Le machine learning au service de la santé et des sciences biomédicales

Guillaume Obozinski, Swiss Data Science Center

Symposium Annuel de Recherche Unisanté, 08.04.2025

The Swiss Data Science Center

Fostering the adoption of data science technologies — from data collection and management to machine learning, AI, and industrialization — serving **academic labs**, **hospitals**, **industry** and **public sector** stakeholders, including cantonal and federal administrations.

From a **Strategic Focus Area** of the ETH Domain to a **National Research Infrastructure** in 2025

Multidisciplinary, impact-driven team of **110+ professionals** across three locations

Applied Research, Innovation, Engineering & Services, Initial and Continuing Education

Health & Biomedical, Climate & Environment, Energy & Sustainability, Large-scale scientific infrastructures

A significant involvement of the SDSC in Health and Biomedical Sciences !

- Supporting 3 SPHN/PHRT National Data Streams projects
  - LUCID: Low value care in medical Hospitalised patients
  - SPO: Swiss Personalised Oncology
  - IICU: Infection in Intensive Care Units

  ➞ automated data workflows, data management, data quality screening, etc

- Swiss Federated European Genome-Phenome Archive (EGA)
- Deploying ML for risk of pressure injuries and recognising sepsis to improve patient care (with CHUV)
- 16 ongoing or complete research collaborations in Health and Biomedical Sciences, e.g.: Activity biomarkers of Parkinsonian neurones (DeepEphys), Opto-acoustic imaging (DLbirhoui), Type-II Diabetes patient trajectories (Polynet).
- Models and online dashboard for Covid epidemic forecasting

Harnessing attention mechanisms to decode chromatin interactions across diverse cell types and conditions

Mutiger project in collaboration with the Computational Cancer Genomics lab at ETH Zurich

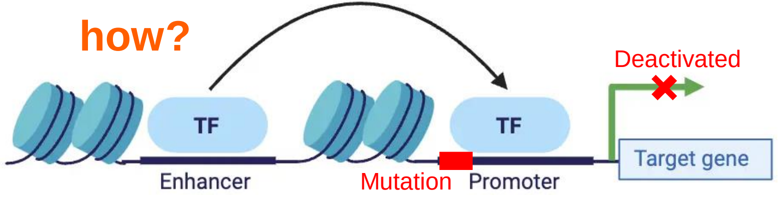
Mutations, Interactions and Gene Interactions

Challenge Non-coding parts of the genome play an important role in gene expression regulation.

Objective ML model for in silico prediction of **Enhancer-Promoter Interactions** using only DNA sequence and chromatin accessibility data (ATAC-seq), that generalises to new cell types.

Impact Decipher the effects of non-coding variants on cancer cells to design personalised treatments

Where and how?



Enhancer - Promoter Interaction (EPI)

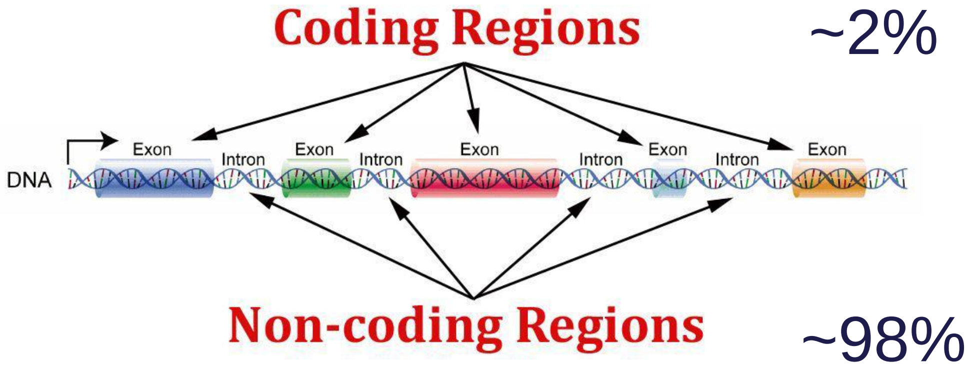
Personalized treatment

Collaborators

Computational Cancer Genomics lab, ETH Zurich
Prof. Valentina Boeva
Aayush Grover

SDSC
Dr Lin Zhang,
Till Muser
Dr Ekaterina Krymova

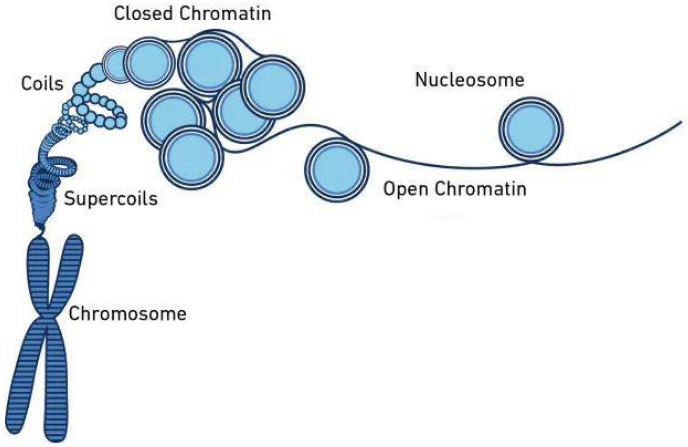
ETH zürich



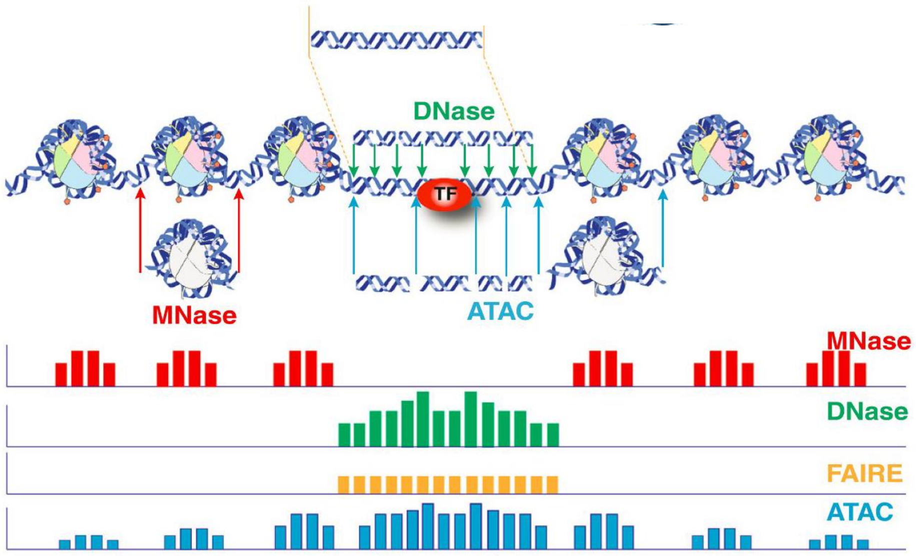
[https://mappingignorance.org/2021/10/04/mutations-in-the-non-coding-genome-contribute-to-autism/]

