GENETIC MARKERS AS TOOLS FOR INDIVIDUALIZED DISEASE MONITORING IN PEDIATRIC ACUTE LYMPHOBLASTIC LEUKEMIA

PhD thesis

Bálint Egyed, MD

Károly Rácz Doctoral School of Clinical Medicine

Semmelweis University

Supervisors: Gábor Kovács, MD, D.Sc
Ágnes Félné Semsei, Ph.D

Official reviewers: Zsolt Nagy, MD, Ph.D
Beáta Soltész, Ph.D

Head of the Complex Examination Committee: Judit Demeter, MD, D.Sc

Members of the Complex Examination Committee:
Gergely Kriván, MD, Ph.D
Hajnalka Andrikovics, MD, Ph.D

Budapest
2022

Table of Contents

List of Abbreviations

ΔFC	expression fold change
ABL1	Abelson murine leukemia viral oncogene homolog 1
ACT	anthracycline-induced cardiotoxicity
ACTA	actin assembly-inducing protein
AKR	aldo-keto reductase
AKR1A1	aldo-keto reductase family 1 member A1
ALL	acute lymphoblastic leukemia
AML	acute myeloid leukemia
AMPK	5' AMP-activated protein kinase
ANT	anthracycline
ARID5B	AT-rich interaction domain 5B
ATP5J	ATP synthase-coupling factor 6, mitochondrial
AUC	area under the curve
BFM	Berlin–Frankfurt–Münster
BM	bone marrow
BN	Bayesian network
BTLA	B and T lymphocyte associated
CBR	carbonyl reductase
CDKN2A/2B	cyclin dependent kinase inhibitor 2A/2B
cDNA	complementary DNA
CEBPE	CCAAT/enhancer binding protein epsilon
CNS	central nervous system
CNS^{-}	CNS-negative
CNS^{+}	CNS-positive
CREBBP	CREB binding protein
CRLF2	cytokine receptor like factor 2
CSF	cerebrospinal fluid
CTCAE	Common Terminology Criteria for Adverse Events
CYP3A4	cytochrome P450 family 3 subfamily A member 4
CYP3A5	cytochrome P450 family 3 subfamily A member 5
DNA	deoxyribonucleic acid
DNR	daunorubicin
DOX	doxorubicin
DRZ	dexrazoxane
EBF1	early B-cell factor 1
ECHO	echocardiography
ETV6	ets variant gene 6 (TEL oncogene)

EV	extracellular vesicle
FC	flow cytometry
GATA4	GATA-binding protein 4
IKZF1	IKAROS family zinc finger 1
ITGA6	integrin alpha-6
JAK1/2	Janus kinase 1/2
KRAS	Kirsten rat sarcoma viral oncogene homolog
LVFS	left ventricular fractional shortening
MAF	minor allele frequency
miR	micro-ribonucleic acid
MLC2	myosin light chain 2
MLL	mixed-lineage leukemia
MM-CK	creatine kinase (skeletal muscle)
MPAL	mixed phenotype acute leukemia
MRD	minimal residual disease
mRNA	messenger ribonucleic acid
NCCR	National Childhood Cancer Registry
NF1	neurofibromin 1
NGS	next generation sequencing
NRAS	neuroblastoma RAS viral (v-ras) oncogene homolog
OR	odds ratio
OSC	osteosarcoma
PAX5	paired box 5
PB	peripheral blood
pB-ALL	precursor B-cell ALL
PC	principal component
PCA	principal component analysis
PCR	polymerase chain reaction
PDGFRB	platelet derived growth factor receptor beta
PGC-1α/β	peroxisome proliferator-activated receptor gamma coactivator 1-alpha/beta
PTPN11	protein tyrosine phosphatase, non-receptor type 11
RAG	recombination-activating gene
RB1	retinoblastoma 1
RNA	ribonucleic acid
ROC	receiver operating characteristic
ROS	reactive oxygen species
RUNX1	Runt-related transcription factor 1
RyR2	ryanodine receptor 2
SE	standard error

SERCA2A	sarco/endoplasmic reticulum Ca^{2+} adenosine triphosphatase-2a
sEV	small extracellular vesicle
SLC28A3	solute carrier family 28 member 3
SMA	spinal muscular atrophy
SNP	single nucleotide polymorphism
SR	sarcoplasmic reticulum
T-ALL	T-cell acute lymphoblastic leukemia
TLDA	TaqMan™ Advanced Low Density miRNA Array
TNNT	troponin T
TOP2	type II topoisomerase
TOX	thymocyte selection associated high mobility group box
TP53	tumor protein p53
VEGFA	vascular endothelial growth factor A
β1-AR	β1-adrenoceptor
β2-AR	β2-adrenoceptor
βAR	β-adrenoceptor
ψLC	pseudo-light chain complex

I. Introduction

The term acute lymphoblastic leukemia (ALL) applies to an immunologically and genetically heterogenous malignant entity of the lymphopoietic system [1]. Contrary to most cancer types, ALL predominantly presents in childhood. Due to clinical trials and first-line chemotherapy over the past few decades, the majority of children with ALL survive and struggle with the problems of returning to community. However, many patients have severe acute or late-onset therapeutic side effects, and the survival rate in some groups (e.g. patients with high-risk tumor genetic features or resistant organ infiltration) is far below the average.

I.1. Significance of ALL and its clinical management in childhood

Despite extensive endeavors towards risk-directed treatment, acute leukemia remains the major cause of death and is responsible for the highest number of disability-adjusted life-years in children and adolescents with cancer [2]. To explore the distribution of the pediatric cases with malignancy in Hungary, the National Childhood Cancer Registry (NCCR) was launched by the Hungarian Pediatric Oncology Network in 1971. According to NCCR data, the incidence of ALL shows a slight elevation year by year [3], which is in parallel with foreign observations [4]. Data from the United States of America denote that the treatment of childhood leukemia costs almost five times as much as hospitalizations for other pediatric conditions, exceeding $50,000 per stay as a financial burden for families as well as national insurance companies [5]. Hereby, precise disease monitoring during the treatment and early recognition of resistant malignant cell clones are prerequisites of advanced leukemia care capable of providing survival rates as high as 85-90% for most cases [6].

The backbone of childhood ALL treatment is still combination chemotherapy, which proves to be effective in 4 of 5 patients to reach long-lasting remission [7]. In the 1990s, the Berlin-Frankfurt-Münster (BFM) regimen was developed, then, adapted in Hungary as well. The cytotoxic scheme starts with an induction phase aiming cytoreduction, eradication of >99% of leukemic blast cells and achievement of complete hematologic remission [6]. Daunorubicin (DNR), a drug from the group of anthracyclines, is a notable part of the initial agents, however, its role seems to be controversial. Some studies

proposed that DNR in induction protocol does not lengthen the duration of remission, but increases the risk of therapy-associated morbidity, thus, certain centers restrain DNR use in the early phase of therapy [8]. The second phase of chemotherapy is pre-symptomatic treatment directed to the central nervous system (CNS) with intrathecal drug administration in order to prevent meningeal relapse. In the absence of the last phase of intravenous chemotherapy, the reinduction regimen, most patient would suffer early, measurable residual disease (MRD)-driven relapse in a few months [9]. The 2 year-long BFM scheme ends with oral maintenance therapy intended to prevent late bone marrow (BM) relapse. The application of similar pediatric ALL chemotherapy guidelines place leukemia care quite far from personalized treatment, but the different intensity of the therapy according to disease progression risk seems to come to the front [10].

Nowadays, the two major challenges in the clinical management of ALL are (*i*) relapsing or refractory disease course and (*ii*) toxic side effects of chemotherapeutic agents which limit efficacy of treatment or quality of life. Firstly, patients with acute leukemia have approximately 10^{12} neoplastic cells at the diagnosis, but up to 10^{10} cells can survive as MRD in the state of clinical and patho-morphological complete remission [11]. The investigation of clonal evolution processes in childhood ALL samples showed that the tumor cell population at relapse usually arises from relapse-fated subclones already existing at the diagnosis [12]. To date, for the follow-up of treatment response and prediction of relapse-free survival, there is no more specific and sensitive marker than measuring MRD [9,13,14]. However, methods for MRD evaluation require regular invasive procedures (e.g. BM biopsy, lumbar puncture) performed under general anesthesia in children and a dedicated laboratory team and equipment to minimalize inter-examiner variability of results. Secondly, organ toxicities of high-dose cytotoxic therapy are inevitable in childhood ALL [7,9,15]. The risk for developing health problems shows an eight-fold increase in pediatric cancer survivors within 30–40 years after diagnosis compared to their siblings; 50% of the survivors experience severe, disabling or life-threatening events, including death by the age of 50 [16,17]. Preventive intrathecal therapy may cause late-onset intracerebral calcification, cortical atrophy or multifocal leukoencephalopathy [9]. Survivors of childhood ALL have a two times risk for memory deficit compared to their healthy siblings [18]. Growth hormone deficiency can also evolve causing central obesity and dyslipidemia in adulthood. The quarter of survivors

have to face osteopenia at the age of 30 [19]. It was recognized early that significant cumulative anthracycline dose causes temporary abnormality in the left ventricular function in over the half of the childhood ALL patients, however, manifest heart failure is rare (1-4%). During the first 30 years after cancer diagnosis, pediatric ALL survivors are nine times as likely to have cerebrovascular events, ten times as likely to have coronary heart disease, and 15 times as likely to have heart failure as their siblings without cancer [20,21].

I.2. Methods for following progressive disease course in ALL

Disease monitoring is based on the pathobiology of ALL cells, which opens the door to tumor cell detection techniques. Exact evolutionary trajectories and leukemia-driving mutational mechanisms are still not fully understood. Yet, it is known, that genetic and immunological alterations specific for ALL occur in a typical order and in typical stages of differentiation, not randomly (Figure 1A) [6,10]. Basically, there are two ways of residual disease monitoring in ALL exploiting these pathobiological manners: cellular and subcellular methods. Examinations on cell level include microscopic (e.g. cytospin preparation or cell counting chamber analysis of cerebrospinal fluid) and cytometric (e.g. multiparametric flow cytometry [FC] of BM aspirate) techniques, while molecular genetic methods (e.g. polymerase chain reaction [PCR] and next generation sequencing [NGS]) take advantage of subcellular information.

I.2.1. Residual disease monitoring

For the last one or two decades, MRD assessment has been part of the routine clinical practice in the developed world. Also in relapsed or refractory ALL patients and patients undergoing stem cell transplantation, MRD diagnostics is guiding treatment decisions. If little residual leukemia is left during childhood ALL treatment, it is most reliably and abundantly found in the BM [22]. Many countries use FC and quantitative PCR based MRD estimation methods, with sensitivities of 10^{-3}–10^{-4} and 10^{-4}–10^{-6}, respectively. The great majority of ALL patients can be examined in a rapid and cost-effective way by FC, which is based on the identification of leukemia-associated immunophenotype. However, standardization of FC is difficult and there is a considerable inter-examiner variability. PCR is a more sensitive and highly reproducible method, which usually shows

antigen receptor and immunoglobulin gene rearrangements or, uncommonly, gene fusions. Disadvantages of PCR method in ALL MRD diagnostics include longer turnaround time (high-cost and labour intensive). False negative results occur with both techniques: MRD estimation may be weakened by drug-induced immunophenotype modulation in FC and by oligoclonality or clonal evolution in PCR [23]. Nowadays, application of high throughput sequencing by NGS methodology is being studied to increase MRD assessment sensitivity and to improve patient stratification [24].

I.2.2. Detection of central nervous system involvement

Leukemic involvement of the CNS represents an outstanding therapeutic problem [25]. Prophylactic regimens against CNS infiltration have become a prerequisite of successful treatment: prior to the introduction of CNS-targeted therapy in the 1960s, symptomatic meningeal leukemia developed in more than 50% of cases and CNS relapse rate was over 65% in childhood ALL [26,27]. Nevertheless, patients who have a history of overt CNS involvement at the initial diagnosis are still susceptible to relapse with only 15–20% 5-year survival rates [28]. According to the modern approach, the CNS invasion can be the ultimate step of the tumor cascade [29], and vice versa, CNS relapse of acute leukemia constitutes an early manifestation of systemic relapse. Moreover, subclones freely traffic between leukemic niches of different organs (Figure 1B) [30]. Thus, neural and BM compartments both need to be monitored to reach proper CNS disease control. As described above, we have very sensitive laboratory methods to monitor MRD in the BM let us able to quantify the leukemic cells after a fall by 4 to 6 logs. However, we lack tools with similar sensitivity regarding the CNS compartment. The most widespread contemporary methods for the assessment of cerebrospinal fluid (CSF) are the conventional cytologic examination of a cytocentrifuge smear and flow cytometry [31]. Especially due to the paucity of cells in the CSF, both methods yield a high proportion (more than 40%) of false-negative reports among patients with proven CNS infiltration by neuroimaging or autopsy [32]. Consequently, a gold standard method for staging the CNS involvement in acute leukemia has not been established yet.

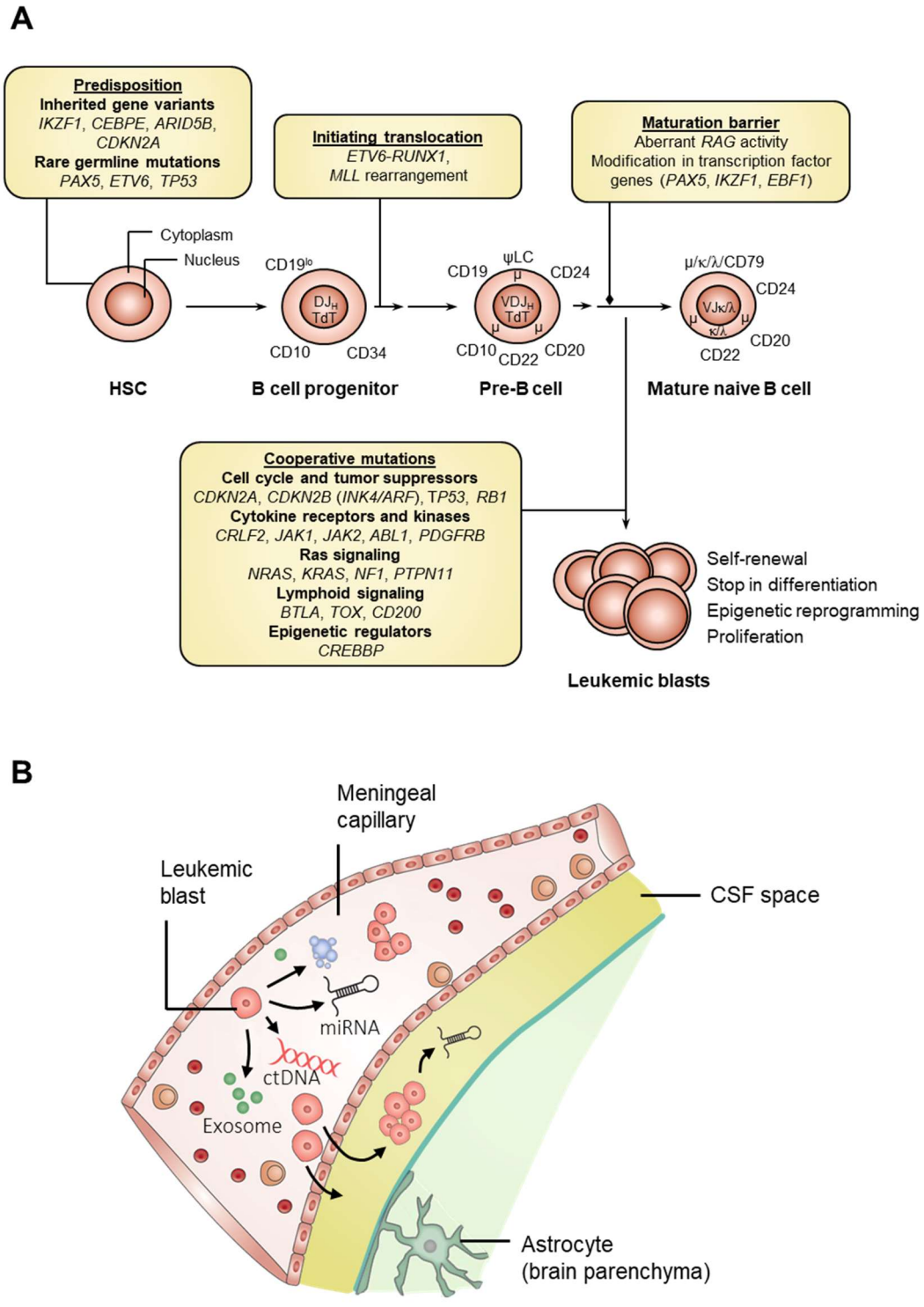

Figure 1. (**A**) Supposed pathogenesis of the precursor B-cell acute lymphoblastic leukemia with possible aberrant genes. (**B**) Way of invading the neuronal niche by leukemic cells. Abbreviations: CD, cluster of differentiation; HSC, hemopoietic stem cell; CSF, cerebrospinal fluid; miRNA, microRNA; ctDNA, cell-free tumour DNA; TdT, terminal deoxynucleotidyl transferase. Gene names are unfolded in 'List of Abbreviations'. The figures are self-created, panel A is based on [6].

I.3. Mechanisms behind and monitoring of permanent side effects of anthracyclines

Anthracyclines (ANTs) are Janus-faced part of chemotherapy regimens against childhood ALL: they are regarded as one of the most effective anticancer drugs decades after their discovery [33], but their use has to be limited due to occasionally life-threatening cardiotoxic side effects. The highly elevated susceptibility of the cardiomyocytes for the anthracycline-related toxic events compared to any other healthy tissues can be explained by subcellular mechanisms (Figure 2).

