nature portfolio

Peer Review File

Chromatin landscape and epigenetic heterogeneity of acute myeloid leukemia

Corresponding Author: Professor Seishi Ogawa

Any redactions in this file are there to maintain patient confidentiality, the confidentiality of unpublished data, or to remove third-party material.

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Referee #1

(Remarks to the Author)
In this submission, Ochi and colleagues present an integrated epigenetic landscape of AML at considerable scale, utilising multi-omics analysis, incorporating ATAC-seq data from 1,563 AML patients, alongside RNA-seq, ChIP-seq, functional drug screening and scRNA/ATAC-seq via Multiomics, on a sequentially decreasing subset of these patients. It represents a tour-de force, both in terms of its integration of multiple genomic techniques, as well as its considerable scale. The starting point is bulk ATAC-seq and a key aspect of the work is the identification of distinct AML subgroups, distinct from genetic and transcription-based classification, based on their ATAC profiles, leading to a new AML classification that sometimes differs from previous classifications (including the diagnostic and prognostic classifications of the WHO, ICC, and ELN). This is hardly surprising and is intellectually attractive, given that genetics presumably defines fixed parameters that are somewhat agnostic to the cell transformed, transcription a static “snap-shot” of cell state that takes into account genetics and cellular provenance, but chromatin accessibility allows a window into cellular history, genetic and epigenetic drivers, as well as the further transcriptional and phenotypic “potential” of the tumour upon perturbation.

This new classification correlates with gene expression profiles, and in fact a gene expression-predicted model for the ATAC subgroups can be constructed using the ClaNC algorithm. In addition, the ATAC-classification also correlates with unique super-enhancer (SEs) usage and ATAC-subgroups appear driven by distinct key transcription factor (TF) networks, as described by gene regulatory analysis of both bulk and scRNA-seq. To demonstrate its clinical utility, the classification also correlates with leukaemia-specific phenotypes (including blast percentage, white blood cell count, and differentiation profiles by FAB), as well as patient overall survival. The study also sheds significant light on the increasingly described inter-tumour heterogeneity described within AML, highlighting differences in ATAC profiles among patients with similar common driver genetic alterations: where differences can be seen not only in the ATAC pattern, but also in the deconvoluted proportion of cells that make up the tumour per ATAC subgroup, and in the recurrent co-mutations that occur with the common drivers within specific ATAC subgroups. This demonstrates the critical role of the epigenome in AML heterogeneity. Potentially it also infers a similar role for the epigenome in treatment resistance. Using an exemplar from the study, the epigenetic classification may also have significant clinical implications, particularly for treatment selection, as demonstrated by subgroup K, which exhibits an unexpected, but, from the ex vivo drug screening, obvious sensitivity to ABL inhibitors. Not to be underestimated for a disease with such a poor overall survival, this study also provides an invaluable and incredibly rich resource of human AML epigenomic data for the research community.

In general, the data are massively comprehensive and highly convincing, the manuscript is clearly and well written and I am very positive about it.

I have the following comments, queries and specific questions:

General.

The breadth and depth of the data is one of the compelling qualities of this study. However, it is incredibly data heavy and in parts, which I will detail, sometimes hard to follow. Anything that the authors can do to make the story more accessible for the

reader and to cement the narrative would be appreciated. For example, in my opinion, the comparisons to the methylation data in the BEATAML cohort (line 255) do not really add anything, other than another layer of the 'omic data. As there is already so much data, I would suggest "less is more" and would look to try and take out superfluous datasets that do not add to the story or stray away from the central narrative. In addition, perhaps a more simplified explanation of the GRN algorithms (ANANSE and SCENIC+) would make this more accessible for a more general reading audience.

Specific comments and questions

1. Stability of the clusters - The authors built an expression-predicted model for the ATAC subgroups, using the ClaNC algorithm. Most subgroups show good accuracy, regardless of the number of patients (i.e group I), except groups O and P, which have accuracies near 0.5 or lower, as shown in ED_Fig.4b. Do the authors have any insights into why these 2 groups are less transcriptionally tractable? Does this relate to heterogeneity within these groups? P is the largest subgroup and judging by its heterogeneous ATAC profile in Fig 1c "ATAC peaks", is it really a single subgroup or should it be split further? In the ATAC profile (Fig.1c), it appears that the samples at the beginning of the group (left side) exhibit slightly different profiles to those at the end (right side), particularly when looking at the red rows at the top and middle of the ATAC profile.

2. For super-enhancers (Figure 2c and text), it is quoted that the top 750 SE were identified. However, on the breakdown in the text and in Fig2 c the numbers given for common (498), partially shared (883) and unique (387), do not add up to 750 (instead 1768). Can this be clarified please.

3. In line 274 it is mentioned that "These subgroup specific SEs were shown to tightly correlate with accessible chromatin and Pol II binding with variable cohesin signals (ED Fig5a-c)". This would be a reasonable description of patterns at all SE - were these subgroup specific SEs more enriched for this pattern? If not then this statement should be removed, clarified or toned down.

4. Utilising the expression-predicted model to assess four external RNAseq adult AML cohorts (ED_Fig.4c), this did not fully recapitulate the classification obtained by the ATAC profile (Fig.1c vs Ext Data 4c). We observed several differences between the 2 classifications: 1) the ratio of patient in each group differs, with subgroup P and L being smaller using the expression-predicted model compared to the ATAC profile, and 2) genetic alterations: group M lacks the DDX41 alteration, which is not maintained on the classification from the expression-predicted model. As this mutation has been relatively recently described -is it simply that this lesion was not on the BeatAML panel?

5. The authors used the ANANSE algorithm to define TFs and gene regulatory networks (GRN) specific to each ATAC subgroup. Why was this methodology chosen over others? I was unable to find the methods used for this analysis (line 301, ED_Fig.6a). The figure legend mentions that the colours represent the regulatory strength of TFs on other TF, is this the meaning of "outdegree" (represented by red colours). This is currently unclear and could be better and more simply explained.

6. The authors performed scRNA/ATAC-seq analysis on 31 AML patients across 14 ATAC subgroups. However, why are ATAC clusters N and P not represented? Did they find it hard to find representative samples for these subgroups?

7. The multiomics needs to be explained better. After quite a few readings it was eventually apparent that the confusing point is the change in nomenclature, between the sc and ATAC clustering and the difference between the named clustering in Fig 4a and Ext Fig 7b (where the UMAP is the same but the nomenclature of the clustering is entirely different). This is confusing, ED_Fig.7b-c are difficult to follow due to changing of names (e.g., cluster A becomes AML6, cluster B becomes AML10). Perhaps labelling A as AML1 and B as AML2 and formally describing the link would make it easier to follow. Additionally, the choice of colours could be improved; using a gradient from grey to black for the normal sample (1 to 4) may help distinguish them more clearly and prevent confusion with group AML12, 13 and 14 for example.

8. The different ATAC patterns and different maturation blocks seen for the subdivisions of NPM1c (patterns D-F) mutations are intriguing. Can the authors speculate further; is this due to their different cellular composition and/or level of maturation block, to their differing patterns of co mutation (D -NPM1c and TET2/IDH1/2, E NPM1c and DNMT3A, (Cohesin and WT1) and F NPM1c and DNMT3A/TET2) to a combination of both or to other factors?

9. The block at the stage of CLP in the group K (RUNX1-mutated) is partiularly intriguing. It is tempting to speculate that these leukeamias may share vulnerabilities with ALL and/or that they have shared vulnerabilities with early normal B-cells? What do the authors speculate on this (please also see point 13 below).

10. I do not quite see the logic of the statement "These findings suggest that the activation of key TFs at critical stages along the dirfferenetiation trajectory may explain the subgroup-specific leukemogenic mechanisms" or find that it is justified by data. Can the authors please clarify.

11. There seems to be some differences between the cohorts discussed (Japanese/Swedish vs BeatAML etc). This is evident in the KM graphs in Fig 5a and Supp Fig 9d as well as the violin plots in Supp Fig 9a. This is not a major issue, as cohorts will differ and it may simply be explained by metrics such as the age and WCC of the cohort, there are some similarities but I do not agree that they are very similar (Age, WCC and Plt count, two of which are major poor prognostic indicators seem to be different). This is important, as it may in turn explain some of the differences outlined in point 1 and I feel should be mentioned.

12. For the improvement in concordance risk in Fig 5c, I understand the meaning of this measure and that it is higher when the ATAC patterns are taken into account, but as I am not a Biostatistician, so: is this effect size significant, or is significance even a measure in these comparisons? It would be good in some way to quantify this statement and to give it some practical relevance.

13. The drug sensitivity comparisons are also intriguing. To me they suggest that there are sensitivities that reflect “cell state” as well as epigenetic state and mutational pattern. The examples given are for sensitivity of MEK1/2 inhibitors in subgroups C, F and H that go beyond the presence of RAS mutations and of ABL inhibitors for the intriguing subgroup K (see above). I note that subgroups C, F and H are all enriched for monocytic cell states and differentiation and that sensitivity of AMLs with such differentiation to MAPK signalling has previously been described (PMID: 35439021). Moreover, subgroup K has shown a differentiation block at the CLP stage and previous experimental literature has demonstrated the critical nature of Abl signalling within the early B-cell compartment (PMID 2065352 and PMID 2065253). It is compelling to think that this vulnerability relates to the cell state as revealed by ATAC. In this regard the RUNX1 mutation may also contribute, as RUNX1 is known to be required for B cell development (PMID: 23704093 and PMID: 15784726).

14. The observation in Fig 5B that the ATAC risk can modulate the prognosis within the ELN subgroups is an important point, but currently it is poorly explained. Could it be explained more clearly please?

15. It would be helpful to the clinician if the insights described here were available within the clinic. In simple terms can the authors speculate on a line-of-sight as to how a more involved and relatively expensive test such as ATAC could be simplified to allow diagnostic usage and utility, perhaps via limited gene expression profiles?

Additional points

Although the data generated are comprehensive, I wonder if we can press the authors to look for more:

More generally, in the scATAC analysis, were you able to identify or speculate on the LSC population within the ATAC subgroups? Was this possible for example within the Muliomics, perhaps by trajectory analysis of scRNASeq? If so, is there a profile difference between the blast population and the LSC population?

I am intrigued by the possibility to look at heterogeneity within this lovely dataset. Inter-tumour heterogeneity (i.e within subgroups) is mentioned and is apparent even visually within the UMAPs (subgroup P for example), although not fully explored. Also have the authors looked at heterogeneity within clusters? In normal blood differentiation, ATAC patterns can give an indication of the cellular potential pathway for differentiation, etc. I suspect that there are similar latent pathways within AML cells that might be determined by ATAC and predictive of the response of cells to perturbations such as therapy. Can the authors expand upon inter and intra-tumour heterogeneity, link this to ATAC subgroup and possibly look for any correlations with differences in outcome, either within or between subgroups? For example, within the larger subgroups does degree of heterogeneity read out into differences in outcome for individual patients?

Minor comments

A. Figure 2a. It would be more helpful if some of the major genes/clusters were named and annotated. This would allow linkage to Figure 2b where programmes are pointed out.

B. Please check the references in the figure legends.

Referee #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #3

(Remarks to the Author)
This study reports a large cohort of AML (and somewhat peripheral) hematologic malignancies through the lens of ATAC-seq analysis, with correlation to RNA-seq, and for subsets of patients, gene mutational analysis, genome sequencing, superenhancer analysis and drug sensitivity testing. The study reports groups of patients defined by ATAC seq, and their correlation with genomic features, expression, TF motifs and drug responsiveness/outcome. Large datasets of AML with multiomic data to rigorously define subgroups and correlation with clinical features are valuable, and this is, in some respects, such a large cohort and the data will likely be extensively utilized. There are several major issues:

The core of the study is that the ATAC approach identifies cases/groups not evident from alternative approaches. From the data presented, my interpretation is that ATAC recapitulates drivers, and identified putative “phenocopies” that are not explored. The claim that only 4 groups are clearly correlated with genetics is an underestimate: at least 10 have striking correlation with genetic features.

There is little exposition in the text about how the groups were defined. Some are distinct, most are not. Algorithms may assign groups but for those where there is overlap (e.g. Fig 1b) how were the groups rigorously defined and benchmarked?

Fig 1C shows that the majority of groups have enrichment for specific drivers. Thus, ATAC appears to indicate known groups, rather than defining new. For cases in these groups lacking an annotated genetic driver, it is not clear which lack complete sequencing, and which have been pursued with comprehensive sequencing to identify an unsuspected driver. This could be very powerful.

Multiple prior studies have used a similar approach, but with methylation profiling, to define groups in AML. The correlation with external datasets is appreciated, but this notwithstanding, the authors should perform methylation profiling of at least a subset of their cases (e.g. EPIC or similar) to provide a reference.

Please provide mode details on how the cohorts were enrolled and treated - particularly important for the outcome analyses.

Were any novel mutations/alterations identified from the genomic analyses? (Lines 156-)

Line 175- (and figs) “unique lineage commitment” is not supported by these data, that show some variation between subtypes”

Several statements are too strong. E.g. line 97 “a definitive role of the epigenome” - rather, it could be a definitive role of the genome, and the epigenome is secondary.

Line 129: “Chromatin” is vague - what is intended here?

ED Fig 4E: (chromatin accessibility and methylation): this does not appear to be a formal correlative analysis as suggested in the text (line 258)

Referee #4

(Remarks to the Author)
In this study Ogawa, Lehmann and colleagues generated a very large ATAC-seq data set from 1563 AML patients drawing from two cohorts, complementing with exome sequencing, RNA sequencing and single-cell RNA/ATAC sequencing. It provides a multi-faceted view of AML. The authors then carried out a number of analyses, identified 16 AML subgroups that were investigated in detailed, characterized super-enhancer mediated gene regulatory networks and regulatory mechanisms, and deconvoluted bulk ATAC-seq data using cell-type specific chromatin profiles determined with scATAC-seq. They further sought to stratify patients and predict drug sensitivity by integrating the clinical features and drug sensitivity information. Overall, this is a quite significant study in terms of the patient number and representation, multiple data modalities, as well as the depth of analysis. I think it will become a landmark study in the field of AML genomics.

There is one area where the authors can greatly improve this manuscript. This study has a lot of data from different assays. The analyses were anchored on chromatin accessibility based on ATAC-seq, which represent the unique aspects of this study in comparison with several previous studies relying on mutations, CNAs, or transcriptome. This is great and I don’t have an issue with that. On the other hand, there have been multiple previous publications of AML focusing on the transcriptome. These studies also classified AML into different subgroups. Subgroups defined by RNA and those by chromatin can partially or completely overlap with each other, but with different names from different studies. This can confuse the researchers or clinicians in the community. Not connecting subgroups identified by RNA-seq and ATAC-seq is a missed opportunity. At the single-cell level, ones can also define cell types or subtypes based on scRNA-seq or scATAC-seq profiles. Based on the existing data from a variety of organs and tissues, in general one can separate single cells at a greater resolution into more subtypes with RNA than with ATAC. This could be due to either technical (spareness of the data) or biological reasons. There are also rare examples of one RNA cluster that can be split into more than one sub-clusters based on ATAC, indicating a greater level of epigenetic heterogeneity not captured by gene expression. With their data, the authors have a great opportunity to define and connect cell types using both RNA and ATAC. Along these lines, I offer a few specific suggestions for improvement:

(1) Cross-Referencing Transcriptome-Based Subgroups. Figure 1c connects a variety of different information to the 16 subgroups defined by ATAC-seq. It would be helpful to define AML subgroups based on transcriptome, either based on the previous studies, or de novo using the RNA-seq data generated from 1563 patients in this study. The RNA-based subgroups can be integrated into this plot, or as a separate supplementary figure. The ATAC subgroups A/B/C/D/I align quite well with specific WHO/ICC subtypes, but the others are more mixed. It would be really interesting to see how well these ATAC subgroups are supported by the transcriptome.

(2) Integrating Single-Cell Clusters Across Modalities. The scRNA-based UMAP in Fig. S7a showed more clusters than the scATAC UMAP in Fig. S7b and Fig. 4a. Aligning, integrating and annotating these cell clusters across the two modalities

can provide more insights on exactly how many cell types or subtypes are there in AML, how each layer of molecular information contribute to the definition, which cell types or subtypes have distinct functions and/or specific contributions to pathology/prognosis.

(3) Cell-Type Specific Gene Regulatory Network Analysis. The gene regulatory networks showed in Fig 3a were identified based on bulk ATAC-seq and RNA-seq data. These networks are likely quite difference once you zoom into individual cell types or subtypes in each group. With the single-cell multiome data from 31 patients, the authors have enough data to run a cell-type specific GRN analysis.

(4) Comparative Predictive Power of RNA- and ATAC-Based Subgroups. Similar to the analyses done in Fig. 5a/b, a parallel analysis using RNA-based subgroups would be really helpful. If the omics information is predictive of survival or informative for patient stratification, for the clinicians the ultimate questions are (i) what is the most predictive; (ii) what is the most practical assay for routine clinical applications. An unbiased comparison between ATAC-seq and RNA-seq subgroups can answer these critical questions. Even if RNA is more informative than ATAC on some aspects, that does not diminish the impacts of this study.

Version 1:

Reviewer comments:

Referee #1

(Remarks to the Author)
The authors are to be commended for their very thorough work in answering our questions and those of the other reviewers. They have answered every significant point that we have raised and particularly related to the heterogeneity of the subgroups have made the analysis more robust in its groupings. Although there is still a huge amount of data in the paper, and indeed this is one of its strong points, the authors have more strictly adhered to the central narrative and as a result, the paper is more easily understood than previously. They have also diligently addressed some of the biological inferences that we for the data and have included some in a more thoughtfully worded discussion section. In light of this significant revision, we find the paper much improved and now appropriate for publication in Nature.

