Library preparation and MiSeq sequencing for the genotyping-by-sequencing of the Huntington disease HTT exon one trinucleotide repeat and the quantification of somatic mosaicism

Marc Ciosi (✉ marc.ciosi@glasgow.ac.uk)
  Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow

Sarah A. Cumming
  Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow

Asma M. Alshammari
  Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow

Efthymia Symeonidi
  Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow

Pawel Herzyk
  Glasgow Polyomics and Institute of Molecular, Cell and Systems Biology', 'College of Medical, Veterinary and Life Sciences, University of Glasgow

David McGuinness
  Glasgow Polyomics and Institute of Molecular, Cell and Systems Biology', 'College of Medical, Veterinary and Life Sciences, University of Glasgow

Julie Galbraith
  Glasgow Polyomics and Institute of Molecular, Cell and Systems Biology', 'College of Medical, Veterinary and Life Sciences, University of Glasgow

Graham Hamilton
  Glasgow Polyomics and Institute of Molecular, Cell and Systems Biology', 'College of Medical, Veterinary and Life Sciences, University of Glasgow

Darren G. Monckton
  Institute of Molecular, Cell and Systems Biology, College of Medical, Veterinary and Life Sciences, University of Glasgow

Method Article

Abstract

Huntington disease \(HD) is an autosomal dominant neurodegenerative disorder caused by the expansion of a CAG repeat in the first exon of the _HTT_ gene. Affected individuals inherit more than 40 repeats and the CAG repeat is genetically unstable in both the germline and soma. Molecular diagnosis and genotyping of the CAG repeat is traditionally performed by estimation of PCR fragment size. However, this approach is complicated by the presence of an adjacent polymorphic CCG repeat and provides no information on the presence of variant repeats, flanking sequence variants or on the degree of somatic mosaicism. To overcome these limitations, we have developed an amplicon-sequencing protocol that allows the sequencing of hundreds of samples in a single MiSeq run. The composition of the _HTT_ exon one trinucleotide repeat locus can be determined from the MiSeq sequencing reads generated. With sufficient sequencing depth, such MiSeq data can also be used to quantify the degree of somatic mosaicism of the _HTT_ CAG repeat in the tissue analysed.

Introduction

Huntington disease \(HD) is caused by the expansion of a CAG repeat in the _HTT_ gene. The number of CAG repeats is the primary determinant of age at onset of motor signs of HD \[1, 2]. The inherited number of CAG repeats is inversely correlated with age at onset with the addition of a single CAG repeat being associated with a reduction in the average age at onset of approximatively 2 years \[1, 3]. HD-causing alleles have > 35 CAG repeats and alleles with > 39 repeats are fully penetrant \[4]. The great majority of affected individuals inherit from 40 to 50 CAG repeats \[2]. The _HTT_ CAG repeat is genetically unstable in the germline and this causes the number of CAG repeats to frequently increase in successive generations \[2]. Germline instability of the _HTT_ CAG repeat thus causes anticipation in HD, _i.e._ earlier onset and increased disease severity in successive generations. The _HTT_ CAG repeat is also genetically unstable at the somatic level \[5]. Increases of tens, hundreds up to thousands of CAG repeats are observed, early in the disease progression, in striatal and cortical neurons \[6]. Genetic testing is used to help the diagnosis when HD is suspected based on typical symptoms or in a predictive manner before clinical onset in individuals at risk of inheriting an HD-causing allele. The molecular genotyping of the CAG repeat is traditionally performed by automated analyses of PCR fragment size \[7-9]. The number of CAG repeats is estimated based on the size of the PCR product measured by electrophoresis under the assumption that all the alleles are of the typical trinucleotide structure \(CAG)$_{Q1}$CAACAGCCGCCA\(CCG)$_{P2}$\(CCT)$_2$ \(where Q1 and P2 are the number of pure CAGs and CCGs respectively). However, the region within and around the _HTT_ CAG/CCG repeats is a hot-spot of single nucleotide variants \(SNV) and alleles with atypical trinucleotide structures have been reported \[10-15]. Five atypical trinucleotide structures have been previously described, all of them associated with SNV within the intervening sequence between the CAG and the CCG tracts \[11]. The most frequent of these variants is a G to A transition in the penultimate CAG of the pure CAG tract \[11, 16] which has been shown to be in complete linkage disequilibrium with a \(CCT)$_3$ variant immediately 3' of the CCG repeat \[16]. The _HTT_ repeat region is thus complex, comprising of a polymorphic mixture of CAG and CAA glutamine codons followed by a similarly complex polymorphic mixture of CCG, CCA and CCT proline codons \[11]. Atypical trinucleotide structures have mainly been identified when the atypical CAG allele could not be amplified using a traditional DNA testing protocol and are thus rarely assessed \[11]. Traditional DNA testing methods thus provide an estimation of the number of CAG repeats, but do not provide information about possible variant repeats and provide only a crude estimate of the degree of somatic mosaicism. To overcome these limitations, we have developed a protocol by which Illumina sequencing libraries of the _HTT_ exon one trinucleotide repeat locus are prepared by a single PCR using locus-specific primers incorporating sequencing adaptors and subsequently sequenced on a MiSeq platform. The number of CAG, CCG repeats, the presence of variant repeats and the genotype of the immediate flanking sequence can then be derived from the

MiSeq sequencing reads generated by this approach. With sufficient sequencing depth, such MiSeq data can also be used to quantify the somatic mosaicism of the _HTT_ CAG repeat in the tissue analysed. The present protocol has the potential to improve HD diagnostics because it can reveal clinically relevant _HTT_ variants that have been largely ignored by commonly used diagnostic protocols and because it does not require internal HD control samples for accurate determination of the number of _HTT_ CAG repeats. Such an approach can significantly improve the accuracy of diagnostics and predictive testing that are based on the CAG repeat genotyping. It also allows the high-throughput quantification of _HTT_ CAG repeat length somatic mosaicism. See figure in Figures section.

