BMC Bioinformatics

SOFTWARE

Open Access

SPICEY: an R package for quantifying tissue specificity from single cell multi-omics data

Georgina Fuentes-Páez[1*], Nacho Molina[2,3], Mireia Ramos-Rodríguez[1] and Lorenzo Pasquali[1*]

*Correspondence:
Georgina Fuentes-Páez
georgina.fuentes@upf.edu
Lorenzo Pasquali
lorenzo.pasquali@upf.edu
[1]Endocrine Regulatory Genomics, Department of Medicine and Life Sciences, Universitat Pompeu Fabra (UPF), 08003 Barcelona, Spain
[2]Institut de Génétique Et de Biologie Moléculaire Et Cellulaire (IGBMC), University of Strasbourg, 67404 Illkirch, France
[3]School of Science and Technology, IE University, 28046 Madrid, Spain

Abstract

Background Single-cell technologies allow detailed mapping of cell type-specific regulatory and transcriptomic landscapes, yet a systematic way to quantify cell type specificity of chromatin accessibility and gene expression remains limited. SPICEY (SPecificity Index for Coding and Epigenetic activitY) is an R package that measures cell-type specificity from single cell multi-omic data.

Results We developed SPICEY, an R package that combines differential and entropy-based metrics to measure cell-type specificity from annotated single-cell accessibility and gene expression data. It computes two indices: RETSI (Regulatory Element cell Type Specificity Index) for chromatin accessibility and GETSI (Gene Expression cell Type Specificity Index) for gene expression. When links between distal chromatin regions and target genes are provided, SPICEY integrates regulatory and transcriptional specificity scores.

Conclusions Applied to human pancreatic islet data, SPICEY identified cell-type-specific gene-regulatory pairs and regulatory features enriched in endocrine cells -including beta cells- providing a framework to dissect cell-type-specific regulatory mechanisms in health and disease.

Keywords Cell-type specificity, Single-cell gene regulation, Epigenomics, Transcriptomics, R package

Background

Advances in single cell technologies, including single cell RNA sequencing (scRNA-seq) and single cell ATAC sequencing (scATAC-seq), have enabled high-resolution profiling of gene expression and chromatin accessibility across diverse cell types within different tissues [1]. These modalities provide complementary views of cellular identity and regulatory state, offering opportunities to dissect cell-type-specific programs in both health and disease. However, while examples of computational tools designed to capture cell-type-specific gene expression are available [2–4], computational frameworks for quantifying the specificity of chromatin accessibility across multiple cell types in a standardized, interpretable, and integrative manner are lacking.

To address this need, we developed SPICEY (SPecificity Index for Coding and Epigenetic activitY), an R package that uses scRNA-seq and scATAC-seq to quantify cell-type specificity of gene regulation (Fig. 1).

To demonstrate the applicability of SPICEY, we consider the pancreatic islet tissue, which exhibits substantial heterogeneity and comprises multiple specialized endocrine cell types, including insulin-producing beta cells, glucagon-secreting alpha cells and somatostatin-producing delta cells, among others. These cell populations, in concert with other tissues, ensure constant blood glucose levels and their dysfunction may lead to disorders such as type 1 and type 2 diabetes [5, 6]. Given that many diabetes-associated genetic variants map to noncoding regions of the genome [7], elucidating the regulatory landscapes that define and maintain cell type-specific functions is a key step towards understanding disease mechanisms and interpreting the cell type-specific impact of noncoding variation [8]. We applied SPICEY to human pancreatic islet data from the Human Pancreas Analysis Program (HPAP) [9, 10], generating integrated measures of gene expression and chromatin accessibility specificity across major islet cell types. By systematically identifying distal chromatin accessible sites and genes with high cell-restricted activity, SPICEY facilitates the interpretation of functional variation in a cell type-resolved context and provides a generalizable framework for studying regulatory architecture in complex tissues.

Methods

Overview

SPICEY is implemented as an R package that quantifies cell-type specificity from transcriptomic and epigenomic data using as input the statistical results derived from differential expression and accessibility analysis of single-cell RNA-seq and/or ATAC-seq data. These differential tests, commonly referred to as marker detection, compare each predefined cell type or cluster against all other cells. The returned statistics, namely

