Univerzita Karlova
Přírodovědecká fakulta

Studijní program: Speciální chemicko-biologické obory
Studijní obor: Molekulární biologie a biochemie organismů

PŘÍRODOVĚDECKÁ
FAKULTA
Univerzita Karlova

Johana Brodská

Single-cell RNA sekvenování v leukémii
Single-cell RNA sequencing in leukemia

Bakalářská práce

Školitelka: doc. MUDr. Eva Froňková, PhD.

Praha, 2022

Prohlášení:

Prohlašuji, že jsem závěrečnou práci zpracovala samostatně a že jsem uvedla všechny použité informační zdroje a literaturu. Tato práce ani její podstatná část nebyla předložena k získání jiného nebo stejného akademického titulu.

V Praze, 01. 08. 2022

Johana Brodská

Poděkování

Děkuji své školitelce Evě Froňkové a také Barboře Brodské za odbornou podporu, Janu Ječmenovi a Gretchen Ječmenové Fernholz za podporu technickou a jazykovou, Jakubu Houškovi za podporu administrativní a v neposlední řadě Ondřeji a Markétě Votrubovým a dalším svým blízkým za nezbytnou podporu duševní.

Abstrakt

Leukémie je nádorové onemocnění krvetvorby postihující celý organismus. V současné době je již dostupných mnoho variant léčby na všechny typy této nemoci, avšak stále není možné pacienta vždy plně uzdravit. Metoda single-cell RNA sekvenování poskytuje nový vhled do heterogenity nádorových i nenádorových buněk v prostředí leukémií. Tato práce si klade za cíl krátce představit tuto metodu a její historii a vyzdvihnout dosavadní poznatky o leukémiích získané její pomocí.

Klíčová slova

leukémie (AML, CML, ALL, CLL), sekvenování, scRNA-seq, buňky, transkriptom, léčba

Abstract

Leukemia is a cancer of hematopoietic cells affecting the whole organism. Currently, there are many treatment options for all disease types, but it is still not always possible to fully cure the patient. The single-cell RNA sequencing method offers a new insight into the heterogeneity of both cancerous and non-cancerous cells in the leukemic environment. This thesis aims to briefly present the method and its history and to highlight current findings about leukemia obtained with the help of it.

Keywords

leukemia (AML, CML, ALL, CLL), sequencing, scRNA-seq, cells, transcriptome, treatment

Table of Contents

I Introduction ... 1

II Sequencing in theory ... 1

Evolution of sequencing – from Sanger to NGS ... 1

Single-cell analysis ... 3

Technology ... 4

BD Rhapsody vs. 10x Genomics ... 6

Multiomics ... 8

III Leukemia ... 8

The four main types of leukemia ... 8

Acute myeloid leukemia (AML) ... 8

Chronic myeloid leukemia (CML) ... 9

Acute lymphocytic leukemia (ALL) ... 9

Chronic lymphocytic leukemia (CLL) ... 10

The use of scRNA-seq in leukemia ... 11

Prognosis, resistance and relapse ... 11

The immune microenvironment ... 13

Other studies ... 15

IV Conclusion ... 16

V Literature ... 17

I INTRODUCTION

Ever since its emergence in 2009,[1] single-cell RNA sequencing (scRNA-seq) has been a promising technology in many fields, and ever since its identification,[2] scientists have been on the hunt for leukemia treatments, each hopefully more effective than the last one. So, when this new tool presented itself, little time was wasted before it was used in leukemia research.

The method of scRNA-seq gives the unique possibility to analyze transcriptome of individual cells without bias, in high resolution and in large quantities over a short amount of time. It allows for cell subtypes to be more precisely identified and later possibly targeted in therapy. Now we can pose and potentially answer questions such as what cell types can be found in a certain sample, what can be observed about cell state switches, how gene-regulatory networks can be used etc.[3]

The scRNA-seq technology is particularly useful in situations where the amount of biological material is scarce, e. g. early embryonic cells[1,4] because, with the use of appropriate modifications, it can provide information about mutations, copy number variants, DNA protein binding, RNA splicing and protein expression in addition to mRNA expression.[5]

II SEQUENCING IN THEORY

Evolution of sequencing – from Sanger to NGS

The complex technology that is scRNA-seq quite obviously did not come from thin air. It is backed by over 70 years of efforts to decode the macromolecules in our cells. Following is a brief history of sequencing.

Interestingly, proteins and RNA were being sequenced before DNA was: Sanger sequenced the amino acid sequence of insulin in the early 1950s by fragmenting and overlapping.[6,7]

Similar approaches were used to sequence RNA in the next decade: fragmentation, separation, degradation and overlap analysis. The first RNA sequence deduced was that of alanine tRNA, 76 nucleotides long.[8]

The first successful attempts to sequence DNA came in the late 1960 – 12 nucleotides of bacteriophage lambda genome by Wu in 1968.[9] In 1973, after two years' worth of work, Maxam and Gilbert sequenced 24 bases of lactose repressor with the help of an RNA transcript.[10] Only three years later, two new methods came to life that allowed the sequencing of hundreds of bases in a matter of hours. These were Coulson&Sanger's chain terminator procedure[11] and Maxam&Gilbert's chemical cleavage procedure,[12] together known as the first-generation sequencing. The method of shotgun sequencing came about in 1979,[13] and by 1987 automated sequencing machines based on the Sanger method were able to generate around 1000 bases a day.[14] The amount of identified sequences was growing rapidly, plus every sequence would grow in value thanks to BLAST[15] and other such tools. From 1982 to 1986 the number of sequenced bases in GenBank went from just over 0.5 million to nearly 10 million. As of June 2022, the statistics declare over 1,4 trillion bases in GenBank (https://www.ncbi.nlm.nih.gov/genbank/statistics/).

More and more improvements were made to sequencing methods, until they allowed scientists to sequence the entire human genome, a project finished in 2004.[16]

From these improvements a new technique emerged that later took over the sequencing world: massively parallel or next-generation sequencing (NGS). It used in vitro amplification (PCR, rolling circle or bridge amplification) rather than bacterial cloning and for the sequencing itself, a combination of biochemical approaches and imaging replaced measuring fragment lengths.[17] There were three main strategies, the best known probably being pyrosequencing.[18]

The need for PCR amplification in sequencing was (and is) not without problems such as errors in copying, loss of information, different biases and more. This gave rise to efforts to develop real-time single-molecule sequencing. Two were successful.[17] The first technique is single molecule real-time sequencing (SMRT), which observes the synthesis mediated by a polymerase optically and in real-time.[19] The second approach is nanopore sequencing, which can be made very portable as it detects electronic signals rather than optical. Both methods have also been able to detect modifications such as methylation.[20]

Today, next-generation sequencing is widely used in many clinical applications, including diagnostics of rare diseases, non-invasive prenatal testing,[21] but mainly in cancer research,[22] allowing for precise determination of prognosis, treatment selection and monitoring.

