Jake Yeung

1 DNA Way, South San Francisco, CA 94080, United States
✉ *jakeyeung@gmail.com*
https://jakeyeung.github.io/

Research Plan

I am a computational biologist working at the interface of statistical/machine learning, functional genomics technologies, and genome biology. I develop computationally-driven experimental designs and biophysically interpretable analysis methods to understand how cells coordinate different genome regulatory layers in healthy states, and how they respond to curative therapy in diseased states.

Professional Preparation

2024 – Present   **Principal Scientist, Group Leader**,
Genentech, Genentech Computational Sciences (gCS),
Analytical Methods.

Sep 2021–2023   **Independent Staff Scientist**,
Institute of Science and Technology (IST) Austria,
Computational biology and machine learning lead,
Two parental leaves 2022-01 and 2023-12.

Mar-Sep 2021   **Machine Learning and Artificial Intelligence Team Leader**,
Wellcome Sanger Institute, University of Cambridge,
Single-cell and spatial genomics.

2020   **Human Frontier Science Program Fellow**,
Hubrecht Institute for Developmental Biology and Stem Cell Research,
Group of Prof. Alexander van Oudenaarden, Single-cell genomics.

January 2019   **PhD in Biotechnology and Bioengineering**,
École Polytechnique Fédérale de Lausanne (EPFL). School of Biotechnology and Bioengineering,
Group of Prof. Felix Naef, Computational Systems Biology.
Committee members: Prof. John Hogenesch, Prof. Erik van Nimwegen, Prof. Joachim Lingner
EPFL Outstanding PhD Thesis Distinction (top 8%)

April 2014   **MSc in Genome Science and Technology**,
University of British Columbia.

June 2012   **BSc in Chemical Engineering**,
University of Waterloo,
Distinction and Dean's Honours.

Manuscripts (* denotes co-first author, bold denotes a top-5 publication)

2023   **J. Yeung*, M. Florescu*, P. Zeller*, B. A. d. Barbanson, Max D. Wellenstein, A. v. Oudenaarden**
scChIX-seq infers dynamic relationships between histone modifications in single cells,
Nature Biotechnology, 10.1038/s41587-022-01560-3.
"Behind the paper" post: https://tinyurl.com/y35yusxp

2022   **P. Zeller*, J. Yeung*, H. V. Gaza, B. A. d. Barbanson, V. Bhardwaj, M. Florescu, A. v. Oudenaarden**
Single-cell sortChIC identifies hierarchical chromatin dynamics during hematopoiesis,
Nature Genetics, https://doi.org/10.1038/s41588-022-01260-3.
Research briefing: https://www.nature.com/articles/s41588-022-01259-w

2021   J. Mermet*, <u>J. Yeung</u>*, F. Naef , *Oscillating and stable genome topologies underlie hepatic physiological rhythms during the circadian cycle*, *PLOS Genetics*, 17(2), e1009350 .

2021   N. Phillips, A. Hugues, <u>J. Yeung</u>, E. Durandau, D. Nicolas, F. Naef , *The circadian oscillator analysed at the single-transcript level*, *Molecular Systems Biology*, 17(3), e10135.

2019   **C. Hor*, <u>J. Yeung</u>*, M. Jan, Y. Emmenegger, J. Hubbard, I. Xenarios, F. Naef, P. Franken**, **Sleep–wake-driven and circadian contributions to daily rhythms in gene expression and chromatin accessibility in the murine cortex**, *Proceedings of the National Academy of Sciences of the United States of America*, 116(51):25773–25783, 2018.

2019   B. Weger, C. Gobet, <u>J. Yeung</u>, E. Martin, S. Jimenez, A. Balvay, A. Foussier, A. Charpagne, C. Jason Chou, F. Naef, F. Gachon, The gut microbiome is required for sex-specific diurnal rhythms of gene expression and metabolism, *Cell Metabolism*, 29(2):362–382.e8, 2019.

2018   F. Gachon, <u>J. Yeung</u>, F. Naef, Cross-regulatory circuits linking inflammation, high-fat diet, and the circadian clock, *Genes & Development*, 32:347–358, 2018.

