Targeted Exon Sequencing by In-Solution Hybrid Selection

Brendan Blumenstiel,[1] Kristian Cibulskis,[1] Sheila Fisher,[1] Matthew DeFelice,[1] Andrew Barry,[1] Tim Fennell,[1] Justin Abreu,[1] Brian Minie,[1] Maura Costello,[1] Geneva Young,[1] Jared Maquire,[1] Andrew Kernytsky,[1] Alexandre Melnikov,[1] Peter Rogov,[1] Andreas Gnirke,[1] and Stacey Gabriel[1]

[1]Broad Institute, Cambridge, Massachusetts

ABSTRACT

This unit describes a protocol for the targeted enrichment of exons from randomly sheared genomic DNA libraries using an in-solution hybrid selection approach for sequencing on an Illumina Genome Analyzer II. The steps for designing and ordering a hybrid selection oligo pool are reviewed, as are critical steps for performing the preparation and hybrid selection of an Illumina paired-end library. Critical parameters, performance metrics, and analysis workflow are discussed. *Curr. Protoc. Hum. Genet.* 66:18.4.1-18.4.24. © 2010 by John Wiley & Sons, Inc.

Keywords: exon sequencing • hybrid selection • mutation discovery • DNA sequencing • targeting

INTRODUCTION

The ability to identify rare polymorphisms in the human genome is crucial for discovering genetic associations and causative mutations related to human disease. With the completion of the Human Genome Project (Lander et al., 2001; *http://www.genome.gov/10001772*), the framework was set for establishing a deep understanding of genomic variation, its structure, and its role in human disease. While genomic sequencing is the most powerful tool for identifying a variety of genetic variants, whole-genome sequencing of thousands of samples remains prohibitively expensive, thus requiring targeted approaches to sequencing genomic regions of interest (Ng et al., 2009).

Traditionally, targeted sequencing has been performed using single-plex PCR-based amplification followed by Sanger sequencing (Sjöblom et al., 2006). For a multitude of reasons including cost and logistic workflow, PCR-based targeting is no longer a cost-effective match for many of the new next-generation sequencing technologies emerging on the market. In recent years, several methods of exon targeting by hybrid selection have been developed by leveraging the massively parallel synthesis of long oligonucleotides on programmable arrays (Li et al., 2008; Gnirke et al., 2009). This relatively inexpensive method for simultaneous synthesis of tens of thousands of unique oligos has led to highly multiplexed methods for exon sequencing.

By moving to array-based oligonucleotides ranging from 60 to 170 bp in length, precise targeting of relatively short exons across many genes is possible. Programmable oligonucleotide microarrays can be used to capture and enrich exons by either solid-phase or solution-based hybrid selection. In the solution-based method developed and implemented at the Broad Institute, PCR-amplified DNA probes are then transcribed into biotinylated RNA, which is hybridized in-solution with a randomly sheared genomic DNA library. Hybridized DNA-RNA duplexes are pulled down using streptavidin-coated magnetic beads. Immobilized beads are then washed, removing non-hybridized DNA.

The remaining captured DNA is subsequently denatured from the immobilized RNA, enriched by PCR, and sequenced on Illumina's Genome Analzyer II (GAII) sequencing system (Gnirke et al., 2009).

Outlined in this unit are the steps for performing solution-based hybrid selection of exons and preparing enriched libraries for paired-end Illumina sequencing on the Illumina GAII. Steps include genomic DNA shearing (Basic Protocol 1); Illumina paired-end library construction (end repair, A base addition, paired-end adapter ligation, PCR enrichment, and clean-up; Basic Protocol 2); hybrid selection (Basic Protocol 3); and library quantification for optimized cluster density using qPCR (Basic Protocol 4). In the Support Protocol, we describe several recommendations for performing read alignment, calculating meaningful hybrid selection metrics, and visualizing and assessing sequence data for overall protocol performance. Specific challenges and points of sensitivity are further discussed, along with specific performance metrics that can be expected by following the published protocols.

STRATEGIC PLANNING

Choosing Targets and Baits

The process for choosing targets is fairly straightforward, and several standard capture panels are commercially available, such as the Agilent SureSelect Human All Exon Kit. In choosing custom targets for hybrid capture, there are two major areas of consideration: target uniqueness and target size.

The genome-wide uniqueness of the capture targets must be considered. If a region is not sufficiently unique in the genome, it may not be able to be aligned uniquely with short reads. Therefore, although the DNA fragments may be physically captured and sequenced, it is not straightforward to analyze the data. A more critical, related problem is targeting regions of high copy number in the genome, such as mitochondrial genes and ALU repeats. Targeting these regions is detrimental, not only because the results are difficult to interpret, but because the high representation of these regions in the DNA causes them to be oversampled. As an example, in one recent capture experiment, 7% of reads mapped to targeted mitochondrial genes, even though those genes, represented only 0.1% of the target set (unpub. observ.).

