

Technical Note

Single Cell Exome

Cells explored. Answers revealed.

Authors

Jon Zawistowski, PhD

Senior Director
Research & Development
BioSkryb Genomics

Swetha Velivela, PhD

Scientist I
Research & Development
BioSkryb Genomics

Isai Salas-Gonzalez, PhD

Computational Biologist
Bioinformatics
BioSkryb Genomics

Jay A.A. West, PhD

President, CEO, Co-founder
BioSkryb Genomics

Siyuan Chen, PhD

Chief Technology Officer
Twist Bioscience

Unprecedented Whole Exome Coverage Uniformity using ResolveDNA WGA and Twist Human Core Exome Panel

Primary Template-directed Amplification (PTA) is a novel and accurate single-cell Whole Genome Amplification (WGA) method for the genomic analysis of single cells and ultra-low input DNA samples. When paired with Twist Human Core Exome panel, ResolveDNA PTA Whole Exome Sequencing method generates high quality data with exome target coverage above 97%.

Introduction

Despite significant advances in genomics analysis methods, hundreds of nanograms to microgram quantities of DNA are still needed for many Next Generation Sequencing (NGS) sample preparation workflows. When working with single cells or other limited-input samples, picogram quantities of genomic DNA have to be amplified through whole genome amplification (WGA) methods to provide the desired amount of input material. Challenges of single cell WGA include low genome coverage, poor coverage uniformity, allelic skewing and cell-to-cell variations. Primary Template-directed Amplification (PTA) is a novel, isothermal method used in ResolveDNA WGA which overcomes these challenges with superior data quality¹.

Combining the revolutionary amplification performance with whole exome enrichment methods enables the screening of all known coding genes within an individual cell. When used with whole exome panels for whole exome sequencing (WES), the high genome coverage and coverage uniformity of ResolveDNA translates to higher exome target coverage, better coverage uniformity, and unprecedented variant calling sensitivity and specificity. This enables researchers to discover new biomarkers while exploring exonic regions of interest with high confidence.

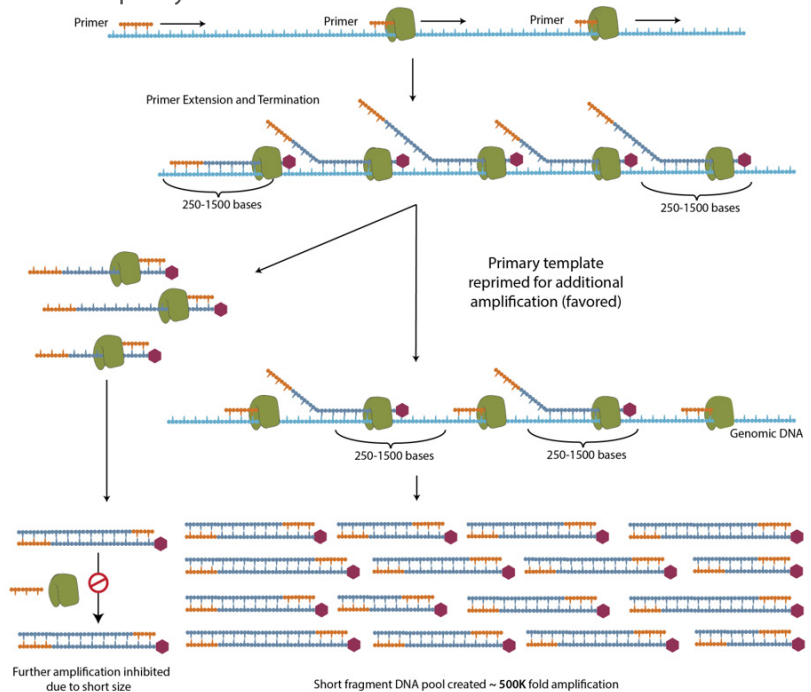


Figure 1. Schematic of PTA mechanism. ResolveDNA WGA is a random-primed, isothermal polymerase driven reaction that provides unbiased amplification by attenuating the size of daughter amplicons and consequentially re-directing primers to the primary template.