2018   <u>J. Yeung</u>, F. Naef, Rhythms of the Genome: Circadian Dynamics from Chromatin Topology, Tissue-Specific Gene Expression, to Behavior, *Trends in Genetics*, 34(12):915–926, 2018.

2018   **J. Mermet*, <u>J. Yeung</u>*, C. Hurni, D. Mauvoisin, K. Gustafson, C. Jouffe, D. Nicolas, Y. Emmenegger, C. Gobet, P. Franken, F. Gachon, F. Naef**, **Clock-dependent chromatin topology modulates circadian transcription and behavior**, *Genes & Development*, 32:347–358, 2018, highlighted by a companion outlook article.

2018   J. Wang*, L. Symul*, <u>J. Yeung</u>, C. Gobet, J. Sobel, S. Luck, P. Westermark, N. Molina, F. Naef, Circadian clock –dependent and –independent post-transcriptional regulation underlies temporal mRNA accumulation in mouse liver, *Proceedings of the National Academy of Sciences of the United States of America*, 115(8):E1916–E1925, 2018.

2018   **<u>J. Yeung</u>*, J. Mermet*, C. Jouffe, J. Marquis, A. Charpagne, F. Gachon, F. Naef**, **Transcription factor activity rhythms and tissue-specific chromatin interactions explain circadian gene expression across organs**, *Genome Research*, 28(2):182–191, 2018, featured on the cover.

2017   R. Shrestha, E. Hodzic, T. Sauerwald, P. Dao, K. Wang, <u>J. Yeung</u>, S. Anderson, F. Vandin, G. Haffari, C. Collins, S. C. Sahinalp, HIT'nDRIVE: Patient-specific multi-driver gene prioritization for precision oncology, *Genome Research*, 27(9):1573–1588, 2017.

2016   J. Mermet, <u>J. Yeung</u>, F. Naef, Systems Chronobiology: Global Analysis of Gene Regulation in a 24-Hour Periodic World, *Cold Spring Harbor Perspectives in Biology*, 9(3) a028720, 2016.

2014   A. W. Wyatt, F. Mo, K. Wang, B. McConeghy, S. Brahmbhatt, L. Jong, D. M. Mitchell, R. L. Johnston, A. Haegert, E. Li, J. Liew, <u>J. Yeung</u>, ..., and C. C. Collins, The inter-tumor transcriptome heterogeneity of high-risk primary prostate cancer, *Genome Biology*, 15(8):426.

2014   R. Shrestha*, E. Hodzic*, <u>J. Yeung</u>*, K. Wang, T. Sauerwald, P. Dao, S. Anderson, H. Beltran, M. Rubin, C. Collins, G. Haffari, S. C. Sahinalp, HIT'nDRIVE: Multi-Driver Gene Prioritization based on Hitting Time, *Research in Computational Molecular Biology*, 293–306, Cham, 2014.

Talks at Conferences, Workshops, and Discussion Groups

2022   Learning relationships between histone modifications in single cells. NeurIPS 2022 Workshop LMRL. Virtual, accepted paper

2022   Inferring relationships between histone modifications in single cells. Single-cell Omics Germany Epigenome 2022. Überherrn Germany

2022   Models to infer gene regulation from single-cell chromatin modification data: BIRS Mathematical models in biology. Banff Canada

