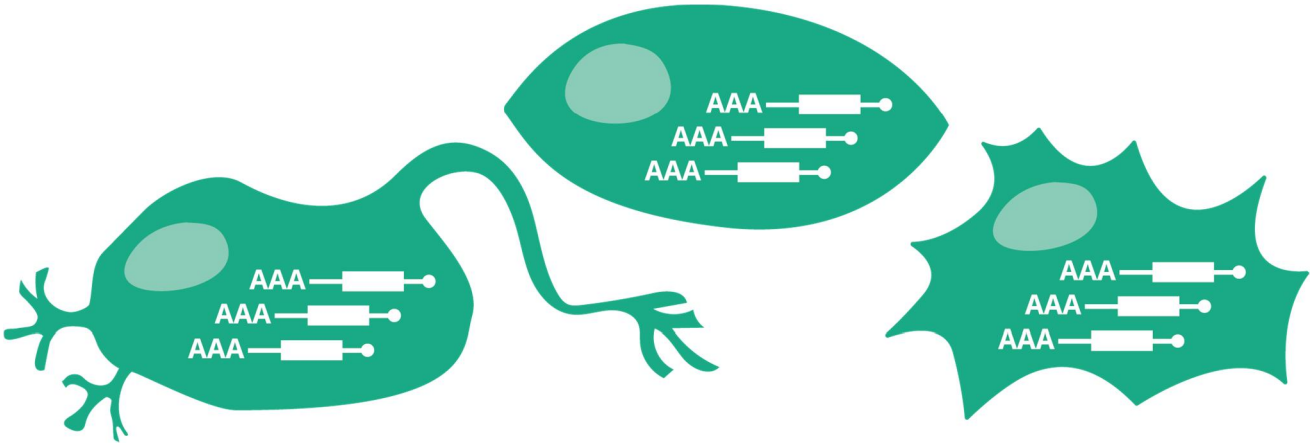
# High throughput analysis of single cell transcriptomes with Dolomite Bio's Nadia Instrument

Encapsulating single cells with barcoded mRNA capture beads on the Nadia Instrument

Version: 3.0

Issue Date: 18 October 2018

Author: JF2




**Dolomite Bio is a brand of Blacktrace Holdings Ltd**
Registered office: Blacktrace Holdings Ltd. 27 Jarman Way, Royston, SG8 5HW, UK
Reg in England No. 04257809

# Contents

# Disclaimer

The Nadia product is for research use only and is not to be used for any other purposes, including, but not limited to, use in drugs, *in vitro* diagnostics, or use in humans. The document and its content are proprietary to Dolomite Bio and is intended only in connection with the products described herein and for no other purposes.

The document and its contents shall not be reproduced, disclosed, or used for any other purpose without written consent of Dolomite Bio. Dolomite Bio does not convey any license under its trademark, copyright, or common-law rights nor similar rights of any third parties by this document.

Dolomite Bio makes no warranty of any kind, either expressed or implied. This includes merchantability for this product, and the fitness of the product for any purpose.

Nadia is a trademark of Dolomite Bio. All other brands and names mentioned within this document are the property of their respective owners.

# Summary

Dolomite Bio has developed the Nadia Instrument, a high-throughput platform for automated droplet generation from multiple biological sample for single cell research. This application note describes the Drop-seq workflow according to Macosko *et. al.* (Macosko E., *et al.* "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanolitre Droplets." *Cell* 161:1202) for the generation of scRNA-seq libraries using the Dolomite Bio Nadia Instrument.

# Introduction

Cell encapsulation within microfluidic droplets is an exciting, cutting-edge technique that, for the first time, enables the analysis of thousands of single cells within one experiment. This unprecedented analytical power enables the discovery of previously unidentified or cryptic cell types in a tissue of interest and allows the characterisation of individual cells within biologically important processes.

Recently, a number of different approaches describing high throughput single cell RNA-Seq were published including Drop-seq (Macosko E., *et al*., "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets." *Cell* 161:1202) as well as inDrop (Klein, AM., *et al*. "Droplet Barcoding for Single-Cell Transcriptomics Applied to Embryonic Stem Cells." *Cell* 161:1187). The inDrop method uses barcoded, deformable beads in lysis buffer that are encapsulated with a cell and a reverse transcription reaction mix. Inside the droplet cells are being lysed, mRNA is bound by the bead and immediately reverse transcribed inside the emulsion.

