Single Molecule Real-Time (SMRT®) Sequencing of Genes Implicated In Autosomal Recessive Diseases

Yan Guo[1], Ali Bashir[2], Edith Gould[2], Ruth Kornreich[2], Lisa Edelmann[2], Tracy Brandt[2], Yumi Kasai[2], Jason Chin[1], Ellen E. Paxinos[1], Andrew Kasarskis[2], Eric Schadt[2]

1) Pacific Biosciences, Menlo Park, CA., USA 2) Mount Sinai School of Medicine, Dept. of Genetics and Genomic Sciences, New York, NY

2012 ASHG Conference, San Francisco, California

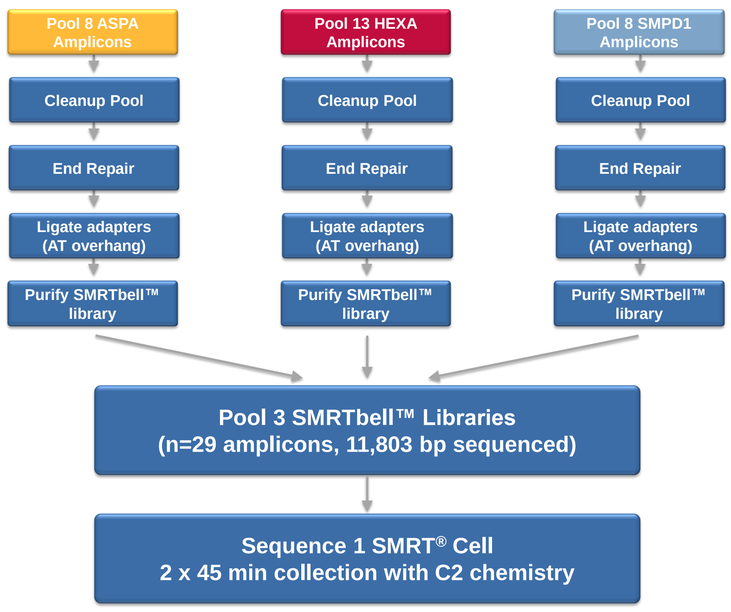
Introduction

- In today's clinical molecular laboratories, disease-causing mutations are mostly detected by genotyping or Sanger sequencing.
- To demonstrate the potential advantages of Single Molecule, Real-Time (SMRT®) sequencing, we conducted a proof-of-concept experiment comparing SMRT sequencing to a clinically-offered test.
- We queried 12,000 bp representing three clinically-relevant genes underlying autosomal recessive diseases in the Ashkenazi Jewish population from samples previously genotyped using multiplex PCR/allele-specific primer extension using the Luminex® FlexMAP™ bead-based approach.

Gene Targets

Disease	Gene	Exons	Amplicons	
			#	Sizes (bp)
Canavan	ASPA	1 – 6	8	401 – 638
Tay-Sachs	HEXA	1 - 14	13	237 - 818
Nieman-Pick	SMPD1	1 – 6	8	425 – 747

SMRT® Sequencing Workflow



SMRT® Sequencing Performance

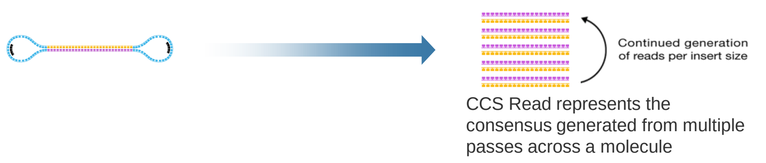
SMRT® sequencing	
# of SMRT® Cells per sample	1
# of Movies per SMRT Cell	2
Readlength	
Mean Mapped Readlength	2,642 bp
95th %ile Mapped Readlength	6,791 bp
Maximum Mapped Readlength	11,094 bp

Standard Metrics	
Accuracy	
Mean Mapped CCS Read Accuracy	98.04%
Mean Mapped Single Pass Accuracy	88.10%
Reads/SMRT® Cell	
Mean # of Mapped CCS Reads	33,574
Mean # of Mapped Total Subreads	315,689
Coverage	
Mean Depth of Coverage	1,828
Mean CCS Read Coverage	240

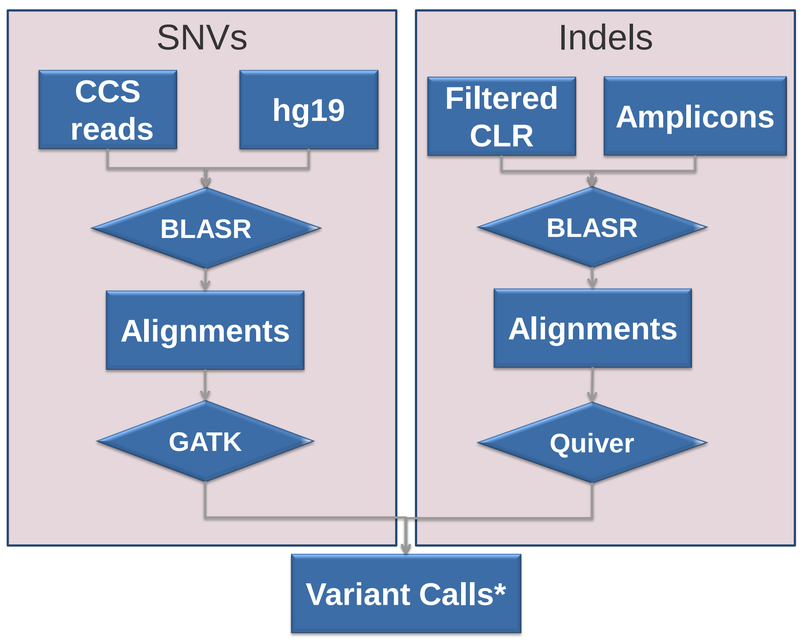
Data and Analysis Workflow

Continuous Long Reads (CLR) data

Circular Consensus Sequencing (CCS) data



Variant Calling Workflow



* confidence cutoff for GATK is >1000, confidence cutoff for Quiver is >90

Detection of Specific Variants

Gene	ASPA	
Variant	p.E285A het[1]	
	Normal	Carrier
Genotyping	N	**Y**
SMRT	N	**Y**

