Target Enrichment Using a Neurology Panel for 12 Barcoded Genomic DNA Samples on the PacBio SMRT Sequencing Platform

Kevin Eng, Walter Lee, Ting Hon, Anand Sethuraman, Jenny Gu, Primo Baybayan, Steve Kujawa

PacBio, Menlo Park, California, USA

Introduction

Target enrichment paired with long read sequencing is a powerful tool for understanding polymorphic SNPs, enabling phasing with tandem repeats and structural variants. With the increasing availability of reference genomes, scientists can easily design a cost-effective, targeted investigation using custom probes specific to regions of interest.

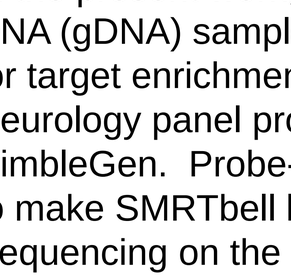
Using PacBio long reads in conjunction with probe capture, we sequenced multi-kilobase enriched regions to fully characterize intronic and exonic regions, distinguish haplotypes, and characterize structural variants. Furthermore, we demonstrated this approach gives access to complex genomic regions that were previously inaccessible with other sequencing platforms.

In the present work, 12 barcoded genomic DNA (gDNA) samples were sheared to 7 kb for target enrichment analysis using the Neurology panel provided by Roche NimbleGen. Probe-captured DNA was used to make SMRTbell libraries for SMRT Sequencing on the PacBio RS II.

Workflow

Multiplex & Capture

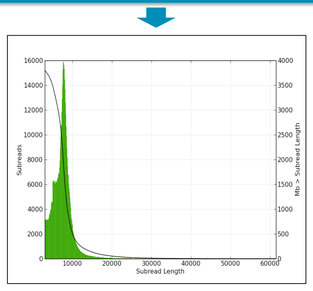
- Shear gDNA to 7 kb
- Linear Barcoded Adapter Ligation
- Pre-Capture Amplification
- Size-Selection
- Hybridization w/ Probes
- Bead Capture & Washing
- Post-Capture Amplification



Sample QC

SMRT Sequencing

- SMRTbell Library Prep
- Size-Selection



PacBio RS II Sequencing
Subread Length Distribution

Data Analysis

- Resequencing
- Demultiplex Samples
- Align to Genome Reference
- Tertiary Analysis Tools

Figure 1. Workflow for multiplex target capture, SMRT Sequencing and data analysis.

Methods

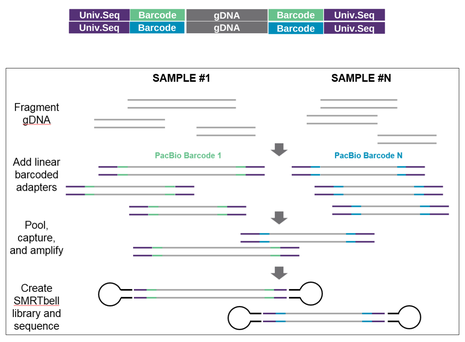


Figure 2. Barcoded linear adapter protocol used for pooling.

Barcode	Cell Line
1	A431
9	HeLa
17	Jurkat
26	K562
38	MCF7
40	Raji
48	NA11922
52	NA18527
58	NA18942
59	NA18484
62	NA18526
70	NA12878

Table 1. List of 12 barcoded cell lines.

Results

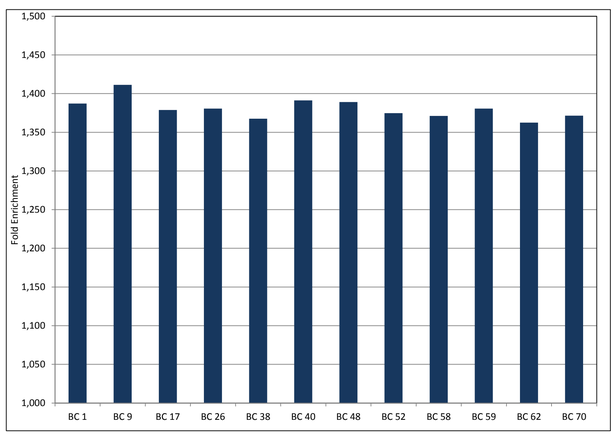


Figure 3. Barcode distribution and fold enrichment. Barcode distribution of the pooled, 12-plex target capture sample.

Sample	On-Target, %	Target Size, Mb	Fold Enrichment	Number of SMRT Cells
12 Plex	68	1.6	1,372	3

Table 2. The average fold enrichment per sample in a 12-plex capture is 1372-fold, with 68% on-target rate within the 1.6 Mb region.

