



Targeted Locus Amplification and Haplotyping

Juliet W. Lefferts, Vera Boersma, Marne C. Hagemeijer,
Karima Hajo, Jeffrey M. Beekman, and Erik Splinter

Abstract

Targeted locus amplification (TLA) allows for the detection of all genetic variation (including structural variation) in a genomic region of interest. As TLA is based on proximity ligation, variants can be linked to each other, thereby enabling allelic phasing and the generation of haplotypes. This allows for the study of genetic variants in an allele-specific manner. Here, we provide a step-by-step protocol for TLA sample preparation and a complete bioinformatics pipeline for the allelic phasing of TLA data. Additionally, to illustrate the protocol, we show the ability of TLA to re-sequence and haplotype the complete cystic fibrosis transmembrane (CFTR) gene (> 200 kb in size) from patient-derived intestinal organoids.

Key words Targeted Locus Amplification (TLA), Haplotyping, Phasing, Genetic variation, Proximity ligation, Next generation sequencing, Organoids, Cystic fibrosis

1 Introduction

Targeted locus amplification (TLA) technology enables the targeted and complete sequencing of any gene locus of interest. TLA uses proximity ligation to connect DNA fragments that originate from the same locus and allele, as these reside in physical proximity in the nucleus (Fig. 1). This enables the detection of genetic variation in a region of interest without detailed prior locus information such as parental sequences [1]. TLA enables the detection of all genetic variations, including single-nucleotide variants (SNVs) and structural variations such as translocations, inversions, deletions, and insertions. Coding and non-coding regions up to hundreds of kilobases in size can be enriched and analyzed. Because TLA is based on proximity ligation, allelic variation can be linked and haplotypes can be reconstructed enabling the analysis of genomic regions in an allele-specific manner. This is particularly

Juliet W. Lefferts and Vera Boersma contributed equally.

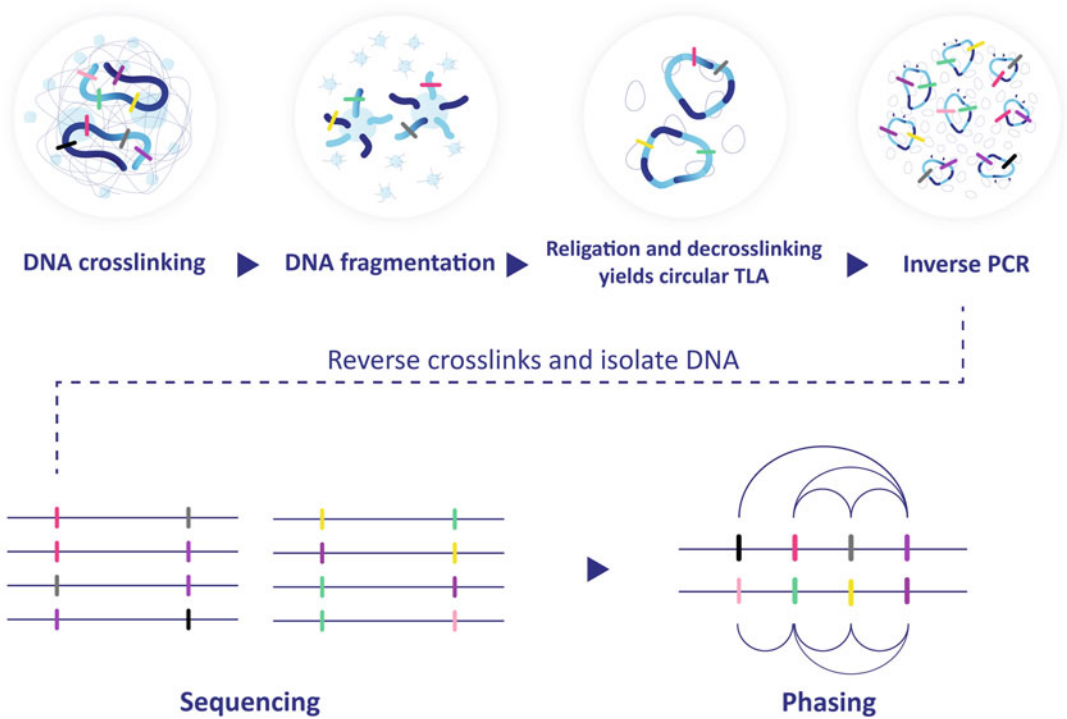


Fig. 1 Schematic overview of the TLA haplotyping workflow. First DNA is crosslinked and fragmented followed by re-ligation and subsequent decrosslinking. The DNA obtained is circularized and amplified by inverse PCR using primers specific for the genetic region of interest. PCR products are sequenced and aligned to a reference genome. Finally allelic phasing is achieved through the linking of heterozygous SNPs

important in the context of studying genetic contributions to disease, as it provides information about the *cis-trans* relationship of genetic variants [1, 2].

During TLA sample preparation, DNA is first crosslinked, followed by fragmentation and re-ligation into random circular DNA fragments. The initial DNA crosslinking step enables the ligation of DNA fragments that were originally in close physical proximity (i.e., the same allele). Ligation products are amplified by inverse PCR, using primers specific for the point or region of interest. PCR products are subsequently sequenced using next-generation sequencing (NGS) or long-read sequencing. Sequencing data is aligned to a reference genome and subsequent linking of heterozygous single nucleotide polymorphisms (SNPs) between sequenced DNA fragments allows for the separation of haplotypes (Fig. 1) [1].

Here, we provide an updated and detailed version of the previously published TLA protocol [3] to enable haplotyping of the complete cystic fibrosis transmembrane conductance regulator (CFTR) gene from cystic fibrosis (CF) patient-derived intestinal organoids. The protocol includes both the sample processing and the bioinformatics pipeline that were applied. CF is a recessive,

monogenic disease and associated with complex *cis-trans* relationships within the CFTR gene and a heterogeneous disease phenotype and therapeutic response. Patient-derived intestinal organoids are a robust model system for studying biological intestinal functions, as has been previously shown for CF and other diseases [4]. The TLA protocol developed here enables haplotyping of genes using a limited amount of organoid starting material (0.5×10^6 cells) and a multiplex PCR strategy to haplotype a relatively large genomic region. Using this procedure, we were able to re-sequence and haplotype the >200 kb genomic region of the CFTR gene. This shows that TLA can haplotype large genetic regions using patient-derived intestinal organoids, which is important for future studies into the relations between large haplotypes and biological functions.

