9%), but higher than tisDNA pairs from concomitant nodal/extranodal and histologically infiltrated bone marrow biopsies (44.4%, IQR: 36.4%-66.7%).

To resolve subclonal dynamics, clonal inference was performed in patients harbouring sample triplicates including 2 tisDNA and a cfDNA from the same timepoint.

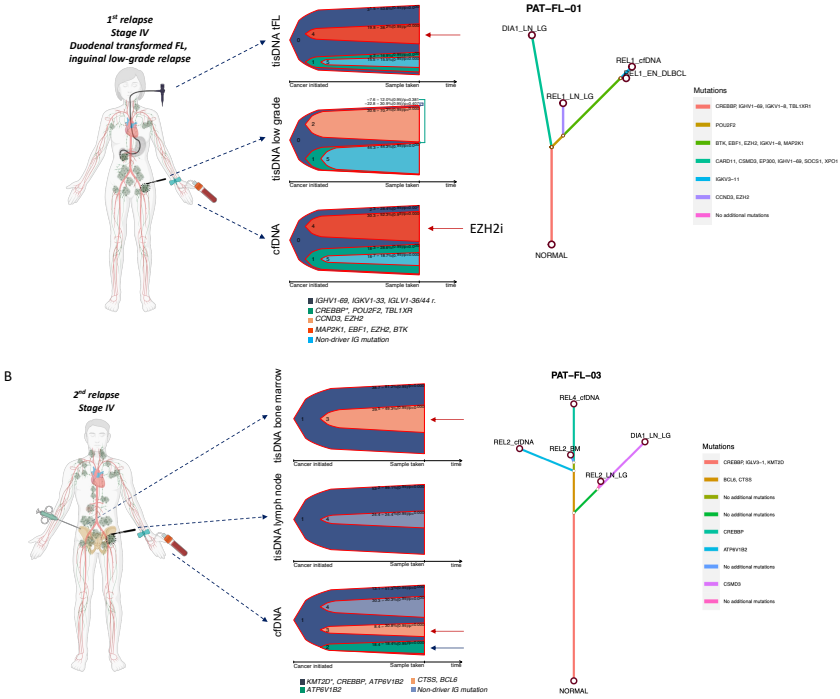

Figure 4. *Spatial dynamics of subclones in FL. A) A representative case showing a spatially restricted subclone*

potentially driving disease progression through the selection of MAP2K1, EBF1, EZH2 and BTK mutations in the transformed duodenal mass readily recovered in the cfDNA compartment. B) A representative case showing an occult subclone harboring an ATP6V1B2 mutation only detected in the cfDNA compartment. Phylogenetic reconstruction shows that the bone marrow and cfDNA sample more readily predicts the genetic background of disease relapse two treatment lines later than nodal tisDNA.

Variants in immunoglobulin genes were also involved in the analysis in order to increase the number of variants leading to more precise clustering. Altogether the 15 samples harboured 208 non-silent variants localized to 90 unique variant positions with median 12 variants per sample (range, 6-28).

Investigating spatial dynamics of subclones based on variant clustering revealed two patterns contributing to spatial heterogeneity. In 4/5 patients a subclone was spatially restricted, meaning that it was only recoverable in one of the tisDNA samples, but readily recovered in the cfDNA compartment (Figure 4.). Additionally, this spatially restricted subclone was shown to harbor the capability of driving disease progression, as evidenced in case #1, where the spatially restricted subclone characteristic of the transformed duodenal mass was harboring a *MAP2K1, EBF1, EZH2* and *BTK* mutation including genes known to be potent drivers of high-grade transformation. Interestingly, in 3/5 patients an occult subclone was also detected, that was only recovered in the cfDNA compartment, but detected in none of tisDNA samples (Figure 4.). Additionally, in case #3 it was shown, that the

cfDNA and bone marrow tisDNA samples were able to capture more accurately the clonal composition of the relapse sample two treatment lines later than the nodal tisDNA.

