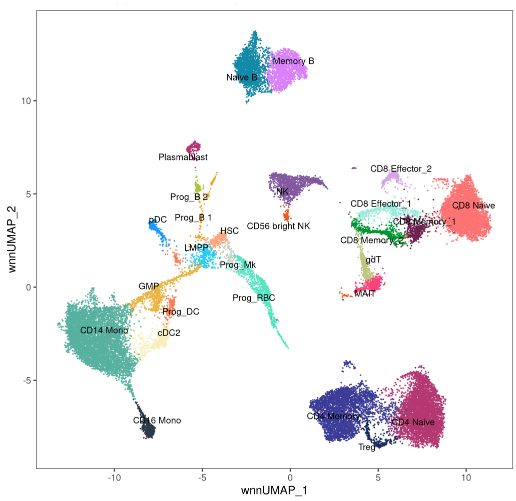
4.3. Lineage and mutation hierarchy

4.3.1. Lineage imputation on AML samples

As a first approach for classifying cell lineage in our scRNA-seq data after QC, we exploited a robust reference dataset of 30,672 BM-MNCs derived from one healthy human donor, profiled using CITE-seq and annotated based on both RNA and protein data(119) (see Materials and methods section, paragraph 3.4.4). Figure 37 shows the multimodal BM reference, which covers the full spectrum of hematopoietic differentiation from HSC to terminally differentiated cells.

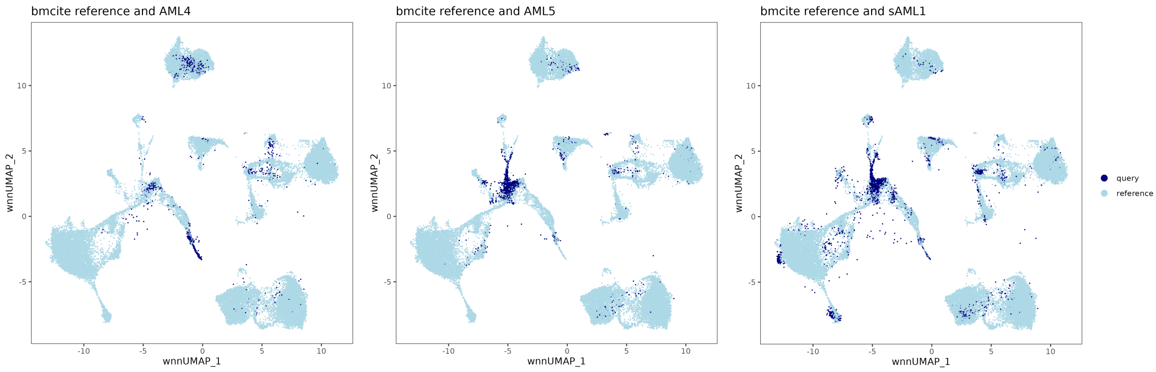
Figure 37. UMAP of reference healthy human BM-MNCs.
The weighted nearest neighbor (WNN) graph weights and combines information from RNA and protein data.



Briefly, we queried each of the three AML samples against the BM reference, by performing anchor-based data integration and lineage labels transferring to annotate AML populations (see Materials and methods, paragraph 3.4.4). It's worth noting that, with the illustrated approach, cells are assigned to a certain lineage based on the similarity to the corresponding counterpart in healthy hematopoiesis, which might lead to misclassification since malignant cells typically show aberrant expression patterns. In AMLs, in particular, leukemic blasts are immature cells whose transcriptional patterns may span across a wide spectrum of normal hematopoietic cells, including HSCs and more differentiated myeloid cells.

Figure 38 shows cells of each query AML mapped onto the coordinates of the BM reference in Figure 37. A variable fraction of cells resulted located in the area corresponding to HSC and undifferentiated progenitors (~10% for AML4, ~75% for AML5 and sAML1), while remaining cells were variably distributed across more differentiated lineages. This picture corresponds to the immunophenotypic profiles of the three samples (see Table 6, chapter 3.1), which showed that blast populations carried rather undifferentiated profiles (CD34 and CD117 positivity in all samples, CD38 positivity in AML5 and sAML1). However, while the immunophenotypically-assessed blast percentage in AML5 and sAML1 was pretty consistent with the fraction of cells mapping to HSC and progenitors in scRNA data (70-80% for both), we noticed a discrepancy for AML4 (80% by immunophenotype and pathology report, 10% in scRNA data), which might be due to hemodilution during BM aspirate.

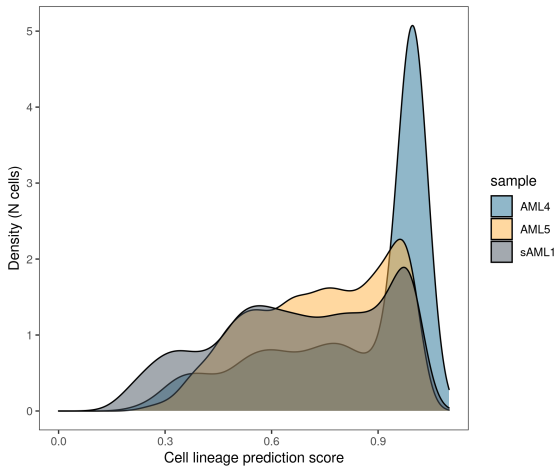
Figure 38. UMAP of AML samples mapped onto a multimodal healthy BM reference.
Single cells of each AML sample (blue) are projected onto the WNN graph of the BM reference (light-blue).



Thus, to investigate the accuracy of our imputation, we assessed the quality of lineage assignment for each cell, by measuring lineage label predictions as elaborated by Stuart et al. and implemented in the *TransferData* function of Seurat. Briefly, lineage label predictions are computed by multiplying the anchor classification matrix (which contains the classification information for each anchor cell in the reference dataset) with the transpose of the weights matrix (which defines the strength of association between each query cell and each anchor). This returns a prediction score for each lineage for every cell in the query dataset, ranging from 0 to 1.

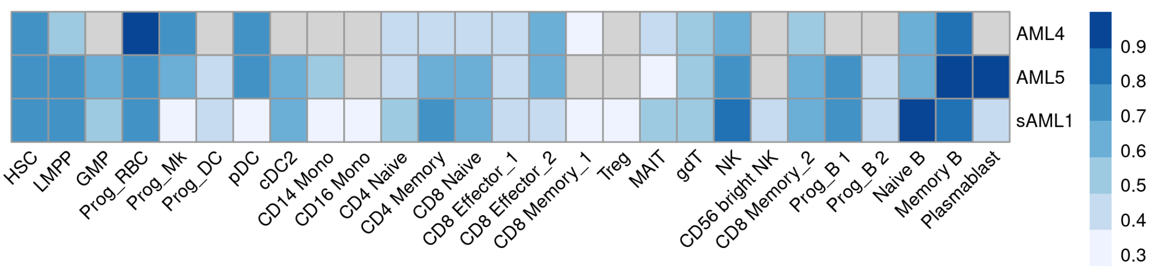
Figure 39 shows the distribution of prediction scores across all cells for the three samples, without accounting for specific lineages. The majority of cells in all samples were assigned a prediction score of >0.5, with the sAML1 sample showing the highest proportion of cells with low prediction score (< 0.35).

Figure 39. Lineage prediction scores by AML sample.
Distribution of lineage prediction scores for all cells in each AML sample, after mapping to the multimodal healthy BM reference.



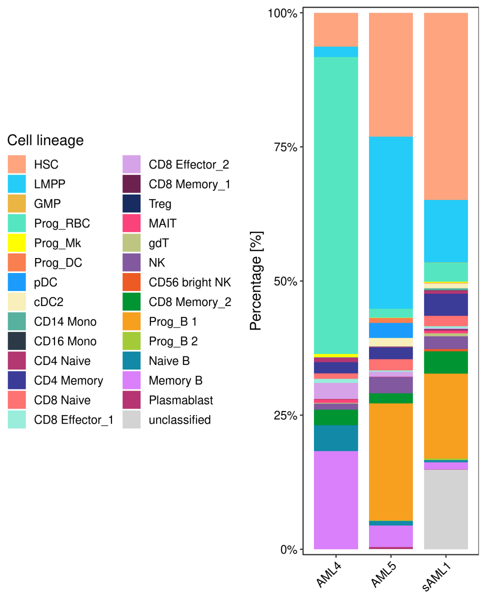
To investigate whether different lineages had been imputed with different accuracy, for each sample we computed the mean prediction score of all cells in each lineage, and compared results across samples by visualization in a heatmap (Figure 40). Strikingly, most lineages (HSC, LMPP, erythroid progenitors, NK cells, naive and memory B cells) had homogeneously high prediction scores across all of the three samples. T cell subsets, instead, had homogeneously lower prediction scores, while differentiated myeloid lineages showed discordant results across the samples, with the lowest prediction scores found in sAML1.

Figure 40. Mean prediction scores by lineage.
The heatmap shows the mean of prediction scores for all cells in each lineage across the three AML samples. Lineages with no cells assigned are labelled in grey.



For cells assigned to the T cell compartment, we accepted lineage imputation regardless of low prediction scores, because T cells may show heterogeneous functional states and overlapping transcriptional features that make difficult to accurately distinguish among different subsets. Instead, remaining cells with prediction score < 0.35 were considered unclassified cells. Results from this analysis (Figure 41) showed that the majority of cells could be assigned to a specific lineage (100% in AML4, 100% in AML5 and 90% in sAML1) and confirmed that approximately 10% of AML4 cells and 50% of AML5 and sAML1 cells belonged to HSC and progenitors lineages. A large proportion (~50%) of AML4 cells was assigned as erythroid progenitors, while other differentiated myeloid lineages were poorly represented. T and B lymphoid subsets made about 30% of AML4 cells and 20% of AML5 and sAML1. Not all lineages of the BM reference database, however, were represented in all samples.

Figure 41. Proportions of hematopoietic lineages across AML samples after mapping to a multimodal healthy BM reference.



Aiming to validate our lineage imputation using an orthogonal method and to further investigate unclassified cells, we assessed the expression of lineage-specific signatures from the BM Human Cell Atlas (HCA) on cells of each imputed lineage (Figures 42-50).

Figure 42. Mean expression of HSC and undifferentiated progenitors HCA signatures.
Median and distribution of average expression of selected HCA signatures across lineages (rows) and samples (columns).

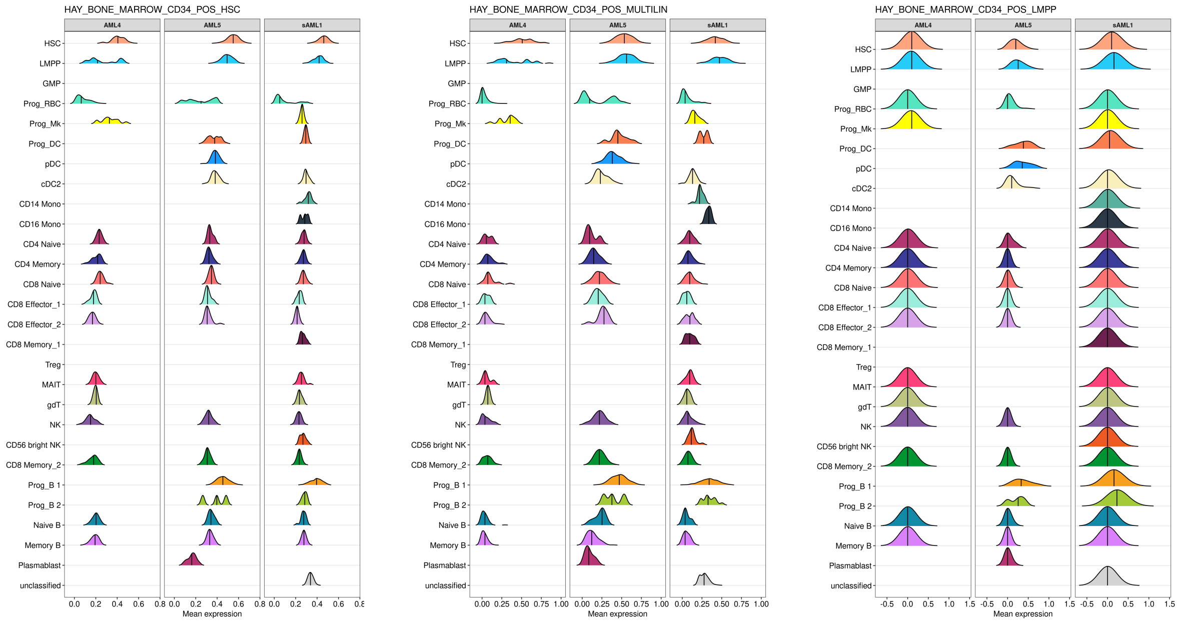


Figure 43. Mean expression of granulocytes-committed HCA signatures.
As in Figure 42.

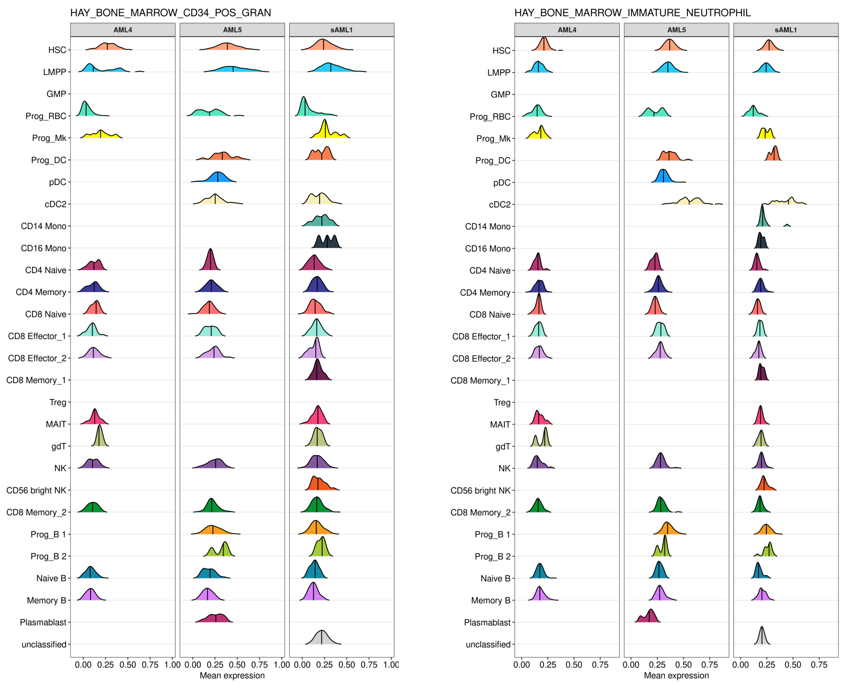


Figure 44. Mean expression of dendritic cells-committed HCA signatures.
As in Figure 42.

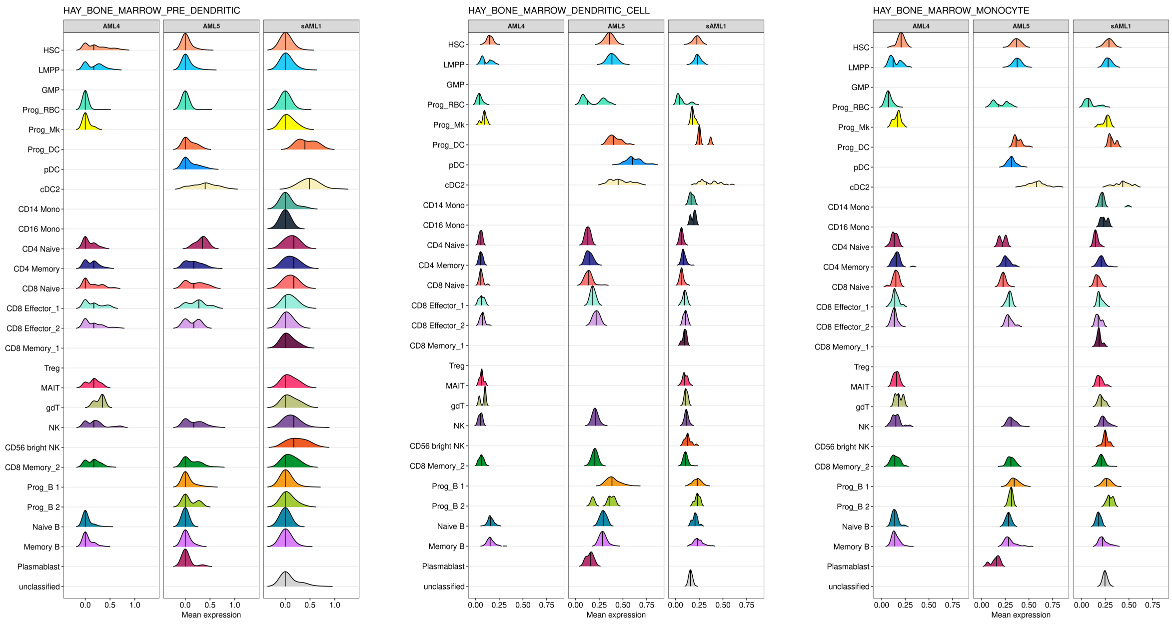


Figure 45. Mean expression of erythroid cells-committed HCA signatures.
As in Figure 42.

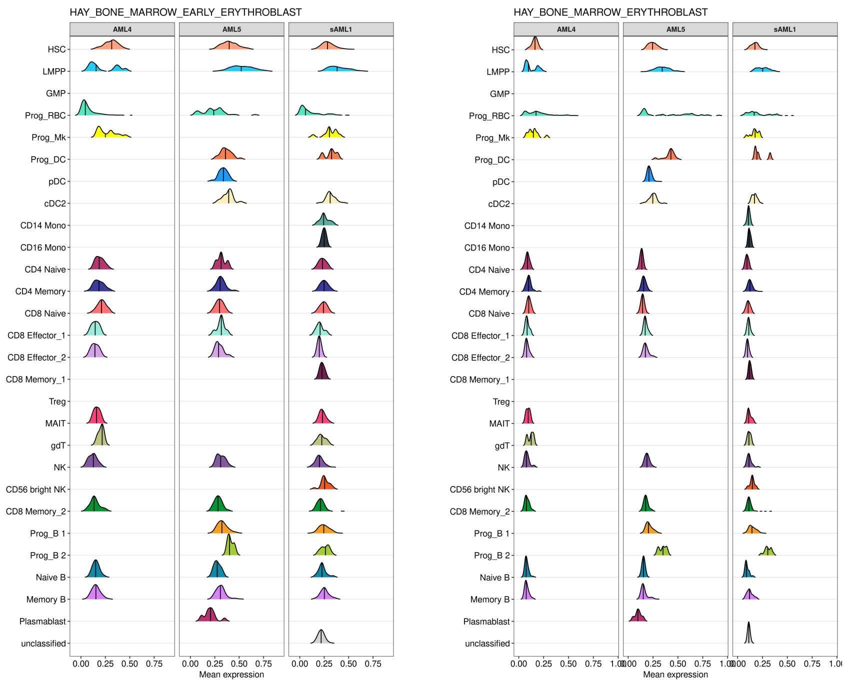


Figure 46. Mean expression of megakaryocytes-committed HCA signatures.
As in Figure 42.