2 Materials

2.1 Equipment

1. 1.5 mL and 2 mL low-binding microcentrifuge tubes (MAX-YMum Recovery™ PCR Tubes, Axygen Scientific).
2. Magnetic rack for Eppendorf Tubes.
3. Microcentrifuge.
4. Tube roller bank.
5. Thermal shaker for Eppendorf Tubes.
6. PCR machine.
7. Agarose gel electrophoresis system.
8. Qubit™ Fluorometer.

2.2 Reagents

1. 10 μM Rho kinase inhibitor (Y27632).
2. 1× TrypLE Express Enzyme (Gibco).
3. 10% FCS/PBS.
4. 37% Formaldehyde solution.
5. 2M Glycine.
6. 1× Smart Cut Buffer (New England Biolabs).
7. 10% SDS in Milli-Q water.
8. 10% Triton X-100.
9. 10 U/μl NlaIII.
10. 10× Ligation buffer: 660 mM TRIS-HCL, 50 mM MgCl₂, 50 mM DTT, and 10 mM ATP, pH 7.5.
11. 5 U/μl T4 DNA Ligase.
12. 10 mg/ml Proteinase K.
13. AMPure XP Beads (Beckman Coulter).

14. 10 mM TRIS, pH 7.5.
15. 10× Smart Cut Buffer (New England Biolabs).
16. 10 U/ μ l NspI.
17. NucleoMag 96 PCR beads (Bioké).
18. 2× Q5 high-fidelity polymerase Master Mix (New England Biolabs).
19. Qubit Broad Range (BR) Assay Kit (ThermoFisher Scientific).

3 Methods

3.1 Organoid Sample Collection

Organoid cultures are maintained following a standard culturing protocol [5]. Fully grown 7-day-old cultures are used. Typically, 12 wells of a 24-well plate will be sufficient for sample input. Organoids are dissociated into single cells before TLA sample preparation is performed.

1. Wash organoids with PBS and collect them in 5 mL TrypLE solution to dissociate into single cells by incubating at 37 °C for 5 min.
2. After 5 min put samples on ice to stop cell dissociation (*see Note 1*).
3. Centrifuge organoids and discard the supernatant. Resuspend cell pellets in freezing medium supplemented with rho-kinase inhibitor to ensure cell viability. Samples can be stored at –80 °C for short-term storage or in liquid nitrogen for long-term storage until TLA sample preparation.

3.2 TLA Sample Preparation

In the following steps, TLA sample preparation is performed (Fig. 2). All steps in this protocol are designed for $0.5\text{--}2 \times 10^6$ cells per sample and are executed at room temperature (18–21 °C) if not indicated otherwise. Please note that all steps prior to PCR amplification are performed *with low-retention plastics* to increase DNA yield (*see Note 2*).

3.2.1 Fixation

In this step, viable cells are washed to remove potential contaminants and crosslinked to connect genetic sequences that are in close proximity to each other.



Fig. 2 Overview of the TLA sample preparation timeline

1. Thaw a vial containing $0.5\text{--}2 \times 10^6$ cells in a $37\text{ }^\circ\text{C}$ water bath.
2. Centrifuge for 2 min at $250 \times g$.
3. Remove supernatant and resuspend the pellet in $100\ \mu\text{l}$ 10% FCS/PBS by pipetting up and down (*see Note 3*). Transfer the cells to a $0.2\ \text{ml}$ low-binding microcentrifuge tube.
4. Centrifuge for 2 min at $250 \times g$.
5. Remove supernatant and resuspend the pellet in $80\ \mu\text{l}$ 10% FCS/PBS.
6. Add $4.3\ \mu\text{l}$ 37% formaldehyde, mix immediately by pipetting five times up and down using a $60\ \mu\text{l}$ sample volume on a hand pipettor. Incubate *exactly* 10 min.
7. Add $15\ \mu\text{l}$ 2 M glycine and mix by inverting the tube $3\times$.
8. Add $50\ \mu\text{l}$ AMPure beads and incubate for 15–30 min at RT while rotating (*see Note 4*).
9. Place the tube in a magnetic rack for ≥ 1 min until the beads are separated.
10. Invert the magnetic rack four times to collect the beads from the lid and incubate for 1 min.
11. Remove the supernatant.
12. Remove the tubes from the magnetic rack and add $180\ \mu\text{l}$ freshly made 80% EtOH and resuspend the beads.
13. Place in the magnetic rack for ≥ 1 min until the beads are separated.
14. Invert the magnetic rack four times to include the beads from the lid, incubate for 1 min, and remove the supernatant.
15. Repeat **Steps 12–14** for two washes total.
16. Remove all supernatant and resuspend the pellet in $50\ \mu\text{l}$ $1\times$ Smart Cut Buffer. Mix by pipetting up and down until a homogeneous suspension is obtained.

**3.2.2 Permeabilization
and First Restriction
Digestion (RE1)**

In this step, the cells are permeabilized to allow the restriction enzyme to reach the DNA.

1. Add $1.5\ \mu\text{l}$ 10% SDS to the suspension and mix by pipetting using a $40\ \mu\text{l}$ sample volume on a hand pipettor. Incubate for 30 min at $65\text{ }^\circ\text{C}$ while shaking at 900 RPM.
2. Spin briefly to collect the sample in the bottom of the tube and add $15\ \mu\text{l}$ 10% Triton X-100.
3. Mix well by pipetting using a $50\ \mu\text{l}$ sample volume on a hand pipettor until a homogeneous suspension is obtained. Incubate 30 min at $37\text{ }^\circ\text{C}$ while shaking at 900 RPM.
 - (a) Optional quality control step: take $5\ \mu\text{L}$ of sample for quality control of the undigested sample (*see Subheading 3.2.9*).

