D7.2

Software to define tumour subclones and association with therapy response

Project number	826121
Project acronym	iPC
Project title	individualizedPaediatricCure: Cloud-based virtual-patient models for precision paediatric oncology
Start date of the project	1st January, 2019
Duration	53 months
Programme	H2020-SC1-DTH-2018-1

Deliverable type	Report
Deliverable reference number	SC1-DTH-07-826121 / D7.2 / 1.0
Work package contributing to the deliverable	WP7
Due date	May, 2022 – M41
Actual submission date	1st June 2022

Responsible organisation	IBM Research
Editor	Marianna Rapsomaniki
Dissemination level	PU
Revision	1.0

Abstract	Flow cytometry is an important diagnostic tool in childhood acute lymphoblastic leukemias (ALL), flow cytometry data analysis is limited by multiple sources of bias and variation. We present a unified machine learning framework for automated analysis of a standardized diagnostic pediatric leukemia staining that can overcome these challenges. We applied our framework in a large cohort of ALL flow cytometry samples and demonstrated how it can robustly extract the frequencies of cell lineage populations with minimal expert intervention. This work provides a proof of concept that our method meets the needs of an automated analysis tool for diagnostic flow cytometry data.
Keywords	acute lymphoblastic leukaemia, machine learning, flow cytometry

The project iPC has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 826121.

Editor

Marianna Rapsomaniki (IBM)

Contributors (ordered according to beneficiary numbers)

Susanne Unger (UZH)

Andrej Lissat (Charite/UZH)

María Rodríguez Martínez (IBM)

Burkhard Becher (UZH)

Jean-Pierre Bourquin (UZH)

Disclaimer

Executive Summary

Flow cytometry is an important diagnostic tool in childhood acute lymphoblastic leukemias (ALL). Despite substantial efforts of staining protocol standardisation over the past decade, analysis of flow data from diagnostic work-ups still requires time-consuming manual gating performed by an expert. So far, staining irregularities and other technical sources of variation have prohibited the successful application of automated flow cytometry data analysis tools.

This report describes the development of a unified machine learning framework for automated analysis of a standardised diagnostic paediatric leukaemia staining. We demonstrate its ability to robustly extract the frequencies of cell lineage populations in a large cohort of bone marrow samples from B cell precursor-ALL patients analysed in the Children's Hospital Zurich by standard diagnostic flow cytometry procedure using two different staining panels. Our model is unaffected by batch effects caused by staining intensity variability, because we avoided data integration and created instead a baseline cell lineage classification model that can accurately detect cell populations of interest in unseen flow cytometry samples. This work provides a proof of concept that our method meets the needs of an automated analysis tool for diagnostic flow cytometry data. It helps to solve the analysis bottleneck by providing a fast, robust and accurate identification of cell populations of interest with little to no expert curation of results.

We are currently writing a manuscript to publish these results. In parallel, we are addressing the internal administrative procedures necessary to release open source code at IBM Research. The code and data supporting this study will be made publicly available upon publication of the article.

Table of Contents

List of Figures

List of Tables

Chapter 1 Introduction

Childhood Acute Lymphoblastic Leukaemia (ALL) is one of the most frequent malignant diseases in childhood and one of the major causes of childhood death [1]. Mutations in lymphoid development genes and in genes associated with apoptosis and proliferation result in malignant transformation of lymphoid precursor cells of B-cell and T-cell phenotype [2], [3]. Flow cytometry is a cornerstone in ALL diagnosis and subtyping, with established flow cytometry panels used in routine diagnostics [4], [5]. Diagnostic staining panels are designed to interrogate cell populations currently considered relevant for diagnosis [4]. Currently, thousands of flow cytometry files from both diagnostic and research studies from leukaemia patients are stored in hospital databases. However, despite being a rich source of information that exceeds the current diagnostic purposes, their use in research is limited. A major impediment is the challenging analysis of flow cytometry data produced in a diagnostic unit. Even though the processing, staining procedure and acquisition of samples is standardised in diagnostic units, staining variability is introduced by reasons which cannot be prevented in daily routine work, such as manual work by technicians, antibody lot-to-lot inconsistency, or change of instruments over time. Importantly, in a clinical setting new files are constantly generated from newly diagnosed patients and patient samples are processed immediately when they are available. This implies that a set of files which are generated on one day represent an individual batch. All these factors introduce pronounced systematic biases, observed as batch effects in the measurements, which greatly impact the reproducibility and robustness of results, and can potentially lead to misleading conclusions [6].

As a result, available flow cytometry data analysis methods cannot account for the needs of data acquired in diagnostic units during large clinical studies, which often include hundreds of patients and millions of high-dimensional single-cell profiles. Currently, identification of cell populations of interest in these patients is traditionally performed in each sample individually through manual gating. Manual gating of hundreds of files independently is not only time consuming, but also prone to the subjectivity of the technician, introducing more batch effects. To circumvent gating, many available data analysis approaches use a clustering method [7] to analyse the integrated dataset, generated by concatenating all individual .fcs files. This approach is not only prohibitive in terms of computational resources, but clustering approaches are extremely sensitive to batch effects. An appealing alternative is normalisation of files to correct for batch effects. However, flow cytometry normalisation methods (e.g. CytoNorm [8]) require a shared control across all batches, which is not feasible in a clinical setting. Furthermore, biologically relevant information can be lost, especially for malignant cells whose expression pattern varies due to biological and not technical reasons.

In this project we address the above challenges and create an automated immune cell classification method based on application of machine learning to treat the flow cytometry measurements. We apply our method to a large cohort of paediatric ALL patients. The main idea behind our method was to create an annotated reference ALL dataset, acquired by selecting and annotating few but highly representative ALL samples capturing ALL heterogeneity. These samples are then used to train a neural network classifier that can then identify cell lineages of interest in individual ALL samples in little time, eliminating the need for manual gating and greatly speeding up the process. Next, we provide detailed information on the available data, methods, and models, and show how our approach performs in unseen samples from the paediatric ALL cohort.

