Adaptation of MAS-Seq/PacBio and ONT long read technology to the ResolveOME™ multiomic workflow for single-cell transcript isoform and DNA amplicon interrogation

Victor J. Weigman, Martin A. Newman, Isai Salas-Gonzalez, Jon S. Zawistowski

BioSkryb Genomics, Durham, NC, USA

Introduction

The ability to accurately detect genomic structural variation (SV) is a crucial spoke in deciphering the multifaceted mechanisms of oncogenesis occurring in concert. There is often ambiguity in calling SV with short-read technologies, at both DNA and RNA levels of resolution, which can begin to be mitigated by long-read sequencing. Ascertainment of differential transcript isoform utilization (DTU) in disease or drug-resistant states also stands to directly benefit from long read and alleviate short read ambiguity. Here we embarked on studies to contrast SV and DTU calling between short and long read sequencing, at the single-cell level, with ResolveDNA genomic amplification and ResolveOME joint genomic/transcriptomic profiling [1]. We employed two long-read platforms, PacBio Revio and Oxford Nanopore Technologies (ONT) PromethION to report DNA- or RNA-level single-cell data using a cell line panel of diverse types with diverse states of copy number alteration. Transcriptomically, we devised a MAS-Seq strategy with a 16-plex of matched parental and drug-resistant triple-negative breast cancer or acute myeloid leukemia single cells as well as euploid HG001 cells to compare the full-length ResolveOME cDNA directly sequenced by the long-read platforms or fragmented for short-read readouts. Genomically, we enriched for long ResolveDNA or ResolveOME amplification products and either enzymatically fragmented these for short-read sequencing or directly sequenced using both long-read platforms. Dependent on the researcher's biological interest, we demonstrate here the feasibility of a forked ResolveDNA/OME workflow that presents a long-read option.

Results – RNA Expression

Considering the ResolveOME RNA amplification approach, sections up to 5kb (limit of reverse transcriptase) are readily observed. We wanted to observe how well amplification of native RNA species would work when protocol was not fragmented and fed to long read sequencing.

The overall quantification work here covers basic transcript and gene detection and scope of this work does not serve to suggest best methods with each technology, just leverage most popular methods, based on publication. This is an initial survey and methods of normalization and batch correction will need to be honed to provide best mechanism of comparison.

Figure 2: Principal Components of Detected Transcripts. PC1 and PC2 loadings are shown for Illumina (green), ONT (red) and PacBio (blue) with corresponding area clouds.

	Transcripts Detected	% Tx Detected	Total Genes	Tx / Gene	20% Tx Len	50% Tx Len	80% Tx Len	Tx Max*
ONT	29,699	24.2%	14,439	2.05686	641	1,561	2,973	5,171
ILMN	116,574	94.8%	30,167	3.864289	568	1,162	2,987	---
PB	13,175	10.7%	5,656	2.329385	1,062	2,076	3,722	6,538

*Maximum transcript is a 95% quantile of total lengths detected. For Illumina could not be assembled.

Table 2. Performance Across Technologies. Overall summary across cell types showing what is detected across technologies.

We were successful in being able to get full-length reads across the full isoform sequence identified in Table 2 for long-read sequences. Detection of isoforms with at least one long read increases these max over 10kb, with multiple examples.

Comparability of these transcripts across technologies follows trends seen in single cell expression sequencing and within our Illumina workflow, we observe sparsity average of 87%.

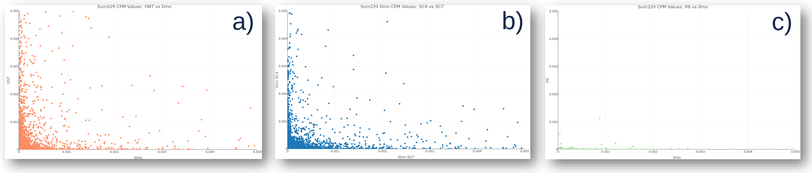

Figure 3. Recall of expected Fusions in Sum229. Comparisons of a) ONT vs. Illumina, b)Illumina against itself and c)Illumina vs. PacBio. Individual cells showed similar trends within cell group