Experimental cell-specific profiling of Enhancer-Promoter Interactions (EPI)

Open Chromatin

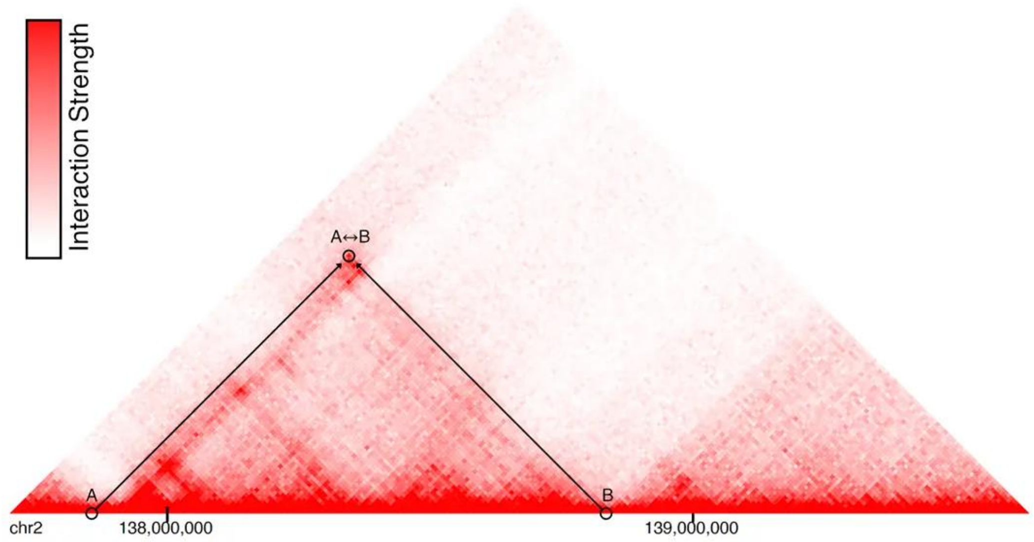


[https://www.10xgenomics.com/blog/]



[https://www.michaelchimenti.com/2020/09/atac-seq-best-practices-tips/]

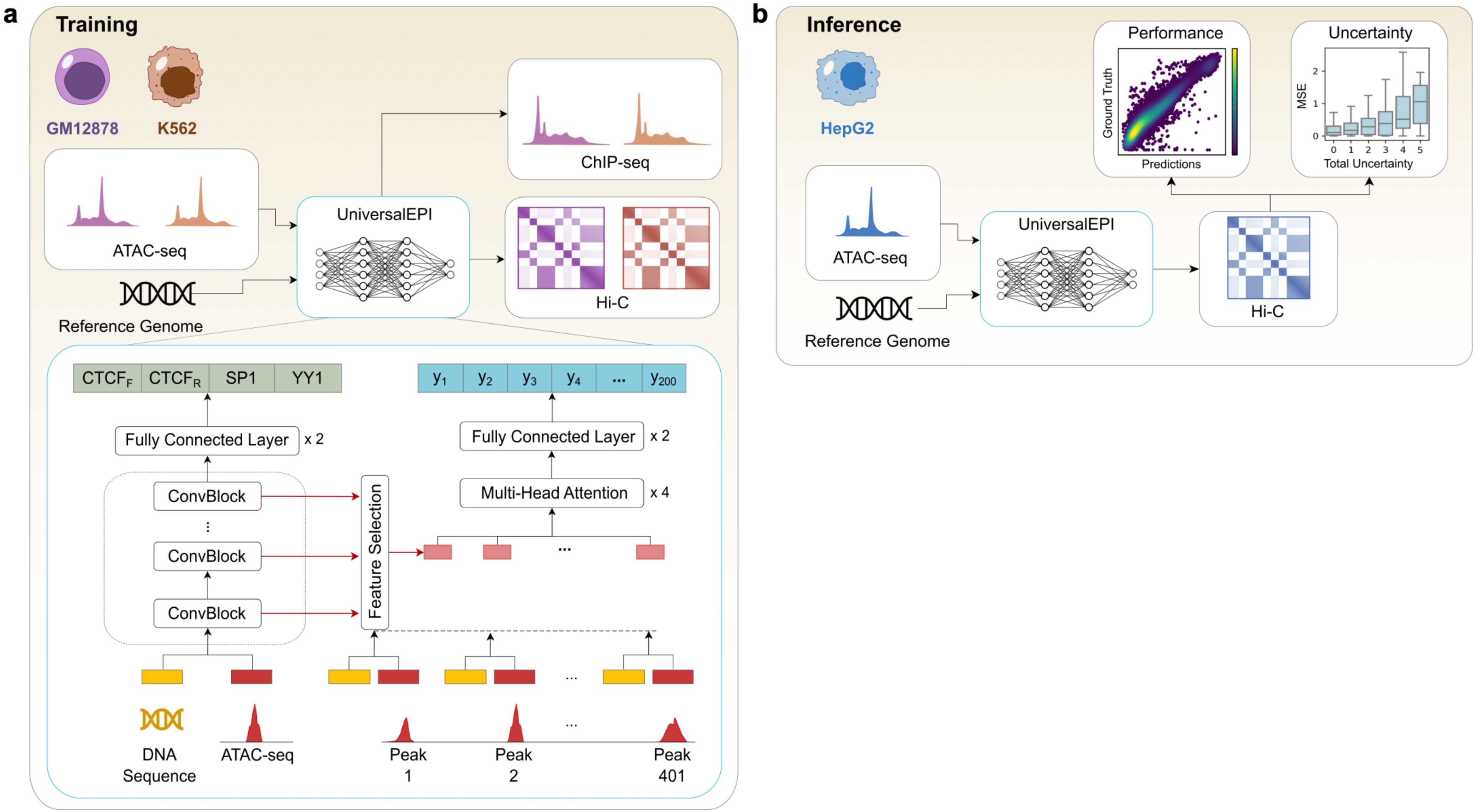
Hi-C



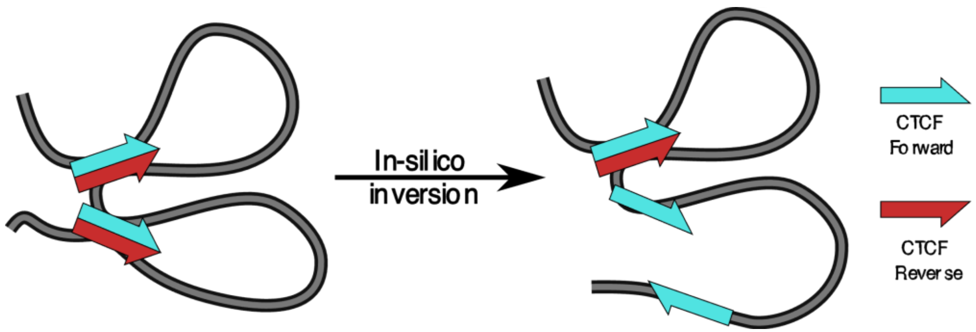
High-throughput Chromosome Conformation Capture (Hi-C) is unfortunately costly and labor intensive

