ICLR 2025 Workshop Proposal
2nd Machine Learning for Genomics Explorations (MLGenX)

Tagline. Bringing together communities at the intersection of machine learning and genomics to discuss areas of interaction and explore future research possibilities. This year, we are introducing a special track dedicated to empowering biomedical discovery through AI agents and foundation models.

Abstract. Our limited understanding of the biological mechanisms underlying diseases remains a critical bottleneck in drug discovery. As a result, we often lack insights into why patients develop specific conditions, leading to the failure of many drug candidates in clinical trials. Recent advancements in genomics platforms and the emergence of diverse omics datasets have sparked increasing interest in this field. The primary objective of this workshop is to bridge the gap between machine learning and genomics, emphasizing target identification and emerging drug modalities such as gene and cell therapies and RNA-based drugs. By fostering interdisciplinary collaboration, we aim to advance the integration of these disciplines and accelerate innovation in drug discovery.

Format. In-Person[1].

1 Workshop Motivation and Description

The main objective of this workshop is to bridge the gap between machine learning (ML) and (functional) genomics, focusing on the new era of biology and drug discovery. We aim to explore the complex challenges of modern drug development, particularly in identifying biological targets that play a critical role in modulating diseases. This exploration has the potential to pave the way for impactful machine learning applications that can accelerate the development of new therapeutics, including cell therapy (Amini et al., 2022), gene therapy (Bulaklak & Gersbach, 2020), and RNA/DNA vaccines (Dailey et al., 2023).

Building on these fundamental problems, the community recognizes that traditional approaches to foundation models, such as large language models (LLMs), may limit the ability to generate novel biological hypotheses (Gao et al., 2024). Unlike generating text, which adheres to semantic and syntactic rules, creating new scientific hypotheses and automated wet lab experiments requires a grounding in reasoning, planning, and scientific knowledge (Wu et al., 2024). Therefore, there is a need to explore and discuss the potential of realizing "AI scientists" as hybrid systems—AI agents enhanced by human oversight, LLMs, machine learning models, and experimental tools (Huang et al., 2024). If successful, these AI agents in biology should be capable of formulating biomedical hypotheses, critically evaluating them, characterizing their uncertainty, and adapting to new biological insights (Richard et al., 2024). This adaptability will ensure that AI agents remain relevant in the face of rapidly evolving biological data while balancing the incorporation of new findings with the retention of established knowledge (Chaves et al., 2024).

Through this workshop, participants will gain valuable insights into the synergies between ML and genomics research, and help to refine the next-generation of applied and theoretical ML methods for target identification.

---

[1] The talks and panel discussion will be live on SlideLive.

1.1 Scopes

This year, the workshop will feature three distinct tracks designed to welcome a diverse array of researchers in the field of machine learning and biology: the **Main Track** including application and ML topics, the **Special Track on Foundation Models and Agentic AI**, and the **Tiny Papers Track**. By providing dedicated spaces for applied ML in biology, machine learning experts, and beginner researchers, we aim to foster collaboration and knowledge exchange across various levels of expertise and research interests.

1.1.1 Main Track

Machine Learning. The distinctive challenges posed by high-throughput omics data necessitate both the development of innovative deep learning architectures and present an intricate landscape for established machine learning domains, including interpretability, causality, generative modeling, and representation learning.

Applications. Topics in genomics that are essential for tackling the unique challenges presented in target identification:

- **Design of regulatory sequence elements:** ML for DNA, RNA, and Cell and Gene Therapeutics will be explored, leveraging cutting-edge generative models and advanced optimization techniques to optimize biological sequences:
  a. Optimizing UTR/codons to enhance translational efficiency for mRNA vaccines
  b. Multi-omics-based sequence design
  c. Modeling long-range genomic sequence interactions
  d. Tissue/cell-type specific sequence design
  e. AI-based CRISPR design
- **Inferring cellular communication via cell states and organization in tissues:** Causal representation learning to model cell states and cellular communities.
  a. Multi-omics data integration (single cell, spatial transcriptomics)
  b. Cell-cell interactions inference
  c. Mechanistic modeling of cells in their context to infer cellular function
- **Perturbative biology:** Interpretable models to understand cellular responses to perturbations.
  a. Translating genetic perturbations to understandable and actionable molecular changes
  b. Integrating multimodal perturbation readouts (transcriptomic and phenotypic) to better characterize the broader molecular effects
  c. Generalizability of perturbation predictive models across cell lines and cellular contexts
- **Dynamic system simulations:** Modeling and simulating the behavior of biological systems over time to gain early clinical insights and understand disease mechanisms.
  a. Tissue-scale and multicellular perturbation simulations
  b. digital twins in clinical trial design
  c. Causal discovery for treatment optimization
  d. Longitudinal modeling of treatment effects

