Archetypal analysis of endogenous antigen-specific CD8 T cell response

Aarthi Venkat[1], Martina Damo[2], Nikhil S Joshi[2,5], Smita Krishnaswamy[1,3,4,5]

[1]Computational Biology and Bioinformatics Program, Yale University [2]Department of Immunobiology, Yale University [3]Department of Genetics, Yale University
[4]Department of Computer Science, Yale University [5]Yale Cancer Center

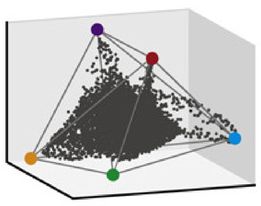
ABSTRACT

Advances in single-cell technologies, such as scRNA-seq, have enabled comprehensive profiling of immune states from a variety of tissues in health and disease. Such developments have highlighted the broad range of differentiation programs driving complex immunological responses, as well as the continuous nature of differentiation in multiple cell types. This has resulted in the emergence of a paradigm that reimagines single cells as existing on a non-uniform continuum of cell states, where cells can be described by a combination of the "extreme states". Such a framework requires a corresponding computational paradigm that can capture transitional cell states along this continuum. Archetypal analysis is a factor analysis method that describes each observation in a dataset as a convex combination of "archetypes", or pure types within the dataset. This enables the characterization of each cell by its relation to each of the archetypes.

Here, we present AANet, a deep neural network framework for identifying the archetypes in the dataset for archetypal analysis. AANet uses an autoencoder with a novel regularization on the latent layer such that the model learns the optimal transformation to represent the data in the bounds of a convex hull. This shows improvement over existing methods, which either incorrectly assume a linear relationship between features or apply fixed non-linear transformations to the data with no guarantee that they will approximate well to a simplex. We demonstrate the utility of AANet using a novel mouse model (iNversion INduced Joined neoAntigen, NINJA). NINJA mice bypass central tolerance mechanisms, enabling study of endogenous T cell responses in peripheral tissues. We define archetypes of endogenous antigen-specific CD8 T cell response in the context of liver-specific tolerance. We show that AANet enables the investigation and comparison of modulation of endogenous T cell response following antigen encounter.

AAnet provides an alternative to clustering for single-cell analysis

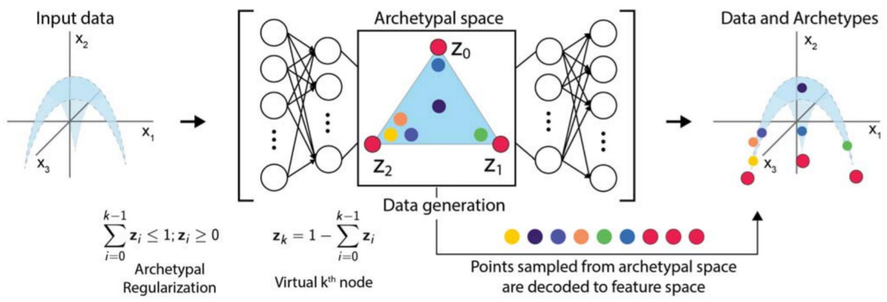
How do we study data in a way that captures a continuum of cell states?



Archetypal analysis describes each observation as a combination of "archetypes", or extreme states in the dataset.

How do we identify archetypes?

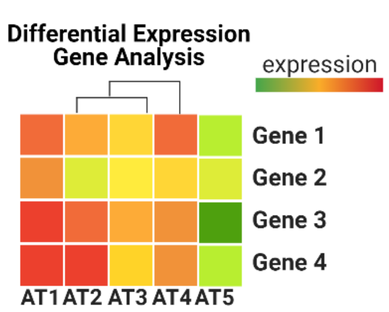
AAnet reformulates archetypal analysis with the goal of learning the ideal transformation of the data into an archetypal space bound by a simplex.

