The Doctoral School in Health Sciences (DSHealth)
Doctoral Programme in Biomedicine (DPBM)

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

Supervised by

Docent Outi Kilpivaara, Ph.D.
Department of Medical and Clinical Genetics,
Research Programs Unit,
Applied Tumor Genomics,
Faculty of Medicine,
University of Helsinki, Helsinki, Finland

Docent Ulla Wartiovaara-Kautto, M.D., Ph.D.
Department of Hematology,
Helsinki University Hospital,
Comprehensive Cancer Center and
University of Helsinki, Helsinki, Finland

Reviewed by

Docent Katri Pylkäs, Ph.D.
Laboratory of Cancer Genetics and Tumor Biology,
Cancer and Translational Medicine Research Unit,
Biocenter Oulu,
Faculty of Medicine,
University of Oulu, Oulu, Finland

Docent Mikko Arvas, Ph.D.
Research and Development,
Finnish Red Cross Blood Service
Helsinki, Finland

Official opponent

Docent Taru Kuittinen, M.D., Ph.D.
Department of Medicine, Hematology
Kuopio University Hospital, Kuopio, Finland

ISBN 978-951-51-5215-2 (paperback)
ISBN 978-951-51-5216-9 (PDF)
http://ethesis.helsinki.fi
Unigrafia
Helsinki 2019

Oivalle

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:

I Wartiovaara-Kautto U*, **Hirvonen EAM***, Pitkänen E, Heckman C, Saarela J, Kettunen K, Porkka K, Kilpivaara O. Germline alterations in a consecutive series of acute myeloid leukemia. *Leukemia* 2018, 32(10):2282-2285. doi: 10.1038/s41375-018-0049-5.

II **Hirvonen EAM**, Pitkänen E, Hemminki K, Aaltonen LA, Kilpivaara O. Whole-exome sequencing identifies novel candidate predisposition genes for familial polycythemia vera. *Hum Genomic*s 2017, 11(1):6, doi: 10.1186/s40246-017-0102-x.

III **Hirvonen EAM**, Peuhkuri S, Norberg A, Degerman S, Hannula-Jouppi K, Välimaa H, Kilpivaara O*, Wartiovaara-Kautto U*. Characterization of an X-chromosome-linked telomere biology disorder in females with *DKC1* mutation. *Leukemia* 2019, 33(1):275-278. doi: 10.1038/s41375-018-0243-5.

*These authors contributed equally to this work.

ABBREVIATIONS

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

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
RUNXT1	Runx1 translocation partner 1
SAMD9	Sterile alpha motif domain containing 9
SAMD9L	Sterile alpha motif domain containing 9 like
SBDS	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 *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).[3] BM is located in the medullary cavity of vascularized, innervated bone and contains many different hematopoietic and non-hematopoietic cell types.[4] 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)[3] (**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.[5]

Hematopoiesis has classically been described as a cellular hierarchy with HSCs residing at the apex of the pyramid.[6] 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.[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.[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.[9,10] Several transcription factors (TF) are essential in the early regulation of hematopoietic cell fate decisions as well.[11] Overall, a great number of signaling pathways interact with each other in a well-organized manner in hematopoiesis.

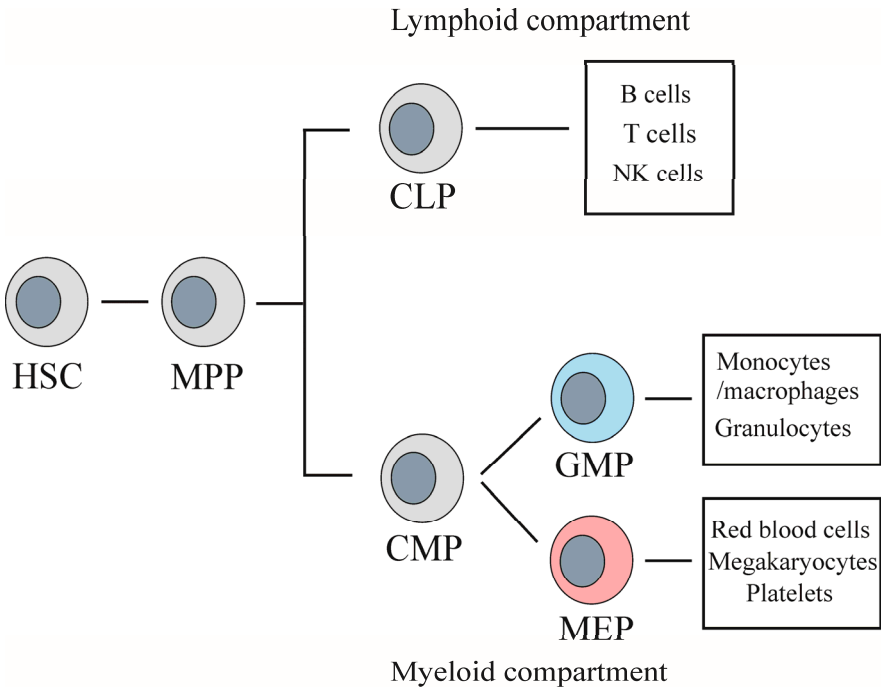

Figure 1. Simplified view of adult hematopoiesis. HSC, hematopoietic stem cell; MPP, multipotent progenitor; CLP, common lymphoid progenitor; CMP, common myeloid progenitor; GMP, granulocyte/macrophage progenitor; MEP, megakaryocyte/erythroid progenitor.

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.[4,9] 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.[4] In response to hematopoietic stress the niche can shift to extra-medullary sites.[4]

A BM niche is highly vascularized due to the vast production of hematopoietic cells. Perivascular stromal cells, including different mesenchymal cells[12-14] and CXCL12 (stromal-derived factor-1, SDF-1) abundant reticular (CAR) cells[15] 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.[16] Sinusoids compose a complex network in the BM cavity, and sinusoidal endothelial cells support HSC proliferation.[17] 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.[18-20] Osteoblasts in the endosteum, 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.[21] Furthermore, the sympathetic nervous system participates in the regulation of HSC function as well.[22]

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.[23,24] 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.[25] 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.[4]

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.[26] 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.[27-31] 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.[32]

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.[33-35] Early precursor cells develop into myeloblasts, and further into promyelocytes.[34] 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.[36] 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.[36] 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.[27] Tissue oxygenation affects the number of red blood cells produced, which is regulated by hypoxia-inducible factors (HIF), HIF-1 and HIF-2.[37] 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.[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,[38,39] which have the potential to differentiate into any of the lymphoid lineages.[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.[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.[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.[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.[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.[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.[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.[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.[47]

Only a small proportion of the alterations in HSCs are pathogenic driver mutations.[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.[55-57] Leukemia stem cells (LSC) are known to share several characteristics with normal HSCs,[58] albeit they can create niches that disrupt the normal HSC behavior in the microenvironment to favor their own expansion.[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).[62] For example, genetic alterations in osteogenic cells in the endosteal compartment may lead to MDS, and further to secondary leukemia.[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.[62] Distinct signals in these niches might also affect malignant transformation in many ways.[9,64] The BM microenvironment is hypoxic in myeloid malignancies; the level of oxygen modulates cell activity and function.[65]

2 Hereditary predisposition to cancer

Pathogenic germline mutations in cancer genes, especially in tumor suppressors capable of preventing tumorigenesis, cause predisposition to malignancies.[66] 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.[66,67] 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.[68] 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.[69] 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.[70] More than 200 genes have been recognized as drivers in common cancers; however, one single mutated gene is not yet enough to cause cancer.[71]

Cancer genes can be classified into oncogenes and tumor suppressors.[72,73] 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.[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.[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.[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.[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.[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.[66] The identification of these cancer predisposition genes and mutations, too, has a huge impact on clinical diagnosis, treatment, and possible prevention of cancer.

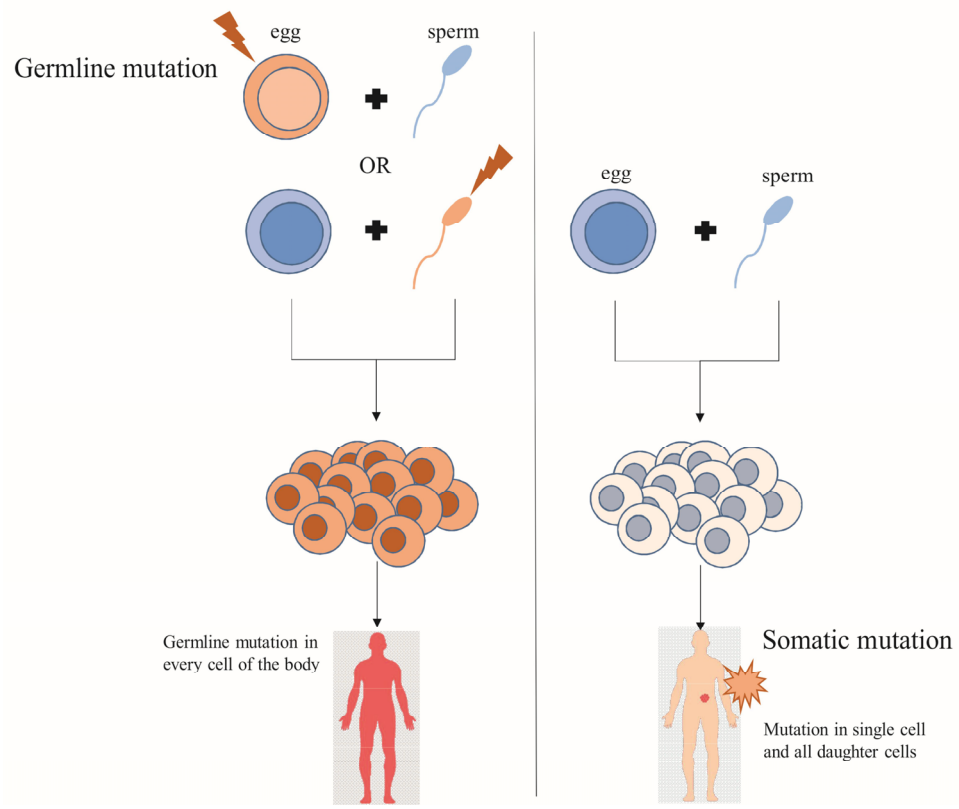

Figure 2. Germline mutations are hereditary and present in every cell of the body, since they occur in the sex cells or gametes that participate in fertilization. In contrast, somatic mutations result from changes in DNA of somatic cells of the body and are only transferred to daughter cells that form after cell division.

2.2 Hereditary cancer syndromes

Germline mutations cause hereditary predisposition to cancer.[66,67] 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*.[76] It encodes the protein p53, which functions in e.g. DNA damage repair, maintaining genomic stability, and cell-cycle arrest. Penetrance 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.[77] Another well-known cancer predisposition syndrome is Lynch syndrome, driven by mutations in DNA mismatch repair genes including *MSH2*, *MLH1*, *MSH6*, and *PMS2*, with risk of especially early colorectal cancer development.[78] In addition, e.g. PTEN hamartoma tumor syndrome with germline mutations in the tumor suppressor *PTEN* is one of the cancer predisposition syndromes.[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.[80] HSPCs accumulate mutations throughout life. Most of these somatic mutations are passengers, meaning that they have no contribution to clonal expansion.[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.[80] Functional and genetic changes contributing to the development of myeloid malignancies have been noted in BM niche cells as well.[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.[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.[84] Pathogenic germline mutations have been described in myeloid malignancies in several genes, of which *RUNX1* was the first in 1999.[85] Since then, multiple genes have been identified to associate with predisposition to MDS/AML. Germline mutations in e.g. *ANKRD26*, *CEBPA*, *GATA2*, *ETV6*, *TP53*, *BRCA1/2*, *DDX41*, *ELANE*, *SAMD9*, *SAMD9L*, and *SRP72*, in addition to certain inherited BMF syndromes, are frequently found in hereditary HM patients.[84,86,87] The affected proteins are involved in multiple functions including transcription, telomere maintenance, DNA repair, RNA processing, and inflammation.[86]

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 (**Table 1**).[84]

Table 1. Classification of myeloid neoplasms with germline predisposition (WHO 2016).

Myeloid neoplasms with germline predisposition without a preexisting disorder or organ dysfunction
AML with germline *CEBPA* mutation
Myeloid neoplasms with germline *DDX41* mutation
Myeloid neoplasms with germline predisposition and preexisting platelet disorder
Myeloid neoplasms with germline *RUNX1* mutation
Myeloid neoplasms with germline *ANKRD26* mutation
Myeloid neoplasms with germline *ETV6* mutation
Myeloid neoplasms with germline predisposition and other organ dysfunction
Myeloid neoplasms with germline *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

*Adapted from publications [2,84].

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.[88,89] MPNs are divided into subcategories: chronic myeloid leukemia (CML), which is characterized by the *BCR-ABL* oncogene fusion (Philadelphia chromosome, Ph-positive), and Ph-negative disorders named polycythemia vera (PV), essential thrombocythemia (ET), primary myelofibrosis (PMF), and prefibrotic PMF.[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.[90] Prefibrotic myelofibrosis, characterized by granulocytic and megakaryocytic proliferation and lack of fibrosis in the BM, represents an early phase of myelofibrosis.[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.[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.[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.[94] Transitions among these diseases are also observed occasionally.[83] Prior to leukemic transformation, MPNs often progress to BM fibrosis, which can present with very severe symptoms even at the preleukemic stage.[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.[96] 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.[97] *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.[90] 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.[98] 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.[99-102] It is present in ~98% of PV patients and about 50% of ET and PMF patients.[103-105] 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.[106,107] Somatic mutations in *CALR* exon 9 are found in 20-35% of ET and PMF cases.[108,109] Furthermore, *MPL* mutants are found approximately in 3-4% of ET and 6-7% of PMF cases.[110-113] These mutants are restricted to MPL activation, which explains why they are found only in ET and PMF, and not in PV.[97] 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".[114] 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.[115,116] In addition, mutations and somatic loss of heterozygosity in *TP53* is strongly associated with leukemic transformation.[98] When diagnosed with an MPN following the WHO diagnostic criteria,[2] 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.[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 *JAK2* mutation.[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.[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.[96]

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

3.2.2 Genetic predisposition to MPNs

The mutations in *JAK2*, *CALR*, and *MPL* probably are not causative mutations but mainly drive the disease phenotype – a preexisting germline factor likely predisposes to a clonal MPN.[105] Familial clustering of MPNs is well recognized; about 7% of cases involve germline predisposition.[90] Familial MPN has been described with dominant-autosomal

transmission,[124,125] and a shared susceptibility factor among the family members likely exists.[116] 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.[126-129] The G allele at rs10974944 increases an individual's risk of developing an MPN 2.8-fold (population attributable risk 46%).[127] 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.[130-132] Furthermore, germline mutations in *JAK2*[133] and *MPL*[134] have been described in triple-negative ET and PMF, and *RBBP6* has been suggested as a candidate gene for MPN susceptibility.[135] Duplication of *ATG2B* and *GSKIP* also predisposes to MPN development.[136] 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.[137-140]

3.3 Myelodysplastic syndromes

MDS are a heterogeneous group of clonal BM diseases arising from the expansion of mutated HSCs.[141] 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.[2,48] 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.[142]

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;[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.[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.[49-51,144] Somatic mutations in *SF3B1*, *TET2*, *DNMT3A,* and *ASXL1* are the most commonly described abnormalities in MDS cases.[142,145,146] MDS is defined as preleukemia since it frequently progresses to sAML: about 20-30% of MDS patients develop AML.[147,148] Hematopoiesis-disrupting mutations in nine specific genes, including four spliceosome genes, occur more frequently in MDS than *de novo* AML (mutated in 60-70% and 5-10% cases, respectively).[148,149]

3.3.2 Genetic predisposition to MDS

Despite most cases being sporadic, predisposition to MDS can be hereditary.[150] One of the best-characterized MDS predisposition syndromes is familial platelet disorder with an autosomal dominant inheritance pattern caused by heterozygous germline *RUNX1* mutations.[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).[148,151-153] In addition, individuals with e.g. a germline *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.[154] Additionally, *SAMD9* and *SAMD9L* mutations predispose to MDS/AML.[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, **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.[157,158] They compete for BM niche occupancy with normal cells and disrupt hematopoiesis.[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.[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.[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.[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.[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.[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.[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.[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.[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.[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.[170,171] Prognosis for relapsed patients is poor.

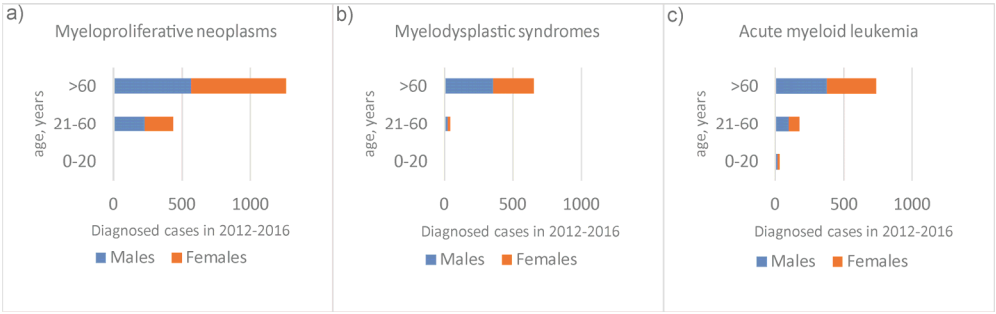

Figure 3. New myeloid disease cases in Finland in years 2012-2016. **a)** MPN, **b)** MDS, and **c)** AML cases. Finnish Cancer Registry, https://tilastot/syoparekisteri.fi/syovat, data from 2018-09-11, version 2019-02-12-002.

