



HiPiler

Visual Exploration of Large Genome Interaction Matrices With Interactive Small Multiples.

Fritz Lekschas¹, Benjamin Bach¹, Peter Kerpedjiev², Nils Gehlenborg², and Hanspeter Pfister¹
¹Harvard John A. Paulson School of Engineering and Applied Sciences and ²Harvard Medical School

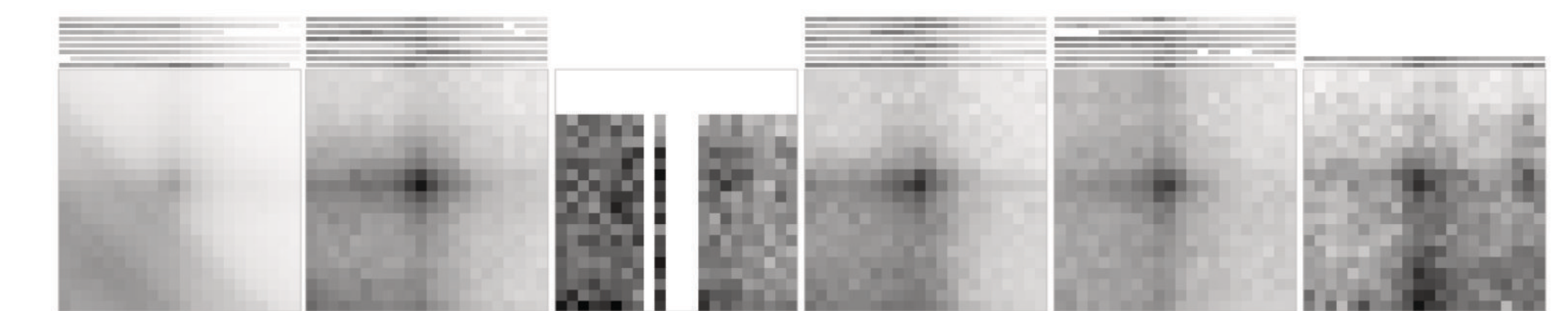


HiPiler lets you explore any kind of 2D feature and is based on your metrics and attributes.

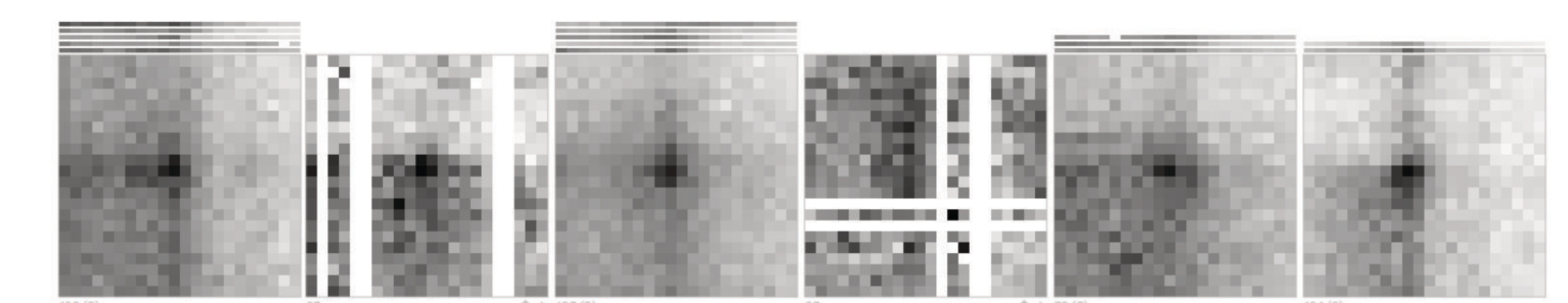
ABSTRACT

HiPiler [1] is an interactive web application for exploring and visualizing many features in large genome interaction matrices. Traditional matrix aggregation or pan-and-zoom interfaces largely fail in supporting dynamic search, inspection, and comparison of thousands of local features (loops, TADs, etc.). HiPiler represents features as interactive thumbnail-like snippets. Snippets can be laid out and aggregated automatically based on their data and derived attributes. They are linked back to the genome matrix, visualized with HiGlass [2], for spatial context.