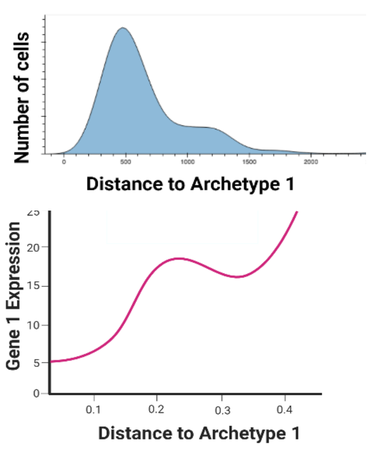


AAnet algorithm:

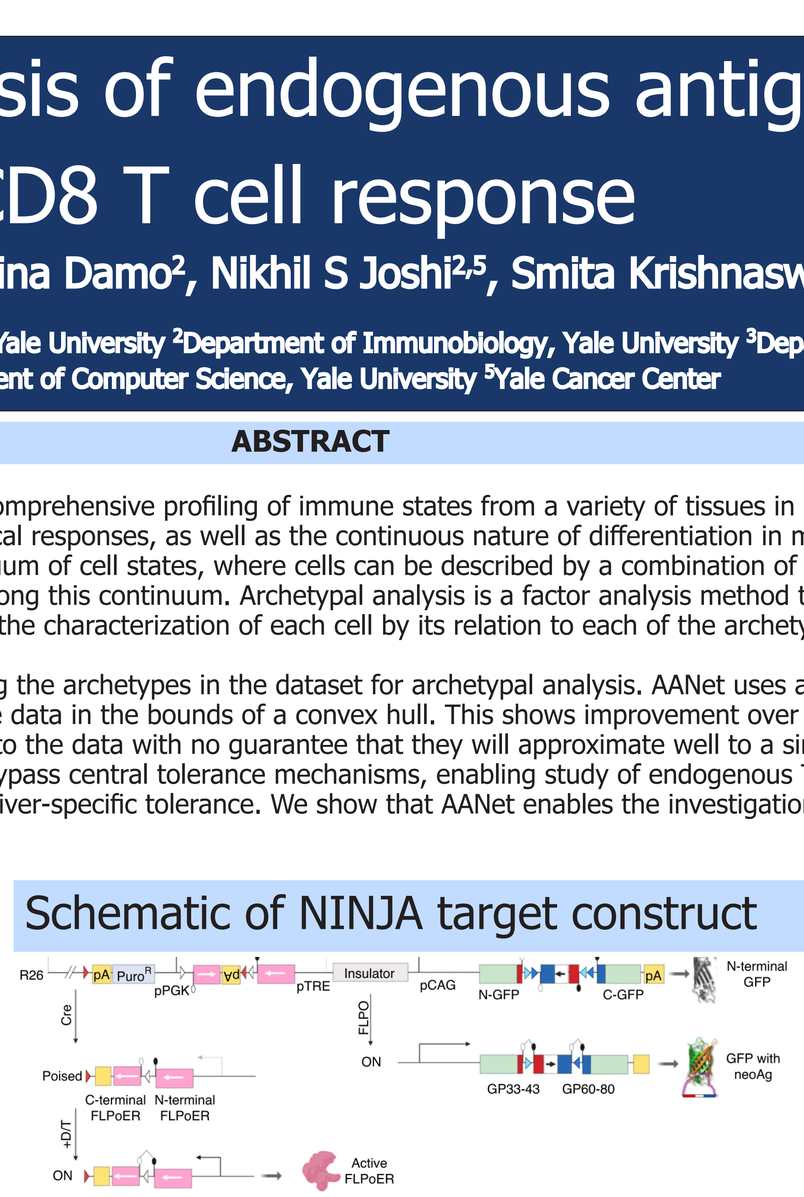
1. Autoencoder with k-dimensional latent space **Z**, where k equals the number of archetypes chosen by the user.
2. The archetypal space is regularized such that single activations of each dimension correspond to archetypes. All observations are bound by a k-dimensional simplex.
3. AAnet learns the optimal transformation from feature space **X** to **Z** and inverse function from **Z** to **X** while preserving the underlying data geometry.

