scNMT-seq brainstorming debrief

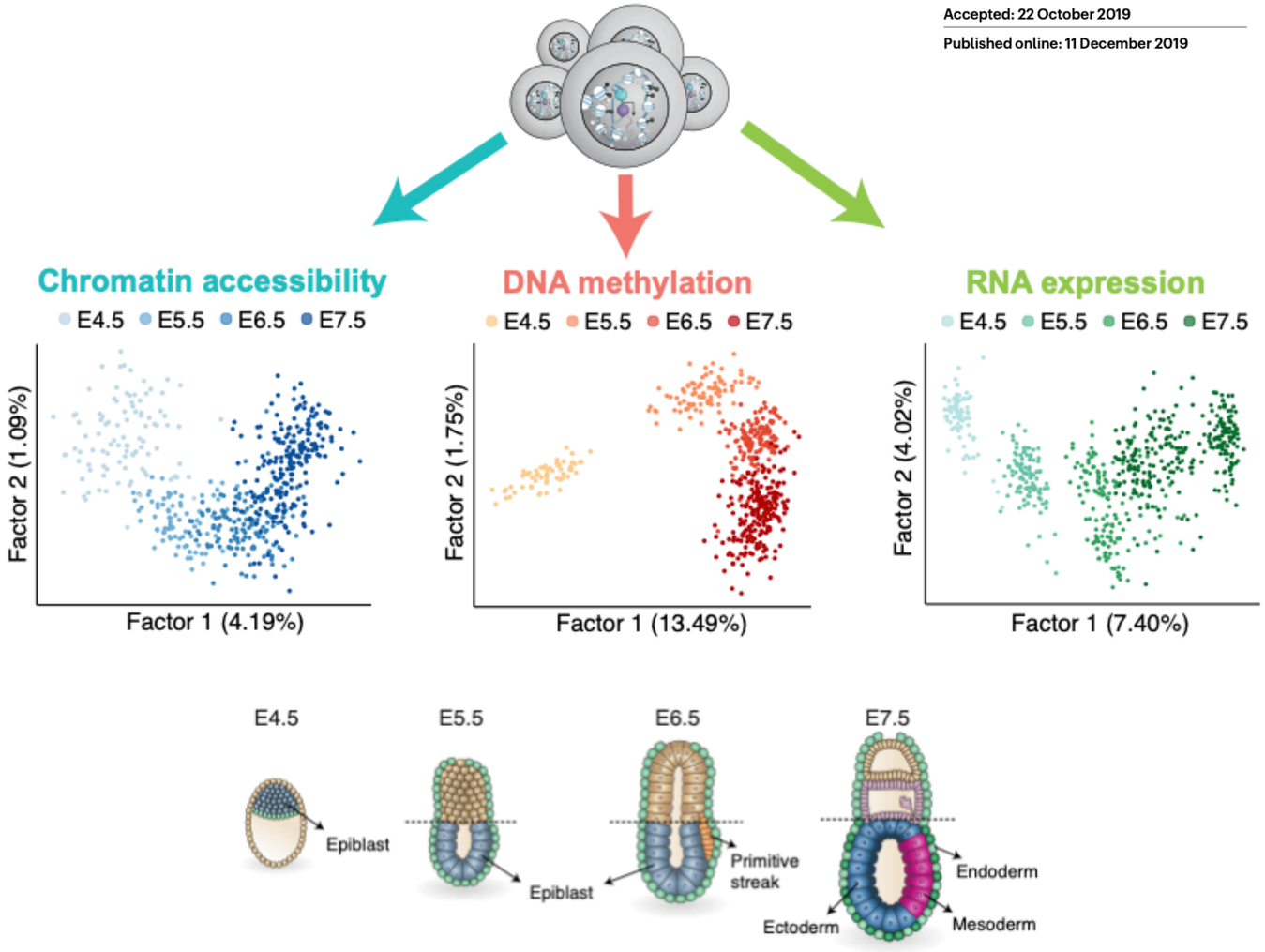
Reference	Omics	Cell type	Number of cells	Organism
Angermueller2017 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4770512/	RNA + MET	ESCs *(in vitro)*	~90	Mouse
Guo2017 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5539349/	MET + ACC	Preimplantation *(in vivo)*	~90	Mouse
Rulands2018 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6066359/	RNA + MET	Postimplantation *(in vivo)*	~150	Mouse
Clark2018 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5823944/	RNA + MET + ACC	ESCs *(in vitro)*	~90	Mouse
Clark2018 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6761124/	RNA + MET	Muscle stem cells *(in vitro)*	~350	Mouse
Linker2019 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371455/	RNA + MET	iPSC differentiation *(in vitro)*	~180	Human
Argelaguet2019 https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6924995/	RNA + MET + ACC	gastrulation *(in vivo)*	~800	Mouse
Luo2020 https://www.biorxiv.org/content/10.1101/2019.12.11.873398v1	RNA + MET + ACC (and other combinations)	Frontal cortex *(in vivo)*	>3000	Human

