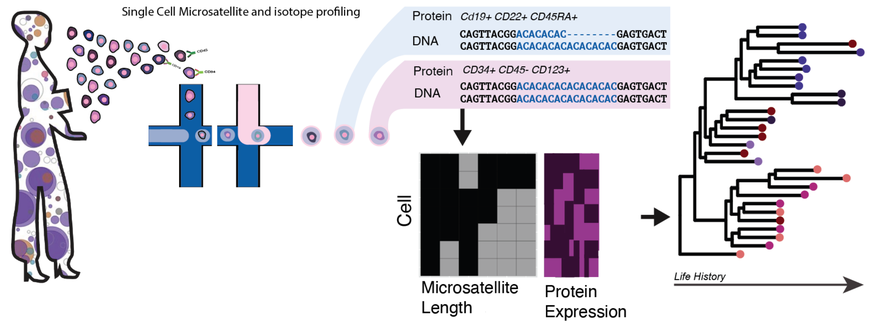
Retrospective lineage tracing and phenotypic profiling in human tissues by droplet single cell microsatellite sequencing

Nathaniel D. Omans[1,2,3], Tamara Prieto[2,3], John Zinno[2,3], Joshua Schiffman[2,3], Jake Qiu[2,3], Shu Wang[4], Lucy A. Godley[5], Dan A. Landau[2,3]

Introduction



Motivation

- Lineage tracing is a powerful approach to study developmental and somatic evolution processes, especially when combined with simultaneous phenotypic profiling.
- Existing methods rely on prospective tracking of synthetic DNA barcodes, precluding use in primary human samples.
- Phylogenetic modeling of somatic mutations offers the ability to retrospectively infer lineages in humans and reveal historic population dynamics, including timing of driver events, population sizes, and growth rates.
- Microsatellites are a valuable marker of lineage histories due to their high mutation rate

Impact

- We developed a single cell DNA sequencing method for the retrospective lineage of thousands of cells. Our method takes advantage of the high somatic mutation rate of microsatellite regions. We can accurately recapitulate the lineage histories of human tissues from a sequencing panel of 1100 microsatellites.
- We extend this method to detect cell states by using DNA-conjugated antibodies for 45 cell surface proteins that define hematopoietic lineages, enabling the identification of heritable cell states across somatic lineages

Findings

- Microsatellite phylogenies recapitulate the ground truth evolutionary history of an *in vitro* model of clonal evolution.
- Phylogenetic branch lengths of clones within the tree accurately reflect the experimental timing of clonal divergences and expansions.
- The microsatellite phylogeny constructed from a patient with a non-cancerous hematopoietic *DNMT3A* mutation accurately reflects the divergence of the mutated stem cell clone. The information density of the microsatellite tree grants inferential power beyond *DNMT3A* genotype alone.
- Cells within the *DNMT3A* mutated clade shows a robust myeloid skew and are associated with the strong heritability of CD71, CD141, and CD49d expression.

Methods

Microsatellite Panel and sequencing:
We developed a targeted DNA sequencing panel of 1121 loci covering 1100 highly mutable microsatellites and 12 known hematopoietic driver mutations. The microsatellite panel was designed for the Mission Bio Tapestri platform, which uses microfluidic cell isolation and targeted DNA amplicon sequencing of about 5000 cells.

Microsatellite genotyping and Tree reconstruction:
Sequencing reads were processed using a custom pipeline (github.com/omansn/flankton). Microsatellites were genotyped by assigning reads to loci using the unique pair of flanking regions of each microsatellite and to cells by a cell barcode. For each cell and locus, microsatellite repeat lengths were detected using Phobos and genotypes were assigned based on the two most frequently observed repeat lengths in each cell and locus. Trees were constructed using Neighbor Joining with distances as the sum of differences in microsatellite lengths for each locus between pairs of cells.

In vitro model of clonal evolution:
A parental population of the microsatellite unstable DLD1 cell line was grown for 30 days and split for cryopreservation. Remaining cells were isolated and single cells were plated for clonal expansion. A series of expansions followed by single cell cloning and cryopreservation was repeated 4 times over a total 115 days, resulting in a clonal structure of known pedigree (fig 1.A.). 8 clones were pooled and processed on the Mission Bio Tapestri. Clone-specific microsatellite genotypes were detected with WGS of individual clones and annotated on the cell tree for clonal identification (fig 1.B).