Reagents

PCR: - MiSeq-compatible PCR primers, _i.e._ already incorporating index and sequencing adaptors. We ordered primers from Sigma Aldrich and the most basic purification \(Desalt) was used. - Nuclease-free water \(Qiagen, 129115) - Taq DNA Polymerase from _Thermus aquaticus_ \(Sigma-Aldrich, D4545) - DMSO \(Sigma-Aldrich, D8418, for molecular biology) - 10X Custom Master Mix \(Thermo Scientific, ABgene UK, cat. no. SM-0005; 450 mM Tris-HCl \(pH 8.8), 110 mM \($(NH_4)_2SO_4$, 45 mM $MgCl_2$, 1.13 mg/ml BSA \(non-acetylated), 44 µM EDTA, 10 mM each of dATP, dCTP, dGTP and dTTP) − see "Supplementary document 1":https://www.nature.com/protocolexchange/system/uploads/6461/original/10xCustomMasterMiX-NoTaq-SM0005-ThermoScientific.pdf?1519668033 for product information from Thermo Scientific - 2-Mercaptoethanol \(βME) \(Sigma-Aldrich, M3148, BioReagent, for molecular biology, 99%) - 10X 'Custom PCR Master Mix + βME' made up by adding 4.8 µl of βME to 1 ml of 10X Custom Master Mix \(The concentration of βME in the 10X 'Custom PCR Master Mix + βME' is thus 0.48% \(v/v)) - 96-well plates AMPure XP size selection and PCR clean-up: - 1.5 ml DNA LoBind tubes \(Eppendorf) - Ethanol - Nuclease-free water \(Qiagen, Cat No./ID: 129115) - AMPure XP Beads \(Beckman Coulter) Quality control of the size selection and PCR clean-up: - Qubit dsDNA HS assay kit \(ThermoFisher, Q32851) - Agilent High Sensitivity DNA Kit \(Agilent, 5067-4626) Sequencing library quantification by qPCR: - KAPA Library Quantification Kit Illumina® Platforms, Complete kit \(Universal) \(KapaBiosystems, Kit Code: KK4824, Roche Cat. No: 07960140001) MiSeq sequencing - MiSeq Reagent Kit v3 \(600-cycle) \(Illumina, MS-102-3003) - PhiX Control v3 \(Illumina, FC-110-3001), _i.e._ a reliable, adapter-ligated library used as a control for Illumina sequencing runs

Equipment

- Standard molecular biology lab equipment - Pre-PCR 12-channel and 8-channel p10 pipettes - Post-PCR 8-channel p10 pipette - PCR machine with 96-well block - Dyna Mag-2 magnetic stand \(Life Technologies, Catalog number: 12321D) - Qubit 3.0 Fluorometer \(Life Technologies) - Bioanalyzer \(Agilent) - qPCR machine \(Step One Plus, Applied Biosystems) - Illumina MiSeq

Procedure

A. Overview: The sequence coding for the HTT polyglutamine and polyproline tracts is amplified from genomic DNA using MiSeq-compatible PCR primers \(_i.e._ PCR primers including _HTT_ locus-specific primers and the complete sequence of a barcoded Illumina adapter − "Figure 1":https://www.nature.com/protocolexchange/system/uploads/6459/original/LibraryPreparationByPCRforHTT.png?1530611667). This makes the PCR product obtained directly sequenceable. Post-PCR, a fraction of each PCR product are pooled together and cleaned-up using AMPure XP beads. This PCR clean-up step also allows removal of primer dimers. The sequencing library is then quality controlled using Qubit, Bioanalyzer and qPCR and sequenced

on the MiSeq platform. **B. MiSeq-compatible PCR primer design** The MiSeq-compatible PCR primers were designed based on the structure of the TruSeq combinatorial dual \(CD) index adapters \(p23 "Supplementary document 2":https://www.nature.com/protocolexchange/system/uploads/6661/original/SupplementaryDocument2-illumina-adapter-sequences-Feb2018.pdf?1530635414) with the addition of spacers between the sequencing primer binding site and the locus-specific primer \(Figure 1). These spacers are used to increase nucleotide diversity at the beginning of the sequencing reads. Spacers on the P5 primer enable cluster detection and validation on the MiSeq flow cell which is not possible without enough nucleotide diversity. This is required for the smooth sequencing of amplicon libraries prepared with the same locus-specific primer pair \[_e.g._ 17]. Spacers on the P7 primer should improve sequencing quality \[18]. The _HTT_ locus-specific primers used are HS319F \(5'-GCGACCCTGGAAAAGCTGATGA-3') and 33935.5 \(5'-AGCAGCGGCTGTGCCTGC-3') which respectively bind 26 bp 5' upstream of the CAG repeat and 26 bp 3' downstream of the CCG repeat. The protocol has been validated using the two following i5 and i7 index sets: - the TruSeq CD set of 20 indexes \(eight i5 and 12 i7 - p24 "Supplementary document 2":https://www.nature.com/protocolexchange/system/uploads/6661/original/SupplementaryDocument2-illumina-adapter-sequences-Feb2018.pdf?1530635414) to process up to 96 samples per MiSeq run \(see Table 1 for the full sequence of the MiSeq-compatible primers including TruSeq CD indexes); - the Nextera XT Kit v2 set of 40 indexes \(16 i5 and 24 i7 − p17 to 19 "Supplementary document 2":https://www.nature.com/protocolexchange/system/uploads/6661/original/SupplementaryDocument2-illumina-adapter-sequences-Feb2018.pdf?1530635414) to process up to 384 samples per MiSeq run \(see Table 2 for the full sequence of the MiSeq-compatible primers including Nextera XT Kit v2 indexes). The TruSeq CD set of primers \(Table 1) corresponds to an earlier version of the primer design that only allowed the sequencing of up to 96 samples per MiSeq run and did not include spacers on the P7 primers. The Nextera XT Kit v2 set of primers corresponds to the latest version of the primer design for which the following criteria were used in the spacer design: • Spacers on the P5 primer \(0 to 7 bases) were manually designed to maximise nucleotide diversity within the first 25 cycles of sequencing. We have limited ourselves to using seven base spacers on the i5 primer in order not to use too much read length. • Spacers on the P7 primer \(0 to 2 bases) were manually designed to stagger the sequencing within the trinucleotide part of the amplicons. • Spacers were associated with i5 and i7 indexes to minimise differences in melting temperature \(T_m) between MiSeq-compatible primer pairs \(spacers with higher T_m with indexes with low T_m, spacers with average T_m with indexes with average T_m, spacers with low T_m with indexes with higher T_m). This was done to minimise differences in PCR yield between primer pairs because they are all used at the same annealing temperature during PCR \(see PCR conditions below) although they have slightly different T_m. **C. Sequencing library preparation by a single PCR amplification** PCR amplifications are performed in 15 µl reactions containing 20 ng of genomic DNA \(diluted in nuclease-free water), 10% \(v/v) DMSO, 0.666 µM of each MiSeq-compatible PCR primer, 1X 'Custom PCR Master Mix + βME' and 1 U of Taq DNA polymerase. PCRs are manually prepared in 96-well plates using multichannel pipettes \(8 i5 indexes x 12 i7 index = 96 index pairs for each plate. 96 index pairs in total for the TruSeq CD index set and 384 index pairs in total for the Nextera XT Kit v2 index sets) as illustrated in "figure 2":https://www.nature.com/protocolexchange/system/uploads/6667/original/Figure2_indexesInPCRplates.png?1530612191. A PCR plate is prepared for 94 DNA samples of interest and two controls \(for each plate, these two controls are: one no-template controls \(NTC) and one HD positive control, _i.e._ a DNA sample from an individual carrying an HD-causing allele of known size) For each 96-well plate the PCRs are set up as follows: 1) Add 2 µl of a working solution \(5 µM) of each primer to each well of the PCR plate using an 8-channel p10 pipette for the P5 primers and a 12-channel p10 pipette for the P7 primers \(according to the plate maps shown in figure 2) 2) Add 5 µl

