High Performance Solution-Based Target Selection Using Individually Synthesized Oligonucleotide Capture Probes

Mirna Jarosz, Zac Zwirko, Doron Lipson, Garrett Frampton, Roman Yelensky, Alex Parker, Maureen Cronin
Foundation Medicine, Inc., Cambridge, MA

Introduction

The availability of solution-based genomic target selection techniques has enabled rapid development of targeted sequencing applications, some of which have led to the introduction of clinical sequencing tests. Commercialized hybridization capture reagents are based on array-synthesized oligonucleotides, which are converted to biotinylated DNA or RNA probes ("baits"). However, all methods of generating these complex pools of probes face performance challenges, for example capturing high-GC content targets.

We present an alternative approach using individually synthesized, 5'-biotinylated oligonucleotides for capturing a target region of ~130kb representing 57 clinically relevant and actionable cancer-related genes.

Materials and Methods

Highly complex ligation-based sequencing libraries were prepared and then amplified by PCR using HiFi polymerase from Kapa Biosystems. In-solution hybrid selection using a 24hr hybridization time was performed with 2μg of library input and either a set of 1,000 5'-biotinylated 120nt DNA oligos spanning ~130kb of target territory or a set of 369 5'-biotinylated 120nt DNA oligos spanning ~29kb of target territory (representing the complete coding sequence of 6 genes). The 5'-biotinylated 120nt oligo baits were supplied by Integrated DNA Technologies (IDT) and were synthesized with IDT's high-fidelity synthesis process (Ultramer™ Oligos). For the 1,000 oligo set, captures were also performed with hybridization times reduced to 2.5 hrs.

5'-biotinylated 120nt oligonucleotide baits were also spiked into array-derived RNA baits to enhance coverage of high-GC regions. For these captures, either 1,000 oligo baits (~133kb target territory) or 3 oligo baits (1 exon) were mixed with the array/RNA baits; otherwise the conditions were equivalent to Foundation Medicine's standard array/RNA bait captures.

All samples were sequenced on an Illumina HiSeq™ 2000 platform using 49x49 paired-end reads.

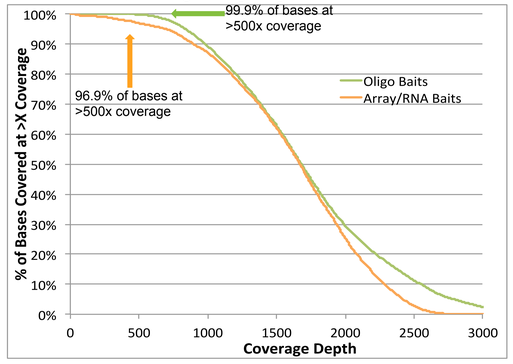
Results

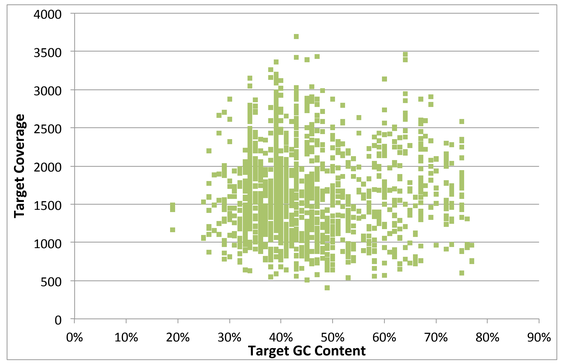
Oligo bait captures produce a high level of enrichment, even for small territories

Bait Set	Target Territory	Fold Enrichment
Array/RNA Baits	1.1 Mb	1,665x
1,000 Oligos	133 kb	4,888x
369 Oligos	29 kb	21,661x

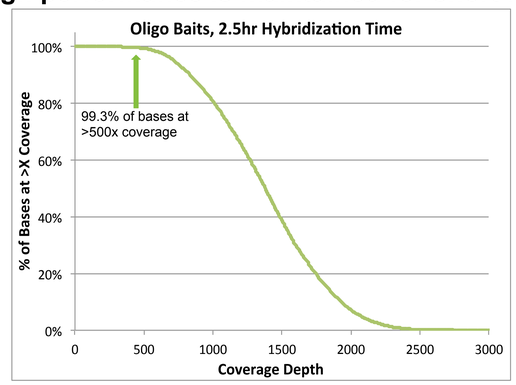
Oligo bait captures yield high coverage over the entire targeted region with minimal GC bias

