

CUTseq is a versatile method for preparing multiplexed DNA sequencing libraries from low-input samples

Xiaolu Zhang

Karolinska Institutet

Silvano Garnerone

Karolinska Institutet

Michele Simonetti

Karolinska Institutet

Marcin Nicoś

Karolinska Institutet

Luuk Harbers

Karolinska Institutet

Caterina Marchiò

Department of Medical Sciences University of Turin

Reza Mirzazadeh

Karolinska Institutet

Tiziana Venesio

Pathology Unit Candiolo Cancer Institute-FPO-IRCCS 10060 Candiolo (TO), Italy

Anna Sapino

Department of Medical Sciences University of Turin

Johan Hartman

Karolinska Institutet

Magda Bienko (✉ magda.bienko@scilifelab.se)

Karolinska Institutet

Nicola Crosetto (✉ Nicola.Crosetto@ki.se)

Karolinska Institutet

Keywords: Next-generation sequencing, FFPE tissue, PCR, whole-genome amplification, CUTseq, T7 phage promoter

Posted Date: May 10th, 2019

DOI: <https://doi.org/10.21203/rs.2.1742/v1>

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

[Read Full License](#)

Version of Record: A version of this preprint was published at Nature Communications on October 18th, 2019. See the published version at <https://doi.org/10.1038/s41467-019-12570-2>.

Abstract

Current multiplexing strategies for massively parallel sequencing of genomic DNA mainly rely on library indexing in the final steps of library preparation. This procedure is costly and time-consuming because a single library must be produced separately for each sample. Furthermore, library preparation is challenging in the case of low-input fixed samples, such as DNA extracted from formalin-fixed paraffin-embedded (FFPE) tissues. Here, we describe CUTseq, a method that uses restriction enzymes and in vitro transcription to barcode and amplify genomic DNA prior to library construction. We thoroughly validate CUTseq and demonstrate its applicability to both genome and exome sequencing, enabling multi-region genome profiling within single stained FFPE tissue sections, to assess intratumor heterogeneity at high spatial resolution. In conclusion, CUTseq is a versatile and cost-effective method for multiplexed DNA sequencing library preparation that can find numerous applications in research and diagnostics.

Introduction

In the past decade, next-generation sequencing (NGS) technologies have become widely available in diagnostics and research laboratories. During this time, the number of methodologies for preparing DNA libraries for NGS has greatly expanded, while the cost of sequencing has exponentially dropped. In spite of this progress, sequencing of multiple samples in parallel remains costly, mainly due to the way in which multiplexing is achieved. Typically, this is done by indexing libraries prepared from individual samples, followed by pooling together multiple libraries in the same sequencing run. This means that all the steps in the library preparation procedure must be repeated for every sample to be sequenced, which is labor-intensive and multiplies the cost of reagents. Furthermore, accurate normalisation of library concentration is necessary before multiple libraries can be pooled together, which is not always possible and requires additional reagents. In contrast, being able to directly barcode genomic DNA prior to library construction, followed by pooling of differentially barcoded samples into a single library, should enable high levels of multiplexing at much lower cost. An example of application which would greatly benefit from improved solutions for NGS library multiplexing is multi-region sequencing of tumor samples. In this approach, DNA is extracted from multiple regions within the same tumor mass, or from multiple tumor sites in the same patient, and a library is prepared from each region. Multi-region tumor sequencing has been successfully used to assess levels of intratumor heterogeneity and to reconstruct evolutionary tumor histories in different cancer types. One limitation of current multi-region sequencing approaches is the size of the regions examined, which must be sufficiently large to enable the recovery of enough DNA to construct a library from every region separately. This precludes the possibility of examining a larger number of smaller regions, for example within a single tissue section, which would enable assessing intratumor heterogeneity at much higher spatial resolution. This, together with the high cost associated to constructing a single library for every region sampled, currently limits the applicability of multi-region sequencing in routine cancer diagnostics. Several approaches have been developed to barcode genomic DNA (gDNA) as well as to amplify sub-nanogram amounts of gDNA prior to library preparation. Direct incorporation of sequencing adapters into gDNA by engineered transposases allows rapid library

