THE UNIVERSITY OF BRITISH COLUMBIA

Curriculum Vitae for Faculty Members

Date: April 18, 2024 **Initials: GRS**

1. **SURNAME:** Schiebinger **FIRST NAME :** Geoffrey
   MIDDLE NAME(S): Robert
2. **DEPARTMENT/SCHOOL:** Mathematics
3. **FACULTY:** Science
4. **PRESENT RANK:** Assistant Professor **SINCE:** 2019
5. **POST-SECONDARY EDUCATION**

   (a) *Degrees:*

University or Institution	Degree	Subject Area	Dates
UC Berkeley	Ph.D.	Statistics	2016
Stanford University	M.S.	Electrical Engineering	2011
Stanford University	B.S.	Mathematics	2011

   (b) *Title of Dissertation and Name of Supervisor*
   Title: "Sparse Inverse Problems: The Mathematics of Precision Measurement"
   Supervisor: Benjamin Recht

6. **EMPLOYMENT RECORD**

   (a) *Prior to coming to UBC:*

University, Company or Organization	Rank or Title	Dates
Massachusetts Institute of Technology	Postdoctoral Fellow	2016 - 2019
Broad Institute of MIT and Harvard	Postdoctoral Fellow	2016 - 2019

   (b) *At UBC:*

Rank or Title	Dates
Assistant Professor	2019 - current.

7. **LEAVES OF ABSENCE**

University, Company or Organization at which leave was taken	Type of leave	Dates
UBC	Parental	03/2021 – 05/2021
UBC	Parental	06/2023 – 08/2023

8. **TEACHING**

   (a) *Areas of special interest and accomplishments:* Mathematical foundations of data science, mathematical biology, probability, statistics, optimization, optimal transport, single cell analysis

   (b) *Courses taught at UBC:*

Session	Course number	Scheduled hours	Class size	Hours taught			
				Lectures	Tutorials	Labs	Other

2019-20 WT1	MATH 612D	10	3 hrs/wk	
2019-20 WT2	MATH 318	105	3 hrs/wk	
2020-21 WT1	MATH 612D	16	3 hrs/wk	
2020-21 WT2	MATH 318	100	3 hrs/wk	taught half
2021-22 WT2	MATH 318	100	3 hrs/wk	
2022-23 WT1	BME 371	100	3 hrs/wk	
2022-23 WT1	MATH 612D	17	3 hrs/wk	
2023-24 WT2	MATH 318	100	3 hrs/wk	

(c) *Graduate/undergraduate students supervised and/or co-supervised:*

Student name	**Program type**	**Dates**	**Principal supervisor**	**Co-supervisor(s)**
Zhang, Stephen	M.Sc.	2019-2021	GRS	
Matsumoto, Tim	M.Sc.	2020-current	GRS	
Bonham-Carter, Rebecca	M.Sc.	2020-current	GRS	
Zand, Roomina	M.Sc.	2020-current	GRS	
Greenstreet, Laura	USRA	2020	GRS	
Muglich, Darius	USRA	2020	GRS	
Afanassiev, Anton	Ph.D.	2021-current	GRS	
Gadhiwala, Nitya	M.Sc.	2021-2023		GRS and Omer Angel
Gadhiwala, Nitya	Ph.D.	2023-current		GRS and Omer Angel
Kubal, Sharvaj	M.Sc.	2021-2023		GRS and Yaniv Plan
Kubal, Sharvaj	Ph.D.	2023-current		GRS and Yaniv Plan
Doebeli, Carlos	USRA	2021	GRS	
Cai, Zhenglun	Directed Study	2021	GRS	
Chakraborty, Parajit	M.Sc.	2022-current		GRS and Omer Angel
Zhang, Irena	M.Sc.	2022-current	GRS	
Ma, Yujia	WLIURA	2022	GRS	
Boyle, Cole	M.Sc.	2023-current	GRS	

Student awards

- Stephen Zhang won the Best Poster Prize in the 2020 Society for Mathematical Biology meeting.
- Becca Bonham Carter won the 2022 Dr. Deepak Kaura award for students conducting interdisciplinary research in applied mathematics and medicine.

(d) *Student committees and thesis reading:*

Student name	**Degree**	**Role**	**Department**	**Date**
Johnson, Jeanette	B.S.	Oral exam committee member	Microbiology and Immun.	Dec, 2019
Sullivan, Kaitlin	M.S.	Oral exam committee member	Neuroscience	2020 - 2022
Salehi, Sohrab	Ph.D.	University examiner	Statistics	2021
Pattie Ye	M.S.	graduate supervisory committee	Bioinformatics	2022 - 2024
Brett Kiyota	M.S.	graduate supervisory committee	GSAT	2022 - 2024
Rafi Abdul	Ph.D.	graduate supervisory committee	SBME	2022 - 2023
Andrea Cossa	Ph.D.	graduate supervisory committee	Istituto Europeo di Oncologia	2021 - 2024

(e) *Continuing education activities:*

(f) *Visiting lecturer (indicate university/organization and dates):*

(g) *Course development:*

- In Fall 2019, I created a new graduate course on Single Cell Analysis (Math 612D), and I taught it for a second time in Fall 2020. It has been well attended, with 10 students in 2019 and 16 students in 2020 registered. Regularly over 20 attended lectures. The students came from diverse programs including Mathematics, Statistics, Computer Science, Zoology, Developmental Biology, Biomedical Engineering, and Physiology. The course covers foundational mathematical tools that are useful in analyzing high-dimensional single-cell datasets, and modelling developmental stochastic processes. We cover basic probability theory, statistical inference, convex optimization, Markov stochastic processes, and advanced topics in optimal transport. This course was offered again in Fall 2022 as a PIMS Network Course.
- I also modified the content of BME 371 Transport Phenomena in Cells and Tissues to include material on optimal transport.

(h) *Employees supervised:*

Employee	Type	Dates	Co-supervisor(s)
Greenstreet, Laura	USRA	2020	–
Muglich, Darius	USRA	2020	–
Matsumoto, Tim	WLI	2020	–
Afanassiev, Anton	USRA	2020	–
Lavenant, Hugo	PDF	2019 - 2020	Y.H. Kim, B. Pass
Heitz, Matthieu	PDF	2021 - current	–
Warren, Andrew	PDF	2022 - current	Y.H. Kim
Deb, Nabarun	PDF	2022 - 2023	Y.H. Kim
Ventre, Elias	PDF	2022 - 2024	–
Doebeli, Carlos	USRA	2021	GRS
Yao, Rentian	PDF	2024 - current	–
Zhao, Wenjun	PDF	2024 - 2025	Y.H. Kim and K Dao Duc

RA – research assistant, USRA – Undergraduate Student Research Awards, PDF – post-doctoral fellow, WLI – work learn international

Postdoctoral placement

- Hugo Lavenant is now an Assistant Professor in the Department of Decision Sciences of Bocconi University in Milan, Italy.