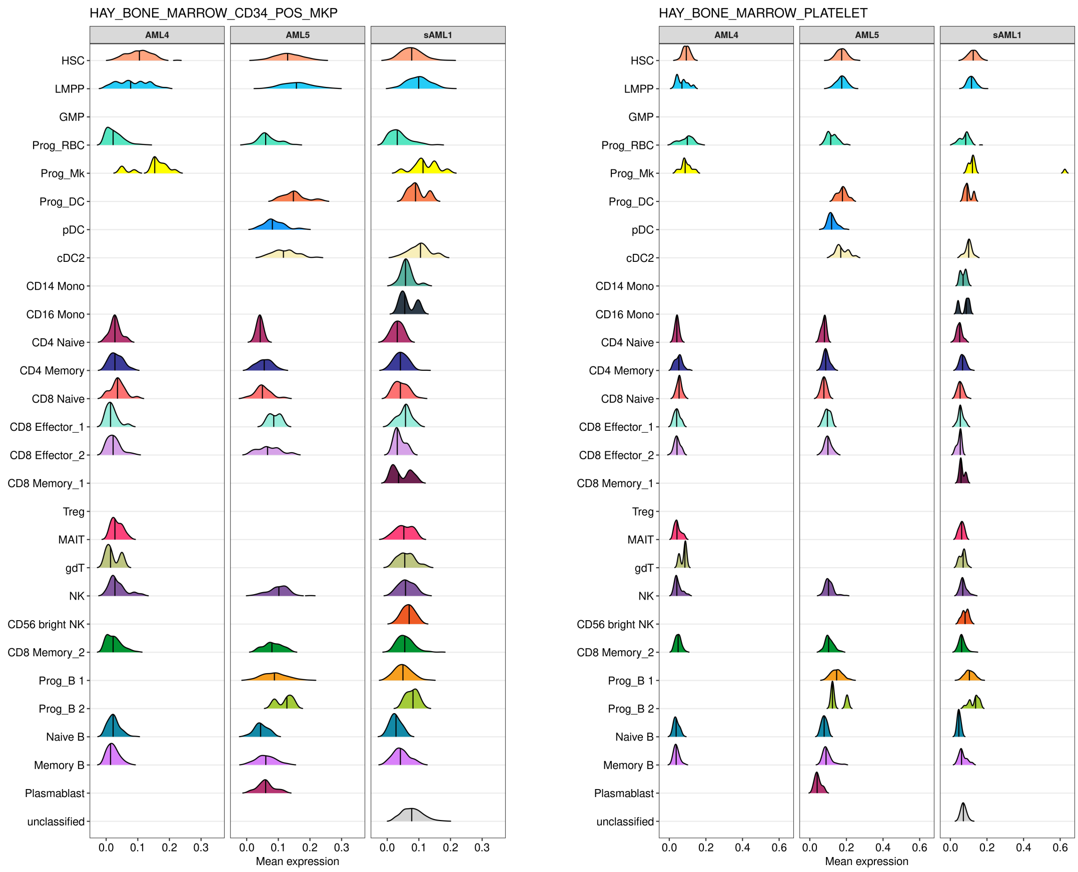


Figure 47. Mean expression of stromal cells HCA signatures.
As in Figure 42.

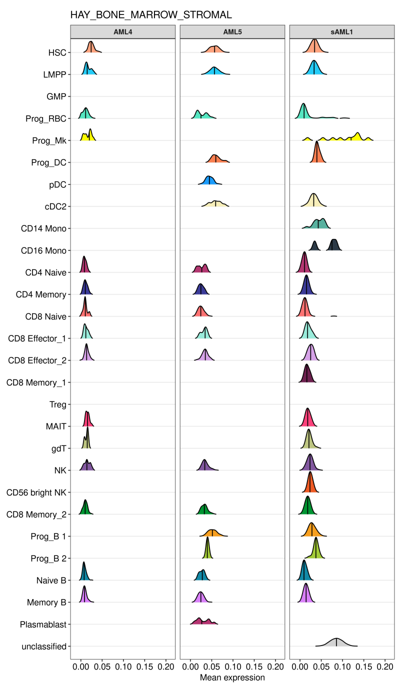


Figure 48. Mean expression of T and NK cells HCA signatures.
As in Figure 42.

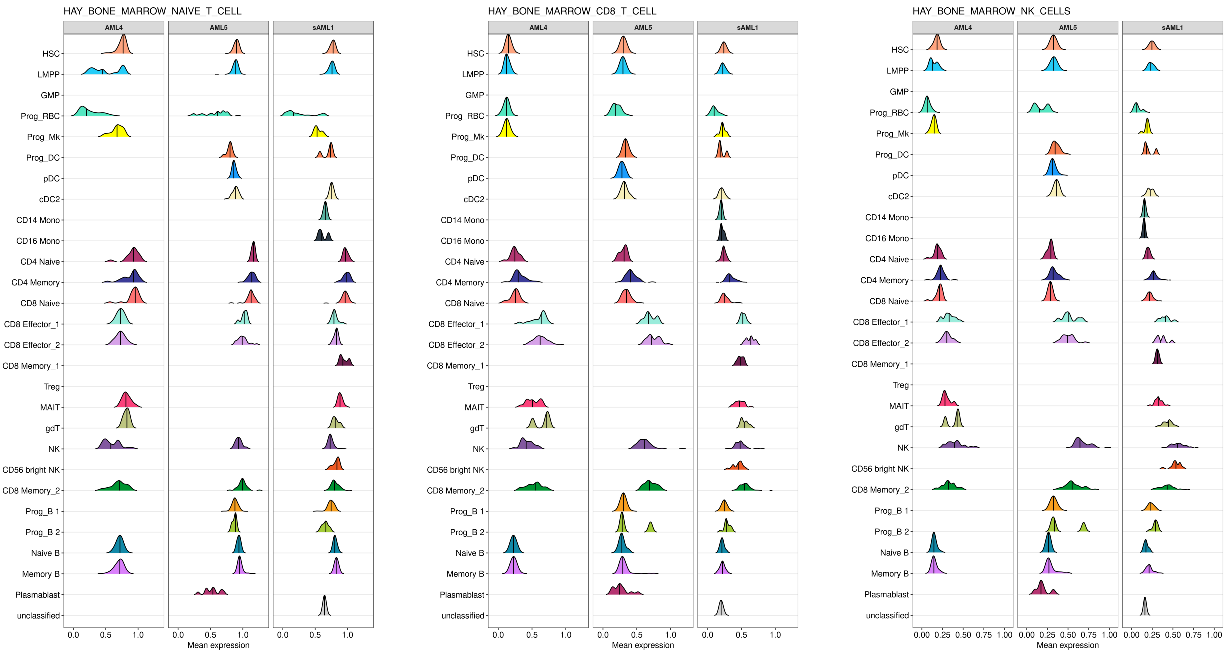


Figure 49. Mean expression of immature B cells HCA signatures.
As in Figure 42.

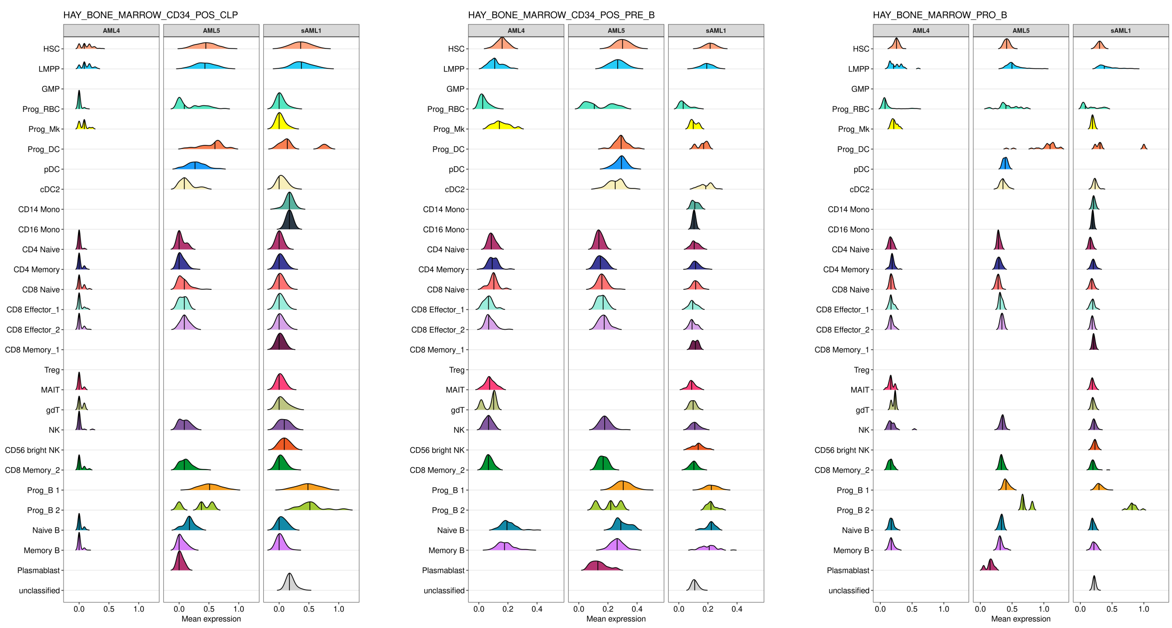
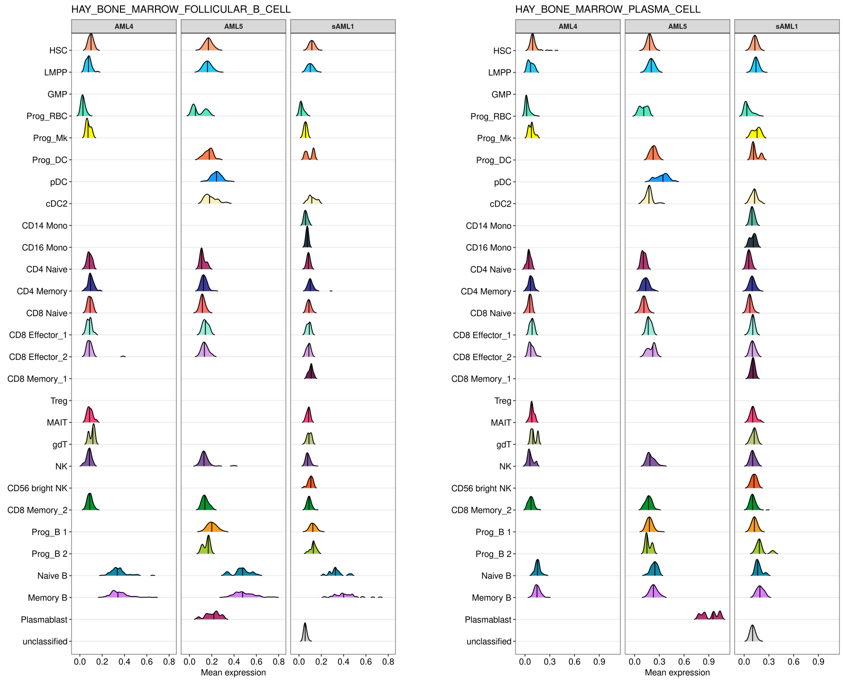
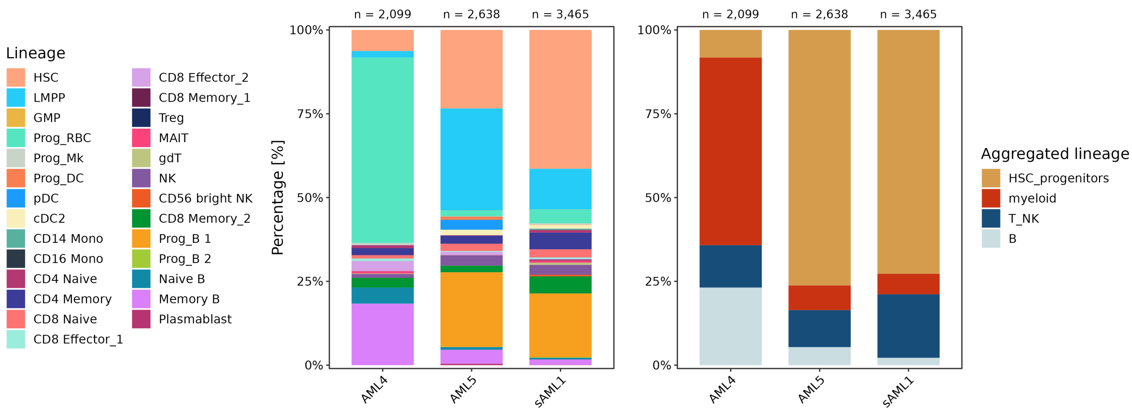


Figure 50. Mean expression of mature B cells HCA signatures.
As in Figure 42.



Upon checking BM Human Cell Atlas signatures, we generally found a good correspondence between our original imputation and the expression of expected signatures. These include both discrete and transitioning states associated to early progenitors and committed precursors, providing an opportunity to appreciate the heterogeneity of cells classified as HSC or immature progenitors. In particular, in AML5 and sAML1, cells imputed as HSC, LMPP, progenitors B1 and B2 showed overlapping expression of signatures related to HSC, undifferentiated progenitors and immature B cells, which suggests that the most undifferentiated cells (i.e., the putative malignant pool) aberrantly express immature B cell markers. In the case of AML4, we observed a discrepancy between the blast percentage assessed by immunophenotype (~80%) and the proportion of cells imputed as HSC and LMPP by scRNA data, which was much lower; as the same sample showed a high prevalence of mature erythroid cells, we wondered whether the stem/progenitor-like populations might also express signatures of more differentiated precursors. However, we did not observe any significant expression overlap, indicating HSC/LMPP and mature erythroid cells are two distinct populations in this sample. We interpreted the discrepancy between immunophenotype data and scRNA as a possible consequence of hemodilution during BM aspirate. Unclassified cells, strikingly, mostly expressed features linked to stromal cells. Therefore, we discarded unclassified cells from our final single-cell dataset, and maintained the original lineage imputation for all remaining cells; the final proportions of each hematopoietic lineage across the three AMLs showed little variation as compared to results before validation (Figure 51, left panel). Given the heterogeneity of cell lineage representation across samples, for the purpose of downstream analyses we re-grouped cells according to broader lineage categories: "HSC and progenitors" (HSPC) (including cells imputed as HSC, LMPP, GMP, progenitor B1 and B2); "myeloid" (including red blood cell progenitors, megakaryocyte progenitors, dendritic cells, CD14 and CD16 monocytes); "T and NK cells" (including CD4 naive cells, CD4 memory cells, CD8 naive cells, CD8 effector cells 1/2, CD8 memory cells 1/2, T regulatory cells, mucosal associated invariant T cells, $\gamma\delta$T cells, NK cells, CD56 bright NK), and "B cells" (including naive B cells, memory B cells, and plasmablasts) (Figure 51, right panel). A strikingly high proportion of HSC and progenitors could be observed in AML5 and sAML1 (in accordance with morphologically assessed BM blast percentage, see Table 6 in chapter 3.1), while AML4 showed more differentiated cells of the myeloid lineage (mostly mature erythroid progenitors). T, NK and B lymphoid populations were represented in all samples, accounting for the tumor immune milieu.

Figure 51. Validated hematopoietic lineages across AML samples.
The barplots show the proportions of validated hematopoietic lineages defined as of the BM reference (left) and by aggregated lineage categories (right).

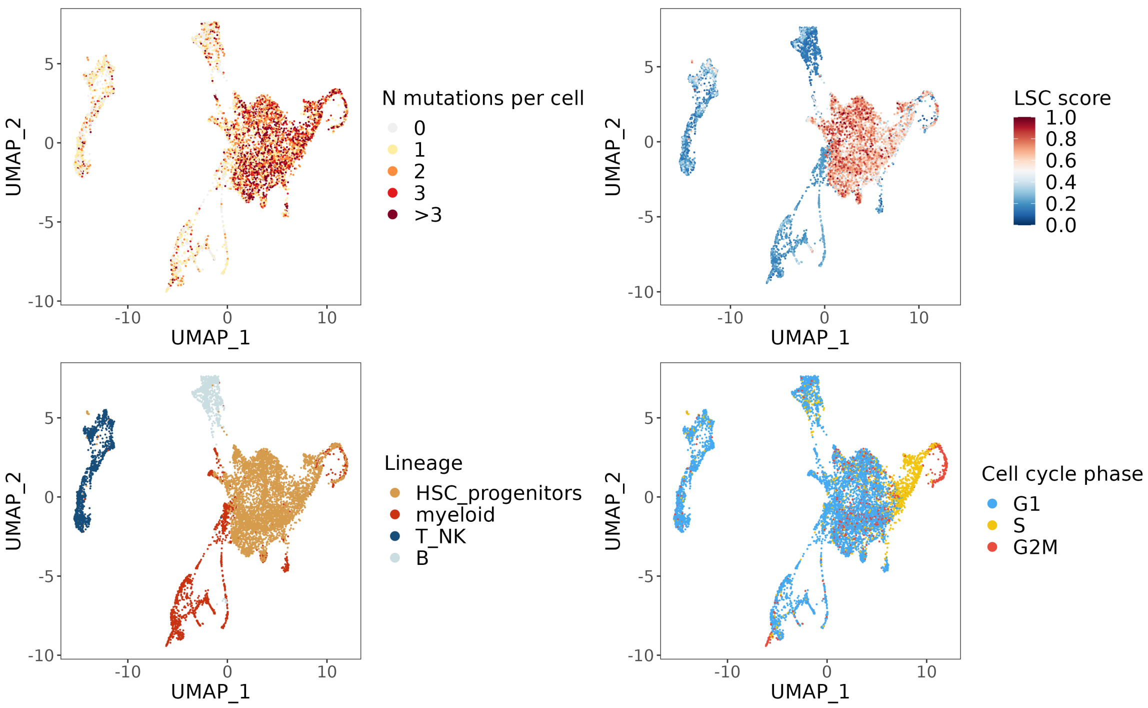


4.3.2 Identification of bona fide malignant cells

One obvious advantage of coupling transcriptional profiles with mutation analyses at single-cell level is the potential to facilitate identification of the malignant compartment of tumor samples,, which is preliminary to investigate the interplay between tumor and microenvironmental cells. AML malignant cells are intrinsically difficult to distinguish from residual hematopoiesis, both phenotypically and genetically. Phenotypically, AML cells are considered relatively immature cells. The presence of a HSC-like phenotype, however, is *per se* not sufficient to identify malignant cells due to their wide spectrum of HSPC-to-myeloid differentiation and, eventually, expression of aberrant lineage features. Genetically, leukemic blasts frequently coexist with apparently normal hematopoietic cells carrying clonal genetic alterations, as it happens in cases of AML evolving from CHIP or MDS.

