APPLICATION NOTE

Single-cell gene fusion and rare event discovery with ribodepletion-enhanced 10x–ONT sequencing.

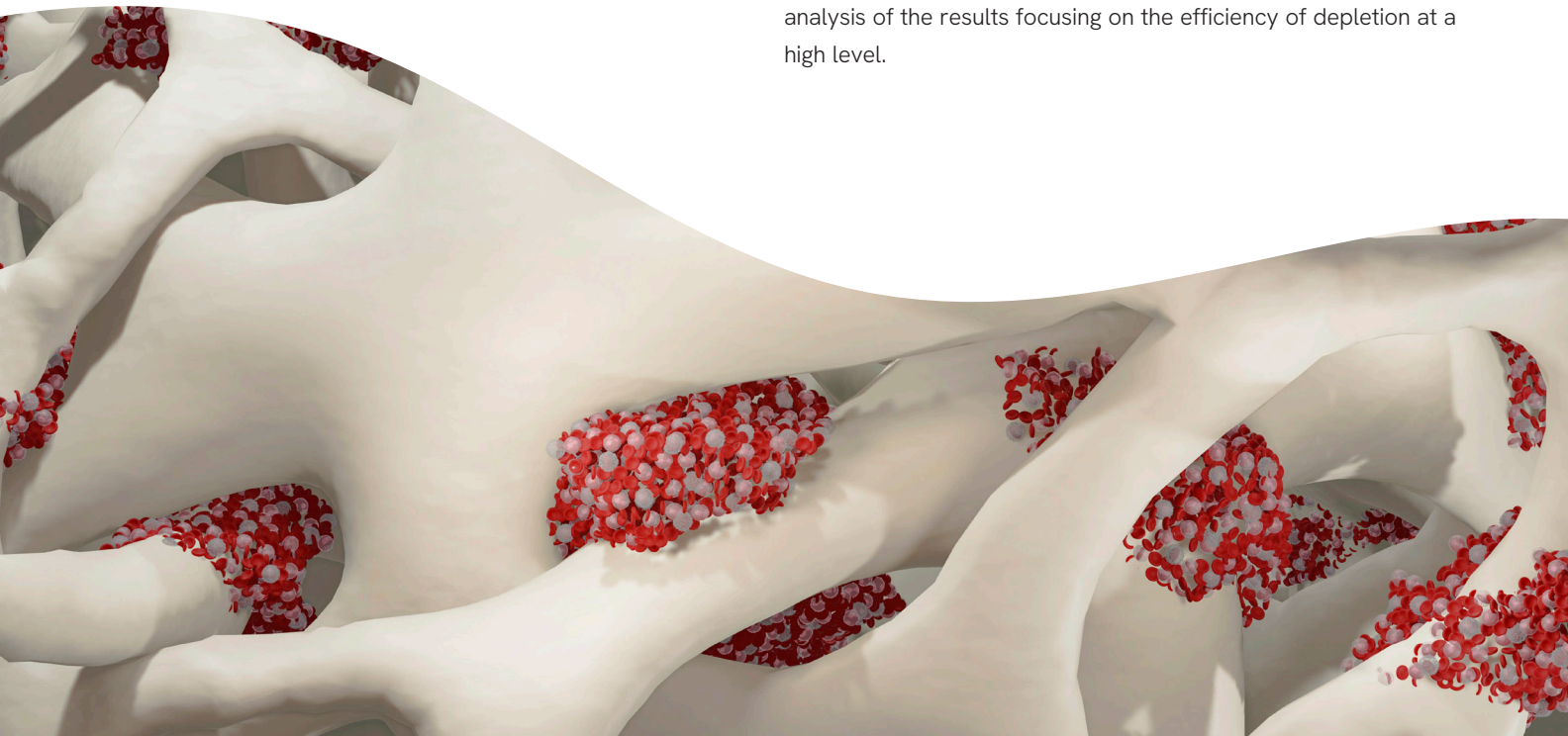
Introduction

Many cancer-driving gene fusions, cryptic splice variants and out-of-frame V(D)J rearrangements escape detection in short-read single-cell RNA-seq[1]. The 3′ and 5′ chemistries from 10x Genomics capture only the ends of transcripts, and typical Illumina® reads seldom span the partner junction.

