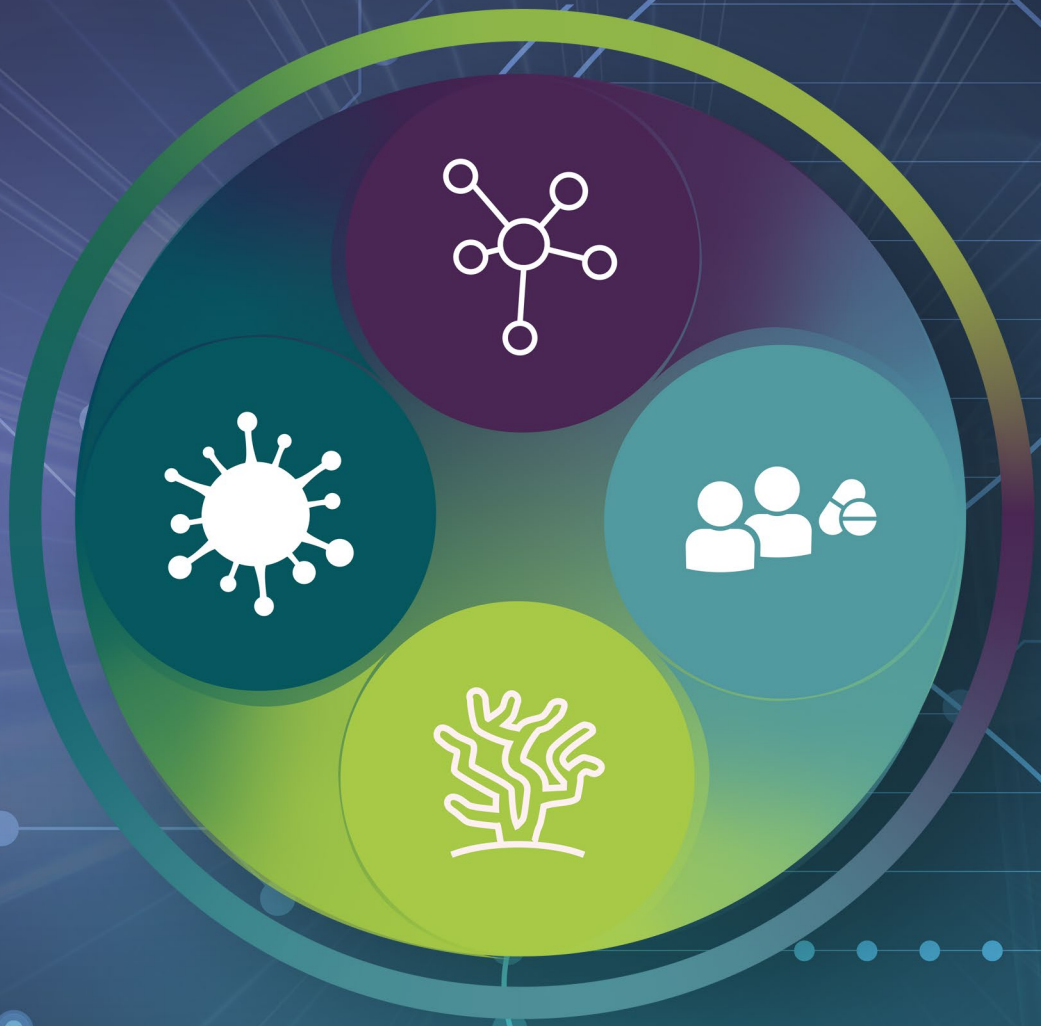
Annual DDLS conference 2026

Poster abstracts



SciLifeLab

Knut and Alice Wallenberg Foundation

Contents

1. envoi: Automated feature extraction of geospatial data for ecological analysis and modeling 5

2. Advancing FAIR genomics with open-source solutions: The Swedish Reference Genome Portal and DivBase 5

3. Sweden AI Factory 6

4. Metabolic differences in the maternal gut microbiome precede the birth of large-for-gestational-age infants: A nested case-control study 7

5. Mapping drug-induced spatial reorganization in cancer-fibroblast co-cultures using cell-cell interaction graphs 7

6. Differences in the faecal microbiome of obese and non-obese pregnant women: a matched cohort study in Sweden 8

7. If I If I have a urinary tract infection caused by Escherichia coli should I worry about my dog? 9

8. Gut microbiota compositional trajectories explain drug response in Type 2 Diabetes clusters 10

9. KaroSpace: a rapid-access framework for interactive exploration of multi-sample spatial omics data 10

10. Organismal development as a source of evolutionary predictability 11

11. Proteomics Platform 11

12. Environmental and Lifestyle Exposures Linked to the Temporal Variability of Gut Microbial Species in Healthy Adults 11

13. Adaptive segmentation of 3D cryo-EM reconstructions using feature-based clustering 12

14. Artificial intelligence-driven diagnostic support for antibiotic-resistant infections 12

15. Conserved mycobacteriophage protein uses structural mimicry to displace host initiation factors 13

16. Temporal Dynamics of the Gut Microbiota and Host Metabolic Regulation in Type 2 Diabetes 14

17. Beyond the Bump: Oral health as an overlooked dimension of perinatal care. 14

18. SHE: Structural History of Eukarya 15

19. A Paradoxical Success Story of Polyploids 16

20. FEGA Sweden – Repository for sharing Swedish biomedical research data 17

21. Beyond FID: How should we evaluate Generative Models for Fluorescence Microscopy? 17

22. Resolving membrane packing and lipid phase behaviour using Excitation-Emission Matrix spectroscopy 18

23. NBIS: a national bioinformatics infrastructure at a university near you! 18

24. Impact of input dimensionality on data-driven prediction of carotid plaque transcriptomics from non-invasive CT Angiography 18

25. Exploring Fragile-Site Genomic Variation as Endogenous Lineage Tracing Markers in Neuroblastoma 20

26. Single-cell, clonal and spatial atlases of cranial placodes illuminate their specification and evolution 21

27. Fit for which purpose? Evaluating differentially private synthetic clinical cohorts on public data: utility per use, attack resistance, and a certificate that prints its contraindications 21

28. Evolutionary Dynamics of Adherence-Mediating Fimbriae in Pathogenic Bacteria 22

29. SciLifeLab OMERO: A Collaborative HPC-Enabled Platform for Data-Driven Bioimaging Research 22

30. Charting cell differentiation in single-cell omics data via transcription-dynamics-informed optimal transport 23

31. Autonomous closed-loop cell biology with a general-purpose coding agent 23

32. LADDER: Context-aware gene set annotation with LLM-based hypothesis generation and literature-grounded validation 24

33. Targeting mitochondrial reprogramming as a new vulnerability in lethal prostate cancer 24

34. Bridging Genes, Drugs, and Behavior: Multiscale Effects of Psilocybin in C57BL/6 Mice 25

35. FIREFLY: Framework for Interactive Rendering, Exploration and Feature Extraction of Light-Sheet Microscopy 26

36. Cognitive relevance of gradient-aligned connectivity changes in aging and Alzheimer's disease 26

37. Rapid and flexible gene cluster family delineation with IGUA 28

38. Scalable And Reference-Free Single-Cell Metagenome Analysis with Bascet and Zorn *28*

39. Cyclodextrins as a tool to study lipid number asymmetry in model membranes 30

40. MembraneFinder: Annotation-free detection and quantitative analysis of lipid membranes in cryo-EM images 30

41. Spatial Omics 31

42.. Plasmid-mediated bacteriocin dynamics during early-life gut microbiome assembly 31

43. Pan-Neurological Disease Profiling of the Human Circulating Proteome 32

44. DeepRepurpose: A Spatial Biology-Informed and Deep Learning-Driven Drug Repurposing Framework 33

45. Evolutive history of the Striped-backed antbird (Thamnophilidae: Myrmorchilus strigilatus) 33

46. Mechanisms of membrane deformation by the Caveolin-1 complex 34

47. Accelerate your infectious disease research 35

48. Time-Resolved, Multi-Channel Live-Cell Imaging and AI-Driven Phenotypic Profiling of Drug Responses in Ovarian Cancer 35

49. Clinical Genomics 36

50. Shiga toxin acquisition in the extraintestinal escherichia coli ST117 lineage through putative prophage transfer 36

51. Global Population Structure of Enterotoxigenic Escherichia coli from a One Health Perspective 37

52. To eat, or not to eat? Host repertoire of a tropical butterfly 38

53. Hypothesis-driven spatial metabolomics: an open-source toolkit for multimodal interrogation of bioactive lipid pathways 38

54. Tracking Biodiversity Change Through Time and Space: Sweden's 60-Year Airborne eDNA Archive 39

55. A Multi-Omics Signature of the Pre-Metastatic Liver in Pancreatic Cancer 40

56. strobealign long-read alignment enable fast and accurate Structural Variant Detection using Long Reads 41

57. From samples to results: BioEngine for browser-based bioimage AI 41

58. Transcriptionally defined AML cell states associate with treatment response and microenvironmental remodeling 42

59. Beyond Cell Counts and Section Planes: Graph-based Modelling and 3D Reconstruction of the Tumour Microenvironment 42

60. Improved interpretation of pathogenic variants by pediatric transcript expression from developmental GTEx 43

61. Decoding treatment resistance in Acute Myeloid Leukaemia(AML) 44

62. SciLifeLab Cellular & Molecular Imaging Platform 44

63. Promoter-proximal pausing defines distinct modes of stress-response transcription across multicellular organisms 45

64. Spatial pattern based peak selection for high-dimensional mass spectrometry imaging 46

65. (G)Host plants of tropical caterpillars 46

66. Safe Colab enables AI agent–guided zero-trust analysis and federated learning on sensitive biomedical data 47

67. Conformal prediction as a solution to distribution shifts in medical image analysis 47

68. The Instrument That Refuses: Physical Constraint Checks for AI-Agent Laboratory Automation 48

69. Genomics platform 48

70. Somatic mitochondrial variants in blood are associated with increased incidence of hematologic malignancies 48

71. The United Structures of Pathology 49

72. Genomic surveillance of the 2025-2026 anthrax outbreak in Kruger National Park, South Africa 49

73. Fake It Till You Make It: Realistic Emulation of Cryo-EM Micrographs Using GANs 50

74. Berzelius: Sweden's National AI Supercomputer for Data-Driven Science 50

75. Identification of new drivers of Clonal Hematopoiesis using Whole-Genome Sequencing 51

76. Leveraging mass spectrometry data for predicting chemical persistence and mobility in water 51

77. The distribution and role of species-level low prevalence taxa in the human gut microbiota 52

78. Metabolomics platform 53

1. envoi: Automated feature extraction of geospatial data for ecological analysis and modeling

feature extraction, remote sensing, environmental data, ecoinformatics

Adrian Baggström, Jakob Nyström, Miguel Redondo, Tobias Andermann

Ecological and biogeographic analyses depend on linking species occurrence data to environmental context, yet compiling variables such as elevation, temperature, and land cover for each occurrence record remains a slow and error-prone task. Researchers routinely stitch together datasets from multiple sources, handle reprojections, and reconcile mismatched formats and resolutions before any modelling can begin. We present envoi, an open-source tool that replaces this fragmented workflow with a single unified interface, available as both a Python package and a web application. Users provide occurrence points in Darwin Core format together with a short configuration, and envoi returns either spatial summary statistics or analysis-ready raster tiles. The package supports data extraction from Google Earth Engine and locally stored datasets through the same interface, drawing on a built-in catalogue of nearly 100 environmental variables from 26 globally available sources. By automating this process, envoi lowers the technical barrier to geospatial data access, letting researchers of any expertise level focus on ecological analysis rather than data engineering. envoi is developed at the Biodiversity Data Lab, Uppsala University. The package is openly available on PyPI as envoi-geospatial, and with source code and documentation on GitHub.

---

2. Advancing FAIR genomics with open-source solutions: The Swedish Reference Genome Portal and DivBase

genomics, biodiversity, open science, data management, FAIR data

Angela P. Fuentes-Pardo, Department of Immunology, Genetics and Pathology, Science for Life Laboratory Data Centre, Uppsala University, Uppsala, Sweden. Daniel P. Brink, Department of Immunology, Genetics and Pathology, Science for Life Laboratory Data Centre, Uppsala University, Uppsala, Sweden. Rory Crean, Department of Immunology, Genetics and Pathology, Science for Life Laboratory Data Centre, Uppsala University, Uppsala, Sweden. Quentin Ågren, Kungliga Tekniska högskolan, Science for Life Laboratory, Data Centre, quentina@kth.se, Stockholm, Sweden. Andreas Wallberg, Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala, Sweden. Hanna Kultima, Department of Immunology, Genetics and Pathology, Science for Life Laboratory Data Centre, Uppsala University, Uppsala, Sweden. Johan Rung, Department of Immunology, Genetics and Pathology, Science for Life Laboratory Data Centre, Uppsala University, Uppsala, Sweden. Henrik Lantz, National Bioinformatics Infrastructure Sweden, Science for Life Laboratory, Uppsala University, Uppsala, Sweden.

The rapid growth of genomics data poses challenges for research data management, including fragmentation, limited accessibility, and insufficient metadata integration. These issues hinder reproducibility, reuse, and compliance with Open Sciences and FAIR (Findable, Accessible, Interoperable, Reusable) principles, particularly in distributed and collaborative research environments involving multiple teams and stakeholders.

Here, we introduce two complementary technical solutions developed by Science for Life Laboratory (SciLifeLab), Sweden's national research infrastructure for life sciences, to address these challenges. First, the Swedish Reference Genome Portal provides an open-

access, user-friendly platform for visualizing and disseminating high-quality genomes and annotations generated by researchers affiliated with Swedish institutions. By lowering barriers to genomic data access and exploration, the portal enhances discoverability and long-term usability of genomic resources, thereby supporting open science and collaboration.

Second, DivBase (Genomic diversity database) addresses the management of large and complex population genomics datasets and their associated metadata. In collaborative environments, such datasets are often dispersed and poorly linked across teams, increasing risks of data loss and limiting reuse. DivBase serves as a local repository where VCF files and sample metadata are securely stored, organized, and shared among collaborators. It provides a centralized system with versioning, metadata integration, and query capabilities, facilitating collaborative analysis and downstream submission to long-term archives.

Together, these solutions exemplify how national research infrastructures can operationalize FAIR principles and enable more efficient, transparent, and reproducible genomics research.

---

3. Sweden AI Factory
AI, Artificial Intelligence, Machine Learning, Life Sciences

Brito, António, Sweden AI Factory NAISS Linköping University

The Sweden AI Factory is part of the EU's ambitious AI Factories initiative which currently includes 19 different sites across Europe. All these Factories offer access to AI-optimised supercomputers and services such as training and support to SMEs, startups, researchers, and government agencies. The aim is to advance innovation, boost the adoption of AI and make Europe a leading AI continent.

The Sweden AI Factory offers hands-on guidance from experienced AI experts. Our experts support you throughout your AI journey from defining the right use case to deploying and scaling solutions. We combine technical expertise, access to advanced infrastructure, and practical guidance to help you succeed. The service hub of the Sweden AI Factory brings deep domain expertise across five focus areas: Materials Science, Autonomous Systems, Gaming, Trustworthy AI, and Life Sciences.

The Life Sciences team of the Sweden AI Factory currently contains 9 experts spread out across Sweden. We bring expertise in various fields including but not limited to medical imaging, drug discovery, precision medicine and molecular life sciences. We are ready to offer guidance and support for AI development, from research to real-world implementation.

The Sweden AI Factory is hosted by NAISS (the National Academic Infrastructure for Supercomputing in Sweden) in partnership with RISE Research Institutes of Sweden. Funding is shared between Sweden (through the Swedish Research Council and Vinnova, Sweden's Innovation Agency) and the EuroHPC Joint Undertaking, which is the EU's supercomputer organisation. The Sweden AI Factory is a fully funded project and in practice this means that our services are free. We only ask for commitment and acknowledgement in return.

4. Metabolic differences in the maternal gut microbiome precede the birth of large-for-gestational-age infants: A nested case-control study
Large-for-gestational-age, Fetal overgrowth, Maternal gut microbiome, Pregnancy, Anaerostipes hadrus

Anusha T. Antony, Fatemeh Mohseni, Gaël Toubon, Unnur Guðnadóttir, Lars Engstrand, Emma Fransson, Ina Schuppe-Koistinen, Nele Brusselaers, Luisa W. Hugerth

Neonatal size impacts newborn health and is influenced by various prenatal factors. This study explores whether the maternal gut microbiome, which affects maternal metabolism, also contributes to excessive fetal growth. We conducted a matched case-control study (1:2) within the SweMaMi cohort (2017–2021), comparing gut microbiome profiles of mothers delivering large-for-gestational-age (LGA; >90th percentile) versus appropriate-for-gestational-age (AGA; 10th–90th percentile) infants at early (TP1:10–20 weeks; LGA n = 245, AGA n = 490) and late pregnancy (TP2:28–32 weeks; LGA n = 157, AGA n = 252) using shotgun metagenomics. Longitudinal within-group changes in gut microbiome diversity from TP1 to TP2 were examined to compare temporal patterns between groups. All analyses were adjusted for maternal age, pre-pregnancy body mass index, parity, and diet category. At TP1, species richness and Shannon diversity were higher in LGA mothers ($p < 0.05$), and community composition differed between the groups ($p < 0.05$). However, these differences were not maintained at TP2. Shannon diversity and evenness decreased more in LGA mothers from TP1 to TP2 ($p < 0.05$). LGA mothers showed a higher prevalence of Oscillospiraceae species, while AGA mothers had higher abundance of Anaerostipes hadrus and an unclassified Lachnospiraceae species. Functional profiling indicated enrichment of proteolytic modules in LGA and short-chain fatty acid production in AGA at TP1, with Anaerostipes hadrus contributing to AGA modules. Early pregnancy gut microbiome diversity and function may influence fetal overgrowth risk, although microbiome profiles converged later. LGA-associated microbiomes appeared more geared towards energy acquisition (proteolysis) and less towards short-chain fatty acid fermentation.