Chapter 2 Materials and Methods

2.1 Sample collection, processing, and flow cytometry

Fresh Li-Heparin- or EDTA-anticoagulated bone marrow aspiration samples were obtained as part of routine diagnostic workup when a haematological malignancy was suspected. Samples obtained in seven local hospitals in Switzerland were transferred to the diagnostic unit of the Children's Hospital Zurich for a centralised flow cytometric evaluation. If immediate processing was not possible, samples were stored at room temperature (Li-Heparin-anticoagulated for max 72 h, EDTA-anticoagulated for max. 12 h). Samples taken between November 2010 and July 2019 were included in the study. Samples were stained according to local protocols based on the European Group for the Immunological Characterization of Leukemias (EGIL) consensus guidelines [4]. In brief, cells were stained with surface antibodies in Horizon Brilliant Stain buffer (BD) for 15-30 min. Subsequently, samples were fixed with IntraSure KIT (BD) according to the manufacturer's protocol, except that the last fixation step in 1% PFA was omitted. Cells were directly resuspended in BD CellWash (BD) and analysed the same day. Until 10th of January 2017, the standardised EuroFlow 8-colour Acute Leukaemia Orientation Tube (ALOT) [9] panel was used and samples were acquired on a BD Canto II. From 31st of January 2018 onwards, the EuroFlow 8-colour ALOT panel was replaced by an in-house established 12-colour panel (ZH ALOT) (Table 1) and samples were acquired on a BD LSRII Fortessa. At least 30,000 events in the WBC gate (defined by SSC/FSC) were recorded.

	FITC	**PerCP -Cy5.5**	**PE**		**PE-Cy7**	**APC**		**APC-H7**	**PB**	**Pacific Orange**		
EuroFlow ALOT	MPO#	CD34	cyCD79#		CD19	CD7		sCD3	cyCD3#	CD45		
	FITC	**PerCP -Cy5.5**	**PE**	**PE-CF594**	**PE-Cy7**	**APC**	**APC-R700**	**APC-H7**	**PB**	**V500**	**BV605**	**BV786**
ZH ALOT	MPO#	CD34	cyCD79#	CD9	CD19	CD7	CD33	CD20	cyCD3#	CD45	CD5	sCD3

Table 1: Design of Euroflow 8-colour ALOT panel and ZH 12-colour ALOT panel. sCD3: surface CD3, cyCD3: cytoplasmatic CD3. #: intracellular staining.

2.2 Data analysis in diagnostics unit and flow cytometry expert laboratory

Before the acquisition of samples, BD Cytometer Setup & Tracking Research Beads (BD) were run to reproducibly set up the cytometer from day to day. Compensation matrix was calculated using BD CompBeads (BD). Data from acquired samples were exported as .fcs files. In the diagnostic unit, files were analysed with BD FACSDiva Software (Version 8.0) by trained staff members. The FlowJo Software (Version 10.8.1) was used for flow cytometry analysis by an experienced researcher. Cell frequencies were assessed by manual gating.

Chapter 3 Automated immune cell classification for paediatric ALL

In this report we present our work on developing a machine learning framework for automated analysis of a standardised diagnostic childhood lymphoblastic leukaemia staining to extract the frequencies of lineage populations. Since B-cell precursor (BCP) ALL represents the main phenotype of childhood leukaemia, our analysis focused on that particular disease type [10]. Our method was applied to bone marrow samples of a paediatric BCP-ALL cohort enrolled into AIOEP BFM 2009 and AIOEP BFM 2017 clinical trial from 2016 – 2019 in and reported to the Swiss AIOEP BFM coordinating clinical trial centre at Children's Hospital Zurich. An overview of the proposed framework is given in Figure 1. Briefly, the framework consists of a *learning step* (black arrows in Figure 1), where few representative patient samples are preprocessed, clustered, and annotated to construct a reference data which is then used to train a neural network model able to classify flow cytometry profiles into known cell lineages, and an *inference step*, where the remaining unseen samples are preprocessed and fed into the trained classifier to identify their cell lineage composition (green arrows in Figure 1). Details about the individual steps follow in the next sections.

