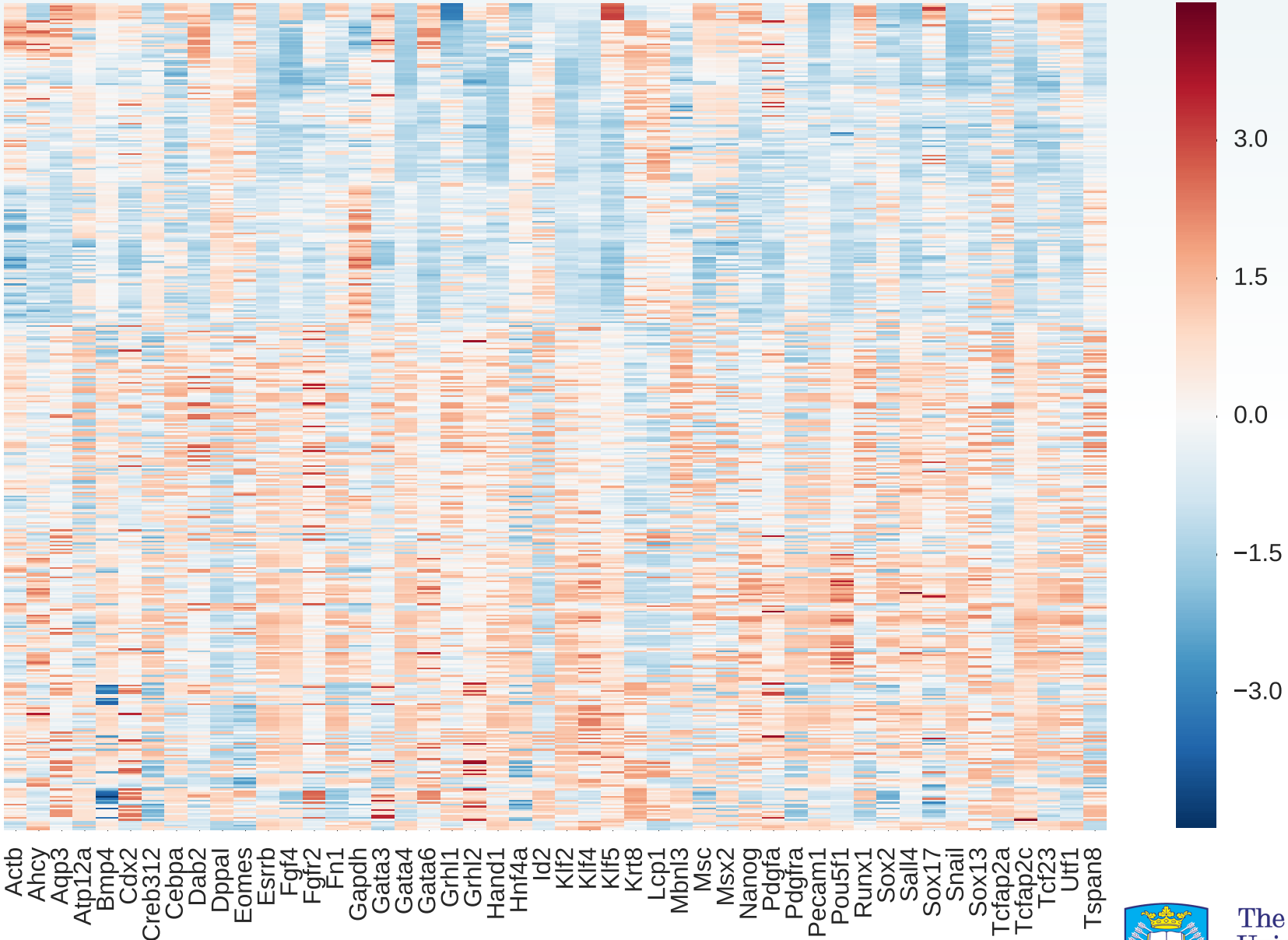
Probabilistic Dimensionality Reduction for Gene Expression Landscape Exploration