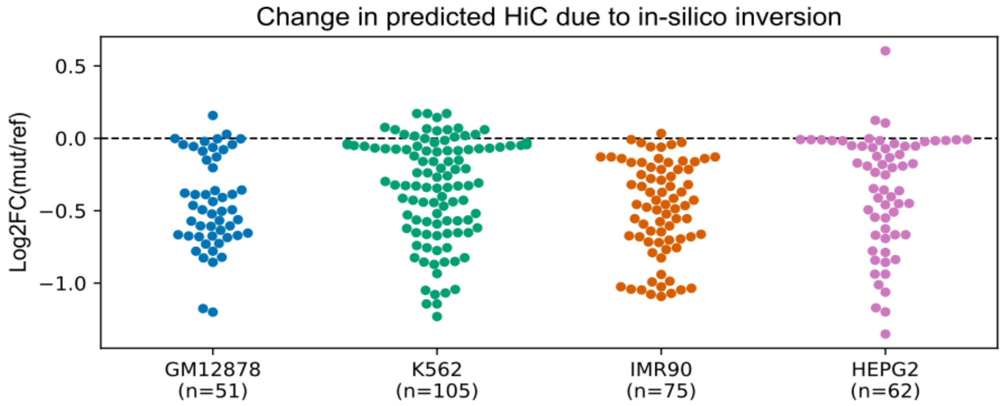
- ➞ High-throughput genetic perturbations impossible
- ➞ Motivates in-silico methods

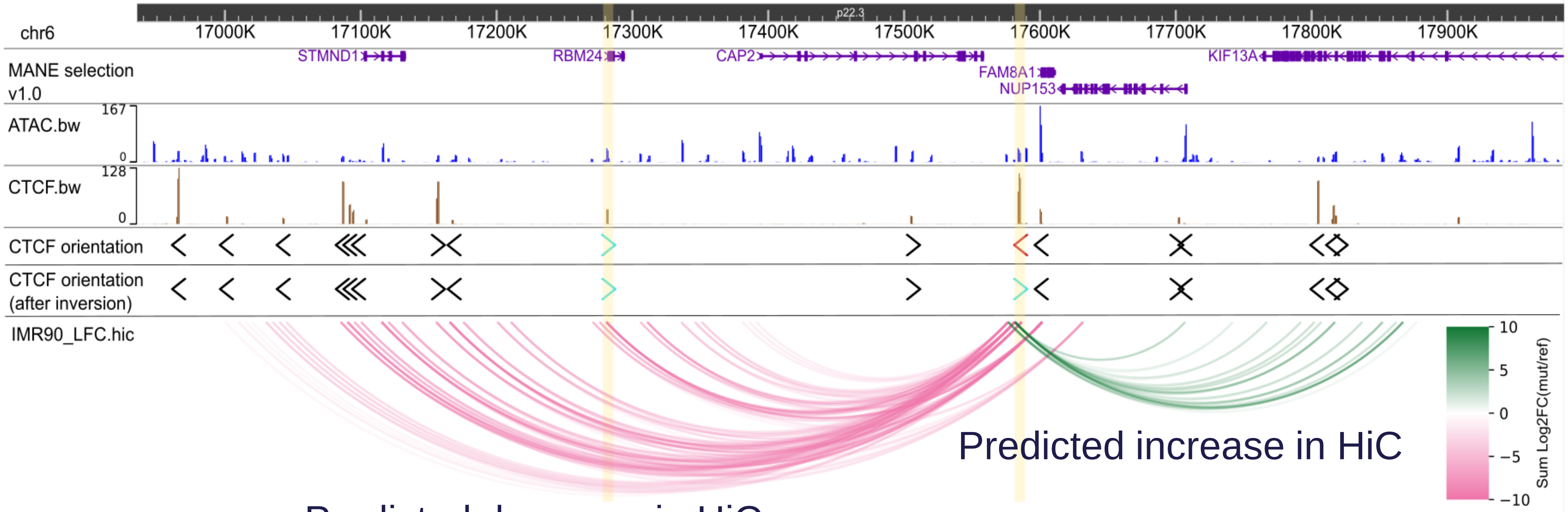
UniversalEPI: an attention-based framework for predicting EPIs