4.3.5. Prognostic value of ctDNA analysis in R/R FL

Median pretreatment ctDNA level was $1.44x10^3$ human genomic equivalent/ml plasma. Investigating the prognostic value of pretreatment ctDNA level we found that treatment response in the subsequent treatment line was significantly associated with pretreatment ctDNA level (Figure 5.). On the other hand, pretreatment ctDNA level was not associated with inferior progression free survival (Figure 5.).

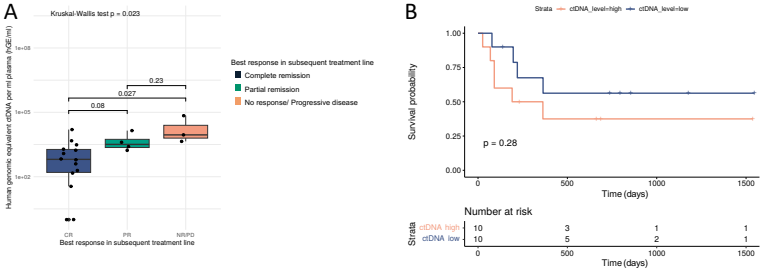

Figure 5. *Prognostic value of pretreatment ctDNA level in R/R FL. A) Pretreatment ctDNA level is significantly associated with treatment response in the subsequent treatment line. B) Pretreatment ctDNA level is not associated with progression free survival in our cohort.*

5. CONCLUSIONS

Novel findings of my thesis are the following:

1. We observed *TNFRSF14* mutations in PCFCL and identified PCFCL cases with complete loss of functional *TNFRSF14* as a consequence of combined 1p36 loss and *TNFRSF14* mutation for the first time.
2. We demonstrated the feasibility of lcWGS for the genome-wide copy number profiling in indolent lymphomas for the first time in a cohort of 28 PCFCL and 64 FL samples.
3. We identified differential distribution of 18q21.33 amplification, 13q14.11-q14.2 deletion and 10q23.32 deletion between PCFCL and FL using lcWGS.
4. We observed for the first time, that higher copy number burden, including proportion of genome altered, as well as number of CNAs were significantly higher in patients developing distant cutaneous spread in PCFCL. Additionally, we identified 2p22.2-p15, 12q13.13-q14.3 and 3q23-q24 amplifications and 9p21.3 deletions as potential biomarkers of distant cutaneous spread.
5. We identified 2p22.2-p15 amplification as an early alteration associated with aggressive disease course analyzing serial tissue samples from six PCFCL patients.
6. We compared the concordance of somatic alterations in FL including cfDNA, lymph node, extranodal and bone marrow samples.
7. We demonstrated two patterns of subclonal dynamics using multiregional sequencing in FL including spatially restricted subclones and occult subclones only

recoverable in cfDNA analysis clarifying the pathogenesis of observed discordance in cfDNA-tisDNA pairs in previous studies

8. We provided preliminary evidence that pretreatment ctDNA level has prognostic value in FL.

9. BIBLIOGRAPHY OF THE CANDIDATE'S PUBLICATIONS

RELATED TO THE THESIS:

1, Bátai, B., Kiss, L., Varga, L., Nagy, Á., Household, J., Baker, A-M., László, T., Udvari, A., Horváth, R., Nagy, T., Csomor, J., Szakonyi, J., Schneider, T., Graham, T.A., Alpár, D., Fitzgibbon, J., Szepesi, Á., Bödör, C. Profiling of copy number alterations using low- coverage whole-genome sequencing informs differential diagnosis and prognosis in primary cutaneous follicle center lymphoma, Modern Pathology 2024, 37(5): 100465.

2, Gango, A.*, Batai, B.*, Varga, M., Kapczar, D., Papp, G., Marschalko, M., Kuroli, E., Schneider, T., Csomor, J., Matolcsy, A., Bodor, C., Szepesi, A. Concomitant 1p36 deletion and TNFRSF14 mutations in primary cutaneous follicle center lymphoma frequently expressing high levels of EZH2 protein. Virchows Archiv 2018, 473(4): 453- 462. *These two authors contributed equally to this work.