In contrast, the Drop-seq protocol developed by Macosko *et al*., encapsulates single cells alongside non-deformable beads in droplets (Figure 1). The beads have surfaces coated with oligonucleotides containing a (dT)30 stretch allowing for the capture of polyadenylated mRNA, which after emulsion breakage is subsequently reverse transcribed into cDNA. Each bead also carries a unique DNA barcode that identifies not only individual cells, but also individual mRNA transcripts derived from any given cell.

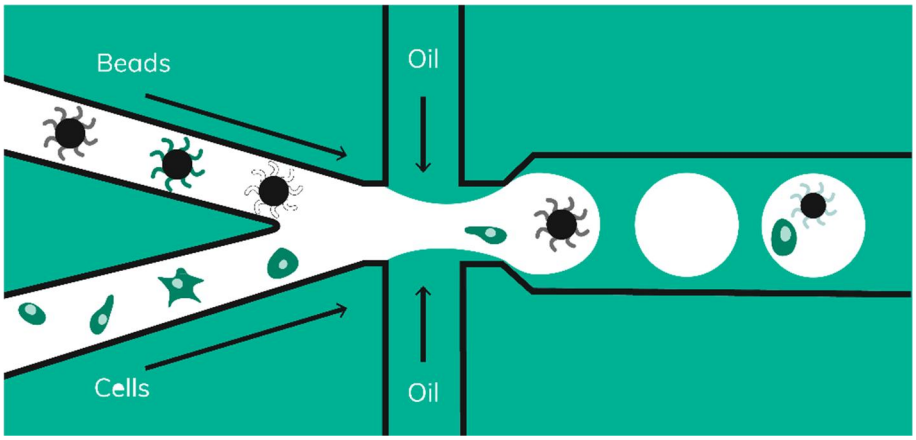



**Figure 1** Schematic of the encapsulation of barcoded mRNA capture beads with single cells. Tens of thousands of single cells are co-encapsulated with uniquely barcoded beads. The cells are lysed inside the droplets and their mRNA content is captured on the beads. These beads are subsequently recovered and processed for downstream analysis using NGS technologies.

With the increasing availability of commercial instruments, automated high-throughput single cell analysis is fast becoming a mature technology. Microfluidic systems first built by engineers as prototypes to carry out cell encapsulation were neither robust nor suitable for use in day-to-day research. Dolomite Bio, a world leader in the design and manufacture of microfluidic systems and components, developed the Nadia Instrument for easy-to-use, high throughput analysis of single cell samples. Nadia allows the encapsulation of up to 8 single cell samples in parallel using a fully automated process, including chilling and homogenisation of samples.

To ensure truly singe cell analysis, high-throughput systems such as Dolomite Bio's Nadia Instrument need to ensure the encapsulation of only one cell alongside a single bead within a droplet. This can generally be achieved through the dilution of cell and bead suspensions. However, to achieve this goal a fine balance between quality (fewer cells but data of higher quality) and quantity (higher throughput but data of lower quality) needs to be struck.

On Nadia, a standard 20-minutes run will yield more than 6000 individual cellular transcriptomes. While this setting is routinely used in Drop-seq on Nadia, the Nadia Instrument also offers the possibility to adjust for smaller sample quantities by reducing run time and sample volume in case of limited input materials.

This application note describes results obtained from the encapsulation of single cells with barcoded mRNA capture beads for single-cell RNA sequencing using the Drop-seq protocol on Nadia.

# Materials and Methods

**Droplet system.** Dolomite Bio's Nadia Instrument (Figure 2) is designed to allow high throughput analysis of single cells using droplet microfluidics. It produces highly monodispersed droplets using three independent, ultra-smooth pressure pumps. The Nadia Instrument has been designed for ease-of-use guiding the user through a chosen application with step-by-step instructions that are clear and effortless to follow. Samples are chilled and stirred during the encapsulation process ensuring high quality and evenly suspended cell and bead suspensions. Runs on the Nadia instrument can be customised by allowing the user to change sample volume, sample number and cell stirrer speed to adapt for different experimental requirements.


**Figure 2** Nadia Instrument

**Preparation and encapsulation of beads and cells.** A detailed description of material and methods used in this application note can be found in the Drop-seq on Nadia protocol. Please contact DolomiteBio@syrris.zohosupport.com for further information.