SCM-seq provided us with the opportunity to adress this question by integrating lineage information with numbers of gene mutations *per* cell and specific transcriptional features (namely, LSC features and cell cycle phases). In Figure 52, we highlighted genotyped cells from the three samples upon integration in the same UMAP.

Figure 52. Identification of *bona fide* malignant AML cells.
The UMAP plots show the integration of genotyped cells from the three AML samples, colored by number of mutations *per* cell (top left), aggregated lineage (bottom left), LSC score (top right), and cell cycle phase (bottom right).

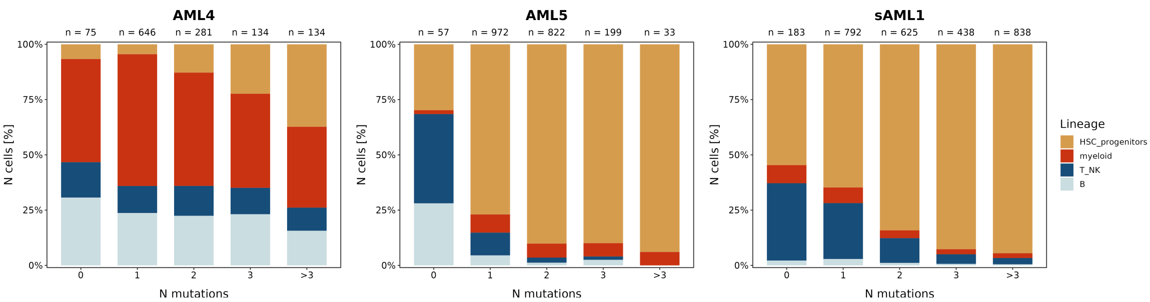


Mutated cells were represented in all the imputed lineages. Cells with more mutations, however, tended to cluster in a specific area of the UMAP (top left panel), suggesting commonalities in their transcriptional profiles. Strikingly, the same cells overlapped with cells imputed as HSC and progenitors (HSPCs, bottom left panel) and distinctively overexpressed a transcriptional signature capturing the core biological properties of functionally validated LSC (top right panel). Notably, within the same HSPC cluster, cells with less intense expression of the LSC signature were associated to cells in the S or G2M phases of cell cycle (bottom right panel). Thus, a large fraction of cells (about 50% of all samples) tend to cluster within the same transcriptional space, express HSPC and LSC transcriptional features and accumulate higher numbers of mutations, suggesting that cells imputed as HSC and progenitors can be *bona fide* considered as the malignant compartment of the AML ecosystem.

4.3.4 Genetic complexity and hierarchy across cell lineages

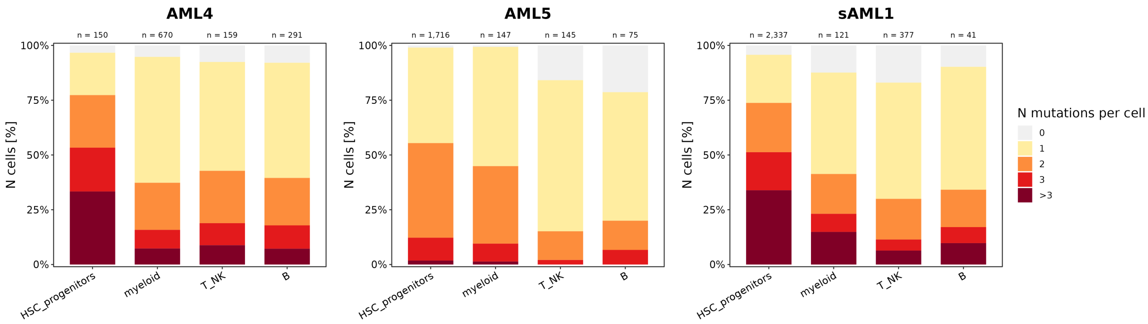
To further investigate the relationships between phenotypic and genetic heterogeneities, we analyzed the lineage architecture of genotyped cells grouped by increasing numbers of mutations *per* cell (Figure 53). Strikingly, in each AML sample, all hematopoietic lineages were represented in all cell groups, regardless of numbers of mutations. Their relative proportion, however, differed across groups, with progressive decrease of lineage heterogeneity and increase of HSPCs representation from non-mutated cells to cells with the highest numbers of mutations.

Figure 53. Relationship between hematopoietic lineages and numbers of mutations *per* cell.
The barplots show the proportions of hematopoietic lineages for genotyped cells grouped by increasing numbers of mutations *per* cell.



Indeed, as expected, HSPCs showed the highest genetic complexity, i.e. the highest proportion of cells bearing 3 or >3 mutations (Figure 54). Strikingly, however, the vast majority of cells belonging to differentiated lineages also held at least one mutation (>75%), with significant fractions also showing 2, 3 or even >3 mutations (e.g. 6-8% of T NK and B cells in sAML1).

Figure 54. Relationship between number of mutations *per* cell and hematopoietic lineage.
The barplots show the proportions of numbers of mutations *per* cell for genotyped cells grouped by aggregated hematopoietic lineages.



The finding of mutations in non-leukemic lymphoid cells may reflect the presence of residual CH, as previously reported(59). Although cells with myeloid differentiated features might be residual CH as well, the presence of increasing numbers of mutations in this lineage compartment might also indicate an intermediate phase between pre-leukemic and overt malignancy, especially in the context of *SRSF2*-mutated AMLs that are typically associated to MDS-like features. In either cases, mutations are expected to occur more frequently in genes involved in epigenetic and/or splicing regulation, while mutations in signaling pathways are typically found in late AML subclones. To investigate the frequency and co-occurrences of gene mutations across the various hematopoietic lineages, we reconstructed the frequency of somatic variants for genotyped cells of each lineage independently (HSPCs, differentiated myeloid cells and the immune compartment, i.e., T, NK and mature B cells) and visualized result by heatmaps. Overall, as shown in Figure 55, 56 and 57, myeloid and immune cells recapitulated the genetic hierarchy observed in HSPCs. Surprisingly, we also found rare immune cells bearing mutations usually not associated to CH (e.g., *CEBPA* in AML4 and *FLT3* in sAML1).

Figure 55. Mutation hierarchies across hematopoietic lineages (sample AML4).
The heatmaps show the presence of a mutation (red bar) for each gene variant (rows) in each genotyped cell (columns), for each aggregated lineage. Mutations are ordered by decreasing frequency (independently assessed for each lineage).

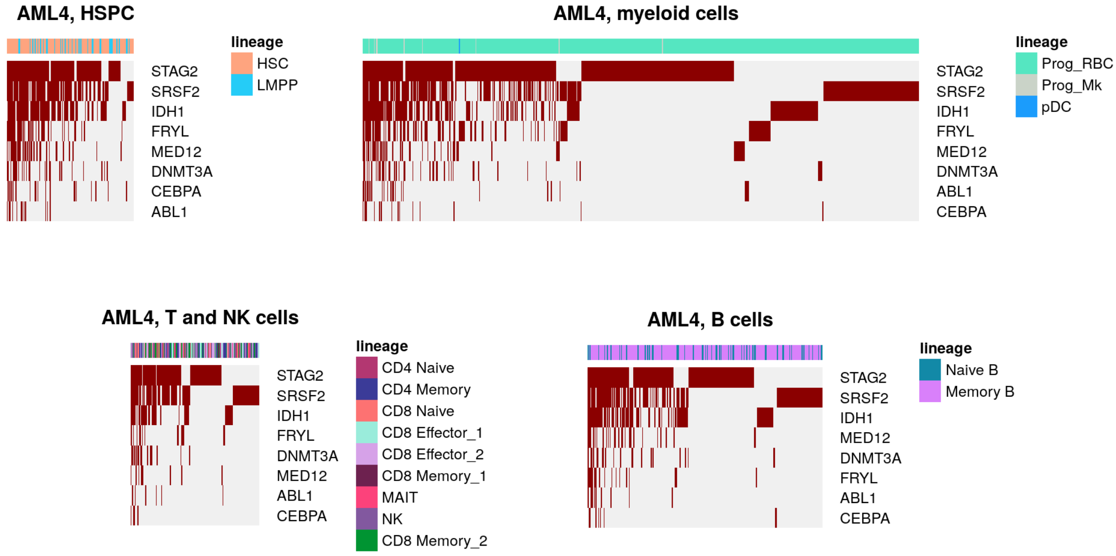


Figure 56. Mutation hierarchies across hematopoietic lineages (sample AML5).
As in Figure 55.

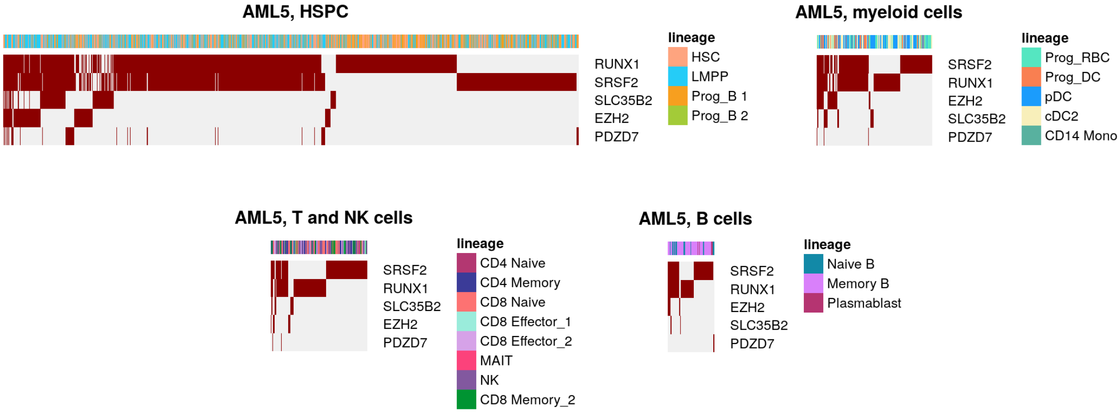
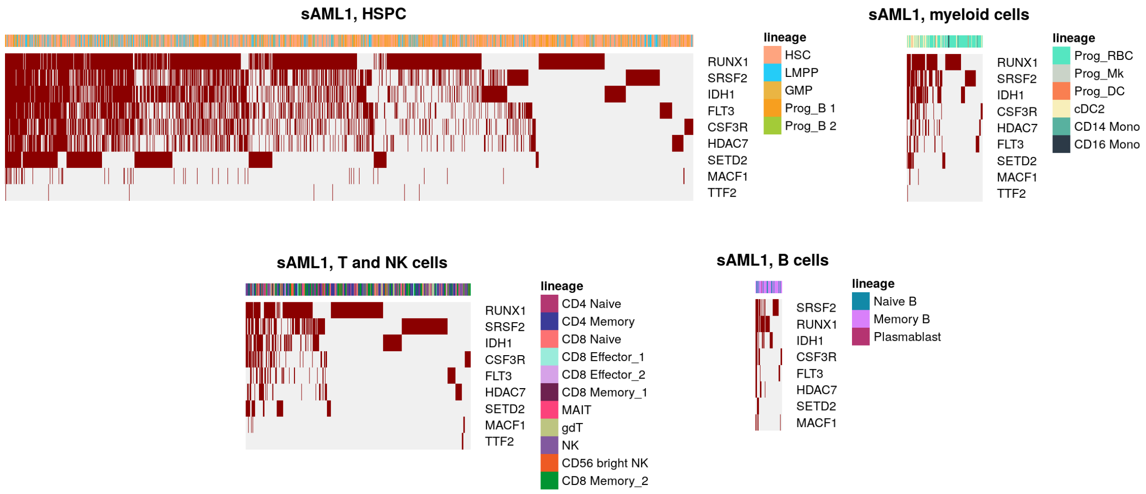
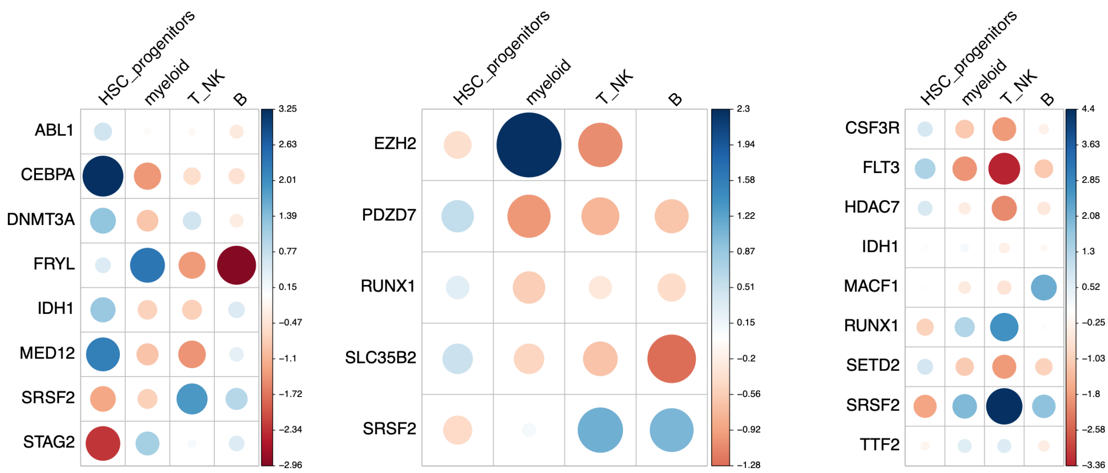


Figure 57. Mutation hierarchies across hematopoietic lineages (sample sAML1).
As in Figure 55.



Thus, we investigated whether specific gene mutations were preferentially enriched in specific lineage cell compartments. To this end, we compared the proportions of cell lineages represented in all cells bearing a given gene mutation, across all gene mutations in each AML sample (Figure 58). As a result, we found that some gene mutations were more strongly associated to specific lineages than others (AML4, p = 0.00003202 by Chi-square test; AML5, p = 0.3673 by Fisher's exact test; sAML1, p = 0.0004998 by Fisher's exact test). Specifically, gene mutations with higher VAF in our experimental data and previously reported to occur in CH (e.g., *STAG2*, *SRSF2, DNMT3A*) tended to show a positive association with lymphoid and differentiated myeloid cells, while gene mutations described to occur at later stages in leukemia development (e.g., *CEBPA*, *CSF3R, FLT3*) were positively associated with HSPCs.

Figure 58. Associations between gene mutations and hematopoietic lineages.
The circles represent Pearson residuals for significant Chi-square or Fisher's exact tests, colored by positive (blue) or negative (red) associations between a given gene mutation and a given lineage. For each association, the size of the circles is proportional to the amount of contribution to the difference between expected and observed values.

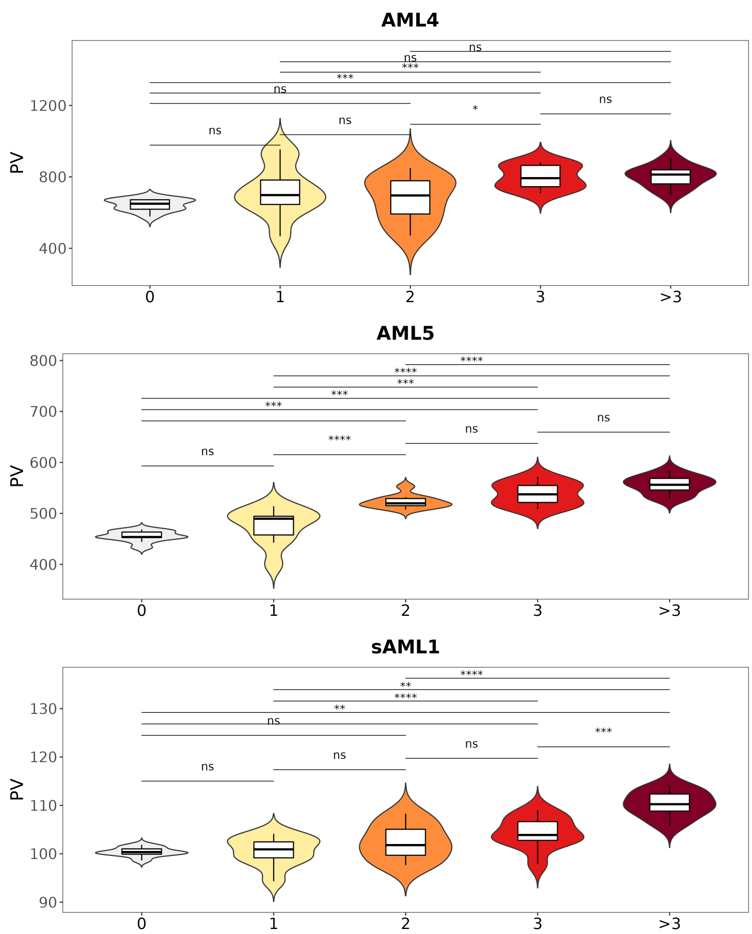


4.4. Genetic complexity and phenotypic diversity

4.4.1 Increasing genetic complexity in AML HSPCs is associated to increasing transcriptional heterogeneity and functional consequences

One key question regarding AML intra-tumor genetic and phenotypic heterogeneity is whether cells with high mutational burden show distinct functional properties as compared to non-mutated cells or cells with few mutations and, if so, whether this is independent of mutation combinations. To assess and quantify the diversity of AML transcriptional phenotypes based on genetic complexity, we selected genotyped HSPCs (i.e., the malignant compartment, to exclude any lineage or differentiation-related bias), grouped them by total number of mutations (0, 1, 2, 3 and >3, respectively) and computed their PV (i.e., the pseudo-determinant of gene expression covariance, an indirect measure of transcriptional identity; see Materials and methods, paragraph 3.4.6). In particular, larger PV in one cell group as compared to another shows the independency of active transcriptional programs, suggesting activation of additional mechanisms and pathways. We repeated PV computation 10 times to increase the robustness of the process.

Figure 59. Phenotypic volumes of HSPC by increasing numbers of mutations *per* cell.
Violin and boxplots represent the range of computed PVs across 10 repetitions. Two-sided Mann Whitney U test. ns = non significant, * = p<0.05, ** = p<0.01, *** = p<0.001.



Strikingly, we observed a significant stepwise increase in PVs for each increase in numbers of mutations *per* cell (Figure 59), suggesting that higher genetic complexity is associated to increased heterogeneity of active transcriptional programs regardless of underlying genotype combinations.

