Interpreting single-cell messages in normal and aberrant hematopoiesis with the Cell Marker Accordion

Emma Busarello[1,*,✉], Giulia Biancon[2,*], Fabio Lauria[3], Zuhairia Ibnat[1], Christian Ramirez[1], Gabriele Tomè[1,3], Kristin R. Aass[4], Jennifer VanOudenhove[2], Therese Standal[4], Gabriella Viero[3], Stephanie Halene[2,✉] and Toma Tebaldi[1,2,✉]

1 Laboratory of RNA and Disease Data Science, Department of Cellular, Computational and Integrative Biology (CIBIO), University of Trento, Trento, Italy

2 Section of Hematology, Department of Internal Medicine, Yale Comprehensive Cancer Center, Yale University School of Medicine, New Haven, CT, USA

3 Institute of Biophysics, CNR Unit at Trento, Italy

4 Norwegian University of Science and Technology (NTNU), Trondheim, Norway

** Equally contributed*

Correspondence to: emma.busarello@unitn.it, stephanie.halene@yale.edu, toma.tebaldi@unitn.it

Disclosure of Conflicts of Interest

S.H., consultancy, Forma Therapeutics. Other authors declare no competing financial interests.

Abstract

Single-cell technologies offer a unique opportunity to explore cellular heterogeneity in hematopoiesis, reveal malignant hematopoietic cells with clinically significant features and measure gene signatures linked to pathological pathways. However, reliable identification of cell types is a crucial bottleneck in single-cell analysis. Available databases contain dissimilar nomenclature and non-concurrent marker sets, leading to inconsistent annotations and poor interpretability. Furthermore, current tools focus mostly on physiological cell types, lacking extensive applicability in disease.

We developed the Cell Marker Accordion, a user-friendly platform for the automatic annotation and biological interpretation of single-cell populations based on consistency weighted markers. We validated our approach on peripheral blood and bone marrow single-cell datasets, using surface markers and expert-based annotation as the ground truth. In all cases, we significantly improved the accuracy in identifying cell types with respect to any single source database.

Moreover, the Cell Marker Accordion can identify disease-critical cells and pathological processes, extracting potential biomarkers in a wide variety of contexts in human and murine single-cell datasets. It characterizes leukemia stem cell subtypes, including therapy-resistant cells in acute myeloid leukemia patients; it identifies malignant plasma cells in multiple myeloma samples; it dissects cell type alterations in splicing factor-mutant cells from myelodysplastic syndrome patients; it discovers activation of innate immunity pathways in bone marrow from mice treated with METTL3 inhibitors.

The breadth of these applications elevates the Cell Marker Accordion as a flexible, faithful and standardized tool to annotate and interpret hematopoietic populations in single-cell datasets focused on the study of hematopoietic development and disease.

Statement of significance

We developed the Cell Marker Accordion, a user-friendly platform to annotate and interpret single-cell data in normal and aberrant hematopoiesis. We a) significantly improve in annotation accuracy; b) provide detailed information on genes that influence annotation outcomes; c) identify disease-critical cells, pathological processes and potential expression biomarkers in different contexts.

Introduction

Single-cell RNA sequencing (scRNA-seq) characterizes the transcriptome of each individual cell in large populations. This high-throughput approach is the ideal choice to reveal the heterogeneous landscape of normal and aberrant hematopoiesis[1,2], composed of cells characterized by differing self-renewal capacity, multipotent potential and high plasticity, and involved in infections and other diseases controlling immune responses[3–5].

With the enormous opportunities offered by single-cell technologies, a new set of challenges is rapidly emerging in data analysis and interpretation. Accurate and reliable annotation of cell types is key to derive faithful biological conclusions. In fact, robustness in identifying cell types is an essential prerequisite for studying hematologic disorders, to discern disease-critical cells, characterized by aberrant cell states responsible for disease progression and therapy resistance[6]. In addition, measuring the single-cell activity of gene signatures, or modules, associated with pathologically relevant pathways is fundamental to unravel pathogenic mechanisms in aberrant cells[7] as well as to discover potential disease biomarkers[8].

Identification of cell populations within single-cell data can be executed manually or automatically[9]. Manual annotation, based on the investigator's knowledge or derived from published literature is generally subjective and often non-reproducible due to lack of standards. Many computational tools perform automatic annotation by correlating reference expression data or by transferring labels from other single-cell datasets[10–14].These approaches require reliable transcriptome profiles of purified cells or high-quality annotated single-cell data[15]. However, such reference datasets are not easily available, especially for pathological samples; they can lack the cell populations of interest, and might be susceptible to technical specificities such as platform or sequencing strategy[16]. Alternatively, automatic annotation can be achieved by employing predefined sets of cell marker genes[17–22]. The majority of current tools require the user to provide a collection of markers, a process prone to bias[12,21,23]. We show that currently available gene marker databases are extremely heterogeneous, contain different marker sets for the same cell type, and use a non-standard nomenclature and classification, leading to inconsistent annotation of cell populations in scRNA-seq data and poor interpretability of results. Furthermore, current tools and resources focus mostly on physiological cell types, limiting the identification of disease-critical cells.

To address these issues and improve the interpretation of normal and aberrant hematopoietic cell types in single-cell data we developed the Cell Marker Accordion, an easily accessible and well-documented platform constituted by an interactive R Shiny web application requiring no programming skills, and an R package to automate the annotation.

The Cell Marker Accordion database is built upon multiple published databases of human and mouse gene markers for hematopoietic cell types[24–30], standard collections of widely used cell sorting markers (Abcam and Thermo Fisher Scientific) and literature-based marker genes associated with disease critical cells in aberrant hematopoiesis in leukemia and myeloma. The Accordion database allows marker genes to be weighted not only by their specificity, but also by their evidence consistency scores (ECs), measuring the agreement of different annotation sources. The Cell Marker Accordion web interface permits to explore the integrated collection of human and mouse marker genes and to easily browse hierarchies of hematopoietic cell types following the Cell Ontology structure in order to obtain the desired level of resolution.

The Cell Marker Accordion R package allows to automatically annotate healthy and aberrant populations in single cell datasets, exploiting positive and negative markers from either the built-in Accordion database, or any gene signature of interest provided by users. Genes, cell types or pathways that mostly influence annotation results can be easily accessed and visualized to allow the transparent interpretation of results.

We benchmarked the Cell Marker Accordion on peripheral blood and bone marrow single-cell datasets, using surface markers and expert-based annotation as the ground truth. In all cases, we significantly improved the annotation accuracy with respect to any available single-source database. Moreover, we show that the Cell Marker Accordion can be used to identify pathological processes and disease-critical cells: leukemia stem cell subtypes, including: therapy-resistant cells, in acute myeloid leukemia patients[31–34]; malignant plasma cells in multiple myeloma samples[35–37]; cell type alterations driven by pathologically relevant mutations in myelodysplastic syndromes[38,39]; activation of innate immunity pathways in bone marrow from mice with Mettl3 deletion or treated with METTL3 inhibitors[40,41].

The Cell Marker Accordion is a user-friendly, flexible and comprehensive tool that improves the annotation and interpretation of both physiological and pathological hematopoietic populations with single-cell resolution.

Methods

Data sources of the Cell Marker Accordion database

The Cell Marker Accordion database was constructed by considering multiple published marker gene databases (CellMarker[24], PanglaoDB[25], GeneMarkeR[26], ASCT+B[27], MSigDB[28], Azimuth[29], CellTypist[42]) and collections of cell sorting markers (Abcam; Thermo Fisher Scientific). First, to have a quantitative measure of marker genes overlap between available databases, we computed the jaccard similarity. Next, we considered human and mouse marker genes that are associated with hematopoietic cell lineages. Both positive and negative markers, when present, were selected.

Next, database integration was performed. Marker genes' nomenclature was standardized to ensure the most recent approved version of gene symbols. HUGO Gene Nomenclature Committee (2022) and Mouse Genome Informatics (v. 6.21) resources were employed to standardize human and mouse gene names, respectively. The collected databases, as well as the cell sorting markers' repositories, report different cell type labels, which makes information integration unfeasible. For this reason, annotations were first standardized by mapping initial cell types label to the Cell Ontology http://obofoundry.org/ontology/cl.html)[42]. Cell hierarchy information was extracted considering the "hematopoietic cell" node as the root and all its descendants. Gene markers associated with hematologic disease-critical cells were collected through a literature search. All annotation sources and references are reported in the Cell Marker Accordion database. The Disease Ontology

(https://disease-ontology.org/) was exploited to standardize disease names and IDs. Overall, the Cell Marker Accordion includes a comprehensive set of 5878 marker genes associated with 140 standardized hematopoietic cell types for the analysis of human samples and 2175 marker genes associated with 97 cell types for the analysis of mouse samples.

Definition of integration scores for marker genes in the Cell Marker Accordion database

To rank and filter marker genes, the specificity and the evidence consensus (EC) score were computed for each marker. The specificity ranges from 0 to 1 and reflects how many cell types a gene is marker for. It is calculated separately for positive and negative markers in human and mouse as:

Specificity of marker *x*= 1/Number of cell types which have gene *x* as a marker

A high score indicates that marker *x* is highly specific for a certain cell type, while low scores are associated with markers spread among multiple cell types.

The EC score evaluates the agreement of different annotation sources and can be used as a measure of marker robustness and reliability. It is calculated as follows:

EC*x,y*=Number of sources which have gene *x* as a marker of cell type *y*

A high EC score indicates more consensus of marker *x* among several sources, while low scores are associated with markers that are present only in few sources.

Implementation of the Cell Marker Accordion R package for automatic annotation

We developed an R package to automatically identify cell type, cell cycle stage and pathway activation in single-cell RNA-seq data (https://github.com/TebaldiLab/cellmarkeraccordion). Users can annotate clusters or cells exploiting the built-in Cell Marker Accordion database with the *accordion()* function or can provide custom sets of markers using the *accordion_custom()* function. Both functions take as input a Seurat object (versions 4 or 5[29]) and execute the following operations. First, if no prior normalization and scaling steps have been performed, the single-cell expression matrix is normalized and scaled only on input marker genes. Based on the input, specificity and EC scores are computed, scaled, log transformed and multiplied to obtain a comprehensive weight for each marker gene. Next, each scaled gene expression level is multiplied by this weight, obtaining a gene weighted expression score for each cell. For each cell type the normalized sum of all associated marker genes is calculated by summing, cell by cell, the weighted expression score divided by the square root of the weighted sum. This step leads to a *ct* x *n* enrichment score matrix, where rows represent cell types and columns represent cells. For each individual cell, the highest *ct* score is used to assign the corresponding cell type. The score per cluster is performed by calculating for each row the third quartile across cells corresponding to a particular cluster *cl*. The cell type with the maximum score is then assigned to the cluster *cl* (**Supplementary Figure 1**).
By default, the gene impact score is calculated as the third-quantile of the distribution of the gene score for each cell type. The cell type impact score is calculated by default as the third-quantile of the distribution of the cell type score for each cluster.