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.[82] Recurrent abnormalities in *de novo* AML include e.g. fusion gene forming translocations or inversions (*PML-RARA*, *RUNX1-RUNXT1*, *MLL1*, *MYH11-CBFB*) and *FLT3* internal tandem duplication, which have been recognized for decades and are used as diagnostic and prognostic markers (FAB sybtypes).[172] Overall, the driver mutations can be different for distinct AML subtypes.[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.[173,174] Genes associated with DNA methylation (*DNMT3A*, *IDH1/2*, *TET2*), transcription regulation and signaling activation (*ASXL1*, *RUNX1*, *CEBPA*, *FLT3*, *N/KRAS*), and nucleophosmin (*NPM1*) are recurrently mutated:[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.[178] RNA splicing factors are mutated in about 10% of the cases.[178] The co-occurrence of mutations can have an impact on overall survival.[149]. Intriguingly, AML with normal cytogenetics comprise almost half of all cases.[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.[84] Inherited germline mutations in e.g. *GATA2*, *RUNX1*, or *CEBPA* predispose to AML development.[28,181,182] These transcription factors are important for myeloid proliferation and differentiation. For example, a heterozygous mutation in *GATA2* confers a 70% risk of developing MDS/AML, and the progression to AML is associated with cytopenias and other somatic mutations.[154,183] Germline mutation in *RUNX1* causes familial platelet disorder with a risk of about 40% of developing AML.[85] Familial AML with a germline *CEBPA* mutation has nearly complete penetrance for AML development through autosomal dominant inheritance.[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[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).[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*.[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).[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.[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.[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.[193] 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.[194] Some patients may develop life-threatening infections or leukemia at younger age, whereas some remain asymptomatic for several years or decades.[194,195] However, almost all patients suffer from peripheral blood monocyte, B, and NK cell cytopenias,[196-198] and the majority will develop myeloid malignancy.[194] Other inherited BMF syndromes such as severe congenital neutropenia and thrombocytopenia also increase AML risk, though the age of onset is later.[199]

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.[191] 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.[200] 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.[201] BMF affects 80-90% of DC cases by the age of 30 years, and the only curative treatment for BMF is allogeneic HSC transplant.[202] The incidence of DC is only 1 in 1,000,000 individuals.[201]

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.[203] Abnormal shortened telomeres lead to enhanced aging of cells, but in some situations can also help the cells to become malignant.[204] The genetic basis is undetectable in about one third of DC cases.[201]

The X-linked form of DC is caused by mutations in the X-chromosomal gene *DKC1*, which are the most frequent mutations appearing in almost one third of patients.[205] 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.[206] XCI can be incomplete in humans, meaning that some genes are expressed from both the active and the inactive X-chromosomes.[205] Female *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.[207] 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.[208] 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.[209] 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.[210] 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.[209] Optimally, NGS can provide the correct diagnosis for patients by identifying the disease-causing mutations.[211] 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.[212] Gene panels can help in making a rapid diagnosis and therapeutic decisions in many genetic disorders.[213] 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-chromosome[216] 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.[221-223] These studies led the 'cancer stem cell' concept and changed the way we view cellular epigenetic plasticity within a genetically homogeneous clone.[217] In addition, many other remarkable discoveries such as stem cell transplants and combination chemotherapy have come from leukemia trials.[217]

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.[224] The first remarkable whole exome studies covering myeloid malignancies revealed many recurrent pathogenic mutations[81,225-227] 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.[148] The availability of NGS platforms has also enabled individualized diagnostic evaluations and tailored treatment strategies.[84] 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%.[228] This methodology is useful in, for example, the early detection of leukemia relapse and for monitoring minimal residual disease.[229,230] 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,[231] 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[231] 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[232,233].

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% CO_2.

3.2 Commercial cell lines (III)

SET-2 essential thrombocythemia 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 earlier[234] 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 Insitutet.

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.[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.[235] Panel A represented known leukemia predisposing

genes[236], and Panel B included genes implicated in DNA repair and solid tumor predisposition (**Figure 4**).

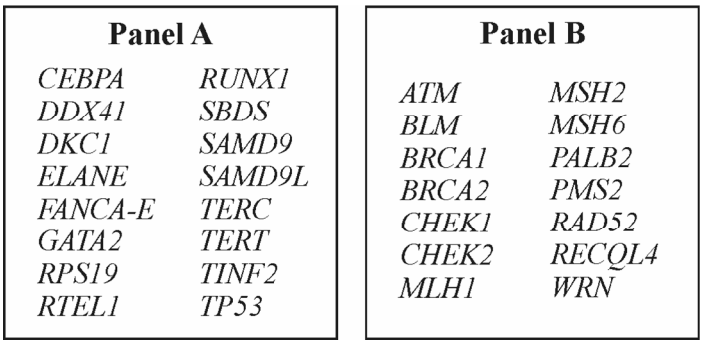

Figure 4. Gene panels representing known leukemia predisposing genes (Panel A), and genes implicated in DNA repair (Panel B) used in Study I.

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.[232,233] Relative telomere length (RTL) for the

brothers was measured previously.[237] 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[231] 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[238] with several online *in silico* prediction tools, PolyPhen-2,[239] SIFT,[240] and ClinVar:[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),[242] and the Human Gene Mutation Database (HGMD).[243] In Study II, the effect of variants was predicted with PolyPhen-2 and SIFT.

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/µL 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 *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: *ATM, BLM, BRCA1, CHEK1, CHEK2, DDX41, FANCA, FANCE, GATA2, MLH1, PALB2, PMS2, RTEL1, SAMD9L, SBDS*, and *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)[244] in *DDX41*, and a homozygous splice-site mutation (c.238 + 2T>C)[245] in *SBDS*. The start-lost mutation in *DDX41* is categorized as a risk factor for AML. The patient with the *DDX41* mutation was 64 years old and did not have any family history of cancer. The splice-site mutation in *SBDS* is considered to be a risk factor as well. The *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.

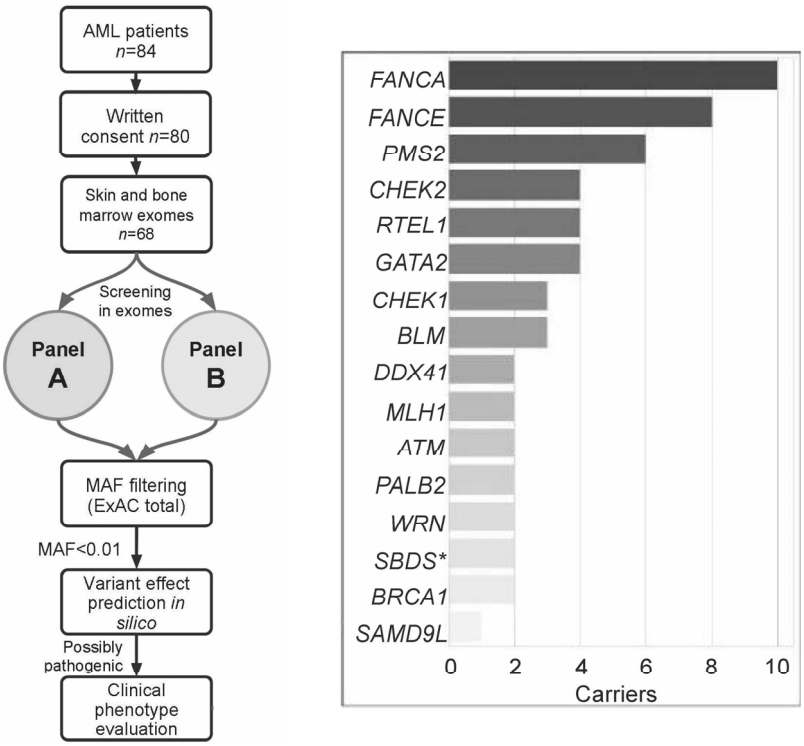

Figure 5. Flowchart describing the AML study and analysis, and the number of patients carrying variants in 16 genes. An asterisk denotes a heterozygous and homozygous case. Adapted and reproduced with permission from Wartiovaara-Kautto *et al.*, *Leukemia*, 2018.

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.Met1Ile 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.

Patient number	Gender	Gene	Location	Variant	MAF ExAC/SISu (total/Finns)
2203	F	*CHEK2*	*22:29091856delG*	c.1230delG (p.Thr410MetfsTer15)	0.00182/ 0.00872
3626	M	*CHEK2*	*22:29121087A>G*	c.470T>C (p.Ile157Thr)	0.00410/ 0.02437
4257	M	*FANCA*	*16:89825107G>T*	c.2859C>A (p.Asp953Glu)	0.00138/ 0.00615
4257	M	*SBDS**	*7:66459197T>C**	c.258+2T>C* (p.Gln86 spl2*)	-/ 0.00974
4311	M	*FANCE*	*6:35427531T>C*	c.1310T>C (p.Met437Thr)	0.00208/ 0.01463
4336	F	*FANCA*	*16:89871796G>A*	c.601C>T (p.Pro201Ser)	0.00557/ 0.05899
4347	F	*MLH1*	*3:37059009A>G*	c.803A>G (p.Glu268Gly)	0.00017/ 0.00143
4347	F	*PALB2*	*16:23614979C>T*	c.3362G>A (p.Gly1121Asp)	8.297e-06/ 4.766e-05
4368	M	*FANCA*	*16:89865516T>A*	c.971A>T (p.Ter317Tyr)	0.00082/ 0.00005
4368	M	*CHEK2*	*22:29121087A>G*	c.470T>C (p.Ile157Thr)	0.00410/ 0.02437
4374	M	*WRN*	*8:30958442T>G*	c.2059T>G (p.Leu687Val)	0.00097/ 0.00934
4411	M	*FANCE*	*6:35423528C>T*	c.253C>T (p.Pro85Ser)	0.00175/ 0.01792
4543	F	*FANCE*	*6:35427531T>C*	c.1310T>C (p.Met437Thr)	0.00208/ 0.01463
4583	F	*BLM*	*15:91341543A>C*	c.3334A>C (p.Asn1112His)	-/ -
4583	F	*CHEK1*	*11:125495891G>A*	c.236G>A (p.Trp79Ter / UTR)	0.00141/ 0.00238
4619	F	*BRCA1*	*17:41226488C>A*	c.4535G>T (p.Ser1512Ile)	0.00215/ 0.001192
4619	F	*SBDS*	*7:66460335C>T*	c.70G>A (p.Gly24Arg)	-/ 9.540e-05
4624	F	*FANCA*	*16:89805301G>C*	c.4249C>G (p.His1417Asp)	0.00340/ 0.00987
4624	F	*PALB2*	*16:23635370C>T*	c.2794G>A (p.Val932Met)	0.00597/ 0.01665
4624	F	*WRN*	*8:31012183C>T*	c.3731C>T (p.Thr1244Met)	0.00010/ 0.00143
4653	M	*GATA2*	*3:128204731C>T*	c.710G>A (p.Gly237Asp)	0.00229/ 0.03673
4686	M	*FANCA*	*16:89871796G>A*	c.601C>T (p.Pro201Ser)	0.00557/ 0.05899
4739	F	*DDX41*	*5:176943769T>G*	c.95T>G (p.Asp32Ala (Asp50Ala))	0.00012/ 0.00176
4739	F	*FANCE*	*6:35423528C>T*	c.253C>T (p.Pro85Ser)	0.00175/ 0.01792
4740	F	*FANCE*	*6:35427531T>C*	c.1310T>C (p.Met437Thr)	0.00208/ 0.01463
4767	F	*PMS2*	*7:6045634T>C*	c.52A>G (p.Ile18Val)	0.00908/ 0.02480
4771	F	*CHEK2*	*22:29121087A>G*	c.470T>C (p.Ile157Thr)	0.00410/ 0.02437†
4771	F	*RTEL1*	*20:62321690delAGA*	c.2310del3 (p.Glu795del)	0.00532/ 0.01551†
4868	M	*BLM*	*15:91304473C>T*	c.1871C>T (p.Gln624Ter)	9.483e-06/ 5.289e-05
4898	F	*FANCA*	*16:89805301G>C*	c.4249C>G (p.His1417Asp)	0.00340/ 0.00987
4898	F	*BRCA1*	*17:41226488C>A*	c.4535G>T (p.Ser1512Ile)	0.00215/ 0.001192
4898	F	*GATA2*	*3:128204731C>T*	c.710G>A (p.Gly237Asp)	0.00229/ 0.03673
4908	F	*PMS2*	*7:6026886C>G*	c.1510G>C (p.Glu504Gln)	2.476e-05/ -†
4980	F	*BRCA1*	*17:41243835G>A*	c.3713C>T (p.Pro1238Leu)	0.00016/ 0.00143
4991	M	*PMS2*	*7:6045634T>C*	c.52A>G (p.Ile18Val)	0.00908/ 0.02480
5034	F	*FANCA*	*16:89871796G>A*	c.601C>T (p.Pro201Ser)	0.00557/ 0.05899
5132	M	*DDX41*	*5:176943944C>T*	c.3G>A (p.Met1Ile start lost)	5.108e-05/ 5.842e-05
5200	F	*PMS2*	*7:6045634T>C*	c.52A>G (p.Ile18Val)	0.00908/ 0.02480
5220	M	*CHEK1*	*11:125495891G>A*	c.236G>A (p.Trp79Ter / UTR)	0.00141/ 0.00238
5232	M	*FANCE*	*6:35427531T>C*	c.1310T>C (p.Met437Thr)	0.00208/ 0.01463
5232	M	*FANCA*	*16:89871796G>A*	c.601C>T (p.Pro201Ser)	0.00557/ 0.05899
5305	M	*CHEK2*	*22:29121087A>G*	c.470T>C (p.Ile157Thr)	0.00410/ 0.02437
5686	M	*ATM*	*11:108216611C>T*	c.8560C>T (p.Arg2854Cys)	0.00018/ 9.534e-05
5690	M	*BLM*	*15:91312417C>A*	c.2362C>A (p.Leu788Ile)	0.00101/ 0.00539
5750	M	*GATA2*	*3:128204731C>T*	c.710G>A (p.Gly237Asp)	0.00229/ 0.03673
5750	M	*MLH1*	*3:37035057G>T*	c.19G>T (p.Val7Phe)	6.594e-05/ 0.00100
5776	M	*FANCA*	*16:89805301G>C*	c.4249C>G (p.His1417Asp)	0.00340/ 0.00987
5776	M	*FANCE*	*6:35423528C>T*	c.253C>T (p.Pro85Ser)	0.00175/ 0.01792
5806	F	*RTEL1*	*20:62322188G>T*	c.2444G>T (p.Ser839Ile)	0.00177/ 0.00687
5897	M	*FANCE*	*6:35423528C>T*	c.253C>T (p.Pro85Ser)	0.00175/ 0.01792
5919	F	*RTEL1*	*20:62309621T>C*	c.1031T>C (p.Met344Thr)	0.00109/ 0.00241
6088	F	*PMS2*	*7:6045634T>C*	c.52A>G (p.Ile18Val)	0.00908/ 0.02480
6187	F	*FANCA*	*16:89871796G>A*	c.601C>T (p.Pro201Ser)	0.00557/ 0.05899
6246	F	*CHEK1*	*11:125495732A>G*	c.77A>G (p.Asn26Ser)	0.00224/ 0.00976
6246	F	*FANCA*	*16:89871796G>A*	c.601C>T (p.Pro201Ser)	0.00557/ 0.05899
6246	F	*PMS2*	*7:6045634T>C*	c.52A>G (p.Ile18Val)	0.00908/ 0.02480 ;
6246	F	*SAMD9L*	*7:92761636A>G*	c.3649T>C (p.Phe1217Leu)	4.125e-05/ 0.00048
6247	M	*RTEL1*	*20:62321690delAGA*	c.2310del3 (p.Glu795del)	0.00532/ 0.01551
6310	F	*GATA2*	*3:128204731C>T*	c.710G>A (p.Gly237Asp)	0.00229/ 0.03673

We identified a VUS in *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 *CHEK1, PMS2*, and *FANCA* genes. The *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 *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 *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**.

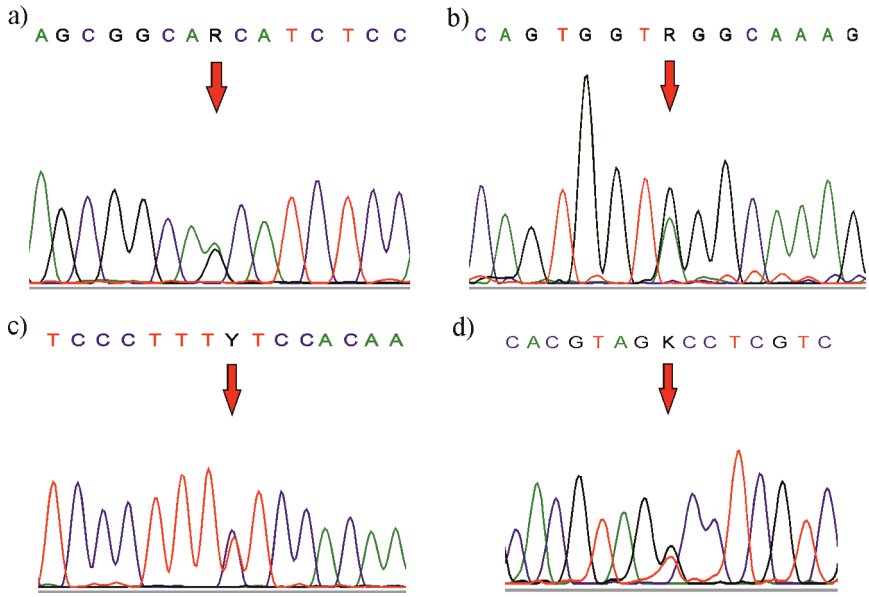

Figure 6. Sequencing chromatograms of germline variants of unknown significance in *CHEK1*, *SAMD9L*, and *DDX41*. **a**) *CHEK1* c.77A>G (p.Asn26Ser), **b**) *CHEK1* c.236G>A (p.Trp79Ter/UTR), **c**) *SAMD9L* c.3649T>C (p.Phe1217Leu), and **d**) *DDX41* c.95T>G (p.Asp32Ala/Asp50Ala).

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.

