CARTEpigenoQC: A Quality Control Toolkit for CAR-T Single-Cell Epigenomic Data

2025-07-26

[1] University of Sydney

0.1 Summary

CARTEpigenoQC is an R-based toolkit designed to streamline quality control (QC) for single-cell epigenomic datasets involving Chimeric Antigen Receptor (CAR)-engineered T cells. With the growing application of scATAC-seq, scCUT&Tag, and scBS-seq to characterize CAR-T cell states (Satpathy et al. 2019), it has become critical to perform customized QC that not only addresses standard metrics like FRiP (Fraction of Reads in Peaks) and TSS enrichment, but also directly detects signal from CAR vector insertion sites.

CARTEpigenoQC supports both 10x Genomics and non-10x data formats and produces HTML and PNG summary outputs suited for exploratory analysis and regulatory-grade preclinical reporting. It is intended to assist researchers, core facilities, and translational immunologists in ensuring the validity of single-cell epigenomic profiling of engineered T cells (Finck and al. 2022; Robinson, Nguyen, and Sadelain 2023).

0.2 Statement of Need

Despite the rise in single-cell chromatin profiling technologies in immunotherapy, current QC pipelines do not provide functionality to assess CAR vector integration in scATAC-seq or related datasets. Existing tools like ArchR (Granja et al. 2021), Signac (Stuart et al. 2021), or SnapATAC (Fang et al. 2021) focus on general chromatin metrics but lack transgene-specific support. CARTEpigenoQC fills this gap by integrating standard QC (e.g., FRiP, duplication, peak overlap) with direct inspection of vector signal such as EF1, PGK1, or TRAC locus insertions that are common in CAR constructs (Eyquem et al. 2017).

This tool was inspired by the increasing use of vector-based gene therapies and the demand for preclinical quality pipelines (Maude et al. 2018). It is lightweight, fast, and does not require complex dependencies. It generates clean, reproducible Markdown-based reports suitable for clinical handoff, regulatory review, or publication.

0.3 Features

- FRiP calculation per cell using fragment overlaps with peak regions.
- Visualization of CAR vector site coverage across single cells.
- Markdown-based report with interactive and static plots.
- Compatible with 10x Genomics `fragments.tsv.gz`, ArchR fragments, and BED-formatted CAR insertion sites.
- Designed for reproducibility and integration into preclinical immunotherapy pipelines.

0.4 Example Usage

CARTEpigenoQC includes a `run_qc.R` script and report template. When supplied with fragment files and a BED file of CAR insertion coordinates, it calculates QC metrics and generates a self-contained HTML report.

```
Rscript scripts/run_qc.R \
  --fragments data/pbmc_10k_fragments.tsv.gz \
```

```
--car_sites data/CAR_standard_insertion_sites.bed \
--sample PBMC_10k
```