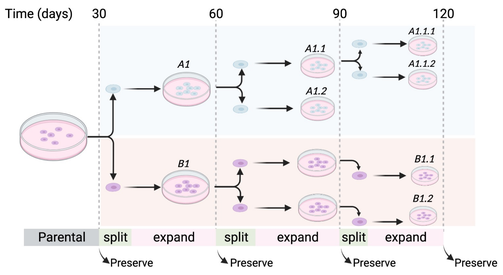
Clonal Hematopoiesis Patient Sample:
A bone marrow aspiration from a patient with DNMT3A mutated clonal hematopoiesis of Indeterminant Potential was collected under the supervision of Dr. Lucy Godley. Fresh aspirate was sorted for CD34+/CD34- hematopoietic populations. 50% CD34+ and 50% CD34- cells were pooled and stained with the totalseq-D heme cell surface marker panel (BioLegend) and immediately processed on the Mission Bio Tapestri.

Defining Heritability with PATH: *(for more information see Josh Schiffman's Poster (244) in Session III)*
Heritability of individual cell surface markers and DNMT3A mutation status were calculated as the cross correlation of these features on the phylogenetic tree structure (fig 2.D). Similarly, cell surface markers were combined into lineage modules based on canonical cell markers. We used Seurat's AddGeneModuleScore function to calculate a score for Lymphoid, Myeloid, and Stem modules in each cell (fig 2.E).

Microsatellite Phylogenies recapitulate somatic evolution *in vitro*

Experimental Model

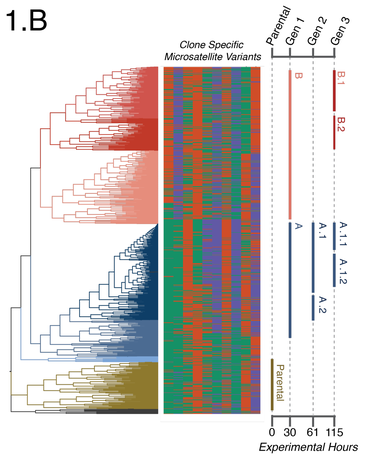
1.A



1.A. A schematic of the in vitro experiment to produce a ground truth clonal tree.

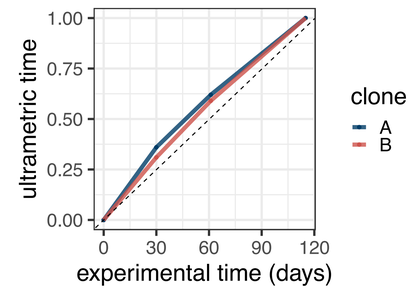
Phylogenies recapitulate ground truth evolutionary relationships and timing of clonal divergences

1.B



1.B. A microsatellite tree of an *in vitro* model of clonal evolution. A heatmap of known clone-specific microsatellite genotypes reflects phylogenetic accuracy. Clone identities inferred from known markers are annotated on the right.

1.C

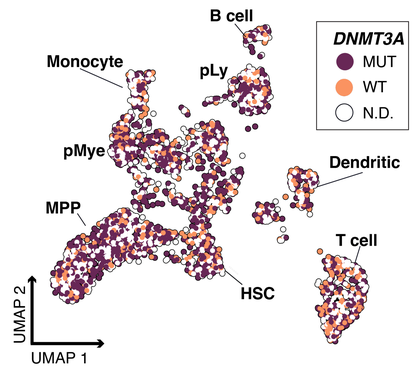


1.C. Phylogenetic branching times of individual clones on the phylogenetic tree are proportional to the ground truth experimental branching times