Data challenge

Multi-omics profiling of mouse gastrulation at single-cell resolution

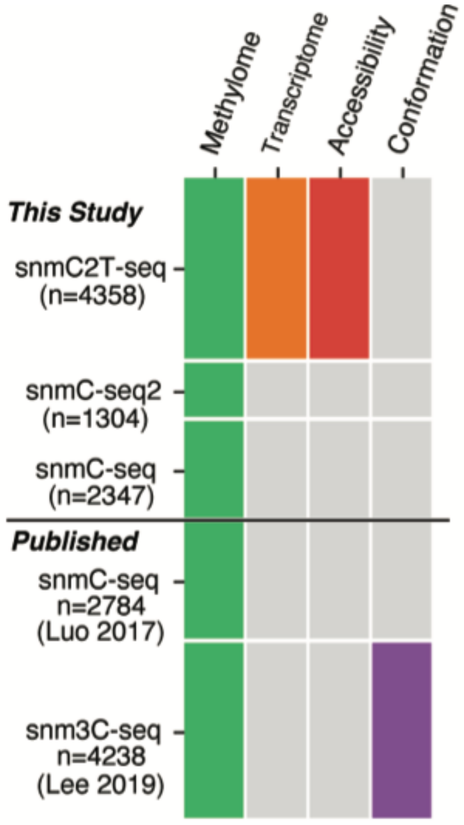
https://doi.org/10.1038/s41586-019-1825-8
Received: 18 October 2018
Accepted: 22 October 2019
Published online: 11 December 2019

Ricard Argelaguet[1,17], Stephen J. Clark[2,17]*, Hisham Mohammed[2,17], L. Carine Stapel[2,17], Christel Krueger[2], Chantriolnt-Andreas Kapourani[3,4], Ivan Imaz-Rosshandler[5,6], Tim Lohoff[2,5], Yunlong Xiang[7,8], Courtney W. Hanna[2,9], Sebastien Smallwood[2], Ximena Ibarra-Soria[10], Florian Buettner[11], Guido Sanguinetti[3], Wei Xie[7,8], Felix Krueger[12], Berthold Göttgens[5,6], Peter J. Rugg-Gunn[2,5,6,9], Gavin Kelsey[2,9], Wendy Dean[13], Jennifer Nichols[5], Oliver Stegle[1,14,15]*, John C. Marioni[1,10,16]* & Wolf Reik[2,9,16]*

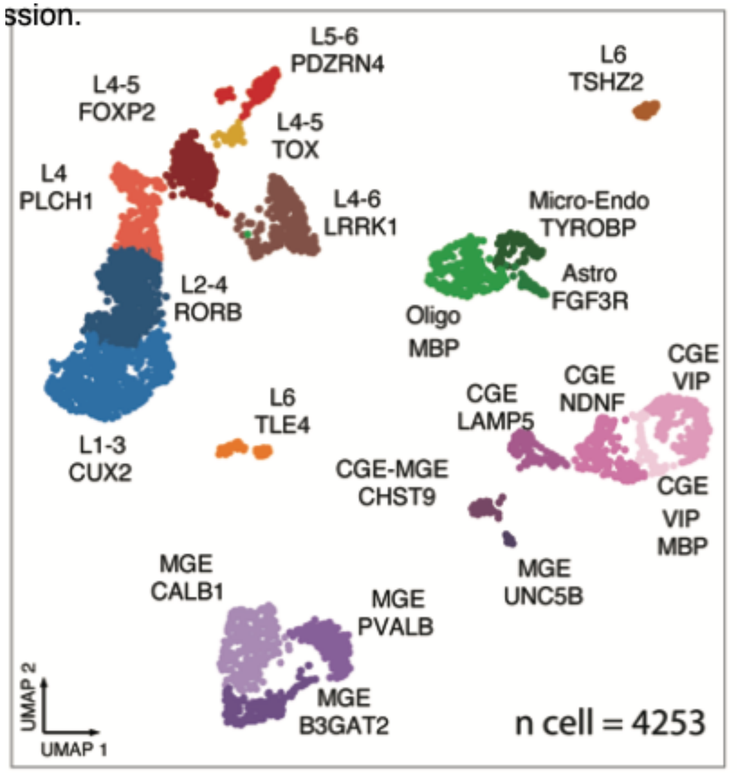


Single nucleus multi-omics links human cortical cell regulatory genome diversity to disease risk variants

