



Published in final edited form as:

*Curr Protoc Hum Genet.* 2010 July ; CHAPTER: Unit–18.3. doi:10.1002/0471142905.hg1803s66.

## Targeted Enrichment of Specific Regions in the Human Genome by Array Hybridization

Catherine Igartua<sup>1</sup>, Emily H. Turner<sup>1</sup>, Sarah B. Ng<sup>1</sup>, Emily Hodges<sup>2,3</sup>, Gregory J. Hannon<sup>2,3</sup>, Arindam Bhattacharjee<sup>4</sup>, Mark J. Rieder<sup>1</sup>, Deborah A. Nickerson<sup>1</sup>, and Jay Shendure<sup>1</sup>

<sup>1</sup> Department of Genome Sciences, University of Washington, 1705 NE Pacific St, Seattle, WA 98195

<sup>2</sup> Watson School of Biological Sciences, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724

<sup>3</sup> Howard Hughes Medical Institute, Cold Spring Harbor Laboratory, 1 Bungtown Road, Cold Spring Harbor, New York 11724

<sup>4</sup> Agilent Technologies, Inc. 5301 Stevens Creek Boulevard, Santa Clara, California 95051

### Abstract

While whole genome resequencing remains expensive, genomic partitioning provides an affordable means of targeting sequence efforts towards regions of high interest. There are several competitive methods for targeted capture; these include molecular inversion probes, microdroplet-segregated multiplex PCR, and on-array or in-solution capture-by-hybridization. Enrichment of the human exome by array hybridization has been successfully applied to pinpoint the causative allele of Mendelian disorders. This protocol focuses on the application of Agilent IM arrays for capture-by-hybridization and sequencing on the Illumina platform, though the library preparation method may be adaptable to other vendor's array platforms and sequencing technologies.

### Keywords

Resequencing; exome; hybridization; targeted enrichment

### Unit Introduction

The identification of sequence-level variation in individuals is of immense utility in genetics, and is of particular interest in the discovery of disease genes and specific mutant alleles. Generating complete genome sequences is increasingly feasible on a large-scale with the introduction of “second-generation” sequencers, which are able to generate sequence data in a cost-effective, massively parallel fashion. However, although the cost of sequencing has dropped dramatically, it is still not low enough for whole-genome resequencing to be done routinely by many laboratories, even on a small-scale. However, for many projects, there is significant interest using new sequencing platforms to perform variation discovery in targeted subsets of the genome. A variety of methods to accomplish “genome partitioning” or “targeted capture” have been described (9), and include several that are based on hybridization (1–6) and others that include enzymatic activity to confer specificity (7,8).

Here we describe the protocol that we have applied to enrich for all protein-coding regions in the human genome (the “exome”), using hybridization on microarrays. The basic protocol has two main components: construction of a shotgun sequencing library, and enrichment of that library for sequences of interest via hybridization to a programmable microarray.

## Strategic Planning

This protocol largely relies on standard molecular biology techniques; however, special attention is required to limit cross-contamination between samples. Large amounts (~20 µg) of shotgun library are produced for each sample, which can easily contaminate gloves and workspaces. If samples are being prepared in parallel, careful segregation of different sample libraries is required.

The flowchart shown in Figure 1 provides a suggested daily schedule for this protocol. Several breakpoints are indicated in the main protocol to aid in scheduling.

## Basic Protocol Materials List

### Solutions and Reagents

5 µg of genomic DNA

1X Tris-EDTA, pH 8.0 (e.g., Fisher BioReagents, BP2473-1)

10X Next End-Repair Buffer (NEB)

Next End-Repair Enzyme Mix (NEB)

Agencourt AMPure XP

10X PCR buffer w/15mM MgCl<sub>2</sub> (Applied Biosystems N8080160)

dATP (100mM)

Taq (5U/µl)

T4 ligase (e.g., Enzymatics L603-HC-L)

10X T4 ligase buffer (e.g., Enzymatics L603-HC-L)

Certified low-range ultra agarose (e.g., Biorad)

10 mg/ml ethidium bromide

1X TAE buffer (0.04M Tris-acetate pH 8, 1mM EDTA)

5X gel loading buffer (e.g., QIAgen 239901)

100 bp DNA ladder (e.g., NEB N3231)

Qiaquick Gel Extraction kit (Qiagen)

Hot Start Phusion polymerase (2U/µl) (Finnzymes)

5X Phusion buffer (Finnzymes)

25 mM dNTPs

Qiaquick PCR Purification kit (Qiagen)  
10X Oligo aCGH/ChIP-on-Chip Blocking Agent (Agilent)  
2X Hi-RPM Hybridization Buffer (Agilent)  
Cot1 human DNA (1mg/ml)  
DNA 1000 Bioanalyzer kit (Agilent)  
DNA High Sensitivity Bioanalyzer kit (Agilent)  
Oligo aCGH/ChIP-on-chip Hybridization Kit (Agilent)