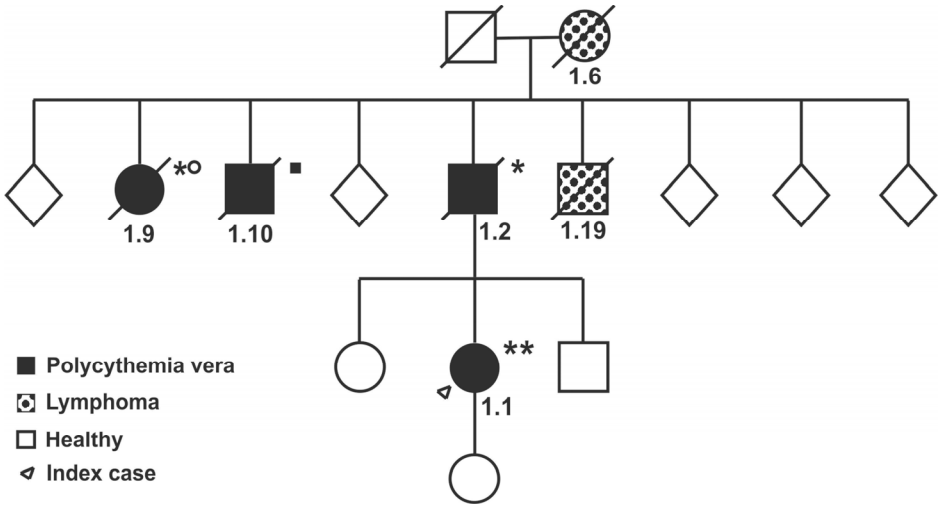

Figure 7. Pedigree of the Finnish family with four polycythemia vera patients and two patients with lymphoma. Family members with exomes available are marked with an asterisk and the individual used for validation with a small black square. Two asterisks denote both exome (germline) and whole genome (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.

Gene	Genomic location	Variation, cDNA	Variation, protein
ZXDC	Chr3: 126189754	c.1254C>G	p.F418L
ATN1	Chr12: 7046361	c.1931C>G	p.P644R
LRRC3	Chr21: 45877228	c.701G>A, rs148872771	p.R234Q
GNL3	Chr3: 52727477	c.1241A>G	p.Y414C
MDC1	Chr6: 30679188	c.2221+1G>T	spl1
ITPR3	Chr6: 33635026	c.1672C>T, rs780906252	p.R558C
FAM135A	Chr6: 71190668	c.1221G>A, rs143901584	p.V203M
SLC2A12	Chr6: 134312391	c.1756C>T, rs200847615	p.P586S
WDR86	Chr7: 151097265	c.226G>A, rs199824863	p.D76N
CSMD1	Chr8: 3165238	c.3929C>T	p.A1310V
SLC24A2	Chr9: 19786283	c.582A>G, rs368590535	p.I194M
ITPKC	Chr19: 41224132	c.1092C>G, rs143757004	p.D364E

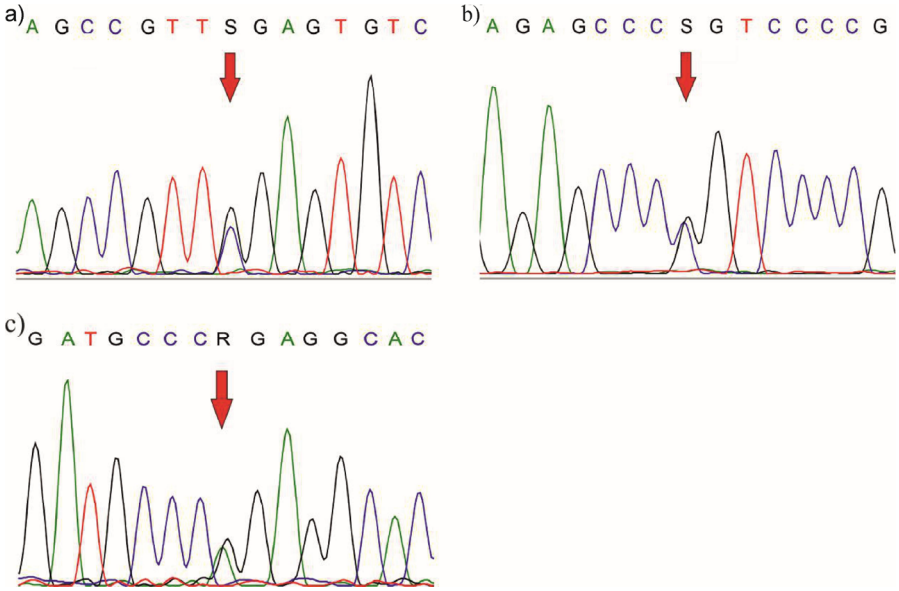

Figure 8. Sequencing chromatograms of novel PV predisposition variants **a)** *ZXDC* c.1254C>G (p.Phe418Leu), **b)** *ATN1* c.1931C>G (p.Pro644Arg), and **c)** *LRRC3* c.701G>A (p.Arg234Gln).

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.Ala412Val) 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 *DKC1* insertion (III)

3.1 Mutant allele burden and XCI in female *DKC1* mutation carriers

A family with two diseased brothers with an insertion c.1218_1219insCAG in the X-chromosomal gene *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 *DKC1* mutation carriers (**Figure 9**). *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.

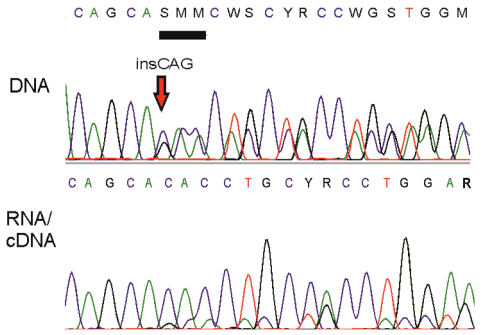

Figure 9. Sequencing chromatograms of aunt 1 showing the heterozygous *DKC1* mutation (DNA). The result was same in all three females. The mutated allele was expressed at the RNA/cDNA level (aunt 1, buccal mucosa). Adapted and reproduced with permission from Hirvonen *et al.*, *Leukemia*, 2019.

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).

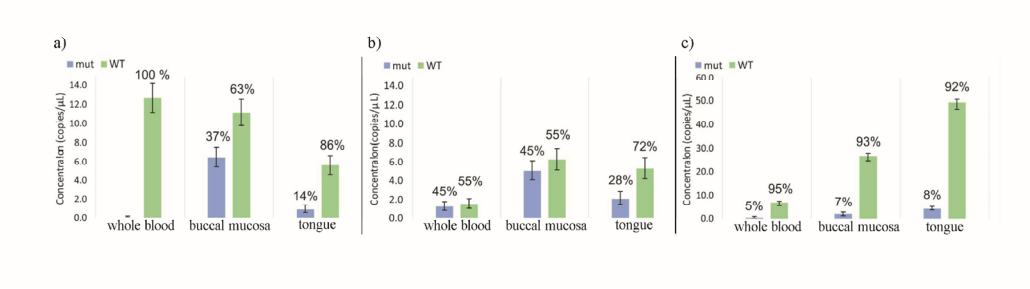

Figure 10. ddPCR analysis showed varying expression of mutant (blue) and wild type (green) *DKC1* expression in different tissues of the **a**) mother, **b**) aunt1, and **c**) aunt 2. Concentration is presented as copies of input cDNA per µL. Adapted and reproduced with permission from Hirvonen *et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

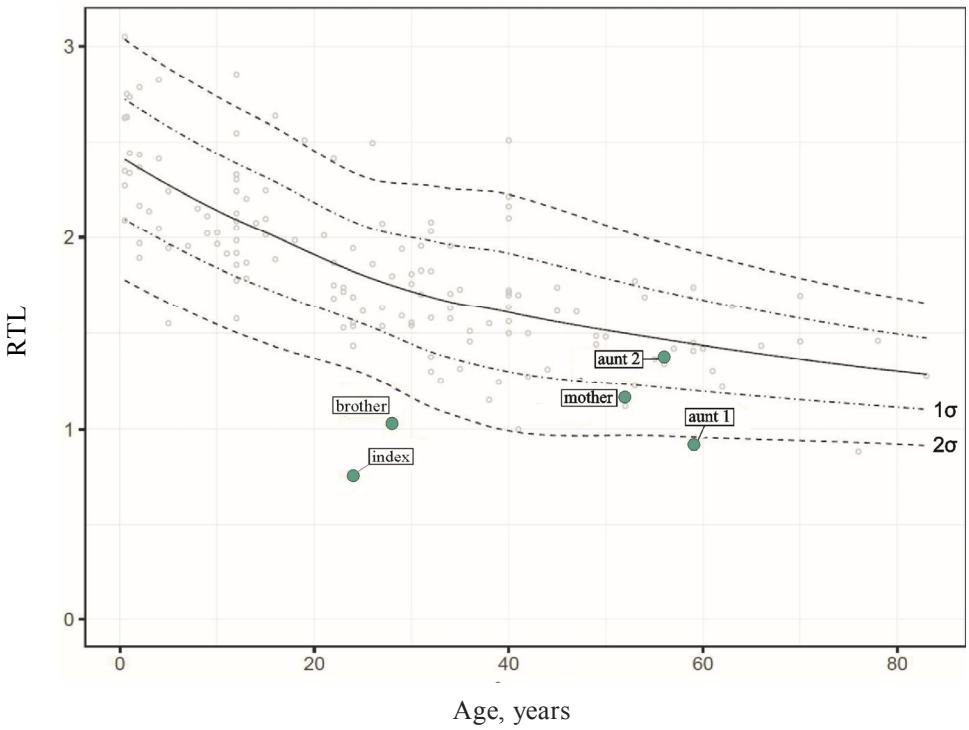

Figure 11. Relationship between age and relative telomere lengths (RTL) of blood cells in DC males and *DKC1* mutation carrier females. Both males and two of the females (mother and aunt 1) had shortened telomere lengths. Adapted and reproduced with permission from Hirvonen *et al., Leukemia*, 2019.

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.

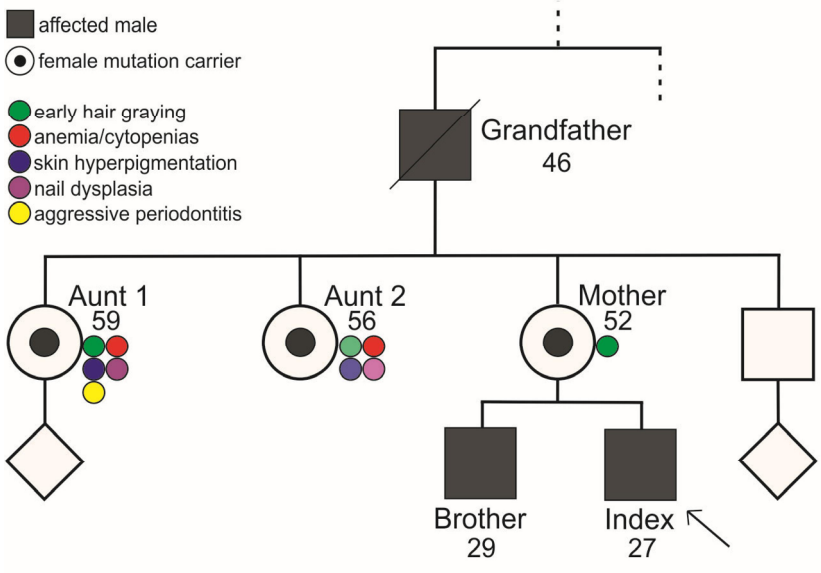

Figure 12. Pedigree and clinical features of c.1218_1219insCAG mutation carrier females. Circles with dots denote mutation carrier females and dark squares represent males with DC. The index case is marked with an arrow, and the ages at the time of examination are shown. Adapted and reproduced with permission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

DISCUSSION

Most cancer types show familial aggregation,[246,247] which together with early age of onset, multiple primary cancers, and specific non-malignant manifestations represent clinical signs of hereditary cancer syndromes.[248] Up to 10% of cancers show familial inheritance,[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.[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.[66] This has also been detected in myeloid malignancies with either germline or somatic mutations in e.g. *RUNX1* and *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,[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.

1 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 *GATA2*,[28] *CEBPA*,[182] *RUNX1*,[85] and *DDX41* mutations.[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.[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.[252,253] The FA pathway intersects with other repair processes and operates in repairing interstrand crosslink lesions in DNA.[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.[252,254] Up to ninety percent of FA patients present with BMF.[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: *DDX41* (p.Met1Ile; start lost);[244] and *SBDS* (c.258+2T>C).[245] The start lost mutation in *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.[238] They recommend the use of standard terminology: pathogenic, likely pathogenic, uncertain significance, likely benign, and benign.[238] 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.[258]

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.[259] Germline mutations in tumor suppressors are known to predispose to cancer because only one additional mutational event is required for tumorigenesis,[68] 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,[155,260,261] and the connection between heterozygous germline missense mutations and familial form of MDS has been shown.[261] *SAMD9L* is conserved and shares 62% sequence identity with the *SAMD9* gene, which is a myeloid tumor suppressor as well.[262] 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.[155,156] 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*.

1.2 *CHEK1*– novel candidate for AML predisposition

Defective DNA repair is one of the hallmarks of cancer.[263] 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.[264] 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.[2]

Three patients in our study were identified to carry a rare VUS in the DNA repair gene *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 *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.[265] 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 *CHEK1* in AML cells predicts reduced overall, event-free, and relapse-free survival.[266,267] *CHEK1* mutations are rare

in cancer, as cells with defective CHK1 are eliminated during tumorigenesis. CHK1 also seems to favor cell proliferation.[265] Nevertheless, LoF mutations have been reported in some cancer types.[268] Earlier studies suggest that *CHEK1* haploinsufficiency may promote cancer, particularly if other mutations are present in relevant genes.[265] 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.[105] 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.[126-129] 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.[126-128] Additional variants in other genes have been found, most notably in *TERT*,[130,131] 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.[269] It cannot, however, be considered as a lineage-determining TF. ATN1 is a nuclear transcriptional corepressor, mainly associated with neurogenerative diseases,[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.[126-129] However, PV is a very rare disease, and the *JAK2* 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.[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.[275] This is most importantly due to bone marrow failure and defects in telomere maintenance and function, which results in exceedingly short telomeres.[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 *DKC1* in X-chromosome. Dyskerin functions in stabilizing the telomerase RNA, thus maintaining its activity.[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.[278] To date, only few reports on female *DKC1* pathogenic mutation carriers exist.[278-281] Here, we studied three symptomatic sisters with a novel heterozygous germline *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.[237]

3.1 Mutant *DKC1* allele expression and telomere length

In order to investigate the XCI status and mutant allele expression in the females, we compared the *DKC1* mutant and wild type allele expression levels in different tissues by utilizing droplet digital PCR. The mutant *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.[282] Families with individuals displaying variable symptoms and telomere lengths despite carrying the same mutations have been reported.[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 *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 *DKC1* mutation carriers.[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 *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 *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.[284] 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.[285] 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.

GERMLINE DIAGNOSTICS IN HEMATOLOGICAL MALIGNANCIES	**NGS techniques**	**Patient**	**Family & other individuals with same germline mutation**	**Therapy strategies**
ADVANTAGES	Identification of new predisposition genes, mutations, and syndromes Relatively fast Increasing knowledge	Finding a reason for the disease Precision medicine, individualized therapy	Early detection of malignancy Follow-up Genetic counseling and risk assessment Increasing knowledge	Precision medicine, tailored treatments Development of new therapy options
CHALLENGES	Interpretation of the variants Lack of knowledge Errors Transition of genetic discoveries to therapeutics	Anxiety and other psychosocial issues Not necessarily present with suspicious family history or symptoms Risks	Anxiety and other psychosocial issues Penetrance cannot necessarily be predicted Risks	Leukemia cannot be prevented – little advances in therapy development due to disease heterogeneity

CONCLUDING REMARKS

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, Katariina 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

1. Gunz FW, Gunz JP, Veale AM, Chapman CJ, Houston IB. Familial leukaemia: a study of 909 families. *Scand J Haematol* 15:117-131, 1975.

2. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 127:2391-2405, 2016.

3. Orkin SH. Diversification of haematopoietic stem cells to specific lineages. *Nat Rev Genet* 1:57-64, 2000.

4. Morrison SJ, Scadden DT. The bone marrow niche for haematopoietic stem cells. *Nature* 505:327-334, 2014.

5. Rieger MA, Schroeder T. Hematopoiesis. *Cold Spring Harb Perspect Biol* 4:2012.

6. Till JE, Mcculloch EA. Hemopoietic stem cell differentiation. *Biochim Biophys Acta* 605:431-459, 1980.

7. Ogawa M. Differentiation and proliferation of hematopoietic stem cells. *Blood* 81:2844-2853, 1993.

8. Metcalf D. Hematopoietic cytokines. *Blood* 111:485-491, 2008.

9. Boulais PE, Frenette PS. Making sense of hematopoietic stem cell niches. *Blood* 125:2621-2629, 2015.

10. Robb L. Cytokine receptors and hematopoietic differentiation. *Oncogene* 26:6715-6723, 2007.

11. Alharbi RA, Pettengell R, Pandha HS, Morgan R. The role of HOX genes in normal hematopoiesis and acute leukemia. *Leukemia* 27:1000-1008, 2013.

12. Hu X, Garcia M, Weng L, Jung X, Murakami JL, Kumar B, Warden CD, Todorov I, Chen CC. Identification of a common mesenchymal stromal progenitor for the adult haematopoietic niche. *Nat Commun* 7:13095, 2016.

13. Mendez-Ferrer S, Michurina TV, Ferraro F, Mazloom AR, Macarthur BD, Lira SA, Scadden DT, Ma'ayan A, Enikolopov GN, Frenette PS. Mesenchymal and haematopoietic stem cells form a unique bone marrow niche. *Nature* 466:829-834, 2010.

14. Zhou BO, Yue R, Murphy MM, Peyer JG, Morrison SJ. Leptin-receptor-expressing mesenchymal stromal cells represent the main source of bone formed by adult bone marrow. *Cell Stem Cell* 15:154-168, 2014.

15. Sugiyama T, Kohara H, Noda M, Nagasawa T. Maintenance of the hematopoietic stem cell pool by CXCL12-CXCR4 chemokine signaling in bone marrow stromal cell niches. *Immunity* 25:977-988, 2006.

16. Asada N, Takeishi S, Frenette PS. Complexity of bone marrow hematopoietic stem cell niche. *Int J Hematol* 106:45-54, 2017.

17. Kunisaki Y, Bruns I, Scheiermann C, Ahmed J, Pinho S, Zhang D, Mizoguchi T, Wei Q, Lucas D, Ito K, Mar JC, Bergman A, Frenette PS. Arteriolar niches maintain haematopoietic stem cell quiescence. *Nature* 502:637-643, 2013.

