

Analysis of microsatellite instability intensity in single-cell resolution with scMnT reveals tumor heterogeneity in colorectal cancer

Gyumin Park¹ and Jihwan Park^{1,2,*}

¹School of Life Sciences, Gwangju Institute of Science and Technology (GIST), Gwangju 61005, Republic of Korea

²Eyoncell, Gwangju 61005, Republic of Korea

*Corresponding author. School of Life Sciences, Gwangju Institute of Science and Technology (GIST), Jihwan Park, Gwangju 61005, Republic of Korea. E-mail: jihwan.park@gist.ac.kr

Abstract

Although microsatellite instability (MSI) is traditionally regarded as a binary trait, growing evidence indicates that it exists on a continuum, with higher intensity of MSI correlating with better prognosis. However, accurate quantification of MSI intensity is challenging due to the heterogeneous cellular composition of the tumor microenvironment. To this end, we introduce scMnT, a bioinformatic method for identifying MSI cells at single-cell resolution and quantifying MSI intensity based on microsatellite allele lengths. Applying scMnT to scRNA-seq datasets produced robust results and revealed substantial variation in MSI intensity among colorectal cancer patients, with higher MSI intensity correlating with improved responses to immunotherapy. Our approach provides a novel framework for analyzing MSI at the single-cell level, allowing new opportunities for MSI research. scMnT is publicly available at <https://github.com/18parkky/scMnT>.

Keywords bioinformatics, colorectal cancer, single-cell analysis, single-cell RNA-sequencing, microsatellite instability, immune checkpoint inhibitor

Introduction

MSI is a genetic hypermutation phenotype characterized by an increased mutation rate in the microsatellite regions due to DNA mismatch repair (MMR) deficiency. MSI is frequently observed in gastric, breast, endometrial, cervical, and especially colorectal cancers (CRCs) [1, 2]. Importantly, MSI tumors exhibit better response to immune checkpoint blockade (ICB) therapy, underscoring MSI as a critical biomarker for therapeutic screening [3–5]. While MSI is often regarded as a binary phenotype, studies have shown that MSI can be measured as a quantitative phenotype, where higher MSI intensity has been associated with increased tumor-infiltrating T lymphocytes and better response to PD-1 blockade [6, 7]. These previous studies highlight the importance of quantifying MSI intensity, rather than merely detecting MSI, to better stratify patients eligible for ICB therapy.

Despite growing recognition of MSI as a quantitative phenotype, current detection methods fall short in this regard. PCR-based MSI detection is limited by the small number of loci assessed, while immunohistochemistry (IHC) only detects the presence or absence of MMR proteins. Although next-generation sequencing (NGS)-based tools such as MSIsensor-pro provide the percentage of unstable loci, which can be used as a proxy for MSI score, they do not account

for the presence of non-malignant cells (e.g. normal cells) nor do they distinguish the varying MSI intensity between cancer cells [8–11]. Among several possible improvements, selectively analyzing tumor cells using single-cell resolution techniques could enhance the quantification of MSI intensity. Indeed, single-cell RNA sequencing (scRNA-seq) has become indispensable for cancer research, enabling the study of tumor heterogeneity at unprecedented resolution [12–14].

A recently introduced method, MSIsensor-RNA, enables MSI detection from RNA-based sequencing data, including single-cell RNA-seq [15]. Although it demonstrates high accuracy for tumor-level MSI classification, MSIsensor-RNA infers MSI indirectly from the expression patterns of MSI-associated gene signatures rather than from microsatellite instability itself. As a result, non-malignant cell populations—such as immune and stromal cells—within MSI tumors are also assigned elevated MSI scores, reflecting the tumor microenvironment rather than cell-intrinsic genomic instability. While this behavior may be appropriate for studies focused on MSI-associated transcriptional programs or immune contexture, MSIsensor-RNA does not provide a direct or quantitative measure of MSI intensity and cannot resolve MSI at the single-cell level. Thus, the lack of tools for single-cell resolution MSI analysis highlights the need for new approaches [16].

Received: August 13, 2025. **Revised:** February 21, 2026. **Accepted:** February 26, 2026

© The Author(s) 2026. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact reprints@oup.com for reprints and translation rights for reprints. All other permissions can be obtained through our RightsLink service via the Permissions link on the article page on our site—for further information please contact journals.permissions@oup.com.

Here, we introduce scMnT, a bioinformatic method for analyzing MSI tumor samples from scRNA-seq data. scMnT builds on our previously published microsatellite analysis software, NanoMnT, which accounts for the error profile of Oxford Nanopore sequencing platforms during microsatellite allele calling [17]. scMnT leverages microsatellite length information from reads mapped to mononucleotide-repeat microsatellite regions and groups them by their cell barcodes, enabling the identification of MSI cells at single-cell resolution and precise quantification of MSI intensity. Analysis of publicly available CRC datasets [18, 19] revealed both inter- and intra-tumoral variation in MSI intensity as well as transcriptomic properties associated with MSI intensity. Our work offers a novel method for quantifying MSI intensity and new opportunities analyzing MSI cancer cells at the single-cell level. scMnT is publicly available at GitHub: <https://github.com/18parkky/scMnT>.

Materials and methods

scRNA-seq data preprocessing

For the SG1 and KUL cohort of CRC patients from Joanito et al. [19], raw data were downloaded from the European Genome-phenome Archive (EGA) database under accession numbers [EGAD00001008555](#), [EGAD00001008584](#), and [EGAD00001008585](#). For the CRC patient cohort from Chen et al. [18], raw data were downloaded from the Genome Sequence Archive under accession number [PRJCA018173](#). Gene expression matrices were generated from the raw data using 10x Genomics Cell Ranger v8.0.1 [20], and author-provided cell type annotations were overlaid onto the expression matrices. Droplets with doublet score $> \mu + \sigma$, computed using Scrublet [21], were considered doublets and discarded. Furthermore, droplets with cell barcodes not present in the authors' annotations were excluded, as they had not passed quality control in the original studies.

The remaining droplets were processed using the Scanpy package [22] as follows: (i) droplets with fewer than 300 detected genes and genes expressed in fewer than 10 cells were removed; (ii) UMI counts were normalized using `scanpy.pp.normalize_total` and log-transformed with `scanpy.pp.log1p`; (iii) 2000 highly variable genes, selected using `scanpy.pp.highly_variable_function`, were used for principal component analysis (PCA); and (iv) 40 PCA components were used for UMAP with `scanpy.pp.neighbors` and `scanpy.pp.umap` [23, 24]. For the cancer cell line dataset from Kinker et al. [25], which employed a custom multiplexing strategy, the gene expression matrix was downloaded from the Gene Expression Omnibus (GEO) (accession number [GSE157220](#)) and was used for subsequent analyses.

Reconstruction of single-cell microsatellite allele profiles

Microsatellite allele information was summarized at the read level using the `getAlleleTable` command from NanoMnT v1.0.0, which requires a list of target microsatellite regions and a BAM file as input. A- and T-repeat mononucleotide microsatellite regions ranging from ≥ 10 to ≤ 24 nucleotides in length were identified in the GRCh38 reference genome using `Search for SSR` function in Krait v1.3.3 [26]. BAM files tagged with cell barcodes and unique molecular identifiers generated by Cell Ranger (`possorted_genome_bam.bam`) were used as input for NanoMnT. Reads containing indel errors within the flanking regions of microsatellites were discarded. The remaining

reads were grouped by cell barcodes, and for each cell, the average and standard deviation of the relative microsatellite allele sizes were calculated. These cell-level values were then overlaid onto the preprocessed Scanpy objects for downstream analyses.

scMnT benchmarking analysis

To assess the minimum number of microsatellite loci required per cell to reliably assign an MSI score, we performed a sensitivity, specificity, and precision analysis using the cell-type annotated data of patient CRC2783 from the CRC-SG1 cohort [19]. For each minimum locus threshold ranging from 1 to 50, cells were included in the analysis only if the number of detected STR loci was greater than or equal to the specified threshold. The MSI score cutoff was defined as the 95th percentile of MSI scores from non-epithelial cells, representing the background noise level of MSS populations. Cells with MSI scores above this cutoff were classified as predicted MSI, while those below were classified as predicted MSS. Sensitivity, specificity, and precision were subsequently calculated and visualized.

scMnT's scalability was evaluated using CRC2783 data comprising six samples. The analysis was performed with 4, 8, 16, 32, and 64 CPU cores on a CentOS 7 system with an Intel(R) Xeon(R) Gold 6240 CPU (2.60 GHz).

