A high-plexity exome solution tailored for sensitive variant detection in single cells

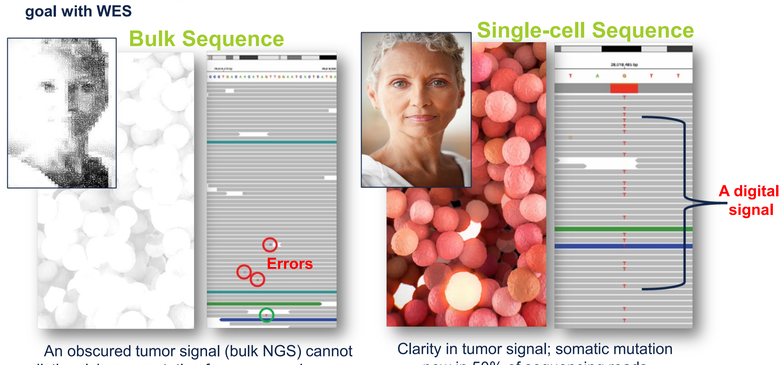
Victor J. Weigman, Durga M. Arvapalli, Laurel Coons, Tia A. Tate, Tatiana V. Morozova, Jesse Ramirez, Subidhya Shrestha, Jon S. Zawistowski

BioSkryb Genomics, Durham, NC, USA

Introduction

The ability to obtain transcriptomic, genomic, and other omic insights at the resolution level of individual cells has revolutionized biological discovery. This exposure of ever-increasing degrees of heterogeneity, masked when performing bulk sequencing, has provided new understanding of gene expression subpopulations within a sample, aided reconstruction of clonal evolution, and guided mechanistic hypotheses for drug-resistant cells. Despite this, adoption of single-cell genomic sequencing has lagged behind single-cell transcriptional analysis, largely because of the expense of whole genome sequencing (WGS) that increases with number of single cells sequenced. In addition, for researchers studying cancer and other diseases, sequence variation found in non-coding and regulatory genomic regions may be of less immediate interest compared to more interpretable or actionable variation found within coding sequences. To address these needs, we present a workflow that couples BioSkryb's ResolveDNA®,[1] Whole Genome Single-Cell Core Kit and ResolveOME™,[2] Whole Genome and Transcriptome Single-Cell Core Kit to third party hybridization capture procedures for downstream whole exome sequencing (WES) of single cells. Highlights of the BioSkryb single-cell exome enrichment workflow include:

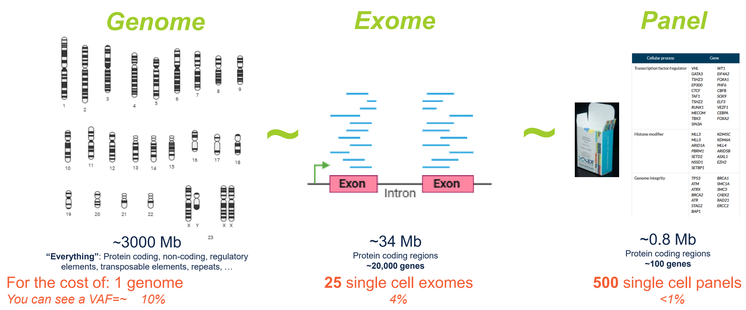
- **Optimized to existing Resolve core chemistries upstream of enrichment**
- **Near-complete and uniform single-cell whole genome amplification creates high quality genomic material for hybrid capture, ensuring optimal target capture (<2.5% of targets with 0 coverage) and reliable variant calling (average sensitivity >90%, average positive predictive value >98%).**
- **96 single-cell libraries in one enrichment reaction streamlines the workflow and saves on enrichment reagent cost**
- **An alternative to single-cell WGS that offers the ability to "right-size" an experiment to the biological goal with WES**



An obscured tumor signal (bulk NGS) cannot distinguish rare mutation from sequencing errors