- Nabarun Deb is now an Assistant Professor of Econometrics and Statistics in the Chicago Booth School of Business at the University of Chicago.

(i) *Other:*

9. SCHOLARLY AND PROFESSIONAL ACTIVITIES

(a) *Areas of special interest and accomplishments*

The following is an excerpt from the attached research statement:

Biology has entered a new era of precision measurement and massive datasets. Techniques like single-cell RNA sequencing (scRNA-seq) and single-cell ATAC-seq have emerged as powerful tools to profile cell states at unprecedented molecular resolution. One of the most exciting prospects associated with this new trove of data is the possibility of studying temporal processes, such as differentiation and development. If we could understand the genetic forces that control embryonic development, then we would have a better idea of how cell types are stabilized throughout adult life and how they destabilize with age or in diseases like cancer.

This would be within reach if we could analyze the dynamic changes in gene expression, as populations develop and subpopulations differentiate. However, this is not directly possible with current measurement technologies because they are destructive (e.g. cells must be lysed to measure expression profiles).

Therefore, we cannot directly observe the waves of transcriptional patterns that dictate changes in cell type. In response, there has been a flurry of recent work on developing methods to infer trajectories from static snapshots of gene expression profiles. However, there is relatively little theoretical understanding of this statistical inverse problem; if we are to rely on trajectory inference to understand disease, develop new therapies, and engineer tissues, we need to know when to trust the results.

My research group is developing a rigorous statistical framework for understanding the developmental trajectories of cells in a dynamically changing, heterogeneous population based on static snapshots along a time-course. The framework is based on a simple hypothesis: over short time-scales cells can only change their expression profile by small amounts. We formulate this in precise mathematical terms using a classical tool called optimal transport (OT), and we propose that this optimal transport hypothesis is one of the first fundamental mathematical principles of developmental biology. Compared to related fields like evolution and population genetics, developmental biology has been relatively non-mathematical. This OT-hypothesis leads to a rigorous mathematical theory of development, broadly interpreted to include any population of cells changing over time (e.g. tumorigenesis, disease progression, aging, wound healing, cellular reprogramming etc). My 2021 CIHR Project Grant on this topic was **ranked first in Canada**, and for this I was **awarded the 2021 Maude Menten Prize in Genetics**.

Please see the research statement at the end of this document for more.

(b) *Research or equivalent grants (indicate under COMP whether grants were obtained competitively (C) or non-competitively (NC))*:

Granting agency	**Subject**	**COMP**	**$ per year**	**Dates**	**Principal investigator**	**Co-investigator(s)**
UT CLIMB	Development of a cell agent-based virtual human lung	C	$100,000 to GRS	2024 - 2027	N. Yachie	GRS, Zandstra, De Boer, Shakiba
Michael Smith Health Research BC	Towards a Mathematical Theory of Development	C	$80,000	2022 - 2027	GRS	
HOPE Wellcome	T-cell induction and lineage tracing	C	$40,000 USD to GRS	2022 - 2023	P. Zandstra	GRS
GenomeBC PIF	A spatial transcriptomics technology of unprecedented scale	C	$250,000	2022 - 2023	GRS and N. Yachie	
	Maud Menten New PI Prize in Genetics	C	$30000	2021 - 2022	GRS	
CIHR Project Grant	Illuminating the genetic forces driving development by profiling with single cell RNA-seq at thousands of timepoints	C	$170,000	2021 - 2026	GRS	N. Yachie and K. Sugioka
CIHR Project Grant	Cytokine networks controlling myeloid cell mediated immunosuppression in colon cancer	C	$164,000	2021 - 2026	Ken Harder	GRS
NSERC Discovery	Towards a mathematical theory of development	C	$41,000	2020 - 2025	GRS	–

NSERC Early Career	Towards a mathematical theory of development	C	$12,500	2020	GRS	–
NFRF Exploration	Towards a mathematical theory of development	C	$125,000	2020 - 2022	GRS	Y.H. Kim (50%)
BWF Career Award at the Scientific Interface	Analyzing developmental processes with optimal transport	C	$166,581	2018 - 2023	GRS	–
STAIR	Illuminating the genetic forces driving development by profiling with single cell transcriptomics at thousands of time-points	NC	$20,000	2020	GRS	Kenji Sugioka (50%)
Chan Zuckerberg BioHub	Chan Zuckerberg Initiative Investigator grant	C	$157,051	2018-2019	Philippe Rigollet	GRS (50%)
NSF	Graduate Research Fellowship	C	$71,317	2011-2013	GRS	–

(c) *Invited presentations* (Conferences, workshops):

1. Statistical Learning and Data Science / Nonparametric Statistics at Columbia University, New York, United States. May 2018.
2. Chan-Zuckerberg Initiative Investigator Meeting, Santa Cruz, United States. April 2018.
3. UCLA Computational Genomics Winter Institute, Los Angeles, United States. February 2018.
4. How to get from A to B: Transitions in Biology Princeton Center for Theoretical Science, Princeton, United States. December 2017.
5. Beyond Convexity: Emerging Challenges in Data Science, Oaxaca, Mexico. October 2017.
6. OTML Workshop at NeurIPS, Vancouver, 2019.
7. LMRL Workshop at NeurIPS, Online, 2020.
8. Molecular Biology Society of Japan Keynote, Online, 2020
9. Society of Mathematical Biology, June 2021.
10. CMS Summer School on Optimal Transport, June 2021.
11. Integrating Single Cell Analysis and Mathematics, December 2021.
12. Workshop on Connections between interacting particle dynamics and data science, Isle of Skye, May 2022 (rescheduled from May 2021 due to Covid).
13. IFML + Kantorovich Initiative Retreat, February. 2023. Seattle, WA
14. RIKEN BDR Symposium 2023 "Transitions in Biological Systems". Kobe, Japan. March 2023.
15. ICERM, Brown, RI, USA. May 2023
16. FOCM, Paris, France. June 2023. (declined due to parental leave)
17. ICIAM workshop on Challenges in single cell data science: theory and application. August 2023.
18. AI and Cell Fate, Beijing, China. October 2023. (declined due to parental leave)
19. WPI PRIMe International Symposium. Osaka, Japan. February 2024.

