CellTracksColab—A platform for compiling, analyzing, and exploring tracking data

Guillaume Jacquemet[1,2,3,4] ✉

[1]Turku Bioscience Centre, University of Turku and Åbo Akademi University, FI-20520 Turku, FI
[2]Faculty of Science and Engineering, Cell Biology, Åbo Akademi University, FI-20520 Turku, FI
[3]Turku Bioimaging, University of Turku and Åbo Akademi University, FI- 20520 Turku, FI
[4]InFLAMES Research Flagship Center, Åbo Akademi University, FI- 20520 Turku, FI

In life sciences, tracking objects from movies enables researchers to quantify the behavior of single particles, organelles, bacteria, cells, and even whole animals. While numerous tools now allow automated tracking from video, a significant challenge persists in compiling, analyzing, and exploring the large datasets generated by these approaches. Here, we introduce CellTracksColab, an online platform tailored to simplify the exploration and analysis of tracking data by leveraging the free, cloud-based computational resources of Google Colab. CellTracksColab facilitates the amalgamation and analysis of tracking results across multiple fields of view, conditions, and repeats, ensuring a holistic dataset overview. CellTracksColab also harnesses the power of high-dimensional data reduction and clustering, using Uniform Manifold Approximation and Projection and Hierarchical Density-Based Spatial Clustering of Applications with Noise, enabling researchers to identify distinct behavioral patterns and trends without bias. CellTracksColab is available for the broader scientific community here.

Cell Migration | Data analysis | Tracking |
Correspondence: ***guillaume.jacquemet@abo.fi***

Introduction

With the evolution of imaging techniques, biological phenomena are acquired at greater speed and resolution, facilitating the recording of dynamic events. In life science, tracking has emerged as an indispensable tool for unparalleled insights into dynamic molecular and cellular behaviors. Parallel to this, segmentation methods relying on machine learning and deep learning are now greatly facilitating the implementation of complex tracking pipelines[1–4]; enabling the quantitative analysis of dynamic behaviors. Yet, as the capabilities of tracking tools have expanded, so too have the challenges associated with analyzing the resulting data.

Traditional analytical approaches, implemented by us and many others, typically reduce these rich datasets to population-level analyses where track metrics are averaged across different conditions. Yet, while practical, such analyses overlook the heterogeneity within biological data. Over the past two years, multiple tools, including CellPhe (an R toolbox[5]), Traject3D (a collection of Matlab scripts[6]), and CellPlato (a Python toolbox[7]), have been designed to harness the high-dimensionality of tracking datasets to assist in the unbiased discovery of rare phenotypes. Yet, these tools remain challenging to implement for users with no or little coding expertise.

In this work, we introduce CellTracksColab, a novice-oriented, cloud-based platform designed to streamline the analysis and exploration of tracking data. Drawing inspiration from ColabFold[8] and ZeroCostDL4Mic[9], CellTracksColab operates within the Google Colaboratory framework (abbreviated as Colab) and provides a simple graphical user interface (GUI) that necessitates only a web browser and a Google account for user access. All computational tasks are conducted in the cloud via Google Colab, eliminating the requirement for software installation or the establishment of complex Python or R environments. Through a simplified workflow, researchers can install essential software dependencies with a few mouse clicks, upload their tracking data, and start with their analyses. CellTracksColab extends beyond visualization and population analyses, empowering researchers to delve into the nuanced dynamics and behaviors encapsulated within their tracking experiments.

Results

Availability, code, and performance. The code for CellTracksColab is publicly available under the MIT license, encouraging broad utilization and adaptation. CellTracksColab's GitHub repository serves as a dynamic platform for tracking the evolution of the code across various versions. Users are encouraged to report issues and suggest features directly through the GitHub interface. To further ensure enduring accessibility, a stable version of the code has been archived on Zenodo. While CellTracksColab is optimized for operation on Google Colab, it can be easily adapted to run on other Jupyter Notebook platforms.

