RPPA-profiling in pediatric acute leukemia: pattern recognition and therapeutic guidance

Hoff, Fieke

DOI:
10.33612/diss.168765254

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2021

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.
Chapter 1

Introduction

This chapter was edited from the following articles:

**Hoff FW**, van Dijk AD*, Kornblau SM. Protein expression in acute leukemia; evaluation of disease- and age-specific expression patterns.

Published in: X Zhan (Ed.), Proteoforms (pp. 62-79). London: IntechOpen, 2020
DOI: 10.5772/intechopen.90329.

Van Dijk AD*, **Hoff FW**, Kornblau SM. Proteomics in acute myeloid leukemia; guidance to targeted therapy selection and risk stratification.


**Hoff FW**, Lu Y, Kornblau SM. Antibody screening.

DOI: 10.1007/978-981-32-9755-5_8

**Hoff FW**, Hu CW, Qutub AA, de Bont ESJM, Horton TM, Kornblau SM. Shining a light on cell survival signaling in leukemia through proteomics: relevance for the clinic

DOI: 10.1080/14789450.2018.1487781

*Contributed equally to the work*
Introduction

Acute leukemia

Pathophysiology
Hematopoiesis is the formation of new blood and immune cells, and is sustained throughout life. It entails the generation of stem cells (capacity of self-renewal), proliferation and maintenance of the multipotent progenitors, and lineage commitment and maturation (differentiation). Multipotent hematopoietic stem cells mainly reside in niches of the bone marrow (BM), and are the apex of a hierarchy of numerous progenitor cell stages. Those progenitor cells differentiate into two major cell lineages; the myeloid and lymphoid lineage. It is a complex and well controlled process that requires support from both cellular and soluble regulator molecules (e.g., growth factors, chemokines and cytokines), provided by the stem cell itself and the surrounding microenvironment (e.g., bone and stroma), but also involves a global change of gene expression.

Although the exact etiology is unknown, acute leukemia results from a series of mutational events that take place during this complex process of hematopoiesis. This genetic reprogramming of the leukemic progenitor cell (“blast”) results in one or more clones of abnormal cells with unlimited proliferative capacity, that are blocked in their ability to differentiate. The accumulation of blasts interferes with the production of mature blood cells, causing neutropenia, thrombocytopenia and anemia, which can be rapidly fatal if left untreated. According to the lineage of origin of the progenitor cells, acute leukemia can be classified as either acute lymphoblastic leukemia (ALL) or acute myeloid leukemia (AML).

Acute lymphoblastic leukemia
ALL forms a heterogeneous group of leukemias composed of immature B-, T-, or uncommon variants (i.e., NK-cells), of which the majority is pre-B ALL (85–90% in children vs. 75% in adults). Chromosomal aberrations are the hallmark of ALL and are often used to categorize patients. In B-ALL, recurrent chromosomal abnormalities are found in 80% of the patients, including numerical and structural changes as translocations, deletions and inversions. The frequency of certain chromosomal abnormalities in B-ALL varies considerably between children and adults. B-ALL is also associated with certain single nucleotide polymorphisms in specific genes, including GATA3, ARID5B, IKZF1, CEBPE and CDKN2A/B. In T-ALL,
chromosomal translations occur less frequently compared to B-ALL (approximately 50–60%) and unlike in B-ALL, their prognostic impact is not well defined and they are not used for risk stratification. They are involved in both the T-cell receptor and the non-T-cell receptor loci on the chromosome or aberrant expression of the transcription factor oncogenes. There is less association with age.

*Acute myeloid leukemia*

Similar to ALL, AML is a genetically heterogeneous disease that, following the hypothesis of Kelly and Gilliland, requires collaboration of at least two classes of mutations; type I and II. Type I mutations, including FLT3, KIT, N-RAS, K-RAS and PTPN11, are involved in proliferation and self-renewal, whereas type II aberrations, including gene translocations affecting hematopoietic transcription factors that result in gene-fusion proteins as PML-RARα, AML1/ETO and CBFβ/MYH11, are mainly involved in maturation arrest and impairment of apoptosis of cells. Although the observation that particular mutations frequently accompany chromosomal rearrangements, ~ 40% of the AML patients are missing mutations in genes involved in signaling pathway components.

