GERMLINE PREDISPOSITION TO CHILDHOOD ACUTE LYMPHOBLASTIC LEUKEMIA AND BONE MARROW FAILURE, AND MITOCHONDRIAL DNA VARIANTS IN LEUKEMIA

Tekla Järviaho
TEKLA JÄRVIAHO

GERMLINE PREDISPOSITION TO CHILDHOOD ACUTE LYMPHOBLASTIC LEUKEMIA AND BONE MARROW FAILURE, AND MITOCHONDRIAL DNA VARIANTS IN LEUKEMIA

Academic dissertation to be presented with the assent of the Doctoral Training Committee of Health and Biosciences of the University of Oulu for public defence in Auditorium 12 of Oulu University Hospital, on 12 October 2018, at 12 noon

UNIVERSITY OF OULU, OULU 2018
Childhood acute lymphoblastic leukemia (ALL) is the most common cancer in children. The overall survival rate has reached to 90%. However, ALL still presents a significant disease burden and is a major cause for deaths in children.

Recently, both inherited germline variants related to ALL susceptibility and somatic genetic variants forming novel subgroups of ALL have been discovered. In this thesis two families with familial ALL were studied. Constitutional heterozygous microdeletion at chromosome 7p12.1p13, including \textit{IKZF1}, was discovered in the first family with intellectual impairment, overgrowth, and susceptibility to childhood ALL. In the second family, constitutional chromosome translocation was revealed in two individuals with childhood ALL and, subsequently, in seven unaffected family members. The balanced reciprocal translocation t(12;14)(p13.2;q23.1) resulted in breakpoints on two genes; \textit{ETV6} on chromosome 12 and \textit{RTN1} on chromosome 14. Only a few familial and sporadic ALL cases with germline variants in either \textit{IKZF1} or \textit{ETV6} have been published, thus supporting the significant role of these constitutional variants in childhood ALL predisposition.

Inherited bone marrow failure syndromes (IBMFS) may predispose to childhood leukemia, including ALL. Two unrelated patients were diagnosed with bone marrow failure without the symptoms of classical IBMFS. Neither patient had any signs of developmental delay or congenital anomalies. Exome sequencing revealed identical c.1457del(p.(Ile 486fs)) mutation on the \textit{ERCC6L2} gene in both patients. A few patients with IBMFS and \textit{ERCC6L2} variants have been described in previous studies. Some of them also had congenital craniofacial anomalies and developmental delay that were not detected in the patients in this thesis.

The ALL cohort study on genetic variation of mitochondrial DNA (mtDNA) included 36 children. Metabolic change where malignant cells uncouple energy production from oxidative phosphorylation (OXPHOS) is one of the established hallmarks of cancer. In the cohort in this study, 22% of patients harbored nonsynonymous variants on mtDNA in the protein-coding genes of OXPHOS enzyme complexes. The somatic non-neutral variants were found in patients with a poor prognosis cytogenetic marker. The results support the hypothesis that cancer cells harbor mtDNA variants that may affect the cell metabolism.

\textit{Keywords:} bone marrow failure syndrome, children, \textit{ERCC6L2}, \textit{ETV6}, germline predisposition, \textit{IKZF1}, leukemia, mitochondrial DNA
Järviaho, Tekla, Perinnöllinen alttius lasten akuutissa lymfoblastileukemiassa ja luuytimen toimintahäiriöissä sekä mitokondrialisen DNA:n muutokset leukemiassa.

Oulun yliopiston tutkijakoulu; Oulun yliopisto, Lääketieteellinen tiedekunta; Medical Research Center Oulu; Biocenter Oulu; Oulun yliopistollinen sairaala

Acta Univ. Oul. D 1485, 2018

Oulun yliopisto, PL 8000, 90014 Oulun yliopisto

Tiivistelmä

Akuutti lymfoblastileukemia (ALL) on lasten yleisin syöpä. Vaikka nykyisin noin 90 prosenttia paranee, ALL aiheuttaa huomattavan paljon sairastavaa ja on merkittävä lasten kuolinsyy.


Potilaskohortti tutkimuksessa tutkittiin mitokondrialisen DNA:n (mtDNA) muutoksia ALL:aan sairastuneilla lapsilla. Syöpäsolut eivät hyödyllä mitokondriisen elektroninsiirtoketjun energian tuotantoon, ja tämä aineenvaihdunnan muutos on tunnustettu syövän ominaisuutena. Tutkimuksessa havaittiin, että 22 prosentilla potilaista ilmeni diagnoosivaiheessa poikkeavia mtDNA:n muutoksia, jotka olivat elektroninsiirtoketjun entsyymien alavyksiköitä koottavissa geneissä. Muutoksia todettiin useimmilla potilailla, joilla oli leukemiasoluissa huonon ennusteen geneetinen tekijä. Havaitut muutokset voivat mahdollisesti vaikuttaa leukemiasolun energia-aineenvaihduntaan.

Asiasanat: ERCC6L2, ETV6, IKZF1, lapset, leukemia, luuytimen toimintahäiriö, mitokondrialinen DNA, perinnöllinen alttius
To my love Jyri
Acknowledgements

This work was conducted at the PEDEGO Research Unit at the University of Oulu from 2012 to 2018. I want to thank all who participated in this project.

I owe my greatest gratitude to my excellent supervisors: Professor Arja Harila-Saari, M.D.; Professor Johanna Uusimaa, M.D.; and Riitta Niinimäki, M.D, Ph.D. You are exceptional role models and I am honored to have had you as my supervisors. You are my “Three Good Fairies.”

Arja, you have taught me about life, research, and being a professional clinician and academic. Johanna, I have learned from you what it takes to be a successful group leader. Your resilience, ambition and attitude to research are admirable. Riitta, you have become not only my supervisor but also a mentor and senior colleague during this Ph.D. process.

I thank Professor Mika Rämet, M.D., the head of the PEDEGO Research Unit, and Docent Päivi Tapanainen, M.D., the chief physician at the Department of Pediatrics, for providing the infrastructure for the research and facilitating the completion of this thesis.

My gratitude goes also to the official reviewers of the thesis: Docent Gisela Barbany, M.D. and Docent Minna Koskenvuo, M.D. I respect your expertise and appreciated your thoughtful comments during the review process. Thank you to my follow-up group members, Docent Petri Kulmala, M.D. and Hanna Juntti, M.D., Ph.D., for your guidance and flexibility along the way.

I thank Professor Ann Nordgren, M.D. from Karolinska Institutet and senior co-author on two of my thesis articles for sharing her expertise in the field and providing resources and infrastructure. I am most grateful to my co-authors from the University of Oulu: Anri Hurme-Niiranen, M.Sc.; Kimmo Halt, M.D.; Docent Reetta Hinttala; Docent Merja Möttönen, M.D.; Heidi Soini, Ph.D.; Docent Jukka Moilanen, M.D.; Docent Eeva-Riitta Savolainen, M.D.; Elisa Rahikkala, M.D., Ph.D.; and Docent Pasi Hirvikoski, M.D. Thank you to all my Swedish co-authors: Vasilios Zachariadis, M.D., Ph.D.; Benedicte Bang, M.D.; Fulya Taylan, Ph.D.; Bianca Tesi, M.D., Ph.D.; Docent Yenan Bryceson; Professor Edvard Smith M.D.; and Samuel Chiang, Ph.D.

I am grateful to our laboratory personnel Pirjo Keränen and Anja Heikkinen for all their help. Thank you to Saara Korpela, who performed some of the laboratory experiments as a part of her studies. I am also very grateful to research nurse Kirsi Kvist-Mäkelä for providing and handling all of the patient samples analyzed in this thesis.
Thanks to my research colleagues at the Clinical Research Center Oulu for support, guidance, and laughter: Anri Hurme-Niiranen, M.Sc.; Laura Kytövuori, Ph.D.; Milla-Riikka Hautakangas, M.D.; Salla Kangas, Ph.D.; Teija Paakkola M.Sc.; Johanna Hynynen, M.D.; Heidi Soini, Ph.D.; Naemeh Nayebzadeh, M.Sc.; and all the rest of you.

I thank my family: my parents Irmeli and Jouko Järviaho; my sister, Johanna Järviaho; and my brother, Ermo Järviaho. Thank you to my extra-parents in Oulu: my aunt Jaana O. Liimatainen, Ph.D., and Petri Salonen. Jaana, I am glad to be the second in our family to defend a thesis. I also express my gratitude to my parents-in-law, Yrjö Harju and Leena Moilan-Harju, for their support, love, and welcome into the family.

My friend and colleague, Jupe Tolonen, M.D., thank you for all the support, all the laughter, and all the love. I truly hope our symbiosis in life will continue for years to come. Your wise words have always guided me. To all my friends—Jenni Marttila, Emmi Ihalainen, Minna Juntunen, Taru Tanhua, Sanna Pesonen, Pieta Myllykangas, and the rest of you—thank you for your friendship, love, and support.

And last but not least, Jyri, you are my guiding light, my love, and my best friend. Without your support, your joyful spirit, inspiring words, and fairy tales I would have not achieved this much in life. You have carried me through this thesis, so I thank you. We make an excellent team.

The work was financially supported by the Cancer Society of Finland; the Alma and K.A. Snellman Foundation; the Maud Kuistila Memorial Foundation; the Finnish Foundation for Pediatric Research; the Väre Foundation for Pediatric Cancer Research; Special State Grants for Health Research in the Department of Pediatrics and Adolescence at Oulu University Hospital, Oulu, Finland; and the Thelma Mäkikyrö Foundation. All of the aforementioned are gratefully acknowledged.

Oulu, August 2018

Tekla Järviaho
## Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABCA13</td>
<td>ATP binding cassette subfamily A member 13</td>
</tr>
<tr>
<td>ABL1</td>
<td>ABL proto-oncogene 1, non-receptor tyrosine kinase</td>
</tr>
<tr>
<td>ADCY1</td>
<td>adenylate cyclase 1</td>
</tr>
<tr>
<td>ADP</td>
<td>adenosine diphosphate</td>
</tr>
<tr>
<td>ALL</td>
<td>acute lymphoblastic leukemia</td>
</tr>
<tr>
<td>AML</td>
<td>acute myeloid leukemia</td>
</tr>
<tr>
<td>ARID5B</td>
<td>AT-rich interaction domain 5B</td>
</tr>
<tr>
<td>ATG</td>
<td>start codon</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>ATP5I</td>
<td>ATP synthase membrane subunit e</td>
</tr>
<tr>
<td>BCP-ALL</td>
<td>B cell precursor acute lymphoblastic leukemia</td>
</tr>
<tr>
<td>BCR</td>
<td>BCR, RhoGEF and GTPase activating protein</td>
</tr>
<tr>
<td>BLAT</td>
<td>basic local alignment tool</td>
</tr>
<tr>
<td>BMFS</td>
<td>bone marrow failure syndrome</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BRCA1/2</td>
<td>BRCA1/2, DNA repair associated</td>
</tr>
<tr>
<td>BWA</td>
<td>Burrows-Wheeler Aligner</td>
</tr>
<tr>
<td>CBS</td>
<td>circular binary segmentation</td>
</tr>
<tr>
<td>CCM2</td>
<td>CCM2 scaffold protein</td>
</tr>
<tr>
<td>CDKN2A/B</td>
<td>cyclin dependent kinase inhibitor 2A/B</td>
</tr>
<tr>
<td>CEBPE</td>
<td>CCAAT enhancer binding protein epsilon</td>
</tr>
<tr>
<td>CMMRD</td>
<td>constitutional mismatch repair deficiency</td>
</tr>
<tr>
<td>CNA</td>
<td>copy number alteration</td>
</tr>
<tr>
<td>CNV</td>
<td>copy number variation</td>
</tr>
<tr>
<td>COBL</td>
<td>cordon-bleu WH2 repeat protein</td>
</tr>
<tr>
<td>CoQ</td>
<td>ubiquinone</td>
</tr>
<tr>
<td>CR1</td>
<td>first clinical remission</td>
</tr>
<tr>
<td>CREBBP</td>
<td>CREB binding protein</td>
</tr>
<tr>
<td>CRLF2</td>
<td>cytokine receptor like factor 2</td>
</tr>
<tr>
<td>CSGE</td>
<td>conformation sensitive gel electrophoresis</td>
</tr>
<tr>
<td>Cyt C</td>
<td>cytochrome c</td>
</tr>
<tr>
<td>C7orf57</td>
<td>chromosome 7 open reading frame 57</td>
</tr>
<tr>
<td>C7orf72</td>
<td>chromosome 7 open reading frame 72</td>
</tr>
<tr>
<td>C9orf3</td>
<td>chromosome 9 open reading frame 3</td>
</tr>
<tr>
<td>DDC</td>
<td>dopa decarboxylase</td>
</tr>
</tbody>
</table>
DDX56  DEAD-box helicase 56
DEAH  aspartate-glutamate-alanine-histidine motif
DKC1  dyskerin pseudouridine synthase 1
D-loop  non-coding displacement region
dx  diagnosis
E2  exon 2
E3  exon 3
E4  exon 4
ERCC6L2  ERCC excision repair 6 like 2
ERG  ERG, ETS transcription factor
ERK  extracellular regulated kinase
ETS  E26 transformation-specific
ETV6  ETS variant 6
EZH2  enhancer of zeste 2 polycomb repressive complex 2 subunit
FA  Fanconi anemia
FAM105B  family with sequence similarity 105, member B
FANC  FA complementation group
FGD5  FYVE, RhoGEF and PH domain containing 5
FIGNL1  fidgetin like 1
FISH  fluorescence in situ hybridization
fs  frameshift
GATA3  GATA binding protein 3
GATK  genome analysis toolkit
GRB10  growth factor receptor bound protein 10
GWAS  genome-wide association studies
H+  proton
H2AFV  H2A histone family member V
H2O  water molecule
HeH  high hyperdiploid
Hb  hemoglobin
HHIP  hedgehog interacting protein
HNSCC  head and neck squamous cell carcinoma
HSCT  hematopoietic stem cell transplantation
HR  high-risk
HUS1  HUS1 checkpoint clamp component
iAMP21  intrachromosomal amplification of chromosome 21
IBMFS  inherited bone marrow failure syndrome
<table>
<thead>
<tr>
<th>Term</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICLs</td>
<td>interstrand crosslinks</td>
</tr>
<tr>
<td>IGFBP1/3</td>
<td>insulin like growth factor binding protein 1/3</td>
</tr>
<tr>
<td>IGV</td>
<td>integrative genomics viewer</td>
</tr>
<tr>
<td>IKZF1</td>
<td>IKAROS family zinc finger 1</td>
</tr>
<tr>
<td>IM</td>
<td>inner membrane</td>
</tr>
<tr>
<td>IMS</td>
<td>inter-membrane space</td>
</tr>
<tr>
<td>IR</td>
<td>intermediate-risk</td>
</tr>
<tr>
<td>ISO-BMI</td>
<td>age and sex adjusted body mass index</td>
</tr>
<tr>
<td>JAK1/2</td>
<td>Janus kinase 1/2</td>
</tr>
<tr>
<td>kb</td>
<td>kilo base pairs</td>
</tr>
<tr>
<td>KMT2A</td>
<td>lysine methyltransferase 2A</td>
</tr>
<tr>
<td>KRAS</td>
<td>KRAS proto-oncogene, GTPase</td>
</tr>
<tr>
<td>LAMA1</td>
<td>laminin subunit alpha 1</td>
</tr>
<tr>
<td>MDS</td>
<td>myelodysplastic syndrome</td>
</tr>
<tr>
<td>MLL</td>
<td>mixed lineage leukemia</td>
</tr>
<tr>
<td>MRD</td>
<td>minimal residual disease</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MT-ATP6</td>
<td>ATP synthase membrane subunit 6</td>
</tr>
<tr>
<td>MT-ATP8</td>
<td>ATP synthase membrane subunit 8</td>
</tr>
<tr>
<td>MT-CO1/2</td>
<td>cytochrome C oxidase I/II</td>
</tr>
<tr>
<td>MT-CO3</td>
<td>cytochrome C oxidase III</td>
</tr>
<tr>
<td>MT-CYB</td>
<td>mitochondrially encoded cytochrome B</td>
</tr>
<tr>
<td>mtDNA</td>
<td>mitochondrial DNA</td>
</tr>
<tr>
<td>MT-ND1</td>
<td>NADH:ubiquinone oxidoreductase core subunit 1</td>
</tr>
<tr>
<td>MT-ND2</td>
<td>NADH:ubiquinone oxidoreductase core subunit 2</td>
</tr>
<tr>
<td>MT-ND4L</td>
<td>NADH:ubiquinone oxidoreductase core subunit 4L</td>
</tr>
<tr>
<td>MT-ND6</td>
<td>NADH:ubiquinone oxidoreductase core subunit 6</td>
</tr>
<tr>
<td>MT-RNR1</td>
<td>mitochondrially encoded 12S RNA</td>
</tr>
<tr>
<td>MT-RNR2</td>
<td>mitochondrially encoded 16S RNA</td>
</tr>
<tr>
<td>MYC</td>
<td>MYC proto-oncogene, bHLH transcription factor</td>
</tr>
<tr>
<td>MYO1G</td>
<td>myosin IG</td>
</tr>
<tr>
<td>NA</td>
<td>data not available</td>
</tr>
<tr>
<td>NACAD</td>
<td>NAC alpha domain containing</td>
</tr>
<tr>
<td>NAD+</td>
<td>nicotinamide adenine dinucleotide, oxidized</td>
</tr>
<tr>
<td>NADH</td>
<td>nicotinamide adenine dinucleotide, reduced</td>
</tr>
<tr>
<td>nDNA</td>
<td>nuclear DNA</td>
</tr>
</tbody>
</table>
NER  nucleotide excision repair
NHP2  NHP2 ribonucleoprotein
NK  natural killer (cell)
NOPHO  the Nordic Society of Paediatric Haematology and Oncology
NOP10  NOP10 ribonucleoprotein
NOS  not otherwise specified
NPC1L1  NPC1 like intracellular cholesterol transporter 1
NRAS  NRAS proto-oncogene, GTPase
NSD1  nuclear receptor binding SET domain protein 1
OGDH  oxoglutarate dehydrogenase
OM  outer membrane
OXPHOS  oxidative phosphorylation
PAX5  paired box 5
PCR  polymerase chain reaction
Ph  Philadelphia chromosome
PHEX  phosphate regulating endopeptidase homolog X-linked
PHF6  PHD finger protein 6
PI4K2A  phosphatidylinositol 4-kinase type 2 alpha
PKD1L1  polycystin 1 like 1, transient receptor potential channel interacting
PNT  pointed domain
PP1A  peptidylprolyl isomerase A
PTEN  phosphatase and tensin homolog
PURB  purine rich element binding protein B
Q  ubiquinone
RAMP3  receptor activity modifying protein 3
Ras  Ras small GTPase
RB1  RB transcriptional corepressor 1
RBC  red blood cell
rCRS  revised Cambridge reference sequence
RDH16  retinol dehydrogenase 16
RFLP  restriction fragment length polymorphism
RFWD2  ring finger and WD repeat domain 2
RHOX2B  Rhox homeobox family member 2B
RIOK1  RIO kinase 1
RPL  ribosomal protein L
RPS  ribosomal protein S
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Full Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>rRNA</td>
<td>ribosomal RNA</td>
</tr>
<tr>
<td>r(21)c</td>
<td>ring chromosome 21</td>
</tr>
<tr>
<td>RUNX1</td>
<td>runt related transcription factor 1</td>
</tr>
<tr>
<td>SALL4</td>
<td>spalt like transcription factor 4</td>
</tr>
<tr>
<td>SBDS</td>
<td>SBDS, ribosome maturation factor</td>
</tr>
<tr>
<td>SD</td>
<td>standard deviation</td>
</tr>
<tr>
<td>SFXN4</td>
<td>sideroflexin 4</td>
</tr>
<tr>
<td>SH2B3</td>
<td>SH2B adaptor protein 3</td>
</tr>
<tr>
<td>SLC2A6</td>
<td>solute carrier family 2 member 6</td>
</tr>
<tr>
<td>SNP</td>
<td>single nucleotide polymorphism</td>
</tr>
<tr>
<td>STR</td>
<td>short tandem repeat</td>
</tr>
<tr>
<td>SR</td>
<td>standard-risk</td>
</tr>
<tr>
<td>SUN3</td>
<td>Sad1 and UNC84 domain containing 3</td>
</tr>
<tr>
<td>TBRG4</td>
<td>transforming growth factor beta regulator 4</td>
</tr>
<tr>
<td>TCA</td>
<td>tricarboxylic acid cycle</td>
</tr>
<tr>
<td>TERC</td>
<td>telomerase RNA component</td>
</tr>
<tr>
<td>TERT</td>
<td>telomerase reverse transcriptase</td>
</tr>
<tr>
<td>THC5</td>
<td>thrombocytopenia 5</td>
</tr>
<tr>
<td>TINF2</td>
<td>TERF1 interacting nuclear factor 2</td>
</tr>
<tr>
<td>TMED4</td>
<td>transmembrane p24 trafficking protein 4</td>
</tr>
<tr>
<td>TNS3</td>
<td>tensin 3</td>
</tr>
<tr>
<td>TP53</td>
<td>tumor protein p53</td>
</tr>
<tr>
<td>TP63</td>
<td>tumor protein p63</td>
</tr>
<tr>
<td>TPK2</td>
<td>tyrosine kinase 2</td>
</tr>
<tr>
<td>UPP1</td>
<td>uridine phosphorylase 1</td>
</tr>
<tr>
<td>VCF</td>
<td>variant call format</td>
</tr>
<tr>
<td>VEP</td>
<td>variant effect predictor</td>
</tr>
<tr>
<td>VWC2</td>
<td>von Willebrand factor C domain containing 2</td>
</tr>
<tr>
<td>WBC</td>
<td>white blood cell</td>
</tr>
<tr>
<td>WGA</td>
<td>whole genome amplification</td>
</tr>
<tr>
<td>WGS</td>
<td>whole genome sequencing</td>
</tr>
<tr>
<td>WHO</td>
<td>World Health Organization</td>
</tr>
<tr>
<td>X</td>
<td>unknown amino acid</td>
</tr>
<tr>
<td>ZC3H18</td>
<td>zinc finger CCCH-type containing 18</td>
</tr>
<tr>
<td>ZCCHC16</td>
<td>zinc finger CCHC-type containing 16</td>
</tr>
<tr>
<td>ZMIZ2</td>
<td>zinc finger MIZ-type containing 2</td>
</tr>
<tr>
<td>ZPBP</td>
<td>zona pellucida binding protein</td>
</tr>
</tbody>
</table>
Amino acids

