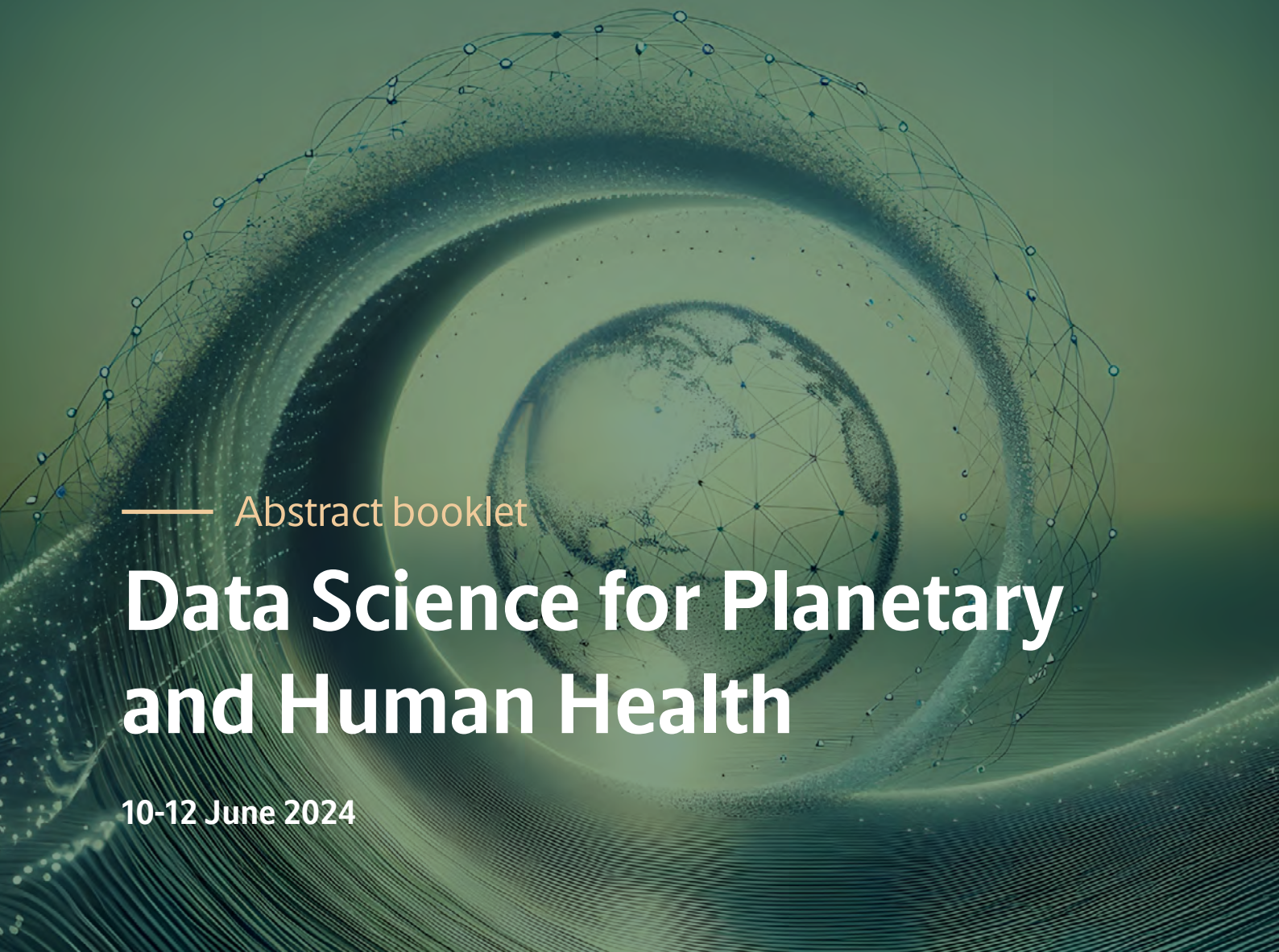
novo nordisk foundation Science Cluster



Abstract booklet

Data Science for Planetary and Human Health

10-12 June 2024

Programme - MONDAY JUNE 10

Time	MONDAY JUNE 10 - Scandic Spectrum
08:00-08.30	Arrival and breakfast
08:30-08:45	**Welcome by Scientific Committee**
	Alberto Santos Delgado, Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark, Denmark **Lea Sommer**, Novonesis, Denmark **Kasper Lage**, Novo Nordisk Foundation Centre for Genomic Mechanisms of Disease at the Broad Institute of MIT and Harvard, US and Denmark
08:45-09:30	**Introduction to Novo Nordisk Foundation Centers**
	Kasper Lage, Novo Nordisk Foundation Centre for Genomic Mechanisms of Disease at the Broad Institute of MIT and Harvard, US and Denmark **Simon Rasmussen**, Novo Nordisk Foundation Centre for Protein Research, University of Copenhagen, Denmark **Tune Pers**, Novo Nordisk Foundation Centre for Basic Metabolic Research, University of Copenhagen, Denmark **Richard Dennis**, Novo Nordisk Foundation Centre for Stem Cell Medicine, reNEW, University of Copenhagen, Denmark **Shilpa Garg,** Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark, Denmark
09:30-10:15	**Session 1**
	KEYNOTE SPEAKER: **Lars Keld Nielsen**, Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark, Denmark Abstract title: The Automated Scientist: the emergence of AI-guided Biofoundries Topic: The future of data and data science and their impact on current world challenges
10:15-10:45	Coffee break

Programme - MONDAY JUNE 10

10:45-12:00	Session 2
	SHORT TALK: Karina Banasik, Copenhagen University Hospital, Amager Hvidovre Hospital, Denmark Abstract title: Multi-Omics studies in Women's Health and reproduction **SHORT TALK: Annelaura Bach Nielsen**, University of Copenhagen, Denmark Abstract title: Improving Clinical Diagnostics by integrating Data Science and Technological Advances **SHORT TALK: Alexander Henriksen**, University of Copenhagen, Denmark Abstract title: ServerEmissions: A tool to estimate resource use of large-scale server-based computations **SHORT TALK: Sumesh Sukumara**, Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark, Denmark Abstract title: Tools for automating sustainability assessment workflows coupled with microbial production data throughout the product development phase. **Chair: Tune Pers, Novo Nordisk Foundation Center for Metabolic Research, University of Copenhagen, Denmark**
12:00-13:00	Lunch
13:00-14:45	**Session 3**
	INVITED SPEAKER: David Ochoa, EMBL-EBI , United Kingdom Abstract title: Open Targets: towards informed target identification and prioritization Topic: Democratizing knowledge by building and sharing data resources **SHORT TALK: Maria Dalby**, Muna Therapeutics, Denmark Abstract title: Cellular resilience to Alzheimer's disease pathology by spatial and single-cell transcriptomics **SHORT TALK: Nils Hofmann**, Technical University of Denmark, Denmark Abstract title: ExpoSeq: simplified analysis of high-throughput sequencing data from antibody discovery campaigns **SHORT TALK: Angel Phanthanourak**, Technical University of Denmark, Denmark Abstract title: KGQC: A Standard Framework for the Evaluation of Biological Knowledge Graph Quality **SHORT TALK: Elysia Gao**, Technical University of Denmark, Denmark Abstract title: Decoding Cell Behavior Patterns in Bioreactors: Interpreting Latent Spaces with Deep Autoencoders **Chair: Simon Rasmusen, Novo Nordisk Foundation Centre for Protein Research, University of Copenhagen, Denmark**

Programme - MONDAY JUNE 10

14:45-15:15	Coffee break
15:15-17:15	**Session 4**
	INVITED SPEAKER: Anika Gupta, Google Ventures, United States Abstract title: Going farther together: effectively communicating innovation in data-driven biomedicine Topic: Communication about data and data science and making them attractive to diverse audiences. **SHORT TALK: Adrian Geissler**, University of Copenhagen, Denmark Abstract title: Show and tell: Visualize comparative genomics RNA structure predictions with Quarto and ObservableJS **SHORT TALK: Anna Svetlova**, Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark Abstract title: Data-Driven Investigation of Secondary Metabolism in Actinomycetes **SHORT TALK: David Westergaard**, Copenhagen University Hospital Hvidovre, Denmark Abstract title: Uncovering the Heritable Components of Multimorbidities and Disease Trajectories: A Nationwide Cohort Study **SHORT TALK: Nadezhda Doncheva**, University of Copenhagen, Denmark Abstract title: Cytoscape stringApp 2.0: Analysis and Visualization of Heterogeneous Biological Networks **Chair: Line K.H. Clemmensen, Department of Applied Mathematics and Computer Science, Technical University of Denmark, Denmark**
17:15-19:00	**Poster session and drinks**

Programme - TUESDAY JUNE 11

Time	TUESDAY JUNE 11 - Scandic Spectrum
08:30-09:00	Arrival and breakfast
09:00-11:00	**Session 5: Panel debate and Group discussion**
	Opportunities for innovation and improvement within the areas of research data management, data science and science communication **Chairs: Alberto Santos Delgado, Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark, Denmark and Lea Sommer, Novonesis, Denmark**
11:00-12:00	**Session 6**
	SHORT TALK: Beatrice Dyring-Andersen, Gentofte Hospital, Denmark Abstract title: The spatial proteome of psoriasis **SHORT TALK: Michel Edwar Mickael**, Institute of Genetics and Animal Biotechnology, Poland Abstract title: Adera2.0 a novel AI-based pipeline for drug repurposing **SHORT TALK: Bunia Adsersen**, Technical University of Denmark, Denmark Abstract title: Design, Production and Analysis of Minibinders against Gremlin-1 **SHORT TALK: Isabel Díaz-Pinés Cort**, University of Copenhagen, Denmark Abstract title: The Inflammatome: A Meta-Analysis of Human Genes Regulated During Inflammation **Chair: Jens Vinge Nygaard, Department of Biological and Chemical Engineering, Medical Biotechnology, Aarhus University, Denmark**
12:00-13:00	Lunch
13:00-14:30	**Session 7**
	INVITED SPEAKER: Niklas Blomberg, IHI JU, Belgium Abstract title: Innovative Health Initiative – Europe's public-private partnership for health research Topic: How to solve the complexity of data FAIRification (making data FAIR) **SHORT TALK: Igor Marín de Mas**, Copenhagen University Hospital, Rigshospitalet, Denmark Abstract title: A Protocol for the Automatic Construction of Highly Curated Genome-Scale Models of Human Metabolism **SHORT TALK: Narod Kebabci**, University College Dublin, Ireland Abstract title: Predicting context-specific synthetic lethality between paralogs in cancer

Programme - TUESDAY JUNE 11

13:00-15:00	Session 7
	SHORT TALK: Ester Milesi, Technical University of Denmark, Denmark Abstract title: Building an Omics Data Infrastructure to Integrate Data Management and Data Science **SHORT TALK: Eleonora Nigro**, Scientitic Illustrator, Denmark Title: From Data to Drawing: A (scientific) illustrated journey **Chair: Richard Dennis, Novo Nordisk Foundation Centre for Stem Cell Medicine, reNEW, University of Copenhagen, Denmark**
15:00 - 15:30	Coffee break
15:30-17:30	**Session 8**
	INVITED SPEAKER: Adriana Tomic, Boston University, United States Abstract title: Creating Smarter Vaccines: Understanding Human Immune System with Artificial Intelligence Topic: Omics technologies and AI with a focus on vaccine development **SHORT TALK: Yesid Cuesta Astroz**, University of Antioquia, Colombia Abstract title: OrthoHPI 2.0: A renewed and extended interactive resource of predicted human-parasite protein-protein interaction networks **SHORT TALK: Julia Villarroel**, ZS Discovery, Denmark Abstract title: Novel algorithm-driven vaccine design framework, built to ensure optimal coverage of both population diversity and pathogen variation, aimed at eliciting strong cellular and humoral response, as well as lasting immune memory **SHORT TALK: Timothy Jenkins**, Technical University of Denmark, Denmark Abstract title: Decode, Design, Deliver: Machine Learning Revolutionises Target Identification and Therapeutic Discovery **SHORT TALK: Boyang Ji**, BioInnovation Institute, Denmark Abstract tile: Modeling the Human Gut Microbiome - Towards a Virtual Environment-Human-Microorganisms Model **Chair: Shilpa Garg, Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark, Denmark**

Programme - TUESDAY JUNE 11 & WEDNESDAY JUNE 12

Time	TUESDAY JUNE 11 - Scandic Spectrum
17:30-19:00	**Poster session and drinks**
19:30-22:00	Dinner w. announcement of poster winners
22:00-00:00	**Party**

Time	WEDNESDAY JUNE 12, Novo Nordisk Foundation, Hellerup
09:15-09:45	Arrival and coffee
09:45-11:00	**Session 9**
	INVITED SPEAKER: Birte Höcker, Bayreuth University, Germany Abstract title: Learning from nature how to design new proteins Topic: Computational and experimental design of protein function **SHORT TALK: Martin Miller**, AstraZeneca, United Kingdom Abstract title: Modelling multi-omic, real-world data reveals immunogenomic drivers of resistance to cancer immunotherapy **SHORT TALK: Francesc Fernández**, Data Science in R&D, Almiral, Spain Abstract title: 3D modelling with VR and indication expansion **Chair: Alberto Santos Delgado, Novo Nordisk Foundation Centre for Biosustainability, Denmark**
11:00-11:30	Coffee break
11:30-12:30	**Session 10**
	INVITED SPEAKER: Lars Juhl Jensen, Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, Denmark Abstract title: Simplifying the complex: Communicating data and concepts to non-expert users Topic: Communication of science network biology and text mining **Chair: Kasper Lage, Novo Nordisk Foundation Centre for Genomic Mechanisms of Disease at the Broad Institute of MIT and Harvard, US and Denmark**

Programme - WEDNESDAY JUNE 12

12:30-13:30	Lunch
13:30-14:30	**Session 11**
	SHORT TALK: Alisa Pavel, Technical University of Denmark, Denmark Abstract title: Knowledge Graphs for Integrated Data Analysis in Toxicology and Pharmacology **SHORT TALK: Naba Al-Sari**, Novo Nordisk, Denmark Abstract title: Identification of non-invasive biomarkers for predictive and prognostic use for end-stage renal disease **SHORT TALK: Lukas Huschet**, Ludwig Maximilian University (LMU). Germany Abstract title: Circadian dynamics of the nuclear proteome in time and space and its modulation by metabolic state **Chair: Kasper Lage, Novo Nordisk Foundation Centre for Genomic Mechanisms of Disease at the Broad Institute of MIT and Harvard, US, and Denmark**
14:30-15:00	Coffee break
15:00-15:45	**Session 12: Keynote speaker 4**
	INVITED SPEAKER: Emma Lundberg, Stanford University and KTH Royal Institute of Technology, US and Sweden Abstract title: From subcellular mapping to modeling the human cell Topic: Combining imaging and spatial multi-omics
15:45-16:15	**Closing remarks by the Scientific Committee**

Speakers

Personal Information

Full name: Adriana Tomic

Position: Professor

Institution: Boston University

Department: Virology, Immunology & Microbiology

Institution country: United States

Abstract

Creating Smarter Vaccines: Understanding Human Immune System with Artificial Intelligence

The development of more effective vaccines necessitates a thorough understanding of the human immune system, achievable through a comprehensive systems-level analysis using high-throughput 'omics' technologies. The primary challenge is converting extensive datasets into meaningful knowledge. We demonstrate that combining 'omics' with Artificial Intelligence (AI) offers a novel approach to grasping the intricate dynamics of host-pathogen interactions, essential for the informed design of the next-generation of vaccines. By transforming vast 'omics' datasets into practical insights, we explore immune response mechanisms to infectious diseases, including influenza and SARS-CoV-2. Our methodology employs multi-omics data analysis and computational modeling to predict immune responses, pinpoint protection markers, and guide new vaccine development. We have uncovered distinct immunity signatures, underscoring the complexity of human protective immunity against significant health threats. By providing global access to our AI platform, we strive to accelerate the understanding of immune processes and the progress of vaccine technology. This approach yields essential knowledge on immune protection and memory, facilitating the creation of more precise and powerful vaccines against infectious diseases.

