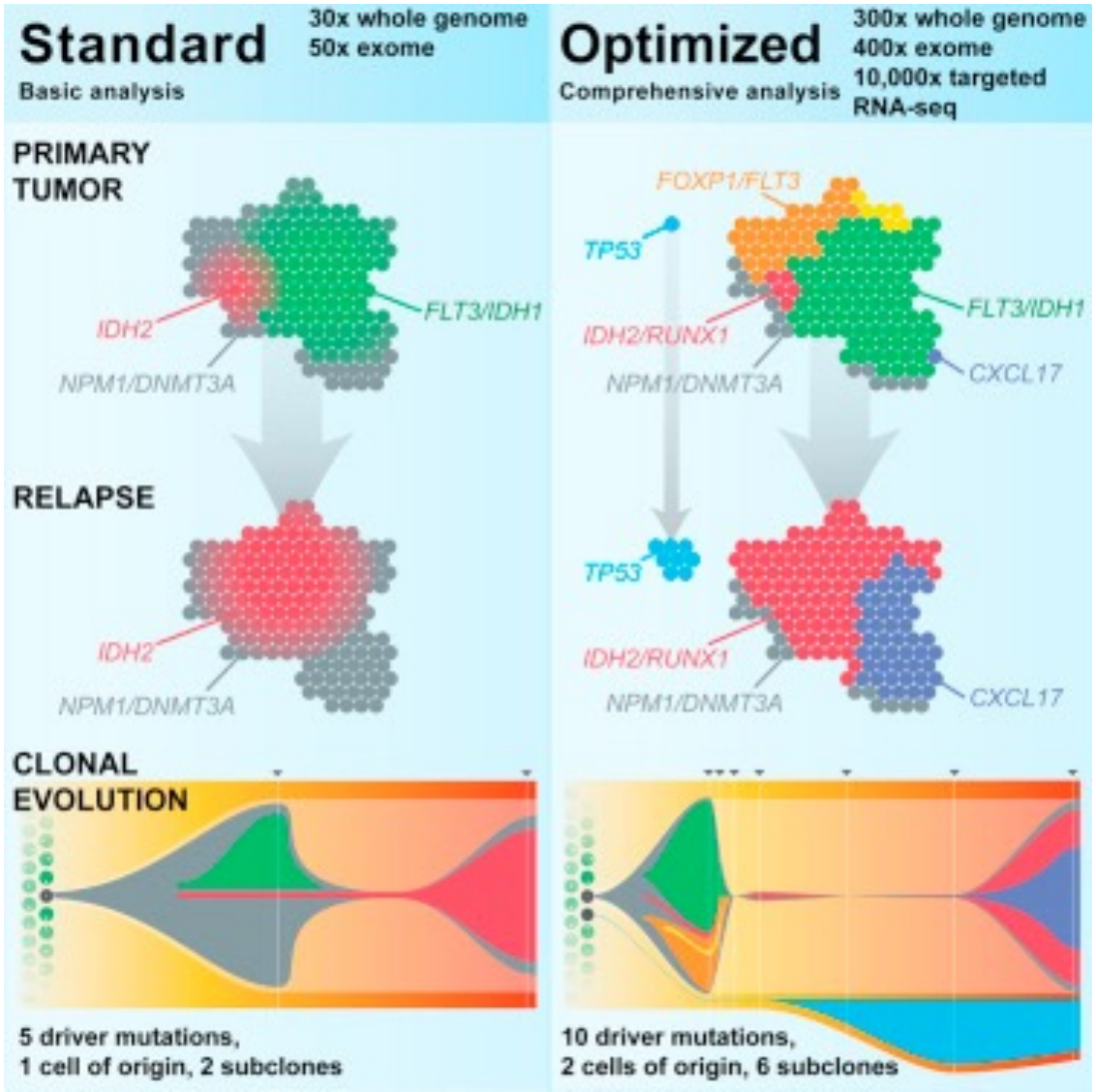
Error-corrected sequencing

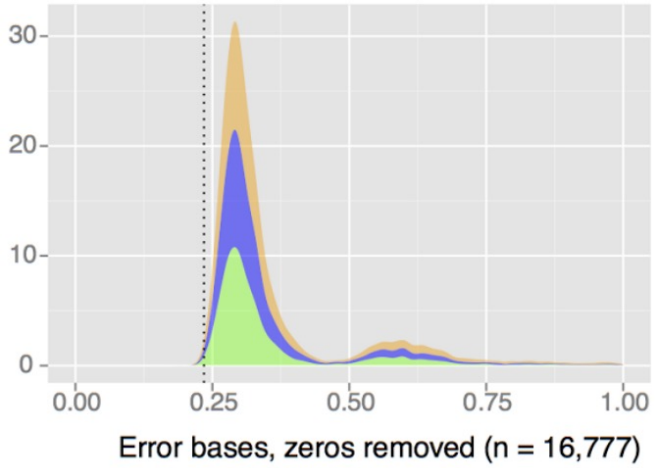
Chris Miller, PhD
c.a.miller@wustl.edu

Sequencing data is noisy



Griffith and Miller et al. Cell Syst. doi: 10.1016/j.cels.2015.08.015

Sequencing data is noisy