Lineage tracing of human blood with a clonal *DNMT3A* mutation

Simultaneous profiling of cell type and genotype

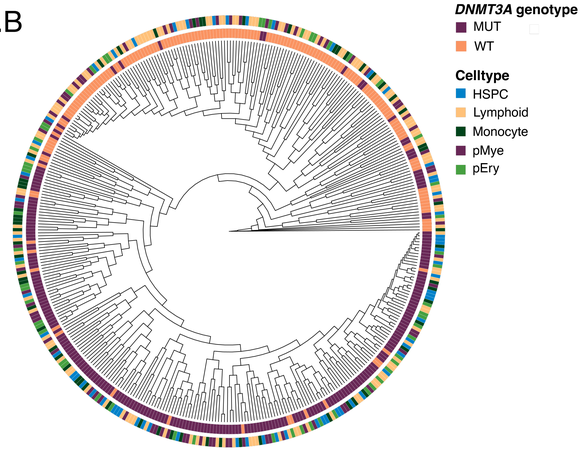
2.A



2.A. A UMAP of 45 cell surface markers. Cell identities are inferred by DNA conjugated antibodies of 45 lineage-specific markers and co-sequenced with the microsatellite panel. *DNMT3A* mutations are observed across hematopoietic lineages.

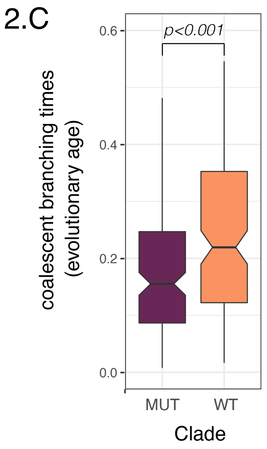
Microsatellite Lineage Tree recapitulates early divergence of DNMT3A mutated cells

2.B



2.B. The microsatellite phylogenetic tree recapitulates the evolutionary divergence of *DNMT3A* mutated and wildtype cells. Cells on the tree are annotated with cell identities and *DNMT3A* genotype

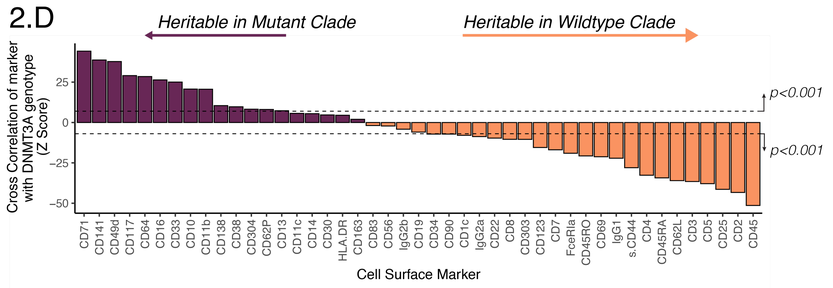
2.C



2.C. The *DNMT3A* mutated clade has shorter branch lengths reflecting its more recent evolution compared to unmutated cells

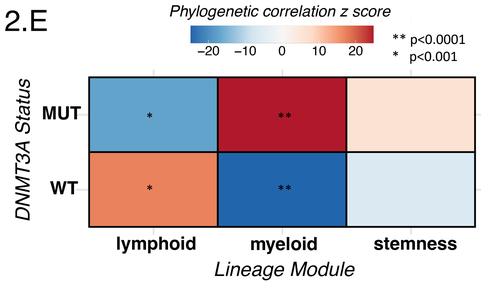
Myeloid phenotypes are evolutionarily heritable and associated with *DNMT3A* mutants

2.D



2.D. Z scores of the cross correlation between *DNMT3A* genotype and individual cell surface markers shows heritability of marker expression within mutated and unmutated clades.

2.E



2.E. Myeloid cell states are heritable within the DNMT3A mutated clade. Stem cell states are not heritable or associated with *DNMT3A* genotype

Affiliations
[1] Tri-Institutional PhD Program in Computational Biology and Medicine, New York, NY, USA.
[2] The Meyer Cancer Center, Weill Cornell Medicine, New York, NY, USA
[3] The New York Genome Center, New York, NY, USA
[4] Mission Bio, Inc. South San Francisco, CA, USA.
[5] Departments of Medicine and Human Genetics, The University of Chicago, Chicago, IL, USA

Training Funds
NYSTEM
T32

First Author Affiliations
Weill Cornell Medicine
NEW YORK GENOME CENTER