Impact of PCR Amplification during Library Preparation on Variant Discovery in Whole Genome Sequencing

Maxwell A. Sherman[1], Tiziana Sanavia[1], Daniel Kwon[1], Rachel Rodin[2], Christopher Walsh[2], Peter Park[1]

[1]Harvard Medical School Department of Biomedical Informatics
[2]Boston Childen's Hospital Departments of Neurology and Pediatrics

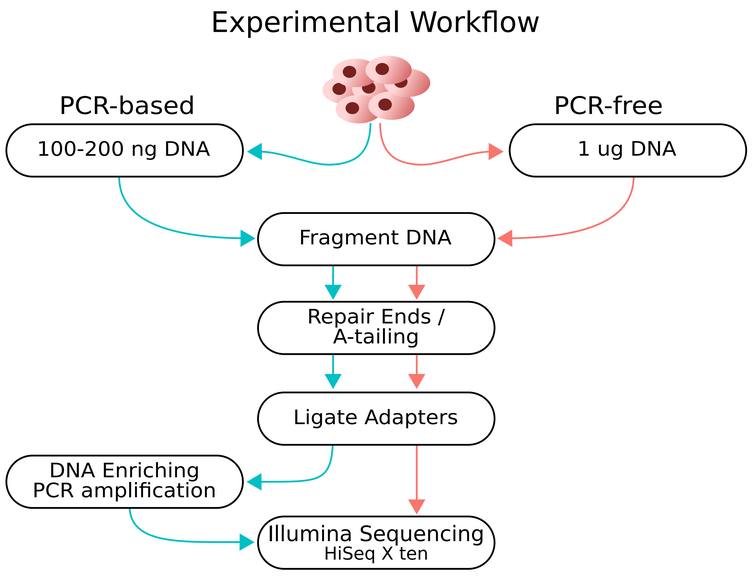
Introduction

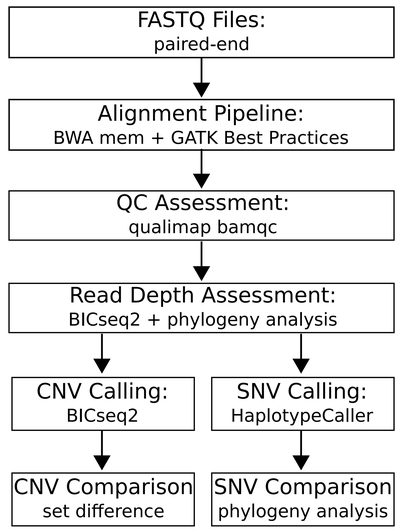
Whole genome sequencing (WGS) commonly utilizes PCR amplification during library preparation. Here we compare the effects of **PCR-based** and **PCR-Free** WGS library preparation methods on:

• Read depth uniformity • CNV detection • Germline SNV identification •

We find PCR-based methods can induce pronounced, non-biological variations in read depth which propogates significant false positive CNV calls and slightly reduces power to call germline SNVs.

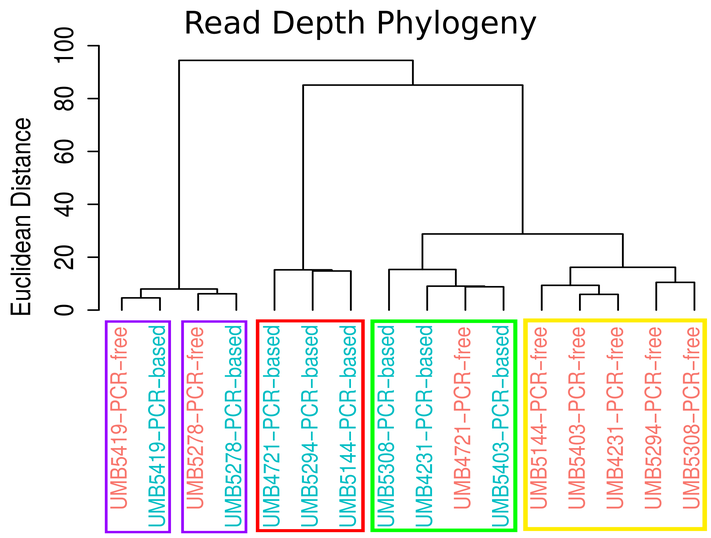
Experimental Design & Analysis Pipeline

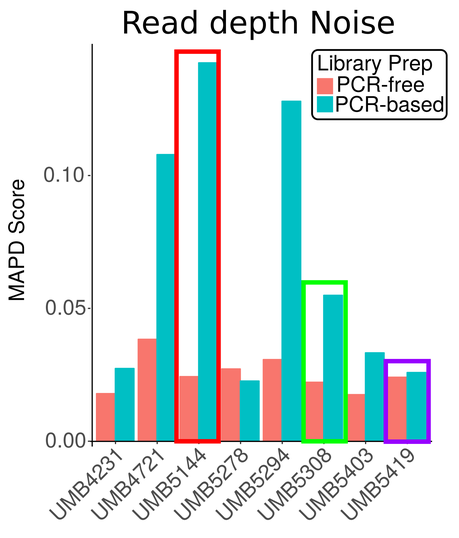




Read depth data was corrected for **mappability**, **GC content**, and **nucleotide composition**.

