

Technical Note

ResolveOME: Comprehensive Multi-omic Single-Cell Analysis

Cells explored. Answers revealed.

Authors

Jeff Blackinton, PhD

Director
New Product Introduction
BioSkryb Genomics

Tatiana Morozova, PhD

Senior Scientist
Research & Development
BioSkryb Genomics

Jon Zawistowski, PhD

Senior Director
Research & Development
BioSkryb Genomics

Isai Sales-Gonzalez, PhD

Computational Biologist
Bioinformatics
BioSkryb Genomics

Durga Arvapalli, PhD

Scientist I
Research & Development
BioSkryb Genomics

Swetha Velivela, PhD

Scientist I
Research & Development
BioSkryb Genomics

Jay A.A. West, PhD

President, CEO, Co-founder
BioSkryb Genomics

The ResolveOME Platform for Comprehensive Analysis of Multi-omic Layers of Single-cell Biology

BioSkryb Genomics has developed a unified system for single cell whole transcriptome and whole genome amplification sequencing analysis. We describe here the overall workflow and essential data metrics enabling the analysis integration of the combined RNA and DNA components of single cells. The system allows comprehensive delineation of genomic variance at both the transcriptional and translational molecular layers of individual cells that comprise mammalian tissues.

Product Highlights

- A multi-omic approach to analyzing the complete transcriptome and genome of single cells
- Whole mRNA transcriptome analysis using full length transcripts
- Whole Genome Amplification (WGA) enabling:
 - Whole Genome Sequencing (WGS)
 - Whole Exome Sequencing (WES)
 - Targeted Enrichment Panels
- Adaptable to targeted protein analysis using Oligo-linked Antibodies

Introduction

The central dogma of molecular biology states that the genome of every single cell contains the complete set of instructions for all biological life. The omic layers, comprised of the entire genome, transcriptome and proteome, synchronize to translate and decode this genetic code to dictate cellular fate. Normal and pathologic cellular development are dependent on the fidelity of this transformative process. They are dependent on the OME, or the sum of ALL of the parts.

Elucidating these individual molecular layers or OMEs has been of significant interest^{1,2} for decades and has driven the development of multiple omic based methodologies, including single cell genome sequencing³⁻⁵, transcriptome sequencing^{6,7} as well as methods for targeted protein detection in single cells⁸⁻¹⁰.

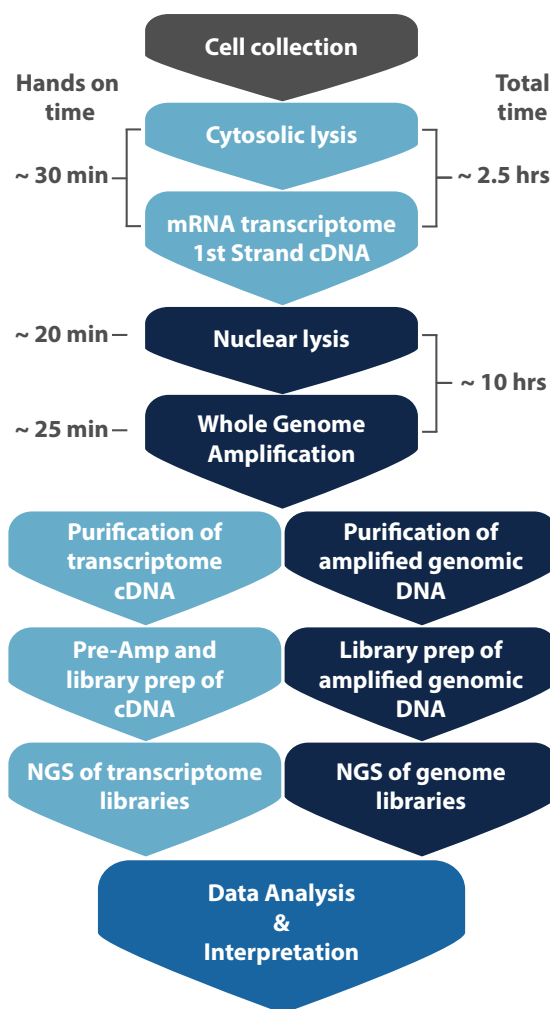


Figure 1. ResolveOME System Workflow enables the integration of first strand cDNA synthesis and genome amplification with downstream separation of the multiomic libraries for NGS analysis.

