SBX sequencing as a flexible RNA-seq toolkit

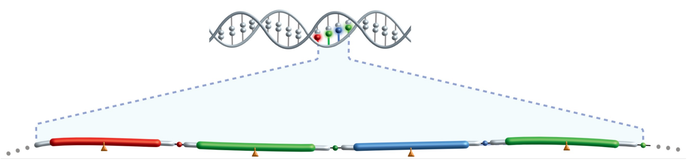
Enabling high throughput single cell transcriptomic profiling at scale

Stephanie Yaung[2], Megan Rogers-Peckham[1], Michael Kartje[2], Hayley M. Bennett[3], Lacey McGee[1], Charlotte Yang[1], Kendall Berg[1], Cynthia Cech[1], Anasha Arryman[1], Ali Carrol[1], Alan Kimura[1], Kevin Pham[1], Meera Balasubramanian[1], Upneet Bala[1], Yui Umezawa[1], Janie Johnson[1], Salka Barrett[1], Joanne Leadbetter[1], Cara Williams[1], Taylor Frese[1], Taylor Lehmann[1], Megan LeProwse[1], Keegan Mcelligott[1], Chen Zhao[2], John Mannion[2], Spyros Darmanis[3], Mark Kokoris[1]

[1]Roche Sequencing Solutions, Inc, Seattle, WA, USA
[2]Roche Sequencing Solutions, Inc. Santa Clara, CA, USA
[3]Genentech, Inc, South San Francisco, CA, USA

Poster #P23.052.B

Introduction to Sequencing By Expansion (SBX)



Sequencing By Expansion (SBX) technology is a novel sequencing approach that uses a biochemical process to encode the sequence of a target nucleic acid molecule into a measurable surrogate polymer called an Xpandomer. Xpandomers encode sequence information into high signal-to-noise reporters, enabling high-fidelity, single-molecule nanopore sequencing using a highly parallel CMOS-based sensor array.

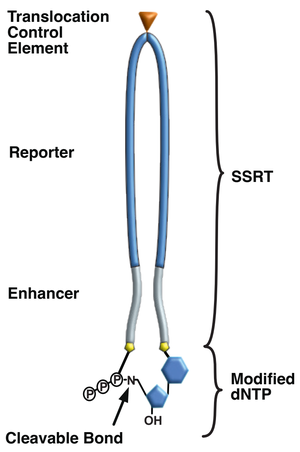


Figure 1. X-NTP Structure

The Expandable Nucleotide Triphosphate (X-NTP) is the building block of the Xpandomer. There are 4 different XNTPs, one for each nucleotide. Highly designed components of the XNTP allow for high signal to noise and non-stochastic translocation control.

The SBX workflow has two components: synthesis and measurement. During synthesis, X-NTPs are used to enzymatically transcribe the template DNA into an Xpandomer molecule.