---

5. Mapping drug-induced spatial reorganization in cancer-fibroblast co-cultures using cell-cell interaction graphs
Co culture cellular network construction and MoA Analysis

Ardan Yılmaz1, Osheen Sharma2*, Golnaz Taheri1†, Brinton Seashore-Ludlow2† 1School of Electrical Engineering and Computer Science, SciLifeLab, KTH Royal Institute of Technology, Stockholm, Sweden 2Department of Oncology-Pathology, Karolinska Institutet, Science for Life Laboratory, Stockholm, Sweden *Co-first authors †Co-senior authors*

High-content imaging has become a powerful approach for profiling drug-induced phenotypic changes, yet conventional single-cell morphological profiling often treats cells as isolated units. In multicellular co-culture systems, this can overlook an important dimension of drug response like the reorganization of spatial relationships between interacting cell populations. Here, we construct Voronoi/Delaunay-based cell-cell interaction graphs from nuclei coordinates extracted from CellProfiler outputs to capture multicellular organization in cancer and fibroblast co-cultures across multiple drug classes tested at five different concentrations. We use these graph representations to evaluate whether local interaction patterns can recover broad phenotypic responses to drug perturbation. Treated wells are assigned to nine phenotypic categories describing cancer and fibroblast drug responses, including cell killing, morphological change, and unaffected conditions, based on CellProfiler-derived morphology and viability measurements. For each category,

we generate representative interaction networks and extract network-level features describing viability, cell-type interaction composition, spatial organization, and community structure. We then compare network features from treated wells with control baselines to quantify how each perturbation alters multicellular organization. In parallel, we examine the distribution of feature changes across control and treated wells to determine whether phenotypic categories show reproducible spatial signatures. Preliminary results indicate that Voronoi-based interaction networks capture consistent patterns associated with broad drug-response classes. These findings suggest that cell–cell interaction graphs could help provide better understanding of how drug perturbations reshape tumor-microenvironment organization, revealing multicellular response patterns and interaction-driven effects that may improve the interpretation of drug mechanisms and support future cancer drug discovery efforts.

---

6. Differences in the faecal microbiome of obese and non-obese pregnant women: a matched cohort study in Sweden

Maternal obesity, gut microbiome

Evangelos Patavoukas 1 2, Bangzhuo Tong 3, Unnur Guðnadóttir 4, Kyriakos Charalampous 3, Nele Brusselaers 4 5, Ina Schuppe-Koistinen 6, Lars Engstrand 6, Emma Fransson 6 7, Eva Wiberg-Itzel 2 8, Luisa Hugerth 9 Affiliations Collapse Affiliations 1Department of Gynecology and Reproductive Medicine, Karolinska University Hospital, Stockholm, 14186, Sweden. 2Department of Clinical Science and Education, Södersjukhuset, Karolinska Institutet, Stockholm, 11883, Sweden. 3Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, 75237, Sweden. 4Department of Women's and Children's Health, Karolinska Institutet, 17165, Solna, Sweden. 5Global Health Institute, University of Antwerp, Antwerp, 2610, Belgium. 6Department of Molecular, Tumour and Cell Biology, Karolinska Institutet, Stockholm, Sweden. 7Department of Women's and Children's Health, Uppsala University, Uppsala, Sweden. 8Department of Obstetrics and Gynecology, Södersjukhuset, Stockholm, 11883, Sweden. 9Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Uppsala, 75237, Sweden. luisa.hugerth@scilifelab.se.

BACKGROUND: Differences in the gut microbiome between lean and obese individuals, even twins, have been recognized for over a decade. The causative role of the microbiome in obesity is known from both mouse and human studies. In parallel, the gut microbiome has been implicated in the most common complications of pregnancy, including preterm birth, gestational diabetes mellitus, and gestational hypertension. Despite obesity being a well-established risk factor for these complications, the composition of the gut microbiome of obese pregnant individuals has not yet been studied. Here, we have examined the differences in the faecal microbiota between lean (n = 746) and obese (n = 254) pregnant women in Sweden, at two timepoints in pregnancy. RESULTS: The differences in the faecal microbiome of one thousand lean and obese persons persist during gestation. Obese individuals have a less diverse and less rich microbiome throughout all trimesters of pregnancy. In the first trimester, 9 species differ significantly between lean and obese individuals, and 35 in the early third trimester, after adjusting for confounders. However, only one species remained significant after further adjusting for bowel transit time, which differed significantly between lean and obese participants. Additionally, obese individuals harbored a consistently lower potential to produce propionate in their gut microbiomes, even after adjustments. CONCLUSIONS: The faecal microbiome adapts to pregnancy with

some key differences between lean and obese mothers, even in late pregnancy. Crucially, there is an over-abundance of opportunistic pathogens in the microbiome of obese pregnant women in the third trimester. This may be a potential underexplored mechanism explaining the increased rates of pregnancy complications among obese patients. Diet, probiotics, or medication interventions to correct the gut microbiome of pregnant obese individuals could potentially improve their pregnancy outcomes.

---

7. If I If I have a urinary tract infection caused by Escherichia coli should I worry about my dog?

E. coli, UPEC, dogs, WGS, machine-learning

Biel Garcias Puigserver1 , Astrid von Mentzer1 1 Deparment of Microbiology and Immunology, University of Gothenburg

Background: Urinary tract infections (UTIs) are among the most common bacterial infections worldwide, leading to extensive antimicrobial use and contributing to antimicrobial resistance (AMR). However, their epidemiology is still largely approached from a human perspective, overlooking the potential role of animals. Dogs, due to their close contact with humans, could represent a reservoir of uropathogenic Escherichia coli.
Methods: We analyzed 3,055 E. coli genomes from canine (n = 2,236) and human (n = 719) urinary isolates available in NCBI, originating from studies without known AMR-related sampling bias. We investigated host-associated genomic markers using sequence typing and linear mixed models and developed ridge regression and random forest models to predict host origin and identify potential markers of cross-host circulation.
Results: Some sequence types (STs) were associated with dogs (ST372, ST998) or humans (ST131, ST73), while others (ST127, ST1193) showed no clear host association, suggesting possible cross-host circulation. Several host-associated genes were identified, including ldrD (a toxin–antitoxin polypeptide system) in dogs and the afa operon (adhesion and urinary colonization) in humans. Interestingly, some of these genes clustered in neighboring genomic regions.

A ridge regression classifier achieved high performance in within-dataset validation (AUC = 0.966, accuracy = 0.93, F1-score = 0.85), classifying 5.9% of human isolates as canine-like. Additionally, 458 isolates (15%) showed an undetermined host origin, particularly among shared (25.2%) and human-associated STs (27.2%). Finally, a random forest model identified the fec operon as the strongest predictor of human host association, suggesting its potential use as a PCR marker to identify canine isolates with human-associated genomic characteristics.

Conclusions: So, should we worry about our dogs? Our results show that, although canine and human urinary E. coli populations have distinct genomic characteristics, some lineages and genetic markers are shared between them. This suggests that dogs could play a role in the circulation of strains associated with human UTIs, although direct transmission cannot be confirmed due to the inherent limitations of the studies based on public available data. The identification of the fec operon as a potential marker of human-associated strains could also help identify canine isolates of interest for further surveillance. Overall, our findings suggest that dogs should not be overlooked when studying the epidemiology of UTIs and their potential contribution to AMR.

8. Gut microbiota compositional trajectories explain drug response in Type 2 Diabetes clusters

Type 2 diabetes, Gut microbiome, Drugs

Sharaf, Chetan, Department of Clinical Microbiology (Climi), Umeå University) (Rosengren, Anders, Department of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg) (Dwibedi, Chinmay, The Laboratory of Molecular Infection Medicine (MIMS), Umeå University | 3Department of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg

Clinical responses to type 2 diabetes therapies exhibit substantial interindividual variability. While multiple stratification frameworks have been developed, their ability to explain differential treatment outcomes remain largely unexplored. We hypothesize that combining pathophysiological disease subtypes with gut microbiota profile will enhance the prediction of drug response and facilitate more effective, personalized interventions to slow disease progression.

In this study, we enrolled 174 individuals classified within the severe clusters of T2D specifically, those with Severe Insulin Resistant Diabetes (SIRD) or Severe Insulin Deficient Diabetes (SIDD). Participants were randomized in an open-label trial to receive either a Glucagon-Like Peptide 1 receptor agonist (GLP1ra) or a Sodium-Glucose Cotransporter 2 inhibitor (SGLT2i) for six months. Gut microbiota, blood glucose levels (HbA1c), and kidney function (eGFR) were profiled at baseline, three months, and six months, while additional secondary clinical parameters including BMI, LDL cholesterol, systolic blood pressure (SBP) and many others, were collected at the baseline and end of the trial.

Our findings demonstrate that individuals across different clusters and treatment arms follow distinct microbiota compositional trajectories over time. These trajectories help explain how each drug influences the broader clinical profile of the individual, suggesting a meaningful interaction between drug mechanism, disease pathophysiology, and the gut microbiota.

---

9. KaroSpace: a rapid-access framework for interactive exploration of multi-sample spatial omics data

Spatial, viewer, multiomics

Christoffer M. Langseth, Bastien Hervé, Hanna P. Piechaczyk, Yuk Kit Lor, Ting Sun, Gonçalo Castelo-Branco

Spatial omics technologies enable high-resolution mapping of molecular and cellular organization within tissues, yet interactive exploration of these data remains challenging due to computational bottlenecks, and reliance on proprietary software infrastructures. We present KaroSpace, a framework for cell-centric exploration of multi-sample and multi-modality spatial omics data, agnostic to upstream analysis pipelines and operating on preprocessed spatial features. By combining a lightweight design with flexible deployment options, KaroSpace allows immediate interactive spatial data exploration, straightforward implementation to the broader scientific community, supports transparent data sharing, and complements existing computational workflows for hypothesis generation and collaborative analysis.

---

10. Organismal development as a source of evolutionary predictability
Development, Dynamical Systems Theory, Machine Learning, Evolution

Netz, Christoph, Scilifelab, Stockholm University; Jirout, Matej, Scilifelab, Stockholm University; Uribarri, Gonzalo, Department of Computer and Systems Sciences, Stockholm University; Milocco, Lisandro, Scilifelab, Stockholm University

Development generates phenotypic variation in a biased way, making some variants more likely to appear than others. Thus, the arrival, not just the survival, of the fittest shapes evolution.

Modeling this bias matters for evolutionary prediction, but it is hard: development is a complex, nonlinear process. However, perturbations from different sources (mutation, environment, noise) have been observed to produce aligned phenotypic effects. This means that variation is effectively low-dimensional, and so may be learnable despite that complexity. In this project, we develop methodologies to learn data-driven models of development from time series, with the goal to leverage different sources perturbations for evolutionary prediction.

---

11. Proteomics Platform
Affinity Proteomics; mass spectrometry; protein biomarkers

Claudia Fredolini

---

12. Environmental and Lifestyle Exposures Linked to the Temporal Variability of Gut Microbial Species in Healthy Adults
Gutmicrobiome, Exposure, Variability, Persistance, Healthy

Talukdar Daizee1,2 , Chanda Deep.1,2, Shternshis Andrey.3, Papazian Stefano.4, Singh Prashant.3 , Dwibedi Chinmay.1,2# 1 Department of Clinical Microbiology, Umeå University, Sweden 2 The Laboratory for Molecular Infection Medicine, Umeå University, Sweden 3 Department Uppsala of Information Technology, Division of Scientific Computing, University, Sweden 4 Department of Environmental Science, Stockholm University, Sweden*

The human gut microbiome is a dynamic ecosystem that contributes to host metabolism and immune homeostasis. Individuals are continuously exposed to environmental and lifestyle-related factors that may shape microbial growth and community composition. We hypothesize that certain exposures exert selective pressures on specific microbial species, contributing to temporal variability within individuals. To identify these factors, we analyzed longitudinal gut microbiome and metabolome profiles from 43 healthy Swedish adults sampled at four time points over a one-year period.

This study aims to identify specific metabolites and exposure-related features associated with the temporal variability of individual gut microbes, and to determine whether distinct exposures or group of exposure classes are linked to overall community composition and its longitudinal trajectory. We identified a high-variability bacterial block composed of taxa that fluctuated markedly over time, consistent with a temporally dynamic microbial component. Several of these bacterial species showed significant longitudinal associations with exposure-linked and host-associated metabolites, including food additives, plant-

derived compounds, and bile acids. Importantly, subject-level trajectory plots for the strongest pairs supported coordinated within-individual temporal patterns, indicating that these associations reflect dynamic coupling rather than static cross-sectional correlation. Machine learning analysis further highlighted food additive-related metabolite among the top predictors of overall microbiome community variation, suggesting that exposure-linked metabolites may contribute both to microbe-specific temporal behavior and to broader community-level ecological effects.

---

13. Adaptive segmentation of 3D cryo-EM reconstructions using feature-based clustering
Cryo-EM, density map segmentation, unsupervised clustering, Gaussian mixture model, feature ablation.

Kuvshinova Dariia, Department of Physics, Chemistry and Biology, SciLifeLab, Linköping University; Forsberg Björn, Department of Physics, Chemistry and Biology, SciLifeLab, Linköping University

Cryo-EM reconstructions contain substantial local variation in density and structural characteristics, making reliable separation of macromolecular regions from solvent challenging. Conventional density-based masking may be sensitive to this variation. As a more adaptive approach, we investigate unsupervised clustering using local patch-level features. A two-component Gaussian mixture model is applied to distinguish macromolecular regions from solvent, while feature ablation is used to identify informative features. The approach is evaluated across diverse cryo-EM reconstructions, and predicted masks are compared with PDB-derived structural references. Results demonstrate that local feature-based clustering enables protein-solvent segmentation of heterogeneous density maps and provides a foundation for exploring more detailed structural segmentation.

---

14. Artificial intelligence-driven diagnostic support for antibiotic-resistant infections
antibiotic resistance, artificial intelligence, conformal prediction

Lund, David, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden/Centre for Antibiotic Resistance Research (CARe), University of Gothenburg, Gothenburg, Sweden Käll, Styrbjörn, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden/ Knightec Group Sweden AB, Gothenburg, Sweden Dedic, Merima, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden Inda-Díaz, Juan S., Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden/Centre for Antibiotic Resistance Research (CARe)/School of Biomedical Sciences, Queensland University of Technology, Brisbane, Australia Hessel, Magnus, Centre for Antibiotic Resistance Research (CARe), University of Gothenburg, Gothenburg, Sweden/Department of Clinical Microbiology, Sahlgrenska University Hospital, Gothenburg, Region Västra Götaland, Sweden Ruhe-van der Werff, Suzanne, Department of Medicine Solna, Division of Infectious Diseases, Karolinska Institutet, Stockholm, Sweden/ Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden Nauclér, Pontus, Department of Medicine Solna, Division of Infectious Diseases, Karolinska Institutet, Stockholm, Sweden/Department of Infectious Diseases, Karolinska University Hospital, Stockholm, Sweden Skovbjerg, Susann, Centre for Antibiotic Resistance Research (CARe), University of Gothenburg, Gothenburg, Sweden/Department of Clinical Microbiology, Sahlgrenska University Hospital, Gothenburg, Region Västra Götaland, Sweden/Department of Infectious Diseases, Institute of Biomedicine, University of Gothenburg, Gothenburg, Sweden

Giske, Christian G., Department of Clinical Microbiology, Karolinska Institutet, Stockholm, Sweden/Department of Laboratory Medicine, Karolinska Institutet, Stockholm, Sweden Kahlmeter, Gunnar, Department of Clinical Microbiology, Central Hospital, Växsjö, Sweden Johnning, Anna, Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden/Centre for Antibiotic Resistance Research (CARe), University of Gothenburg, Gothenburg, Sweden/Department of Systems and Data Analysis, Fraunhofer-Chalmers Centre, Gothenburg, Sweden Kristiansson, Erik, Chalmers University of Technology and University of Gothenburg, Gothenburg, Sweden/Centre for Antibiotic Resistance Research (CARe), University of Gothenburg, Gothenburg, Sweden

Efficient treatment of bacterial infections requires rapid and accurate diagnostics. Current cultivation-based methods can be slow, endangering patients and contributing to the overprescription of broad-spectrum antibiotics. To address this, we have developed a deep learning framework based on transformers and conformal prediction to assess antibiotic susceptibility in clinical isolates using patient data and incomplete diagnostic information. We integrated this framework with a unique dataset representing antibiotic susceptibility test results generated by three major Swedish clinical labs over ten years, including more than 1.4M clinical isolates and more than 7.9M antibiotic susceptibility tests. Our model can predict susceptibility to 40 antibiotics across 11 bacterial species with overall high accuracy even based on very limited information. Moreover, we integrated the framework with conformal prediction to provide an accurate estimate of uncertainty with each prediction, further limiting the number of errors produced by the model. Together, our results demonstrate the potential of artificial intelligence-based decision support as a new means to combat the growing antibiotic resistance crisis.

---

15. Conserved mycobacteriophage protein uses structural mimicry to displace host initiation factors

Mycobacteriophage, Structural mimicry, Transcription factors, RNA Polymerase, Transcriptional host takeover

Parker, David R. — Karolinska Institutet, Science for Life Laboratory, Solna, Sweden Ghosh, Shrestha — Karolinska Institutet, Science for Life Laboratory, Solna, Sweden Dujiskova, Adela — The Rockefeller University, New York, USA Olinares, Paul Dominic B. — The Rockefeller University, New York, USA Danielsson, Sarah Narrowe — Stockholm University, Science for Life Laboratory, Solna, Sweden Prensa, Grisna Isabel — Karolinska Institutet, Science for Life Laboratory, Solna, Sweden Kim, Doyeon — Karolinska Institutet, Science for Life Laboratory, Solna, Sweden Rosenberger, Florian Andreas — Max Planck Institute of Biochemistry, Martinsried, Germany; Karolinska Institutet, Solna, Sweden Elofsson, Arne — Stockholm University, Science for Life Laboratory, Solna, Sweden Mann, Matthias — Max Planck Institute of Biochemistry, Martinsried, Germany; University of Copenhagen, Copenhagen, Denmark Campbell, Elizabeth — The Rockefeller University, New York, USA Fossati, Andrea — Karolinska Institutet, Science for Life Laboratory, Solna, Sweden

Compact viral genomes encode sophisticated strategies to reprogram host gene expression, often by co-opting rather than replacing RNA polymerase. How bacteriophages achieve such control in Mycobacteria remains largely unknown. Here, time-resolved interaction proteomics of mycobacteriophage D29 infection identifies a conserved two-protein module, gp53-gp52, that binds the host RNA polymerase. Affinity purification confirmed association of both proteins with the polymerase, while native mass spectrometry demonstrated that gp53 displaces the host sigma factor. Structural modelling places gp53 at the sigma-binding surface of RNA polymerase, consistent with molecular mimicry of sigma-factor engagement.

The module is conserved throughout Cluster A mycobacteriophages and occurs in fused or split genomic architectures. These findings suggest sigma-factor displacement as a mechanism of transcriptional takeover in viruses of mycobacteria and show how a compact viral module can repurpose an essential host molecular machine.

