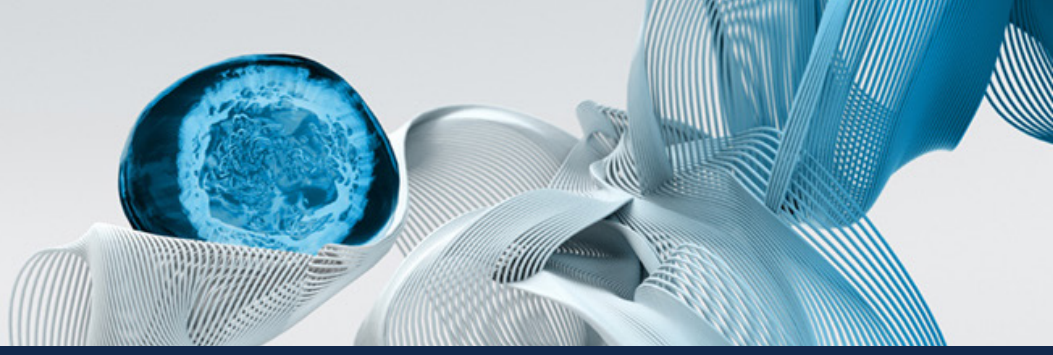
BioSkryb GENOMICS



Technical Note

ResolveDNA Lysis Optimization

Cells explored. Answers revealed.

Authors

Jeff Blackinton, PhD
Director
New Products & Services
BioSkryb Genomics

Swetha Velivela, PhD
Scientist I
Research & Development
BioSkryb Genomics

Tatiana Morozova, PhD
Senior Scientist
Research & Development
BioSkryb Genomics

Gail Joseph, PhD
Manager
Production
BioSkryb Genomics

Isai Salas-Gonzalez, PhD
Computational Biologist
Research & Development
BioSkryb Genomics

Jon Zawistowski, PhD
Senior Director
Research & Development
BioSkryb Genomics

Optimizing single cell genomic sequencing by tuning the lysis conditions of the ResolveDNA Whole Genome Amplification kit

Primary template-directed amplification (PTA), the technology underpinning BioSkryb's ResolveDNA™ Whole Genome Amplification (WGA) kit, employs innovative chemistry to efficiently and evenly amplify DNA from a single cell. This technology has unlocked the powerful combination of near-bulk genomic quality and single cell resolution. A critical step in the protocol is the lysis of the cell, thus enabling access to the genome. Here we report the modification of lysis conditions to tune the resulting genomic data to optimize reaction efficiency, genomic coverage, mutation rate, allelic balance, sensitivity, and/or specificity.

Highlights

- ResolveDNA WGA evenly and efficiently amplifies single cell DNA with near-bulk quality
- Modulating the lysis conditions can optimize the assay for robustness, allelic balance, genomic sensitivity, and mutational profile
- For certain specific applications, users may prefer a short, cold lysis
- In most cases, users will prefer lysing for twenty minutes at room temperature due to improved genomic coverage and data consistency

Introduction

The fundamental requirement for optimal single cell WGA is access to the complete genome in a manner that is compatible with DNA polymerases. To achieve this, most single cell WGA methods utilize an alkaline pH condition that lyses cell membranes and denatures both proteins and DNA. Neutralization with an acid then restores the neutral pH conditions preferred by DNA polymerases.

With this approach, however, rare chemical reactions can spontaneously occur. A well-characterized chemical reaction that can confound genomic analysis is a cytosine deamination, the spontaneous loss of an amino group from the cytosine base that effectively converts it into a uracil base. DNA polymerases read this uracil as a thymidine, and pair an adenine, resulting ultimately in the artificial detection of a thymidine, or a C to T conversion. When this occurs on the non-template strand, it results in the artificial detection of an A instead of G. These false positive variants can be very difficult to decouple from true biological variants. The number of these C to T conversions is primarily driven by time and temperature, which are also the key parameters driving cell lysis and denaturation of proteins and DNA to render the genome accessible.

