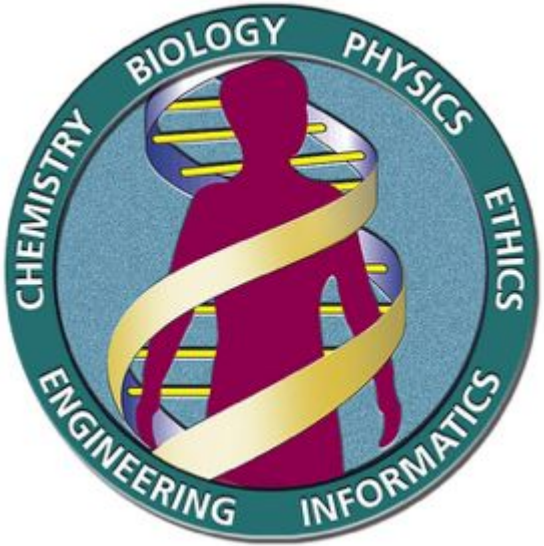
Reference-mapping in single-cell genomics

Jaison Jain
Computational Biologist, Satija Lab

Reference mapping in genomics

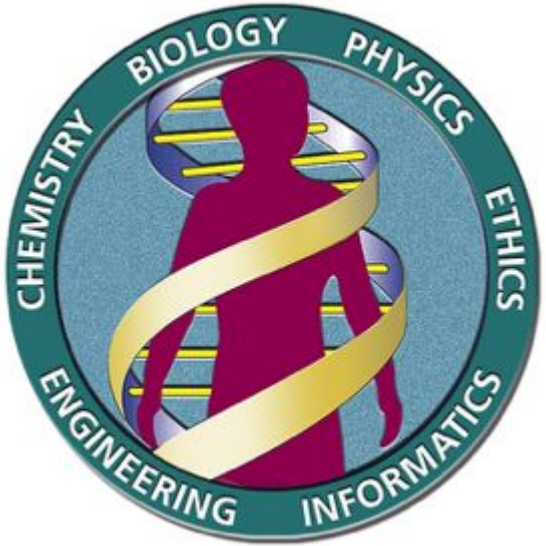
Human Genome Project



- International collaborative effort (~10 years)
- Required high-quality data (~500 bp reads)
- Complex computational tools (“assemblers”)

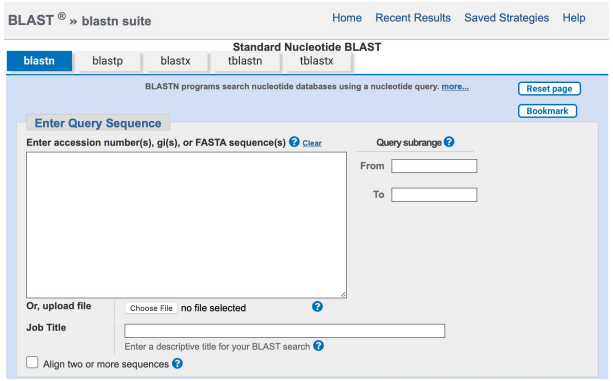
Reference mapping in genomics

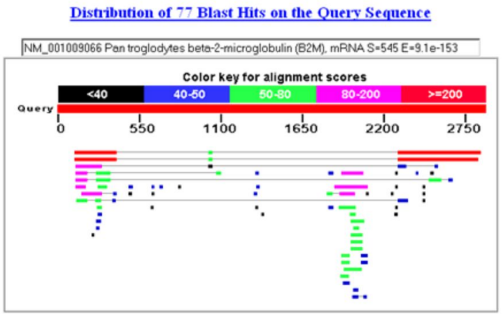
Human Genome Project



- International collaborative effort (~10 years)
- Required high-quality data (~500 bp reads)
- Complex computational tools (“assemblers”)

Can be leveraged to rapidly & reliably annotate new data





Reference mapping in single-cell genomics

A Single-Cell Transcriptome Atlas of the Human Pancreas

Mauro J. Muraro,[1,5] Gitanjali Dharmadhikari,[1,5] Dominic Grün,[1,2] Nathalie Groen,[4] Tim Dielen,[1] Erik Jansen,[1] Leon van Gurp,[1] Marten A. Engelse,[3] Francoise Carlotti,[4] Eelco J.P. de Koning,[1,3,*] and Alexander van Oudenaarden[1,6,*]

Article | Published: 25 March 2020

Construction of a human cell landscape at single-cell level

Xiaoping Han, Ziming Zhou, [...] Guoji Guo

Article | Published: 18 November 2020

A molecular cell atlas of the human lung from single-cell RNA sequencing

Kyle J. Travaglini, Ahmad N. Nabhan, Lolita Penland, Rahul Sinha, Astrid Gillich, Rene V. Sit, Stephen Chang, Stephanie D. Conley, Yasuo Mori, Jun Seita, Gerald J. Berry, Joseph B. Shrager, Ross J. Metzger, Christin S. Kuo, Norma Neff, Irving L. Weissman, Stephen R. Quake & Mark A. Krasnow

The human body at cellular resolution: the NIH Human Biomolecular Atlas Program

HuBMAP Consortium*

Article | Open Access | Published: 24 September 2020

Cells of the adult human heart

Monika Litviňuková, Carlos Talavera-López, [...] Sarah A. Teichmann

- Single-cell data has been generated for nearly every human tissue type
- These datasets may be integrated to form high-quality references (**“reference building”**)
- Information in references may be transferred onto user queries (**“reference mapping”**)