---

16. Temporal Dynamics of the Gut Microbiota and Host Metabolic Regulation in Type 2 Diabetes

Type 2 diabetes; Gut microbiome; Longitudinal stability; Intraclass correlation coeLicient; Biomarker discovery

Chanda D., Sinha N., Rosengren A., Dwibedi C.

Gut microbiota alterations are consistently linked to type 2 diabetes (T2D). However, both T2D and gut microbiome are highly dynamic, with substantial intra-individual temporal variation leading to limited reproducibility of gut microbial markers across cross sectional studies. We hypothesize that longitudinally stable microbial features, identified across independent populations, are essential for discovering robust and clinically relevant T2D biomarkers.
Using matched longitudinal shotgun metagenomic data from Swedish T2D (n=45, 5 timepoints) and healthy (n=75, 4 timepoints) cohorts, we quantified microbial stability using intraclass correlation coeLicients (ICC) and coeLicients of variation (CV). T2D individuals showed significantly higher intra-individual compositional variability, indicating fundamentally less stable gut ecosystems under chronic metabolic stress.
Tracking how species shifted their stability profiles between health and T2D revealed ecologically distinct groups. Some health-associated taxa, including Christensenella minuta and Faecalicatena sp. Marseille-Q4148, lost their intra-individual stable abundance patterns in T2D, becoming fluctuating across timepoints, likely reflecting the volatile gut environment driven by glycemic variability, medication, and chronic inflammation. These taxa may potentially respond to T2D selection pressures and can be targets for intervention. In contrast, other species gained intra-individual stability under T2D conditions, including opportunistic oral bacteria such as Streptococcus mutans and depleted SCFA producers including Ruminococcus lactaris and Ruminococcus bicirculans, representing potentially reliable disease biomarkers. Based on this framework, we integrate multi-kingdom stability profiles with lifestyle, psychosocial, and clinical variables (HbA1c, HOMA-IR) using mixed-eLects models and XGBoost to identify microbiota mechanistically linked to metabolic dysregulation in T2D. Our stabilitystratified framework moves beyond snapshot microbiome comparisons, oLering a path toward reproducible biomarker discovery for precision T2D management.

---

17. Beyond the Bump: Oral health as an overlooked dimension of perinatal care.

Oral health, Pregnancy, Women health, Oral microbiome

Bameri, Delaram, Department of Medical Biochemistry and Microbiology, Uppsala University, Sweden/ Bostanci, Nagihan, Department of Dental Medicine, Karolinska Institutet, Sweden/ Brusselaers, Nele, Department of Women's and Children's Health, Karolinska Institutet, Sweden/ Schuppe-Koistinen, Ina, Department of Microbiology, Tumour and Cell Biology, Karolinska Institutet, Sweden/ Engstrand, Lars, Department of Microbiology, Tumour and Cell Biology, Karolinska Institutet, Sweden/ Hugerth, Luisa W., Department of Medical Biochemistry and Microbiology, Science for Life Laboratory, Uppsala University, Sweden

Background: Oral health is an underappreciated dimension of perinatal care. Pregnancy-related hormonal changes increase susceptibility to gingivitis and periodontitis, and periodontal disease has been associated with adverse pregnancy outcomes such as preterm birth and low birth weight in multiple studies. Despite this elevated risk profile, dental care attendance during pregnancy remains low globally, and most prior studies treat attendance as a single cross-sectional outcome rather than examining longitudinal engagement patterns across the full peripartum window.

Methods: We used the SweMaMi cohort (2017–2022), linked to the Swedish Quality Registry for Caries and Periodontal Disease (SKaPa), to follow 4,136 women across seven peripartum windows spanning 2 years before to 2 years after pregnancy. Women were classified by prior dental visit status (prior visit, n=1,573; no prior visit, n=2,563). Odds of attendance were estimated by logistic regression, visit intensity by negative binomial regression, visit-reason distribution by chi-square, periodontal and caries trajectories by CLMM and GLMM respectively.

Results: Time-adjusted dental visit rates declined from a pre-pregnancy peak of 4.02% per month (6-12 months before conception) to 2.47% per month during pregnancy as the lowest point across the entire window. Women with a prior visit were far more likely to attend during pregnancy (40.6% vs. 14.9%; OR≈3.9, $p<0.001$), and this gap never closed at any later timepoint. Examination/follow-up was the dominant visit reason overall (45.6%) with significant differences between the subgroups at some observed time periods. Active cares were the more common cares sought among the prior-visit group before pregnancy, with a shift towards preventive related cares after delivery. Significantly worsened condition was observed for both periodontal and caries condition after delivery, after adjusting for covariates.

Conclusion: Dental engagement falls at pregnancy when biological risk is highest. This decline is strongly associated with pre-pregnancy attendance history. Oral health conditions worsened after pregnancy, confirming the accumulative pattern of oral disorders and the importance of oral care intervention initiation prior to pregnancy. These findings indicate that the principal barrier is initiation of care, not visit intensity once engaged. Moreover, pregnancy represents a missed prevention window, and efforts to improve perinatal oral health must proactively reach women without established dental care relationships rather than relying on attendance-based outreach alone.

---

18. SHE: Structural History of Eukarya
Protein structure prediction, evolution

Diandra Daumiller, Qiuzhen Li, Patrick Bryant. Science for Life Laboratory, Department of Molecular Biosciences, The Wenner-Gren Institute, Stockholm University

Comparative genomics has traditionally relied on sequence-based markers, yet the global evolutionary landscape of protein architecture remains largely unexplored. We present the Structural History of Eukarya (SHE), a structural similarity-derived phylogeny from all-vs-all structural comparisons on the proteome level of 1,542 eukaryotic species, encompassing nearly 300 trillion pairwise protein alignments. This proteome-scale analysis reveals a bipartite model of eukaryotic evolution: a rigid Strict Core dominated by cytoskeletal architecture, supporting a more plastic Operational Engine centred on translational machinery. We identify lineage-specific accelerations of structural evolution in species such

as birds and ants that are decoupled from proteome expansion. Finally, we demonstrate that proteome-wide structural fidelity enables the functional annotation of uncharacterised proteins and a data-driven framework for model organism selection, replacing heuristic choices with quantitative matching to human biological processes. SHE is freely accessible as an interactive portal: https://she-app.serve.scilifelab.se/

---

19. A Paradoxical Success Story of Polyploids
evolutionary genomics, plant biology, ploypoidy

Pehlivanoglu, Ege (1); Sandeen, Rhianah (1); Renner, Denzel (1); Pan, Xiaoyang (1); Šrámková, Gabriela (2); Kolar, Filip (2,3,4); Koch, Marcus (5,6,7); Mandáková, Terezie (8); Yant, Levi (1)
*Affiliation of all authors * 1- SLU - Linnean Centre for Plant Biology, Department of Plant Biology, Swedish University of Agricultural Sciences, Uppsala, Sweden; 2- Department of Botany, Faculty of Science, Charles University, Prague, Czech Republic 3- Natural History Museum, University of Oslo, Oslo, Norway; 4- Institute of Botany, The Czech Academy of Sciences, Průhonice, Czech Republic; 5- Center for Organismal Studies, Biodiversity and Plant Systematics, Heidelberg University, Heidelberg, Germany; 6- Cluster of Excellence GreenRobust, Heidelberg University, Heidelberg, Germany; 7- Heidelberg Center for the Environment, Heidelberg University, Heidelberg, Germany; 8- Laboratory of Plant Cytogenomics, Central European Institute of Technology (CEITEC), Masaryk University, Brno, Czech Republic*

Whole-genome duplication (WGD) is a macromutation that reshapes evolutionary trajectories by increasing genetic variation, including structural variants (SVs). WGD doubles the number of chromosomes (leading to polyploidy) and hence affects evolutionary dynamics by increasing all mutation classes, reducing the impact of purifying selection, and possibly destabilizing chromatin organization and centromere architecture. Recent longread based population scale work in plants has shown that WGD increases SV accumulation, genetic diversity, and genetic load in both Cochlearia (Hämälä et al., 2024) and Arabidopsis arenosa (Vlček et al., 2025), with the latter study integrating empirical data with forward-time population genetic modelling to take first steps in disentangling the opposing forces caused by WGD at the selective and population levels. These studies used long reads to profile SVs at the population scale, but they did not use referencefree pangenomes, nor did they enable generalization of principles by applying a comparative context.
This newly started PhD project addresses these gaps through comparative pangenomic and multi-omic analyses across paired diploid and autopolyploid cytotypes, initially focusing on A. arenosa and Cochlearia. Leveraging existing long-read pangenomes encompassing hundreds of phased haplomes (built with HiFi, Hi-C, and CiFi data), we will establish computational pipelines to intersect SV landscapes with transcriptomic, chromatin accessibility (ATAC-seq), and 3D genome architecture data. We aim to build predictive, integrative models linking structural genomic features to regulatory outcomes and allele frequency dynamics across ploidy levels. Cross-system comparisons will test whether WGD drives convergent functional signatures, thereby advancing our understanding of polyploid genome evolution. Here, we present the conceptual framework, initial pipeline architecture, and planned analytical strategy for this PhD project.

Hämälä, T., Moore, C., Cowan, L., Carlile, M., Gopaulchan, D., Brandrud, M. K., Birkeland, S., Loose, M., Kolář, F., Koch, M. A., & Yant, L. (2024). Impact of whole-genome duplications on structural variant evolution in Cochlearia. Nature Communications, 15(1), 5377. https://doi.org/10.1038/s41467-024-49679-y Vlček, J., Hämälä, T., Vives Cobo, C., Curran, E., Šrámková, G., Slotte, T., Schmickl, R., Yant, L., & Kolář, F. (2025). Whole-genome duplication

increases genetic diversity and load in outcrossing Arabidopsis arenosa. Proceedings of the National Academy of Sciences, 122(31), e2501739122. https://doi.org/10.1073/pnas.2501739122

---

20. FEGA Sweden – Repository for sharing Swedish biomedical research data
Sensitive human data, FEGA Sweden, Controlled-access sharing, FAIR data standards, Legal Compliance

Elisavet Torstensson-NBIS/UU, Maria Vrettou-NBIS/UU, Britt Andermann-NBIS/UU

Human genomic and phenotypic data are essential for biomedical research, but strict privacy regulations like GDPR limit how sensitive human data can be shared. FEGA Sweden solves this challenge by serving as the Swedish national node within the Federated European Genome-phenome Archive.

Operated by NBIS at SciLifeLab, FEGA Sweden provides a secure platform for the permanent deposit and controlled-access sharing of human biomedical data generated in Sweden. Data remains safely stored within national borders while managed Data Access Committees (DACs) ensure access is granted only to approved researchers. By issuing unique accession numbers, FEGA Sweden enables researchers—including those at Uppsala University and Lund University—to fulfil scientific journal requirements and FAIR data standards without compromising participant privacy or legal compliance.

---

21. Beyond FID: How should we evaluate Generative Models for Fluorescence Microscopy?
Generative models, Microscopy, Evaluation, Metrics

Emma Latron, Sebastian Bauer, Harshith Bachimanchi, Vsevolod Viliuga, Juliette Griffié. All others are affiliated with Scilifelab and the department of biochemistry and biophysics at Stockholm University

Generative models synthesize microscopy images of cells under drug or genetic perturbations, offering a scalable complement to costly wet-lab experiments. However, evaluating the biological accuracy of these generated images remains a challenge. The standard metric, Fréchet Inception Distance (FID), compares real and generated image distributions using features extracted from a model pretrained on natural images. This raises a fundamental question: do these representations capture phenotypic variation, or merely low-level visual realism? FID being the primary metric to evaluate these generative models, ensuring its reliability is crucial; as otherwise models risk producing visually plausible images that lack underlying biological signal, ultimately compromising downstream analysis.
We address the limitations of evaluating generative models in cellular imaging by introducing a benchmarking strategy that quantifies how effectively different feature representations isolate true biological signal from technical noise, demonstrating that models pretrained on natural images are often confounded by domain shifts.

22. Resolving membrane packing and lipid phase behaviour using Excitation-Emission Matrix spectroscopy
Experimental Methods, Data Analysis

Mattiuz Eric, Stockholm University Wagner Julian P., Stockholm University, van Beers Pauline, Stockholm University, Heberle Frederick A., Stockholm University, Doktorova Milka, Stockholm University

Studying biological membrane properties is essential for understanding how cells carry out specific tasks to maintain cellular homeostasis. To study such properties, environment-sensitive fluorescent lipid analogues can be incorporated into the membrane, as they respond to local environment changes with a shift in their emission spectra. This dye behaviour is most commonly quantified through the generalized polarization (GP) metric, which captures the red- and blue-shifts in the emission spectrum of the fluorophore. While GP is a reliable descriptor within a given lipid system, its utility diminishes considerably when comparing across chemically distinct membrane compositions, most likely because it reduces the spectral information to only two fixed wavelengths, limiting its utility in more complex systems. Excitation-emission matrix (EEM) spectroscopy, on the other hand, exploits not only the entire emission but also the entire excitation spectrum of the dye, providing a richer and more complete interpretation of the fluorophore readout.
Here we present a new methodological framework based on EEM spectroscopy, analysed through phasor transformation. By using sine and cosine functions, an EEM-phasor matrix is generated, providing a platform for determining and separating changes in bilayer properties arising from distinct physical contributions, such as chemical effects (packing modulated by cholesterol) and thermal changes (driven by temperature increase). This analysis captures changes in the fluorophore's spectral environment, which reflect the local hydration state of the bilayer membrane, hence a proxy for lipid packing, and can also reveal the coexistence of liquid-ordered and liquid-disordered phases that arise from cholesterol-rich systems.

---

23. NBIS: a national bioinformatics infrastructure at a university near you!
Bioinformatics

Proux-Wéra, Estelle, Department of Biochemistry and Biophysics, National Bioinformatics Infrastructure Sweden Science for Life Laboratory, Stockholm University, Solna, Sweden

NBIS (National Bioinformatics Infrastructure Sweden) provides bioinformatics and data science support to the Swedish life science research community.

We provide a wide spectrum of services, including advanced bioinformatics analysis, bioimage informatics, data management, imaging AI support, systems and tools development, and support to national compute resources.

---

24. Impact of input dimensionality on data-driven prediction of carotid plaque transcriptomics from non-invasive CT Angiography
Atherosclerosis, Deep Learning, Medical Imaging, Transcriptomics

Stamos, Evangelos, Department of Molecular Medicine and Surgery, Karolinska Institutet, Solna, Sweden; Buckler, Andrew, Department of Molecular Medicine and Surgery, Karolinska Institutet, Solna, Sweden; Matic, Ljubica, Department of Molecular Medicine and Surgery, Karolinska

Institutet, Solna, Sweden; Hedin, Ulf, Department of Molecular Medicine and Surgery, Karolinska Institutet, Solna, Sweden and Department of Vascular Surgery, Karolinska University Hospital, Stockholm, Sweden; Marlevi, David, Department of Molecular Medicine and Surgery, Karolinska Institutet, Solna, Sweden and Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, USA

Background: Carotid plaque instability, a major contributor to ischemic stroke, is governed by a complex interplay of morphological features, biomechanical properties and biological processes that together drive plaque and disease progression. In particular, intraplaque molecular and cellular activity underpins developments of so called vulnerable phenotypes, with specific hallmarks (active inflammatory processes; proteolysis and extracellular matrix degradation; angiogenesis and intraplaque hemorrhage; apoptosis and necrotic core expansion; etc.) associated with more risk-prone clinical behaviour. However, in a clinical setting, such information is restricted to post-operative ex-vivo analyses, assessing already explanted plaque samples. As an alternative, recent advances in machine learning offer the ability to learn biological information from medical imaging data, offering means for non-invasive biological phenotyping. However, while primarily developed and showcased for oncological purposes, only partially exploratory investigations exist for atherosclerotic phenotyping. To address this, we here sought to systematically investigate the prediction of carotid plaque transcriptomic profiles from preoperative computed tomography angiography (CTA) using deep learning, exploring performance across dimensional architectures.

Materials and methods: Carotid endarterectomy specimens were identified in the Biobank of Karolinska Endarterectomies with pre-interventional CTA and post-interventional bulk RNA sequencing available (n=104). Reconstructed CTA images were analyzed using the ElucidVivo software, including segmentation and 3D reconstruction of intraplaque morphologies including calcification (CALC), lipid-rich necrotic core (LRNC) and intraplaque hemorrhage (IPH). A compact residual network was trained on three input representations of plaque morphology: quantitative plaque metrics(1D), an axial cross-section of plaque tissue composition at the midpoint of proximal plaque (2D), and a 64^3 voxelized volume with separate CALC, LRNC and IPH channels (3D). Expression of 2593 transcripts was predicted via regression and binary classification (dichotomized at the cohort mean). Models were trained across three output strategies: (i) individually per transcript, (ii) jointly across all 2593 transcripts, and (iii) grouped into four curated biological families comprising 28 transcripts (lipids, matrix, IPH, and CALC). Performance was evaluated using 5-fold cross-validation, assessing correlation coefficients and root-mean-square error (RMSE) for regression, and area under the curve (AUC) for classification, benchmarked against an untrained baseline model of identical architecture.

Results: Across 1D, 2D, and 3D inputs, single-transcript models significantly predicted 97, 35 and 18 transcripts in regression, and 129, 39 and 26 in classification, respectively. On the contrary, joint modeling of all 2593 transcripts yielded no significant predictions. For the 28 family transcripts, biological family-wise modeling outperformed size-matched random transcript groupings, increasing significant predictions from 15 to 30 in 1D regression and from 2 to 16 in 1D classification.

Conclusion and Outlook: While prediction of biological information from medical image data is possible, performance is directly contingent on input data description with distinct differences observed across 1, 2, and 3D networks, with increase in data input substantially

impacting performance. While various reasons may underpin this (dimensionality overfitting; data redundancies; vanishing gradients; etc.), grouping of related transcripts enhanced overall predictive performance across networks, indicating that biologically informed multi-output learning could strengthen non-invasive transcriptomic inference from preoperative CTA.

---

25. Exploring Fragile-Site Genomic Variation as Endogenous Lineage Tracing Markers in Neuroblastoma

Neuroblastoma; Common Fragile Sites; Endogenous Molecular Barcoding; Long-Read Sequencing; Lineage Tracing

Fayazbakhsh, Farzaneh — Swedish University of Agricultural Sciences (SLU), Sweden; Karolinska Institutet, Sweden Liu, Cheng-De (Ernest) — Department of Oncology-Pathology, Karolinska Institutet, Sweden Karlsson, Kasper — Department of Oncology-Pathology, Karolinska Institutet, Sweden Bongcam Rudloff, Erik — Swedish University of Agricultural Sciences (SLU), Sweden

Background: Understanding tumor heterogeneity and the dynamics of cancer cell subpopulations is essential for investigating tumor evolution and treatment response. Endogenous genomic variations, particularly those associated with common fragile sites, may provide naturally occurring molecular barcodes for distinguishing cellular subpopulations without introducing external genetic labels.