To assess which functional pathways are activated by the highest genetic complexity within each lineage, we calculated single-cell average expression levels of selected signatures using the *AddModuleScore* function in Seurat, and correlated each of them with the z-score of mutation burden (Figure 60). Higher numbers of mutations in HSPCs showed a positive, significant correlation with signatures related to cell cycle control, proliferation, response to oxidative stress, RNA splicing regulation, MTORC1 signaling and MYC targets, as well as anti-correlation with inflammatory pathways. Interestingly, we observed similar patterns when comparing HSPCs and more differentiated myeloid progenitors, which is consistent with the hypothesis that the myeloid compartment might include not only residual hematopoiesis, but also pre-malignant cells.

To further confirm the association between specific pathways and increasing mutational burden, we performed differential expression and pathway enrichment analysis between cells with none or 1 mutation vs cells with 3 or >3 mutations, for each sample separately. Pathways enriched with genes overexpressed in HSPCs with high mutation burden were mainly involved in mitosis, cell cycle G1/S phase transition, DNA repair/metabolism, RNA splicing, protein and mitochondrial metabolism, consistently with the above findings. Notably, the 8 genes that were shared across the three samples (Figure 61) pointed to biological functions such as protein translation regulation (*DCTN5*, *EIF3A* and *MRPS9*), post-translational processing and protein metabolism (*RBM10* and *UBA2*), pyrimidine metabolism (*DUT*), mitochondrial integrity and respiratory chain function (*CHCHD3*), and mitosis (*NUP37*). Intriguingly, marker genes of HSPCs with lower mutation burden in AML5 to pathways related to antigen processing and presentation of peptide antigen via MHC class II, suggesting that immunomodulatory properties are confined to cells with low genetic complexity. This observation, however, was limited to one single AML sample.

Figure 60. Relationship between mutation burden and selected transcriptional signatures.

Spearman correlation highlights the associations between increasing numbers of mutations *per* cell and averaged expression of signatures of interest. The color represents the direction of the association (blue for positive, red for negative), while the intensity of the color and the size of the circles is proportional to correlation coefficients. Only significant associations are given ($p < 0.05$).

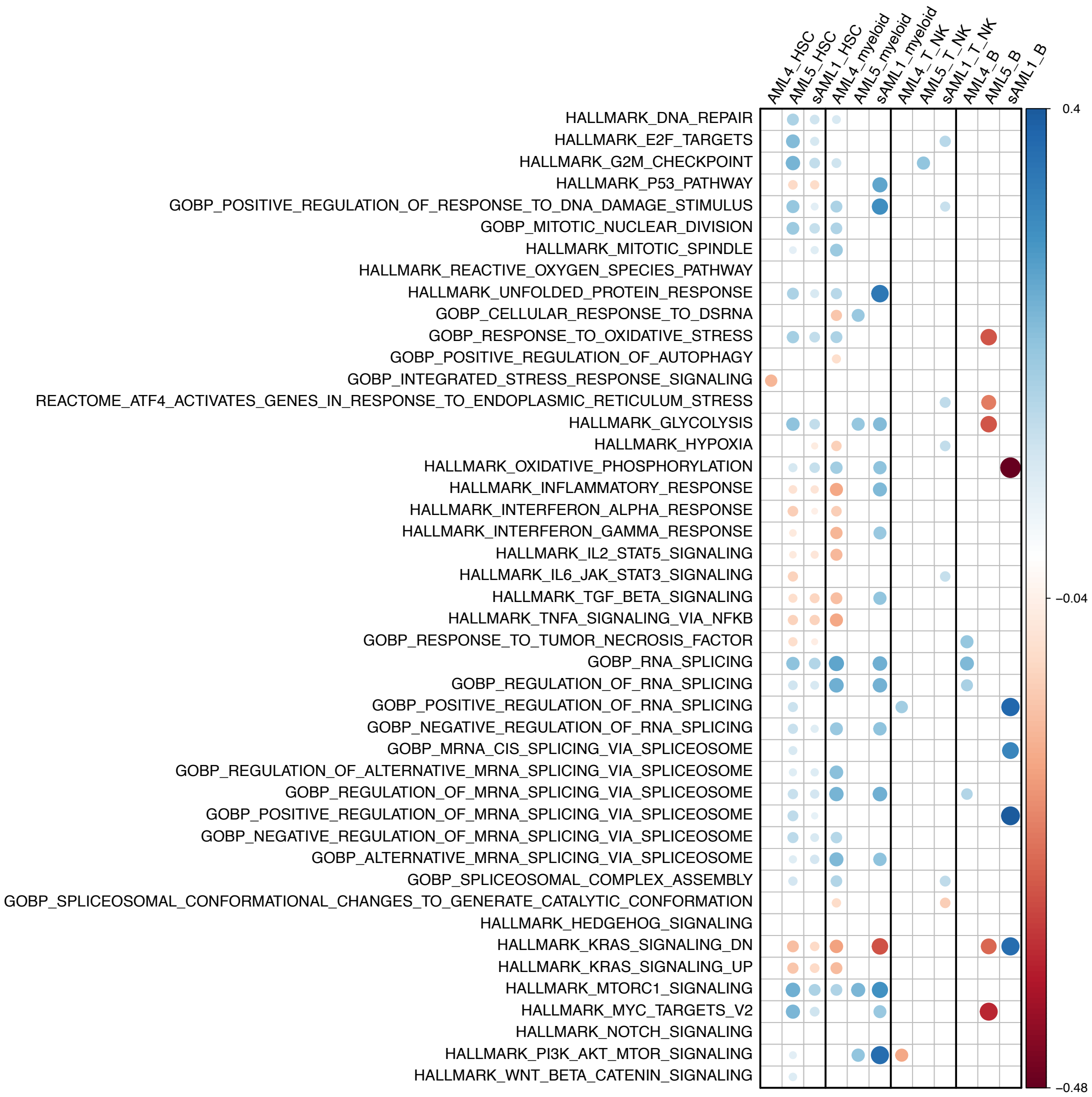
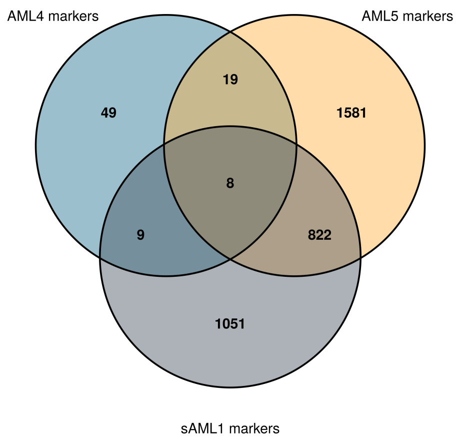


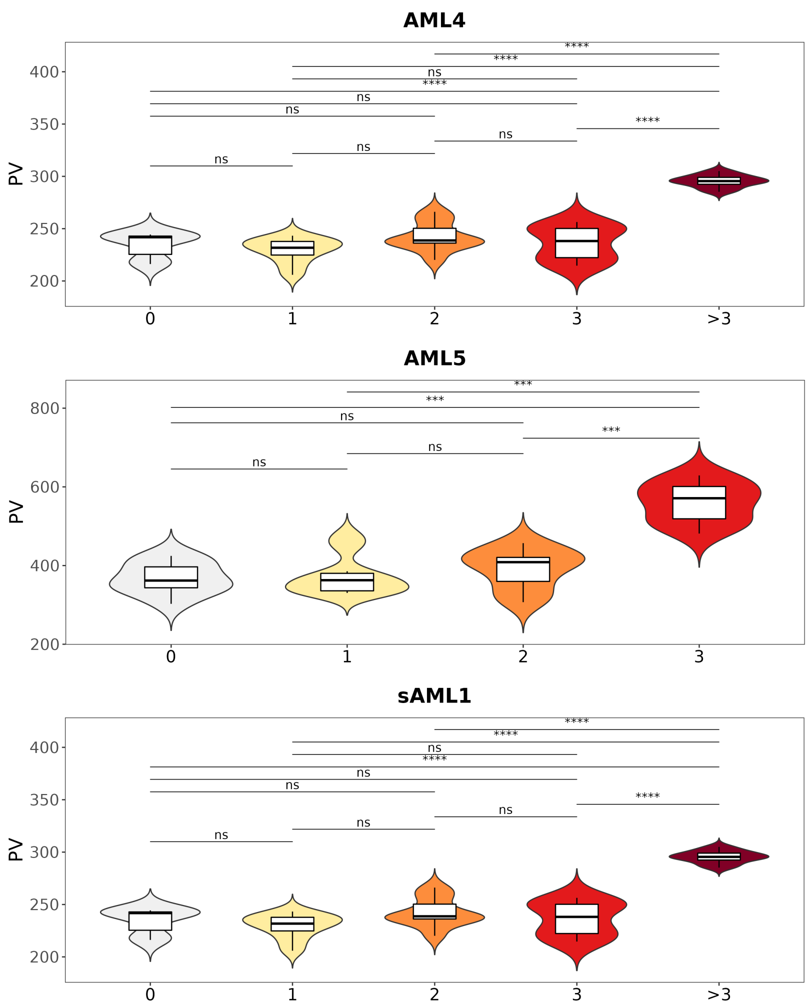
Figure 61. Overlap of genes overexpressed in HSPC with high mutation burden.



4.4.2 Transcriptional profiles of immune populations are poorly affected by genetic complexity

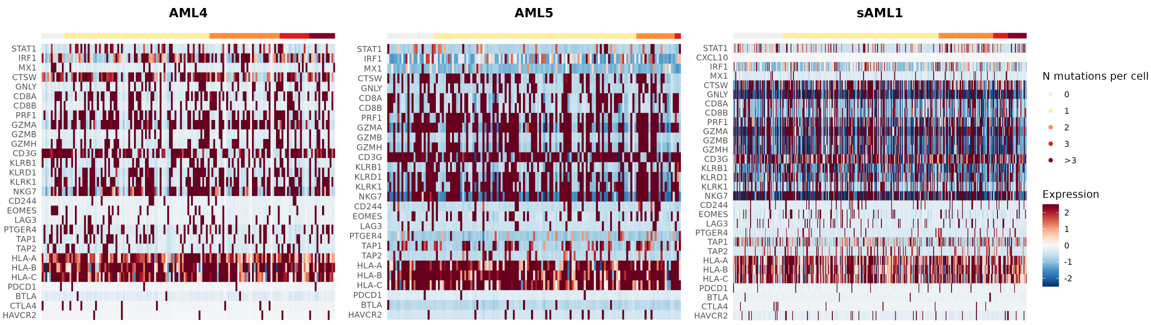
As both innate and adaptive immunity have been demonstrated to sustain the fitness of AML cells, we focused on immune cell subsets (T, NK and mature B cells) to address whether underlying genetic complexity affects their functional profiles. We used again PV as a measure of transcriptional heterogeneity on cell populations with increasing number of mutations, and found that PVs sharply arose in cells with >3 mutations, while cells with lower genetic complexity showed substantial stability (Figure 62). In keeping with this, correlations within the lymphoid lineages were far less frequent and systematic that those scored for HSPC and differentiated myeloid cells (Figure 60 above).

Figure 62. Phenotypic volumes of immune cells by increasing numbers of mutations *per* cell.
As in Figure 59.



Moreover, we analysed T and NK cells for the expression of genes with well-established lineage-specific functions, namely genes associated to interferon stimulation (*STAT1*, *CXCL10*, *IRF1*, *MX1*), cytotoxicity (*CTSW, GNLY, CD8A, CD8B, PRF1, GZMA, GZMB, GZMH, CD3G, KLRB1, KLRD1, KLRK1, NKG7*), T-cell exhaustion (*CD244, EOMES, LAG3, PTGER4*), antigen processing and presentation (*TAP1, TAP2, HLA-A, HLA-B, HLA-C*) and immunotherapy targets (*PDCD1, BTLA, CTLA4, HAVCR2*). Strikingly, we did not find any relevant change across cells bearing different numbers of mutations (Figure 63). Thus, with the limits of the low prevalence of cells with high numbers of mutations, we concluded that the overall impact of genetic complexity on immune functions is poor.

Figure 63. Expression of T/NK-related genes by number of mutations *per* cell.
The heatmaps show z-scored expressions of selected genes relevant to T/NK functions (rows) for each genotyped T/NK cell (columns). The upper bar highlights the number of mutations by which cells are grouped.



4.4.3 Single-cell isoform-level diversity and relationship to genetic complexity

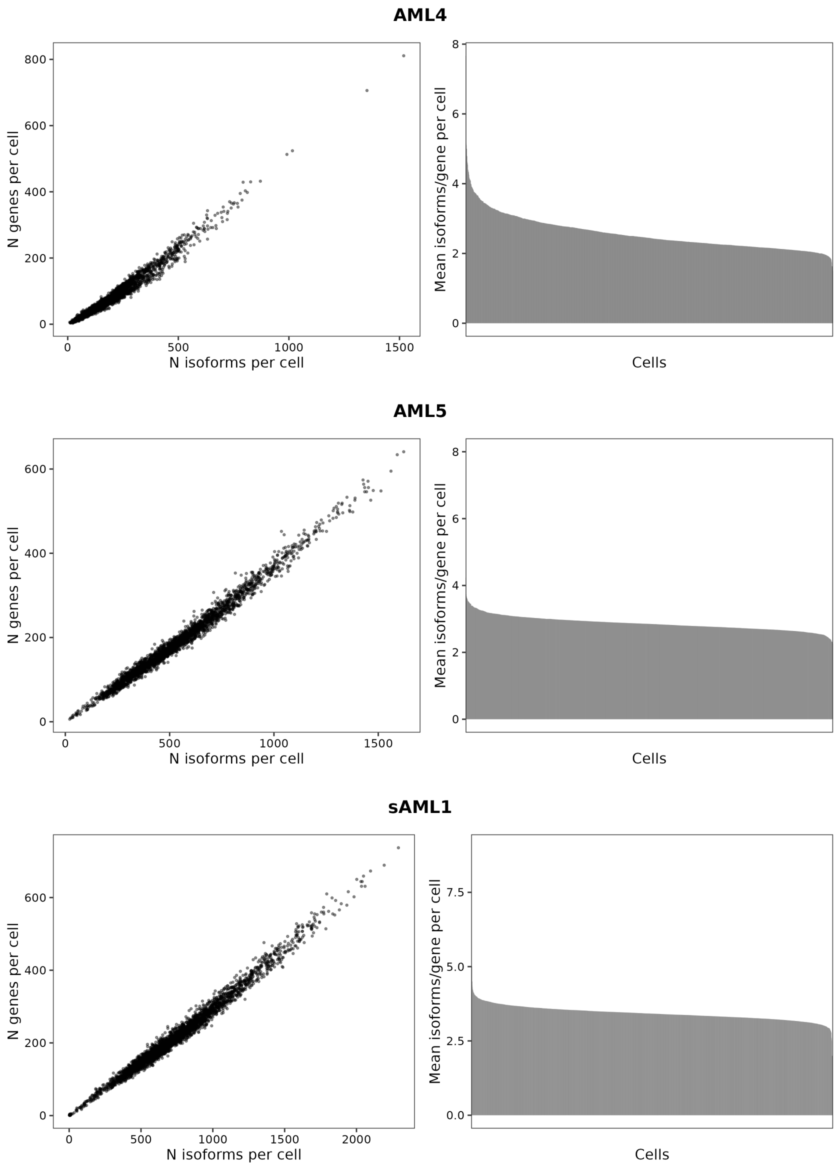
To further dissect the impact of genetic complexity on phenotypic heterogeneity, we focused on the analysis of the mRNA isoform-repertoire at single-cell level. To maximise the accuracy of transcript isoforms annotation and link this information to single cells, we performed whole-transcriptome ONT sequencing of sample-matched barcoded full-length cDNA. A summary of sequencing metrics is presented in paragraph 4.1.1 (Table 11). After CB matching between 10x and ONT datasets, we obtained isoform-*per*-cell matrices spanning quite heterogeneous numbers of isoforms, belonging to 3740, 7009 and 8710 genes, respectively (Table 17). As we have applied a threshold of 10 counts for isoforms inclusion, this estimate of isoform diversity is likely to be conservative. We further subsetted the matrices to include only cells with genotype assignment for at least one gene. Overall, we could integrate genotype, gene expression and isoform profiles on 60.5%, 78.8% and 83% of cells in the three AML samples, respectively.

Table 17. Summary of full-length transcriptome characteristics.

Sample	N cells	N reads (x 10^6)	N genes with multi-exon isoforms	N multi-exon isoforms	N cells after subsetting	N isoforms after subsetting
AML4	2112	31.56	3740	7884	1270	6365
AML5	2794	36.65	7009	18211	2083	14686
sAML1	4201	50.69	8710	26857	2876	22154

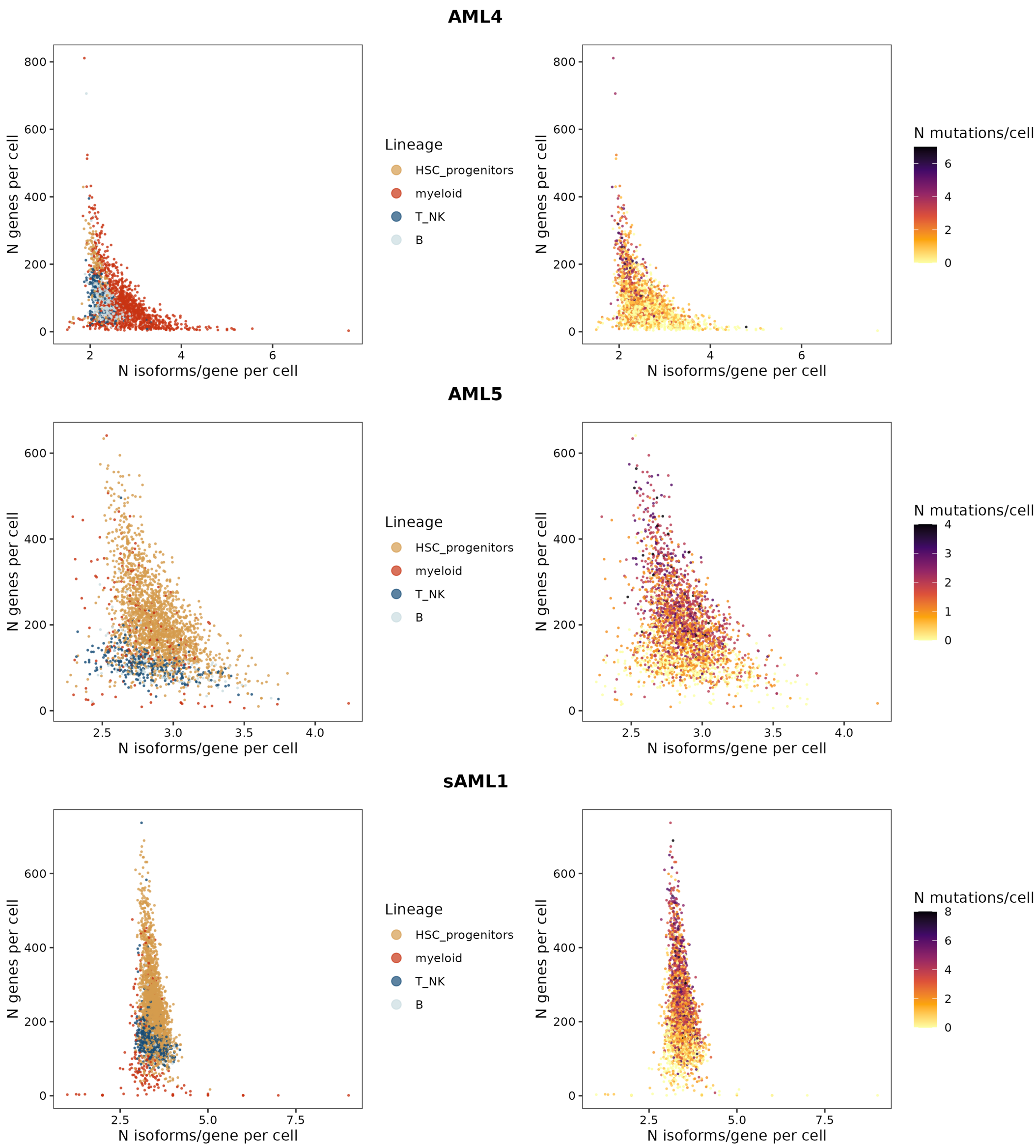
As expected, we found a neat linear relationship between the absolute total number of expressed genes and isoforms *per* cell (Figure 64, left panels). To capture cell-level isoform abundance and identify cells expressing aberrant numbers of isoforms (i.e., numbers of isoforms not expected based on numbers of expressed genes), we normalized the total number of isoforms on the total absolute number of expressed genes, and ranked cells based on this parameter (Figure 64, right panels). For each sample, the majority of cells exhibited mostly subtle, continuous variation in isoform abundance.

