TECHNICAL NOTE

Genome-wide Perturb-Seq on the NovaSeq™ X 25B flow cell

High-throughput single-cell CRISPR interference screening

In collaboration with

SINGLE CELL DISCOVERIES

illumina®

For Research Use Only. Not for use in diagnostic procedures.

Introduction

Functional genomics screens using CRISPR (clustered regularly interspaced short palindromic repeats) perturbation allow researchers to modulate expression of thousands of genes in a single experiment and study phenotypic effects. The Perturb-Seq method uses single-cell RNA sequencing (scRNA-Seq) for CRISPR-based screens yielding information-rich readouts. From the single-cell transcriptional phenotypes, scientists can dissect complex cellular pathways and develop maps of gene and cellular function. Historically, however, workflow limitations and the high cost of single-cell sequencing have constrained the throughput and scale of these powerful studies.[1]

The sequencing capacity of the NovaSeq X Series 25B flow cells and advances in combinatorial indexing technology[2] significantly alleviate these restrictions, making genome-wide Perturb-Seq projects more accessible and efficient. Higher throughputs can accelerate research timelines. Deeper sequencing of more cells, more timepoints, or additional modalities can offer greater precision and more comprehensive insights.

This application note demonstrates a large-scale CRISPR interference (CRISPRi) experiment to repress gene expression across tens of thousands of genes and examine gene expression patterns in approximately a million cells.

Methods

Protocol overview

The high-throughput Perturb-Seq experiment follows a workflow of CRISPRi screen, single-cell library preparation, sequencing, and data analysis (Figure 1). While multiple single-cell sequencing approaches are compatible with Perturb-Seq, this experiment, performed by a contract research organization (Single Cell Discoveries of Utrecht, Netherlands), used combinatorial indexing kits to maximize cell throughput. A CRISPR droplet sequencing (CROP-Seq) guide RNA (gRNA) vector for the CRISPRi screen was used to introduce polyA tails to ensure compatibility with gRNA capture primers from the kit. Following the CRISPRi screen, cells were fixed and stored at -80°C until library preparation. Two different types of barcoded sequencing libraries were generated: an scRNA-Seq library and a CRISPR library. For projects encompassing 100,000s to 1,000,000s of cells, the NovaSeq X Plus System with 25B flow cells is recommended (Table 1).

Table 1: Sample throughput for single-cell experiments

	NovaSeq X System (single flow cell run)			NovaSeq X Plus System (dual flow cell run)		
Flow cell	1.5B	10B	25B	1.5B	10B	25B
No. of cells per run[a]	80K	500K	1.3M	160K	1M	2.6M
Run time	17 hr	18 hr	48 hr[b]	17 hr	18 hr	48 hr[b]

a. Assumes 20,000 reads per cell.
b. 48-hr run time is for 300-cycle runs. Run time for Perturb-Seq (~130 cycles) was < 24 hr.

Figure 1: Perturb-Seq workflow—After CRISPRi screen, fix cells and prepare single-cell RNA-Seq and CRISPR libraries. Sequence on the NovaSeq X Plus System using 25B flow cells and analyze data using single-cell pipelines.

For Research Use Only. Not for use in diagnostic procedures.

CRISPRi screen

A third-party collaborator performed the whole-genome CRISPRi screen in glioblastoma spheroids. A stable dCas9-expressing glioblastoma cell line (PDM-20) was transduced with a CROP-Seq–style lentiviral gRNA pool consisting of 57,624 gRNAs targeting gene expression regulatory elements of approximately 12,000 protein-coding genes. As a negative control, the gRNA pool included 490 gRNAs targeting nonessential genes and 900 nontargeting gRNAs. In addition, 1168 gRNAs targeting essential genes served as positive controls. After enrichment of successfully transduced cells, cells were differentiated to 3D spheroids and further cultured for six days. At the end of the CRISPRi screen, spheroids were harvested, dissociated to single cells, and fixed in 12 batches. Fixed samples were shipped to Single Cell Discoveries for library preparation, sequencing, and data analysis.

scRNA-Seq and CRISPR library preparation

Single Cell Discoveries used three rounds of combinatorial indexing, an instrument-free, plate-based scRNA-Seq approach (Figure 2A). They prepared scRNA-Seq and CRISPR libraries with the ScaleBio CRISPR Guide Enrichment Kit v1.1 (Scale Biosciences, Catalog no. 955100) and the ScaleBio CRISPR Guide Enrichment Extended Throughput Kit v1.1 (Scale Biosciences, Catalog no. 945058) per manufacturer's instructions. They used six out of 12 fixed cell aliquots for library construction by loading into four 96-well first distribution plates and splitting the final pools across ten 96-well final distribution plates, generating ten final scRNA-Seq libraries with ten matching CRISPR libraries with expected capture of 1.25 million cells.