preparation and is the basis of successful commercial solutions such as Nextera from Illumina. However, this approach still requires that individual libraries are generated from each sample, and then pooled together before sequencing. On the other hand, whole-genome amplification methods, such as DOP-PCR, MDA, MALBAC and the more recent SCMDA and LIANTI, achieve direct gDNA barcoding during genome amplification, so that multiple samples can be pooled together into a single multiplexed library. Although such methods are specifically tailored for whole-genome sequencing of single cells, they could, in principle, also be used for other applications, for instance to generate multiplexed libraries for multi-region tumor sequencing in tissue sections. One limitation, however, is that whole-genome amplification requires intact DNA and thus is not compatible with fixed tissue samples, in particular formalin-fixed, paraffin-embedded (FFPE) specimens, which still represent a cornerstone in pathology. Additionally, whole-genome amplification methods are very costly, making them hardly applicable in routine diagnostics. To overcome these limitations, here we develop a method, which we name CUTseq, that combines restriction endonucleases with in vitro transcription, to construct multiplexed DNA libraries for reduced representation genome sequencing of multiple samples in parallel. We show that CUTseq can be used to barcode gDNA extracted from both non-fixed and fixed samples, including old archival FFPE tissue sections. We benchmark CUTseq by comparing it to a widely used method of DNA library preparation, and demonstrate that CUTseq can be used for both genome and exome sequencing, enabling reproducible DNA copy number profiling and single-nucleotide variant calling even in low-input FFPE tissue samples. Lastly, we describe a workflow for rapid and cost-effective preparation of highly multiplexed CUTseq libraries, that can be applied in the context of high-throughput genetic screens and for cell line validation.

Reagents

reagents used only in high-throughput CUTseq

- CUTseq oligonucleotide adapters (for a list of adapters we used, see Supplementary Data)
- Absolute Ethanol (VWR, cat. no. 20816.367)
- Nuclease-free Phosphate-Buffered Saline (10X) pH 7.4 (Thermo Fisher Scientific, cat. no. AM9625)
- Nuclease-free water (Thermo Fisher Scientific, cat. no. 4387936)
- PinPoint Slide DNA Isolation System™ (ZymoResearch, cat. no. D3001)
- Cell Lysis Buffer: Tris-HCl 10mM, NaCl 10mM, EDTA 1mM, Igepal 0.2%, pH 8 @ 25 °C
- Nucleus Break Buffer: Tris-HCl 10mM, NaCl 150mM, EDTA 1mM, SDS 0.3%, pH 8 @ 25 °C
- Tail Buffer: Tris-HCl 10mM, NaCl 100mM, EDTA 50mM, SDS 1%, pH 7.5 @ 25 °C
- Proteinase K, Molecular Biology Grade (NEB, cat. no. P8107S)

