**Molecular inversion probe experimental protocol**

Molecular inversion probes are 70 nucleotide single stranded DNA molecules containing end sequences complementary to the target DNA (40 bases total, most typically 20 bases for each end hybridization sequence), connected by a 30-base linker sequence. The linker region is common to all MIPs and contains two universal PCR primer sites. When the probe is hybridized to the genomic target, there is a gap (112 bases) spanning the genomic region of interest to be captured.

**Protocol overview**

* **Pool all the MIPs together and phosphorylate**

DNA ligase requires a 5’-phosphorylated end to close the circle. Thus, we phosphorylate the probes. For the first time testing a set of MIPs, we naively add equal volumes of each MIP to the pool. With some data from a pilot experiment using a MIP set, we can alter initial volumes of MIPs added to the pool to enhance capture performance of MIPs that did not capture well in the pilot experiment.

* **Capture targeted genomic regions of interest**

Probes are added to the genomic DNA sample. After denaturation, the target-complementary ends of the probe hybridize to the target DNA during an annealing step. DNA polymerase primes off the 3’ end of the probe, incorporating the complement of the target sequence into the probe oligonucleotide. Once the entire target sequence is incorporated, DNA ligase ligates the end of the newly incorporated target sequence with the phosphorylated 5’ end of the probe, resulting in a circularized molecule containing the probe sequence and the target sequence of interest.

* **Remove linear probes and genomic DNA**

Probes that fail to capture their corresponding target region of interest remain linear. Exonuclease treatment removes these non-reacted probes as well as any remaining linear genomic DNA in the reaction.

* **Amplify captured regions of interest and barcode samples**

PCR amplification is performed using primers complimentary to the two universal PCR primer sites that are common to all MIPs. We use barcoded reverse primers so we can combine several samples into a single library.

* **Sequencing preparation and sequencing**

Captured targeted sequences can be identified by sequencing. We have 384 different barcodes, so we can pool up to 384 samples and sequence them together. Sequencing can be performed in a single lane on the HiSeq2000 Illumina sequencer (PE-100) or via a single MiSeq run (PE-151, 310 cycles), depending on how many MIPs we have, how many individual samples we have combined, and the coverage we need.

1. **MIP pooling**

We order the oligos from Integrated DNA Technologies: 70mer oligos, column synthesized at the 25 nanomole scale and hydrated to 100 µM in 1X TE Buffer, pH 8.0.

Probes are initially pooled at equal ratios. Example MIP1714: 19 plates of oligos, we take 5ul of each MIP using 8-channel P10 and pool to one column, then pool that column into a 1.5ml tube. So we have one pool of MIPs per each plate. Then, we take 0.1µl x # MIPs in the pool from each tube and put them into a new tube. We'll phosphorylate this and store the rest in case we need it for future use.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SSC\_EXOME\_MIP | #mips/plate | total volume | amount from each pool µl | Current Conc | Final\_Conc | Final\_Volume |
| 1 | 95 | 475 | 9.5 | 1.052631579 | 0.058343057 | 171.4 |
| 2 | 95 | 475 | 9.5 | 1.052631579 | 0.058343057 |  |
| 3 | 95 | 475 | 9.5 | 1.052631579 | 0.058343057 |  |
| 4 | 93 | 465 | 9.3 | 1.075268817 | 0.058343057 |  |
| 5 | 93 | 465 | 9.3 | 1.075268817 | 0.058343057 |  |
| 6 | 90 | 450 | 9 | 1.111111111 | 0.058343057 |  |
| 7 | 94 | 470 | 9.4 | 1.063829787 | 0.058343057 |  |
| 8 | 88 | 440 | 8.8 | 1.136363636 | 0.058343057 |  |
| 9 | 96 | 480 | 9.6 | 1.041666667 | 0.058343057 |  |
| 10 | 96 | 480 | 9.6 | 1.041666667 | 0.058343057 |  |
| 11 | 90 | 450 | 9 | 1.111111111 | 0.058343057 |  |
| 12 | 94 | 470 | 9.4 | 1.063829787 | 0.058343057 |  |
| 13 | 96 | 480 | 9.6 | 1.041666667 | 0.058343057 |  |
| 14 | 96 | 480 | 9.6 | 1.041666667 | 0.058343057 |  |
| 15 | 96 | 480 | 9.6 | 1.041666667 | 0.058343057 |  |
| 16 | 94 | 470 | 9.4 | 1.063829787 | 0.058343057 |  |
| 17 | 25 | 125 | 2.5 | 4 | 0.058343057 |  |
| 18 | 94 | 470 | 9.4 | 1.063829787 | 0.058343057 |  |
| 19 | 94 | 470 | 9.4 | 1.063829787 | 0.058343057 |  |