of template DNA (solution at 4 ng/µl) from each sample of interest and from the positive control to each well of the 96 well plate 3) Add 5 µl of nuclease-free water to the well of the NTC 4) Prepare the PCR master mix for the 96 reactions (volume of each reagent for one tube: 2.8 µl of nuclease-free water, 1.5 µl of 10X 'Custom PCR Master Mix + βME'*, 1.5 µl of DMSO, 0.2 µl of Taq polymerase at 5 units/µl) * _The concentration of βME in the 10X 'Custom PCR Master Mix + βME' is 0.48% (v/v). This leads to a final βME concentration of 0.048% (v/v) in the 15 µl PCR (as indicated in section Reagents/PCR)_ 5) Add 6 µl of this PCR master mix to each well of the plate 6) Seal the PCR plate 7) Place the PCR plate in a PCR machine and start the PCR. PCR cycling conditions are: (96°C, 5 min) 28 cycles of (96°C, 45 s), (58.5°C, 45 s) and (72°C, 3 min) (72°C, 10 min) **D. Control of the PCR amplification** To make sure that PCR amplification has occurred on each plate, 1 µl of each positive and NTC PCR products is collected after PCR. A 1:10 dilution of the positive control PCR products is prepared from the 1 µl collected and the DNA concentration of these dilutions is then measured using a Qubit fluorometer using the Qubit® dsDNA HS Assay Kit (Life Technologies) following the manufacturer's instructions. The dilutions of the positive control PCR products are then adjusted to 0.5 ng/µl. That dilution of the positive controls and the neat NTCs are then checked by capillary electrophoresis on a Bioanalyzer (Figure 3). Considering the full length of the MiSeq-compatible primers with Nextera XT Index Kit v2 indexes (Table 2), the expected size of the PCR products for alleles of structure $(CAG)_{Q1}CAACAGCCGCCA(CCG)_7(CCT)_2$ is 305 to 314 bp if Q1 = 15, 350 to 359 bp if Q1 = 30, 395 to 404 bp if Q1 = 45 and 440 to 449 bp if Q1 = 60. The Bioanalyzer concentration of the PCR products (300 to 450 bp) for the HD control should be between 0.2 and 1.5 ng/µl. **E. AMPure XP size selection and PCR clean-up** Note: volumes in the section below correspond to 384 PCR products. These volumes would have to be adjusted if processing fewer PCR products. After PCR amplification and the absence of contamination of the NTCs have been confirmed, 5 µl of each of the 384 products (including NTCs) are used for downstream pooling and purification, which leaves approximately 10 µl of the PCR product available in the PCR plates. These 10 µl, remaining in the PCR plates, are stored at 4°C as a back-up. Five µl of each of the PCR products are pooled and mixed together (total volume of the pool for 384 samples = about 1900µl). The pool of about 1900 µl pool is distributed in six DNA LoBind 1.5 ml tubes (300 µl per tube). From that step, DNA LoBind 1.5 ml Tubes are used. Each of the 300 µl aliquots is then purified with AMPure XP beads using a Dyna Mag-2 magnetic stand as follows: **General guidelines:** - The temperature of the beads and of the pool of _HTT_ PCR products are allowed to stabilise at room temperature - All the mixing is performed by gently flushing solutions up and down 10 times. No vortexing, no centrifugation **E.1. 0.6X AMPure XP clean-up to get rid of the primer dimers** This first AMPure XP purification has two aims: (i) to get rid of the primer dimers (size selection) and (ii) to concentrate the sequencing library. At this stage, there are six aliquots of PCR products in six 1.5 ml tubes. - E.1.1. Accurately measure the volume of each aliquot (this is better done directly by aliquoting a known volume of aliquot). The following protocol assumes each aliquot is 300 µl at this stage. Overestimation of the volume of the aliquot of PCR product can lead to a reduction of efficiency in the size selection against smaller fragments during clean-up which can, in turn, result in the carryover of primer dimers - E.1.2. Vortex AMPure XP beads to resuspend - E.1.3. In each of the 1.5 ml DNA LoBind tubes containing the aliquots of _HTT_ PCR products, add 0.6 µl of resuspended AMPure XP beads (180 µl) for each µl of _HTT_ PCR product - E.1.4. Mix well by gently pipetting up and down 10 times - E.1.5. Incubate 5 min at room temperature - E.1.6. Place the tubes on the magnetic stand. Loosen the tube cap (without opening it) to allow easy opening of the cap in subsequent steps (this will avoid disturbing the beads when opening the tube on the magnetic stand) - E.1.7. Leave the tubes on the stand for 8 minutes - E.1.8. Carefully remove the supernatant and transfer it to another tube (as a backup in case something goes wrong). Be careful not to disturb the beads that contain the DNA targets Leave the tubes on the magnetic stand while performing the following steps E.1.9. to E.1.13.: - E.1.9. Add 500 µl of freshly prepared 80% ethanol to each tube without disturbing the beads (the beads must be completely covered by the ethanol. Use more than 500 µl of 80% ethanol if this is not the case) - E.1.10. Incubate at room temperature for 30 seconds - E.1.11.

