UPPSALA UNIVERSITET

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1709

Multi-omics analysis of relapsed acute myeloid leukemia

SVEA STRATMANN

ACTA UNIVERSITATIS UPSALIENSIS UPPSALA 2021

ISSN 1651-6206
ISBN 978-91-513-1093-0
urn:nbn:se:uu:diva-425965

Dissertation presented at Uppsala University to be publicly examined in A1:107a, BMC, Husargatan 3, Uppsala, Friday, 12 February 2021 at 09:00 for the degree of Doctor of Philosophy (Faculty of Medicine). The examination will be conducted in English. Faculty examiner: Dr Caroline Heckman (Institute for Molecular Medicine Finland-FIMM, HiLIFE-Helsinki Institute of Life Science, University of Helsinki, Finland).

Online defence: https://uu-se.zoom.us/s/64187095364

Abstract
Stratmann, S. 2021. Multi-omics analysis of relapsed acute myeloid leukemia. *Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine* 1709. 48 pp. Uppsala: Acta Universitatis Upsaliensis. ISBN 978-91-513-1093-0.

The prognosis for patients suffering from acute myeloid leukemia (AML) remains unsatisfactory and survival is often measured in months. Although the majority of patients achieve complete remission after aggressive treatment, most of them relapse within a few years. Those patients that relapse frequently show accelerated disease progression and therapy resistance and represent the major clinical challenge in AML oncology. The advent of massive parallel sequencing launched the detailed understanding of the molecular basis of AML leukemogenesis, however, studies focused on relapse and primary resistant (R/PR) AML remain sparse.

This thesis explores the spectrum of molecular alterations present in R/PR AML, using a multi-omics analysis approach on sequential primary patient specimens from 48 adult and 25 pediatric R/PR AML patients. In **Paper I** we applied genome wide next generation sequencing to investigate genomic alterations in adult and pediatric R/PR AML. We identified recurrent alterations affecting *MGA*, *ARID1A* and *H3F3A*, specific for adult R/PR AML cases. In addition, we reported previously unappreciated internal tandem duplications in *UBTF*, solely found in pediatric cases. In **Paper II** we showed an association between a pro-inflammatory signature and AML relapse, utilizing transcriptome wide RNA sequencing. Further, through a novel machine learning based analysis we were able to depict gene interactive networks and predictive features in AML relapse. In **Paper III** we performed DNA methylation analysis to further understand transcriptional changes during disease progression that could not be explained by genomic alterations. We identified hypermethylation of promoter associated sites of *RNF180* and *DSC3* at relapse and in treatment resistant AML samples, respectively, as well as concordant downregulated transcription of these genes. In **Paper IV** we were able to confirm some of the above mentioned alterations at the proteomic level by exploiting high resolution mass spectrometry data. In addition, we showed higher levels of mitochondrial related proteins at AML relapse.

In summary, molecular associations identified in this thesis, together with AML-specific neoantigens discovered via a proteogenomic approach in **Paper IV**, predict novel therapeutic targets and/or help to further optimize current treatment schemes. We envision that knowledge gathered through our studies will shed further light on the molecular characteristics underlying disease progression, thus contributing to prolong AML patient survival.

Keywords: Acute myeloid leukemia (AML), relapse and resistance, next generation sequencing (WGS/WES), RNA-sequencing, DNA methylation, HiRIEF LC-MS, proteogenomics

Svea Stratmann, Department of Immunology, Genetics and Pathology, Rudbecklaboratoriet, Uppsala University, SE-751 85 Uppsala, Sweden.

ISSN 1651-6206
ISBN 978-91-513-1093-0
urn:nbn:se:uu:diva-425965 (http://urn.kb.se/resolve?urn=urn:nbn:se:uu:diva-425965)

To my family

Supervisors and examination committee

Main supervisor	**Linda Holmfeldt** Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University (Sweden).
Co-supervisors	**Richard Rosenquist Brandell** Clinical Genetics, Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm (Sweden). **Tobias Sjöblom** Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University (Sweden).
Examiner	**Helena Jernberg Wiklund** Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University (Sweden).
Opponent	**Caroline Heckman** Institute for Molecular Medicine Finland-FIMM, Helsinki Institute of Life Science, University of Helsinki (Finland).
Committee	**Kajsa Paulsson** Department of Laboratory Medicine, Division of Clinical Genetics, Lund University (Sweden). **Sören Lehmann** Department of Medical Sciences; Haematology, Uppsala University (Sweden). **Lars Feuk** Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University (Sweden). **Jean Hausser** C5 Department of Cell and Molecular Biology; SciLife-Lab, Karolinska Institutet Stockholm (Sweden). **Grzegorz Wicher** Department of Immunology, Genetics and Pathology, Science for Life Laboratory, Uppsala University (Sweden).

List of Papers

This thesis is based on the following papers, which are referred to in the text by their Roman numerals.

I. **Stratmann, S.**, Yones, S.A., Mayrhofer, M., Norgren, N., Skaftason, A., Sun, J., Smolinska, K., Komorowski J., Krogh Herlin M., Sundström C., Eriksson A., Höglund M., Palle J., Abrahamsson J., Jahnukainen K., Cheng Munthe-Kaas M., Zeller B., Pokrovskaja Tamm K., Cavelier L., Holmfeldt L. (2020) Genomic characterization of adult and pediatric relapsed acute myeloid leukemia reveals novel therapeutic targets. *Submitted*

II. **Stratmann, S.**, Yones, S.A., Garbulowski, M., Sun, J., Skaftason, A., Mayrhofer, M., Norgren, Krogh Herlin M., Sundström C., Eriksson A., Höglund M., Palle J., Abrahamsson J., Jahnukainen K., Cheng Munthe-Kaas M., Zeller B., Pokrovskaja Tamm K., Cavelier L., Komorowski, J., Holmfeldt L. (2020) Transcriptomic analysis reveals pro-inflammatory signatures associated with acute myeloid leukemia progression. *Manuscript*

III. Dadras, M.S., **Stratmann, S.**, Skaftason, A., Mayrhofer, M., Norgren, Krogh Herlin M., Sundström C., Eriksson A., Höglund M., Palle J., Abrahamsson J., Jahnukainen K., Cheng Munthe-Kaas M., Zeller B., Pokrovskaja Tamm K., Cavelier L., Holmfeldt L. (2020) The DNA methylome of adult and pediatric relapsed acute myeloid leukemia. *Manuscript*

IV. **Stratmann, S.**, Vesterlund, M., Umer, H.M., Skaftason, A., Krogh Herlin M., Sundström C., Eriksson A., Höglund M., Palle J., Abrahamsson J., Jahnukainen K., Cheng Munthe-Kaas M., Zeller B., Pokrovskaja Tamm K., Cavelier L., Lehtiö, J., Holmfeldt L. (2020) Proteogenomic analysis of relapsed acute myeloid leukemia in adults and children. *Manuscript*

Cover: Vector DNA spiral made of polygons and dots.

Adapted, modified and extended

Contents

Abbreviations

AML	Acute myeloid leukemia
APL	Acute promyelocytic leukemia
BM	Bone marrow
BM-control	CD34-expressing bone marrow control sample
CHIP	Clonal hematopoiesis of indeterminate potential
CNA	Copy number alteration
CN-LOH	Copy-neutral loss-of-heterozygosity
CR	Complete remission
DMC	Differentially methylated CpG site
DMR	Differentially methylated region
DNA	Deoxyribonucleic acid
EFS	Event free survival
ELN	European Leukemia Network
HiRIEF LC-MS	High resolution isoelectric focusing mass spectrometry
HSC	Hematopoietic stem cell
HSCT	Hematopoietic stem cell transplantation
InDel	Insertions and deletions
LIC	Leukemia-initiating cell
MDS	Myelodysplastic syndromes
NOPHO	Nordic Society of Paediatric Haematology & Oncology
R/PR AML	Relapse and primary resistant AML
RNA	Ribonucleic acid
RNA-seq	RNA sequencing
SNV	Single nucleotide variant
VAF	Variant allele frequency
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World health organization

Introduction

Hematopoiesis (from the Greek αἷμα "blood" and ποιεῖν "to make") describes the maturation and differentiation of hematopoietic stem cells (HSCs) into diverse mature blood components in higher organisms (**Figure 1**). During this hierarchical process, the cells decrease their self-renewal capacity while increasing their differentiation status and migrate from the bone marrow niche to the peripheral blood and other tissues. During a lifetime, background alterations, so called “passenger events”, can accumulate in each HSC without conferring a malignant effect. Nonetheless, alterations with a selective advantage, so called “driver events”, can disrupt the strongly regulated process of normal hematopoiesis, leading to different kinds of hematological diseases such as leukemia. Leukemia can be further subdivided according to the cellular origin and the disease dynamics. A complex ensemble of (epi-)genomic, transcriptomic, and proteomic alterations and how these alterations interact with the tumor microenvironment is thought to contribute to the initiation, progression and conceivably therapy resistance of hematological tumors.

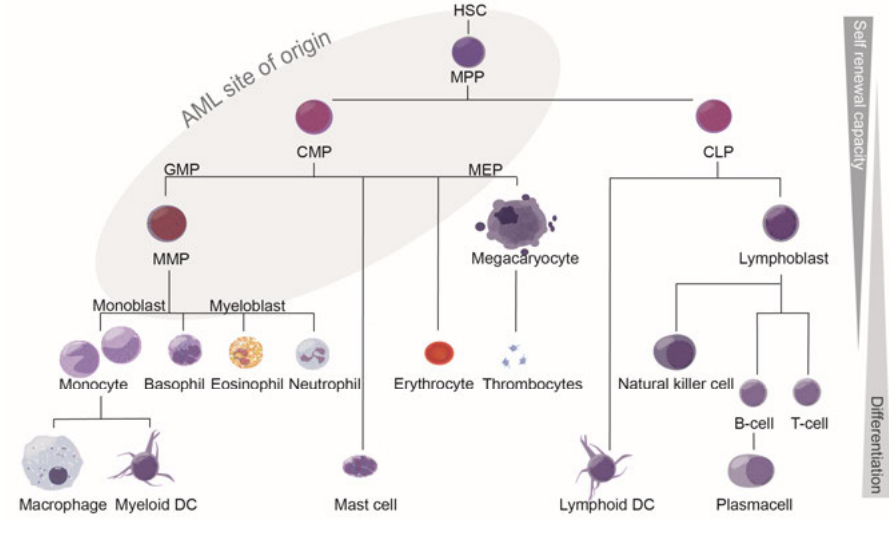

Figure 1. Schematic representation of the hematopoietic tree. AML, Acute myeloid leukemia; CLP, Common lymphoid progenitor; CMP, Common myeloid progenitor; DC, Dendritic cell; GMP, Granulocyte monocyte progenitor; HCS, Hematopoietic stem cell; MEP, Megakaryocyte erythrocyte progenitor; MMP, Myeloid-monocytic progenitor; MPP, Multi-potent progenitor.

Acute myeloid leukemia

Acute myeloid leukemia (AML) arises from the malignant transformation of myeloid stem or progenitor cells, overgrowing functional blood cells in the bone marrow (BM) before infiltrating the peripheral blood and possibly other tissues. AML is the most common form of acute leukemia with an incidence of 3.2 cases per 100,000 population and a higher rate of occurrence in men, based on data collected between 2007 and 2013 in the United States[1]. The incidence for developing AML increases with age and thus also within an aging population, with approximately one (children <19 years), two (adults <65 years) and 20 (adults >65 years) cases per 100,000 population[1]. About 360 cases are diagnosed annually in the northern countries[2]. Although the majority of patients achieve clinical complete remission (CR) after intensive chemotherapy and/or allogeneic hematopoietic stem cell transplantation (HSCT), 40-60% of the adults and 35% of the pediatric patients relapse within two to three years[3-6]. Those patients with recurrent tumors show poorer overall survival, accelerated disease progression and most likely will not respond to conventional treatment. Together, relapse and resistant disease remain the central challenges in the treatment of AML, resulting in poor five year overall survival rates for adult and pediatric patients, 29% and 65% respectively, and the highest percentage of death among leukemias[1,7].

Tumor evolution

> *In each case the fittest will survive, and a race will be eventually produced adapted to the conditions in which it lives* (Alfred R. 1823-1913).

AML tumor cells are thought to originate from hematopoietic stem or progenitor cells that sequentially accumulate alterations, eventually leading to increased self-renewal and proliferative capability (**Figure 2**). Thereafter, the tumor gains further mutations following a linear and/or branching evolutionary pattern resulting in high inter- and intra-tumor heterogeneity and, overall, increasing selective advantage for the tumor cells. Following the Darwinian rules of evolution and survival of the fittest, the tumor evolves. In this scenario, treatment can be seen as a bottleneck event in which sub-clonal populations might harbor mutations with a selective advantage. These cells can expand after intensive treatment and lead to relapse and treatment resistant disease.

