Using 3’ Single Cell Sequencing to Study Patient-Derived Tumor Cells Reveals Enhanced Potential to Study Differential Splicing at the Cell Type Level

Sergey Nemzer[1], Zohar Shipony[2]* , Assaf Wool[1], Loai Naom[1], Gila Lithwick Yanai[2], Eti Meiri[2], Masha Frenkel[1], Zoya Alteber[1], Zurit Levine[1], Eran Ophir[1], Sarah Pollock[2], Roy Granit[1]

[1]Computational Discovery, Compugen LTD, Holon, Israel; [2]Ultima Genomics Inc, Fremont, CA; *Presenting author - Data is presented by Uiltima Genomics on behalf of Compugen LTD

compugen FROM CODE TO CURE®

ULTIMA GENOMICS

Poster #224

Introduction

The UG 100™ sequencing technology stands out for its ability to deliver high-throughput sequencing at a significantly reduced cost and has specific technical differences compared to traditional platforms. We conducted a comparative study involving seven tumor samples from cancer patients to assess the qualitative differences between sequencing technologies for single-cell RNA sequencing (scRNA-seq). The samples were initially processed with 10x Genomics Chromium scRNA-seq, with libraries sequenced on either the Illumina NovaSeq 6000 or converted for sequencing on the Ultima Genomics UG 100™ platform. Our analysis found high consistency in single-cell expression levels across the two platforms (Pearson Correlation, R = 0.98). However, there is a key difference in read alignment, since Illumina reads are paired-end, while UG 100™ reads are single-ended: UG 100™ reads consistently initiated from the 3’ end of mRNA transcripts, whereas reads from Illumina libraries did not consistently reach the 3' ends. This unique feature of UG 100™ enables clearer demarcation of transcript termination. Leveraging this distinct characteristic, we developed a bioinformatic workflow that harnesses the UG 100™ data to detect alternative splicing events and poly-adenylation alterations in thousands of protein-coding genes. This workflow will enable us to enrich our Unigen™ platform with additional information relating to gene splicing and advance our understanding of cancer immunotherapy targets.

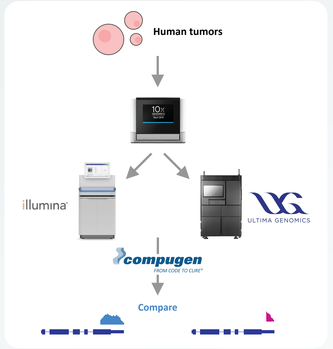
Notably, using UG 100™ data, we identified a truncated variant of CARD8—a core component of the inflammasome complex—suggesting a potential loss of functional activity. Further scRNA-seq analysis revealed that this truncated CARD8 isoform was reduced in macrophages, cells which are typically associated with high inflammasome activity. These findings point to a potential mechanism for regulating inflammasome function via alternative splicing.

Comparative analysis of sequencing platforms

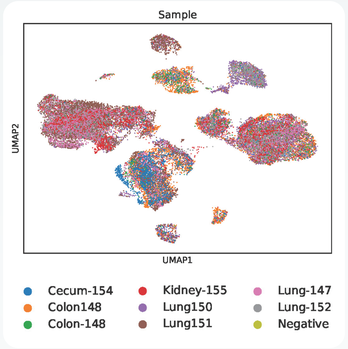
A.

Patient ID	Tumor type	Stage	CD45 enrichment	Hashing
147	Lung	I	+	+
148	Colon	NA	+	+
152	Lung	NA	+	+
154	Cecum	IIB	+	+
155	Kidney	NA	+	+
148	Colon	NA	+	-
150	Lung	IIB	+	-
151	Lung	IIB	+	-

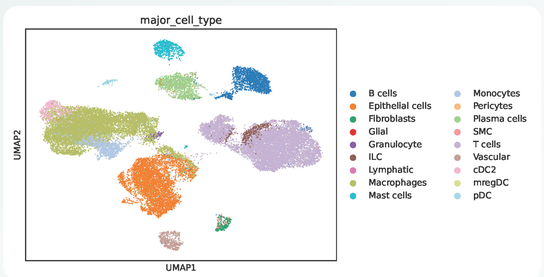
B.



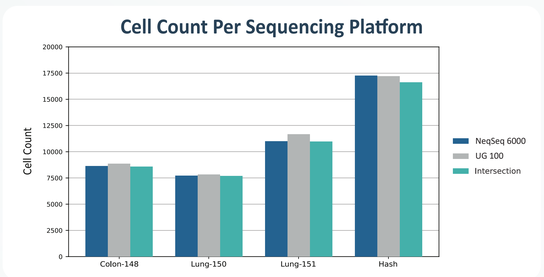
C.



