PHARMACOGENETICS AND CELL BIOLOGY OF PEDIATRIC ACUTE LYMPHOBLASTIC LEUKEMIA

PhD thesis

Judit Csányiné Sági, MD

Doctoral School of Molecular Medicine

Semmelweis University

Supervisor: Ágnes Félné Semsei, Ph.D

Official reviewers: Zsolt Rónai, Ph.D

Zsolt Szelid, MD, Ph.D

Head of the Final Exam Committee: Péter András Lakatos, MD, D.Sc

Members of the Final Exam Committee: András Bors, Ph.D

Anna Sebestyén, Ph.D

Csaba Bödör, Ph.D

Budapest, 2021

1. Introduction

In Europe, an average of 12-15 new patients is diagnosed with malignancies per 100,000 children a year. Among childhood cancers, the most common is acute lymphoblastic leukemia (ALL). Incidence is 1/2000 kids in the 1-15 years age group in developed countries, diagnosis peak is at 2-5 years. Males are more prone than females. Leukemia is the malignant transformation and proliferation of a lymphoid progenitor with differentiation blockade. In Hungary, 50-70 pediatric patients are diagnosed with ALL every year. The two main subgroups of ALL are T-cell and B-cell types. The cure rate is higher than 80-90% in the overall pediatric ALL population. However, complications of the therapy are responsible for the remaining 10-20% unfavorable outcome. Challenges related to ALL therapy are the focus of this thesis. Side effects of the treatment or the return of primary leukemia caused by ineffective chemotherapy reduce survival. Recurrent leukemia cases are called relapses. Relapse cases presented extramedullary in the central nervous system (CNS) or after primary T-cell ALL (T-ALL) are outstanding limitations on survival. Adverse effects of the used drugs for example cardiotoxicity or neurotoxicity are responsible for dose-limitation during treatment or for long-lasting dysfunction of the affected organs. The response to certain drugs or the success of the treatment shows inter-individual differences which can be influenced by inherited germline variants or de novo mutations of the transformed cells. Reliable biomarkers for predicting these complications and for targeted therapy are still needed. After proper validation, biomarkers based on the genetic background of patients could support endeavors of personalized precision medicine also in the care of ALL.

In the present thesis, our aim was to investigate the association of inherited single nucleotide variants with cardio-, and neurotoxicity or first CNS relapse and de novo mutations in the clonal evolution of the first relapse in T-ALL.

In my research works, I aimed to evaluate the role of genetic variants in childhood primary acute lymphoblastic leukemia, its relapses, and osteosarcoma. I studied treatment-related side effects: cardio-, and neurotoxicity and the recurrence of leukemia in the CNS in the presence of heritable SNPs and the relapse of T-ALL in relation with de novo SNVs and InDels, respectively. To achieve my aim, I have searched for possible target genes in the literature and topic-related databases, studied patient's clinical records

retrospectively, built the databank of our biobank, genotyped samples, prepared datasheets for statistical evaluation, contributed to statistical analysis of the data, and wrote research articles. Pharmacogenetic markers have the potential to reform the therapy of pediatric oncology. Predicting possible toxicities or underexposure related to the used chemotherapeutic drugs before their clinical manifestation with biomarkers, could increase the personalized approach and prevent treatment failures. Identification of novel variations for a possible targeted therapy could increase the safety and efficacy of the treatment of ALL and OSC.

2 Objectives

The aims of my research were the followings:

1. Evaluation of single nucleotide polymorphisms as possible pharmacogenetic markers in anthracycline-related cardiotoxicity. To test the association between genotypes of germline SNPs and left ventricular parameters indicating cardiotoxicity induced by anthracyclines appeared during or after the treatment of pediatric primary ALL (T-, and B- cell subtypes) and OSC.

2. Evaluation of single nucleotide polymorphisms as possible pharmacogenetic markers in neurotoxicity. To analyze if germline SNPs of metabolizing enzymes and transporters of the blood-brain barrier have a significant contribution to the appearance of neurotoxic events caused by chemotherapy in pediatric patients with primary ALL (T-, and B-cell subtypes).

3. Evaluation of single nucleotide polymorphisms as possible pharmacogenetic markers in association with the first relapse presented in the central nervous system. To investigate the association between germline SNPs of metabolizing enzymes and transporters of the blood-brain barrier and the presence of relapse in the central nervous system in pediatric patients with ALL (T-, and B-cell subtypes).

4. Searching for somatic mutations as possible prognostic markers in the relapse of T-cell acute lymphoblastic leukemia. To evaluate their role in the survival of relapse-specific SNVs, small insertions, and deletions in T- cell ALL tumor-specific samples collected at initial and relapse time points.

