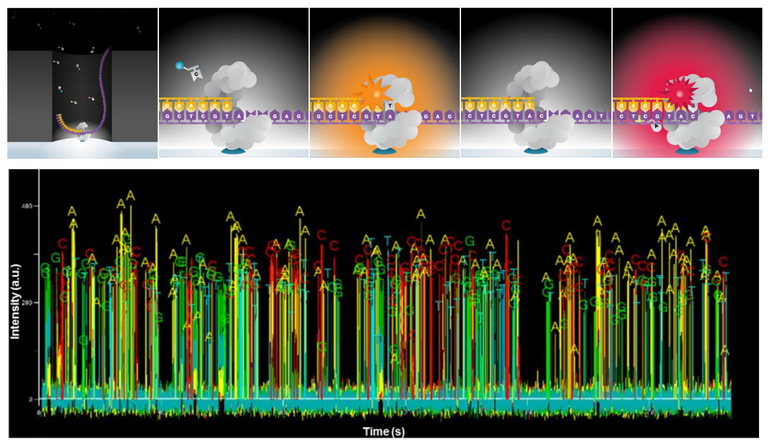
Candidate gene screening using long-read sequencing

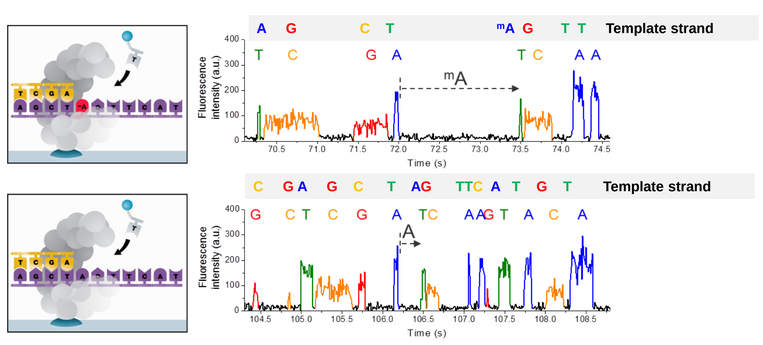
Jenny Ekholm, Ting Hon, Yu-Chih Tsai, David Greenberg, Tyson A. Clark, Steve Kujawa
PacBio, Menlo Park, CA USA

SMRT Sequencing overview

We have developed several candidate gene screening applications for both Neuromuscular and Neurological disorders. The power behind these applications comes from the use of long-read sequencing. It allows us to access previously unresolvable and even unsequencable genomic regions. SMRT Sequencing offers uniform coverage, a lack of sequence context bias, and very high accuracy. In addition, it is also possible to directly detect epigenetic signatures and characterize full-length gene transcripts through assembly-free isoform sequencing.



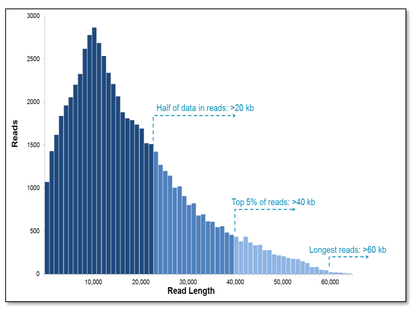
In addition to calling the bases, SMRT Sequencing uses the kinetic information from each nucleotide to distinguish between modified and native bases.



Example: N^6-methyladenine

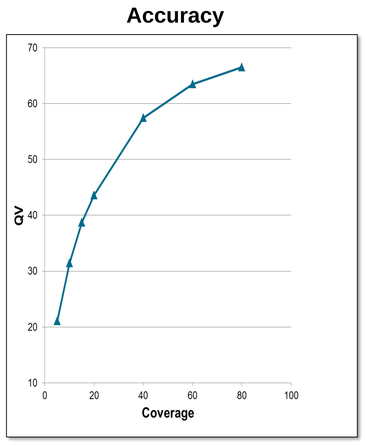
Long-read sequencing data properties

Read Length



Data generated with 20 kb size-selected human library using 6 hour movies with P6-C4 chemistry using PacBio RS II, analyzed with SMRT® Analysis v 2.3. Each SMRT Cell generates ~55,000.