Sequencing

The recommended sequencing depth is at least 20,000 reads per cell for scRNA-Seq libraries and 2000 reads per cell for CRISPR libraries. Smaller QC scRNA-Seq and CRISPR libraries from a subset of pooled PCR wells can be sequenced first using the NextSeq™ 2000 System to estimate the quality of libraries and captured cells and for adjusting the read depth for the full sequencing depending on desired sequencing saturation. Sequencing of full libraries required 25 billion reads for scRNA-Seq libraries (1.25 million cells × 20,000 reads/cell) and 2.5 billion reads for CRISPR libraries (1.25 million cells × 2000 reads/cell).

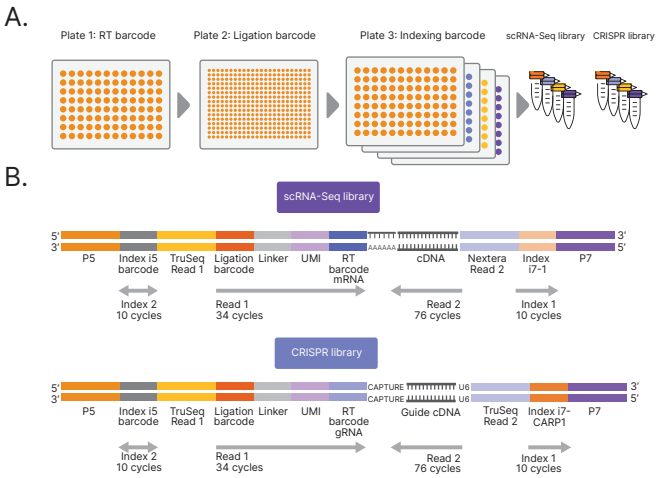

Figure 2: scRNA-Seq and CRISPR library generation—(A) Combinatorial indexing protocol overview and (B) final scRNA-Seq and CRISPR library structures.

Each final scRNA-Seq library was mixed with its matching CRISPR library and sequenced on a separate lane of a 25B flow cell on the NovaSeq X Plus System. To ensure high-quality sequencing, the PhiX spike-in for sequencing such libraries should be 5%–10%. The higher percentage of PhiX spike-in is required due to an invariable linker present between the first and second barcode in Read 1. Libraries were paired-end sequenced following the read configuration: R1: 34, i7: 10, i5: 10, R2: 76 (Figure 2B and Table 2).

Data analysis and visualization

After sequencing, Single Cell Discoveries used the following workflow to process the sequencing data. First, demultiplexed FASTQ files of scRNA-Seq libraries were processed with the designated Nextflow-based ScaleRNA pipeline. The latest version of this pipeline (v1.5.0-beta2) that was used for analyzing the data set includes an updated CellFinder cell-calling algorithm to improve the calling of true barcodes (ie, cells) from the spurious barcodes background. This results in a higher number of true cells recovered compared to the standard cell calling method based on the identified inflection point in barcode rank plot. The demultiplexed FASTQ files of CRISPR libraries were processed with a separate Nextflow-based ScaleCRISPR pipeline. Seurat v5 tool2 is recommended for downstream data analysis and visualizations.[3]

Table 2: Recommended read configuration for scSRNA-Seq and CRISPR libraries

	scRNA-Seq libraries				CRISPR libraries			
	Read 1	i7 index	i5 index	Read 2	Read 1	i7 index	i5 index	Read 2
Purpose	Ligation barcode, UMI, and RT barcode	Library/ plate index	Cell barcode	cDNA insert	Ligation barcode, UMI, and RT barcode	Library/ plate index	Cell barcode	gRNA
Length	34 bp	10 bp	10 bp	76 bp	34 bp	10 bp	10 bp	76 bp

Seurat v5 has been specifically developed to support analysis of data sets spanning millions of cells relying on a new 'sketch'-based approach with minimal memory requirements.

Results

The NovaSeq X 25B flow cells make high-throughput, large-scale Perturb-Seq experiments more accessible and efficient. For this study, sequencing performance met target specifications, delivering billions of high-quality reads across ~1M cells (Table 3). Different libraries loaded on different lanes demonstrate lane-to-lane consistency and comparable detection of genes (Figure 3).