Methods: This study investigated the potential of fragile-site-associated genomic variation as endogenous lineage markers in two neuroblastoma cell lines, SK-N-SH and SK-N-BE(2). Selected genomic regions, including FRA3B, FRA16D, and neuroblastoma-associated loci, were targeted using long-range PCR and Oxford Nanopore long-read sequencing. A computational workflow was applied to characterize sequence variation, identify haplotype patterns, and compare molecular profiles across experimental conditions, including drug-treated and untreated samples.

Results: Sequencing and computational analyses identified distinct sequence-level variation and haplotype patterns within successfully amplified genomic regions. In particular, reproducible haplotype configurations were observed in selected chromosome 11 and chromosome 16 amplicons, revealing molecular heterogeneity within the investigated cell lines. Comparisons between experimental conditions also revealed modest differences in the frequencies of selected variants. These findings suggest that combinations of naturally occurring genomic variants may help distinguish molecular subpopulations, although their stability and lineage specificity require further validation.

Conclusions: This study demonstrates the feasibility of combining targeted long-read sequencing and computational haplotype analysis to investigate endogenous genomic variation in neuroblastoma. The findings provide preliminary support for exploring naturally occurring genomic patterns as potential molecular barcodes and establish a foundation for future studies of cancer cell heterogeneity and lineage dynamics.

26. Single-cell, clonal and spatial atlases of cranial placodes illuminate their specification and evolution
scRNA-seq. Clonal Tracing.

Aliia, Murtazina, Karolinska Institutet. Yuliia, Fatieieva, Medical University Vienna. Felix, Waern, Karolinska Institutet. Helen R ,Maunsell Baylor College of Medicine. Ankita, Thawani, Washington University School of Medicine. Bettina, Semsch, Karolinska Institutet. Johan, Bostroem, Medical University Vienna. Caleb C, Reagor, Karonlinska Institutet. Polina, Kameneva, St. Anna Children's Cancer Research Institute. Karina, Araslanova, Medicanl University Vienna. Sergei, Isaev, Medical University Vienna. Franziska, Schelb, Hopp Children's Cancer Center Heidelberg. Kaj, Fried, Karolinska Institutet. Alek, Erickson, Stockholm University. Alexander, Klimovich, Christain-Albrechts University of Kiel. Adnrea, Streit, King's College. Lena M, Kutscher, Hopp Children's Cancer Center Heidelberg. Iryna, Kozmikova, Institute of Molecular Genetics of the Czech Adademy of Sciences. Zbynek, Kozmik, Institute of Molecular Genetics of the Czech Academy of Sciences. Emmar R, Andersson, Karolinska Institutet. Gerhard, Schlosser, University of Galway. Andrew K, Groves, Baylor College of Medicine. Igor, Adameyko, Medical University Vienna

Cranial placodes are specialized ectodermal thickenings that generate major sensory structures of the vertebrate head, yet their identity, diversification, and evolutionary origin remain unclear. Together with the neural crest, they arise at the neural plate border, a region whose patterning has long been enigmatic. Here, we present an integrated single-cell, spatial, and clonal atlas of the placodes across lineage segregation. Combining single-cell RNA-sequencing, spatial transcriptomics, and high-resolution clonal tracing, we show that placodal and neighboring progenitors form a transcriptional landscape with gradual transitions between domains. Boundary cells co-express markers of adjacent territories, consistent with transient bipotent states. Clonal analysis reveals frequent sharing of progenitors between neighboring placodes, supporting a model of competitive segregation. Comparative analysis with Amphioxus suggests that vertebrate olfactory placodes emerged from an ancestral neuroectoderm that later resolved into distinct neural and olfactory domains, with the olfactory system representing an early specialization of the ancient anterior neural tube.

---

27. Fit for which purpose? Evaluating differentially private synthetic clinical cohorts on public data: utility per use, attack resistance, and a certificate that prints its contraindications
synthetic health data; differential privacy; data quality evaluation; European Health Data Space; secondary use

Chattarki, Gaurav Raj, Synthetix Health, Stockholm

Synthetic health data is increasingly proposed as a route to secondary use under GDPR and the European Health Data Space, yet a synthetic dataset is neither "good" nor "bad" in the abstract: it is fit or unfit for a stated purpose, and generators without a formal privacy guarantee can reproduce real records while passing similarity-based privacy checks. We present an evaluation framework and a worked evaluation for synthetic clinical cohorts generated under ε-differential privacy. The framework orders purposes by tolerance for synthesis error (software testing, trial feasibility, model pre-training, hypothesis generation; and, as contraindications, rare-event detection, safety-signal detection and regulatory evidence), assigns utility measures to each purpose (schema and referential-integrity checks, marginal and pairwise divergence, train-on-synthetic-test-on-real gaps, replication of known associations), and treats privacy as two attacks, membership inference and

linkage, confirmed against a disclosed (ε, δ) budget rather than substituted for it. Applied to a 5,000-person OMOP cohort synthesized from the public CMS DE-SynPUF data at ε = 1.0, $\delta = 10^{-5}$, the evaluation reports 7 of 7 referential-integrity checks passed, composite marginal concordance of 0.91, and a disclosed long-tail cost (Jensen–Shannon divergence 0.20 on infrequent drug codes) where the privacy mechanism operates as intended. Each dataset ships with a certificate stating purpose grades, utility per grade, privacy budget, a tamper-evident provenance record and printed contraindications. Limitations: results are producer-reported on a public synthetic seed; no independent evaluation on real clinical data has been published. We propose university-examined degree projects as the arm's-length evaluation this class of data requires, and invite collaboration on adequacy criteria for synthetic releases under EHDS.

---

28. Evolutionary Dynamics of Adherence-Mediating Fimbriae in Pathogenic Bacteria
bacteria, evolution, genomics, machine learning

Ebert, Gleb; von Mentzer, Astrid; Department of Microbiology and Immunology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg; SciLifeLab, University of Gothenburg; Wallenberg Centre for Molecular and Translational Medicine, University of Gothenburg

Adherence factors play key roles in determining host and tissue specificity of bacteria. In gram-negative bacteria, fimbriae, in particular those assembled by the chaperone–usher (CU) pathway, often act as adherence factors. They are hair-like extra-cellular structures made of protein building blocks and are often important virulence factors in bacterial pathogens. Operons encoding CU fimbriae are highly diverse and are often located on mobile genetic elements, enabling spread to other bacterial strains through horizontal gene transfer (HGT). Despite the importance of CU fimbriae for virulence and as key vaccine targets, our understanding of their evolution and transmission dynamics is limited as research efforts focus on individual CU operons. This knowledge gap has immediate clinical implications, as demonstrated by the recent emergence of a new variant of CU fimbriae that rapidly increased in prevalence during an outbreak of diarrhoeal disease in Bangladesh, highlighting how CU evolution directly impacts disease epidemiology. The availability of millions of bacterial genomes (> 2.6 million bacterial genomes found in GenBank) and advances in machine-learning make comprehensive discovery feasible. We will map chaperone-usher operons across gram-negative bacteria using a machine learning approach for the first time. Previously unknown CU operons will enable us to reconstruct the evolutionary history of CU fimbriae in unprecedented detail. Furthermore, focusing on CU fimbriae found in the Escherichia/Shigella genera will allow us to characterise their transmission dynamics and model their interaction with molecules found on epithelia of clinically and agriculturally relevant hosts.

---

29. SciLifeLab OMERO: A Collaborative HPC-Enabled Platform for Data-Driven Bioimaging Research
bio-image data management

Gökhan Özsari - Chalmers University of Technology, Otoniel Maya - Chalmers University of Technology , Yunqi Shao - Chalmers University of Technology, Sonja Mathias - Uppsala University

The data revolution in the life sciences increasingly relies on close integration with artificial

intelligence (AI) to transform raw data into biological insight—particularly for image data. Processing and analysis pipelines require high-performance computing (HPC) resources to handle today's large file sizes and complex datasets. At the same time, human-in-the-loop workflows continue to depend on intuitive visualization and annotation tools that can scale effectively to large datasets. To meet these competing requirements, collaborative access to image data sets located close to computing resources is needed for all actors in today's interdisciplinary research projects, from data producers to methods developers. Our future service, SciLifeLab OMERO, will provide exactly this. Built on the globally recognized, open-source data management solution OMERO1 for visualizing, managing, and sharing biological microscopy images, SciLifeLab OMERO offers active data storage connected to HPC resources for collaborative research projects using biological image data.

---

30. Charting cell differentiation in single-cell omics data via transcription-dynamics-informed optimal transport

Optimal Transport, RNA velocity, scRNA-seq data

Hahn, Hanns, KI; Tronstad, Magnus, KI; Karlsson, Johan, KTH; Dahlin, Joakim, KI

A central limitation of single cell RNA sequencing (scRNA-seq) is that an individual cell's RNA profile cannot be traced along the differentiation process. To address this, recent methods have employed Optimal Transport (OT) theory by modeling cell differentiation as mass that is transported from a set of initial states to a set of end states, under a certain objective. However, these approaches typically rely on total RNA expression and thereby neglect transcriptional dynamics captured by unspliced and spliced RNA expression. Here we integrate transcriptional dynamics, that can be obtained through RNA velocity estimation methods, into an OT-based trajectory inference method. This provides a more informative objective that accounts for both the overall data distribution and the transcriptional dynamics of single cells, with the potential to improve the reconstruction of progenitor cell trajectories.

---

31. Autonomous closed-loop cell biology with a general-purpose coding agent

Hanzhao Zhang1†, Songtao Cheng1†, Nils Mechtel1, Hugo Dettner Källander1, Wei Ouyang1. 1 Department of Applied Physics, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden. † These authors contributed equally to this work.

A research programme in cell biology runs as a cycle: perturb, image, quantify, fit a model, and let that model set the next perturbation. Each stage is automated; the junctions between them are not, least of all the one returning a fitted model to the next experiment. We report REEF, a laboratory of seven instruments and two supporting services in which a general-purpose coding agent, used without fine-tuning, reaches every instrument and analysis service through one interface, and correctness is enforced by the servers, which refuse incorrect calls before any actuator moves. Carrying one research programme through the cycle, REEF ran a dose–response search that set each round's doses from the preceding round's images and washed each well on its own arrest index rather than on a schedule; traced a tracking failure to the acquisition rather than the analysis and specified the imaging run that cut identity errors from 9.98% to 1.13%; froze and pre-registered a cell-fate model whose ranking transferred to plates prepared a week later; and generated fifteen candidate findings from its own data, each tested in sessions that had not produced it — six were

withdrawn and none survived as novel, which the analysis reported rather than presenting a confirmation as a discovery. It executed 73,777 instrument commands and 871 plate transports across four experiments, including 61.7 h unattended. When the cycle is closed in one substrate, a model the system has trained becomes, without further engineering, the rule that sets its next physical experiment.

---

32. LADDER: Context-aware gene set annotation with LLM-based hypothesis generation and literature-grounded validation
enrichment analysis, context-specific prompting, LLM-as-a-judge

1) Seby, Justin, Karolinska Institutet, 2) Siavelis, Ioannis, Karolinska Institutet, 3) Lehtiö, Janne, Karolinska Institutet

Gene set enrichment methods constrained to curated databases are reliable but cannot extend beyond existing pathways, whereas fully generative approaches using large language models (LLMs) can propose novel associations but are prone to hallucination. We present LADDER (Literature-Assisted Dual-annotation and Evidence-based Reasoning), a context-aware framework that uses structured, scaffolded prompting to generate two complementary LLM annotations for each gene set: one driven by enrichment results and one derived from the gene list alone. An independent LLM-as-a-judge then adjudicates between these candidates using PubMed literature retrieved for the same biological context. Across a panel of context-specific gene sets, LADDER outperformed state-of-the-art LLM-based annotation methods on semantic similarity and ROUGE scores, at roughly two orders of magnitude lower cost. In two omics case studies, LADDER recovered both established and candidate novel context-specific annotations, highlighting targets for experimental follow-up.

---

33. Targeting mitochondrial reprogramming as a new vulnerability in lethal prostate cancer
Prostate cancer, Biomarker discovery, Multi-omcis, Machine Learning, Mitochondrial dysfunction

Rafeletou, Alexandra, Department of Protein Science SciLifeLab KTH Stockholm Sweden, Fransson Jack, Department of Protein Science SciLifeLab KTH Stockholm Sweden, Kiselova, Tatjana, Department of Protein Science SciLifeLab KTH Stockholm Sweden, Lundberg, Arian, Department of Protein Science SciLifeLab KTH Stockholm Sweden

Background & Objective: Prostate cancer (PCa) is highly heterogeneous, and improved prognostic biomarkers, beyond current clinical risk stratification, are needed to better identify patients at risk of disease progression. Altered mitochondrial activity and metabolic reprogramming have been implicated in PCa progression and inflammatory signaling. Here, we used multi-omics data to investigate mitochondrial activity and identify molecular features associated with progression-free survival (PFS).

Methods: RNA-seq, copy number variation (CNV), and DNA methylation data from The Cancer Genome Atlas (TCGA), together with normal prostate tissue data from GTEx, were analyzed.

A transcriptomic mitochondrial activity score (BC-MitoAct) was derived from genes linked to mitochondria activity and/or their gene products localized in mitochondria, using single-sample gene set enrichment analysis (ssGSEA) and normalized against GTEx normal tissue.

Prognostic features were selected using an Elastic-Net Cox model across 200 bootstrap iterations.

We then integrated RNA-seq, CNV, and DNA methylation data using Multi-Omics Factor Analysis (MOFA) to identify latent factors associated with PFS. From the top 5% weighted features of significant MOFA factors (Factors 2 and 9), LASSO-Cox complementary pairs stability selection and multivariate survival analysis yielded three independent prognostic features. These were used to derive a risk-score model evaluated in a test cohort.

Results: At the transcriptional level, BC-MitoAct significantly stratified patients by PFS in the independent test set (p=0.0049), was associated with NCCN risk category, and remained independently associated with PFS in multivariable Cox regression (HR=2.00, p=0.0363).

The integrated multi-omics model also significantly stratified patients by PFS (p<0.0001) and remained an independent predictor after adjustment for Clinical Tumor stage, Gleason score, PSA, and age (HR=1.48, p=0.0382). Gene Set Enrichment Analysis (GSEA) showed enrichment of pro-inflammatory Hallmark pathways in patients with higher predicted risk.

Conclusions: Mitochondrial activity is associated with aggressive disease biology in PCa. Both BC-MitoAct and the integrated multi-omics signature provided prognostic information beyond established clinical risk factors. The association with pro-inflammatory transcriptional programs further links mitochondrial activity to the inflammatory tumor microenvironment and supports its potential relevance for molecular risk stratification in PCa.

---

34. Bridging Genes, Drugs, and Behavior: Multiscale Effects of Psilocybin in C57BL/6 Mice
pharmacology, neuroscience, transcriptomics, psilocybin, behavior

Jakub Mlost, Adam Wojtas, Shaima Islam, Yamen Levain, Zhaokun Wang, Iskra Pollak Dorocic, Science for Life Laboratory, Department of Biochemistry and Biophysics, Stockholm University, Sweden

Psilocybin, a serotonergic psychedelic, induces profound behavioral and molecular changes, yet the direct relationship between altered behavior and underlying transcriptional programs remains poorly understood. Here, we bridge this gap by integrating high-resolution behavioral phenotyping with spatial transcriptomics in the same animals to map psilocybin-induced behavioral states onto their corresponding molecular signatures across key brain regions. Male and female C57BL/6 mice received 0.125, 1, or 3 mg/kg psilocybin i.p. and were subjected to 60-minute Open Field Tests (OFT), followed by brain collection for Visium 10x Spatial Transcriptomics. Leveraging pose estimation (DeepLabCut), classical OFT analysis, and unsupervised behavioral classification (B-SOiD), we uncovered dose-dependent and temporally dynamic effects: 1 mg/kg induced early hyperlocomotion (0-10 min), while 3 mg/kg produced delayed hypolocomotion (20-40 min) accompanied by anxiogenic-like preference towards the periphery of the open field box. B-SOiD identified over 10 distinct behavioral motifs. Motifs including sniffing and rearing were selectively modulated by psilocybin in a dose-dependent manner. Spatial transcriptomics resolved molecular clusters corresponding to cortical layers, hippocampus, amygdala, thalamus, hypothalamus, lateral habenula, and striatum, representing regions critically involved in affective, cognitive, and innate behaviors. Critically, we identified cluster-specific psilocybin-

induced differentially expressed genes (DEGs). The number of DEGs within anatomical structures was positively correlated with the expression of psilocybin's main molecular target, Htr2a. Partial Least Squares analysis uncovered the underlying transcriptomic signature that significantly influenced behavioral patterns, revealing potential molecular substrates underlying psychedelic-induced behavioral states. This integrative approach provides a framework for mechanistically linking the behavioral and transcriptional dimensions of psychoactive compounds.

---

35. FIREFLY: Framework for Interactive Rendering, Exploration and Feature Extraction of Light-Sheet Microscopy

Light-sheet microscopy; Scientific visualization; Self-supervised super-resolution; 3D image reconstruction; Quantitative image analysis

Tember, John, University of Gothenburg; Giovanni, Volpe, University of Gothenburg, Andrei, Chagin, University of Gothenburg

Recent advances in tissue clearing and light-sheet microscopy enable three-dimensional imaging of large biological samples, revealing complex spatial relationships between cells, vasculature and surrounding tissue. However, the resulting datasets present substantial computational challenges for visualization, reconstruction and quantitative analysis.

We present FIREFLY, a Unity-based computational framework that adapts real-time graphics and GPU-accelerated processing to large volumetric microscopy datasets. FIREFLY combines interactive 3D rendering and surface modelling with computational tools for multi-scan alignment and fusion, and machine-learning-based super-resolution.

A central development is a self-supervised diffusion-based super-resolution pipeline for anisotropic light-sheet microscopy volumes. The approach learns directly from acquired data without requiring separate high-resolution training datasets and preserves experimentally measured image planes while reconstructing information between them. Preliminary results using bone marrow data acquired with 10 μm through-plane spacing demonstrate improved structural reconstruction and approximately 40–60% lower reconstruction error compared with standard interpolation.