For comparison with MSIsensor-pro, we applied scMnT to patient 20 (P20) from the Chen et al. dataset [18], as this patient provides a representative MSI tumor profile. Cells underwent Scanpy preprocessing as described in the 'scRNA-seq data preprocessing' Methods section, and subjected to Leiden clustering using `scanpy.tl.leiden` with resolution value set to 0.1, and tumor cells were identified. The BAM files were subset into two, one for normal cells and the other for tumor cells. Each BAM file was processed using the following MSIsensor-pro's `pro` command with default parameters.

SNV analysis using SComatic

Precomputed cell type annotations and Cell Ranger output BAM files from tumor samples of patient SC035 were prepared as inputs for SComatic. The following SComatic scripts were executed in order: `SplitBamCellTypes.py`, `BaseCellCounter.py`, `MergeBaseCellCounts.py`, `BaseCellCalling.step1.py`, `BaseCellCalling.step2.py`. SNVs that passed all filters ('Multiple_cell_types,' 'LC_Upstream,' 'Clustered,' 'LC_Downstream,' 'Min_cell_types,' 'Cell_type_noise,' 'Noisy_site,' 'Multi-allelic') were then processed using SComatic script `SingleCellGenotype.py`. SNVs supported by fewer than three reads or represented in fewer than 100 cells were discarded. The zygosity of each SNV locus was assessed by examining alleles reported by normal cells. Monoallelic loci (i.e. all normal cells reporting a single allele) were selected for downstream analysis. To compare the allele frequencies across the four cell groups in patient SC035 – normal cells and tumor cells from Leiden clusters 0–2 (as shown in [Fig. 2d](#) and [2e](#))– only SNVs represented in at least five cells per group were retained, resulting in 134 SNV loci. Hierarchical clustering was performed using the `clustermap` function from the Python package Seaborn [27], which internally uses Scipy's `cluster.hierarchy.linkage` function [28]. Finally, SNVs whose alternate allele frequency exceeded 0.1 in any of the tumor Leiden cluster were visualized ([Supplementary Fig. S4d](#)). Allele frequencies for all four groups across each SNV locus are provided in [Supplementary Table S4](#).

Classification of patients based on their MSI intensity

MSI CRC patients were classified into three groups—MSI-LI, MSI-MI, or MSI-HI—based on the z-score standardized average MSI scores of their cancer cells. Patients with average scores below -0.5 were classified as MSI-LI, those above 0.5 as MSI-HI, and those with scores between -0.5 and 0.5 as MSI-MI. The average score of each patient is available in [Supplementary Table S2](#).

Identification of differentially expressed genes

Differentially expressed genes (DEGs) between MSI-LI and MSI-HI groups were identified using Scanpy's `rank_genes_groups` function with the Wilcoxon method. Genes with a log fold change >1 or <-1 and an adjusted $P \leq .01$ were considered enriched in MSI-HI or MSI-LI, respectively. Genes meeting these criteria in both the Joanito et al. and Chen et al. datasets ([Fig. 3b](#)) were selected for further analysis.

ORA and GSEA analysis

Both over-representation analysis (ORA) and gene set enrichment analysis (GSEA) were performed using the GSEAPy package [29]. For ORA, we used EnrichR [30] with CellMarker_2024 [31] as the reference database to identify enriched cell-type markers. GSEA was conducted using curated gene sets derived from two studies [19, 32] and CellMarker 2024 ([Supplementary Table S5](#)).

Characterization of DEGs using human protein atlas data

The expression of selected DEGs identified in both datasets ([Fig. 3a](#)) was visualized across enterocytes, goblet cells, and intestinal stem cells using the Human Protein Atlas 'Single Cell Type' dataset [33]. For each gene, normalized transcript per million (nTPM) values were aggregated by cell type and converted into relative proportions to assess its expression distribution across the three cell types.

Regression and statistical analysis

Linear regression analysis and statistical testing (Spearman correlation, Pearson correlation, and Kolmogorov–Smirnov test) has been performed using the Python package Seaborn (v0.13.2).

Data visualization

All data visualizations have been performed using the Python packages Seaborn and Matplotlib (v3.9.2) and were adjusted using Microsoft PowerPoints.

Results

Inference of MSI status in single-cell resolution

An overview of scMnT is provided in [Fig. 1a](#) (see Methods). First, sequencing reads mapped to mononucleotide-repeat microsatellite loci within the genome are collected and assigned to individual cells based on their cell barcodes, enabling the reconstruction of microsatellite profiles at single-cell resolution. Although mononucleotide repeats exhibit higher sequencing error rates compared to dinucleotide repeats, the resulting baseline noise is comparable across cell populations and does not distort relative MSI score differences

between cell types ([Supplementary Fig. S1](#)). Additionally, mononucleotide repeats were prioritized because of their substantially higher abundance compared to dinucleotide repeats within gene bodies (748,880 versus 154,246 in GRCh38 detected using Krait v1.3.3), which improves locus capture under the sparse and expression-constrained conditions of scRNA-seq. Next, the relative microsatellite allele size is measured using the GRCh38 reference allele as a surrogate for the host genotype because genotype information is typically unavailable from scRNA-seq data alone. While this approach does not provide locus-specific genotyping, we have previously observed that it is sufficient for inferring MSI status from sequencing data [17]. Subsequently, the average and standard deviation of relative allele sizes for each cell are calculated. MSI cells typically exhibit decreased average allele sizes and increased variability (higher standard deviation), reflecting the accumulation of deletion mutations in mononucleotide repeats [34]. In contrast, microsatellite stable (MSS) cells maintain allele size distributions that are similar to the reference genome, with low average and low standard deviation values. Finally, the MSI score is calculated as the product of the average and standard deviation, multiplied by -1 , integrating the two metrics into a single composite score. These scores are then used to infer the MSI status of individual cells or user-defined cell clusters. For the latter, scMnT performs this inference by statistically comparing the MSI score distribution of a target cluster (e.g. cancer cells) to that of normal, MSS cells (e.g. immune cells). The comparison is conducted using the Kolmogorov–Smirnov test, and the effect size is reported using Cliff's delta. Clusters with a statistically significant P -value (e.g. $<.01$) and effect size (e.g. $>.4$) are considered MSI.

Applying scMnT to a publicly available cancer cell line scRNA-seq dataset [25] produced robust results, with MSI scores for each cell line aligning well with their known MSI status ([Fig. 1b](#), [Supplementary Fig. S2a](#), [Supplementary Table S1](#)). As expected, we found that both the average and standard deviation values provide valuable information about the MSI status of single cells ([Fig. 1c](#)). MSI cell lines exhibited varying degrees of MSI intensity, whereas MSS cell lines maintained relatively stable MSI intensity near zero, indicating that the observed variability of MSI scores among MSI cell lines likely represent true biological differences ([Supplementary Fig. S2b](#)). Notably, cancer cell lines known to be MSI-L, which is clinically considered to be the same as MSS [1], exhibited varying MSI scores that evidently resembled MSI scores of either MSI or MSS.

While users can test MSI in user-defined clusters of interest, the MSI status of individual cells can also be inferred directly from their MSI scores. In such cases, using a threshold of 1.5 to distinguish MSI from MSS cells yielded the best performance in terms of sensitivity and specificity ([Supplementary Fig. S3a](#)). To determine the minimum number of microsatellite loci required per cell for confident MSI scoring, we performed sensitivity, specificity, and precision analyses using actual CRC patient data (CRC2783) from the CRC-SG1 cohort (Methods, [Supplementary Fig. S3b](#)). All three metrics increased as the minimum number of microsatellite loci per cell was raised, indicating that stricter thresholding leads to more reliable MSI assignments. However, it is important to note that the optimal thresholds likely may vary depending on the single-cell chemistry and sequencing instrument, as these factors determine the number of PCR cycles applied to the cDNA molecules, which are known to introduce significant amount of noise known as PCR stutter [35, 36]. In practice, we recommend users to fine-tune the MSI score threshold by manually inspecting the MSI score distribution between malignant cells and