4. Spin briefly to collect the sample in the bottom of the tube and add 1 μL NlaIII. Mix by pipetting using a 50 μL sample volume on a hand pipettor until a homogeneous suspension is obtained.
5. Incubate overnight at 37 °C while shaking at 900 RPM.

3.2.3 Heat Inactivation and Ligation

In this step, RE1 is heat inactivated and DNA fragments that are connected via crosslinking are ligated.

1. Briefly spin the sample to collect the liquid from the lid and inactivate NlaIII by incubating for 25 min at 65 °C.
 - (a) Optional quality control step: take 5 μL of sample to check if the sample is digested sufficiently (*see* Subheading 3.2.9).
2. Cool the sample down to room temperature by incubation on ice for 5 min and add 10 μL 10 \times Ligation buffer, 22.5 μL MilliQ water, and 1 μL T4 DNA Ligase.
3. Mix by pipetting five times using a 90 μL sample volume on a hand pipettor and incubate at room temperature for a minimum of 1 h.
 - (a) Optional quality control: take 10 μL of sample to check if the digested sample is religated (*see* Subheading 3.2.9).

3.2.4 Reverse-Crosslinking

In this step, the crosslinks of the DNA are reversed.

1. Add 1 μL Proteinase K to the sample and reverse-crosslink overnight at 65 °C, while shaking at 900 RPM.

3.2.5 Purification 1

In this step, the DNA is purified to facilitate the subsequent steps in the protocol.

1. Add 2.5 μL of thoroughly mixed NucleoMag 96 PCR beads and 100 μL 2-Propanol to the 100 μL ligation sample.
2. Mix well and incubate for 15 min while agitating. Alternatively, the sample can be inverted three times every 3 min during incubation (*see* **Note 5**).
3. Place the tube in a magnetic rack for ≥ 1 min, until the beads are separated.
4. Invert the magnetic rack four times to include the beads from the lid and incubate for 1 min or until the beads are separated.
5. Remove the supernatant.
6. Remove the magnet, add 180 μL freshly made 80% EtOH, and resuspend the beads.
7. Place on the magnet for ≥ 1 min, until the beads are separated.
8. Invert the magnetic rack four times to include the beads from the lid, incubate for 1 min, and remove the supernatant.
9. Repeat **Steps 6–8** for two washes total.

10. Air-dry the beads for 10 min on the magnet and remove residual EtOH.
11. Air-dry the beads for an additional 5 min.
12. Remove the magnet and resuspend the beads in 52 μ l of 10 mM Tris-HCl pH 7.5.
13. Place on the magnet for ≥ 1 min, until the beads are separated, and transfer 50 μ l of the eluted sample to a clean 0.2 ml tube.
14. Use 1–2 μ l of the eluted sample to measure the DNA concentration using the Qubit® Fluorometer. Expect a total yield of ~ 2 μ g DNA (*see* **Note 6**).

3.2.6 Second Restriction Digestion (RE2)

In this step, the large DNA molecules are trimmed to a length that is better suitable for PCR amplification.

1. Add 10 μ l 10 \times Smart Cut Buffer and 1 μ l NspI to 2 μ g of isolated DNA sample. Adjust the volume to 100 μ l with MilliQ water. Note: use the whole eluted sample if less than 2 μ g DNA was isolated.
2. Mix by pipetting five times using 90 μ l sample volume on a hand pipettor and incubate for at least 1 h at 37 $^{\circ}$ C.

3.2.7 Ligation

In this step, the DNA is circularized.

1. Briefly spin the tube to collect the liquid from the lid and inactivate RE2 by incubating at 65 $^{\circ}$ C for 25 min.
2. Briefly spin the tube to collect the liquid from the lid and transfer the sample to a 1.5 ml low retention tube and add 40 μ l 2 \times Ligation buffer, 259 μ l MilliQ water, and 1 μ l T4 DNA Ligase.
3. Mix by inverting the sample five times and incubate for at least 1 h at room temperature.

3.2.8 Second Purification

In this step, the DNA is purified to facilitate the subsequent steps in the protocol.

1. Add 10 μ l thoroughly mixed NucleoMag 96 PCR beads and 400 μ l 2-Propanol to the 400 μ l ligation sample.
2. Mix by inverting the sample five times and incubate for 1 h while agitating. Alternatively, invert the sample three times every 10 min during the incubation.
3. Place the tube in a magnetic rack for ≥ 1 min, until the beads are separated.
4. Invert the magnetic rack four times to include the beads from the lid and incubate for 1 min.
5. Remove the supernatant.
6. Remove the magnet, add 900 μ l of freshly made 80% EtOH, and resuspend the beads.

7. Place on the magnet for ≥ 1 min, until the beads are separated.
8. Invert the magnetic rack four times to include the beads from the lid, incubate for 1 min, and remove the supernatant.
9. Repeat **Steps 6–8** for two washes total.
10. Air-dry the beads for 15 min on the magnet, remove residual EtOH after 10 min of incubation.
11. Remove the magnet and resuspend the beads in 52 μL 10 mM TRIS pH 7.5.
12. Place on the magnet for ≥ 1 min, until the beads are separated.
13. Transfer and pool 50 μL of the eluted TLA template per sample in a clean 1.5 ml tube.
14. Use 1–2 μL of the eluted sample to measure the DNA concentration using the Qubit®. Expect a final yield of 0.8–1.5 μg DNA.

3.2.9 Quality Control Samples

To assess the quality of the TLA sample preparation, the control samples “Undigested control,” “Digestion control,” and “Ligation control” can be collected at Subheadings 3.2.2 and 3.2.3 where indicated in the TLA protocol, respectively.

1. Adjust volumes of the quality control samples to 20 μL with MilliQ water.
2. Add 1 μL Proteinase K to the control samples and incubate for at least 1 h at 65 °C.
3. Add 3 μL of 5 \times loading buffer.
4. Run the sample on a 0.6–1% agarose gel, at 100 V for 45–60 min.