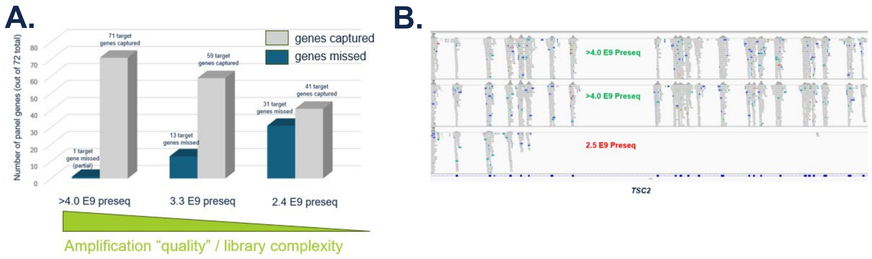
Clarity in tumor signal; somatic mutation now in 50% of sequencing reads

The digital variant-calling signal of the single-cell and its exome relevance. When exposing a single nucleotide variant with low allele frequency, the read structure of single cell is digital, whereby a heterozygous 0/1 reference/alternate allele is present in ~50% of the reads (right). Similarly, a homozygous change (1/1) will be present in all reads. By contrast, a bulk exome (left) will not be able to differentiate a rare variant (green) from a sequencing or amplification error within the averaged background of the sample (red). This affords confidence in variant detection even at a sequencing depth at 10X *or lower* and contrasts from what bulk exome experimentalists are accustomed to with 50-100X read depths.



Genome	Exome	Panel
~3000 Mb	~34 Mb	~0.8 Mb
"**Everything**": Protein coding, non-coding, regulatory elements, transposable elements, repeats, …	Protein coding regions **~20,000 genes**	Protein coding regions **~100 genes**
For the cost of: 1 genome	**25** single cell exomes	**500** single cell panels
You can see a VAF=~ 10%	*4%*	*<1%*

Genomic content "right-sizing" for biological goal. Whole genome sequencing provides the highest level of insight into genomic features beyond protein coding sequence, including regulatory elements and repeats. However, dependent on the experimental or biological question, protein coding regions exposed by exome hybrid capture (WES) or more focused hotspot panels may be preferred for increased confidence in prediction of variant effects mutations that could be therapeutically targeted. For the same all-in cost (sample->sequence) of a bulk whole genome, many more single cells can be processed, providing a richer biological story. The VAF is based on an average WGS of a10B NovaSeq X flow cell, where 25 genomes =~ 625 or =~ 12,500 panels



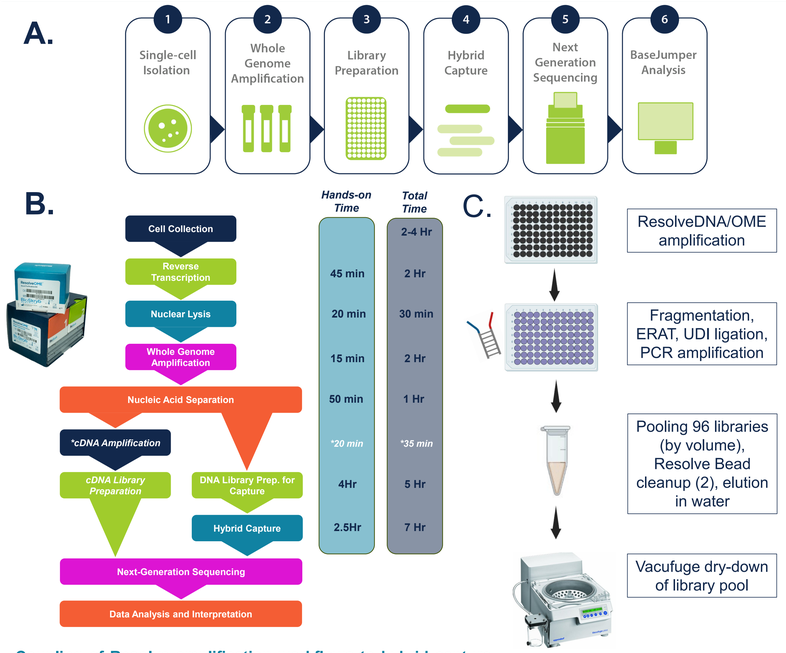
Quality in = quality out: faithful amplification of single-cell genomes is required for robust hybrid capture.