*Epidemiology*

Acute leukemia can occur at every age. The estimated annual incidence of ALL worldwide is 1-5 per 100,000. ALL is equally a disease of children and adults, with 75% of the pediatric cases occurring in children < 6 years of age. A second peak of incidence is seen in adults > 60 years old. ALL is the most common cancer in children, and is four times more common in children than AML. AML is the most common form of acute leukemia in adults with a frequency of 3-5 cases per 100,000 people. The incidence strongly increases with age (median of 65 years).

*Prognosis*

Survival rates for pediatric ALL have improved dramatically over the last two decades, with a current 5-year overall survival (OS) up to 90% of the patients. However, within certain very high-risk subgroups (e.g., infants or children ≥13 years of age) survival is < 80%. Survival in adults with ALL is lower with OS rate of around 45%, and patients above the age of 60 have inferior outcomes with only 10–15% long-term survival. This is, at least partially, due to higher risk of medical comorbidities, the inability to tolerate standard chemotherapy regimens,
and age-related unfavorable intrinsic biology such as Philadelphia chromosome positive, hypodiploidy and complex karyotype. However, as even the adolescents and young adults who lack medical comorbidities do significantly worse compared to their younger counterparts, the different underlying biology must be affecting response to current therapies.

Prognosis of AML is worse with a 35-40% OS for adult patients below the age of 60 years, and only 5-15% for those above the age of 60 years. Outcome in older patients who are unable to receive intensive chemotherapy without unacceptable side effects remains lower, with a median survival of only 5-10 months. Although, survival rates in the pediatric AML population have improved greatly with intensive therapy and stem cell transplant, survival rates remain 65–70%. Curability after relapse is dismal for most, ranging from <10% in adults and children with AML to 57% in late relapse ALL in children.

**Challenges in acute leukemia**

Despite improvement in OS, acute leukemia remains a deadly disease with a significant proportion of the patients who cannot be cured. In addition, treatment-related morbidity and mortality is challenging. To improve therapeutic outcome and to reduce side effects, identification of low- and high-risk patients is crucial. Risk stratification has traditionally been based on clinical features like age, performance status and prior myelodysplasia in adults, complemented by cytogenetic abnormalities, and more recently by incorporation of prognostic information related to specific gene mutations according to the World Health Organization classification system. The transformation of normal cells to leukemia involves the mutations of many genes, some of which are driver mutations, while others contribute to the leukemia phenotype. Incorporating genetic mutation analysis into risk stratification have yielded significant prognostic information and enabled stratification of patients into “favorable”, “intermediate” and “unfavorable” risk groups. However, about half of the AML patients fall in the intermediate risk group and prognosis in this group is highly heterogeneous.

Furthermore, a better understanding of the heterogeneous molecular pathophysiology of acute leukemia is needed to improve therapeutic outcome in high-risk patients and relapsed acute leukemia. Defining what signaling pathways and regulatory network dependencies are crucial to driving the underlying malignancy would facilitate the use of targeted therapies on an individualized basis. The heterogeneity in acute leukemia pathophysiology, arising from the near infinite possible combinations of mutations, creates a major challenge to personalized
medicine. Drug initiatives to target mutations in AML have been successful for specifically targeting the mutant forms of IDH1 and 2,14 and all forms of FLT3 (wildtype and mutated15), as well as for the chromosomal aberration BCR-ABL1 in ALL, but so far have been unsuccessful for several other commonly mutated genes (NPM1, DNMT3a, TET2, ASXL1, RUNX1, etc.),16,17 making the translation of the identification of the mutations that underlie the disease into improved clinical outcomes a largely unfulfilled promise. Also, comparison of AML molecular profiles across age group have shown significant differences in mutated genes, structural variants, and DNA methylation patterns, and several of the most common mutations in adults are far less prevalent in pediatric AML compared to adult AML.18 Given the difference in outcome between pediatric and adult acute leukemia, this raises the question of how leukemia in children differs from adults, with the greater goal of allowing information on biology and clinical utility from one age group to be applicable to the other. For the same reason, comparing AML to ALL, caused by distinct genetic aberrations and associated with a significant different prognosis under the current treatment strategies, would be informative. We will further discuss this in the discussion of this thesis.