(d) *Invited presentations* (seminars, colloquia, lectures):

1. Duke Computational Biology and Bioinformatics. November 2023

2. Osaka University, Japan. August 2023
3. PIMS-NSF Summer School on Optimal Transport. June 2022. Seattle, WA
4. UC Riverside Interdisciplinary Center for Quantitative Modeling in Biology, April 2022.
5. Oxford CSML Seminar, April 2022.
6. UBC IAM Faculty Seminar, September 2020.
7. UBC Life Science Institute Seminar, March 2020 (cancelled due to Covid 19).
8. Yale Applied Math Seminar, March 2020 (cancelled due to Covid 19).
9. UBC Cellular and Physiological Sciences Seminar, Vancouver, Canada. December 2019.
10. UBC MathBio Seminar, Vancouver, Canada. Sept 2019.
11. UBC Statistics Colloquium, Vancouver, Canada. Sept 2019.
12. Duke Statistics Department Colloquium, Durham, United States. Feb 2019.
13. UC Irvine Statistics Seminar, Irvine, United States. Feb 2019.
14. Statistics and Operations Research Seminar, University of North Carolina Chapel Hill, Chapel Hill, United States. Feb 2019.
15. UC Berkeley Biostatistics Seminar, Berkeley, United States. Feb 2019.
16. Department of Bioengineering Seminar, UW Madison, Madison, United States. Feb 2019.
17. Department of Mathematics Seminar, University of British Columbia, Vancouver, Canada. Feb 2019.
18. Stanford Genetics Departmental Colloquium, Stanford, United States. March 2018.
19. Klarman Cell Observatory Scientific Advisory Board Meeting, Cambridge, United States. May 2018.
20. Duke Applied Math and Duke Genome Sciences Joint Colloquium, Durham, United States. January 2018.
21. Harvard Theory Lunch, Harvard, United States. December 2017.
22. Single cell analytics group seminar, MIT, Cambridge, United States. November 2017.
23. Models, Inference, and Algorithms Seminar, MIT, Cambridge, United States. October 2017.
24. Centers of Excellence in Genomic Science 15th Annual Grantee Meeting, Seattle, United States. September 2017.
25. Klarman Cell Observatory Scientific Advisory Board Meeting, Cambridge, United States. May 2017.
26. Laboratory for Information and Decision Systems Seminar, MIT, Cambridge, United States. September 2016.
27. Models, Algorithms, and Inference, Cambridge, United States. February 2016.
28. Stanford Statistics Colloquium, Stanford, United States. October 2015.
29. UW Madison SILO Seminar, Madison, United States. September 2015.
30. Risk Analysis Seminar, Berkeley, United States. April 2014.
31. Berkeley Statistics Annual Research Symposium, Berkeley, United States. March 2014.
32. Computational Algebraic Geometry Seminar, Bonn, Germany. November 2013.

(e) *Contributed presentations* (conferences, workshops):

1. Statistical Challenges in Single Cell Analysis, Ascona, Switzerland. May 2017. (*Prize for "Best Contribution to the Conference"*).

(f) *Conference organization:*

1. Co-organizer, Pacific Interdisciplinary Hub on Optimal Transport Summer School, University of Washington, June 19 — July 1 2022.

(g) *Other:* (e.g. visitors)

10. **SERVICE TO THE UNIVERSITY**

(a) *Memberships on committees, including offices held and dates*
I was on the UPER committee. from 2019 - 2020
I led the UPER sub-committee on Research and Communication.
I wrote the linear algebra qual exams for term 1 and term 2 of Winter 2021-22.
Merit committee, 2023.
Grad admissions committee, 2023
Hiring committee for Michael Smith Labs 2024

(b) *Other service, including dates*
In 2019WT1 and T2, I participated in the cluster hiring initiative with a proposal on **Microbiome Interactions and Synthetic Biology** in collaboration with Stephen Hallam and Lindsay Eltis and others, including Leah Keshet, but our proposal was not selected.

11. SERVICE TO THE COMMUNITY

(a) *Memberships on scholarly societies, including offices held and dates*

(b) *Memberships on other societies, including offices held and dates*

(c) *Memberships on scholarly committees, including offices held and dates*
UPER Committee member, 2019 - 2020.
SBME Awards Committee, 2021 - 2022.

(d) *Memberships on other committees, including offices held and dates*

(e) *Editorships (list journal and dates)*

(f) *Reviewer (journal, agency, etc. including dates)*
Journals:

- Annals of Statistics, 2019,
- Annals of Applied Statistics, 2019,2020.
- PLOS Computational Biology, 2019, 2020.
- Annals of Statistics, 2019.
- Cell Reports, 2019.
- Information and Inference, 2018, 2020.
- Applied and Computational Harmonic Analysis, 2013, 2016.
- FOCM 2020.
- Nature Biotech 2021.
- Bioinformatics, 2022.
- Annals of Applied Probability, 2023.
- Cell Systems, 2024.

Agencies, institutes:

- Agency/inst. name, dates (# of reviews)
- NSERC Discovery Grant, 2020 (1 review).
- NFRF Exploration Grant, 2021 (1 review).

(g) *External examiner (indicate universities and dates)*

(h) *Consultant (indicate organization and dates)*

(i) *Other service to the community*

12. AWARDS AND DISTINCTIONS

(a) *Awards for Teaching (indicate name of award, awarding organizations, date)*

(b) *Awards for Scholarship (indicate name of award, awarding organizations, date)*

1. Michael Smith Health Research BC Award, 2022 ($ 400,000 CAD)

2. Maud Menten New Principal Investigator Prize in Genetics, CIHR, 2021 ($ 30000 CAD). **For ranking first in Canada in the 2021 CIHR Project Grant competition**.
3. Career Award at the Scientific Interface from the Burroughs Wellcome Fund, 2018 ($500000 USD)
4. Invited faculty at the *UCLA Computational Genomics Winter Institute*, 2018.
5. *Best contributed talk* at Statistical Challenges in Single Cell Analysis in Ascona (organized by ETH Zurich), 2017.
6. First place in the Single Molecule Localization Microscopy Challenge. Organized by EPFL. 2016. The third place contestant also used our algorithm.
7. Honorable mention for best student paper award at CAMSAP conference. 2015.
8. NSF Graduate Research Fellowship. 2011 - 2016.
9. VIGRE Berkeley Fellowship.

(c) *Awards for Service (indicate name of award, awarding organizations, date)*

(d) *Other Awards*

13. **OTHER RELEVANT INFORMATION (Maximum One Page)**

THE UNIVERSITY OF BRITISH COLUMBIA

Publications Record

Date: April 18, 2024 **Initials: GRS**

Surname: Schiebinger **First Name:** Geoffrey
Middle Name(s): Robert

ORCID https://orcid.org/0000-0002-8290-7997

1. **REFEREED PUBLICATIONS**

(a) Journals [14 total with 4 as senior author (last named) and 3 as first author]

1. H. Lavenant, S. Zhang, Y.H. Kim and G. Schiebinger
Towards a Mathematical Theory of Trajectory Inference.
Annals of Applied Probability, 34 (1A), 428-500. 2024.

2. L Greenstreet, A Afanassiev, Y Kijima, M Heitz, S Ishiguro, S King, N Yachie, and G Schiebinger
DNA-GPS: A theoretical framework for optics-free spatial genomics and synthesis of current methods
Cell Systems. 14 (10), 844-859. 2023.

3. F Panariello, O Gagliano, C Luni, A Grimaldi, S Angiolillo, W Qin, A Manfredi, P Annunziata, S Slovin, L Vaccaro, S Riccardo, V Bouche, M Dionisi, M Salvi, S Martewicz, M Hu, M Cui, H Stuart, C Laterza, G Baruzzo, G Schiebinger, B Di Camillo, D Cacchiarelli, and N Elvassore
Cellular population dynamics shape the route to human pluripotency.
Nature Communications, 14 (1), 2829. 05/2023.

