Streamlined, Efficient, and Uniform Molecular Inversion Probe Capture for Targeted Sequencing

INVITAE | GoodStart Genetics

Eric D Boyden, Eric F Tsung, David J Maganzini, Gregory J Porreca, Mark A Umbarger

Background

Molecular inversion probes (MIPs) are short oligonucleotides (~50-200 nt) that contain a common "backbone" sequence flanked by unique ends or "targeting arms" designed to be complimentary to genomic regions of interest. MIPs can be used to enrich these targets for sequencing via stringent hybridization, followed by enzymatic gap fill-in to create circular molecules. Next, free MIPs and gDNA are nuclease-digested, and finally PCR using backbone-derived primers is performed to incorporate sample barcodes and sequencing adapters.

MIP-based capture is an attractive alternative sample preparation strategy for sequencing, as it simultaneously enables highly specific target enrichment and efficient library construction with relatively little hands-on labor. However, the adoption of MIP technology has been limited in part due to the technical challenge of designing probes that yield uniform capture and sequencing coverage, as well as the complexity of the experimental workflow. We have optimized the computational MIP design process, and the laboratory workflow, to minimize these barriers to adoption.

Methods

We developed an algorithm to design probes that tile genomic intervals on alternating strands (Figure 1). This reduces the risk of allele dropout, a common pitfall of target enrichment; improves capture efficiency and sequencing uniformity, which reduces sequencing costs and sample input requirements; and generally obviates the need to empirically optimize probe concentrations.

Additionally, we employed unique molecular identifiers (UMIs or UIDs) in the designed MIPs (Figure 1, gray rectangles in MIPs). UMIs, which are short random sequences that label individual target molecules pre-amplification. UMIs facilitate the removal of PCR duplicates, a potentially significant source of bias and noise; improve the detection of low frequency somatic variants by distinguishing them from sequencing errors; and enable the estimation of capture efficiency.

Finally, we developed a streamlined single-tube workflow that is easily automated, utilizes dual barcodes to minimize the risk of barcode swapping, is scalable to run hundreds of samples in parallel, and can be completed in under 12 hours (Figure 2).

Results

Across 5 MIP panels targeting polymorphic loci and gene coding regions throughout the human genome, we observed that greater than 98% of the target bases had coverage depth greater than or equal to 20% of the mean (Figure 3 top). Fold-80 penalties (the fold increase in sequencing depth required to raise the 20th percentile coverage to the mean) ranged from 1.3 to 1.8 across multiple probesets and DNA inputs down to 1 ng (Figure 3 bottom). The uniformity remained fairly constant when the gDNA input was decreased from 100 ng to 1 ng (Figure 3, probeset B).

Using UMI-encoded MIPs, we calculated a median per-locus capture efficiency estimate of ~12% (Figure 4 top), and observed that allele fractions of 5% can be detected reliably using as little as 10 ng of gDNA input (Figure 4 bottom).

Conclusion

Collectively, our highly optimized MIP-based capture technology represents an appealing option for efficient, uniform, streamlined, and scalable targeted sequencing.