EXAMPLES

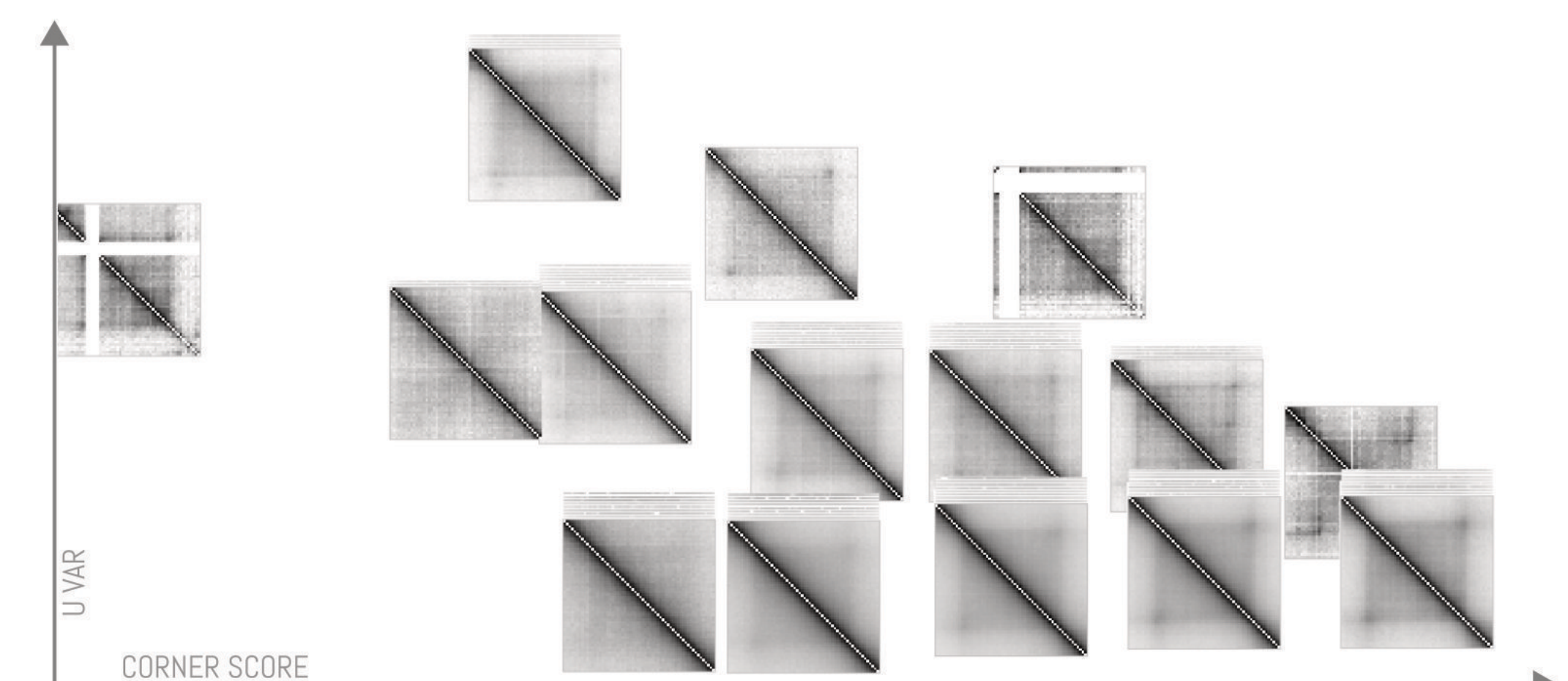


1. Overview of two Hi-C maps (top row) from which features are extracted (highlighted locations). The second row of Hi-C maps shows the detailed neighborhood of one location in Rao et al.'s [3] GM12878 and K562.