1.1.2 Special Track on Foundation Models and Agentic AI

This track focuses on foundation models, i.e. large-scale predictive or generative models trained on extensive datasets, and agentic AI systems for genomics. The topics include:

- Pre-training multi-omics models
- Synthetic data generation and data quality for pre-training, fine-tuning and instruction tuning
- Fine-tuning (SFT, RLHF, RL with lab feedback, ...) on novel tasks
- In-context learning with large-context models
- Reasoning through prompt engineering or architectural design
- Interpretability and uncertainty quantification
- Knowledge retrieval (RAG, knowledge graph, ...)
- Efficient interactive system designs (agents, humans, and biological tools)
- Training/fine-tuning LLM-powered design and planning engine

1.1.3 Tiny papers Track

The tiny papers track aims to provide diverse entry points to research, offering approachable avenues for beginners to engage with the MLGenX community. This track celebrates intermediate breakthroughs in machine learning and promotes efficient dissemination of ideas, findings, and opinions in a concise, two-page format.

1.2 Tentative Schedule

The workshop will feature four invited talks from industry and academic leaders, 30 minutes each (25 minutes presentation, 5 minutes Q&A). To foster interactivity during the workshop, we will have **two 60-minute poster sessions** (one in the morning and one in the afternoon to accommodate time zone constraints), one 60-minute-long **panel discussion** with domain experts, and one 40-minute-long small interactive round table with speakers/panelists/organizers and the audience. Two spotlight presentations of 40 minutes each will feature the most interesting works submitted to the workshop. Each session will feature 4 different papers (10 minutes per paper). *All invited speakers and panelists are confirmed.*

All invited speakers and panelists have been confirmed.

Time	Speaker	Affiliation	Areas of Expertise
Opening remarks			
9:00 - 9:10	Organizers		
Invited Talks I			
9:10 - 9:40	Marinka Zitnik	Harvard University	FMs for Therapeutic Science
9:40 - 10:10	Jure Leskovec	Stanford University	FMs for Perturbation
Spotlight I			
10:10 - 10:50	Accepted spotlights		
10:50 - 11:00	Break		
Poster Session I			
11:00 - 12:00	Accepted papers		
12:00 - 13:00	Lunch Break		
Invited Talks II			
13:00 - 13:30	Mihaela van der Schaar	University of Cambridge	ML for Medicine
13:30 - 14:00	Yun S. Song	UC Berkeley	Genomics Language Models
Spotlight II			
14:00 - 14:40	Accepted spotlights		
14:40 - 14:50	Break		
Expert Panel Discussion			
14:50 - 15:50	Djork-Arne Clevert	Pfizer	Knowledge Graph & AI Agents
	Shekoofeh Azizi	Google DeepMind	Multi-modal LLMs
	Max Welling	CuspAI/UvA	Generative AI
	Limsoon Wong	National University of Singapore	LLMs, Sequence Design
Round Table			
15:50 - 16:30	Speakers/Panelists/Audience/Organizers		
Poster Session II			
16:30 - 17:40	Accepted papers		
Closing remarks			
17:40 - 17:50	Organizers		

1.3 Attendees

We expect approximately 450 participants, which amounts to roughly 5% of the total ICLR attendees. This estimation is based on our experience organizing the 1st MLGenX workshop at ICLR 2024 where 470 people attended the workshop [2].

We expect the workshop to attract researchers both from machine learning and biology who are interested in diverse questions ranging from what ML can do for genomics as well as which datasets and questions in the next generation of therapeutics can contribute to advancing fundamental ML research.

