Manu Saraswat
Email: manu.saraswat10@gmail.com
Heidelberg, DE and New York, USA

Interpretable and Mechanistic AI for Biological Discovery

Computational biologist building interpretable deep learning that links **genetic variation + multi-omics** to **cellular behavior and disease mechanisms**. Mechanistically grounded representation learning— models that encode biological structure, so predictions are explainable and experimentally testable.

7+ years across **MSKCC, EMBL, DKFZ, UBC and Genentech**, working end-to-end from data → model → evaluation → biological insight → experimental validation/translation and collaborating closely with **wet-lab experimentalists and clinicians,** across **cancer plasticity, neurodevelopment, immunology, and regulatory genomics**

Education

10/20 — 07/25 **PhD in Computational Biology**
European Molecular Biology Laboratory (EMBL) and German Cancer Research Center (DKFZ), Heidelberg, Germany

Supervisor: Oliver Stegle, Co-supervisor: Moritz Mall
Thesis: “Decoding gene regulation from single cells to populations with interpretable deep learning”

08/14 — 08/19 **M.Sc.(hons), Mathematics and B.E. (hons), Computer Science.**
Birla Institute of Technology and Science (BITS), Pilani, India
Master’s Thesis: “Predicting transcription initiation at microsatellites using sequence-based deep neural networks”

RESEARCH AND PROFESSIONAL EXPERIENCE

07/25 — present **Postdoctoral Researcher**
Memorial Sloan Kettering Cancer Centre (MSKCC) New York & DKFZ, Heidelberg

Supervisor: Dana Pe'er (MSKCC) and Oliver Stegle (DKFZ)
- Designing interpretable deep learning frameworks to integrate spatial, temporal, and perturbation data at scale to model tumor-immune interactions.
- Collaborating closely with experimental and computational teams to iterate on hypotheses and validation plans

07/23 — 10/23 **Intern | Machine Learning for Regulatory Genomics**
Genentech, Inc. (South San Francisco, CA, USA)

Supervisor: Gokcen Eraslan
- Explored transformer-based approaches to integrate DNA, RNA, and chromatin data from single-cell multi-ome atlases.
- Investigated transfer learning from genomic and single-cell foundation models.

10/19 — 10/20 **Research Software Developer | Full time**
University of British Columbia (UBC), (Vancouver, Canada)
Supervisor: Wyeth Wasserman

- Built interpretable sequence to function deep learning models to predict transcription factor binding and chromatin accessibility.
- Resulted in two co-first author publications in Genome Biology.

10/18 — 07/19 **Master’s Thesis | Computational Regulatory Genomics**
Institut de Génétique Moléculaire de Montpellier (IGMM), (Montpellier, France)
Supervisor: Charles Henri-Lecellier

- Co-first authored a Nature Communications paper on deep learningbased modelling of sequence-driven transcription initiation at microsatellites.

05/18 — 07/18 **Summer Analyst**
Goldman Sachs, Bengaluru, India

- Communicated technical concepts to non-technical stakeholders in a fast-paced environment.

KEY PROJECTS

scDoRI: Single-cell Deep multi-Omics Regulatory Inference
[manuscript under revision in *Nature* | code: github.com/bioFAM/scDoRI]
Interpretable deep learning framework for inferring context-specific gene regulatory networks from multi-omic single-cell data.

- Designed a mechanistically constrained autoencoder incorporating signatures of transcription factor mediated gene regulation in decoding objectives.
- Applied to glioblastoma; identified **MYT1L** as a key plasticity regulator; **validated** in collaboration with wet-lab teams (tumor burden reduction and extended survival in mouse models.

*Demonstrates how mechanistic interpretability enables both **discovery** and **therapeutic translation.***

DeepGenoXcan: Deep Learning for Genetic Variant Effects
[manuscript in preparation | trained on 1000+ genomes]
Donor-level DNA sequence to function prediction, mapping genetic variation to gene expression with interpretability constraints.

- Outperforms linear baselines and advanced sequence models (e.g., Enformer, Borzoi) for predicting gene expression from personalized genomes. expanding disease-gene discovery to previously inaccessible genes.

*Shows how **biologically motivated architectures** can outperform large scale foundation models while maintaining scalability and interpretability.*

ExplaiNN: *interpretable and transparent neural networks for genomics.*
[published, *Genome Biology*]
One of the first fully transparent deep learning architectures for genomics producing motif-level explanations of sequence-to-function predictions.

- Benchmarked across Transcription Factor (TF) binding, chromatin accessibility, enhancer activity, and de novo motif discovery tasks.

Full interpretability without sacrificing performance.

Biologically Relevant Transfer Learning for transcription factor (TF) binding prediction
[published, *Genome Biology*]
Systematic evaluation of how biological priors (TF families, DNA-binding domains, cofactors) affect deep learning training strategies.

- Established benchmarks showing biologically informed pre-training outperforms naive approaches, especially in **low-data regimes**.