Reference mapping in single-cell genomics

Unsupervised analysis

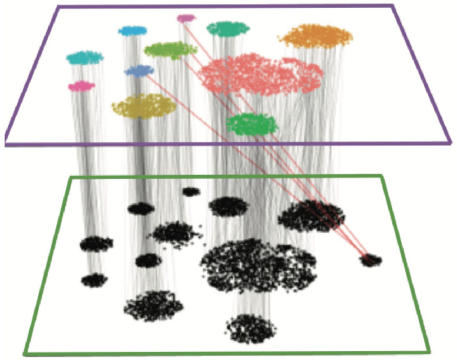
- Quality control
- Normalization and scaling
- Identification of variable genes
- Dimensional reduction (e.g,. PCA)
- Data integration (multiple studies)
- Clustering
- Identification & visualization of marker genes
- Manual cell type annotation

Supervised analysis

- Rapid and fully automated
- Consistent and easily reproducible
- High-resolution cell type annotations
- Improves interpretability of low-quality data

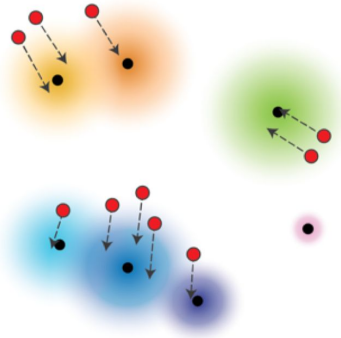
Reference mapping in single-cell genomics

Seurat v4



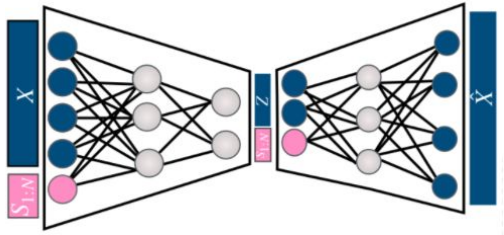
Integrated analysis of multimodal single-cell data. Yuhan Hao, Stephanie Hao [...] Peter Smibert, Rahul Satija. *bioRxiv* 2020.10.12.

Symphony



Efficient and precise single-cell reference atlas mapping with Symphony. Joyce B. Kang [...] Soumya Raychaudhuri. *bioRxiv* 2020.11.18.

scArches

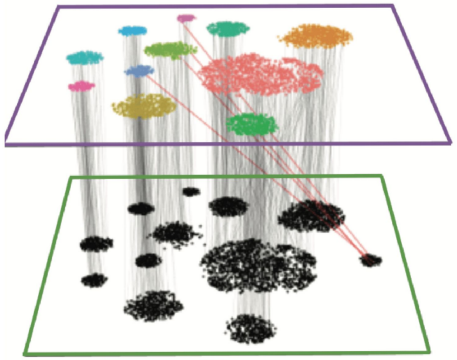


Query to reference single-cell integration with transfer learning. Mohammad Lotfollahi [...] Fabian J. Theis. *bioRxiv* 2020.07.16.

- Demo of **Azimuth**: online GUI for reference-based mapping

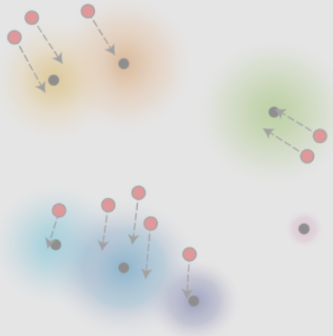
Reference mapping in single-cell genomics

Seurat v4



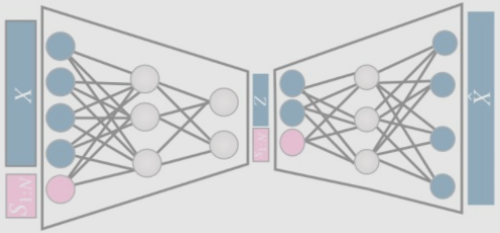
Integrated analysis of multimodal single-cell data. Yuhan Hao, Stephanie Hao [...] Peter Smibert, Rahul Satija. *bioRxiv* 2020.10.12.

Symphony



Efficient and precise single-cell reference atlas mapping with Symphony. Joyce B. Kang [...] Soumya Raychaudhuri. *bioRxiv* 2020.11.18.

scArches



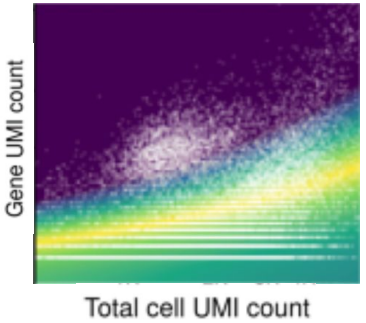
Query to reference single-cell integration with transfer learning. Mohammad Lotfollahi [...] Fabian J. Theis. *bioRxiv* 2020.07.16.

- Demo of **Azimuth**: online GUI for reference-based mapping

Reference mapping with Seurat v4

(1) Query is preprocessed using same SCTransform model as reference

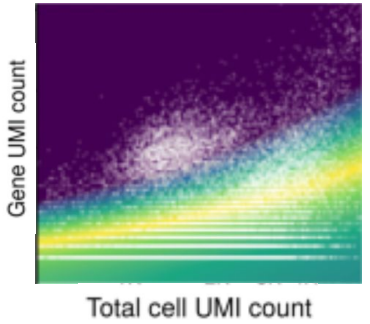
B_0 (intercept), β_1 (log_UMI), Θ (dispersion), μ (mean)



