Integration of the ResolveOME™ multi-omics workflow with IDT xGen™ hybridization capture at single-cell resolution to obtain high quality exome data

BioSkryb GENOMICS

AGBT™

Durga M. Arvapalli, Isai Salas-González, Benjamin Warlick, Viren Amin, Swetha D. Velivela, Katie Kennedy, Jamie Remington, Jeff Blackinton, Victor J. Weigman, Gary L. Harton, Jay A.A. West, Jon S. Zawistowski
BioSkryb Genomics, Durham, NC, USA

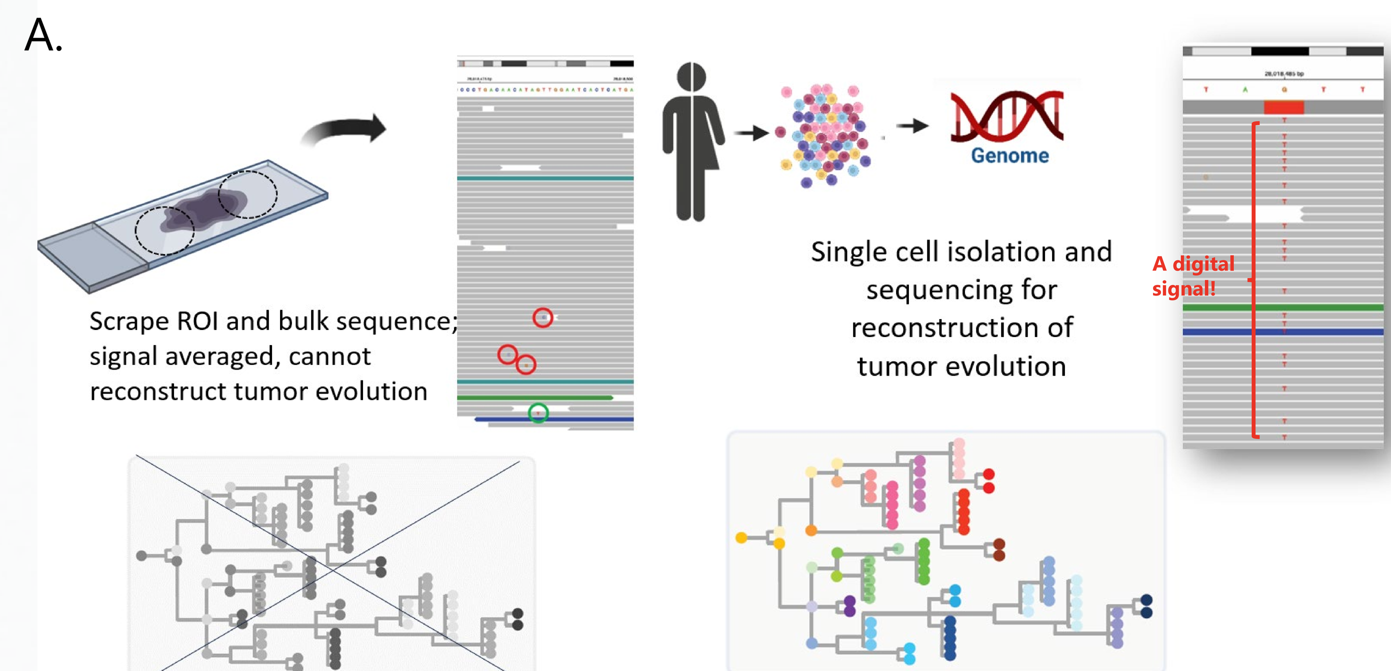
ABSTRACT

Simultaneous examination of multiple -omic layers in single cells offers essential insights to decode the complexity of cancer and understand the pathophysiology of oncogenesis as well as tumor progression. ResolveOME™ amplification chemistry provides these insights through unified whole-genome and full-length transcriptome information from the same single cell. While whole genome sequencing (WGS) provides the complete complement of genomic elements for discovery, whole exome sequencing (WES) offers a cost-effective alternative while still providing the ability to ascertain the entirety of coding sequence.

