BioSkryb GENOMICS

Technical Note
Single-Cell Exome Sequencing

High-Plexity, Single-Cell Exome Hybrid Capture of ResolveDNA® and ResolveOME™ Amplification Products

The ability to obtain transcriptomic, genomic, and other omic insights at the resolution level of individual cells has revolutionized biological discovery. Examining genomes and transcriptomes at single-cell resolution exposes ever-increasing degrees of cellular heterogeneity that are masked when performing bulk sequencing. This has provided new understanding of gene expression subpopulations within a sample, aided reconstruction of clonal evolution, and guided mechanistic hypotheses for drug-resistant persister 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.

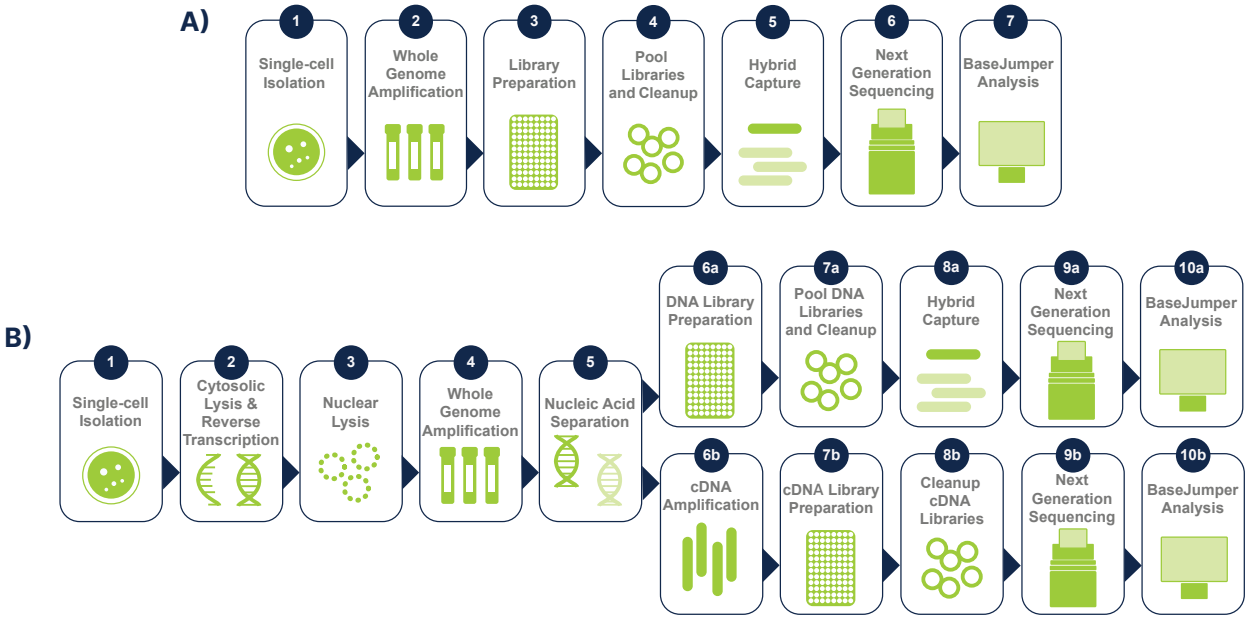

Figure 1: ResolveDNA (A) and ResolveOME (B) workflows coupled to downstream hybrid capture. Following genomic amplification of single cells, BioSkryb UDI ligation-based library preparation for hybrid capture enzymatically fragments the amplification product to an optimal size for exome enrichment. For ResolveDNA, libraries are pooled, bead-purified, and subjected to hybrid capture after vacufuge dry-down. For the ResolveOME multi-omic workflow, the genomic (DNA) fraction similarly undergoes library prep for capture, 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.

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 workflows that couple BioSkryb's ResolveDNA® Whole Genome Single-Cell Core Kit and ResolveOME™ Whole Genome and Transcriptome Single-Cell Core Kit to third party hybridization capture procedures for downstream whole exome sequencing (WES) of single cells (Figure 1). Exome enrichment panels used in this study were the xGen™ Exome Hyb Panel v2 from IDT™ and the Exome 2.0 plus comprehensive spike-in from Twist Bioscience. Highlights of the BioSkryb single-cell exome enrichment workflows include:

- **An alternative to single-cell WGS that offers the ability to "right-size" an experiment to the biological goal with WES.**
- **Works with existing ResolveDNA and ResolveOME Core Kits, including 96- and 384-well formats.**
- **Covers >97.5% of the single cell exome with variant calling sensitivity >90% and positive predictive value (PPV) >98%.**
- **96 single-cell libraries in one enrichment reaction streamlines the workflows and saves on enrichment reagent cost.**

Capturing Distinct Information: The Single-Cell vs. Bulk Exome

For the identification of rare genomic variation, typical measurements of cells in bulk obscure signals and mask the underlying heterogeneity present in a sample. By contrast, when sequencing at the level of the single cell, the signal from even rare variants is unambiguous and "digital", in that a heterozygous change is present in 50% of the reads (Figure 2). This means that the sensitivity of typical somatic exome sequencing can be achieved with multiple cells sequenced at low-depth (<10x) at an overall increased signal.

In addition, with single cell resolution in WES, cataloging co-occurrence of multiple genomic variants in the same individual cell is possible. While a typical bulk exome sequencing experiment can expose a variant with high allele frequency within a sample, defining variant co-occurrences to explore hypotheses about genetic interaction between variant A and variant B across tumor lineage is not possible with bulk sequencing.

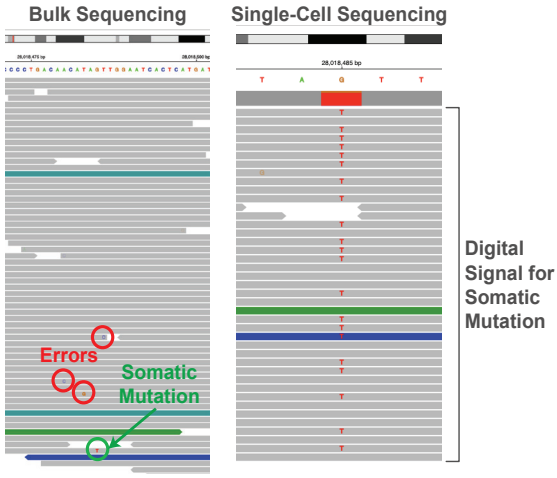

Figure 2: The digital variant-calling signal of the single-cell. When exposing a single nucleotide variant with low allele frequency, the read structure of the single cell is digital, whereby a heterozygous 0/1 reference/alternate allele is present in ~50% of the reads (right). Similarly, a homozygous 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).

Quality In = Quality Out: Faithful Single-Cell Genome Amplification Is Required for Successful Target Capture

There have been limited instances of exome studies, and released commercial protocols, that are single-cell focused. Borgström et al.[1] 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 exome data.

Whole genome amplification performed with ResolveDNA or ResolveOME Single-Cell Core Kits achieves 97% genome recovery and >85% allelic balance[2], ***which provides completely differentiated pre-capture material*** from the amplification methods utilized in the Borgström et al.[1] study. The more faithful genomic amplification provided by ResolveDNA and ResolveOME provides much more genomic "real estate" for capture by the biotinylated probes present in the exome panel. Herein, we demonstrate this quality in = quality out translation for single-cell exome enrichment, with low proportion of zero coverage targets (~0.02), ~75-85% of targets covered at 10X, and >90% median SNV calling sensitivity (Figure 4A, C-D). Table 1 presents the additional post-capture metrics of mean target coverage, allelic balance, GC/AT dropout, and duplication rate.

To best represent cell behavior that would be observed by our single-cell workflows, we show data across 96 cells and called out outliers (orange dots) in Figure 4. While careful cell dispensing and viability selection upstream of plating can reduce their presence, there will always be some outliers out of 96 cells with poor library diversity. Even so, meaningful information can be gained from these cells.

Pushing the Bounds of Capture Plexity while Retaining Variant Calling Performance

In a typical bulk exome enrichment experiment, libraries are subjected to hybrid capture at a plexity range of 6-12. A notable feature of our single-cell workflows are the ability to pool 96 libraries into a single hybrid capture (Figure 3). Through our workflows, the total quantity of library input for hybrid capture has been modified for the higher plexity to quell the increase in duplication rate and other adverse effects that typically accompany increasing plexity while holding hybrid capture panel quantity constant. Hybrid capture at 96-plex saves time and expense associated with multiple captures, without sacrificing performance as is demonstrated by the metrics reported in Figure 4 and Table 1.