In parallel, FIREFLY enables complementary scans acquired from different directions to be aligned and merged into more complete volumetric representations. Ongoing work focuses on validating the biological fidelity of reconstructed data and developing graph-based analysis of bone marrow vasculature to characterize connectivity, branching and spatial relationships.

Together, these developments aim to establish an integrated computational workflow for exploring, enhancing and extracting quantitative biological information from large-scale light-sheet microscopy datasets.

---

36. Cognitive relevance of gradient-aligned connectivity changes in aging and Alzheimer's disease

Alzheimer's disease; aging; fMRI

Rittmo Jonathan Department of Clinical Sciences Malmö, Faculty of Medicine, SciLifeLab,

Lund University, Lund, Sweden Franzmeier Nicolai Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, 81377 Munich, Germany; Munich Cluster for Systems Neurology (SyNergy), 81377 Munich, Germany; Department of Psychiatry and Neurochemistry, Institute of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, 405 30 Mölndal and Gothenburg, Sweden Strandberg Olof Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden Chauveau Léa Department of Clinical Sciences Malmö, Faculty of Medicine, SciLifeLab, Lund University, Lund, Sweden Satterthwaite Theodore D. Penn Lifespan Informatics and Neuroimaging Center (PennLINC), Philadelphia, PA, United States; Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, United States; Penn-Children's Hospital of Philadelphia Lifespan Brain Institute, University of Pennsylvania, Philadelphia, PA 19104; Data Science for Integrated Diagnostics (AI2D), Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA Wisse Laura EM Department of Diagnostic Radiology, Department of Clinical Sciences Lund, Lund University, Lund, Sweden Spotorno Nicola Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden Behjat Harry H Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden Dehsarvi Amir Institute for Stroke and Dementia Research (ISD), University Hospital, LMU Munich, 81377 Munich, Germany van Westen Danielle Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden Anijärv Toomas Erik Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden Palmqvist Sebastian Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden; Memory Clinic, Skåne University Hospital, SE-205 02 Malmö, Sweden Janelidze Shorena Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden Stomrud Erik Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden; Memory Clinic, Skåne University Hospital, SE-205 02 Malmö, Sweden Ossenkoppele Rik Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden; Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, the Netherlands; Amsterdam Neuroscience, Neurodegeneration, Amsterdam, the Netherlands Mattsson-Carlgren Niklas Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden; Memory Clinic, Skåne University Hospital, SE-205 02 Malmö, Sweden Hansson Oskar Clinical Memory Research Unit, Department of Clinical Sciences Malmö, Faculty of Medicine, Lund University, Lund, Sweden Vogel Jacob W Department of Clinical Sciences Malmö, Faculty of Medicine, SciLifeLab, Lund University, Lund, Sweden

Aging and Alzheimer's disease (AD) both alter resting-state functional connectivity (FC), yet region-based studies report mixed hyper- and hypoconnectivity patterns without a unified model. This may partly be explained by non-linear FC trajectories observed across the AD spectrum, suggesting stage-dependent shifts in network demands. Furthermore, because aging and AD disproportionately affect executive and memory functioning, respectively, differences in cognitive demands may contribute to differential effects on FC. Recent work suggests that these heterogeneous findings can be understood within the framework of functional axes or 'gradients' of brain organization. AD pathology and aging show dissociable FC reconfigurations aligned with two major axes: a sensory–association (SA) axis for AD and a representational–executive (RE) axis for normal aging. What drives this systematic alignment, and whether and how it relates to cognitive change in both conditions, are unanswered questions. Here, we illustrate robust effect patterns of aging

and AD that align with these functional axes and aim to elucidate underlying factors that may be associated with these alignments. We jointly model age, AD pathology and other age related factors; evaluate whether memory and executive performance map differentially onto the gradients; and use parcel-wise mediation to determine whether gradient-aligned FC reconfiguration mediates the effects of age and AD on memory vs. executive function.

---

37. Rapid and flexible gene cluster family delineation with IGUA
biosynthetic gene clusters; secondary metabolites; method development

Larralde, Martin, EMBL Heidelberg; Blom, Josefin, Umeå University; Gourlé, Hadrien, Umeå University; Carroll, Laura, Umeå University; Zeller, Georg, EMBL Heidelberg

Biosynthetic gene clusters (BGCs), the genes necessary for the synthesis of prokaryotic secondary metabolites, are a valuable resource for natural product-based drug discovery. A crucial step in a BGC research pipeline is the clustering of BGCs into gene cluster families (GCFs), as this can be used to reduce dataset redundancy and for inferring the function of unknown BGCs. However, available software for GCF delineation suffers from a lack of scalability, which restricts usage on large-scale BGC datasets. Here, we introduce IGUA (Iterative Gene clUster Analysis; https://github.com/zellerlab/IGUA), a novel tool for GCF delineation. IGUA executes three major steps to rapidly generate GCFs: fragment removal, nucleotide-level deduplication and protein-level clustering. When tested on a dataset of manually curated BGC-to-GCF mappings, IGUA performed GCF delineation with comparable accuracy to gold standard software. Moreover, it outperformed the state-of-the-art tools in terms of memory requirements and computational time, by a factor of 5 and 10 respectively, when clustering datasets of >13 500 BGCs on 28 CPUs. We thereafter demonstrated the scalability of IGUA by clustering a dataset of 2 829 071 BGCs to generate 56 960 GCFs, representing the largest clustered BGC dataset to date. In contrast to previous tools, IGUA does not rely on Pfam domains for its clustering process and can therefore be used to delineate gene clusters beyond BGCs. Consequently, we used IGUA to cluster a dataset of >10 000 secretion systems, showcasing the versatility of the computational tool. Thus, IGUA is a novel, flexible and scalable method for GCF delineation, which, together with rapid in silico-based tools for BGC prediction, truly enables large-scale exploration of microbial biosynthetic potential.

---

38. Scalable And Reference-Free Single-Cell Metagenome Analysis with Bascet and Zorn
Single-cell Metagenomics; Bioinformatics

Gourlé, Hadrien (Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden; Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden; Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden); Yakovenko, Iryna (Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden; Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden; Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden; Department of Molecular Biology, SciLifeLab, Umeå University, Umeå, Sweden); Verma, Jyoti (Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden; Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden; Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden; Department of Molecular Biology, SciLifeLab, Umeå University, Umeå, Sweden; Integrated Science Lab (IceLab), Umeå University, Umeå, Sweden; Department of Odontology, Umeå University, Umeå, Sweden);

Dicken, Julian (Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden; Department of Molecular Biology, SciLifeLab, Umeå University, Umeå, Sweden); Albrecht, Florian (Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden; Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden; Department of Molecular Biology, SciLifeLab, Umeå University, Umeå, Sweden; Integrated Science Lab (IceLab), Umeå University, Umeå, Sweden); Stenlund, Arthur (Department of Molecular Biology, SciLifeLab, Umeå University, Umeå, Sweden); Sharaf, Chetan (Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden; Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden; Department of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; The Wallenberg Laboratory, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden); Dwibedi, Chinmay (Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden; Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden; Department of Neuroscience and Physiology, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; The Wallenberg Laboratory, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden); Mažutis, Linas (Department of Molecular Biology, SciLifeLab, Umeå University, Umeå, Sweden; Institute of Biotechnology, Life Sciences Centre, Vilnius University, Vilnius, Lithuania); Normark, Johan (Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden; Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden; Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden); Sheng, Nongfei (Department of Odontology, Umeå University, Umeå, Sweden); Strömberg, Nicklas (Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden; Department of Odontology, Umeå University, Umeå, Sweden); Löfstedt, Tommy (Department of Computing Science, Umeå University, Umeå, Sweden); Carroll, Laura M. (Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden; Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Umeå, Sweden; Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden; Integrated Science Lab (IceLab), Umeå University, Umeå, Sweden); Henriksson, Johan (Umeå Centre for Microbial Research (UCMR), Umeå University, Umeå, Sweden; Department of Molecular Biology, SciLifeLab, Umeå University, Umeå, Sweden; Integrated Science Lab (IceLab), Umeå University, Umeå, Sweden).

Single-cell whole-genome sequencing of microbial communities is an emerging method that enables de novo genomic characterization of individual cells without prior knowledge of community composition, but no tools exist for analysis at the scale required for community-representative characterisation. We introduce Bascet (Rust command-line suite) and Zorn (R package; http://zorn.henlab.org/), enabling de novo single-cell microbial genomics via indexed per-cell file formats, countsketch projection of k-mer spectra for sparse genome comparison, and HPC-compatible horizontal partitioning designed to scale to millions of cells. Evaluated on >50,000 bacterial cells from a ten-species mock community, Bascet's countsketch-based clustering achieves higher species-level accuracy than KRAKEN-based clustering (99.2% vs 93.8%), while requiring no reference database. Bascet is 8× faster than competing pipelines, and >200× faster than naïve KRAKEN-based annotation. Applied to human saliva, Bascet and Zorn recover single-amplified genomes from thousands of prokaryotic cells, demonstrating utility on a complex, real-world microbiome.

39. Cyclodextrins as a tool to study lipid number asymmetry in model membranes
membrane biophysics, lipid bilayers, membrane asymmetry, liposomes, molecular dynamics simulations

Wagner, Julian P., Department of Biochemistry and Biophysics, Science for Life Laboratory (SciLifeLab), Stockholm University, Stockholm, Sweden; Bergner, Leon, Department of Biochemistry and Biophysics, Science for Life Laboratory (SciLifeLab), Stockholm University, Stockholm, Sweden; Reagle, Tyler, Department of Molecular Physiology and Biological Physics, University of Virginia, VA, USA; Doktorova, Milka, Department of Biochemistry and Biophysics, Science for Life Laboratory (SciLifeLab), Stockholm University, Stockholm, Sweden

Lipid bilayers act as functional barriers between subcellular compartments and between the cell's cytoplasm and the extracellular environment. A fundamental characteristic of the mammalian plasma membrane is its leaflet asymmetry. On the one hand, this asymmetry refers to differences in lipid composition between the two leaflets. On the other hand, emerging evidence has revealed a new dimension of asymmetry, namely phospholipid abundance (or number asymmetry). Both of these asymmetries can act as critical modulators of the membrane's biophysical properties such as lipid packing, fluidity, and tension, which play important roles in various physiological processes. However, how number asymmetry mechanistically influences bilayer properties remains poorly understood. Here, we use model membranes (liposomes) and cyclodextrins, which can selectively remove phospholipids from the outer leaflet of the lipid bilayer, to generate and study the biophysical effects of lipid number imbalances. To this end, we first identify a regime where lipids are extracted but the vesicles remain relatively stable through liposome solubilization experiments. We show that this regime varies depending on the type of lipid as well as the curvature of the liposomes. Using environment-sensitive fluorescent probes, we find that increasing number asymmetry is accompanied by specific changes in the lipid packing of the outer leaflet. As a complementary approach, we characterize the stability of number-asymmetric bilayers using molecular dynamics simulations. These studies show that at large imbalances, the deformation of the bilayer is determined largely by the packing density of the overpopulated leaflet, which becomes more compressed with increasing number imbalance. Applying net bilayer tension suppresses this effect and restores a flat morphology by shifting leaflet packing away from a critical threshold. Prompted by these in silico findings, we apply net tension to the liposomes via fixed osmotic gradients and examine their resulting tolerance to lipid number imbalance. Our results demonstrate that positive and negative net tensions have opposite effects on the onset and breakpoint of vesicle solubilization. Finally, using a theoretical model backed by the simulations, we explain the experimental trends by stretching-dominated bilayer instability at the onset of solubilization and compression-dominated collapse at the breakpoint. In summary, our integrative approach using experiments, molecular dynamics simulations and theory reveals mechanistic insights into how number asymmetry shapes bilayer properties.

---

40. MembraneFinder: Annotation-free detection and quantitative analysis of lipid membranes in cryo-EM images
Cryo-EM, MD simulations, Machine learning

Sharma, Karan, Stockholm University, Sweden; Laabadi, Aymane, EMINES, Morocco; Cowgill, John, Stockholm University, Sweden; Howard, Rebecca, Stockholm University, Sweden; Heberle, Frederick, Stockholm University, Sweden; Lindahl, Erik, Stockholm University, Sweden; Doktorova, Milka, Stockholm University, Sweden

Automated and accurate lipid membrane detection in cryo-electron microscopy (cryo-EM) images is essential for the detailed structural characterization of both biological and model membranes. Existing deep-learning tools often fail to generalize across variable imaging conditions and require tedious parameter tuning, model retraining, and manual annotation. To address these limitations, we present MembraneFinder, a comprehensive framework for lipid membrane detection in single-projection cryo-EM images, trained exclusively on physics-informed synthetic data. The complete framework is split across three integrated pipelines. First, a synthetic data generator utilizes molecular dynamics simulations to produce large-scale batches of synthetic cryo-EM images spanning variable imaging conditions, noise distributions, vesicle sizes, and morphologies. The generated images also include multi-lamellar vesicles with realistic spacing and angular deformations in the overall shape of the membrane, as well as simulated contaminants and background structures. Crucially, the synthetic data allows for generation of precise ground-truth masks for the membranes in the images, completely eliminating the need for manual annotation. Second, a model training and inference pipeline uses these image-mask pairs to train a U-Net++ model for binary membrane segmentation. We validated the trained model on publicly available datasets from EMPIAR as well as diverse datasets collected in-house, where it achieved higher membrane detection rates than widely used general-purpose cryo-EM software packages. Finally, the analysis pipeline extracts quantitative structural insights from the segmented outputs, enabling measurement not only of readily visible features like membrane shape and curvature but also local bilayer properties like thickness, which can vary spatially as a result of interactions with proteins and other particles of interest. Publicly available via a web interface, MembraneFinder thus provides an efficient and versatile solution for membrane analysis in 2D cryo-EM images.

---

41. Spatial Omics
Spatial biology; spatial transcriptomics; spatial proteomics; spatial metabolomics

Anja Mezger, Anna Nilsson, Charlotte Stadler, Katarina Tiklova, Chika Yokota, SciLifeLab.

The Spatial Omics Platform offers researchers cutting-edge technologies to map the spatial organization of cells and molecules within complex tissues. By integrating profiling of transcripts, proteins, lipids, glycans, and metabolites, the platform enables comprehensive multi-modal studies. It supports projects that combine multiple spatial omics methods, providing coordinated access, expert guidance, and state-of-the-art tools to advance research in tissue biology and disease.

---

42. Plasmid-mediated bacteriocin dynamics during early-life gut microbiome assembly
Infant gut microbiome; Plasmidome; Microbiome assembly; Bacteriocins

Drane, Kezia, Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden; West, Christina E., Department of Clinical Sciences, Pediatrics, Umeå University, Umeå, Sweden; Dwibedi, Chinmay, Department of Clinical Microbiology, SciLifeLab, Umeå University, Umeå, Sweden, and Laboratory for Molecular Infection Medicine Sweden (MIMS), Umeå University, Sweden.

Plasmids enable the horizontal transfer of traits that may influence bacterial adaptation and microbial interactions. However, resolving relationships between plasmids, their bacterial hosts, and associated functions remains a major computational challenge in metagenomic

datasets. This project proposes a framework to investigate plasmid dynamics in the developing human gut microbiome, using plasmid-encoded bacteriocins as a model system for studying the potential contribution of mobile traits to microbial competition.
The project will leverage longitudinal metagenomic datasets from the NorthPop cohort, comprising repeated gut microbiome samples collected across early childhood. The core objective is to develop analytical workflows to infer relationships between plasmids, their bacterial hosts, and associated traits, as well as to characterise their persistence and turnover over time.

The analytical framework will integrate metagenomic assembly, graph-based and compositional inference, longitudinal co-abundance analysis, and statistical and machine learning approaches. These analyses will examine whether plasmid-encoded bacteriocins are associated with bacterial persistence, turnover, coexistence, and community stability during early-life microbiome assembly. Complementary evidence will support host inference, with uncertainty in plasmid reconstruction and host assignment explicitly considered.

The project aims to generate reproducible computational workflows, a catalogue of putative plasmid-encoded bacteriocins with inferred host associations, and models linking mobile traits to longitudinal microbiome dynamics. Together, these outputs will provide a foundation for investigating the ecological roles of plasmid-associated functions in the developing infant gut.

---

43. Pan-Neurological Disease Profiling of the Human Circulating Proteome
Circulating proteome, Neurological disease, Biomarkers, Proteomic signatures, Disease classification

Antonopoulos Konstantinos, KTH Royal Institute of Technology, Sweden; Lilja Sofia, KTH Royal Institute of Technology, Sweden; Sankaran Sathyanarajanan, KTH Royal Institute of Technology, Sweden; Aksoylu Inci, KTH Royal Institute of Technology, Sweden; Bueno Alvéz María, KTH Royal Institute of Technology, Sweden; Johnson Sterling, University of Wisconsin- Madison, USA; Zetterberg Henrik, Göteborgs Universitet, Sweden; Graff Caroline, Karolinska Institutet, Sweden; Svenningsson Per, Karolinska Institutet, Sweden; Ingre Caroline, Karolinska Institutet, Sweden; Uhlén Mathias, KTH Royal Institute of Technology, Sweden; Sjöstedt Evelina, KTH Royal Institute of Technology, Sweden; Nilsson Peter, KTH Royal Institute of Technology, Sweden; Edfors Fredrik, KTH Royal Institute of Technology, Sweden; Månberg Anna, KTH Royal Institute of Technology, Sweden.

Neurological disorders exhibit substantial clinical and biological heterogeneity, motivating the search for molecular signatures that capture disease-associated changes across diagnostic groups. We profiled the circulating proteome of 896 patients with Alzheimer's disease (AD), Parkinson's disease (PD), frontotemporal dementia (FTD), or amyotrophic lateral sclerosis (ALS), alongside 548 healthy controls, using the Olink HT platform. Differential abundance analyses and classification models were used to characterize disease-associated protein patterns and assess their potential for distinguishing neurological conditions from controls. Across diseases, a relatively small subset of proteins provided strong classification performance, suggesting that compact plasma-protein signatures may capture aspects of disease biology. We further examined the influence of age and sex on disease-associated proteomic patterns. Findings from PD and ALS, together with selected controls, were independently evaluated using the SomaScan 11k platform. Together, these analyses provide a cross-disease view of circulating proteomic variation in

neurological disease and highlight protein signatures for further investigation.

---

44. DeepRepurpose: A Spatial Biology-Informed and Deep Learning-Driven Drug Repurposing Framework

Deep Learning, Drug Repurposing, Spatial Transcriptomics

Kalaitzidis Konstantinos (Merck & DBB SU), Wang Zhaokun (DBB SU), Guala Dimitri (Merck & DBB SU), and Sonnhammer Erik (DBB SU).