4. TM Nolan, N Vukašinović, CW Hsu, J Zhang, I Vanhoutte, R Shahan, et al
Brassinosteroid gene regulatory networks at cellular resolution in the Arabidopsis root
Science 379 (6639). 01/2023.

5. R Shahan, CW Hsu, TM. Nolan, BJ. Cole, I W. Taylor, L Greenstreet, S Zhang, A Afanassiev, A H Cornelia Vlot, G Schiebinger, P N. Benfey, and U Ohler
A single cell Arabidopsis root atlas reveals developmental trajectories in wild type and cell identity mutants.
Developmental Cell, 57 (4), 543-560. 2022.

6. G Schiebinger
Reconstructing developmental landscapes and trajectories from single-cell data *Current Opinion in Systems Biology*, 27, 100351. 2021.

7. S. Zhang, A. Afanassiev, L. Greenstreet, T. Matsumoto, and G. Schiebinger
Optimal transport analysis reveals trajectories in steady-state systems.
PLOS Computational Biology, 17 (12) 2021.

8. A. Forrow and G. Schiebinger
LineageOT is a Unified framework for lineage tracing and trajectory inference.
Nature Communications, 12 (1). 2021.

9. AJ Massri, L Greenstreet, A Afanassiev, A Berrio, GA Wray, G Schiebinger, and DR McClay
Developmental Single-cell transcriptomics in the Lytechinus variegatus Sea Urchin Embryo.
Development, 148 (19) 2021.

10. G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, et. al.,
Optimal-transport analysis of single-cell gene expression across time sheds light on reprogramming.
Cell, 176 (4), 928-943. 2019.
700 citations.

11. N. Boyd, G. Schiebinger and B. Recht.
The Alternating Descent Conditional Gradient Method for Sparse Inverse Problems.
SIAM Journal on Optimization, 27 (2), 616-639. 2017.

12. G. Schiebinger, E. Robeva and B. Recht.
Superresolution without Separation. *Information and Inference,* 7 (1), 1-30. 2017.

13. G. Schiebinger, M. J. Wainwright and B. Yu.
The Geometry of Kernelized Spectral Clustering.
Annals of Statistics, 43 (2) 819-846, 2016.

14. A. Guntuboyina, S. Saha and G. Schiebinger. (alphabetical order)
Sharp Inequalities for f-divergences.
IEEE Transactions on Information Theory, 60 (1), 104-121. 2014.

15. L. A. Warren, D. J. Rossi, G. Schiebinger, I. L. Weissman, S. K. Kim and S. R. Quake.
Transcriptional instability is not a universal attribute of aging.
Aging Cell, 6 (6), 775-782. 2007.

(b) Conference Proceedings

1. L Chizat, S Zhang, M Heitz, G Schiebinger
Trajectory inference via mean-field langevin in path space
Advances in Neural Information Processing Systems. 35, 16731-16742. 2022.
2. A. Forrow, J.C. Hutter, M. Nitzan, P. Rigollet, G. Schiebinger, and J. Weed.
Statistical Optimal Transport via Factored Couplings.
AI Stats, 2019.
3. M.E. Shiffman, W. Stephenson, G. Schiebinger, T. Campbell, J. Huggins, A. Regev, and T. Broderick.
Probabilistic reconstruction of cellular differentiation trees from single-cell RNA-seq data.
NeurIPS Bayesian Nonparametrics Workshop, 2017.
4. A short version of **Superresolution without Separation** appeared in CAMSAP 2015. (full version above).

5. A short version of **The Alternating Descent Conditional Gradient Method for Sparse Inverse Problems** appeared in CAMSAP 2015. (full version above).

2. **NON-REFEREED PUBLICATIONS**

3. **BOOKS**

(a) Authored

(b) Edited

(c) Chapter: **Methodologies for Following EMT In Vivo at Single Cell Resolution**. AJ. Massri, G. Schiebinger, A Berrio, L Wang, GA. Wray, DR. McClay

4. **PATENTS**

(a) U.S. Provisional Patent Application No.: 63/322,386, filed March 22, 2022.

Title: A DNA-based global positioning system

Inventors: Geoffrey Schiebinger, Anton Afanassiev, Yusuke Kijima, Laura Greenstreet, Nozomu Yachie, Matthieu Heitz

(b) U.S. Provisional Patent Application No. 62/561,047, filed September 20, 2017.

Title: Methods and Systems for Reconstruction of Developmental Landscapes by Optimal Transport Analysis

Inventors: Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Aviv Regev, Eric S. Lander, Philippe Rigollet

5. **SPECIAL COPYRIGHTS**
6. **ARTISTIC WORKS, PERFORMANCES, DESIGNS**
7. **OTHER WORKS**
8. **WORK SUBMITTED (including publisher and date of submission)**

(a) M Heitz, Y Ma, S Kubal, G Schiebinger
Spatial transcriptomics bring new challenges and opportunities for trajectory inference
Invited to submit to *Annual Review of Cell and Developmental Biology.* Submitted 11/2023.

(b) Charlotte Bunne, Geoffrey Schiebinger, Andreas Krause, Aviv Regev, Marco Cuturi
Optimal transport for single-cell and spatial omics.
Invited to submit to *Nature Methods.* Submitted 10/2023.

(c) N Deb, YH Kim, S Pal, G Schiebinger
Wasserstein mirror gradient flow as the limit of the Sinkhorn algorithm.
Submitted to *Annals of Probability* 7/2023.
arXiv preprint arXiv:2307.16421

(d) B Bonham-Carter, G Schiebinger
Cellular proliferation biases clonal lineage tracing and trajectory inference.
Submitted to *Bioinformatics.* 10/2023. Major Revision 01/2024.
bioRxiv, 2023.07. 20.549801

(e) G. Mordant, T. Matsumoto, S. Zhang, and G. Schiebinger
Manifold learning with sparse regularised optimal transport.
Submitted to *JMLR*, 10/2023.
arXiv preprint arXiv:2307.09816

(f) E Ventre, A Forrow, N Gadhiwala, P Chakraborty, O Angel, G Schiebinger
Trajectory inference for a branching SDE model of cell differentiation.
Submitted to *Annals of Applied Probability*, 09/2023.
arXiv preprint arXiv:2307.07687

(g) YS Michaels, MC Major, B Bonham-Carter, J Zhang, T Heydari, JM Edgar, et al
Time-and lineage-resolved transcriptional profiling uncovers gene expression programs and clonal relationships that underlie human T lineage specification
Submitted to *Science Immunology*, 10/2023.
bioRxiv, 2023.10. 06.561277

(h) YC Cheng, Y Zhang, S Tripathi, H BV, MK Jolly, G Schiebinger, H Levine, ...
Reconstruction of single cell lineage trajectories and identification of diversity in fates during the epithelial-to-mesenchymal transition
Submitted to *Developmental Cell*, 11/2023.
bioRxiv, 2023.09. 19.558325

(i) H Li, J Ezike, A Afanassiev, L Greenstreet, S Zhang, J Whangbo, V L. Butty, E Moiso, G G. Connelly, V Morris, D Wang, G Q. Daley, S Garg, S T. Chou, A Regev, E Lummertz da Rocha, G Schiebinger, and R. G. Rowe
Hematopoiesis at single cell resolution spanning human development and maturation.
Submitted to *Cell*, 10/2023.