Under the constraints of the free version of Google Colab, the performance of CellTracksColab is bound by the resources available. However, optimization of the underlying code has been executed to ensure maximal efficiency in resource utilization. For instance, we quickly analyzed more than 50,000 tracks (> 3 million objects from 117 videos) using CellTracksColab and the free version of Google Colab. CellTracksColab could accommodate one of our larger datasets encompassing over 536,000 tracks (> 56 million objects from 300 videos), but this required the enhanced capabilities of Google Colab Plus, acquired through a subscription.

Data input and structure. CellTracksColab ingests tracking data stored in CSV files, each containing the results from

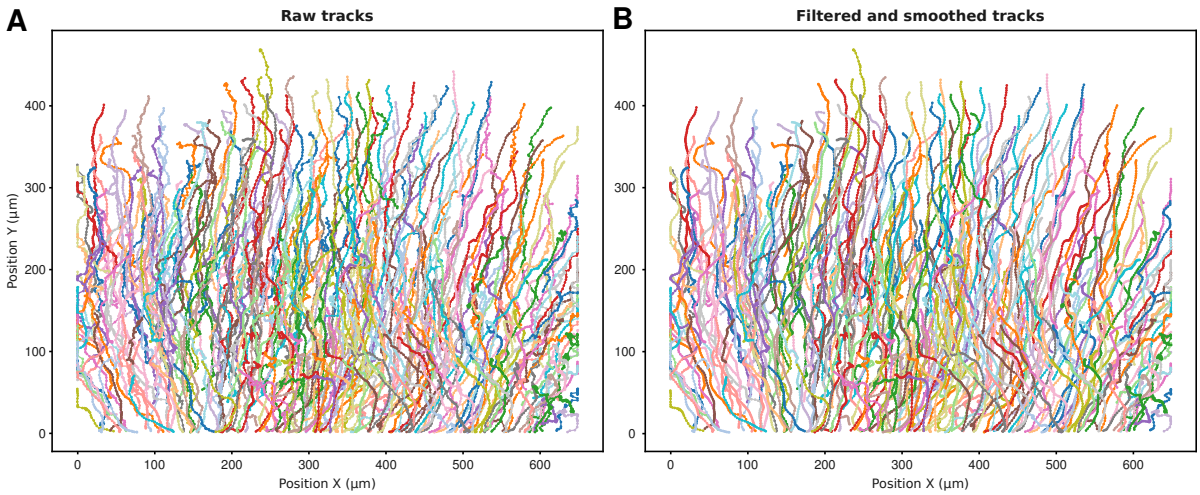

Fig. 1. Track visualization and filtering. Example of track display before (**A**) and after (**B**) smoothing and filtering. The tracks originate from a video of migrating breast cancer cells tracked from their nuclei using TrackMate.

a single tracked video. These files should be methodically organized within a two-tiered folder hierarchy. The primary folder's name delineates the experimental conditions, while the subfolders correspond to the biological repeats. This systematic hierarchy facilitates a comprehensive assessment of data variability and heterogeneity across the entire dataset. One major limitation of CellTracksColab is that it does not support track splitting or merging.

Presently, CellTracksColab comprises two distinct notebooks. One notebook can import data produced from a broad spectrum of tracking tools. The subsequent notebook, titled CellTracksColab - TrackMate, facilitates the direct importation of tracking data generated by the TrackMate batch processor 1. We focused on TrackMate due to its popularity and user-friendly nature. Detailed information regarding the required data input is accessible on our Wiki page.

Track visualization and filtering. CellTracksColab enables users to visualize tracks corresponding to each loaded CSV file (Fig. 1A). Inspired by MotilityLab 10, CellTracksColab offers the possibility to filter tracks based on their length and smooth tracks using moving averages (Fig. 1B). Track smoothing can prove to be particularly beneficial before the computation of directionality metrics, especially when the tracked object exhibits jitteriness (e.g., nuclei) and the user's interest lies in discerning the overall movement of the cell.