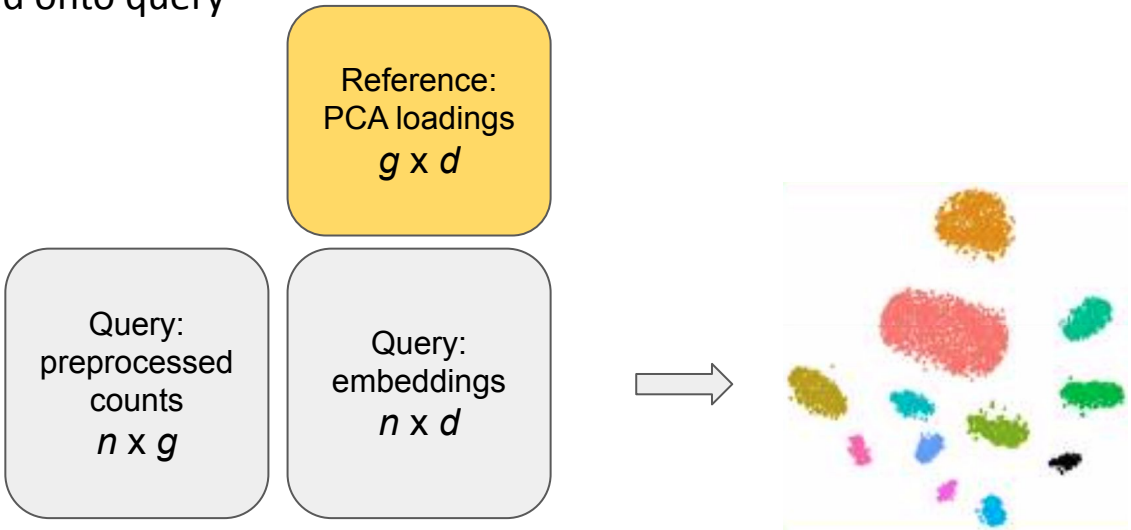
Reference mapping with Seurat v4

(1) Query is preprocessed using same SCTransform model as reference

B_0 (intercept), β_1 (log_UMI), Θ (dispersion), μ (mean)

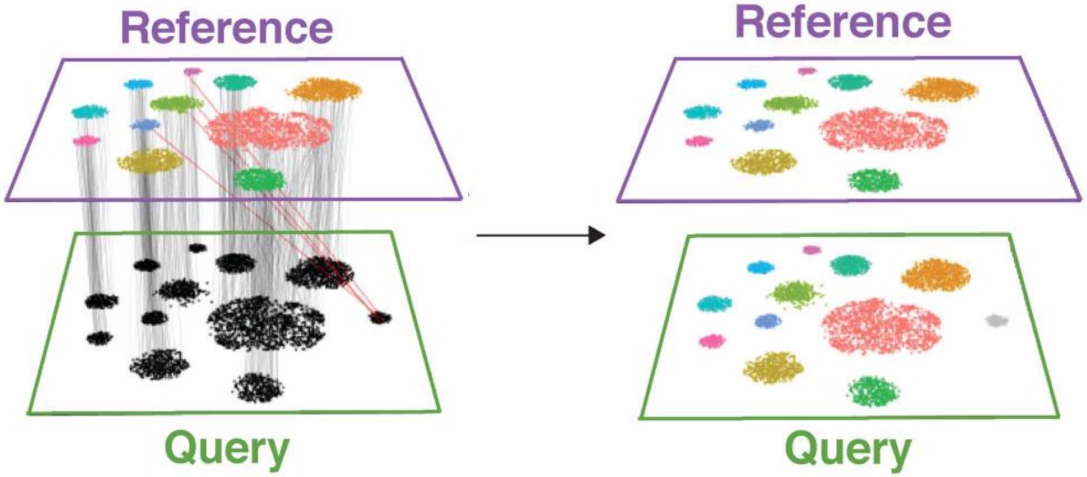


(2) Dimension reduction model (e.g, PCA, SPCA) from reference is projected onto query



Reference mapping with Seurat v4

(3) For each query cell, we can assign reference cell “anchors” if they satisfy the mutual nearest neighbor condition.

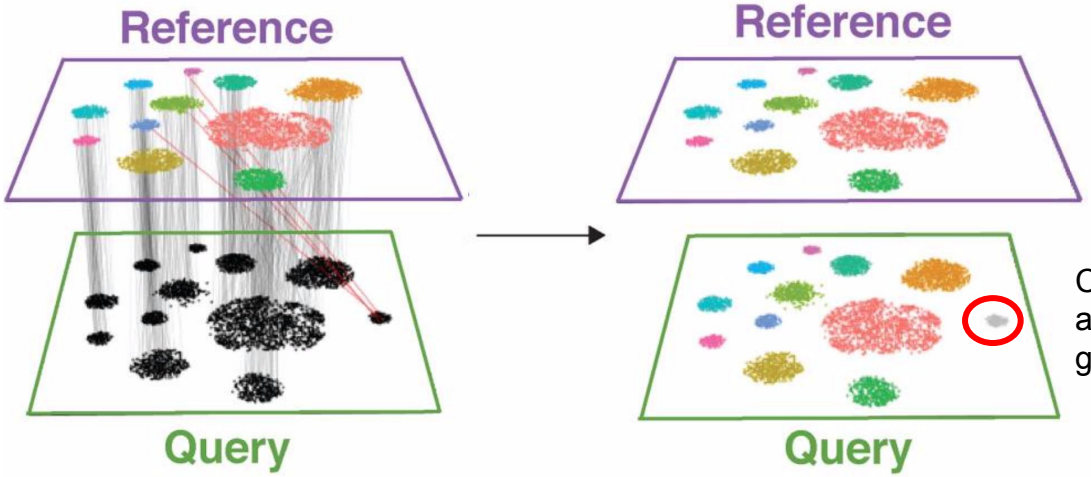


4) Each query cell can also be assigned anchor “weights”, reflecting the relative importance of nearby anchors.

5) Labels from reference can be transferred to a given query cell as a weighted sum.

Reference mapping with Seurat v4

(3) For each query cell, we can assign reference cell “anchors” if they satisfy the mutual nearest neighbor condition.