0.5 Example Outputs

0.5.1 FRiP Score Distribution

This histogram visualizes the distribution of FRiP (Fraction of Reads in Peaks) across all cells. A peak around 0.2–0.4 is expected for high-quality scATAC-seq data (Chen et al. 2019). A right-skewed distribution is typical, and cells with FRiP < 0.2 may be filtered out in downstream analyses.
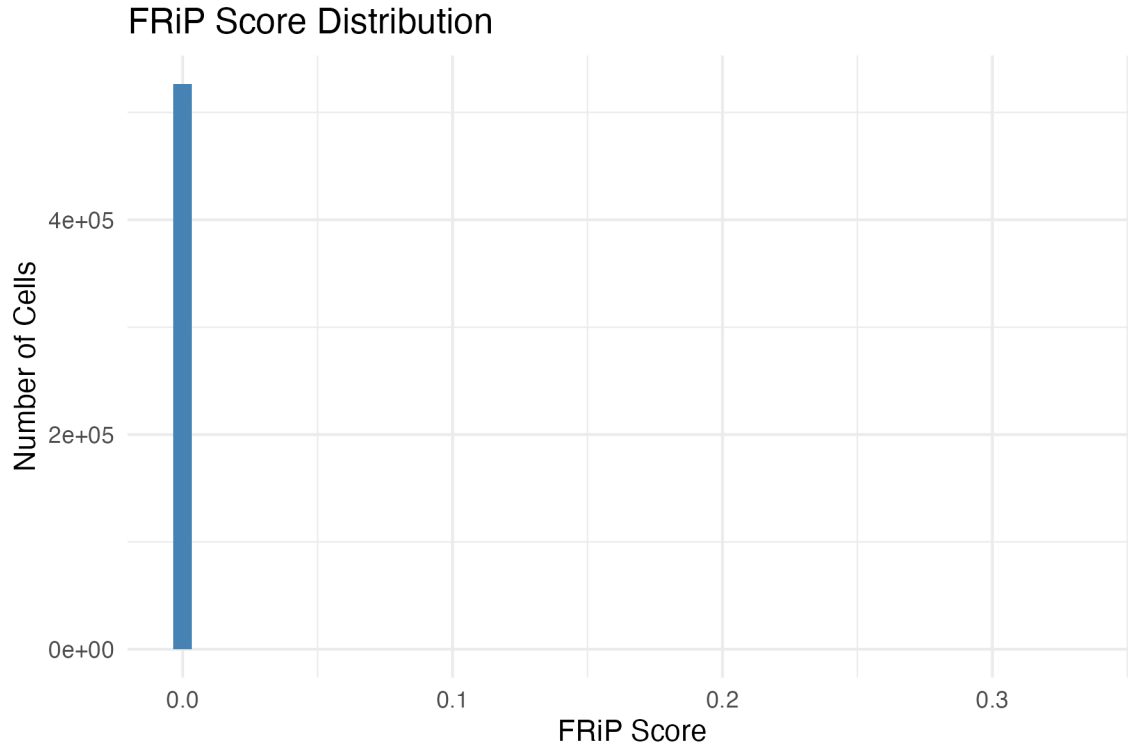

Figure 1: FRiP Histogram

0.5.2 Top Cells by FRiP

This table lists the top 20 cells with the highest FRiP scores. These cells often exhibit clearer chromatin signatures and are more informative for downstream applications like motif enrichment or regulatory network analysis.

0.5.3 CAR Insertion Site Coverage Heatmap

The heatmap shows per-cell read coverage across CAR vector insertion sites (e.g., EF1a, PGK1, TRAC). Each row represents a cell barcode, and each column represents a CAR target site. This allows visualization of which cells are likely to carry vector insertions and where signal enrichment occurs, helping to verify construct delivery or identify mosaicism in editing (Torres et al. 2022).

0.5.4 Top CAR Sites by Total Coverage

This table summarizes the top 10 CAR insertion sites ranked by the total number of reads mapped to each site across all cells. It helps identify the most active or consistently covered transgene targets, which is critical

|	barcode	total	in_peaks	frip
1	ACAGAAACAGGTCTGC-1	3	1	0.3333333333
2	ACTGTCCCATAGACGG-1	6	1	0.1666666667
3	GCAGCTGTCAATTGGC-1	17	1	0.0588235294
4	TTGTCTAAGCAGTAGC-1	24	1	0.0416666667
5	AGTTACGGTATCTGCA-1	1457	2	0.0013726836
6	CAGTATGAGTCGTATC-1	781	1	0.0012804097
7	ACTATTCTCGGATAAA-1	1695	2	0.0011799410
8	TCAGTCCGTGGACGAT-1	957	1	0.0010449321
9	AATGCCAGTGGCCTCA-1	966	1	0.0010351967
10	AATGGCTGTCCCTTTG-1	1002	1	0.0009980040
11	GCAGATTAGCCCATGC-1	1009	1	0.0009910803
12	AGTCAACAGGCGTCCT-1	1020	1	0.0009803922
13	TGCCTGTTCGCAGATT-1	1021	1	0.0009794319
14	CCCTCTCTCTTAACGG-1	1059	1	0.0009442871
15	CACATGACATATTGGC-1	1064	1	0.0009398496
16	CGCTATCTCGCTCTAC-1	1094	1	0.0009140768
17	TTAACGGAGTCGATAA-1	2228	2	0.0008976661
18	GAAGAGCCAAAGCATA-1	1208	1	0.0008278146
19	CCCGTTAAGGCAATTA-1	1234	1	0.0008103728
20	CAACGTAGTAAGTCTC-1	1243	1	0.0008045053