Population analysis of track metrics. Upon aggregating data from multiple fields of view, CellTracksColab can compute various track metrics or import them directly from prior analyses conducted in the tracking software. CellTracksColab's modular design (Jupyter Notebook) significantly simplifies the process for advanced users to incorporate additional analysis modules.

Users can generate boxplots illustrating the distribution of various track metrics of interest (Fig. 2). Additionally, several pertinent statistical metrics are calculated. Among these is Cohen's d value, which quantifies the effect size between groups, aiding in discerning the degree of variance between conditions. A notable advantage of Cohen's d value is its insensitivity to effect sizes. CellTracksColab also calculates p-values employing a randomization test to examine the null hypothesis (assessing whether the conditions share the same distribution), both with and without Bonferroni Correction (which adjusts significance thresholds for multiple comparisons) (Fig. 2). Importantly, PDF files of all plots and CSV files encapsulating all plot data are exported, enabling users to replot the results using their preferred software.

Assessing experimental variability. Ensuring data reliability is essential for generating meaningful results from large datasets. CellTracksColab computes similarities among fields of view (FOVs), experimental conditions, and repeats using user-defined track metrics. These results are then visualized using hierarchical clustering, presented as dendrograms (Fig. 3). These dendrograms serve two essential functions. They allow for comparing the similarity within FOVs of the same experimental condition against those from different ones (Fig. 3A) and highlight how conditions and repeats cluster together (Fig. 3B). These analyses also pinpoint potential outliers (among FOVs or biological repeats) that could skew the results. Users can select from various similarity metrics, including cosine, euclidean, and Jaccard, and decide on the hierarchical clustering linkage method.

Exploring tracking data using dimensionality reduction and clustering. Inspired by the analysis pipeline introduced by CellPlato[7], CellTracksColab employs the synergistic capabilities of Uniform Manifold Approximation and Projection (UMAP) and Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN) to explore the inherent heterogeneity within tracking datasets unbiasedly. Utilizing UMAP, the platform condenses various track metrics, transforming the dataset into a more accessible and interpretable visual form (Fig. 4A). This adaptable dimensionality reduction allows for 1-dimension (D), 2D, or 3D representations to suit multiple visualization needs. Following this reduction, HDBSCAN clusters the tracks (based on the UMAP or the raw data), effectively revealing distinct