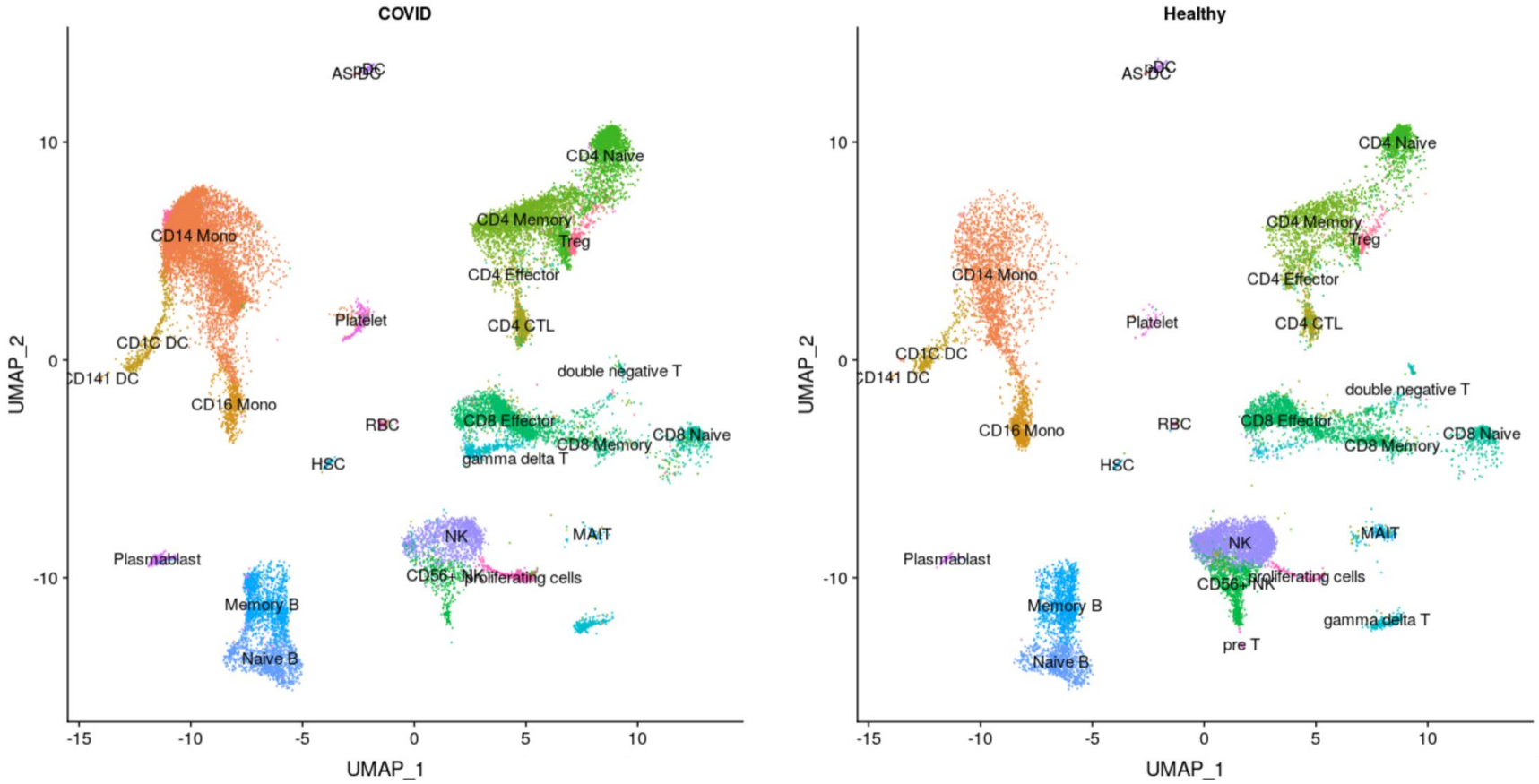


Compute mapping score and/or prediction score to gauge quality of mapping

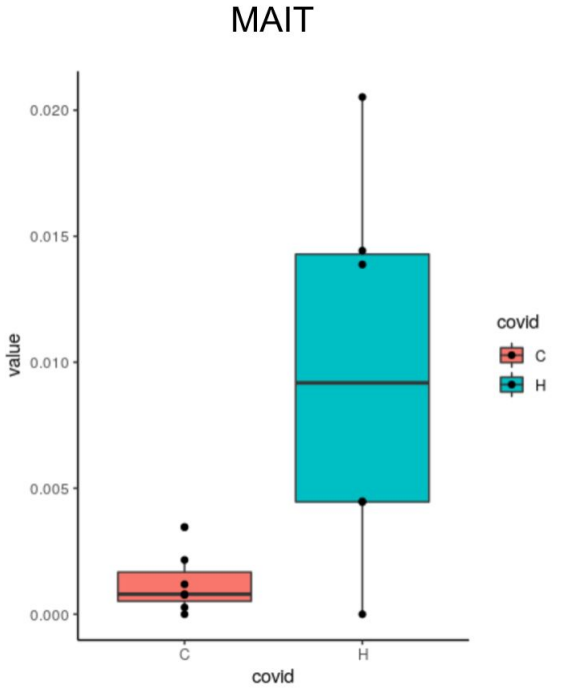
4) Each query cell can also be assigned anchor “weights”, reflecting the relative importance of nearby anchors.

