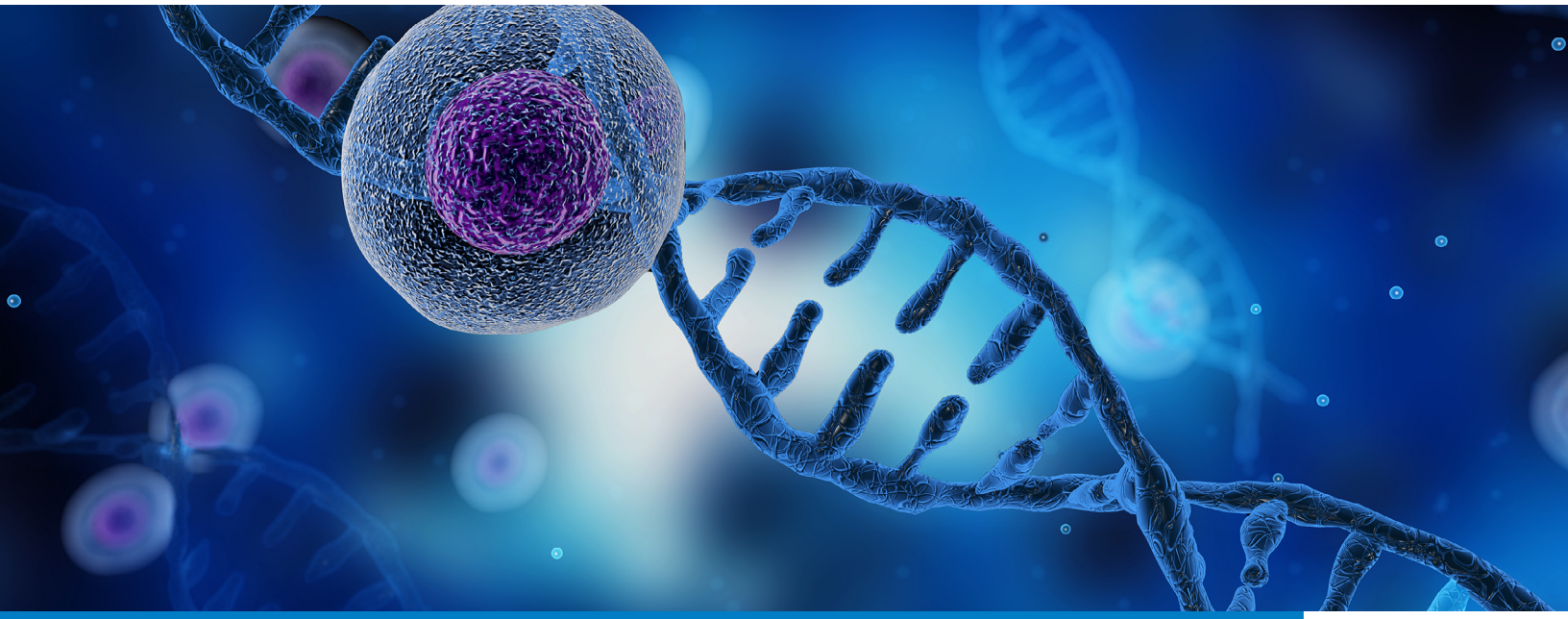
Case study

A novel high-throughput method for single-cell alternative splicing analysis



Highlights:

- Identified gene expression changes, differentially spliced genes, and alternative splicing events in a *DUX4* knockdown cell model provided by the Fred Hutchinson Cancer Center
- Overcame the limitations of droplet-based methods for the discovery of mRNA splice variants with full-length single-cell transcriptome library prep
- Detected on average 9,000 genes per individual cell with high-th roughput, high-sensitivity single-cell RNA-seq for downstream gene expression and alternative splicing analyses

Introduction

Alternative splicing of pre-mRNA allows multiple proteins to be expressed from a single gene. Studying alternative splicing at single-cell resolution is essential for uncovering cellular diversity, understanding disease mechanisms, tracking developmental changes, and discovering novel splice variants. The detection of splice variants is typically performed using RNA-sequencing approaches; however, the use of end-counting technologies, which rarely span splice junctions and only capture the 3′ end or 5′ end of transcripts, limits the discovery of novel variants (Jiang and Chen 2021). Full-length single-cell RNA-seq (scRNA-seq) surpasses these limitations by providing full gene-body coverage, greatly increasing the detection of splice junctions and identification of isoforms. Throughput is an important consideration; performing full-length scRNA-seq techniques at high throughput improves the identification and quantification of rare or low-abundance splice variants that may be missed in lower-throughput analyses.