**Cell viability.** To demonstrate that gentle stirring employed to maintain cells in suspension exerted no adverse effects, cellular viability was tested using 0.4% Trypan Blue Solution (#15250061, ThermoFisher Scientific). MEF cells at a density of 6400 cells/µl in a 1:1 solution of PBS-BSA + 0.4% Trypan Blue were stirred in the Nadia Instrument for 30 min at 4˚C and compared to cells that had been stored on ice for 30 min without stirring. Cell viability was measured by determining the number of viable cells (unstained cells).

**Preparation and NGS-analysis of single cell DNA libraries.** The steps following cells and beads encapsulation on the Nadia Instrument were conducted as described in the Drop-seq on Nadia protocol. Aliquots of 2,000 beads were used in 6 PCR reactions, which corresponded to approximately 600 STAMPs (Single-cell Transcriptome Attached to MicroParticles). The quality of DNA following PCR amplification and tagmentation was determined using a BioAnalyzer.

**Illumina sequencing and bioinformatic pipeline.** The tagmented DNA libraries were sequenced on Illumina's NextSeq 500 instrument using a 2x75 bp paired-end run. We configured read lengths of 26 bp for read1, 8 bp for the index read and 116 bp for read 2. For data-analysis, the "dropSeqPipe" (https://github.com/Hoohm/dropSeqPipe) a computational pipeline developed by the McCarroll Lab at Harvard was used. The analysis was performed as described in Macosko *et al*., 2015.

# Results

In order to assess the performance of the Nadia instrument as well as the overall workflow efficiency, a series of mixed-species experiments was performed using the Drop-seq on Nadia protocol. Factors that will influence the overall workflow efficiency and data quality include emulsion quality, PCR duplication rate, gene capture rate and reproducibility. The quality of the emulsion and the cell suspension will impact the doublet rate, which is the percentage of barcoded beads that have captured more than a single cell. The PCR duplication rate is defined by the ratio of detected UMIs (Unique Molecular Identifier) to reads per cell. Gene capture rate is defined by the number of genes detected at a given sequencing depth, which is indicative of the efficiency of the overall protocol. Lastly, comparing results from experiments carried out on different days by different users will give an indication of the reproducibility.

**Flowing beads and cells.** The Nadia Instrument contains built-in stirrers that gently agitate the cell and bead suspensions. This ensures that samples stay in suspension for the duration of the run. In addition, beads and cells are evenly encapsulated reducing the chance of either cell or bead doublets. The stirrers avoid cell and bead damage by rotating gently and away from the walls or the bottom of the chambers. To evaluate emulsion quality, monodispersity was estimated by measuring the average droplet size and the rate of encapsulation of beads (Figure 3).

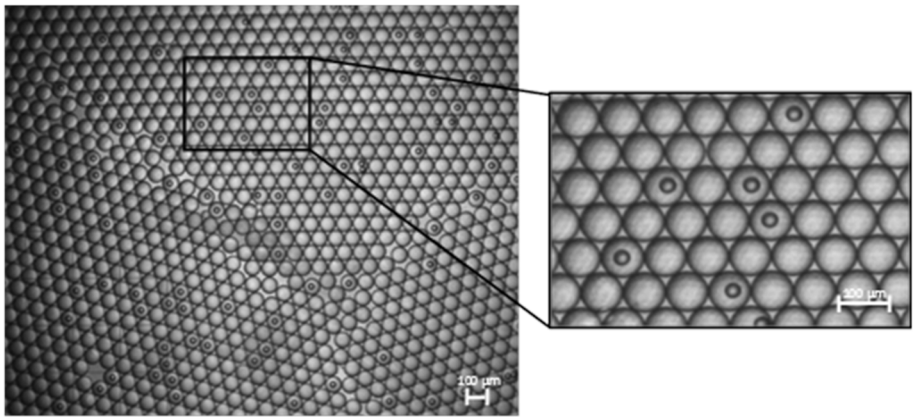



**Figure 3** Droplets produced during a Drop-seq run on the Nadia Instrument.

The droplet quality was analysed over 30 runs performed either in-house or at customers' sites. The average droplet size was found to be 81 µm with a CV (Coefficient of Variation) of 4 %. The rate of bead encapsulation was determined by "number of droplets containing a bead" divided by the "total number of droplets". On average, 11 % of all droplets contained a bead which is consistent with a theoretical encapsulation rate of 10%.

Cell viability was tested to demonstrate that gentle stirring employed to maintain cells in suspension exerted no adverse effects. Cells that had been stirred on Nadia for 30 min at 4˚C were compared to cells that had been stored on ice for 30 min without stirring. No significant difference was observed between those two experiments, with 91 % of the cells viable after stirring and 90 % viable when stored on ice without stirring.