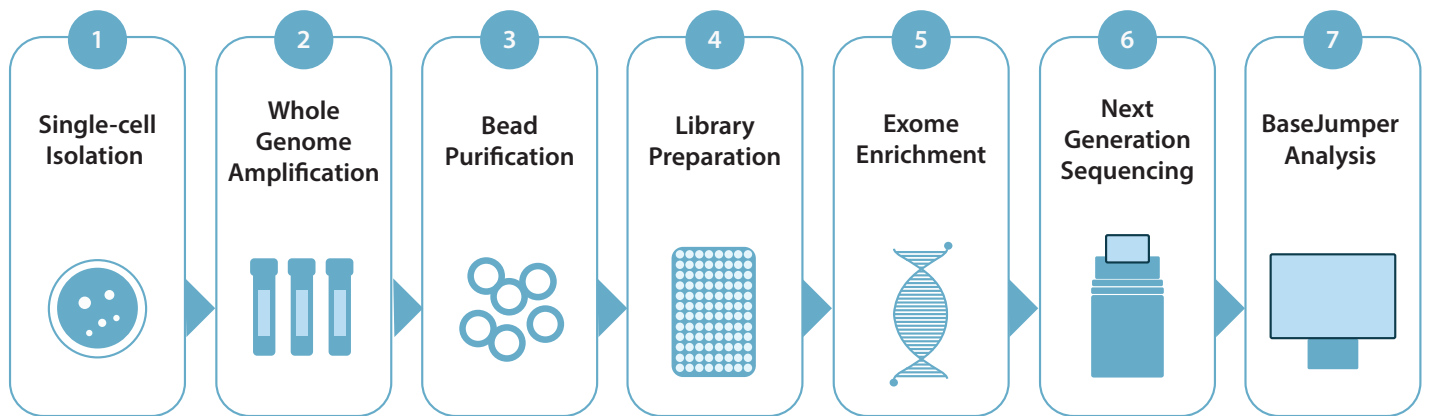


Figure 2. BioSkrby Genomics ResolveDNA paired with Twist Biosciences Exome enrichment Workflow: Isolated single-cells undergo a workflow that integrates Primary Template-directed Amplification with library preparation of purified amplification products, followed by exome enrichment, sequencing, and analysis with BaseJumper software.

Twist BioSciences Exome Enrichment

Twist Human Core Exome is an exome panel consisting of 190,213 non-overlapping regions covering 34Mb of Consensus Coding Sequencing (CCDS) or 99% of ClinVar variants. The regions were defined using the most current January 2017 freeze of the hg38 build. High fidelity double stranded 120 nt DNA probes were synthesized using Twist Bioscience's silicon-based technology. A quality control step assessing probe representation and uniformity is used to ensure uniform enrichment of target regions. The Twist Human Core Exome panel boasts a high on-target rate and coverage uniformity with more than 85% of bases exhibiting >30x coverage.

In this Tech Note, we present data from the single cell PTA-WES workflow using ResolveDNA WGA kits and Twist library preparation and whole exome enrichment kits (Figure 2). As both PTA and Twist Human Core Exome panel are known for high genome coverage, outstanding coverage uniformity and allelic balance, we saw high quality exome coverage, uniformity and allelic balance data from ResolveDNA Twist WES.

Materials and Methods

Primary Template-directed Amplification was conducted on single NA12878 cells collected and stained with Calcein AM and PI dye for live cell sorting. The sorted Calcein AM positive and PI negative single cells in 96 well LoBind plate were then processed with PTA using ResolveDNA WGA kit (PN100068 or PN100136)². PTA product was purified with ResolveDNA Bead Purification kit (PN100121 or PN100182). Purified PTA products were quantified with Qubit HS dsDNA Assay kit (ThermoScientific) and DNA size profiling was performed with D5000 HS Screenshotapes on TapeStation 4200 (Agilent Technologies).

Library Preparation and Target Enrichment were prepared in duplicates with Twist Library Preparation EF kit (PN100253) using 50 ng PTA products from six single NA12878 cells and one 100 pg gDNA control. A bulk control library was also prepared using 100 ng gDNA as input for each duplicate run. PTA products were fragmented for 4 minutes, much shorter

than the recommended 22 minutes, due to the small size of PTA amplicons (<10kb). As a control, the intact gDNA was fragmented for 22 minutes as suggested by Twist protocol. Target enrichment was performed using 100 ng libraries with Twist Human Core Exome kit (PN101217) following the manufacturer's instructions.