Personal Information

Full name: Anika Gupta
Position: Principal
Institution: Google Ventures
Department: Life Sciences
Institution country: United States

Abstract

Going farther together: effectively communicating innovation in data-driven biomedicine

Advances in data science are rapidly expanding our understanding of and ability to improve human health. From the molecular to the population level, we can increasingly pinpoint which factors—inherited or environmental—associate with and may even cause disease. While such findings can transform the way medicines and public health guidelines are developed, far too often, they are opaque in their presentation. At baseline, scientific reports and publications are laden with technical jargon. Further, while all research findings entail some degree of uncertainty, such uncertainty is often poorly communicated. The consequence: the potential for confusion or mistrust by the public, the ultimate beneficiary of such research. Our best hope at incorporating learnings from cutting-edge research into society is to motivate the "why," clearly delineate the "what and how," and drive home the "so what"—establishing a shared vocabulary from the start, while also leaving room for conclusions to be updated as new data are incorporated. Drawing from Dr. Gupta's podcast, The Data Pulse, which delves into leading innovations from over 20 experts in data-driven biomedicine, this talk will walk through a framework for anyone hoping to effectively communicate research to a broader audience. Only by bringing everyone in can we enable the uptake and acceleration of the kind of progress that only a sliver of the population may envision today.

Personal Information

Full name: Birte Höcker

Position: Professor

Institution: Bayreuth University

Department: Biochemistry

Institution country: Germany

Abstract

Learning from nature how to design new proteins

Proteins are the machines of life. They are ubiquitous and diverse macromolecules that are essential for all cellular processes. Nature has generated this impressive set of proteins through evolution. Many protein structures and even more protein sequences are known by now. This enormous set of data can be used on the one hand to learn about the evolutionary history and how different proteins came about. On the other hand, we can extract information to be applied in the design of new tailor-made proteins. The ability to design custom proteins, such as reagent antibodies, biosensors or enzymes, is a major goal in protein biochemistry and will be necessary to tackle global challenges that we face today in ecology, biotechnology and medicine. Here I will discuss advantages and difficulties of different approaches and show some highlights from our work on designing new complex proteins.

Personal Information

Full name: David Ochoa

Position: Open Targets Platform Coordinator

Institution: EMBL-EBI

Department: Open Targets

Institution country: United Kingdom

Abstract

Open Targets: towards informed target identification and prioritisation

The identification of the right drug target for the right indication constitutes a critical step in the development of safe and effective medicines. The Open Targets public-private partnership aims to better understand the many factors influencing the selection targets providing a systematic view on disease causality and target modulation. In this talk we will walk through some of the factors that lead to successful targets in different therapeutic areas with a particular focus on understanding how we could inform future medicines.

Personal Information

Full name: Emma Lundberg

Position: Associate Professor

Institution: Stanford University & KTH, Royal Institute of Technology

Department: Bioengineering

Institution country: United States and Sweden

Abstract

From subcellular mapping to modeling the human cell

Biological systems are functionally defined by the nature, amount and spatial location of the totality of their proteins. We have generated an image-based map of the subcellular distribution of the human proteome and showed that there is great complexity to the subcellular organization of the cell giving rise to potential pleiotropic effects. As much as half of all proteins localize to multiple compartments and around 20% of the human proteome shows temporal variability. Our temporal mapping results shows that cell cycle progression explains less than half of all temporal protein variability, and that most cycling proteins are regulated post-translationally, rather than by transcriptomic cycling. This work is critically dependent on computational image analysis, and I will discuss machine learning approaches embedding of spatial subcellular patterns and how such embeddings as well as generative AI can be used to build multi-scale models of cell architecture. In summary, I will demonstrate the importance of spatial proteomics data for improved single cell biology and present how the freely available Human Protein Atlas database (www.proteinatlas.org) can be used as a resource for life science.

Personal Information

Full name: Lars Juhl Jensen

Position: Professor

Institution: University of Copenhagen

Department: Novo Nordisk Foundation Center for Protein Research

Institution country: Denmark

Abstract

Simplifying the complex: Communicating data and concepts to non-expert users

Data scientists often find themselves needing to explain complex data and analysis methods to non-experts. As a developer of data science tools intended for analysis and visualization of omics data, such as the STRING database of protein interactions, I find myself in this situation particularly often. In this presentation, I will focus on two aspects of data-science communication, namely the design of data visualization tools and the use of social media to teach them to a wide audience.

Data visualization: When it comes to omics data, good visualization is essential to tell "data stories" and to discover them in the first place. This inherently involves simplification; if you try to show everything, you fail at communicating anything. However, giving users the ability to drill down to the original evidence — as exemplified by the STRING evidence viewers — is crucial for building trust.

Teaching data science: Experts often lose their audience by assuming too much, using technical terminology, providing unnecessary detail. Avoiding these mistakes is especially important if communicating on social where the attention span is very short. However, if done right, a YouTube channel can be a very powerful way to teach orders of magnitude more potential users of your tools than otherwise possible.

Personal Information

Full name: Lars Keld Nielsen

Position: Scientific Director

Institution: Technical University of Denmark

Department: Center for Biosustainability

Institution country: Denmark

Abstract

The Automated Scientist: the emergence of AI-guided Biofoundries

Biofoundries emerged during the 2010s as an Industry 4.0 solution to accelerating cell factory engineering. Early efforts focused on automating the parallel construction of 100s cell factories. It soon became evident that the real bottleneck in strain design is the efficient management and exploitation of large volumes of heterogeneous data. This led to a second generation of data centric biofoundries to enhance capture and utilization of data. Advances in generative AI is expected to lead to a third generation of biofoundries, where the design, build, test cycle is largely guided by AI. Whereas large language models already enable de novo design of proteins, we are yet to develop the equivalent for pathways and cells. As the number of closed genome sequences remains limited, it will be critical to develop a detailed understanding of the "grammar" of genomes to guide AI. The development of knowledgebases for key organisms will enable contextualization of experimental data in the near term and hopefully guide the development of a genomic grammar.

Personal Information

Full name: Niklas Blomberg

Position: Executive Director, Innovative Health Initiative

Institution: IHI JU

Institution country: Belgium

Abstract

Innovative Health Initiative – Europe's public-private partnership for health research

IHI is a true partnership of public and industry actors – each of our projects draws half of their funding from our industry members and the other half from European public funds. This makes the IHI programme unique globally, there is no other programme that drives interdisciplinary research at this level – each of IHI's project teams are composed of experts from all sectors: medical researchers, healthcare practitioners, patients, and health industry.

Medical science is becoming increasingly interdisciplinary and IHI's ambition is to bring together expertise from many fields – life sciences, biotechnology, technology but also social and behavioural sciences to address Europe's health challenges. The programme funds a large cross-sectorial portfolio of projects, for instance projects that bring together expertise in medical and pharmaceutical research with health technologies and artificial intelligence to address novel modes for cancer treatments, or projects that work with healthcare practitioners to assess the capability of novel technologies to reduce stress and cognitive load of healthcare staff working in intensive care units.

The intervention will present an overview of the programme and the opportunities for public and private research organisations and healthcare actors to participate as well as the opportunities for partnerships with other research funders and foundations. Two key topics for discussions are the opportunities around regulatory science – ensuring timely development of the safe healthcare solutions of tomorrow and the making most of Europe's data for advancing health research.

Attendees

#001

Personal Information

Full name: Adrian Geissler

Position: Postdoc

Department: Department of Veterinary and Animal Sciences

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Show and tell: Visualize comparative genomics RNA structure predictions with Quarto and ObservableJS

Effective science communication requires both easy access to and visualization of computational results. In the realm of big data, presenting results with data volume exceeding simple Excel files can be achieved with interactive webpages. However, setting up webpages without advanced programming skills or lacking server infrastructure can be a challenging task. An open-source solution to present data in a reproducible manner without dedicated web programming skills is the publishing system Quarto, which can produce documents, slideshows, books, and websites. Quarto supports multiple programming languages, such as Python, R, Julia, and JavaScript. Quarto facilitates the interchange of data between these languages, and supports the ObservableJS framework for creating interactive data visualization without dedicated JavaScript programming needs and complex server requirements.

Here, we present how Quarto and ObservableJS have been utilized to present results from a large-scale comparative genomics study in >200 cyanobacterial genomes that predicted >400 novel RNA structure motifs. This data representation allows a user to filter motifs by tentative biological pathway or phylogenetic associations and to view the detailed scores and structure of each prediction. Further, we integrated a more advanced JavaScript library for displaying structural alignments directly in the browser with JalViewJS.

Our showcase demonstrates a presentation of computational results to which otherwise a non-technical or specialized user would not have had access, while the implementation could be transferred to display results from other bioinformatics analyses.

#002

Personal Information

Full name: Agnete Troen Lundgaard

Position: Postdoc

Department: Department of Obstetrics and Gynecology

Institution: Copenhagen University Hospital Hvidovre

Institution country: Denmark

Abstract

Immune Changes in Pregnancy: Associations with Pre-existing Conditions and Obstetrical Complications at the 20th Gestational Week - A Prospective Cohort Study

Pregnancy-related diseases and outcomes such as gestational diabetes mellitus (GDM), pre-eclampsia, and preterm birth pose serious threats to maternal and fetal health. The identification of early biomarkers linked to specific molecular mechanisms is critical to improve early detection and intervention.

We analyzed the associations between 47 proteins involved in inflammation, chemotaxis, angiogenesis, and immune system regulation, maternal and neonatal health outcomes, and the baseline characteristics and pre-existing conditions (diseases and obstetric history) of the mother in a prospective cohort of 1,049 pregnant women around the 20th gestational week. Bayesian linear models were used to determine the effect of risk factors and disease history on biomarker levels. Bayesian cause-specific parametric proportional hazards models together with Shapley additive explanation (SHAP) scores were used to determine the predictive potential of biomarkers for future maternal and neonatal outcomes.

We identified distinct biomarker patterns associated with risk factors, including smoking, ethnicity, and COVID-19 exposure. Obstetric complications including GDM and preterm labor in a prior pregnancy were associated with multiple pro-inflammatory cytokines and growth factors. The inclusion of biomarkers was found to improve the prediction of GDM and pre-eclampsia to a larger extent than commonly used clinical measures.

Our study provides novel insights into the interplay between preexisting conditions and immune dysregulation in pregnancy. These findings contribute to the understanding of the etiology of obstetric complications and identify novel biomarkers in early pregnancy.

#003

Personal Information

Full name: Ajuna Azad

Position: Head of Research Data Management

Department: DTU Bioengineering

Institution: Technical University of Denmark (DTU)

Institution country: Denmark

Abstract

Empowering FAIR Research: A Multi-Step Approach to Data-Driven Innovation

Research Data Management initiative at DTU Bioengineering is aimed at both enhancing data management practices and fostering digital innovation. We have formulated a five step approach to establish a robust data storage, an adaptable data lifecycle and a FAIR data management system adhering to institutional data policies.

1. Infrastructure Development:We are establishing a multi-cloud computing infrastructure tailored to the needs of our researchers. This will empower researchers with scalable resources(eg: for deploying and hosting machine learning tools), thus facilitating computation-intensive tasks and enabling IoTs for automated labs.

2. Skills Enhancement:Comprehensive training and courses are being organised to upskill researchers. Focus areas include both technical skills like cloud computing, GitHub, machine learning; and also data management- aiming to catalyse innovation and enhance reproducibility of scientific findings.

3. Establishing more Data Governance Layers: We are strengthening our data governance by recognising 'data ambassadors' at various levels of department to join our core which includes RDM, IT security, and administrative staff including GDPR functions.

4. Automate Submission of DMPs: We are working on a submission portal for Data Management Plans (DMPs) to facilitate seamless compliance and promote data stewardship throughout the research lifecycle.

5. Community Engagement:We are committed to fostering a vibrant data science community through our Data Science Club. Organizing community-building events, informative talks on applications in biotechnology and biomedical projects will nurture a culture of collaboration and knowledge exchange.

#004

Personal Information

Full name: Alba Refoyo Martinez

Position: Data Scientist

Department: Center for Health Data Science

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Reduce data chaos with the Health Data Science Sandbox: practical RDM tips for academic researchers

The National Health Data Science Sandbox is a collaborative national initiative that engages with a network of health data science experts to develop educational resources hosted on academic supercomputers. These resources are freely accessible to teachers, students, and researchers across Denmark. They include specialized training modules packaged in a portable containerized format for easy access and use, as well as computing support and setup services for courses and workshops. We believe efficient research data management (RDM) is crucial for maximizing scientific discovery and innovation, particularly in bioinformatics. Inspired by FAIR principles (Findability, Accessibility, Interoperability, and Reuse of digital assets) and Open Science principles, we aim to overcome the challenges posed by disorganized and outdated data repositories within research labs. We are compiling a comprehensive repository of reading material and best practices for RDM implementation in bioinformatics, including metadata standards, software for reproducibility and shareability, and the utilization of community-curated pipelines. Our ultimate goal is to develop foundational packages to support the adoption of RDM for Next-Generation Sequencing (NGS) data within academic research labs.

#005

Personal Information

Full name: Albert Palleja Caro

Position: Senior data scientist

Department: Informatics platform

Institution: Novo Nordisk Foundation Center for Biosustainablility

Institution country: Denmark

Abstract

A graph-based framework to analyze microbial associations in the wastewater treatment microbiome

Wastewater treatment (WWT) is the process of removing contaminants from used water before it is discharged back into the environment, which contributes to address water scarcity and to protect aquatic ecosystems. Recent advances in high-throughput omics technologies have facilitated the study of microbiomes from complex environmental samples such as WWT samples. These meta-omics datasets have been stored in databases such as MGnify and others, following different standards and providing different biological information. As a result, a comprehensive study of an environmental microbiome requires integrating data from various studies and meta-omics technologies, as well as biological knowledge to interpret these data. Here, we investigated the microbiome of treated wastewater to build an open-source knowledge graph (KG) that integrates WWT metagenomic and metatranscriptomic datasets with their biological context, including physicochemical conditions, metabolic pathways, and additional metadata. We developed a workflow to collect metagenomic datasets from MGnify and infer potential interactions among microorganisms through microbial association networks (MANs). MANs and additional metadata were integrated into a KG that enabled the identification of microbial communities and their possible roles in WWT. Moreover, this project provides easy access to the WWT microbiome information through an user-friendly interface that queries and summarizes MANs signatures. Although this project ocuses on WWT datasets, the workflow is easily adaptable to the analysis of other biomes.

#006

Personal Information

Full name: Albert Thor Thorhallsson

Position: Postdoc

Department: Department of Biotechnology and Biomedicine

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Active site modifications and its effect on transfucosylation in GH29B: A computational perspective with experimental confirmation

An oligosaccharide is a saccharide polymer containing a small number of monosaccharides. Oligosaccharides can have many functions in human health. Enzyme catalysed transglycosylation is one way of producing oligosaccharides from natural substrates. Engineering glycosidase hydrolases (GHs) to improve transglycosylation yields, is a promising avenue for high-value/low-cost oligosaccharides production. Molecular dynamics (MD) simulations were used as an effective tool to study the structure and dynamics of GHs of interest, belonging to the GH29 subfamily B, and revealed potential candidates for mutated variants with more/less transglycosylation yields, depending on the desired effect of mutation. Experimental validation of the predicted variants resulted in excellent agreement with the computational prediction.

#007

Personal Information

Full name: Alexander G. B. Grønning

Position: Postdoc

Department: Novo Nordisk Foundation Center for Metabolic Research

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Discovering functionally intra-connected modules from dimensionality reduced data

The use of dimensionality reduction techniques is an unavoidable step when analyzing high-dimensional omics data. A lower-dimensional representation of a dataset can serve as a platform to more clearly understand the analyzed data. For instance, it provides a means for gaining a clear visual overview and reduces noise to improve clustering and classification performance.

A wealth of computational methods have been developed to extract data items that contribute to explaining the separation of samples in dimensionality-reduced data. However, current algorithmic solutions fall short when attempting to identify subtle or gradual molecular and functional differences between dataset samples along a distinct path or axis in the dimensionality-reduced data.

Revealing such trends is pivotal when, for instance, seeking to functionally distinguish between differentiating cells or tissue samples from patients with varying disease progression.