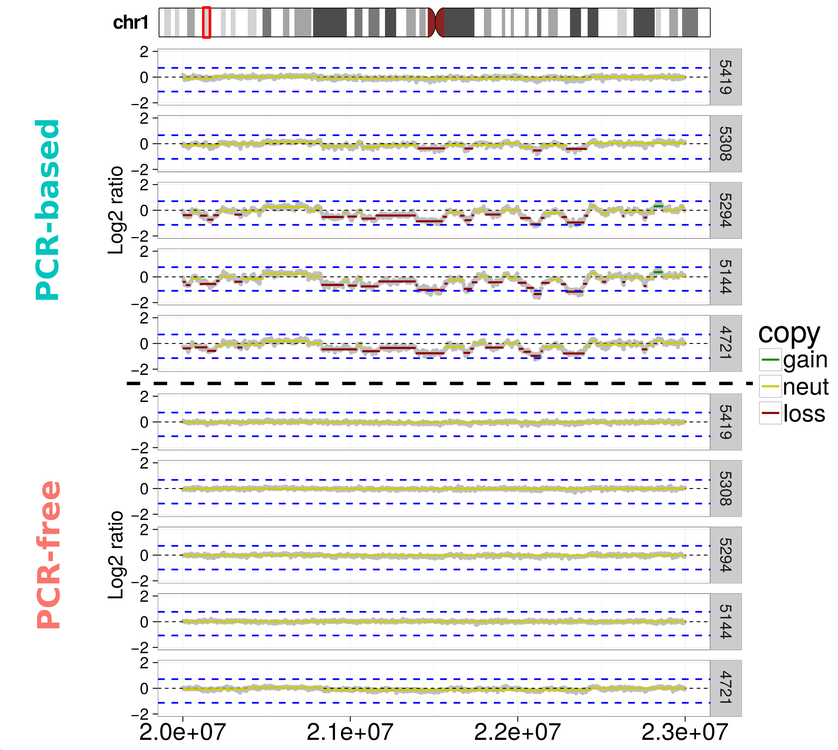
PCR-Based Library Preparation Induces Read Depth Noise

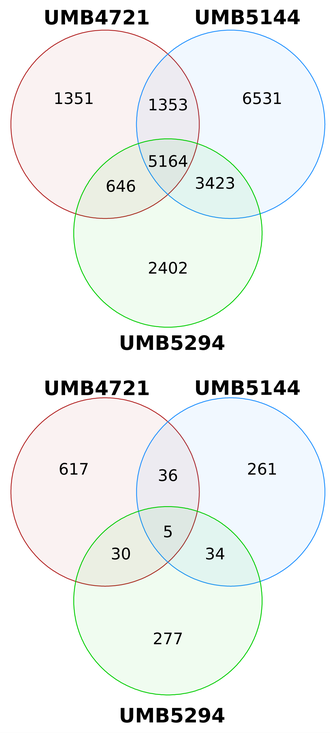




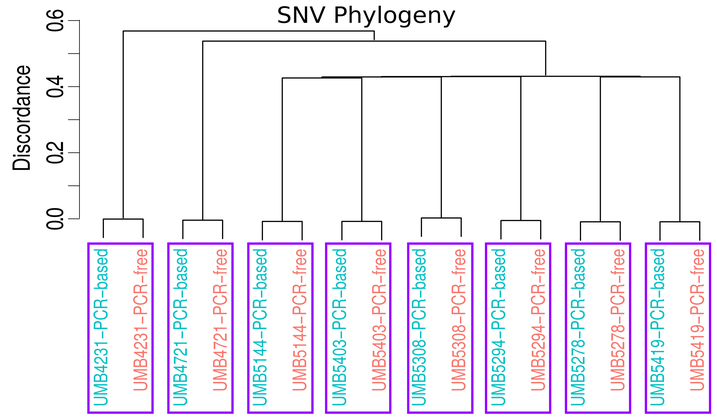
Purple: correct clustering; Red: PCR-based high noise; Green: PCR-based moderate noise; Yellow: PCR-free

PCR-Based Preparation Induces Recurrent CNV Artifact





Germline SNV Calling is Consistent Across Libraries



$$\text{discordance} = \frac{(A \cup B) - (A \cap B)}{A \cup B}$$

where A and B are SNV call sets.

- Discordance between biological replicates $= 0.31 \pm 0.098\%$ (~ 250 / 90000 variants)
- Significantly more unique calls (1420/2214) are found in PCR-free samples ($p < 0.01$, Welch T-test).
- PCR-based libs have alternate reads at sites unique to PCR-free, but number of alt reads is low.

Conclusions & Recommendations

1. PCR-based library preparation can introduce uncorrectable read depth variation into sequencing data.
2. This RD variability drives significant false positive CNV calls.
3. PCR amplification slightly reduces the power to call SNVs.
   Hypothesis: PCR amplification may unevely amplify genomic loci.

Recommendation 1: use PCR-free library prep whenever possible.
Recommendation 2: be cautious when comparing data prepared with different library methods.

Acknowledgements

- Alison Barton and Carl Vitzhum
- The Brain Somatic Mosaicism (BSM) Consortium
- The National Institute of Mental Health
- Manton Center for Orphan Disease Research