Fluctuating methylation clocks and mutational frequencies lead to patient-specific inference of CHIP subclone dynamics

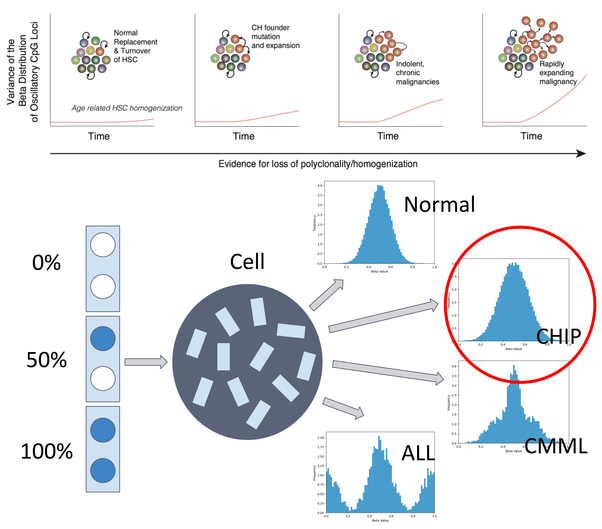
Prateek Anand[1], Ryan Schenck[2], and Christina Curtis[2]

[1]University of California, Los Angeles, Los Angeles, CA; [2]Stanford University, Stanford, CA

Background

Clonal hematopoiesis of indeterminate potential (CHIP) is a condition characterized by the presence of genetically distinct populations of blood cells in the bone marrow and peripheral blood. While CHIP is not considered a malignancy, it is a known pre-cancerous state that has been associated with an increased risk of progression to a hematological malignancy. Individuals with CHIP are also at risk of other adverse outcomes, such as cardiovascular events and developing epithelial cancers. Due to CHIP's widespread prevalence within aging individuals, it is paramount to stratify patients that may progress towards cancer.

Mutational data alone is insufficient to imply cancer or progression status. Although CHIP is associated with somatic driver mutations in HSCs that are shared with certain myeloid malignancies, patients with CHIP do not exhibit any signs of malignancies. Further, very few cases develop into malignancies. We need additional data to better address CHIP at early stages.

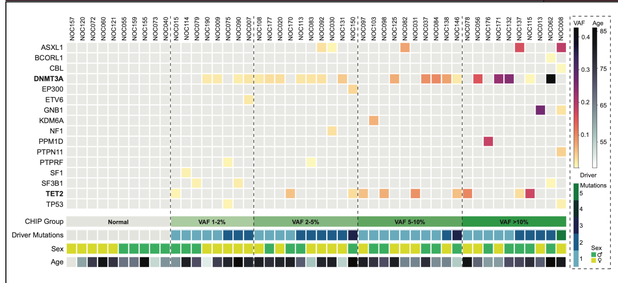


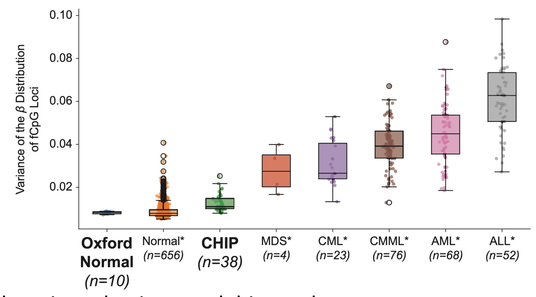
Fluctuating methylation clocks[1] (FMCs) contain more information about recent clonal dynamics, clonal conversions, and cell turnover. FMCs refer to certain CpG loci that exhibit oscillating methylation patterns. They can be used to detect stem cell numbers and their replacement rates. **Top:** A healthy HSC crypt acquires a CH driver mutation. As homogenization progresses, fCpG loci become more concentrated around methylation states of 0%, 50%, and 100%, leading to a more "W-shaped" methylation distribution (**Bottom**). This project builds off the previous FMC diagnostic tool we developed for CH.

Aims: Using both sources of data

- characterize discovery cohort
- build/refine inference model to obtain induction time and expansion rate for samples
- test on validation cohort and compare with previous findings[2] that use only mutational data

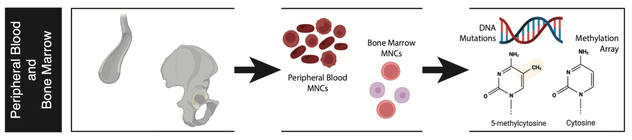
Results





Left: Characteristics of the discovery cohort. Patients undergoing elective total hip replacement surgery were diagnosed as either normal or CH based on the VAF of putative driver mutations. **Right:** When mutations leading to CH occur, the variance of the fCpG methylation distribution increases. As expected, CHIP variance falls close to normal. Since CHIP is potentially the earliest stage of progression, we want to infer the expansion rate of the subclone as soon as it is detected and get an accurate risk assessment.