Expect the undigested control sample to appear as a single >10 kb band, the digestion control sample as a smear from 0.3–2 kb, and the ligation control sample to run >5 kb (Fig. 3). In our experience, it is advisable to only proceed from here on if quality control samples are as expected. However, with low DNA quantities, the digested control can be difficult to visualize on a gel. If quality controls are not as expected, see **Note 7** for troubleshooting.

3.3 PCR

3.3.1 Primer Design

Primers are designed for inverse PCR within a NlaIII digestion fragment with either the forward or reverse primer approximately 50 bp upstream of a common SNV, which itself should be approximately 30 bp from an NlaIII digestion site. For CFTR-specific primer sets, see Table 1.

1. Prepare the Multiplex Primer mix by combining 500 μL MilliQ water and 10 μL of each 100 μM primer.

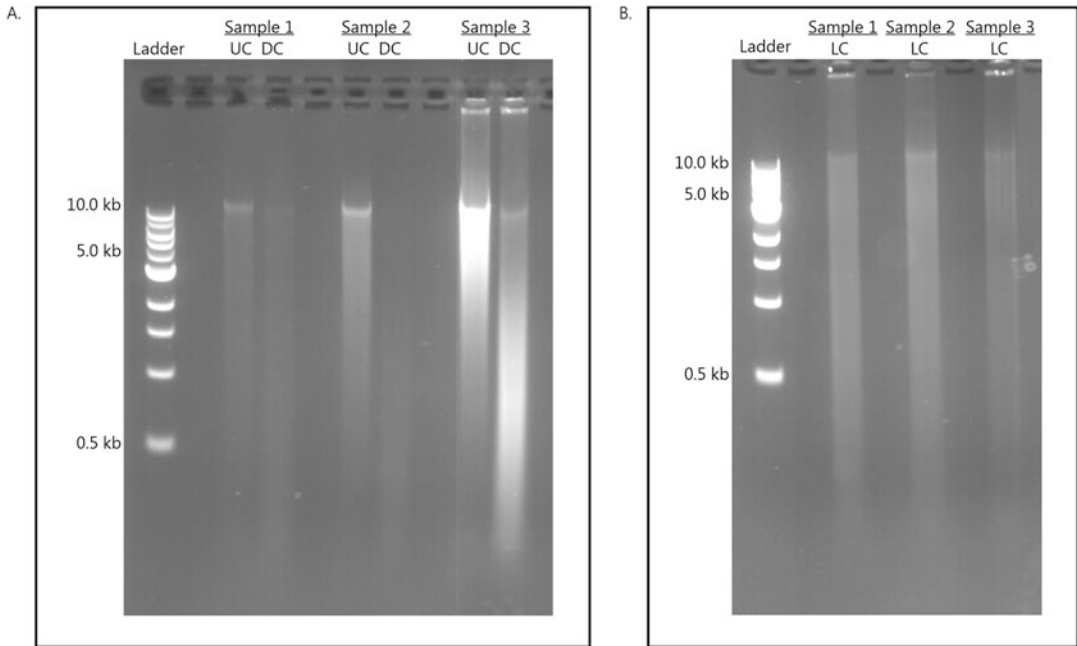


Fig. 3 Gel images of TLA quality controls. (a) the undigested (UC) and digested control (DC). (b) Ligation control (LC) of three example samples. Undigested controls show a single >10 kb band, while the digested controls show a smear between 0.3 and 2 kb. Digested smears can be difficult to visualize when imaging samples with low sample input. For ligation controls, the digested smear migrates to a higher weight, showing a band >5 kb

3.3.2 TLA Multiplex PCR

In this step, the TLA Multiplex PCR amplification is performed.

1. Set up the TLA PCR reaction by mixing 400 ng TLA Template, 50 μ l Q5 High-Fidelity 2 \times Master Mix, 5 μ l Multiplex Primer mix, and MilliQ water to a final volume of 90 μ l.
2. Mix by pipetting five times using 90 μ l sample volume on a hand pipettor and divide the sample over four PCR wells with 25 μ l in each well.
3. Seal and quick spin the PCR plate.
4. Run the following (Touch-Down) PCR program:
 - Step 1: 98 $^{\circ}$ C for 30 s.
 - Step 2: 98 $^{\circ}$ C for 10 s.
 - Step 3: 67 $^{\circ}$ C- $>$ 59.5 $^{\circ}$ C (-0.5 $^{\circ}$ C/cycle) for 30 s.
 - Step 4: 65 $^{\circ}$ C for 3.5 min.
 Repeat to Step 2 for 15 additional cycles, 16 cycles in total.
 - Step 5: 65 $^{\circ}$ C for 5 min.
 - Step 6: 12 $^{\circ}$ C forever.

