OPEN ACCESS Freely available online

Immunogenetics: Open Access

Research Article

Gene Model Correction for PVRIG in Single Cell and Bulk Sequencing Data Enables Accurate Detection and Study of its Functional Relevance

Sergey Nemzer[1], Niv Sabath[1], Assaf Wool[1], Zoya Alteber[1], Hirofumi Ando[2], Amanda Nickles-Fader[2], Tian-Li Wang[2], Ie-Ming Shih[2], Drew M. Pardoll[2], Sudipto Ganguly[2], Yaron Turpaz[1], Zurit Levine[1], Roy Z Granit[1*]

[1]Department of Developmental Biology and Cancer Research, The Hebrew University of Jerusalem, Jerusalem, Israel; [2]Department of Immunology, Johns Hopkins University, School of Medicine, Baltimore, USA

ABSTRACT

Single cell RNA sequencing (scRNA-seq) has gained increased popularity in recent years and has revolutionized the study of cell populations; however, this technology presents several caveats regarding specific gene expression measurement. Here we examine the expression levels of several immune checkpoint genes, which are currently assessed in clinical studies. We find that unlike in most bulk sequencing studies, PVRIG, a novel immune-modulatory receptor in the DNAM-1 axis, suffers from poor detection in 10x Chromium scRNA-seq and other types of assays that utilize the GENCODE transcriptomic reference (gene model). We show that the default GENCODE gene model, typically used in the analysis of such data, is incorrect in the PVRIG genomic region and demonstrate that fixing the gene model recovers genuine PVRIG expression levels. We explore computational strategies for resolving multi-gene mapped reads, such as those implemented in RSEM and STARsolo and find that they provide a partial solution to the problem. Our study provides means to better interrogate the expression of PVRIG in scRNA-seq and emphasizes the importance of optimizing gene models and alignment algorithms to enable accurate gene expression measurement in scRNA-seq and bulk sequencing. The methodology applied here for PVRIG can be applied to other genes with similar issues.

Keywords: Gene model; PVRIG; Gene expression; T-cell; scRNA-seq

INTRODUCTION

PVRIG is a novel immune-modulatory receptor in the DNAM-1 axis [1], PVRIG directly inhibits T and NK cell activation while also competing with the co-activating receptor DNAM-1, for the binding of a shared ligand, PVRL2, which is expressed on tumor and antigen-presenting cells. An antibody blocking the PVRIG/ PVRL2 interaction was shown to increase CD8+ T-cell and NK cell activity in preclinical studies [2] and is currently being evaluated in early clinical trials (NCT04570839, NCT03667716). The advancement of scRNA-seq methods has enabled a more accurate characterization of PVRIG expression and of other drug targets within different cell types. Consequently, we have shown that PVRIG expression is uniquely enriched on early-differentiated T Stem-Cell Memory (TSCM) CD8+ T cells [3], unlike other immune checkpoints such as CTLA4, PD1, LAG3, TIM3 and TIGIT which are more highly expressed on exhausted CD8+ T cells [4]. This unique expression pattern in TSCM suggests a potential for improved clinical response, as these cells have been shown to possess an increased capacity to expand and self-renew to generate waves of activated T-cells [5,6] and their increased levels were found to be associated with a positive response to immunotherapy treatment [7].

In recent years droplet-based single-cell RNA sequencing has gained increased popularity within the scientific community, mainly utilizing the 10x Genomics (10x) Chromium platform [8]. In support of this platform 10x have developed the CellRanger software suite which allows easy and uniform conversion of raw sequencing data into a cell-by-gene count matrix [9]. One of the

Correspondence to: Roy Z Granit, Department of Developmental Biology and Cancer Research, The Hebrew University of Jerusalem, Jerusalem, Israel; E-mail: royg@cgen.com

Received: 24-Jan-2024, Manuscript No. IGOA-24-29326; **Editor assigned:** 26-Jan-2024, PreQC No. IGOA-24-29326 (PQ); **Reviewed:** 09-Feb-2024, QC No. IGOA-24-29326; **Revised:** 14-Mar-2025, Manuscript No. IGOA-24-29326 (R); **Published:** 21-Mar-2025, DOI: 10.35248/IGOA.25.10.249

Citation: Nemzer S, Sabath N, Wool A, Alteber Z, Ando H, Nickles-Fader A, et al. (2024) Gene Model Correction for PVRIG in Single Cell and Bulk Sequencing Data Enables Accurate Detection and Study of its Functional Relevance. Immunogenet Open Access. 10:249.

innovations implemented in droplet technology is the use of 3' or 5' biased mRNA sequencing which allows considerable cost saving (less than $1 per cell) and thus enables the profiling of tens or hundreds of thousands of cells; greatly promoting the study of cell populations and tissue composition [10]. However, probing the expression of individual genes using this technology might pose challenges associated with the relative sparsity of the data [11].