Figure 2: Top FRiP Table
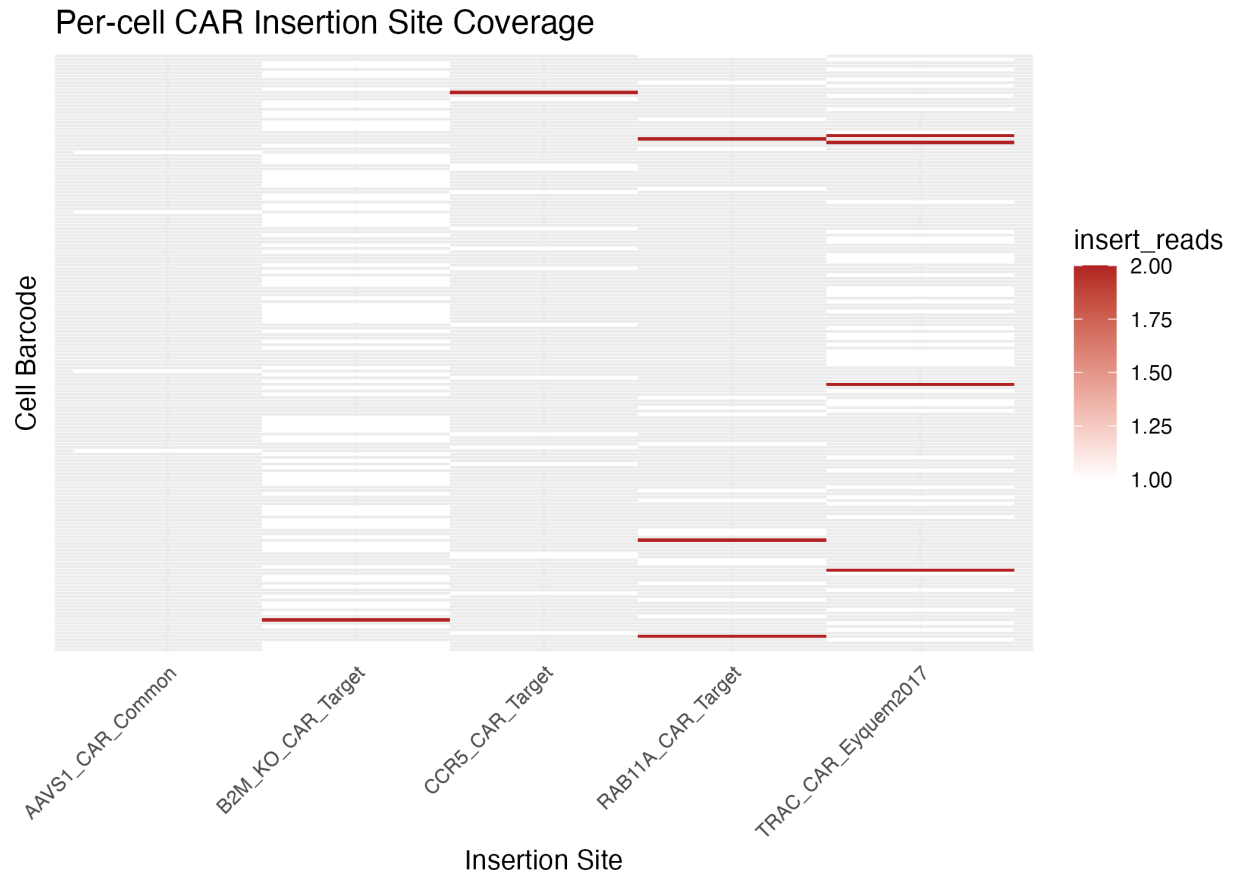
Figure 3: CAR Insertion Heatmap

3

in validating vector design and insertion stability (Frangieh et al. 2021).