The latent space has semantic structure and can be used for data exploration and data generation.





Schematic of NINJA target construct

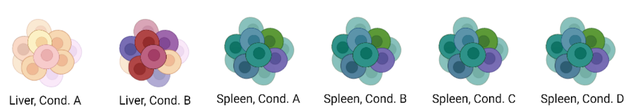


NINJA enables study of mechanisms of peripheral tolerance induction

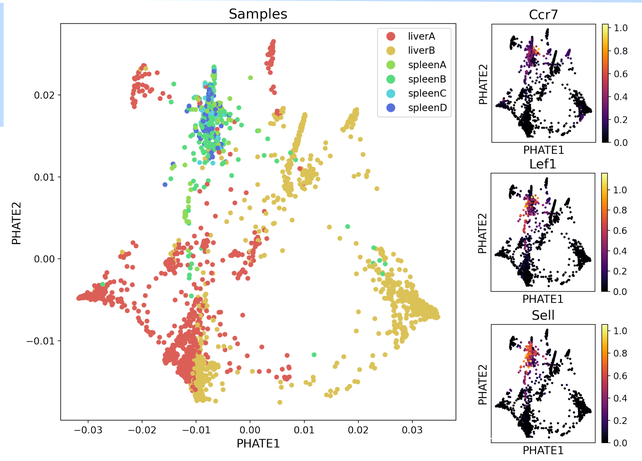
NINJA x mice carrying Cre transgene expressed under Albumin promoter. After administration of D/T, hepatocytes express antigen GP33.

Condition A: NINJA-albumin-Cre mice without D/T treatment.
Condition B: NINJA-albumin-Cre mice with D/T treatment.
Condition C: NINJA mice without Cre without D/T treatment.
Condition D: NINJA mice without Cre with D/T treatment.