The total size of the targets to be captured has an effect on the efficiency of the hybrid selection, with smaller target sets causing a smaller fraction of reads to align to the target. With large target sets, such as whole-exome capture, over 80% of the reads typically align to the desired target. However, with smaller sets of a few hundred genes, often only 50% to 70% of reads may align to the target. Once a set of targets is chosen, baits are typically tiled across the target with a small overlap between baits, as seen in Figure 18.4.1. The figure also illustrates the nomenclature commonly used to refer to regions surrounding the targets and baits.

Biotinylated RNA Baits

Solution-based hybrid selection involves the hybridization of a prepared paired-end Illumina library ("pond") with a pool of biotinylated RNA ("baits"). These RNA baits are generated from unique oligonucleotides synthesized on an Agilent programmable DNA microarray. Up to 55,000 unique oligos can be synthesized simultaneously; they are 150-200 bp in length and include 15-bp universal PCR primer sites at the extreme ends. Following synthesis, the oligos are stripped from the array substrate and are universally PCR amplified into double-stranded DNA. A second round of PCR incorporates a T7 promoter site into the amplicon, which is used to transcribe the DNA into single-stranded, biotinylated RNAs. This process has recently been commercialized by Agilent Technologies and is currently being marketed as the SureSelect Target

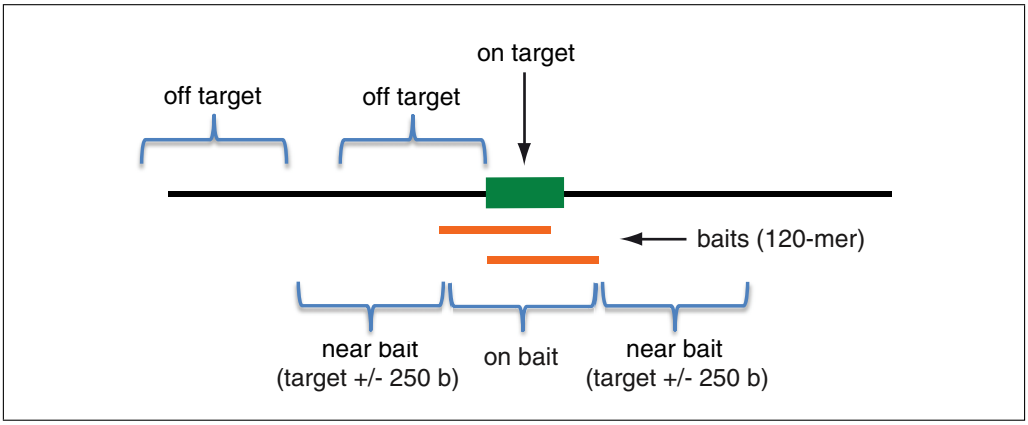

Figure 18.4.1 Targets, baits, and nomenclature. Sequencing reads can fall into several categories depending on where they align along a targeted region of the genome. Bases aligning to the exact targeted sequence are considered "on target." Because RNA bait sequences can hang off the ends of the actual target, aligned bases can be "off target" but "on bait." Additionally, because randomly sheared fragments vary in size, it is realistic to expect a proportion of aligned bases to be "near bait," which is considered ±250 bp of the bait sequence. Metrics calculating the percentage of bases falling into these categories are helpful in understanding the performance of a hybrid selection experiment. For the color version of the figure, go to *http://www.currentprotocols.com/protocol/hg1804*.

Enrichment System. Agilent has developed a streamlined web interface for uploading custom probe sequences that can be synthesized and manufactured into a ready-to-use biotinylated RNA pool (*https://earray.chem.agilent.com/earray*).

DNA Quality and Quantity

DNA quality and quantity must be considered when compiling a cohort for hybrid selection sequencing. If available, DNA samples with more than 3 µg of high-quality DNA should be used. Although whole-genome DNA extracted from cell lines and blood is preferred, whole-genome amplified DNA can be used as long as the starting DNA is not highly degraded. Before beginning a hybrid selection study, all samples should be quantified and an aliquot from each sample should be assessed for quality by gel electrophoresis or bioanalyzer.

DNA FRAGMENTATION

BASIC PROTOCOL 1

Genomic DNA must be fragmented in order to capture and sequence exons. First, because the goal is to sequence only exons and as little background genome as possible, DNA must be fragmented to a size that allows maximum sequence coverage of targeted exons with minimal sequencing of neighboring intronic regions. Because exons average ~160 bp in length, shearing DNA to a mean length of ~150 bp enables the efficient capture and sequencing of these small target regions. Second, for optimal clonal cluster amplification on the flow cell, DNA fragments should range from 200 to 500 bp in length. With a tight fragment size distribution, uniformly sized clusters are more easily differentiated from one another on the GAII, ultimately increasing sequence yields.