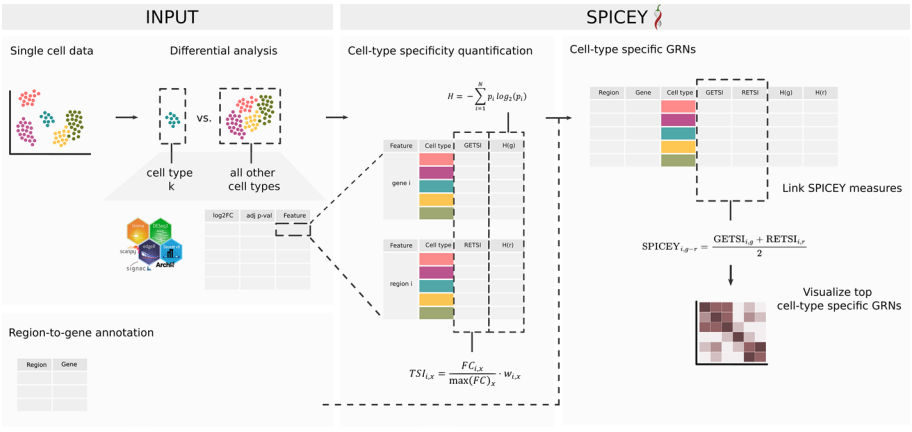

Fig. 1 SPICEY-based inference of regulatory specificity. Schematic overview of the SPICEY framework for quantifying and integrating cell-type-specific gene regulatory activity. SPICEY takes as input per-cell-type marker detection results for all regions and genes. For each gene and regulatory region, SPICEY integrates fold-change estimates with statistical significance to compute tissue specificity indices, namely the Gene Expression cell Type Specificity Index (GETSI) and the Regulatory Element cell Type Specificity Index (RETSI). In parallel, entropy-based measures are calculated to quantify the degree of expression or accessibility dispersion across cell types. If a link between regulatory regions and putative target genes is provided, RETSI and GETSI scores are subsequently combined to derive a unified SPICEY score for each gene, capturing its tissue-specificity across regulatory and transcriptional layers

fold-change and associated *p*-values, are then used by SPICEY to compute per-feature cell type-specificity scores (Fig. 1).

Differential analyses in single-cell datasets are sensitive to the underlying statistical framework. In particular, approaches that treat individual cells as independent replicates may violate independence assumptions, resulting in pseudoreplication, inflated significance estimates, and spurious detection of differentially expressed or accessible features [11]. To accommodate user preferences and evolving best practices, SPICEY is intentionally method-agnostic and can accept input from any statistical framework as long as they return a fold-change and a *p*-value, including those derived from bulk transcriptomic or epigenomic data across different cell types. Thus, SPICEY computes quantitative indices that capture the degree to which each feature is restricted or not to specific cell types.

Algorithm

1. Cell type-specificity indices

SPICEY computes a tissue specificity index (TSI) in a per-feature and per-cell-type manner using the following formula:

$$\mathrm{TSI}_{i,x} = \frac{\mathrm{FC}_{i,x}}{\max(\mathrm{FC})_x} \cdot W_{i,x}$$

where:

i is a specific cell type
x is a specific feature
$\mathrm{FC}_{i,x}$ is the fold-change for feature *x* in cell type *i*
$\mathrm{w}_{i,x}$ is the weight for feature *x* in cell type *i* defined as the negative $\log_{10}$ of the adjusted *p*-value, rescaled to the interval [0,1] to allow comparability across features. Specifically:

$$w_{i,x} = \mathrm{rescale}(-\log_{10}(p\mathrm{adj}_{i,x}), [0, 1])$$

Additionally, adjusted *p*-values equal to zero are replaced by the smallest non-zero adjusted *p*-value present in that specific cell-type. Rescaling ensures that the most statistically significant features receive weights close to 1, while less significant features are assigned lower weights, providing a continuous measure of confidence for specificity scoring.

This formulation integrates both the magnitude of change and the statistical confidence to quantify how selectively a feature is accessible or a gene is expressed in a given cell type, yielding a specificity score ranging from 0 to 1. Higher values indicate greater specificity in the given cell type. When tissue specificity is quantified using chromatin accessibility, the TSI is referred to as RETSI (Regulatory Element cell Type Specificity Index), whereas when gene expression is used, the TSI is referred to as GETSI (Gene Expression cell Type Specificity Index).

2. Entropy-based specificity indices

SPICEY also quantifies feature distribution skewness using Shannon entropy. For each region (RETSI) or gene (GETSI), the tissue specificity indexes are normalized into a probability distribution (p_i) across all N cell types:

$$\{p_i\}_{i=1}^{N}$$

And entropy is calculated as follows:

$$H = -\sum_{i=1}^{N} p_i \log_2(p_i)$$

To ensure comparability across datasets, entropy is then normalized as:

$$H_{\text{norm}} = 1 - e^{H}$$

This yields scores ranging from 0 to 1, where values close to 1 indicate widespread accessibility or expression across cell types, and values near 0 denote specificity to one or few cell types.

3. Integrating RETSI and GETSI

To assess the relationship between cell-type specificity of regulatory elements and gene expression, RETSI and GETSI scores can be correlated using any gene-region annotation data. For the following analyses we decided to use co-accessibility links derived from single cell accessibility data using tools such as Cicero [12]. If that information is not available, regions can be annotated to the nearest gene. The combined score linking the open chromatin site with the associated gene, derived from integrating RETSI and GETSI, is referred to as SPICEY score.

Processing of input data

We applied SPICEY to different human single-cell transcriptomic and chromatin accessibility datasets to test its performance across biologically distinct contexts. We analyzed human pancreatic islets as an example of a disease-relevant tissue with a well-defined cellular heterogeneity, and peripheral blood mononuclear cells (PBMCs) as an independent and widely used reference system. These datasets allowed us to evaluate the robustness, general applicability, and biological interpretability of SPICEY-derived cell-type-specific metrics.