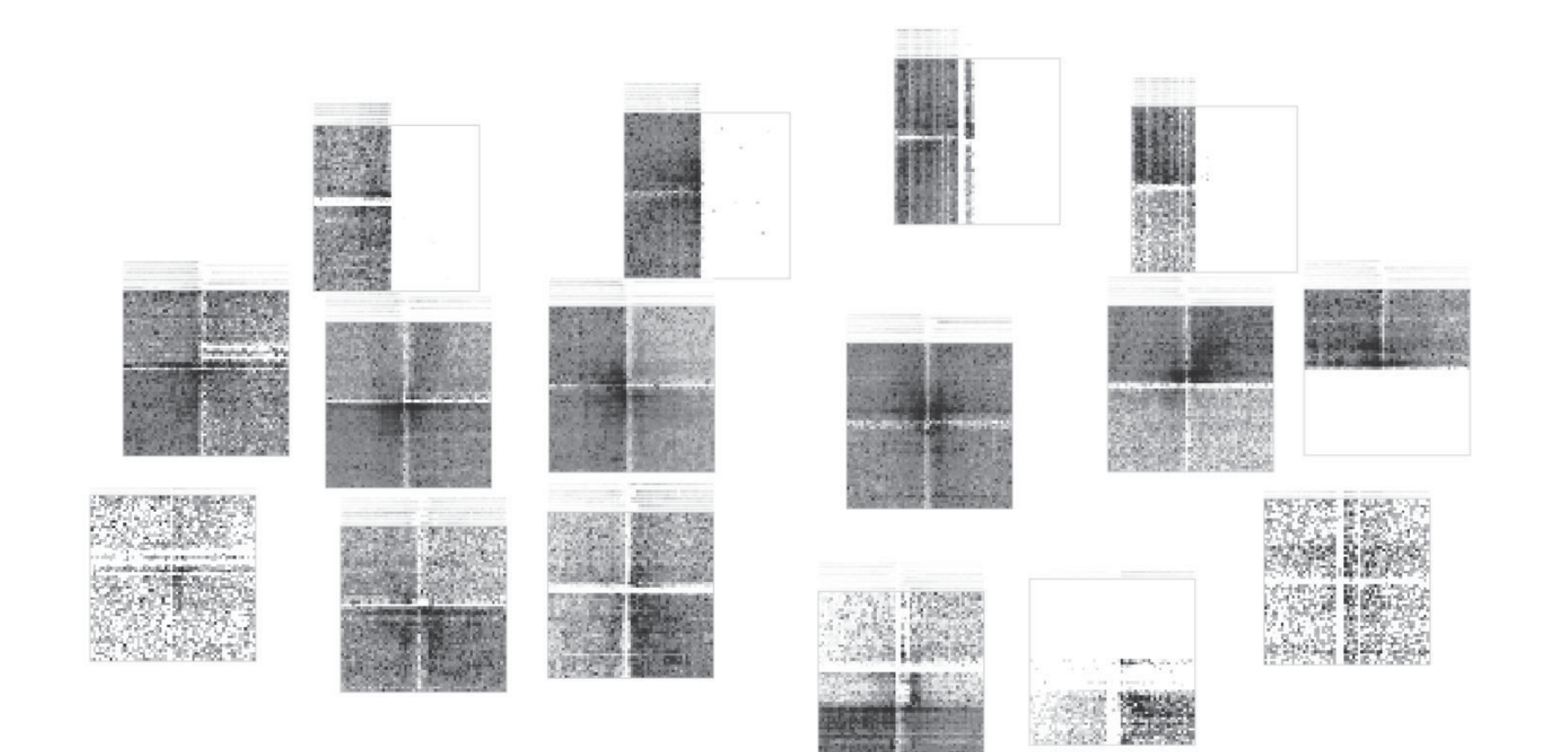


2. Telomeric regions are visualized as thumbnails, which can be filtered, arranged, grouped, and aggregated into piles. This example demonstrates multi-dimensional clustering with auto-grouping by spatial distance. An interesting group is enlarged.

Overview of loop calls from Rao et al. [3] of chromosome 22 ordered by their distance to the diagonal and grouped by similarity (k-means clustering) for initial exploration.



Analysis of TAD calls from Rao et al. [3] of chromosome 4 along corner score and U var from Arrowhead [3] for performance and parameter estimation of the algorithm. Snippets have been auto-piled by distance and log transformed for clarity.