1.4 Diversity

Our dedication to diversity, balance, equality, and inclusion is not only represented in the gender- balanced organizing team (including senior and junior researchers across several continents and institutions) but also in the invited speakers, panelists, reviewers, and the workshop's organizers:

- The organizing committee includes representation from different affiliations (academia and industry), seniority, geographic location (Europe, USA), gender, and ethnicity.

---

[2] The number of attendees is based on the Whova app.

- The speakers and panelists are chosen from different fields of expertise (both biology and ML), various institutions (both academia and industry), and different seniority levels, genders, and geographic locations.
- The program committee members are also chosen with demographic, expertise, and seniority diversity in mind to minimize the risk of biased judgment of the submissions.
- We offer partial funding for the attendance of two participants from underrepresented countries.
- We offer three distinct tracks—the main track, the special track on foundation models and agentic AI, and the tiny papers track—to encourage diverse participation from academia, industry, and early-stage researchers. By welcoming submissions from individuals with a wide range of backgrounds and expertise, we aim to create an inclusive environment that fosters collaboration and innovation across all levels of the ML and biology research communities.

1.5 Accessibility, Advertisement, and Website

While the workshop will primarily be held **in person**, we will also utilize a range of digital tools to *engage with our online audience* and enhance interactivity throughout the event. These tools include Gathertown for poster sessions and break discussions, as well as virtual Q&As conducted on Zoom. Moreover, our website will serve as a central platform for disseminating the call for papers, promoting the workshop, and providing early access to the planned agenda and talk titles. This enables attendees to make choices about their attendance based on the content schedule in an accessible manner. We will promote the workshop in advance on our website, via our social media channels, and through collaborations with industry and academic partners to attract a diverse community of researchers interested in machine learning for target identification.

1.6 Related Workshops

While there are areas of overlap with recent workshops, MLGenX will focus on emphasizing specific open problems in the field of genomics. Unlike previous workshops that were closer to drug discovery and structural biology, our workshop focuses primarily on high-throughput omics techniques, e.g. single-cell and perturbation data, which play a pivotal role in bridging genomics research and machine learning. More specifically, distinctions between our proposed workshop and related ones are as follows:

Drug Discovery: ICLR workshop on Machine Learning for Drug Discovery, NeurIPS Workshop on Machine Learning in Structural Biology (MLSB), and NeurIPS Workshop on New Frontiers of AI for Drug Discovery and Development (AI4D3).
The proposed MLGenX workshop distinguishes itself by delving into new and comprehensive topics. Unlike MLDD, MLSB, and AI4D3, which primarily focus on molecular design in drug discovery, MLGenX places a strong emphasis on *target identification*. This focus is pivotal not only as the initial step in molecule design but also for shaping the next generation of therapeutic domains, including RNA/DNA Vaccines, cell therapy, and gene therapy. From a machine learning perspective, MLGenX stands out by emphasizing foundation models applied to large-scale genetics and genomics data, causality, interpretability, and generalizability. This sets it apart from the MLDD, MLSB, and AI4D3 workshops, which primarily focus on active learning, graph neural networks, and generative models, respectively.

Computational Biology: NeurIPS Workshop on Generative AI and Biology (GenBio), ICML Workshop on Computational Biology (CompBio 2023), and NeurIPS Workshop on Learning Meaningful Representations of Life (LMRL).

The proposed MLGenX workshop on target identification stands apart from the previous workshops on computational biology because of its distinct focus and specialized objective. In fact, while past workshops in computational biology have explored the broader spectrum of computational techniques applied to various facets of biology, this workshop puts a strong focus on the three specific areas that connect deep learning to the field of target discovery which is so central in healthcare. From a machine learning perspective, MLGenX shares some similarities with representation learning in LMRL and generative models in GenBio. However, this workshop delves into the foundation models, generalizability, and effective methods for achieving *disentangled* representation learning and interpretability —challenging machine learning topics that we plan to thoroughly discuss during this event.