The Cell Marker Accordion Shiny app

The R Shiny tool is available at https://rdds.it/CellMarkerAccordion/. The Shiny app incorporates reactive programming allowing users to access the Cell Marker Accordion database and to retrieve marker genes that are specific for their selected cell types. Furthermore, when users choose genes of interest on the marker gene tab, the tool interactively retrieves the standardized cell types associated with the selected genes.

Validation of the Cell Marker Accordion

To validate the Cell Marker Accordion, we exploited four different single-cell and multi-omics datasets listed in **Supplementary Table 1** (see **Supplementary Methods**). For this benchmark, ScType[43] was selected among currently available marker-based automatic annotation tools such as SCINA[23], clustifyR[12], scCATCH[18], scSorter[21]. The choice of scType is based on its possibility to use both positive and negative markers, and its improved performance, in terms of both annotation accuracy and running time, with respect to the other tools (tested on the Zheng et al., 2017[44] dataset). To measure performance accuracy we used the F1-score which is the harmonic mean of precision and sensitivity.

MDS single cell dataset

Bone marrow samples from MDS patients (**Supplementary Table 2**) were processed as previously reported in Biancon et al., 2022[44]. Briefly, viable (7-AADneg) CD34pos cells were sorted by the Yale Flow Cytometry facility on the FACSAria instrument (BD Biosciences) and subsequently processed for scRNA-seq library preparation by the Yale Center for Genome Analysis using Chromium Next GEM Single Cell 5' kit v2 (10x Genomics). A total of 64915 sequenced cells were used for downstream analysis, with an average of 71828 reads per cell and 3476 genes per cell. Cell-variant assignment, based on U2AF1 S34F mutation calling in the 21:43104346-43104346 locus, was performed with VarTrix v1.1.19 (https://github.com/10XGenomics/vartrix). Single cell expression data analysis was performed as described in supplementary methods.

Results

Widespread heterogeneity across annotation sources leads to inconsistent cell type annotation

To unravel information discrepancies across currently available gene marker databases, we automatically annotated a published scRNA-seq dataset of human bone marrow[45], extracting marker genes from CellMarker[24] and Panglao DB[25], two of the most comprehensive databases for cell type markers (**Figure 1A**). Cell type annotation was inconsistent between the two sources, showing divergent cell types (for example, "platelet" with CellMarker and "natural killer cell" with PanglaoDB) or using different nomenclature (for example "Red blood cell (erythrocyte)" and "erythroid progenitor cell"). The maximum Jaccard similarity index between cell types is 0.23, unfolding very high discrepancies in terms of marker genes shared among these resources (**Figure 1B**).
To extend this initial observation, we systematically explored the heterogeneity of seven available marker genes databases over common cell types. The comparison showed a low consistency between databases, with an average Jaccard similarity of 0.04 and a maximum of 0.12 (**Figure 1C**), highlighting how the annotation of single-cell populations is affected by selecting a specific marker gene database. These results show that different marker genes databases inevitably lead to inconsistent interpretations of the biological meaning of single-cell data and pinpoint the existence of an open issue with profound consequences for data mining.

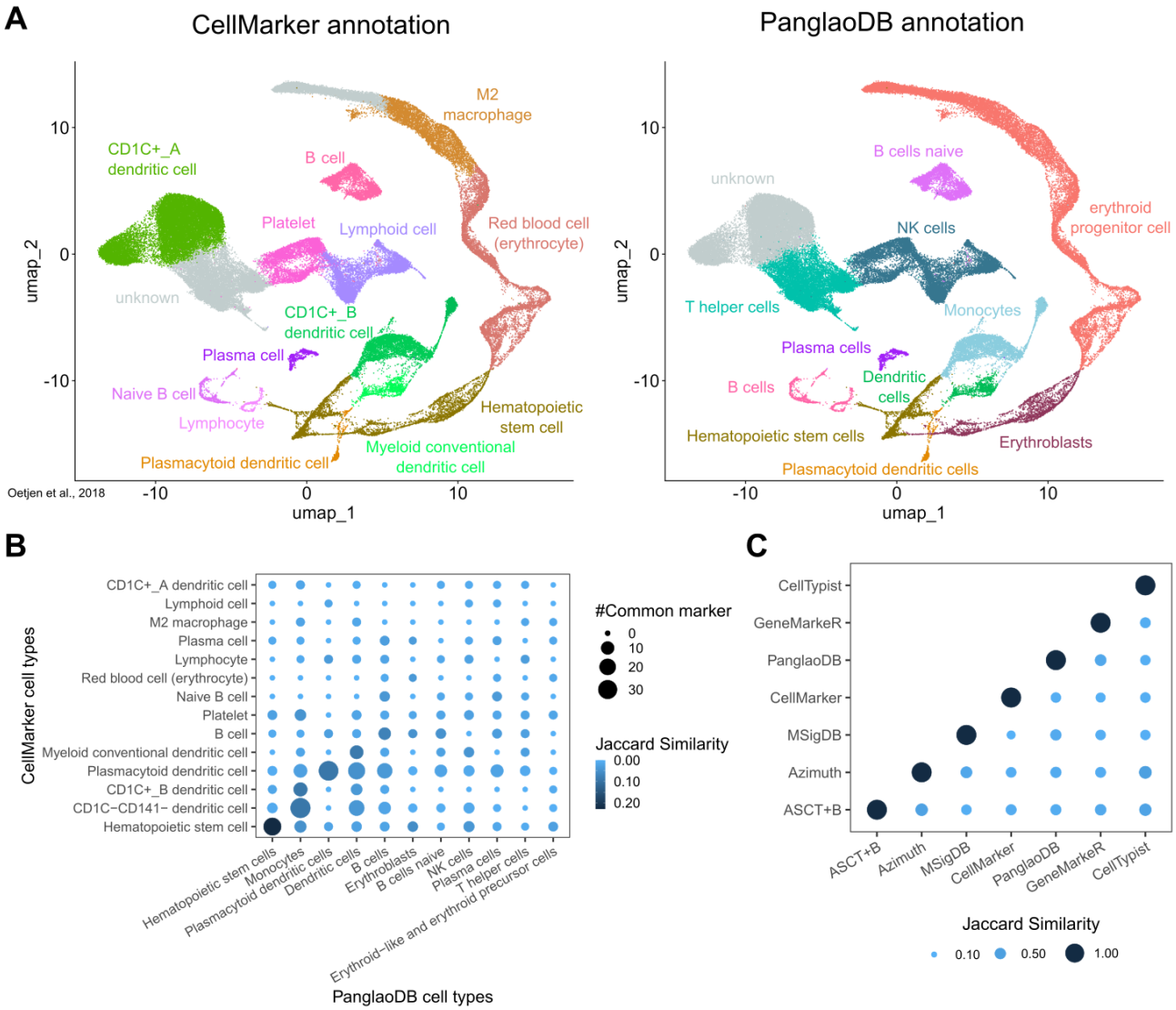

Figure 1. Heterogeneity in annotation sources leads to inconsistent single-cell annotations. *(A) Cell types identification by automatic annotation with ScType in the Oetjen et al., 2018[45] bone marrow dataset, using markers from CellMarker (left) and PanglaoDB (right) as input. (B) Overlap between marker genes from CellMarker (y-axis) and PanglaoDB (x-axis). The dot color represents the Jaccard similarity index and the dot size indicates the number of common markers in each cell type pair. (C) Comparison of cell type markers in published databases. The dot color and size indicate the Jaccard similarity index between each database pair, calculated as the average of common cell types.*

The Cell Marker Accordion: a user-friendly platform for the annotation and interpretation of single-cell populations

To address the need of robust and reproducible identification of hematopoietic cell types in single-cell datasets, we developed the Cell Marker Accordion, comprising a gene marker database, an R shiny web app and an R package for automatically annotating and interpreting single-cell populations.

We built the Accordion database by integrating multiple marker genes database and cell sorting markers sources, distinguishing positive from negative markers (**Figure 2A)**. Label standardization was performed by mapping the initial nomenclature to the Cell Ontology terms. Next, via database

integration we obtained a comprehensive set of hematopoietic cell type specific marker genes for both human and mouse. Importantly, in the Accordion database marker genes are weighted by their specificity, indicating whether a gene is a marker for different cell types, and by their evidence consistency score, measuring the agreement of different annotation sources (see **Methods**).

The user-friendly and interactive Accordion Shiny web interface permits to easily retrieve lists of marker genes associated with input cell types and *vice versa*, starting from a list of candidate genes to obtain the matching cell types (**Figure 2B right**). Hierarchies of hematopoietic cell types can be easily browsed following the Cell Ontology structure in order to obtain the desired level of resolution in the markers. Users can upload their custom sets of genes to either update the repository or obtain the closest associated cell type, with no need for programming skills.

Finally, the Accordion R package allows to automatically annotate healthy and disease-critical cell populations based on the built-in Accordion gene marker database, weighting the markers according to their evidence consistency and specificity score (**Figure 2B left**). The automatic annotation can be easily integrated in a Seurat analysis workflow[29], requiring as input only the count matrix or a Seurat object. Built-in lists of positive and negative cell cycle markers can be used to assign the appropriate cell cycle phase to each cell, or to evaluate quiescence. Any annotation procedure can be easily enhanced by including custom gene lists associated with cell types, specific pathways or signatures of interest. Importantly, the Cell Marker Accordion implements novel options to explore annotation results by inspecting the top marker genes that most significantly determined the ranking of the candidate cell type for each cell, or for each cluster of cells. The distribution of all cell types competing for the same annotation can be evaluated by inspecting their position along the cell ontology tree (**Supplementary Figure 2**).