	site	total_reads
1	B2M_KO_CAR_Target	91
2	TRAC_CAR_Eyquem2017	53
3	RAB11A_CAR_Target	24
4	CCR5_CAR_Target	18
5	AAVS1_CAR_Common	4
6	NA	NA
7	NA	NA
8	NA	NA
9	NA	NA
10	NA	NA

Figure 4: Top CAR Sites

0.6 Repository and Installation

The source code is hosted on GitHub:
https://github.com/biosciences/CARTEpigenoQC

Installation:

```
devtools::install_github("biosciences/CARTEpigenoQC")
```

0.7 Acknowledgements

The author thanks collaborators at Westmead Institute for Medical Research for insights into CAR-T clinical pipelines, and the Sydney Informatics Hub for pipeline development guidance.

References

Chen, Hanqing, Caleb A Lareau, Thomas Andreani, Matthew E Vinyard, Sandra P Garcia, Kyle Clement, Miguel A Andrade-Navarro, and Jason D Buenrostro. 2019. "High-Throughput Sequencing of the Transcriptome and Epigenome of Single Cells." *Nature Reviews Genetics* 20 (10): 631–47.

Eyquem, Julien, Jorge Mansilla-Soto, Theodoros Giavridis, Stefan JC van der Stegen, Mohamad Hamieh, Kristen M Cunanan, Alexander Odak, Mehmet Gonen, and Michel Sadelain. 2017. "Targeting a CAR to the TRAC Locus with CRISPR/Cas9 Enhances Tumour Rejection." *Nature* 543 (7643): 113–17.

Fang, Rui, Sebastian Preissl, Yanxiao Li, Xinyi Hou, Julian Lucero, Xinyi Wang, Amir Motamedi, et al. 2021. "SnapATAC: A Comprehensive Analysis Package for Single Cell ATAC-Seq." *bioRxiv*. https://doi.org/10.1101/615179.

Finck, Alexander, and et al. 2022. "Epigenetic and Transcriptional Profiling of CAR t Cells Reveals Insights into t Cell Exhaustion and Persistence." *Nature Medicine* 28 (3): 467–79.

Frangieh, Cyrus J, John C Melms, Pratik I Thakore, and et al. 2021. "Multimodal Pooled Perturb-Seq Reveals Synergistic and Antagonistic Gene Modules in the Immune Response." *Cell* 184 (3): 882–895.e21.

Granja, Jeffrey M, M Ryan Corces, Sarah E Pierce, Seyed T Bagdatli, Harinder Choudhry, Howard Y Chang, and William J Greenleaf. 2021. "ArchR Is a Scalable Software Package for Integrative Single-Cell Chromatin Accessibility Analysis." *Nature Genetics* 53 (3): 403–11.

Maude, Shannon L, Theodore W Laetsch, Jochen Buechner, Silvia Rives, Michel Boyer, Henri Bittencourt, Peter Bader, et al. 2018. "Tisagenlecleucel in Children and Young Adults with b-Cell Lymphoblastic Leukemia." *New England Journal of Medicine* 378 (5): 439–48.

Robinson, Kimberly, Mai Nguyen, and Michel Sadelain. 2023. "Single-Cell Epigenomic Profiling in Immunotherapy: New Tools and Perspectives." *Trends in Immunology* 44 (2): 115–29.

Satpathy, Ansuman T, Jeffrey M Granja, Kathryn E Yost, Yuhan Qi, Francesco Meschi, Gavin P McDermott, Ben N Olsen, et al. 2019. "Massively Parallel Single-Cell Chromatin Landscapes of Human Immune Cell Development and Intratumoral t Cell Exhaustion." *Nature Biotechnology* 37 (8): 925–36.

Stuart, Tim, Avi Srivastava, Shila Madad, Caleb A Lareau, and Rahul Satija. 2021. "Signac: Multimodal Single-Cell Chromatin Analysis with r." *bioRxiv.* https://doi.org/10.1101/2020.11.09.373613.

Torres, David, Ankit Tiwari, Jinyu Du, et al. 2022. "Comprehensive Single-Cell CRISPR Screening Reveals a Novel Vector Integration Bias in t Cells." *Nature Biotechnology* 40: 1636–45.