3 Methods

3.1. Patients

Hungarian pediatric patients diagnosed with acute lympoblastic leukemia or osteosarcoma were enrolled in the cardiotoxicity, neurotoxicity and CNS relapse studies (n=661, n=626 and n=533, respectively). All of them were treated in one of the 6 pediatric oncology centers of Hungary. Their DNA was stored in the Hungarian Pediatric Oncohematology Biobank at the Semmelweis University, Department of Genetics, Cell- and Immunobiology due to the successful collaboration with the Hungarian Pediatric Oncology Network. Retrospective clinical data collection was conducted from medical records of enrolled patients or from the National Pediatric Cancer Registry of Hungary. Chemotherapeutic drugs and doses used in treatment protocols were summarized in previous articles of our research group. Further patients data were collected and analyzed in international collaborations. The researches were conducted according to the principles expressed in the Declaration of Helsinki. Written informed consent was requested from all patients or the parents or guardians of the minors involved in the studies. The ethical committees of the participating countries approved the study.

Cardiotoxicity study group

Patients aged 0-18 years and diagnosed between 1989 and 2015 with ALL or OSC in Hungary were enrolled into the cardiotoxicity study group. Exclusion criteria were preceding cardiac problems, congenital cardiac abnormalities, or inherited genetic diseases which could influence the cardiac and genetic markers of the individuals (n=19). Dosages of anthracyclines were administered in the intravenous part of the chemotherapy. Low and intermediate-risk patients received a cumulative dose between 180-240 mg/m^2 while patients with high risk or relapse got a cumulative dose between 240-380 mg /m^2. Cumulative doxorubicin doses of OSC patients were 360 mg/m^2 for standard-risk or 180

mg/m^2 for high-risk. Twenty-nine percent of the cohort was treated with 12 Gy cranial radiotherapy (regarding the corresponding BFM protocol). Patients were checked for left ventricular function after anthracycline therapy regularly. Echocardiography (ECHO) was used to control these parameters. Left ventricular end-diastolic-diameter (LVEDD) and left ventricular end-systolic diameter (LVESD) data were collected retrospectively for the study. Left ventricular ejection fraction (EF) and left ventricular fractional shortening (FS) were calculated as EF = (LVEDD3-LVESD3)/LVEDD3 and FS = (LVEDD-LVESD)/ LVEDD and included in the analysis.

Clinical checkpoints for echocardiography were before the start of the therapy, during and after the intravenous treatment on several occasions, repeatedly. In this study, patients were grouped by follow-up categories and their FS and EF values were analyzed within the group, respectively. These follow-up categories were: 1) at the diagnosis (used as a control) (n=387); 2) in acute phase: during the intensive chemotherapy phase (n=280); 3) during oral maintenance chemotherapy (n=49); 4) at the end of the treatment (n=315), which is after the oral maintenance chemotherapy period completed 2 or 3 years after the diagnosis; 5) from the end of the treatment until 5 years after the diagnosis (n=264); 6) 5– 10 years after the diagnosis (n=301); 7) 10–15 years after the diagnosis (n=152); 8) more than 15 years after the diagnosis (n=32). During the long follow-up, not all ECHO data were accessible. Always the latest ECHO data were included in the analysis of each follow-up category. For the case-control study, the worst recorded ECHO data were used by every patient. Echocardiograms with FS ≤ 28% were determined as pathological FS cases (n=20), controls had FS > 28% throughout the follow-up time (n=641). The change in the FS value was studied as a dichotomous variable: the difference of FS at diagnosis from the end of the treatment was registered. The change in FS value from the diagnosis to the last follow-up was also analyzed. Decreasing FS cases (n=105, n=170) were studied in comparison with increasing FS cases (n=94, n=152), respectively. In this analysis, power was ≥75% for all of the results.