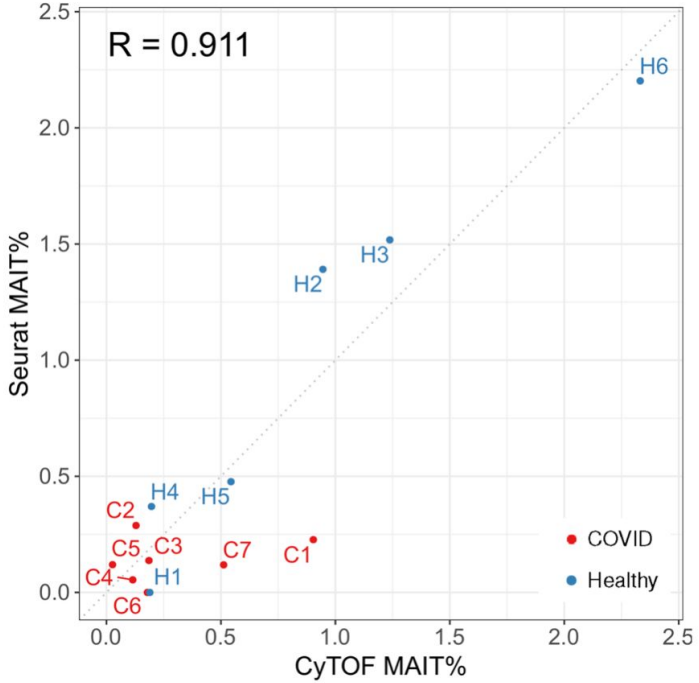
5) Labels from reference can be transferred to a given query cell as a weighted sum.

Reference mapping with Seurat v4



Reference mapping with Seurat v4

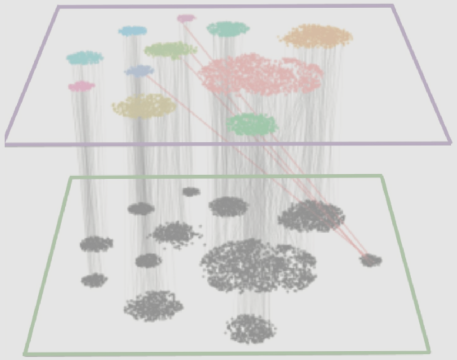




Hao*, Hao* et al, bioRxiv, 2020

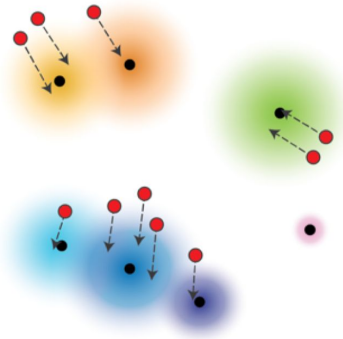
Reference mapping in single-cell genomics

Seurat v4



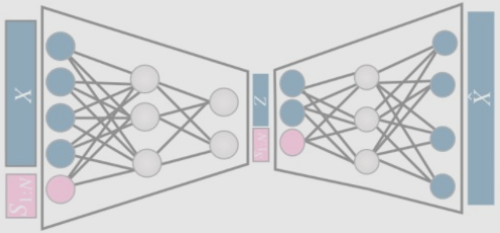
Integrated analysis of multimodal single-cell data. Yuhan Hao, Stephanie Hao [...] Peter Smibert, Rahul Satija. *bioRxiv* 2020.10.12.

Symphony



Efficient and precise single-cell reference atlas mapping with Symphony. Joyce B. Kang [...] Soumya Raychaudhuri. *bioRxiv* 2020.11.18.

scArches



Query to reference single-cell integration with transfer learning. Mohammad Lotfollahi [...] Fabian J. Theis. *bioRxiv* 2020.07.16.

- Demo of **Azimuth**: online GUI for reference-based mapping

Reference mapping with Symphony

(1) Query is preprocessed using same variable features and scaling parameters as reference
μ (mean), σ (standard deviation)

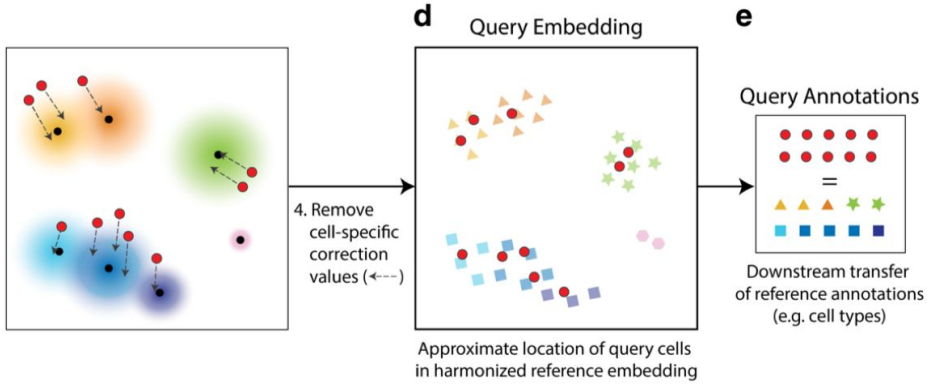
(2) Dimension reduction model (PCA) from reference is projected onto query

(3) Cells “soft-assigned” to reference clusters

(4) Query batch effects per cluster are modelled.

(5) Query data is batch corrected.

(6) Labels are transferred from reference to query using consensus of 5 nearest neighbors



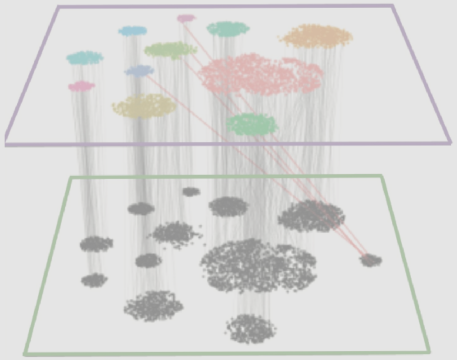
Reference mapping with Symphony

- Symphony as a 2-step anchoring procedure

Symphony	Seurat
Query scaled by reference means and standard deviations Reference PCA model is projected onto query Query cells are “soft-assigned” to reference clusters --> Query cells batch-corrected --> Query cells anchored to 5 NN	Query preprocessed with reference SCTransform model Reference PCA model is projected onto query Query cells get “weighted anchors” to reference cells