**Preparation of single cell cDNA libraries for NGS-analysis.** After recovery of the emulsion from the instrument beads were treated as described in the Drop-seq on Nadia protocol. After PCR amplification (Figure 4 A) and tagmentation (Figure 4 B) a proportion of the cDNA library was analysed on a BioAnalyzer to estimate size

distribution and overall quality. The size distribution of the PCR products ranged from 600 to 2500 bp which correlated with a mammalian cDNA library with an expected range of 400 to 1000 bp.

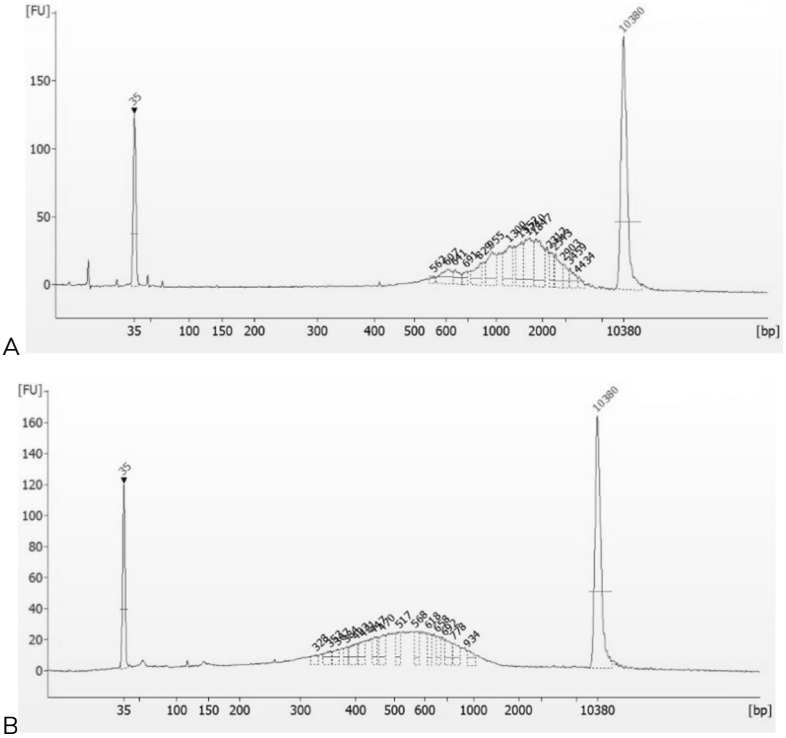



**Figure 4** BioAnalyzer results before (A) and after tagmentation (B).

**Performance of the Drop-seq protocol on the Nadia Instrument.** The quality of the sequencing data was assessed based on knee- and Barnyard plots. The knee plot allows for the estimation of the number of productive beads per sample (i.e. beads that have captured mRNA derived from a cell). Each bead can be associated with a unique cell barcode and a corresponding number of NGS reads. The cell barcodes are then arranged in descending order of reads and visualised alongside the cumulative fraction of reads. In the plot described below, 12,000 beads were used for amplification of the cDNA library. At a dilution of 1 cell in 20 droplets, ~5% of beads should have captured mRNA from a cell yielding a theoretical 600 cellular transcriptomes (STAMPS). From the plot below, an inflection point can be observed at 537 STAMPS a number which is consistent with the 600 cells predicted by Poisson statistics (Figure 5).

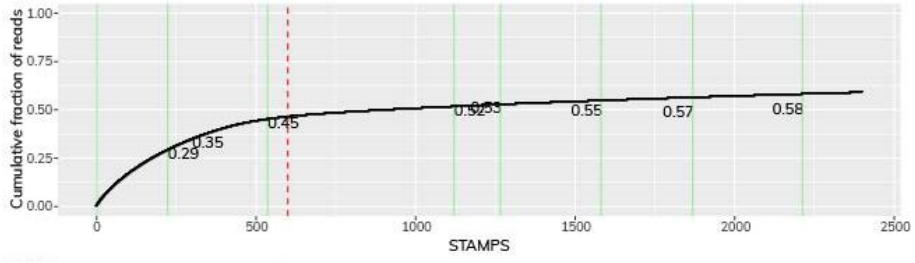



**Figure 5** A knee plot comparing cell barcodes in descending order of reads alongside the cumulative fraction of reads.