### Special Consumables and Equipment

1M DNA Capture Microarray (Agilent, PN: G3358A, AMADID: 027616).  
DNA Microarray Hybridization Chamber (Agilent)  
DNA Microarray Hybridization Oven (Agilent)  
Secure Seal Tabs (Grace Bio Labs)

## Basic Protocol: Genomic Enrichment by Array Hybridization

### Design of Capture Arrays

Capture arrays can be designed against any set of genomic subsequences of interest, including the entire exome, with the caveat that repetitive regions of the genome are likely to be problematic. Programmable arrays are commercially available, e.g. arrays with densities as high as 1 million features per array and individual oligo lengths of up to 60 bp can be ordered from Agilent. Guidelines that have been successful for us include the following. In general, hybridization probes should be designed to be evenly spaced across each target region. Other general probe design constraints include: (1) to be relatively unique, such that the average occurrence of each 15-mer in the probe sequence is less than  $10^{-8}$ , (2) to be between 20 and 60 bases in length, with preference for longer probes, and (3) to have a calculated melting temperature ( $T_m$ )  $\geq 69$  °C, with preference for higher  $T_m$  values. The protocol described here was evaluated on array designs that followed these guidelines.

### Prepare Working Solutions of Primers, Adaptors and Blocking Oligos

#### Protocol

1. Resuspend the oligos listed in Table 1 with 1x TE buffer to a final concentration of 100  $\mu$ M.
2. For oligos EXOME\_BULK\_FWD (100 $\mu$ M) and EXOME\_BULK\_REV (100 $\mu$ M) do a 1:10 dilution to create a 10 $\mu$ M in 100  $\mu$ l stock of each oligo.
3. To prepare a 50  $\mu$ M stock solution of the paired-end adaptor, combine equal volumes of Exome\_Adapt\_Hi (100  $\mu$ M) and Exome\_Adapt\_Lo (100  $\mu$ M) in a strip tube.
4. Denature the mixture in a thermal cycler at 95°C for 5 minutes. Turn the thermal cycler off and allow the mixture to anneal while slowly cooling to room temperature (30–45 minutes).

## Fragment Genomic DNA with Covaris

**Materials required**—Covaris (E-series or S-series)

Covaris microtubes

1x TE buffer

5 µg input genomic DNA

**Protocol**—Each sample requires 5 µg of high quality gDNA as input. Check the concentration of the input sample DNA by Picogreen or Nanodrop assays.

1. Dilute 5 µg of gDNA in total volume of 85 µL in 1x TE buffer.
2. Load samples into Covaris tubes, recording its positions in the Covaris tube plate. Ensure there are no air bubbles in the bottom of the tubes after loading the samples. In the case of air bubbles, gently tap the plate to bring the solution to the bottom of the tube.
3. Shear the samples on the Covaris under the following settings:  
Duty Cyclor: 20% Intensity: 5 Cycles/Burst: 200 Time: 180 seconds
4. Transfer the samples to a half skirted 300 µL 96 well plate and centrifuge. There is no cleanup after this step.

## End Repair Fragmented DNA

**Materials required**—10X Next End-Repair Buffer (NEB)

Next End-Repair Enzyme Mix (NEB)

Agencourt AMPure XP

Thermal cycler

Magnetic plate

½ skirted 300 µl PCR plate

### Protocol

1. Prepare a master mix based on Table 2, scaling the volumes for the number of samples being prepared.
2. Aliquot 15 µL of master mix into each well containing 85 µL of fragmented DNA, mix and spin down. Each well will contain 100 µL total volume.
3. React on thermal cycler for 30 minutes at 20°C.
4. Transfer 180 µL of room temperature AMPure beads to each well containing end-repaired DNA sample, mix well and incubate for 5 minutes at room temperature.
5. Place the reaction plate onto a magnet plate for 2 minutes to separate beads from the solution. Visually confirm that the beads have moved to the side of the plate. Aspirate the 280 µL of clear solution from the reaction plate and discard.
6. Dispense 200 µl of 70% ethanol to each well of the reaction plate and incubate for 30 seconds at room temperature. Aspirate out the ethanol and discard. Repeat for a total of two washes.

7. Take the reaction plate off the magnet plate, add 45  $\mu\text{L}$  of ultrapure water to each well and mix. Place the reaction plate onto the magnet plate for 1 minute to separate the beads. Transfer the 45  $\mu\text{L}$  of supernatant to a new standard 96 well plate.
8. The DNA is now sheared, end repaired and purified. This is a good stopping point if you need to pause the experiment. Store at  $-20^{\circ}\text{C}$ .

