NOVEL INSIGHTS INTO GENETIC PREDISPOSITION TO MYELOID MALIGNANCIES

Elina Hirvonen

Faculty of Medicine
Department of Medical and Clinical Genetics, Medicum
Applied Tumor Genomics, Research Programs Unit
University of Helsinki
Helsinki, Finland

Academic dissertation

To be presented for public discussion with the permission of the Faculty of Medicine of the University of Helsinki, in Biomedicum Lecture Hall 2, Haartmaninkatu 8, Helsinki, on the 10th of May 2019, at 1.00 pm

Helsinki, 2019
CONTENTS

LIST OF ORIGINAL PUBLICATIONS ................................................................. 7
ABBREVIATIONS .................................................................................................. 8
GENE AND PROTEIN NAMES ........................................................................ 9
ABSTRACT ........................................................................................................... 10
INTRODUCTION .................................................................................................. 12

REVIEW OF THE LITERATURE ........................................................................ 13
1 Hematopoiesis in adults ............................................................................ 13
  1.1 Bone marrow niche ............................................................................. 15
    1.1.1 Production of myeloid and red blood cells ............................... 16
    1.1.2 Lymphopoiesis .......................................................................... 17
  1.3 Dysregulation of hematopoiesis ......................................................... 18
2 Hereditary predisposition to cancer ......................................................... 20
  2.1 Cancer genes and mutations ............................................................. 20
  2.2 Hereditary cancer syndromes ........................................................... 22
3 Myeloid malignancies ................................................................................. 23
  3.1 Genetic predisposition to myeloid malignancies .............................. 24
  3.2 Myeloproliferative neoplasms ......................................................... 25
    3.2.1 MPN pathogenesis ................................................................. 26
    3.2.2 Genetic predisposition to MPNs .............................................. 27
  3.3 Myelodysplastic syndromes ............................................................. 28
    3.3.1 MDS pathogenesis ................................................................. 29
    3.3.2 Genetic predisposition to MDS ............................................... 29
  3.4 Acute myeloid leukemia ................................................................. 30
    3.4.1 AML pathogenesis ................................................................. 31
    3.4.2 Genetic predisposition to AML ............................................... 32
  3.5 Hereditary bone marrow failure syndromes .................................... 33
    3.5.1 Dyskeratosis congenita – a BMF syndrome with leukemia
        predisposition .............................................................................. 34
4 Modern molecular techniques in genomic research ............................ 35
  4.1 Next-generation sequencing ........................................................... 36
  4.2 Genomics in leukemia research ..................................................... 37

AIMS OF THE STUDY ...................................................................................... 39
MATERIALS AND METHODS ................................................................. 40
1 Study subjects ................................................................................ 40
   1.1 Acute myeloid leukemia patients (I) ........................................ 40
   1.2 Polycythemia vera patients (II) .............................................. 41
   1.3 Dyskeratosis congenita study subjects (III) ............................ 41
   1.4 Population controls (I-III) .................................................... 42
2 Sample processing ...................................................................... 42
   2.1 DNA and RNA extraction (I-III) ........................................... 42
3 Cell lines ..................................................................................... 43
   3.1 EBV-transformation and culture of lymphocytes (III) .......... 43
   3.2 Commercial cell lines (III) ................................................... 43
4 Next generation sequencing (NGS) .............................................. 43
   4.1 Exome sequencing (I, II) ..................................................... 43
   4.2 Whole-genome sequencing (II) .......................................... 44
   4.3 Variant calling (I, II) ......................................................... 44
   4.4 Gene panels (I, III) .......................................................... 44
5 Sanger sequencing validation (I-III) ............................................ 45
6 Telomere length analysis (III) ...................................................... 45
7 Droplet digital PCR (III) ............................................................ 46
8 Online databases and in silico variant prediction tools (I, II) .... 46
9 Statistical analyses (I, III) .......................................................... 47

RESULTS .......................................................................................... 48
1 Identification of germline alterations in AML patients (I) ............. 48
   1.1 Variants of unknown significance ......................................... 48
   1.2 Germline variants with adverse or potentially pathogenic function 48
   1.3 Clinical and family history examination in conjunction with germline exome sequencing ........................................... 52
2 Exome sequencing analysis of familial PV (II) ............................ 52
   2.1 Germline variants in ZXDC, ATN1, and LRRC3 ................. 52
   2.2 Screening of frequent germline alterations in PV patients ..... 55
   2.3 Detection of somatic variants and LOH ............................... 55
3 Female carriers of germline X-chromosomal DKC1 insertion (III) ... 56
   3.1 Mutant allele burden and XCI in female DKC1 mutation carriers.. 56
   3.2 Telomere length and clinical features of the females .......... 57

DISCUSSION .................................................................................. 60
1 Germline alterations in AML patients ........................................ 61
   1.1 Variants of uncertain significance ....................................... 62
1.2 CHEK1 – novel candidate for AML predisposition..........................64
2 Candidate predisposition genes for familial PV.................................65
3 Females with germline mutation in X-chromosomal telomere gene......67
  3.1 Mutant DKC1 allele expression and telomere length .....................68
4 Leukemia germline research and diagnostics now and in future .........69

CONCLUDING REMARKS ........................................................................72

ACKNOWLEDGEMENTS .........................................................................74

REFERENCES ........................................................................................76

Appendix ....................................................................................................95
LIST OF ORIGINAL PUBLICATIONS

This thesis is based on the following publications that are referred to in the text by the Roman numerals I-III:


*These authors contributed equally to this work.
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AML</td>
<td>Acute myeloid leukemia</td>
</tr>
<tr>
<td>APL</td>
<td>Acute promyelocytic leukemia</td>
</tr>
<tr>
<td>BM</td>
<td>Bone marrow</td>
</tr>
<tr>
<td>BMF</td>
<td>Bone marrow failure</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary DNA</td>
</tr>
<tr>
<td>CHIP</td>
<td>Clonal hematopoiesis of indeterminate potential</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CLP</td>
<td>Common lymphoid progenitor</td>
</tr>
<tr>
<td>CML</td>
<td>Chronic myeloid leukemia</td>
</tr>
<tr>
<td>CMP</td>
<td>Common myeloid progenitor</td>
</tr>
<tr>
<td>DC</td>
<td>Dyskeratosis congenita</td>
</tr>
<tr>
<td>dd-PCR</td>
<td>Droplet digital polymerase chain reaction</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein-Barr virus</td>
</tr>
<tr>
<td>ET</td>
<td>Essential thrombocytopenia</td>
</tr>
<tr>
<td>ExAC</td>
<td>The Exome Aggregation Consortium</td>
</tr>
<tr>
<td>FA</td>
<td>Fanconi anemia</td>
</tr>
<tr>
<td>FAB</td>
<td>French-American-British classification of AML</td>
</tr>
<tr>
<td>FFPE</td>
<td>Formalin-fixed paraffin embedded</td>
</tr>
<tr>
<td>FHRB</td>
<td>Finnish Hematology Registry and Biobank</td>
</tr>
<tr>
<td>FIMM</td>
<td>Finnish Institute for Molecular Medicine</td>
</tr>
<tr>
<td>FuGu</td>
<td>Functional Genomics Unit</td>
</tr>
<tr>
<td>GMP</td>
<td>Granulocyte/macrophage progenitor</td>
</tr>
<tr>
<td>GoF</td>
<td>Gain-of-function</td>
</tr>
<tr>
<td>HM</td>
<td>Hematological malignancy</td>
</tr>
<tr>
<td>HSC</td>
<td>Hematopoietic stem cell</td>
</tr>
<tr>
<td>HSPC</td>
<td>Hematopoietic stem and progenitor cell</td>
</tr>
<tr>
<td>HUH</td>
<td>Helsinki University Hospital</td>
</tr>
<tr>
<td>LSC</td>
<td>Leukemic stem cell</td>
</tr>
<tr>
<td>LoF</td>
<td>Loss-of-function</td>
</tr>
<tr>
<td>MDS</td>
<td>Myelodysplastic syndrome</td>
</tr>
<tr>
<td>MEP</td>
<td>Megakaryocyte/erythroid progenitor</td>
</tr>
<tr>
<td>MPN</td>
<td>Myeloproliferative neoplasm</td>
</tr>
<tr>
<td>MPP</td>
<td>Multi-potent progenitor</td>
</tr>
<tr>
<td>NGS</td>
<td>Next-generation sequencing</td>
</tr>
<tr>
<td>NK</td>
<td>Natural killer cell</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>PB</td>
<td>Peripheral blood</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>Ph</td>
<td>Philadelphia chromosome</td>
</tr>
<tr>
<td>PMF</td>
<td>Primary myelofibrosis</td>
</tr>
<tr>
<td>PV</td>
<td>Polycythemia vera</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic acid</td>
</tr>
<tr>
<td>sAML</td>
<td>Secondary acute myeloid leukemia</td>
</tr>
<tr>
<td>SISu</td>
<td>Sequencing Initiative Suomi database</td>
</tr>
<tr>
<td>TBD</td>
<td>Telomere biology disorder</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>UTR</td>
<td>Untranslated region</td>
</tr>
<tr>
<td>VUS</td>
<td>Variant of unknown or uncertain significance</td>
</tr>
<tr>
<td>WHO</td>
<td>The World Health Organization</td>
</tr>
<tr>
<td>XCI</td>
<td>X-chromosome inactivation</td>
</tr>
</tbody>
</table>
GENE AND PROTEIN NAMES

ABL1  ABL proto-oncogene 1, non-receptor tyrosine kinase
ATG2B  Autophagy related 2B
ATM  ATM serine/threonine kinase
ATN1  Atrophin 1
BCR  BCR, RhoGEF and GTPase activating protein-
BLM  Bloom syndrome RecQ like helicase
BRCA1  BRCA1, DNA repair associated
BRCA2  BRCA2, DNA repair associated
CEBPA  CCAAT enhancer binding protein alpha
CHEK1  Checkpoint kinase 1
CHEK2  Checkpoint kinase 2
DDX41  DEAD-box helicase 41
DKC1  Dyskerin pseudouridine synthase 1
ELANE  Elastase, neutrophil expressed
EPO  Erythropoietin
FANCA-E  FA complementation group A-E
FLT3  FMS-like tyrosine kinase receptor 3
GATA2  GATA binding protein 2
G-CSF  Granulocyte-colony stimulating factor
GSKIP  GSK3B interacting protein
HBS1L-MYB  HBS1 like translational GTPase/MYB proto-oncogene
HIF  Hypoxia-inducible factor
IDH  Isocitrate dehydrogenase
IL  Interleukine
JAK2  Janus kinase 2
LRRC3  Leucine rich repeat 3
MECOM  MDS1 and EVI1 complex locus
MLH1  MutL homolog 1
MLL1  Mixed lineage leukemia protein-1
MSH2  MutS homolog 2
MSH6  MutS homolog 6
MYH11-CBFB  Myosin heavy chain 11/Core binding factor subunit beta
PALB2  Partner and localizer of BRCA2
PML-RARA  Promyelocytic leukemia/retinoic acid receptor alpha
PMS2  PMS1 homolog 2, mismatch repair system component
RAD52  RAD52 homolog, DNA repair protein
RBBP6  RB binding protein 6
RECQL4  RecQ like helicase 4
RPS19  Ribosomal protein S19
RUNX1  Runt related transcription factor 1
RUNX1T1  Runx1 translocation partner 1
SAMD9  Sterile alpha motif domain containing 9
SAMD9L  Sterile alpha motif domain containing 9 like
SBD5  Shwachman-Bodian-Diamond syndrome protein
SH2B2  SH2B adaptor protein 2
SPI1  Spi-1 proto-oncogene
STAT  Signal transducer and activator of transcription
TERC  Telomerase RNA component
TERT  Telomerase reverse transcriptase
TINF2  TERF1 interacting nuclear factor 2
TP53  Tumor protein 53
TPO  Thrombopoietin
WRN  Werner syndrome RecQ like helicase
ZXDC  ZXD family zinc finger C

Gene names and symbols are italicized in the text.
ABSTRACT

Familial clustering in hematological malignancies is a well-recognized phenomenon, and patients with germline predisposition are diagnosed with increasing frequency. Many genes and inherited syndromes have been identified as predisposing factors to acute myeloid leukemia (AML) and other myeloid disorders. Nevertheless, individuals with hereditary predisposition to hematological malignancy still continue to be underdiagnosed, and in some cases of familial myeloid diseases, the germline cause is unknown. Identification of inherited, disease-predisposing mutations is important for the screening of family members and other individuals at higher risk of developing a myeloid malignancy.

The first aim of this study was to discover the germline mutations in genes associated with AML pathogenesis and DNA repair in a consecutive, unselected series of 80 Finnish adult AML patients. We identified 34 variants of uncertain significance (VUS) or mutations in 16 genes in 42/68 (62%) patients with exomes available. Two variants in the genes DDX41 and SBDS have previously been reported in myeloid malignancies, but most of the identified gene alterations were found in DNA repair genes. The fraction of potentially pathogenic mutations in the patient series was 9%. Sixty percent of the study patients had a first or second-degree relative with a malignancy, and the VUS or mutation carriers more often had a positive family history of malignancies compared to non-carriers. This study implicates that germline defects possibly associated with AML can also be identified in older cases without a known family history of cancer. The results also suggest a novel candidate gene, CHEK1, in AML predisposition.

The second aim of this study was to identify novel candidate predisposition genes to a myeloproliferative neoplasm named polycythemia vera (PV) by conducting an exome sequencing analysis of three individuals in a Finnish family with four diagnosed PV patients. Three variants that may predispose to PV in this family and were shared by all the patients were identified in the genes ZXDC, ATN1, and LRRC3. Of these, ZXDC appeared the most
interesting candidate since it encodes a transcription factor that regulates gene transcription in myeloid cell differentiation. The variants were screened in eight other patients in six families with PV clustering, but those patients did not carry the variants.

The third aim of this study was to investigate an X-chromosome-linked telomere biology disorder (TBD) in females with a heterozygous germline $DKC1$ mutation. TBDs predispose to cancer, including hematological malignancies. We studied three female mutation carrier siblings with dyskeratosis congenita (DC)-like manifestations; usually, X-chromosome inactivation (XCI) silences the defective X-chromosome in females, protecting them from symptoms. Droplet digital PCR was utilized in examining the XCI status and mutant allele expression in different tissue samples. The results showed expression of both alleles in blood in two out of three symptomatic females. Further, only two of the females showed shortened telomere length, suggesting that the relationship between telomere length and the severity of symptoms is not straightforward in mutation carrier females.
INTRODUCTION

Molecular biology research has always been crucial in hematology. Leukemia studies have had a remarkable role in advancing understanding of cancer biology and providing new innovative treatment options. Leukemia differs from other types of cancer in many ways, which makes it an attractive target for research. Firstly, the tissue is easily accessible. Secondly, leukemia has a relatively simple set of genetic aberrations and karyotype compared to many other cancer types; however, the molecular basis has only recently become clearer. The development of massively parallel next-generation sequencing (NGS) technologies has revolutionized cancer genomics research and enhanced the generation of targeted therapies and precision medicine.

Acute myeloid leukemia (AML) is the most common leukemia type in adults. Despite most cases being sporadic, families with AML and other myeloid malignancies have been documented for decades.\(^1\) Many genes and inherited syndromes are known to associate with predisposition to myeloid disorders. NGS technologies have facilitated the identification of genetic mutations, and individuals with germline predisposition to hematological myeloid malignancies are recognized at an accelerating pace. Thus, the World Health Organization (WHO) included familial myeloid malignancies in its leukemia classification scheme in the year 2016.\(^2\) However, inherited predisposition to hematological malignancies is still underdiagnosed in a subset of cases, and in some familial myeloid neoplasia patients the germline cause is not known.