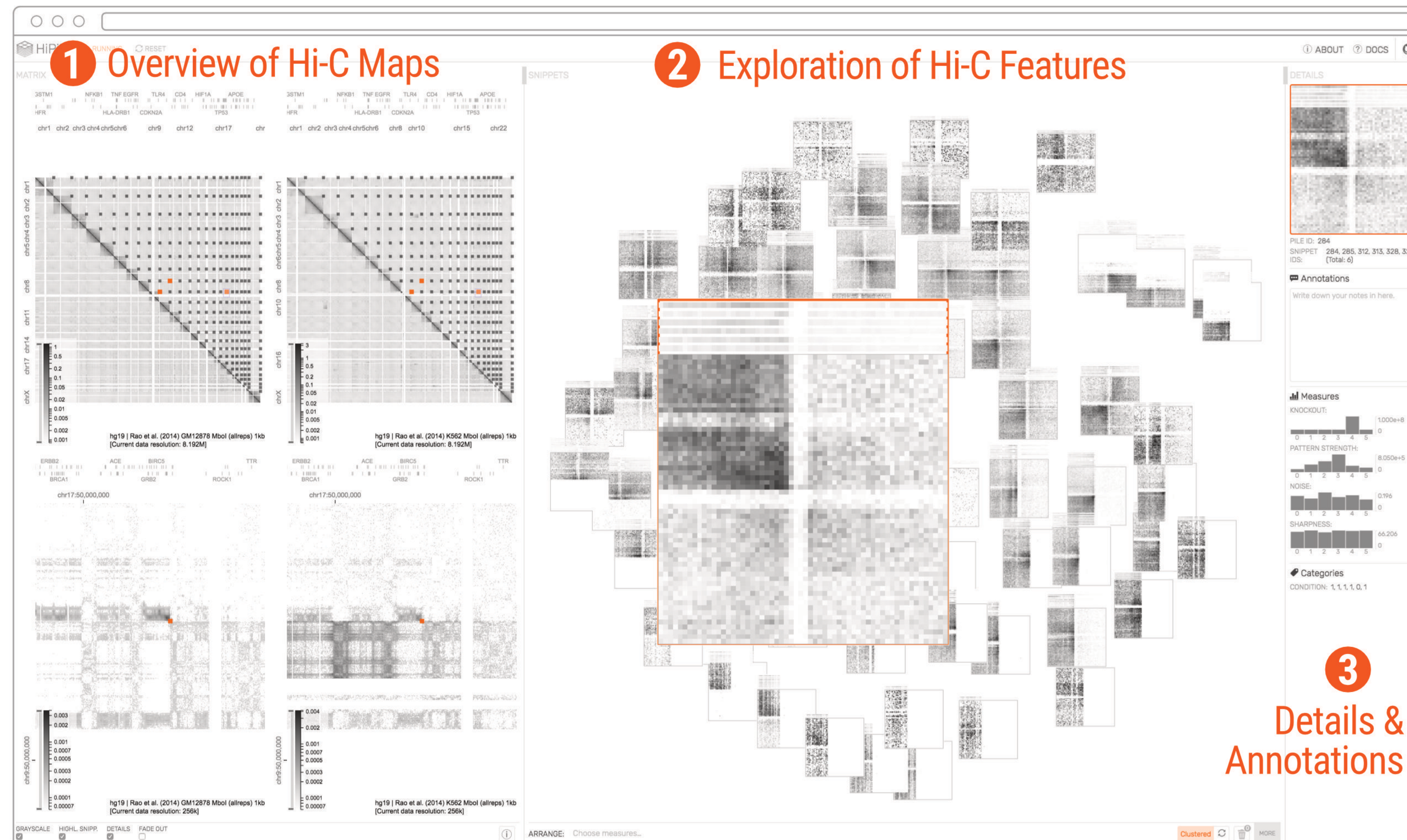
Variances in telomeric regions in Rao et al. [3] GM12878 and K562 grouped into piles and laid out by pattern-driven clustering. The colors encode the variance (log-transformed s.d.) within piles. Darker colors indicate higher variance.