Neurotoxicity study group

Our research group (Medical Genomics Research Group, Semmelweis University, Department of Genetics, Cell- and Immunobiology) studied chemotherapy-related ATE among patients diagnosed with ALL in the period of 1995-2005. In my research, I extended this cohort with additional ATE cases among patients diagnosed with ALL

between 2005 and 2015. The newly collected population and the original one were analyzed together to test associations between SNPs of blood-brain barrier enzymes and transporters (n=580) (*Discovery cohort*). Our previously published results with *ABCB1* rs1045642 and with its combination with *ABCG2* rs2231142 and ATE and further candidate SNPs were investigated on the extended population. To validate the association of five chosen SNPs with ATE, a European international collaboration was organized. Matched case-control cohort was set up with Austrian, Czech, and Northern (Nordic Society of Pediatric Hematology and Oncology (NOPHO) Group) patients collected from national study groups. The same enrolment criteria were used for all of the study groups when selecting patients for the *Joined validation cohort*. The *Joined validation cohort* included 107 ATE cases and 211 controls. ATE and its subphenotypes were also studied on the total *Combined cohort* of ATE which contained a matched Hungarian population and the *Joined validation cohort*. Studying complications of the central nervous system in pediatric ALL, cohorts included children with 0-18 years of age at the time of diagnosis (1-18 years for neurotoxicity study; 0-18 years for relapse study) from Hungary, Austria, Czech Republic and countries of the NOPHO Group (Denmark, Norway, Sweden, Finland, Iceland, Lithuania, Estonia). Clinical data collection source were data collection sheets of the PdL 'Retrospective Investigation of Children with ALL/LBL with Central Neurotoxicity Related to Therapy' study (with complements to Christina Halsey and the Ponte di Legno Toxicity Working Group). Complications of the central nervous system in ALL, such as acute encephalopathy (AE) and CNS relapse were in the focus of this study. AE included adverse CNS symptoms at least grade 3 regarding the Common Terminology Criteria for Adverse Events (CTCAE) v.4.0. Acute symptoms, appearing from the start of the treatment until 3 weeks after the i.v. chemotherapy, were analyzed. Exclusion criteria for neurotoxicity study were any previous chemotherapy, any major deviations from ALL therapy, previous CNS diseases, uncertain or mild neurologic symptoms.

AE cases were CNS-linked cerebrovascular events, like infections or leukemia not in remission, systemic metabolic disturbances (e.g. hepatic encephalopathy, hypoglycemia, or diabetic ketoacidosis), or the result of insufficient CNS circulation (e.g. hypertensive encephalopathy, increased intracranial pressure, hypotension, or hypoxia).

Subphenotype groups of AE were also investigated. Events with no known secondary etiology were called acute toxic encephalopathy (ATE) and were considered as direct adverse effects of chemotherapeutic drugs affecting the blood-brain barrier. CNS symptoms with known causes were excluded from this cohort. AE and ATE cases were categorized into overlapping Delphi consensus phenotypes as a stroke-like syndrome (SLS), seizures with no other symptoms, depressed level of consciousness, posterior reversible encephalopathy syndrome (PRES) defined by the PdL Toxicity Working Group in 2016. These phenotypes could appear with or without known origin which was the basis of their classification in the analysis. Two controls were matched to one case in the total cohort. Controls were pediatric patients with ALL who experienced none of these events, had no comorbidities, medical history, or co-medication that may have influenced the occurrence of CNS complications or drug pharmacokinetics.

AE events occurred during the whole intravenous therapy which was divided into 4 treatment phases in the analyses: 'Induction-like' (including steroid, L-asparaginase, vincristine, +/- anthracycline); 'High-dose methotrexate' (including methotrexate 2-5 g/m2 iv with mercaptopurine); 'BFM-consolidation' (including cyclophosphamide, cytarabine, 6-mercaptopurine or thioguanine) and 'Other' (e.g. BFM high risk (HR) 1, HR2, HR3 cycles).

CNS relapse study group

CNS relapse was studied on a non-matched population of Hungarian patients as a discovery population. After this, we established a further international cohort for studying CNS relapse on a second *Combined cohort* with patients of the four study groups. Into the CNS relapse study patients were enrolled with 1st relapse affecting only the CNS as isolated cases, and combined CNS relapse cases with additional medullary or extramedullary ALL. In the analysis two non-relapsed ALL cases and one isolated BM first relapse case were controls matched to one CNS relapse case. Into the *Combined cohort* of CNS relapse 86 CNS relapse cases (isolated or combined), 105 isolated bone-marrow (BM) relapse cases, and 129 relapse-free controls were enrolled and three controls per case were matched. The number of patients in the validation cohort was designed based on the results found on the discovery population with the statistical power of 0.8.

T-ALL study group

Samples at the time of initial diagnosis were collected from 147 patients with T-cell acute lymphoblastic leukemia. These samples were called initial samples. Further non-matched 66 samples of patients with T-ALL were collected at the diagnosis of relapse. Patients with primary leukemia or relapse were treated according to ALL-BFM 2000, AIEOP-BFM ALL 2009, or ALL-REZ BFM 2002 protocols. The study was approved by the institutional review boards of the Charité Universitätsmedizin Berlin and the Medical Faculty Heidelberg. Informed consent was obtained in accordance with the Declaration of Helsinki.