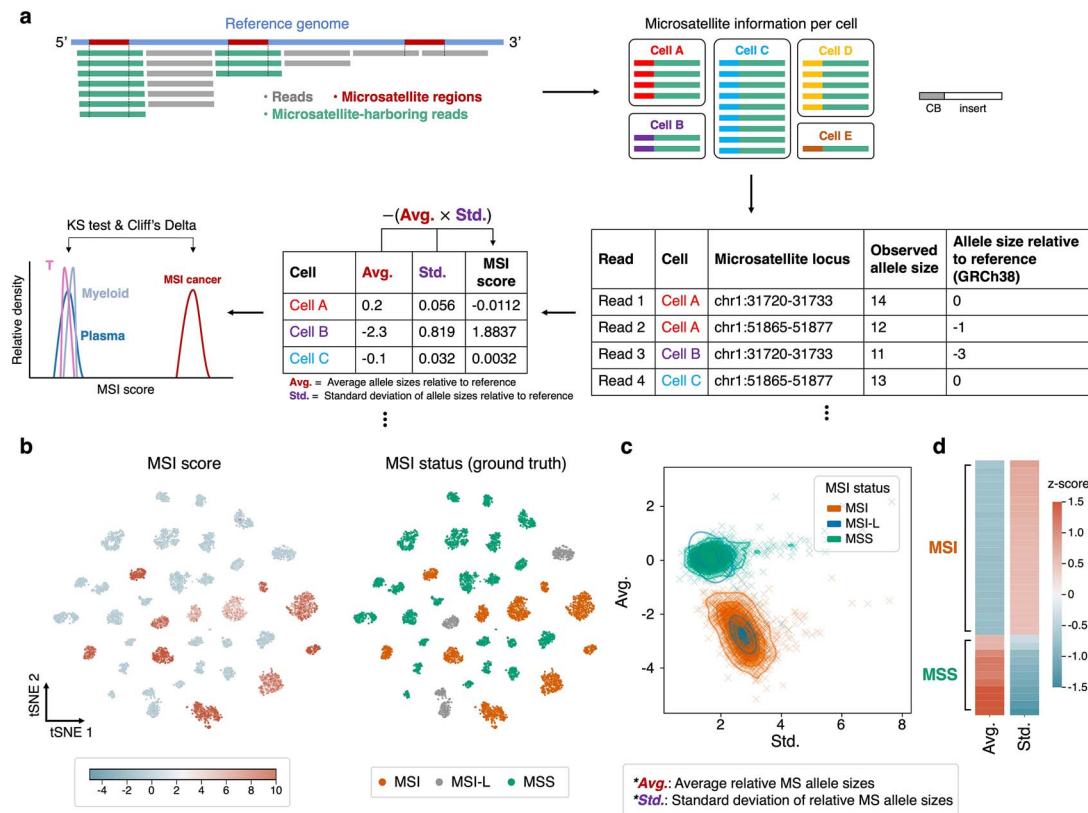


Figure 1 Overview of scMnT and its application to cancer cell line data. a, schematic diagram depicting the workflow of scMnT. From the BAM file generated by cell ranger, sequencing reads that map to mononucleotide-repeat microsatellite regions are grouped by cell barcode (CB). The allele size (i.e. length of the microsatellite) relative to the reference genome (GRCh38) is determined for each read, which is used to calculate the MSI score for each cell. Using user-provided cell type annotations, the MSI scores can be compared among cell types using the Kolmogorov–Smirnov (KS) test and Cliff's Delta values. b, t-SNE plot of cancer cells from a scRNA-seq dataset from Kinker et al. (2020) showing accurate MSI identification results from scMnT. Ground-truth MSI status for each cell line was obtained from Cellosaurus and literatures. c, scatter plot of cancer cells based on the standard deviation of their relative microsatellite allele sizes (x-axis) and their average relative microsatellite allele sizes (y-axis), demonstrating that both metrics contribute to describing the MSI status of cells. Each 'x' represents a cell. d, heatmap of each cancer cell line colored by their average- (left) and their standard deviation- (right) of the relative microsatellite allele sizes. Each bar represents a cancer cell line.

normal cells. Furthermore, scMnT efficiently utilizes multiple CPU cores for parallel processing, completing the analysis of six samples in 122 min with 32 cores (Supplementary Fig. S3c).

To ensure that the number of detected microsatellite loci per cell does not lead to bias in MSI score, we inspected the relationship between the two metrics (Supplementary Fig. S3d). Regardless of the number of loci detected, MSS cells consistently exhibited low MSI scores. In contrast, MSI cells showed a weak but significant correlation between MSI score and the number of detected loci. Nonetheless, this effect did not impact the overall MSI intensity of cell lines, indicating that MSI classification remains robust (Supplementary Fig. S3e).

Variations in MSI intensity among CRC patients

We applied our approach to a scRNA-seq dataset of 10 CRC patients from the CRC-SG1 cohort and KUL cohort [19]. UMAP visualization of cells after preprocessing with Scanpy [22] (Methods) revealed that all cells except for subsets of epithelial cells displayed low MSI scores (Fig. 2a, Supplementary Fig. S4a). Sub-clustering of epithelial cells revealed multiple patient-specific clusters—representing the malignant cells—as well as a single cluster containing cells from

multiple patients, which represent the normal intestinal epithelial cells. Again, substantial variation in MSI intensity was observed among the MSI patients, but not in MSS patients, further supporting the quantitative properties of MSI phenotype (Fig. 2b). Furthermore, MSI scores were markedly higher in longer microsatellites, as expected given that increased length is known to promote strand slippage. Notably, normal cells and MSS cancer cells also exhibited elevated MSI scores in microsatellites longer than 20 bp, likely due to PCR stutter artifacts, which are similarly correlated with microsatellite length.

In patient SC035, cancer cells initially formed two distinct clusters in the UMAP embedding (Fig. 2a), prompting a more detailed analysis by sub-clustering the cancer cells separately. A refined UMAP embedding revealed three distinct cancer cell clusters with varying MSI intensities (Fig. 2d, Supplementary Fig. S4b). Suspecting that these clusters may reflect polyclonal origins of CRC, we analyzed their single-nucleotide variants (SNVs) using Somatic [37]. Hierarchical clustering based on SNV features further supported the hypothesis that these clusters arose from distinct clones (Methods, Supplementary Fig. S4c, 2d). Moreover, we verified that this observation was not an artifact by applying MSIsensor-pro (using pseudo-bulks from each SC035 cancer subcluster as inputs), which produced results (Fig. 2e) concordant