- Phenol: Chloroform: Isoamyl Alcohol 25:24:1 (Sigma, cat. no. P2069-100ML)
- Glycogen from mussels (Sigma, cat. no. 10901393001)
- Sodium Acetate, pH 5.5 (Life Technologies, cat. no. AM9740)
- # Vapor-Lock (Qiagen, cat. no. 981611)
- CutSmart® buffer (NEB, cat. no. B7204S)
- HindIII-HF® 20U/ml (NEB, cat. no. R3104L)
- T4 DNA Ligase (Thermo Fisher Scientific, cat. no. EL0014)
- # T4 DNA Rapid Ligase (Thermo Fisher Scientific, cat. no. K1423)
- UltraPure™ BSA (50 mg/ml) (Thermo Fisher Scientific, cat. no. AM2616)
- Adenosine 5'- Triphosphate (ATP) Solution Set (NEB, cat. no. P0756L)
- Alkaline Phosphatase, Calf Intestinal (Promega, cat. no. M1821)
- W&B Buffer: Tris-HCl 10mM, NaCl 1M, Igepal 0.2%, pH 8 @ 25 °C
- MEGAscript® T7 Transcription Kit (Thermo Fisher Scientific, cat. no. AM1334-5)
- DNase I, RNase-free (Thermo Fisher Scientific, cat. no. AM2222)
- RA3 adaptor and RTP, RP1 and RPI primers (custom-made by Integrated DNA Technologies Inc. based on the sequences in the TruSeq Small RNA Library Preparation kit, Illumina)
- RNaseOUT™ Recombinant Ribonuclease Inhibitor (Invitrogen, cat. no. 10777-019)
- T4 RNA ligase 2, truncated (NEB, cat. no. M0242L)
- SuperScript® IV Reverse Transcriptase (Thermo Fisher Scientific, cat. no. 18090200)
- NEBNext Ultrall Q5 PCR Mastermix (NEB, cat. no. M0544L)
- Agencourt RNAClean XP with Scalable throughput (Beckman Coulter, cat. no. A63987)
- Agencourt AMPure XP (Beckman Coulter, cat. no. A63880)
- NEBNext® Ultra™ II DNA Library Prep Kit for Illumina (NEB, cat. no. E7645S)
- SureSelect XT HS Reagent Kit (Agilent, cat. no. G9704)

- Qubit® dsDNA HS Assay Kit (Thermo Fisher Scientific, cat. no. Q32851)
- High Sensitivity DNA Kit (Agilent, cat. no. 5067-4627)

Equipment

- Cell counter (for example, Countess II FL Automated Cell Counter, Thermo Fisher Scientific)
- Incubator (for example, Binder incubator, Model KB 53)
- Tabletop centrifuge (for example, Eppendorf® Microcentrifuge 5424)
- # I-DOT One MC (Dispendix GmbH, Stuttgart, Germany)
- Thermoshaker (for example, Eppendorf® Thermomixer Compact)
- PCR cycler
- Sonication device (for example, ME220 Focused-ultrasonicator, Covaris)
- SpeedVac Vacuum Concentrator (for example, Savant™ SpeedVac™ DNA 130 Integrated Vacuum Concentrator System, Thermo Fisher Scientific, cat. no. DNA130-230)
- DynaMag™-2 Magnet (Thermo Fisher Scientific, cat. no. 12321D)
- Qubit® 2.0 Fluorometer (Thermo Fisher Scientific, cat. no. Q32866)
- Bioanalyzer 2100 (Agilent, cat. no. G2943CA)

Procedure

DAY 1

1. Genomic DNA (gDNA) extraction

Note: the following is a procedure to isolate gDNA from small regions within single FFPE tissue sections using the PinPoint Slide DNA Isolation System™. For tissue sections or cells, any isolation protocol yielding high-quality gDNA, including commercially available silica-based kits, is compatible with the subsequent steps of CUTseq.

1. Apply a little amount of PinPoint glue onto a small region of interest of a FFPE tissue section previously stained with hematoxylin-eosin and imaged

Note: this step is recommended to confirm the presence of tumor cells within the tissue region

2. Air-dry the PinPoint glue for at least 30 min at room temperature
3. Cut the solidified glue using a disposable insulin needle, then transfer the pellet into a LoBind tube pre-filled with 100 µl of tail buffer plus 10 µl of proteinase K
4. Incubate on a thermomixer for 18 hours @ 55 °C, shaking at 800 rpm

DAY 2

1. Add 10 µl of fresh proteinase K
2. Incubate for 1 hour @ 55 °C
3. Transfer the sample to a PCR thermocycler, and incubate for 10 min @ 95 °C to inactivate proteinase K
4. Purify gDNA using phenol-chloroform

Note: in our experience, phenol-chloroform extraction yields gDNA of the highest quality. However, we have also successfully used silica columns provided in the PinPoint Slide DNA Isolation System™

