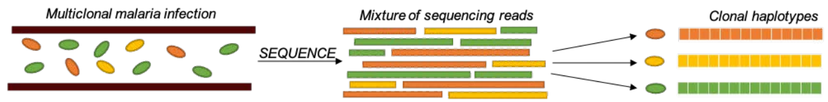
Deconvolving multiclonal malaria infections using Oxford Nanopore long read sequencing

Somya Mehra[1], Zahra Razook[1,2], Alyssa Barry[1,2]

1. Life Sciences Discipline, Burnet Institute, Melbourne
2. Institute for Mental and Physical Health and Clinical Translation (IMPACT), School of Medicine, Deakin University, Geelong

Multiclonal Malaria Infections

- **Malaria** is a deadly tropical disease, causing 405,000 deaths in 2018 alone[1]
- **Multiclonal infections,** which involve the co-circulation of multiple genetically distinct parasite clones, are common, particularly in high transmission settings[2]
- **Deconvolving** clonal haplotypes remains an analytic challenge, with implications for estimating drug and treatment efficacy
- **Long read sequencing** can aid haplotype reconstruction



A Simple Haplotype Clustering Method

1. Focus on highly heterozygous regions of parasite genome

- Align reads against parasite genome and extract high-quality primary alignments, treating each read as a separate haplotype

2. Read editing to remove sequencing artefacts and retain informative variation

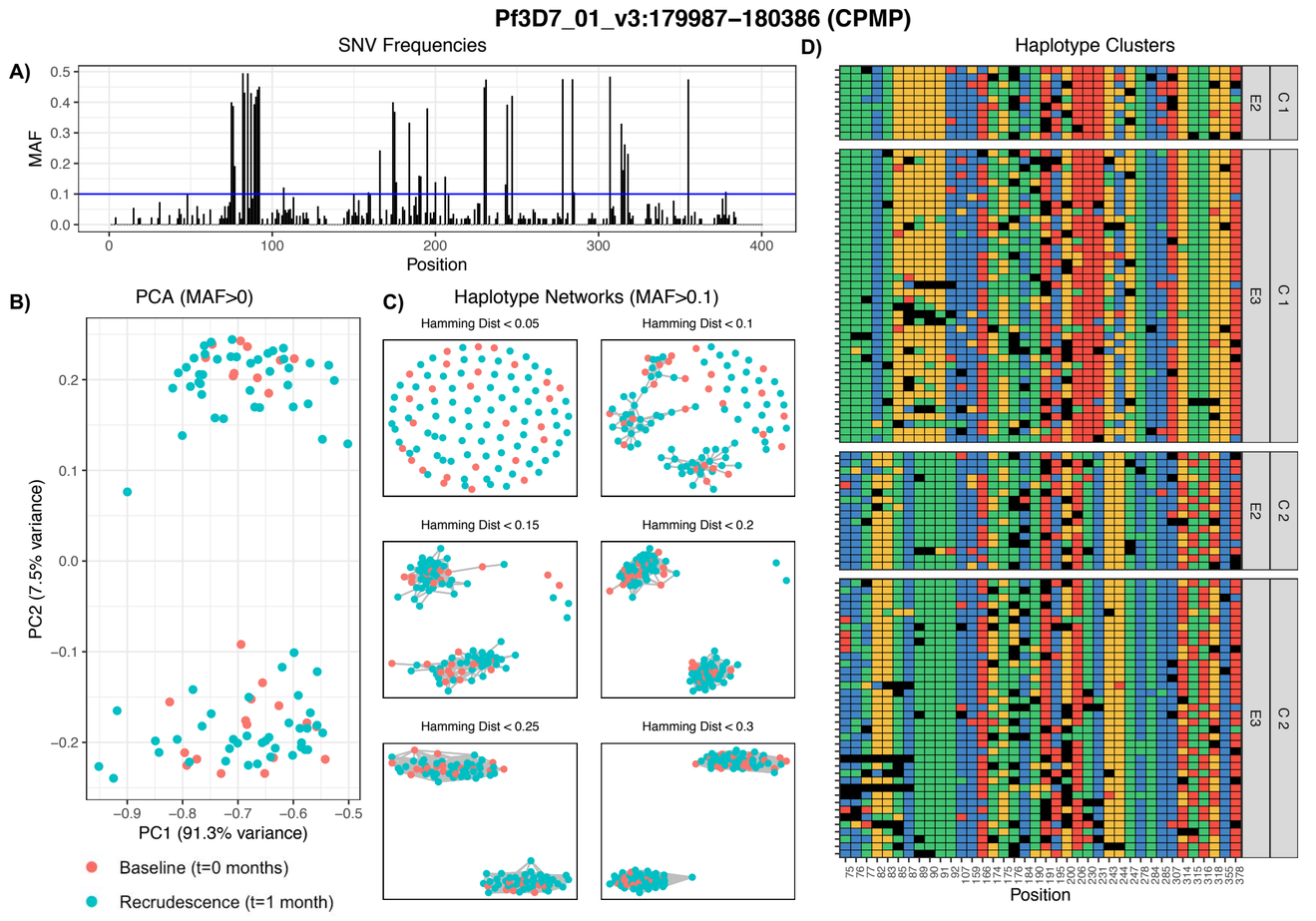
- Focus on single nucleotide variation (SNVs) since Oxford Nanopore long reads exhibit extensive indel error[3]
- Iterate through CIGAR strings encoding alignment information
  - Remove insertions, mark deletions and soft-clipped bases with Ns
  - Remove reads with indels >10bp or >25% loci marked with Ns
- Reduce background noise – focus on informative SNVs (MAF>0.1)

3. Network analysis to assign haplotype clusters

- Calculate pairwise Hamming distances between haplotypes using informative SNVs
- Construct networks based on adjacency matrices with a range of cutoffs
- Examine component structure in networks across a range of thresholds to characterise variation within and between haplotype clusters

Case Study: Treatment Failure in a Returning Traveler

Genomic profiling confirms a **recrudescence**, resulting from the expansion of the original parasite population within the patient at baseline.



Definition of distinct clones within a clinical case of P. falciparum malaria using long read sequencing data. *Haplotypes (polished and filtered individual reads) were extracted for cpmp gene[4] (PF3D7_0104100, region Pf3D7_01_v3:179987−180386) and PCA and network analysis done to identify unique clones. A) SNV frequencies across the region of interest B) Principal components analysis based on pairwise Hamming distances calculated using all polymorphic loci C) Haplotype networks based on informative loci (MAF>0.1), showing connections between haplotypes with Hamming distances below the specified threshold D) Assignment of haplotypes into clonal clusters, based on component structure in the haplotype networks.*

References

1. WHO (2017). *World Malaria Report,*
2. Touray et al. (2020). *Scientific Reports* 10(1):1-8
3. Tyler et al (2017). *Scientific Reports* 8(1):1-12
4. Lerch et al. (2017). *BMC Genomics* 18(1):864