<table>
<thead>
<tr>
<th>Letter</th>
<th>Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>A, Ala</td>
<td>Alanine</td>
</tr>
<tr>
<td>C, Cys</td>
<td>Cysteine</td>
</tr>
<tr>
<td>D, Asp</td>
<td>Aspartate</td>
</tr>
<tr>
<td>E, Glu</td>
<td>Glutamate</td>
</tr>
<tr>
<td>F, Phe</td>
<td>Phenylalanine</td>
</tr>
<tr>
<td>G, Gly</td>
<td>Glycine</td>
</tr>
<tr>
<td>H, His</td>
<td>Histidine</td>
</tr>
<tr>
<td>I, Ile</td>
<td>Isoleucine</td>
</tr>
<tr>
<td>K, Lys</td>
<td>Lysine</td>
</tr>
<tr>
<td>L, Leu</td>
<td>Leucine</td>
</tr>
<tr>
<td>M, Met</td>
<td>Methione</td>
</tr>
<tr>
<td>N, Asn</td>
<td>Asparagine</td>
</tr>
<tr>
<td>P, Pro</td>
<td>Proline</td>
</tr>
<tr>
<td>Q, Gln</td>
<td>Glutamine</td>
</tr>
<tr>
<td>R, Arg</td>
<td>Arginine</td>
</tr>
<tr>
<td>S, Ser</td>
<td>Serine</td>
</tr>
<tr>
<td>T, Thr</td>
<td>Threonine</td>
</tr>
<tr>
<td>V, Val</td>
<td>Valine</td>
</tr>
<tr>
<td>W, Trp</td>
<td>Tryptophan</td>
</tr>
<tr>
<td>Y, Tyr</td>
<td>Tyrosine</td>
</tr>
</tbody>
</table>
List of original publications

This thesis is based on the following publications, which are referred throughout the text by their Roman numerals:


*= equal contribution; **= equal contribution for senior authors
# Table of contents

Abstract

Tiiivistelmä

Acknowledgements 9

Abbreviations 11

List of original publications 17

Table of contents 19

1 Introduction 23

2 Review of the literature 25

2.1 Childhood acute lymphoblastic leukemia .............................................. 25

2.1.1 Incidence and prognosis ................................................................. 25

2.1.2 Treatment of childhood ALL .......................................................... 25

2.1.3 Immunophenotype ........................................................................ 26

2.2 Genetic etiology of ALL ................................................................. 26

2.2.1 Somatic aberrations .................................................................... 26

2.2.2 Inherited susceptibility in ALL ..................................................... 29

2.2.3 Genetic pathway of leukemogenesis in ALL ............................... 31

2.3 Bone marrow failure syndrome ....................................................... 33

2.3.1 Inherited bone marrow failure syndromes ................................ 33

2.3.2 Genetic mechanisms of IBMFS .................................................... 36

2.4 Mitochondrial genome and cancer .................................................... 37

2.4.1 Mitochondria and their function ................................................... 37

2.4.2 The mitochondrial genome ........................................................... 39

2.4.3 The Warburg hypothesis and metabolism in cancer cells .......... 39

2.4.4 MtDNA and cancer .................................................................... 41

3 Aims of the study 43

4 Patients, materials, and methods 45

4.1 Patients ............................................................................................... 45

4.1.1 Patients with familial ALL (I, II) .................................................. 45

4.1.2 Patients with bone marrow failure (III) ....................................... 45

4.1.3 Patient cohort in the mtDNA sequence variation study

(IV) ......................................................................................................... 46

4.1.4 Controls (IV) ............................................................................... 46

4.2 Ethical considerations (I–IV) ............................................................. 46

4.3 Molecular methods ............................................................................. 47

4.3.1 Familial ALL (I, II) .................................................................... 47

19
5 Results

5.1 Familial ALL and intellectual impairment syndrome caused by constitutional microdeletion of 7p12.1p13, including \textit{IKZF1} (I) ........................................... 55
   5.1.1 Overgrowth and intellectual impairment syndrome with susceptibility to childhood ALL (I) ......................................................... 55
   5.1.2 Microdeletion of 7p12.1p13, including \textit{IKZF1}, causes developmental delay syndrome with predisposition to childhood ALL (I) ......................................................... 57
   5.1.3 Characteristics of the leukemia with germline \textit{IKZF1} deletion (I) .......................................................... 58
   5.1.4 Expression analysis of the \textit{IKZF1} (I) ......................................................... 58
   5.1.5 Multi-parametric flow cytometry (I) ......................................................... 58

5.2 Familial ALL caused by constitutional chromosome translocation t(12;14)(p13.2;q23.1) disrupting \textit{ETV6} (II) ........................................... 59
   5.2.1 The family (II) .............................................................................. 59
   5.2.2 The constitutional chromosome translocation (II) ........................ 59
   5.2.3 Characteristics of leukemia with constitutional chromosome translocation disrupting \textit{ETV6} (II) ........................................... 62
   5.2.4 Expression analysis (II) .......................................................... 62

5.3 \textit{ERCC6L2} gene mutation causes inherited bone marrow failure syndrome (III) ................................................................................. 63
   5.3.1 Patients with bone marrow failure without congenital anomalies (III) .............................................................................. 63
   5.3.2 Truncating mutation of \textit{ERCC6L2} results in bone marrow failure (III) .......................................................... 65

5.4 Mitochondrial DNA sequence variation in patients with childhood ALL (IV) .......................................................... 65
   5.4.1 Somatic novel nonneutral mtDNA variants (IV) ............................. 65
   5.4.2 Non-neutral mtDNA variants associated with poor prognosis (IV) .......................................................... 67
   5.4.3 Mitochondrial haplogroups and detection of external DNA from blood transfusions (IV) .......................................................... 68

6 Discussion

6.1 Germline variants of \textit{IKZF1} and \textit{ETV6} predispose to childhood ALL .............................................................................. 71
6.1.1 Germline deletion of 7p12.1p13 causes intellectual impairment syndrome and susceptibility to childhood ALL ................................................................. 71
6.1.2 Predisposition to childhood BCP-ALL caused by a constitutional chromosome translocation disrupting ETV6 ............ 72
6.1.3 Germline predisposition to ALL; and future perspectives ........ 74
6.2 ERCC6L2 causes IBMFS without congenital anomalies ............ 75
6.2.1 Truncating c.1457del mutation in the ERCC6L2 gene ............ 75
6.2.2 ERCC6L2 and IBMFS .............................................................. 75
6.2.3 Further studies ................................................................. 76
6.3 Novel non-neutral mtDNA variants in the OXPHOS proteins detected in childhood ALL patients .................................................. 77
6.3.1 Main findings ................................................................. 77
6.3.2 Mutations on OXPHOS subunits ........................................ 77
6.3.3 Further studies .............................................................. 78
6.4 Strengths and limitations of the study ................................... 79
6.4.1 Strengths of the study ................................................... 79
6.4.2 Limitations of the study ................................................ 80
6.4.3 Future ........................................................................... 80
7 Conclusions .......................... 83
References .......................... 85
Original publications ......... 97
1 Introduction

Childhood acute lymphoblastic leukemia (ALL) is the most common cancer in children that has a high cure rate of almost 90% in high-income countries (Hunger & Mullighan, 2015). The etiology of the disease has been largely unknown, but recent breakthroughs in genetics have implicated novel somatic and inherited variants associated with disease progression, chemotherapy response, and inherited susceptibility to ALL (Pui, Yang et al., 2015).

Germline predisposition to childhood cancer, including ALL, has become more evident (Gröbner et al., 2018; Ripperger et al., 2017). J. Zhang et al. (2015) reported 4.4% of patients to have a pathogenic germline variant in known cancer-causing genes in a cohort of 588 childhood leukemia patients. Reports on familial ALL have revealed inherited variants in various genes that predispose to the disease, including *ETV6* (Melazzini et al., 2016; Moriyama et al., 2015; Noetzli et al., 2015; Topka et al., 2015; M. Y. Zhang et al., 2015), *IKZF1* (Churchman et al., 2018; Hoshino et al., 2017; Kuehn et al., 2016; Mendoza-Londono, Kashork, Shaffer, Krance, & Plon, 2005) and *PAX5* (Auer et al., 2014; Shah et al., 2013).

Congenital syndromes, including Noonan syndrome and other RASopathies (Cavé et al., 2016), and inherited bone marrow failure syndromes (IBMFS) may predispose to childhood ALL (Wegman-Ostrosky & Savage, 2017). However, more often, the leukemias associated with IBMFS are acute myeloid leukemia (AML) (Shimamura & Alter, 2010). IBMFS are caused by genetic variants on key cellular pathways, including DNA damage response and ribosome and telomere maintenance (Wegman-Ostrosky & Savage, 2017). Various genes, including *FANCA*, *BRCA1*, and *ERCC6L2*, are associated with the syndromes (Bluteau et al., 2017). The altered genes are essential for cell survival and development, and thus, they are also fundamental in carcinogenesis, explaining the increased risk for malignancies with IBMFS (Wegman-Ostrosky & Savage, 2017).

One of the established hallmarks of cancer by Hanahan & Weinberg (2011) is the metabolic change in the cancer cell. Malignant cells rely on glycolysis and do not use the oxidative phosphorylation (OXPHOS) in the mitochondrion to produce energy as normal cells do (Hanahan & Weinberg, 2011). Low OXPHOS activity and increased glycolysis are crucial for the growth and survival of a cancer cell (Coller, 2014). Mitochondrial DNA (mtDNA) encodes 13 essential subunits of OXPHOS enzyme complexes, and somatic mtDNA variants have been reported in various cancers (Cook & Higuchi, 2012; Dasgupta & Hertweck, 2017). Few studies on mtDNA variation in childhood ALL have been published. One such study
revealed variation on the non-coding regulatory displacement loop (D-loop) of the mtDNA (Kwok, Quah, Ariffin, Tay, & Yeoh, 2011).

In this thesis, two families with familial ALL were studied to determine the inherited genetic variant predisposing to childhood ALL. Familial ALL is rare (Kharazmi et al., 2012), and it provides unique opportunities for studying the genetic etiology of ALL. As IBMFS may predispose to childhood leukemia, two unrelated children with progressive bone marrow failure with no congenital anomalies or signs of classical IBMFS were studied to establish the cause of syndrome in these patients. Because metabolism and the role of mitochondria are important in carcinogenesis and the knowledge of mtDNA coding region variants in childhood ALL is limited, the whole mtDNA in 36 childhood ALL patients was analyzed.
2  Review of the literature

2.1  Childhood acute lymphoblastic leukemia

2.1.1  Incidence and prognosis

The most common cancer in childhood is acute lymphoblastic leukemia (ALL) with peak incidence in children 3–5 years old (Hunger & Mullighan, 2015). In Finland, the incidence in children and adolescents 0–19 years is 40–50 new cases per year (Engholm et al., 2010). The first clinical remission for five patients, although temporary, was achieved by Sydney Farber in 1948 (Farber, Diamond, Mercer, Sylvester Jr, & Wolff, 1948). Since then, the cure rate has reached 90% in high-income countries because of multidrug chemotherapy, improved supportive care, international collaborations on treatment guidelines, and increased knowledge about the biology of the childhood ALL subtypes (Inaba, Greaves, & Mullighan, 2013).

2.1.2  Treatment of childhood ALL

Childhood ALL chemotherapy lasts 2.5 years with the current Nordic treatment protocol, the Nordic Society of Paediatric and Haematology (NOPHO) ALL2008 protocol, which is composed of induction, consolidation, and maintenance therapy (Toft et al., 2018). Patients are assigned to three different risk groups, standard risk (SR), intermediate risk (IR) and high risk (HR), which receive variable levels of intensity of chemotherapy based on host and leukemia characteristics (Toft et al., 2013). Allogeneic hematopoietic stem cell transplantation (HSCT) is used for some high-risk and relapsed patients (Toft et al., 2018). The risk stratification characteristics include the white blood cell count in peripheral blood at diagnosis, immunophenotype and genetic abnormalities of the leukemic blast cell, the presence of extramedullary disease, and the early response to chemotherapy based on minimal residual disease (MRD) (Inaba et al., 2013). The intensive chemotherapy may cause treatment-related toxicity, exposure for secondary malignancies, and late effects that impair the quality of life (Dixon et al., 2018; Robison, 2011).