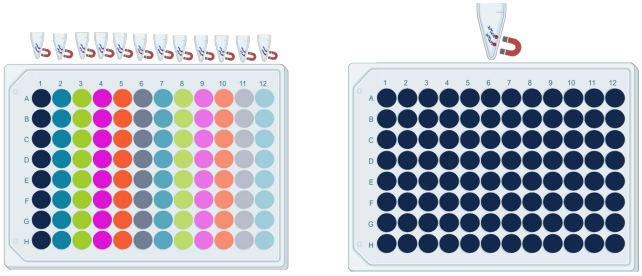

Figure 3: Single-cell exome plexity maximization. Exome enrichment of a plate of 96, single-cell ResolveDNA or ResolveOME reactions can be performed in one tube (right). This contrasts with default recommendations for plexity for bulk exome enrichment whereby a series of parallel 8-plex reactions would be required (left). Figure created with BioRender.com.

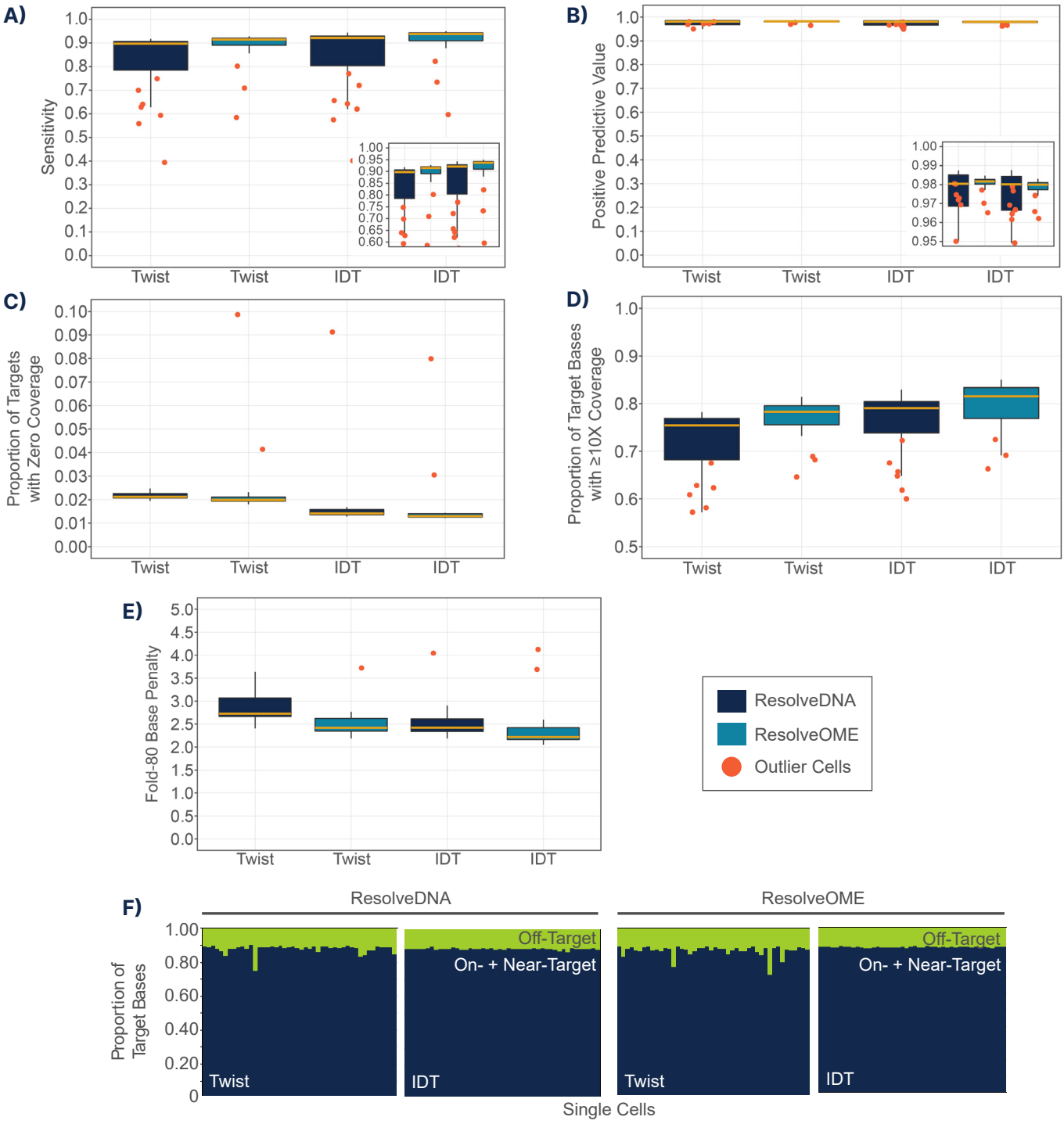

Figure 4: 96-plex single-cell exome performance. For either Twist 2.0 + comprehensive spike-in or IDT xGen v2 exomes, ResolveDNA single cells are represented by dark blue box plots and ResolveOME single cells are represented by light blue box plots. Yellow lines in box plots represent median values. Orange points represent single cells with subpar library complexity, showcasing potential impacts of this pooling strategy. Insets provide higher scale resolution in the context of the outlier cells with subpar library complexity. All data are down-sampled to 20M total paired-end/10M clusters (150 bp) per single-cell library. All data shown were generated using the 96-well formats of ResolveDNA and ResolveOME Core Kits. **Variant calling metrics:** for Genome in a Bottle HG002 cells **(A)** sensitivity (TP/(TP+FN)) and **(B)** positive predictive values (TP/(TP+FP)) are shown. **Capture performance:** The fraction of exome targets that are not represented by sequence reads are shown **(C)** as well as the fraction of target bases at or exceeding 10 fold coverage **(D)**. Fold-80 base penalty is plotted as an indicator of capture uniformity: the value represents the fold sequence coverage additionally needed to reach 80% of the mean coverage for all targets **(E)**. The proportion of on+near target bases, where near-target is defined with +/- 250 bp padding, vs off-target bases are presented in **(F)**. TP = true positive, FN = false negative, FP = false positive.

Metric	BioSkryb Platform	Twist, Mean ± SD	IDT, Mean ± SD
Target Coverage	ResolveDNA	21.39 ± 1.35	22.76 ± 0.85
	ResolveOME	21.52 ± 1.08	22.01 ± 3.37
Allelic Balance	ResolveDNA	0.67 ± 0.24	0.69 + 0.24
	ResolveOME	0.79 ± 0.13	0.80 ± 0.14
Proportion Duplication	ResolveDNA	0.05 ± 0.02	0.12 ± 0.03
	ResolveOME	0.05 ± 0.01	0.15 ± 0.06
GC Dropout	ResolveDNA	16.22 ± 1.27	9.96 ± 1.04
	ResolveOME	13.47 ± 0.64	7.50 ± 0.86
AT Dropout	ResolveDNA	0.35 ± 0.06	0.28 ± 0.17
	ResolveOME	0.48 ± 0.05	0.74 ± 0.25