The aim of this study was to identify germline alterations possibly predisposing to myeloid malignancies, especially AML and a myeloproliferative neoplasm named polycythemia vera (PV). In addition, telomere biology disorders (TBD) are known to predispose to myeloid malignancies due to bone marrow defects; one of the aims was to characterize an inherited X-linked TBD in female \textit{DKC1} mutation carriers. The main methods used were exome sequencing, in which biobank samples were utilized, and PCR-based methods.
REVIEW OF THE LITERATURE

1 Hematopoiesis in adults

Human hematopoiesis is a complex process producing up to one trillion new, mature blood cells every day in a healthy individual. The volume of whole blood is approximately five liters, consisting of the liquid component called plasma and circulating cells. Blood cells can be divided into three major groups. Red blood cells (erythrocytes) provide oxygen and carbon dioxide transport; platelets (thrombocytes) form blood clots and heal wounds; and white blood cells (leukocytes) protect the body against invading pathogens and other infections. Hematopoiesis takes place in the soft fatty tissue called bone marrow (BM). BM is located in the medullary cavity of vascularized, innervated bone and contains many different hematopoietic and non-hematopoietic cell types. In adults, the hematopoietically active marrow is localized in the shoulder and pelvic girdles, sternum, ribs, vertebrae, and lower skull. All blood cells originate from pluripotent self-renewing hematopoietic stem cells (HSCs), which undergo an asymmetric cell division resulting in an identical daughter cell and a multipotent progenitor cell (MPP). MPPs further commit to myeloid lineage differentiation (common myeloid progenitor, CMP) or lymphoid lineage differentiation (common lymphoid progenitor, CLP). CMPs give rise to cells of the myeloid lineage including granulocytes, macrophages, erythrocytes, and platelets whereas CLPs differentiate into long-lived cells of adaptive immunity (T- and B-cells) or cells of innate immunity (natural killer cells, NK). (Figure 1). Cell production is rapidly responsive to external or internal changes such as infections or anemia, and the lifetime of mature cell types ranges from hours to years.

Hematopoiesis has classically been described as a cellular hierarchy with HSCs residing at the apex of the pyramid. As hematopoiesis proceeds, the cells gradually lose their differentiation potential and finally become
committed to a single cell lineage; at each developmental stage the genes associated with the particular pathway remain expressed or upregulated, whereas the genes specifying the other lineages are silenced.\textsuperscript{7} Hematopoietic cytokines, hormones, and other factors regulate many steps of hematopoietic cell production and function. These factors can be lineage-specific or regulate cells in multiple lineages.\textsuperscript{8} Certain cytokines also prevent the cells from undergoing exhaustion, which is crucial for blood system homeostasis, or trigger DNA repair mechanisms in response to DNA damage.\textsuperscript{9,10} Several transcription factors (TF) are essential in the early regulation of hematopoietic cell fate decisions as well.\textsuperscript{11} Overall, a great number of signaling pathways interact with each other in a well-organized manner in hematopoiesis.

![Figure 1. Simplified view of adult hematopoiesis. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte/macrophage progenitor; MEP, megakaryocyte/erythroid progenitor.](image-url)
1.1 Bone marrow niche

HSCs proliferate and differentiate into mature cells in the extravascular spaces between marrow sinuses in BM. Special local BM microenvironments, termed niches, are composed of different types of cells and molecules (interleukins, interferons, chemokines, and tumor necrosis factors) that regulate HSC function and behavior in both homeostasis and pathological conditions. Distinct hematopoietic progenitors and HSCs have distinct niches in the bone marrow, and adhesion receptors play a key role in the localization and release of maturing cells from these niches. In response to hematopoietic stress the niche can shift to extra-medullary sites.

A BM niche is highly vascularized due to the vast production of hematopoietic cells. Perivascular stromal cells, including different mesenchymal cells and CXCL12 (stromal-derived factor-1, SDF-1) abundant reticular (CAR) cells act as key components in the HSC niche by expressing high levels of major niche factors. The arteries in the BM-penetrating bone canal branch into smaller arterioles, which further transit to venous sinusoids through which mature blood cells leave the BM into systemic circulation. Sinusoids compose a complex network in the BM cavity, and sinusoidal endothelial cells support HSC proliferation. Mature blood cell types such as macrophages and megakaryocytes residing in the BM and adipocytes also contribute to HSC quiescence or proliferation by cytokine secretion. Osteoblasts in the endosteam, the interface of the bone and marrow, secrete cytokines and growth factors that maintain the stem cell and progenitor population pool in BM. They also mediate the migration of HSCs. Furthermore, the sympathetic nervous system participates in the regulation of HSC function as well.

HSCs are not randomly distributed in the BM. The level of oxygen modulates cell activity and function; hypoxic response in regulating the quiescence of HSCs is of great importance and is thought to protect long-term HSCs from DNA damage with only a few of them entering the DNA
synthesis and proliferation phase of the cell cycle. In addition, quiescent HSCs with a low level of reactive oxygen species (ROS) are localized near less permeable arterioles, whereas the more permeable sinusoids promote stem cell activation and are the site for leukocyte trafficking. The mechanisms by which niche cells regulate HSCs are thus complex and diverse, and abnormalities in the BM microenvironment or disruption of the niche regulation may initiate or collaborate in the development of hematologic malignancies.

1.1.1 Production of myeloid and red blood cells

Myelopoiesis is the production of innate immune cells that develop from a common myeloid progenitor in the BM. CMPs are considered to undergo restriction into granulocyte/macrophage progenitors (GMPs) and megakaryocyte/erythroid progenitors (MEPs) that gradually differentiate into the mature cells lineages of granulocytes and monocytes/macrophages, and platelets and red blood cells, respectively. Myelopoiesis is tightly regulated. For example, the TFs encoded by the genes GATA1, GATA2, SPI1 (PU.1), CEBPA, and RUNX1 drive the differentiation and commitment of myeloid specific cell types at specific stages of maturation. Additionally, myelopoietic cytokines including stem cell factor (SCF), FMS-like tyrosine kinase 3 ligand (FLT3-L), granulocyte-macrophage colony-stimulating factor (GM-CSF), and interleukins (IL)-3 and IL-6, are important in the maintenance and self-renewal of hematopoietic stem and progenitor cells (HSPC), and in the different steps of myeloid cell maturation.

Three types of granulocytes with different roles in the immune system are present in humans: neutrophils, basophils, and eosinophils. Granulocytes synthesize proteins and store them in cytoplasmic granules, which are used in e.g. microbial killing. Early precursor cells develop into myeloblasts, and further into promyelocytes. These cells continue along the maturation process, and the mature granulocytes reside in the blood for an average of 10 hours before leaving the circulation and heading to sites of inflammation.
Pathogen-digesting macrophages are derived from monocytes that have entered the tissues for maturation process after circulating in the bloodstream for one day. The most abundant cell type in the blood is erythrocytes. They deliver hemoglobin-bound oxygen to body tissues via the circulatory system and carry some of the waste product, carbon dioxide, back from the tissues. Similar to other blood lineage cell production, the production of red blood cells (erythropoiesis) is also tightly regulated. An MEP undergoes a series of divisions and maturation steps, finally resulting in erythroblasts. They lose their nucleus and become reticulocytes that leave the BM to circulation. Reticulocytes lose their ribosomes and mitochondria and finally mature into red blood cells. A hormonal regulator called erythropoietin (EPO), which is primarily produced in the kidneys, controls the production of erythrocytes. Another principal TF in erythropoiesis is GATA1, which together with EPO influences the function and development of early progenitor to late erythroblasts. Tissue oxygenation affects the number of red blood cells produced, which is regulated by hypoxia-inducible factors (HIF), HIF-1 and HIF-2. Platelets (thrombocytes), which originate from the same progenitor cells as erythrocytes, are small cell fragments derived from megakaryocytes and play a crucial role in blood clotting. Thrombopoietin (TPO) hormone, produced in the liver, affects platelet production. EPO and TPO are examples of endocrine signaling, meaning that the molecules circulate in the blood and can have an effect far from the production site, whereas many cytokines are produced and presented locally in the BM (autocrine or paracrine signaling).

1.1.2 Lymphopoiesis

Lymphopoiesis produces the infection-fighting B and T-lymphocytes, NK cells, and a proportion of dendritic cells. Lymphocytes make up the majority of lymphoid tissue, which is found in e.g. lymph nodes, the spleen, and the thymus gland. B-lymphocytes function in humoral, adaptive immunity by protecting the body against invaders with specific antibodies, whereas T-lymphocytes serve in cell-mediated, cytotoxic adaptive immunity; they attack infected cells and tumors and help in the regulation of the immune
system. NK cells function in cell-mediated cytotoxic innate immunity.\textsuperscript{38} Like myelopoiesis, lymphopoiesis is also tightly regulated by cytokines and other factors. The Ikaros gene family of TFs and PU.1 act in parallel in controlling the transition of HSCs into CLPs,\textsuperscript{38,39} which have the potential to differentiate into any of the lymphoid lineages.\textsuperscript{38,40,41} B and T-cells also form subsets of memory cells that maintain the ability to rapidly reactivate upon restimulation with the same antigens.\textsuperscript{42}

1.3 Dysregulation of hematopoiesis

Hematopoiesis is normally regulated through cytokines that bind to their receptors in the cell membrane, which results in activation of intracellular signaling cascades. Three important signal transduction pathways with key roles in cell proliferation and differentiation include the Janus kinase/signal transducer and activator of transcription (JAK-STAT), phosphatidylinositol 3-kinase/protein kinase B/mammalian target of rapamycin (PI3K/AKT/mTOR), and mitogen activated protein kinase (MAPK) pathways.\textsuperscript{43-45} Mutations can alter these signaling events remarkably. Activating mutations in genes encoding receptor tyrosine kinases can cause constant proliferation of the target cells. In addition, deletions in tumor suppressor genes may result in inability to maintain homeostasis. Thus, dysregulation of hematopoiesis can lead to hematopoietic deficiency, immunodeficiency or hematological cancer.

Somatic mutations in genes associated with myeloid malignancies in HSCs and clonal hematopoiesis are often seen at low frequency in aging healthy individuals (age-related clonal hematopoiesis); a stem cell has gained a growth advantage leading to clonal expansion of the cells.\textsuperscript{46} Sometimes the clone will progress to a premalignant state to meet the diagnostic criteria of clonal hematopoiesis of indeterminate potential (CHIP), which is diagnosed when a somatic mutation with a mutant allele fraction of at least 2\% is present in the peripheral blood without any evidence of hematological malignancies.\textsuperscript{46,47} CHIP is associated with an increased risk of developing
hematologic malignancies such as the “preleukemia” known as myelodysplastic syndrome (MDS) or a blood cancer.\textsuperscript{47,48} Up to 10\% of individuals over the age of 65 years show recurrent somatic mutations associated with myeloid malignancies, and two of the most recurrently mutated genes encode the epigenetic regulators DNMT3A and TET2.\textsuperscript{49-51} The presence of these initial mutations in the premalignant landscape can be detected years before a potential AML diagnosis, which suggests a period of latency preceding AML.\textsuperscript{52,53} Although the risk of acquiring clonal expansion increases during aging, most individuals will never develop MDS or AML; the acquisition of additional somatic mutations is needed for malignant transition.\textsuperscript{47}

Only a small proportion of the alterations in HSCs are pathogenic driver mutations.\textsuperscript{54} Pathogenic mutations affect the stem cells’ ability to differentiate into mature cells and lead to their transformation into preleukemic cells. These preleukemic cells are further transformed into leukemic cells when additional mutations occur.\textsuperscript{55-57} Leukemia stem cells (LSC) are known to share several characteristics with normal HSCs,\textsuperscript{58} albeit they can create niches that disrupt the normal HSC behavior in the microenvironment to favor their own expansion.\textsuperscript{59-61}

Little is known about the impact of other cellular components in BM niches on leukemic transformation. However, morphological and functional changes have been reported in BM stromal cells in patients with MDS, AML, and primary myelofibrosis (PMF).\textsuperscript{62} For example, genetic alterations in osteogenic cells in the endosteal compartment may lead to MDS, and further to secondary leukemia.\textsuperscript{63} Alterations of the microenvironment can promote myeloproliferative neoplasms (MPN) as well. The progression of PMF, which is one of the MPNs, is thought to remodel the BM niche in a way that leads to the impairment of normal hematopoiesis and favoring LSC function.\textsuperscript{62} Distinct signals in these niches might also affect malignant transformation in many ways.\textsuperscript{9,64} The BM microenvironment is hypoxic in myeloid malignancies; the level of oxygen modulates cell activity and function.\textsuperscript{65}
2 Hereditary predisposition to cancer

Pathogenic germline mutations in cancer genes, especially in tumor suppressors capable of preventing tumorigenesis, cause predisposition to malignancies. More than 100 genes have been identified that predispose to cancer, and about 5-10% of cancers have a heritable component; the majority of them are inherited in an autosomal dominant manner with varying penetrance. This means that hereditary diseases are not always expressed in the same way in every individual carrying the same mutation. Patients with a germline mutation in a cancer gene have a highly or moderately increased risk of developing cancer. Two independent mutations are required for tumor development according to Knudson’s “two-hit” model: in the hereditary form, the first mutation is inherited in the germline, and the second event occurs somatically. Thus, individuals with a germline mutation often develop cancer at a younger age since only one additional somatic mutation is required for tumorigenesis. They also often develop multiple tumors during their lifetime. Also, loss of just one copy of a tumor suppressor gene is occasionally enough to provide a growth advantage to a cell.

2.1 Cancer genes and mutations

A malignant tumor originates from a single cell and evolves through clonal expansion. The acquisition of somatic mutations in the cell provides a growth advantage, which further drives the tumor progression. DNA damage and errors can arise during DNA replication or be caused by external mutagens. If not repaired, the mutations will be present in all the cell’s offspring. More than 200 genes have been recognized as drivers in common cancers; however, one single mutated gene is not yet enough to cause cancer.

Cancer genes can be classified into oncogenes and tumor suppressors. Dominantly acting oncogenes encode factors that control cell proliferation or apoptosis, and in cancer they are activated via gain-of-function (GoF)
mutations, which leads to constitutive activation of the gene. In contrast, recessively acting tumor suppressor genes are repressed via loss-of-function (LoF) mutations.\textsuperscript{72,73} Both alleles need to be inactivated for a tumor suppressor to contribute to tumor development. The most common mechanism for the inactivation of a wild type allele of a tumor suppressor gene is loss of heterozygosity (LOH) due to chromosomal deletion, uniparental disomy or mitotic recombination.\textsuperscript{74} Some tumor suppressors may display haploinsufficiency (loss of one copy of the gene drives tumorigenesis). Tumor suppressor genes can be further classified into subgroups based on their functions. Caretakers act in maintaining genomic stability and gatekeepers affect cellular proliferation and prevent tumor growth.\textsuperscript{72,73} In addition, a third group of tumor suppressor genes called landscapers helps in creating microenvironments that control cell growth and promote tissue homeostasis by regulating e.g. cellular adhesion markers and growth factors.\textsuperscript{75}

Cancer genetics is research of two genomes: germline and cancer (somatic) mutations (Figure 2). By comparing the DNA in cancer cells with that in normal cells, the genetic changes in the cancer cells can be identified. In general, cancer cells have more genetic changes than normal ones. As the cancerous tumor continues to grow, additional changes will occur, meaning that cancer cells may have different mutations within the same tumor.\textsuperscript{71} The genetic information from the tumor genome can further have an impact on the therapy selected for treatment. Tumor sequencing can, however, also reveal the presence of inherited mutations. A number of known germline and tumor-mutated cancer genes overlap: almost 50 of the genes with known somatic driver mutations in cancer are also included within the group of more than 100 cancer predisposition genes.\textsuperscript{66} The identification of these cancer predisposition genes and mutations, too, has a huge impact on clinical diagnosis, treatment, and possible prevention of cancer.
Figure 2. Germline mutations are hereditary and present in every cell of the body, since they occur in the sex cells or gametes that participate in fertilization. In contrast, somatic mutations result from changes in DNA of somatic cells of the body and are only transferred to daughter cells that form after cell division.

2.2 Hereditary cancer syndromes

Germline mutations cause hereditary predisposition to cancer. The majority of these inherited mutations are LoF mutations in tumor suppressor genes that often predispose to specific tumor types, even though many of the tumor suppressor genes function in almost every single cell of the body. One of the best-recognized cancer predisposition syndromes is Li-Fraumeni syndrome, which is due to germline LoF mutations in an important tumor
suppressor gene TP53.\textsuperscript{76} It encodes the protein p53, which functions in e.g. DNA damage repair, maintaining genomic stability, and cell-cycle arrest. Penetration in Li-Fraumeni syndrome is almost complete, and the patients have an extremely high lifetime cumulative risk of developing multiple malignancies. The most frequent cancer types in patients with Li-Fraumeni syndrome include sarcomas, breast cancer, brain tumors, and acute leukemias.\textsuperscript{77} Another well-known cancer predisposition syndrome is Lynch syndrome, driven by mutations in DNA mismatch repair genes including \textit{MSH2}, \textit{MLH1}, \textit{MSH6}, and \textit{PMS2}, with risk of especially early colorectal cancer development.\textsuperscript{78} In addition, e.g. PTEN hamartoma tumor syndrome with germline mutations in the tumor suppressor \textit{PTEN} is one of the cancer predisposition syndromes.\textsuperscript{79} Despite the rarity of these syndromes, they are clinically relevant in directing cancer prevention options.