18. Naveiras O, Nardi V, Wenzel PL, Hauschka PV, Fahey F, Daley GQ. Bone-marrow adipocytes as negative regulators of the haematopoietic microenvironment. *Nature* 460:259-263, 2009.

19. Bruns I, Lucas D, Pinho S, Ahmed J, Lambert MP, Kunisaki Y, Scheiermann C, Schiff L, Poncz M, Bergman A, Frenette PS. Megakaryocytes regulate hematopoietic stem cell quiescence through CXCL4 secretion. *Nat Med* 20:1315-1320, 2014.

20. Winkler IG, Sims NA, Pettit AR, Barbier V, Nowlan B, Helwani F, Poulton IJ, Van Rooijen N, Alexander KA, Raggatt LJ, Levesque JP. Bone marrow macrophages maintain hematopoietic stem cell (HSC) niches and their depletion mobilizes HSCs. *Blood* 116:4815-4828, 2010.

21. Le PM, Andreeff M, Battula VL. Osteogenic niche in the regulation of normal hematopoiesis and leukemogenesis. *Haematologica*; 10.3324/haematol.2018.1970042018.

22. Katayama Y, Battista M, Kao WM, Hidalgo A, Peired AJ, Thomas SA, Frenette PS. Signals from the sympathetic nervous system regulate hematopoietic stem cell egress from bone marrow. *Cell* 124:407-421, 2006.

23. Imanirad P, Dzierzak E. Hypoxia and HIFs in regulating the development of the hematopoietic system. *Blood Cells Mol Dis* 51:256-263, 2013.

24. Wilson A, Laurenti E, Oser G, Van Der Wath RC, Blanco-Bose W, Jaworski M, Offner S, Dunant CF, Eshkind L, Bockamp E, Lio P, Macdonald HR, Trumpp A. Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. *Cell* 135:1118-1129, 2008.

25. Itkin T, Gur-Cohen S, Spencer JA, Schajnovitz A, Ramasamy SK, Kusumbe AP, Ledergor G, Jung Y, Milo I, Poulos MG, Kalinkovich A, Ludin A, Kollet O, Shakhar G, Butler JM, Rafii S, Adams RH, Scadden DT, Lin CP, Lapidot T. Distinct bone marrow blood vessels differentially regulate haematopoiesis. *Nature* 532:323-328, 2016.

26. Akashi K, Traver D, Miyamoto T, Weissman IL. A clonogenic common myeloid progenitor that gives rise to all myeloid lineages. *Nature* 404:193-197, 2000.

27. Gregory T, Yu C, Ma A, Orkin SH, Blobel GA, Weiss MJ. GATA-1 and erythropoietin cooperate to promote erythroid cell survival by regulating bcl-xL expression. *Blood* 94:87-96, 1999.

28. Hahn CN, Chong CE, Carmichael CL, Wilkins EJ, Brautigan PJ, Li XC, Babic M, Lin M, Carmagnac A, Lee YK, Kok CH, Gagliardi L, Friend KL, Ekert PG, Butcher CM, Brown AL, Lewis ID, To LB, Timms AE, Storek J, Moore S, Altree M, Escher R, Bardy PG, Suthers GK, D'andrea RJ, Horwitz MS, Scott HS. Heritable GATA2 mutations associated with familial myelodysplastic syndrome and acute myeloid leukemia. *Nat Genet* 43:1012-1017, 2011.

29. Lloberas J, Soler C, Celada A. The key role of PU.1/SPI-1 in B cells, myeloid cells and macrophages. *Immunol Today* 20:184-189, 1999.

30. Friedman AD. C/EBPalpha in normal and malignant myelopoiesis. *Int J Hematol* 101:330-341, 2015.

31. Challen GA, Goodell MA. Runx1 isoforms show differential expression patterns during hematopoietic development but have similar functional effects in adult hematopoietic stem cells. *Exp Hematol* 38:403-416, 2010.

32. Herault A, Binnewies M, Leong S, Calero-Nieto FJ, Zhang SY, Kang YA, Wang X, Pietras EM, Chu SH, Barry-Holson K, Armstrong S, Gottgens B, Passegue E. Myeloid progenitor cluster formation drives emergency and leukaemic myelopoiesis. *Nature* 544:53-58, 2017.

33. Bainton DF, Farquhar MG. Origin of granules in polymorphonuclear leukocytes. Two types derived from opposite faces of the Golgi complex in developing granulocytes. *J Cell Biol* 28:277-301, 1966.

34. Bainton DF, Ullyot JL, Farquhar MG. The development of neutrophilic polymorphonuclear leukocytes in human bone marrow. *J Exp Med* 134:907-934, 1971.

35. Borregaard N, Sorensen OE, Theilgaard-Monch K. Neutrophil granules: a library of innate immunity proteins. *Trends Immunol* 28:340-345, 2007.

36.

37. Wenger RH. Cellular adaptation to hypoxia: O2-sensing protein hydroxylases, hypoxia-inducible transcription factors, and O2-regulated gene expression. *FASEB J* 16:1151-1162, 2002.

38. Blom B, Spits H. Development of human lymphoid cells. *Annu Rev Immunol* 24:287-320, 2006.

39. Nutt SL, Kee BL. The transcriptional regulation of B cell lineage commitment. *Immunity* 26:715-725, 2007.

40. Kondo M, Weissman IL, Akashi K. Identification of clonogenic common lymphoid progenitors in mouse bone marrow. *Cell* 91:661-672, 1997.

41. Serwold T, Ehrlich LI, Weissman IL. Reductive isolation from bone marrow and blood implicates common lymphoid progenitors as the major source of thymopoiesis. *Blood* 113:807-815, 2009.

42. Masopust D, Schenkel JM. The integration of T cell migration, differentiation and function. *Nat Rev Immunol* 13:309-320, 2013.

43. Stark GR, Darnell JE, Jr. The JAK-STAT pathway at twenty. *Immunity* 36:503-514, 2012.

44. Shenoy AR, Kirschnek S, Hacker G. IL-15 regulates Bcl-2 family members Bim and Mcl-1 through JAK/STAT and PI3K/AKT pathways in T cells. *Eur J Immunol* 44:2500-2507, 2014.

45. Geest CR, Coffer PJ. MAPK signaling pathways in the regulation of hematopoiesis. *J Leukoc Biol* 86:237-250, 2009.

46. Busque L, Buscarlet M, Mollica L, Levine RL. Concise Review: Age-Related Clonal Hematopoiesis: Stem Cells Tempting the Devil. *Stem Cells* 36:1287-1294, 2018.

47. Steensma DP, Bejar R, Jaiswal S, Lindsley RC, Sekeres MA, Hasserjian RP, Ebert BL. Clonal hematopoiesis of indeterminate potential and its distinction from myelodysplastic syndromes. *Blood* 126:9-16, 2015.

48. Sperling AS, Gibson CJ, Ebert BL. The genetics of myelodysplastic syndrome: from clonal haematopoiesis to secondary leukaemia. *Nat Rev Cancer* 17:5-19, 2017.

49. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, Chambert K, Mick E, Neale BM, Fromer M, Purcell SM, Svantesson O, Landen M, Hoglund M, Lehmann S, Gabriel SB, Moran JL, Lander ES, Sullivan PF, Sklar P, Gronberg H, Hultman CM, Mccarroll SA. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N Engl J Med* 371:2477-2487, 2014.

50. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, Lindsley RC, Mermel CH, Burtt N, Chavez A, Higgins JM, Moltchanov V, Kuo FC, Kluk MJ, Henderson B, Kinnunen L, Koistinen HA, Ladenvall C, Getz G, Correa A, Banahan BF, Gabriel S, Kathiresan S, Stringham HM, Mccarthy MI, Boehnke M, Tuomilehto J, Haiman C, Groop L, Atzmon G, Wilson JG, Neuberg D, Altshuler D, Ebert BL. Age-related clonal hematopoiesis associated with adverse outcomes. *N Engl J Med* 371:2488-2498, 2014.

51. Xie M, Lu C, Wang J, Mclellan MD, Johnson KJ, Wendl MC, Mcmichael JF, Schmidt HK, Yellapantula V, Miller CA, Ozenberger BA, Welch JS, Link DC, Walter MJ, Mardis ER, Dipersio JF, Chen F, Wilson RK, Ley TJ, Ding L. Age-related mutations associated with clonal hematopoietic expansion and malignancies. *Nat Med* 20:1472-1478, 2014.

52. Desai P, Mencia-Trinchant N, Savenkov O, Simon MS, Cheang G, Lee S, Samuel M, Ritchie EK, Guzman ML, Ballman KV, Roboz GJ, Hassane DC. Somatic mutations precede acute myeloid leukemia years before diagnosis. *Nat Med* 24:1015-1023, 2018.

53. Abelson S, Collord G, Ng SWK, Weissbrod O, Mendelson Cohen N, Niemeyer E, Barda N, Zuzarte PC, Heisler L, Sundaravadanam Y, Luben R, Hayat S, Wang TT, Zhao Z, Cirlan I, Pugh TJ, Soave D, Ng K, Latimer C, Hardy C, Raine K, Jones D, Hoult D, Britten A, Mcpherson JD, Johansson M, Mbabaali F, Eagles J, Miller JK, Pasternack D, Timms L, Krzyzanowski P, Awadalla P, Costa R, Segal E, Bratman SV, Beer P, Behjati S, Martincorena I, Wang JCY, Bowles KM, Quiros JR, Karakatsani A, La Vecchia C, Trichopoulou A, Salamanca-Fernandez E, Huerta JM, Barricarte A, Travis RC, Tumino R, Masala G, Boeing H, Panico S, Kaaks R, Kramer A, Sieri S, Riboli E, Vineis P, Foll M, Mckay J, Polidoro S, Sala N, Khaw KT, Vermeulen R, Campbell PJ, Papaemmanuil E, Minden MD, Tanay A, Balicer RD, Wareham NJ, Gerstung M, Dick JE, Brennan P, Vassiliou GS, Shlush LI. Prediction of acute myeloid leukaemia risk in healthy individuals. *Nature* 559:400-404, 2018.

54. Adamson JW, Fialkow PJ, Murphy S, Prchal JF, Steinmann L. Polycythemia vera: stem-cell and probable clonal origin of the disease. *N Engl J Med* 295:913-916, 1976.

55. Walter MJ, Shen D, Ding L, Shao J, Koboldt DC, Chen K, Larson DE, Mclellan MD, Dooling D, Abbott R, Fulton R, Magrini V, Schmidt H, Kalicki-Veizer J, O'laughlin M, Fan X, Grillot M, Witowski S, Heath S, Frater JL, Eades W, Tomasson M, Westervelt P, Dipersio JF, Link DC, Mardis ER, Ley TJ, Wilson RK, Graubert TA. Clonal architecture of secondary acute myeloid leukemia. *N Engl J Med* 366:1090-1098, 2012.

56. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ, Jr., Laird PW, Baty JD, Fulton LL, Fulton R, Heath SE, Kalicki-Veizer J, Kandoth C, Klco JM, Koboldt DC, Kanchi KL, Kulkarni S, Lamprecht TL, Larson DE, Lin L, Lu C, Mclellan MD, Mcmichael JF, Payton J, Schmidt H, Spencer DH, Tomasson MH, Wallis JW, Wartman LD, Watson MA, Welch J, Wendl MC, Ally A, Balasundaram M, Birol I, Butterfield Y, Chiu R, Chu A, Chuah E, Chun HJ, Corbett R, Dhalla N, Guin R, He A, Hirst C, Hirst M, Holt RA, Jones S, Karsan A, Lee D, Li HI, Marra MA, Mayo M, Moore RA, Mungall K, Parker J, Pleasance E, Plettner P, Schein J, Stoll D, Swanson L, Tam A, Thiessen N, Varhol R, Wye N, Zhao Y, Gabriel S, Getz G, Sougnez C, Zou L, Leiserson MD, Vandin F, Wu HT, Applebaum F, Baylin SB, Akbani R, Broom BM, Chen K, Motter TC, Nguyen K, Weinstein JN, Zhang N, Ferguson ML, Adams C, Black A, Bowen J, Gastier-Foster J, Grossman T, Lichtenberg T, Wise L, Davidsen T, Demchok JA, Shaw KR, Sheth M, Sofia HJ, Yang L, Downing JR, Eley G. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368:2059-2074, 2013.

57. Shlush LI, Zandi S, Mitchell A, Chen WC, Brandwein JM, Gupta V, Kennedy JA, Schimmer AD, Schuh AC, Yee KW, Mcleod JL, Doedens M, Medeiros JJ, Marke R, Kim HJ, Lee K, Mcpherson JD, Hudson TJ, Consortium HP-LGP, Brown AM, Yousif F, Trinh QM, Stein LD, Minden MD, Wang JC, Dick JE. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature* 506:328-333, 2014.

58. Kreso A, Dick JE. Evolution of the cancer stem cell model. *Cell Stem Cell* 14:275-291, 2014.

59. Colmone A, Amorim M, Pontier AL, Wang S, Jablonski E, Sipkins DA. Leukemic cells create bone marrow niches that disrupt the behavior of normal hematopoietic progenitor cells. *Science* 322:1861-1865, 2008.

60. Hawkins ED, Duarte D, Akinduro O, Khorshed RA, Passaro D, Nowicka M, Straszkowski L, Scott MK, Rothery S, Ruivo N, Foster K, Waibel M, Johnstone RW, Harrison SJ, Westerman DA, Quach H, Gribben J, Robinson MD, Purton LE, Bonnet D, Lo Celso C. T-cell acute leukaemia exhibits dynamic interactions with bone marrow microenvironments. *Nature* 538:518-522, 2016.

61. Kumar B, Garcia M, Weng L, Jung X, Murakami JL, Hu X, Mcdonald T, Lin A, Kumar AR, Digiusto DL, Stein AS, Pullarkat VA, Hui SK, Carlesso N, Kuo YH, Bhatia R, Marcucci G, Chen CC. Acute myeloid leukemia transforms the bone marrow niche into a leukemia-permissive microenvironment through exosome secretion. *Leukemia* 32:575-587, 2018.

62. Schepers K, Campbell TB, Passegue E. Normal and leukemic stem cell niches: insights and therapeutic opportunities. *Cell Stem Cell* 16:254-267, 2015.

63. Raaijmakers MH, Mukherjee S, Guo S, Zhang S, Kobayashi T, Schoonmaker JA, Ebert BL, Al-Shahrour F, Hasserjian RP, Scadden EO, Aung Z, Matza M, Merkenschlager M, Lin C, Rommens JM, Scadden DT. Bone progenitor dysfunction induces myelodysplasia and secondary leukaemia. *Nature* 464:852-857, 2010.

64. Krause DS, Fulzele K, Catic A, Sun CC, Dombkowski D, Hurley MP, Lezeau S, Attar E, Wu JY, Lin HY, Divieti-Pajevic P, Hasserjian RP, Schipani E, Van Etten RA, Scadden DT. Differential regulation of myeloid leukemias by the bone marrow microenvironment. *Nat Med* 19:1513-1517, 2013.

65. Rieger CT, Fiegl M. Microenvironmental oxygen partial pressure in acute myeloid leukemia: Is there really a role for hypoxia? *Exp Hematol* 44:578-582, 2016.

66. Rahman N. Realizing the promise of cancer predisposition genes. *Nature* 505:302-308, 2014.

67. Nagy R, Sweet K, Eng C. Highly penetrant hereditary cancer syndromes. *Oncogene* 23:6445-6470, 2004.

68. Knudson AG. Two genetic hits (more or less) to cancer. *Nat Rev Cancer* 1:157-162, 2001.

69. Nowell PC. The clonal evolution of tumor cell populations. *Science* 194:23-28, 1976.

70. Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature* 458:719-724, 2009.

71. Vogelstein B, Kinzler KW. The Path to Cancer --Three Strikes and You're Out. *N Engl J Med* 373:1895-1898, 2015.

72. Vogelstein B, Kinzler KW. Cancer genes and the pathways they control. *Nat Med* 10:789-799, 2004.

73. Kinzler KW, Vogelstein B. Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature* 386:761, 763, 1997.

74. Ryland GL, Doyle MA, Goode D, Boyle SE, Choong DY, Rowley SM, Li J, Australian Ovarian Cancer Study G, Bowtell DD, Tothill RW, Campbell IG, Gorringe KL. Loss of heterozygosity: what is it good for? *BMC Med Genomics* 8:45, 2015.

75. Macleod K. Tumor suppressor genes. *Curr Opin Genet Dev* 10:81-93, 2000.

76. Stacey SN, Sulem P, Jonasdottir A, Masson G, Gudmundsson J, Gudbjartsson DF, Magnusson OT, Gudjonsson SA, Sigurgeirsson B, Thorisdottir K, Ragnarsson R, Benediktsdottir KR, Nexo BA, Tjonneland A, Overvad K, Rudnai P, Gurzau E, Koppova K, Hemminki K, Corredera C, Fuentelsaz V, Grasa P, Navarrete S, Fuertes F, Garcia-Prats MD, Sanambrosio E, Panadero A, De Juan A, Garcia A, Rivera F, Planelles D, Soriano V, Requena C, Aben KK, Van Rossum MM, Cremers RG, Van Oort IM, Van Spronsen DJ, Schalken JA, Peters WH, Helfand BT, Donovan JL, Hamdy FC, Badescu D, Codreanu O, Jinga M, Csiki IE, Constantinescu V, Badea P, Mates IN, Dinu DE, Constantin A, Mates D, Kristjansdottir S, Agnarsson BA, Jonsson E, Barkardottir RB, Einarsson GV, Sigurdsson F, Moller PH, Stefansson T, Valdimarsson T, Johannsson OT, Sigurdsson H, Jonsson T, Jonasson JG, Tryggvadottir L, Rice T, Hansen HM, Xiao Y, Lachance DH, Bp ON, Kosel ML, Decker PA, Thorleifsson G, Johannsdottir H, Helgadottir HT, Sigurdsson A, Steinthorsdottir V, Lindblom A, Swedish Low-Risk Colorectal Cancer Study G, Sandler RS, Keku TO, Banasik K, Jorgensen T, Witte DR, Hansen T, Pedersen O, Jinga V, Neal DE, Catalona WJ, Wrensch M, Wiencke J, Jenkins RB, Nagore E, Vogel U, Kiemeney LA, Kumar R, Mayordomo JI, Olafsson JH, Kong A, Thorsteinsdottir U, Rafnar T, Stefansson K. A germline variant in the TP53 polyadenylation signal confers cancer susceptibility. *Nat Genet* 43:1098-1103, 2011.