MATERIALS AND METHODS

scRNA-seq processing

Single-cell RNA-seq raw data was downloaded from GEO and processed using CellRanger 7.0.0 count option employing three gene model references: 1) GENCODE (default), 2) Model from Pool et al., and 3) a specific model where only STAG3-209 transcript was removed (CGEN model). For Smart-seq2 raw data was not available and thus processed count matrices were used. CellRanger output data or count-metrics were explored and further processed using scanpy 1.9.1 while employing standard preprocessing steps and clustering. Doublets were removed using Scrublet 0.2.3 with default parameters. Clusters were classified based on the expression of known markers (CD8 T-cells: CD3E +CD3D+TRAC+CD8A+CD4-, CD4 T-cells: CD3E+CD3D +TRAC+CD4+CD8A-, Treg: CD3E+CD3D+TRAC +CD4+CD8A-FOXP3+, NK:CD3E-CD3D-NRC1+). For multi-mapping analysis (STARsolo) STAR 2.7.10a was used employing either uniform or EM options; to generate filtered cells matrices 'STAR ~runMode soloCellFiltering' was used. Long-read sequencing (PacBio) was aligned using STAR.

Bulk RNA-Seq processing

Bulk RNAseq data was downloaded from GEO and processed through STAR 2.7.9a and RSEM 1.3.3 (STAR-based) using the two gene model references as described above.

Generation of corrected gene model

To generate a corrected gene model the reference provided as default by 10x, GRCh38 (GENCODE v32/Ensembl 98), was downloaded. Rows relating to STAG3-209 were deleted from the GFT file and new reference was generated using 'cellranger mkref' for CellRanger scRNAseq or using 'STAR ~runMode genome Generate'for bulk and STARsolo analysis.

Flow cytometry and scRNAseq

Tumor biospecimens from patients with High-Grade Serous Ovarian Cancer (HGSOC) were processed within 6 hours of surgical resection. Dissociation was done using human tumor dissociation kit (Cat# 130-095-929, Miltenyi Biotec, Bergisch Gladbach, Germany) and gentleMACS™ Octo Dissociator with heaters as per manufacturer's instruction. PBMCs were enriched by density-gradient centrifugation with Ficoll-Paque Plus (17-1440-02, Cytiva).

Flow cytometry for PVRIG and TIGIT expression on viable CD8 and CD4 T cells was done as previously described. For sorting prior to scRNAseq, 106 cells per sample were stained with the Live/Dead aqua viability dye (Invitrogen Cat# L34957), followed by Fc receptor blocking. Cells were then stained with FITC-conjugated anti-human CD45 (clone HI30, BioLegend). Viable CD45+ cells were sorted on a BD FACSAriaTM Fusion sorter.

Approximately 16,500 immune cells per patient were loaded on a 10x Genomics Chromium platform for targeted recovery of ~10,000 cells. Libraries were sequenced on a NovaSeq 6000 system.