Chongyuan Luo, Hanqing Liu, Fangming Xie, Ethan J. Armand, Kimberly Siletti, Trygve E. Bakken, Rongxin Fang, Wayne I. Doyle, Rebecca D. Hodge, Lijuan Hu, Bang-An Wang, Zhuzhu Zhang, Sebastian Preissl, Dong-Sung Lee, Jingtian Zhou, Sheng-Yong Niu, Rosa Castanon, Anna Bartlett, Angeline Rivkin, Xinxin Wang, Jacinta Lucero, Joseph R. Nery, David A. Davis, Deborah C. Mash, Jesse R. Dixon, Sten Linnarsson, Ed Lein, M. Margarita Behrens, Bing Ren, Eran A. Mukamel, Joseph R. Ecker



ssion.



Hackaton applications

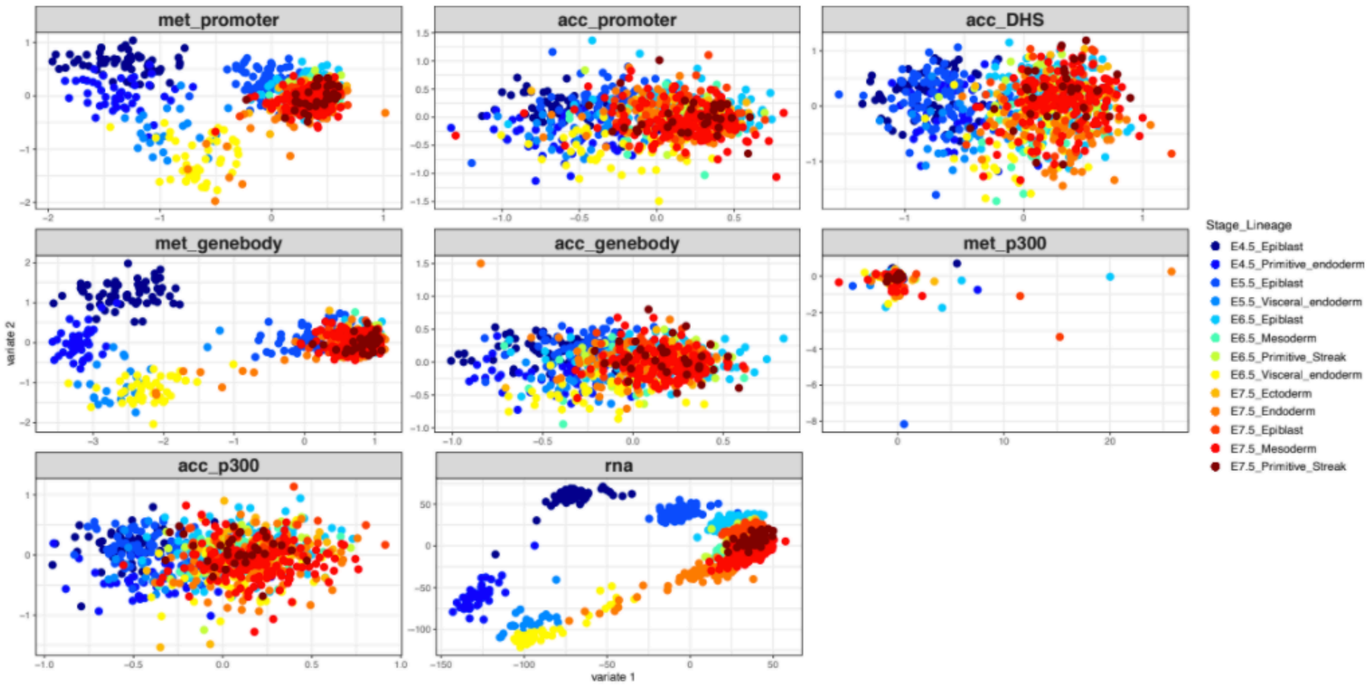
Application of LIGER to the scNMT-seq data set was (partially) unsuccessful. Why?