Referee #2

(Remarks to the Author)
I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

Referee #3

(Remarks to the Author)
The authors have made several improvements to the manuscripts, including clarification of methods and incorporating additional methylation data. There is a persisting major concern that in part arises, potentially from the comprehensiveness of genomic profiling, and then the downstream implications.
The authors persist in saying that only 4 groups are associated with genomic alterations (raising complete correlation as being required), and conversely, that several groups have no correlation and are novel groups that transcend (e.g.) the ELN classification. As stated in my original review, and as acknowledged by the authors in response (and to a limited degree, in the revised MS), many of the groups show quite strong associations with genomic alterations, even if this is incomplete. This may reflect several factors, including the genomic complexity of AML (there may be multiple mutations in single or multiple genes in a case), and the adequacy of genomic profiling. From their response and the revised manuscript, it is not clear if there has been comprehensive sequencing (e.g. WGS + WTS) for cases in a "less completely" associated group, and the putative novel groups. This is really important. Multiple recent large scale genomic studies have identified unexpected drivers, including in the non-coding genome, when subjected to comprehensive sequencing. Here the authors state their analysis has been limited to known cancer drivers, and using a candidate (coding) sequencing approach, and what appears to be relatively sparse SNP sampling for copy number analysis. To this reviewer, this is not adequate to resolve this point. AML is a genomic disease, in which somatic mutations often affect regulators of the epigenome. It would thus be expected that any putative novel group may indeed have distinct drivers, or distinct patterns of alteration of drivers (indeed the authors do deploy this strategy in a limited way, for example for the NPM<1 + TET2/IDH group).

As a downstream consideration - how do the authors expect these data will be translated, when clinical ATAC sequencing is unlikely to be clinically implemented?

Referee #4

(Remarks to the Author)
In this revision the authors have provided a significant amount of new experimental (DNA methylation) data, added many analyses, and improved the writing. which together fully addressed my concerns. These efforts have lifted the quality of this manuscript above the standard for Nature, so I'm supportive of accepting it for publication.

One small request: in Supplementary Fig. 19, I think adding the AUC curves would be quite helpful, because the prediction accuracies can depend on the specific cutoffs of choice.

Version 2:

Reviewer comments:

Referee #3

(Remarks to the Author)
In this revised version of the manuscript, the authors have been quite responsive to this reviewer providing more clarity and additional data and analysis for (e.g.) RNAseq and methylation data, and integration thereof. The study will be a valuable resource. There are persisting concerns. The main is the claim that the study has identified new subtypes of AML that transcend existing classification schema (and through the outcome analysis, the suggestion that this classification should replace existing schema). The authors response is somewhat contradictory in parts both with itself and the manuscript in terms of correlation with genetic alterations. The report 4 total, although this is really 3 as inspection of the data indicate that the association with CEBP alterations is not as absolute as the fusions; most of the other ATAC groups have clear incomplete associations with mutations, or groups of mutations, as they describe in the MS. The main outcome associations favorable and unfavorable that they report with the ATAC groups correlate with strong prognostic genomic features (fusions for good outcome, MLLr, TP53 etc for poor). Yes, there are are cases in several of the groups that do not appear explained by the genomic data, but the request for detailed WGS (at a minimum) for cases in these groups was not fully responded to: specifically, it is not clear which cases were sequenced, and focused genotyping of known drivers does not address the issue (of the need for discovery).
Thus, while the dataset is large and valuable to the field, it is somewhat incremental on prior studies that have used gene expression, or methylation, to define groups of AML.
The authors have provided additional data and analysis for a variety of approaches: superenhancer analysis, TF networks, sc profiling and mapping to the Zeng/Dick map: but these are observational and in each case raise notions that are left completely unexplored to demonstrate a role if a factor in biology (or response).
The outcome analyses are interesting, as they raise the notion that there are subsets of cases that are not fully resolved with the WHO or ICC classifications, but it is a missed opportunity that the authors do not fully explore their own mutational data to determine which schema might be optimally prognostic - more nuanced analysis of the mutational data and combinations thereof, or ATAC?

Some observations are not novel or referenced (e.g. RUNX1 and lymphoid features).
The close correlation with methylation is notable, and raises the question as to which epigenetic classifier might be optimal for grouping/subtyping.

The mutational data could be explored in more detail. One notable example is the TP53 mutant groups where there are associations with inferred developmental stage, and also lineage of leukemia. There are known associations between TP53 status, VAF/zygosity, and leukemia phenotype (e.g. erythroid). These features: VAF, sequence of acquisition, and mutational constellations, would be valuable to mine and associate with ATAC group in more detail.

The expanded therapeutic data are welcome and a major feature of the study.

Version 3:

Reviewer comments:

Referee #3

(Remarks to the Author)
My key concern is not fully resolved. I appreciate that the WGS data was analyzed genome wide, but my question was to address the relationship of genomic driver and ATAC group to (1) determine the drivers in groups where there is not a complete 1:1 relationship of driver to ATAC group and (2) to determine the best outcome model (as the authors claim that this represents a new framework for classification. The authors report that the WGS group is a subset of the overall group, and direct to ST3, but most cases in this table appear to have a canonical driver. The request was systematic WGS of those cases in each ATAC group that do NOT have a canonical driver.

Version 4:

Reviewer comments:

Referee #3

(Remarks to the Author)
The authors show that the existing detected constellations of mutations imperfectly predict ATAC-defined groups, and analysis of the limited additional WGS data in cases lacking a known driver identifies driver in some, but not all cases. It is helpful information, but does not fully resolve the question as to whether full discovery of these cases, with systematic WGS/RNA-seq/ATAC could more fully resolve the drivers of AML.

As an additional note, the additional analyses do not appear to have been incorporated into the manuscript.

Point-by-point responses to the reviewers' comments.

Responses to Reviewer #1

>>> Reviewer #1

General comments:

In this submission, Ochi and colleagues present an integrated epigenetic landscape of AML at considerable scale, utilising multi-omics analysis, incorporating ATAC-seq data from 1,563 AML patients, alongside RNA-seq, ChIP-seq, functional drug screening and scRNA/ATAC-seq via Multiomics, on a sequentially decreasing subset of these patients. It represents a tour-de force, both in terms of its integration of multiple genomic techniques, as well as its considerable scale. The starting point is bulk ATAC-seq and a key aspect of the work is the identification of distinct AML subgroups, distinct from genetic and transcription-based classification, based on their ATAC profiles, leading to a new AML classification that sometimes differs from previous classifications (including the diagnostic and prognostic classifications of the WHO, ICC, and ELN). This is hardly surprising and is intellectually attractive, given that genetics presumably defines fixed parameters that are somewhat agnostic to the cell transformed, transcription a static "snap-shot" of cell state that takes into account genetics and cellular provenance, but chromatin accessibility allows a window into cellular history, genetic and epigenetic drivers, as well as the further transcriptional and phenotypic "potential" of the tumour upon perturbation.

This new classification correlates with gene expression profiles, and in fact a gene expression-predicted model for the ATAC subgroups can be constructed using the ClaNC algorithm. In addition, the ATAC-classification also correlates with unique super-enhancer (SEs) usage and ATAC-subgroups appear driven by distinct key transcription factor (TF) networks, as described by gene regulatory analysis of both bulk and scRNA-seq. To demonstrate its clinical utility, the classification also correlates with leukaemia-specific phenotypes (including blast percentage, white blood cell count, and differentiation profiles by FAB), as well as patient overall survival. The study also sheds significant light on the increasingly described inter-tumour heterogeneity described within AML, highlighting differences in ATAC profiles among patients with similar common driver genetic alterations: where differences can be seen not only in the ATAC pattern, but also in the deconvoluted proportion of cells that make up the tumour per ATAC subgroup, and in the recurrent co-mutations that occur with the common drivers within specific ATAC subgroups. This demonstrates the critical role of the epigenome in AML heterogeneity. Potentially it also infers a similar role for the epigenome in treatment resistance. Using an exemplar from the study, the epigenetic classification may also have significant clinical implications, particularly for treatment selection, as demonstrated by subgroup K, which exhibits an unexpected, but, from the ex vivo drug screening, obvious sensitivity to ABL inhibitors. Not to be underestimated for a disease with such a poor overall survival, this study also provides an invaluable and incredibly rich resource of human AML epigenomic data for the research community.

In general, the data are massively comprehensive and highly convincing, the manuscript is clearly and

well written and I am very positive about it.

I have the following comments, queries and specific questions:

(REPLY)

We sincerely thank the reviewer for the thoughtful and constructive comments, as well as the positive evaluation of our manuscript. In response, we have carefully revised the manuscript and conducted additional analyses to address all points raised by the reviewers. Below, we provide a detailed, point-by-point response outlining the changes made and the rationale behind them.

General

The breadth and depth of the data is one of the compelling qualities of this study. However, it is incredibly data heavy and in parts, which I will detail, sometimes hard to follow. Anything that the authors can do to make the story more accessible for the reader and to cement the narrative would be appreciated.

(REPLY)

Thank you for highlighting the breadth and depth of our study and for your constructive suggestion to improve its readability. We recognize that the large volume of data may at times have made the manuscript difficult to follow. In response to your detailed feedback, we have revised the text to provide a clearer and more cohesive narrative. In particular, we improved the descriptions of ANANSE, SCENIC+, and the refinement of ELN risk prediction, as outlined below.

For example, in my opinion, the comparisons to the methylation data in the BEATAML cohort (line 255) do not really add anything, other than another layer of the 'omic data. As there is already so much data, I would suggest "less is more" and would look to try and take out superfluous datasets that do not add to the story or stray away from the central narrative.

(REPLY)

Thank you for this comment. We agree with the reviewer. In response to the reviewer's suggestion, we removed the comparison with the BEAT AML methylation data to streamline the narrative and avoid adding superfluous layers of information.

Meanwhile, another reviewer (Reviewer#3) requested the methylation analysis using "our own" cohort and in response to this reviewer, we newly performed DNA methylation profiling using the Illumina EPIC array on 424 samples from our own cohort and analysed subgroup-specific DNA methylation. This was partly because we wanted to answer all our reviewers' requests as sincerely as possible and also because we believe that this newly generated dataset will provide a valuable resource for the broader research community. Through the analysis, we can directly correlate the methylation and ATAC data at the peak level, confirming the role of methylation in the regulation of chromatin accessibility and the epigenome.

In addition, perhaps a more simplified explanation of the GRN algorithms (ANANSE and SCENIC+) would make this more accessible for a more general reading audience.

(REPLY)

Thank you for this suggestion. In response, we have revised the main text, METHODs, and the legend of **Fig. 3a and ED Fig 6** to provide more accessible explanations of the GRN algorithms. While the algorithms themselves are inherently complex, we have simplified the descriptions in the main text as much as possible, and directed readers to the **METHODS** and figure legends for additional details. In particular, we emphasized that both algorithms integrate chromatin accessibility and gene expression to infer subgroup-specific TF–regulatory networks.

Lines 319–324 (for ANANSE):

"Because TFs are key regulators of gene expression and are frequently deregulated in AML[37], we analyzed TF-centered gene regulatory networks (GRNs) characteristic of each ATAC subgroup using ANANSE[38]. This software integrates chromatin accessibility and gene expression to construct TF networks, linking a TF to its potential targets when binding motifs are predicted at accessible regulatory regions, and reporting each TF's overall influence on gene expression (Influence score) (METHODS; ED Fig. 6a)."

Lines 392-395 (for SCENIC+):

"Finally, we investigated the activity of TFs across ATAC subgroups at single-cell resolution using multiome scRNA/ATAC-seq data and SCENIC+ software[46]. SCENIC+ infers TF networks by integrating gene expression and chromatin accessibility from the same single cells, in a manner analogous to ANANSE but applied to single-cell data (METHODS; ED Fig. 8a)."

Specific comments and questions

1. Stability of the clusters - The authors built an expression-predicted model for the ATAC subgroups, using the ClaNC algorithm. Most subgroups show good accuracy, regardless of the number of patients (i.e group I), except groups O and P, which have accuracies near 0.5 or lower, as shown in ED_Fig.4b.

1a. Do the authors have any insights into why these 2 groups are less transcriptionally tractable?

1b. Does this relate to heterogeneity within these groups?

1c. P is the largest subgroup and judging by its heterogeneous ATAC profile in Fig 1c "ATAC peaks", is it really a single subgroup or should it be split further? In the ATAC profile (Fig.1c), it appears that the samples at the beginning of the group (left side) exhibit slightly different profiles to those at the end (right side), particularly when looking at the red rows at the top and middle of the ATAC profile.

(REPLY)

1a. Thank you for this comment, which is also related to the comment from reviewer#3. We speculate that the reduced predictability for subgroups O and P stem from the inherent limitation in consistently detecting these two subgroups in ATAC-based clustering. To test this, we performed a resampling analysis, in which ATAC-based clustering was repeated 100 times using 90% of randomly selected samples to evaluate the stability of ATAC-based clustering. As shown in new **SFig. 4**, we obtained a high Adjusted Rand Index (ARI) (mean = 0.79) **(a)** and a consistency score (mean = 0.97) **(b)**, indicating high global stability of our ATAC clustering. However, when subgroup-level consistency scores were assessed,

O and P subgroups exhibited significantly lower values (mean: 0.66 and 0.59 for O and P, respectively) **(c)**, compared to other subgroups, where scores were nearly complete (subgroups A, B, C, D, and I) or very high (>0.9) **(c)**. This appears to be caused by reassignment of many samples from these subgroups to neighboring subgroups (e.g., P→N; O→N/K), likely reflecting ATAC profile similarity **(d)**.

1b and 1c. To assess intra-subgroup heterogeneity in ATAC-based clustering, we performed higher-resolution clustering by adjusting the clustering parameters to allow for increased cluster numbers. As the reviewer expected, subgroups O and P were split into two subgroups (O-1 and O-2 and P-1 and P-2) along with subgroups D and L (to D-1/D-2 and L-1/L2) **(SFigs. 5)**. This did not affect the overall stability **(SFig 6a-b)** and the improved clustering stability was observed for some subgroups J, K, L-2, and O-2. However, this improvement was offset by reduced stability in others (E, L-1, M, and O-1) with no improvement for subgroup P (P-1/P-2) **(SFig. 6c–d)**. Moreover, reproducibility of RNA-based prediction was substantially decreased **(SFig. 6e-f)**. These results indicate that intragroup-heterogeneity does exist and may partially explain the poor predictability of some subgroups (O and P) in RNA-based prediction model, but the refinement of clustering resolution does not result in the overall improvement of the stability with rather reduced predictability of RNA-based prediction of ATAC subgroups.

We added the results of the benchmark studies of our ATAC-based clustering, in which we have briefly mentioned the lower stability of subgroups O and P in the revised manuscript as follows:

Lines 194-199:

"The reproducibility of our clustering was benchmarked by high Adjusted Rand Index (ARI) (mean=0.79) and co-clustering accuracy (mean=0.97) in resampling analysis (SFig. 4a-c), although the consistency was significantly lower in subgroups O and P (SFig. 4d). A higher-resolution clustering revealed finer structures but resulted in a reduced stability (SFig. 5–6). Therefore, we retained the original 16-subgroup framework."

Lines 271-278:

"Given distinct gene expression profiles across subgroups, we built an expression-based prediction model for ATAC subgroups using the Classification to Nearest Centroids (ClaNC) algorithm[32] (ED_Fig. 4a), which achieved high accuracy, except for subgroups O and P showing lower cluster stability (SFig. 6e)."

2. For super-enhancers (Figure 2c and text), it is quoted that the top 750 SE were identified. However, on the breakdown in the text and in Fig2 c the numbers given for common (498), partially shared (883) and unique (387), do not add up to 750 (instead 1768). Can this be clarified please.

(REPLY)

We apologize for this confusion. To clarify, we selected the top 750 super-enhancers (SEs) within each ATAC subgroup, not a single set of 750 across all subgroups. We then merged SEs across all subgroups and removed duplicates to obtain a non-redundant union of 1,718 SE loci. These were classified as common (n = 498; present in ≥12 subgroups), partially shared (n = 833; present in 2–11 subgroups), and unique (n = 387; present in a single subgroup), which sum to 1,718. We have revised the text to make this explicit and corrected the figure caption accordingly.

Lines 288–294:

"For this purpose, we first identified all enhancers recurrently found in 234 AML samples using H3K27ac ChIP-seq, ranked them by average H3K27ac signals within each subgroup, and identified the top 750 as SEs within each subgroup. After merging the SEs identified across all subgroups, a total of 1,718 non-redundant SE loci were obtained, which were classified as common (n=498, found in ≥12 subgroups), partially shared (n=833, found in 2–11 subgroups), or unique (n=387, found in a single subgroup) (Fig. 2c, and STable 6)."

3. In line 274 it is mentioned that "These subgroup specific SEs were shown to tightly correlate with accessible chromatin and Pol II binding with variable cohesin signals (ED Fig5a-c)". This would be a reasonable description of patterns at all SE -were these subgroup specific SEs more enriched for this pattern? If not then this statement should be removed, clarified or toned down.

(REPLY)

No, these subgroup-specific SEs were not necessarily more associated with accessible chromatin and Pol II binding; instead, the association with these chromatin features is a general feature of SEs but not specific to subgroup-specific SEs. We are sorry for this confusion. According to the reviewer's suggestion, we removed this statement.

4. Utilising the expression-predicted model to assess four external RNAseq adult AML cohorts (ED_Fig.4c), this did not fully recapitulate the classification obtained by the ATAC profile (Fig.1c vs Ext Data 4c). We observed several differences between the 2 classifications: 1) the ratio of patient in each group differs, with subgroup P and L being smaller using the expression-predicted model compared to the ATAC profile, and 2) genetic alterations: group M lacks the DDX41 alteration, which is not maintained on the classification from the expression-predicted model. As this mutation has been relatively recently described -is it simply that this lesion was not on the BeatAML panel?

(REPLY)

Thank you for this comment. The reviewer's concerns provided an opportunity to test the performance of our prediction model.

1. The different representation of subgroups P and L in the external cohort could be explained by a) the real difference between the original and external cohort and/or b) the intrinsic poor performance of our model to predict some subgroups. To test this we first compared the subgroup composition within the original cohorts, i.e., between Swedish and Japanese cohorts. As shown in **Reviewers' Fig. R1a**, we saw substantial differences in some groups, suggesting the subgroup representation may differ significantly depending on the cohort tested. Second, as shown in **Reviewers' Fig. R1b** (see also below), applying our prediction model to the original cohort revealed poorer prediction of subgroups P, M, H, and to a lesser extent E and L. Thus, we speculate that the lower representation of P and L in the external cohort could be explained by the intrinsic nature of our prediction model, and may also reflect the real difference of the representation.