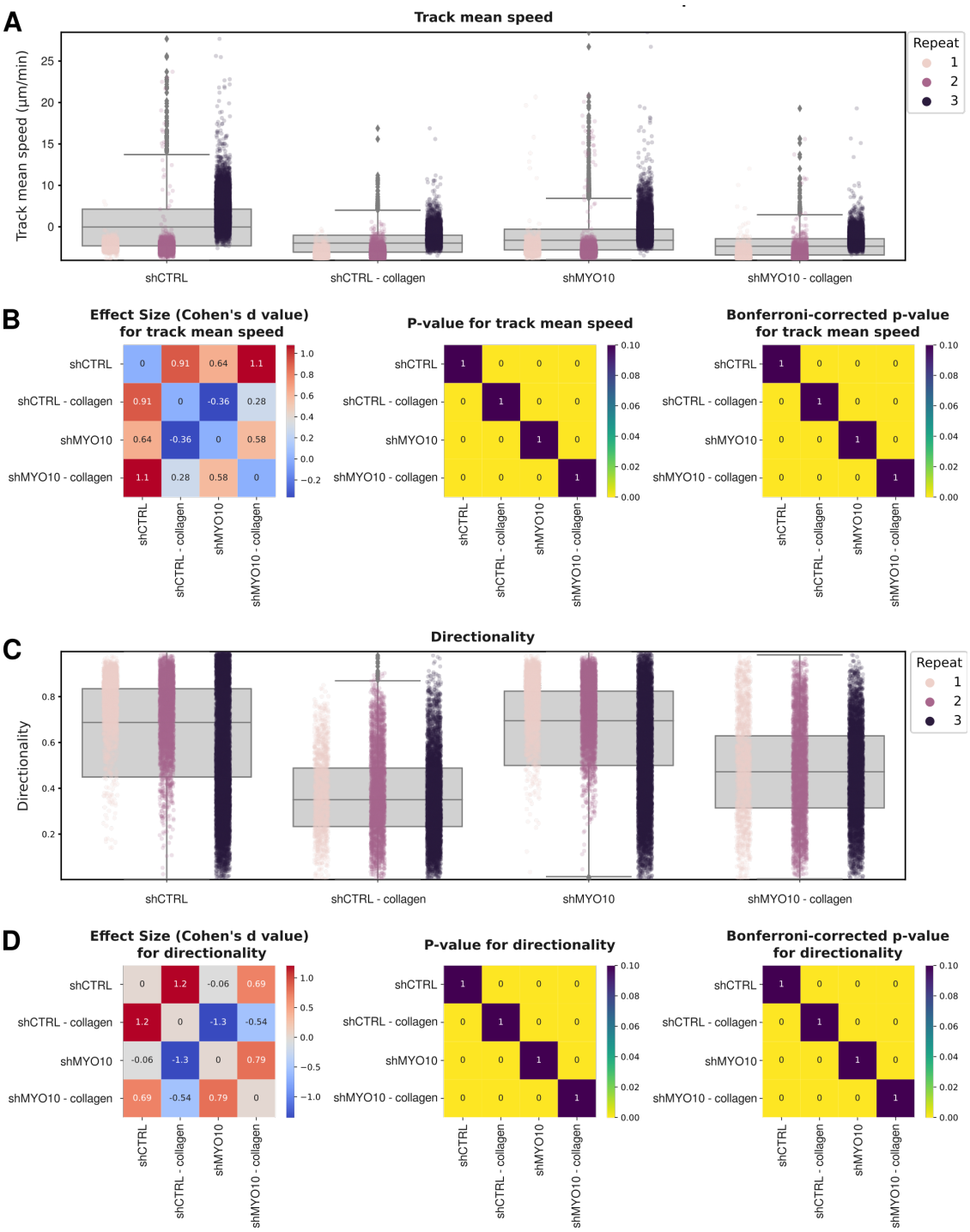

Fig. 2. Population analysis of track metrics and associated statistical analyses. Example of track metric plots and corresponding statistical analyses generated by CellTracksColab. In this dataset, breast cancer cells expressing a CTRL shRNA or a shRNA targeting MYO10 migrate either beneath a collagen gel or within standard media[11]. Displayed track metrics encompass track mean speed (**A**, **B**) and directionality (**C**, **D**). (**A**, **C**) Tukey boxplots of the specified track metric. Whiskers, depicted as vertical lines, extend to data points not exceeding 1.5× the interquartile range from the box. Individual data points from each biological repeat are color-coded. (**B**, **D**) Mirrored heatmaps displaying Cohen's d value alongside the p-values derived from randomization tests (both with and without Bonferroni Correction) across the dataset's conditions.