In addition to genetic mutations, there are many other events that influence tumor development and drug resistance; expression and translational control from small RNA (usually non-coding RNA; e.g., microRNA, small interfering RNA),19 environmental effects from mesenchymal stromal cells,20 and both local and distant chemokine and cytokine contributions,21 which all affect how emerging leukemia cells develop and behave. The combined contributions from these external effectors arise from outside the leukemic cell, and are therefore independent of the genetic events that drive the leukemic cell. This further complicates translating knowledge of molecular events into effective therapies. A means to understand the integrated effect of internal genetic events and external biology influences on the leukemic cell is therefore required to optimally understand how to perturb the malignant cell physiology for maximal therapeutic benefit.

Heterogeneity does not only exist between patients, but also within the leukemic blasts of an individual patients. Growing evidence supports the role of clonal heterogeneity in therapeutic resistance, recurrence and poor outcome in leukemia. The ability to infer tumor heterogeneity from bulk leukemic cells is inherently limited, and should be evaluated by using single cells approaches.22 Moreover, single cell analysis will enable reconstruction of the mutational revolution, that had likely resulted in the clonal evolution of the leukemic cell, and could reveal
differences between the hematopoietic stem cells and the more matured progenitor cells. Previous studies have shown that hematopoietic stem cells or leukemic stem cells have different patterns of gene/protein expression and activation, compared to more differentiated progeny that define their unique biology.\textsuperscript{23,24} Although not studied in this thesis, we will further discuss the application of single cell proteomics in the discussion of this thesis.

**Proteomics**

*Why proteins?*

Proteins are the central effectors that drive cell function and are the combined consequences of the genetic and epigenetic events, along with influences from the BM microenvironment. The application of proteomics could potentially measure and assess the summary effects (i.e., expression and activation) of the upstream mutational events, which in turn could enhance risk stratification and prognostication. Although there are many mutations possible for each gene, there are in essence three net effects on protein function: a protein can 1) gain or 2) lose normal function, potentially to a variable degree (complete, or partial gain/loss of function), or 3) gain a new function. The plurality of genomic alterations raises important questions, including which of these multiple genomic alterations in an individual patient’s tumor is driving the cancer and should therefore be the target for therapy? Most current therapies do not target the tumor genome directly, but are directed at protein targets. These targets are often cell surface receptors, protein signaling molecules regulating apoptosis and growth, or protein kinases and/or their substrates. Hence, knowing that an individual mutation exists may not fully translate into a definable effect on the cell biology, especially when it may occur along with other mutations that are also affecting cell biology. This complicates knowing how to leverage the presence of a mutation into a decision on what therapy to use.

While gene expression profiling has the ability to comprehensively define the level of mRNA expression, it cannot serve as a surrogate for protein measurement as transcription has only limited correlation with translation.\textsuperscript{25-27} A given mRNA level can be translated to produce more than one protein due to alternative splicing, and since genomic techniques cannot assess posttranslational modifications (PTM), protein activation states cannot be measured. Proteomic methods can infer protein activation states by assessing proteins using antibodies specific for protein PTM or in mass spectrometry (MS) through recognition of the mass changes associated with the PTM (e.g., addition of a phosphate group to a tyrosine, serine or threonine, or a sugar
molecule etc.). Therefore, the assessment of protein abundance and PTM activation states has the potential to provide information representing the summary of genetic, epigenetic, and environmental actions on the cell. By summarizing the effects of genetic and epigenetic information, proteomics could provide both prognostic information and therapeutic guidance.

