Workshop Summary - Motivation & Themes

Since the last time that the LMRL workshop was held at NeurIPS 2022 (https://www.lmrl.org/), interest in representation learning for biology has surged, with new ideas challenging traditional approaches and sparking discussions on how best to capture the complexity of biological systems through machine learning. The availability of large-scale public DNA and RNA sequencing, protein sequences and 3D structures, mass spectrometry, and cell painting datasets (JUMP-CP, RxRx3, Human Cell Atlas) has fueled the development of numerous large-scale “foundation models” for biological data (Rozenblatt-Rosen et al. 2021; Fay et al. 2023; Chandrasekaran et al. 2023). These models aim to extract “meaningful” representations from noisy, raw and unstructured high-dimensional data to address a variety of biological questions.

In 2023 and 2024 alone, the community has produced a number of papers dedicated to crafting foundation models for biological *data* across multiple modalities, such as DNA and RNA sequencing, cell morphology images, digital pathology, protein sequences, and more (Theodoris et al. 2023; Cui et al. 2024; Rosen et al. 2023; Y. Chen and Zou 2024; Thapa et al. 2024; Nguyen et al. 2024; R. J. Chen et al. 2024, Ming Y. et al. 2024, Xu H. et al 2024). In some cases, techniques that were first developed for text and natural images have been successfully ported to biological data leading to maps of cellular morphological variation to modeling phenotypic (Kraus et al. 2024) and functional perturbations to uncover cell’s gene regulatory networks (Zhang et al. 2024). However, simple baselines have also been shown to outperform foundation models in settings such as 0-shot learning (Kedzierska et al. 2023), distribution shifts (Wenteler et al. 2024) and predicting the outcomes of unseen experiments (Kernfeld et al 2024), where we might have hoped that large-scale pre-training would lead to better performance.

Given this disconnect between modalities where large scale representation learning has led to significant improvements, and those where scale has not lived up to its promise, the AIxBio community has two important questions to answer: (i) what data, models and algorithms do we need to ensure that we *extract meaningful representations* (sufficient for their intended applications); and (ii) what are the *appropriate methods for evaluating the quality of these embeddings*, both in terms of the richness of information they capture, and their ability to generalize and improve performance on downstream tasks?

Recent studies have begun to explore these topics (Boiarsky et al. 2023; Kedzierska et al. 2023; Liu et al. 2024), and we believe that the early stage of this field presents a remarkable opportunity to foster discussion, collaboration, and insight sharing through our workshop on **“Learning Meaningful Representations of Life”**. Our agenda will encourage discussion both about new methods for representation learning in biology as well as biologically relevant & substantive evaluations to probe the generalization capabilities of the learned representations. Building upon the themes of previous years, the workshop will focus on multiple layers of biological information: genomes, molecules, cells, phenotype and beyond.

It is essential for such *“meaningful representations”* to not only generalize across modalities but also to capture biological information across different scales, from subcellular to multi-cellular and organism-wide processes. Harmonizing representations from molecules, proteins, cells, and tissues enables in-silico simulation of biological processes, interactions, and causal mechanisms, ultimately building towards a foundation model of AI-powered virtual cell (Bunne et al. 2024) , *i.e.* universal simulators of cellular function and behavior. We have made a concerted effort to ensure our speakers and panelists represent expertise in learning across scales -- Alex Rives on large language models for protein biology (Hayes et al. 2024), Emma Lundberg on subcellular protein localization via the Cell Atlas (Uhlén et al. 2015), Jean-Philippe Vert on single-cell genomics (Meng-Papaxanthos et al. 2023), Faisal Mahmood (R. J. Chen et al. 2024) on foundation models in computational pathology, and Mor Nitzan on cellular responses and interactions (Piran et al. 2024).

For the LMRL workshop at ICLR 2025, our objectives are (i) to convene those engaged in learning representations within and across different modalities of biological data, (ii) to discuss cutting-edge methods for assessing and measuring the significance of learned biological representations, (iii) to create a platform for developing open-source standardization of datasets and evaluation metrics for benchmarking new methods, and (iv) to envisage potential real-world problems that could be solved with improved strategies for learning meaningful representations of life. We believe that ICLR, as a premier AI conference with a strong emphasis on cutting-edge advances in representation learning, provides the ideal platform for uniting researchers at the forefront of AIxBio and fostering breakthroughs in how we understand and apply biological data representations to real-world challenges.

Invited Speakers

We have extended our invitations to 5 keynote speakers covering their specific research area, all of whom have **already accepted our invitation** as well as an **in-person attendance** of our proposed workshop**:**

- **Alexander Rives** (Assistant Professor | Massachusetts Institute of Technology (MIT), USA & Core Member | Broad Institute of MIT and Harvard, USA & Chief Scientist | EvolutionaryScale.ai)
  - Research Area & Workshop Fit: *Large language model-assisted novel protein design at evolutionary scale*
  - Bio: In his research, Alex is focused on AI for scientific understanding, discovery, and design for biology. He is widely acknowledged for previously leading Meta AI's protein folding team as a New York University graduate student, where he founded and led the EvolutionaryScale team that develops large language models for protein design modeling. To our workshop, Alex brings his valuable experience from an industrial and academic research setup, with a unique perspective on meaningfulness from an evolutionary angle.