Figure 64. Relationship between expressed genes and expressed isoforms *per* cell.
Total numbers of expressed genes and expressed isoforms *per* cell (left) and isoform abundance *per* cell (normalized by number of expressed genes) (right).



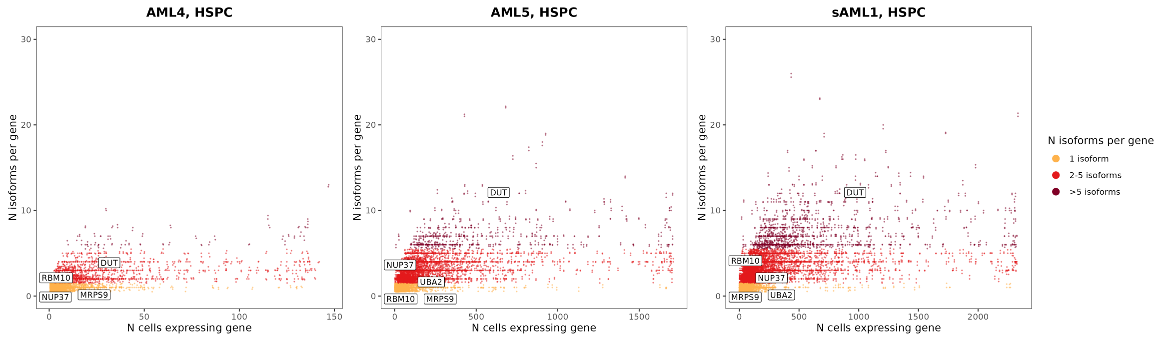
We then asked whether isoform diversity might correlate with cell lineage and mutation burden, which we investigated by plotting genotyped cells based on isoform abundance and number of expressed genes (Figure 65). For each of the three AML samples, we uncovered a high number of cells with heterogeneous isoform abundance, but relatively few expressed genes, and progressively fewer cells with lower isoform abundance, but higher numbers of expressed genes. Most of the cells were distributed along a continuum between these two states and, notably, cell lineage and mutation burden clustered accordingly. In particular, HSPCs with low mutation burden and differentiated myeloid and lymphoid cells tended to use higher numbers of isoforms for relatively few expressed genes, while HSPCs with high mutation burden expressed more genes with limited isoform abundance.

Figure 65. Single-cell isoform diversity.
Cell-level relationship between isoform abundance and expressed genes by cell lineage (left) and mutation burden (right).



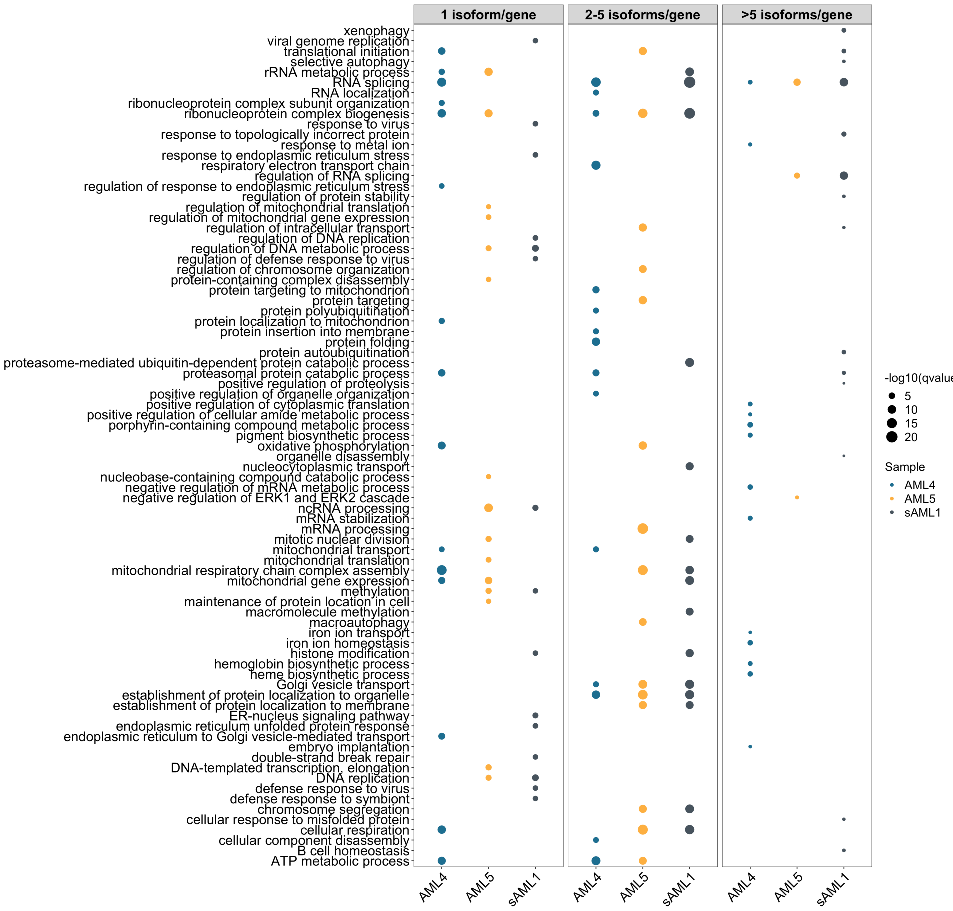
This result suggests that increasing mutation burden in AML HSPCs, despite higher numbers of expressed genes, is associated to the use of a progressively restricted repertoire of isoforms, possibly indicating a functional selection. To further delve into this finding and link it to the distinct gene expression profile we observed in cells with high mutation burden, we investigated genes expressed in HSPCs based on numbers of cells expressing any given gene, and the corresponding numbers of unique isoforms for that gene (Figure 66). Most of the genes expressed in HSPCs were scored in a relatively low number of cells, including genes overexpressed in HSPCs with high mutation burden (see paragraph 4.4.1). Notably and consistently with the profile shown in Figure 65, the number of isoforms expressed for these genes was generally low.

Figure 66. Gene-level isoform diversity in AML HSPCs.
Gene-level relationship between number of cells expressing the gene and number of unique isoforms *per* gene.



Genes expressed in high numbers of cells were often conserved across the three samples (88.4%, 54.3% and 49.6% of shared genes by sample) and mainly encoded for ribosomial or translation-related proteins when represented by few isoforms, while HLA and immune-related genes displayed higher isoform diversity. Instead, the majority of genes expressed in lower numbers of cells often showed sample-specific representation, although some pathways resulted enriched across all the three samples, i.e. RNA splicing, protein metabolism, mitochondrial functions, mitosis, response to unfolded protein and endoplasmic reticulum stress. We also evaluated whether certain pathways tended to be enriched in genes expressed with higher isoform diversity, and found that genes of RNA splicing-related pathways were consistently represented with even more than 5 isoforms in all samples (Figure 67). Instead, pathways related to mitosis, mitochondrial functions and protein metabolism were more often represented by genes expressed with fewer isoforms.

Figure 67. Isoform diversity-based pathway enrichment.
Top 15 GO terms (biological process category, ranked by FDR) enriched in genes expressed with 1, 2-5 or >5 isoforms, respectively. The size of the circles corresponds to FDRs.



Finally, we investigated whether the varying patterns of isoform diversity were associated to the presence of mutations of *SRSF2*, a gene that encodes for a spliceosome factor and is mutated in all of the three AML samples analyzed. To this end, we compared the number of genes expressed by 1, 2-5 or >5 unique isoforms in cells bearing the mutation vs wild-type cells within each lineage, in order to account for possible differentiation-related differences. Results showed that, in all lineages, *SRSF2*-mutated cells carried significantly higher proportions of genes expressed with more than one isoform as compared to wild-type cells (Figure 68). Furthermore, isoforms expressed in mutated cells less often matched the corresponding reference transcript at all splice junctions, but were more often classified as novel (Figure 69), suggesting that the presence of a *SRSF2*-mutation increases the diversity of expression repertoire. However, we could not score major differences in terms of predicted coding potential (Figure 70) or non-sense mediated decay (Figure 71).

Figure 68. Isoform diversity in *SRSF2*-mutated vs wild-type cells by lineage.
The barplots show the numbers of genes expressed in mutated and wild-type cells in each aggregated lineage, categorized by number of expressed isoforms *per* gene. Two-sided Mann Whitney U test. ns = non significant, * = p<0.05, ** = p<0.01, *** = p<0.001.

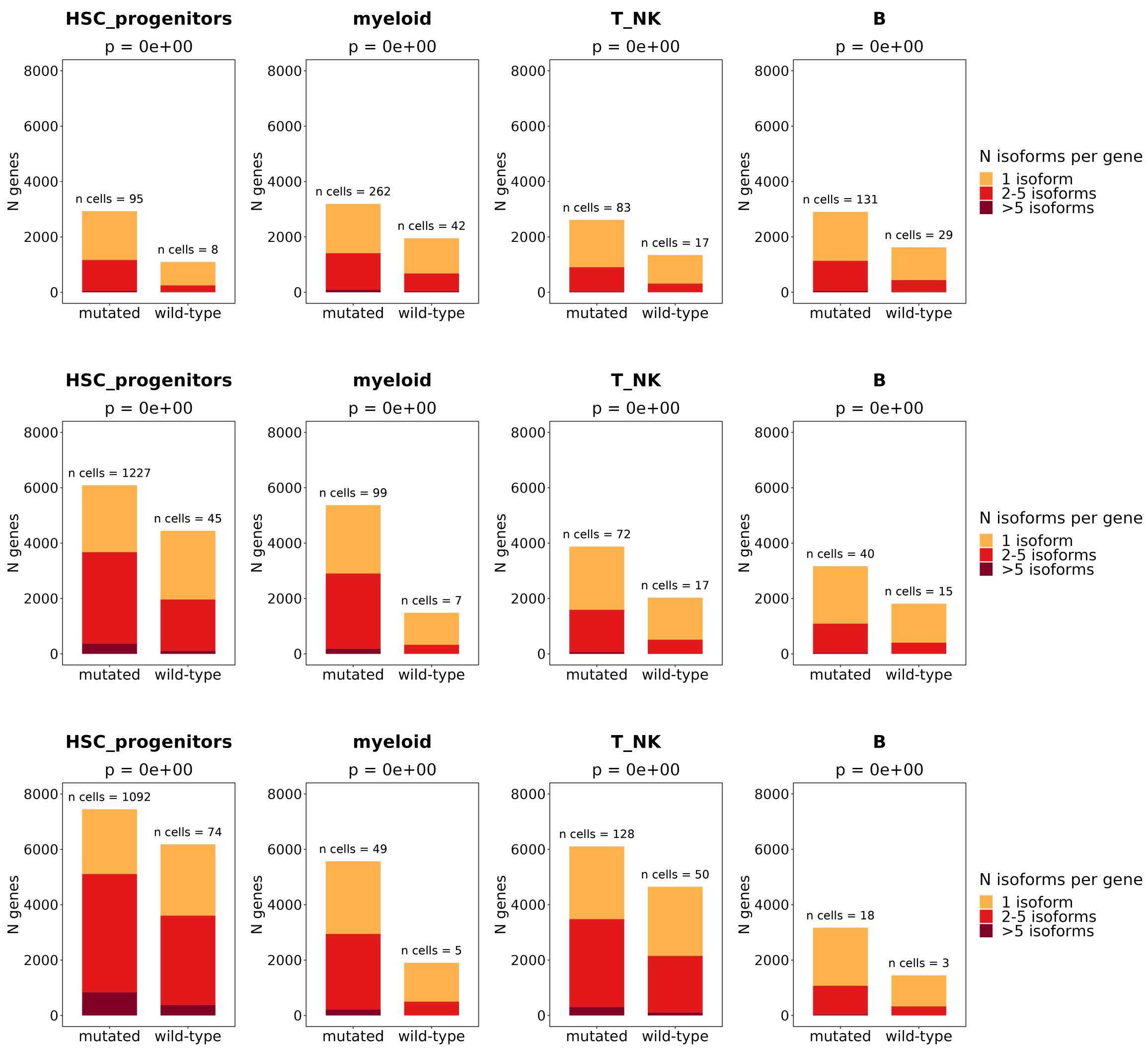


Figure 69. Isoform structural variation in *SRSF2*-mutated vs wild-type cells by lineage. The barplots show the numbers of isoforms expressed in mutated and wild-type cells in each aggregated lineage, according to legend categories. Two-sided Mann Whitney U test. ns = non significant, * = p<0.05, ** = p<0.01, *** = p<0.001.

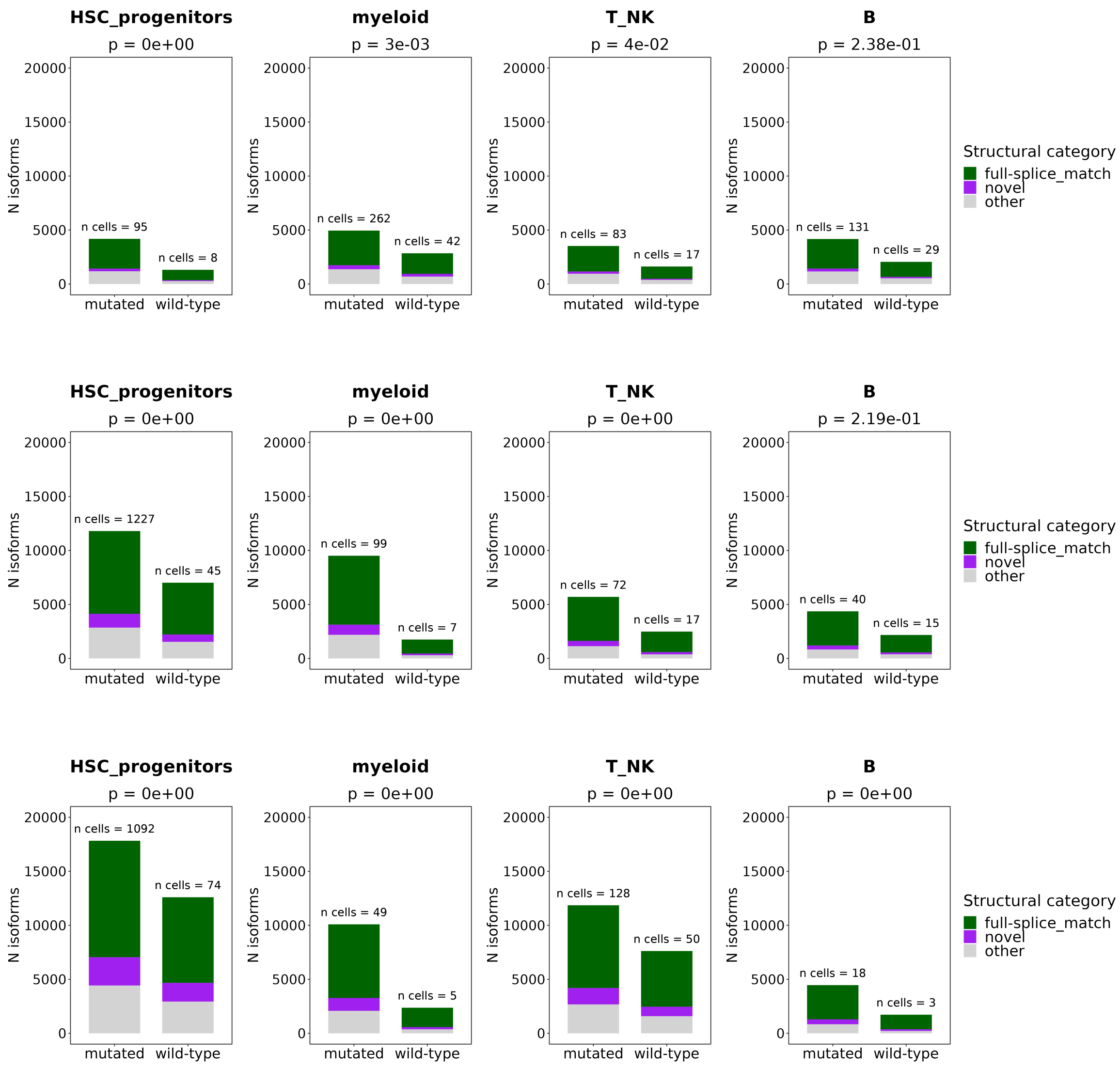


Figure 70. Isoform coding potential in *SRSF2*-mutated vs wild-type cells by lineage.
As in Figure 69.