FUNDING

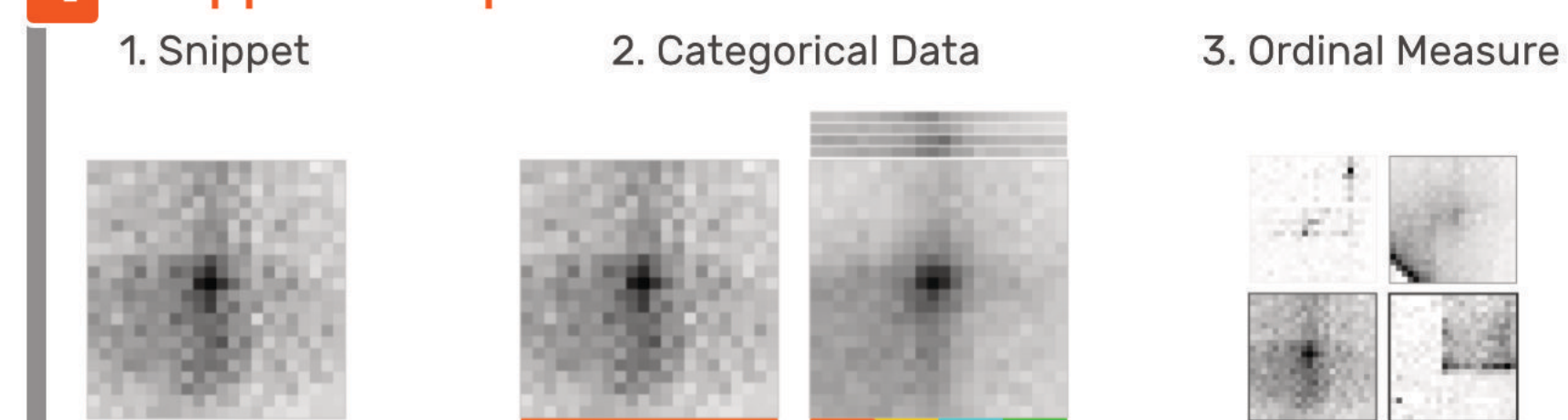
This work was performed as part of the 4DN DCIC at Harvard John A. Paulson School of Engineering and Applied Sciences and Harvard Medical School and is in part supported by NIH Grant #U01 CA200059 and R00 HG007583.

REFERENCES

- [1] Lekschas, et al. "HiPiler: Visual Exploration of Large Genome Interaction Matrices with Interactive Small Multiples." IEEE Transactions on Visualization and Computer Graphics, 2017.
 [2] Kerpedjiev et al., "HiGlass: Web-based visual comparison and exploration of genome interaction maps." bioRxiv, 2017.
 [3] Rao et al., "A 3D Map of the Human Genome at Kilobase Resolution Reveals Principles of Chromatin Looping," Cell, vol. 159, no. 7, pp. 1665–1680, Dec. 2014.