- **Kevin Tsia** (Professor & Program Director | The University of Hong Kong, Hong Kong)
  - Research Area & Workshop Fit: *Developing dynamic optical tools for new frontiers in biomedicine & the clinic*
  - Bio: Kevin's lab is developing advanced optical technologies to solve critical problems in modern biomedicine and clinical practices, such as liquid biopsy, blood analysis and cancer screening. His biotechnology innovations present tools on how to dynamically represent fast-changing systems over time, such as high-speed optical methods for neural imaging. He uniquely combines microfluidics with bioinformatics techniques for interpretable and accurate detection of tumor-like cells in a label-free way.

- **Faisal Mahmood** (Associate Professor | Division of Medical Sciences, Harvard University, USA)
  - Research Area & Workshop Fit: *Multimodal and generative AI for cancer diagnosis & biomarker discovery*
  - Bio: Faisal's lab is pioneering the use of machine learning and medical image analysis to streamline cancer diagnostics and improve clinical outcomes. His research centers on developing AI-driven tools that assist pathologists, ensuring more consistent and accurate cancer diagnosis. Faisal also focuses on creating novel gen-AI algorithms that identify clinically significant biomarkers tied to patient responses to specific therapies. By integrating multimodal data, including imaging, genomics, and patient histories, his work advances precision medicine, offering insights that enhance the diagnosis, prognosis, and treatment of cancer.

- **Yixin Wang** (Assistant Professor | Michigan Institute for Data Science & University of Michigan, USA)
  - Research Area & Workshop Fit: *Developing causal & probabilistic understanding of biological data*
  - Bio: Yixin's research focuses on developing statistical and computational methods for analyzing biological data, with a particular interest in identifiable representation learning. Her work spans various areas, including causal inference, statistics, probabilistic modeling and latent variable models with applications to computational biology. Yixin's theoretical background will enhance the workshop by her contribution of her

causal definition of what it means for learnt representations to be meaningful.

- **Juan Caicedo** (Assistant Professor | Morgridge Institute & University of Wisconsin-Madison, USA)
  - Research Area & Workshop Fit: *Advancing the representation learning field towards biomedical imaging*
  - Bio: Juan's expertise lies in the fields of computer vision and machine learning for creating computational models of phenotypic variation using microscopy imaging and machine learning to observe the complexity of cellular structures and tissue organization. Juan will introduce how he uses machine learning to capture complex phenotypes in space and time, to allow for quantitative inference with complex phenotypes, and to enable precise and sensitive disentanglement of disease impact and treatment effects at the cellular level.

Invited Panelists

To foster thought-provoking discussions & insights from the field, we plan to host a panel discussion on the topic of *"Balancing Rigor, Interpretability, and Innovation: Desiderata for Biological Representation Learning,"* featuring 6 following speakers, all of whom have **already accepted our in-person invitation** to attend the panel * :

- **Emma Lundberg** (Professor | Stanford University, USA & KTH Royal Institute of Technology, Sweden)
  - Research Area & Workshop Fit: *Large-scale human atlas creation in spatial proteomics & cell biology*
  - Bio: Emma is a trained cell biologist and the Director of Cell Profiling at the Science for Life Laboratory. Her research combines spatial proteomics with cell biology, making use of an antibody-based approach to assess fundamental aspects of human biology at single-cell level. Her aim is to utilize AI modeling to understand how human proteins are distributed in time and space, how variations and deviations in localization can contribute to cell type specificity as well as disease phenotypes.

- **Tim Stuart** (Principal Scientist | Genome Institute of Singapore (GIS A-STAR), Singapore)
  - Research Area & Workshop Fit: *Integrative, multimodal and scalable single-cell analysis with omics data*
  - Bio: Through extensive clinical collaborations, Tim's group is applying cutting-edge methodologies to study how mutations in epigenomic regulators impact the state of human blood cells in leukemia, lymphoma, or myeloma. Passionate about bringing AI to patients, Tim develops and maintains specialized scientific software for the processing, analysis, interpretation, and visualization of single-cell genomic data.

- **Jean-Philippe Vert** (Chief Research & Development Officer | Owkin & Co-founder | Bioptimus, France)
  - Research Area & Workshop Fit: *Drug discovery & diagnostics via universal AI foundation models for biology*
  - Bio: Jean-Philippe is a prominent leader in the field of machine learning and computational biology, focused on transforming medicine through AI-powered solutions. He directs groundbreaking efforts to use AI for optimizing treatment discovery, particularly in cancer, by analyzing complex biological data to guide personalized therapies. Through the Bioptimus pioneering initiative, Jean-Philippe creates universal AI foundation models for biology designed to accelerate breakthroughs in biomedicine and drug development.

- **Mor Nitzan** (Assistant Professor | The Hebrew University of Jerusalem, Israel)
  - Research Area & Workshop Fit: *Single-cell network dynamics & cell interactions in collaborative ecosystems*
  - Bio: Mor's group is interested in how local changes, responses and interactions between cells lead to global coordinated effects, such as multicellular self-organization. She studies how to infer and design optimal interactions on single-cell level to reveal biological processes and contexts that cellular populations are driven by. Mor implements geometrical structures to determine the global states of (dys-)functioning tissues.