RESULTS

PVRIG is poorly detected in 10x chromium scRNA-seq data

While studying the expression of T-cell checkpoints in scRNA-seq data, unexpectedly, we found that PVRIG expression in droplet-based data (most commonly 10x chromium) is often considerably lower versus what is observed in full-length scRNA-seq (e.g., well-based Smart-Seq) or in bulk RNA-seq (Figures 1a and 1b) and does not align with flow cytometry data for surface protein expression. We hypothesized that the observed reduced expression of PVRIG in existing scRNA-seq datasets is a technical artifact stemming from droplet-based scRNA-seq. The literature describes three potential culprits for reduced gene expression in droplet-based scRNA-seq: (1) Reads mapping immediately 3' to known gene boundaries due to poor 3' UTR annotation; (2) Intronic reads stemming from unannotated exons or pre-mRNA; (3) Discarded reads by read count software (e.g. CellRanger, RSEM) due to multi-gene mapping to paralogous genes or read-through transcripts. We thus examined mapped reads within and around the PVRIG gene locus using bulk RNA-seq, 3' and 5' droplet-based scRNA-seq and long-read scRNA-seq from PBMC and sorted T cells data (Figures 1c and 1d). We did not observe a substantial number of reads downstream of the 3' end of the gene, suggesting that the cause is not an unidentified longer 3' UTR (Figure 1d). We also did not find an unusually high number of reads mapped to the intronic regions of the gene or to other loci in the genome. In contrast, we identified an overlapping transcript, STAG3-209 (ENST00000451963) that overlaps the entire PVRIG gene region (Figures 1c and 1d). This transcript is curated under the GENCODE gene model, which relies on Ensembl transcripts and is the default used by CellRanger. STAG3-209 stems from a readthrough of the STAG3 gene upstream of PVRIG, its supporting evidence consists of only one sequence originating from testis tissue (Figure S1a,b) and it is predicted to undergo non-sense mediated decay (NMD) (ENST00000451963.1 Ensembl Entry). Importantly, the weak supporting evidence of STAG3-209 resulted in the exclusion of this transcript from the RefSeq gene model (Figure S1c).

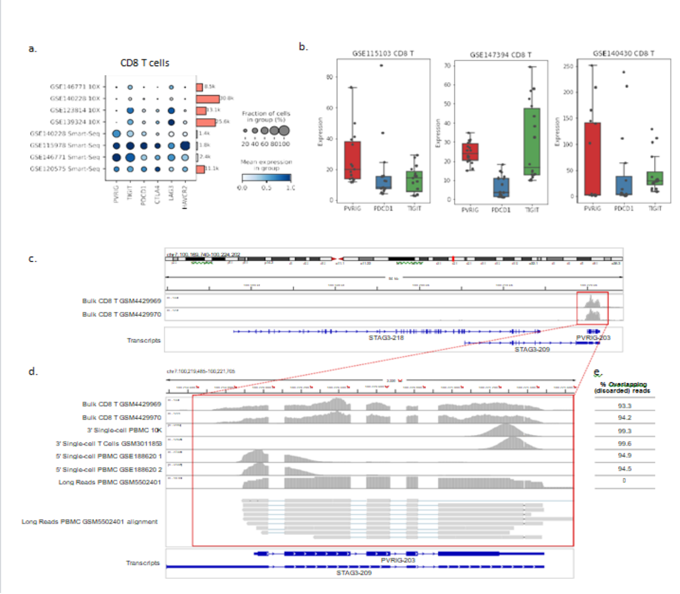

Figure 1: PVRIG is poorly detected in 10x chromium data and overlaps with STAG3-209. a) Dotplot showing the expression of PVRIG and additional T-cell checkpoints in several 10x datasets (top) versus smart-seq2 datasets (bottom) collected from TME. Bars indicate number of cells in each sample. b) Boxplots showing the expression levels of PVRIG, TIGIT and PD-1 across several CD8 T-cell datasets as measured using bulk RNA-seq. GSE140430 isolated CD8 T-cells from TME, rest are from blood. c) The STAG3-PVRIG genomic locus. Histograms feature the distribution of reads in bulk RNA-seq data of isolated CD8 T-cells highlighting that PVRIG is expressed in such cells (red box). Gene structures appear below in blue. d) Closer view of PVRIG locus. Histograms highlight the distribution of reads in the indicated datasets. e) Fraction of reads that map to both PVRIG and STAG3-209 and are thus discarded.

Little evidence supports STAG3-209, suggesting it could be discarded

The exons and introns of the canonical PVRIG (PVRIG-203) are almost entirely identical to that of the overlapping predicted STAG3-209 except for the second PVRIG intron, 82 bp length, and a slightly shorter (by two bp) 3' end of STAG3-209 (Figure 1d). We thus examined the mapping of reads from several bulk RNA-seq of sorted CD8 T cells around the second intron of PVRIG. We found that in all samples the reads follow the exact gene structure of PVRIG and almost no reads (except stochastic noise-level reads) are mapped to the intron that is part of the predicted STAG3-209 transcript, whereas several reads in each sample are mapped to the junction between exon 2 and exon 3 in PVRIG (Figure 1c). Thus, at least in CD8 T cells, all reads mapped to the PVRIG-STAG3 overlap are most likely to originate from the PVRIG mRNA. In addition, we examined exon-exon junctions from the cancer genome atlas (TCGA), data spanning over 12K samples and could not detect junctions that support the expression of the STAG3-209-PVRIG read through while junctions supporting the PVRIG gene were readily found (data not shown).