Background. Developing a new drug costs approximately $900 million in direct R&D. Repurposing previously approved drugs is an attractive, data-driven strategy to reduce development timelines. Current state-of-the-art perturbation response models, such as PDGrapher, learn drug effects on gene regulatory networks but are restricted to non-spatial data, discarding the tissue context in which disease arises. With DeepRepurpose, we aim to extend perturbation modelling to spatial transcriptomics (ST) in support of drug repurposing.

Methods. We trained a variational autoencoder (VAE) on the Sci-Plex3 dataset (188 compounds, four doses, three cancer cell lines) to predict treated expression from an unperturbed cell, a drug, and a dose. After mitochondrial and gene-count quality control and selection of highly variable genes, the encoder maps each cell to a drug-independent latent cell state. Our FiLM-based conditioning module combines a drug embedding with a saturating dose function to modulate the decoder. We tested our model's generalisation by holding out an entire cell line and comparing predictions with real treated cells against a VAE with conditioning disabled as well as a random baseline. Candidate drugs are ranked by how far their predicted effect moves a diseased profile toward the healthy state.

Results. Our preliminary results show that the conditioning module learned drug and dose-dependent structure, as low doses cluster near the origin, while high doses spread along drug-specific directions. On the held-out cell line, the model achieved moderate per-cell Pearson correlation and R^2, well above the random baseline. However, the vanilla VAE, which never sees drug identity, matched it on R^2 and exceeded it on Pearson. This indicates that a shared, drug-agnostic treatment response may dominate these metrics.

Conclusion. Conditioning captures drug and dose structure, but control-referenced metrics cannot yet demonstrate drug-specific learning. Next steps are evaluation that isolates drug-specific effects, a transfer-learning bridge onto liver cancer ST, and in silico screening of existing compounds for reversal of spatially resolved disease signatures.

---

45. Evolutive history of the Striped-backed antbird (Thamnophilidae: Myrmorchilus strigilatus)

Neotropical Seasonally Dry Forests; Myrmorchilus; Caatinga; Chaco; Thamnophilidae

Viturino, Lays; Mariz, Daniele; Lima, Hevana; Couto, Giovanna; Naka, Luciano

The role of ecological factors in the origin and diversification of the tropical biota represents one of the most complex subjects in evolutive biology, whereas much debate has surrounded the Neotropical Seasonally Dry Forests. Here, we investigate the evolutionary and biogeographical relationship of the two largest patches of these forests:

The Caatinga and the Chaco. We use ultraconserved elements (UCEs) and mitogenome from 35 individuals to bring light into the evolutive history of the Striped-back antbird (Myrmorchilus strigilatus), a species with two recognized subspecies almost perfectly restricted to these two dry forests. Bayesian structuring and phylogenomic analyses shows a clear separation between the populations restricted to the Caatinga and those to the Chaco, which may have occurred 0.84 (0.74 – 0.96) Mya. Within the caatinga, we also find two genetically structured groups of individuals, which seem to be separated by the São Franciso River. These results suggest a recent connection between the Caatinga and the Chaco, while also highlighting a complex evolutionary history in the heterogeneous Caatinga. Additionally, more studies should be done focusing on the demographic history and on cryptic lineages within the Caatinga.

---

46. Mechanisms of membrane deformation by the Caveolin-1 complex
molecular dynamics simulations, coarse-grained

(Bergner, Leon, Stockholm University/SciLifeLab), (Connolly, Sarah M., University of Michigan), (Ke, Xiao, University of California), (Sharma, Karan, Stockholm University/SciLifeLab), (Aland, Sebastian, TU Bergakademie Freiberg), (Kenworthy, Anne M., University of Virginia), (Mosalaganti, Shyamal, University of Michigan), (Ohi, Melanie D., University of Michigan), (Rangamani, Padmini, University of California San Diego), (Doktorova, Milka, Stockholm University/SciLifeLab)

Caveolin is an integral monotopic membrane protein associated with curved regions of cell membranes. In mammalian cells caveolins oligomerize into large complexes which together with another group of proteins called cavins form small invaginations of the membrane, so called caveolae. However, experimental studies using model membranes have demonstrated the ability of the human Caveolin-1 complex (7sc0) to induce curved regions of the membrane without cavins. This membrane shaping ability is lacking in caveolin from the species S. purpuratus (9dn0) and S. rosetta (9dn1). Motivated by these experimental results, we used computational methods to examine the membrane shaping ability of the human Caveolin-1 complex by comparing lipid-protein interactions between these three species. The dynamics of the caveolin complexes embedded in simple bilayers were studied using coarse-grained molecular dynamics simulations. Similar computational studies have indicated a correlation between curved regions and the splaying of the lipid chains in the distal leaflet right across from the complex. The results from our simulations agree with the experimental observations in that 7sc0 does associate with regions of high curvature while 9dn0 and 9dn1 do not, but we did not see the expected dependence on lipid splay in the distal leaflet. Instead, analysis of the lipid-protein interactions in the proximal protein-containing leaflet indicated differences between the three caveolin variants. For 7sc0, the leaflet thins and the proximal lipids tilt with their tails below the complex bottom surface close to the boundary. This is not observed for the other two variants. We further developed a model based on the protein shape and lipid-protein contact angle to explain the different complex-dependent shapes of the bilayer deformation. Finally, detailed hydrophobicity analysis of the structure of each complex showed that both 9dn0 and 9dn1 have a clear boundary between hydrophobic and hydrophilic regions preventing lipids from sliding down the rim of the disk-shaped complex. This boundary is not present for 7sc0, allowing the proximal lipids to instead change their relative positioning and more easily accommodate bilayer curvature. These findings help further our understanding of the membrane shaping ability of the human Caveolin-1 complex.

47. Accelerate your infectious disease research
Pathogens, infectious disease, RDM, pandemic preparedness

Liane Hughes1,2, Nalina Hamsaiyni Venkatesh3,4, Senthilkumar Panneerselvam1,2, Abdullah Aziz3,4, Paul Dulaud3,4, Kazi Jahurul Islam1,2, Rickard Hammarén1,2, Hanna Kultima1,2, & Johan Rung1,2 (1. Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden, 2. Science for Life Laboratory (SciLifeLab), Uppsala University, Uppsala, Sweden, 3.High Performance Computing Center North, Umeå University, Umeå, Sweden, 4.Science for Life Laboratory (SciLifeLab), Umeå University, Umeå Sweden).

The Swedish Pathogens Portal works closely with the infectious disease research community to accelerate research. The portal is continuously updated and improved in accordance with the needs of the community. Features typically enable the community to find relevant resources more quickly, raise the profile of Swedish research efforts, and ensure that data is as FAIR (Findable, Interoperable, Accessible, and Reusable) and open as possible.

The latest version of the portal was released in October, and represents the biggest change to the site to date. The changes reflect some of the feedback from the community, and they include:

• An updated interface to improve user experience.
• More streamlined advice related to FAIR data management.
• The ability to dynamically gather and collate data on a greater diversity of pathogens and topics.
• More powerful data dashboards, that enable more flexibility for how data can be displayed and interacted with.
• The ability for research groups to update/add data directly within the site.

More updates are planned in the coming months, including new dashboards, further improvements to existing elements, and new sections. Feedback from the community is always warmly welcomed!

Hundreds of research groups have already benefited from working with the portal, by getting support with data sharing, and/or having their work directly featured on the site. Maybe your group could be next?

---

48. Time-Resolved, Multi-Channel Live-Cell Imaging and AI-Driven Phenotypic Profiling of Drug Responses in Ovarian Cancer
1) Live-cell imaging 2) Phenotypic profiling 3) Ovarian cancer 4) Drug screening 5) AI/ML image analysis

Hansen, Mads Ringholm, Karolinska Institutet, Sweden; Ariyaberg, Taka, Uppsala University, Sweden; Dunlop, Nils, Chalmers University of Technology, Sweden; Ghanbari, Elham, Chalmers University of Technology, Sweden; Babaeivavdare, Mehdi, Chalmers University of Technology, Sweden; Cropsal, Télio, Chalmers University of Technology, Sweden; Karakulev, Aleksandr, Uppsala University, Sweden; Horváth, Csongor, Uppsala University, Sweden; Mapitigama, Dhanushki, Uppsala University, Sweden; Panahi, Ashkan, Chalmers University of Technology, Sweden; Singh, Prashant, Uppsala University, Sweden; Mercado Oropeza, Rocio, Chalmers

University of Technology, Sweden; Spjuth, Ola, Uppsala University, Sweden; Seashore-Ludlow, Brinton, Karolinska Institutet, Sweden.
Preclinical drug discovery platforms often fail to capture the complex and evolving treatment responses observed in solid tumors, contributing to high failure rates among oncology drug candidates. Conventional screening approaches rely on static, endpoint-based measurements that provide only a snapshot of cellular behavior and fail to capture the dynamic adaptations that drive therapeutic resistance and disease progression. This limitation is particularly relevant in ovarian cancer, where pronounced heterogeneity and rapid emergence of resistance complicate treatment prediction. Within the TIMED (Time-resolved Imaging and Multi-channel Evaluation of Cellular Dynamics) consortium, we aim to develop a time-resolved, AI-enabled phenotypic profiling framework that combines high-throughput multiplexed live-cell imaging with AI/ML-based image analysis to quantify dynamic cellular responses to drug perturbation. As the first component of this framework, we aim to establish a washout-based morphological profiling assay capable of capturing cellular state changes following drug withdrawal. The assay uses a high-throughput imaging platform to enable frequent, high-density temporal imaging, with cellular phenotypes profiled through multiplexed live-cell staining of distinct subcellular compartments over periods of up to 72 hours. The workflow is being optimized in ovarian cancer cell lines, with planned extension to primary patient samples.

We established a multiplexed live-cell staining panel that maintains growth rates comparable to unstained, DMSO-treated controls, indicating minimal physiological perturbation while supporting longitudinal imaging for up to 72 hours. This optimized panel provides a foundation for time-resolved morphological profiling of drug responses and enables characterization of phenotypic changes beyond conventional static endpoints. Next, an annotated compound library will be screened to determine whether the assay captures expected drug-induced phenotypic differences and to further establish its potential for time-resolved characterization of cancer cell responses.

---

49. Clinical Genomics

genomics, clinical, cancer, rare diseases, microbiology

Dávila Marcela, Clinical Genomics

The Clinical Genomics (CG) platform was established in 2014 to meet the increasing demands by the translational research community to leverage genomic technologies towards improved diagnostics and treatment. Today, the platform has full national outreach with nodes at all seven Swedish medical faculties, working in close collaboration with the respective university hospital. Given that Swedish healthcare is organized into autonomous regions, the national coverage is critical for ensuring equal access to services provided by the platform and accelerating the implementation of new diagnostics in healthcare throughout the country. Our MISSION includes three key objectives: Provide world-class support for translational research and clinical studies, adapt and optimize new genomic technologies for tomorrow's healthcare and contribute to a direct patient and societal impact by promoting precision medicine.

---

50. Shiga toxin acquisition in the extraintestinal escherichia coli ST117 lineage through putative prophage transfer

Extraintestinal pathogenic Escherichia coli, Phylogenetics, Ancestral state reconstruction, Prophages

Marina Masliakova, Astrid von Mentzer; Department of Microbiology and Immunology, Sahlgrenska Academy, University of Gothenburg; SciLifeLab; The Wallenberg Centre for Molecular and Translational Medicine; Centre for Antibiotic Resistance Research (CARe)

Escherichia coli is a genetically versatile species in which horizontal gene transfer continually reshuffles virulence determinants, giving rise to novel pathogenic lineages. ST117 is a globally distributed extraintestinal pathogenic E. coli (ExPEC) lineage recovered from both poultry and humans, with genomic evidence pointing to a hypothesized zoonotic, poultry-to-human direction of spread and a frequent burden of multidrug resistance, including to critically important antimicrobials. Shiga toxin (stx), the defining marker of Shiga toxin–producing E. coli (STEC), has not previously been reported in ST117. Its acquisition could mark a shift toward intestinal pathogenesis and a further dimension of human adaptation. A dataset of >3,000 ST117 genomes was assembled from the AllTheBacteria collection, of which >100 carry gene(s) encoding stx and are predominantly human-derived, contrasting with the largely poultry-associated background population. Phylogenetic reconstruction, together with antimicrobial-resistance and virulence profiling, resolves the stx-positive isolates into four distinct clades. Based on a whole-genome SNP alignment of a representative subset, Bayesian molecular dating was applied to infer when, and in which host, these lineages arose and to identify the most recent common ancestor of each stx-positive clade. Prophages across the ST117 population are now being annotated and characterized to localize stx and, through comparative analysis against candidate donor lineages using complete assemblies, to test whether stx was acquired via prophage from another E. coli lineage. By reconstructing the timing, host context and mechanism of stx acquisition, this work clarifies how an extraintestinal lineage gains intestinal virulence traits and whether this reflects ongoing adaptation to humans, informing surveillance of emerging hybrid pathogens at the animal–human interface.

---

51. Global Population Structure of Enterotoxigenic Escherichia coli from a One Health Perspective

enterotoxigenic Escherichia coli, One Health, phylogenetics

Grimmer, Marlene, University of Gothenburg; Astrid von Mentzer, University of Gothenburg, SciLifeLab

Enterotoxigenic Escherichia coli (ETEC) is a major bacterial pathogen, causing severe diarrheal disease in a variety of hosts. Infection and pathogenesis depend on several virulence factors such as enterotoxins and colonization factors (CFs). While some of them are expressed in diverse genomic backgrounds, stable phylogenetic lineages have been described that carry major human virulence factor profiles. However, existing surveillance efforts and research have focused mostly on ETEC from single hosts. We compiled 12,900 publicly available ETEC genomes and associated metadata from human, animal, and environmental sources. Phylogenetic analyses of the global population structure show that most of the previously described lineages are conserved with stable and host-specific virulence factor profiles. However, the presence of smaller lineages, containing isolates from humans, animals, and environmental sources, with mixed virulence factor profiles, suggest closer similarity between different host isolates than previously thought. Additionally, antimicrobial resistance (AMR) genes show an overall high resistance burden and accumulation of resistance determinants against multiple antimicrobial classes in some lineages. This study provides a global ETEC population structure from a One Health perspective

by integrating human, animal, and environmental isolates. This framework helps identify epidemiologically important lineages, host-associated traits, virulence factors, and antimicrobial resistance genes. These findings may support genomic surveillance and inform the prioritization of vaccine targets in evolving ETEC populations.

---

52. To eat, or not to eat? Host repertoire of a tropical butterfly

interactions; host repertoire; acceptance; growth rate; Lepidoptera: Nymphalidae, Nymphalini

Bovolenta Martina 1,2; Katharina Schneider 1,2; Sille Holm 1,2; André V. L. Freitas 3; Mariana P. Braga 1,2; 1 Department of Ecology, Swedish University of Agricultural Sciences, SLU, Uppsala, Sweden; 2 Science for Life (SciLifelab), Swedish University of Agricultural Sciences, Uppsala, Sweden; 3 Departamento de Biologia Animal and Museu de Zoologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Brazil

Ecological specialisation appears to vary along the latitudinal gradient, with species interactions being more specialised in tropical regions compared to temperate ones. However, the latitudinal specialisation gradient remains controversial, likely due to data bias in tropical regions.

In order to investigate host-plant specialisation in a tropical butterfly, we explored the host repertoire of Vanessa braziliensis (the Brazilian Painted Lady). We tested caterpillar acceptance, survival, and performance on plants belonging to genera and families used by closely related temperate species within the tribe Nymphalini.

Our results show that this tropical butterfly can accept a limited number of plant families, namely Asteraceae and Urticaceae. However, only a subset of these plants supported caterpillar survival, and caterpillar performance varied depending on the plant species. Overall, caterpillars showed higher survival rates and higher performance on plants belonging to the Asteraceae family. These results suggest that V. braziliensis is a specialised butterfly with higher survival rates and performance on Asteraceae than on other plant families, including the ancestral host of the tribe Nymphalini, Urticaceae.

The genetic mechanisms underlying host plant use in V. braziliensis and other tropical butterfly species remain largely unknown. Therefore, investigating the genetic basis of host use may provide important insights into the latitudinal specialisation gradient.

---

53. Hypothesis-driven spatial metabolomics: an open-source toolkit for multimodal interrogation of bioactive lipid pathways

Spatial, metabolomics, multimodal, lipids.

Smith, Matthew, 1Institute of Environmental Medicine, Karolinska Institute. Firsova, Alexandra, 2Department of Molecular Biosciences, Stockholm University. Samakovlis, Christos, Department of Molecular Biosciences, Stockholm University, Stockholm. Wheelock, Craig, 1Institute of Environmental Medicine, Karolinska Institute.

Spatial metabolomics is increasingly used to localize bioactive metabolites within tissue, yet most studies rely on broad, untargeted analyses interpreted post hoc against disease states.We argue this limits mechanistic insight: untargeted methods cannot reliably

detect and resolve isomers of potent lipid mediators (eicosanoids) at biologically relevant concentrations, and they spread measurement effort across thousands of features that are largely unrelated to the biological model.

We propose that targeted, hypothesis-driven, multimodal spatial metabolomics is a more tractable framework; focusing on fewer biological features measured across complementary methods spanning the central dogma of biology. In this framework each feature is chosen based on biological relevance and analytical reasoning determined by the preceding modality (e.g., bulk metabolomics to select targets for spatial metabolomics; spatial metabolomics to select cell type or enzyme panel). This results in fewer measurements but greater biological context per measurement which is directly interpretable against the study design. Additionally, it offers generally increased confidence in metabolite identifications based on biological trackability (e.g., enzyme with downstream metabolite) and orthogonal analytics (LC separation of isomers in the same model as for spatial analyses).

To support this workflow, we have developed four open-source tools.
MRManalyzeR processes targeted LC-MS/MS data (S/N and blank filtering, batch correction, PCA / volcano / heatmap HTML reports), and is used to assess MRM transition fidelity, resolve isomers, and inform target selection for spatial assays.
quantMSImageR processes DESI-MRM images, including hot/cold spot removal, normalization, calibration-based quantification, region-of-interest integration, and HTML reporting of ion images, metabolite ratios, and inter-metabolite correlations.
tissUUmaps helpers (developed with NBIS) provides interactive, browser-based visualization that integrates spatial metabolite data (from DESI-MRM) with co-registered H&E and spatial transcriptomics; the easily shared outputs enable non-specialist collaborators to explore and interpret the data directly.
corrMSI enables ROI-resolved comparisons and bivariate spatial autocorrelation (e.g., local Moran statistics) across processed spatial datasets. Specifically, this includes gaussian decay from cell types, niches or ROI for regional comparisons and enables correlation analysis between different modalities (e.g., metabolite v ROI or metabolite v cell type) to determine spatial regulation between sample groups (e.g., case/control).