Max Zwießele

The University of Sheffield
Department of Computerscience

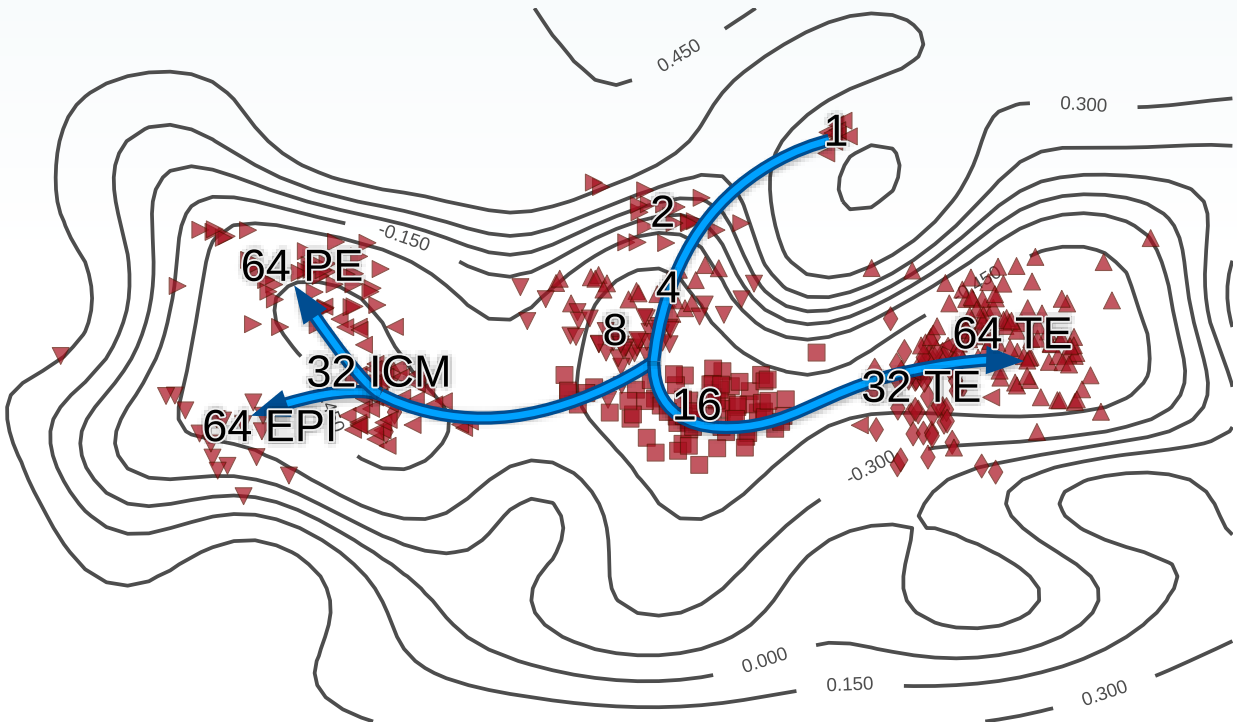
May 18, 2016

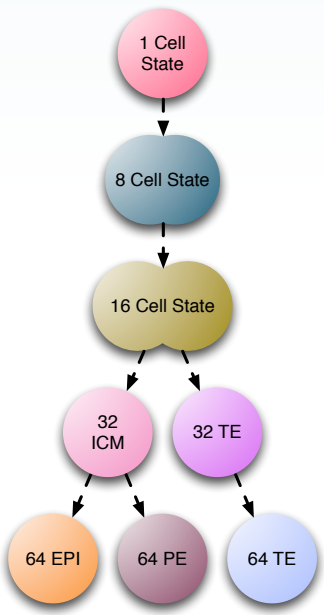
The University Of Sheffield.



The University Of Sheffield.