With each new method, many commercial versions compete for researchers' favor. Currently, the most successful company in the world of sequencing is Illumina: it is nearing a monopoly.

Single-cell analysis

It has long been known that the sequence of DNA is not the only defining factor in the characterization of a cell. The transcriptomic profile plays a major role in cell identity, and analyzing it is crucial to answering questions such as how a cell is defined or what the difference is between a normal and malignant cell.[23]

The transcriptome is also complex just as it is important. In humans, 80% of the genome is transcribed, only 2% is protein-coding, but even the transcripts which do not code proteins are important: apart from tRNA and rRNA, there are long non-coding RNAs, microRNAs, small interfering RNAs, and other forms that all play significant roles in the regulation of gene expression. Then, there is added complexity in the existence of isoforms, etc.[24] It is not yet possible to sequence single cell RNA directly —it first needs to be converted to cDNA and then amplified. Both of these steps may cause some bias and loss of information. Furthermore, most transcripts are only present in a few copies in one cell, and even cells in the same microenvironment may have different transcription levels.[25]

Micromanipulation was the first technique used to isolate single cells, and it is still utilized when the number of cells is limited or when the cells are fragile.[26]

While it is possible to observe every cell under a microscope, this technique is very time- and labor-intensive. Fluorescence-activated cell sorting (or FACS) has a much higher throughput and allows sorting based on different features such as size or surface markers. Before the development of the current devices for cell processing, the cells were usually sorted into 96- or 384-cell well plates.[26]

The most recent addition to the field is microfluidics devices. They allow the automated capture of single cells. However, this technique performs poorly on primary and rare cells, e. g. circulating tumor cells. Another disadvantage is the price of these special devices.[26]

More context can be obtained from combination methods, such as Genome and Transcriptome Sequencing (or DR-seq).[27,28] Also, methods observing the methylome of single cells have been developed, and some researchers have already combined all three methods in so-called triple omics.[29,30] To go even further, the spatial relationship of the transcriptome can be explored, but that is a topic outside the scope of this thesis.[31]

The first scRNA-seq method was presented in 2009.[1] More details on it are presented in the following chapter. The field is rapidly developing, but limitations still remain: relying on reverse transcription prior to sequencing, the necessity of amplification, end biases, isolating single cells from samples and effective analysis of large amounts of data. Long-read sequencing technologies exist, but they still need to convert RNA to cDNA.[26]

Technology

To sequence mRNA from a single cell, we need to be able to first capture the single cell and then amplify the minute amount of mRNA present. The process is usually as follows: first, the cell is captured and lysed, then its mRNA is reverse transcribed to cDNA and finally the cDNA is amplified to prepare the sequencing library.[3]

The first scRNA-seq method was presented in 2009 and was named the Tang method.[1,4] The protocol was based on that for single cell microarray analysis. The cells were manually sorted under a microscope, lysed and the RNA was converted into cDNA using oligo-dT primers with a specific anchor sequence (UP1). The second strand was created by polyadenylation of the first cDNA strand which was then transcribed using an oligo-dT, this time with UP2 (specific anchor sequence at 5' end). UP1 and UP2 allowed amplification by PCR. Unfortunately, this method had a strong 3'-end bias, and it did not allow for studying transcriptional start sites or analyzing splice variants. The PCR amplification also meant the loss of strand specificity.[1,4]

Since then, many new systems were developed and presented under different names (Smart-seq, CEL-seq, MARS-seq, InDrop, DropSeq, etc).[26] Today, two main systems remain on the market: BD Rhapsody (from Becton Dickinson) and Chromium X (from 10x Genomics), with Chromium being increasingly more dominant.

In general, the protocol remained unchanged since Tang’s method, the steps are only being optimized. The following Figure 1 shows an example of a present-day scRNA-seq experiment.

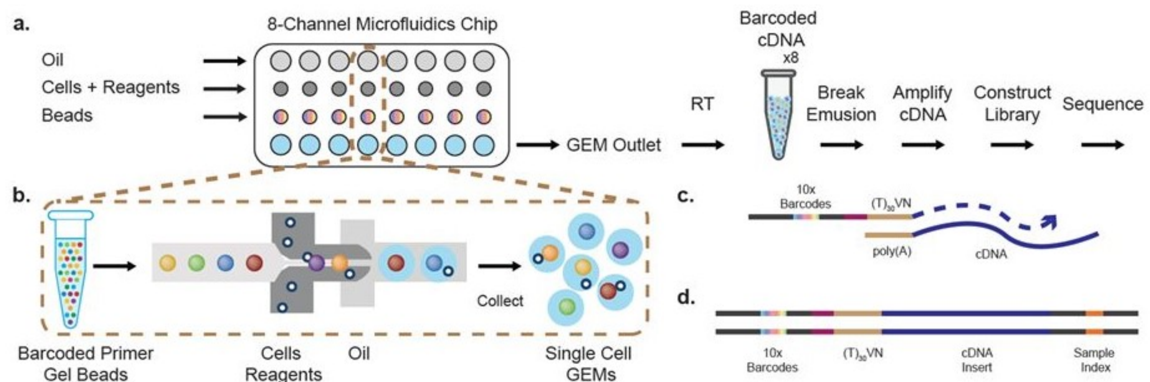

Figure 1: Experiment workflow. a) An overview of the whole process. GEM=Gel Bead in Emulsion: an oil droplet containing a cell attached to a barcoded gel bead. RT= reverse transcription. b) A detail of the process inside a microfluidics chip. c) Amplification of polyadenylated cDNA. d) The structure of a complete molecule: double-stranded barcoded cDNA. VN=a primer, 10x Barcodes=DNA cell tag, sample index=identification for the sequencing machine. Source: https://dnatech.genomecenter.ucdavis.edu/.

