Fisher et al. Genome Biology 2011, 12:R1
http://genomebiology.com/2011/12/1/R1

Genome Biology

METHOD Open Access

A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries

Sheila Fisher[1], Andrew Barry[1], Justin Abreu[1], Brian Minie[1], Jillian Nolan[1], Toni M Delorey[1], Geneva Young[1], Timothy J Fennell[1], Alexander Allen[1], Lauren Ambrogio[1], Aaron M Berlin[2], Brendan Blumenstiel[3], Kristian Cibulskis[3], Dennis Friedrich[1], Ryan Johnson[1], Frank Juhn[4], Brian Reilly[1], Ramy Shammas[1], John Stalker[1], Sean M Sykes[2], Jon Thompson[1], John Walsh[1], Andrew Zimmer[1], Zac Zwirko[1,4], Stacey Gabriel[2], Robert Nicol[1], Chad Nusbaum[2*]

Abstract

Genome targeting methods enable cost-effective capture of specific subsets of the genome for sequencing. We present here an automated, highly scalable method for carrying out the Solution Hybrid Selection capture approach that provides a dramatic increase in scale and throughput of sequence-ready libraries produced. Significant process improvements and a series of in-process quality control checkpoints are also added. These process improvements can also be used in a manual version of the protocol.

Background

The cost of DNA sequencing continues to fall, driven by ongoing innovation in sequencing technology [1-4]. As a result, it has recently become feasible to sequence non-trivial numbers of whole human genomes [3,5-10]. Many more such projects are planned and commercial genome sequencing services are now becoming available [11,12]. At the same time, there is growing interest in sequencing specific portions of genomes, and several affordable methods for sample preparation of targeted regions have been recently published [13-17]. Key applications for targeted approaches include sequencing of exons or sets of protein-coding genes implicated in specific diseases [18-21], whole human exome sequencing (for example, in cancer or disease cohorts) [22-24] (reviewed in [25]), and resequencing of specific regions as a follow-up to genome-wide association studies [26]. The economics of whole exome sequencing have made targeted enrichment approaches an attractive option for discovery of rare mutations in a variety of diseases as the price tag is substantially lower than for sequencing an entire human genome. For example, using list prices and including the targeted capture step, the all-in cost of sequencing a whole exome (roughly 30 Mb), is 13-fold less than for the whole genome (Table S1a in Additional file 1). This translates directly into a budget that can include more than ten times as many samples, greatly increasing the statistical power of the data to be generated. The effect is even greater for smaller sequencing targets, which further scale down the required sequencing, although costs of targeting scale down more slowly. Ultimately, as long as the expense of the required sample preparation does not dominate, targeting will continue to be a cost-effective approach. To date, however, no targeting method has been described that can handle the many thousands of samples that are becoming available. To fill this need, we set out to develop such a method.

Solution hybrid selection (SHS), developed by Gnirke *et al.* [14], was created as a tool to cheaply and effectively target multiple regions in the genome in a way that is compatible with next generation sequencing technologies (Figure 1). The published protocol performs well in terms of efficiency of enrichment (selectivity), reproducibility, evenness of coverage, and sensitivity to detect single-base changes [14]. Using this method, a single technician can process six samples simultaneously from genomic DNA to sequence-ready library in

* Correspondence: chad@broadinstitute.org
[2]Genome Sequencing and Analysis Program, Broad Institute of MIT and Harvard, 320 Charles Street, Cambridge, MA 02141, USA
Full list of author information is available at the end of the article