To aid researches in tasks like these, we are developing a webtool that allows users to generate a new axis through the dimensionality reduced data. The generated axis will provide means forcomparing the molecules of the analyzed dataset and for extracting functionally intra-connected modules whose gradual expression change help drive the separation of the samples along the applied axis.

#008

Personal Information

Full name: Alexander Pil Henriksen

Position: PhD student

Department: Novo Nordisk Foundation Center for Protein Research

Institution: University of Copenhagen

Institution country: Denmark

Abstract

ServerEmissions: A tool to estimate resource use of large-scale server-based computations.

Computational research is an approach growing in popularity across many different scientific domains. In life science and medical science specifically, computational approaches have long been used to analyse biological data such as DNA fragments, and are now increasingly being used for predictive approaches, for example through the use of machine learning. While these approaches may provide insights and results more quickly and with fewer resources than more conventional lab and clinical studies, the resource use associated with large-scale computations shouldn't be overlooked. Globally, the use of computational infrastructure such as data centers and cloud services is growing rapidly. Some estimate that already, IT services and data centers emit more carbon than the aviation industry.

A first step towards less resource-intensive computation is to measure current energy consumption. Already, there exist a number of carbon and energy trackers that can be used to track energy use and emissions from large-scale computations. We complement these existing tools with our tool, ServerEmissions, which tracks resource use tied to computations on web servers using the Torque queuing system. ServerEmissions is easy to use, can be run online and can be customised to various geographical locations to account for differences in carbon intensity of electricity.

#009

Personal Information

Full name: Alisa Pavel

Position: Postdoc

Department: Department of Applied Mathematics and Computer Science Statistics and Data Analysis

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Knowledge Graphs for Integrated Data Analysis in Toxicology and Pharmacology

Large scale Big Data integration and analysis pervades the realm of life sciences, presenting both significant potential and challenges. In contrast to other disciplines the fragmented nature of the research field and sub-fields result in diverse data, knowledge, and standards, challenging integrated analysis and knowledge transfer across sub-fields. However, the integration of diverse data and data types is essential for comprehensive modeling of complex biological processes, for example in drug development and chemical safety assessment.

Knowledge Graphs (KGs) are one approach trying to make the life science data landscape integrable and analyzable by serving as knowledge bases, data analysis engines, and discovery systems. By functioning as an integrated and unified data model, KGs enhances data re-usability and informativity, while their link-oriented structure enables the inference of new insights, bridging fragmented data and fostering a more comprehensive understanding of biological processes by being analyzable with systems and network biology approaches.

We have created a multi-billion data point KG and used its data provisioning and discovering abilities in combination with network analysis methodologies across multiple tasks, such as for drug repositioning, transcriptomics analysis and chemical safety assessment.

#010

Personal Information

Full name: Angel Phanthanourak
Position: Research Assistant
Department: Novo Nordisk Foundation Centre for Biosustainability
Institution: Technical University of Denmark
Institution country: Denmark

Abstract

KGQC: A Standard Framework for the Evaluation of Biological Knowledge Graph Quality

Since 2012, Knowledge Graphs (KGs) have proven to be valuable tools for integrating heterogeneous data through graph-based approaches. Using graph structures, KGs can enhance the analysis of diverse datasets and increase the performance of machine learning (ML) tasks. To date, over 50 biology- and biomedicine-specific KGs have been proposed for solving challenges such as predicting protein-protein and microbial interactions, and as part of drug discovery pipelines. As AI development priorities shift from model optimization to enhancing data quality and quantity, the need for high-quality KGs and standards when building models is becoming essential. However, despite the surge and downstream impact of biological KGs, there are no standards for the evaluation and reporting of data quality making it difficult to compare KGs for reuse for ML tasks. Here, we propose the Knowledge Graph Quality Control (KGQC) framework, both a standard and a tool for assessing the quality of biological KGs during the graph construction process. The first of 3 KGQC modules reports structure-based evaluations, such as topological analyses. Module 2 evaluates the KG accuracy and completeness on domain-relevant topics, such as its ability to identify symptoms of a given disease. The third module focuses on downstream applications of the KG, calculating performance metrics in specific tasks such as drug repurposing prediction using benchmark datasets. KGQC is a python library interoperable across graph database software such as Neo4j and ArangoDB, and it aims to become a standard for the production and dissemination of KG quality metrics to enhance the reliability of results obtained downstream.

#011

Personal Information

Full name: Angelica Prado
Position: Data Standards Developer
Department: Data Systems and Standards
Institution: Novo Nordisk
Institution country: Denmark

Abstract

Considerations for standardisation and integration of omics datasets from the clinical research and regulatory perspectives

Omics data plays an important role in the understanding of metabolic processes across different scientific fields. Despite the wide use of these technologies both in industry and academy, we still face challenges especially when it comes to reusability of data across institutions due to the lack of a commonly adopted standardisation system.Therefore, any metadata standardisation initiative should consider potential sources of variability both in the experimental and computational phases of omics data generation. In this presentation we explore the main challenges for reusability and interoperability of omics data in terms of standardisation and data systems and suggest fundamental considerations for the creation of a common standardisation system from the clinical and regulatory perspectives aiming to ensure accessibility and reusability of complex and large omics datasets.

#012

Personal Information

Full name: Anna Svetlova

Position: PhD Student

Department: Natural Product Genome Mining

Institution: Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark

Institution country: Denmark

Abstract

Data-Driven Investigation of Secondary Metabolism in Actinomycetes

Streptomycesare Gram-positive bacteria known to produce a variety of bioactive compounds that benefit planetary and human health, for example antibiotics, anti-cancer, anti-inflammatory agents, pesticides, herbicides, and plant growth promoters. The genus Streptomycesis incredibly diverse, with thousands of known species. Streptomycesremain largely underexplored, with each species having the capability toproduce on average 30 potentially valuable secondary metabolites.

A major hurdle in unlocking the full potential of Streptomycesis the poor expression of biosynthetic gene clusters (BGCs) responsible for secondary metabolite production. This PhD project aims to contribute novel insights into the regulatory mechanisms of Streptomyces, paving the way for more effective manipulation and optimization of bioactive molecule production in Actinomycetes. Through an integrative multi-omics approach encompassing genomics, transcriptomics, and metabolomics data for a diverse set of 140 Actinomycetesstrains, we aim to investigate the regulatory mechanisms influencing the expression of BGCs and the production of secondary metabolites. Notably, this dataset stands out for Actinobacteria, as transcriptomics data collection traditionally centers on a small set of strains.

In summary, this project aims to study Streptomycesregulatory mechanisms through a multi-omics approach, providing a data-driven perspective on improving bioactive molecule production. The unique multi-omics dataset holds promise for advancing Actinomycetesunderstanding and contributing to the broader microbial secondary metabolite production landscape.

#013

Personal Information

Full name: Annelaura Bach Nielsen

Position: Postdoc

Department: Novo Nordisk Foundation Center for Protein Research

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Bridging Clinical Diagnostics with Data Science and Mass Spectrometry Advances: Collaborative Insights from Hospital and Academia

This collaboration between between Bispebjerg Hospital and the Mann Group at University of Copenhagen leverages machine learning and state of the art data science tools to analyze high-sensitivity mass spectrometry data from patient body fluid samples, with the goal of providing fast and accurate diagnostic support in the clinic.

Over the past 5 years we've established a new platform for advancing clinical decision support by setting up a proteomics facility at Bispebjerg Hospital, Copenhagen, Denmark. Diagnostic challenges identified by clinical experts at hospital departments across Copenhagen are brought to the facility for evaluation. After careful project assessment and agreement, patient material are collected and subsequently measured and analyzed at our clinical proteomics facility. To date, we have measured more than 14500 proteomes across 47 projects.

The collaboration between the hospital and a cutting-edge research lab within the field of proteomics benefits from the proximity of clinical expertise and the easy translation of advances in mass spectrometry and data science.

An example of an ongoing project within the facility is one that focuses on Lyme Neuroborreliosis diagnosis using plasma instead of CSF, where mass spectrometry coupled with machine learning has identified distinctive proteomic patterns, offering a less invasive and fast alternative for early detection.

This cross-sectional collaboration exemplifies the potential of merging clinical and data science expertise to advance diagnostic capabilities. Our findings will not only improve patient outcomes but also highlight the importance of interdisciplinary partnerships in healthcare innovation.

#014

Personal Information

Full name: Beatrice Dyring-Andersen

Position: Associate Professor

Department: Department of Dermatology and Allergy

Institution: Gentofte Hospital

Institution country: Danmark

Abstract

The spatial proteome of psoriasis

Analyzing the proteome of psoriasis provides the opportunity to characterize it at a deeper level, potentially leading to a broader understanding of the pathology of the disease. In this study, we used laser-capture microscopy to microdissect skin layers from both lesional and non-lesional psoriatic skin of eight patients with plaque psoriasis, and eight healthy donors. The proteomes of the samples were analyzed by mass spectrometry using data-independent acquisition. Furthermore, we cultured primary fibroblasts from lesional psoriatic skin and compared their proteomes to fibroblasts from healthy human skin. In total, we identified 6834 proteins from the microdissected skin layers (outer epidermis, inner epidermis, dermis and subcutis). The most notable differences were found in the inner epidermis when comparing lesional and non-lesional psoriatic skin (1651 differentially expressed proteins (DEPs), FDR<0.05), with high expression of alarmins, S100 proteins, and the IL-36 pathway. Additionally, our comparison of fibroblasts from psoriatic and healthy skin identified 3300 DEPs (FDR<0.05) including proteins involved in regulation of the cell cycle, TGF-beta production, and cell migration. These findings suggest dermal fibroblasts may potentiate inflammatory responses by recruiting immune cells.

In conclusion, our study underscores the potential of proteomic analysis and laser-capture microdissection to deepen our understanding of psoriasis pathology.

#015

Personal Information

Full name: Blaine Gabriel Fritz

Position: Assistant Professor

Department: Department of Immunology and Microbiology

Institution: Institute for Health and Medical Science

Institution country: Denmark

Abstract

Longitudinal, multi-omic analysis of lesion development and treatment in atopic dermatitis

Atopic dermatitis (AD) is a common, chronic inflammatory skin condition characterized by recurring, itchy lesions. AD's pathophysiology involves a complex interplay of host immune response, inflammation, bacterial infection, environmental factors and timing. The temporal relationships between changes in the host gene-expression, microbiome, and how they relate to a flare are not understood. This study followed a group of 40 participants for two weeks, starting with the development of a flare (Week 1) and throughout the application of topical treatments (Week 2). Paired samples were collected from lesional and non-lesional skin areas for RNA-sequencing (n=530), 16S amplicon sequencing (n=889), lipid-omics (n=857), cytokine/natural-moisturizing factor (NMF) measurements (n = 889), as well as full, clinical metadata. The aim of this study is to use this large, multi-modal dataset to identify global, temporal effects of changes in bacterial composition on host inflammation (and vice-versa). It also aims to explore the influence of specific bacterial species on flare development and treatment efficacy, while establishing a temporal map of patterns of changes in cytokine, lipid, and microbial abundances that characterize the development of AD flares.

#016

Personal Information

Full name: Boyang Ji

Position: Scientific Data Developer

Department: -

Institution: BioInnovation Institute

Institution country: Denmark

Abstract

Modeling the Human Gut Microbiome - Towards a Virtual Environment-Human-Microorganisms Model

Advances in metagenome sequencing of the human microbiome have provided a basis of new insights and revealed a close association of this complex ecosystem with a range of human diseases. However, there is limited knowledge about how the different members of the microbial community interact with each other and with the host, and we still lack basic mechanistic understanding of these interactions related to health and disease. Mathematical modelling had been demonstrated to be highly advantageous for gaining insights into the dynamics and interactions of complex systems in recent years, and several modelling approaches have been developed to enhance our understanding of the microbiome. Here, we will mainly discuss the constraint-based modeling approaches and their applications in human microbiome modeling. In addition, we will discuss how multi-omics data, modeling approaches and machine learning technology together can be applied to advance our understanding of environment-human-microorganisms interactions.

#017

Personal Information

Full name: Bunia Ingrid Adsersen

Position: BSc. Student in Life Science Engineering

Department: DTU Bioengineering

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Design, Production and Analysis of Minibinders against Gremlin-1

In the dynamic landscape of biomedical research, the fusion of machine learning (ML) and molecular biology has widened possibilities for scientific exploration. In our collaboration between DTU and Novo Nordisk, we explore de novo protein design for Gremlin-1 affinity reagents using generative AI. Gremlin-1 is a multifaceted protein implicated in conditions such as fibrosis, cancer, and diabetic nephropathy, making it a significant target for therapeutic intervention. Our project aims to generate ML-guided minibinders —small, engineered molecules—that effectively bind to Gremlin-1. ML algorithms guide binder optimization, enhancing specificity and affinity. Minibinders hold immense promise for biomarker assays, potentially revolutionizing disease diagnosis, prognosis, and therapeutic monitoring. The project involves minibinder design, protein expression, and functional characterization, all valuable methods in biotechnological research and development. We recognize that data is necessary for scientific progress. Our project emphasizes robust data management practices, ensuring the integrity, accessibility, and reproducibility of experimental results.

Our approach transcends disciplinary boundaries, weaving together diverse fields. We bridge the gap between wet-lab experiments and computational predictions. ML models aid in predicting minibinder-Gremlin-1 interactions, streamlining the design process in development-time and costs. By showcasing versatile technologies, methods, and algorithms presented by renowned researchers across academia and industry, we connect disciplines and institutions for de novo protein design.

#018

Personal Information

Full name: Carlos G. Acevedo-Rocha

Position: Senior Researcher (Group Leader)

Department: Novo Nordisk Foundation Centre for Biosustainability

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

EnzymeML: A markup language for FAIR enzymatic data.

The design of biocatalytic processes involves a large number of variables including enzyme kinetic parameters, reaction conditions and modelling method. Recently, the EnzymeML markup language was developed to make enzyme data findable, accessible, interoperable and reusable (FAIR). Here, we will show an example on how enzymatic data and metadata are collected and analysed. Standardisation of enzyme data using frameworks like EnzymeML is key to enhance data reproducibility and enable the full potential of machine learning for the data-driven design of biocatalytic processes.

#019

Personal Information

Full name: Cecilie Møller-Jensen

Position: BSc. Student in Biomedical Engineering

Department: Novo Nordisk Foundation Centre for Biosustainability

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Design, Production and Analysis of Minibinders against Gremlin-1

In the dynamic landscape of biomedical research, the fusion of machine learning (ML) and molecular biology has widened possibilities for scientific exploration. In our collaboration between DTU and Novo Nordisk, we explore de novo protein design for Gremlin-1 affinity reagents using generative AI. Gremlin-1 is a multifaceted protein implicated in conditions such as fibrosis, cancer, and diabetic nephropathy, making it a significant target for therapeutic intervention. Our project aims to generate ML-guided minibinders —small, engineered molecules—that effectively bind to Gremlin-1. ML algorithms guide binder optimization, enhancing specificity and affinity. Minibinders hold immense promise for biomarker assays, potentially revolutionizing disease diagnosis, prognosis, and therapeutic monitoring. The project involves minibinder design, protein expression, and functional characterization, all valuable methods in biotechnological research and development.
We recognize that data is necessary for scientific progress. Our project emphasizes robust data management practices, ensuring the integrity, accessibility, and reproducibility of experimental results.

Our approach transcends disciplinary boundaries, weaving together diverse fields. We bridge the gap between wet-lab experiments and computational predictions. ML models aid in predicting minibinder-Gremlin-1 interactions, streamlining the design process in development-time and costs. By showcasing versatile technologies, methods, and algorithms presented by renowned researchers across academia and industry, we connect disciplines and institutions for de novo protein design.

#020

Personal Information

Full name: David Westergaard
Position: Senior Researcher
Department: Department of Gynecology and Obstetrics
Institution: Copenhagen University Hospital Hvidovre
Institution country: Denmark

Abstract

Uncovering the Heritable Components of Multimorbidities and Disease Trajectories: A Nationwide Cohort Study

Quantifying the contribution of genetics and environmental effects on disease initiation and progression, as well as the shared genetics of different diseases, is vital for the understanding of the disease etiology of multimorbidities. In this study, we leverage nationwide Danish registries to provide a granular atlas of the genetic origin of disease phenotypes for a cohort of all Danes 1978-2018 with partially known pedigree (n = 6.3 million). We estimate the heritability and genetic correlation between thousands of disease phenotypes using a novel approach that can be scaled to nationwide data. Our findings confirm the importance of genetics for a number of known associations and increase the resolution of heritability by adding numerous novel associations, some of which point to shared biologically origin of different phenotypes. We also establish the heritability of disease trajectories and the importance of sex-specific genetic contributions.

