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.

Download date: 07-12-2022
Chapter 11

General discussion and future perspectives
General discussion

Acute leukemia is the result of genetic, epigenetic and microenvironmental factors. As these co-occur in a near infinite number of combinations, it makes almost each patient genetically unique. The molecular consequences of genetic and epigenetic events are mediated by the expression and function of proteins, and most drugs that are currently under development target proteins. In this thesis, we hypothesized that this molecular heterogeneity would coalesce into a more finite number of protein expression patterns ("Protein Expression Signatures"; SIGs), and that these patterns would in turn reveal key dependencies that could aid in risk stratification and therapy selection.

In line with this thesis, this discussion is divided into three parts:

- Proteomics: methodological issues of sample handling;
- Protein expression patterns in acute leukemia;
- Clinical utility: guidance to therapy selection.

Part I: Proteomics: methodological issues of sample handling

In Part I we discussed several methodological issues of sample handling. Because RPPA is a very sensitive proteomic high-throughput technology, it is important that preanalytical handling does not affect protein expression, as this could potentially result in confounding experimental results. We concluded that variables including shipping (e.g., in the setting of multi-site clinical trials), transit time (≤ 72 hours), and temperature, only had minimal effects on proteins expression if processing of samples occurred carefully and if other conditions were held constant. Furthermore, we showed that preservation cell tubes (Cell Save (CS) vs. heparin) did not change protein expression in cells that were not yet exposed to chemotherapy. However, in leukemic cells that were collected after exposure to chemotherapy, dramatic changes in protein expression levels were observed between samples collected in CS vs. heparin tubes. We hypothesize that the protein expression stability is higher in CS tubes, as CS tubes “fix” the sample and so prevent further effects of drug therapy during the process of storage, shipping and processing. While this study was performed with relatively few patient samples and antibodies targeting a limit panel of total- and post-translational modification (PTM)-proteins, we concluded that analysis of samples collected at multiple sites as part of a clinical trial and shipped to a central site is safe. This finding is very important for future proteomic studies that
evaluate samples collected from clinical trials, as those samples often come from multiple institutions across the USA or different countries in the world. However, to maintain the best sample conditions, we still recommend that samples should be exposed to a little stress as possible, and that samples should be collected in CS tubes, refrigerated prior to shipping and covered by ice packs during shipment, and that transit time should be as short as possible. A more thorough assessment is required for definitive conclusions.

Our second study assessed the effect of mycoplasma infection and treatment in cell cultures of leukemia cell lines. Although protein expression was not globally affected by current mycoplasma infection or after eradication by antibiotic therapy, levels of individual proteins were significantly affected. Most of the altered proteins were involved in apoptotic-signaling pathways. While it may have been expected that the infection itself would affect protein expression (e.g., due to a lack of energy as the infected cell has to share its metabolism with the mycoplasma), our results also suggested that once the mycoplasma is eliminated, their metabolism remains altered and is only partially restored. A caveat is that the mycoplasma uninfected cell lines often originated from different laboratories than the infected and post-treatment samples, and were therefore possibly cultured under different circumstances. A study that begins with uninfected cell lines, and then artificially infects the cell lines, while maintaining identical culture conditions, will likely result in a more accurate description of the effect for mycoplasma infection. Based on our results, we encourage regular testing of cell cultures for mycoplasma and to treat cell lines if cultures test positive to avoid having the consequences of altered expression of these proteins confounding the experimental results. Yet, if possible, additional cell culturing for more than two weeks after completion of the antibiotic treatment, or taking a clean cell line aliquot, should also be considered.

Given the high dependency of RPPA on antibodies, we also reviewed how antibodies are currently validated. A review of the existing literature shows that antibodies often do not detect what they were supposed to detect, that they did not function robustly across different sample types, only work in some applications but not in others, were not stable over time, and did not always yield reproducible results over different batches. Several approaches to validate antibodies exist, varying from genetic strategies that knock out or overexpress the protein of interest, to demonstrating correlation with none antibody-based approaches. Validation, which is defined as “the process of demonstrating that an analytical procedure is suitable for its intended purpose”, is often inconsistent and no standardized guidelines exist. As validation of
individual antibodies requires a lot of time and money, we proposed to accurately document the results once validation of a antibodies has been done, and to make this available to other researchers. This should include basic facts (e.g., name, target, manufacturer, etc.), as well as standardized PTM-convention, usage facts (e.g., how was the antibody validated, for which sample type and application, etc.), and use of a rating metric resulting in final confidence scores (e.g., how specific was the antibody, etc.). Although this would potentially have significant utility, it would require a serious effort in terms of workload and would require large financial support to establish and maintain. However, knowing that an antibody would not work in an application or sample type of interest, on the other hand, would save effort to attempt to validate an antibody that in the end will not meet the criteria of validation.