5. Add 120 µl of phenol/chloroform pre-warmed at room temperature, and shake vigorously
6. Centrifuge at ~20,000 x g for 5 min at room temperature
7. Collect the upper phase (~100 µl) and transfer it to a new LoBind tube
8. Add an equal volume of chloroform and shake vigorously
9. Centrifuge at ~20,000 x g for 5min at room temperature
10. Collect the upper phase (~60 µl) and transfer it to a new LoBind tube
11. Add 3.7 µl of glycogen per 100 µl of solution, and vortex vigorously to mix
12. Add Na-Acetate 3M, pH 5.5 to a final concentration of 0.3 M, and vortex vigorously to mix
13. Add 2.5 volumes of ice-cold 100% ethanol, then immediately vortex vigorously to mix
14. Incubate for 16–18 hours @ –80 °C

DAY 3

1. Centrifuge at ~20,000 x g for 1 hour @ 4 °C
2. Gently discard the supernatant
3. Wash the DNA pellet with 500 µl of ice-cold 70% ethanol, by vortexing vigorously
4. Centrifuge at ~20,000 x g for 15 min @ 4 °C
5. Discard the supernatant and wash the pellet again with 500 µl of ice-cold 70% ethanol
6. Centrifuge at ~20,000 x g for 15 min @ 4 °C
7. Discard the supernatant
8. Air-dry the DNA pellet

Note: avoid over-drying the DNA pellet, as this may result in low DNA yield

9. Resuspend the pellet in 8 µl of nuclease-free water

Breakpoint: if needed, at this point the samples can be stored @ –20 °C for several months

DAY 3

2. DNA Digestion

Note: here we describe the use of HindIII to digest gDNA, since this enzyme was used for most of the experiments described in this manuscript. However, different restriction enzymes can be used at this point to digest the purified gDNA. The choice of enzyme depends on the application and desired resolution. For example, for applications that require high genome coverage, such as exome sequencing, a 4-cutter is preferable. On the other hand, 6-cutters are better suited for applications for which shallow genome sequencing is sufficient, for instance DNA copy number profiling at low resolution. Additional aspects to consider when choosing which restriction enzyme to use are the distribution of recognition sites along the genome of interest (see **Supplementary Figure 1**), the sensitivity of different enzymes to methylation in the recognition sequence, as well as the cost of purchasing the enzymes from commercial providers. A list of enzymes that we recommend for CUTseq and are available through NEW ENGLAND BioLabs is provided in **Supplementary Table 1**.

Standard CUTseq

1. Transfer 8 µl of purified gDNA into a 0.5 ml LoBind tube

Note: when processing multiple samples to be pooled into the same library, PCR tube strips or 96-well plates can be used, depending on the number of samples

2. Add 2 µl of digestion mix:

CutSmart buffer 10X 1 µl

HindIII-HF 1 µl

3. Perform the following steps in a PCR thermocycler with the lid set @ 50 °C:

1. 37 °C 16-18 h

2. 80 °C 20 min

3. 4 °C Hold

High-throughput CUTseq

Note: the following protocol was implemented on the I-DOT One nanodispensing system (Dispendix, Germany). Other systems may also be used; however, volumes might have to be adjusted depending on the technical specifications of each instrument.

Dispense with I-DOT

1. Dispense 5 µl of Vapor Lock per well in a 96- or 384-well plate

Note: 384-well plates are better suited when dispensing with nanoliter volumes as the conical bottom and size of the wells allows to more easily visualize the dispensed droplets inside the Vapor Lock phase.