We illustrate the workflow in a mouse HDM/curdlan model of allergic airway inflammation, targeted eicosanoid imaging integrated with spatial transcriptomics, DAPI, and H&E staining enabled the spatial regulation of COX-pathway biosynthesis to be investigated in relation to recruited immune populations and cellular niches and linking these molecular shifts to airway proximity.

---

54. Tracking Biodiversity Change Through Time and Space: Sweden's 60-Year Airborne eDNA Archive

eDNA, air filters, biodiversity, genomics, national resource

Hawgood Michael; Johansson Anna-Mia; Sandström Anton; Andersson Bea Angelica; Svensson Daniel; Lozano Villegas Jose Antonio; Brindefalk Björn; Cunico Marco; Jessen Henrik; Dvirnas Albertas; Sten Åslund Matilda; Brodin Tomas; Brännström Åke; Singh Navinder; Stenberg Per

Biodiversity is changing rapidly, but understanding these changes requires observations that span decades, landscapes and ecosystems. Sweden has a unique resource for doing this: a growing 60-year archive of weekly air samples collected at six monitoring stations across

the country.

The Swedish Biodiversity in Time and Space (SweBITS) project is using airborne environmental DNA (eDNA) to transform this historical archive into a long-term record of biodiversity change. DNA from plants, fungi, microorganisms and animals is captured on air filters and can be extracted and sequenced, allowing us to investigate what organisms were present in the air and how these communities have changed over time.

The archive currently contains more than 20,000 samples collected since the 1960s, with more than 40 TB of raw sequencing data already generated. Weekly sampling across all six stations continues to extend this record, providing an ever-growing resource for investigating biodiversity change across Sweden. By combining these genomic observations with climate, land-use and other environmental data, SweBITS is investigating how biodiversity has responded to environmental change across Sweden.

Our analyses demonstrate the potential of this approach to reveal both long-term and seasonal patterns in biodiversity, including changes associated with landscape transformation and climate. By connecting observations from individual air samples to patterns across decades and across the country, SweBITS provides a new perspective on biodiversity monitoring and allows us to revisit the past and watch ecosystems change in almost real time.

This poster will introduce the SweBITS project and the valuable national resource it provides. The aim is to inform the research community about this growing resource, connect with researchers interested in using the data, and explore new opportunities for collaboration across biodiversity, environmental change and ecosystem resilience.

---

55. A Multi-Omics Signature of the Pre-Metastatic Liver in Pancreatic Cancer

Pancreatic cancer; Liver metastasis; Pre-metastatic niche; Extracellular vesicles; Multi-omics

Mohamed Elhusseiny, Alexander Sandberg, Jonas Burman, Carolin Jönsson, Hakon Andersen Blomstrand, Karin Johansen, Constantinos Zambirinis, Per Sandström, Bergthor Björnsson, Linda Bojmar

The liver is the most common site of distant metastasis in pancreatic (PDAC) cancer, yet our ability to identify patients at increased risk of liver metastasis remains limited. Emerging evidence suggests that the liver undergoes molecular and cellular changes before clinically detectable metastases develop. Characterizing this pre-metastatic state may therefore reveal prognostic biomarkers and provide new opportunities for early risk stratification.

We applied a data-driven multi-omics approach to characterize the pre-metastatic liver in patients with PDAC cancer. Liver tissue collected at surgery was profiled using transcriptomics and proteomic approaches together with tissue staining to capture molecular and cellular features associated with subsequent metastatic outcome. In parallel, circulating extracellular vesicles (EVs) were analyzed to investigate whether systemic tumor-associated signals reflect alterations observed within the liver. Data across these complementary layers were integrated to identify molecular patterns associated with liver metastatic risk and to develop a candidate multi-omic signature.

Integrated analysis revealed coordinated molecular alterations within the liver associated with metastatic outcome, involving changes detectable across transcriptomic, proteomic and tissue-level measurements. Comparison with circulating EV profiles identified overlapping and complementary signals, suggesting that components of the pre-metastatic liver phenotype may be detectable in circulation. These findings provide the basis for a combined tissue- and blood-derived signature for stratifying patients according to their risk of developing liver metastasis.

Our study demonstrates the potential of integrating liver tissue multi-omics with circulating EVs to characterize the pre-metastatic environment in pancreatic cancer. Linking local changes in the future metastatic organ with minimally invasive circulating biomarkers may provide a new framework for predicting liver metastasis before it becomes clinically detectable.

---

56. strobealign long-read alignment enable fast and accurate Structural Variant Detection using Long Reads

Long read mapping, alignment, SV calling, Structural Variants

Buchin Nicolas Stockholm University, Martin Marcel NBIS, Ameur Adam Uppsala Genome Center, Sahlin Kristoffer Stockholm University

We present a fast long-read mapper, an extension of strobealign, that enables accurate classification of structural variants (SVs). Telling a tandem duplication from an insertion matters for gene dosage and copy number in cancer, yet current long-read aligners often describe a duplicated copy as new inserted sequence, so downstream callers report the wrong variant type. Strobealign instead reports a read crossing an SV as separate alignments, one per region of the reference it covers, so an extra copy appears as a second alignment over the same region, a pattern SV callers like SVIM recognize. On simulated reads from the SV benchmark HG002 genome, strobealign identified duplications more completely and more precisely than the widely used aligners while keeping insertion accuracy high. It also gave the best overall SV detection, and mapped a human genome in less than half of other popular tools time.

---

57. From samples to results: BioEngine for browser-based bioimage AI

Bioimage analysis, BioImage Model Zoo, Foundation models, Smart microscopy, Core facility

Mechtel, Nils, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden; Dettner Källander, Hugo, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden; Gómez-de-Mariscal, Estibaliz, NOVA Institute for Medical Systems Biology (NIMSB), Lisboa, Portugal; Beuttenmüller, Fynn, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany; AI4Life Consortium, Horizon Europe Programme https://ai4life.eurobioimaging.eu/; Ouyang, Wei, Science for Life Laboratory, KTH Royal Institute of Technology, Stockholm, Sweden.

Biologists want results, not files. Yet most bioimage AI lives behind Python environments, GPU drivers, and command lines that few microscopists have time to learn, leaving data unanalysed. BioEngine is an open-source execution layer that brings the BioImage Model Zoo and foundation models such as Cellpose-SAM into the browser. A microscopist or facility user uploads their images and receives results in the same session, with no local installation. BioEngine apps let labs

and facilities deploy custom analysis workflows with their own web interfaces, each adapted to a specific modality or sample type. The same engine powers model screening, browser-based fine-tuning, and live inference for smart microscopy, on hardware from a laptop to an HPC cluster. A community deployment at https://bioimage.io already lets users run any BioImage Model Zoo model and fine-tune Cellpose-SAM in the browser. We show how a biologist or facility can ship their own analysis on the same infrastructure.

---

58. Transcriptionally defined AML cell states associate with treatment response and microenvironmental remodeling

precision medicine, leukemic stem cells, cell states

Nona Struyf, Karolinska Institutet, Leonard Hartmanis, Karolinska Institutet, Lucia Rico Pizarro, Karolinska Institutet, Albin Österroos, Uppsala University Hospital, Anna Bohlin, Karolinska Institutet, Sofia Bengtzén, Karolinska Institutet, Sören Lehmann, Karolinska Institutet, Olli Kallioniemi, Karolinska Institutet and Tom Erkers, Karolinska Institutet

Acute myeloid leukemia (AML) is heterogeneous, and therapy resistance is often attributed to leukemic stem cells (LSCs). We integrated scRNA-seq and CITE-seq from diagnostic AML samples (n=6) with bulk RNA-seq (n=448), ex vivo drug sensitivity (n=124), plasma proteomics (n=38), clinical data (n=183), and longitudinal single-cell datasets to define functional AML cell states and their therapeutic and microenvironmental associations. Nine conserved states were identified spanning progenitor-like, stromal-like, monocytic, and mixed stromal/monocytic programs, with stemness distributed across multiple states. Distinct states showed differential drug sensitivity, including convergent BH3-mimetic resistance in monocytic states and selective Wee1/PLK1 inhibitor sensitivity in stress-adapted progenitors. This stress-adapted state associated with adverse clinical features and expanded after venetoclax-based therapy. Monocytic states formed immunosuppressive communication hubs, while stromal and progenitor states engaged survival and cytoprotective signaling. These findings establish functional AML states beyond the classical LSC paradigm and identify candidate vulnerabilities for state-directed therapeutic strategies.

---

59. Beyond Cell Counts and Section Planes: Graph-based Modelling and 3D Reconstruction of the Tumour Microenvironment

Survival prediction, 3D reconstruction, Tumour microenvironment, Multiplex immunofluorescence microscopy, Proteomics

Edgren Schüllerqvist, Olle, Uppsala University; Baumann, Jens, Uppsala University; Ma, Junming, Uppsala University; Nordling, Love, Uppsala University; Yu, Hui, Uppsala University; Strell, Carina, Uppsala University; Lindblad, Joakim, Uppsala University; Mezheyeuski, Artur, Vall d'Hebron Institute of Oncology; Micke, Patrick, Uppsala University; Sladoje, Nataša, Uppsala University

The tumour microenvironment has emerged as a promising source of prognostic biomarkers. Fully leveraging its potential requires methods that capture complex interactions between cell types. Moreover, these interactions take place in three dimensions, yet tissue is routinely analysed in single 2D sections, and volumetric imaging remains slow, costly and not always compatible with FFPE clinical material. We address these challenges in two research directions based on multiplex immunofluorescence imaging of tissue microarray (TMA) cores.

To capture complex cell interactions, we developed a hierarchical graph neural network-

based framework that predicts clinical outcomes from 2D cell graphs. In a retrospective cohort of 298 non-small cell lung cancer patients, our model stratifies patients into risk groups with significantly different survival (HR 1.87, $p < 0.001$), extracting prognostic information beyond cell densities and clinical cancer stage. Our ongoing work refines the graph-based methodology and scales training to a multi-centre, pan-cancer collection of more than 3,000 patients, towards robust and generalisable risk stratification.

To quantify how well single sections represent the tissue, we reconstruct FFPE TMA cores in 3D by aligning consecutive sections with dense deformable registration. Across 30 lung and breast cancer cores (20 sections each), single 2D sections give a systematically biased picture of the cellular landscape captured in 3D, overestimating nearest-neighbour distances and underestimating neighbourhood entropy and cell contacts.

As graph-based modelling extends naturally to 3D, the two directions can be combined to enable outcome prediction from reconstructed tissue.

---

60. Improved interpretation of pathogenic variants by pediatric transcript expression from developmental GTEx

genomics splicing rare-disease

Philipp Rentzsch1, Winona Oliveros1, Josie Gleeson1, Laura Domenech2 , François Aguet2 , Kristin Ardlie2 , The developmental GTEx consortium, Tuuli Lappalainen1,3, Sanna Gudmundsson1,2,4 . 1 Science for Life Laboratory & Department of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden 2 Broad Institute of MIT and Harvard, Cambridge, MA, USA 3 New York Genome Center, New York, NY, USA 4 Department of Molecular Medicine and Surgery, Karolinska Institutet, Stockholm, Sweden.

Genomic variant interpretation has been shown to improve from filtering variants in non-expressed exons. This is implemented in the widely used per-base expression metric “proportion expressed across transcripts” (pext) calculated from GTEx multi-tissue RNA-seq data. However, this data from adult donors creates a blind spot for variants in isoforms that are exclusively or predominantly expressed during early stages of human development. Consequently, some established pathogenic variants currently fall into “low pext” regions.

Here, we expand the pext framework using the newly available developmental GTEx (dGTEx) data from infants aged 0-3 years and adolescents aged 15-18 years. Analyzing 3,320,226 clinically relevant SNVs from ClinVar, we show that dGTEx pext scores are highly correlated to pext from adults ($R2 = 0.88$). However, for 7.0% of pathogenic/likely pathogenic and 4.4% of all ClinVar variants infant pext is above the commonly used threshold of 0.2 for putative clinical relevance, while adult pext falls under the same threshold or is not captured. While much of these gains are driven by technical improvements in the RNA sequencing protocol, 24% of these variants can be considered expressed only in the pediatric and not in the adolescent age group. Projecting our findings to fetal data from non-human primates suggests additional transcript expression differences and increasing usefulness of pext as a variant annotation tool when considering prenatal expression.

Our results underscore the necessity of incorporating functional data from different developmental stages into variant annotation. An updated pext metric offers an opportunity to refine diagnostic tools for interpreting Mendelian disorders.

61. Decoding treatment resistance in Acute Myeloid Leukaemia (AML)
Acute myeloid leukaemia, Single-cell RNA sequencing, Drug resistance, Transcriptional regulation

Qianyue Li, Ilona Szczot, Antonio Lentini - SciLifeLab & Linköping University, Sweden

Acute myeloid leukaemia (AML) is an aggressive haematological malignancy with poor longterm survival. Development of treatment resistance remains a major cause of mortality in the disease. However, treatment resistance frequently occurs without acquisition of new driver mutations, implicating transcriptional reprogramming and shifts in cellular state as contributors to treatment resistance. Investigating these changes has been challenging due to the heterogeneity of AML, limiting inference of transcriptional changes in malignant cells using conventional bulk RNA sequencing. Although single-cell RNA sequencing (scRNA-seq) offers a cell-type-resolved alternative, paired diagnosis–relapse single-cell datasets remain rare, and computational methods for isolating relapse-specific transcriptional programs from those single-cell data have not previously been applied in AML treatment resistance.
Here, I address these gaps using paired scRNA-seq data from four AML patients sampled at diagnosis and relapse/refractory timepoints. Cell type annotation, genotype-supported sample identity verification, and malignant-like compartment definition were performed to establish a reliable analytical foundation. Critically, genotype-based analysis identified a likely sample mislabelling in one patient sample. Cell-type-resolved differential expression confirmed that relapse-associated transcriptional changes were concentrated in the malignant-like myeloid compartment. A deep generative model that mathematically separates variation shared across conditions from variation specifically enriched in one condition, was then applied to the confirmed AML diagnosis–relapse pair to extract shared and relapse-enriched transcriptional variation. Pseudotime analysis of the shared latent space recovered a biologically coherent myeloid differentiation trajectory, whereas model-based differential expression of the salient latent space yielded a 51-gene relapse-associated signature.

External validation in diagnostic samples from the BEAT-AML patient cohort (n = 450) demonstrated that high signature activity at initial diagnosis was significantly associated with shorter overall survival. When combined with patient risk classification, the signature provided additional prognostic stratification within the clinically heterogeneous intermediate risk group, separating patients with markedly different outcomes. These findings establish a framework for extracting clinically relevant relapse-enriched transcriptional programs from longitudinal single-cell data and demonstrate that such signatures can complement existing genomic risk classifiers in AML. As larger longitudinal single-cell cohorts become available, this framework provides a scalable basis for identifying recurrent relapse-associated programs across genetically diverse AML patients.

---

62. SciLifeLab Cellular & Molecular Imaging Platform
imaging, infrastructure, microscopy, structure

Blom, Hans, KTH; Brismar, Hjalmar, KTH; Camacho, Rafael, GU; Carroni, Marta, SU; Coincon, Mathieu, SU; Fernandez-Rodriguez, Julia, GU; Gonnin, Lorene, UmU; Hauff, Nils, UmU; Henriksson, Sara, UmU; Howard, Rebecca J, KTH; Lundmark, Richard, UmU; Phan, Thi Ngoc Nhu, GU; Sandblad, Linda, UmU; Sandin, Sara, UmU; Timpert, Gerrit, SU; von Heijne, Gunnar, SU

Cellular and molecular imaging enables the visualization of biological systems from atomistic to tissue levels. Typical methods involve the transmission, absorption, reflection, or refraction of a focused beam of visible light or electrons by a microscopic sample of interest. In the life sciences, such samples may be biological macromolecules, cells, small organs or organisms of interest to biomedical or planetary biology research. At SciLifeLab, imaging staff assist users from academic, industry, and clinical settings to identify and implement relevant techniques for sample preparation and for the collection, processing, analysis, and sharing of imaging data, with a focus on methods and instruments not generally accessible to individual research groups or local facilities. Key SciLifeLab imaging services include cryogenic electronic microscopy (cryo-EM) and tomography (cryo-ET), super-resolution and other advanced methods in fluorescence microscopy, focused ion-beam milling scanning electron microscopy (FIB-SEM), correlative array tomography (CAT) and nanoscale secondary ion mass spectrometry (nanoSIMS), along with a range of imaging modalities in support of structural biology and spatial omics.

---

63. Promoter-proximal pausing defines distinct modes of stress-response transcription across multicellular organisms

Promoter-proximal pausing, PRO-seq, Nextflow, transcription

Aktay-Kiselman, Serhat, KTH, Viharvaara, Anniina, KTH

Promoter-proximal pausing is a central regulatory step in metazoan transcription, but how it varies across multicellular organisms and shapes stress responses has not been examined within a unified analytical framework. We developed TrackTx, a GMM-based nascent transcription pipeline with spike-in normalization, and applied it to PRO-seq and GRO-seq datasets from H. sapiens, D. melanogaster, and A. thaliana across stress and perturbation conditions.

Median pausing indices (PI) differ by more than 10-fold across species: highest in human K562 cells, intermediate in Drosophila, and ~15-fold lower in Arabidopsis. In K562 cells under celastrol-induced proteotoxic stress, housekeeping gene promoters continue to accumulate Pol II as gene body output declines, consistent with genome-wide impairment of pause release rather than gene-specific repression. Pre-stress PI at heat shock protein genes quantitatively predicts subsequent induction magnitude, supporting a graded model of transcriptional poising.

In Drosophila, NELFe depletion shows that NELF restrains bidirectional Pol II initiation genome-wide. Removal of NELFe nearly doubles divergent transcription sites (10,521 to 20,411) while reducing median PI, suggesting that NELF influences initiation breadth as well as pause duration.

In Arabidopsis, where promoter-proximal pausing is largely absent, TrackTx identifies ~2,200 divergent transcription sites per condition, including sites at unannotated loci. Prior heat stress increases the number of active sites by 14% without a corresponding shift in PI, indicating that stress memory-associated regulation operates through changes in initiation rather than pause release.

Across all three organisms, pausing prevalence is associated with distinct patterns of transcriptional response to stress: systems with strong pausing show changes consistent with altered pause release, whereas those with weak pausing show differences in initiation

site usage.