Proteomic methodologies
Over the years many methods have been employed that look at individual proteins (i.e., western blotting, enzyme-linked immunosorbent assays). Proteomics, the study of multiple proteins simultaneously, allows a broader picture of the protein composition. The two most frequent high-throughput methodologies currently used in leukemia research are MS and the reverse phase protein array (RPPA).\textsuperscript{28,29}

Mass spectrometry
Traditionally, MS-based approaches are used to identify and quantify proteins from highly complex biological samples. MS starts with the ionization of samples which breaks down the sample into charged fragments (ions). Those ions can be sorted and measured by accelerating and subjecting them to an electric or magnetic field based on their mass-to-charge ratio (m/z). Identification of the protein follows based on the abundance of those m/z-fragments.\textsuperscript{30}

Globally, proteins can be ionized with two distinct methods: matrix assisted laser desorption/ionization (MALDI) and electrospray ionization (ESI). In MALDI the protein sample is mixed with an energy absorbing matrix. Irradiation of this matrix causes vaporization of the matrix together with the sample, resulting in the formation of ions.\textsuperscript{31} ESI creates ions using electrospray to dissolve the protein lysate, by applying high-voltage to the dissolvent to create an aerosol of small charged fragments.\textsuperscript{32} When a protein sample is highly complex, samples may require separation prior to MS analysis using 1D or 2D gel electrophoresis, high-pressure liquid chromatography, or gas chromatography to maximize the sensitivity. Because different variants of a protein are derived from a single gene, they often contain homologous sequence regions, and because of the digestion step, information about the relationship between amino acid sequence could be lost.

Reverse phase protein array
RPPA is an antibody-based proteomic approach (Figure 1). The name “reverse phase” indicates that the antigens (protein lysates) are printed on the array and subsequently probed with a
primary and a secondary antibody that target specific proteins (or protein variants) in the solid phase lysates. This is the “reverse” of traditional immunoassays in which specific antibodies that are immobilized on the solid phase of the array to capture the antigen of interest. Because RPPA requires the availability of highly-specific antibodies, it cannot be used as a de novo protein discovery platform.

**Figure 1. RPPA flow chart.** Figure illustrating the full RPPA workflow starting with protein lysate preparation of acute leukemia samples, printing of the lysates in five serial dilutions along with negative and positive control samples, staining of the slides (primary, secondary antibodies, signal detection reagents), readout of spot intensities, data normalization, background correction, and finally data analysis.

Briefly, proteins extracted from patient tissues or cultured cells are denatured followed by serial dilution (in order to define antigen-antibody reaction in a linear range for accurate quantification). Serially diluted cellular proteins are arrayed on nitrocellulose-coated slides (>1000 samples can be printed on a single slide) and probed with validated antibodies that recognize signaling molecules in their functional state. Those signals are captured by a tyramide dye deposition and a 3,3'-diaminobenzidine colorimetric reaction or by infrared fluorescence labeling. Data are collected and quantitative analysis is performed using custom software. Features of the software include automated spot identification, background correction (i.e., topographical normalization33), controlling for location, serial dilution-signal intensity curve construction (i.e., “SuperCurve” normalization34), and concentration determination (i.e., loading control35). The values derived from the slope and intercept are expressed relative to standard control cell lysates or control peptides on the array. These values
indicate the levels of protein expression and modification (e.g., phosphorylation, cleavage of histone modification based on antibody specificity). If samples are printed on different slides, replicates-based normalization is used to align samples from two different slides using replicate samples that are printed on the two slides.

Advantages of the RPPA technology

RPPA is a high-throughput methodology and has the big advantage that, in contrast to MS, it can measure the expression levels of thousands of samples simultaneously. This enables assessment of correlation between protein expression and multiple patients and disease characteristics, making it a suitable approach for clinical applications. In addition, RPPA is a cost-effective and sample sparing technique, as it requires only a minimal amount of protein samples (approximately $3 \times 10^9$ cells for 400 proteins). It has a high sensitivity (can detect proteins in the picomole-to-femtogram range) and an excellent precision (coefficient of variation < 15%). Our group and others have shown the high inter- and intra-slide reproducibility, precision of sample spotting, throughput, and reliability of RPPA. As the technology is based on an approach were samples are first printed on a large number of slides that are subsequently analyzed by using antibodies, it is also possible to generate and store additional slides (sample arrays), so that further analysis later on can be performed when new affinity reagents become available or when new hypotheses need to be tested using the same samples.