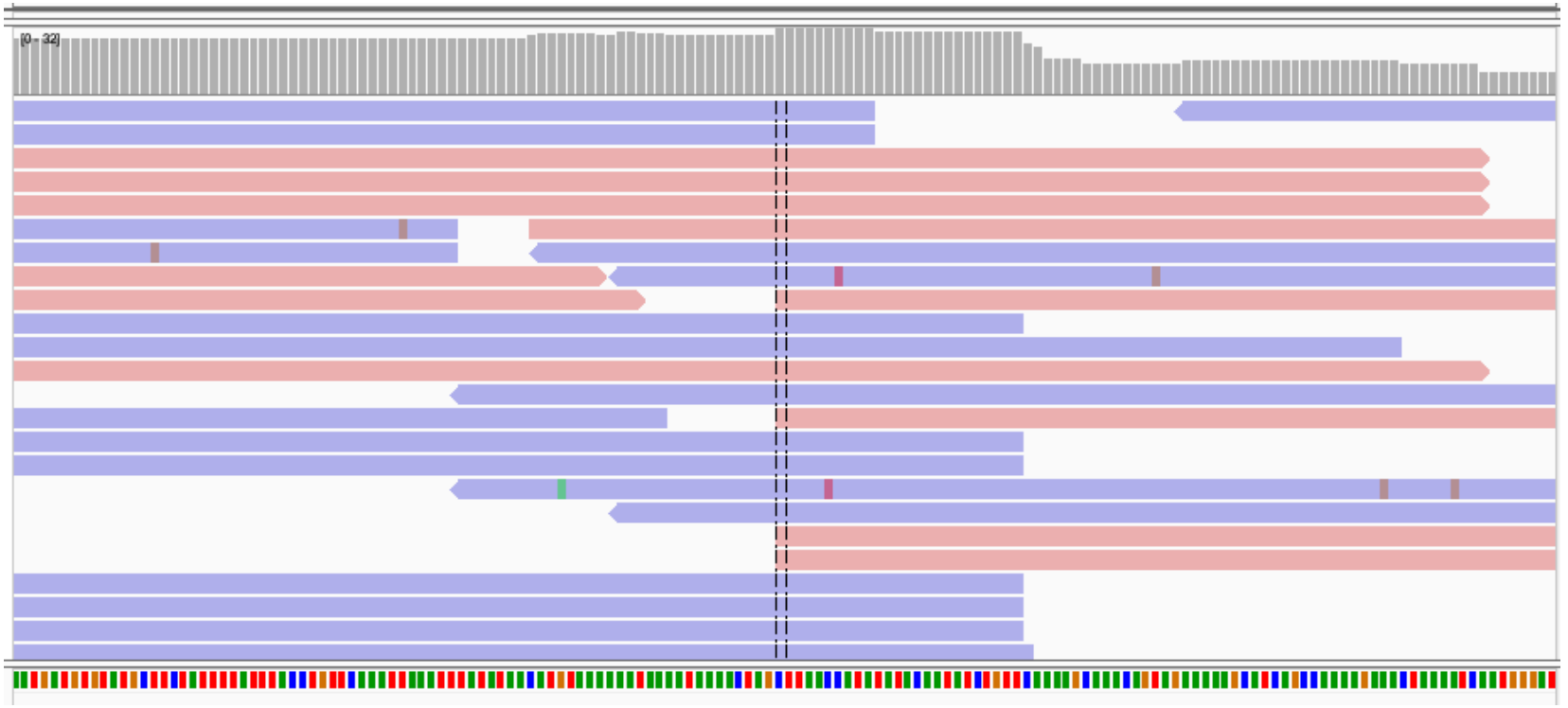


Griffith and Miller et al. Cell Syst. doi: 10.1016/j.cels.2015.08.015

Sources of error

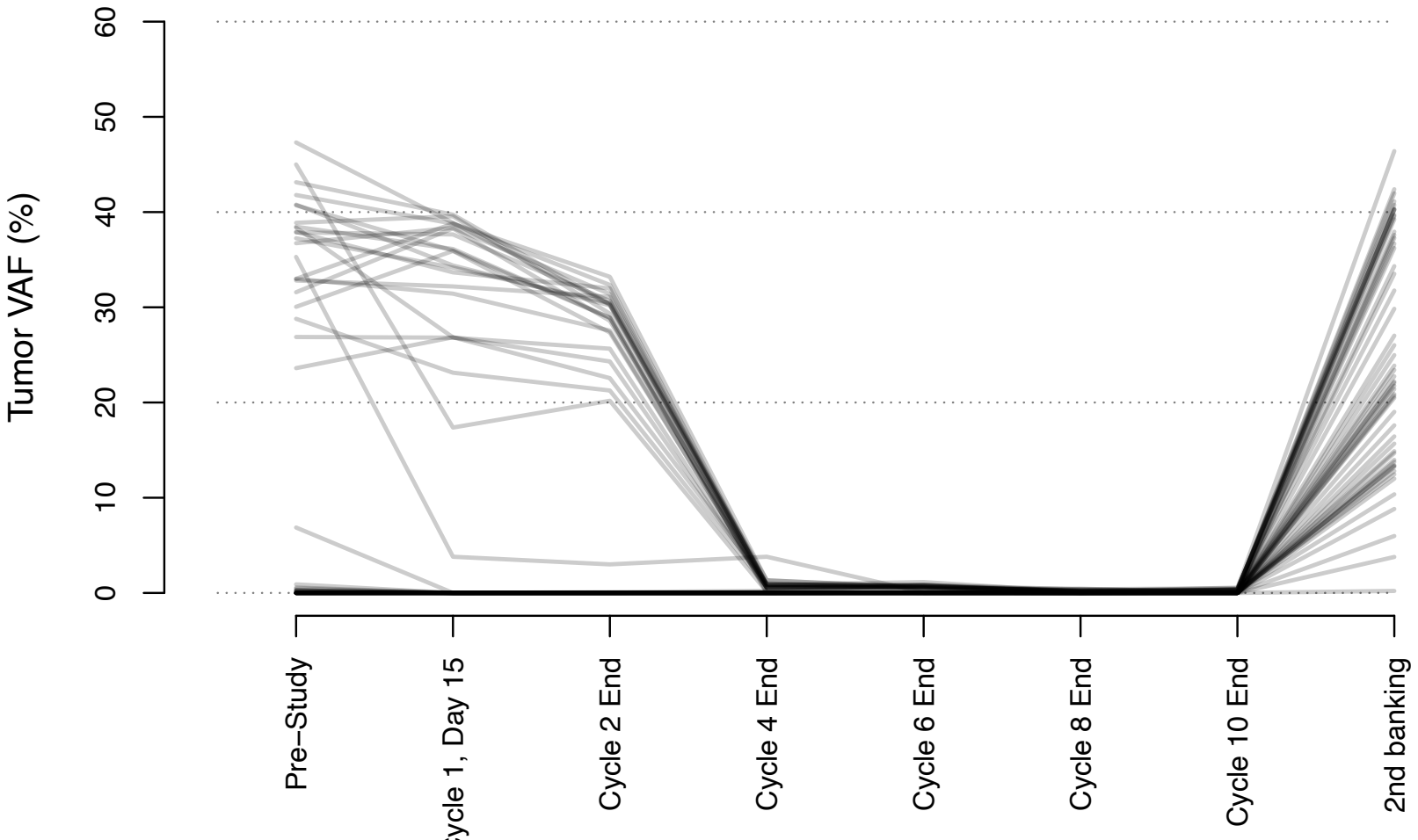
- DNA damage during handling/prep (FFPE)
- PCR Amplification steps
- Base calling on the instrument

