ASHG 2017: #1492W

Development of a Dual Platform Strategy for Targeted DNA Sequencing in Genetic Screening

Cassie Schumacher, Ashley Wood, Sukhinder Sandhu, Justin Lenhart, Laurie Kurihara, Timothy Harkins, Vladimir Makarov

Swift Biosciences, 58 Parkland Plaza, Suite 100, Ann Arbor, MI 48103

Swift BIOSCIENCES™

Abstract

The use of next-generation sequencing (NGS) to assess genetic variation in genes involved in both inherited diseases, as well as cancer, is critical. Traditional NGS assays utilize short-read chemistry to enable sequencing of biomarkers in accessible parts of the genome, devoid of difficult-to-sequence motifs and high homology to pseudogenes. Newer long-read assays offer additional insight into the genetic architecture of known genes including haplotype structure, detection of large insertions/deletions, variants present in regulatory elements, and copy number variations. Both massively parallel screening techniques are economical, and a turn-around time of less than 48 hours makes them ideal for high-throughput use. To assess the utility of both short- and long-read platforms, Swift Biosciences has developed single-tube, multiplexed amplicon assays for both CFTR and BRCA1/BRCA2 (BRCA) genes.

DNA from MUTCF-2 (Coriell Institute) with known CFTR variants was used in both the short- and long-read assays. In the short-read assay, an 87-amplicon panel comprehensively covered all exons (~10kb) in CFTR. 10 ng of DNA was used as input for multiplexed PCR, and the products were adapted and sequenced on an Illumina® MiniSeq®. The long-read assay covered all the same exons, and 6 complete intronic regions, using 21 amplicons (~54kb). 25 ng of DNA was used as input for the multiplexed PCR, barcoded for sample multiplexing, and adapted for Pacific Biosciences® (PacBio®) sequencing on an RS II. To assess the BRCA genes, DNA from BC01 (Coriell Institute) was used as input into the 246-amplicon (~23kb) short-read panel which comprehensively covers all BRCA exons. A 35-amplicon panel (~86kb), was used to cover these same exons and 20 complete intronic regions with the long-read assay. Both assays were adapted and sequenced as described for CFTR.

Libraries from both short-read assays were sequenced with > 95% on-target and coverage uniformity. Based on the quality of the consensus sequence and the low number of multiplexed amplicons, the long-read assays require up to 10x less sequencing reads per amplicon to call the same variants as the short-read assays. The same number of variants were detected in both assays, despite the difference in sequencing depth. The long-read assay identified indels, such as a 40 bp deletion in BRCA, that are challenging to detect with short reads, indicating that large insertions/deletions are more robustly detectable with long-read technology.