Here we extend ResolveOME™ to a robust hybrid capture workflow to examine whole exomes. The core ResolveOME™ workflow is unification of template-switching single-cell RNAseq chemistry with ResolveDNA® whole genome amplification (WGA) technology, whereby the cDNA synthesized is separated from the genomic amplification fraction through affinity purification. The ResolveDNA® Library Preparation kit with enzymatic fragmentation was used to create libraries from the amplified genomic fraction. Prior to performing hybrid capture, low pass sequencing ensured high library complexity as assessed by the preseq algorithm. The libraries were pooled and enriched using the IDT xGen™ Exome Research Panel v2. The pool was 2X150 sequenced on an Illumina NextSeq 1000 instrument. The samples were downsampled to 40 million reads per library where analysis using BaseJumper™ software showed target base coverage of 98% at 1X and 96% at 10X. The single-cell SNV calling metrics showed excellent precision and sensitivity of 99% and 92% respectively with robust allelic balance across single cells and on target bases over 70%.

We demonstrated the successful coupling of the ResolveOME™ multi-omics workflow with IDT xGen™ Research Panel v2 hybridization capture of genomic libraries, providing the researcher a WES option in addition to WGS for maximizing cost and efficiency, with the key attributes of robust exomic metrics.

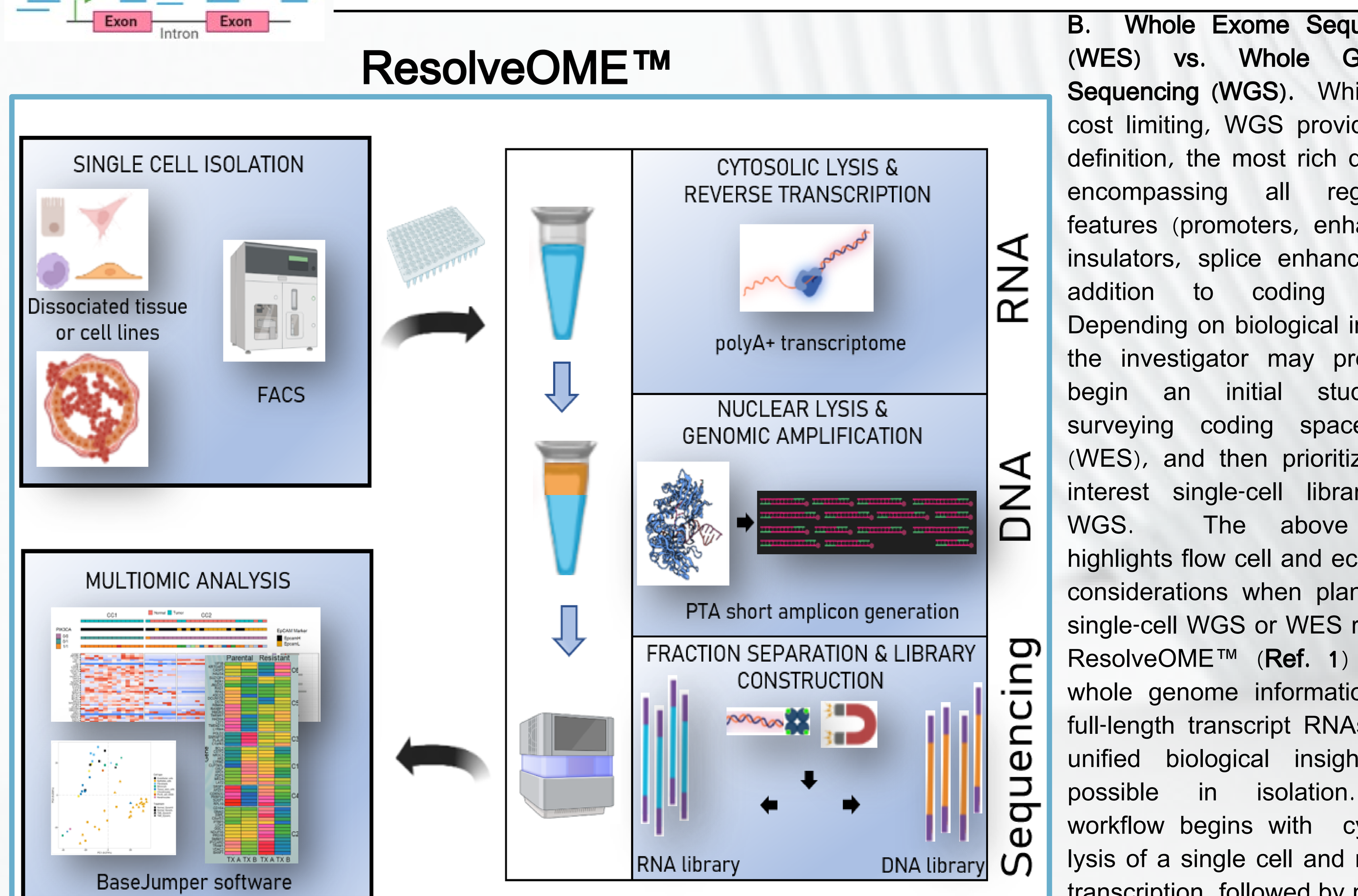
INTRODUCTION AND BACKGROUND



A. Bulk sequencing cannot resolve low-frequency molecular lesions driving tumor evolution. Classically, tissue regions of interest have been captured by physical scraping or by laser capture microdissection followed by genomic DNA isolation, library construction and sequencing. This signal-averaged approach is unable to distinguish sequencing errors from a rare somatic allele (left). By contrast, single cell isolation from biopsies or automated cell picking approaches provides the power to resolve rare variants; offering a digital and unambiguous signa (right).

B. WES / WGS (Exon, Intron, Exon)

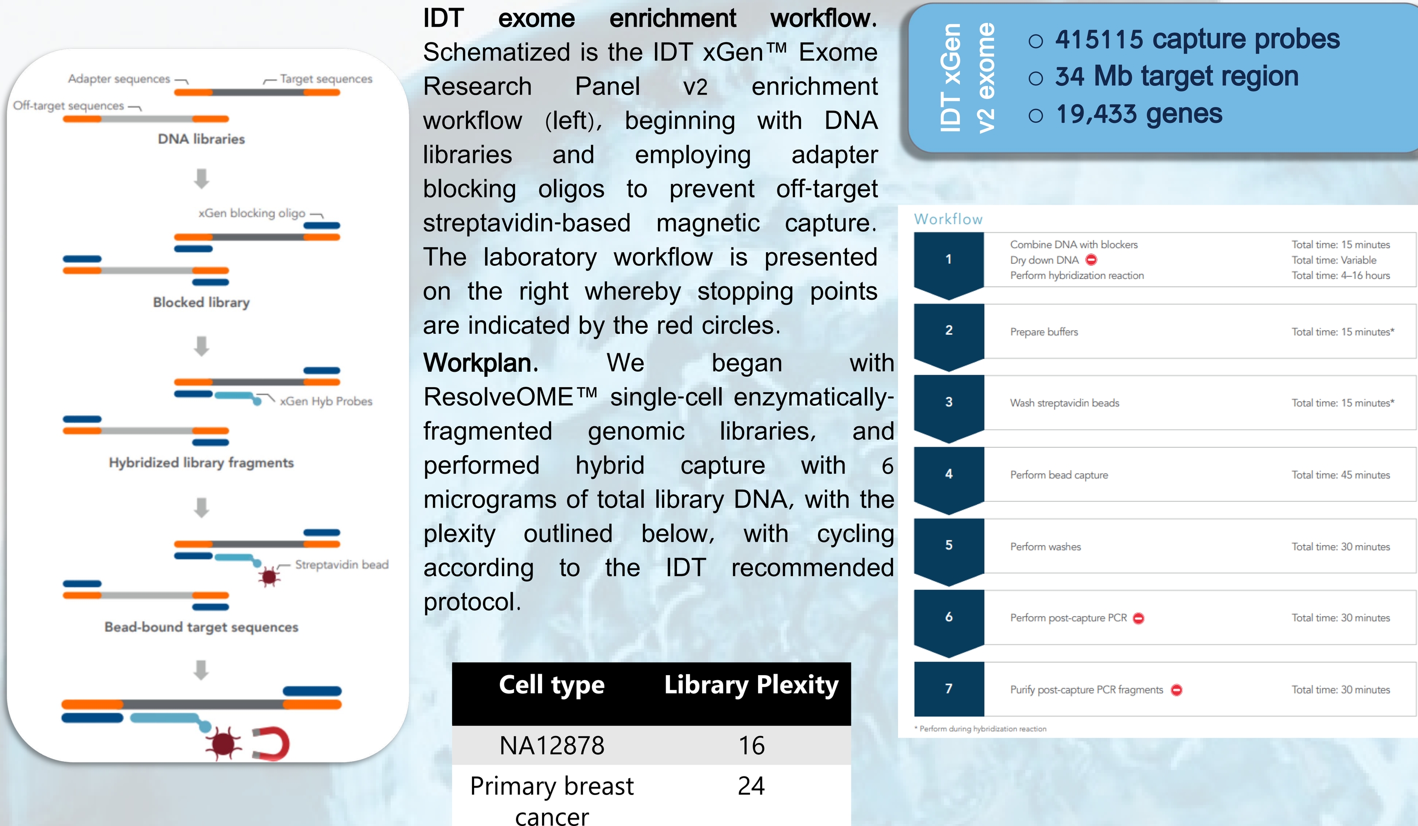
	Genomic features captured	Functional elements captured	Megabases	Coverage required	Total reads (M) required per cell	Flow cell cost per cell (NextSeq 1000)	Flow cell cost per cell (NovaSeq 6000)	Number of cells per NextSeq 1000 high output flow cell	Number of cells per NovaSeq S4 flow cell
WES	exonic	protein coding	34	50X	20	~$90	~$15	40	1000
WGS	exonic intronic intergenic	protein coding + regulatory	3200	10-30X	150-450	~$1800-3600	~$250-800	1-2	40