2. As the reviewer suspected, *DDX41* mutations were not systematically assessed or reported in the external datasets, including TCGA and the BeatAML cohorts (n=566/1,079), likely due to the relatively recent recognition of germline *DDX41* mutations (Polprasert et al., Cancer Cell 2015, PMID: 25920683). Only 9 cases in Subgroup M were informative for the *DDX41* mutation status, which prevented unbiased evaluation of the enrichment of *DDX41* mutations in this subgroup.

Nevertheless, overall, RNA-based prediction reproduced the subgroup features observed in the original ATAC classification, supporting the robustness of our ATAC-based classification. In line with the reviewer's comment #11, we also revised the main text as follows:

Lines 282-285:

"The predictions captured the overall patterns of gene mutation enrichment and clinical features, such as age, WBC, and blast percentage, characteristic of each subgroup, despite some deviations, thereby supporting the robustness and validity of our ATAC-based classification (ED_Fig. 4b, SFig 10-11)."

5. The authors used the ANANSE algorithm to define TFs and gene regulatory networks (GRN) specific to each ATAC subgroup.

5-a. Why was this methodology chosen over others?

(REPLY)

We chose the ANANSE because, to our knowledge, this is the only model which provides the capability of inferring enhancer-based gene regulatory networks by integrating both chromatin accessibility (from bulk ATAC-seq) and gene expression (from bulk RNA-seq) to infer GRNs. Unlike other methods, which rely solely on transcriptomic data, ANANSE enables the construction of GRNs rooted in chromatin-level regulatory potential, making full use of our large-scale ATAC-seq dataset. This enhancer-centric framework is particularly suitable for studying subgroup-specific transcriptional regulation in AML, where epigenetic dysregulation plays a central role in disease heterogeneity. Moreover, since we later conducted super-enhancer analysis and integrated these results with TF regulatory activity, using an enhancer-aware GRN approach allowed for a seamless and biologically coherent integration. We added the following explanations of the reason why we used ANANSE:

Lines 319-324:

"Because TFs are key regulators of gene expression and are frequently deregulated in AML[37], we analyzed TF-centered gene regulatory networks (GRNs) characteristic of each ATAC subgroup using ANANSE[38]. This software integrates chromatin accessibility and gene expression to construct TF networks, linking a TF to its potential targets when binding motifs are predicted at accessible regulatory regions, and reporting each TF's overall influence on gene expression (Influence score) (METHODS; ED Fig. 6a)."

5-b. I was unable to find the methods used for this analysis (line 301, ED_Fig.6a). The figure legend mentions that the colours represent the regulatory strength of TFs on other TF, is this the meaning of

"outdegree" (represented by red colours)? This is currently unclear and could be better and more simply explained.

(REPLY)

Thank you for these comments. The methods for this analysis were described in lines 710–720 of the original manuscript (now lines 883–895 in the revision). As the reviewer expected, the color intensity in **ED_Fig. 6b** represents the "outdegree" of each transcription factor (TF), i.e., the cumulative regulatory strength from that TF to other TFs within the predicted gene regulatory network inferred by ANANSE. In response to the reviewer's criticism, we fully revised Methods and the legends, as follows:

Lines 883-895 (METHODS):

"**GRN analysis by bulk ATAC- and RNA-seq.** GRNs were inferred using the ANANSE software[38]. Briefly, this software integrates chromatin accessibility and gene expression data to infer enhancer-based GRNs. Input data included the consensus ATAC peak set, a merged ATAC-seq BAM file, mean RNA expression (log_2 CPM), and the default motif database (GimmeMotifs[76]).

For each TF, ANANSE estimates genome-wide binding potential, predicts target genes, and incorporates expression levels of the TF and its targets to model regulatory interactions. The regulatory importance of each TF is quantified by two key metrics: outdegree (the number and strength of predicted regulatory connections) and linkscore (the strength of each individual connection). GRNs were constructed for each ATAC subgroup and for the entire AML cohort. To identify subgroup-specific regulators, we compared each subgroup GRN to the global AML GRN and calculated TF influence scores, which measure how much a TF explains expression differences between the two groups (ED Fig. 6a). The top 20 differential TFs per subgroup were visualized."

(Figure legend for ED_Fig. 6b):

"Subgroup-specific TF regulatory networks inferred by ANANSE. Color intensity indicates the outdegree, i.e., the total regulatory influence of each TF on other TFs. Edge width represents the linkscore, indicating the regulatory strength between TF pairs."

6. The authors performed scRNA/ATAC-seq analysis on 31 AML patients across 14 ATAC subgroups. However, why are ATAC clusters N and P not represented? Did they find it hard to find representative samples for these subgroups?

(REPLY)

Thank you for this suggestion. In the original manuscript, we could not include single-cell analysis for clusters N and P, due to the time-related constraint rather than sample availability. In this revision, we performed additional scRNA/ATAC-seq data for these samples, bringing the total to 36 AML (and 4 remission) samples. This expanded dataset now includes representative cases from all 16 ATAC subgroups, including N and P. Accordingly, we have fully revised all relevant texts and figures, including the main text, Methods, and Figures, to reflect this comprehensive update (**Fig. 4, ED_Fig. 7-8**). Specifically, we newly reveal that *TP53*-related subgroups (N, O, and P) showed distinct differentiation patterns in the trajectory analysis (**ED_Fig. 7e**). We have modified the manuscript accordingly:

Lines 348-350:

"We analyzed 281,167 mononuclear cells from 36 AML samples across all 16 ATAC subgroups along with four complete remission samples as normal controls (Fig. 4a, ED_Fig. 7a, SFig. 13-14, STable 7)"

Lines 378-380:

"Among the *TP53*-related subgroups, subgroup O was arrested at the HSC stage, whereas subgroups N and P exhibited differentiation toward the erythroid and monocytic lineages, respectively."

7. The multiomics needs to be explained better. After quite a few readings it was eventually apparent that the confusing point is the change in nomenclature, between the sc and ATAC clustering and the difference between the named clustering in Fig 4a and Ext Fig 7b (where the UMAP is the same but the nomenclature of the clustering is entirely different). This is confusing, ED_Fig.7b-c are difficult to follow due to changing of names (e.g., cluster A becomes AML6, cluster B becomes AML10).

7-a. Perhaps labelling A as AML1 and B as AML2 and formally describing the link would make it easier to follow.

7-b. Additionally, the choice of colours could be improved; using a gradient from grey to black for the normal sample (1 to 4) may help distinguish them more clearly and prevent confusion with group AML12, 13 and 14 for example.

(REPLY)

We appreciate the reviewer's comments and suggestions. Indeed, the UMAPs in **Fig. 4a** and **ED_Fig. 7b** are identical. The confusion appeared to arise from the different focus on cell labeling and nomenclature of clusters between both figures. To clarify, in **Fig. 4a**, cells were color-coded by sample, reflecting their bulk ATAC subgroup. In contrast, **ED_Fig. 7b** shows scATAC-based clusters (AML01, AML02, etc.), where cells from the same sample may be distributed across multiple clusters (see also ED Fig. 7c). This is more or less happening in virtually all samples as shown in **ED_Fig. 7c**. For this reason, the nomenclature of scATAC-based clusters in **ED_Fig. 7b** cannot be matched to that of bulk ATAC-based clusters and the clusters were annotated formally as AML01, AML02, AML03, etc. By contrast, in **Fig. 4a**, bulk ATAC subgroup is reflected in the sample name, like A (*PML::RARA*)-1, A (*PML::RARA*)-2, …, K (*RUNX1*)-1, K (*RUNX1*)-2. Thus, in the revised manuscript, we retain this distinct nomenclature in **Fig. 4a** and **ED_Fig. 7b**, while we modified colors used in both figures so that the colors indicate the bulk ATAC subgroups of the samples (in **Fig. 4a**) or reflect the major cell population (in **ED_Fig. 7b**). We also adjusted the colors for the normal samples (1 to 4) (**Fig. 4a**) and for the normal cell lineages (**ED_Fig. 7b**), according to the reviewer's suggestion.

We believe that these changes improve clarity while preserving the conceptual distinction between bulk-derived subgrouping and single-cell clustering.

8. The different ATAC patterns and different maturation blocks seen for the subdivisions of NPM1c (patterns D-F) mutations are intriguing. Can the authors speculate further; is this due to their different cellular composition and/or level of maturation block, to their differing patterns of co mutation (D -

NPM1c and TET2/IDH1/2, E NPM1c and DNMT3A, (Cohesin and WT1) and F NPM1c and DNMT3A/TET2) to a combination of both or to other factors?

(REPLY)

Thank you for these insightful comments. In scATAC-Seq analysis, most samples formed a distinct cell cluster, where leukemic cells from the same ATAC subgroup were clustered closely (**FIg. 4a**). This indicates that all leukemic cells in a given cluster share a consistent ATAC profile despite heterogeneity in their differentiation states (**Fig. 4c-d**). Such a profile is shared within the same ATAC subgroup but distinct from those from different subgroups, and based on those profiles, ATAC subgroups (A-P) are separated from each other. In this point of view, *NPM1*-mutated samples were thought to be further subdivided into three distinct subgroups (D-F) due to the difference in their unique ATAC (therefore, epigenomic) profile shared by all cells, including leukemia stem cells, within each subgroup. We speculate that such unique ATAC (epigenetic) profiles were determined by their cell-of-origin, patterns of co-mutations and copy number alterations, and other primary epigenetic factors, such as stem cell aging (PMID:8782459), chronic inflammation (PMID: 38917807), infections (PMID:27102489), and metabolic insults (PMID:34272515), dictating differentiation potential/maturation block, and distinct cellular components of each AML subgroup. These considerations apply not only to *NPM1*-mutated AML but also more broadly to other ATAC subgroups, and we have added this discussion in the revised manuscript as follows:

Lines 475-482:

"The validity of these distinct ATAC subgroups was further supported by scATAC-seq clustering, in which samples from the same subgroup showed a strong tendency to cluster closely together (Fig. 4a). This suggests that all leukemic cells within a subgroup, including LSCs, share a consistent ATAC profile that is distinct from the profiles of other subgroups. We speculate that such a profile, likely shaped by gene mutations in combination with non-genetic mechanisms such as stem-cell aging[13], cell-of-origin[12], and inflammatory[14] or metabolic insults[16], may dictate subgroup-specific differentiation trajectories, maturation blocks, and other leukemia phenotypes in ways that cannot be explained solely by genetics."

9. The block at the stage of CLP in the group K (RUNX1-mutated) is partiularly intriguing. It is tempting to speculate that these leukeamias may share vulnerabilities with ALL and/or that they have shared vulnerabilities with early normal B-cells? What do the authors speculate on this (please also see point 13 below).

(REPLY)

Thank you for the insightful comments and questions. We agree with the reviewer's speculation. In scRNA-seq analysis, subgroup K (*RUNX1*-mutated) exhibits a maturation block at the early B-cell stage or CLP-like cells (Fig. 4c), likely due to loss of RUNX1 function, a key transcription factor for early B-cell development (PMIDs 14966519, 23704093). The pronounced sensitivity of this AML subgroup to ABL inhibitors, even in the absence of canonical *ABL* fusions or mutations, is consistent with previous reports that ABL signaling plays a key role in early B-cell development (PMIDs 2065352, 2065353, 23704093,

15784726). We therefore speculate that the sensitivity of subgroup K AML samples to ABL inhibitors may reflect their phenotypic (epigenetic) similarity to normal early B-cells showing an inherent dependence on ABL signaling. We briefly mentioned this speculation in the revised manuscript, citing the relevant literatures:

Lines 453-460:

"In scRNA-seq analysis, subgroup K (*RUNX1*-mutated) exhibited a maturation block at the early B-cell stage or CLP-like cells (Fig. 4c), most likely due to loss of RUNX1 function, a key transcription factor for early B-cell development[51]. The pronounced sensitivity of this AML subgroup to ABL inhibitors, even in the absence of canonical *ABL* fusions or mutations, is consistent with previous reports that ABL signaling plays a key role in early B-cell development[52,53]. We therefore speculate that the sensitivity of subgroup K samples to ABL inhibitors may reflect their phenotypic (epigenetic) similarity to normal early B-cells showing an inherent dependence on ABL signaling."

10. I do not quite see the logic of the statement "These findings suggest that the activation of key TFs at critical stages along the dirfferenetiation trajectory may explain the subgroup-specific leukemogenic mechanisms" or find that it is justified by data. Can the authors please clarify.

(REPLY)

In this section, by combining SCENIC+ and pseudotime analyses, we demonstrated that key TFs are often shared between several ATAC subgroups but show distinct temporal expression patterns across those subgroups (**Fig. 4e**). For example, HOXA9 was a key TF consistently activated throughout the myeloid differentiation in HOX-related subgroups (D-G) but peaked at different stages according to subgroups (**Fig. 4e**). By contrast, although expressed throughout the myeloid differentiation in subgroup K, *BCL11A* and *IRF8* expression were differentially peaked in progenitor and mature stages, respectively. Another notable example was SPI1 (or PU.1), which showed a distinct expression peak in the stem-cell stage in subgroup G but mature cell stage in F and J. Based on these observations, we speculate that not only key TFs but also their temporal expression profile along the differentiation trajectory might be important in subgroup-specific leukemogenic mechanisms. However, we agree that these observations may not be sufficient to support the statement. We therefore attenuate the original statement.

The revised sentence reads:

Lines 405-408:

"These findings suggest that not only subgroup-specific key TFs but also their temporal activity profiles along the myeloid differentiation might be important to understand the subgroup-specific leukemogenic mechanism."

11. There seems to be some differences between the cohorts discussed (Japanese/Swedish vs BeatAML etc). This is evident in the KM graphs in Fig 5a and Supp Fig 9d as well as the violin plots in Supp Fig 9a. This is not a major issue, as cohorts will differ and it may simply be explained by metrics such as the age and WCC of the cohort, there are some similarities but I do not agree that they are very similar (Age,

WCC and Plt count, two of which are major poor prognostic indicators seem to be different). This is important, as it may in turn explain some of the differences outlined in point 1 and I feel should be mentioned.

(REPLY)

In response to the reviewer's comment, we quantitatively evaluated the correlations of major clinical parameters between the present and BeatAML cohorts for each ATAC subgroup. As shown in **SFig. 10b**, parameters of ATAC subgroups in the two cohorts showed overall good correlations, supporting the reproducibility of the ATAC subgroups across cohorts. Together with the similar co-mutation profile in the inferred ATAC subgroups, this supports the robustness of our ATAC classification. Nevertheless, we do agree that substantial deviations were still observed for some parameters in certain subgroups and correlations were weaker for hemoglobin values and platelet counts. We have revised the manuscript accordingly: the comparison of clinical parameters now appears immediately after the RNA-based inference of ATAC subgroups in the BeatAML cohort. We also briefly mention these deviations in the main text for clarity.

Lines 282-285:

"The predictions captured the overall patterns of gene mutation enrichment and clinical features, such as age, WBC, and blast percentage, characteristic of each subgroup, despite some deviations, thereby supporting the robustness and validity of our ATAC-based classification (ED_Fig. 4b, SFig 10-11)."

12. For the improvement in concordance risk in Fig 5c, I understand the meaning of this measure and that it is higher when the ATAC patterns are taken into account, but as I am not a Biostatistician, so: is this effect size significant, or is significance even a measure in these comparisons? It would be good in some way to quantify this statement and to give it some practical relevance.

(REPLY)

Thank you for raising this issue. Yes, this effect size is statistically significant and clinically meaningful. In our study, we assessed the statistical significance of improved c-index (Δc-index) by bootstrap resampling (1,000 iterations), which is a well-established method for evaluating the uncertainty of performance metrics in survival models (PMID: 35941135). However, rather than relying on P-values, which tend to converge to zero with increasing resampling iterations when the true Δc-index is positive, we should have reported the estimated Δc-index alongside its 95% confidence interval (95%CI), which more appropriately quantifies model improvement. In the Swedish cohort, the inclusion of ATAC-defined subgroups in the ELN-based model yielded a Δc-index of 0.039 (95% CI: 0.018–0.060). Importantly, the increased c-index was confirmed in the Japanese cohort, in which Δc-index of similar size 0.036 (95%CI: 0.0097-0.060) was observed, confirming the improved model by adding ATAC risk subgroups in the ELN-based prognostic model. We believe that this size of Δc-index (>0.03) has been accepted as clinically significant in many clinically relevant studies (PMID: 35941135). We are sorry for this confusing use of P value, and have revised the manuscript, just reporting Δc-index and its 95% confidence interval.

Lines 429-432:

"Moreover, the improvement of the original ELN model was demonstrated in the independent Swedish and Japanese cohorts, in which significant increases in the concordance index (Δc-index) of clinical relevance[47] were observed (0.039 and 0.036), respectively (Fig. 5c)."

13. The drug sensitivity comparisons are also intriguing. To me they suggest that there are sensitivities that reflect "cell state" as well as epigenetic state and mutational pattern. The examples given are for sensitivity of MEK1/2 inhibitors in subgroups C, F and H that go beyond the presence of RAS mutations and of ABL inhibitors for the intriguing subgroup K (see above). I note that subgroups C, F and H are all enriched for monocytic cell states and differentiation and that sensitivity of AMLs with such differentiation to MAPK signalling has previously been described (PMID: 35439021). Moreover, subgroup K has shown a differentiation block at the CLP stage and previous experimental literature has demonstrated the critical nature of Abl signalling within the early B-cell compartment (PMID 2065352 and PMID 2065253. It is compelling to think that this vulnerability relates to the cell state as revealed by ATAC. In this regard the RUNX1 mutation may also contribute, as RUNX1 is known to be required for B cell development (PMID: 23704093 and PMID: 15784726).

(REPLY)

According to the reviewer's suggestions, we fully rewrote this paragraph as follows. Thank you for these insightful comments.