Preprint with full author list available at https://www.medrxiv.org/content/10.1101/2023.02.08.23285642v1

#021

Personal Information

Full name: Ding He

Position: Research Data Steward

Department: Research Data Management

Institution: Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark

Institution country: Denmark

Abstract

Research Data Catalog – A FAIRer Data Portal at NNF Center for Biosustainability

Research data are the core for institutes and universities. Increasing demands to research data management (RDM) foster the need for tools to automate and digitalise workflows to support research projects. Adhering to best practices, policies and regulations is also an essential and integral part of such workflows. Importantly, research data should remain "FAIRly" operational even before openly accessible. To this end, we have built a Data Catalog (beta) on the foundation of the CKAN, an open-source data management system that serves as a software platform for managing research projects and data. Currently, it has:

- DTU single sign-on (SSO) integration
- Customed life science metadata schemas
- Customized dynamic workflows that standardize and streamline regulatory processes and domain-specific metadata registration
- Linked with the laboratory information management system (LIMS – Benchling)
- Backend storage linking to a Microsoft Azure Data Lake
- A flexible, semi-automatic generated data management plan

We are actively developing this product and plan to improve upon:

- Comprehensive UI/UX
- Integration with DTU infrastructure such as DTUDOCX and DTU Data
- Integration with Benchling and Microsoft Azure Data Lake
- Integration with automatic data science processing pipelines
- Integration with instrument-specific and public repository metadata

With the Data Catalog, our goal is to embed FAIR data practices in researchers' daily data operations with as effortless as possible. This should enable easier collaborations and data discovery, and ultimately to gain more value from the data we are generating.

#022

Personal Information

Full name: Elysia Gao

Position: Master's Student / Research Assistant

Department: Novo Nordisk Foundation Centre for Biosustainability

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Decoding Cell Behavior in Bioreactors: A Deep Learning Approach with a Focus on Latent Space Interpretability

Computational Fluid Dynamics (CFD) produces complex, high-dimensional time series data on bioreactor cell behavior, necessitating advanced methods for deriving clear biological insights. Our research employs machine and deep learning to simplify cell behavior analysis while enhancing latent space interpretability in autoencoders.

In order to study cell behavior in bioreactors, our strategy focuses on reducing the dimensions of metabolic data followed by downstream clustering and classification to understand patterns between cell lifelines and cyclical patterns within cell lifelines. Deep Autoencoders are utilized to feature engineer, reduce dimensions, while representing inputs well followed by k-means clustering and random forest classification to ultimately pinpoint crucial metabolic states, offering new insights that were difficult to achieve with traditional data analysis methods. Additionally, Dynamic Time Warping clustering allows us to group cells according to their metabolic behavior over time, highlighting diversity among cells. Despite their utility, autoencoders' latent spaces can be opaque. We use perturbation analysis and SHAP values for feature importance to demystify latent features' influences and metabolic state impacts.

Overall, these methods allow great reduction in data size while maintaining interpretability and not only improve our comprehension of heterogeneity within cell populations but also identify potential markers for evaluating bioreactor performance. Our findings provide a more detailed view of cell activity in bioreactors, allowing for better interpretability and new biological insights.

#023

Personal Information

Full name: Eric Bautista Farrerons

Position: MSc student

Department: DTU Health Tech

Institution: Denmark Technical University

Institution country: Denmark

Abstract

Integrating Genome-Scale Metabolic Models with Patient Plasma Metabolome to Study Endothelial Metabolism In Situ

Patient blood samples are invaluable in clinical omics databases, yet current methodologies often fail to fully uncover the molecular mechanisms driving patient pathology. While Genome-scale Metabolic Models (GEMs) show promise in systems medicine by integrating various omics data, having only exometabolomic data remains a limiting factor. To address this gap, we introduce a comprehensive pipeline integrating GEMs with patient plasma metabolome. This pipeline constructs case-specific GEMs using literature-based and patient-specific metabolomic data. Novel computational methods, including adaptive sampling and an in-house developed algorithm for the rational exploration of the sampled space of solutions, enhance integration accuracy while improving computational performance. Model characterization involves task analysis in combination with clustering methods to identify critical cellular functions. The new pipeline was applied to a cohort of trauma patients to investigates shock-induced endotheliopathy using patient plasma metabolome data. By analyzing endothelial cell metabolism comprehensively, the pipeline identified critical substrates and contributed to the development of targeted therapeutic strategies. Our study demonstrates the efficacy of integrating patient plasma metabolome data into computational models to analyze endothelial cell metabolism in disease contexts. This approach offers a deeper understanding of metabolic dysregulations and provides insights into diseases with metabolic components and potential treatments.

#024

Personal Information

Full name: Francesc Fernández

Position: Director

Department: -

Institution: Almirall

Institution country: Denmark

Abstract

Data Science in R&D at Almirall

At Almirall, we are a pharmaceutical company fully dedicated to medical dermatology with an ambition to become a global leader in the field. With an R&D pipeline solely focused on medical dermatology, we work closely with leading experts around the globe to innovate and develop novel technologies such as AI-based drug discovery, immunology, biological treatments, and mRNA (amongst others) we are at the forefront of advancing the scientific knowledge of skin and skin conditions. In this talk I will describe the main activities that the Data Science department works on to support the R&D activities in Almirall. This includes the contributions of the Data Science department in the Discovery area such as target identification, design of new molecules and indication expansion. Other applications of the Data Science department in Development will also be discussed such as the usage of digital sensors to better characterize the patients or the potential identification and usage of digital biomarkers.

#025

Personal Information

Full name: Hans Roubos
Position: Senior Science Fellow
Department: Science & Research
Institution: dsm-firmenich
Institution country: The Netherlands

Abstract

FAIR Data and Omics Technologies for a Deeper Dive into the Microbiome

FAIR data principles and omics technologies offer a holistic view into the world of the microbiome, unveiling unprecedented insights with diverse applications. In the realm of animal nutrition and health, these advancements empower researchers to evaluate the symbiotic relationships between microbial communities and host organisms, optimizing feed formulations and preventive health strategies for livestock. Similarly, in human nutrition and health, leveraging FAIR data and omics technologies provides a nuanced understanding of how the microbiome influences metabolism and immune function. This knowledge fuels the development of dietary supplements and probiotics to promote wellness and combat a spectrum of health conditions. Moreover, in personal care, this intersection enables the formulation of microbiome-friendly products that preserve skin ecosystems, promoting a balanced microbiota and enhancing overall skin health. In essence, FAIR data and omics technologies pave the way for a deeper exploration of the microbiome's multifaceted role across diverse domains, from animal and human health to personal care innovations.

#026

Personal Information

Full name: Ester Milesi

Position: LIMS administrator

Department: The Novo Nordisk Foundation Center for Biosustainability

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Building an Omics Data Infrastructure to Integrate Data Management and Data Science.

Advances in high-throughput technologies are allowing the generation of large amounts of data, and its usage is becoming a requirement for biological discovery. Additionally, the generated data can have multiple lives beyond the original purpose, increasing the importance of data reusability. However, the lack of standardized procedures in the capture and storage of OMICS data and related metadata makes data reusability difficult and introduces inconsistencies causing unnecessary delays in the data analysis.

To address these challenges, we propose an integrated data infrastructure designed to streamline the entire life cycle of OMICS data, from raw acquisition to interpretation. This infrastructure offers researchers a user-friendly platform to create projects and associated experiments metadata, that are then linked to the generated data. The capture of structured and comprehensive metadata allows automated and replicable analysis of data, making use of community workflow management tools.

The described infrastructure adapts to diverse computational environments, ensuring its scalability across different research settings. By providing standardized procedures, easy-to-use interfaces and robust data lineage, our proposed infrastructure aims to enhance data management efficiency and accelerate biological discoveries.

#027

Personal Information

Full name: Henry Webel

Position: Postdoc

Department: Novo Nordisk Foundation Center for Basic Metabolic Research

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Mass spectrometry-based proteomics data from thousands of HeLa control samples

We provide a curated, large scale, label free mass spectrometry-based proteomics data set derived from HeLa cell lines for general purpose machine learning and analysis. Data access and filtering is a tedious task, which takes up considerable amounts of time for researchers. Therefore we provide machine based metadata for easy selection and overview along the 7,444 raw files and MaxQuant search output. For convenience, we provide three filtered and aggregated development datasets on the protein groups, peptides and precursors level. Next to providing easy to access training data, we provide a SDRF file annotating each raw file with instrument settings allowing automated reprocessing. We encourage others to enlarge this data set by instrument runs of further HeLa samples from different machine types by providing our workflows and analysis scripts.

#028

Personal Information

Full name: Igor Marin de Mas

Position: Senior Researcher and PI

Department: -

Institution: Copenhagen University Hospital, Rigshospitalet

Institution country: Denmark

Abstract

A Protocol for the Automatic Construction of Highly Curated Genome-Scale Models of Human Metabolism

Genome-scale metabolic models (GEMs) are a powerful tool for understanding human metabolism from a holistic perspective, with high relevance in the study of many diseases and in the metabolic engineering of human cell lines. However, GEM building can be challenging due to the limitations of either automated processes that lack manual refinement and result in inaccurate models or manual curation which is a time-consuming process that limits the continuous update of reliable GEMs. To overcome these limitations, we present a novel algorithm-aided protocol that facilitates continuous updating of highly curated GEMs. The algorithm enables the automatic curation and/or expansion of existing GEMs or generates a highly curated metabolic network based on the current information retrieved from multiple databases, such as KEGG, Reactome, BiGG, and HMR, in real time. This tool was applied to the latest reconstruction of human metabolism (Human1) generating a series of the human GEMs that improve and expand the reference model and generate the most extensive and comprehensive general reconstruction of human metabolism to date. The tool presented here goes beyond the current state-of-the-art and paves the way for the automatic reconstruction of a highly curated up-to-date GEM. It provides a clear and transparent step-by-step view of the data treatment process to generate highly curated and functional GEMs in an unsupervised manner. This tool has high potential in computational biology as well as in multiple fields of biological science where metabolism is relevant, such as biomedicine, biotechnology, and bioengineering

#029

Personal Information

Full name: Isabel Díaz-Pinés Cort

Position: PhD student

Department: Costerton Biofilm Center, Department of Immunology and Microbiology

Institution: University of Copenhagen

Institution country: Denmark

Abstract

The Inflammatome: A Meta-Analysis of Human Genes Regulated During Inflammation

Inflammation, a hallmark of human pathology, is a complex biological process that triggers changes in the expression of numerous genes. These changes can dominate the statistical signal detected in differential expression analysis of case-control studies and obscure the more interesting tissue- or disease-specific changes in gene expression. In this work, we exploit the inflammatory signal in publicly available clinical transcriptomics and proteomics datasets to derive the inflammatome, a set of transcripts and proteins consistently upregulated in various inflammatory conditions affecting a range of tissues compared to healthy controls. We propose an inflammation score based on the expression of the inflammatome and demonstrate its correlation to clinical disease severity scores in independent datasets and new diseases, showcasing its robustness. The inflammatome provides an objective proxy for disease severity and will help future studies distinguish disease-specific changes in gene expression from those solely caused by inflammatory processes, thereby aiding biomarker and drug target discovery.

#030

Personal Information

Full name: Jakob Berg Jespersen
Position: Data Scientist
Department: Novo Nordisk Foundation Centre for Biosustainability
Institution: Technical University of Denmark
Institution country: Denmark

Abstract

Automating an R package with Nextflow for seamless mass spec report generation from label free output using the MS-DAP pipeline seamlessly and with no need for user interaction.

Advancement in Science is dependent on processing of data to knowledge. Mass spectrometry is an abundant technique, and thus many tools exist for data interpretation. One such tool is the MS-DAP pipeline which turns output from label free mass spec experiments into comprehensive PDF reports. Usage of the R package does however require bits of interaction in an R session. We have here written a fully automated wrapper in Nextflow to launch the otherwise interactive parts of this R package. By leveraging on the Nextflow capabilities, and designing scripts with this in mind, the pipeline can be deployed on a variety of hardware and cloud systems with convenient ways of reproducibility locally, or remotely in a very scalable way. We have successfully configured the pipeline to run on Microsoft Azure Batch Service and are working to make data submission as seamless as possible.
Running containerized scripts that are all available in open-source repositories, reduces friction that can otherwise be involved in installing and potentially debugging software of various versions.

Attached relevant imagery is a flow chart of data processing by our Nextflow pipeline in the popular Nextflow metro map style.

References:

F. Koopmans, et al. MS-DAP Platform for Downstream Data Analysis of Label-Free Proteomics Uncovers Optimal Workflows in Benchmark Data Sets and Increased Sensitivity in Analysis of Alzheimer's Biomarker Data. J. Proteome Res. 2023, 22, 2, 374–386 (2022) doi:10.1021/acs.jproteome.2c00513
P. Di Tommaso, et al. Nextflow enables reproducible computational workflows. Nature Biotechnology 35, 316–319 (2017) doi:10.1038/nbt.3820

#031

Personal Information

Full name: Jakub Vašíček

Position: PhD student

Department: Department of Clinical Science

Institution: University of Bergen

Institution country: Norway

Abstract

Haplotypes and Human Diversity in Proteomics

Genomic research has long benefited from using diverse population panels, increasing the statistical power of association studies for admixed populations. However, proteomic workflows often project all data to a set of reference protein sequences. Consequently, we obscure a portion of the proteome, restricting our ability to fully analyze complex samples. Moreover, we risk introducing a bias against populations with a different haplotypic structure.

Alleles co-occurring in the protein-coding regions of the same gene produce a unique protein sequence - protein haplotype. These haplotypes are present in biological samples, and detectable by mass spectrometry. We have demonstrated that thousands of amino acid substitutions can be discovered in a single sample, sometimes featuring alleles in linkage disequilibrium within the same peptide after a tryptic digestion of the protein. We have recently released ProHap, a bioinformatic pipeline that allows building proteomic databases from genetic reference panels.

We generated proteomic databases from the 1000 Genomes Project and showed that participants of the African superpopulation diverge from the reference proteome more than others, while all the included ancestry groups show notable differences from the reference proteome. ProHap alleviates this bias by creating databases that capture the diversity of human proteomes and allows the fair competition of protein haplotypes during proteomic searches. The pipeline can be run on public as well as local reference panels, with great flexibility in terms of types of genetic variants and haplotype frequency, empowering researchers to tailor their proteomic studies to populations.

#032

Personal Information

Full name: Jean-Marie Mouillon

Position: Science Manager

Department: Production Strain Technology

Institution: Novonesis

Institution country: Denmark

Abstract

Feeding the world on CO2 → acetate → food

According to a recent report led by the UN, more than 250 million individuals faced severe hunger in 2022. To address this growing issue, it is vital that weestablisha robust bioeconomy to combat significant global challenges such as malnutrition, biodiversity loss, and climate change. With this goal in mind,the Bill& Melinda Gates and the Novo Nordisk Foundation are funding a consortium where Power to X and biotechnology are joining forces to use CO2 to produce human food protein. Here, CO2 will be converted to acetate by electrolysis and acetate will serve as a fermentation substrate for protein production for food. Large-scale microbial fermentations can produce massive quantities of protein, decoupled from land-useand serve as an alternative to animal protein.

For decades, industrial production strains, such as filamentous fungi, used for fermentation have been fine-tuned for growth on glucose. However, transitioning to acetate as a carbon source demands a complete rewiring of their central metabolism. To accelerate the development of these production strains, we employ adaptive evolution and genome-scale metabolic modeling. Population genomics, transcriptomics, proteomics, and metabolomics are employed to understand the key changes driving phenotypic traits linked to acetate growth and increased production titer. Finally, this multivariate data will be combined with ML and refine mechanistic models, enabling the identification of genetic targets for metabolic engineering and the development of stable production strains.

Ultimately, thetechnology developed in this project has broad perspectives for production of other biomolecules,chemicalsandbiofuels.

