Apr 13, 2020

# An error-corrected panel-based next-generation sequencing assay for ultra-sensitive somatic mutation detection in ctDNA

## Heidi Fettke[1]

[1]Department of Medicine, School of Clinical Sciences, Monash University, Australia

1   Works for me        dx.doi.org/10.17504/protocols.io.bb3siqne

Heidi Fettke ⚡

## ABSTRACT

Circulating tumour DNA (ctDNA), fragmented DNA shed into the circulation from tumour cells, holds the potential to become a minimally-invasive biopsy material with which to interrogate the tumour landscape. However, ctDNA can be highly diluted in normal cell-free DNA (cfDNA) (released into the blood primarily by hematopoietic cells), and thus robust liquid biopsy assays with high analytical sensitivity are required for ctDNA to reach its full clinical potential in molecular genotyping.

This protocol uses optimised off-the-shelf reagents in order to generate indexed and molecularly-tagged cfDNA and germline DNA libraries compatible with Illumina short-read sequencing. This protocol involves the hybridisation-capture of targeted regions of the genome using RNA baits - and has been used previously with a 42-gene panel 447 kbp in size. Also described here is the bioinformatic tools used for data processing and variant identification.

The combination of deep-sequencing and error-correction based on molecular barcoding allows high sensitivity and specificity to be generated at low variant frequencies.

## GUIDELINES

Always try to use lobind plasticware to reduce loss of cfDNA during processing (this includes pipette tips)

Always use pure, nuclease-free water (and PBS if necessary)

Always try to keep Pre-PCR components separate from Post-PCR components. Note: Post-PCR = final stage of library generation

You can store your cfDNA in TE buffer rather than water but keep in mind where you may need to concentrate your samples

| Blood collection | 5m |
| --- | --- |

1   Collect whole peripheral blood in either an EDTA tube or a dedicated cfDNA stabilising tube such as a Streck Cell-Free DNA BCT tube depending on the estimated time to processing (within 4 hours for an EDTA tube otherwise use a stabilising tube). 21-23 gauge needle or larger preferred to reduce chances of haemolysis. Invert tube ~5 times, then try to prevent any further agitation during transit if possible.

| Blood processing | 40m |
| --- | --- |

2   Centrifuge blood tubes at ⊛**1900 x g, 4°C 00:10:00** to facilitate the initial separation of plasma from other blood components.

3   Separate plasma into 🧪**5 ml** eppendorf tubes, being careful not to draw up the buffy coat layer below.

At the same time, collect the buffy coat layer (approx 🧪**1 ml**) below the plasma into a cryovial and store at ❄**-80 °C** until required.

4   Further clarify the plasma via centrifugation at ⊛**16000 x g, 4°C 00:10:00**.

Draw up supernatant into a fresh 🧪**5 ml** eppendorf tube, leaving any pellet behind to be discarded.

Re-centrifuge the supernatant to ensure the absence of a pellet.

Aliquot the plasma into cryovials and store at ❄**-80 °C** required.

---

| cfDNA extraction | 3h |
|---|---|

5   Thaw plasma on ice.

Extract cfDNA from plasma using the QIAamp circulating nucleic acid kit (Qiagen) as per the manufacturer's instructions. Elute into 🧪**50 µl** buffer AVE. Pass the eluate over the column again to maximise yield after a second on-column incubation period of ⏱**00:01:00**.

Note: Try to avoid extracting cfDNA from more than 6 patient samples at once in order to avoid cross-contamination as much as possible.

6   Remove contaminating large molecular weight DNA fragments (that can arise due to cellular lysis during the blood draw or transporting process:

- Incubate the 🧪**50 µl** eluate with 🧪**25 µl** ampureXP paramagnetic beads (Beckman Coulter) (a 0.5:1 v/v ratio) at room temperature for ⏱**00:05:00**.