B. Whole Exome Sequencing (WES) vs. Whole Genome Sequencing (WGS). While still cost limiting, WGS provides, by definition, the most rich dataset, encompassing all regulatory features (promoters, enhancers, insulators, splice enhancers) in addition to coding exons. Depending on biological interest, the investigator may prefer to begin an initial study by surveying coding space only (WES), and then prioritize high interest single-cell libraries to WGS. The above chart highlights flow cell and economic considerations when planning a single-cell WGS or WES run. C. ResolveOME™ (**Ref. 1**) unifies whole genome information with full-length transcript RNAseq for unified biological insights not possible in isolation. The workflow begins with cytosolic lysis of a single cell and reverse transcription, followed by nuclear lysis and primary template-directed amplification (**PTA, Ref. 2**) of the genome in the same reaction tube. PTA generates short amplicons that limit reamplification; driving the primers back to the template of interest and resulting in uniform, near-complete genomic coverage with high allelic balance—allowing single nucleotide and structural variation to be called with confidence. The reverse transcription products are affinity purified away from the genomic amplification products and two distinct sets of libraries, RNA and DNA, are created.

Goal: Provide a robust exome enrichment option to ResolveOME™ unified single-cell genomics and transcriptomics
Scope: Benchmark performance in NA12878 cells and extend workflow to primary breast cancer cells