Table 1
CFTR-specific TLA primer sets

1	CCTATTTGCAATCAGTGTGCATCATG	ATATGCAACCTTCTCCATTATCAAAAATATT
2	ACTGGCTTGAGGTAGGCATG	ACCTAGAATGTTTAAATGCACAGTAAGG
3	GGAGGCAGAAGAAGGATTGTGA	ATTCTGCCTTATCCTCAACAGTTAGAA
4	AGAAGATTAAACAAGTCTTGAATCTGTC TC	CAGTGTCTCCCAGAGAAAATGCC
5	TCCTTGAATGGTAACAGCTCCAT	AAATTAGCACTCCGAAAACCCC
6	ATAAACTGCATTTTTAGTTTGTAACATG	TTCGGTATGAAAAATTCTAATAATAAGA TCATCTGCAA
7	ATATTTTCAGTAGATATAGCAAACAG TTAACATG	TGTTTTTTCACATTTATCCCTATTGATAC TTGT
8	CCATTGATACAGAAAGTAATA TTACACCATC	CACAAGTCACTGCTTAGGATAAAGG
9	TTGTTTATTGCTCCAAGAGAGTCATA	TTTGTATTGGTTGCCAGCAATT
10	AGTATTATGATAGAGGTTACAGAGAAGG	TTCTGTTCTCAGTTTTCTGGATTATG
11	CATTCCCCTGATGAAAGCTCTC	TATATCTGACTACAAGGAGAAAGACTAA TAGGA
12	ATGTTGTGGTTAGAAGTTTATGAAATA TTGTTATTCA	GCTAGAACACTAAGCACTACGC
13	TCTTCCTCCTCAGCTGACTCAACAT	AAAGGTCAACTGTACAATAACTTGGCTA
14	GGATTCTAACTAAGGCAATTTGTCTAC	CCAAAGTGCCAAAAACCTAGC
15	AGTAACCTTACTTATATCTCAAGTACA TAGGATTCT	CCTCACAAAACCCAACATCTTCA
16	TTTAGGCTCAGGACTGAGAAATTCA	CCACAACACTACCCTGGAAGGTATTTAAA
17	AGAACATTCTTACCTCTGCCAG	GGGATGTATACATATATATGCACACAC
18	GCTGGATCTCTGTCTTATCCTGTT	AGGACACCCCTGGGTATGT
19	ATGTGAATTTAGATGTGGGCATG	CAACTGTGATTTTTGTAGTTCTTTTTATA TTTAGATC
20	ATCTATTTAATTCTTTGCCCAACT	CCGAACACAGAAGCTAAAGAG
21	CTGTTGATTCGTGATTTCAATTTCATG	GAATACTATGAATAACTCTACGTTGACATA TTTG
22	CGCCAGGTATTTTAGGGTACTCC	GGTTCATTGATGGTGCTTTCTTGC
23	GCCAAAACGTTGAAGCAACC	TCCTCGGCGTTTTAGAATTTGGC
24	CCTTTATAAACAGGATCATATCCTTTTCGC	TGAGCCCATTTGAATCACCATC
25	TGACCAAGTGACAGTCACTGAG	CAGGGACCTTGTCCTCTGG
26	ACTATTGACTTTTCCCAGACA	ACTATTGACTTTTCCCAGACA
27	CAAGTTGTTTGAGATCTTGT	TAACGTTTTGACAACCAACA
28	AGCTTTGAAAGAGGAGGATT	CTGACCTACTTCCCTCTCAT
29	CCACTGGTGACAGGATAAAA	AGGAAAAGCAAAGAGTGTCT

3.3.3 TLA Multiplex PCR Cleanup

In this step, the PCR products are purified.

1. Pool the four 25 μ l PCR reactions into a clean 1.5 ml Safe-Lock tube.
2. To the 100 μ l PCR sample add 100 μ l isopropanol and 5 μ l NucleoMag PCR beads, mix by inverting the sample five times, and incubate 15 min while agitating.
3. Place the tube in a magnetic rack for ≥ 3 min, until the beads are separated.
4. Invert the magnetic rack four times to include the beads from the lid and incubate for 1 min.
5. Remove the supernatant.
6. Remove the magnet, add 900 μ l freshly made 80% EtOH, and resuspend the beads.
7. Place on the magnet for ≥ 1 min, until the beads are separated.
8. Invert the magnetic rack four times to include the beads from the lid, incubate for 1 min, and remove the supernatant.
9. Repeat **Steps 6–8** for two washes total.
10. Remove residual EtOH and air-dry the beads for 5 min.
11. Remove the magnet and resuspend the beads in 320 μ l 10 mM TRIS pH 7.5 (10 μ l/primer set +10%).
12. Place on the magnet for ≥ 1 min, until the beads are separated.
13. Transfer 310 μ l of the eluted PCR product to a clean 1.5 ml tube.
14. Samples are now ready for the second PCR amplification.

3.3.4 TLA PCR 2

In this step, the second TLA PCR amplification is performed using single primer sets.

1. Set up the TLA PCR reaction in separate wells for each primer set. To each well add 10 μ l Multiplex PCR Template from Multiplex PCR product, 12.5 μ l Q5 High-Fidelity 2 \times Master Mix, and 2.5 μ l Single Primer mix (*see* Table 1 for which primers to use on each multiplex mix; 10 μ M concentration of primers).
2. Mix by pipetting five times using a multichannel pipettor set at a sample volume of 15 μ l.
3. Seal and quick spin the PCR plate.
4. Run the following PCR program:
Step 1: 98 $^{\circ}$ C for 30 s.
Step 2: 98 $^{\circ}$ C for 10 s.
Step 3: 60 $^{\circ}$ C for 30 s.

Step 4: 65 °C for 3.5 min.

Repeat to Step 2 for 19 additional cycles, 20 cycles total.

Step 5: 65 °C for 5 min.

Step 6: 12 °C forever.

3.3.5 TLA PCR Cleanup

In this step, the PCR products are pooled and purified.

1. Measure the DNA concentration using 2–4 µl of each PCR product using Qubit® BR.
2. Pool equal amounts of DNA of each PCR product into a 1.5 ml Eppendorf tube.
3. To the pooled PCR sample add a 1× volume of isopropanol and 5 µl of NucleoMag 96 PCR beads, mix by inverting the sample five times, and incubate 15 min while agitating.
4. Place the tube in a magnetic rack for ≥ 3 min, until the beads are separated.
5. Invert the magnetic rack four times to include the beads from the lid and incubate for 1 min.
6. Remove the supernatant.
7. Remove the magnet, add 900 µl freshly made 80% EtOH, and resuspend the beads.
8. Place on the magnet for ≥ 1 min, until the beads are separated.
9. Invert the magnetic rack four times to include the beads from the lid, incubate for 1 min, and remove the supernatant.
10. Repeat **Steps 7–9** for two washes total.
11. Remove residual EtOH and air-dry the beads for 5 min.
12. Remove the magnet and resuspend the beads in 105 µl 10 mM TRIS pH 7.5.
13. Place on the magnet for ≥ 1 min, until the beads are separated.
14. Transfer 100 µl of the eluted PCR product to a clean 1.5 ml tube.
15. Use 2–4 µl eluted sample to measure the DNA concentration using the Qubit® BR.
16. Samples are now ready for NGS library preparation.