1. **MIP phosphorylation**

Example MIP1714:

|  |  |
| --- | --- |
| 174.1 | µl MIPs |
| 20.5 | µl 10X T4 DNA Ligase Buffer with 10mM ATP |
| 10.4 | µl T4 PNK |
| 205 |  |

Split the volume in two PCR tubes so we can do it in a PCR machine:

|  |  |
| --- | --- |
| 37°C, 45min |  |
| 65°C, 20min |  |
| 4°C, forever |  |

**We can calculate the concentration of our MIP pool stock:**

|  |  |
| --- | --- |
| Vi x Ci = Vf x Cf |  |
| 0.1µL x 100µM = 205µl x Cf | |
| **Cf = 0.04878 µM** |  |

1. **MIP capture**

In one step we are hybridizing the probe to the DNA, filling the gap with a polymerase, and closing the circle with a ligase. We have done the capture with different conditions: amount of DNA, amount of probes, and time of incubation. In the MIP1714 example we were using 100ng of DNA, ratio MIP:DNA copies was 200:1, and the time of incubation 23 hours. For the *SRGAP2*/*RH* MIPs, we used an 800:1 MIP:gDNA ratio. NOTE: During this step and all subsequent steps involving use of a thermocycler, take great care to seal your plate very well – if not well-sealed, you risk losing all your DNA!

**Calculations of the pool of MIPs:**

We need to calculate the µl of MIP pool that we need to do the capture reaction. Let’s do the example of MIP1714 where we captured 100ng of DNA, the ratio of MIP copies :DNA copies was 200:1, and the concentration of our MIP pool is 0.04878 pmols:

1ng DNA = 330 haploid genome copies

100 ng DNA = 33000 haploid genome copies

Copies of MIP needed: 200 x 33000 = 6600000 MIP copies

Now we can transform number of copies to picomoles, using Avogadro’s number: 6.02214 x 1023

6600000 molecules x 1 mole x 1x1012 picomoles = 1.1 x 10-5 pmol

6.02214 x 1023 molecules1 mole

Our MIP pool concentration is 0.04878 pmols, so we can calculate the µl to take of this pool for the capture reaction:

1.1 x 10-5 pmol x 1ul = 2.25 x 10-4 µl

0.04878pmol

If the volume is too small, we can dilute our stock of MIPs to a lower concentration so we can pipette a higher volume. For example doing a 1/5 dil from the stock, and then a 1/100 dil from the 1/5 dil.

**MIP pool normalization:**

The first time using a MIP pool, we test it running just a few samples to evaluate uniformity. Then we can rebalance the MIP pool by adding more of the MIPs that didn’t perform well in the test run. In the example MIP1714; we pool all the “good” MIPs in one tube, and all the “bad MIPs” in another tube, and we phosphorylate them separately. Then we pool at a ratio 50:1 (bad:good). We’ll use the final concentration of the good MIPs in the normalized pool to calculate the µl of MIPs to add to the capture.

So… we take 0.1µl of each MIP to make each pool, and then we phosphorylate separately in a final volume of 100µl, both pools we’ll be at the same concentration of 0.1µM (Vi x Ci = Vf x Cf; 0.1µL x 100µM = 100µl x Cf; Cf = 0.1 µM).

We can make a new rebalanced pool with 1µl of the good pool and 50µl of the bad pool. The new concentration will be: Vi x Ci = Vf x Cf; 1µL x 0.1µM = 51µl x Cf; Cf = 0.0019607µM. We’ll use this concentration to calculate the µL of the rebalanced pool to add to the capture mix.

**Capture Mix:**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1X Mix |  |  |
| Ampligase Buffer | 2.5 | µl |  |
| MIP pool | X | µl |  |
| dNTP 0.25mM | 0.032 | µl |  |
| Stoffel Taq 10U/ul | 0.04 | µl |  |
| Ampligase 100U/ul | 0.01 | µl |  |
| H2O | Up to 15 | µl |  |
|  |  |  |  |
| Add DNA (10ng/ul) | 10 |  |  |
| Final Volume | 25 |  |  |

**New Capture Mix (Klentaq):**

|  |  |  |  |
| --- | --- | --- | --- |
|  | 1X Mix |  |  |
| Ampligase Buffer | 2.5 | µl |  |
| MIP pool | X | µl |  |
| dNTP 0.25mM | 0.032 | µl |  |
| Hemo Klentaq | 0.32 | µl |  |
| Ampligase 100U/ul | 0.01 | µl |  |
| H2O | Up to 15 | µl |  |
|  |  |  |  |
| Add DNA (10ng/ul) | 10 |  |  |
| Final Volume | 25 |  |  |