A major improvement over the last decade has been the use of MRD measurement in monitoring chemotherapy response that has become the main
criterion for risk assessment (Toft et al., 2018). Regardless of morphological remission, blast cells can be detected in the bone marrow with either conventional flow cytometry using cell surface markers or with more advanced molecular genetic methods of sequencing immunoglobulin and/or T cell receptor genes (Borowitz et al., 2008; Pui, Pei et al., 2015). Despite improved risk stratification the prediction of relapse on the individual level is difficult, and the prognosis for relapsed ALL has remained poor (Borowitz et al., 2008; Nguyen et al., 2008).

### 2.1.3 Immunophenotype

ALL arises from malignant lymphoid blast cells of the bone marrow and can affect either the B or T lineage cells. The two different lineages are referred to as B cell precursor ALL (BCP-ALL) or T-ALL, with 85% of the cases being BCP-ALL. B or T lineage ALL are further divided into smaller groups based on the molecular and genetic characteristics of the malignant blast cell, i.e., cytogenetics. (Bhojwani, Yang, & Pui, 2015)

### 2.2 Genetic etiology of ALL

#### 2.2.1 Somatic aberrations

*Conventional genetic aberrations*

In the latest World Health Organization (WHO) classification of hematopoietic malignancies by Arber et al. (2016), B-ALL was divided into eight subgroups based on genetic aberrations, which are somatic, i.e., acquired, variants and are detected in the leukemic blasts: chromosomal translocation t(12;21) (ETV6-RUNX1), hyperdiploidy (>50 chromosomes), Philadelphia chromosome (Ph) translocation t(9;22) (BCR-ABL1), KMT2A (MLL) rearrangement, hypodiploidy (<44 chromosomes), t(1;19) and t(5;14), and not otherwise specified (NOS) groups, in addition to two provisional subgroups that require further research: BCR-ABL1-like (Ph-like) and the intrachromosomal amplification of chromosome 21 (iAMP21). T-ALL has two provisional subgroups: early T-cell precursor ALL and natural killer cell (NK) lymphoblastic leukemia or lymphoma (Arber et al., 2016).

The conventional genetic subgroups are an important part of risk stratification (Bhojwani et al., 2015). The most common cytogenetic aberrations—hyperdiploidy
and *ETV6-RUNX1*, which together cover roughly 50% of BCP-ALL cases—are both markers for good prognosis (Inaba et al., 2013). The frequency of cytogenetic aberrations in BCP-ALL are depicted in Figure 1. Hypodiploidy, which may be a manifestation of Li-Fraumeni syndrome with constitutional *TP53* mutation (Holmfeldt et al., 2013), and *KMT2A* gene-rearranged ALL, which is usually initiated in utero and affects infants, both have significantly inferior prognosis (Mullighan, 2012).

**Novel genetic variants**

In addition to the WHO classification, there are novel genetic subtypes with reported risk assessment, e.g., dicentric chromosome or dic(9;20) in BCP-ALL associated with an unfavorable prognosis, which may be overcome by intensified chemotherapy (Forestier et al., 2008; Pichler et al., 2010).

Many novel genetic variants are focal deletions involving single genes that are tumor suppressors or lymphoid signaling molecules and are involved in lymphoid development and differentiation, transcription, and cell-cycle regulation (Hunger & Mullighan, 2015; Inaba et al., 2013). Deletions of *IKZF1*, which encodes a lymphoid transcription factor IKAROS, are frequent in *BCR-ABL1* and *BCR-ABL1*-like ALL as well in other BCP-ALL subtypes, and they are related to poor prognosis with an increased risk of relapse (Kuiper et al., 2010; Mullighan et al., 2009). Another gene encoding transcriptional factor essential in lymphoid development is *CREBBP*, which is somatically mutated in ALL and associated with an increased risk of relapse (Mullighan et al., 2011).

Deletions or mutations of *PAX5*, which is a transcription factor involved in lymphoid B cell development, are common in BCP-ALL (Nebral et al., 2009). Recently, somatic intragenic amplification or gain in copy number of the *PAX5* gene was proposed for creating a novel subtype representing 1% of BCP-ALL (Schwab et al., 2017). The *PAX5* amplification was associated with *CDKN2A/B* deletions (Schwab et al., 2017), which result in deregulation of the tumor suppressor p53 encoded by *TP53* (Eswaran et al., 2015). Focal deletions of *ETV6*, which is involved in lymphoid development and is a tumor suppressor, have been frequently detected in BCP-ALL, particularly with *ETV6-RUNX1* fusion as additional deletion (Irvin et al., 2003).
Fig. 1. Distribution of the cytogenetic subtypes of childhood BCP-ALL. The pie chart represents the subgroups defined by WHO, excluding t(5;14) (Arber et al., 2016). NOS, not otherwise specified. Adapted from Bhojwani et al. (2015).
2.2.2 Inherited susceptibility in ALL

Constitutional single-gene variants and chromosomal aberrations

In two recent reports of large childhood cancer cohorts, genetic predisposition accounted for 8.5% (J. Zhang et al., 2015) and 6% (Gröbner et al., 2018) on the basis of the mutations in the known cancer-causing genes. Rare germline mutations in several genes have been reported to significantly predispose to childhood ALL, including ETV6 (Moriyama et al., 2015; Noetzli et al., 2015; M. Y. Zhang et al., 2015), IKZF1 (Churchman et al., 2018; Kuehn et al., 2016), PAX5 (Auer et al., 2014; Shah et al., 2013), SH2B3 (Perez-Garcia et al., 2013), RUNX1 (Linden, Schnittger, Groll, Juergens, & Rossig, 2010), and TYK2 (Waanders et al., 2017). Most of the genes, for example IKZF1, ETV6, RUNX1 and PAX5, have somatic mutations that are frequently detected in ALL (Mullighan, 2012).

Constitutional TP53 gene mutations are a representation of Li Fraumeni syndrome, which is an inherited cancer predisposition syndrome (Kamihara, Rana, & Garber, 2014). The hypodiploid subtype of childhood ALL is one manifestation of the syndrome (Holmfeldt et al., 2013; Powell et al., 2013; Qian et al., 2018). Germline BRCA2 mutations, which are usually associated with an increased breast and ovarian cancer risk, have been implicated in T-ALL (Wagner et al., 2004) and reported in B cell ALL in recent publications of cancer predisposition genes in pediatric cancer (Gröbner et al., 2018; J. Zhang et al., 2015). J. Zhang et al. (2015) also noted an association between BRCA1 germline mutations and childhood leukemia. This is reasonable because in BRCA1/2 are included in the DNA repair pathways, and BRCA1/2 mutations result in Fanconi anemia (Howlett et al., 2002; Mamrak, Shimamura, & Howlett, 2017).

Constitutional chromosomal aberrations have been reported to have a susceptibility to childhood ALL. For example, ring chromosome 21 (r(21)c) predisposes to iAMP21 subtype of ALL (Yilong Li et al., 2014). In addition, Robertsonian chromosome rob(15;21)(q10;q10) has a high genetic predisposition to iAMP21 subtype (Harrison & Schwab, 2016).

Genome-wide association studies

In genome-wide association studies (GWAS), common variants in several genes, e.g., IKZF1, ARID5B, CEBPE, CDKN2A, PI4K2A, GATA3, and TP63, have been
associated with an increased risk of developing ALL (Ellinghaus et al., 2012; Migliorini et al., 2013; Papaemmanuil et al., 2009; Sherborne et al., 2010). Although the variants are common, the genes are frequently involved somatically in childhood ALL (Moriyama, Relling, & Yang, 2015), and germline variation have been reported, at least with IKZF1 (Churchman et al., 2018). A recent study reported two novel susceptibility loci: at 8q24.21, a long non-coding RNA, associated with BCP-ALL; and at 2q22.3, an intergenic region, associated with ETV6-RUNX1 subtype (Vijayakrishnan et al., 2018).

**Congenital syndromes and disorders**

Rare congenital syndromes and disorders have a susceptibility to childhood cancer, including ALL, however, these account for only a small proportion of childhood ALL cases (Ripperger et al., 2017). The most dominantly predisposing syndrome is Down syndrome (trisomy of chromosome 21), which presents an approximately 20-fold increased risk for ALL (Izraeli, Vora, Zwaan, & Whitlock, 2014). Another congenital developmental delay syndrome, Noonan syndrome, which is a type of RASopathy caused by variants involved in the Ras-extracellular signal-regulated kinase (RAS-ERK) pathway (Schubbert, Shannon, & Bollag, 2007), predisposes to ALL (Kratz et al., 2015).

Other congenital developmental delay syndromes moderately associated with an increased prevalence of ALL include Rubinstein-Taybi syndrome caused by the CREBBP gene (Iyer, Oezdag, & Caldas, 2004; Jonas, Heilbron, & Ablin, 1978), Sotos syndrome caused by NSD1 (Tatton-Brown & Rahman, 2013), Weaver syndrome caused by EZH2 (Basel-Vanagaite, 2010), and Börjeson-Forssman-Lehmann syndrome caused by PHF6 (Chao et al., 2010).

In addition to developmental delay syndromes, inherited DNA repair syndromes, e.g., ataxia telangiectasia, Fanconi anemia, Bloom syndrome, Nijmegen breakage syndrome, and constitutional mismatch repair deficiency (CMMRD), predispose to childhood ALL (Kratz, Stanulla, & Cavé, 2016). Fanconi anemia is a subtype of IBMFS, in which the production of blood cells is dysfunctional, and it can lead to acute leukemia (Mamrak et al., 2017). Other bone marrow failure syndromes also predispose to acute leukemia, predominantly acute myeloid leukemia (AML), but in a few cases to ALL (Bluteau et al., 2017).
2.2.3 Genetic pathway of leukemogenesis in ALL

Recurrent genetic aberrations, such as *ETV6-RUNX1* or hyperdiploidy, which are fundamental in ALL, have been detected early in utero and in the neonatal blood spots of children who later developed ALL (Greaves & Wiemels, 2003). These genetic lesions are considered the initiating event in leukemogenesis (Bhojwani et al., 2015; Mullighan, 2012). The importance of the initiating genetic aberrations is also evidenced by the low genomic instability and few additional somatic events detected in childhood ALL compared to that in other pediatric cancers and in adult malignancies (Gröbner et al., 2018; Xiaotu Ma et al., 2018).

However, the chromosomal aberrations, such as *ETV6-RUNX1* and hyperdiploidy, are in and of themselves inadequate for rendering overt disease and explaining the etiology and genetic heterogeneity of ALL (Bhojwani et al., 2015; Mullighan, 2012). Despite the genetic aberrations detected in utero, there is a highly variable latency period that lasts many years before overt disease is detected (Greaves & Wiemels, 2003). Hence, secondary somatic mutations, which are usually focal mutations in single genes, are presumably required for leukemogenesis and the postnatal promotion of ALL (Inaba et al., 2013). A possible exception is *KMT2A*-rearranged leukemia, which may not have additional genetic alterations and has a short latency, given that most patients are diagnosed before they are 1 year old (Mullighan, 2012).

The underlying host genome variation or germline mutations affect ALL progression, possibly enabling the initiating genomic events to occur prenatally in the lymphoid precursor cells (Bhojwani et al., 2015; Hunger & Mullighan, 2015). Subsequent additional somatic mutations in the key lymphoid genes are essential for leukemogenesis in most ALL subtypes (Mullighan, 2012). This proposed mechanism of leukemogenesis is depicted in Figure 2. The theory is further supported by the fact that many inherited genetic variants predispose to certain genetic subtypes of ALL. For example *TP63* variants and 2q22.3 locus associated with *ETV6-RUNX1* (Ellinghaus et al., 2012; Vijayakrishnan et al., 2018), germline *ETV6* variants have been associated with the hyperdiploid subtype (Duployez et al., 2018; Moriyama et al., 2015), and *TP53* mutations have been detected in hypodiploid ALL (Holmfeldt et al., 2013).
Fig. 2. Pathogenesis of childhood ALL. Adapted from Bhojwani et al. (2015) and Hunger & Mullighan (2015).
2.3 Bone marrow failure syndrome

Bone marrow failure syndrome (BMFS) is characterized by the inadequacy of the bone marrow to produce blood cells including, red blood cells (RBCs), white blood cells (WBCs), and platelets, that results in cytopenia of one or two lineages or pancytopenia of all three cell lines, i.e., aplastic anemia (Adam et al., 2017). BMFS can be acquired or it can be inherited i.e. IBMFS, which is a heterogeneous disorder often characterized by additional congenital abnormalities (Dokal & Vulliamy, 2010). The non-hematological abnormalities can include developmental delay, craniofacial malformations, and systemic organ dysfunctions (Savage & Dufour, 2017).

The phenotypes may vary considerably among patients, and the clinical manifestations, which can range from mild to severe, may appear in early childhood or in adulthood (Dokal & Vulliamy, 2010). In addition to the associated congenital abnormalities, IBMFS may often predispose to solid tumors and leukemias, most predominantly premalignant myelodysplastic syndrome (MDS) and AML, however, ALL can be seen also in IBMFS (Shimamura & Alter, 2010). Bone marrow failure is usually treated with HSCT to render adequate blood cells, however, this may not prevent later malignancies or internal organ failures (Dalle & Latour, 2016).

2.3.1 Inherited bone marrow failure syndromes

The IBMFS are a large group of disorders. The classical IBMFS are Fanconi anemia, Dyskeratosis Congenita, Diamond-Blackfan anemia, and Shwachman-Diamond Syndrome (Dokal & Vulliamy, 2010). In addition, constitutional thrombocytopenia, or platelet disorder, is one type of IBMFS (Bluteau et al., 2017) that predisposes to acute leukemia and other hematological malignancies, including thrombocytopenia 5 (THC5), which is caused by the germline ETV6 variants associated with pediatric ALL (Noetzli et al., 2015; Topka et al., 2015; M. Y. Zhang et al., 2015) or the germline RUNX1 mutations that cause familial platelet disorder, which is associated with myeloid malignancy (Schlegelberger & Heller, 2017).

Fanconi anemia, which may predispose to ALL (Wagner et al., 2004), is the most common IBMFS and the first to be suspected in children with bone marrow failure after the exclusion of acute leukemia and MDS (Shimamura & Alter, 2010). Currently, more than 20 genes are known to cause Fanconi anemia, which is
characterized by pancytopenia, congenital abnormalities, slow growth rate with skeletal anomalies, and increased risk for hematological and solid tumors, including AML, ALL, head and neck tumors, neuroblastoma, and hepatoblastoma (Mamrak et al., 2017; Shimamura & Alter, 2010).

Dyskeratosis Congenita causes aplastic anemia associated with clinical symptoms, such as dystrophic nails, aberrant skin pigmentation, and mucosal leukoplakia, in addition to internal organ problems in some cases (Shimamura & Alter, 2010). The severity of the syndrome varies according to the underlying genetic cause. For example, the severe form observed in early childhood is caused by mutations in the X chromosome-linked DKC1 gene, and thus affects boys only (Townsley, Dumitriu, & Young, 2014). As with other IBMFS, the most common malignancy associated with Dyskeratosis Congenita is AML although solid tumors, including head and neck squamous cell carcinoma (HNSCC) and colorectal cancer, are also related to the syndrome (Adam et al., 2017).

Diamond-Blackfan anemia results in cytopenia and an aberrant formation of the erythroid cells in early childhood, and it may cause growth restriction, craniofacial and skeletal anomalies, and internal organ dysfunction (Adam et al., 2017). It has susceptibility to MDS, acute leukemia, and osteosarcoma (Wegman-Ostrosky & Savage, 2017). Shwachman-Diamond Syndrome is associated with bone marrow failure with aplastic anemia and an insufficiency of the exocrine pancreas, and it causes predisposition to hematological malignancies (Dokal & Vulliamy, 2010). Short stature and skeletal and internal organ anomalies are rarely detected in Shwachman-Diamond Syndrome patients (Shimamura & Alter, 2010).