The ResolveDNA WGA kit uses Primary Template-directed Amplification to generate the highest quality single cell data on the market, with quality and coverage rivaling bulk DNA preparations. With this high resolution of the method comes a better ability to detect mutations, including false positive mutations. The standard protocol for ResolveDNA WGA employs a 10 minute lysis on ice; other single cell WGA methods use notably higher temperatures and/or times. Therefore, given the exceptional resolution of the ResolveDNA WGA kit, we sought to assess PTA performance in varying lysis conditions to guide the selection of an optimal approach for a particular application.

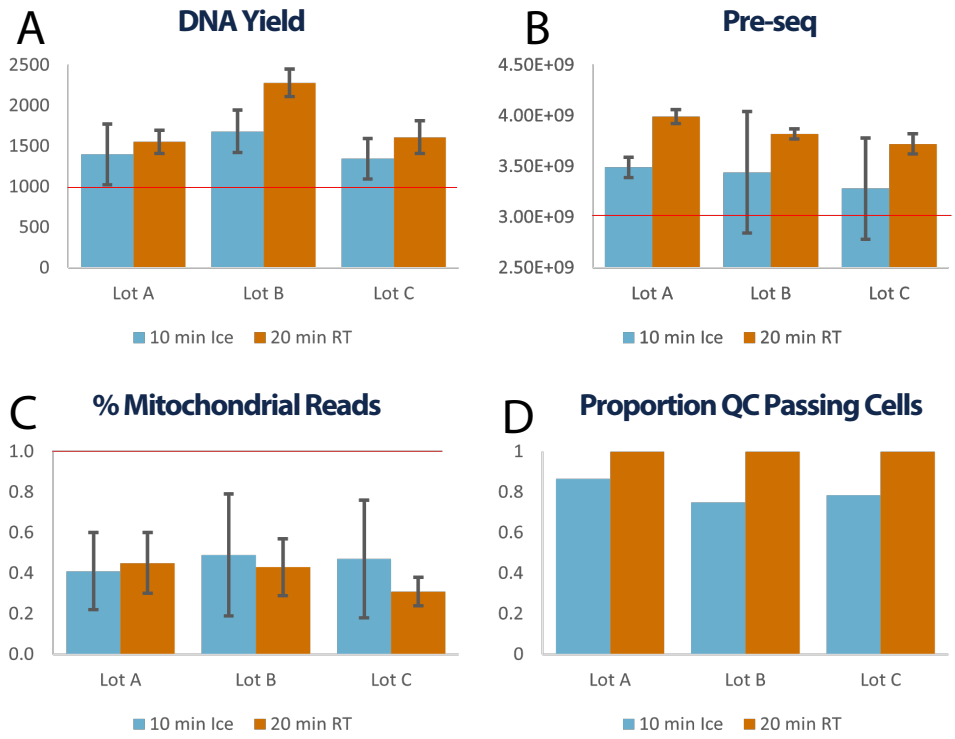


Figure 1. Improved performance with increased lysis time and temperature across lots of ResolveDNA WGA kit. All the key QC metrics for single cell PTA reaction performance improve with enhanced lysis. Red lines indicate minimum/maximum quality specifications and error bars represent standard deviation. (A) The amplified DNA yield and (B) genome diversity as measured by Pre-seq score both increase and become less variable across all lots. (C) Similarly, the percentage of mitochondrial reads, which is preferably as low as possible mostly decreased and also became less variable. (D) Overall, the proportion of cells that passed BioSkryb's stringent QC specifications increased from approximately 80% to 100%.