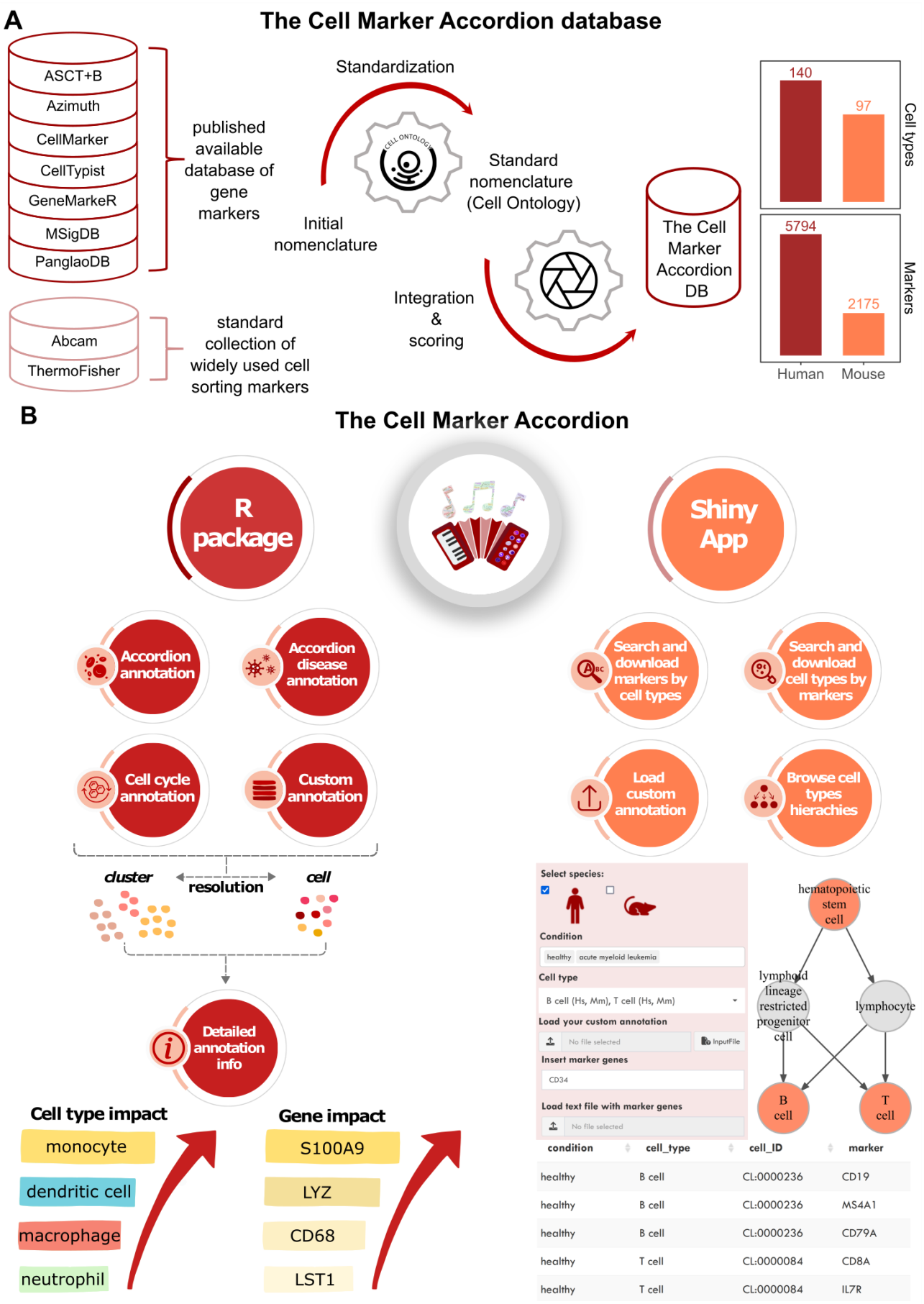
A
The Cell Marker Accordion database
ASCT+B
Azimuth
CellMarker
CellTypist
GeneMarkeR
MSigDB
PanglaoDB
published available database of gene markers
Abcam
ThermoFisher
standard collection of widely used cell sorting markers
Standardization
Initial nomenclature
Standard nomenclature (Cell Ontology)
Integration & scoring
The Cell Marker Accordion DB
140
97
Cell types
5794
2175
Markers
Human
Mouse
B
The Cell Marker Accordion
R package
Shiny App
Accordion annotation
Accordion disease annotation
Cell cycle annotation
Custom annotation
Search and download markers by cell types
Search and download cell types by markers
Load custom annotation
Browse cell types hierachies
cluster
resolution
cell
Detailed annotation info
Cell type impact
monocyte
dendritic cell
macrophage
neutrophil
Gene impact
S100A9
LYZ
CD68
LST1
Select species:
Condition
healthy
acute myeloid leukemia
Cell type
B cell (Hs, Mm), T cell (Hs, Mm)
Load your custom annotation
No file selected
InputFile
Insert marker genes
CD34
Load text file with marker genes
No file selected
hematopoietic stem cell
lymphoid lineage restricted progenitor cell
lymphocyte
B cell
T cell
condition
cell_type
cell_ID
marker
healthy B cell CL:0000236 CD19
healthy B cell CL:0000236 MS4A1
healthy B cell CL:0000236 CD79A
healthy T cell CL:0000084 CD8A
healthy T cell CL:0000084 IL7R

Figure 2. The Cell Marker Accordion: a user-friendly platform for annotating and interpreting single-cell populations*. (A) Workflow for building the Cell Marker Accordion database. The resulting number of cell types and the number of markers for both human and mouse are reported. (B) Overview of the main functionalities of the Cell Marker Accordion R package and Shiny app.*

The Cell Marker Accordion improves the annotation of hematopoietic cell types in complex single-cell multiomics

To validate the Cell Marker Accordion, we undertook a benchmark study to compare its annotation performance against other databases in multiple published single-cell studies, with an increasing degree of complexity and number of annotation challenges (**Figure 3**).

First, we exploited a dataset acquired from fluorescent antibody sorted (FACS) blood cells based on 15 cell surface markers, resulting in 10 different populations separately profiled via single-cell RNA-seq[44] with 94655 total cells (**Figure 3A**). Next, we selected two human bone marrow datasets, obtained via similar multi-omics methods (CITE-seq and Abseq), that simultaneously captured RNA and protein expressions[46,47]. In the first case, 25 barcoded antibodies were used to quantify surface proteins and identify 14 different cell types, in 77534 total cells (**Figure 3B**). The second bone marrow dataset comprised 13159 cells, classified into 24 different cell types according to the expression of 97 barcoded antibodies (**Figure 3C**). In these three datasets we considered surface markers as the ground truth to evaluate and compare annotation results. Additionally, we included a multi-study single-cell RNA-seq dataset with ~500,000 cells, obtained from 25 diverse cell types of the human immune system (**Figure 3D**). For our purpose, we considered a subset of the dataset and analyzed 149204 cells, for which a manually expert-based cell types annotation was provided and used as the ground truth (https://www.ebi.ac.uk/gxa/sc/experiments/E-HCAD-4).

To test the accuracy of the Cell Marker Accordion we compared its cell type assignment and annotation against those obtained with marker genes from individual sources [24–30,] . Performance of each annotation was assessed using two metrics: the percentage of cells correctly annotated with respect to the ground truth, and the corresponding F1 scores (see **Methods**). Notably, in all cases the Cell Marker Accordion significantly showed improved accuracy, with an average increase of approximately 10% in the number of correctly identified cells and in F1 scores with respect to any of the single sources marker sets (**Figure 3, right panels, Supplementary Figure 3** for the results in each cell type). Together, these benchmarking results highlight the Cell Marker Accordion's utility as a novel tool to obtain a more robust, consistent and highly interpretable annotation of hematopoietic populations in single-cell data.