In addition to those three studies, another (potential) issue that recently appeared, is the use of different detection methodologies on the slides (e.g., fluorescent dyes or colorimetric means). Concomitantly, different scanners are used to detect the subsequently different types of signals. This difference in methodology could confound studies that aim to perform inter-array comparisons, requiring alignment of the different arrays (e.g., comparing expression in chronic leukemia vs. acute leukemia, or pediatric leukemia vs. adult leukemia). Slide alignment uses an approach called “replicate-based normalization”,¹ which requires a minimum number of replicated samples printed on both arrays to be. With this approach, we were previously able to align arrays (that used similar detection methodologies) printed with pediatric acute myeloid leukemia (AML) and T-cell acute lymphoblastic leukemia (T-ALL) samples (manuscript in preparation). As one of the companies stopped producing di- amino-benzidine (DAB) substrate kits, we recently (2019) had to switch from using DAB-horseradish peroxidase (HRP) detection (DAB substrate reacts with HRP to yield an insoluble brown-colored product) to infrared detection. Although this will not affect the analysis of that particular array, as expression of all samples on that array is measured similarly, and RPPA measures relative expression (i.e., expression in one sample compared to expression in another sample), rather than absolute expression, this could potentially complicate the analysis of alignment different arrays that uses different protein expression detection methods. While not covered in this thesis, future analysis is needed to evaluate whether we can compare expression between arrays that either used DAB and infrared, and if not, how we can correct for this so that we can. 
Part II: Protein expression patterns in acute leukemia

In part II of this thesis, we identified the existence of SIGs in pediatric ALL, pediatric AML, and adult acute promyelocytic leukemia (APL), by applying the “MetaGalaxy” approach. The MetaGalaxy approach is a multi-step method that starts with defining “Protein Clusters” (i.e., subset of patients with similar (correlated) expression profiles within a “Protein Functional Group” (PFG) (i.e., groups of functionally related proteins based on existing knowledge from the literature), and from there searches for higher order structures between protein clusters, that allow classification of patients.

While acute leukemia is a genetically heterogeneously disease, arising from the near infinite possible combinations of mutations, we have learned that this genetic chaos coalescences into a finite number of SIGs, by showing that each complex genetic leukemia patient can be reduced to a limited number of combinations of protein constellations (CON) (i.e., strong correlation between different protein clusters from various PFG). An allegory to explain this is a menu from a Chinese restaurant (devised by prof. Kornblau, who used this figuration to explain me his rational behind the MetaGalaxy approach at the very beginning of my Ph.D.) (Figure 1). For instance, a “functional group” is a group of functionally (like) items, in this example Chinese dishes: soups, appetizes, noodles, pork, vegetables, etc. A constellation is a collection of functional group patterns that show strong correlation with each other; or a pattern of menu items that get ordered with each other. In an average family with four members (mom, dad, son, daughter), each member places very similar orders every time after time they visit a Chinese restaurant. If you apply this metric to multiple families, a pattern of recurrent associations can be observed (i.e., constellation), and based on those constellations, people can be classified into subgroups (i.e., signatures).

If you then annotate those groups, you will likely find associations with signatures. Some combinations of constellations are always ordered by women, while other are associated with children or vegans, etc. So, the complexity of food ordering (i.e., theoretically, almost each person can order a unique combination of dishes) can be reduced to a finite number of patterns (i.e., the combination of food ordering occurs following a recurrent pattern, rather than completely at random).
To translate this back to acute leukemia, dishes are replaced by proteins, and Chinese restaurant visitors by patients. Indeed, we found that the assembly of molecularly complex patients resulted in a limited number of recurrent patterns of protein clusters (i.e., CONs), and that combinations of those CONs defined a finite number of SIGs. This shows 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”, six or ten of biological capabilities that need to be fulfilled during the development of cancer).2–3 So proteomics allows us to be reductionistic and to classify patients into groups, rather than as a series of one.

In Chapter 5, 6, and 7 of this thesis, we showed that pediatric ALL and pediatric AML all could be defined by finite number of SIGs (ranging from 7-9 across the subtypes of pediatric
acute leukemia). While several CONs were shared among the SIGs, each pattern of recurrent combination of CONs was specific to a given SIG. This suggests that whereas subgroups of acute leukemia overall express protein expression profiles that separate one from the other, there are expression patterns that share mutual use of some proteins between the SIGs. This was not only observed within one disease, but also between subtypes of diseases. For instance, in Chapter 5, we compared protein expression patterns in pediatric T-cell ALL (T-ALL) to pediatric pre-B-cell ALL (B-ALL), and in Chapter 6, we compared pediatric ALL to pediatric AML. In both studies, we found that the SIGs were largely disease-specific, but that overlapping CONs existed that were expressed in subsets of both diseases. This confirms that AML, B-ALL and T-ALL are indeed the result of different underlying biology, but that there are shared protein dysregulations within some pathways, needed for the cell to become leukemic. Moreover, in Chapter 8, we showed that APL, a subtype of AML, could, overall, not be separated from non-APL AML based on proteomics, but that ~10% of the individual proteins were differentially expressed.

A notable finding in our pediatric T-ALL study, which included samples drawn from an area in Southern Texas with a large Hispanic population, was the association of certain SIGs that were enriched or depleted for the Hispanic-ethnicity. Preliminary results in a second T-ALL RPPA with 358 patient samples (data not presented in this thesis, manuscript in preparation) identified 9 SIGs, and again found clustering of Hispanics, with overrepresentation in SIG-5 (43% vs. 21% overall). This SIG was associated with unfavorable outcome. Numerous studies have shown an inferior outcome for patients with a Hispanic-ethnicity, but it is uncertain whether this arises from a different pathophysiology or from socioeconomic factors. In part, because non-Hispanic populations are disproportionally underrepresented in cancer genomic studies. Our study suggested that for some Hispanic patients, this difference in outcome could arise from underlying differences in the pathophysiology of their leukemia. So far, no major cytogenetic aberrations are more or less associated with ethnicity in T-ALL. However, several studies, including a recent study by Qian et al. showed that single-nucleotide polymorphisms (SNP) of leukemia risk genes (e.g., ERG, GATA3, ARID5B) have higher-risk allele frequencies in Hispanics. This may suggest that there could be a substantial number of other genetic variants/loci contributing to racial/ethnic disparities in ALL, but that these are not yet identified. SNPs are variants in the genome present with a minor allele frequency ≥1% (and thus not defined as genetic mutation which occur in a frequency of <1%), that fall within a coding or non-coding sequence of a gene. They can change the amino acid sequence of the
produced protein. If certain SNPs are more associated with Hispanics vs. non-Hispanics, this can explain why we found protein expression pattern associated with Hispanics, and this could have an effect on the chemotherapy sensitivity (and thus on survival rates). Genome-wide studies comparing genotype frequencies between Hispanic and non-Hispanics are needed to confirm this hypothesis, as well as larger proteomic studies to assess if and how the proteomic landscape differs, but also if SNPs affect proteins in a consistent and predictive way. This, combined together with outcome data or drug sensitivity tests, could potentially tell us more about whether the inferior outcome in Hispanic indeed is related to underlying differences in biology, and if so, how treatment strategies should be adapted (e.g., more intensive chemotherapy, use of particular (combinational) targeted therapy based on SIG suggested dependencies, etc.).