Sequencing data is noisy

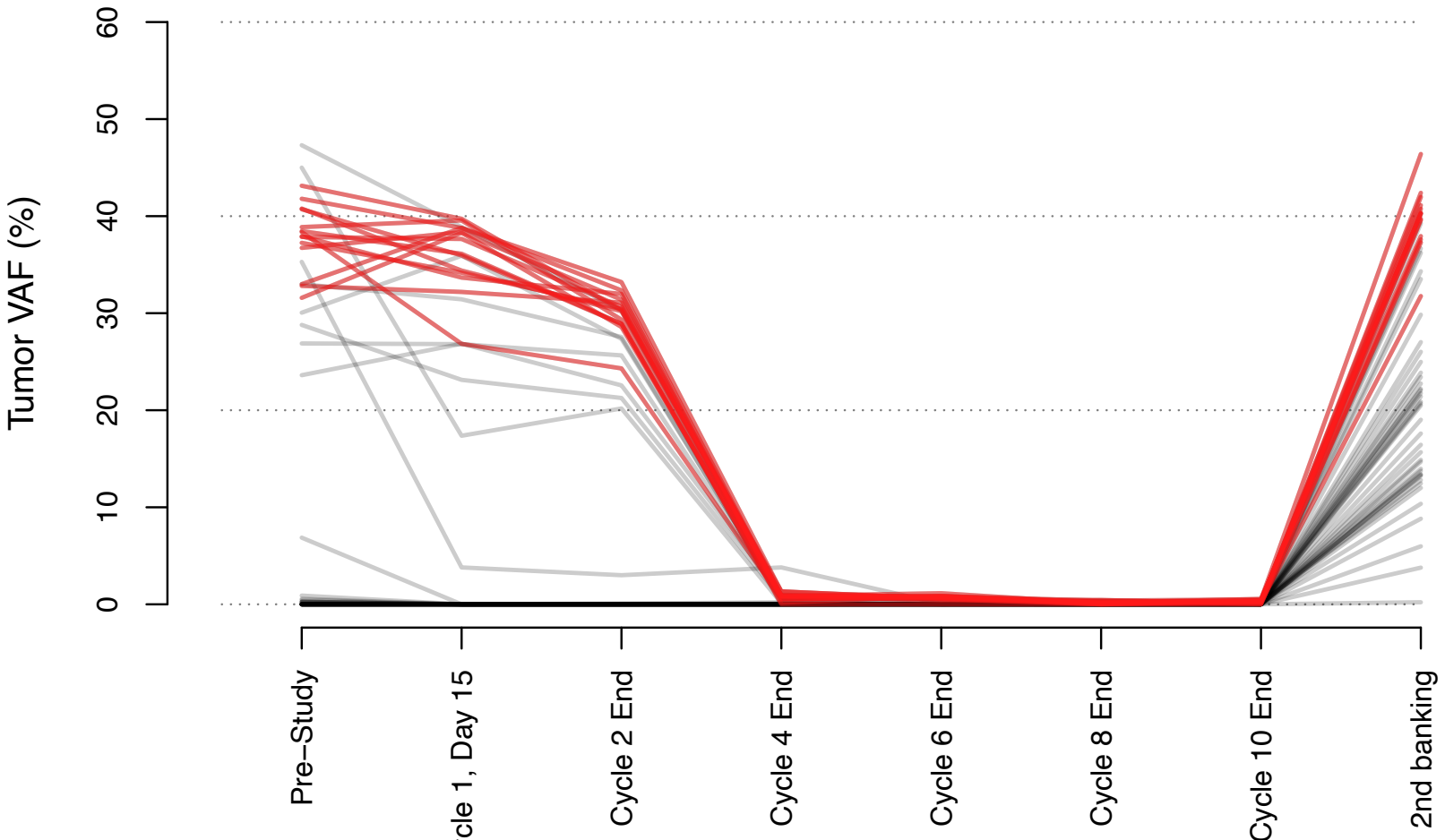


Errors are a fact of life

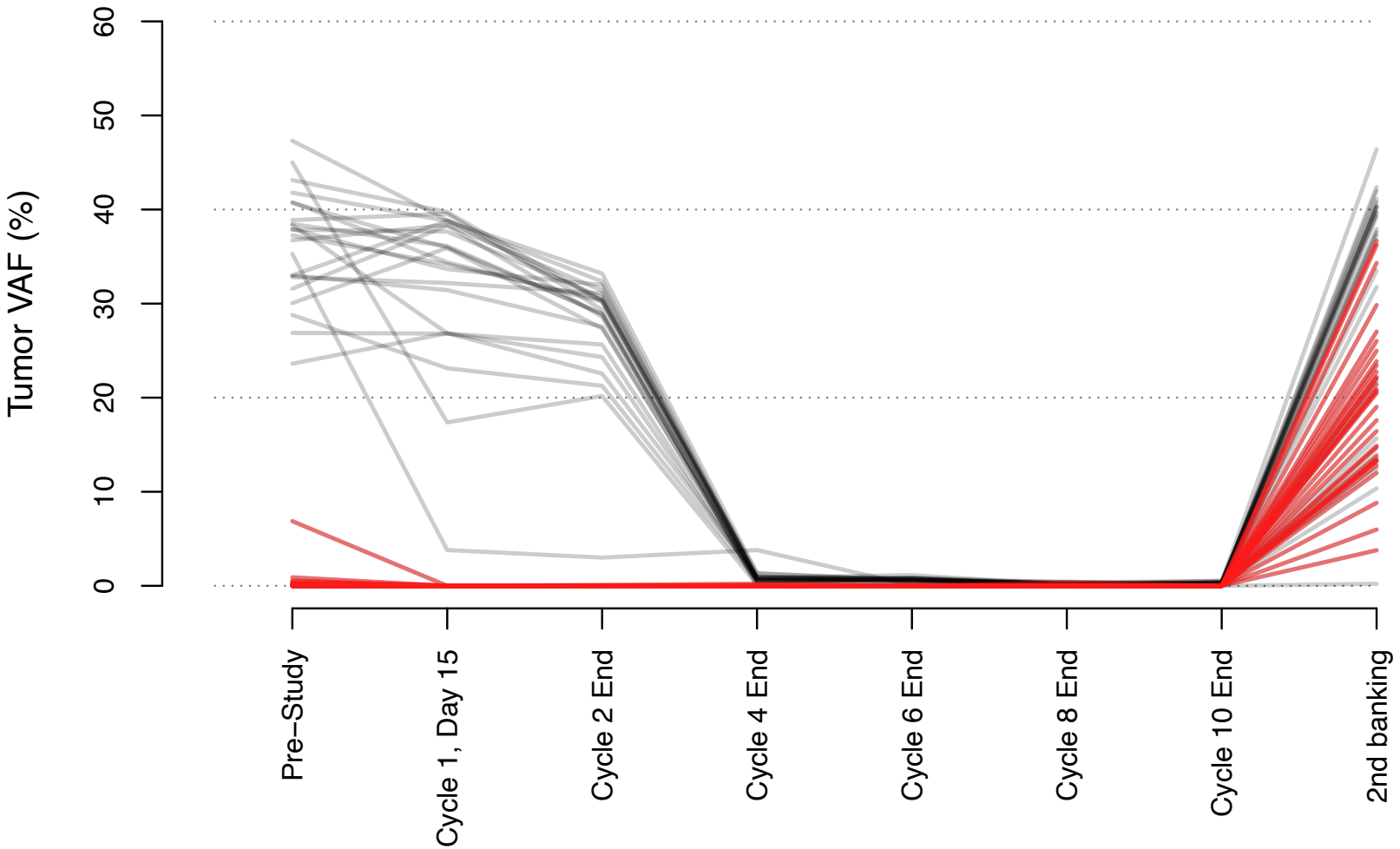
MDS response to decitabine



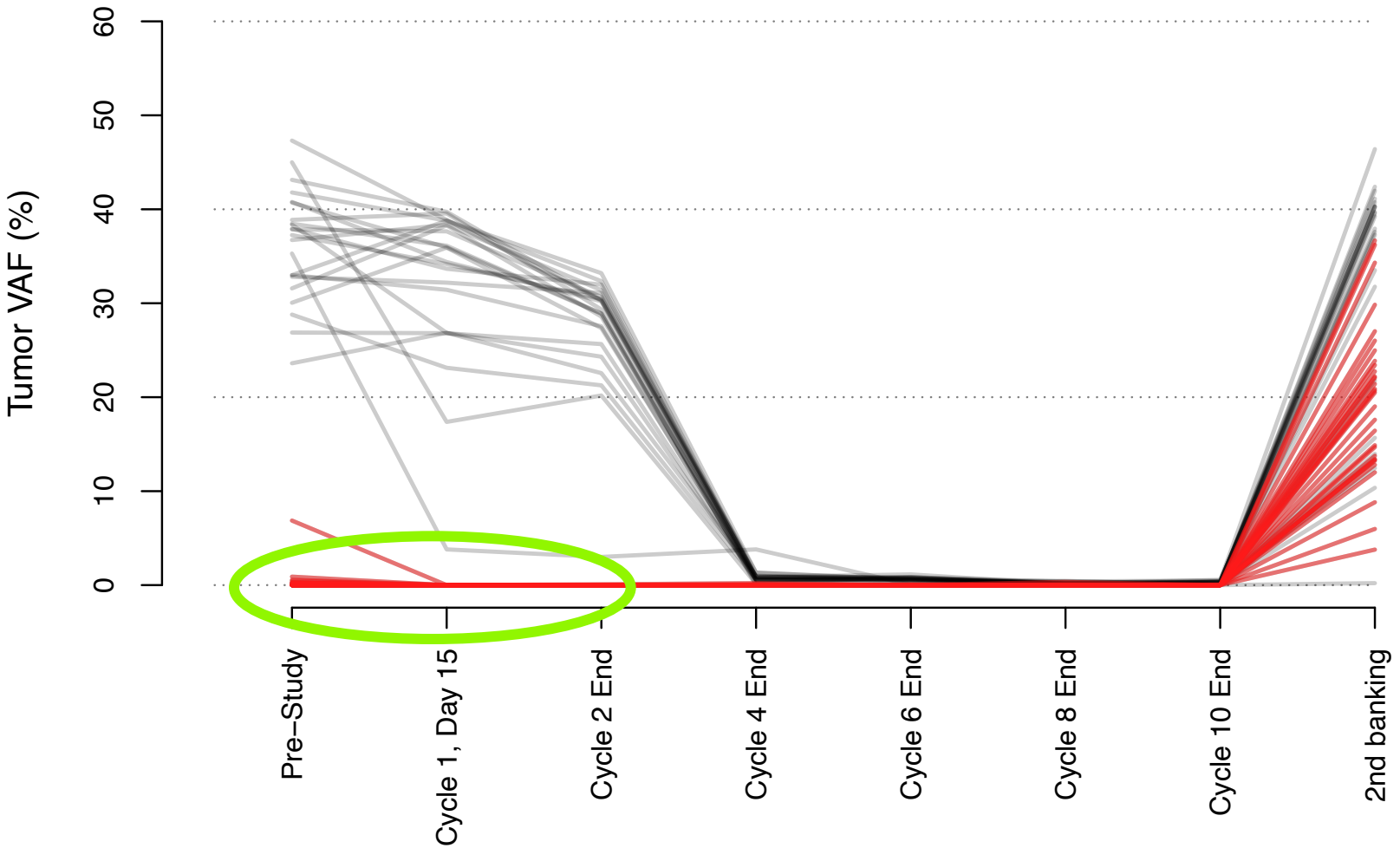
MDS response to decitabine



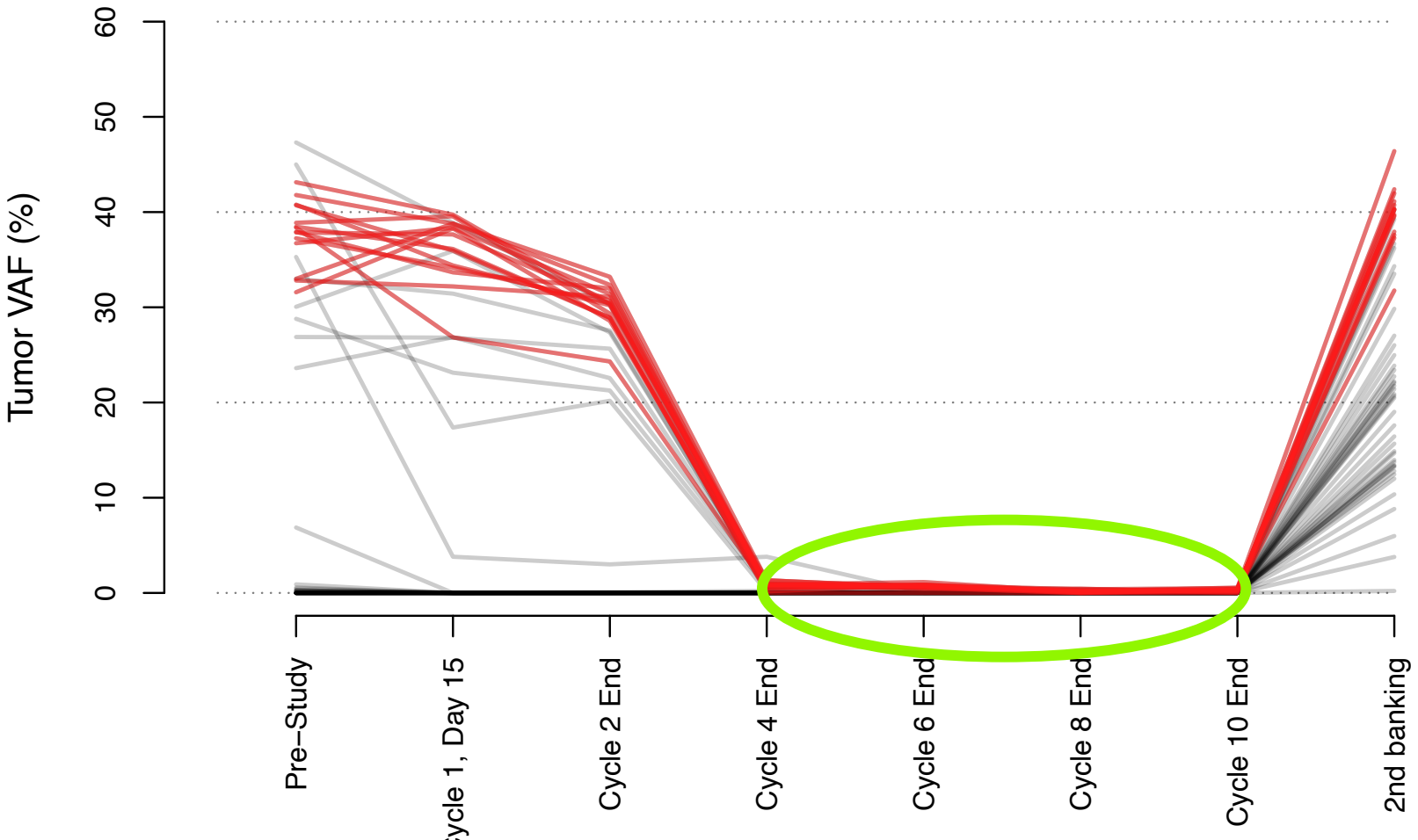
MDS response to decitabine



MDS response to decitabine



MDS response to decitabine

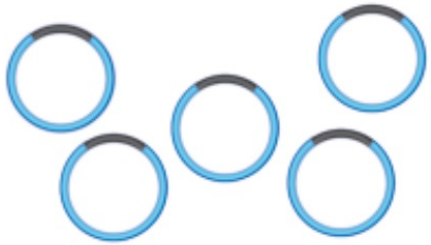


Error-corrected sequencing

1) Digest genomic DNA.

2) Hybridize the HaloPlex probe library in presence of the Indexing Primer Cassette. Hybridization results in gDNA fragment circularization and incorporation of indexes and Illumina sequencing motifs.

3) Capture target DNA-probe hybrids. Biotinylation of probe DNA allows capture using streptavidin-coated magnetic beads.



4) PCR amplify targeted fragments to produce a sequencing-ready, target-enriched sample.