However, combining these methodologies, particularly with high coverage genome analysis, has been challenging in terms of the data quality¹⁰. The ResolveOME product solution was developed to resolve this challenge. The workflow (Figure 1) begins with the isolation of single cells, prior to cytosolic lysis and reverse transcription. Once complete, the same samples are processed using Primary Template-directed Amplification (PTA) to perform whole genome amplification. After bead-based separation and amplification of the transcriptome cDNA pool, distinct libraries from both the transcriptome and genome fraction are prepared for downstream next generation sequencing and analysis.

ResolveOME is built upon the foundation of PTA. Previous methodologies for whole genome amplification of single cells and picogram (pg) quantities of DNA, such as MDA, have not been able to provide the breadth and uniformity of genomic coverage required for robust variant detection³⁻⁵. With the development of PTA¹¹, individual cells and low DNA inputs can now be amplified by ResolveDNA with high uniformity, providing a broad sequencing breadth and the highest variant calling sensitivity.

Building on the performance of the PTA chemistry, BioSkrbyb Genomics has integrated the ability to perform first strand cDNA synthesis of the entire transcriptome (Figure 2). After single cell isolation in a standard microtiter plate, the BioSkrbyb Resolve RT chemistry is applied to each cell to perform cytosolic lysis and reverse transcription of the mRNA from each individual cell (Figure 2A). The synthesized cDNA molecule of each transcript remains in the sample during the nuclear lysis and subsequent steps to amplify the genomes of each cell. Similar to the ResolveDNA product, the genome of each cell is then denatured in preparation for random priming based genome amplification (Figure 2B). PTA utilizes isothermal amplification and proprietary termination chemistry to restrict amplicon size, preferentially redirecting random primers to the primary template (Figure 2B). This is a critical feature of the ResolveOME chemistry system, which minimizes recopying of not only synthesized genome amplicons, but the amplicons generated from the initial template switch-based first strand cDNA synthesis.

The result of the combined processes is a pool of amplified genome which also contains the complete diversity of cDNA in a cell. These combined pools are then separated using a proprietary magnetic bead based reagent system to isolate the cDNAs, and remove any contaminating amplified genomic DNA from the transcriptome pool. Having created these two distinct fractions, the genomic DNA and the transcriptomic cDNA, the samples are then converted to sequencing libraries using well defined processes for transcriptome (including pre-amplification) and genome libraries using the optimized ResolveOME library preparation system.