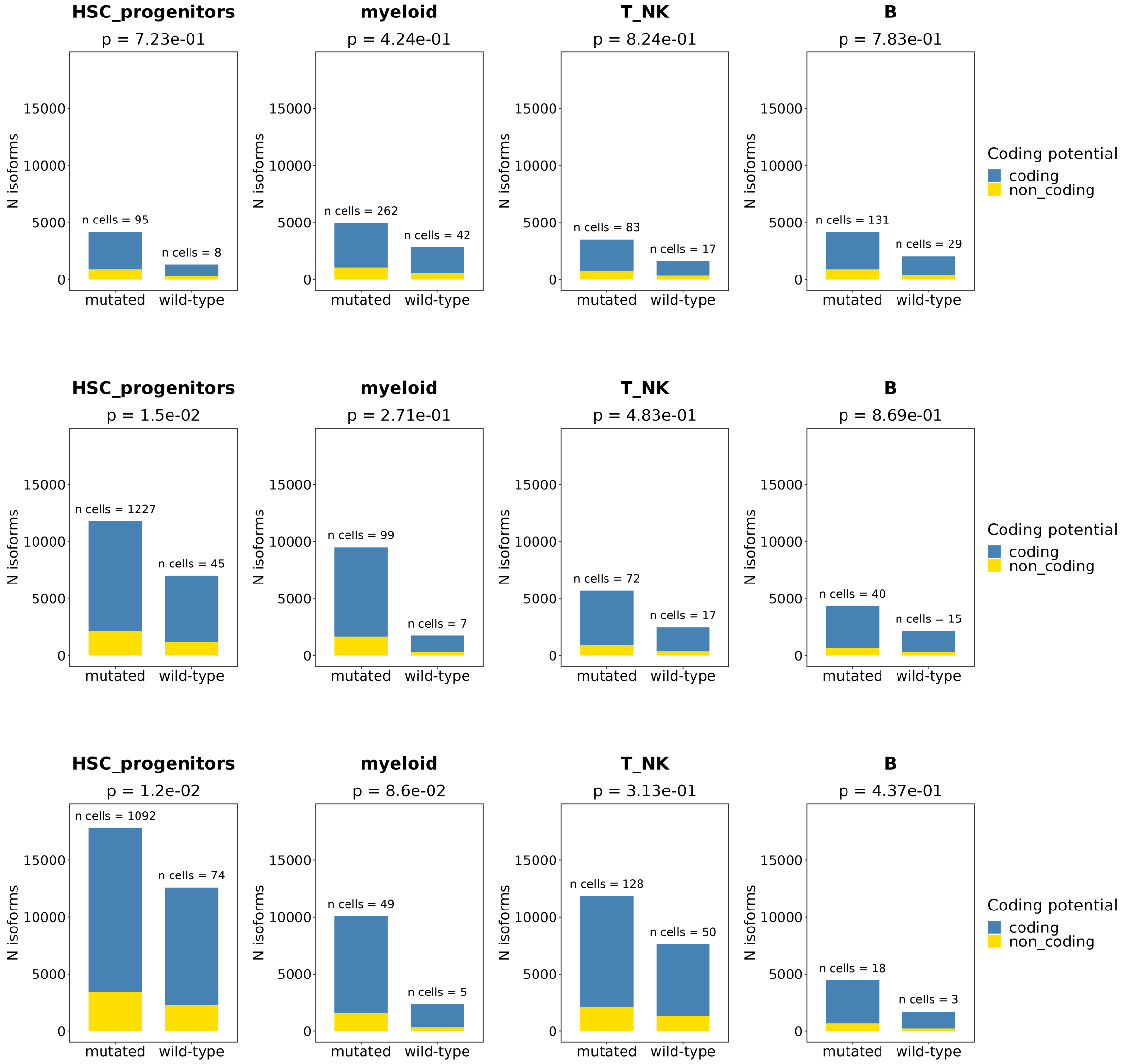
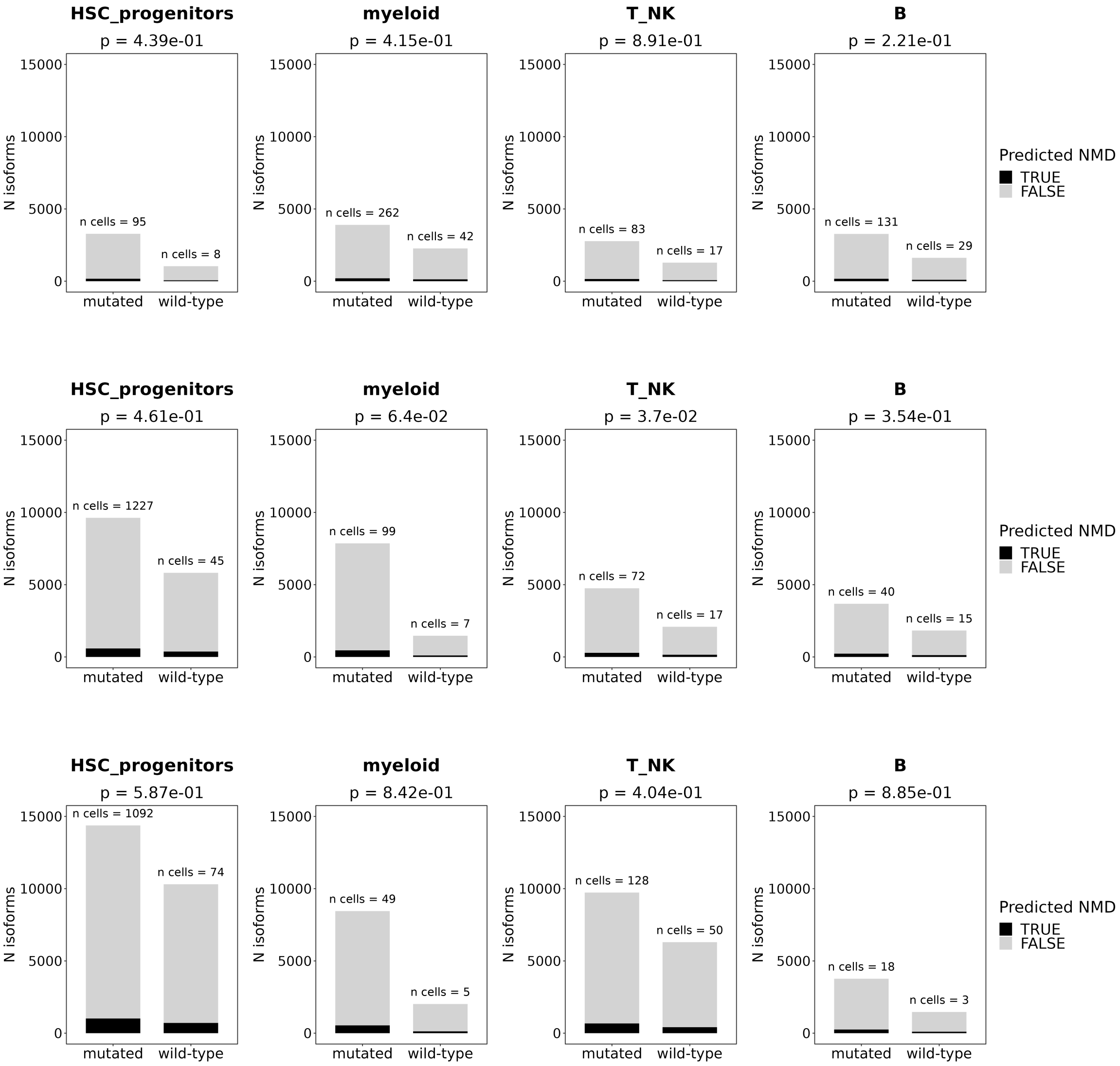


Figure 71. Isoform non-sense mediated decay in *SRSF2*-mutated vs wild-type cells by lineage.
As in Figure 69.



5. Discussion

Evolutionary dynamics and treatment resistance in AMLs are both heavily impacted by intra-tumor heterogeneity, which stems from distinct and interconnected cellular and biological levels. Solving the paths linking genetic events and phenotypic diversity might be of considerable importance to scan vulnerabilities for the development of meaningful therapeutic targets. This task, however, can be efficiently achieved only by applying multiomics approaches that integrate different layers of information at single-cell level. Ideally, any meaningful approach should be capable to profile high numbers of cells, thus allowing to capture both the malignant pool and the less-represented tumor-associated immune milieu, thus enabling a broad, ecosystem-wide characterization of different cancer traits.

Currently used droplet-based short-read scRNA-seq protocols achieve high cell throughput and are suitable to study heterogeneous cell populations and associated functional profiles. Yet, due to 3′ or 5′ end bias and consequent lack of transcript coverage of short-read sequencing (100), these protocols preclude reliable evaluation of other information, including expressed somatic mutations, whose impact in AML is well described, and transcript isoforms features, which might provide further insights into the phenotypic heterogeneity of AML. Long-read sequencing generates full-length transcript information in single cells and can overcome these limitations(126,152,153).

In this work we have developed SCM-seq, a multiomics method that combines the high-throughput of the short-read Chromium 10x platform, which is exploited for the isolation of single cells and mRNA sequencing, with the whole-transcript resolution provided by parallel ONT sequencing of full-length cDNA molecules. In particular, we took advantage of long-read ONT sequencing to access cell features that could not be reliably scored with short reads, namely transcript isoforms (from whole transcriptome sequencing) and somatic variants (from target enrichment of known mutated regions).

In recent years, a handful of experimental approaches have been devised to couple short-read scRNA-seq with long-read sequencing(126,133,154,143,153) but, to our knowledge, there is only one published example exploiting a similar approach for the joint single-cell analysis of gene mutations, expression and isoforms profiles(143). However, in comparison to SCM-seq, this method was severely impaired in the ability to capture biological complexity because the integrated analysis was only performed on a subsample of the total number of sequenced cells (10-20%). SCM-seq, instead, allowed to analyze genotype, gene expression and isoform profiles on 60.5%, 78.8% and 83% of cells in three AML samples, respectively. In principle, the SCM-seq platform can also be exploited to analyze simultaneously the T- (TCR) and B- (BCR) receptor profiles, thus offering a mean to efficiently investigate the connections between genetic profiles, immune clonotypes and functional responses within the tumor ecosystem. The sensitivity of our multiomics characterization (i.e., number of cells scored by multiple omics) also fairly compares to other recently published methods devised to couple transcriptional and mutational information(132,133,155–157).

We showed correspondence of transcriptional features between the two sequence datasets (10x and ONT), which was instrumental to prove the reliability and consistence of our integration analysis. With respect to somatic mutation detection, ONT sequencing data are known to be less accurate than Illumina, yet our data indicate that an average error rate of 5% at the variant position (across all cells) does not preclude mutation analysis, although this has to be assessed on a gene-by-gene basis. In fact, SCM-seq is inherently limited by the level of expression of the mutated gene, a limit that in our study allowed to score around half of the originally targeted mutations. This limit is attenuated by the questionable relevance of non-expressed mutations, as confirmed by our observation that the non-expressed and thus non-detected mutations involved genes that have not been described as AML drivers (i.e., are passenger mutations). Mutations transcribed into mRNA are indeed more likely to be translated, thus directly affecting cellular phenotypes and, eventually, actionability. For the expressed mutations, we have maximized coverage at the variant position by target enrichment, in order to build consensus sequences and contrast the possibly low expression patterns of mutated genes and the relatively low accuracy of ONT sequencing. Notably, this approach "lowered" the overall error rate at the variant position to <5% on average. Differences in read depth at the level of single-cells, however, cause a non-homogeneous distribution of the error rate, thus limiting the performance of our genotype imputation for rare cells, especially non-tumoral cells. As a future improvement, to mitigate the influence of varying levels of target gene expression in the detection of mutant alleles, we will perform genotype imputation after collapsing cells to single amplicon UMIs. Despite these limits, however, for several variants we were able to score both mutant and wild-type cells with high confidence, which is the basis for the systematic analyses of genotype-phenotype interactions.

Mutant cells were identified with high sensitivity and good correspondence with WES VAFs. Notably, we could score at least two mutations in more than half of mutant cells, a result that fairly compares to previous studies(132,133,155–157) and allowed us to stratify groups of cells based on their genetic complexity (i.e., numbers and combination of co-occurring mutations). AML cells with transcriptional features of HSCs/progenitors accumulated higher numbers of mutations and shared transcriptional features of LSCs, thus enabling identification of the putative malignant compartment. However, we found that mutations were also present in cells with transcriptional features of all lineages, including differentiated myeloid cells and lymphocytes. Importantly, the frequency of somatic variants in each lineage recapitulated the genetic hierarchy observed in HSCs, which supports a model of pre-leukemia clonal evolution in which serial mutations accumulate in self-renewing HSCs(31). We also observed that some mutations showed stronger association to specific lineages. In fact, gene mutations with higher VAFs and previously reported to occur in CH (e.g., *STAG2*, *SRSF2, DNMT3A*) tended to associate with lymphoid and differentiated-myeloid cells, consistently with earlier expansion in the non-leukemic compartment.

In healthy individuals with CH, somatic mutations are nearly always present in circulating innate immune cells and, less frequently and at lower VAFs, in B and T lymphocytes(59). It is still unclear whether and how mutations in the immune populations affect their cytotoxic vs immunosuppressive profiles, possibly altering surveillance against emerging tumor cells and response to immunotherapies. In our study, we did not identify significant changes in the expression of genes related to key immune functions in T and NK cells with increasing mutation complexity. Possibly, at the time of clinical diagnosis the immune microenvironment is irredeemably primed by the immunomodulatory properties of overt AML, which makes difficult scoring distinct functional subsets. In future analyses, we plan to investigate these aspects by systematic comparisons of the expression profiles of T and B lymphocytes from AML and normal BM. Intriguingly, we found rare cells in both the T and B immune compartment bearing more than 2-3 mutations and selectively associated to higher transcriptional heterogeneity, suggesting more advanced clonal stage than CH and distinct functional properties within the immune milieu. Reportedly, *DNMT3A*-mutated AMLs have been shown to originate from lympho-myeloid CH in up to 25% of cases(158), and it is possible that this phenomenon might involve also other gene mutations. One study showed that pre-leukemic clones (as defined by scRNA-seq and clonal tracking) contributed not only to HSC-like cells, but also to erythroid and lymphoid lineages(159). To confirm the lymphoid identity of these populations and study in deep their functional profiles, we will further validate their lineage by analysing TCR or BCR from the available long-read sequence.

With respect to the HSC/progenitor AML population, we have exploited the mutation detection sensitivity of SCM-seq to analyze the functional impact of cell-level mutation burden, regardless of the presence of specific combinations of mutations. This may become a critical trait of AMLs, since it provides a functional measure of the genetic heterogeneity that can be used to compare different leukemia samples. In previously published studies, the throughput of single-cell approaches for the parallel investigation of mutation and gene expression profiles was generally too low to enable such a characterization. Our results indicate that increasing mutation burden is systematically associated to increasing transcriptional heterogeneity, suggesting activation of independent expression programs upon serial acquisition of mutations and the existence of functional differences between early (i.e., with less mutations) vs late (i.e., with more mutations) cell clones. Further analyses is needed to investigate which gene modules are involved and if mutation-specific trait prevail. In the present study, hyper-mutated cells showed distinct transcriptional features as compared to low- or non-mutant cells, namely upregulation of genes and signatures related to cell-cycle control and proliferation, response to oxidative stress, DNA damage and repair, RNA splicing regulation, protein metabolism, MTORC1 signaling and MYC targets. Notably, these functional themes have been already described in AMLs and associated to adverse features and chemoresistance(160,161). We also observed homogeneous anti-correlation with inflammatory pathways, which reinforces the possibility that early and late cell clones exhibit distinct functional properties.

The identification of gene expression programs is typically instrumental and widely used to define phenotypic cell features. However, due to variable regulatory mechanisms including AS, gene expression alone is inherently limited in representing the actual functional impact of a gene's product, because transcript isoforms originating from the same mRNA may exhibit distinct - or even opposing - effects (162,163). To refine our analyses of the phenotypic heterogeneity in AML samples, and gain insights into transcriptional adaptation to genetic complexity, we coupled analyses of single-cell mutation burden to isoform profiling and diversity. Strikingly, we found that HSC/progenitor-like AML cells with high mutation burden displayed limited isoform abundance in proportion to the high number of expressed genes, indicating a progressively restricted repertoire of isoform generation in the presence of increasing mutation burden. Importantly, we could link this finding to the distinct gene expression profile we observed in cells with high mutation burden, which suggests the HSPC-like AML cells with highest genetic complexity undergo functional selection.

Finally, we have preliminarily explored the role of *SRSF2* mutations, which were shared between the three AML samples included in this work. *SRSF2* encodes for a spliceosome factor that is mutated in 10%-14% of AML patients and 20%-30% of MDS. The presence of mutations in this and other spliceosome factors is associated with adverse outcomes, including higher risk of MDS-to-AML progression and relapse after treatment(78). Although mutated *SRSF2* alone is not sufficient to promote leukemogenesis in *in vivo* model systems(164,165), it is still regarded as an ideal therapeutic target, due to its prognostic impact and their putative role in AML maintenance (these mutations are acquired early in leukemogenesis and often persists after treatment)(91,92). However, splicing is a fundamental biological process that also involves normal tissues, and the therapeutic window for splicing modulation may be narrow(166). More work is needed to understand the role and mechanisms of splicing abnormalities in AMLs, either mutation-dependent or independent. As a preliminary analysis, we assessed whether the presence of a mutation in *SRSF2* is associated to any change in isoforms diversity, and found that *SRSF2*-mutated AML cells carried significantly higher proportions of genes expressed with more than one isoform, as compared to *SRSF2*-wild-type AML cells, with high frequency of novel or alternative isoforms in the mutated cells. This is in line with previous studies reporting that splicing factor mutations do not cause loss of gene function (which may be lethal), but rather alter splicing preferences(164). Notably, we observed the same splicing pattern in all hematopoietic lineages. More investigation are needed to assess which specific splicing features distinguish mutated vs wild-type cells, as well their targets, in both leukemic and immune populations.

In conclusion, we have set up a single-cell multiomics method that allows integration of different sources of AML diversity. To show the potential of our approach, we exploited SCM-seq to highlight the relationships and relevance of genetic complexity and functional traits within the AML ecosystem, which, although preliminary, contributed solving the manifold paths of intra-tumor heterogeneity into common vulnerabilities. Importantly, in addition to the already mentioned analyses, the SCM-seq platform will enable us to pursue further studies. First, we will systematically test groups of cells bearing specific gene mutations or combinations with known prognostic value, in order to assess their functional impact and expand our view on possibly vulnerable pathways. On the same cells, we will exploit the availability of isoform-level characterization to study whether isoform diversity activates specific expression programs, along with differential transcript usage and associated splicing events. Finally, we will use isoform sequences to perform in silico translation and molecular docking simulations on predicted peptides, with the aim to screen their binding affinity for candidate drugs and immunotherapeutics. Such integrated approach will assist the effort of redirecting the hurdle of intra-tumor heterogeneity toward meaningful precision medicine strategies.

