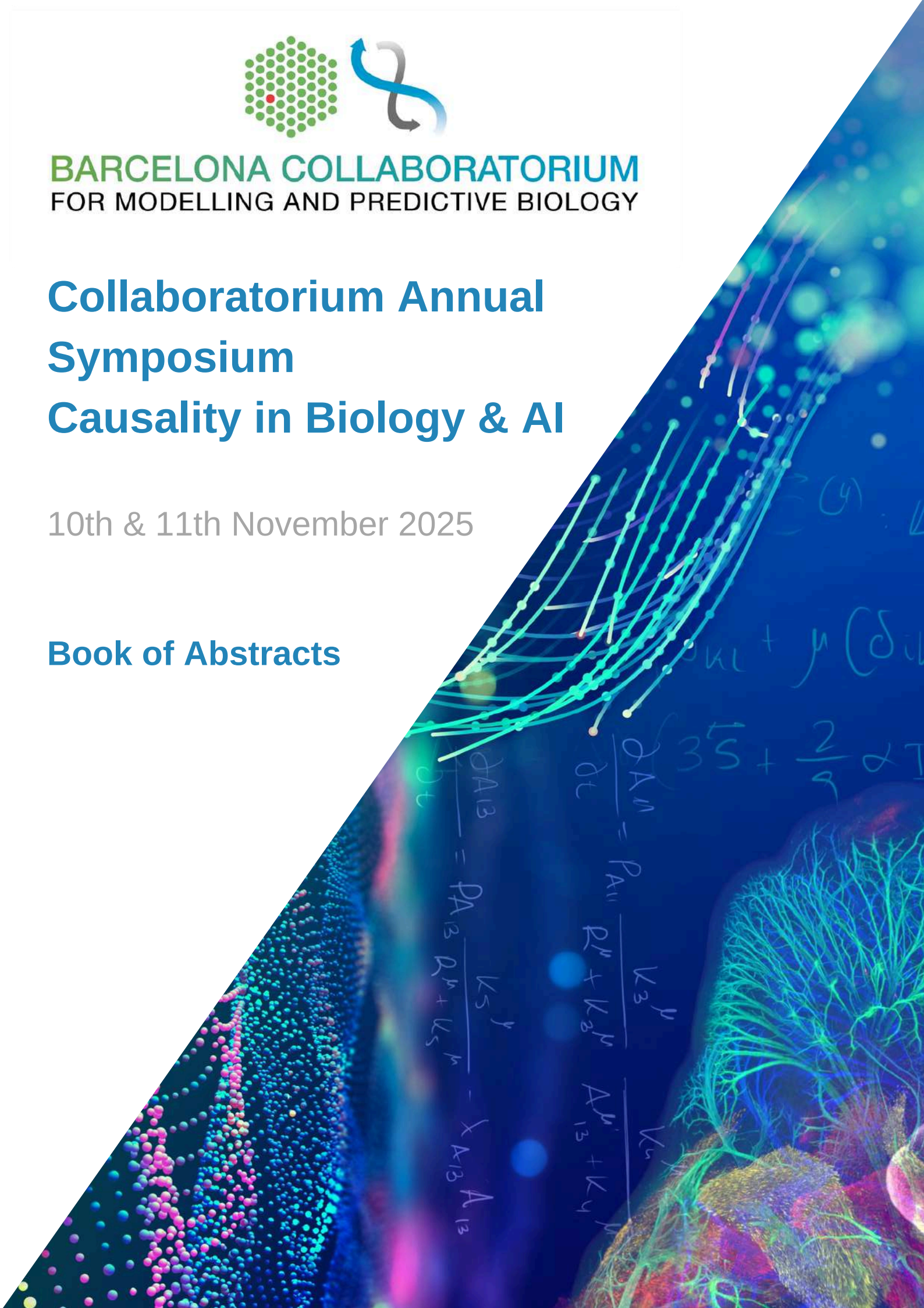
BARCELONA COLLABORATORIUM
FOR MODELLING AND PREDICTIVE BIOLOGY

Collaboratorium Annual Symposium
Causality in Biology & AI

10th & 11th November 2025

Book of Abstracts



SUPPORTED BY:

FUNDERS:

PROGRAM

DAY 1 - NOVEMBER 10th

14:15. Welcome & Presentation James Sharpe (Collaboratorium – EMBL Barcelona); Mafalda Dias (CRG)

Session I: Causality Concepts — Chair: Mafalda Dias

14:30. Can modelling predict life?
Denis Noble (University of Oxford)

15:10. Causal Learning for Complex Cellular Systems
James DiFrisco (Francis Crick Institute)

16:20. Emergent (Non-)Causal Abstractions Over Causal Structure in Mesoscale Biological Systems
Jonas Hartmann (UCL)

17:00. Downwards Causality from Tissues to Genes
James Sharpe (EMBL Barcelona)