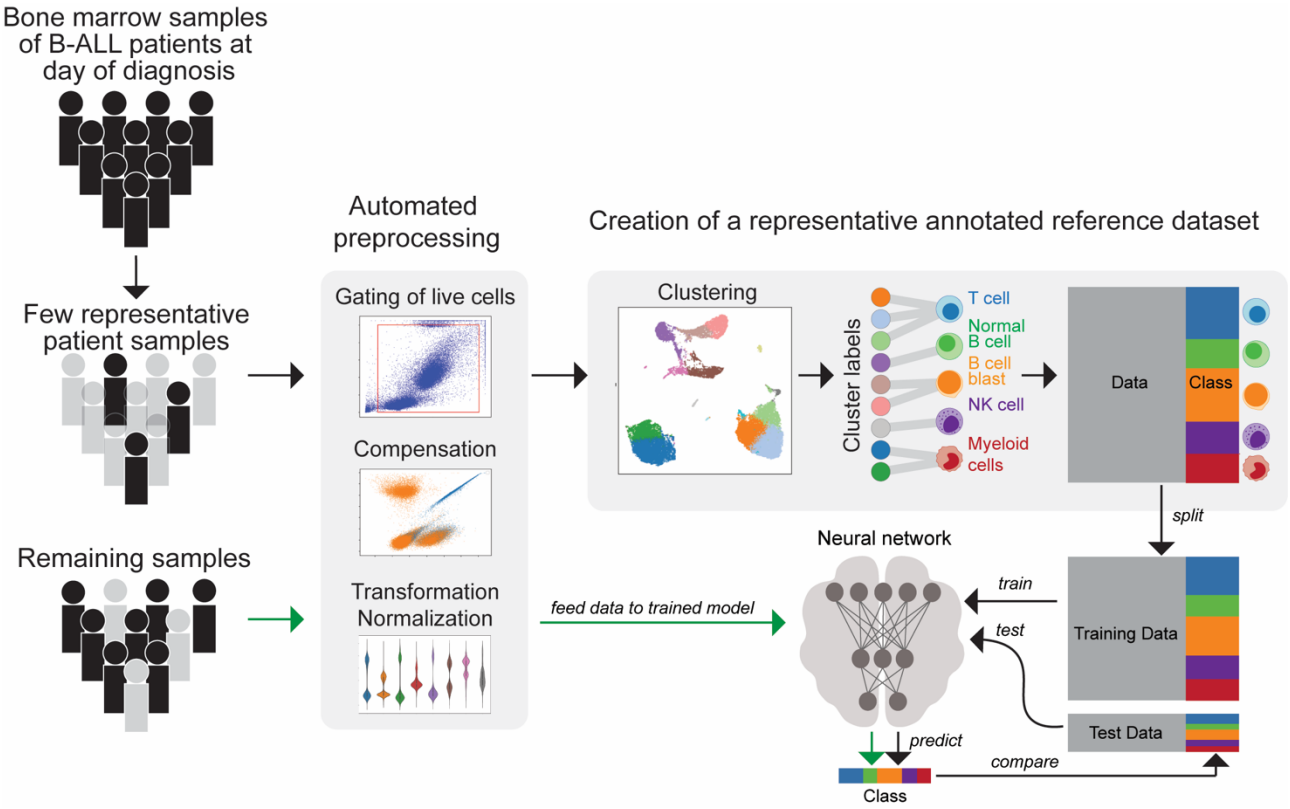

Figure 1: Schematic overview of automated immune cell classification using flow cytometry bone marrow measurements.

3.1 Dataset description

To develop our framework, we exploited a dataset of n=187 flow cytometry samples at the time of diagnosis from B-ALL patients enrolled in clinical studies and treated in seven hospitals in Switzerland from both the EuroFlow ALOT and the Zurich (ZH) ALOT (Table 1). As shown in Table 2, these panels contain lineage markers which show specific and constant expression patterns on non-malignant B cells, T cells, NK cells and Myeloid cells in the bone marrow. Cells which do not match the criteria are considered to be doublets or debris and can be classified as "undefined". For each panel, out of these 187 samples, 5 were selected to build a reference dataset. Important criteria for selection of these files were to capture both hospital or origin variability and disease heterogeneity (e.g., files that contained leukemic blasts with variable CD34 and CD45 expression). For the ZH ALOT, we excluded markers CD9, CD34 and cyCD3 from the entire analysis due to signal artefacts caused by off-scale signal or incorrect compensation in some files. We verified that the signal artefacts did not impact the signal in the other channels. Flow cytometry measurements were acquired as described in Chapter 2.

	B cells	**T cells**	**NK cells**	**Myeloid cells**	**B cell blasts**
CD45	**pos**	**pos**	**pos**	**pos**	variable
CD19	**pos**	neg	neg	neg	variable
CD79a	**pos**	neg	neg	neg	variable
CD20	neg → **pos**	neg	neg	neg	variable
CD7	neg	**pos**	**pos**	neg	neg
Surface CD3	neg	**pos**	neg	neg	neg
Cytoplasmatic CD3	neg	**pos**	neg	neg	neg
MPO	neg	neg	neg	**pos**	neg
CD33	neg	neg	neg	**pos**	neg
CD34	**pos** → neg	neg	neg	**pos**	variable
CD5	neg → **pos** → neg	**pos**	neg	neg	variable

Table 2: Expression of lineage markers on healthy B cells, B cell leukemic blasts, T cells, NK cells and Myeloid cells. Pos/neg: positive/negative expression of the marker in the cell lineage of interest.

3.2 Data preprocessing

The first step in our proposed pipeline is the automatic preprocessing of the .fcs files, which consist of the following substeps:

3.2.1 Gating of live cells

After loading the measurements in the .fcs files, we need to gate the live cells and remove any debris, dead cells and doublets. This step is typically achieved by gating the live cells on the FSC-A, SS-A plot, however the gating strategy needs to be adapted to each file individually. To automate this process, after removing all negative values from the FSC-A channel, we fit the FSC-A distribution, and identify the first local minimum of the distribution plot. This value serves as the low gate threshold of our FSC-A channel. We then set the upper threshold to 250,000 following current standard practices in the field, and also set the SSC-A gate to a low value of 0 and a high value of 250,000.

Example results of this automated gating strategy for 4 individual files with pronounced discrepancies are seen in Figure 2.

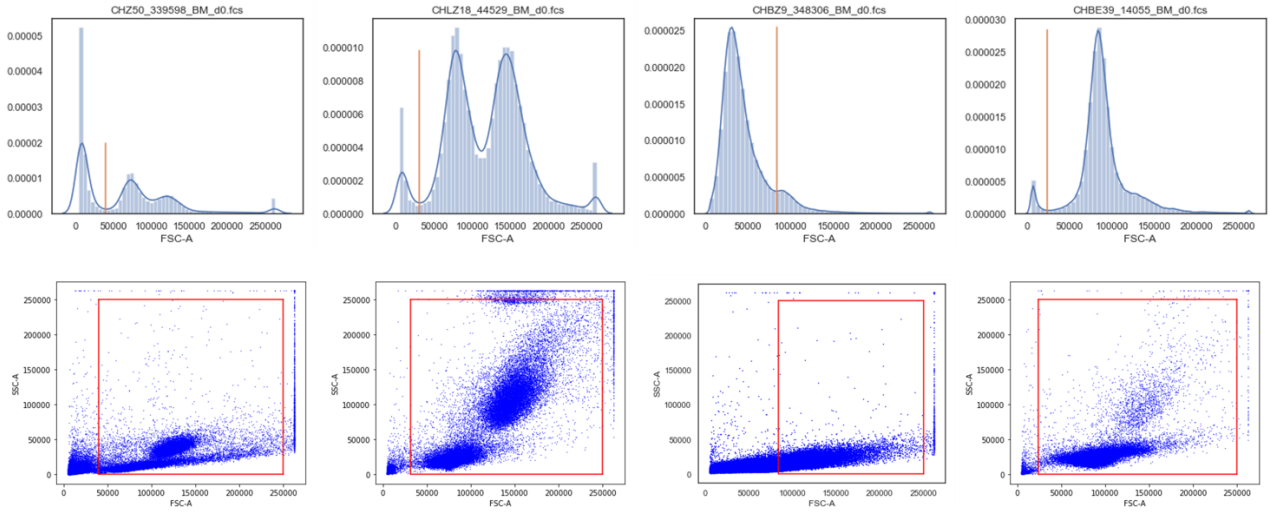

Figure 2: Automatic live cell gating of 4 individual .fcs files (columns) where the FSC-A channel distribution (top row) and the FSC-A, SSC-A scatter plot is shown. Orange vertical lines (top row) indicate low FSC-A gate and red rectangles indicate selected gate on both channels.