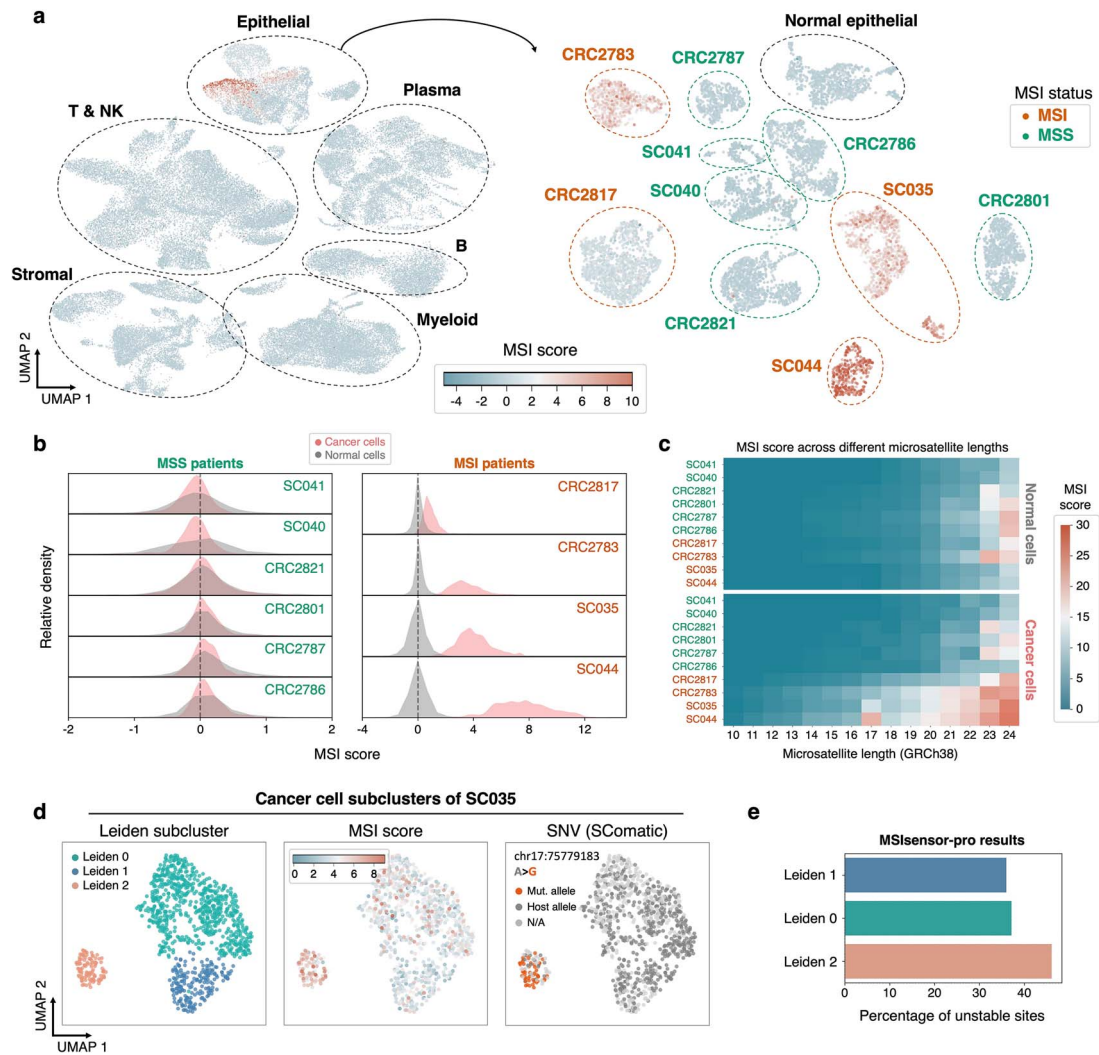


Figure 2 Inter- and intra-tumoral heterogeneity in MSI intensity across colorectal cancer patients. a, UMAP visualization of all cells (left) and epithelial cells (right) from a colorectal cancer (CRC) patient dataset from Joanito et al. (2022). Only cells with at least 10 reads mapped to microsatellite regions are shown, and cells are colored by their MSI scores. b, kernel density estimate (KDE) plot of MSI scores of normal cells (gray) and cancer cells (red). c, heatmap showing the average MSI scores across different microsatellite lengths in normal cells (top) and cancer cells (bottom). d, UMAP visualization of cancer cells from patient SC035, colored by their Leiden cluster (left), MSI scores (middle), and SNV (right). e, MSIsensor-pro results of each SC035 cancer cell Leiden cluster.

with those from scMnT (Supplementary Fig. S4e). These results suggest that intratumoral heterogeneity may extend to variations in MSI intensity.

Transcriptomic signatures correlated with MSI intensity

While high MSI intensity is associated with favorable prognosis and improved response to ICB treatments [6, 7], whether MSI intensity is driven by biological factors or determined randomly remains unclear. To investigate this, we sought transcriptomic signatures correlated with MSI intensity. We expanded the number of MSI patients for analysis by incorporating an additional comprehensive CRC scRNA-seq dataset from a recent study by Chen et al. [18], which consists of 10 MSI patients and includes patient metadata on ICB response and tumor regression ratios (Supplementary Fig. S5a, b).

Once more, we observed substantial variability in MSI intensity among CRC patients (Supplementary Fig. S5c). Patients were classified into MSI-LI, MSI-MI, and MSI-HI (not to be confused with traditional definition of MSI-L/MSI-H) based on their MSI intensity (Methods, Supplementary Fig. S5d, Supplementary Table S2), and differential gene expression analysis was performed between MSI-LI and MSI-HI cancer cells (Fig. 3a). Genes enriched for MSI-LI cancer cells included well known markers of goblet cells (e.g. *MUC2*, *ATOH1*) and enterocytes (e.g. *TFF1*, *PIGR*), whereas genes enriched for MSI-HI cancer cells were associated with intestinal stem cell signature (e.g. *ASCL2*, *RGMB*, *SOX4*, *LEF1*) (Supplementary Fig. S5e).

Using genes that were identified as DEGs in both datasets (Fig. 3b, Supplementary Table S3) as inputs, ORA produced similar results (Fig. 3c, Methods). We also noticed that MSI-LI cancer cells were characterized by high expression of S100 family genes, which are associated with fetal signatures [38, 39]. Given that CRC arising from sessile serrated lesions were found to be strongly associated with

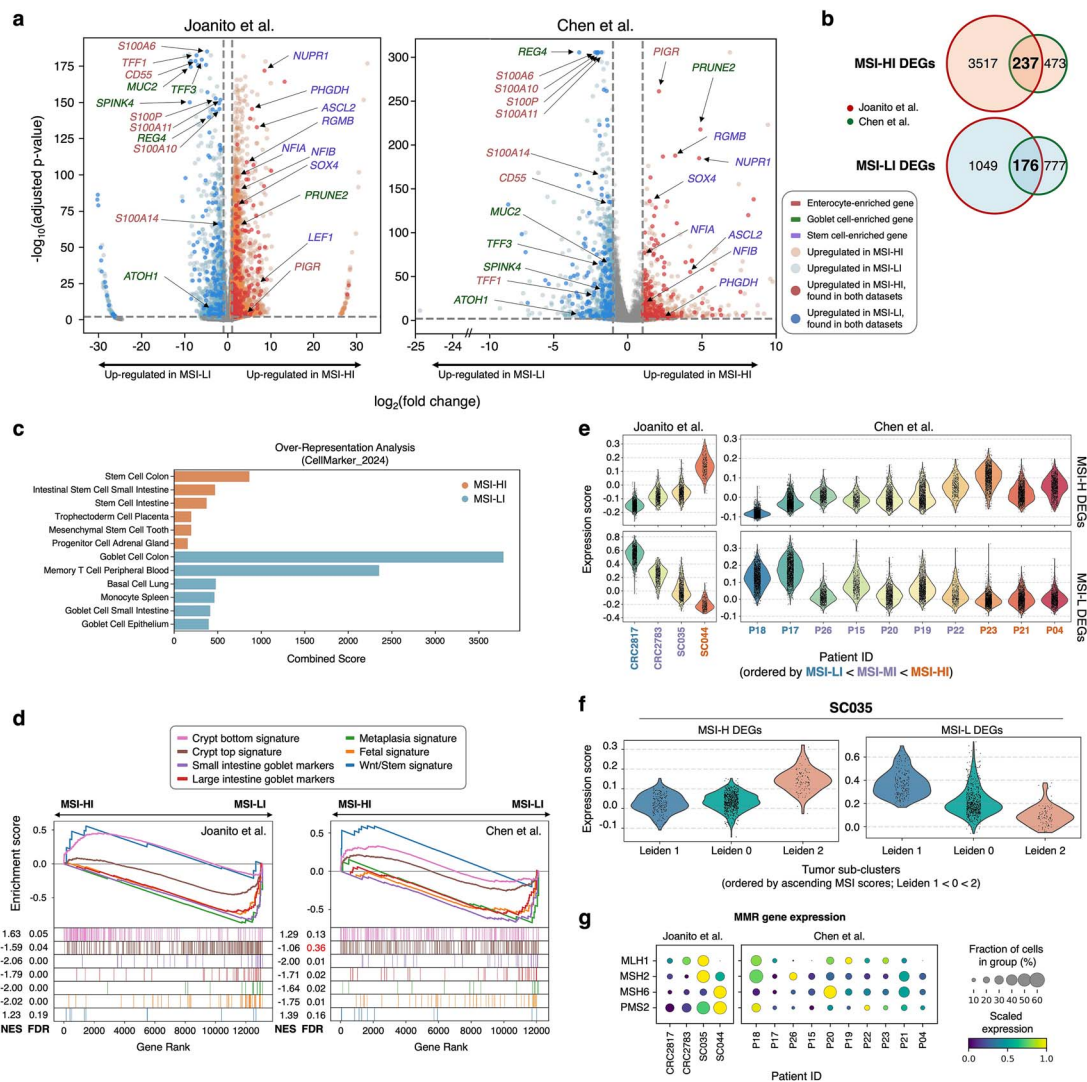


Figure 3 Tumor-intrinsic transcriptomic signatures associated with MSI intensity. **a**, volcano plot displaying DEGs in MSI-LI and MSI-HI cancer cells from the Joanito et al. (2022) dataset (left) and Chen et al. (2024) dataset (right). **b**, Venn diagram showing the number of significant DEGs identified in cancer cells from the two datasets. **c**, Bar plot visualizing the overrepresentation analysis (ORA) results for DEGs identified in cancer cells from both datasets. **d**, gene set enrichment analysis (GSEA) visualization of DEGs identified in cancer cells from both datasets, using gene sets from the literature. **e**, violin plot showing the average expression of MSI-HI and MSI-LI-associated DEGs in cancer cells across all patients. **f**, violin plot showing the average expression of MSI-HI and MSI-LI-associated DEGs in cancer cells of patient SC035. **g**, dot plot visualizing the expression of MMR genes in cancer cells across all patients.

gastric metaplasia and fetal signatures [40], we performed gene set enrichment analysis (GSEA) to test for relevant pathways (Fig. 3d, Methods). Interestingly, both metaplasia, fetal signatures and crypt top signatures were enriched in MSI-LI cancer cells, while Wnt/stem and crypt bottom signatures were enriched in MSI-HI cancer cells.