77. Correa H. Li-Fraumeni Syndrome. *J Pediatr Genet* 5:84-88, 2016.

78. Glaire MA, Brown M, Church DN, Tomlinson I. Cancer predisposition syndromes: lessons for truly precision medicine. *J Pathol* 241:226-235, 2017.

79. Hobert JA, Eng C. PTEN hamartoma tumor syndrome: an overview. *Genet Med* 11:687-694, 2009.

80. Murati A, Brecqueville M, Devillier R, Mozziconacci MJ, Gelsi-Boyer V, Birnbaum D. Myeloid malignancies: mutations, models and management. *BMC Cancer* 12:304, 2012.

81. Mardis ER, Ding L, Dooling DJ, Larson DE, Mclellan MD, Chen K, Koboldt DC, Fulton RS, Delehaunty KD, Mcgrath SD, Fulton LA, Locke DP, Magrini VJ, Abbott RM, Vickery TL, Reed JS, Robinson JS, Wylie T, Smith SM, Carmichael L, Eldred JM, Harris CC, Walker J, Peck JB, Du F, Dukes AF, Sanderson GE, Brummett AM, Clark E, Mcmichael JF, Meyer RJ, Schindler JK, Pohl CS, Wallis JW, Shi X, Lin L, Schmidt H, Tang Y, Haipek C, Wiechert ME, Ivy JV, Kalicki J, Elliott G, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson MA, Baty J, Heath S, Shannon WD, Nagarajan R, Link DC, Walter MJ, Graubert TA, Dipersio JF, Wilson RK, Ley TJ. Recurring mutations found by sequencing an acute myeloid leukemia genome. *N Engl J Med* 361:1058-1066, 2009.

82. Welch JS, Ley TJ, Link DC, Miller CA, Larson DE, Koboldt DC, Wartman LD, Lamprecht TL, Liu F, Xia J, Kandoth C, Fulton RS, Mclellan MD, Dooling DJ, Wallis JW, Chen K, Harris CC, Schmidt HK, Kalicki-Veizer JM, Lu C, Zhang Q, Lin L, O'laughlin MD, Mcmichael JF, Delehaunty KD, Fulton LA, Magrini VJ, Mcgrath SD, Demeter RT, Vickery TL, Hundal J, Cook LL, Swift GW, Reed JP, Alldredge PA, Wylie TN, Walker JR, Watson MA, Heath SE, Shannon WD, Varghese N, Nagarajan R, Payton JE, Baty JD, Kulkarni S, Klco JM, Tomasson MH, Westervelt P, Walter MJ, Graubert TA, Dipersio JF, Ding L, Mardis ER, Wilson RK. The origin and evolution of mutations in acute myeloid leukemia. *Cell* 150:264-278, 2012.

83. Korn C, Mendez-Ferrer S. Myeloid malignancies and the microenvironment. *Blood* 129:811-822, 2017.

84. Godley LA, Shimamura A. Genetic predisposition to hematologic malignancies: management and surveillance. *Blood* 130:424-432, 2017.

85. Song WJ, Sullivan MG, Legare RD, Hutchings S, Tan X, Kufrin D, Ratajczak J, Resende IC, Haworth C, Hock R, Loh M, Felix C, Roy DC, Busque L, Kurnit D, Willman C, Gewirtz AM, Speck NA, Bushweller JH, Li FP, Gardiner K, Poncz M, Maris JM, Gilliland DG. Haploinsufficiency of CBFA2 causes familial thrombocytopenia with propensity to develop acute myelogenous leukaemia. *Nat Genet* 23:166-175, 1999.

86. Akpan IJ, Osman AEG, Drazer MW, Godley LA. Hereditary Myelodysplastic Syndrome and Acute Myeloid Leukemia: Diagnosis, Questions, and Controversies. *Curr Hematol Malig Rep*; 10.1007/s11899-018-0473-72018.

87. Babushok DV, Bessler M, Olson TS. Genetic predisposition to myelodysplastic syndrome and acute myeloid leukemia in children and young adults. *Leuk Lymphoma* 57:520-536, 2016.

88. Delhommeau F, Pisani DF, James C, Casadevall N, Constantinescu S, Vainchenker W. Oncogenic mechanisms in myeloproliferative disorders. *Cell Mol Life Sci* 63:2939-2953, 2006.

89. Campbell PJ, Green AR. The myeloproliferative disorders. *N Engl J Med* 355:2452-2466, 2006.

90. Spivak JL. Myeloproliferative Neoplasms. *N Engl J Med* 377:895-896, 2017.

91. Rumi E, Cazzola M. Diagnosis, risk stratification, and response evaluation in classical myeloproliferative neoplasms. *Blood* 129:680-692, 2017.

92. Yogarajah M, Tefferi A. Leukemic Transformation in Myeloproliferative Neoplasms: A Literature Review on Risk, Characteristics, and Outcome. *Mayo Clin Proc* 92:1118-1128, 2017.

93. Tefferi A, Guglielmelli P, Larson DR, Finke C, Wassie EA, Pieri L, Gangat N, Fjerza R, Belachew AA, Lasho TL, Ketterling RP, Hanson CA, Rambaldi A, Finazzi G, Thiele J, Barbui T, Pardanani A, Vannucchi AM. Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. *Blood* 124:2507-2513; quiz 2615, 2014.

94. Tefferi A, Rumi E, Finazzi G, Gisslinger H, Vannucchi AM, Rodeghiero F, Randi ML, Vaidya R, Cazzola M, Rambaldi A, Gisslinger B, Pieri L, Ruggeri M, Bertozzi I, Sulai NH, Casetti I, Carobbio A, Jeryczynski G, Larson DR, Mullauer L, Pardanani A, Thiele J, Passamonti F, Barbui T. Survival and prognosis among 1545 patients with contemporary polycythemia vera: an international study. *Leukemia* 27:1874-1881, 2013.

95. Abdel-Wahab OI, Levine RL. Primary myelofibrosis: update on definition, pathogenesis, and treatment. *Annu Rev Med* 60:233-245, 2009.

96. Tefferi A. Myeloproliferative neoplasms: A decade of discoveries and treatment advances. *Am J Hematol* 91:50-58, 2016.

97. Vainchenker W, Kralovics R. Genetic basis and molecular pathophysiology of classical myeloproliferative neoplasms. *Blood* 129:667-679, 2017.

98. Vainchenker W, Delhommeau F, Constantinescu SN, Bernard OA. New mutations and pathogenesis of myeloproliferative neoplasms. *Blood* 118:1723-1735, 2011.

99. Kralovics R, Passamonti F, Buser AS, Teo SS, Tiedt R, Passweg JR, Tichelli A, Cazzola M, Skoda RC. A gain-of-function mutation of JAK2 in myeloproliferative disorders. *N Engl J Med* 352:1779-1790, 2005.

100. Baxter EJ, Scott LM, Campbell PJ, East C, Fourouclas N, Swanton S, Vassiliou GS, Bench AJ, Boyd EM, Curtin N, Scott MA, Erber WN, Green AR, Cancer Genome P. Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders. *Lancet* 365:1054-1061, 2005.

101. James C, Ugo V, Le Couedic JP, Staerk J, Delhommeau F, Lacout C, Garcon L, Raslova H, Berger R, Bennaceur-Griscelli A, Villeval JL, Constantinescu SN, Casadevall N, Vainchenker W. A unique clonal JAK2 mutation leading to constitutive signalling causes polycythaemia vera. *Nature* 434:1144-1148, 2005.

102. Levine RL, Wadleigh M, Cools J, Ebert BL, Wernig G, Huntly BJ, Boggon TJ, Wlodarska I, Clark JJ, Moore S, Adelsperger J, Koo S, Lee JC, Gabriel S, Mercher T, D'andrea A, Frohling S, Dohner K, Marynen P, Vandenberghe P, Mesa RA, Tefferi A, Griffin JD, Eck MJ, Sellers WR, Meyerson M, Golub TR, Lee SJ, Gilliland DG. Activating mutation in the tyrosine kinase JAK2 in polycythemia vera, essential thrombocythemia, and myeloid metaplasia with myelofibrosis. *Cancer Cell* 7:387-397, 2005.

103. Levine RL, Belisle C, Wadleigh M, Zahrieh D, Lee S, Chagnon P, Gilliland DG, Busque L. X-inactivation-based clonality analysis and quantitative JAK2V617F assessment reveal a strong association between clonality and

JAK2V617F in PV but not ET/MMM, and identifies a subset of JAK2V617F-negative ET and MMM patients with clonal hematopoiesis. *Blood* 107:4139-4141, 2006.

104. Lippert E, Boissinot M, Kralovics R, Girodon F, Dobo I, Praloran V, Boiret-Dupre N, Skoda RC, Hermouet S. The JAK2-V617F mutation is frequently present at diagnosis in patients with essential thrombocythemia and polycythemia vera. *Blood* 108:1865-1867, 2006.

105. Tashi T, Swierczek S, Prchal JT. Familial MPN Predisposition. *Curr Hematol Malig Rep* 12:442-447, 2017.

106. Kralovics R, Teo SS, Li S, Theocharides A, Buser AS, Tichelli A, Skoda RC. Acquisition of the V617F mutation of JAK2 is a late genetic event in a subset of patients with myeloproliferative disorders. *Blood* 108:1377-1380, 2006.

107. Schaub FX, Jager R, Looser R, Hao-Shen H, Hermouet S, Girodon F, Tichelli A, Gisslinger H, Kralovics R, Skoda RC. Clonal analysis of deletions on chromosome 20q and JAK2-V617F in MPD suggests that del20q acts independently and is not one of the predisposing mutations for JAK2-V617F. *Blood* 113:2022-2027, 2009.

108. Klampfl T, Gisslinger H, Harutyunyan AS, Nivarthi H, Rumi E, Milosevic JD, Them NC, Berg T, Gisslinger B, Pietra D, Chen D, Vladimer GI, Bagienski K, Milanesi C, Casetti IC, Sant'antonio E, Ferretti V, Elena C, Schischlik F, Cleary C, Six M, Schalling M, Schonegger A, Bock C, Malcovati L, Pascutto C, Superti-Furga G, Cazzola M, Kralovics R. Somatic mutations of calreticulin in myeloproliferative neoplasms. *N Engl J Med* 369:2379-2390, 2013.

109. Nangalia J, Massie CE, Baxter EJ, Nice FL, Gundem G, Wedge DC, Avezov E, Li J, Kollmann K, Kent DG, Aziz A, Godfrey AL, Hinton J, Martincorena I, Van Loo P, Jones AV, Guglielmelli P, Tarpey P, Harding HP, Fitzpatrick JD, Goudie CT, Ortmann CA, Loughran SJ, Raine K, Jones DR, Butler AP, Teague JW, O'meara S, Mclaren S, Bianchi M, Silber Y, Dimitropoulou D, Bloxham D, Mudie L, Maddison M, Robinson B, Keohane C, Maclean C, Hill K, Orchard K, Tauro S, Du MQ, Greaves M, Bowen D, Huntly BJP, Harrison CN, Cross NCP, Ron D, Vannucchi AM, Papaemmanuil E, Campbell PJ, Green AR. Somatic CALR mutations in myeloproliferative neoplasms with nonmutated JAK2. *N Engl J Med* 369:2391-2405, 2013.

110. Pikman Y, Lee BH, Mercher T, Mcdowell E, Ebert BL, Gozo M, Cuker A, Wernig G, Moore S, Galinsky I, Deangelo DJ, Clark JJ, Lee SJ, Golub TR, Wadleigh M, Gilliland DG, Levine RL. MPLW515L is a novel somatic activating mutation in myelofibrosis with myeloid metaplasia. *PLoS Med* 3:e270, 2006.

111. Beer PA, Campbell PJ, Scott LM, Bench AJ, Erber WN, Bareford D, Wilkins BS, Reilly JT, Hasselbalch HC, Bowman R, Wheatley K, Buck G, Harrison CN, Green AR. MPL mutations in myeloproliferative disorders: analysis of the PT-1 cohort. *Blood* 112:141-149, 2008.

112. Boyd EM, Bench AJ, Goday-Fernandez A, Anand S, Vaghela KJ, Beer P, Scott MA, Bareford D, Green AR, Huntly B, Erber WN. Clinical utility of routine MPL exon 10 analysis in the diagnosis of essential thrombocythaemia and primary myelofibrosis. *Br J Haematol* 149:250-257, 2010.

113. Chaligne R, Tonetti C, Besancenot R, Roy L, Marty C, Mossuz P, Kiladjian JJ, Socie G, Bordessoule D, Le Bousse-Kerdiles MC, Vainchenker W, Giraudier S. New mutations of MPL in primitive myelofibrosis: only the MPL W515 mutations promote a G1/S-phase transition. *Leukemia* 22:1557-1566, 2008.

114. Tefferi A, Guglielmelli P, Larson DR, Finke C, Wassie EA, Pieri L, Gangat N, Fjerza R, Belachew AA, Lasho TL, Ketterling RP, Hanson CA, Rambaldi A, Finazzi G, Thiele J, Barbui T, Pardanani A, Vannucchi AM. Long-term survival and blast transformation in molecularly annotated essential thrombocythemia, polycythemia vera, and myelofibrosis. *Blood* 124:2507-2513; quiz 2615, 2014.

115. Lundberg P, Karow A, Nienhold R, Looser R, Hao-Shen H, Nissen I, Girsberger S, Lehmann T, Passweg J, Stern M, Beisel C, Kralovics R, Skoda RC. Clonal evolution and clinical correlates of somatic mutations in myeloproliferative neoplasms. *Blood* 123:2220-2228, 2014.

116. Wang L, Swierczek SI, Drummond J, Hickman K, Kim SJ, Walker K, Doddapaneni H, Muzny DM, Gibbs RA, Wheeler DA, Prchal JT. Whole-exome sequencing of polycythemia vera revealed novel driver genes and somatic mutation shared by T cells and granulocytes. *Leukemia* 28:935-938, 2014.

117. Stein BL, Saraf S, Sobol U, Halpern A, Shammo J, Rondelli D, Michaelis L, Odenike O, Rademaker A, Zakarija A, Mcmahon B, Spivak JL, Moliterno AR. Age-related differences in disease characteristics and clinical outcomes in polycythemia vera. *Leuk Lymphoma* 54:1989-1995, 2013.

118. Stein BL, Oh ST, Berenzon D, Hobbs GS, Kremyanskaya M, Rampal RK, Abboud CN, Adler K, Heaney ML, Jabbour EJ, Komrokji RS, Moliterno AR, Ritchie EK, Rice L, Mascarenhas J, Hoffman R. Polycythemia Vera: An Appraisal of the Biology and Management 10 Years After the Discovery of JAK2 V617F. *J Clin Oncol* 33:3953-3960, 2015.

119. Nussenzveig RH, Swierczek SI, Jelinek J, Gaikwad A, Liu E, Verstovsek S, Prchal JF, Prchal JT. Polycythemia vera is not initiated by JAK2V617F mutation. *Exp Hematol* 35:32-38, 2007.

120. Scott LM, Tong W, Levine RL, Scott MA, Beer PA, Stratton MR, Futreal PA, Erber WN, Mcmullin MF, Harrison CN, Warren AJ, Gilliland DG, Lodish HF, Green AR. JAK2 exon 12 mutations in polycythemia vera and idiopathic erythrocytosis. *N Engl J Med* 356:459-468, 2007.

121. Pietra D, Li S, Brisci A, Passamonti F, Rumi E, Theocharides A, Ferrari M, Gisslinger H, Kralovics R, Cremonesi L, Skoda R, Cazzola M. Somatic mutations of JAK2 exon 12 in patients with JAK2 (V617F)-negative myeloproliferative disorders. *Blood* 111:1686-1689, 2008.

122. Passamonti F, Elena C, Schnittger S, Skoda RC, Green AR, Girodon F, Kiladjian JJ, Mcmullin MF, Ruggeri M, Besses C, Vannucchi AM, Lippert E, Gisslinger H, Rumi E, Lehmann T, Ortmann CA, Pietra D, Pascutto C, Haferlach T, Cazzola M. Molecular and clinical features of the myeloproliferative neoplasm associated with JAK2 exon 12 mutations. *Blood* 117:2813-2816, 2011.

123. Kralovics R, Guan Y, Prchal JT. Acquired uniparental disomy of chromosome 9p is a frequent stem cell defect in polycythemia vera. *Exp Hematol* 30:229-236, 2002.

124. Bellanne-Chantelot C, Chaumarel I, Labopin M, Bellanger F, Barbu V, De Toma C, Delhommeau F, Casadevall N, Vainchenker W, Thomas G, Najman A. Genetic and clinical implications of the Val617Phe JAK2 mutation in 72 families with myeloproliferative disorders. *Blood* 108:346-352, 2006.

125. Olcaydu D, Rumi E, Harutyunyan A, Passamonti F, Pietra D, Pascutto C, Berg T, Jager R, Hammond E, Cazzola M, Kralovics R. The role of the JAK2 GGCC haplotype and the TET2 gene in familial myeloproliferative neoplasms. *Haematologica* 96:367-374, 2011.

126. Jones AV, Chase A, Silver RT, Oscier D, Zoi K, Wang YL, Cario H, Pahl HL, Collins A, Reiter A, Grand F, Cross NC. JAK2 haplotype is a major risk factor for the development of myeloproliferative neoplasms. *Nat Genet* 41:446-449, 2009.

127. Kilpivaara O, Mukherjee S, Schram AM, Wadleigh M, Mullally A, Ebert BL, Bass A, Marubayashi S, Heguy A, Garcia-Manero G, Kantarjian H, Offit K, Stone RM, Gilliland DG, Klein RJ, Levine RL. A germline JAK2 SNP is associated with predisposition to the development of JAK2(V617F)-positive myeloproliferative neoplasms. *Nat Genet* 41:455-459, 2009.

128. Olcaydu D, Harutyunyan A, Jager R, Berg T, Gisslinger B, Pabinger I, Gisslinger H, Kralovics R. A common JAK2 haplotype confers susceptibility to myeloproliferative neoplasms. *Nat Genet* 41:450-454, 2009.