3.2. Laboratory approaches

DNA isolation

DNA was isolated from mononuclear cells extracted from BM or peripheral blood samples collected at initial diagnosis or remission. DNA isolation was performed using Qiagen isolation kits according to the manufacturer's instructions (QIAmp DNA Blood Midi or Maxi Kit, Qiagen, Hilden, Germany) or the Gentra Puregene Cell Kit (Qiagen, Hilden, Germany). For one isolated extramedullary relapse case, DNA was extracted from a lymph node.

SNP selection

Seventy SNPs in 26 genes were selected for genotyping. Based on previous publications, we collected variants in transporters or metabolizing enzymes related to chemotherapeutic drugs and further variants of potential new candidate genes. The functionality of SNPs was prioritized during selection: non-synonymous SNPs, SNPs in the promoter and the 3'-UTR (3′-untranslated region) region, synonymous SNPs, and intronic SNPs. Minor allele frequency data were established with the HapMap database No. 27 and the CEU population (CEPH: Utah residents with ancestry from northern and western Europe).

Genotyping of the Hungarian samples

DNA samples of Hungarian patients were genotyped using TaqMan® OpenArray™ Genotyping System (Thermo Fisher Scientific, Waltham, MA, USA) by Zsolt Rónai and his workgroup at the Semmelweis University, Department of Medical Chemistry, Molecular Biology and Pathobiochemistry. The results were visualized using the TaqMan Genotyper SoftwareTM (Applied Biosystems). A portion of the Hungarian samples was genotyped using KASPar (KBioscience Competitive Allele-Specific Polymerase chain reaction)-on-Demand prevalidated assays (LGC Genomics, Berlin, Germany) on 7900HT Fast Real-Time PCR System (Thermo Fisher Scientific Waltham, MA, USA) following the manufacturer's instructions by our group at the Semmelweis University, Department of Genetics, Cell- and Immunobiology. Genotypes of the samples collected internationally were provided by collaborators.

Selection of genes for target sequencing in samples of patients with T-ALL

We wanted to evaluate repetitive T-ALL relapse-specific somatic alterations in association with prognosis and clinical features of the studied patient group. During target design, we collected candidate genes from the literature. We searched genes in association with T-cell leukemia, which have been found in initial and relapsed disease or only in relapse. We excluded the too large genes and the results of cell line studies. We included genes from our previous results, as well. With these screening, we could collect 324 possible candidates for further analysis. I have searched for genes in the literature, looked hotspots of the selected genes in COSMIC, ranked them with Endeavour, and finally looked for the function and role of these genes in GeneCards.

Mutation identification in T-ALL samples

After the gene selection step, we designed a Haloplex Panel to target regions of interest. Haloplex Target Enrichment Kit (Agilent, Santa Clara, CA, USA) was used for library preparation and covered the selected genes, 3.04 Mbp target sequence. After DNA quantification (Qubit dsDNA BR Assay kit (Life Technologies, Darmstadt, Germany), 112.5 ng genomic DNA was used. Libraries were pooled in batches and were sequenced as 100 bp paired reads on one lane using an Illumina HiSeq 2000 instrument (Illumina, San Diego, CA, USA). VarScan26 was used to detect both SNVs and small insertions and

deletions. To control the NGS (next-generation sequencing) results, we also used conventional Sanger-sequencing of the *NOTCH1* and *PTEN* genes in 144 patients.