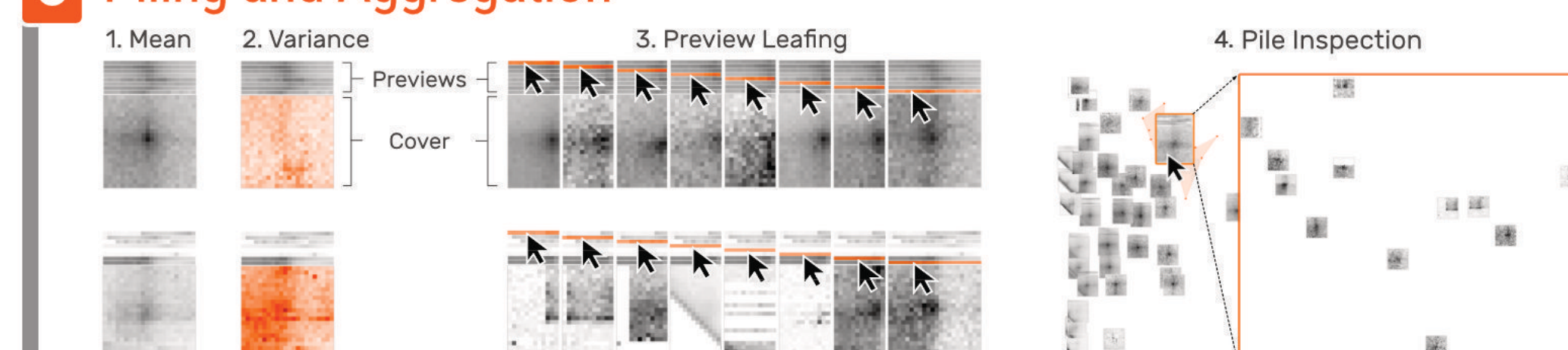
4 Snippet Metaphor



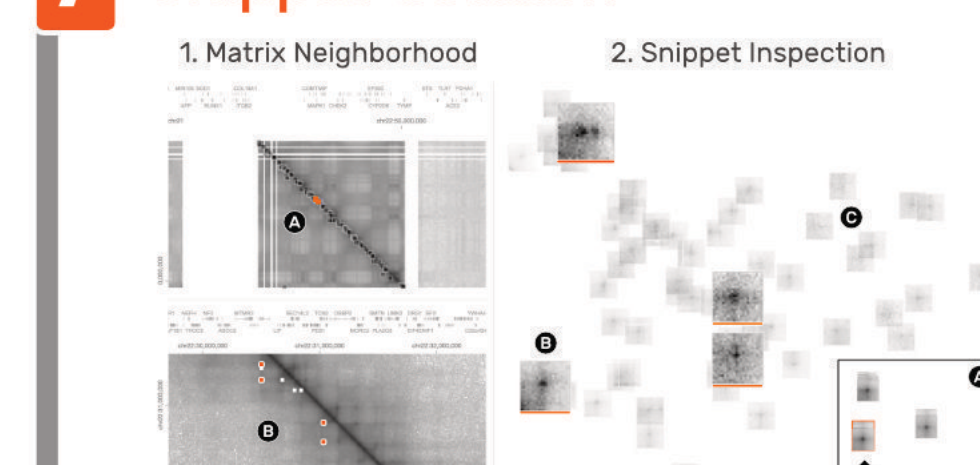
5 Snippet Layouts



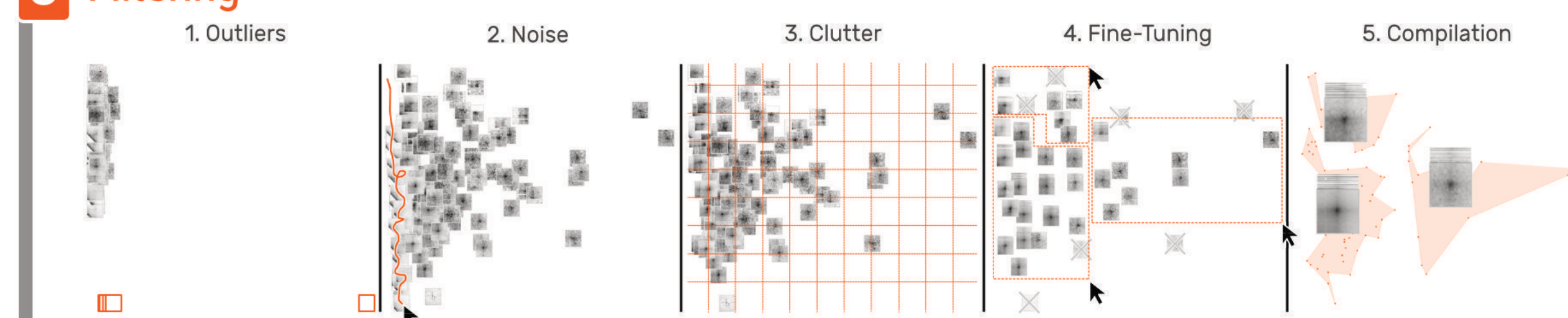
6 Piling and Aggregation



7 Snippet Context



8 Filtering



ACKNOWLEDGEMENTS

We are thankful to N. Abdennur, B. Alver, H. Belaghzal, A. van den Berg, J. Dekker, G. Fudenberg, J. Gibcus, A. Goloborodko, D. Gorkin, M. Imakaev, Y. Liu, L. Mirny, J. Nübler, P. Park, H. Strobel, and S. Wang for their invaluable feedback during the development of HiPiler. This project wouldn't have been possible without your input.