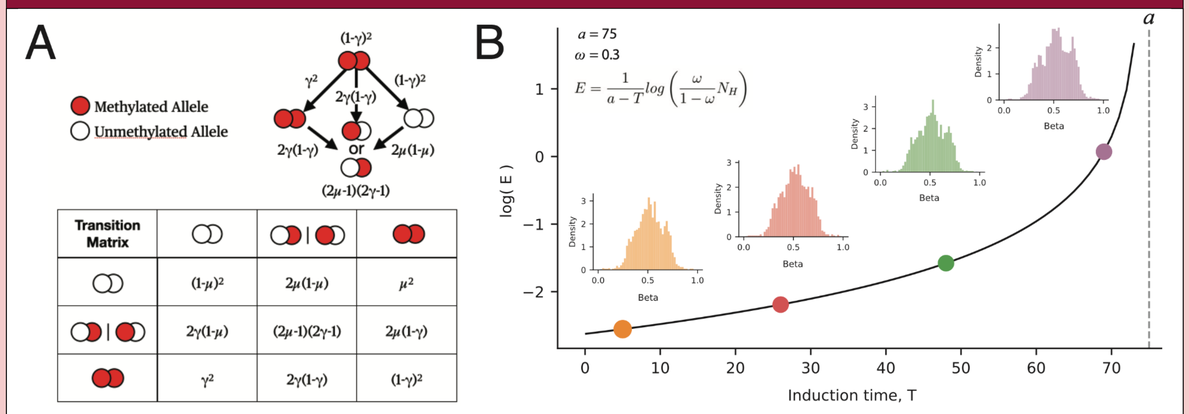
Data



Targeted DNA sequencing was performed on bone marrow mononuclear cell and peripheral blood granulocyte DNA samples. Samples were annotated as having CH based on the presence of driver mutations in bone marrow sequencing (in genes DNMT3A, TET2 , etc.).

DNA methylation analysis was performed on samples of peripheral blood granulocyte DNA, selecting samples with CH variants in a range of VAFs and controls without CH.

Model

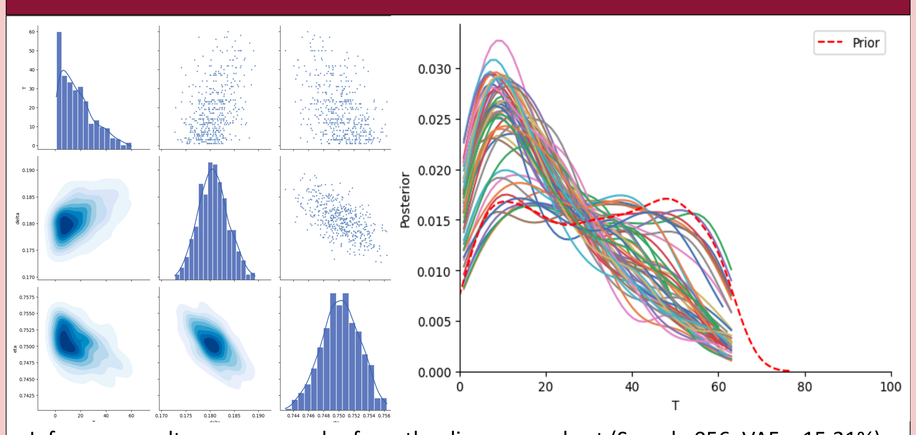


The model used in the inference framework. **A:** The transition state matrix showing the Markov Process for the transitions between methylation status of a loci with two alleles. μ is the methylation rate and γ is the demethylation rate. **B:** Relating expansion rate to induction time of the driver mutation. Using the fCpG methylation distribution allows us to visualize this curve: indolent subclones arise earlier and grow slower, while aggressive subclones arise later and grow faster. This finding motivates the possibility of inferring T and E.