3.3. Statistical analysis

In the Hungarian population, significant violation from Hardy-Weinberg equilibrium was tested with online software https://ihg.gsf.de/cgi-bin/hw/hwa1.pl. A multi-adjusted general linear model was used to study associations of SNPs and left ventricular parameters as continuous variables (such as FS or EF). Alterations of FS were studied as categorical variables using univariate or multivariate logistic regression. Potential confounders were age at the time of diagnosis (years), gender (male-female), chemotherapy protocols (before 2000, after 2000 and OSC protocols; also reflects radiotherapy), risk groups (standard, intermediate, high-risk), and cumulative dose of anthracycline ($\leq$ or $>$ 240 mg/m^2). Conditional logistic regression models were applied to study the association of genotype frequencies and the presence of neurotoxic events (case-control study design). Cox proportional hazard regression models for nested case-control data were used to analyze overall survival (OS) and event-free survival (EFS) in every CNS event cohort. Confidence intervals (CI) of odds ratio (OR) or hazard ratio (HR) were calculated at the 95% level. Confounders of the CNS study were gender, age at diagnosis, ALL phenotype, risk group, intravenous MTX doses, study group, cycles of treatment. We studied the genotypes separately (11 vs. 12; 11 vs. 22), using dominant (11 vs. 12/22) or recessive (11/12 vs. 22) models. Common homozygotes were described with 11 and the rare homozygotes with 22, which was supposed to be the dominant allele. For the correction of multiple comparisons the Benjamini-Hochberg false discovery rate (FDR) method with a type I error rate of 10 % or 13% was used, respectively. Alpha levels of $p \leq 8.90E\text{-}03$ or $p \leq 1.13E\text{-}02$ were considered significant after FDR correction. EF and FS are given always with the standard error (SE) of the estimate of the mean. Analyses and figures were prepared in IBM SPSS Statistics 23.0 or 25.0 (IBM Corporation, Armonk, NY, USA) and RStudio Version 1.0.136 or version Version 3.6.3 (RStudio, Boston, MA, USA) programs. The clogit function of the survival package of R was used for the conditional logistic regression analyses. MultipleNCC package was applied for the Cox proportional hazards regression analyses for nested case-control data. Power analysis was performed with PS: Power and Sample Size Calculation 3.1.2 or was estimated at a

significance level of 0.05 using SPSS Statistics 23.0 or 25.0 programs. In the T-ALL population, statistical analyses were performed using GraphPad Prism 6.0, or R (Package: Survival).

4. Results

4.1. Genotype data for studying cardiotoxicity and CNS complications

In the Hungarian population, 70 SNPs were genotyped. Genotyping was not successful for 3 SNPs and genotype distribution was not in Hardy-Weinberg equilibrium in case of one SNP (*AKR1A1* (Aldo-Keto Reductase Family 1 Member A1) rs2934859. These 4 SNPs were excluded from further analyses.

4.1.1 Single nucleotide polymorphisms associated with cardiotoxicity

Sixty-six SNPs were studied in association with left ventricular function parameters (ejection fraction (EF), fractional shortening (FS)) of pediatric patients with ALL or OSC. Case-control and follow-up types of studies were used to explore the genetic background of EF and FS. In my dissertation, I focus on the follow-up analysis of the total, of the ALL subpopulation and the alteration of FS. *ABCC2* rs3740066 was associated with worse FS in the acute phase of the therapy in the total cohort and at 5–10 years after the diagnosis, and also in this 5-10 years phase in the ALL subgroup (p=0.00738, p=0.000711, p=0.0045 respectively). *NQO1* (NAD(P)H Quinone Dehydrogenase 1) rs1043470 associated significantly with worse FS on the total population in these phases and in the acute phase in the ALL subpopulation (p=0.00428, p=0.00582, p=0.0026, respectively). *SLC22A6* (Solute Carrier Family 22 Member 6) gene rs6591722 was associated with lower mean FS at 5–10 years after the diagnosis on the total population and patients with ALL (p=0.00171, p=0.0059, respectively). Ejection fraction results showed the same tendency as FS. Alterations of FS were studied in association with SNPs observing two intervals: between the time of diagnosis and the end of the treatment or the last follow-up. In the ALL subpopulation, SNPs were found in significant association with cardiac parameters, however, these associations were not significant on the total cohort. *CYP3A4* (Cytochrome P450 Family 3 Subfamily A Member 4) rs3735451, *CYP3A5* rs776746 were in association with a shift in FS values analyzing the period

between the time of diagnosis and the end of the treatment (p=0.0057, p=0.0038, respectively). Studying the time of diagnosis and the last registered FS value, *NQO1* rs1043470 was in association with FS alteration (p=0.0089). The change in FS was below or over 3% in 67–71% in all groups.

4.1.2 Single nucleotide polymorphisms associated with central nervous system toxicity

We evaluated the association between 60 SNPs in 20 genes in the transport and metabolism of chemotherapeutic drugs and CNS toxicity. The presence of AE and ATE in Hungarian pediatric patients with ALL (discovery population) were studied in the case-control analysis. We validated the results (5 SNPs) related to CNS toxicity on an international validation cohort. CNS relapse and the same SNP-set were investigated on the cohort of all nations (Austria, Czech Republic, Hungary, and NOPHO Group).

In the *Hungarian discovery cohort*, AE and ATE significantly associated with *ABCB1* rs1045642 (p=0.011, p=0.047). The association both with AE and ATE remained significant when analyzing *ABCB1* rs1045642 in combination with *ABCG2* rs2231142 (p=0.010, p=0.003). ATE associated with further *ABCB1* SNPs: rs1128503 and rs2032582 (p=0.043, p=0.026, respectively). *GSTP1* rs1695 was also in association with both AE and ATE (p=0.0005, p=0.004).