Next, we calculated module scores for MSI-HI DEGs and MSI-LI DEGs across cancer cells from individual patients and found a positive relationship between MSI-HI DEG module scores and MSI intensity, and an inverse relationship for MSI-LI DEG module scores (Fig. 3e). Notably, a similar pattern was observed within the cancer subclusters of patient SC035 (Fig. 3f). Additionally, we examined the expression of MLH1 and other MMR genes in cancer cells, as MLH1 loss—often due to promoter hypermethylation—is a hallmark of sporadic MSI CRC [41–43]. Strikingly, cancer cells of several patients displayed mRNA

expression of MLH1, whereas patients with the highest MSI intensity (SC044 and P04) did not (Fig. 3g). Other important MMR genes were detectable across all patients.

High MSI intensity is associated with better ICB response and higher baseline T cell abundance

In agreement with previous reports [6, 7], we also observed that patients who achieved complete response (CR) showed higher MSI scores than patients who had partial response (PR) (Fig. 4a). Furthermore, tumor regression ratios were also correlated with the patient MSI scores, though the robustness of correlation analysis was limited by the number of patients available.

Examination of baseline cellular abundance of various cell types within the tumor microenvironment (TME) across the three MSI

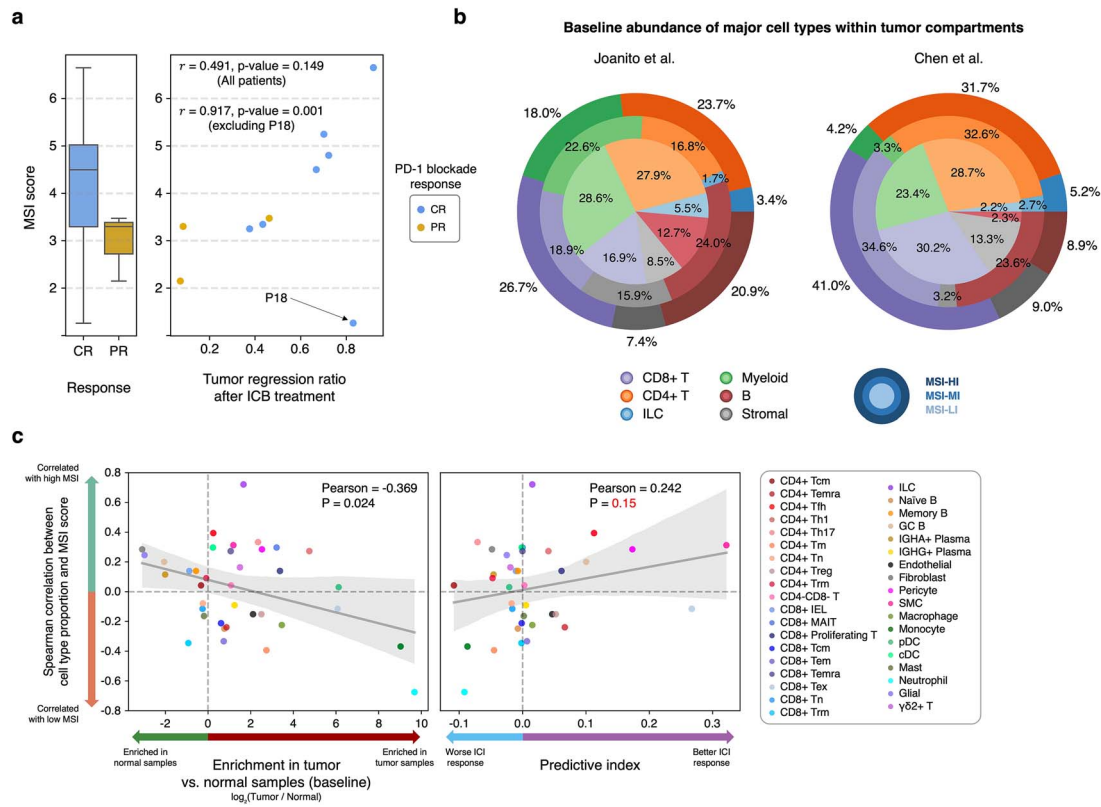


Figure 4 Association of MSI score with response to immune checkpoint blockade therapy and baseline abundance of various cell types. **a**, association of MSI score and response to ICB therapy. Box plot comparing the distribution of MSI scores (of cancer cells) between patients who displayed complete response versus patients who displayed partial response (left) and scatter plot displaying the relationship between the tumor regression ratio following ICB therapy and MSI score of each patient (right). **b**, comparison of baseline abundance of major cell types within tumor environments between MSI-LI (inner), MSI-MI (middle), and MSI-HI (outer) patients. **c**, scatter plots showing, for each cell type, the spearman correlation between its proportion and the MSI score (y-axis) against two variables: Sample enrichment (left, x-axis) and predictive index (right, x-axis).

intensity classes revealed cell types whose abundance correlated with MSI intensity (Fig. 4b). In both datasets, CD8+ T cells were more abundant in tumors with higher MSI intensity, following the order of MSI-HI > MSI-MI > MSI-LI, suggesting that elevated immune infiltration may underlie the improved ICB response seen in MSI-HI CRC. Additionally, MSI-HI and MSI-MI tumors had higher baseline proportions of B cells compared to MSI-LI tumors, consistent with previous findings linking B cell presence to enhanced immunotherapy responses [44–47]. Conversely, MSI-LI tumors displayed a greater abundance of myeloid cells relative to MSI-HI and MSI-MI tumors. No other cell types showed obvious associations with MSI intensity.

To further investigate these relationships, we leveraged the high-resolution cell-type annotations from Chen et al. [18] to explore associations between baseline TME composition and MSI intensity at a finer granularity (Fig. 4c). Tumor-enriched cell types (e.g. CD8+ Tex, CD4+ Treg, endothelial cells) were modestly enriched in tumors with lower MSI intensities, whereas normal tissue-enriched cell types (e.g. GC B, Memory B) were enriched in tumors with higher MSI intensities. Additionally, cell types enriched in tumors with higher MSI intensities exhibited a higher predictive index, which reflects the correlation between baseline cellular proportions and tumor size changes following ICB treatment [47]. This suggests that higher MSI intensities are associated with a TME favorable to ICB response, although the association did not reach statistical significance ($P = .15$).

Comparison of scMnT with MSIsensor-pro

Finally, we compared scMnT with MSIsensor-pro, a well-established software for detecting MSI status from next-generation sequencing (NGS) data, including RNA sequencing, whole-exome sequencing, whole-genome sequencing, and targeted sequencing [10]. MSIsensor-pro identifies unstable microsatellite loci from raw sequencing files (BAM files) and does not require a matched normal sample. For application to single-cell data, pseudobulk BAM files can be generated from clusters of interest (e.g. Leiden clusters) and analyzed individually using MSIsensor-pro. Our comparison shows that both tools can clearly distinguish MSI cells from normal cells, although with differing sensitivity (Supplementary Fig. S6). A key advantage of scMnT is that it does not require prior identification of cell populations of interest, whereas MSIsensor-pro relies on predefined pseudobulk groupings.

Discussion

In this study, we introduced scMnT, which is designed to (i) identify MSI cells at single-cell resolution and (ii) quantify MSI intensity by analyzing the mean and standard deviation of microsatellite allele lengths. Using scRNA-seq datasets of CRC cohorts, we validated the robustness of our method by accurately distinguishing MSI cancer cells from MSS cells and quantitatively measuring the MSI intensity

of individual cells. These results demonstrate that scMnT reliably detects MSI cells at single-cell resolution, enabling analysis of heterogeneous samples containing both MSI and MSS cells. scMnT also clarifies the MSI status of 'MSI-L' cancer cells, which are clinically classified as MSS, enabling definitive assessment of their microsatellite instability at the single-cell level. Our method also addresses a key challenge in scRNA-seq studies: annotating cancer cells. Unlike normal cell types, cancer cells often lack well-defined marker genes [48], complicating their identification—especially for MSI cancers, which typically do not exhibit common genomic features such as copy number variations [49]. scMnT thus provides a complementary strategy to improve cancer cell annotation in single-cell datasets. In summary, scMnT offers a robust and scalable framework for MSI detection in single-cell transcriptomic data, enhancing our ability to study intratumoral heterogeneity and cancer cell identity in MSI-driven tumors.