1.7 Tentative Timeline

- **Call for papers:** 15 December 2024
- **Submission deadline:** 3 February 2025
- **Reviewing period:** 4 February - 25 February 2025
- **Notification:** 5 March 2025

2 Workshop Organizers

Fabian Theis - Director of the Institute of Computational Biology and Professor at TUM Mathematics and Life Sciences, Germany (✉ fabian.theis@helmholtz-munich.de)
Fabian conducts research in the field of computational biology. The main focus of his work is the application of machine learning methods to biological questions, in particular as a means of modeling cell heterogeneities on the basis of single cell analyses and also of integrating "omics" data into systems medicine approaches. Since 2013 he has been a Full Professor of biomathematics at TUM, where he holds the Chair of Mathematical Modeling of Biological Systems, and director of the Institute of Computational Biology at the Helmholtz Zentrum München.
Previously organized workshops: Fabian organized multiple leading workshops on computational biology such as Workshop on Single Cell Genomics meets Data Science (2022) and Workshop on Computational Single Cell Genomics (2019). He was also a co-organizer of MLGenX 2024.

Aviv Regev - Executive Vice President, Genentech, USA (✉ regev.aviv@gene.com)
Aviv is a computational biologist and systems biologist and Executive Vice President and Head of Genentech Research and Early Development in Genentech/Roche. She is a former core member of the Broad Institute of MIT and Harvard and former professor at the Department of Biology of the Massachusetts Institute of Technology. Regev is a pioneer of single cell genomics and of computational and systems biology of gene regulatory circuits. She co-founded and co-leads the Human Cell Atlas project.
Previously organized workshops: She has also played a role in organizing several conferences and workshops, including AI for Science: Progress and Promises (NeurIPS 2022), Human Cell Atlas General Meetings, and the yearly Single Cell Genomics Conference (SCG). She was also a co-organizer of MLGenX 2024.

Arman Hassanzadeh - Software Engineer, Google, USA (✉ armanihm@google.com)
Arman is a software engineer at Google. His work is centered around developing multi-modal large language models. Prior to Google, Arman was a PhD student at Texas A&M University working with Nick Duffield and Mingyuan Zhou (UT Austin). His primary research interests are generative models, graph machine learning and Bayesian statistics.
Previously organized workshops: Arman was co-organizer of MLGenX 2024.

Mengdi Wang - Professor, Princeton University(✉ mengdiw@princeton.edu)
Mengdi is an associate professor in the ECE Department and the Center for Statistics and Machine Learning at Princeton University. She was a visiting research scientist at DeepMind. Her research focuses on RL and

generative AI for science. Mengdi has received numerous awards, including the Google Faculty Award in 2017, the MIT Tech Review 35-Under-35 Innovation Award in 2018, the WAIC YunFan Award in 2022, and AACC's Donald Eckman Award in 2024. *Previously organized conferences/workshops:* Mengdi serves as a Program Chair for ICLR 2023 and Senior AC for Neurips, ICML, COLT. She served as an organizer in New Frontiers of AI for Drug Discovery and Development at NeurIPS 2023.

Tommaso Biancalani - Director and Distinguished Scientist, Genentech Computational Sciences, USA (✉ biancalt@gene.com)
Tommaso is the head of the BRAID department (Biology Research — AI Development) which is part of the Computational Biology and Translation pillar within the Genentech Computational Science organization. The core mission of the BRAID team is to bridge foundational machine learning research with biology, with emphasis on target discovery. Prior to joining Genentech in 2021, Tommaso was at the Broad Institute of MIT and Harvard where he led a team working on the Human Cell Atlas project. Tommaso trained as a theoretical statistical physicist and completed his post-doctoral training at the Carl Woese Institute of Genomics and MIT.
Previously organized workshops: Tommaso served as a co-organizer for MLGenX 2024.