Proteome differs from transcriptome

The human genome is the total amount of DNA that each cell in the body contains. The central dogma of molecular biology: DNA is transcribed into messenger RNA (mRNA), which is then translated into proteins (DNA \(\rightarrow\) mRNA \(\rightarrow\) protein), has traditionally assumed an implied correlation between mRNA levels and protein expression, but we now know that this view was overly simplistic. Epigenetics, environmental influences, mRNA editing, alternative splicing, noncoding RNA’s, and several other processes, acting together determine the level of expression of an individual protein. Each of these processes can be determined using currently available technologies. Also, the effect of mutations (e.g., insertions, deletions, translocations, inversions, etc.) not only depends on the type of mutation, but also on the location in the mRNA (e.g., non-coding vs. coding regions, epigenetic regulators, transcription factors), resulting in loss or gain of function and regulation of the proteins. Additionally, after translation of the RNA transcript, proteins undergo multiple PTMs (e.g., phosphorylation, acetylation, glycosylation, ubiquitination, cleavage etc.) affecting the protein function and activation, localization, lifespan and activity. All of these events individually, and the sum of them in combination, degrade the correlation between raw mRNA levels and protein level and functional status.

In Chapter 7, we measured protein expression of 296 protein targets, including 229 different antibodies directed against total proteins in pediatric AML. As we also had RNA-sequencing data available for 205 of the 229 total proteins, we were able to correlate mRNA expression to protein abundance levels in 390 patients. Similar to previous studies,⁴ ⁷ we confirmed a low
correlation between mRNA and total protein levels of R=0.17. Further stressing the difference between mRNA and protein, the protein expression SIGs could not be reproduced by RNA-sequencing data. However, while RNA levels did not correlate with protein data, we found associations between SIGs and cytogenetics and mutation states, which make sense as “driver” mutations would be expected to have a defining effect on the biology/protein expression, even though combinations of events might further alter some of these signals. Patients with the MLL-rearrangement tended to fall more frequently into one SIG, though, MLL-rearrangement did not exclusively result into this SIG. In contrast, t(8;21) and inv(16), both characterized by aberrations at core-binding factor genes, were not characterized by overlapping CONs. Again, this reflects our hypothesis that genomics has pleiotropic effects on the proteomic phenotype (e.g., by the effect epigenetic modification and environmental circumstances), and that genomic diversity can result in a common protein profile. What this could mean for future treatment strategies will be discussed in Part III.

**Integromics**

Although, in this thesis, we only measured protein expression, a future challenge would be to integrate different technologies, each measuring one of the mentioned processes. This approach is defined under the term “integromics”. For each process, the effect on a pathway can be determined, and the combined analysis can then show how a pathway is affected by different processes. Studying the DNA (e.g., whole genome sequencing) combined with technologies that assess histone methylation or acetylation (e.g., ATAC-sequencing) can predict whether a region of the DNA is silenced or transcribed, and sequencing of small RNA can identify over- or under expression of non-coding RNA fragments (e.g., microRNA, small interfering RNA) that can silence the DNA transcription. Thus, while you would expect activation of certain pathways based on activating gene mutations, this expectation can completely change after analyzing the integrated data of different processes.

What can we learn from this, that cannot be learned from proteomics alone? Knowing that a protein has a low expression (based on proteomics), leads to the inquiry of identifying the cause of this observed decrease in protein expression (and how this expression and/or function can be restored). If ATAC-sequencing shows evidence of epigenetic repression, then the expression can hypothetically be reestablished by treatment with hypomethylating agents (or demethylating agents), which inhibit DNA methylation, or by using histone deacetylase inhibitors that inhibit histone deacetylases, and so potentially restore protein expression by
remodeling the chromatin structure. However, if epigenetic studies do not point toward epigenetic repression, but whole exome sequencing identifies mutations associated with a loss of function, then epigenetic modifiers will not be of any benefit, and restoring the protein function is much more difficult. The classic example of replacement of a lost function has been successfully demonstrated in APL, where RARα in the fusion gene cannot reach the nucleus, resulting in a loss of function, but all-trans retinoid acid can replace this lost function, enabling differentiation and growth arrest of promyelocytes. This technically difficult therapeutic goal of replacing a lost function is also currently being tested by using mimetic agents. The use of a small activation RNA (saRNA), in which short double-stranded RNA (dsRNA) is used to restore gene expression of the CCAAT/enhancer binding protein alpha (C/EBPα), or the saRNA use in prostate cancer cell lines that reactivates the deleted KLF4 transcription factor, resulting in inhibition of cell proliferation and survival.