Carefully remove and discard the supernatant - E.1.12. Repeat steps E.1.9. to E.1.11. once - E.1.13. Leave the tubes on the magnetic stand with the lid open and air-dry the beads for 5 minutes. During drying, collect excess ethanol with a small tip and a 10 µl pipette **Caution: Do not over dry the beads. This may result in lower recovery of DNA target. The surface of overdried beads looks cracked.** - E.1.14. Remove the tubes from the magnetic stand. Elute DNA target by adding 45 µl of nuclease free water to the beads. First, use the water to detach the beads from the side of the tube by pipetting the water onto the bead pellet several times. When all the beads and water are at the bottom of the tube, mix well by gently pipetting up and down 10 times - E.1.15. Incubate the tubes at room temperature for 2 min **out** of the magnetic stand - E.1.16. Place the tubes in the magnetic stand. Loosen the tube cap \(without opening it) to allow easy opening of the cap in subsequent steps \(this will avoid disturbing the beads when opening the tube on the magnetic stand) - E.1.17. Incubate 5 min at room temperature - E.1.18. Without disturbing the bead pellet which is on the side of the tubes, carefully collect 40 to 43 µl of supernatant, _i.e._ eluted DNA, from each tube - E.1.19. Collect the supernatant from each of the six tubes and pool the six supernatants in a single new 1.5 ml DNA LoBind tube - E.1.20. Thoroughly mix the pool of these six supernatants by gently pipetting up and down **E.2. 1.4X AMPure XP clean-up to concentrate the sequencing library** At this stage, the sequencing library is contained in a single 1.5 ml tube and the total volume of the sequencing library is 240 to 260 µl. Take 50% of the eluted DNA \(about 130 µl) and put aside and store at 4°C as a backup. A second round of AMPure XP clean-up \(using 1.4 µl of the bead solution for each µl of sequencing library to concentrate) is performed on the other 50% \(about 130 µl) to increase the DNA concentration of the MiSeq library and further clean it up. - E.2.1. Vortex AMPure XP beads to resuspend - E.2.2. In the 1.5 ml DNA LoBind tube containing the _HTT_ sequencing library \(130 µl), add 182 µl \(1.4X) of resuspended AMPure XP beads - E.2.3. Mix well by gently pipetting up and down 10 times - E.2.4. Incubate 5 min at room temperature - E.2.5. Place the tube on the magnetic stand. Loosen the tube cap \(without opening it) to allow easy opening of the cap in subsequent steps \(this will avoid disturbing the beads when opening the tube on the magnetic stand) - E.2.6. Leave the tube on the stand for 8 minutes - E.2.7. Carefully remove the supernatant and transfer it to another tube \(as a backup in case something goes wrong). Be careful not to disturb the beads that contain the DNA targets Leave the tube on the magnetic stand while performing the following steps E.2.8 to E.2.12: - E.2.8. Add 200 µl of freshly prepared 80% ethanol to the tube without disturbing the beads \(the beads must be completely covered by the ethanol. Use more than 200 µl of 80% ethanol if this is not the case) - E.2.9. Incubate at room temperature for 30 seconds - E.2.10. Carefully remove and discard the supernatant - E.2.11. Repeat steps E.2.8. to E.2.10. once - E.2.12. Leave the tube on the magnetic stand with the lid open and air-dry beads for 5 minutes. During drying, collect excess ethanol with a small tip and a 10 µl pipette **Caution: Do not over dry the beads. This may result in lower recovery of DNA target. The surface of overdried beads looks cracked.** - E.2.13. Remove the tube from the magnetic stand. Elute the target DNA by adding 90 µl of nuclease free water to the beads. First, use the water to detach the beads from the side of the tube by pipetting the water and pouring it onto the bead pellet several times. When all the beads and water are at the bottom of the tube, mix well by gently pipetting up and down 10 times - E.2.14. Incubate the tube at room temperature for 2 min **out** of the magnetic stand - E.2.15. Place the tube in the magnetic stand. Loosen the tube cap \(without opening it) to allow easy opening of the cap in subsequent steps \(this will avoid disturbing the beads when opening the tube on the magnetic stand) - E.2.16. Incubate 5 min at room temperature - E.2.17. Without disturbing the bead pellet which is on the side of the tube, carefully collect 85 µl of the clear supernatant **Note: NO beads should be collected. It is safe to transfer a volume which is 5 µl smaller than the elution volume. If beads are collected when trying to collect 85 µl, replace the supernatant in the tube, put the tube back on the magnetic stand for 3 to 5 min and then collect less than 85 µl of clear supernatant in order not to collect any beads.** - E.2.18. Transfer the clear supernatant to a new 1.5 ml DNA LoBind tube **F. Quality control of the size selection and PCR clean-up** The DNA concentration of the purified library is then measured using a Qubit fluorometer using the Qubit® dsDNA HS Assay Kit following the manufacturer's instructions. This quantification is

used to prepare a dilution of the library that can be run on a Bioanalyzer. The sequencing library, diluted to 0.5 ng/μl, is then run on a Bioanalyzer to: \(i) check the library is of the expected size; \(ii) check the library is free of primer dimers; and \(iii) estimate the average size of the library \(Figure 4). Based on Bioanalyzer estimates, the aim is to have a library of at least 30 μl at 10nM. Note: the above QC **SHOULD NOT** be used as a basis for loading the library on the MiSeq. We have **ONLY** been able to properly load our _HTT_ libraries on MiSeq based on library quantification by qPCR. **G. Sequencing library quantification by qPCR** The quantity of sequenceable molecules in the sequencing library is then measured by qPCR using the KAPA Library Quantification Kits \(KAPABIOSYSTEMS). - G1. Using the concentration \(in ng/μl) estimated by Qubit and the average fragment size estimated by the Bioanalyzer, a rough molarity is calculated. This rough molarity estimate is used to dilute the sequencing library to approximately 8 pM within the range of the qPCR standards. - G2. The qPCR is performed as instructed in the Kapa Quantification user manual \(using high ROX that is the dye recommended for our qPCR machine). All samples and standards are run in triplicate. - G3. The concentration of the sequencing library is then calculated and corrected for size using the average fragment size estimated by the Bioanalyzer. **H. MiSeq Sequencing** The library is then sequenced following Illumina guidelines for an amplicon MiSeq run, using MiSeq Reagent Kit v3 with a cluster density of 1000k cluster / mm^2 but with 5% PhiX spike-in \(the PhiX library allows increasing nucleotide diversity during the run and also serves as a sequencing control). The sequencing library is then diluted to 4 nM \(based on the qPCR molarity estimate) and 5% PhiX library are added. The mix "sequencing library + PhiX" is then denatured according to the Illumina protocol \(MiSeq System, Denature and Dilute Libraries guide) and loaded at 12 pM onto the MiSeq. We used the MiSeq Control software version 2.5.05. If the samples correspond to individuals that are not affected by HD then the 600 cycles of sequencing can be used to produce two reads of 300 nt per cluster. However, if some samples correspond to individuals affected by HD or suspected to be affected, the 600 cycles of sequencing should be used to produce a forward read \(_i.e._ Read 1) of 400 nt and a reverse \(_i.e._ Read 2) of 200 nt. After the run \(65 h) the MiSeq Reporter software \(version 2.5.1) is used to demultiplex the reads corresponding to the index pair used \(96 TruSeq CD or 384 Nextera XT Index Kit v2 index pairs). This demultiplexing is done using the default demultiplexing parameters of the MiSeq Reporter software, _i.e._ "clusters are assigned to a sample when the index sequence matches exactly or there is up to a single mismatch per index" \(as indicated in the MiSeq Reporter Software Guide). The MiSeq Reporter software outputs the sequencing reads in .fastq files, two \(Read 1 and Read 2) for each of the 96 or 384 indexes \(5 to 120 MB per file, ~10 GB per run) as well as two \(Read 1 and Read 2) undetermined reads \(reads corresponding to the PhiX control library, which is not indexed, and reads for which the indexes could not be identified).