To determine the doublet rate in the mixed-species experiment a Barnyard plot was used. This plot depicts STAMPs as individual data points each associated with a number of human or mouse transcripts (Figure 6). For the dataset shown, 261 STAMPs were associated with only mouse transcripts 258 with only human transcripts and 18 STAMPs carried transcripts from both species equating to a doublet rate of 6.6 %.

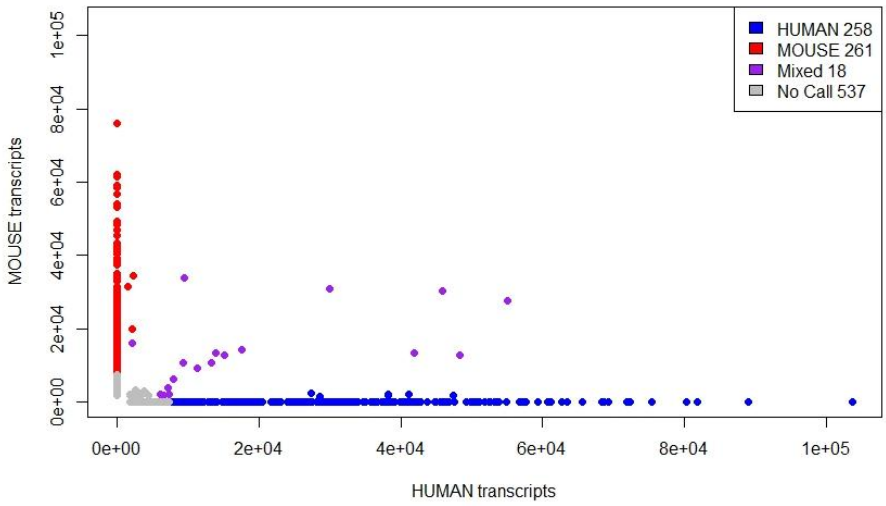



**Figure 6** A Barnyard plot depicting STAMPs as individual data points each associated with a number of human or mouse transcripts

### Evaluation of PCR duplication rate and gene capture efficiency

To evaluate PCR duplication rate and gene capture efficiency a set of four independent Drop-seq on Nadia runs were compared to data published by Macosko *et. al.* 2015. The dataset consisted of two replicates using 200 cells (STAMPS) at low sequencing depth (average of ~20,000 reads/ cell), and two replicates using 600 cells (STAMPS) one at an average depth of 50,000 reads/ cell and the other one at a higher depth of~120,000 reads/ cell. An overview of this comparison is given in Table 1. It summarizes the number of cells as determined with knee plots, NGS reads per cell, doublet rate and median number of genes and UMIs detected.

Overall, the doublet rates were comparable across all samples with an average doublet rate of about 7% which is also comparable to previously the published Macosko dataset. The median of detected genes and UMIs depended upon the number of cells analysed and on the sequencing depth per cell.

***Table 1*** Comparison of a set of four independent Drop-seq on Nadia runs to published data

| | | | | Median | |
|---|---|---|---|---|---|
| **Sample** | **# Cells** | **NGS reads/cell** | **Doublet rate** | **Genes** | **UMIs** |
| Nadia 200 R1 | 272 | 19,757 | 7.3 % | 2,728 | 6,413 |
| Nadia 200 R2 | 185 | 26,224 | 6.4 % | 3,107 | 6,299 |
| Nadia 600 R1 | 537 | 121,975 | 6.6 % | 5,525 | 23,233 |
| Nadia 600 R2 | 654 | 50,024 | 7.8 % | 4,132 | 11,704 |
| Macosko 1000 | 1135 | 124,161 | 7.0 % | 4,868 | 16,295 |

To evaluate the gene capture efficiencies, cells analysed for each of the five datasets were plotted according to the number of genes and reads detected per cell (Figure 7). Figure 7 shows that the number of genes detected per cell increased with sequencing depth. While efficient gene capture was observed even at low sequencing depth, thereby reducing cost for NGS analysis, increased sequencing depth will in turn allow the detection of less well represented transcripts.