[1] HGVS NM_000049.2: c.854A>C; build 37.3, Chr17: 3402294

Gene	HEXA	
Variant	c.1421+1G>C het[2]	
	Normal	Carrier
Genotyping	N	**Y**
SMRT	N	**Y**

[2] HGVS NM_000520.4: c.1421+1G>C; build 37.3, Chr15: 72638575

Gene	SMPD1	
Variant	c.996delC het[3]	
	Normal	Carrier
Genotyping	N	**Y**
SMRT	N	**Y**

[3] HGVS NM_000543.4: c.996delC; build 37.3 Chr11: 6413291

Other Variation

dbSNP	rs12948217[4]		rs71467507[5]	
Carrier Status	Normal	Carrier	Normal	Carrier
SMRT	N	Y	N	Y

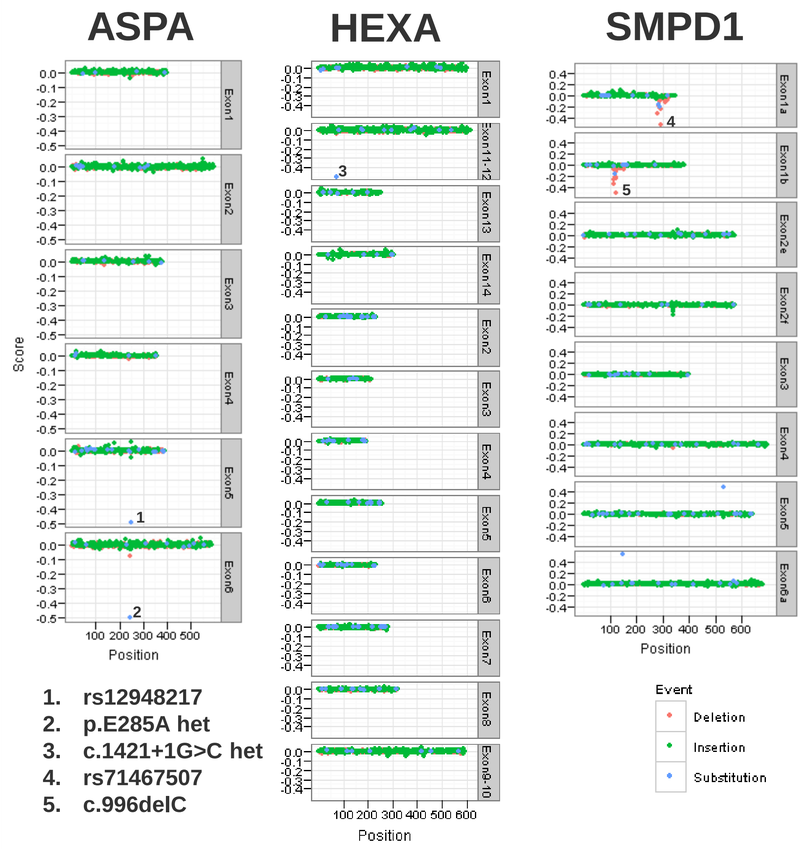
[4] HGVS NC_000017.10:g.3397702C>T; build 37.3, Chr17: 3397702

[5] HGVS NC_000011.9:g.6411930C>T; build 37.3, Chr11: 6411935

[6] HGVS NC_000011.9:g.6412931T>C; build 37.3,Chr11:6412931

[7] HGVS NC_000011.9:g.6415882G>A; build 37.3,chr11:6,415,882

Normal vs. Carrier Sample Comparison



Comparison of normal and carrier sequencing highlights variants. For each amplicon, the score at any base position corresponds to the ratio of bases supporting the reference of the normal sample minus the ratio of bases supporting the reference in the carrier. Positive values indicate variation in the normal sample. Negative values correspond to variation in the carrier sample.

Quiver Detects Homopolymer Deletions

Region in SMPD1 in which Quiver correctly calls homopolymer deletion. Naïve alignment strategies lead to difficulty in placing deletions within homopolymers.

Conclusions

- From a single SMRT® Cell, we queried ~12,000 bases from 29 amplicons of 3 genes with high coverage, and can easily called known variants.
- A single SMRT Cell can be sequenced in 90 minutes, suggesting a rapid, cost-effective method for routine sequencing entire genes.
- Quiver (https://github.com/PacificBiosciences/GenomicConsensus) allows indel calling even in homopolymers.
- Using SMRT Sequencing, we were able to detect variants not assayed with standard genotyping panels.

Reference:
Mauricio O Carneiro[1*], Carsten Russ[2], Michael G Ross[2], Stacey B Gabriel[1], Chad Nusbaum[2] and Mark A DePristo[1] **Pacific biosciences sequencing technology for genotyping and variation discovery in human data** Carneiro et al. BMC Genomics 2012, 13:375 http://www.biomedcentral.com/1471-2164/13/375

Acknowledgements

The authors would like to thank Anand Sethuraman and David Alexander for their generous help in running QUIVER and generating plots for the poster.