Lines 442-463:

"By contrast, in other cases, the sensitivities did not seem to be determined solely by gene mutations but also by epigenetic state captured by ATAC-based classification. Sensitivity to MEK1/2 inhibitors, such as selumetinib and trametinib, was observed in subgroups C, F, and H, which was expected from the enrichment of RAS pathway mutations in these subtypes. However, the sensitivity was not confined to samples with RAS pathway mutations but also observed in those lacking RAS pathway mutations (Fig. 5d-e, ED_Fig. 10c). Given that all these ATAC subgroups are characterized by monocytic differentiation, this suggests that some epigenetic states common to these subgroups that dictate their monocytic differentiation underlies the sensitivity to MEK1/2 inhibitors, which is in agreement with a previous report showing the sensitivity of AMLs with monocytic differentiation to MAPK signaling[50]. Another notable example was an unexpected sensitivity of subgroup K to multiple ABL inhibitors (Fig. 5f, ED_Fig. 10d). In scRNA-seq analysis, subgroup K (*RUNX1*-mutated) exhibited a maturation block at the early B-cell stage or CLP-like cells (Fig. 4c), most likely due to loss of RUNX1 function, a key transcription factor for early B-cell development[51]. The pronounced sensitivity of this AML subgroup to ABL inhibitors, even in the absence of canonical *ABL* fusions or mutations, is consistent with previous reports that ABL signaling plays a key role in early B-cell development[52,53]. We therefore speculate that the sensitivity of subgroup K samples to ABL inhibitors may reflect their phenotypic (epigenetic) similarity to normal early B-cells showing an inherent dependence on ABL signaling. Nevertheless, it should also be noted that *RUNX1*-mutated samples in other ATAC subgroups did not exhibit similar sensitivities (ED_Fig. 10e),

highlighting the subgroup-dependent role of *RUNX1* mutations, most likely explained by the subgroup-specific epigenomic profile."

14. The observation in Fig 5B that the ATAC risk can modulate the prognosis within the ELN subgroups is an important point, but currently it is poorly explained. Could it be explained more clearly please?

(REPLY)

Thank you for pointing this out, and we are sorry for the insufficient explanation. To clarify, within each ELN risk category we first used multivariable Cox regression analyses incorporating both ATAC subgroups and known clinical covariates, such as age, sex, blood counts, and blast count, to identify the "high-risk ATAC subgroups" significantly and independently associated with OS (ED Fig. 9f). We then stratified patients in each ELN category into two groups according to the presence or absence of these risk ATAC subgroups and compared OS in KM analysis. Significance was evaluated using the Logrank test. As shown in Fig. 5b, this yielded two prognostically distinct groups within each ELN category.

We revised the main text and Method accordingly:

Main (Lines 422-429):

"Importantly, the ATAC-based classification was shown to improve the prognostication based on the ELN risk stratification. To demonstrate this, we first performed multivariable Cox regression analyses within each ELN risk category (favorable, intermediate, adverse) to identify the ATAC subgroups significantly associated with OS within each ELN category, by incorporating ATAC subgroups together with known clinical parameters as covariates (ED Fig. 9e-f). We then evaluated their effects on OS in each ELN category by stratifying each ELN category by the presence or absence of the risk ATAC groups. As shown in Fig. 5b, patients in each ELN category were further separated into two subcategories with different OS."

Methods (Lines 1043-1055):

"To assess whether ATAC subgroups could further stratify prognosis within established genomic risk categories, we performed multivariable Cox proportional hazards regression analyses separately within each ELN 2022 risk group (favorable, intermediate, and adverse). These models included clinical covariates (age, sex, WBC, hemoglobin level, platelet count, blast percentage) and ATAC subgroup membership as independent variables. Subgroups that were significantly associated with worse OS within each ELN stratum were designated as high-risk ATAC subgroups. We then stratified patients in each ELN category into two groups according to the presence or absence of these risk ATAC subgroups and compared OS in KM analysis. To evaluate the additive prognostic value of ATAC subgroups, we built a Cox model for ELN 2022 risk classification and compared it to a combined model of ELN and ATAC risk groups (Fig. 5b). Model performance was assessed by calculating the c-index using 1,000 bootstrap replicates in both the training (Swedish) and validation (Japanese) cohorts. The difference in c-index between the two models was used to evaluate the significance and robustness of the improvement."

15. It would be helpful to the clinician if the insights described here were available within the clinic. In

simple terms can the authors speculate on a line-of-sight as to how a more involved and relatively expensive test such as ATAC could be simplified to allow diagnostic usage and utility, perhaps via limited gene expression profiles?

(REPLY)

Thank you for the important comment. We suggest two potentially practical applications of our findings on ATAC subgroups. First, we newly demonstrated that subgroup K can be reliably predicted by measuring expression of 30 genes with positive and negative predictive values of 0.62 and 0.99, respectively (**SFig. 19b**). Similarly, high-risk ATAC subgroups for each ELN category can be predicted by seeing the expression of 30 genes (**SFig. 19a**). These findings suggest that we could potentially translate the current findings into bed-side clinics through limited gene expression profiling.

We fully rewrote the last paragraph of Discussion as follows:

Lines 493-501:

"Finally, our ATAC-based epigenetic classification also has important clinical implications. It has prognostic value independent of the conventional ELN classification and predicts new drug sensitivities not previously recognized. Incorporating risk ATAC subgroups significantly improved ELN-based prognostication, and MEK and ABL inhibitors may have a role in the treatment of particular ATAC subtypes. These findings provide a framework for personalized therapeutic strategies, once the ATAC-seq-based classification is implemented through more feasible diagnostic platforms. In this regard, we have developed reliable models predicting the risk ATAC subgroups based on the expression of only 30 genes, which could contribute to improved prognostication and drug selection (SFig. 19)."

Additional points

Although the data generated are comprehensive, I wonder if we can press the authors to look for more: More generally, in the scATAC analysis, were you able to identify or speculate on the LSC population within the ATAC subgroups? Was this possible for example within the Muliomics, perhaps by trajectory analysis of scRNASeq? If so, is there a profile difference between the blast population and the LSC population?

(REPLY)

To investigate the LSC population at the single-cell level, we evaluated LSC scores using previously reported LSC-related gene signatures (Ng et al., Nature 2016, PMID: 27926740; Zeng et al., Blood Cancer Discov 2025, PMID: 40294241) in our scRNA/ATAC-seq dataset. We then assessed the relationship between LSC scores, and differentiation states inferred by pseudotime analysis and ATAC subgroups. We found a strong inverse correlation between LSC scores and pseudotime across all 16 ATAC subgroups. Regardless of ATAC subgroups, cells with higher LSC scores were consistently enriched in the early stages of the differentiation trajectory, supporting that LSCs are characterized by HSC-like gene expression (**SFig. 17a**). Given the distinct lineage commitment along the differentiation trajectory across different ATAC subgroups, this may suggest a large variation in LSCs numbers (**SFig. 17b**), which however, needs further validations in functional assays.

We included these analyses in the revised manuscript:

Lines 384-391:

"Leukemic stem cells (LSCs), defined by stem cell–associated gene expression programs and thought to sustain disease propagation and relapse in AML, were computationally assessed at the single-cell level using established LSC signatures (LSC score)[44,45]. High LSC scores consistently mapped to early stages of the differentiation trajectory across all ATAC subgroups, suggesting that LSC-like cells are present within each epigenetic context (SFig 17a). Given the distinct lineage commitment along the differentiation trajectory across different ATAC subgroups, this may suggest a large variation in LSC numbers (SFig. 17b), which however, needs further validations in functional assays."

I am intrigued by the possibility to look at heterogeneity within this lovely dataset. Inter-tumour heterogeneity (i.e within subgroups) is mentioned and is apparent even visually within the UMAPs (subgroup P for example), although not fully explored. Also have the authors looked at heterogeneity within clusters?

(REPLY)

As detailed in our responses to point 1, we explored intra-subgroup heterogeneity using higher-resolution re-clustering analyses. This revealed substructure within some subgroups, including D, L, O, and P, each of which was further divided into two subgroups, D-1/D-2, L-1/L-2, O-1/O-2, and P1/P2, with distinct mutation enrichments (e.g., *TP53* mutations enriched in P-1) (**SFigs. 5-6**). However, this caused significant reduction of clustering stability and compromised reproducibility in many subgroups (see also Response to point 1). Thus, we decided to keep the original 16 ATAC subgroups as the primary framework. Nevertheless, our high-resolution analyses showed that ATAC-seq has the capacity to resolve finer layers of epigenetic heterogeneity within these major subgroups. We believe this also demonstrates the versatility of ATAC-based profiling in defining stable AML subgroups.

In normal blood differentiation, ATAC patterns can give an indication of the cellular potential pathway for differentiation, etc. I suspect that there are similar latent pathways within AML cells that might be determined by ATAC and predictive of the response of cells to perturbations such as therapy. Can the authors expand upon inter and intra-tumour heterogeneity, link this to ATAC subgroup and possibly look for any correlations with differences in outcome, either within or between subgroups? For example, within the larger subgroups does degree of heterogeneity read out into differences in outcome for individual patients?

(REPLY)

In response to the reviewer's suggestion, we investigated the relationship between chromatin-based heterogeneity and clinical outcome. For this purpose, we used the deconvolution data of 13 hematopoietic cell-type fractions derived from bulk ATAC-seq (Fig. 1c) and quantified their prognostic impact using a LASSO Cox regression model, deriving a deconvolution-based risk score for each patient (ED Fig. 9a).

Of interest, primitive compartments, such as HSC, were associated with poorer prognosis, whereas progenitor compartments, including CMP and GMP, correlated with favorable outcomes. This observation is in line with previous notions that LSC signatures, which share transcriptional programs with normal HSCs, are linked to adverse prognosis in AML (Ng et al., Nature 2016, PMID 27926740).
Patients with higher deconvolution-based risk scores exhibited significantly worse OS (**ED_Fig. 9b**). Notably, the distribution of this score varied among ATAC subgroups (**ED_Fig. 9c**), and the median risk score for each subgroup was inversely correlated with its corresponding 5-year overall survival rate (**ED_Fig. 9d**). These results suggest that differences in differentiation patterns, captured through chromatin accessibility, may underlie the distinct clinical outcomes observed between subgroups.
We have incorporated these analyses into the revised manuscript:

Lines 415-421:

"Of interest, we found that ATAC-inferred differentiation patterns significantly correlated with these outcome differences. To see this, we constructed a deconvolution-based risk score by calculating the weighted sum of cell-type fractions, using coefficients derived from a LASSO model trained on OS (ED Fig. 9a). This score was significantly associated with OS, varied across ATAC subgroups, and was higher in subgroups with poor prognosis (ED Fig. 9b-d), suggesting that subgroup-specific differentiation states, as reflected in chromatin accessibility, may partly explain prognostic heterogeneity."

Minor comments

A. Figure 2a. It would be more helpful if some of the major genes/clusters were named and annotated. This would allow linkage to Figure 2b where programmes are pointed out.

(REPLY)

Thank you for this suggestion. Based on the reviewer's suggestion, we have revised **Fig. 2a** by annotating representative genes that best highlight inter-subgroup differences, particularly genes associated with hematopoietic differentiation (e.g., CD14, CD34) and leukemia-related transcriptional programs (e.g., HOXA9, MEIS1). These genes were selected based on known biological relevance and their subgroup-specific expression patterns in our dataset.

B. Please check the references in the figure legends.

(REPLY)

Thank you for pointing this out. We carefully reviewed all references cited in the figure legends and updated them.

Responses to Reviewer #2

>>> Reviewer #2

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

(REPLY)

Thank you very much for your review on our manuscript. Please find our point-by-point responses in other sections.

Responses to Reviewer #3

>>> Reviewer #3

General comments:

This study reports a large cohort of AML (and somewhat peripheral) hematologic malignancies through the lens of ATAC-seq analysis, with correlation to RNA-seq, and for subsets of patients, gene mutational analysis, genome sequencing, superenhancer analysis and drug sensitivity testing. The study reports groups of patients defined by ATAC seq, and their correlation with genomic features, expression, TF motifs and drug responsiveness/outcome. Large datasets of AML with multiomic data to rigorously define subgroups and correlation with clinical features are valuable, and this is, in some respects, such a large cohort and the data will likely be extensively utilized. There are several major issues:

(REPLY)

Thank you very much for your helpful comments on our manuscript. In response to your suggestions, we have conducted additional experiments/analyses and revised the text. As the reviewer correctly noted, this study characterizes a large AML cohort "through the lens of ATAC-seq analysis." We do not claim that this represents the only possible classification of AML; rather, our aim is to provide an alternative framework that reveals novel aspects of AML biology not captured by genetic analyses alone.

The core of the study is that the ATAC approach identifies cases/groups not evident from alternative approaches. From the data presented, my interpretation is that ATAC recapitulates drivers, and identified putative "phenocopies" that are not explored. The claim that only 4 groups are clearly correlated with genetics is an underestimate: at least 10 have striking correlation with genetic features.

(REPLY)

We appreciate the opportunity to clarify this important point. We completely agree with the reviewer that striking correlation with genetic features was not confined to 4 groups but found in most ATAC subgroups, including subgroups A (with *PML::RARA*), B (with *RUNX1::RUNX1T1*), C (with *CBFB::MYH11*), D, E, and F (*NPM1* mutations), G (KMT2Ar), H(mutated *RUNX1, TET2, RAS,* etc), I (bZIP *CEBPA* mutations), J (other *CEBPA* mutations with *TET2* mutations), K (mutated *RUNX1* and splicing factors), L (*IDH1/2* mutations), M (*DDX41* mutations), and N, O, and P (*TP53* mutations). However, among these, only 4 (A, B, C, and I) perfectly matched subgroups defined in ICC or WHO (2022) classification. In other subgroups, the correlated mutations were not specific enough to capture the ATAC-defined subgroups that showed distinct features of cell differentiation program, gene expression, and prognoses (non-bZIP *CEBPA* mutations, and *DDX41* and *TP53* mutations) or do not conform to known standalone AML subtypes (H, K, L, and M).

1) *NPM1*-mutated AML, defined as a single entity in the ICC and WHO 2022 classifications, is further subdivided by ATAC into three subgroups (D, E, F), which differ in chromatin accessibility, differentiation state (D/E: immature; F: more differentiated), co-mutation patterns (D: *TET2/IDH* and splicing factor mutations; E: *DNMT3A*; F: *DNMT3A* plus RAS), gene expression (**Fig. 2a**), and prognosis (F > D > E, **Fig. 5a**).

2) In other subgroups (G, H, and K-P), no single mutation or combination of mutations can define these subgroups, because they are not specific to these subgroups but also found across multiple subgroups. This highlights a common problem in AML classification based solely on gene mutations, as exemplified by the statement of the WHO authors, *"RUNX1* mutations in AML overlap with such a broad range of defining molecular features that it was determined to lack enough specificity to define a standalone AML type (WHO 2022)." Consistent with this, *RUNX1* mutations were highly enriched in subgroup K, but this subgroup accounts for only 31% of all *RUNX1* mutations with remaining mutations widely distributed across other subgroups. Importantly, this subgroup is characterized by high sensitivity to ABL inhibitors, which is not found in *RUNX1*-mutated cases outside this subgroup. Likewise, other ATAC subgroups are also characterized by distinct profiles of gene expression, cell differentiation, survival, and drug sensitivities, yet cannot be identified by gene mutations alone. Based on these observations, we suggest that ATAC subgroups reveal an additional layer of heterogeneity not detected by gene mutations.

We have revised the manuscript as follows:

Lines 202-212:

"These 'epigenetic' subgroups showed a unique enrichment of gene mutations and other genetic lesions (Fig. 1c) and significantly overlapped with known genetic subtypes of AML in the WHO and ICC classifications[5,6], indicating a strong link between ATAC subgroups and genetic abnormalities (ED_Fig. 2g-i). However, except for four subgroups defined by common gene fusions, i.e., *PML::RARA* (subgroup A), *RUNX1::RUNX1T1* (subgroup B), and *CBFB::MYH11* (subgroup C) or by *CEBPA* in-frame basic leucine zipper (bZIP) mutations (subgroup I), they were not perfectly aligned with existing genetic subgroups. Instead, they showed a complex mapping pattern, where a single epigenetic subgroup was projected onto multiple genetic subgroups in the WHO and ICC classifications, or vice versa (ED_Fig. 2h-i), representing novel AML subtypes not previously defined by the genomic classification."

Lines 470-475:

"A central finding of our study is the identification of distinct epigenetic subgroups of AML using large-scale ATAC data associated with unique clinical and genetic features. Despite their striking correlations with common driver mutations in AML, most ATAC subgroups have not been identified by gene mutations alone or recognized as standalone subtypes in conventional genetic classifications, but represent novel AML subtypes, unveiling epigenetic heterogeneity of AML."

There is little exposition in the text about how the groups were defined. Some are distinct, most are not. Algorithms may assign groups but for those where there is overlap (e.g. Fig 1b) how were the groups rigorously defined and benchmarked?

(REPLY)

We thank the reviewer for this important point. As described in the Methods (Lines 836-849), the ATAC subgroups were defined using Leiden clustering on the top 50 principal components derived from the ATAC-seq signal matrix (based on the top 3,000 variable peaks). The clustering was performed in the

high-dimensional PCA space, not in the two-dimensional UMAP shown in Fig. 1b. UMAP is a nonlinear visualization method that may distort global distances to emphasize local relationships; thus, some apparent overlaps in UMAP do not necessarily indicate a lack of separation in the true clustering space. We also shared the reviewer's concern and validated the stability and reproducibility of our clustering by resampling analysis (100 iterations, each using 90% of the cohort). Across these resamples, we calculated the Adjusted Rand Index (ARI), co-clustering accuracy, and subgroup-specific consistency scores. We obtained a high mean ARI (0.79) (**SFig. 4a**), indicating global stability. Pairwise co-clustering accuracy had a median of 0.97 **(SFig. 4b)**, showing that most samples consistently cluster together. Most subgroups showed high consistency scores, although subgroups O and P had lower scores (median 0.76 and 0.64, respectively; **SFig. 4c–d**).

We have revised the main text accordingly:

Lines 194-197:

"The reproducibility of our clustering was benchmarked by high Adjusted Rand Index (ARI) (mean=0.79) and co-clustering accuracy (mean=0.97) in resampling analysis (SFig. 4a-c), although the consistency was significantly lower in subgroups O and P (SFig. 4d)."

Fig 1C shows that the majority of groups have enrichment for specific drivers. Thus, ATAC appears to indicate known groups, rather than defining new. For cases in these groups lacking an annotated genetic driver, it is not clear which lack complete sequencing, and which have been pursued with comprehensive sequencing to identify an unsuspected driver. This could be very powerful.