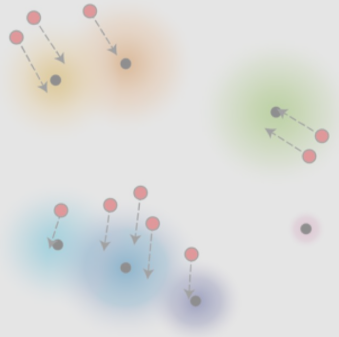
Reference mapping in single-cell genomics

Seurat v4



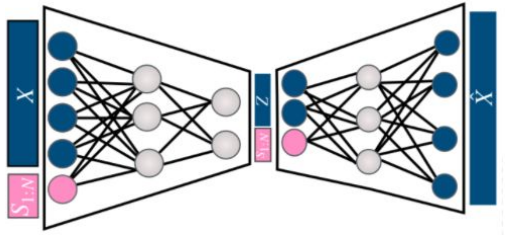
Integrated analysis of multimodal single-cell data. Yuhan Hao, Stephanie Hao [...] Peter Smibert, Rahul Satija. *bioRxiv* 2020.10.12.

Symphony



Efficient and precise single-cell reference atlas mapping with Symphony. Joyce B. Kang [...] Soumya Raychaudhuri. *bioRxiv* 2020.11.18.

scArches



Query to reference single-cell integration with transfer learning. Mohammad Lotfollahi [...] Fabian J. Theis. *bioRxiv* 2020.07.16.