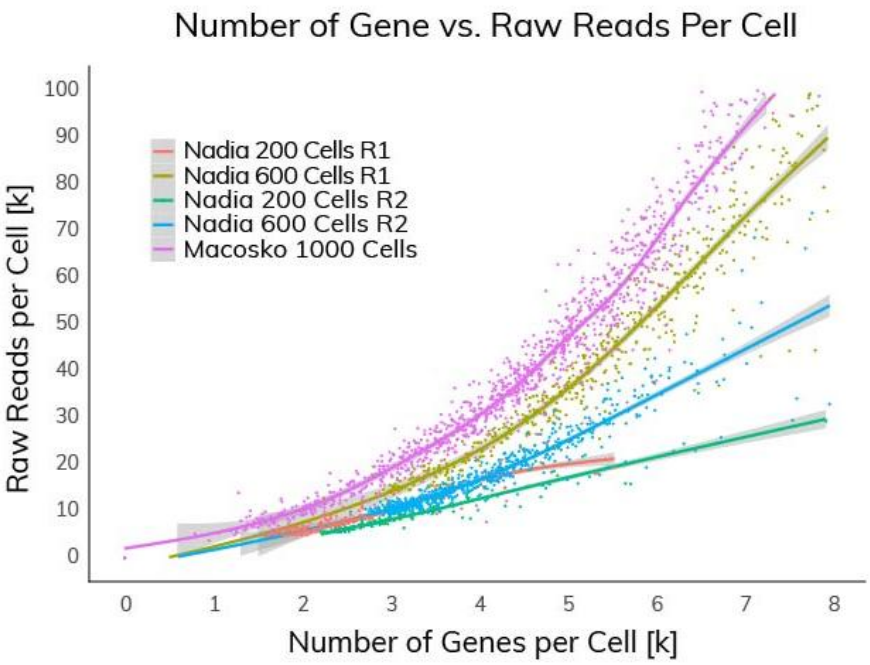



***Figure 7*** The number of detected genes was compared to the read depth across all samples.

Plotting the number of detected UMIs against reads per cell gives an indication of the PCR duplication rate during cDNA amplification (Figure 8). The black line in this plot reflects the ideal ratio of 1 UMI per read. All Drop-seq on Nadia datasets aligned very closely to this ideal line, however, the degree of deviation depended also upon sequencing depth.

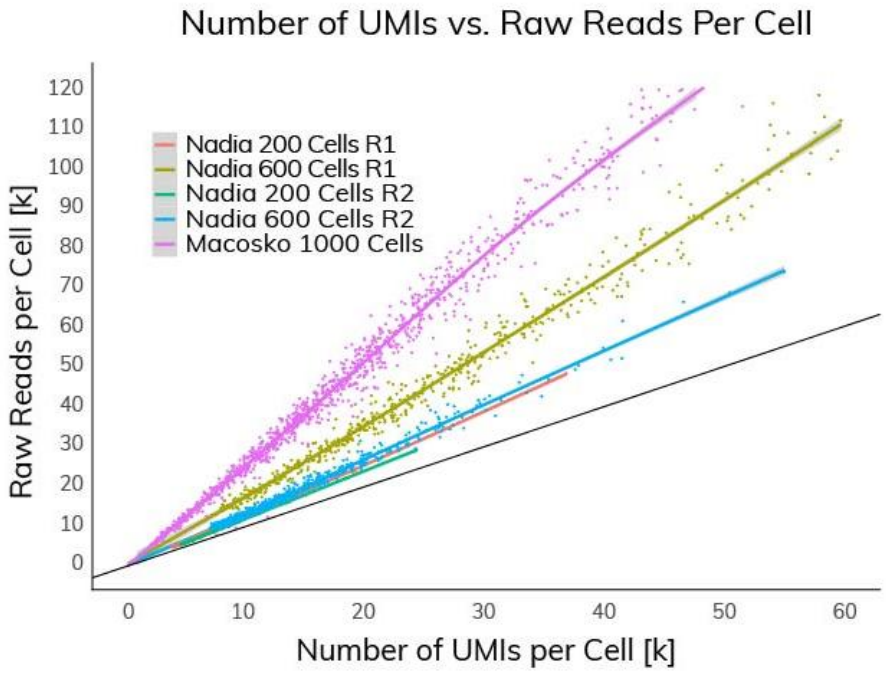



**Figure 8** The number of detected UMIs was compared to the read depth per cell across samples. The black line represents the ideal ratio of 1 UMI per 1 read, indicating no PCR duplicates.

**Reproducibility of Drop-seq data on the Nadia instrument.** Data described in this study were generated by different users on different days. Comparing these experiments enabled assessment of the inter-experiment as well as inter-user variability of samples processed with the Nadia instrument. Pairwise-comparisons between all datasets using the "number of UMIs detected per gene across all cells" were performed (Figure 9). R values ranged from 0.84 to 0.93 across all datasets indicating high reproducibility between users and across replicates.