129. Trifa AP, Cucuianu A, Petrov L, Urian L, Militaru MS, Dima D, Pop IV, Popp RA. The G allele of the JAK2 rs10974944 SNP, part of JAK2 46/1 haplotype, is strongly associated with JAK2 V617F-positive myeloproliferative neoplasms. *Ann Hematol* 89:979-983, 2010.

130. Tapper W, Jones AV, Kralovics R, Harutyunyan AS, Zoi K, Leung W, Godfrey AL, Guglielmelli P, Callaway A, Ward D, Aranaz P, White HE, Waghorn K, Lin F, Chase A, Baxter EJ, Maclean C, Nangalia J, Chen E, Evans P, Short M, Jack A, Wallis L, Oscier D, Duncombe AS, Schuh A, Mead AJ, Griffiths M, Ewing J, Gale RE, Schnittger S, Haferlach T, Stegelmann F, Dohner K, Grallert H, Strauch K, Tanaka T, Bandinelli S, Giannopoulos A, Pieri L, Mannarelli C, Gisslinger H, Barosi G, Cazzola M, Reiter A, Harrison C, Campbell P, Green AR, Vannucchi A, Cross NC. Genetic variation at MECOM, TERT, JAK2 and HBS1L-MYB predisposes to myeloproliferative neoplasms. *Nat Commun* 6:6691, 2015.

131. Oddsson A, Kristinsson SY, Helgason H, Gudbjartsson DF, Masson G, Sigurdsson A, Jonasdottir A, Jonasdottir A, Steingrimsdottir H, Vidarsson B, Reykdal S, Eyjolfsson GI, Olafsson I, Onundarson PT, Runarsson G, Sigurdardottir O, Kong A, Rafnar T, Sulem P, Thorsteinsdottir U, Stefansson K. The germline sequence variant rs2736100_C in TERT associates with myeloproliferative neoplasms. *Leukemia* 28:1371-1374, 2014.

132. Hinds DA, Barnholt KE, Mesa RA, Kiefer AK, Do CB, Eriksson N, Mountain JL, Francke U, Tung JY, Nguyen HM, Zhang H, Gojenola L, Zehnder JL, Gotlib J. Germ line variants predispose to both JAK2 V617F clonal hematopoiesis and myeloproliferative neoplasms. *Blood* 128:1121-1128, 2016.

133. Marty C, Saint-Martin C, Pecquet C, Grosjean S, Saliba J, Mouton C, Leroy E, Harutyunyan AS, Abgrall JF, Favier R, Toussaint A, Solary E, Kralovics R, Constantinescu SN, Najman A, Vainchenker W, Plo I, Bellanne-Chantelot C. Germ-line JAK2 mutations in the kinase domain are responsible for hereditary thrombocytosis and are resistant to JAK2 and HSP90 inhibitors. *Blood* 123:1372-1383, 2014.

134. Ding J, Komatsu H, Wakita A, Kato-Uranishi M, Ito M, Satoh A, Tsuboi K, Nitta M, Miyazaki H, Iida S, Ueda R. Familial essential thrombocythemia associated with a dominant-positive activating mutation of the c-MPL gene, which encodes for the receptor for thrombopoietin. *Blood* 103:4198-4200, 2004.

135. Harutyunyan AS, Giambruno R, Krendl C, Stukalov A, Klampfl T, Berg T, Chen D, Milosevic Feenstra JD, Jager R, Gisslinger B, Gisslinger H, Rumi E, Passamonti F, Pietra D, Muller AC, Parapatics K, Breitwieser FP, Herrmann R, Colinge J, Bennett KL, Superti-Furga G, Cazzola M, Hammond E, Kralovics R. Germline RBBP6 mutations in familial myeloproliferative neoplasms. *Blood* 127:362-365, 2016.

136. Saliba J, Saint-Martin C, Di Stefano A, Lenglet G, Marty C, Keren B, Pasquier F, Valle VD, Secardin L, Leroy G, Mahfoudhi E, Grosjean S, Droin N, Diop M, Dessen P, Charrier S, Palazzo A, Merlevede J, Meniane JC, Delaunay-Darivon C, Fuseau P, Isnard F, Casadevall N, Solary E, Debili N, Bernard OA, Raslova H, Najman A, Vainchenker W, Bellanne-Chantelot C, Plo I. Germline duplication of ATG2B and GSKIP predisposes to familial myeloid malignancies. *Nat Genet* 47:1131-1140, 2015.

137. Cario H, Goerttler PS, Steimle C, Levine RL, Pahl HL. The JAK2V617F mutation is acquired secondary to the predisposing alteration in familial polycythaemia vera. *Br J Haematol* 130:800-801, 2005.

138. Hemminki K, Jiang Y. Familial polycythemia vera: results from the Swedish Family-Cancer Database. *Leukemia* 15:1313-1315, 2001.

139. Kralovics R, Stockton DW, Prchal JT. Clonal hematopoiesis in familial polycythemia vera suggests the involvement of multiple mutational events in the early pathogenesis of the disease. *Blood* 102:3793-3796, 2003.

140. Lanikova L, Babosova O, Swierczek S, Wang L, Wheeler DA, Divoky V, Korinek V, Prchal JT. Coexistence of gain-of-function JAK2 germ line mutations with JAK2V617F in polycythemia vera. *Blood* 128:2266-2270, 2016.

141. Janssen JW, Buschle M, Layton M, Drexler HG, Lyons J, Van Den Berghe H, Heimpel H, Kubanek B, Kleihauer E, Mufti GJ, Et Al. Clonal analysis of myelodysplastic syndromes: evidence of multipotent stem cell origin. *Blood* 73:248-254, 1989.

142. Visconte V, Tiu RV, Rogers HJ. Pathogenesis of myelodysplastic syndromes: an overview of molecular and non-molecular aspects of the disease. *Blood Res* 49:216-227, 2014.

143. Bejar R. Prognostic models in myelodysplastic syndromes. *Hematology Am Soc Hematol Educ Program* 2013:504-510, 2013.

144. Busque L, Patel JP, Figueroa ME, Vasanthakumar A, Provost S, Hamilou Z, Mollica L, Li J, Viale A, Heguy A, Hassimi M, Socci N, Bhatt PK, Gonen M, Mason CE, Melnick A, Godley LA, Brennan CW, Abdel-Wahab O, Levine RL. Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis. *Nat Genet* 44:1179-1181, 2012.

145. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, Schnittger S, Sanada M, Kon A, Alpermann T, Yoshida K, Roller A, Nadarajah N, Shiraishi Y, Shiozawa Y, Chiba K, Tanaka H, Koeffler HP, Klein HU, Dugas M, Aburatani H, Kohlmann A, Miyano S, Haferlach C, Kern W, Ogawa S. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. *Leukemia* 28:241-247, 2014.

146. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, Yoon CJ, Ellis P, Wedge DC, Pellagatti A, Shlien A, Groves MJ, Forbes SA, Raine K, Hinton J, Mudie LJ, Mclaren S, Hardy C, Latimer C, Della Porta MG, O'meara S, Ambaglio I, Galli A, Butler AP, Walldin G, Teague JW, Quek L, Sternberg A, Gambacorti-Passerini C, Cross NC, Green AR, Boultwood J, Vyas P, Hellstrom-Lindberg E, Bowen D, Cazzola M, Stratton MR, Campbell PJ, Chronic Myeloid Disorders Working Group of the International Cancer Genome C. Clinical and biological implications of driver mutations in myelodysplastic syndromes. *Blood* 122:3616-3627; quiz 3699, 2013.

147. Chen J, Kao YR, Sun D, Todorova TI, Reynolds D, Narayanagari SR, Montagna C, Will B, Verma A, Steidl AU. Myelodysplastic syndrome progression to acute myeloid leukemia at the stem cell level. *Nat Med*; 10.1038/s41591-018-0267-42018.

148. Koeffler HP, Leong G. Preleukemia: one name, many meanings. *Leukemia* 31:534-542, 2017.

149. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N, Gundem G, Van Loo P, Martincorena I, Ganly P, Mudie L, Mclaren S, O'meara S, Raine K, Jones DR, Teague JW, Butler AP, Greaves MF, Ganser A, Dohner K, Schlenk RF, Dohner H, Campbell PJ. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 374:2209-2221, 2016.

150. Goldin LR, Kristinsson SY, Liang XS, Derolf AR, Landgren O, Bjorkholm M. Familial aggregation of acute myeloid leukemia and myelodysplastic syndromes. *J Clin Oncol* 30:179-183, 2012.

151. Babushok DV, Bessler M. Genetic predisposition syndromes: when should they be considered in the work-up of MDS? *Best Pract Res Clin Haematol* 28:55-68, 2015.

152. Nickels EM, Soodalter J, Churpek JE, Godley LA. Recognizing familial myeloid leukemia in adults. *Ther Adv Hematol* 4:254-269, 2013.

153. Bannon SA, Dinardo CD. Hereditary Predispositions to Myelodysplastic Syndrome. *Int J Mol Sci* 17:2016.

154. Wang X, Muramatsu H, Okuno Y, Sakaguchi H, Yoshida K, Kawashima N, Xu Y, Shiraishi Y, Chiba K, Tanaka H, Saito S, Nakazawa Y, Masunari T, Hirose T, Elmahdi S, Narita A, Doisaki S, Ismael O, Makishima H, Hama A, Miyano S, Takahashi Y, Ogawa S, Kojima S. GATA2 and secondary mutations in familial myelodysplastic syndromes and pediatric myeloid malignancies. *Haematologica* 100:e398-401, 2015.

155. Chen DH, Below JE, Shimamura A, Keel SB, Matsushita M, Wolff J, Sul Y, Bonkowski E, Castella M, Taniguchi T, Nickerson D, Papayannopoulou T, Bird TD, Raskind WH. Ataxia-Pancytopenia Syndrome Is Caused by Missense Mutations in SAMD9L. *Am J Hum Genet* 98:1146-1158, 2016.

156. Narumi S, Amano N, Ishii T, Katsumata N, Muroya K, Adachi M, Toyoshima K, Tanaka Y, Fukuzawa R, Miyako K, Kinjo S, Ohga S, Ihara K, Inoue H, Kinjo T, Hara T, Kohno M, Yamada S, Urano H, Kitagawa Y, Tsugawa K, Higa A, Miyawaki M, Okutani T, Kizaki Z, Hamada H, Kihara M, Shiga K, Yamaguchi T, Kenmochi M, Kitajima H, Fukami M, Shimizu A, Kudoh J, Shibata S, Okano H, Miyake N, Matsumoto N, Hasegawa T. SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. *Nat Genet* 48:792-797, 2016.

157. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *N Engl J Med* 373:1136-1152, 2015.

158. Estey E, Dohner H. Acute myeloid leukaemia. *Lancet* 368:1894-1907, 2006.

159. Glait-Santar C, Desmond R, Feng X, Bat T, Chen J, Heuston E, Mizukawa B, Mulloy JC, Bodine DM, Larochelle A, Dunbar CE. Functional Niche Competition Between Normal Hematopoietic Stem and Progenitor Cells and Myeloid Leukemia Cells. *Stem Cells* 33:3635-3642, 2015.

160. Miraki-Moud F, Anjos-Afonso F, Hodby KA, Griessinger E, Rosignoli G, Lillington D, Jia L, Davies JK, Cavenagh J, Smith M, Oakervee H, Agrawal S, Gribben JG, Bonnet D, Taussig DC. Acute myeloid leukemia does not deplete normal hematopoietic stem cells but induces cytopenias by impeding their differentiation. *Proc Natl Acad Sci U S A* 110:13576-13581, 2013.

161. Estey EH. Acute myeloid leukemia: 2014 update on risk-stratification and management. *Am J Hematol* 89:1063-1081, 2014.

162. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, Sultan C. Proposed revised criteria for the classification of acute myeloid leukemia. A report of the French-American-British Cooperative Group. *Ann Intern Med* 103:620-625, 1985.

163. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, Sultan C. Proposal for the recognition of minimally differentiated acute myeloid leukaemia (AML-MO). *Br J Haematol* 78:325-329, 1991.

164. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, Sultan C. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. *Br J Haematol* 33:451-458, 1976.

165. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood* 100:2292-2302, 2002.

166. Vardiman JW, Thiele J, Arber DA, Brunning RD, Borowitz MJ, Porwit A, Harris NL, Le Beau MM, Hellstrom-Lindberg E, Tefferi A, Bloomfield CD. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood* 114:937-951, 2009.

167. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T, Dombret H, Ebert BL, Fenaux P, Larson RA, Levine RL, Lo-Coco F, Naoe T, Niederwieser D, Ossenkoppele GJ, Sanz M, Sierra J, Tallman MS, Tien HF, Wei AH, Lowenberg B, Bloomfield CD. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood* 129:424-447, 2017.

168. Dinardo CD, Stein EM, De Botton S, Roboz GJ, Altman JK, Mims AS, Swords R, Collins RH, Mannis GN, Pollyea DA, Donnellan W, Fathi AT, Pigneux A, Erba HP, Prince GT, Stein AS, Uy GL, Foran JM, Traer E, Stuart RK, Arellano ML, Slack JL, Sekeres MA, Willekens C, Choe S, Wang H, Zhang V, Yen KE, Kapsalis SM, Yang H, Dai D, Fan B, Goldwasser M, Liu H, Agresta S, Wu B, Attar EC, Tallman MS, Stone RM, Kantarjian HM. Durable Remissions with Ivosidenib in IDH1-Mutated Relapsed or Refractory AML. *N Engl J Med* 378:2386-2398, 2018.

169. Stone RM, Mandrekar SJ, Sanford BL, Laumann K, Geyer S, Bloomfield CD, Thiede C, Prior TW, Dohner K, Marcucci G, Lo-Coco F, Klisovic RB, Wei A, Sierra J, Sanz MA, Brandwein JM, De Witte T, Niederwieser D, Appelbaum FR, Medeiros BC, Tallman MS, Krauter J, Schlenk RF, Ganser A, Serve H, Ehninger G, Amadori S, Larson RA, Dohner H. Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3 Mutation. *N Engl J Med* 377:454-464, 2017.

170. Stein EM, Tallman MS. Emerging therapeutic drugs for AML. *Blood* 127:71-78, 2016.

171. Podoltsev NA, Stahl M, Zeidan AM, Gore SD. Selecting initial treatment of acute myeloid leukaemia in older adults. *Blood Rev* 31:43-62, 2017.

172. Mrozek K, Marcucci G, Paschka P, Whitman SP, Bloomfield CD. Clinical relevance of mutations and gene-expression changes in adult acute myeloid leukemia with normal cytogenetics: are we ready for a prognostically prioritized molecular classification? *Blood* 109:431-448, 2007.

173. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, Robertson A, Hoadley K, Triche TJ, Jr., Laird PW, Baty JD, Fulton LL, Fulton R, Heath SE, Kalicki-Veizer J, Kandoth C, Klco JM, Koboldt DC, Kanchi KL, Kulkarni S, Lamprecht TL, Larson DE, Lin L, Lu C, Mclellan MD, Mcmichael JF, Payton J, Schmidt H, Spencer DH, Tomasson MH, Wallis JW, Wartman LD, Watson MA, Welch J, Wendl MC, Ally A, Balasundaram M, Birol I, Butterfield Y, Chiu R, Chu A, Chuah E, Chun HJ, Corbett R, Dhalla N, Guin R, He A, Hirst C, Hirst M, Holt RA, Jones S, Karsan A, Lee D, Li HI, Marra MA, Mayo M, Moore RA, Mungall K, Parker J, Pleasance E, Plettner P, Schein J, Stoll D, Swanson L, Tam A, Thiessen N, Varhol R, Wye N, Zhao Y, Gabriel S, Getz G, Sougnez C, Zou L, Leiserson MD, Vandin F, Wu HT, Applebaum F, Baylin SB, Akbani R, Broom BM, Chen K, Motter TC, Nguyen K, Weinstein JN, Zhang N, Ferguson ML, Adams C, Black A, Bowen J, Gastier-Foster J, Grossman T, Lichtenberg T, Wise L, Davidsen T, Demchok JA, Shaw KR, Sheth M, Sofia HJ, Yang L, Downing JR, Eley G. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med* 368:2059-2074, 2013.

174. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N, Gundem G, Van Loo P, Martincorena I, Ganly P, Mudie L, Mclaren S, O'meara S, Raine K, Jones DR, Teague JW, Butler AP, Greaves MF, Ganser A, Dohner K, Schlenk RF, Dohner H, Campbell PJ. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med* 374:2209-2221, 2016.

175. Betz BL, Hess JL. Acute myeloid leukemia diagnosis in the 21st century. *Arch Pathol Lab Med* 134:1427-1433, 2010.

176. Ley TJ, Ding L, Walter MJ, Mclellan MD, Lamprecht T, Larson DE, Kandoth C, Payton JE, Baty J, Welch J, Harris CC, Lichti CF, Townsend RR, Fulton RS, Dooling DJ, Koboldt DC, Schmidt H, Zhang Q, Osborne JR, Lin L, O'laughlin M, Mcmichael JF, Delehaunty KD, Mcgrath SD, Fulton LA, Magrini VJ, Vickery TL, Hundal J, Cook LL, Conyers JJ, Swift GW, Reed JP, Alldredge PA, Wylie T, Walker J, Kalicki J, Watson MA, Heath S, Shannon WD, Varghese N, Nagarajan R, Westervelt P, Tomasson MH, Link DC, Graubert TA, Dipersio JF, Mardis ER, Wilson RK. DNMT3A mutations in acute myeloid leukemia. *N Engl J Med* 363:2424-2433, 2010.

177. Patel JP, Gonen M, Figueroa ME, Fernandez H, Sun Z, Racevskis J, Van Vlierberghe P, Dolgalev I, Thomas S, Aminova O, Huberman K, Cheng J, Viale A, Socci ND, Heguy A, Cherry A, Vance G, Higgins RR, Ketterling RP, Gallagher RE, Litzow M, Van Den Brink MR, Lazarus HM, Rowe JM, Luger S, Ferrando A, Paietta E, Tallman MS, Melnick A, Abdel-Wahab O, Levine RL. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med* 366:1079-1089, 2012.

178. Dinardo CD, Cortes JE. Mutations in AML: prognostic and therapeutic implications. *Hematology Am Soc Hematol Educ Program* 2016:348-355, 2016.