3.2.2 Compensation, transformation, and normalisation

After gating, remaining outlier cells were removed by identifying the top 0.1 percentile of the values for each channel and removing the cells that fall on this percentile in at least one channel. For the ZH ALOT, we additionally removed the top 0.7 percentile of CD34 channel due to strong presence of outliers. Then, the .fcs files were compensated using a spillover matrix with manually curated spillover coefficients. The measurements were transformed using an inverse hyperbolic sine (asinh) transformation, following standard practice in flow cytometry data analysis, where, for each protein vector y we have:

$$asinhy = ln\left(\frac{y}{c} + \sqrt{\frac{y^2}{c} + 1}\right)$$

where c is the cofactor, fixed here to 500. Finally, the transformed measurements were scaled using a standard min-max normalisation.

3.3 Data annotation

After automatic data preprocessing, the data matrices of all 5 input files were independently clustered using the popular community detection algorithm Phenograph [11] with a Euclidean distance metric and *k*=200 nearest neighbours. The results of the clustering were assessed visually using a Uniform Manifold Approximation and Projection (UMAP) [12] dimensionality reduction approach and the clusters were annotated as T cells, normal B cells, NK cells, Myeloid/monocyte cells, Leukemic Blasts or Undefined, based on prior knowledge of protein expression in these cell lineages (Table 2). The results of this process for one of the input samples are shown graphically in Figure 3.

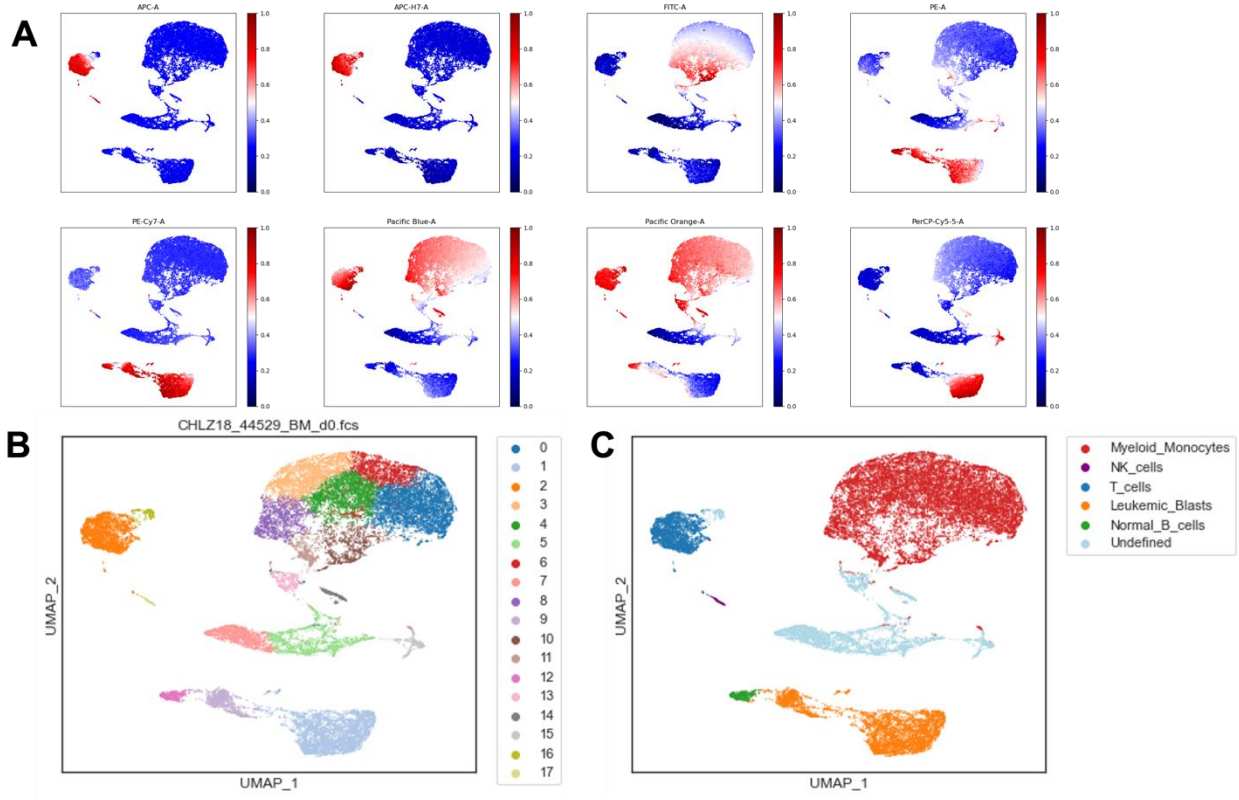

Figure 3: Annotation process of one training sample.

(A) UMAP projection colored by protein intensity for all 8 channels of the EuroFlow ALOT. (B) Corresponding cluster labels as identified by Phenograph. (C) Results of annotations and assigning clustering labels to known cell lineages.

This process was repeated and carefully assessed for all 5 files. Finally, all 5 reference files together with their cell lineage labels are concatenated to create the reference dataset. The same process was repeated for the ZH ALOT samples.