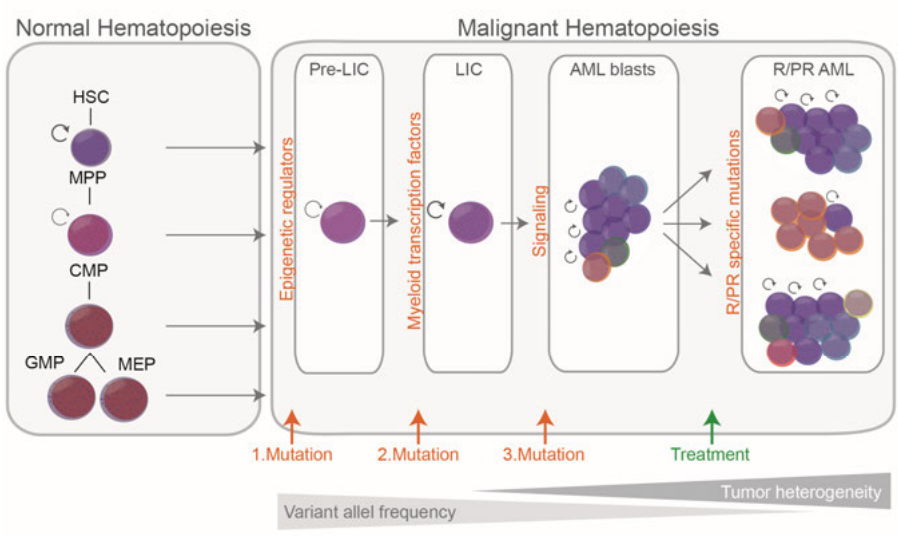

Figure 2. Clonal evolution from a genetic point of view. Depicted is a schematic view of the clonal evolution for adult *de novo* AML from the pre leukemia initiating phase up to relapsed disease. AML, Acute myeloid leukemia; HCS, Hematopoietic stem cell; MPP, Multi-potent progenitor; CMP, Common myeloid progenitor; GMP, Granulocyte monocyte progenitor; MEP, Megakaryocyte erythrocyte progenitor; LIC, Leukemia initiating cell; R/PR, Relapse and primary resistant AML.

Disease initiation

The hypothesis of a leukemia-initiating cell (LIC) or leukemic stem cell was raised almost 50 years ago, initiated by the identification of normal HSCs[8,9]. In the early 1990s, the observation that a small subpopulation of AML cells harbored colony forming potential and could be engrafted in xenotransplanted mouse models, further proved the cancer stem cell hypothesis[10,11]. Since then, it could be demonstrated that this ancestral sub-population of tumor cells frequently share characteristics with normal HSCs, such as increased self-renewal capacity, cell cycle quiescence and the ability to give rise to additional stem cells as well as more differentiated blast cells. Beyond these similarities, LICs are defined by increased drug resistance and the expression of CD34 surface antigens in combination with the lack of CD38 and lineage antigens (Lin-; e.g. CD19 and CD3; Ref.[12,13]). However, more recent studies have reported a greater variety among repopulating cells, for instance double positive cells (CD34+, CD38+) and even cells negative for CD34 (Ref.[14,15]).

Clonal evolution can be evaluated by following leukemic engraftment in immunodeficient mouse models or by investigating the allelic frequency of somatic alterations within a tumor population. Early events during tumor evolution are present at high variant allele frequency (VAF; ~50%), meaning that

the majority of cells within the tumor harbor that specific mutation. These mutations are referred to as clonal- or driving mutations. It has been shown that alterations in epigenetic regulators, such as *DNMT3A, TET2, IDH2, ASXL1* or *KMT2A* are mostly found to be clonal in adult patients with AML and therefore indicate early events[16,17]. These alterations might induce epigenetic instability and a multilineage repopulation advantage in the affected cells, and thus can result in a pool of pre-leukemic founder clones. This mutational status is referred to as clonal hematopoiesis of indeterminate potential (CHIP). Interestingly, these early aberrations can be found in 95% of the healthy population over the age of 60 years at very low but increasing VAFs[18]. Although patients affected by CHIP show an increased risk for developing hematological malignancies, CHIP status alone is not sufficient to create LICs as demonstrated by studies that reported CHIP 1) in healthy individuals that never developed AML, 2) during CR in patients that did not experience relapse, as well as in 3) mouse xenograft models that lack engraftment potential[7]. Pre-LICs with self-renewal advantage accumulate further aberrations, such as *NPM1*-mutations, lesions in hematopoietic transcription factors, or in splicing machinery components, resulting in distorted myeloid differentiation capacity and successively complete LICs[16,17,19,20].

Tumor progression and therapy resistance

Subsequently, these LICs frequently accumulate lesions in signaling pathways, such as receptor tyrosine kinase- or RAS-signaling pathways, leading to clonal expansion and to a non-engrafting blast population[16,17,21]. Adult patients that harbor common AML fusion genes or mutations in the tumor suppressor gene *TP53*, as well as most pediatric patients including patients with a germline predisposition for AML, commonly follow a slightly different type of clonal evolution that lacks CHIP-associated mutations during disease onset[19,22,23]. During further disease progression, some cells might seed sub-clonal populations that harbor additional mutations. These mutations might not be of importance during tumor formation but lead to an increased ability for the tumor to adapt and survive challenges in the tumor microenvironment, such as immune surveillance, energy shortage or intensive treatment.

Primary resistant disease (also called primary refractory AML) refers to AML patients that failed to achieve CR after intensive treatment. AML relapse describes disease recurrence after the patient had achieved clinical CR and can arise from the expansion of a sub-clone that was present at diagnosis or further evolution of a LIC. The same early lesions reported in pre-LICs commonly persist in CR and are thought to be primed to acquire additional mutations that help the tumor cells to expand after chemotherapy depletion[16,24-26] (**Figure 2**). Furthermore, the allele frequency of initiating driver events increases during disease progression, with a 10- to 100-fold higher frequency during relapse,

and this increase correlates with decreased event free survival (EFS) as well as overall survival[27]. Cytogenetic lesions and mutations affecting *TP53* have also frequently been found to be stable or gained during disease progression, with the later implicated in increased therapy resistance[28]. In contrast to stable early events, late events are frequently eradicated by treatment and are thought to be able to substitute for one another, as reported for alterations in signaling pathways[26,29,30]. Furthermore, each patient might gain relapse specific mutations, dominantly transversions, resulting in an even higher clonal diversity among relapsed AML and the increased likelihood of treatment resistance[24].

In cancer cells, this resistance to therapeutic drugs can be acquired through different strategies, including 1) altered drug intake into the tumor cells, 2) evasion into the bone marrow niche, 3) adaptation of the drug metabolism, and 4) bypass pathways and quiescence[31,32]. Mutations that are gained or remain stable between diagnostic- and relapse clones are likely to contribute to treatment resistance. However, the precise molecular mechanisms of how AML cells evade treatment remains to be elucidated.

Molecular heterogeneity

Massive parallel DNA-sequencing has uncovered the genomic landscape at AML diagnosis, leading to improved risk classification, minimal residual disease monitoring and treatment decision making. AML is a highly heterogeneous disease with genetic and non-genetic alterations known to contribute to disease characteristics through functional synergism. To date, more than 100[33] altered genes have been cataloged in AML leukemogenesis, with the most recurrent gene mutations and cytogenetic alterations summarized in **Table 1**. An average of 13 gene mutations, 1.5 fusion events and one copy-number alteration were found per adult *de novo* AML, which is a relatively low mutational rate compared to most other forms of cancer[22].

Beyond the next generation sequencing based investigations that have built the basis to formulate evolutionary hypotheses and to exploit genomic heterogeneity in AML, clonal evolution may also take place on the transcriptomic, epigenomic, proteomic and metabolomic level (exemplified by[34-40]). Research in these fields has gained popularity over the last years due to wider access to the required analysis methods. However, these results are often less validated and such approaches have not yet been implemented into clinical settings.

Table 1. Recurrently mutated genes and genomic alterations in adult *de novo* AML.

Functional group	Mutated gene/ Genomic aberration	Encoded protein	Mutational frequency* (%)
Nucleophosmin	*NPM1*	Nucleophosmin 1	25-30
DNA methylation	*DNMT3A*	DNA Methyltransferase 3α	25-30
	TET2	Tet Methylcytosine Dioxygenase 2	8-20
	IDH1/2	Isocitrate Dehydrogenase 1/2	5-10/15-20
(Myeloid)	*RUNX1*	Runt Related Transcription Factor 1	5-15
Transcription factors	*CEBPA*	CCAAT Enhancer Binding Protein α	6-10
	GATA1/2	GATA Binding Protein 1/2	5
	ETV6	ETS Variant 6	2
Signaling	*FLT3*	Fms Related Tyrosine Kinase 3	25-40
	NRAS	NRAS Proto-Oncogene, GTPase	12-22
	KIT	KIT Proto-Oncogene Receptor Tyrosine Kinase	4-10
	PTPN11	Protein Tyrosine Phosphatase, Non-Receptor Type 11	4-10
	KRAS	KRAS Proto-Oncogene, GTPase	6
	NF1	Neurofibromin 1	2
Tumor suppressors	*WT1*	Wilms Tumor 1	6-13
	TP53	Tumor Protein P53	6-10
	PHF6	PHD Finger Protein 6	2
Chromatin modifiers	*ASXL1/2*	Additional Sex Combs Like 1/2	5-20
	BCOR/L1	BCL6 Corepressor/Ligand 1	4-7
	KMT2A	Lysine Methyltransferase 2A	6
	EZH2	Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit	4
Spliceosome	*SRSF1/2/6*	Serine And Arginine Rich Splicing Factor 1/2/6	10

	U2AF1	U2 Small Nuclear RNA Auxiliary Factor 1	4
	SF3B1	Splicing Factor 3b Subunit 1	3
Cohesin complex	*STAG2*	Stromal Antigen 2	6
	RAD21	RAD21 Cohesin Complex Component	5
	SMC1A/3	Structural maintenance of chromosomes 1A/3	3
Gene fusions	t(8;21)	RUNX1-RUNX1T1	4-15
	t(11;#)	KMT2A-fusion partner	5-10
	inv(16)/t(16;16)	CBFB-MYH11	6
	t(5;11)	NUP98-NSD1	1-2
CNA/CN-LOH	-5/del(5q)	Including: *NPM1*	8
	-7/del(7q)	Including: *EZH2*	6-10
	-17/abn(17p)	Including: *NF1, SRSF1/2, TP53*	5
	LOH/+21q	Including: *RUNX1, U2AF1*	3

(*), Adapted from[20,22,41,42]; (#), different fusion partners; (-), Copy number loss; (+), Copy number gain; abn, Abnormal; CNA, Copy number alteration; CN-LOH, Copy-neutral loss-of-heterozygosity; del, Deletion; inv, Inversion; t, Translocation.

Epigenetic changes, such as differential DNA methylation, histone tail modification or RNA interference, control gene expression and play an important role during normal, as well as malignant, hematopoiesis. Loss-of-function mutations in genes regulating DNA methylation are found among the most frequent alterations in adult AML (**Table 1**). DNA-methyltransferases (e.g. DNMT3A) are "writers" that introduce a methyl group to the fifth carbon of a cytosine ring, while methylcytosine dioxygenases (e.g. TET2) are "erasers", which together with isocitrate dehydrogenases (e.g. IDH1/2) play a role in DNA demethylation. In addition, genes regulating histone modifications, such as *KMT2A* and *ASXL1/2*, have been frequently found altered in AML (**Table 1**). Their gene products are involved in chromatin remodeling and thus are responsible for the maintenance of both epigenetic activation and silencing. Finally, alterations in genes associated with the CTCF and cohesin complex, which regulate the three-dimensional conformation of chromatin, have been reported in AML[43]. As described above, several genes involved in epigenetic regulation are found to be altered in AML patients and often cause an unfavorable outcome[3,20,22]. Nevertheless, a number of patients do not harbor any of these genomic alterations, indicating independent epigenetic modifications[38]. Global methylation studies in AML have shown that CpG islands are enriched for hypermethylation while distant regions are generally hypomethylated when compared to normal CD34-expressing cells[44]. Differential DNA methylation analysis in AML revealed that promoters are mainly hypermethylated, leading to transcriptional silencing, while enhancers show both, AML specific hyper- and hypomethylation[45]. Among the most epigenetically altered regions in AML are regions annotated to genes encoding lineage determining transcription factors, including *WT1, SOX1, GATA1/2, RUNX1* and *HOX*-gene family members[43,44,46,47]. In summary, the AML epigenome is highly heterogeneous, which might reflect the loss of epigenetic control and thus confer an increased possibility for the tumor to evolve[48]. In line with that, it has been demonstrated that intensive treatment increases the plasticity of the epigenetic state of AML cells, which in turn is linked to adverse outcome[38,48-50]. As epigenetic marks are reversible and alterations are often found early during leukemogenesis, they present highly interesting targeted treatment options.