### A-Tail DNA

#### Materials required—10X PCR buffer w/15mM $\text{MgCl}_2$

dATP (100mM)

Taq (5U/ $\mu\text{L}$ )

Agencourt AMPure XP

Thermal cycler

Magnetic plate

#### Protocol

1. Prepare a master mix based on Table 3, scaling the volumes for the number of samples being prepared.
2. Aliquot 6  $\mu\text{L}$  of master mix to each well containing end-repaired DNA, mix and spin down. The final volume will be 50  $\mu\text{L}$  in each well.
3. React on the thermal cycler at  $70^{\circ}\text{C}$  for 20 minutes.
4. Transfer 90  $\mu\text{L}$  of room temperature AMPure beads to each well containing the A-tailed DNA sample, mix well and incubate for 5 minutes at room temperature.
5. Place the reaction plate onto a magnet plate for 2 minutes to separate beads from the solution. Visually confirm that the beads have moved to the side of the plate. Aspirate the 140  $\mu\text{L}$  of clear solution from the reaction plate and discard.
6. Dispense 200  $\mu\text{L}$  of 70% ethanol to each well of the reaction plate and incubate for 30 seconds at room temperature. Aspirate out the ethanol and discard. Repeat for a total of two washes.
7. Take the reaction plate off the magnet plate, add 37  $\mu\text{L}$  of ultrapure water to each well and mix. Place the reaction plate onto the magnet plate for 1 minute to separate the beads. Transfer the 37  $\mu\text{L}$  of supernatant to a new standard 96 well PCR plate.
8. Perform a picogreen assay to quantify DNA. Samples must be at least 20 ng/ $\mu\text{L}$ .

### Ligate Paired-end Adaptors to DNA

#### Materials required—T4 ligase (e.g. NEB, Enzymatics)

10X T4 ligase buffer (e.g. NEB, Enzymatics)

Annealed paired-end adaptor (50 $\mu\text{M}$ )

Agencourt AMPure XP

Thermal cycler

Magnetic plate

### Protocol

1. Prepare a master mix based on Table 4, scaling the volumes for the number of samples being prepared.
2. Aliquot 15  $\mu\text{L}$  of master mix to each well containing A-tailed sample, mix and spin down. Each well will contain 50  $\mu\text{L}$ .
3. React at room temperature for 15 minutes.
4. Transfer 90  $\mu\text{L}$  of room-temperature AMPure beads to each well containing the adaptor ligated DNA sample, mix well and incubate for 5 minutes at room temperature.
5. Place the reaction plate onto a magnet plate for 2 minutes to separate beads from the solution. Visually confirm that the beads have moved to the side of the plate. Aspirate the 140  $\mu\text{L}$  of clear solution from the reaction plate and discard.
6. Dispense 200  $\mu\text{L}$  of 70% ethanol to each well of the reaction plate and incubate for 30 seconds at room temperature. Aspirate out the ethanol and discard. Repeat for a total of two washes.
7. Take the reaction plate off the magnet plate, add 32  $\mu\text{L}$  of ultrapure water to each well and mix. Place the reaction plate onto the magnet plate for 1 minute to separate the beads. Transfer the 32  $\mu\text{L}$  of supernatant to a new standard 96 well plate.

The DNA is now sheared, end repaired, A-tailed, adaptor ligated and purified. This is a good stopping point if you need to pause the experiment. Store at  $-20^{\circ}\text{C}$ .

### Size Select 250–350bp DNA

**Special considerations**—Size selection is optional. Performing size selection will result in a narrower insert size range, while performing a gel-free protocol will produce a more complex final library.

**Materials required**—Certified low-range ultra agarose (e.g. Biorad)

Comb: 1mm(d)  $\times$  mm (w)  $\times$  10mm (h) (to accommodate 40 $\mu\text{L}$  sample/well)

10 mg/ml ethidium bromide

1x TAE buffer (0.04M Tris-acetate pH 8, 1mM EDTA)

5X gel loading buffer

100 bp DNA ladder

UV light box

Microcentrifuge

Razor blades

15mL Tubes

Clear plastic film

Isopropanol

Qiaquick Gel Extraction kit (Qiagen)

1. Prepare a 2% low melting agarose gel in a final volume of 200 ml of 1x TAE buffer. (200 ml TAE, 4 g Agarose. Four samples can fit on one gel.)
2. Microwave for about 2 minutes or until agarose has fully dissolved.
3. Add 8  $\mu$ l of EtBr (0.4  $\mu$ g/ml final concentration), pour gel onto the casting tray and place the wide comb in the upper section.
4. Once the gel has solidified, place the gel onto the gel box and pour fresh 1x TAE buffer until the gel is covered.
5. Dispense 10  $\mu$ l of 4x loading buffer to each 30  $\mu$ l of DNA sample.
6. For each sample, load 40  $\mu$ l of sample, skip a well, then load ladder. Make sure the samples are loaded asymmetrically on the gel so you can distinguish left from right on the gel image.
7. Run electrophoresis at 100 V for 120 minutes.
8. Place agarose gel on saran wrap and view the gel on the UV light box.
9. For each sample, using the 100 bp ladder as reference, cut the region of interest between 250–350bp. Use a clean razor blade while cutting each sample and place each slice in a separate 15ml tube.
10. Weigh each gel slice and add 4x the volume of Qiagen Buffer QG to the sample (e.g. 400 mg slice, add 1600  $\mu$ l Buffer QG). Allow the gel to melt at room temperature for 15 minutes vortexing frequently. Ensure the gel is fully dissolved before proceeding.
11. While gel is melting, heat 120  $\mu$ l of Qiagen Buffer EB per sample to 65°C.
12. Add 1 gel volume of isopropanol to the gel solution and mix (e.g. 400mg slice, 400  $\mu$ l isopropanol).
13. If the slice weighed more than 400 mg, split the solution across two columns. If the volume is greater than 750  $\mu$ l, perform multiple centrifugations. Load the solution to the Qiagen column and centrifuge for 1 minute.
14. Discard flow-through and add 500  $\mu$ l Buffer QG to each column. Centrifuge 1 minute and discard flow-through.
15. Add 750  $\mu$ l Buffer PE to each column. Let the column stand for 2 minutes before centrifuging. Centrifuge 1 minute and discard flow-through. Centrifuge again for 1 minute.
16. Place the column in a clean labeled 1.5 ml micro-centrifuge tube. Add 50  $\mu$ l of Qiagen Buffer EB that has been heated to 65°C to each column. Let the column sit for 1 minute before centrifuging for 1 minute.