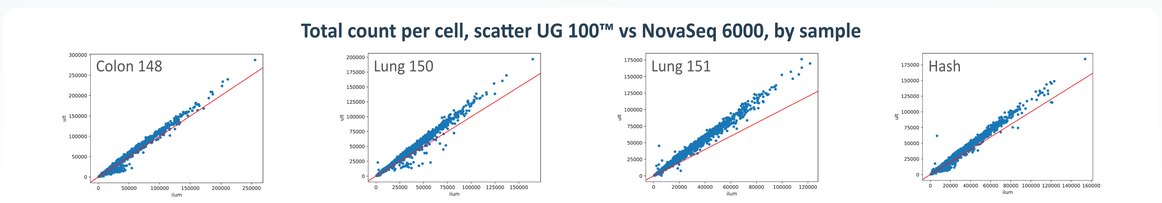
D.



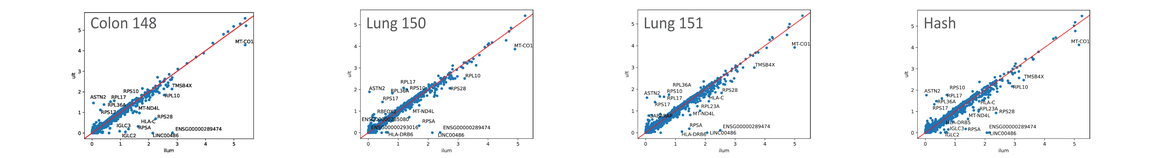
E.



F.



G.

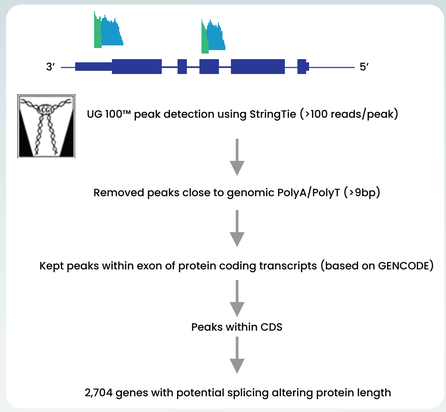


Multi-step workflow for single-cell analysis of human tumors comparing Illumina and Ultima Sequencing Platforms. A. Table summarizing sample metadata. Each row represents a patient tumor sample, detailing the tumor type, stage (if available), CD45+ immune cell enrichment status, and whether hashing technology was applied for sample multiplexing. **B.** Experimental and analytical pipeline for single-cell sequencing. Tumor samples were processed for CD45+ enrichment, followed by single cell library preparation and using 10x Genomics Chromium X. Sequencing was conducted using both Illumina and UG 100™ platforms. Comparative data analysis was conducted using Compugen’s computational platform for identifying tumor-specific and immune-related features. **C.** UMAP visualization of single-cell RNA-sequencing data, colored by tumor sample origin, including lung. **D.** UMAP visualization of single-cell RNA-sequencing data, colored by cell type **E.** Comparison of the total number of cells detected across different tumor samples (Colon-148, Lung-150, Lung-151) and hash-enriched samples. Bars represent the number of cells captured by Illumina (blue), Ultima (orange), and both technologies combined (green). The consistency of cell detection highlights the robustness of both sequencing platforms. **F.** Scatterplots comparing total counts per gene between UG 100™ and NovaSeq 6000 sequencing technologies for each sample. The plots demonstrate a strong correlation, indicating high concordance in gene expression quantification across platforms. **G.** Scatterplots showing log-normalized gene counts for Illumina versus Ultima by sample. Each gene is represented as a point, with high agreement observed for most genes. Deviations from the diagonal indicate genes with platform-specific differences (mainly ribosomal and mitochondrial genes with little importance) , labeled for identification.

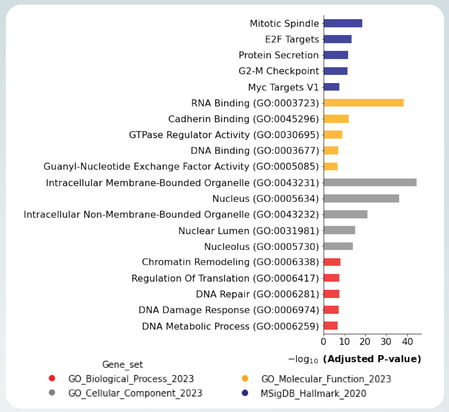
UG 100™ technical properties provide a unique opportunity to study gene structure

UG 100™ reads consistently start from the 3’ end of mRNA transcripts, unlike NovaSeq reads. This unique feature allows for clearer transcript termination. Leveraging this, we developed a computational pipeline to detect alternative splicing and poly-adenylation alterations in protein-coding genes.

A.



B.