Recently, IBMFS cases with mutations in the ERCC6L2 gene (Table 1), which has a role in DNA damage response, were published (Bluteau et al., 2017; Shabanova et al., 2018; Tummala et al., 2014; S. Zhang et al., 2016). Some of the patients had developmental delay and craniofacial anomalies, including microcephaly, however, those patients were from consanguineous families (Shabanova et al., 2018; Tummala et al., 2014; S. Zhang et al., 2016).
Table 1. Clinical details of patients with IBMFS caused by ERCC6L2 mutation

<table>
<thead>
<tr>
<th>Reference</th>
<th>ERCC6L2 variant</th>
<th>Age at onset</th>
<th>Sex</th>
<th>Hb</th>
<th>WBCs</th>
<th>Platelets</th>
<th>Clinical features</th>
<th>First-cousin parents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tummala et al. (2014)</td>
<td>p.R655*</td>
<td>12</td>
<td>M</td>
<td>99 g/l</td>
<td>2.4 x 10^9/l</td>
<td>9 x 10^9/l</td>
<td>Developmental delay, microcephaly, facial dysmorphia</td>
<td>yes</td>
</tr>
<tr>
<td>Tummala et al. (2014)</td>
<td>p.T413Cfs*2</td>
<td>19</td>
<td>F</td>
<td>93 g/l</td>
<td>5.3 x 10^9/l</td>
<td>33 x 10^9/l</td>
<td>Developmental delay, microcephaly, floppy in infancy</td>
<td>yes</td>
</tr>
<tr>
<td>Zhang et al. (2016)</td>
<td>p.T413Cfs*2</td>
<td>13</td>
<td>M</td>
<td>90 g/l</td>
<td>1.47 x 10^9/l</td>
<td>4 x 10^9/l</td>
<td>Developmental delay, microcephaly, vascular anomalies in the brain, renal anomalies</td>
<td>yes</td>
</tr>
<tr>
<td>Shabanova et al.</td>
<td>p.R563*</td>
<td>7</td>
<td>F</td>
<td>133 g/l</td>
<td>3 x 10^9/l</td>
<td>58 x 10^9/l</td>
<td>Developmental delay, microcephaly, ataxia, nystagmus, volume loss of the brain</td>
<td>yes</td>
</tr>
<tr>
<td>Bluteau et al. (2017)</td>
<td>p.E729fs*14</td>
<td>7</td>
<td>M</td>
<td>114 g/l</td>
<td>&lt;1.5 x 10^9/l</td>
<td>64 x 10^9/l</td>
<td>—</td>
<td>no</td>
</tr>
<tr>
<td>Bluteau et al. (2017)</td>
<td>p.E729fs*14</td>
<td>13</td>
<td>F</td>
<td>&lt;120 g/l</td>
<td>&lt;1.5 x 10^9/l</td>
<td>&lt;150 x 10^9/l</td>
<td>—</td>
<td>no</td>
</tr>
<tr>
<td>Bluteau et al. (2017)</td>
<td>p.Q502X; p.R1266X</td>
<td>22</td>
<td>F</td>
<td>119 g/l</td>
<td>0.4 x 10^9/l</td>
<td>107 x 10^9/l</td>
<td>Monosomy of chromosome 7 in bone marrow</td>
<td>no</td>
</tr>
<tr>
<td>Bluteau et al. (2017)</td>
<td>p.Q502X; p.R1266X</td>
<td>18</td>
<td>F</td>
<td>129 g/l</td>
<td>1.6 x 10^9/l</td>
<td>101 x 10^9/l</td>
<td>—</td>
<td>no</td>
</tr>
<tr>
<td>Bluteau et al. (2017)</td>
<td>p.R655*</td>
<td>2</td>
<td>M</td>
<td>109 g/l</td>
<td>1.0 x 10^9/l</td>
<td>48 x 10^9/l</td>
<td>Facial dysmorphia</td>
<td>no</td>
</tr>
<tr>
<td>Bluteau et al. (2017)</td>
<td>p.D283N</td>
<td>22</td>
<td>F</td>
<td>107 g/l</td>
<td>0.1 x 10^9/l</td>
<td>38 x 10^9/l</td>
<td>Died at 43 years old after AML</td>
<td>no</td>
</tr>
</tbody>
</table>

1Age in years, 2The number of codons before stop codon, 3absolute neutrophil count, asterisk (*), stop codon; fs, frameshift; g, grams; Hb, hemoglobin; l, liter; WBCs, white blood cells, X, unknown amino acid
2.3.2 Genetic mechanisms of IBMFS

The IBMFS are caused by germline genetic mutations in the key cellular pathways required for development and cellular homeostasis, including DNA damage response, ribosome biogenesis, and telomere function and maintenance (Table 2.) (Adam et al., 2017). The same cellular pathways and genes are involved in carcinogenesis; thus, most of the IBMFS predispose to hematological and solid tumors (Wegman-Ostrosky & Savage, 2017).

The genes involved in Fanconi anemia are part of the FA/BRCA-pathway, which has a role in DNA repair against the DNA interstrand crosslinks (ICLs) obstructing DNA replication and RNA transcription (Wegman-Ostrosky & Savage, 2017). Most of the genes are part of the protein complex FA core complex, which is recruited to the site of the DNA damage caused by ICLs (Mamrak et al., 2017). The rest of the genes function together with the FA complex or are involved in homologous recombination, an important mechanism of the FA/BRCA pathway (Wegman-Ostrosky & Savage, 2017).

Ercc6L2, the mutations of which cause IBMFS, is also involved in DNA damage response through the nucleotide excision repair (NER) pathway (Tummala et al., 2014). Tummala et al. (2014) demonstrated impaired response against the DNA-damaging agent mitomycin C, which also causes ICLs (Mamrak et al., 2017). Notably, the ICL repair and the NER pathway, which prevents single-strand DNA damage, have been shown to share common functions (Mouw & D’Andrea, 2014).

Diamond-Blackfan anemia and Shwachman-Diamond syndrome are both caused by impaired ribosome biogenesis (Shimamura & Alter, 2010). Ribosomes, which are two-piece complexes with ribosomal RNA (rRNA) and multiple proteins, are involved in protein translation from the messenger RNA (mRNA) (Ruggero & Shimamura, 2014). The most important gene causing Shwachman-Diamond syndrome is the SBDS gene, which has an important role in the maturation of the large 60S ribosomal subunit, and multiple genes, of which the most commonly mutated is RPS19 of the small ribosomal subunit, are known to cause Diamond-Blackfan anemia (Dokal & Vulliamy, 2010). Currently, at least 15 genes are related to the syndrome, some of them encoding proteins of the large ribosomal subunit (Savage & Dufour, 2017).

Dyskeratosis Congenita is a syndrome caused by faulted telomere function and maintenance resulting from mutations in the following genes: DKC1, TERC, TERT, NOP10, NHP2, and TINF2 (Dokal & Vulliamy, 2010). Telomeres are regions at the
end of each chromosome that become shorter with every cell division and, together with the telomerase repair complex, which in turn mediate the elongation of telomeres, enable DNA replication (Towsley et al., 2014). As indicated by Towsley et al. (2014), the genes causing Dyskeratosis Congenita impair the function of telomerase, and this results in a reduced number of hematopoietic stem cells. Some of the genes are involved in ribosome biogenesis, and notably, the telomerase repair complex is a ribonucleoprotein enzyme complex (Ruggero & Shimamura, 2014; Townsley et al., 2014).

2.4 Mitochondrial genome and cancer

2.4.1 Mitochondria and their function

Mitochondria are energy-producing cell organelles, each of which has a double-layered lipid membrane with outer (OM) and inner membrane (IM); space inside the IM, i.e., the matrix; and an intermembrane space (IMS) between the OM and the IM (Fig 3.) (Ylikallio & Suomalainen, 2012). Most of the metabolic pathways of the cell, including tricarboxylic acid cycle (TCA) in the matrix and oxidative phosphorylation (OXPHOS) on the IM and its surroundings, occur within the mitochondria, and, with abundant molecular oxygen, TCA and OXPHOS function together to produce energy in a form of adenosine triphosphate (ATP) (Gorman et al., 2016).

The TCA cycle breaks down metabolic cell nutrients, and it reduces NAD+ to NADH and fumarate to succinate, that are later oxidized by OXPHOS complex I and II, respectively, and the electrons are donated to OXPHOS (Suomalainen & Battersby, 2018). Ubiquinone (CoQ), cytochrome c (Cyt C), and complex III and IV continue the electron transport and oxidoreduction reactions to form the proton (H+) gradient on the IM (Fernández-Vizarra, Tiranti, & Zeviani, 2009). This electrical gradient is finally utilized by complex V to phosphorylate adenosine diphosphate (ADP) to ATP (Fernández-Vizarra et al., 2009). Molecular oxygen (O2), which is required for the OXPHOS function, is eventually reduced to water molecule (H2O) (Gorman et al., 2016).
Table 2. The most common IBMFS and their biological mechanism and clinical features

<table>
<thead>
<tr>
<th>Biological mechanism of the IBMFS</th>
<th>Gene</th>
<th>Predominant cytopenia</th>
<th>Frequent congenital abnormalities</th>
<th>Frequent malignancies</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA damage response</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ERCC6L2</td>
<td>ERCC6L2</td>
<td>thrombocytopenia progressing to pancytopenia</td>
<td>developmental delay, craniofacial anomalies(^1)</td>
<td>AML(^2)</td>
</tr>
<tr>
<td>Fanconi anemia</td>
<td>FANCA/B/C/D2/E/F/G/H/J/L/M/N/O/P/Q/R/T/U/V, BRCA1/2</td>
<td>pancytopenia</td>
<td>short stature, dyspigmentation, skeletal and internal organ anomalies</td>
<td>ALL, AML, HNSCC, neuroblastoma, liver tumors</td>
</tr>
<tr>
<td>Telomere maintenance</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dyskeratosis Congenita</td>
<td>DKC1, TERC, TERT, NOP10, NHP2, TINF2</td>
<td>pancytopenia</td>
<td>dermatologic triad: nail dystrophy, dyspigmentation, and mucosal leukoplakia; internal organ problems</td>
<td>AML, HNSCC, colorectal cancer,</td>
</tr>
<tr>
<td>Diamond-Blackfan anemia</td>
<td>RPS19/24/17/7/10/26/27/29, RPLS/5/11/26/15/27/35</td>
<td>anemia</td>
<td>craniofacial, skeletal, cardiac, and genitourinary anomalies</td>
<td>hematological malignancies, osteosarcoma</td>
</tr>
<tr>
<td>Shwachman-Diamond Syndrome</td>
<td>SBDS</td>
<td>pancytopenia</td>
<td>exocrine pancreatic insufficiency,</td>
<td>hematological malignancies</td>
</tr>
<tr>
<td>Hematopoiesis</td>
<td>RUNX1, THC5</td>
<td>RUNX1, ETV6</td>
<td>thrombocytopenia</td>
<td>—, —</td>
</tr>
</tbody>
</table>

\(^1\)Patients are from consanguineous families. \(^2\)One patient reported by Bluteau et al. (2017).
2.4.2 The mitochondrial genome

Mitochondria have a matrilineally inherited circular genome known as mtDNA, a double-stranded and 16,569 base pair (bp) long DNA molecule (Fig. 3), and each mitochondrion has multiple copies of mtDNA (Ylikallio & Suomalainen, 2012). Two mtDNA sequences may vary from each other with few nucleotides, and when different mtDNA genomes form a mixture, this is called heteroplasmy, or homoplasmy when all of the mtDNA have the identical nucleotide sequence (Gorman et al., 2016).

The mitochondrial genome encodes 13 protein-coding genes for the subunits of five OXPHOS enzyme complexes, for which the remaining subunits are encoded by nuclear DNA (nDNA) (Fernández-Vizarra et al., 2009). The mtDNA has also a non-coding regulatory displacement loop, i.e., D-loop, it encodes two ribosomal RNAs and several transfer RNAs (tRNA) for different amino acids, and has a total of 37 genes (Gorman et al., 2016). The mtDNA is categorized into haplogroups or phylogenetic cluster groups, which are based on specific nucleotide variations of the mitochondrial genome, and they are important for interpreting mtDNA mutations (Weissensteiner, Pacher et al., 2016).

2.4.3 The Warburg hypothesis and metabolism in cancer cells

Hanahan & Weinberg (2011) published the hallmarks of cancer, and one of the hallmarks is the altered energy metabolism of the cancer cell. Under hypoxic conditions, glycolysis takes over energy production in normal cells, albeit with a lower ATP yield. In contrast, malignant cells use glycolysis for energy production even in the presence of oxygen (Hanahan & Weinberg, 2011). The phenomenon, known as the Warburg effect or aerobic glycolysis, was published in the 1950s by Otto Warburg, who hypothesized that dysfunctional mitochondria were the cause (Warburg, 1956).

Pyruvate, provided by glycolysis in the cytosol, is imported to the mitochondria for energy production via the TCA and sequential OXPHOS in normal cells (Ylikallio & Suomalainen, 2012). In fast-proliferating cancer cells, pyruvate is uncoupled from the mitochondria possibly to provide substrates for macromolecule synthesis; thus, glycolysis could provide advantage for cancer cells despite lower energy production (Vander Heiden, Cantley, & Thompson, 2009).
Fig. 3. Mitochondria, mtDNA, and simplified OXPHOS complexes. The protein-coding genes are color-coded and named without the MT prefix. The red and black mtDNA within the mitochondrion represent heteroplasy. On the mtDNA, the three-letter abbreviations are amino acids affiliated with tRNA genes. Adapted from Ylikallio & Suomalainen (2012) and Gorman et al. (2016).
2.4.4 MtDNA and cancer

Somatic mtDNA mutations have been detected in various cancer types, mostly in solid tumors, and the reported variants have been mainly non-synonymous in the protein-coding genes of OXPHOS enzyme complexes (Dasgupta & Hertweck, 2017). Low OXPHOS activity has been reported in aggressive cancer types (Simonnet et al., 2002; Xiao et al., 2013). In addition, the enhancement of OXPHOS activity can impair the growth and viability of cancer cells (Dickerson, Jauregui, & Teng, 2017; Schulz et al., 2006). Thus, mtDNA mutations on OXPHOS protein-coding genes could contribute to tumor progression (Tan, Baty, & Berridge, 2014).

Variation of the mtDNA in the regulatory D-loop area have been reported in childhood ALL (Kwok et al., 2011). It has been proposed that carcinogenesis induce mutations on the D-loop, a mutational hotspot of mtDNA; thus, resulting in impaired mtDNA maintenance and subsequent mutations on the protein-coding genes (Cook & Higuchi, 2012). MtDNA mutations on the D-loop have also been detected in adult AML (Silkjaer et al., 2013) and childhood AML (Sharawat, Bakhshi, Vishnubhatla, & Bakhshi, 2010). In addition, variants on the subunits of OXPHOS proteins encoded by both mtDNA or nDNA have been detected in AML (Damm et al., 2012; Silkjaer, Nyvold, Juhl - Christensen, Hokland, & Nørgaard, 2013). The variants were implicated to correlate with prognosis, however, controversially to either with favorable (Damm et al., 2012) or with inferior survival (Silkjaer et al., 2013).
3 Aims of the study

Pediatric ALL has reached an excellent overall survival rate of nearly 90% through better care, extensive research, and international collaborations. Despite improved risk stratification, relapse is not predictable, and patients who relapse have poor prognoses. The genetic etiology of ALL is still largely unknown, and the genetic factors behind relapse or treatment failure are poorly understood.

Familial ALL, where more than one related individual develops ALL, are rare. Thus, these patients provide possibilities for studying leukemia etiology and cancer predisposition. The genetic etiology of IBMFS, which may predispose to ALL, is quickly broadening with high-throughput sequencing. Understanding the genetic cause is crucial for patients with familial ALL and those with IBMFS to receive appropriate genetic counselling.

The mitochondrial genome is an interesting target for studying in leukemia. The balance between OXPHOS and glycolysis is important for cancer cell viability, and aggressive cancer growth is associated with the metabolic shift toward glycolysis. Thus, leukemic blast cells may include mtDNA variants affecting the protein-encoding genes of the OXPHOS enzyme complexes.

The specific aims of the research were:

1. To discover the underlying germline predisposition to childhood ALL in the familial cases (I, II).
2. To determine the probable genetic cause of BMFS in patients with bone marrow failure without signs of classical IBMFS (III).
3. To determine the existence of mtDNA sequence variation in childhood ALL and the ways in which such variation may correlate with prognosis (IV).
4 Patients, materials, and methods

4.1 Patients

4.1.1 Patients with familial ALL (I, II)

Two families with familial ALL were studied for potential genetic predisposition to leukemia. It is rare that related individuals, either first- or second-degree relatives from the same family are diagnosed with ALL; therefore, these families were studied. Both families included two patients diagnosed with childhood ALL. The patients were treated at the Oulu University Hospital in Oulu, Finland.

The first family included two siblings, a boy and a girl, with childhood ALL and unknown intellectual disability syndrome, including overgrowth and dysmorphic facial features. The mother of the siblings displayed the characteristic features, which were similar to those exhibited by the children, and learning difficulties, but no hematological malignancy. The family also included a healthy father and three healthy siblings. The mother had lost one child to sudden infant death syndrome and had had one miscarriage. DNA was extracted from the bone marrow and the peripheral blood of the affected siblings and from the peripheral blood of the mother, the father, and one healthy sibling in accordance with the standard methods. No other relatives were available for genetic analysis.

In the other family, a girl and her maternal aunt had been both diagnosed with childhood ALL. Routine cytogenetic analysis, performed according to the NOPHO treatment protocol used at the time of diagnosis (Schmiegelow et al., 2010; Toft et al., 2018), revealed constitutional chromosome translocation in both patients. This led to further genetic testing of the family. In total, 11 individuals were enrolled for karyotyping, which was performed at the Oulu University Hospital in Oulu, Finland. The Giemsa banding technique for karyotyping was used in accordance with the standard methods. DNA was extracted from the bone marrow and the peripheral blood of both affected females and from the peripheral blood of the additional family members included in the study by the standard procedure.