FURTHER PUBLICATIONS:

3, Bedics, G., Szőke, P., Bátai, B., Nagy, T., Papp, G., Kránitz, N., Rajnai, H., Reiniger, L., Bödör, Cs., Scheich, B. Novel, clinically relevant genomic patterns identified by comprehensive genomic profiling in ATRX-deficient IDH-wildtype adult high-grade gliomas. Scientific Reports. 2023;13(1):18436.

4, Bedics, G., Egyed, B., Kotmayer, L., Benard-Slagter, A., de Groot, K., Bekő, A., Hegyi, L. L., Bátai, B., Krizsán, Sz., Kriván, G., Erdélyi, D. J., Müller, J., Haltrich, I., Kajtár, B., Pajor, L., Vojcek, Á., Ottóffy, G., Ujfalusi, A., Szegedi, I., Tiszlavicz, L. Gy., Bartyik, K., Csanádi, K., Péter, Gy., Simon, R., Hauser, P., Kelemen, Á., Sebestyén, E., Jakab, Zs., Matolcsy, A., Kiss, Cs., Kovács, G., Savola, S., Bödör, Cs., Alpár, D. PersonALL: a genetic scoring guide for personalized risk assessment in pediatric B-cell precursor acute lymphoblastic leukemia. British Journal of Cancer. 2023; 129(3):455-465.

5, Nagy, Á., Bátai, B., Kiss, L., Gróf, S., Király, P. A., Jóna, Á., Demeter, J., Sánta, H., Bátai, Á., Pettendi, P., Szendrei, T., Plander, M., Körösmezey, G., Alizadeh, H., Kajtár, B., Méhes, G., Krenács, L, Timár, B., Csomor, J., Tóth, E., Schneider, T., Mikala, G., Matolcsy, A., Alpár, D., Masszi, A.*, Bödör, Cs.* Parallel testing of liquid biopsy (ctDNA) and tissue biopsy samples reveals a higher frequency of EZH2 mutations in follicular lymphoma. Journal of Internal Medicine. 2023; 294: 295–313.

6, Varga, A., Nguyen, M. T., Pénzes, K., Bátai, B., Gyulavári, P., Gurbi, B., Murányi, J., Csermely, P., Csala, M., Vántus, T., Sőti, Cs. Protein Kinase D3 (PKD3) Requires Hsp90 for Stability and Promotion of Prostate Cancer Cell Migration. Cells. 2023, 12(2), 212.

7, Nagy, Á., Bátai, B., Balogh, A., Illés, S., Mikala, G., Nagy, N., Kiss, L., Kotmayer, L., Matolcsy, A., Alpár, D., Masszi, T., Masszi, A., Bödör Cs. Quantitative Analysis and Monitoring of

EZH2 Mutations Using Liquid Biopsy in Follicular Lymphoma. Genes (Basel) 2020, 11(7): E785.

8, Bödör, Cs., Alpár, D., Marosvári, D., Galik, B., Rajnai, H., Bátai, B., Nagy, Á., Kajtár, B., Burján, A., Deák, B., Schneider, T., Alizadeh, H., Matolcsy, A., Brandner, S., Storhoff, J., Chen, N., Liu, M.D., Ghali, N., Csala, I., Bagó G., A., Gyenesei, A., Reiniger L. Molecular subtypes and genomic profile of primary central nervous system lymphoma. Journal of Neuropathology and Experimental Neurology 2020, 79(2): 176–183.

9, Richárd Kiss, Gergő Papp, Szilvia Krizsán, Lili Kotmayer, Ambrus Gángó, Noémi Nagy, Bence Bátai, Zoltán Mátrai, Csaba Bödör, Donát Alpár. Screening for genomic copy number alterations in chronic lymphocytic leukemia using multiplex ligation- dependent probe amplification. [Hungarian] Hematológia-Transzfuziológia 2018, 51(1): 31-40.