We also performed a series of analyses to elucidate biological factors associated with MSI intensity. Through comparative analyses of MSI-LI and MSI-HI cancer cells, we found that MSI-LI tumors exhibited strong fetal, metaplastic, crypt-bottom, and goblet cell signatures, whereas MSI-HI tumors were enriched for stem, Wnt, and crypt top signatures. These results suggest that MSI intensity may be shaped by biological factors rather than arising purely from stochastic processes. A previous comprehensive study reported that CRCs originating from differentiated intestinal epithelial cells (enterocytes and goblet cells) display strong gastric metaplasia signatures, whereas those arising from intestinal stem cells are characterized by prominent Wnt signaling [40]. In light of this context, our results may indicate a potential link between the cell of origin and MSI intensity.

As others have reported, we corroborate the existence of inter-tumoral heterogeneity in MSI intensity among CRC patients and its correlation with ICB treatment response. However, notable variations persist. For instance, patient 18 from the Chen et al. dataset exhibited a high tumor regression ratio despite showing the lowest MSI intensity (Fig. 4a). It is important to recognize that therapeutic outcomes are influenced by multiple factors beyond MSI intensity, including HLA genotype [50, 51], baseline immune infiltration [18, 52], defects in antigen presentation machinery [53, 54], Wnt pathway activation [55], presence of immunosuppressive cells [56, 57], and the degree of intratumoral heterogeneity [58, 59]. Nonetheless, patients with high MSI intensity generally exhibit better responses to ICB treatment and therefore should be prioritized for ICB treatment. This is of particular clinical relevance since around half of MSI patients fail to achieve a durable response to ICB treatment [58]. A more nuanced, quantitative understanding of MSI intensity—rather than relying on a binary classification—may help improve patient selection for immunotherapy. Furthermore, we emphasize that MSI intensity is not synonymous with tumor mutational burden (TMB). Although both metrics reflect the somatic mutational load, MSI intensity specifically captures indel-driven frameshift mutations, which are more likely to produce novel immunogenic neoantigens. In contrast, TMB accounts for both single nucleotide variants (SNVs) and indels, with SNVs often generating mutations less likely to elicit strong immune responses.

A particularly intriguing observation from our analysis was the detectable mRNA expression of MLH1 in MSI cancer cells of several MSI patients (Fig. 3g). While loss-of-function or hypomorphic

mutations in MMR genes may explain this observation, such mutations are relatively uncommon in sporadic MSI CRC. Instead, most cases arise from epigenetic silencing of the MLH1 promoter via hypermethylation [60–62]. Thus, the presence of MLH1 transcripts suggests that MMR deficiency may not be a fixed state, but rather a reversible or heterogeneous phenotype. This raises the possibility that the duration or timing of MMR deficiency could influence accumulation of MSI events and therefore drive the intertumoral variation of MSI intensity of MSI tumors.

We acknowledge several limitations of this study. scMnT requires the raw sequencing files (e.g. FASTQ or BAM) and relies on a small subset of reads that align to microsatellite regions within intragenic areas, which may not be consistently captured across different cancer types. We recommend users without access to raw sequencing files to use transcriptome-based approaches, such as MSIsensor-RNA [15]. Furthermore, while scMnT demonstrates accurate MSI detection using mononucleotide-repeat loci, comparable performance could not be achieved with dinucleotide-repeat loci. Given the potential relevance of dinucleotide repeats in the context of WRN inhibitor sensitivity for MSI cancers [63], this represents a limitation in those specific applications. Future works will be aimed to identify dinucleotide repeat loci that are suitable for these purposes. Additionally, although we have proposed several biological hypotheses—including a potential link between MSI intensity and CRC biology, as well as the possibility of MMR deficiency reversion—these findings were derived retrospectively from a relatively small patient cohort. Therefore, prospective validation in larger cohorts will be necessary to substantiate these observations. Another important limitation is that we have not evaluated scMnT on datasets generated using single-cell protocols other than those from 10x Genomics. This represents a relevant limitation, as different single-cell chemistries employ distinct strategies for mRNA capture and cDNA amplification—most notably the use of poly(A)-specific versus random priming—which can result in substantially different pools of captured molecules and, consequently, distinct sets of microsatellite loci. Therefore, scMnT in its current form should be used for analyzing scRNA-seq data generated using 10x Genomics chemistries.

While we demonstrated that scMnT reliably identifies MSI cells using the Kolmogorov–Smirnov test between user-defined cell groups (e.g. cell types or Leiden clusters), we were unable to establish a precise MSI score threshold for identifying patients eligible for ICB therapy. This threshold is likely to vary depending on sequencing protocols (e.g. library preparation chemistry and sequencing platform) and cancer type. For instance, 10x 3' single-cell chemistries preferentially capture transcripts containing microsatellite loci located in the final exon, whereas 5' chemistries are enriched for loci near transcription start sites. These differences cause each chemistry type to cover largely distinct sets of microsatellite loci, resulting in divergent MSI score profiles. In addition, different sequencing platforms can introduce unique artifact patterns [62], adding another layer of variability to scMnT analysis. Nonetheless, despite the datasets analyzed in this study being generated using different chemistries and platforms (Kinker et al.: 10x Chromium Single Cell 3' v2, Illumina NextSeq; Joanito et al.: 10x Chromium Single Cell 5', Illumina NextSeq 500, HiSeq 4000 or NovaSeq 6000, and Chen et al.: 10x Chromium Single Cell 5', Illumina NovaSeq), scMnT consistently produced reliable and biologically coherent results. This suggests that scMnT is robust across diverse library preparation chemistries and sequencing configurations. Even

so, future studies should aim to standardize sequencing parameters within each cancer type to establish consistent MSI score thresholds and enable cross-study comparability.

Moreover, given the strong performance of scMnT on the dataset from Kinker et al., the method can be readily applied to other malignancies characterized by frequent MSI, such as endometrial, breast, and ovarian cancers, where single cell sequencing data are increasingly available. Extending scMnT to these tumor types may help reveal cancer-specific patterns of MSI heterogeneity and offer broader insights into the role of microsatellite instability across diverse oncogenic contexts. Finally, with the continued advancement of long-read sequencing technologies and its application to single-cell transcriptomics, single-cell analysis of various genomic events, including MSI, is expected to become more accessible. When combined with analyses uniquely enabled by long-read sequencing (such as isoform-level characterization [64] and transposable element profiling [65]) these approaches will provide valuable insights into cellular heterogeneity and genome dynamics. We anticipate that scMnT will serve as a foundation for future studies investigating the biological relevance of MSI heterogeneity and its impact on cancer evolution and immunotherapy response.

Key Points

- Microsatellite instability (MSI) is better characterized as a continuous phenotype, which scMnT can measure from single cell RNA-sequencing data.
- MSI intensity measured by scMnT can be used to detect MSI cells in single-cell resolution and is positively correlated with response to immune checkpoint inhibitor therapy.
- scMnT reveals substantial intertumoral and intratumoral tumor heterogeneity of MSI intensity.
- Cancer cells with high MSI intensity exhibit high stem signatures, whereas those with low MSI intensity exhibit high enterocyte and goblet cell signatures.
- Tumors with high MSI intensity show higher proportions of lymphoid cells and lower proportions of myeloid cells compared to tumors with low MSI intensity.

Acknowledgements

We thank the authors of the two studies for granting access to the raw sequencing data from their cohorts. This study was supported by the National Research Foundation of Korea (NRF) (RS-2024-00335026, RS-2024-00403622), Korea Technology and Information Promotion Agency for SMEs (RS-2024-00506966), and a grant of the Korea-US Collaborative Research Fund (KUCRF), funded by the Ministry of Science and ICT and Ministry of Health & Welfare, Republic of Korea (RS-2024-00466906).

Author contributions

Both G.P. and J.P. conceptualized the study and prepared the original manuscript. G.P. designed and implemented scMnT, led the formal analysis and visualization. J.P. supervised the study and reviewed the manuscript.

Supplementary material

Supplementary material is available at *Briefings in Bioinformatics* online.

Data availability

All raw sequencing data analyzed in this study are publicly available in the European Genome-phenome Archive (EGA) database (accession number [EGAD00001008555](#) and [EGAD00001008585](#)), Genome Sequence Archive (accession number [PRJCA08173](#)), and Gene Expression Omnibus (GEO) (accession number [GSE157220](#)). The processed Scanpy object files are available on Zenodo (<https://zenodo.org/records/15478777>).

Code availability

Codes for scMnT and tutorials for running scMnT were deposited at GitHub and are publicly available (<https://github.com/18parkky/scMnT>). Codes used for the analyses in this study were also deposited at GitHub and are available (https://github.com/18parkky/CRC_MSI_intensity_analysis or <https://zenodo.org/records/15422474>).