- **Theofanis Karaletsos** (Head of Artificial Intelligence for Science | Chan Zuckerberg Initiative (CZI), USA)
  - Research Area & Workshop Fit: *Accelerated development of virtual cell models at a non-profit organization*

  - Bio: Theo leads CZI's AI strategy for science, with deep focus on the organization's vision to develop virtual cell models, utilizing one of the largest high-performance AI computing systems for non-profit life science research. Scientifically, Theo is interested in laying the foundations for exploring the molecular underpinnings of human health and disease, enabled by scalable and uncertainty-aware AI for life science research.

- **Javier Alvarez-Valle** (Senior Director of Biomedical Imaging | Microsoft Health Futures, UK)
  - Research Area & Workshop Fit: *Multimodal AI for clinically relevant vision-language radiology applications*
  - Bio: Javier leads the Biomedical Imaging Team, spearheading projects focused on using AI to revolutionize healthcare. His work evolves around multimodal AI technology to assist radiologists, improving patient care and empowering healthcare professionals. Javier also drives an open-source deep learning toolkit InnerEye, designed to democratize medical imaging AI by providing tools for the automatic analysis of 3D medical images. His team's deployed innovations significantly accelerate radiotherapy planning for faster treatments.

* *Note: We recognize that travel logistics can be challenging, and while we are prioritizing in-person attendance from all panelists, we understand that confirmations may depend on availability closer to the date. Should any panelist face unforeseen travel complications, we will prioritize ensuring a robust in-person panel for the best experience possible for our attendees.*

Preliminary Agenda & Thematic Areas

By maintaining a streamlined program with **five keynote speakers** and subsequent **spotlight talks** highlighting the best posters per theme (chosen to match the keynote speakers' areas of research), we aim to keep the workshop focused and elevate the prestige of the spotlight sessions to showcase the most significant contributions within each area. Likewise, our 2-hours long **poster session** is aimed to ensure a high level of impact and engagement for all attendees, which will be furthermore enhanced via their questions to our **six invited panelists**. Moveover, three tiny paper submissions will be included to present their **tiny paper pitches.**

Table 1 | Preliminary workshop agenda.

Time	Duration*	Activity	Presenter	Theme
8:30 - 8:45	15 mins	Opening remarks	Lead organizers	Welcome + Sponsors Announcement
8:45 - 9:25	40 mins	Keynote 1	**Juan Caicedo**	Advancing the representation learning field towards biomedical imaging
9:25 - 9:40	15 mins	Spotlight 1		
9:40 - 9:55	15 mins	Break		
9:55 - 10:35	40 mins	Keynote 2	**Kevin Tsia**	Developing dynamic optical tools for new frontiers in biomedicine & the clinic
10:35 - 10:50	15 mins	Spotlight 2		
10:50 - 11:05	15 mins	Break		
11:05 - 11:45	40 mins	Keynote 3	**Faisal Mahmood**	Multimodal and generative AI for cancer diagnosis & biomarker discovery
11:45 - 12:00	3 x 5 mins	Tiny Paper Pitches		
12:00 - 14:30	120 mins	Lunch break + poster session		
14:30 - 15:10	40 mins	Keynote 4	**Yixin Wang**	Developing causal & probabilistic understanding of biological data
15:10 - 15:25	15 mins	Spotlight 3		
15:25 - 15:40	15 mins	Break		
15:40 - 16:20	40 mins	Keynote 5	**Alexander Rives**	Large language model-assisted novel protein design at evolutionary scale
16:20 - 16:35	15 mins	Spotlight 4		
16:35 - 16:45	10 mins	Break		

16:45 - 17:45	60 mins	Panel Discussion	**Emma Lundberg** **Tim Stuart** **Jean-Phillippe Vert** **Mor Nitzan** **Theofanis Karaletsos** **Javier Alvarez-Valle**	*“Balancing Rigor, Interpretability, and Innovation: Desiderata for Biological Representation Learning”*
17:45 - 18:00	15 mins	Closing remarks	Sponsors’ reps	Award ceremony

***** All time slots include the time for questions from the audience.

Paper Tracks

We invite topics ranging across different areas of biology (AI for omics analysis, protein design, cell profiling, plant biology, microbiome research, tissue and systems biology, clinical and/or diagnostic tool development, etc.), and invite submissions from other sciences. We propose this workshop primarily to attract submissions not native to ICLR and **bring people from science to an AI-oriented conference**. Therefore, a misfit of the biological topic won’t in itself stand as a basis for rejection of the workshop submission. However, to align all submissions to the topic of the workshop and to encourage deeper reflection on the relevance and impact of their work, we require each submission to include a brief **"meaningfulness statement"**. This brief statement (100 words max) should address the question of *"How does this work help us learn meaningful representations of life?"* and it is aimed to ensure that the contributors thoughtfully consider the significance, impact and relevance of their work to the workshop. Importantly, this statement will not be visible to reviewers nor influence the acceptance decision, but it serves to discourage the resubmission of previously rejected papers without proper refinement.

We recognize that workshops often attract work-in-progress submissions that may not be ready for a full conference paper. To accommodate varying levels of progress, we offer **two submission tracks**:

Full paper track | We invite submissions of 4–8 pages, excluding references. These papers should outline the biological problem, the AI methods used, implementation details, benchmarking metrics, and comparisons to related work. Based on past LMRL experience, we expect over 150 submissions and plan to accommodate up to 80 posters. The top three ranked papers will be selected for a spotlight in the main program as "Contributed Talks", and all full papers will compete for various awards.