Because most short reads do not fall within exon-intron junctions, we also examined long-read data based on PacBio sequencing of T-cells [12]. We found that all reads spanning the second intron of PVRIG do not contain the retained-intron which is found in STAG3-209 and there are no junctions that support STAG3-209 (Figure 1d). Thus, at least in these datasets, there is no support for the expression of STAG3-209.

Gene model correction enables accurate detection of PVRIG expression

In light of different technical issue that can potentially hinder the expression of specific genes in scRNAseq, Pool et al recently proposed a modified gene reference model in which read through transcripts were removed; including deletion of STAG3-209 [13]. In order to verify the specificity of the change we also generated a gene model in which we removed only the STAG3-209 read through and validated that only PVRIG expression is altered by the modification (Figure S2). To test the effect of the modified references, we ran the standard cell ranger pipeline using these improved gene models on 10x Chromium public dataset released by 10x which contains PBMCs as well isolated CD8 T-cells from 4 donors that were confirmed by bulk RNA-seq to express PVRIG (Figure 2 and 1b). Indeed, when we used the improved gene models, we found elevated PVRIG expression in CD8+ T cells and NK populations previously found to express it (Figure 2). As expected, expression of other T cell checkpoints was not altered by the correction as expected. Moreover, the differential expression of PVRIG across cell types following the correction reflects protein levels of PVRIG previously assessed by flow-cytometry in human normal and tumor tissue (Figure 2). To further validate our correction, we generated scRNA-seq profiles of immune populations isolated from the TME of patients with ovarian cancer. These samples had been previously analyzed by flow cytometry as PVRIG-positive (Figure 3). We found that following the correction PVRIG demonstrated solid expression levels and patterns that match those previously described, namely high expression in NK and TSCM cells (Figures 3a and 3b).

Furthermore, we noted that the mRNA levels of PVRIG correlated with the protein levels (though with low number of supporting samples) measured by FACS in cells obtained from the same patient-derived samples (Figures 3c and 3d). Together, our findings suggest that the discarding of reads by cell ranger due to multi gene mapping was responsible for the observed decreased expression.

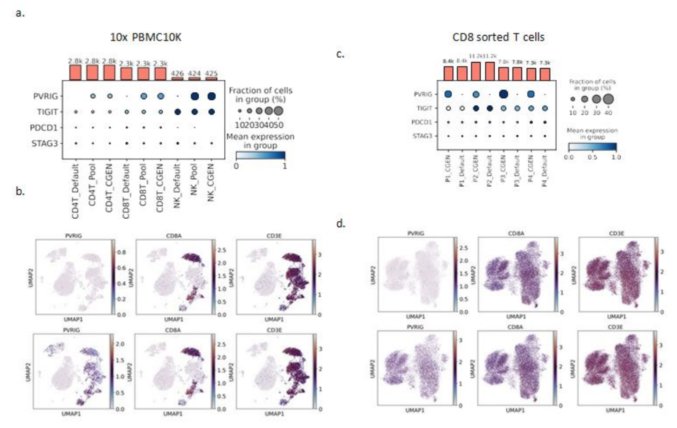

Figure 2: Gene model correction allows proper detection of PVRIG. a) Dotplot showing the expression of PVRIG and additional T-cell checkpoints as well as STAG3 in 10x PBMC10K dataset processed using various gene models subset for T and NK cells. Default-GENCODE, Pool-previously published, broadly corrected reference. CGE-gene model sepecifically corrected in the PVRIG gene locus. Bars indicate number of cells in each sample. b) UMAP showing the expression of PVRIG, CD8A and CD3E in the 10x PBMC10K dataset using the default gene model (top row) and the corrected gene model (bottom row). c,d) Same as a,b but for CD8+CD95+ sorted T-cells from 4 healthy donors (GSE147397).