4.1.2 Patients with bone marrow failure (III)

Two index patients, a male and a female from unrelated families, were treated for a previously uncharacterized bone marrow failure at the Oulu University Hospital.
Both patients presented with the first symptoms at the age of 8 years. The DNA for genetic analysis from the patients, the parents, and one sibling of each index patient was extracted from the blood in accordance with standard hospital methods.

### 4.1.3 Patient cohort in the mtDNA sequence variation study (IV)

The patient cohort included 36 patients with childhood ALL, of whom 17 were deceased at the time of the mtDNA analysis. The patients were treated at the Oulu University Hospital from 1992 to 2011 in accordance with the NOPHO ALL treatment protocols (Schmiegelow et al., 2010; Toft et al., 2018).

The DNA samples were collected at diagnosis, during and at the end of treatment or at relapse, and before potential stem cell transplantation. The relapse samples were obtained between days 28 and 1,018 after diagnosis, with day 700 being the median. The patients are referred to as A (surviving) and B (deceased). The DNA was extracted from the blood and the bone marrow by the standard procedure. The exception was four samples of DNA that were extracted using a QiaAmp DNA Micro Kit (Qiagen, Hilden, Germany).

### 4.1.4 Controls (IV)

Study IV had a control population of 480 Finnish individuals whose DNA samples had been obtained from the Finnish Red Cross offices in the capitals of the provinces of northern Ostrobothnia, central Ostrobothnia, Kainuu, and northern Savo, as was previously described (Finnilä, Lehtonen, & Majamaa, 2001).

### 4.2 Ethical considerations (I–IV)

Studies I and II were approved by the regional ethical committees of the Northern Ostrobothnia Hospital District, Karolinska Institutet, and Karolinska University Hospital. Studies III and IV were approved by the regional ethics committee of the Northern Ostrobothnia Hospital District. Written informed consent for participation in the study and for the publication of photographs was obtained from each individual or the legal guardian as appropriate.

All studies were conducted in accordance with the 1975 Declaration of Helsinki, as revised in 2000.
4.3 Molecular methods

4.3.1 Familial ALL (I, II)

Single nucleotide polymorphism (SNP) array (I, II)

A high-resolution SNP genotyping array HumanOmni2.5+Exome-8v1 or HumanOmni2.5 (Illumina, San Diego) was used in the study I. DNA from leukemic and remission bone marrow of both affected siblings, together with germline DNA of the mother, the father, and one healthy sibling, were analyzed for germline copy number variations (CNVs) and somatic copy number alterations (CNAs). By using both diagnostic and remission DNA for the affected, one can compare the germline variants (remission) to leukemic i.e. somatic variants (diagnostic). In the study II DNA was extracted from bone marrow at leukemia diagnosis for both females to determine the karyotype of leukemic cells using Infinium Omni2.5Exome BeadChip (Illumina, San Diego, CA, USA) to analyze CNAs.

CNVs are constitutional variation on the host genome and CNAs refer to somatic or acquired variation. In the study I, CNVs were called with Tumor Aberration Prediction Suite (TAPS) (Rasmussen et al., 2011) and CNAs with both TAPS and allele-specific copy number analysis of tumors (ASCAT) (Van Loo et al., 2010) using R software. In the study II, CNAs were called using circular binary segmentation (CBS) algorithm (Venkatraman & Olshen, 2007) and TAPS (Rasmussen et al., 2011) as implemented in R software.

A panel of internal controls were used to normalize probe intensities from our analysis to normal diploid samples, to generate log transformed ratios (log2 ratios, LRR), which were centered at zero for diploid samples. Variants less than 20 kilobase pairs (kb) were filtered out in the study I and for the study II, and the criteria was set at ten consecutive supporting probes. All putative variants were visualized and manually reviewed using Integrative genomics viewer (IGV) (Robinson et al., 2011).

All coordinates within the thesis refer to genome build GRCh37/hg19.

Whole Genome Sequencing and bioinformatics analysis (II)

In the study II whole genome sequencing (WGS) was performed on DNA from peripheral blood at remission for both affected females with leukemia and on
germline DNA from peripheral blood of the unaffected carrier (mother of the young female). With WGS the nucleotide sequence of the entire human genome can be analyzed, and with the method we could define the exact coordinates of the constitutional translocation detected by karyotyping. WGS was also performed on DNA from bone marrow at leukemia diagnosis for both affected females to investigate somatic events in the leukemic clones.

Sequencing was performed on Illumina HiSeq X platform (Illumina Inc, San Diego, CA, USA). Libraries were prepared from the genomic DNA using the Illumina TruSeq PCR-free kit with a mean insert size of >350 bp, which resulted on average of 500 million sequences per sample (range between 386M and 740M) with a mean coverage of 40x. An in-house pipeline developed by Science for Life Laboratory (Stockholm, Sweden) was used to map the reads. Data were aligned to the reference genome using Burrows-Wheeler Aligner (BWA; v0.7.12) (Li & Durbin, 2009) and the data was refined with Genome Analysis Toolkit (GATK; Broad Institute, Cambridge, MA, USA; v3.3-0-geee94ec) (McKenna et al., 2010). The quality control was checked with Qualimap (v2.0) (Okonechnikov, Conesa, & García-Alcalde, 2015).

Somatic variants were identified using somatic SNP and indel caller MuTect2 (Cibulskis et al., 2013). Variant files were manipulated using CatVariants and SelectVariants tools in GATK. The somatic variants were functionally annotated using Variant Effect Predictor (VEP; version 89) (McLaren et al., 2010), loaded into a database using GEMINI (v0.20.0) (Paila, Chapman, Kirchner, & Quinlan, 2013), and visualized in the IGV (Robinson et al., 2011).

Structural variants were detected using the FindSV pipeline (https://github.com/J35P312/FindSV) which utilize CNVnator v0.3.2 and TIDDDT (Abyzov, Urban, Snyder, & Gerstein, 2011; Eisfeldt, Vezzi, Olason, Nilsson, & Lindstrand, 2017). The Variant Call Format (VCF) for structural variants were annotated by VEP and sorted based on a local structural variant frequency database consisting of 400 patient samples. The exact coordinates of the structural variants were identified by aligning the split reads to the reference genome using BLAT (BLAST-like alignment tool) (Altschul et al., 1997; Kent, 2002).

Sanger sequencing (I, II)

In both studies Sanger sequencing was used to confirm the results from SNP array (I) or WGS (II). In the study I, the coding regions of IKZF1 (NM_006060; the 8-exon isoform Ik1) were amplified from leukemic bone marrow DNA of the affected
female sibling to detect somatic mutations. In the study II, primers were designed based on WGS data of the translocation breakpoints and in silico predictions of fusion gene transcripts. In addition, coding regions of ETV6 from diagnostic and remission bone marrow samples of both affected females and germline DNA from unaffected carrier (mother of the affected young female) were sequenced to determine the presence of somatic variants.

In both studies, sequencing was performed in an ABI3730xl DNA Analyzer (Applied Biosystems) using BigDye Terminator v3.1 (Applied Biosystems).

Expression analysis (I, II)

In studies I and II, quantitative real-time PCR (qPCR) was performed to assess the expression levels of the genes in interest using Power SYBRGreen (Life Technologies) on the QuantStudio 7 Flex Real-time PCR System (Applied Biosystems). All samples were run in triplicate, averaged and normalized against human ribosomal 18S rRNA, which was used as an internal control for each sample.

In the study I, the expression levels of IKZF1 for the mother, the affected female and one healthy sibling, and, also, for four controls was determined from RNA. First-strand cDNA was synthesized with 200 ng of total RNA using random primers and a SuperScript VILO cDNA synthesis kit (Invitrogen).

In the study II, the expression levels of wild-type ETV6 and RTN1 mRNA, and fusion transcripts ETV6-RTN1 and RTN1-ETV6 were studied by qPCR. Expression levels were analyzed in germline samples for the affected females with ALL and one unaffected female carrier, and in leukemia samples for affected females.

For germline expression, total RNA from peripheral blood from six unrelated healthy normal controls, three sex- and age-matched controls to each of the analyzed family member, were used. For the wild-type expression of ETV6 and RTN1 in leukemia, total RNA from diagnostic bone marrow samples from the two affected individuals and eight unrelated cases of sporadic BCP-ALL with different somatic aberrations; ETV6 deletion with or without translocation ETV6-RUNX1, and cases wild-type for ETV6.

Flow Cytometry (I)

In the study I, a multi-parametric flow cytometry analysis was performed on the affected female sibling and the mother to broadly phenotype the immune system and functionally assess the natural killer (NK) and cytotoxic T cell activity. The
patient and control samples were shipped from Oulu, Finland (at room temperature) for analysis to Assistant Professor Yena Bryceson’s laboratory at Center for Hematology and Regenerative Medicine (HERM), Karolinska Institutet, Stockholm, Sweden.

Peripheral blood mononuclear cells (PBMCs) were obtained from heparinized blood samples via Ficoll gradient centrifugation. Intracellular levels of perforin, granzyme A and granzyme B were quantified in NK cells. The exocytosis function of both NK (CD3−CD56dim) and cytotoxic T lymphocytes (CD3+CD8+CD57+) were quantified as previously published (Chiang et al., 2013). Briefly, PBMCs were incubated with either K562 cells, P815 cells alone, P815 cells with anti-CD16 antibody (clone 3G8) or P815 cells with anti-CD3 antibody (clone S4.1). Cytotoxic lymphocyte responses were quantified by surface staining of CD107a. The cytotoxicity of NK cells was evaluated with a standard 4-hour 51Cr release assay at various effector to target ratios, as previously described (Chiang et al., 2013).

The antibodies used were anti-CD3 (clone S4.1), anti-CD8 (RPA-T8), anti-CD56 (NCAM16.2), anti-CD57 (HCD57), anti-CD107a (H4A3), anti-perforin (dG9), anti-granzyme A (CB9), and anti-granzyme B (Gb11).

4.3.2 Bone marrow failure study (III)

For the male index patient and his family, whole exome sequencing was performed by Genome Diagnostics Nijmegen (Nijmegen, Netherlands), and for the female index patient and her family, it was performed by Centogene (Rostock, Germany). Exome sequencing is a method to screen the protein coding areas of the human genome. Peripheral blood cell counts were also analyzed for the parents of both families.

4.3.3 MtDNA sequence variation study (IV)

DNA amplification

The non-coding displacement region (D-loop) of the mtDNA and the coding region with overlapping fragments were amplified via polymerase chain reaction (PCR) using Phire Hot Start II DNA Polymerase (Thermo Fisher Scientific) in accordance with the manufacturer’s instructions. For the four samples for which DNA was extracted with a different method, the DNA was amplified with the GenomePlex®
Complete Whole Genome Amplification (WGA) Kit (Sigma-Aldrich, St Louis, MO, USA), in accordance with the manufacturer’s instructions, to overcome technical issues. Finally, the PCR products of these four samples were purified using the GeneJET PCR Purification Kit (Thermo Fisher Scientific, Waltham, MA, USA).

**Conformation sensitive gel electrophoresis (CSGE)**

The amplified DNA was used for heteroduplex and homoduplex formations where the DNA samples were mixed with the same sample or control having a known mtDNA sequence. The DNA double strands in the mixture were denatured at 95 °C and allowed to anneal again at 68 °C. If the DNA strands did not match, a mismatched binding happened during annealing and resulted in conformation change on the DNA double strand known as heteroduplex. Homoduplex formation occurs between two identical DNA strands. Duplexes were visualized upon CSGE, where the migration of DNA fragments on polyacrylamide gel were different based on the conformation. This method, which was described by Finnilä et al. (2001), can reveal heteroplasmic mtDNA variants within the patient sample and show differences between the patient and the control sample.

After electrophoresis, the gel was stained with 2.5 µl SYBR® Gold Nucleic Acid Gel Stain (Thermo Fisher Scientific) diluted in 25 ml TE buffer (10 mM Tris-HCl; pH 8.0, 1 mM EDTA) for visualizing with Safe Imager™ Blue-Light Transilluminator (Invitrogen, Carlsbad, CA, USA).

**Sanger sequencing**

The D-loop segment and the fragments of coding mtDNA, visually estimated to be different on the CSGE, were subsequently analyzed with Sanger sequencing using an Applied Biosystems 3500xL Genetic Analyzer with a Big-Dye Terminator v1.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). All of the sequencing reactions were carried out by the Biocenter Oulu sequencing core facility in Oulu, Finland.

**Restriction fragment length polymorphism (RFLP)**

The novel and rare mtDNA variants discovered through previous methods were confirmed using RFLP. Enzymes, which digest specific DNA nucleotide sequences,
are used for cutting the DNA strands with or without the specific mtDNA variant. The cut DNA fragments are then visualized either on 2% agarose or 3% MetaPhor® agarose gel. FastDigest enzymes (Thermo Fisher Scientific) were used, and the reactions were performed according to the manufacturer’s instructions.

Identification of the mtDNA variants

The mtDNA sequences were aligned with the revised Cambridge reference sequence (rCRS) (NC_012920.1) (Andrews et al., 1999). The following databases were used in data analysis: A Human Mitochondrial Genome Database (MITOMAP) (Lott et al., 2013), Human Mitochondrial Database (HmtDB) (Clima et al., 2016), Human Mitochondrial Genome Database (mtDB) (Ingman & Gyllensten, 2006), and Human Mitochondrial Genome Polymorphism Database (The GiiB-JST mtSNP) (Tanaka, Takeyasu, Fuku, Guo, & Kurata, 2004; Tanaka et al., 2004).

Novel mtDNA mutations were further analyzed to determine their pathological effect by using the following databases: SNAP2 (Hecht, Bromberg, & Rost, 2015) and Polyphen-2 (Adzhubei et al., 2010). In silico analysis of the protein structure and the effect of the mtDNA variant on protein conformation were estimated using Phyre2 (Kelley, Mezulis, Yates, Wass, & Sternberg, 2015).

Haplogroup determination and phylogenetic analysis

A phylogenetic approach was used to interpret the mtDNA sequence data and to determine the mtDNA haplogroups. The phylogenetic networks were based on the median algorithm (Bandelt, Forster, Sykes, & Richards, 1995), and they were drawn on the basis of the Finnilä et al. (2001) publication. The haplogroups were defined using HaploGrep (van Oven, 2015; Weissensteiner et al., 2016) on the basis of the mtDNA polymorphisms. The frequencies of the haplogroups were compared to the controls. A two-tailed Fisher’s exact test was used to determine the statistically significant differences. The non-haplogroup specific mtDNA variants, which were potentially leukemia-associated, were identified after determining the haplogroup. The heteroplasmy of the variants was determined by CSGE, Sanger sequencing, and RFLP.
Satellite marker analysis

Satellite marker analysis was performed to rule out the presence of external DNA from the patient DNA samples, given that some DNA samples had mixed haplogroup-specific mtDNA variants that raised the suspicion. The analysis was performed using an AuthentiFiler™ PCR Amplification Kit (Life Technologies, Carlsbad, CA, USA), which is a short tandem repeat (STR) multiplex PCR assay that amplifies nine STR loci and the gender-determining marker amelogenin. Amplification was performed with a GeneAmp® PCR System 9700 Thermal Cycler (Thermo Fisher Scientific), and electrophoresis was carried out using the Applied Biosystems 3500xL Genetic Analyzer (Applied Biosystems). The data were analyzed using GeneMapper® Software version 5.0.
5 Results

5.1 Familial ALL and intellectual impairment syndrome caused by constitutional microdeletion of 7p12.1p13, including IKZF1 (I)

5.1.1 Overgrowth and intellectual impairment syndrome with susceptibility to childhood ALL (I)

The family

Two siblings, an elder brother (II:3), and his younger sister (II:6) were diagnosed with childhood ALL. From early childhood, both siblings presented with symptoms of an unknown intellectual disability syndrome. The symptoms included impaired motor and cognitive development, as well as overgrowth with a high body mass index and increased head circumference. The siblings also had characteristic dysmorphic facial features. The mother (I:2) of the children presented with similar features and learning difficulties but no hematological malignancy (Table 3). The mother also has three healthy normally developed children (Fig. 4). Her fourth child died of sudden infant death syndrome at 8 months old, and she has had one miscarriage. No other family member—not the maternal grandparents, the mother’s nine siblings, or the close relatives—has been diagnosed with dysmorphic features, intellectual impairment, or a hematological malignancy.