In this study, we used a combination of high-throughput full-length scRNA-seq and an alternative splicing analysis pipeline to demonstrate the simultaneous detection of splicing events and differential gene expression in a cell model of double homeobox protein 4 (*DUX4*) knockdown. In this novel method, treated cells were first processed

with the Takara Bio SMART-Seq® Pro Application Kit (SSPro), an end-to-end solution for full-length RNA-seq library prep and integrated data analysis of single-cell transcriptomes. Then, scRNA-seq data processed by Takara Bio Cogent™ NGS tools were used as input for the MARVEL pipeline, which generated splicing and gene expression information. The detection of *DUX4*-specific splicing events reveals this approach as a simple yet powerful discovery tool.

Results

Exceptional sensitivity in high-throughput full-length scRNA-seq

To analyze *DUX4* expression-associated alternative splicing events, the human G401 cell line was treated with doxorubicin to induce *DUX4* expression. Then, endogenous *DUX4* mRNA was knocked down with antisense gapmers in these cells ("G401_DUX4_KD"). Wild-type control cells ("G401_CNTRL_KD") were treated with doxorubicin and control gapmers. Samples were kindly provided by the Stephen Tapscott lab at the Fred Hutchinson Cancer Center. These cells were dispensed onto chips using the ICELL8® cx Single-Cell System, generating 600+ single-cell candidates for each cell type. Full-length RNA-seq libraries were prepared using the SMART-Seq Pro Application Kit and sequenced on the Illumina® NextSeq® 550 system (2 x 75 bp reads) with a high-output kit (3.7 x 10^5 reads/barcode). Sequencing data were processed using the Cogent NGS Analysis Pipeline (CogentAP).

Using Cogent NGS Discovery Software (CogentDS) to visualize the output files, a median of 8,705 and 8,933 genes per single cell were identified in the G401_CNTRL_KD cells and G401_DUX4_KD cells, respectively (see **Figure 1** and **Table 1**).