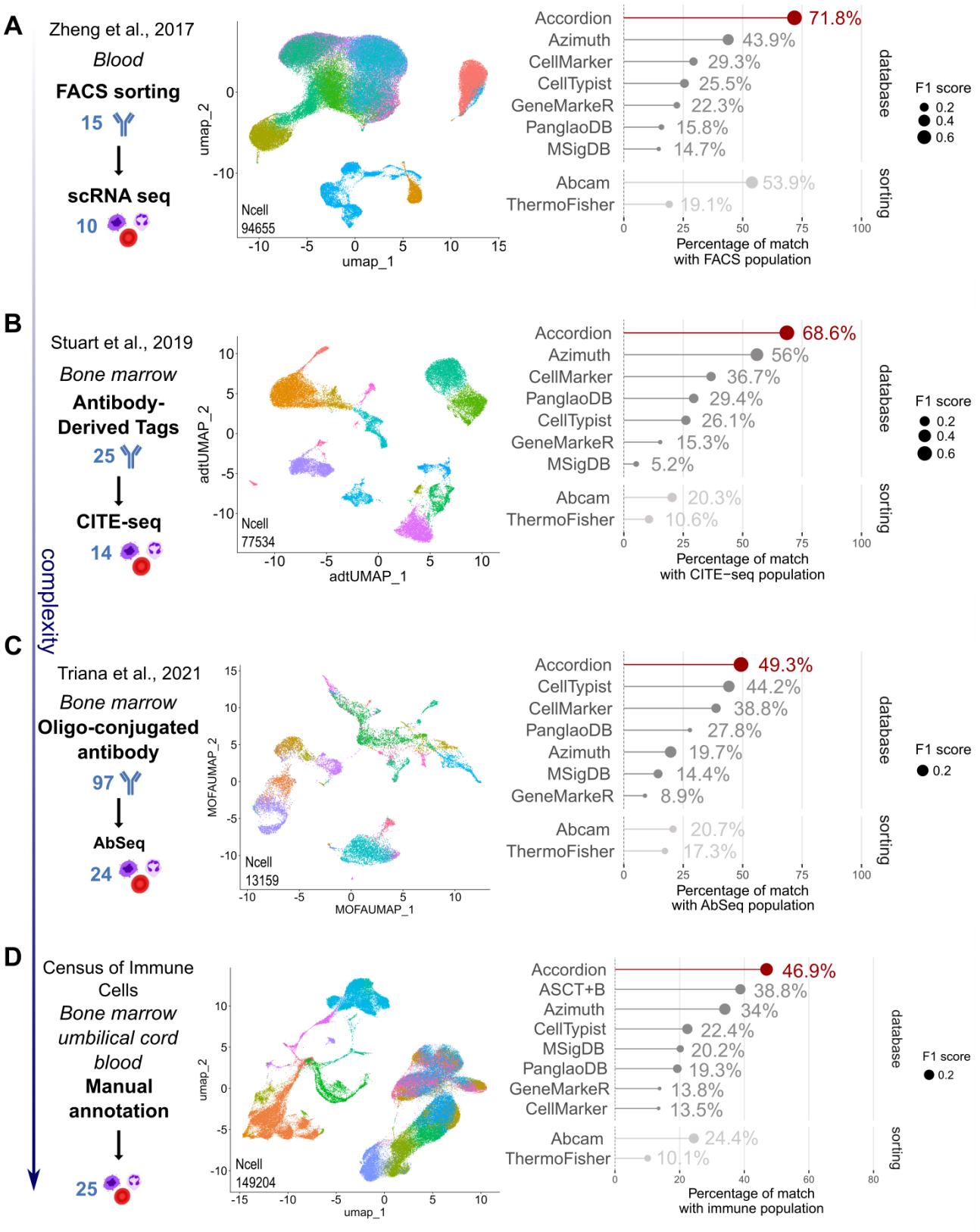

Figure 3. The Cell Marker Accordion improves the annotation of hematopoietic cell types in complex single-cell multiomics. *(A-D) Annotation with the Cell Marker Accordion of single-cell datasets with increasing complexity (left column) and performance comparison with other annotation resources, including databases and collections of sorting markers (right column) (A) Dataset of PBMC FACS sorted cells separately profiled with single-cell RNA-seq. 15 surface antibodies were*

used to sort 10 different cell types, used as the ground truth. The Accordion Annotation Performance, measured as the percentage of cells corresponding to the ground truth and corresponding F1 score (dot size), is compared against other resources (B) Human bone marrow dataset obtained with CITE-seq multi-modal approach (25 barcoded antibodies were used to quantify surface proteins and identify 14 different cell types, considered as the ground truth). (C) Human bone marrow dataset obtained with Ab-seq multi-modal approach (97 barcoded antibodies were used to quantify surface proteins and identify 24 different cell types, considered as the ground truth). (D) Single-cell RNA-seq dataset of human cells from bone marrow and umbilical cord blood. Expert-based manual annotation identified 25 different cell types, considered as the ground truth.

The Cell Marker Accordion identifies disease-critical cells in aberrant hematopoiesis

Blood cancers, like many other types of tumors, are populated by disease-critical cells, characterized by altered states and aberrant gene expression and play a central role in disease progression and treatment response[31]. The persistence of a selective subset of malignant cells has been considered the underlying cause of the high relapse rates commonly observed in patients with hematologic malignancies[48–50]. Identifying and characterizing disease-critical cells in cancer patients is pivotal to improve diagnosis towards interceptive medicine[8], to understand pathogenesis and therapy resistance mechanisms, and to develop novel therapies able to specifically target and eradicate cancer initiating cells while minimizing adverse effects on healthy cells. To expand the Cell Marker Accordion to the analysis of hematologic cancers, in addition to the "healthy" collection we discussed so far, we created a "disease" section. We collected and integrated marker genes associated with disease-critical cells found in the most common blood cancer types (**Figure 4A**). To obtain a standardized and consistent vocabulary of cancer types we mapped disease terms to the Disease Ontology (https://disease-ontology.org/).

Notably, we collected more than 50 markers associated with acute myeloid leukemia (AML), which is one of the most common types of acute leukemia in adults[51]. A major challenge in the treatment of acute myeloid leukemia is the survival of a few therapy-resistant cells. These cells, known as leukemia stem or initiating cells (LSCs or LICs), which are one of the the key factors contributing to disease progression and relapse[31,48,52,53]. To show the potential of the Cell Marker Accordion in identifying disease-critical cells in human blood cancers, we analyzed a published scRNA-seq dataset of CD34+ bone marrow cells from 5 healthy controls and 14 acute myeloid leukemia patients[54]. First, healthy cell types were annotated (**Figure 4B**). Next, by exploiting the leukemia stem cell marker genes, the Cell Marker Accordion was able to assign a LSC score for each cell in healthy donors and AML patients (**Figure 4C**). Notably, we observed an accumulation of malignant cells, especially in progenitor and monocyte populations (**Figure 4C**). Our accuracy is proven by the fact that we obtained an overall significant increase in the LSC score in AML patients with respect to healthy controls. (**Figure 4D**). To extend our analysis to the context of therapies, we took advantage of another published scRNA-seq dataset of human bone marrow, with sequential samples at diagnosis and relapse from patients treated with the BCL-2 inhibitor venetoclax[31,55]. As for the previous dataset, we first annotated healthy cell types (**Figure 4E**). LSCs were identified at diagnosis and relapse and consistent with published results, we found cells with high LSC scores in the progenitor and monocyte populations at diagnosis (**Figure 4F**). Comparing the LSC score distribution between diagnosis and relapse, we observed a significant overall increase of malignant stem cells at diagnosis compared to relapse (**Figure 4G**). However the progenitor population at relapse doesn't contribute to the LSC score, implying that venetoclax-based treatment is able to target and eradicate most of the LSC with progenitor features. Instead, LSCs with monocytic phenotype persist after therapy, confirming, as previously proposed, that the mechanism of

resistance to venetoclax resides in a monocytic LSC population. These results suggest that the presence of malignant stem cell heterogeneity play a significant role in treatment response and disease progression[31,55]. To further characterize the properties of leukemia stem cells in AML patients, we extracted the core genes that define these malignant cells and drive their identification by the Cell Marker Accordion. We indeed were able to extract altered gene signatures associated with LSCs specific to either the progenitor and monocyte populations.(**Figure 4H**).

We further showed the potential of the Cell Marker Accordion in identifying disease-critical cells by for myeloma plasma cells in single-cell datasets from patients with multiple myeloma (MM) (**Supplementary Figure 4**). We exploited a published scRNA-seq dataset of bone marrow from 11 healthy controls and 12 multiple myeloma patients[56]. After annotation of healthy cell types, (**Supplementary Figure 4A**) we identified malignant plasma cells (**Supplementary Figure 4B**), with a significantly higher score in MM patients (**Supplementary Figure 4C**). These cells were clustered in patient-specific groups, suggesting distinct clonotypes (**Supplementary Figure 4D-E**). Also in this case, we were able to extract genes with the highest impact in defining myeloma plasma cells (**Supplementary Figure 4F**).

These results provide robust evidence about the potential of the Cell Marker Accordion to identify malignant cells with deviant states with respect to their physiological counterparts and to investigate disease mechanisms by extracting altered gene signatures in the quest for biomarker discovery.

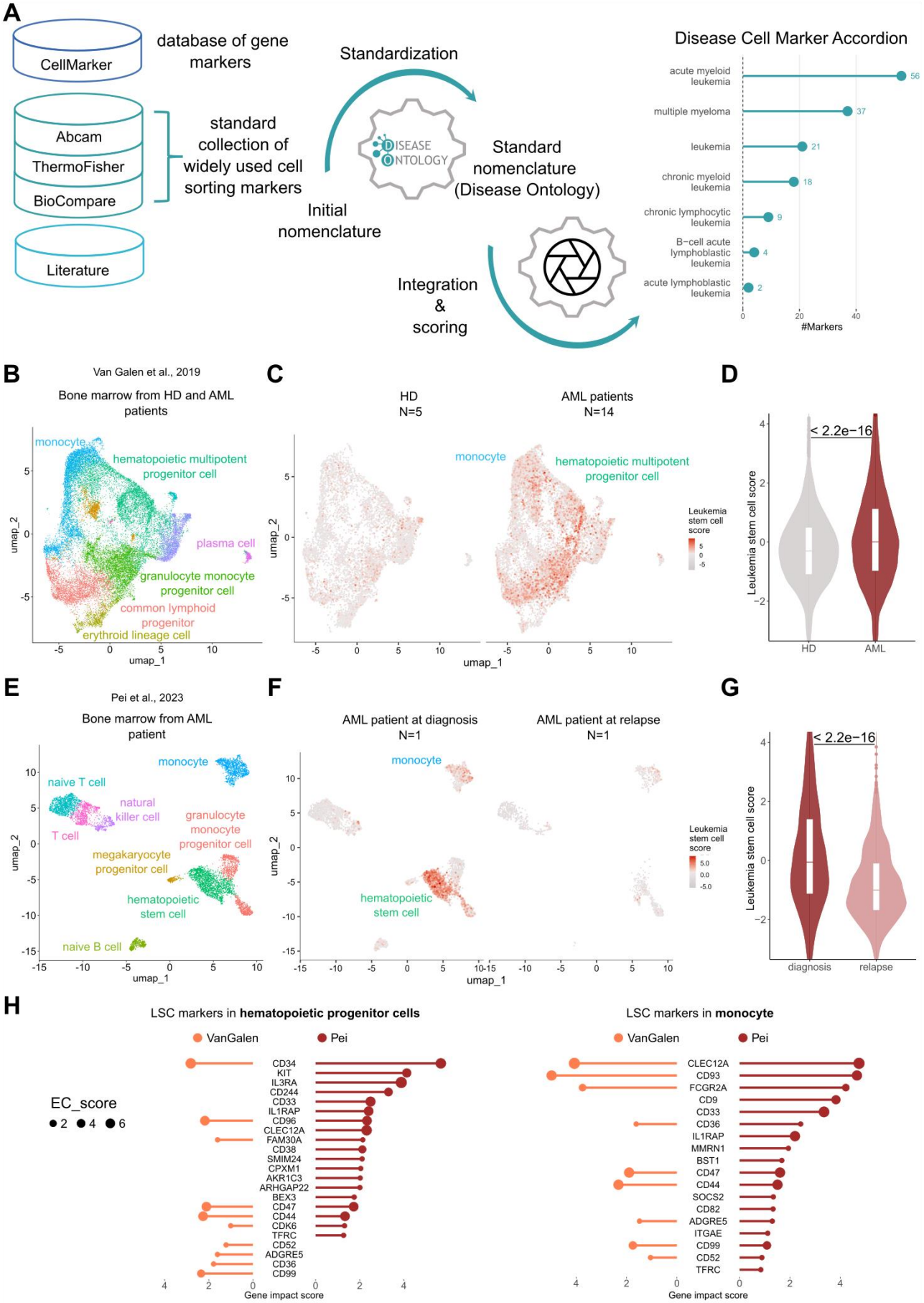
A
CellMarker
database of gene markers
Abcam
ThermoFisher
BioCompare
standard collection of widely used cell sorting markers
Literature
Standardization
Initial nomenclature
Standard nomenclature (Disease Ontology)
Integration & scoring
Disease Cell Marker Accordion
acute myeloid leukemia
multiple myeloma
leukemia
chronic myeloid leukemia
chronic lymphocytic leukemia
B−cell acute lymphoblastic leukemia
acute lymphoblastic leukemia
#Markers
B
Van Galen et al., 2019
Bone marrow from HD and AML patients
monocyte
hematopoietic multipotent progenitor cell
plasma cell
granulocyte monocyte progenitor cell
common lymphoid progenitor
erythroid lineage cell
umap_1
umap_2
C
HD N=5
AML patients N=14
Leukemia stem cell score
D
< 2.2e−16
Leukemia stem cell score
HD
AML
E
Pei et al., 2023
Bone marrow from AML patient
naive T cell
natural killer cell
T cell
megakaryocyte progenitor cell
hematopoietic stem cell
naive B cell
F
AML patient at diagnosis N=1
AML patient at relapse N=1
G
diagnosis
relapse
H
LSC markers in hematopoietic progenitor cells
LSC markers in monocyte
VanGalen
Pei
EC_score
Gene impact score