3.4 Sequencing

Library preparation is dependent on the sequencing platform. Here, libraries are prepared using the Illumina Tagmentation DNA prep kit with Nextera CD indexes and sequenced using the Illumina NextSeq500 system. The number of reads required for total coverage depends on the size of the genomic region of interest and the sequencing strategy. In this case, to cover the CFTR region of over 200 kb using Illumina PE150 sequencing, approximately 10 million reads per sample are required.

3.5 Bioinformatic Analysis and Haplotyping

TLA sequencing data derived from next-generation sequencing platforms such as Illumina MiniSeq, MiSeq, or HiSeq are aligned to the hg19 human reference genome (UCSC release GRCh37) using the BWA MEM algorithm [6]. Small-nucleotide variants (SNVs) are called using samtools mpileup [7].

For haplotyping, a similar workflow as described by Vermeulen et al. [8] is used. In short, heterozygous SNVs are selected as SNVs with a frequency of 15–85% and a minimal coverage of 25×. Only point mutations (but not small insertions or deletions) are selected for haplotyping. After SNV calling, all reads that have at least two small variants are selected, thereby creating links between SNVs. From these reads, the SNV with the highest number of links to other SNVs is selected. From this SNV, an initial seed haplotype is built by adding the SNVs directly linked to this SNV. Twenty-five loops of this are performed, each time extending the current haplotype by adding novel SNVs that are linked to SNVs present in the current seed haplotype. For each round, the linkage strength of each SNV within the current haplotypes is calculated. Also, during each round the ratio between how strongly an SNV can be linked to each of the opposing haplotypes is calculated for each SNV in the haplotypes. SNVs that do not pass the threshold criteria of the maximum allowed ratio are dropped from the current haplotype and re-evaluated in a later round. The stringency of these criteria slowly drops with each consecutive round of SNV linkage to the haplotype seed. In the first 20 iterations, only SNVs where both variants are linked to opposite haplotypes are accepted. In the last five iterations, SNVs where one variant is linked to one haplotype are accepted and phasing of the other variant to the opposing haplotype is assumed. Care should be taken that these last phased SNVs are less reliable and would require independent validation to confirm their phasing. All information and statistics of each phased allele is saved into an extensive table as an output file. Graphs showing all individual linkages used to construct the haplotypes are produced as well.

Lastly, the reads in the original alignment file are phased based on SNVs in the reads and to which allele they belong. SNVs of interest that could not be phased by the tool can still be visually inspected using these split alignment files in a genome browser.

3.6 Data Interpretation

Coverage of the total sequencing data can be visualized in a genome browser to confirm if TLA sample preparation was successful in obtaining enrichment of the region of interest. Here, Fig. 4a shows enrichment of TLA fragments along the position of the CFTR gene (117,086,331–117,336,482) on chromosome 7. Zooming in on this specific region shows the TLA coverage along the complete 200 kb CFTR gene, including the promoter region (Fig. 4b). Spider plots show the linkage between phased heterozygous SNPs on each allele. In our example, 112 and 105 SNPs are linked in

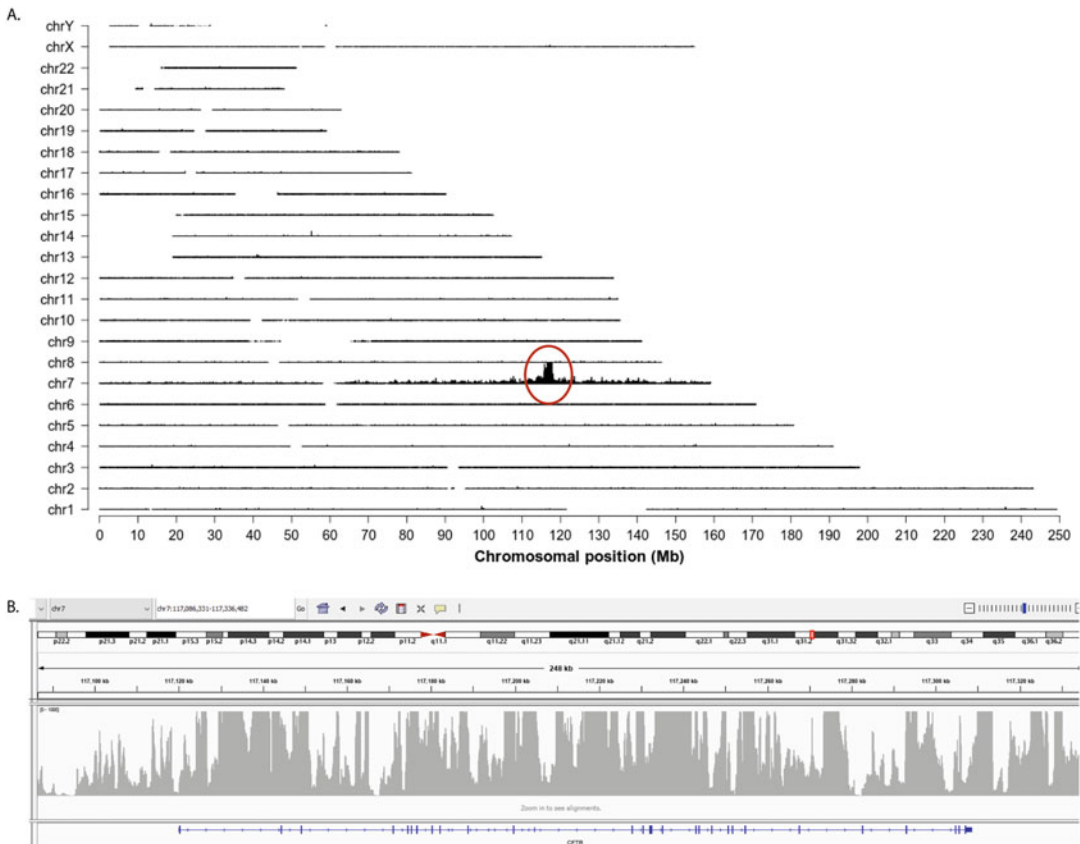
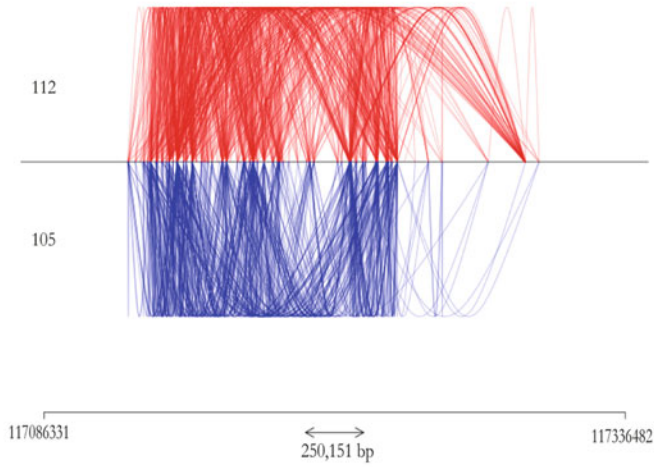