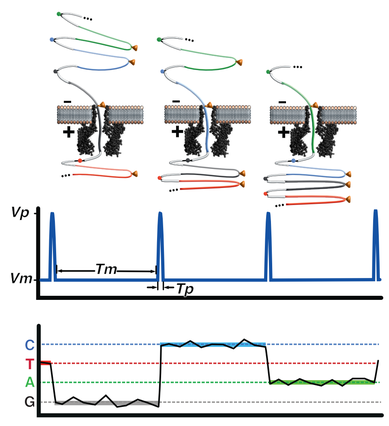


Figure 2. Xpandomer Measurement

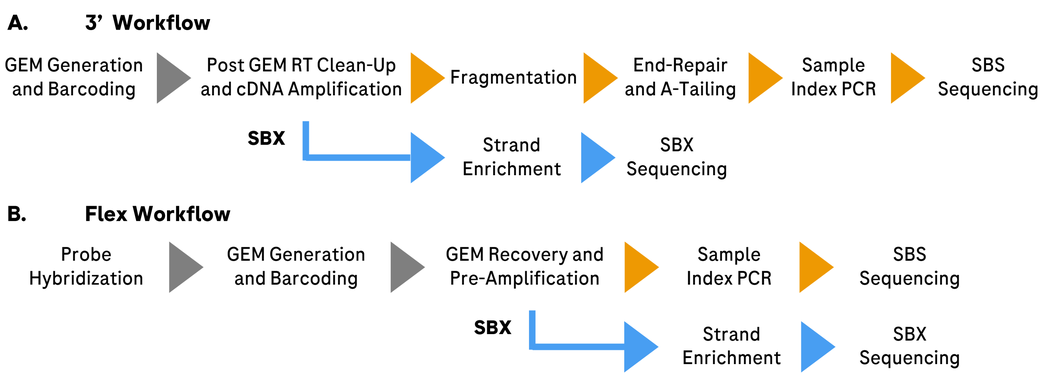
After synthesis, the resulting Xpandomer is sequenced by stepwise translocation through a nanopore.

Specialized components of the structure are designed to maximize capture rate and translocation efficiency through the pore.

Materials & Methods

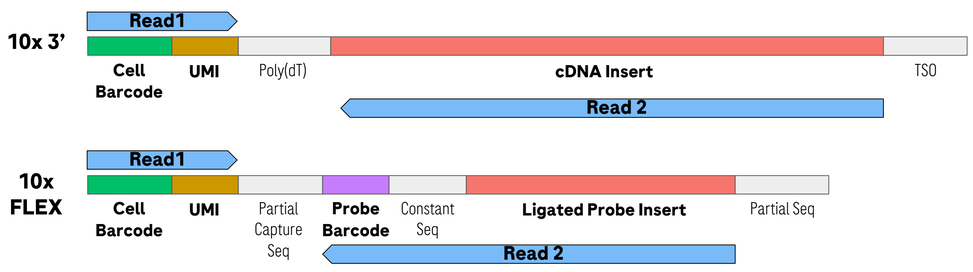
Commercially sourced frozen human peripheral blood mononuclear cells (PBMCs), obtained under Institutional Review Board/Research Ethics Committee (IRB/REC) approval, were used for this study. Before each experiment, cryopreserved aliquots of these cells were thawed and processed with either the NextGEM Flex kit or the NextGEM 3' kit from 10x Genomics. Libraries were created targeting 10k cells and prepared for SBX sequencing using an optimized workflow.

Figure 3. SBX entry into standard 10x Genomics workflows



Sequencing data was pre-processed to mimic paired-end reads expected by traditional bioinformatics pipelines.

Figure 4. SBX read processing



Throughput was assessed using the number of usable reads attained, defined as reads with valid barcodes that confidently mapped to the reference transcriptome or probe set. For Flex data, probes not initially mapped in CellRanger due to indels were realigned.

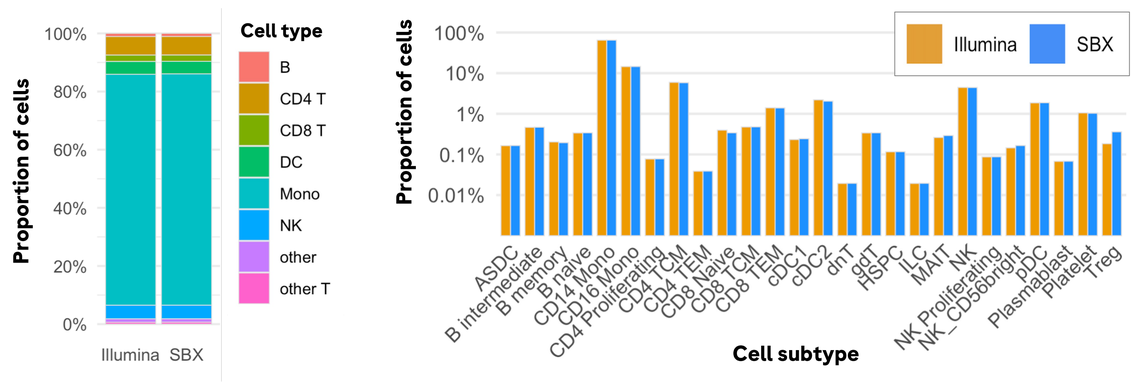
Results - 10x Genomics Flex

For the 10x Genomics Flex kit, we obtained 13B usable reads in one hour of SBX sequencing, which was greater than 30X throughput per hour compared to Illumina. SBX and Illumina sequencing resulted in very similar expression levels and annotated cell types.