Dual Platform Approach

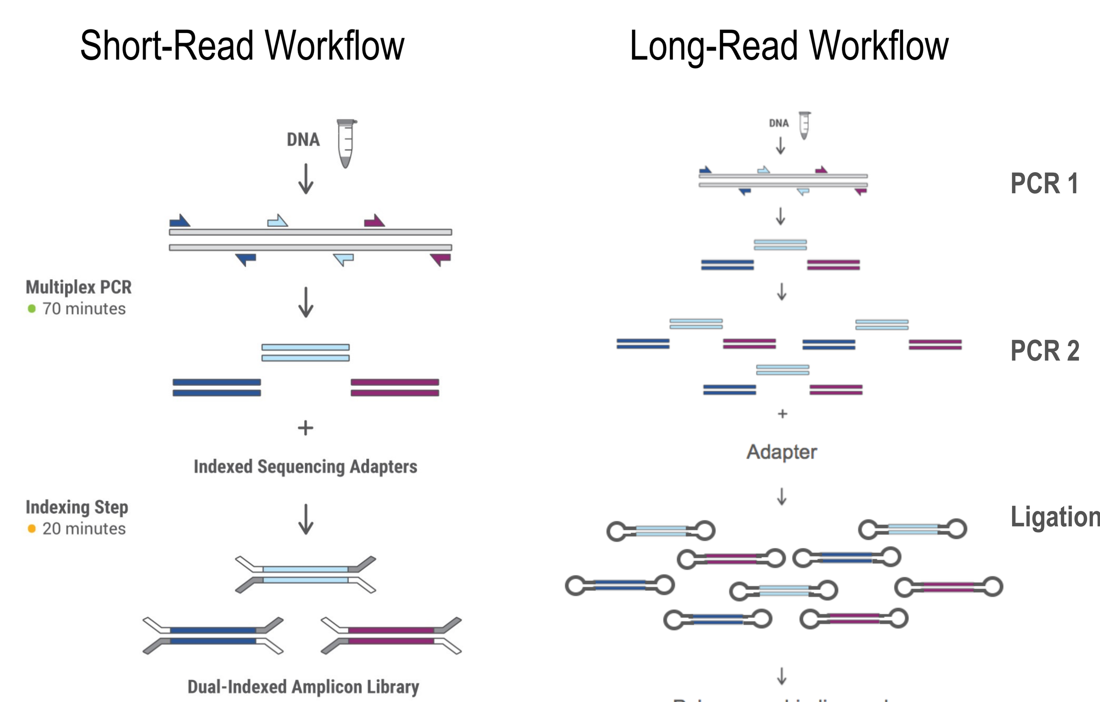
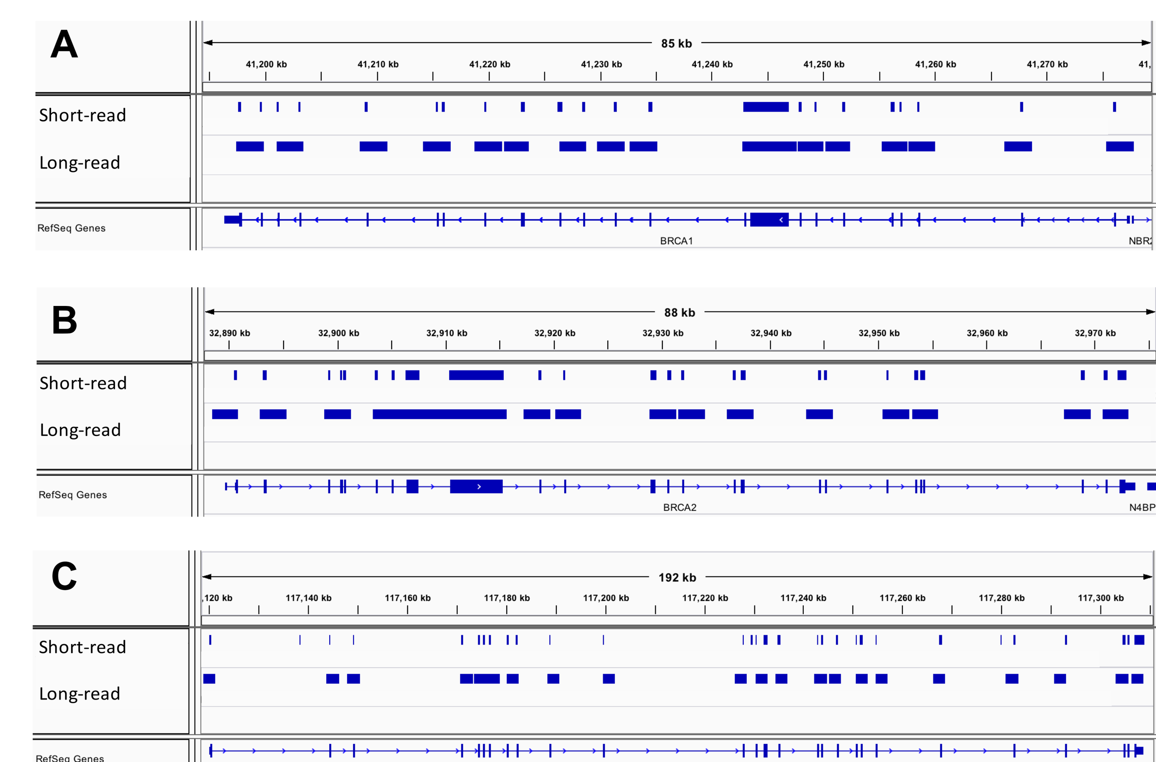


Figure 1. Short- and long-read workflows. (Left) The short-read workflow consists of two steps: multiplexed PCR followed by an indexing step to barcode and adapt the PCR products for Illumina sequencing. (Right) The long-read workflow consists of three steps, two PCR steps to amplify the target sequences and a ligation step to barcode and adapt the PCR products for PacBio sequencing.