(REPLY)

Thank you for raising these points.

For the first, we respectfully disagree that ATAC merely reflects known groups. With the exception of four subgroups (A, B, C, and J), most ATAC subgroups did not correspond to known genetically defined entities and therefore represent novel AML subtypes. This point is also addressed in our response to the previous comment.

For the second, **all cases** (n = 1,563) underwent deep targeted sequencing (>1,000× coverage) with a panel covering 331 known or candidate driver genes, together with 1,158–1,317 SNP loci for allelic imbalance and genome-wide CNA assessment (Fig. 1a). This approach has been extensively validated in our prior publications (PMIDs: 39715854, 36989067, 34239136) and provides highly sensitive and reliable detection of driver mutations, structural variants, and CNAs. In addition, for cases with both targeted sequencing and WGS (n = 213), comparative analysis showed that nearly all known driver mutations were consistently captured by targeted sequencing (Reviewers' Fig. 2). WGS occasionally identified additional lesions, such as *KMT2A* and *NUP98* rearrangements, but these were relatively rare and did not alter the overall subgroup structures or biological interpretations. These results underscore the comprehensiveness and robustness of our sequencing approach. A summary of the omics data obtained for each patient is provided in Supplementary Table 3.

Multiple prior studies have used a similar approach, but with methylation profiling, to define groups in AML. The correlation with external datasets is appreciated, but this notwithstanding, the authors should perform methylation profiling of at least a subset of their cases (e.g. EPIC or similar) to provide a reference.

(REPLY)

Thank you for this constructive suggestion. In response, we newly performed DNA methylation profiling using the Illumina MethylationEPIC array for 424 AML cases in our cohort (**Fig. 1a**). This analysis revealed that each ATAC-defined subgroup exhibits a distinct DNA methylation profile (**SFig. 3**), further supporting the biological and clinical relevance of our chromatin-based classification.

To formally assess the relationship between chromatin accessibility and DNA methylation, we examined genomic regions where ATAC peaks overlapped with CpG sites covered by the EPIC array (**SFig. 3b**). We then performed statistical correlation analyses, revealing a negative correlation between chromatin accessibility and DNA methylation in all subgroups, with a median Pearson's *r* of –0.49 (**SFig. 3c**). These results confirm that the two epigenetic layers are tightly coupled in AML.

Meanwhile, we also acknowledge that ATAC peaks and CpG sites do not always co-localize. Therefore, ATAC-seq and DNA methylation profiling provide complementary information, each capturing different aspects of the regulatory landscape. We anticipate that our dataset will serve as a valuable reference for future integrative epigenomic studies, especially in understanding regulatory contexts where either modality may offer distinct biological insights.

This new analysis has been added to the main text:

Lines 168-174:

"For all patients, we performed targeted-capture sequencing (target-seq) of 331 known/candidate driver genes and 1,158–1,317 single nucleotide polymorphism (SNP) loci for genomic copy-number detection. In a subset of patients, RNA-seq, whole-genome sequencing (WGS), array-based DNA methylation analysis, chromatin immunoprecipitation sequencing (ChIP-seq), and single-cell RNA and ATAC sequencing (scRNA/ATAC-seq) were also performed (Fig. 1a and STables 1-4)."

Lines 191-194:

"As expected, DNA methylation profiling in a subset of samples (n=424) revealed a distinct methylation pattern in each ATAC subgroup (SFig. 3a), where chromatin accessibility and DNA methylation levels were significantly and inversely correlated across all subgroups with a median Pearson's *r* of -0.49 (*P*<0.001) (SFig. 3b-c)."

Lines 465-467:

"We conducted an integrated analysis of the AML epigenome, generating a large dataset comprising ATAC-seq, DNA methylation, ChIP-seq, and multi-omics scRNA/ATAC-seq data (Fig. 1a)."

We have also modified the Methods section, accordingly.

Please provide mode details on how the cohorts were enrolled and treated - particularly important for the outcome analyses.

(REPLY)

Thank you for this important point. We have revised the manuscript to provide more detailed information on patient enrollment and treatment protocols. Specifically, we now clarify that patients were consecutively diagnosed and enrolled between 1997 and 2022 in Sweden and between 2011 and 2022 in Japan. All samples analyzed in this study were collected at the time of initial diagnosis before treatment. While no preselection criteria were applied with respect to clinical characteristics, sample inclusion was contingent upon the availability of viable cryopreserved tumor cells suitable for ATAC-seq. Diagnoses were made at each participating institution at the time of initial presentation in accordance with the WHO classification in use at the time. Treatment was administered according to institutional standards of care, with a subset of patients participating in clinical trials. Detailed information on diagnosis, treatment, and clinical annotations for each cohort is summarized in the updated **STable 1**. These revisions are now reflected in the updated Methods section:

Lines 711-728:

"A total of 1,563 patients diagnosed with AML and related neoplasms were consecutively enrolled from the participating institutes and hospitals between 1997 and 2022 in Sweden and 2011 and 2022 in Japan. While no preselection criteria were applied with respect to clinical characteristics, sample inclusion was contingent upon the availability of viable cryopreserved tumor cells suitable for ATAC-seq for consecutively diagnosed AML patients. Diagnoses were made at each participating institution and hospital in accordance with the WHO classification in use at the time. In Sweden, patients were registered through the national AML registry, which prospectively collects clinical and genomic data on all newly diagnosed cases. In Japan, patients were diagnosed and treated at participating hospitals, with biospecimens and clinical information sent to and managed by the Kyoto University biobank, where clinical annotations were updated annually. Treatment was administered according to institutional standards of care and, for Swedish patients, according to the National Swedish Treatment Guidelines for AML, with a subset of patients participating in clinical trials. All samples analyzed in this study were collected at the time of initial diagnosis before treatment. Detailed clinical annotations, including diagnosis, demographic and laboratory data, treatment regimens, and outcomes, were extracted from electronic medical records and the Swedish AML Registry. No statistical methods were used to predetermine sample size. Patient characteristics and diagnosis of patients are summarized in Stables 1-2."

Were any novel mutations/alterations identified from the genomic analyses? (Lines 156-)

(REPLY)

No, we did not identify novel mutations/alterations. Our target panel included several candidate novel driver genes. During the course of this study, however, all potentially novel driver mutations/alterations were independently reported, including *MED12* (1.6%)*, PHIP* (1.3%)*, MYB* (1.2%)*,* and *UBTF* (0.8%). Nevertheless, owing to the large cohort size, our study enabled a more precise evaluation of their frequencies. Accordingly, we have revised the text as follows:

Lines 176-178:

"Less common driver mutations, including those more recently reported, were also identified, including *MED12, PHIP*[24], *MYB*[25], *and UBTF*[26]."

Line 175- (and figs) "unique lineage commitment" is not supported by these data, that show some variation between subtypes"

(REPLY)

We agree that the original wording "unique lineage commitment" may have overstated the interpretation of the data. In light of this, we have revised the text to describe these patterns more carefully, now referring to them as "**characteristic patterns of lineage bias**" rather than unique commitments (line 200). We believe this wording more accurately reflects the probabilistic nature of the deconvolution analysis and the partial overlaps observed between subtypes.

Several statements are too strong. E.g. line 97 "a definitive role of the epigenome" - rather, it could be a definitive role of the genome, and the epigenome is secondary.

(REPLY)

We understand the reviewer's concern. Our intention was to emphasize the critical role of the epigenome in cell fate and phenotype determination in AML. To avoid potential overstatement, we have revised the wording from "the definitive role of the epigenome" to "the key role of the epigenome" (Line 112).

Line 129: "Chromatin" is vague - what is intended here?

(REPLY)

We agree that the original wording "chromatin" was vague and potentially confusing. Our intention was to emphasize chromatin modifications, in contrast to DNA methylation. We therefore avoided the vague term "chromatin" and explicitly used "chromatin modifications" instead. The paragraph has been extensively revised as follows:

Lines 146-153:

"Chromatin modifications represent another layer of the epigenome. In concert with DNA methylation and other components of the gene expression machinery, such as enhancers and transcription factors (TFs), chromatin modifications play a pivotal role in the regulation of gene expression[11]. However, compared to DNA methylation, chromatin modifications have been less intensively studied in cancer using large sample sets because they entail a wide variety of histone modifications and higher-order structural variations, which require complex experimental platforms for evaluation. In this study, therefore, we focused on chromatin accessibility."

ED Fig 4E: (chromatin accessibility and methylation): this does not appear to be a formal correlative analysis as suggested in the text (line 258)

(REPLY)

We appreciate the reviewer's comment. In response, we have now formally conducted a statistical correlation analysis between chromatin accessibility and DNA methylation using the newly generated methylation profiles from 424 AML samples analyzed by the Illumina EPIC array (**SFig. 3**). Specifically, we statistically evaluated the correlation between ATAC-seq signals and DNA methylation levels (beta values) at CpG loci where ATAC peaks overlapped with EPIC probes targeting CpG regions in each subgroup. We found a consistent and significant inverse correlation between chromatin accessibility and DNA methylation across all subgroups, with a median Pearson's *r* of –0.49 (**SFig. 3c**). These results reinforce the biological validity of chromatin accessibility as a key readout of epigenetic regulation in AML. The text and figure legend have been revised accordingly.

Lines 191-194:

"As expected, DNA methylation profiling in a subset of samples (n=424) revealed a distinct methylation pattern in each ATAC subgroup (SFig. 3a), where chromatin accessibility and DNA methylation levels were significantly and inversely correlated across all subgroups with a median Pearson's *r* of -0.49 (*P*<0.001) (SFig. 3b-c)."

Responses to Reviewer #4

>>> Reviewer #4

In this study Ogawa, Lehmann and colleagues generated a very large ATAC-seq data set from 1563 AML patients drawing from two cohorts, complementing with exome sequencing, RNA sequencing and single-cell RNA/ATAC sequencing. It provides a multi-faceted view of AML. The authors then carried out a number of analyses, identified 16 AML subgroups that were investigated in detailed, characterized super-enhancer mediated gene regulatory networks and regulatory mechanisms, and deconvoluted bulk ATAC-seq data using cell-type specific chromatin profiles determined with scATAC-seq. They further sought to stratify patients and predict drug sensitivity by integrating the clinical features and drug sensitivity information. Overall, this is a quite significant study in terms of the patient number and representation, multiple data modalities, as well as the depth of analysis. I think it will become a landmark study in the field of AML genomics.

(REPLY)

We sincerely thank the reviewer for his/her positive evaluation and constructive comments on our manuscript.

There is one area where the authors can greatly improve this manuscript. This study has a lot of data from different assays. The analyses were anchored on chromatin accessibility based on ATAC-seq, which represent the unique aspects of this study in comparison with several previous studies relying on mutations, CNAs, or transcriptome. This is great and I don't have an issue with that. On the other hand, there have been multiple previous publications of AML focusing on the transcriptome. These studies also classified AML into different subgroups. Subgroups defined by RNA and those by chromatin can partially or completely overlap with each other, but with different names from different studies. This can confuse the researchers or clinicians in the community. Not connecting subgroups identified by RNA-seq and ATAC-seq is a missed opportunity. At the single-cell level, ones can also define cell types or subtypes based on scRNA-seq or scATAC-seq profiles. Based on the existing data from a variety of organs and tissues, in general one can separate single cells at a greater resolution into more subtypes with RNA than with ATAC. This could be due to either technical (spareness of the data) or biological reasons. There are also rare examples of one RNA cluster that can be split into more than one sub-clusters based on ATAC, indicating a greater level of epigenetic heterogeneity not captured by gene expression. With their data, the authors have a great opportunity to define and connect cell types using both RNA and ATAC. Along these lines, I offer a few specific suggestions for improvement:

(REPLY)

Thank you very much for this thoughtful and constructive comment. We fully agree that systematically connecting ATAC-based and transcriptome-based classifications will improve the clarity and relevance of our study.

In response, we performed additional integrative analyses linking ATAC subgroups with RNA-based clusters, highlighting both overlapping and distinct features between these classifications. At the single-

cell level, we also compared scRNA-seq and scATAC-seq clustering structures, illustrating how these modalities complement each other in capturing AML heterogeneity. These additions help clarify how chromatin accessibility and gene expression provide complementary layers of information and strengthen the biological interpretation of our subgrouping framework.
We have provided detailed responses to each of your specific suggestions as shown below.

(1) Cross-Referencing Transcriptome-Based Subgroups. Figure 1c connects a variety of different information to the 16 subgroups defined by ATAC-seq. It would be helpful to define AML subgroups based on transcriptome, either based on the previous studies, or de novo using the RNA-seq data generated from 1563 patients in this study. The RNA-based subgroups can be integrated into this plot, or as a separate supplementary figure. The ATAC subgroups A/B/C/D/I align quite well with specific WHO/ICC subtypes, but the others are more mixed. It would be really interesting to see how well these ATAC subgroups are supported by the transcriptome.

(REPLY)

According to the reviewer's suggestion, we performed unsupervised clustering of the 1,398 AML cases with available RNA-seq data, using the same Leiden strategy as for the ATAC-based analysis. This identified 18 distinct clusters, which we systematically compared with the ATAC subgroups (**SFig. 7**). Approximately half of the ATAC subgroups (A–F and K) were largely recapitulated in RNA-based clustering (#1–6, #10), while the majority of ATAC subgroups I and J were collapsed into a single RNA-defined cluster (#9) and most of the other ATAC-defined subgroups were projected into multiple RNA-defined clusters, where no correlations were observed (**SFig. 7b**). Compared with ATAC-based clustering, *NPM1*-mutated and *KMT2A*-rearranged cases were less consistently grouped by RNA. These results suggest that although highly correlated, chromatin accessibility and gene expression capture distinct aspects of cellular phenotypes, providing complementary layers of information for understanding AML heterogeneity. These complementary perspectives underscore the value of integrating both modalities to define biologically distinct AML subtypes.
We have added a description of these results to the revised manuscript:

Lines 215-222:

"We also performed RNA-based clustering, which has been conventionally used to understand the heterogeneity of AML[10]. It revealed 18 distinct subgroups (cluster 1-18), where significant correlations with known driver mutations were evident but less conspicuous (SFig. 7a). While half of the subgroups in both clustering approaches were highly concordant, the remaining subgroups were poorly correlated and unique to each clustering (SFig. 7b), suggesting that chromatin accessibility and gene expression are similar but distinct measures of cell states revealing complementary layers of AML heterogeneity."

(2) Integrating Single-Cell Clusters Across Modalities. The scRNA-based UMAP in Fig. S7a showed more clusters than the scATAC UMAP in Fig. S7b and Fig. 4a. Aligning, integrating and annotating these cell clusters across the two modalities can provide more insights on exactly how many cell types or subtypes

are there in AML, how each layer of molecular information contribute to the definition, which cell types or subtypes have distinct functions and/or specific contributions to pathology/prognosis.

(REPLY)

In response to the reviewer's suggestion, we made a direct comparison between scRNA- and scATAC-based clustering across our expanded dataset of 36 AML samples, now including newly analyzed 5 samples from subgroups H, N, and P, uniformly represent all ATAC subgroups. Including the data from four normal (remission) samples, both clustering modalities produced similar numbers of AML clusters (21 for scRNA-seq and 20 for scATAC-seq) (**Fig. ED_Fig. 7b and SFig. 15a**). Overall, these clusters in both clustering showed a concordant clustering pattern and largely aligned one by one, as shown in red (**SFig. 15b**). However, some scRNA clusters were degenerated, comprising cells from multiple ATAC clusters and vice versa. For example, RNA6 and AML08 comprised cells from multiple ATAC and RNA clusters, respectively. These results in single cell analysis are in line with the observations in bulk RNA and ATAC sequencing, suggesting that although highly correlated, chromatin accessibility and gene expression features reveal distinct aspects of AML pathogenesis. Unfortunately, however, the limited number of samples analyzed by scSeq precluded a reliable analysis of the impact of clusters on different modalities on pathology/prognosis.

Lines 356-361:

"A similar co-clustering pattern was also observed in scRNA-seq analysis. Both scRNA and scATAC clusters were largely concordant but some clusters in both modalities (e.g., RNA6 and AML08) were dispersed across clusters in the alternative modality, containing cells from multiple clusters, supporting the complementary roles of ATAC and RNA profiles underlying AML heterogeneity (ED_Fig. 7a, SFig. 15), as discussed in bulk RNA-based clustering (SFig. 7)."

(3) Cell-Type Specific Gene Regulatory Network Analysis. The gene regulatory networks showed in Fig 3a were identified based on bulk ATAC-seq and RNA-seq data. These networks are likely quite difference once you zoom into individual cell types or subtypes in each group. With the single-cell multiome data from 31 patients, the authors have enough data to run a cell-type specific GRN analysis.

(REPLY)

In our original analysis, we performed SCENIC+ on bulk ATAC-defined subgroups, which revealed distinct subgroup-specific transcription factor (TF) activities. In response to the reviewer's comment, we further applied SCENIC+ to the scATAC-defined cell-type clusters.

As these scATAC clusters closely corresponded to the bulk ATAC subgroups, we confirmed that the distinct TF activity landscapes observed at the subgroup level were similarly evident at the cell-type level, reinforcing the biological significance of these regulatory programs (**Reviewers' Fig. 3**). Since the scATAC clustering yielded 25 clusters, providing finer resolution than the 16 ATAC subgroups, we were able to explore TF activities with higher resolution, uncovering additional regulatory variation between related cell types within subgroups. As the scATAC clusters largely corresponded to bulk ATAC subgroups and we want to focus on the central finding, ATAC subgroups, we kept the original SCENIC+

analysis showing the ATAC subgroup-specific GRNs in the paper (ED_Fig 8d).

(4) Comparative Predictive Power of RNA- and ATAC-Based Subgroups. Similar to the analyses done in Fig. 5a/b, a parallel analysis using RNA-based subgroups would be really helpful. If the omics information is predictive of survival or informative for patient stratification, for the clinicians the ultimate questions are (i) what is the most predictive; (ii) what is the most practical assay for routine clinical applications. An unbiased comparison between ATAC-seq and RNA-seq subgroups can answer these critical questions. Even if RNA is more informative than ATAC on some aspects, that does not diminish the impacts of this study.