- Gene body/promoter epigenomic state does not align well with expression state
- Different preprocessing strategies do not change result

from Josh Welsh talk

Hackaton applications

Application of PLS to the scNMT-seq data set was (partially) unsuccessful. Why?



from Al JalalAbadi's talk

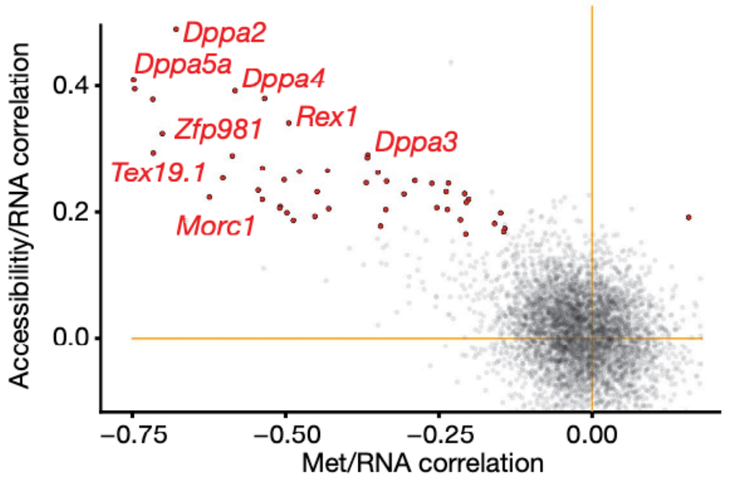
Hackaton applications

Computational challenges in scNMT-seq

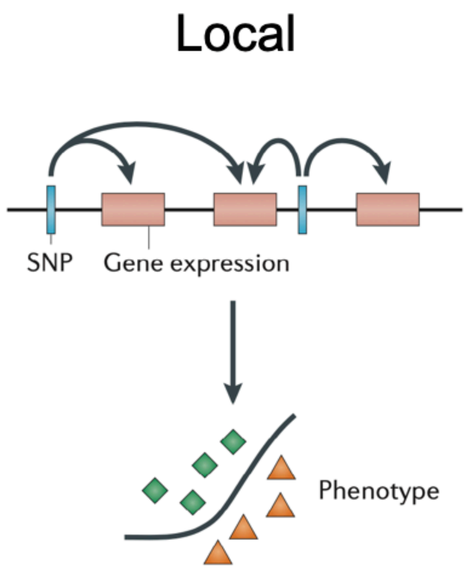
- Epigenetic readouts are extremely sparse (>80% of CpG sites not observed per cell). Integrative methods must handle NAs!

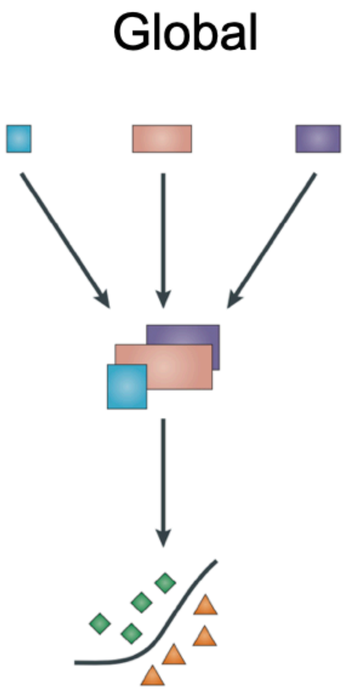
- Non-gaussian observations:
  - binary at the CpG level
  - binomial at the genomic feature level

- In embryonic stages, the relationship between mRNA expression and DNA methylation is less pronounced than in somatic tissues -> polycomb repression via histone marks



Integration strategies





Open questions

Global analysis

- How to perform dimensionality reduction with the DNA methylation data?
- What genomic contexts to use for DNA methylation quantification?
- Can we do transfer learning of epigenetic measurements onto large scRNA-seq atlas?
- How to deal with the feature imbalance between data modalities in integrative methods?

Local analysis

- How to link epigenetic features to genes?
- How to impute DNA methylation data?
- How to model (non-linear) epigenetic dynamics across pseudotime?

How to perform dimensionality reduction with DNA methylation data?

- PCA/NMF works well with continuous data. DNA methylation rates are approximately continuous with bulk measurements, but not with sparse single-cell DNA methylation data.
- Suggestions:
  - Binary distance metrics followed by MDS
  - GLM-PCA (https://cran.r-project.org/web/packages/glmpca/)
  - LSI and topic modelling (http://andrewjohnhill.com/blog/2019/05/06/dimensionality-reduction-for-scatac-data/)

What genomic contexts to use for DNA methylation quantification?