Challenges of RPPA in acute leukemia

The major advantage of studying leukemia is that primary leukemic samples can be easily isolated from peripheral blood (PB) samples or BM aspirates that are obtained during diagnostic procedures, without performing additional bone marrows on patients. In order to obtain a representative analysis of the leukemic cells, correct cell purification and sample handling (i.e., those changes occurring between sample collection and assay processing) remain challenging.

To acquire a representative protein lysate of the leukemic blasts, Ficoll separation is applied to remove neutrophils and red blood cells from the collected samples. Ficoll separation is an isolation and enrichment technique of the mononuclear cell fraction that uses Ficoll-Paque media (which has a density of 1.077 and 1.084 g/ml) and centrifugation to separate cell
subtypes and to remove dead cells based on cell density (lymphocytes, monocytes, and platelets are not dense enough to penetrate into the Ficoll-Paque media layers). During one of the first RPPAs that our group generated, contaminating non-tumor cells were adequately removed from the collected samples, and the purified samples were cryopreserved so that protein lysates could be made at a later time point. However, after thawing the cryopreserved cells, a proportion of the cells had presumably died because of stress caused by the cold and DMSO media used to cryopreserve cells, whereas Ficoll separation was not applied a second time. As a result, when protein expression levels were measured, unrealistic high expression of apoptotic proteins were detected, likely as a signal from the dead cells.

Even when Ficoll separation was applied adequately, cryopreservation then led to a second problem of an altered set of cells that you are profiling. Our group observed that when samples were not prepared from fresh materials at the day of collection, but from cells that were first cryopreserved, measurement of protein expression resulted in abnormal high signals of several other signaling proteins. The AML719 array included many samples of which more than half were prepared from cells that were first cryopreserved, and the remaining samples from cells that were immediately processed after collection of the cells (“fresh”). Comparison of protein expression from fresh and cryopreserved samples resulted in striking differences. At the individual protein level, 85.5% (197/228) of the proteins showed statistically significant differences in expression between fresh and cryopreserved samples. Investigation of the expression differences at the protein pathway level resulted in expression patterns in 23 of 31 protein functional groups that were only seen in cryopreserved samples, but not in fresh samples. Since fresh samples were uninfluenced by the potential effects of cryopreservation or thawing, we assumed that they represent a more accurate picture of the AML biology compared to cryopreserved samples. Therefore, future arrays are now restricted to samples that were prepared fresh.

Because RPPA can assess protein abundance and activation states in large numbers of samples using small amounts of material, this method is ideal for use in multi-institution clinical trials. However, in order to obtain those samples, samples often have to be shipped from one institution, where cells were collected, to another where protein lysates are prepared, or where lysates are analyzed. As there is concern that preanalytical handling variables can affect the integrity of protein concentrations, we will seek to define the variability in preanalytical handling variables to determine if instability in protein expression would adversely affect
protein assessment. In Chapter 2, we will address several questions covering possible handling effects, including the effect of transit time, shipping, temperature, different types of collection tubes, and the effect of chemotherapy.

Another major challenge that often occurs during cell culture is mycoplasma infection. Despite significant improvements in diagnostic and therapeutic possibilities to detect and eliminate mycoplasma contamination, it is still present in 5-30% of the cell lines.\textsuperscript{38,39} As limited knowledge is available on the effects of mycoplasma on protein expression, we aim to determine protein expression levels in mycoplasma infected and non-infected cell lines along with post treatment mycoplasma-free versions. Answers to these questions will be discussed in Chapter 3.