Dimensionality Reduction as Landscape Discovery





State of the Art

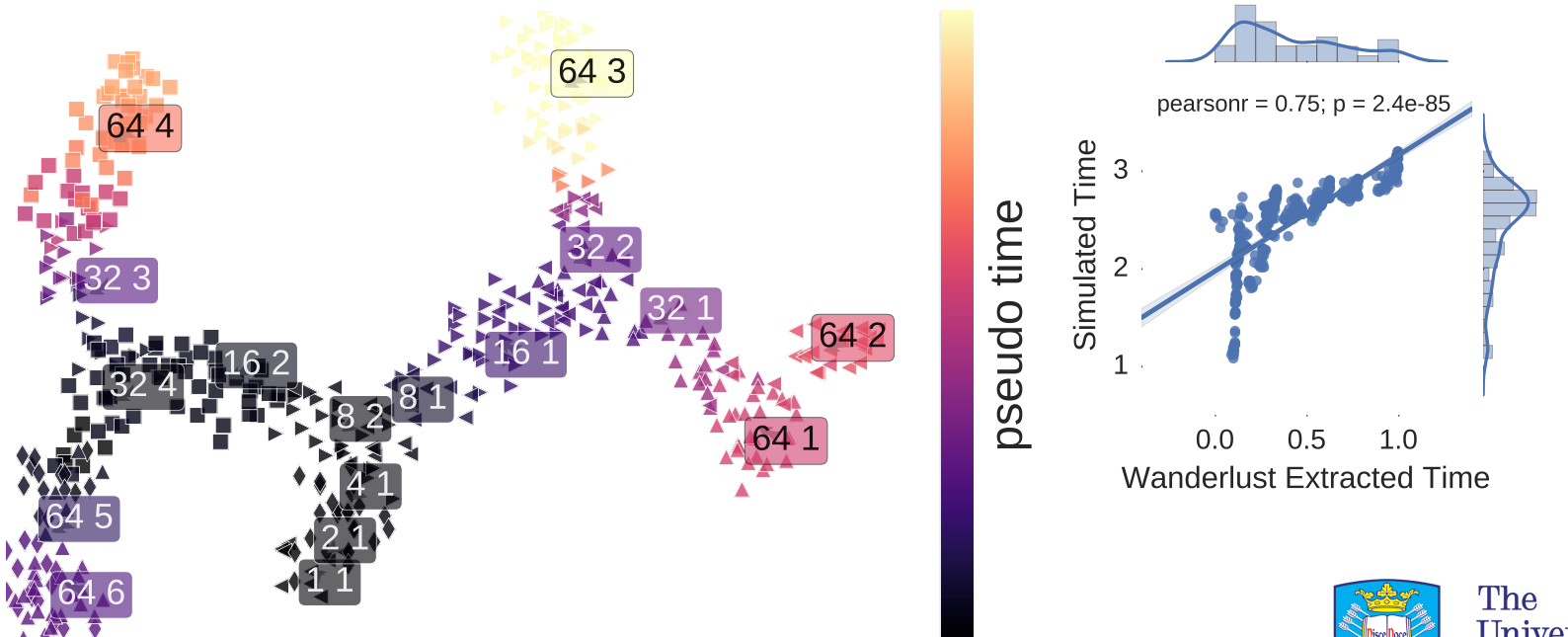
Two step approach:

- Lower dimensional representation (usually 2D).
- Given Starting Cell.
- Find ordering in representation, following
  - Minimal spanning Tree (e.g. Monocle [Trapnell et al., 2014]).
  - K-Nearest-Neighbour graph (e.g. Wanderlust [Bendall et al., 2014]).
- Post process ordering, smoothing (Diffusion map), branching ([Setty et al., 2016]), shortcut detection etc.

State of the Art

Wanderlust [Bendall et al., 2014]

- t-(distributed) Stochastic Neighberhood Embedding
- Follow "waypoints" along trajectory.
- Find smooth trajectory by majority vote of randomly generated KNN (sub-)graphs.
- Find branching structure [Setty et al., 2016]