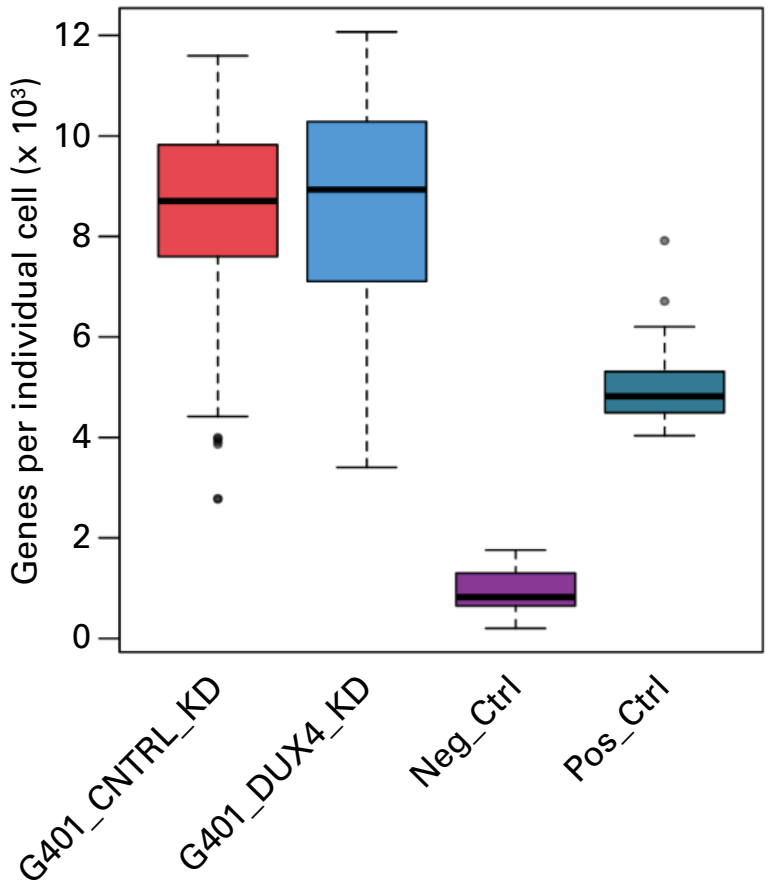


Figure 1. Sensitivity of SSPro demonstrated in a human G401 cell line. Box-and-whisker plots illustrate the total number of genes identified in control (G401_CNTRL_KD) and experimental (G401_DUX4_KD) cells, plus a positive control (K562 RNA provided with the SSPro kit) and a negative control (no cells). Full-length scRNA-seq data were processed using CogentAP and visualized using CogentDS. A total of 642 and 621 cells were sequenced from G401_CNTRL_KD and G401_DUX4_KD cultures, respectively.

	Average # genes per cell	Median # genes per cell
G401_CNTRL_KD	8,540	8,705
RPK 10,000 >0.1	3,709	3,727
RPK 10,000 >1.0	574	586
G401_DUX4_KD	8,628	8,933
RPK 10,000 >0.1	3,866	3,864
RPK 10,000 >1.0	567	578
Neg_Ctrl	951	823
RPK 10,000 >0.1	951	823
RPK 10,000 >1.0	643	677
Pos_Ctrl	5,009	4,818
RPK 10,000 >0.1	2,937	2,900
RPK 10,000 >1.0	684	686

Table 1. Mean and median gene counts detected using SSPro.

Analysis of differentially expressed genes

scRNA-seq data were next analyzed to characterize the impact of *DUX4* downregulation on gene expression changes. Using the R package MARVEL (Wen, Mead, and Thongjuea 2023) to analyze gene expression matrix files generated by Cogent software, a total of 1,143 upregulated and 892 downregulated genes were identified in G401_DUX4_KD cells compared to G401_CNTRL_KD cells. Genes with a greater than fivefold change in expression are illustrated in **Figure 2**.