A computational pipeline to detect alternative splicing using UG 100™ reads. A. Description of the computational pipeline, briefly: using the StringTie assembler (V2.2.3) we detected peaks supported by more than 100 reads, then removed any peaks situated within nine base pairs of genomic polyA or polyT tracts. Next, we retained only those peaks falling within exons of protein-coding transcripts (based on GENCODE) that lie in the coding sequence. From this refined set, we identified 2,704 genes whose splicing patterns may alter protein length. **B.** Gene-set enrichment analysis of the genes predicted to undergo alternative splicing which alters the protein.

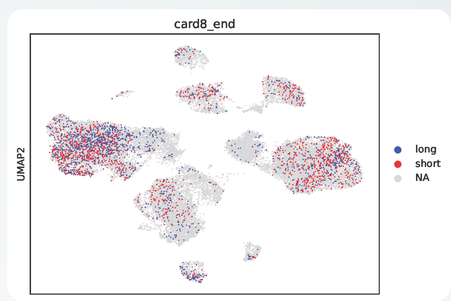
Differential splicing of the CARD8 gene suggests potential regulatory mechanism of inflammasome activity

We decided to focus on one of the genes identified in our analysis—CARD8, a component of the inflammasome complex, which is s a multiprotein complex that detects pathogenic signals and activates inflammatory responses. Our sequence data indicate the presence of two CARD8 isoforms, a long and a short. UG 100™ reads enable the clear and accurate detection of the isoforms.

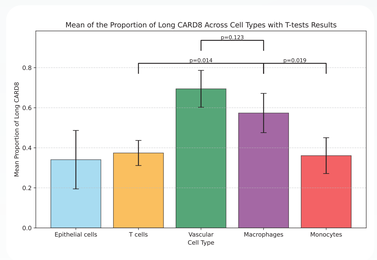
A.



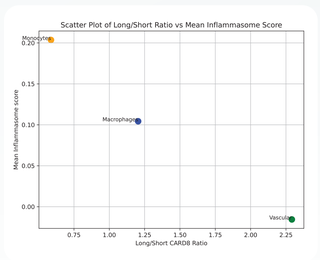
B.



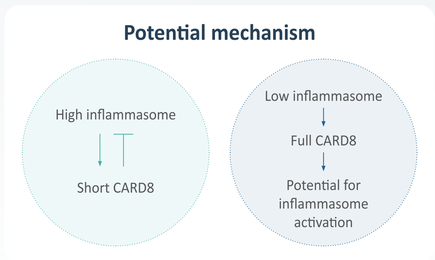
C.



D.



E.



Analysis of CARD8 Isoforms and Their Association with Cell Types and Inflammasome Activity. A. IGV visualization of the CARD8 gene locus. The genomic alignment tracks for CRC1 and Lung 1/2 samples sequenced with Ultima and Illumina platforms display expression profiles across the CARD8 gene. The zoomed-in view highlights two major CARD8 isoforms detected and how the UG 100™ reads allow to precisely determine the transcript and differentiate between the long form (537 amino acids) and the short form (70 amino acids). **B.** UMAP visualization of single cell expression of the two CARD8 isoforms. **C.** Mean proportions of the long CARD8 isoform across different cell types. Bar plots represent epithelial cells, T cells, vascular cell types, macrophages, and monocytes, with statistical comparisons (t-tests) indicated by brackets and associated p-values. Vascular cells and macrophages show significantly higher proportions of the long CARD8 isoform compared to other cell types. **D.** Scatter plot showing the relationship between the long-to-short CARD8 isoform ratio and mean inflammasome scores across cell types. Macrophages and vascular cells display distinct clusters, with macrophages showing a higher inflammasome score. The plot suggests a positive association between the CARD8 isoform ratio and inflammasome activity calculated using scanpy score_genes function with the following genes: CASP1, CASP4, CASP5, PYCARD , GSDMD **E.** Proposed regulatory mechanism: In cells with high inflammasome signature expression, inflammation may be suppressed through the production of inactive CARD8. In cells with low inflammasome activity, this regulatory mechanism is absent, potentially allowing for rapid inflammasome activation.

Conclusion

Our comparative analysis of NovaSeq and UG 100™ single-cell RNA sequencing highlights the pivotal advantage of 3’ end-specific reads from the UG 100™ platform for precisely resolving alternative splicing and polyadenylation events. Notably, this capability enabled the identification of a truncated CARD8 isoform that may modulate inflammasome activity in specific cell types, potentially functioning as a negative feedback mechanism. These findings underscore the broader power of such data to uncover novel molecular events with substantial implications for tumor biology and immune regulation. By incorporating this enhanced resolution into single-cell transcriptomics, we establish a robust framework for investigating alternative splicing as a regulatory mechanism.

By leveraging these methodologies, we can develop a proprietary 'splicing-aware' gene sequencing database, integrate it into our Unigen™ platform, and utilize it to further identify novel drug targets and advance our understanding of cancer immunotherapy biology.