On the other hand, a protein with a high expression can be inhibited at the protein level by small molecule inhibitors. The paradigm for this is the use of tyrosine kinase inhibitors (e.g., imatinib, nilotinib and dasatinib) to suppress the constitutively activated ABL kinase activity seen in leukemia harboring the Philadelphia-chromosome. Also, delivery of small interfering RNA (siRNA) via nanoparticles has shown promise, although they often have a weak stability, and clinical usage has been limited by less than optimal pharmacokinetics and biodistribution of siRNA. Another relatively new approach is the use of PROteolysis TArgeting Chimeras (PROTACs) as emerging anticancer therapeutics. PROTACs are used to degrade proteins that are important for tumorigenesis. They are molecules consisting of one ligand that binds to the protein of interest, and a second ligand that binds to an E3 ubiquitin ligase, and so cause ubiquitination and degradation of the protein by the ubiquitin-proteasome system.

**Tumor heterogeneity**

One of the limitations of this thesis is that we only looked at bulk leukemic samples, rather than at single cells. Heterogeneity exists not only between different acute leukemia patients (“inter-leukemia heterogeneity”), as demonstrated by the classification of leukemia subtypes (e.g., World Health Organization classification), and by recognition of our identified SIGs, but also exist within an individual leukemia case (“intra-leukemia heterogeneity”).

Acute leukemia is clonal disorder that is the consequence of multiple genetic mutations that accumulate throughout life and throughout the life of the disease (new and separate clones in
the tumor evolve over time).\textsuperscript{13} The classical model of the development of acute leukemia starts with initial “pre-leukemic” mutations. While most mutations can be repaired, or are temporarily present and disappear when the cell dies (“passenger mutations”), others confer an advantageous phenotype (“driver mutations”) resulting in a clonal advantage over the non-mutated cells. Over time, these pre-leukemic clones can obtain additional mutations that lead to the transformation into acute leukemia. This clonal evolution can be seen as a multistep process that develops in space and over time. It is conceivable that many factors influence the timing of the appearance of the mutation and the positive selection of such a clone. Next generation sequencing technologies as well as single cell sequencing have revealed leukemic complexity, whereby multiple genetically distinct subclones can co-exist within one patient (“clonal heterogeneity”). Measuring protein expression in bulk leukemic cells, which takes the average of all cells, does not differentiate between expression in individual cells, and therefore misses distinct proteomic profiles that are the result of clonal heterogeneity.

Clonal heterogeneity is one of the current challenges in treatment strategies of acute leukemia, as resistant subclones can escape chemotherapy and give relapse of disease.\textsuperscript{14} Sequencing studies have shown that the primary \textit{de novo} clone, often differs from the relapsing clone. This highlights that future research is needed to develop proteomic methodologies that can elucidate the leukemic biological processes at the single cell level. In addition, single cell studies can aid to reconstruct the mutational histories characterized by linear and branching patterns of clonal evolution. The ultimate goal would be to provide patients with individualized, combinational, subclone-specific targeted therapies.

Another limitation that applies to bulk studies is the “\textit{Simpson’s paradox}”.\textsuperscript{15} Though it was expected that the same genome and the same environment give rise to identical proteomes, it is known that individual cells differ in their proteome, with sometimes different functional consequences. Simpson’s paradox describes that as it is likely true that if you take a population-average of protein levels, there seems to be a particular correlation between proteins. However, when you look at the individual cells, proteins can have the same correlation in one cell, but the complete inverse relationship in the other. Taken together, two main issues arise when studying bulk cells are: 1) different clones result in different proteomes, and 2) cells differ in their proteome, even if they contain an identical genome. The environment of the cells also likely plays a big role; some cells are floating in the blood, others trapped in a niche in the
marrow, what chemokines, cytokines and other neighboring cells are interacting with the blast to affect its proteome?

If we could study thousands of individual cells, from different locations, then it would be possible to model the direct and indirect interactions and correlation between proteins with a certain probability. In particular when performed in samples collected from large clinical trials, single cell proteomics would allow systematic investigation of predictive and prognostic impact of clonal diversity acute leukemia. More recently proteomic approaches that allow for single cell analyses have started to emerge (e.g., CyTOF), however there remains a tradeoff between the number of proteins that can be measured and the methodology selected. RPPA allows for no limit (provided enough slides can be printed and antibodies validated), while CyTOF is currently limited to about 50 protein measurements per cell, also limited by antibody validation. Thus, a broader profile of an individual cell still remains technically beyond reach as of today.