17:40. Integrating Biology Across Scales: How Causal is "Causal"?
Ava Khamseh (University of Edinburgh)

18:20. *End of Day I — James Sharpe (EMBL Barcelona) and reception at the Collaboratorium at 19:00*

PROGRAM

DAY 2 - NOVEMBER 11th

09:15. Summary of Day 1 & Introduction to Day 2 Mafalda Dias (CRG)

Session II: Causality at Different Scales of Biology — Chair: James Sharpe

09:30. Causal Inference in Molecular Systems
Gaudenz Danuser (Institute for Human Biology)

10:10. Sequence-based interpretable dimensionality reduction decodes cellular plasticity
Oliver Stegle (EMBL Heidelberg)

11:20. Host-Microbiome Interactions: Leveraging Genetic Variation in the Host to Better Understand Causality
Amelie Baud (CRG)

12:00. What causes inter-individual variation in lifespan?
Nicholas Edward Stroustrup (CRG)

12:40. Flash Talks — Tanmayee Narendra, Vasilis F. Ntasis, Júlia Vicens Figueres, Dimitris Volteras

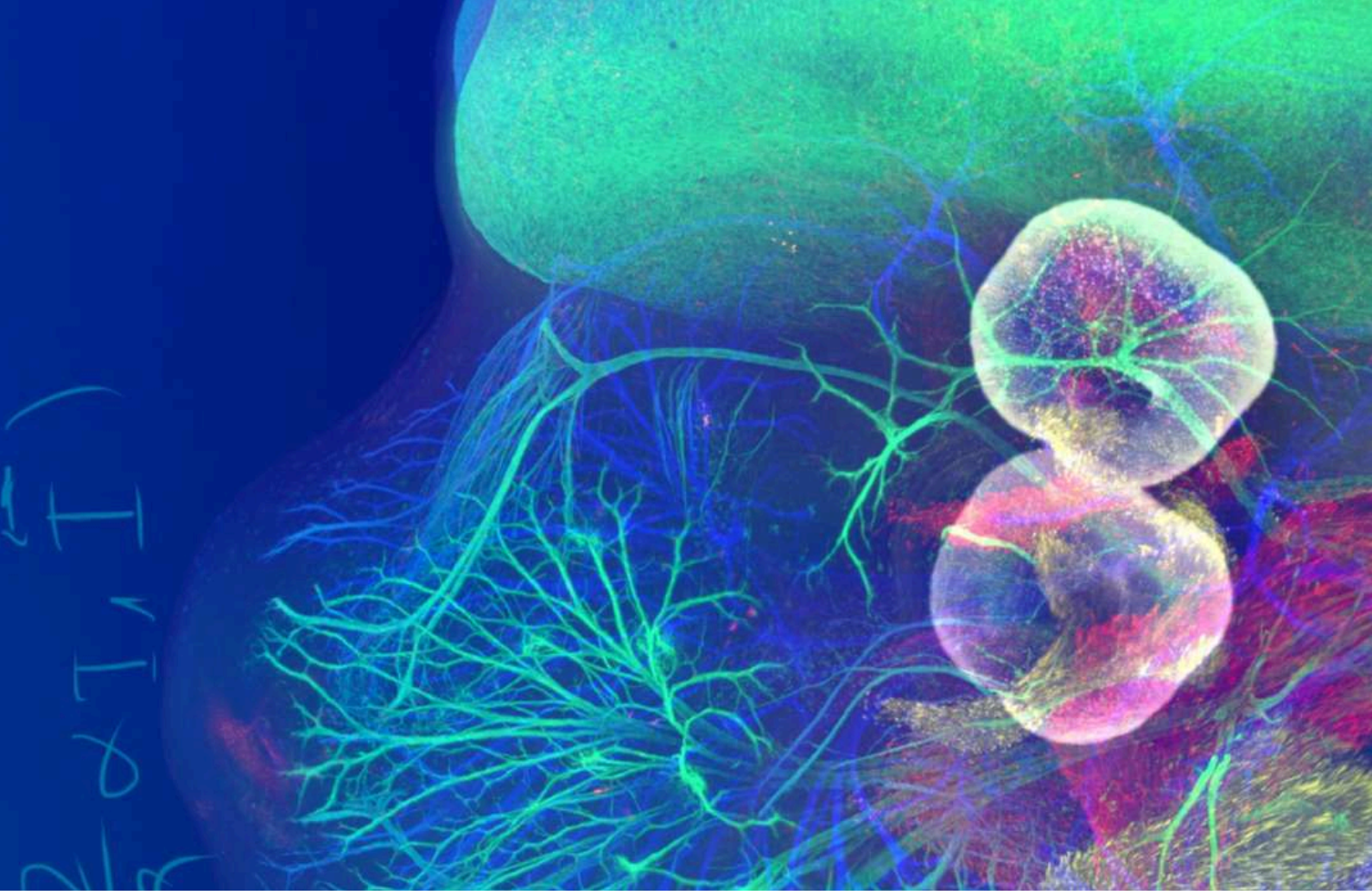
Session III: Causality in AI Models — Chair: Gaudenz Danuser