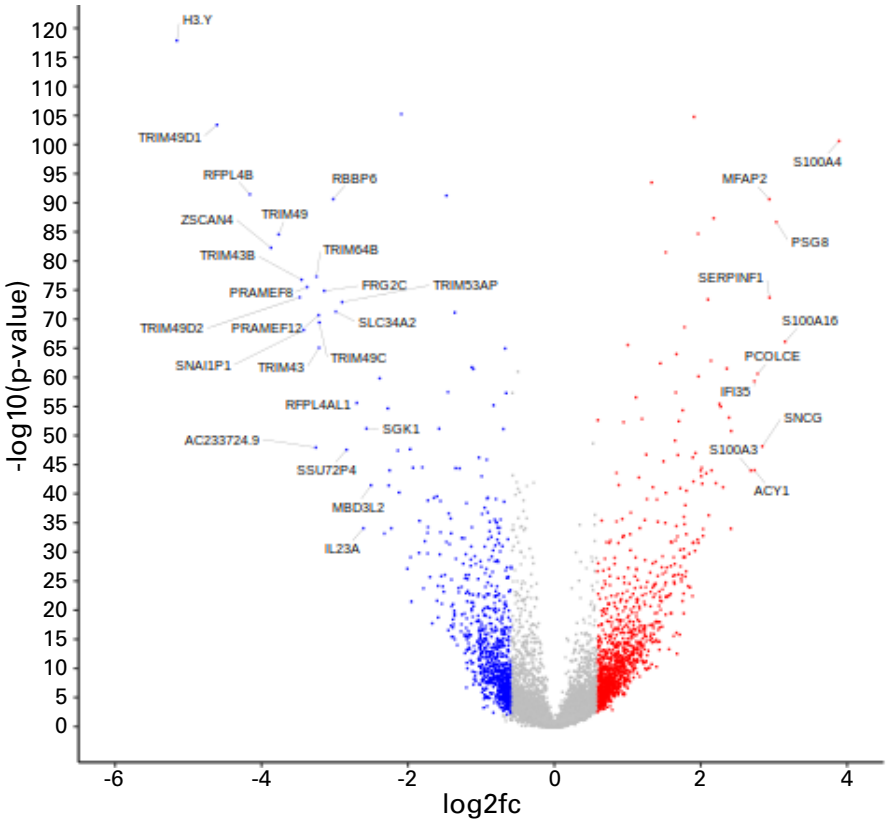


Figure 2. Characterization of *DUX4* knockdown-associated differential gene expression. Volcano plot illustrates differentially expressed genes (blue = downregulated genes; red = upregulated genes) in G401_DUX4_KD cells compared to G401_CNTRL_KD cells (log2FC threshold >0.5; adjusted P value <0.01). All genes for which log2FC exceeded 2.5 are labeled in the plot.

Identification of alternative splicing events

There are seven splicing patterns (Liang et al. 2021), which together contribute to transcript isoforms for 95% of human genes: skipped exons (SE), mutually exclusive exons (MXE), retained intron (RI), alternative 5′ splice site (A5SS), alternative 3′ splice site (A3SS), alternative first exon (AFE), and alternative last exon (ALE) (**Figure 3, Panel A**). SSPro detected 6 out of 7 different alternative splicing patterns (**Figure 3, Panel B**) across 192 differentially spliced genes (**Figure 3, Panel C**).