- Unsupervised:
  - genome-wide running window (bins)

- Supervised:
  - Using a reference of DHS peaks
  - Define chromatin compartments using multiple sources of epigenetic information (histone marks, etc.)

How to impute DNA methylation data?

Method | Open Access | Published: 11 April 2017

DeepCpG: accurate prediction of single-cell DNA methylation states using deep learning

Christof Angermueller, Heather J. Lee, Wolf Reik & Oliver Stegle

Genome Biology **18**, Article number: 67 (2017) | Cite this article

29k Accesses | **95** Citations | **151** Altmetric | Metrics

Melissa: Bayesian clustering and imputation of single-cell methylomes

ChantrioInt-Andreas Kapourani & Guido Sanguinetti

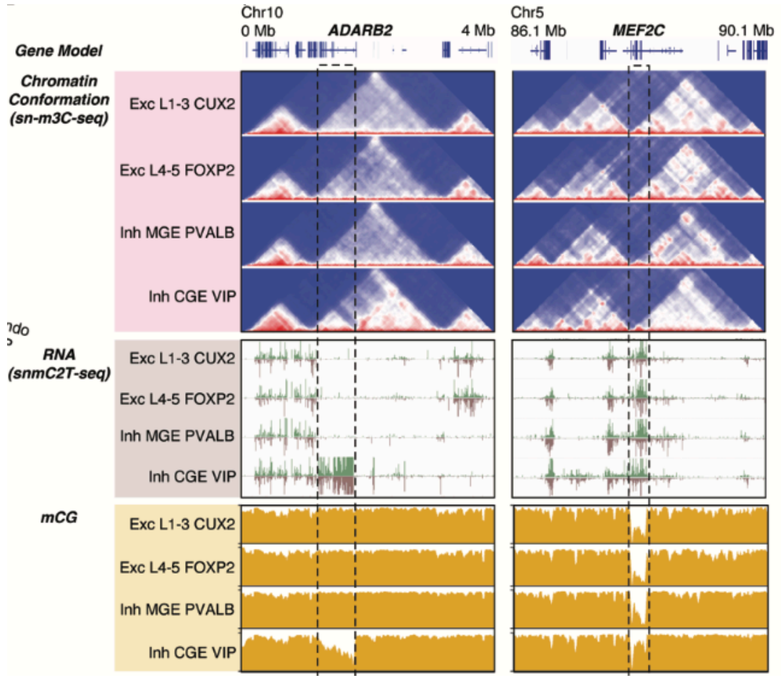
Genome Biology **20**, Article number: 61 (2019) | Cite this article

3192 Accesses | **1** Citations | **37** Altmetric | Metrics

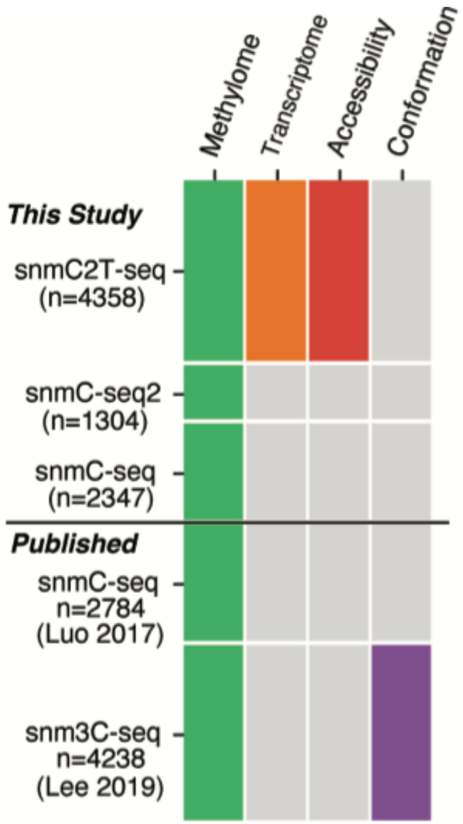
Very important to model cell-to-cell heterogeneity, otherwise you homogeneise differences between cell types

How to link epigenetic features to genes?