Methods

To evaluate the impact of modulating lysis conditions, we performed ResolveDNA Whole Genome Amplification on single NA12878 cells while varying the lysis time (10 vs. 20 minutes) and temperature (on ice vs. room temperature). In three different lots, the standard protocol lysis time of 10 minutes on ice was compared to a 20 minute lysis time at room temperature. In a fourth lot, a full factor design of experiment (DOE) was executed to evaluate the relative contribution of lysis time and temperature. Samples were evaluated first for amplified DNA yield, with a specification that a single cell reaction must generate greater than 1 µg DNA after amplification. Subsequently, sequencing libraries were created using the ResolveDNA Library Preparation kit and sequenced on an Illumina MiniSeq sequencer to a depth of 2 million reads to evaluate the sample quality. This data was analyzed using the percentage of mitochondrial reads to assess cell and nuclear lysis quality, and the pre-seq software package[1] which uses the diversity of genomic data to extrapolate the number of discoverable unique bases; for a fully recovered genome, the Pre-seq value is greater than the size of the human genome (3.1 billion bases). From the DOE samples, libraries from four cells from each condition that passed all specifications and were selected as representative of their

Lysis Time	Lysis Temp	# of PTA reactions	DNA Yield Avg (ng, ±SD)	% Cells >1µg DNA Yield	# of samples sequenced	Pre-seq Avg	% Passing Calls (Seq)	% Mitochondrial Reads
10	Ice	51	1184 (±399)	76%	8	3.44e9 (±5e8)	75%	0.56 (±0.22%)
10	RT	16	1108 (±177)	81%	8	3.77e9 (±7e7)	100%	0.34 (±0.13%)
20	Ice	15	1461 (±233)	93%	8	3.52e9 (±7e8)	88%	0.43 (±0.18%)
20	RT	16	1573 (±205)	100%	8	3.85e9 (±7e7)	100%	0.33 (±0.19%)

Table 1. Longer and warmer lysis improves early PTA quality metrics in a fourth ResolveDNA WGA kit lot (Lot D).
Yield increases, yield variation decreases, and the percentage of cells that pass specifications increases with increased lysis time and/or temperature. Additionally, increased lysis time and/or temperature increases the diversity of resulting sequencing libraries and a reduction in the proportion of mitochondrial reads. All of these results suggest that increased lysis time or temperature results in an additive improvement in reaction quality.

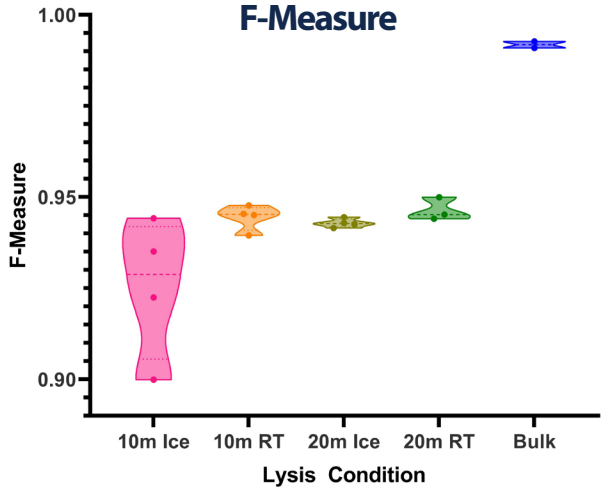


Figure 2. Increased lysis time and/or temperature enhances sensitivity and precision. F-measure, a calculation of the harmonic mean between sensitivity and precision, is subtly lower when lysing for 10 minutes on ice. Other methods at best approach 0.7[2].

