Sequencing the Previously Unsequenceable Using Amplification-free Targeted Enrichment Powered by CRISPR/Cas9

Jenny Ekholm, Yu-Chih Tsai, Ian McLaughlin, Brendan Galvin and Janet Ziegle
PacBio, 1305 O'Brien Drive, Menlo Park, CA 94025

Introduction

Genomic regions with extreme base composition bias and repetitive sequences have long proven challenging for targeted enrichment methods, as they rely upon some form of amplification. Similarly, most DNA sequencing technologies struggle to faithfully sequence regions of low complexity. This has especially been true for repeat expansion disorders such as Fragile X syndrome, Huntington's disease and various Ataxias, where the repetitive elements range from several hundreds of bases to tens of kilobases.

We have developed a robust, amplification-free targeted enrichment technique, called No-Amp Targeted Sequencing, that employs the CRISPR/Cas9 system. In conjunction with Single Molecule, Real-Time (SMRT) Sequencing, which delivers long reads spanning the entire repeat expansion, high consensus accuracy, and uniform coverage, these previously inaccessible regions are now accessible. This method is completely amplification-free, therefore removing any PCR errors and biases from the experiment. Furthermore, this technique also preserves native DNA molecules, allowing for direct detection and characterization of epigenetic signatures.

The No-Amp method is a two-day protocol, compatible with multiplexing of multiple targets and samples in a single reaction, using as little as 1 µg of genomic DNA input per sample. We have successfully targeted a number of repeat expansion disorder loci (*HTT*, *FMR1*, *ATXN10*, *C9orf72*) with alleles as long as >2700 repeat unites (>13 kb). Using the No-Amp method we have isolated hundreds of individual on-target molecules, allowing for reliable repeat size estimation, mosaicism detection and identification of interruption sequences – all aspects of repeat expansion disorders which are important for better understanding the underlying disease mechanisms.

No-Amp Targeted Sequencing Method

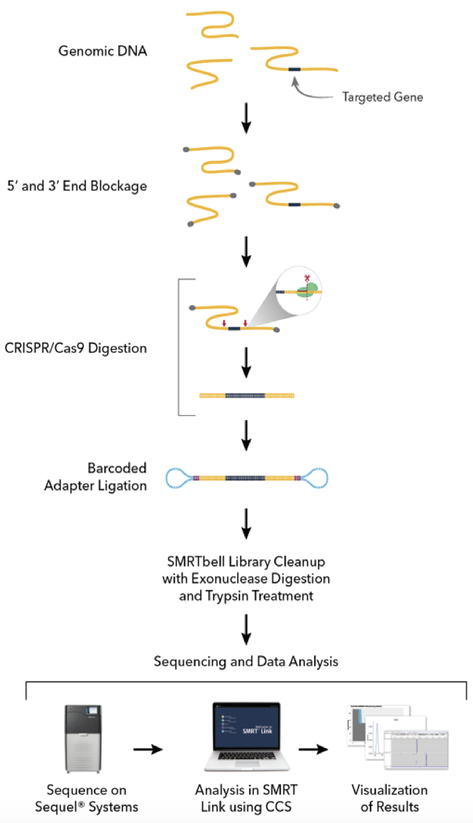


Figure 1. Overview of the No-Amp workflow. The first step of the No-Amp workflow is to block the 5' and 3' ends of your gDNA. By designing guide RNAs flanking each end of your region/s of interest the CRISPR/Cas9 system will cleave the double stranded gDNA and leave the ends available for sequencing adapter ligation. Before sequencing the library is cleaned up.

Both Sequel Systems are compatible with the No-Amp method. You can multiplex both target regions as well as samples. The required input gDNA is 5-10 µg/SMRT Cell, which can be subdivided across multiplexed samples. The insert sizes can range between 4-6+kb. The expected on-target yield is 100s CCS reads per 1-2 µg sample and the on-target rate is 40-60% per SMRT Cell after demultiplexing which translates to enrichment factors of 10,000s-100,000s fold.

The analysis pipeline consists of CCS analysis in SMRT Link and command-line based visual reporting tools for repeat count, mosaicism characterization and identification of sequence interruptions.

