SINGULAR GENOMICS

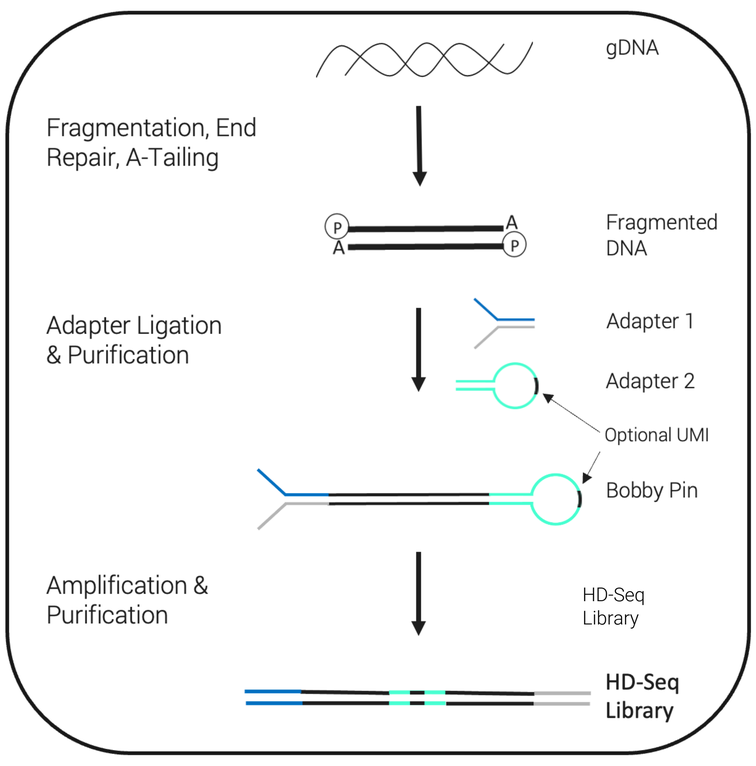
HD-Seq: A Novel Method for High Accuracy Sequencing

Sabrina Shore, Andrew King, Elizabeth A LaMarca, Kenneth Gouin III, Ryan Shultzaberger, Timothy Looney, Daan Witters, Martin M Fabani, Eli N Glezer
Singular Genomics Systems, Inc., San Diego, California

Abstract

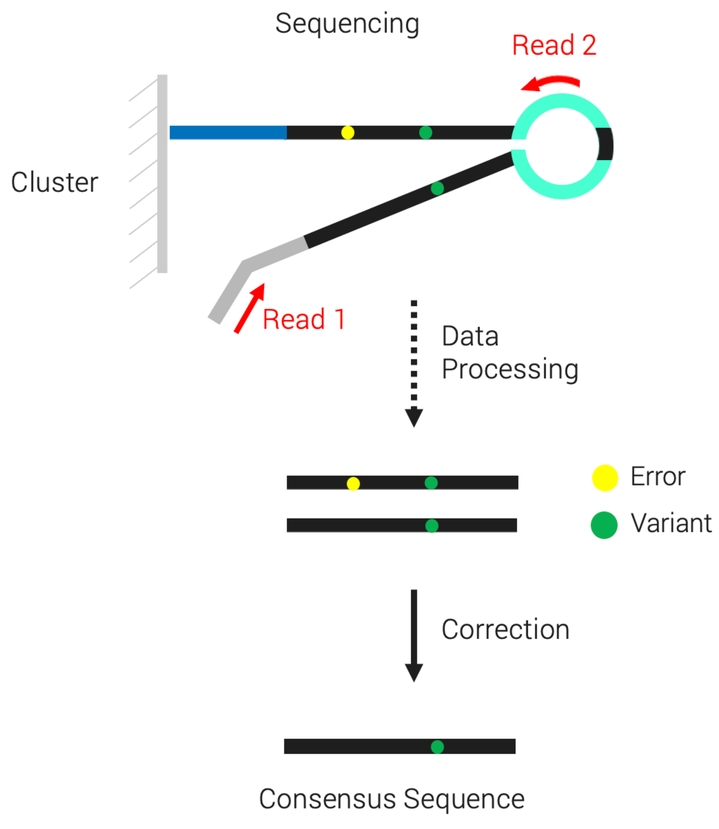
Next-generation sequencing (NGS) has enabled rapid progress in cancer diagnostics. However, challenges remain around detection of somatic variants at allele frequencies similar or below the error rate of on-market DNA sequencers (typically a~10^{-3}) when using standard library preparation methodologies. Here, we introduce HD-Seq™, a novel library preparation and sequencing method that allows for efficient and high-accuracy sequencing, achieving error rates in the range of ~10^{-5}-10^{-6}. The HD-Seq method relies on the physical linkage of original dsDNA templates in a sample, a connection that is maintained throughout library preparation, clustering and sequencing. By requiring consensus between the sequences of the linked strands, errors present in one of the original strands or introduced during amplification are readily identified and eliminated. In addition, error correction efficiency is maximized by guaranteeing that complementary strands of each molecule are represented in the sequence output.

Methods – Library Preparation



Library preparation for the HD-Seq method is similar to library preparation for standard libraries, except for use of two different adapters and an extra purification step to select for the desired bobby pin targets. Typically, 50 ng of fragmented gDNA (150bp mean size, Covaris) is end-repaired and A-tailed (Quantabio SparQ DNA library prep kit), ligated to Singular Genomics Adapters 1 and 2, and purified. Libraries are then amplified with 10 cycles of PCR and purified. If desired, libraries may be enriched using commercially available target capture kits (with minor modification, not shown).