2. @ Add 350 nl per well of gDNA per well

3. @ Add 150 nl per well of digestion mix:

CutSmart buffer 10X 50 nl

HindIII-HF 100nl

4. Perform the following steps in a thermo incubator

1. 37 ° 30 min

2. 80 °C 20 min

3. 4 °C Hold

DAY 4 (DAY 3 for high-throughput CUTseq)

3. Ligation of CUTseq adapters

Standard CUTseq

1. Add 1 µl per sample of 0.33 µM HindIII adapter with the desired barcode sequence
2. Place the sample(s) on ice
3. To each sample, add 19 µl of ligation mix:

Nuclease-free water 12 µl

T4 ligase buffer 10X 3 µl

ATP 10 mM 2.4 µl

BSA 50 mg/ml 0.6 µl

T4 ligase highly conc. 1 µl

4. In a PCR thermocycler with the lid set @ 20 °C, perform the following steps:

1. 16 °C 16-18 h

2. 65 °C 10 min

3. 4 °C Hold

High-throughput CUTseq

1. @ Add 300 nl of 33 nM CUTseq adapter per well

2. @ Add 700 nl of ligation mix per well:

Nuclease-free water 50 nl

T4 ligase buffer 5X 300 nl

ATP 10 mM 120 nl

BSA 50 mg/ml 30 nl

T4 rapid ligase 200 nl

3. Incubate at room temperature for 30 min

4. Pool the volume in multiple wells into one 1.5 ml LoBind tube

Note: in order to be pooled together, multiple wells must have different sample barcodes. The number of wells pooled together depends on the total number of samples, and on the desired complexity of the final library. In general, libraries containing many samples will need to be sequenced deeply in order for each sample to receive a sufficient number of reads. In our experience, we have successfully sequenced libraries containing up to 96 different barcodes in a single NextSeq 500 run, obtaining a number of reads per sample sufficient to perform reproducible DNA copy number profiling (see **Figure 3**).

DAY 5 (DAY 3 for high-throughput CUTseq)

4. DNA cleanup

Note: at this point, the same procedure is followed for single-sample and multiplexed libraries. Below, we describe a standard DNA precipitation cleanup procedure. However, silica-based columns or AMPure XP beads can also be used, according to the manufacturer's instructions.

1. Measure the sample volume, then add the following reagents:

Glycogen 20 mg/ml 3.7 µl per 100 µl

Na-Acetate 3 M, pH 5.5 Up to 0.3 M final conc.

2. Add 2.5 volumes of ice-cold 100% ethanol
3. Vortex 5 seconds at max. power
4. Incubate for 18 hours @ -80 °C

Note: incubation @ -20 °C is also fine

DAY 6 (DAY 4 for high-throughput CUTseq)

5. Centrifuge at ~20,000 x g for 1 hour @ 4 °C
6. Discard the supernatant
7. Wash the DNA pellet with 500 µl of ice-cold 70% ethanol
8. Centrifuge at ~20,000 x g for 15 min @ 4 °C
9. Discard the supernatant and wash the pellet again with 500 µl of ice-cold 70% ethanol

10. Centrifuge at ~20,000 x g for 15 min @ 4 °C

11. Discard the supernatant

12. Air-dry the pellet

Note: avoid over-drying the DNA pellet, as this may result in low DNA yield

10. Resuspend the pellet in 50 µl of nuclease-free water

Breakpoint: if needed, at this point the samples can be stored at –20 °C for several months

DAY 6 (DAY 4 for high-throughput CUTseq)

5. Sonication

Note: in our experience, DNA fragments of 200–300 bp are needed to obtain high-quality libraries and sequencing results. To sonicate gDNA, we typically use a Covaris ME220 Focused-ultrasonicator with microTUBE-50 AFA Fiber Screw-Cap tubes or microTUBE-15 AFA Beads Screw-Cap tubes, with a target peak of 200 bp. Note that we sonicate gDNA independently of the restriction enzyme used, as well as gDNA extracted from FFPE samples, since we have found that this results in higher-quality libraries.

6. SpeedVac

Transfer sample from covaris tube to new 0.5 ml DNA LoBind tube, proceed to SpeedVac Vacuum Concentrator with a normal heating mode till the moment when the volume of sample is entirely evaporated.