BRCA1, BRCA2, CFTR Panel Specifications



D

Panel	# of Amplicons	Amplicon Size (kb)	Panel Size/ Gene Size (kb)	Tiling?
BRCA1/2 - Short	246	0.15	23/165	Yes
BRCA1/2 - Long	36	2.5	90/165	Yes
CFTR - Short	87	0.14	10/189	Yes
CFTR - Long	21	2.5	53/189	No

Figure 2. Top panels show amplicon coverage of the exons and intronic regions of BRCA1 (A), BRCA2 (B) and CFTR (C). Within each panel, top track shows short-read coverage, middle track shows long-read coverage, and bottom track shows location of the exons. Panels are screen shots from the Integrative Genome Viewer (IGV, Broad Institute). Table lists specifications for short- and long-read assays for both BRCA1/2 and CFTR. Overlapping, tiled regions exist in the BRCA1/2 panel.

Sequencing Metrics

BRCA1/2						
	# Aligned Reads		% On Target		Coverage Uniformity	
DNA	Short-Read	Long-Read	Short-Read	Long-Read	Short-Read	Long-Read
NA14623	2552437	6203	99%	98%	100%	100%
NA14624	2763174	6061	98%	98%	100%	100%
NA14626	2495232	6497	98%	98%	100%	97%
NA13705	2031846	2416	98%	98%	99%	100%
NA13715	1777741	2070	98%	98%	100%	97%
NA14090	1488832	2449	98%	98%	96%	100%
NA14094	1582618	2356	98%	98%	100%	97%
NA14638	1484737	2526	98%	98%	86%	100%
NA14634	2748198	2686	98%	98%	100%	97%
NA14636	2683304	3128	98%	98%	100%	100%
NA14637	2389220	3010	99%	98%	100%	100%
NA14170	1970282	3381	99%	98%	100%	97%

CFTR						
	# Aligned Reads		% On Target		Coverage Uniformity	
DNA	Short-Read	Long-Read	Short-Read	Long-Read	Short-Read	Long-Read
NA07552	496886	4005	99.9%	98%	100%	95%
NA08338	360324	2923	99.9%	98%	100%	100%
NA11282	1166387	2432	99.6%	98%	97%	100%
NA01531	406033	1672	99.9%	98%	100%	90%
NA07441	287558	2104	99.9%	98%	100%	100%
NA11275	409414	2094	99.9%	98%	100%	100%
NA11277	384345	3740	99.9%	98%	100%	100%
NA11280	383071	2117	99.9%	98%	99%	100%
NA11281	1178533	2677	99.7%	98%	94%	95%
NA11283	390719	2459	99.8%	98%	100%	95%
NA11284	532043	1830	99.8%	98%	99%	100%
NA11472	15561963	1656	99.9%	98%	100%	95%

Figure 3. Sequencing metrics for short- and long-read assays for both BRCA1/2 and CFTR. Tables list the number of aligned reads, percent of reads on target, and the coverage uniformity in both assay types. Left table shows data from BRCA1/2 assay while right table shows data from CFTR assay. BRCA1/2 DNA samples are a subset of the BC01 gene mutation panel; CFTR DNA samples are a subset of the MUTCF-2 gene mutation panel (Coriell Institute). All short-read samples were sequenced on an Illumina MiniSeq; BRCA1/2 long-read samples were sequenced on a PacBio Sequel™ while CFTR long-read samples were sequenced on a PacBio RS II. Aligned reads for all long-read samples are derived from CCS read numbers.