3.4 Training of cell lineage classifier

The integrated and labelled dataset is then used to train a neural network classifier that can automatically predict cell lineages from the input measurements. We first randomly split the dataset into a training and test set (70% and 30% of the data, respectively). The neural network model architecture consisted of 1 input layer of 8 neurons, two hidden layers of 30 neurons each and one output layer of 6 neurons, equal to the 6 classes in the data (Figure 4 A). All layers are fully connected and use a Rectified Linear Unit (ReLU) activation function, apart from the output layer that uses a sigmoid activation function. The classifier was trained using the Adam optimizer [13] with a learning rate of 0.001 and a batch size of 256. Its performance was evaluated using a weighted cross-entropy loss function that uses the inverse class frequencies as class weights. The weighted loss function helped tackle the class imbalance found in our data, where some cell lineages (e.g., NK cells) are found in much smaller proportions than others. Training was finalised after at most 500 epochs; we

also used an early stopping criterion by terminating training when the model's performance failed to improve for 20 consecutive epochs.

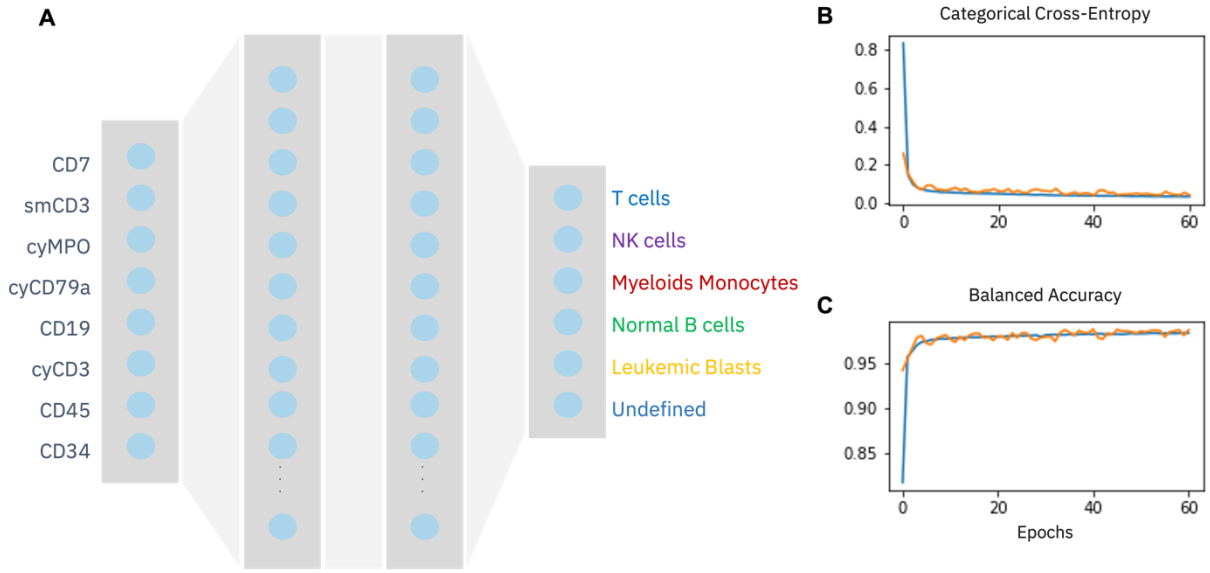

Figure 4: Model architecture and training results of cell lineage neural network classifier.

(A) Model architecture of the neural network. The neural network consists of 1 input layer of 8 neurons, two hidden layers of 30 neurons each and 1 output layer of 6 neurons, equal to the 6 classes in the data. All layers are fully connected using a Rectified Linear Unit (ReLU) activation function, apart from the output layer that uses a sigmoid activation function. The classifier was trained using the Adam optimizer [13] with a learning rate of 0.001 and a batch size of 256. (B-C) Evolution of the categorical cross entropy loss function and the balanced accuracy in the training and test data (blue and orange curves, respectively). Training was terminated after 60 epochs due to early stopping.

To examine the model's performance, we first assessed the evolution of the loss function and the balanced accuracy during training (Figure 4 B-C), where we observed that the model converged after approximately 60 epochs, reaching a very high balanced accuracy of approximately 0.98. We also assessed the performance of the classifier per class, where we observed that the model's F1-score (harmonic mean of precision and recall) for all cell lineages was higher than 0.90.

	Precision	Recall	F1-Score	Cell counts
Leukemic Blasts	1.00	0.99	0.99	22730
Myeloid	1.00	0.99	0.99	7700
NK cells	0.94	1.00	0.97	237
Normal B cells	0.83	0.98	0.90	1115
T cells	1.00	0.98	0.99	6017
Undefined	0.97	0.99	0.98	8073

Figure 5: Performance of the classifier for each cell lineage in terms of precision, recall and F1-Score.

3.5 Testing of cell lineage classifier in unseen samples

Since the model's performance on the test data from the 5 representative files was deemed satisfactory, we next applied it to the remaining samples. These samples were first preprocessed following the automated strategy of Section 3.2 and then fed into the trained classifier. The results for 3 example unseen datasets can be seen in Figure 6, which shows UMAP embeddings of the high-dimensional measurements coloured by the model's predictions. Since this data has not been annotated, we cannot derive a quantitative measurement of the model's performance. However, although the model was never given the UMAP embedding as an input, we observe that the clusters within these embeddings are remarkably homogeneous with respect to their class labels, a result that we further validated by assessing the marker expression in each UMAP cluster. Taken together, these results strengthen our confidence in the model's ability to accurately identify cell lineages from unseen measurements.