Recent studies combining 10x Genomics with long-read Oxford Nanopore Technologies (ONT) have reported that 17–22 % of malignant cells in solid tumors express at least one chimeric transcript, uncovering classical oncogenic drivers alongside clone-specific fusions that may act as neoantigens[2]. High-throughput targeted long-read single-cell protocols have also demonstrated improved resolution of lymphocyte clonal landscapes[3], underscoring the value of read length for rare-event discovery.

Despite these advances, detection remains challenging because gene-fusion mRNA is not abundant, and ribosomal and mitochondrial fragments often monopolize pore time. Ribodepletion of rRNA and/or other RNAs that are not relevant should increase probability of detection, but the inputs required for hybridization and RNase H-based depletion methods are too high for single cell applications[4].

To tackle this bottleneck, we performed a pilot experiment inserting a Cas9-based depletion step between 10x Genomics® cDNA amplification and the ONT® ligation protocol and assessed its impact downstream. We present in this note a preliminary analysis of the results focusing on the efficiency of depletion at a high level.



Methods

Two bone marrow samples from cancer patients were collected. They were depleted of erythrocyte progenitors via FACS and processed with the 3′ GEX Next GEM protocol of 10x Genomics.

The two independent 10x runs (3′ v3.1), using bone-marrow mononuclear cells were processed side-by-side. 20 ng of amplified 10x cDNA underwent Cas9-based ribodepletion (NEXTFLEX™ Cas9-gRNA rRNA Depletion Enzyme and NEXTFLEX Cas9-gRNA Mito Depletion Enzyme; Revvity), while 10 ng of amplified 10x cDNA proceeded directly into the ONT ligation V14 kit, according to manufacturer's instructions. Final libraries were pooled and sequenced on an ONT® PromethION® sequencer (Figure 1).

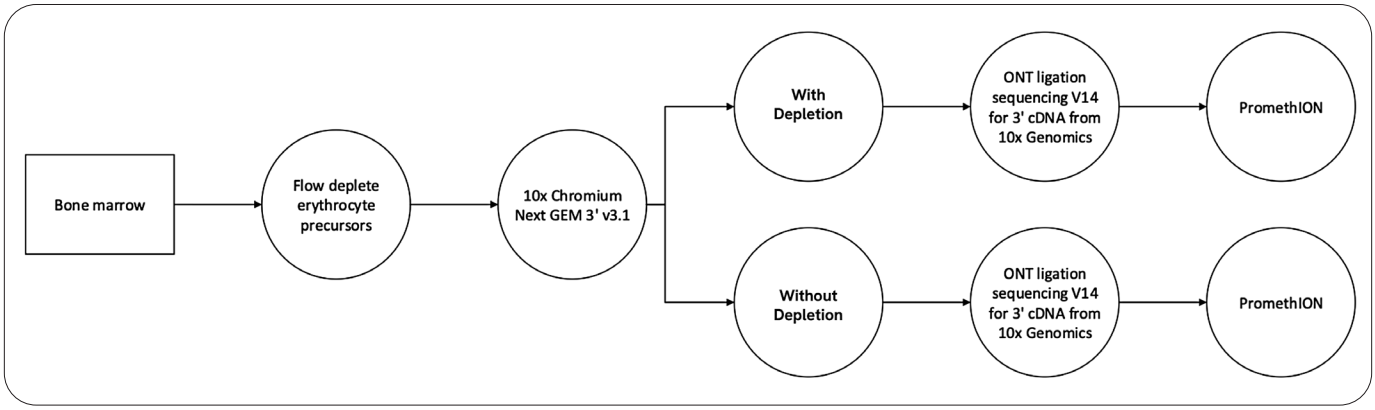


Figure 1: Experimental design used in the pilot study described.

NEXTFLEX Cas9-gRNA rRNA Depletion Enzyme contains a set of optimized guides to remove ribosomal transcripts, whereas the NEXTFLEX Cas9-gRNA rRNA Depletion Enzyme is programmed to remove mitochondrial reads. Both were used in this pilot.

For both cDNA samples, sequencing data from libraries prepped for Illumina sequencing, non depleted were also available. For creating the UMAPs and cell type annotation, cells from the ONT® libraries were filtered based on the following criteria from the Illumina libraries. 1) Cells with low counts,2) cells with high mitochondrial content and 3) putative doublets were filtered out.