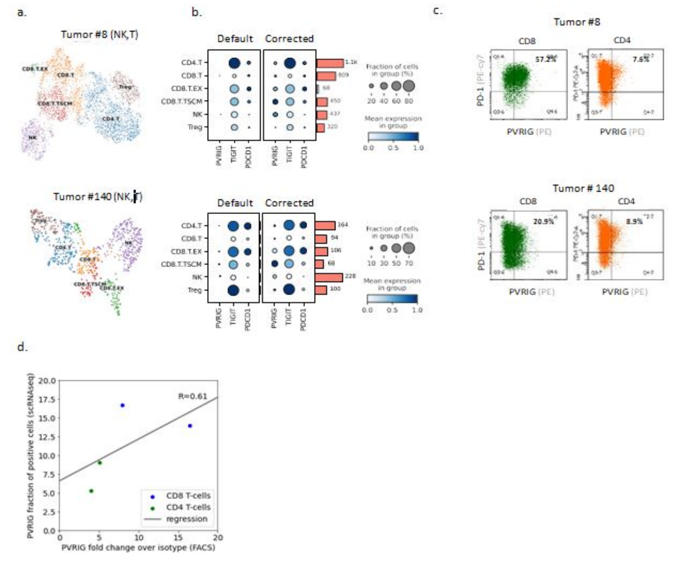

Figure 3: Corrected PVRIG expression in ovarian tumor samples demonstrates a correlation with protein expression. a) UMAPs showing the population distribution in two ovarian tumor samples subset for T and NK cells. b) Dotplot showing the expression of PVRIG and additional T-cell checkpoints in ovarian tumor samples processed using the default or corrected gene model subset for T and NK cells. Bars indicate number of cells in each sample. c) Matched FACS analysis showing the expression of PVRIG and PD-1 in CD8 or CD4 T cells. d) Correlation between PVRIG-positive fraction as measured using FACS (X-axis) to its parallel in the scRNA-seq data following the gene model correction. Linear-regression line is shown in gray as well as the Pearson correlation coefficient.

Gene model correctness determines accurate expression in bulk RNA-seq data

Considering our findings, we were interested to uncover the underlying cause which allows PVRIG to be readily detected in Smart-seq2 and bulk sequencing while it is poorly detected in 10x Chromium data. Hence, we analyzed bulk sequencing using the STAR aligner, which is used by CellRanger and using the GENCODE reference or our corrected gene model and found that PVRIG could only be detected upon using our model (Figure 4a). Thus, we concluded that the underlying reason for poor detection of PVRIG does not stem from 3' or 5' bias used in 10x droplet-based single cells data, but due to the gene model used by default which contains the STAG3-209 transcript. We have examined several publications that demonstrated PVRIG expression in Smart-seq2 and noted that they indeed employed gene models that do not contain STAG3-209.

Multimap algorithms partially rescue the expression of PVRIG

The issue of multi-mapped reads was extensively studied for full-length RNA-seq data and algorithms that distribute the multi-mapped reads between their alignments based on the uniquely mapped read ratio have been developed [14]. We therefore analyzed bulk data of CD8 T-cells using the RSEM read estimator [15] that employs STAR aligner data as input and contains algorithms to handle multi-mapped transcripts. We noted that indeed RSEM was able to recover PVRIG gene expression in bulk RNAseq (Figure 4b) regardless of the gene model used. More recently, efforts were made to resolve the issue of multi-mapped reads in scRNA-seq data as well [16,17]. To test the feasibility of such algorithms to correct the PVRIG-STAG3-209 multi-mapping we ran STARsolo on data which was aligned to the default gene model, using Expectation Maximization (EM) or uniform distribution as means to correct multi-gene mapping. We noted that the tool was indeed able to improve the detection of PVRIG, to similar levels obtained by employing gene model correction (Figures 4c and 4d). Further expression analysis demonstrated that the multi-map correction applied using STARsolo also increased the expression of STAG-3 in a way that is likely to be an artifact (Figure 4d). Though STARsolo was able to improve the expression of PVRIG it is apparent that this feature is not yet fully supported as the package currently generates a non-filtered matrix which holds the multi- mapping data, forcing the user to conduct additional non-trivial processing steps.