(j) R. Wilder Scott; Martin Arostegui; Lesley Ann Hill; Amanda YuanYuan Yang; Stephen Zhang; Alyssa Zhao; Geoff Schiebinger; Tully Michael Underhill
A single cell epigenomic and transcriptomic atlas of murine mesenchymal stromal cells
Submitted to *Cell*. 03/2023.

(k) Vijay Kuchroo, Yu Hou, Martin LaFleur, Linglin Huang, Conner Lambden, Pratiksha Thakore, Kathryn Geiger-Schuller, Ruihan Tang, Jingwen Shi, Rocky Barilla, Ayshwarya Subramanian, Antonia Wallrapp, Hee Sun Choi, Yoon-Chul Kye, Orr Ashenbrg, Geoffrey Schiebinger, John Doench, Aviv Regev, Arlene Sharpe
CRISPR screens reveal neuropeptide signaling orchestrates T helper cell differentiation.
Submitted to *Cell*. 08/2022.

9. **WORK IN PROGRESS (including degree of completion)**

- **The asteroid belt is an entropic phenomenon**, 75%, submitting to *Nature* as single-author paper.
- **Illuminating the genetic forces driving development in C. elegans and mouse by profiling with scRNA-seq at thousands of time-points**, with Kenji Sugioka (UBC Zoology) and Nozomu Yachie (UBC Biomedical Engineering), 30%.

Towards a Mathematical Theory of Development

Biology has entered a new era of precision measurement and massive datasets. Techniques like single-cell RNA sequencing (scRNA-seq) and single-cell ATAC-seq have emerged as powerful tools to profile cell states at unprecedented molecular resolution. One of the most exciting prospects associated with this new trove of data is the possibility of studying temporal processes, such as differentiation and development. If we could understand the genetic forces that control embryonic development, then we would have a better idea of how cell types are stabilized throughout adult life and how they destabilize with age or in diseases like cancer.

This would be within reach if we could analyze the dynamic changes in gene expression, as populations develop and subpopulations differentiate. However, this is not directly possible with current measurement technologies because they are destructive (e.g. cells must be lysed to measure expression profiles). Therefore, we cannot directly observe the waves of transcriptional patterns that dictate changes in cell type. In response, there has been a flurry of recent work on developing methods to infer trajectories from static snapshots of gene expression profiles (e.g. [1], [2], [3], [4]). However, there is relatively little theoretical understanding of this statistical inverse problem; if we are to rely on trajectory inference to understand disease, develop new therapies, and engineer tissues, we need to know when to trust the results.

My research group is developing a rigorous statistical framework for understanding the developmental trajectories of cells in a dynamically changing, heterogeneous population based on static snapshots along a time-course. The framework is based on a simple hypothesis: over short time-scales cells can only change their expression profile by small amounts. We formulate this in precise mathematical terms using a classical tool called *optimal transport (OT)*, and we propose that this ***optimal transport hypothesis*** **is one of the first fundamental mathematical principles of developmental biology.** Compared to related fields like evolution and population genetics, developmental biology has been relatively non-mathematical. This OT-hypothesis leads to a rigorous mathematical theory of development, **broadly interpreted to include any population of cells changing over time (e.g. tumorigenesis, disease progression, aging, wound healing, cellular reprogramming** etc).

Research Accomplishments

I formulated the OT hypothesis during my postdoctoral studies with Lander, Regev and Rigollet at MIT. We were studying stem cell reprogramming with scRNA-seq, and wished to recover developmental trajectories from snapshots of gene expression profiles collected along a time-course of cellular reprogramming [1]. The gene expression vector of a cell is a 20,000-dimensional vector which encodes the number of molecules of RNA in the cell for each gene. Over time, as cells turn genes on or off to accomplish various tasks, cells trace out trajectories through gene expression space. With scRNA-seq, we can take a large population of cells and measure their positions in gene expression space. However, this process kills the cells, so we attempt to infer trajectories from static snapshots collected along a time-course. In **Figure 1**, we know that the entire green population gives rise to the entire red population, and we would like to infer that the left subpopulation of green cells at time t_1 gives rise to the left subpopulations of red cells at time t_2.

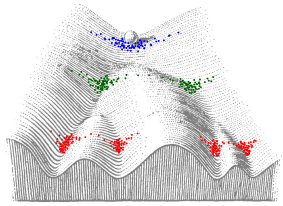

Figure 1: Sampling Waddington's landscape. Cells collected at three distinct time-points are shown in blue, green, and red.

A developing population of cells can be modeled with a continuous time Markov stochastic process over a space of cell states (e.g. gene expression space). We are given samples from the marginals of the stochastic process at various time-points. Crucially, samples from different time-points are independent, so, it is difficult to learn the transition kernel of the Markov process without further assumptions. The

OT hypothesis states that the transition kernel for this Markov process agrees with optimal transport over short time-scales. We have recently shown that this holds for stochastic differential equations with conservative drift (i.e. the drift is the gradient of a potential function as in Fig 1), and that the inverse problem of recovering trajectories can be solved efficiently through convex optimization [5].

We have recently tested the OT hypothesis in diverse systems including induced pluripotent stem cell (iPSC) reprogramming in mice [1], sea urchin embryonic development [6], Arabidopsis root growth [7], [8], [9], human hematopoiesis [10], and reprogramming human cells. Other groups have recently applied OT to study high-plasticity states in lung cancer evolution [11], lineage plasticity in distal lung progenitors [12], and trajectories of aging [13]. See Aim 2b below for a summary of collaborations in progress and future plans.

In each of these collaborations, we found that **OT is predictive and robust**. For example, when we analyzed *L. variegatus* sea urchin development with Gregory Wray and David McClay at Duke [6], we were able to rediscover the vast majority of classically known regulators (e.g. 18 of 21 for endoderm and 13 out of 14 for skeletogenic cells). Similarly, in *Arabidopsis*, which we analyzed with Philip Benfey and Uwe Ohler, we found that OT was able to identify both known developmental regulators and also novel candidates which we verified experimentally [8], [9].