Timing

- PCR set-up for one 96-well plate: 1 h - PCR: ~2.5 h - Sample pooling, size selection and PCR clean-up: 1 h - Quality control of the size selection and PCR clean-up: 1.5 h - Sequencing library quantification and quality control by capillary electrophoresis and qPCR: 5h - MiSeq sequencing: 65 h

Troubleshooting

See "Figure5/Table 3":https://www.nature.com/protocolexchange/system/uploads/6897/original/Table3.png?1530812300.

Anticipated Results

Anticipated results - D. Control of the PCR amplification: see Figure 3 - E. AMPure XP size selection and PCR clean-up: see Figure 4 - F. Quality control of the size selection and PCR clean-up: see Figure 4 C and F \(NB: there should be no primer dimers and the peak ratio \(non-HD-causing alleles / HD-causing alleles) should be < 0.7) - Expected number of reads produced by a MiSeq run: ~16 million - Expected mean number of reads per sample \(when sequencing 384 samples per MiSeq run): $_{40,000}$ \(SD 45,000) - Failure rate \(when sequencing 384 blood DNA samples per MiSeq run): $_{1\%}$ \(_i.e._ not enough reads produced to confidently call a genotype for 1% of the samples) - 400 nt forward reads should allow the sequencing of HD-causing alleles containing up to ~100 CAG repeats - Expected non-Index Reads error rate \(_i.e._ error rate on PhiX reads): ~5% - Expected % of PhiX bases ≥ Q30 over the first 300 bp of the forward read \(_i.e._ Read 1): > 85% \(minimum warranted by Illumina when using the MiSeq Reagent Kit v3: > 70%) - Expected % of PhiX bases ≥ Q30 over the first 100 bp of the reverse read \(_i.e._ Read 2): > 80% \(minimum warranted by Illumina when using the MiSeq Reagent Kit v3: > 70%)

References

1. Andrew SE _et al_: The relationship between trinucleotide \(CAG) repeat length and clinical features of Huntington's disease. _Nature genetics_, 4\(4):398-403 \(1993). 2. Duyao M _et al_: Trinucleotide repeat length instability and age of onset in Huntington's disease. _Nature genetics_, 4\(4):387-392 \(1993). 3. Gusella JF _et al_: Huntington's disease: seeing the pathogenic process through a genetic lens. _Trends in biochemical sciences_, 31\(9):533-540 \(2006). 4. Warby SC _et al_: Huntington Disease, In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® \[Internet]. Seattle \(WA): University of Washington, Seattle; 1993-2018. Available from: https://www.ncbi.nlm.nih.gov/books/NBK1305/. \(1998 Oct 23 \[Updated 2014 Dec 11]). 5. Telenius H _et al_: Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. _Nature genetics_, 6\(4):409-414 \(1994). 6. Kennedy L _et al_: Dramatic tissue-specific mutation length increases are an early molecular event in Huntington disease pathogenesis. _Hum Mol Genet_, 12\(24):3359-3367 \(2003). 7. Jama M _et al_: Triplet repeat primed PCR simplifies testing for Huntington disease. _The Journal of molecular diagnostics_, 15\(2):255-262 \(2013). 8. Warner JP _et al_: A new polymerase chain reaction \(PCR) assay for the trinucleotide repeat that is unstable and expanded on Huntington's disease chromosomes. _Molecular and cellular probes_, 7\(3):235-239 \(1993). 9. Warner JP _et al_: A general method for the detection of large CAG repeat expansions by fluorescent PCR. _Journal of medical genetics_, 33\(12):1022-1026 \(1996). 10. Goldberg YP _et al_: Increased instability of intermediate alleles in families with sporadic Huntington disease compared to similar sized intermediate alleles in the general population. _Hum Mol Genet_, 4\(10):1911-1918 \(1995). 11. Yu S _et al_: Polymorphisms in the CAG repeat-a source of error in Huntington disease DNA testing. _Clinical genetics_, 58\(6):469-472 \(2000). 12. Gellera C _et al_: Errors in Huntington disease diagnostic test caused by trinucleotide deletion in the _IT15_ gene. _American journal of human genetics_, 59\(2):475-477 \(1996). 13. Margolis RL _et al_: Two novel single-base-pair substitutions adjacent to the CAG repeat in the huntington disease gene \(_IT15_): implications for diagnostic testing. _American journal of human genetics_, 64\(1):323-326 \(1999). 14. Williams LC _et al_: Null alleles at the Huntington disease locus: implications for diagnostics and CAG repeat instability. _Genetic testing_, 4\(1):55-60 \(2000). 15. Chong SS _et al_: Contribution of DNA sequence and CAG size to mutation frequencies of intermediate alleles for Huntington disease: evidence from single sperm analyses. _Hum Mol Genet_, 6\(2):301-309 \(1997). 16. Pêcheux C _et al_: Sequence analysis of the CCG polymorphic region adjacent to the CAG triplet repeat of the HD gene in normal and HD chromosomes. _Journal of medical genetics_, 32\(5):399-400 \(1995). 17. Fadrosh DW _et al_: An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. _Microbiome_, 2\(1):6 \(2014). 18. Bradley IM _et al_: Design and Evaluation of Illumina MiSeq-Compatible, 18S

rRNA Gene-Specific Primers for Improved Characterization of Mixed Phototrophic Communities. _Applied and environmental microbiology_, 82\(19):5878-5891 \(2016).