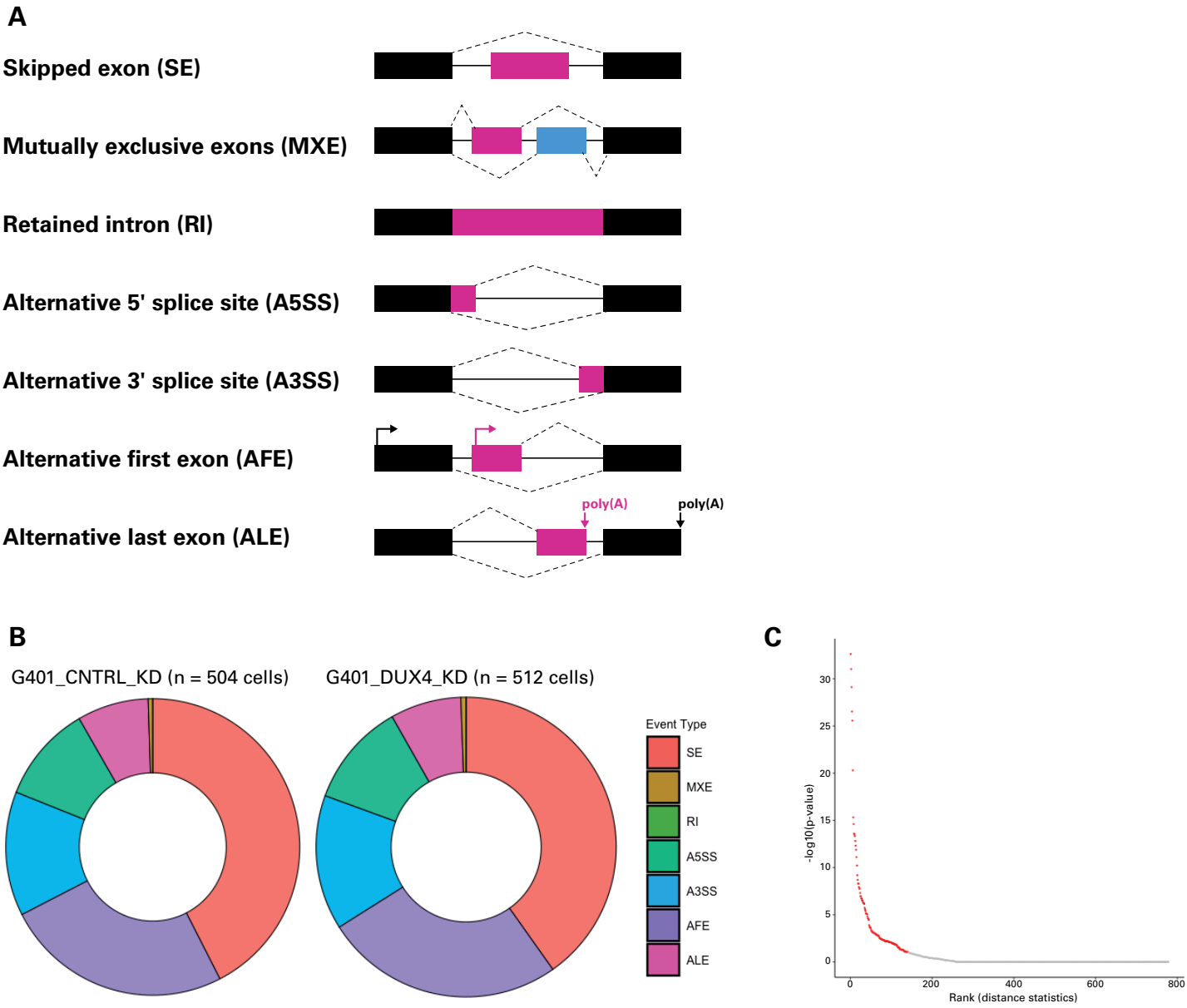


Figure 3. Detection and characterization of alternative splicing events across cell lines. Seven different alternative splicing patterns (**Panel A**) were analyzed in SSPro data, with skipped exons being the most common event across both G401_CTRL_KD and G401_DUX4_KD cell lines (**Panel B**). Differential splicing analysis of genes detected in at least 50 individual cells across both cell lines revealed 192 differentially spliced genes (**Panel C**; FDR <0.10, red dots). This splicing distance plot, a ranked plot described by MARVEL, is based on distance statistics (Anderson-Darling method). The Figure in **Panel A** was adapted from MARVEL documentation, available at https://wenweixiong.github.io/MARVEL_Plate.html

Using MARVEL (see **Figure 4**) to analyze all seven splicing patterns, 2,398 and 1,781 total alternative splicing events were identified in G401_CTRL_KD and G401_DUX4_KD cells, respectively (**Table 2**).