No-Amp Protocol Overview

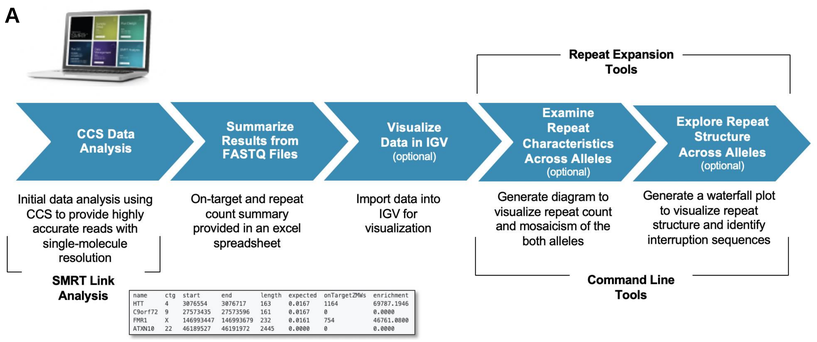
Step	Incubation Time	Hands-On Time	Day
gDNA Treatment: Dephosphorylation + Klenow Fill-In	85 min	20 min	Day 1
CRISPR-Cas9 Digestion	80 min	20 min	
1 AMPure PB Purification	–	30 min	
Adapter Ligation	75 min	10 min	
1 AMPure PB Purification	–	30 min	
Optional Stop Point			
Exonuclease Digestion + Trypsin Treatment	80 min	10 min	
2 AMPure PB Purifications	–	60 min	
Optional Stop Point			
Sequencing Complex Preparation	302 min	50 min	Day 2
Total	*10 hr 22 min*	*3 hr 50 min*	

6 hr 22 min incubation
3 hr 20 min hands-on

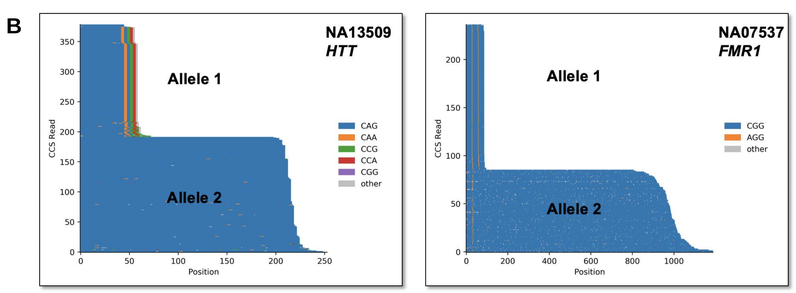
Figure 2. **No-Amp protocol overview.** The subsequent steps for the No-Amp workflow is outlined on the left, along with two optional stopping points. The incubation and hands-on time needed for each step in shown in the middle. By using either one of the stopping points the workflow can be split up in 2 days.