All cells are inspected under a microscope to assess the cell viability. FACS can help to discern between live and dead cells and different cell sizes. Oligo-dT primer is used to remove tRNA and rRNA as it binds polyA tails. However, this means that only mRNA is captured and lncRNA, miRNA and other non-polyadenylated RNAs are also removed from the research.[26]

After the RNA is captured and amplified, a library for sequencing is prepared. Library preparation can significantly influence the cost of an experiment. Currently, the sequencing libraries are prepared to be compatible with the Illumina NGS method. Nanopore sequencing is an ultrafast and cheap alternative method for sequencing both DNA and RNA, though both can only be single-stranded.[32] In theory, both RNA and DNA can be nanopore sequenced, but thus far, there has been more of a focus on DNA. This method could solve many of the current limitations of library preparation for RNA-seq: estimating of the amount of unamplified mRNA, revealing splice isoforms or post-transcriptional modifications, exploring the entire transcriptome as opposed to just polyA RNA etc. Sequencing long RNA molecules should also be possible.[33]

Quality is controlled with various tools throughout the experiment. The final control involves a comparison with bulk-level control data, although caution is needed to avoid

removing rare cells.[3] A brief comparison of bulk and single cell RNA sequencing is in the following Figure 2.

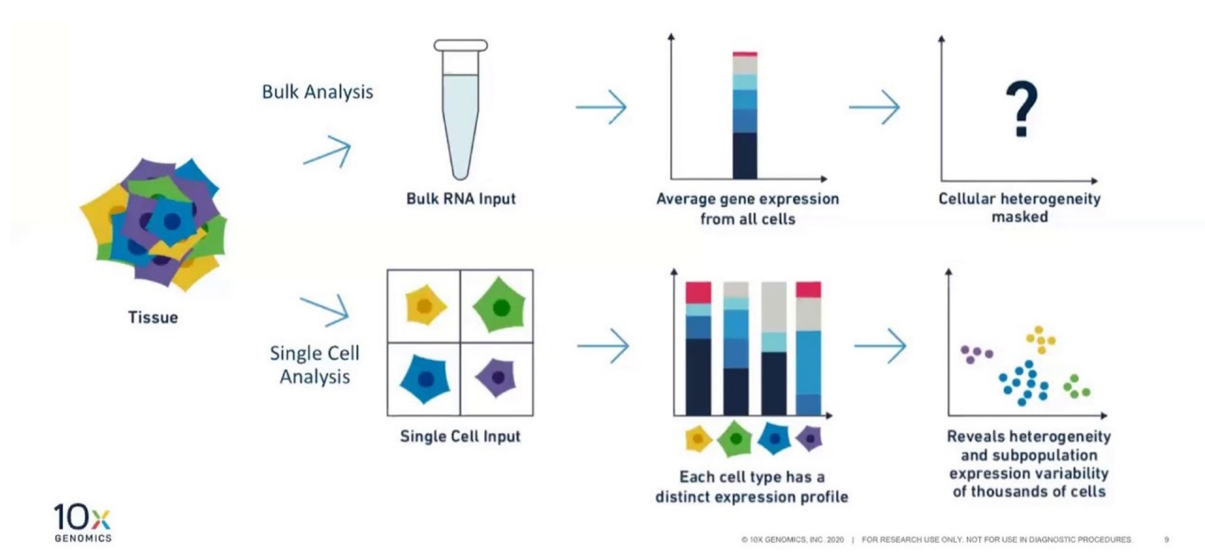

Figure 2: the advantages of single cell RNA sequencing in comparison to bulk RNA sequencing. Source: https://www.10xgenomics.com.

To improve the quantitative analysis of scRNA-seq data, we can use unique molecular identifiers. These are random sequences attached to each mRNA molecule and they allow to identify sequenced reads coming from the same molecule, amplified by PCR.[34]

BD Rhapsody vs. 10x Genomics

As mentioned above, the two players on the market are 10x Genomics with Chromium platforms and BD with Rhapsody instrument. Their strategies are quite similar in the basics. Both work in nanoliter-volume compartments where all cDNA is marked with a common barcode attached to a bead.

The beads are coated with oligonucleotides composed of a universal PCR primer, a cell label, a unique molecular index and an oligodT sequence. Therefore, each molecule has a cell label shared with other molecules from one cell and a label unique for itself.

Both platforms are able to simultaneously analyze RNA and surface proteins with the help of special oligonucleotide-labelled antibodies.

The major difference between the two methods is how they combine the mRNAs and the labelled beads (Fig. 3a, b). 10x Genomics uses GEMs: Gel Bead-In-Emulsion. The technology mixes the beads and cells in a microfluidic device and then encloses both in an oil drop.

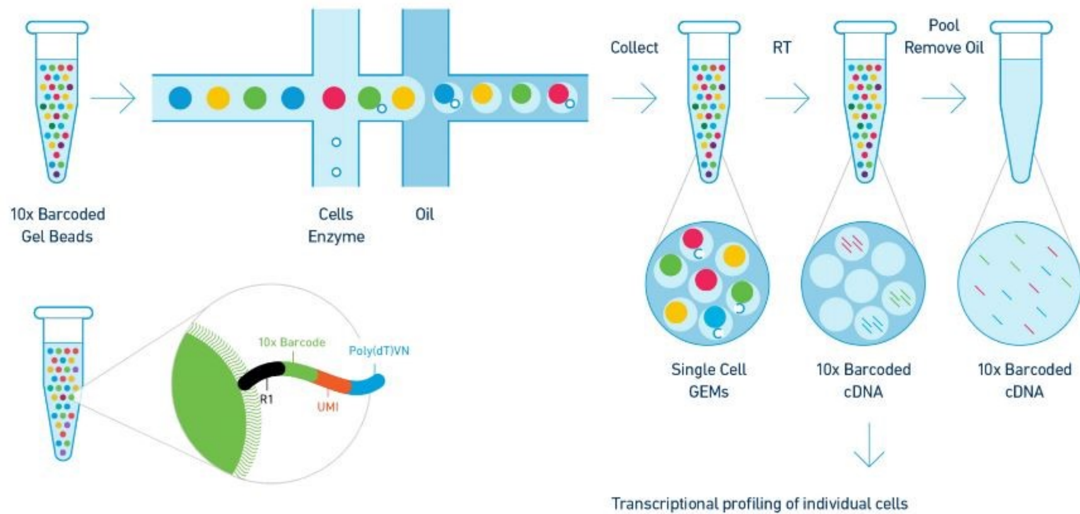

Figure 3a: 10x Genomics Chromium system. Source: https://www.10xgenomics.com.

BD is microwell-based: first, cells get randomly distributed into wells, then the beads (which are magnetic in this case) are loaded on top. Cells are lysed and mRNA gets hybridized onto the bead in its well. In both systems, cells are only loaded in amounts that allow avoiding multiplets. This results in only about 10% of the beads being used.