Human islets

To evaluate the performance of SPICEY in a biologically relevant context, we applied the R package to single cell data sets derived from non-diabetic human pancreatic islets generated by the Human Pancreas Analysis Program (HPAP) [9, 10]. Specifically we analyzed single cell data (scRNA-seq and scATAC-seq) from five donors (Supplementary Table 1), which showed comparable total cell numbers and similar distributions across the major islet cell types. We retained only those cell types represented by at least 100 cells to ensure robust downstream analysis (Supplementary Fig. 1).

ATAC and RNA data were processed using standard Signac v.1.15.0 [13] and Seurat v.5.3.0 [14] pipelines, respectively. Samples from different donors were integrated using harmony v.1.2.3 [15]. Cell types were annotated using either accessibility or expression of canonical islet markers [16]. Only cell types with ≥ 100 cells/nuclei per donor were retained. To create an integrated UMAP of RNA and ATAC profiles, cross-modality anchors were identified using *FindTransferAnchors*, and cell-type labels were transferred from scRNA-seq to scATAC-seq using *TransferData* function from Seurat v5, providing consistent annotations across modalities.

SPICEY computes tissue-specific metrics using the statistical estimates -$\log_2$FC and multiple testing adjusted *p*-values- derived from marker detection of genes and accessible regions. We obtained these metrics using Seurat's *FindMarkers* function with the Wilcoxon rank sum test, comparing each cell type against all others. To maximize sensitivity, we only tested features detected in at least 1% of cells. All returned features for each annotated cell type are used as input for SPICEY, without imposing any *p*-value or log-fold change cutoff.

Co-accessible region pairs were identified from the scATAC-seq data using Cicero [12]. Only high-confidence co-accessibility links (score > 0.5) connecting distal open chromatin sites to protein-coding promoters were retained. These distal-promoter co-accessibility links were used to associate open chromatin sites with putative target genes.

Peripheral blood mononuclear cells (PBMCs)

To further assess robustness and general applicability of SPICEY across distinct biological contexts, we applied the same framework to a publicly available single-cell multiome dataset (scRNA + snATAC-seq) of peripheral blood mononuclear cells (PBMCs). This human PBMC multiome dataset was originally generated by 10 × Genomics (*10 k Human PBMCs, Multiome v1.0, Chromium X*) (https://www.10xgenomics.com/resources/datasets/) and was accessed in this study via the *pbmcMultiome.SeuratData* repository using the SeuratData package (satijalab, 2019/2025), which provides pre-processed and curated matched transcriptomic and chromatin accessibility profiles from the same cells. After filtering low-quality cells, the dataset comprised 10,970 paired multiome cells with reference cell-type annotations.

Both modalities were processed independently using standard workflows implemented in Signac v.1.15.0 [13] and Seurat v.5.3.0 [14]. To annotate cell types in the dataset we transferred cell labels from an existing PBMC reference dataset [17] using *FindTransferAnchors* and *TransferData* function from Seurat v5, providing consistent annotations across modalities following the same procedures as we did for the Human Pancreas Analysis Program (HPAP) datasets.

Differential testing analyses were conducted using the same statistical framework, parameters, and inclusion criteria as described for the HPAP datasets above. Co-accessible region pairs were identified following the same procedure used for the HPAP dataset.

The resulting summary statistics -$\log_2$FC and multiple testing adjusted *p*-values- were used as input for SPICEY to compute gene and regulatory-level cell-type specificity metrics.

Benchmarking SPICEY against other methods

To assess the performance of SPICEY in quantifying cell-type specificity, we benchmarked it against three established approaches: dreamlet [2], deCS [3] and scanpy [4]. All analyses were conducted on annotated single-cell RNA-seq data from the Human Pancreas Analysis Program (HPAP). For dreamlet, single-cell counts were aggregated into pseudo-bulk profiles for each cell type using the *aggregateToPseudoBulk* function from the dreamlet R package. Cell-type-specificity scores were subsequently computed with *cellTypeSpecificity* and normalized to a 0–1 scale for comparability. For deCS, the Seurat object was processed following standard workflow: normalization was performed with *NormalizeData*, highly variable genes were identified with *FindVariableFeatures*, and data was scaled using *ScaleData* functions from Seurat. Dimensionality reduction and clustering were performed using *RunPCA*, *FindNeighbors*, *FindClusters* and *RunUMAP* functions from the Seurat R package. Cluster-average expression values were obtained using the *AverageExpression* function from Seurat. Scores were calculated for all marker genes identified with *FindAllMarker*s function and linearly rescaled to generate per-gene, per-cell-type specificity scores between 0 and 1. For Scanpy, normalized gene-ranking results generated with *rank_genes_groups* were retrieved. Raw scores were rescaled to a 0–1 interval to allow direct comparison across methods. SPICEY specificity scores (GETSI) were computed as described above, integrating fold-change and multiple-testing-adjusted *p*-values from differential expression analysis to generate a per-gene, per-cell-type index, ranging values from 0 to 1. All features from all datasets were annotated according to literature-derived tissue-specific and ubiquitous genes [18, 19]. Discriminative performance of each method was evaluated using the delta area under the curve (ΔAUC), defined as the difference between cumulative distributions of specificity scores for ubiquitous versus tissue-specific genes. For each method, empirical cumulative distribution functions were computed separately for tissue-specific and ubiquitous genes using the *ecdf* function from the stats R package. Partial AUCs were calculated with *cumsum* and *diff* functions from base R, and ΔAUC was defined as the difference between cumulative sums for ubiquitous and tissue-specific genes. Positive ΔAUC values indicate that tissue-specific genes are assigned higher specificity scores than ubiquitous genes, reflecting superior discriminative capacity.