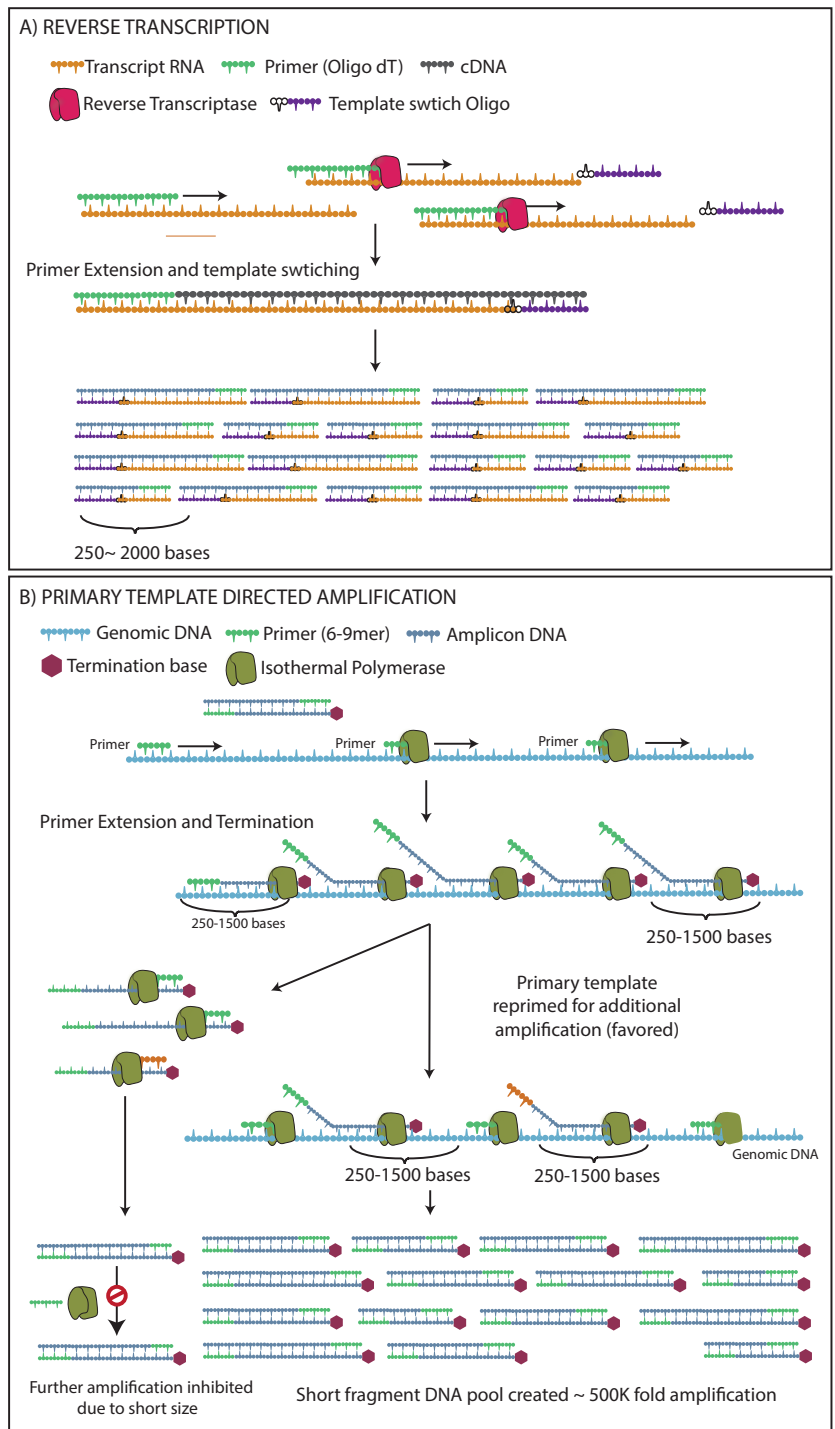


Figure 2. ResolveOME™ Amplification process. Isolated single cells first undergo 1st strand cDNA synthesis of cytosolic mRNA molecules existing within each cell using a reverse transcription process (A). When this process is complete primary template amplification of the nuclear fraction is conducted (B). First strand cDNA products are then enriched from the pool of amplified genomic DNA, prior to library preparation using the BioSkrbyb ResolveOME library preparation system, next generation sequencing, and analysis with BaseJumper software.