There have been limited instances of exome studies, and released commercial protocols, that are single-cell focused. Borgstrom et al.[3] performed a comparative study of single-cell genomic amplification methodologies (Ampli1, MALBAC, REPLI-g, PicoPLEX) each coupled to downstream Illumina TruSeq exome capture. The study demonstrated that the incomplete and uneven allelic representation inherent in the amplification methods tested impacted hybrid capture and resulted in target region coverages ranging from 7 to 68% for single cell exomes. This is in contrast to 90% target recovery observed with bulk exome controls . The relatively low target coverage translated to, at best, 25% of single nucleotide variants (SNVs) observed in the bulk controls being detected in the single-cell exomes. Whole genome amplification performed with ResolveDNA or ResolveOME Single-Cell Core Kits achieves 97% genome recovery and >85% allelic balance, which provides completely differentiated pre-capture material from the amplification methods utilized in the Borgstrom et al.[3] study

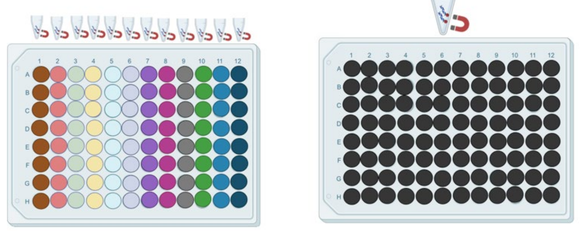
To further demonstrate the quality-in = quality-out concept, we have modeled-in poor genome recovery by the inclusion of subpar amplification samples (Preseq[4] count <3.5E9) in the same enrichment plex with high quality amplification samples. The manifestation of poor library complexity on each capture metric is shown with subpar single cells depicted with orange icons in each plot (see Selected (A), (B),

We have also demonstrated the direct consequence of single-cell amplification quality and subsequent library complexity by performing hybrid capture (72 gene panel) downstream of ResolveDNA amplification. The number of genes missed in the panel correspondingly increased as a function of decreasing library complexity (A) of different single cells. An example of this can be seen with an IGV read pileup view of a gene in the panel (*TSC2*) of single cells harboring robust library complexity (green) vs poor library complexity (red): the majority of exonic coverage is missing for *TSC2* in the cell with poor genome amplification (B).

Methods



Coupling of Resolve amplification workflows to hybrid capture. Following genomic amplification of single cells, BioSkryb UDI ligation-based library preparation for hybrid capture enzymatically fragments the amplification product to an optimal size for exome enrichment. For ResolveDNA (A), libraries are pooled, bead-purified, and subjected to hybrid capture after vacufuge dry-down. For the ResolveOME multi-omic workflow (B), the genomic (DNA) fraction similarly undergoes library prep for capture (C), pooling and exome enrichment, while the transcriptome (RNA) fraction libraries are prepared as per the default workflow and sequenced separately from the exome captured libraries. BaseJumper® software powers pre-capture QC (if desired) and post-capture analysis of genomic amplification libraries, in addition to transcriptional analysis of ResolveOME RNA libraries.