Transcriptomic profiling imparts the great advantage of comprehensively reflecting the function of genetic and epigenetic alterations at the gene expression level. In addition, the approach enables the detection of alternative transcripts, derived through alternative splicing and fusion transcripts, both of which have been described to contribute to tumor pathogenesis in different cancers. An enormous number of predictive risk scores have been proposed upon differential gene expression analyses in AML, mainly in the comparison to healthy controls or other leukemic subtypes. Although, serving as a promising tool for AML prognostication, only a few gene expression profiles are well acknowledged biomarkers. For instance, high expression of *HOX*-gene

family members has been associated with poor outcome and decreased therapy sensitivity in AML, in keeping with their role in stemness maintenance[51,52]. Further, elevated expression of *BAALC, ERG, MN1, PRAME, CD34* and *WT1* have been frequently identified in AML patients with adverse outcome[53,54]. In addition, certain non-coding RNAs (small- and long noncoding RNAs) and circular RNAs have been associated with tumor specific regulation in AML cells.

Beyond the potential of genetic (and transcriptomic) information to guide AML diagnosis and risk classification, as described in more detail below, their direct role in disease pathogenesis cannot always be confidently predicted. Proteins, however, represent the biomolecules that directly perform biological processes and are the active targets of most drugs. Previous proteomic studies mainly used antibody-based analysis methods. Although delivering high detection specificity, their main disadvantage lies in the limited multiplex capability. High throughput proteomic approaches, such as mass spectrometry based techniques, have revolutionized the proteomic research field and made it possible to add another omics level to unbiased and integrative cancer studies. Although global proteomic studies in AML have gained popularity since 2017, they remain limited. Proteomic profiling in AML could successfully be used to classify AML patients into risk groups, while instantly providing treatment guidance[55]. Further, a tumor specific OXPHOS status and changes in energy metabolism, RNA-processing, and cellular trafficking have been repeatedly associated with AML progression and patient survival[56-58]. In addition, post translational modification signatures were suggested as promising predictors for therapy response (as reviewed in[59]). Taken together, proteomic investigations may provide a more functional understanding of AML progression and enable improved prediction and validation of potential biomarkers and novel treatment options.

Notwithstanding characteristics of the tumor cells themselves, their interaction with the microenvironment plays a crucial role during all aspects of tumor evolution. The BM microenvironment represents a protective niche for hematopoietic tumor cells where they come in close proximity with remaining normal HSCs, non-hematopoietic cells and non-cellular components. S. Paget described in 1889 this phenomenon as the “seed and soil” hypothesis: The seed describes the tumor cell with its malignant variations, while the soil represents the microenvironment sustaining tumor growth and metastasis[60]. Recent studies strengthened this hypothesis and demonstrated a bi-directional cross-talk between tumor cells and their microenvironment (as reviewed in[61]).

This enormous clonal diversity and the continued lack of knowledge about underlying molecular mechanisms that can explain relapse and resistance, make AML treatment challenging and urgently require further exploration.

Clinical aspect of AML

Diagnosis

As the name AML indicates, typically the onset of symptoms is acute, occurring a few weeks to months prior to diagnosis. Symptoms are primarily the result of the expansion of immature malignant blast cells at the expense of mature functional blood cells, leading to a variety of systematic consequences such as neutropenia, anemia, fever, increased infection risk, weight loss and fatigue[3,62].

Morphological characterization of BM or peripheral blood cells is the first commonly applied step for AML diagnosis. A count of at least 20% myeloid blasts, including myeloblasts, monoblasts and megakaryoblasts, leads to the verification of suspected AML.

Flow cytometric characterization (expression of at least two of the following markers: CD13, CD33, CDw65, CD117, MPO) validates the myeloid origin of the disease and, together with metaphase karyotyping, adds information to further classify AML into clinically relevant subgroups. The French-American-British classification[3,63] has widely been used to subgroup AML tumors according to their blast morphology, but has had little to no independent prognostic or therapy leading relevance. One exception is the morphological diagnosis of M3 classified AML, followed by the validation of the gene fusion *PML-RARA* by real-time polymerase chain reaction and fluorescence in-situ hybridization, describing the entity of acute promyelocytic leukemia (APL), which subsequently leads to adjusted clinical guidelines and treatment protocols.

Historically AML has been classified by the combined information determined through the above mentioned clinical tests. However, cytogenetically normal AML accounts for the largest subset of AML and requires further molecular characterization[22,64]. Over the last decade, a shift towards clinical diagnosis led by causative genomic alterations has occurred, and a large number of genetic and transcriptomic biomarkers have been proposed since then. Nevertheless, only a few genetic markers (e.g. *NPM1, CEBPA, RUNX1, FLT3, TP53*), as well as defined gene fusions (e.g. *RUNX1-RUNX1T1, CBFB-MYH11, PML-RARA, MLLT3-KMT2A*) are commonly incorporated in clinical routine diagnostic and prognostic schemes.

Disease classification

Following diagnosis, AML is classified according to the World Health Organization (WHO) classification system combining clinical features with morphological, immunophenotype and genetic data[62,65]. According to the WHO system, AML is classified into seven main groups with clinical significance:

- AML with recurrent genetic alterations,
- AML with myelodysplasia-related changes,
- Therapy-related AML and myelodysplastic syndromes (MDS),
- Myeloid sarcoma,
- Myeloid proliferations related to Down syndrome,
- AML not otherwise specified,
- AML or MDS with germline predisposition

Current and future studies will further increase the molecular knowledge of underlying pathway alterations, aiming for a more detailed classification system with subsequently enhanced targeted therapeutic options.

Prognostic models

The prognosis in AML is dependent upon patient-based as well as tumor-related factors. Due to a higher frequency of unfavorable mutations, comorbidities, and poorer treatment tolerability, higher age is the most significant prognostic factor for adverse outcome. Therapy related AML as well as a history of MDS and high minimal residual disease are further factors that predict poor prognosis. The length of CR correlates with prognosis, whereas relapse at any stage indicates adverse outcome. The European Leukemia Network (ELN) genetic risk stratification model is widely recognized, grouping adult patients younger than 60 years into a favorable, intermediate or adverse risk group and subsequently guiding treatment decisions (**Table 2**; Ref.[3]). A recent revision could validate the applicability of the model also for patients above 60 years of age, although it was less robust for patients older than 75 years[66]. Mutations in *TP53* and a complex karyotype define the worst prognosis within the adverse risk group, while core-binding factor AML (*RUNX1-RUNX1T1* or *CBFB-MYH11*) accounts for the most favorable prediction. Further, *DNMT3A* and *WT1* mutations have frequently been suggested as additional markers for adverse outcome[66,67].

Table 2. Staging of AML according to the ELN-risk stratification[3].

Risk category	Genetic abnormality
Favorable	t(8;21)(q22;q22.1); *RUNX1-RUNX1T1*
	inv(16)(p13.1q22)/t(16;16)(p13.1;q22); *CBFB-MYH11*
	Mutated *NPM1* without *FLT3*-ITD or *FLT3*-ITD^{low}
	Biallelic mutated *CEBPA*
Intermediate	Mutated *NPM1 FLT3*-ITD^{high}
	Wild-type *NPM1* without *FLT3*-ITD or *FLT3*-ITD$^{low, \#}$
	t(9;11)(p21.3;q23.3); *MLLT3-KMT2A**
	Entities not classified as favorable or adverse
Adverse	t(6;9)(p23;q34.1); *DEK-NUP214*
	t(variable fusion partner;11q23.3); *KMT2A*-rearranged
	t(9;22)(p34.1;q11.2); *BCR-ABL1*
	inv(3)(q21.3q26.2), t(3;3)(q21.3;q26.2); *GATA2, MECOM*
	(-5)/del(5q) or (-7) or (-17)/abn(17p)
	Complex karyotype (≥ three unrelated abnormalities)
	Wild-type *NPM1 FLT3*-ITD^{high}
	Mutated *RUNX1*$^{¤}$, *ASXL1*$^{¤}$ or *TP53*

(#), Without adverse-risk genetic lesions, high/low = high/low variant allele frequency; (*), Preceding over rare, concurrent adverse-risk gene mutations; (¤), Marker should not be used when coexisting with favorable-risk AML subtypes; (-), Copy number loss; abn, Abnormal; del, Deletion; ELN, European Leukemia Network; inv, Inversion; ITD, Internal tandem duplication; t, Translocation.

Although the differences between the type and frequency of cytogenetic aberrations, between children and adults, are well described, most genetic classification schemes in AML are optimized for adult patients. Thus, the Nordic Society for Pediatric Hematology and Oncology (NOPHO) adjusted the ELN-classification system for pediatric patients resulting in significantly improved remission rates[7,68]. Approximately 40% of newly diagnosed patients do not, however, harbor any of the above mentioned causative genomic alterations[20]. These patients, as well as elderly patients and patients suffering from relapsed and/or resistant tumors, require further molecular characterization and refinement of the current prediction model.

Current and future treatment options

Due to the high tumor heterogeneity and to the importance of preventing relapse, the decision about the best therapeutic options is challenging, highlighting the need for improved tools in precision medicine. To date, the gold standard in the treatment of newly diagnosed AML patients is intensive chemotherapy, divided into remission induction- followed by consolidation therapy. The treatment regimens for fit adults, as specified below, and pediatric patients (described in detail in[4,7]) are largely comparable. However, three patient

groups are treated differently: 1) patients harboring high risk genetic lesions, who should receive HSCT, 2) elderly patients with comorbidity, advised to enrollment into a clinical trial or palliative care, and 3) patients diagnosed with APL, eligible for targeted treatment with all trans retinoic acid. The latter shows the great potential of precision treatment, leading to remission rates of over 90% (Ref.[3]).

Induction chemotherapy is mostly unaffected by individual patient factors, aiming at the reduction of leukemic cells and restoration of normal blood functions. For nearly 50 years, induction therapy has remained widely unchanged, consisting of the antimetabolite cytarabine in combination with an anthracycline (such as Daunorubicin) in a 7+3 combination (**Figure 3**). In the majority of cases two cycles within 21 days are applied to achieve CR, determined by a blast percentage of less than 5%. Although the majority of patients achieve CR after induction chemotherapy, at least half of the adult- and 35% of the pediatric patients eventually relapse[3-6].

In contrast to induction chemotherapy, the choice of consolidation therapy depends on an individual patient´s risk factor and genomic signatures, aimed at relapse prevention. Four to six weeks after CR, patients grouped into the favorable risk group receive consolidation chemotherapy with or without the combination of targeted drugs. Consolidation chemotherapy consists of two to four cycles of cytarabine with at least one cycle being of intermediate- or high-dose. Patients younger than 75 years of age, with adverse subtype AML should receive HSCT during their first CR. The decision of whether to apply HSCT is based on a risk-benefit ratio, taking into consideration the risk of treatment related death against the risk of disease recurrence. To date, the best salvage treatment option for relapse and primary resistant (R/PR) AML is HSCT and/or the enrollment into a clinical trial. In addition, Ivosidenib and Enasidenib are approved targeted inhibitors for *IDH1*- and *IDH2* mutated R/PR AML, respectively[69,70].

The great success of frontline trans retinoic acid administration for APL patients has encouraged the exploration of novel (targeted) therapies for AML, with many of these having entered different stages of clinical trials[33]. Due to the broad scope of conventional chemotherapy via the inhibition of DNA synthesis and cell division in tumor cells, these drugs also understandably affect normal cells, potentially leading to severe side effects. In contrast, novel targeted drugs are tailored to certain patient groups, and thus harbor a great potential for improving both the quality of life as well as quantity of life. To date, Midostaurin, a multi-kinase inhibitor, is the only approved targeted drug, used for the treatment of *de novo* (*FLT3* mutation positive) AML patients, in combination with standard chemotherapy[71] (**Figure 3**). In addition, Venetoclax, a BCL2-inhibitor as well as Glasdegib, a sonic hedgehog pathway inhibitor,

have recently been granted approval for patients ineligible for standard intensive treatment[72-74]. Small molecule inhibitors targeting for example KIT, TP53 or RAS are under investigation and show promising results as combinatorial treatment, although they are predetermined for patients harboring the respective mutation. Epigenetic therapies, including hypomethylating agents, histone deacetylase inhibitors and DOTL1- or LSD1 inhibitors show great potential for defined AML-subgroups (as reviewed in[75]). In addition, recent preclinical studies have shown increasing efficiency of immunotherapeutic concepts, using PD1/PDL1 inhibitors or genetically modified natural killer cells or T-cells (as reviewed in[75,76]). Finally, Gemtuzumab ozogamicin, an anti-CD33 monoclonal antibody linked to calicheamicin, was the first antibody-drug conjugate to be approved in AML for the treatment of CD33-positive AML patients[77].

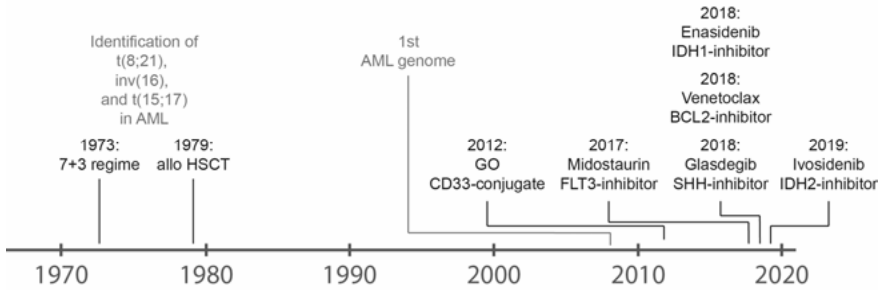

Figure 3. Timeline of clinical approaches in the treatment of AML. 7+3 regime, 7-day continuous infusion of cytarabine plus daunorubicin the first three days; Allo HSCT, Allogeneic hematopoietic stem cell transplantation; GO, Gemtuzumab ozogamicin; inv, Inversion; SHH, Sonic hedgehog signaling pathway; t, Translocation.