We investigated the above results with the presence of ATE in an international validation study. In the *Joined cohort* (Austria, Czech Republic, NOPHO Group) we found only *GSTP1* rs1695 in an opposite association with ATE compared to results of the *Hungarian cohort* (p=0.029).

Studying the *Combined cohort* of every patient, ATE was not in association with any of the studied SNPs. However, the extended number of patients allowed the analysis of subpopulations of ATE in this cohort. Seizure was associated with *ABCB1* rs1045642, rs1128503 and rs2032582 polymorphisms, respectively (p=0.011, p=0.034, p=0.019) in a subpopulation of 44 patients with seizure and 89 matched controls. Analyzing only seizure cases in the therapy phase, in Induction-like chemotherapy cycles (n=28/57) associations remained unambiguous for the same SNPs (p=0.010, p=0.027, p=0.007). We investigated also the survival and effect of the 5 SNPs on the Hungarian and *Combined*

AE and ATE cohorts. Worse OS was associated with AE in Hungarian cohort (n=76/626) (p=0.005, HR=2.51, CI 95% (1.32-4.76)). Two neurotoxicity-induced deaths were registered in our database of 82 AE cases (9.5% of all exits).

Survival of AE or ATE Hungarian populations were influenced by *CYP3A5* rs4646450 and *CYP3A4* rs3735451. *CYP3A5* rs4646450 associated with worse OS and EFS in AE and ATE cohorts (p=0.0001, HR=2.80, CI 95% (1.70-4.60); p=0.0003, HR=2.27, CI 95% (1.46-3.53); p=0.001, HR=2.43, CI 95% (1.43-4.13); p=0.004, HR=2.00, CI 95% (1.25-3.18), respectively). *CYP3A4* rs3735451 associated also with worse OS and EFS in AE and ATE cohorts (p=0.007, HR=5.15, CI 95% (1.56-17.00), p=0.001, HR=5.54, CI 95% (1.97-15.56), p=0.004, HR=5.91, CI 95% (1.77-19.72), p=0.001, HR=6.20, CI 95% (2.19-17.52), respectively).

OS in *Combined matched cohort* (n=29/81) of ATE patients showed significant association with *GSTP1* rs1695 also when evaluating only ATE cases in Induction- like phase (n=18/49) (p=0.005, HR=0.23, CI 95% (0.08-0.64); p=0.005; HR=0.22, CI 95% (0.08-0.63)). OS associated with rs1695 in the seizure subgroup (n=14/37), as well (p=0.007; HR=0.15, CI 95% (0.04-0.59). These associations remained significant after FDR correction.

4.1.3 Single nucleotide polymorphisms associated with risk of CNS relapse

CNS relapse was analyzed on the non-matched population of Hungarian patients as a discovery population. We analyzed the impact of SNPs on CNS relapse comparing patients with isolated or combined CNS relapse to non-relapsed controls. We have found *CEP72* rs12522955, *SLC22A7* rs4149178, *CYP3A4* rs3735451, *ABCC1* rs3743527, *BCL2* rs4987853, *CYP3A5* rs4646450, *SLC22A7* rs4149178 in significant association with CNS relapse (p=0.009, OR=0.07, CI 95% (0.01-0.52), p=0.01, OR=4.01, CI 95% (1.40-11.52), p=0.005, OR=13.12, CI 95% (2.21-78.06), p=0.008, OR=0.41, CI 95% (0.21-0.79), p=0.007, OR=4.28, CI 95% (1.48-12.41), p=0.004, OR=2.34, CI 95% (1.32-4.17), p=0.002, OR=8.07, CI 95% (2.23-29.19), respectively).Next to this OS and EFS were calculated in this population. *CYP3A5* rs776746 was in association with OS, *CYP3A5* rs776746 and *SLC22A7* rs4149178 were in association with EFS in the Hungarian cohort(p=0.0003, HR=6.76 CI 95% (2.40-19.04), p=0.0006, HR=6.05 CI 95% (2.17-16.86), p=0.0008, HR=3.84 CI 95% (1.75-8.44), respectively).

We analyzed the impact of SNPs on CNS relapse comparing patients with isolated or combined CNS relapse to non-relapsed controls on the *Combined cohort* (n=86/129) also. *ABCB1* rs2032582 and rs1128503 showed significant association with CNS relapse when relapse cases were compared to patients having no relapse (p=0.019, p=0.038). After FDR correction these associations were not significant. OS and EFS of the *Combined cohort* with CNS-relapse did not show association with any of the SNPs.