3 Myeloid malignancies

Myeloid malignancies are clonal hematopoietic disorders resulting from genetic and epigenetic alterations that disturb normal processes in HSPCs.\textsuperscript{80} HSPCs accumulate mutations throughout life. Most of these somatic mutations are passengers, meaning that they have no contribution to clonal expansion.\textsuperscript{81,82} Instead, pathogenic mutations in genes encoding signaling pathway proteins, TFs, epigenetic regulators, tumor suppressors, and components of the spliceosome lead to excessive proliferation, abnormal self-renewal, and differentiation defects in the HSPCs.\textsuperscript{80} Functional and genetic changes contributing to the development of myeloid malignancies have been noted in BM niche cells as well.\textsuperscript{83} Myeloid malignancies comprise mainly chronic stages such as MPNs and MDS, and acute (AML) stages. The World Health Organization (WHO) system classifies myeloid malignancies into these major categories based on peripheral blood (PB) counts and smear analysis, BM morphology, karyotype, and genetic tests.\textsuperscript{2}
3.1 Genetic predisposition to myeloid malignancies

Familial clustering of hematological malignancies (HM) has been reported for decades.\(^1\) Usually the first member of the family has already developed an HM, and testing of additional family members identifies the same mutation but they are not yet diagnosed with cancer.\(^8\)\(^4\) Pathogenic germline mutations have been described in myeloid malignancies in several genes, of which \textit{RUNX1} was the first in 1999.\(^8\)\(^5\) Since then, multiple genes have been identified to associate with predisposition to MDS/AML. Germline mutations in e.g. \textit{ANKRD26}, \textit{CEBPA}, \textit{GATA2}, \textit{ETV6}, \textit{TP53}, \textit{BRCA1/2}, \textit{DDX41}, \textit{ELANE}, \textit{SAMD9}, \textit{SAMD9L}, and \textit{SRP72}, in addition to certain inherited BMF syndromes, are frequently found in hereditary HM patients.\(^8\)\(^4\)\(^,\)\(^8\)\(^6\)\(^,\)\(^8\)\(^7\) The affected proteins are involved in multiple functions including transcription, telomere maintenance, DNA repair, RNA processing, and inflammation.\(^8\)\(^6\)

As individuals with germline predisposition to myeloid HMs are diagnosed at an increasing frequency, WHO included familial hematological myeloid malignancies in its leukemia classification scheme in the year 2016.\(^2\) Familial myeloid disorders are classified into distinct subtypes (\textbf{Table 1}).\(^8\)\(^4\)

\begin{table}[h]
\centering
\begin{tabular}{|l|}
\hline
\textbf{Myeloid neoplasms with germline predisposition without a preexisting disorder or organ dysfunction} \\
AML with germline \textit{CEBPA} mutation \\
Myeloid neoplasms with germline \textit{DDX41} mutation \\
\hline
\textbf{Myeloid neoplasms with germline predisposition and preexisting platelet disorder} \\
Myeloid neoplasms with germline \textit{RUNX1} mutation \\
Myeloid neoplasms with germline \textit{ANKRD26} mutation \\
Myeloid neoplasms with germline \textit{ETV6} mutation \\
\hline
\textbf{Myeloid neoplasms with germline predisposition and other organ dysfunction} \\
Myeloid neoplasms with germline \textit{GATA2} mutation \\
Myeloid neoplasms associated with BM failure syndromes \\
Myeloid neoplasms associated with telomere biology disorders \\
Juvenile myelomonocytic leukemia associated with neurofibromatosis, Noonan syndrome or Noonan syndrome-like disorders \\
Myeloid neoplasms associated with Down syndrome \\
\hline
\end{tabular}
\caption{Classification of myeloid neoplasms with germline predisposition (WHO 2016).}
\end{table}

\(^*\)Adapted from publications\(^2\),\(^8\)\(^4\).
3.2 Myeloproliferative neoplasms

MPNs are clonal, chronic HSC disorders with abnormal hematopoietic proliferation and an increased tendency toward leukemic transformation. The somatic initial mutation in HSCs results in the excessive production of one or more types of terminally differentiated myeloid lineage cells due to the hypersensitivity for cytokine regulation and the absence of feedback regulation by mature cells.\textsuperscript{88,89} MPNs are divided into subcategories: chronic myeloid leukemia (CML), which is characterized by the \textit{BCR-ABL} oncogene fusion (Philadelphia chromosome, Ph-positive), and Ph-negative disorders named polycythemia vera (PV), essential thrombocytemia (ET), primary myelofibrosis (PMF), and prefibrotic PMF.\textsuperscript{2} Of Ph-negative MPNs, PV is primarily associated with a high number of red blood cells, whereas ET patients have high platelet counts. Patients with PMF usually have high counts of granulocytes, BM failure resulting from reticulin or collagen fibrosis, and extramedullary hematopoiesis in the spleen and liver.\textsuperscript{90} Prefibrotic myelofibrosis, characterized by granulocytic and megakaryocytic proliferation and lack of fibrosis in the BM, represents an early phase of myelofibrosis.\textsuperscript{91} These cellular defects can cause health problems such as thrombosis, blood clotting and weakness. An overlap among the Ph-negative diseases exists in terms of pathophysiological and molecular features, but diagnostic accuracy is prognostically relevant due to the higher tendency of leukemic transformation of certain MPNs.\textsuperscript{2} The risk of leukemia development is highest in PMF with an incidence of 10\% to 20\% during the first decade, and lowest in ET with an incidence of less than 1\% in the first 10 years.\textsuperscript{92,93} The tendency of leukemic transformation in PV patients is approximately 2\% in the first 10 years of the disease and 8\% at 20 years.\textsuperscript{94} Transitions among these diseases are also observed occasionally.\textsuperscript{83} Prior to leukemic transformation, MPNs often progress to BM fibrosis, which can present with very severe symptoms even at the preleukemic stage.\textsuperscript{95} However, many patients with MPNs have a normal life span if the disease is properly treated and controlled. Treatment in PV and ET is primarily directed at the prevention of thrombohemorrhagic complications and to relieve the symptoms, but no current drug therapy is curative. Some PMF patients receive stem cell
transplants and have the possibility of being cured. The overall incidence of MPNs in Finland is approximately 6 in 100,000 individuals (Finnish Cancer Registry, Figure 3).

3.2.1 MPN pathogenesis

MPNs share mutations that abnormally activate the cytokine receptor/Janus Kinase 2 (JAK2) pathway and their downstream effectors, STATs. JAK2, calreticulin (CALR), and the myeloproliferative leukemia virus gene (MPL) are the most recurrently mutated genes in MPN patients: a somatic driver mutation in one of the three genes is present in the majority of cases. JAK2 is a kinase that binds to myeloid receptors such as erythropoietin receptor (EPO-R), MPL receptor, which affects platelet production (also named thrombopoietin receptor, TPO-R), and G-CSF receptor, affecting granulocyte production. Thus, JAK2 plays a major activating role in myeloid signaling. An acquired mutation JAK2 V617F, which causes constitutive activation of the JAK2 kinase and leads to abnormal signaling, is the major alteration in all three Ph-negative MPNs. It is present in ~98% of PV patients and about 50% of ET and PMF patients. However, although JAK2 mutations seem to be the phenotypic drivers in many MPN cases, clonality and other mutations apparently precedes the acquisition of the JAK2 V617F mutation. Somatic mutations in CALR exon 9 are found in 20-35% of ET and PMF cases. Furthermore, MPL mutants are found approximately in 3-4% of ET and 6-7% of PMF cases. These mutants are restricted to MPL activation, which explains why they are found only in ET and PMF, and not in PV. Overall, pathogenic driver mutations are identified in about 90% of the ET and PMF cases – the rest of the patients do not express any of the three mutations and are referred to as being “triple-negative”. Other recurrently mutated genes in MPNs include the same epigenetic regulators as frequently mutated in other myeloid clonal disorders: TET2, DNMT3A, ASXL1, and other genes participating in hematopoietic signaling pathways. In addition, mutations and somatic loss of heterozygosity in TP53 is strongly associated with leukemic transformation. When diagnosed with an MPN following the WHO diagnostic criteria, most patients are older than 60 years of age, and
survival depends on the severity of the disease. The most severe disease-related complications are thrombosis, stroke, and hemorrhage. Due to potential complications, the life expectancy is slightly reduced when compared to the general population; e.g. thrombosis occurs more frequently in younger PV patients compared to older individuals with the disease.\textsuperscript{117}

**Polycythemia vera**

PV belongs to the group of MPNs and is characterized by excessive production of mature erythrocytes. The major diagnostic criteria for PV include high hemoglobin, high hematocrit or increased red cell mass, BM morphology showing hypercellularity, and the presence of a \textit{JAK2} mutation.\textsuperscript{2} The symptoms include pruritus, fatigue, and splenomegaly. The risk of thrombosis and progression to secondary AML (sAML), as in other MPNs, is also increased.\textsuperscript{118} Traditionally, low-risk patients are treated with aspirin and phlebotomy; high-risk patients receive cytoreductive therapy. Current drug therapy is not curative or capable of preventing disease progression.\textsuperscript{96}

Clonal origin of PV has been acknowledged for decades.\textsuperscript{54} Most PV patients have gained the somatic \textit{JAK2} V617F mutation in exon 14. However, this particular mutation does not initiate PV. Instead it arises as a secondary genetic event.\textsuperscript{119} In the rare \textit{JAK2} V617F-negative PV (~2% of patients), somatic GoF mutations in exon 12 have been found.\textsuperscript{120,121} These mutations are not associated with ET or PMF.\textsuperscript{122} In approximately 50% of PV patients the \textit{JAK2} locus on chromosome 9p is affected by LOH.\textsuperscript{123}

3.2.2 Genetic predisposition to MPNs

The mutations in \textit{JAK2}, \textit{CALR}, and \textit{MPL} probably are not causative mutations but mainly drive the disease phenotype – a preexisting germline factor likely predisposes to a clonal MPN.\textsuperscript{105} Familial clustering of MPNs is well recognized; about 7% of cases involve germline predisposition.\textsuperscript{90} Familial MPN has been described with dominant-autosomal
transmission, and a shared susceptibility factor among the family members likely exists. Certain SNVs and germline alterations increase the probability of developing an MPN. For example, a common JAK2 haplotype 46/1, or ‘GGCC’ haplotype (rs10974944), is associated with an increased risk of an in cis JAK2-activating mutation. The G allele at rs10974944 increases an individual’s risk of developing an MPN 2.8-fold (population attributable risk 46%). Other variants, including a TERT SNV (rs2736100_C) and several common genetic polymorphisms in e.g. the MECOM, HBS1L-MYB, SH2B2, ATM, and CHEK2 genes, also associate with myeloid hematopoiesis activation or predispose to JAK2 V617F-positive MPNs. Furthermore, germline mutations in JAK2 and MPL have been described in triple-negative ET and PMF, and RBBP6 has been suggested as a candidate gene for MPN susceptibility. Duplication of ATG2B and GSKIP also predisposes to MPN development. In addition to the JAK2 46/1 haplotype and other aforementioned MPN-predisposing factors, certain germline JAK2 mutations are predicted to possibly precede the acquisition of the JAK2 V617F mutation in familial PV.

3.3 Myelodysplastic syndromes

MDS are a heterogeneous group of clonal BM diseases arising from the expansion of mutated HSCs. An MDS is characterized by ineffective hematopoiesis and clonal karyotypic abnormalities, manifested by the presence of peripheral blood cytopenias, dysplastic cellular morphology, and increased risk of developing AML. Most MDS cases are sporadic, and the median age at the diagnosis is older than 70 years. The incidence of MDS is about 2 in 100,000 individuals in Finland (Finnish Cancer Registry, Figure 3). MDS is treated with immunomodulatory agents and hypomethylating therapy, but all patients will eventually lose their response to therapy. The overall survival is poor, and the only potentially curative treatment option is HSC transplantation.
3.3.1 MDS pathogenesis

Genetic defects such as chromosomal aberrations (translocations, inversions, and deletions) and copy-number alterations are frequent in MDS. These abnormalities provide prognostic value;\textsuperscript{143} acquired HSC cytogenetic aberrations are one of the major risk factors. The most common abnormality in adults is deletion of chromosome 5q, followed by the loss of chromosome 7. The disorder has been associated with recurrent somatic mutations in more than 50 different genes encoding proteins of diverse functions.\textsuperscript{48} These mutations are not, however, limited to individuals with MDS or other myeloid neoplasms; they can also be detected in healthy people with normal blood counts. Nevertheless, the presence of these mutations increases the risk of developing an MDS as well as leading to higher mortality.\textsuperscript{49-51,144} Somatic mutations in \textit{SF3B1}, \textit{TET2}, \textit{DNMT3A}, and \textit{ASXL1} are the most commonly described abnormalities in MDS cases.\textsuperscript{142,145,146} MDS is defined as preleukemia since it frequently progresses to sAML: about 20-30\% of MDS patients develop AML.\textsuperscript{147,148} Hematopoiesis-disrupting mutations in nine specific genes, including four spliceosome genes, occur more frequently in MDS than \textit{de novo} AML (mutated in 60-70\% and 5-10\% cases, respectively).\textsuperscript{148,149}

3.3.2 Genetic predisposition to MDS

Despite most cases being sporadic, predisposition to MDS can be hereditary.\textsuperscript{150} One of the best-characterized MDS predisposition syndromes is familial platelet disorder with an autosomal dominant inheritance pattern caused by heterozygous germline \textit{RUNX1} mutations.\textsuperscript{85} A number of inherited BMF syndromes also predispose to both MDS and AML. These include e.g. dyskeratosis congenita, which is a disorder of telomere maintenance (30\% risk of MDS/AML), and many other syndromes such as Diamond-Blackfan anemia (20\% risk), Fanconi anemia (40\% risk), severe congenital neutropenia (20-40\% risk), Schwachman-Diamond syndrome (10-35\% risk), and Li-Fraumeni syndrome (5-7\% risk).\textsuperscript{148,151-153} In addition, individuals with e.g. a germline \textit{GATA2} mutation have a significantly increased risk of developing
MDS/AML, and the MDS phase in these patients is often characterized as chronic myelomonocytic leukemia.\textsuperscript{154} Additionally, \textit{SAMD9} and \textit{SAMD9L} mutations predispose to MDS/AML.\textsuperscript{155,156}

### 3.4 Acute myeloid leukemia

AML is one of the most aggressive hematological malignancies and also the most common malignant myeloid disorder in adults, affecting almost 200 individuals in Finland annually (Finnish Cancer Registry, \textbf{Figure 3}). The prevalence increases with age; the median age for AML patients at the time of diagnosis is about 70 years. AML is characterized by infiltration of the BM and blood by clonal, proliferative and abnormal hematopoietic progenitor cells (blasts) that fail to differentiate into mature myeloid cells. In addition to the block in differentiation, the progenitor cells acquire resistance for apoptosis and an increased proliferation rate.\textsuperscript{157,158} They compete for BM niche occupancy with normal cells and disrupt hematopoiesis.\textsuperscript{159} The replacement of normal blood cells with leukemic blasts causes cytopenias, frequent infections, bleeding, and BM failure. AML can lead to death in a few weeks if not treated.\textsuperscript{160,161} Leukemic cells can also escape from BM to the blood and further infiltrate other organs such as the lungs or the central nervous system.\textsuperscript{161} Risk factors for developing AML are e.g. exposure to ionizing radiation, cytotoxic chemotherapy (usually treated for a solid cancer; therapy-related AML), and benzene.\textsuperscript{158}

AML is diagnosed when at least 20% of the cells in the BM or PB are defined as immature blasts of myeloid origin based on morphological examination.\textsuperscript{2} AML cases were classically divided into different subtypes (M0-M7) according to the FAB system, based on the differentiation stage of leukemic cells, which was used for over two decades from the year 1976 to 2001.\textsuperscript{162-164} In the year 2002, it was replaced by the WHO classification, updated in 2008 and 2016, which takes into account cytogenetic alterations and mutations and provides better prognostic value.\textsuperscript{2,165,166}
The prognosis for adult patients is associated with genetic aberrations. The risk classification of the European LeukemiaNet (ELN), which is based on WHO classification, divides AML cases into three cytogenetic risk groups according to the karyotype: favorable, intermediate, and adverse.\textsuperscript{167} The standard treatment of intensive induction chemotherapy aims at complete remission, and if achieved, the patients of intermediate or adverse risk should receive allogeneic HSC transplantation.\textsuperscript{157} Increased knowledge on the genetic background of AML has led to the development of new therapeutic options. Various novel agents (e.g. IDH- and FLT3-inhibitors) have proved promising and showed improvement in terms of overall remission and survival.\textsuperscript{168-170} Despite advances in therapeutic development, AML remains challenging to treat because of its heterogeneity; only 30-40% of patients younger than 60 years of age survive more than 5 years. In older patients, the median overall survival is often less than one year, since they are usually unfit for intensive chemotherapy and thus are treated with lower-intensity treatment.\textsuperscript{170,171} Prognosis for relapsed patients is poor.