CellRanger and throughput metrics	Illumina	SBX
Valid barcodes	98.4%	99.8%
Reads confidently mapped to probe set	96.7%	100%
Cells called	10,347	10,314
Useable reads per run (total reads in molecule_info.h5)	9.5 B (NovaSeqX 10B 2X100)	13.3 B (8M array)
Instrument time per run	22 hours	1 hour
Useable reads in one hour	432 M	13.3 B (30X)

Figure 5. Comparison of cell barcodes and cell types

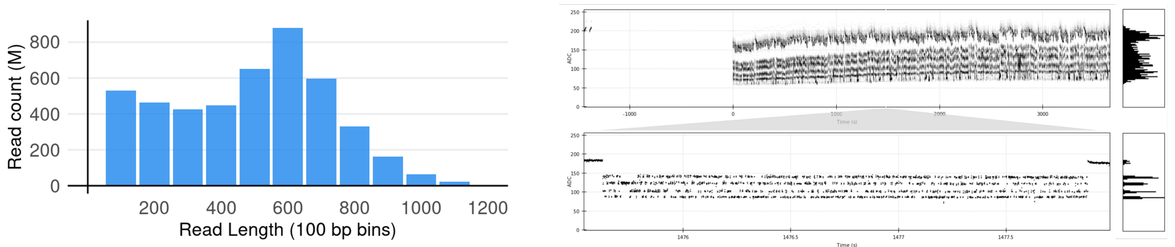
Analysis of the same 10k human PBMC sample sequenced with SBX and Illumina showed 99.2% agreement in detected cell barcodes. For common cell barcodes, transcript abundance was highly correlated (Pearson's r = 0.9996) and cell type assignments were in 99.7% agreement for cell types and 99.1% agreement in cell subtypes. Results shown are for SBX downsampled to the same number of reads as the Illumina run; results for SBX non-downsampled were nearly identical.



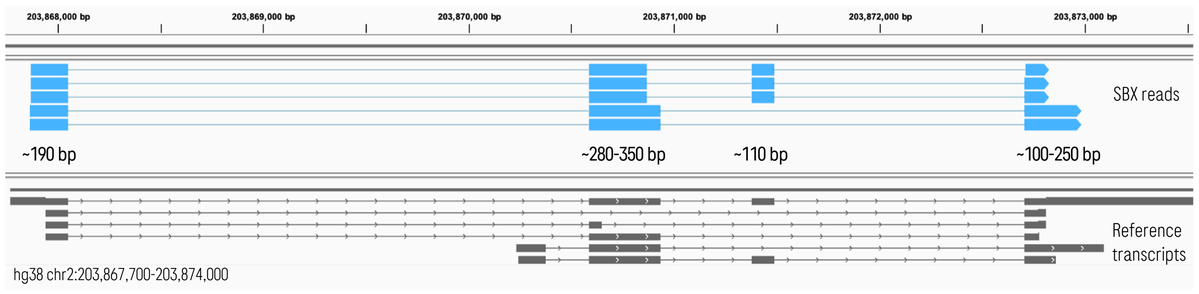
Results - 10x Genomics 3'

For the 10x Genomics 3' kit, we obtained 2.4B usable reads in one hour of SBX sequencing, which was 5.6X throughput per hour compared to an Illumina NovaSeqX run (25B 2X150). SBX read lengths were substantially longer, with 2B reads greater than 600bp. Downstream cell type assignments were highly similar when downsampling SBX data to the same number of reads as Illumina.

Figure 6. Read length and Isoform Detection



Longer SBX reads enable clearer resolution of RNA isoforms, such as for CTLA4.



Conclusions

SBX sequencing demonstrated massively high-throughput across all tested kits, recovering billions of usable reads, whilst maintaining single-cell metrics comparable to those generated with Illumina sequencing. SBX's workflow presents a solution for the increasing need for higher sequencing throughput in single-cell profiling experiments, thus facilitating the resolution of cellular subpopulations and the discovery of novel biological insights. *Please contact lead author at: stephanie.yaung@roche.com*