Conclusions

- Modified ResolveOME protocol generated high quality DNA and RNA data simultaneously in both short and long read methods
- We can detect a very high number of isoforms / cell across sequencing technologies
- The workflow is robust to detect fusions present within the cell
- Additional modifications will be needed to better sync with long-read workflows

Methods

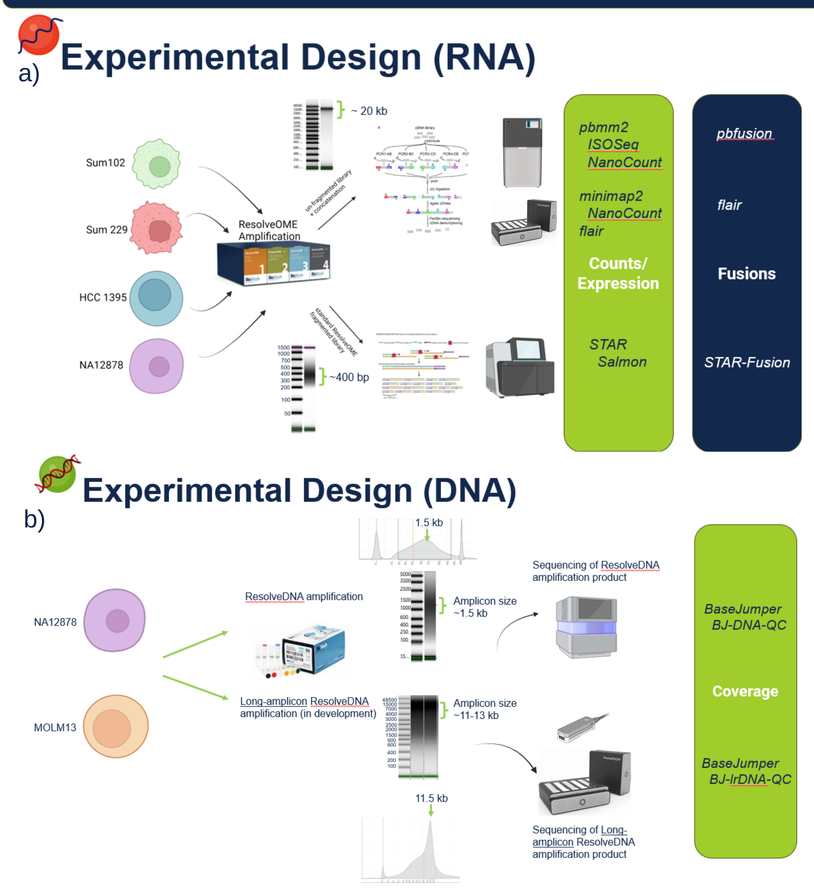

Figure 1: Modified ResolveOME workflow. The ResolveOME workflow was modified to remove fragmentation portion for long read sequencing methodologies but kept consistent for short read in **A)** the RNA fraction and **B)** DNA Fraction for cell lines. To take advantage of the longer reads, transcript concatenation was performed prior to sequencing, outlined in [2].

a)

Cell Line	Molecule	Technology	Instrument	Filtered Reads	Mean Read Length	Total Bases
NA12878	RNA	PacBio	Sequel	405,424	830	342,953,712
Molm13-P	RNA	PacBio	Sequel	155,991	767	122,686,857
Molm13-R	RNA	PacBio	Sequel	155,509	779	124,147,978
HCC1395	RNA	PacBio	Sequel	501,680	806	413,276,050
HCC1395BL	RNA	PacBio	Sequel	531,868	850	476,338,915
Sum102	RNA	PacBio	Sequel	498,383	870	441,543,559
Sum229	RNA	PacBio	Sequel	646,608	910	589,755,546
Molm13-P	RNA	ONT	PromethION	6,269,973	875	5,310,994,232
Molm13-R	RNA	ONT	PromethION	6,239,669	868	5,268,440,571
Sum102	RNA	ONT	PromethION	5,493,784	995	5,477,186,705
Sum229	RNA	ONT	PromethION	4,339,812	1,021	4,431,702,511
HCC1395	RNA	Illumina	NextSeq	1,586,271	51	80,173,924
HCC1395BL	RNA	Illumina	NextSeq	1,367,657	51	69,140,133
Molm13-P	RNA	Illumina	NextSeq	1,617,707	51	81,814,741
Molm13-R	RNA	Illumina	NextSeq	1,626,143	51	82,253,615
Sum102	RNA	Illumina	NextSeq	1,623,036	51	82,109,708
Sum229	RNA	Illumina	NextSeq	1,611,439	51	81,492,947