Tiny paper track | We also invite 1–2 page submissions to encourage participation from under-represented and under-funded communities. These "extended abstracts" should focus on the methodology or novel approach to solving a biological problem, even if results are not yet available. Lack of results will not be grounds for rejection, but the proposed methodology must be well-defined and framed within the biological context. We expect 50+ submissions and plan to accept 20 "tiny posters" (smaller format, *e.g.* A1). Additionally, three "Tiny Paper Pitches" will be selected for short spotlight talks in the main program.

Table 2 | Comparison of the ‘full’ and ‘tiny’ paper tracks.

Criteria	**Item**	**Paper Track**	
		Full	**Tiny**
Submission	Meaningfulness statement	Required	Required
	Biological areas	Any	Any
	Page limit	4-8 pages	1-2 pages
	Paper structure (comparable to)	Conference submission	Extended abstract

	Expected no. of submissions	Over 150 papers	Over 50 papers
	Expected no. of acceptances	~80 posters	~20 posters
Reviewing	Reviewer guidance	Stricter	Relaxed
	Reviewer nomination	Required	Optional
	Double-blind review	Yes	Yes
On the day	Number of spotlights	4 talks	3 talks
	Spotlight talk duration	15 mins	5 mins
	Poster size	A0	A1

Both paper tracks will share the same submission deadline and will undergo review by a common program committee. The key differences are: (i) full paper submissions will be required to nominate at least one reviewer, while this will be optional for tiny paper submissions, and (ii) review criteria will be slightly less stringent for the smaller, work-in-progress papers. We are drawing inspiration from the Women in Machine Learning (WiML) workshop guidelines at NeurIPS 2024. Reviewers will be informed of the track under which each submission was received and, in special cases, may recommend moving a submission between tracks. All submissions are **by default non-archival**, and papers will be posted online unless explicitly requested otherwise. In exceptional circumstances where authors are unable to attend the workshop in person, **pre-recorded talks** will be enabled for sharing with the workshop audience.

Submission Timeline

To ensure timely submissions and decision communication to the authors, we have devised a preliminary timeline shared for both full & tiny paper tracks (**recommended deadlines** by the ICLR workshop guidelines):

- Workshop Acceptance Notification: 2 December 2024
- Social Media Campaign to Announce Workshop: 8 December 2024
- Paper Track Criteria Published (on the website): 16 December 2024
- Social Media Campaign to Invite Submissions: 13 January 2025
- Paper Submission Deadline: **3 February 2025, 11.59pm AoE**
- Reviewer Assignment Deadline: 7 February 2025
- Review Deadline: 28th February 2025
- Decision Notification + Spotlight Talk Invitations: **5 March 2025, 11.59pm AoE**
- Camera-Ready Deadline: 12 March 2025
- Electronic Poster Upload Deadline: 19 March 2025
- Import Workshop Program and Accepted Papers to iclr.cc: **27 March 2025, 11.59pm AoE**

Award Sponsors

Valence labs (https://www.valencelabs.com/) have already confirmed their continued excitement for the workshop and offered a financial sponsorship to “Outstanding Poster Award” winners, which we envisage to present in several categories, e.g. “Programme Committee Award“ and “Attendees' Choice Award” received by popular vote. We have also reached out to local AI research groups at the National University of Singapore (NUS) and the Genome Institute of Singapore (GIS) to explore potential collaborations and ways to increase local participation.

Workshop Attendees

Building on the success of the past LMRL workshops and the growing interest in AIxBio sessions at major AI conferences, we anticipate approximately 300 attendees at peak times. While tracking exact numbers at an open venue is challenging, we estimate that up to 500 individuals could benefit from our diverse program throughout the day, including poster presenters. We are committed to enabling remote participation, with plans to (i) livestream all talks and sessions for virtual attendees, and (ii) record the presentations for later access via the ICLR platform and/or our custom LMRL website (https://www.lmrl.org/). This option will be particularly valuable for participants facing exceptional circumstances, such as visa issues or inability to travel to Singapore due to underfunding. Additionally, we may consider documenting the outcomes of the workshop in a correspondence letter or review article, or a less formal series of blog posts to be communicated over social media.

Related Workshops

This year's edition of the workshop draws inspiration from the previous NeurIPS LMRL workshops held in 2021 and 2022 and more recent, similarly themed workshops including the 2023 ICML Workshop of Computational Biology and 2024 ICLR MLGenX workshop. Unlike other ICLR workshops that focus on specific AI methodologies or challenges (e.g. causal representation learning; distribution shift; zero-shot/few-shot learning for foundation models), LMRL distinguishes itself by uniquely bringing together scientists using diverse AI methodologies to share a common goal of building virtual representations of biological systems.

Our emphasis on the challenges and opportunities in biology—such as interpretability, limited training data, real-world validation, and expensive label acquisition—sets this workshop apart. Compared to previous ICLR workshops with a biology focus, *e.g.* Generative and Experimental Perspectives for Biomolecular Design (GemBio) or Machine Learning for Genomics Explorations (MLGenX), the LMRL workshop (i) has historically attracted more researchers studying genomics and cellular phenotypes beyond structural biology, and (ii) highlights the importance of learning meaningful representations that provide biological insights rather than just improving task performance.