(REPLY)

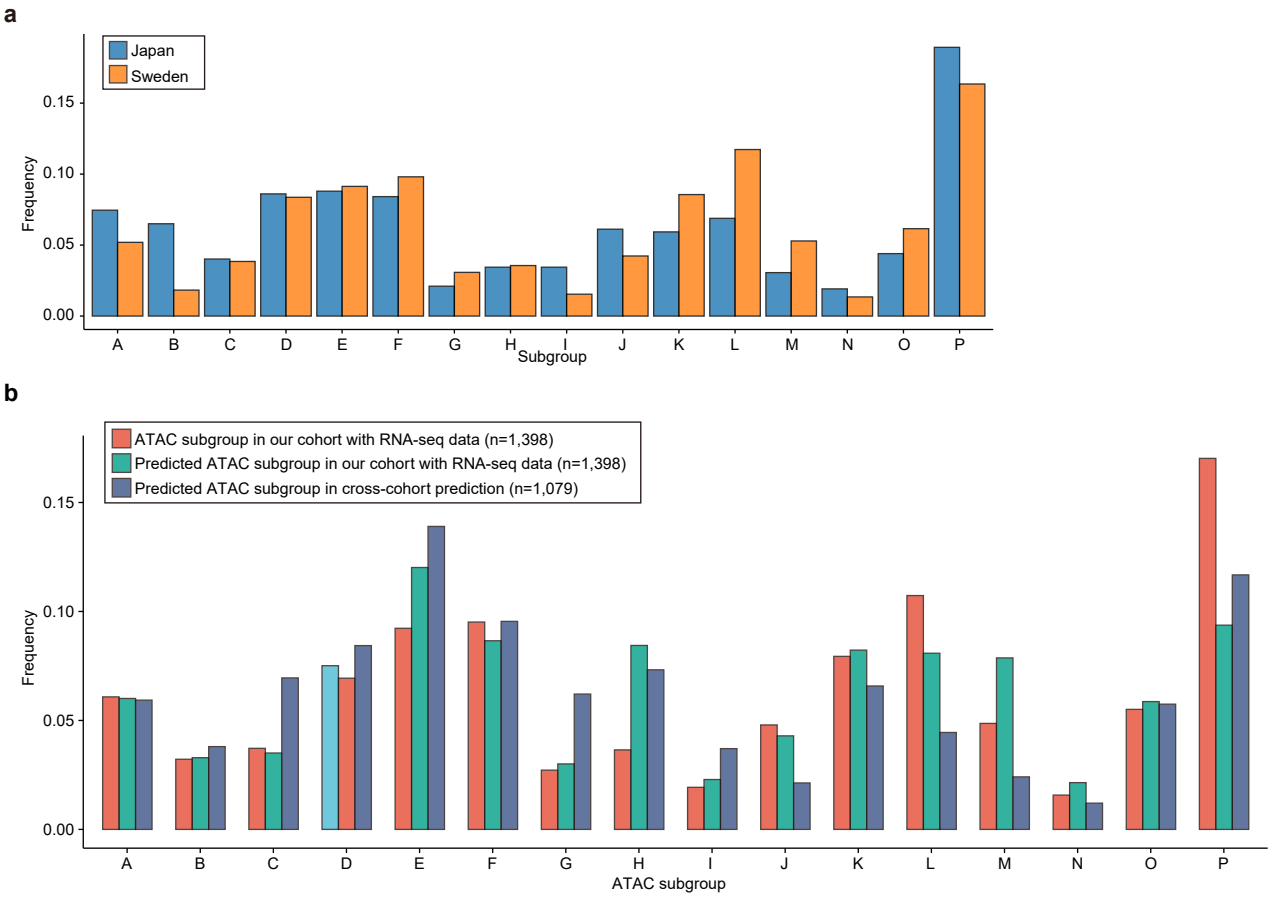
In response to the reviewer's suggestion, we investigated a prognostic value of RNA-based subgroups by performing a parallel analysis done in Fig. 5a/b (**SFig 18b-d**): Within each ELN risk category, we used multivariable Cox regression analyses incorporating both RNA clusters and clinical covariates (age, sex, and blood counts) to identify the "high-risk RNA clusters" associated with OS (**SFig 18b**). We then stratified patients in each ELN category into two groups according to the presence or absence of these high-risk RNA clusters and compared OS (**SFig 18c**). However, unlike the case with ATAC-based subgroups, incorporating high-risk RNA cluster information did not significantly improve the risk stratification of ELN, as shown in **SFig 18d**.

We have included the results in the revised manuscript as follows:

Lines 432-433:

"Of note, such improvement of prognostication was not obtained by RNA-based subgroups (SFig. 18)."

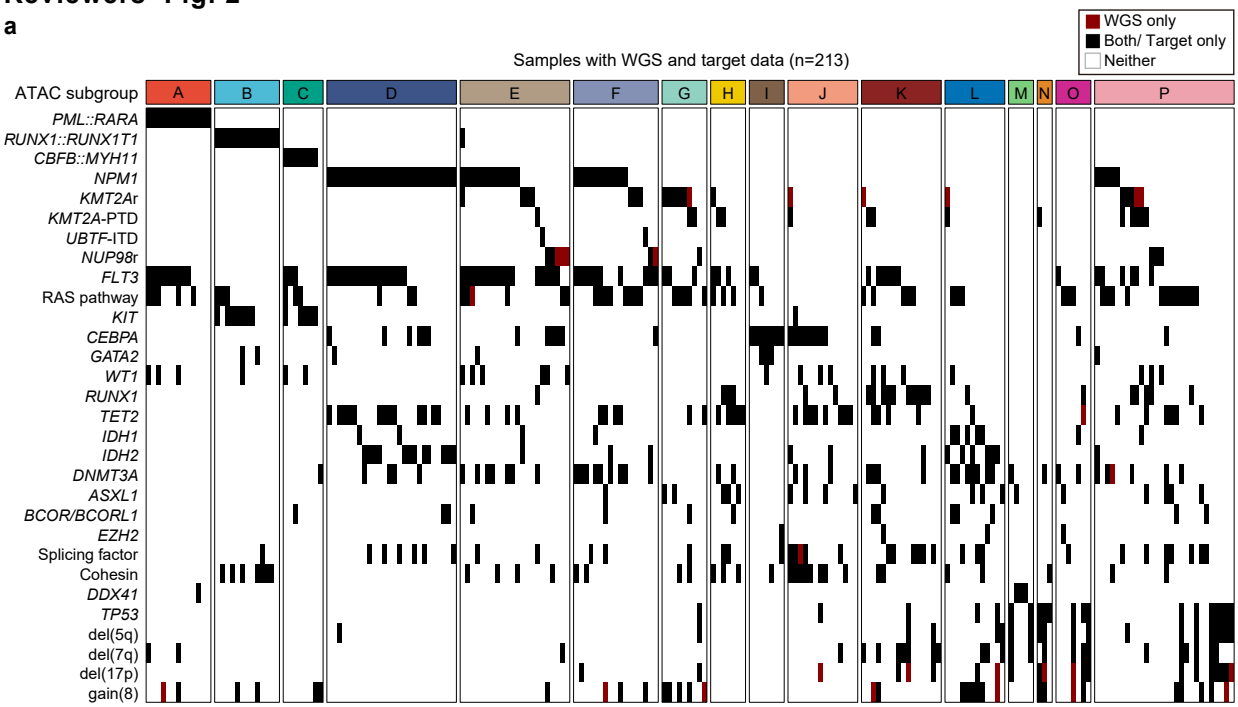
Reviewers' Fig. 1



Reviewers' Fig. 1 Frequency of predicted ATAC subgroups in our and external cohorts. **a.** Frequencies of ATAC subgroups identified in our cohort and of predicted ATAC subgroups in our cohort and external cohorts. **b.** Frequencies of ATAC subgroups in Japanese and Swedish cohorts.

Reviewers' Fig. 2

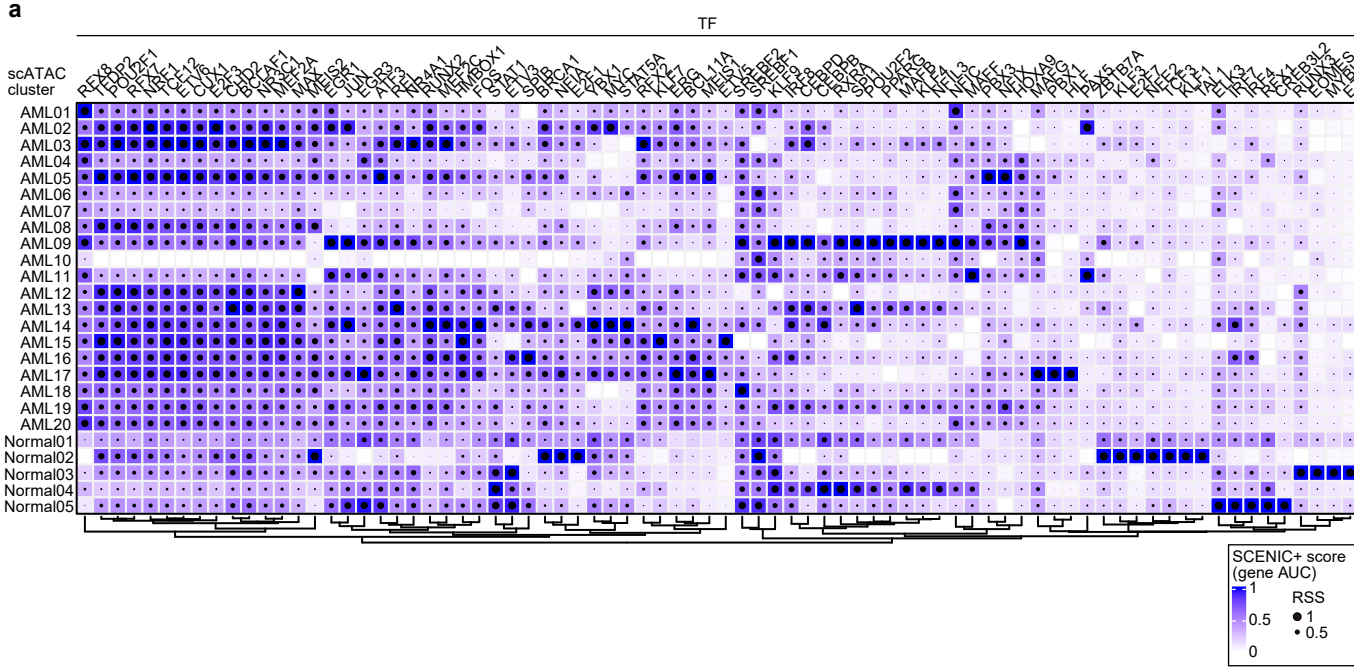
a



Reviewers' Fig. 2 Summary of genetic abnormalities by targeted-capture sequencing and whole-genome sequencing. **a.** Summary of genetic alterations identified by targeted-capture sequencing and WGS. Alterations detected by WGS but not by targeted-capture sequencing are highlighted in red.

Reviewers' Fig. 3

a



Reviewers' Fig. 3 TF activities across cell types defined by scATAC-based clustering. **a.** Heatmap summarizing activities of TFs that were identified in both ANANSE and SCENIC+ analysis (ED_Fig 8d). Color and circle size indicate the SCENIC+ score (gene-based AUC) and regulon specificity scores (RSS) in SCENIC+[46] analysis for each TF. Each row represents the scATAC cluster.

Point-by-point responses to the reviewers' comments.

Responses to Reviewer #1

>>> Reviewer #1

The authors are to be commended for their very thorough work in answering our questions and those of the other reviewers. They have answered every significant point that we have raised and particularly related to the heterogeneity of the subgroups have made the analysis more robust in its groupings. Although there is still a huge amount of data in the paper, and indeed this is one of its strong points, the authors have more strictly adhered to the central narrative and as a result, the paper is more easily understood than previously. They have also diligently addressed some of the biological inferences that we for the data and have included some in a more thoughtfully worded discussion section. In light of this significant revision, we find the paper much improved and now appropriate for publication in Nature.

(REPLY)

We sincerely thank the reviewer for the constructive comments and positive evaluation of our manuscript.

Responses to Reviewer #2

>>> Reviewer #2

I co-reviewed this manuscript with one of the reviewers who provided the listed reports.

(REPLY)

Thank you very much for your review on our manuscript.

Responses to Reviewer #3

>>> Reviewer #3

The authors have made several improvements to the manuscripts, including clarification of methods and incorporating additional methylation data.

(REPLY)

Thank you very much for your review on our manuscript. Please find our point-by-point responses below to address your remaining concerns.

There is a persisting major concern that in part arises, potentially from the comprehensiveness of genomic profiling, and then the downstream implications. The authors persist in saying that only 4 groups are associated with genomic alterations (raising complete correlation as being required), and conversely, that several groups have no correlation and are novel groups that transcend (e.g.) the ELN classification. As stated in my original review, and as acknowledged by the authors in response (and to a limited degree, in the revised MS), many of the groups show quite strong associations with genomic alterations, even if this is incomplete.

(REPLY)

Thank you very much for the opportunity to clarify this important point. We respectfully disagree with the interpretation that we state that only 4 groups are associated with genomic alterations (raising complete correlation as being required). On the contrary, we fully agree that *all* subgroups show strong associations with genomic alterations, even if they are incomplete. We have clearly stated this in both the revised manuscript and our previous response, where we repeatedly emphasized that this strong correlation of mutations in each subgroup is one of the cardinal features of ATAC-based subgroups, even though the association is incomplete in most subgroups. That said, we believe that it would be appropriate and important to clarify that, to the best of our current knowledge, these subgroups cannot be defined solely by genomic alterations because of the lack of specificity of the enrichment, except for the 4 subgroups (A-C, and I).

Abstract, lines 107-110:

"By integrating multi-omics analyses of genome, transcriptome, and major histone marks, we show that these epigenetic subgroups exhibit unique features in clinical presentation, gene mutations, differentiation states, gene expression, and super-enhancer profiles, which are validated across independent cohorts."

Main text, lines 202-206:
"These 'epigenetic' subgroups showed a unique enrichment of gene mutations and other genetic lesions (Fig. 1c) and significantly overlapped with known genetic subtypes of AML in the WHO and ICC classifications[5,6], indicating a strong link between ATAC subgroups and genetic abnormalities (ED_Fig. 2g-i)."

Discussion, lines 470-471:
"A central finding of our study is the identification of distinct epigenetic subgroups of AML using large-scale ATAC data associated with unique clinical and genetic features."

Response
"We completely agree with the reviewer that striking correlation with genetic features was not confined to 4 groups but found in most ATAC subgroups, including subgroups A (with *PML::RARA*), B (with *RUNX1::RUNX1T1*), C (with *CBFB::MYH11*), D, E, and F (*NPM1* mutations), G (KMT2Ar), H(mutated *RUNX1, TET2, RAS,* etc), I (bZIP *CEBPA* mutations), J (other *CEBPA* mutations with *TET2* mutations), K (mutated *RUNX1* and splicing factors), L (*IDH1/2* mutations), M (*DDX41* mutations), and N, O, and P (*TP53* mutations)."

This may reflect several factors, including the genomic complexity of AML (there may be multiple mutations in single or multiple genes in a case), and the adequacy of genomic profiling. From their response and the revised manuscript, it is not clear if there has been comprehensive sequencing (e.g. WGS + WTS) for cases in a "less completely" associated group, and the putative novel groups. This is really important. Multiple recent large scale genomic studies have identified unexpected drivers, including in the non-coding genome, when subjected to comprehensive sequencing. Here the authors state their analysis has been limited to known cancer drivers, and using a candidate (coding) sequencing approach, and what appears to be relatively sparse SNP sampling for copy number analysis. To this reviewer, this is not adequate to resolve this point. AML is a genomic disease, in which somatic mutations often affect regulators of the epigenome. It would thus be expected that any putative novel group may indeed have distinct drivers, or distinct patterns of alteration of drivers (indeed the authors do deploy this strategy in a limited way, for example for the NPM<1 + TET2/IDH group).
(REPLY)
As mentioned in our response to the reviewer#3, our analysis was not confined to known cancer drivers and copy number alterations through targeted capture sequencing but also included WGS for a subset of cases (n=213), covering 7.0-26.5% of each subgroup. However, in the previous revision, we did not include the analysis of previously unreported alterations, because they were mostly very rare and their functional relevance was often unclear.

To address the reviewer's concern, we have now performed a more comprehensive analysis of genomic alterations in this subset of AML, including both known and previously unreported drivers, using the data from our ongoing large-scale WGS study including 1,173 cases with myeloid neoplasms (578 AML and 595 MDS), which included the 213 AML cases from the current study (Redacted). WTS w as also available for 436 cases (including 137 of the 213 AML cases).

1) Evaluation of known driver alterations:
Across the 213 cases, we detected a total of 753 mutations/alterations in known driver genes using e ither targeted sequencing or WGS, of which only 49 (6.5%) were newly detected by WGS/WTS, i ncluding deep intronic mutations causing mis-splicing and other structural variations (SVs) disrupting c oding sequences (**Redacted**). These findings suggest that targeted sequencing already captured most o f the "known" driver alterations and that even combining these newly detected "known" drive r alterations, none of the known driver alterations, nor their combinations, were specific enough t o uniquely define these ATAC subtypes except for subgroups A, B, and C, or explained incomplet e associations.

2) Exploration of novel drivers using large-scale WGS:
We then investigated alterations *outside of known driver genes* using WGS data. Our analysis of 1,173 cases not only recapitulated all previously reported driver alterations (PMIDs: 23634996, 27276561, 30333627) but also detected a total of 37 candidates for novel driver genes that were affected by SNVs and small indels, focal amplifications and deletions as small as 1 kb, and large SVs causing novel gene

fusions, gene disruptions, and enhancer hijacking (Redacted). However, the frequency of abnormalities in these novel driver genes was uniformly very low (<1%) and only 12 novel alterations outside of kno wn driver alterations were newly detected in the 213 cases (Redacted), which precluded the identificati on of a set of driver mutations or their combinations that could uniquely define these ATAC subgroups other than A-C and I.