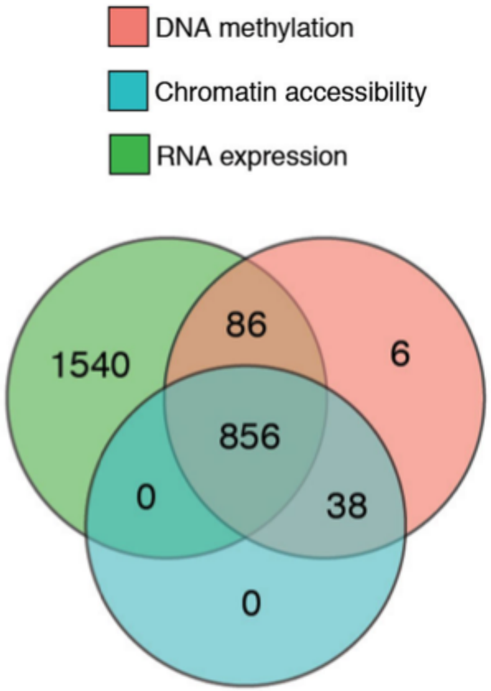
- Simplest approach is to link via proximal associations (in *cis)*
- Promoter capture Hi-C data sets would enable targetedly probe distal associations *(a*ttempted with scNMT-seq data in Luo2020)



Mosaic integration (i.e. transfer learning?)

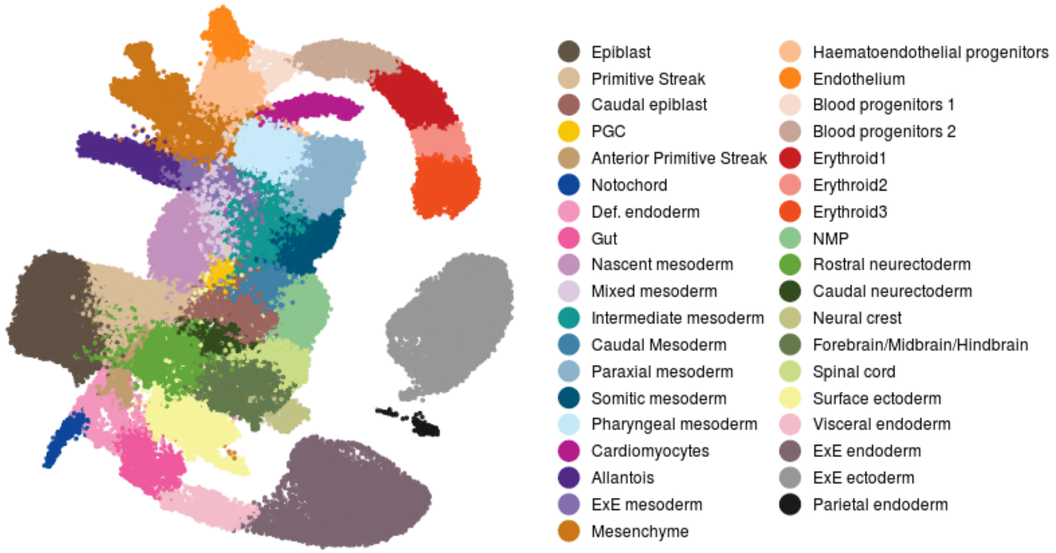


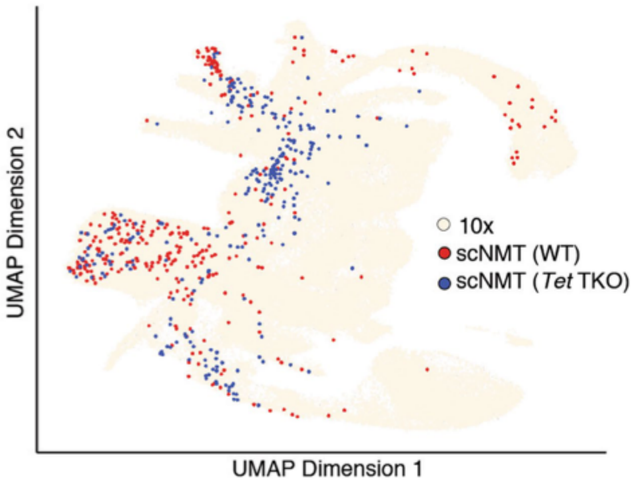
Luo2019



Argelaguet2019

Exploit the RNA as common coordinate framework to map epigenetic profiles onto large-scale scRNA-seq atlas





scRNA-seq atlas (>1e5 cells)

scNMT-seq (<1e3 cells)