96 Resolve reactions
12 X 8 capture

96 Resolve reactions
1 X 96 capture

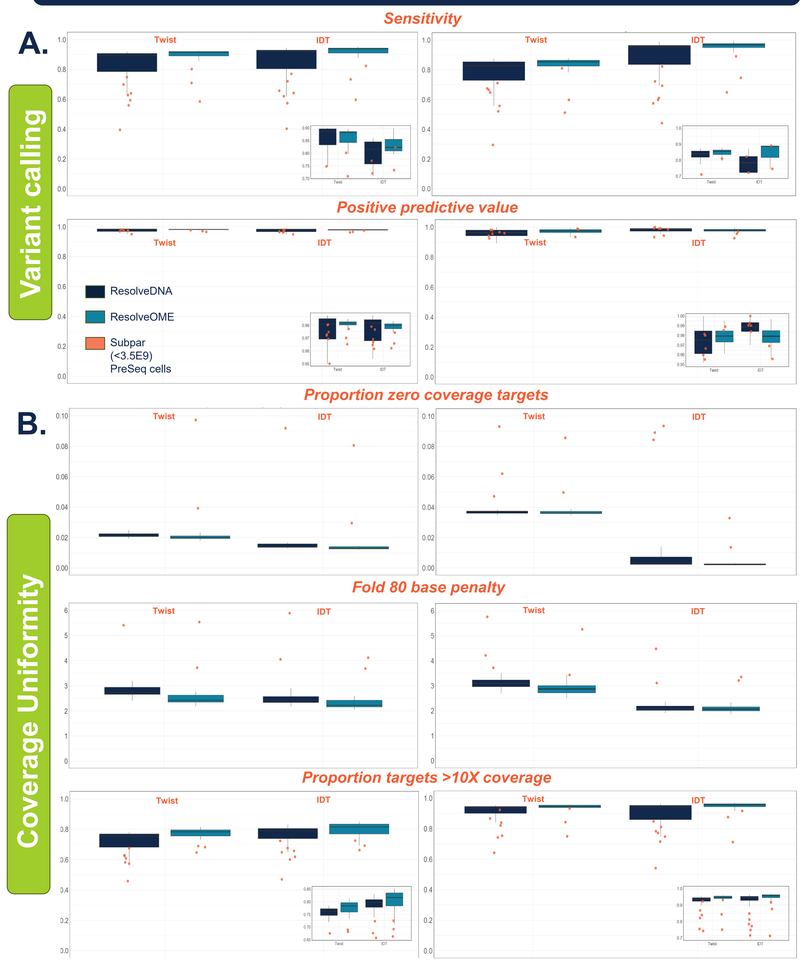
Pushing the bounds of capture plexity while retaining variant calling performance. In a prototypical bulk exome enrichment experiment, libraries are subjected to hybrid capture at a plexity range of 6-12. For the single-cell researcher, this plexity range is not ideal as one scales the number of single cells desired to be screened for coding sequence variation. Instead of requiring 8-16 separate enrichment reactions, including their associated washes, we have created a simplified workflow to enrich 96 single-cell libraries at once.

Table 1: Capture parameter and sequencing recommendations as a function of plexity

Vendor	Panel	Vendor Protocol Version Used	Hybridization Time (Hours)	Number of Post-Capture Amplification Cycles	Plexity Used for Hybrid Capture	Total Pooled Library Input	Sequencing Parameters (Illumina)
Twist Bioscience	Exome 2.0 + Comp. spike-in	Twist Target Enrichment Fast Hybridization Protocol (version: DOC-001066 REV 4.0)	4	8	96	8 µg (2x vendor recommended)	20M paired-end reads (10M clusters) per library X 96 libraries = **1920M paired-end reads (960M clusters) per pool**
					48	8 µg (2x vendor recommended)	20M paired-end reads (10M clusters) per library X 48 libraries = **960M paired-end reads (480M clusters) per pool**
IDT™	xGen™ Exome Hyb Panel v2*	xGen Hybridization Capture of DNA Libraries Protocol "Tube Protocol" (version:8)	4	6	96	12 µg (2x vendor recommended)	20M paired-end reads (10M clusters) per library X 96 libraries = **1920M paired-end reads (960M clusters) per pool**
					48	6 µg (1x vendor recommended)	20M paired-end reads (10M clusters) per library X 48 libraries = **960M paired-end reads (480M clusters) per pool**
Twist Bioscience	Alliance CNTG Hereditary Oncology Panel (0.2 MB)	Twist Target Enrichment Fast Hybridization Protocol (version: DOC-001066 REV 4.0)	4	13	96	8 µg (2x vendor recommended)	1M paired-end reads (500K clusters) per library X 96 libraries = **96M paired-end reads (48M clusters) per pool**
					48	8 µg (2x vendor recommended)	1M paired-end reads (500K clusters) per library X 48 libraries = **48M paired-end reads (24M clusters) per pool**
IDT™	xGen™ Pan-Cancer Hybridization Panel (0.8 MB)	xGen Hybridization Capture of DNA Libraries Protocol "Tube Protocol" (version:8)	16	10	96	12 µg (2x vendor recommended)	1M paired-end reads (500K clusters) per library X 96 libraries = **96M paired-end reads (48M clusters) per pool**
					48	12 µg (2x vendor recommended)	1M paired-end reads (500K clusters) per library X 48 libraries = **48M paired-end reads (24M clusters) per pool**

For two biotinylated exome panels, Twist Bioscience Exome 2.0 + Comprehensive Spike-In or IDT xGen v2, optimal pooled library input, hybridization time, and PCR cycling is recommended above as a function of plexity. Enrichment parameters are additionally included for the indicated small oncology panels. Sequencing depth recommendations are listed for the given plexity. The user guides for enrichment were followed according to the manufacturers' protocols with the exception of the modifications listed in the table above.