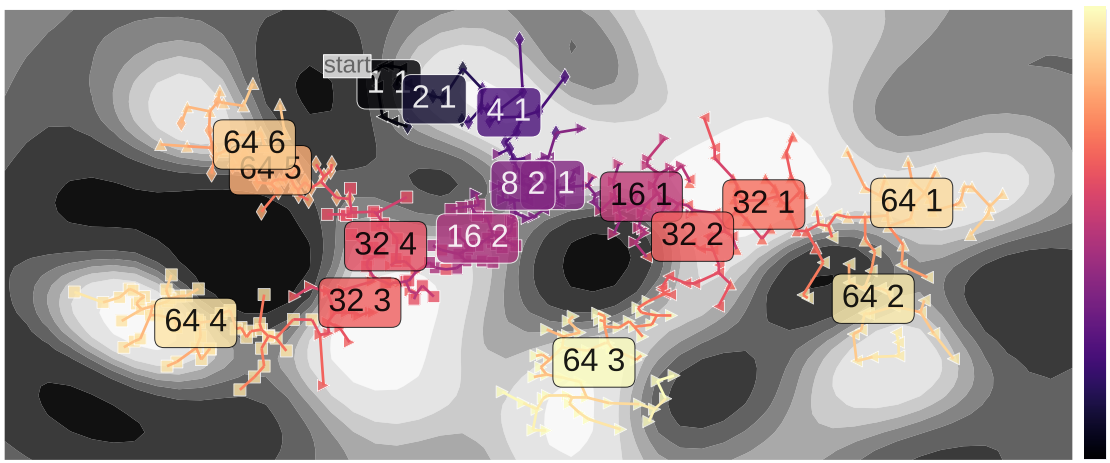


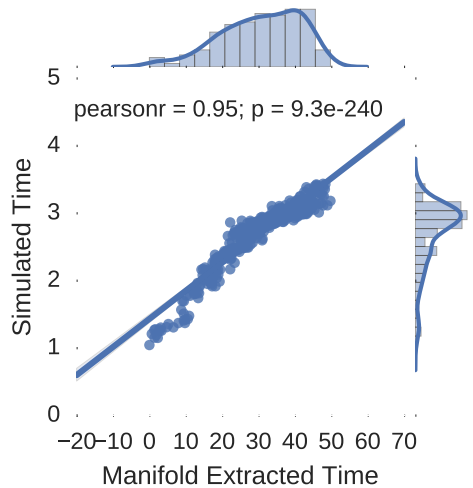
The University Of Sheffield.

Probabilistic Waddington's Landscape

topslam

- Distances along topography of landscape.
- Probabilistic correction for distortions in landscape.
- Graph to extract time line.