\textbf{Methodological challenges of proteomics}

The first question that can be answered with the use of proteomic data is whether individual proteins convey prognostic information. Previously, our group has performed RPPA on 511 adult AML patient samples, and identified several protein, including Forkhead O Transcription Factor 3A (FOXO3A),\textsuperscript{40} Bromodomain Containing 4 (BRD4) (ASH 2017 #3794), Absent, Small, Or Homeotic Discs-like Protein (ASH2L),\textsuperscript{41} Tripartite Motif Containing 62 (TRIM62),\textsuperscript{42} Friend Leukemia Virus Integration 1 (FLI1),\textsuperscript{43} protein phosphatase 2A (PP2A) regulatory subunit B55α,\textsuperscript{44} and phosphorylation of Glycogen Synthase Kinase 3 α/β (GSKα/β)\textsuperscript{45} as prognostic factors associated with survival.

As it is the net consequence of the combined influences of all the proteins that determines the summed effect on the cell, rather than a single protein, over the years, this focus has developed from analyzing individual proteins to a more systematic biological approach, capable of simultaneously looking at hundreds of proteins. This approach started with analyzing proteins in the context of a \textbf{“Protein Functional Group”} (PFG). A PFG is formed by proteins that are known to be functionally related from the literature, or based on strong correlation with the RPPA data set, and so takes previously known relationships into account which the traditional unsupervised hierarchical clustering does not. Within each PFG, a mathematical algorithm (Progeny Clustering\textsuperscript{46}) was applied to identify an optimal number of \textbf{“Protein Clusters”}; a subset of cases with similar (correlated) expression of core protein functional group components. As an example, van Dijk et al. analyzed the \textquotedblleft Histone modification\textquotedblright PFG, formed
by 20 histone modification proteins.\textsuperscript{47} They identified five different protein clusters associated with varied expression or activation states relative to normal CD34\textsuperscript{+} cells. Among these clusters, the “on”-state (i.e., expression above that of the normal CD34\textsuperscript{+} cells) of the histone modifier proteins was associated with a poor outcome compared to AML cells like normal CD34\textsuperscript{+} and “off” states. Another example showed the nine proteins related to TP53, and identified seven protein clusters which were prognostic for survival.\textsuperscript{48}

Next, to obtain a cohesive understanding of the proteomic heterogeneities from the RPPA data across multiple levels (i.e., PFG, protein clusters), our group developed a novel computational procedure called the “MetaGalaxy” analysis (Figure 2).\textsuperscript{49} The MetaGalaxy analysis is a multi-step approach, that starts with the analysis of individual proteins in the context of the described PFG. To search for higher order structures between protein clusters, protein clusters from each PFG were organized in a binary matrix for each patient (i.e., “1” if a patient was a member of a given protein cluster, “0” if a patient was not a member). Using the Block Clustering algorithm, we were able to recognize the existence of recurrences between protein clusters. Those correlations were defined as “Protein Constellation” (CON); a group of protein clusters from various protein functional groups that were strongly correlated with each other. The unique and selected combinations of CONs than enabled characterization of AML patient subpopulation (i.e., “Protein Expression Signature” (SIG)).

We hypothesize that the genetic complexity of the leukemic cells, ultimately results in a constrained number of protein patterns pathway utilizations, needed for the cell to become leukemic (similar to the “Hallmarks of Cancer”).\textsuperscript{50,51} The “Hallmarks of Cancer” are a conceptual framework of 6 (and later 10) biological capabilities acquired during the multistep development of human tumors.\textsuperscript{50,51} Though all malignancies share those same hallmarks of cancer, heterogeneities between patients complicate patients’ response to therapy and subject them to varied outcomes. If this idea is correct, then each combination of genetic events, regardless of their direct downstream consequence, must somehow meet each of the hallmarks. We hypothesize that those combinations of CONs together represent the quantitative hallmarks in acute leukemia, and can significantly accelerate the identification and development of new therapeutic targets.
Figure 2. Computation work flow of the “MetaGalaxy” analysis. (A) Scheme showing the overall computational analysis, starting with individual proteins that are allocated into PFG. Protein clusters are identified within a PFG and correlated to form a CON. (B) Pathway analysis. Protein clusters were identified for each PFG and correlations between outcome, clinical and laboratory variables were determined. PCA was performed to visualize the distribution between the protein clusters relative to the normal CD34+ samples. Networks were generated to reveal different activation states, and to show the relative protein expression levels in a given protein clusters. (C) “Meta-galaxy” analysis. Co-clustering algorithm was performed a binary matrix indicating protein cluster membership for each patient. This identified the existence of CONs (horizontally). Patients that express similar combinations of CONs were defined as a SIG (vertically). Correlation between SIGs and outcome, clinical and laboratory variables were determined. Proteins that were significantly over or under expressed compared to CD34+ cells were identified for each CON and SIG.
Scope of the thesis