During CR, the careful monitoring of minimal residual disease is evolving to follow therapy sensitivity, increase the specificity of prognosis, and to detect relapse at an early stage[78,79]. Here, high resolution next generation sequencing approaches are desirable for detecting submicroscopic levels of disease, and need to be combined with optimized biomarkers, which are stable during disease progression and generally absent among the healthy population.

Together, relapse and treatment resistance represent the major clinical challenge in AML oncology. The molecular characteristics enabling malignant cells to escape treatment of (chemotherapeutic) drugs have yet to be elucidated, although being imperative to improve long term remission and subsequently patient survival.

Present investigations

Thesis aim

Relapse and treatment resistance in AML remain challenging and are the main reasons for overall poor survival. To this end, the work presented in this thesis is aimed at the exploration of the multi-omics landscape of relapsing and primary resistant (R/PR) AML, incorporating knowledge gathered from next generation sequencing, DNA methylation analysis, and proteogenomic data. In detail, the aims were as follows:

Paper I:
To investigate recurrent genomic alterations in adult and pediatric R/PR AML patients that may explain disease progression.

Paper II:
To characterize changes in the transcriptome of AML patients over the course of the disease, incorporating machine learning analysis as well as information regarding the patient-matched genomic background.

Paper III:
To determine the role of the DNA methylome and the correlation with expression patterns during disease progression and therapy resistance in AML.

Paper IV:
To delineate the proteome of R/PR AML, as affected by underlying genomic and transcriptomic alterations.

Patient and control specimens

We investigated the molecular characteristics of sequential primary patient specimens from 48 adult and 25 pediatric non-APL AML patients from the Nordic countries, all of which relapsed or had primary resistant disease. The study cohort comprised samples collected at diagnosis (n=52), relapse (n=80), as well as primary resistant specimens (n=6; **Table 3**).

All patients in this study were diagnosed according to the WHO[62,65,80] criteria and classified following the ELN-risk classification[3] for adult patients and the NOPHO-DBH AML 2012 Protocol (EudraCT Number 2012-002934-35) for pediatric AML. Inclusion criteria were the availability of R/PR AML material of sufficient quality and yield via Uppsala Biobank and Karolinska Institute Biobank, collected from 1995 through 2016. Sixty-six patients were diagnosed with *de novo* AML. The remaining seven patients had a prior diagnosis of MDS, suffered from therapy related AML, or therapy related MDS (n=2, n=3, n=2, respectively). Informed consent was obtained from each patient or their guardians according to the Declaration of Helsinki and study approval was acquired from the Uppsala Ethical Review Board (Sweden) and the Regional Ethical Committee South-East (Norway).

Alongside, we analyzed complete remission specimens (n=18) and/or normal BM-derived stromal cells (n=44 cases) as a source of germline DNA for analysis at the genomic level. Further, we included CD34-expressing BM control samples (referred to as BM-controls) from five distinct healthy donors as reference for the transcriptomic, epigenomic and proteomic studies.

Mononuclear cells were enriched through Ficoll gradient centrifugation and cryopreserved until use. Six additional samples were obtained as cell pellets. Cryopreserved AML specimens with leukemic cell content below 80% and sufficient amount of starting material were purified by immune-based depletion (CD3, CD19, CD14, CD235a) of non-tumor cells. Nucleic acids and protein were obtained from tumor samples and control specimens via the AllPrep DNA/RNA/Protein Kit (Qiagen) according to the manufacturer's protocol, incorporating DNase I treatment. Protein pellets were dissolved in lysis buffer (4% SDS, 20mM HEPES pH 7.6, 1mM DTT), as required for downstream analysis. Extracted DNA was quantified by fluorometric measurements (Quant-iT Broad-Range dsDNA Assay, Life Technologies), and qualified by NanoDrop 2000 (Thermo Scientific) and agarose gel electrophoresis. Total RNA was quantified by NanoDrop 2000 and qualified by automated electrophoresis using a TapeStation 4200 (Agilent Technologies). Protein concentration was quantified utilizing the BCA Protein Assay Kit (Thermo Fisher) with reducing agent compatibility.

Table 3. AML R/PR study cohort.

Patient \ Sample	D	R1	R1-P	R2	R2-P	R3	PR	CR/BMS
AML001		G/T/E/P						
AML002	G/T/E/P	G/T/E/P						G
AML003		G/T/E/P						G
AML004							G/T/E/P	G
AML005						G/T/E/P		G
AML006	G/E/P			G/T/E/P				G
AML008	G/T/E/P						G/T/E/P	
AML009	G/T/E/P	G/T/E/P						G
AML011	G/T/E/P	G/T/E/P						G
AML012		G/T/E/P						G
AML013	G/E	G/T/E/P						G
AML014	G/E	G/T/E/P						G
AML015	G/T/E/P					G/T/E/P		G
AML016							G/T/E/P	
AML017		G/T/E/P						
AML020	G/T/E/P						G/T/E/P	G
AML021		G/T/E/P						G
AML022				G/T/E/P				
AML023	G/T/E/P	G/T/E/P						G
AML024	G/T/E/P	G/T/E/P						G
AML025		G/T/E/P						
AML026		G/T/E/P						G
AML027				G/T/E/P				G
AML028	G/T/E/P						G/T/E/P	G
AML029				G/T/E/P				G
AML033				G/T/E/P				G
AML034		G/T/E/P						
AML035	G/T/E/P	G/T/E/P						
AML036	G/T/E/P	G/T/E/P						G
AML037	G/T/E/P	G/T/E/P						G
AML038		G/T/E/P						G
AML039	G/E	G/E						G
AML040	G/T/E/P	G/T/E/P						G
AML041*	G/E	G/E						G
AML042	G/T/E/P	G/T/P		G/E				G
AML043	G/T/E/P	G/T/E/P						G
AML044	G/T/E/P					G/T/E/P		G
AML045	G/T/E/P	G/T/E/P						
AML047	G/T/E/P			G/T/E/P				G
AML048		G/T/E/P		G/T/E/P				G
AML049	G/E	G/T/E/P						G
AML050	G/T/E/P	G/T/E/P		G/T/E/P				G
AML051	G/T/E/P	G/T/E/P		G/T/E				G
AML052	G/T/E/P	G/T/E/P						G
AML054	G/T/E/P	G/T/E/P						G
AML056		G/T/E/P						G
AML057	G/T/E/P	G/T/E/P						G
AML069*	G/T/E/P	G/T/E/P		G/T/E/P		G/T/E/P		G
AML070*	G/T/E/P	G/T/E/P						G
AML071*	G/T/E/P	G/T/E/P						G
AML072*	G/T/E/P	G/T/E/P		G/T/E/P	G/T/E/P			G
AML073*	G/T/E/P		G/T/E/P					
AML074*	G/T/E/P	G/T/E/P	G/T/E/P					G
AML075*	G/T/E/P	G/T/E/P						G
AML076*	G/T/E/P	G/T/E/P		G/T/E/P	G/T/E/P			G
AML080*		G/T/E/P						G
AML081*	G/T/E/P	G/T/E/P						G
AML082*	G/T/E/P	G/T/E/P						G
AML083*	G/E/P	G/T/E/P						G
AML085*	G/T/E/P	G/T/E/P						G
AML087*		G/T/E/P	G/T/E/P					G
AML089*	G	G						G
AML090*		G/T/E/P						G
AML091*	G/T/E/P		G/T/E/P					
AML092*	G/T/E/P	G/T/E/P		G				G
AML093*	G/T/E/P						G/T/E/P	G
AML094	G/T/E/P	G/T/E/P						G
AML095	G/T/E/P	G/T/E/P						G
AML097*	G/T/E/P	G/T/E/P						G
AML098*	G/E	G/T/E/P						G
AML100*	G/T/E	G/T/E						G
AML101*	G/T/E/P	G/T/E/P						G
AML103*	G/T/E	G/T/E/P						G
n=73	n=52	n=56	n=4	n=14	n=2	n=4	n=6	n=62

Data availability

- G Genomic (WGS/WES)
- T Transcriptomic (RNA-seq)
- E Epigenomic (DNA-methylation array)
- P Proteomic (HiRIEF LC-MS)

Sequencing method

- G WGS
- G WES

Cohort

- * Pediatric

Sample type

Diagnosis
Relapse
Relapse **P**ersistant
Primary **R**esistant
Complete **R**emission
Bone **M**arrow derived **S**tromal cells

The genome of relapsed and resistant AML

Exploration of the genomic landscape of pre-treatment AML has gained great popularity over the last decade. However, longitudinal studies in AML are scarce and mostly limited to gene panels, whole exome sequencing (WES) or a very small cohort size[24,27,29,38,40,81,82]. Detailed whole genome characterization of R/PR AML is essential for improved prognostication, disease monitoring and tailored treatment. With this in mind, we performed a comprehensive next generation sequencing analysis of all 73 R/PR AML patients included in our study cohort (**Paper I** and **Figure 4**).

Whole genome sequencing (WGS; HiSeq X, Illumina) was carried out for a total of 111 AML specimens as well as for patient-matched normal DNA from 60 patients (**Table 3**). For 99 of the AML samples, we aimed for 90X genomic coverage (mean coverage 114X), while 12 AML specimens, with limited DNA material, were sequenced aiming for 30X genomic coverage (mean coverage 39X). Also, patient-matched normal DNA was sequenced with an aim of 30X coverage (mean coverage 38X). WES (Ion Proton, Thermo Fisher Scientific) was carried out for an additional 27 AML specimens from 20 different patients, as well as for one patient-matched normal DNA sample (**Table 3**), with all samples analyzed by WES been lacking patient-matched normal DNA or sufficient amount of DNA for WGS. The mean coverage reached 131X.

Subsequently, genomic variant calling and annotation were performed using the Sarek[83] pipeline. The calling of somatic single nucleotide variants (SNVs) and small insertions and deletions (InDels; ≤50 nucleotides) was performed by Strelka[84]. Analysis of copy number alterations (CNAs) and copy-neutral loss-of-heterozygosity (CN-LOH) was conducted using ASCAT[85] and these results were subsequently validated utilizing CNVkit[86-88]. Somatic structural variants including large InDels (≥50 nucleotides) were called by Manta[89], though this was only performed for samples analyzed by WGS.

Our findings from the first multi-whole genome sequencing study of longitudinal adult and pediatric AML specimens revealed great plasticity during leukemic progression. Fifty-four percent of SNVs and small InDels persisted after disease relapse, 34% were gained after treatment, with the remainder being lost during disease progression. Relapse specific variants were predominately represented by transversions, as previously descibed[24,82]. In addition, larger structural variants (>50bp) and chromosomal gains and losses were mostly stable or gained during disease progression. Many of these structural variants and R/PR specific mutations frequently cannot be detected confidently using WES or targeted gene panels. Thus, we highlight the importance of applying WGS to fully elucidate the genomic variations present in R/PR AML.

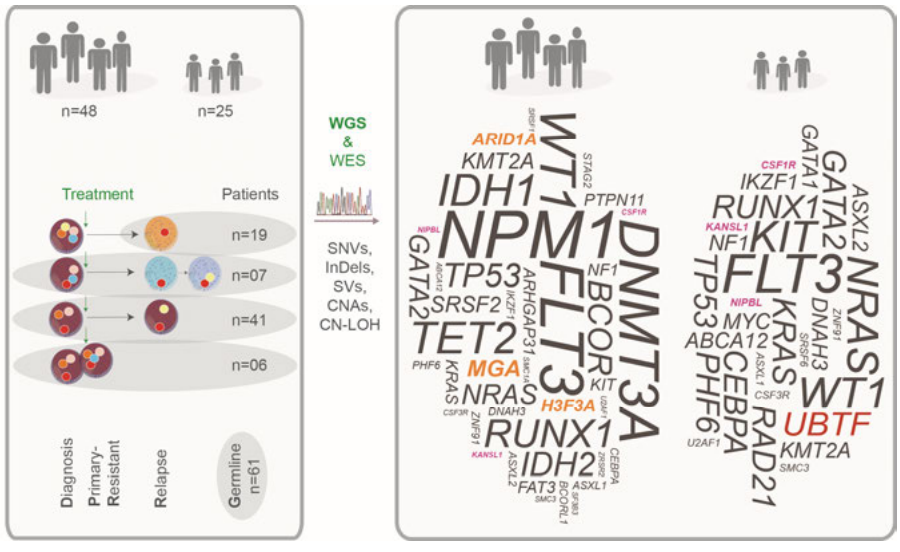

Figure 4. The genome of relapsed and primary resistant AML – Schematic view. CN-LOH, Copy-neutral loss-of-heterozygosity; CNA, Copy number alteration; InDel, Insertions and deletions; SNV, Single nucleotide variant; SV, Structural variant; WES, Whole exome sequencing; WGS, Whole genome sequencing.