14:30. Causal Representation Learning: From Biomarkers to Mechanisms
Caroline Uhler (Broad Institute of MIT & Harvard)

15:10. Explainable AI for Protein Design
Noelia Ferruz (CRG)

16:20. Causal Molecular Design
Eli Weinstein (DTU, Denmark)

17:00. *Closing Remarks* James Sharpe (Collaboratorium – EMBL Barcelona)



Speakers

BARCELONA COLLABORATORIUM
FOR MODELLING AND PREDICTIVE BIOLOGY

SCIENTIFIC ORGANISERS

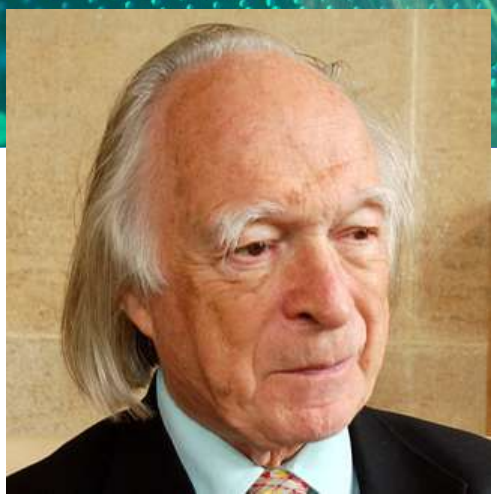
GAUDENZ DANUSER
Co-Director, Computational Biology, Institute for Human Biology (Roche); Head, Organoid Systems Biology

JAMES SHARPE
Head of EMBL Barcelona; Group Leader (Sharpe Group); Collaboratorium Co-Chair

MAFALDA DIAS
Group Leader, Dias & Frazer Lab, Centre for Genomic Regulation (CRG)

ROSA MARTÍNEZ-CORRAL
Previously Independent Fellow (group-leader level), Martinez Lab, Centre for Genomic Regulation (CRG)
Associate Professor, University Pompeu Fabra (UPF)

BARCELONA COLLABORATORIUM
FOR MODELLING AND PREDICTIVE BIOLOGY



DENIS NOBLE

University of Oxford

Can modelling predict life?

When the Human Genome Project first announced the complete sequencing of the human Genome in 2001, the prediction (Collins, 1999) was that within 10 years, we would not only be able to predict disease states from genomes, but also discover the cures for most diseases. 25 years later, neither has been achieved. Instead, we have discovered that the great majority of association scores between genes and disease states are surprisingly small. So small that, in order to try to get predictive results, we need to add association scores together to produce a polygenic score. My group's work on modelling the heart pacemaker (Noble et al 1992) showed both why most association scores are so low, and why the outcome is not predictive (Hingorani et al 2023). The reason is that association scores do not reveal causation. In the heart modelling work an association score of around 10% for a given gene product may hide causation of as much as 85%. Hillenmeyer et al (2008) showed the same phenomenon when examing the 6000 genes in yeast. Most show a zero association score under optimal physiological conditions, yet strong causality is revealed when important nutrients are deficient in the environment.

It is time therefore to abandon the idea that genes are the blueprint for life (Noble, 2024), and to replace the Central Dogma of Molecular Biology with causal diagrams of functional physiological networks (Noble, 2025a, b) that show how organisms use stochasticity at molecular and other lower levels of organisation to generate functional physiological outcomes. With that information, physioloical multi-scale modelling may develop the solution to the gene-centric impasse in finding cures for complex polygenic diseases.

JAMES DIFRISCO

Francis Crick Institute

Causal Learning for Complex Cellular Systems

In contrast to description and prediction, causal understanding enables intervention and control. Yet efforts to reach causal understanding in biological systems quickly run up against their astounding complexity. This talk will provide an overview of the central problem facing efforts to discover causal mechanisms of development and evolution in conditions of biological complexity, and will outline how these problems might be addressed. Themes to be addressed include variational versus mechanistic causation, ‘sloppy’ models, modularity, downward causation, the distinctive causal role of genes, and the prospects for AI-guided mechanistic research.

JONAS HARTMANN

UCL

Emergent (Non-)Causal Abstractions Over Causal Structure in Mesoscale Biological Systems

A major aim of cell and developmental biology is to understand the complex causal structure of living systems above the molecular scale. However, even if we were to chart the complete set of causal interactions at play within a given model system, the resulting map would be so complicated as to be largely inaccessible to human reasoning and so context-sensitive that it would resist generalization to different conditions, other systems of interest, or engineering problems. This motivates our search for novel non-standard abstractions that enable the distillation of comparably simple and general models which nonetheless capture a biological system's salient causal structure.