**Figure 3.** New myeloid disease cases in Finland in years 2012-2016. \textbf{a}) MPN, \textbf{b}) MDS, and \textbf{c}) AML cases. Finnish Cancer Registry, https://tilastot/syoparekisteri.fi/syovat, data from 2018-09-11, version 2019-02-12-002.

### 3.4.1 AML pathogenesis

AML represents a genetically heterogeneous disease. The patients are estimated to harbor fewer mutations than most individuals with other cancer types – usually a few, or only one or two additional mutations are needed for the clone to become malignant. The complexity and number of genetic
aberrations, however, tend to increase during the disease evolution. Most mutations identified in LSCs are random events that had occurred before acquiring the leukemia-initiating mutation.\textsuperscript{82} Recurrent abnormalities in \textit{de novo} AML include e.g. fusion gene forming translocations or inversions (\textit{PML-RARA}, \textit{RUNXI-RUNXT1}, \textit{MLL1}, \textit{MYH11-CBFB}) and \textit{FLT3} internal tandem duplication, which have been recognized for decades and are used as diagnostic and prognostic markers (FAB subtypes).\textsuperscript{172} Overall, the driver mutations can be different for distinct AML subtypes.\textsuperscript{82} Both targeted and genome wide sequencing within the last decade have revealed several diagnostically and prognostically informative mutations in AML, and the understanding of the genomic landscape has massively improved.\textsuperscript{173,174} Genes associated with DNA methylation (\textit{DNMT3A}, \textit{IDH1/2}, \textit{TET2}), transcription regulation and signaling activation (\textit{ASXL1}, \textit{RUNXI}, \textit{CEBPA}, \textit{FLT3}, \textit{N/KRAS}), and nucleophosmin (\textit{NPM1}) are recurrently mutated:\textsuperscript{81,175-177} about two-thirds of patients have acquired mutations in genes affecting signaling pathways, whereas mutations within epigenetic modifiers are identified in less than half of patients.\textsuperscript{178} RNA splicing factors are mutated in about 10\% of the cases.\textsuperscript{178} The co-occurrence of mutations can have an impact on overall survival.\textsuperscript{149} Intriguingly, AML with normal cytogenetics comprise almost half of all cases.\textsuperscript{179,180}

### 3.4.2 Genetic predisposition to AML

Many leukemia predisposition genes are known, and certain monogenic inherited disorders or BMF syndromes have an increased tendency towards AML transformation.\textsuperscript{84} Inherited germline mutations in e.g. \textit{GATA2}, \textit{RUNXI}, or \textit{CEBPA} predispose to AML development.\textsuperscript{28,181,182} These transcription factors are important for myeloid proliferation and differentiation. For example, a heterozygous mutation in \textit{GATA2} confers a 70\% risk of developing MDS/AML, and the progression to AML is associated with cytopenias and other somatic mutations.\textsuperscript{154,183} Germline mutation in \textit{RUNXI} causes familial platelet disorder with a risk of about 40\% of developing AML.\textsuperscript{85} Familial AML with a germline \textit{CEBPA} mutation has nearly complete penetrance for AML development through autosomal dominant inheritance.\textsuperscript{182,184} Telomere
biology disorders due to a mutation in e.g. TERC or TERT, or ribosomopathies with defects in ribosome biogenesis and function are also categorized as MDS/AML predisposition syndromes.\textsuperscript{185,186} (Table 1). Individuals with sAML often have worse prognosis, and their event-free survival is much lower than that of de novo AML patients (4.2 months vs. 15.7 months, respectively).\textsuperscript{187} In addition to the aforementioned disorders, myeloid leukemia predisposition disorders can be due to germline mutations in genes including ANKRD26, DDX41, ELANE, SAMD9, SAMD9L, and SRP72.\textsuperscript{155,156,188-190}

3.5 Hereditary bone marrow failure syndromes

Inherited bone marrow failure syndromes are a heterogeneous group of disorders characterized by BMF, cancer predisposition, and a variety of congenital anomalies. BMF is due to insufficient production of one or more major hematopoietic lineages, which leads to cytopenias. One cause for the development of BMF syndromes is mutations in telomere biology genes that lead to defects in the telomere complex. These diseases, including dyskeratosis congenita (DC), are called telomere biology disorders (TBD).\textsuperscript{191} Additionally, mutations in ribosomal genes can cause BMF syndromes such as Diamond-Blackfan anemia (DBA), Shwachman-Diamond syndrome (SDS), and cartilage hypoplasia, which are thus called ribosomopathies. Mutations in ribosomal proteins and in proteins functioning in ribosomal RNA processing cause impaired ribosome assembly and function, leading to distinct clinical phenotypes, most often BMF.\textsuperscript{186}

One of the best-recognized AML-predisposing syndromes is Fanconi anemia, an autosomal and X-linked recessive disorder. Due to hypersensitivity to DNA-cross-linking agents and thus a huge number of chromosomal abnormalities, it also predisposes to solid tumors and developmental anomalies.\textsuperscript{192} Another relatively well-known MDS/AML-predisposing syndrome is a multi-organ protean disorder called GATA2 deficiency, which is due to heterozygous germline mutations in GATA2. The mutations are transmitted with autosomal dominant inheritance, causing loss of function of the mutated allele that leads to haploinsufficiency; one normal
allele is not sufficient to produce the wild type phenotype. The disease has a broad phenotype including multi-lineage cytopenias, immunodeficiency with susceptibility to human papillomavirus and nontuberculous mycobacteria, vascular and lymphatic dysfunction, and pulmonary alveolar proteinosis (PAP). Clinical symptoms, age, and clinical severity vary markedly. Some patients may develop life-threatening infections or leukemia at younger age, whereas some remain asymptomatic for several years or decades. However, almost all patients suffer from peripheral blood monocyte, B, and NK cell cytopenias, and the majority will develop myeloid malignancy. Other inherited BMF syndromes such as severe congenital neutropenia and thrombocytopenia also increase AML risk, though the age of onset is later.

3.5.1 Dyskeratosis congenita – a BMF syndrome with leukemia predisposition

Dyskeratosis congenita (DC) is an inherited BMF and cancer predisposition syndrome caused by germline mutations in telomere biology genes. The mutations cause exceedingly short telomeres, which especially affects rapidly renewing tissues such as epithelial and blood cells. Classically, patients present with the mucocutaneous triad of abnormal skin pigmentation, nail dystrophy, and oral premalignant leukoplakia. They are predisposed to cancer, especially hematologic malignancies and squamous cell carcinomas of the head and neck. The risk of developing MDS/AML is significantly high. The severity of symptoms varies between patients and even between individuals with the same mutation. The leading cause of death and premature morbidity of DC patients is most commonly BMF because of the reduction in mature blood cells and opportunistic infections. BMF affects 80-90% of DC cases by the age of 30 years, and the only curative treatment for BMF is allogeneic HSC transplant. The incidence of DC is only 1 in 1,000,000 individuals.

DC is genetically heterogeneous and can follow an X-linked recessive, autosomal dominant, or autosomal recessive inheritance pattern. Germline mutations in more than ten telomere biology genes are recognized, which
lead to shortened telomere lengths (TL) in patients. Abnormal shortened telomeres lead to enhanced aging of cells, but in some situations can also help the cells to become malignant. The genetic basis is undetectable in about one third of DC cases.

The X-linked form of DC is caused by mutations in the X-chromosomal gene $\text{DKC1}$, which are the most frequent mutations appearing in almost one third of patients. Females with a mutation on the X-chromosome are often considered as asymptomatic mutation carriers due to X-chromosome inactivation (XCI), by which one of the two X-chromosomes, usually the defective one, is silenced to balance the expression dosage between females and males. Normally, XCI skews as women age, which causes unequal distribution of the alleles; skewing approaches 20% of those who are in their thirties and 40% of those over 60 years of age. XCI can be incomplete in humans, meaning that some genes are expressed from both the active and the inactive X-chromosomes. Female $\text{DKC1}$ mutations carriers do not usually develop DC-like symptoms due to skewed XCI.

4 Modern molecular techniques in genomic research

Nucleic acid (DNA or RNA) sequencing is a method for deciphering the exact order of nucleotides in a given molecule. Sequencing of the first human genome took $3$ billion and thirteen years, and was completed in 2003. The Human Genome Project was completed with Sanger sequencing, which is also called first-generation sequencing. The method was developed in 1975 and was used as the gold standard for sequencing for about 25 years. The demand for faster and cheaper sequencing has increased since, leading to the development of massively parallel, next-generation sequencing (NGS) methods. NGS platforms provide the possibility to quickly sequence millions of DNA fragments from a single sample. The increasing use of massively parallel NGS technology in the past decade has facilitated the finding of mutations in both research and clinical settings. The molecular basis of
leukemia is much better known than any other form of cancer, primarily because of the availability of malignant cells. Leukemia also has a relatively simple genome and karyotype compared to many other cancer types, and it does not form solid tumors. The continuation of major advances in discovering new mechanisms of tumorigenesis and new potential therapeutic targets has been made in the era of NGS. It has enabled the rapid discovery of, for example, recognizable heritable HM syndromes, and has deepened the understanding of molecular mechanisms underlying these malignancies.

4.1 Next-generation sequencing

The creation of NGS platforms has enabled affordable and fast sequencing in research and clinical laboratories. An entire human genome, which comprises three billion bases and more than 20,000 coding genes, can currently be sequenced within one day. Each base in the genome is sequenced multiple times, which provides high depth in the sequencing data. Thus, NGS can capture a broader range of mutations and does not depend on the pre-knowledge of the genetic region, unlike Sanger sequencing. The three main NGS applications include whole-genome sequencing, exome sequencing, and targeted panel sequencing.

Exome sequencing can reveal the mutational events occurring in gene-coding regions, which comprise about 1% of the genome, hence making it more cost-effective and affordable than sequencing the whole genome. Optimally, NGS can provide the correct diagnosis for patients by identifying the disease-causing mutations. Targeted sequencing of specific genes or genomic regions can be utilized in cases when the suspected disease or condition is identified and the most likely causal genes are known. Gene panels can help in making a rapid diagnosis and therapeutic decisions in many genetic disorders. RNA-sequencing (RNA-seq) technology, NGS of RNA, has improved in recent years, and gene expression studies using RNA-seq have replaced most microarray studies. Furthermore, single cell RNA-sequencing
(scRNA-seq) is one of the newest NGS methods, providing the possibility to identify the transcriptome in one single cell.214

NGS platforms have enabled the production of an enormous amount of data and information from cancer genomes. The Cancer Genome Atlas (TCGA) project started in 2006 with the goal of identifying genetic mutations responsible for cancer by using large-scale sequencing. The Pan-Cancer Atlas was published in 2018, covering 10,000 tumors from 33 types of cancer.215 Overall, the development of the NGS era has revolutionized cancer heterogeneity research and enhanced the generation of precision medicine and targeted therapies.

4.2 Genomics in leukemia research

The understanding of cancer biology has massively evolved during the last decades largely due to blood cancer research, which has provided many important insights and new treatment options. Identification of the first specific chromosomal abnormality consistently associated with a certain cancer type, the ‘Philadelphia chromosome’ in CML, was a major breakthrough. Recognition of the chromosomal translocation and fusion gene BCR-ABL1 underlying the Ph-chromosome216 by improved methods led to the development of therapeutic targeting via kinase-inhibitors. The first recombinant fusion genes were sequenced in blood cancers as well.217

AML is not a disease caused by hundreds of mutations, but only a few.82 The genomes and karyotypes of leukemia and lymphoma are relatively simple compared to many other cancer types. The first sequencing, copy number, and genome-wide expression analyses were all done in acute leukemia samples.218-220 Clonal evolution and the complexity of cancer has become clearer as high-throughput single-cell genetic methods have improved, and leukemia has had a key role in the phylogeny and branching studies. Precursor clones and their genetic lesions were first deciphered for acute leukemia.217 The theory of stem cells as cellular drivers of cancer and as key targets for therapy was endorsed in leukemia research, and the genetic and functional diversity in the stem cells
of individual patients was detected in acute leukemia. These studies led the ‘cancer stem cell’ concept and changed the way we view cellular epigenetic plasticity within a genetically homogeneous clone. In addition, many other remarkable discoveries such as stem cell transplants and combination chemotherapy have come from leukemia trials.

Conventional cytogenetics was the standard diagnostic tool for a long time. Fluorescence in situ hybridization (FISH) and microarray-based techniques were utilized as well, but they only allow the identification of large chromosomal abnormalities. Since the improvement of NGS technologies, candidate gene sequencing and direct sequencing have enabled the discovery of many mutations in AML patients. Whole-exome or genome sequencing provides more insights into the origin and evolution of AML mutations. The first remarkable whole exome studies covering myeloid malignancies revealed many recurrent pathogenic mutations and led the way to further studies and findings. For example, the major mutations affecting patients with MDS, whose disease evolves from MDS to sAML, and those with de novo AML, are fairly clear. The availability of NGS platforms has also enabled individualized diagnostic evaluations and tailored treatment strategies. Besides NGS, other technologies have been developed. For example, droplet digital PCR (dd-PCR) technology provides ultrahigh sensitivity and very high precision; it can detect a mutated allele frequency as small as 0.001%. This methodology is useful in, for example, the early detection of leukemia relapse and for monitoring minimal residual disease. Overall, leukemia research has revolutionized the understanding of cancer genetics.
AIMS OF THE STUDY

This thesis concentrates on malignancies of the bone marrow, especially the germline genetics of acute myeloid leukemia, a myeloproliferative neoplasm named polycythemia vera, and a leukemia-predisposing inherited telomere biology disorder dyskeratosis congenita. PV patients have an increased risk of transforming to AML. Most DC patients develop bone marrow failure, which also remarkably predisposes to AML. The specific aims were:

1. To analyze predisposing mutations to acute myeloid leukemia in genes previously implicated in AML or solid cancer predisposition, and to evaluate the patients’ clinical phenotypes and family history with respect to the germline variant analysis.

2. To identify novel polycythemia vera candidate predisposition genes and variants in a family with four PV patients.

3. To characterize the molecular and clinical details of X-chromosomal DKC1 female mutation carriers with DC-like symptoms in a family with three affected males.
MATERIALS AND METHODS

1 Study subjects

Studies I and II concerned germline exome variant analysis of Finnish AML and PV patients, respectively. Skin biopsies from AML patients, and formalin-fixed paraffin embedded (FFPE) blocks, buccal swab samples, or peripheral blood (PB) were collected from PV patients. For Study III, PB and oral tissue samples (buccal mucosa, tongue tissue) were obtained from X-chromosomal DKC1 mutation carrier females. All new diagnoses are registered in the national hematological registry (Finnish Hematology Registry, FHR). The studies were approved by the Helsinki University Hospital (HUH) Ethics Committees (#408/13/03/03/2009, #239/13/03/00/2010, #303/13/03/01/2011, and #206/13/03/03/2016) in compliance with the Declaration of Helsinki. All study participants who were alive at the time of the study gave written informed consent.