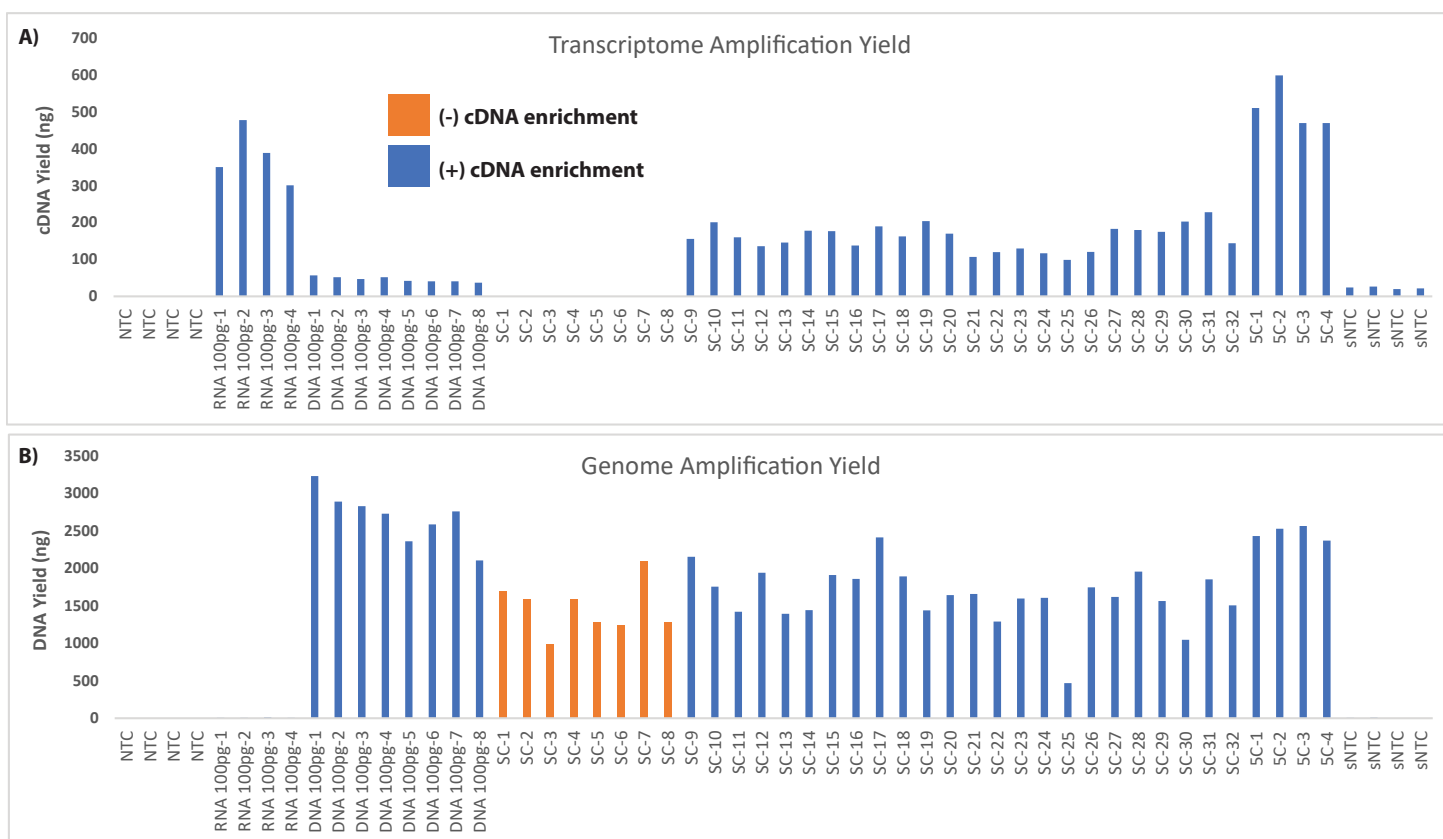


Figure 3. Amplification yield of Transcriptome and Genome from individual GM12878 cells. To assess ResolveOME chemistry performance, samples are first quantified after combined transcriptome and genome amplification. In (A), the yield of pre-amplified transcriptome cDNA is assessed with both RNA and DNA as well as single cells, and wells that contain 5 cells. No template controls were deposited during the chemistry run (NTC) while the FACS sorting negative controls (sNTC) were deposited in parallel with cell sorting. The yield of amplified genomic DNA (B) is assessed from the same cells. A comparison was made to determine the effect of cDNA enrichment on the output of either the transcriptome cDNA or amplified genomic DNA yield (orange bars). We found the use of the enrichment process was critical for achieving acceptable transcriptome cDNA yield, while differences in amplified Genomic DNA were not observed.

This system in its entirety allows the analysis of full length mRNA transcript-derived cDNA molecules as well as the ability to analyze the entire genome of each cell. While WGS analysis may be a desirable downstream next generation sequencing application, the genomic fraction can be alternatively enriched for WES or used for targeted panel enrichment^{12,13}. The use of various enrichment strategies is the same process that has been employed for the integration of the ResolveDNA WGA with various commercial genome enrichment products. Protocols for these application workflows can be found at BioSkrby.com.

Experimental Methods and Results

To demonstrate the efficiency of amplification we first analyzed DNA yield (Figure 3) and fragment sizes (data not shown) of the resultant amplification process. To determine performance of the combined genome/transcriptome workflow we first utilized the 1000 Genomes cell line GM12878. While this cell line has fewer active mRNA transcripts, the use of this highly defined genome allows the assessment of allelic balance, precision and sensitivity of single nucleotide variation (SNV). We found that the ResolveOME process allows the creation of separate and distinct pools of transcriptome cDNA and amplified genomic DNA (Figure 3) that allowed downstream library creation for next generation sequencing. We found the average single cell

yield of transcriptome cDNA was ~150 ng after preamplification, while the same single cells typically generated ~1500 ng of amplified genomic DNA. We noted some degree of RT activity of purified DNA (Figure 3A). In these reactions (DNA 100 pg), DNA is available during the RT step, unlike the nuclear DNA of single cells where nuclear lysis occurs later. The generation of amplified product is a result of the promiscuity of the reverse transcriptase, which can and will use DNA as template in the absence of template mRNA molecules. This is less prevalent in the RT process of single cells in ResolveOME, as the nuclear genomic DNA is not denatured prior to the RT step. An important feature of the chemistry is the limited amount of amplification of the transcriptome cDNA from bulk RNA during PTA (Figure 3B). This is a result of the mechanism of PTA, which has limited recopying of templates that are smaller in fragment size.