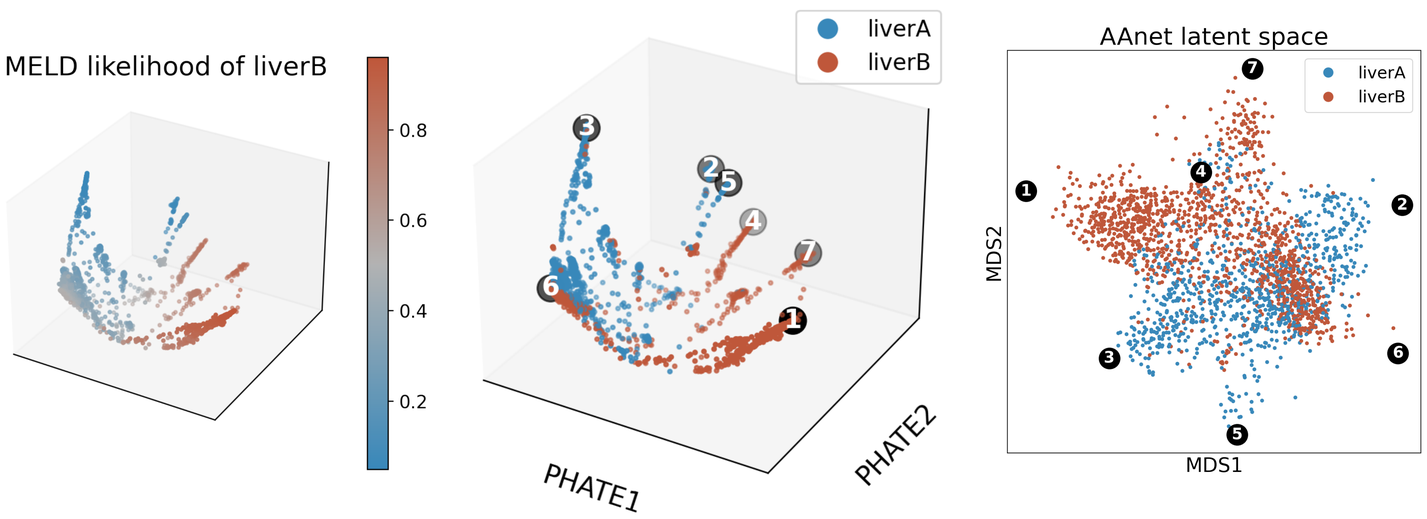
Dataset

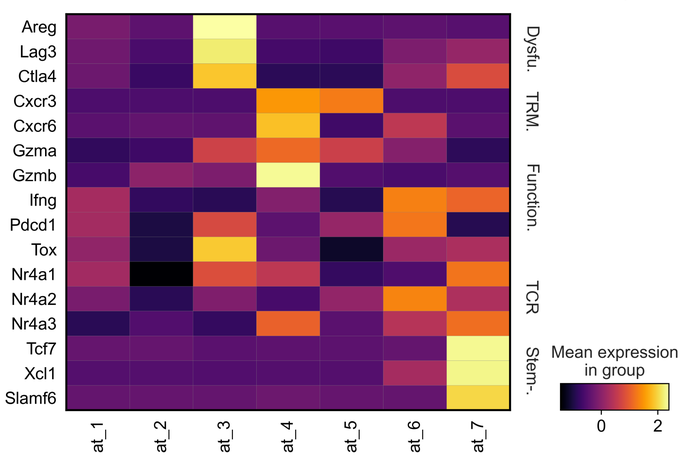


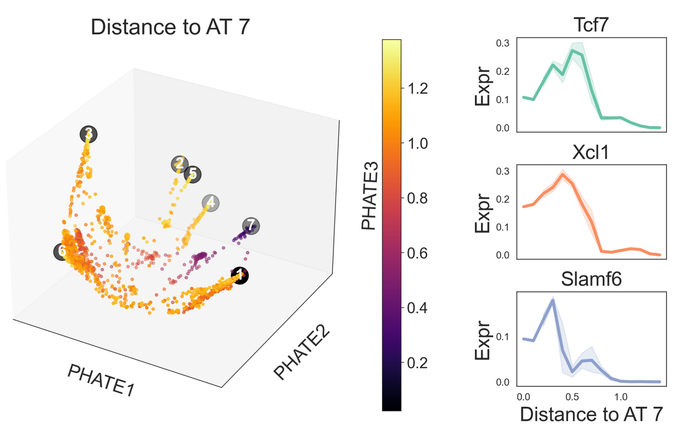
NINJA-albumin-Cre model is specific to the liver



AAnet captures signatures of CD8 T cell peripheral tolerance in NINJA mice







REFERENCES

AAnet: D. v. Dijk, D. B. Burkhardt, M. Amodio, A. Tong, G. Wolf and S. Krishnaswamy, "Finding Archetypal Spaces Using Neural Networks," 2019 IEEE International Conference on Big Data (Big Data), Los Angeles, CA, USA, 2019, pp. 2634-2643.

NINJA: Damo, M., Fitzgerald, B., Lu, Y. et al. Inducible de novo expression of neoantigens in tumor cells and mice. Nat Biotechnol 39, 64–73 (2021).

PHATE: Moon, K. R. et al. Visualizing Transitions and Structure for Biological Data Exploration. bioRxiv 120378 (2018). doi:10.1101/120378

MAGIC: Dijk, D. van et al. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. Cell 174, 716-729.e27 (2018).

MELD: Burkhardt, D.B., Stanley, J.S., Tong, A. et al. Quantifying the effect of experimental perturbations at single-cell resolution. Nat Biotechnol (2021). https://doi.org/10.1038/s41587-020-00803-5

All code and methods freely available on GitHub

github.com/KrishnaswamyLab

This work is supported in part by the Gruber Foundation