1.1 Acute myeloid leukemia patients (I)

The study material consisted of a consecutive series of primary and secondary AML patients diagnosed in the HUH region in the years 2015-2016. Written informed consent was obtained from 80/84 patients. Sixty out of the 80 AML patient samples (skin and BM) were originally collected for germline filtering of somatic exome sequencing analysis in conjunction with an AML diagnosis at the Institute for Molecular Medicine Finland (FIMM). Additional exome data was produced from eight patients for this study at the Functional Genomics Unit (FuGu). The skin and BM samples for sequencing were obtained from the Finnish Hematology Registry and Clinical Biobank (FHRB). As twelve AML patient samples were not available in FHRB, exome sequencing was not feasible in these patients. Clinical data (family history, immune deficiencies other than hematological diseases, cytopenias, and detailed laboratory characteristics) was extracted from FHR and patient records.
1.2 Polycythemia vera patients (II)

The primary study material consisted of a Finnish family with four diagnosed PV patients in two generations. Buccal swabs and PB were available from the index patient (1.1), who was diagnosed with PV at the age of 36, and with myelofibrosis at the age of 47 years. Only FFPE blocks were available from the three other family members with PV (1.2, 1.9, 1.10). The father (1.2) of the index case was diagnosed with PV at the age of 48; the aunt (1.9) was diagnosed with PV and acute leukemia at the age of 91; and the uncle (1.10) was diagnosed with PV at the age of 83. Germline DNA was available from one of two lymphoma patients (1.19) of the family, who was diagnosed with nodular lymphocyte-predominant Hodgkin lymphoma (NLPHL). The other lymphoma patient (1.6) was diagnosed with differentiated diffuse lymphocytic lymphoma at the age of 89. The daughter of the index patient was 31 years old and healthy at the time of the study. The second sample set consisted of six other Finnish families with two first-degree relative PV cases in five families, and two more distant relatives in the sixth family. FFPE blocks were available from eight of these patients. Samples were derived after signed informed consent was obtained or after authorization from the National Supervisory Authority for Welfare and Health.

1.3 Dyskeratosis congenita study subjects (III)

Two brothers were identified with a novel X-chromosomal DKC1 mutation c.1218_1219insCAG, p.(Asp406_Ser407insGln) in a Finnish family, resulting in shortened telomeres and manifestation of DC. Three female siblings, the mother, and two aunts (aunts 1 and 2) of the brothers, suffered from DC-related symptoms. They all had had early hair graying. In addition, both aunts had anemias, skin hyperpigmentation, and nail dysplasia. Clinical and laboratory status regarding mucocutaneous manifestations, teeth, complete blood counts, telomere lengths, liver, lungs, and so on, were examined. PB and tissue samples from the tongue and buccal mucosa were collected for research laboratory analyses for all the three females and one of the two diseased brothers. Aunt 1 was 59 years of age; aunt 2 was 56; and the mother of the brothers was 52 years of age at the time of examination. Informed consent was obtained from all study participants.
1.4 Population controls (I-III)

In Study I, variants were filtered against The Exome Aggregation Consortium (ExAC) dataset of 60,706 unrelated individuals, setting the MAF below 0.01 in the total population. The Sequencing Initiative Suomi (SISu) database (University of Helsinki, Finland, URL: http://sisuproject.fi [SISu v4.1, (5/2017 accessed)]) was utilized for examining the variant frequency in Finns. The population controls used in Study II were an in-house control set of 542 Finns (93 whole-genome sequenced individuals from the 1000 Genomes Project, 402 whole-genome sequenced individuals from Kuusamo, Finland (SISu), and 47 uterine leiomyoma normal controls). ExAC Finnish population data was further utilized for variant filtration, setting the MAF below 0.001. In Study III, the relative telomere length of the peripheral blood leukocyte samples from DC patients was compared to 143 normal controls.

2 Sample processing

2.1 DNA and RNA extraction (I-III)

Genomic DNA from skin biopsies was extracted with the DNeasy Blood and Tissue kit according to the manufacturer’s protocol (Qiagen) (I).

DNA was extracted from FFPE blocks with a standard phenol-chloroform method or with the NucleoSpin DNA FFPE XS kit (Macherey-Nagel), and from the buccal swab sample with the QIAmp DNA Mini kit (Qiagen). DNA from the blood sample was extracted with the standard non-enzymatic TKM buffer-proteinase K method (II) or with the Nucleospin DNA Blood XL kit (Macherey-Nagel) (III). DNA for telomere length analysis (III) was extracted from PB white cells with standard protocols as previously described.

RNA was extracted from whole blood samples with the NucleoSpin RNA Blood kit (Macherey-Nagel), and reverse-transcribed into complementary DNA (cDNA) with Promega M-MLV Reverse Transcriptase (Thermo Scientific) according to the manufacturers’ protocols (II, III). Oral tissues
(buccal mucosa and tongue) were stored in RNAlater solution (Invitrogen, Thermo Fisher Scientific) until RNA extraction. RNA from oral tissues and EBV-transformed lymphocyte cells was extracted using the NucleoSpin RNA set for Nucleozol kit (Macherey-Nagel).

3 Cell lines

3.1 EBV-transformation and culture of lymphocytes (III)

PB from patients was collected in EDTA tubes. Peripheral blood mononuclear cells (PBMCs) were extracted with Ficoll density gradient centrifugation (Histopaque 1077; MP Biomedicals). The cells were infected with Epstein-Barr virus (EBV) for immortalization and Sandimmun (Novartis, 50μg/ml) was added as an immunosuppressant. Suspended cells were cultured in RPMI 1640 (Lonza) with 20% heat-inactivated fetal bovine serum (FBS), 5% Glutamax (Gibco), and 5% antibiotics (penicillin and streptomycin) in 37°C, 5% CO₂.

3.2 Commercial cell lines (III)

SET-2 essential thrombocytemia cell line (DSMZ, ACC 608) was used as a control in the ddPCR experiments. SET-2 cells were cultured under the same conditions as the EBV-transformed patient-derived lymphocytes.

4 Next generation sequencing (NGS)

4.1 Exome sequencing (I, II)

In Study I, the exome data were produced at FIMM or FuGu, and in Study II at Karolinska Institutet, Stockholm, Sweden. Genomic DNA libraries were prepared at FIMM as described earlier²³⁴ and in FuGu with the KAPA Hyper Prep kit (Roche). In Study II, the libraries were prepared at Karolinska Institutet using the NEBNext DNA Library Prep Reagent Set for Illumina (New England Biolabs Ltd.), and exonic regions were enriched using the
Agilent Sure SelectXT Human All Exon V4+UTRs 50Mb kit (Agilent Technologies). Paired-end short read sequencing at Karolinska Institutet was performed on the HiSeq 2000 (Illumina) sequencer. Exomes produced at FIMM were captured using the Nimblegen SeqCap EZ v2 (Roche NimbleGen), Agilent SureSelect v5 Exome or Agilent SureSelect XT Clinical Research Exome (Agilent) capture kits and were sequenced using the HiSeq 1500 or 2500 instruments (Illumina). Exomes produced at FuGu were captured with MedExome kit (Roche) and sequencing was conducted using a HiSeq sequencer (Illumina).

4.2 Whole-genome sequencing (II)

The DNA library preparation for whole-genome sequencing was performed with the KAPA Hyper Prep Kit (KAPA Biosystems) and paired-end short read sequencing was accomplished with a HiSeq 4000 (Illumina) at Karolinska Institutet.

4.3 Variant calling (I, II)

After sequencing, the processed and filtered reads were aligned to the human reference genome GRCh37. Paired reads were used for variant calling with SAMtools mpileup v0.1.19 (http://htslib.org/) or Genome Analysis Toolkit (GATK) HaplotypeCaller (https://software.broadinstitute.org/gatk/documentation/tooldocs/current/org_broadinstitute_gatk_tools_walkers_haplotypecaller_HaplotypeCaller.php). The variants in the exome and genome sequencing data were visually analyzed with an analysis and visualization tool developed in-house called BasePlayer.\textsuperscript{235} Requirements for calling a variant were a minimum coverage of four reads and the mutated allele present in at least 20% of the reads. Of the intronic variants, only those located at splice sites (+/- 1-2 nucleotides from exon boundaries) were included in the initial variant lists.

4.4 Gene panels (I, III)

In study I, thirty-four genes were screened for AML germline exon mutations with BasePlayer.\textsuperscript{235} Panel A represented known leukemia predisposing
genes, and Panel B included genes implicated in DNA repair and solid tumor predisposition (Figure 4).

<table>
<thead>
<tr>
<th>Panel A</th>
<th>Panel B</th>
</tr>
</thead>
<tbody>
<tr>
<td>CEBPA</td>
<td>ATM</td>
</tr>
<tr>
<td>DDX41</td>
<td>MSH2</td>
</tr>
<tr>
<td>DKC1</td>
<td>BLM</td>
</tr>
<tr>
<td>ELANE</td>
<td>BRCA1</td>
</tr>
<tr>
<td>FANCA-E</td>
<td>BRCA2</td>
</tr>
<tr>
<td>GATA2</td>
<td>CHEK1</td>
</tr>
<tr>
<td>RPS19</td>
<td>RAD52</td>
</tr>
<tr>
<td>RTEL1</td>
<td>CHEK2</td>
</tr>
<tr>
<td></td>
<td>RECQL4</td>
</tr>
<tr>
<td></td>
<td>MLH1</td>
</tr>
<tr>
<td></td>
<td>WRN</td>
</tr>
</tbody>
</table>

Figure 4. Gene panels representing known leukemia predisposing genes (Panel A), and genes implicated in DNA repair (Panel B) used in Study I.

In study III, a clinically validated in-house NGS gene panel was utilized to detect somatic variants in genes typically seen in myeloid neoplasms for deciphering the potential development of clonal hematopoiesis in two females suffering from anemia. The panel includes exons and exon-intron boundaries and mutational hotspots of the genes presented in Supplementary Table 1.

5 Sanger sequencing validation (I-III)

Direct Sanger sequencing was utilized for germline variant validation and screening. DNA and cDNA samples were amplified with polymerase chain reaction (PCR) using standard protocols. The PCR products were purified with A’SAP (ArcticZymes) enzymes. The ABI BigDye Terminator 3.1 Cycle sequencing kit (Applied Biosystems) was used in the sequencing, and capillary electrophoresis was performed on an ABI3730xl DNA Analyzer (Applied Biosystems) at FIMM. Sequence histograms were analyzed with FinchTV (Geospiza).

6 Telomere length analysis (III)

DNA extracted from PB white cells was analyzed by the quantitative PCR method as previously described. Relative telomere length (RTL) for the
brothers was measured previously. Each DNA sample was analyzed in triplicate in separate telomere (T) and single copy gene (S) reactions using an ABI 7900HT instrument (Applied Biosystems). T/S values were calculated as \(2^{-\Delta Ct}\), where \(\Delta Ct = Ct_T - Ct_S\). The RTL value was generated by dividing the T/S value of the sample with the T/S value of DNA from a reference cell line (CCRF-CEM). The RTL value of each sample was plotted against the individual’s age and compared to 143 normal controls (age 0-83 years).

7 Droplet digital PCR (III)

The probes were designed to detect the \(DKC1\) mutated allele expression (c.1218_1219insCAG; FAM fluorescence signal) and the wild type (WT; HEX fluorescence signal) reference. Amplifications were performed in a 20 μl reaction containing 10 μl of 2X ddPCR Super Mix for probes (No dUTP, Bio-Rad Laboratories), which consisted of 1 μl of each target and reference amplification primer/probe mix (final concentration 900 nmol primers/250 nmol probe, respectively), 2 μl of cDNA template (40 ng of initial RNA), and 6 μl of sterile distilled water. Droplets were generated using the Automatic Droplet Generator QX200 (Bio-Rad Laboratories). The PCR amplification cycles were as follows: 1 cycle of 95°C (2°C/s ramp) for 10 min, 40 cycles of 94°C (2°C/s ramp) for 30 sec and 55°C (2°C/s ramp) for 1 min, followed by 1 cycle of 98°C (2°C/s ramp) for 10 min. The samples were held at 4°C until further processing. Amplified droplets were read with a QX200 droplet reader (Bio-Rad Laboratories). QuantaSoft Analysis Pro software v.1.0.596 (Bio-Rad Laboratories) was used for data visualization and wells with less than 8000 droplets were not examined. SET-2 cell line was used as a mutant-negative control and the cDNA of Study III index case’s brother was used as a mutant-positive control.

8 Online databases and in silico variant prediction tools (I, II)

In Studies I and II, total and Finnish population data in ExAC\(^2\) were utilized for MAF inspection, respectively. In Study I, MAF in the Finnish population
was checked in the SISu database. The effects of variants in Study I were evaluated according to the ACMG/AMP guidelines\textsuperscript{238} with several online \textit{in silico} prediction tools, PolyPhen-2,\textsuperscript{239} SIFT,\textsuperscript{240} and ClinVar:\textsuperscript{241} pathogenic or conflicting missense variants, nonsense variants, small deletions, and splice-site variants were included in the analysis. Pathogenicity of the variants was also predicted with Intervar (http://wintervar.wglab.org/), the Rare Exome Variant Ensemble Learner (REVEL),\textsuperscript{242} and the Human Gene Mutation Database (HGMD).\textsuperscript{243} In Study II, the effect of variants was predicted with PolyPhen-2 and SIFT.

\textbf{9 Statistical analyses (I, III)}

In Study I, statistical analyses were performed using the R environment (http://www.r-project.org/). Statistical differences in the mean age of the variant carriers and non-carriers was determined with Student’s \( t \) test. Fisher’s exact test was utilized for the calculation of Odds Ratio (OR) and 95\% confidence intervals for positive family history of malignancies of variant carriers vs. non-carriers. A \( P \)-value of \(<0.05\) was considered statistically significant. In Study III, the QuantaSoft Analysis Pro software was utilized for calculating the positive and negative droplets for each fluorophore (mutant-positive and negative, respectively), fitting the positive droplets to a Poisson distribution, and determining the original concentration in copies/\( \mu \)L of input sample. The 95\% confidence intervals were used.
RESULTS

1 Identification of germline alterations in AML patients (I)

1.1 Variants of unknown significance

Germline exome sequencing was conducted in 68/80 Finnish AML patients (85%). The study flow is presented in Figure 5. Gene panels A and B consisted of 34 genes implicated in AML (A), and DNA repair and solid tumors (B). A systematic analysis of the germline exomes revealed 34 rare single-nucleotide variants and short insertions or deletions (indels) in panel genes predicted to be functionally adverse \textit{in silico} in 42/68 (62%) of the patients (Table 2). The variants identified in the study were detected mostly in the genes from panel B, representing DNA repair genes. Variants of unknown significance (VUS) were found in the 16 following genes: \textit{ATM}, \textit{BLM}, \textit{BRCA1}, \textit{CHEK1}, \textit{CHEK2}, \textit{DDX41}, \textit{FANCA}, \textit{FANCE}, \textit{GATA2}, \textit{MLH1}, \textit{PALB2}, \textit{PMS2}, \textit{RTEL1}, \textit{SAMD9L}, \textit{SBDS}, and \textit{WRN} (Figure 5).

1.2 Germline variants with adverse or potentially pathogenic function

The frequency of likely pathogenic germline variants in our series is 9% (95% Jeffreys credible interval 4–17%). Two of the 34 variants (6%) were previously reported as AML-predisposing mutations: a start-lost mutation c.3G>A (p.Met1Ile)\textsuperscript{244} in \textit{DDX41}, and a homozygous splice-site mutation (c.238 + 2T>C)\textsuperscript{245} in \textit{SBDS}. The start-lost mutation in \textit{DDX41} is categorized as a risk factor for AML. The patient with the \textit{DDX41} mutation was 64 years old and did not have any family history of cancer. The splice-site mutation in \textit{SBDS} is considered to be a risk factor as well. The \textit{SBDS} mutation carrier was
a 70-year-old male diagnosed with MDS one year prior to AML but had no history of known symptoms linked to SDS, which usually develops due to SBDS mutations. He carried monosomy 7 in his BM cells.

**Figure 5.** Flowchart describing the AML study and analysis, and the number of patients carrying variants in 16 genes. An asterisk denotes a heterozygous and homozygous case. Adapted and reproduced with permission from Wartiovaara-Kauto et al., Leukemia, 2018.
Table 2. Variants detected in gene panels A and B in the germline exomes of 68 AML patients. Previously reported risk factors in SBDS (c.258+2T>C) and DDX41 (p.Met1lle start lost) are colored with light red. Novel candidate variants in CHEK1, SAMD9L, and DDX41 VUS are colored with grey. Homozygosity is marked with an asterisk. Genome assembly GRCh37.