179. Shen Y, Zhu YM, Fan X, Shi JY, Wang QR, Yan XJ, Gu ZH, Wang YY, Chen B, Jiang CL, Yan H, Chen FF, Chen HM, Chen Z, Jin J, Chen SJ. Gene mutation patterns and their prognostic impact in a cohort of 1185 patients with acute myeloid leukemia. *Blood* 118:5593-5603, 2011.

180. Kansal R. Acute myeloid leukemia in the era of precision medicine: recent advances in diagnostic classification and risk stratification. *Cancer Biol Med* 13:41-54, 2016.

181. Wlodarski MW, Hirabayashi S, Pastor V, Stary J, Hasle H, Masetti R, Dworzak M, Schmugge M, Van Den Heuvel-Eibrink M, Ussowicz M, De Moerloose B, Catala A, Smith OP, Sedlacek P, Lankester AC, Zecca M, Bordon V, Matthes-Martin S, Abrahamsson J, Kuhl JS, Sykora KW, Albert MH, Przychodzien B, Maciejewski JP, Schwarz S, Gohring G, Schlegelberger B, Cseh A, Noellke P, Yoshimi A, Locatelli F, Baumann I, Strahm B, Niemeyer CM, Ewog MDS. Prevalence, clinical characteristics, and prognosis of GATA2-related myelodysplastic syndromes in children and adolescents. *Blood* 127:1387-1397; quiz 1518, 2016.

182. Smith ML, Cavenagh JD, Lister TA, Fitzgibbon J. Mutation of CEBPA in familial acute myeloid leukemia. *N Engl J Med* 351:2403-2407, 2004.

183. West RR, Hsu AP, Holland SM, Cuellar-Rodriguez J, Hickstein DD. Acquired ASXL1 mutations are common in patients with inherited GATA2 mutations and correlate with myeloid transformation. *Haematologica* 99:276-281, 2014.

184. Sellick GS, Spendlove HE, Catovsky D, Pritchard-Jones K, Houlston RS. Further evidence that germline CEBPA mutations cause dominant inheritance of acute myeloid leukaemia. *Leukemia* 19:1276-1278, 2005.

185. Kirwan M, Vulliamy T, Marrone A, Walne AJ, Beswick R, Hillmen P, Kelly R, Stewart A, Bowen D, Schonland SO, Whittle AM, Mcverry A, Gilleece M, Dokal I. Defining the pathogenic role of telomerase mutations in myelodysplastic syndrome and acute myeloid leukemia. *Hum Mutat* 30:1567-1573, 2009.

186. Narla A, Ebert BL. Ribosomopathies: human disorders of ribosome dysfunction. *Blood* 115:3196-3205, 2010.

187. Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL, Pigneux A, Wetzler M, Stuart RK, Erba HP, Damon LE, Powell BL, Lindeman N, Steensma DP, Wadleigh M, Deangelo DJ, Neuberg D, Stone RM, Ebert BL. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. *Blood* 125:1367-1376, 2015.

188. Shimamura A, Alter BP. Pathophysiology and management of inherited bone marrow failure syndromes. *Blood Rev* 24:101-122, 2010.

189. Feurstein S, Drazer MW, Godley LA. Genetic predisposition to leukemia and other hematologic malignancies. *Semin Oncol* 43:598-608, 2016.

190. Porter CC. Germ line mutations associated with leukemias. *Hematology Am Soc Hematol Educ Program* 2016:302-308, 2016.

191. Savage SA. Beginning at the ends: telomeres and human disease. *F1000Res* 7:2018.

192. Chirnomas SD, Kupfer GM. The inherited bone marrow failure syndromes. *Pediatr Clin North Am* 60:1291-1310, 2013.

193. Hsu AP, Sampaio EP, Khan J, Calvo KR, Lemieux JE, Patel SY, Frucht DM, Vinh DC, Auth RD, Freeman AF, Olivier KN, Uzel G, Zerbe CS, Spalding C, Pittaluga S, Raffeld M, Kuhns DB, Ding L, Paulson ML, Marciano BE, Gea-Banacloche JC, Orange JS, Cuellar-Rodriguez J, Hickstein DD, Holland SM. Mutations in GATA2 are associated with the autosomal dominant and sporadic monocytopenia and mycobacterial infection (MonoMAC) syndrome. *Blood* 118:2653-2655, 2011.

194. Spinner MA, Sanchez LA, Hsu AP, Shaw PA, Zerbe CS, Calvo KR, Arthur DC, Gu W, Gould CM, Brewer CC, Cowen EW, Freeman AF, Olivier KN, Uzel G, Zelazny AM, Daub JR, Spalding CD, Claypool RJ, Giri NK, Alter BP, Mace EM, Orange JS, Cuellar-Rodriguez J, Hickstein DD, Holland SM. GATA2 deficiency: a protean disorder of hematopoiesis, lymphatics, and immunity. *Blood* 123:809-821, 2014.

195. Mutsaers PG, Van De Loosdrecht AA, Tawana K, Bodor C, Fitzgibbon J, Menko FH. Highly variable clinical manifestations in a large family with a novel GATA2 mutation. *Leukemia* 27:2247-2248, 2013.

196. Bigley V, Collin M. Dendritic cell, monocyte, B and NK lymphoid deficiency defines the lost lineages of a new GATA-2 dependent myelodysplastic syndrome. *Haematologica* 96:1081-1083, 2011.

197. Dickinson RE, Griffin H, Bigley V, Reynard LN, Hussain R, Haniffa M, Lakey JH, Rahman T, Wang XN, Mcgovern N, Pagan S, Cookson S, Mcdonald D, Chua I, Wallis J, Cant A, Wright M, Keavney B, Chinnery PF, Loughlin J, Hambleton S, Santibanez-Koref M, Collin M. Exome sequencing identifies GATA-2 mutation as the cause of dendritic cell, monocyte, B and NK lymphoid deficiency. *Blood* 118:2656-2658, 2011.

198. Vinh DC, Patel SY, Uzel G, Anderson VL, Freeman AF, Olivier KN, Spalding C, Hughes S, Pittaluga S, Raffeld M, Sorbara LR, Elloumi HZ, Kuhns DB, Turner ML, Cowen EW, Fink D, Long-Priel D, Hsu AP, Ding L, Paulson ML, Whitney AR, Sampaio EP, Frucht DM, Deleo FR, Holland SM. Autosomal dominant and sporadic monocytopenia with susceptibility to mycobacteria, fungi, papillomaviruses, and myelodysplasia. *Blood* 115:1519-1529, 2010.

199. Alter BP, Giri N, Savage SA, Peters JA, Loud JT, Leathwood L, Carr AG, Greene MH, Rosenberg PS. Malignancies and survival patterns in the National Cancer Institute inherited bone marrow failure syndromes cohort study. *Br J Haematol* 150:179-188, 2010.

200. Alter BP, Giri N, Savage SA, Rosenberg PS. Cancer in dyskeratosis congenita. *Blood* 113:6549-6557, 2009.

201. Fernandez Garcia MS, Teruya-Feldstein J. The diagnosis and treatment of dyskeratosis congenita: a review. *J Blood Med* 5:157-167, 2014.

202. De La Fuente J, Dokal I. Dyskeratosis congenita: advances in the understanding of the telomerase defect and the role of stem cell transplantation. *Pediatr Transplant* 11:584-594, 2007.

203. Nelson ND, Bertuch AA. Dyskeratosis congenita as a disorder of telomere maintenance. *Mutat Res* 730:43-51, 2012.

204. Aviv A, Anderson JJ, Shay JW. Mutations, Cancer and the Telomere Length Paradox. *Trends Cancer* 3:253-258, 2017.

205. Tukiainen T, Villani AC, Yen A, Rivas MA, Marshall JL, Satija R, Aguirre M, Gauthier L, Fleharty M, Kirby A, Cummings BB, Castel SE, Karczewski KJ, Aguet F, Byrnes A, Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, Statistical Methods Groups-Analysis Working G, Enhancing GG, Fund NIHC, Nih/Nci, Nih/Nhgri, Nih/Nimh, Nih/Nida, Biospecimen Collection Source Site N, Biospecimen Collection Source Site R, Biospecimen Core Resource V, Brain Bank Repository-University of Miami Brain Endowment B, Leidos Biomedical-Project M, Study E, Genome Browser Data I, Visualization EBI, Genome Browser Data I, Visualization-Ucsc Genomics Institute UOCSC, Lappalainen T, Regev A, Ardlie KG, Hacohen N, Macarthur DG. Landscape of X chromosome inactivation across human tissues. *Nature* 550:244-248, 2017.

206. Gale RE, Fielding AK, Harrison CN, Linch DC. Acquired skewing of X-chromosome inactivation patterns in myeloid cells of the elderly suggests stochastic clonal loss with age. *Br J Haematol* 98:512-519, 1997.

207. International Human Genome Sequencing C. Finishing the euchromatic sequence of the human genome. *Nature* 431:931-945, 2004.

208. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci U S A* 74:5463-5467, 1977.

209. Grada A, Weinbrecht K. Next-generation sequencing: methodology and application. *J Invest Dermatol* 133:e11, 2013.

210. Behjati S, Tarpey PS. What is next generation sequencing? *Arch Dis Child Educ Pract Ed* 98:236-238, 2013.

211. Cullinane AR, Vilboux T, O'brien K, Curry JA, Maynard DM, Carlson-Donohoe H, Ciccone C, Program NCS, Markello TC, Gunay-Aygun M, Huizing M, Gahl WA. Homozygosity mapping and whole-exome sequencing to detect SLC45A2 and G6PC3 mutations in a single patient with oculocutaneous albinism and neutropenia. *J Invest Dermatol* 131:2017-2025, 2011.

212. Xuan J, Yu Y, Qing T, Guo L, Shi L. Next-generation sequencing in the clinic: promises and challenges. *Cancer Lett* 340:284-295, 2013.

213. Rehm HL. Disease-targeted sequencing: a cornerstone in the clinic. *Nat Rev Genet* 14:295-300, 2013.

214. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Exp Mol Med* 50:96, 2018.

215. Hoadley KA, Yau C, Hinoue T, Wolf DM, Lazar AJ, Drill E, Shen R, Taylor AM, Cherniack AD, Thorsson V, Akbani R, Bowlby R, Wong CK, Wiznerowicz M, Sanchez-Vega F, Robertson AG, Schneider BG, Lawrence MS, Noushmehr H, Malta TM, Cancer Genome Atlas N, Stuart JM, Benz CC, Laird PW. Cell-of-Origin Patterns Dominate the Molecular Classification of 10,000 Tumors from 33 Types of Cancer. *Cell* 173:291-304 e296, 2018.

216. Heisterkamp N, Stephenson JR, Groffen J, Hansen PF, De Klein A, Bartram CR, Grosveld G. Localization of the c-ab1 oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. *Nature* 306:239-242, 1983.

217. Greaves M. Leukaemia 'firsts' in cancer research and treatment. *Nat Rev Cancer* 16:163-172, 2016.

218. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531-537, 1999.

219. Mullighan CG, Goorha S, Radtke I, Miller CB, Coustan-Smith E, Dalton JD, Girtman K, Mathew S, Ma J, Pounds SB, Su X, Pui CH, Relling MV, Evans WE, Shurtleff SA, Downing JR. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446:758-764, 2007.

220. Ley TJ, Mardis ER, Ding L, Fulton B, Mclellan MD, Chen K, Dooling D, Dunford-Shore BH, Mcgrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, Dipersio JF, Wilson RK. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 456:66-72, 2008.

221. Anderson K, Lutz C, Van Delft FW, Bateman CM, Guo Y, Colman SM, Kempski H, Moorman AV, Titley I, Swansbury J, Kearney L, Enver T, Greaves M. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature* 469:356-361, 2011.

222. Klco JM, Spencer DH, Miller CA, Griffith M, Lamprecht TL, O'laughlin M, Fronick C, Magrini V, Demeter RT, Fulton RS, Eades WC, Link DC, Graubert TA, Walter MJ, Mardis ER, Dipersio JF, Wilson RK, Ley TJ. Functional heterogeneity of genetically defined subclones in acute myeloid leukemia. *Cancer Cell* 25:379-392, 2014.

223. Notta F, Mullighan CG, Wang JC, Poeppl A, Doulatov S, Phillips LA, Ma J, Minden MD, Downing JR, Dick JE. Evolution of human BCR-ABL1 lymphoblastic leukaemia-initiating cells. *Nature* 469:362-367, 2011.

224. Kim J. Unravelling the genomic landscape of leukemia using NGS techniques: the challenge remains. *Blood Res* 52:237-239, 2017.

225. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, Ritchey JK, Young MA, Lamprecht T, Mclellan MD, Mcmichael JF, Wallis JW, Lu C, Shen D, Harris CC, Dooling DJ, Fulton RS, Fulton LL, Chen K, Schmidt H, Kalicki-Veizer J, Magrini VJ, Cook L, Mcgrath SD, Vickery TL, Wendl MC, Heath S, Watson MA, Link DC, Tomasson MH, Shannon WD, Payton JE, Kulkarni S, Westervelt P, Walter MJ, Graubert TA, Mardis ER, Wilson RK, Dipersio JF. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature* 481:506-510, 2012.

226. Yan XJ, Xu J, Gu ZH, Pan CM, Lu G, Shen Y, Shi JY, Zhu YM, Tang L, Zhang XW, Liang WX, Mi JQ, Song HD, Li KQ, Chen Z, Chen SJ. Exome sequencing identifies somatic mutations of DNA methyltransferase gene DNMT3A in acute monocytic leukemia. *Nat Genet* 43:309-315, 2011.

227. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, Sato Y, Sato-Otsubo A, Kon A, Nagasaki M, Chalkidis G, Suzuki Y, Shiosaka M, Kawahata R, Yamaguchi T, Otsu M, Obara N, Sakata-Yanagimoto M, Ishiyama K, Mori H, Nolte F, Hofmann WK, Miyawaki S, Sugano S, Haferlach C, Koeffler HP, Shih LY, Haferlach T, Chiba S, Nakauchi H, Miyano S, Ogawa S. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature* 478:64-69, 2011.

228. Hindson BJ, Ness KD, Masquelier DA, Belgrader P, Heredia NJ, Makarewicz AJ, Bright IJ, Lucero MY, Hiddessen AL, Legler TC, Kitano TK, Hodel MR, Petersen JF, Wyatt PW, Steenblock ER, Shah PH, Bousse LJ, Troup CB, Mellen JC, Wittmann DK, Erndt NG, Cauley TH, Koehler RT, So AP, Dube S, Rose KA, Montesclaros L, Wang S, Stumbo DP, Hodges SP, Romine S, Milanovich FP, White HE, Regan JF, Karlin-Neumann GA, Hindson CM, Saxonov S, Colston BW. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem* 83:8604-8610, 2011.

229. Brambati C, Galbiati S, Xue E, Toffalori C, Crucitti L, Greco R, Sala E, Crippa A, Chiesa L, Soriani N, Mazzi B, Tresoldi C, Stanghellini MT, Peccatori J, Carrabba MG, Bernardi M, Ferrari M, Lampasona V, Ciceri F, Vago L. Droplet digital polymerase chain reaction for DNMT3A and IDH1/2 mutations to improve early detection of acute myeloid leukemia relapse after allogeneic hematopoietic stem cell transplantation. *Haematologica* 101:e157-161, 2016.

230. Brunetti C, Anelli L, Zagaria A, Minervini A, Minervini CF, Casieri P, Coccaro N, Cumbo C, Tota G, Impera L, Orsini P, Specchia G, Albano F. Droplet Digital PCR Is a Reliable Tool for Monitoring Minimal Residual Disease in Acute Promyelocytic Leukemia. *J Mol Diagn* 19:437-444, 2017.

231. Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, O'donnell-Luria AH, Ware JS, Hill AJ, Cummings BB, Tukiainen T, Birnbaum DP, Kosmicki JA, Duncan LE, Estrada K, Zhao F, Zou J, Pierce-Hoffman E, Berghout J, Cooper DN, Deflaux N, Depristo M, Do R, Flannick J, Fromer M, Gauthier L, Goldstein J, Gupta N, Howrigan D, Kiezun A, Kurki MI, Moonshine AL, Natarajan P, Orozco L, Peloso GM, Poplin R, Rivas MA, Ruano-Rubio V, Rose SA, Ruderfer DM, Shakir K, Stenson PD, Stevens C, Thomas BP, Tiao G, Tusie-Luna MT, Weisburd B, Won HH, Yu D, Altshuler DM, Ardissino D, Boehnke M, Danesh J, Donnelly S, Elosua R, Florez JC, Gabriel SB, Getz G, Glatt SJ, Hultman CM, Kathiresan S, Laakso M, Mccarroll S, Mccarthy MI, Mcgovern D, Mcpherson R, Neale BM, Palotie A, Purcell SM, Saleheen D, Scharf JM, Sklar P, Sullivan PF, Tuomilehto J, Tsuang MT, Watkins HC, Wilson JG, Daly MJ, Macarthur DG, Exome Aggregation C. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 536:285-291, 2016.

232. Cawthon RM. Telomere measurement by quantitative PCR. *Nucleic Acids Res* 30:e47, 2002.

233. Norberg A, Rosen A, Raaschou-Jensen K, Kjeldsen L, Moilanen JS, Paulsson-Karlsson Y, Baliakas P, Lohi O, Ahmed A, Kittang AO, Larsson P, Roos G, Degerman S, Hultdin M. Novel variants in Nordic patients referred for genetic testing of telomere-related disorders. *Eur J Hum Genet* 26:858-867, 2018.

234. Koskela HL, Eldfors S, Ellonen P, Van Adrichem AJ, Kuusanmaki H, Andersson EI, Lagstrom S, Clemente MJ, Olson T, Jalkanen SE, Majumder MM, Almusa H, Edgren H, Lepisto M, Mattila P, Guinta K, Koistinen P, Kuittinen T, Penttinen K, Parsons A, Knowles J, Saarela J, Wennerberg K, Kallioniemi O, Porkka K, Loughran TP,

Jr., Heckman CA, Maciejewski JP, Mustjoki S. Somatic STAT3 mutations in large granular lymphocytic leukemia. *N Engl J Med* 366:1905-1913, 2012.