Accuracy



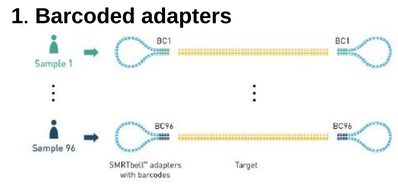
E. coli 20 kb-insert library, SMRT Analysis v2.3

Targeted sequencing and multiplexing

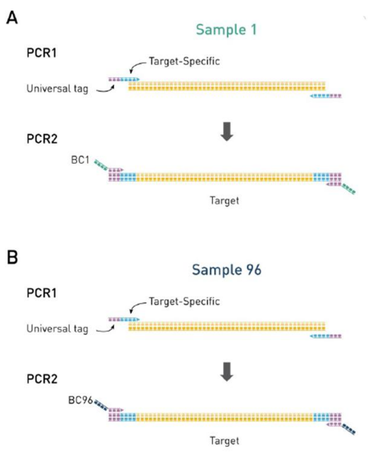
Targeted sequencing workflow

Barcoding options

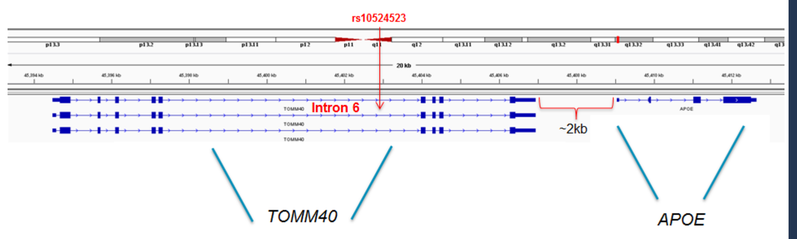
1. Barcoded adapters



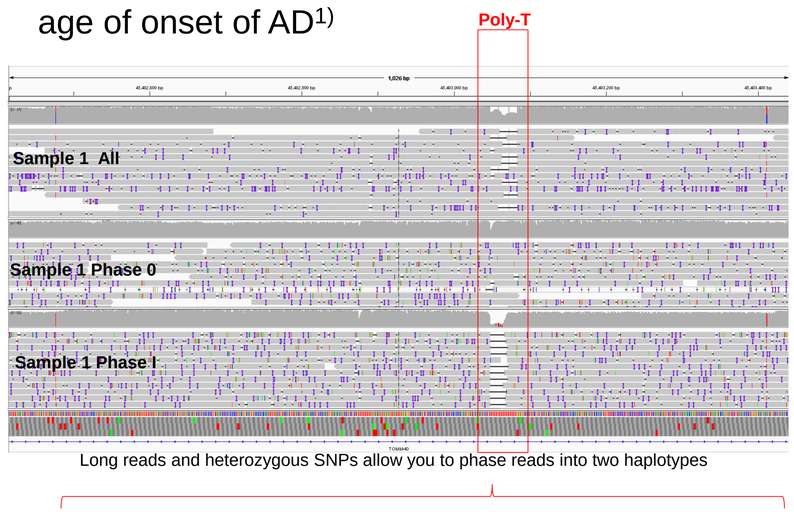
2. Barcoded Universal Primer



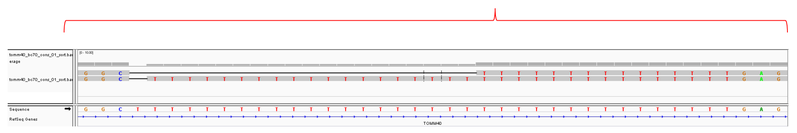
Alzheimer's disease (AD): *TOMM40* gene



A variable poly-T repeat at the rs10524523 SNP within intron 6 of the *TOMM40* gene that in combination with *APOE3* allele will affect the age of onset of AD[1)]

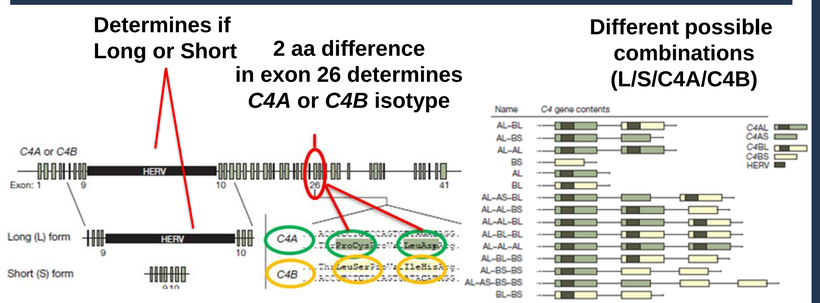


Long reads and heterozygous SNPs allow you to phase reads into two haplotypes



We successfully captured and sequenced the associated poly-T repeat in the *TOMM40* gene and were able to determine that the sample had one short (15) allele and one Very Long (34) allele.