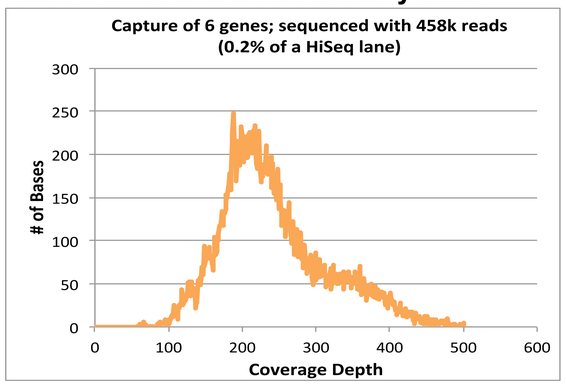
Data shown for 1,000 oligos targeting 133kb





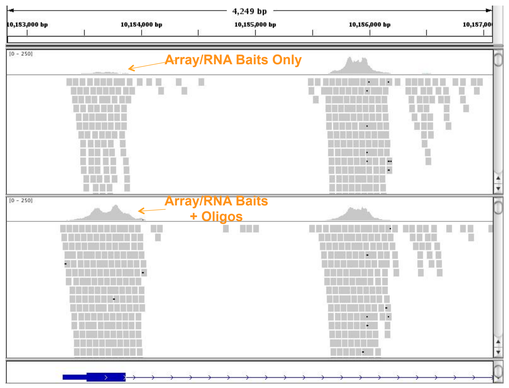
High performance is maintained for short hybridization times and small territory sizes



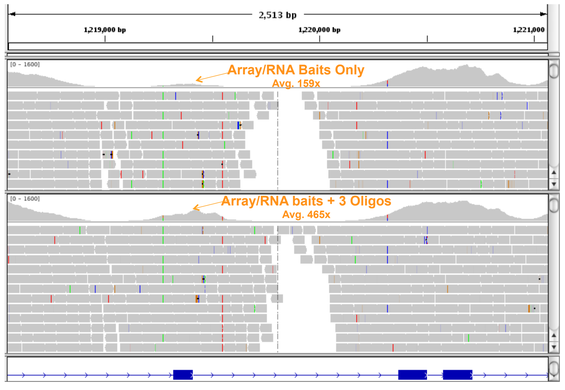


Oligo baits can be added to array-derived RNA bait reactions to enhance coverage

Mixing a 1000 oligo set (133kb) with a 1.1MB array/RNA bait set enhances coverage of many GC rich targets (e.g. 1st exons)



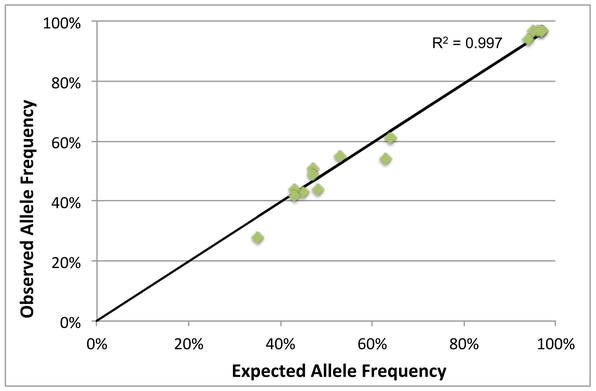
Adding 3 oligos targeting a single GC rich exon to a 1.1MB array/RNAt bait set enhances coverage of that exon



Capture of insertions and Deletions

For tumor cell lines captured with oligo baits, insertions and deletions are detected at the expected frequencies

Range of insertion lengths: 1 to 35
Range of deletion lengths: 1 to 36



Conclusions

- Individually synthesized DNA baits can be used successfully for targeted ultra-deep Next Generation Sequencing. We demonstrate:
  - High level of enrichment (~5,000-fold +)
  - Minimal GC bias
  - Ultra-deep coverage of the entire targeted region
  - Hybridizations in a few hours, not days
  - High level of performance even for very small target areas (e.g. 6 genes)
  - Efficient capture even of IN/DEL containing alleles
- Individually synthesized DNA baits can be spiked into array-derived RNA baits to rescue low coverage regions

References

Ross JS, et al. (2010) Am J Clin Pathol 134(3): 478-90

Gnirke A, et al. (2009) Nature Biotechnology 27: 182 -189

Fisher, S et al. (2011) Genome Biology 12: R1

Dahl F, et al. (2007) PNAS 104(22): 9387-939

Summerer D. (2009) Genomics 94(6): 363-368

Porreca GJ, (2007) Nature Methods 4: 931 – 936

Forbes SA,. (2011) Nucleic Acids Res. D945-50