### Amplify shotgun library with bulk PCR

**Materials Required**—Hot Start Phusion polymerase (2U/ $\mu$ l) (Finnzymes)

5X Phusion buffer (Finnzymes)

25 mM dNTPs

EXOME\_BULK\_FWD (100 $\mu$ M)

EXOME\_BULK\_REV(100 $\mu$ M)

Qiaquick PCR Purification kit (Qiagen)

Thermal Cycler

Microcentrifuge

### Protocol

1. Dilute each size-selected sample to 100  $\mu$ L with EB.
2. Make the master mix described in Table 5, scaling as necessary for the number of samples being captured.
3. Aliquot 45.7  $\mu$ L of the master mix to the no-template control well, and add 4.3  $\mu$ L of water to that well.
4. Add 1100  $\mu$ L master mix to the 100  $\mu$ L of sample from step 1, mix and aliquot 50  $\mu$ L across 23 wells.
5. Amplify the sample in a thermal cycler with the following program:
  - a. 98°C, 0:30
  - b. 98°C, 0:10
  - c. 65°C, 0:30
  - d. 72°C, 0:60
  - e. Go to Step b, 17 cycles
  - f. 72°C, 2:00
  - g. 16°C, forever
6. Pool the reaction volumes from each sample well into a 15 ml tube. Add 6 ml of Buffer PB to each tube.
7. Aliquot 750  $\mu$ l of the sample/PB mixture into each of four Qiagen QIAquick columns. Centrifuge the column for 30 seconds and discard eluant. Repeat until all the solution has been loaded.
8. Aliquot 750  $\mu$ l Buffer PE to each column. Centrifuge for 30 seconds and discard eluant. Centrifuge for an additional 2 minutes to remove all traces of Buffer PE.
9. Transfer the tube to a clean 1.5 ml tube. Aliquot 30  $\mu$ l Buffer EB to each column. Centrifuge for 30 seconds.
10. Pool each 32  $\mu$ l aliquot from the sample for a total volume of 128  $\mu$ l.

### Quality Checkpoint – Bioanalyzer assay to confirm sample size and concentration

**Materials Required**—Agilent DNA 1000 kit

Agilent Bioanalyzer



## Protocol

1. Run the sample on the Agilent DNA 1000 bioanalyzer.

Twelve samples can be run at one time. Given that the sample's size selection was 250–350 bp, the bioanalyzer's electropherograms should have a peak that averages at 300 bp. Look for unligated adaptor bands around 50 base pairs and for overamplification humps at a size range higher than the size selection (e.g. 500bp).

2. Calculate the mass of DNA in the sample by multiplying the concentration acquired from the bioanalyzer by the total elution volume from the previous step (128  $\mu$ l). The concentration of the sample must be at least 5  $\mu$ g to continue.

If quantifying using a nanodrop instead of the bioanalyzer or a pico-green assays aim for 15–20  $\mu$ g of DNA.

## Array hybridization

**Materials Required**—EXOME\_BLOCK\_F (100  $\mu$ M)

EXOME\_BLOCK\_R (100  $\mu$ M)

EXOME\_BLOCK\_F\_rev (100  $\mu$ M)

EXOME\_BLOCK\_R\_rev (100  $\mu$ M)

10X Oligo aCGH/ChIP-on-Chip Blocking Agent (Agilent)

2X Hi-RPM Hybridization Buffer (Agilent)

Cot1 human DNA (1mg/ml)

Thermal cycler

1M Exome Agilent Microarray

Gaskets (Agilent)

Chambers (Agilent)

Oven (Agilent)

## Protocol

1. Make the master mix described in Table 6 scaling as necessary for the number of samples being captured.
2. Add a minimum of 5  $\mu$ g of template DNA in 120  $\mu$ l of water to the master mix for array hybridization. The ideal loading mass is 20  $\mu$ g of library.
3. Split the volume of 520  $\mu$ l into 4 wells of 130 $\mu$ l each.
4. Run in a thermal cycler using the following protocol:
  - a. 95°C, 3:00
  - b. 37°C, 30:00

Keep the sample in the thermal cycler at 37°C until it is ready to be loaded on the chip.

