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

Jenny Ekhholm, Yu-Chih Tsai, Ian McLaughlin, Brendan Galvin and Janet Ziegel
PacBio, 1305 O’Brien Drive, Menlo Park, CA 94025

Introduction

Genomic regions with extreme base composition bias and repetitive sequences have long been 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 approachable. 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, TCFC4, C9orf72) with alleles as long as >2700 repeat units (>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 Overview

A. Drawing an Enrichment Region

B. Designing NanoGuide Ligation Adapters

C. Constructing the NanoGuide

D. Target Capture

Figure 1. Overview of the No-Amp workflow. (A) The first step of the No-Amp sequencing is to construct the SMRTbell library by shearing the DNA using a restriction enzyme that does not cut within the target region. This is followed by No-Amp adapter ligation. These adapters lack the sequencing polymerase binding sites and can therefore not be used for sequencing. (B) The NanoGuide is designed to target the 5’ end of the region of interest where the Cas9 will do a double-stranded digest. The fragments containing the target will now have a free 3’ end where the sequencing adapter can be added. (C) In order to remove background, two to four restriction enzymes and exonucleases that do not cut out the region of interest will be used. (D) The SMRTbell libraries will then be sequenced on a Sequel System and analyzed in SMRT Link using consensus sequencing (CCS) analysis. The results can then be viewed in IGV, or using command-line tools that allow repeat length for both alleles, characterization of mosaicism and interruption sequences.

Sequencing of Repeat Expansion Loci

Table 1. Targeted disease loci. Three repeat expansion disease loci were targeted using the No-Amp method.

<table>
<thead>
<tr>
<th>Target Gene</th>
<th>Associated Disease</th>
<th>Chr Repeat</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTT</td>
<td>Huntington's Disease</td>
<td>4</td>
</tr>
<tr>
<td>FMR1</td>
<td>Fragile X syndrome</td>
<td>15</td>
</tr>
<tr>
<td>ATXN10</td>
<td>Spinocerebellar Ataxia Type 10</td>
<td>21</td>
</tr>
</tbody>
</table>

Targeted sequencing using the No-Amp method was carried out on Coriell cell lines from patients with known repeat expansion disease loci for HTT and FMR1. We used 2 µg of input DNA per sample and multiplexed 6 samples using barcodes as well as the 2 disease loci on one SMRT Cell.

Figure 2. Effect of DNA quality on yield. PacBio QC traces for DNA samples NA13509 and NA14044. The lower yield of NA13509 is accompanied by reduced on-target reads and an imbalance in the ratio of target reads to background. (A) shows the target reads per sample and (B) shows the total number of target reads per sample.

Impact of gDNA Quality on Results

Sample quality has the biggest effect on target recovery. The figures below show the Fmto Pulse traces for samples NA13509 and NA14044. The lower quality (degraded) DNA sample NA14044 results in fewer on-target reads and an imbalance in the ratio of normal to expanded allele counts. While the higher quality DNA for sample NA13509 yielded twice as many on-target reads and an even balance between the normal and expanded alleles.

Visualization of Results

Table 2. Comparative results for the Coriell samples. The tables above show the number of on-target reads for the Core samples and Coriell samples for targets (A) HTT and (B) FMR1. Results are normalized to 20 g SMRT cell.

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>NA13509</th>
<th>NA14044</th>
<th>NA20241</th>
<th>NA20236</th>
</tr>
</thead>
<tbody>
<tr>
<td>HTT</td>
<td>(B)</td>
<td>(A)</td>
<td>(A)</td>
<td>(B)</td>
</tr>
<tr>
<td>Total</td>
<td>3895</td>
<td>3154</td>
<td>3515</td>
<td>3155</td>
</tr>
<tr>
<td>Normal</td>
<td>265</td>
<td>272</td>
<td>278</td>
<td>274</td>
</tr>
<tr>
<td>Expanded</td>
<td>3632</td>
<td>3058</td>
<td>3237</td>
<td>3065</td>
</tr>
</tbody>
</table>

Figure 3. Repeat size and characterization mosaicism. Using command-line tools the repeat size distribution for both alleles can be seen for sample NA12226 for FMR1. (A) Shows the target repeat length of both allele bases. (B) Shows the target repeat length in terms of repeat copies.

Figure 4. IGV visualization of reads for NA70537 for FMR1. The data can be imported into the IGV viewer to visualize all reads for both alleles (20:306-395) of the target region.

References


For Reprint Use Only. Not for use in diagnostic procedures. © Copyright 2019 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, the PacBio logo, and Sequel are trademarks of Pacific Biosciences. SMRTLink and SMRTCell are trademarks of Sequel Systems. NGS go and RepliGator are trademarks of Sequel. FMR1 Poly and Fragile X Analysis are trademarks of Agilent Technologies, Inc. All other trademarks are the sole property of their respective owners.