Among recurrent relapse specific alterations we revealed mutations in *CSF1R* (2.7% of our relapse cases), which have previously not been reported in *de novo* AML. In addition, alterations in *ARID1A* (6.3%) and *MGA* (10.4%) were found to be recurrent in adult R/PR AML, both of which had ben identified at very low mutational frequencies in previous studies solely focused on pre-treatment specimens. Both CSF1R as well as ARID1A represent potential novel treatment options in R/PR AML. Receptor tyrosine kinase inhibitors could be used for *CSF1R*-mutated cases[90], while bromodomain and extraterminal domain inhibitors have been suggested as a therapeutic option for non-AML-*ARID1A*-mutated tumors[91]. Further, we reported novel specific differences in the mutational spectrum between pediatric versus adult R/PR AML, with recurrent internal tandem duplications in *UBTF*, encoding Upstream binding transcription factor, found solely in pediatric AML (n=3; 12.0%), while mutations affecting *H3F3A* (6.3%), *ARID1A* and *MGA* were specific for adult R/PR AML. Also, pediatric R/PR cases harbored a substantially higher frequency of mutations in cohesin-associated genes (20% of cases).

Despite the relatively low frequency of several of the reported mutations described in this study, their identification indicates important roles during disease progression and/or therapy resistance, and they are thus of great interest in the setting of personalized medicine.

The transcriptome of relapsed and resistant AML

Although genetic abnormalities that drive AML initiation have been investigated extensively, the risk stratification, especially for patients with no prognostic genetic aberrations and for relapse prediction, is still poor, indicating an interplay of molecular alterations at different omic-levels. RNA-sequencing (RNA-seq) provides a comprehensive picture of the cellular transcriptome, which enables a more functional view of underlying (epi-)genetic alterations. Former transcriptomic studies in AML have mainly been focused on gene expression signatures at diagnosis, however, global RNA-seq studies on longitudinal samples are largely missing or lack crucial genomic background information[92-94]. To this end, we applied RNA-seq, combining mutational detection, gene expression analysis, and machine learning based studies, on our R/PR AML cohort (**Paper II** and **Figure 5**).

RNA-seq was performed on 122 tumor specimens from 70 AML patients (n=47 adults, n=23 children; **Table 3**) and five BM-controls from healthy individuals (with two technical replicates each). Library preparation (Illumina TruSeq Stranded total RNA [ribosomal depletion] library kit) and RNA-seq (Illumina HiSeq2500 and/or Illumina NovaSeq6000) were carried out by the SNP&SEQ Technology platform, SciLifeLab, NGI Uppsala, Sweden. SNVs and small InDels (<50 bp) were called by HaplotypeCaller (GATK) using default settings for RNA-seq data. Fusion transcripts were retrieved via STAR-Fusion[95]. Gene counts were generated using FeatureCounts[96] and filtered towards expressed, protein-coding genes and normalized by the trimmed mean of M-values (TMM[97]; normalization Z-score [mean=0, var=1]) followed by log2-transformation. Subsequently, normalized counts were used for differential gene expression analyses carried out using Qlucore omics explorer 3.6. Resulting data were compared to two adult- and one pediatric validation cohorts.

To investigate the potential of using RNA-seq for variant and fusion detection, we compared our patient-matched genomic and transcriptomic results. We could verify all genomic aberrations leading to in-frame gene fusions that we had previously detected using WGS. The validation of SNVs and small InDels at the transcriptomic level was, however, less satisfying. Although 91% of variants located in regions with sufficient read coverage could be validated, we experienced the following challenges: 1) ~40% of all protein coding variants detected via WGS/WES were located in regions lacking read coverage by RNA-seq and 2) approximately 50% of the remaining variants were missed by automated variant calling using HaplotypeCaller and only identified by manual inspection of the RNA-seq reads. Thus, increased coverage and improved bioinformatic tools are required to allow for the full potential of transcriptomic sequencing, including variant detection.

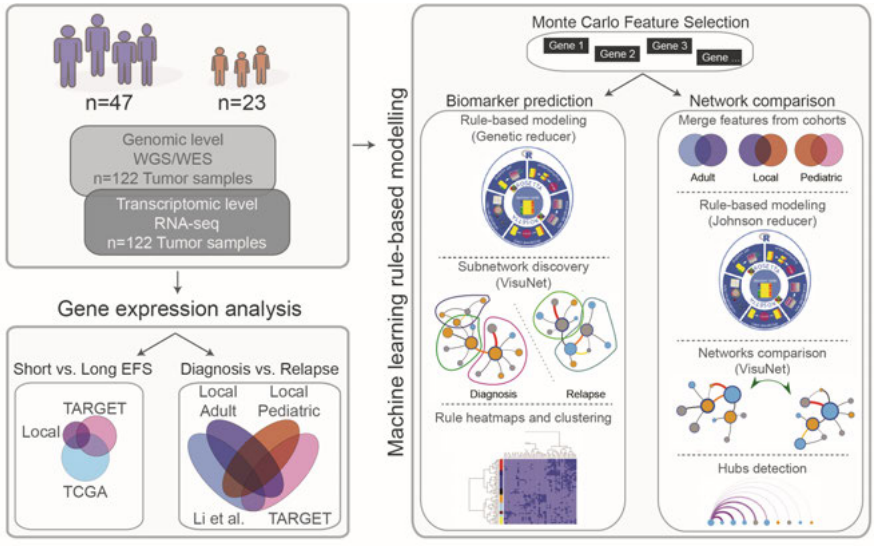

Figure 5. Principal workflow of transcriptomic based feature prediction. EFS, Event free survival; RNA-seq, RNA sequencing; WES, Whole exome sequencing; WGS, Whole genome sequencing.

Next, we performed differential gene expression analysis to determine relapse specific and prognostic factors in AML. We detected an association between short EFS and elevated expression levels of *GLI2* and *IL1R1*, as well as with downregulation of *ST18*. These findings could be validated in two independent larger AML cohorts (adult cohort [TCGA[22]], pediatric cohort [TARGET phs000465]), showing significant differences in regards to EFS, but also in overall survival, in correlation with differential gene expression of the respective genes. Further, relapse samples showed significantly altered expression levels of, for instance, *CR1*, *INSR* and *DPEP1*, compared to their pre-treated counterparts. Many of the alterations, identified herein, in the expression profile of R/PR AML are expected to promote a pro-inflammatory tumor environment and suggest their potential for targeted therapy.

In addition, we applied machine learning-based modeling to identify co-predictive biomarkers (**Figure 5**). First, relapse predictive features were selected using a Monte Carlo Feature Selection algorithm[98]. Next, these features were used to construct rule-based models and subsequently co-predictive features were estimated utilizing machine learning approaches (the rough set-based theory framework R.ROSETTA[99]). Through this analysis we identified a relapse associated network represented by *CD6*-overexpression and *INSR*-downregulation in adult AML samples. In addition, restored high expression of *NFATC4* and *KATNAL2* were associated with relapse as predicted through network comparison between our pediatric cohort and the TARGET cohort.

The DNA methylome of relapsed and resistant AML

In an attempt to further understand transcriptional changes during disease progression and therapy resistance that could not be explained by genomic alterations, we performed DNA methylation analysis as presented in **Paper III** (**Figure 6**). Most epigenetic based studies in AML have been focused on the exploration of epigenetic therapeutic options or the role of genetic mutations in genes known to act as epigenetic regulators (as reviewed in[100,101]). However, global epigenetic research on paired diagnosis and R/PR AML samples is limited to a few studies. Li et al. compared DNA methylation patterns ("eloci") between paired diagnosis-relapse samples from 138 adult patients, with matching genomic and transcriptomic data available for just eight of those patients[38]. Kroeger et al. studied differential DNA methylation in a cohort comprising 30 diagnosis-relapse pairs, but focused on merely nine loci[102]. Finally, Zampini et al. investigated global DNA methylation changes in longitudinal samples from six pediatric patients characterized by a t(8;21)*RUNX1-RUNX1T1*-rearrangement[103].

To shed further light on epigenetic changes that might be involved in relapse and resistance in adult and pediatric AML, we performed a DNA methylation analysis of more than 850 thousand sites, carried out on 132 leukemic specimens from 72 patients, including 50 diagnosis-R/PR pairs (**Table 3**), and five individual BM-controls. Matched next generation sequencing data were available for all samples, while transcriptomic data were available for all except nine. Bisulfite conversion (EZ DNA MethylationTM Kit from Zymo) and subsequent DNA methylation analysis (Infinium MethylationEPIC BeadChip, Illumina) were performed by the SNP&SEQ Technology platform, SciLifeLab, Uppsala, Sweden. Resulting raw data were normalized against two internal controls and background corrected using GenomeStudio. CpG sites were annotated to RefSeq genes and CpG islands according to the Infinium MethylationEPIC manifest file and genome build 37. Differential DNA methylation analysis was carried out for specific promoter-associated sites (differentially methylated CpG [DMC] sites; average difference ≥ 0.2) as well as on a regional level (differentially methylated region [DMR][104], minimum absolute $\Delta\beta$-value of 0.2). Finally, we evaluated the correlation between DMCs and DMRs with matching expression data of the annotated genes.

We identified hypermethylation of promoter associated CpG sites of the tumor suppressor gene *RNF180* at relapse. Interestingly, relapse samples that were characterized by treatment resistance overall showed higher methylation levels than their treatment sensitive counterparts. In line with these results, we found significantly lower expression of *RNF180* among relapse samples in comparison with the expression at diagnosis. Together, these results indicate towards a novel role of *RNF180* during AML tumor progression and therapy

resistance. In a comparison between pediatric patient-matched treatment sensitive and treatment resistant samples, we found an association between promoter hypermethylation of the cadherin superfamily member *DSC3* and treatment response. Elevated methylation levels of eight CpG sites in the promoter region of *DSC3*, and a correlation with significantly lower *DSC3* expression, were linked to therapy resistance.

In addition, we reported DMRs annotated to *OXGR1* and *EGFR* in pediatric relapse AML samples, without any apparent association to gene expression, suggesting a potential role of these methylation marks as novel biomarkers.

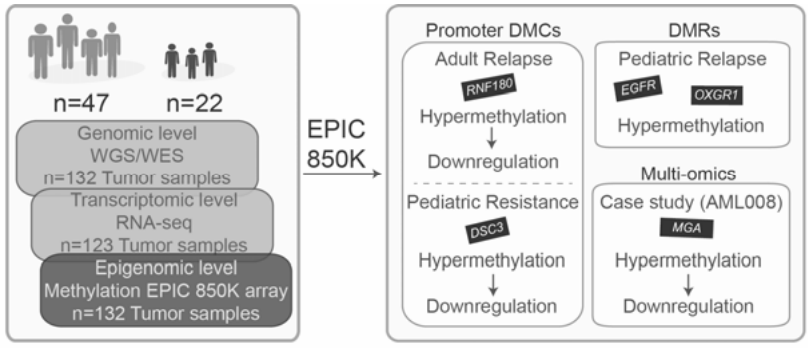

Figure 6. The search for DMCs and DMRs in R/PR AML- schematic view. DMC, Differentially methylated CpG site; DMR, Differentially methylated region; RNA-seq, RNA sequencing; WES, Whole exome sequencing; WGS, Whole genome sequencing.

Finally, we initiated our integrative analysis using a case study of a hypodiploid primary resistant AML case (AML008). Hypodiploidy in AML is rare, and due to the distinct transcriptomic and methylomic characteristics of this patient, samples were analyzed separately from the rest of the cohort. Through our transcriptomic study, we identified the silencing of *MGA*, which encodes a suppressor of MYC-associated transcription, in the diagnosis and primary resistant sample of patient AML008, which could not be explained by any genetic alterations. Here we showed promoter associated hypermethylation of *MGA*, which was shown to be specific for AML008 in comparison with our adult R/PR AML cohort. In addition, we previously identified recurrent inactivating mutations of *MGA* in R/PR AML (10.4% of adult cases) using WES and/or WGS. Taken together, our data suggest a previously unappreciated role of altered *MGA*-expression and function during tumor progression in AML.

Our results further delineate a role of aberrant DNA methylation during disease progression and therapy resistance in AML. Further unbiased integrative data mining approaches will possibly provide additional key insights into the pathophysiology of R/PR AML.

The proteome of relapsed and resistant AML

Finally, as described in **Paper IV**, we performed proteomic studies, yet again on our R/PR AML cohort (**Figure 7**). Mass spectrometry is a powerful tool that enables increasingly detailed information for a more functional characterization of tumor cells. Previous proteomic based studies in AML have mainly been focused on early diagnostic or prognostic markers or have been used to scrutinize novel treatments[105,106]. To the best of our knowledge, to-date, the only published study on global proteomic changes in relapsed AML is based on diagnosis-relapse pairs from just seven patients[107].