Acknowledgements
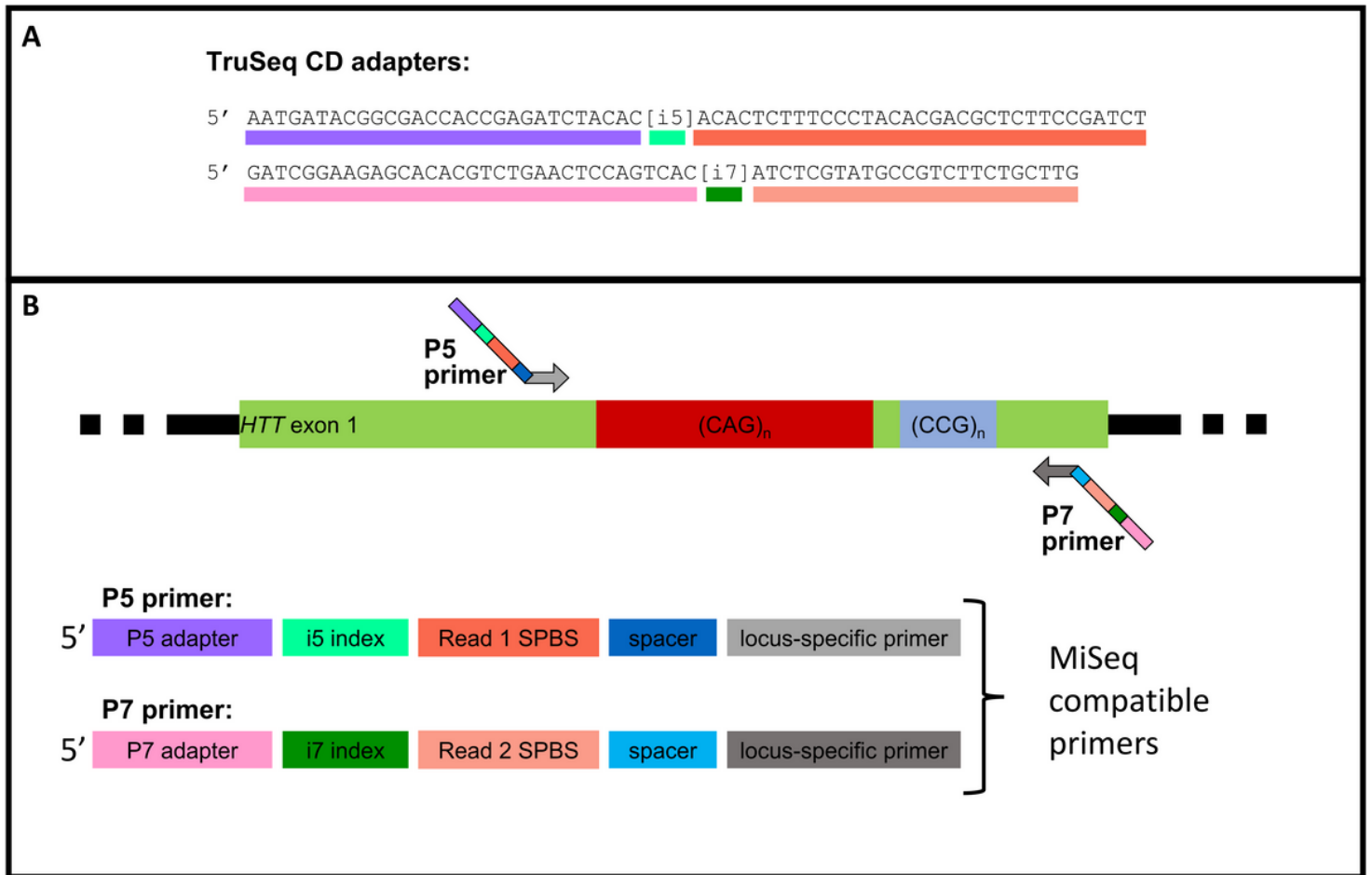
Figures

Figure 1

Library preparation by PCR using locus-specific primers incorporating MiSeq adapters for the sequencing of HTT exon one trinucleotide repeat locus. A: TruSeq CD adapters. i5 and i7: indexes B: Design of MiSeq-compatible primers for the MiSeq sequencing of the HTT CAG/CCG repeat region. (CAG)n: CAG repeat. (CCG)n: CCG repeat. SPBS: Sequencing primer binding site.

A   TruSeq CD index set

i7→	1	2	3	4	5	6	7	8	9	10	11	12
i5↓	D701	D702	D703	D704	D705	D706	D707	D708	D709	D710	D711	D712
1 D501												
2 D502												
3 D503												
4 D504												
5 D505												
6 D506												
7 D507												
8 D508												

B   Nextera XT Index Kit v2

plate A

i7→	1	2	3	4	5	6	7	8	9	10	11	12
i5↓	N701	N702	N703	N704	N705	N706	N707	N710	N711	N712	N714	N715
1 S502												
2 S503												
3 S505												
4 S506												
5 S507												
6 S508												
7 S510												
8 S511												

plate B

i7→	13	14	15	16	17	18	19	20	21	22	23	24
i5↓	N716	N718	N719	N720	N721	N722	N723	N724	N726	N727	N728	N729
1 S502												
2 S503												
3 S505												
4 S506												
5 S507												
6 S508												
7 S510												
8 S511												

plate C

i7→	1	2	3	4	5	6	7	8	9	10	11	12
i5↓	N701	N702	N703	N704	N705	N706	N707	N710	N711	N712	N714	N715
9 S513												
10 S515												
11 S516												
12 S517												
13 S518												
14 S520												
15 S521												
16 S522												

plate D

i7→	13	14	15	16	17	18	19	20	21	22	23	24
i5↓	N716	N718	N719	N720	N721	N722	N723	N724	N726	N727	N728	N729
9 S513												
10 S515												
11 S516												
12 S517												
13 S518												
14 S520												
15 S521												
16 S522												