While we believe our proposal stands out, we welcome feedback from attendees over past years and recognize that participants may wish to attend sessions across multiple workshops. To facilitate this, we kindly request that the ICLR workshop committee either schedule our workshop on a separate day from other biology-themed workshops, or co-locate related sessions in nearby rooms, if venue capacity permits. This arrangement will allow participants to move between rooms based on their interests and expertise, enhancing their engagement with concurrent workshops.

Organizing Committee

Kristina Ulicna (kristina@valencelabs.com) is a Research Scientist at Valence Labs, powered by Recursion Pharma, specializing in **AI-driven explainable single-cell representations from latent image embeddings**. Previously, she was a post-doctoral research associate at The Alan Turing Institute, working on spatial patterning in imaging data. She holds a PhD in time-lapse microscopy image analysis from UCL under Alan Lowe's supervision, a BSc in Biomedical Science from King's College London, and interned at MIT, Cancer Research UK, and Microsoft Research Cambridge. *Previous organizing experience:* Kristina served as a (super)-volunteer to help co-organize the Women in Machine Learning (WiML) workshop at both ICML 2023 and NeurIPS 2023.

Rebecca Boiarsky (rboiar@mit.edu) is a PhD candidate at MIT, focusing on **representation learning for single-cell RNA-sequencing data with an emphasis on applications in cancer and translational medicine**. She is co-advised by David Sontag at MIT CSAIL and Gad Getz at the Broad Institute of MIT and Harvard. Before pursuing her Ph.D., she worked in the Molecular Profiling department at Regeneron Pharmaceuticals. Rebecca holds a Master's degree in Biomedical Engineering from Columbia University and a Bachelor's degree in Physics from Yeshiva University. *Previous organizing experience:* Rebecca co-organised the NeurIPS LMRL in 2022.

Eeshaan Jain (eeshaan.jain@epfl.ch) is a first year PhD student at the Department of Computer Science, EPFL, advised by Charlotte Bunne. Prior to joining EPFL, he obtained his Bachelor degree in Electrical Engineering and Masters degree in AI from IIT Bombay. His current research focuses on **multi-scale representation learning for biomedical datasets and drug discovery**, building on his past work experience from Google Research & Sony.

Till Richter (till.richter@helmholtz-munich.de) is a final year PhD candidate at Technische Universität München (TUM) and Helmholtz Munich, advised by Fabian Theis and Niki Kilbertus. His research focuses on **large-scale and mechanistic models for machine learning in single-cell genomics**. Till holds an MSc degree in Robotics, Cognition, Intelligence from TUM and a BSc degree in Industrial Engineering from Hannover University, Germany. *Previous organizing experience:* Till co-organised the Explainable AI workshop at ECCB in 2024.

Giovanni Palla (giovanni.palla@helmholtz-munich.de) is an AI/ML Research Scientist led by Theofanis Karaletsos at the Chan Zuckerberg Initiative (CZI), California, USA. He obtained his PhD at the Munich School for Data Science in Fabian Theis's lab, at Helmholtz Munich, Germany. He holds a BSc from the University of Trento and a MSc from Utrecht University. His research interests include **AI applications to spatial biology, with a focus on multimodal and generative modeling**. He co-founded and contributes to scverse, and is an Add-on Fellow of the Joachim Herz Stiftung. *Previous organizing experience:* scverse hackathons across Austria, UK & Germany.

Jason Hartford (jason@valencelabs.com) is Dame Kathleen Ollerenshaw Fellow at the University of Manchester and Research Unit Lead and Staff Research Scientist at Valence Labs. Previously, he was a postdoctoral fellow with Yoshua Bengio at Mila & Université de Montréal. Before joining Mila, he completed his PhD at the University of British Columbia with Kevin Leyton-Brown. His research focuses on **developing new techniques for causal representation learning and causal inference**. *Previous organizing experience:* Jason is a co-organizer of the Causality, Abstraction, Representation, and Extrapolation (CARE) reading groups as a virtual seminar series.

Oren Kraus (oren.kraus@recursion.com) is an Associate Director of Machine Learning at Recursion Pharma, where his team **develops foundation models for cell imaging and gene expression data**. Before joining Recursion, Oren founded Phenomic AI, a startup leveraging computer vision and phenotypic screening to develop cancer therapeutics. During his PhD at the University of Toronto, Oren developed deep learning models for cell-microscopy data and interned at Apple and Borealis AI. He's passionate about building machine learning models to decode biology. *Previous organizing experience:* Oren co-organised the CytoData Conference in 2018.

Aleksandrina Goeva (aleksandrina.goeva@utoronto.ca) is an Assistant Professor at the Donnelly Centre and a Faculty Affiliate of the Vector Institute at the University of Toronto. Her research group **develops machine learning models to define principles of intercellular interactions to improve human health**. Aleks was a post-doc in the Macosko Lab at the Broad Institute and got her PhD in Statistics at Boston University with Henry Lam and Eric Kolaczyk. *Previous organizing experience:* Aleks co-organised the latest editions of NeurIPS LMRL in 2021 and 2022 and the Models, Inferences & Algorithms Initiative at the Broad Institute 2017-2024.