Figure 4. The Cell Marker Accordion identifies leukemia stem cell subtypes in acute myeloid leukemia patients. *(A) Workflow for building the Cell Marker Accordion Disease database. The resulting number of disease-critical cells markers associated with different hematologic malignancies are reported. (B) Accordion annotation of human bone marrow cells from healthy donors (HD) and acute myeloid leukemia (AML) patients[54]. (C) Identification of leukemia stem cells (LSCs) with the Accordion Disease. Cells are colored according to the LSC score in HD and AML patients. (D) Distribution of LSC score in the HDs and AML patients. A significant increase is observed in AML patients with respect to HDs. (E) Accordion annotation of human bone marrow cells from healthy donors (HD) at diagnosis and relapse time point[57]. (F) Identification of leukemia stem cells (LSCs) with the Accordion Disease. Cells are colored according to the LSC score in AML patients at diagnosis and relapse time points. (G) Distribution of LSC score in the AML patient at diagnosis and after venetoclax treatment. A significant increase is observed at diagnosis with respect to the relapse time point. (H) Comparison of marker genes with the highest impact in defining leukemia stem cells in the two leukemia datasets, among progenitor cells and monocytes respectively.*

The Cell Marker Accordion identifies altered cell type composition in patients with splicing factor mutant myelodysplastic syndromes

Mutations in splicing factor (SF) genes are prevalent in approximately 50% of patients with Myelodysplasia (MDS) and Acute Myeloid Leukemia (AML)[58–60]. These mutations, especially the U2AF1 mutations, are linked to a high risk of AML transformation and to decreased survival rates[61–67]. To explore the molecular mechanisms and biological implications that drive the clonal advantage of SF mutant cells over their wildtype counterparts, we conducted single-cell RNA sequencing on CD34+ cells from MDS patients, either without SF mutations (n=5) or with the U2AF1 S34F mutation (n=3). From a total of 62496 high quality cells (see **Methods**), we performed cell type identification with the Cell Marker Accordion (**Figure 5A**). The majority of resulting cell types were related to blood progenitor cells, in line with the CD34+ cell sorting. To investigate the impact of the U2AF1 S34F splicing factor mutation, we compared cell type composition between U2AF1 WT and mutant patients (**Figure 5B**). Interestingly, we observed an increase in hematopoietic multipotent progenitors, common lymphoid progenitors, monocytes and plasmacytoid dendritic cells, with, in parallel, a decrease in granulocyte-monocyte progenitors, megakaryocyte-erythroid progenitors, megakaryocyte progenitors, mast cells and erythroid lineage cells (**Figure 5B**). These results are consistent with the lineage-specific alterations induced by U2AF1 S34F, with impaired erythroid and granulomonocytic differentiation[39,68]. Moreover, these results suggest that the U2AF1 S34F mutation drives a monocytic phenotype associated with poor clinical outcomes[57].

By single-cell mutation calling on reads mapping to the U2AF1 locus (see **Methods**), we classified each cell from U2AF1 S34F patient samples as either WT or S34F (**Figure 5C**). Notably, we observed that different cell types were characterized by various degrees of mutant cells, ranging from 5% to 32% (**Figure 5D**). This data confirm a myelo-monocytic shift with reduction in megakaryocyte and erythroid lineage priming within hematopoietic stem and progenitor cells from patients with S34F mutant MDS (**Figure 5C**); accumulation of mutant cells within the megakaryocytic and erythroid lineage suggests a differentiation defect conferred specifically by the S34F mutation (**Figure 5D**).

Overall, these results demonstrate that the Cell Marker Accordion can be effectively used to identify and dissect cell-type variations driven by pathologically relevant mutations.

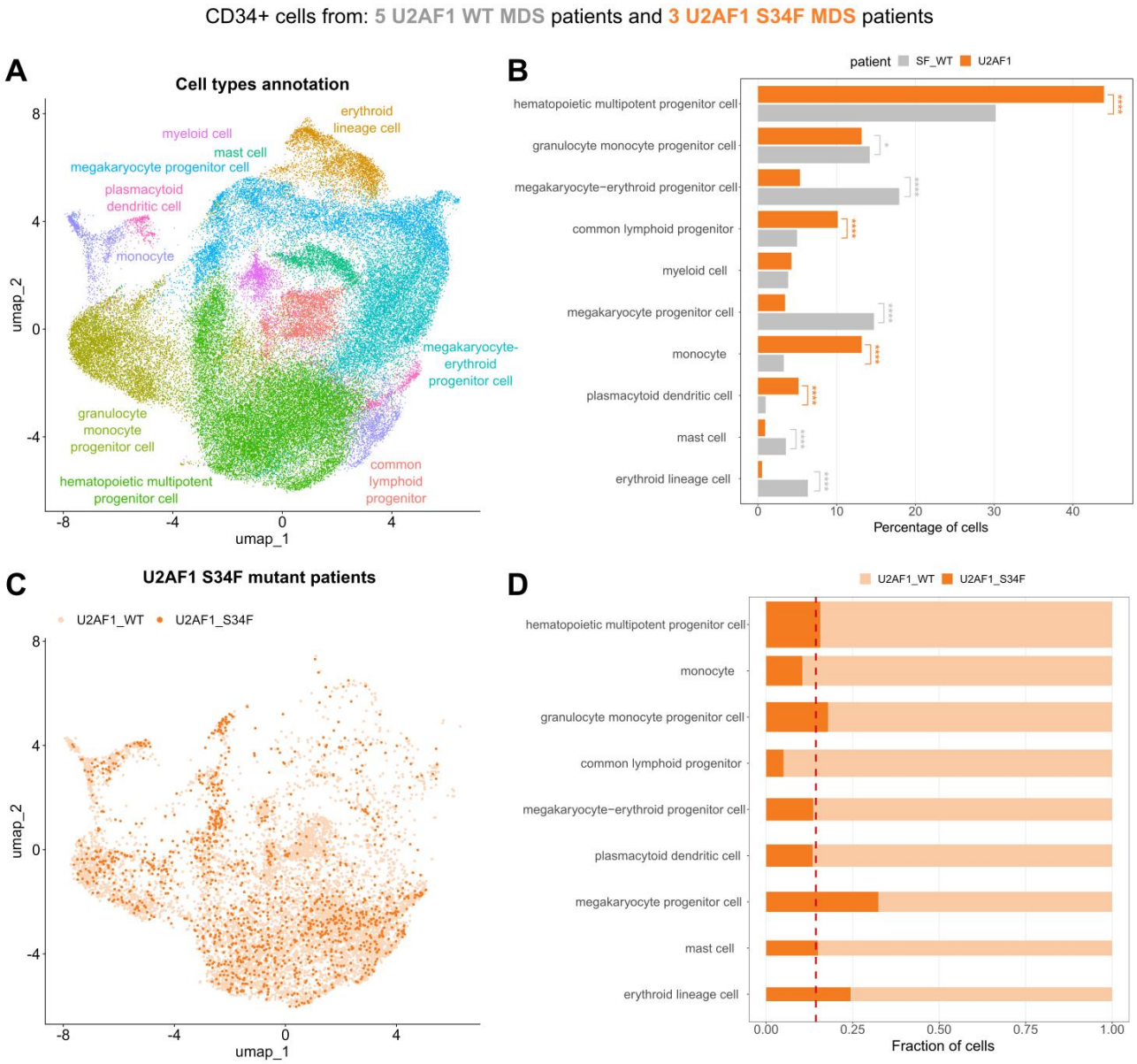

Figure 5. The Cell Marker Accordion identifies cell type alterations in splicing factor mutant cells from patients with myelodysplastic syndromes. *(A) Accordion cell types annotation of MDS patients with and without U2AF1 S34F mutation. (B) Changes in the abundance of hematopoietic cell types among conditions. Orange bars represent patients with U2AF1 S34F mutations and gray bars represent patients without splicing factor mutations. (C) Color-code representation of U2AF1 WT and S34F cells in S34F mutant patients. (D) Fraction of mutant (dark orange) and WT cells (light orange) in each cell type. The width of the bar is proportional to the average number of cells in each population. The dashed line represents the average number of mutant cells across all cell types in U2AF1 S34F patients.*

The Cell Marker Accordion identifies activation of innate immunity pathways in mouse bone marrow

N6-methyladenosine (m6A) is the most abundant eukaryotic internal mRNA modification and exerts significant influence in RNA biology[69–71]. This modification plays important roles in normal hematopoiesis and alterations in m6A metabolism are strongly associated with acute myeloid

leukemia pathogenesis, characterized by the overexpression of the m6A methyltransferase METTL3[72,73]. For this reason, pharmacological inhibition of METTL3 has been proposed as a therapeutic strategy to treat leukemias[74].