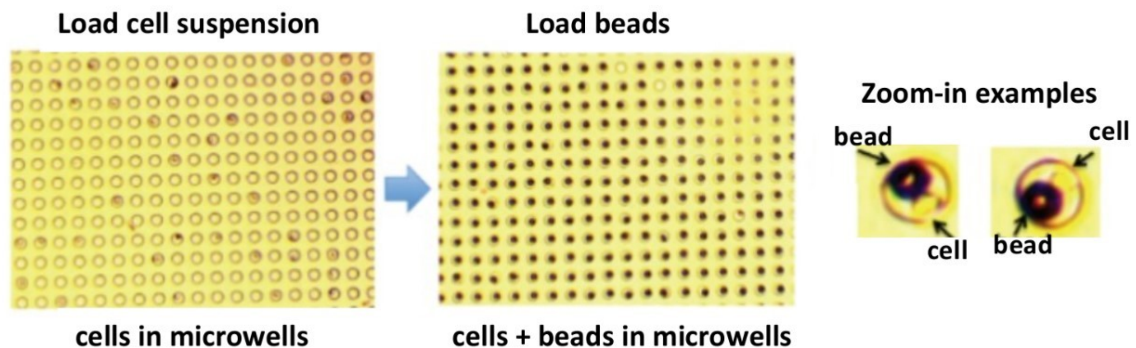

Figure 3b: BD Rhapsody system [26]

Another difference is that Chromium uses the template-switch oligo (TSO) system to obtain full-length cDNA, while BD only offers fragment cDNA with random primers. However, the full reads from Chromium are later sheared anyway.

The price of these two techniques is not very different. The Chromium system requires higher cell viability.

Multiomics

It is now possible to quantify not only mRNA with scRNA-seq but also proteins. An example of this method is AbSeq technology from BD Rhapsody, which uses antibodies conjugated with a DNA oligonucleotide barcode. A similar approach is used in a technology from 10x Genomics under the name CiteSeq (Figure 4).[35]

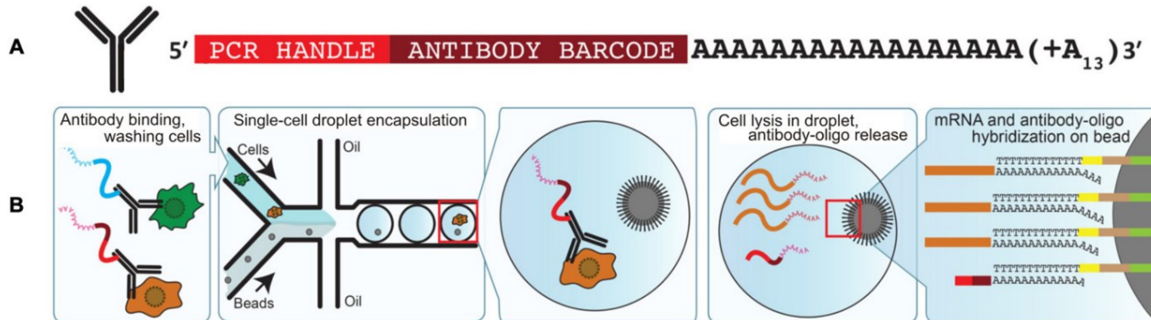

Figure 4: Multiomics scheme for CiteSeq. Source: https://www.10xgenomics.com.

III LEUKEMIA

The four main types of leukemia

Leukemia is a cancer of the blood, or more precisely of hematopoietic cells. While the malignant transformation occurs in a line of blood cells, the entire body is affected, especially the immune system, since the blood cells mediate the immune response. There are many types of malignant diseases originating in hematopoietic cells, but in this thesis, I only focused on the four main types of leukemia which are further described below.

Acute myeloid leukemia (AML)

AML is a disease of the myeloid lineage of blood cells. It occurs mostly due to genetic alterations in the cell precursors which result in overproduction of clonal progenitor cells.[36] The initial symptoms vary greatly in different patients and include infections, fatigue, fevers, paleness, etc. The most important step in the diagnosis is bone marrow examination.[37]

The current treatment protocols usually define three risk profile categories: favorable, intermediate, and adverse with the overall survival rates at 5 years being 64%, 41% and 11% respectively.[38]

A so-called “7+3” induction treatment has been the most common choice for the past 50 years. 80% of favorable-risk patients and 50-60% of those with intermediate risk respond completely to it, but it only works in 40% of adverse risk patients.[39] With almost half of adverse risk patients relapsing even after hematopoietic stem cell transplantation, the need is dire for further research and improvement of therapy options.[37]

After induction treatment, there are two main choices for post-remission therapy: cytotoxic chemotherapy or allogenic hematopoietic stem cell transplantation. The latter is a risky choice with an increased risk of morbidity and non-relapse mortality, but it decreases the risk of relapse.[37] Since many patients are elderly and often have comorbidities, these intensive types of treatment may not be recommended, and a non-curative option to alleviate symptoms is chosen.

Chronic myeloid leukemia (CML)

CML is characterized by the presence of Philadelphia chromosome: a balanced translocation between ch9 and ch22.[40] The result is a fusion gene called *BCR-ABL1*, and its protein (tyrosine kinase Bcr-Abl1) stimulates uncontrollable cell division.

There are three stages of the disease: the first, chronic phase with easily treatable symptoms, which, if untreated or resistant to treatment, can progress to acceleration phase (second stage) and later into a blast crisis (third stage). The last stage presents as AML in ½ cases, ALL in ¼ cases and combined or undefined phenotype in the rest.[41]

The diagnosis can be made after the occurrence of certain symptoms (with fatigue, night sweats, weight loss and splenomegaly among the common ones) or during a routine screening with an increased number of white cells.[42]

Today, CML in the chronic phase is routinely treated with tyrosine kinase inhibitors (TKIs), which need to be taken indefinitely, but allow most patients to reach a healthy-like life expectancy.[41]

Acute lymphocytic leukemia (ALL)

ALL stems from mutations affecting differentiation and proliferation of lymphoid cell precursors, 75% of cases in adults developing in the B-cell lineage. As well as with AML,

the basic therapy has not changed much despite advances in research: it consists of multi-agent chemotherapy and allogenic stem cell transplantation in eligible patients.[43]

According to data from the American Cancer Society, 80% of cases develop in children and a second peak of cases appears in patients around 50 years of age. Treatment is usually very effective in children, but much less so in adults. Patients over 60 in particular usually have a very poor prognosis.[44]

Some genetic syndromes have been identified as predispositions for some ALL cases (e. g. Down's syndrome[45]) as well as other factors such as HIV,[46] but mostly the disease is a consequence of a de novo malignancy. The presence of the Philadelphia chromosome t(9;22) has the greatest impact on prognosis from the cytogenetic aberrations.[43]

Studies have shown that pediatric-inspired treatment works best for patients younger than 45. Over that age, the occurrence of chemotherapy-related events was too high.[47]

For a long time, the only treatment offering long-term survival in patients with poor response to standard chemotherapy was Allo-SCT. Tyrosine kinase inhibitors were a turning point in treating Ph-positive ALL,[48] and some ten years ago, monoclonal antibody treatments gave rise to new hope [49] and dramatically improved prognosis in patients with poor response to standard treatment. The latest addition, CAR-T therapy (using genetically modified T-cells recognizing surface antigens specific for ALL), is also very successful.[50]

Allo-SCT is still recommended to all high-risk children and young adults with an available donor.