b)

Cell Line	Molecule	Technology	Instrument	Filtered Reads	Mean Read Length	Total Bases
Molm13-P	DNA	ONT	MinION	16,714	4,087	68,100,685
Molm13-R	DNA	ONT	MinION	18,557	4,050	74,804,473
NA12878	DNA	ONT	MinION	21,131	3,860	79,225,205
Molm13-P	DNA	ONT	PromethION	249,875	3,703	926,133,060
Molm13-R	DNA	ONT	PromethION	334,349	3,837	1,276,675,470
NA12878	DNA	ONT	PromethION	270,150	3,363	886,593,341
HCC1395	DNA	Illumina	NextSeq	1,753,243	75	129,740,001
HCC1395BL	DNA	Illumina	NextSeq	1,750,239	75	129,517,714
Molm13-P	DNA	Illumina	NextSeq	1,751,328	75	129,598,235
Molm13-R	DNA	Illumina	NextSeq	1,774,782	75	131,333,896
Sum102	DNA	Illumina	NextSeq	1,758,148	75	130,102,961
Sum229	DNA	Illumina	NextSeq	1,776,179	75	131,437,209

Table 1. Sequencing Results in Design. Sequencing reads were determined to provide a roughly similar size of sequencing we perform in standard protocols. All values are summaries over several cells. **A)** Metrics for samples in RNA workflow and **B)** DNA workflow

Results – RNA Fusion Calling

For this part of the analysis reads from Table 1A were use with PacBio, ONT and Illumina reads. *STAR-Fusion* was used for Illumina, *Flair* for ONT and *pbfusion* for PacBio reads with recommended settings and workflow. NA12878 was omitted (DepMap[3] does not report any).

Quite unexpectedly, we were unable to detect fusions in long-read technologies, despite counts observed for most of the genes that are involved in the fusion. More analytical investigation will need to be performed, With fragments up to 20kb, we are carrying through fusion products.

	Expected*	Illumina	PacBio	ONT
HCC 1395	36	EIF3K – CYP39A FTH1—MT-CO2 PLA2R1—RBMS1	--	--
Sum102	2	ERP44—RIN2 UCK2—GLRX2	--	--
Sum 229	9	HNRNPA2B1--RPL4	--	--

*expected from DepMap fusions.csv with at least 1 spanning read

Table 3. Recall of expected Fusions. Comparisons of fusions returned from fusion-calling pipelines from individual cells sequenced at depths in Table 1A

Results – DNA Copy Number

One of the hallmarks of the ResolveOME DNA arm is robustness at copy number calling[1]. The two cell lines show expected behavior in short read but appear to be more noisy with longer read data, even with the increased reads from PromethION.

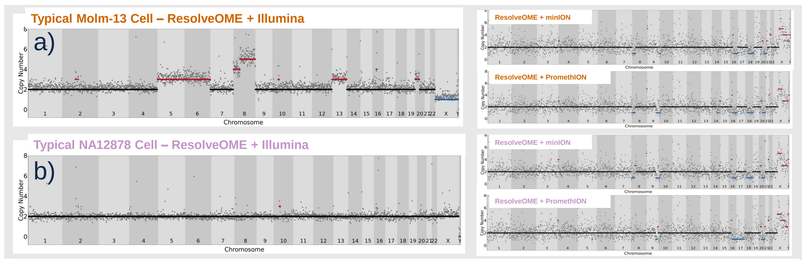

Figure 4. Performance of Copy Number with different libraries. a) Molm13 cells have expected amplifications found with short-read workflow b)NA12878 showing their quiescence.