Results

To evaluate the performance of SPICEY in a biologically relevant context, we applied the R package to five paired scATAC-seq and scRNA-seq datasets from healthy human pancreatic islets generated by the Human Pancreas Analysis Program (HPAP) [9, 10]. SPICEY was applied to marker detection outputs of annotated islet populations, including beta, alpha, delta, gamma, epsilon, ductal, acinar, endothelial, stellate, and immune cells (Fig. 2A).

To annotate regulatory elements to genes, we used co-accessible links, allowing assignment of distal chromatin peaks to their putative target genes based on correlated accessible patterns (see Methods). The resulting distributions of RETSI and GETSI values are depicted in Fig. 2B. Both measures exhibited a broad range of values across cell types, reflecting heterogeneous regulatory and transcriptional landscapes within the pancreatic islets. Of note, the number of cells captured for each cell type may influence these distributions, as cell types with fewer cells (e.g., gamma, epsilon) tend to show narrower

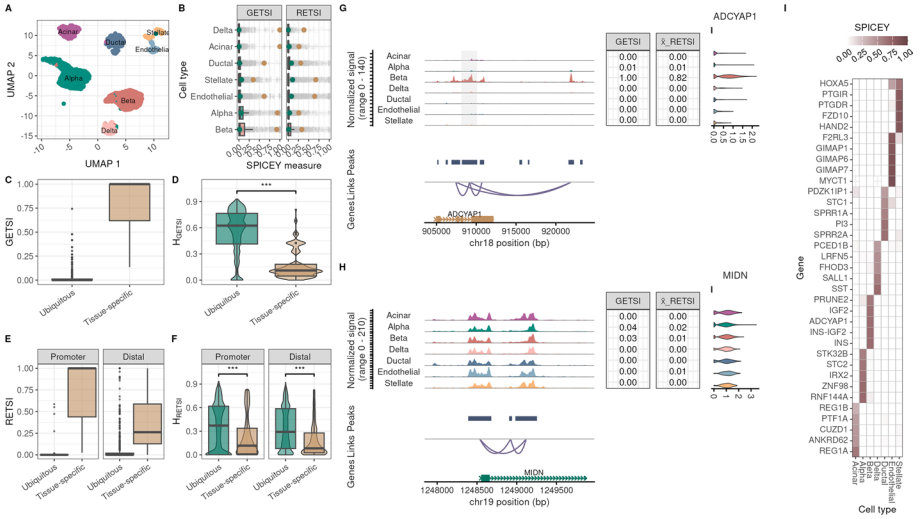

Fig. 2 SPICEY scores reveal cell-type-specific regulatory activity in human pancreatic islets. **A** Joint UMAP of RNA and ATAC profiles from five human islet samples. Each point represents a cell/nucleus, colored by annotated cell type. **B** Distributions of GETSI and RETSI values across annotated human pancreatic islet cell types. Each grey point represents a gene (GETSI) or open chromatin site (RETSI), indicating the degree of cell-type specificity in expression or chromatin accessibility, respectively. Colored dots represent the mean SPICEY scores for literature-derived tissue-specific (brown) and ubiquitous (green) genes or elements within each cell type and metric. **C**-**F** Boxplots and violin plots illustrating the distribution of GETSI **C**-**D** and RETSI **E**–**F** SPICEY scores for ubiquitous and tissue-specific genes. For ubiquitous features, scores are averaged across all cell types. For tissue-specific features, the annotated scores correspond to each marker-defined cell type. **C** and **E** depict GETSI and RETSI scores while **D** and **F** represent entropy distributions for gene expression and chromatin accessibility respectively. RETSI scores and entropy measures are stratified by promoter and distal annotations. Tissue-specific features exhibit significantly lower entropy and higher RETSI and GETSI scores indicating greater cell-type restriction (**$p < 0.001$, Wilcoxon test). Promoters represent open chromatin sites overlapping Transcription Start Sites (TSS) of the selected genes. Distal sites correspond to regions linked by co-accessibility to the selected gene promoters. **G**-**H** ADCYAP1 and *MIDN* loci as representative examples of cell-type specific and ubiquitously expressed genes across cell types. Tracks display normalized chromatin accessibility signals across cell types, together with peak locations and co-accessibility links connecting distal elements to gene promoters. For *ADCYAP1* (tissue-specific gene), only high-confidence co-accessibility links were shown (link score > 0.9), to highlight the most strongly connected regulatory interactions, whereas for *MIDN* (ubiquitous gene) all detected links are considered. Gene-level GETSI values, mean RETSI values ($\bar{x}$ RETSI) across regulatory elements within each locus, and gene expression violin plots across cell types are shown alongside each locus. Tissue-specific loci display elevated SPICEY scores (GETSI and $\bar{x}$ RETSI) and higher expression in the corresponding cell type compared with all others, whereas ubiquitously expressed loci exhibit uniformly lower and evenly distributed SPICEY scores and expression across cell types. **I** Heatmap of top genes by SPICEY measures across cell types. The heatmap displays the mean combined SPICEY scores for the top 5 genes per cell type, derived from RETSI and GETSI specificity measures. RETSI and GETSI were averaged to obtain a combined specificity score. The color scale displays the magnitude of the combined SPICEY score, with darker dyes indicating higher cell type specificity