As the molecular and genetic alterations are very diverse, a single agent will not be effective for all patients. Therefore, the best answer is probably personalized and targeted therapy and careful monitoring of minimal residual disease (MRD) response after each block of therapy.[43]

Chronic lymphocytic leukemia (CLL)

CLL is the most common adult leukemia in the western world.[51] The biggest risk factor is a family history of hematological diseases.[52] The prevalence also increases with age, and the vast majority of patients are over 65 years old when diagnosed. Men are affected more often than women. More specifically, the data holds true for the USA, taken from SEER Cancer Statistics Explorer Network, https://seer.cancer.gov.

There is no aberration specific to CLL, but different aberrations can be detected in 80% of patients.[53] The most common one and also that associated with a favorable prognosis (if isolated) is deletion in the long arm of chromosome 13.[54]

CLL is associated with reduced immune function and therefore higher susceptibility to infections. This holds true even after effective treatment.[55] Since the situation worsens with advanced disease, infections represent the main cause of death. Autoimmune manifestations can also be present.[56] Moreover, patients with CLL are at a higher risk of developing a second cancer.[57]

Early treatment is rarely performed as it can in fact be detrimental and induce relapses. Rather, long-term follow up is advised.[51]

The use of scRNA-seq in leukemia

Leukemia and other hematological diseases are highly heterogenic and have diverse populations of cells. ScRNA-seq is a good tool for better understanding the processes involved in the development and pathogenesis of the diseases and for improving diagnostic and therapeutic approaches. Additional information can be obtained by analyzing proteins along with scRNA-seq,[58] as described above.

Most studies today only provide a “proof of concept”, meaning that the results support previous theories. The most researched leukemia type seems to be AML, probably due to its heterogeneity and possible therapeutic implications.

(Note: the categories in this chapter are not in any way official and were created just for the purpose of arranging this thesis.)

Prognosis, resistance and relapse

Despite great advancements already made in researching treatments, many patients are still resistant to therapy or relapse after some time. The scRNA-seq has been very helpful in discovering new targets for therapy or indicators for possible resistance. Following are some examples of such discoveries.

Mixed lineage leukemia-rearranged infant ALL (MLL-r iALL) is a disease with simultaneous expression of both lymphoid and myeloid of surface markers with a very poor prognosis: survival rates are 30-40%. One group sampled for the study consisted of patients that did

not relapse within 7 years and the second of those who relapsed within 2 years after diagnosis.[59] First, cells from bone marrow biopsies were sequenced. The result clustered cells mostly according to individual patients. This supports the knowledge that it is a very heterogeneous disease that will likely require personalized approach.[59]

As prednisone is an important part of the usual ALL treatment, the one-week response to it has been a major factor in risk assessment and relapse prediction. Gene signatures have been identified that point to resistance or sensitivity to prednisone (which is a continuum rather than a binary division). Seventy-eight genes that were upregulated and 370 genes that were downregulated formed two modules used to classify cells as either sensitive or resistant. Using this classification system on single-cell results from individual patients, it was found that non-relapsing patients had a higher number of sensitive-predicted cells, and the opposite was true for relapsing patients. Further experiments on peripheral blood (PB) samples offered results consistent with previous findings.[59]

Applying the result to attempt relapse prediction in the same patient cohort, the authors found a strong correlation between relapse occurrence and a higher ratio of resistant cells being present among PB blasts.

Gene Ontology enrichment showed higher metabolic activity in cells with higher predicted sensitivity, and that the opposite is also true: resistant-predicted cells are more quiescent or dormant.

The study was focused on response to prednisone, but it appears that the findings are likely to apply to chemotherapy in general.[59]

Zhao et al., who studied CLL, found three genes that had a fluctuating level of expression dependent on the state of the disease: increasing in accelerated phase, subsiding after therapy but increasing again at relapse.[60]

For the next study, also on CLL, samples were obtained from a patient with CLL at diagnosis and later at relapse that occurred after standard treatment. Results showed a sub-clone of cells that was already present at diagnosis, even though in very low numbers, and was significantly enriched in the sample taken after the relapse. This sub-clone lost the expression of CD19 and showed other molecular changes, such as 17p and 3p deletion and altered regulation of some apoptosis-related genes. After this change, the cells from the sub-clone showed many plasma cell-like features.[61]

As the aforementioned changes are not of common occurrence in CLL, the findings may help with predicting a possible relapse. Nonetheless, the presence of the observed clone is not the only player in a potential relapse, the role of epigenetic factors is also nonnegligible.[61]

In AML, Jiang et al. identified three (intrapatient) populations of cells, one of which was chemotherapy resistant. This population was found in higher proportions in patients with poor therapy outcomes, thus rendering the results clinically significant and offering a tool for better predicting potential relapses. The study results also contradict previous findings that state that chemotherapy does not induce mutations and the evolution of new clones but rather selects already existing resistant clones.[62] In this case, results suggest that, in fact, both scenarios are possible.[63]

It has also been shown that the risk of hematopoietic stem cells (HSCs) being malignantly transformed increases with age. Adelman et al. compared enriched HSC from young and older donors and found that there were new genetic changes occurring rather than preexisting subpopulation expanding,[64] adding another voice to the debate between "new mutations occurring" and "preexisting clones expanding" already mentioned above.

The immune microenvironment

Leukemia is not an isolated disease and therefore, it also affects non-malignant cells, especially those with a role in the immune system. The scRNA-seq technology finds its use in this field as well.

A study by Guo et al. focused on the microenvironment of AML which is crucial in understanding the disease better.[65] Specifically, it explored the diversity of patients' macrophages and dendritic cells and identified some unique immunity-related cell types.