#033

Personal Information

Full name: Jesper Lauridsen

Position: BSc Student

Department: DTU Compute

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Advancing InstaNovo: Database-free de novo Proteomic Analysis for Post-Translational Modifications

In the field of mass spectrometry-based proteomics, the commonly used target-decoy (TD)database strategy presents inherent drawbacks. Because it relies on a database search,identification of sequences, different to the ones already known, is difficult. This hinders acomprehensive proteomic analysis. It also presents a significant computational cost -particularly if the search includes Post-Translational Modifications (PTMs).

InstaNovois a database-free model which significantly exceedsstate-of-the-artperformance on de novo peptide prediction. This projectseeksto further developInstaNovoby investigating whether training on a larger dataset or with more parameters will increaseperformance. It will also investigate the feasibility of detecting PTMs by increasing the modelvocabulary. Potentially, data from different fragmentation methods could be included, and themodel could be fine-tuned to the specific task of PTM detection.

#034

Personal Information

Full name: John Shorter
Position: Assistant Professor
Department: Department of Science and the Environment
Institution: Roskilde University
Institution country: Denmark

Abstract

Genome-wide association study using health registers in harmonized Nordic cohorts reveals different genetic architecture between early- and late-onset depression

Background:
Major depressive disorder (MDD) is a common and heterogeneous mental disorder. Heterogeneity in symptoms, severity, and response to treatment hampers our ability to provide a biological characterization of MDD. Studying specific subtypes of more homogeneous patient groups may increase our chances of identifying the underlying genetic causes of the disorder and facilitate our efforts of designing targeted treatment strategies.

Methods:
We investigated differences in the genetic architecture of early onset (eoMDD) and late onset (loMDD) depression; two MDD subtypes with potential etiological differences. We performed a meta-analysis using nine cohorts from five countries with detailed electronic healthcare registries and accumulated around 46K early onset (eoMDD, ≤25 years old at diagnosis) and 37K late onset (loMDD, ≥50 years at diagnosis) cases to compare the genetic architecture between these two age-related subtypes of MDD.

Results:
We found that eoMDD and loMDD shared no genome-wide significant loci ($P<5x10-8$), that genes associated with eoMDD have a role in neurodevelopment and suicide, that the SNP-heritability for eoMDD was nearly twice as high as loMDD, the two subtypes only had a modest genetic correlation, and the polygenic score for eoMDD is a significant predictor for suicide.

Conclusion:
This study is the first to identify genetic loci differentially associated with early versus late-onset MDD. The development and distribution of software for statistical genetics analysis, alongside with their respective harmonized reference data and scripts was critical for this analysis. These data management tools are publicly available through GitHub.

#035

Personal Information

Full name: Jonathan Funk

Position: PhD student

Department: Antibody Technologies & Novo Nordisk Foundation Center for Biosustainability

Institution: DTU Bioengineering

Institution country: Denmark

Abstract

ProteusAI: Bridging Machine Learning and Protein Engineering Through User-Friendly Applications

The centrepiece of the project is an intuitive app, designed to serve as a bridge between complex ML methods and protein research. The app is built to support a wide array of users, particularly those with little to no computational training, allowing them to perform Machine Learning Guided Directed Evolution (MLDE), and de novo protein design workflows without the need to code. For MLDE workflows, users are presented with two options: 1) to train ML models from existing experimental data, and 2) to use ML models to guide the design of an initial mutant library. Both options offer the possibility to fine tune powerful ML models, such as pre-trained Protein Language Models (PLMs) with millions of model parameters. Access to large pre-trained ML models enables the extraction of complex features from data to make stronger predictions for future experiments. For protein design workflows, the app offers access to geometric deep learning models, to optimize sequences for a given structure. These models are often used to enhance the stability or expressibility of proteins. The protein design workflows allow users to provide additional knowledge, such as evolutionary conservation patterns or the annotation of active site residues which should not be re-designed.

The current goal for ProteusAI is to publish the application as a web application with the additional option to run the app on a local computer. This development strategy addresses a significant barrier in the adoption of computational methods and simultaneously allows users to run the application locally if they have data security concerns.

#036

Personal Information

Full name: Julia Villarroel

Position: Senior Bioinformatician

Department: Discovery

Institution: ZS Discovery

Institution country: Denmark

Abstract

Novel algorithm-driven vaccine design framework, built to ensure optimal coverage of both population diversity and pathogen variation, aimed at eliciting strong cellular and humoral response, as well as lasting immune memory.

The vaccine framework aims to elicit strong and relevant immune responses in diverse populations and protect against most infectious variants using minimal active components.

We begin by defining the HLA allele distribution of the target population using resources like allelefrequencies.net, and by accounting for the known genetic variation of the pathogen.

In the context of viral infections, we select proteins essential for initiating and maintaining infection, categorizing them as either external/secreted or exclusively intracellular.

MHC binding prediction algorithms are used to identify CD8+ T cell epitopes from all proteins that bind to HLA-encoded MHC class I, and CD4+ T cell epitopes from external/secreted proteins that bind to MHC class II.

Ideally, predicted T-cell epitopes undergo in vitro validation to confirm their immunogenicity. Subsequently, an algorithm iteratively selects validated epitopes for MHC classes I and II, optimizing for population coverage based on HLA allele frequencies, and optimal coverage of pathogen genetic variation. For MHC class II, selected epitopes are elongated to 30 amino acids to resemble native protein folding, inducing B-cell responses and enhancing T-cell memory.

Patent pending, WO2023021056A1.

#037

Personal Information

Full name: Kai Blin

Position: Senior Researcher

Department: Novo Nordisk Foundation Center for Biosustainability

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Maintaining and Improving Community-Driven Databases: Lessons from the Minimum Information about Biosynthetic Gene Clusters (MIBiG) Database

Small bioactive compounds produced by microbes form the basis of almost three quarters of drugs, as well as about half of the agrochemicals currently on the market. Traditionally, new compounds were discovered using a "find and grind" workflow of extracting from natural sources, chemically isolating compounds, followed by purification and activity testing. With the wide availability of genome data, this approach is now routinely complemented by mining genomes and metagenomes to identify natural product biosynthetic pathways. Software tools to perform this genome mining have existed for over a decade. Since its 2011 release, the antiSMASH tool developed in our group has established itself as the gold standard tool for microbial natural product genome mining.

From the beginning, antiSMASH provided a way to compare identified gene clusters to a collection of known gene clusters. It quickly became obvious that there was a wide community interest in making a curated list of known gene clusters accessible, and in 2015, over 150 members of the community came together to create an annotation standard, and collect over 1,000 literature-backed, experimentally verified gene clusters. In the past years, that database has grown to contain over 2,500 entries through community-run annotation events, called annotathons. We are currently preparing for the fourth major release of the database, with over 300 people signed up for this round of annotathons.

Here I will share the challenges and opportunities we have encountered in the roughly ten years of organising a community-annotation-based database.

#038

Personal Information

Full name: Karina Banasik

Position: Associate Professor, Visiting professor

Department: -

Institution: Novo Nordisk Foundation Center for Protein Research

Institution country: Denmark

Abstract

Multi-Omics studies in Women's Health and reproduction

Women's health and reproduction faces unique challenges due to complex interplay between genetics, environment, and hormonal fluctuations. Traditional single-omics approaches often provide limited insights. In our projects we deploy a wealth of multi-omics strategies, integrating data from genomics, transcriptomics, proteomics, and metabolomics, to advance our understanding of women's health and disease. I will present some of our ongoing studies and address challenges associated with study design, data integration, standardization, and the need for studies encompassing diverse populations. Multi-omics studies, although expensive to conduct, hold immense promise for revolutionizing women's healthcare by enabling comprehensive analysis of complex biological processes and paving the way for precision womens health.

#039

Personal Information

Full name: Kübra Altinel

Position: Postdoc

Department: Faculty of Health and Medical Sciences

Institution: Biotech Research and Innovation Centre (BRIC), University of Copenhagen

Institution country: Denmark

Abstract

Searching for the specialised ribosomes with deep learning

Ribosomes are the main drivers of protein synthesis as they translate the genetic information into proteins. They are large macro-molecules, consisting of 4 ribosomal RNAs (rRNAs) and 80 ribosomal proteins (RPs). Despite their importance as an essential hub in gene expression and their high level of complexity, they are commonly regarded as neutral apparatuses of translation with the same overall composition and no intrinsic regulatory capacity. However, the proof for heterogeneous ribosomes is accumulating: ribosomes can have heterogeneity based on their variation in rRNA sequence, protein composition, and RNA modification patterns. Whether heterogeneous ribosomes acquire specialised translational activity is debated, as their significance in correct execution of gene expression programs remains to be understood. Here, we aim to investigate the individual and combinatorial impact of 2'-O-methylation (2'-O-me) on ribosome function and translation by using single molecule nanopore sequencing and a semi-supervised variational autoencoder (SS-VAE) written in the deep probabilistic programming language Pyro. By categorising the ribosomes into subtypes based on their 2'-O-me combination patterns, this work will expand our knowledge of if and how ribosome subtypes conduct specialised translation. As altered translational programs are essential for the process of differentiation, we focus here on early stages of organismal development as modelled through the directed differentiation of human embryonic stem cells. We anticipate that the results will enhance our knowledge of ribosome mediated translational control and central mechanisms of cell fate decision making.

#040

Personal Information

Full name: Laura Machado

Position: MSc Bioinformatics

Department: Biosustain & Bioengineering

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Machine learning-guided library design for protein engineering

Protein engineering (PE) holds transformative potential across diverse fields, including biosustainability, food production, therapeutics, and diagnostics. The goal of the engineering endeavor is to find variants with improved fitness: efficacy of performing a desired function. The process of identifying optimized protein variants involves designing, synthesizing, and testing protein variant libraries. The design task is non-trivial, as it is infeasible to test all possible combinations in a laboratory setting. This is because the number of possible amino acid arrangements scales exponentially with the protein length: the number of possibilities for a 100 amino acid protein would exceed the number of atoms in the observable universe. Machine Learning (ML) tools can be used to address this combinatorial challenge if used to guiding the decisions of which variants to try in the lab. We can use available data on the single mutations effects to train ML models to estimate the mutational effects on protein fitness on higher order mutants. In the context of MLPE, sampling refers to the process of selecting a subset of protein variants to test and analyze. The goal of this project is to identify optimal sampling strategies, that can be used in conjunction with mutational effect prediction to guide the library design for future experiments. By exploring different ML models and sampling strategies we aim to reduce the minimum necessary number of experiments, leading to more efficient and sustainable research practices.

#041

Personal Information

Full name: Laura Pikkupeura

Position: Postdoc

Department: Novo Nordisk Foundation Center for Basic Metabolic Research

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Cellular and genetic heterogeneity in steatotic liver disease development and progression

Steatotic liver disease is the most common cause of liver-related illness and death worldwide and is linked to obesity, diabetes, and alcohol use, but little is known of why patients follow very different paths - some develop cirrhosis, while others never experience worsening despite decades with steatosis. The cellular and genetic underpinnings of the heterogeneity among patients remain poorly characterised. To disentangle this heterogeneity in SLD development, we integrate in-depth multimodal single cell characterization of liver biopsies and population-scale human genetic studies in deeply phenotype cohorts.

We have generated the largest multi-modal single-cell atlas of gene regulation (gene expression and chromatin accessibility) in the human liver consisting of 300.000 cells from 60 biopsies from patients at different stages of SLD to investigate the disease progression in terms of changes in cell type abundance, transcription and chromatin accessibility.

Combining the cell type-specific chromatin accessibility data and genetic data, we have calculated cell-type specific polygenic scores for liver related traits. Rather than a genome-wide polygenic score (PGS), we can use the chromatin accessibility to calculate a polygenic score for each cell type, allowing for multi-dimensional stratification of patients and identification of individuals with genetic variants predominantly affecting specific cell types. We have applied the cell type-specific PGSs to patient and population cohorts with detailed molecular and clinical phenotypes to investigate phenotypic consequences of the cell type-specific genetic risk underlying SLD heterogeneity.

#042

Personal Information

Full name: Luca Gaessler

Position: Research Assistant

Department: Oncology-Pathology

Institution: Karolinska Institute

Institution country: Sweden

Abstract

Proteomic deconvolution of the therapy response in HER2-positive breast cancer patients

(Ongoing research, not published)

According to the WHO, breast cancer caused almost 700,000 documented global casualties in 2020 and showed the highest incidence of all cancer types in that year. Current treatment decisions for breast cancer patients consider the presence of tyrosine kinase receptor ERBB2 (HER2) as a specific target on tumor cell surfaces.

However, only approximately half of the patients respond to the standard therapy for HER2-positive breast cancer, chemotherapy combined with anti-HER2 targeted drugs. Here, precision medicine presents an effective approach to improve therapy decisions by considering each patient's molecular response to drug treatment. Therefore, the primary goal of my Master's thesis project is to employ proteomic methods to identify biomarkers that can differentiate responders from non-responders.

In this context, I will utilize breast cancer samples obtained from a translational clinical trial to further the biological understanding of the anti-HER2 treatment response. The trial has collected biopsies at different disease stages, providing a unique opportunity to dynamically track the patients' responses over time. Insights at the proteome level will be complemented by DNA and RNA sequencing data to provide further understanding of genotype-phenotype associations.

One of the long-term goals of this project is to develop proteomic assays for the detection of biomarkers and their practical implementation in a clinical setting. In future, information derived from various systematic assays can assist clinicians in making more informed treatment decisions, thereby minimizing the risk of adverse effects associated with cancer treatment.

#043

Personal Information

Full name: Lukas Huschet

Position: PhD student

Department: Systems Chronobiology

Institution: Ludwig Maximilian University (LMU)

Institution country: Germany

Abstract

Circadian dynamics of the nuclear proteome in time and space and its modulation by metabolic state

The liver, as one of the key organs regulating metabolism, has a prominent circadian control, ranging from glucose and lipid regulation to detoxification and bile acid metabolism (Reinke and Asher 2016). Because of this wide spectrum of metabolic functions, the liver and its clock are susceptible to changes that lead to pathology when external alterations occur. Evidence in mice has shown that a high fat diet (HFD), disturbs the clock of the liver at the transcriptional and metabolic level (Eckel-Mahan et al. 2013), but little is known in regards to protein level in the context of metabolic disruption. Because of this, we sought to investigate the daily dynamics of the mouse liver nuclear proteome under a HFD regime as a model of metabolic disruption. To investigate this, we collected livers across a 48-hour time course from mice fed a control diet (CD) or HFD for 10 weeks, and, to determine nuclear protein rhythms, we combined our optimized nuclei enrichment protocol with state-of-the-art mass spectrometry-based quantitative proteomics using data independent acquisition methods. Our preliminary results show that CD-fed animals have significantly less oscillating nuclear proteins compared to HFD-fed animals(1059 vs 1721)and that some of the cycling proteins present in both diets change their rhythmic properties, indicating the ability of metabolic disruption to alter the oscillating nuclear proteome of the mouse liver.

#044

Personal Information

Full name: Luke W. Johnston
Position: Team Leader
Department: NA
Institution: Steno Diabetes Center Aarhus
Institution country: Denmark

Abstract

The Seedcase Project: A software- and training-based initiative to develop data engineering tools and skills for open and FAIR infrastructures of research health data

Across much of science, we suffer from a lack of software and data engineering resources, skills, and research-specific tools. This has negative impacts on our abilities to generate reliable results and be productive and effective as researchers. Historically, there has been no or limited sources of funding that allow allocating enough resources for personnel with these technical skills. This is especially an issue for small- to mid-sized clinical and health research groups. Part of the issue is a lack of a general awareness by researchers, funding agencies, and organizations on the importance of software and data engineering in research, as well as the lack of systemic incentive structures to dedicate resources and time to these practices. This has led to an overall low capacity for these skills within research settings. The consequence of which is the many organizational and technical challenges with managing, sharing, building, and using research data, not least of which is the impact on our ability to follow FAIR and open science practices.

The aim of the Seedcase Project (<seedcase-project.org>) is two-fold: To develop software to build and manage infrastructures for data designed by researchers, for researchers; and, to create training material to upskill researchers in modern software and data engineering practices to more effectively tackle the challenges we currently face and will continue to face in the future. Ultimately, we work towards helping researchers do “better science in less time”, not only those in academic settings in Denmark but in industry and across the globe.