Assessing dynamic change
In addition to measuring protein expression in the treatment naive leukemic cells, another consideration is to measure (changes in) protein expression and activation (i.e., post chemotherapy initiation, defined as “post-chemotherapy”) during a cycle of chemotherapy, immunotherapy or antibody-based therapy (e.g., bispecific T-cell engagers; BiTEs). We hypothesize that the change in expression in cells to stress (i.e., treatment) can be a key to whether a cell is responsive or resistant. In response to therapy, cancer cells may reduce their dependence on a particular hallmark capability or (protein) signaling pathway, becoming more dependent on another, representing a form of acquired drug resistance. Looking at post-treatment expression and activation of proteins may provide insights into biological effects of drugs, and mechanisms of drug resistance. This can either be done from static expression levels post-treatment at a given time point, or from the dynamic change in expression during treatment (i.e., expression post-treatment minus expression pre-treatment). Particularly, in acute leukemia, where blood can easily be drawn from the patient without performing any additional invasive procedures, expression can be measured at several time points during treatment. This dynamic adaptability then holds additional promise for improved identification of key targets on an individualized basis.
While this approach is conceptually simple, in reality it is much more complicated. In Chapter 7 we did a first attempt in assessing the dynamic change in ~400 pediatric AML patients. We collected paired samples pre-treatment, and 10 hours and 24 hours post-chemotherapy. Although we had expected to find patterns specific to SIGs (i.e., baseline protein expression levels) and treatment (i.e., ADE vs. ADEB), changes were limited to DNA damage, cell cycle regulation, protein translation and biogenesis and histone modification were seen, and were non-specific to subgroups of patients or to treatment. We think that the time point of measuring the expression is crucial. Although a significant percentage of the cells had probably died because of the chemotherapy (at 10 and 24 hours), these observed changes in protein profiles likely reflects the presence of many cells that were doomed to die (as these patients achieved remission, the blasts living at the time of collection must have died by the time of response recognition), but which were trying to repair DNA damage, or to control cell cycle progression to buy time for repair (NB. we only selected living cells). However, if we had collected samples at an earlier time point, this may had given the impression that the chemotherapy did not (yet) work or had no effect on protein expression and activation, as there was not enough time for the cells after initial chemotherapy exposure to alter their protein expression. Collecting samples at an even later would complicate sample collection, as most cells had already died, making it difficult to isolate sufficient numbers of living leukemic cells and to distinguish them from their normal, regenerating, counterparts.

Despite the ability of chemotherapy to kill the vast majority of leukemic cells, the rare leukemic stem cell that survives the chemotherapy (“survivor cells”), and that is responsible for the outgrowth of the leukemia cells which is manifested as relapse or primary resistant disease, is the cell from which we can potentially gather the most information. With current therapy, most patients with AML experience an approximately 2^+ log cell kill, with 82% achieving aplastic marrows on day 14.16 But, how can we identify the single cell that is going to survive from the one that will die in the next few days, and what is the right time to look at this residual leukemic cell or at the normal regenerating stem cell? A systematic experiment may be to take random selections of leukemia cells throughout the course of chemotherapy (e.g., time of diagnosis, day 1, 4, 8, 21, 30, etc.) and to perform single-cell analysis. We can take a final sample around day 30, when minimal/measurable residual disease (MRD) is measured at the end of induction I, and when the majority of patients has a blast count of < 0.1% (defined as MRD-). An additional sample can be taken after induction II or at time of relapse diagnosis. An elegant approach to isolate this particular (resistant) cell of interest, present with a very low frequency
in those collected samples, would be to use Probe-Seq.\textsuperscript{17} Probe-Seq uses a fluorescent in situ hybridization (FISH) method based upon a new probe design: Serial Amplification By Exchange Reaction (SABER). SABER-FISH uses 20–40 nucleotide oligonucleotides that are complementary to the RNA species of interest and are optimized for minimal off-target binding, and provides amplified fluorescent detection of RNA molecules. It can be used to mark cell populations based on combinatorial RNA expression profiles. Previously-identified markers can thus be targeted by gene-specific probe sets to isolate specific cell types by FACS. So, if we know that a leukemia has a specific translocation or mutated gene, we can isolate the rare, survivor cell that still has this aberration present. Those identified and isolated cells, could then be further analyzed using single-cell sequencing experiments or single-cell proteomics.

Comparison of protein profiles between cell populations (e.g., sensitive vs. resistant), would raise new biological questions about why cells behave differently, why, or how, cells are able to circumvent chemotherapy, and what treatment should be used to kill those cells. Which cells are dying first and which cells can still be detected at time of complete remission (i.e., MRD-)? Are we able to detect a consistent pattern across cells and patients? Which (treatment naïve) proteomic profiles are associated with relapse? Can we \textit{a priori} identify those, and thus predict who likely will relapse? Can we treat those patients differentially? Are there particular proteins that pop-up after relapse or after treatment? Are those proteins specific to certain induction or consolidation therapy? Are there existing drugs that target those proteins? Can we combine those drugs with current treatment strategies?

\textit{Studying leukemic stem cells (LSC) vs. leukemic blasts?}

Blood cells are organized as a coordinated hierarchy with hematopoietic stem cells (HSC) at the root. They have the capacity to differentiate into all types of blood cells, possess self-renewal and are mostly quiescence (they divide infrequently to protect the HSC from exhaustion and to prevent accumulation of mutations). Acute leukemia is thought to arise from the LSC, and as relapse or primary refractory arise from the chemoresistant LSC. In particular because LSC/HSC are quiescent, they are inherently resistant to most forms of chemotherapy. Therefore, studying the proteomics of this LSC may be another important approach to better understand the pathophysiology of acute leukemia, especially if compared to the normal HSC or the leukemic blast with the goal of developing therapies targeted to the LSC.
Inopportune, the cell markers that differentiate LSCs from other cells may differ between patients (hence, a functional definition would be more reliable), and LSCs only occur with a frequency of 1 in 10,000 to 1 to 5 million in AML. This makes isolation of enough cells for experimental analyses of the stem cell challenging, deterring most from studying these cells (for instance, MS requires around $1 \times 10^7$ cells, and RPPA requires approximately $3 \times 10^5$ cells, which would then require too many starting cells to enable analysis). Again, single-cell technologies may be a fitting approach to use. Several studies have previously shown that LSCs have different patterns of gene/protein expression and activation, compared to healthy HSCs or more differentiated progeny that define their unique biology.\textsuperscript{18,19} The use of LSCs will provide insight into the biological differences that distinguish LSCs from other populations, and could identify (proteomic) characteristics of the LSC that are therapeutically targetable. If the bulk blast population is studied, the identified features are less likely to accurately reflect the biology of the more clinically relevant LSC, and will by definition miss those changes found exclusively in the LSC. A study by Gentles et al. compared gene expression profiles derived from transcriptional data from LSC-enriched and leukemia-progenitor cells from the same AML patient samples. They identified an LSC-signature based on highly expressed genes. Among healthy samples, this signature was high in HSCs and in multipotent progenitor cells compared with more mature myeloid cell populations, suggesting that the LSC signature was shared with normal HSCs. Using four public data sets, they found that these LSC-enriched genes were predictive of primary refractoriness to chemotherapy, OS, EFS and relapse-free survival in cytogenetically normal patients.\textsuperscript{20} Eppert et al. identified an LSC signatures (based on comparison of gene expression between LSC and non-LSC fractions) that could predict OS and EFS independent of other known prognostic factors.\textsuperscript{21} However, the main caveat of these studies is the lack of a native microenvironment. Other questions to consider in review are; can we use this LSC information to monitor patients in remission? Can we target LSC without affecting the normal HSC?