5. Assemble the gasket onto the chamber.
6. Transfer hybridizing solution to gasket pipetting sample down the middle of the gasket. Ensure that no sample spillover occurs, and pipette slowly to minimize bubbles.
7. Load the array chip to the chamber ensuring the numerical barcode is up and the Agilent label is down.
8. Assemble chamber, verify the bubble is mobile and place it in rotating oven at 65°C for 66–72 hours.

### **Wash Array and Elute Captured DNA**

#### **Materials Required**—Wash Buffer 1 (Agilent)

Wash Buffer 2 (Agilent)

15 ml tubes

Secure Seal Gaskets

Secure Seal dots

Elution buffer

Shaker

Tweezers

Heat block (with thermometer)

Water bath

#### **Protocol**

##### **Set up for each array**

1. Set heat block to 95°C and water bath to 37°C.  
If using a heat block that can accommodate tubes, partially fill the wells with water.
2. Fill 2x 1.5 ml tubes with Elution Buffer and place on heat block at 95°C.
3. Fill 2x 50ml tubes of Wash Buffer 1 (Agilent).
4. Fill 1x 50ml tube of Wash Buffer 2 (Agilent) and place on water bath at 37°C.
5. Label elution tubes. There are two elutions and each requires two 1.5 ml tubes.

##### **Array Wash**

6. Disassemble array chamber.
7. Submerge array and gasket in Agilent Wash Buffer 1. Quickly pry the array/gasket assembly apart with tweezers and transfer the array into the second tube of Agilent Wash Buffer 1. Shake the tube at room temperature for 10 minutes.
8. Move array into Wash Buffer 2 (at 37°C) for 5 minutes.

## Elution

9. Remove array from Agilent Wash Buffer 2.
10. Add Secure Chamber against the array's Agilent side, sealing the edges by sliding the array on the heat block.
11. Seal the port farthest from the Agilent label with a Secure Seal dot.
12. Fill the array with the preheated 95°C EB (approximately 800 µL, prevent leaks and bubbles).
13. Seal the second port with a Secure Seal dot (leave overhang to facilitate removal).
14. Place on heat block at 95°C for 5 minutes.
15. Remove the Secure Seal and aspirate eluant into 2 tubes (approximately 400 µL each).
16. Refill Secure Chamber, re-tab Secure Seal, and repeat Steps 10–15 for Elution 2.

## Ethanol Precipitation

### Materials Required—100% cold ethanol

70% cold ethanol

Glycogen (5mg/ml)

Ammonium acetate NH<sub>4</sub>OAc (7.5 M)

4°C Microcentrifuge

Speed Vacuum

### Protocol

1. Add the solutions described in Table 7 to each tube. Each tube will have an approximate final volume of 1.5ml.
2. Place tubes at –80°C for 90 minutes to overnight.
3. Spin in a 4°C mini-centrifuge at maximum speed for 30 minutes.
4. Remove supernatant.
5. Add 750 µl of 75% cold ethanol and spin at maximum speed for 5 minutes.
6. Remove supernatant (750 µl) and repeat Step 5.
7. Vacuum concentrate samples at 30°C for 15 minutes or until dried.
8. Resuspend pellet in 10µL EB for 30 minutes, vortex and spin down.
9. Pool both elution tubes into one (20 µL per sample/elution).
10. Freeze elution 2 as a back up sample and continue with elution 1.

## Amplify Captured DNA by Real-time PCR

**Material Required**—EXOME\_SEQ\_ FWD (10µM)

EXOME\_SEQ\_ REV (10µM)

2x iTaq Supermix

Agencourt AMPure XP

Real Time Thermal Cycler (e.g. Bio Rad)

Magnetic Plate

1. Prepare a master mix based on Table 8, scaling the volumes for the number of samples being prepared plus a no template control.
2. Aliquot 30  $\mu\text{L}$  of master mix into each a well containing 20  $\mu\text{L}$  of water for the no template control.
3. Combine the master mix with the DNA sample, and aliquot to wells. Each well will contain 50  $\mu\text{L}$  of solution.
4. Run samples in a RT-PCR thermal cycler with the following program:
  - a. 95°C, 5:00
  - b. 95°C, 0:30
  - c. 55°C, 2:00
  - d. 72°C, 2:00
  - e. Read plate
  - f. 72°C, 0:15
  - g. Go to Step b, 29 cycles
  - h. 72°C, 10:00
  - i. 16°C, forever

5. Watch the amplification profile and remove samples from the instrument when the exponential curve reaches a plateau.

This program is designed so that a sample can be taken out mid-run immediately after a plate read. Step f allows the application to pause the run to take the sample out before cycling around and denaturing the sample. This program is designed to a total of 30 cycles to allow all possible amplification scenarios. Elution 1 samples should peak around cycle 15 and elution 2 samples around cycle 20.