We have recently developed one such abstraction, the Core & Periphery (C&P) hypothesis, which proposes a decomposition of complex biological systems into a highly versatile system core and a complementary system periphery that configures or “programs” the core to perform one particular function from its vast repertoire. By facilitating variation, multi-functionality, and plasticity, versatile cores can become widely reused and deeply entrenched in evolution, which confers generality to both the theoretical principles and the biological mechanisms that underpin them. We found that this abstraction can be fruitfully applied to reframe a variety of complex biological phenomena, including actomyosin dynamics, Turing patterning, and different modes of collective cell behavior.

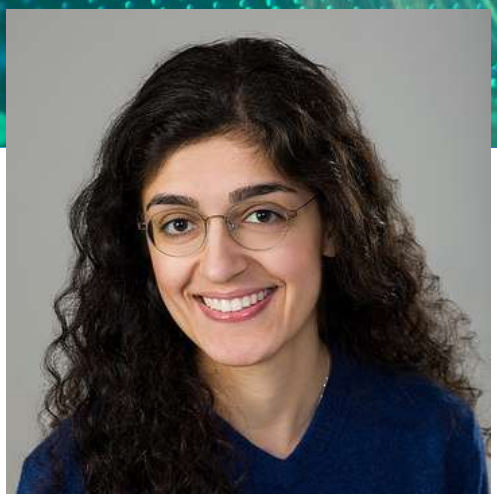
Here, I will introduce the C&P hypothesis along with some of its key predictions and implications, and I will discuss how it might relate to established frameworks of causation in cell and developmental biology. Time permitting, I will also provide glimpses into ongoing work to formalize and model C&P, and/or into another non-standard abstraction we are currently exploring.

JAMES SHARPE

EMBL Barcelona

Downwards Causality from Tissues to Genes

A common theme within modern biology is the notion that control logic resides at the level of molecular networks. For example, we understand that the circuitry of GRNs (gene regulatory networks) which is an important part of the decision-making process for cells, appears analogous to the control logic of electronic circuits designed to achieve a specific task (albeit analogue circuitry, rather than digital). It is also clear however that large multicellular systems (such as tissues, organs or even whole organisms) must exhibit control logic in which macroscopic features (such as size and shape of an organ) directly feeds back to control molecular events (such as regulation of genes that promote growth). This has been labelled as “downward causation” by Denis Noble and others, and although this is an indisputable phenomenon it has received far less attention in the life sciences. I will discuss our work in modelling the development of the mouse limb bud, and how this should help us to directly study downwards causation from tissues to genes.



AVA KHAMSEH

University of Edinburgh

Integrating Biology Across Scales: How Causal is "Causal"?

Genomic medicine goes beyond identification of risk factors and aims at pinpointing underlying causal mechanisms, often at various biological scales.

This requires to close in on the cell (sub)type and state of origin, together with its underlying molecular mechanisms. In the first part of this talk, we introduce Stator, a data-driven methodology that identifies cell (sub)types and states without relying on cells' local proximity in transcriptome space, in contrast to methods such as clustering. Stator labels the same single cell multiply, not just by type and sub-type, but also by state such as activation, maturity or cell cycle sub-phase, through deriving higher-order (beyond pairwise) gene expression dependencies from a sparse gene-by-cell expression matrix. We exemplify Stator's fine resolution in liver disease/cancer to reveal expression programmes that identify relevant disease cell types/states, consequences of their progression, and generation of hypotheses on their corresponding causal molecular mechanisms for experimental validation.

In the second part of the talk, we present a novel approach, integrating population genetics, functional genomics and targeted machine learning (TarGene), to quantify epistatic contributions to human traits via transcription factor mechanisms. By taking experimentally verified differentially binding variants across 9 nuclear hormone receptors as candidates and using UK Biobank data across 768 traits, we reveal, for the first time, hundreds of epistatic interactions as causal candidates for these transcription factor mechanisms of disease.

We will discuss why integrating biology across scales is an essential ingredient for revealing causal mechanism of disease.

BARCELONA COLLABORATORIUM
FOR MODELLING AND PREDICTIVE BIOLOGY

GAUDENZ DANUSER

Institute for Human Biology

Of Cause and Effect in rapidly self-organizing Systems

Determining causal structure is the wholly grail of most scientific inquiries. When possible, the structure is reconstructed via interventions with individual components and interpretation of the ensuing system responses in a directed graph of cause-effect relations. In many cases, however, it is technically not feasible to apply well-controlled interventions. Instead, causal structure must be inferred from observational data. Moreover, experimental intervention generates limited information for causal analysis in systems with significant nonlinearity and redundancy between components. Many molecular systems in biology fall in this category. While molecular and genetic engineering provides a powerful means to perturb specific components or component interactions, the deduction of cause and effect is challenged by compensatory responses.