Firstly, anthracycline-induced cardiotoxicity (ACT) is provoked by the mitochondrial dysfunction of cardiac muscle tissue. Mitochondrial density in cardiomyocytes is high in order to satisfy the need for oxidative metabolism (especially beta-oxidation of fatty acids) [34]. ANTs bind to cardiolipin with high affinity, thus, block the respiratory chain [35]. In addition, fatty acid oxidation is suppressed by ANT-mediated inhibition of adenosine monophosphate-activated protein kinase (AMPK) as well as peroxisome proliferator-activated receptor gamma coactivator 1-alpha (PGC-1α) [36,37]. Secondly, the disruption of calcium homeostasis is also a reason of ACT. The myofibrillar calcium supply is disenabled by the downregulation of sarco/endoplasmic reticulum Ca^{2+} adenosine triphosphatase-2A (SERCA2A), phospholamban and ryanodine receptor 2 (RyR2) as a response to ANTs [38,39]. The anthracycline-induced cytoplasmatic calcium overload leads to left ventricular stiffness in diastole, thus, asymptomatic diastolic dysfunction (weakened relaxation) is supposed to be the earliest sign of ACT [40,41]. GATA-binding protein 4 (GATA4) transcription factor is also downregulated causing the deficit of contractile proteins (e.g. myosin light chain, alpha-actinin, troponin) [42], which serves as a basis for secondary systolic dysfunction [43]. Thirdly, the production of reactive oxygen species (ROS) plays a role in the cellular susceptibility for ACT. Antioxidant enzyme (e.g. superoxide dismutase, catalase) concentration is relatively low in cardiomyocytes, which makes these cells susceptible for ROS-induced injury [44]. Hydrogen peroxide and lipid peroxidation products increase myocardial aldo-keto reductase (AKR) and carbonyl reductase (CBR) levels. These enzymes, besides detoxifying lipid aldehydes, convert ANTs to secondary alcohol metabolites (e.g. doxorubicinol), which form long half-life reservoirs in cardiomyocytes and are responsible for slow progression of ACT [45]. Fourthly, β-adrenergic receptor dysregulation is an emerging mechanism behind ACT development. In a study by

Bernstein et al, β2-adrenoceptor (β2-AR) knock-out animals died suddenly after ANT infusion, which was preventable by β1-adrenoceptor (β1-AR) inactivation [46]. These results suggest the cardioprotective role of β2-ARs and the pro-cardiotoxic effect of β1-ARs. ANTs probably upregulate both adrenoceptor family [47]. Fifthly, inhibited cardiac muscle renewal may raise ACT risk at the stem cell level. Cardiac progenitor cells are destroyed by ANTs, presumably via inhibition of neuregulin-I/ErbB signaling pathway [41,48]. This effect is particularly relevant in children with physiologically developing heart.

ACT is a progressive and irreversible disorder. At the end of ANT therapy, a slightly decreased left ventricular contractility can be measured with damaged ejection, but this is well-compensated by heart remodeling for years. Later, persistent myocardial stress from various source causes cell loss with fibrosis and ventricular wall thinning, which can be quantified by significantly decreased left ventricular fractional shortening (LVFS) [49]. LVFS reduction seen by echocardiography (ECHO) is related to already necrotized cardiac muscle volume, so it is considered as a late marker for ACT and is not suitable for indicating the need for rapid therapeutic interventions due to toxicity [50]. Despite this fact, LVFS monitoring by ECHO was the standard method for ACT follow-up in Hungarian pediatric ALL patients in the last two decades in parallel with previous international convention [51].

I.4. Promising genetic markers in blood cancer monitoring and toxicity prediction

In 2005, a group reported the first study showing the prognostic and diagnostic importance of micro-ribonucleic acids (miRs) [52], which opened out this field of biomarker research. MiRs belong to the class of highly conserved small non-coding RNAs that play key regulatory role in a wide range of biological processes such as proliferation, differentiation and survival [53]. MiRs regulate more than half of the human genes [54]. Altered expression patterns of miRs have been increasingly recognized associating with progression in various types of tumors including acute leukemia [55]. MiRs can act as oncogenes (oncomiRs) or tumor suppressors and are involved in a variety of pathways disturbed in cancer [56]. The deregulation of miRs in disease conditions can be harnessed as potential new-generation therapeutics.

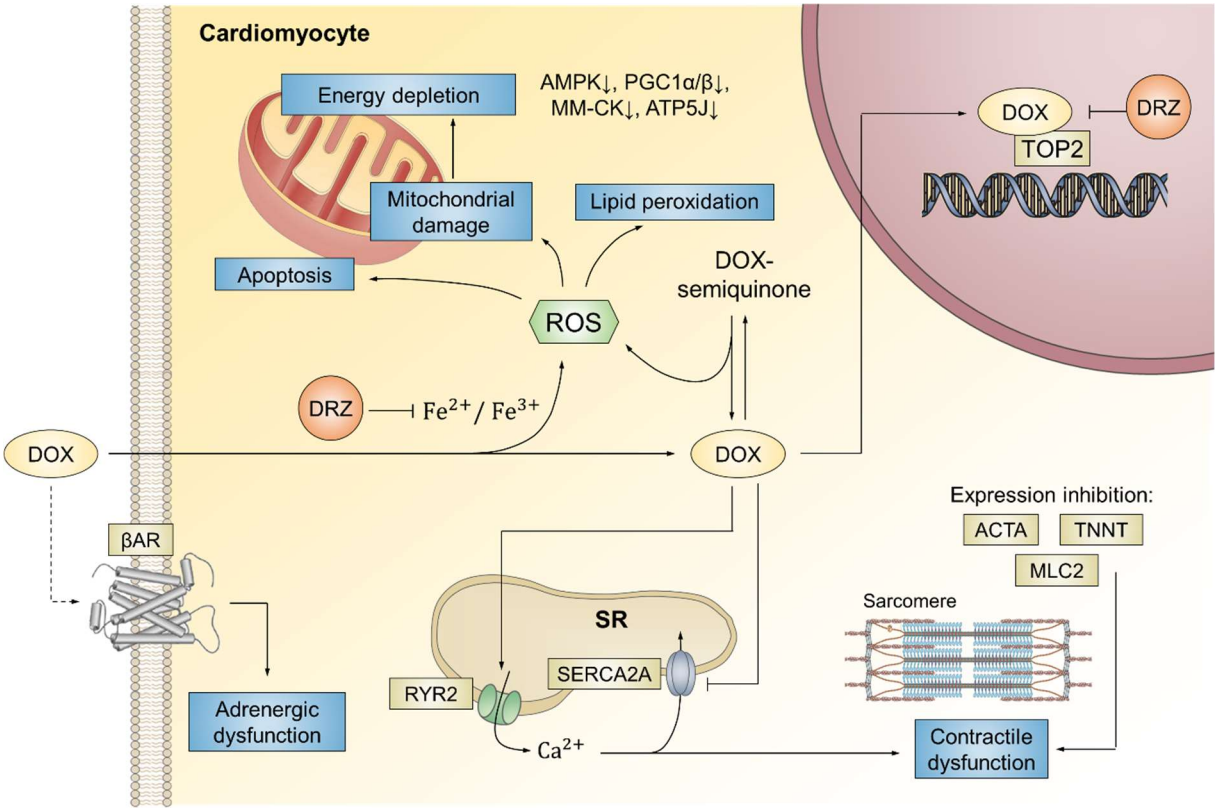

Figure 2. Molecular mechanisms of cardiotoxic side effect induced by anthracyclines and salvage points exploited by the cardioprotective agent dexrazoxane. Abbreviations: DOX, doxorubicin; DRZ, dexrazoxane; ROS; reactive oxygen species; SR, sarcoplasmic reticulum. Protein names are unfolded in 'List of Abbreviations'. Own representation based on [20].

As a very universal phenomenon, malignant cells secrete higher amounts of miRs than healthy cells or tissues [57]. This, together with the selective over-secretion of some miRs, their easy distribution among body fluid spaces and their stability makes miRs good candidates for cancer biomarkers [58]. For miRs, one way to provide protection from RNases is the encapsulation into extracellular vesicles (EVs). Cells in the human body continuously discharge large amounts of vesicles 30–1000 nm in diameter into the extracellular space. EVs are important factors of intercellular communication: they transfer DNA, coding and regulatory RNAs, lipids and proteins in the microenvironment as well as among very distant tissues [59]. It is widely accepted that EV production increases in the tumor microenvironment, hence, these particles may facilitate proper 'cancer cell biopsy'. In studies performed on different leukemia patients, presence of EVs

originating from leukemic cells was shown and certain alterations in the EVs correlated with the hijacking of disease niches and treatment efficacies [60].

Measuring miR and EV miR content of liquid biopsy specimens from leukemia patients may open new trends in determining the burden and the spatial heterogeneity of the disease. Previously, we designed a study to establish a miR-based molecular ALL MRD screen method which indicates post-diagnosis BM MRD by measuring miR levels in peripheral blood (PB) [61], but MRD in the CNS compartment also have to be monitored by novel sensitive CSF markers. Generally, eligible MRD markers needs to be standardizable and identify impending relapse, thus, enable early intervention [62].

Despite the fact that anthracycline-induced cardiac injury is well known, the toxicity is unpredictable [63–65]. There are patients with cardiac problems who received very low doses of anthracyclines while others were administered with high doses and escaped the side effect. The variable development of anthracycline cardiotoxicity suggests that the genetic background of the patients is important in this side effect [66]. Inherited differences in DNA sequence contribute to phenotypic variation, influencing an individual's risk of disease. Most human sequence variation is attributable to single nucleotide polymorphisms (SNPs), which may occur every 100 to 300 bases. In contrast to mutable genetic markers, such as microsatellites, SNPs have a low rate of recurrent mutation, making them stable indicators [67]. SNPs may influence promoter activity interfering gene expression, mRNA stability and subcellular localization of mRNAs or proteins, hence, may produce disease. New, useful SNP markers for medical testing can force a way of safer individualized medication.

II. Objectives

Individualized care is a high-priority endeavour in current clinical management of childhood ALL. Newly identified genetic markers guiding the physician's decision in a progressive disease course or a chemotherapy-induced toxic event may contribute to better survival rates and/or quality of life after recovery from leukemia. Our studies searched the answers for dilemmas and questions as follows:

1. *High-throughput screening of a selected set of microRNAs (miRs) in cell-free specimens produced from peripheral blood (PB) and cerebrospinal fluid (CSF) samples from children with ALL to investigate their expression patterns.*
   1.1. Can we setup a reliable laboratory workflow to measure the expression of leukemia-specific miRs in diagnostic PB samples?
   1.2. Are there measurable leukemia-specific miRs in CSF samples and can they show altered expression in patients with central nervous system (CNS) leukemia compared to cases without CNS disease?

2. *The investigation and validation of miRs as novel CSF markers of CNS involvement in various sample types of children with ALL.*
   Can we distinguish between CNS-positive (CNS^+; unambiguously identified blasts in CSF) and CNS-negative (CNS^-; blast-free CSF) patients by miR expression in CSF samples of an extended patient cohort? How do CSF miR levels change with time, in parallel with intrathecal chemotherapy progress? Can we follow the therapeutic response of CNS compartment this way? Does miR-based CSF diagnostics of CNS involvement provide advantage in sensitivity and specificity compared to contemporary cytological methods?

3. *Assessment of PB and bone marrow (BM) miR expression as indicator of CNS manifestation of ALL.*
   PB sampling and BM aspiration remain routine procedures during childhood ALL treatment, therefore, a PB- or BM-derived marker of CNS leukemia could pave the way for a less invasive monitoring of the CNS compartment.

4. *Preliminary examination of the CSF small extracellular vesicle (sEV) content in CNS leukemia.*

   Can we demonstrate sEVs in CSF samples of children with ALL? Is there any difference in sEV density between CSF samples from CNS^+ and CNS^- patients?

5. *Pharmacogenetic study of anthracycline-induced cardiotoxicity (ACT) in a considerable Hungarian pediatric ALL cohort.*

   5.1. What are the trends in long-term left ventricular function changes of Hungarian ALL survivors who received ANTs during cancer treatment?

   5.2. Can we identify single nucleotide polymorphisms (SNPs) which predispose to or defend from a pathological decline in left ventricular function of the heart established by echocardiography? Are there any haplotype blocks (ie. combination of SNPs inherited together) of the examined SNPs which influence left ventricular function in the population of ALL survivors?

   5.3. How does the interaction network or priority order of SNPs relevant in ACT look? Can we outline a causality structure between clinical and genetic factors influencing the risk of ACT?

III. Results

With an aim of finding novel genetic markers that can help the field of pediatric leukemia care, we measured the expression level of 47 miRs and genotyped 70 SNPs. Genetic data was evaluated using regression models, for such analyses the statistical power exceeded 75%.

III.1. High-throughput screening of liquid biopsy samples from a discovery cohort of children with ALL

Between October 2015 and August 2019, we collected PB, bone marrow (BM) and cerebrospinal fluid (CSF) from 186 consecutive, unselected patients (aged ≤ 18 years) diagnosed with acute leukemia or concomitant relapse in two Hungarian pediatric hematology centers (Semmelweis University 2nd Department of Pediatrics and Heim Pal National Pediatric Institute). Our first aim was to setup a high-throughput miR profiling method in cell-free liquid specimens (e.g. platelet-free plasma, CSF after cell removal) produced from the collected biofluids. Expression profiling was performed using Custom TaqMan^{TM} Advanced Low Density miRNA Array (TLDA) cards (Thermo Fisher Scientific, Waltham, MA, USA) [61]. A set of candidate reference/housekeeping (15) and leukemia-specific (32) miRs were quantified [61,68]. Only those disease-relevant miRs were included in miR profiling which were found to be overexpressed in leukemic blasts in at least two previously published studies from at least two different laboratories.

III.1.1. Peripheral blood (PB) miR expression pattern of diagnostic ALL samples

Based on our sample bank, firstly, we studied 15 pediatric patients with *de novo* and 5 independent patients with 1st relapse of precursor B-cell ALL (pB-ALL), detailed in Table 1. Only three cytogenetic subgroups were included in order to minimize tumor biological confounders: hyperdiploidy, *ETV6*/*RUNX1* fusion, normal karyotype.

The expression of candidate miRs in diagnostic (day 0) PB platelet-free plasma (PFP) samples is shown in Figure 3. As a quality control of this initial screen, four samples were measured in duplicates on TLDA cards and these pairs showed a very high correlation (Pearson's r = 0.981 [CI 95% 0.974–0.986], $p < 10^{-16}$). Highest stability miRs were selected based on average pairwise variation estimation between all candidate miRs using

geNorm and NormFinder algorithms [69,70]. This way, miR-484 and miR-532-5p were identified as feasible reference miRs. All of the 32 leukemia-related miRs were measurably expressed in the samples, but none of them was significantly different between *de novo* and relapsed ALL groups. Thus, PB PFP expression levels of this selected set of miRs is inapplicable to determine what high-burden disease progression stage (newly diagnosed or relapsing disease) are the patients in. Furthermore, continuing this study, we analyzed and found association between intra-patient PB PFP miR expression changes and decreasing leukemic cell burden in the PB and the BM, in parallel with systemic chemotherapy progress [61]. These 'MRD-indicator' miRs were significantly overexpressed in the PB PFP of patients with ALL compared to the PB PFP of healthy controls. This line of our research was thoroughly demonstrated and discussed in the doctoral thesis of my colleague, Andrea Rzepiel [71]. Nevertheless, the results of this first stage of our miR study regarding PB PFP expression patterns took effects on experiments in the CSF samples with less abundant cell and RNA content: miR expression in a cell-free specimen produced from a biofluid from children with ALL had the potential to indicate blast burden in the certain body compartment.

Table 1. Basic characteristics of childhood acute lymphoblastic leukemia patients involved in high-throughput microRNA study on diagnostic samples. Normal karyotype (†) means no alteration with FISH (fluorescence *in situ* hybridization), DNA index is diploid, and cytogenetics is normal or unsuccessful. Abbreviation: ALL, acute lymphoblastic leukemia.

Subgroups	Median age at diagnosis in years (range)	Number of patients (% of males)
De novo ALL		
Hyperdiploid karyotype	3.1 (2.0-3.7)	3 (33%)
t(12;21) translocation	3.6 (3.1-4.6)	6 (50%)
Normal karyotype†	3.1 (1.5-5.3)	6 (67%)
Relapsed ALL	10.4 (6.3-18.8)	5 (60%)

III.1.2. Differences in CSF miR expression between CNS^+ and CNS^- ALL patients of the discovery cohort found by high-throughput array

Firstly, we examined whether miRs were detectable in frozen and then thawed CSF samples. Interestingly, in direct measurements with spectrophotometry, the concentration of unamplified complementary DNA (cDNA) prepared from total miR was slightly

higher in CNS^- patients than in the CNS^+ group (mean ± standard error: 1222.3 ± 19.4 ng/µl in CNS^- group vs. 1126.8 ± 9.1 ng/µl in CNS^+ group; p=0.002).

TLDA cards was used to screen the miR profile of CSF samples derived from 4–4 CNS^+ and CNS^- patients with pB-ALL (Table 2). In 83% of miRs, the standard error (SE) of the average C_t values in sample duplicates remained under 2 (range: 0.05–4.69). All miRs were quantified at two different treatment checkpoints: (i) at the diagnosis of both CNS^+ and CNS^- patients and (ii) on the 15th follow-up day of CNS^+ patients, see Figure 4A.

Expression fold changes (relative to normalizer miR-532-5p) of each miR in each sample were evaluated by principal component analysis (PCA) with a goal of identifying miRs that could potentially influence the classification of the patients by their CNS status. It was observed that three members of the miR-181-family (miR-181a-5p, miR-181b-5p, miR-181c-5p) had consistent and considerable positive contributions to all the first three principal components (PCs) and their aggregated contributions were above the expected average cut-off value (data shown in [68]). The first three PCs are included to reach a total sum of at least 70% of the original variation (PC1: 36.8%; PC2: 22.9%; PC3: 16.4%). The PCA biplot of PC1 and PC2 (Figure 4B) shows the clusters of CNS^+ and CNS^- diagnostic samples with clustering performed at a confidence interval of 95%. Small angles between the miR-181-family members on loading plot of PC1 and PC2 (Figure 4B) imply strong positive pairwise correlations.

A decrease in CSF miR levels by the 15th day of the treatment was also found in CNS^+ patients. Among all tested RNAs, levels of miR-181a-5p, miR-181b-5p and miR-181c-5p showed the strongest reductions, as their expression fold changes (ΔFC) were 36.2, 299.4 and 39.8 times lower than at the diagnosis, respectively (see more information in the additional file of [68]). In accordance with the previously mentioned results, miR-181a-5p and miR-181b-5p were selected for further validation.

Table 2. Subgroups and case numbers in cerebrospinal fluid microRNA marker study with acute leukemia patients. Abbreviations: CNS, central nervous system; ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; miR, microRNA; MPAL, mixed phenotype acute leukemia; CSF, cerebrospinal fluid; BM, bone marrow; PB, peripheral blood.