\textit{Comparison of proteomics in acute leukemia across age}

The incidence of acute leukemia increases with age, but affects individuals of all ages. The cure fraction is substantially higher for pediatric ALL and AML compared to that seen in adults. This is partially due to the correlation between increasing rates of medical comorbidities and age, which renders many adults (especially the oldest) unable to tolerate standard chemotherapy regimens. As adolescents and young adults who lack medical comorbidities do also significantly worse compared to their younger counterparts, this suggests that there are
age-related differences in the intrinsic biology of “older” leukemic cells making them more resistant. Recently, the COG–National Cancer Institute (NCI) Therapeutically Applicable Research to Generate Effective Treatments (TARGET) AML initiative performed an extensive study in pediatric AML cases by whole-genome, targeted DNA, mRNA and microRNA sequencing and CpG methylation profiling (proteomics were not studied). Through comparison of AML molecular profiles across age groups (range 0-39 yrs.), they showed that significant differences in mutated genes, structural variants and DNA methylation patterns distinguish AML in infants, children, adolescents and adults. Although a similar study has not been performed in ALL, studies have shown that the distribution of ALL subtype varies across age. To our knowledge, no proteomic studies exist that assesses the proteomic landscape across age.

From the perspective of this thesis, a proteomic comparison between pediatric and adult acute leukemia patients, as well as between ALL and AML or even chronic leukemias, may aid to unravel the biological pathogenesis, and reveal similarities and dissimilarities that can propose therapeutics that target. Our group has RPPA data available from pediatric ALL \( (n = 266, n = 73) \) and AML \( (n = 500, n = 95) \), adult ALL \( (n = 92, n = 196, n = 130) \), AML \( (n = 205, n = 850) \), APL \( (n = 20) \), CLL \( (n = 876) \), and MDS \( (n = 194) \). Preliminary results comparing adult T-ALL to pediatric T-ALL showed no patterns exclusively specific to adults or pediatric patients, though several signatures were dominated by the younger patients. A similar analysis between pediatric and adult AML has not yet been done. However, based on the analysis in the 500 pediatric AML patients (Chapter 7) that found a certain association with infants (age ≤ 1 year) and CON-5 (SIG-7, 8, 9), we expect to find at least some differences in protein profiles. In addition, preliminary results that compared AML to ALL in pediatric patients and in adult patients, found that AML and ALL are proteomically as well as genetically different diseases. Three separate studies (pediatric AML \( (n = 95) \) vs. pediatric ALL \( (n = 73) \) (Chapter 6), adult AML \( (n = 241) \) vs. adult ALL \( (n = 130) \) (manuscript in preparation), pediatric AML \( (n = 500) \) vs. pediatric/ adult T-ALL \( (n = 358) \) (manuscript in preparation)) all showed that half of the CONs were shared between the two diseases, while the remaining CONs were either AML- or ALL-dominant. On the other hand, the SIGs, which are defined based on strong co-associations between the CONs, were almost all specific to a single acute leukemia subtype, this implies that while many pathwayutilizations are used in common, the overall system biology is distinct. For instance, in the 858 pediatric AML and pediatric/adult ALL samples, 14 CONs were found: two were T-ALL-dominant, five were AML-dominant, and the remaining seven
were shared between both diseases. In contrast, out of the 17 identified SIGs, only three SIGs had cases of both AML and T-ALL. From this data, we hope to learn if proteins allow for information on clinical utility from one disease to be applicable to the other? Hypothetically, treatment that works in one age-subgroup or disease, could also be effective other subgroups that share protein patterns. Are protein expression patterns prognostic for outcome? Are their protein profiles predictive of bortezomib sensitivity? Are those similar in T-ALL and AML?

**Part III: Clinical utility: guidance to therapy selection**

Although **Part II** already covered a bit of the clinical implications of the results presented in this thesis, in **Part III** we discussed the potential clinical implications of our studies in more depth.