References

1. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood. 2016.
2. Arber DA, Orazi A, Hasserjian RP, Borowitz MJ, Calvo KR, Kvasnicka HM, et al. International Consensus Classification of Myeloid Neoplasms and Acute Leukemias: integrating morphologic, clinical, and genomic data. Blood. 2022 Sep 15;140(11):1200–28.
3. Döhner H, Wei AH, Appelbaum FR, Craddock C, DiNardo CD, Dombret H, et al. Diagnosis and Management of AML in Adults: 2022 ELN Recommendations from an International Expert Panel. Blood. 2022 Jul 7;blood.2022016867.
4. Dong Y, Shi O, Zeng Q, Lu X, Wang W, Li Y, et al. Leukemia incidence trends at the global, regional, and national level between 1990 and 2017. Exp Hematol Oncol. 2020 Dec;9(1):14.
5. Heuser M, Ofran Y, Boissel N, Brunet Mauri S, Craddock C, Janssen J, et al. Acute myeloid leukaemia in adult patients: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. Ann Oncol. 2020 Jun;31(6):697–712.
6. Lindsley RC, Mar BG, Mazzola E, Grauman PV, Shareef S, Allen SL, et al. Acute myeloid leukemia ontogeny is defined by distinct somatic mutations. Blood. 2015;
7. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med. 2016;
8. Bamopoulos SA, Batcha AMN, Jurinovic V, Rothenberg-Thurley M, Janke H, Ksienzyk B, et al. Clinical presentation and differential splicing of SRSF2, U2AF1 and SF3B1 mutations in patients with acute myeloid leukemia. Leukemia. 2020 Oct;34(10):2621–34.
9. Herold T, Rothenberg-Thurley M, Grunwald VV, Janke H, Goerlich D, Sauerland MC, et al. Validation and refinement of the revised 2017 European LeukemiaNet genetic risk stratification of acute myeloid leukemia. Leukemia. 2020 Dec;34(12):3161–72.
10. Lancet JE, Uy GL, Cortes JE, Newell LF, Lin TL, Ritchie EK, et al. Cpx-351 (cytarabine and daunorubicin) liposome for injection versus conventional cytarabine plus daunorubicin in older patients with newly diagnosed secondary acute myeloid leukemia. In: Journal of Clinical Oncology. 2018.
11. Stone RM, Mandrekar SJ, Sanford BL, Laumann K, Geyer S, Bloomfield CD, et al. Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a *FLT3* Mutation. N Engl J Med. 2017 Aug 3;377(5):454–64.
12. Cortes JE, Khaled S, Martinelli G, Perl AE, Ganguly S, Russell N, et al. Quizartinib versus salvage chemotherapy in relapsed or refractory FLT3-ITD acute myeloid leukaemia (QuANTUM-R): a multicentre, randomised, controlled, open-label, phase 3 trial. Lancet Oncol. 2019 Jul;20(7):984–97.
13. Schlenk RF, Paschka P, Krzykalla J, Weber D, Kapp-Schwoerer S, Gaidzik VI, et al. Gemtuzumab Ozogamicin in *NPM1* -Mutated Acute Myeloid Leukemia: Early Results From the Prospective Randomized AMLSG 09-09 Phase III Study. J Clin Oncol. 2020 Feb 20;38(6):623–32.
14. Kantarjian HM, Thomas XG, Dmoszynska A, Wierzbowska A, Mazur G, Mayer J, et al. Multicenter, Randomized, Open-Label, Phase III Trial of Decitabine Versus Patient Choice, With Physician Advice, of Either Supportive Care or Low-Dose Cytarabine for the Treatment of Older Patients With Newly Diagnosed Acute Myeloid Leukemia. J Clin Oncol. 2012 Jul 20;30(21):2670–7.
15. Dombret H, Seymour JF, Butrym A, Wierzbowska A, Selleslag D, Jang JH, et al. International phase 3 study of azacitidine vs conventional care regimens in older patients with newly diagnosed AML with >30% blasts. Blood. 2015 Jul 16;126(3):291–9.
16. DiNardo CD, Pratz K, Pullarkat V, Jonas BA, Arellano M, Becker PS, et al. Venetoclax combined with decitabine or azacitidine in treatment-naive, elderly patients with acute myeloid leukemia. Blood. 2019;
17. Montesinos P, Recher C, Vives S, Zarzycka E, Wang J, Bertani G, et al. Ivosidenib and Azacitidine in *IDH1* -Mutated Acute Myeloid Leukemia. N Engl J Med. 2022 Apr 21;386(16):1519–31.
18. Cornelissen JJ, Blaise D. Hematopoietic stem cell transplantation for patients with AML in first complete remission. Blood. 2016 Jan 7;127(1):62–70.
19. Schuurhuis GJ, Heuser M, Freeman S, Béné MC, Buccisano F, Cloos J, et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. Blood. 2018 Mar 22;131(12):1275–91.

20. Heuser M, Freeman SD, Ossenkoppele GJ, Buccisano F, Hourigan CS, Ngai LL, et al. 2021 Update on MRD in acute myeloid leukemia: a consensus document from the European LeukemiaNet MRD Working Party. Blood. 2021 Dec 30;138(26):2753–67.
21. Lussana F, Caprioli C, Stefanoni P, Pavoni C, Spinelli O, Buklijas K, et al. Molecular Detection of Minimal Residual Disease before Allogeneic Stem Cell Transplantation Predicts a High Incidence of Early Relapse in Adult Patients with NPM1 Positive Acute Myeloid Leukemia. Cancers. 2019 Sep 28;11(10):1455.
22. Zilberberg J, Feinman R, Korngold R. Strategies for the Identification of T Cell–Recognized Tumor Antigens in Hematological Malignancies for Improved Graft-versus-Tumor Responses after Allogeneic Blood and Marrow Transplantation. Biol Blood Marrow Transplant. 2015 Jun;21(6):1000–7.
23. Huls G, Chitu DA, Havelange V, Jongen-Lavrencic M, van de Loosdrecht AA, Biemond BJ, et al. Azacitidine maintenance after intensive chemotherapy improves DFS in older AML patients. Blood. 2019 Mar 28;133(13):1457–64.
24. Wei AH, Döhner H, Pocock C, Montesinos P, Afanasyev B, Dombret H, et al. Oral Azacitidine Maintenance Therapy for Acute Myeloid Leukemia in First Remission. N Engl J Med. 2020 Dec 24;383(26):2526–37.
25. Döhner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. Longo DL, editor. N Engl J Med. 2015 Sep 17;373(12):1136–52.
26. Perl AE, Larson RA, Podoltsev NA, Strickland S, Wang ES, Atallah E, et al. Follow-up of patients with R/R *FLT3*- mutation–positive AML treated with gilteritinib in the phase 3 ADMIRAL trial. Blood. 2022 Jun 9;139(23):3366–75.
27. DiNardo CD, Stein EM, de Botton S, Roboz GJ, Altman JK, Mims AS, et al. Durable Remissions with Ivosidenib in *IDH1* -Mutated Relapsed or Refractory AML. N Engl J Med. 2018 Jun 21;378(25):2386–98.
28. Stein EM, DiNardo CD, Pollyea DA, Fathi AT, Roboz GJ, Altman JK, et al. Enasidenib in mutant IDH2 relapsed or refractory acute myeloid leukemia. Blood. 2017 Aug 10;130(6):722–31.
29. Shlush LI, Mitchell A, Heisler L, Abelson S, Ng SWK, Trotman-Grant A, et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. Nature. 2017 Jul;547(7661):104–8.
30. Lichtenegger FS, Krupka C, Haubner S, Köhnke T, Subklewe M. Recent developments in immunotherapy of acute myeloid leukemia. J Hematol OncolJ Hematol Oncol. 2017 Dec;10(1):142.
31. Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR, et al. Clonal Evolution of Preleukemic Hematopoietic Stem Cells Precedes Human Acute Myeloid Leukemia. Sci Transl Med [Internet]. 2012 Aug 29 [cited 2022 Sep 9];4(149). Available from: https://www.science.org/doi/10.1126/scitranslmed.3004315
32. Desai P, Mencia-Trinchant N, Savenkov O, Simon MS, Cheang G, Lee S, et al. Somatic mutations precede acute myeloid leukemia years before diagnosis. Nat Med. 2018 Jul;24(7):1015–23.
33. The Cancer Genome Atlas Research Network. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. N Engl J Med. 2013 May 30;368(22):2059–74.
34. Metzeler KH, Herold T, Rothenberg-Thurley M, Amler S, Sauerland MC, Görlich D, et al. Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia. In: Blood. 2016.
35. Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, et al. Functional genomic landscape of acute myeloid leukaemia. Nature. 2018 Oct;562(7728):526–31.
36. Bullinger L, Döhner K, Dohner H. Genomics of acute myeloid leukemia diagnosis and pathways. Journal of Clinical Oncology. 2017.
37. Galanis A, Ma H, Rajkhowa T, Ramachandran A, Small D, Cortes J, et al. Crenolanib is a potent inhibitor of FLT3 with activity against resistance-conferring point mutants. Blood. 2014 Jan 2;123(1):94–100.
38. Borthakur G, Popplewell L, Boyiadzis M, Foran J, Platzbecker U, Vey N, et al. Activity of the oral mitogen-activated protein kinase kinase inhibitor trametinib in *RAS* -mutant relapsed or refractory myeloid malignancies: Trametinib in *RAS* -Mutant Malignancies. Cancer. 2016 Jun 15;122(12):1871–9.
39. Seiler M, Yoshimi A, Darman R, Chan B, Keaney G, Thomas M, et al. H3B-8800, an orally available small-molecule splicing modulator, induces lethality in spliceosome-mutant cancers. Nat Med. 2018 Apr;24(4):497–504.
40. Hamard PJ, Santiago GE, Liu F, Karl DL, Martinez C, Man N, et al. PRMT5 Regulates DNA Repair by Controlling the Alternative Splicing of Histone-Modifying Enzymes. Cell Rep. 2018 Sep;24(10):2643–57.

41. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012 Jan;481(7382):506–10.
42. Li S, Garrett-Bakelman FE, Chung SS, Sanders MA, Hricik T, Rapaport F, et al. Distinct evolution and dynamics of epigenetic and genetic heterogeneity in acute myeloid leukemia. Nat Med. 2016 Jul;22(7):792–9.
43. Morita K, Wang F, Jahn K, Hu T, Tanaka T, Sasaki Y, et al. Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics. Nat Commun. 2020 Dec;11(1):5327.
44. Daver N, Cortes J, Ravandi F, Patel KP, Burger JA, Konopleva M, et al. Secondary mutations as mediators of resistance to targeted therapy in leukemia. Blood. 2015 May 21;125(21):3236–45.
45. Joshi SK, Nechiporuk T, Bottomly D, Piehowski PD, Reisz JA, Pittsenbarger J, et al. The AML microenvironment catalyzes a stepwise evolution to gilteritinib resistance. Cancer Cell. 2021;39(7):999-1014.e8.
46. Klco JM, Spencer DH, Miller CA, Griffith M, Lamprecht TL, O'Laughlin M, et al. Functional Heterogeneity of Genetically Defined Subclones in Acute Myeloid Leukemia. Cancer Cell. 2014 Mar;25(3):379–92.
47. McKerrell T, Park N, Chi J, Collord G, Moreno T, Ponstingl H, et al. JAK2 V617F hematopoietic clones are present several years prior to MPN diagnosis and follow different expansion kinetics. Blood Adv. 2017 Jun 13;1(14):968–71.
48. Young AL, Challen GA, Birmann BM, Druley TE. Clonal haematopoiesis harbouring AML-associated mutations is ubiquitous in healthy adults. Nat Commun. 2016 Nov;7(1):12484.
49. Lachowiez CA, Loghavi S, Furudate K, Montalban-Bravo G, Maiti A, Kadia T, et al. Impact of splicing mutations in acute myeloid leukemia treated with hypomethylating agents combined with venetoclax. Blood Adv. 2021 Apr 27;5(8):2173–83.
50. Pollyea DA, Jordan CT. Therapeutic targeting of acute myeloid leukemia stem cells. Blood. 2017 Mar 23;129(12):1627–35.
51. Kreso A, Dick JE. Evolution of the Cancer Stem Cell Model. Cell Stem Cell. 2014 Mar;14(3):275–91.
52. Ng SWK, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature. 2016 Dec 15;540(7633):433–7.
53. Fennell KA, Vassiliadis D, Lam EYN, Martelotto LG, Balic JJ, Hollizeck S, et al. Non-genetic determinants of malignant clonal fitness at single-cell resolution. Nature. 2022 Jan 6;601(7891):125–31.
54. Zeng AGX, Bansal S, Jin L, Mitchell A, Chen WC, Abbas HA, et al. A cellular hierarchy framework for understanding heterogeneity and predicting drug response in acute myeloid leukemia. Nat Med. 2022 Jun;28(6):1212–23.
55. Tettamanti S, Pievani A, Biondi A, Dotti G, Serafini M. Catch me if you can: how AML and its niche escape immunotherapy. Leukemia. 2021;(July):1–10.
56. Vago L, Gojo I. Immune escape and immunotherapy of acute myeloid leukemia. J Clin Invest. 2020;130(4):1552–64.
57. Agarwal P, Isringhausen S, Li H, Paterson AJ, He J, Gomariz Á, et al. Mesenchymal Niche-Specific Expression of Cxcl12 Controls Quiescence of Treatment-Resistant Leukemia Stem Cells. Cell Stem Cell. 2019 May;24(5):769-784.e6.
58. Trowbridge JJ, Starczynowski DT. Innate immune pathways and inflammation in hematopoietic aging, clonal hematopoiesis, and MDS. J Exp Med. 2021;218(7):1–15.
59. Arends CM, Galan-Sousa J, Hoyer K, Chan W, Jäger M, Yoshida K, et al. Hematopoietic lineage distribution and evolutionary dynamics of clonal hematopoiesis. Leukemia. 2018 Sep;32(9):1908–19.
60. Cai Z, Kotzin JJ, Ramdas B, Chen S, Nelanuthala S, Palam LR, et al. Inhibition of Inflammatory Signaling in Tet2 Mutant Preleukemic Cells Mitigates Stress-Induced Abnormalities and Clonal Hematopoiesis. Cell Stem Cell. 2018 Dec;23(6):833-849.e5.
61. Meisel M, Hinterleitner R, Pacis A, Chen L, Earley ZM, Mayassi T, et al. Microbial signals drive pre-leukaemic myeloproliferation in a Tet2-deficient host. Nature. 2018 May;557(7706):580–4.
62. Warren JT, Link DC. Clonal hematopoiesis and risk for hematologic malignancy. Blood. 2020;136(14):1599–605.
63. Sloand EM, Melenhorst JJ, Tucker ZCG, Pfannes L, Brenchley JM, Yong A, et al. T-cell immune responses to Wilms tumor 1 protein in myelodysplasia responsive to immunosuppressive therapy. Blood. 2011 Mar 3;117(9):2691–9.

64. Elias S, Yamin R, Golomb L, Tsukerman P, Stanietsky-Kaynan N, Ben-Yehuda D, et al. Immune evasion by oncogenic proteins of acute myeloid leukemia. Blood. 2014 Mar 6;123(10):1535–43.
65. Prestipino A, Emhardt AJ, Aumann K, O'Sullivan D, Gorantla SP, Duquesne S, et al. Oncogenic JAK2 V617F causes PD-L1 expression, mediating immune escape in myeloproliferative neoplasms. Sci Transl Med. 2018 Feb;10(429):eaam7729.
66. Holmström MO, Riley CH, Svane IM, Hasselbalch HC, Andersen MH. The CALR exon 9 mutations are shared neoantigens in patients with CALR mutant chronic myeloproliferative neoplasms. Leukemia. 2016 Dec;30(12):2413–6.
67. Adamia S, Bar-Natan M, Haibe-Kains B, Pilarski PM, Bach C, Pevzner S, et al. NOTCH2 and FLT3 gene mis-splicings are common events in patients with acute myeloid leukemia (AML): new potential targets in AML. Blood. 2014 May 1;123(18):2816–25.
68. Miles LA, Bowman RL, Merlinsky TR, Csete IS, Ooi AT, Durruthy-Durruthy R, et al. Single-cell mutation analysis of clonal evolution in myeloid malignancies. Nature. 2020;587(7834):477–82.
69. Christopher MJ, Petti AA, Rettig MP, Miller CA, Chendamarai E, Duncavage EJ, et al. Immune Escape of Relapsed AML Cells after Allogeneic Transplantation. N Engl J Med. 2018 Dec 13;379(24):2330–41.
70. Toffalori C, Zito L, Gambacorta V, Riba M, Oliveira G, Bucci G, et al. Immune signature drives leukemia escape and relapse after hematopoietic cell transplantation. Nat Med. 2019 Apr;25(4):603–11.
71. Boddu P, Kantarjian H, Garcia-Manero G, Allison J, Sharma P, Daver N. The emerging role of immune checkpoint based approaches in AML and MDS. Leuk Lymphoma. 2018;59(4):790–802.
72. Wirth TC, Kühnel F. Neoantigen Targeting—Dawn of a New Era in Cancer Immunotherapy? Front Immunol. 2017 Dec 19;8:1848.
73. Vadakekolathu J, Minden MD, Hood T, Church SE, Reeder S, Altmann H, et al. Immune landscapes predict chemotherapy resistance and immunotherapy response in acute myeloid leukemia. Sci Transl Med. 2020 Jun 3;12(546):eaaz0463.
74. Rutella S, Vadakekolathu J, Mazziotta F, Reeder S, Yau TO, Mukhopadhyay R, et al. Signatures of immune senescence predict outcomes and define checkpoint blockade-unresponsive microenvironments in acute myeloid leukemia [Internet]. Hematology; 2022 Feb [cited 2022 Sep 8]. Available from: http://medrxiv.org/lookup/doi/10.1101/2022.02.08.22270578
75. Chiara Caprioli, Federico Lussana, Silvia Salmoiraghi, Roberta Cavagna, Ksenija Buklijas, Lara Elidi, et al. Clinical significance of chromatin-spliceosome acute myeloid leukemia: a report from the Northern Italy Leukemia Group (NILG) randomized trial 02/06. Haematologica. 2020 Aug 27;106(10):2578–87.
76. Wahl MC, Lührmann R. SnapShot: Spliceosome Dynamics I. Cell. 2015 Jun;161(6):1474-1474.e1.
77. Wilkinson ME, Charenton C, Nagai K. RNA Splicing by the Spliceosome. Annu Rev Biochem. 2020 Jun 20;89(1):359–88.
78. Chen S, Benbarche S, Abdel-Wahab O. Splicing factor mutations in hematologic malignancies. Blood. 2021 Aug 26;138(8):599–612.
79. Papaemmanuil E, Gerstung M, Malcovati L, Tauro S, Gundem G, Van Loo P, et al. Clinical and biological implications of driver mutations in myelodysplastic syndromes. Blood. 2013;
80. Yoshida K, Sanada M, Shiraishi Y, Nowak D, Nagata Y, Yamamoto R, et al. Frequent pathway mutations of splicing machinery in myelodysplasia. Nature. 2011 Oct;478(7367):64–9.
81. Seiler M, Peng S, Agrawal AA, Palacino J, Teng T, Zhu P, et al. Somatic Mutational Landscape of Splicing Factor Genes and Their Functional Consequences across 33 Cancer Types. Cell Rep. 2018 Apr;23(1):282-296.e4.
82. Yoshimi A, Lin KT, Wiseman DH, Rahman MA, Pastore A, Wang B, et al. Coordinated alterations in RNA splicing and epigenetic regulation drive leukaemogenesis. Nature. 2019 Oct 10;574(7777):273–7.
83. Chen L, Chen JY, Huang YJ, Gu Y, Qiu J, Qian H, et al. The Augmented R-Loop Is a Unifying Mechanism for Myelodysplastic Syndromes Induced by High-Risk Splicing Factor Mutations. Mol Cell. 2018 Feb;69(3):412-425.e6.
84. Flynn RL, Zou L. ATR: a master conductor of cellular responses to DNA replication stress. Trends Biochem Sci. 2011 Mar;36(3):133–40.
85. Rahman MA, Lin KT, Bradley RK, Abdel-Wahab O, Krainer AR. Recurrent SRSF2 mutations in MDS affect both splicing and NMD. Genes Dev. 2020 Mar 1;34(5–6):413–27.
86. Climente-González H, Porta-Pardo E, Godzik A, Eyras E. The Functional Impact of Alternative Splicing in Cancer. Cell Rep. 2017 Aug;20(9):2215–26.