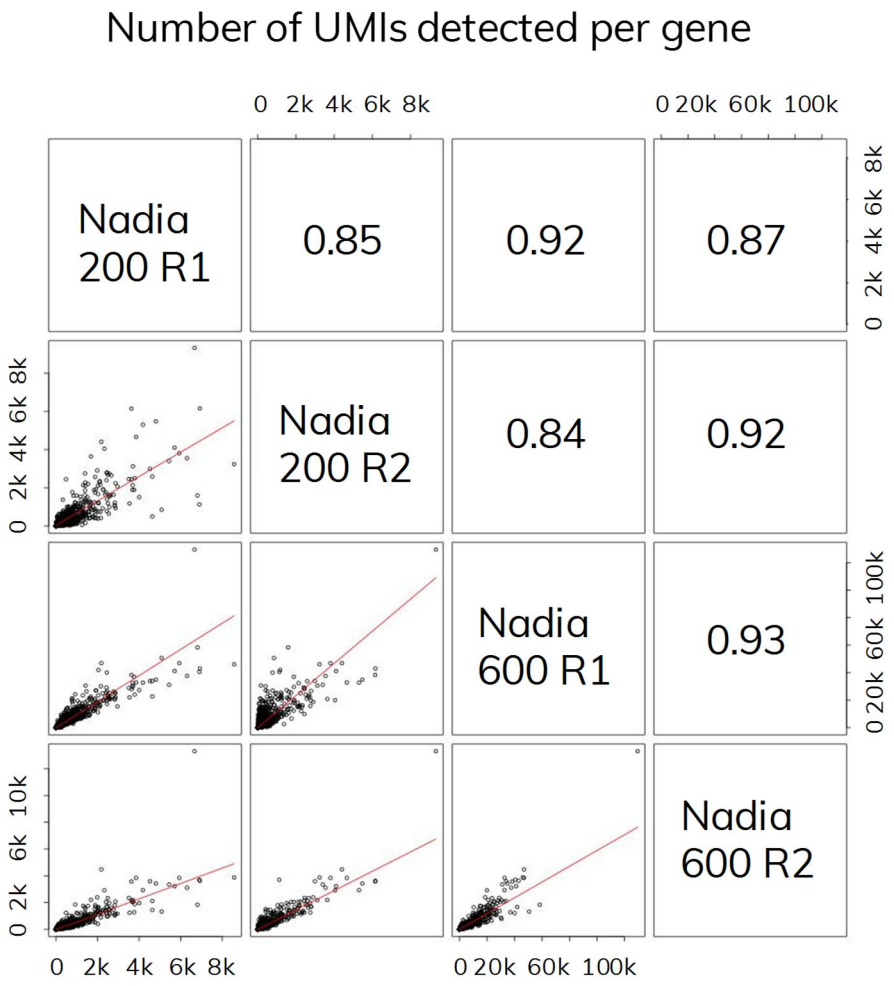



**Figure 9** This plot shows the correlation of four datasets from different users generated on different days on the Nadia instrument. Compared are the numbers of UMIs detected per gene.

# Discussion

The development of high-throughput single cell analysis gives, for the first time, insight into gene expression profiles of thousands of cells at the single cell level to determine tissue heterogeneity or identify new or mutated cell clusters. It is thus unsurprising that many scientists are eager to adopt this methodology for their research.

We have shown that the Nadia Instrument can efficiently, cost effectively and reproducibly encapsulate cells alongside barcoded mRNA capture beads. Nadia produces highly monodisperse droplets with a cell capture rate of 10 %. We have also shown that gentle stirring on Nadia, both maintains beads and cells in suspension and facilitates encapsulation without damaging cells. Furthermore, stirring ensures an even encapsulation of cells and beads as the cell doublet rate was found to be as low as 6.7 % with over 6,000 cells captured in a complete run.