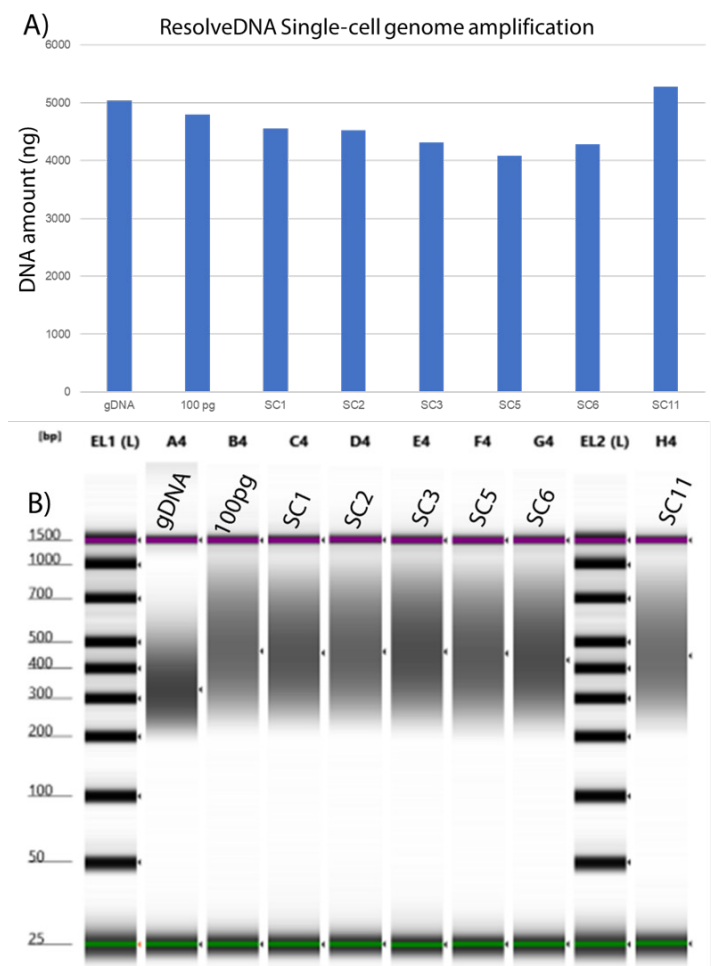


Figure 3. Single NA12878 cells generate robust amplified genomes. In (A) individual cells generated >1000 ng of amplified DNA from each cell. Libraries prepared (B) using the Twist Biosciences library preparation kit yield exome enrichment libraries with an average of ~450 bp.

Sequencing and Data Analysis was completed on libraries made from six single cell PTA products which were pooled and loaded on a MiniSeq at 2.2 pM. Approximately 5 million PE 75 bp reads were generated per library. Preseq count and ChrM metrics were called and compared to historical data for quality assessment. Enriched libraries were pooled and sequenced on NovaSeq using S4 flowcell. Approximately 40 million PE 150bp reads were generated per library. After demultiplexing and quality trimming raw reads were aligned to the hg38 reference genome using BWA-MEM. Duplicates, reads with mapq<20 and unmapped pairs were filtered out using SAMtools. Enrichment metrics were assessed using Picard. Variant calling was carried out using the GATK method (v4). All data was downsampled to 40 million reads per library.

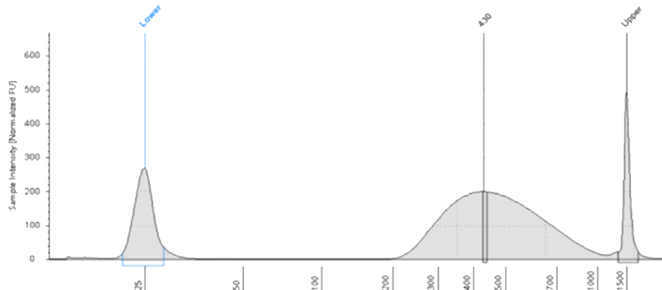
Results

Single Cell PTA Quality can be assessed by PTA yield, DNA size profiling and low pass sequencing metrics. Single cells used in this experiment generated higher than 1000 ng of DNA, indicating a successful amplification (Figure 3). The size of amplified DNA ranged from 100 bp to 10 kb and peaked between 800 bp to 1.5 kb, matching historical data for high quality PTA products.

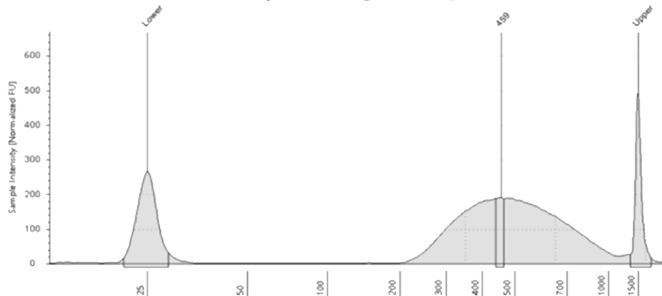
Libraries from these high quality PTA products ranged in average size from 408 bp to 466 bp, optimal for whole exome data analysis. These libraries were sequenced on a MiniSeq alongside bulk DNA control for ~5 million PE 75 bp reads per cell/sample. All twelve cells generated PreSeq count higher than 3.9 Gb and ChrM lower than 1%, demonstrating high data quality from the perspective of library complexity and cell lysis.