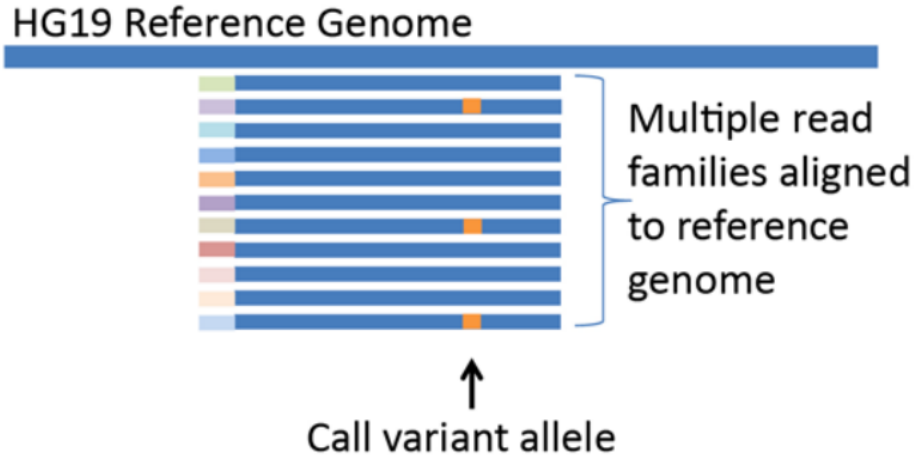
https://www.agilent.com/cs/library/usermanuals/public/G9900-90001_HaloPlex_Target_Enrichment%20_Manual.pdf

Barcoded sequencing

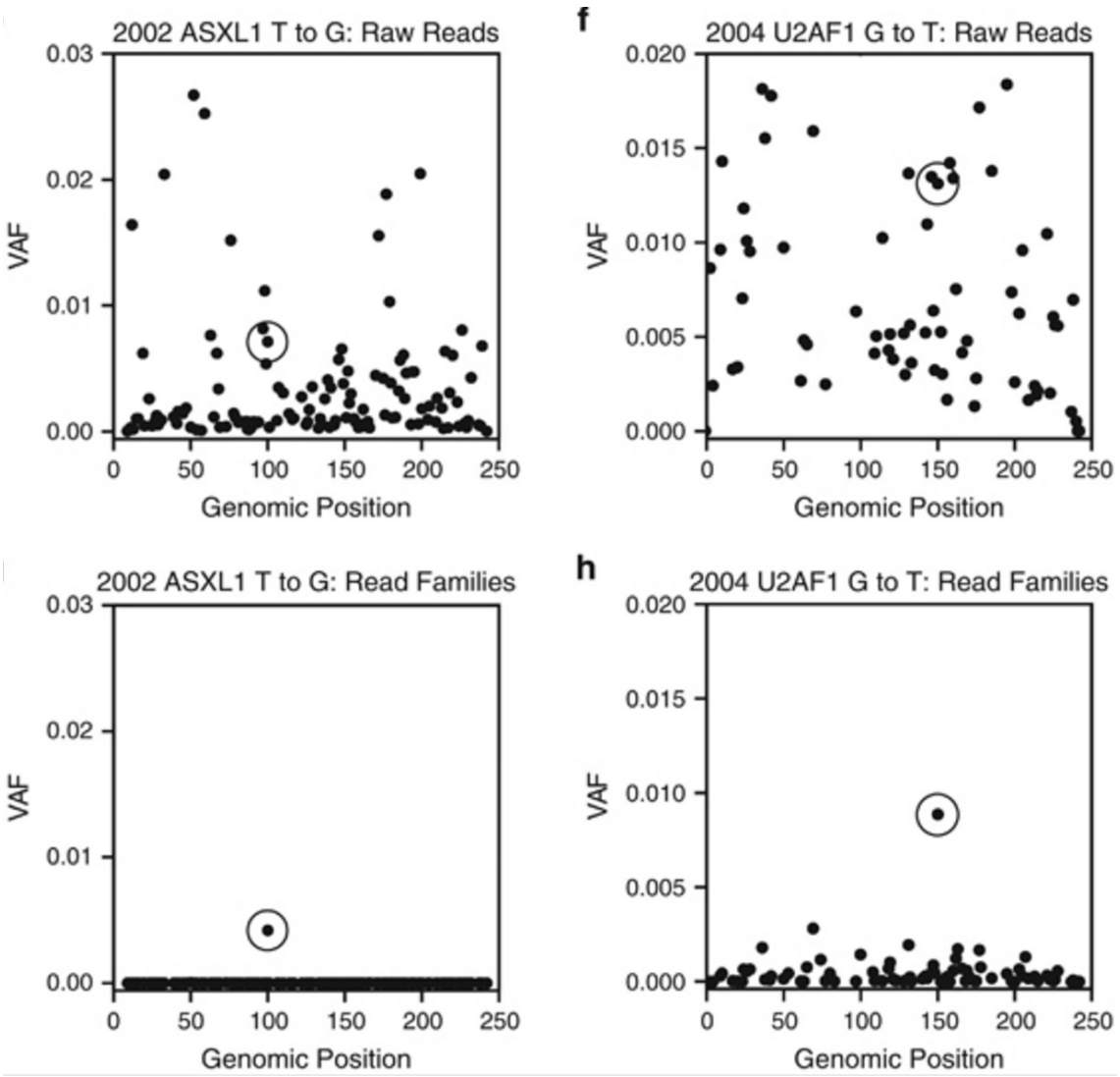
d. Generate read families

e. Create error-corrected consensus-sequence (ECCS)