WORKFLOW

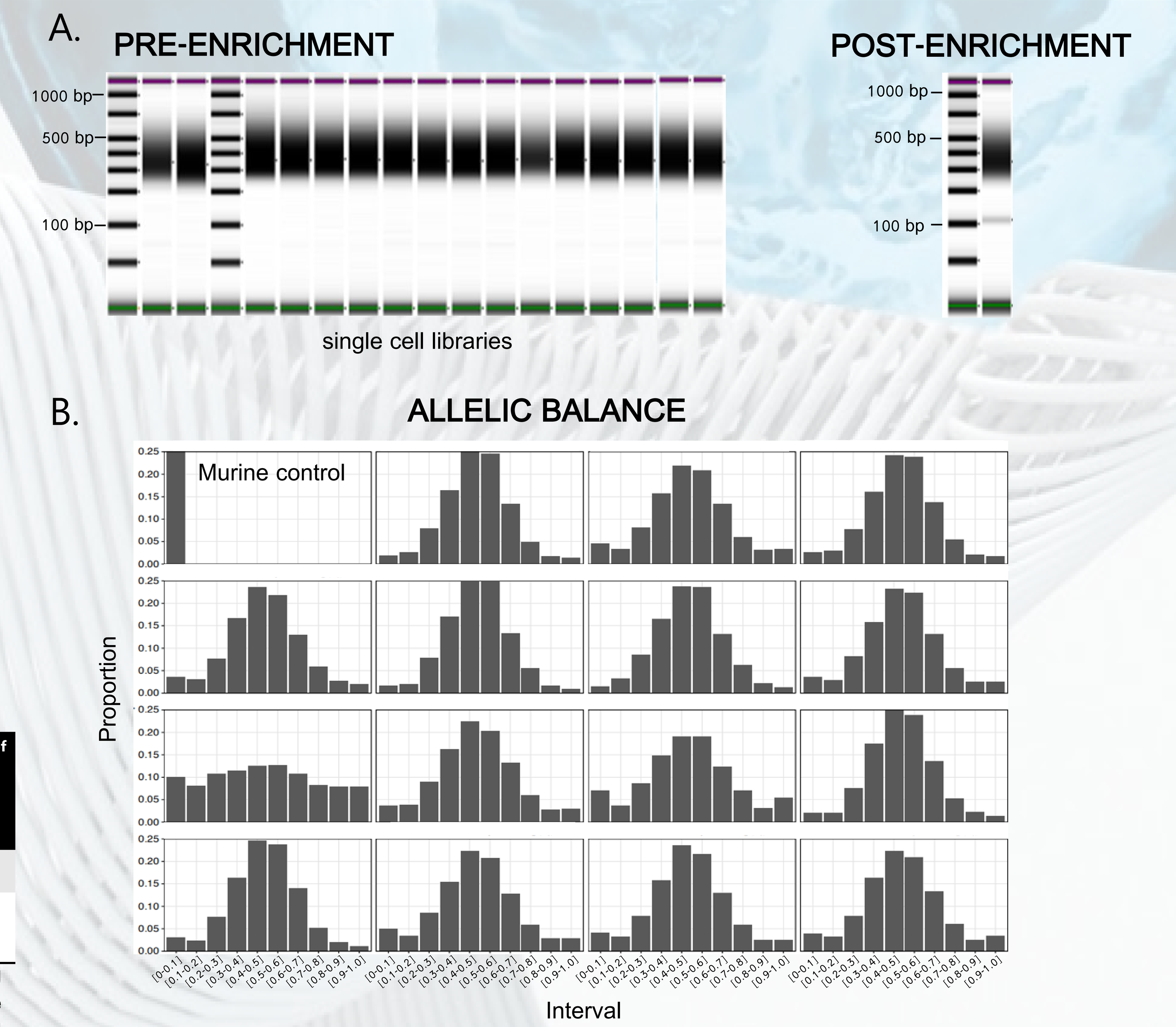


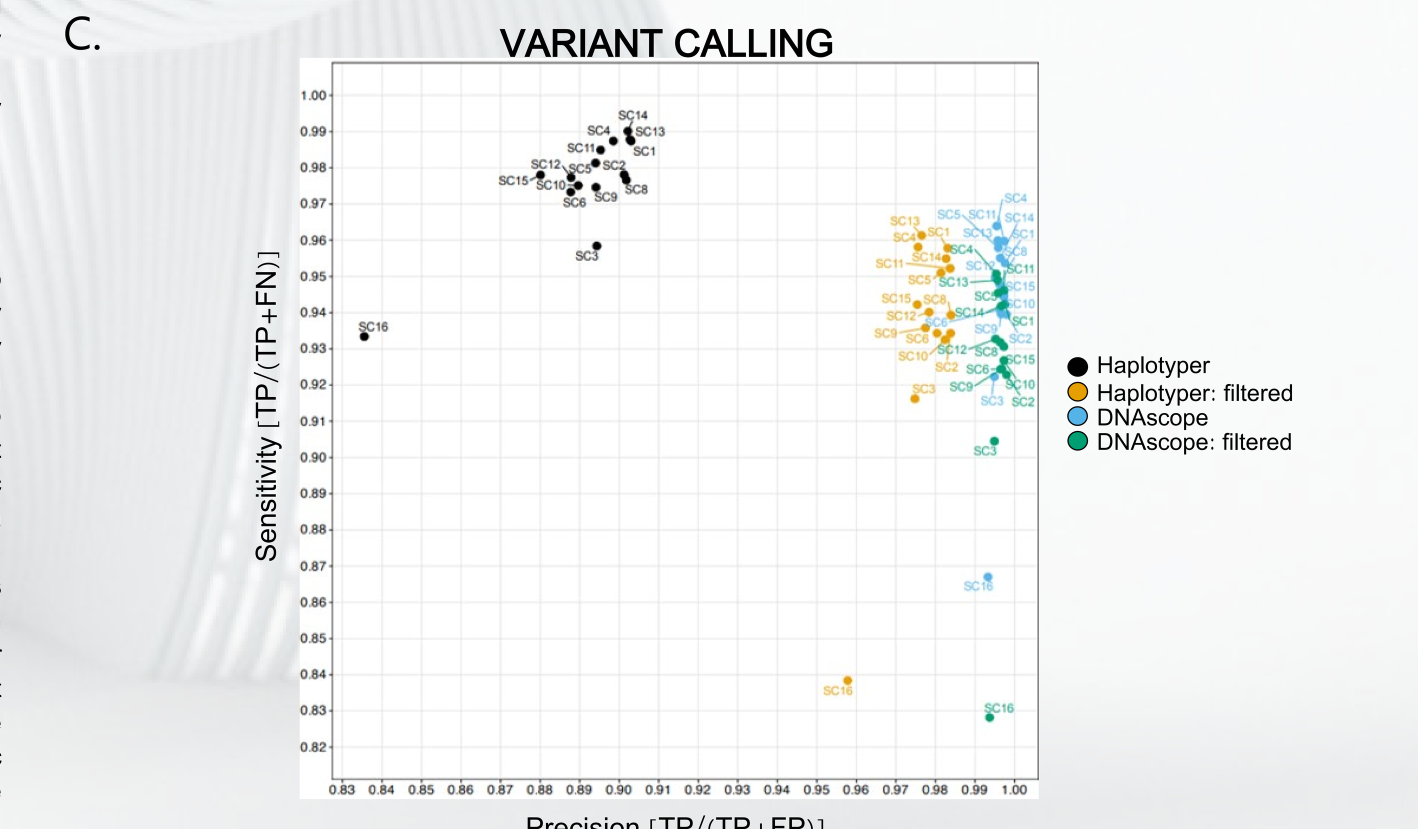
IDT exome enrichment workflow. Schematized is the IDT xGen™ Exome Research Panel v2 enrichment workflow (left), beginning with DNA libraries and employing adapter blocking oligos to prevent off-target streptavidin-based magnetic capture. The laboratory workflow is presented on the right whereby stopping points are indicated by the red circles.