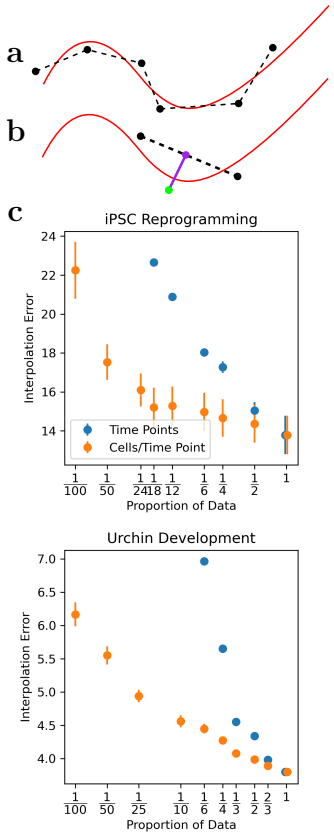

Figure 2: (a) Development traces out a curve in the space of probability distributions. Black dots indicate sampled populations at various time-points, which we connect using optimal transport (dashed lines). (b) Geodesic interpolation compares held out data (green) to mid-point of segment (purple) to quantify performance. (c) Saturation analysis for iPSC reprogramming and urchin embryonic development. Subsampling time-points (blue) causes interpolation to degrade more quickly compared to subsampling cells (orange).

Developmental curves and the optimal design of experiments. As a population of cells changes over time, it traces out a curve in *the space of probability distributions* (red curve in **Fig 2a**). The OT hypothesis can be interpreted geometrically: developmental curves are '*locally geodesic*' with respect to the optimal transport metric.

When we sample cells at a time-point with scRNA-seq, the empirical distribution of cells forms a "noisy measurement" of a point along the curve (black dots in **Fig 2a**). The number of cells sampled determines the "noise-level" of the time-point measurement: the more cells we sequence at a time-point, the more precisely we can localize the curve at that one position. We can then connect consecutive time points with optimal transport (**Fig 2a**, dashed lines), as proposed in [1]. Through *geodesic interpolation* (**Fig 2b**), we can quantify performance by comparing the midpoint of a line-segment (purple point) to held out data (green point).

A **key outcome** of this viewpoint is a **paradigm shift in the design of experiments**: while the number of cells per study has increased dramatically with droplet-based scRNA-seq [14], [15], the number of time points in time-course studies of development has not increased by nearly the same amount. For example, a recent high-profile study on mouse embryonic development profiled an impressive one million cells, but over only 5 developmental time-points [16]. While high-resolution sampling of many time-points is practically very difficult, the theory recommends collecting more time-points with fewer cells per time-point. To illustrate this, we analyzed the data from iPSC reprogramming [1] and sea urchin embryonic development [6]. Similar to analyzing saturation of reads in sequencing, we examined the saturation levels for

both temporal resolution and numbers of cells (**Fig 2c**). We found that geodesic interpolation results degraded more rapidly when time-points were removed, compared to down-sampling cells [17]. Therefore, it seems better to collect more time-points, with fewer cells per time-point. With this motivation, we are currently collaborating with experimentalists to **develop a new technology to collect thousands of time-points with scRNA-seq**. My 2021 CIHR (Canadian Institutes of Health Research) Project Grant on this topic was **ranked first in Canada**, and for this I was awarded the 2021 Maude Menten Prize in Genetics. See Aim 2 for more on this new measurement technology.

Research Plans

AIM 1: *Build mathematical theory of development based on the OT hypothesis*

We have recently analyzed trajectory inference as a statistical inverse problem, and we have proven that a consistent estimator can be formulated through convex optimization [5]. We consider the problem of recovering the law of a stochastic differential equation from samples of the marginals at various time-points. By law, we mean the probability distribution on paths induced by the SDE. While this is not possible in general, we prove that for SDEs with conservative drift (i.e. the drift is the gradient of a potential function), the curve of marginals uniquely determines the trajectories.

The most basic estimator would "connect-the-dots" with optimal transport (Fig 3, red curve). We construct an estimator which generalizes this concept by trading off data fitting with regularization. Roughly speaking, the regularization term measures the total (squared) length of the curve, and the data fitting is measured by likelihood (Fig 3, blue curve). More precisely, the optimization variable is a stochastic process and the regularization term is the relative entropy to a reference process, which we take to be the Wiener process (with diffusion equal to the diffusion of the SDE). This relative entropy minimization is equivalent to entropy-regularized optimal transport between the marginals of the reconstructed process (i.e. we smoothly "connect-the-dots" with Schrödinger bridges to form the blue curve in Fig 3).

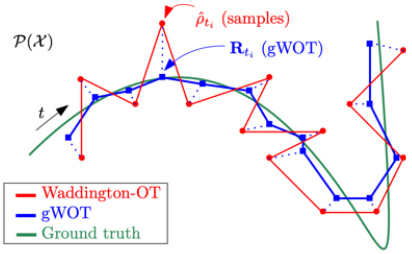

Figure 3: A developmental process (green curve) is sampled at various time-points (red dots). The blue curve shows a regularized estimate.

The resulting optimization problem is convex, yet infinite dimensional. In our first paper on the topic, we proved that the optimal solution yields a consistent estimator: it recovers the true law on paths in the limit of infinitely many time-points [5]. Interestingly, we only require one sample at each time-point for consistency. This has been published in *Annals of Applied Probability* [18]. In a second paper on the topic, we design an algorithm to solve the optimization problem via mean-field Langevin dynamics [19].

There are several interesting open questions in this area. First, we are working to establish finite-sample rates of convergence. This holds the key to answering biological questions like: *What time-resolution do we need to reconstruct a developmental curve? What is the optimal set of time-points to sample? How many cells should we sample at each time-point?* We suspect the rates will depend on the 'curvature' of the developmental progression and one should sample more frequently over periods of rapid change. While estimating Wasserstein distances can suffer from the curse of dimensionality, entropic regularization is known to yield parametric rates [20]. We are also eager to explore notions of sparsity in this context (e.g. a sparse law on paths for the SDE). Along similar lines, we could replace the Wiener process in the entropy minimization with something more biological (e.g. a process parameterized by a gene regulatory network which drives trajectories through gene expression space).

Second, we envision interesting extensions of this framework to regressions of families of developmental curves. For example, we can model the *evolution* of development with a branching

process in the space of developmental curves (branching is a speciation event, each leaf of the tree is the developmental curve of a species living today). After inferring developmental trajectories for various species, we could model the phylogeny and attempt to infer ancestral developmental curves, similar to how ancestral DNA sequences are inferred. Beyond discrete phylogenies, one can also consider continuously parameterized families of curves: for example, consider wound healing or disease progression parameterized by age. This is a one-dimensional family of curves. It would be interesting to share information across ages and simultaneously infer all curves in a single regression framework. Finally, one could also imagine other applications of trajectory inference beyond cellular development. For example, in the design of clinical trials, when longitudinal data is not always available for every patient, one could leverage OT to infer trajectories and therefore obtain pseudo-longitudinal data from static snapshots.