<table>
<thead>
<tr>
<th>Patient number</th>
<th>Gender</th>
<th>Gene</th>
<th>Location</th>
<th>Variant</th>
<th>MAF ExAC/SBSa (total/Finnis)</th>
</tr>
</thead>
<tbody>
<tr>
<td>230C</td>
<td>F</td>
<td>CHEK2</td>
<td>c.1229delG</td>
<td>0.0082/0.0087</td>
<td></td>
</tr>
<tr>
<td>3626M</td>
<td>M</td>
<td>CHEK3</td>
<td>c.470T&gt;C</td>
<td>0.0041/0.0043</td>
<td></td>
</tr>
<tr>
<td>4257M</td>
<td>F</td>
<td>FANCc</td>
<td>16:89625470G&gt;T</td>
<td>0.00138/0.00015</td>
<td></td>
</tr>
<tr>
<td>4257M</td>
<td>M</td>
<td>SBDS*</td>
<td>7:664591977C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>431M</td>
<td>M</td>
<td>FANCE</td>
<td>6:33427531T&gt;C</td>
<td>0.0028/0.00005</td>
<td></td>
</tr>
<tr>
<td>4336F</td>
<td>F</td>
<td>FANCE</td>
<td>16:89671796G&gt;A</td>
<td>0.00557/0.00557</td>
<td></td>
</tr>
<tr>
<td>454F</td>
<td>F</td>
<td>MLH1</td>
<td>3:27059060A&gt;G</td>
<td>0.00017/0.00017</td>
<td></td>
</tr>
<tr>
<td>454F</td>
<td>F</td>
<td>PALB2</td>
<td>16:23614979C&gt;T</td>
<td>0.00897/0.00897</td>
<td></td>
</tr>
<tr>
<td>458M</td>
<td>M</td>
<td>PALB2</td>
<td>16:99805351G&gt;A</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>458M</td>
<td>M</td>
<td>CHEK2</td>
<td>22:29121867G&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>473R</td>
<td>M</td>
<td>WRN</td>
<td>8:31812183C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>465M</td>
<td>M</td>
<td>GATA2</td>
<td>3:128204731C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>468M</td>
<td>M</td>
<td>FANCA</td>
<td>16:89671796G&gt;A</td>
<td>0.00557/0.00557</td>
<td></td>
</tr>
<tr>
<td>479F</td>
<td>F</td>
<td>DDX41</td>
<td>1:517646797G&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>479F</td>
<td>F</td>
<td>FANCE</td>
<td>6:33423528C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>479F</td>
<td>F</td>
<td>FANCE</td>
<td>6:33427531T&gt;C</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>476M</td>
<td>M</td>
<td>PM52</td>
<td>7:60456347C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>477F</td>
<td>F</td>
<td>CHEK2</td>
<td>22:29121867G&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>477F</td>
<td>F</td>
<td>RETEL1</td>
<td>20:002442000C&gt;CCA</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>480M</td>
<td>M</td>
<td>SBDS</td>
<td>15:9194473C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>489F</td>
<td>F</td>
<td>FANCE</td>
<td>16:89639510G&gt;C</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>496F</td>
<td>F</td>
<td>FANCE</td>
<td>16:89691666G&gt;A</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>498F</td>
<td>F</td>
<td>GATA2</td>
<td>3:128204731C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>498F</td>
<td>F</td>
<td>PM52</td>
<td>7:60262866G&gt;C</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>499M</td>
<td>M</td>
<td>HBC1A</td>
<td>17:41254835G&gt;A</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>500M</td>
<td>M</td>
<td>PM52</td>
<td>7:60456347C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>503F</td>
<td>F</td>
<td>FANCE</td>
<td>16:89671796G&gt;A</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>512M</td>
<td>M</td>
<td>DDX41</td>
<td>1:517646797G&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>520F</td>
<td>F</td>
<td>PM52</td>
<td>7:60456347C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>522M</td>
<td>M</td>
<td>CHEK1</td>
<td>11:125495891G&gt;A</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>523M</td>
<td>M</td>
<td>FANCE</td>
<td>6:33427531T&gt;C</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>523M</td>
<td>M</td>
<td>FANCE</td>
<td>16:89671796G&gt;A</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>532M</td>
<td>M</td>
<td>CHEK3</td>
<td>22:29121867G&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>532M</td>
<td>M</td>
<td>ATM</td>
<td>11:128451976C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>535M</td>
<td>M</td>
<td>GATA2</td>
<td>3:128204731C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>535M</td>
<td>M</td>
<td>MLH1</td>
<td>3:117056357G&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>576F</td>
<td>F</td>
<td>FANCE</td>
<td>10:89895051C&gt;C</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>577F</td>
<td>F</td>
<td>FANCE</td>
<td>6:33423528C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>580F</td>
<td>F</td>
<td>RETEL1</td>
<td>20:025242000C&gt;CCA</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>580F</td>
<td>F</td>
<td>HBC1A</td>
<td>17:41254835G&gt;A</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>591F</td>
<td>F</td>
<td>PM52</td>
<td>7:60456347C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>591F</td>
<td>F</td>
<td>FANCE</td>
<td>16:89671796G&gt;A</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>594F</td>
<td>F</td>
<td>PM52</td>
<td>7:60456347C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>594F</td>
<td>F</td>
<td>PM52</td>
<td>7:60456347C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>594F</td>
<td>F</td>
<td>PM52</td>
<td>7:60456347C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>594F</td>
<td>F</td>
<td>PM52</td>
<td>7:60456347C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
<tr>
<td>594F</td>
<td>F</td>
<td>PM52</td>
<td>7:60456347C&gt;T</td>
<td>0.00974/0.00974</td>
<td></td>
</tr>
</tbody>
</table>
We identified a VUS in \textit{SAMD9L}, c.3649T>C (p.Phe1217Leu). The allele frequency in Finns is 0.00048 and it is even rarer in other populations. The variant carrier suffered from acute promyelocytic leukemia (APL). He had a positive family history of malignancies and also carries variants in the \textit{CHEK1}, \textit{PMS2}, and \textit{FANCA} genes. The \textit{CHEK1} variant c.77A>G (p.Asn26Ser) in his germline is of unknown significance, and the allele frequency in Finns is 0.00976. Two patients were identified to carry c.236G>A (p.Trp79Ter/UTR) in \textit{CHEK1}, which leads to a termination codon depending on the transcript. In addition, one patient had a previously unreported rare variant in an AML-associated gene \textit{DDX41}, c.95T>G (p.Asp32Ala/Asp50Ala), with varying prediction depending on the transcript as well. The sequencing chromatograms of the four variants are depicted in Figure 6.

\textbf{Figure 6.} Sequencing chromatograms of germline variants of unknown significance in \textit{CHEK1}, \textit{SAMD9L}, and \textit{DDX41}. \textbf{a)} \textit{CHEK1} c.77A>G (p.Asn26Ser), \textbf{b)} \textit{CHEK1} c.236G>A (p.Trp79Ter/UTR), \textbf{c)} \textit{SAMD9L} c.3649T>C (p.Phe1217Leu), and \textbf{d)} \textit{DDX41} c.95T>G (p.Asp32Ala/Asp50Ala).
**1.3 Clinical and family history examination in conjunction with germline exome sequencing**

Full clinical history and detailed information on disease characteristics were obtained from all patients. A self-reported family history of malignancies was collected from 62/80 (78%) patients. Thirty-seven out of the 62 patients (60%) had a first- or second-degree family member with a malignancy, of which eight (13%) had a hematological malignancy. The variant carriers more often had a positive family history of malignancies compared with non-carriers (OR=3.2, 95% CI [1.2, 8.5], \( P=0.01 \)). The difference between the mean age of 63 years in variant carriers vs. 59 years in non-carriers was not statistically significant (Student’s \( t \) = -1.08, \( P=0.3 \)). Eight of the patients with exomes available for analysis had a solid tumor malignancy prior to AML, and all of them carried germline variants. Only some of these patients had received chemotherapy or radiation therapy. Two-thirds of the patients with a previous history of solid cancer or hematological malignancy and germline variants were deceased by the time of analysis.

**2 Exome sequencing analysis of familial PV (II)**

**2.1 Germline variants in ZXDC, ATN1, and LRRC3**

Germline exome sequencing was performed for three PV patients (1.1, 1.2, 1.9) in a Finnish family with four patients diagnosed with PV (Figure 7). The exome analysis detected 12 shared variants with MAF < 0.001 in the Finnish population (ExAC), which were predicted to be damaging in silico (Table 3). The variants detected in the Finnish controls (\( N=542 \)) were removed (MAF < 0.2%, 95% CI [0, 0.05%]). The 12 variants (11 missense and one
splice site variants) included in the analysis were further validated in a fourth affected family member (1.10) by Sanger sequencing. From these variants predicted as damaging, 1.10 carried three rare SNVs: c.1254C>G (p.Phe418Leu) in ZXDC; c.1931C>G (p.Pro644Arg) in ATN1; and rs148872771, c.701G>A (p.Arg234Gln) in LRRC3 (Figure 8). One of the two lymphoma patients in the family (1.19) was identified to carry the variant in LRRC3. The three variants were checked in the germlines of eight PV patients from six other Finnish families, but the variants were not observed. In addition to the possibly damaging variants, a rare SNP rs144332650, c.2912C>G (p.Ala971Gly) predicted as benign in an X-chromosomal gene BCORL1 was identified in all four PV patients in the family. PV patients in six other families did not carry this variant.

Figure 7. Pedigree of the Finnish family with four polycythemia vera patients and two patients with lymphoma. Family members with exomes available are marked with an asterisk and the individual used for validation with a small black square. Two asterisks denote both exome (germline) and whole genome (peripheral blood DNA) data available. Reproduced with permission from Hirvonen et al., Hum Genomics, 2017.
Table 3. Identified germline variants shared in PV patients of the family. The three variants shared by all four PV patients are colored with grey. Adapted and reproduced with permission from Hirvonen et al., Hum Genomics, 2017. Genome assembly GRCh37.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Genomic location</th>
<th>Variation, cDNA</th>
<th>Variation, protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>ZXDC</td>
<td>Chr3: 126189754</td>
<td>c.1254C&gt;G</td>
<td>p.F418L</td>
</tr>
<tr>
<td>ATN1</td>
<td>Chr12: 7046361</td>
<td>c.1931C&gt;G</td>
<td>p.P644R</td>
</tr>
<tr>
<td>LRR3C</td>
<td>Chr21: 45877228</td>
<td>c.701G&gt;A, rs148872771</td>
<td>p.R234Q</td>
</tr>
<tr>
<td>GNL3</td>
<td>Chr3: 52727477</td>
<td>c.1241A&gt;G</td>
<td>p.Y414C</td>
</tr>
<tr>
<td>MDC1</td>
<td>Chr6: 30679188</td>
<td>c.2221+1G&gt;T</td>
<td>spl1</td>
</tr>
<tr>
<td>ITPR3</td>
<td>Chr6: 33635026</td>
<td>c.1672C&gt;T, rs780906252</td>
<td>p.R558C</td>
</tr>
<tr>
<td>FAM135A</td>
<td>Chr6: 71190668</td>
<td>c.1221G&gt;A, rs143901584</td>
<td>p.V203M</td>
</tr>
<tr>
<td>SLC2A12</td>
<td>Chr6: 134312391</td>
<td>c.1756C&gt;T, rs200847615</td>
<td>p.P586S</td>
</tr>
<tr>
<td>WDR86</td>
<td>Chr7: 151097265</td>
<td>c.226G&gt;A, rs199824863</td>
<td>p.D76N</td>
</tr>
<tr>
<td>CSMD1</td>
<td>Chr8: 3165238</td>
<td>c.3929C&gt;T</td>
<td>p.A1310V</td>
</tr>
<tr>
<td>SLC24A2</td>
<td>Chr9: 19786283</td>
<td>c.582A&gt;G, rs368590535</td>
<td>p.I194M</td>
</tr>
<tr>
<td>ITPKC</td>
<td>Chr19: 41224132</td>
<td>c.1092C&gt;G, rs143757004</td>
<td>p.D364E</td>
</tr>
</tbody>
</table>

Figure 8. Sequencing chromatograms of novel PV predisposition variants a) ZXDC c.1254C>G (p.Phe418Leu), b) ATN1 c.1931C>G (p.Pro644Arg), and c) LRRC3 c.701G>A (p.Arg234Gln).
2.2 Screening of frequent germline alterations in PV patients

A predisposition allele in *JAK2*, rs10974944, was identified in all the PV patients in the family. Patients 1.1, 1.2, and 1.10 were homozygous for the risk variant in their germline (GG genotype), whereas 1.9 was heterozygous (CG). All eight PV patients from the six other families also carried the SNP: two of them were heterozygous, and six were homozygous. An MPN-predisposing variant in *TERT*, rs2736100, was also checked, revealing homozygosity in all the PV patients in the studied family (CC genotype). Other MPN-predisposing mutations such as duplication of *ATG2B* and *GSKIP* or mutations in *RBBP6* were not identified.

2.3 Detection of somatic variants and LOH

Peripheral blood genomic DNA of the index case was sequenced. The most frequent somatic mutation in PV, *JAK2*V617F, was identified in the index patient’s blood sample, showing LOH, which was also studied in the 12 damaging gene variants detected by exome sequencing. Only the germline variants c.582A>G (p.Ile194Met) in *SLC24A2* and c.3929C>T (p.Ala1310Val) in *CSMD1* showed clear LOH in the index case’s blood sample. The index case carried two missense variants predicted as possibly damaging, c.680C>T (p.Thr227Met) in *FLT3* and c.5162 T>G (p.Leu1721Trp) in *TET2*, which are known MPN-associated genes. In addition, we identified the possibly damaging missense variants c.3263C>T (p.Ser1088Phe) and c.1235C>T (p.Alac412Val) in *FANCA*. Sanger sequencing of the cDNA of the index case identified the expression of the X-chromosomal *BCORL1* variant.
3 Female carriers of germline X-chromosomal \textit{DKC1} insertion (III)

3.1 Mutant allele burden and XCI in female \textit{DKC1} mutation carriers

A family with two diseased brothers with an insertion c.1218_1219insCAG in the X-chromosomal gene \textit{DKC1} and three females (mother and aunts 1 and 2) with DC-like symptoms was investigated. By sequencing DNA from peripheral blood, we confirmed all females to be heterozygous \textit{DKC1} mutation carriers (Figure 9). \textit{DKC1} gene expression levels and distribution of mutated and wild type alleles were studied with ddPCR in blood, buccal mucosa, tongue, and EBV-transformed lymphocytes to study X-inactivation in different tissues.

\textbf{Figure 9.} Sequencing chromatograms of aunt 1 showing the heterozygous \textit{DKC1} mutation (DNA). The result was same in all three females. The mutated allele was expressed at the RNA/cDNA level (aunt 1, buccal mucosa). Adapted and reproduced with permission from Hirvonen \textit{et al.}, Leukemia, 2019.
The allele burden varied depending on the individual and tissue (Figure 10): the mutant allele burden in blood was 0% in the mother, 45% in aunt 1, and 5% in aunt 2. This could not be seen with Sanger sequencing in all the samples (not shown). In buccal mucosa samples, the fractions were 37%, 45%, and 7%, respectively. Additionally, tongue samples showed varying amounts of mutant *DKC1* allele expression: 14% in the mother, 28% in aunt 1, and 8% in aunt 2. The EBV-transformed lymphocytes showed extremely high overall *DKC1* expression (>98% WT *DKC1*, not shown).

![Figure 10](image.png)

**Figure 10.** ddPCR analysis showed varying expression of mutant (blue) and wild type (green) *DKC1* expression in different tissues of the a) mother, b) aunt1, and c) aunt 2. Concentration is presented as copies of input cDNA per μL. Adapted and reproduced with permission from Hirvonen et al., Leukemia, 2019.

3.2 Telomere length and clinical features of the females

Telomere lengths were measured using a quantitative PCR-based method that showed shortened telomeres in two out of the three females (the mother and aunt 1) compared to age-matched controls. The telomeres in the two brothers were previously shown to be very short (<5th percentile). Aunt 2 had normal telomere lengths (Figure 11).
Figure 11. Relationship between age and relative telomere lengths (RTL) of blood cells in DC males and DKCI mutation carrier females. Both males and two of the females (mother and aunt 1) had shortened telomere lengths. Adapted and reproduced with permission from Hirvonen et al., Leukemia, 2019.

The clinical and laboratory status regarding the wide spectrum of potential DC symptoms was examined (Figure 12). Mucocutaneous and dental problems as well as hair, skin, and nail abnormalities are potential signs of DC. Aunt 1 had suffered from aggressive periodontitis and due to this had lost all her teeth by the age of 52 years. The mother and aunt 2 had no dental diseases reported that associate with DC. In addition, none of the studied females presented with oral mucosal abnormalities. Aunt 1 with shortened telomeres showed nail problems, skin pigmentation, and hair abnormalities compatible with DC. She had onset of premature graying of the hair already in her teenage years. Aunt 2 also showed abnormalities compatible with DC: her nails were abnormal.
and she had mild skin hyperpigmentation and hyperkeratosis. She had premature graying of the hair at 30 years of age. The mother of the brothers also had onset of premature graying of the hair in her teenage years. She had shortened telomeres in blood cells but did not show classical DC-associated skin abnormalities.