Curing AML in a larger proportion of patients requires recognition of features that identify patients for personalized targeted therapy. The heterogeneity in acute leukemia pathophysiology, arising from the near infinite possible combinations of mutations, creates a major challenge to personalized medicine. Proteins are the central effectors of both genetic and external forces that drive cell function. In this thesis, we hypothesize that the genetic and environmental variability present in leukemia coalesces into a more finite number of protein expression patterns, and that these expression patterns reveal key protein dependencies that could aid classification and identify therapeutic targets.

We used the RPPA methodology to study protein expression abundance and activation in acute leukemia patient samples. However, as there are many (external) factors that can potentially affect protein expression of the leukemic cells, we will start in Chapter 2 with evaluating the effect of preanalytical variables on protein expression. Factors that will be discussed are shipping (shipping vs. local processing), temperature (4°C vs. ambient temperature), collection tube type (heparin versus CS preservation tubes), treatment (pre- vs. post-chemotherapy) and transit time. In Chapter 3, we will evaluate the effect of mycoplasma contamination in cell lines and mycoplasma eradication antibiotics on protein expression in leukemic cell lines.

Because RPPA is an antibody-based approach, the reliability of the RPPA methodology is highly dependent on the quality of the antibodies that are used. Unfortunately, universally accepted guidelines or standardized methods for validating antibodies for use on RPPA have not yet been established. In Chapter 4, we will discuss the current need for better antibody validation in general as well as provide information about how RPPA antibodies are currently validated.

In Chapter 5, 6, 7 and 8 we will discuss the results of the MetaGalaxy methodology in pediatric ALL, pediatric AML, and adult acute promyelocytic leukemia with the goal of identifying expression patterns that 1) classify patients based on proteomics, 2) enhance risk stratification and 3) recognize additional targets for potential combinational therapy for subgroups of leukemia patients.
In Chapter 9, we described a role of phosphorylation of heat shock factor 1 (HSF1-pSer\textsuperscript{326}) in the treatment of pediatric AML patients with standard chemotherapy plus the proteasome inhibitor bortezomib. Samples were collected from patients that participated in the Children’s Oncology Group (COG) AAML1031 phase 3 randomized clinical trial that evaluated the effect of standard therapy vs. standard therapy plus the proteasome inhibitor bortezomib. Although this trial evaluating the effect of bortezomib was unsuccessful overall, we identified a potential therapeutic benefit of bortezomib in a selected subgroup of patients and identified proteins modulating this effect. Chapter 10 will review the applicability of proteomics in acute leukemia, and will in particular focus on how this can be translated to the clinic.

In Chapter 11 we will provide our interpretation of the results presented in this thesis and discuss the developments that are likely to be important in the future. All results are summarized in an English and Dutch Summary in Chapter 12 of this thesis.

In addition to all published data, we have contributed to build a web portal (https://leukemiaatlas.org) to make the full analysis results accessible to people worldwide. The portal enables researchers to investigate protein expression patterns within AML cells in the context of patients’ clinical and genetic features as well as protein association networks.
Reference list


35. Neeley ES, Kornblau SM, Coombes KR, Baggerly KA. Variable slope normalization
Part I

Proteomics: methodological issues of sample handling