Questions/approaches for Mosaic integration

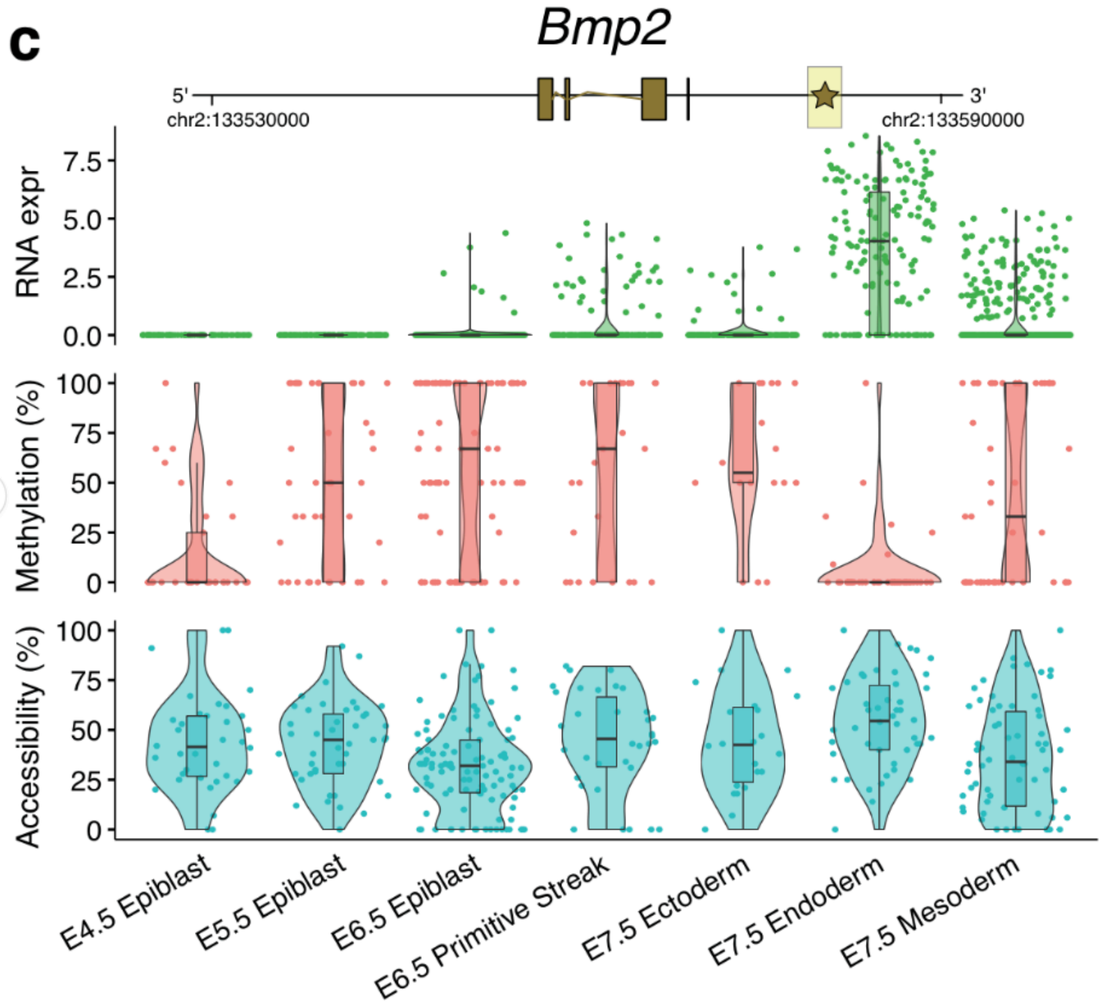
- What is the relative information content and biological content in each data modality? Which omic is better as the “anchor”? mRNA?
- How predictive is RNA from ATAC? and viceversa?
- Existing approaches already exploit a common feature space for data integration: LIGER
- Transfer learning approaches can be adopted here: ProjectR

Non-linear modelling of epigenetic profiles

Two modes for possible non-linear action:

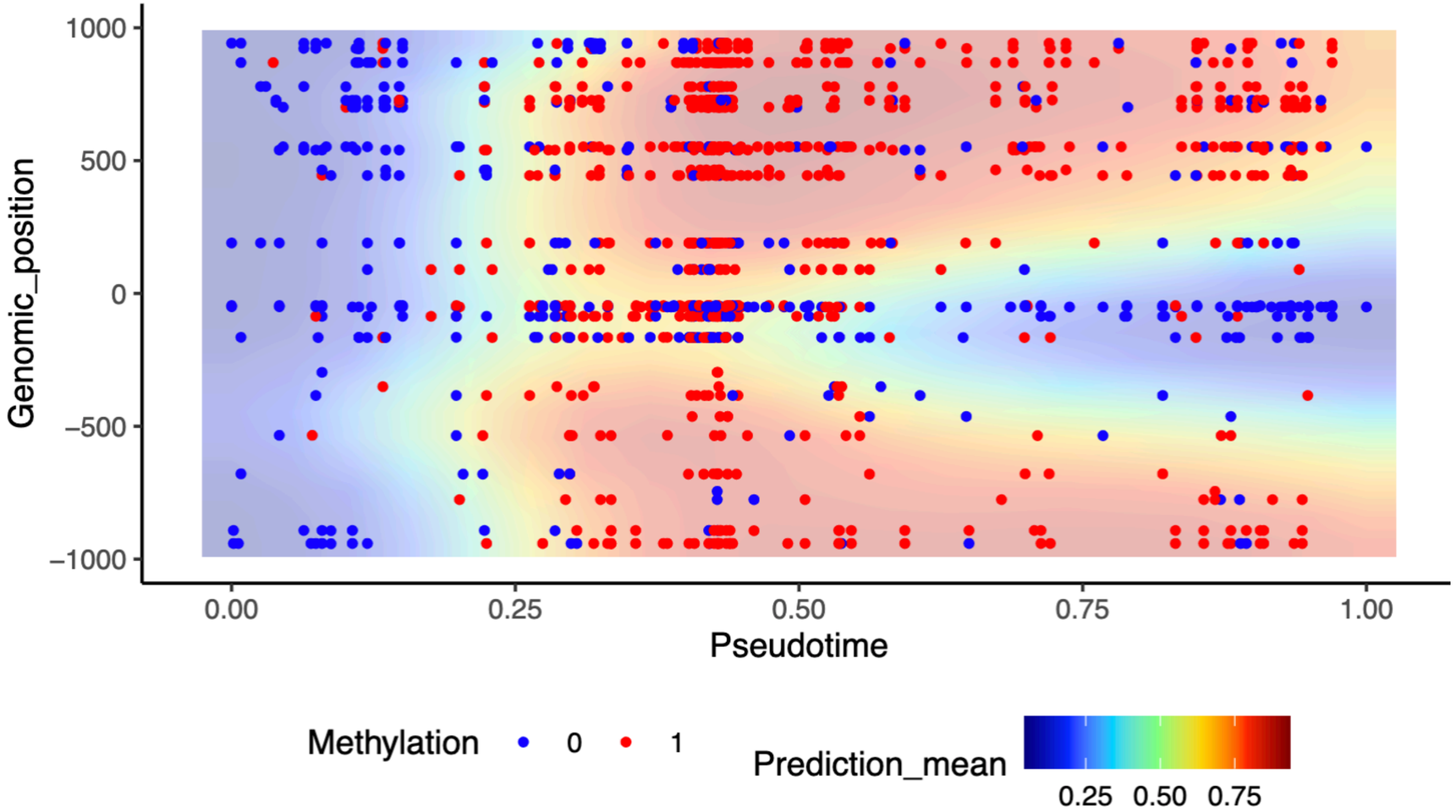
- Local modelling of epigenetic dynamics
- Global modelling for dimensionality reduction

Local non-linear modeling for epigenetic dynamics



Gaussian process classification model

- Covariates: pseudotime + genomic location
- Predictors: CpG methylation



Global non-linear modeling for dimensionality reduction

Article | Published: 30 November 2018

Deep generative modeling for single-cell transcriptomics

Romain Lopez, Jeffrey Regier, Michael B. Cole, Michael I. Jordan & Nir Yosef

Nature Methods **15**, 1053–1058(2018) | Cite this article

22k Accesses | **80** Citations | **129** Altmetric | Metrics

---

A joint model of unpaired data from scRNA-seq and spatial transcriptomics for imputing missing gene expression measurements

---

Romain Lopez [1 *] **Achille Nazaret** [1 2 *] **Maxime Langevin** [1 2 *] **Jules Samaran** [1 3 *] **Jeffrey Regier** [1 *]
Michael I. Jordan [1] **Nir Yosef** [1]

<u>Open questions:</u>

- How to benchmark non-linear models if the ground truth is defined using linear models (i.e. PCA)?
- Deep learning models can be useful with large amounts of data. Fine for mRNA but still tricky for DNA methylation
- Non-linearity may be less important than other modeling choices like normalization

Output

- Table with studies for benchmarking
- Box with open questions
- Figure of taxonomy of methods from Josh’s talk?
- Figure on global versus local integration
- Figure on mosaic integration