In each of the four discussed studies, SIGs were significantly associated with several demographic and disease features. To simplify the usage of the prognostic information from the protein SIGs for clinical use, these were reduced to three groups: “favorable”, “intermediate”, “unfavorable”, similar to what is traditionally done with cytogenetics. In addition, when we performed survival analysis restricted to the AAML1031 low- or high-risk groups, SIGs had a more statistically significant prognostic impact. Identically, in our APL study, two SIGs were identified of which one contained all \( n = 4/4 \) relapse cases, suggesting that high-risk APL could be identified and stratified from low-risk APL patients based on their proteomic profiles. Again, this suggests that proteomics in combination with cytogenetics (APL is diagnosed based on cell morphology (i.e., FAB M3), or the molecular presence of the PML-RAR\( \alpha \) fusion protein \(((15;17) \text{ detected by PCR, conventional cytogenetics or an abnormal protein nuclear appearance (POD test)})\) can enhance risk stratification, by identifying high-risk patients in a considered low-risk group (i.e., APL is the most favorable subtype of AML). Each study suggests that an improved risk stratification schema would be created by the addition of the proteomic information to the features currently used for clinical risk assessment. Being able to better predict outcome based on features assessable at the time of presentation, or in an early stage of the disease, can led to more individualized treatment strategies. For instance, could we recognize which patients have a lower/higher risk of relapse and adjust their therapy accordingly, by giving less or more respectively, thereby minimizing late chemotherapy side effects by delivering less chemotherapy to those that do not need as
much, or decrease the relapse and death rate in others, through the judicious use of additional therapy (e.g., transplant in CR1)?

One of the other main goals in this thesis was to perform proteomics that could guide therapy selection. How do we know which drug to test, and in which combination? In Part I we have discussed that genetic chaos can be reduced to a finite number of SIGs, which allowed us to look at patients in groups, rather than as series of one. Can we find treatment that works in one protein expression profile, but not in another? Given that the majority of drugs that are currently tested in clinical trials target proteins, we hypothesized that proteins with an abnormal expression could potentially function as indicator for (combinational) targeted therapy across subgroups of patients (i.e., proteins with higher expression could be a rational target for inhibition, while proteins with lower expression could act as target for replacement or re-activation). This can result in rational selection of therapeutic agents to test, either as single agent or in combination, rather than “randomly” selecting drugs in the absence of clinical features previously identified as conferring sensitivity/resistance (e.g., estrogen receptor (ER) blockers in ER+ breast cancer). Even when a drug lacks efficacy as single agent, it may be of great benefit when used in combination with other chemotherapeutic agents. Hypothetical protein targets can be identified from either a CON (and can thus also be a target across other subgroups) or a SIG. Again, as we are aiming at proteins and not at genetic events, there is a limited number of patterns that tell us in which direction we should target.

We were able to identify proteins that were significantly changed in each individual SIGs relative to our normal CD34+ samples, as well as proteins that were statistically higher or lower expressed across all SIGs, which we defined as “universal”. In Chapter 7, we generated a list of these proteins in combination with a drug that is currently tested in a clinical trial or already FDA-approved, targeting the protein of interest. The observation that common driver events like MLL-rearrangement tended to fall in a given SIG with a frequency above expected (84% in SIG-7), hints that there are key proteins or pathways within the majority of the MLL-leukemias. If we can identify those, this can be an indicator of novel drugs that can be tested in combination with the current Menin-MLL inhibitors to improve response and to prevent resistance. As Menin-MLL inhibitors seem to be working in MLL-mutated AML, and as MLL has a characteristic SIG, this may also suggest that the non-MLL patients in this SIG would respond as well (if validated then is may be possible to use proteomics to define a MLL-like
SIG, similar to the existence of a Ph+-like patients in ALL, and thereby improve outcomes. Unfortunately, due to the Covid-19 pandemic, we were unable to test a couple of those protein targets in the laboratory, and to validate our hypothesis.

Experimental drug screen and RPPA
Performing large-scale ex vivo drug screens using primary patient samples that we studied on the RPPA could be a first useful step in the validation of our hypothesis. Precision medicine platforms could be performed to predict responses to potential therapies, and could ultimately determine which drugs or drug combinations could be most successful. This can work in two different ways: 1) start with rational selected drugs (based on the RPPA expression data in expression in treatment naïve cells) and test those individually as well as in combination with other agents, or 2) apply randomly selected drugs to the patient samples and see if there is a correlation between sensitivity and RPPA protein expression data. However, given the limited number of primary patient cells that is available, this will limit how many drugs can be tested, indicating that blindly testing (of all combinations) of drugs is not possible and that a rational selection, based on proteomic results, would be the best approach. A second drawback of ex vivo drug screens is that cells are not in their native environment, lacking the support from regulator molecules. Although this can partially be overcome by adding media components, cytokines/growth factor and stromal components, or by using PDX mice which have their own bone marrow microenvironment, it is still simplified version of reality.

In addition, screening the abundance and activation state for targets and pathways of interest in response to drug therapy, including off-target activity, can be performed on samples collected during the course of treatment, at the population, or single cell level. This can potentially profile the mechanism of action of the drug, and might identify pathways of drug resistance, relapse and sensitivity, and characterize mechanism of action of synergistic dose-ratio combinations. As an example, the Combinatorial Adaptive Response Therapy platform was developed to use RPPA to characterize and target adaptive responses to therapeutic stresses. They showed that rapid changes in tumors reflect adaptive resistance to targeted therapies, but also found therapeutically tractable vulnerabilities that could be overcome by selection rational combinations of drugs.
Bortezomib beneficial in subsets of patients