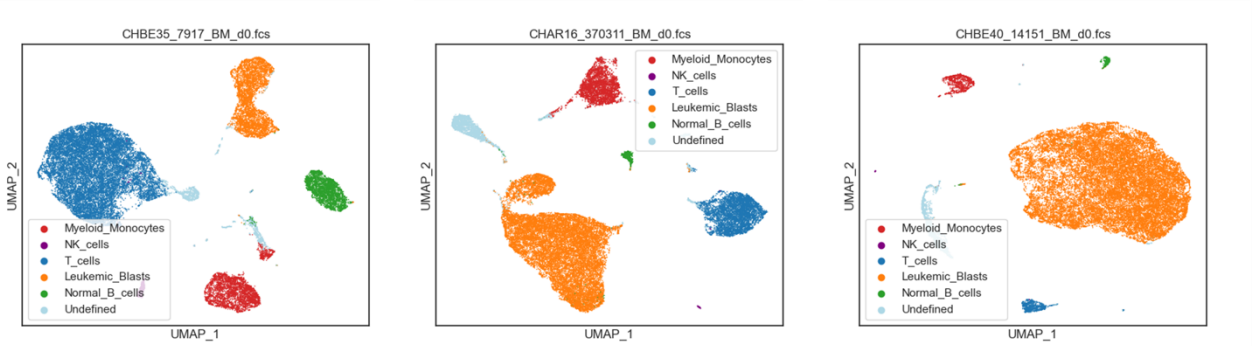

Figure 6: Performance of the model in 3 unseen samples.

For each sample, a UMAP plot of the measurements coloured by the prediction of the classifier is shown.

Chapter 4 Association of cell lineage frequencies with clinical data

After demonstrating that the cell lineage classifier can accurately detect cell lineages of interest, we feed all remaining unseen samples to predict, for each one, its cell lineage composition. This process was performed for both the EuroFlow and ZH ALOT models, resulting in an integrated dataset where each patient is represented by a vector of cell lineage frequencies. We then proceeded to cluster both rows and columns of this integrated dataset using a hierarchical clustering method with an average linkage. To compute distances between the rows and columns we used a correlation metric and a Jensen-Shannon divergence score, respectively. The results of this process can be seen in Figure 7, where we also overlay various metadata on the top of the heatmap. We observe that the ALOT label (EuroFlow or ZH) is randomly distributed across all clusters, indicating that there are no batch effects associated with the ALOT in the integrated dataset. The same is true for the City of Origin label. Apart from a small cluster on the left that is characterised by low blast frequency, most of the samples have a high blast frequency that dominates the other populations. We also do not observe a clustering of the high risk or relapse samples, which again appear to be randomly distributed across all clusters high for blasts.

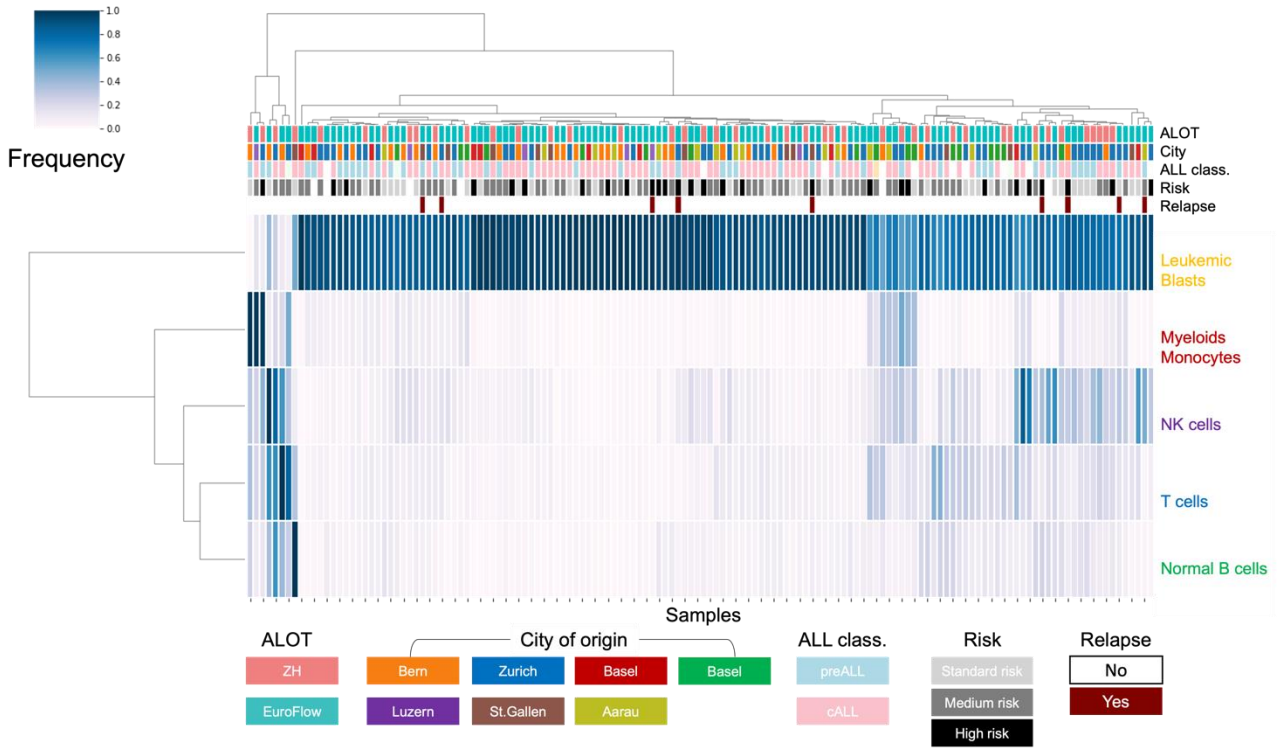

Figure 7: Clustering of cell lineage frequencies per sample and its associations with various ALL metadata.

In the clustered heatmap, columns indicate patient samples, rows correspond to cell lineages of interest and colour intensity indicates the frequency of cell lineage. Relevant ALL metadata are shown on top of the heatmap, with their exact values given in the bottom legend.

We then proceeded to focus on the percentage of leukemic blasts with respect to different disease metadata. We observe that there is no association between blast frequency and risk stratification or relapse, but samples that exhibit a minimal residual disease slow early response (MRD SER) at day 33 appear to all be very high in terms of blast frequency (Figure 8).