Cohorts

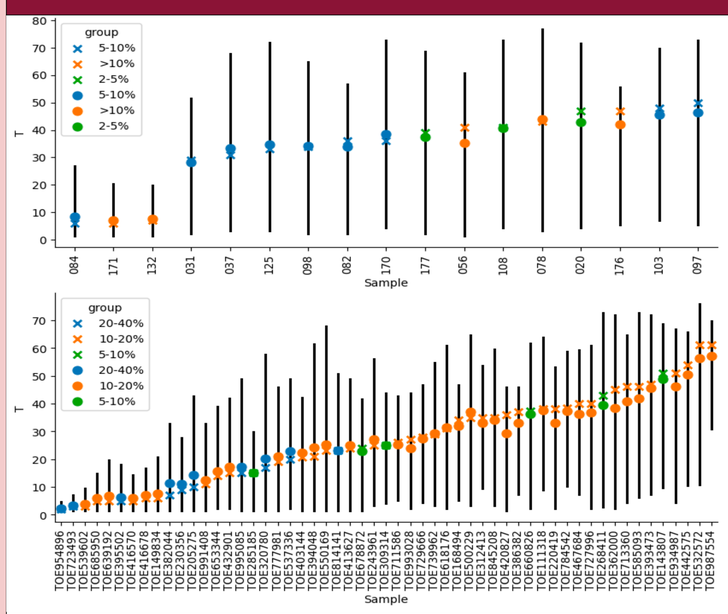
Cohort	Full Name	Samples	# CHIP Samples	Female/ Male
MARCH (Discovery)	Mechanisms of Age-Related Clonal Hematopoiesis	48	38	26/22
MESA (Validation)	Multi-Ethnic Study of Atherosclerosis	936	31	495/441
JHS (Validation)	Jackson Heart Study	1732	68	1081/651
WHI (Validation)	Women's Heart Study	1428	103	394/1034
FHS (Validation)	Framingham Heart Study	1815	111	567/1248

Inference



Inference results on an exemplar from the discovery cohort (Sample 056, VAF = 15.31%). **Left:** Pairplot displaying parameter densities. T is induction time. Delta and Eta are corrective parameters used in the linear transformation needed for data comparison to the original model. **Right:** Iterations of the inference on T using an uninformative prior $T \sim U(1, \text{Age} - 5)$, converging to a result between 0 - 20. This is biologically expected.

Results



Top: Confidence intervals of induction times for samples in the discovery cohort. "x" is the median and "o" is the mean. **Bottom:** Same plot for the validation cohort. 55 CHIP samples remained after passing through VAF filtering and subsetting for only one driver mutation.

Methods

Utilizing PyABC (approximate bayesian computation) we infer induction time given the patient's age, VAF driver frequency, and fCpG methylation distribution.

Inference Parameters

Max Iterations	40
Error Threshold	0.000001
Efficiency Threshold	0.001
Simulations/Iteration	500
Initial Population	500
fCpG Loci	2000

Conclusions

- Clonal hematopoiesis is heterogeneous and can be reliably detected using FMC's regardless of the driver mutation
- Given the nature of tumorigenesis, it is difficult to deconvolve the effects of multiple driver mutations/nested subclonal populations
  - The model can be adjusted to account for additional subclones arising from an initial subclone
- Inferences are reliable for samples above the clinically significant VAF threshold of 5%
  - Samples that fall under 5% often have uninformative inferences or are biased by the prior

Future Directions

One issue that we noticed with the inference procedure is that the acceptance rate threshold is often reached before the epsilon threshold. This suggests that the method is sampling inefficiently and therefore needs exponentially more simulations to accept the required amount. There are two possible avenues to consider:

- **Use PyABC's adaptive distance function.** Rather than staying fixed on RMSE (mean + variance), this feature will adjust the distance metric throughout the iterations.
- **Change the model to infer methylation and demethylation rates (μ/γ) separately for each sample.** This could potentially help the inferences because it is likely that μ and γ vary slightly between patients, providing more individualized accuracy.

In the near future, we hope to obtain tighter posterior distributions when resolving these parameters. Finally, we would also like to compare these results with published findings that only utilize mutational data.

Literature Cited

1. Gabbutt, C., Schenck, R.O., Weisenberger, D.J. *et al.* Fluctuating methylation clocks for cell lineage tracing at high temporal resolution in human tissues. *Nat Biotechnol* 40, 720–730 (2022). https://doi.org/10.1038/s41587-021-01109-w
2. Weinstock, J.S., Gopakumar, J., Burugula, B.B. *et al.* Aberrant activation of TCL1A promotes stem cell expansion in clonal haematopoiesis. *Nature* 616, 755–763 (2023). https://doi.org/10.1038/s41586-023-05806-1

Acknowledgments

I would like to thank my mentor, Ryan, for the constant guidance throughout this project. His effort helped me quickly learn the project background and overcome any obstacles we encountered. I also greatly appreciate the Curtis Lab for providing me a welcoming and intellectual environment to collaborate with.

Program supported by the National Cancer Institute