Fig. 4. Pedigree of the family (I).
Table 3. Characteristic features of the affected siblings and their parents (I)

<table>
<thead>
<tr>
<th>Clinical feature</th>
<th>Affected male (II:3)</th>
<th>Affected female (II:6)</th>
<th>Mother (I:2)</th>
<th>Father (I:1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>7p12.1p13 del</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>BCP-ALL</td>
<td>+</td>
<td>+</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Intellectual disability</td>
<td>+</td>
<td></td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Large head circumference (SD)</td>
<td>+3</td>
<td>+2</td>
<td>not confirmed</td>
<td>—</td>
</tr>
<tr>
<td>Tall stature (SD)</td>
<td>+2</td>
<td>+1.5</td>
<td>+0.5</td>
<td>0</td>
</tr>
<tr>
<td>Overweight</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>ISO-BMI</td>
<td>29 (6)</td>
<td>32.3 (4)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hypertelorism</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>Broad face and forehead</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>Arched flared eyebrows</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>Prominent nasal bridge</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>Broad nasal tip</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>Narrow palpebral fissures</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>Thin upper lip</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
<tr>
<td>Long philtrum</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>—</td>
</tr>
</tbody>
</table>

1The mother had learning difficulties, 2Highest ISO-BMI: the age when observed is in parentheses, SD, standard deviation; ISO-BMI, age- and sex-adjusted body mass index

Individuals with BCP-ALL

The male sibling (II:3) was diagnosed with BCP-ALL at the age of 11 years, and the female sibling (II:6) was 5 years old at diagnosis. Routine diagnostic cytogenetic workups revealed no risk-stratifying aberrations in the few metaphases available for analysis. In addition, the WBC count was low (Table 4.) The treatment for both children was initiated in accordance with the prevailing NOPHO standard risk protocol at that time. The male sibling responded poorly to induction therapy, and the protocol was changed to the HR protocol at day 14 after diagnosis. He underwent an allogeneic HSCT from a sibling donor (not analyzed in the study) in the first complete clinical remission (CR1). Unfortunately, he suffered a bone marrow relapse and died 18 months after the primary diagnosis. The female sibling is in CR1 10 years after diagnosis.
Table 4. Clinical details of the leukemia cases with germline **IKZF1** deletion (I)

<table>
<thead>
<tr>
<th>Clinical feature</th>
<th>Affected male (II:3)</th>
<th>Affected female (II:6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at dx</td>
<td>11</td>
<td>5</td>
</tr>
<tr>
<td>Immunophenotype</td>
<td>BCP-ALL</td>
<td>BCP-ALL</td>
</tr>
<tr>
<td>Highest WBC count at dx (x 10^9/l)</td>
<td>3.1</td>
<td>3</td>
</tr>
<tr>
<td>Highest blast count at dx (%) PB</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>Highest blast count at dx (%) BM</td>
<td>99</td>
<td>70</td>
</tr>
<tr>
<td>NOPHO Treatment protocol</td>
<td>ALL-2000, HR¹</td>
<td>ALL-2000, SR</td>
</tr>
<tr>
<td>Primary event</td>
<td>Relapse</td>
<td>CR1</td>
</tr>
<tr>
<td>Severe adverse effect</td>
<td>Osteonecrosis</td>
<td>—</td>
</tr>
<tr>
<td>Cause of death</td>
<td>Bone marrow relapse after HSCT</td>
<td>—</td>
</tr>
</tbody>
</table>

¹Upgraded from SR on day 14.

BCP-ALL, B-cell precursor acute lymphoblastic leukemia; BM, bone marrow; CR1, complete first remission; dx, diagnosis; HR, High risk; HSCT, hematopoietic stem cell transplantation; NA, data not available; NOPHO, Nordic Society of Paediatric Oncology; PB, peripheral blood; SR, Standard risk; WBC, white blood cell

5.1.2 **Microdeletion of 7p12.1p13, including **IKZF1**, causes developmental delay syndrome with predisposition to childhood ALL (I)**

The genotyping SNP array performed on the affected siblings, the mother, the healthy father, and one healthy sibling revealed a heterozygous germline deletion of 6.7 Mb at chromosome 7p12.1p13 (chr7:44,547,861–51,219,568) in the affected individuals (I:2, II:3 and II:6) but not in the healthy individuals.

The deletion involves 31 protein coding genes: NPC1L1, DDX56, TMED4, OGDH, ZMIZ2, PPLA, H2AFV, PURB, MYO1G, CCM2, NACD, TBRG4, RAMP3, ADCY1, IGFBP1, IGFBP3, TNS3, HUS1, PKD1L1, SUN3, C7orf57, UPP1, ABCA13, VWC2, ZPBP, C7orf72, IKZF1, FITNLI, DDC, GRB10, and COBL.

As was previously mentioned, somatic **IKZF1** deletions are frequent in BCP-ALL (Mullighan et al., 2009); therefore, it was considered the primary candidate gene for leukemia predisposition in this family. However, the syndromic features are most likely the result of the synergist effect of multiple gene deletions within the 6.7 Mb deletion.

The affected individuals had no symptoms of visceral heterotaxy, which has been described in patients with homozygous **PKD1L1** deletions (Vetrini et al.,
2016), or Aromatic L-Amino Acid Decarboxylase Deficiency, a syndrome related to deletions of DDC (Pons et al., 2004). A heterozygous deletion of the CCM2 gene has been associated with cerebral cavernous malformations (Liquori et al., 2003), a syndrome for which the affected individuals showed no signs, as evidenced by the normal findings on a computed tomography scan (II:3) or magnetic resonance imaging (II:6) of the brain.

5.1.3 Characteristics of the leukemia with germline IKZF1 deletion (I)

In addition to the detected germline 7p12.1p13 deletion, the SNP array did not reveal any structural genetic aberrations, such as high hyperdiploidy or hypodiploidy, in the leukemic cells. This is in line with the diagnostic cytogenetic analysis showing a normal karyotype in both leukemia cases.

In the male sibling’s diagnostic sample, however, the SNP array did reveal somatic focal deletions in the known leukemia-associated genes, including \textit{PA\textsc{x}5}, \textit{PTEN}, \textit{RB1}, and \textit{CREBBP}. In addition, a second deletion on the \textit{IKZF1} in the male sibling’s leukemia was discovered. This was a small biallelic deletion on the promoter and the first non-coding exon regions of \textit{IKZF1}. Notably, no deletions or gains in the known leukemia genes were detected in the female sibling’s leukemia sample. Furthermore, Sanger sequencing of the intact \textit{IKZF1} allele revealed no aberrations on the exons also covering the exon-intron boundaries.

5.1.4 Expression analysis of the IKZF1 (I)

Expression analysis by qPCR revealed lower levels of \textit{IKZF1} transcripts in the peripheral blood of the affected female sibling and the mother than in the peripheral blood of the healthy sibling and healthy controls. The analysis could not be performed for the male sibling because genetic material for qPCR was unavailable.

5.1.5 Multi-parametric flow cytometry (I)

No immunodeficiency was reported in the family, but a lower \textit{IKZF1} expression was detected in the affected female sibling and the mother. It was therefore hypothesized that subclinical defects in the B cell maturation could be detected because of the low expression. Therefore, a multi-parametric flow cytometry assay was performed to broadly phenotype the immune system. The functional assessment of NK and cytotoxic T cell activity was normal, with normal cell counts.
for both individuals. Moreover, the results revealed no apparent defect in the B cell or other lymphocyte cell counts or in B cell differentiation.

5.2 Familial ALL caused by constitutional chromosome translocation t(12;14)(p13.2;q23.1) disrupting ETV6 (II)

5.2.1 The family (II)

An index female proband was diagnosed with BCP-ALL at the age of 12 years and was treated in accordance with the NOPHO ALL-2000 SR protocol. She is in CR1 4 years past diagnosis. Previously, her maternal aunt, currently in CR1 23 years after diagnosis, had been diagnosed with childhood ALL at 8 years old. The maternal aunt was assigned to the NOPHO ALL-1992 HR protocol because myeloid lineage cell surface markers were detected on the blast cells at diagnosis. No cytopenia, aberrations of the RBCs, or other hematological malignancies were reported in the family.

In addition to the two family members with BCP-ALL, three others had been diagnosed with solid tumors: mesothelioma and prostate cancer in the confirmed translocation carriers, and an uncharacterized renal tumor in a boy whose father was a confirmed translocation carrier.

5.2.2 The constitutional chromosome translocation (II)

Routine diagnostic karyotyping revealed a constitutional balanced reciprocal translocation t(12;14)(p13.2;q23.1) in both individuals with leukemia (Fig. 5). The translocation was first detected in the maternal aunt and suspected to be constitutional because it was present in the leukemia blasts and the normal patient cells. No other somatic structural chromosome aberrations were detected in either patient at diagnosis. Karyotyping of nine other family members revealed seven individuals as carriers and two as wild type for the translocation.

The specific breakpoints of the t(12;14)(p13.2;q23.1) were characterized by WGS. This denoted them on intron 1 of the ETV6 (chromosome 12) and intron 1 of the RTN1 (chromosome 14). Identical breakpoints were detected in three analyzed carriers (the proband, her mother, and her maternal aunt) and confirmed by Sanger sequencing. The translocation creates two fusion genes: ETV6-RTN1 and RTN1-ETV6, which both harbor a frameshift leading to premature stop codon and
abolishing functional protein domains (Figs. 6 & 7). The fusions were confirmed to be actively transcribed by PCR.

Fig. 5. Karyotyping showing the constitutional balanced reciprocal chromosome translocation t(12;14) of a female individual.
Fig. 6. Schematic figure of the fusion mRNA transcripts ETV6-RTN1 (A) and RTN1-ETV6 (B). A. The fusion breakpoint is in intron 1 after the ETV6 exon 1. The fusion mRNA continues from RTN1 exon 2 but with a frameshift resulting in a stop codon. B. RTN1-ETV6 fusion breaks at the intron 1 after exon 1 of RTN1. The transcription continues from the exon 2 of ETV6, but the fusion leads to premature stop codon in the exon 4 of ETV6. ATG, start codon; E2/3/4, exon 2, 3, or 4.

Fig. 7. Schematic figure of the translocation breakpoints on the protein sequence. A. ETV6 protein, which is 452 amino acids long, with functional domains PNT and ETS. B. RTN1 protein (776 amino acids) and the functional domain Reticulon.
5.2.3 Characteristics of leukemia with constitutional chromosome translocation disrupting ETV6 (II)

Genotyping SNP array on the diagnostic bone marrow samples for both leukemia cases revealed a high hyperdiploid (HeH) karyotype despite the detection of normal karyotype through routine cytogenetic analysis at diagnosis (Table 5.). No CNAs (gains or deletions) were detected in ETV6, RTN1, or the leukemia-associated genes. Sanger sequencing of the intact ETV6 allele revealed no mutations in the leukemia samples.

Somatic aberrations were detected with WGS from the diagnostic bone marrow DNA in both leukemia cases. The aberrations were RIOK1, RFWD2, SALL4, ATP5I, FGD5, HHIP, SLC2A6, SFXN4, LAMA1, and FAM105B genes in the female index patient’s leukemia sample and the NRAS, C9orf3, RDH16, ZC3H18, PHEX, ZCCHC16, and RHOXF2B genes in the maternal aunt’s leukemia sample. Thirteen of the genes, of which two were identical variants (NM_002524.4:c.34G>A in NRAS and 27 kb WAC-AS1 deletion), were previously reported as somatic mutations in childhood ALL by the Pediatric Cancer Data Portal (PeCan) (St. Jude Children’s Research Hospital [2015–2018]; https://pecan.stjude.cloud/home. Accessed 7 August 2018.)

5.2.4 Expression analysis (II)

The expression levels of the wild-type ETV6 and RTN1 transcripts were analyzed by qPCR (the peripheral blood samples from the BCP-ALL cases at remission and the germline sample of the unaffected mother of the proband). The expression of wild-type ETV6 was decreased for the proband but, in contrast, increased in the maternal aunt. The unaffected mother of the proband showed expression levels comparable to those of the normal controls. Similar expression levels were detected for wild-type RTN1.

The wild-type ETV6 expression was at the same level in both leukemia cases, the sporadic BCP-ALL controls with somatic heterozygous ETV6 deletions, and the cases with both ETV6-RUNX1 translocation and ETV6 deletion. The ETV6 expression levels in the sporadic cases without somatic ETV6 aberration varied and overlapped with the aforementioned cases. The expression of wild-type RTN1 was low in all of the tested leukemic samples, and it did not differ significantly between the leukemia cases and the sporadic controls in this study.

62
Finally, expression of the fusion transcripts was detected in the proband, the maternal aunt, and the unaffected mother.

### Table 5. Clinical details of the leukemia cases with constitutional t(12;14) disrupting ETV6 (II)

<table>
<thead>
<tr>
<th>Clinical feature</th>
<th>Affected maternal aunt</th>
<th>Affected young female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age at dx</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>Immunophenotype</td>
<td>BCP-ALL</td>
<td>BCP-ALL</td>
</tr>
<tr>
<td>Karyotype</td>
<td>57,XX,t(12;14)(p13.2;q23.1)</td>
<td>54,XX,t(12;14)(p13.2;q23.1)</td>
</tr>
<tr>
<td>Highest WBC count at dx (x 10^9/l)</td>
<td>2.8</td>
<td>4.4</td>
</tr>
<tr>
<td>Highest blast count at dx (%)</td>
<td>0</td>
<td>10</td>
</tr>
<tr>
<td>Highest blast count at dx (%), BM</td>
<td>95</td>
<td>95</td>
</tr>
<tr>
<td>NOPHO Treatment protocol</td>
<td>ALL-1992, HR</td>
<td>ALL-2008, SR</td>
</tr>
<tr>
<td>Primary event</td>
<td>CR1</td>
<td>CR1</td>
</tr>
<tr>
<td>Severe adverse effect</td>
<td>Osteonecrosis</td>
<td>Osteonecrosis, vincristine neuropathy, osteoporosis with compression fractures</td>
</tr>
</tbody>
</table>

1 Additional chromosomes: +X, +X, +6, +10, +der(14)t(12;14)(p13.2;q23.1), +14, +17, +18, +18, +21, +21, +21
2 Additional chromosomes: +X, +4, +6, +der(14)t(12;14)(p13.2;q23.1), +17, +18, +21, +21

BCP-ALL, B-cell precursor acute lymphoblastic leukemia; BM, bone marrow; CR1, complete first remission; dx, diagnosis; HR, high risk; HSCT, hematopoietic stem cell transplantation; NA, data not available; NOPHO, Nordic Society of Paediatric Oncology; PB, peripheral blood; SR, standard risk; WBC, white blood cell

### 5.3 ERCC6L2 gene mutation causes inherited bone marrow failure syndrome (III)

#### 5.3.1 Patients with bone marrow failure without congenital anomalies (III)

Two children from unrelated families with no consanguineous marriages presented with symptoms of bone marrow failure without congenital anomalies (Table 6.).

The first male index patient had bruising and thrombocytopenia at the age of 8 years and became dependent on platelet transfusions. He had macrocytosis and vitamin B12 deficiency, but the macrocytosis was unresponsive to vitamin substitution. At age 9, the first male index patient developed neutropenia and became dependent on RBC transfusions at 13. The histology of the bone marrow
trephine revealed severe hypoplasia. An allogeneic HSCT from an unrelated donor was performed when the patient was 15 years old because of dependency on platelet and RBC transfusions after which the blood cell counts normalized. He had experienced pubertal delay, and at 17 years old, he was short (−2.7 SD). This was in line with his mother’s development. Late-maturing, she had her first menstruation at 14. No signs of learning difficulties have emerged in the index patient.

The second female index patient presented with thrombocytopenia and eosinophilia at the age of eight years. The bone marrow failure progressed with the histology showing reduced cellularity, and the patient developed anemia and leucopenia at 11. Her vitamin B12 level was slightly above the reference value, and she had no macrocytosis. Her bone marrow failure has not progressed to aplastic anemia, and she does not require transfusions. She had no congenital anomalies or signs of learning difficulties. Moreover, her growth and puberty have progressed normally.

The brother of the second index patient was discovered to be carrying the same homozygous *ERCC6L2* mutation when he was examined as a possible sibling donor for allogeneic HSCT for his sister. The brother is 10 years old and has had cytopenia of two blood cell lineages: thrombocytopenia (platelets $97 \times 10^9/l$) and mild leucopenia ($4.3 \times 10^9/l$). His hemoglobin is in the normal range. He is otherwise healthy and has no signs of developmental delay or congenital abnormalities. In addition, his growth is normal. He is subject to clinical follow-up because of his blood counts.

### Table 6. Clinical details of the two index patients with bone marrow failure (III)

<table>
<thead>
<tr>
<th>Patient</th>
<th>Age at presentation</th>
<th>Hb</th>
<th>WBCs</th>
<th>Platelets</th>
<th>First-cousin parents</th>
<th>Developmental delay or craniofacial anomalies</th>
<th>Other clinical features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male index</td>
<td>8 years</td>
<td>81 g/l$^1$</td>
<td>$2.3 \times 10^9/l$</td>
<td>$5 \times 10^9/l$</td>
<td>no</td>
<td>—</td>
<td>Short stature, pubertal delay, vitamin B12 deficiency</td>
</tr>
<tr>
<td>Female index</td>
<td>8 years</td>
<td>106 g/l$^2$</td>
<td>$2.0 \times 10^9/l$</td>
<td>$23 \times 10^9/l$</td>
<td>no</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

$^1$Blood cell counts at age 15. $^2$Blood cell counts at age 11, g, grams; Hb, hemoglobin; l, liter; WBCs, white blood cells;
5.3.2 Truncating mutation of ERCC6L2 results in bone marrow failure (III)

A homozygous deletion c.1457del (p.(Ile486fs)) (rs768081343 [GRCh37/hg19]) in the ERCC6L2 gene, which is located on chromosome 9, was found with exome sequencing in the index patients (Fig. 8). The deletion was heterozygous in the parents of both families, and it has an allele frequency of 5.2 per 1,000 in the Finnish population (http://gnomad.broadinstitute.org/). The mutation leads to a frameshift at isoleucine 486 and results in a premature stop codon at lysine 520 (NCBI RefSeq NM_020207.4 and NP_064592.2, isoform 1). The premature truncating mutation results in a loss of the catalytic helicase superfamily domain of the ERCC6L2 protein (Fig. 8).