Twist WES enrichment demonstrated effective creation of pooled libraries. Two pools were created, both with an average insert size of ~250-300 bp (~430-450 bp including adapters), as seen by the TapeStation analysis (Figure 4). Enrichments yielded similar fragment size from the library preparation indicating little if any skewing or sample loss during the enrichment process. In addition, the enrichment yielded significantly greater product than required for a single sequencing run. This allowed for library preservation and pool storage for follow up analysis.

Twist enriched library - 1- Avg 430 bp



Twist enriched library - 2- Avg 459 bp



Twist target enriched library total yield (ng)

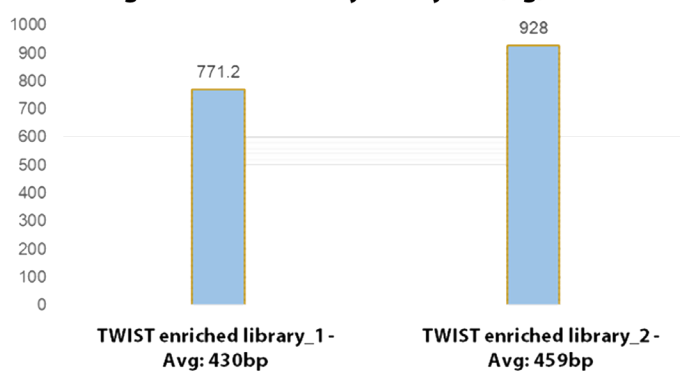


Figure 4. Twist Biosciences exome enriched single-cell amplified genomes. Enriched samples again demonstrated adequate fragmentation and insert size for downstream NGS analysis. Total insert size after purification and adaptation using the stubby adapters ranged from 430-460 bp with a yield of greater to 750ng total DNA.

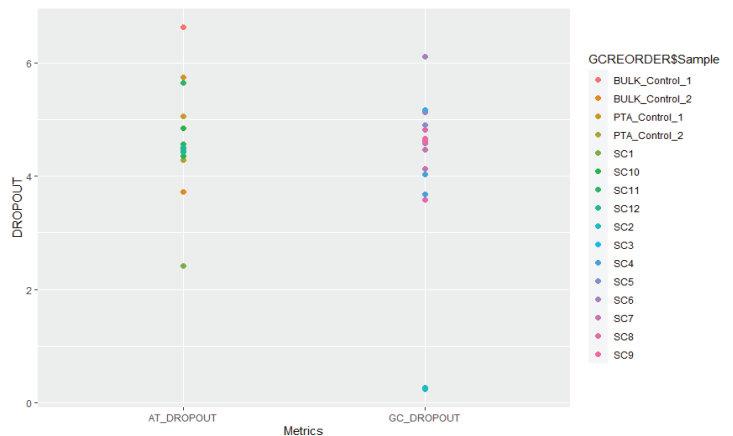


Figure 5. Single NA12878 cells demonstrate low GC bias: Analysis of both AT and GC regions of the genome demonstrate similar allelic dropout rates. We found the average dropout rate in single cells were comparable, 4.5% vs 4.7%, respectively, demonstrating GC rich and AT rich region are uniformly enriched with minimal GC bias.

Twist WES Coverage and Uniformity.

After 8-plex whole exome capturing, the captured libraries were sequenced on a NovaSeq for ~40 million PE 150bp reads. We first determined if the coverage variation was a function of the base content using the GC content to determine allelic dropout (Figure 5). The average GC_DROPOUT and AT_DROPOUT in single cells were comparable, 4.5% vs 4.7%, respectively, suggesting a both GC rich and AT rich region are uniformly enriched with minimal

GC bias. The average Fold 80 Base Penalty of the 12 single cells is 2.39, with a tight range between 2.29 and 2.76. Notably, scPTA-WES using Twist Human Core Exome panel achieved an average of 98% of the bases covered at 1x and 92% of the bases covered at 10x or higher (Figure 6B).

Another parameter indicating high quality data is the percent of bases that have greater than 30X coverage. In this experiment, in the majority of the single cells, typically 85% of bases had >30x coverage, exceeding most known published results using other WGA and whole exome capture methods.