Schizophrenia: *C4* gene



<u>The *C4* structural variant is highly complex:[2)]</u>

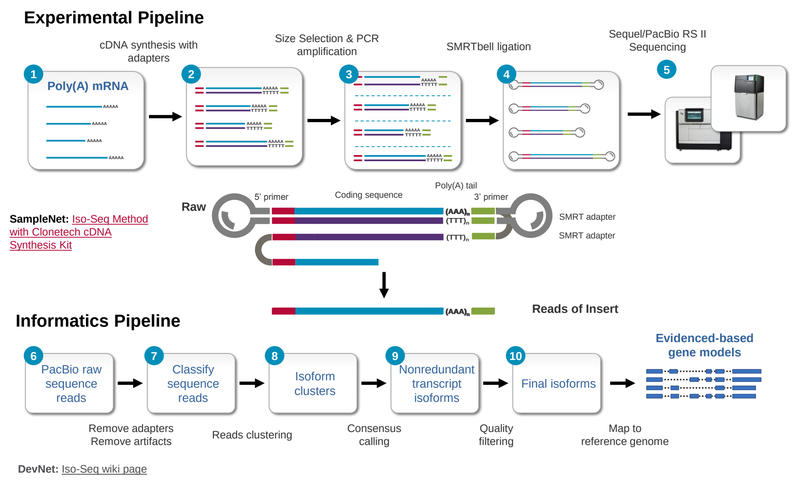
- Two functionally distinct genes (isotypes); *C4A* and *C4B*
- Both isotypes can have 1 - 3 functional copies
- A human endogenous retroviral (HERV) insertion in intron 9 changes the length of the gene

C4 plays a role in signaling which connections between neurons should be “pruned” or removed, as the brain develops after childhood. And the more *C4* was present, the higher the risk of developing schizophrenia. Certain versions of the *C4* gene seem to increase people’s risk for developing schizophrenia by 27 to 50 percent.

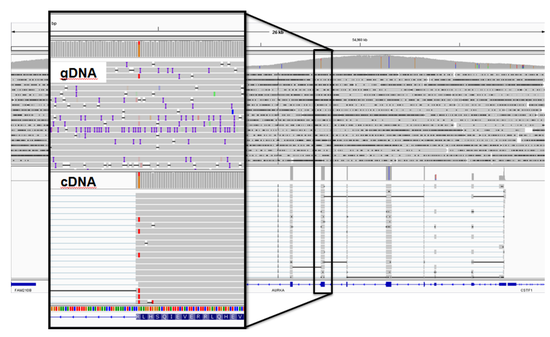
We successfully captured and sequenced the *C4* gene as part of a MHC capture panel. We were able to see the two different *C4* isotypes (*C4A* and *C4B*) as well as seeing the 7 kb HERV insertion in intron 9.

Example: Isotype C4A/L was successfully called for a sample using long-read sequencing