f. Align ECCSs

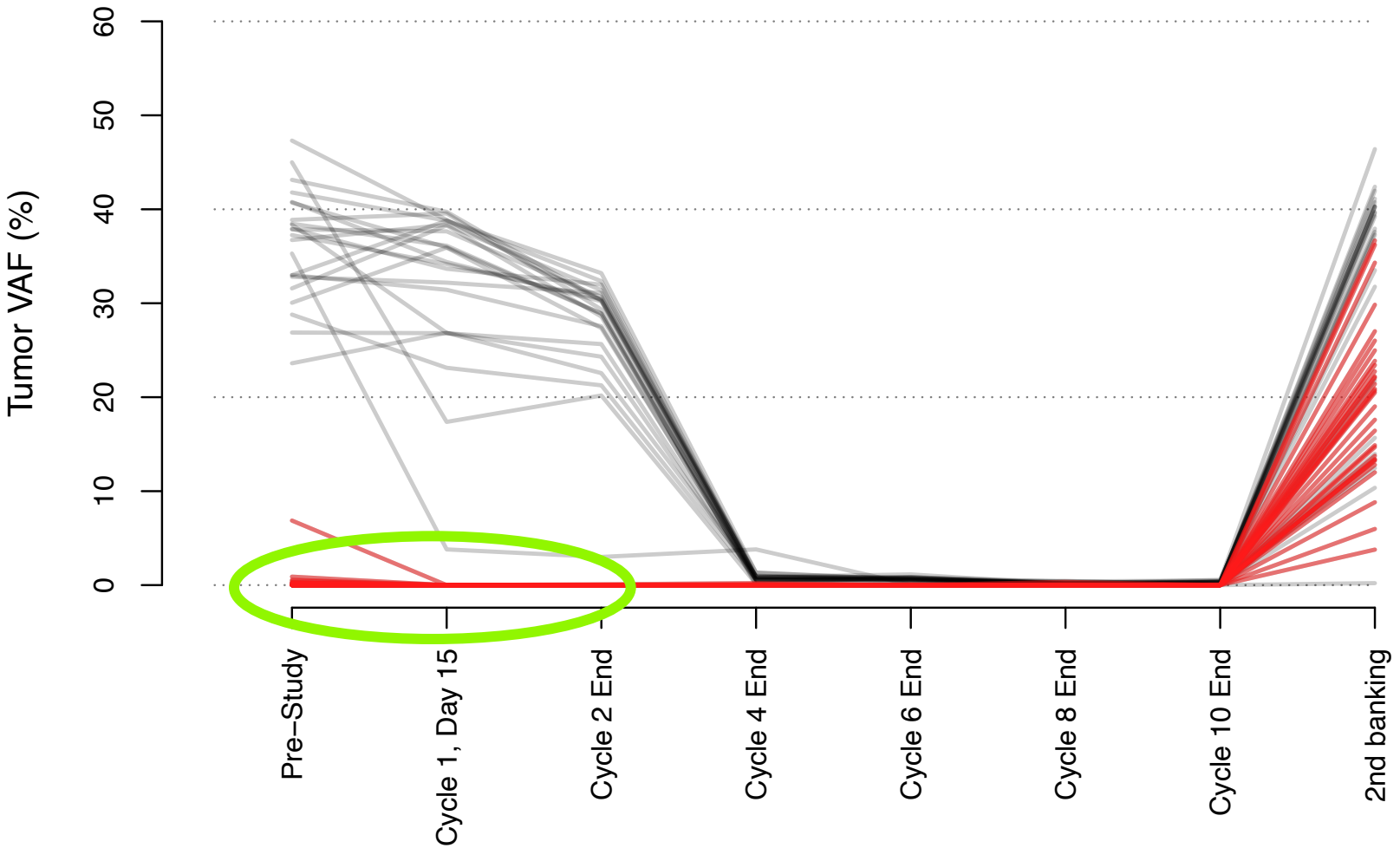


Ultradeep barcoded sequencing

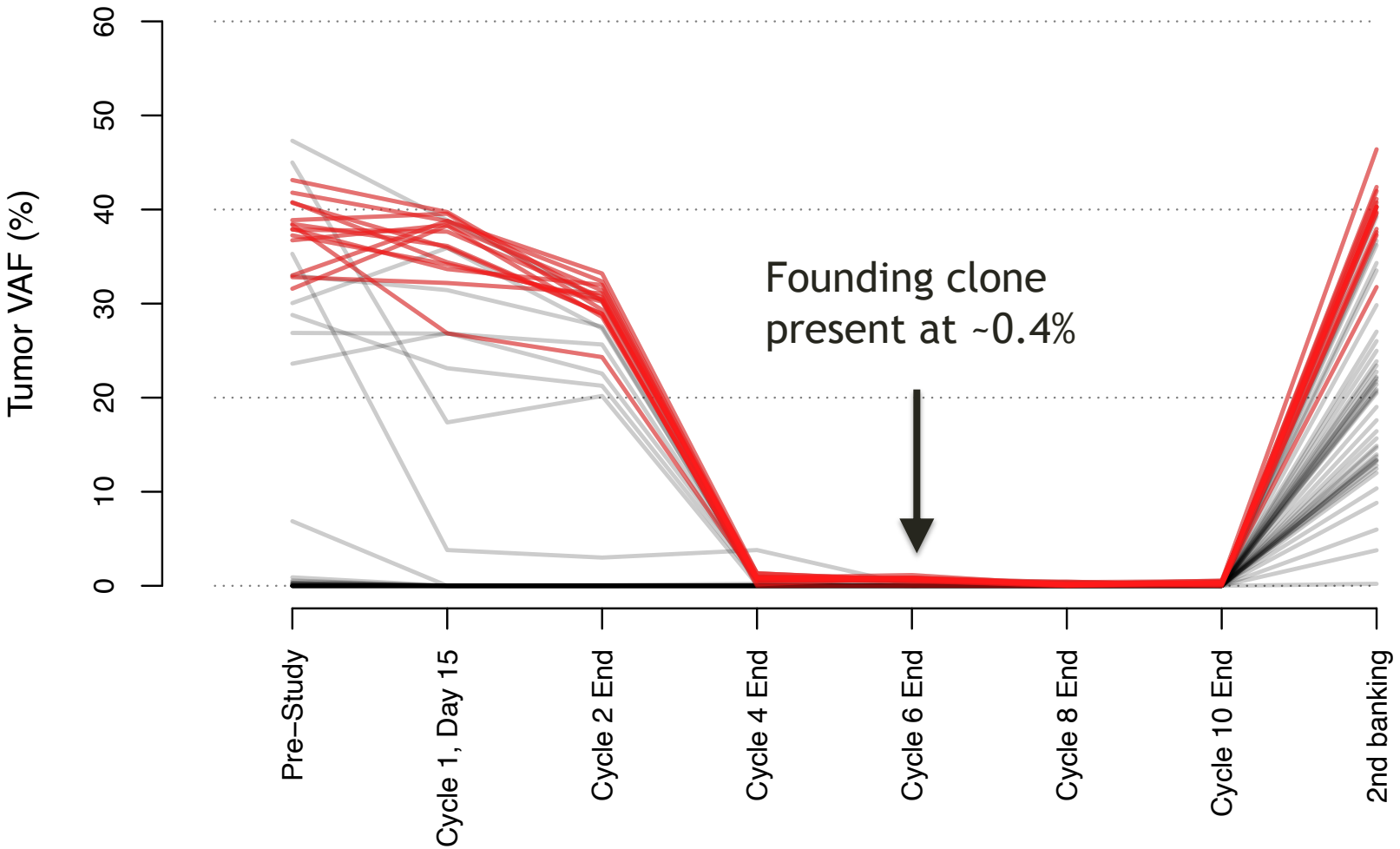


Young, et al. 2015 doi:10.1038/leu.2015.17

MDS response to decitabine



Ultradeep barcoded sequencing



Key tools for working with error-corrected reads

- analysis-workflows repo
  - pipelines > alignment_umi_molecular.cwl
  - pipelines > alignment_umi_duplex.cwl

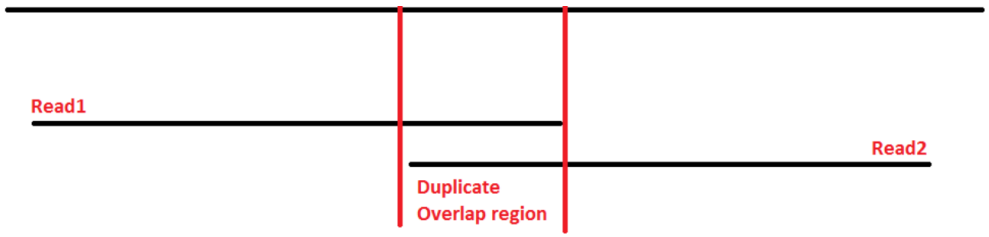
- fgbio tools
  - https://fulcrumgenomics.github.io/fgbio/

Alignment

- fgbio: ExtractUmisFromBam
  - Pulls UMIs out of sequence and into a tag (or tags)
- Align with BWA
- Picard – MergeBamAlignment
  - Takes the aligned bam, and re-attaches those UMI tags to the reads

Collapsing/Assembly

- fgbio: GroupReadsByUmi
  - re-sorting the bam to make all reads from the same molecule adjacent
- fgbio: CallMolecularConsensusReads
  - collapse/assemble them into one read per molecule
  - Sets some tags with stats – depth/error rate/etc
- fgbio: FilterConsensusReads
  - Filter based on above tags
- fgbio: ClipBam
  - Remove overlaps