Here we analyzed the proteome of 78 leukemic specimens (**Table 3**) and five distinct BM-controls by high resolution isoelectric focusing mass spectrometry (HiRIEF LC-MS[108]). The method was chosen as it allows results with higher resolution compared with conventional mass spectrometry approaches, and therefore makes it possible to detect also smaller variations between protein profiles as well as low abundant proteins. Approximately 100µg of protein per sample were digested, labeled (Isobaric Mass Tag Labeling) and pooled into ten sets, each including a BM-control sample or a technical replicate, together with up to eight tumor samples and a pooled sample. Peptides were prefractionated following their isoelectronic point on a wide range HiRIEF gel strip (pH 3-10, 72 fractions per strip). Each fraction was further analyzed using LC-MS on an LTQ Orbitrap Velos (Thermo Scientific). Each peptide was subsequently mapped, by comparing the results to a database containing isoelectric point prediction of genome sequencing data. The resulting data were filtered towards proteins that had been quantified in every adult or, pediatric sample, respectively. The filtered data was subsequently used to perform quantitative protein abundance analysis utilizing Qlucore omics explorer 3.6. Finally, we compared the derived peptides to our transcriptomic data in a proteogenomic approach, to investigate novel peptides which are the result of translation of regions thought to be silent or non-coding (**Figure 7**).

We were able to confirm some of the main findings from our transcriptomic investigation (Paper II) at the proteomic level, including the association of a pro-inflammatory signature with AML progression. Downregulation of the anti-inflammatory complement receptor CR1 and distinct upregulation of various granzymes were identified among relapse samples. In addition, lower expression levels of INSR at relapse compared to diagnosis, could be verified at the proteomic level in adult patient samples. Further, we detected higher protein levels of a large number of mitochondrial ribosomal proteins (MRPL/S) and several proteins that are part of the mitochondrial respiratory chain complex (NDUFA/B/C) in AML relapse samples, resulting in a proteome enriched for pathways involved in mitochondrial translation, peptide biosynthesis and mitochondrial respiratory chain complex assembly. Finally, we identified an

association between higher levels of the pre-mRNA splicing factor SRSF9 and chemotherapy resistance in AML.

Next, we compared the protein profiles of samples associated with short versus long EFS. This analysis identified elevated protein levels of ANXA3 in adult and pediatric samples associated with short EFS. Kaplan-Mayer curves further revealed the correlation of high ANXA3 protein levels and short overall survival. Protein profiles enriched for ANXA3 have previously been reported in different cancers, other than AML, where it has been linked to increased proliferation and chemotherapy resistance[109,110].

Through the integration of our proteomic and transcriptomic data, we identified a total of 381 novel peptides and protein products annotated to regions previously described as untranslated. These non-canonical and novel peptides, so called neoantigens, present a large unexplored repertoire in the search for biomarkers and tumor specific therapeutic targets, thus building the foundation for further exploration.

In conclusion, we could show specific protein profiles associated with relapse and treatment resistance in AML. This knowledge, together with information derived through the integration of the proteomic data with additional patient matched omics-data, such as novel peptides, can be used to improve the quality of future and current treatments for R/PR AML patients.

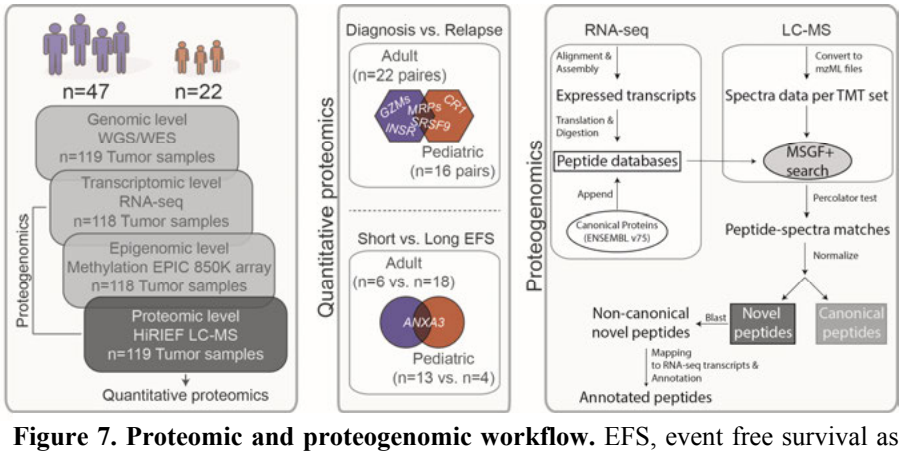

Figure 7. Proteomic and proteogenomic workflow. EFS, event free survival as time to first relapse; HiRIEF LC-MS, High resolution isoelectric focusing liquid chromatography mass spectrometry; MS-GF+ (v10072); RNA-seq, RNA sequencing; TMT, TMT10plex reagent (Thermo Scientific); WES, Whole exome sequencing; WGS, Whole genome sequencing.

Concluding remarks and future perspectives

The first historical footprint of leukemia dates back to 1811, when Peter Cullen published his investigation on serum blood with a "milky" appearance[111]. Driven by biological curiosity and the wish to cure this life threatening disease, our knowledge has increased tremendously over these last 200 years. Due to the ease of access to the blood compartment, leukemia studies have revolutionized molecular cancer research. Thanks to the great development in high-throughput molecular techniques, especially over the last decade, we are now able to detect, classify and treat most patients in a much improved way. However, the increasing number of active cancer patients in an aging society and the continued reality of adverse outcomes, especially for older AML patients, urgently speak to the need for further advancements.

Tumor initiation, progression and resistance can be depicted by the Darwinian rules of evolution and survival of the fittest and might, in part, also be comparable to the evolution of multi-resistant pathogens. Following this hypothesis, a cancer initiation cell gains tumor beneficial growth and survival mutations over time and upon environmental pressure. In this scenario, treatment leads to the accumulation of sub-clonal populations that might harbor some alterations with a selective advantage, but might also cause additional alterations as a consequence of its mutagenic effects. The treatment survival of some tumor cells is thought to be attributable to a synergism of cell autonomous molecular alterations and the tumor microenvironment[61]. These frequently resistant cells can expand after intensive treatment and lead to recurrent disease.

Next generation sequencing efforts have enhanced our molecular knowledge of AML, enabling the detection of "driver" mutations and have led to improved classification systems and treatment schemes[3,69,70,112]. These developments are accountable for complete remission being achieved in the majority of patients subjected to intensive chemotherapy or HSCT. However, 40-60% of adults and 35% of children experience disease relapse within three years (Ref.[3-6]). Until recently, the majority of molecular studies on AML have mainly been focused on the characterization of genomic and transcriptomic alterations present at disease onset. Nevertheless, as outlined above, a large proportion of AML patients with initial therapy response ultimately relapse and eventually succumb due to resistant tumors. To date, however, the precise

molecular alterations driving relapse and resistance in AML remain incompletely defined, although imperative to overcome this enormous clinical challenge.

With this in mind, this work was conducted to shed further light on the multi-level molecular characteristics that might explain relapse and resistance in AML. To the best of our knowledge, this is the largest multi-omics study of longitudinal AML patient specimens (adult n=48, pediatric n=25), incorporating data derived from next generation sequencing, DNA methylation and high resolution mass spectrometry analyses. We identified novel R/PR specific recurrent genomic alterations, as well as unexpectedly higher frequencies of those that have previously been reported in diagnosis-only studies (affecting e.g. *MGA*, *CSF1R*, *ARID1A*, *H3F3A* and *UBTF*; **Paper I**). At the level of transcriptomic alterations, we identified an association between a pro-inflammatory signature and AML relapse (**Paper II**). Further, by machine learning based analysis, we depicted interactive networks of relapse specific features and potential novel biomarkers. Via DNA methylation analysis, we were able to explain changes in expression patters that could not be explained by underlying genomic alterations. For instance, we detected hypermethylation in the promoter region of *MGA*, which in turn could explain the silencing of that gene in the corresponding samples, uncovered by RNA-seq. In addition, relapse associated hypermethylation and concordant downregulation were identified for the tumor suppressor gene *RNF180* and the cadherin superfamily member *DSC3* (**Paper III**). Finally, we were able to validate some of the above mentioned molecular changes at the proteomic level and additionally found a strong correlation between the overexpression of mitochondrial related proteins and AML relapse (**Paper IV**).

In conclusion, our results highlight the importance of complementary study approaches to fully reveal the biological differences between leukemic blasts at relapse compared to their pre-treatment counterparts. Although functional studies are to be awaited, each of these presented datasets contributes independently to the better understanding of the molecular landscape at relapse or primary resistance in AML. Molecular associations identified in this study, together with AML-specific neoantigens (**Paper IV**), represent the foundation for an exploration and development of novel therapeutic alternatives and for an effort to maximize the benefits of current treatments, altogether with the aim of improving outcome for R/PR AML patients. The next challenge lies in the assessment of the full integration of these multi-omics datasets.

Svensk sammanfattning

Akut myeloisk leukemi (AML) är en blodsjukdom karaktäriserad av överproduktion av omogna vita blodkroppar i benmärgen, vilket resulterar i nedsatt blodfunktion. Trots bästa möjliga behandling, som leder till att de flesta patienter uppnår fullständig remission, återinsjuknar en stor andel av patienterna med AML och de slutar svara på behandlingen. För att hitta nya behandlingsalternativ eller förbättra nuvarande behandlingsscheman behövs det mer kunskap kring de bakomliggande orsakerna till leukemicellernas uppkomst och speciellt deras fortlevnad och resistensutveckling.

AML karaktäriseras av specifika förändringar i leukemicellerna och deras tumörmikromiljö. Vissa förändringar i leukemicellernas arvsmassa associeras med sämre svar på behandlingen och större risk för återfall, men det är till stor del oklart hur de kan leda till detta.

Syftet med denna studie var att finna de förändringar som orsakar vidareutveckling av leukemin vid återfall, samt de förändringar som leukemicellerna ofta förvärvar för att motstå cancerbehandlingen. För detta ändamål genomförde vi så kallad ”multi-omics”-analys, vilket i vårt fall innefattade helgenomssekvenseringdata, transkriptomiska data, DNA-metyleringsdata och proteomikdata baserat på leukemiceller från 48 vuxna och 25 barn som drabbats av AML-återfall eller hade primär behandlingsresistens. Genom detta har vi hittat tidigare okända förändringar på alla ”omics”-nivåer, vilket skulle kunna tyda på att dessa förändringar är av betydelse för sjukdomens progression och resistensutveckling. Var och en av dessa presenterade datamängder, härledda från en separat ”omics-nivå”, bidrar oberoende till en bättre förståelse av det molekylära landskapet vid återfall eller primär resistens vid AML. I framtiden kommer utmaningen vara att integrera all information från de olika ”omics”-nivåerna och att vidare dechiffrera koden som ligger bakom leukemins utveckling och överlevnad.

Dessa kunskaper kommer sannolikt att öka vår förståelse för hur leukemicellerna styrs och därmed ge oss kunskap som kan användas för att förbättra behandlingen av denna svårbehandlade cancertyp.

Deutsche Zusammenfassung

Akute myeloische Leukämie (AML) ist eine bösartige Erkrankung des Blutsystems, die zu einer unkontrollierten Vermehrung von unreifen Blutzellen führt. AML betrifft größtenteils ältere Menschen, weshalb die Anzahl an Neuerkrankungen in einer alternden Gesellschaft stetig zunimmt. Innerhalb der letzten zwei Jahrzehnte, haben intensive Forschungsarbeit und Errungenschaften neuer, hochauflösender Analysetechniken dazu beigetragen, den Ursprung und das Fortschreiten dieser Krankheit besser zu verstehen. Dank dieser Fortschritte können die meisten Patienten heutzutage mit Hilfe von Chemotherapie oder einer Knochenmarksspende zunächst klinisch geheilt werden. Dennoch ist die Langzeitprognose für AML Patienten unzufriedenstellend und die Lebenserwartung beträgt zumeist nur weninge Monate. Der Grund, und damit die größte klinische Herrausforderung liegt darin, dass der Großteil der Patienten innerhalb weniger Jahre einen Rückfall erleidet. Rückfallpatienten sprechen, auf Grund von Resistenzentwicklungen, selten auf eine erneute Behandlung an und diese Patienten erwarten daher geringe Heilungschancen. Um neue Behandlungsalternativen zu erforschen und die bestehenden Behandlungsöglichkeiten verbessern zu können, ist detailliertes molekularbiologisches Wissen über die zugrundeliegenden Ursachen für die Entwicklung von Leukämiezellen erforderlich.

Das Ziel dieser Arbeit bestand darin, das Spektrum der molekularen Veränderungen in AML-Zellen detailliert zu untersuchen. Zu diesem Zweck wurde eine sogenannte "Multi-Omics" -Analyse durchgeführt, welche genomische und transkriptomische Daten (DNA- und RNA-Sequenzierung), sowie epigenomische (DNA-Methylierung) und proteomische (Massenspektrometrie) Daten von Leukämiezellen von 48 Erwachsenen und 25 Kindern mit AML-Rückfall umfasste. Durch diese Analysen konnten zuvor unbekannte Veränderungen auf allen vier "Omics" -Ebenen gefunden werden, die darauf hinweisen für den Krankheitsfortschritt und die Resistenzentwicklung von Bedeutung zu sein. Jeder dieser Datensätze, der von einer separaten "Omics-Ebene" stammt, trägt unabhängig zu einem verbesserten Verständnis des molekularen Krankheitsbildes von AML im Falle eines Rückfalls oder einer Resistenz bei. In Zukunft wird die Herausforderung darin bestehen, diese unterschiedlichen "Omics" -Ebenen zu integrieren um den Code hinter der Entwicklung und dem Überleben von Leukämiezellen weiter zu entschlüsseln.