- separate beads on a magnetic rack for ⏱**00:02:00** and draw off the supernatant into a new eppendorf tube. The beads now contain large DNA fragments (over 500 nt) and can be discarded

- to the supernatant, add a further 🧪**50 µl** ampure XP beads (1:1 v/v) to the supernatant and incubate for ⏱**00:10:00** at room temperature in order to capture the remaining cfDNA fragments

- separate beads on a magnetic rack for ⏱**00:02:00** and discard the supernatant

- keeping the beads on the magnetic rack, wash twice with 🧪**200 µl** 80% (v/v) pure ethanol

- after discarding the ethanol, allow the beads to dry for between ⏱**00:05:00** to ⏱**00:10:00** (do not leave for so long that cracks appear)

- elute the cfDNA from the beads by mixing well with 🧪**13 µl** pure water and incubating at room temperature for ⏱**00:10:00**

- place tube back on the magnetic rack to separate the beads, drawing up the eluate and storing in a new eppendorf tube

7

Assess size distribution, quality and quantity of resulting cfDNA using automated gel electrophoresis (2200 TapeStation and Expert software, Agilent) and the Qubit dsDNA high-sensitivity assay (Thermo Fisher Scientific). When using Qubit, it is a good idea to always use a control i.e. 🧪**1 µl** of standard #2

Note: cfDNA can be stored at either ❄**4 °C** or ❄**-20 °C**

## gDNA extraction          1h

8   Thaw buffy coat at room temperature.
    Extract germline DNA from buffy coat (containing peripheral blood mononuclear cells) using the QIAamp DNA blood mini kit
    (Qiagen) and elute into 🔲**200 µl** buffer AE.

9   Shear extracted gDNA to 150-200 bp in length using ultrasonication (Covaris), with the following settings:
    Duty Factor: 10%
    Peak Incident Power: 175
    Cycles per Burst: 200
    Treatment Time: 360 seconds
    Bath temperature: 4-8ºC

10  Quantify and qualify fragmented gDNA using automated gel electrophoresis (2200 TapeStation and Expert software, Agilent)
    and the Qubit dsDNA high-sensitivity assay (Thermo Fisher Scientific).

## Library construction          2h

11  Use the ThruPLEX Tag-seq kit as per the manufacturer's protocol to generate molecularly-tagged libraries from both cfDNA and
    gDNA. The ideal input is the kit maximum of 🔲**50 ng** as it will maximise the number of unique DNA fragments and improve
    chances of low frequency variant identification; however as little as 🔲**10 ng** can also be used.

12  Clean-up the library fragments by using a 1:1 (v/v) ratio of AmpureXP beads (Beckman Coulter) and eluting in 🔲**33 µl** pure
    water. Quantify using the qubit as before, and confirm library size (main peak should have moved from ~170 bp to ~340 bp.

13  If you have >400 ng of library, proceed to the next step. If not, perform 1-2 cycles of PCR enrichment following the protocol in
    Step 16 below.

## Hybridisation capture     4d

14    Use the Agilent SureSelect target enrichment system (Integrated Sciences) to capture target regions of interest using a library of biotinylated RNA baits complementary to the regions of interest (Integrated Sciences). In addition to the reagents required, you will also need 4 blocking oligonucleotides (IDT) to allow compatability with the SureSelect libraries:

Thruplex p5-UMI-11nt ext
AATGATACGGCGACCACCGAGATCTACACAGGCGAAGACACTCTTTCCCTACACGACGCTCTTC
CGATCTXXXXXXXTCAGTAGCTCA/3block/

Thruplex p5-UMI-8nt ext
AATGATACGGCGACCACCGAGATCTACACAGGCGAAGACACTCTTTCCCTACACGACGCTCTTC
CGATCTXXXXXXGTAGCTCA/3block/

Thruplex p7-UMI-11nt ext
CAAGCAGAAGACGGCATACGAGATXXXXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
CTXXXXXXXTCAGTAGCTCA/3block/

Thruplex p7-UMI-8nt ext
CAAGCAGAAGACGGCATACGAGATXXXXXXXXXGTGACTGGAGTTCAGACGTGTGCTCTTCCGAT
CTXXXXXXGTAGCTCA/3block/

Note: 'X' represents either sample-unique barcode or unique molecular identifier (UMI) sequence regions.

15    Follow the manufactuer's protocol for the first capture, incubating the baits with the libraries for ⏱ **24:00:00** at 🌡 **65 °C** ; with the following alterations:

1. Replace SureSelect Block 3 with 🔲**0.5 µl** of each of the indexing blockers, and 0 🔲**0.6 µl** pure water (total volume of blocking mix is 🔲**7.6 µl** )