The AML-induced processes and interactions in the bone marrow microenvironment led to possible evasion of immune surveillance and exhaustion of T-cells. The results are summarized in the following Table 1.[65]

Cluster	Subset	Representative prognostic genes	Prognosis
Mast cell	–	Mast cell-gene signature (215 genes)	Good
DC	**CD206^{+} DC subset**	CD206^{+} DC-gene signature (111 genes), *MRC1*, *TNFSF8*	Poor
	–	*CX3CR1*, *TGFB1*, *CLEC7A*, *ITGAX*, *ITGB2*	Poor
Mono/Mac	**MARCO^{high} subset**	MARCO^{high} subset-gene signature (203 genes)	Poor
	–	*CCL22*, *CD163*, *ITGAM*, *CCL5*	Poor
	–	*MMP9*	Good
T	**Treg**	Treg subset-gene signature (173 genes)	Poor
	Dysfunctional/exhausted T	Dysfunctional/exhausted-gene set (*LAG3*, *TIGIT*, *CTLA4*, *HAVCR2*, *TOX*, *PDCD1*, *CD274*, *PDCD1LG2*)	Poor
	–	*CD274*, *PDCD1LG2*, *PDCD1*, *BATF*	Poor

Table 1: Summary of the major cell subsets found in AML microenvironment and their impact on prognosis.[65]

A study by van Galen et al. had a similar focus: combining scRNA-seq and genotyping to elucidate the ecosystem of AML. The authors were able to put six malignant cell types along on the continuum from HSC to differentiated myeloid cells. Depending on the genotype of the disease, the subclones were present in vastly different amounts. The following investigation of the opposite ends of the differentiation spectrum revealed dysregulation of transcription in stem-like AML cells and the ability to affect immunity in monocyte-like AML.[66]

Adding to findings in AML environment, A. Crinier et al. first identified three populations of natural killer (NK) cells in healthy bone marrow: NK1, NK2 and their precursors NK0. Then, however, no distinct clusters were found in AML patients, as the samples had different transcriptomic profiles. Patients' cells showed altered expression levels of genes related to cytotoxicity, activating effector receptors (lower than normal), inhibition and transcription factors for maturation and survival (higher than normal).

Apart from the above, the researchers also observed a correlation between higher levels of CD160 (an activating molecule) expression in NK cells and better survival. [67]

Overall, it is clear that AML substantially alters NK cell function, hence weakening the immune system.

Focusing on patients with CML in treatment-free remission, G. Yu et al. identified a subset of NK cells that did not have a counterpart in healthy donors: platelet-NK aggregation cells. Other than that, no difference was found in differentiation or classification of NK cells between patients and healthy controls.

Since only a small number of samples was used, the results need confirmation in future studies, but it seems that discontinuation a successful TKI treatment can mostly restore the functionality of NK cells[68], which are then effective at eliminating residual acute myeloid cells.[69]

In ALL, 11 T cell clusters in samples from healthy donors have been discovered that could also be found in B-ALL patients and generally showed higher activation levels, and then two more subsets that were only present in patient samples were identified. These were characterized as exhausted and did not have counterparts in healthy donors.

Further findings from this study point to biomarkers that could prove useful for diagnosis and immunotherapy and show that there are a few highly dominant clones in each patient – this suggests high levels of clonal expansion.

The results may prove helpful in advancing CAR-T based therapy by providing insight into the variety of T-cells, for example by preventing preparing therapeutics from dysfunctional T cells.[70]

Other studies

Below are some more studies, whose results provide information outside the two previous categories.

By comparing samples from a healthy individual and an AML patient, it has been confirmed that alternative polyadenylation (=APA, a regulatory tool for gene expression, translation efficiency and mRNA stability) is involved in leukemia development. APA was found to be abundant in pathways associated with AML development.[71]

Analyzing 87 genes in purified CD34+ cells from three samples from patients with CML (256 cells total), the researchers found a group of 12 genes, some related to pluripotency and embryogenesis, that were closely connected and expressed in 22 cells from the 256. Using pseudotime analysis, these were put onto a “continuum of transitional states”. Cells were grouped into 4 clusters and 7 stem cell states were identified in these clusters.[72]

De Bie et al. observed the chronological order of mutations appearing during T-ALL.[73]

One of the latest studies has explored leukemic cell populations in pediatric B-ALL at diagnosis, in residual cells after therapy and in relapse. The researchers paired scRNA-seq with single cell B-cell-receptor sequencing and discovered that at diagnosis, three out of the four patients examined were lacking cells with BCR expression, which points to the possibility of altered V(D)J rearrangement. They also noted that relapsed patients' leukemic cells were more poorly differentiated than those from recently diagnosed patients. As for the cell profile at the stage of minimal residual disease, these were in very heterogeneous states, but the upregulation of the hypoxia pathway genes has been identified as a common denominator and as such a possible treatment target. [74]

IV CONCLUSION

From one of the first studies, that "only" identified 11 presumed blasts from 20 sample cells and pointed out the problems to overcome, [75] the research has come a long way. In this thesis, I have mostly focused on those that provide a general insight into each of the four main leukemia subtypes to give an overview of what is there to be found. There is a number of studies with very specific goals and results, for example studying resistance to a specific drug or putting a rare genotype in the spotlight.[76–78]

Since scRNA-seq is still not a method affordable for routine practice, it cannot be used as a diagnostic or monitoring tool nearly often enough. This would be especially helpful in personalizing therapy for patients with the always-different AML. Despite that, the combined effort of researchers has already provided new possible drug targets and a robust base for further explorations of the role of immunity and the microenvironment of leukemia. Furthermore, with the advent of multiomics and spatial transcriptomics, even more discoveries are to be expected, for example in the realm of epigenetic regulation.

In short, it is irrefutable that scRNA-seq is a tool with a long and successful future ahead.

V LITERATURE

* značí sekundární citaci

1. Tang, F. *et al.* mRNA-Seq whole-transcriptome analysis of a single cell. *Nature Methods* **6**, 377–382 (2009).
2. Kampen, K. R. The discovery and early understanding of leukemia. *Leukemia Research* **36**, 6–13 (2012).

*3. Kolodziejczyk, A. A., Kim, J. K., Svensson, V., Marioni, J. C. & Teichmann, S. A. The Technology and Biology of Single-Cell RNA Sequencing. *Molecular Cell* **58**, 610–620 (2015).

4. Tang, F. *et al.* RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature Protocols* **5**, 516–535 (2010).

*5. Gao, C., Zhang, M. & Chen, L. The Comparison of Two Single-cell Sequencing Platforms: BD Rhapsody and 10x Genomics Chromium. *Current Genomics* **21**, 602–609 (2020).