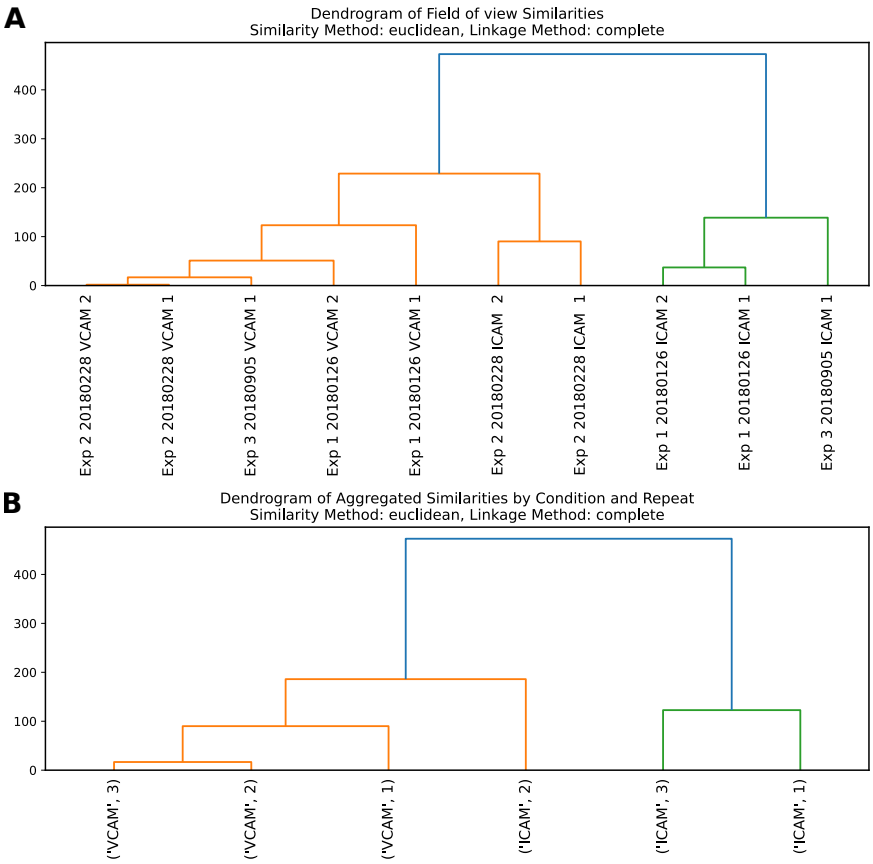

Fig. 3. Example of dendrograms generated by CellTracksColab to assess experimental variability. In this dataset, T-cells were allowed to migrate on a VCAM or ICAM-coated substrate[12]. Cells were then automatically tracked using StarDist and TrackMate. (**A**-**B**) Dendrograms represent the dataset's hierarchical clustering, with similarities determined using the Euclidean similarity metric across all available track metrics. The chosen linkage method was 'complete'. (**A**) Hierarchical clustering across the ten available FOVs. Notably, two FOVs from the ICAM condition show clustering in alignment with the VCAM FOVs. (**B**) Hierarchical clustering of the dataset when segregated by conditions and repeats. Interestingly, the ICAM's second repeat exhibits clustering tendencies similar to the VCAM repeats.

subpopulations or subtle behaviors embedded within the data (Fig. 4B).

To explain the nature of these clusters, CellTracksColab can plot track metrics for each cluster (Fig. 4C). In addition, users can generate a heatmap, normalized via Z scores, depicting the variation of all track metrics across clusters, providing a panoramic view of data heterogeneity (Fig. 4D). Moreover, HDBSCAN can identify exemplar tracks for each cluster, allowing researchers to translate them into tangible examples. These tracks can then be plotted (Fig. 4E) or, when the raw movie data are available, be dynamically visualized in the movies (Fig. 4F). Finally, CellTracksColab facilitates the generation of a 'fingerprint', representing cluster distributions across experimental conditions. This unbiased snapshot can be used for benchmarking and nuanced comparative studies (Fig. 4G).

Conclusion

CellTracksColab offers a platform for comprehensively analyzing tracking data in life sciences. Operating within the Google Colaboratory framework, it provides a cloud-based solution accessible via a standard web browser, eliminating the need for intricate software installations. CellTracksColab's intuitive GUI ensures the tool remains user-friendly, bridging the gap for individuals with limited coding expertise. The platform integrates multiple functionalities, encompassing track visualization, population analysis, statistical assessments, and exploring tracking data through dimensionality reduction techniques. CellTracksColab is optimized for handling large datasets, and its structured data ingestion method via a two-tiered folder hierarchy ensures organized data input. CellTracksColab offers adaptability, making it a valuable tool in life sciences research.