DC also affects the cells forming in the bone marrow. Both aunts suffered from anemia. Bone marrow examination was performed for aunt 2, showing decreased cellularity without signs of dysplasia or an excess of blasts. To decipher the potential development of clonal hematopoiesis in the two aunts, an in-house NGS panel was analyzed on their peripheral blood to detect somatic variants in genes involved in myeloid neoplasms. No pathogenic mutations were detected.

Figure 12. Pedigree and clinical features of c.1218_1219insCAG mutation carrier females. Circles with dots denote mutation carrier females and dark squares represent males with DC. The index case is marked with an arrow, and the ages at the time of examination are shown. Adapted and reproduced with permission from Hirvonen et al., Leukemia, 2019.
DISCUSSION

Most cancer types show familial aggregation,\textsuperscript{246,247} which together with early age of onset, multiple primary cancers, and specific non-malignant manifestations represent clinical signs of hereditary cancer syndromes.\textsuperscript{248} Up to 10\% of cancers show familial inheritance,\textsuperscript{67} and while this is a small fraction of the overall cancer burden, the molecular genetic findings in these cases have changed family counseling and management of the disease; these findings have also shed light on the regulatory pathways playing important roles in sporadic tumor development.\textsuperscript{248} Some of the known germline and tumor mutated cancer genes overlap: a fraction of somatically mutated cancer genes confer susceptibility to cancer if mutated in the germline, and a number of the germline mutated cancer genes can contribute to cancer development when mutations occur in the tumor.\textsuperscript{66} This has also been detected in myeloid malignancies with either germline or somatic mutations in e.g. $\textit{RUNX1}$ and $\textit{GATA2}$: the genes are recurrently mutated in AML as secondary events, but if mutated in the germline, they remarkably predispose to myeloid malignancies. Hematological malignancies with inherited predisposition have been recognized for decades,\textsuperscript{1} but the molecular basis has only recently become clearer. Individuals with germline predisposition to hematological malignancies are diagnosed at an accelerating pace. However, individuals with hereditary HM still continue to be underdiagnosed, and in a number of familial myeloid neoplasia cases, the germline cause is unknown. In this thesis, we studied the germline alterations that potentially predispose to myeloid malignancies. We aimed to analyze germline mutations in AML in an unselected series of patients; identify novel candidate susceptibility genes for familial PV, which also predisposes to AML; and characterize X-linked DC in mutation carrier females at potential risk for a myeloid malignancy. Telomere biology disorders are known to predispose to hematological disorders due to bone marrow defects.
Germline alterations in AML patients

Germline predisposition to myeloid malignancies has long been recognized in children, and inherited forms of leukemia were long thought of as solely pediatric diseases. Today, familial clustering of AML is well recognized, and many individuals with a predisposing genetic alteration can develop MDS or AML later in adulthood. Several inherited BMF syndromes and gene mutations have been associated with MDS and AML, including \textit{GATA2}\textsuperscript{,28}, \textit{CEBPA}\textsuperscript{,182}, \textit{RUNX1}\textsuperscript{,85} and \textit{DDX41} mutations.\textsuperscript{249} Furthermore, germline mutations in myeloid malignancies have also been identified in genes classically implicated in solid tumor predisposition genes such as DNA repair genes.\textsuperscript{250,251} For example, Fanconi anemia, a rare primarily autosomal recessive cancer-prone syndrome, is due to mutations in several different genes that encode proteins interacting in a DNA repair pathway known as the FA pathway. It comprises 19 FA proteins and many other associated proteins.\textsuperscript{252,253} The FA pathway intersects with other repair processes and operates in repairing interstrand crosslink lesions in DNA.\textsuperscript{254} In patients with biallelic mutations in FA complex genes, DNA damage is not repaired efficiently, resulting in chromosomal aberrations and uncontrolled cell growth or inability to make new DNA molecules. In particular, quickly dividing cells, such as BM cells, are affected. As a consequence, the individuals develop AML, MDS, aplastic anemia, and solid tumors, which are major complications identified in patients.\textsuperscript{252,254} Up to ninety percent of FA patients present with BMF.\textsuperscript{255} Thus, the role of DNA repair pathways is major in genome maintenance and cancer development, including in AML.

The frequency of predisposing variants is unknown in an unselected, consecutive series of unrelated AML patients. Hence, we screened the germline variations in 34 leukemia and solid tumor predisposing genes in a consecutive series of AML patients. Thirty-four rare SNVs and short insertion or deletion variants with predicted functionally adverse consequences were detected in 42/68 (62%) germline exomes in 16 genes, of which most are associated with DNA repair. Only two of the variants were previously reported as AML-predisposing mutations: \textit{DDX41} (p.Met1Ile; start lost);\textsuperscript{244} and \textit{SBDS} (c.258+2T>C).\textsuperscript{245} The start lost mutation in \textit{DDX41} was identified in a 64-
year-old patient with no family history of cancer, which was intriguing. Another patient (57y), with a positive family history of cancer, was identified to carry a previously unreported rare variant p.Asp32Ala/Asp50Ala depending on the transcript, with conflicting functional prediction. DDX41 mutations usually result in myeloid neoplasia in older adults.249 The SBDS mutation was identified in a 70-year-old male with MDS one year prior to diagnosis with AML. Biallelic mutations in SBDS are associated with a multiorgan disorder called Shwachman–Diamond syndrome (SDS). Patients with SDS typically present with increased risk of BM dysfunction, MDS, and acute leukemias, but also with pancreatic insufficiency and short stature.245 A subset of individuals lack syndromic features, and thus, the syndrome may be underdiagnosed,256 which was also reflected in our AML series: the patient carried the characteristic monosomy 7 in his BM cells, but did not show any known symptoms linked to SDS.

Overall, the frequency of mutations in DNA repair genes increases due to chemotherapy or radiation therapy, which predisposes to secondary cancers. The incidence of these therapy-related and secondary cancers is increasing, since the treatment and survival rates are better after the preceding malignancy.257 All eight patients in our study with a solid tumor prior to AML carried germline variants, which were mostly identified in the DNA repair genes (Panel B). Only five of these patients had had chemotherapy or radiation therapy, which suggests the possibility of a second, not secondary, cancer. By the time of analysis, about two-thirds of the patients with a previous solid cancer or HM and with germline variants detected were deceased, which may suggest that therapy-related AML and AML with preceding MDS or MPN can develop with high risk, but the possible contribution of germline alterations should not be ignored. Further studies in other, more extensive patient series are warranted to accurately compare the clustering of germline variants in secondary and second cancers with high-risk somatic gene changes and primary AML.

1.1 Variants of uncertain significance

Current guidelines for the interpretation of sequence variants recommended by the American College of Medical Genetics and Genomics and the
Association for Molecular Pathology (ACMG/AMP) provide a useful standard for differentiation of mutations and less significant variants in human diseases. They recommend the use of standard terminology: pathogenic, likely pathogenic, uncertain significance, likely benign, and benign. Sometimes, however, a novel variant of unknown significance becomes suspicious and calls for reclassification as a likely pathogenic mutation, particularly when a patient’s personal or family history raises suspicions of inherited predisposition to a disorder.

In our study, we identified a rare variant in *SAMD9L*, predicted to be a variant of uncertain significance (p.Phe1217Leu) located in the C-terminal region of the gene in a patient diagnosed with APL. *SAMD9L* acts as a myeloid tumor suppressor, but the general function and effect on hematopoiesis is not well known. Germline mutations in tumor suppressors are known to predispose to cancer because only one additional mutational event is required for tumorigenesis, or through e.g. loss of heterozygosity or uniparental disomy. Mutations particularly in the C-terminal region of *SAMD9L* have only recently been associated with the development of MDS together with preceding cytopenia, immunodeficiency, and neurological symptoms, and the connection between heterozygous germline missense mutations and familial form of MDS has been shown. *SAMD9L* is conserved and shares 62% sequence identity with the *SAMD9* gene, which is a myeloid tumor suppressor as well. Both genes are located in the region of chromosome 7 that is commonly lost in myeloid malignancies. The development of chromosome 7 deletion consistently removes the allele with the mutated *SAMD9*/*SAMD9L*, which is termed “adaption by aneuploidy”; HSPCs that eliminate *SAMD9* or *SAMD9L* GoF mutations gain a competitive advantage. Most of the previously reported patients with a *SAMD9L* mutation carried somatic chromosome 7 abnormalities. The patient with the *SAMD9L* VUS in our study, however, lacked somatic changes in chromosome 7. The role of this rare variant in the causality of APL thus remains unknown. Intriguingly, the patient has a family history of malignancies, and in addition to the *SAMD9L* variant, also carries VUS in the DNA repair genes *CHEK1*, *FANCA*, and *PMS2*. 
Defective DNA repair is one of the hallmarks of cancer. DNA repair genes are often mutated in malignant tumors and are the genetic events most commonly involved in hereditary cancers. It is possible that the prevalence of germline mutations in different cancer types is higher than previously estimated, and the clinical implications of many DNA repair genes still remain unknown. These genes are also mentioned in the latest WHO classification concerning myeloid malignancies, which refers to the great importance of screening DNA repair germline mutations in addition to somatic defects.

Three patients in our study were identified to carry a rare VUS in the DNA repair gene $\text{CHEK1}$: one patient with a missense variant p.Asn26Ser; and two patients with p.Trp79Ter/UTR, leading to a termination codon depending on the transcript. One patient with the early termination codon mutation was relatively young at the time of diagnosis (55y), and also had family history of breast cancer. However, the other patient with the same variant had no family history of malignancies and was closer to the median age of approximately 70 years at the time of AML diagnosis (66y). The third patient with a missense variant had a positive family history of malignancies, and was only 28 years old when diagnosed with APL. None of these three patients had had previous malignancies or chemotherapy, which would have implicated the development of therapy-related AML or secondary cancer.

The highly conserved gene $\text{CHEK1}$ on chromosome 11 encodes a serine/threonine kinase CHK1, which is a key component in the DNA damage response. CHK1 regulates cell cycle checkpoints and prevents cells with DNA damage from entering mitosis, but also contributes to cell proliferation and survival. Thus, targeted therapy to CHK1 and its related signaling have become an area of great interest in oncology. CHK1 also provides a prognostic indicator of survival and a promising therapeutic target in AML; high expression of $\text{CHEK1}$ in AML cells predicts reduced overall, event-free, and relapse-free survival. $\text{CHEK1}$ mutations are rare
in cancer, as cells with defective CHK1 are eliminated during tumorigenesis. CHK1 also seems to favor cell proliferation. Nevertheless, LoF mutations have been reported in some cancer types. Earlier studies suggest that CHEK1 haploinsufficiency may promote cancer, particularly if other mutations are present in relevant genes. In our study, the patient with CHEK1 variant p.Asn26Ser carried also other VUS in the genes SAMD9L, FANCA, and PMS2. In conclusion, although we did not detect any shared clinical features for the three CHEK1 mutation carrier patients in addition to AML diagnosis, we would like to suggest CHEK1 as a novel candidate gene for AML susceptibility.

2 Candidate predisposition genes for familial PV

Familial clustering of PV is occasionally observed. A preexisting germline predisposition factor is thought to exist; however, high-penetrance predisposition genes to PV have not been clearly defined to date. PV arises from a somatic mutation in JAK2 in HSC, which results in hyper-activated signaling and massively increased production of mature erythrocytes. These JAK2 mutations are not founder mutations, but mainly drive the PV phenotype. Identification of potential predisposing gene alterations is important for families with PV susceptibility. PV can occasionally transform into other, more severe, myeloid malignancies.

Most studies have only identified candidate variants and haplotypes in familial MPN cases. The most significant finding so far has been the JAK2 46/1 haplotype, which favors the acquisition of the most frequent mutation JAK2 V617F in MPNs. The exact reason for this is not known, but it is hypothesized that the inherited haplotype may be genetically unstable, thus acquiring the somatic in cis V617F mutation faster or that the V617F clone gains a selective advantage for the 46/1 haplotype and not the wild type haplotype. Additional variants in other genes have been found, most notably in TERT, but these still do not fully explain the family clustering of PV and other MPNs.
We utilized exome sequencing to study the predisposition to PV in a Finnish family with four diagnosed PV patients. This is a rather exceptional family, as PV is very rare and most cases are sporadic. The index case was younger (36 years of age) than PV patients typically are at the time of diagnosis, which is common in familial PV cases. She progressed to myelofibrosis eleven years after PV diagnosis. Our study is the first report on Finnish familial PV cases. We identified three novel candidate variants with adverse functional predictions in **ZXDC** (p.Phe418Leu), **ATN1** (p.Pro644Arg), and **LRRC3** (p.Arg234Gln) shared by all four patients. Of these three, the transcription factor **ZXDC** appears the most relevant candidate. It regulates the transcription of genes involved in differentiation and inflammatory response in a myeloid cell, especially in monocytes.\(^\text{269}\) It cannot, however, be considered as a lineage-determining TF. **ATN1** is a nuclear transcriptional corepressor, mainly associated with neurogenerative diseases,\(^\text{270}\) and the **LRRC3** variant was identified also in one lymphoma patient of the family, indicating that it probably is not responsible for PV predisposition exclusively. We further screened the three SNVs from eight other patients in six families with PV clustering, but they did not carry these variants. The risk haplotype 46/1 was identified in every PV patient in the study; nine of them (75%) were homozygous, and three (25%) were heterozygous. The G allele increases the risk of developing an MPN 2.8-fold and strongly associates with the V617F mutation, which is detected in almost every PV case.\(^\text{126-129}\) However, PV is a very rare disease, and the **JAK2** V617F haplotype is detected in more than one fourth of the population, suggesting that it cannot be the only cause for the disease development. All the patients in our study were **JAK2** V617F mutation-positive, as expected.

To conclude, the identification of new hereditary gene variants and mutations may lead to screening of family members and other individuals at high risk of developing PV or other myeloid malignancies. Both diseased and asymptomatic family members should be interviewed about the family history to improve in making diagnoses. The rare variants identified in our exome study may predispose to PV development in the studied family, although they were not identified in other patients; the whole genes were not screened but only the specific SNV locations. All in all, exome sequencing provides an
excellent tool for the sequence analysis of gene-coding regions of the genome, although, for example, copy number alterations, larger genomic rearrangements or aberrations in non-coding regions cannot be identified solely in exome data. In addition, it is not known if a germline mutation leading to cytokine dysregulation in the BM microenvironment has an impact on the PV development, though it is known that many pro-inflammatory cytokines are dysregulated in MPN patients and drive the clonal evolution.\textsuperscript{271-274} More studies are warranted to gain better insights into familial clustering of PV and other MPNs.

3 Females with germline mutation in X-chromosomal telomere gene

Patients with dyskeratosis congenita are at risk of developing cancer, especially hematological malignancies and squamous cell carcinomas of the head and neck. The risk for MDS is remarkably high – over 500-fold compared to healthy population.\textsuperscript{275} This is most importantly due to bone marrow failure and defects in telomere maintenance and function, which results in exceedingly short telomeres.\textsuperscript{191} DC is one of the most severe telomere biology disorders, typically presenting in childhood as BMF. The most frequently mutated gene is the dyskerin-encoding \textit{DKC1} in X-chromosome. Dyskerin functions in stabilizing the telomerase RNA, thus maintaining its activity.\textsuperscript{276,277} Hence, with X-linked recessive pattern of inheritance it mostly affects males. Females are assumed to be protected from the disease by X-inactivation. Skewed XCI has been thought to function especially well in hematopoietic tissues in females.\textsuperscript{278} To date, only few reports on female \textit{DKC1} pathogenic mutation carriers exist.\textsuperscript{278-281} Here, we studied three symptomatic sisters with a novel heterozygous germline \textit{DKC1} insertion of three nucleotides. They showed characteristic DC-like symptoms such as early greying of the hair, nail dysplasia, and skin hyperpigmentation. The same mutation was identified in two brothers in the family, who were diagnosed with DC and had short telomeres.\textsuperscript{237}
3.1 Mutant \textit{DKC1} allele expression and telomere length

In order to investigate the XCI status and mutant allele expression in the females, we compared the \textit{DKC1} mutant and wild type allele expression levels in different tissues by utilizing droplet digital PCR. The mutant \textit{DKC1} allele expression varied especially in the blood: the mother of the brothers showed 0\% expression of the mutant allele, whereas one of the two aunts (aunt 1) showed up to 45\% expression. The expression does not fully correlate with the telomere length analysis; the mother had shortened telomeres in her blood cells but no expression of the mutated allele. However, it has been shown that there is not a strict relationship between TL and severity of symptoms at the individual level.\textsuperscript{282} Families with individuals displaying variable symptoms and telomere lengths despite carrying the same mutations have been reported.\textsuperscript{233,283} Aunt 2 in our study had normal telomere lengths but showed 5\% mutant allele expression in her blood cells. Despite X-linked inheritance, two out of three females in our study had shortened telomere lengths, which refers to a TBD. Furthermore, we detected both mutant and wild type \textit{DKC1} expression in blood in two females, who were both diagnosed with marginal macrocytic anemia, suggesting that skewed XCI does not necessarily protect females from hematological symptoms. Germline mosaicism and epigenetics, in addition to XCI, have been suggested to contribute to DC-like phenotypes in female \textit{DKC1} mutation carriers.\textsuperscript{278,281} Even though we detected both alleles to be present in all the tissue types and having their unique distribution patterns, it is likely that signals come from different cells expressing either mutant or wild type \textit{DKC1} alleles, not both. Whether the allele distribution pattern arises by coincidence or is a controlled phenomenon remains unknown.