Sara Mostafavi - Associate Professor, University of Washington, USA (✉ saramos@cs.washington.edu)
Sara Mostafavi is an Associate Professor at the Paul Allen School of Computer Science and Engineering at the University of Washington (UW). Before joining UW, she was an Assistant Professor in the Departments of Statistics and Medical Genetics at the University of British Columbia (UBC) and a faculty member at the Vector Institute. She was the recipient of a Canada Research Chair (CRC II) in Computational Biology (2015-2020), and a Canada CIFAR Chair in Artificial Intelligence (CIFAR-AI). Sara completed her postdoctoral fellowship at Stanford University with Daphne Koller and earned her PhD in Computer Science from the University of Toronto, where she developed machine learning methods for predicting gene function. *Previously organized conferences:* Sara is co-founder of the Machine Learning for Computational Biology (MLCB) Conference

Ehsan Hajiramezanali - Principal AI Research Scientist, Genentech, USA (✉ hajiramezanali.ehsan@gene.com)
Ehsan is a principal AI research scientist at Genentech in the DELTA team within the BRAID (Biology Research — AI Development) department. Before that, he was an AI research scientist at AstraZeneca. Ehsan received his PhD from Texas A&M University working with Xiaoning Qian and Mingyuan Zhou (UT Austin). His research lies at the intersection of machine learning and Bayesian statistics. He is interested in probabilistic methods, generative models, representation learning, relational learning, and multi-domain learning. *Previously organized workshops:* ICLR 2023 Workshop on Machine Learning for Drug Discovery, MARBLE 2023 at ECML-PKDD, and MLGenX 2024.

3 Invited Speakers and Panelists

Marinka Zitnik. Marinka Zitnik is an Assistant Professor at Harvard University in the Department of Biomedical Informatics. Dr. Zitnik is Associate Faculty at the Kempner Institute for the Study of Natural and Artificial Intelligence, Broad Institute of MIT and Harvard, and Harvard Data Science. Dr. Zitnik investigates the foundations of AI to enhance scientific discovery and facilitate individualized diagnosis and treatment in medicine. Before joining Harvard, she was a postdoctoral scholar in Computer Science at Stanford University. She was also a member of the Chan Zuckerberg Biohub at Stanford. She received her bachelor's degree, double majoring in computer science and mathematics, and then graduated with a Ph.D. in Computer Science from University of Ljubljana just three years later while also researching at Imperial College London, University of Toronto, Baylor College of Medicine, and Stanford University. Dr. Zitnik is an ELLIS Scholar in the European Laboratory for Learning and Intelligent Systems (ELLIS) Society. She is a member of the Science Working Group at NASA Space Biology. She co-founded Therapeutics Data Commons and is the faculty lead of the AI4Science initiative. She was named Kavli Fellow 2023 by the US National Academy of Sciences.

Jure Leskovec. Jure Leskovec is a Professor of Computer Science at Stanford University. His primary research domain lies in applied machine learning for large interconnected systems, with a specific focus on modeling complex, richly-labeled relational structures, graphs, and networks across systems of all scales. These systems range from the interactions of proteins within a cell to the interactions between individuals in a society. Jure's research has broad applications, encompassing commonsense reasoning, recommender systems, computational social science, and computational biology, with a special emphasis on drug discovery.

Mihaela van der Schaar. Mihaela van der Schaar is the John Humphrey Plummer Professor of Machine Learning, Artificial Intelligence and Medicine at the University of Cambridge. In addition to leading the van der Schaar Lab, Mihaela is founder and director of the Cambridge Centre for AI in Medicine (CCAIM). Mihaela was elected IEEE Fellow in 2009 and Fellow of the Royal Society in 2024. She has received numerous awards, including the Johann Anton Merck Award (2024), the Oon Prize on Preventative Medicine from the University of Cambridge (2018), a National Science Foundation CAREER Award (2004), 3 IBM Faculty Awards, the IBM Exploratory Stream Analytics Innovation Award, the Philips Make a Difference Award and several best paper awards, including the IEEE Darlington Award. She was a Turing Fellow at The Alan Turing Institute in London between 2016 and 2024. She has made over 45 contributions to international standards for which she received 3 ISO Awards. In 2019, a Nesta report determined that Mihaela was the most-cited female AI researcher in the U.K.

Yun S. Song. Yun S. Song is a professor in Electrical Engineering & Computer Science (EECS) and Statistics. He holds BS degrees in mathematics and physics from MIT and earned his PhD in physics from Stanford University. His research is centered on computational and mathematical biology, where he develops computational tools and statistical methods to advance biomedical research. In addition to his work in data analysis and interpretation, Professor Song is deeply interested in machine learning, combinatorial optimization, algorithms, and Monte Carlo methods.