Sequencing Quality

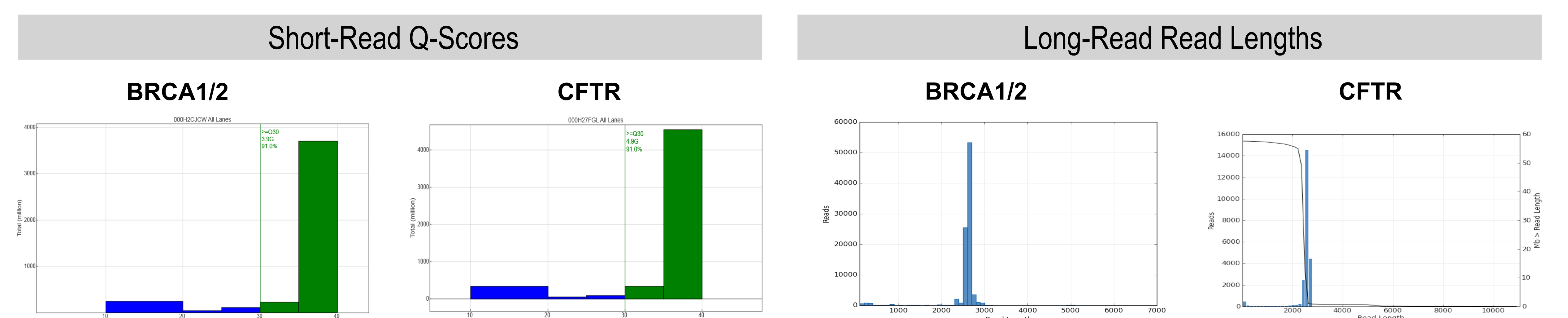


Figure 4. Sequencing quality across panels and platforms. To show quality of sequence obtained for each panel and platform, Q-scores for short-read sequencing and read lengths for long-read sequencing were examined. Both BRCA1/2 and CFTR had Q-scores above 90% for the short-read panels, while the read lengths for both the BRCA1/2 and CFTR long-read panels showed max peaks near 2500 bases, indicating that the full amplicon length was successfully sequenced.

Variant Sequencing Using Short- and Long-Read Technology

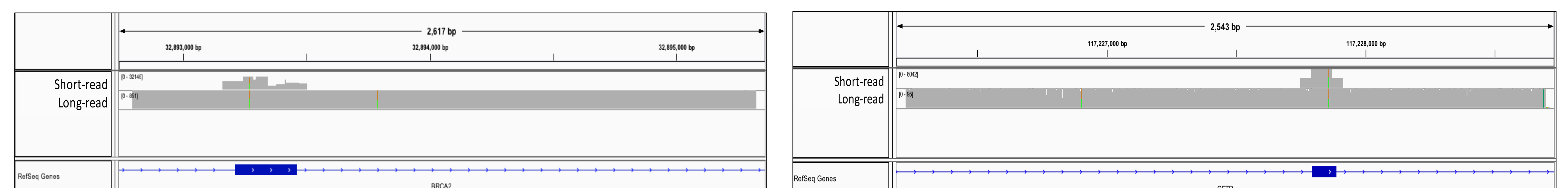


Figure 5. Sequence analysis using IGV. After data analysis, BAM files were loaded into IGV to visually inspect known variants. (Left) The known variant in NA14623 in BRCA2 is shown in the left panel, while the known variant in NA08338 in CFTR is shown in the right panel. The top track in each window shows the coverage obtained from the short-read panel while the bottom track shows the coverage obtained by the long-read panel. Note that the mutation of interest in the coding region is seen using both techniques in both genes, but an additional variant can be seen in the neighboring intron of both genes using the long-read technology.