To characterize the effect of m6A modulation on hematopoietic populations, we applied the Cell Marker Accordion to two murine single-cell datasets obtained from the bone marrow of Mettl3 conditional knockout mice[75] (**Figure 6 A-D**) and from mice upon pharmacological inhibition of METTL3 with STM2457[76] (**Figure 6 E-H**). We performed cell type annotation and compared cell type compositions (**Figure 6B** and **6F**). Interestingly, in both datasets, we observed an increase in hematopoietic stem cells and megakaryocyte progenitors, together with a decrease in erythroid progenitors upon Mettl3 deletion or inhibition (**Figure 6C** and **6G**). These observations are in line with the original publications and with results obtained in previous studies[40]. Next, we performed cell cycle annotation based on lists of cell-type phase-specific positive and negative markers (**Figure 6B** and **6F, right panels**). With this procedure, we were able to detect cycling changes in specific hematopoietic cell types, in particular an increase of G0 cells among hematopoietic stem cells and megakaryocyte progenitors (**Figure 6D** and **6H**).

Two recent studies by us and others turned the spotlight on aberrant activation of innate immune pathways as a consequence of response to the deletion of the m6A methyltransferase Mettl3 or pharmacological inhibition, mediated by the formation of aberrant endogenous double stranded RNAs[40,41]. To explore the impact of the knockout and the inhibition of Mettl3 on immunity in single cell datasets, the Cell Marker Accordion computed an “innate immune response” score based on the activation of genes associated with this signature (**Supplementary File 1**). Notably, both in the case of Mettl3 KO and drug-induced Mettl3 inhibition, we obtained a significant increase in innate immune response score with respect to the control condition (**Figure 6I**). In addition, by extracting genes that mostly influence the immune response score, we found a subset that exhibits consistent activation in both murine models, as well as sets of genes that are specifically activated in response to either the knockout or the pharmacological inhibition of Mettl3 in murine hematopoietic stem and progenitor cells (**Figure 6L**).

Overall, these results demonstrate that the Cell Marker Accordion can be effectively used to characterize pathologically relevant pathways in disease or pharmacological treatment models.

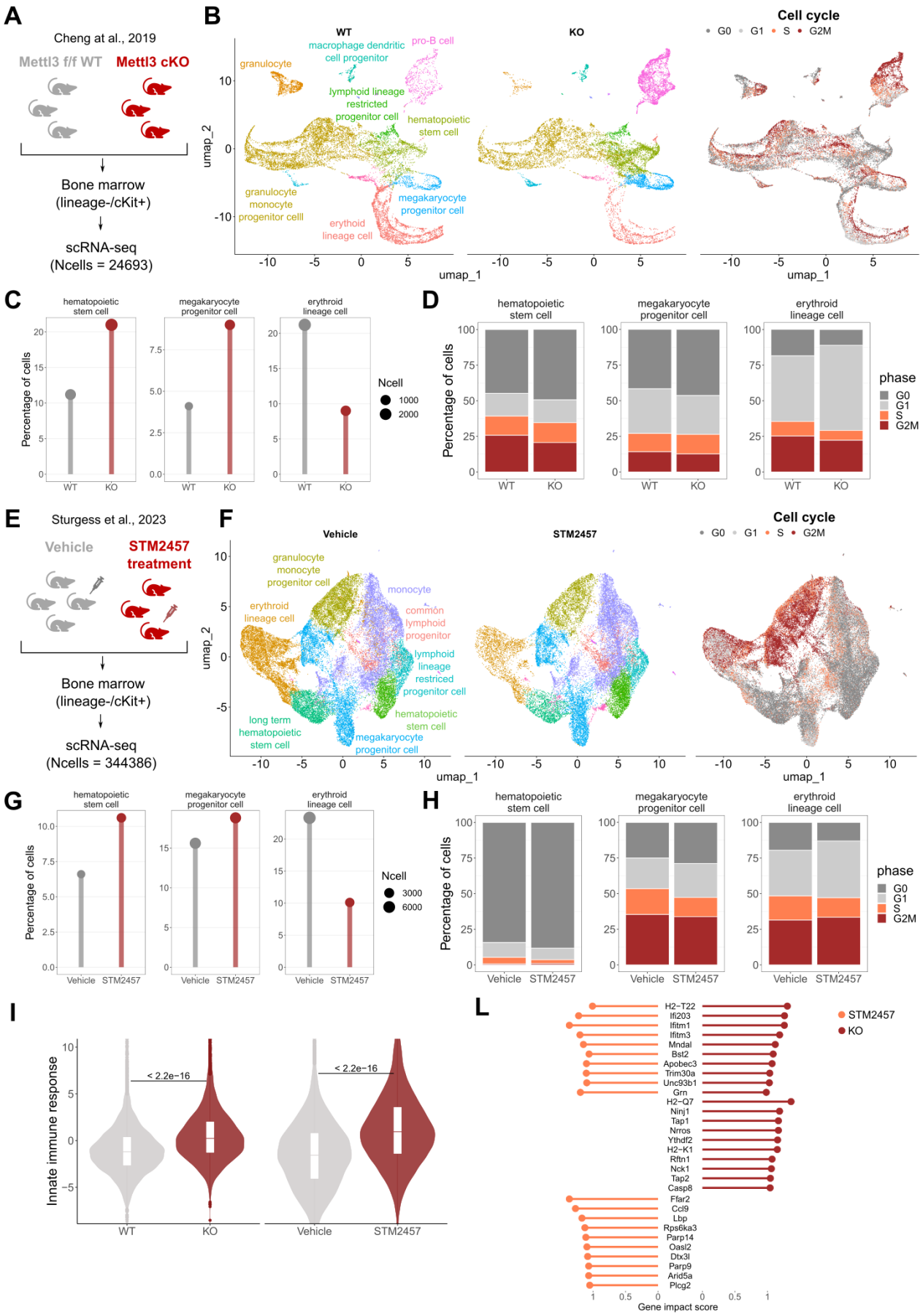
A
Cheng at al., 2019
Mettl3 f/f WT
Mettl3 cKO
Bone marrow
(lineage-/cKit+)
scRNA-seq
(Ncells = 24693)
B
WT
KO
Cell cycle
G0 G1 S G2M
granulocyte
macrophage dendritic cell progenitor
pro-B cell
lymphoid lineage restricted progenitor cell
hematopoietic stem cell
granulocyte monocyte progenitor celll
megakaryocyte progenitor cell
erythoid lineage cell
umap_2
umap_1
C
hematopoietic stem cell
megakaryocyte progenitor cell
erythroid lineage cell
Percentage of cells
Ncell
1000
2000
WT
KO
D
phase
G0
G1
S
G2M
E
Sturgess et al., 2023
Vehicle
STM2457 treatment
Bone marrow
(lineage-/cKit+)
scRNA-seq
(Ncells = 344386)
F
Vehicle
STM2457
Cell cycle
G0 G1 S G2M
granulocyte monocyte progenitor cell
monocyte
erythroid lineage cell
common lymphoid progenitor
lymphoid lineage restriced progenitor cell
hematopoietic stem cell
long term hematopoietic stem cell
megakaryocyte progenitor cell
umap_2
umap_1
G
Ncell
3000
6000
Vehicle
STM2457
H
phase
G0
G1
S
G2M
I
Innate immune response
< 2.2e−16
WT
KO
Vehicle
STM2457
L
STM2457
KO
H2−T22
Ifi203
Ifitm1
Ifitm3
Mndal
Bst2
Apobec3
Trim30a
Unc93b1
Grn
H2−Q7
Ninj1
Tap1
Nrros
Ythdf2
H2−K1
Rftn1
Nck1
Tap2
Casp8
Ffar2
Ccl9
Lbp
Rps6ka3
Parp14
Oasl2
Dtx3l
Parp9
Arid5a
Plcg2
Gene impact score

Figure 6. The Cell Marker Accordion identifies activation of innate immunity pathways in mice bone marrow. *(A) Schematic diagram of the single-cell experimental design of Cheng et al., 2019*[75] *dataset, comparing bone marrow from Mettl3 KO and WT mice. (B) Accordion cell types annotation of WT and KO mice and identification of cell cycle phase, based on lists of phase-specific markers. (C) Changes in the abundance of specific hematopoietic cell types upon Mettl3 KO. The increase in hematopoietic stem cells and megakaryocytes, with the parallel decrease of erythroid progenitors, is consistent with literature. (D) Cell type specific variations in cell cycle between WT and Mettl3 KO bone marrows. (E) Schematic diagram of the Mettl3 inhibition experimental design of Sturgess et al., 2023*[76] *dataset. (F) Accordion cell types annotation of mice treated with STM2457 METTL3 inhibitor and vehicle treated mice and identification of cell cycle phase. (G) Changes in the abundance of specific cell types between STM2457 and vehicle mice, consistent with changes observed in panel C. (H) Cell type specific variations of cell cycle between STM2457 and vehicle mice. (I) Significant increase of the "innate immune response" signature in Mettl3 KO and STM2457 treated cells, consistent with innate immunity activation observed in Gao et al., 2020*[40]*. (L) Genes involved in "innate immune response" pathways and showing the highest impact score in Mettl3 KO or STM2457 treated cells.*

Discussion

Accurate identification of cell types and states within heterogeneous and complex tissues is a prerequisite for comprehensive exploration and interpretation of single-cell data to provide biological insights, yet it is a challenging step in the single-cell analysis workflow[10].

Here we present the Cell Marker Accordion, a user-friendly platform encompassing an interactive R Shiny web application and an R package designed to automatically identify and interpret single-cell populations in both physiological and pathological conditions. With respect to the majority of existing computational methods[17–22], the Cell Marker Accordion not only provides the users with an accurate annotation of hematopoietic cell types, but it is also able to detect disease-critical cells and pinpoint altered pathways in aberrant conditions, including cell cycle and quiescence analysis. The Cell Marker Accordion combines both positive and negative markers, providing a more specific and unambiguous annotation. Hematopoietic cell types can be easily browsed following the Cell Ontology hierarchy, to obtain the desired level of resolution. With respect to existing tools, the Accordion weights markers not only on their specificity but also on their consistency among resources, allowing a more robust cell type identification. Moreover, the Accordion allows the inclusion of customized annotations, by incorporating any weighted signature of interest. The biological interpretation of results is straightforward and especially transparent since the Accordion provides detailed information and graphics on genes, cell types or pathways that exert the most significant influence on annotation outcomes.