To overcome this fundamental problem, we adopt strategies of statistical causal inference. One popular family of methods relies on statistical time series analysis and the notion of Granger causality, which prescribes that a cause-effect relation between two components will exist if the past of the former is indispensable for predicting the future of the latter. We exploit live cell imaging to visualize the dynamics of molecular processes and develop advanced computer vision methods to transform 2D and 3D image data into mathematical representations of spatially resolved time series conducive to causal inference. We exploit the same representations also to infer cause-effect relations under the ‘treatment-response’ paradigm, which offers more flexibility in systems with a spatiotemporally transient causal structure.

Key to this application is the definition of ‘treatment’ events in observational data, which we accomplish by Hidden-Markov modeling of transient co-fluctuations between component time-series that may be causally coupled. This presentation will outline a number of recent efforts to overcome the limitation of experimental perturbation in determining causal structure in molecular systems with a high degree of nonlinearity and redundancy and describe how these computational tools are leading us to unexpected discoveries of cell regulation.

BARCELONA COLLABORATORIUM
FOR MODELLING AND PREDICTIVE BIOLOGY

OLIVER STEGLE

EMBL Heidelberg

Sequence-based interpretable dimensionality reduction decodes cellular plasticity

AI is a foundational tool for decoding human sequence variation, allowing to pin-point regulatory variants and predict their effects across molecular layers, cell types and cell states. However, translating sequence models into downstream insights has remained challenging. I will describe advances how to combine sequence-based modelling with multi-modal manifold learning, thereby predicting and understanding cellular and organismal traits. Among others, I will present applications for decoding cellular plasticity in human tumors, where our framework uncovers new insights into the regulatory logic that promotes or restricts cell state transitions.

AMELIE BAUD

CRG

Host-Microbiome Interactions: Leveraging Genetic Variation in the Host to Better Understand Causality

Causal effects of the microbiome on health: hype or reality? To distinguish between correlation and causation in an observational setting, we combine the use of outbred rodents housed in controlled laboratory conditions and of host genetic variants serving as causal anchors. In this talk, I will present the series of empirical findings that led us to investigate the impact of the gut microbiome on health — starting with evidence that the immune status of (clean) laboratory mice is affected by the genetic makeup of their cage mates (how?) — and present data suggesting that the reach of the microbiome may extend even further than we thought…

NICHOLAS EDWARD STROUSTRUP

CRG

Systems Biology research programme

Aging involves functional declines across multiple spatial scales, with molecules, cells, tissues, and organs changing together over time. Complex functional interdependencies across these levels are a fundamental barrier to distinguishing between the causes and effects of aging.

In this talk, I will discuss our development of systematic approaches that use population asynchrony to map multi-scale gene-regulatory networks in aging. We focus on a specific aging phenotype: the more than 70% of human lifespan variance that cannot be explained by heritable factors. In C. elegans, we discover that incomplete coordination between organs, rather than individual genes, is the largest source of inter-individual heterogeneity in aging. Screening for genes that influence inter-organ coordination during aging, we identify a diverse set of molecular mechanisms that contribute to age-associated heterogeneity in gene-expression, healthspan, and lifespan. In particular, we identify a set of pleiotropic genes whose perturbation, even late in life, dramatically reduces lifespan variance—providing new experimental entry points for studying how noise propagates across multicellular, aging systems.

CAROLINE UHLER

Broad Institute of MIT & Harvard

Causal Representation Learning: From Biomarkers to Mechanisms

An exciting opportunity at the intersection of the biomedical sciences and machine learning stems from the growing availability of large-scale multi-modal data (imaging-based and sequencing-based, observational and perturbational, at the single-cell level, tissue-level, and organism-level). Traditional representation learning methods, although often highly successful in predictive tasks, do not generally elucidate underlying causal mechanisms. I will present a statistical and computational framework for causal representation learning and its applications towards identifying novel disease biomarkers as well as inferring gene regulation in different disease contexts.

BARCELONA COLLABORATORIUM
FOR MODELLING AND PREDICTIVE BIOLOGY

NOELIA FERRUZ

CRG

Explainable AI for Protein Design

Protein design has the potential to solve myriad biotechnological challenges but has traditionally suffered from low success rates. In recent years, the field has witnessed an explosion of AI methods, including protein language models (pLMs), enabling the design of functional proteins at unprecedented rates. These models are end-to-end differentiable, complex nonlinear functions $f: X \to Y$ that map inputs to outputs, yet their mappings are essentially black boxes, making it difficult to explain why a model makes a particular decision. pLMs are no exception; their decision processes remain poorly understood. We recently reviewed eXplainable AI (XAI) techniques applicable to pLMs and categorized them by the input features they analyze. We further argue that these XAI methods play five roles: Evaluator, Multitasker, Coach, Engineer, and Teacher, with only the Evaluator widely adopted, while the Teacher role, which could yield biological insight, remains largely unused. Among these methods, we train sparse autoencoders to expose interpretable features and steer generation. In particular, we train SAEs on the ZymCTRL pLM and use α-amylase deep-mutational data to select activity-linked features; clamping beneficial features, ablating detrimental ones, and introducing a new MSA-steering strategy to raise predicted activity, with MSA-steering yielding the largest shift (median 1.995 vs. 1.045 base). Together, these results show how XAI can be used in the context of protein engineering, but we underscore the need for more efforts and methods to allow the interpretation of these models.

