

Somatic mutation landscapes at single-molecule resolution

Stefanie V. Lensing (✉ sl18@sanger.ac.uk)

Wellcome Sanger Institute

Peter Ellis

Inivata

Federico Abascal

Wellcome Sanger Institute <https://orcid.org/0000-0002-6201-1587>

Iñigo Martincorena

Wellcome Sanger Institute <https://orcid.org/0000-0003-1122-4416>

Robert J. Osborne (✉ r.osborne@biofidelity.com)

Biofidelity <https://orcid.org/0000-0002-1914-1239>

Method Article

Keywords: NanoSeq, Mutation, Cancer, Duplex-Sequencing, BotSeqS,

Posted Date: April 28th, 2021

DOI: <https://doi.org/10.21203/rs.3.pex-1298/v1>

License:   This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Somatic mutations drive cancer development and may contribute to ageing and other diseases. Yet, the difficulty of detecting mutations present only in single cells or small clones has limited our knowledge of somatic mutagenesis to a minority of tissues. To overcome these limitations, we introduce nanorate sequencing (NanoSeq), a new duplex sequencing protocol with error rates <5 errors per billion base pairs in single DNA molecules from cell populations. The version of the protocol described here uses clean genome fragmentation with a restriction enzyme to prevent end-repair-associated errors and ddBTPs/dATPs during A-tailing to prevent nick extension. Both changes reduce the error rate of standard duplex sequencing protocols by preventing the fixation of DNA damage into both strands of DNA molecules during library preparation. We also use qPCR quantification of the library prior to amplification to optimise the complexity of the sequencing library given the desired sequencing coverage, maximising duplex coverage. The sample preparation protocol takes between 1 and 2 days, depending on the number of samples processed. The bioinformatic protocol is described in:

<https://github.com/cancerit/NanoSeq>

https://github.com/fa8sanger/NanoSeq_Paper_Code

Introduction

Reagents

NEBNext® Ultra™ II Q5® Master Mix (NEB: M0544L)

HpyCH4V (NEB: R0620L)

T4 DNA Ligase (NEB: M0202L)

dATP Solution (NEB: N0440S)

xGen CS Adapter - Tech Access (IDT: 1080799)

CutSmart® Buffer (NEB: B7204S)

NEBuffer™ 4 (NEB: B7004S)

ATP (Thermo Scientific: R0441)

ddNTP Set, 5 mM solutions (ddCTP, ddGTP, ddTTP) (GE Healthcare: 27204501)

Klenow Fragment (3'→5' exo-) (M0212L)

Nuclease-free water

Agencourt Ampure XP beads (Beckman Coulter: A63882)

75% Ethanol

KAPA library quantification kit (KK4835)

100 μM NanoqPCR1 primer (HPLC: 5'ACACTCTTTCCCTACACGAC3')

100 μM NanoqPCR2 primer (HPLC: 5'GTGACTGGAGTTCAGACGTG3')

UDI containing PCR primers (dried):

i5: AATGATACGGCGACCACCGAGATCTACAC[barcode]ACACTCTTTCCCTACACGACGCTCTTCCGATC*T

i7: CAAGCAGAAGACGGCATAACGAGAT[barcode]GTGACTGGAGTTCAGACGTGTGCTCTTCCGATC*T

*phosphorothioate bond

AccuClear Ultra High Sensitivity dsDNA Quantification kit (Biotium: 31028)

Equipment

Full set of pipettes and tips

Plate magnet

Eppendorf twin.tec plates and plate seals

384 well plate and optical seal

Thermocycler

Lightcycler e.g. Roche 480 Lightcycler

Method of DNA quantification e.g. bioanalyzer/plate reader

Sequencer e.g. Nova-seq

Procedure

Library Preparation

1. Prepare DNA samples and make each sample up to 20 μL in a 96 well twin.tec Eppendorf plate (e.g. extracted genomic DNA).
2. Perform an Ampure bead clean-up: per sample, mix 50 μL NFW with 50 μL Ampure beads and add 100 μL of this 50:50 mix to each 20 μL sample. Mix well by pipetting up and down and allow DNA to bind to beads, wash twice with 75% EtOH and re-suspend beads in 20 μL NFW.
3. Prepare a fragmentation mix:

10X CutSmart® Buffer	2.5 μL
NFW	2 μL
HpyCH4V (5U/ μL)	0.5 μL

4. Add the 5 μL fragmentation mix to the 20 μL bead suspension (this is an on-bead digestion).
5. Incubate at 37 $^{\circ}\text{C}$ for 15 min on a thermocycler
6. Perform an Ampure bead clean-up: add 62.5 μL Ampure XP beads to each 25 μL sample. Mix well by pipetting up and down and allow DNA to bind to beads, wash twice with 75% EtOH and elute in 15 μL NFW.
7. Make up a 1 mM dATP/ddBTP mix (combine 2.5 μL 100 mM dATP, 50 μL 5 mM ddCTP, 50 μL 5 mM ddTTP, 50 μL 5 mM ddGTP and 97.5 μL NFW).
8. Prepare A-tailing mix:

10X NEBuffer™ 4	1.5 μL
NFW	1.85 μL
1 mM dATP/ddBTP mix	1.5 μL
Klenow Fragment (3'→5' exo-)	0.15 μL

9. Add 5 μL A-tailing mix to 10 μL of the cleaned-up fragmentation product.
10. Incubate at 37 $^{\circ}\text{C}$ for 30 min on a thermocycler

11. Prepare ligation mix:

10X NEBuffer™ 4	2.24 µL
NFW	15.53 µL
10 mM ATP	3.74 µL
15 µM xGen Duplex Seq Adapters	0.33 µL
400 U/µL T4 DNA ligase	0.56 µL

12. Add 22.4 µL ligation mix to the 15 µL A-tail product

13. Incubate at 20 °C for 20 min on a thermocycler

14. Perform an Ampure bead clean-up: add 37.4 µL Ampure XP beads to each 37.4 µL sample. Mix well by pipetting up and down and allow DNA to bind to beads, wash twice with 75% EtOH and elute in 50 µL NFW.

DNA Quantification by qPCR

15. Take a KAPA library quantification kit (KK4835). Add the supplied primer premix to the supplied KAPA SYBR FAST master mix. In addition add 20 µL of 100 µM NanoqPCR1 primer (HPLC: 5'ACACTCTTCCCTACACGAC3') and 20 µL of 100 µM NanoqPCR2 primer (HPLC: 5'GTGACTGGAGTTCAGACGTG3') to the KAPA SYBR FAST master mix.

16. Dilute a fraction of each sample 1 in 500 using NFW and set up triplicate 10 µL qPCR reactions (6 µL master mix, 2 µL sample/standard, 2 µL water) in a 384 well plate.

17. Run samples on a qPCR machine e.g. Roche 480 Lightcycler

18. Perform analysis: Absolute quantification (2nd Derivative Maximum Method) with the high sensitivity algorithm).

19. Determine the nM (fmol/ µL) concentration per sample as follows: Mean of sample concentration x dilution factor (500) x 452/573/1000 (where 452 is the size of the standard in bp and 573 is the proxy for the average fragment length of the library in bp). We multiply this value by 1.5 to correct for the performance between different thermocyclers within the laboratory.

20. Dilute your sample to your desired fmol input amount in 25 μ L using NFW to achieve duplicate rates close to the optimal 81% duplicate rate. For instance, if working with 150 bp paired reads, take 0.3 fmol per 15x sequencing (the equivalent of 150 million read pairs). For 30x, take 0.6 fmol, and so on. We estimate that, under optimal dilution and after applying all the bioinformatic mapping quality filters, a 26x standard coverage (260 million read pairs) will result in 1x effective duplex coverage (3 billion duplex calls).

Sequencing of matched normals

Calling somatic mutations require accurate filtration of germline SNPs, which is achieved by looking at a matched normal. A cost-efficient approach to obtain a matched normal for a NanoSeq sample is to sequence an undiluted NanoSeq library from the same individual. In some cases, when there is no clonality, the undiluted DNA can be obtained from the very same library that is going to be diluted. Sequencing NanoSeq libraries as matched normals concentrates the coverage on the 30% of the genome covered by NanoSeq. We found that at least 2 fmol and the equivalent to 8x whole-genome sequencing allow obtaining a matched normal with enough depth of coverage.

PCR amplification

21. Add 25 μ L NEBNext Ultra II Q5 Master Mix to UDI containing PCR primers (*we purchase PCR plates that contain 0.1 nmol of dry i5 primer and 0.1 nmol of dry i7 primer in each well). Add 25 μ L sample, mix and thermocycle:

Step 1: 98 $^{\circ}$ C 30 seconds

Step 2: 98 $^{\circ}$ C 10 seconds

Step 3: 65 $^{\circ}$ C 75 seconds

Step 4: Return to Step 2, X times

Step 5: 65 $^{\circ}$ C for 5 min

Step 6: Hold at 4 $^{\circ}$ C.

*We use dry primers. When using re-suspended primers adjust your sample volume to allow for the addition of primers.

Note, in order to correct between different thermocyclers in our laboratory we have a 1.5X correction factor in our calculations. Thus our 0.3 fmol input translates to 0.2 fmol input if the correction factor is not applied.

The number of PCR cycles is dependent upon the input: 0.1 fmol (0.066 fmol) = 16 cycles, 0.3 fmol (0.2 fmol) = 14 cycles, 0.6 fmol (0.4 fmol) = 13 cycles, 5 fmol (3.33 fmol) = 10 cycles.

22. Perform two consecutive 0.7X Ampure bead clean-ups and elute in 50 μ L NFW.

23. Quantify samples e.g. using AccuClear Ultra High Sensitivity dsDNA Quantification kit (Biotium) or bioanalyzer.

24. Pool samples together for sequencing.

25. Run pool on bioanalyser and adjust the pool concentration.

26. Sequence on preferred illumina sequencer (e.g. Nova-Seq using 150PE read lengths).

27. Sequencing depth is determined by the fmol input into PCR e.g. 0.3 fmol input requires 15X whole genome coverage and 0.6 fmol input requires 30X whole genome coverage.