or less variable specificity ranges, whereas more abundant populations (e.g., beta, alpha) display broader distributions.

We further examined SPICEY scores at promoter and distal chromatin accessible sites associated with a list of literature-derived ubiquitous [19] and tissue-specific genes [16, 18, 20–26] (Fig. 2C and E) (Supplementary Tables 2 and 3). For each ubiquitous gene or accessible region, GETSI and RETSI values were averaged across all cell types. For tissue-specific features, scores were evaluated exclusively within the annotated marker cell type (e.g., *INS* specificity was assessed in beta cells but not in other cell types). We found that promoters and accessible sites linked to known cell-type-specific genes exhibited significantly higher indices compared to their ubiquitous counterparts, manifesting the assumption that cell-type-specific gene expression is coupled with cell-type-specific

cis-regulatory activity. These results demonstrate that SPICEY reliably captures expected patterns of regulatory and transcriptional specificity within the human islet context.

To evaluate the global distribution of chromatin accessibility and gene expression across cell types, we computed entropy scores for all features (Fig. 2D and F). Ubiquitously expressed genes and open chromatin sites displayed high entropy, showing relatively uniform activity across different cell types. In contrast, tissue-specific features were characterized by significantly lower entropy values (*Wilcoxon test,* $p < 0.001$), indicating a bias toward cell-type specific activity. These findings validate the entropy-based component of SPICEY as an additional robust measure of cell-type specificity.

To assess the biological relevance and discriminative power of SPICEY, we computed a combined SPICEY score by averaging RETSI and GETSI values for each pair of linked gene and chromatin accessible sites, unifying both chromatin accessibility and gene expression tissue specificity (see Methods). Next, we generated a heatmap of the top 5 genes per cell type ranked by these combined SPICEY scores, showcasing coordinated regulatory and transcriptional specificity (Fig. 2I). Of note, using this approach, SPICEY effectively captures well known cell type specific genes across both endocrine and non-endocrine populations among the top scored genes. For example, *INS* and *SST*, characteristic hormones secreted specifically by beta and delta cells [27], respectively, were among the top hits for these cell types. SPICEY also highlighted other known markers such as *ADCYAP1* in beta cells [28] and *IRX2* transcription factor in alpha cells [29]. In non-endocrine populations, SPICEY identified canonical acinar markers such as *REG1A/B* and *PI3* ductal cell marker [29]. We highlight *ADCYAP1* as a representative example of high SPICEY score for which we provide the visualization of chromatin accessibility profile across cell types. *ADCYAP1* displays strong and beta cell exclusive accessibility and expression (Fig. 2G), resulting in a high SPICEY score specifically in beta cells. In contrast, the ubiquitously expressed gene *MIDN* exhibits broadly shared accessibility and transcriptional activity across cell types (Fig. 2H), consistent with uniformly low SPICEY scores. Collectively, these data demonstrate that SPICEY effectively captures both regulatory and transcriptional specificity.

To further evaluate the robustness of SPICEY, we applied the R package to an additional multiome dataset to ensure that the observed results were not dataset-dependent. Using an independent dataset generated from processed peripheral blood mononuclear cells (PBMCs), SPICEY successfully distinguished ubiquitous from cell-type-specific regulatory programs (Supplementary Fig. 2).

We additionally assessed SPICEY's ability to discriminate between ubiquitous and cell-type-specific transcription, comparing its performance with that of other computational tools. When applied to the Human Pancreas Analysis Program (HPAP) dataset, SPICEY demonstrated clearly superior discriminative capacity (Supplementary Fig. 3). Importantly, this comparison was limited to gene expression cell-type specificity, as, to our knowledge, no tools currently exist for computing cell-type-specific scores for chromatin accessibility features from multiome and/or scATAC-seq data.

Collectively, these results establish SPICEY as a quantitative and biologically interpretable framework for measuring cell-type specificity at both epigenetic and transcriptomic levels. By leveraging integrated single cell information, SPICEY enables the identification of open chromatin sites and genes that underlie functional heterogeneity within the pancreatic islet, but also may uncover novel candidates for pancreatic identity

and function regulation. This highlights the value of integrating chromatin accessibility and gene expression to refine tissue-specific regulatory networks and to facilitate the identification and prioritization of candidate genes for targeted functional studies.

