Supplementary Discussion

Single cell DNA sequencing affords capabilities with respect to genomic profiling not previously enabled by bulk sequencing approaches. The power to identify mutations that are driving AML pathogenesis and to simultaneously elucidate the mutational representation of clones and their relative contribution to disease repertoire offers substantive biologic and potential therapeutic insight. While the ability to profile mutational spectra and immunophenotype at single cell resolution is powerful, our analyses have limitations which will be attenuated as single cell DNA sequencing increases in throughput (to 100,000-1,000,000 cells/sample). Technical advances with unique molecular identifiers will help resolve further analysis into mutation order and architecture through accurate assessment of very small clones.

We analyzed each patient at a single timepoint, and analysis of serial samples will delineate how clonal evolution changes during disease progression and/or in response to anti-leukemic therapies. Moreover, our analyses of clones and trajectories assume that the clones expand and contract without inter-clonal interactions. There is growing evidence that mutant clones can interact with other existing clones and could potentially affect the growth and fitness of each other. The observation that the majority of AML samples are comprised of 1-2 dominant clones adds an important layer to this possibility, such that dominant clone(s) may outcompete minor clones through increased proliferation/self-renewal or through active, cell non-autonomous suppression of less fit minor clones as has been shown in murine models of CML where mutant cells can suppress the fitness of wild-type cells[1].

Serial sampling of patients across disease progression and transformation, such as from MPN/MDS to AML, allows a look at clonal evolution which drives clinically relevant changes in disease state. Moreover, serial sampling of patients treated with anti-cancer therapies, such as IDH1/2 inhibitors or FLT3 inhibition, can elucidate resistance mechanisms and/or combinations of mutations that allow for escape from therapeutic inhibition. For example, previous single cell DNA sequencing work has identified the emergence of *FLT3/RAS* co-mutant clones that lead to resistance to FLT3 inhibition in *FLT3*-mutant AML patients[2]. Here we identify that not all *RAS*-pathway mutations drive resistance, as we observed suppression of a *U2AF1/STAG2/FLT3/KRAS* mutant clone, while the *U2AF1/KRAS* double mutant clone emerged following therapy. Moreover, we observed emergence of *FLT3*-wild-type, *RAS* mutant cells during FLT3 inhibitor therapy, particularly in the setting of oncogenic *KRAS*. While it is unlikely that genetics alone will determine the efficacy and resistance to targeted therapy, we highlight this as an instance where clone level information provides critical context which is not seen with bulk sequencing. Further study of serial samples combining DNA+Protein single cell sequencing will provide greater insights into the consequences of targeted therapy on clonal representation and evolution, particularly for therapeutic modalities that are thought to function, in part through, enforcing differentiation. The novel insights gained by single cell resolution of AML clonality have great potential impact for developing new therapeutic agents/combinations aimed to target different AML clones.

We used simultaneous cell surface protein expression and DNA genotyping to detect changes in immunophenotype based on specific genotypes, such as somatic *RAS* mutations. This technology has the capability to be expanded to 30+ cell surface proteins that can functionally identify many

populations in normal and malignant hematopoiesis, allowing for granular mapping of specific mutations and mutational combinations across the hematopoietic landscape[3,4]. In addition, by tracking dominant clones with particular immunophenotypes during disease progression and/or during therapeutic intervention, researchers may be able to identify immunophenotypic populations amenable to cell-surface targeting therapies including cellular therapies. These studies will likely require large patient numbers to identify true patterns for targeting, but the potential to affect the AML treatment landscape. Single cell DNA sequencing may have further use and applicability such that clone-specific gene expression, intracellular protein expression, or multi-omic analyses may allow for more detailed studies of clone-specific phenotypes.

Although we focused on SNV and small indels with our custom amplicon panel, the Tapestri platform can be adapted to analyze single cell copy number variations (CNVs), which will be critical for the next set of studies. The technology will be informative for cancer types with frequent complex chromosomal abnormalities, such as breast cancer, renal cell carcinoma, or colorectal cancer[5-8]. We focused on normal karyotype MPN/AML samples, but clonal analysis of *P53*-mutant/complex karyotype AML with integrated single cell mutational/copy number profiling will provide a glimpse into the clone-specific biology of this adverse risk AML subset and empower comparisons between normal karyotype and complex karyotype AML at a clonal level. These studies, in particular, will be strengthened by simultaneous gene expression or immunophenotypic analysis with DNA genotyping/CNV analysis at single cell resolution.