The ERCC6L2 gene was recently linked to IBMFS and, in addition, it was associated with developmental delay and microcephaly in patients with first-cousin parents (Bluteau et al., 2017; Shabanova et al., 2018; Tummala et al., 2014; S. Zhang et al., 2016). The patients in this study presented no signs of developmental delay or microcephaly. The wild-type ERCC6L2 protein and the previously published variants, together with the deletion discovered in this study, are depicted in Figure 8.

5.4 Mitochondrial DNA sequence variation in patients with childhood ALL (IV)

5.4.1 Somatic novel nonneutral mtDNA variants (IV)

Eight heteroplasmic non-neutral variants of mtDNA were found in eight patients in the childhood ALL cohort. The mutations were in the genes encoding subunits of the OXPHOS enzyme complexes as follows: MT-ND1, MT-ND2, MT-ND4L, and MT-ND6 of complex I; MT-CO3 of complex IV; and MT-ATP6 and MT-ATP8 of complex V (Fig. 9). In addition, two novel synonymous variants, m.4475T>C and m.15355G>T were detected at diagnosis in two deceased patients. Six of the mutations were missense mutations resulting in amino acid transition, and two were nonsense mutations leading to a premature stop codon on the mRNA.
Fig. 8. Schematic presentation of the wild-type ERCC6L2 protein and the mutations leading to IBMFS including those in the present study. A. Wild-type protein of 1561 amino acids with locations of three previously published mutations and the functional domains, ATP-binding, and C-terminal helicases. B. The truncating mutation discovered in this study. C.–F. The truncating mutations published by Tummala et al. (2014) and Zhang et al. (2016) (C), by Bluteau et al. (2017) (D), by Tummala et al. (2014) and Bluteau et al. (2017) (E), and by Shabanova et al. (2018) (F). The figure was drawn using Illustrator for Biological sequences (Liu et al., 2015). fs, frameshift; NCBI, National Center for Biotechnology Information *, stop codon.
5.4.2 Non-neutral mtDNA variants associated with poor prognosis (IV)

Six of eight non-neutral variants were detected in deceased patients, of whom four had a poor prognosis genetic marker (Table 7.), two had \textit{BCR-ABL1} translocations, one had a \textit{KMT2A} rearrangement, and one (the male sibling in Study I) had \textit{IKZF1} deletion. Notably, only one patient had a genetic poor prognosis marker without mtDNA mutation.

The mutations were detected in the diagnostic samples (patients 2, 4, 6, 7, and 8) only at relapse (patients 1 and 5) or at diagnosis and the day 66 sample (patient 3), indicating that they were leukemia-associated, i.e., somatic. The prediction databases for the damaging effects of the mtDNA variants, SNAP2 and Polyphen-2, revealed the most damaging prediction for variants in deceased patients 2 and 3 with \textit{BCR-ABL1} aberration and \textit{IKZF1} deletion, respectively. The most benign prediction was for the surviving patient and the m.10578A>G (p.M37V; \textit{MT-ND4L}) variant.
Phyre2, which estimates the effect on protein conformation, revealed no change for either variant in the MT-CO3 gene. Two novel mtDNA nonsense mutations, which lead to truncated mRNA, were found in two deceased patients in the subunits of complex V: ATP synthase protein 8 and ATP synthase subunit a (patients 1 and 6 [Table 7]).

5.4.3 Mitochondrial haplogroups and detection of external DNA from blood transfusions (IV)

Mitochondrial haplogroup K was overrepresented in the childhood ALL cohort as compared to the previously published population cohort of 480 Finnish individuals. Four childhood ALL patients (11%) had haplogroup K in comparison to 12 individuals (2.5%) in the population cohort (Fisher’s exact test, 2-tailed \( P = 0.0199 \)).

The phylogenetic approach was used in this study; therefore, the haplogroups and the haplogroup specific variants were first determined in the patients through the use of HaploGrep. The variants associated with the haplogroups were not treated as somatic variants, i.e., leukemia-associated. With the phylogenetic approach, heteroplasmic mtDNA variants that define separate haplogroups were detected and this led to the discovery of eight patients with a mixed population of haplogroup variants.

In a previous study, external donor mtDNA from platelet transfusions has been shown to mimic mtDNA variants and to interfere with DNA analysis for three weeks after transfusions (Meierhofer et al., 2006). Satellite marker analysis was performed to rule out external DNA contamination, and no external genomic DNA was detected. However, the method cannot exclude external mtDNA contamination as the amount of mtDNA is more abundant than that of nDNA because each human cell has multiple copies of mtDNA but a single nDNA (Ylikallio & Suomalainen, 2012). Thus, heteroplasmic haplogroup-specific mtDNA variants could be detected despite the intact satellite marker results.

Three patients (3, 5, and 8) had both mixed populations of mtDNA and novel mtDNA mutations. The mutations were detected in the bone marrow samples specifically. It can be postulated that they would originate from leukemic cells in contrast to the external DNA, which was detected mainly in the blood samples.
Table 7. Clinical details of the patients and features of the heteroplasmic nonneutral mtDNA mutations

<table>
<thead>
<tr>
<th>Patient</th>
<th>Immuno-phenotype</th>
<th>Cytogenetics</th>
<th>Treatment protocol</th>
<th>Primary event</th>
<th>Cause of death</th>
<th>mtDNA variant</th>
<th>Gene</th>
<th>Amino acid transition</th>
<th>Prediction databases</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>BCP-ALL</td>
<td>t(9;22)(BCR-ABL1)</td>
<td>ALL-1992, HR</td>
<td>Relapse</td>
<td>Relapse after HSCT</td>
<td>m.14600G&gt;A MT-ND6</td>
<td>p.P25L</td>
<td>74</td>
<td>1.00</td>
</tr>
<tr>
<td>3</td>
<td>BCP-ALL</td>
<td>IKZF1 deletion</td>
<td>ALL-2000, HR</td>
<td>Relapse</td>
<td>Relapse after HSCT</td>
<td>m.5013T&gt;C MT-ND2</td>
<td>p.S182P</td>
<td>95</td>
<td>0.998</td>
</tr>
<tr>
<td>4</td>
<td>T-ALL</td>
<td>—</td>
<td>ALL-2000, HR</td>
<td>Relapse</td>
<td>—</td>
<td>m.10578A&gt;G MT-ND4L</td>
<td>p.M37V</td>
<td>10</td>
<td>0.00</td>
</tr>
<tr>
<td>5</td>
<td>BCP-ALL</td>
<td>—</td>
<td>ALL-2000, IR</td>
<td>Relapse</td>
<td>Refractory disease</td>
<td>m.3967G&gt;T MT-ND1</td>
<td>p.A221S</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>6</td>
<td>BCP-ALL</td>
<td>t(4;11)(KMT2A rearrangement)</td>
<td>ALL-2000, IR</td>
<td>Relapse</td>
<td>Secondary graft failure</td>
<td>m.8729G&gt;A MT-ATP6</td>
<td>p.W688Ter</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>7</td>
<td>BCP-ALL</td>
<td>t(9;22)(BCR-ABL1)</td>
<td>ALL-2000, HR</td>
<td>Relapse</td>
<td>Relapse after HSCT</td>
<td>m.9918G&gt;A MT-CO3</td>
<td>p.A238T</td>
<td>22</td>
<td>1.00</td>
</tr>
<tr>
<td>8</td>
<td>BCP-ALL</td>
<td>t(12;21)(ETV6-RUNX1)</td>
<td>ALL-2008, SR</td>
<td>Relapse</td>
<td>—</td>
<td>m.9583C&gt;T MT-CO3</td>
<td>p.P126L</td>
<td>39</td>
<td>1.00</td>
</tr>
</tbody>
</table>

1The score between 0 and 100 is damaging, and the score below zero is neutral. 2The score 0.00 equals benign, and 1.00 is the most damaging. M, Methionine; Y, Tyrosine; L, Leucine; Ter, stop codon; P, Proline; S, Serine; V, Valine; A, Alanine; W, Tryptophan; T, Threonine
6 Discussion

6.1 Germline variants of IKZF1 and ETV6 predispose to childhood ALL

6.1.1 Germline deletion of 7p12.1p13 causes intellectual impairment syndrome and susceptibility to childhood ALL

Germline microdeletion of chromosome 7, including the IKZF1 gene, was detected in two siblings with maternally inherited novel syndrome with overgrowth, distinct facial features, intellectual impairment, and susceptibility to childhood ALL. IKZF1 is frequently somatically deleted in BCP-ALL, and the deletions are associated with inferior survival (Mullighan et al., 2009). Rare germline IKZF1 variants have been recently shown to segregate with hypogammaglobulinemia, low B cell counts, and childhood ALL (Churchman et al., 2018; Kuehn et al., 2016). IKZF1 encodes the IKAROS transcription factor, which has an essential role in hematopoiesis by controlling chromatin remodeling and directing lymphoid differentiation (T. Yoshida & Georgopoulos, 2014).

The deceased male sibling had an additional IKZF1 deletion, which resulted in the homozygous deletion of the promoter and first exon regions. He had a slow response to induction therapy and underwent HSCT but suffered a relapse and died from the disease. The additional IKZF1 deletion may have had a role in the disease progression. Somatic focal deletions in leukemia-associated genes, such as PTEN, PAX5, CREBBP, and RB1, were also detected in the deceased male at diagnosis but not in the female sibling.

In this family, the germline deletion of IKZF1 appeared to predispose to childhood ALL with an autosomal dominant negative pattern with incomplete penetrance. Previous studies have documented heterozygous missense or loss-of-function germline IKZF1 variants in 53 familial cases in which eight developed ALL (Churchman et al., 2018; Hoshino et al., 2017; Kuehn et al., 2016; Mendoza Londono et al., 2005; N. Yoshida et al., 2017), in addition to 43 sporadic cases accounting for 0.9% of the patient cohort in the report by Churchman et al. (2018). Our findings suggest that not only missense or nonsense variants in IKZF1 but also deletions of the whole gene may predispose to ALL.

The patients in this study had no hypogammaglobulinemia or immunodeficiency. This was in contrast to previously published IKZF1 associated
with low B cell counts and hypogammaglobulinemia (Churchman et al., 2018; Hoshino et al., 2017; Kuehn et al., 2016). The affected female and the mother had normal function of the immune system cells, including the B, T, and NK cells, with normal cell counts detected with flow cytometry. Constitutional variants of \textit{IKZF1} do not always lead to defects in lymphoid cell development and maturation.

The syndrome, which was detected in the two siblings and the mother but with milder symptoms, included overgrowth, intellectual impairment, dysmorphic facial features, and susceptibility to childhood ALL. The evident cause was the microdeletion of 7p12.1p13 with the synergist effect of 31 deleted genes. Notably, the deletion included an imprinted \textit{GRB10} gene, which encodes a growth factor receptor-binding protein (Vecchione, Marchese, Henry, Rotin, & Morrione, 2003) and, when maternally inherited, could contribute to the overgrowth characteristic based on murine models (Charalambous et al., 2003). A previous study reported one patient with overgrowth, intellectual disability, childhood BCP-ALL, and Greig cephalopolysyndactyly caused by a large 7p deletion, including \textit{IKZF1} and \textit{GLI3} (Mendoza-Londono et al., 2005). The patients in this study did not have \textit{GLI3} deletion.

\textbf{6.1.2 Predisposition to childhood BCP-ALL caused by a constitutional chromosome translocation disrupting \textit{ETV6}}

In the other familial ALL family, a constitutional balanced reciprocal translocation t(12;14)(p13.2;q23.1), which was detected in two related females with childhood BCP-ALL, disrupted the \textit{ETV6} and \textit{RTN1} genes. This study focused on the \textit{ETV6} gene because of its documented role in normal hematopoiesis and leukemia. \textit{ETV6} has an important role in embryonic development and hematopoiesis (Wang et al., 1997; Wang et al., 1998). Somatic \textit{ETV6} translocations with several partner genes are found in various hematological malignancies (De Braekeleer et al., 2012), with the \textit{ETV6}-RUNXI being the most common structural lesion detected in BCP-ALL (Mullighan, 2012). Inherited variants of \textit{ETV6} have been found to segregate with thrombocytopenia and hematological malignancies (Melazzini et al., 2016; Noetzli et al., 2015; Topka et al., 2015; M. Y. Zhang et al., 2015). Furthermore, 31 distinct constitutional \textit{ETV6} variants in familial ALL and sporadic BCP-ALL cases (Moriyama et al., 2015).

Currently, 49 cases with childhood ALL and germline \textit{ETV6} variants have been published. Of these, 14 were familial cases from 10 families (Duployez et al., 2018; Melazzini et al., 2016; Moriyama et al., 2015; Noetzli et al., 2015; Topka et al., 2015).
Three sporadic cases of childhood ALL with constitutional ETV6 variants have been reported to develop secondary malignancies after chemotherapy (Junk et al. 2017). In addition, Dirse et al. (2018) and Poggi et al. (2017) have reported patients with germline ETV6 variants but no cases of ALL.

The family described in this thesis is the first reported family for whom predisposition to BCP-ALL is caused by a constitutional structural chromosome aberration involving the ETV6 gene. In the previously reported ETV6 variants, the mutations resulted in an altered ETV6 protein, which had a dominant-negative effect on the wild-type ETV6 (Moriyama et al., 2015; Nishii et al., 2016; Noetzli et al., 2015; M. Y. Zhang et al., 2015). Thus, in the present study the eliminating effect of the translocation on the ETV6 protein and its functional domains generates a different mechanism for leukemogenesis.

The BCP-ALL cases with germline ETV6 mutations have been characterized by HeH karyotype and older age at onset, along with a WBC count less than 50 × 10^9/l at diagnosis (Duployez et al., 2018; Moriyama et al., 2015; Nishii et al., 2016). Both of the leukemia cases in this study had duplications of the translocated chromosome derivative, der(14)t(12;14)(p13.2;q23.1), in addition to other chromosomes typically involved in hyperdiploidy (Paulsson et al., 2015). The patients were 8 and 12 at diagnosis and had WBC counts of 2.8 and 4.4 × 10^9/l. Patients with HeH karyotype have peak incidence at ages 3–5 (Paulsson, 2016). In contrast, patients with germline ETV6 variants and HeH BCP-ALL were, on average, 10 years old at presentation (Moriyama et al., 2015; Nishii et al., 2016).

The WGS showed no somatic variants in ETV6 or RTNI. Additional somatic variants in other genes were detected, but none was found to be shared by the affected females. Thus, the two leukemia cases in this family might have developed differently despite the common predisposing chromosome translocation. However, both leukemias were HeH karyotype, which was similar to the previously reported cases with germline ETV6 variants. This suggests that germline ETV6 variants may predispose specifically to HeH ALL. The receptor tyrosine kinase-RAS signaling pathway is frequently mutated in HeH BCP-ALL (Paulsson et al., 2015), and the maternal aunt had somatic missense mutation of NRAS in the leukemic blasts at diagnosis. In patients with germline ETV6 variants, additional somatic events are most likely required for leukemia development, as evident for example by the incomplete penetrance detected in this family, but the role of somatic events needs to be further studied.
6.1.3 Germline predisposition to ALL; and future perspectives

The conventional and risk-stratifying genetic aberrations of ALL blast cells, including \textit{ETV6-RUNX1} and hyperdiploidy, are considered the initiating events in leukemogenesis (Bhojwani et al., 2015; Hunger & Mullighan, 2015). Gröbner et al. (2018) and Ma et al. (2018) reported a lower mutation frequency in childhood ALL than in adult cancers. This supports the importance of recurrent genetic aberrations as an initiating event. With the mounting evidence of novel cytogenetic subgroups without any conventional somatic aberrations, e.g., \textit{PAX5} amplification (Schwab et al., 2017), these genetic variants may be more essential in leukemogenesis than was previously estimated. They represent initiating lesions rather than being only secondary events.

In addition, germline predisposition and the important role of the host genome in ALL development have become evident (Hunger & Mullighan, 2015; Inaba et al., 2013). Rare germline variants in the genes that affect lymphoid development or in the genes that are known to predispose to cancer have been associated with a high risk of the development of childhood ALL (Churchman et al., 2018; Gröbner et al., 2018; Moriyama et al., 2015; J. Zhang et al., 2015). Furthermore, the genetic risk loci for childhood ALL have been established with GWAS. However, achieving adequate statistical power in these studies is challenging because ALL is quite rare in the general population (Moriyama et al., 2015). The germline predisposition in childhood cancer has been estimated at approximately 7% (Gröbner et al., 2018; J. Zhang et al., 2015), and 4.4% in ALL (J. Zhang et al., 2015); however, this assessment is based on distinguished cancer-causing genes. Through GWAS, two novel susceptibility loci were recently detected at the intergenic regions of the genome (Vijayakrishnan et al., 2018). These findings suggest that the proportion of germline predisposition to childhood cancer is probably higher when all of the events on the whole human genome are taken into account. For example, epigenetics has provided novel tools for characterizing and classifying brain tumors (David Capper et al., 2018). These DNA methylation profiles, which are the most studied epigenetic modifications, have also been recognized in childhood ALL (Nordlund & Syvänen, 2017).