235. Katainen R, Donner I, Cajuso T, Kaasinen E, Palin K, Makinen V, Aaltonen LA, Pitkanen E. Discovery of potential causative mutations in human coding and noncoding genome with the interactive software BasePlayer. *Nat Protoc* 13:2580-2600, 2018.

236. Churpek JE, Pyrtel K, Kanchi KL, Shao J, Koboldt D, Miller CA, Shen D, Fulton R, O'laughlin M, Fronick C, Pusic I, Uy GL, Braunstein EM, Levis M, Ross J, Elliott K, Heath S, Jiang A, Westervelt P, Dipersio JF, Link DC, Walter MJ, Welch J, Wilson R, Ley TJ, Godley LA, Graubert TA. Genomic analysis of germ line and somatic variants in familial myelodysplasia/acute myeloid leukemia. *Blood* 126:2484-2490, 2015.

237. Trotta L, Norberg A, Taskinen M, Beziat V, Degerman S, Wartiovaara-Kautto U, Valimaa H, Jahnukainen K, Casanova JL, Seppanen M, Saarela J, Koskenvuo M, Martelius T. Diagnostics of rare disorders: whole-exome sequencing deciphering locus heterogeneity in telomere biology disorders. *Orphanet J Rare Dis* 13:139, 2018.

238. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, Grody WW, Hegde M, Lyon E, Spector E, Voelkerding K, Rehm HL, Committee ALQA. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 17:405-424, 2015.

239. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. *Nat Methods* 7:248-249, 2010.

240. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc* 4:1073-1081, 2009.

241. Landrum MJ, Lee JM, Benson M, Brown GR, Chao C, Chitipiralla S, Gu B, Hart J, Hoffman D, Jang W, Karapetyan K, Katz K, Liu C, Maddipatla Z, Malheiro A, Mcdaniel K, Ovetsky M, Riley G, Zhou G, Holmes JB, Kattman BL, Maglott DR. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res* 46:D1062-D1067, 2018.

242. Ioannidis NM, Rothstein JH, Pejaver V, Middha S, Mcdonnell SK, Baheti S, Musolf A, Li Q, Holzinger E, Karyadi D, Cannon-Albright LA, Teerlink CC, Stanford JL, Isaacs WB, Xu J, Cooney KA, Lange EM, Schleutker J, Carpten JD, Powell IJ, Cussenot O, Cancel-Tassin G, Giles GG, Macinnis RJ, Maier C, Hsieh CL, Wiklund F, Catalona WJ, Foulkes WD, Mandal D, Eeles RA, Kote-Jarai Z, Bustamante CD, Schaid DJ, Hastie T, Ostrander EA, Bailey-Wilson JE, Radivojac P, Thibodeau SN, Whittemore AS, Sieh W. REVEL: An Ensemble Method for Predicting the Pathogenicity of Rare Missense Variants. *Am J Hum Genet* 99:877-885, 2016.

243. Stenson PD, Mort M, Ball EV, Evans K, Hayden M, Heywood S, Hussain M, Phillips AD, Cooper DN. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. *Hum Genet* 136:665-677, 2017.

244. Cardoso SR, Ryan G, Walne AJ, Ellison A, Lowe R, Tummala H, Rio-Machin A, Collopy L, Al Seraihi A, Wallis Y, Page P, Akiki S, Fitzgibbon J, Vulliamy T, Dokal I. Germline heterozygous DDX41 variants in a subset of familial myelodysplasia and acute myeloid leukemia. *Leukemia* 30:2083-2086, 2016.

245. Boocock GR, Morrison JA, Popovic M, Richards N, Ellis L, Durie PR, Rommens JM. Mutations in SBDS are associated with Shwachman-Diamond syndrome. *Nat Genet* 33:97-101, 2003.

246. Goldgar DE, Easton DF, Cannon-Albright LA, Skolnick MH. Systematic population-based assessment of cancer risk in first-degree relatives of cancer probands. *J Natl Cancer Inst* 86:1600-1608, 1994.

247. Hemminki K, Vaittinen P, Kyyronen P. Age-specific familial risks in common cancers of the offspring. *Int J Cancer* 78:172-175, 1998.

248. Nagy R, Sweet K, Eng C. Highly penetrant hereditary cancer syndromes. *Oncogene* 23:6445-6470, 2004.

249. Polprasert C, Schulze I, Sekeres MA, Makishima H, Przychodzen B, Hosono N, Singh J, Padgett RA, Gu X, Phillips JG, Clemente M, Parker Y, Lindner D, Dienes B, Jankowsky E, Sauntharajah Y, Du Y, Oakley K, Nguyen N, Mukherjee S, Pabst C, Godley LA, Churpek JE, Pollyea DA, Krug U, Berdel WE, Klein HU, Dugas M, Shiraishi

Y, Chiba K, Tanaka H, Miyano S, Yoshida K, Ogawa S, Muller-Tidow C, Maciejewski JP. Inherited and Somatic Defects in DDX41 in Myeloid Neoplasms. *Cancer Cell* 27:658-670, 2015.

250. Zhang J, Walsh MF, Wu G, Edmonson MN, Gruber TA, Easton J, Hedges D, Ma X, Zhou X, Yergeau DA, Wilkinson MR, Vadodaria B, Chen X, Mcgee RB, Hines-Dowell S, Nuccio R, Quinn E, Shurtleff SA, Rusch M, Patel A, Becksfort JB, Wang S, Weaver MS, Ding L, Mardis ER, Wilson RK, Gajjar A, Ellison DW, Pappo AS, Pui CH, Nichols KE, Downing JR. Germline Mutations in Predisposition Genes in Pediatric Cancer. *N Engl J Med* 373:2336-2346, 2015.

251. Churpek JE, Marquez R, Neistadt B, Claussen K, Lee MK, Churpek MM, Huo D, Weiner H, Bannerjee M, Godley LA, Le Beau MM, Pritchard CC, Walsh T, King MC, Olopade OI, Larson RA. Inherited mutations in cancer susceptibility genes are common among survivors of breast cancer who develop therapy-related leukemia. *Cancer* 122:304-311, 2016.

252. Alter BP. Fanconi anemia and the development of leukemia. *Best Pract Res Clin Haematol* 27:214-221, 2014.

253. Ceccaldi R, Sarangi P, D'andrea AD. The Fanconi anaemia pathway: new players and new functions. *Nat Rev Mol Cell Biol* 17:337-349, 2016.

254. Kitao H, Takata M. Fanconi anemia: a disorder defective in the DNA damage response. *Int J Hematol* 93:417-424, 2011.

255. Velez-Ruelas MA, Martinez-Jaramillo G, Arana-Trejo RM, Mayani H. Hematopoietic changes during progression from Fanconi anemia into acute myeloid leukemia: case report and brief review of the literature. *Hematology* 11:331-334, 2006.

256. Rider LG, Aggarwal R, Pistorio A, Bayat N, Erman B, Feldman BM, Huber AM, Cimaz R, Cuttica RJ, De Oliveira SK, Lindsley CB, Pilkington CA, Punaro M, Ravelli A, Reed AM, Rouster-Stevens K, Van Royen-Kerkhof A, Dressler F, Saad Magalhaes C, Constantin T, Davidson JE, Magnusson B, Russo R, Villa L, Rinaldi M, Rockette H, Lachenbruch PA, Miller FW, Vencovsky J, Ruperto N, International Myositis A, Clinical Studies G, The Paediatric Rheumatology International Trials O. 2016 American College of Rheumatology/European League Against Rheumatism Criteria for Minimal, Moderate, and Major Clinical Response in Juvenile Dermatomyositis: An International Myositis Assessment and Clinical Studies Group/Paediatric Rheumatology International Trials Organisation Collaborative Initiative. *Ann Rheum Dis* 76:782-791, 2017.

257. Churpek JE, Marquez R, Neistadt B, Claussen K, Lee MK, Churpek MM, Huo D, Weiner H, Bannerjee M, Godley LA, Le Beau MM, Pritchard CC, Walsh T, King MC, Olopade OI, Larson RA. Inherited mutations in cancer susceptibility genes are common among survivors of breast cancer who develop therapy-related leukemia. *Cancer* 122:304-311, 2016.

258. Brown AL, Churpek JE, Malcovati L, Dohner H, Godley LA. Recognition of familial myeloid neoplasia in adults. *Semin Hematol* 54:60-68, 2017.

259. Nagamachi A, Matsui H, Asou H, Ozaki Y, Aki D, Kanai A, Takubo K, Suda T, Nakamura T, Wolff L, Honda H, Inaba T. Haploinsufficiency of SAMD9L, an endosome fusion facilitator, causes myeloid malignancies in mice mimicking human diseases with monosomy 7. *Cancer Cell* 24:305-317, 2013.

260. Tesi B, Davidsson J, Voss M, Rahikkala E, Holmes TD, Chiang SCC, Komulainen-Ebrahim J, Gorcenco S, Rundberg Nilsson A, Ripperger T, Kokkonen H, Bryder D, Fioretos T, Henter JI, Mottonen M, Niinimaki R, Nilsson L, Pronk CJ, Puschmann A, Qian H, Uusimaa J, Moilanen J, Tedgard U, Cammenga J, Bryceson YT. Gain-of-function SAMD9L mutations cause a syndrome of cytopenia, immunodeficiency, MDS, and neurological symptoms. *Blood* 129:2266-2279, 2017.

261. Pastor VB, Sahoo SS, Boklan J, Schwabe GC, Saribeyoglu E, Strahm B, Lebrecht D, Voss M, Bryceson YT, Erlacher M, Ehninger G, Niewisch M, Schlegelberger B, Baumann I, Achermann JC, Shimamura A, Hochrein J, Tedgard U, Nilsson L, Hasle H, Boerries M, Busch H, Niemeyer CM, Wlodarski MW. Constitutional SAMD9L mutations cause familial myelodysplastic syndrome and transient monosomy 7. *Haematologica* 103:427-437, 2018.

262. Schwartz JR, Wang S, Ma J, Lamprecht T, Walsh M, Song G, Raimondi SC, Wu G, Walsh MF, Mcgee RB, Kesserwan C, Nichols KE, Cauff BE, Ribeiro RC, Wlodarski M, Klco JM. Germline SAMD9 mutation in siblings with monosomy 7 and myelodysplastic syndrome. *Leukemia* 31:1827-1830, 2017.

263. Hanahan D, Weinberg RA. Hallmarks of cancer: the next generation. *Cell* 144:646-674, 2011.

264. Romero-Laorden N, Castro E. Inherited mutations in DNA repair genes and cancer risk. *Curr Probl Cancer* 41:251-264, 2017.

265. Patil M, Pabla N, Dong Z. Checkpoint kinase 1 in DNA damage response and cell cycle regulation. *Cell Mol Life Sci* 70:4009-4021, 2013.

266. David L, Fernandez-Vidal A, Bertoli S, Grgurevic S, Lepage B, Deshaies D, Prade N, Cartel M, Larrue C, Sarry JE, Delabesse E, Cazaux C, Didier C, Recher C, Manenti S, Hoffmann JS. CHK1 as a therapeutic target to bypass chemoresistance in AML. *Sci Signal* 9:ra90, 2016.

267. Chamoun K, Borthakur G. Investigational CHK1 inhibitors in early stage clinical trials for acute myeloid leukemia. *Expert Opin Investig Drugs*; 10.1080/13543784.2018.15084481-6, 2018.

268. Bertoni F, Codegoni AM, Furlan D, Tibiletti MG, Capella C, Broggini M. CHK1 frameshift mutations in genetically unstable colorectal and endometrial cancers. *Genes Chromosomes Cancer* 26:176-180, 1999.

269. Ramsey JE, Fontes JD. The zinc finger transcription factor ZXDC activates CCL2 gene expression by opposing BCL6-mediated repression. *Mol Immunol* 56:768-780, 2013.

270. Wang L, Tsai CC. Atrophin proteins: an overview of a new class of nuclear receptor corepressors. *Nucl Recept Signal* 6:e009, 2008.

271. Fleischman AG, Aichberger KJ, Luty SB, Bumm TG, Petersen CL, Doratotaj S, Vasudevan KB, Latocha DH, Yang F, Press RD, Loriaux MM, Pahl HL, Silver RT, Agarwal A, O'hare T, Druker BJ, Bagby GC, Deininger MW. TNFalpha facilitates clonal expansion of JAK2V617F positive cells in myeloproliferative neoplasms. *Blood* 118:6392-6398, 2011.

272. Hasselbalch HC. Chronic inflammation as a promotor of mutagenesis in essential thrombocythemia, polycythemia vera and myelofibrosis. A human inflammation model for cancer development? *Leuk Res* 37:214-220, 2013.

273. Pourcelot E, Trocme C, Mondet J, Bailly S, Toussaint B, Mossuz P. Cytokine profiles in polycythemia vera and essential thrombocythemia patients: clinical implications. *Exp Hematol* 42:360-368, 2014.

274. Tefferi A, Vaidya R, Caramazza D, Finke C, Lasho T, Pardanani A. Circulating interleukin (IL)-8, IL-2R, IL-12, and IL-15 levels are independently prognostic in primary myelofibrosis: a comprehensive cytokine profiling study. *J Clin Oncol* 29:1356-1363, 2011.

275. Alter BP, Giri N, Savage SA, Rosenberg PS. Cancer in the National Cancer Institute inherited bone marrow failure syndrome cohort after fifteen years of follow-up. *Haematologica* 103:30-39, 2018.

276. Mitchell JR, Wood E, Collins K. A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 402:551-555, 1999.

277. Mitchell JR, Cheng J, Collins K. A box H/ACA small nucleolar RNA-like domain at the human telomerase RNA 3' end. *Mol Cell Biol* 19:567-576, 1999.

278. Alder JK, Parry EM, Yegnasubramanian S, Wagner CL, Lieblich LM, Auerbach R, Auerbach AD, Wheelan SJ, Armanios M. Telomere phenotypes in females with heterozygous mutations in the dyskeratosis congenita 1 (DKC1) gene. *Hum Mutat* 34:1481-1485, 2013.

279. Devriendt K, Matthijs G, Legius E, Schollen E, Blockmans D, Van Geet C, Degreef H, Cassiman JJ, Fryns JP. Skewed X-chromosome inactivation in female carriers of dyskeratosis congenita. *Am J Hum Genet* 60:581-587, 1997.

280. Vulliamy TJ, Knight SW, Dokal I, Mason PJ. Skewed X-inactivation in carriers of X-linked dyskeratosis congenita. *Blood* 90:2213-2216, 1997.

281. Xu J, Khincha PP, Giri N, Alter BP, Savage SA, Wong JM. Investigation of chromosome X inactivation and clinical phenotypes in female carriers of DKC1 mutations. *Am J Hematol* 91:1215-1220, 2016.

282. Vulliamy TJ, Kirwan MJ, Beswick R, Hossain U, Baqai C, Ratcliffe A, Marsh J, Walne A, Dokal I. Differences in disease severity but similar telomere lengths in genetic subgroups of patients with telomerase and shelterin mutations. *PLoS One* 6:e24383, 2011.

283. Alder JK, Hanumanthu VS, Strong MA, Dezern AE, Stanley SE, Takemoto CM, Danilova L, Applegate CD, Bolton SG, Mohr DW, Brodsky RA, Casella JF, Greider CW, Jackson JB, Armanios M. Diagnostic utility of telomere length testing in a hospital-based setting. *Proc Natl Acad Sci U S A* 115:E2358-E2365, 2018.

284. Zomnir MG, Lipkin L, Pacula M, Meneses ED, Macleay A, Duraisamy S, Nadhamuni N, Al Turki SH, Zheng Z, Rivera M, Nardi V, Dias-Santagata D, Iafrate AJ, Le LP, Lennerz JK. Artificial Intelligence Approach for Variant Reporting. *JCO Clin Cancer Inform* 2018:2018.

285. Taeubner J, Wieczorek D, Yasin L, Brozou T, Borkhardt A, Kuhlen M. Penetrance and Expressivity in Inherited Cancer Predisposing Syndromes. *Trends Cancer* 4:718-728, 2018.

Appendix

Supplementary Table 1. (III: Myeloid gene panel)

Supplementary Table 1. Myeloid gene panel (III).

Gene (exons and exon-intron boundaries covered)	
ASXL1	ASXL transcriptional regulator 1
BCOR	BCL6 corepressor
CDKN2A	cyclin-dependent kinase inhibitor 2A
CEBPA	CCAAT enhancer binding protein alpha
CREBBP	CREB binding protein
CUX1	Cut like homeobox 1
DNMT3A	DNA (cytosine-5)-methyltransferase 3A
EP300	E1A binding protein P300
ETV6	ETS variant 6
EZH2	Enhancer of zeste 2
GATA2	GATA-binding factor 2
KDM6A	Lysine demethylase 6A
NF1	Neurofibromin 1
PHF6	PHD finger protein 6
RAD21	RAD21 cohesin complex component
SETD2	SET domain containing 2
STAG2	Stromal antigen 2
TET2	Tet methylcytosine dioxygenase 2
TP53	Tumor protein 53
ZRSR2	Zinc finger CCCH-type, RNA binding motif and
Gene (mutational hotspots covered)	
BRAF	B-Raf proto-oncogene, serine/threonine kinase
CALR	Calreticulin
CBL	Cbl proto-oncogene
CSF3R	Colony stimulating factor 3 receptor
FLT3	Fms related tyrosine kinase 3
GATA1	GATA binding protein 1
IDH1	Isocitrate dehydrogenase (NADP(+)) 1, cytosolic
IDH2	Isocitrate dehydrogenase (NADP(+)) 2,
JAK2	Janus kinase 2
KIT	KIT proto-oncogene receptor tyrosine kinase
KRAS	KRAS proto-oncogene, GTPase
MPL	MPL proto-oncogene, thrombopoietin receptor
NPM1	Nucleophosmin 1
PDGFRA	Platelet derived growth factor receptor alpha
PTPN11	Protein tyrosine phosphatase, non-receptor type 11
RUNX1	Runt related transcription factor 1
SETBP1	SET binding protein 1
SF3B1	Splicing factor 3b subunit 1
SMC1A	Structural maintenance of chromosomes 1A
SMC3	Structural maintenance of chromosomes 3
SRSF2	Serine and arginine rich splicing factor 2
U2AF1	U2 small nuclear RNA auxiliary factor 1
WT1	Wilms tumor 1