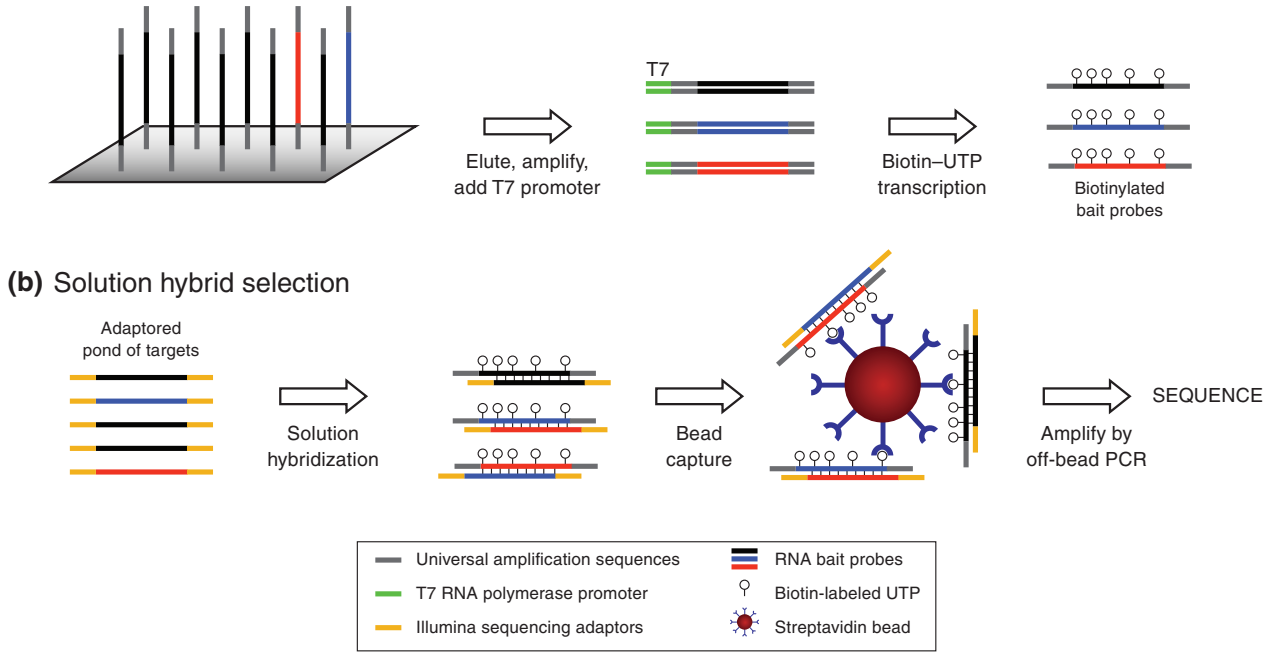

Figure 1 Overview of the hybrid selection method. Two specific sequencing targets and their respective capture baits are indicated in blue and red. **(a)** Generation of RNA bait capture probes. 150mer oligos are synthesized on array in batches of 55,000 and cleaved off. They are made double stranded by PCR amplification and tailed with a T7 RNA polymerase promoter, and RNA capture baits are made by transcription in the presence of biotinylated UTP. **(b)** Solution hybrid selection. RNA baits (from the top line) are mixed with a size selected pond library of fragments modified with sequencing adaptors. Hybridized fragments are then captured to streptavidin beads and eluted by the with-bead protocol for sequencing. See text for details.

approximately one week. This process was designed purely as a series of liquid handling steps and incubations, with the specific intention of making it amenable to scale up and automation. Given the demonstrated success of this and other methods, demand for targeted sequencing has increased sharply. To accommodate the increased demand, keep costs down, and limit the requirements for human labor, we have adapted SHS to an automated high-throughput process. This SHS method includes improvements designed to increase the efficiency of the target selection process through optimization of reactions and automation of the library and capture procedures using liquid handling robots. Several aspects of this method, in particular the 'with-bead' sample preparation method, are amenable to sample preparation steps for a range of next generation sequencing applications, including alternative in-solution and solid-phase capture strategies.

To support high-throughput SHS for targeted sequencing, we set out to devise a laboratory process that would handle very large numbers of samples in parallel for targeting and preparation of sequence-ready libraries at a low cost per sample. This process was designed to carry out whole exome targeting but also yields good results in targeting subsets of genes or regions for resequencing. Results described here come from whole exome targeting using the Agilent SureSelect Human All Exon v2 kit, which is a commercially available implementation of the optimized capture reagent we have described previously [14].

A number of challenges were overcome in developing a robust, automated, and highly scalable process for selection of exomes and other targets. Beyond the need for processing large numbers of samples, modifications of the protocol were made to achieve or maintain the following: elimination of manual, agarose gel-based size selection, which has now been replaced by fully automated, bead-based steps; high selectivity, with a high number of sequenced bases on or near the target region of interest; evenness of sequence coverage among captured targets, avoiding highly overrepresented targets and dropouts; high library complexity, or low molecular duplication, so that libraries contain large numbers of unique genome fragments; reproducibility, so that performance of the process is highly predictable; low cost of the targeting process relative to sequencing; detailed process tracking to reduce errors and provide sample history; quality control checkpoints built into the

process to identify poor performers prior to sequencing; and limited human labor.