In-silico genomic perturbation



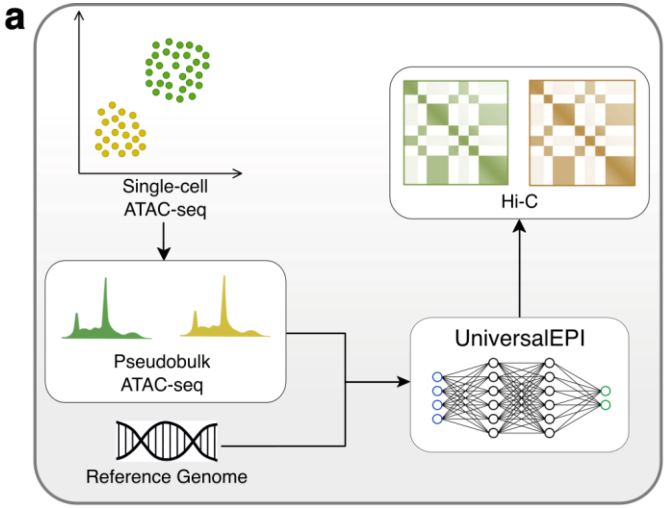


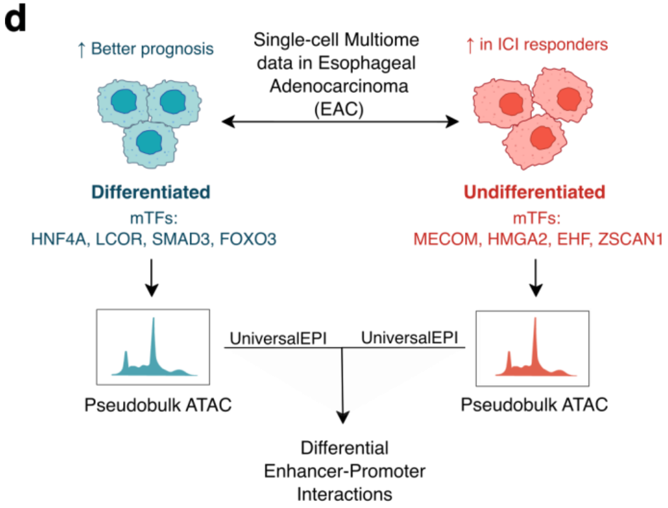


Predicted increase in HiC

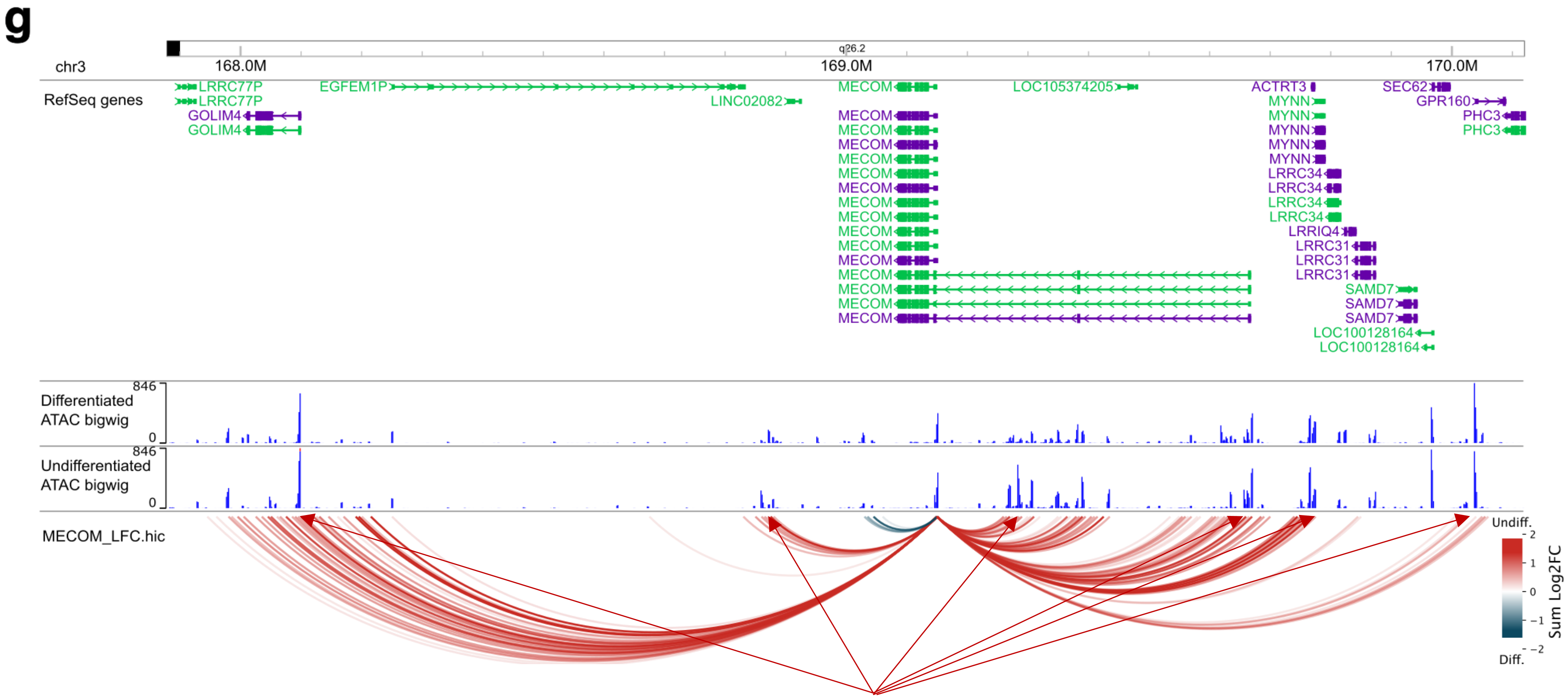
Predicted decrease in HiC

Cell-type specific prediction using pseudo-bulk single-cell ATAC-seq





MECOM gene: over-expression linked with uncontrolled cell growth and reduced apoptosis



Higher promoter activity in the undifferentiated cells

Summary

The proposed ML algorithm enables

- Accurate Hi-C prediction of gene regulatory interactions
- Quantification of uncertainty on predictions
- In-silico genetic perturbations to study effects of non-coding variants
- Chromatin structure prediction in complex tissues using single-cell data

A neural network model to learn myotomal innervation for individual patients

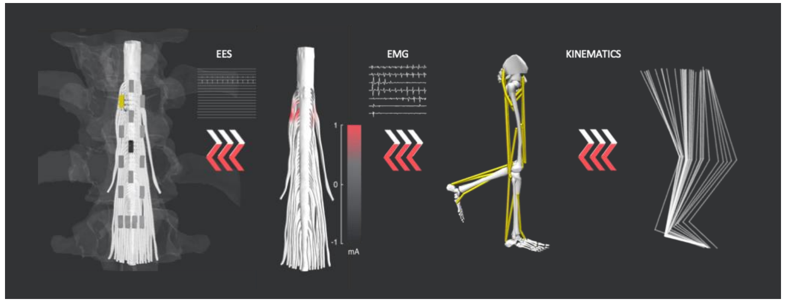
STIMO project in collaboration with .NeuroRestore, EPFL, CHUV, and UNIL

A neural network model to learn myotomal innervation for individual patients

Challenge Restoring movement after **spinal cord injury** requires **personalized mapping** between nerve stimulation and muscle activation.

Objective Build a model that predicts muscle activity from simulated nerve responses, while **learning patient-specific myotomal innervation**.

Impact Guide individualized spinal cord stimulation therapies to **improve motor recovery**.