Workplan. We began with ResolveOME™ single-cell enzymatically-fragmented genomic libraries, and performed hybrid capture with 6 micrograms of total library DNA, with the plexity outlined below, with cycling according to the IDT recommended protocol.

Cell type	Library Plexity
NA12878	16
Primary breast cancer	24

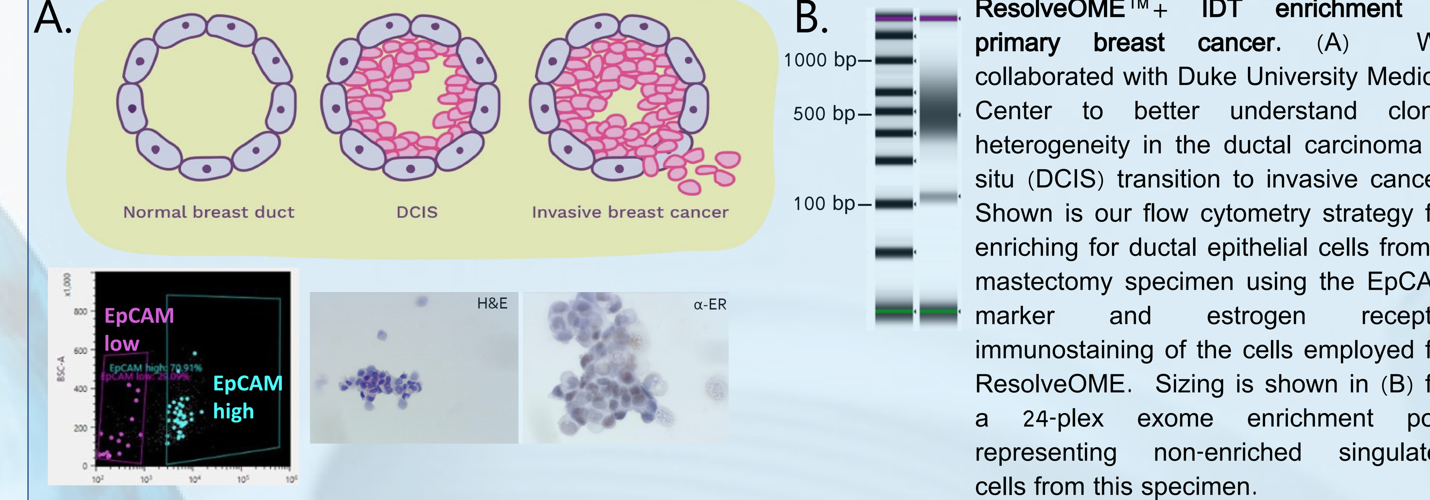
RESULTS



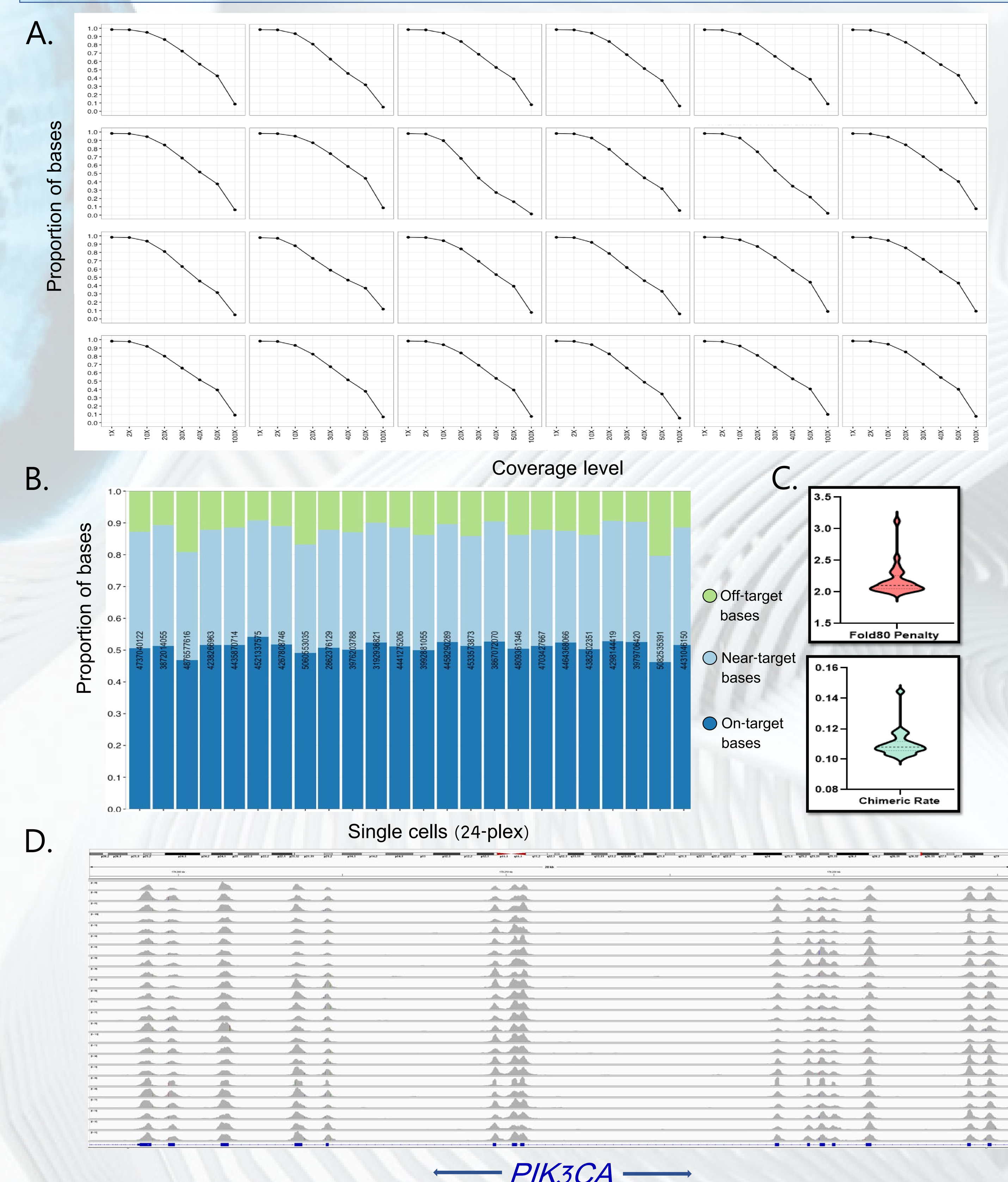


Performance benchmarking with NA12878 B lymphocytes. (A). HS D1000 Agilent TapeStation gel images showing size distribution of adapted, fragmented pre-enrichment libraries (left) and the size distribution of the exome-enriched 16-plex pool—with a mean size of 375-400 bp. The enriched pool was 2X150 sequenced on Illumina NextSeq 1000 and total reads downsampled to 40M. Allelic balance for each single-cell library within the pool was first assessed (B), with histograms of allelic frequency showing uniform representation of both alleles for 14/15 single cells plotted above. Murine gDNA run through the ResolveOME™ + IDT hybrid capture workflow served as a negative control (upper left). Precision and sensitivity of single nucleotide variant calling was assessed using two Sentieon variant callers, Haplotyper and DNAscope, in the presence or absence of filtering: filtering corresponded to high quality variants established using variant allele depth, with homozygous variants requiring at least 80% of read depth support, while for heterozygous sites, only variants with allele depth between 20% and 80% were considered. Both callers revealed high sensitivity (>92%) and high precision (>98%). DNAscope calling yielded overall higher precision than calls made by Haplotyper.