Djork-Arne Clevert. Djork-Arné Clevert is the Head of Machine Learning Research and Vice President at Pfizer, where he oversees pioneering efforts in applying AI and machine learning to pharmaceutical research and drug discovery. With a deep passion for advancing computational methods, he has a strong interest in AI-driven molecular design, including generative models, and the development of novel architectures like graph neural networks. Under his leadership, Pfizer's machine learning research focuses on predictive modeling, optimizing drug development, and improving patient outcomes. Clevert's work exemplifies the transformative potential of AI in healthcare, pushing the boundaries of innovation in medicine. He is a member of the European Lab for Learning and Intelligent Systems (ELLIS).

Shekoofeh Azizi. Shekoofeh Azizi is a staff research scientist and a research lead at Google DeepMind. Her research is focused on developing approaches that facilitate the translation of AI solutions into tangible clinical impact. She is particularly interested in designing foundation models for biomedical applications and I have been leading multiple major efforts in this area, including the moonshot project behind Med-PaLM and Med-PaLM 2, Google's flagship medical large language models. She is also lead researcher of Med-PaLM M the first demonstration of a generalist biomedical AI. Her research has been published in well-regarded journals and conferences such as Nature, Nature Medicine, Nature Biomedical Engineering, and CVPR and it serves as the cornerstone for multiple medical device products undergoing clinical trials at Google. Her works have been recognized by multiple awards including the Governor General's Canada Academic Gold Medal for contributions in improving diagnostic ultrasound.

Max Welling. Max Welling is a research chair in Machine Learning at the University of Amsterdam and a Distinguished Scientist at MSR. He is a fellow at the Canadian Institute for Advanced Research (CIFAR) and the European Lab for Learning and Intelligent Systems (ELLIS) where he also serves on the founding board. His previous appointments include VP at Qualcomm Technologies, professor at UC Irvine, postdoc at U. Toronto and UCL under supervision of prof. Geoffrey Hinton, and postdoc at Caltech under supervision of prof. Pietro Perona. He finished his PhD in theoretical high energy physics under supervision of Nobel laureate prof. Gerard 't Hooft. Max Welling has served as associate editor in chief of IEEE TPAMI from 2011-2015, he serves on the advisory board of the Neurips foundation since 2015 and has been program chair and general chair of Neurips in 2013 and 2014 respectively. He was also program chair of AISTATS in 2009 and ECCV in 2016 and general chair of MIDL 2018. Max Welling is recipient of the ECCV Koenderink Prize in 2010 and the ICML Test of Time award in 2021. He directs the Amsterdam Machine Learning Lab (AMLAB) and co-directs the Qualcomm-UvA deep learning lab (QUVA) and the Bosch-UvA Deep Learning lab (DELTA).

Limsoon Wong. Wong Limsoon is Kwan-Im-Thong-Hood-Cho-Temple Professor in the School of Computing at the National University of Singapore (NUS). He was also a professor (now honorary) of pathology in the Yong Loo Lin School of Medicine at NUS. Before coming to NUS, he was the Deputy Executive Director for Research at A*STAR's Institute for Infocomm Research. He currently works mostly on knowledge discovery technologies and their application to biomedicine. He has also done, in the earlier part of his career, significant research in database query language theory and finite model theory, as well as significant development work in broad-scale data integration systems. Limsoon is a Fellow of the ACM, inducted for his contributions to database theory and computational biology. Some of his other awards include the 2003 FEER Asian Innovation Gold Award for his work on treatment optimization of childhood leukemias, the 2006 Singapore Youth Award Medal of Commendation for his sustained contributions to science and technology, and the ICDT 2014 Test of Time Award for his work on naturally embedded query languages. He co-founded Molecular Connections in India, and has served as its chairman for over a decade, seeing the company growing progressively to some 2,000 information curators, software engineers, research scientists.

4 Advisory Committee

We received invaluable support and guidance from numerous domain experts from both academia and industry, in addition to our core organizers.