Characterize Structural Variation

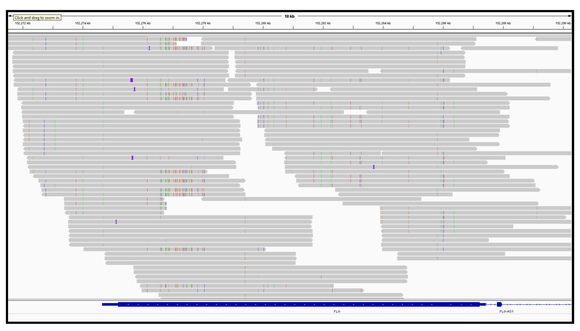


Figure 4A (above). Filaggrin (*FLG*) gene exon 3 spans ~12 kb and contains multiple 1 kb repeats. (A) An IGV screenshot of a captured sample from the pool shows uniform coverage across the region and unambiguous read mapping of the 7 kb fragments.

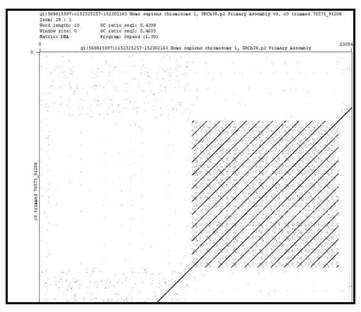


Figure 4B (left). Dot plot of the alignment of the consensus capture sequence to the reference illustrates this sample has twelve 1 kb repeats.

Intronic and Exonic Regions

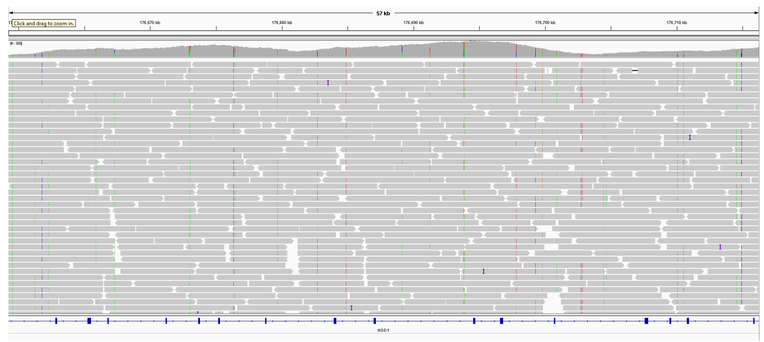


Figure 5. NimbleGen neurology exome panel captures 57 kb region of *NSD1* on chromosome 5. IGV screenshot showing reads mapping to *NSD1*. Even though probes target only exons, capturing longer fragments (7 kb) provides complete coverage of both intronic and exonic regions, showing heterozygous SNPs distributed across the 57 kb region.

Phasing

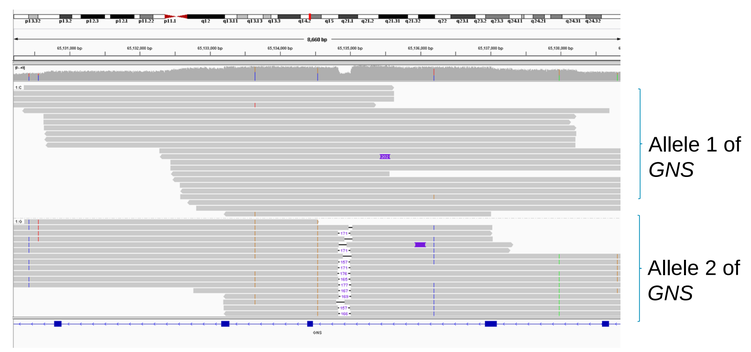


Figure 6. Long reads allow phasing of heterozygous SNPs, resulting in two fully resolved haplotypes.

Conclusions

- Use of linear barcode adapters allows pooling of up to 12 samples prior to capture to reduce overall cost per sample.
- SMRT Sequencing results of NimbleGen neurology panel capture demonstrate even coverage across sample barcodes and over multi-kilobase regions of the genome.
- Long reads can phase SNPs and structural variants, interrogate both intronic and exonic regions, and reveal complex structural variation.

Resources/ Acknowledgements

NimbleGen target capture library generation protocol:

http://www.pacb.com/wp-content/uploads/Procedure-Checklist-Target-Sequence-Capture-Roche-NimbleGen-SeqCapEZ-Library-PacBioBarcodedAdapters.pdf

Application Note: Using NimbleGen target capture library with PacBio sequencing:

http://www.pacb.com/wp-content/uploads/multiplex-target-enrichment-barcoded-multi-kilobase-fragments-probe-based-capture-technologies.pdf

Acknowledgements

The authors would like to thank our collaborators at Roche NimbleGen who have helped to generate this data.