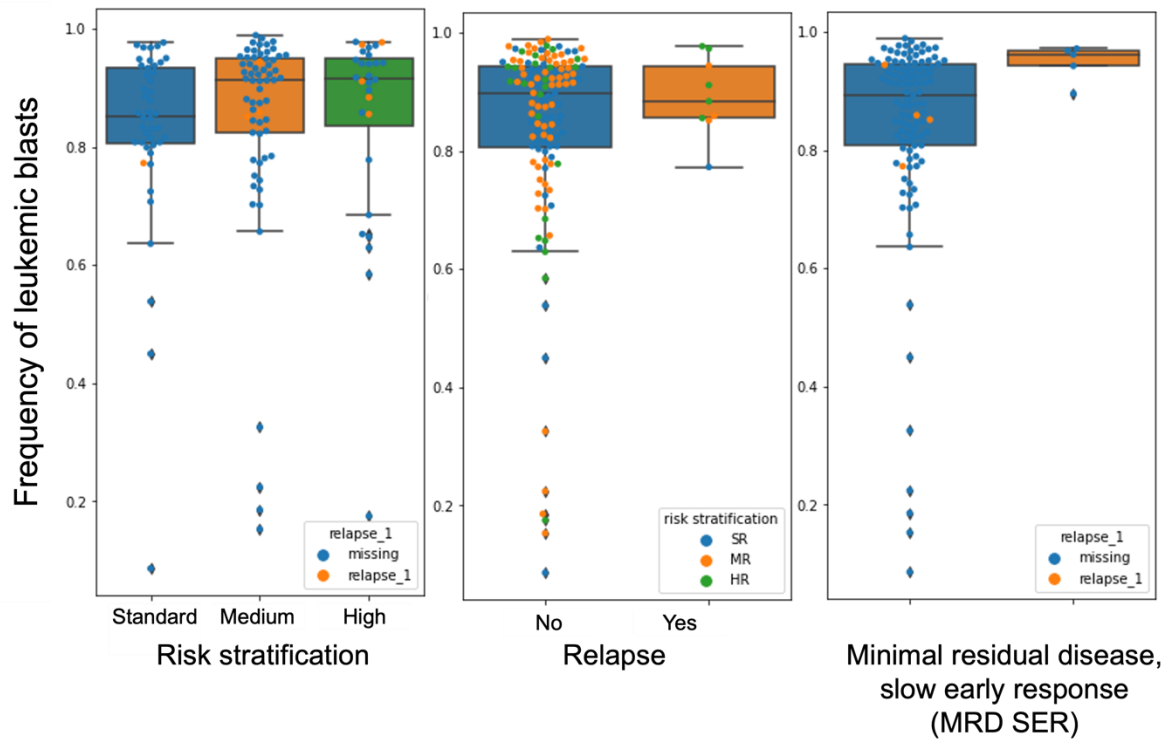

Figure 8: Association of frequency of leukemic blasts and various disease metadata.

We then focused on the nature of the blast population, which, as previously explained, exhibits variable expression of CD34. We quantified the percentage of blasts that are positive for CD34 and, this time, we observe that the frequency of CD34 positive blasts is associated both with the risk stratification and MRD SER, with high-risk samples and MRD SER positive samples high for CD34 blast frequency.

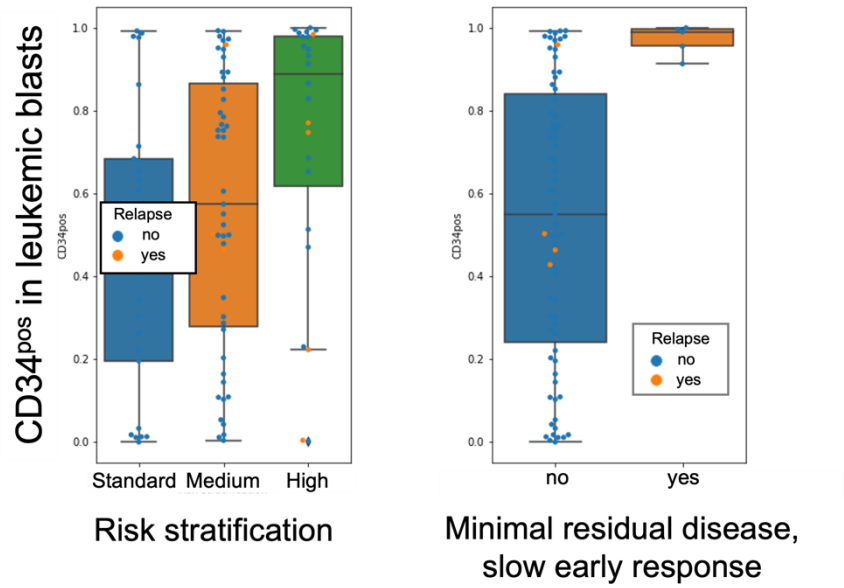

Figure 9: Association of percentage of CD34 positive leukemic blasts with disease metadata.

Chapter 5 Summary and Conclusion

This report describes the development of a unified machine learning framework for automated analysis of a standardised diagnostic paediatric leukaemia staining to extract the frequencies of cell lineage populations. The framework exploits a large cohort of bone marrow samples from paediatric B-ALL patients analysed in the clinic by standard diagnostic flow cytometry analysis using two different staining panels. Our work overcomes the need for manual gating and addresses the issues of staining irregularities and other sources of batch effects. This is achieved by circumventing the need to integrate the data and creating instead a baseline cell lineage classification model that can accurately detect cell populations of interest in unseen flow cytometry samples. Our model provides a proof of concept that our method can deal with current challenges in diagnostic flow cytometry data analysis and allows a fast, robust and accurate identification of cell populations of interest with little to no expert curation of results.

As a future step, we are currently working on benchmarking our method against other known flow cytometry automated gating approaches (e.g., DeepCyTOF [14], flowDensity [15], OpenCyto [16], FlowMap-FR [17]). We are also working towards validating our method using a second, independent B-ALL cohort, hoping to show that it can generalise equally well when data from a distinct unseen cohort are used.