Table 3: Sequencing metrics for Perturb-Seq on NovaSeq X 25B flow cells

Output	25.7 billion read pairs (before filtering)
	~23 billion read pairs (after filtering)
No. of recovered cells	1.2 million cells (before filtering)
	832,432 high-quality cells (after filtering)
Median transcripts per cell	7794 transcripts
Median genes per cell	4123 genes
Average reads per cell	~27,000 reads
Reads mapped to genome	> 90%

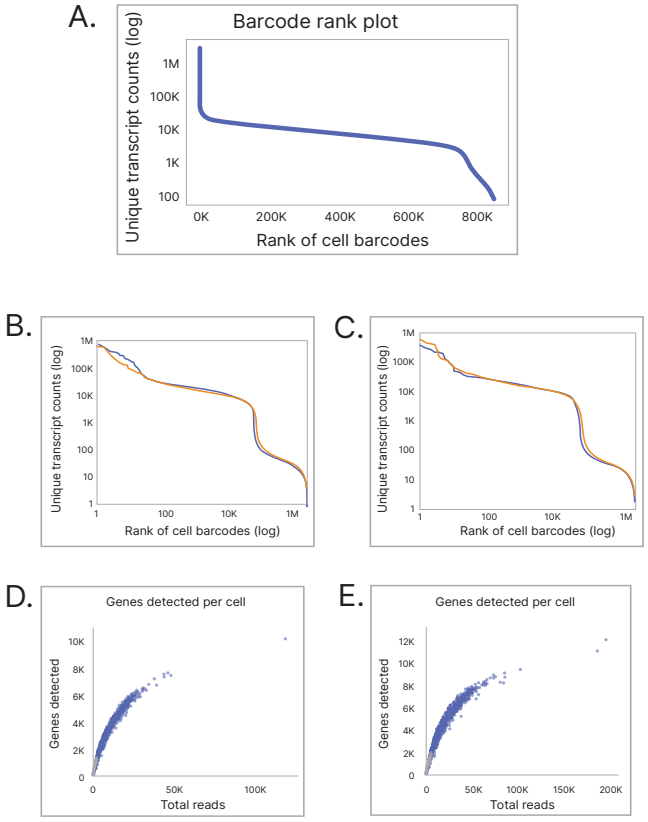

Figure 3: Sequencing metrics for Perturb-Seq on NovaSeq X 25B flow cells—(A) Plot of cell barcodes ranked by number of unique transcripts. (B, C) Different libraries on different lanes show comparable data quality. Barcode rank plots for (B) library 1 on lane 1 and (C) library 2 on lane 2, each with two sample batches (orange and blue). (D, E) Comparison between genes detected per cell vs read depth across two different libraries: (D) library 1 on lane 1 and (E) library 2 on lane 2.

For Research Use Only. Not for use in diagnostic procedures.

Resulting scRNA-seq data can inform deep characterization of cellular phenotypes. For example, for initial data exploration, Uniform manifold approximation and projection (UMAP) plotting can help visualize clustering of cells based on similarities or differences in their transcriptomes. For this data set, after mapping and cell calling, Single Cell Discoveries identified 832,432 cells that formed nine separate clusters (0-8) based on their transcriptomic profiles (Figure 4A). Other recorded information about the samples can be projected on the UMAP to check for possible confounding variables that might drive cell clustering. Here, as an example, cells were labeled based on their fixation batch origin to showcase that fixation batches do not contribute to differences in gene expression (Figure 4B).

In addition, a cell's CRISPR perturbation status can be linked to its transcriptome and visualized in the scRNA-Seq UMAP space. Cells can be labeled based on their detected gRNAs identity to explore if a particular predefined gRNA group drives the formation of a specific cluster. Here, four gRNA groups were used to visualize in UMAP depending on whether the gRNAs were nontargeting or target glioblastoma nonessential genes, essential genes or genes with an unknown effect on glioblastoma proliferation/survival (Figure 5).