The genetic predisposition to childhood ALL has become more evident, and germline DNA is analyzed in the clinic when, on the basis of the subtype of ALL or a family history, the suspicion of a constitutional genetic cause. Clinics and working groups are initiating screening and surveillance for various predisposing gene variants and patient groups (Porter et al., 2017; Schütte et al., 2016). However,
in many cases, e.g., with germline ETV6 mutations and THC5, further research is still needed for the development of appropriate guidelines (Porter et al., 2017). It is hoped that the results on genetic predisposition will further broaden the knowledge about the genetic mechanisms behind ALL development. Ultimately, the treatment of ALL and other pediatric cancers should be personalized precision medicine and based on the genetic variants in the patient and in the cancer cells.

6.2 **ERCC6L2 causes IBMFS without congenital anomalies**

6.2.1 **Truncating c.1457del mutation in the ERCC6L2 gene**

An identical truncating mutation c.1457del(p.Ile486fs) of ERCC6L2 was discovered in two children, from unrelated families, with progressive BMFS and the brother of the female index patient. The mutation results in a frameshift and significantly truncated protein, which, if translated, would be missing the functional helicase domain. The patients had inherited bone marrow failure and presented no signs of any congenital anomalies or symptoms of developmental delay. The sole finding was pubertal delay, with short stature in the male index patient.

6.2.2 **ERCC6L2 and IBMFS**

Recently, mutations in the ERCC6L2 have been associated with IBMFS with developmental delay and congenital anomalies, including craniofacial abnormalities (Bluteau et al., 2017; Shabanova et al., 2018; Tummala et al., 2014; S. Zhang et al., 2016). The patients with congenital anomalies, developmental delay, and microcephaly all had first-cousin parents (Shabanova et al., 2018; Tummala et al., 2014; S. Zhang et al., 2016). Bluteau et al. (2017) published one case with craniofacial abnormalities, specifically facial dysmorphia, without consanguinity. In addition, six patients without consanguinity and ERCC6L2 mutation showed no neurological defects or congenital anomalies (Bluteau et al., 2017). The case with developmental delay in the Bluteau et al. (2017) report had been previously published by Zhang et al. (2016). The patients in this study had normal school performance and showed no signs of learning difficulties or developmental delay. Thus, the ERCC6L2 mutations might not cause neurological defects and developmental delay. Although only ERCC6L2 mutations were reported the in
patients with consanguinity (Shabanova et al., 2018; Tummala et al., 2014; S. Zhang et al., 2016), other genetic or epigenetic effects of consanguinity leading to developmental disorder cannot be ruled out.

ERCC6L2 is a member of the SNF2 family of helicase-like proteins, which function by using ATP to unwind and to remodel chromatin structures (Ryan & Owen-Hughes, 2011). The function of ERCC6L2 is related to the DNA damage response by the nucleotide excision repair (NER) mechanism, and the ERCC6L2 knock-down cells exhibited impaired fitness after cytotoxic substances specific to the NER pathway (Tummala et al., 2014). The NER pathway protects against single-strand covalent bond formation between nucleotides, that could be caused by exposure to UV irradiation, and environmental cytotoxic and chemotherapeutic agents (Mouw & D’Andrea, 2014). Zhang et al. (2016) demonstrated a more generalized response to DNA damage that includes double-strand breaks. In addition, ERCC6L2 has been implicated to in mitochondrial function, and after cytotoxic stress, the ERCC6L2 protein was detected to localize within mitochondria (Tummala et al., 2014).

The ERCC6L2 protein has an N-terminal helicase that binds ATP, and it has an aspartate-glutamate-alanine-histidine (DEAH) motif and the C-terminal helicase, which is a catalytic domain (Tummala et al., 2014). Notably, five of eight reported mutations, including the one in this study, were truncating variants abolishing the C-terminal catalytic helicase domain of ERCC6L2 (Bluteau et al., 2017; Shabanova et al., 2018; Tummala et al., 2014; S. Zhang et al., 2016). Tummala et al. (2014) showed that the truncated proteins formed aggregates and were mislocalized within the cell to the endoplasmic reticulum (ER), lysosomes, and autophagic vacuoles, thus impairing functionality of ERCC6L2.

Only one mutation has been reported to be located on the N-terminal ATP-binding helicase (Bluteau et al., 2017). Notably, the mutation c.G847A(p.D283N) was discovered in a female patient who later developed AML and died from the disease but no functional studies were performed by Bluteau et al. (2017).

6.2.3 Further studies

Currently, 12 cases of ERCC6L2 mutations and IBMFS have been published. Some have congenital anomalies and development delay. Neither the index patients nor the six patients published by Bluteau et al. (2017) had congenital anomalies or intellectual disability. Therefore, further studies and a larger cohort of patients with
bone marrow failure are needed for determining whether developmental delay is truly associated with $ERCC6L2$ variants.

Only one patient has been reported with hematological malignancy (AML) and $ERCC6L2$ mutation (Bluteau et al., 2017); however, the cancer predisposing effect of the $ERCC6L2$ gene is unknown. It would be important to determine the frequency of the underlying $ERCC6L2$ variants in acute leukemia patients, especially if preceding cytopenia has been observed.

### 6.3 Novel non-neutral mtDNA variants in the OXPHOS proteins detected in childhood ALL patients

#### 6.3.1 Main findings

Eight heteroplasmic non-neutral mutations, of which seven were novel, were discovered in the protein-coding genes of OXPHOS enzyme complexes in 22% of childhood ALL patients. The variants were either missense or nonsense variants, thus potentially damaging to the subunits of the OXPHOS enzymes. The somatic mtDNA mutations were often detected in patients who have relapsed and those with a genetic poor prognosis marker. Each of four of five patients with poor prognosis cytogenetic markers ($BCR-ABL1$, $KMT2A$ rearrangement, or $IKZF1$ deletion) had a nonsynonymous mtDNA variant. Furthermore, six of eight variants were detected in relapsed patients who had deceased.

The haplogroup K1 was overrepresented in the leukemia cohort. The phylogenetic approach, used for data analysis, revealed mixed populations of mtDNA in eight patients. The most likely cause was external DNA from blood transfusions.

#### 6.3.2 Mutations on OXPHOS subunits

The non-neutral mtDNA variants were detected in the genes encoding the subunits of the OXPHOS enzyme complex I ($MT-ND1$, $MT-ND2$, $MT-ND4L$, and $MT-ND6$), complex IV ($MT-CO3$), and complex V ($MT-ATP6$ and $MT-ATP8$). The metabolic shift towards glycolysis has been proved to result from the intricate reprogramming of the cancer cell by oncogenes and tumor suppressors, e.g., $TP53$, $RAS$, and $MYC$, instead of damaged mitochondria (Hanahan & Weinberg, 2011; Tan et al., 2014). Cancer cells rely on glycolysis, so if the OXPHOS is uncoupled from energy
production (Vander Heiden et al., 2009), the subunits could be vulnerable for mutations. The detected mtDNA mutations are most likely secondary effects of leukemogenesis, potentially resulting from the genomic instability of the cancer cell (Dasgupta & Hertweck, 2017) rather than driving mutations.

Notably, the mutations were on the OXPHOS enzyme complexes and were detected mainly in patients who have relapsed and patients with poor prognosis cytogenetic markers, including BRC-ABL1, KMT2A rearrangement, and IKZF1 deletion. One of the detected somatic mutations was found in the diagnostic bone marrow sample of a deceased patient, m.14600G>A. The mutation has been previously associated with Leigh disease and optic neuropathy, and has been proved to be pathogenic by decreasing the activity of complex I (Lin et al., 2012; Malfatti et al., 2007). The viability and aggressiveness of the cancer cell have been associated with low OXPHOS activity (Dasgupta & Hertweck, 2017). Indeed, furthermore, increased glycolysis correlates with poor prognosis in various cancers (Coller, 2014). Thus, the mutations in the subunits of OXPHOS enzymes could prove advantageous for the blast cells. To prove this, further functional studies should be performed.

6.3.3 Further studies

Normoxic glycolysis has been observed in blast cells derived from childhood ALL patients, (Boag et al., 2006), and increased glycolysis rates were associated with glucocorticoid resistance (Beesley et al., 2009; Hulleman et al., 2009). Thus, further studies of OXPHOS and glycolysis activity in relapsed childhood ALL patient-derived blast cells would be of interest. Recent studies have shown that the downregulation of glycolysis and the enhancement of OXPHOS could provide novel chemotherapeutic targets (Dickerson et al., 2017), which would also then be implicated in childhood ALL.

The patient cohort in this study, 36 patients, was small and biased toward poor-prognosis patients because 17 were deceased and there IR and HR patients were overrepresented. Thus, further studies are required to prove the connection between OXPHOS mutations and poor prognosis. Phylogenetic approach revealed that haplogroup K was overrepresented in the childhood ALL cohort, a finding that requires a larger cohort for proving the significance. Haplogroups have been associated with the increased risk of various cancers (Van Gisbergen et al., 2015). The phylogenetic approach revealed contamination from the blood transfusions, both platelets, and the RBCs. MtDNA from the platelet transfusions
has been previously shown to interfere data analysis (Meierhofer et al., 2006). Flawed data interpretation and laboratory errors have been reported in studies on the role of mtDNA in cancer (Salas et al., 2005; Yao, Macaulay, Kivisild, Zhang, & Bandelt, 2003). In previous studies on leukemia, some of the mtDNA variants that were reported as significant have been common polymorphisms and mutational hotspots; thus, the conclusions are somewhat unreliable. For example, m.152T>C is a common polymorphism on the D-loop, and it has been associated with good prognosis. (Kwok et al., 2011).

Whole genome and exome sequencing are the current methods for mtDNA analysis (Weissensteiner, Forer et al., 2016). It is important to note the risk of external mtDNA from blood transfusions. The phylogenetic approach for interpreting results and ensuring adequate coverage for WGS is key notes in future mtDNA studies with leukemia. The systematic phylogenetic analysis of mtDNA proved to be a prudent method for avoiding the common problems in mtDNA and cancer studies.

6.4 Strengths and limitations of the study

6.4.1 Strengths of the study

Germline predisposition studies

The ability to identify families with familial ALL and to recruit family members to participate in studies were the strengths of this study. Familial ALL is rare; thus, the identification of these families is crucial for studies of the underlying reason for susceptibility. Studies on familial ALL provide new information about ALL development and create important possibilities for developing diagnostic and surveillance protocols. The discovery of constitutional \textit{IKZF} and \textit{ETV6} variants in BCP-ALL, together with the findings in previously published cases, increases the current knowledge of not only leukemogenesis but also predisposition. The potential screening and surveillance of patients with germline variants requires further research and an adequate number of patients for developing appropriate guidelines. Because ALL is rare and cancer predisposition is even rarer and often caused by incomplete penetrance, these families are invaluable.
Another strength of these studies is the variable molecular methods that were used and the combination of laboratory, bioinformatic, and clinical expertise that was applied—all of which are important in translation research.

**Mitochondrial DNA variants in ALL**

The mitochondrial genome has its own characteristics with haplogroups, heteroplasmy, high variability, and a higher mutation rate than that of nDNA. Therefore, the interpretation of mtDNA results is difficult and requires expertise. The strength of this study was the expertise in the study group for mitochondrial genome research. Previous studies on the role of mtDNA in cancer have been criticized for false data interpretations.

### 6.4.2 Limitations of the study

The results of the studies of *IKZF1* and *ETV6* gene expression based on quantitative PCR should be interpreted with caution. RNA expression is very dependent on cell type, tissue and cell cycle; therefore, the comparison of samples from different patients and controls is not always reliable.

No functional studies on the protein level and with patient-derived cells were performed. This is a clear limitation of this thesis. For the family with constitutional chromosome translocation disrupting *ETV6*, proving the hypothesis of haploinsufficiency on protein level would have been interesting. Functional studies could have eliminated the problems with qPCR. In the mtDNA study, functional analyses of OXPHOS enzyme activities would have been interesting for understanding the effect of the novel mtDNA mutations.

The size of the study population, 36 patients, was a limitation in the mitochondrial genome study. Reliable conclusions on haplogroup frequencies and the association of mtDNA mutations with inferior prognosis are therefore difficult to make. Further studies on larger patient cohorts are needed. Another limitation was the retrospective nature of the study; thus, some patients had received transfusions before sampling.

### 6.4.3 Future

Systematic sample collection for WGS for ALL patients at diagnosis and at remission is ongoing at the clinics. This systematic genetic analysis will be part of
the new international treatment guidelines. The genetic testing for bone marrow failures is already an important part of the diagnostics at the clinics. These routine analyses will provide a great deal of genetic data and will require appropriate databases for information sharing. The WGS data also includes the mtDNA sequence, and this could provide interesting further studies. Novel and available genetic data will provide more information, including the pathogenesis and predisposition, on the diseases. This thesis is one piece of the genetic puzzle of leukemia and BMFS. It is hoped that in the future, all the pieces of the puzzle are discovered so that patients can benefit from better treatment options and cures.
7 Conclusions

The aims of the present study were to discover the genetic predisposing variants in ALL and IBMFS, and to determine pathogenesis of these diseases. The discovery of genetic variants will provide novel tools for diagnostics, treatment, and risk stratification. The goal is to have personalized care for each child based on the host and leukemia genetics.

The following conclusions can be drawn from this work:

1. Susceptibility to familial ALL was caused by inherited germline variants in the two families studied. Microdeletion in the chromosome 7p included the \textit{IKZF1} gene, which caused predisposition to BCP-ALL in the siblings with intellectual disability, cranio-facial anomalies, and overgrowth. The other predisposing constitutional variant was suggested to be \textit{ETV6}, which was affected by a constitutional structural chromosome translocation detected in two females with HeH karyotype BCP-ALL.

2. A homozygous c.1457del(p.Ile486fs) mutation of the \textit{ERCC6L2} gene was discovered in both of the patients who had no signs of classical IBMFS or neurological findings, developmental delay, or craniofacial anomalies. Only six other cases with solitary IBMFS and \textit{ERCC6L2} mutation have been published.

3. Mitochondrial DNA sequence variation was detected in childhood ALL patients. Of the cohort, 22% had non-neutral variants on the genes encoding subunits of the OXPHOS enzyme complexes. Despite the small cohort with an overrepresentation of IR and HR patients, the mtDNA mutations appeared to be associated with poor prognosis cytogenetic markers.
References


Hecht, M., Bromberg, Y., & Rost, B. (2015). Better prediction of functional effects for sequence variants. BMC Genomics, 16(8), S1.


Original publications


* = equal contribution; ** = equal contribution for senior authors

Reprinted with permission from John Wiley and Sons (I, III, IV).

Original publications are not included in the electronic version of the dissertation.
<table>
<thead>
<tr>
<th>Book Order</th>
<th>Title and Authors</th>
</tr>
</thead>
<tbody>
<tr>
<td>1471</td>
<td>Haapanen, Henri (2018) Preconditioning against ischemic injury of the central nervous system in aortic surgery: an experimental study in a porcine model with remote ischemic preconditioning and diazoxide</td>
</tr>
<tr>
<td>1472</td>
<td>Mahlman, Mari (2018) Genetic background and antenatal risk factors of bronchopulmonary dysplasia</td>
</tr>
<tr>
<td>1473</td>
<td>Lantto, Ulla (2018) Etiology and outcome of PFAPA (periodic fever, aphthous stomatitis, pharyngitis and adenitis) syndrome among patients operated with tonsillectomy in childhood</td>
</tr>
<tr>
<td>1474</td>
<td>Hintsala, Heidi (2018) Cardiovascular responses to cold exposure in untreated hypertension</td>
</tr>
<tr>
<td>1475</td>
<td>Alakortes, Jaana (2018) Social-emotional and behavioral development problems in 1 to 2-year-old children in Northern Finland: reports of mothers, fathers and healthcare professionals</td>
</tr>
<tr>
<td>1477</td>
<td>Honkila, Minna (2018) Chlamydia trachomatis infections in neonates and infants</td>
</tr>
<tr>
<td>1478</td>
<td>Krooks, Laura (2018) Malocclusions in relation to facial soft tissue characteristics, facial aesthetics and temporomandibular disorders in the Northern Finland Birth Cohort 1966</td>
</tr>
<tr>
<td>1479</td>
<td>Hoque Apu, Ehsanul (2018) Migration and invasion pattern analysis of oral cancer cells in vitro</td>
</tr>
<tr>
<td>1480</td>
<td>Härkin, Pia (2018) Closure of patent ductus arteriosus in very preterm infants: Potential role of paracetamol and consequences of current treatments</td>
</tr>
<tr>
<td>1481</td>
<td>Kelloniemi, Annina (2018) Novel factors regulating cardiac remodeling in experimental models of cardiac hypertrophy and failure</td>
</tr>
<tr>
<td>1483</td>
<td>Isohoookana, Joel (2018) Emerging novel prognostic markers in pancreatic ductal adenocarcinoma</td>
</tr>
<tr>
<td>1484</td>
<td>Toukola, Tomi (2018) Physical exercise and sudden cardiac death: Characteristics and risk factors</td>
</tr>
</tbody>
</table>

Book orders:
Granum: Virtual book store
http://granum.uta.fi/granum/
GERMLINE PREDISPOSITION TO CHILDHOOD ACUTE LYMPHOBLASTIC LEUKEMIA AND BONE MARROW FAILURE, AND MITOCHONDRIAL DNA VARIANTS IN LEUKEMIA