6. Sanger, F. & Tuppy, H. The Amino-acid Sequence in the Phenylalanyl Chain of Insulin 1. *Biochemical J* **49**, 463–81 (1951).
7. SANGER, F. & TUPPY, H. The amino-acid sequence in the phenylalanyl chain of insulin. 2. The investigation of peptides from enzymic hydrolysates. *Biochem J* **49**, 481–490 (1951).
8. Zajonc, R. B. Structure of a Ribonucleic Acid. *Science (1979)* **149**, 269–274 (1965).
9. Wu, R. & Kaiser, A. D. Structure and base sequence in the cohesive ends of bacteriophage lambda DNA. *Journal of Molecular Biology* **35**, 523–537 (1968).
10. Gilbert, W. & Maxam, A. The nucleotide sequence of the lac operator. *Proc Natl Acad Sci U S A* **70**, 3581–3584 (1973).
11. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* **74**, 5463–5467 (1977).
12. Maxam, A. M. & Gilbert, W. A new method for sequencing DNA. *Proc Natl Acad Sci U S A* **74**, 560–564 (1977).
13. Road, H. A strategy of DNA sequencing employing computer programs. *Nucleic Acids Research* **6**, 2601–2610 (1979).
14. Smith, L. M. *et al.* Fluoresence detection in automated DNA sequence analysis. *Nature* **321**, 674–679 (1986).
15. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403–410 (1990).
16. Hattori, M. Finishing the euchromatic sequence of the human genome. *Tanpakushitsu Kakusan Koso* **50**, 162–168 (2005).

*17. Shendure, J. *et al.* DNA sequencing at 40: Past, present and future. *Nature* **550**, (2017).

18. Ronaghi, M., Karamohamed, S., Pettersson, B., Uhlén, M. & Nyrén, P. Real-time DNA sequencing using detection of pyrophosphate release. *Analytical Biochemistry* **242**, 84–89 (1996).

19. Mccarthy, A. Third generation DNA sequencing: Pacific biosciences' single molecule real time technology. *Chemistry and Biology* **17**, 675–676 (2010).
20. Flusberg, B. A. *et al.* Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nature Methods* **7**, 461–465 (2010).
21. Chiu, R. W. K. *et al.* Noninvasive prenatal diagnosis of fetal chromosomal aneuploidy by massively parallel genomic sequencing of DNA in maternal plasma. *Proc Natl Acad Sci U S A* **105**, 20458–20463 (2008).
22. Wood, L. D. *et al.* The genomic landscapes of human breast and colorectal cancers. *Science (1979)* **318**, 1108–1113 (2007).
23. Saliba, A. E., Westermann, A. J., Gorski, S. A. & Vogel, J. Single-cell RNA-seq: Advances and future challenges. *Nucleic Acids Research* **42**, 8845–8860 (2014).
24. Hangauer, M. J., Vaughn, I. W. & McManus, M. T. Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLoS Genetics* **9**, (2013).
25. Chang, H. H., Hemberg, M., Barahona, M., Ingber, D. E. & Huang, S. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* **453**, 544–547 (2008).
*26. Picelli, S. Single-cell RNA-sequencing: The future of genome biology is now. *RNA Biology* **14**, 637–650 (2017).
27. Macaulay, I. C. *et al.* Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nature Protocols* **11**, 2081–2103 (2016).
28. Dey, S. S., Kester, L., Spanjaard, B., Bienko, M. & van Oudenaarden, A. Integrated genome and transcriptome sequencing of the same cell. *Nature Biotechnology* **33**, 285–289 (2015).
29. Guo, H. *et al.* Single-Cell methylome landscapes of mouse embryonic stem cells and early embryos analyzed using reduced representation bisulfite sequencing. *Genome Research* **23**, 2126–2135 (2013).
30. Hou, Y. *et al.* Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Research* **26**, 304–319 (2016).
31. Ke, R. *et al.* In situ sequencing for RNA analysis in preserved tissue and cells. *Nature Methods* **10**, 857–860 (2013).
32. Kasianowicz, J. J., Brandin, E., Branton, D. & Deamer, D. W. Characterization of individual polynucleotide molecules using a membrane channel. *Proc Natl Acad Sci U S A* **93**, 13770–13773 (1996).
33. Cracknell, J. A., Japrung, D. & Bayley, H. Translocating kilobase RNA through the staphylococcal α-hemolysin nanopore. *Nano Letters* **13**, 2500–2505 (2013).
34. Fu, G. K., Hu, J., Wang, P. H. & Fodor, S. P. A. Counting individual DNA molecules by the stochastic attachment of diverse labels. *Proc Natl Acad Sci U S A* **108**, 9026–9031 (2011).
35. Stoeckius, M. *et al.* Cell Hashing with barcoded antibodies enables multiplexing and doublet detection for single cell genomics. *Genome Biology* **19**, 1–12 (2018).
36. Grimwade, D., Ivey, A. & Huntly, B. J. P. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. *Blood* **127**, 29–41 (2016).
*37. Pelcovits, A. & Niroula, R. Acute Myeloid Leukemia: A Review. *R I Med J (2013)* **103**, 38–40 (2020).