Acknowledgements

This work was carried out at the Department of Immunology, Genetics and Pathology, Science for Life Laboratory at Uppsala University, Sweden. The following chapter is dedicated to all of those that made this result possible, lightened the path and filled up the unexpectedly long journey with joy.

First of all I would like to thank my main supervisor **Linda Holmfeldt**, for giving me the opportunity to be her first PhD student. I have learned a tremendous amount from you during the last years and I am enormously thankful that you were (literally) always there to help and give good advice, meanwhile giving the freedom and support to evolve and improve independently.

To my co-supervisors **Richard Rosenquist Brandell** and **Tobias Sjöblom** for guiding me during the time until my PhD defense. Thank you **Christer Sundström** for your support and supervision regarding the morphological examination of leukemic specimens. I also would like to thank **Panagiotis Baliakas**, **Anna Eriksson**, **Martin Höglund** and **Morten Krogh Herlin** for the support with patient related clinical data and clinical questions.

To the former and present members of the hematological groups, **Alba**, **Albin**, **Emma**, **Larry**, **Lesley**, **Linda**, **Lotta**, **My**, **Patrick**, **Sujata** and **Viktor** for the friendly and helpful working environment and fika sessions, especially when starting up our group at IGP. A special thanks to my group members **Henrik**, **Ren**, **Jitong**, **Sara**, **Mahsa** and **Mateusz**. It was always interesting to brainstorm with you and it felt good to occasionally share the office with somebody. Thank you **Aron**, **Markus**, **Nina**, **Husen** and **Mattias** for analyzing the tremendous amount of raw data. I can honestly say I would have been lost without your help. To **Diego**, for patiently answering all my additional "sequencing" questions. "Thank you, **Ross,** for proofreading, this report.!"

To the members of the PhD student council for all the good times inside and outside of Rudbeck. No matter what others have to say, we organized exquisite conferences! I am enormously thankful for all the help and guidance I have received from the PhD student administrators **Christina**, **Helene** and **Ulrica**.

Thank you (Gracias) **Andrea**, **Leonor**, **Vero** and (Dziękuję) **Magda** because of you I found a new home and family in Sweden. I miss you a lot. Uppsala is

not the same with you being so far away, but you will always be close to my heart! To **Johanna** and **Antonia** for exploring a new chapter of our lives together. Thank you **Anne** for sharing the passion to make Rudbeck a better (organized) place. Ordnung muss sein! And thanks to my friends in Germany, in particular **Hanna**, **Maren**, **Kathi**, **Judith**, **Mara**, **Anna**, **Sarah**, **Jana**, **Tina**, **Tine** und **Kristina** who always believed in me and loved me, even though I was the only nature science nerd among all the social scientists and teachers. Thank you for being so patient with me during the last few months. You make me miss Heinsberg. Danke.

Thank you **Adam** for being at my side no matter how the weather looked like and how many kilometers have kept us apart for quite some time. You definitely became a master chef over this last year. Kocham cię. Mein herzlichster Dank geht an meine Familie **Janis**, **Malte**, **Alisa** und besonders an **Mama** und **Ralle**. Vielen Dank für eure Liebe, Unterstützung, Ermutigung und Versorgungs-Pakete. Remember: Miles apart but close at heart! And finally, thank you **Maja**! You made these last two years of my time as a PhD student additionally sleepless but most of all incredible joyful.

Thank you.

This work was further supported by: Maria Lindström (Clinical Pathology, Uppsala University Hospital), Clinical Genomics and Clinical Genetics Uppsala (Science for Life Laboratory), and BioVis (Department of Immunology, Genetics and Pathology; Uppsala University) through technical support; U-CAN, Clinical Pathology, Uppsala University Hospital (Sweden), and Nordic Society of Paediatric Haematology and Oncology (NOPHO) by providing patient material; SNP&SEQ Technology Platform in Uppsala, part of the National Genomics Infrastructure (NGI) Sweden and Science for Life Laboratory (supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation) by performing sequencing analyses; Uppsala Genome Center, part of NGI Sweden (supported by the Swedish Council for Research Infrastructures and Uppsala University, and is hosted by Science for Life Laboratory) by performing whole exome sequencing; Clinical Proteomics Mass Spectrometry facility, Karolinska Institut/ Karolinska University Hospital/ Science for Life Laboratory by performing mass spectrometry analysis; SNIC (Uppsala Multidisciplinary Center for Advanced Computational Science (UPPMAX)) and SciLifeLab-WABI bioinformatics support at Uppsala University by providing computational support.

Financial support was granted by Knut and Alice Wallenberg's Foundation, The Swedish Research Council, The Swedish Childhood Cancer Foundation, The Swedish Cancer Society, and The Kjell and Märta Beijer Foundation.

References

1. Howlader N NA, Krapcho M, Miller D, Brest A, Yu M, Ruhl J, Tatalovich Z, Mariotto A, Lewis DR, Chen HS, Feuer EJ, Cronin KA. SEER Cancer Statistics Review (CSR) 1975-2016. SEER web site: National Cancer Institute; 2018.
2. NORDCAN. Association of the Nordic Cancer Registries. Vol. 2020: World Health Organization.
3. Dohner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424-447.
4. Karlsson L, Forestier E, Hasle H, et al. Outcome after intensive reinduction therapy and allogeneic stem cell transplant in paediatric relapsed acute myeloid leukaemia. *Br J Haematol*. 2017;178(4):592-602.
5. Verma D, Kantarjian H, Faderl S, et al. Late relapses in acute myeloid leukemia: analysis of characteristics and outcome. *Leuk Lymphoma*. 2010;51(5):778-782.
6. Bejanyan N, Weisdorf DJ, Logan BR, et al. Survival of patients with acute myeloid leukemia relapsing after allogeneic hematopoietic cell transplantation: a center for international blood and marrow transplant research study. *Biol Blood Marrow Transplant*. 2015;21(3):454-459.
7. Abrahamsson J, Forestier E, Heldrup J, et al. Response-guided induction therapy in pediatric acute myeloid leukemia with excellent remission rate. *Journal of clinical oncology : official journal of the American Society of Clinical Oncology*. 2011;29(3):310-315.
8. Killmann SA. Preleukemia: does it exist? *Nouv Rev Fr Hematol Blood Cells*. 1976;17(1-2):81-105.
9. Till JE, Mc CE. A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res*. 1961;14:213-222.
10. Lapidot T, Sirard C, Vormoor J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature*. 1994;367(6464):645-648.
11. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. *Nature medicine*. 1997;3(7):730-737.
12. Civin CI, Almeida-Porada G, Lee MJ, Olweus J, Terstappen LW, Zanjani ED. Sustained, retransplantable, multilineage engraftment of highly purified adult human bone marrow stem cells in vivo. *Blood*. 1996;88(11):4102-4109.
13. Terwijn M, Zeijlemaker W, Kelder A, et al. Leukemic stem cell frequency: a strong biomarker for clinical outcome in acute myeloid leukemia. *PLoS One*. 2014;9(9):e107587.
14. Taussig DC, Miraki-Moud F, Anjos-Afonso F, et al. Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. *Blood*. 2008;112(3):568-575.

15. Herrmann H, Sadovnik I, Eisenwort G, et al. Delineation of target expression profiles in CD34+/CD38- and CD34+/CD38+ stem and progenitor cells in AML and CML. *Blood Adv*. 2020;4(20):5118-5132.
16. Corces-Zimmerman MR, Hong WJ, Weissman IL, Medeiros BC, Majeti R. Preleukemic mutations in human acute myeloid leukemia affect epigenetic regulators and persist in remission. *Proc Natl Acad Sci U S A*. 2014;111(7):2548-2553.
17. Shlush LI, Zandi S, Mitchell A, et al. Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia. *Nature*. 2014;506(7488):328-333.
18. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. *Nat Commun*. 2016;7:12484.
19. Hirsch P, Zhang Y, Tang R, et al. Genetic hierarchy and temporal variegation in the clonal history of acute myeloid leukaemia. *Nat Commun*. 2016;7:12475.
20. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med*. 2016;374(23):2209-2221.
21. Sandén C, Lilljebjörn H, Orsmark Pietras C, et al. Clonal competition within complex evolutionary hierarchies shapes AML over time. *Nature Communications*. 2020;11(1):579.
22. Cancer Genome Atlas Research N, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med*. 2013;368(22):2059-2074.
23. Barwe SP, Sidhu I, Kolb EA, Gopalakrishnapillai A. Modeling Transient Abnormal Myelopoiesis Using Induced Pluripotent Stem Cells and CRISPR/Cas9 Technology. *Mol Ther Methods Clin Dev*. 2020;19:201-209.
24. Ding L, Ley TJ, Larson DE, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. *Nature*. 2012;481(7382):506-510.
25. Garg M, Nagata Y, Kanojia D, et al. Profiling of somatic mutations in acute myeloid leukemia with FLT3-ITD at diagnosis and relapse. *Blood*. 2015;126(22):2491-2501.
26. Buelow DR, Pounds SB, Wang YD, et al. Uncovering the Genomic Landscape in Newly Diagnosed and Relapsed Pediatric Cytogenetically Normal FLT3-ITD AML. *Clin Transl Sci*. 2019;12(6):641-647.
27. Ho TC, LaMere M, Stevens BM, et al. Evolution of acute myelogenous leukemia stem cell properties after treatment and progression. *Blood*. 2016;128(13):1671-1678.
28. McNeer NA, Philip J, Geiger H, et al. Genetic mechanisms of primary chemotherapy resistance in pediatric acute myeloid leukemia. *Leukemia*. 2019;33(8):1934-1943.
29. Morita K, Wang F, Jahn K, et al. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. *Nature Communications*. 2020;11(1):5327.
30. Farrar JE, Schuback HL, Ries RE, et al. Genomic Profiling of Pediatric Acute Myeloid Leukemia Reveals a Changing Mutational Landscape from Disease Diagnosis to Relapse. *Cancer Res*. 2016;76(8):2197-2205.
31. Ross DD. Novel mechanisms of drug resistance in leukemia. *Leukemia*. 2000;14(3):467-473.

32. Wang A, Zhong H. Roles of the bone marrow niche in hematopoiesis, leukemogenesis, and chemotherapy resistance in acute myeloid leukemia. *Hematology*. 2018;23(10):729-739.
33. My Cancer Genome.
34. George J, Uyar A, Young K, et al. Leukaemia cell of origin identified by chromatin landscape of bulk tumour cells. *Nature Communications*. 2016;7(1):12166.
35. Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nature Communications*. 2015;6(1):8489.
36. Bonardi F, Fusetti F, Deelen P, van Gosliga D, Vellenga E, Schuringa JJ. A proteomics and transcriptomics approach to identify leukemic stem cell (LSC) markers. *Mol Cell Proteomics*. 2013;12(3):626-637.
37. Song BH, Son SY, Kim HK, et al. Profiling of Metabolic Differences between Hematopoietic Stem Cells and Acute/Chronic Myeloid Leukemia. *Metabolites*. 2020;10(11).
38. Li S, Garrett-Bakelman FE, Chung SS, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. *Nat Med*. 2016;22(7):792-799.
39. Bachas C, Schuurhuis GJ, Zwaan CM, et al. Gene expression profiles associated with pediatric relapsed AML. *PLoS One*. 2015;10(4):e0121730.
40. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. 2018;562(7728):526-531.
41. Kishtagari A, Levine RL, Viny AD. Driver mutations in acute myeloid leukemia. *Current Opinion in Hematology*. 2020;27(2):49-57.
42. Metzeler KH, Herold T, Rothenberg-Thurley M, et al. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. *Blood*. 2016;128(5):686-698.
43. Mujahed H, Miliara S, Neddermeyer A, et al. AML displays increased CTCF occupancy associated with aberrant gene expression and transcription factor binding. *Blood*. 2020;136(3):339-352.
44. Qu Y, Lennartsson A, Gaidzik VI, et al. Differential methylation in CN-AML preferentially targets non-CGI regions and is dictated by DNMT3A mutational status and associated with predominant hypomethylation of HOX genes. *Epigenetics*. 2014;9(8):1108-1119.
45. Qu Y, Siggens L, Cordeddu L, et al. Cancer-specific changes in DNA methylation reveal aberrant silencing and activation of enhancers in leukemia. *Blood*. 2017;129(7):e13-e25.
46. Bullinger L, Ehrich M, Döhner K, et al. Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. *Blood*. 2010;115(3):636-642.
47. Spencer DH, Young MA, Lamprecht TL, et al. Epigenomic analysis of the HOX gene loci reveals mechanisms that may control canonical expression patterns in AML and normal hematopoietic cells. *Leukemia*. 2015;29(6):1279-1289.
48. Li S, Chen X, Wang J, et al. Somatic Mutations Drive Specific, but Reversible, Epigenetic Heterogeneity States in AML. *Cancer Discovery*. 2020.
49. Li S, Garrett-Bakelman F, Perl AE, et al. Dynamic evolution of clonal epialleles revealed by methclone. *Genome Biology*. 2014;15(9):472.
50. Luskin MR, Gimotty PA, Smith C, et al. A clinical measure of DNA methylation predicts outcome in de novo acute myeloid leukemia. *JCI Insight*. 2016;1(9).