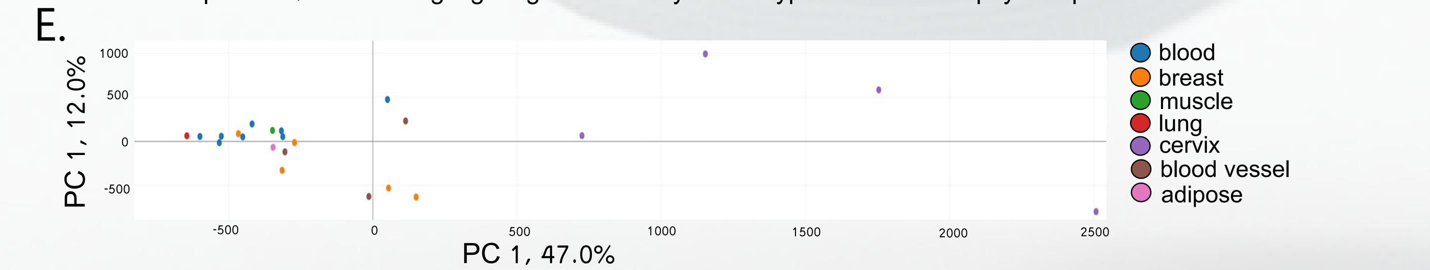
RESULTS



ResolveOME™+ IDT enrichment in primary breast cancer. (A) We collaborated with Duke University Medical Center to better understand clonal heterogeneity in the ductal carcinoma in situ (DCIS) transition to invasive cancer. Shown is our flow cytometry strategy for enriching for ductal epithelial cells from a mastectomy specimen using the EpCAM marker and estrogen receptor immunostaining of the cells employed for ResolveOME. Sizing is shown in (B) for a 24-plex exome enrichment pool representing non-enriched singulated cells from this specimen.



IDT xGen™ Exome Research Panel v2 performance in the ResolveOME™ workflow with individual primary breast cancer cells. (A) Fraction of targeted bases covered at a range of coverage depths for each individual cell of a 24-plex exome enrichment. (B) Proportion and absolute number of on-/off-/near-target bases for each individual single-cell library. (C) Mean Fold80 penalty (top) as an assessment of enrichment uniformity and mean chimeric rate (bottom). (D) Integrative genome viewer tracks visualizing exon alignment and lack of intronic read background for the *PIK3CA* gene. (E) Corresponding transcriptional signatures for the single cell exomes. Shown is principal component analysis from the BaseJumper™ software platform of the breast cancer single cell transcriptomes and tissue type predictions based on these signatures—highlighting the expected presence of a breast tissue signature for multiple cells, but also highlighting the diversity of cell type within the biopsy sample.



Summary and Conclusions

- ResolveOME™ amplification exposes critical interrelationships between the genome and transcriptome in individual cells. The genomics can be accurately ascertained at the exome level in addition to whole-genome.
- The attributes of high allelic balance, sensitivity and precision of single nucleotide variant calling provided by primary template-directed genomic amplification are crucial for downstream high-quality exome enrichment.
- IDT hybrid capture with ResolveOME™ genomic libraries affords robust single nucleotide variant calling (>92% sensitivity, >99% precision), uniform capture (mean Fold80 Penalty of 2.1), and >90% of target bases at 10Xcoverage—all in conjunction with the transcriptome from the same cell.

Acknowledgments

We thank Shelley Hwang and Jeffrey Marks of Duke University Medical Center for collaboration on Ductal Carcinoma In Situ: IRB: PRO00034242 "Biologic Characterization of the Breast Cancer Tumor Microenvironment". We thank the patients involved in this study for the gracious gift of tissue to fuel discovery.

References

1) Zawistowski, J.S. *et al.* Unifying genomics and transcriptomics in single cells with ResolveOME amplification chemistry to illuminate oncogenic and drug resistance mechanisms. bioRxiv preprint (2022) doi: https://doi.org/10.1101/2022.04.29.489440
2) 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).

Cells explored. Answers revealed.