This project is funded by an NNF grant (number NNF21OC0069462).

#045

Personal Information

Full name: Malte Thodberg

Position: Postdoc

Department: Novo Nordisk Foundation Center for Basic Metabolic Research

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Partitioning the Genetic Risk of Obesity and Type 2 Diabetes Using Single-cell Multi-omics of the Gastrointestinal Tract.

The worldwide increase in the prevalence or obesity and Type 2 Diabetes (T2D) constitute a major threat to global health, however the complex nature of these diseases is severely hampering the development of prognostic tools and clinical interventions. Both T2D and obesity have a strong genetic component, yet the functional impact of the identified genetic variants in terms of causal genes, cell types and organs remains largely unknown.

The gastrointestinal (GI) tract is located at the junction between the gut microbiome and the gut-brain axis, which links intestinal function to the central nervous system resulting in changes in appetite and behavior. Due to the complex regionality and cellular composition of the GI-tract, it is largely unknown to what extent genetics play a role in metabolic diseases mediated by the GI-tract, in particular obesity and T2D.

We are performing single-cell sequencing on human biopsies from multiple locations in the colon using the 10x multiome assay for gene expression and chromatin accessibility. This allows us to map the location, activity and interaction of promoters and enhancers in individual cell types and integrate these with genome-wide association studies (GWASs) of obesity and T2D.

Using the resulting multi-omic and cell type-specific map of gene regulation, we can i) partition the genetic risk of obesity and T2D across cell types, ii) prioritize causal genetic variants and genes and iii) perform data-driven stratification of obesity and T2D patients based on their cell type-specific polygenic risks of T2D. Ultimately, this will lead to novel molecular clinical targets and non-invasive prognostic tools for obesity and T2D.

#046

Personal Information

Full name: Maria Dalby
Position: Director
Department: Bioinformatics
Institution: Muna Therapeutics
Institution country: Denmark

Abstract

Cellular resilience to Alzheimer's disease pathology by spatial and single-cell transcriptomics

Alzheimer's disease (AD) is characterized by neuropathological hallmarks including amyloid plaques and tau tangles and neurodegeneration in the frontal cortex. While there is a strong association of tau pathology with neurodegeneration and cognitive decline, around 30% of amyloid plaque-carrying seniors remain cognitively healthy. We hypothesize that protective cellular mechanisms upstream of tau phosphorylation and neurodegeneration may underlie this resilience.

We apply spatial transcriptomics, histology for amyloid beta and pTau, and single nuclei RNA sequencing on postmortem prefrontal cortex from >50 individuals, characterized by amyloid pathology with (AD+Dem) or without (AD-Dem) cognitive deficits. Findings are validated using Xenium to reveal the cellular resolution of transcripts of interest and their spatial relationship to pathology in situ.

We identified 54 cellular subpopulations and several gene co-expression networks that were differentially responsive to amyloid beta and pTau pathology in the presence or absence of dementia, which may play important roles in cellular resilience to amyloid-induced tau pathology.

Spatial transcriptomics from human brain tissue is a powerful approach to identify cellular mechanisms of resilience towards neuropathology. These mechanisms represent a rich source for the identification of novel therapeutic targets for the treatment of AD

#047

Personal Information

Full name: Marietta Kokla

Position: Postdoc

Department: Klin Translationel Forskning

Institution: Steno Diabetes Center

Institution country: Denmark

Abstract

Understanding the Endocannabinoid System and its Role in Diabetic Neuropathy of T1D Patients

Background:
Type 1 diabetes (T1D) is a severe autoimmune disease affecting millions worldwide, characterized by the immune system's attack on insulin-producing pancreatic cells. Diabetic sensorimotor polyneuropathy (DSPN) is a common complication T1D, impairing peripheral nerves and sensory functions. Ongoing research focuses on Endocannabinoids (eCBs), which are endogenous lipids that activate cannabinoid receptors that regulate various physiological processes, including pain sensation, inflammation, and glucose metabolism.Aim:This study aims to identify novel endocannabinoid system (ECS) biomarkers for DSPN.

Methods:
Serum samples from T1D patients with and without DSPN analyzed with Targeted Liquid Chromatography-Mass Spectrometry (LC-MS). Additionally, proteomics data were integrated for pathway enrichment. Statistical analysis explored associations between endocannabinoids, metabolites, proteomic profiles, and clinical parameters, providing insights into DSPN mechanisms.Results:Our results revealed differences in the serum levels of certain endocannabinoids, lipids, and hormones between T1D patients with DSPN and those without DSPN. Several of these lipids and hormones were found to be significantly correlated with the presence of DSPN in T1D patients, which suggests potential mechanisms underlying neuropathic complications.

Conclusion:
These findings highlight the importance of the endocannabinoid system in the pathogenesis of DSPN and provide novel insights into the metabolic dysregulation associated with this complication. Further research is needed to validate these findings and explore their clinical implications for diagnosing and managing neuropathic complications

#048

Personal Information

Full name: Martin Miller
Position: Senior Director
Department: Oncology Data Science
Institution: AstraZeneca
Institution country: United Kingdom

Abstract

Modelling multi-omic, real-world data reveals immunogenomic drivers of resistance to cancer immunotherapy

Why some cancer patients fail or have short lived response to immune checkpoint blockade (ICB) immunotherapy remains largely unknown. At AstraZeneca's Oncology Data Science, we committed to unlock the potential of AI/ML-driven data science to address such hard to answer questions. Here, we applied ML and bioinformatic methods to analyse clinical endpoints together with >10.000 of DNA and RNA profiled samples in the Tempus real-world database to identify the immunogenomic drivers of acquired and primary resistance to ICB in lung, breast, bladder and head and neck cancers. Post-ICB, acquired resistant patients showed a significantly inflamed tumour microenvironment (TME) characterised by higher estimation of infiltration of T cells and myeloid cells and higher activation of interferon gamma signalling as compared to primary resistant patients. ML-based pseudotime analysis of bulk tumour RNA confirmed this observation in independent cohorts. In addition, in post-ICB acquired resistance, we observed selection for mutations in genes involved in known immunomodulatory pathways including Hedgehog and Notch pathways. In summary, acquired and primary ICB resistant patients have distinct clinical and molecular features at progression. Their tumours' TME is fundamentally different with acquired resistance TMEs being infiltrated with immune cells albeit escaped post progression. In addition, ICB selects mutations in immune escape pathways. This data science-based, multi-modal analysis of post therapy biopsies has given insights for patient selection strategies and provides rational into combination treatment options for acquired resistant patients.

#049

Personal Information

Full name: Matin Nuhamunada
Position: PhD Student
Department: Novo Nordisk Foundation Centre for Biosustainability
Institution: Technical University of Denmark
Institution country: Denmark

Abstract

Leveraging Modern Data Stack in Box for Natural Product Genome Mining in Small-Scale and Private Strain Collection

The advent of third-generation sequencing technologies, especially Oxford Nanopore Technology, enables individual researchers and small laboratories to affordably create and manage private microbial strain collections. This shift promises to accelerate natural product discovery by facilitating the mining of biosynthetic gene clusters (BGCs) from genomic sequences, an important step in unlocking novel pharmaceuticals, agrochemicals, and other industrially relevant compounds. As researchers embark on building and analyzing their own private collections, the challenge extends beyond managing large-scale public genomic datasets but also in providing solutions that cater to the analysis of smaller, more focused collections. Here, we present BGCFlow, a comprehensive genome mining workflow for the analysis of bacterial pangenomes. BGCFlow integrates a "modern data stack in a box," leveraging tools such as dbt, DuckDB, and Metabase to offer streamlined data engineering pipeline and efficient platform for the exploration and management of private strain collections. Each tool is selected for its unique capabilities: dbt for transforming data with simplicity and reproducibility, duckdb for its lightweight, in-process SQL database that facilitates fast analytical queries, and metabase for its user-friendly interface allowing both data scientists and lab researchers to visualize and interact with their data. By doing so, we aim to bridge the gap between the potential of genome mining and the practicalities of conducting such research at a scale that is both manageable and accessible to a broader scientific community.

#050

Personal Information

Full name: Matthias Mattanovich

Position: Postdoc

Department: Novo Nordisk Foundation Center for Basic Metabolic Research, SUND

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Comparison of approaches to multi-omics data analysis for systems biology: lipolysis in brown adipocytes

The integration of omics technologies into molecular biological research has transformed it into a field where a system can be investigated as a whole by combining information from multiple biochemical layers. This extension of traditional molecular biology, where various data sources are combined, offers novel insights into complex molecular interactions. To illustrate different approaches to how multi-omics data can be integrated, we present illustrative analyses of murine brown adipocytes undergoing thermogenic lipolysis. Furthermore, a preliminary classification of approaches, ranging from correlation-based techniques to network-driven interpretations, to multi-omics data analysis are presented.

#051

Personal Information

Full name: Michaël Pierrelée

Position: Postdoctoral researcher

Department: Novo Nordisk Foundation Centre for Biosustainability

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Metagenomics-based metabolic models enabling optimization of environmental biotechnologies – the case of methanotrophy

In Denmark, urban areas produced 2.32 million of m3 of wastewater. As it contains high levels of nutrients (eg nitrogen) and organic micropollutants (eg pharmaceutical compounds), wastewater can harm ecosystems and human health. To remove them, public services rely on complex microbial communities in treatment plants, each species being specialized in one or multiple degradation steps. For example, methanotrophic communities degrade micropollutants while metabolizing methane.

Yet, it is unclear which microbes are the main contributors to biodegradation and which conditions maximize it. It results in consuming more energy and chemicals for alternative treatments, while wastewater production is still increasing. Therefore, the interactions between microbes need to be modeled.

Normally, information from omics data tends to be neglected for process design and optimization. Static models are applied with few species or metabolic reactions. Indeed, flux balance analysis (FBA), that solves metabolic models, assumes steady-state and is complex with multiple species.

In this project, we aim to identify key interactions between species by developing dynamic hybrid models that combine metabolic and macroscopic reactions. We will apply them to methanotrophic communities, which only use methane as an energy and carbon source. To build metabolic models solved by FBA, we will use shotgun metagenomics, time-course metatranscriptomics and bulk pollutant concentrations. Then, through structural sensitivity analysis we will select key metabolic reactions to build hybrid models suitable for process design. Finally, our model will be validated for systems treating reals streams.

#052

Personal Information

Full name: Michel Edwar Mickael

Position: Associate Professor

Department: Experimental genomics

Institution: Institute of Genetics and Animal Biotechnology

Institution country: Poland

Abstract

Adera2.0 a novel AI-based pipeline for drug repurposing.

Drug repurposing is a crucial method for discovering new uses for known medications in neuroimmunological (NI) investigations. This approach allows researchers to bypass the lengthy and costly process of developing new drugs from scratch. By repurposing existing medications, scientists can expedite the discovery of treatments for neuroimmunological conditions, potentially leading to faster advancements in patient care. Additionally, drug repurposing may uncover unexpected benefits and applications for medications that were originally developed for other purposes. Several techniques exist that can be used for drug repurposing. However, current methods suffer from a high false positive rate. To address this issue, a deep neural network named Adera2.0 was developed to perform drug repurposing. The workflow uses three deep learning networks: an encoder for embedding text into matrices, a convolution network for predicting answers in the form of embedded matrices, and a new network for extracting compound names from relevant sentences. This deep neural network, which consists of an RNN neural network and a leaky ReLU, achieved 0.0001 loss and 67% sensitivity. Adera2.0's ability to predict NI drug usage was validated against gold-standard datasets, demonstrating its ability to repurpose drug candidates and shorten the development of the drug cycle. Our development sheds light on the possibilities of AI applications in the process of text mining for drug repurposing.

#053

Personal Information

Full name: Morten Bahrt Haulrig

Position: PhD student

Department: Department of Allergy and Dermatology, Herlev and Gentofte Hospital

Institution: Department of Allergy and Dermatology, Herlev and Gentofte Hospital

Institution country: Denmark

Abstract

Exploring vitiligo: The proteomic characteristics of lesional and non-lesional skin.

Background:
Vitiligo is an immune-mediated skin disease that can lead to a decline in quality of life and self-esteem. The current treatments are inefficient, only provide short-term symptomatic relief and their specific treatment effects are largely unknown. Understanding key vitiligo disease mechanisms and the specific proteomic changes in the skin following treatment is essential to develop better treatments, since the current knowledge on the pathophysiology in vitiligo is only partly understood. In this context, mass-spectrometry (MS) based proteomics provide opportunities for optimal characterisation of skin proteins, disease predicting biomarkers, and potential future treatment targets. The potential of MS-based proteomics has not yet been examined in vitiligo. Therefore, this study examined the proteome of vitiligo by using MS-based proteomics to identify and quantify proteomic differences between lesional and non-lesional vitiligo skin.

Objective:
To identify and quantify the proteome of vitiligo by comparing lesional skin, non-lesional skin and skin from healthy controls.

Methods:
63 participants with non-segmental vitiligo were included in this study. Two 2mm skin punch biopsies were collected from lesional and non-lesional skin in each participant. A 2mm skin punch biopsy was collected from 12 age-, sex- and anatomy-matched healthy controls. Each sample was prepared for MS-based proteomics at our

laboratory at Gentofte Hospital. The MS analysis is currently being conducted at PRI, University of Copenhagen.

Results:
We have not yet received the raw data to analyse. Hence, this study does not have any results to present yet.

#054

Personal Information

Full name: Naba Al-Sari
Position: Postdoc
Department: Computational Biomarker Discovery
Institution: Novo Nordisk
Institution country: Denmark

Abstract

Identification of non-invasive biomarkers for predictive and prognostic use for end-stage renal disease

Background:
Individuals with long-standing diabetes can experience damage to small microvascular blood vessels leading to Chronic Kidney Disease (CKD) and increased mortality. CKD impacts 40% of people living with diabetes. Small molecule profiling in blood, coupled with computational biology and Machine Learning (ML) can identify prognostic biomarkers for early diagnosis.

Methods:
Using human-relevant in-vitro models of kidney function and disease mechanism, we elucidate Mode of Action (MoA) to relevant biomarkers in patients. This ensures translatability and back-translation. Prognostic ML models will be computed with clinical risk factors and blood-derived molecular omics data to estimate an individual risk scoring for progression to CKD as we have done previously.

Findings:
We aim to develop highly accurate new ML models for better identification of patients at high risk of progression to end-stage renal disease. We are also expecting the models to highlight the importance of exciting clinical biomarkers for CKD such as albuminuria and eGFR.

Interpretation:
Individuals with diabetes and CKD are often diagnosed late.Our new ML tools and prognostic biomarkers would help reach patients earlier, initiate drug development, and drive quicker trials. By replacing creatinine as surrogate trial endpoints, a better understanding of what drives CKD progression, and clinical response to current treatments could facilitate the identification of new MoAs. This will facilitate the implementation of precision diagnosis in the clinic in the future.

#055

Personal Information

Full name: Nadezhda Doncheva

Position: Assistant Professor

Department: Novo Nordisk Foundation Center for Protein Research

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Cytoscape stringApp 2.0: Analysis and Visualization of Heterogeneous Biological Networks

Complex biological systems are often represented as biological networks containing several types of molecular entities. Analyzing and visualizing such networks can advance our knowledge of the underlying cellular mechanisms. This can be accomplished with online databases and software tools like STRING, Cytoscape, and its many apps. Specifically, Cytoscape stringApp has focused on providing intra-species protein–protein interactions from STRING to facilitate the interpretation of data from omics studies, such as proteomics and transcriptomics.

Here, we highlight the latest stringApp and STRING functionality that opens new avenues for exploring results from high-throughput experiments. Version 2.0 of stringApp greatly improves the support for heterogeneous networks, thus making it possible to create networks that contain both proteins and interactions from STRING as well as other molecular entities or associations from external sources. We exemplify this on a published SARS-CoV-2 interactome in Cytoscape using stringApp. Thereby, we also showcase the new group-wise enrichment analysis, which can be performed automatically on several subnetworks, as well as the retrieval of both functional associations and physical interactions from STRING. Finally, the latest stringApp version offers an improved user interface as well as support for cross-species queries to retrieve protein–protein interactions between eukaryotic parasites and their hosts.