51. Bullinger L, Dohner K, Bair E, et al. Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia. *New England Journal of Medicine*. 2004;350(16):1605-1616.
52. Andreeff M, Ruvolo V, Gadgil S, et al. HOX expression patterns identify a common signature for favorable AML. *Leukemia*. 2008;22(11):2041-2047.
53. Metzeler KH, Dufour A, Benthaus T, et al. ERG expression is an independent prognostic factor and allows refined risk stratification in cytogenetically normal acute myeloid leukemia: a comprehensive analysis of ERG, MN1, and BAALC transcript levels using oligonucleotide microarrays. *J Clin Oncol*. 2009;27(30):5031-5038.
54. Zhu Y-M, Wang P-P, Huang J-Y, et al. Gene mutational pattern and expression level in 560 acute myeloid leukemia patients and their clinical relevance. *Journal of Translational Medicine*. 2017;15(1):178.
55. Kornblau SM, Tibes R, Qiu YH, et al. Functional proteomic profiling of AML predicts response and survival. *Blood*. 2009;113(1):154-164.
56. Farge T, Saland E, de Toni F, et al. Chemotherapy-Resistant Human Acute Myeloid Leukemia Cells Are Not Enriched for Leukemic Stem Cells but Require Oxidative Metabolism. *Cancer Discov*. 2017;7(7):716-735.
57. Raffel S, Klimmeck D, Falcone M, et al. Quantitative proteomics reveals specific metabolic features of acute myeloid leukemia stem cells. *Blood*. 2020;136(13):1507-1519.
58. Aasebø E, Berven FS, Bartaula-Brevik S, et al. Proteome and Phosphoproteome Changes Associated with Prognosis in Acute Myeloid Leukemia. *Cancers (Basel)*. 2020;12(3).
59. Aasebø E, Forthun RB, Berven F, Selheim F, Hernandez-Valladares M. Global Cell Proteome Profiling, Phospho-signaling and Quantitative Proteomics for Identification of New Biomarkers in Acute Myeloid Leukemia Patients. *Curr Pharm Biotechnol*. 2016;17(1):52-70.
60. Paget S. The distribution of secondary growths in cancer of the breast. 1889. *Cancer Metastasis Rev*. 1989;8(2):98-101.
61. Duarte D, Hawkins ED, Lo Celso C. The interplay of leukemia cells and the bone marrow microenvironment. *Blood*. 2018;131(14):1507-1511.
62. Arber DA, Orazi A, Hasserjian R, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016;127(20):2391-2405.
63. Bennett JM, Catovsky D, Daniel M-T, et al. Proposals for the Classification of the Acute Leukaemias French-American-British (FAB) Co-operative Group. *British Journal of Haematology*. 1976;33(4):451-458.
64. Patel JP, Gönen M, Figueroa ME, et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med*. 2012;366(12):1079-1089.
65. Vardiman JW, Thiele J, Arber DA, et al. The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia: rationale and important changes. *Blood*. 2009;114(5):937-951.
66. Herold T, Rothenberg-Thurley M, Grunwald VV, et al. Validation and refinement of the revised 2017 European LeukemiaNet genetic risk stratification of acute myeloid leukemia. *Leukemia*. 2020.
67. Eisfeld A-K, Kohlschmidt J, Mims A, et al. Additional gene mutations may refine the 2017 European LeukemiaNet classification in adult patients with de novo acute myeloid leukemia aged <60 years. *Leukemia*. 2020.

68. Sandahl JD, Kjeldsen E, Abrahamsson J, et al. The applicability of the WHO classification in paediatric AML. A NOPHO-AML study. *Br J Haematol*. 2015;169(6):859-867.
69. Stein EM, DiNardo CD, Pollyea DA, et al. Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. *Blood*. 2017;130(6):722-731.
70. DiNardo CD, Stein EM, de Botton S, et al. Durable Remissions with Ivosidenib in IDH1-Mutated Relapsed or Refractory AML. *N Engl J Med*. 2018;378(25):2386-2398.
71. Levis M. Midostaurin approved for FLT3-mutated AML. *Blood*. 2017;129(26):3403-3406.
72. Cortes JE, Heidel FH, Hellmann A, et al. Randomized comparison of low dose cytarabine with or without glasdegib in patients with newly diagnosed acute myeloid leukemia or high-risk myelodysplastic syndrome. *Leukemia*. 2019;33(2):379-389.
73. DiNardo CD, Pratz K, Pullarkat V, et al. Venetoclax combined with decitabine or azacitidine in treatment-naive, elderly patients with acute myeloid leukemia. *Blood*. 2019;133(1):7-17.
74. DiNardo CD, Pratz KW, Letai A, et al. Safety and preliminary efficacy of venetoclax with decitabine or azacitidine in elderly patients with previously untreated acute myeloid leukaemia: a non-randomised, open-label, phase 1b study. *Lancet Oncol*. 2018;19(2):216-228.
75. Hou HA, Tien HF. Genomic landscape in acute myeloid leukemia and its implications in risk classification and targeted therapies. *J Biomed Sci*. 2020;27(1):81.
76. Bachiller M, Battram AM, Perez-Amill L, Martín-Antonio B. Natural Killer Cells in Immunotherapy: Are We Nearly There? *Cancers*. 2020;12(11):3139.
77. Castaigne S, Pautas C, Terré C, et al. Effect of gemtuzumab ozogamicin on survival of adult patients with de-novo acute myeloid leukaemia (ALFA-0701): a randomised, open-label, phase 3 study. *Lancet*. 2012;379(9825):1508-1516.
78. Jongen-Lavrencic M, Grob T, Hanekamp D, et al. Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *N Engl J Med*. 2018;378(13):1189-1199.
79. Klco JM, Miller CA, Griffith M, et al. Association Between Mutation Clearance After Induction Therapy and Outcomes in Acute Myeloid Leukemia. *Jama*. 2015;314(8):811-822.
80. Vardiman JW, Harris NL, Brunning RD. The World Health Organization (WHO) classification of the myeloid neoplasms. *Blood*. 2002;100(7):2292-2302.
81. Onecha E, Rapado I, Morales ML, et al. Monitoring of clonal evolution of acute myeloid leukemia identifies the leukemia subtype, clinical outcome and potential new drug targets for post-remission strategies or relapse. *Haematologica*. 2020.
82. Greif PA, Hartmann L, Vosberg S, et al. Evolution of Cytogenetically Normal Acute Myeloid Leukemia During Therapy and Relapse: An Exome Sequencing Study of 50 Patients. *Clin Cancer Res*. 2018;24(7):1716-1726.
83. Garcia M, Juhos S, Larsson M, et al. Sarek: A portable workflow for whole-genome sequencing analysis of germline and somatic variants. *F1000Res*. 2020;9:63.
84. Saunders CT, Wong WS, Swamy S, Becq J, Murray LJ, Cheetham RK. Strelka: accurate somatic small-variant calling from sequenced tumor-normal sample pairs. *Bioinformatics*. 2012;28(14):1811-1817.

85. Van Loo P, Nordgard SH, Lingjaerde OC, et al. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A*. 2010;107(39):16910-16915.
86. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. *PLoS Comput Biol*. 2016;12(4):e1004873.
87. Olshen AB, Bengtsson H, Neuvial P, Spellman PT, Olshen RA, Seshan VE. Parent-specific copy number in paired tumor-normal studies using circular binary segmentation. *Bioinformatics*. 2011;27(15):2038-2046.
88. Venkatraman ES, Olshen AB. A faster circular binary segmentation algorithm for the analysis of array CGH data. *Bioinformatics*. 2007;23(6):657-663.
89. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220-1222.
90. Edwards DKt, Watanabe-Smith K, Rofelty A, et al. CSF1R inhibitors exhibit antitumor activity in acute myeloid leukemia by blocking paracrine signals from support cells. *Blood*. 2019;133(6):588-599.
91. Nagarajan S, Rao SV, Sutton J, et al. ARID1A influences HDAC1/BRD4 activity, intrinsic proliferative capacity and breast cancer treatment response. *Nat Genet*. 2020;52(2):187-197.
92. Wu J, Xiao Y, Sun J, et al. A single-cell survey of cellular hierarchy in acute myeloid leukemia. *J Hematol Oncol*. 2020;13(1):128.
93. Patel C, Stenke L, Varma S, et al. Multidrug resistance in relapsed acute myeloid leukemia: evidence of biological heterogeneity. *Cancer*. 2013;119(16):3076-3083.
94. Hackl H, Steinleitner K, Lind K, et al. A gene expression profile associated with relapse of cytogenetically normal acute myeloid leukemia is enriched for leukemia stem cell genes. *Leuk Lymphoma*. 2015;56(4):1126-1128.
95. Haas BJ, Dobin A, Li B, Stransky N, Pochet N, Regev A. Accuracy assessment of fusion transcript detection via read-mapping and de novo fusion transcript assembly-based methods. *Genome Biology*. 2019;20(1):213.
96. Liao Y, Smyth GK, Shi W. featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*. 2014;30(7):923-930.
97. Robinson MD, Oshlack A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biology*. 2010;11(3):R25.
98. Draminski M, Rada-Iglesias A, Enroth S, Wadelius C, Koronacki J, Komorowski J. Monte Carlo feature selection for supervised classification. *Bioinformatics*. 2008;24(1):110-117.
99. Garbulowski M, Diamanti K, Smolińska K, et al. R.ROSETTA: an interpretable machine learning framework. *bioRxiv*. 2020:625905.
100. Boila LD, Sengupta A. Evolving insights on histone methylome regulation in human acute myeloid leukemia pathogenesis and targeted therapy. *Experimental Hematology*. 2020.
101. Hackl H, Astanina K, Wieser R. Molecular and genetic alterations associated with therapy resistance and relapse of acute myeloid leukemia. *Journal of Hematology & Oncology*. 2017;10(1):51.
102. Kroeger H, Jelinek J, Estécio MRH, et al. Aberrant CpG island methylation in acute myeloid leukemia is accentuated at relapse. *Blood*. 2008;112(4):1366-1373.

103. Zampini M, Tregnago C, Bisio V, et al. Epigenetic heterogeneity affects the risk of relapse in children with t(8;21)RUNX1-RUNX1T1-rearranged AML. *Leukemia*. 2018;32(5):1124-1134.
104. Peters TJ, Buckley MJ, Statham AL, et al. De novo identification of differentially methylated regions in the human genome. *Epigenetics & Chromatin*. 2015;8(1):6.
105. Tong J, Helmy M, Cavalli FM, et al. Integrated analysis of proteome, phosphotyrosine-proteome, tyrosine-kinome, and tyrosine-phosphatome in acute myeloid leukemia. *Proteomics*. 2017;17(6).
106. Foss EJ, Radulovic D, Stirewalt DL, et al. Proteomic classification of acute leukemias by alignment-based quantitation of LC-MS/MS data sets. *J Proteome Res*. 2012;11(10):5005-5010.
107. Aasebø E, Berven FS, Hovland R, et al. The Progression of Acute Myeloid Leukemia from First Diagnosis to Chemoresistant Relapse: A Comparison of Proteomic and Phosphoproteomic Profiles. *Cancers (Basel)*. 2020;12(6).
108. Branca RM, Orre LM, Johansson HJ, et al. HiRIEF LC-MS enables deep proteome coverage and unbiased proteogenomics. *Nat Methods*. 2014;11(1):59-62.
109. Du R, Liu B, Zhou L, et al. Downregulation of annexin A3 inhibits tumor metastasis and decreases drug resistance in breast cancer. *Cell Death & Disease*. 2018;9(2):126.
110. Zhou T, Liu S, Yang L, Ju Y, Li C. The expression of ANXA3 and its relationship with the occurrence and development of breast cancer. *J buon*. 2018;23(3):713-719.
111. Cullen P. Case of Splenitis Acutus. *Edinb Med Surg J*. 1811;7(26):169-171.
112. Stone RM, Mandrekar SJ, Sanford BL, et al. Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3 Mutation. *N Engl J Med*. 2017;377(5):454-464.

Acta Universitatis Upsaliensis

Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine 1709

Editor: The Dean of the Faculty of Medicine

A doctoral dissertation from the Faculty of Medicine, Uppsala University, is usually a summary of a number of papers. A few copies of the complete dissertation are kept at major Swedish research libraries, while the summary alone is distributed internationally through the series Digital Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine. (Prior to January, 2005, the series was published under the title "Comprehensive Summaries of Uppsala Dissertations from the Faculty of Medicine".)

Distribution: publications.uu.se
urn:nbn:se:uu:diva-425965

ACTA
UNIVERSITATIS
UPSALIENSIS
UPPSALA
2021