In PCR machine:

|  |
| --- |
| 95°C, 10 min |
| 60°C, 23 hours |
| 10°C lid off-set |

1. **Exo treatment**

|  |  |  |
| --- | --- | --- |
|  | **Master Mix (ul)** |  |
| EXO I | 0.5 µl |  |
| EXO III | 0.5 µl |  |
| Ampligase buffer 10X | 0.2 µl |  |
| H2O | 0.8 µl |  |
| TOTAL | 2 |  |

Cool down plates on ice before adding mix, use cold blocks. I put the mix in an 8 strip-tubes first, and use a multichannel to add the 2µl to the captured plates.

In PCR machine:

|  |  |  |
| --- | --- | --- |
| 37°C, 45 minutes |  |  |
| 95°C, 2 minutes (exo inactivation) | |  |
| 4°C, forever |  |  |

1. **Test RT-PCR**

The first time we are doing the experiment on a particular sample set, we need to do a RT-PCR to check at what cycle the PCR gets to the plateau. NOTE: it is very important to do this once for each new sample set – for example, the HapMap DNA samples behaved differently than a Troina sample tested. Thus, the cycle value I calculated for the HapMap set is not applicable to the Troina set.

|  |  |  |
| --- | --- | --- |
|  | **Mix (ul)** |  |
| 2X iProof | 12.5 | µl |
| Primer SLXA\_PE\_MIPBC\_FOR  (100 uM) | 0.125 | µl |
| Sybr Green 100x | 0.125 | µl |
| H2O | 6 | µl |
| TOTAL | 18.75 | µl |
|  |  |  |
| barcode Rv (10 uM) | 1.25 | µl |
| template MIP reaction | 5 | µl |
| Total volume reaction | 25 | µl |
|  |  |  |
|  |  |  |
|  |  |  |
| Thermocycler: |  |  |
| 98°C, 30" |  |  |
| 98°C, 10" |  |  |
| 60°C, 30" | until it gets to the plateau | |
| 72°C, 30" | 20sec, read, 10s | |
| 72°C, 2' |  |  |
| 4°C, forever |  |  |
|  |  |  |

1. **PCR**

Now that we know the cycles to do, we can do it in

a regular PCR machine. Be sure to record which barcode set you use so you can later interpret the results!

|  |  |
| --- | --- |
|  | **Mix (ul)** |
| 2X iProof | 12.5 |
| Primer SLXA\_PE\_MIPBC\_FOR  (100 uM) | 0.125 |
| H2O | 6.125 |
| TOTAL | 18.75 |
|  |  |
| barcode Rv (10 uM) | 1.25 |
| template MIP reaction | 5 |
| Total volume reaction | 25 |
|  |  |
| Thermocycler: |  |
| 98°C, 30" |  |
| 98°C, 10" |  |
| 60°C, 30" | X cycles |
| 72°C, 30" | (20 for MIP1714) |
| 72°C, 2' |  |
| 4°C, forever |  |

1. **Gel, pooling, clean up**

Run gel with a few samples of each plate to make sure we have the desired product.

Pool 5 µl from each sample together. I pool first in one column with multichannel, and then pool the 8 wells in a 1.5ml tube.

Clean up pool with Agencourt beads: we can play with the amount of beads so we can get rid of lower bands that sometimes appear (especially when using a bigger pool of probes). (NOTE: Allow Agencourt beads to come to room temperature before using them; waiting 30 mins after taking them out of the fridge is a good rule of thumb; mix to ensure no beads stick to bottom of bottle)

Example *SRGAP2*/*RH* MIPs:

-pool 5 uL of PCR reaction from each sample (96 samples) to make a total sample volume of 480 uL

-add 1.8 uL beads per 1 uL sample pool to sample pool (add 864 uL beads to sample pool in this case); vortex and spin down

-incubate sample pool with added beads at room temperature for 10 minutes

-place tube with sample pool and beads onto room temperature magnet plate and wait 5 minutes

-pipette out clear solution from sample pool

-add enough 70% EtOH to fully cover beads (1 mL worked well in this case); incubate 30 s

-pipette out EtOH and repeat EtOH wash (above step)