- Demo of **Azimuth**: online GUI for reference-based mapping

Reference mapping with scArches

Reference building

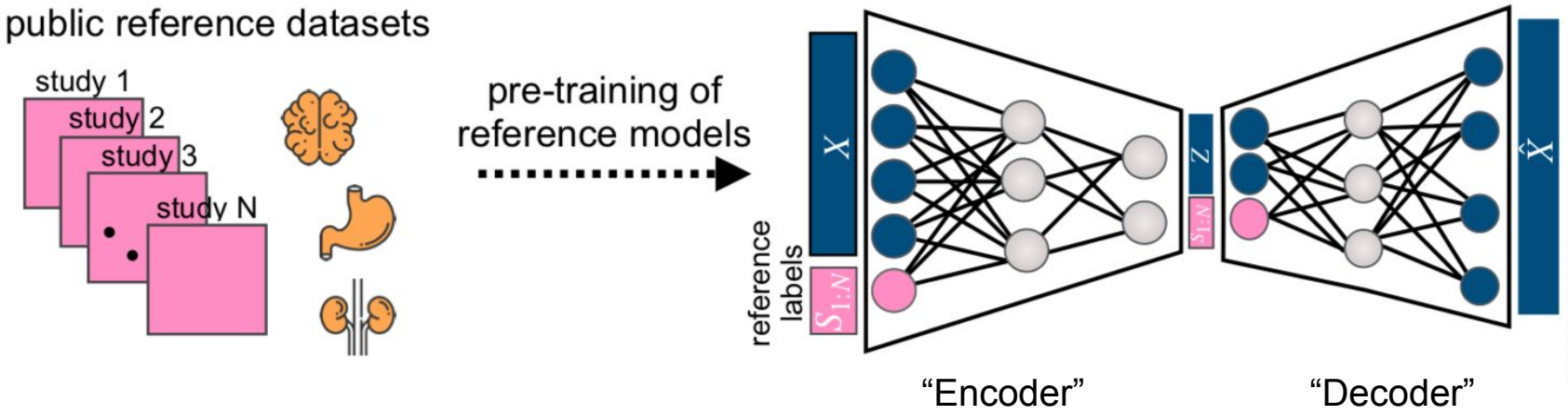


Fig 1a

Reference mapping with scArches

Reference building

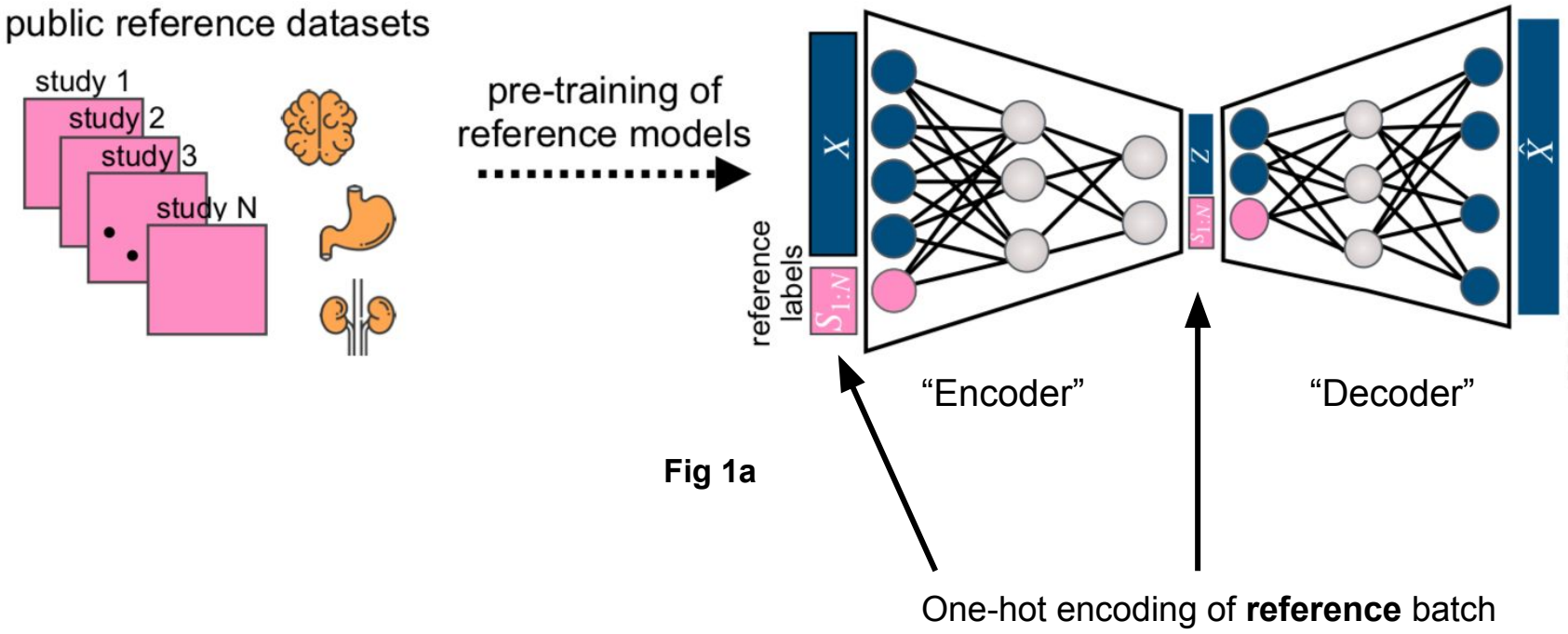
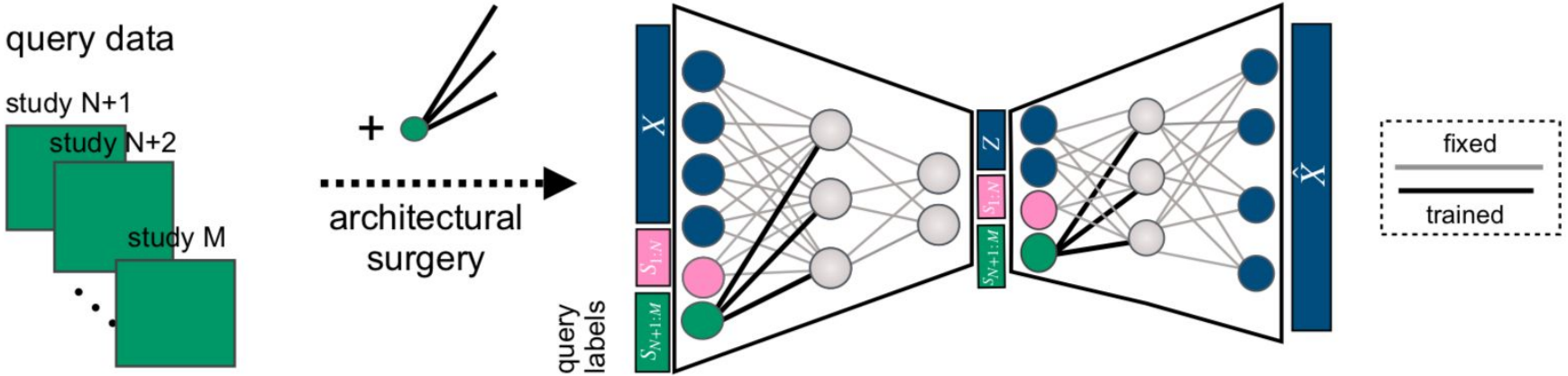