2021	Uncovering multi-modal chromatin states in single cells. EMBO: Multiomics to Mechanisms: Challenges in Data Integration (virtual)
2021	Deconvolving multiplexed histone modifications in single cells. CSHL: Biology of Genomes (virtual)
2021	Statistical challenges in single-cell epigenomics. Hubrecht single-cell workshop (virtual)
2020	Using machine learning to make maps of stem cell differentiation. Machine learning in Physics, Chemistry and Materials discussion group (MLDG). Cambridge UK
2019	Profiling histone modifications in single cells reveals the dynamics and control of blood cell differentiation. EMBO Workshop: Single Cell Biology. Tokyo Japan
2018	Clock-dependent chromatin topology modulates circadian transcription and behavior. EMBL: From Functional Genomics to Systems Biology. Heidelberg Germany
2018	Dynamic gene expression and regulatory outputs of the mouse cortex. Society for Sleep medicine, Sleep Research and Circadian rhythms. Basel Switzerland
2018	Clock-dependent chromatin topology modulates circadian transcription and behavior. CSHL: Biology of Genomes. Cold Spring Harbor USA
2018	Transcription factor activity rhythms and tissue-specific chromatin interactions orchestrate circadian gene expression across organs. Genomic Regulation 2018. Haute-Nendaz, Switzerland
2017	Interaction of tissue-specific and circadian transcriptional regulatory networks. European Biological Rhythms Society. Amsterdam, Netherlands
2017	Tissue-specific circadian transcriptional regulation. Gordon Conference Seminars on Chronobiology. Stowe, USA
2016	Integrated analysis reveals distinct roles of promoters underlying rhythmic gene expression. Gene Regulation and Promoter Meeting: Eukaryotic Promoter Database 30th Anniversary Symposium. Lausanne, Switzerland
2016	Integrating functional genomics data reveals tissue-dependent mechanisms underlying circadian gene expression. Society for Research on Biological Rhythms. Tampa Bay, USA

Grants and Awards

2022	Invitation to the 72nd Lindau Nobel Laureate Meeting 2023
2019	**Human Frontier Science Program (HFSP) Fellowship** (3-years)
2018	Swiss National Science Foundation Postdoctoral Fellowship (1.5-years)
2018	EPFL Outstanding PhD Thesis Distinction (top 8%)
2018	European Molecular Biology Organization (EMBO) Long-Term Fellowship Finalist (withdrawn after accepting HFSP fellowship)
2015	Natural Resources and Engineering Research Council of Canada Doctoral Fellowship (3 years)
2012	Genome Science and Technology Graduate Award (2 years)
2012	First Prize Walmart Green Student Challenge Competition ($60000 award)

Teaching and Mentoring Experience

2021 - Present	**Co-supervising PhD students**, *IST Austria*, Jessica Kirchner, Puzhen Xia.
2021	**Supervising postdoctoral fellows**, *Wellcome Sanger Institute*, Fabiola Curion.
2019	**Episystem Workshop: Single-Cell Epigenetics**, *Hubrecht Institute*, Developed and led practical sessions to teach analysis of single-cell epigenetics data. Link to course material: https://github.com/avolab/episystem-workshop

2014-2018	**Computational and Mathematical Modeling in Biology**, *EPFL*, Developed and supervised exercises for the third-year undergraduate bioengineering course.
2018	**Supervising Master's Projects in Computational Biology**, *EPFL*, Moritz Durtschi: Gaussian process latent variable models for inferring gene expression pseudotime. Alice Hugues: Stochastic trajectory and noise characteristics in gene expression.

Genomics and Machine Learning Software and Packages

Author
scChIX (https://github.com/jakeyeung/scChIX). Statistical inference method for mapping multiple histone modifications in single cells.

sortChIC (https://github.com/jakeyeung/sortchicAnalysis). Statistical analyses of sortChIC data.

CircadianRNASeq (https://github.com/jakeyeung/TissueCiradianAnalysis). Complex-valued matrix decomposition method for extracting periodic signal from temporal gene expression data.

SleepDepAnalysis (https://github.com/jakeyeung/SleepDepAnalysis). Statistical model for integrating RNA-seq with electroencephalography (EEG) time-series data.

Circadian4Cseq (https://github.com/jakeyeung/Circadian4Cseq). Analysis of time-resolved chromatin conformation data.

Contributor
gpjax (https://github.com/JaxGaussianProcesses/GPJax). JAX Gaussian Processes.

glmpca (https://cran.r-project.org/web/packages/glmpca/index.html). Generalized linear model extensions to principal component analysis.

SingleCellMultiOmics (https://pypi.org/project/singlecellmultiomics). General framework for processing single-cell genomics and epigenomics data.