Recently, the COG performed a large phase 3 randomized clinical trial (AAML1031) that evaluated the effect of bortezomib in combination with standard ADE therapy (ADEB) vs. ADE therapy alone. When their interim analysis showed that 3-year OS and EFS did not improve with the addition of bortezomib they closed the trial early to further patient entry, and concluded that bortezomib should not be added to standard chemotherapy in the treatment of children with de novo AML. We hypothesized that while the bortezomib failed to improve outcome in pediatric AML overall, there are likely subgroups of patients that did benefit from the addition of bortezomib. If the frequency of patients that are susceptible to a drug is sufficiently low in the general pediatric AML population, then there is a high chance that the activity of a highly efficacious agent in subgroups can easily be missed if the correct patients did not make it onto the right arm of the trial, and if outcome in other subgroups is worsened (as patients cancel each other out). We observed that some SIGs and individual protein expression patterns did benefit from the addition of bortezomib. We found that HSF1-pSer$^{326}$ was prognostic in patients treated with ADEB. Patients with middle- or high-HSF-pSer$^{326}$ expression had a step-wise decrease in OS and EFS, and an increase in RR compared to those with low-HSF1-pSer$^{326}$. Multivariate analysis identified low-HSF1-pSer$^{326}$ as an independent favorable prognostic variable. When comparing outcome in patients treated with ADE vs. ADEB, patients with low-HSF1-pSer$^{326}$ significantly benefitted from additional bortezomib. Laboratory experiments using a HSF1 gene variants that mimicked phosphorylated and nonphosphorylated HSF1-pSer$^{326}$ showed that those with decreased phosphorylation were significantly more sensitive to bortezomib, but not to the chemotherapy agent etoposide (a component of the ADE regimen) that does not work though the proteasome.

A similar finding was observed with the AML MetaGalaxy analysis where we found three of the nine SIGs with improved outcome after the addition of bortezomib. Two of these SIGs, and one SIG in which a similar but less strong trend was observed, all shared an identical CON, that was not shared by any of the other SIGs. Also, while overall CEBPA-mutated patients did not benefit from bortezomib-addition, we identified that CEBPA-mutated patients with a particular protein profile did benefit. Preliminary results in T-ALL, identified one SIGs with an improve outcome after ADEB vs. ADE, and one SIG that did significantly worse after additional bortezomib. Currently we are analyzing how protein patterns associated with these two SIGs differ. We will correlate the protein data with mutational analysis and RNA-sequencing, as we wonder if protein differences between these two SIGs are already visible on
mRNA levels, as well as if and how mRNA levels are distinctive between both. Improved subgroup profiling will enable better patient selection for the addition of bortezomib to their chemotherapy. Future studies are needed to better understand why some patients are sensitive to bortezomib, while other are not or are even harmed.

In our T-ALL cohort, Hispanic patients did slightly, but not-significantly, better with bortezomib addition. In contrast, Hispanics in SIG-5 (which was enriched for Hispanics) did worse after treatment with bortezomib vs. standard ALL chemotherapy. This implies that if we can identify non-SIG-5 Hispanics (88% of the Hispanics in our study) prior to chemotherapy, and treat those with bortezomib, we may be able to significantly improve outcome in this historically more chemoresistant population (EFS at 4-year 66% vs. 90% for non-SIG-5 Hispanic patients, p = 0.05). However, more extensive studies are required for definitive conclusions.

**Development of protein assessment kits to classify acute leukemia**

Once you have identified what patient or leukemia characteristics make a patient susceptible to a particular drug, the next step is to find a means to identify those patients to facilitate therapy selection. If rapid enough, this could be used to select initial induction therapy, or if leukocytosis mandates urgent initiation of therapy, this could lead to the addition of appropriate agents later on during the induction chemotherapy cycle. Development of protein assessment kits that quickly evaluate protein levels of specific proteins could help. For instance, enzyme-linked immunosorbent assay, immunohistochemistry staining or another protein-based method that is quick and has a good sensitivity and specificity could be used to quantify protein expression and to rapidly identify patients with low- vs. high-HSF1-pSer^{326}, or could be used for a limited panel of proteins that enable classification into one of the SIGs. The development and validation of these kits is an area for future work by our group.

**Targeted clinical trials**

If the application of a protein classification kit is successful, this can identify bortezomib sensitive patient cohorts (or other drugs in the future) prior to therapy, and would be a rational set up for a new clinical trial that pre-selects bortezomib sensitive patients. For instance, if 30% of the 600 pediatric AML patients that are diagnosed annually in the United States would qualify for bortezomib treatment based on their low-HSF1-pSer^{326} levels, and if 5-year OS of these patients increases with 15%, then out of the 200 patients treated with bortezomib, 30
additional patients will be alive after 5 years. For the other patients that do not meet the criteria of low-HSF1-pSer\textsuperscript{326}, development of a HSF1-(phosphorylation) inhibitor should be continued. This inhibitor could be used in combination with bortezomib and in turn sensitize high-HSF1-pSer\textsuperscript{326} patients, or augment sensitivity of those 30% who are already sensitive. The same strategy can potentially be applied to the SIGs, where 34% of the AML patient fell in one of the three bortezomib sensitive SIGs.

**Future perspective**

Analysis of the proteome could enhance our knowledge of leukemia and its treatment in several ways. Currently, several prognostic factors are used in risk stratification for leukemia. However, these markers are imperfect, with many patient relapses coming from those in low and standard risk groups. In this thesis we have demonstrated that proteomics improves risk stratification by identifying risk groups within risk groups defined by traditional cytogenetic information. Incorporation of this into clinical practice, after validation, could allow patients with low-risk disease to be less intensively treated, whereas high-risk patients could be treated with more intensive treatment protocols. Also, development of protein assessment kits that quickly evaluate protein levels of specific prognostic proteins, could help guide personalized therapy by rational selecting the combinational agents, something that is already emerging based on mutations or, in other disease, based on proteomics. Future studies analyzing single cells may have even greater potential, as it also allows to study intra-leukemia heterogeneity.
Reference list