The clinical and molecular examination of one family can significantly increase our knowledge on rare disease features. Our study strengthens the finding that telomere lengths can vary between individuals with the same mutation in the same family, and does not strictly correlate with DC-like symptoms. In this study, we expanded the clinical spectrum of \textit{DKC1}-linked TBD in women and revealed the molecular roots of the patients’ symptoms.
Long-term follow-up of the females carrying mutations in telomere biology genes is important. The health problems in females, too, may be severe. Thus, early recognition of all patients with DC or other TBD is essential.

4 Leukemia germline research and diagnostics now and in the future

Transferring NGS methods and germline analysis from the research laboratory to the clinical setting can be challenging. One of the main goals is to successfully transmit genetic discoveries into therapeutic interventions. Whole-exome sequencing is the most inclusive approach in cancer genome research; however it has some major limitations. For example, it “eliminates” more than 98% of the genome, which brings certain risks to the analysis via the inability to screen anything other than the coding regions. Due to this, the best approach for the optimization of genome analysis has to be considered. Furthermore, the detection of variants is relatively straightforward, whereas the interpretation of variants and their consequences still remains challenging. Variants of unknown significance can be difficult to analyze due to a lack of knowledge and evidence for both the scientific community and clinicians, which patients can sometimes find shocking. Occasionally, a somatic mutation is accidentally considered to be a germline mutation or vice versa. Artificial intelligence (AI) may help in prediction and decision making, and by early detection, potentially minimize disease severity. Collecting high-quality sample material and clinically annotated data in e.g. biobanks secures the possibility of using even larger data sets in future research.

The identification and knowledge of germline variants and mutations in patients with AML or leukemia-predisposing diseases or syndromes has both advantages and challenges (Table 4). Reduced or incomplete penetrance in individuals with a germline mutation may explain why inherited diseases are in some cases transferred through unaffected parents and thus are not recognized, but also why clinically healthy family members
can carry potentially pathogenic variants without clinical symptoms. In addition to the patients themselves, the knowledge about the germline variants can affect the family members and may lead to increased anxiety and stress. Despite this, mutation carriers should be encouraged to obtain genetic counseling. Leukemia unfortunately cannot be prevented despite the advances in therapy development, which even further increases the anxiety in these families with a clustering of myeloid malignancies. However, the incorporation of both somatic and germline information into patient care is important in the era of precision medicine and tailored treatment strategies – especially when considering an HSC transplant donor.
Table 4. Advantages and challenges in germline diagnostics of hematological malignancies.

<table>
<thead>
<tr>
<th>ADVANTAGES</th>
<th>NGS techniques</th>
<th>Patient</th>
<th>Family &amp; other individuals with same germline mutation</th>
<th>Therapy strategies</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identification of new predisposition genes, mutations, and syndromes</td>
<td>Finding a reason for the disease</td>
<td>Early detection of malignancy</td>
<td>Precision medicine, tailored treatments</td>
<td></td>
</tr>
<tr>
<td>Relatively fast</td>
<td>Precision medicine, individualized therapy</td>
<td>Follow-up</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Increasing knowledge</td>
<td>Increasing knowledge</td>
<td>Genetic counseling and risk assessment</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHALLENGES</th>
<th>Interpretation of the variants</th>
<th>Anxiety and other psychosocial issues</th>
<th>Anxiety and other psychosocial issues</th>
<th>Leukemia cannot be prevented – little advances in therapy development due to disease heterogeneity</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lack of knowledge</td>
<td>Not necessarily present with suspicious family history or symptoms</td>
<td>Penetration cannot necessarily be predicted</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Errors</td>
<td>Risks</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Transition of genetic discoveries to therapeutics</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

NGS: Next-Generation Sequencing
Germline predisposition to myeloid malignancies is well known, but there are still clinical challenges in the recognition of familial leukemia and other hematological malignancies. Patients may be unaware of their predisposition and be asymptomatic mutation carriers. On the other hand, the other mutation carriers in the family may already manifest symptoms in their younger years. Thus, wide variation in age of onset, disease latency and outcome, and variable penetrance of germline mutations can complicate the recognition of familial aggregation. Understanding the recognized syndromes is critical for clinicians to have a high index of suspicion and to have knowledge about how to manage patients with germline mutations.

Targeted sequencing methods are routinely used in leukemia diagnostics; however, in germline screening they are still rarely used in the clinic. The identification of predisposing genes and mutations is important for families with susceptibility for hematological malignancies. Knowledge about specific predisposing alterations enables the screening of other individuals at higher risk as well, which may affect, for example, the selection of transplant donors or therapy options. NGS technologies, such as whole-exome sequencing, enable the identification of sequence variants, although larger structural variants, non-coding mutations, and copy number alterations are difficult or impossible to detect with exome sequencing. In addition, variants of uncertain significance may complicate genetic counseling and management.

The active gathering of family history is important for detecting inherited predisposition to malignancies. Germline exome analysis in conjunction with somatic exome sequencing may be justifiable. Inherited germline variants potentially associated with e.g. AML may also be identified in older cases without known family history. Also, clinical and molecular examination of only one family can significantly increase our knowledge.
on rare disease features, and follow-up of both the symptomatic and asymptomatic individuals with a predisposing germline variant is essential despite the possible anxiety in the patient and their family members. Therefore, clinicians are encouraged to integrate interpreted germline data into patient care, and the complexity of genetic information requires active collaboration between researchers, hematologists, and counseling units.
ACKNOWLEDGEMENTS

This thesis was carried out in the Hematological Genetics Lab at the Department of Medical and Clinical Genetics and Research Programs Unit, Faculty of Medicine, University of Helsinki during 2015-2019. I wish to acknowledge the Doctoral Programme in Biomedicine (DPBM) for their high-quality education and creating an inspiring atmosphere. This thesis was financially supported by grants from the Finnish Cultural Foundation, the Biomedicum Helsinki Foundation, the Orion Research Foundation, the Cancer Society of Finland, and the Academy of Finland. I also wish to thank the Institute for Molecular Medicine Finland, and the Biomedicum Functional Genomics Unit for their excellent services.

I wish to acknowledge all the co-authors contributing to this PhD project and clinicians for providing invaluable sample material and their expertise: Prof. Kimmo Porkka, MD, PhD, Prof. Lauri A. Aaltonen, MD, PhD, Prof. Kari Hemminki, MD, PhD, Esa Pitkänen, PhD, Saara Peuhkuri, MB, Hanna Välimaa, MD, PhD, Katriina Hannula-Jouppi, MD, PhD, Sofie Degerman, PhD, Anna Norberg, PhD, Caroline Heckman, PhD, Kaisa Kettunen, PhD, and Janna Saarela, MD, PhD. Riku Katainen, M.Sc. is thanked for the technical help with BasePlayer.

My supervisors, docents Outi Kilpivaara, PhD, and Ulla Wartiovaara-Kautto, MD, PhD, are sincerely thanked for the invaluable opportunity to work as a member of the hematological genetics research group and for offering such an interesting PhD project. Your enthusiasm towards science is inspiring and I have learned a countless number of things from both of you. Outi, thank you for believing in me, especially in the beginning of my doctoral studies when I knew practically nothing about cancer genetics. Ulla, thank you for guiding me into the fascinating world of hematology and for showing what passion for science is. It has been a pleasure to be part of this lab.

The pre-examiners of my thesis, docents Katri Pylkäs, PhD, and Mikko Arvas, PhD, are appreciated for their critical review of this thesis and their helpful comments and suggestions. Likewise, I wish to sincerely thank my thesis committee members, docents Maija Wolf, PhD, and Miina Ollikainen, PhD, for their support and advice during this project. I would like to thank Diana Cousminer for the language revision of my thesis.
I express my gratitude to all former and present members of the KiVa lab for sharing the good (and not so good) moments. I wish to thank Terhi A. for the friendship and endless discussions about everything and anything (especially the thorough analysis of reality TV series!), and for all the fun conference trips. Without you this PhD journey would have been so much less fun. Tuomas, thank you for the best laughs and advice on basically everything. Inari, Suvi, and Atte, thanks for (peer) support and all the fun moments inside and outside the lab. Kati, thank you for all the advice and for answering my stupid questions. Annukka, Lotta, and Pihla are thanked for the help and technical expertise. I would like to wish Marja, Anna, Susanna, Sara, and Terhi F. good luck with the doctoral studies. Nilla, Pia, and Miika are thanked for their expertise and knowledge on various things. I have been privileged to work with all of you!

I wish to thank all my friends outside work for reminding me of all the truly important things in life. The University gang Elina N, Marianne, Ansku, Maiju, Sanna, Mia, Anu, Lissu, Laura, and all the great people I bumped into during the years; thank you for all the fun adventures outside “scientific world”. You are golden. Special thanks goes to Elina N. for the invaluable peer support and discussions about the pros and cons (mostly cons) of doctoral studies. The Kouvola people Emmi, Heli, Päivi, Milla, Piia, Heidi, Kaisa, and Mia: no matter the time or distance, friendship is forever. Thank you for growing up with me and always believing in me. I wish to thank Sonja for all the laughs in Ilosaarirock festival in those endless summer nights when everything felt possible.

Finally, I would like to thank my family; my mother Päivi and father Otto, who is no longer with us, for support and always letting me choose my own path; and my grandparents Aila and Pekka for everything. Salla, thank you for being the best sister ever. Without you this world would not make sense. My dear nephew Oiva, thank you for making me laugh through the darkest days and for reminding me of all the wonders of the world. Of the seven billion smiles on this planet, yours is my favorite. Above all, I would like to thank my boyfriend Pauli for all the support throughout my doctoral studies (and through basically everything) and for believing in me when I didn’t. You remind me of what really matters in life.

I express my deepest gratitude to all the patients who participated in this study.

Helsinki, April 2019

Elina Hirvonen
REFERENCES


Appendix

Supplementary Table 1. (III: Myeloid gene panel)
Supplementary Table 1. Myeloid gene panel (III).

<table>
<thead>
<tr>
<th>Gene (exons and exon-intron boundaries covered)</th>
<th>Gene (mutational hotspots covered)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ASXL1</strong></td>
<td>ASXL transcriptional regulator 1</td>
</tr>
<tr>
<td><strong>BCOR</strong></td>
<td>BCL6 corepressor</td>
</tr>
<tr>
<td><strong>CDKN2A</strong></td>
<td>cyclin-dependent kinase inhibitor 2A</td>
</tr>
<tr>
<td><strong>CEBPA</strong></td>
<td>CCAAT enhancer binding protein alpha</td>
</tr>
<tr>
<td><strong>CREBBP</strong></td>
<td>CREB binding protein</td>
</tr>
<tr>
<td><strong>CUX1</strong></td>
<td>Cut like homeobox 1</td>
</tr>
<tr>
<td><strong>DNMT3A</strong></td>
<td>DNA (cytosine-5)-methyltransferase 3A</td>
</tr>
<tr>
<td><strong>EP300</strong></td>
<td>E1A binding protein P300</td>
</tr>
<tr>
<td><strong>ETV6</strong></td>
<td>ETS variant 6</td>
</tr>
<tr>
<td><strong>EZH2</strong></td>
<td>Enhancer of zeste 2</td>
</tr>
<tr>
<td><strong>GATA2</strong></td>
<td>GATA-binding factor 2</td>
</tr>
<tr>
<td><strong>KDM6A</strong></td>
<td>Lysine demethylase 6A</td>
</tr>
<tr>
<td><strong>NF1</strong></td>
<td>Neurofibromin 1</td>
</tr>
<tr>
<td><strong>PHF6</strong></td>
<td>PHD finger protein 6</td>
</tr>
<tr>
<td><strong>RAD21</strong></td>
<td>RAD21 cohesin complex component</td>
</tr>
<tr>
<td><strong>SETD2</strong></td>
<td>SET domain containing 2</td>
</tr>
<tr>
<td><strong>STAG2</strong></td>
<td>Stromal antigen 2</td>
</tr>
<tr>
<td><strong>TET2</strong></td>
<td>Tet methylcytosine dioxygenase 2</td>
</tr>
<tr>
<td><strong>TP53</strong></td>
<td>Tumor protein 53</td>
</tr>
<tr>
<td><strong>ZRSR2</strong></td>
<td>Zinc finger CCCH-type, RNA binding motif and</td>
</tr>
</tbody>
</table>

**Gene (mutational hotspots covered)**

<table>
<thead>
<tr>
<th>Gene</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>BRAF</strong></td>
<td>B-Raf proto-oncogene, serine/threonine kinase</td>
</tr>
<tr>
<td><strong>CALR</strong></td>
<td>Calreticulin</td>
</tr>
<tr>
<td><strong>CBL</strong></td>
<td>Cbl proto-oncogene</td>
</tr>
<tr>
<td><strong>CSF3R</strong></td>
<td>Colony stimulating factor 3 receptor</td>
</tr>
<tr>
<td><strong>FLT3</strong></td>
<td>Fms related tyrosine kinase 3</td>
</tr>
<tr>
<td><strong>GATA1</strong></td>
<td>GATA binding protein 1</td>
</tr>
<tr>
<td><strong>IDH1</strong></td>
<td>Isocitrate dehydrogenase (NADP(+)) 1, cytosolic</td>
</tr>
<tr>
<td><strong>IDH2</strong></td>
<td>Isocitrate dehydrogenase (NADP(+)) 2,</td>
</tr>
<tr>
<td><strong>JAK2</strong></td>
<td>Janus kinase 2</td>
</tr>
<tr>
<td><strong>KIT</strong></td>
<td>KIT proto-oncogene receptor tyrosine kinase</td>
</tr>
<tr>
<td><strong>KRAS</strong></td>
<td>KRAS proto-oncogene, GTPase</td>
</tr>
<tr>
<td><strong>MPL</strong></td>
<td>MPL proto-oncogene, thrombopoietin receptor</td>
</tr>
<tr>
<td><strong>NPM1</strong></td>
<td>Nucleophosmin 1</td>
</tr>
<tr>
<td><strong>PDGFRα</strong></td>
<td>Platelet derived growth factor receptor alpha</td>
</tr>
<tr>
<td><strong>PTPN11</strong></td>
<td>Protein tyrosine phosphatase, non-receptor type 11</td>
</tr>
<tr>
<td><strong>RUNX1</strong></td>
<td>Runt related transcription factor 1</td>
</tr>
<tr>
<td><strong>SETBP1</strong></td>
<td>SET binding protein 1</td>
</tr>
<tr>
<td><strong>SF3B1</strong></td>
<td>Splicing factor 3b subunit 1</td>
</tr>
<tr>
<td><strong>SMC1A</strong></td>
<td>Structural maintenance of chromosomes 1A</td>
</tr>
<tr>
<td><strong>SMC3</strong></td>
<td>Structural maintenance of chromosomes 3</td>
</tr>
<tr>
<td><strong>SRSF2</strong></td>
<td>Serine and arginine rich splicing factor 2</td>
</tr>
<tr>
<td><strong>U2AF1</strong></td>
<td>U2 small nuclear RNA auxiliary factor 1</td>
</tr>
<tr>
<td><strong>WT1</strong></td>
<td>Wilms tumor 1</td>
</tr>
</tbody>
</table>