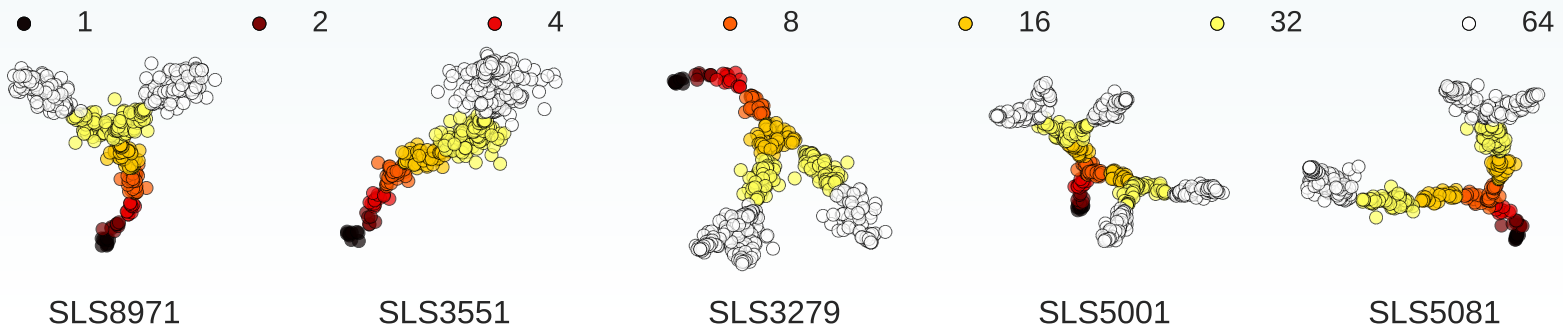


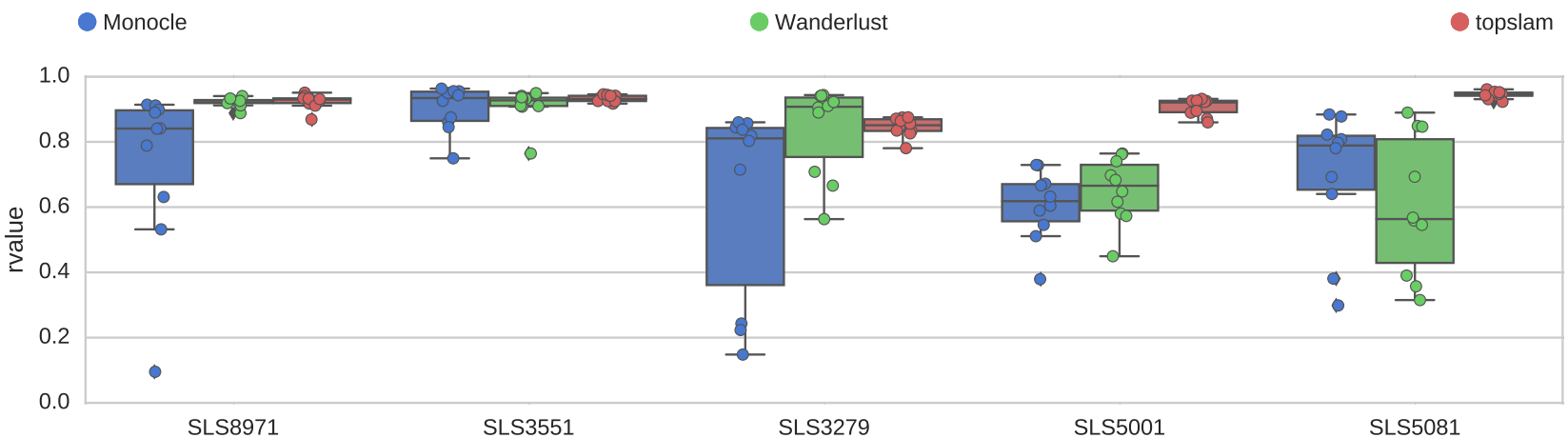


The University Of Sheffield.

Probabilistic Waddington's Landscape

topslam





Correlation Coefficients ρ between Simulated and Extracted Time.

The University Of Sheffield.

Takehome

Pseudotime Extraction

- Dimensionality reduction technique introduces errors.
- Correction usually involves heuristics or additional data.
- Principled correction by probabilistic interpretation of Waddington's landscape [Waddington, 1966].

Takehome

Publication Pending

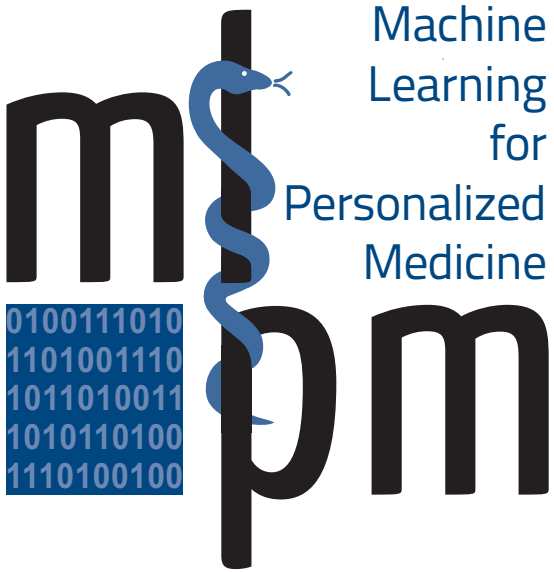
Python package topslam (based on GPy [GPy, 2012]):

https://github.com/mzwiessele/topslam

```
@Misc{topslam2016,
  author = {{Max Zwiessele}},
  title = {{topslam}: Probabilistic Epigenetic Landscapes
          for Single Cell Gene Expression Experiments},
  howpublished = {\url{
         https://github.com/mzwiessele/topslam
         }},
  year = {since 2016}
}
```

Acknowledgements

I am grateful for financial support from the European Union 7th Framework Programme through the Marie Curie Initial Training Network "Machine Learning for Personalized Medicine" MLPM2012, Grant No. 316861.



The University Of Sheffield.

Takehome

References I

Bendall, S. C., Davis, K. L., Amir, E.-a. D., Tadmor, M. D., Simonds, E. F., Chen, T. J., Shenfeld, D. K., Nolan, G. P., and Peer, D. (2014).
Single-cell trajectory detection uncovers progression and regulatory coordination in human B cell development.
Cell, 157(3):714–725.

GPy (since 2012).
GPy: A Gaussian process framework in python.
`http://github.com/SheffieldML/GPy`.

Setty, M., Tadmor, M. D., Reich-Zeliger, S., Angel, O., Salame, T. M., Kathail, P., Choi, K., Bendall, S., Friedman, N., and Pe'er, D. (2016).
Wishbone identifies bifurcating developmental trajectories from single-cell data.
Nature Biotechnology.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., Lennon, N. J., Livak, K. J., Mikkelsen, T. S., and Rinn, J. L. (2014).
The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.
Nature biotechnology, 32(4):381–386.

Waddington, C. (1966).
Principles of development and differentiation, CH Waddington.
Current concepts in biology series.

The University Of Sheffield.