6. Transfer 90  $\mu\text{L}$  of room temperature AMPure beads to each well containing the RT amplified DNA sample, mix well and incubate for 5 minutes at room temperature.
7. Place the reaction plate onto a magnet plate for 2 minutes to separate beads from the solution. Visually confirm that the beads have moved to the side of the plate. Aspirate the 140  $\mu\text{L}$  of clear solution from the reaction plate and discard.
8. Dispense 200  $\mu\text{L}$  of 70% ethanol to each well of the reaction plate and incubate for 30 seconds at room temperature. Aspirate out the ethanol and discard. Repeat for a total of two washes.
9. Take the reaction plate off the magnet plate, add 32  $\mu\text{L}$  of ultrapure water to each well and mix. Place the reaction plate onto the magnet plate for 1 minute to separate the beads. Transfer the 32  $\mu\text{L}$  of supernatant to a new standard 96 well PCR plate.

## Quality Control Assay – Bioanalyzer for sample size and concentration

1. Run the sample on the Agilent DNA 1000 bioanalyzer.

Note the average base pair size, the concentration of the sample as it will be important for sequencing calculations. Also look for adaptor and overamplification artifacts.

2. Add 3  $\mu\text{L}$  1% Tween to the finished library. The sample is now ready for sequencing.

## Commentary

### Background Information

With the development of next-generation sequencing platforms, the per-base cost of sequencing has dropped by several orders of magnitude. Several different capture strategies exist that allow for enrichment of mammalian genomes for regions of interest. These methods can be broadly divided into those that rely on an enzymatic step to achieve specificity (i.e. molecular inversion probes, multiplex PCR), and those that rely purely on hybridization to oligos complementary to sequences of interest. Hybridization-based methods can be further divided into in-solution and on-array methods. Performance parameters of relevance include input DNA requirements, capture uniformity, specificity, allelic bias, maintenance of sample complexity, scalability and cost (9).

The success of this array capture protocol relies on a strategy that carries forward the entire sample. Unlike traditional molecular biology protocols, where aliquots of sample are loaded to PCR, for example, the entire reaction volume is used for all steps to maintain final library complexity. In general, this protocol could be adapted to work with alternative sequencing methods by modifying the adaptor and PCR primer sequences, and with alternative capture reagents, by adjusting hybridization conditions to reflect the volume and temperature needs of different platforms.

### Critical Parameters

**Start input DNA**—Currently 5  $\mu\text{g}$  is the lowest gDNA input recommended. If 10  $\mu\text{g}$  is available, better results may be obtained using this higher amount. The protocol for 10  $\mu\text{g}$  would be the same, with the exception of end repair. The end repair reaction is optimized for 5  $\mu\text{g}$ ; doubling the input DNA would require doubling the end repair reaction.

**Reagent Purity – (Reagent Aliquots and Cross contamination)**—Reagent purity and avoiding cross contamination are crucial factors of this protocol. Aliquot all reagents to single-use volumes and use PCR grade water.

There are multiple amplifications steps in this protocol and micrograms of DNA for a given sample are passed from one step to the next. When capturing multiple samples at once, samples must be handled carefully. Leave blank wells on agarose gels and between samples on plates where possible.

**Magnetic Beads vs. Qiagen Columns**—AMPure magnetic beads and Qiagen columns can be used interchangeably as purification steps without a significant difference in concentration yield. In a high-throughput setting, beads are cost- and time-efficient. However, magnetic beads require a minimum elution volume of 30  $\mu\text{l}$ . Since AMPure purification calls for 1.8  $\mu\text{l}$  of beads per 1  $\mu\text{l}$  of reaction, a 70  $\mu\text{l}$  reaction is the maximum volume that can be completed in a standard PCR plate. Half skirted 300  $\mu\text{l}$  plates are a

solution; however, half skirted plates cannot be gripped or moved on some robotic platforms.

**Adaptor Contamination**—Achieving a robust T4 ligase reaction in the adaptor ligation is a key step. Unligated adaptors can be carried through to sequencing, reducing useful data output. If adaptor bands remain present in the final sample, titrating the adaptor ligation reaction can be helpful.

The blockers used during the array hybridization are designed with the purpose of blocking adaptor and primer sequences, and their inclusion has been demonstrated to be critical. We speculate that they prevent common adaptor sequences of different fragments in the genomic library from annealing to each other, i.e. “daisy-chaining”, which would be predicted to reduce specificity.

**Array Handling**—When eluting samples off of arrays controlling the heat block to 95°C with a thermometer is crucial, especially if using a digital heat block with a built in thermometer. If using blocks for micro-centrifuge tubes or strip-tubes, partially fill the wells with distilled water to create steam contact with the array.

**Overamplification – RT PCR Machine with Pause Feature vs. Thermocycler**—The PCR amplifications are key steps in the protocol that determine the sequence complexity of the samples. This protocol employs a “no molecule left behind strategy”, in which the entire sample mixture is carried forward into each PCR to ensure maximal library complexity. Overamplifying the sample will result in an over representation of some sequences. This artifact may create a bias in the library, thus necessitating more sequencing than would be otherwise required. Overamplification also makes it more difficult to quantify samples, as some fraction of the amplified product remains single stranded.