Figure 2

Arrangement of i5 and i7 indexes at PCR set-up in 96-well plates. A: using the 96 index combinations of the TruSeq CD index set. B: Using the 384 index combinations of the Nextera XT Index Kit v2.
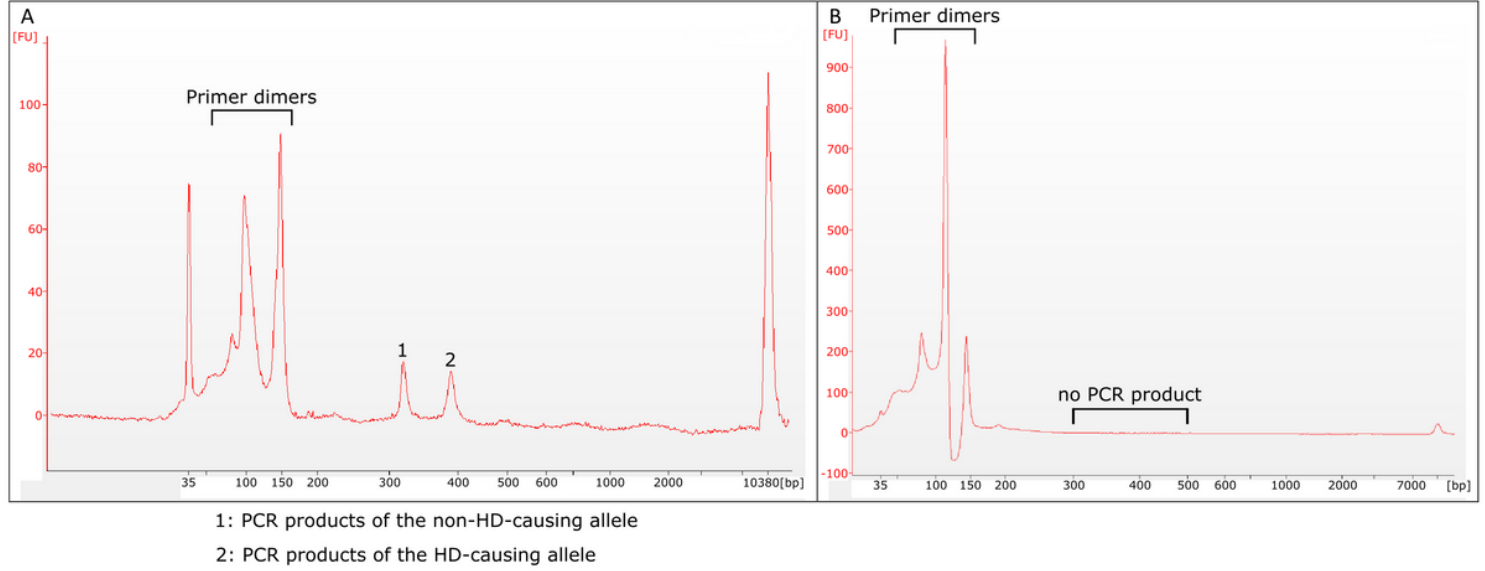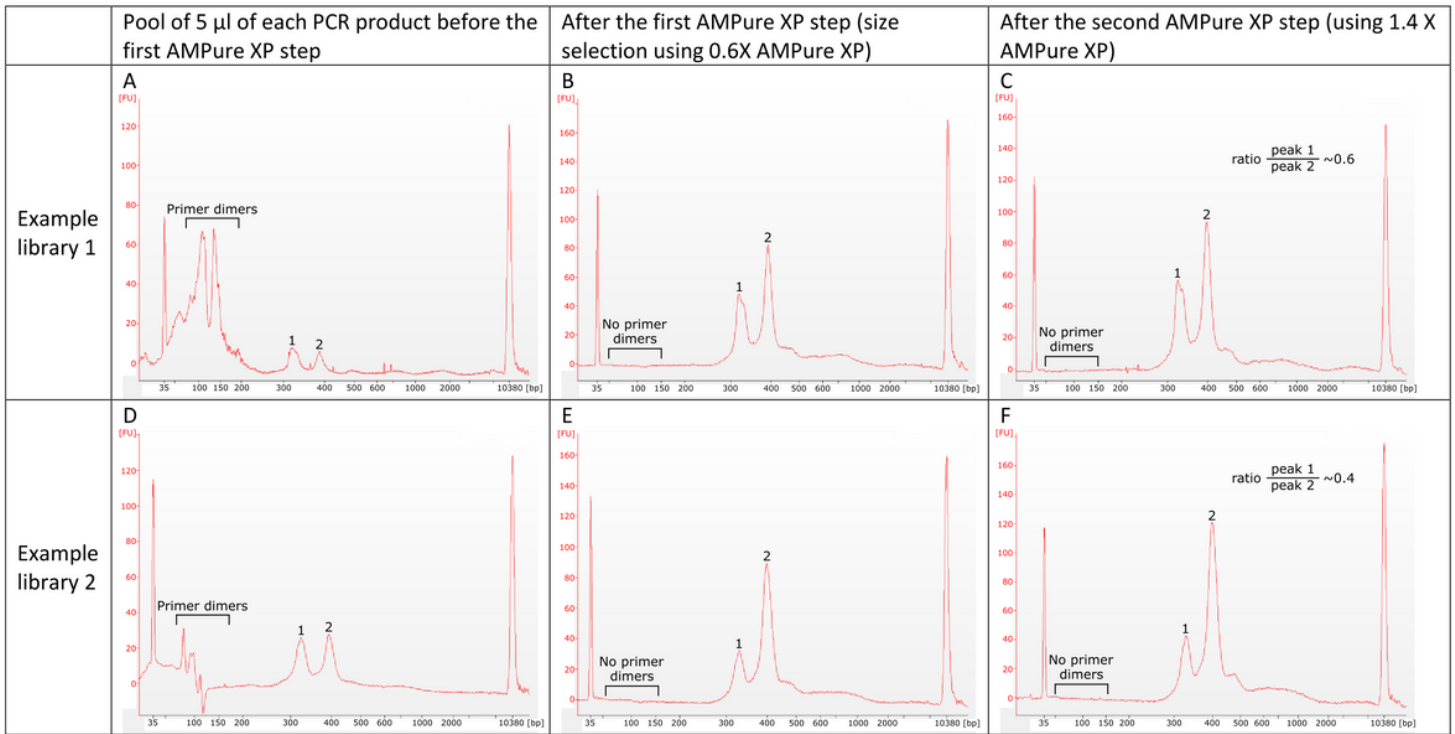
A
[FU]

Primer dimers

1
2

35   100 150 200   300   400   500 600   1000   2000   10380[bp]

1: PCR products of the non-HD-causing allele
2: PCR products of the HD-causing allele

B   Primer dimers
[FU]

no PCR product

35   100 150 200   300   400   500 600   1000   2000   7000   [bp]

Figure 3

Example of Bioanalyzer traces for the PCR controls. A: Positive HD control: primer dimers between 50 and ~160 bp; 1: PCR products of the non-HD-causing allele around 320 bp; 2: PCR products of the HD-causing allele around 390 bp. B: No template control: primer dimers between ~50 and ~160 bp, no PCR product between 300 and 500 bp.