A key feature in the ResolveOME process is the bead based enrichment of the transcriptome-derived cDNA which occurs at the completion of the genomic amplification process. We found the enrichment process was critical to achieve acceptable yield for the transcriptome-derived cDNA (Figure 3A), while not appearing to cause a loss of amplified genomic DNA yield (Figure 3B). While some variability in cDNA or amplified genomic DNA yield is expected in the amplification of these two omic pools from a single cell, we noted high yield consistency,

with only one cell (Figure 3B - SC-25) that had lower yield suggesting a potential FACS drop out or user error. Overall, the ResolveOME system is able to produce pools of amplified cDNA and genomic DNA with sufficient quantity to prepare next generation sequencing libraries for downstream analysis. Using the unified workflow, the transcriptome cDNA and amplified genomic DNA products from the same single cells were used for library preparation to compare the performance of a range of cell types, the detection of expressed genes and, to calculate the allelic balance, precision and sensitivity of SNV using the reference GM12878 genome.

Transcriptome Sequencing Gene Detection

The ability to combine the transcriptome analysis with the genome analysis is dependent on high data quality for these two molecular layers. A key metric is the number of unique genes detected within a single cell. To create this assessment we used SALMON, a bioinformatic algorithm to compare genes detected between cell types using the ResolveOME chemistry. Using the GM12878 line as a basis, we compared the number of genes detected to several cell lines (Figure 4). We found the number of genes detected in GM12878 cells using the ResolveOME chemistry ranged from 2500 to approximately 4500. In alternative cell types, particularly those that are induced

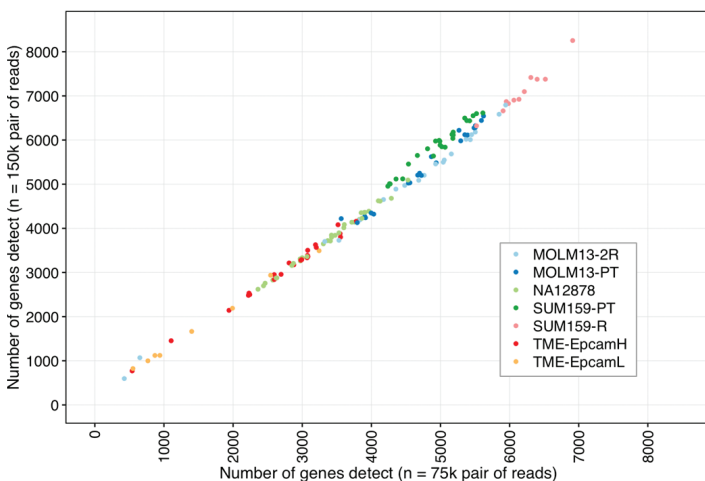


Figure 4. Detected Genes as Measured by SALMON. Unique gene detection by ResolveOME for a range of cell types and conditions was measured to determine sequencing performance.

to be drug resistant, the number of unique genes detected ranged from ~4000-7000. This was compared further to a clinical breast cancer sample that had been cryopreserved and remained uncultured and found the range of gene detection to be 500-4000 genes. Taken together, the data suggest unifying the transcriptome and genome does not require a sacrifice in data quality. Rather, the analysis of full length transcripts provides high sensitivity and dynamic range while allowing the delineation of splice variation and gene fusions within a specific transcript, even using fragile and precious clinical samples.

Genome Sequencing Precision and Sensitivity

The integration of the PTA with template switching based transcriptome cDNA preparation allows a plethora of downstream analysis options. Some of these downstream bioinformatics workflows may include combining copy number variation with a set of expressed genes which may be either deleted or amplified at the chromosome level. At a basic level the ability to define cell phenotype and state while delineating the structure of the genome is critical for various forms of pathology. Other applications may include the analysis of oncogenic SNV drivers of pathology combined with structurally modified mRNA molecules. The possibilities through these combined modalities are abundant. However, to conduct the synergized analysis, the data quality must be high.

The metrics indicating the highest quality data in genome sequencing relate to SNV sensitivity and precision. This is true for both the sequencing of large populations of cells (referred here as bulk) as well as in the genome of a single cell. Given that an individual cell typically contains one genome (diploid), about 6.6 pg of DNA, the amplification of this genome material must preserve the allelic balance. This is a measure of the uniformity or evenness of the amplification process. PTA has the highest allelic balance of any developed whole genome amplification chemistry¹¹. The development of the ResolveOME system has the same requirements. Measuring the allelic balance (Figure 5) of single cells is the best method to determine amplification bias. Using the same cells (Figure 3), libraries were created and QC'd prior to high depth sequencing. To ensure adequate genomic coverage and uniformity, each library was initially sequenced to a minimum of 2M total reads to establish its Preseq count. This is an estimation of library complexity that predicts the ability to call variants with high sensitivity and precision at the depths normally required for accurate analysis¹⁴. Only those libraries