References

1. Li K, Luo H, Huang L, *et al.* Microsatellite instability: a review of what the oncologist should know. *Cancer Cell Int* 2020;**20**:16. <https://doi.org/10.1186/s12935-019-1091-8>.
2. Dudley JC, Lin MT, le DT, *et al.* Microsatellite instability as a biomarker for PD-1 blockade. *Clin Cancer Res* 2016;**22**:813–20. <https://doi.org/10.1158/1078-0432.CCR-15-1678>.
3. Chang L, Chang M, Chang HM, *et al.* Microsatellite instability: a predictive biomarker for cancer immunotherapy. *Appl Immunohistochem Mol Morphol* 2018;**26**:e15–21. <https://doi.org/10.1097/PAI.0000000000000575>.
4. Le DT, Durham JN, Smith KN, *et al.* Mismatch repair deficiency predicts response of solid tumors to PD-1 blockade. *Science* 2017;**357**:409–13. <https://doi.org/10.1126/science.aan6733>.
5. Marabelle A, O'Malley DM, Hendifar AE, *et al.* Pembrolizumab in microsatellite-instability-high and mismatch-repair-deficient advanced solid tumors: updated results of the KEYNOTE-158 trial. *Nat Cancer* 2025;**6**:253–8. <https://doi.org/10.1038/s43018-024-00894-y>.
6. Mandal R, Samstein RM, Lee KW, *et al.* Genetic diversity of tumors with mismatch repair deficiency influences anti-PD-1 immunotherapy response. *Science* 2019;**364**:485–91. <https://doi.org/10.1126/science.aau0447>.
7. Wang QX, Qu CH, Gao YH, *et al.* The degree of microsatellite instability predicts response to PD-1 blockade immunotherapy in mismatch repair-deficient/microsatellite instability-high colorectal cancers. *Exp Hematol Oncol* 2021;**10**:2. <https://doi.org/10.1186/s40164-020-00193-z>.
8. Ziegler J, Hechtman JF, Rana S, *et al.* A deep multiple instance learning framework improves microsatellite instability detection from tumor next generation sequencing. *Nat Commun* 2025;**16**:136. <https://doi.org/10.1038/s41467-024-54970-z>.
9. Han XY, Zhang S, Zhou DC, *et al.* MSIsensor-ct: microsatellite instability detection using cfDNA sequencing data. *Brief Bioinform* 2021;**22**:bbaa402. <https://doi.org/10.1093/bib/bbaa402>.
10. Jia P, Yang X, Guo L, *et al.* MSIsensor-pro: fast, accurate, and matched-normal-sample-free detection of microsatellite instability. *Genomics*

- Proteomics Bioinformatics* 2020;**18**:65–71. <https://doi.org/10.1016/j.gpb.2020.02.001>.
11. Kautto EA, Bonneville R, Miya J, *et al.* Performance evaluation for rapid detection of pan-cancer microsatellite instability with MAN-TIS. *Oncotarget* 2017;**8**:7452–63. <https://doi.org/10.18632/oncotarget.13918>.
 12. Luo H, Xia X, Huang LB, *et al.* Pan-cancer single-cell analysis reveals the heterogeneity and plasticity of cancer-associated fibroblasts in the tumor microenvironment. *Nat Commun* 2022;**13**:6619. <https://doi.org/10.1038/s41467-022-34395-2>.
 13. Li PH, Kong XY, He YZ, *et al.* Recent developments in application of single-cell RNA sequencing in the tumour immune microenvironment and cancer therapy. *Mil Med Res* 2022;**9**:52. <https://doi.org/10.1186/s40779-022-00414-y>.
 14. Bae SG, Kim HJ, Kim MY, *et al.* Identification of cell type-specific effects of DNMT3A mutations on relapse in acute myeloid leukemia. *Mol Cells* 2023;**46**:611–26. <https://doi.org/10.14348/molcells.2023.0093>.
 15. Jia P, Yang X, Yang X, *et al.* MSIsensor-RNA: microsatellite instability detection for bulk and single-cell gene expression data. *Genomics Proteomics Bioinformatics* 2024;**22**:qzae004. <https://doi.org/10.1093/gpbjnl/qzae004>.
 16. Shao DD, Kriz AJ, Snellings DA, *et al.* Advances in single-cell DNA sequencing enable insights into human somatic mosaicism. *Nat Rev Genet* 2025;**26**:761–74. <https://doi.org/10.1038/s41576-025-00832-3>.
 17. Park G, An H, Luo H, *et al.* NanoMnT: An STR analysis tool for Oxford nanopore sequencing data driven by a comprehensive analysis of error profile in STR regions. *Gigascience* 2025;**14**:giaf013. <https://doi.org/10.1093/gigascience/giaf013>.
 18. Chen Y, Wang D, Li Y, *et al.* Spatiotemporal single-cell analysis decodes cellular dynamics underlying different responses to immunotherapy in colorectal cancer. *Cancer Cell* 2024;**42**:1268–1285.e7. <https://doi.org/10.1016/j.ccell.2024.06.009>.
 19. Joanito I, Wirapati P, Zhao N, *et al.* Single-cell and bulk transcriptome sequencing identifies two epithelial tumor cell states and refines the consensus molecular classification of colorectal cancer. *Nat Genet* 2022;**54**:963–75. <https://doi.org/10.1038/s41588-022-01100-4>.
 20. Zheng GX, Terry JM, Belgrader P, *et al.* Massively parallel digital transcriptional profiling of single cells. *Nat Commun* 2017;**8**:14049. <https://doi.org/10.1038/ncomms14049>.
 21. Wolock SL, Lopez R, Klein AM. Scrublet: computational identification of cell doublets in single-cell transcriptomic data. *Cell Syst* 2019;**8**:281–291.e9. <https://doi.org/10.1016/j.cels.2018.11.005>.
 22. Wolf FA, Angerer P, Theis FJ. SCANPY: Large-scale single-cell gene expression data analysis. *Genome Biol* 2018;**19**:15. <https://doi.org/10.1186/s13059-017-1382-0>.
 23. Kim GD, Lim C, Park J. A practical handbook on single-cell RNA sequencing data quality control and downstream analysis. *Mol Cells* 2024;**47**:100103. <https://doi.org/10.1016/j.mocell.2024.100103>.
 24. Kim GD, Shin SI, Jung SW, *et al.* Cell type- and age-specific expression of lncRNAs across kidney cell types. *J Am Soc Nephrol* 2024;**35**:870–85. <https://doi.org/10.1681/ASN.0000000000000354>.
 25. Kinker GS, Greenwald AC, Tal R, *et al.* Pan-cancer single-cell RNA-seq identifies recurring programs of cellular heterogeneity. *Nat Genet* 2020;**52**:1208–18. <https://doi.org/10.1038/s41588-020-00726-6>.
 26. Du L, Zhang C, Liu Q, *et al.* Krait: An ultrafast tool for genome-wide survey of microsatellites and primer design. *Bioinformatics* 2018;**34**:681–3. <https://doi.org/10.1093/bioinformatics/btx665>.
 27. Waskom ML. Seaborn: Statistical data visualization. *J Open Source Softw* 2021;**6**:3021. <https://doi.org/10.21105/joss.03021>.
 28. Virtanen P, Gommers R, Oliphant TE, *et al.* SciPy 1.0: fundamental algorithms for scientific computing in python. *Nat Methods* 2020;**17**:261–72. <https://doi.org/10.1038/s41592-019-0686-2>.
 29. Fang Z, Liu X, Peltz G. GSEAPy: a comprehensive package for performing gene set enrichment analysis in python. *Bioinformatics* 2023;**39**:btac757. <https://doi.org/10.1093/bioinformatics/btac757>.
 30. Xie Z, Bailey A, Kuleshov MV, *et al.* Gene set knowledge discovery with Enrichr. *Curr Protoc* 2021;**1**:e90. <https://doi.org/10.1002/cpz1.90>.
 31. Hu C, Li T, Xu Y, *et al.* CellMarker 2.0: An updated database of manually curated cell markers in human/mouse and web tools based on scRNA-seq data. *Nucleic Acids Res* 2023;**51**:D870–6. <https://doi.org/10.1093/nar/gkac947>.
 32. Kosinski C, Li VSW, Chan ASY, *et al.* Gene expression patterns of human colon tops and basal crypts and BMP antagonists as intestinal stem cell niche factors. *Proc Natl Acad Sci U S A* 2007;**104**:15418–23. <https://doi.org/10.1073/pnas.0707210104>.
 33. Alam O. A single-cell-type transcriptomics map of human tissues. *Nat Genet* 2021;**53**:1275.
 34. Aska EM, Zagidullin B, Pitkanen E, *et al.* Single-cell mononucleotide microsatellite analysis reveals differential insertion-deletion dynamics in mouse T cells. *Front Genet* 2022;**13**:913163. <https://doi.org/10.3389/fgene.2022.913163>.
 35. Kistler L, Johnson SM, Irwin MT, *et al.* A massively parallel strategy for STR marker development, capture, and genotyping. *Nucleic Acids Res* 2017;**45**:e142. <https://doi.org/10.1093/nar/gkx574>.
 36. Raz O, Biezuner T, Spiro A, *et al.* Short tandem repeat stutter model inferred from direct measurement of in vitro stutter noise. *Nucleic Acids Res* 2019;**47**:2436–45. <https://doi.org/10.1093/nar/gky1318>.
 37. Muyas F, Sauer CM, Valle-Inclán JE, *et al.* De novo detection of somatic mutations in high-throughput single-cell profiling data sets. *Nat Biotechnol* 2024;**42**:758–67. <https://doi.org/10.1038/s41587-023-01863-z>.
 38. Park YK, Franklin JL, Settle SH, *et al.* Gene expression profile analysis of mouse colon embryonic development. *Genesis* 2005;**41**:1–12. <https://doi.org/10.1002/gene.20088>.
 39. Mustata RC, Vasile G, Fernandez-Vallone V, *et al.* Identification of Lgr5-independent spheroid-generating progenitors of the mouse fetal intestinal epithelium. *Cell Rep* 2013;**5**:421–32. <https://doi.org/10.1016/j.celrep.2013.09.005>.
 40. Chen B, Scurrah C'R, McKinley ET, *et al.* Differential pre-malignant programs and microenvironment chart distinct paths to malignancy in human colorectal polyps. *Cell* 2021;**184**:6262–6280.e26. <https://doi.org/10.1016/j.cell.2021.11.031>.
 41. Simpkins SB, Bocker T, Swisher EM, *et al.* MLH1 promoter methylation and gene silencing is the primary cause of microsatellite instability in sporadic endometrial cancers. *Hum Mol Genet* 1999;**8**:661–6. <https://doi.org/10.1093/hmg/8.4.661>.
 42. Tariq K, Ghias K, Tariq K, *et al.* Colorectal cancer carcinogenesis: a review of mechanisms. *Cancer Biol Med* 2016;**13**:120–35. <https://doi.org/10.20892/j.issn.2095-3941.2015.0103>.
 43. Gkekas I, Jan N, Kaprio T, *et al.* Sporadic deficient mismatch repair in colorectal cancer increases the risk for non-colorectal malignancy: a European multicenter cohort study. *J Surg Oncol* 2024;**129**:1295–304. <https://doi.org/10.1002/jso.27619>.
 44. Cabrita R, Lauss M, Sanna A, *et al.* Tertiary lymphoid structures improve immunotherapy and survival in melanoma. *Nature* 2020;**577**:561–5. <https://doi.org/10.1038/s41586-019-1914-8>.