Table 1: Additional performance characteristics of post-capture ResolveDNA/OME libraries. In addition to the mean target coverage achieved at 20M total paired-end reads, the table highlights allelic balance, which is the ability to detect both alleles of heterozygous sites in the HG002 genome. The value is determined using the variant allele frequency range of 0.2-0.8. The proportion duplication reflects the total proportion of post-sequencing and capture duplicates (optical + PCR duplicates). The Picard Metrics GC dropout value represents the percentage of reads that should have mapped to regions of >50% GC content but rather mapped elsewhere, while AT dropout represents the percentage of reads that should have mapped to regions of <50% GC content but mapped elsewhere. SD = standard deviation. All data shown were generated using the 96-well formats of ResolveDNA and ResolveOME Core Kits.

Sequencing Recommendations for BioSkryb's Single-Cell WES Workflows

In arriving at a sequencing recommendation for each single cell library (and thus the entire enrichment plex), we considered maximizing target coverage while simultaneously mitigating duplication rate. For a single-cell library within our defined 96-plex, we recommend 20M total paired-end reads / 10M clusters (150 bp), such that the mean depth of coverage averages 20x and duplication rate remains <15%.

While summing this read requirement across 96 cells requires nearly 2 billion reads and represents a significant fraction of a lane on a high output sequencer, the information obtained from single-cell WES is completely differentiated from typical bulk WES. Single-cell WES allows identification of co-occurrence of genomic variants and exposure of rare variants, all in addition to confirmation of a prevalent, truncal variant within a tumor sample.

Our 96-plex single-cell exome workflows generate sizable and rich datasets that require approachable and facile analysis solutions—with the distinctions that single-cell exome analysis brings relative to bulk exome analysis. BaseJumper® software enables the researcher, if desired, to perform quality control of pre-enrichment libraries, including preseq estimation of library complexity. The suite of post capture metrics presented in this note can be generated using BaseJumper's BJ-WES pipeline, including filtering and stratification of called variants across the cohort of single cells of the plex.