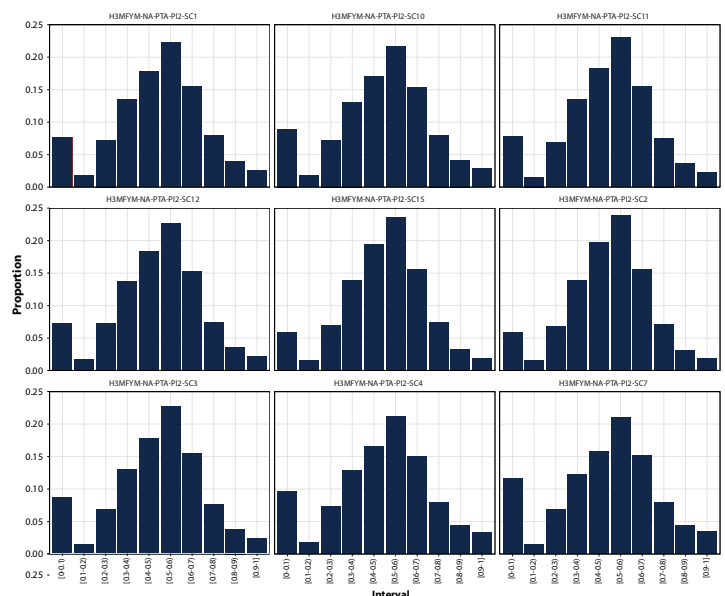


Figure 5. Allelic balance of heterozygous sites was measured in whole-genome reactions using alternate allele frequency sites with at least 2X coverage.

that achieved a Preseq count $>3.5 \text{ E9}$ were used for sequencing at full depth. WGS libraries were sequenced at 2 X 150 on NovaSeq 6000 S4 flowcells.

A key factor driving the precision and sensitivity in the detection of SNV within single cells and low input samples is the ability to maintain strict allelic balance of heterozygous sites within the genome. Results show that allelic balance was high in amplified genomes from single cell samples from ResolveOME (Figure 5) and preserved over all samples. The data demonstrates the ResolveOME WGA process integrated with ResolveOME library preparation system, maintains a true representation of the allele frequency from samples with a single intact genome.

Ultimately, the performance of whole-genome amplified libraries is determined by how faithfully they represent the genome being amplified. Reference alleles for GM12878 were therefore evaluated in PTA libraries using the BaseSpace Variant Calling Assessment Tool with the Genome in a Bottle Consortium's v3.3.2 truth set.

Importantly, variant detection amongst heterozygous alleles showed limited allelic bias (Figure 5), and called single nucleotide variants with high precision and sensitivity (Figure 6). To interpret sensitivity and precision, ResolveOME single-cell samples analyzed by WGS were compared to 1, 3, and 5 GM12878 cells using the ResolveDNA workflow, as well as to 50 and 100 ng of purified genomic GM12878 DNA. ResolveOME single-cell WGS samples overall generated allele detection sensitivities typically greater than 91% (with the exception of one drop out cell - intentionally included which had a low preseq value-Figure 6A) where the single cell had the lowest sensitivity of 91%. We found the sensitivity and precision of single cells processed with the ResolveOME system to be comparable to the same cells processed with the ResolveDNA and library preparation using Illumina DNA prep (genome only). Not surprisingly, samples which contained greater than one cell (3 & 5 cell) and those using 10 and 20 cell equivalents (50 & 100 pg) of genomic DNA had slightly higher sensitivity and precision. Single-cell libraries prepared using the ResolveOME amplification and library preparation system had equivalent performance as those processed with ResolveDNA and library preparation using the Illumina DNA prep kit.

Taken together, these data demonstrate the combination of uniform amplification and ligation-based library preparation allow for exceptional genome recovery and allelic balance, even when combined with the first strand mRNA transcriptome in the ResolveOME workflow. This enables highly sensitive and precise single nucleotide variation detection, at the resolution of a single genome within an individual cell.

The development of the ResolveOME chemistry allows a new class of single-cell analysis. This comprehensive view of the macromolecular dynamics allows further definition and delineation of the factors that influence clonal heterogeneity within a cellular population. This is both true in the process of normal development and the evolution of disease states. The chemistry system also maintains a significant amount of flexibility to enable the addition of new modalities, such as CITE-seq⁹ to probe the expression of translated surface marker proteins. Protein state can then be delineated using genome based mutation profiling with the BioSkryb BaseJumper analytics platform. BaseJumper uses curated databases (ClinVar, COSMIC) to enable the global annotation of a coding change to provide the inference on the structure and function of every single protein in the genome. This data can be further used to predict functional changes in protein activity and the risk profile of the modification. This is accomplished using well established algorithms (SIFT, PROVEAN, FATHOM, etc.) to determine if a functional change is likely, and whether or not the non-synonymous change is tolerated, neutral or deleterious to cell function. Taken together, ResolveOME enables the study of the multi-omic spectrum of individual cells.