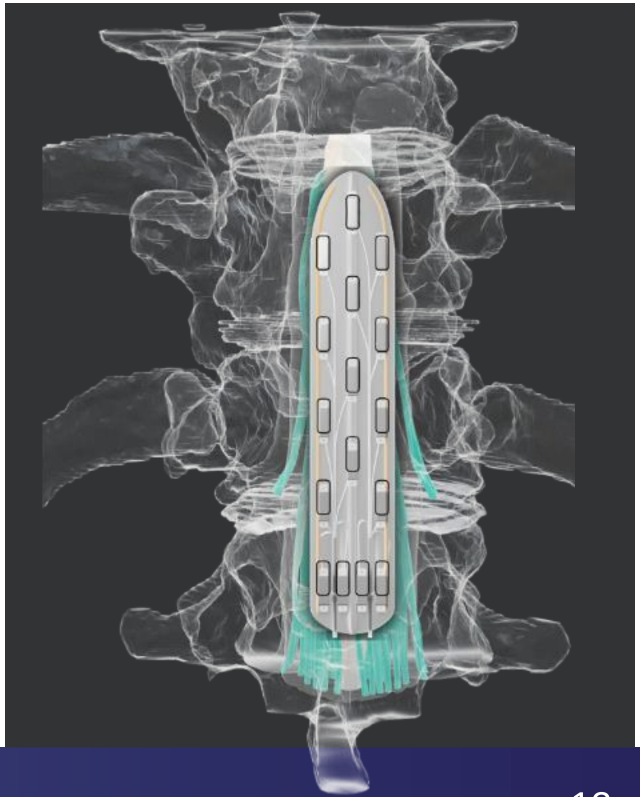


project funded by
Strategic Focus Area
**Personalized Health
and Related Technologies**

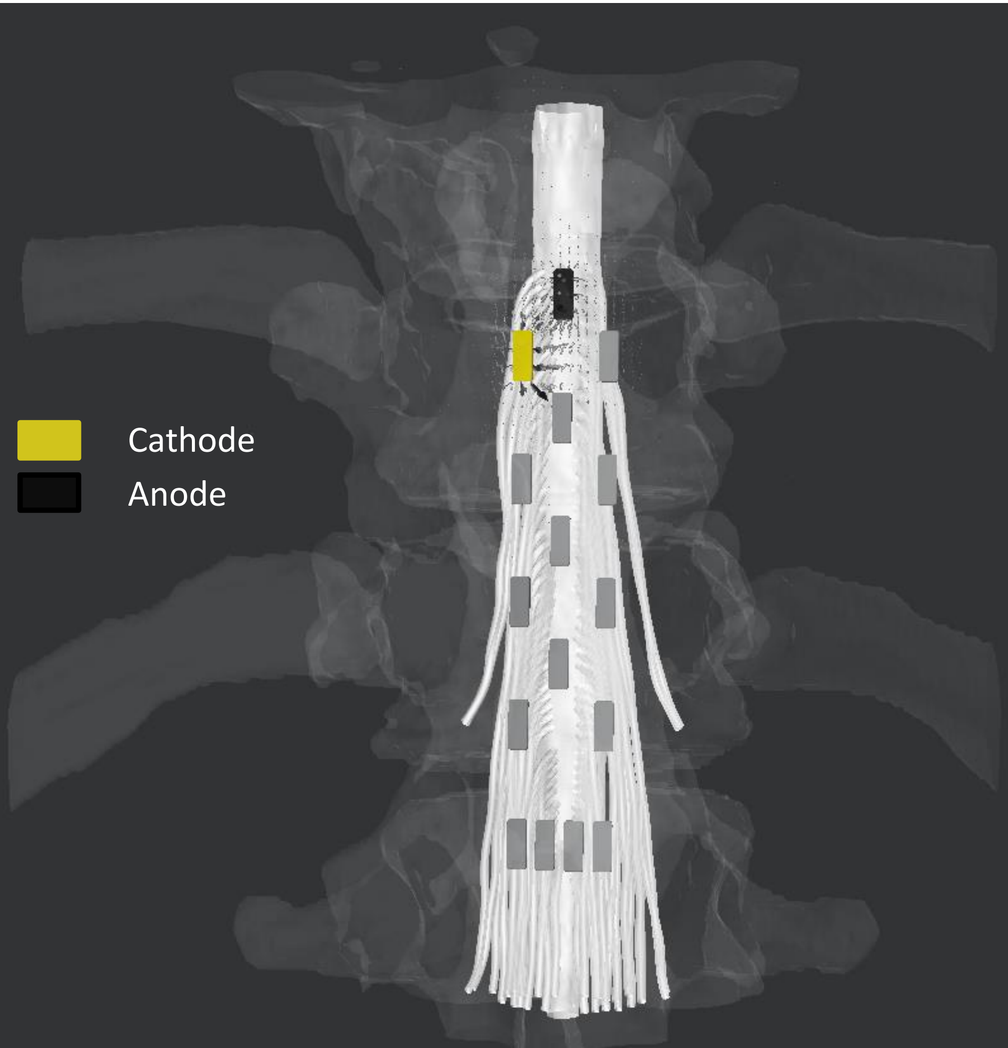
Collaborators

**.NeuroRestore
EPFL, CHUV, and UNIL**
Prof. Grégoire Courtine
Prof. Henri Lorach
Dr. Grégory Dumont
Dr. Alice Bruel

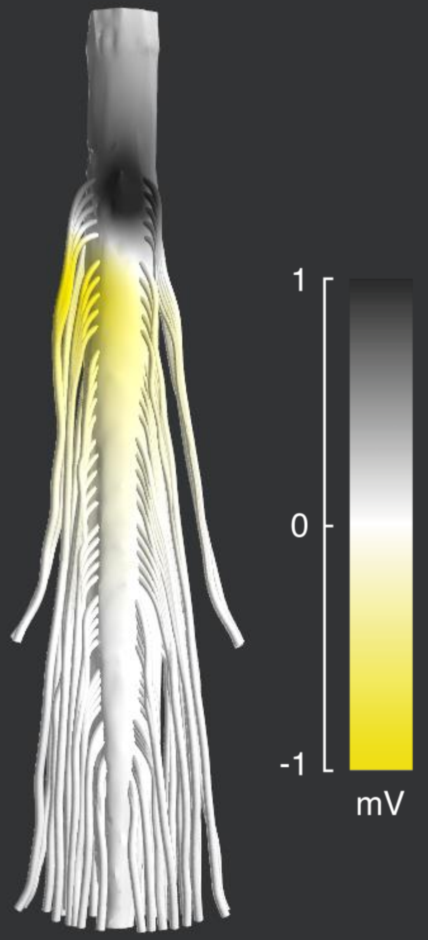
SDSC
Dr. Oleg Bakhteev,
Dr. Leonid Iosipoi
Prof. Guillaume Obozinski
Prof. Mathieu Salzmann



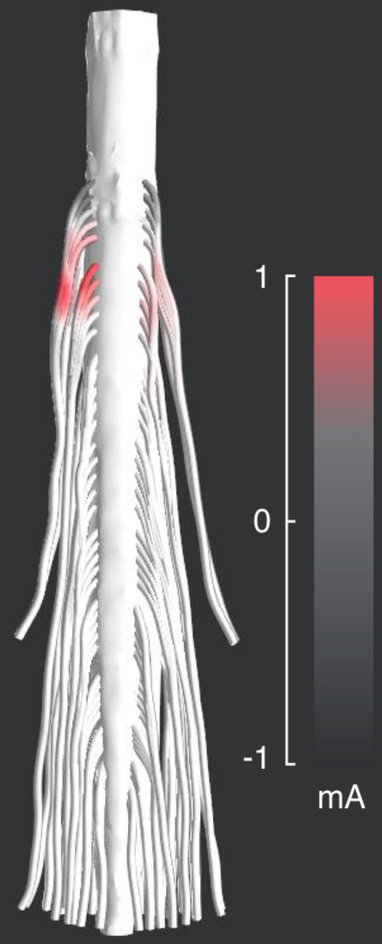
From the electrosimulation to the neural current