3) Interpretation and limitations:
We are aware that the algorithms and pipelines currently used to detect novel driver alterations based on WGS/WTS are not perfect but have their own limitations, and that we cannot completely exclude t he possibility that there remain some unknown drivers not detected even in the large-scale WGS analysis. Nevertheless, based on the results from (1) and (2), it would be very unlikely, to the best of our current knowledge, that the incomplete associations between ATAC subgroups and enriched driver alterations can be fully resolved or that any putative novel group may have distinct drivers, or distinct patterns of alteration of drivers even in this small subset. This is mainly due to the lack of specificity of common mutations and the rarity of newly identified driver alterations. A representative example is s hown for subgroup K, enriched for *RUNX1* mutations, in which WGS failed to newly identify *RUNX1*-in volving alterations in cases with no *RUNX1* lesions in targeted sequencing (Redacted and 1c).

4) Role of the epigenome:
Another, and not mutually exclusive, possibility would be epigenomic alterations. The epigenome represents a major determinant of tissue/lineage commitment and plasticity even with an identical genome sequence, and can be reprogrammed during the establishment of iPS cells, underscoring its role in determining and maintaining cellular phenotypes. Recent studies also suggest that epigenetic context (such as cell of origin) can also influence disease phenotypes, including AML heterogeneity (PMIDs: 21248838, 28729483, 35022204).

5) Conclusion:
Regardless of whether these ATAC subgroups are defined solely by genetic alterations or whether epigenomic changes may also play a primary role, which we believe is beyond the scope of this study, the novelty of our study lies in the identification of unique ATAC subgroups that, to the best of our knowledge, cannot be defined solely by known genetic abnormalities. As the reviewer correctly noted, the ATAC-based classification reveals an alternative aspect of AML through the lens of epigenome. These ATAC subgroups are characterized by distinctive enrichment of gene mutations, cell differentiation /maturation states, prognostic values, and drug sensitivity and thereby, highlighting the relevance of epigenomic classification in understanding of AML pathogenesis, developing novel therapeutics, and improving risk prediction and stratification.

Based on the above results and considerations, we revised the statement in the second paragraph of Discussion as follows:

"Despite their striking correlations with common driver mutations in AML, most ATAC subgroups are not uniquely defined solely by known genomic alterations or their combinations due to their lack of specificity to these subtypes, although we could not completely exclude the presence of some unknown genetic lesions explaining these subtypes."

As a downstream consideration - how do the authors expect these data will be translated, when clinical ATAC sequencing is unlikely to be clinically implemented?
(REPLY)
As we have already addressed the same point in the first revision in response to reviewer #1, we suggest gene expression-based diagnosis of ATAC subgroups even when ATAC-seq is not feasible in the clinics. For example, we demonstrated that subgroup K can be reliably predicted by measuring expression of 30 genes (SFig. 19b, STable 11). Similarly, high-risk ATAC subgroups for each ELN category can be predicted with the expression of 30 genes (SFig. 19a, STable 11). These findings suggest that we could potentially translate the current findings into the clinics through limited gene expression profiling. This was briefly discussed in the last paragraph in the previous manuscript.

"Finally, our ATAC-based epigenetic classification also has important clinical implications. It has

prognostic value independent of the conventional ELN classification and predicts new drug sensitivities not previously recognized. Incorporating risk ATAC subgroups significantly improved ELN-based prognostication, and MEK and ABL inhibitors may have a role in the treatment of particular ATAC subtypes. These findings provide a framework for personalized therapeutic strategies, once the ATAC-seq-based classification is implemented through more feasible diagnostic platforms. In this regard, we have developed reliable models predicting the risk ATAC subgroups based on the expression of only 30 genes, which could contribute to improved prognostication and drug selection (SFig. 19, STable 11)."

Responses to Reviewer #4

>>> Reviewer #4

In this revision the authors have provided a significant amount of new experimental (DNA methylation) data, added many analyses, and improved the writing. which together fully addressed my concerns. These efforts have lifted the quality of this manuscript above the standard for Nature, so I'm supportive of accepting it for publication.

(REPLY)

We sincerely thank the reviewer for the positive evaluation and constructive comments on our manuscript.

One small request: in Supplementary Fig. 19, I think adding the AUC curves would be quite helpful, because the prediction accuracies can depend on the specific cutoffs of choice.

(REPLY)

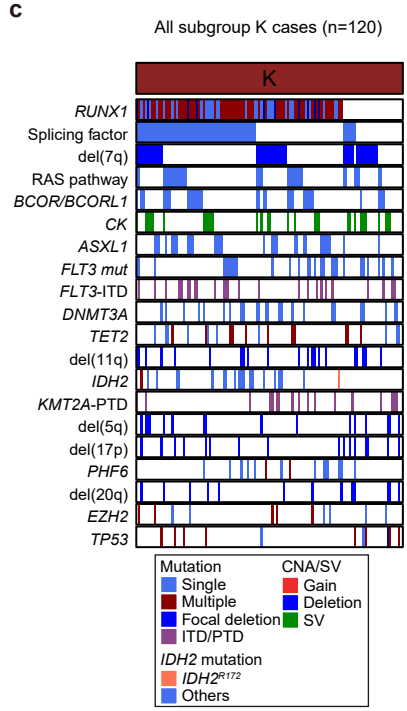
Thank you for your comment. In Supplementary Fig. 19, we employed a centroid-based prediction approach to assign samples to the subgroups without using a probability score or adjustable threshold. As such, this method provides categorical predictions rather than continuous scores. Therefore, standard AUC analysis is not directly applicable in this context, as there is no adjustable threshold to compute the accuracy.

Reviewers' Fig. 1

[PANEL REDACTED]

[PANEL REDACTED]

c



Reviewers' Fig. 1 Summary of WGS analysis and detailed mutation profiles in subgroup K.
a. Overview of WGS and ATAC cohorts (left) and genetic alterations in 213 AML cases with available WGS data (right heatmap). In the WGS analysis, both known and previously unknown genetic alterations were included. For CNAs and SVs, those with at least one breakpoint located within gene bodies were counted. For enhancer hijacking events, SVs with elevated expression of proximal genes were identified by our own pipeline. Gene alterations detected by targeted-capture sequencing are shown in gray, whereas those identified only by WGS (but not by targeted-capture sequencing) are highlighted in other colors. Data types obtained for each sample are indicated in gray at the top of the heatmap. Cases without known drivers in targeted-capture sequencing but any in WGS are shown in black on the bottom.
b. Summary of all gene alterations found in subgroup K with available WGS data. Legends are same with a.
c. Summary of the top 20 genetic alterations identified by targeted-capture sequencing complemented with RNA-seq and WGS in subgroup K cases (n = 120) (related to Fig. 1c, focusing on subgroup K).

Figure Redacted

Point-by-point responses to the reviewers' comments.

We thank the reviewers for their careful and constructive evaluation of our manuscript. In response to the editor's guidance, we clarify that our whole-genome sequencing (WGS) analysis was not limited to known driver alterations, but was performed in an unbiased, genome-wide manner to examine the association between genomic alterations and ATAC-defined subgroups.

Responses to Reviewer #3

>>> Reviewer #3

In this revised version of the manuscript, the authors have been quite responsive to this reviewer providing more clarity and additional data and analysis for (e.g.) RNAseq and methylation data, and integration thereof. The study will be a valuable resource.

(REPLY)

Thank you very much for your review on our manuscript. Please find our point-by-point responses below to address your remaining concerns.

There are persisting concerns. The main is the claim that the study has identified new subtypes of AML that transcend existing classification schema (and through the outcome analysis, the suggestion that this classification should replace existing schema). The authors response is somewhat contradictory in parts both with itself and the manuscript in terms of correlation with genetic alterations. The report 4 total, although this is really 3 as inspection of the data indicate that the association with CEBP alterations is not as absolute as the fusions; most of the other ATAC groups have clear incomplete associations with mutations, or groups of mutations, as they describe in the MS. The main outcome associations favorable and unfavorable that they report with the ATAC groups correlate with strong prognostic genomic features (fusions for good outcome, MLLr, TP53 etc for poor). Yes, there are are cases in several of the groups that do not appear explained by the genomic data, but the request for detailed WGS (at a minimum) for cases in these groups was not fully responded to: specifically, it is not clear which cases were sequenced, and focused genotyping of known drivers does not address the issue (of the need for discovery). Thus, while the dataset is large and valuable to the field, it is somewhat incremental on prior studies that have used gene expression, or methylation, to define groups of AML.

(REPLY)

We thank the reviewer for the opportunity to further clarify this important point, which we had addressed in the previous revision but may not have described with sufficient clarity.

In the current study, WGS was performed for 213 AML cases spanning all ATAC subgroups (7.0–26.5% per subgroup) (Supplementary Table 3), and these data were analyzed in a genome-wide, unbiased manner rather than being restricted to known driver alterations (Reviewers' Figs. 1-2). Importantly, these 213 cases represent a subset of our ongoing large-scale WGS study comprising more than 1,000 cases with myeloid neoplasms, which enabled comprehensive evaluation of both known and previously unreported genomic alteration (Redacted).

Across the WGS-analyzed cases, recurrent alterations newly identified exclusively by WGS accounted for only a small fraction of all events (6.5%), and incorporation of these alterations did not materially change the relationship between genomic alterations and ATAC-defined subgroups that we previously reported based on targeted sequencing ((Redacted)). While three ATAC subgroups (A–C) showed near o ne-to-one correspondence with recurrent gene fusions, no additional ATAC subgroup—including subgro up K—could be uniquely defined by newly identified coding or non-coding alterations, even in the co ntext of large-scale WGS analysis ((Redacted)).

These findings indicate that the incomplete correspondence between ATAC subgroups and genomic alterations is not primarily due to insufficient genomic profiling. Rather, they support the concept that chromatin accessibility captures biologically meaningful heterogeneity that is not fully encoded by genomic alterations alone. Accordingly, the conceptual novelty of this study lies not in identifying additional genome-defined entities, but in demonstrating that epigenomic profiling can reveal novel

subgroups and refine heterogeneity within existing genetic subtypes of AML.

The authors have provided additional data and analysis for a variety of approaches: superenhancer analysis, TF networks, sc profiling and mapping to the Zeng/Dick map: but these are observational and in each case raise notions that are left completely unexplored to demonstrate a role if a factor in biology (or response).

(REPLY)

We agree with the reviewer that functional validation of subgroup-specific transcription factors and regulatory programs represents an important future direction. As stated in the Discussion, the analyses of super-enhancers, transcription factor networks, and single-cell mappings in the present study are intended to define subgroup-specific regulatory landscapes and generate testable biological hypotheses. Comprehensive mechanistic dissection of these factors, including experimental perturbation studies, would require substantial additional experimental work and falls outside the scope of the current study, which is focused on establishing an epigenomic classification framework for AML.

The outcome analyses are interesting, as they raise the notion that there are subsets of cases that are not fully resolved with the WHO or ICC classifications, but it is a missed opportunity that the authors do not fully explore their own mutational data to determine which schema might be optimally prognostic - more nuanced analysis of the mutational data and combinations thereof, or ATAC?

(REPLY)

In this study, we intentionally used the ELN classification as the current clinical standard for prognostic risk stratification and demonstrated that ATAC-based subgroups provide additional prognostic information beyond ELN. Our goal was not to propose immediate replacement of existing genome-based classifications, but rather to show that epigenomic information adds orthogonal and clinically relevant insights.

Determining whether prognosis is best predicted by genome-based, epigenome-based, or integrated models will require larger prospective studies designed specifically for this purpose. We believe such investigations are important but beyond the scope of the present work.

Some observations are not novel or referenced (e.g. RUNX1 and lymphoid features).

(REPLY)

While previous studies have reported associations between *RUNX1* mutations and lymphoid features, our findings are distinct in several important aspects. In our cohort, only approximately 30% of *RUNX1*-mutated cases fell into ATAC subgroup K, and lymphoid features were specific to this epigenomically defined subgroup rather than to *RUNX1* mutation status per se.

Indeed, *RUNX1* mutation alone has not been sufficient to define a distinct clinico-pathological entity, as reflected in the current WHO and ICC classifications, which do not recognize *RUNX1*-mutated AML as a standalone entity. Our study demonstrates for the first time that a specific epigenomic pattern, rather than the presence of *RUNX1* mutation itself, delineates a subset of AML with characteristic lymphoid features, highlighting the added value and novelty of epigenomic classification.

The close correlation with methylation is notable, and raises the question as to which epigenetic classifier might be optimal for grouping/subtyping.

(REPLY)

The present work is deliberately focused on chromatin accessibility as a regulatory layer that directly captures transcriptional potential. DNA methylation profiling was leveraged to assess epigenetic concordance and robustness of the ATAC-defined subgroups, rather than to nominate an alternative epigenetic classification scheme.

Comparative benchmarking of different epigenetic modalities would require a study explicitly designed for that purpose, addressing a conceptually distinct question from the chromatin-centered framework developed here. Importantly, the strong concordance observed between chromatin accessibility and

DNA methylation supports the biological robustness of the ATAC-based subgroups, rather than suggesting redundancy between epigenetic layers.

The mutational data could be explored in more detail. One notable example is the TP53 mutant groups where there are associations with inferred developmental stage, and also lineage of leukemia. There are known associations between TP53 status, VAF/zygosity, and leukemia phenotype (e.g. erythroid). These features: VAF, sequence of acquisition, and mutational constellations, would be valuable to mine and associate with ATAC group in more detail.

(REPLY)

We agree that detailed analyses of *TP53* allelic status, clonal architecture, and mutational constellations represent future research directions. While such analyses are highly important, the primary aim of the present study was to establish an epigenomic framework for AML classification and to characterize its clinical and biological associations. Extensive dissection of genomic substructures within individual ATAC subgroups would detract from this central narrative and is therefore outside the scope of the current manuscript.

The expanded therapeutic data are welcome and a major feature of the study.

(REPLY)

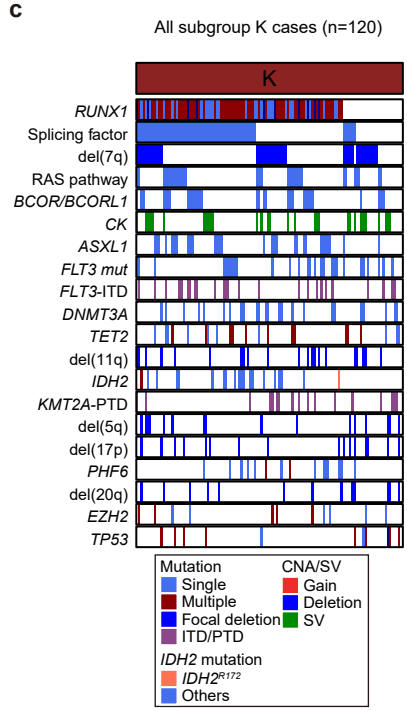
We thank the reviewer for this positive assessment. We expanded the therapeutic analyses in response to the reviewers' earlier comments, which we believe has enhanced the clinical relevance of the ATAC-based classification without altering the central conclusions of the study.

Reviewers' Fig. 1

[PANEL REDACTED]

[PANEL REDACTED]

c



Reviewers' Fig. 1 Summary of WGS analysis and detailed mutation profiles in subgroup K.
a. Overview of WGS and ATAC cohorts (left) and genetic alterations in 213 AML cases with available WGS data (right heatmap). In the WGS analysis, both known and previously unknown genetic alterations were included. For CNAs and SVs, those with at least one breakpoint located within gene bodies were counted. For enhancer hijacking events, SVs with elevated expression of proximal genes were identified by our own pipeline. Gene alterations detected by targeted-capture sequencing are shown in gray, whereas those identified only by WGS (but not by targeted-capture sequencing) are highlighted in other colors. Data types obtained for each sample are indicated in gray at the top of the heatmap. Cases without known drivers in targeted-capture sequencing but any in WGS are shown in black on the bottom.
b. Summary of all gene alterations found in subgroup K with available WGS data. Legends are same with a.
c. Summary of the top 20 genetic alterations identified by targeted-capture sequencing complemented with RNA-seq and WGS in subgroup K cases (n = 120) (related to Fig. 1c, focusing on subgroup K).

Figure Redacted

Point-by-point responses to the reviewers' comments.

Responses to Reviewer #3

>>> Reviewer #3

My key concern is not fully resolved. I appreciate that the WGS data was analyzed genome wide, but my question was to address the relationship of genomic driver and ATAC group to (1) determine the drivers in groups where there is not a complete 1:1 relationship of driver to ATAC group and (2) to determine the best outcome model (as the authors claim that this represents a new framework for classification.

(REPLY)

We thank the reviewer for the opportunity to clarify this important point. In response to the reviewer's request to address the relationship between genomic drivers and ATAC subgroups, we systematically examined whether genomic alterations, individually or in combination, could define, predict, or otherwise account for the observed ATAC subgroup structure.

1. To this end, we employed decision tree models (Breiman, 1984, Classification and Regression Trees) to test whether mutation, copy-number, and structural-variant features could robustly classify each ATAC subgroup. For each ATAC subgroup, one-versus-rest models were constructed and optimized using balanced accuracy and F1-score as performance metrics. Both indices approach 1.0 when the model achieves a complete 1:1 correspondence with a given ATAC subgroup. Models were trained on 80% of cases and validated on the remaining 20%. These models allowed us to exhaustively test whether such models could be identified by progressively increasing model complexity, defined as maximum tree depth.

   Reviewer Only Fig. 1a shows representative models, which captured Subgroups B, I, and L at maximum depth 3, with the corresponding sensitivities and positive predictive values (PPVs). While Subgroup B was captured with near-complete accuracy, Subgroups I and L failed to achieve a complete 1:1 correspondence and did not reach high balanced accuracy or F1-score. For the remaining subgroups except for fusion-defined subgroups A and C, model performance was substantially lower **(Reviewer Only Fig. 1b)** and did not meaningfully improve with increased tree depth up to 12 **(Reviewer Only Fig. 1c)**. These analyses demonstrate that, within the limits of current genome-wide knowledge, genomic drivers or their combinations cannot robustly determine most ATAC subgroups.

   Regarding point (1), these analyses demonstrate that, for ATAC subgroups lacking a complete 1:1 correspondence with genomic alterations, no individual driver or combination of drivers can be robustly identified that uniquely defines these groups based on current genomic knowledge.