SNV Calling Sensitivity and Specificity

To investigate whether scPTA-WES maintained the high single nucleotide variants (SNVs) calling sensitivity and specificity, scPTA-WES data were analyzed using BaseJumper SNV calling software. scPTA-WES achieved SNV calling sensitivity above 90% and specificity above 90%.

Conclusion

Single cell PTA-WES using ResolveDNA WGA and Twist Human Core Exome panel revealed high quality data featuring:

- Typical zero target coverage of < 2%
- Higher than 85% of bases with 30x coverage
- High single cell exome coverage uniformity demonstrated by Fold 80 base penalty and GC/AT dropout rate
- Exceptional single cell SNV calling sensitivity and specificity

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, 2021. 118(24).
2. BioSkryb Genomics, i., ResolveDNA™ Whole Genome Amplification Kit For high-quality single-cell and low-input DNA amplification, in www.bioskryb.com, B. Genomics, Editor. 2021, BioSkryb Genomics: Durham, NC. USA.

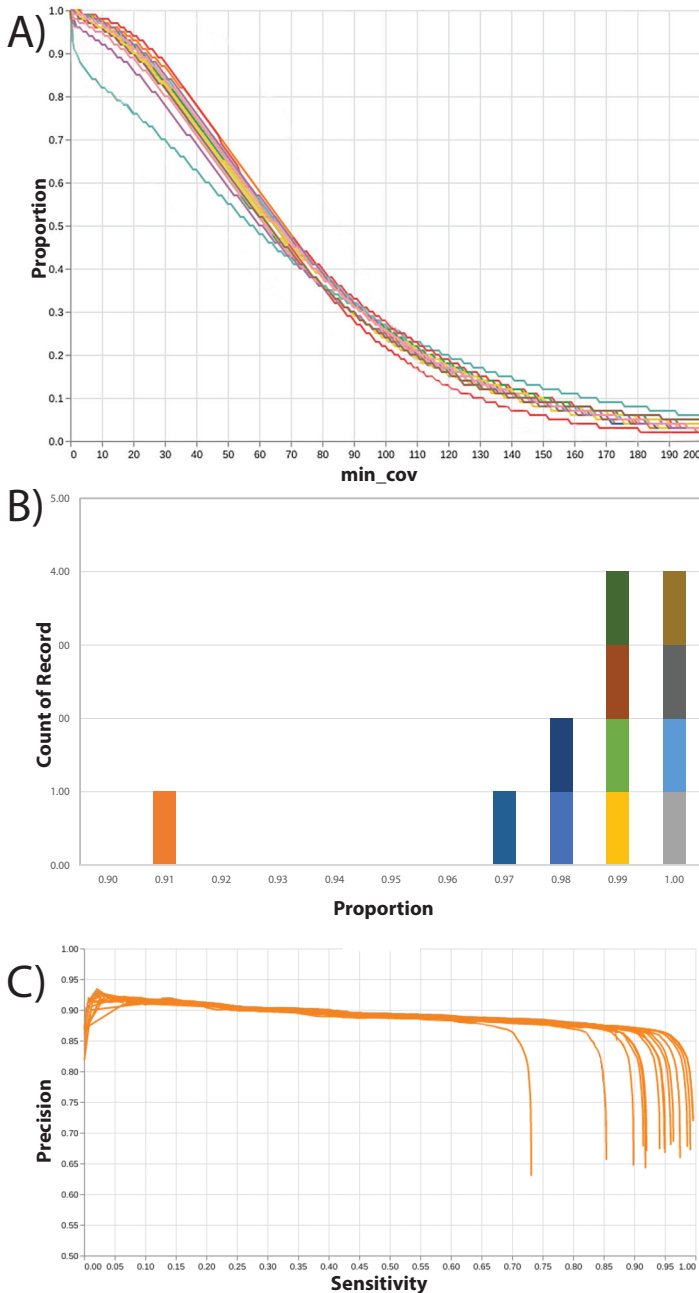


Figure 6. Coverage uniformity, sensitivity and specificity of BioSkryb Genomics ResolveDNA paired with Twist Biosciences Exome. In (A) coverage of libraries prepared from isolated single cells after ResolveDNA WGA, library preparation, exome enrichment, and sequencing demonstrated 98% exome coverage at 1 read/base (1X), greater than 93% at 10 reads/base (10X), and 85% at 30 reads/base (30X). Of the cells analyzed 11/12 had coverage (B) greater than 97% bases with 1 read/base (1X). This translated to 11/12 samples with SNV sensitivity between 82-95% at a precision of 86% (C).

Published by:

BioSkryb
GENOMICS

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

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

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