Results

Bone marrow samples are listed in the figures as B1 and B2. The number of reads obtained for each of the conditions tested differed (Table 1). Despite an uneven read depth and consequently less transcripts and genes for the non-depleted libraries, a similar number of cells was detected.

Table 1: Reads and number of cells obtained for each of the conditions tested.

Name	Number of reads	Number of cells
B1 no depletion	44,933,035	12,315
B1 with depletion	92,432,704	12,141
B2 no depletion	57,677,533	13,061
B2 with depletion	116,047,720	12,574

First, we looked at nFeature_RNA and nCount_RNA plots. They represent the breadth of the captured transcriptome and the per-cell total UMI counts, respectively, and they are often used as quality check. Too few genes/UMIs per cell hint at empty or damaged droplets, whereas too many flag potential doublets. Even though the non-depleted libraries had a lower sequencing depth and consequently lower UMI and gene counts (Figure 2), this did not affect downstream analyses, as is explained below.

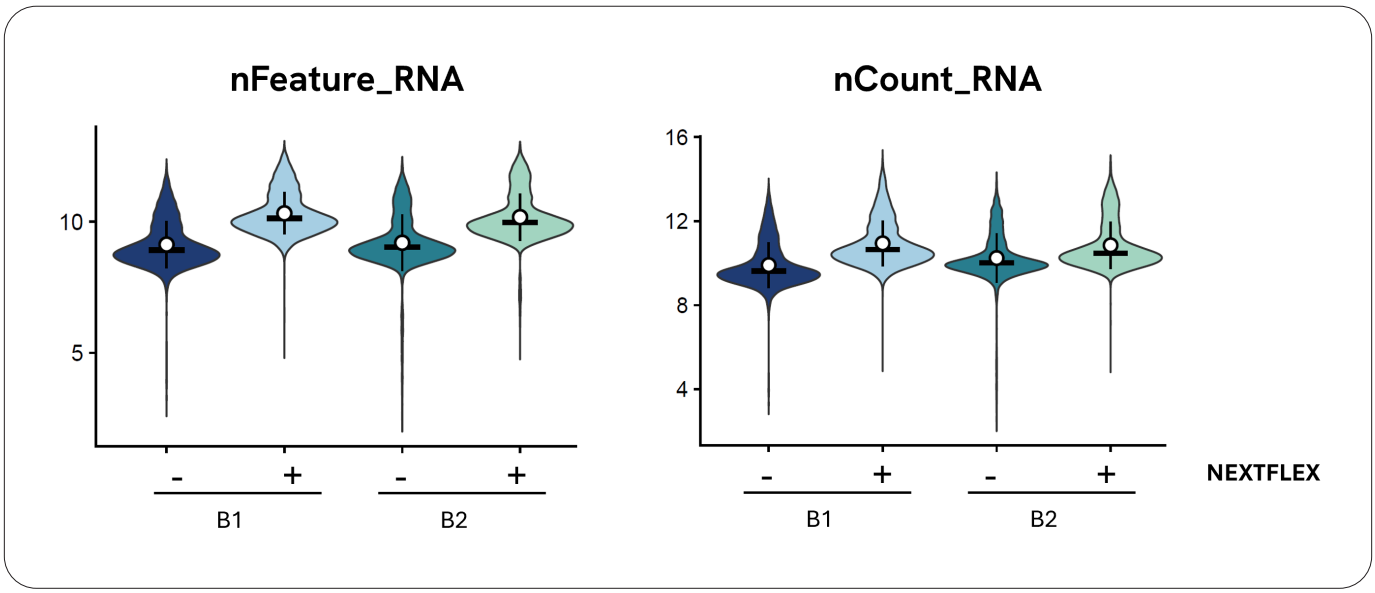


Figure 2: Log2 transformed nFeature and nCount plots for the different conditions tested are presented. Differences observed reflect uneven read depth rather than significant differences across sample pairs. The black horizontal bar represents the median and the white dot with vertical line the mean with standard deviation.

We looked at the percentage of a cell's total UMI counts that map to ribosomal RNA and mitochondrial genes. High values (e.g. > 40–50 %) flags cells whose libraries are dominated by uninformative or stress-related transcripts. Depleted libraries showed a significant reduction in the percentage of uninformative reads, in blood and in bone marrow (Figure 3).