Running samples on a real-time PCR machine offers the advantage of visually verifying amplification based on individual sample curves. When the sample curve hits a plateau, the sample must be taken out at the end of that cycle. Real-time PCR machines with a pause feature where the lid can be opened, allowing removal of the sample mid-run. Some real-time machines have a delayed real-time analysis and will not be useful for stopping a sample from overamplification.

Samples that are captured in parallel through under the same conditions should behave within a few cycles from each other. Once the number of cycles is optimized for a specific input gDNA extraction method, a standard thermal cycler can replace the RT-PCR machine. When using 5 µg of gDNA the amplification should peak, on average, at 15 cycles.

**Bioanalyzer vs. Gel and Quantification**—Analysis of the library by Bioanalyzer allows direct determination of the DNA library insert size range and concentration, and is the recommended quality control method. However, a combination of gel electrophoresis and DNA quantitation (eg Nanodrop or Picogreen) methods could be used to give a similar, albeit less precise, quality control result.

**Troubleshooting**—The key failure points in this protocol are loss of DNA during the library preparation steps, and poor recovery during array elution. Care should be taken to ensure that elutions off AMPure or Qiagen columns are performed carefully. Spot-checking DNA concentrations throughout the protocol can help identify areas of loss. During array elution, it is important to heat the array and elution buffer to 95°C. Many heat blocks do not maintain temperatures evenly over long periods of time; frequent verification of the temperature will ensure full denaturation of the sample from the array. Small bubbles should

be visible during the elution; turning and tapping the array during elution steps will ensure the entire surface is reached by the elution buffer. Care should be taken to avoid touching the active surface (Agilent label side) with fingertips or pipette tips. Any contact will result in the loss of DNA captured by probes in the affected region.

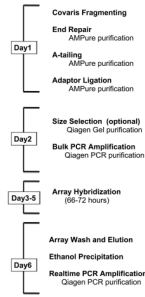
**Anticipated Results**—Assuming a starting input of 5  $\mu\text{g}$  high-quality genomic DNA, the shotgun library preparation (Fragmentation – Bulk PCR) should produce 10–40  $\mu\text{g}$  of library for hybridization to the array. The final PCR should produce  $\sim 30$   $\text{ng}/\mu\text{L}$  of product, which should be sufficient for multiple sequencing lanes.

## Acknowledgments

Support for the development of this protocol was provided in part by grants from the NHLBI (5R01HL094976 to JS and DAN) and NHGRI (5R21HG004749 to JS).

## References

1. Albert TJ, Molla MN, Muzny DM, Nazareth L, Wheeler D, Song X, Richmond TA, Middle CM, Rodesch MJ, Packard CJ, Weinstock GM, Gibbs RA. Direct selection of human genomic loci by microarray hybridization. *Nat Methods*. 2007; 4:903–5. [PubMed: 17934467]
2. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol*. 2009; 27:182–9. [PubMed: 19182786]
3. Hodges E, Rooks M, Xuan Z, Bhattacharjee A, Benjamin Gordon D, Brizuela L, Richard McCombie W, Hannon GJ. Hybrid selection of discrete genomic intervals on custom-designed microarrays for massively parallel sequencing. *Nat Protoc*. 2009; 4:960–74. [PubMed: 19478811]
4. Hodges E, Smith AD, Kendall J, Xuan Z, Ravi K, Rooks M, Zhang MQ, Ye K, Bhattacharjee A, Brizuela L, McCombie WR, Wigler M, Hannon GJ, Hicks JB. High definition profiling of mammalian DNA methylation by array capture and single molecule bisulfite sequencing. *Genome Res*. 2009; 19:1593–605. [PubMed: 19581485]
5. Hodges E, Xuan Z, Balija V, Kramer M, Molla MN, Smith SW, Middle CM, Rodesch MJ, Albert TJ, Hannon GJ, McCombie WR. Genome-wide in situ exon capture for selective resequencing. *Nat Genet*. 2007; 39:1522–7. [PubMed: 17982454]
6. Okou DT, Steinberg KM, Middle C, Cutler DJ, Albert TJ, Zwick ME. Microarray-based genomic selection for high-throughput resequencing. *Nat Methods*. 2007; 4:907–9. [PubMed: 17934469]
7. Porreca GJ, Zhang K, Li JB, Xie B, Austin D, Vassallo SL, LeProust EM, Peck BJ, Emig CJ, Dahl F, Gao Y, Church GM, Shendure J. Multiplex amplification of large sets of human exons. *Nat Methods*. 2007; 4:931–6. [PubMed: 17934468]
8. Turner EH, Lee C, Ng SB, Nickerson DA, Shendure J. Massively parallel exon capture and library-free resequencing across 16 genomes. *Nat Methods*. 2009; 6:315–6. [PubMed: 19349981]
9. Turner EH, Ng SB, Nickerson DA, Shendure J. Methods for genomic partitioning. *Annu Rev Genomics Hum Genet*. 2009; 10:263–84. [PubMed: 19630561]