Charlotte Bunne (charlotte.bunne@epfl.ch) is an Assistant Professor at EPFL in the Computer Science and Life Sciences Department, leading an AI in Molecular Medicine (AIMM) group. Her research aims to **advance personalized medicine by utilizing machine learning and large-scale biomedical data**. Before, she was a PostDoc at Genentech and Stanford and completed a PhD in Computer Science at ETH Zurich working with Andreas Krause and Marco Cuturi. During her graduate studies, she was a visiting researcher at the Broad Institute of MIT and Harvard hosted by Anne Carpenter and Shantanu Singh and worked with Stefanie Jegelka at MIT. *Previous organizing experience:* Charlotte co-organized the Optimal Transport and Machine Learning Workshop at NeurIPS 2021 & 2023, the Diffusion Model Workshop at NeurIPS 2023, and the New Frontiers in Learning, Control, and Dynamical Systems Workshop at ICML 2023, and the MLGenX Workshop at ICLR 2024.

Fabian Theis (fabian.theis@helmholtz-munich.de) is the Director of the Computational Health Center and the Institute for Computational Biology at Helmholtz Munich & Professor at TUM Mathematics and Life Sciences, Germany. He uses **AI to study human cells, focusing on heterogeneities of cellular interactions and disease mechanisms via integrating "omics" data into systems medicine approaches**. With two Master's degrees in Mathematics and Physics and doctorates in Physics and Computer Science, Fabian has led research groups the Max Planck Institute and Helmholtz Munich. He also chairs the Mathematical Modeling of Biological Systems program at the Technical University of Munich and coordinates initiatives at HelmholtzAI. *Previous organizing experience:* Fabian co-organized multiple leading events such as 2024 ICLR MLGenX, Workshop on Single Cell Genomics meets Data Science (2022) and Workshop on Computational Single Cell Genomics (2019).

Committee Structures

Our **Organizing Committee** (OC) is intentionally larger to ensure a diverse representation of institutions, geographies, affiliations, seniority, and gender balance. This approach comprehensively covers many relevant research topics, creating an inclusive platform for attendees to submit their papers. By embracing diversity within the OC, we aim to enhance the experience for all participants. Our team has extensive experience to this workshop, having previously served as invited speakers, poster award winners, and organizers of similar workshops at ICLR, NeurIPS, and other major AIxBio conferences. Importantly, our entire OC is dedicated solely to proposing and organizing this workshop for ICLR 2025, reflecting our full commitment to this event.

Our **Programme Committee** (PC) currently consists of over 50 members from more than 15 institutions worldwide, recruited through our extensive connections in the field. Given LMRL's past successes and the rising popularity of ICLR workshops, excitement about our proposal is growing within our network. As the PC member list (see **Table 3** below) continues to expand, we will actively manage volunteer involvement to ensure broad topic coverage and a range of reviewer seniority for each submission. Each full paper track submission will require the nomination of *at least* one reviewer, allowing our PC to focus on reviewing submissions for the tiny paper track, which demands less time and seniority.

With the addition of nominated reviewers, our PC structure will ensure timely, comprehensive, and thoughtful reviews, with a maximum of five submissions per reviewer and a minimum of two reviews per submission in either paper track. To manage conflicts of interest, our extensive PC will guarantee that no OC member is involved in writing or curating the paper review process. We will implement a double-blind review policy and manage conflicts using the standard email domain matching system through OpenReview.

Diversity Statement

Diverse Committees Structure | Diversity is essential to our organizing team, fostering innovation, creativity, inclusivity, and cross-talk between academia and industry. We take pride in establishing a **gender-balanced** organizing committee, complemented by a diverse lineup of speakers and panelists for our workshop. Our team includes individuals across **various career stages**, from PhD students in their first to final years, to Assistant, Associate and Full Professors actively engaged in academia and university environments, as well as early-career Research Scientists and Division Directors in **research-driven industries**. With 10 organizers from **8 affiliated institutions** spanning 5 distinct countries, we ensure a broad range of perspectives. Moreover, the scale and richness of our programme committee's affiliations and career stages will help us bring together a diverse community and provide insightful, domain-specific reviews.

Diverse Speaker & Panelist Line-Up | For this workshop, we have curated an exceptionally diverse lineup of researchers who will share their insights on a **wide array of AI research topics**. Our speakers and panelists will cover methodologies such as large language models (LLMs), protein design, disentanglement, clinical applications, bioengineering tool development, spatial and temporal imaging, cell phenotyping and morphological profiling, causality, statistical modeling, building large-scale datasets of subcellular information, protein localization tasks, citizen-driven data labeling and annotations, (epi-)genomics research in clinical settings, single-cell approaches, uncertainty quantification, and abstractions of the virtual cell. We are proud to have created **gender-balanced** and **ethnically diverse lineups** with a mix of seniority (principal investigators to industry executives) and **institutional diversity**, encompassing industrial (budding biotech companies to established technological giants) to academic and non-profit affiliations (universities and independent institutes). Prioritizing **geographical diversity**, we have successfully nominated speakers from institutions across the US, Europe, and the Middle East. We also recognize the unique opportunity presented by the conference in Singapore to involve **local researchers** from Southeast Asia and Singapore, who are at the forefront of the meaningful AI representation learning community.