			Sample numbers			
			Discovery cohort	**Validation cohort**		
Total number of examined samples			20	138		
Control group (sample type, number of individuals)			PB, n=10	CSF, n=6		
Diagnosis	**Treatment guidelines**	**Number of patients (% of males)**	**CSF**	**CSF**	**BM**	**PB**
CNS^+ patient group						
De novo ALL	ALL IC BFM 2009, Interfant 2006	7 (71%)	5	18	9	10
Relapsed ALL	ALL REZ BFM 2002	2 (50%)	1	3	4	3
De novo AML	AML BFM 98	2 (0%)	0	4	3	4
MPAL	ALL IC BFM 2009	1 (100%)	0	2	3	3
CNS^- patient group						
De novo ALL	ALL IC BFM 2009	12 (58%)	4	13	10	9
Relapsed ALL	ALL REZ BFM 2002	2 (0%)	0	4	5	5
De novo AML	AML BFM 98	2 (50%)	0	3	5	4
MPAL	ALL IC BFM 2009	1 (100%)	0	2	3	3

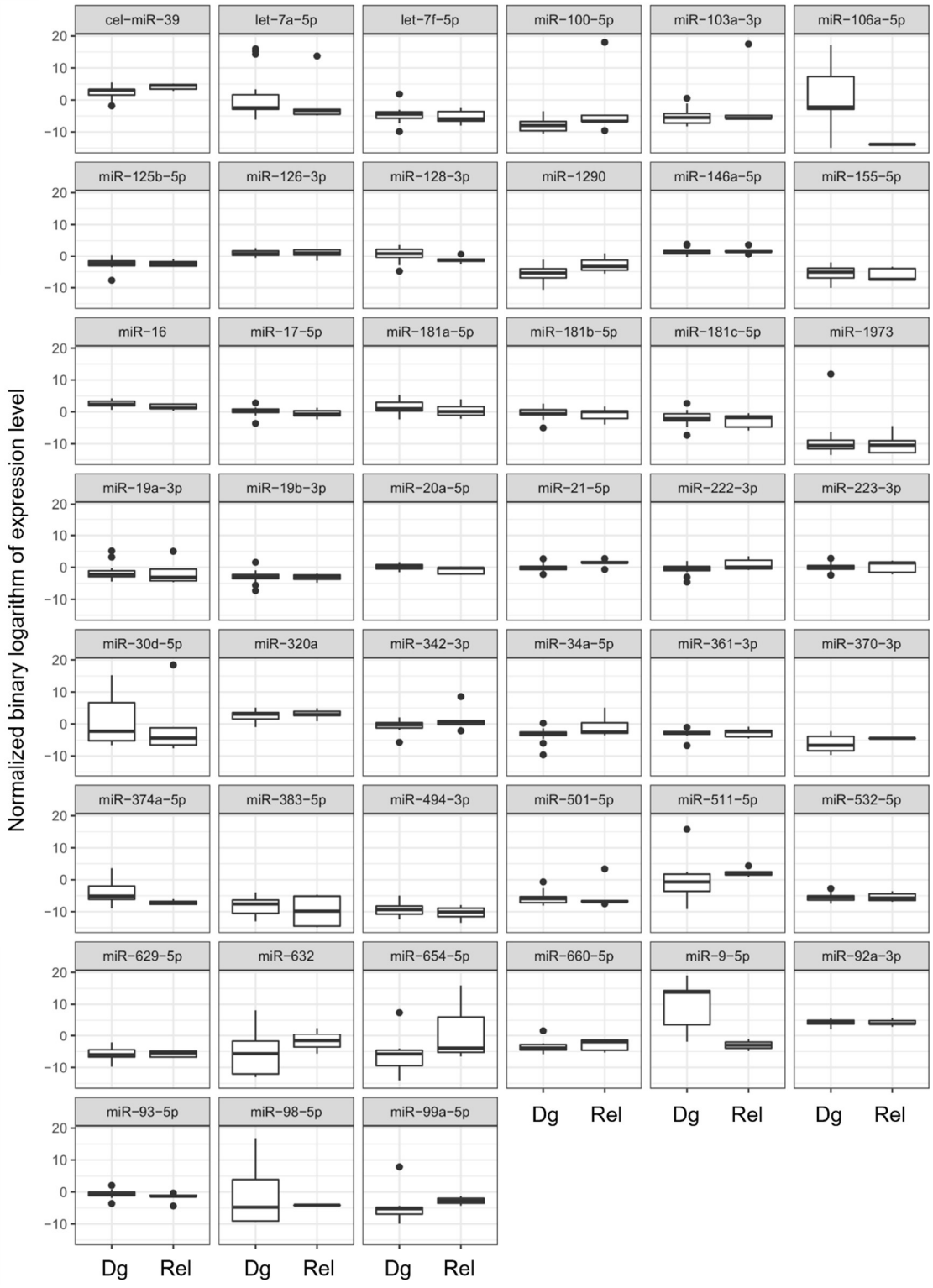

Figure 3. MicroRNA expression in peripheral blood platelet-free plasma of *de novo* and relapsed ALL patients at the day of diagnosis. Box: the 2nd and 3rd quartiles; thick line in the box: median; lower whiskers: minimal value if there are no low outrange values, or Q1−1.5×IQR; upper whiskers: maximum value if there is no upper outlier, or Q3+1.5×IQR. Abbreviations: Dg, diagnosis; Rel, relapse; IQR, interquartile range. Own representation, published in [61].

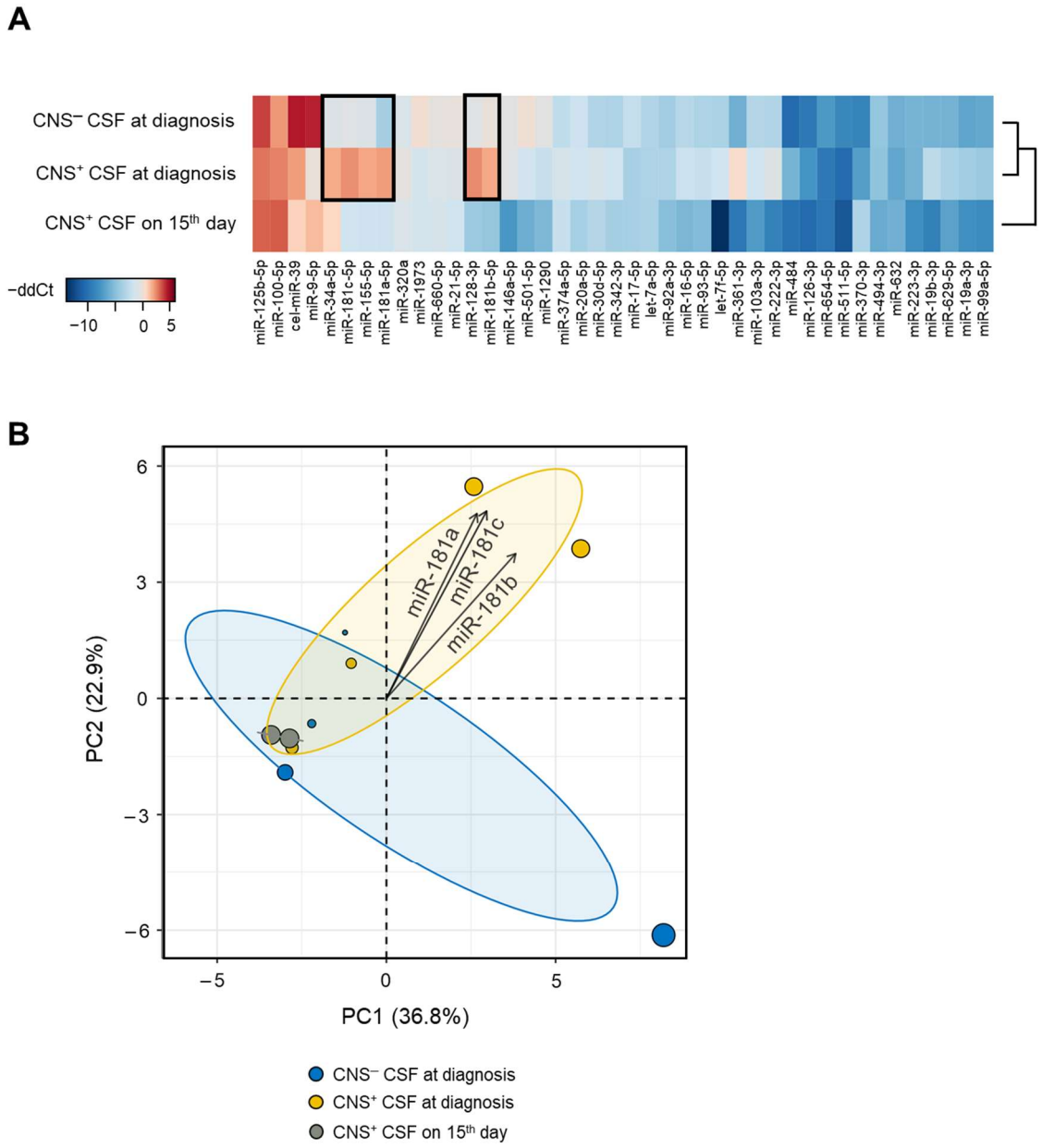

Figure 4. MiR-181-family has consistent contribution to altered microRNA expression in cerebrospinal fluid samples in discovery cohort. (**A**) Heatmap shows the overexpressed (red) and the downregulated (blue) microRNAs (miRs) in distinct sample types. Gross divergence in miR pattern of patients with or without initial CNS disease is highlighted with black frames. Mean relative expression levels were determined by ddCt algorithm, using miR-532-5p as reference miR and PB samples of control individuals as reference samples. (**B**) Biplot derived from principle component analysis shows clusters of CNS-positive and CNS-negative diagnostic samples at a confidence level of 95%. Corresponding loading plot (vectors) displays miR-181-family members. The cosine between two vectors approximates the correlation between miR variables. Position of vectors indicates considerable positive contributions to first and second principal component. CNS, central nervous system; CSF, cerebrospinal fluid; ddCt, delta-delta Ct; miR, microRNA; PC, principal component. Own representation, published in [68].

III.2. Diagnosis of CNS leukemia on the basis of miR-181a-5p level in CSF of patients with ALL

Relying on the TLDA card-based findings (section III.1.2.), two candidate markers (miR-181a-5p, miR-181b-5p) were tested using 49 CSF samples in a validation cohort consisted of 24 patients with lymphoid, myeloid and mixed phenotype acute leukemia (Table 2). We were able to confirm the high expression of miR-181a-5p in diagnostic CSF samples in a partly independent set of 8 CNS^+ ALL patients compared with 10 $CNS^–$ ALL patients by conventional qPCR, see Figure 5A. However, similar results were not found with adding acute myeloid leukemia (AML) or mixed phenotype acute leukemia (MPAL) patients to the analysis. In ALL patients, miR-181a-5p expression levels conferred a more than 52-fold increased risk for CNS leukemia (CNS^+ vs. $CNS^–$ patients: ΔFC=52.30, p=1.49E-4). The miR-181a-5p relative expression level in CSF were independent of B- or T-cell immunophenotypes of ALL. Patients with overt CSF blastosis (CSF with a cell count of >5/µl and blasts in excess on the cytospin slide) had the most outstanding miR-181a-5p expressions within the CNS^+ ALL group. Median miR-181a-5p expression in ALL patients without CNS leukemia did not reach the median miR level detected in a group of spinal muscular atrophy (SMA) patients (n=6) considered as reference sample cohort (Figure 5A). There were no significant expression differences in the case of miR-181b-5p in the validation cohort.

Regarding ALL patients in the validation cohort, the diagnostic efficacy of miR-181a-5p in CSF was evaluated by receiver operating characteristic (ROC) method. This novel marker yielded a sensitivity of 90.0% and specificity of 87.5% (area under the ROC curve, AUC: 92.5%) for CNS disease at a cut-off ΔFC value of 3.7, as shown in Figure 6. To compare the clinical applicability of this miR marker with conventional cytomorphology technique, the latter one was also assessed based on previous reports [72] and signed as a ROC-curve on Figure 6.

To address the issue of treatment-induced alteration dynamics of miR expression in time, we determined the miR-181a-5p expression level in CSF samples of ALL patients during the induction chemotherapy. MiR levels were measured at three time points: at diagnosis (before any drug administration), on the 15th day of therapy and on the 33rd day of therapy. We managed to obtain an extra sample from a patient (P2, see Table 1 in [68]) with massive CNS involvement at 45th day of therapy and samples at exceptional

therapy days (15th, 26th and 40th) from a patient (P12) with fatal disease course. Figure 5B shows changes in relative expression level of CSF miR-181a-5p in CNS^+ and CNS^- ALL groups and, separately, in patient P12. The average level of miR-181a-5p decreased by 96.7% among CNS^+ patients (P12 was excluded from this analysis) by the 33rd day of therapy, while it remained around the expression detected in diagnostic sample at our later time points in patients without CNS leukemia. Interestingly, patient P12's miR-181a-5p expression has risen in parallel with clinically observed disease progression. Patient P12 could not finish the induction chemotherapy and was treated in the intensive care unit until death. In a patient with CNS+ relapse (P1, $\Delta FC=2.40$) we observed a more than 2-fold rise in miR-181a-5p expression by the 15th treatment day from the initial level. This change was not seen in relapsed patients with blast-free CSF (P17 and P18; $\Delta FC=0.06$ and $\Delta FC=0.83$, respectively).

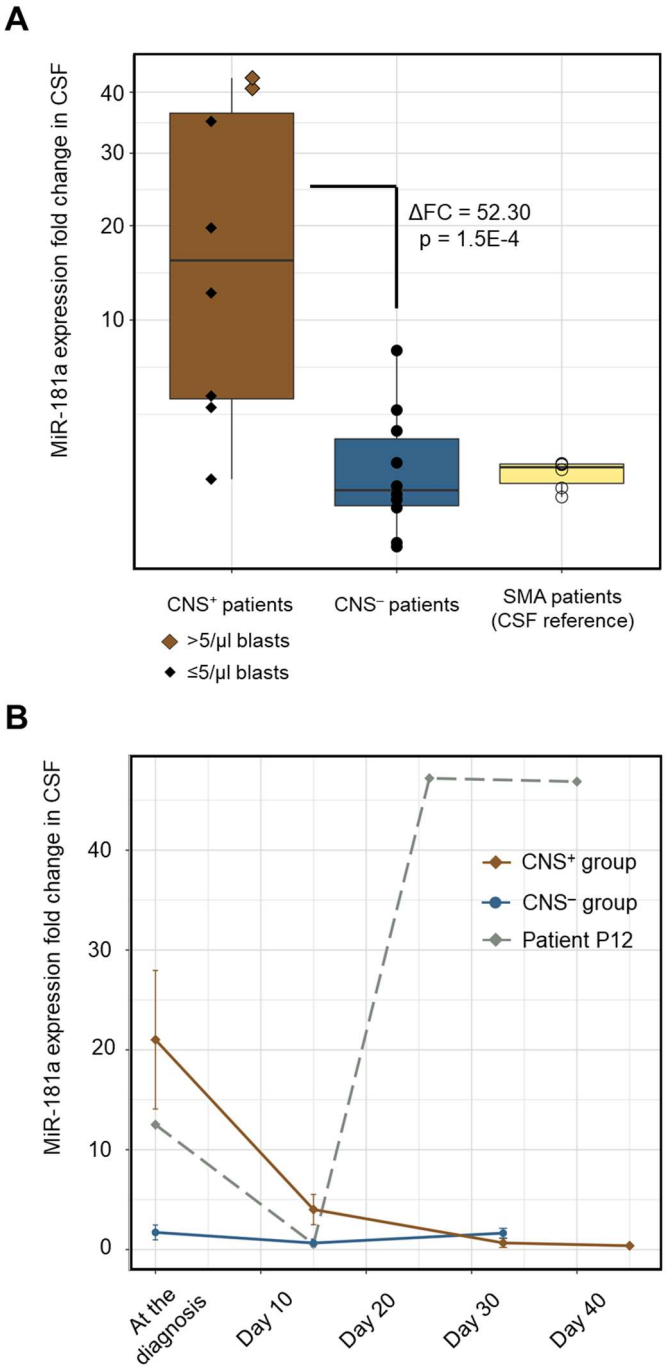

Figure 5. MiR-181a-5p level is a central nervous system involvement indicator in pediatric acute lymphoblastic leukemia. (**A**) Each dot indicates the miR-181a-5p expression of a patient. Horizontal line in box plots represents median fold change (FC), extent of boxes indicates upper and lower quartiles, whiskers show maximum and minimum values. (**B**) Lines are dedicated to follow the FC alterations in patient groups with and without meningeal leukemia and in a highlighted patient with progressive disease course during induction chemotherapy. Points indicate average FC values, while whiskers mark standard errors. CSF, cerebrospinal fluid; CNS, central nervous system; SMA, spinal muscular atrophy; ΔFC, difference in fold change. Own representation, published in [68].

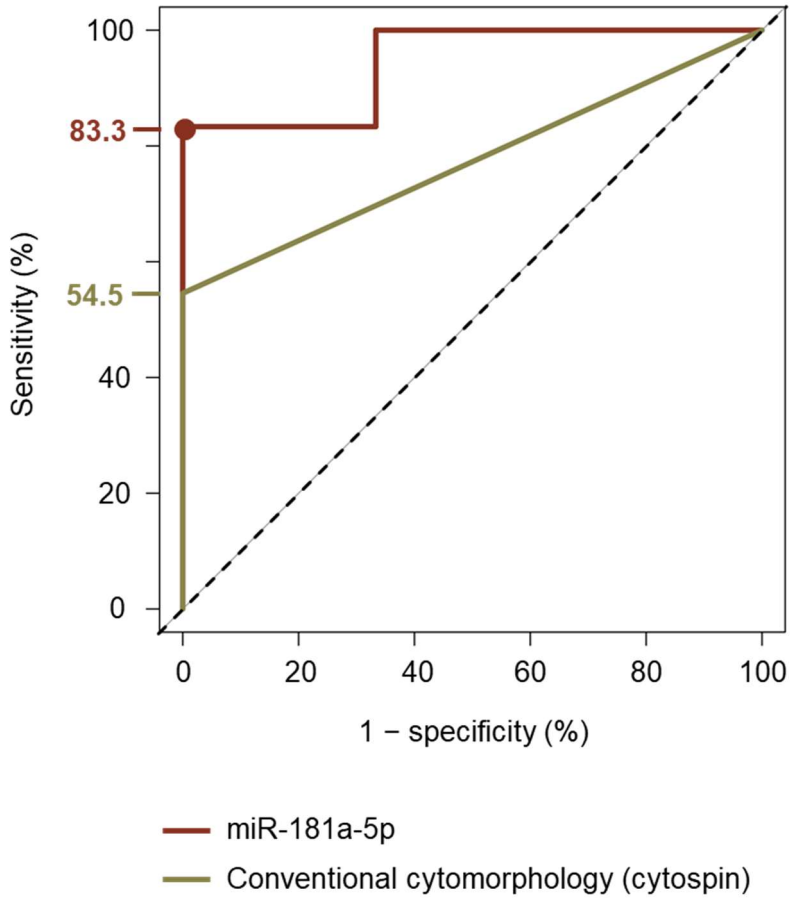

Figure 6. Comparison of receiver operating characteristic (ROC) curves of miR-181a-5p and conventional cytospin methods. Curve of conventional cytomorphology was produced based on information found in the literature (reviewed in de Graaf et al. 2011), where diagnostic efficacy of cytomorphology was assessed being aware of flow cytometry data. Own representation, published in [68].