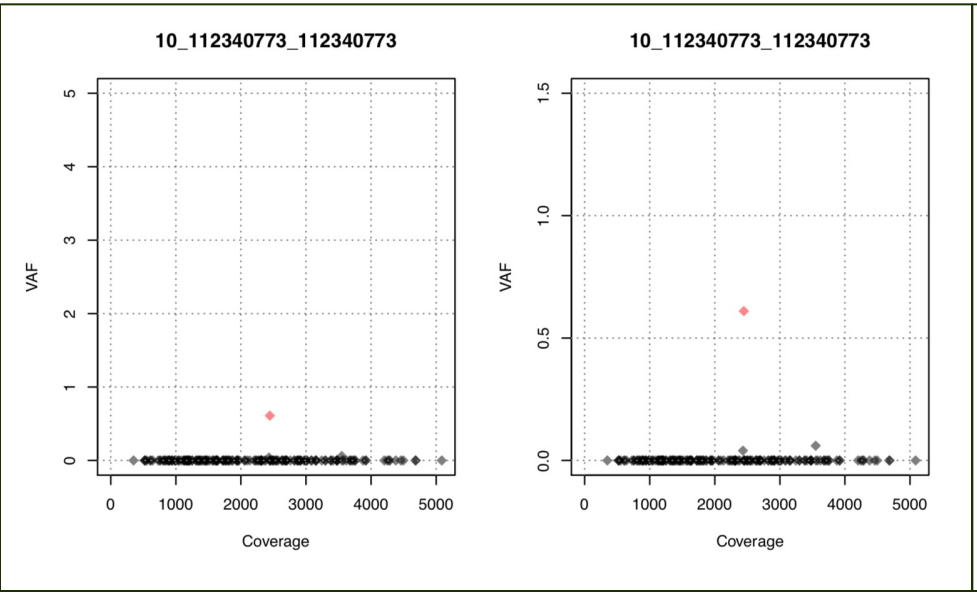


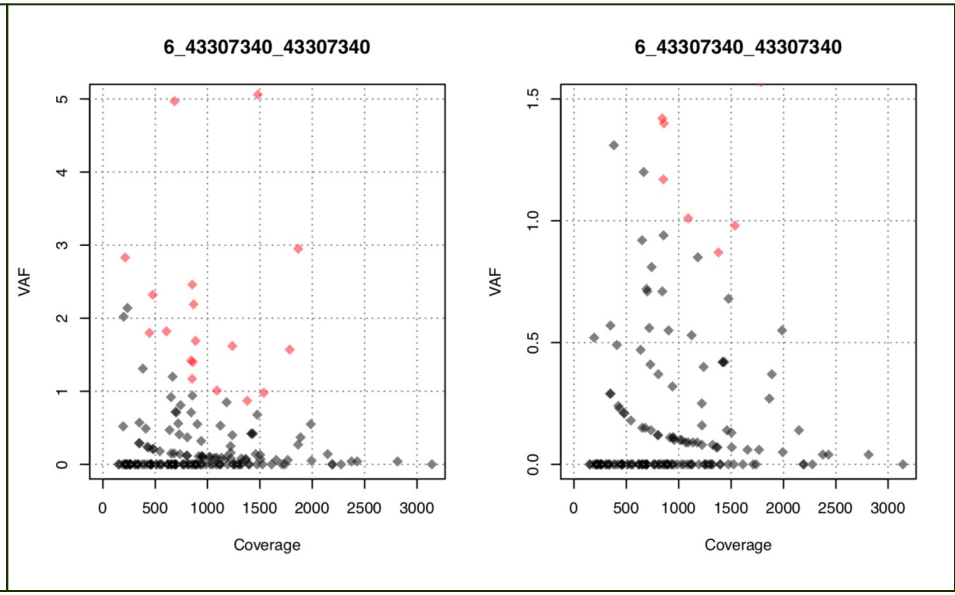
Results

- A bam with very low error levels
- Can be run through standard variant callers
  - Parameters may need to be tweaked!!

Results

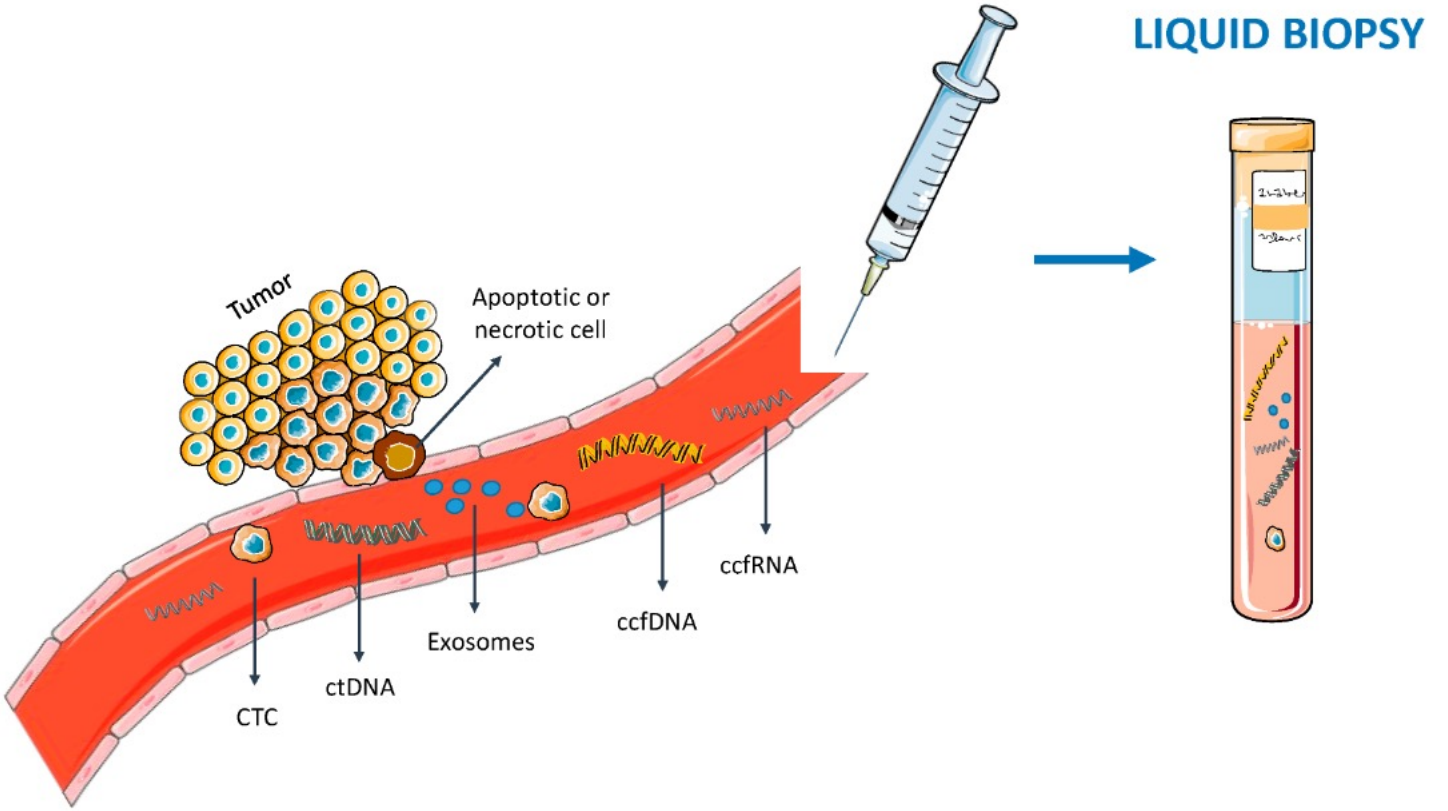
- Further filtering may be necessary
- Consider site-specific error rates
  - Iterative calling