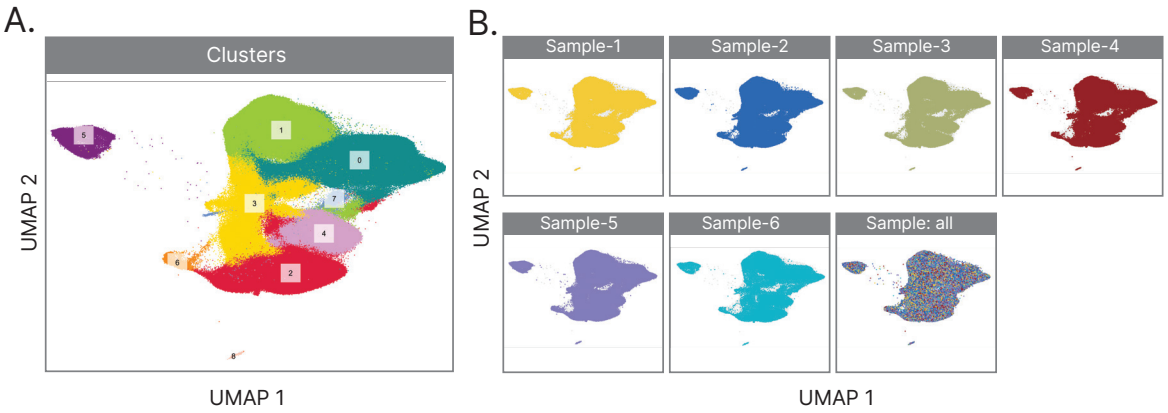

Figure 4: Single cell RNA expression UMAP of filtered final 832,432 cells—(A) Cells were clustered based on the similarity of their gene expression profiles resulting in identification of nine distinct cell clusters. (B) Fixation batches do not contribute to gene expression differences between cells. Fixation sample identity was projected on top of the RNA expression UMAP showing that cells from different fixation batches contribute to all identified clusters.

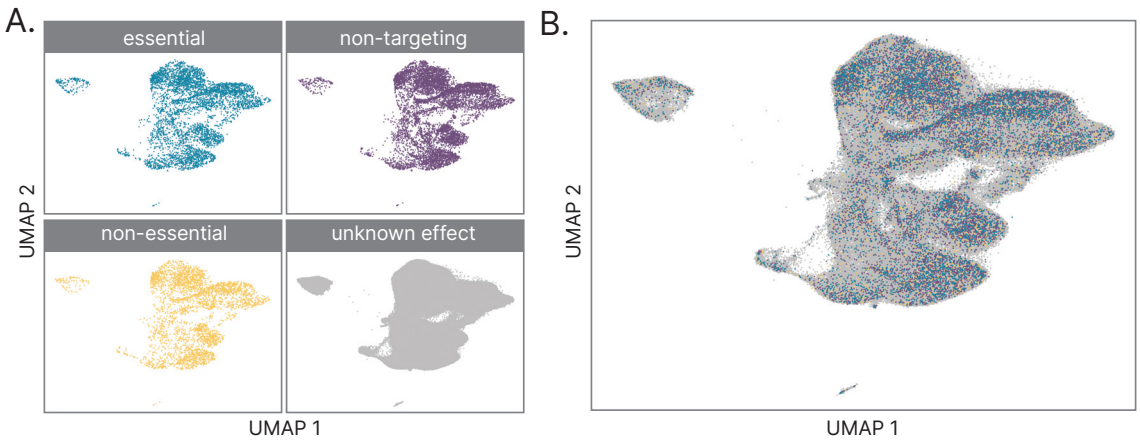

Figure 5: Connecting cell CRISPRi perturbation status to its gene expression—(A) UMAP projection of single-cell RNA expression data colored by the gRNA targeting group (nontargeting, nonessential, essential, and unknown effect). (B) Merged UMAP of all gRNA targeting groups showing that gRNAs from different targeting groups are present in all identified cell clusters.

Summary

The NovaSeq X 25B flow cells enable high-throughput, genome-scale Perturb-Seq experiments. The insights gained from CRISPR-based single-cell functional genomics screens will be enhanced by deeper sequencing and measuring more modalities of more cells at more time points. Dual flow cell capability on the NovaSeq X Plus System can further improve efficiency and extend the discovery power of Perturb-Seq studies.

Learn more

NovaSeq X and NovaSeq X Plus Sequencing Systems

High-accuracy next-generation sequencing with the NovaSeq X Series technical note

Single Cell Discoveries

References

1. Replogle JM, Saunders RA, Pogson AN, et al. Mapping information-rich genotype-phenotype landscapes with genome-scale Perturb-seq. *Cell*. 2022;185(14):2559-2575.e28. doi:10.1016/j.cell.2022.05.013
2. Cao J, Spielmann M, Qiu X, et al. The single-cell transcriptional landscape of mammalian organogenesis. *Nature*. 2019;566(7745):496-502. doi:10.1038/s41586-019-0969-x
3. Hao Y, Stuart T, Kowalski MH, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. 2024;42(2):293-304. doi:10.1038/s41587-023-01767-y

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

© 2024 Illumina, Inc. All rights reserved. All trademarks are the property of Illumina, Inc. or their respective owners.
For specific trademark information, see www.illumina.com/company/legal.html.
M-GL-02890 v1.0