4.2. Inverse association between CNS toxicity and relapse

Inverse association of two SNPs of *ABCB1* (rs1128503 and rs2032582) with chemotherapy-related adverse neurological events (seizure, n=44/89) and CNS relapse (n=86/129) was found on the *Combined cohorts* of ATE and CNS relapse. Patients having risk for toxicity-related seizures had a lower chance for CNS relapse and vice-versa.

4.3. Genetic events associated with risk of T-ALL relapse and survival

To identify T-ALL relapse-specific genes, the Haloplex target capture technique was used. The designed Haloplex panel contained exons of 324 genes; from these 313 genes were sufficiently covered on the T-ALL samples. The average coverage of the exons was 424 (median = 417), and 97% of them were covered more than 30-fold. We have found SNVs and small insertions and deletions (InDels) in the targeted sequences. Filtering out the false positives were a crucial point of the study. We excluded variants with allele frequency >1% and found in 1000 Genome Project release 2014 or in dbsnp138. Integrative Genomic Viewer (IGV) was used for the quality control of every suspected variant and the found sequencing artifacts were excluded. In the T-ALL population results of NGS of *NOTCH1* and *PTEN* genes was compared to Sanger-sequencing; the sensitivity of the NGS was above 90% for both genes.

The mean number of single nucleotide variations (SNVs) or small insertions and deletions (InDels) per sample was seven. Frequencies were similar in primary and relapsed samples regarding SNVs found in leukemia-associated genes. InDels were present more often in relapse samples which assumed that InDels may be acquired by chemotherapy. Cytosine deamination (exchanges of C>T) was the most frequent mutagenic event among nonsynonymous alterations. The majority of mutations were supposed to be heterozygote. Previously identified leukemia driver genes showed mutation densities similar to the

known. SNVs were found most often in *NOTCH1, PHF6, FBXW7,* and *PTEN* (55-12%); alterations in *DNM2, XIRP2,* and *CDH23* were found in more than 10% of the samples. Our aim was to find new genes in the development of T-ALL relapse and possible prognostic markers. To distinguish between driver and passenger genes, we have determined the mutation density (mutations per Mbp): the number of nonsynonymous SNVs and InDels were divided by the length of the targeted exons per gene. The mean density was 1.8/Mbp. The majority of the genes were below this value. We could confirm driver genes in this setting such as *NOTCH1*, *PTEN,* or *NT5C2*. In *NT5C2* 20 SNVs were detected in 24% of all relapse samples and one SNV in one initial diagnosis sample (p= 0.0001, Fisher's exact test). *NT5C2* had no impact on survival in this analysis, although, mutations of *NT5C2* were more often carried by patients with early relapses (p=0.01, Fisher's exact test).

5. Conclusions

1. Chemotherapeutic drug-metabolizing enzyme and transporter SNPs have a possible role in the prevalence of treatment-related cardiotoxicity in ALL (T-, and B-cell subtype) or OSC. Potential biomarkers can be the genetic variations in *ABCC2, NQO1, SLC22A6* predisposing for decreasing FS values. SNPs in *CYP3A4, CYP3A5* could be indicators for further FS changes.

2. Chemotherapeutic drug-metabolizing enzyme and transporter SNPs have a possible role in the prevalence of treatment-related neurological complications in ALL (T-, and B-cell subtype): including toxicity or CNS relapse. Significantly associated SNPs in carrier genes were *ABCB1*, *GSTP1* in association with neurotoxic events, AE, or ATE. *CYP3A4, CYP3A5, GSTP1* influenced prognosis in these populations.

3. Chemotherapeutic drug-metabolizing enzyme and transporter SNPs were studied with CNS relapse in *ABCB1*, *ABCC1*, *BCL2, CEP72*, *CYP3A4*, *CYP3A5*, *SLC22A7.* They influenced the appearance of CNS relapse or survival of the studied groups. Inverse association between the function of SNPs predisposing for

toxicity or relapse in ALL (T-, and B-cell subtype) was proved through two *ABCB1* SNPs.

4. *NT5C2* most probably contributes to T-ALL relapse.