Biological domain expertise can directly improve ML model training strategies.

SELECTED PUBLICATIONS

CO-FIRST CONTRIBUTIONS

2025, ***bioRxiv*** ***Decoding Plasticity Regulators and Transition Trajectories in Glioblastoma with Single-Cell multi-omics***
Manu Saraswat*, Laura Rueda-Gensini*, Elisa Heinzelmann*,

Tannia Gracia*, Fani Memi* et. al.

2023, ***Genome Biology*** ***ExplaiNN: interpretable and transparent neural networks for genomics.***
Gherman Novakovsky*, Oriol Fornes*, **Manu Saraswat***, Sara Mostafavi, WW Wasserman.

2021, ***Genome Biology*** ***Biologically relevant transfer learning improves transcription factor binding prediction.***
Gherman Novakovsky*, **Manu Saraswat***, Oriol Fornes*, Sara Mostafavi, WW Wasserman.

2021, ***Nature Communications*** ***Discovery of widespread transcription initiation at microsatellites predictable by sequence-based deep neural network.***
Mathys Grapotte*, **Manu Saraswat***, Chloé Bessière*, FANTOM Consortium, Laurent Brehelin, Charles-Henri Lecellier.

OTHER CONTRIBUTIONS

2025, ***bioRxiv*** ***High-throughput single-cell CRISPRi screens stratify neurodevelopmental functions of schizophrenia-associated genes***
. Umut Yildiz, Annique Claringbould, Mikael Marttinen, Victor Campos-Fornes, Mantha Lamprousi, **Manu Saraswat**, Mathias Saver, Daria Bunina, Michael Dorrity, Judith B Zaugg, Kyung-Min Noh.

* Equal contribution

TECHNICAL SKILLS & TOOLS

ML/AI	VAEs, CNNs, Transformers, Representation Learning, Interpretable ML, Causal Inference, Factor models, Model evaluation and benchmarking.
Scale	HPC/GPU pipelines across thousands of genomes & millions of cells; Docker for workflow reproducibility.
Biology	Single cell and spatial multi-omics, Gene Regulation, Population genetics, Variant Effect Mapping, Cancer Plasticity, Immunogenomics (Basic), Brain Organoids.
Tools	Python, PyTorch, Scanpy, AnnData

SELECTED PRESENTATIONS

INVITED RESEARCH TALKS

2025	Genentech (Host: Avantika Lal), South San Francsico
2025	New York Genome Centre (Host: Neville Sanjana), New York
2025	Dana Farber Cancer Institute / Harvard (Host: Sasha Gusev), Boston
2025	MILA (Multi-omics Reading group), Virtual
2025	Stanford University (Anshul Kundaje Lab), Virtual
2024	Sanger Institute (Roser Vento-Tormo lab), Virtual

CONFERENCE ORAL PRESENTATIONS

09/2025 **Keystone Symposia: AI in Molecular Biology**, ***Santa Fe, NM, USA.***
Dissecting Cellular Plasticity in Glioblastoma via Deep Learning of Single-Cell Gene Regulatory Networks.

07/2025	**Leena Peltonen School of Human Genetics, *Cambridge, UK.*** Sequence-to-Expression Mapping in Personalized Genomes using Interpretable Deep Learning.
2022	**RECOMB/ISCB Conference on Regulatory and Systems Genomics with DREAM Challenges, *Las Vegas, USA.*** scDoRI: Gene regulatory inference from single-cell multi-omics data using interpretable deep learning.
2020	**Machine Learning in Computational Biology (MLCB) , *virtual.*** Convolutional Additive Models: a fully interpretable approach to deep learning in genomics.

SCHOLARSHIPS & AWARDS

2024	**Collaboration Funding Program for interdisciplinary projects (DKFZ) | EUR 20,000** Project Title: Integrating Single-Cell Perturbations and Causal Inference to refine Gene Regulatory Networks in Neurodevelopment. Applicants: Manu Saraswat, Laura Rueda Gensini, Moritz Mall, Oliver Stegle
2020	**DKFZ International PhD Fellowship | 3-Year funding** Competitive PhD fellowship

PROFESSIONAL ACTIVITIES

PEER REVIEW

2025	**Machine Learning for Genomics Exploration (MLGenX), ICLR**
2024, 2025	**Machine Learning in Computational and Systems Biology (MLCSB) at ISMB**
2021-2025	**Machine Learning in Computational Biology (MLCB) Conference**

MENTORING & COMMUNITY

2024- present	**Jiawei Luo, PhD candidate at DKFZ (Computer Science)**
2024	**Ian Dirk Fichtner, Master's Thesis (Molecular Biotechnology)**
2022-2023	**Katharina Mikulik, Master's Thesis (Molecular Biosciences)**
2019-2023	**Queer in AI** — Organiser (ICML Queer in AI workshop) and co-author on community research