Simulation of current with a physics-based numerical model



Voltage

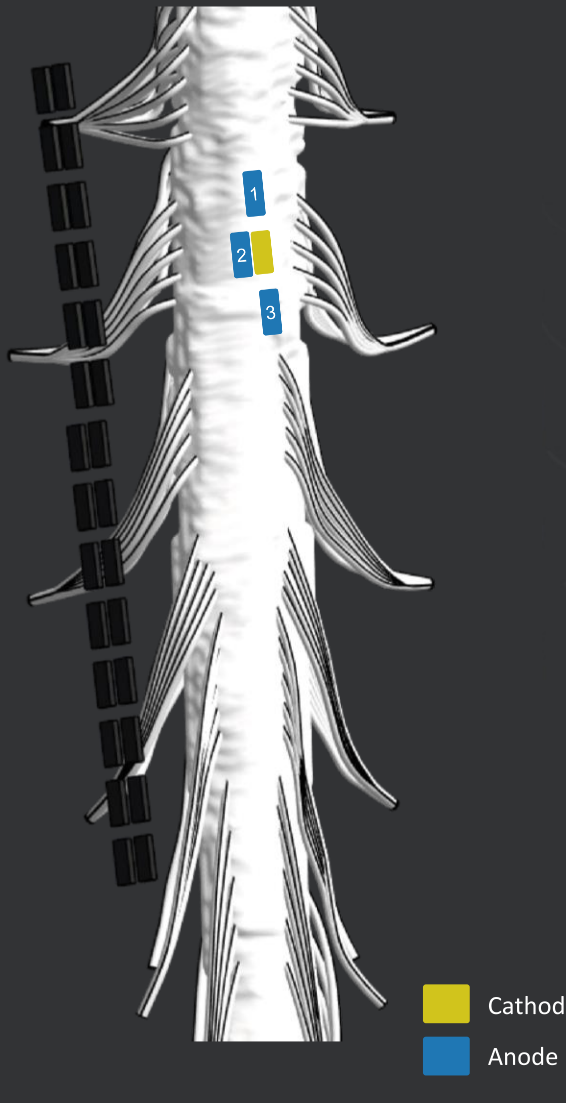


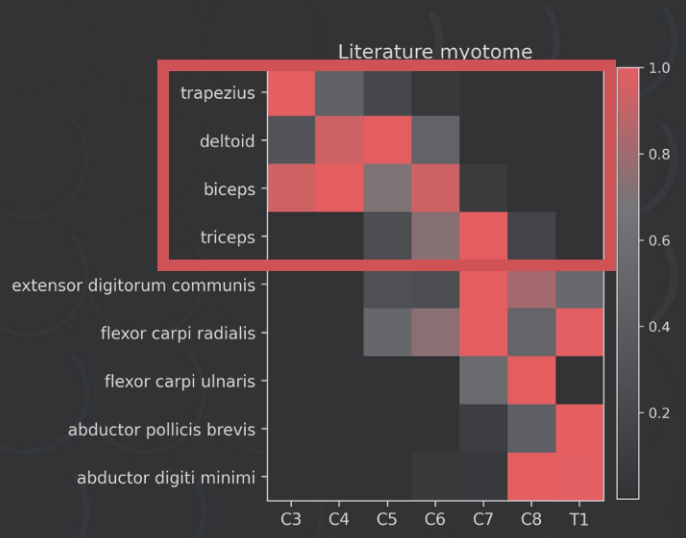
Neural current

Upper limb data and objective

Upper limb data:

- 3 days of recordings
- Every cathode placement was tested (28 possible positions)
- Frequency: 30 Hz
  Amplitude: ramp 0-5mA
- Electromyograms (EMGs) for 18 upper limb muscles
- Tested anodes are chosen adjacent to the cathode