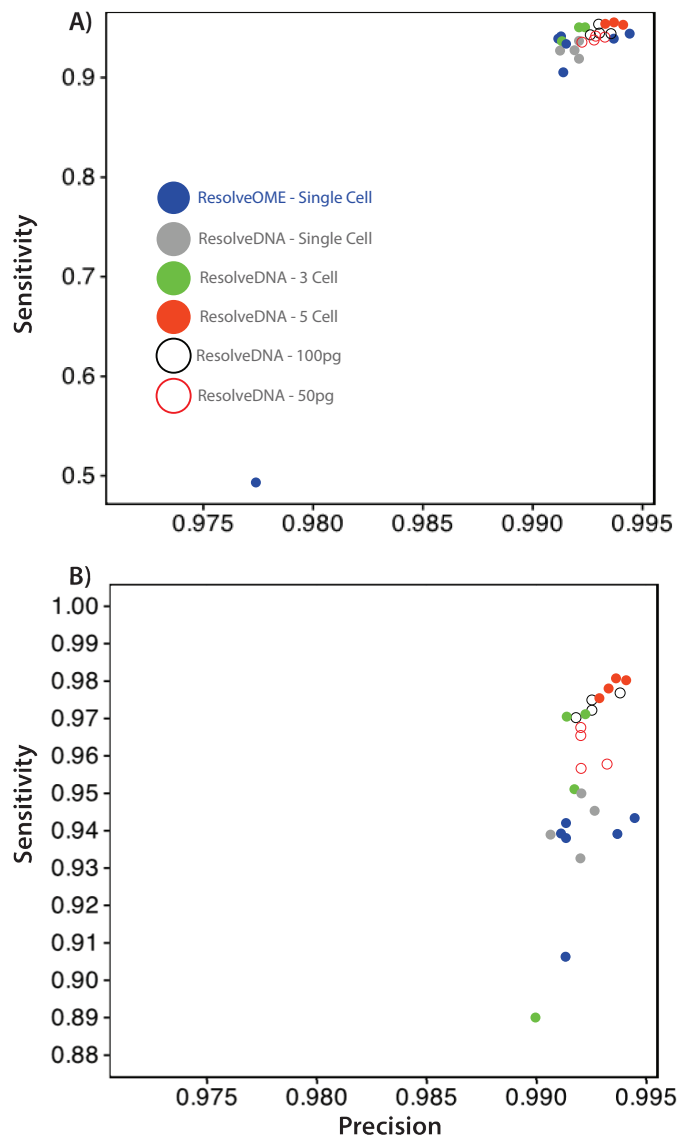


Figure 6. Precision and sensitivity of single nucleotide variation detection was measured in samples containing a single cell/ reaction. WGS data indicated precision and sensitivity exceed 99% and 91%, respectively, where 5/6 cells displayed sensitivity of exceeding 93%.

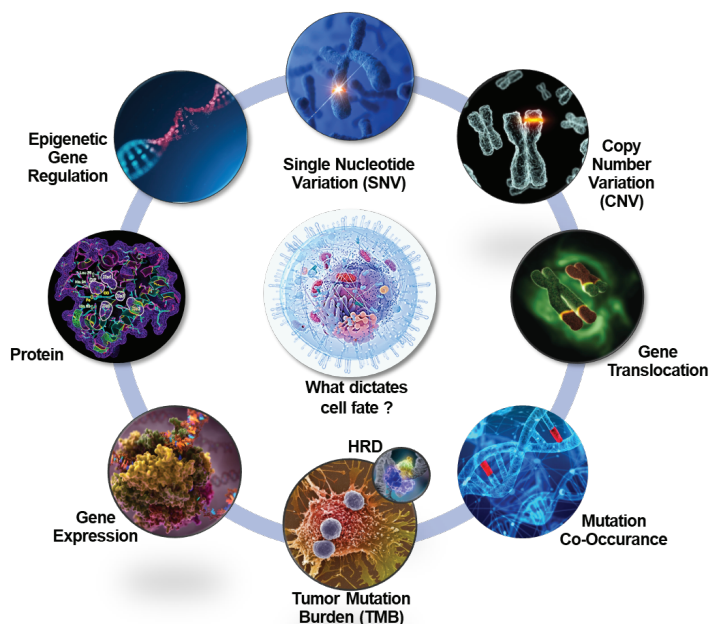


Figure 7. Molecular Factors Affecting Cell Fate. Many factors contribute to cell fate including SNV, CNV, translocation, mutation co-occurrence and mutation rate, as well as gene expression and regulation, protein levels and functional state.