Figure 1. Strand-Alternating Probe Design

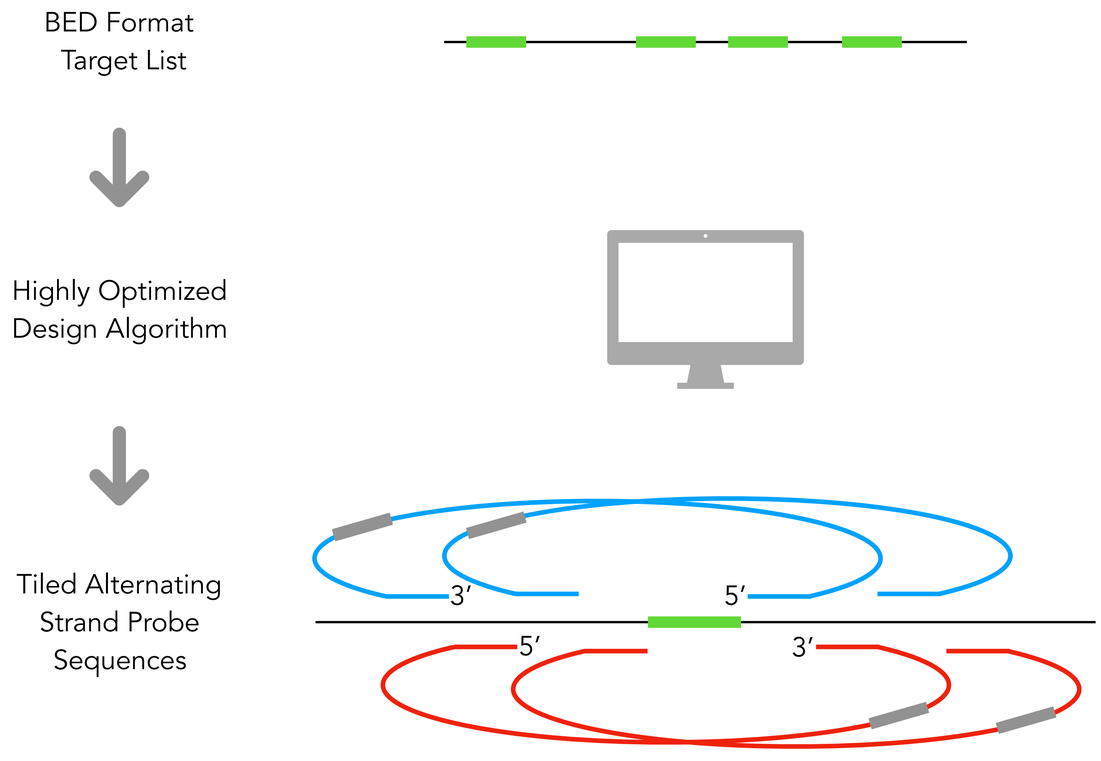


Figure 2. Automated Single-Tube Workflow

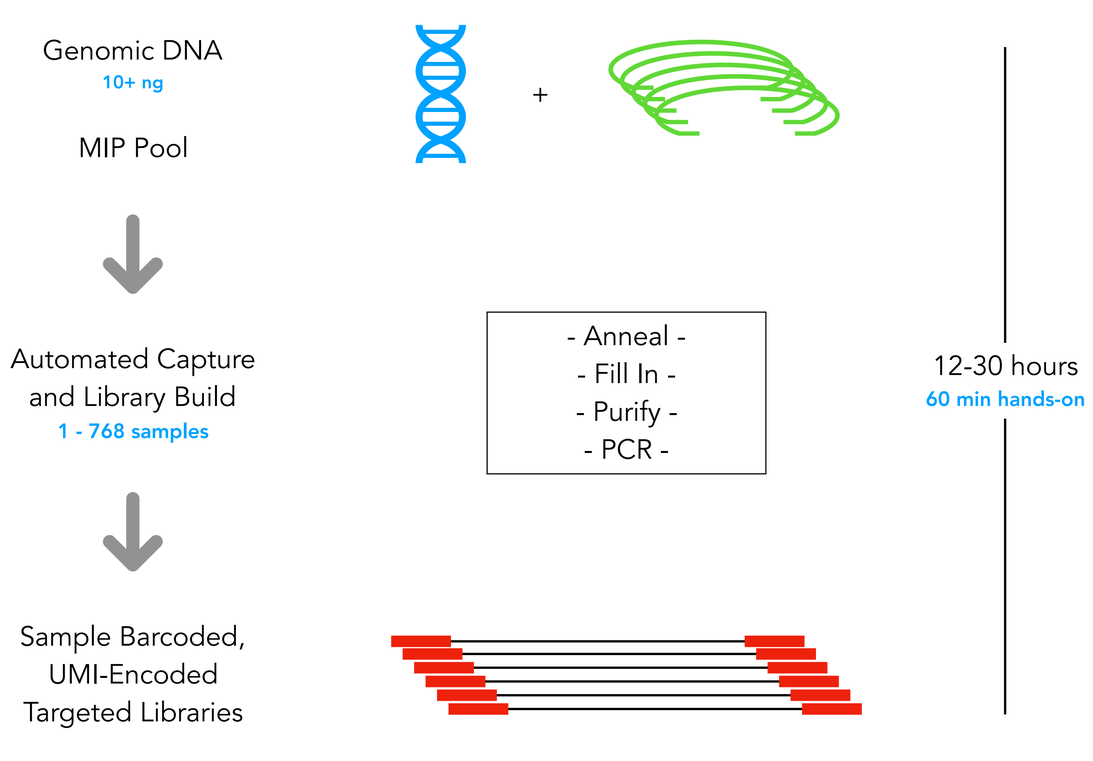
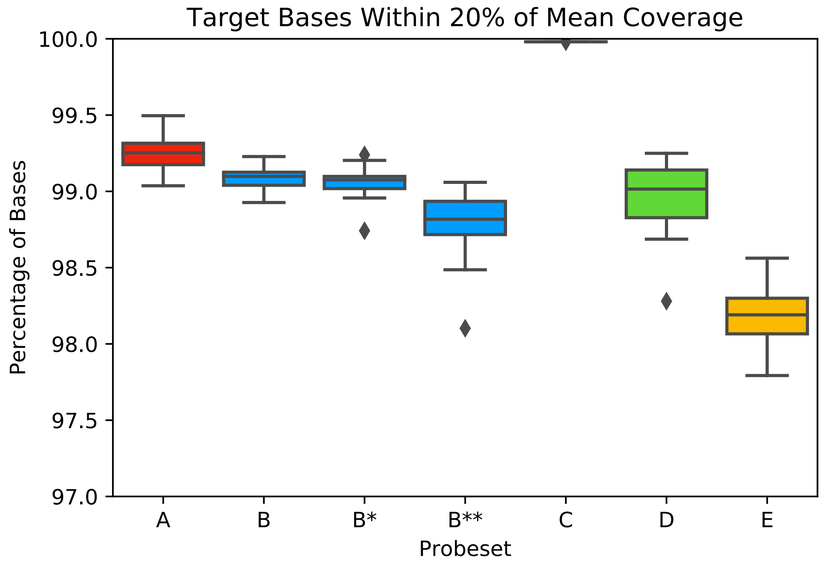
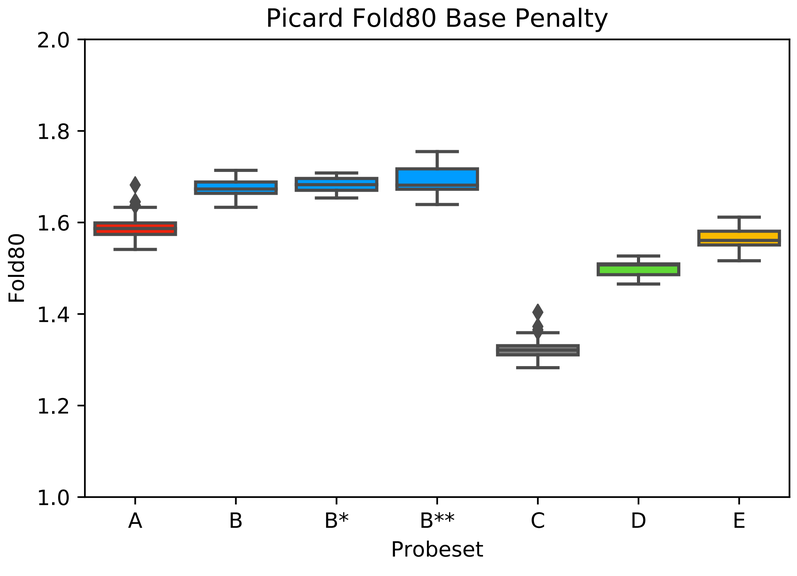


Figure 3. Target Depth Uniformity for Inputs as Low as 1 ng





Probeset B was used to perform capture reactions with varying amounts of input DNA:

B = 100 ng
B* = 10 ng
B** = 1 ng

Figure 4. Capture Efficiency and Low-Frequency Variant Detection