Acknowledgements

CellTracksColab is inspired by several key projects in cell tracking and analysis. We acknowledge the influential contributions of Traject3d, CellPhe, CelltrackR, the MotilityLab website, and Cellplato on Zenodo. The innovative use of UMAP and HDBSCAN for analyzing tracking data, as featured in CellTracksColab, was first introduced in CellPlato.

I thank Hellyeh Hamidi for providing feedback on this manuscript. This study was supported by the Research Council of Finland (338537 to G.J.), the Sigrid Juselius Foundation (to G.J.), the Cancer Society of Finland (Syöpäjärjestöt; to G.J.), and the Solutions for Health strategic funding to Åbo Akademi University (to G.J.). This research was supported by the InFLAMES Flagship Programme of the Academy of Finland (decision numbers: 337530, 337531, 357910, and 357911).

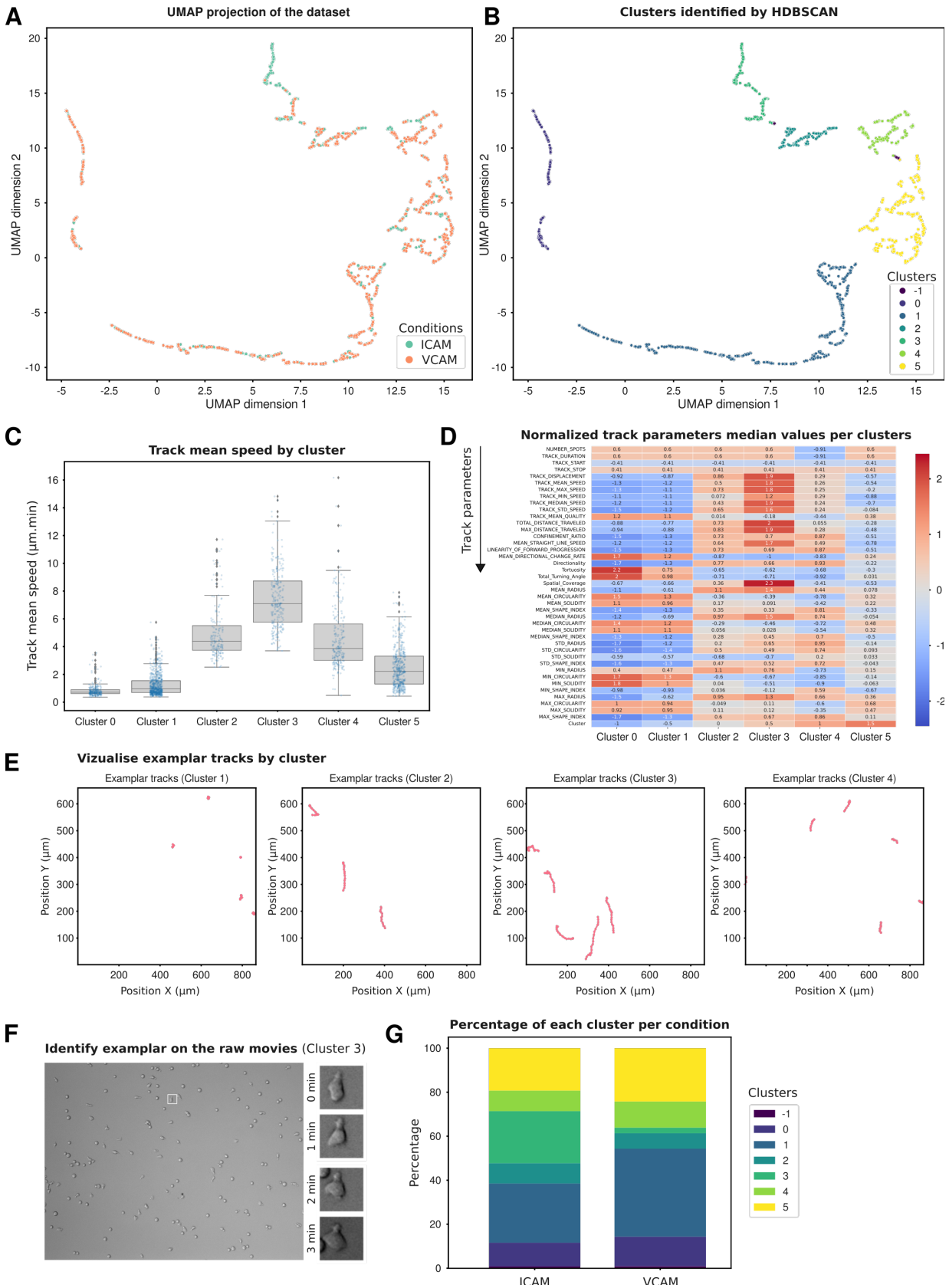

Fig. 4. Exploration of tracking data using UMAP and HDBSCAN. (**A**-**G**) Analysis was performed on the T-cell dataset previously introduced in Fig. 3. (**A**) 2D UMAP projection of the entire dataset, using all available track metrics for dimensionality reduction. Data points are color-coded based on VCAM and ICAM conditions. (**B**) Resultant clusters from the HDBSCAN analysis on the 2D UMAP projection. Euclidean distance served as the metric for clustering. Each identified cluster is uniquely color-coded, with tracks not assigned to any cluster denoted as -1. (**C**) The 'mean speed' metric for each track is plotted per cluster. (**D**) Heatmap representation, normalized using Z scores, displaying variations in all track metrics among the clusters. This provides a comprehensive view of data heterogeneity across clusters. (**E**) Exemplar tracks representing clusters 1, 2, 3, and 4 are displayed. (**F**) Overlay of an exemplar track (from cluster 3) onto the corresponding raw video frame, pinpointing its position and trajectory. (**G**) Fingerprint plot showcasing the distribution percentage of each cluster across different conditions.