ELI WEINSTEIN

DTU

Causal Molecular Design

Advances in ML-driven molecular design hold great promise for human therapeutics. But existing tools focus on modeling laboratory proxies for therapeutic efficacy, such as structure or binding, which may not generalize to clinical outcomes. In this talk, we leverage natural experiments in humans to learn the effects of novel proteins on patient outcomes. We focus on T cell receptors (TCRs), using patient repertoire and clinical data. To correct for bias and confounding, we introduce novel causal machine learning methods, which can exploit V(D)J recombination as a source of randomization. The result is a neural network that estimates the effect of interventions on patients with a specific TCR sequence. Overall, our results bridge causal and molecular machine learning to develop a new approach to patient-centered protein design. This talk is based on joint work with David Blei and Elizabeth Wood.

FLASH TALK PROGRAM

Inferring causal, time-varying gene regulatory networks from single-cell temporal transcriptomic data

Tanmayee Narendra | University of Dundee

Trajectory-based causality analysis reveals bidirectional regulation between histone PTMs and gene expression

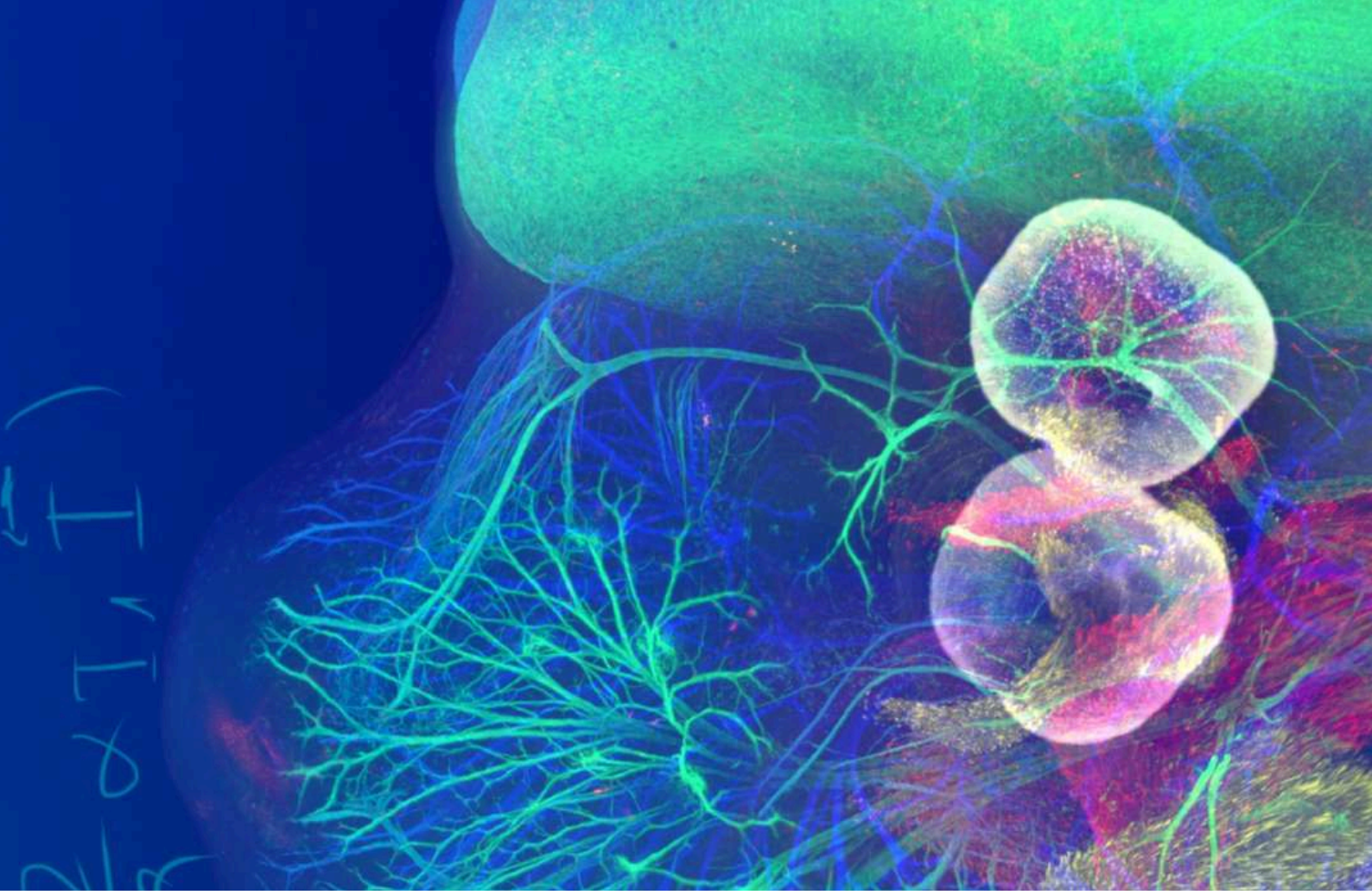
Vasilis F. Ntasis | Centre for Genomic Regulation (CRG)