Aïcha BenTaieb - Director Of ML For Clinical Sciences, Genentech, USA (✉ bentaieb.aicha@gene.com)
Aicha BenTaieb is Director Of ML For Clinical Sciences at Genentech leading a research group focused on developing novel machine learning techniques for clinical trial design in the BRAID department. Prior to joining Genentech, she was the director of the Pathology AI team at Tempus Labs and a scientist at Roche Tissue Diagnostics.

Gabriele Scalia - Principal Research Scientist II, Genentech, USA (✉ scalia.gabriele@gene.com)
Gabriele is an AI research scientist at Genentech, where he leads the DELTA research group focused on developing new methods to support biology and biomedicine discovery. Before that, he was at the Broad Institute of MIT and Harvard, developing novel deep learning methods for genomics applications. He received his PhD in Computer Science and Engineering from the Polytechnic University of Milan, while conducting research at MIT as a visiting researcher. His research interests include graph learning, generative modeling, uncertainty estimation, and multi-modal learning.

5 Program Committee

We have assembled a diverse program committee for the workshop from various levels of seniority, fields of expertise, gender, and ethnic background to represent different views in the reviewing process. We will ensure the review load will not exceed more than 3 papers per reviewer. This helps us to provide constructive and detailed feedback to the authors. If the workshop is accepted, we will announce the call for papers by *15 December 2024* with the submission deadline of 3 February 2025. The reviews will be available to the authors by 5 March 2025. It will be communicated that we will *not accept* submissions that have previously been published. Moreover, the reviewers will be instructed to ensure the novelty of the submissions in their scores and mark any previously published work as such. To prevent conflict of interests, we will ask both authors and reviewers to provide their affiliations and also existing conflict with institutions, or individuals.

We will use Openreview as the submission and reviewing system. We will put the accepted papers on the workshop website and will communicate explicitly that the accepted papers are *non-archival* and can be published elsewhere in the future. Here are the tentative PC members (reviewers):

- Z. Liu (BeiGene)
- A. Banerjee (Illumina)
- S. Sankarapandian (Calico Labs)
- M. Wang (Purdue University)
- H. Sun (Stanford University)
- N. Jaydip Gandhi (Dartmouth College)
- O. Tastan (Microsoft Research)
- Y. Jiang (Yale University)
- S. Sivanandan (Insitro)
- M. Jindal (Trine University)
- J. Shah (Dolby Laboratories)
- L. Jiang (EPFL)
- M. Rohbeck (DKFZ)
- L. Cao (Pfizer)
- R. Lyu (CMU)
- A. Wu (MIT)
- R. Qureshi (MD Anderson Cancer Center)
- A. Saadat (EPFL)
- S. Sadhuka (MIT)
- C. Wan (Duke University)
- M. Koido (University of Tokyo)
- S. Batista (UCLA)
- M. Jia (University of Pittsburgh)
- C. Yuan (City University of New York)
- L. Li (Rice University)
- A. BenTaieb (Genentech)
- S. Mourragui (Ensocell)
- S. Niyakan (Texas A&M University)
- M. Yang (CMU)
- O. Ramos (Cold Spring Harbor Laboratory)
- H. Huang (Genentech)
- Y. Rosen (Stanford University)
- U. Hazra (Georgia Institute of Technology)
- P. Ghari (UC Irvine)
- L. Moukheiber (MIT)
- A. Xie (MSK Cancer Centre)
- T. Ucar (AstraZeneca)
- S. Jayasundara (Purdue University)
- A. Hashemi (Harvard University)
- A. Agrawal (Stanford University)
- H. Jeong (MIT)
- Y. Fu (Northwestern University)
- A. Lin (Harvard University)
- Y. Liu (Ohio State University)
- X. Huang (Calico Labs)
- W. Guo (UC Davis)
- C. Ye (UC Berkeley)
- R. Lopez (Stanford University)
- R. Littman (Genentech)
- Z. Lu (Genentech)
- B. Zhang (Rice University)
- C. Hu (Mayo Clinic)
- W. Guo (UCLA)
- P. Avdeyev (UT Southwestern Medical Center)
- A. Andersson (Genentech)
- X. Zhang (University of Minnesota)
- L. Lorch (EPFL)
- S. Ghosal (Broad Institute)
- P. Mohseni (Texas A&M University)
- S. Nair (Stanford University)
- D. Cakmakci (McGill University)
- E. Nguyen (Stanford University)
- C. Wang (MIT)
- G. Murtaza (Brown University)
- R. Zhang (University of Chicago)
- A. Lobley (University of London)
- Y. Chen (University of Maryland)
- L. Mao (Georgia Institute of Technology)
- W. Qiu (University of Washington)
- N. Wan (Nuanced Health)
- T. Yu (EPFL)
- K. Ni (CMU)
- N. Janakiraman (Google)
- S. Joshi (Columbia University)
- S. Thapa (Intel)
- Z. Pan (Purdue University)
- S. Muller (International Business Machines)
- Y. Wang (ByteDance)
- Y. Yang (University of Michigan)
- A. Feller (UT Austin)
- A. Turcan (CMU)
- R. Rastogi (UC Berkeley)
- X. Liu (Princeton University)
- A. Jain (Salesforce)
- A. Tseng (Genentech)
- Y. Yu (City University)
- M. Jain (Université de Montréal)
- G. Mishra (Northeastern University)
- Y. Lee (Genentech)
- Y. Xiao (UCLA)
- S. Maleki (Genentech)
- Y. Wang (UCB Biosciences)
- X. Tu (University of Washington)
- Y. Annadani (TUM)
- N. Diamant (Genentech)
- S. Banerjee (Amazon)