No-Amp Analysis Overview

No-Amp analysis pipeline



Visualize Repeat Structure and Identify Interruption Sequences



Repeat count and characterization of mosaicism

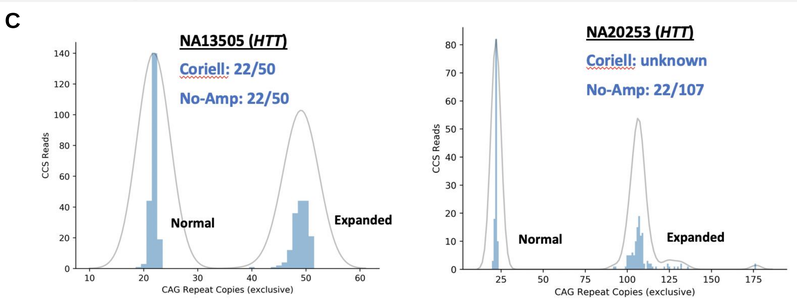
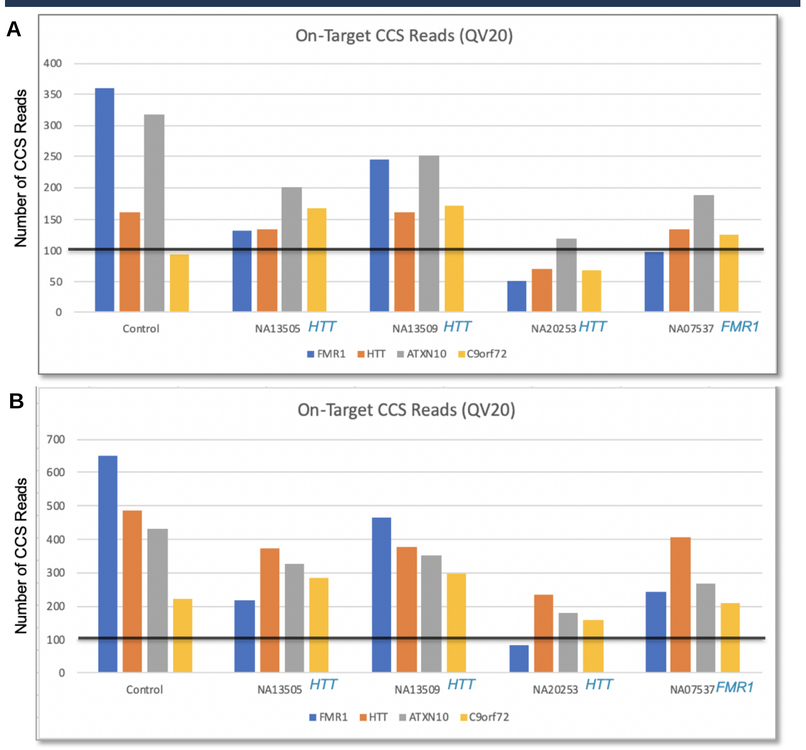


Figure 3. **Overview of the No-Amp analysis pipeline.** (**A**) After sequencing, CCS analysis is performed on SMRT Link analysis software. The output files can be used to summarize the data and view the data in the Integrated Genome Viewer (IGV). Command-line analysis can be used to look at (**B**) repeat structures to identify interruption sequences and (**C**) repeat count and mosaicism characterization.

No-Amp Sequencing Results



C

Target	Run	Control	NA13505 HTT (Male)	NA13509 HTT (Female)	NA20253 HTT (Male)	NA07537 FMR1 (Female)
FMR1	A/B	360 / 648	132 / 219	245 / 466	50 / 82	97 / 242
HTT	A/B	160 / 487	134 / 372	161 / 379	70 / 233	133/ 408
ATXN10	A/B	318 / 433	201 / 328	252 / 352	118 / 182	189 / 268
C9orf72	A/B	92 / 220	167 / 285	171 / 297	68 / 160	135 / 210

Figure 4. **No-Amp Sequencing Results.** 4 targets (*HTT, FMR1, ATXN10, C9orf72*) and 5 samples (1 control, 3 Coriell samples with *HTT* expansion, 1 Coriell sample with *FMR1*) were multiplexed in the sequencing runs. (**A**) 1 µg/sample (5 µg total) was used as input gDNA. On-target rate was 30-65% translating to enrichment factors of 23,717-237,976. (**B**) 2 µg/sample (10 µg total) was used as input gDNA. On-target rate was 30-62% translating to enrichment factors of 35,676-234,585. (**C**) The repeat counts are shown for each target and sample for both runs A and B.

Conclusion

No-Amp Targeted Sequencing - amplification-free targeted enrichment of hard-to-amplify regions using CRISPR/Cas9

- No PCR-bias or errors

SMRT Sequencing provides

- Long reads to span the region of interest
- High accuracy to accurately call every base
- Uniform coverage to access previously unseqenceable regions
- Single molecule resolution that enables characterization of mosaicism

Repeat Expansion Disorders - holistically capture all needed information in one experiment

- Accurately repeat count for both alleles
- Identification of interruption sequences
- Characterization of mosaicism

References

- Learn more and see a complete publication list: pacb.com/NoAmp
- Visit pacb.com/documentation to download:
  - No-Amp sequencing library preparation protocol
  - No-Amp analysis protocol
  - No-Amp reference guide on how to design new targets
- Connect with a PacBio Scientist: pacb.com/contact-us