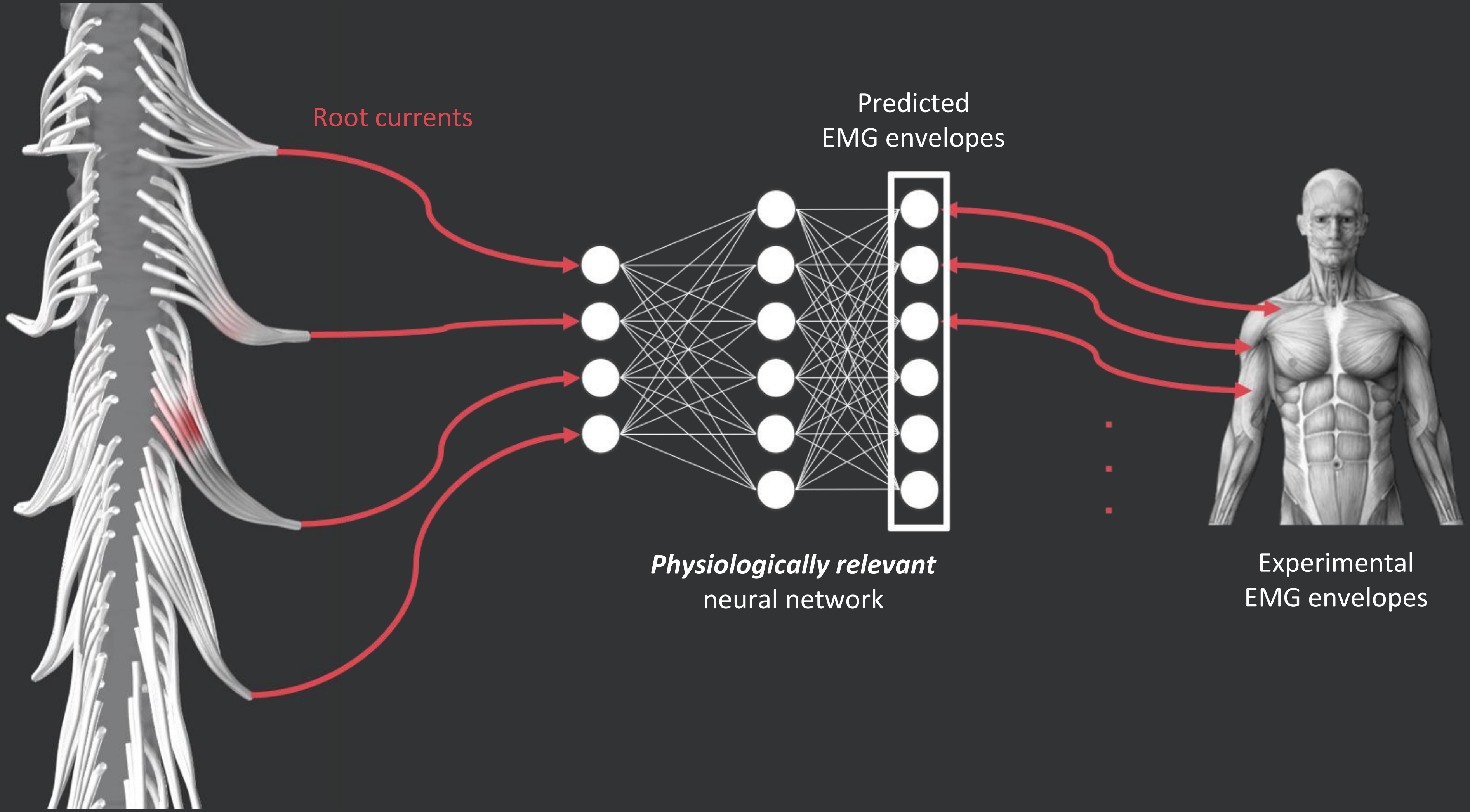




Heuristic map of myotomal innervation in humans using direct intraoperative nerve root stimulation
Journal of Neurosurgery: Spine / Volume 15 / July 2011

An interpretable model to learn the contributions of rootlet stimulation to individual muscle responses