Modeling the bacterial response to antibiotics using PINNs

Júlia Vicens Figueres | Universitat Pompeu Fabra (UPF)

Inferring causal gene regulatory relationships from time-resolved single-cell transcriptomics

Dimitris Volteras | Francis Crick Institute & Imperial College London



SELECTED POSTERS

BARCELONA COLLABORATORIUM
FOR MODELLING AND PREDICTIVE BIOLOGY

POSTERS

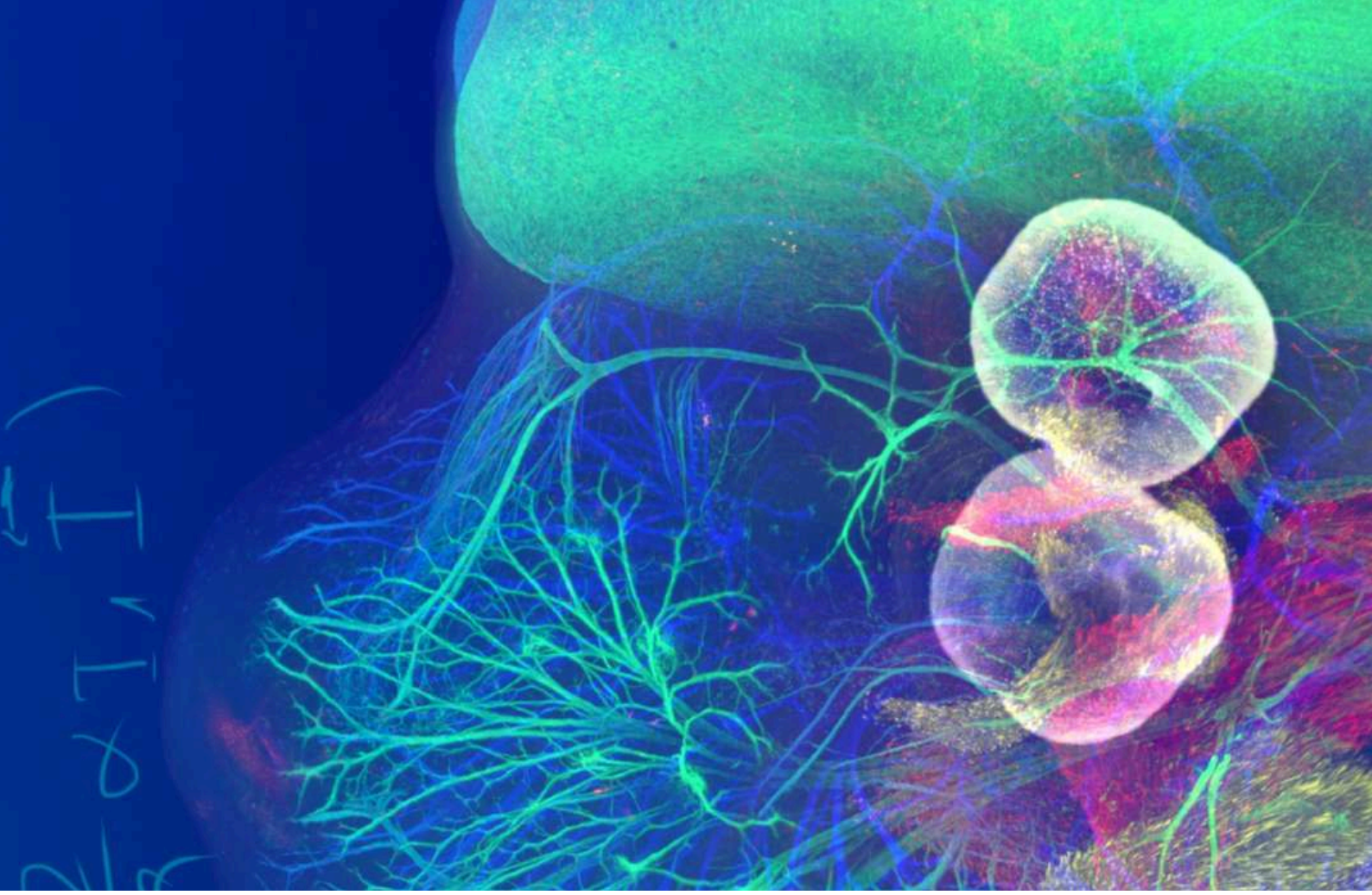
#	First name	Last name	Institute	Title
1	Judith	Ballmann	Goethe University Frankfurt	Rotational dynamics in pancreatic organoids: a theoretical framework towards cell-driven self-organization
2	José	Carbonell-Caballero	Barcelona Supercomputing Center	Towards a hybrid framework for inferring dynamic signalling models from RNA-Seq data
3	Jaime	Casado Garcia-Consuegra	IBEC	In Situ Tracking of Clonal Evolution and Phenotypic Heterogeneity in Tumors by Spatial Epitope Barcoding
4	Robert	Castelo	Universitat Pompeu Fabra	Interventional idlBNs in DAG-space
5	Dylan	Dalton Martínez	IRB	Unveiling Disease Connections through Molecular Signatures
6	Aurora	Desideri Perea	IRB Barcelona	Expanding the small molecule repertoire towards precision drugs through AI generative models
7	Matteo	Di Vincenzo	UPF - Melis	In Silico Evolution of Lentiviral Vectors via Direct Preference Optimization (DPO)
8	Juan	Fernandez Dourado	UPF	Bioprospecting, AI generation and characterization of mammalian hyper-active retroelements for RNA-based gene writing
9	Mirko	Francesconi	ENS-lyon	Integrative transcriptomic analysis reveals uncoupling in tissue-specific physiological ages between genetically identical individuals, across environments and perturbations
10	Mariona Montserrat	Fucho Rius	Universitat Politècnica de Catalunya	A Dynamical Systems Approach to Combined Bacteriophage–Antibiotic Therapies
11	Othmane	Hayoun Mya	Barcelona Supercomputing Center	Multi-Scale Network Analysis Reveals Sequence-Dependent Synergy in PI3K-MEK Inhibitor Combinations
12	Akke Mats	Houben	University of Barcelona	Circular causality in neural networks: uniting pattern formation and network dynamics
13	Kye	Hunter	CRG	Rugged yet accessible fitness landscapes: The 2-parameter Rough Mount Fuji
14	Rob	Jelier	KU Leuven	Causal inference by physical modeling: nematic structures in the C. elegans zygote