AIM 1b: *Develop unified framework for lineage tracing and trajectory inference*. New measurement technologies (including from collaborator Yachie [21], [22]) make it possible to simultaneously trace cell lineage and measure cell state. However, the computational approaches for trajectory inference and lineage tracing have been approached from separate directions. We have recently made progress towards a unified framework for lineage tracing and trajectory inference [23], [24]. In the first work [24] we have shown that OT trajectory inference can be improved with lineage information (**Fig 4**). The main idea is to leverage a lineage tree to adjust cell states (**Fig 4c**) before connecting them to their putative ancestors with OT (**Fig 4d**). In recent work, with Omer Angel, we have shown that lineage tracing can also be used to incorporate automatic estimates of cellular proliferation [23]. Ultimately, we aim to extend these approaches to share information over time (the curves in Fig 3 would then be curves in the space of trees). The first step is to replace the reference process in the Schrodinger problem with branching Brownian motion, which has recently been accomplished by my former postdoc, Hugo Lavenant [25].

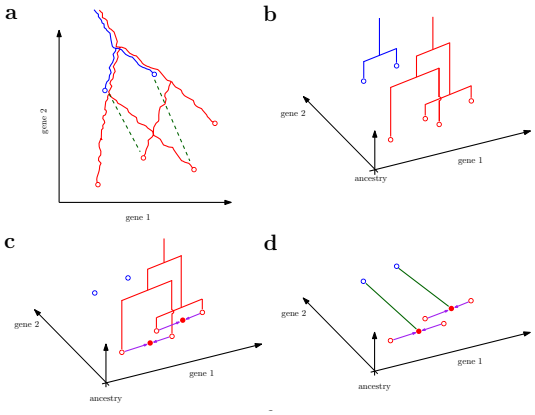

Figure 4: (a) Two developmental processes are stopped at different time-points (red and blue). Green lines show incorrect trajectories, inferred without lineage information. (b) Lineage information, collected simultaneously with cell state, is shown on the vertical axis. (c-d) The two steps of our preliminary method.

AIM 2: *Profile development with scRNA-seq at thousands of time-points and test the OT hypothesis.* Our theory motivates the collection of high-density time-courses (see Fig 3). We propose a simple method for embryo barcoding to easily profile thousands of embryos in a single experiment. The idea is to cross two parents with slightly different genotypes in an organism that can produce large numbers of offspring (like urchin, fly, *C. elegans*, etc). We will dissociate all embryos in parallel, sample cells with scRNA-seq, and detect genotypes from the scRNA-seq data. This will allow us to cluster cells according to their embryo of origin. We will initially use *C. elegans* embryos as a model system. *C. elegans* has an invariant cell lineage and its full lineage has been already described [26]. This aim will be supported by collaborators Nozomu Yachie and Kenji Sugioka, and funded by our CIHR Project Grant. We are also collaborating with Dave McClay and Greg Wray from Duke to test this in urchin and with Steve Plotkin from UBC to test this in ctenophora, the most evolutionary ancient multicellular animal.

AIM 2b: *Test the OT hypothesis in diverse biological settings*: We are collaborating with numerous groups of experimentalists to analyze diverse developmental processes. 1) With the first group (Wray

and McClay at Duke), we will analyze a pair of sea urchin developmental time-courses and compare developmental trajectories across species. We will collect thousands of time-points leveraging Aims 1 and 2. 2) With Peter Zandstra's lab at UBC, we will elucidate the events along a time-course of T-cell induction. 3) With Ryan Flannigan's lab at Vancouver General Hospital, we will compare the developmental progression of spermatogenesis in healthy vs diseased (collapsed) patients. 4) With Sam Aparicio's lab (BC Cancer), we will analyze a time-course of tumor progression in mice. 5) With Fadi Lakkis and Khodor Abou-Daya (U Pitt), we will leverage a combined time-course of single cell RNA-seq and ATAC-seq to shed light on monocyte differentiation. 6) With Philip Benfey's lab at Duke, we are analyzing several mutant knockout atlases using StationaryOT, a variant of our OT framework designed for systems in equilibrium. 7) With Nozomu Yachie's Lab at UBC, we aim to collect and analyze a lineage-tracing time-course in mice (in addition to the *C. elegans* experiments described above). 8) Pamela Hoodless's lab at BC Cancer is planning to perform time-course single-cell RNA sequencing measurements of hepatic organoid development. I will support the prediction of cell differentiation trajectories using WaddingtonOT. 9) In collaboration with Ken Harder, we will analyze cytokine networks controlling myeloid cell mediated immunosuppression in colon cancer. Schiebinger is co-PI on Harder's CIHR Project Grant, awarded June 2021. In the first six collaborations, data has already been provided.

AIM 3: *Develop theory and methods for spatiotemporal trajectory inference*

With collaborator Nozomu Yachie (UBC SBME), we are developing a new measurement technology for large-scale spatial transcriptomics (ST), which can measure gene expression across whole organs. (Ordinary scRNA-seq loses the spatial context of cells in tissues). Existing ST technologies have struggled to capture large field-of-view and have mostly been restricted to capturing two-dimensional images from tissue slices. Our technology, called DNA-GPS [27] leverages concepts from manifold learning to dramatically reduce the difficulty of the measurement process. Our key idea is to randomly distribute DNA barcodes throughout the tissue sample. These stick to cells and are captured and sequenced together with the rest of the genes. Each cell is then equipped with two high-dimensional vectors: the ordinary gene expression vector describing the cell state, and the artificial DNA barcode vector describing the number of copies of each DNA barcode species captured by the cell. Cells close in physical space will capture similar counts of DNA barcodes; therefore, this process embeds the physical tissue as a low-dimensional manifold in a high-dimensional DNA-barcode space. We have demonstrated through simulations that a manifold learning algorithm called UMAP can accurately recover cellular positions (**Fig 5**), and we have used the simulations to optimize the experimental design (e.g., *how many DNA barcodes and what depth of sequencing do we need and how should DNA barcodes be distributed over space*?).

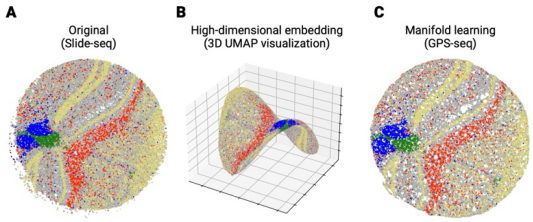

Figure 5: (A) Original slide-seq spatial transcriptomic image of mouse cortex. Colors represent cell types. (B) Tissue slice embedded in high-dimensional space of DNA barcodes. (C) DNA-GPS reconstruction.

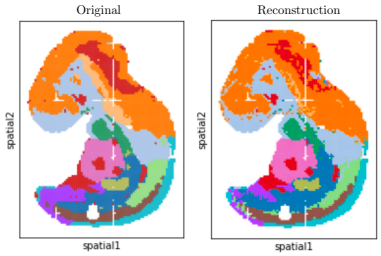

Figure 6: Original image (left) and reconstruction (right) colored by cell type.