DAY 6 (DAY 4 for high-throughput CUTseq)

7. *In vitro* transcription

1. Resuspend the sample with 8 µl nuclease-free water

2. Add the following reagents on ice:

rATP+rUTP+rGTP+rCTP* 8 µl

T7 polymerase buffer 10X 2 µl

T7 polymerase 1.5 µl

RNaseOUT™ Recombinant Ribonuclease Inhibitor 40U 0.5 µl

*Prepared from separate rNTP solutions provided with the MEGAscript® T7 Transcription Kit

3. Incubate for 14 hours @ 37 °C in a PCR thermocycler with the lid set @ 70 °C

DAY 7 (DAY 5 for high-throughput CUTseq)

8. RNA cleanup

1. Add 1 µl of DNase I RNase-free to the IVT product
2. Incubate for 15 min @ 37 °C
3. Bring up the volume to 30 µl by adding 9 µl nuclease-free water, then mix with 54 µl (1.8x) of RNAClean XP beads pre-warmed at room temperature
4. Mix thoroughly and incubate for 10 min at room temperature
5. Place the sample on a magnetic stand
6. Incubate for at least 5 min until the liquid appears clear
7. Remove and discard the supernatant
8. Wash the beads twice with 200 µl of freshly prepared 70% ice-cold ethanol
9. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

10. Remove the sample from the magnetic stand
11. Resuspend the beads in 8 µl of nuclease-free water
12. Incubate for 2 min at room temperature
13. Place the sample back on the magnetic stand
14. Incubate for at least 5 min until the liquid appears clear
15. Transfer 7.8 µl of supernatant to a new 0.5 µl DNA LoBind tube

Note: the following steps until section 11. are adapted from the TruSeq Small RNA Library Preparation protocol from Illumina. *Steps on ice.

DAY 7 (DAY 5 for high-throughput CUTseq)

9. RA3 adapter ligation

1. * Add 1 µl of 10 µM RA3 adapter
2. Incubate for 2 min @ 70 °C in a PCR thermocycler, then immediately place sample on ice
3. * Add 3.2 µl of the following mix:

RNA ligase buffer 1.2 µl

RNaseOUT™ Recombinant Ribonuclease Inhibitor 40U 1 µl

T4 RNA ligase truncated 1 µl

4. Incubate for 2 hours @ 25 °C in a PCR thermocycler with the lid set @ 30 °C

DAY 7 (DAY 5 for high-throughput CUTseq)

10. Reverse transcription (1st strand synthesis)

1. * Add 2 µl per sample of RTP primer
2. In a PCR thermocycler, incubate for 2 min @ 70 °C
3. * Quickly transfer the sample to ice
4. * Add 11 µl of the following mix:

1st stand buffer 4 µl

dNTPs @ 25 mM 1 µl

100 mM DTT 2 µl

RNaseOUT™ Recombinant Ribonuclease Inhibitor 40 U 2 µl

SuperScript IV reverse transcriptase 2 µl

5. Incubate for 1 hour @ 50°C in a PCR thermocycler with the lid set @ 50 °C

Breakpoint: if needed, at this point the samples can be stored @ –20 °C for several months

DAY 7 (DAY 5 for high-throughput CUTseq)

11. Library indexing and amplification

1. * Add 4 µl per sample of the desired indexed Illumina primer

2. * Add 71 µl of the following mix:

Nuclease-free water 17 µl

NEBNext® Ultra™ II PCR Master Mix 50 µl

RP1 primer 4 µl

3. In a PCR thermocycler perform the following cycles:

1. 98 °C 30 sec

2. 98 °C 10 sec

3. 65 °C 75 sec

GOTO step 2, 8-20 times

4. 65 °C 5 min

5. 4 °C Hold

Note: the number of PCR cycles needs to be adjusted depending on the gDNA input in the IVT reaction. The following table shows the recommended number of PCR cycles for various gDNA inputs, based on our experience:

Input gDNA (ng)