Methods – Sequencing & Bioinformatics



Methods – Sequencing & Bioinformatics

HD-Seq libraries are sequenced on the Singular Genomics G4™ Sequencing Platform with standard paired-read 150-cycle format and analyzed with a custom bioinformatics pipeline. Briefly, reads are quality filtered, paired and adapter-trimmed. Read 1 (R1) and Read 2 (R2) are aligned to each other and non-overlapping portions of the read pair are trimmed. Discordant basecalls between R1 and R2 within the pairwise alignment are assumed to reflect library preparation or sequencing errors and are masked (called 'N'). After error correction by sequence consensus, reads are aligned to the corresponding reference genome and variants called (if applicable).

Results – HD-Seq Error Correction Enables >Q50 Accuracies

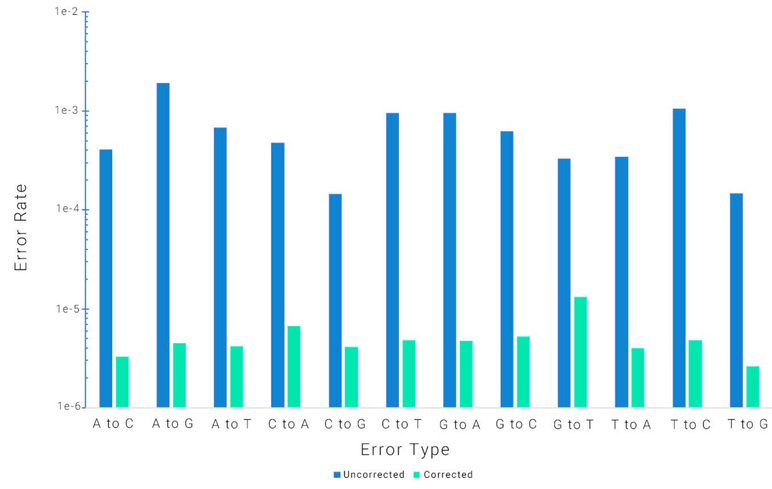


Figure 1. The typical mismatch rate observed in single-pass (non-consensus) sequencing is approximately 10^{-3}. The HD-Seq method enables >100-fold lower error rates by building a consensus sequence from linked strands.

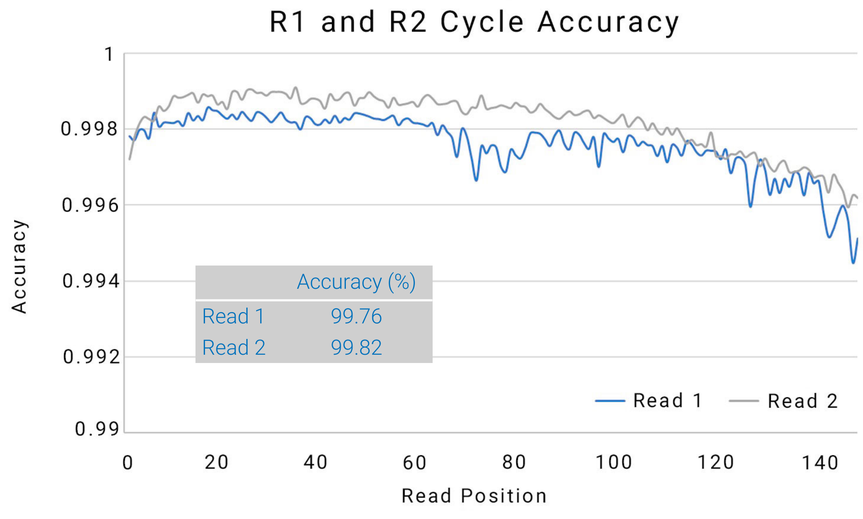


Figure 2. Prior to HD-Seq correction, Read 1 and Read 2 from a representative HD-Seq experiment (Salmonella gDNA input) show mean accuracies of 99.76% and 99.82%, respectively.

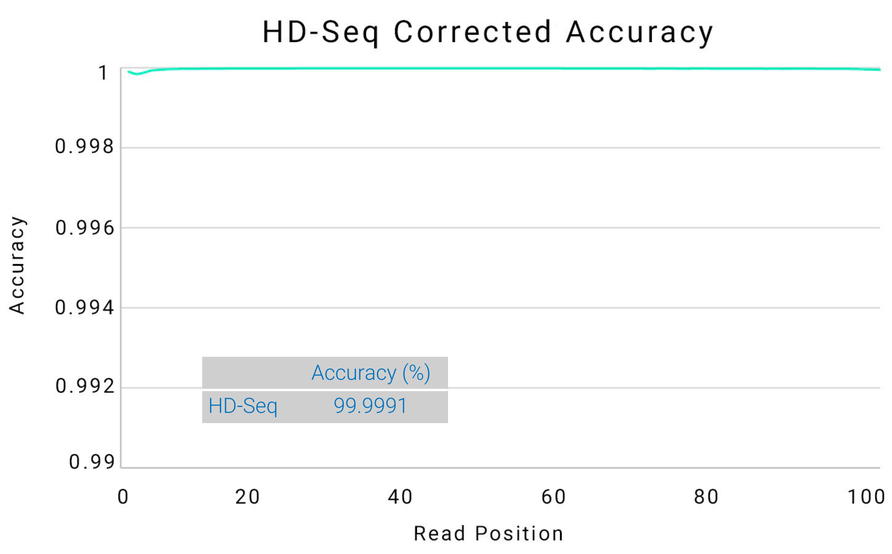


Figure 3. When reads are combined to build a consensus sequence, the accuracy increases to 99.9991%.

Conclusion

The HD-Seq approach is a novel library preparation and sequencing method that relies on linking the original paired strands in a dsDNA fragment. DNA errors affecting one of the two strands can be easily identified during sequencing of the corresponding linked strands. The G4 Sequencing Platform's high single-pass accuracy enables >Q50 accuracies with the novel HD-seq methodology.