Applications

- Circulating tumor DNA (ctDNA)

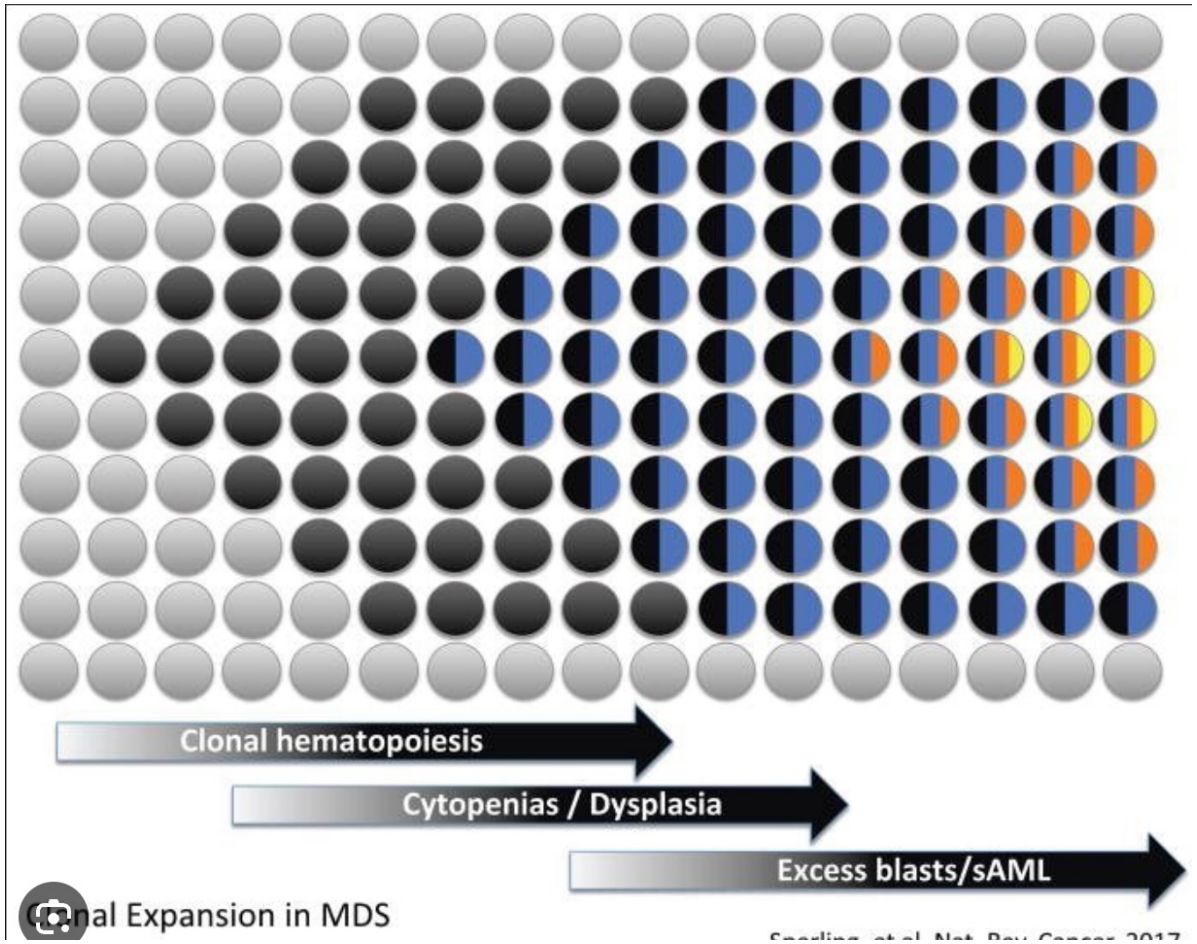


Early detection!

https://www.mdpi.com/2073-4409/9/3/624

Problems

- Clonal Hematopoeisis/Somatic mosaicism



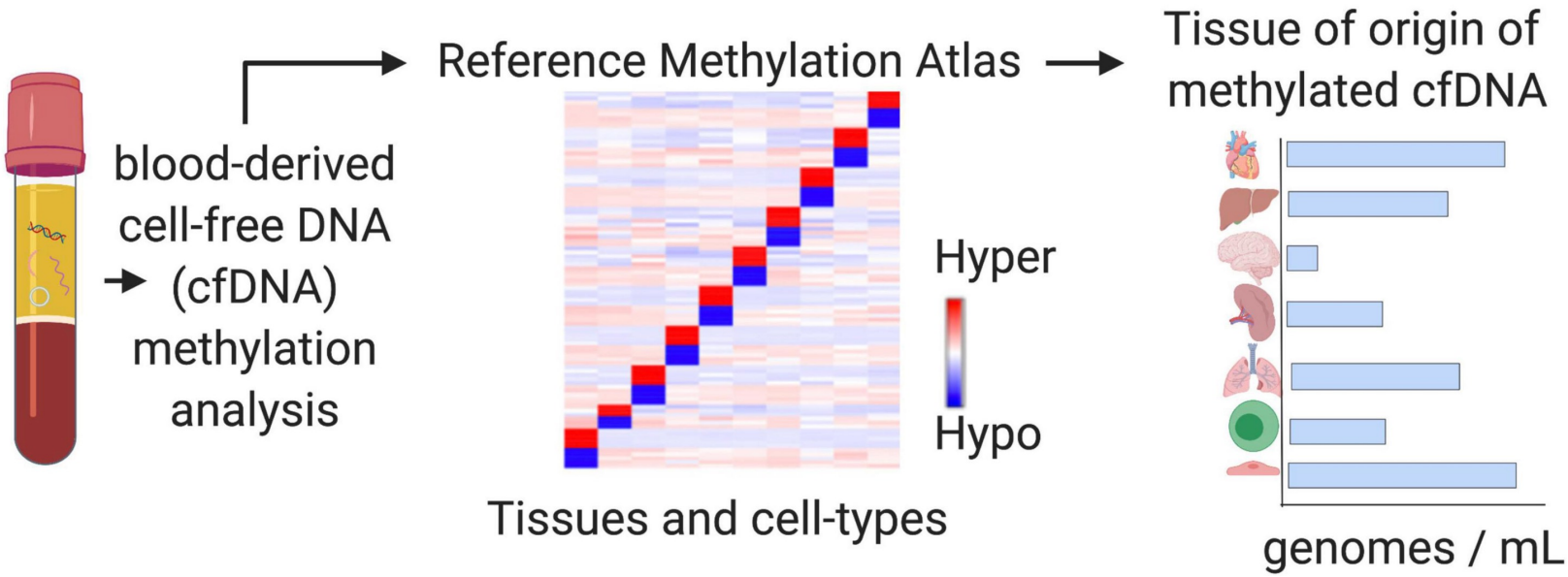
Sperling, et al., Nature Rev. Cancer, 2017

Screening

- High false positive rate – then what?!

Source identification

- Tissue specific methylation patterns



https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2021.671057/full

Applications

- Tracking residual disease
  - much closer to "prime time"
- Off-target effects of therapies
  - chemo-induced damage
  - off-target CRISPR/gene editing
- Microbial species
- When you need to find "needle in a haystack"