PCR cycles

500-700 ng gDNA input: 6 cycles

300-400 ng gDNA input: 7 cycles

100-200 ng gDNA input: 8 cycles

50-100 ng gDNA input: 9 cycles

30 ng gDNA input: 10 cycles

15 ng gDNA input: 11 cycles

7.5 ng gDNA input: 12 cycles

3.8 ng gDNA input: 13 cycles

1.9 ng gDNA input: 14 cycles

DAY 7 (DAY 5 for high-throughput CUTseq)

12. Library size selection and purification

1. Transfer each sample into a new 1.5 ml LoBind tube
2. Add 70 μ l (0.7x) of AMPure XP beads pre-warmed at room temperature
3. Mix thoroughly and incubate for 10 min at room temperature
4. Place the sample on a magnetic stand
5. Incubate for at least 5 min until the liquid appears clear
6. Carefully collect the supernatant and transfer it to a new 1.5 ml LoBind tube
7. Add 20 μ l (0.2x) of AMPure XP beads pre-warmed at room temperature
8. Mix thoroughly and incubate for 5 min at room temperature
9. Place the sample on a magnetic stand
10. Remove and discard the supernatant
11. Wash the beads twice with 200 μ l of freshly prepared 80% ice-cold ethanol
12. Air-dry the beads at room temperature

Note: do not dry the beads for more than 5–8 min, since this may result in low DNA yield

13. Remove the sample from the magnetic stand

14. Resuspend the beads in 10–20 µl of nuclease-free water
15. Incubate for 2 min at room temperature
16. Place the sample back on the magnetic stand
17. Incubate for at least 5 min until the liquid appears clear
18. Transfer the supernatant to a new 1.5 µl DNA LoBind tube
19. Check the library concentration using Qubit
20. Check the library quality and size on a Bioanalyzer using a high sensitivity DNA kit
21. Store the library @ –20 °C

Troubleshooting

Time Taken

Anticipated Results

Tagged cut sites by CutSeq will generate high quality sequencing libraries with fragment size of 200-300 bp, that can be checked on Bioanalyzer 2100 (Agilent Technologies).

References

1. Reuter, J. A., Spacek, D. V. & Snyder, M. P. High-throughput sequencing technologies. *Mol. Cell* 58, 586–597 (2015).
2. Goodwin, S., McPherson, J. D. & McCombie, W. R. Coming of age: ten years of next-generation sequencing technologies. *Nat. Rev. Genet.* 17, 333–351 (2016).
3. McGranahan, N. & Swanton, C. Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future. *Cell* 168, 613–628 (2017).
4. Telenius, H. et al. Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13, 718–725 (1992).

5. Dean, F. B. et al. Comprehensive human genome amplification using multiple displacement amplification. *Proc. Natl. Acad. Sci.* 99, 5261–5266 (2002).
6. Zong, C., Lu, S., Chapman, A. R. & Xie, X. S. Genome-Wide Detection of Single Nucleotide and Copy Number Variations of a Single Human Cell. *Science* 338, 1622–1626 (2012).
7. Dong, X. et al. Accurate identification of single-nucleotide variants in whole-genome-amplified single cells. *Nat. Methods* 14, 491–493 (2017).
8. Chen, C. et al. Single-Cell Whole Genome Analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science* 356, 189–194 (2017).
9. Kivioja, T. et al. Counting absolute numbers of molecules using unique molecular identifiers. *Nat. Methods* 9, 72–74 (2011).
10. Ullah, I. et al. Evolutionary history of metastatic breast cancer reveals minimal seeding from axillary lymph nodes. *J. Clin. Invest.* 128, 1355–1370 (2018).
11. Forbes, S. A. et al. COSMIC: somatic cancer genetics at high-resolution. *Nucleic Acids Res.* 45, D777–D783 (2017).
12. Nik-Zainal, S. et al. Landscape of somatic mutations in 560 breast cancer whole-genome sequences. *Nature* 534, 47–54 (2016).
13. Kandoth, C. et al. Mutational landscape and significance across 12 major cancer types. *Nature* 502, 333–339 (2013).
14. Baird, N. A. et al. Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *PLOS ONE* 3, e3376 (2008).
15. Davey, J. W. et al. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* 12, 499–510 (2011).
16. Kosicki, M., Tomberg, K. & Bradley, A. Repair of double-strand breaks induced by