45. Petitprez F, de Reyniès A, Keung EZ, *et al.* B cells are associated with survival and immunotherapy response in sarcoma. *Nature* 2020;**577**:556–60. <https://doi.org/10.1038/s41586-019-1906-8>.
46. Helmink BA, Reddy SM, Gao J, *et al.* B cells and tertiary lymphoid structures promote immunotherapy response. *Nature* 2020;**577**:549–55. <https://doi.org/10.1038/s41586-019-1922-8>.
47. Zhang Y, Chen H, Mo H, *et al.* Single-cell analyses reveal key immune cell subsets associated with response to PD-L1 blockade in triple-negative breast cancer. *Cancer Cell* 2021;**39**:1578–1593.e8. <https://doi.org/10.1016/j.ccell.2021.09.010>.
48. Fan J, Slowikowski K, Zhang F. Single-cell transcriptomics in cancer: computational challenges and opportunities. *Exp Mol Med* 2020;**52**:1452–65. <https://doi.org/10.1038/s12276-020-0422-0>.
49. Cornish AJ, Gruber AJ, Kinnarsley B, *et al.* The genomic landscape of 2,023 colorectal cancers. *Nature* 2024;**633**:127–36. <https://doi.org/10.1038/s41586-024-07747-9>.
50. Chowell D, Morris LGT, Grigg CM, *et al.* Patient HLA class I genotype influences cancer response to checkpoint blockade immunotherapy. *Science* 2018;**359**:582–7. <https://doi.org/10.1126/science.aao4572>.
51. Naranbhai V, Viard M, Dean M, *et al.* HLA-A*03 and response to immune checkpoint blockade in cancer: An epidemiological biomarker study. *Lancet Oncol* 2022;**23**:172–84. [https://doi.org/10.1016/S1470-2045\(21\)00582-9](https://doi.org/10.1016/S1470-2045(21)00582-9).
52. Mispelbaum R, Hattenhauer ST, Held SAE, *et al.* Baseline immune signature score of Tregs x HLA-DR(+)CD4(+) T cells x PD1(+)CD8(+) T cells predicts outcome to immunotherapy in cancer patients. *Front Immunol* 2022;**13**:1054161. <https://doi.org/10.3389/fimmu.2022.1054161>.
53. Gettinger S, Choi J, Hastings K, *et al.* Impaired HLA class I antigen processing and presentation as a mechanism of acquired resistance to immune checkpoint inhibitors in lung cancer. *Cancer Discov* 2017;**7**:1420–35. <https://doi.org/10.1158/2159-8290.CD-17-0593>.
54. Sade-Feldman M, Jiao YJ, Chen JH, *et al.* Resistance to checkpoint blockade therapy through inactivation of antigen presentation. *Nat Commun* 2017;**8**:1136. <https://doi.org/10.1038/s41467-017-01062-w>.
55. Luke JJ, Bao R, Sweis RF, *et al.* WNT/beta-catenin pathway activation correlates with immune exclusion across human cancers. *Clin Cancer Res* 2019;**25**:3074–83. <https://doi.org/10.1158/1078-0432.CCR-18-1942>.
56. Lasser SA, Ozbay Kurt FG, Arkhypov I, *et al.* Myeloid-derived suppressor cells in cancer and cancer therapy. *Nat Rev Clin Oncol* 2024;**21**:147–64. <https://doi.org/10.1038/s41571-023-00846-y>.
57. Yang D, Liu J, Qian H, *et al.* Cancer-associated fibroblasts: from basic science to anticancer therapy. *Exp Mol Med* 2023;**55**:1322–32. <https://doi.org/10.1038/s12276-023-01013-0>.
58. Westcott PMK, Muiyas F, Hauck H, *et al.* Mismatch repair deficiency is not sufficient to elicit tumor immunogenicity. *Nat Genet* 2023;**55**:1686–95. <https://doi.org/10.1038/s41588-023-01499-4>.
59. Lin Z, Meng X, Wen J, *et al.* Intratumor heterogeneity correlates with reduced immune activity and worse survival in melanoma patients. *Front Oncol* 2020;**10**:596493. <https://doi.org/10.3389/fonc.2020.596493>.
60. Weisenberger DJ, Siegmund KD, Campan M, *et al.* CpG island methylator phenotype underlies sporadic microsatellite instability and is tightly associated with BRAF mutation in colorectal cancer. *Nat Genet* 2006;**38**:787–93. <https://doi.org/10.1038/ng1834>.
61. Boland CR, Goel A. Microsatellite instability in colorectal cancer. *Gastroenterology* 2010;**138**:2073–2087.e3. <https://doi.org/10.1053/j.gastro.2009.12.064>.
62. Nguyen HT, Duong HQ. The molecular characteristics of colorectal cancer: implications for diagnosis and therapy. *Oncol Lett* 2018;**16**:9–18. <https://doi.org/10.3892/ol.2018.8679>.
63. Ferretti S, Hamon J, de Kanter R, *et al.* Discovery of WRN inhibitor HRO761 with synthetic lethality in MSI cancers. *Nature* 2024;**629**:443–9. <https://doi.org/10.1038/s41586-024-07350-y>.
64. An H, Choi SY, Yi J, *et al.* Mouse single-cell long-read splicing atlas by ouro-seq. 2025. <https://doi.org/10.1101/2025.01.17.633678>.
65. Lim C, An H, Park J. Identification of cell-type-specific, transcriptionally active transposable elements using long-read RNA-sequencing data-based comprehensive annotation. *Genomics Inform* 2025;**23**:17. <https://doi.org/10.1186/s44342-025-00048-1>.