We present here a scalable, automated SHS method that operates at a throughput far higher than achieved by other methods. The process can also be carried out by hand using a multichannel pipetter. This method has not only been scaled but also optimized to improve selectivity and evenness of target coverage and to minimize artifactual duplication to consistently deliver greater than 94% of the alignable exome (Additional file 2). The automated protocol has a capacity to process over 1,200 SHS samples in less than a week with four technicians (one technician can generate 1,200 pond libraries per week, and three technicians can each generate 384 SHS captures per week). For ease of explanation, we employ a fishing-based terminology in SHS, where the biotinylated RNA capture reagent is referred to as the 'bait', the genomic DNA library from which targets are captured as the 'pond' in which we are 'fishing', and the DNA targets from the pond that are captured by the bait are referred to as the 'catch'.

Results

Building a high-throughput solution hybrid selection process

SHS is a method used to selectively enrich for regions of interest within the human genome [14] (Figure 1). Briefly, a library (or 'pond') of adapter-ligated fragments of randomly sheared DNA is hybridized to biotinylated RNA (or 'baits') that are complementary to the target sequences. Hybridized molecules (the 'catch') are then captured using streptavidin-coated beads. Once the captured DNA fragments are PCR amplified off the capture reagent, they are available to be sequenced using next generation sequencing technologies. The standard SHS protocol was redesigned from a manual, bench scale process to an automated process, in much the same way as our recent work to scale library construction for 454 sequencing [27], and is capable of far greater throughput than demonstrated for other methods (Additional file 2). A series of process innovations were required to facilitate reimplementation of this process at large scale. In particular, all manual pipetting steps were converted to automation-amenable liquid handling steps, and these liquid handling steps were extensively optimized to maximize yield efficiency. As part of this, the electrophoretic size selection step has been replaced by fully automated bead-based sizing. Other optimizations are described below. Table 1 shows a comparison of the original published method and the new protocol with a description of each step and the improvements in the new method. Table 2 describes a set of key sequencing metrics by which we measure SHS process performance.

The automated SHS process is implemented on the Bravo liquid handling workstation (Agilent Automation Solutions), a commercially available small-footprint, liquid handling platform, but can be implemented on many commercially available liquid handlers. The process can also be carried out manually using a multichannel pipette. An overview map of the process can be found in Additional file 3 and the manual protocol version can be found in Additional file 4.

Optimization of acoustic shearing

The process begins with fragmentation of genomic DNA using the Covaris E210 adaptive focused acoustics instrument. Maximizing the yield of DNA fragments in the desired size range is a key step in minimizing overall sample loss. The Covaris E210 instrument focuses acoustic energy into a small, localized zone to create cavitation, thereby producing breaks in double-stranded DNA. A number of variables control mean fragment length and distribution, including duty cycle, cycles per burst, and time. The Covaris adaptive focused acoustics system has several advantages over other methods such as nebulization or hydrodynamic force. First, DNA is sheared in a small closed environment and is not handled in large volume vessels or in tubing, greatly reducing sample loss. Second, the closed, independent vessels greatly reduce sample cross-contamination. Third, the Covaris machine can operate automatically on up to 96 samples per run, eliminating significant sample handling labor and eliminating shearing as a process bottleneck. Fourth, improvements to the shearing protocol in combination with removal of small fragments in subsequent bead-based clean up steps (see below) eliminates the need to size select and extract samples from agarose gels, a critical bottleneck in the overall process.

Shearing performance was extensively optimized for increased sample yield, narrower insert size distribution, and robust and reproducible handling of large numbers of samples in parallel. Optimizations focused on the following factors: shearing volume, tube type, elimination of tube breakage, shearing pulse time, water degassing, and positioning of tubes in the water bath (see Materials and methods for details). In order to accommodate automated handling of the samples, volumes were reduced from 100 μl to 50 μl without any effect on shearing profiles or sample loss (Additional file 5). Importantly, proper fit of the shearing rack (Covaris, catalogue number 500111) into custom adapters (see Additional file 6 for CAD drawing) prevents movement, allowing transfers to occur via automated liquid handling. In addition, specific tubes available from Covaris (Covaris, catalogue number 500114) virtually eliminated the problem of tube breakage. Only a single sample in

Table 1 Comparison of standard versus improved solution hybrid selection methods

	Manual standard SHS protocol		Automated improved SHS protocol	
Process step	**Standard method**	**Drawbacks**	**Improved method**	**Advantages**
Shearing of genomic DNA	Covaris S2	Single sample	Optimized Covaris E210	Multi-sample, improved yield, tight size range
Enzymatic cleanups	Individual spin columns	Low throughput, 50 to 60% recovery, manual	'With-bead' SPRI	High throughput, 80 to 90% recovery, automated
Solution hybrid selection capture	Manual, column-based	Labor intensive (6 samples/FTE/week)	Fully automated	Walkaway, high throughput (1,200 samples/4FTE/week)
Final PCR enrichment	Denature, followed by PCR	Sample loss through transfers	Direct 'off-bead' PCR	Improved final yield
In process quality control checkpoints	Agilent Bioanalyzer	Limited visibility until sequence results	Many	In process results: key predictors of sample, library and sequencing quality