Conclusion

Until now, there have been limited workflows to enable concerted single-cell exome studies. The high-plexity, single-cell workflows presented in this technical note open the door for exposing the heterogeneity of tumor samples and other applications at a new pace. Our workflows are importantly rooted in the strengths of ResolveDNA and ResolveOME genome amplification upstream of exome

enrichment that provide the highest chance of uncovering the genomic variation present at single-cell resolution.

Methods

Figure 1 highlights the BioSkryb single-cell end-to-end exome workflows for both ResolveDNA genomic and ResolveOME multiomic amplification modalities.

Single Cell Preparation

HG002 B-lymphoblasts were employed for this study due to their extensive sequencing characterization by the Genome in a Bottle Consortium[3], allowing for ascertainment of post-hybrid capture SNV calling performance. HG002 cells were arrayed by FACS into 96-well PCR plates with wells containing BioSkryb Cell Buffer with two modes of staining to ensure the viability of cells dispensed: propidium iodide for negative selection of cells with compromised membrane integrity and Calcein-AM which is dependent on endogenous esterase activity of the cells for positive selection. Dispensed plates of HG002 cells were frozen at -80°C until amplification reaction setup.

ResolveDNA and ResolveOME Amplification

ResolveDNA genomic amplification or ResolveOME multiomic amplification was performed according to the respective user guides and ResolveDNA or ResolveOME genomic amplification products were subsequently subjected to a hybrid capture-specific BioSkryb library preparation protocol (included in User Guides: TAS-081, TAS-082). RNA libraries of the ResolveOME multiomic workflow, not to be subjected to hybrid capture, were prepared separately using the standard library preparation protocol of the ResolveOME user guide.

Library Preparation for Hybrid Capture

20 ng of ResolveDNA/OME genomic amplification product was input into the BioSkryb library preparation protocol for downstream hybrid capture (included in User Guides: TAS-081, TAS-082). Via enzymatic fragmentation, this modified protocol generates smaller-sized libraries suitable for hybrid capture and provides the ability to pool libraries after unique dual index (UDI) duplex adapter ligation and PCR amplification for en masse bead clean up steps in preparation for hybrid capture of the pool. The library pool was eluted in RT-PCR grade water, as emphasized in the user guides, to avoid compromising capture performance.

Hybrid Capture

In defining the composition of the enrichment plexes for this study, we intentionally included 20% of the single-cell libraries per plex with subpar (incomplete or non-uniform) genomic amplification as defined by the preseq[4] estimate of library diversity less than <3.5 E9. The rationale for this design was to reflect, in an exome pool, what a typical ResolveDNA/OME experiment may look like and account for some failures at the level of amplification, whether stemming from compromised cell integrity or technical/user errors. In addition, this provided the opportunity to highlight the consequences of a subpar amplified genome at the level of target capture and contrast that with single-cell libraries of optimal library diversity in the same enrichment.

We utilized the manufacturers' protocols for exome enrichment for 1) Twist Exome 2.0 + comprehensive spike-in and 2) IDT xGen v2 exome with the exception of a modified library pool input recommendation. Optimization for 96-plexity requires 8 microgram library pool input for Twist and 12 microgram library pool input for IDT; each representing 2x the manufacturer's recommended quantity. Four

hour hybridization was performed for both Twist and IDT workflows.

Sequencing

Sequencing of each 96-plex enrichment was conducted on an Illumina NovaSeq X instrument, targeting 20M total paired-end 150 bp reads / 10M clusters 150 bp reads per single-cell library equating to 96 X 20M = 1.92B total reads per 96 plex. To ensure this minimum read depth per single-cell library, the total number of reads was exceeded and then computationally down-sampled to the recommended 20M total reads per library.

Analysis

BaseJumper analysis pipelines were utilized to ascertain pre-enrichment libraries (BJ-DNA-QC) and post-capture performance (BJ-WES). These methods are available through our cloud portal (basejumper.bioskryb.com) and online (https://github.com/BioSkryb).

References

1. Borgström E, et al. PLOS One 2017, 12(2):e0171566. DOI 10.1371/journal.pone.0171566
2. Data on file.
3. Zook JM, et al. Sci Data 3, 160025 (2016). https://doi.org/10.1038/sdata.2016.25
4. Daley T and Smith AD. Nature Methods 2013, 10:4, 325-327. DOI 10.1038/nmeth.2375

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

Published by:

BioSkryb GENOMICS

2810 Meridian Pkwy, Suite 110, Durham, NC 27713
bioskryb.com

Version: TAS_084.1 | 07/2025