	Pool of 5 µl of each PCR product before the first AMPure XP step	After the first AMPure XP step (size selection using 0.6X AMPure XP)	After the second AMPure XP step (using 1.4 X AMPure XP)
Example library 1	A	B	C
Example library 2	D	E	F

1: PCR products of the non-HD-causing allele   2: PCR products of the HD-causing allele

Figure 4

Example of bioanalyzer traces for the pool of PCR products at the different steps of the AMPure XP size selection and PCR clean-up. A, B and C: example library 1. D, E and F: example library 2. A and D: before the first AMPure XP step. B and E: After the first AMPure XP step (size selection using 0.6X AMPure XP). C and F: After the second AMPure XP step (using 1.4 X AMPure XP). Peaks corresponding to the primer dimers: ~50 to ~160 bp. 1: peak corresponding to the PCR products of the non-HD-causing allele: ~320 to ~340 bp. 2: peak corresponding to the PCR products of the HD-causing allele: ~380 to ~400 bp. Expected ratio peak1/peak2 < 0.7.

Step	Problem	Possible reason	Solution
E. and F.	Primer dimers are still present after size selection and PCR clean-up*	Something went wrong with the size-selection	make sure to use 0.6X AMPure XP beads in step E.1 (beads should be very well homogenised before pipetting, the volume of solution to which to apply size selection should be accurately known) and repeat step E and F
E. and F.	Primer dimers are still present after size selection and PCR clean-up*	Higher salt concentration in the sample on which size-selection is applied (can, for example, happen if a different PCR master mix is used)	1) Use the specified 'Custom PCR Master Mix + βME' 2) Run a gradient of AMPure XP beads concentration around 0.6X to find the concentration that facilitates removal rid of the primer dimers** 3) Purify the library by gel extraction**
E. and F.	Primer dimers are still present after size selection and PCR clean-up	Lower quality or quantity of the DNA template leading to a higher relative quantity of primer dimers in the PCR product	1) Do an additional size selection step with 0.6X AMPure XP beads (similar to E.1) on the library to get rid of the primer dimers** 2) Purify the library by gel extraction** 3) Use more or better-quality template DNA and start over
E. and F.	The library has been lost somehow during step E.	Used less than 0.6X AMPure XP beads in step E.1.3	Do a 2X AMPure XP clean-up on the supernatant that was kept as a backup at E.1.8. and start over with E.1. on the eluted DNA
G	The standard deviation of a library concentration estimated by qPCR is too high	Poor pipetting	Repeat qPCR
G	Library molarity is out of the standards' range	Estimated molarity (based on Qubit concentration and Bioanalyzer sizing) and/or dilution is incorrect	Repeat qPCR using a different dilution
H	Number of clusters is too high or too low	Concentration estimated by qPCR is wrong or error in diluting and loading	Sequencing reads may still be usable and the MiSeq run would just have produced less reads than expected. If no data or not enough data, the sequencing library can be rerun on a new MiSeq run correcting for the wrong loading
Data analysis post-H	Read length distributions with multiple peaks or very broad peaks	The DNA template was prepared from a cell line	Probably OK for genotyping but difficult to use to quantify somatic mosaicism. Best to use DNA from primary tissues (e.g. whole blood DNA), see figure S2 from associated paper (Ciosi et al.) for typical read length distribution obtained from blood DNA

Figure 5

Table 3: Troubleshooting. *: should also be associated with a ratio peak1/peak2 higher than expected (Figure 4C & F). **: Different size-selection and PCR clean-up procedures may not affect the genotyping of the HTT exon one repeat but are likely to have an impact on the degree of somatic mosaicism quantified from the read length distribution obtained.
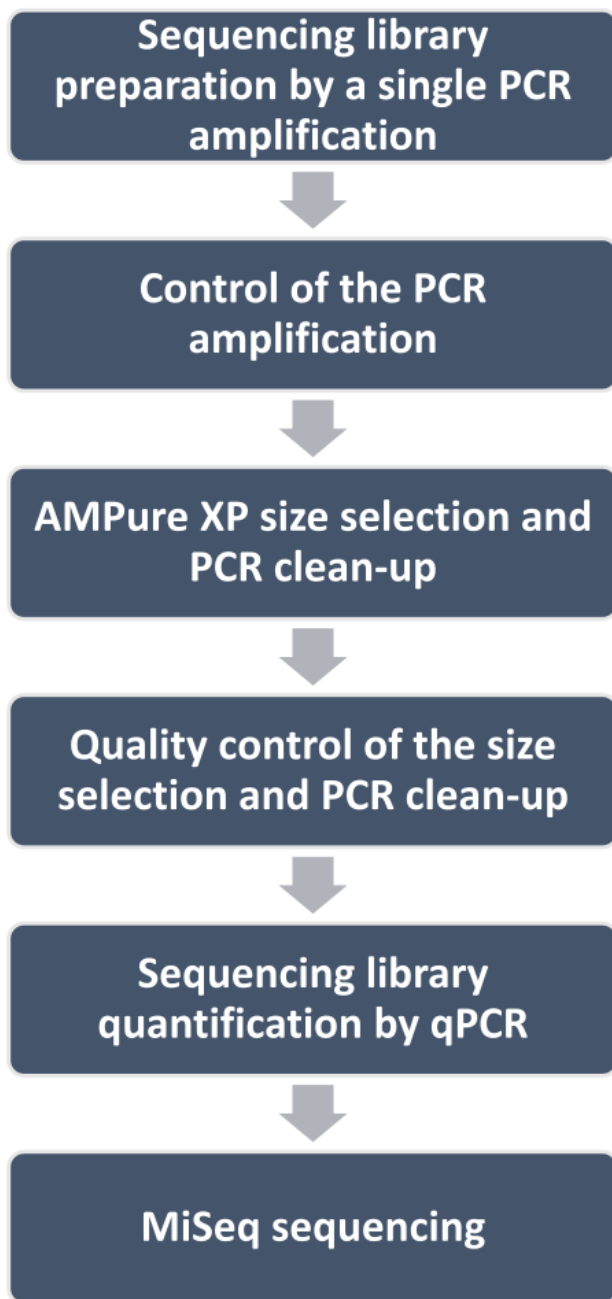
Figure 6

FlowDiagramFlow diagram of the protocol

Supplementary Files

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

- Table1TruSeqCDprimers.xlsx
- 10xCustomMasterMiXNoTaqSM0005ThermoScientific.pdf
- SupplementaryDocument2illuminaadaptersequencesFeb2018.pdf

- Table2NexteraV2primers.xlsx