FTE, full time employee; SHS, solution hybrid selection; SPRI, solid phase reversible immobilization.

the most recent 5,000 processed suffered a broken tube. Through a systematically designed and controlled set of experiments, optimal pulse time parameters were chosen to provide a mean fragment length of 150 bp with a range of 75 to 300 bp (Materials and methods). Additional file 5 shows the contrast between unoptimized and optimized size profiles of sheared DNA. In addition to regular maintenance, careful degassing of the water bath and proper water levels are critical for reproducible results. In a nondegassed water bath dissolved oxygen reduces cavitation and disperses energy, reducing shearing efficiency.

Modified bead-based cleanups enable scale-up to 96 wells

A key requirement in scaling SHS was to implement processing of samples in a standard 96-well microtiter plate. This was facilitated by development of a novel modification to solid-phase reversible immobilization (SPRI) magnetic bead reaction cleanup methodology [27,28] we have termed 'with-bead' SPRI (Figure 2), which is highly scalable due to its amenability to liquid handling automation. Implementation of with-bead SPRI in SHS offers significant advantages. First, it replaces single tube spin-column-based cleanups with liquid handling-compatible magnetic bead-based cleanups; second, it enables selection of molecular weight ranges, eliminating the need for agarose gel-based sizing; third, it simplifies the process by allowing elimination or combining of several steps, which results in a higher overall DNA yield.

Table 2 Automated solution hybrid selection performance

Performance factor	3 μg input average (*n* = 1,117 exomes)
Median target coverage	131.0×
Percentage bases > 2×	96.0%
Percentage bases > 10×	91.9%
Percentage bases > 20×	87.6%
Percentage selected bases (on target)	83.7%
Percentage duplicated reads	4.4%
Fold 80 penalty[a]	3.17
Estimated library size of captured fragments	278 million

See Additional file 12 for metric definitions. [a]Fold 80 penalty is a measure of the non-uniformity of sequence coverage, defined as the amount of additional coverage (in fold coverage of the genome) required so that 80% of the target bases will be covered at the current mean coverage (see Additional file 12 for details).

The innovation of the with-bead SPRI method is as follows. Rather than employing a series of discrete cleanup steps in the library construction process, the cleanups are effectively integrated. The SPRI beads are added to the sample after the shearing step, and remain in the reaction vessel throughout the sample preparation protocol. By allowing each cleanup step to employ the same beads, the with-bead method greatly reduces the number of liquid transfer steps required. The 'cleaned up' DNA is then eluted at the conclusion of the process. This methodology increases the overall DNA yield (Figure 3), primarily because it allowed us to eliminate six of the ten sample transfer steps, avoiding the loss of DNA sticking to the sides of the vessel or loss of volume in pipetting. Briefly, following each process step, DNA is selectively bound to the iron beads, already present, through the addition of a 20% polyethylene glycol (PEG), 2.5 M NaCl buffer. The mixture is placed on a magnet, which pulls the beads and bound DNA to the sides of the well so that the reagents, washes and/or unwanted fragments can be removed with the supernatant. Molecular weight exclusion, which is essentially a size selection, of unwanted lower molecular weight DNA fragments can be controlled through the volume of the PEG NaCl buffer that is added to the reaction, changing the final concentration of PEG in the resulting mixture and altering the size range of fragments bound to the beads [27,28]. DNA fragments that have been cleaned or size selected are eluted from the beads, ready