AIM 3b: *ST denoising and trajectory inference*. With colleagues Yaniv Plan and Michael Friedlander, we are developing a compressed-sensing approach for denoising ST images. Large-scale ST will require unprecedented sequencing depth. We tested whether we could down-sample the sequencing reads from published ST data [28] and obtain similar results. We found that low-rank regularization on the gene expression matrix, together with total-variation regularization of the ST images allows us to down-sample sequencing reads to 10% of the original depth and recover similar

images (Fig 6). Finally, we envision that we can perform trajectory inference on time-courses of ST images by incorporating spatial position into the cost function of optimal transport. For example, by generalizing our regression approach illustrated in Fig 3, we envision that we could infer the developmental curve of 3D ST from a large number of 2D images from randomly oriented slices.

References

[1] G. Schiebinger *et al.*, "Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming.," *Cell*, vol. 176, no. 4, pp. 928-943.e22, Feb. 2019, doi: 10.1016/j.cell.2019.01.006.

[2] J. A. Farrell, Y. Wang, S. J. Riesenfeld, K. Shekhar, A. Regev, and A. F. Schier, "Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis," *Science*, vol. 360, no. 6392, p. eaar3131, Jun. 2018, doi: 10.1126/science.aar3131.

[3] F. A. Wolf *et al.*, "PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells," *Genome Biology*, vol. 20, no. 1, p. 59, 2019, doi: 10.1186/s13059-019-1663-x.

[4] C. Trapnell *et al.*, "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells," *Nature Biotechnology*, vol. 32, no. 4, pp. 381–386, 2014, doi: 10.1038/nbt.2859.

[5] H. Lavenant, S. Zhang, Y.-H. Kim, and G. Schiebinger, "Towards a mathematical theory of trajectory inference." 2021.

[6] A. J. Massri *et al.*, "Developmental Single-cell transcriptomics in the Lytechinus variegatus Sea Urchin Embryo," *Development*, 2021, doi: 10.1101/2020.11.12.380675.

[7] S. Zhang, A. Afanassiev, L. Greenstreet, T. Matsumoto, and G. Schiebinger, "Optimal transport analysis reveals trajectories in steady-state systems," *PLoS Comput Biol*, 2021, doi: 10.1101/2021.03.02.433630.

[8] T. M. Nolan *et al.*, "Brassinosteroid gene regulatory networks at cellular resolution in the Arabidopsis root," *Science (1979)*, vol. 379, no. 6639, Mar. 2023, doi: 10.1126/science.adf4721.

[9] A. R. Shahan *et al.*, "A single cell Arabidopsis root atlas reveals developmental trajectories in wild type and cell identity mutants," *Dev Cell*, 2022.

[10] H. Li *et al.*, "Hematopoiesis at single cell resolution spanning human development and maturation," *Bioarxiv*, pp. 1–17, 2021.

[11] N. D. Marjanovic *et al.*, "Emergence of a High-Plasticity Cell State during Lung Cancer Evolution," *Cancer Cell*, vol. 38, no. 2, pp. 229-246.e13, 2020, doi: 10.1016/j.ccell.2020.06.012.

[12] K. Hurley *et al.*, "Reconstructed Single-Cell Fate Trajectories Define Lineage Plasticity Windows during Differentiation of Human PSC-Derived Distal Lung Progenitors," *Cell Stem Cell*, vol. 26, no. 4, pp. 593-608.e8, 2020, doi: 10.1016/j.stem.2019.12.009.

[13] J. C. Kimmel, L. Penland, N. D. Rubinstein, D. G. Hendrickson, D. R. Kelley, and A. Z. Rosenthal, "Murine single-cell RNA-seq reveals cell-identity- And tissue-specific trajectories of aging," *Genome Research*, vol. 29, no. 12, pp. 2088–2103, 2019, doi: 10.1101/gr.253880.119.

[14] A. M. Klein *et al.*, "Droplet barcoding for single-cell transcriptomics applied to embryonic stem cells.," *Cell*, vol. 161, no. 5, pp. 1187–1201, May 2015, doi: 10.1016/j.cell.2015.04.044.

[15] E. Z. Macosko *et al.*, "Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets," *Cell*, vol. 161, no. 5, pp. 1202–1214, May 2015, doi: 10.1016/j.cell.2015.05.002.
[16] J. Cao *et al.*, "The single-cell transcriptional landscape of mammalian organogenesis," *Nature*, vol. 566, no. 7745, pp. 496–502, 2019, doi: 10.1038/s41586-019-0969-x.
[17] G. Schiebinger, "Reconstructing developmental landscapes and trajectories from single-cell data," *Current Opinion in Systems Biology*, vol. 27, p. 100351, 2021, doi: 10.1016/j.coisb.2021.06.002.
[18] H. Lavenant, S. Zhang, Y.-H. Kim, and G. Schiebinger, "Toward a mathematical theory of trajectory inference," *The Annals of Applied Probability*, vol. 34, no. 1A, pp. 428–500, Feb. 2024, doi: 10.1214/23-AAP1969.
[19] L. Chizat, S. Zhang, M. Heitz, and G. Schiebinger, "Trajectory Inference via Mean-field Langevin in Path Space," 2022.
[20] A. Genevay, L. Chizat, F. Bach, M. Cuturi, and G. Peyré, "Sample complexity of sinkhorn divergences," in *AISTATS 2019 - 22nd International Conference on Artificial Intelligence and Statistics*, 2020.
[21] R. C. Sakata *et al.*, "Base editors for simultaneous introduction of C-to-T and A-to-G mutations," *Nature Biotechnology*, vol. 38, no. 7, pp. 865–869, 2020, doi: 10.1038/s41587-020-0509-0.
[22] N. Masuyama, H. Mori, and N. Yachie, "DNA barcodes evolve for high-resolution cell lineage tracing," *Current Opinion in Chemical Biology*, vol. 52, pp. 63–71, 2019, doi: https://doi.org/10.1016/j.cbpa.2019.05.014.
[23] E. Ventre, A. Forrow, N. Gadhiwala, P. Chakraborty, O. Angel, and G. Schiebinger, "Trajectory inference for a branching SDE model of cell differentiation," Jul. 2023, [Online]. Available: http://arxiv.org/abs/2307.07687
[24] A. Forrow and G. Schiebinger, "LineageOT is a Unified Framework for Lineage Tracing and Trajectory Inference," *Nat Commun*, 2021, doi: 10.1101/2020.07.31.231621.
[25] A. Baradat and H. Lavenant, "Regularized unbalanced optimal transport as entropy minimization with respect to branching Brownian motion," 2021.
[26] J. E. Sulston, E. Schierenberg, J. G. White, and J. N. Thomson, "The embryonic cell lineage of the nematode Caenorhabditis elegans," *Developmental Biology*, vol. 100, no. 1, pp. 64–119, 1983, doi: https://doi.org/10.1016/0012-1606(83)90201-4.
[27] L. Greenstreet *et al.*, "The DNA-based global positioning system—a theoretical framework for large-scale spatial genomics," *bioRxiv*, p. 2022.03.22.485380, 2022.
[28] A. Chen *et al.*, "Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays," *Cell*, vol. 185, no. 10, pp. 1777-1792.e21, 2022, doi: 10.1016/j.cell.2022.04.003.