Bibliography

1. Ershov, D. et al. TrackMate 7: integrating state-of-the-art segmentation algorithms into tracking pipelines. Nat. Methods 19, 829–832 (2022).

2. Aragaki, H., Ogoh, K., Kondo, Y. and Aoki, K. LIM Tracker: a software package for cell tracking and analysis with advanced interactivity. Sci. Rep. 12, 2702 (2022).

3. Ulicna, K., Vallardi, G., Charras, G. and Lowe, A. R. Automated Deep Lineage Tree Analysis Using a Bayesian Single Cell Tracking Approach. Front. Comput. Sci. 3, (2021).

4. Sugawara, K., Çevrim, Ç. and Averof, M. Tracking cell lineages in 3D by incremental deep learning. eLife 11, e69380 (2022).

5. Wiggins, L. et al. The CellPhe toolkit for cell phenotyping using time-lapse imaging and pattern recognition. Nat. Commun. 14, 1854 (2023).

6. Freckmann, E. C. et al. Traject3d allows label-free identification of distinct co-occurring phenotypes within 3D culture by live imaging. Nat. Commun. 13, 5317 (2022). 7. Michael-shannon/cellPLATO: cellPLATO. doi:10.5281/zenodo.8096717.

8. Mirdita, M. et al. ColabFold: making protein folding accessible to all. Nat. Methods 19, 679–682 (2022).

9. von Chamier, L. et al. Democratising deep learning for microscopy with ZeroCostDL4Mic. Nat. Commun. 12, 2276 (2021).

10. Wortel, I. M. N., Dannenberg, K., Berry, J. C., Miller, M. J. and Textor, J. CelltrackR: an R package for fast and flexible analysis of immune cell migration data. http://biorxiv.org/lookup/doi/10.1101/670505 (2019) doi:10.1101/670505.

11. Peuhu, E. et al. MYO10-filopodia support basement membranes at pre-invasive tumor boundaries. Dev. Cell 57, 2350-2364.e7 (2022).

12. Roy, N. H. et al. LFA-1 signals to promote actin polymerization and upstream migration in T cells. J. Cell Sci. 133, (2020).