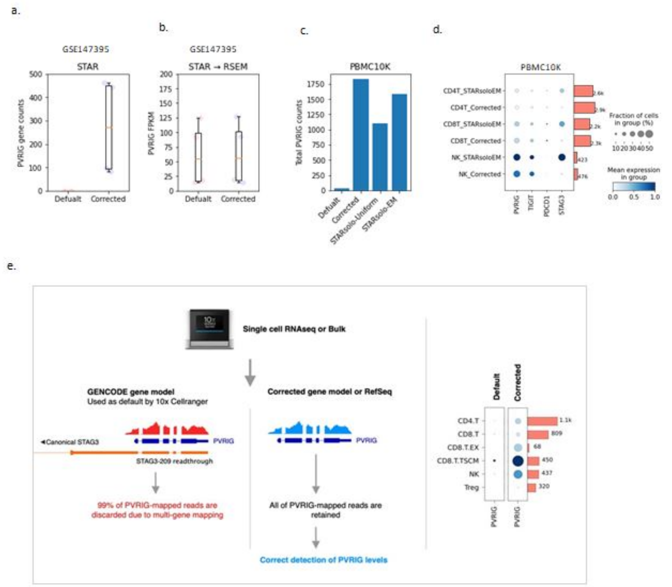

Figure 4: Correction using multimappers and summary illustration. a) Boxplot showing the number of PVRIG read counts per sample in bulk sequencing data of CD8 T-cells using the default or corrected gene model. b) Same as in a, but following multimap correction using RSEM. c) Barplot showing the number of PVRIG counts in PBMC10K dataset processed using the default, corrected gene model or following multi-mapping correction using STARsolo. d) DotPlot showing the expression of PVRIG, TIGIT, PD-1 and STAG3 in NK/T populations in the PBMC10K dataset processed using the corrected gene model or using STARsolo-EM method. e) Summarizing illustration; using the GENCODE gene model reads are discarded due to multi-mapping, following the correction of the gene model PVRIG expression is correctly measured.

DISCUSSION

One of the strengths of scientific investigation is the great diversity of approaches and highly distributed manner by which results are repeated using different methodologies. 10x Chromium revolutionized the field of scRNA-seq, allowing a uniform analysis of the outputs, which greatly contributes to the reproducibility of the results across different laboratories. However, as we demonstrate in our study, the decisions made by these tools, such as the selection of reference gene models, could have broad implications. We show that gene models should not be taken for granted when studying the expression of specific genes and emphasize the importance of refining such models. Our work reveals that PVRIG expression is underestimated in many single-cell and potentially also in bulk, datasets due to technical reasons that stem from the usage of the GENCODE gene model; that includes the rare STAG3-209 read through which almost entirely overlaps with PVRIG. While we point the spotlight on PVRIG, due to our specific scientific interest, similar phenomena are likely to affect additional genes as well. As the field of scRNA continues to evolve it is essential for the scientific community to address the inclusion of rare read through transcripts and correct annotation of 3' UTRs. While a holistic, genome-wide, solution such as the one suggested by Pool et al., may be useful, they still require further verification and refinement as they have broad implications on the expression of many genes.

Alternately, algorithms for handling multi-mapped reads count could be integrated into cell ranger, yet these also require rigorous testing and as we show in our study, currently available tools require further development before they become standard. Meanwhile, we suggest employing our generated gene model which we demonstrated to faithfully correct the expression of PVRIG in a specific manner that does not alter the expression of additional genes. Using our corrected reference, we show that PVRIG expression levels in CD8 T-cells located at the periphery and in primary tumor samples are comparable to other immune-checkpoints, such as TIGIT and PD-1. Thus, we believe that studies which profiled T-cells using 10x Chromium or other methods that have used the GENCODE reference should re-evaluate the expression of PVRIG and other genes that might suffer from a similar phenomenon. Future works and cell atlases should also carefully consider which gene model is utilized. This will hopefully lead to further establishment and understanding of the role of PVRIG in the context of immune-oncology and will support ongoing clinical efforts.

CONCLUSION

In conclusion, the implementation of gene model correction for PVRIG in both single-cell and bulk sequencing datasets represents a significant advancement in genomic research. By refining the accuracy of gene annotations and improving the quality of sequencing data, this approach enhances our ability to detect and analyze PVRIG'S functional roles. This refinement not only facilitates more precise understanding of PVRIG'S biological functions but also paves the way for more reliable investigations into its potential implications in various diseases and therapeutic contexts. Ultimately, the enhanced accuracy in gene modeling contributes to a deeper and more nuanced comprehension of gene function, which is crucial for advancing both fundamental research and clinical applications.

DATA AVAILABILITY

The public scRNAseq datasets used in the study: GSE146771, GSE140228, GSE123814, GSE136324, GSE140228, GSE115978, GSE120575, GSE188620, GSM5502401 and 10x.