Discussion

SPICEY is a robust and versatile R package designed to quantify cell type-specificity from chromatin accessibility and gene expression single cell data. By integrating open chromatin and transcriptomic profiles, SPICEY provides a unified framework to characterize per feature cell-type specificity through two complementary metrics: RETSI and GETSI. These indices combine the results of differential marker testing with entropy-based quantification to highlight features with selective accessibility or expression across diverse cellular populations. When an accessible site to gene annotation is provided, RETSI and GETSI are combined into a SPICEY score, which integrates per-gene transcriptional and regulatory specificity and facilitates the prioritization of regulatory elements and candidate genes with cell-type-specific functions.

Importantly, SPICEY is agnostic to the differential method used to identify marker features or genes, and thus, can be applied in conjunction with the more widely used methods such as Signac [13], Seurat [14], scran [30], edgeR [31], and DESeq2 [32]. This ensures compatibility with existing pipelines, facilitates broader adoption across the single cell research community and allows SPICEY to readily integrate with novel methods as statistical approaches continue to evolve.

Application of SPICEY to scATAC-seq and scRNA-seq data from human pancreatic islets demonstrates its ability to recover biologically meaningful patterns of tissue specificity, accurately capturing distinct molecular signatures of both endocrine and non-endocrine cell types. This performance is recapitulated in an independent dataset of PBMCs, indicating that SPICEY generalizes across tissues.

Entropy-based analyses further supports the robustness of these scores, with tissue-specific promoters and distal chromatin accessible sites showing significantly lower entropy than ubiquitous elements, reflecting tighter regulatory control and restriction to defined cellular contexts. This dual specificity quantification enables deeper exploration of gene regulation, particularly in tissues with complex cellular heterogeneity like the pancreas.

Importantly, SPICEY offers an integrative view of transcription and chromatin activity that help reveal cell-type-restricted genes with coordinated regulatory and transcriptional activity.

Of note, SPICEY relies on predefined cell-type annotations and does not correct for potential misannotations; therefore, inaccuracies in the input labels may propagate to the resulting GETSI and RETSI scores. In addition, these scores capture relative cell-type specificity, meaning that cell types characterized by broadly distributed transcriptional or regulatory programs may exhibit lower specificity values even when annotations are correct. Finally, while the RETSI score quantifies cell-type-specific chromatin accessibility patterns, it does not directly imply functional enhancer activity, and additional experimental validation is required to confirm the regulatory roles of the identified regions.

Conclusions

In summary, SPICEY provides a flexible, user-friendly platform for quantifying and visualizing regulatory and transcriptional specificity in heterogeneous tissues. By integrating single-cell chromatin accessibility and gene expression data, it enables systematic dissection of cell-type-specific regulatory programs and supports the interpretation of cellular heterogeneity in diverse tissues and experimental settings, including disease models, differentiation trajectories or perturbed systems.

Abbreviations

SPICEY	SPecificity Index for Coding and Epigenetic activitY
RETSI	Regulatory element cell type specificity index
GETSI	Gene expression cell type specificity index
HPAP	Human pancreas analysis program
scRNA-seq	Single cell RNA sequencing
scATAC-seq	Single cell ATAC sequencing
FC	Fold-change
padj	*P*-value adjusted for multiple testing
DEGs	Differentially expressed genes
DARs	Differential accessible regions
TSS	Transcription start site
UMAP	Uniform manifold approximation
WNN	Weighted nearest neighbor

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s12859-026-06418-y.

Supplementary material

Supplementary Figure 1

Supplementary Figure 2

Supplementary Figure 3

Supplementary Table 1

Supplementary Table 2

Supplementary Table 3

Supplementary Table 4

Acknowledgements

This manuscript used data acquired from the database (https://hpap.pmacs.upenn.edu/) of the Human Pancreas Analysis Program (HPAP; RRID:SCR_016202; PMID: 31127054; PMID: 36206763). HPAP is part of a Human Islet Research Network (RRID:SCR_014393) consortium (UC4-DK112217, U01-DK123594, UC4-DK112232, and U01-DK123716). We thank the MELIS-UPF Scientific Computing Core Facility and its staff for their contribution to this publication.

Author contributions

GFP designed and programmed SPICEY. GFP and MRR tested and validated the software. LP, MRR, and NM provided supervision and critical guidance throughout the project. LP secured funding acquisition. GFP, MRR, and LP wrote and edited the manuscript, with input from all authors. All authors read and approved the final version of the manuscript.

Funding

This work was supported by "la Caixa" Foundation, LCF-PR-HR24-00150 and by the Spanish Ministry of Science and Innovation projects PID2020-117099RB-I00, PID2023-151556OB-I00 and CNS2024-154742 under the FPI predoctoral fellowship (PRE2021-098216), funded by MICIU/AEI/https://doi.org/10.13039/501100011033 and, as appropriate, by "ESF Investing in your future", by "ESF+" or by "European Union NextGenerationEU/PRTR". M.R.-R. is supported by the EFSD/Lilly Young Investigator Award. This work was made possible through the "Unidad de Excelencia María de Maeztu" CEX2024-001431-M, funded by MICIU/AEI/https://doi.org/10.13039/501100011033.