Results – DNA Genome Coverage

In comparison of genome coverage between multiple technologies, we wanted to look at features that typically convey good QC and compare between ONT (red: circles-MinION/squares-PromethION) and Illumina (green triangles). Very consistent were mapping, and genome coverage per sequenced base).

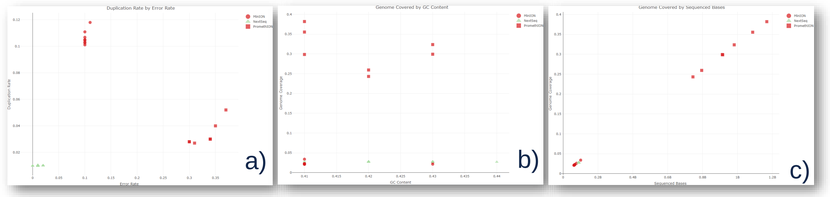

Figure 5. DNA Alignment QC. Comparisons of common alignment metrics. a) Duplication Rates vs. Error Rates, b) Genome Covered by GC Content and c) Genome Coverage by Sequenced bases

Sample	Sequencer	Mapped Reads	Sequenced Bases	Dup Rate	Mean Insert	GC %	Error Rate	Mean Coverage	Coverage Stdev	%chrM
Molm13-P	MinION	1,212,772	65,487,597	0.103	4,087	41.51%	0.107	0.022	0.211	0.31%
Molm13-R	MinION	1,487,569	71,519,490	0.108	4,050	41.68%	0.111	0.024	0.272	1.10%
NA12878	MinION	1,358,147	76,504,840	0.102	3,860	41.00%	0.103	0.026	0.212	0.11%
Molm13-P	PromethION	38,408,553	914,724,483	0.328	3,703	42.08%	0.033	0.299	0.970	0.28%
Molm13-R	PromethION	48,754,041	1,074,513,216	0.354	3,837	42.24%	0.041	0.353	2.155	1.21%
NA12878	PromethION	36,106,963	854,926,380	0.305	2,933	41.58%	0.027	0.279	0.770	0.10%
Molm13-P	NextSeq	1,790,753	88,762,158	0.007	295	42.44%	0.010	0.028	0.689	1.14%
Molm13-R	NextSeq	1,801,828	89,309,775	0.006	299	42.04%	0.010	0.028	0.446	0.86%
NA12878	NextSeq	1,799,236	88,887,303	0.011	280	42.43%	0.010	0.028	0.696	1.70%

Table 4. Sequencing Results. Sequencing reads were determined to provide a roughly similar size of sequencing we perform in standard protocols. All values are summaries over several cells. **A)** Metrics for samples in RNA workflow and **B)** DNA workflow

References

1, Jon S. Zawistowski, Isai Salas-González, Tatiana V. Morozova et al. Unifying genomics and transcriptomics in single cells with ResolveOME amplification chemistry to illuminate oncogenic and drug resistance mechanisms. bioRxiv 2022.04.29
2. Al'Khafaji, A.M., Smith, J.T., Garimella, K.V. et al. High-throughput RNA isoform sequencing using programmed cDNA concatenation. Nat Biotechnology (2023).
3. DepMap Portal. Fusion data pulled from Fusions Public 23Q4 https://depmap.org/portal/download/custom/.

Acknowledgements

We thank Julien Curaba, Julian Abery, and Hamed Bostan of Eremid Genomic Services (Kannapolis, NC) for generation and computational demultiplexing of .MAS-seq libraries.
We thank Piotr Mieczkowski, Tara Skelly, and Hemant Kelkar at the University of North Carolina School of Medicine High-Throughput Sequencing Facility for ONT data and processing.