Diverse Authors & Poster Presenters | We remain committed to diversity, equity, and inclusion, believing that our workshop represents a significant opportunity for PhD candidates, post-doctoral researchers, and industry professionals to submit their work with the intention to (i) receive high-quality feedback, (ii) present at an event featuring world-renowned speakers and attendees, and (iii) contribute to discussions on advancing the field collectively. To additionally encourage underrepresented groups to submit their work to either paper track, we will advertise through affinity group email lists (WiML, Black in AI, MLDS-Africa, etc.) to foster a truly inclusive environment that embraces a variety of perspectives. Additionally, we plan to reach a broad range of researchers through a coordinated social media campaign to (i) announce workshop acceptance and (ii) invite submissions.

Diverse Attendees & Representatives | We believe in fostering meaningful networking and collaboration opportunities throughout the workshop, aiming for mutual engagement rather than a one-sided experience. It is essential for us to encourage poster and spotlight presenters to actively engage with established leaders and innovators in the field, ensuring their contributions receive ample attention and constructive feedback. To this end, we are inviting prominent industry leaders to join us for the entire day of the workshop. We will finalize these nominations as the date approaches, recognizing that travel arrangements for these attendees may not yet be fully confirmed. In the meantime, we are actively generating support and excitement for this workshop. Our outreach efforts will include a social media campaign aimed at attracting high-profile individuals from industrial research institutions to engage with our poster presenters. We understand that not all poster presenters may be pursuing an academic career; therefore, providing opportunities to connect with industry professionals is crucial for fostering relevant networking that meets their diverse interests and career aspirations.

Table 3 | Confirmed Programme Committee Members.

Name	Affiliation
Xinming Tu	University of Washington
Kexin Huang	Stanford University
Yanay Rosen	Stanford University
Kasia Kedzierska	Wellcome Trust Centre
Michael Vinyard	Broad Institute of MIT & Harvard
Ruitong Li	Broad Institute of MIT & Harvard
Alejandro Tejada-Lapuerta	Hemholtz Munich
Anna Schaar	Bioptimus
Berton Earnshaw	Valence Labs
Michael Craig	Valence Labs

ICLR 2025 Proposed Workshop: Learning Meaningful Representations of Life (LMRL)

Cristian Gabelini	Valence Labs
Frederick Wenkel	Valence Labs
Dominique Beaini	Valence Labs
Shawn Whitfield	Valence Labs
Hanene Benyedder	Valence Labs
Semih Canturk	Valence Labs
Konstantin Donhauser	Valence Labs
Ihab Bendidi	Ecole Normale Superieure PSL
Auguste Genovesio	Ecole Normale Superieure PSL
Thouis “Ray” Jones	Broad Institute
Elizabeth Wood	Jura Bio
Johann Wenckstern	EPFL
Stefan Stark	ETH Zurich
Frederike Lübeck	ETH Zurich
Vignesh Ram Somnath	ETH Zurich
César Miguel Valdez Córdova	Mila
Matthew Scicluna	Mila
Joseph Viviano	Mila
Velina Kozareva	MIT
Delaram Pouyabahar	University of Toronto
Duncan Forster	University of Toronto
Sam Cooper	Phenomic AI
Elias Williams	Phenomic AI
Conor Tilinghast	Recursion
Safiye Celik	Recursion
Youssef Barhomi	Recursion
Stephen MacKinnon	Recursion
Kyle Hansen	Recursion
Ben Fogelson	Recursion
John Urbanik	Recursion
Peter McLean	Recursion
Marissa Saunders	Recursion
Nathan Lazar	Recursion
Saber Saberian	Recursion
Grant Watson	Recursion
Rory Blucher	Recursion
Jay Huang	Recursion
Arin Minasian	Recursion
Kian Kenyon-Dean	Recursion

References

Boiarsky, Rebecca, Nalini Singh, Alejandro Buendia, Gad Getz, and David Sontag. (2023) “A Deep Dive into Single-Cell RNA Sequencing Foundation Models.” *bioRxiv*. https://doi.org/10.1101/2023.10.19.563100.

Bunne, Charlotte, Yusuf Roohani, Yanay Rosen, Ankit Gupta, Xikun Zhang, Marcel Roed, Theo Alexandrov et al. (2024) "How to Build the Virtual Cell with Artificial Intelligence: Priorities and Opportunities." *arXiv preprint arXiv:2409.11654*.

Chandrasekaran, Srinivas Niranj, Jeanelle Ackerman, Eric Alix, D. Michael Ando, John Arevalo, Melissa Bennion, Nicolas Boisseau, et al. (2023) “JUMP Cell Painting Dataset: Morphological Impact of 136,000 Chemical and Genetic Perturbations.” *bioRxiv*. https://doi.org/10.1101/2023.03.23.534023.

Chen, Richard J., Tong Ding, Ming Y. Lu, Drew F. K. Williamson, Guillaume Jaume, Andrew H. Song, Bowen Chen, et al. (2024) “Towards a General-Purpose Foundation Model for Computational Pathology.” *Nature Medicine* 30 (3): 850–62.