PBMC10K (10k Human PBMCs, 3' v3.1, Chromium Controller). Bulk RNAseq datasets: GSE115103, GSE140430, GSE147394, GSM4429969, GSM4429970.

COMPETING INTERESTS

S.N.,N.S.,A.W.,Z.A.,Z.L.,Y.T.,R.Z.G. are employees of Compugen LTD. D.M.P. is on the scientific advisory board of Compugen and receives from Compugen research support.

ACKNOWLEDGMENTS

We thank Eran Ophir for reviewing the manuscript and insightful comments.

REFERENCES

1. Alteber Z, Kotturi MF, Whelan S, Ganguly S, Weyl E, Pardoll DM, et al. Therapeutic targeting of checkpoint receptors within the DNAM1 Axis. Cancer Discov. 2021;11(5):1040-1051.
2. Whelan S, Ophir E, Kotturi MF, Levy O, Ganguly S, Leung L, et al. PVRIG and PVRL2 are induced in cancer and inhibit CD8+ T-cell function. Cancer Immunol Res. 2019;7(2):257-268.
3. Alteber Z, Cojocaru G, Frenkel M, Weyl E, Sabath N, Wool A, et al. 252 Novel DNAM-1 axis member, PVRIG, is potentially a dominant checkpoint involved in stem-like memory T cells-dendritic cell interaction. J ImmunoTher Cancer. 2021;9(Suppl 2):A272–A273.
4. Wherry EJ, Kurachi M. Molecular and cellular insights into T cell exhaustion. Nat Rev Immunol. 2015;15(8):486–499.
5. Klebanoff CA, Gattinoni L, Torabi-Parizi P, Kerstann K, Cardones AR, Finkelstein SE, et al. Central memory self/tumor-reactive CD8+ T cells confer superior antitumor immunity compared with effector memory T cells. Proc Natl Acad Sci USA. 2005;102(27):9571-9576.
6. Krishna S, Lowery FJ, Copeland AR, Bahadiroglu E, Mukherjee R, Jia L, et al. Stem-like CD8 T cells mediate response of adoptive cell immunotherapy against human cancer. Science. 2020;370(6522): 1328–1334.
7. Sade-Feldman M, Yizhak K, Bjorgaard SL, Ray JP, de Boer CG, Jenkins RW, et al. Defining T cell states associated with response to checkpoint immunotherapy in melanoma. Cell. 2018;175(4): 998-1013.
8. Svensson V, da Veiga Beltrame E, Pachter L. A curated database reveals trends in single-cell transcriptomics. Database (Oxford). 2020:baaa073.
9. Zheng GX, Terry JM, Belgrader P, Ryvkin P, Bent ZW, Wilson R, et al. Massively parallel digital transcriptional profiling of single cells. Nat Commun. 2017;8:14049.
10. Zhang X, Li T, Liu F, Chen Y, Yao J, Li Z, et al. Comparative analysis of droplet-based ultra-high-throughput single-cell RNA-Seq systems. Mol Cell. 2019;73(1):130-142.e5.
11. Wang X, He Y, Zhang Q, Ren X, Zhang Z. Direct comparative analyses of 10x genomics chromium and smart-seq2. Genomics Proteomics Bioinformatics. 2021;19(2):253-266.
12. Tian M, Cheuk AT, Wei JS, Abdelmaksoud A, Chou HC, Milewski D, et al. An optimized bicistronic chimeric antigen receptor against GPC2 or CD276 overcomes heterogeneous expression in neuroblastoma. J Clin Invest. 2022;132(16):e155621.
13. Pool AH, Poldsam H, Chen S, Thomson M, Oka Y. Enhanced recovery of single-cell RNA-sequencing reads for missing gene expression data. BioRxiv. 2022:2022-2024.
14. Deschamps-Francoeur G, Simoneau J, Scott MS. Handling multi-mapped reads in RNA-seq. Comput Struct Biotechnol J. 2020;18:1569-1576.
15. Li B, Dewey CN. RSEM: Accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics. 2011;12(1):323.
16. Kaminow B, Yunusov D, Dobin A. STARsolo: Accurate, fast and versatile mapping/quantification of single-cell and single-nucleus RNA-seq data. BiorXiv. 2021;05.
17. Srivastava A, Malik L, Smith T, Sudbery I, Patro R. Alevin efficiently estimates accurate gene abundances from dscRNA-seq data. Genome Biol. 2019;20(1):65.