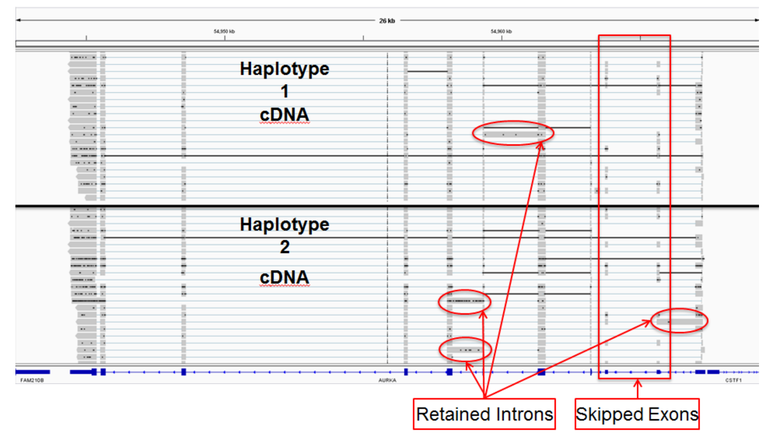
Gene isoform characterization



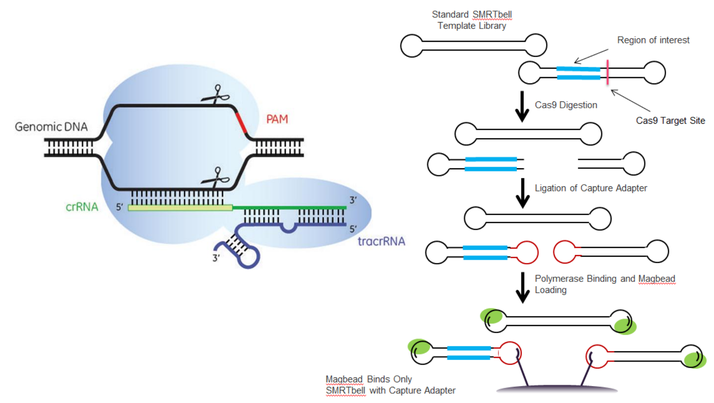
gDNA & Transcripts from SK-BR-3 Cell Line Captured with NimbleGen Oncology Panel - example *AURKA* gene



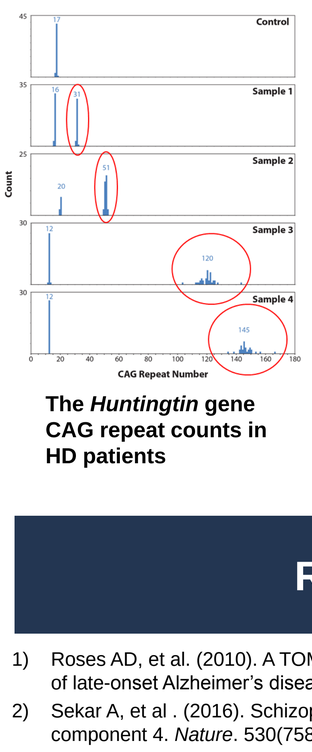
Phased Transcripts reveal retained introns and skipped exons



Amplification-free targeted sequencing using CRISPR/Cas9

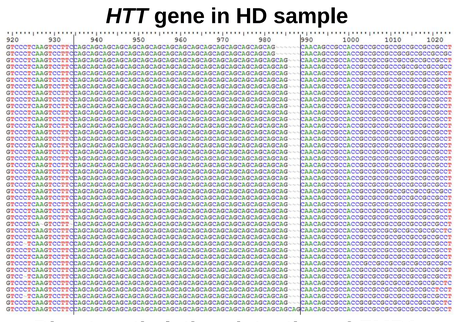


Repeat expansion disorders are challenging to interrogate due to the long repetitive regions. Using CRISRP/Cas9 we are able to access the repeat counts, interruption sequences as well as epigenetic information without introducing PCR bias.

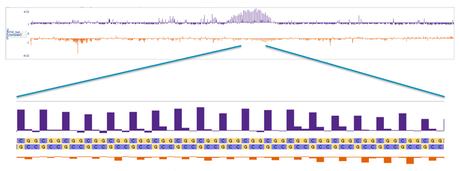


The *Huntingtin* gene CAG repeat counts in HD patients

***HTT* gene in HD sample**



Direct methylation detection of *FMR1* pre-mutation sample



CGG repeat region appears to be heavily methylated (5mC)

References

1) Roses AD, et al. (2010). A TOMM40 variable-length polymorphism predicts the age of late-onset Alzheimer’s disease. Pharmacogenomics J. 10(5): 375-84
2) Sekar A, et al . (2016). Schizophrenia risk from complex variation of complement component 4. *Nature*. 530(7589):177-83

For Research Use Only. Not for use in diagnostics procedures. © Copyright 2016 by Pacific Biosciences of California, Inc. All rights reserved. Pacific Biosciences, the Pacific Biosciences logo, PacBio, SMRT, SMRTbell, Iso-Seq, and Sequel are trademarks of Pacific Biosciences. BluePippin and SageELF are trademarks of Sage Science. NGS-go and NGSengine are trademarks of GenDx. All other trademarks are the sole property of their respective owners.