We validated the accuracy of the Cell Marker Accordion on peripheral blood and bone marrow cell populations considering surface markers and expert-based annotations as the reference[44,46,47]. Across all these datasets, the Cell Marker Accordion notably enhanced the precision in identifying cell types compared to any single-source database available[17–22]. With the increase of sample complexity and number of antibodies in the panel, we noticed that the overall classification performance decreased. This can be attributed to the use of single-cell measurements as the ground truth in the benchmark: data are affected by overall noise, and certain sub-populations, defined by a limited number of surface markers with low transcript expression levels, are generally challenging to detect. Nevertheless, our approach and output can be leveraged to address technical imprecisions,

possible label inaccuracies or hidden heterogeneity in cell clusters, for example by checking the consistency of all cell types competing for the same annotation along the Cell Ontology hierarchy.

Accurate identification of cell types is a fundamental requirement for investigating hematologic disorders. Blood cancers are populated by disease-critical cells, characterized by aberrant gene expression, which play a central role in disease progression and treatment response[31]. Current tools focus mostly on physiological cell types or attempt to distinguish malignant vs non-malignant cells through SNV calling[43,77], lacking specific characterization of abnormal cell states by expression. To fill this gap, the Cell Marker Accordion includes weighted collections of marker genes associated with disease-critical cells in the most common blood cancer types, mostly linked to acute myeloid leukemia (AML), multiple myeloma (MM) and lymphoma. In AML and MM patients, the persistence of therapy-resistant leukemia stem cells and malignant plasma cells respectively, inevitably increases the risk of relapse and reduces overall treatment effectiveness[31,48,52,53]. By exploiting scRNA-seq datasets from AML and MM patients, we showed that the Cell Marker Accordion effectively identifies aberrant cell types and extracts altered gene signatures. Identifying and characterizing disease-critical cells is pivotal for improving diagnosis and interceptive medicine[8], for understanding pathogenesis and therapy resistance mechanisms, for identifying biomarkers and for developing effective therapies minimizing adverse effects on healthy cells.

Besides the identification of disease-critical cells, the Cell Marker Accordion can be applied to the study and characterization of pathological processes. We demonstrated this in the context of myelodysplastic syndromes (MDS), where mutations in splicing factors (SF) genes such as U2AF1 are prevalent in approximately 50% of patients[58–60] and linked to decreased survival rates[61–67]. By applying the Cell Marker Accordion to single-cell data that we generated from bone marrow of a small cohort of MDS patients, we revealed skewing in the hematopoietic lineages in patients with U2AF1 S34F mutation. In particular, we observed impaired erythroid and granulomonocytic differentiation[39,68], pointing out the impact of pathologically relevant splicing factor mutation on ineffective hematopoiesis and clonal advantage. This approach could be extended by tracking additional MDS mutations to determine their effect at various stages in differentiation.

Finally, we used the Cell Marker Accordion to dissect the effects of m6A RNA modification[69–71] and modulation of the METTL3 methyltransferase on hematopoiesis in murine models. Alterations in m6A have been strongly associated with acute myeloid leukemia pathogenesis[72,73], and pharmacological inhibition of METTL3 has been proposed as a therapeutic strategy[74]. The Cell Marker Accordion identified cell cycle changes and activation of immune response pathways in specific hematopoietic cell types, consistent with the formation of aberrant endogenous dsRNAs upon METTL3 depletion[40,41]. We extracted gene signatures activated in response to either the knockout or drug-mediated inhibition of METTL3 or both to demonstrate that the Cell Marker Accordion can be utilized to characterize pathologically relevant pathways in disease or pharmacological treatment models.

Fast-forward advances in single-cell techniques are expected to provide increasingly accurate and comprehensive measurements of single-cell populations. The Cell Marker Accordion is designed to accommodate updates and new sources of information, aiming for a more precise and refined cell type identification in diverse contexts and across different types of data. Possible extensions of the Cell Marker Accordion include single-cell approaches profiling chromatin accessibility[78] and spatially resolved data, by sequencing or imaging approaches. In this context, the Accordion could be used to identify immune cells in spatially resolved data obtained from tissue slices, such as lymph nodes or bone marrow, to dissect architectural heterogeneity and disease microenvironment[79,80].

In conclusion, the Cell Marker Accordion is a user-friendly and flexible tool that can be exploited to improve the annotation and interpretation of hematopoietic populations in single-cell datasets focused on the study of disease.

Acknowledgements

We thank Guilin Wang, Christopher Castaldi and the Yale Center for Genome Analysis for scRNA-seq guidance. We thank Lesley Devine and the Yale Flow Cytometry Facility for guidance in cell sorting. We thank all our patients and all clinical staff for their help with patient recruitment. This study was funded by AIRC under MFAG 2020 (ID. 24883 project) to T.T. S.H. was supported by NIH/NIDDK R01DK124788, NIH/NCI R01CA266604, NIH/NCI R01CA222518, NIH/NCI R01CA253981, The Frederick A. Deluca Foundation and the Edward P. Evans Foundation. G.B was supported by the American Society of Hematology Scholar Award and by the Edward P. Evans Foundation EvansMDS Young Investigator Award.

Authorship

Conceptualization: E.B., G.B., S.H. and T.T.; methodology: E.B., G.B., F.L., Z.I., C.R., G.T., G.V., S.H. and T.T.; investigation: E.B., G.B., K.R.A, J.V., T.S., S.H. and T.T.; formal analysis: E.B., G.B., F.L., Z.I., C.R., G.T., S.H., T.T.; resources: S.H, T.T.; validation: E.B., G.B., K.R.A, J.V, T.S., S.H.,T.T; writing: E.B., G.B., S.H. and T.T.; visualization: E.B., G.B., S.H. and T.T.; supervision: S.H. and T.T.; project administration: S.H. and T.T.; funding acquisition: S.H. and T.T.

References

1. Svensson V, Vento-Tormo R, Teichmann SA. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc*. 2018;13(4):599–604.
2. Pellin D, Loperfido M, Baricordi C, et al. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nature Communications 2019 10:1*. 2019;10(1):1–15.
3. Wilson NK, Göttgens B. Single-Cell Sequencing in Normal and Malignant Hematopoiesis. *Hemasphere*. 2018;2(2):e34.
4. Zhang P, Li X, Pan C, et al. Single-cell RNA sequencing to track novel perspectives in HSC heterogeneity. *Stem Cell Research & Therapy 2022 13:1*. 2022;13(1):1–15.
5. Monga I, Kaur K, Dhanda SK. Revisiting hematopoiesis: applications of the bulk and single-cell transcriptomics dissecting transcriptional heterogeneity in hematopoietic stem cells. *Brief Funct Genomics*. 2022;21(3):159–176.
6. Lei KF, Ho YC, Huang CH, Huang CH, Pai PC. Characterization of stem cell-like property in cancer cells based on single-cell impedance measurement in a microfluidic platform. *Talanta*. 2021;229:.
7. Baslan T, Hicks J. Unravelling biology and shifting paradigms in cancer with single-cell sequencing. *Nat Rev Cancer*. 2017;17(9):557–569.
8. Rajewsky N, Almouzni G, Gorski SA, et al. LifeTime and improving European healthcare through cell-based interceptive medicine. *Nature 2020 587:7834*. 2020;587(7834):377–386.

9. Abdelaal T, Michielsen L, Cats D, et al. A comparison of automatic cell identification methods for single-cell RNA sequencing data. *Genome Biol*. 2019;20(1):1–19.

10. Pasquini G, Rojo Arias JE, Schäfer P, Busskamp V. Automated methods for cell type annotation on scRNA-seq data. *Comput Struct Biotechnol J*. 2021;19:961–969.

11. Ekiz HA, Conley CJ, Stephens WZ, O'Connell RM. CIPR: a web-based R/shiny app and R package to annotate cell clusters in single cell RNA sequencing experiments. *BMC Bioinformatics*. 2020;21(1):191.

12. Riemondy KA, Fu R, Gillen AE, et al. clustifyr: An R package for automated single-cell RNA sequencing cluster classification. *F1000Res*. 2020;9:.

13. Kiselev VY, Yiu A, Hemberg M. scmap: projection of single-cell RNA-seq data across data sets. *Nat Methods*. 2018;15(5):359–362.

14. Aran D, Looney AP, Liu L, et al. Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nature Immunology 2019 20:2*. 2019;20(2):163–172.

15. Clarke ZA, Andrews TS, Atif J, et al. Tutorial: guidelines for annotating single-cell transcriptomic maps using automated and manual methods. *Nat Protoc*. 2021;16(6):2749–2764.

16. Wang X, He Y, Zhang Q, Ren X, Zhang Z. Direct Comparative Analyses of 10X Genomics Chromium and Smart-seq2. *Genomics Proteomics Bioinformatics*. 2021;19(2):253–266.

17. Pliner HA, Shendure J, Trapnell C. Supervised classification enables rapid annotation of cell atlases. *Nat Methods*. 2019;16(10):983–986.

18. Shao X, Liao J, Lu X, et al. scCATCH: Automatic Annotation on Cell Types of Clusters from Single-Cell RNA Sequencing Data. *iScience*. 2020;23(3):.

19. Zhang AW, O'Flanagan C, Chavez EA, et al. Probabilistic cell-type assignment of single-cell RNA-seq for tumor microenvironment profiling. *Nat Methods*. 2019;16(10):1007–1015.

20. Wei Z, Zhang S. CALLR: a semi-supervised cell-type annotation method for single-cell RNA sequencing data. *Bioinformatics*. 2021;37(Supplement_1):i51–i58.

21. Guo H, Li J. scSorter: assigning cells to known cell types according to marker genes. *Genome Biol*. 2021;22(1):1–18.

22. Chen Y, Zhang S. Automatic Cell Type Annotation Using Marker Genes for Single-Cell RNA Sequencing Data. *Biomolecules*. 2022;12(10):.

23. Zhang Z, Luo D, Zhong X, et al. SCINA: A Semi-Supervised Subtyping Algorithm of Single Cells and Bulk Samples. *Genes (Basel)*. 2019;10:531.