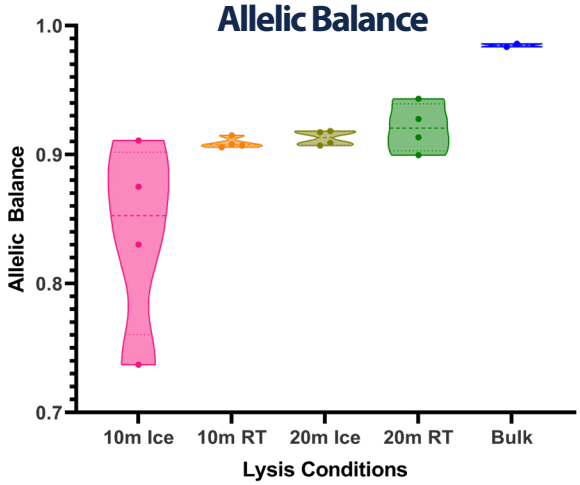


Figure 3. Allelic balance with increased lysis time and/or temperature. The proportion of heterozygous SNPs with greater than 10% representation of each allele is greater than 90% when enhancing the lysis. It should be noted that the allelic balance at 10 minutes on ice is still market leading performance for single cells by a large margin[2].

respective conditions in quality were submitted for whole genome sequencing at a depth of ~180 million 2x150 paired end reads (~15X coverage).

To determine genomic coverage, sensitivity, specificity, and allelic balance, we first aligned the reads from each cell to the GRCh38 reference genome, removed duplicates, and then performed joint-genotyping using a software-optimized version of GATK[4]. The sensitivity and precision of single-nucleotide variant calling was computed using the genome-in-a-bottle (GIAB) variant call-set as the ground truth for each sample[3]. The precision and recall of variant calling was determined for each sample at the values that maximized the harmonic mean (F-measure).

Allelic balance was computed in each sample by first computing the proportion of reads supporting the alternate allele at each known heterozygous site, as defined by the GIAB. An allelic balance value was then assigned to each cell by measuring the proportion of heterozygous sites that had between 10% and 90% alternate allele frequencies.

Results

Across four different lots, we observed clear improvements in metrics that predict ResolveDNA WGA reaction quality in our standard QC process. In three lots, we compared 10 minutes on ice to 20 minutes at room temperature. In virtually all cases, average DNA yield and Pre-seq values increased and became less variable, while the percentage of mitochondrial reads decreased (Figure 1). In a fourth lot, we observed the metrics to improve in an additive fashion when isolating the contributions of lysis time or temperature (Table 1).

The QC testing results, however, are merely predictors of the whole genome sequencing quality. To assess the quality of the whole genomes of these single cells, four representative cells from each condition were submitted for sequencing at a depth of ~15x coverage. We calculated several metrics in these samples, including F-measure (the harmonic mean between sensitivity and precision), allelic balance, and the types and rates of various mutations observed. We observed ResolveDNA WGA samples have remarkable SNV-calling sensitivity and specificity based on the F-measures observed (Figure 2). Across the board, over 90% of SNVs could be determined with high precision.

This contrasts to typically less than 50%, and up to 70% at best, for competitive methods[2]. A subtle but increasing trend, however, was observed as lysis time and temperature increased. A similar phenomenon was observed in the dataset comparing allelic balance across conditions, though the difference between the 10 minutes on ice condition and the other conditions was more pronounced (Figure 3). In heterozygous SNPs with greater than 6 reads, over 90% contained both alleles represented at greater than 10% in the three enhanced lysis conditions. These data sup-

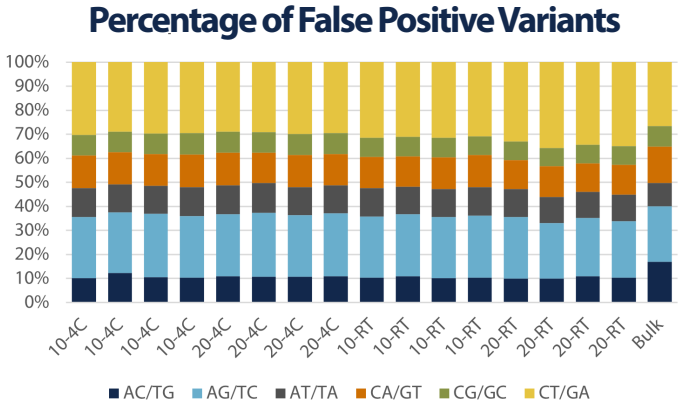


Figure 4. The percentage of false positive variants by mutation type. False positive variants were determined in reference to the genome in a bottle NA12878 reference[3] and classified by mutation type. Since complementary strands cannot be discriminated, A to C mutations, for example, are classified with T to G. The proportions of mutational types remain quite consistent across the lysis conditions, with a subtle increase in the C to T variant observable (from 29.5% to 34.4%) as the lysis conditions increase in time and temp.