Chen, Yiqun, and James Zou. (2024) “GenePT: A Simple But Effective Foundation Model for Genes and Cells Built From ChatGPT.” *bioRxiv : The Preprint Server for Biology*, March. https://doi.org/10.1101/2023.10.16.562533.

Cui, Haotian, Chloe Wang, Hassaan Maan, Kuan Pang, Fengning Luo, Nan Duan, and Bo Wang. (2024) “scGPT: Toward Building a Foundation Model for Single-Cell Multi-Omics Using Generative AI.” *Nature Methods*, February. https://doi.org/10.1038/s41592-024-02201-0.

Fay, Marta M., Oren Kraus, Mason Victors, Lakshmanan Arumugam, Kamal Vuggumudi, John Urbanik, Kyle Hansen, et al. (2023) “RxRx3: Phenomics Map of Biology.” *bioRxiv*. https://doi.org/10.1101/2023.02.07.527350.

Hayes, Tomas, Roshan Rao, Halil Akin, Nicholas J. Sofroniew, Deniz Oktay, Zeming Lin, Robert Verkuil et al. (2024) "Simulating 500 million years of evolution with a language model." *bioRxiv*: 2024-07.

Kedzierska, Kasia Z., Lorin Crawford, Ava P. Amini, and Alex X. Lu. (2023) “Assessing the Limits of Zero-Shot Foundation Models in Single-Cell Biology.” *bioRxiv*. https://doi.org/10.1101/2023.10.16.561085.

Kraus, Oren, Kian Kenyon-Dean, Saber Saberian, Maryam Fallah, Peter McLean et al. (2024) “Masked Autoencoders for Microscopy are Scalable Learners of Cellular Biology”. *arXiv:* 2404.10242, cs.CV https://arxiv.org/abs/2404.10242

Liu, Tianyu, Kexing Li, Yuge Wang, Hongyu Li, and Hongyu Zhao. (2024) “Evaluating the Utilities of Foundation Models in Single-Cell Data Analysis.” *bioRxiv*. https://doi.org/10.1101/2023.09.08.555192.

Lu, Ming Y., Bowen Chen, Drew FK Williamson, Richard J. Chen, Ivy Liang, Tong Ding, Guillaume Jaume et al. "A visual-language foundation model for computational pathology." *Nature Medicine* 30, no. 3 (2024): 863-874.

Meng-Papaxanthos, Laetitia, Ran Zhang, Gang Li, Marco Cuturi, William Stafford Noble, and Jean-Philippe Vert. (2023) "LSMMD-MA: scaling multimodal data integration for single-cell genomics data analysis." *Bioinformatics* 39, no. 7: btad420.

Nguyen, E., M. Poli, M. G. Durrant, A. W. Thomas, and B. Kang. (2024) “Sequence Modeling and Design from Molecular to Genome Scale with Evo.” *bioRxiv*. https://www.biorxiv.org/content/10.1101/2024.02.27.582234

Piran, Zoe, and Mor Nitzan. "SiFT: uncovering hidden biological processes by probabilistic filtering of single-cell data." *Nature Communications* 15, no. 1 (2024): 760.

Rosen, Yanay, Yusuf Roohani, Ayush Agarwal, Leon Samotorčan, Tabula Sapiens Consortium, Stephen R. Quake, and Jure Leskovec. (2023) “Universal Cell Embeddings: A Foundation Model for Cell Biology.” *bioRxiv*. https://doi.org/10.1101/2023.11.28.568918.

Rozenblatt-Rosen, Orit, Jay W. Shin, Jennifer E. Rood, Anna Hupalowska, Human Cell Atlas Standards and Technology Working Group, Aviv Regev, and Holger Heyn. (2021) “Building a High-Quality Human Cell Atlas.” *Nature Biotechnology* 39 (2): 149–53.

Thapa, Rahul, Bryan He, Magnus Ruud Kjaer, Hyatt Moore, Gauri Ganjoo, Emmanuel Mignot, and James

Zou. (2024) “SleepFM: Multi-Modal Representation Learning for Sleep across Brain Activity, ECG and Respiratory Signals,” May. http://arxiv.org/abs/2405.17766.

Theodoris, Christina V., Ling Xiao, Anant Chopra, Mark D. Chaffin, Zeina R. Al Sayed, Matthew C. Hill, Helene Mantineo, et al. (2023) “Transfer Learning Enables Predictions in Network Biology.” *Nature* 618 (7965): 616–24.

Uhlén, Mathias, Linn Fagerberg, Björn M. Hallström, Cecilia Lindskog, Per Oksvold, Adil Mardinoglu, Åsa Sivertsson et al. (2015) "Tissue-based map of the human proteome." *Science* 347, no. 6220: 1260419.

Xu, Hanwen, Naoto Usuyama, Jaspreet Bagga, Sheng Zhang, Rajesh Rao, Tristan Naumann, Cliff Wong et al. "A whole-slide foundation model for digital pathology from real-world data." *Nature* (2024): 1-8.

Zhang, Xinyi, Saradha Venkatachalapathy, Daniel Paysan, Paulina Schaerer, Claudio Tripodo, Caroline Uhler, GV Shivashankar (2024) “Unsupervised representation learning of chromatin images identifies changes in cell state and tissue organization in DCIS” *Nat Commun* 15, 6112 (2024). https://doi.org/10.1038/s41467-024-50285-1