Swiss Data Science Center

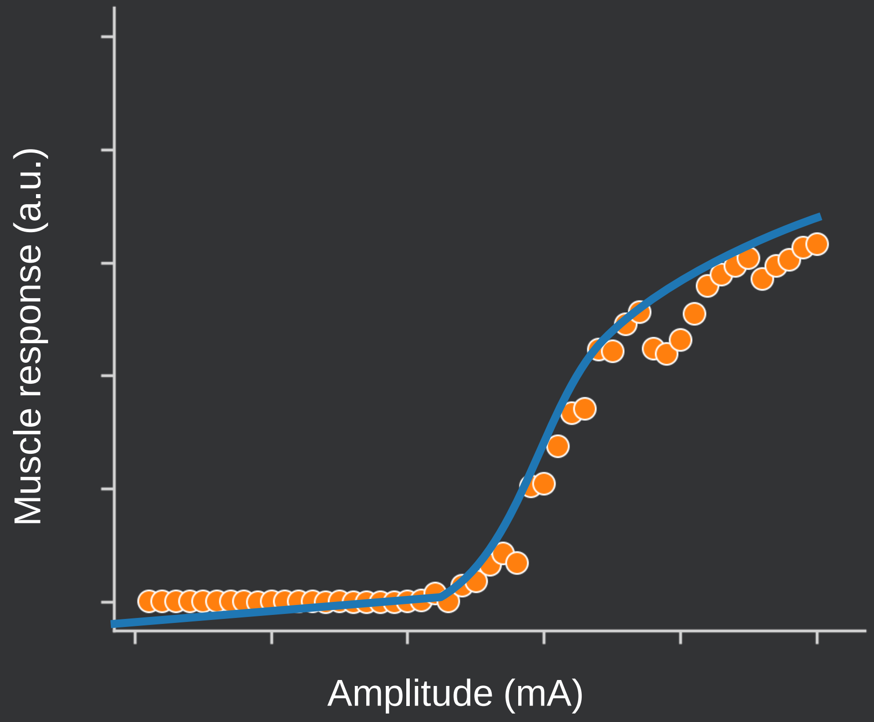


Biophysical simulation

A physiologically relevant model

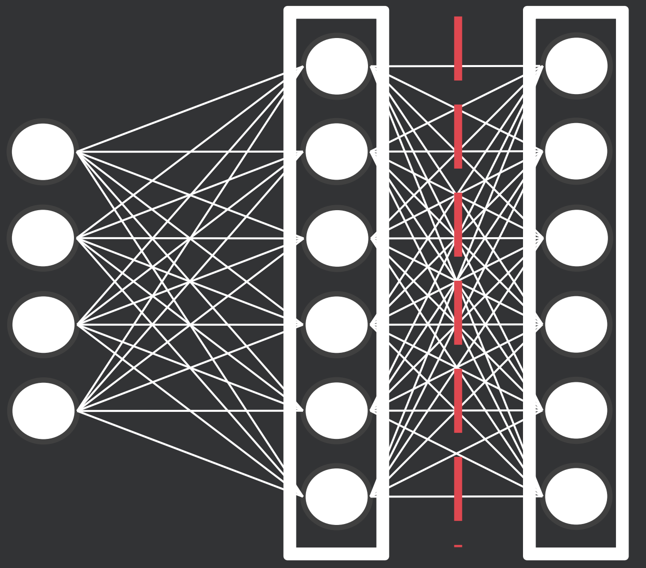
Sigmoid-shaped response

Match physiological FI-curves



Roots contribute additively to muscle response with

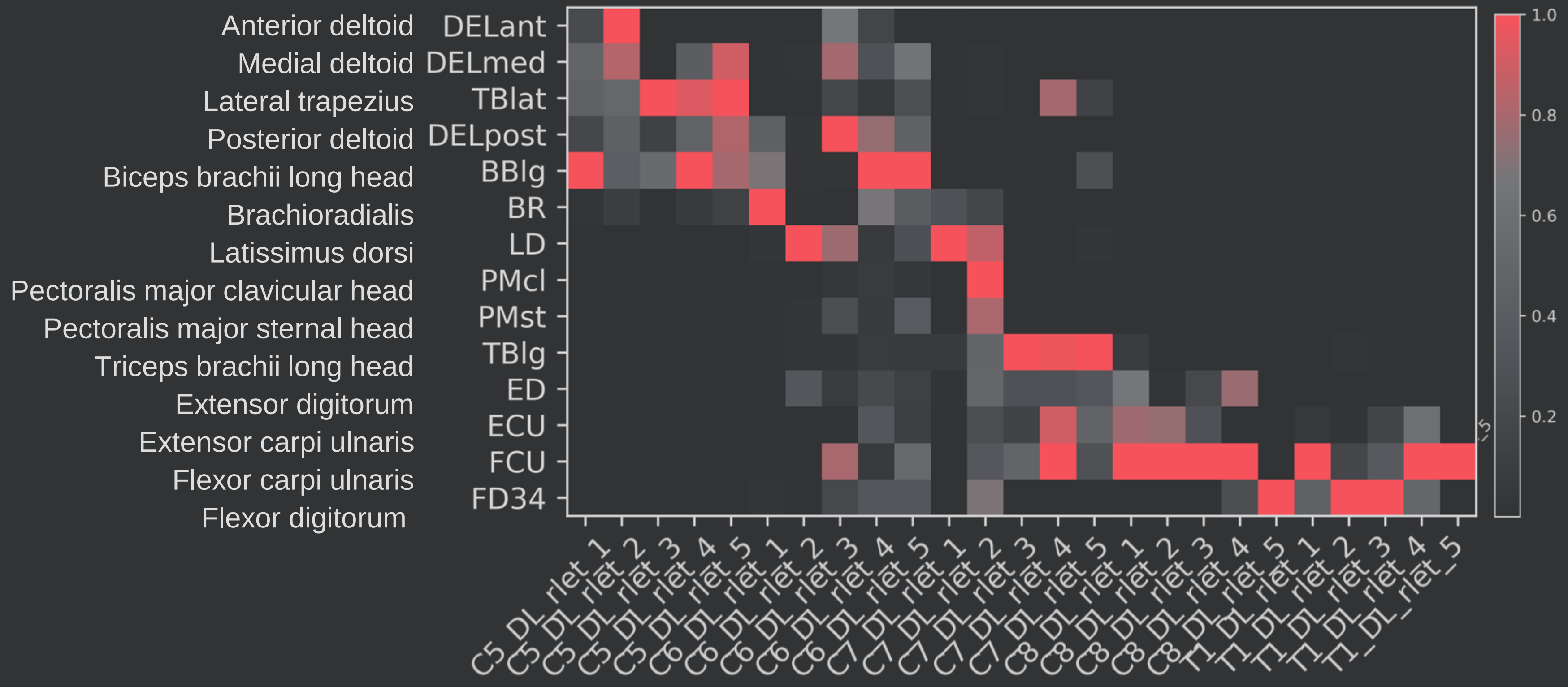
Positive weights



Root activations layer

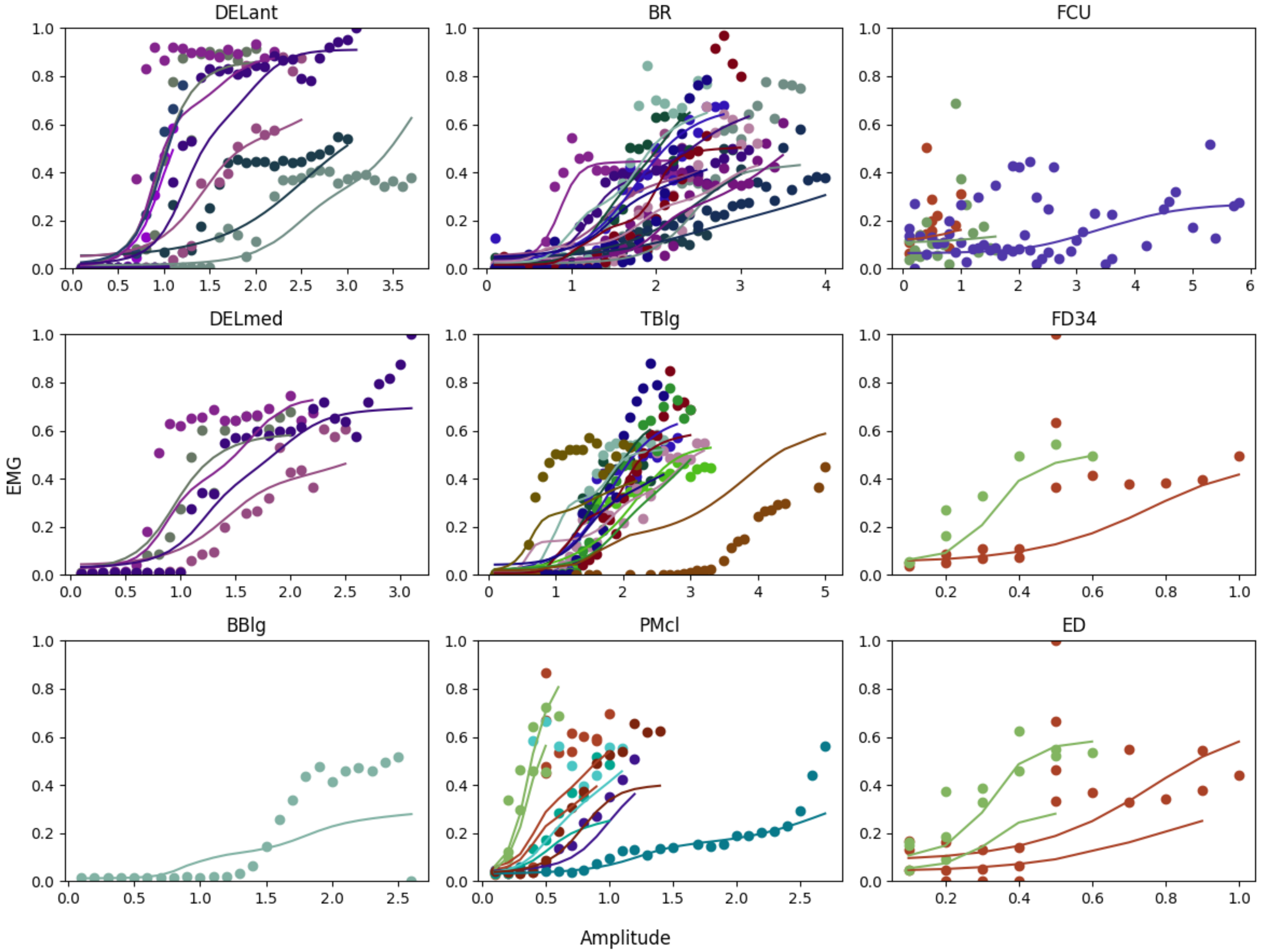
Muscle layer

Results: Estimated connectivity



Frequency of rootlet selection in the model

Observed vs predicted EMG peak intensity for different electrode configuration and amplitude of stimulation



- predicted (solid line) *vs*
- observed (dots)
  EMG maximal response

for configurations unseen in training

as a function of

- stimulation amplitude (x-axis)
- electrode configuration (color)

Conclusion

The proposed algorithm aimed at mapping muscle response to individual rootlet stimulation

Remaining challenges:

- Rootlet activity strongly correlated
- Limited amount of data per patient
- In general, a limited number of configurations are (can be) tested
- Sources of noise and variability (change of sensor placement, patient fatigue)

General conclusion

- Many settings in HBMS where sample size is small but variables/features high
  - ➞ A strong need for specific ML models using precise domain knowledge
- Transformer architectures find applications in many domains
- Uncertainty quantification is key for health and biomedical sciences

Thank you!

www.datascience.ch