Several methods for randomly shearing DNA are in use today, including nebulization using compressed air, sonication, and hydro-shearing. These methods typically produce a wide size distribution and often require the use of a preparative gel and size selection to obtain the tight size distribution preferred for exon hybrid selection. To eliminate material loss and the time-consuming process of gel-based size selection, we routinely use a more recently developed DNA shearing technology called Adaptive Focused Acoustics (AFA). The Covaris S-series Sample Prep Station is highly adjustable and allows genomic

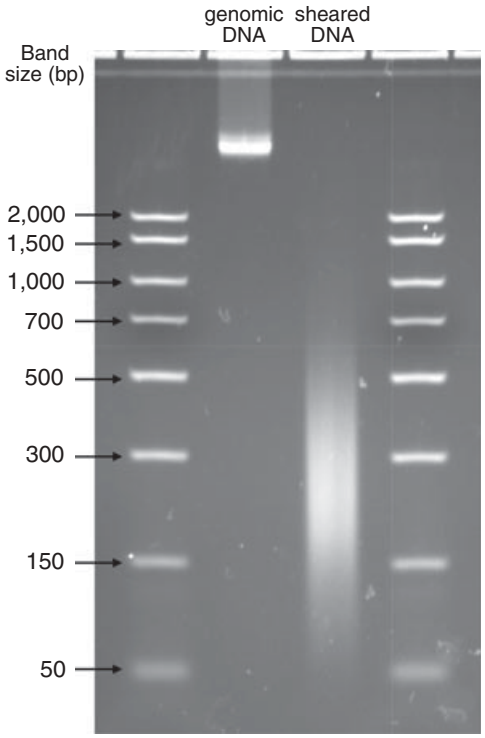

Figure 18.4.2 Sheared genomic DNA size distribution. High-quality genomic DNA was sheared using the Covaris instrument. Unsheared gDNA (100 ng) and sheared DNA (200 ng) were run in parallel on a 2% agarose gel. After shearing, the bulk of the fragments should run between ~100 and 400 bp.

DNA to be sheared into a tight band averaging 150 bp with a distribution of ~100 to 400 bp (Fig. 18.4.2). The Covaris instrument uses adjustable acoustic energy that is focused into a glass vial containing the diluted DNA samples. The focused energy creates tiny bubbles that constantly collapse in a process called cavitation, which shears the DNA. By adjusting the energy level and the exposure time, genomic DNA can be sheared to many size distributions.

Materials

DNA sample (e.g., see *APPENDIX 3B*)
Nuclease-free water
70% (v/v) ethanol

NanoDrop ND-1000 spectrophotometer
Covaris S-2 Sample Preparation System
VWR circulating chiller
Covaris shearing vial (6 × 16−mm AFA fiber vial; cat. no. 520045)
1.5-ml microcentrifuge tube
Agencourt AMPure XP kit (Beckman Coulter, cat. no. A63881)
Magnetic separator (DynaMag Spin Magnet, Invitrogen, cat. no. 123-20D)

Additional reagents and equipment for DNA quantitation (*APPENDIX 3D*) and agarose gel electrophoresis (*UNIT 2.7*)

Dilute DNA sample

1. Prepare a dilution of DNA sample at a concentration of 3 µg in 100 µl nuclease-free water (~30 ng/µl).

2. Confirm DNA concentration by absorption at 260 nm on a Nanodrop ND-1000 spectrophotometer (*APPENDIX 3D*), using nuclease-free water to blank the instrument.

Shear DNA

3. Fill a Covaris water bath to the fill line, adjust circulating chiller bath to 4°C, and begin degassing. Allow system to chill and degas for 20 min or more.
4. Pipet 100 µl DNA sample through the split septum cap of a shearing vial, insert vial into holder, and place holder into position.
5. Adjust shearing parameters and run program as follows:

Duty cycle	10%
Intensity	5%
Cycler/burst	200
Mode	frequency sweeping
# Cycles	3.

6. Pipet the sheared sample from the vial to a clean 1.5-ml microcentrifuge tube.

Clean DNA using AMPure XP beads

7. Allow AMPure beads to equilibrate to room temperature ~20 min.

   For additional information about using AMPure beads, see manufacturer's instructions.

8. Gently shake the bottle to resuspend any beads that may have settled and ensure that mixture is homogeneous.
9. Slowly add 1.8× volume (180 µl) of beads to the sheared DNA.
10. Vortex bead/reaction mixture for 10 sec or until the mixture is homogeneous. Incubate 5 min at room temperature.
11. Place the tube on a magnetic separator and allow the beads to separate out of solution for 2 min until the solution appears clear.
12. With the tube still on the magnet, slowly pipet off and discard the supernatant.
13. Gently pipet 500 µl of 70% ethanol into the tube, being careful not to disturb beads. Let stand 30 sec and then remove and discard the ethanol wash.
14. Repeat wash, being sure to remove all ethanol after the second wash.
15. With tube still on the magnet, allow beads to air dry for 2 min.

    Do not allow beads to over dry and appear cracked, as this will greatly reduce DNA recovery.

Elute DNA

16. Remove tube from magnet and add 32 µl nuclease-free water to elute DNA.
17. Briefly vortex to ensure all beads come in contact with eluant.
18. Place tube on magnetic separator and allow beads to separate for 1 min until liquid is clear.
19. Carefully pipet the eluate to a new labeled tube. Store at −20°C until end repair step.

Check DNA fragment size

20. Run 2 µl of eluate on a 2% agarose gel to ensure correct fragment distribution.

    The smear should be from ~100 to 400 bp with a peak around 150 to 200 bp.

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.