38. Grimwade, D. *et al.* The importance of diagnostic cytogenetics on outcome in AML: Analysis of 1,612 patients entered into the MRC AML 10 trial. *Blood* **92**, 2322–2333 (1998).
39. Fernandez, H. F. *et al.* Anthracycline Dose Intensification in Acute Myeloid Leukemia. *New England Journal of Medicine* **361**, 1249–1259 (2009).
40. Nurmi, E.; Rantala, M. A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nat. Phys. Sci.* **241**, 20 (1973).
41. Apperley, J. F. Chronic myeloid leukaemia. *The Lancet* **385**, 1447–1459 (2015).
42. Jabbour, E. & Kantarjian, H. Chronic myeloid leukemia: 2012 update on diagnosis, monitoring, and management. *American Journal of Hematology* **87**, 1037–1045 (2012).
*43. Terwilliger, T. & Abdul-Hay, M. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J* **7**, (2017).
44. Rowe, J. M. Prognostic factors in adult acute lymphoblastic leukaemia. *British Journal of Haematology* **150**, 389–405 (2010).
45. Chessells, J. M. *et al.* Down's syndrome and acute lymphoblastic leukaemia : clinical features and response to treatment. 321–325 (2001).
46. Gérinière, L. *et al.* Heterogeneity of acute lymphoblastic leukemia in HIV-seropositive patients. *Annals of Oncology* **5**, 437–440 (1994).
47. Huguet, F. *et al.* Pediatric-inspired therapy in adults with philadelphia chromosome-negative acute lymphoblastic leukemia: The GRAALL-2003 study. *Journal of Clinical Oncology* **27**, 911–918 (2009).
48. Thomas, D. A. *et al.* Treatment of Philadelphia chromosome-positive acute lymphocytic leukemia with hyper-CVAD and imatinib mesylate. *Blood* **103**, 4396–4407 (2004).
49. Nagorsen, D., Kufer, P., Baeuerle, P. A. & Bargou, R. Blinatumomab: A historical perspective. *Pharmacology and Therapeutics* **136**, 334–342 (2012).
50. Sterner, R. C. & Sterner, R. M. CAR-T cell therapy: current limitations and potential strategies. *Blood Cancer Journal* **11**, (2021).
*51. Scarfò, L., Ferreri, A. J. M. & Ghia, P. Chronic lymphocytic leukaemia. *Critical Reviews in Oncology/Hematology* **104**, 169–182 (2016).
52. Slager, S. L., Caporaso, N. E., de Sanjose, S. & Goldin, L. R. Genetic susceptibility to chronic lymphocytic leukemia. *Seminars in Hematology* **50**, 296–302 (2013).
53. Döhner, H. *et al.* Genomic Aberrations and Survival in Chronic Lymphocytic Leukemia. *New England Journal of Medicine* **343**, 1910–1916 (2000).
54. Cimmino, A. *et al.* miR-15 and miR-16 induce apoptosis by targeting BCL2. *Proc Natl Acad Sci U S A* **102**, 13944–13949 (2005).
55. Morrison, V. A. Infectious complications of chronic lymphocytic leukaemia: Pathogenesis, spectrum of infection, preventive approaches. *Best Practice and Research: Clinical Haematology* **23**, 145–153 (2010).
56. Hodgson, K., Ferrer, G., Pereira, A., Moreno, C. & Montserrat, E. Autoimmune cytopenia in chronic lymphocytic leukaemia: Diagnosis and treatment. *British Journal of Haematology* **154**, 14–22 (2011).
57. Benjamini, O. *et al.* Second cancers in patients with chronic lymphocytic leukemia who received frontline fludarabine, cyclophosphamide and rituximab therapy:

Distribution and clinical outcomes. *Leukemia and Lymphoma* **56**, 1643–1650 (2015).

*58. Zhu, Y., Huang, Y., Tan, Y., Zhao, W. & Tian, Q. Single-Cell RNA Sequencing in Hematological Diseases. *Proteomics* vol. 20 Preprint at https://doi.org/10.1002/pmic.201900228 (2020).

59. Candelli, T. *et al.* Identification and characterization of relapse-initiating cells in MLL-rearranged infant ALL by single-cell transcriptomics. *Leukemia* **36**, (2022).

60. Zhao, Z. *et al.* Evolution of multiple cell clones over a 29-year period of a CLL patient. *Nat Commun* **7**, 13765 (2016).

61. Ostasov, P. *et al.* Evolution of Advanced Chronic Lymphoid Leukemia Unveiled by Single-Cell Transcriptomics: A Case Report. *Frontiers in Oncology* **10**, (2020).

62. Shlush, L. I. *et al.* Tracing the origins of relapse in acute myeloid leukaemia to stem cells. *Nature* **547**, 104–108 (2017).

63. Jiang, L. *et al.* Multidimensional study of the heterogeneity of leukemia cells in t(8;21) acute myelogenous leukemia identifies the subtype with poor outcome. *Proc Natl Acad Sci U S A* **117**, 20117–20126 (2020).

64. Adelman, E. R. *et al.* Aging human hematopoietic stem cells manifest profound epigenetic reprogramming of enhancers that may predispose to leukemia. *Cancer Discovery* **9**, 1080–1101 (2019).

65. Guo, R. *et al.* Single-cell map of diverse immune phenotypes in the acute myeloid leukemia microenvironment. *Biomarker Research* **9**, 1–16 (2021).

66. van Galen, P. *et al.* Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell* **176**, 1265-1281.e24 (2019).

67. Crinier, A. *et al.* Single-cell profiling reveals the trajectories of natural killer cell differentiation in bone marrow and a stress signature induced by acute myeloid leukemia. *Cellular and Molecular Immunology* **18**, 1290–1304 (2021).

68. Yu, G. *et al.* Single-cell RNA sequencing to explore composition of peripheral blood NK cells in patients with chronic myeloid leukemia in treatment-free remission. *Leukemia and Lymphoma* **0**, 1–12 (2022).

69. Stojanovic, A., Correia, M. P. & Cerwenka, A. Shaping of NK cell responses by the tumor microenvironment. *Cancer Microenvironment* **6**, 135–146 (2013).

70. Wang, X. *et al.* Single-Cell RNA-Seq of T Cells in B-ALL Patients Reveals an Exhausted Subset with Remarkable Heterogeneity. *Advanced Science* **8**, 1–12 (2021).

71. Ye, C., Zhou, Q., Hong, Y. & Li, Q. Q. Role of alternative polyadenylation dynamics in acute myeloid leukaemia at single-cell resolution. *RNA Biology* **16**, 785–797 (2019).

72. Pagliaro, S. *et al.* Single-Cell Transcriptome in Chronic Myeloid Leukemia: Pseudotime Analysis Reveals Evidence of Embryonic and Transitional Stem Cell States. *Experimental Hematology* **85**, 47-56.e2 (2020).

73. de Bie, J. *et al.* Single-cell sequencing reveals the origin and the order of mutation acquisition in T-cell acute lymphoblastic leukemia. *Leukemia* **32**, 1358–1369 (2018).

74. Zhang, Y. *et al.* Elucidating minimal residual disease of paediatric B-cell acute lymphoblastic leukaemia by single-cell analysis. *Nature Cell Biology* **24**, 242–252 (2022).

75. Yan, B. *et al.* Single-cell genomic profiling of acute myeloid leukemia for clinical use: A pilot study. *Oncology Letters* **13**, 1625–1630 (2017).

76. Zhang, W. *et al.* Single cell sequencing reveals cell populations that predict primary resistance to imatinib in chronic myeloid leukemia. *Aging* **12**, 25337–25355 (2020).
77. Rendeiro, A. F. *et al.* Chromatin mapping and single-cell immune profiling define the temporal dynamics of ibrutinib response in CLL. *Nature Communications* **11**, 1–14 (2020).
78. Johnston, G. *et al.* Nascent transcript and single-cell RNA-seq analysis defines the mechanism of action of the LSD1 inhibitor INCB059872 in myeloid leukemia. *Gene* **752**, 144758 (2020).