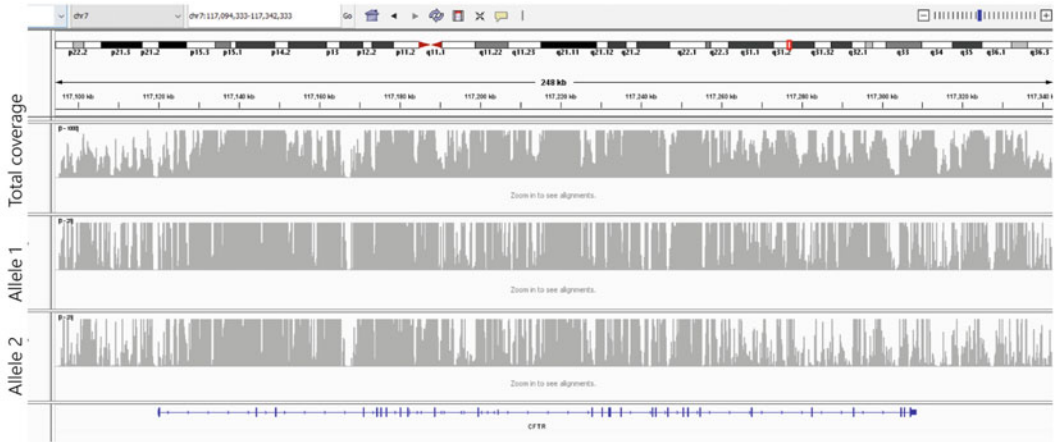
Fig. 4 Total sequencing coverage of TLA sequencing data, obtained from CF patient-derived intestinal organoids. **(a)** Overview of the total coverage along all the chromosomes; results clearly show increased signal along the position of the CFTR gene at chromosome 7. **(b)** Zooming in to the CFTR locus shows that TLA data covers the complete CFTR gene (human chr7:117,086,331-117,336,482). Y-axis is limited to 1000×

allele 1 and allele 2, respectively, as shown in Fig. 5a. Individual alleles and variation can also be inspected using this data. Investigation of the individual alleles shows coverage of individual haplotypes along the complete CFTR gene (Fig. 5b). In our example, the three-nucleotide deletion causing the F508del mutation and a C > G SNP causing the Y1092X mutation are found and assigned to allele 2 and allele 1, respectively, among other variations (Fig. 5c). From these data, we are able to determine that the two CF-causing mutations that are over 50 kb apart are present *in trans* in this sample thereby showing the ability of TLA to study allele-specific variation over a large genomic distance. This technique is not only interesting to study known and novel mutations both in coding- and non-coding regions, but is especially interesting to study complex alleles, where one allele harbors multiple mutations that in combination can alter gene expression or function.

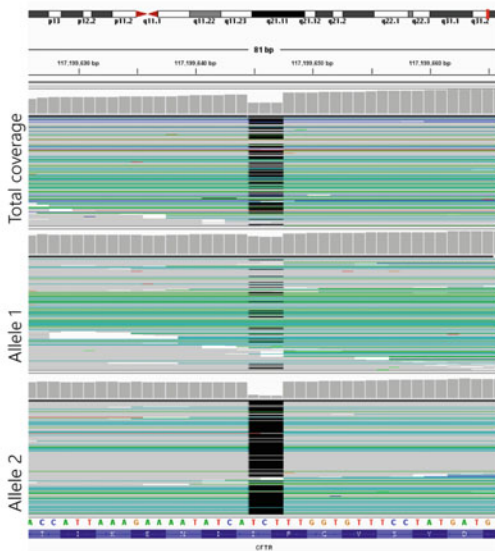
A.



B.



C.



D.

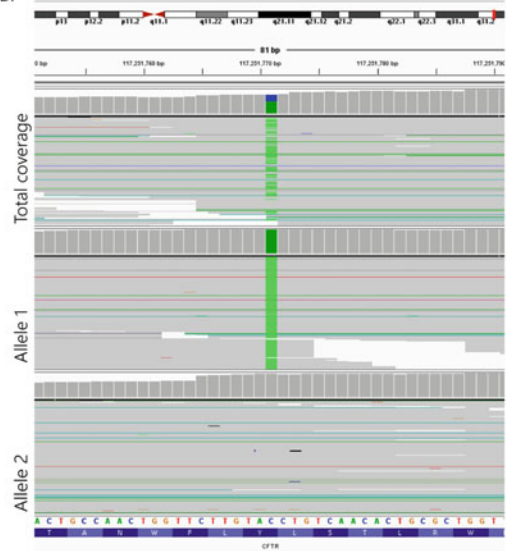


Fig. 5 Example of TLA data of CFTR F508del/Y1092X patient-derived intestinal organoids. **(a)** Spider plot of the split alleles showing the links between SNPs across the CFTR locus human chr7: 117,086,331-117,336,482. The number of SNPs used per allele for phasing are shown on the left. The red color represents allele 1 and the blue color represents allele 2. **(b)** From top to bottom the panels show: TLA sequence coverage of the original

4 Notes

The following notes refer to specific steps in the TLA protocol as indicated in the text.