port the increased library diversity and sequencing quality when utilizing longer lysis times at higher temperatures.

The expected trade-off of this increase in quality, however, is the possibility for an increased mutation rate. Alternative WGA methods report a breadth of C to T false positive variant proportions, from approximately 30% to over 50%[2]. To evaluate the impact of these alternative lysis conditions on the false positive variant rate, we identified variants in our dataset relative to the GIAB reference for NA12878[3]. These variants were segregated based on base transition and paired with their non-reference strand alternates. The proportion of variants of each type were evaluated (Figure 4). The 10 minute on ice condition averaged 29.5% of variants from C to T transitions, equivalent to other WGA approaches and close to bulk clonal sequencing[2]. A similar rate was observed lysing for 20 minutes on ice, while the rate increased to 31.1% when lysing at room temperature for 10 min. With the harshest lysis of 20 minutes at room temperature, 34.4% of variants derived from C to T transitions. While this is a measurable increase, it remains dramatically lower than several alternative WGA approaches[2].

Conclusions

First and foremost, this nuanced evaluation of data quality is only feasible using PTA technology and the ResolveDNA WGA kit. The combination of genomic coverage, allelic balance, and SNP sensitivity and specificity is unmatched among single cell WGA approaches[2].

Secondly, the selection of appropriate methods is crucial to attaining the goals of a given project. The execution of a WGA reaction is no exception. We recommend that investigators carefully consider the goals of the project when determining the study lysis conditions as particular choices can enhance particular outcomes. We expect, for example, that researchers focusing on studying mechanisms of mutation, or identifying *de novo* or individual somatic mutations between cells would prefer low temperature lysis, as the benefit of the greatest confidence in the unique mutations identified outweighs the cost of a small fraction of poor-quality reactions and a subtle decrease in sensitivity and allelic balance.

For researchers interested in understanding cancer, sample in variants in a mixed cell population, structural variation, and for many other applications, the benefits of improved genomic coverage, SNP sensitivity, and allelic balance will outweigh the cost of a few additional, mitigatable false positive mutations. This understanding empowers investigators to select the lysis conditions that best align with their experimental objectives and generate the highest possible quality data to address critical biological questions.

References

1 T. Daley, A. D. Smith, Predicting the molecular complexity of sequencing libraries. Nat. Methods. 10, 325–327 (2013).

2 V. Gonzalez-Pena, S. Natarajan, Y. Xia, D. Klein, R. Carter, Y. Pang, B. Shaner, K. Annu, D. Putnam, W. Chen, J. Connelly, S. Pruett-Miller, X. Chen, J. Easton, C. Gawad, Accurate genomic variant detection in single cells with primary template-directed amplification. Proc. Natl. Acad. Sci. 118 (2021), doi:10/gkt3xd.

3 Zook JM, McDaniel J, Olson ND, Wagner J, Parikh H, Heaton H, Irvine SA, Trigg L, Truty R, McLean CY, De La Vega FM, Xiao C, Sherry S, Salit M. An open resource for accurately benchmarking small variant and reference calls. Nat Biotechnol. 37, 561-566. (2019), doi: 10.1038/s41587-019-0074-6.

4 Sentieon (2018). Sentieon DNASeq. https://www.sentieon.com/products/.

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

Published by:

BioSkryb GENOMICS

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

TAS.025, 01/2022