2. Regarding the issue of (2) determining the best outcome model, we demonstrated that the conventional ELN prognostic model based on genomic drivers is significantly improved by incorporating high-risk ATAC subgroups (Fig. 5b,c). Because ATAC subgroups are not fully explained by genomic drivers, this supports the clinical relevance of ATAC subgroups in AML prognostication. However, our intent was not to construct or optimize a definitive outcome model integrating all genomic and epigenomic variables, which would require substantially larger cohorts with complete multi-omic profiling. Such comprehensive model optimization, while important, falls beyond the scope of the present study.

The authors report that the WGS group is a subset of the overall group, and direct to ST3, but most cases in this table appear to have a canonical driver. The request was systematic WGS of those cases in each ATAC group that do NOT have a canonical driver.

(REPLY)

We thank the reviewer for this important clarification. In our cohort of 1,563 AML cases, whole-genome sequencing (WGS) was restricted to a subset of cases (n=213, 13.6%), primarily due to limited sample availability, most often resulting from the lack of high-quality matched germline control DNA. As a result, WGS was not systematically performed in all cases lacking canonical drivers. Nevertheless, the data from this subset allowed us to assess the reviewer's concern within this group.

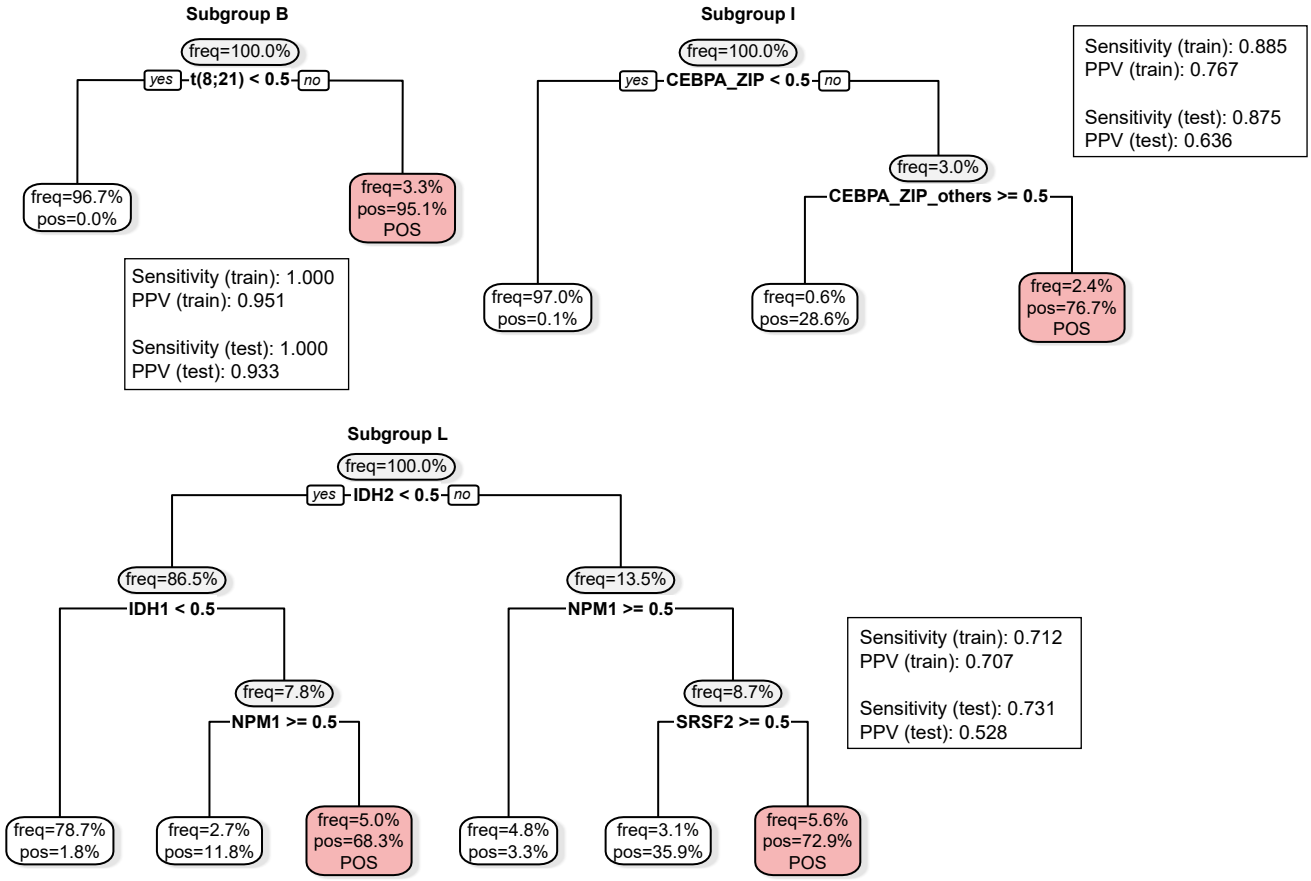
Within this subset, WGS revealed additional driver alterations that were not detected by targeted sequencing in 69 cases (32.4%). However, with the exception of *KMT2A* and *NUP98* rearrangements (three cases each in subgroups P and E), these additional alterations were observed in only one or two cases within a given ATAC s ubgroup and were too infrequent to explain cases that lacked known subgroup-defining drivers. For example, s ubgroups J and K are highly enriched for *CEBPA* and *RUNX1* mutations, respectively. In each subgroup, six ca ses lacked the corresponding canonical mutations based on targeted sequencing. These cases were fully ass essed using WGS data, but even using WGS data, no recurrent alternative driver mutations accounting for the m were identified **(Redacted)**. Taken together, even in this subset of AML cases comprehensivel y analyzed using WGS, we could not identify subgroup-defining genomic drivers that account for the cases lack ing canonical driver mutations in targeted sequencing.

We were unable to perform WGS analysis for the remaining 1,350 cases. However, in the large-scale WGS we performed for an independent set of myeloid neoplasm samples, including 578 AML and 595 MDS cases, the frequency of mutations in newly identified drivers was uniformly rare (<1%) and therefore unlikely to represent recurrent, subgroup-defining genomic drivers.

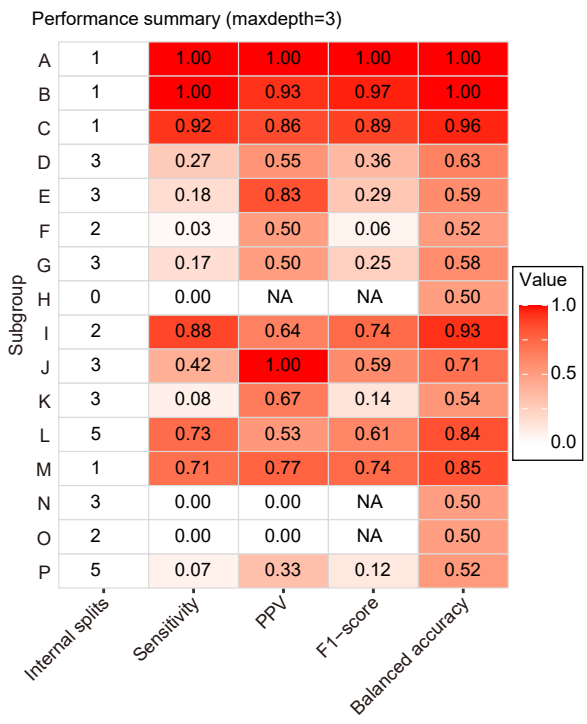
Last but not least, we sincerely thank the reviewer for his/her thoughtful suggestions and questions, which provided valuable opportunities to further clarify the relationship between ATAC subgroups and genomic drivers. We do not intend to claim that ATAC-based classification supersedes existing genomic classifications. Rather, we aim to demonstrate that ATAC-based classification provides complementary and biologically informative insights into AML pathogenesis that are not fully captured by genomic alterations alone. We hope that the revised analyses and clarifications satisfactorily address the reviewer's concerns.

Reviewer Only Fig. 1

a



b



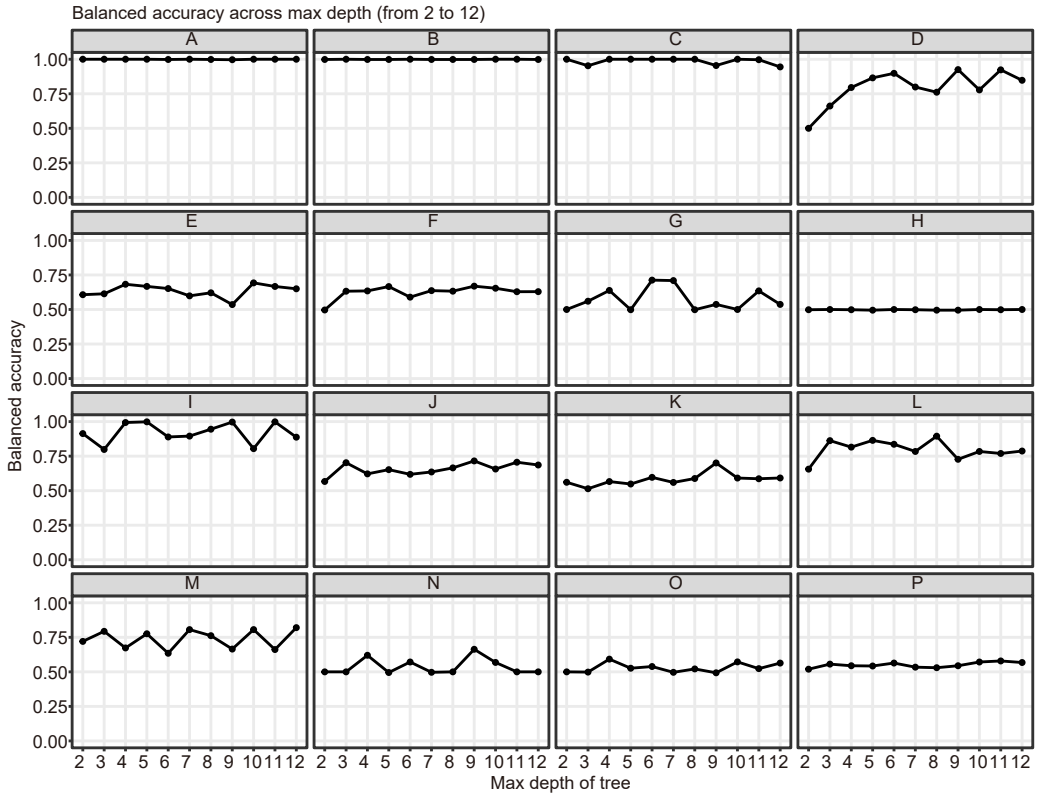
Reviewer Only Fig. 1 Genome-based decision tree analysis of ATAC subgroups.
a. Representative genome-based decision trees for ATAC subgroups. Genome-based decision tree classifiers were constructed to assess whether ATAC-defined subgroups can be predicted from genomic alterations. For each ATAC subgroup (A–P), a binary one-versus-rest classifier was trained using gene mutations, copy number alterations, and structural variants as binary input features. Models were trained using a CART framework, with samples randomly split into training (80%) and test (20%) sets. Example decision trees are shown for ATAC subgroups with relatively higher predictability (subgroups B, I, and L), with the maximum tree depth fixed at 3. Internal nodes indicate genomic features used for splitting, with thresholds representing presence (≥0.5) or absence (<0.5) of the alteration. At each node, *freq* denotes the proportion of samples reaching that node, and *pos* denotes the proportion of samples belonging to the target subgroup. Terminal nodes labeled “POS” represent leaves classified as the positive class. Sensitivity and positive predictive value (PPV) are reported for both training and test sets. Except for subgroup B, which is expected to be predictable based on known defining alterations, even relatively well-performing subgroups show only modest sensitivity and/or PPV, indicating limited identification by genome-based rules.

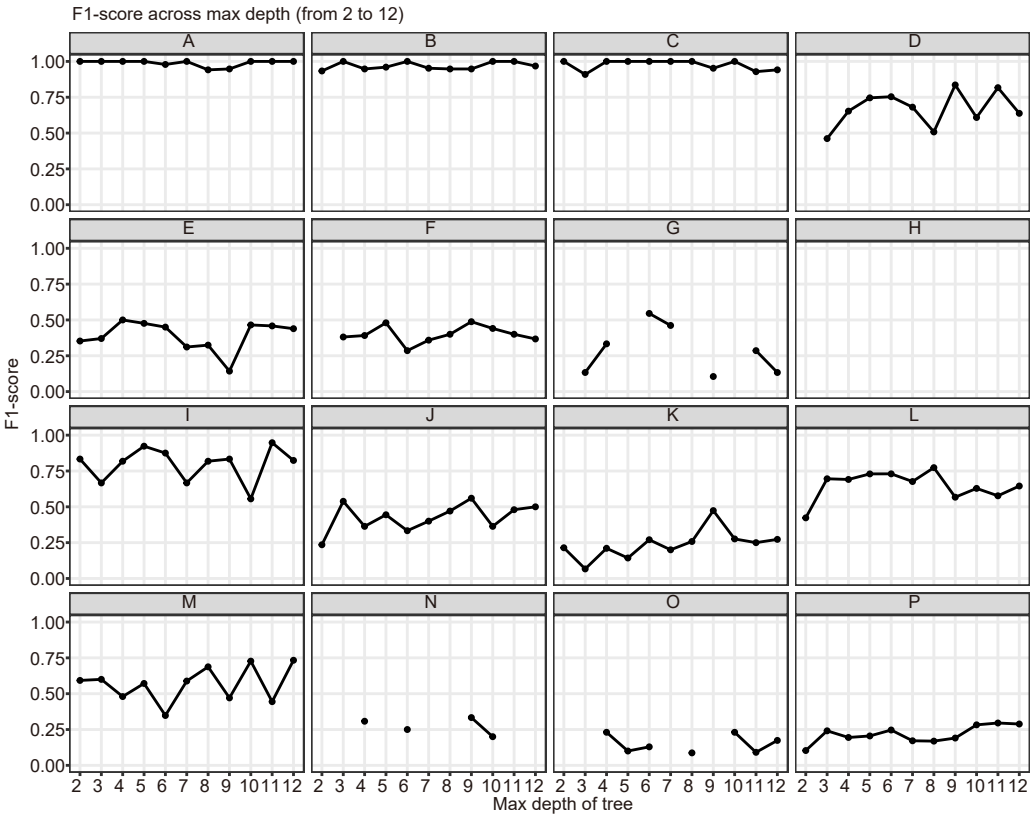
b. Performance summary of genome-based decision tree classifiers for individual ATAC subgroups. A summary matrix shows model complexity and predictive performance for each ATAC subgroup. The leftmost column indicates the number of internal splits in the final tree, reflecting model complexity, and the remaining columns show sensitivity, PPV, F1-score, and balanced accuracy evaluated on the test set. Balanced accuracy is defined as the mean of sensitivity and specificity and accounts for class imbalance, whereas the F1-score represents the harmonic mean of sensitivity and PPV. Color intensity indicates metric values from 0 to 1. Overall, only a small number of subgroups achieve high performance with shallow trees, highlighting the limited predictability of most ATAC subgroups using genome-based rules.

c. Effect of increasing tree depth on predictive performance. Balanced accuracy (top) and F1-score (bottom) are plotted as a function of increasing maximum tree depth (2–12) for each ATAC subgroup. Increasing model complexity yields only marginal performance gains for most subgroups, indicating that deeper trees do not substantially improve subgroup prediction.

Reviewer Only Fig. 1 (continued)

c





Redacted

Point-by-point responses to the reviewer's comments.

>>> Reviewer #3

The authors show that the existing detected constellations of mutations imperfectly predict ATAC-defined groups, and analysis of the limited additional WGS data in cases lacking a known driver identifies driver in some, but not all cases. It is helpful information, but does not fully resolve the question as to whether full discovery of these cases, with systematic WGS/RNA-seq/ATAC could more fully resolve the drivers of AML.

(REPLY)

We thank the reviewer for this important comment.

As noted in the previous review rounds, WGS was performed in 213 AML cases that had not previously been characterized at this level in the literature, together with integrated RNA-seq and ATAC-seq profiling. These data represent a systematic multi-omic analysis at a scale sufficient to assess whether more comprehensive genomic discovery would resolve the drivers of AML.

Within this cohort, even in cases comprehensively analyzed using WGS, we could not identify subgroup-defining genomic drivers that account for cases lacking canonical driver mutations in targeted sequencing. This indicates that the incomplete correspondence between ATAC subgroups and genomic alterations is not primarily due to insufficient genomic profiling. Rather, our results support the concept that chromatin accessibility captures biologically meaningful heterogeneity that is not fully encoded by genomic alterations alone.

We agree with the reviewer that, in principle, larger cohorts or future technologies may identify additional rare or currently unrecognized drivers. However, based on the present large-scale, integrated multi-omic dataset, such drivers, if they exist, are unlikely to represent recurrent, subgroup-defining events. A comprehensive effort to identify extremely rare or novel mechanisms would require substantially larger cohorts and therefore falls beyond the scope of the present study.

We have added the description to discuss the points as follows:

Lines 443-448:

"Meanwhile, it should be emphasized that the ATAC-based classification does not replace conventional genetic classifications; instead, it provides an alternative view of AML heterogeneity through the lens of the epigenome that is not captured by genetic approaches alone. Furthermore, we cannot entirely exclude the possibility that as-yet unidentified gene mutations may underlie or help define these ATAC subtypes."

As an additional note, the additional analyses do not appear to have been incorporated into the manuscript.

(REPLY)

We thank the reviewer for this comment. The additional analyses have now been incorporated into the manuscript and Supplementary Information.

Specifically, the relevant results are presented in Supplementary Figs. 8 and 9, and the corresponding text has been added as follows:

Lines 187-192:

"In fact, even after systematic and exhaustive decision-tree analyses using known driver lesions, we failed to identify drivers or their combinations that could uniquely define ATAC subgroups except for subgroups A–C (and partially subgroup I) (SFig. 7). Such drivers or combinations thereof were not identified by rigorously

interrogating previously unknown drivers using WGS, even within a subset (n=213) of our cases, as demonstrated in detail in subgroups J and K (SFig. 8).”