III.3. Assessment of bone marrow and peripheral blood miR-181-family expression to indicate CNS status in ALL

We studied whether the expression levels of miR-181a-5p and miR-181b-5p in PB and BM samples could provide diagnostic or predictive information on CNS leukemia involvement. Initial PB miR-181a-5p level significantly fell down by the 15th and the 33rd day of treatment in ALL patients regardless of CNS status (by day 15: $\Delta FC=-6.42$, $p=0.02$; by day 33: $\Delta FC=7.95$, $p=0.03$). However, there was no difference in miR-181a-5p expression at diagnosis between CNS^+ and CNS^- ALL groups in PB.

MiR-181a-5p expression levels in BM highly exceeded the corresponding CSF expression in CNS^- ALL patients at the diagnosis ($\Delta FC=21.97$, $p=0.007$), but there was

no such relation in the CNS^+ group. Initial BM expression of miR-181a-5p distinguished only those patients with CNS^+ and CNS^- ALL who were characterized by precursor B-cell immunophenotype (ΔFC=9.18, p=0.04). However, miR-181a-5p level in PB of control individuals was lower than in BM of pB-ALL patients in both CNS status groups (CNS^+ status: ΔFC=55.82, p=0.001; CNS^- status: ΔFC=6.08, p=0.04). Decrease in miR-181a-5p level from the diagnosis of ALL by the 33rd day of therapy was also detectable in BM samples not only among CNS^+ patients (ΔFC=-187.96, p=0.006), but also in the CNS^- group (ΔFC=-8.54, p=0.002).

III.4. Atypical small extracellular vesicles in CSF of ALL patients with CNS involvement

Whether extracellular vesicle (EV) production from parameningeal leukemic infiltration can be demonstrated by CSF sampling is not yet known. Immunomagnetic bead separation technique was applied to compare the number of $CD63^+$ particles (typically endosome-derived small EVs) in CSF samples of CNS^+ and CNS^- ALL patients (n=2 and n=2, respectively). Their measurable amount was very low and no difference was found in the proportion of $CD63^+$ beads (mean±SE: 1.0±0.2% in CNS^+ group vs. 1.1±0.1% in CNS^- group). Next, from both CNS^+ and CNS^- patient groups, a representative pB-ALL and a representative T-cell ALL (T-ALL) CSF sample processed by ultracentrifugation (nonselective for cluster of differentiation markers) were supervised by immunolabelling for CD63 and CD81 markers using transmission electron microscopy (TEM). Interestingly, a considerable difference was observed in the density of vesicular elements between CNS^+ and CNS^- patients, as shown in Figure 7. High amount of EVs in the CNS^+ pB-ALL sample mostly showed $CD63^-/CD81^-$ characteristics (Figure 7A), whilst, there were low number of EVs in the CNS^- pB-ALL sample (Figure 7B). CSF samples of T-ALL cases showed quite similar pattern, but in this immunophenotype group we found a $CD63^-/CD81^-$ conglomerate of EVs in the CNS^- sample as well (Figure 7C-D). Presence of $CD63^+$ and $CD81^+$ small EVs in CSF, even if a low number, were confirmed in both CNS groups and immunophenotypes.

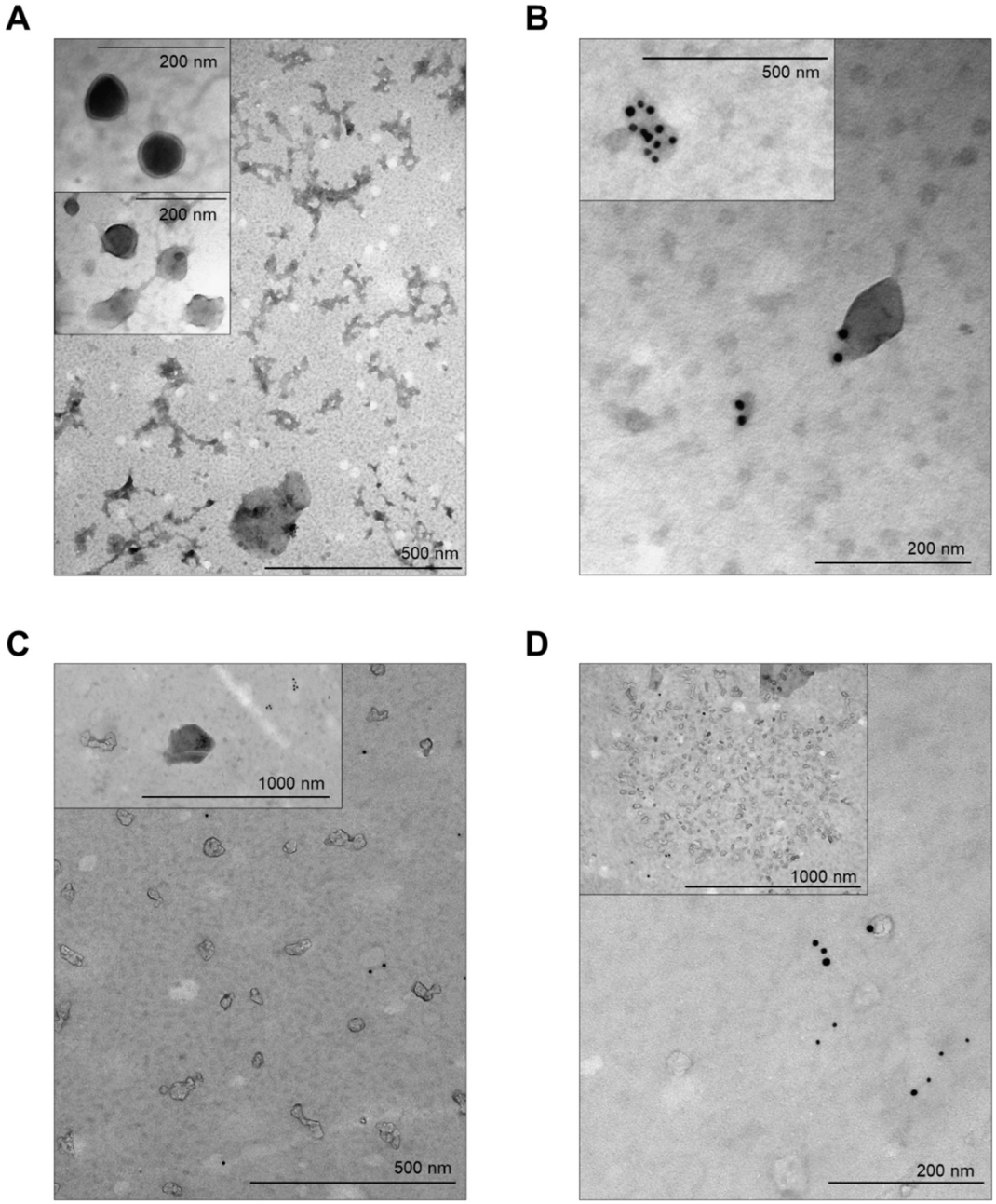

Figure 7. Atypical extracellular vesicles in the cerebrospinal fluid of patients with different central nervous system (CNS) status. (**A**) Transmission electron microscopy (TEM) images of a cerebrospinal fluid (CSF) sample from a CNS^+ patient with precursor B-cell acute lymphoblastic leukemia (pB-ALL) depict several small extracellular vesicles (EVs). Inserted panels show the absence of CD63/CD81 immunogold labelling which is a key feature of these EVs. (**B**) TEM images taken from the CSF sample of a patient with CNS^- pB-ALL. Appearance of gold particles refer to CD63 (10 nm) and CD81 (5 nm) positivity of the EVs. (**C**) Small EVs in a patient with relapsed CNS^+ T-cell ALL show mainly the absence of CD63/CD81 immunogold labelling. (**D**) $CD63^+/CD81^+$ small EVs in CSF from a patient with T-cell ALL without CNS involvement. Conversely, inserted slide shows a conglomerate of EVs without typical markers of exosomes originating from multivesicular body. Own photographs, published in [68].

III.5. Retrospective pharmacogenetic analysis of anthracycline-induced cardiotoxicity (ACT) in a considerable Hungarian childhood cancer population

In our pharmacogenetic study, altogether 70 SNPs were genotyped (see more information in the additional file of [73]). Genotype distributions were in Hardy-Weinberg equilibrium except for one SNP (*AKR1A1* rs2934859) which was excluded from the analysis. Genotyping was unsuccessful in the case of three SNPs. Thus, the genotyping results of 66 SNPs were used for the evaluations. Minor allele frequencies (MAF) in our population were found to be more than 7% for all of the SNPs. The children with ALL (n=622) involved in this study had undergone chemotherapy between 1989 and 2015 in six treatment centers of the Hungarian Pediatric Oncology Network. To raise a validation cohort with patients also treated with ANTs according to a different chemotherapy scheme, this ALL population was extended with a group of pediatric osteosarcoma (OSC) patients (n=39) in case-control analysis of ACT [73]. Detailed patient characteristics are shown in [73].

III.5.1. Tendentious decrease in the heart function in typical time points after anthracycline (ANT) exposure

Left ventricular fractional shortening (LVFS) measurements by ECHO were performed before the initiation of therapy, several times during the treatment and annually after finishing treatment. LVFS data were analyzed in follow-up categories as shown in Table 3. Not all of the ECHO records were available, because of the retrospective data collection. Only the latest ECHO of each patient was used in each follow-up category. Individual LVFS values were followed and the intra-patient alteration was calculated in every category compared to the individual value at diagnosis. These LVFS differences were then dichotomized (‘decreased’ or ‘not decreased’) patient by patient (Table 3). Odds ratios (OR) were calculated from a baseline point of view represented by the second follow-up category (‘<1 year from diagnosis’). Thus, OR > 1 means an increased odd to decreased LVFS, while OR < 1 refers to better expected heart function compared to echocardiographic state in the first year of chemotherapy. As seen in Table 3, active ANT treatment end point (3rd follow-up category, 'early-onset' ACT) and >15 years from cancer diagnosis (8th follow-up category, 'late-onset' ACT) seem to pose a statistically non-significant risk for worsened ejection capacity of the heart.

Table 3. Follow-up categories with echocardiography parameters. The decrease of LVFS was calculated patient by patient in every category compared to the individual value at diagnosis if these data were available. [1] Number of patients with a decreased LVFS per number of patients with a not decreased LVFS. [2] Compared to the second category ('<1 yr from diagnosis'). Abbreviations: CI, confidence interval; LVFS, left ventricular fractional shortening; OR, odds ratio; n, number; SD, standard deviation.

Follow-up category	**Patients with ALL**	**Patients with OSC**	**Total population**		
	n (mean LVFS ± SD)	n (mean LVFS ± SD)	n (mean LVFS ± SD)	Decreased | not decreased LVFS, n[1]	OR (95% CI) [2]
At the diagnosis	358 41.5 ±6.1	29 39.6 ±4.4	387 41.4 ±6.0		
<1 yr from diagnosis	275 40.4 ±6.1	5 40.2 ±5.3	280 40.4 ±6.1	104 | 83	1.0
1-2 yr from diagnosis	46 41.4 ±6.0	3 39.9 ±2.3	49 41.3 ±5.9	19 | 10	1.5 (0.6-3.4)
End of the treatment	287 40.0 ±5.6	28 38.4 ±6.3	315 39.9 ±5.7	105 | 98	0.9 (0.6-1.3)
2-5 yr from diagnosis	229 40.4 ±5.7	35 38.1 ±5.2	264 40.1 ±5.7	77 | 73	0.8 (0.5-1.3)
5-10 yr from diagnosis	265 40.1 ±5.5	36 40.3 ±5.6	301 40.1 ±5.5	70 | 76	0.7 (0.5-1.1)
10-15 yr from diagnosis	133 40.4 ±5.4	19 39.9 ±5.2	152 40.3 ±5.3	24 | 36	0.5 (0.3-1.0)
>15 yr from diagnosis	24 37.6 ±7.5	8 40.7 ±6.1	32 38.4 ±7.2	5 | 3	1.3 (0.3-5.7)

III.5.2. Case-control and haplotype analysis of pathological left ventricular function induced by anthracycline use in pediatric ALL

The worst heart function of each patient was used to define patients for the case-control study. Cases were those who had echocardiograms with LVFS ≤ 28% (pathological LVFS) at any time point during the follow-up (n=20). Patients who received the same chemotherapy but never had LVFS ≤ 28% were regarded as controls (n=641). To assess the association of the genotypes with the ACT, the genotype and allele frequencies in the two groups were compared. As multiple comparisons were conducted, we applied the Benjamini-Hochberg false discovery rate (FDR) method with type I error rate of 10% as correction of p-values, thus, p = 8.90E-03 was considered as the upper limit of statistical significance.

Multi-adjusted logistic regression was applied to analyze the full cohort. The analyses were adjusted for potential confounders, which were age at the time of diagnosis (years), gender (male or female), chemotherapy protocols (before 2000, after 2000 and OSC protocols; also reflects radiotherapy), risk groups (standard, intermediate, high risk) and cumulative dose of ANTs (≤ or > 240 mg/m2). The case-control analysis was performed for ALL patients in case of all SNPs. Risk of pathological LVFS was significantly influenced by SNPs in *CYP3A5* and *SLC28A3* genes. *CYP3A5* rs4646450 TT was 17% among cases and 3% in ALL patients without pathological FS (p = 5.60E-03; OR = 6.94 [1.76-27.39]). *SLC28A3* rs7853758 AA was 12% in cases, while only 1% among controls with ALL (p = 6.50E-03; OR = 11.56 [1.98-67.45]). These two SNPs were analyzed in the whole population including both ALL and OSC patients.

The genotype distribution of the *CYP3A5* rs4646450 differed significantly between cases and controls in the combined cohort (ALL and OSC patients) as well (p = 4.81E-03; OR = 7.25 [1.83–28.78]). Among cases (n=20) 15% had TT genotype while this value was 2.8% in controls. The genotype distribution of the *SLC28A3* rs7853758 SNP was not different between cases (10.5%) and controls (1.3%) in the combined cohort (p = 1.00E-02; OR = 9.837 [1.73–56.02]) if considering the adjusted p value (Benjamini-Hochberg false discovery rate method).

Subsequently, it was investigated whether *CYP3A5* rs4646450 was associated with cardiotoxicity in various subpopulations determined by clinical characteristics of the patients (Figure 8). Univariate logistic regression was performed on all of the patient

subgroups displayed in Figure 8. As a result of this analysis, the *CYP3A5* rs4646450 rare homozygous TT genotype showed an association with cardiotoxicity in patients with ALL (p = 7.00E-03; OR = 6.56 [1.68-25.71]). Similar association was found when analyzing only male patients (p = 4.00E-03; OR = 13.45 [2.26-80.1]), intermediate-risk patients (p = 2.00E-04; OR = 23.34 [4.46-122.07]), patients who received higher ANT doses (p = 5.00E-03; OR = 20.83 [2.49−174.3]) or patients who suffered relapse (p = 6.00E-03; OR = 48.00 [2.96−778.53]).

Haplotype analyses were carried out to study the association of haplotype blocks of genes in ACT, in other words, to identify combinations of genetic variations having an effect on pathological decrease of LVFS greater than SNPs have separately. Estimated haplotype frequencies in the case and control groups and haplotype-specific OR were determined by Haploview 4.1 software [74]. One haplotype block (TA) of *CYP3A5* proved to influence the ACT in the ALL cohort (p = 7.71E-2; χ^2 test in Haploview software). Rs4646450 and rs776746, members of this haplotype block (Figure 9), showed moderate linkage (D' = 0.9; r^2 = 0.48; LOD = 65.03)*. The effect of TA haploblock on cardiotoxic risk was nominally significant when involving clinical confounders and patients with ALL as well as OSC (Table 4).

* The term of 'linkage disequilibrium' (LD) characterizes the odd of two alleles to be inherited together in a population. The 'linkage disequilibrium' denomination refers to the non-random distribution of two linked (not independently descended) loci in the population compared to each other, that is, the loci are said to be in disequilibrium. Two major coefficients quantifying LD are normalized LD coefficient (D') and Pearson coefficient of correlation (r^2). Possible values of D' and r^2 have to be in a range of [0;1], where 1 refers to complete linkage of the loci. Logarithm of odds (LOD): is a statistical estimate of whether two loci are linked in comparison with the odd of the zero linkage between these loci.

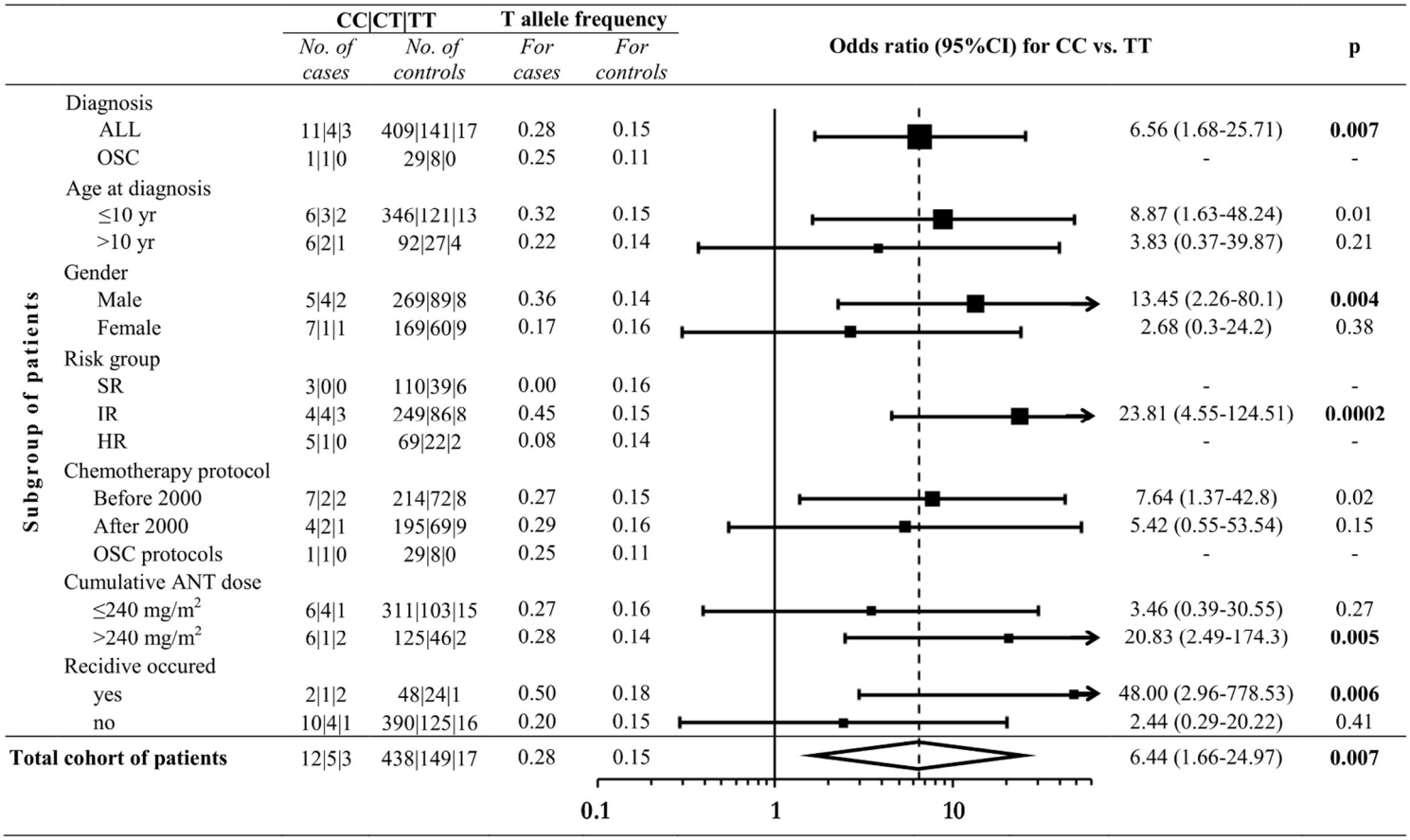

Figure 8. Odds ratios for cardiotoxicity associated with the *CYP3A5* rs4646450 genotype among subgroups of patients and in the whole cohort as well. Subpopulations are determined based on the following factors: diagnosis, age at diagnosis, gender, risk group, chemotherapy protocol, cumulative ANT dose, relapse occurred. Black boxes represent OR, the number of cases is proportional with the width of the boxes. The lengths of the horizontal lines depict the 95% confidence intervals. Analysis of OR was not accomplished if the number of cases was 0. Abbreviations: ALL, acute lymphoblastic leukemia; OSC, osteosarcoma; No, number; yr, year; SR, standard-risk; IR, intermediate-risk; HR, high-risk; OR, odds ratio. Own representation, published in [73].