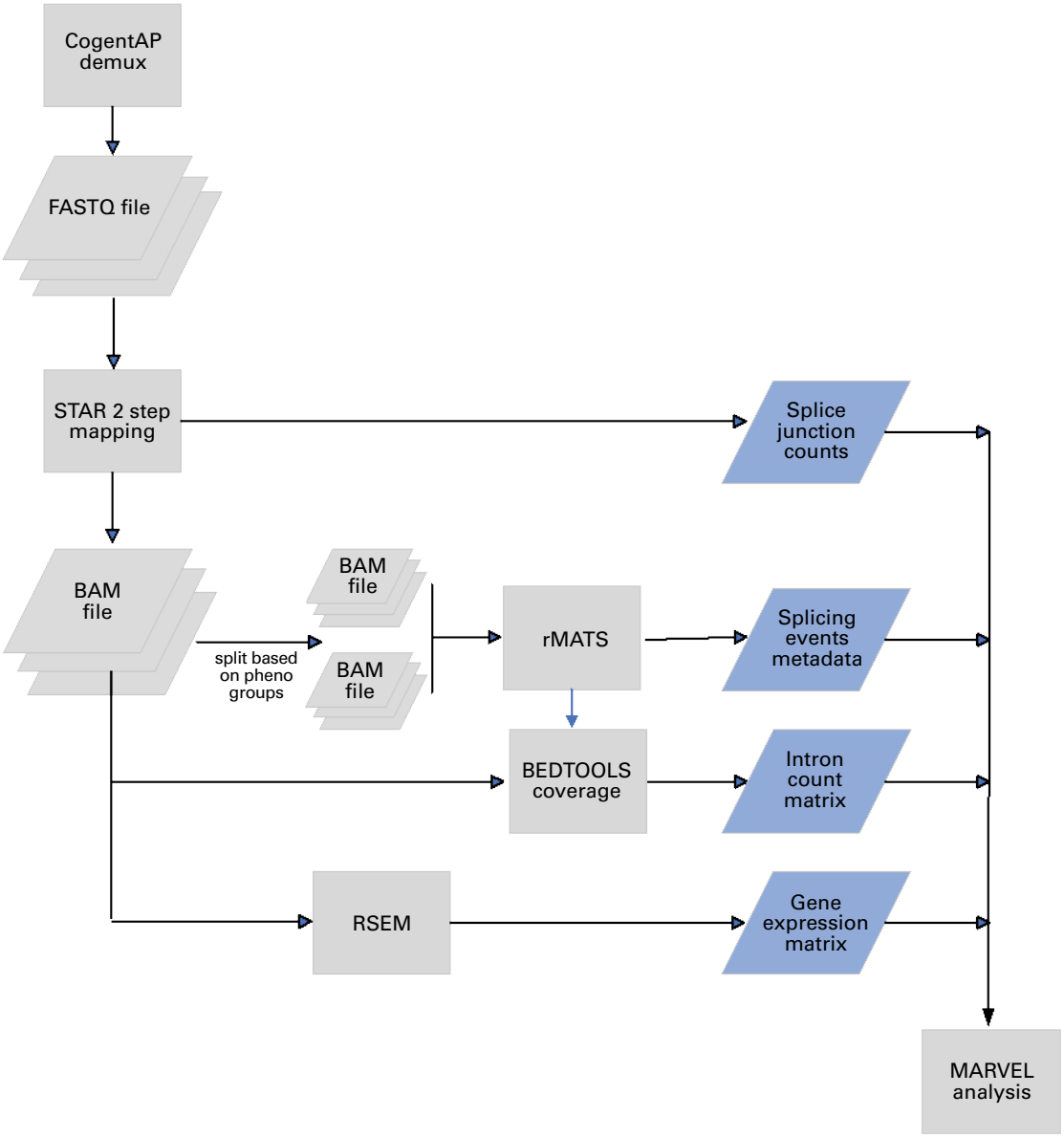


Figure 4. Alternative splicing analysis workflow. Flowchart illustrating the process for analyzing alternative splicing events from SSPro data using CogentAP and the MARVEL R package. First, sequences were demultiplexed and barcodes assigned (one barcode per individual cell) using CogentAP. Reads were trimmed using Trim Galore 0.6.5 and aligned against the reference genome using STAR in 2-pass mode. Output files included the gene expression matrix, splice junction counts, and splicing events metadata generated by rMATS; these files were used as inputs for alternative splicing analysis with MARVEL.

AS events identified in each group	Average # of AS events
G401_CNTRL_KD only	2,398
G401_DUX4_KD only	1,781
Both cell lines	1,743

Table 2. Alternative splicing (AS) events detected across cell lines.

Conclusions

Splicing isoform analysis performed using the SSPro/MARVEL method revealed gene expression patterns, differentially spliced genes, and alternative splicing events in control and G401 *DUX4* KD cells, suggesting a promising pool of candidates from which to identify novel *DUX4* expression-related biomarkers.

References

Jiang, W. and Chen, L. Alternative splicing: Human disease and quantitative analysis from high-throughput sequencing. *Comput. Struct. Biotechnol. J.* **19**, 183–195 (2021).

Liang, J. et al. The Alternatively Spliced Isoforms of Key Molecules in the cGAS-STING Signaling Pathway. *Front. Immunol.* **12**, 771744 (2021).

Wen, W. X., Mead, A. J. & Thongjuea, S. MARVEL: an integrated alternative splicing analysis platform for single-cell RNA sequencing data. *Nucleic Acids Res.* **51**, e29–e29 (2023).

PRODUCTS

Product	Package Size	Cat. #
SMART-Seq Pro Application Kit - 2 Chip	2 Chips	640257

Takara Bio USA, Inc. provides kits, reagents, instruments, and services that help researchers explore questions about gene discovery, regulation, and function. Our mission is to develop high-quality innovative tools and services to accelerate discovery.

Takara Bio USA, Inc.
United States/Canada: +1.800.662.2566 • Asia Pacific: +1.650.919.7300 • Europe: +33.(0)1.3904.6880 • Japan: +81.(0)77.565.6999

10.24 US 634070