#056

Personal Information

Full name: Narod Kebabci

Position: PhD Student

Department: UCD Conway Institute

Institution: University College Dublin

Institution country: Ireland

Abstract

Predicting context-specific synthetic lethality between paralogs in cancer

Exploiting synthetic lethal interactions is a promising approach for the development of new targeted cancer therapies. Paralogs, genes that arise from gene duplication, are enriched in synthetic lethal interactions, making them a valuable source of therapeutic targets. However, to date, only a minority of paralog pairs have been experimentally tested for synthetic lethality, and many pairs have been identified as synthetic lethal only in specific contexts. Consequently, there is a need for computational tools to prioritize those paralog pairs most likely to be synthetic lethal and to identify the contexts in which these synthetic lethal interactions will operate. To address these challenges, we collated large-scale CRISPR screens of paralog pairs and annotated them with context-specific information (including expression, mutation, and gene essentiality). We found that single gene essentiality and gene expression are among the features that are most predictive of context-specific synthetic lethality between paralog pairs. Combining multiple features, we developed a random forest classifier that distinguishes SL pairs from non-SL pairs with a prediction performance, ROC AUC of ~0.92. Our classifier outperforms individual features and existing classifiers that ignore context and may help to identify those synthetic lethal pairs that are most likely to operate in specific cancer types of interest. Overall, our study highlights the potential of paralog-based synthetic lethality as a broadly applicable approach for targeting gene loss in cancer and sheds light on features that make paralog pairs likely to be synthetic lethal in specific contexts.

#057

Personal Information

Full name: Nils Hofmann
Position: Master student
Department: DTU Bioengineering
Institution: Technical University of Denmark
Institution country: Denmark

Abstract

ExpoSeq: simplified analysis of high-throughput sequencing data from antibody discovery campaigns

High-throughput sequencing (HTS) offers a modern, fast, and explorative solution to unveil the full potential of display techniques, like antibody phage display, in molecular biology. However, a significant challenge lies in the processing and analysis of such data. Furthermore, there is a notable absence of open-access user-friendly software tools that can be utilized by scientists lacking programming expertise. Here, we present ExpoSeq as an easy-to-use tool to explore, process, and visualize HTS data from antibody discovery campaigns like an expert while only requiring a beginner's knowledge.

#058

Personal Information

Full name: Pascal N. Timshel
Position: Bioinformatics Manager
Department: -
Institution: ZS
Institution country: Denmark

Abstract

Uncover protein functions and disease hypotheses with ZS Revelen

ZS Revelen
ZS Revelen enables the discovery of novel, disease-specific molecular relationships by integrating omics data with network biology. Users can explore and analyze multiomics datasets, gaining deeper insights into target-disease connections in a biological context.

Read more at www.zs-revelen.com

About ZS - who we are and what we do
Drug discovery: We accelerate drug discovery by uncovering novel aspects of disease biology to identify new targets. We help identify biomarkers and design and execute experiments, while applying in silico methods to deliver therapies to market faster.

Data and knowledge management: We help clients advance research by using data and AI platforms for in silico science. We leverage ML, customized bioinformatics analyses and multiomics platforms to make research data accessible and actionable.

Read more at
www.zs.com/solutions/life-sciences-randd-and-medical/zs-discovery

#059

Personal Information

Full name: Pernille Kjersgaard Bech

Position: Research Scientist

Department: Production Strain Technology

Institution: Novonesis

Institution country: Denmark

Abstract

Feeding the world on CO2 → acetate → food

According to a recent report led by the UN, more than 250 million individuals faced severe hunger in 2022. To address this growing issue, it is vital that we establish a robust bioeconomy to combat significant global challenges such as malnutrition, biodiversity loss, and climate change. With this goal in mind, the Bill & Melinda Gates and the Novo Nordisk Foundation are funding a consortium where Power to X and biotechnology are joining forces to use CO2 to produce human food protein. Here, CO2 will be converted to acetate by electrolysis and acetate will serve as a fermentation substrate for protein production for food. Large-scale microbial fermentations can produce massive quantities of protein, decoupled from land-use and serve as an alternative to animal protein.

For decades, industrial production strains, such as filamentous fungi, used for fermentation have been fine-tuned for growth on glucose. However, transitioning to acetate as a carbon source demands a complete rewiring of their central metabolism. To accelerate the development of these production strains, we employ adaptive evolution and genome-scale metabolic modeling. Population genomics, transcriptomics, proteomics, and metabolomics are employed to understand the key changes driving phenotypic traits linked to acetate growth and increased production titer. Finally, this multivariate data will be combined with ML and refine mechanistic models, enabling the identification of genetic targets for metabolic engineering and the development of stable production strains.

Ultimately, the technology developed in this project has broad perspectives for production of other biomolecules, chemicals and biofuels.

#060

Personal Information

Full name: Pia Egerup

Position: MD, PhD Student

Department: Department of Obstetrics and Gynecology

Institution: Copenhagen University Hospital Hvidovre

Institution country: Danmark

Abstract

The Danish Health Registers provide a unique opportunity to examine the association between pregnancy loss and later disease

Numerous validated and comprehensive national registers exist in Denmark (Figure 1). All Danish citizens are at birth or emigration provided with a unique central personal identification (CPR) number, which are registered in all contacts with the health care system. All information regarding hospitalization, medication, morbidity and mortality are recorded in the Danish health registers under the CPR-number allowing cross-referencing between registers at an individual level. The Danish National Health Register (DNHR) was established in 1977 and contains information of date, cause and type of all hospitalizations and surgical procedures. All diagnoses are registered according to International Classification of Diseases (ICD). The Danish Medical Birth Register (DMBR) includes all children born in Denmark and their parents since 1973. The Danish National Prescription Register (DNPR) contains information on all redeemed prescriptions since 1995.

Based on these registers we have conducted several studies investigating pregnancy loss and the association with different diseases. First, we identified all pregnancies and pregnancy outcomes for all Danish women by using DNHR and DMBR. Secondly, we identified a case population of women with the specific disease through DNHR and DNPR. Controls consisted of women from the general Danish population without the disease. Logistic regression models provided odds ratios for different diseases with different number of pregnancy losses.

Our studies found a significant association between pregnancy loss and different diseases (e.g. type 2 diabetes, asthma, endometriosis, endocrine disease), which increased with increasing number of losses.

#061

Personal Information

Full name: Rebecca Kirsch

Position: Researcher

Department: Faculty of Health Sciences, University of Copenhagen

Institution: Novo Nordisk Foundation Center for Protein Research

Institution country: Denmark

Abstract

Keeping Database Resources Alive for a Decade

The academic system with high personnel turnover, a constant pressure towards focusing on novel research, and a lack of funding for infrastructure is not conducive to the long-term maintenance of tools and data resources. Yet, since 2013 the JensenLab has been developing and successfully maintaining a number of resources that integrate known and predicted associations between proteins and other biomedical entities. COMPARTMENTS specializes on subcellular localization, TISSUES focuses on protein localization in the body, and DISEASES links proteins with associated medical conditions. We collect, score, and integrate evidence for these different types of associations from databases of curated knowledge and experimental data, computational predictions, and results from text mining of the complete body of publicly available biomedical literature.

For 11 years, the underlying evidence has been updated on a weekly basis. The resources have a combined monthly count of about 1500 visits, and the underlying data is used by many more via both the STRING database and Cytoscape stringApp.

What allowed us to keep each of our resources functional, relevant, and up-to-date for this long is streamlined pipelines, organization into a common backend database, and a flexible frontend Python framework that serves all the different web resources. This way, maintenance is minimal, and individual datasets can be updated easily. At the same time, we can deliver a range of different data types while keeping each resource tailored towards answering a specific user question.

#062

Personal Information

Full name: Sofie Agerbæk

Position: PhD Student

Department: Department of dermatology

Institution: Gentofte Hospital

Institution country: Denmark

Abstract

Characterising the skin mycobiome

More than 25% of the population suffer from an inflammatory skin disease such as psoriasis, atopic dermatitis and fungal infections. Fungal infections are frequent and often affect people who are otherwise healthy. Currently, fungal infections are diagnosed using microscopic examination and cultivation of the fungi, which in some cases takes weeks and specificity relies on the skill of a clinician. Some faster and more specific molecular methods have been explored to identify isolate fungi, but none of these are currently used-in clinic and often require cultivation before sending the samples to a central analysis laboratory. In this project we wish to explore the applicability of using proteomes measured by Liquid-Chromatography tandem Mass Spectrometry (LC-MS/MS) for identification of fungi on the skin. Initially we will analyze the genomes of a cohort of common skin infecting fungi to characterize these and build gene libraries for the proteome analysis. After this we will resolve their proteomes at multiple depths and characterize these thoroughly. Ultimately, we will include a cohort of clinical skin samples from infected and healthy skin to examine the quantifiable fungal proteins. We will explore the possibility of creating a model based on the unique protein expression profiles of the isolate fungi to identify specific fungi present on the clinical skin samples. This has the potential to increase precision and significantly shorten the time from consultation to diagnosis and administration of appropriate treatment to patients.

#063

Personal Information

Full name: Sophia Metz

Position: Postdoc

Department: Novo Nordisk Foundation Center for Basic Metabolic Research

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Mapping genetic effects of cardiometabolic traits on stem cell commitment and early adipogenesis

Adipose tissue (AT) is crucial for systemic energy balance and whole-body metabolism with distinct depot specific regulatory effects. Alterations in the quantity and size of adipocytes influences the microenvironment within adipose tissue, leading to changes in the secretion of adipokines, local hypoxia, adipocyte death, and lipid turnover. Hence, adipose tissue remodeling, including differentiation of new adipocytes from adipose tissue derived stem cells (ASC), are a crucial determinant of systemic metabolic health. Nevertheless, the relationship between depot specific ASC commitment and metabolic health is not yet understood.

Genome-wide association studies (GWAS) have identified numerous genetic loci linked to cardiometabolic health, and AT regulation and thereby harbor information for novel targets of ASC commitment. In this project, I aim to construct disease relevant polygenic risk scores (PRS) from previously published GWAS and run correlation analysis with adipocyte specific markers of early adipogenesis. I will integrate DNA sequencing data, with RNAseq, ATAC, and LipocyteProfiler data at differentiation day 0 and 3 from subcutaneous and visceral AT biopsy samples. This will allow for the first time, to identify disease relevant loci that potentially regulate ASC commitment with physiological consequences. Eventually, I will be able to identify gene targets for functional analyses, and furthermore compile a comprehensive landscape of factors involved in depot specific AT renewal and regulation.

#064

Personal Information

Full name: Stefan E Seemann

Position: Associate Professor

Department: Department of Veterinary and Animal Sciences

Institution: Faculty of Health and Medical Sciences

Institution country: Denmark

Abstract

Analysis of full length 16S and 18S rRNA amplicons and metagenomics for enrichment of microbial communities

Datasets of high-throughput short read sequencing of 16S rRNA gene amlicons, metagenomics and metatranscriptomics are commonplace and common bioinformatic analysis pipelines have been established. Recently, full length sequencing by the Nanopore technology has entered the game. Its main advantages are real time sequencing and the direct alignment to a reference species database. However, there is currently no single recommended analysis pipeline available partly due to the rapid evolution of Nanopore chemistry.

Both methane (CH4) and carbon dioxide (CO2) are greenhouse gases that occur as often undesirable waste products in various industries. We have explored how these gases may be consumed by microbial communities for potential conversion into useful biomass. Bioreactors with natural seawater was enriched with inorganic nutrients and injected with a continuous gas stream containing CH4 and CO2, either in the dark or in the light. The microbial diversity was characterized using chemotaxonomic markers (pigments and isoprenoid quinones), 16S and 18S ribosomal RNA gene amplicons and metagenomes sequenced by Nanopore technology. We discuss different data analysis workflows and taxonomic reference databases to characterize the species specificity of Nanopore sequencing and to identify the enriched microbial communities under different conditions.

#065

Personal Information

Full name: Stephen Tan
Position: Principal Laboratory Automation Engineer
Department: Technology
Institution: Lawrence Berkeley National Laboratory
Institution country: United States

Abstract

"The Road from Data Generation to Integration: Lessons from a Laboratory Automation Engineer"

Facing the challenge of creating a data pipeline with an underspecified scope, I, Stephen Tan, am navigating these uncharted waters as the Principal Laboratory Automation Engineer at a Lawrence Berkeley National Laboratory facility. Our team is dedicated to scaling up bioreactors for a range of projects, where each multi-day experiment, termed a "campaign," relies on a manually completed excel batch record for planning. Despite the continuous data flow from these bioreactors, our current data management process is manual, slow, and prone to errors. While reactor conditions are monitored in real-time, meaningful insights and adjustments are dependent on manual analysis by a process engineer for subsequent campaigns.

Embarking on the journey to construct a data pipeline presents a unique learning curve, especially without prior experience. However, the foundational elements are in place: we have a database ready to receive data, marking our finish line, and our bioreactors are primed to generate this data, defining our starting point. The core challenge lies in bridging these two points seamlessly. Through this endeavor, I aim to unravel the intricacies of data pipeline construction and share valuable lessons learned, hoping to empower others to implement similar solutions in their workplaces. This abstract aims not only to outline the challenges faced but also to kindle interest in the transformative potential of effective data management and the journey from raw data to actionable insights.

#066

Personal Information

Full name: Sumesh Sukumara

Position: Senior Researcher and Group Leader

Department: Novo Nordisk Foundation Centre for Biosustainability

Institution: Technical University of Denmark

Institution country: Denmark

Abstract

Tools for automating sustainability assessment workflows coupled with microbial production data throughout the product development phase.

Humanity has exceeded nature's carrying capacity, which could result in irreversible changes to the environment and global Earth systems, partly due to the over exploitation of non-renewable resources. Therefore, industries worldwide are venturing into possibilities of transforming processes from unsustainable sources to sustainable feedstocks to produce chemicals, food, and other functional materials. This transition towards sustainable development requires microbial products with better sustainability performance. Measuring the sustainability performance of these products demands the deployment of a comprehensive assessment framework to quantify the impacts of the research at various product development phases.

This contribution provides perspectives on successfully integrating environmental sustainability, techno-economic viability, and societal impact assessments to guide product innovation in microbial fermentation-based products. Efforts on quantifying the sustainability KPIs of bio-chemicals are analyzed early in the R&D to highlight impacts, challenges, and opportunities for guiding bio-based products. The research will showcase how production-level simulations can feed broader techno-economic and environmental impact assessment tools such as Life Cycle Assessment (LCA) and planetary environmental and social impact assessment frameworks. Being relevant to the bio-based community, methodologies, and applied cases will be showcased.

#067

Personal Information

Full name: Teddy Groves

Position: Researcher

Department: Quantitative Modelling of Cell Metabolism

Institution: Novo Nordisk Foundation Center for Biosustainability, Technical University of Denmark

Institution country: Denmark

Abstract

Maud: quantitative modelling of metabolic networks with Bayesian kinetic models

We will present Maud, a free and open source command line application that implements Bayesian statistical inference for kinetic models of biochemical metabolic reaction networks.

Using Maud it is possible to infer the mechanistic causes of differences in metabolism between experimental conditions while respecting physical laws and accommodating information from quantitative multi-omics datasets and the kinetic parameter literature. This results in previously inaccessible insights about processes like the Warburg effect and C1 metabolism that are relevant both to human health and to the design of sustainable biological production systems.

We will use a simple case study to illustrate how users can quickly and easily implement their own Bayesian kinetic models, covering how to prepare a suitable dataset, how to install and run Maud and how to interpret its results.

We will particularly emphasise aspects of our work that can be transferred to similar data-based computational biology projects. Specifically, we will explain Maud's approach to data modelling, validation and serialisation, how we achieved a convenient installation workflow and user interface despite the inclusion of low-level, platform-dependent dependencies, and how we addressed challenges involved in software development in an academic setting.

Finally, we will discuss how Maud encourages reproducibility and FAIR principles, and how we have approached the difficult task of visualising and communicating the output of Bayesian kinetic models.