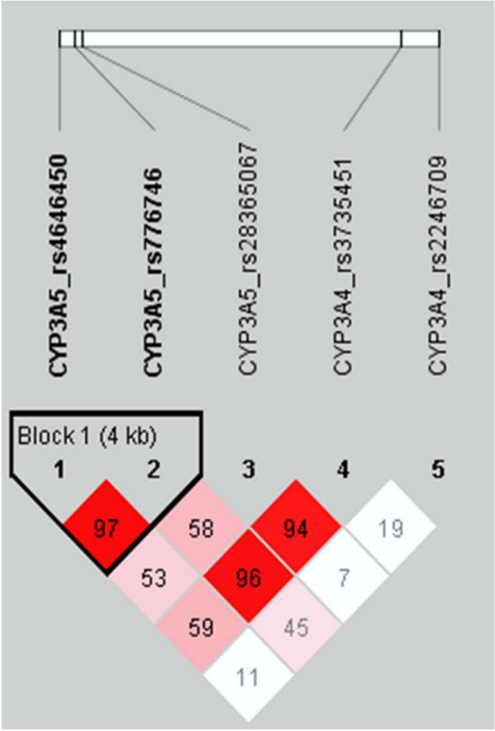

Figure 9. D'/LOD diagram resulted from haplotype analysis of *CYP3A5* and *CYP3A4* genetic variations studied in this thesis. These two genes sit near to each other on human chromosome 7. The pairwise linkage disequilibrium is expressed by D' (normalized LD coefficient, values between 0 and 1). Numbers displayed in squares are D' values multiplied by 100. Background color of squares changes in parallel with the LOD value: the darker the red color is, the stronger the linkage is (in case of LOD<2 the background color is white). Abbreviations: LD, linkage disequilibrium; LOD, logarithm of odds. Own representation that was not published before.

Table 4. Risk of anthracycline-induced cardiotoxicity considering the haplotype blocks of *CYP3A5*. *Multivariate logistic regression, regression model includes the interaction of rs4646450 and rs776746. Abbreviations: OR, odds ratio; CI, confidence interval; SNP, single nucleotide polymorphism.

Gene	SNP	Haplotype	Incidence (%)	Frequency in		p value*	OR (95% CI)*
				cases	controls		
CYP3A5	rs4646450 rs776746	CG	84.5%	0.763	0.847	0.200	2.14 (0.67−6.83)
	rs4646450 rs776746	**TA**	8.2%	**0.184**	**0.079**	**3.05E-2**	**14.26 (1.28−158.54)**
	rs4646450 rs776746	TG	7.1%	0.053	0.205	0.107	6.18 (0.67−56.68)

III.5.3. Potential interactions in SNP–SNP and SNP–clinical variable networks determining ACT risk

Widely known frequentist statistical approaches (e.g. logistic regression used in our case-control analysis above) are unable to quantify how confident one should be that a given SNP is truly associated with a phenotype (e.g. cardiotoxicity). Conversely, Bayesian statistics provides a rational and quantitative way to incorporate biological information, and it can allow for a range of possible genetic models in a single analysis [75]. Bayesian methods rely on the interchangeability of cause and effect (antecedent and

outcome). To date, it is accepted that genetic variations are not independent entities when determining a phenotype, they are rather parts of a hierarchical network. In our analysis, we searched for a directed acyclic graph, so called Bayesian network (BN), which describes the causality structure of pathological heart function and it's strongly relevant factors (SNPs and clinical variables). Thus, vertices of the graph are the random variables, such as SNPs or clinical factors. Based on their probability distribution (*a priori* knowledge; e.g. relative frequency of an SNP's genotypes in the population), Bayesian statistics can tell us the dependency of variables connected with edges in the graph (*a posteriori* knowledge; e.g. an SNP's genotype frequencies can only be determined when the genotype is known at another SNP). As an example, a BN is shown in Figure 10. In our study, the most probable BN with *a posteriori* probabilities (P) of the vertices and their connections (edges) was identified using Bayesian network based Bayesian multilevel analysis of relevance (BN-BMLA) method [76]. Values for posterior probability of strong relevance (P) range from 0 to 1, where $P = 1$ means that the probability of the given variable is 100% relevant with respect to the case-control status. Then, it was tested whether interactions exist between strongly relevant variables in respect of ACT as an endpoint. Interaction means that a variable (e.g. an SNP) can only be part of relevant variables' set, if another variable is already the part of this set.

BN-BMLA was performed for SNPs in the *ABCB1*, *ABCC1*, *ABCC2*, *ABCG2*, *AKR1A1*, *AKR1C3*, *CYP3A4*, *CYP3A5*, *GSTP1*, *HAS3*, *NQO1*, *NQO2*, *RARG*, *SLC22A17*, *SLC22A6*, *SLC22A7*, *SLC22A8* and *SLC28A3* genes along with cofactors. Our analyses revealed potentially strongly relevant effects of an SNP in gene *CYP3A5* (rs776746, $P = 0.42$), two SNPs in gene *NQO1* (rs1043470 and rs1469908, $P = 0.42$ and 0.34, respectively), two SNPs in gene *SLC28A3* (rs7853758 and rs885004, $P = 0.55$ and 0.36, respectively), and several cofactors (age at the time of diagnosis, $P = 0.72$; gender, $P = 0.44$; risk group, $P = 0.73$; diagnosis [ALL vs. OSC], $P = 0.8$ and cumulative dose of anthracycline, $P = 0.64$). BN structured from these vertices is shown is Figure 11. Besides, several interaction effects were found between the variables (Figure 12). Among these, the two SNPs (rs7853758 and rs885004) in gene *SLC28A3* showed the strongest interaction. It may suggest the refinement of the previously reported predisposing role of *SLC28A3* rs7853758 A allele in ACT by their interaction with the rs7853758 polymorphic locus.

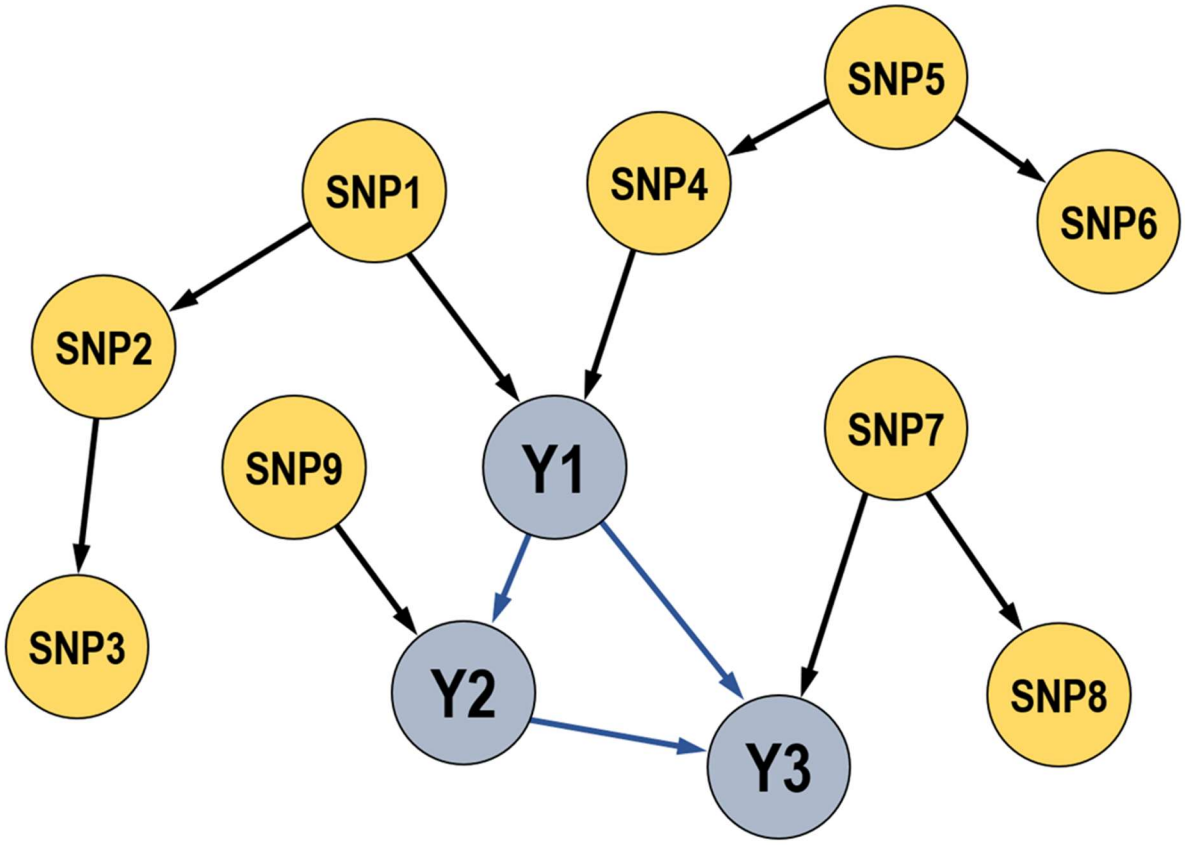

Figure 10. Relation types between variables in a Bayesian network (BN). The directed acyclic graph (DAG) depicted in the figure represents the structure of a BN. Yellow nodes mark genetic variations (single nucleotide polymorphisms, SNPs), grey nodes are featured target variables (e.g clinical covariants). Edges connecting vertices represent linkage between two random variables. Pairwise relation types: (*i*) direct causal relevance (e.g. between SNP4 and Y1, having common edges); (*ii*) pure interactionist relevance (e.g. between Y1 and SNP9, having a common child [Y2]); (*iii*) strong relevance (e.g. variable set consisting of Y2's parents, its children, and the other parents of its children is {Y1,SNP9,Y3,SNP7}); (*iv*) transitive causal relevance (e.g. between Y3 and SNP5, as two directed paths can connect them); (*v*) confounded relevance (e.g. between SNP3 and Y2, as they are connected by undirected path); (*vi*) association (e.g. between SNP3 and Y2 or between SNP5 and Y3). Markov blanket (MB): a variable's MB includes those variables correlate directly with this variable, thus, hide other in-model variable's effect from this variable. For example, MB of Y1 incudes variable set {SNP1,SNP4,Y2,Y3,SNP9,SNP7}. As values of this set are known, Y1 becomes independent of other nodes of the BN. Own representation based on our research group's previous work: [77].

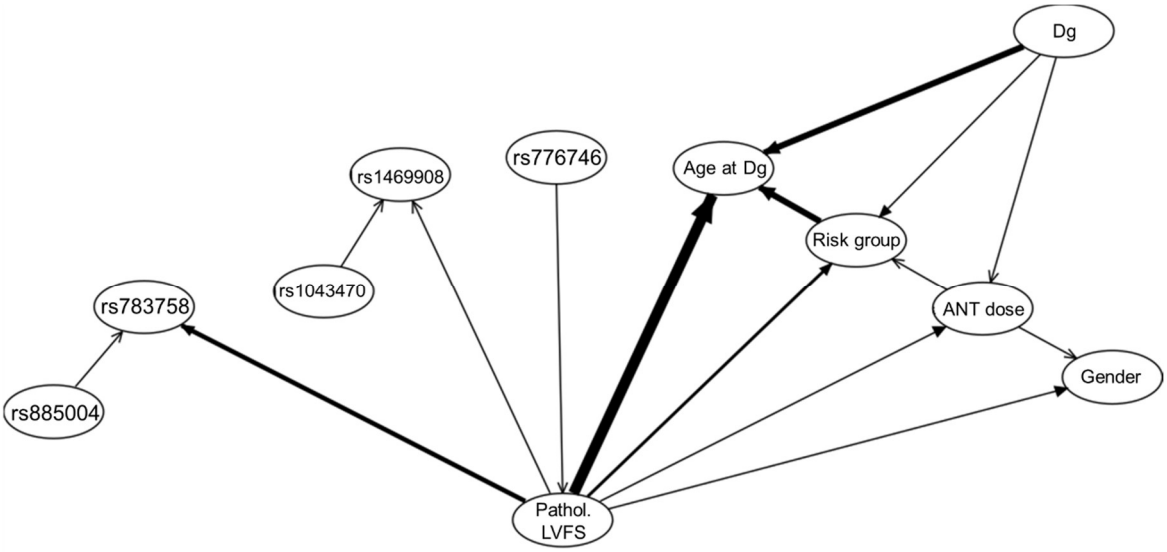

Figure 11. Relations of pathological left ventricular fractional shortening (LVFS) in the most probable Bayesian network. Only strongly relevant variables are displayed. Thickness of arrows is directly proportional to *posterior* probabilities. Abbreviations: ANT, anthracycline; Dg, diagnosis. Own representation that was not published before.

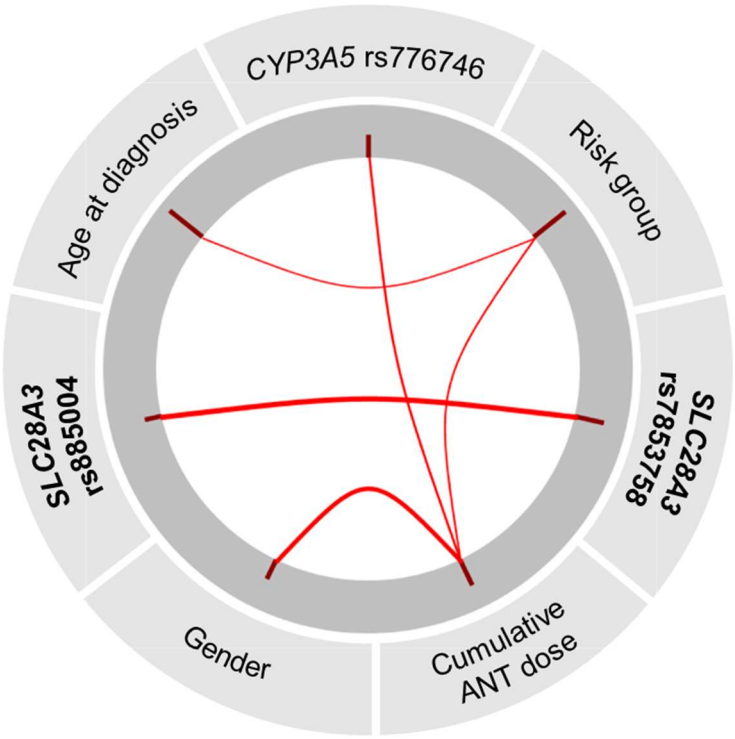

Figure 12. The most probable interactions between strongly relevant variables in anthracycline-induced cardiotoxicity by BN-BMLA. Red lines connect the members of and interaction. Thickness of lines is directly proportional to the probability of the interaction. Abbreviations: ANT, anthracycline. Gene names are unfolded in 'List of Abbreviations'. Own representation that was not published before.

IV. Discussion

Contrary to the overall survival rate of 80-90% among children with ALL, relapsing disease course, leukemic infiltration of the central nervous system (CNS) niche or treatment-limiting toxicities during chemotherapy are associated major therapeutic obstacles. Novel biomarkers (e.g. microRNAs, EVs or SNPs) assisting these challenging situations can contribute to better clinical care if standardized laboratory workflows are established.

IV.1. Peripheral blood (PB) miR expression in diagnostic childhood ALL samples (related to section III.1.1.)

MicroRNAs (miRs) were found to be highly stable in all body fluids, hence, could be utilized as clinical biomarkers [78]. Platelet-free plasma (PFP) prepared from PB samples of pB-ALL patients was analyzed, because platelets are the largest source of miRs in the blood which could have an interfering effect on our study [79,80]. We are not aware of any previous study which would have evaluated PFP miR expression in pediatric ALL. According to our results detailed in my thesis, candidate miR expressions did not differentiate between diagnostic, high blast burden PB PFP samples (*de novo* and relapsed disease periods), but their levels were well quantifiable in these cell-free liquid biopsy specimens. This led our research group to investigate (*i*) whether PB PFP miR expressions show MRD-like alterations during ALL treatment [61,71] and (*ii*) whether other biofluids (e.g. CSF) from other disease niches (e.g. CNS) also contain these blast-specific miRs [68].

Previously, based on platelet-containing plasma measurements, it was proposed that the risk of relapse in childhood ALL can be predicted by miR expression results [81,82]. This group found differences in miR expressions at the diagnosis of *de novo* versus relapsed ALL, but most leukemia-relevant miRs showed similar expression in those groups and differed uniformly from samples collected in complete remission. The set of examined miRs was partly overlapping with our study. With the use of PFP samples, we ruled out the influence of abundant platelet-derived miRs in PB plasma. As we could not separate the blast-rich stages of the patients examined by miR profiles, we rather support the investigation of treatment-related (chemotherapy-induced) changes in miR levels as

indicators of the disease course. A longitudinal assessment of expression levels may correlate with intra-individual treatment response and could be utilized in MRD estimation or early detection of molecular relapse, as it was scrutinized in ALL cells by other authors [83].