- W. Connell (UC San Francisco)
- Y. Jin (Harvard University)
- Y. Wu (City University of New York)

References

Leila Amini, Sara K Silbert, Shannon L Maude, Loretta J Nastoupil, Carlos A Ramos, Renier J Brentjens, Craig S Sauter, Nirali N Shah, and Mohamed Abou-el Enein. Preparing for car t cell therapy: patient selection, bridging therapies and lymphodepletion. *Nature Reviews Clinical Oncology*, 19(5):342–355, 2022.

Karen Bulaklak and Charles A Gersbach. The once and future gene therapy. *Nature communications*, 11 (1):5820, 2020.

Juan Manuel Zambrano Chaves, Eric Wang, Tao Tu, Eeshit Dhaval Vaishnav, Byron Lee, S Sara Mahdavi, Christopher Semturs, David Fleet, Vivek Natarajan, and Shekoofeh Azizi. Tx-llm: A large language model for therapeutics. *arXiv preprint arXiv:2406.06316*, 2024.

Gabrielle P Dailey, Erika J Crosby, and Zachary C Hartman. Cancer vaccine strategies using self-replicating rna viral platforms. *Cancer Gene Therapy*, 30(6):794–802, 2023.

Shanghua Gao, Ada Fang, Yepeng Huang, Valentina Giunchiglia, Ayush Noori, Jonathan Richard Schwarz, Yasha Ektefaie, Jovana Kondic, and Marinka Zitnik. Empowering biomedical discovery with ai agents. *arXiv preprint arXiv:2404.02831*, 2024.

Kaixuan Huang, Yuanhao Qu, Henry Cousins, William A Johnson, Di Yin, Mihir Shah, Denny Zhou, Russ Altman, Mengdi Wang, and Le Cong. Crispr-gpt: An llm agent for automated design of gene-editing experiments. *arXiv preprint arXiv:2404.18021*, 2024.

Guillaume Richard, Bernardo P de Almeida, Hugo Dalla-Torre, Christopher Blum, Lorenz Hexemer, Priyanka Pandey, Stefan Laurent, Marie P Lopez, Alexander Laterre, Maren Lang, et al. Chatnt: A multimodal conversational agent for dna, rna and protein tasks. *bioRxiv*, pp. 2024–04, 2024.

Shirley Wu, Shiyu Zhao, Qian Huang, Kexin Huang, Michihiro Yasunaga, Kaidi Cao, Vassilis N Ioannidis, Karthik Subbian, Jure Leskovec, and James Zou. Avatar: Optimizing llm agents for tool-assisted knowledge retrieval. *arXiv preprint arXiv:2406.11200*, 2024.