Data availability

The single-cell RNA-seq and ATAC-seq datasets analyzed in this study were obtained from the Human Pancreas Analysis Program (HPAP) [9, 10] and are publicly accessible through the HPAP data portal (https://hpap.pmacs.upenn.edu). The PBMC single-cell multiome (scRNA and snATAC-seq) dataset was obtained from the ***pbmcMultiome.SeuratData*** repository via the SeuratData package and originates from a publicly released 10×Genomics dataset (https://www.10xgenomics.com/datasets) was obtained from the ***SeuratData*** repository (***pbmcMultiome***). This dataset provides pre-processed and annotated matched scRNA-seq and scATAC-seq profiles from the same cells. All processed data, intermediate files and code required to reproduce the analyses and figures presented in this study is available at Zenodo (https://doi.org/10.5281/zenodo.18482899).

Declarations

Ethics approval and consent to participate
This study used publicly available, de-identified human pancreatic islet single-cell datasets from the Human Pancreas Analysis Program (HPAP). All data were collected with informed consent under protocols approved by the respective institutional review boards of the contributing institutions. No additional ethical approval was required for the analyses conducted in this study.

Consent for publication
Not applicable.

Competing interests
The authors declare no competing interests.

Received: 13 November 2025 / Accepted: 3 March 2026
Published online: 18 March 2026

References

1. Lim J, Park C, Kim M, Kim H, Kim J, Lee D-S. Advances in single-cell omics and multiomics for high-resolution molecular profiling. Exp Mol Med. 2024;56(3):515–26. https://doi.org/10.1038/s12276-024-01186-2.
2. Hoffman GE, Lee D, Bendl J, Prashant NM, Hong A, Casey C, et al. Efficient differential expression analysis of large-scale single cell transcriptomics data using dreamlet. bioRxiv. 2024. https://doi.org/10.1101/2023.03.17.533005.
3. Pei G, Yan F, Simon LM, Dai Y, Jia P, Zhao Z. deCS: A tool for systematic cell type annotations of single-cell RNA sequencing data among human tissues. Genomics Proteomics Bioinformatics. 2023;21(2):370–84. https://doi.org/10.1016/j.gpb.2022.04.001.
4. Wolf FA, Angerer P, Theis FJ. SCANPY: large-scale single-cell gene expression data analysis. Genome Biol. 2018;19(1):15. https://doi.org/10.1186/s13059-017-1382-0.
5. Eizirik DL, Pasquali L, Cnop M. Pancreatic β-cells in type 1 and type 2 diabetes mellitus: different pathways to failure. Nat Rev Endocrinol. 2020;16(7):Article 7. https://doi.org/10.1038/s41574-020-0355-7.
6. Röder PV, Wu B, Liu Y, Han W. Pancreatic regulation of glucose homeostasis. Exp Mol Med. 2016;48(3):e219–e219. https://doi.org/10.1038/emm.2016.6.
7. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. Proc Natl Acad Sci U S A. 2009;106(23):9362–7. https://doi.org/10.1073/pnas.0903103106.
8. Ramos-Rodríguez M, Pérez-González B, Pasquali L. The β-cell genomic landscape in T1D: implications for disease pathogenesis. Curr DiabRep. 2021;21(1):1. https://doi.org/10.1007/s11892-020-01370-4.
9. Fasolino M, Schwartz GW, Patil AR, Mongia A, Golson ML, Wang YJ, et al. Single-cell multi-omics analysis of human pancreatic islets reveals novel cellular states in type 1 diabetes. Nat Metab. 2022;4(2):284–99. https://doi.org/10.1038/s42255-022-00531-x.
10. Kaestner KH, Powers AC, Naji A, HPAP Consortium, Atkinson MA. NIH initiative to improve understanding of the pancreas, islet, and autoimmunity in type 1 diabetes: the Human Pancreas Analysis Program (HPAP). Diabetes. 2019;68(7):1394–402. https://doi.org/10.2337/db19-0058.
11. Squair JW, Gautier M, Kathe C, Anderson MA, James ND, Hutson TH, et al. Confronting false discoveries in single-cell differential expression. Nat Commun. 2021;12(1):5692. https://doi.org/10.1038/s41467-021-25960-2.
12. Pliner HA, Packer JS, McFaline-Figueroa JL, Cusanovich DA, Daza RM, Aghamirzaie D, et al. Cicero predicts cis-regulatory DNA interactions from single-cell chromatin accessibility data. Mol Cell. 2018;71(5):858-871.e8. https://doi.org/10.1016/j.molcel.2018.06.044.
13. Stuart T, Srivastava A, Madad S, Lareau CA, Satija R. Single-cell chromatin state analysis with Signac. Nat Methods. 2021;18(11):1333–41. https://doi.org/10.1038/s41592-021-01282-5.
14. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. 2024;42(2):293–304. https://doi.org/10.1038/s41587-023-01767-y.
15. Korsunsky I, Millard N, Fan J, Slowikowski K, Zhang F, Wei K, et al. Fast, sensitive and accurate integration of single-cell data with Harmony. Nat Methods. 2019;16(12):1289–96. https://doi.org/10.1038/s41592-019-0619-0.
16. Muraro MJ, Dharmadhikari G, Grün D, Groen N, Dielen T, Jansen E, et al. A single-cell transcriptome atlas of the human pancreas. Cell Syst. 2016;3(4):385-394.e3. https://doi.org/10.1016/j.cels.2016.09.002.
17. Hao Y, Hao S, Andersen-Nissen E, Mauck WM, Zheng S, Butler A, et al. Integrated analysis of multimodal single-cell data. Cell. 2021;184(13):3573-3587.e29. https://doi.org/10.1016/j.cell.2021.04.048.
18. Bonnycastle LL, Gildea DE, Yan T, Narisu N, Swift AJ, Wolfsberg TG, et al. Single-cell transcriptomics from human pancreatic islets: sample preparation matters. Biol Methods Protocols. 2019;4(1):bpz019. https://doi.org/10.1093/biomethods/bpz019.
19. Hsiao L-L, Dangond F, Yoshida T, Hong R, Jensen RV, Misra J, et al. A compendium of gene expression in normal human tissues. Physiol Genomics. 2001;7(2):97–104. https://doi.org/10.1152/physiolgenomics.00040.2001.
20. Chiou J, Geusz RJ, Okino M-L, Han JY, Miller M, Melton R, et al. Interpreting type 1 diabetes risk with genetics and single-cell epigenomics. Nature. 2021;594(7863):398–402. https://doi.org/10.1038/s41586-021-03552-w.
21. Chiou J, Zeng C, Cheng Z, Han JY, Schlichting M, Miller M, et al. Single-cell chromatin accessibility identifies pancreatic islet cell type– and state-specific regulatory programs of diabetes risk. Nat Genet. 2021;53(4):455–66. https://doi.org/10.1038/s41588-021-00823-0.
22. Franzén O, Gan L-M, Björkegren JLM. PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. Database. 2019;2019:baz046. https://doi.org/10.1093/database/baz046.
23. Hashimoto Y, Greene C, Munnich A, Campbell M. The CLDN5 gene at the blood-brain barrier in health and disease. Fluids Barriers CNS. 2023;20(1):22. https://doi.org/10.1186/s12987-023-00424-5.