IV.2. MicroRNAs as markers of CNS involvement of childhood ALL (related to sections III.1.2.-III.3.)

We still lack methods to follow up CNS leukemia with appropriate sensitivity. There are two major clinical enigmas related to CNS disease: (*i*) how can we better screen occult meningeal leukemia and thereby prescribe adequate CNS-directed treatment for those who need it, and (*ii*) may we apply novel molecularly targeted therapies against CNS infiltration in the era of precision oncology? Whether the initial CNS leukemia burden or its dynamics during treatment have more prognostic value is not fully understood. To date, only one study has reported that miRs can be measured in the CSF collected from children with leukemia [84].

CSF miR-181a-5p expression levels were unexpectedly high in patients with cytologically confirmed CNS^+ ALL, while levels in those lacking meningeal involvement were similar to the levels of control individuals. MiR-181a serves as a regulator of normal hematopoiesis and its disruption has been linked to various types of cancer, including hematological malignancies [85]. It is one of the most abundant miRs in lymphocytes [86]. In acute leukemia, the exact role of miR-181a is still unclear as contradicting articles claim its oncogene as well as tumor suppressor effects [87–90]. Besides, several mechanisms of the spread of ALL cells to the CNS have been revealed recently and can be influenced by miR-181a regulatory actions. Of particular interest to our findings, overexpression of vascular endothelial growth factor A (VEGFA), a miR-181a regulated hypoxia-responsive gene, was shown to provide survival advantage to the CNS-penetrating ALL cells in the hypoxic, nutrient-poor microenvironment of the CNS. This phenomenon was described in both primary pre-B ALL cells isolated from the CSF of children with CNS involvement and a mouse primograft model for CNS^+ ALL [91–93]. The positive correlation between the miR-181a expression and VEGFA production in a non-epithelial tumor tissue was proposed through the scrutiny of human chondrosarcoma cells [94]. A detailed explanation how miR-181a promotes the SRC/VEGF signaling

pathway was shown in preclinical models of colorectal cancer [95]. A similar mechanism might work in ALL, too. Yet, the nature of CNS disease evolution is not fully understood. While the VEGFA mechanism suggests increased vascular permeability and reduced barrier function of endothelium to make cell migration possible into the CNS, a novel study claims that the whole process happens abluminally, along emissary bridging vessels based on integrin alpha-6 (ITGA6)–laminin interaction [96]. Interestingly, ITGA6 may interact with miR-181a according to the mirDIP v4.1 integrative miR target predictor database [97].

Whether the ability to invade the CNS is a generic property of all ALL blasts or only a subpopulation of cells can selectively enter this compartment is a subject of recent debate. We found that BM miR-181a-5p expression of CNS– patients (*i*) exceeded the average level of CNS^- CSF and control samples and (*ii*) responded to the chemotherapy with significant reduction, similarly to the CNS^+ BM samples. Thus, our data tend to support the theory that ALL cells in the BM compartment have a common ability to reach meningeal surfaces, as previously proposed [98,99]. Also, the seeding of CNS leukemia may depend on a dynamic relationship between extramedullary microenvironment and blast cells which is influenced by regulatory elements (such as miR-181a, possibly), rather than predetermined cell-intrinsic and clone-specific 'CNS-tropic' features (e.g. cell surface protein composition). Basically, it was out of our aims to identify the cellular origin of miR-181a-5p in this study setting. Yet, there are some factors which suggest the examination of leukemic blasts as potential miR-181 secretors: (*i*) miR-181a is a well-known small RNA produced by lymphoid cells [86]; (*ii*) miR-181a was found in significant amount in blast-rich BM samples, regardless of CNS status; (*iii*) in CNS^- CSF samples, where blasts were absent, the average expression of miR-181a was very low. However, it is also a possibility that miR-181a originates from non-leukemic cells communicating with leukemic blasts (e.g. stromal cells in BM or resident cells in CNS leukemic niche) and this phenomenon might have been found in our samples. There are plenty of evidence that the miR-181-family plays role in the modulation of tissue differentiation, remodeling and degeneration in the CNS, moreover, it has anti-inflammatory effect in the context of neuroinflammation [100].

According to our data, miR-181a-5p may provide about 35% benefit of sensitivity compared to the cytospin method when diagnosing ALL infiltration in the CNS at the

initiation of chemotherapy. However, an article described that occult meningeal disease during therapy resulted in significantly higher rates of relapses and death in pediatric ALL, while no such influences were found with initial CSF evaluation [101]. Thus, the alteration of initial miR-181a-5p level might be considered to estimate prognosis.

Candidate miR expression analysis of PB samples identified no benefit of measuring miR-181-family levels to diagnose CNS involvement. In contrast, we showed the possible role of diagnostic BM miR-181a-5p expression in determining CNS status, however, it was demonstrated solely in pB-ALL patients. This finding might propose a less invasive alternative to lumbar puncture as BM aspiration is also performed at CSF sampling time points in the clinical routine. However, higher BM miR-181a-5p level may be associated with general miRNomic disturbance under leukemic conditions as CNS^{-} patients showed elevated relative expression as well. We rather suggest the investigation of miR-181a-5p level alteration in time as it may monitor the capacity of CNS-attacking blasts of the BM compartment. To indicate subsequent CNS relapse by BM examination, cellular proteins [102] and miR pattern [103] in BM cells of ALL patients have already been identified, but venipuncture-based CNS involvement markers have not been published before.

IV.3. Extracellular vesicles (EVs) in CSF of children with CNS leukemia (related to section III.4.)

CSF-based EV studies in leukemia are in their early infancy. According to our information, this is the first time when small EV composition of CSF was assessed in relation with CNS disease state in acute leukemia patients. Atypical small EVs were found in unexpectedly high density in CNS^{+} diagnostic CSF samples. The phenotype of these vesicles is not fully understood. They were negative for CD63 and CD81, which contradict the multivesicular body origin [104]. Nonexosomal subpopulation of small EVs was previously demonstrated [105], and most of the EVs in CNS^{+} CSF reached the lower limit of the microvesicle size range. Otherwise, it is unlikely that these bodies are lipoproteins with EV-like appearance as CSF lipoprotein content is minimal [106]. In all, this experiment with CSF sEVs is definitely preliminary and needs further investigation.

IV.4. *CYP3A5* and *SLC28A3* SNPs in anthracycline-induced cardiotoxicity (ACT) risk prediction (related to section III.5.)

In our joined (patients with ALL or OSC) population the intronic *CYP3A5* rs4646450 SNP's TT genotype was associated with low LVFS (<28%) with a significantly higher OR value in males indicating gender related differences. Both the male and female gender has been already shown as risk factors for developing cardiotoxicity [107–109]. *CYP3A5* gene (7q21.1) is a member of cytochrome P450 proteins involved in drug metabolism, synthesis of steroids and lipids. *CYP3A5* is expressed in the liver and also in extrahepatic tissues, e.g. in intestines. Genetic variability of *CYP3A5* is high; it is not expressed in 20% of African and 80% of Caucasian population. Previous studies in the literature did not find association of *CYP3A5* rs4646450 with cardiac parameters [110]. Currently, the function of this SNP is not known, however, in a study rs4646450 was in correlation with reduced protein-expression and activity of CYP3A4 in human liver [111].

Our BN-BMLA analysis revealed potential strongly relevant effect of rs776746 SNP in gene *CYP3A5* with respect to the case-control status of the patients. SNPs in *CYP3A5* (rs776746 and rs10264272) may modify its alternative splicing and protein truncation, which can result in a less active *CYP3A5* [112]. Huang et al. studied the CYP3A5 enzyme activity with the presence of its different gene polymorphisms in pediatric ALL patients. They revealed that patients with rs776746 G allele (*CYP3A5*3*, c.6986A>G) had lower enzyme activity. Rs776746 genotype was in association with the mRNA expression, daunorubicin plasma concentration and adverse drug reactions. In this investigation the AUC of daunorubicin was higher in children with cardiotoxicity [113]. Interestingly, contradictory results were seen in diffuse large B-cell lymphoma patients as *CYP3A5* rs776746 AG/AA seemed to increase the risk of grade 2-4 cardiac toxicity compared to GG genotype [114].

The genotype distribution of the *SLC28A3* rs7853758 was also significantly different between cases and controls in our multivariate logistic regression analysis. In many studies *SLC28A3* rs7853758 has been shown to be a very important protective genetic marker against ACT, its minor allele (A) found more often in controls than in patient cases [66,115]. This SNP was recommended for clinical use in pharmacogenetic testing before using doxorubicin or daunorubicin in the treatment of pediatric cancer patients treatment [116]. This correlation for chronic cardiotoxicity was not significant in another

cohort [117], nor the RICOVER-60 trial (elder patients with aggressive B-cell lymphoma) found association with adverse cardiac reactions [118]. In contrast to these analyses, in our cohort the SLC28A3 rs7853758 AA genotype was more frequent among cases. However, not only in the case-control study but also in the BN-BMLA we could confirm the importance of this variant which needs further validation in larger cohorts.

IV.5. Biases in our genetic marker studies of childhood ALL

Several limitations should be noted when interpreting our results. Lacking standards regarding normalizing miRs and only partial correlation of various PCR based methods which are widely known difficulties of RNA and miR laboratory methodologies contribute to the limitations of our work. Since CNS involvement is a rare condition in childhood leukemia, the difficulty of obtaining CNS^{+} samples resulted in low patient numbers. Our results should be verified in larger cohorts before concluding any practical benefit. Another bias was the selection of patients with and without the meningeal disease, as it relied on conventional cytologic assessment tools lacking appropriate sensitivity.

In our pharmacogenetic study, not all of the ECHO records were available due to retrospective data collection. Therefore, the analysis of ECHO data was not possible for each year; categories of follow-up had to be generated. Only the data of the latest ECHO of each patient were used in each follow-up category, the redundant echocardiography measurements were excluded. Controls have a relatively wide range (1-7 occasions) of echocardiographic assessments in our cohort (21% of the patients had only one ECHO and 60% of patients had 3 or more ECHOs). It is also notable that the average LVFS slightly decreased at the population level during our follow-up time (Table 3), however, it was still within the normal range. According to former Common Terminology Criteria for Adverse Events (CTCAE), the left ventricular systolic dysfunction is considered when LVFS value is below 30% [116]. Previous studies confirmed that a subclinical change in LVFS has also a prognostic value on later development of heart failure [119,120]. Yet, we applied a quite low limit value of pathological LVFS (≤28%) in our case-control study, because we tried to find association between manifest systolic function impairment and SNPs. On the other side, there was a need for the genetic association analysis of subclinical changes in LVFS as well, which was covered by the next step of our study [73].

V. Conclusions

In this thesis, identification of stable nucleic acid-based markers associated with leukemia severity and treatment-induced toxicity was in the focus. Experiments and analyses are based on a continuously growing biobank containing more than 800 DNA and more than 300 RNA samples from children diagnosed with acute leukemia in Hungary. Our results propose the potential advantage of measuring microRNA (miR) expression and genotyping single nucleotide polymorphisms (SNPs) in patients with ALL in order to monitor CNS migration of leukemia cells and to predict cardiotoxic events, respectively. Conclusions of my work are as follows:

1. *De novo* and relapsed ALL at the time of diagnosis, two high-burden disease states, have similar expression pattern of leukemia-specific miRs in peripheral blood PFP samples.
2. MiR-181 family is a measurable leukemia-specific group of miRs in the CSF showing different expression in CNS^+ and CNS^- patient groups.
3. MiR-181a-5p level in CSF may indicate CNS infiltration of ALL with a high sensitivity at the time point of leukemia diagnosis and might provide a longitudinal marker for CNS disease monitoring.
4. Atypical, non-exosome small EVs are highly abundant in CSF samples from CNS^+ patients compared to CNS^- cases.
5. In a substantial Hungarian pediatric ALL cohort, *CYP3A5* rs4646450 TT genotype was significantly more frequent in those children with pathological LVFS (<28%) measured on follow-up echocardiography visits.
6. *CYP3A5* rs776746 seems to sit on the same haplotype as *CYP3A5* rs4646450, and rs776746 is part of the strongly relevant factor's network of anthracycline-induced cardiotoxicity (ACT).
7. An interaction between *SLC28A3* rs7853758 and *SLC28A3* rs885004 polymorphisms was observed, which complicates the previously described role of *SLC28A3* rs7853758 in ACT by other authors.

VI. Summary

Residual disease-based treatment intensity in leukemia-affected body compartments, such as the central nervous system (CNS), has resulted in improved survival in childhood acute lymphoblastic leukemia (ALL). However, procedures applied to obtain samples from patients are mainly invasive and sample analysis techniques are sometimes unreliable. MicroRNAs (miRs) easily traffic among fluid spaces, thus, can be suitable candidates for liquid biopsy purposes. Childhood ALL survivors need to face possible chemotherapy-induced organ toxicities, which may worsen their quality of life. Singe nucleotide polymorphisms (SNPs) may predict the individual risk for such toxicities.

My thesis focused on identifying nucleic acid-based markers indicating CNS disease or associating with anthracycline-induced cardiotoxicity (ACT) in childhood ALL. Expression of more than 40 leukemia-specific miRs were screened on serial cerebrospinal fluid (CSF), bone marrow (BM) and peripheral blood samples (n=174) with custom TaqMan^{TM} low-density array cards, and results were validated by unique qPCR runs. The main reservoir of miRs in body fluids, extracellular vesicles (EVs) were then characterized in CSF samples. Associations between longitudinal heart function data and genotypes of >60 SNPs determined by OpenArray^{TM} technology were analyzed in a substantial cohort (n=661).

The role of miR-181-family was identified in clustering CSF samples from patients with or without CNS disease (CNS^+ vs. CNS^-). MiR-181a-5p expression was about 52 times higher in CSF samples of CNS^+ ALL patients compared to CNS^- cases, furthermore, CNS^+ precursor B cell subgroup also had ninefold higher miR-181a-5p levels in their BM. CNS^+ CSF samples showed atypical $CD63^-/CD81^-$ small extracellular vesicles in high density, which were absent in CNS^- samples. *CYP3A5* rs4646450 TT was six times more frequent among ALL cases with pathological heart function than in patients with normal ejection values. *SLC28A3* rs7853758 and rs885004 are proved to be interacting SNPs in determining ACT risk.

After validating in extensive cohorts, quantification of miR-181a-5p might provide a novel tool to monitor CNS disease course and further adjust CNS-directed therapy. Genetic variants in transporters and metabolic enzymes might modulate the individual risk to cardiac toxicity after ALL treatment.

VII. References

1. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW. (2016) The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood, 127(20): 2391–2405.
2. Force LM, Abdollahpour I, Advani SM, Agius D, Ahmadian E, Alahdab F, Alam T, Alebel A, Alipour V, Allen CA, Almasi-Hashiani A, Alvarez EM, Amini S, Amoako YA, Anber NH, Arabloo J, Artaman A, Atique S, Awasthi A, Bagherzadeh M, Basaleem H, Bekru ET, Bijani A, Bogale KA, Car M, Carvalho F, Castro C, Catalá-López F, Chu DT, Costa VM, Darwish AH, Demeke FM, Demis AB, Demoz GT, Do HP, Murray CJL, Bhakta N, Fitzmaurice C. (2019) The global burden of childhood and adolescent cancer in 2017: an analysis of the Global Burden of Disease Study 2017. Lancet Oncol, 20(9): 1211–1225.
3. Garami M, Schuler D, Jakab Z. (2014) Az Országos Gyermektumor Regiszter jelentősége a gyermekonkológiai ellátásban. Orv Hetil, 155(19): 732–739.
4. Howlader N, Noone A, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis D, Chen H, Feuer E, Cronin K. SEER Cancer Statistics Review, 1975-2017, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2017/, based on November 2019 SEER data submission, posted to the SEER web site, April 2020.
5. Price R, Stranges E, Elixhauser A. (2012) Pediatric Cancer Hospitalizations, 2009: Agency for Healthcare Research and Quality Healthcare Cost and Utilization Project.
6. Inaba H, Greaves M, Mullighan CG. (2013) Acute lymphoblastic leukaemia. Lancet, 381(9881): 1943–1955.
7. Whitlock J, Gaynon P. Acute Lymphoblastic Leukemia in Children. In: Greer J, Foerster J, Rodgers G, Paraskevas F, Glader B, Arber D, Means R, editors. Wintrobe's Clinical Hematology, 12th ed. Philadelphia, PA: Lippincott Williams & Wilkins; p. 2009: 1889–1917.
8. Eden TOB, Pieters R, Richards S, and Childhood Acute Lymphoblastic Leukaemia Collaborative Group (CALLCG). (2009) Beneficial and harmful effects of

anthracyclines in the treatment of childhood acute lymphoblastic leukaemia: A systematic review and meta-analysis. Br J Haematol, 145(3): 376–388.

9. Margolin J, Rabin K, Steuber C, Poplack D. Acute lymphoblastic leukemia. In: Pizzo P, Poplack D, editors. Principles and Practice of Pediatric Oncology, Philadelphia, PA: Lippincott Williams & Wilkins; p. 518–564.
10. Hunger SP, Mullighan CG. (2015) Acute Lymphoblastic Leukemia in Children. Longo DL, editor. N Engl J Med, 373(16): 1541–1552.
11. Campana D, Pui CH. (1995) Detection of minimal residual disease in acute leukemia: Methodologic advances and clinical significance. Blood, 85(6): 1416–1434.
12. Waanders E, Gu Z, Dobson SM, Antić Ž, Crawford JC, Ma X, Edmonson MN, Payne-Turner D, van de Vorst M, Jongmans MCJ, McGuire I, Zhou X, Wang J, Shi L, Pounds S, Pei D, Cheng C, Song G, Fan Y, Shao Y, Rusch M, McCastlain K, Yu J, van Boxtel R, Blokzijl F, Iacobucci I, Roberts KG, Wen J, Wu G, Ma J, Easton J, Neale G, Olsen SR, Nichols KE, Pui CH, Zhang J, Evans WE, Relling MV, Yang JJ, Thomas PG, Dick JE, Kuiper RP, Mullighan CG. (2020) Mutational Landscape and Patterns of Clonal Evolution in Relapsed Pediatric Acute Lymphoblastic Leukemia. Blood Cancer Discov, 1(1): 96–111.
13. Eckert C, Hagedorn N, Sramkova L, Mann G, Panzer-Grümayer R, Peters C, Bourquin JP, Klingebiel T, Borkhardt A, Cario G, Alten J, Escherich G, Astrahantseff K, Seeger K, Henze G, Von Stackelberg A. (2015) Monitoring minimal residual disease in children with high-risk relapses of acute lymphoblastic leukemia: Prognostic relevance of early and late assessment. Leukemia, 29(8): 1648–1655.
14. Gaipa G, Basso G, Biondi A, Campana D. (2013) Detection of minimal residual disease in pediatric acute lymphoblastic leukemia. Cytom Part B - Clin Cytom, 84(6): 359–369.
15. Pui C-H, Evans WE. (2006) Treatment of Acute Lymphoblastic Leukemia. N Engl J Med, 354(2): 166–178.
16. Sundberg KK, Doukkali E, Lampic C, Eriksson LE, Arvidson J, Wettergren L. (2010) Long-term survivors of childhood cancer report quality of life and health status in parity with a comparison group. Pediatr Blood Cancer, 55(2): 337–343.