Exome Results *Panel*



Selected post-capture performance metrics. HG002 single cells that were amplified by either ResolveDNA (dark blue) or ResolveOME multi-omic amplification (light blue), were subjected to 96-plex hybridization capture with either Twist Exome 2.0 + Comprehensive Spike-In (left) or IDT xGen v2 exome (right) using the conditions listed in Table 1. All single-cell libraries were down sampled to 20M total 150 bp paired end reads / 10 M single reads. Sensitivity and positive predictive value are shown as variant calling metrics in (A) while capture performance and uniformity are presented in (B). On-target analysis is quantified below for each 96-plex, whereby near-target padding is defined as 250 bp (C). Table 2 highlights additional post-capture metrics, whereby subpar PreSeq samples were not included in the mean + SD calculations.

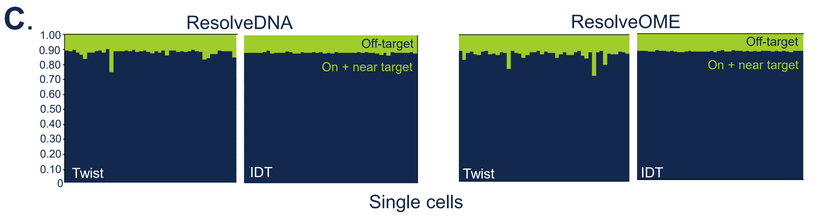


Table 2: Additional post-capture metrics of ResolveDNA/OME single cell libraries

Metric	Resolve Platform	Mean + SD: Twist	Mean + SD: IDT
Mean target coverage	ResolveDNA	21.42 ± 1.38	19.1 ± 1.14
	ResolveOME	21.48 ± 1.1	18.17 ± 3.36
Allelic balance	ResolveDNA	0.73 ± 0.2	0.73 ± 0.2
	ResolveOME	0.81 ± 0.06	0.8 ± 0.1
Proportion duplication	ResolveDNA	0.05 ± 0.02	0.13 ± 0.03
	ResolveOME	0.05 ± 0.01	0.15 ± 0.05
GC dropout	ResolveDNA	16.06 ± 1.24	9.86 ± 0.9
	ResolveOME	13.46 ± 0.62	7.58 ± 0.85
AT dropout	ResolveDNA	0.35 ± 0.06	0.35 ± 0.19
	ResolveOME	0.48 ± 0.05	0.8 ± 0.24

Conclusions

- A single-cell exome workflow coupling Resolve single cell amplification to hybrid capture was developed to ensure robust variant calling and to provide an alternative to WGS-level resolution
- PCR duplication rate associated with increasing plexity was mitigated while achieving ~20X mean target coverage at a recommended sequencing depth of 20M total paired end reads / library
- Plexity maximization (96) streamlines workflow from bulk-centric protocols and allows more cells to be assessed for biological applications focused on protein coding regions

References

1. Gonzalez-Pena V *et al.* Accurate genomic variant detection in single cells with primary template-directed amplification. *Proc Natl Acad Sci U S A* 118, e2024176118 (2021).
2. Marks JR *et al.* Unifying comprehensive genomics and transcriptomics in individual cells to illuminate oncogenic and drug resistance mechanisms. bioRxiv (2023) DOI 10.1101/2022.04.29.489440
3. Borgstrom E *et al.* Comparison of whole genome amplification techniques for human single cell exome sequencing. PLOS One 2017, 12(2):e0171566. DOI 10.1371/journal.pone.0171566
4. Daley T, Smith AD. Predicting the molecular complexity of sequencing libraries. Nature Methods 2013,10:4, 325-327. DOI 10.1038/nmeth.2375