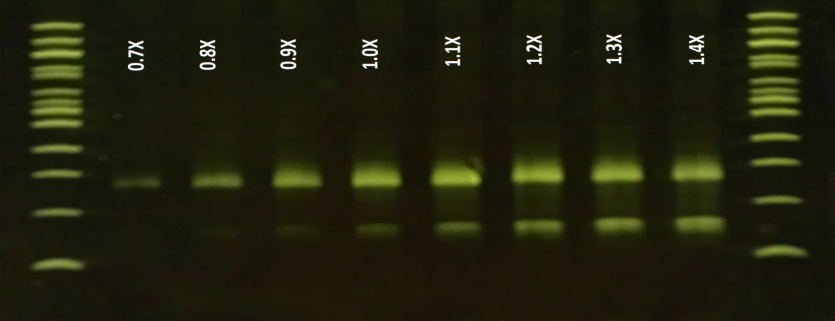
-pipette out EtOH and allow beads to dry for 5 minutes

-remove tube with beads from magnet plate and add 100 uL EB (elution buffer)

-mix by pipetting up and down at least 10 times

-transfer EB with beads to magnet plate and wait at least 1 min for beads to separate from EB

-transfer EB (which contains your DNA) to a new tube for storage



Our band

Band to get rid of

Run gel after Agencourt cleanup.

Prepare MegaPools combining different pools. Ex: MIP1714; Megapools of two pools, so 192 samples.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Sample | Conc1 | Conc2 | Avg | # of Samples | ng/sample\*ul | Normalize to Plate 2 | ul | Barcode |
| Plate 1 | 50.40 | 50.58 | 50.49 | 82 | 0.615731707 | 0.602179887 | 15.05 | Barcode Pl1 |
| Plate 2 | 36.52 | 34.67 | 35.595 | 96 | 0.37078125 | 1 | 25.00 | Barcode Pl2 |

The MIP library is now ready to be sequenced. For MIP1714 we were doing one lane HiSeq 2000 Illumina.

1. **SEQUENCING CONSIDERATIONS**

Which sequencing platform you use depends on your desired number of reads. The MiSeq platform run for 310 cycles at 151 bp paired end mode generates ~10 million paired end reads (1 paired end read consists of a 1st read, an index read, and a 2nd read), ~7 million of which are good (based on comparing total and mapped paired end reads from the *SRGAP2*/*RH* MIP experiment). For the SRGAP2/RH MIPs, there were 257 MIPs total and 96 individuals total, so the approximate expected read count per individual per MIP is:

7000000 good reads / (96 individuals \* 257 MIPs) = ~280 good reads / (individual \* MIP)

For paralog-specific copy number genotyping, this number worked very well. In practice, some MIPs captured poorly and had far fewer than this number of mapped reads. Others had far more mapped reads. Rebalancing probe concentration during pooling should help make total MIP counts more uniform between MIPs. It might be possible to still get good results with fewer reads; however, a good rule of thumb for this application is that you should sequence such that you will receive an expectation of at least 280 good reads / (individual \* MIP).

**REAGENTS**

|  |  |  |
| --- | --- | --- |
| **Item** | **Company** | **Catalog#** |
| T4 Polynucleotide Kinase, 2,500 units 10,000 units/ml | NEB | M0201L |
| Ampligase 10X Reaction Buffer (5ml) | Epicentre/Illumina | A1905B |
| AmpliTaq® DNA Polymerase, Stoffel Fragment, 1000 units | Applied Biosystems | N8080038 |
| Ampligase® DNA Ligase | Epicentre/Illumina | A0110K |
| Exonuclease I (E. coli) 15,000 units; 20,000 U/mL | NEB | M0293L |
| Exonuclease III (E. coli), 25,000units at 100k units/ml | NEB | M0206L |
| iProof HF Master Mix (2 tubes) | Bio-Rad | 172-5310 |
| Agencourt AMPure XP – PCR purification | Beckman Coulter Genomics Inc. |  |
| SYBR® Green I Nucleic Acid Gel Stain - 10,000X concentrate in DMSO | Invitrogen | S-7563 |

**PCR PRIMERS:**

Primer SLXA\_PE\_MIPBC\_FOR

AATGATACGGCGACCACCGAGATCTACACATACGAGATCCGTAATCGGGAAGCTGAAG

Primer MIPBC\_RV (N: bp of barcode sequence)

CAAGCAGAAGACGGCATACGAGATNNNNNNNNACACGCACGATCCGACGGTAGTGT

**SEQUENCING PRIMERS:**

Primer MIPBC\_SEQ\_FOR (for/r1, corresponding to read 1 of Illumina output)

CATACGAGATCCGTAATCGGGAAGCTGAAG

Primer MIPBC\_SEQ\_REV (rev/r2, corresponding to read 3 of Illumina output)

ACACGCACGATCCGACGGTAGTGT

MIPBC\_SEQ\_INDX (index/ind, corresponding to read 2 of Illumina output)

ACACTACCGTCGGATCGTGCGTGT

**VISUAL OVERVIEW:**