17. Robison LL. (2011) Late effects of acute lymphoblastic leukemia therapy in patients diagnosed at 0-20 years of age. Hematology Am Soc Hematol Educ Program, 2011(1): 238–242.
18. Diller L. (2011) Adult Primary Care after Childhood Acute Lymphoblastic Leukemia. N Engl J Med, 365(15): 1417–1424.
19. Thomas IH, Donohue JE, Ness KK, Dengel DR, Baker KS, Gurney JG. (2008) Bone mineral density in young adult survivors of acute lymphoblastic leukemia. Cancer, 113(11): 3248–3256.
20. Lipshultz SE, Cochran TR, Franco VI, Miller TL. (2013) Treatment-related cardiotoxicity in survivors of childhood cancer. Nat Rev Clin Oncol, 10(12): 697–710.
21. Oeffinger KC, Mertens AC, Sklar CA, Kawashima T, Hudson MM, Meadows AT, Friedman DL, Marina N, Hobbie W, Kadan-Lottick NS, Schwartz CL, Leisenring W, Robison LL. (2006) Chronic Health Conditions in Adult Survivors of Childhood Cancer. N Engl J Med, 355(15): 1572–1582.
22. Van Dongen JJM, Van Der Velden VHJ, Brüggemann M, Orfao A. (2015) Minimal residual disease diagnostics in acute lymphoblastic leukemia: Need for sensitive, fast, and standardized technologies. Blood, 125(26): 3996–4009.
23. Bálint E, Gábor K, Nóra K, Andrea R, Judit CS, János ED, Judit M, Ágnes FS. (2018) New and traditional directions in the biology and management of childhood acute lymphoblastic leukemia. Orv Hetil, 159(20): 786–797.
24. Wood B, Wu D, Crossley B, Dai Y, Williamson D, Gawad C, Borowitz MJ, Devidas M, Maloney KW, Larsen E, Winick N, Raetz E, Carroll WL, Hunger SP, Loh ML, Robins H, Kirsch I. (2018) Measurable residual disease detection by high-throughput sequencing improves risk stratification for pediatric B-ALL. Blood, 131(12): 1350–1359.
25. Pui CH, Howard SC. (2008) Current management and challenges of malignant disease in the CNS in paediatric leukaemia. Lancet Oncol, 9(3): 257–268.
26. Evans AE, Gilbert ES, Zandstra R. (1970) The increasing incidence of central nervous system leukemia in children. Children's cancer study group a. Cancer, 26(2): 404–409.
27. Laningham FH, Kun LE, Reddick WE, Ogg RJ, Morris EB, Pui CH. (2007)

Childhood central nervous system leukemia: Historical perspectives, current therapy, and acute neurological sequelae. Neuroradiology, 49(11): 873–888.

28. Nguyen K, Devidas M, Cheng SC, La M, Raetz EA, Carroll WL, Winick NJ, Hunger SP, Gaynon PS, Loh ML. (2008) Factors influencing survival after relapse from acute lymphoblastic leukemia: A Children's Oncology Group study. Leukemia, 22(12): 2142–2150.
29. Heidari N, Saki N, De Filippis L, Shahjahani M, Teimouri A, Ahmadzadeh A. (2016) Central nervous system niche involvement in the leukemia. Clin Transl Oncol, 18(3): 240–250.
30. Barredo JC, Devidas M, Lauer SJ, Billett A, Marymont MA, Pullen J, Camitta B, Winick N, Carroll W, Ritchey AK. (2006) Isolated CNS relapse of acute lymphoblastic leukemia treated with intensive systemic chemotherapy and delayed CNS radiation: A pediatric oncology group study. J Clin Oncol, 24(19): 3142–3149.
31. Ranta S, Nilsson F, Harila-Saari A, Saft L, Tani E, Söderhäll S, Porwit A, Hultdin M, Noren-Nyström U, Heyman M. (2015) Detection of central nervous system involvement in childhood acute lymphoblastic leukemia by cytomorphology and flow cytometry of the cerebrospinal fluid. Pediatr Blood Cancer, 62(6): 951–956.
32. Crespo-Solis E, López-Karpovitch X, Higuera J, Vega-Ramos B. (2012) Diagnosis of acute leukemia in cerebrospinal fluid (CSF-Acute Leukemia). Curr Oncol Rep, 14(5): 369–378.
33. Weiss R. (1992) The anthracyclines: will we ever find a better doxorubicin? Semin Oncol, 19(6): 670–686.
34. Lopaschuk GD, Jaswal JS. (2010) Energy metabolic phenotype of the cardiomyocyte during development, differentiation, and postnatal maturation. J Cardiovasc Pharmacol, 56(2): 130–140.
35. Volkova M, Russell R. (2012) Anthracycline Cardiotoxicity: Prevalence, Pathogenesis and Treatment. Curr Cardiol Rev, 7(4): 214–220.
36. Gratia S, Kay L, Potenza L, Seffouh A, Novel-Chate V, Schnebelen C, Sestili P, Schlattner U, Tokarska-Schlattner M. (2012) Inhibition of AMPK signalling by doxorubicin: At the crossroads of the cardiac responses to energetic, oxidative, and genotoxic stress. Cardiovasc Res, 95(3): 290–299.

37. Yang Y, Zhang H, Li X, Yang T, Jiang Q. (2015) Effects of PPARα/PGC-1α on the energy metabolism remodeling and apoptosis in the doxorubicin induced mice cardiomyocytes in vitro. Int J Clin Exp Pathol, 8(10): 12216–12224.
38. Zhang Y, Chen Y, Zhang M, Tang Y, Xie Y, Huang X, Li Y. (2014) Doxorubicin Induces Sarcoplasmic Reticulum Calcium Regulation Dysfunction via the Decrease of SERCA2 and Phospholamban Expressions in Rats. Cell Biochem Biophys, 70(3): 1791–1798.
39. Gambliel HA, Burke BE, Cusack BJ, Zhang YL, Olson RD, Gambliel HA, Burke BE, Cusack BJ, Zhang YL, Olson RD, Cusack BJ, Olson RD, Walsh GM, Olson RD, Mushlin PS. (2002) Doxorubicin and C-13 deoxydoxorubicin effects on ryanodine receptor gene expression. Biochem Biophys Res Commun, 291(3): 433–438.
40. Gianni L, Herman EH, Lipshultz SE, Minotti G, Sarvazyan N, Sawyer DB. (2008) Anthracycline cardiotoxicity: From bench to bedside. J Clin Oncol, 26(22): 3777–3784.
41. Sadurska E. (2015) Current Views on Anthracycline Cardiotoxicity in Childhood Cancer Survivors. Pediatr Cardiol, 36(6): 1112–1119.
42. Trachtenberg BH, Landy DC, Franco VI, Henkel JM, Pearson EJ, Miller TL, Lipshultz SE. (2011) Anthracycline-associated cardiotoxicity in survivors of childhood cancer. Pediatr Cardiol, 32(3): 342–353.
43. Minotti G, Salvatorelli E, Menna P. (2010) Pharmacological foundations of cardio-oncology. J Pharmacol Exp Ther, 334(1): 2–8.
44. Elliott P. (2006) Pathogenesis of Cardiotoxicity Induced by Anthracyclines. Semin Oncol, 33(Suppl. 8): 2–7.
45. Mordente A, Meucci E, Silvestrini A, Martorana G, Giardina B. (2009) New Developments in Anthracycline-Induced Cardiotoxicity. Curr Med Chem, 16(13): 1656–1672.
46. Bernstein D, Fajardo G, Zhao M, Urashima T, Powers J, Berry G, Kobilka BK. (2005) Differential cardioprotective/cardiotoxic effects mediated by β-adrenergic receptor subtypes. Am J Physiol Circ Physiol, 289(6): H2441–2449.
47. Roca-Alonso L, Castellano L, Mills A, Dabrowska AF, Sikkel MB, Pellegrino L, Jacob J, Frampton AE, Krell J, Coombes RC, Harding SE, Lyon AR, Stebbing J.

(2015) Myocardial MiR-30 downregulation triggered by doxorubicin drives alterations in β-adrenergic signaling and enhances apoptosis. Cell Death Dis, 6(5): e1754–e1754.

48. De Angelis A, Piegari E, Cappetta D, Marino L, Filippelli A, Berrino L, Ferreira-Martins J, Zheng H, Hosoda T, Rota M, Urbanek K, Kajstura J, Leri A, Rossi F, Anversa P. (2010) Anthracycline cardiomyopathy is mediated by depletion of the cardiac stem cell pool and is rescued by restoration of progenitor cell function. Circulation, 121(2): 276–292.
49. Lipshultz SE. (2006) Exposure to Anthracyclines During Childhood Causes Cardiac Injury. Semin Oncol, 33(SUPPL. 8): 8–14.
50. Ewer MS, Lenihan DJ. (2008) Left ventricular ejection fraction and cardiotoxicity: Is our ear really to the ground? J Clin Oncol, 26(8): 1201–1203.
51. Armenian SH, Hudson MM, Mulder RL, Chen MH, Constine LS, Dwyer M, Nathan PC, Tissing WJE, Shankar S, Sieswerda E, Skinner R, Steinberger J, van Dalen EC, van der Pal H, Wallace WH, Levitt G, Kremer LCM. (2015) Recommendations for cardiomyopathy surveillance for survivors of childhood cancer: A report from the International Late Effects of Childhood Cancer Guideline Harmonization Group. Lancet Oncol, 16(3): e123–136.
52. Calin GA, Ferracin M, Cimmino A, Di Leva G, Shimizu M, Wojcik SE, Iorio M V., Visone R, Sever NI, Fabbri M, Iuliano R, Palumbo T, Pichiorri F, Roldo C, Garzon R, Sevignani C, Rassenti L, Alder H, Volinia S, Liu CG, Kipps TJ, Negrini M, Croce CM. (2005) A MicroRNA Signature Associated with Prognosis and Progression in Chronic Lymphocytic Leukemia. N Engl J Med, 353(17): 1793–1801.
53. Rupaimoole R, Slack FJ. (2017) MicroRNA therapeutics: Towards a new era for the management of cancer and other diseases. Nat Rev Drug Discov, 16(3): 203–221.
54. Umerez M, Garcia-Obregon S, Martin-Guerrero I, Astigarraga I, Gutierrez-Camino A, Garcia-Orad A. (2018) Role of miRNAs in treatment response and toxicity of childhood acute lymphoblastic leukemia. Pharmacogenomics, 19(4): 361–373.
55. Lim EL, Trinh DL, Ries RE, Wang J, Gerbing RB, Ma Y, Topham J, Hughes M,

Pleasance E, Mungall AJ, Moore R, Zhao Y, Aplenc R, Sung L, Kolb EA, Gamis A, Smith M, Gerhard DS, Alonzo TA, Meshinchi S, Marra MA. (2017) MicroRNA expression-based model indicates event-free survival in pediatric acute myeloid leukemia. J Clin Oncol, 35(35): 3964–3977.

56. Garzon R, Calin GA, Croce CM. (2009) MicroRNAs in cancer. Annu Rev Med, 60: 167–179.
57. Schwarzenbach H, Nishida N, Calin GA, Pantel K. (2014) Clinical relevance of circulating cell-free microRNAs in cancer. Nat Rev Clin Oncol, 11(3): 145–156.
58. Leidinger P, Backes C, Meder B, Meese E, Keller A. (2014) The human miRNA repertoire of different blood compounds. BMC Genomics, 15(1): 474.
59. Tkach M, Théry C. (2016) Communication by Extracellular Vesicles: Where We Are and Where We Need to Go. Cell, 164(6): 1226–1232.
60. Johnson SM, Dempsey C, Chadwick A, Harrison S, Liu J, Di Y, McGinn OJ, Fiorillo M, Sotgia F, Lisanti MP, Parihar M, Krishnan S, Saha V. (2016) Metabolic reprogramming of bone marrow stromal cells by leukemic extracellular vesicles in acute lymphoblastic leukemia. Blood, 128(3): 453.
61. Rzepiel A, Kutszegi N, Gézsi A, Sági JC, Egyed B, Péter G, Butz H, Nyírő G, Müller J, Kovács GT, Szalai C, Semsei ÁF, Erdélyi DJ. (2019) Circulating microRNAs as minimal residual disease biomarkers in childhood acute lymphoblastic leukemia. J Transl Med, 17: 372.
62. Brüggemann M, Raff T, Kneba M. (2012) Has MRD monitoring superseded other prognostic factors in adult ALL? Blood, 120(23): 4470–4781.
63. Lipshultz SE, Adams MJ. (2010) Cardiotoxicity after childhood cancer: Beginning with the end in mind. J Clin Oncol, 28(8): 1276–1281.
64. Hershman DL, Neugut AI. (2008) Anthracycline Cardiotoxicity: One Size Does Not Fit All! J Natl Cancer Inst, 100(15): 1046–1047.
65. Mody R, Li S, Dover DC, Sallan S, Leisenring W, Oeffinger KC, Yasui Y, Robison LL, Neglia JP. (2008) Twenty-five-year follow-up among survivors of childhood acute lymphoblastic leukemia: A report from the Childhood Cancer Survivor Study. Blood, 111(12): 5515–5523.
66. Visscher H, Ross CJD, Rassekh SR, Barhdadi A, Dubé MP, Al-Saloos H, Sandor GS, Caron HN, Van Dalen EC, Kremer LC, Van Der Pal HJ, Brown AMK, Rogers

PC, Phillips MS, Rieder MJ, Carleton BC, Hayden MR. (2012) Pharmacogenomic prediction of anthracycline-induced cardiotoxicity in children. J Clin Oncol, 30(13): 1422–1428.

67. The International SNP Map Working Group. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature, 409: 928–933.
68. Egyed B, Kutszegi N, Sági JC, Gézsi A, Rzepiel A, Visnovitz T, Lorincz P, Müller J, Zombori M, Szalai C, Erdélyi DJ, Kovács GT, Semsei ÁF. (2020) MicroRNA-181a as novel liquid biopsy marker of central nervous system involvement in pediatric acute lymphoblastic leukemia. J Transl Med, 18(1): 250.
69. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, Speleman F. (2002) Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. Genome Biol, 3(7): research0034.1.
70. Andersen CL, Jensen JL, Ørntoft TF. (2004) Normalization of real-time quantitative reverse transcription-PCR data: A model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. Cancer Res, 64(15): 5245–5250.
71. Rzepiel A. (2020) Circulating microRNAs as minimal residual disease biomarkers in childhood acute lymphoblastic leukaemia. Semmelweis University Károly Rácz Doctoral School of Clinical Medicine.
72. de Graaf MT, de Jongste AHC, Kraan J, Boonstra JG, Smitt PAES, Gratama JW. (2011) Flow cytometric characterization of cerebrospinal fluid cells. Cytom Part B Clin Cytom, 80B(5): 271–281.
73. Sági JC, Egyed B, Kelemen A, Kutszegi N, Hegyi M, Gézsi A, Herlitschke MA, Rzepiel A, Fodor LE, Ottóffy G, Kovács GT, Erdélyi DJ, Szalai C, Semsei ÁF. (2018) Possible roles of genetic variations in chemotherapy related cardiotoxicity in pediatric acute lymphoblastic leukemia and osteosarcoma. BMC Cancer, 18(1): 704.
74. Barrett JC, Fry B, Maller J, Daly MJ. (2005) Haploview: Analysis and visualization of LD and haplotype maps. Bioinformatics, 21(2): 263–265.
75. Stephens M, Balding DJ. (2009) Bayesian statistical methods for genetic

association studies. Nat Rev Genet, 10(10): 681–690.

76. Antal P, Hullám G, Hajós G, Sárközy P, Gézsi A, Szalai C. Bayesian, Systems-based, Multilevel Analysis of Associations for Complex Phenotypes: from Interpretation to Decision. In: Sinoquet C, Mourad R, editors. Probabilistic Graphical Models for Genetics, Genomics and Postgenomics. Oxford University Press, 2014.

77. Ungvári I, Hullám G, Antal P, Kiszel PS, Gézsi A, Hadadi É, Virág V, Hajós G, Millinghoffer A, Nagy A, Kiss A, Semsei ÁF, Temesi G, Melegh B, Kisfali P, Széll M, Bikov A, Gálffy G, Tamási L, Falus A, Szalai C. (2012) Evaluation of a partial genome screening of two asthma susceptibility regions using bayesian network based bayesian multilevel analysis of relevance. PLoS One, 7(3): 33573.

78. Grasedieck S, Sorrentino A, Langer C, Buske C, Döhner H, Mertens D, Kuchenbauer F. (2013) Circulating microRNAs in hematological diseases: Principles, challenges, and perspectives. Blood, 121(25): 4977–4984.

79. Bray PF, McKenzie SE, Edelstein LC, Nagalla S, Delgrosso K, Ertel A, Kupper J, Jing Y, Londin E, Loher P, Chen HW, Fortina P, Rigoutsos I. (2013) The complex transcriptional landscape of the anucleate human platelet. BMC Genomics, 14(1): 1.

80. Mitchell AJ, Gray WD, Hayek SS, Ko YA, Thomas S, Rooney K, Awad M, Roback JD, Quyyumi A, Searles CD. (2016) Platelets confound the measurement of extracellular miRNA in archived plasma. Sci Rep, 6.

81. Han B-W, Feng D-D, Li Z-G, Luo X-Q, Zhang H, Li X-J, Zhang X-J, Zheng L-L, Zeng C-W, Lin K-Y, Zhang P, Xu L, Chen Y-Q. (2011) A set of miRNAs that involve in the pathways of drug resistance and leukemic stem-cell differentiation is associated with the risk of relapse and glucocorticoid response in childhood ALL. Hum Mol Genet, 20(24): 4903–4915.

82. Jiang Q, Lu X, Huang P, Gao C, Zhao X, Xing T, Li G, Bao S, Zheng H, Raimondo F Di. (2018) Expression of miR-652-3p and Effect on Apoptosis and Drug Sensitivity in Pediatric Acute Lymphoblastic Leukemia. Biomed Res Int, 5724686.