Sets of 200 and 600 captured cells were processed and analysed using Illumina's NGS to evaluate PCR duplication as well as cell capture rate. This data was compared to a 1000 STAMP dataset published by Macosko *et. al.* in 2015 which was used as a benchmark to determine the quality of Drop-seq data obtained from the Nadia Instrument. The datasets were evaluated based on their doublet rate, capture rate of UMIs per read as well as number of genes detected. On average, all single cell samples processed on Nadia showed a doublet rate of 7 % which was consistent with the doublet rate observed for Macosko's dataset. In addition to the quality of droplet formation, cell doublet rate is also highly dependent on the cell preparation techniques and cell concentrations used. Macosko and colleagues found that cell doublet rates ranged from 0.36% to 11.3% when using cell sample concentrations of between 12.5 cells/µl to 100 cells/µl (Macosko *et. al.*, 2015). Nadia works with cell concentrations of 300 cells/µl, reducing the overall run time while at the same time ensuring that the doublet rate remains low. Shorter run times lowers the risk of cells dying or altering their transcriptomes due to stress. This risk is further reduced by the in-built temperature control in the Nadia instrument which keeps samples at a constant temperature of 4˚C during a run.

Another important measurement to assess the quality as well as the efficiency of an scRNA-seq run is the number of UMIs and genes detected per cell at a given sequencing depth. These numbers will depend upon a variety of factors that include but are not limited to, the rate at which the mRNA capture beads are able to bind transcripts, cell lysis, efficiency of reverse transcription and amplification of the library as well as sequencing depth per cell.

Due to a similar sequencing depth the "Nadia 600 R1" and the "Macosko 1000 STAMP" datasets are most suited for comparison of the above-mentioned factors. With Nadia, we detected on average ~20,000 UMIs per cell which is significantly more than the ~16,000 UMIs in Macosko's dataset. Furthermore, we detected over 5500 genes per cell using the Nadia workflow which is significantly more than the 4800 in Macosko's dataset. In this regard, the performance of Nadia is comparable or modestly superior to other commercial platforms. This may be due to the manner in which RT (reverse transcription) is performed in Drop-seq, i.e. outside of the droplet and thereby separated from cell lysis. This offers the possibility to use very strong lysis agents that increase the likelihood of complete lysis for a wider range of cell types. Furthermore, it also eliminates any adverse effects a lysis buffer can have on the efficiency of the RT reaction. Interestingly, we could also show that with Nadia we were able to detect gene numbers as high as 2000 – 3000 at low sequencing depths of 20,000 reads per cell. This is especially important should a quick characterisation of samples or cost-effective sequencing of large sample numbers be required.

UMIs present on each oligo attached to a bead, have proven a useful tool when differentiating between numbers of transcripts and PCR duplicates. Ideally, one would expect only a single read per UMI detected if no PCR duplication occurred. Thus, the closer the data points of a sample are to this threshold the lower the PCR duplication rate. All datasets generated on the Nadia instrument were very closely aligned to this ideal threshold indicating that duplication was very low during PCR which in turn will significantly reduce sequencing cost. Furthermore, we observed an influence of the sequencing depth on the ideal 1:1 ratio. As the deeper a dataset had been sequenced the more it deviated away from this ideal 1:1 ratio.

Finally, we evaluated reproducibility of data produced on the Nadia instrument. Here, we looked not only at samples processed on different days but also by different users. R values from pairwise-comparisons across all datasets indicated a high level of inter-sample and inter-user reproducibility.

# Conclusion

Using Drop-seq on Nadia, we obtained high-quality single cell cDNA libraries that were further analysed to assess the performance of the Nadia Instrument as well as the overall workflow efficiency. To that end we evaluated emulsion quality, PCR duplication and gene capture rates. We demonstrated that Drop-seq on Nadia not only yielded low doublet rates, but also resulted in high gene capture as well as low PCR duplication rates making Drop-seq on Nadia a highly cost-effective approach for single cell transcriptome analysis.

# Product Information

| Instrument/Consumables | Order Number |
|---|---|
| Nadia Instrument | 3200590 |
| Nadia Training Cartridge | 3200605 |
| Nadia Cartridge for scRNA-Seq and sNuc-Seq - 8 Runs (8x1) | 3200648 |
| Nadia Cartridge for scRNA-Seq and sNuc-Seq - 8 Runs (2x2 & 1x4) | 3200649 |
| Nadia Cartridge for scRNA-Seq and sNuc-Seq - 8 Runs (1x8) | 3200650 |
| Nadia Cartridge for scRNA-Seq and sNuc-Seq - 40 Runs (40x1) | 3200651 |
| Nadia Cartridge for scRNA-Seq and sNuc-Seq - 40 Runs (10x2 & 5x4) | 3200652 |
| Nadia Cartridge for scRNA-Seq and sNuc-Seq - 40 Runs (5x8) | 3200653 |