Fig 1a

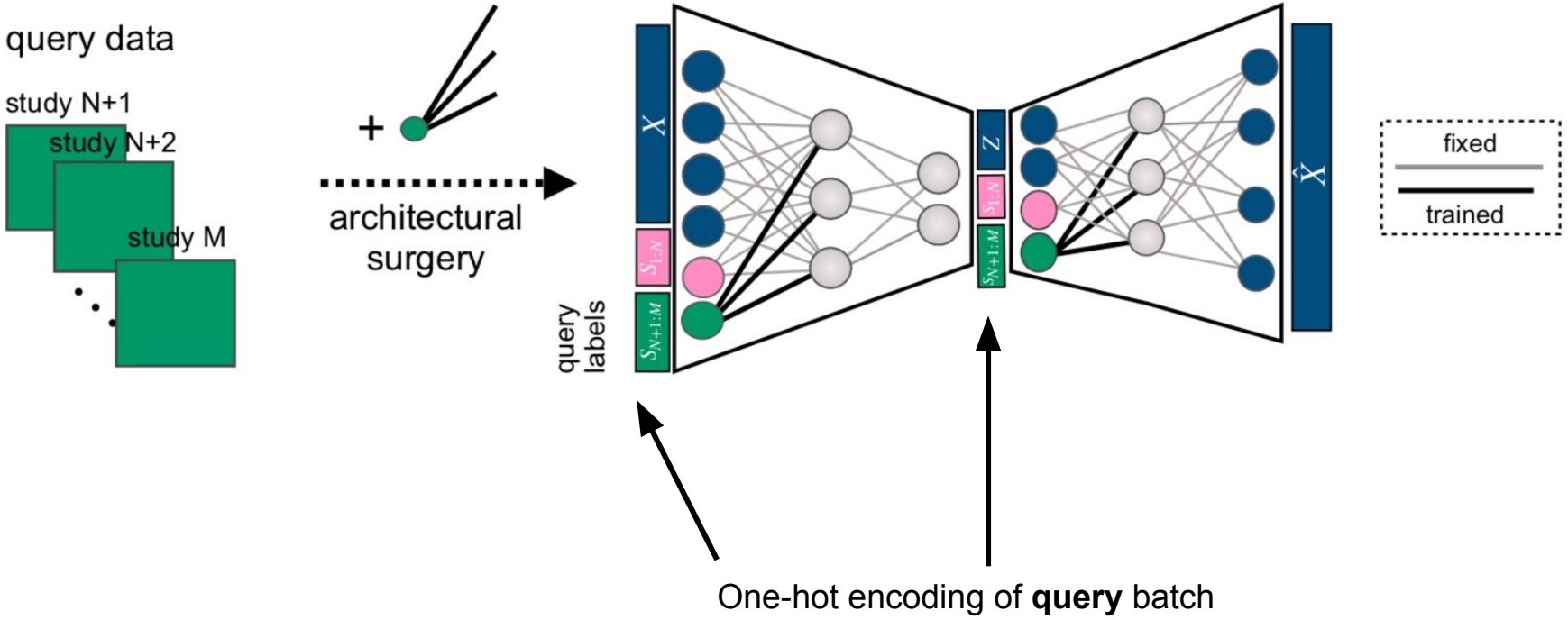
Reference mapping with scArches

Reference mapping



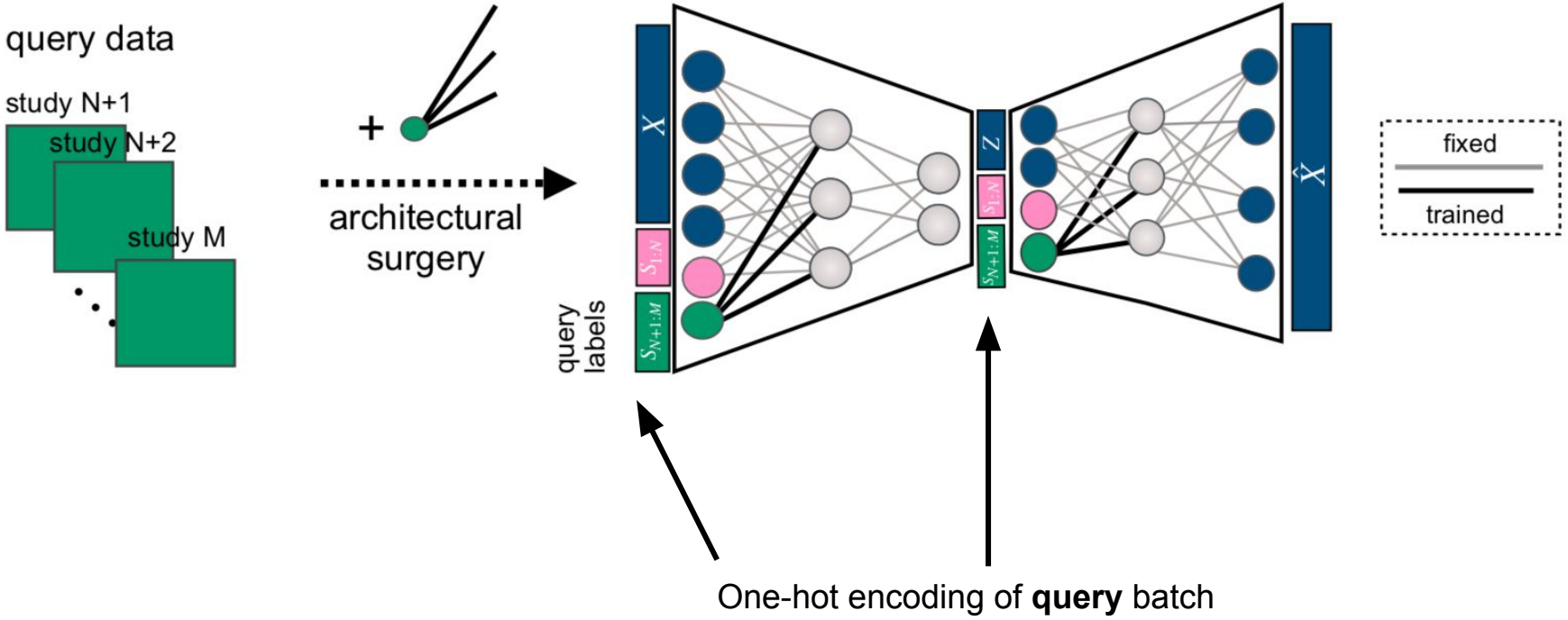
Reference mapping with scArches

Reference mapping



Reference mapping with scArches

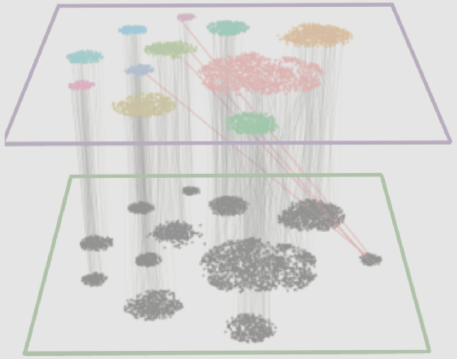
Reference mapping



- **Nearest neighbor-based approach for label transfer**

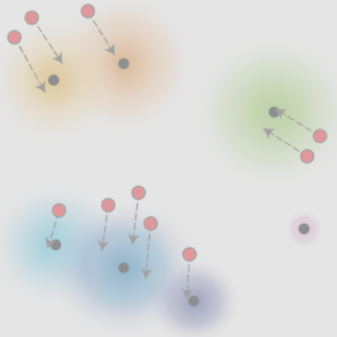
Reference mapping in single-cell genomics

Seurat v4



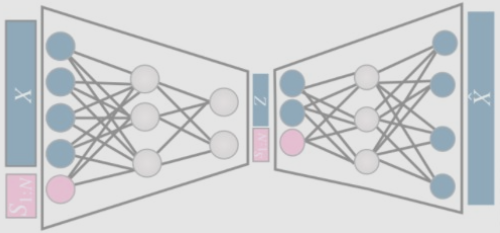
Integrated analysis of multimodal single-cell data. Yuhan Hao, Stephanie Hao [...] Peter Smibert, Rahul Satija. *bioRxiv* 2020.10.12.

Symphony



Efficient and precise single-cell reference atlas mapping with Symphony. Joyce B. Kang [...] Soumya Raychaudhuri. *bioRxiv* 2020.11.18.

scArches



Query to reference single-cell integration with transfer learning. Mohammad Lotfollahi [...] Fabian J. Theis. *bioRxiv* 2020.07.16.

- Demo of **Azimuth**: online GUI for reference-based mapping