SUMMARY

The creation of the ResolveOME system allows comprehensive analysis of the transcriptome and genome in parallel from the same cell. Providing high resolution accuracy of genome analysis down to the single base level combined with the comprehensive full length mRNA transcriptome enables the understanding of interplay of these omic layers within and between individual cells. Our current understanding, the multi-ome is limited by the ability to analyze the complete complement of these macromolecular layers in the same single cell. Conventional multi-omic methods are restricted to decoding a tiny portion of the genome, end-counting a fraction of an expressed gene, or a limited population of proteins. The ResolveOME system provides the highest resolution multi-omic view of an individual cell. The method enables a non-biased approach to detect the expressed full-length transcripts within a single cell and allow the analysis of the complete or targeted genome (i.e. exome) of the same cell. Taking advantage of this flexibility the user can select a subset of analyses vectors, such as expression, copy number variation, and single nucleotide variation to allow novel discovery of new biomarkers and potentially therapeutic targets. The ResolveOME system enables a newfound level of multi-omic breadth, which will enable discovery over a wide set of application areas, including oncology, neurology, immunology, reproductive medicine, toxicology, and cardiac research.

References:

1. Yu, K.H. and M. Snyder, Omics Profiling in Precision Oncology. *Mol Cell Proteomics*, 2016. 15(8): p. 2525-36.
2. Liu, J., et al., Applications of Single-Cell Omics in Tumor Immunology. *Front Immunol*, 2021. 12: p. 697412.
3. Chen, M., et al., Comparison of multiple displacement amplification (MDA) and multiple annealing and looping-based amplification cycles (MALBAC) in single-cell sequencing. *PLoS One*, 2014. 9(12): p. e114520.
4. Chen, C., et al., Single-cell whole-genome analyses by Linear Amplification via Transposon Insertion (LIANTI). *Science*, 2017. 356(6334): p. 189-194.
5. Dean, F.B., et al., Comprehensive human genome amplification using multiple displacement amplification. *Proc Natl Acad Sci U S A*, 2002. 99(8): p. 5261-6.
6. Pollen, A.A., et al., Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. *Nat Biotechnol*, 2014. 32(10): p. 1053-8.
7. Nowakowski, T.J., et al., Spatiotemporal gene expression trajectories reveal developmental hierarchies of the human cortex. *Science*, 2017. 358(6368): p. 1318-1323.
8. Swanson, E., et al., Simultaneous trimodal single-cell measurement of transcripts, epitopes, and chromatin accessibility using TEA-seq. *Elife*, 2021. 10.
9. Stoeckius, M., et al., Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods*, 2017. 14(9): p. 865-868.
10. Macaulay, I.C., et al., Separation and parallel sequencing of the genomes and transcriptomes of single cells using G&T-seq. *Nat Protoc*, 2016. 11(11): p. 2081-103.
11. Gonzalez-Pena, V., et al., Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc Natl Acad Sci U S A*, 2021. 118(24).
12. Zawistowski, J., et al., ResolveDNA Integration with Illumina DNA Prep and DNA Prep with Enrichment to Enable Single-cell Genomics. 2022, BioSkrby Genomics. p. 1-5.
13. Zawistowski, J., et al., Unprecedented Whole Exome Coverage Uniformity using ResolveDNA WGA and Twist Human Core Exome Panel, in www.bioskrby.com, B. Genomics, Editor. 2022.
14. Daley, T. and A.D. Smith, Modeling genome coverage in single-cell sequencing. *Bioinformatics*, 2014. 30(22): p. 3159-65.

For more information or technical assistance:
info@bioskrby.com

Published by:

BioSkrby
GENOMICS

2810 Meridian Parkway, Suite 110
Durham, NC 27713
www.bioskrby.com

All data on file.

For Research Use Only. Not for use in diagnostic procedures.
BIOSKRYB, RESOLVEDNA and RESOLVEOME are trademarks of BioSkrby, Inc.
All other product names and trademarks are the property of their respective owners.

TAS-007, 02/2022