POSTERS

#	First name	Last name	Institute	Title
15	Maria	Kelly	Centre for Genomic Regulation	Evolution of Mutational Susceptibility to Disease
16	Tim	Liebisch	EMBL Barcelona	Data driven modelling of limb bud growth and morphogenesis
17	Gabriele	Malagoli	Helmholtz Munich	Geometry aware graph attention networks to explain single-cell chromatin state and gene expression
18	Pavan	Matani	Centre for Genomic Regulation	Phenotyping the social behaviour of mice using computer vision
19	Antoni	Matyjaszkiewicz	EMBL Barcelona	LimbNET: collaborative platform for simulating spatial patterns of gene networks in limb development
20	Mikel	Ocio-Moliner	Universitat de Barcelona	Burst Initiation Points displacement upon electrical stimulation
21	Elena	Pareja-Lorente	IRB Barcelona	Foundation models for predicting cellular responses to perturbations.
22	Matteo	Zambon	CRG	Decoding gene essentiality from cross-species transcriptional constraints
23	Charles	Pugh	CRG	Do in-silico methods for function optimization resemble evolutionary processes?
24	David	Ricote	Universisad Complutense de Madrid	Modeling multiple causality in extended heredity: the sisterhood of genotokens
25	Marco	Ruscone	Barcelona Supercomputing Center	Intelligent Tool Orchestration for Rapid Mechanistic Model Prototyping: MCP Servers as AI-Biology Interfaces
26	Kshitij	Sinha	CRG Barcelona	Structural Causal Modeling for Gene Regulatory Networks Using Dual Interventions in C. elegans
27	Oleksandra	Soldatkina	BSC	Multimodal data analysis reveals asynchronous aging dynamics across female reproductive system
28	Carlotta	Viana	CRG	Tracing the hidden signature: Deep Learning-based detection of leukemia through chromatin architecture in confocal and super-resolution imaging

POSTERS - Flash Talks

#	First name	Last name	Institute	Title
29	Tanmayee	Narendra	University of Dundee	Inferring causal, time-varying gene regulatory networks from single-cell temporal transcriptomic data
30	Vasilis	Ntasis	Centre for Genomic Regulation (CRG)	Trajectory-based causality analysis reveals bidirectional regulation between histone PTMs and gene expression
31	Júlia Vicens	Figueres	Universitat Pompeu Fabra (UPF)	Modeling the bacterial response to antibiotics using PINNs
32	Dimitris	Volteras	Francis Crick Institute & Imperial College	Inferring causal gene regulatory relationships from time-resolved single-cell transcriptomics



Event management

ANNA MARTA BORRELL PAGES (CRG)

BASTIEN DEBIÈVE (EMBL Barcelona)