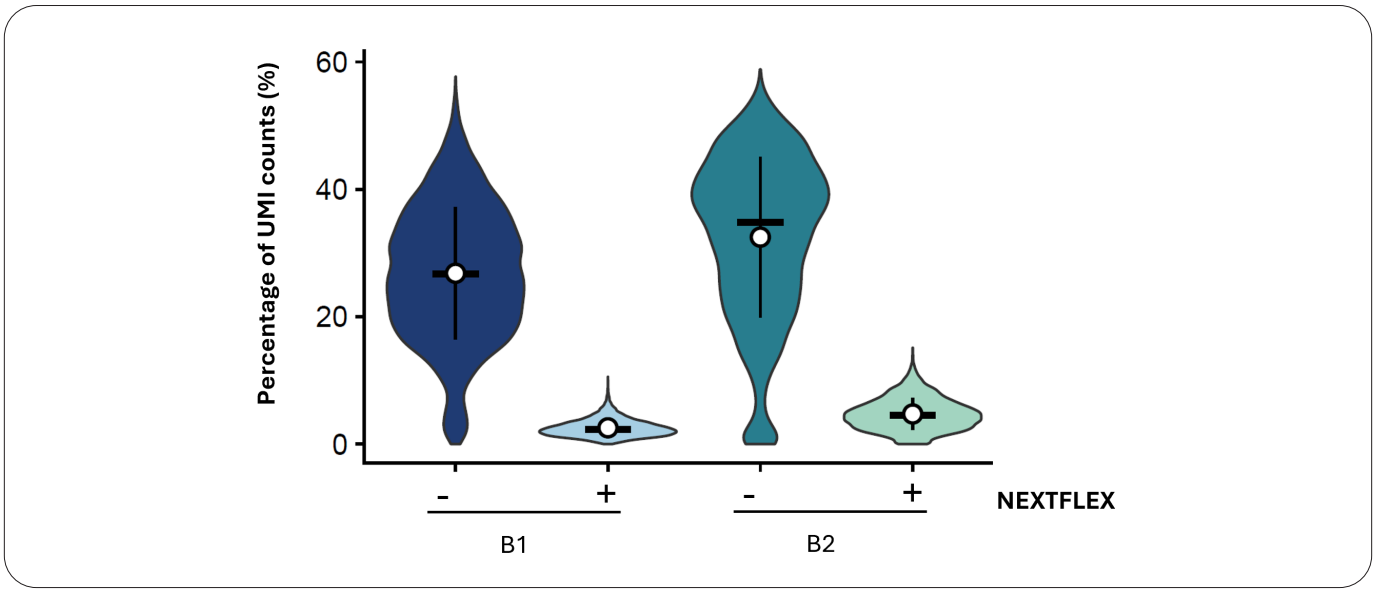


Figure 3: Samples that were treated (+) with the NEXTFLEX depletion enzymes show a lower percentage of UMI counts mapping to ribosomal and mitochondrial transcripts. The black horizontal bar represents the median and the white dot with vertical line the mean with standard deviation.

In single cell transcriptomics sample integration refers to the computational process of combining data from multiple experiments into a unified representation, minimizing technical differences and preserving true biological variation. We integrated the data of all the samples (B1, B2 with and without depletion). We observed that integration removes technical variation due to the NEXTFLEX technology but keeps the sample-specific biological interpretation the same (Figure 4).

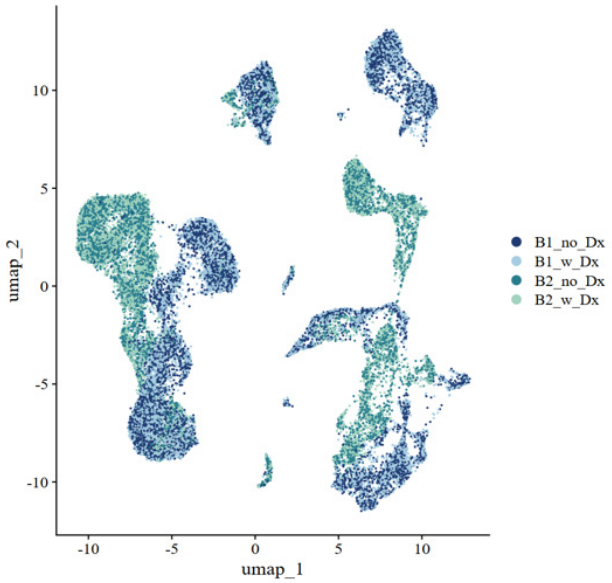


Figure 4: After integration data appears in clear sample-specific clusters populated by both the depleted and non-depleted condition.