24. Segerstolpe Å, Palasantza A, Eliasson P, Andersson E-M, Andréasson A-C, Sun X, et al. Single-cell transcriptome profiling of human pancreatic islets in health and type 2 diabetes. Cell Metab. 2016;24(4):593–607. https://doi.org/10.1016/j.cmet.2016.08.020.
25. Stankovic S, Shekari S, Huang QQ, Gardner EJ, Ivarsdottir EV, Owens NDL, et al. Genetic links between ovarian ageing, cancer risk and de novo mutation rates. Nature. 2024;633(8030):608–14. https://doi.org/10.1038/s41586-024-07931-x.
26. Wang Y, Liang Y, Xu H, Zhang X, Mao T, Cui J, et al. Single-cell analysis of pancreatic ductal adenocarcinoma identifies a novel fibroblast subtype associated with poor prognosis but better immunotherapy response. Cell Discov. 2021;7(1):36. https://doi.org/10.1038/s41421-021-00271-4.
27. Hauge-Evans AC, King AJ, Carmignac D, Richardson CC, Robinson ICAF, Low MJ, et al. Somatostatin secreted by islet delta-cells fulfills multiple roles as a paracrine regulator of islet function. Diabetes. 2009;58(2):403–11. https://doi.org/10.2337/db08-0792.
28. Yamamoto K, Hashimoto H, Tomimoto S, Shintani N, Miyazaki J, Tashiro F, et al. Overexpression of PACAP in transgenic mouse pancreatic beta-cells enhances insulin secretion and ameliorates streptozotocin-induced diabetes. Diabetes. 2003;52(5):1155–62. https://doi.org/10.2337/diabetes.52.5.1155.
29. Elgamal RM, Kudtarkar P, Melton RL, Mummey HM, Benaglio P, Okino M-L, et al. An integrated map of cell type-specific gene expression in pancreatic islets. Diabetes. 2023;72(11):1719–28. https://doi.org/10.2337/db23-0130.
30. Lun ATL, McCarthy DJ, Marioni JC. A step-by-step workflow for low-level analysis of single-cell RNA-seq data with Bioconductor. F1000Research. 2016;5:2122. https://doi.org/10.12688/f1000research.9501.2.
31. Chen Y, Chen L, Lun ATL, Baldoni PL, Smyth GK. edgeR v4: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets. Nucleic Acids Res. 2025;53(2):gkaf018. https://doi.org/10.1093/nar/gkaf018.
32. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 2014;15(12):550. https://doi.org/10.1186/s13059-014-0550-8.

Publisher's Note