87. Jayasinghe RG, Cao S, Gao Q, Wendl MC, Vo NS, Reynolds SM, et al. Systematic Analysis of Splice-Site-Creating Mutations in Cancer. Cell Rep. 2018 Apr;23(1):270-281.e3.
88. Kahles A, Lehmann KV, Toussaint NC, Hüser M, Stark SG, Sachsenberg T, et al. Comprehensive Analysis of Alternative Splicing Across Tumors from 8,705 Patients. Cancer Cell. 2018 Aug 13;34(2):211-224.e6.
89. Crews LA, Balaian L, Delos Santos NP, Leu HS, Court AC, Lazzari E, et al. RNA Splicing Modulation Selectively Impairs Leukemia Stem Cell Maintenance in Secondary Human AML. Cell Stem Cell. 2016 Nov;19(5):599–612.
90. Anande G, Deshpande NP, Mareschal S, Batcha AMN, Hampton HR, Herold T, et al. RNA Splicing Alterations Induce a Cellular Stress Response Associated with Poor Prognosis in Acute Myeloid Leukemia. Clin Cancer Res Off J Am Assoc Cancer Res. 2020 Jul 15;26(14):3597–607.
91. Wu SJ, Kuo YY, Hou HA, Li LY, Tseng MH, Huang CF, et al. The clinical implication of SRSF2 mutation in patients with myelodysplastic syndrome and its stability during disease evolution. Blood. 2012 Oct 11;120(15):3106–11.
92. Mossner M, Jann JC, Wittig J, Nolte F, Fey S, Nowak V, et al. Mutational hierarchies in myelodysplastic syndromes dynamically adapt and evolve upon therapy response and failure. Blood. 2016 Sep 1;128(9):1246–59.
93. Lee SCW, Dvinge H, Kim E, Cho H, Micol JB, Chung YR, et al. Modulation of splicing catalysis for therapeutic targeting of leukemia with mutations in genes encoding spliceosomal proteins. Nat Med. 2016 Jun;22(6):672–8.
94. Lee SCW, Abdel-Wahab O. Therapeutic targeting of splicing in cancer. Nat Med. 2016 Sep;22(9):976–86.
95. Steensma DP, Wermke M, Klimek VM, Greenberg PL, Font P, Komrokji RS, et al. Phase I First-in-Human Dose Escalation Study of the oral SF3B1 modulator H3B-8800 in myeloid neoplasms. Leukemia. 2021 Dec;35(12):3542–50.
96. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, et al. Robust enumeration of cell subsets from tissue expression profiles. Nat Methods. 2015 May;12(5):453–7.
97. Navin NE. Cancer genomics: one cell at a time. Genome Biol. 2014 Aug;15(8):452.
98. Giladi A, Amit I. Single-Cell Genomics: A Stepping Stone for Future Immunology Discoveries. Cell. 2018;172(1–2):14–21.
99. Fan J, Slowikowski K, Zhang F. Single-cell transcriptomics in cancer: computational challenges and opportunities. Exp Mol Med. 2020;52(9):1452–65.
100. Zheng GXY, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017 Apr;8(1):14049.
101. Deng Q, Ramskold D, Reinius B, Sandberg R. Single-Cell RNA-Seq Reveals Dynamic, Random Monoallelic Gene Expression in Mammalian Cells. Science. 2014 Jan 10;343(6167):193–6.
102. Macosko EZ, Basu A, Satija R, Nemesh J, Shekhar K, Goldman M, et al. Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell. 2015 May;161(5):1202–14.
103. Paul F, Arkin Y, Giladi A, Jaitin DA, Kenigsberg E, Keren-Shaul H, et al. Transcriptional Heterogeneity and Lineage Commitment in Myeloid Progenitors. Cell. 2015 Dec;163(7):1663–77.
104. Ziegenhain C, Vieth B, Parekh S, Reinius B, Guillaumet-Adkins A, Smets M, et al. Comparative Analysis of Single-Cell RNA Sequencing Methods. Mol Cell. 2017 Feb;65(4):631-643.e4.
105. Ramsköld D, Luo S, Wang YC, Li R, Deng Q, Faridani OR, et al. Full-length mRNA-Seq from single-cell levels of RNA and individual circulating tumor cells. Nat Biotechnol. 2012 Aug;30(8):777–82.
106. Picelli S, Faridani OR, Björklund ÅK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014 Jan;9(1):171–81.
107. Shalek AK, Satija R, Adiconis X, Gertner RS, Gaublomme JT, Raychowdhury R, et al. Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. Nature. 2013 Jun 13;498(7453):236–40.
108. Eltahla AA, Rizzetto S, Pirozyan MR, Betz-Stablein BD, Venturi V, Kedzierska K, et al. Linking the T cell receptor to the single cell transcriptome in antigen-specific human T cells. Immunol Cell Biol. 2016 Jul;94(6):604–11.
109. Afik S, Yates KB, Bi K, Darko S, Godec J, Gerdemann U, et al. Targeted reconstruction of T cell receptor sequence from single cell RNA-seq links CDR3 length to T cell differentiation state. Nucleic Acids Res. 2017 Sep 19;45(16):e148–e148.

110. Rizzetto S, Koppstein DNP, Samir J, Singh M, Reed JH, Cai CH, et al. B-cell receptor reconstruction from single-cell RNA-seq with VDJPuzzle. Kelso J, editor. Bioinformatics. 2018 Aug 15;34(16):2846–7.
111. Islam S, Zeisel A, Joost S, La Manno G, Zajac P, Kasper M, et al. Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods. 2014 Feb;11(2):163–6.
112. Haghverdi L, Buettner F, Theis FJ. Diffusion maps for high-dimensional single-cell analysis of differentiation data. Bioinformatics. 2015 Sep 15;31(18):2989–98.
113. Luecken MD, Theis FJ. Current best practices in single-cell RNA-seq analysis: a tutorial. Mol Syst Biol [Internet]. 2019 Jun [cited 2021 Sep 29];15(6). Available from: https://onlinelibrary.wiley.com/doi/10.15252/msb.20188746
114. Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, et al. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. Nat Biotechnol. 2014 Apr;32(4):381–6.
115. Pellegrino M, Sciambi A, Treusch S, Durruthy-Durruthy R, Gokhale K, Jacob J, et al. High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. Genome Res. 2018 Sep;28(9):1345–52.
116. van den Brink SC, Sage F, Vértesy Á, Spanjaard B, Peterson-Maduro J, Baron CS, et al. Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. Nat Methods. 2017 Oct;14(10):935–6.
117. Lähnemann D, Köster J, Szczurek E, McCarthy DJ, Hicks SC, Robinson MD, et al. Eleven grand challenges in single-cell data science. Genome Biol. 2020 Dec;21(1):31.
118. Butler A, Hoffman P, Smibert P, Papalexi E, Satija R. Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat Biotechnol. 2018 May;36(5):411–20.
119. Stuart T, Butler A, Hoffman P, Hafemeister C, Papalexi E, Mauck WM, et al. Comprehensive Integration of Single-Cell Data. Cell. 2019 Jun;177(7):1888-1902.e21.
120. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021 Jun;184(13):3573-3587.e29.
121. Argelaguet R, Cuomo ASE, Stegle O, Marioni JC. Computational principles and challenges in single-cell data integration. Nat Biotechnol [Internet]. 2021 May 3 [cited 2021 Sep 29]; Available from: http://www.nature.com/articles/s41587-021-00895-7
122. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, et al. The Human Cell Atlas [Internet]. Cell Biology; 2017 May [cited 2021 Dec 17]. Available from: http://biorxiv.org/lookup/doi/10.1101/121202
123. Hay SB, Ferchen K, Chetal K, Grimes HL, Salomonis N. The Human Cell Atlas bone marrow single-cell interactive web portal. Exp Hematol. 2018;68:51–61.
124. Macaulay IC, Haerty W, Kumar P, Li YI, Hu TX, Teng MJ, et al. G&T-seq: parallel sequencing of single-cell genomes and transcriptomes. Nat Methods. 2015 Jun;12(6):519–22.
125. Dey SS, Kester L, Spanjaard B, Bienko M, van Oudenaarden A. Integrated genome and transcriptome sequencing of the same cell. Nat Biotechnol. 2015 Mar;33(3):285–9.
126. Singh M, Al-Eryani G, Carswell S, Ferguson JM, Blackburn J, Barton K, et al. High-throughput targeted long-read single cell sequencing reveals the clonal and transcriptional landscape of lymphocytes. Nat Commun [Internet]. 2019;10(1). Available from: http://dx.doi.org/10.1038/s41467-019-11049-4
127. Zhang Z, Xiong D, Wang X, Liu H, Wang T. Mapping the functional landscape of T cell receptor repertoires by single-T cell transcriptomics. Nat Methods. 2021;18(1):92–9.
128. Stoeckius M, Hafemeister C, Stephenson W, Houck-Loomis B, Chattopadhyay PK, Swerdlow H, et al. Simultaneous epitope and transcriptome measurement in single cells. Nat Methods. 2017 Sep;14(9):865–8.
129. Peterson VM, Zhang KX, Kumar N, Wong J, Li L, Wilson DC, et al. Multiplexed quantification of proteins and transcripts in single cells. Nat Biotechnol. 2017 Oct;35(10):936–9.
130. Frei AP, Bava FA, Zunder ER, Hsieh EWY, Chen SY, Nolan GP, et al. Highly multiplexed simultaneous detection of RNAs and proteins in single cells. Nat Methods. 2016 Mar;13(3):269–75.
131. Gerlach JanP, van Buggenum JAG, Tanis SEJ, Hogeweg M, Heuts BMH, Muraro MJ, et al. Combined quantification of intracellular (phospho-)proteins and transcriptomics from fixed single cells. Sci Rep. 2019 Dec;9(1):1469.
132. Giustacchini A, Thongjuea S, Barkas N, Woll PS, Povinelli BJ, Booth CAG, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. Nat Med. 2017;23(6):692–702.

133. van Galen P, Hovestadt V, Wadsworth MH, Hughes TK, Griffin GK, Battaglia S, et al. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. Cell. 2019;176(6):1265-1281.e24.
134. Germain PL, Lun A, Macnair W, Robinson MD. Doublet identification in single-cell sequencing data using scDblFinder. F1000Research. 2021 Sep 28;10:979.
135. Satija R, Farrell JA, Gennert D, Schier AF, Regev A. Spatial reconstruction of single-cell gene expression data. Nat Biotechnol. 2015 May;33(5):495–502.
136. Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 2016 Apr 8;352(6282):189–96.
137. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J Integr Biol. 2012 May;16(5):284–7.
138. Yu G, Li F, Qin Y, Bo X, Wu Y, Wang S. GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. Bioinformatics. 2010 Apr 1;26(7):976–8.
139. Azizi E, Carr AJ, Plitas G, Cornish AE, Konopacki C, Prabhakaran S, et al. Single-Cell Map of Diverse Immune Phenotypes in the Breast Tumor Microenvironment. Cell. 2018 Aug;174(5):1293-1308.e36.
140. Mujahed H, Jansson M, Bengtzén S, Lehamnn S. Bone marrow stroma cells derived from mononuclear cells at diagnosis as a source of germline control DNA for determination of somatic mutations in acute myeloid leukemia. Blood Cancer J. 2017 Oct;7(10):e616–e616.
141. Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, et al. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. Curr Protoc Bioinforma [Internet]. 2013 Oct [cited 2022 Oct 6];43(1). Available from: https://onlinelibrary.wiley.com/doi/10.1002/0471250953.bi1110s43
142. Leger A, Leonardi T. pycoQC, interactive quality control for Oxford Nanopore Sequencing. J Open Source Softw. 2019 Feb 28;4(34):1236.
143. Tian L, Jabbari JS, Thijssen R, Gouil Q, Amarasinghe SL, Voogd O, et al. Comprehensive characterization of single-cell full-length isoforms in human and mouse with long-read sequencing. Genome Biol. 2021 Dec;22(1):310.
144. Tian L, Su S, Dong X, Amann-Zalcenstein D, Biben C, Seidi A, et al. scPipe: A flexible R/Bioconductor preprocessing pipeline for single-cell RNA-sequencing data. Pertea M, editor. PLOS Comput Biol. 2018 Aug 10;14(8):e1006361.
145. Hummel M, Bonnin S, Lowy E, Roma G. TEQC: an R package for quality control in target capture experiments. Bioinformatics. 2011 May 1;27(9):1316–7.
146. Kono N, Arakawa K. Nanopore sequencing: Review of potential applications in functional genomics. Dev Growth Differ. 2019 Jun;61(5):316–26.
147. Ginart P, Kalish JM, Jiang CL, Yu AC, Bartolomei MS, Raj A. Visualizing allele-specific expression in single cells reveals epigenetic mosaicism in an *H19* loss-of-imprinting mutant. Genes Dev. 2016 Mar 1;30(5):567–78.
148. Tardaguila M, de la Fuente L, Marti C, Pereira C, Pardo-Palacios FJ, del Risco H, et al. SQANTI: extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. Genome Res. 2018 Mar;28(3):396–411.
149. Tang AD, Soulette CM, van Baren MJ, Hart K, Hrabeta-Robinson E, Wu CJ, et al. Full-length transcript characterization of SF3B1 mutation in chronic lymphocytic leukemia reveals downregulation of retained introns. Nat Commun. 2020 Dec;11(1):1438.
150. Tamborero D, Gonzalez-Perez A, Perez-Llamas C, Deu-Pons J, Kandoth C, Reimand J, et al. Comprehensive identification of mutational cancer driver genes across 12 tumor types. Sci Rep. 2013 Dec 20;3(1):2650.
151. Melloni GE, Ogier AG, de Pretis S, Mazzarella L, Pelizzola M, Pelicci P, et al. DOTS-Finder: a comprehensive tool for assessing driver genes in cancer genomes. Genome Med. 2014;6(6):44.
152. Gupta I, Collier PG, Haase B, Mahfouz A, Joglekar A, Floyd T, et al. Single-cell isoform RNA sequencing characterizes isoforms in thousands of cerebellar cells. Nat Biotechnol. 2018 Dec;36(12):1197–202.
153. Volden R, Vollmers C. Single-cell isoform analysis in human immune cells. Genome Biol. 2022 Dec;23(1):47.
154. Lebrigand K, Magnone V, Barbry P, Waldmann R. High throughput error corrected Nanopore single cell transcriptome sequencing. Nat Commun. 2020 Dec;11(1):4025.
155. Petti AA, Williams SR, Miller CA, Fiddes IT, Srivatsan SN, Chen DY, et al. A general approach for detecting expressed mutations in AML cells using single cell RNA-sequencing. Nat Commun. 2019 Dec;10(1):3660.

156. Nam AS, Kim KT, Chaligne R, Izzo F, Ang C, Taylor J, et al. Somatic mutations and cell identity linked by Genotyping of Transcriptomes. Nature. 2019 Jul 18;571(7765):355–60.
157. Rodriguez-Meira A, O'Sullivan J, Rahman H, Mead AJ. TARGET-Seq: A Protocol for High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. STAR Protoc. 2020 Dec;1(3):100125.
158. Thol F, Klesse S, Köhler L, Gabdoulline R, Kloos A, Liebich A, et al. Acute myeloid leukemia derived from lympho-myeloid clonal hematopoiesis. Leukemia. 2017;
159. Velten L, Story BA, Hernández-Malmierca P, Raffel S, Leonce DR, Milbank J, et al. Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. Nat Commun. 2021 Dec;12(1):1366.
160. Farge T, Saland E, de Toni F, Aroua N, Hosseini M, Perry R, et al. Chemotherapy-Resistant Human Acute Myeloid Leukemia Cells Are Not Enriched for Leukemic Stem Cells but Require Oxidative Metabolism. Cancer Discov. 2017 Jul 1;7(7):716–35.
161. Esposito MT, So CWE. DNA damage accumulation and repair defects in acute myeloid leukemia: implications for pathogenesis, disease progression, and chemotherapy resistance. Chromosoma. 2014 Dec;123(6):545–61.
162. Clark TA, Schweitzer AC, Chen TX, Staples MK, Lu G, Wang H, et al. Discovery of tissue-specific exons using comprehensive human exon microarrays. Genome Biol. 2007;8(4):R64.
163. Yi L, Pimentel H, Bray NL, Pachter L. Gene-level differential analysis at transcript-level resolution. Genome Biol. 2018 Dec;19(1):53.
164. Kim E, Ilagan JO, Liang Y, Daubner GM, Lee SCW, Ramakrishnan A, et al. SRSF2 Mutations Contribute to Myelodysplasia by Mutant-Specific Effects on Exon Recognition. Cancer Cell. 2015 May;27(5):617–30.
165. Kon A, Yamazaki S, Nannya Y, Kataoka K, Ota Y, Nakagawa MM, et al. Physiological Srsf2 P95H expression causes impaired hematopoietic stem cell functions and aberrant RNA splicing in mice. Blood. 2018 Feb 8;131(6):621–35.
166. Brunner AM, Steensma DP. Targeting Aberrant Splicing in Myelodysplastic Syndromes. Hematol Oncol Clin North Am. 2020 Apr;34(2):379–91.