2. Use 1 ul of capture library instead of the 🔲**2 µl** suggested for libraries <3 Mb

3. Perform 4 washes with 🔲**400 µl** of Wash Buffer 2, incubated for ⏱**00:20:00** each time

4. Re-suspend library in 🔲**14 µl** pure water

16 Perform PCR enrichment of the capture libraries using all 🗔**14 µl** of the captured library as follows:

| Reagent | Volume for 1 reaction |
|---|---|
| Nuclease-free water | 19.5µl |
| 5x Herculase Rxn Buffer (clear cap) | 10µl |
| Herculase II Fusion DNA Polymerase (red cap) | 1µl |
| 100mM dNTP Mix (green cap) | 0.5µl |
| 10µM Illumina P5 primer | 2.5µl |
| 10µM Illumina P7 primer | 2.5µl |
| **Total** | **36µl** |

Run the following steps on the thermal cycler:

| Step | Number of cycles | Temperature | Time |
|---|---|---|---|
| Step 1 | 1 | 98°C | 2 mins |
| Step 2 | 16 | 98°C | 30 secs |
| | | 57°C | 30 secs |
| | | 72°C | 1 min |
| Step 3 | 1 | 72°C | 10 mins |
| Step 4 | 1 | 4°C | Hold |

17 Complete the capture protocol by cleaning up the PCR enriched libraries with AmpureXP beads and elute in 33 ul pure water. Quantitate libraries using the Qubit as above. Libraries can be stored at ৪ **4 °C** for up to one week or at ৪ **-20 °C** for longer periods.

18 Perform DCAP-Seq by repeating the hybridisation capture and PCR enrichment procedures above with the following changes:
   1. Incubate library with probes for ⏱**05:00:00** at ৪ **65 °C** instead of 24 hours
   2. 5 PCR amplification cycles instead of 16
   3. Resuspend final libraries in 🗔**33 µl** pure water

Library sequencing and data processing

19 Generate between 50 -70 million 100-150 bp paired end reads (for cfDNA), and approximately 1 million reads (for gDNA) on the Illumina HiSeq2500 on rapid mode.

20 Demultiplex libraries using per sample indices and map to the human reference genome (hg19) using BWA-MEM (v 0.7.12) and converted to BAM files for further analysis using Samtools (v 1.6).

21 Generate read families and perform error-correction using the molecular barcodes using Connor (v 0.5). Only reads with perfectly matching barcodes and at least 3 replicates should be clustered together as a family to increase the validity of each base call. A consensus frequency of at least 0.66 is required to generate a collapsed read.

## Variant identification and filtering

**22**      Identify somatic variants by using VarDict (v 1.5.8) in paired samples mode using standard filter settings, with the exception of minmum variant frequency, which is reduced to 0.001%.

**23**      Filter resulting VCF files using VarSeq (Golden Helix) using the following settings:

Somatic variants
1. Status matches 'strong somatic'
2. Variant not located in intergenic regions
3. Variant must either be loss-of-function or a missense variant with a CADD PHRED score of 20 or greater
4. Variant not present in gnomAD Exomes Variant Frequencies v 2.1.1 (BROAD) database at greater than 0.001 frequency, or is missing altogether
5. Minimum read depth of 500X
6. Minimum variant allele frequency of 0.0025
7. Minimum of 3 supporting reads
8. Variant must have at least 1 supporting reverse read and 1 supporting forward read

Germline variants
1. Status matches 'germline'
2. Minimum read depth of 8X in the germline sample

**24**      Assess all remaining variants on Integrative Genomics Viewer (IGV) in order to remove any sequencing artefacts or common variants that may have been missed by gnomAD

## CNV identification and filtering

**25**      Generate CNV calls per patient sample using VarSeq CNV Caller ( VS-CNV) using the following settings:

1. Set up the reference files by importing BAM files from germline (normal) samples that have undergone sequencing using the same targeted panel and methodology as the tumour samples. At least 30 references are preferred. Pre-deduplicated files may be preferred if post-connor depth is low
2. Open VarSeq project and calculate coverage regions using the BED file used for panel design
3. Generate CNV calls: as standard settings are not sensitive enough to allow cfDNA analysis, custom parameters need to be set on the CNV calling algorithm:
- maximum allowable difference between tumour and reference samples <30%
- mean z-score for duplications reduced to 1.2; and mean ratio reduced to 1.3
- mean z-score for heterozygous deletions reduced to -1.2; and mean ratio reduced to 0.85
4. Additional filters used are as follows, but may need to be optimised for different assays:
- Duplications require a p-value <0.0001 AND a  z-score of >0.25 AND a minimum target depth of 50
- Deletions required to fulfill one of the following:
   Have p-value <0.00001 AND ratio of <0.87
   Have p-value <0.05 AND z-score <-0.85 AND ratio of <0.7