6. Bibliography of the candidate's publications

Publications related to the thesis

Sági, J.C.; Gézsi, A.; Egyed, B.; Jakab, Z.; Benedek, N.; Attarbaschi, A.; Köhrer, S.; Sipek, J.; Winkowska, L.; Zaliova, M.; et al. Pharmacogenetics of the central nervous system—toxicity and relapse affecting the cns in pediatric acute lymphoblastic leukemia. Cancers (Basel). 2021, 13, doi:10.3390/cancers13102333.
IF 6.639 (2020)

Sági, J.C.*; Egyed, B.*; Kelemen, A.; Kutszegi, N.; Hegyi, M.; Gézsi, A.; Herlitschke, M.A.; Rzepiel, A.; Fodor, L.E.; Ottóffy, G.; et al. Possible roles of genetic variations in chemotherapy related cardiotoxicity in pediatric acute lymphoblastic leukemia and osteosarcoma. BMC Cancer 2018, 18, 1–14, doi:10.1186/s12885-018-4629-6.
IF 2.933

Richter-Pechańska, P.; Kunz, J.B.; Hof, J.; Zimmermann, M.; Rausch, T.; Bandapalli, O.R.; Orlova, E.; Scapinello, G.; **Sagi, J.C.**; Stanulla, M.; et al. Identification of a genetically defined ultra-high-risk group in relapsed pediatric T-lymphoblastic leukemia. Blood Cancer J. 2017, 7, doi:10.1038/bcj.2017.3.
IF 8.125

Publications not related to the thesis

Kutszegi, N.; Semsei, Á.F.; Gézsi, A.; **Sági, J.C.**; Nagy, V.; Csordás, K.; Jakab, Z.; Lautner-Csorba, O.; Gábor, K.M.; Kovács, G.T.; et al. Subgroups of paediatric acute lymphoblastic leukaemia might differ significantly in genetic predisposition to asparaginase hypersensitivity. PLoS One 2015, 10, 1–12, doi:10.1371/journal.pone.0140136.

IF 3.057

Csányiné Sági, J.; Félné Semsei, Á.; Erdélyi, D. Gyermekkori Leukémia. Élet és Tudomány 2015.

Sági, J.C.; Kutszegi, N.; Kelemen, A.; Fodor, L.E.; Gézsi, A.; Kovács, G.T.; Erdélyi, D.J.; Szalai, C.; Semsei, Á.F. Pharmacogenetics of anthracyclines. Pharmacogenomics. 2016, 1075–87, doi:10.2217/pgs-2016-0036.
IF 2.35

Rzepiel, A.; Kutszegi, N.; **Sági, J.C.**; Andrea, K.; Semsei, Á.F.; Buzás, E.; Erdélyi, D.J. Extracelluláris Vesiculák És Hematológiai Malignitásokban Játszott Szerepök. Orv. Hetil. 2016, 157, 1379–1384, doi:10.1556/650.2016.30532.
IF 0.349

Kutszegi, N.; Yang, X.; Gézsi, A.; Schermann, G.; Erdélyi, D.J.; Semsei, Á.F.; Gábor, K.M.; **Sági, J.C.**; Kovács, G.T.; Falus, A.; et al. HLA-DRB1*07:01-HLA-DQA1*02:01-HLADQB1* 02:02 haplotype is associated with a high risk of asparaginase hypersensitivity in acute lymphoblastic leukemia. Haematologica 2017, 102, 1578–1586, doi:10.3324/haematol.2017.168211.
IF 9.09

Egyed, B.; Kovács, G.; Kutszegi, N.; Rzepiel, A.; **Csányiné Sági, J.**; Erdélyi, D.J.; Müller, J.; Félné Semsei, Á. New and traditional directions in the biology and management of childhood acute lymphoblastic leukemia. Orv. Hetil. 2018, 159, 786–797, doi:10.1556/650.2018.31059.
IF 0.322

Rzepiel, A.; Kutszegi, N.; Gézsi, A.; **Sági, J.C.**; Egyed, B.; Péter, G.; Butz, H.; Nyíro, G.; Müller, J.; Kovács, G.T.; et al. Circulating microRNAs as minimal residual disease biomarkers in childhood acute lymphoblastic leukemia. J. Transl. Med. 2019, 17, 1–16, doi:10.1186/s12967-019-2114-x.

IF 4.124

Egyed, B.; Kutszegi, N.**; Ságí, J.C.**; Gézsi, A.; Rzepiel, A.; Visnovitz, T.; Lorincz, P.; Müller, J.; Zombori, M.; Szalai, C.; et al. MicroRNA-181a as novel liquid biopsy marker of central nervous system involvement in pediatric acute lymphoblastic leukemia. J. Transl. Med. 2020, 18, 1–12, doi:10.1186/s12967-020-02415-8.
IF 5.531

Cumulative IF: 42.762
First author IF: 11.922