---

64. Spatial pattern based peak selection for high-dimensional mass spectrometry imaging
Mass Spectrometry Imaging, Multimodal Image Registration

Masood, Sharjeel, Uppsala University and Sintorn, Ida-Maria, Uppsala University

Registering and analyzing Mass Spectrometry Imaging (MSI) datasets can be a computationally expensive process, due to their large file sizes. However, many m/z peaks contain noisy or uninformative data. By introducing a preprocessing step to isolate m/z peaks with meaningful spatial patterns, we can filter out background noise. This can improve the registration speeds and limits the post-registration analysis to only the m/z values with interesting spatial patterns. We use a combination of Moran's I and HDBSCAN as our pre-processing framework. Moran's I identifies m/z values with non-random spatial patterns, while HDBSCAN groups m/z images with similar spatial patterns into clusters. This results in a small number of clusters that need to be analyzed from the MSI sample with 100000+ m/z values

---

65. (G)Host plants of tropical caterpillars
network; interaction; host repertoire

Authors: Holm, Sille1,2,, Freitas, André V. L.3, Alexandersson Ros, Karin1,2, Bovolenta, Martina1,2, Schneider, Katharina1,2 and Braga, Mariana P1,2 1 Department of Ecology, Swedish University of Agricultural Sciences, SLU, Uppsala, Sweden 2 Science for Life (SciLifelab), Swedish University of Agricultural Sciences, Uppsala, Sweden 3 Departamento de Biologia Animal and Museu de Zoologia, Instituto de Biologia, Universidade Estadual de Campinas, Campinas, Brazil *Corresponding author: Sille Holm (sille.holm@slu.se)*

Studying global ecological interaction patterns is becoming increasingly sophisticated by integrating phylogenies and advanced mathematical modelling, yielding increasingly powerful insights into the past, present, and future of species and their ecological interactions. However, all of these methods are haunted by a persistent problem: the scarcity of empirical input data for tropical species. Many tropical species remain little more than shadows in the dataset, leaving substantial uncertainty in the reported patterns despite the sophisticated methods.

Butterflies and moths (Lepidoptera) are among the most diverse groups of plant-eating (herbivorous) insects in terrestrial ecosystems. Yet, for most tropical species, we still do not know which plants their juvenile stages, known as the caterpillars, feed on. Both the caterpillars and their host plants are like ghosts hidden deep within tropical forests, rarely observed and even more rarely recorded in scientific datasets.

Uncovering these hidden interactions is laborious, requiring both field observations and extensive laboratory rearing of caterpillars to validate those observations. Nevertheless, we are dedicated to busting these data ghosts. This poster takes you to the tropical frontlines of Lepidoptera research, where ghost interactions gradually become tangible data points that improve our ability to understand present-day global herbivore-plant interaction patterns, reconstruct past ecological networks, and predict how these interactions may change in the future.

66. Safe Colab enables AI agent–guided zero-trust analysis and federated learning on sensitive biomedical data

*Hugo Dettner Källander1†, Joanna Hård1†, Simin Zhang1, Nils Mechtel1, Adam Ameur2, Emre Sören3, Wei Ouyang1**

Biomedical discovery increasingly depends on clinical, genomic and multi-omic datasets that cannot readily be shared across institutional boundaries. Here we present Safe Colab, an AI agent–guided zero-trust framework for collaborative analysis of biomedical data without transferring the underlying records. In Safe Colab, natural-language requests submitted by analysts for specific data-analysis objectives are translated into executable code and run locally at the data-owner site, while a two-stage Guardian Agent reviews both submitted code and released outputs against data-owner-defined sensitivity contracts.

We evaluated Safe Colab across transcriptomic, pharmacogenomic and clinical machine-learning workflows. In lung adenocarcinoma RNA-seq, the system enabled differential expression analysis while returning only approved aggregate outputs. In Swedish population pharmacogenomics, we deployed a public Safe Colab portal over the SweGen resource that exposes governed analysis of 20,179 clinically relevant variants across 12 chromosomes without exposing the underlying records. In a simulated two-site sepsis study, the same guardian-audited interface supported federated learning across hospital-partitioned cohorts of 20,336 and 20,000 intensive-care stays, yielding held-out AUCs of 0.901/0.915 for a multilayer perceptron and 0.920/0.933 for an FT-Transformer across the two sites. All eight crafted policy-violating submissions in an adversarial validation suite were blocked by the Guardian review pipeline. These results establish Safe Colab as a practical framework for governed, multi-institutional analysis and model development on sensitive biomedical data without direct data sharing.

---

67. Conformal prediction as a solution to distribution shifts in medical image analysis

Conformal prediction, cancer AI, digital pathology, distribution shift

Boman, Sol Erika, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden; Mulliqi, Nita, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden; Liu, Luana Xuan, Department of Medical Epidemiology and Biostatistics, SciLifeLab, Karolinska Institutet, Sweden;Eklund, Martin, Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Sweden; Kartasalo, Kimmo, Department of Medical Epidemiology and Biostatistics, SciLifeLab, Karolinska Institutet, Sweden

Before medical AI systems can be widely deployed in clinics, there is a need for uncertainty quantification and improved handling of distribution shifts. Conformal prediction (CP) is an uncertainty quantification method that is highly useful but requires datasets to exhibit exchangeability, something that is broken by distribution shifts. In this work, we employ CP on real-world pathology image data to investigate distribution shifts' impact on the validity of CP. We find that the main driver for loss of exchangeability is disagreements between pathologists, rather than medical differences between patients.

68. The Instrument That Refuses: Physical Constraint Checks for AI-Agent Laboratory Automation

AI agents; laboratory automation; liquid handling; physical constraints; safety

AI agents can generate laboratory robot protocols, but physical constraints require explicit enforcement. We study AI-agent control of a Hamilton liquid-handling robot through constraint checks, offline feedback-driven protocol correction, audits of existing checks, and evaluation of visual sensing limitations. In an offline named-refusal loop, 29 of 30 initially violating protocols were corrected. A paired comparison of diagnostic advice and diagnostic refusal corrected 30 of 30 protocols in each arm, showing no additional correction benefit from refusal framing in this test. In a separate water-only hardware demonstration, a client-side gate refused an incorrect channel-to-row mapping before dispatch; a corrected protocol then completed 22 steps. This demonstration covers one constraint and does not establish complete mediation. The work highlights the value of actionable feedback and checks placed on the command path, while distinguishing visual observations from decisions about whether an action is safe.

---

69. Genomics platform

Tuuli Lappalainen

---

70. Somatic mitochondrial variants in blood are associated with increased incidence of hematologic malignancies

Clonal Hematopoiesis; Hematological Cancers; Mitochondria DNA; Genomics

Peniche, Tânia 1; - Huang, Yajuan 1; - Niroula, Abhishek 1,2; 1 - Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden; 2 - Science for Life Laboratory, University of Gothenburg, Gothenburg, Sweden

Clonal hematopoiesis of indeterminate potential (CHIP), defined by the clonal expansion of blood cells carrying pathogenic cancer driver mutation, is associated with an elevated risk of hematological malignancies. Yet only a minority of individuals with CHIP develop a hematological malignancy during their life. Novel markers of malignant transformation could facilitate risk stratification of CHIP-positive individuals. Recently, heteroplasmic mitochondrial DNA (mtDNA) variants have been linked to increased risk of hematological malignancies. While the majority of heteroplasmic mtDNA variants are assumed to be somatic, the landscape of inherited and acquired mtDNA variants in blood remains unexplored. In this study, we characterized somatic and germline mtDNA heteroplasmic variants in blood using paired whole genome sequencing data from blood and solid tumor samples in Genomics England 100 thousand genomes project and from maternally related individuals in UK Biobank. As expected, the frequency of somatic variants increased with age, but the frequency of germline variants remained stable across age groups. Furthermore, somatic mtDNA variants significantly co-occurred with CHIP and were associated with an elevated risk of hematologic malignancies. In contrast, the germline mtDNA variants were not associated with CHIP or incident hematologic malignancies. While both somatic

mtDNA variants and CH mutations were independently associated with higher incidence of myeloid and lymphoid malignancies, the presence of both somatic mtDNA variants and CH mutations amplified the risk significantly. These findings provide new insights into the landscape of somatic mtDNA variants in normal blood samples and highlight the potential utility of mtDNA variants as novel markers for detecting both myeloid and lymphoid clonal hematopoiesis and hematological malignancies.

Acknowledgements: The study was supported by funding from the SciLifeLab & Wallenberg Data Driven Life Science Program (KAW 2020.0239), Gunnar Nilsson Cancerstiftelse (GN-2022-14 - (266)), Swedish Cancer Foundation (22.2362Pj), and the Swedish Research Council (2023-03131). This research was made possible through access to the National Genomic Research Library (NGRL), a secure database and bioresource managed by Genomics England in partnership with the NHS. This research was partly conducted using the UK Biobank Resource under Application Number 186376.

---

71. The United Structures of Pathology

Spatial transcriptomics; Pathological structures; Tissue organization; Joint analysis; Open-access data

Taobo Hu; Mengping Long; Mats Nilsson

Spatial transcriptomics links molecular profiles to tissue architecture, creating new opportunities to study pathological structures in their spatial context. However, observations from open-access datasets are often considered separately, limiting comparisons across studies. We analyzed diverse open-access spatial transcriptomics datasets and identified a broad range of pathological structures. Bringing these structures together allows their spatial organization and molecular characteristics to be examined across datasets within a shared analytical perspective. This structure-centered approach provides a basis for investigating similarities and differences in pathological tissue organization while retaining the context of each study. Our findings support the use of pathological structures as a common unit for the joint analysis of spatial transcriptomics data. Together, this work connects observations from diverse open-access resources and lays the groundwork for a more unified understanding of pathology through spatial molecular data.

---

72. Genomic surveillance of the 2025-2026 anthrax outbreak in Kruger National Park, South Africa

Bacillus anthracis, Outbreak surveillance, Genomic Epidemiology, Phylogenomics

Ramnath, Vignesh, Umeå University, Sweden; Matle, Itumeleng, Onderstepoort Veterinary Research, South Africa; Carroll, Laura, Umeå University, Sweden

Anthrax is a severe zoonotic disease caused primarily by Bacillus anthracis, a pathogen of major public health and biosafety concern. Its environmentally persistent spores can drive recurrent outbreaks in susceptible wildlife and livestock, particularly in endemic regions of sub-Saharan Africa.

We investigated a recent anthrax outbreak among wildlife in Kruger National Park (KNP), South Africa, using whole-genome sequencing (WGS) and phylogenomic approaches. KNP outbreak isolates were characterized using bacterial WGS analysis, including assembly,

annotation, quality assessment, typing, classification, and virulence-factor detection. We compared the KNP outbreak genomes with publicly available, global B. anthracis genomes and associated sample metadata, to place the outbreak within its broader genomic and epidemiological context. Phylogenomic methods based on single-nucleotide differences across shared genomic regions resolved relationships at progressively finer scales, from global B. anthracis diversity to cluster, clade, and outbreak levels.

All outbreak-associated isolates were placed within the expected African cluster and the same clade, where they grouped with closely related historical South African and KNP isolates. Within the outbreak, most isolates formed a tight central group, while several basal or nested sublineages revealed limited internal heterogeneity. Together, these findings support a closely related outbreak-associated population within a broader local genomic background, while indicating that genomic relatedness alone does not fully distinguish outbreak-associated from historical isolates.

This multi-resolution phylogenomic approach contextualized the KNP outbreak within broader African B. anthracis diversity and revealed a closely related but heterogeneous outbreak-associated population.

---

73. Fake It Till You Make It: Realistic Emulation of Cryo-EM Micrographs Using GANs
cryo-EM; synthetic data generation; generative adversarial networks; benchmarking

Kuksin Vsevolod, Linköping University; Forsberg Björn, Linköping University

Single-particle cryo-electron microscopy (cryo-EM) reconstruction aims to recover three-dimensional macromolecular structures from extremely low signal-to-noise ratio images. Development and benchmarking of these methods would benefit from the complementary forward problem: generating realistic cryo-EM datasets from known structures at known positions and orientations, providing direct ground truth for reconstruction and particle-picking methods.

Real experimental micrographs, however, contain complex contributions from vitreous ice, imaging effects, and spatially correlated noise that are not fully captured by either physics-based simulators or current generative approaches. Moreover, there is no established set of metrics for defining simulation realism across different downstream applications.

We present GroundControl (GC), a framework for benchmarking cryo-EM micrograph generators against experimental and simulated data. We use GC to characterize limitations of existing simulation strategies and guide the development of EMulation, a generative architecture for producing realistic micrographs from atomic models. We further investigate adversarial training strategies and complementary loss functions aimed at improving training stability, reducing generator bias, and controlling non-physical artifacts in the generated images.

---

74. Berzelius: Sweden's National AI Supercomputer for Data-Driven Science
Berzelius; Artificial intelligence; GPU computing; High-performance computing; Data-driven life science

Chaki, Soumi; Gu, Xuan. National Academic Infrastructure for Supercomputing in Sweden (NAISS)

Berzelius is Sweden's national supercomputer for Artificial Intelligence (AI) and data-driven science. Donated by the Knut and Alice Wallenberg Foundation (KAW) and operated by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), Berzelius provides researchers with large-scale computational performance, modern GPU hardware, and an optimized software environment for AI research and high-throughput data analysis.

With 110 NVIDIA DGX A100 and H200 systems connected through high-bandwidth NVIDIA Quantum InfiniBand, Berzelius delivers up to 720 PFLOPS of theoretical AI performance. Its high-speed solid-state storage, powered by VAST Data and DDN AI400X, supports multi-petabyte data volumes and enables efficient workflows for large-scale training, data processing, and scientific simulations.

Since its launch, Berzelius has supported more than 1,000 research projects across Sweden. A dedicated team of system and application experts supports users through onboarding, consultation, and training. Regular workshops and user events help both new and experienced researchers make effective use of the system, adopt best practices, and optimize their workflows for large-scale AI and data-driven research.

---

75. Identification of new drivers of Clonal Hematopoiesis using Whole-Genome Sequencing

Clonal hematopoiesis; whole-genome sequencing; somatic mutations; driver mutations; hematological malignancy

Yajuan Huang1, Tânia Peniche1, Abhishek Niroula1,2 Affiliations: 1Department of Medical Biochemistry and Cell Biology, Institute of Biomedicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden 2Science for Life Laboratory, University of Gothenburg, Gothenburg, Sweden

Clonal hematopoiesis (CH) is characterized by the expansion of hematopoietic clones carrying somatic alterations. The prevalence of CH increases with age and is associated with a higher risk of blood cancer. Several driver genes recurrently mutated in CH are known, including DNMT3A, TET2, ASXL1, and JAK2. However, driver variants underlying the majority of the blood cell clones remain unknown. To identify new drivers of CH, we analyzed somatic variants in 13,906 whole-genome sequences from the Genomics England 100,000 genomes project. We developed a workflow to deplete germline variants and technical artifacts while enriching for somatic variants. The somatic variants identified using this workflow enabled driver-independent CH inference, which identified CH both with and without known drivers. We identified CH in 24.4% of all individuals, which is 2.7 times higher than based on CH with previously known driver mutations. Somatic variants in these CH clones were subsequently used to identify drivers of CH. In addition to previously known drivers of CH, we identified 13 new candidate genes, which include genes involved in hematopoiesis. We are currently evaluating the role of these new candidates in hematological malignancies. Our driver-independent CH inference workflow facilitates the detection of CH beyond clones carrying known drivers and facilitates the discovery of novel drivers.

---

76. Leveraging mass spectrometry data for predicting chemical persistence and mobility in water

Mass spectra, prediction, persistence, mobility, non-targeted screening

Yingxiao Yan1, May Britt Rian1, Anneli Kruve1,2, Jonathan W. Martin1 1Department of Environmental Science, Stockholm University 2Department of Chemistry, Stockholm University

Despite increasing awareness, the cumulative risk of persistent mobile chemicals to water resources remains difficult to quantify, in large part due to a lack of experimental persistence (P) and mobility (M) data for tens-of-thousands of substances detected in surface waters and municipal wastewater. Here, we present a machine learning framework that leverages data-rich high resolution mass spectral information from non-target analyses to predict chemical persistence and mobility for known and unknown substances. Using publicly available spectra and P/M classifications, models were trained and performance were compared using either spectral embeddings or SIRIUS derived molecular fingerprints. Across multiple machine learning algorithms, fingerprint based models, in which features with higher variable importance reflect molecular structures with known links to P and M, consistently outperformed embeddings, achieving lower balanced error rates. External comparison using hundreds of chemicals from wastewater treatment plant datasets further demonstrated that models based on molecular fingerprints showed lower balanced error rate for persistence prediction using both positive and negative mode spectra, and for mobility prediction using positive mode spectra. Overall, this study demonstrates a rapid approach to screen for persistent and mobile chemicals in water without the limiting steps of identification and experimental testing, thereby supporting feature prioritization in non target screening.

---

77. The distribution and role of species-level low prevalence taxa in the human gut microbiota

low prevalence taxa; gut microbiota; meta-analyses

Yingyu Mu, Department of Clinical Microbiology, Umeå University

The human gut microbiome exhibits a pronounced long-tail distribution, comprising a small number of species consistently detected across individuals and a much larger number of species-level low-prevalence taxa (LPTs) detected in only a subset of individuals. Although LPTs account for a substantial fraction of microbial diversity, their population structure, ecological organization, and functional significance remain poorly understood.

To address this knowledge gap, we integrated metagenomic data from 25 human gut microbiome cohorts spanning diverse geographic regions, lifestyles, and health conditions. Using a framework combining intraclass correlation coefficient (ICC) analysis, centered log-ratio (CLR) transformation, network inference, and machine-learning approaches, we systematically characterized the structure of LPTs and investigated their ecological and population-level patterns.

We identified three major LPTs categories: 184 globally stable LPTs, 192 cohort-specific LPTs, and 6,811 unclassified LPTs requiring further characterization. Network analysis revealed that globally stable LPTs were organized into ten ecological modules, suggesting the existence of conserved interaction networks among rare taxa across human populations. Machine-learning models based on cohort-specific LPTs accurately distinguished industrialized populations, whereas predictive performance was substantially lower for traditional populations, indicating greater ecological heterogeneity and less consistent LPTs assemblages. Ongoing analyses are applying additional supervised and

unsupervised learning approaches to further resolve the large unclassified LPTs fraction.

This work provides a population-scale framework for understanding the structure and ecological roles of LPTs, extends the current focus beyond the core microbiome, and advances our understanding of how rare microbial species contribute to the organization and function of the human gut ecosystem.

---

78. Metabolomics platform

Hans Stenlund