Chapter 6 Future work

The software presented in this deliverable has been developed and optimised using B-ALL samples, as the data was readily available in the iPC consortium through the labs of UZH partners Profs. Burkhard Becher and Jean-Pierre. Bourquin. We are currently writing a manuscript to publish the software and data. We will provide an update and short summary about the status of the manuscript, data and code release in the final periodic report.

In a parallel collaboration with Prof. Burkhard Becher, we are adapting the presented methodology to the analysis of bone marrow AML samples, where we aim at characterising the cellular and molecular patterns that distinguish diagnostic, remission and relapse samples from the same patient. In future work, we will extend the analysis to include solid tumours. We are currently discussing with other iPC partners data generation for solid tumours.

Chapter 7 List of Abbreviations

Abbreviation	Translation
ALL	Acute lymphoblastic leukemia
ALOT	Acute Leukemia Orientation Tube
BCP-ALL	B cell precursor acute lymphoblastic leukemia
ReLU	Rectified Linear Unit
UMAP	Uniform Manifold Approximation and Projection

Chapter 8 Bibliography

[1] P. S. Gaynon, “Childhood acute lymphoblastic leukaemia and relapse,” *Br. J. Haematol.*, vol. 131, no. 5, pp. 579–587, 2005, doi: 10.1111/j.1365-2141.2005.05773.x.

[2] M. Greaves, “A causal mechanism for childhood acute lymphoblastic leukaemia,” *Nat. Rev. Cancer*, vol. 18, no. 8, Art. no. 8, Aug. 2018, doi: 10.1038/s41568-018-0015-6.

[3] M. F. Greaves and J. Wiemels, “Origins of chromosome translocations in childhood leukaemia,” *Nat. Rev. Cancer*, vol. 3, no. 9, Art. no. 9, Sep. 2003, doi: 10.1038/nrc1164.

[4] M. N. Dworzak *et al.*, “AIEOP-BFM Consensus Guidelines 2016 for Flow Cytometric Immunophenotyping of Pediatric Acute Lymphoblastic Leukemia,” *Cytometry B Clin. Cytom.*, vol. 94, no. 1, pp. 82–93, 2018, doi: 10.1002/cyto.b.21518.

[5] R. Ratei *et al.*, “Lineage classification of childhood acute lymphoblastic leukemia according to the EGIL recommendations: results of the ALL-BFM 2000 trial,” *Klin. Padiatr.*, vol. 225 Suppl 1, pp. S34-39, May 2013, doi: 10.1055/s-0033-1337961.

[6] D. F. Ransohoff, “Lessons from controversy: ovarian cancer screening and serum proteomics,” *J. Natl. Cancer Inst.*, vol. 97, no. 4, pp. 315–319, Feb. 2005, doi: 10.1093/jnci/dji054.

[7] L. M. Weber and M. D. Robinson, “Comparison of clustering methods for high-dimensional single-cell flow and mass cytometry data,” *Cytometry A*, vol. 89, no. 12, pp. 1084–1096, Dec. 2016, doi: 10.1002/cyto.a.23030.

[8] S. Van Gassen, B. Gaudilliere, M. S. Angst, Y. Saeys, and N. Aghaeepour, “CytoNorm: A Normalization Algorithm for Cytometry Data,” *Cytometry A*, vol. 97, no. 3, pp. 268–278, 2020, doi: 10.1002/cyto.a.23904.

[9] J. J. M. van Dongen *et al.*, “EuroFlow antibody panels for standardized n-dimensional flow cytometric immunophenotyping of normal, reactive and malignant leukocytes,” *Leukemia*, vol. 26, no. 9, Art. no. 9, Sep. 2012, doi: 10.1038/leu.2012.120.

[10] A. Möricke *et al.*, “Risk-adjusted therapy of acute lymphoblastic leukemia can decrease treatment burden and improve survival: treatment results of 2169 unselected pediatric and adolescent patients enrolled in the trial ALL-BFM 95,” *Blood*, vol. 111, no. 9, pp. 4477–4489, May 2008, doi: 10.1182/blood-2007-09-112920.

[11] J. H. Levine *et al.*, “Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis,” *Cell*, vol. 162, no. 1, pp. 184–197, Jul. 2015, doi: 10.1016/j.cell.2015.05.047.

[12] L. McInnes, J. Healy, and J. Melville, “Umap: Uniform manifold approximation and projection for dimension reduction,” *ArXiv Prepr. ArXiv180203426*, 2018.

[13] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *ArXiv14126980 Cs*, Jan. 2017, Accessed: Oct. 15, 2021. [Online]. Available: http://arxiv.org/abs/1412.6980

[14] H. Li *et al.*, “Gating mass cytometry data by deep learning,” *Bioinformatics*, vol. 33, no. 21, pp. 3423–3430, Nov. 2017, doi: 10.1093/bioinformatics/btx448.

[15] M. Malek, M. J. Taghiyar, L. Chong, G. Finak, R. Gottardo, and R. R. Brinkman, “flowDensity: reproducing manual gating of flow cytometry data by automated density-based cell population identification,” *Bioinformatics*, vol. 31, no. 4, pp. 606–607, Feb. 2015, doi: 10.1093/bioinformatics/btu677.

[16] G. Finak *et al.*, “OpenCyto: An Open Source Infrastructure for Scalable, Robust, Reproducible, and Automated, End-to-End Flow Cytometry Data Analysis,” *PLOS Comput. Biol.*, vol. 10, no. 8, p. e1003806, Aug. 2014, doi: 10.1371/journal.pcbi.1003806.

[17] C. Hsiao, M. Liu, R. Stanton, M. McGee, Y. Qian, and R. H. Scheuermann, “Mapping cell populations in flow cytometry data for cross-sample comparison using the Friedman-Rafsky test statistic as a distance measure,” *Cytom. Part J. Int. Soc. Anal. Cytol.*, vol. 89, no. 1, pp. 71–88, Jan. 2016, doi: 10.1002/cyto.a.22735.