**Figure 1.**  
Basic Flowchart for Genomic Enrichment by Array Hybridization



**Table 1**

## Oligonucleotide sequences

EXOME_BULK_FWD	ACA CTC TTT CCC TAC ACG ACG* C
EXOME_BULK_REV	CTC GGC ATT CCT GCT GAA C*C
EXOME_BLOCK_F	CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT CCG ATC* T
EXOME_BLOCK_R	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC* T
EXOME_BLOCK_F_rev	AGA TCG GAA GAG CGT CGT GTA GGG AAA GAG TGT
EXOME_BLOCK_R_rev	AGA TCG GAA GAG CGG TTC AGC AGG AAT GCC GAG
EXOME_SEQ_AMP_FWD	CAA GCA GAA GAC GGC ATA CGA GAT CGG TCT CGG CAT TCC TGC TGA ACC GCT CTT CCG ATC* T
EXOME_SEQ_AMP_REV	AAT GAT ACG GCG ACC ACC GAG ATC TAC ACT CTT TCC CTA CAC GAC GCT CTT CCG ATC* T
EXOME_ADAPT_HI	ACA CTC TTT CCC TAC ACG ACG CTC TTC CGA TC*T
EXOME_ADAPT_LO	/5Phos/GAT CGG AAG AGC GGT TCA GCA GGA ATG CCG AG

**Table 2**

## Master Mix for End Repair (per sample)

<b>Component</b>	<b>Volume</b>
10X Next End-Repair Buffer (NEB)	10 $\mu$ l
Next End-Repair Enzyme Mix (NEB)	5 $\mu$ l
TOTAL	15 $\mu$ l

**Table 3**

Master Mix for A-Tail (per sample)

<b>Component</b>	<b>Volume</b>
10X PCR buffer w/15mM MgCl <sub>2</sub>	5 $\mu$ l
dATP (100mM)	0.5 $\mu$ l
Taq (5U/ $\mu$ l)	0.5 $\mu$ l
<b>TOTAL</b>	<b>6 <math>\mu</math>l</b>

**Table 4**

Adaptor Ligation Master Mix (per sample):

<b>Component</b>	<b>Volume</b>
Annealed paired-end adaptor (50 $\mu$ M)	5 $\mu$ l
10X T4 Ligation Buffer	5 $\mu$ l
T4 ligase (400 U/ $\mu$ L)	5 $\mu$ l
<b>TOTAL</b>	<b>15 <math>\mu</math>l</b>

**Table 5**

## Master mix for Pre-array bulk Amplification (per sample)

Number of reactions (23 DNA, 1 no template control)	24
Total volume with DNA template	1200 $\mu$ l
<b>Component</b>	<b>Volume</b>
Water	823.2 $\mu$ l
5x Phusion buffer (5x $\rightarrow$ 1x)	240 $\mu$ l
25mM dNTPs [25mM $\rightarrow$ 200 $\mu$ M]	9.6 $\mu$ l
EXOME_BULK_FWD [100 $\mu$ M $\rightarrow$ 500nM]	6 $\mu$ l
EXOME_BULK_REV [100 $\mu$ M $\rightarrow$ 500nM]	6 $\mu$ l
Phusion HotStart (2U/ $\mu$ l $\rightarrow$ 0.02U/ $\mu$ l)	12 $\mu$ l
<b>TOTAL</b>	<b>1100 <math>\mu</math>l</b>

**Table 6**

Master mix for array hybridization (per sample)

<b>Component</b>	<b>Volume</b>
EXOME_BLOCK_F (100 $\mu$ M)	3.6 $\mu$ l
EXOME_BLOCK_R (100 $\mu$ M)	3.6 $\mu$ l
EXOME_BLOCK_F_rev (100 $\mu$ M)	3.6 $\mu$ l
EXOME_BLOCK_R_rev (100 $\mu$ M)	3.6 $\mu$ l
Cot1 human DNA (1mg/ml)	50 $\mu$ l
10X Blocking Agent (Agilent)	52 $\mu$ l
2X Hybridization Buffer (Agilent)	260 $\mu$ l
<b>TOTAL</b>	<b>400 <math>\mu</math>l</b>

**Table 7**

## Ethanol Precipitation Recipe (per elution)

<b>Component</b>	<b>Volume</b>
1/10 vol NH <sub>4</sub> OAC (7.5 M)	40 μL
1/100 vol glycogen (5mg/ml)	4 μL
2.5 vol 100% cold ethanol	1000 μL
<b>TOTAL</b>	<b>1044 μL</b>

**Table 8**

Master mix for Pre-sequencing Amplification (per sample)

<b>Component</b>	<b>Volume</b>
Water	3 $\mu$ L
2x iTaq Supermix	25 $\mu$ L
EXOME_SEQ_FWD [10 $\mu$ M $\rightarrow$ 200nM]	1 $\mu$ L
EXOME_SEQ_REV [10 $\mu$ M $\rightarrow$ 200nM]	1 $\mu$ L
<b>TOTAL</b>	<b>30 <math>\mu</math>L</b>