24. Zhang X, Lan Y, Xu J, et al. CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Res*. 2019;47(D1):D721–D728.

25. Franzén O, Gan LM, Björkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database*. 2019;2019(1):46.

26. Paisley BM, Liu Y. GeneMarkeR: A Database and User Interface for scRNA-seq Marker Genes. *Front Genet*. 2021;12:763431.

27. Börner K, Teichmann SA, Quardokus EM, et al. Anatomical structures, cell types and biomarkers of the Human Reference Atlas. *Nature Cell Biology 2021 23:11*. 2021;23(11):1117–1128.

28. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst*. 2015;1(6):417.

29. Hao Y, Hao S, Andersen-Nissen E, et al. Integrated analysis of multimodal single-cell data. *Cell*. 2021;184(13):3573-3587.e29.

30. Domínguez Conde C, Xu C, Jarvis LB, et al. Cross-tissue immune cell analysis reveals tissue-specific features in humans. *Science*. 2022;376(6594):.

31. Stelmach P, Trumpp A. Leukemic stem cells and therapy resistance in acute myeloid leukemia. *Haematologica*. 2023;108(2):353–366.

32. Pollyea DA, Jordan CT. Therapeutic targeting of acute myeloid leukemia stem cells. *Blood*. 2017;129(12):1627–1635.

33. Chan WI, Huntly BJP. Leukemia stem cells in acute myeloid leukemia. *Semin Oncol*. 2008;35(4):326–335.

34. Thomas D, Majeti R. Biology and relevance of human acute myeloid leukemia stem cells. *Blood*. 2017;129(12):1577–1585.

35. Frigyesi I, Adolfsson J, Ali M, et al. Robust isolation of malignant plasma cells in multiple myeloma. *Blood*. 2014;123(9):1336–1340.

36. Flores-Montero J, de Tute R, Paiva B, et al. Immunophenotype of normal vs. myeloma plasma cells: Toward antibody panel specifications for MRD detection in multiple myeloma. *Cytometry B Clin Cytom*. 2016;90(1):61–72.

37. Hideshima T, Bergsagel PL, Kuehl WM, Anderson KC. Advances in biology of multiple myeloma: clinical applications. *Blood*. 2004;104(3):607–618.

38. Shastri A, Will B, Steidl U, Verma A. Stem and progenitor cell alterations in myelodysplastic syndromes. *Blood*. 2017;129(12):1586–1594.

39. Ganan-Gomez I, Yang H, Ma F, et al. Stem cell architecture drives myelodysplastic syndrome progression and predicts response to venetoclax-based therapy. *Nature Medicine 2022 28:3*. 2022;28(3):557–567.

40. Gao Y, Vasic R, Song Y, et al. m6A Modification Prevents Formation of Endogenous Double-Stranded RNAs and Deleterious Innate Immune Responses during Hematopoietic Development. *Immunity*. 2020;52(6):1007-1021.e8.

41. Winkler R, Gillis E, Lasman L, et al. m6A modification controls the innate immune response to infection by targeting type I interferons. *Nat Immunol*. 2019;20(2):173–182.

42. Diehl AD, Meehan TF, Bradford YM, et al. The Cell Ontology 2016: enhanced content, modularization, and ontology interoperability. *J Biomed Semantics*. 2016;7(1):.

43. Ianevski A, Giri AK, Aittokallio T. Fully-automated and ultra-fast cell-type identification using specific marker combinations from single-cell transcriptomic data. *Nature Communications 2022 13:1*. 2022;13(1):1–10.

44. Zheng GXY, Terry JM, Belgrader P, et al. Massively parallel digital transcriptional profiling of single cells. *Nature Communications 2017 8:1*. 2017;8(1):1–12.

45. Oetjen KA, Lindblad KE, Goswami M, et al. Human bone marrow assessment by single-cell RNA sequencing, mass cytometry, and flow cytometry. *JCI Insight*. 2018;3(23):.

46. Stuart T, Butler A, Hoffman P, et al. Comprehensive Integration of Single-Cell Data. *Cell*. 2019;177(7):1888-1902.e21.

47. Triana S, Vonficht D, Jopp-Saile L, et al. Single-cell proteo-genomic reference maps of the hematopoietic system enable the purification and massive profiling of precisely defined cell states. *Nature Immunology 2021 22:12*. 2021;22(12):1577–1589.

48. van Gils N, Denkers F, Smit L. Escape From Treatment; the Different Faces of Leukemic Stem Cells and Therapy Resistance in Acute Myeloid Leukemia. *Front Oncol*. 2021;11:.

49. Jongen-Lavrencic M, Grob T, Hanekamp D, et al. Molecular Minimal Residual Disease in Acute Myeloid Leukemia. *N Engl J Med*. 2018;378(13):1189–1199.

50. Terwijn M, Kelder A, Huijgens PC, et al. High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. *J Clin Oncol*. 2013;31(31):3889–3897.

51. Shimony S, Stahl M, Stone RM. Acute myeloid leukemia: 2023 update on diagnosis, risk-stratification, and management. *Am J Hematol*. 2023;98(3):502–526.

52. Hanekamp D, Cloos J, Schuurhuis GJ. Leukemic stem cells: identification and clinical application. *Int J Hematol*. 2017;105(5):549–557.

53. Barreto IV, Pessoa FMC de P, Machado CB, et al. Leukemic Stem Cell: A Mini-Review on Clinical Perspectives. *Front Oncol*. 2022;12:931050.

54. van Galen P, Hovestadt V, Wadsworth MH, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell*. 2019;176(6):1265-1281.e24.

55. Beneyto-Calabuig S, Merbach AK, Kniffka JA, et al. Clonally resolved single-cell multi-omics identifies routes of cellular differentiation in acute myeloid leukemia. *Cell Stem Cell*. 2023;30(5):706-721.e8.

56. Ledergor G, Weiner A, Zada M, et al. Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. *Nat Med*. 2018;24(12):1867–1876.

57. Pei S, Shelton IT, Gillen AE, et al. A Novel Type of Monocytic Leukemia Stem Cell Revealed by the Clinical Use of Venetoclax-Based Therapy. *Cancer Discov*. 2023;13(9):2032–2049.

58. Kennedy JA, Ebert BL. Clinical Implications of Genetic Mutations in Myelodysplastic Syndrome. *J Clin Oncol*. 2017;35(9):968–974.

59. Visconte V, Nakashima MO, Rogers HJ. Mutations in Splicing Factor Genes in Myeloid Malignancies: Significance and Impact on Clinical Features. *Cancers (Basel)*. 2019;11(12):.

60. Yoshida K, Sanada M, Shiraishi Y, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. *Nature*. 2011;478(7367):64–69.

61. Bejar R, Stevenson KE, Caughey BA, et al. Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. *J Clin Oncol*. 2012;30(27):3376–3382.

62. Makishima H, Visconte V, Sakaguchi H, et al. Mutations in the spliceosome machinery, a novel and ubiquitous pathway in leukemogenesis. *Blood*. 2012;119(14):3203–3210.

63. Graubert TA, Shen D, Ding L, et al. Recurrent mutations in the U2AF1 splicing factor in myelodysplastic syndromes. *Nat Genet*. 2011;44(1):53–57.

64. Ogawa S. Genetics of MDS. *Blood*. 2019;133(10):1049–1059.

65. Patnaik MM, Lasho TL, Finke CM, et al. Spliceosome mutations involving SRSF2, SF3B1, and U2AF35 in chronic myelomonocytic leukemia: prevalence, clinical correlates, and prognostic relevance. *Am J Hematol*. 2013;88(3):201–206.

66. Thol F, Kade S, Schlarmann C, et al. Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. *Blood*. 2012;119(15):3578–3584.

67. Walter MJ, Shen D, Shao J, et al. Clonal diversity of recurrently mutated genes in myelodysplastic syndromes. *Leukemia*. 2013;27(6):1275–1282.

68. Yip BH, Steeples V, Repapi E, et al. The U2AF1S34F mutation induces lineage-specific splicing alterations in myelodysplastic syndromes. *J Clin Invest*. 2017;127(6):2206–2221.

69. Desrosiers R, Friderici K, Rottman F. Identification of methylated nucleosides in messenger RNA from Novikoff hepatoma cells. *Proc Natl Acad Sci U S A*. 1974;71(10):3971–3975.

70. Frye M, Harada BT, Behm M, He C. RNA modifications modulate gene expression during development. *Science*. 2018;361(6409):1346–1349.

71. Liu N, Dai Q, Zheng G, et al. N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature*. 2015;518(7540):560–564.

72. Nagy RM, Mohamed AAEH, El-Gamal RAER, Ibrahim SAM, Pessar SA. Methyltransferase-like 3 gene (METTL3) expression and prognostic impact in acute myeloid leukemia patients. *Egyptian Journal of Medical Human Genetics*. 2022;23(1):1–13.

73. Li M, Ye J, Xia Y, et al. METTL3 mediates chemoresistance by enhancing AML homing and engraftment via ITGA4. *Leukemia 2022 36:11*. 2022;36(11):2586–2595.

74. Yankova E, Blackaby W, Albertella M, et al. Small-molecule inhibition of METTL3 as a strategy against myeloid leukaemia. *Nature*. 2021;593(7860):597–601.

75. Cheng Y, Luo H, Izzo F, et al. m6A RNA Methylation Maintains Hematopoietic Stem Cell Identity and Symmetric Commitment. *Cell Rep*. 2019;28(7):1703-1716.e6.

76. Sturgess K, Yankova E, Vijayabaskar MS, et al. Pharmacological inhibition of METTL3 impacts specific haematopoietic lineages. *Leukemia 2023 37:10*. 2023;37(10):2133–2137.

77. Hu C, Li T, Xu Y, et al. CellMarker 2.0: an updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res*. 2023;51(D1):D870–D876.

78. Ranzoni AM, Tangherloni A, Berest I, et al. Integrative Single-Cell RNA-Seq and ATAC-Seq Analysis of Human Developmental Hematopoiesis. *Cell Stem Cell*. 2021;28(3):472-487.e7.

79. Liu Y, Yang M, Deng Y, et al. High-Spatial-Resolution Multi-Omics Sequencing via Deterministic Barcoding in Tissue. *Cell*. 2020;183(6):1665-1681.e18.

80. Deng Y, Bartosovic M, Ma S, et al. Spatial profiling of chromatin accessibility in mouse and human tissues. *Nature 2022 609:7926*. 2022;609(7926):375–383.