For these cDNAs we also created libraries that were sequenced with an Illumina® (non-depleted) platform. When we perform cell typing on the non-integrated data by label transfer from a bone marrow reference dataset (REF) and compare this to the Illumina® sequenced data ("ground truth") we found that the cell typing is highly similar, with over 95% overlap (Figure 5). This shows that the depletion had little effect on cell typing and that reliable data can be obtained using NEXTFLEX depletion.

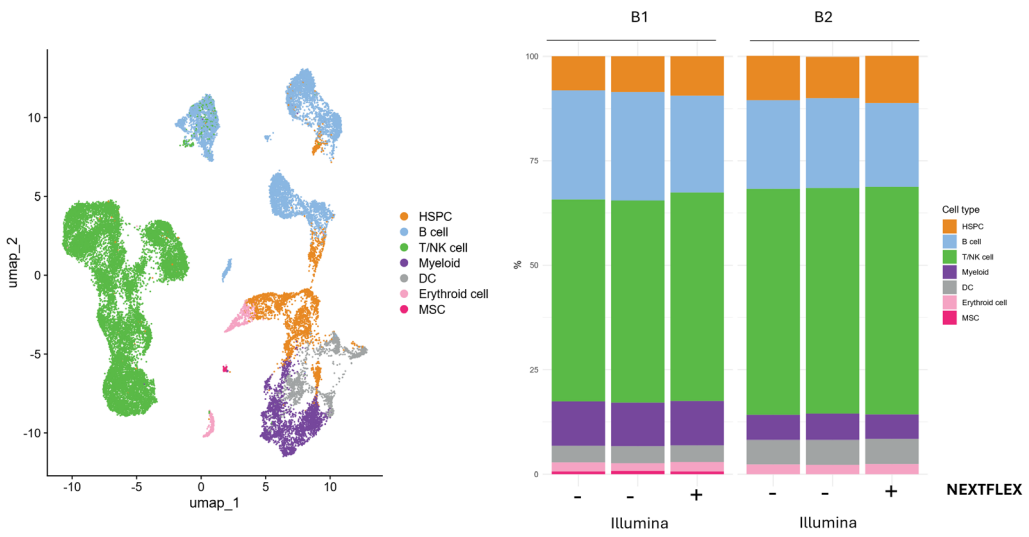


Figure 5: A similar cell typing was observed between the depleted ONT® libraries, non-depleted ONT® data and the Illumina® data, demonstrating that NEXTFLEX technology does not affect cell typing.

Discussion

In the experiments performed we find > 6-fold depletion of the NEXTFLEX depletion enzymes targets on average. Importantly, samples that were NEXTFLEX treated integrated with their own non-treated sample and the NEXTFLEX treatment did not affect cell type annotation. This shows that the addition of a Cas9-based ribodepletion step markedly boosts information density in 10x-ONT single-cell workflows, facilitating the detection of rare transcripts such as those corresponding to gene fusions.

References

1. Tian L, Jabbari JS, Thijssen R, et al. Comprehensive characterization of single-cell full-length isoforms in human and mouse with long-read sequencing. *Genome Biology*. 2021;22:310. doi:10.1186/s13059-021-02525-6.
2. Shiau C-K, Lu L, Kieser R, et al. High-throughput single-cell long-read sequencing analyses of same-cell genotypes and phenotypes in human tumors. *Nature Communications*. 2023;14:4124. doi:10.1038/s41467-023-39813-7
3. Davidson NM, Chen Y, Sadras T, et al. JAFFAL: detecting fusion genes with long-read transcriptome sequencing. *Genome Biology*. 2022;23:10. doi:10.1186/s13059-021-02588-5.
4. Haile S, Corbett RD, O'Neill K, et al. Adaptable and comprehensive approaches for long-read nanopore sequencing of polyadenylated and non-polyadenylated RNAs. *Frontiers in Genetics*. 2024;15:1466338. doi:10.3389/fgene.2024.1466338.