1. Generation of single cells from organoid cultures reduces cell viability. Make sure to stop cell dissociation in TrypLE solution after 5 min of incubation to maximize sample quality.
2. All steps prior to PCR are performed with low-retention plastics, both pipette tips and Eppendorf tubes, to increase DNA yield.
3. The TLA protocol has been optimized for sample quantities between 0.5 and 1 million cells. Typically, after centrifugation this amount of cells yields a small but visible pellet. If not, this is indicative of too little starting material. *Were the cells viably frozen and thawed? Was the correct amount of starting material used?*

Cell viability is an important parameter impacting sample quality. Typically, (extreme) clumping of the cell pellet is observed when (part of) the sample has been lysed indicating poor cell viability. *Were the cells viably frozen and thawed?*

4. Here the use of Ampure XP beads deviates from its regular use for the purifying of DNA. However, we found that the magnetic beads bind crosslinked cells with high affinity. Using them enables the collection of cells based on magnetic separation which is more robust and user-friendly than centrifugation using samples with low cell numbers.
5. At this step, DNA of relatively high amounts and molecular weight is isolated. This could induce the aggregation of the purification beads but does not impact on sample quality.
6. For most (mammalian) sample types, around 2 μg of DNA is isolated when starting with 0.5–2 million cells. Dependent on the size of the genome (e.g., tumor samples, different organisms), these values can vary. The protocol can be continued with less than 2 μg DNA, although an impact on sample quality cannot be excluded. *Is a deviating genome size to be expected in the processed sample? Was the protocol started with the correct sample amount? Was the cell pellet observed throughout all relevant steps? Were the purification beads not over dried when isolating the DNA?*

Fig. 5 (continued) data (total coverage) and split alleles across the CFTR locus. Y-axes are limited to 25 \times . (**c–d**) Visual inspection of mutations of interest in IGV (Integrative Genomics Viewer) (**c**) along the CFTR F508del mutations and (**d**) spanning the Y1092X mutations. Black lines represent a deletion in the sequence and green bases indicate a misaligned base (SNP). This confirms the presence of the F508del and Y1092X mutation in trans in this organoid sample

7. At certain steps throughout the protocol, observations can be made that are indicative of the sample quality. Below, these steps have been listed including an explanation and suggestions for future experiments.

Degraded Undigested Control

TLA cannot be applied on samples containing degraded DNA. If the digestion and ligation controls were degraded: *Were viable cells used for the procedure? When dissociating organoids into single cells, make sure to stop dissociating in TrypLE after 5 min to prevent sample degradation.*

Digestion Control Sample Is Larger Than 2 kb

To obtain high quality TLA samples, it is important to not under-digest the sample using NlaIII. The presence of 10–20% of the sample > 2 kb in size typically does not affect the quality much; with higher percentages this cannot be guaranteed. The digestion efficiency can be modified by adjusting the incubation time at the permeabilization step. Most cell types require 30 min incubation at 65 °C, but it might be required to perform a titration series to obtain the most optimal results.

If incomplete digestion is observed, add NlaIII and repeat the digestion step. Spin down the sample and discard the supernatant and resuspend the sample pellet in Smart Cut Buffer and add 1 µL NlaIII. Incubate at 37 °C while shaking at 900 rpm for an additional 3–12 h.

The Ligation Control Is Below 5 kb in Size

Typically, after the first ligation step, an increase in molecular weight is observed yielding DNA molecules larger than 5 kb. When this increase is absent, it is highly likely that the TLA sample quality is not optimal. Multiple factors can be the cause of this observation:

- Sample amount. *Was the correct amount of starting material used?*
- Poor sample quality. *Were viable cells used at the start of the protocol? Does the digestion control show a size distribution between 0.3 and 2 kb?*
- Ineffective ligation. *Were the ligase (LIG) and ligation buffer (10× Smart Cut Buffer) subjected to temperatures above the indicated range and or subjected to more than three freeze and thaw cycles?*

In the case of incomplete re-ligation, increase the ligation time up to 3 h. Alternatively, spin the sample down, remove the supernatant, and repeat ligation by adding ligase buffer and ligase.

Acknowledgments

This work was funded by Health Holland grant no. 171114 and the Dutch CF foundation (NCFS) HIT-CF2 program.

References

1. De Vree PJP, De Wit E, Yilmaz M et al (2014) Targeted sequencing by proximity ligation for comprehensive variant detection and local haplotyping. *Nat Biotechnol* 32:1019–1025. <https://doi.org/10.1038/nbt.2959>
2. Snyder MW, Adey A, Kitzman JO, Shendure J (2015) Haplotype-resolved genome sequencing: experimental methods and applications. *Nat Rev Genet* 16:344–358. <https://doi.org/10.1038/nrg3903>
3. Hottentot QP, Van MM, Splinter E, White SJ (2017) Targeted locus Amplification and next-generation sequencing. *Methods Mol Biol* 1492:185–196. <https://doi.org/10.1007/978-1-4939-6442-0>
4. Dekkers JF, Berkers G, Kruisselbrink E et al (2016) Characterizing responses to CFTR-modulating drugs using rectal organoids derived from subjects with cystic fibrosis. *Sci Transl Med* 8:1–12. <https://doi.org/10.1126/scitranslmed.aad8278>
5. Vonk AM, van Mourik P, Ramalho AS et al (2020) Protocol for application, standardization and validation of the Forskolin-induced swelling assay in cystic fibrosis human colon organoids. *STAR Protoc* 1:100019. <https://doi.org/10.1016/j.xpro.2020.100019>
6. Li H, Durbin R (2009) Fast and accurate short read alignment with burrows-wheeler transform. *Bioinformatics* 25:1754–1760. <https://doi.org/10.1093/bioinformatics/btp324>
7. Li H, Handsaker B, Wysoker A et al (2009) The sequence alignment/map format and SAMtools. *Bioinformatics* 25:2078–2079. <https://doi.org/10.1093/bioinformatics/btp352>
8. Vermeulen C, Geeven G, de Wit E et al (2017) Sensitive monogenic noninvasive prenatal diagnosis by targeted Haplotyping. *Am J Hum Genet* 101:326–339. <https://doi.org/10.1016/j.ajhg.2017.07.012>

Open Access This chapter is licensed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