- CRISPR–Cas9 leads to large deletions and complex rearrangements. *Nat. Biotechnol.* 36, 765–771 (2018).
17. Adikusuma, F. et al. Large deletions induced by Cas9 cleavage. *Nature* 560, E8 (2018).
18. Masters, J. R. Cell-line authentication: End the scandal of false cell lines. *Nature* 492, 186 (2012).
19. Sommer, C., Straehle, C., Köthe, U. & Hamprecht, F. A. Ilastik: Interactive learning and segmentation toolkit. in 2011 IEEE International Symposium on Biomedical Imaging: From Nano to Macro 230–233 (2011). doi:10.1109/ISBI.2011.5872394
20. Schindelin, J. et al. Fiji: an open-source platform for biological-image analysis. *Nat. Methods* 9, 676–682 (2012).
21. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *ArXiv13033997 Q-Bio* (2013).
22. Smith, T., Heger, A. & Sudbery, I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. *Genome Res.* 27, 491–499 (2017).
23. Rausch, T., Fritz, M. H.-Y., Korbel, J. O. & Benes, V. Alfred: Interactive multi-sample BAM alignment statistics, feature counting and feature annotation for long- and short-read sequencing. *Bioinforma. Oxf. Engl.* (2018).
24. Scheinin, I. et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res.* 24, 2022–2032 (2014).
25. Lai, Z. et al. VarDict: a novel and versatile variant caller for next-generation sequencing in cancer research. *Nucleic Acids Res.* 44, e108 (2016).
26. Cock, P. J. A. et al. Biopython: freely available Python tools for computational molecular

biology and bioinformatics. *Bioinforma. Oxf. Engl.* 25, 1422–1423 (2009).

27. Quinlan, A. R. & Hall, I. M. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinforma. Oxf. Engl.* 26, 841–842 (2010).

Acknowledgements

We thank Bastiaan Spanjaard and Alexander van Oudenaarden for helping during the initial stages of CUTseq implementation. This work was supported by a postdoctoral scholarship from the Swedish Society for Medical Research (SSMF) to X.Z.; by a Mobilnosc Plus scholarship from the Polish Ministry of Science and High Education (1622/MOB/V/2017/0) to M.N.; by grants from Italian Ministry of Health, Ricerca Corrente 2017, AIRC 5xMille MCO Extension Program (# 9970), FPRC 5xMille Ministero Salute 2014 and FPRC 5xMille Ministero Salute 2013 to A.S.; by additional grants from SSMF, the Swedish Cancer Fund, Stockholm Cancer Fund and Stockholm County Council to J.H.; by grants from the Science for Life Laboratory, the Swedish Research Council (621-2014-5503), and the Ragnar Söderberg Foundation to M.B.; and by grants from the Swedish Research Council (521-2014-2866), the Swedish Cancer Research Foundation (CAN 2015/585), the Ragnar Söderberg Foundation, the Swedish Foundation for Strategic Research (BD15-0095), and the Strategic Research Programme in Cancer (StratCan) at Karolinska Institutet to N.C.

Author contributions

Conceptualization: X.Z., M.B., N.C. Clinical samples: C.M., T.V., A.S. Data curation: X.Z., S.G., M.S. Formal analysis: S.G., L.H. Funding acquisition: N.C., M.B. Investigation: X.Z., M.S., M.N., C.M., T.V. Methodology: X.Z., N.C., M.B. Project administration: N.C. Software: S.G. Supervision: N.C., M.B. Validation: X.Z., M.N., M.S., R.M. Visualization: S.G., L.H., X.Z., M.S., N.C., M.B. Writing: N.C., X.Z.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [supplement1.docx](#)
- [supplement2.xlsx](#)