Variant Analysis Using Short- and Long-Read Technology

BRCA1/2					
DNA	Gene	Variant	Expected AF	Observed AF Short-Read	Observed AF Long-Read
NA14623	BRCA2	TYR42CYS	0.50	0.53	0.47
NA14624	BRCA2	5946delCT	0.50	0.52	0.49
NA14626	BRCA2	LYS3326TER	0.50	0.50	0.54
NA13705	BRCA1	4-BP DEL, FS1252TER	0.50	0.31	0.45
NA13715	BRCA1	1-BP INS, 5382C	0.50	0.47	0.43
NA14090	BRCA1	2-BP DEL, 185AG	0.50	0.46	0.46
NA14094	BRCA1	40-BP DEL, FS397TER	0.50	0.44	0.56
NA14638	BRCA1	IVS5-11T>G	0.50	0.56	0.49
NA14634	BRCA1	4-BP DEL, FS1364TER	0.50	0.51	0.51
NA14636	BRCA1	5677insA	0.50	0.53	0.53
NA14637	BRCA1	ARG1443TER	0.50	0.47	0.48
NA14170	BRCA2	1-BP DEL, 6174T, FS	0.50	0.55	0.44

CFTR								
DNA	Mutation Allele 1	Expected AF	Observed AF Short-Read	Observed AF Long-Read	Mutation Allele 2	Expected AF	Observed AF Short-Read	Observed AF Long-Read
NA07552	PHE508DEL (rs113993960, CTT>-)	0.5	0.52	0.5	1789C>T(Arg553>TER) (rs74597325)	0.5	0.51	0.63
NA08338	GLY551ASP (rs75527207)	0.5	0.52	0.62	N/A	N/A	N/A	N/A
NA11282	621+1G>T	0.5	0.49	0.49	GLY85GLU (G>A nucleotide 386, exon 3)	0.5	0.46	0.44
NA01531	PHE508DEL	1	1	1	PHE508DEL	1	1	1
NA07441	IVS16,G>A,+1 (3120+1G>A)	0.5	0.5	0.49	621+1G>T	0.5	0.51	0.52
NA11275	1-BP DEL, 3659C (exon 22)	0.5	0.52	0.43	PHE508DEL	0.5	0.49	0.08
NA11277	ILE507DEL	0.5	0.54	0.45	N/A	N/A	N/A	N/A
NA11280	621+1G>T	0.5	0.49	0.43	711+1G>T	0.5	0.52	0.24
NA11281	621+1G>T	0.5	0.48	0.42	PHE508DEL	0.5	0.53	0.75
NA11283	ALA455GLU (1496 C>A)	0.5	0.51	0.5	PHE508DEL	0.5	0.5	0.5
NA11284	ARG560THR (1811 G>C)	0.5	0.52	0.45	PHE508DEL	0.5	0.52	0.13
NA11472	GLY1349ASP (G>A)	0.5	0.5	0.38	ASN1303LYS (C>G)	0.5	0.51	0.68

Figure 6. Variant calling across panels and platforms. Known variants for all 12 samples in both BRCA1/2 and CFTR genes are detailed. The expected allele frequency (AF) is given, along with the observed AF in both the short-read and long-read assay. All known variants were identified using both techniques.

Conclusion

- Accel-Amplicon™ panels from Swift Biosciences can be used on short- and long-read sequencing technologies for variant detection.
- Short-read amplicon panels are useful tools to interrogate variants of known significance in coding regions and intron/exon boundaries.
- Long-read amplicon panels are useful for full gene coverage to not only analyze variants in the coding region, but to also probe neighboring introns which can be difficult to target with short-read amplicons due to repetitive regions and low complexity motifs.
- Long-read technology provides more accurate alignment to identify structural variants because it does not rely on short-read assembly.
- Both short- and long-read amplicon panel workflows can be completed in a single day and have the ability to process samples in a high-throughput manner.

Swift Biosciences, Inc.
58 Parkland Plaza, Suite 100 • Ann Arbor, MI 48103 wwwswiftbiosci.com

© 2017, Swift Biosciences, Inc. The Swift logo and Accel-Amplicon are trademarks of Swift Biosciences. This product is for Research Use Only. Not for use in diagnostic procedures. Illumina and MiniSeq are registered trademarks and Sequel is a trademark of Illumina, Inc. Pacific Biosciences and PacBio are registered trademarks of Pacific Biosciences of California, Inc. 17-1765, 10/17

www.swiftbiosci.com