83. Nemes K, Csóka M, Nagy N, Márk Á, Váradi Z, Dankó T, Kovács G, Kopper L, Sebestyén A. (2015) Expression of certain leukemia/lymphoma related

microRNAs and its correlation with prognosis in childhood acute lymphoblastic leukemia. Pathol Oncol Res, 21(3): 597–604.

84. Shalaby T, Fiaschetti G, Baulande S, Gerber NU, Baumgartner M, Grotzer MA. (2015) Detection and quantification of extracellular microRNAs in medulloblastoma. J Cancer Metastasis Treat, 1(2): 67–75.
85. Lin S, Pan L, Guo S, Wu J, Li J, Wang J-C, Wang S. (2013) Prognostic Role of MicroRNA-181a/b in Hematological Malignancies: A Meta-Analysis. PLoS One, 8(3): e59532.
86. Ye Z, Li G, Kim C, Hu B, Jadhav RR, Weyand CM, Goronzy JJ. (2018) Regulation of miR-181a expression in T cell aging. Nat Commun, 9(1): 1–11.
87. Verduci L, Azzalin G, Gioiosa S, Carissimi C, Laudadio I, Fulci V, Macino G. (2015) MicroRNA-181a enhances cell proliferation in acute lymphoblastic leukemia by targeting EGR1. Leuk Res, 39(4): 479–485.
88. Nabhan M, Louka ML, Khairy E, Tash F, Ali-Labib R, El-Habashy S. (2017) MicroRNA-181a and its target Smad 7 as potential biomarkers for tracking child acute lymphoblastic leukemia. Gene, 628: 253–258.
89. Liu X, Liao W, Peng H, Luo X, Luo Z, Jiang H, Xu L. (2016) miR-181a promotes G1/S transition and cell proliferation in pediatric acute myeloid leukemia by targeting ATM. J Cancer Res Clin Oncol, 142(1): 77–87.
90. Marcucci G, Radmacher MD, Maharry K, Mrózek K, Ruppert AS, Paschka P, Vukosavljevic T, Whitman SP, Baldus CD, Langer C, Liu C-G, Carroll AJ, Powell BL, Garzon R, Croce CM, Kolitz JE, Caligiuri MA, Larson RA, Bloomfield CD. (2008) MicroRNA Expression in Cytogenetically Normal Acute Myeloid Leukemia. N Engl J Med, 358(18): 1919–1928.
91. Münch V, Trentin L, Herzig J, Krauss J, Kestler HA, te Kronnie G, Debatin K-M, Meyer LH. (2015) Migration of Acute Lymphoblastic Leukemia Cells into the Central Nervous System Is Regulated By VEGF. Blood, 126(23): 2634–2634.
92. Kato I, Nishinaka Y, Nakamura M, Akarca AU, Niwa A, Ozawa H, Yoshida K, Mori M, Wang D, Morita M, Ueno H, Shiozawa Y, Shiraishi Y, Miyano S, Gupta R, Umeda K, Watanabe K, Koh K, Adachi S, Heike T, Saito MK, Sanada M, Ogawa S, Marafioti T, Watanabe A, Nakahata T, Enver T. (2017) Hypoxic adaptation of leukemic cells infiltrating the CNS affords a therapeutic strategy

targeting VEGFA. Blood, 129(23): 3126–3129.

93. Münch V, Trentin L, Herzig J, Demir S, Seyfried F, Kraus JM, Kestler HA, Köhler R, Barth TFE, Te Kronnie G, Debatin KM, Meyer LH. (2017) Central nervous system involvement in acute lymphoblastic leukemia is mediated by vascular endothelial growth factor. Blood, 130(5): 643–654.

94. Sun X, Wei L, Chen Q, Terek RM. (2015) MicroRNA Regulates Vascular Endothelial Growth Factor Expression in Chondrosarcoma Cells. Clin Orthop Relat Res, 473(3): 907–913.

95. Sun W, Wang X, Li J, You C, Lu P, Feng H, Kong Y, Zhang H, Liu Y, Jiao R, Chen X, Ba Y. (2018) MicroRNA-181a promotes angiogenesis in colorectal cancer by targeting SRCIN1 to promote the SRC/VEGF signaling pathway article. Cell Death Dis, 9(4): 1–13.

96. Yao H, Price TT, Cantelli G, Ngo B, Warner MJ, Olivere L, Ridge SM, Jablonski EM, Therrien J, Tannheimer S, McCall CM, Chenn A, Sipkins DA. (2018) Leukaemia hijacks a neural mechanism to invade the central nervous system. Nature, 560(7716): 55–60.

97. Tokar T, Pastrello C, Rossos AEM, Abovsky M, Hauschild AC, Tsay M, Lu R, Jurisica I. (2018) MirDIP 4.1 - Integrative database of human microRNA target predictions. Nucleic Acids Res, 46(D1): D360–370.

98. Williams MTS, Yousafzai YM, Elder A, Rehe K, Bomken S, Frishman-Levy L, Tavor S, Sinclair P, Dormon K, Masic D, Perry T, Weston VJ, Kearns P, Blair H, Russell LJ, Heidenreich O, Irving JAE, Izraeli S, Vormoor J, Graham GJ, Halsey C. (2016) The ability to cross the blood-cerebrospinal fluid barrier is a generic property of acute lymphoblastic leukemia blasts. Blood, 127(16): 1998–2006.

99. Bartram J, Goulden N, Wright G, Adams S, Brooks T, Edwards D, Inglott S, Yousafzai Y, Hubank M, Halsey C. (2018) High throughput sequencing in acute lymphoblastic leukemia reveals clonal architecture of central nervous system and bone marrow compartments. Haematologica, 103(3): e110–114.

100. Indrieri A, Carrella S, Carotenuto P, Banfi S, Franco B. (2020) The pervasive role of the miR-181 family in development, neurodegeneration, and cancer. Int J Mol Sci, 21(6): 2092.

101. Martínez-Laperche C, Gómez-García AM, Lassaletta Á, Moscardó C, Vivanco JL,

Molina J, Fuster JL, Couselo JM, De Toledo JS, Bureo E, Madero L, Ramírez M. (2013) Detection of occult cerebrospinal fluid involvement during maintenance therapy identifies a group of children with acute lymphoblastic leukemia at high risk for relapse. Am J Hematol, 88(5): 359–364.

102. van der Velden VHJ, de Launaij D, de Vries JF, de Haas V, Sonneveld E, Voerman JSA, de Bie M, Revesz T, Avigad S, Yeoh AEJ, Swagemakers SMA, Eckert C, Pieters R, van Dongen JJM. (2016) New cellular markers at diagnosis are associated with isolated central nervous system relapse in paediatric B-cell precursor acute lymphoblastic leukaemia. Br J Haematol, 172(5): 769–781.
103. Zhang H, Luo XQ, Zhang P, Huang L Bin, Zheng YS, Wu J, Zhou H, Qu LH, Xu L, Chen YQ. (2009) MicroRNA patterns associated with clinical prognostic parameters and CNS relapse prediction in pediatric acute leukemia. PLoS One, 4(11): e7826.
104. Andreu Z, Yáñez-Mó M. (2014) Tetraspanins in extracellular vesicle formation and function. Front Immunol, 5: 442.
105. Kowal J, Arras G, Colombo M, Jouve M, Morath JP, Primdal-Bengtson B, Dingli F, Loew D, Tkach M, Théry C. (2016) Proteomic comparison defines novel markers to characterize heterogeneous populations of extracellular vesicle subtypes. Proc Natl Acad Sci U S A, 113(8): E968–977.
106. Wang H, Eckel RH. (2014) What are lipoproteins doing in the brain? Trends Endocrinol Metab, 25(1): 8–14.
107. Lipshultz SE, Lipsitz SR, Mone SM, Goorin AM, Sallan SE, Sanders SP, Orav EJ, Gelber RD, Colan SD. (1995) Female Sex and Higher Drug Dose as Risk Factors for Late Cardiotoxic Effects of Doxorubicin Therapy for Childhood Cancer. N Engl J Med, 332(26): 1738–1744.
108. Hequet O, Le QH, Moullet I, Pauli E, Salles G, Espinouse D, Dumontet C, Thieblemont C, Arnaud P, Antal D, Bouafia F, Coiffier B. (2004) Subclinical late cardiomyopathy after doxorubicin therapy for lymphoma in adults. J Clin Oncol, 22(10): 1864–1871.
109. Mulrooney DA, Armstrong GT, Huang S, Ness KK, Ehrhardt MJ, Joshi VM, Plana JC, Soliman EZ, Green DM, Srivastava D, Santucci A, Krasin MJ, Robison LL, Hudson MM. (2016) Cardiac outcomes in adult survivors of childhood cancer

exposed to cardiotoxic therapy. Ann Intern Med, 164(2): 93–101.

110. Sági JC, Kutszegi N, Kelemen A, Fodor LE, Gézsi A, Kovács GT, Erdélyi DJ, Szalai C, Semsei ÁF. (2016) Pharmacogenetics of anthracyclines. Pharmacogenomics, 17(9): 1075–1087.
111. Klein K, Thomas M, Winter S, Nussler AK, Niemi M, Schwab M, Zanger UM. (2012) PPARA: A novel genetic determinant of CYP3A4 in vitro and in vivo. Clin Pharmacol Ther, 91(6): 1044–1052.
112. Ross C, Visscher H, Rassekh S, Castro-Pastrana L, Shereck E, Carleton B, Hayden M. (2011) Pharmacogenomis of serious adverse drug reactions in pediatric oncology. J Popul Ther Clin Pharmacol, 18(1): e134-151.
113. Huang Z, Wang J, Qian J, Li Y, Xu Z, Chen M, Tong H. (2017) Effects of cytochrome P450 family 3 subfamily a member 5 gene polymorphisms on daunorubicin metabolism and adverse reactions in patients with acute leukemia. Mol Med Rep, 15(6): 3493–3498.
114. Rossi D, Rasi S, Franceschetti S, Capello D, Castelli A, De Paoli L, Ramponi A, Chiappella A, Pogliani EM, Vitolo U, Kwee I, Bertoni F, Conconi A, Gaidano G. (2009) Analysis of the host pharmacogenetic background for prediction of outcome and toxicity in diffuse large B-cell lymphoma treated with R-CHOP21. Leukemia, 23(6): 1118–1126.
115. Visscher H, Ross CJD, Rassekh SR, Sandor GSS, Caron HN, van Dalen EC, Kremer LC, van der Pal HJ, Rogers PC, Rieder MJ, Carleton BC, Hayden MR. (2013) Validation of variants in SLC28A3 and UGT1A6 as genetic markers predictive of anthracycline-induced cardiotoxicity in children. Pediatr Blood Cancer, 60(8): 1375–1381.
116. Aminkeng F, Ross CJD, Rassekh SR, Hwang S, Rieder MJ, Bhavsar AP, Smith A, Sanatani S, Gelmon KA, Bernstein D, Hayden MR, Amstutz U, Carleton BC, CPNDS Clinical Practice Recommendations Group. (2016) Recommendations for genetic testing to reduce the incidence of anthracycline-induced cardiotoxicity. Br J Clin Pharmacol, 82(3): 683–695.
117. Hertz DL, Caram M V., Kidwell KM, Thibert JN, Gersch C, Seewald NJ, Smerage J, Rubenfire M, Henry NL, Cooney KA, Leja M, Griggs JJ, Rae JM. (2016) Evidence for association of SNPs in ABCB1 and CBR3, but not RAC2, NCF4,

SLC28A3 or TOP2B, with chronic cardiotoxicity in a cohort of breast cancer patients treated with anthracyclines. Pharmacogenomics, 17(3): 231–240.

118. Reichwagen A, Ziepert M, Kreuz M, Gödtel-Armbrust U, Rixecker T, Poeschel V, Reza Toliat M, Nürnberg P, Tzvetkov M, Deng S, Trümper L, Hasenfuss G, Pfreundschuh M, Wojnowski L. (2015) Association of NADPH oxidase polymorphisms with anthracycline-induced cardiotoxicity in the RICOVER-60 trial of patients with aggressive CD20+ B-cell lymphoma. Pharmacogenomics, 16(4): 361–372.
119. Lipshultz SE, Alvarez JA, Scully RE. (2008) Anthracycline associated cardiotoxicity in survivors of childhood cancer. Heart, 94(4): 525–533.
120. Kremer LCM, van der Pal HJH, Offringa M, van Dalen EC, Voute PA. (2002) Frequency and risk factors of subclinical cardiotoxicity after anthracycline therapy in children: a systematic review. Ann Oncol, 13(6): 819–829.

VIII. Bibliography of the candidate's publications

VIII.1. Own papers taking the basis of this dissertation

Egyed B, Kutszegi N, Sági JC, Gézsi A, Rzepiel A, Visnovitz T, Lorincz P, Müller J, Zombori M, Szalai C, Erdélyi DJ, Kovács GT*, Semsei ÁF*. (2020) MicroRNA-181a as novel liquid biopsy marker of central nervous system involvement in pediatric acute lymphoblastic leukemia. *J Transl Med*, 18(1): 250.

* Co-last authors.

D1 publication in *Medicine (miscellaneous)* category, impact factor: 5.531

Sági JC*, **Egyed B***, Kelemen A, Kutszegi N, Hegyi M, Gézsi A, Herlitschke MA, Rzepiel A, Fodor LE, Ottóffy G, Kovács GT, Erdélyi DJ, Szalai C, Semsei ÁF. (2018) Possible roles of genetic variations in chemotherapy related cardiotoxicity in pediatric acute lymphoblastic leukemia and osteosarcoma. *BMC Cancer*, 18(1): 704.

* Co-first authors.

Q1 publication in *Oncology* category, impact factor: 2.933

Rzepiel A, Kutszegi N, Gézsi A, Sági JC, **Egyed B**, Péter G, Butz H, Nyírő G, Müller J, Kovács GT, Szalai C, Semsei ÁF, Erdélyi DJ. (2019) Circulating microRNAs as minimal residual disease biomarkers in childhood acute lymphoblastic leukemia. *J Transl Med*, 17: 372.

D1 publication in *Medicine (miscellaneous)* category, impact factor: 4.124

VIII.2. Other own papers in the dissertation's topic

Egyed B, Horváth A, Semsei ÁF, Szalai C, Müller J, Erdélyi DJ, Kovács GT. (2022) Co-detection of VEGF-A and its regulator, microRNA-181a, may indicate central nervous system involvement in pediatric leukemia. *Pathol Oncol Res*, 28:1610096.

Impact factor (2021): 2.874

Sági JC, Gézsi A, **Egyed B**, Jakab Z, Benedek N, Attarbaschi A, Köhrer S, Sipek J, Winkowska L, Zaliova M, Anastasopoulou S, Wolthers BO, Ranta S, Szalai C, Kovács GT, Semsei ÁF*, Erdélyi DJ*. (2021) Pharmacogenetics of the central nervous system - toxicity and relapse affecting the CNS in pediatric acute lymphoblastic leukemia. *Cancers*, 13:2333.

* Co-last authors.
Impact factor: 6.575

Egyed B, Kovács GT, Kutszegi N, Rzepiel A, Sági JC, Erdélyi DJ, Müller J, Semsei ÁF. (2018) New and traditional directions in the biology and management of childhood acute lymphoblastic leukemia. *Orv Hetil*, 159(20), 786-797.
Impact factor: 0.564

VIII.3. Other own papers out of the dissertation's topic

Alpár D*, **Egyed B***, Bödör C, Kovács GT. (2021) Single-cell sequencing: biological insight and potential clinical implications in pediatric leukemia. *Cancers*, 13:5658.
* Co-first authors.
Impact factor: 6.575

Constatin T, Andrási N, Ponyi A, Goschler Á, Ablonczy L, Kincs J, Csóka M, **Egyed B**, Horváth Z, Kalocsai K, Káposzta R, Kardics K, Kemény V, Mosdósi B, Pék T, Szabó Z, Tóth A, Tory K, Tölgyesi A, Ónozó B, Vágó H, Vilmányi C, Weiser P, Szekanecz Z, Kovács GT, Szabó A. (2021) Diagnosis and treatment of paediatric inflammatory multisystem syndrome. *Orv Hetil*, 162(17):652-667.
Impact factor: 0.707

Kiss R, Gángó A, Benard-Slagter A, **Egyed B**, Haltrich I, Hegyi L, de Groot K, Király PA, Krizsán S, Kajtár B, Pikó H, Pajor L, Vojcek Á, Matolcsy A, Kovács GT, Szuhai K, Savola S, Bödör C*, Alpár D*. (2020) Comprehensive profiling of disease-relevant copy number aberrations for advanced clinical diagnostics of pediatric acute lymphoblastic leukemia. *Mod Pathol*, 33:812-824.
* Co-last authors.
Impact factor: 7.842

Total cumulative impact factor of the candidate: 37.725

IX. Acknowledgements

I owe a great debt of gratitude to my wife, Krisztina, my son, Barnabás, my mother and my whole beloved family for their support and help. They tolerated the difficulties of my work and stood by me in hard times. I think of the area of Zala county with endless love, where I had the opportunity to grow up in an incentive environment, between people who valued intellectual creative work. I am grateful to my first biology and chemistry teachers, Mariann Járiné Lévai, Katalin Tölgyesné Kovács and László Halmi, who formed my interest by their view of nature and life.

It was quite a matter of chance that I joined the Hematology Working Group in the 2nd Department of Pediatrics at Semmelweis University in 2015, as a medical student. The precise and structured operation of the Hematology Department to care and cure pediatric leukemia patients with the best available tools always inspired me. I will be eternally grateful to Prof. Gábor Kovács, my supervisor and department head, who always helped my scientific work and my medical studies during the last six years and guaranteed the conditions for the progress of my work without any request. I have learnt a lot from him at both professional and human levels. I appreciate his thoughtfulness and the opportunity to meet the Hungarian Pediatric Oncology Network very early in my career. I say thank to Ágnes Félné Semsei, my co-supervisor, for being in my corner during the accomplishment of this thesis. She helped me designing, performing and interpreting my experiments, while I could learn how scientific work should be conducted. I will always be grateful for the endless hours she has spent with answering my questions.

Most of my experiments were based on a continuously growing biobank founded by Dániel Erdélyi. I say thank for his trust, knowledge and helpfulness, as he was also available to answer my questions. I am grateful to Judit Müller for her knowledge and openness she always showed me. I say thank to my colleagues, Judit Csányiné Sági and Nóra Kutszegi, because I could discuss my questions with them as researchers of the same scientific area. I am grateful to András Gézsi, who was also available to answer my biostatistics questions. I am grateful to Prof. Csaba Szalai, who always supported my work. I say thank to Mónika Sándorné Vángor and Andrea Rzepiel for their precise work in the laboratory and their help in picking laboratory skills up. I am grateful to the nurses of the Hematology Department for their generous work and help in sample collection.