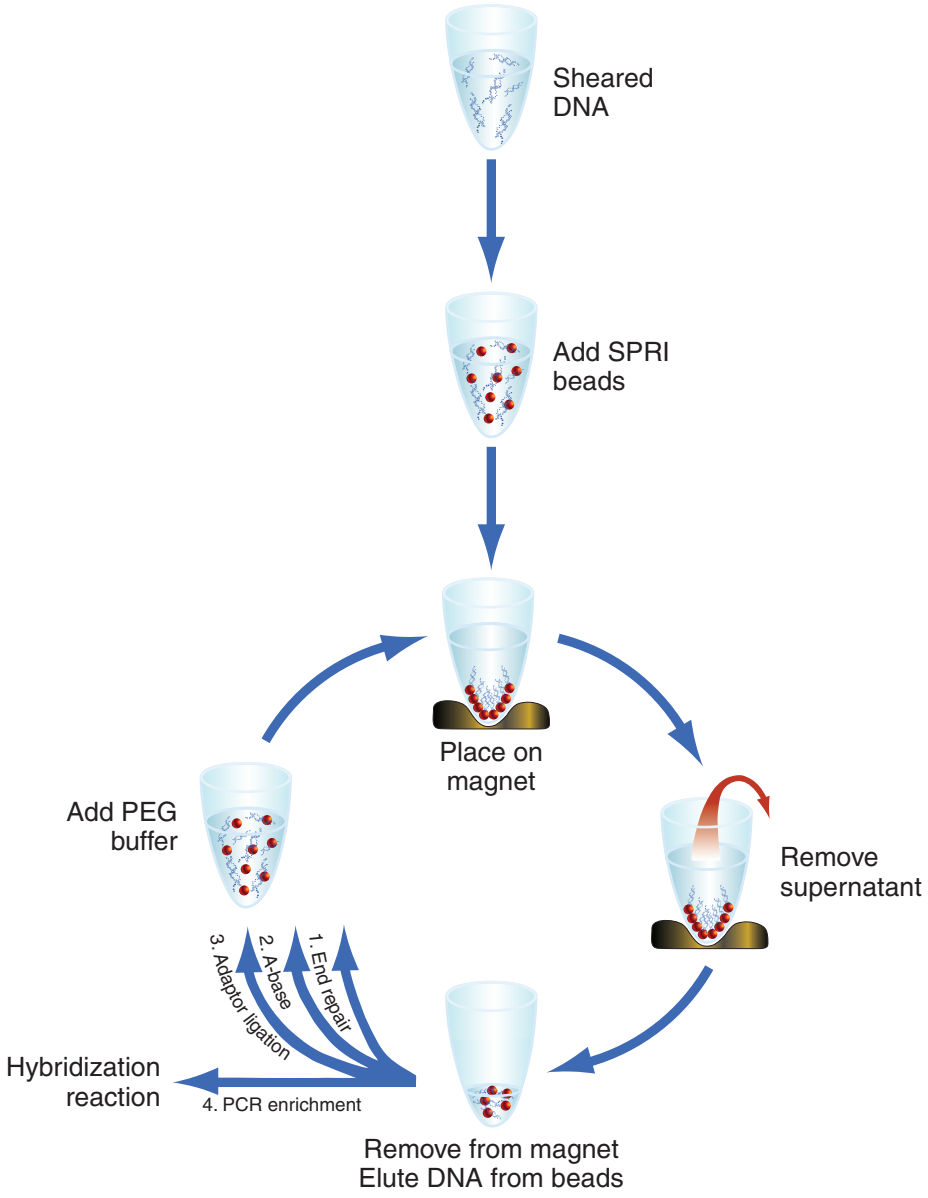

Figure 2 With-bead SPRI method for pond library construction. SPRI magnetic beads are added to the sheared DNA sample. DNA is selectively bound to SPRI beads, which are immobilized when the sample plate is placed on a magnet, leaving other molecules in the liquid phase. The liquid phase is removed and discarded. The sample plate is then removed from the magnet and DNA is eluted from the beads. Library construction master mixes are then added to eluant/bead solution. The DNA and SPRI beads then pass through three cycles of reaction, binding to beads (in the presence of polyethylene glycol (PEG)/NaCl solution) and cleanup/washing. The cycles carry out end repair, A-base addition and adaptor ligation, respectively. A final elution is then followed by PCR amplification.

for the next step; however, the eluate is not transferred into a new reaction vessel. Rather, the reagents for the next step are added directly to the reaction vessel containing samples and beads. The presence of beads does not interfere with any of the steps in the process (Table 3). This with-bead protocol has greatly increased the number of unique fragments entering the pond PCR step, increasing the complexity of libraries made by roughly 12-fold (Table 3).

This increase in yield with the with-bead SPRI protocol has the added benefits of reducing both the input DNA requirement to the process and the number of PCR cycles required. Efficient with-bead targeted captures can be achieved with pond libraries made with as little as 100 ng of input DNA and six to eight cycles of PCR, a major improvement over the commercialized SHS method, which requires 3 μg of starting genomic DNA and 14 cycles (Table 3). We note here that PCR

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.