#068

Personal Information

Full name: Thomas Vain
Position: Digital Lab Manager
Department: Technology
Institution: Technical University of Denmark
Institution country: Denmark

Abstract

Bridging Plant Science and Data Analytics for Planetary Health: Thomas Vain's Innovative Research Journey

Thomas Vain, an adept plant science researcher, seamlessly integrates data science into his work, driving advancements in agricultural sustainability and human health. With expertise in plant physiology and agro-chemical interactions, Thomas Vain works pioneers image processing techniques for high-throughput plant phenotyping. today, as a digital lab manager, he utilizes, electronics, computer sciences and big data analytics to optimize agricultural practices, emphasizing ecosystem resilience and public health. Thomas Vain's interdisciplinary approach highlights the transformative potential of data-driven solutions in addressing complex challenges at the intersection of agriculture, ecology, and human well-being.

#069

Personal Information

Full name: Timothy Jenkins
Position: Asst. Prof. & Head of Data Science
Department: DTU Bioengineering
Institution: DTU
Institution country: Denmark

Abstract

Decode, Design, Deliver: Machine Learning Revolutionises Target Identification and Therapeutic Discovery

Machine learning (ML) is ushering in a new era in target identification and therapeutic discovery, with profound implications across various scientific domains. In the realm of proteomics, we have leveraged the power of ML to develop a leading-edge deep learning model called InstaNovo. This model enables high-precisionde novo peptide sequencing, eliminating many of the constraints of conventional methods and opening up new opportunities in antibody sequencing, identification of neo-epitopes in cancer, and the exploration of the dark proteome. However, beyond target identification, ML is also proving promising in the rapid discovery and development of therapeutics. Particularly with the rise of generative de novo protein design, design of functional binders entirely in silico has been brought within reach. We have taken advantage of these developments, to design minibinders (small binding proteins primarily comprised of beta sheets and alpha helices) that can neutralise snake venom toxins. Though further validation is needed, these findings hold promise for the rapid development of next-generation therapeutics.

#070

Personal Information

Full name: Txell Amigó

Position: Master's Student

Department: Multi-omics Network Analytics Research Group

Institution: Novo Nordisk Foundation Center for Biosustainability Technical University of Denmark

Institution country: Denmark

Abstract

FermentDB: A Database for High-cell density Fermentations

Industrial fermentation leverages microorganisms for the synthesis of valuable products. These processes serve as a pivotal technological asset for reducing our dependence on chemicals and products derived from fossil fuels. Employing high-density fermentation strategies proves to be a cost-effective approach for achieving optimal yields in biomass, extracellular metabolites, intracellular components, or modified substrates. The rate in which fermented products are generated is contingent upon various factors such as concentration of microorganisms, cell density, cellular components, enzymes, temperature, and pH. These systems are difficult to design, operate, and scale up and down, which has a direct impact in crucial parameters such as product titer, rate, and yield. The absence of high-quality fermentation data hinders optimization and prediction strategies that could otherwise mitigate these challenges. Our goal is to establish a standardized and aggregated database, consolidating high-density fermentation data for accessibility within the scientific community.
The database will harmonize a previously generated comprehensive dataset from more than 600 fermentations using different strains under varied conditions to produce melatonin and tryptophan. Furthermore, this resource will allow the analysis and visualization of these experiments as well as streamlining the integration of new fermentation datasets. By compiling comprehensive data, we aim to expedite the development of "fermenterphiles" and enhance bioproduct production, marking a significant stride toward sustainable industrial practices."

#071

Personal Information

Full name: Yesid Cuesta Astroz
Position: Assistant Professor
Department: School of microbiology
Institution: University of Antioquia
Institution country: Colombia

Abstract

OrthoHPI 2.0: A renewed and extended interactive resource of predicted human-parasite protein-protein interaction networks

Infectious diseases are one of the leading causes of death worldwide, particularly in low-income countries, especially in young children. Among the pathogenic infections, the parasitic ones have a great socioeconomic impact as they are responsible for infections in humans, animals, and plants. According to the WHO, in 2015 more than two billion people were infected by parasites, the majority of them in developing countries. Unfortunately, combatting the parasitic diseases with the available drugs has been a challenge due to the acquired drug resistance by the parasites. The study of molecular human-parasite interactions is essential to understanding parasitic infection and adaptation and contributes to the development of new treatments. To this end, we previously developed OrthoHPI (PMID:30815000 www.ncbi.nlm.nih.gov/pubmed/30815000; https://orthohpi.jensenlab.org), a resource that provides homology-derived predictions of host-parasite protein-protein interactions (PPI). This work shows a renewed and extended OrthoHPI by integrating new versions of databases, predictors, and proteomes of 24 eukaryotic parasites, allowing the comparison of the resulting PPI networks, extracting biological insights from data, and generating new hypotheses. Apart from the previous features, in this new version we included single-cell RNA-seq datasets from the human protein atlas and structural predictions using AlphaFold to add a single-cell and molecular context of the interactions. The predicted networks, single-cell data and the structures can be visualized and downloaded at https://orthohpi.streamlit.app

#072

Personal Information

Full name: Yun Huang

Position: Postdoc

Department: Novo Nordisk Foundation Center for Basic Metabolic Research

Institution: University of Copenhagen

Institution country: Denmark

Abstract

Genetic determinants of the plasma lipidome and relations to cardiometabolic risk in children and adolescents

Several lipid species have been associated with type 2 diabetes, cardiovascular disease, and steatotic liver disease. While the genetic architecture of the lipidome has been investigated in adults, the genetic determinants of lipid species and their association with cardiometabolic risk in children and adolescents remain understudied.

We measured 227plasmalipid species in 1,149 children and adolescents(44.8% boys)with a median age of 11.2 years from an obesity clinic and the general population.We identified 37 independent genome-wide significant loci ($P < 5 \times 10–8$) for 52 lipid species, including loci associated with sphingolipids (SPTLC3, SYNE2, ATP10D), phosphatidylethanolamine (LIPC), phosphatidylinositol (MBOAT7), and 9 previously unreported loci. One-sample Mendelian randomization (MR) analysis using 5 loci with biological plausibility on individual-level data identified positive causal associations between ceramide and liver enzymes, sphingomyelin and hemoglobin A1c (HbA1c), and phosphatidylethanolamine and high-sensitivity C-reactive protein in children and adolescents. Two-sample MR using summary statistics showed consistent results with one-sample MR, and indicated additional causal links, specifically between ceramide and higher HbA1c levels, and phosphatidylinositol withhigher liver enzymes.

Our results provide insights into the genetic determinants of plasma lipid species in children and adolescents, emphasizing causal links between specific risk lipids and cardiometabolic risk profiles, which may provide further insights into targeted intervention strategies.

Participants

Adrian Geissler	Postdoc	University of Copenhagen
Adriana Tomic	Professor	Boston Univesity
Agnete Troen Lundgaard	Postdoc	Copenhagen University Hospital Hvidovre
Ajuna Azad	Head of Research Data Management	Technical University of Denmark
Alba Refoyo Martinez	Data Scientist	University of Copenhagen
Albert Palleja Caro	Senior data scientist	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Albert Thor Thorhallsson	Postdoc	Technical University of Denmark
Alberto Santos Delgado	Senior Researcher	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Alexander G. B. Grønning	Postdoc	University of Copenhagen
Alexander Pil Henriksen	PhD student	University of Copenhagen
Alisa Pavel	Postdoc	Technical University of Denmark
Angel Phanthanourak	Research Assistant	Technical University of Denmark
Angelica Prado	Data Standards Developer	Novo Nordisk A/S
Anika Gupta	Principal	Google Ventures
Anna Svetlova	PhD Student	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Anne Wenzel	Staff Scientist	University of Copenhagen
Annelaura Bach Nielsen	Postdoc	University of Copenhagen

Full Name	Position	Institution
Beatrice Dyring-Andersen	Associate Professor	Gentofte Hospital
Birte Höcker	Professor	Bayreuth University
Blaine Gabriel Fritz	Assistant Professor	Institute for Health and Medical Science, University of Copenhagen
Bogumil Zimon	Bioinformatician	Cirqle Biomedical
Boyang Ji	Data Scientist	BioInnovation Institute
Bunia Ingrid Adsersen	BSc. Student in Life Science Engineering	Technical University of Denmark
Carl Malina	Research Scientist	Carlsberg Research Laboratory
Carlos G. Acevedo-Rocha	Senior Researcher (Group Leader)	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Carmen Saenz	Senior Bioinformatician	Novonesis
Cecilie Møller-Jensen	BSc. Student in Biomedical Engineering	Technical University of Denmark
Clare Cox	Project Vice President, Obesity Drug Development	Novo Nordisk A/S
Daniel Vik	Senior Data Scientist	Amgen
David Lennox-Hvenekilde	Senior Scientist	Enduro Genetics
David Ochoa	Open Targets Platform Coordinator	EMBL-EBI
David Westergaard	Senior Researcher	Copenhagen University Hospital Hvidovre
Ding He	Research Data Steward	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark

Full Name	Position	Institution
Elenora Nigro	Scientific Illustrator	
Elysia Gao	Master's Student / Research Assistant	Technical University of Denmark
Emma Lundberg	Associate Professor	Stanford University and KTH Royal Institute of Technology
Eric Bautista Farrerons	MSc student	Technical University of Denmark
Ester Milesi	LIMS administrator	Technical University of Denmark
Eva Mechlenburg-Vrou	Project Specialist	Novo Nordisk Foundation
Fatemeh Ghanami	Data Engineer	Technical University of Denmark
Francesc Fernandez	Senior Director R&D Data Science	Almirall
Francesco Russo	Bioinformatics Specialist	H. Lundbeck A/S
Georgia Mitsa	Scientist - Proteomics	Acrivon AB
Gianluca Mazzoni	Senior Research Scientist	Novo Nordisk A/S
Gustav Næs	Student Assistant	Novo Nordisk Foundation
Hans Roubos	Senior Science Fellow	DSM-Firmenich
Helle Krogh Pedersen	Director of Integrativ Systems Biology	Clinical Microbiomics
Henry Webel	Postdoc	University of Copenhagen
Igor Marin de Mas	Senior Researcher and PI	Copenhagen University Hospital, Rigshospitalet
Isabel Díaz-Pinés Cort	PhD student	University of Copenhagen
Ivan Tomic		Boston Univesity

Full Name	Position	Institution
Jakob Berg Jespersen	Data Scientist	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Jakub Vašíček	PhD student	University of Bergen
Jean-Marie Mouillon	Science Manager	Novonesis
Jens Frickel	Computational biologist	Rejuvenate Biomed
Jens Vinge Nygaard	Professor	Department of Biological and Chemical Engineering , Medical Biotechnology, Aarhus University
Jesper Lauridsen	BSc Student	Technical University of Denmark
John Shorter	Assistant Professor	Roskilde University
Jonathan Funk	PhD student	DTU Bioengineering
Jonathan Robinson	Data Engineer	BioInnovation Institute
Jose Alejandro Romero Herrera	Principal Bioinfomatic	H. Lundbeck A/S
Josefine Meineche	Research assistant	University of Copenhagen
Juan Lafuente-Barquero	Assistant Professor	Biotech Research and Innovation Centre (BRIC), University of Copenhagen
Julia Villarroel	Senior Bioinformatician	ZS Discovery
Kai Blin	Senior Researcher	Technical University of Denmark
Karen Sømmergaard	Project Specialist	Novo Nordisk Foundation
Karina Banasik	Senior Researcher	Copenhagen University Hospital,Hvidovre

Full Name	Position	Institution
Kasper Lage	Managing Director	"Novo Nordisk Foundation Centre for Genomic Mechanisms of Disease at the Broad Institute of MIT and Harvard"
Kübra Altinel	Postdoc	Biotech Research and Innovation Centre (BRIC), University of Copenhagen
Lars Juhl Jensen	Professor	University of Copenhagen
Lars Keld Nielsen	Scientific Director	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Laura Machado	MSc Bioinformatics	Technical University of Denmark
Laura Pikkupeura	Postdoc	University of Copenhagen
Lea Harder	Senior Research Scientist	Novo Nordisk A/S
Lea Mette Madsen Sommer	Dept. Manager	Novonesis
Leonie Boender-van Dijk	Scientist (PhD)	DSM-Firmenich
Line Clemmensen	Associate Professor	Department of Applied Mathematics and Computer Science, Technical University of Denmark
Lorenzo Guerci	PhD student	Technical University of Denmark
Luca Gaessler	Master's student	Karolinska Institute
Lukas Huschet	PhD student	Ludwig Maximilian University (LMU)
Luke W. Johnston	Team Leader	Steno Diabetes Center Aarhus
Malte Thodberg	Postdoc	University of Copenhagen
Maria Dalby	Director	Muna Therapeutics

Full Name	Position	Institution
Maria Novosolov	Data Scientist	Clinical Microbiomics
Marietta Kokla	Postdoc	Steno Diabetes Center
Martin Miller	Senior Director	AstraZeneca
Martin S. Borchert	Director, R&D	Novonesis
Matin Nuhamunada	PhD Student	Technical University of Denmark
Matthias Mattanovich	Postdoc	University of Copenhagen
Michaël Pierrelée	Postdoctoral researcher	Technical University of Denmark
Michel Edwar Mickael	Associate Professor	Institute of Genetics and Animal Biotechnology
Morten Bahrt Haulrig	PhD student	Department of Allergy and Dermatology, Herlev and Gentofte Hospital
Naba Al-Sari	Postdoc	Novo Nordisk A/S
Nadezhda Doncheva	Assistant Professor	University of Copenhagen
Naixin Zhang	Postdoc	University of Copenhagen
Narod Kebabci	PhD Student	University College Dublin
Niklas Blomberg	Executive Director	Innovative Health Initiative
Nils Hofmann	Master student	Technical University of Denmark
Oksana Lukjancenko	Senior bioinformatician	Clinical Microbiomics
Pascal N. Timshel	Bioinformatics Manager	ZS Dsicovery
Patrizia Buldo	Scientific Manager, Biotech	Novo Nordisk Foundation

Full Name	Position	Institution
Pernille Kjersgaard Bech	Research Scientist	Novonesis
Pia Egerup	MD, PhD Student	Copenhagen University Hospital Hvidovre
Rebecca Kirsch	Researcher	Novo Nordisk Foundation Centre for Protein Research, University of Copenhagen
Richard Dennis	Research Consultant Data Management	Novo Nordisk Foundation Centre for Stem Cell Medicine, reNEW, University of Copenhagen
Sergio Oller Moreno	Senior Data Engineering	Almirall
Sergio Picart-Armada	Senior AI/ML Scientist	Almirall
Shanshan He	Research Assistant	University of Copenhagen
Shilpa Garg	Senior Researcher and Group Leader	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Simon Rasmussen	Associate Professor and Group Leader	Novo Nordisk Foundation Centre for Protein Research, University of Copenhagen
Sofia Norås	Senior Project Manager	Novo Nordisk Foundation
Sofie Agerbæk	PhD Student	Gentofte Hospital
Sophia Metz	Postdoc	University of Copenhagen
Stefan E Seemann	Associate Professor	Faculty of Health and Medical Sciences, University of Copenhagen
Stephan Malzacher	Scientific Data Engineer	Novonesis
Stephen Tan	Principal Laboratory Automation Engineer	Lawrence Berkeley National Laboratory

Full Name	Position	Institution
Sumesh Sukumara	Senior Researcher and Group Leader	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Sünje Johanna Pamp	Excutive Assistant to VP in R&D	Novonesis
Søren Karst	Senior Scientist	Novonesis
Teddy Groves	Researcher	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Thomas Vain	Digital Lab Manager	Technical University of Denmark
Timothy Jenkins	Asst. Prof. & Head of Data Science	Technical University of Denmark
Torben Vedel Borchert	Scientific Director, Biotech	Novo Nordisk Foundation
Tune H Pers	Associate Professor and Group Leader	Novo Nordisk Foundation, Center for Basic Metabolic Research, University of Copenhagen
Txell Amigó	Master's Student	Novo Nordisk Foundation Centre for Biosustainability, Technical University of Denmark
Ulla Gro Jensen	Senior Scientific Lead, Ph.D., Nat-Tech	Novo Nordisk Foundation
Ulrik Nicolai de Lichtenberg	Senior Scientific Manager, PhD, Nat-Tech	Novo Nordisk Foundation
Yesid Cuesta Astroz	Assistant Professor	University of Antioquia
Yun Huang	Postdoc	University of Copenhagen