The clinical applicability of single cell DNA sequencing still requires further studies including detailed studies of homogeneous clinical trial cohorts, and studies of large panels of serial samples

from patients with changes in disease state (MPN/MDS->AML) or pre/post anti-cancer therapies. However, single cell profiling as presented here is impactful for many types of biological inquiries. Assessments of clonality and the identification of immunophenotypic changes based on mutation combinations will not just be insightful for hematologic malignancies but also for solid tumors. Studies could identify mutational co-occurrences and dominant clones that have not been previously delineated from bulk sequencing, similar to our study described here. This knowledge in solid tumors and other heme malignancies could allow for creation of better disease models, discovery of disease initiating cells, and/or targetable cell populations based on clone-specific mutations, surface marker expression and/or gene expression.

Single cell DNA sequencing, while highly valuable, is not without limitations. The low sequencing yield from cell input has the potential to miss rare clones and variants. Depending on the biological or clinical question, loss of rare clones/variants can confound study results. While we removed low cell yield samples (<100 cells) from our analysis and/or repeated low cell yield samples by sequencing another aliquot from the same time point, this obstacle can also be prevented by performing studies with high cell number samples to ensure optimal cell loading and encapsulation. Additionally, allelic dropout and potential doublets could lead to further confounding results, which is discussed further in the main text[9]. Inclusion of these estimates as transition probabilities in the Markov model will provide greater resolution in clonal trajectories. However, discerning these technical artifacts from copy number alterations or loss of heterozygosity requires expanded panel design and incorporation of unique molecular identifiers in library preparation. Phylogenetic modeling approaches incorporating these uncertainties have been developed, and likely have application here even with reduced genome coverage[10-13]. Despite

these important limitations, we posit that the use of single cell genotypic analysis provides a view into AML pathogenesis not possible with bulk genomic analysis, and provides new insights into the pathogenesis of myeloid transformation and leukemia evolution which can be investigated using models which assess AML as a multi-clonal, evolving disease state.

References

1 Schepers, K. *et al.* Myeloproliferative neoplasia remodels the endosteal bone marrow niche into a self-reinforcing leukemic niche. *Cell Stem Cell* **13**, 285-299, doi:10.1016/j.stem.2013.06.009 (2013).

2 McMahon, C. M. *et al.* Clonal Selection with RAS Pathway Activation Mediates Secondary Clinical Resistance to Selective FLT3 Inhibition in Acute Myeloid Leukemia. *Cancer Discov* **9**, 1050-1063, doi:10.1158/2159-8290.CD-18-1453 (2019).

3 Notta, F. *et al.* Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* **351**, aab2116, doi:10.1126/science.aab2116 (2016).

4 Majeti, R., Park, C. Y. & Weissman, I. L. Identification of a hierarchy of multipotent hematopoietic progenitors in human cord blood. *Cell Stem Cell* **1**, 635-645, doi:10.1016/j.stem.2007.10.001 (2007).

5 Frohling, S. & Dohner, H. Chromosomal abnormalities in cancer. *N Engl J Med* **359**, 722-734, doi:10.1056/NEJMra0803109 (2008).

6 Albertson, D. G., Collins, C., McCormick, F. & Gray, J. W. Chromosome aberrations in solid tumors. *Nat Genet* **34**, 369-376, doi:10.1038/ng1215 (2003).

7 Cancer Genome Atlas, N. Comprehensive molecular characterization of human colon and rectal cancer. *Nature* **487**, 330-337, doi:10.1038/nature11252 (2012).

8 Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61-70, doi:10.1038/nature11412 (2012).

9 Pellegrino, M. *et al.* High-throughput single-cell DNA sequencing of acute myeloid leukemia tumors with droplet microfluidics. *Genome Res* **28**, 1345-1352, doi:10.1101/gr.232272.117 (2018).

10 Malikic, S., Jahn, K., Kuipers, J., Sahinalp, S. C. & Beerenwinkel, N. Integrative inference of subclonal tumour evolution from single-cell and bulk sequencing data. *Nat Commun* **10**, 2750, doi:10.1038/s41467-019-10737-5 (2019).

11 Zafar, H., Tzen, A., Navin, N., Chen, K. & Nakhleh, L. SiFit: inferring tumor trees from single-cell sequencing data under finite-sites models. *Genome Biol* **18**, 178, doi:10.1186/s13059-017-1311-2 (2017).

12 Ross, E. M. & Markowetz, F. OncoNEM: inferring tumor evolution from single-cell sequencing data. *Genome Biol* **17**, 69, doi:10.1186/s13059-016-0929-9 (2016).

13 Jahn, K., Kuipers, J. & Beerenwinkel, N. Tree inference for single-cell data. *Genome Biol* **17**, 86, doi:10.1186/s13059-016-0936-x (2016).